NSTB

National SCADA Test Bed

enhancing control systems security in the energy sector

SAND2007-7018C

National Laboratories

Threat Discovery Using Graph Analysis

Jason Stamp
Sandia National Laboratories, USA
jestamp@sandia.gov

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

U.S. Department of Energy
Office of Electricity Delivery
and Energy Reliability

Introduction

Background:

- Apply limited resources to protect
- Questions include:
  - Which vulnerabilities, if exploited, would be particularly damaging?
  - How likely is it that adversary will exploit a given vulnerability class (has interest and capability)?

Project Objective:

Help analysts in assessing the threat situation by providing capability to systematically and comprehensively address adversary interest/capability question

Introduction

- Systems of interest are often:
  - Large, dynamical networks (of people, organizations, information, resources, …) motivated to operate covertly
  - Observable only via subtle signatures embedded in massive, incomplete, noisy, largely irrelevant data
- Questions of interest:
  - Usually require discovery/analysis of complex relationships in data
  - Often aimed at reducing surprise

"I'm sure it's all right – it's a horse you have to worry about."

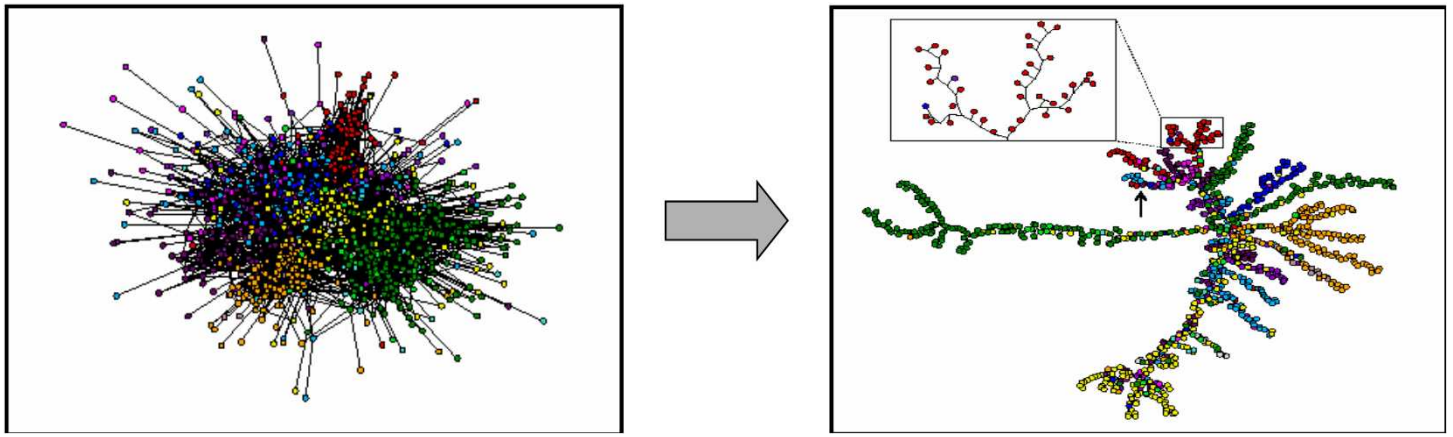
Discovery Objectives

- Develop real-time vulnerability analysis

  How likely the vulnerability has been identified by an adversary and the adversary is discussing an exploitation
- Use graph-based analysis to discover relationships in data
  - Use semantics to identify relationships
  - Vertex or node is equivalent to a data source (although not all sources are created equal - authoritative vs. non authoritative)
  - An edge is an association with multiple data sources
- Analyze and evaluate data, from plausible data associations
- Review viable scenarios, and search on derived approach
- Signature detection to analyze competing scenarios or hypotheses

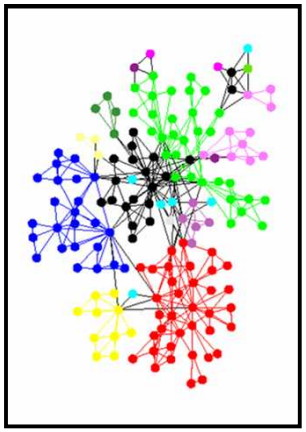
Analysis of Agent Transaction Graphs

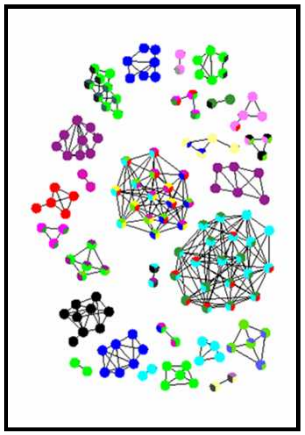
Topologies and dynamics of agents' transactions graphs (e.g., email from-to/co-recipient graphs) contain significant information and can be exploited using graph analysis methods.

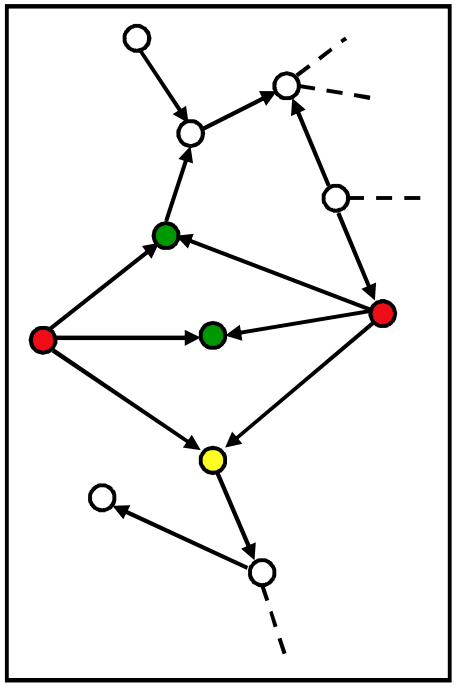


Analysis of Agent Transaction Graphs

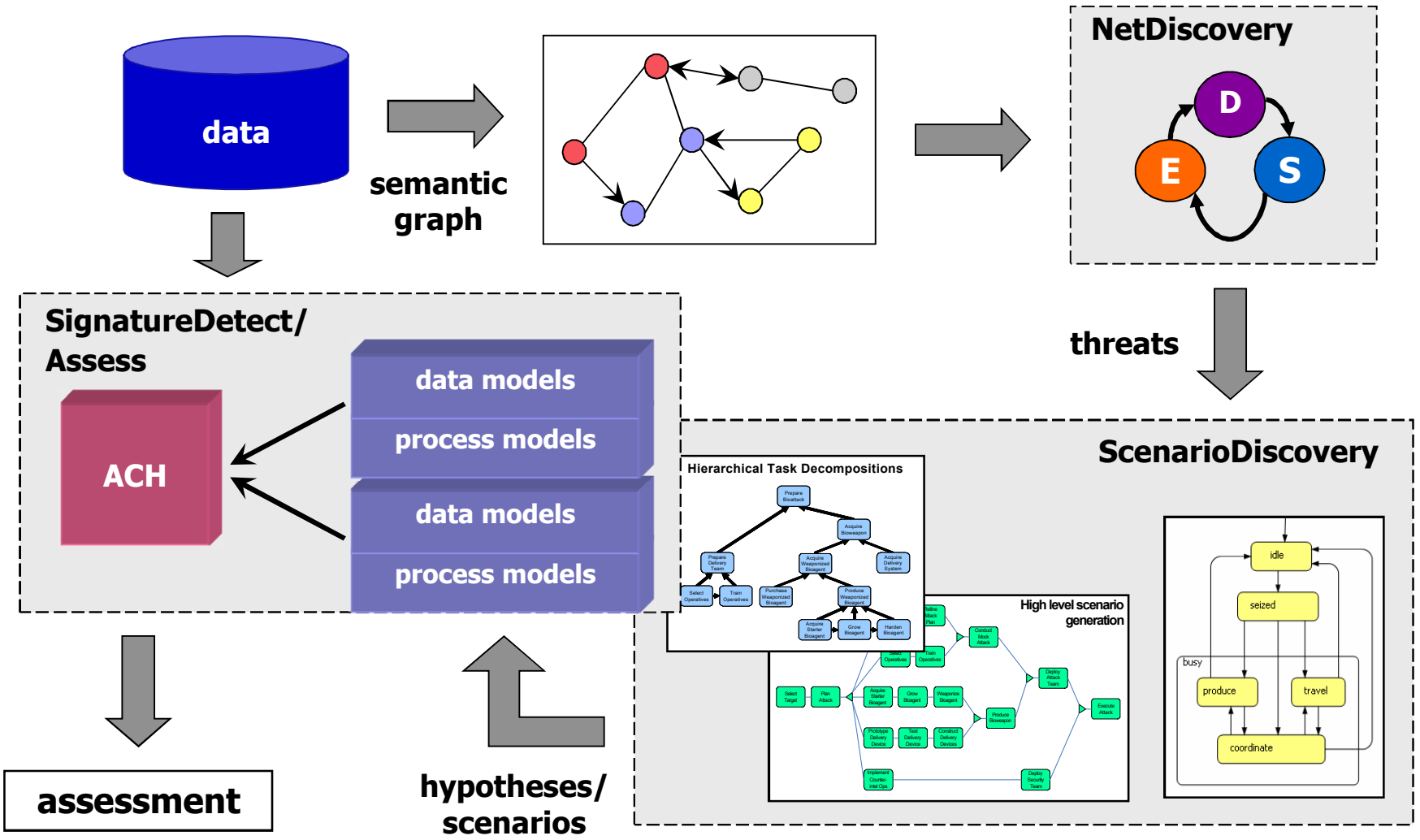
- Approach: apply motif discovery and dynamics analysis to agent transactions graphs
- Result: identified key agents and collaborating groups in organizations (e.g., terror networks, universities, firms)





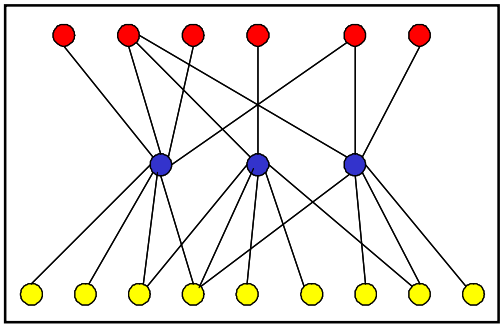


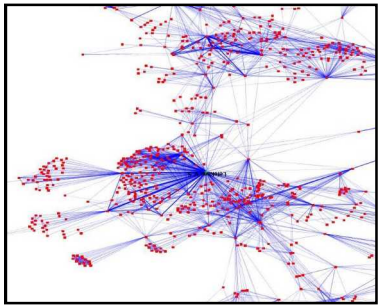
Threat Discovery and Analysis Framework

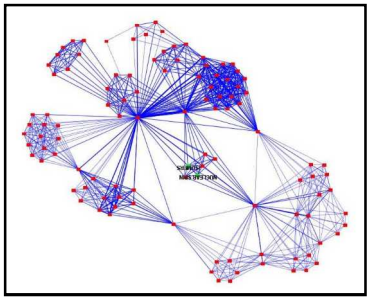


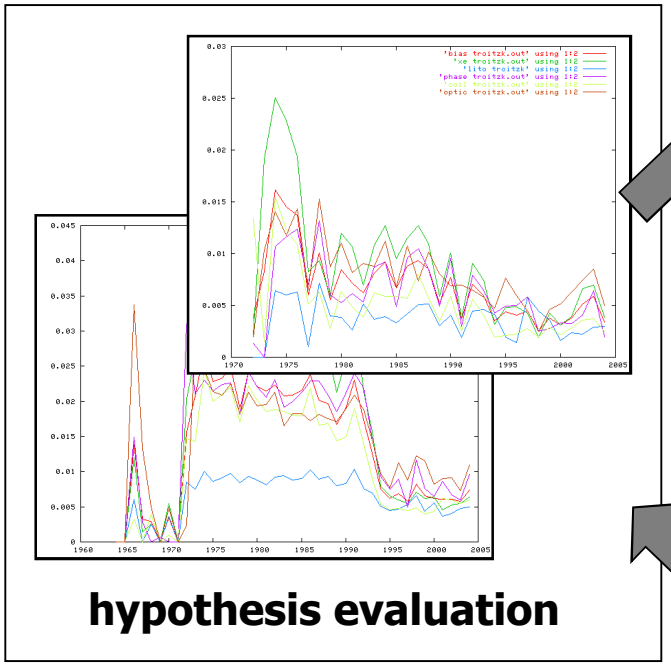
Examples: Threat Discovery

Network-based discovery

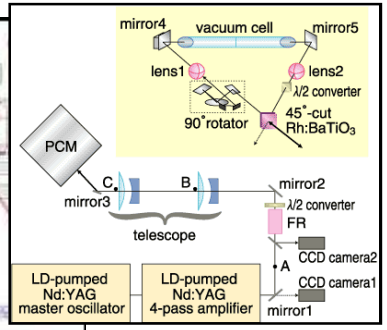








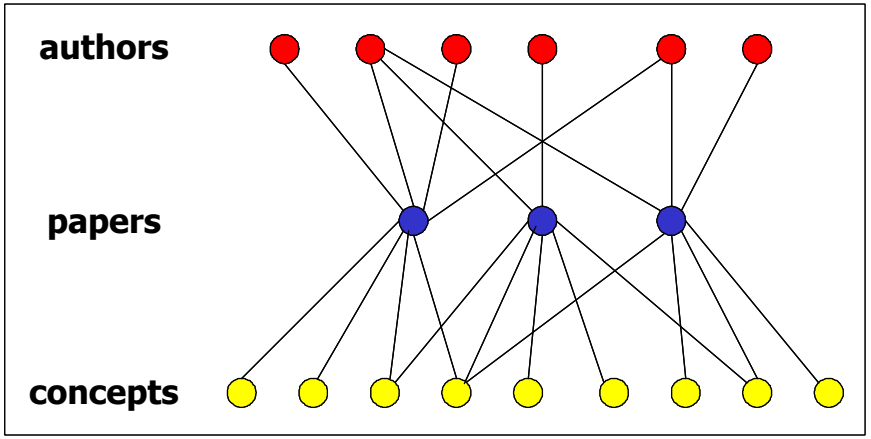
hypothesis evaluation



SME analysis

Example: Clandestine Bioweapons Program

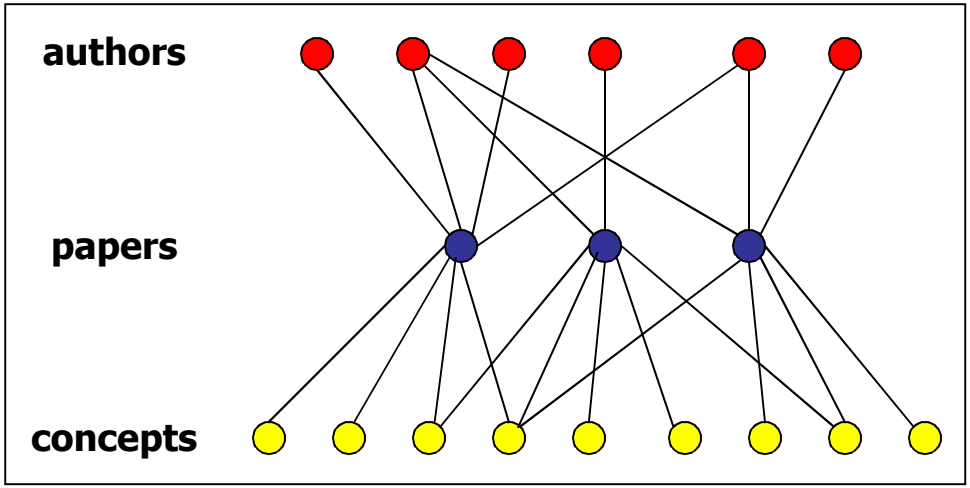
- Basic idea:
  - Scientific publications provide genuine (albeit distorted) information regarding scientific activity
  - Accurately assessing activity in presence of distortion requires quantitative/comprehensive perspective: network dynamics view
- Approach: Assess scientific activity by integrating publication data and research activity models within network dynamics framework



Analysis of Publication Networks

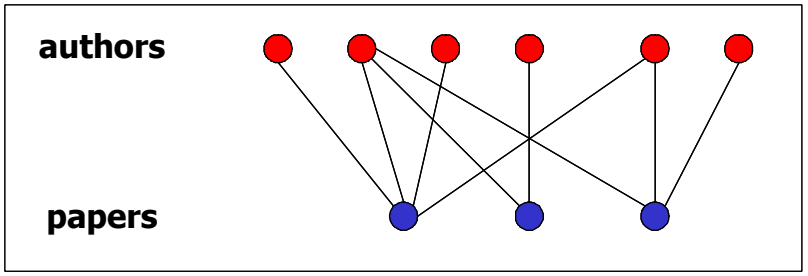
Consider a network representation of publication data and the following two natural subsystems:

- paper/concept networks
- author/paper networks



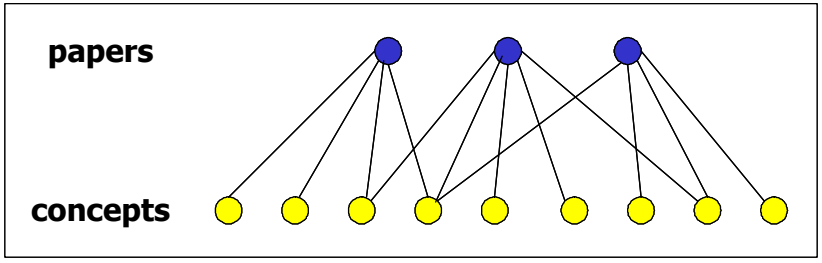
Analysis of Author / Paper Networks

- Model:
  - Evolving network that captures underlying research group structure and growth characteristics
  - Reproduces “stylized facts” of real networks (e.g., author/paper and paper/author distributions, clustering, diameter)
- Sample analysis goals: detect significant events; identify key scientists and collaborating groups



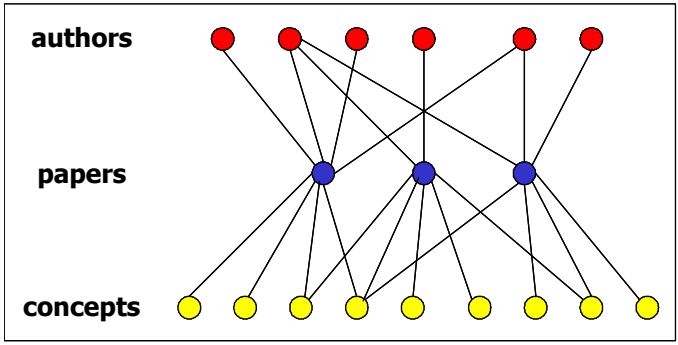
Analysis of Paper / Concept Networks

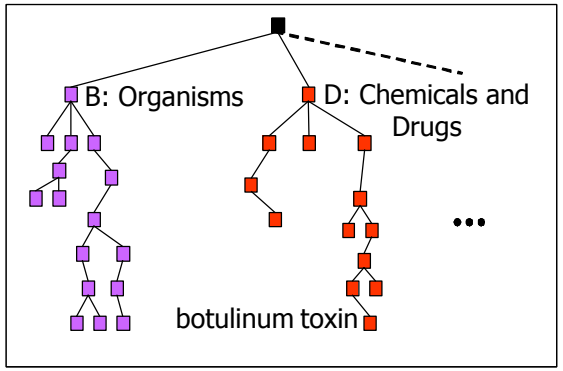
- Model:
  - characterization of paper-paper distance (MeSH taxonomy, concept overlap, etc.)
  - characterization of concept-concept distance (concept co-occurrence)
- Goals: identify interesting science (e.g., by identifying group for which “distant” concepts are close) and research trends



Scientific Publications Approach

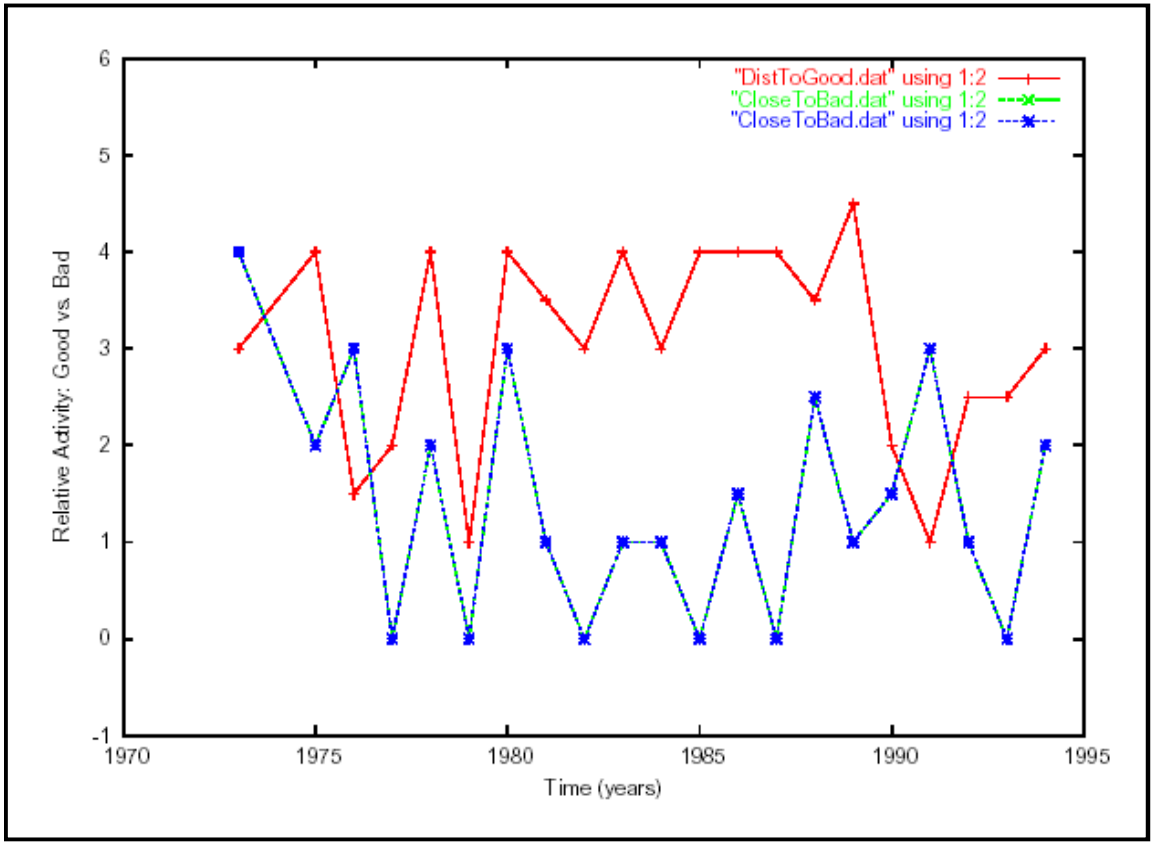
- Grow co-authorship graph and identify collaborating groups
- Assemble papers for interesting groups and extract MeSH terms
- Exploit MeSH taxonomic hierarchy to compute “distances” between group MeSH terms and {MeSH}C, {MeSH}B1, …
- Estimate group interest/ activity using time series of MeSH distances.





Analysis of BW Using Publication Networks

Sample results: detection of offensive bioweapons research activity despite aggressive D&D



Nontraditional NW Threat Assessment

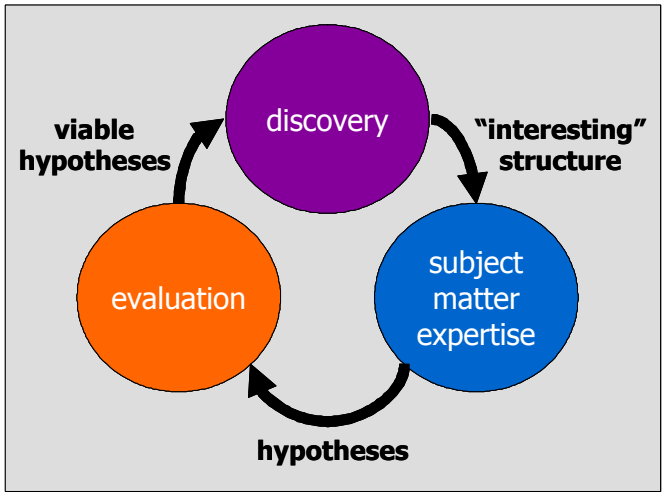
- Materials exist which represent attractive alternatives to the traditional elements used in NW (particularly for substate groups)
  - weapon-related properties are comparable to 235U
  - possess advantages from perspective of material acquisition
- Basic questions:

  1. Are substate groups interested in nontraditional approaches to NW (NT-NW) and can they discover such approaches?
  2. How could substate group produce such a NW?
  3. What are signatures associated with above activities?

Analytic Framework

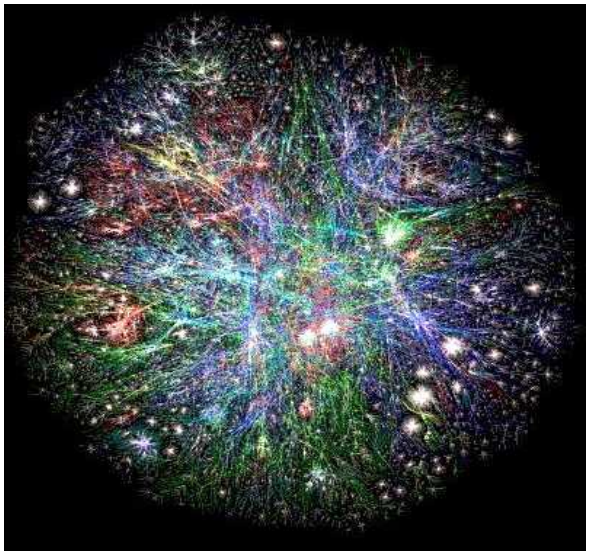
Given general threat domain to be explored (e.g., nuclear terrorism) generate specific, high quality hypotheses via discovery cycle of:

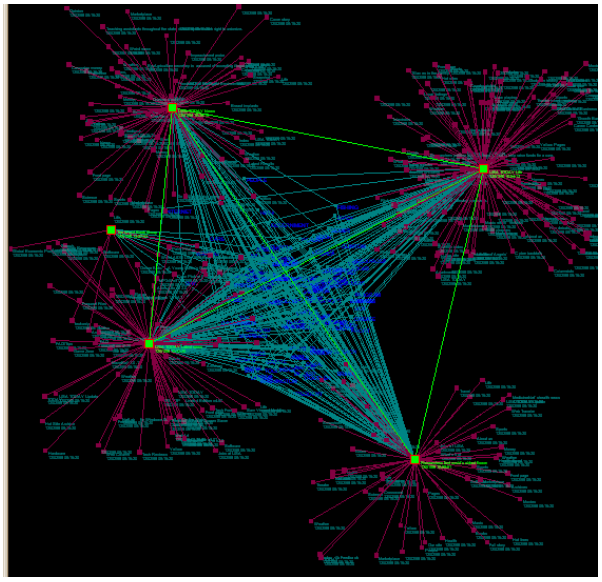
- network-based discovery
- SME analysis
- hypothesis evaluation / refinement



Automated Web Exploration

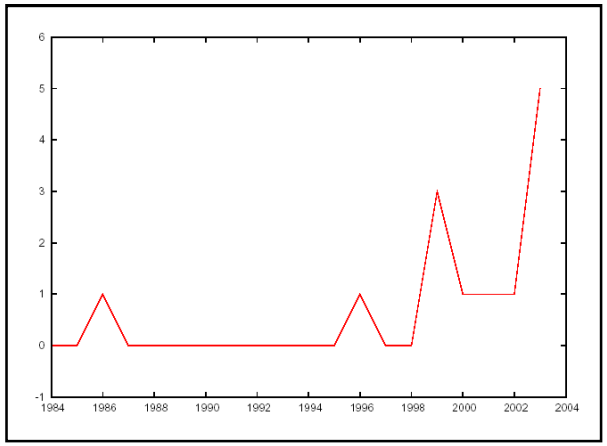
- Select web page “seeds”
- Assemble web page sets using web crawler/page filter
- Generate concept/web page graph
- Analyze concept/web page graph:
  - characterize web page network (e.g., extract communities, identify authoritative pages)
  - detect “interesting” capability (e.g., through structured anomaly detection)

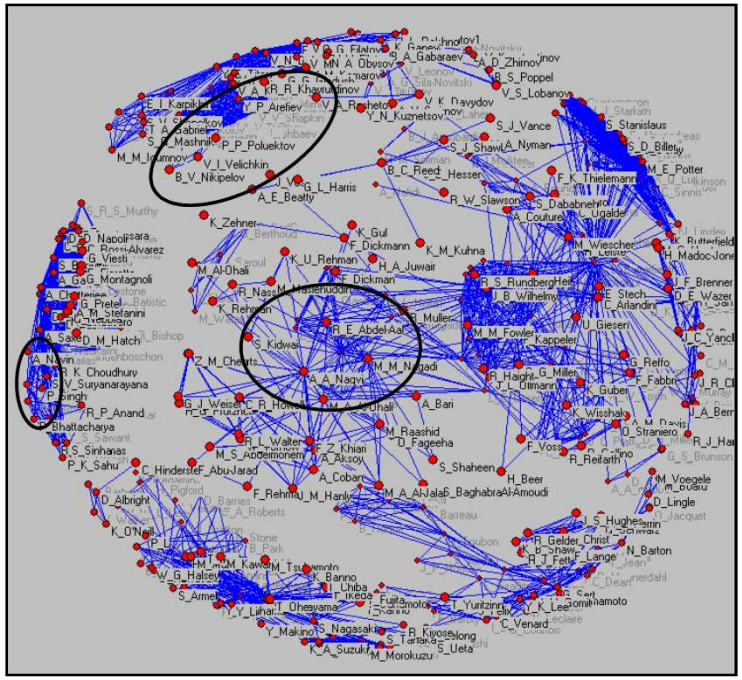




NT-NW Threat Assessment: Discoverability

- Automated whole document collection exploration yields
- Confirmation of feasibility/validity of X-based NW idea
- Detailed information on material availability / weaponization, scientists / groups with expertise, and significant events





NT-NW Threat Assessment: Interest

Computational exploration of DarkWeb archive (via web crawler, simple translation tool, webpage assessment algorithm) uncovers authoritative websites expressing such interest, including one site encouraging search for nuclear bomb materials which are “effective alternative[s] to uranium and available on the market”

Application to Cyber Threat

- Capability to be delivered: an analyst-support tool which takes as input large class of infrastructure cyber-vulnerabilities and discovers / assesses evidence that adversary is interested in, or capable of, exploiting such vulnerabilities
- Note: interest will be discovered/assessed via exploration and analysis of “buzz” in authoritative sources; capability will be judged by considering both inherent expertise and ease with which relevant information can be discovered by non-experts
- Data: data sets to be used for threat discovery / assessment are expected to include www archives, intelligence reporting, and various open sources