

Structural Patterns in Link Analysis (US Customs)

Daniel M. Venese

703-983-6137 • venese@mitre.org

USCS Office of Advanced Technology/\$50K

The logo for the MITRE Technology Program, featuring a stylized graphic of stacked blocks in yellow, orange, and blue to the left of the text.

MITRE
Technology
Program

Problem

- **Current methods and tools for link analysis have not been demonstrated to scale to large databases, have simplistic ways of categorizing relationships**
- **COTS data mining tools geared to finding large segments of a population (markets) are less suitable for incorporating domain knowledge required to profile minuscule segments of populations**
- **Critical shortcoming is finding suspect relationships in vast databases of routine transactions in an automated manner**
- **Hard to integrate knowledge from arbitrary set of COTS and GOTS analysis tools**



Background

- Threat from terrorist and major criminal organizations is perceived to be increasing, suspicious patterns can be identified from existing intel, DOD, federal law enforcement, and commercial databases
- Counter terrorism/criminal enterprise environment
 - Needle in haystack problem: highly skewed distributions
 - Sophisticated adversary, unlimited funds
 - Instant reaction to successful methods
 - Constantly changing patterns
- Better techniques needed to exploit large databases of millions of records: air travel passenger records, border crossings, alien visas, cargo imports and exports

Objectives



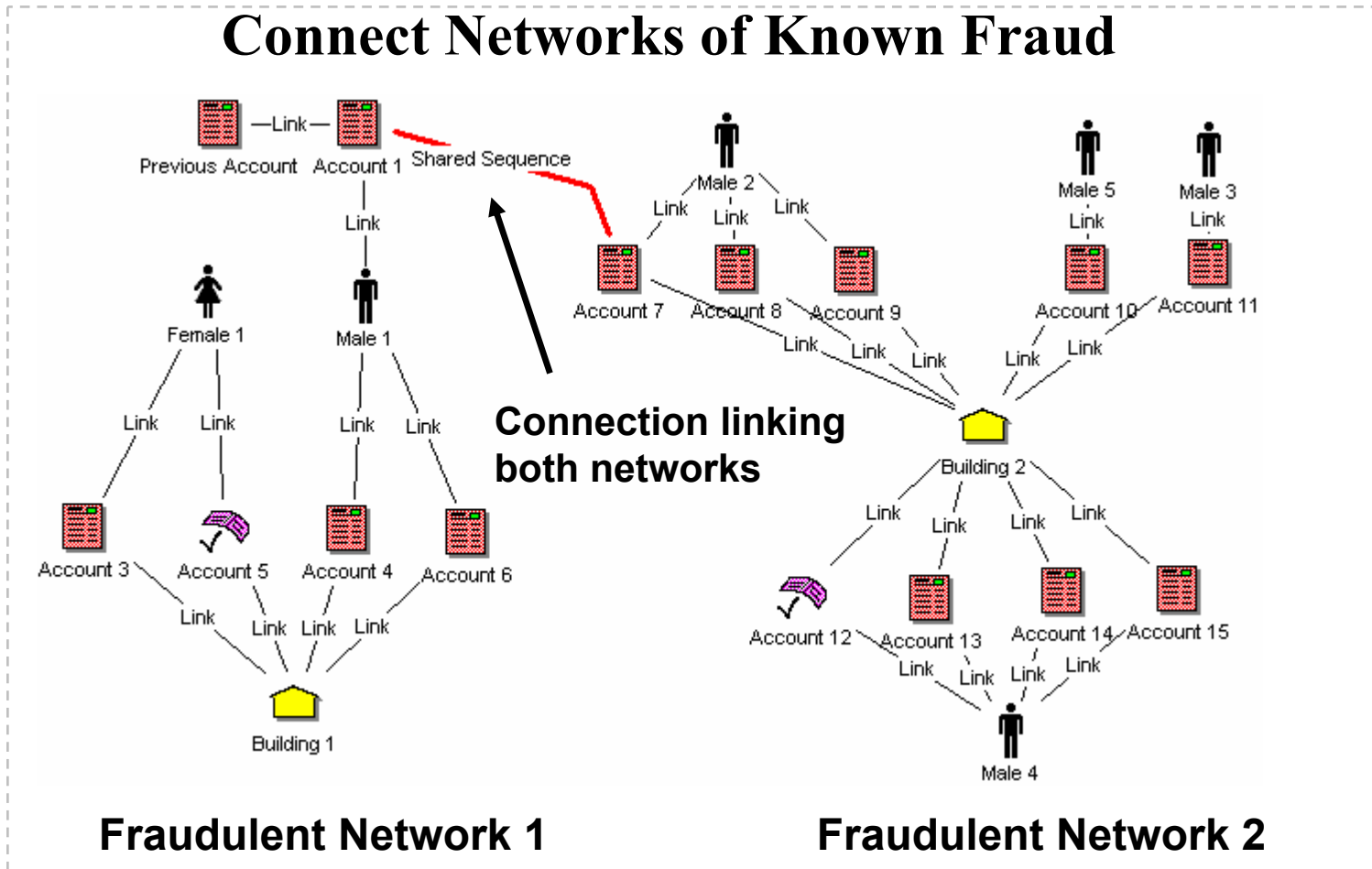
- Find patterns of suspicious relationships in large, structured databases of routine transactions using databases from the USCS
- Demonstrate scalability of link analysis techniques to databases containing millions of records
- Investigate the use of machine learning methods to automate the discovery of link subgraphs that might identify criminal behavior and terrorist suspects
- Derive quantitative metrics to guide application of link technology to sponsor environments

Activities

- **Build large-scale knowledge repository with real databases**
- **Experiment with approaches for representation of explicit and implicit associations**
- **Use automated data analysis to discover non-obvious chains of relationships among individuals and organizations, routine and criminal events**
- **Experiment with approaches for link traversal and searching**
- **Find instances of social networks (cliques), identify & classify suspect networks to find other instances**

Highlight

Using Automated Techniques to Identify and Connect Networks of Known Fraud



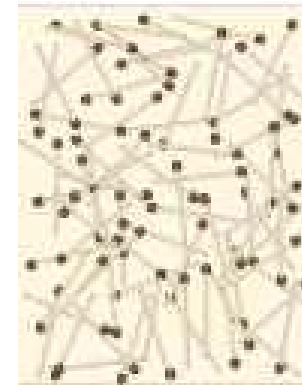
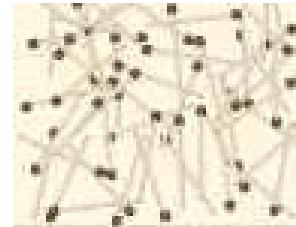
Fraudulent Network 1

Fraudulent Network 2

Demonstration



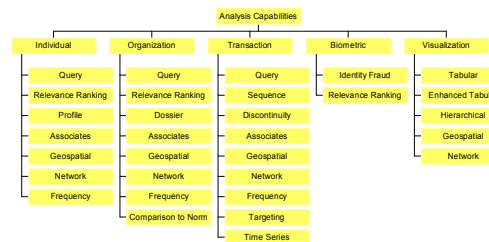
Automated detection of links



Finding networks similar to training examples



Finding relationship between two nodes

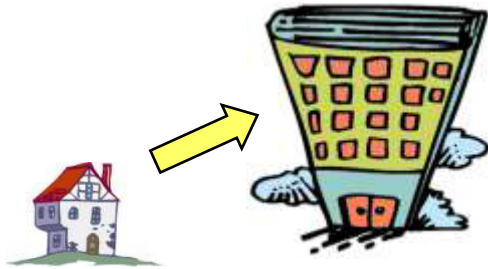


Finding associates to depth N

Impacts

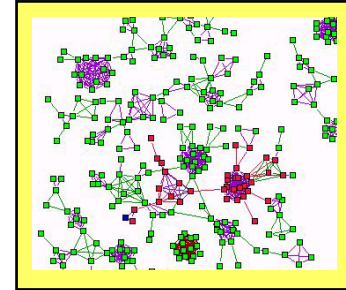
- **Homeland defense to address the problem of terrorist actions against the U.S. is a national priority. This research is intended to explore new methodologies for discovering terrorist organizations and other criminal behavior.**
- **Scalability of link analysis to structured databases similar to those of major federal law enforcement agencies**
- **Use of a link analysis metaphor for integrating arbitrary COTS and GOTS analysis tools**
- **Feasibility of automated techniques for discovery of suspect networks**
- **Guidance on modeling subtle relationships, link traversal, and addressing problem of link explosion**

Future Plans

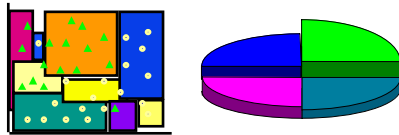


Explore link repository scalability issues

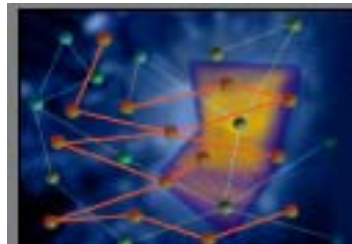
**Link
Repository**



Improve methods for dynamic creation of composite links



Continue analysis of algorithms for finding suspect networks



Improve performance for traversal and analysis of information space

