

Combining Data and Knowledge in Graph Analysis

Eric Bloedorn
703-983-5274 • bloedorn@mitre.org

Dave DeBarr
703-983-6231 • debarr@mitre.org



MITRE
Technology
Program

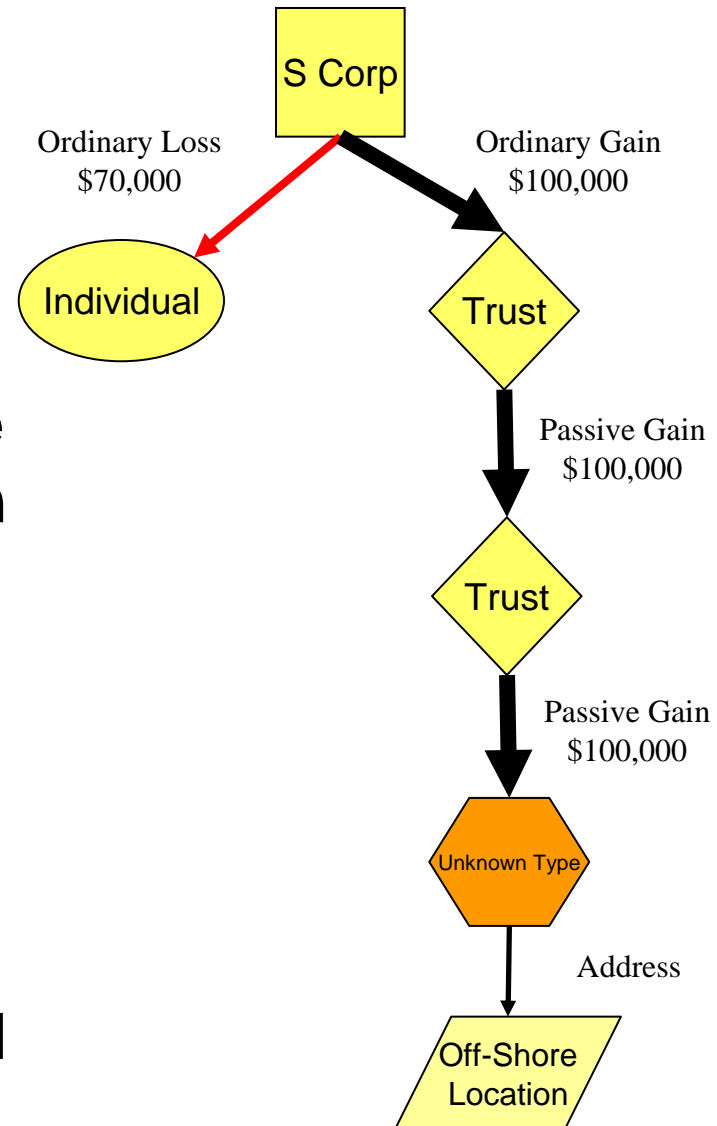
MITRE Sponsored Research

Problem

- **Data analysis tools today do not make effective use of information about relationships between entities.**
 - **Link analysis techniques assume a starting point of a “known suspect.”**
 - **Graph analysis techniques show promise, but...**
 - **There are no COTS tools for automatically analyzing large graph representations.**
 - **Limited “researchware” tools have several problems, including issues related to invalid assumptions and scalability.**

Background

- A graph is a collection of points and the links that connect them together.
- A graph-based data structure is a mathematical abstraction that allows for the automated analysis of relationships between entities.
- Graphs are important structures in numerous domains, from financial fraud to counterterrorism and biology.



Objective

- **Extend known automated techniques with domain knowledge in order to exploit information about relationships between entities**
- **Three specific challenges:**
 - 1. Inexact graph matching**
 - 2. Frequent substructure discovery and conceptual clustering**
 - 3. Active learning to support subgraph classification**

Activities

■ Obtaining and Representing Data

- large-scale simulation model for terrorist financing
- Predictive Toxicology
- KDDCup (authors and papers)
- graphgen graph generator (UT-Austin)

■ Building a Tool for Frequent Subgraph Discovery

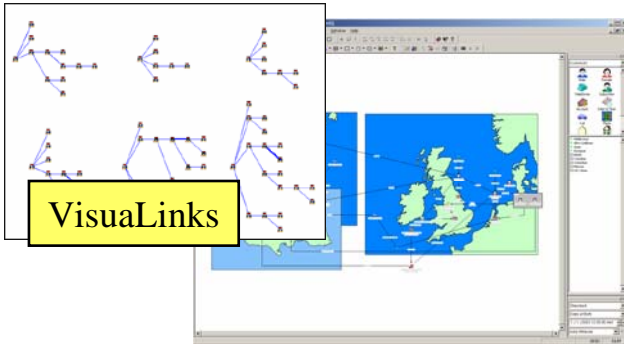
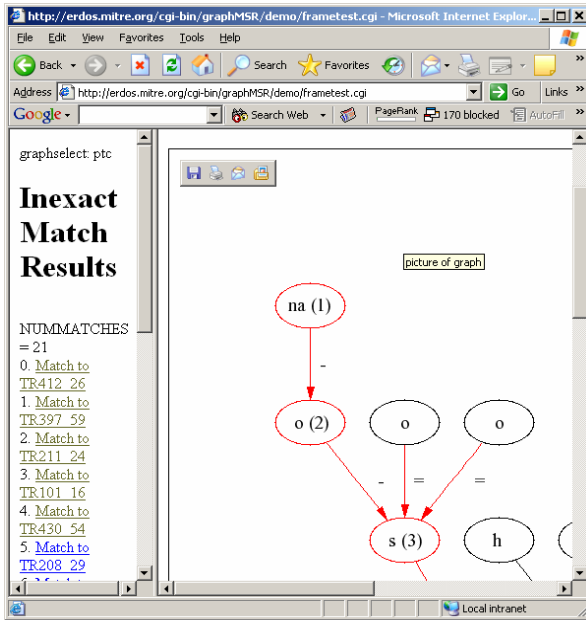
■ Building a GUI Tool for Graph Querying

■ Building a Tool for Inexact Graph Matching

Identifying Terrorist Financing

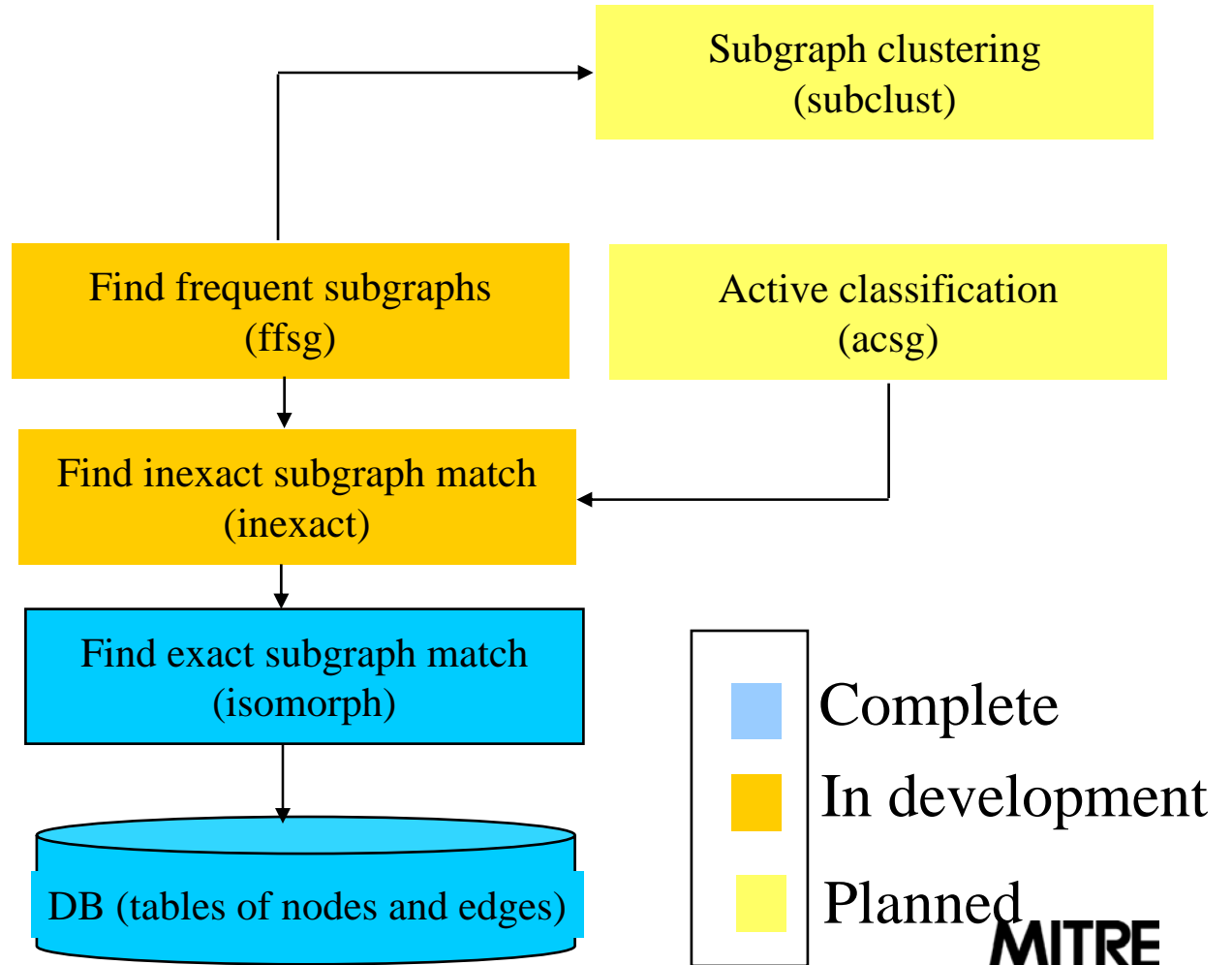
- **Definition**: “Persons or entities providing material support or resources to designated foreign terrorist organizations, or attempting or conspiring to do so.” - SAR Activity Review Trends, Tips & Issues Aug '04
- One estimate is that over a ten-year period al-Qaeda received between **\$300 million** and **\$500 million**, averaging US\$30 to US\$50 million a year.
 - Ronald Noble, Secretary General of Interpol
- Many approaches to concealing money flows:
 - Various types of fraud (credit card, charity)
 - Drug smuggling
 - Informal Value Transfer Systems (e.g., Hawala)

Highlight

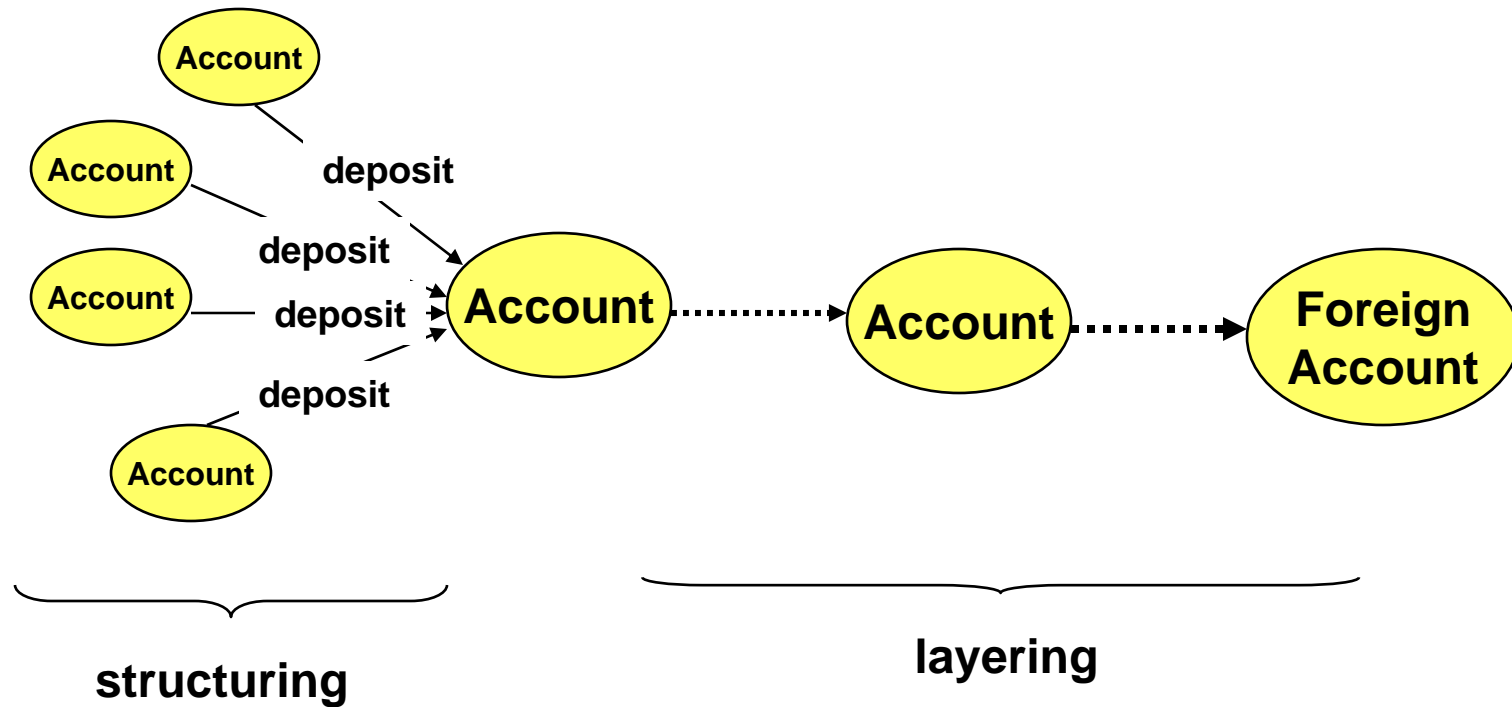


Analyst Notebook

Result



Highlight



Example Terrorist Financing Network

Impacts

- **Started applying code to sponsor work**
- **Collaboration with FinCEN**

Future Plans

- **Algorithm for active learning**
- **Application paper(s)**
 - **SIGKDD, AAI**
- **Further investigation of potential sponsor applications**