Chapter 6

BORDER AND TRANSPORTATION SECURITY

Chapter Overview

Terrorists enter a targeted country through an air, land, or sea port of entry. Criminals in narcotics rings travel across borders to purchase, carry, distribute, and sell drugs. Creating a “smart border” can greatly improve the government’s counter-terrorism and crime-fighting capabilities, where information from borders, customs, departments of transportation, and local law enforcement are integrated and analyzed to help locate terrorists or criminals. We present our “BorderSafe” project for cross-jurisdictional information sharing and criminal network analysis as an example for creating a smart and safe border.

6.1 Case Study 5: Enhancing "BorderSafe" Information Sharing

The BorderSafe project is a collaborative research effort involving the University of Arizona's Artificial Intelligence Lab, law enforcement agencies, including the Tucson Police Department (TPD), Pima County Sheriff's Department (PCSD), and Tucson Customs and Border Protection (CBP), as well as San Diego ARJIS (Automated Regional Justice Information Systems, a regional consortium of 50+ public safety agencies), San Diego Supercomputer Center (SDSC), and the Corporation for National Research Initiative (CNRI).

In this study our objective was to share and analyze structured, authoritative data from TPD, PCSD, and a limited dataset from CBP containing license plate data of border crossing vehicles. Tables 6-1 and 6-2 present the statistics of the three datasets. TPD's and PCSD's jurisdictions represent a shared community of citizens in Tucson and southern Arizona. They also share intertwined communities of criminals in these areas. We found a substantial amount of data overlap among these datasets. Around 7% of vehicles involved in gang-related, violent, and narcotics crimes were registered outside of Arizona. More than 483,000 people appeared in both the TPD and PCSD datasets. That represented 36% of the TPD records and 37% of the PCSD records. These statistics strongly suggest that sharing information across jurisdictions could help catch criminals.

Table 6-1. Statistics regarding the TPD and PCSD datasets.

	TPD	**PCSD**
Number of recorded incidents	2.84 million	2.18 million
Number of persons	1.35 million	1.31 million
Number of vehicles	62,656	520,539

Table 6-2. CBP border crossing dataset.

Item	**Number**
Number of records	1,125,155
Number of distinct vehicles	226,207
Number of plates issued in AZ	130,195
Number of plates issued in CA	5,546
Number of plates issued in Mexico	90,466

We employed the federation approach for data integration. At the schema level, we adopted the COPLINK schema as the global schema and developed a transformation mechanism to reconcile the database structure and semantics from a particular database into the global schema. Data were then mapped or transformed to allow shared query processing. In our datasets, the establishment of automated transformation procedures for putting legacy PCSD and TPD records into COPLINK format resolved most structural and semantic difference issues.

At the instance level, each dataset had a unique key assigned to each person or vehicle, but these unique keys did not match across datasets. To address this problem, vehicles were then matched between datasets based on their license plate numbers.

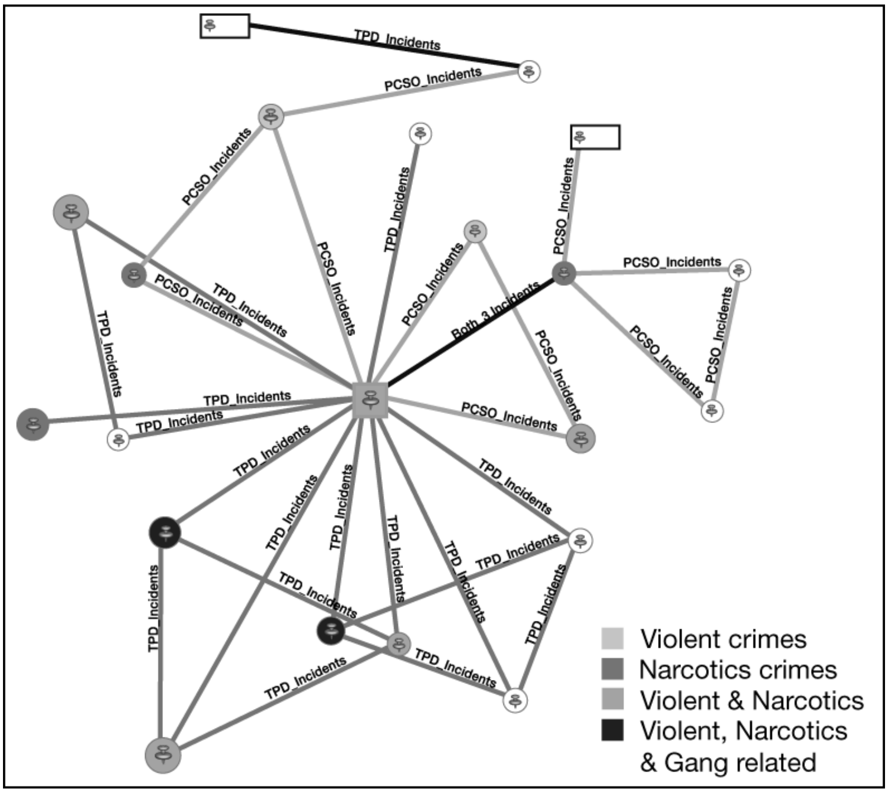

Figure 6-1. A sample criminal network based on integrated data from multiple sources. (In original figures, border crossing plates are outlined in red. Associations found in the TPD data are blue, PCSD links are green, and when a link is found in both sets the link is colored red. Colors are not shown here.)

For people match, we based the matching on input from domain experts and assumed that all records with the same first name, last name, and DOB represented the same person. These heuristics were not perfect; a few incorrect matches resulted and certainly many correct matches might have been missed.

We generated and visualized several criminal networks based on integrated data. We extracted associations between a set of criminals and vehicles from crime incident records. A link was created when two or more criminals or vehicles were listed in the same incident record. In network visualization we differentiated entity types by shape, key attributes by node color, level of activeness (measured by number of crimes committed) as node size, data source by link color, and some details in link text or roll-over tool tips. Figure 6-1 shows a network connecting a known narcotics dealer to a border crossing license plate.

A qualitative field study was conducted to assess the potential of our data integration approach. We received positive feedback regarding the value of integrated data. The crime analysts who participated in the evaluation study expressed confidence that including border crossing information into their investigations would promptly result in arrests and convictions. One analyst commented: "I had no idea that these types of complex networks involving Mexican plates even existed in the Tucson area. That information could subsequently be used to focus and direct law enforcement resources and investigations."

6.2 Case Study 6: Topological Analysis of Cross-Jurisdictional Criminal Networks

A criminal activity network (CAN) is a network of interconnected criminals, vehicles, and locations based on law enforcement records. The networks can be augmented with data from sources like transportation systems and motor vehicle division data. These networks allow us to analyze and visualize information that is helpful for identifying suspicious vehicles and people at the border or around critical infrastructures. Criminal activity networks can contain information from multiple sources and be used to identify relationships between people and vehicles that are unknown to a single jurisdiction (Chen et al., 2004). As a result, cross-jurisdictional information sharing and triangulation can help generate better investigative leads and strengthen legal cases against criminals.

Criminal activity networks can be large and complex (particularly in a cross-jurisdictional environment) and can be better analyzed if we study their topological properties. Topological properties describe the network as a whole and help us better understand its governing mechanisms. Topological properties can also be used to quantify the advantages of data sharing to law enforcement and transportation security agencies/personnel. In addition, understanding the properties of CANs can help design better analysis tools to assist in identifying potentially dangerous vehicles and people. In this study, we examine the topological properties of, and explore important research questions related to, cross-jurisdictional criminal activity networks.

Complex networks of individuals and other entities have been traditionally studied under the random graph theory. However, later studies suggested that real-world complex networks may not be random but may be governed by certain organizing principles (Albert & Barabasi, 2002). This prompted the study of real-world networks. These studies have explored the topology, evolution and growth, robustness and attack tolerance, and other properties of networks.

The datasets used in this study are available to us through the DHS-funded BorderSafe project. To study criminal activity networks we used police incident reports from the Tucson Police Department (TPD) and the Pima County Sheriff's Department (PCSD) from 1990 – 2002.

This testbed was used to extract narcotics networks that consisted of vehicles and individuals as nodes and police incidents as edges between them. Individuals were included as nodes in the network if they were wanted, suspected, arrested, or had a warrant for arrest in a narcotics crime. Vehicles were included as nodes in the network if they had been involved with a suspect in a narcotics crime. Two nodes were connected by an edge if they were in the same incident involving a narcotics or narcotics-related crime. Table 6-3 presents the basic statistics of the narcotics networks extracted from TPD and PCSD's records.

A giant component containing the majority of the nodes emerges from both networks. This is not uncommon; other social and affiliation networks that have been studied before exhibit this structure. The giant component, in this case, is a large group of individuals linked by narcotics crimes. In addition, we find that the second largest component is significantly smaller, suggesting that other much smaller groups of people exist in both jurisdictions.

Table 6-3. Basic statistics of TPD and PCSD narcotics networks.

	TPD	**PCSD**
Nodes	31,478 individuals	11,173 individuals
Edges	82,696	67,106
Giant component	22,393 (70%)	10,610 (94%)
2^{nd} largest component	41	103
Associated border crossing vehicles	6,927	2,979

Table 6-4. Small world properties of narcotics networks. Values in parenthesis are values for a random network of the same size and average degree.

	TPD	PCSD
Clustering Coefficient	0.39 (1.39 x 10^{-4})	0.53 (4.08 x 10^{-4})
Average Shortest Path Length (L)	5.09 (8.80)	4.62 (6.32)
Diameter	22	23

The small-world and scale-free properties of these and other networks shown later are studied by using the giant component. The small-world properties of both networks are shown in Table 6-4.

The narcotics networks in both jurisdictions can be classified as small-world networks since their clustering coefficients are much higher than comparable random graphs, and they have a small average shortest path length (L) relative to their size. The high clustering coefficient suggests that criminals show a tendency to form circles of associates who partner in crimes. According to domain experts this is not unusual in narcotics networks, where individuals tend to have circles of trust that include friends and family members. This property is advantageous to law enforcement because it helps them form strong conspiracy cases against other members of the group. A small L implies a faster flow of information (e.g., news of police raids) and goods (e.g., drugs) in the network. However, short paths could be advantageous for law enforcement too. Investigators search for associations among criminals to form a case against them. They suggest that shorter association paths between criminals generate better and higher quality investigative leads. Table 6-5 presents the scale-free properties of both networks.

Table 6-5. Scale free properties of narcotics networks.

	TPD	PCSD
Average Degree, <k> (average number of partners in crime)	3.12	4.33
Maximum Degree (largest number of partners in crime)	84	96
Exponent, γ	1.3	0.85
Cutoff, κ	17.24	16.71

The narcotics networks have degree distributions that follow the truncated power law, which classifies them as scale-free networks. This implies that a large number of nodes have low degrees as shown by the slow rate of decay (exponents of 0.85 – 1.3) at low values of k. This is not unexpected since high degrees attract more attention from law enforcement authorities, so having fewer associates is beneficial. However, it is worth pointing out that the degree of a node in these narcotics networks is also restricted by the fact that we are considering only narcotics-related crimes (to extract 'pure' narcotics networks). If other common crimes like traffic citations are included, then degrees are likely to be greater. Thus, the exponent (γ) value can be affected by the methods used for network extraction. The truncated power law distribution fits both curves better (R^2 = 93%) than the power law distribution (R^2 = 85%, 87%). This suggests that as the degree (k) increases, the probability of having k links ($p(k)$) decreases. This might indicate a cost or trust constraint to growth.

Table 6-6 shows the topological properties of the TPD narcotics network when it is augmented with associations found in PCSD data. No additional individuals from PCSD data were added.

Table 6-6. Topological statistics on adding associations (found in PCSD data) between the individuals in the TPD narcotics network. Values in parenthesis are for the original TPD network.

Giant component	27,700 (22,393)
Edges	98,763 (70,079)
Associated border crossing vehicles	8,975 (6,927)
Clustering coefficient	0.36 (0.39)
Average Shortest Path Length (L)	8.54 (5.09)
Diameter	24 (22)
Average degree, <k>	3.56 (3.12)
Maximum degree	96 (84)
Exponent, γ	1.01 (1.3)
Cutoff, κ	16.39 (17.24)

In Table 6-6, we see that the size of the giant component in the TPD narcotics network increases. Nodes that were previously thought to be disconnected from the main network got connected. Since we added only associations, it is clear that PCSD data contained associations between individuals in TPD data that TPD was not aware of. The increase in the number of edges shows that previously unknown associations between existing and new nodes were added. From a total of 28,684 new relationships added, 6,300 (statistic not in Table 6-6) were between existing criminals in the TPD narcotics network. These new associations between existing people help form a stronger case against criminals. The increase in the number of nodes and associations is a convincing example of the advantage of sharing data between jurisdictions.

The topological properties have important implications for law enforcement and hence transportation security. We found that a single jurisdiction may contain incomplete information on criminals and cross-jurisdictional data provides an increased number of high-quality investigative leads. The inclusion of vehicular data in criminal activity networks had clear advantages. Vehicles provided new investigative leads that can be used to target individuals and vehicles that might pose a threat to the security of the border and transportation infrastructure.

6.3 Future Directions

Many federal agencies are directly involved in protecting the safety of the U.S. borders (e.g., Immigration, Customs and Border Protection, Transportation Security Agency, Border Patrols, etc). Although most of these previously disparate agencies are consolidated under the Department of Homeland Security, significant cultural, organizational, and information technology barriers exist. In addition, other federal agencies, such as the FBI and the CIA, and local jurisdictions also hold critical information about selected border crossers and vehicles that may be of relevance. With the passage of the USA PATRIOT Act and the continuous re-structuring within DHS, we hope to see a more cooperative relationship among these agencies for information sharing.

However, information sharing without careful data mining research or civil liberties considerations will only cause information overload and potential misuse. Unreasonable border protection and immigration policy could also severely affect international trading. Much policy and technical research is needed for border and transportation security in this era of increasing globalization.

6.4 Questions for Discussion

1. How can information technology research and civil liberties considerations be balanced in the context of border and transportation security?
2. What are some ways to collaborate with various DHS agencies in border and transportation security research? What are the relevant DHS funding programs?
3. What are some of the other relevant technologies and techniques for border and transportation security research?