Detecting Anomalies in Cargo Using Graph Properties

William Eberle and Lawrence Holder

Department of Computer Science and Engineering
University of Texas at Arlington
Box 19015, Arlington, TX 76019-0015
{eberle, holder}@cse.uta.edu

The ability to mine relational data has become important in several domains (e.g., counter-terrorism), and a graph-based representation of this data has proven useful in detecting various relational, structural patterns [1]. Here, we analyze the use of *graph properties* as a method for uncovering anomalies in data represented as a graph.

Graph Properties. While our initial research examined many of the basic graph properties, only a few of them proved to be insightful as to the structure of a graph for anomaly detection purposes. For the *average shortest path length L*, we used the Floyd-Warshall all-pairs algorithm. For a measurement of *density*, we chose to use a definition that is commonly used when defining social networks [4]. For *connectedness*, we used a definition that Broder et al. [2] defined in their paper. For some of the more complex graph properties, we investigated two measurements. First, there is the maximum *eigenvalue* of a graph [5]. Another, which was used in identifying e-mail "spammers", is the *graph clustering coefficient* [3].

Synthetic Results. For each of our tests, we created 6 different graph size types consisting of approximately 35, 100, 400, 1000, and 2000 vertices, and another being a dense graph of 100 vertices and 1000 edges. For each of these increment sizes, we created 30 non-anomalous graphs. We then generated 30 anomalous graphs for each of the graph types and for each of the following structural anomalies: add substructure, remove substructure, move edge, and add isolated substructure. The *density* of small graphs lessens when an anomalous substructure is connected to existing vertices in the graph. This makes sense, as the ratio of actual vertices and edges to the number of *possible* pairs would increase, resulting in a lower density. This also explains why the density of graphs that contain *isolated* substructures is less, due to containing unconnected vertices. Also, the removal of a substructure results in a wide deviation in the density measurement. The *connectedness* of the smaller graphs varies for each of the different types of anomalies. The insertion and isolation anomalies result in lower values, and insertion of an isolated substructure has an even greater variation on the measurement. The same behavior is also found in dense graphs. Changes in the *clustering coefficient* on smaller graphs are only evident for inserted isolated anomalous substructures and the anomaly of moved edges. This variance, because of the moved edges, is significant due to the way the deviation changes. As the graphs get larger, the distribution still holds, but the coefficient of the graphs with moved edges increases significantly. The *average path length* and *eigenvalue* metrics behave similarly to the above metrics, except that they are better indicators of inserted substructures and moved edges.

S. Mehrotra et al. (Eds.): ISI 2006, LNCS 3975, pp. 728 – 730, 2006.

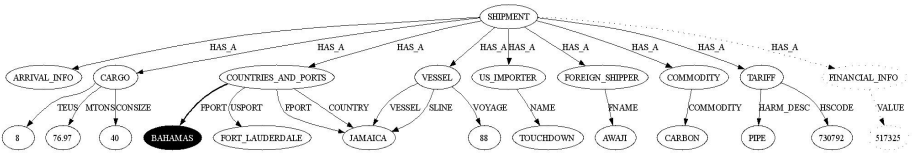

Fig. 1. Graph representation of cargo and anomaly (insertion in bold, removals as dotted lines)

Cargo Results. This data set consists of cargo shipments that represent imported items from foreign countries to the U.S. The anomalies that we introduced into the cargo data consist of two scenarios. The first anomaly represents drug smuggling [6], whereby the perpetrators attempt to smuggle the contraband into the U.S. without disclosing some financial information about the shipment. Also, an extra port was traversed in-route. While the shipment looked for the most part like containers of toys, food, and bicycles from Jamaica, there were a couple of structural alterations. Fig. 1 shows a graphical representation of a shipment (as a substructure in the entire graph) that contains the anomaly. For the second anomaly representing an arms shipment [7], similar to the first anomaly, there is certain manifest information not consistent with other similar (but legal) shipments. In addition, the original port of departure (in this case, China) is removed from the manifest. Again, these are all structural changes in the graph representation of the cargo data.

For both of these anomalies, there are no significant deviations displayed using the average shortest path or eigenvalue metrics. However, there are visible differences for the density, connectedness and clustering coefficient measurements. Another encouraging metric that can be used is the *combination* of individual measurements to provide a clearer view. For instance, when combining the density, connectedness and clustering coefficient measurements, we get values for the drug smuggling and arms shipment scenarios that clearly indicate anomalies. Similar results are evident when applying different combinations on the synthetic data sets.

Conclusion. We show that differences in graph properties between normal graphs and those intentionally altered can detect anomalies. While the changes vary based on the type of modification, they can be combined to clarify what is occurring, as was shown in the results from the issue of analyzing cargo containers for illegal, and possibly terrorist-related, shipments.

References

1. Holder, L., Cook, D., Coble, J., and Mukherjee, M.: Graph-Based Relational Learning with Application to Security. Fundamenta Informaticae Special Issue on Mining Graphs, Trees and Sequences, Vol. 66, Number 1-2. (2005) 83-101
2. Broder, A. et al.: Graph Structure in the Web. Computer Networks. Vol. 33. (2000) 309-320
3. Boykin, P. and Roychowdhury, V.: Leveraging Social Networks to Fight Spam. IEEE Computer, April 2005, Vol. 38, Number 4. (2005) 61-67
4. Scott, J.: Social Network Analysis: A Handbook. SAGE Publications, Second Edition. (2000) 72-78

5. Chung, F., Lu, L., and Vu, V.: Eigenvalues of Random Power Law Graphs. Annals of Combinatorics 7, 2003. (2003) 21-33
6. U.S. Customs Service: 1,754 Pounds of Marijuana Seized in Cargo Container at Port Everglades. November 6, 2000. (http://www.cbp.gov/hot-new/pressrel/2000/1106-01.htm)
7. Mae Dey Newsletter: Customs Seizes Weapons. Vol. 23, Issue 4, August/September (2003)