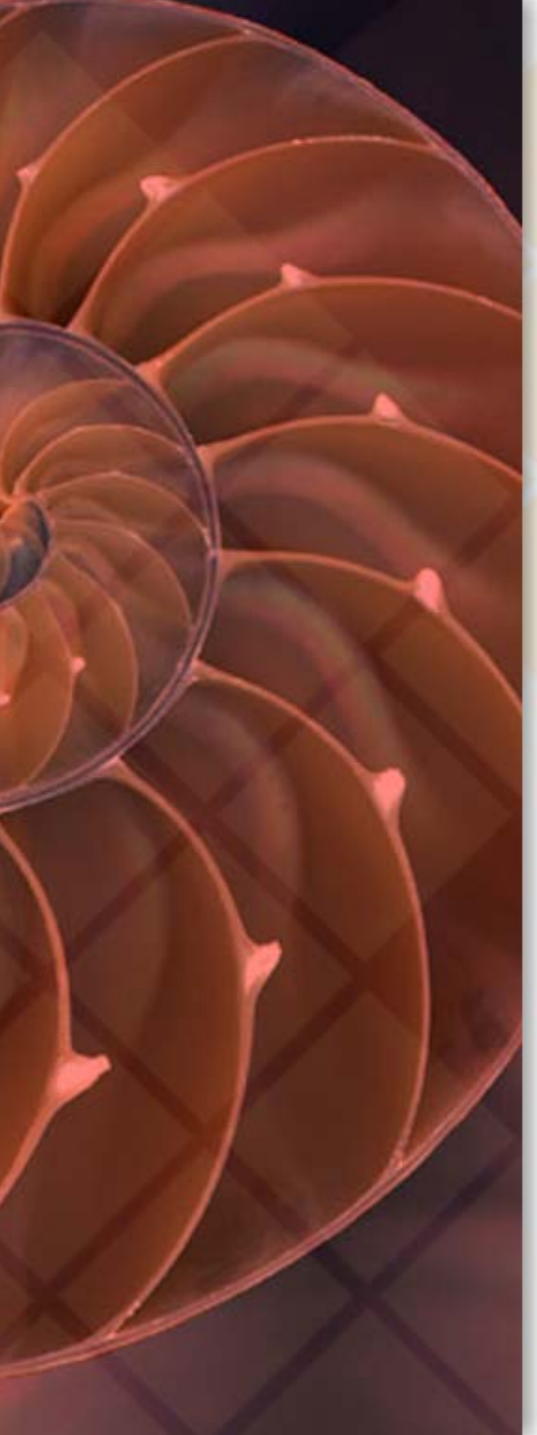


21st CENTURY TECHNOLOGIES

Self-Learning Ontologies

Presented to the 25th Soar Workshop
Ann Arbor, MI
June 15-17, 2005

Tim Darr, Ph. D.
University of Michigan AI Lab '96

Introduction – 21st Century Technologies

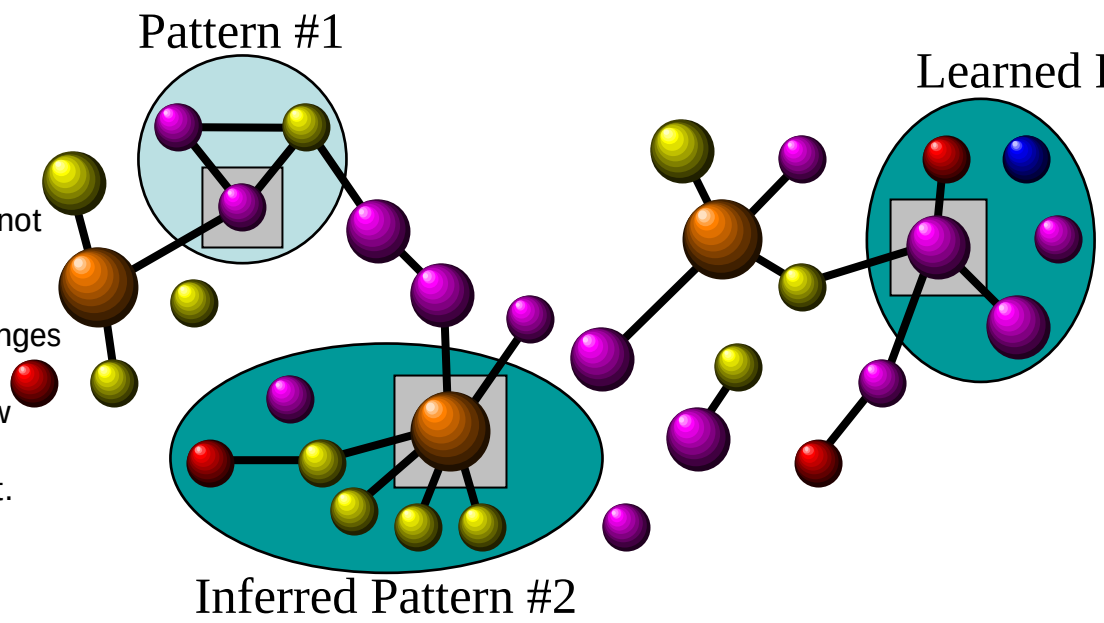
- R&D company in Austin, TX
  - 40 employees and growing
  - 50% of employees hold MS or PhD in Math, CS, or EE
  - First major contract award in 1998
- Much of our work has been done under contract for
  - US Army
  - USAF
  - ONR
  - Intelligence Community
  - **D**efense **A**dvanced **R**esearch **P**rojects **A**gency (DARPA)
  - **D**epartment of **H**omeland **S**ecurity (DHS)
- Founders
  - Sherry Marcus, Ph.D., MIT
  - Darrin Taylor, Sc.D., MIT

Introduction –21st Century Technologies

- Core Competencies
  - **Link Analysis**
  - **Graph-based pattern recognition**
  - **Social network analysis (SNA)**
  - Statistical Pattern Recognition
  - Data mining
  - Predictive planning
  - Natural Language Processing and Text Extraction
  - Fielded Applications for:
    - Defense
    - Intelligence
    - Homeland security

Definition of Link Analysis

Link Analysis is about Making Connections that represent *Meaningful Links* between Data Elements to detect Complex Relational Structures indicative of Patterns of Interest.



Patterns:

Cargo being shipped
From country that does not produce said cargo

Pattern of Shipment Changes

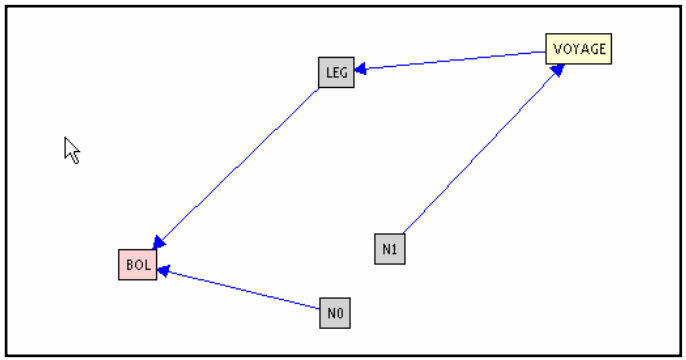
Commodity has a new "notify party"
For just one shipment.

DSB Study on Transnational Threats: making of *connections* between otherwise meaningless bits of information is at the core of (transnational) threat analysis."

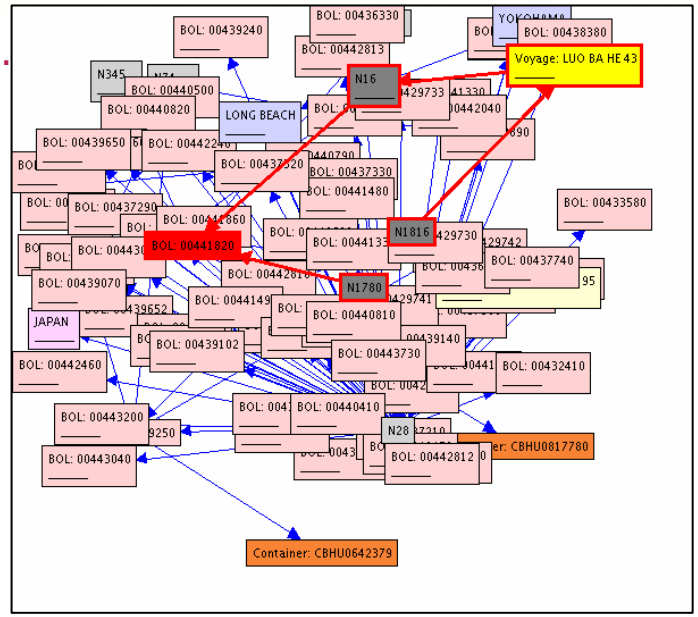
Connecting the Dots is Easy;
Deciding Which Dots to Connect Is Hard

Graph-Based Pattern Recognition

- Nodes represent things
  - People, organizations, objects, events, alerts, packets, etc.
- Edges represent relationships
  - Communication, friend, participant, owner, etc.



Graph Pattern



Graph Match

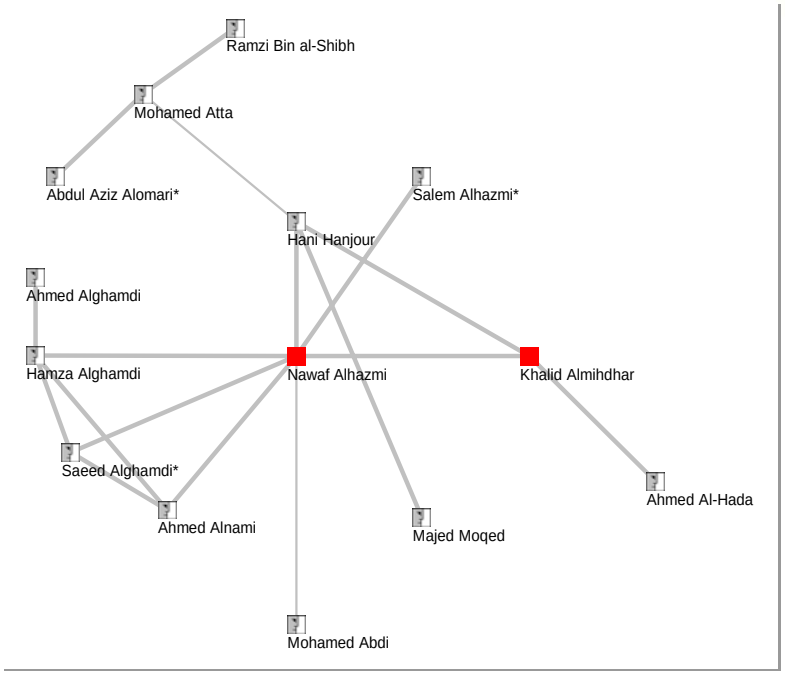
- Given:
  - A pattern graph that defines a threat
  - An evidence graph that records observed activity
- Graph matching lets you
  - Correlate events to quickly isolate true threats from normal activity
  - Develop higher-level situational awareness

Social Network Analysis (SNA)

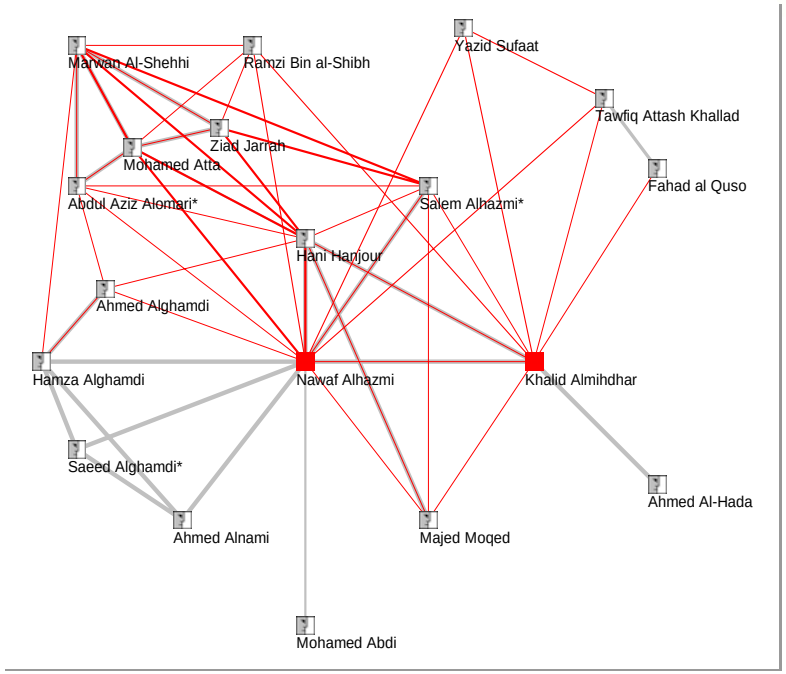
- Originally formulated for human social interaction
- Metric values for normal and abnormal behavior are usually different
  - We can use metric values to classify observed activity
    - Average path length
    - “Ring-like-ness”
    - Centrality / betweenness of nodes
    - Clique-ish-ness
- Many reasons for “abnormal” social behavior
  - Dysfunctional organization
  - Covert organization
    - Legitimate
    - Terrorist
    - Criminal
  - Unexpected organization
    - Terrorist activity within normal human social activity
    - Coordinated attacks within normal network activity

Dormant vs. Active Networks

Dormant 911 Network
(around 2 suspects)



Active 911 Network
(around 2 suspects)

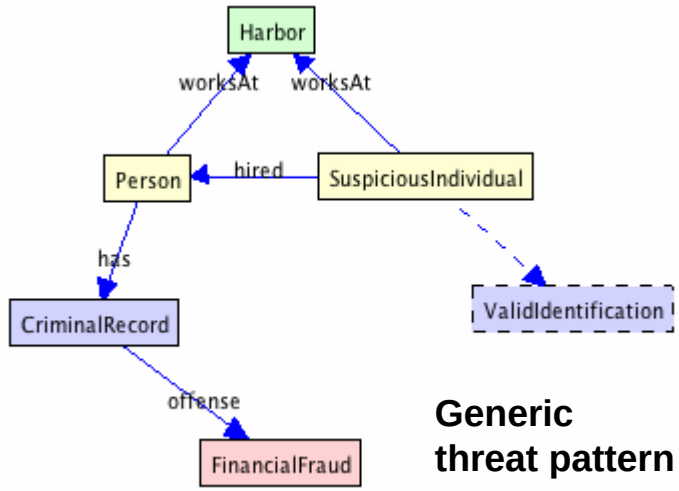


- **Approach:**
  - ✓ Calculate Salient Social Metrics from data.
  - ✓ Use Social Network Analysis (SNA) to identify threat signatures.
  - ✓ This method would have been successful in detecting active 911 network.

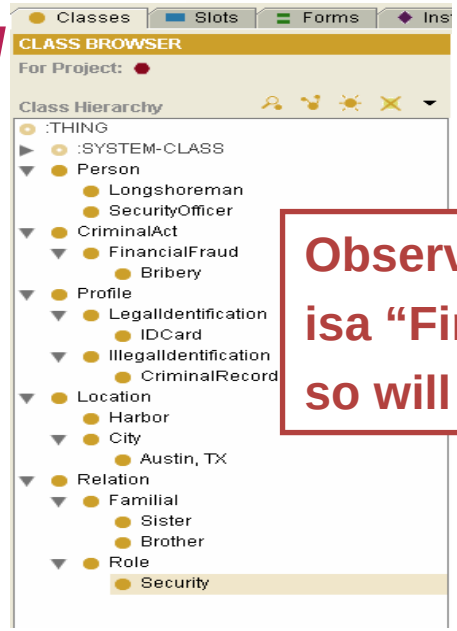
- *Observable Social Metrics:*
  - Group size: 15 ↑ 19
  - Link Count: 20 ↑ 49
  - Density: 19% ↑ 29%
  - Path Length: 2.6 ↓ 1.9
  - Closed Trios: 41% ↑ 60%
  - Appearance of Hubs

Ontologies in Graph Matching

- Traditional Ontologies
  - Define type-subtype relationships with multiple inheritance
  - Patterns refer to general event types instead of specific events
  - ***Ontologies provide linkage between observed data and patterns of interest***



Generic threat pattern



Observed data "Bribery" isa "FinancialFraud", so will match.

Terrorist Ontology

What is a “Self-Learning” Ontology?

- Advanced ontologies to provide more powerful ontological capabilities
- Ontologies based on cognitive systems
  - Domain-specific ontologies
    - Terrorists, terror groups, methods of financing
  - Ontologies as domain experts
  - Learn new patterns of behavior
    - Social network analysis
- Declarative representations
  - Encode knowledge in a declarative form
  - Infer categorization of entities
  - Easily respond to new facts
    - Streaming data ingest
- Ontology fusion
  - Fuse multiple representations
  - Inference over fused representations
  - Take advantage of strengths of different ontological representations and inference mechanisms

Evaluation

- Golden Nuggets
  - None
- Coal
  - No working system - YET
- For further evaluation or questions contact:
  - Dr. Tim Darr
    - tdarr@21technologies.com
  - Dr. Seth Greenblatt
    - sgreenblatt@21technologies.com