Chapter 7

DOMESTIC COUNTER-TERRORISM

Chapter Overview

As terrorists may be involved in local crimes, state and local law enforcement agencies are contributing to national security by investigating and prosecuting crimes. Terrorism, like gangs and narcotics trafficking, is treated as a type of organized crime in which multiple offenders cooperate to carry out offenses. Information technologies that help find cooperative relationships between criminals and their interactive patterns would also be helpful in analyzing terrorism. Through three case studies in this section, we show how criminal association information can be extracted from large volumes of data and how structural patterns in criminal or terrorist organizations can be discovered.

7.1 Case Study 7: COPLINK Detect

Crime analysts and detectives search for criminal associations to develop investigative leads. However, because association information is not directly available in most existing law enforcement and intelligence databases and manual searching is extremely time-consuming, automatic identification of relationships among criminal entities may significantly speed up crime investigations. COPLINK Detect is a system that automatically extracts criminal element relationships from large volumes of crime incident data (Hauck et al., 2002).

Our data were structured crime incident records stored in TPD databases. TPD's current record management system (RMS) consists of more than 1.5 million crime incident records that contain details from criminal events spanning the period from 1986 to 2004. Although investigators can access the RMS to tie together information, they must manually search the RMS for connections or existing relationships.

We used the concept space approach (Chen and Lynch, 1992) to identify relationships between entities of interest. Concept space analysis is a type of co-occurrence analysis used in information retrieval. The resulting network-like concept space holds all possible associations between terms, which means that the system retains and ranks every existing link between every pair of concepts. In COPLINK Detect, detailed incident records served as the underlying space, while concepts derive from the meaningful terms that occur in each incident. Concept space analysis easily identifies relevant terms and their degree of relationship to the search term. The system output includes relevant terms ranked in the order of their degree of association, thereby distinguishing the most relevant terms from inconsequential terms. From a crime investigation standpoint, concept space analysis can help investigators link known entities to other related entities that might contain useful information for further investigation, such as people and vehicles related to a given suspect. It is considered an example of entity association mining (Lin and Brown, 2003).

Information related to a suspect can direct an investigation to expand in the right direction, but revealing relationships among data in one particular incident might fail to capture those relationships from the entire database. In effect, investigators need to review all incident reports related to a suspect, which can be tedious work. The COPLINK Detect system introduces concept space as an alternative method that captures the relationships between four types of entities (person, organization, location, and vehicle) in the entire database. COPLINK Detect also offers an easy-to-use user interface and allows searching for relationships among the four types of entities.

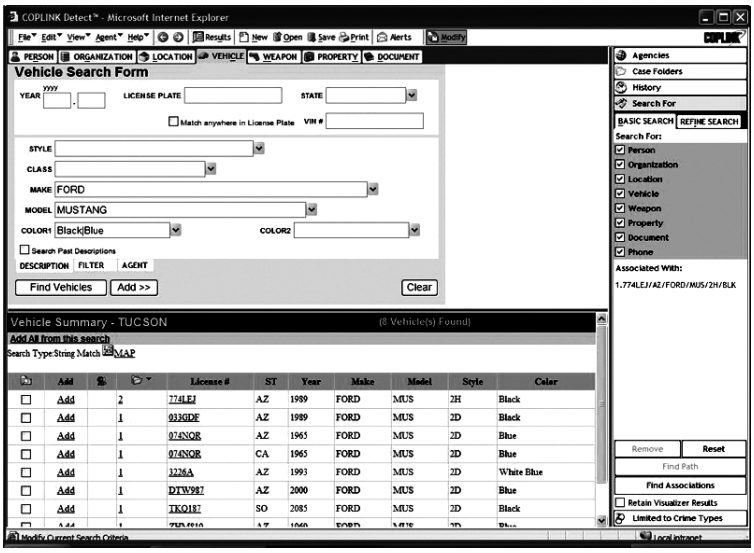

Figure 7-1. COPLINK Detect interface showing sample research results.
Figure 7-1a. COPLINK Detect vehicle search screen. "Vehicle" is one of the information types users can enter as a search term. After adding the search terms to the Associated With box, the user selects the Find Associations button to retrieve associates.

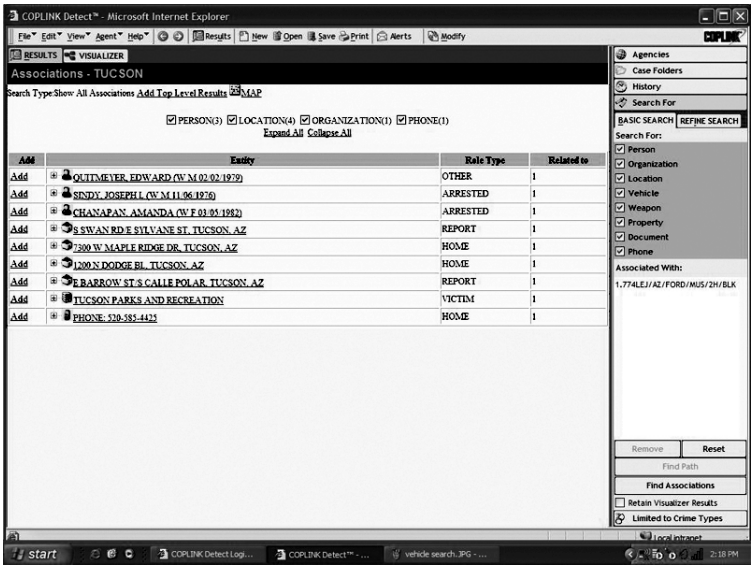

Figure 7-1b. Associations screen. The application can return elements for each of the eight information object types: Person, Organization, Location, Vehicle, Weapon, Property, Document, and Phone. Here the search has return results for three persons, four locations, one organization, and one phone number.

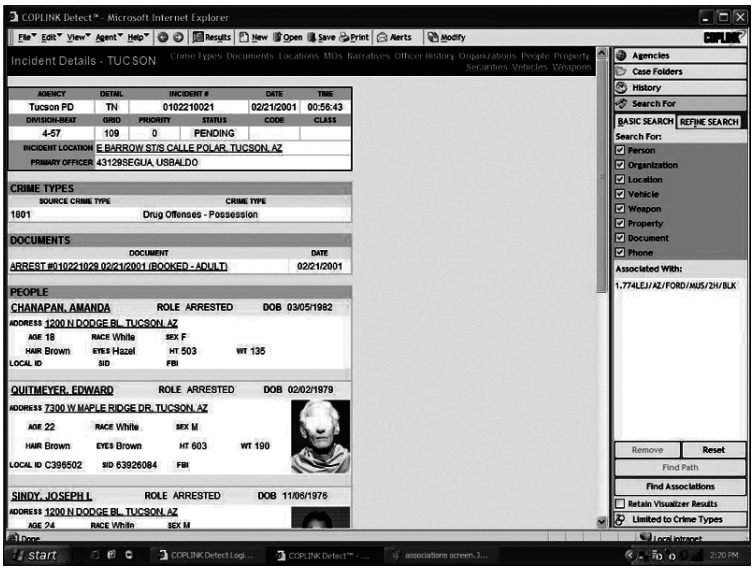

Figure 7-1c. Incident Details screen. In this example, the detective can view the details of one prior incident in the database, including the role and home address of each person involved. (All criminal data shown here are scrubbed.)

Figure 7-1 presents the COPLINK Detect interface showing sample search results of vehicles, relations, and crime case details (Hauck et al., 2002).

We conducted user studies to evaluate the performance and usefulness of COPLINK Detect. Twelve crime analysts and detectives participated in the field study during a four-week period. Three major areas were identified where COPLINK Detect provided improved support for crime investigation:

- *Link analysis*. Participants indicated that COPLINK Detect served as a powerful tool for acquiring criminal association information. They cited its value in helping determine the presence or absence of links between people, places, vehicles, and other entity types in investigating a crime.
- *Interface design*. In general, users reported that they found the COPLINK Detect interface easy to use. Officers noted that the graphical user interface and use of color to distinguish different entity types provided a more intuitive visualization than traditional text-based record management systems.
- *Operating efficiency*. In a direct comparison of 15 searches, using COPLINK Detect required an average of 30 minutes less per search than did a benchmark record management system (20 minutes vs. 50 minutes).

7.2 Case Study 8: Criminal Network Mining

Because organized crimes are carried out by networked offenders, investigation of organized crimes naturally depends on network analysis approaches. Grounded on social network analysis (SNA) methodology, our criminal network structure mining research aims to help intelligence and security agencies extract valuable knowledge regarding criminal or terrorist organizations by identifying the central members, subgroups, and network structure (Xu and Chen, Forthcoming).

Two datasets from TPD were used in the study. (1) A gang network: the list of gang members consisted of 16 offenders who had been under investigation in the first quarter of 2002. These gang members had been involved in 72 crime incidents of various types (e.g., theft, burglary, aggravated assault, drug offense, etc.) since 1985. We used the concept space approach and generated links between criminals who had committed crimes together, resulting in a network of 164 members. (2) A narcotics network: the list for the narcotics network consisted of 71 criminal names. A sergeant from the Gang Unit had been studying the activities of these criminals since 1995. Because most of them had committed crimes related to methamphetamines, the sergeant called this network the "Meth World." These offenders had been involved in 1,206 incidents since 1983. A network of 744 members was generated.

We employed SNA approaches to extract structural patterns in our criminal networks.

- *Network Partition.* We employed hierarchical clustering, namely the complete-link algorithm, to partition a network into subgroups based on relational strength. Clusters obtained represent subgroups. To employ the algorithm, we first transformed co-occurrence weights generated in the previous phrase into distances/dissimilarities. The distance between two clusters was defined as the distance between the pair of nodes drawn from each cluster that was farthest apart. The algorithm worked by merging the two nearest clusters into one cluster at each step and eventually formed a cluster hierarchy. The resulting cluster hierarchy specified groupings of network members at different granularity levels. At lower levels of the hierarchy, clusters (subgroups) tended to be smaller and group members were more closely related. At higher levels of the hierarchy, subgroups are large and group members might be loosely related.
- *Centrality Measures.* We used all three centrality measures to identify central members in a given subgroup. The degree of a node could be obtained by counting the total number of links it had to all the other

group members. A node's score of betweenness and closeness required the computation of shortest paths (geodesics) using Dijkstra's algorithm (1959).

- *Blockmodeling.* At a given level of a cluster hierarchy, we compared between-group link densities with the network's overall link density to determine the presence or absence of between-group relationships.
- *Visualization.* To map a criminal network onto a two-dimensional display, we employed Multi-Dimensional Scaling (MDS) to generate x-y coordinates for each member in a network. We chose Torgerson's classical MDS algorithm (Torgerson, 1952) since distances transformed from co-occurrence weights were quantitative data.

A graphical user interface was provided to visualize criminal networks. Figure 7-2 shows the screenshot of our prototype system. In this example, each node was labeled with the name of the criminal it represented. Criminal names were scrubbed for data confidentiality. A straight line connecting two nodes indicated that two corresponding criminals committed crimes together and thus were related. To find subgroups and interaction patterns between groups, a user could adjust the "level of abstraction" slider at the bottom of the panel. A high level of abstraction corresponded with a high distance level in the cluster hierarchy. Group members' rankings in centrality are listed in a table.

We conducted a qualitative study recently to evaluate the prototype system. We presented the two testing networks to domain experts at TPD and received encouraging feedback:

- *Subgroups detected were mostly correct*. Our domain experts checked and validated the members in each group. These groups had different characteristics with different specialties or crime preferences. We also found that although relationships in our networks were extracted based on crime incidents, they reflected true relationships between criminals such as friendship, kinship, and even conflicts.
- *Centrality measures provided ways of identifying key members in a network.* According to our domain experts, betweenness was a reliable measure to identify gatekeepers between subgroups. However, degree sometimes identified wrong leaders because the criminals with the most connections to others may not always be the leaders. Leaders may be smart enough to hide behind other criminals to avoid police contact.

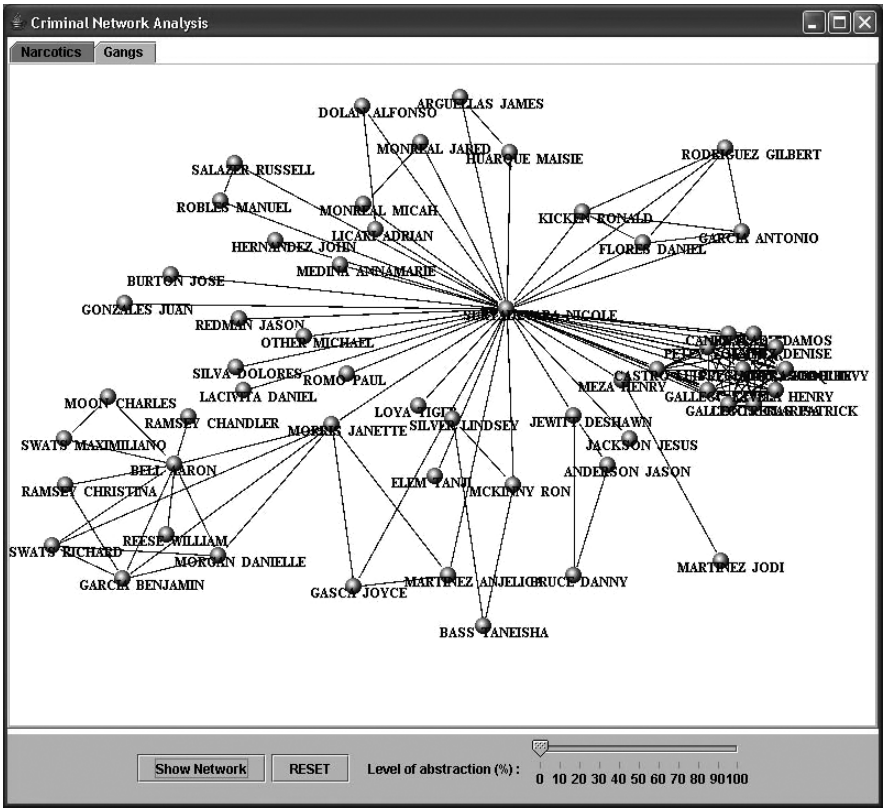

Figure 7-2. An SNA-based system for criminal network analysis and visualization.
Figure 7-2a. A 57-member criminal network. Each node is labeled using the name of the criminal it represents. Lines represent the relationships between criminals.

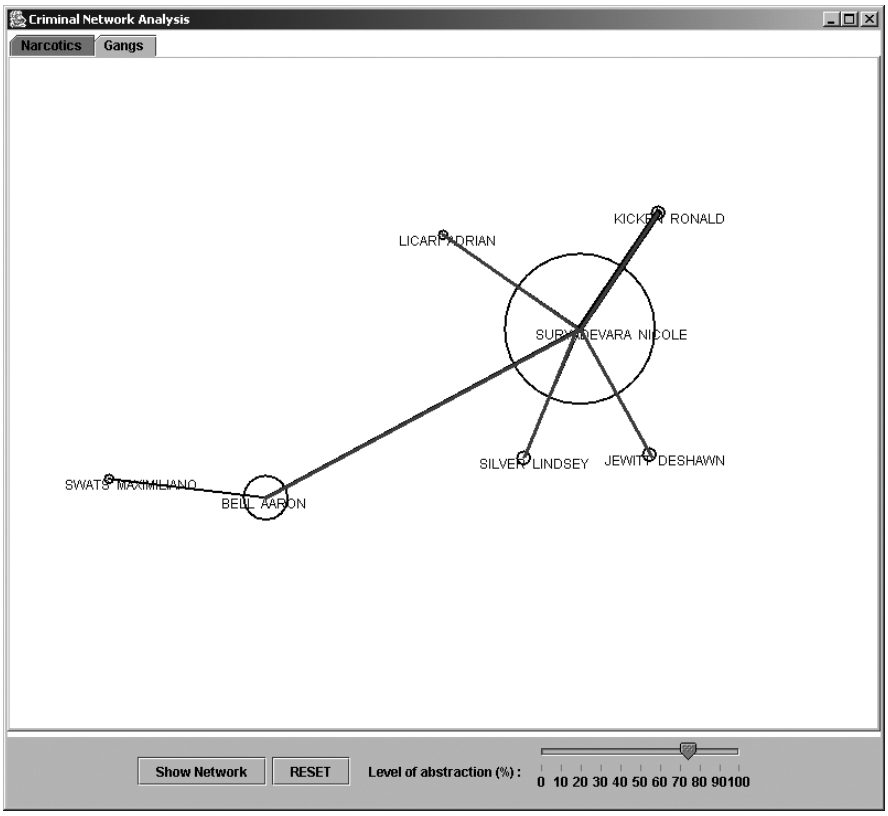

Figure 7-2b. The reduced structure of the network. Each circle represents one subgroup labeled by its leader's name. The size of the circle is proportional to the number of criminals in the group. A line represents a relationship between two groups. The thickness represents the strength of the relationship.

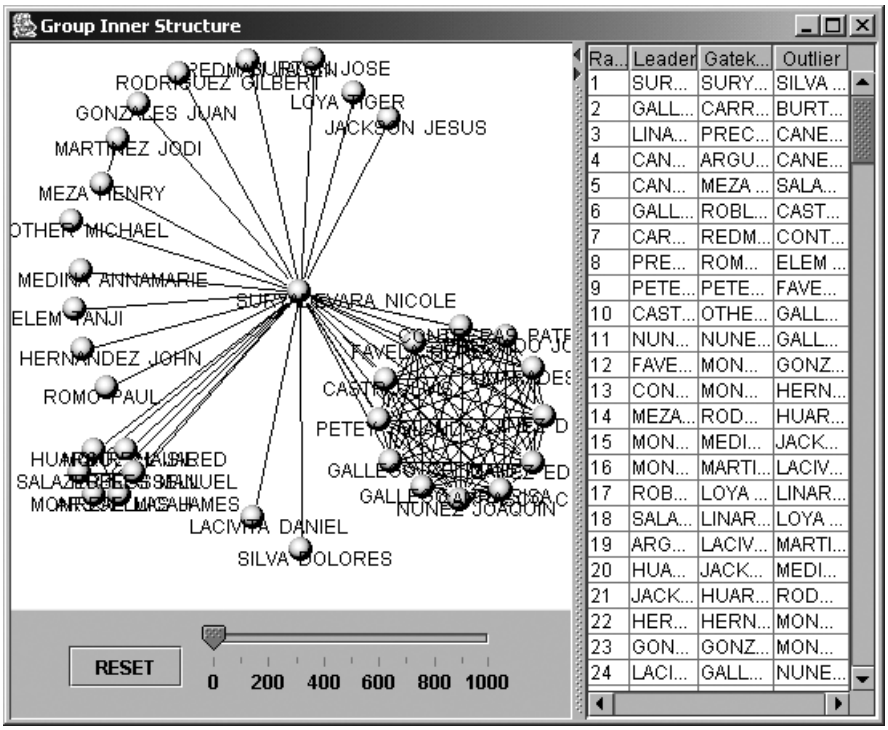

Figure 7-2c. The inner structure of the biggest group (the relationships between group members). Centrality rankings of members in this group are listed in a table at the right-hand side.

- *Interaction patterns identified could help reveal relationships that previously had been overlooked*. Our system could generate the "big picture" for a complex network. As a result some relationships between criminal groups that were overlooked before could become easier to identify.
- *Saving investigation time*. Our domain experts obtained knowledge about the gang and narcotics organizations during several years of work. Using information gathered from a large number of arrests and interviews, they built the networks incrementally by linking new criminals to known gangs in the network and then studying the organization of these networks. Because there was no structural analysis tool available, they did all this work by hand. With the help of our system, they expected substantial time could be saved in network creation and structural analysis.

- *Saving training time for new investigators*. New investigators who did not have sufficient knowledge of criminal organizations and individuals could use the system to grasp the essence of the network and crime history quickly. They would not have to spend a significant amount of time studying hundreds of incident reports.
- *Helping prove guilt of criminals in court*. The relationships discovered between individual criminals and criminal groups would be helpful for proving guilt when presented at court for prosecution.

7.3 Case Study 9: Domestic Extremist Groups on the Web

Although not as well-known as some of the international terrorist organizations, the extremist and hate groups within the United States also pose a significant threat to our national security. Recently, these groups have been intensively utilizing the Internet to advance their causes. Thus, to understand how the domestic extremist and hate groups develop their web presence is very important in addressing domestic terrorism threats. This study proposes the development of systematic methodologies to capture domestic extremist and hate groups' web site data and support subsequent analyses. In this study, we aim to answer the following research questions:

- What are the most appropriate techniques for collecting high-quality web pages of domestic extremist and hate groups?
- What are the systematic procedures for analyzing and visualizing the content of these individual web sites?

We propose a sequence of semi-automated methods to study domestic extremist and hate group content on the web. First, we employ a semi-automatic procedure to harvest and construct a high-quality domestic terrorist web site collection. We then perform hyperlink analysis based on a clustering algorithm to reveal the relationships between these groups. Lastly, we conduct an attribute-based content analysis to determine how these groups use the web for their purposes. Because the procedure adopted in this study is similar to that reported in Case Study 3, Jihad on the Web, we only summarize selected interesting results below.

- *Collection Building:* We manually extracted a set of URLs from relevant literature. In particular, the web sites of the "Southern Poverty Law Center" (SPLC, www.splcenter.org) and the Anti-Defamation League (ADL, www.adl.org) are authoritative sources for identifying domestic extremists and hate groups. A total of 266 seed URLs were identified in SPLC and the ADL web sites as well as in the Google directory. A

backlink expansion of this initial set was performed and the count increased to 386 URLs. The resulting set of URLs is validated by an expert. A total of 97 URLs were deemed relevant. We then spidered and downloaded all of the web documents within the identified web sites. As a result, our final collection contained about 400,000 documents.

- *Hyperlink Analysis:* Using the MDS algorithm (Torgerson, 1952), we visualize the hidden hyperlinked communities among 35 web sites randomly retrieved from our collection. Several communities are identified in the network shown in Figure 7-3. The top-left side of the network shows the "Neo-Confederate" cluster, which mainly consists of the web sites of new confederate organizations in the Southern states. They espouse a separatist ideology, promoting the establishment of an independent state in the south. In addition, they share elements of white supremacy ideas with other non-neo-confederate racist organizations such as the KKK. A cluster of web sites of white supremacists occupies the top-right corner of the network, including: Stormfront, White Aryan Resistance, etc. Christian Identity, Militia, and Eco-Terrorism clusters were also identified.
- *Content Analysis:* We asked our domain experts to review each web site in our collection and record the presence of low-level attributes based on an eight-attribute coding scheme: Sharing Ideology, Propaganda (Insiders), Recruitment and Training, Command and Control, Virtual Community, Propaganda (Outsiders), Fundraising, and Communications. For instance, the web page of "Nation of Islam" contains recordings of the organization's leaders (for their followers). The presence of these recordings contributes to the web site's content richness and is coded under the "Propaganda (Insiders)" attribute. Our web coding scheme is similar in nature to the one developed by Demchak et al. (2000) for coding government web site characteristics.

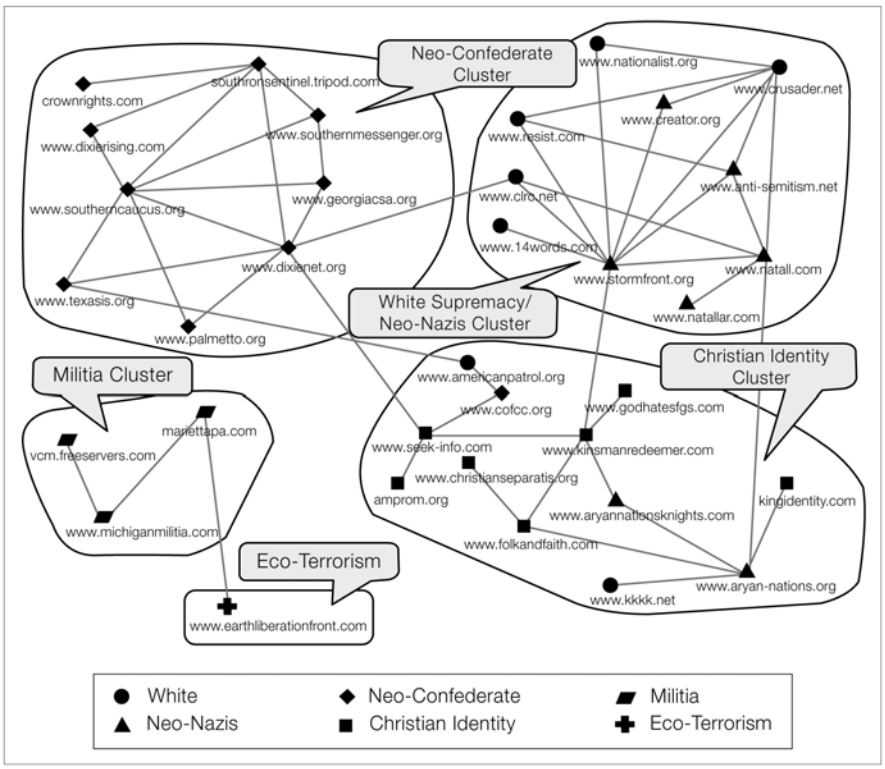

Figure 7-3. Web community visualization of selected domestic extremist and hate groups.

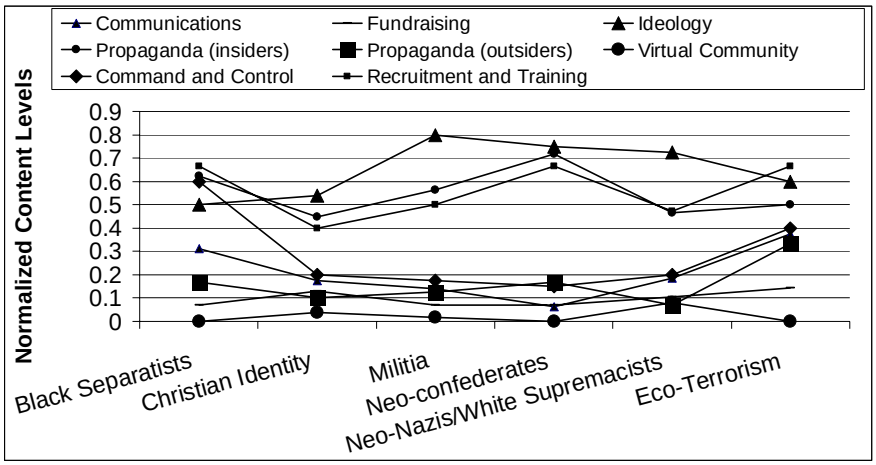

Figure 7-4. Content analysis of selected web sites of domestic extremist and hate groups.

The manual coding of the attributes in a web site takes about 45 minutes. After completing coding for the web sites in our collection, we compared the content of each of the six domestic extremist and hate groups as shown in Figure 7-4. "Sharing Ideology" is the attribute with the highest frequency of occurrence in these web sites. It encapsulates all communication media devoted to portraying the goals of the terrorist group, defining its general policies, and presenting the foundational ideology. In addition, "Propaganda (Insiders)" and "Recruitment and Training" are widely used by all groups on their web sites.

Another interesting observation is the low presence of "Propaganda (Outsiders)," with the exception of Eco-terrorism/Animal Rights groups, which are considered to have a much wider audience than the racist groups, who have a more targeted audience. Much research is still needed for systematic understanding of how domestic extremist and hate groups use the web to promote their causes.

7.4 Case Study 10: Topological Analysis of Dark Networks

Large-scale networks such as scientific collaboration networks, the World Wide Web, the Internet, electric power grids, food webs, and metabolic networks are surprisingly similar in topology (e.g., power-law degree distribution), leading to a conjecture that complex systems are governed by the same self-organizing principle (Albert & Barabasi, 2002). Although the topological properties of these networks have been discovered, the structures of dark (covert, illegal) networks are largely unknown due to the difficulty of collecting and accessing reliable data (Krebs, 2001). Do dark networks share the same topological properties with other types of networks? Do they follow the same organizing principle? How do they achieve efficiency under constant surveillance and threat from authorities? We report in this study the topological properties of several covert criminal- or terrorist-related networks. We hope not only to contribute to general knowledge of the topological properties of complex systems in a hostile environment but also to provide authorities with insights regarding disruptive strategies.

Three models have been employed to characterize complex networks: random graph model, small-world model, and scale-free model (Albert & Barabasi, 2002). Most complex systems are not random but are governed by certain organizing principles encoded in the topology of the networks. A small-world network has a significantly larger clustering coefficient than its random model counterpart while maintaining a relatively small average path

length. The large clustering coefficient indicates that there is a high tendency for nodes to form communities and groups. Scale-free networks, on the other hand, are characterized by the power-law degree distribution, meaning that while a large number of nodes in the network have just a few links, a small fraction of the nodes have a large number of links. It is believed that scale-free networks evolve following the self-organizing principle, where growth and preferential attachment play a key role for the emergence of the power-law degree distribution.

We studied the topology of four covert networks: the Global Salafi Jihad (GSJ) terrorist network (Sageman, 2004), a narcotics trafficking criminal network (Xu & Chen, 2003; Xu & Chen, Forthcoming) whose members mainly deal with methamphetamines, a gang criminal network, and a terrorist web site network (Chen *et al.*, 2004). The 366-member GSJ network was constructed based entirely on open source data but all nodes and links were examined and carefully validated by a domain expert. The "Meth World" consists of 1,349 criminals who were involved in methamphetamine-related crimes in Tucson, Arizona, between 1985 and 2002. The gang network consists of 3,917 criminals who were involved in gang-related crimes in Tucson between 1985 and 2002. In the two criminal networks, two members are connected if they committed at least one crime together. Based on reliable governmental sources, we also identified 104 web sites created by four major international terrorist groups. Hyperlinks were used as between-site relations.

Each network contains many small components and a single giant component (see Table 7-1). We focused only on the giant component in these networks and performed topology analysis and robustness analysis. We found that all these networks are small worlds (see Table 7-2). The average path lengths and diameters of these networks are small with respect to their network sizes. Thus, a terrorist or criminal can connect with any other member in a network through just a few mediators. In addition, these networks are quite sparse with very low link density. These two properties have important implications to the efficiency of transmission of goods and information within the networks. Because the risk of being detected by authorities increases as more people are involved, the small path length and link sparseness can help lower risks and enhance efficiency. In addition, we calculated the path length of a node to a central node, a measure which is called "Erdős number" in the network of mathematicians. This measure is also related to the closeness centrality, defined as the total path length from a specific node to all other nodes in a network. We found that members in the criminal and terrorist networks are extremely close to their leaders. The terrorists in the GSJ network are on average only 2.5 steps away from bin Laden, meaning that bin Laden's command can reach an arbitrary member

through only two mediators. Similarly, the average path length to the leader in the Meth World is only 3.9. Such a short chain of command means communication efficiency. However, special attention should be paid to the Dark Web. Despite its small size (80), the average path length is 4.70, larger than that (4.20) of the GSJ network, which has almost 9 times more nodes. Since hyperlinks help visitors navigate between web pages, and because terrorist web sites are often used for soliciting new members and donations, the relatively long path length may be due to the reluctance of terrorist groups to share potential resources with other terrorist groups.

The other small-world topology, high clustering coefficient, is also present in these dark networks. The clustering coefficients of these four networks are significantly higher than those of random graph counterparts. Previous studies have also shown the evidence of groups and teams in these networks. In these groups and teams, members tend to have denser and stronger relations with one another. The communication between group members becomes more efficient, making a crime or an attack easier to plan, organize, and execute.

In addition, these dark networks are scale-free systems. The three human networks have an exponentially truncated power-law degree distribution (see Table 7-1 and Figure 7-5). Different from other types of networks whose exponents usually are between 2.0 and 3.0, the exponents of dark networks are fairly small. The degree distribution decays much more slowly for small degrees than for that of other types of networks, indicating a higher frequency for small degrees.

Table 7-1. The statistics and parameters in the exponentially truncated power-law degree distribution of the dark networks**.**

	GSJ	**Meth World**	**Gang Network**	**Dark Web**
Number of Nodes	366	1349	3917	104
Number of Links	1247	4784	9051	156
Size of Giant Component	356 (97.3%)	924 (68.5%)	2231 (57.0%)	80 (77.9%)
Link Density	0.02	0.01	0.003	0.05
Average Degree, <k>	6.97	4.62	2.87	1.94
Exponent, γ	0.67	1.41	1.11	1.33
Cutoff, κ	15.35	23.60	14.65	34.59

At the same time, the exponential cutoff implies that the distribution for large degrees decays faster than is expected for a power-law distribution, preventing the emergence of large hubs which have many links.

Two possible reasons have been suggested that may attenuate the effect of growth and preferential attachment: (a) the aging effect: as time progresses some older nodes may stop receiving new links, and (b) the cost effect: as maintaining links induces costs (Hummon, 2000), there is a constraint on the maximum number of links a node can have. We believe that the aging effect does exist in the dark networks. In the Meth World, for example, some criminals who were present in the network several years ago may have become inactive due to arrest or death, and thus could not receive new links even though they are still included in the network. Moreover, the cost of links takes the form of risks. Under constant threat from authorities, criminals or terrorists may avoid attaching to too many people, limiting the effects of preferential attachment.

Table 7-2. The small-world properties of the dark networks**.**

	GSJ		**Meth World**		**Gang Network**		**Dark Web**	
	Real	Random	Real	Random	Real	Random	Real	Random
Average Path Length	4.20	3.23	6.49	4.52	9.56	6.23	4.70	3.35
Diameter	9	6.00	17	9.57	22	16.40	12	13.16
Clustering Coefficient	0.55	0.2×10^{-1}	0.60	0.5×10^{-1}	0.68	0.6×10^{-3}	0.47	0.1×10^{-1}

Evidence has shown that hubs in criminal networks may not be the real leaders. Another possible constraint on preferential attachment is trust (Krebs, 2001). This constraint is especially common in the GSJ network where the terrorists preferred to attach to those who were their relatives, friends, or religious partners.

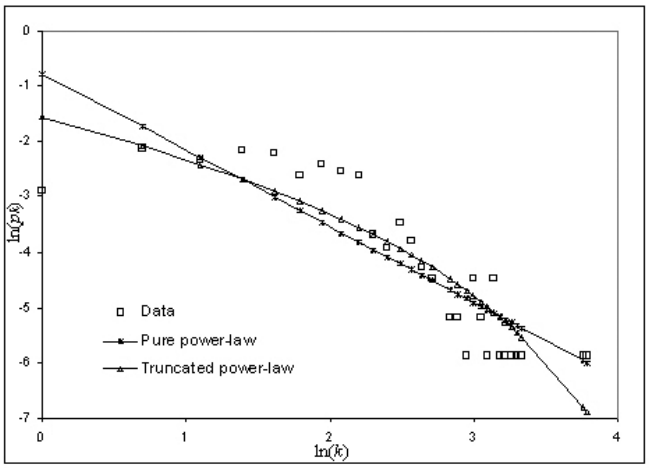

Figure 7-5. The degree distributions (x-axis: ln(k); y-axis: ln(p_k)).
Figure 7-5a. The degree distribution of the GSJ network.

.

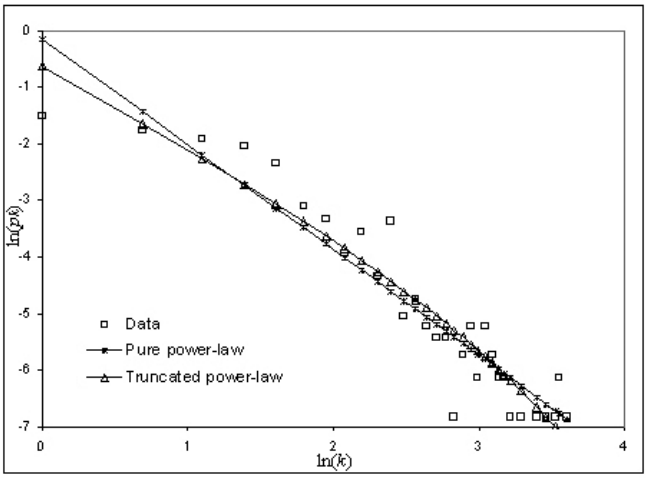

Figure 7-5b. The degree distribution of the Meth World.

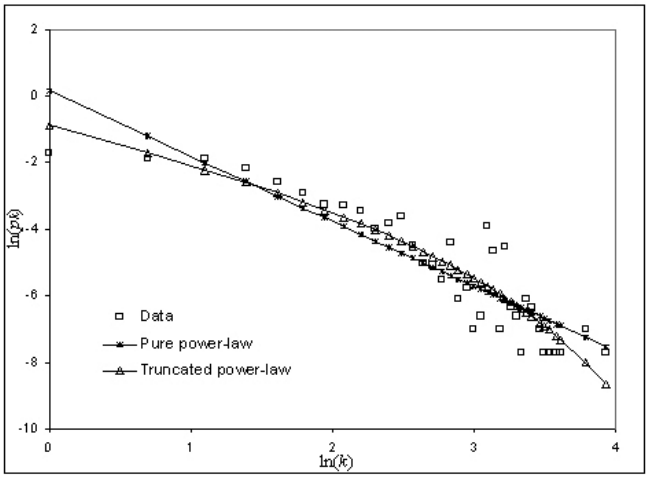

Figure 7-5c. The degree distribution of the gang network.

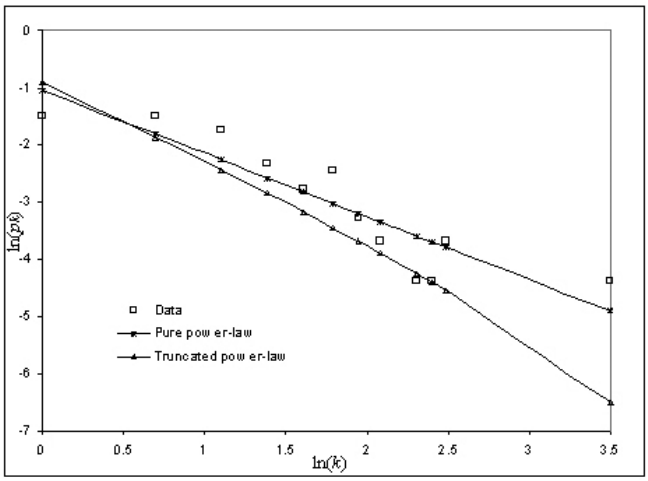

Figure 7-5d. The degree distribution of the Dark Web.

7.5 Future Directions

As terrorists may be involved in local crimes and conduct their acts in local jurisdictions, state and local law enforcement agencies are crucial in fighting terrorism. International terrorists may also leverage the covert gang, narcotic, and people smuggling networks to evade law enforcement investigation and capture. The tragic event of the Oklahoma City bombing shows the continuous threat of domestic extremist and terrorist groups. Information technologies can help find relationships between criminals and (domestic and international) terrorists.

We believe network mining, network visualization, and network topological analysis techniques are invaluable for understanding criminal and terrorist networks. Instead of studying single criminal entities or pairs of entities, we need to study criminals and terrorists using a network perspective. However, we also see a pressing need to develop new theories and principles based on the unique characteristics of the "dark networks" – covertness effect, law enforcement disruption effect, criminal aging effect, religious influence effect, etc. Criminology, criminal psychology, public policy, and terrorism research can potentially contribute to the study of dark networks.

7.6 Questions for Discussion

1. What are important readings and techniques in social network analysis, network learning, and network topological analysis that are suited for criminal and terrorist network analysis?
2. What are some ways to partner with domestic law enforcement and public safety agencies in ISI research?
3. What are some ways to develop or obtain research testbeds for domestic counter-terrorism research?