Welcome to the CASOS Summer Institute

Prof. Kathleen M. Carley

kathleen.carley@cs.cmu.edu

Overview

- Social Networks and Dynamic Network Analysis
- Focus on ability to use tools
  - Get the data in!
    - Importing, Creating, Cleaning
    - Text Analysis – AutoMap and NetMapper
    - Social Media - Twitter
  - Analyze the data! – ORA
    - Networks – social networks, semantic networks, sensor networks ...
    - Key nodes, Groups, Topology
    - Compare and Contrast
    - Geo-spatial, Trails
    - Dynamics, Over Time analytics and Change Detection
  - Forecast and Predict!
    - Immediate Impact analysis – ORA-NetScenes
    - Near term analysis - ORA (and Construct underneath)
    - Simulation – Construct
  - Get the results out!
    - Exporting: CSV, HTML, export to other network tools, pdfs, tiff, ...
Logistics

- Facilities
- Internet access
  - Wired and wireless
- CMU CASOS team
- Get tools on machine – breaks & lunch time
- Tools
  - [www.casos.cs.cmu.edu](http://www.casos.cs.cmu.edu)
- DVD's
  - Slides
  - Papers
  - Data
  - Tools
- Reception
- Sign ups – see handout
- Posters

Today

- Overview
- The big picture
- Field history
- How everything fits together
- From social networks to dynamic networks
- Dynamic Network Analysis
  - Key metrics
  - Intro to ORA
  - Groups
  - Importing structured data
Social Networks / Network Science
Fast-Growing Scientific Area

• In past 30 years growth has become exponential
• Popular culture
  – Games, plays, television
  – Forbes, Fortune, NY Times
• Business Practitioners
  – New concepts, tools for management consultants
  – New organizational forms; knowledge management
  – New applications such as search technologies
• Academia
  – Multiple fields from linguistics to AIDS research to political science to sociology to computer science
• Government
  – Multiple applications including counter-narcotics, counter-terrorism, command and control

6 Degrees of Separation
Social Networks

A social network is a description of the social structure at a particular point in time in terms of the actors (e.g., individuals or organizations) and the links among them.

A social network indicates the ways in which the actors are connected through various social familiarities ranging from casual acquaintance to close familiar bonds.

Networking

Networking: the process of building, maintaining, pruning, and using your communication network.

Principles:
- Develop access to useful information
- Develop a base of operations (power)
- Create opportunities for information confirmation
- Inhibit the flow of negative information
Definitions
From the NRC

- **Social Network**
  - The network of people to people, organizations to organizations mapping who knows, works with, communicates with whom
  - E.g., the data in facebook, the Enron email corpus, the NGO alliance for disaster response

- **Social Network Analysis**
  - The process of analyzing a social network and identifying key actors, groups, vulnerabilities and redundancies, and changes in these

- **Social Network Analysis Tools**
  - The set of tools, technologies, metrics, models and visualization techniques used for social network analysis – includes data extraction tools, link analysis, statistical techniques, and graph theory techniques

- **Social Network Theory**
  - The set of theories for forecasting, reasoning about, understanding how the social network forms, is maintained, evolves and the role of social networking tools, media, stress, etc. in effecting the emergence of, utilization of, and change in the social network

- **Social Networking**
  - The process of creating, maintaining, altering your network and using your network to gain resources or influence, mobilize activity, etc.
  - E.g., the process of twittering and blogging

- **Social Networking Tools**
  - A set of computational techniques that enable individuals and groups to engage in social networking, and to monitor and interact with their “ego” social networks i.e., the set of others with whom they are connected and the connections among those others.
  - E.g., facebook, twitter

Traditional Social Network Analysis

- Social network analysis [SNA] is the detection, tracking and analysis of links (social relations and flows) between people, groups, organizations, computers or other information/knowledge processing entities.

- The nodes in the network are the people and groups while the ties show how the nodes are linked.
Networks are Ubiquitous

Nodes
- People
- Topics
- Units of action
- Countries
- Hashtags
- Departments
- Resources
- Ideas or Skills
- Events

Ties Between Nodes (links)
- Transfer of resources
- Authority lines
- Association or affiliation
- Alliance
- Substitution
- Proximity
- What do you do
- Who do you like

Nodes
- People
- Topics
- Units of action
- Countries
- Hashtags
- Departments
- Resources
- Ideas or Skills
- Events

Connect & Dis-Connect the Dots!

<table>
<thead>
<tr>
<th>Degree</th>
<th>Betweenness</th>
<th>Closeness</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.417 Mohamed Atta</td>
<td>0.334 Nawaf Alhazmi</td>
</tr>
<tr>
<td>2</td>
<td>0.389 Marwan Al-Shehhi</td>
<td>0.318 Mohamed Atta</td>
</tr>
<tr>
<td>3</td>
<td>0.278 Hani Hanjour</td>
<td>0.227 Hani Hanjour</td>
</tr>
<tr>
<td>4</td>
<td>0.278 Nawaf Alhazmi</td>
<td>0.158 Marwan Al-Shehhi</td>
</tr>
</tbody>
</table>

Standard Social Network Measures
Why is Network Analysis Important

- Networks give precise formal definition to aspects of the socio-cultural-technical-political environment.
- Relations constrain and enable behavior:
  - Organization's inter-organizational position controls access to resources and information.
  - Manager's perception of inter-organizational network influences alliance decisions such as mergers and joint ventures, maneuvering of supply chain.
  - Individual's position in network influences chance of promotion.
  - Networks affect the flow of rumors, adoption of new technology, disease spread, ...
  - Employee networks influence: giving, hoarding, creating, influencing or accumulating information, resources, power... and so organizational behavior.
  - Managers use social networks...for knowledge management, early warning, control, outreach, planning, make decisions under uncertainty.
  - Topology of routers, electrical grid, internet backbone, ... affect dispersion of viruses and response; facilitate or degrade detection, survivability ...
  - Expertise involves building more densely connected information and transactive knowledge networks.

So – why is this hard?

- The Network:
  - Vast quantities of data.
  - Multi-mode – people, events, etc.
  - Multi-plex – many connections e.g. financial and authority.
  - Geo-temporal
- The Information:
  - Intentional misinformation – e.g., aliases.
  - Inaccurate information – e.g., typos.
  - Out-of-date information.
  - Incomplete information.
  - Inconsistent information.
  - Varying levels of resolution in spatially and temporally.
- Dynamic:
  - Learning.
  - Recruitment.
  - Attrition.
  - Daily activity...
Setting the Context

1940’s & before

- Sociology
- Organization Theory
- Communication
- Bibliometrics
- Electrical Grid
- Router topology
- Economics

Traditional Social Network Analysis
- Network Text Analysis
- Physics
- The “new” network science
- Dynamic Network Analysis
- Data Science

Big Networks
- Heuristics
- Dynamic Communication Media
- Sensor

Small Networks
- Animal Networks
- Groups
- Social Ties
- People Organizations Animals
- Static Questionnaires Observation

- Modern Link Analysis
- Relational Databases
- Machine learning
- Agent technology
- Computer Science
- Big Networks
- High Dimensional Data Science

Evolutionary Themes
**What is Dynamic Network Analysis?**

- The study of how entities are constrained and enabled by the relations among them and the process effecting change in these relations
- Combines social networks analysis, link analysis, multi-agent modeling, machine learning, graph theory, and non-parametric statistics
- Complex Meta-Networks: multiple networks, multiple types of nodes, multiple relations
- Key Issues: Scalability, Robustness, Flexibility, Error
  - Relations among nodes are flexible and vary in strength and certainty
  - Node membership may be questionable
  - Networks may be large $10^6$ nodes
  - Classes of data may not be discoverable

---

**Dynamic Network Analysis**

- **The Network Perspective**
  - It’s not just the elements (composition) of a system, but how they are put together
  - non-reductionist, holistic
- What are networks and how do you analyze them?
- Social Network Analysis, Link Analysis, Network Text Analysis, Dynamic Network Analysis
- Network Elites
- Groups and clustering
- Consensus and networks
- Network Topology
- Compare and contrast networks
- Network dynamics
- Network Visualization
Where are Dynamic Network Analysis Models Used

- Designing adaptive teams
  - Command and Control, organizational teams
- Evaluating organizational structures and changes like downsizing
  - E.g., hospitals, health departments, Medical Informatics
- Estimating effectiveness and adaptability of new structure
  - E.g., SSG – Comcargru, Army Unit of Action, CPOF (IRAQ), FAA
- Estimating size, shape and vulnerabilities in organizational designs and covert networks
  - E.g., NASA, Counter-terrorism, drug, terrorist, tax-avoiders
- Network management and IT intervention/effectiveness analysis
  - E.g., NASA, Knowledge Wall in JTF, supply chains, various companies
- Impact analysis of actions
  - E.g. Dark networks, spreading fake news,
- Deterrence
  - E.g., weaponized biological or chemical attacks, global cyber
- Identifying key actors and emergent groups
  - E.g., Counter terrorism, Health Units, Merchant Marine, Aviation
- Prevention and intervention
  - E.g., IRS tax avoidance interventions

Typical Mainstream Data Structure

Variables
(attributes)

<table>
<thead>
<tr>
<th>Cases (individuals)</th>
<th>Variables</th>
<th>Age</th>
<th>Gen</th>
<th>Education</th>
<th>Income</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>1001</td>
<td></td>
<td>28</td>
<td>0</td>
<td>&lt; highschool</td>
<td>28,000</td>
<td>PA</td>
</tr>
<tr>
<td>1002</td>
<td></td>
<td>35</td>
<td>1</td>
<td>highschool</td>
<td>54,000</td>
<td>MA</td>
</tr>
<tr>
<td>1003</td>
<td></td>
<td>26</td>
<td>0</td>
<td>&lt; highschool</td>
<td>26,000</td>
<td>MA</td>
</tr>
<tr>
<td>1004</td>
<td></td>
<td>40</td>
<td>1</td>
<td>Bachelors</td>
<td>65,000</td>
<td>MA</td>
</tr>
<tr>
<td>1005</td>
<td></td>
<td>24</td>
<td>0</td>
<td>highschool</td>
<td>27,500</td>
<td>PA</td>
</tr>
<tr>
<td>1006</td>
<td></td>
<td>55</td>
<td>1</td>
<td>Ph.D.</td>
<td>82,800</td>
<td>PA</td>
</tr>
<tr>
<td>1007</td>
<td></td>
<td>31</td>
<td>1</td>
<td>Ph.D.</td>
<td>73,000</td>
<td>MA</td>
</tr>
<tr>
<td>1008</td>
<td></td>
<td>M</td>
<td>0</td>
<td>highschool</td>
<td>33,500</td>
<td>PA</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Analysis consists of correlating attributes, regression, anova...
The Network Perspective

<table>
<thead>
<tr>
<th>Standard Statistics</th>
<th>Social Network Analysis</th>
<th>Dynamic Network Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attributes</td>
<td>Relations</td>
<td>Relations + Attributes</td>
</tr>
<tr>
<td>Atomistic</td>
<td>Interdependence</td>
<td></td>
</tr>
<tr>
<td>Actors as independent</td>
<td>Actors constrained and enabled by links</td>
<td>Actors constrained and enabled by links</td>
</tr>
<tr>
<td>Actor state matters</td>
<td>Actor state irrelevant</td>
<td>Actor state impacts perception of and use of links</td>
</tr>
</tbody>
</table>

Discovery of HIV: Sexual contacts among gay men w/ unusual cancers, traced by Bill Durow of the CDC

Network data is about relations

<table>
<thead>
<tr>
<th></th>
<th>Jenny</th>
<th>Martin</th>
<th>Chloe</th>
<th>Ramesh</th>
<th>Lenore</th>
<th>Stephen</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jenny</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Martin</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Chloe</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ramesh</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Lenore</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Stephen</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
### Network Data Structures

#### Adjacency matrices

**Friendship**

<table>
<thead>
<tr>
<th></th>
<th>Jenny</th>
<th>Martin</th>
<th>Chloe</th>
<th>Ramesh</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jenny</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Martin</td>
<td>1</td>
<td>-</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Chloe</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>Ramesh</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>-</td>
</tr>
</tbody>
</table>

**Shared Concepts**

<table>
<thead>
<tr>
<th></th>
<th>Jenny</th>
<th>Martin</th>
<th>Chloe</th>
<th>Ramesh</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jenny</td>
<td>-</td>
<td>5</td>
<td>35</td>
<td>21</td>
</tr>
<tr>
<td>Martin</td>
<td>5</td>
<td>-</td>
<td>3</td>
<td>24</td>
</tr>
<tr>
<td>Chloe</td>
<td>35</td>
<td>3</td>
<td>-</td>
<td>3</td>
</tr>
<tr>
<td>Ramesh</td>
<td>21</td>
<td>24</td>
<td>3</td>
<td>-</td>
</tr>
</tbody>
</table>

#### Incidence matrix

**Friendship**

<table>
<thead>
<tr>
<th></th>
<th>Jenny-Martin</th>
<th>Jenny-Chloe</th>
<th>Jenny-Ramesh</th>
<th>Martin-Chloe</th>
<th>Martin-Ramesh</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jenny</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Martin</td>
<td>1</td>
<td>3</td>
<td>24</td>
<td>0</td>
<td>21</td>
</tr>
<tr>
<td>Chloe</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>24</td>
</tr>
<tr>
<td>Ramesh</td>
<td>0</td>
<td>21</td>
<td>3</td>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>

**Shared Concepts**

<table>
<thead>
<tr>
<th></th>
<th>Jenny-Martin</th>
<th>Jenny-Chloe</th>
<th>Jenny-Ramesh</th>
<th>Martin-Chloe</th>
<th>Martin-Ramesh</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jenny</td>
<td>1</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Martin</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Chloe</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

---

**Note** – Network data violates standard statistical assumptions i.e., that cases are iid

---

### Dynamic Network Data

#### Time 1

**Friendship**

<table>
<thead>
<tr>
<th></th>
<th>Jenny</th>
<th>Martin</th>
<th>Chloe</th>
<th>Ramesh</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jenny</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Martin</td>
<td>1</td>
<td>-</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Chloe</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>Ramesh</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>-</td>
</tr>
</tbody>
</table>

**Shared Concepts**

<table>
<thead>
<tr>
<th></th>
<th>Jenny</th>
<th>Martin</th>
<th>Chloe</th>
<th>Ramesh</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jenny</td>
<td>-</td>
<td>5</td>
<td>35</td>
<td>21</td>
</tr>
<tr>
<td>Martin</td>
<td>5</td>
<td>-</td>
<td>3</td>
<td>24</td>
</tr>
<tr>
<td>Chloe</td>
<td>35</td>
<td>3</td>
<td>-</td>
<td>3</td>
</tr>
<tr>
<td>Ramesh</td>
<td>21</td>
<td>24</td>
<td>3</td>
<td>-</td>
</tr>
</tbody>
</table>

#### Time 2

**Friendship**

<table>
<thead>
<tr>
<th></th>
<th>Jenny</th>
<th>Martin</th>
<th>Chloe</th>
<th>Ramesh</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jenny</td>
<td>-</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Martin</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Chloe</td>
<td>1</td>
<td>0</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>Ramesh</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>-</td>
</tr>
</tbody>
</table>

**Shared Concepts**

<table>
<thead>
<tr>
<th></th>
<th>Jenny</th>
<th>Martin</th>
<th>Chloe</th>
<th>Ramesh</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jenny</td>
<td>10</td>
<td>40</td>
<td>21</td>
<td>10</td>
</tr>
<tr>
<td>Martin</td>
<td>10</td>
<td>30</td>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td>Chloe</td>
<td>30</td>
<td>20</td>
<td>30</td>
<td>-</td>
</tr>
<tr>
<td>Ramesh</td>
<td>30</td>
<td>20</td>
<td>30</td>
<td>-</td>
</tr>
</tbody>
</table>

- **Meta-Network**
  - Multi-mode (different kinds of entities)
  - Multi-Link (different kind of connections)
  - Multi-Level (different aggregations)

- **Connecting**
  - Who, what, how, why, when, where

- **Often available as trails**
  - E.g. who was where when
  - Co-presence becomes interaction data
"Injects Drugs With" Relation

- African-American
- Puerto-Rican
- US White

Drug using relations in Hartford, CT

Illustrative Networks

- High School Dating
- Physicist Collaborations
- Contagion of TB
- Fresh Water Food Web
- Sexual Contacts
- The Internet
- Topic Network (Email)
- Email Profile
- al Qaida 2004

Nodes have attributes that matter
**Twitter Example - Benghazi**

Some Nodes Stand Out

**Network Effects**

<table>
<thead>
<tr>
<th>Access</th>
<th>Control of Flows</th>
<th>Influence</th>
</tr>
</thead>
<tbody>
<tr>
<td>The more people you are connected to the more you can know</td>
<td>The more you are on the path between people the more you can control</td>
<td>The more you are connected to others who are connected to each other the more influence they have on you and you on them</td>
</tr>
</tbody>
</table>
Network Analysis Enables Management and Disruption

A vulnerability to exploit!

This guy needs help!

Non-Lethal Targeting

- Target differentiation
- Objective targeting
- Faster
- New insights

Max Disruption
Betweenness Centrality

Max Information
Closeness Centrality

Low value for Disruption

High value for Information

Numerical values of node importance based on network structure

Resource Broker
Exclusivity/Specialization
Organizations are Formal Relations

- Lines of Authority
- Information passes up
  - With information loss
- Commands pass down
  - With intent loss
- Lateral communication is for coordination
- This information is codified
  - E.g. organizational charts, doctrine ...

Organizations have Formal and Informal Structures

Each person is embedded in many networks
Two Ways to Think About Networks

What it means for the individual
• Your position in the network impacts:
  – Promotions
  – Access to information and resources
  – Health
  – Beliefs and voting behavior
  – Etc.

What it means for the organization
• The network topology of the group impacts:
  – Efficiency
  – Resiliency
  – Speed with which information diffuses
  – Performance caps
  – Points of vulnerability

Ego Network
• Ego – the node you are examining
• Alters – the nodes directly connected to ego
• The set of ties directly to/from ego to/from an alter
• + (sometimes) the set of ties among ego’s alters
1-mode ego network

Carter Administration meetings

Year 1

Year 4

Drug Networks

Normal Person

Cocaine User

People with Different Roles have Different Networks
We pull out – the Meta-Network

Meta-Network: multi-mode, multi-plex, multi-level

<table>
<thead>
<tr>
<th></th>
<th>People</th>
<th>Expertise</th>
<th>Activities</th>
<th>Locations</th>
</tr>
</thead>
<tbody>
<tr>
<td>People</td>
<td>Social Network</td>
<td>Knowledge Network</td>
<td>Assignment Network</td>
<td>Presence Network</td>
</tr>
<tr>
<td>Expertise</td>
<td></td>
<td>Information Network</td>
<td>Needs Network</td>
<td>Availability Network</td>
</tr>
<tr>
<td>Activities</td>
<td></td>
<td>Precedence Network</td>
<td></td>
<td>Happenings Network</td>
</tr>
<tr>
<td>Locations</td>
<td></td>
<td></td>
<td></td>
<td>Border Network</td>
</tr>
</tbody>
</table>
Sphere of Influence

- All ties to/from ego and the connections among the associated nodes in any meta-network
- Extension of ego net idea to meta-network
- Level = the selection path length allowed between an ego and alters in defining the size of the sphere
  - E.g., level 1 – all nodes one away from ego and the connections among them - typical ego net
  - E.g., level 2 – all nodes two away from ego and the connections among them
How to Influence

- Who are they connected to
- What groups are they in
- What do they know
- What resources do they control
- What activities are they involved in

```
<table>
<thead>
<tr>
<th>Node Type</th>
<th>Size</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>agent</td>
<td>8</td>
<td>+50%</td>
</tr>
<tr>
<td>knowledge</td>
<td>2</td>
<td>+50%</td>
</tr>
<tr>
<td>location</td>
<td>2</td>
<td>+3%</td>
</tr>
<tr>
<td>resource</td>
<td>1</td>
<td>+25%</td>
</tr>
<tr>
<td>task</td>
<td>1</td>
<td>+20%</td>
</tr>
</tbody>
</table>
```

**ORA:** Sphere of Influence

<table>
<thead>
<tr>
<th>Name</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>alias</td>
<td>wadih_hage</td>
</tr>
<tr>
<td>hostility_level</td>
<td>1</td>
</tr>
<tr>
<td>joined_al_qaeda</td>
<td>1989</td>
</tr>
<tr>
<td>left_al_qaeda</td>
<td>1998</td>
</tr>
<tr>
<td>nationality</td>
<td>lebanese</td>
</tr>
<tr>
<td>nationality_relation</td>
<td>hostile</td>
</tr>
<tr>
<td>nyl</td>
<td>prosecutor</td>
</tr>
<tr>
<td>source_date</td>
<td>2006-05-10</td>
</tr>
<tr>
<td>suspected_terrorist</td>
<td>yes</td>
</tr>
</tbody>
</table>

Most similar other – jamal al-fadil
Network Analysis

- Derives from graph theory
- Set of measures on graphs or networks
- Graph - binary matrix
- Network - weighted matrix
Terminology

- **Node**
  - entity, <person>, dot, point, actor

- **Tie**
  - relation, link, edge, connection, <friendship>
  - vary in strength (weight), direction, type, confidence (another weight)

  Caveat: 80-90% of work/measures uses binary data

```
0 1 0 1 0
1 0 0 1 0
1 0 0 0 1
0 0 1 0 1
0 1 1 0 0
```

```
0 2 0 5 0
1 0 1 0 1
1 0 0 7 0
0 0 8 0 1
0 2 4 0 0
```

- **Mode** – types of nodes (number of entity classes)
  - Traditional SNA uses 1 mode data (sometimes 2)
  - Traditional Link analysis uses multi-mode

- **Multi-plex** – types of links
  - A multi-plex data set has multiple relations among nodes of the same mode/entity class
  - Most SNA data sets of single plex
  - Traditional Link analysis uses multi-plex data

- “Way” means dimensions: rows, columns, levels, etc.
  - Most SNA data sets are 2-way (row by column)
  - Most over time data sets are 3 way (1 matrix per time)

- **Meta-Network**
  - A set of networks defined over multiple entity classes, both multi-mode and multi-plex
  - Can be multi-way also

- E.g., 3-way, 1-mode, single-plex
  - Perceived social networks (CSS)
    - CSS – cognitive social structure
    - Each person gives their perception of who knows whom
    - Transactive memory of social relations

- E.g., 3-way, 3-mode, multi-plex
  - Transactive memory (over actors, knowledge, tasks) for existent and desired relations
Terminology

- Rij means i sends relation R to j, i is row and j is column
- A tie is defined as existing if Rij != 0
  - Caveat: most work does not differentiate a missing tie from a known non-existent tie
- Symmetry
  - This is defined about the diagonal
  - A network is symmetric if Rij = Rji

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Symmetry achieved by Union of Rij and Rji

Conventions

- Most measures assume the diagonal is 0
- If multiple types of ties, they are typically binarized, summed, and re-binarized
  - Binarization typically uses a > mean or in top third rule
- Many measures assume data is symmetric
  - Let R' be the new matrix and R the original
  - Strong agreement (local aggregation)
    - R'ij = R'ji = 1 if either Rij & Rji = 1
  - Weak agreement
    - R'ij = R'ji = 1 if either Rij or Rji = 1
- Only square matrices are analyzed
  - Common: Drop columns for nodes for which there is missing information
  - Alternative: Include columns but fill with 0’s
  - Alternative: Assume missing data is random and fill it to reflect the average
Network Measures – Analysis Levels

- Network (complete graph) level
  - E.g., density
  - Is it easier to disrupt a cellular or hierarchical structure?
  - Use: Characterizing topology, comparing groups, high level change

- Dyadic level
  - E.g., frequency
  - Is there a pattern that money launderers follow?
  - Use: Locating trails

- Node level
  - E.g., centralities
  - Who has the power?
  - Use: Identifying key actors, events, resources ...

Individual v. Network Level

- Individual behaviors are not independent of the network within which the behaviors occur.
- Individual network position is not independent of the network structure – being central in a centralized network is different than being central in a decentralized network.
Key Graph Theoretic Concepts

- Directed –versus- undirected
  - Directed – commands
  - Undirected – works with
- Strength
  - Frequency of interaction
  - Distance
- Adjacency
  - Equivalent matrix
- Walk; length
  - Unrestricted; number of ties
- Path
  - Do not repeat a node
- Trail
  - Do not repeat a tie
- Distance
  - Shortest path (geodesic)

FLOW: Walks, Trails, Paths

- Path: can’t repeat node
  - 1-2-3-4-5-6-7-8
  - Not 7-1-2-3-7-4
- Trail: can’t repeat line
  - 1-2-3-1-7-8
  - Not 7-1-2-7-1-4
- Walk: unrestricted
  - 1-2-3-1-2-7-1-7-1
- But – different things flow differently through networks
  - Gift process
  - Currency process
  - Transport process
  - Postal process
  - Gossip process
  - E-mail process
  - Infection process
  - Influence process
Flow Terminology

- Path: the set of nodes and edges needed to get from one node to another
  - Paths for B: BA, BD, BDC, BDCE ...
  - Path length: the number of steps/links
    - Popularly known as degree – 6 degrees of separation
    - BA 1, BDC 2, BDCE 3

- Cycle: a path that crosses the same node
  - BDCEB

- Distance between two nodes is the length of the shortest path (aka geodesic)

- Shortest path: the path between two nodes that has the fewest links

Ego Nets

- Ego Network:
  - All ties to/from ego and the connections among the associated nodes in a two-way, single mode, single plex data set
  - Be careful – some authors only use the ties to/from B
  - Nora – finds these limited ego nets
  - By convention – typically only used for actor or organization entity classes

- B’s ego Net: 0 1 0 1 0
  - A – B – C
  - D – E
  - A – B – D
  - A – B – E
  - B – C – D – E

- A’s ego Net: 1 0 0 1 0
  - A – B
  - A – D
  - A – E

- B’s ego Net: 1 0 0 0 1
  - B – A
  - B – C
  - B – D
  - B – E

- A’s ego Net: 0 0 1 0 1
  - A – B
  - A – D
  - A – E

- B’s ego Net: 0 1 1 0 0
  - B – A
  - B – E
Meta Network Approach to Organizational Representation

<table>
<thead>
<tr>
<th>People Relation</th>
<th>Knowledge Relation</th>
<th>Resources Relation</th>
<th>Tasks Relation</th>
<th>Organizations Relation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social Network</td>
<td>Knowledge Network</td>
<td>Capabilities Network</td>
<td>Assignment Network</td>
<td>Work Network</td>
</tr>
<tr>
<td>Who knows who</td>
<td>Who knows what</td>
<td>Who has what resource</td>
<td>Who does what</td>
<td>Who works where</td>
</tr>
<tr>
<td>Knowledge</td>
<td>Information Network</td>
<td>Skills Network</td>
<td>Needs Network</td>
<td>Competency Network</td>
</tr>
<tr>
<td>Relation</td>
<td>What informs what</td>
<td>What knowledge is needed to use what resource</td>
<td>What knowledge is needed to do that task</td>
<td>What knowledge is where</td>
</tr>
<tr>
<td>Resources Relation</td>
<td>Substitution Network</td>
<td>Requirements Network</td>
<td>Capital Network</td>
<td>What resources are where</td>
</tr>
<tr>
<td>What resources can be substituted for which</td>
<td>What resources are needed to do that task</td>
<td>What resources are where</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tasks Relation</td>
<td>Procedure Network</td>
<td>Market Network</td>
<td>Inter-Organizational Network</td>
<td></td>
</tr>
<tr>
<td>What tasks must be done before which</td>
<td>What tasks are done where</td>
<td>Which organizations link with which</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
SNA Insufficient

- Centralities
  - Communication
    - Degree – most connected
    - Betweenness – most paths
- Exclusivities
  - Expertise
    - Knowledge – special expertise
    - Task – special experience
- Demands/Loads
  - Roles
    - Cognitive demand – emergent leader
    - Workload

Metrics come in Modes!

- 1 Mode Metrics
  - E.g. all centralities
  - Assumes network is square – and rows=columns e.g. social network
- 2 Mode Metrics
  - E.g. all redundancies
  - Assumes you have two types of nodes e.g. agents by events
- N Mode Metrics
  - E.g. cognitive demand
  - Assumes you have N types of nodes e.g. entire meta-network
Two-Mode Data

- “Indirect” Connection
  - Co-Publishing, memberships in clubs, boards of companies
  - Events
- Shared ...
  - Knowledge, Resources
  - Tasks, Locations

Structural Analysis

- Overall Process
- Tools
- Applications
The Overall Network Analysis Process

Get the Data       Analyze & Visualize       Forecast

The Basic Work Flow

STEP 1 – Data Collection
Socio-Cultural Data is Every Where

• Unstructured
  – Text – e.g. interviews, news articles, blogs, email
  – Various on-line sources

• Semi-structured
  – Blogs
  – Emails
  – Crowd-sourced

• Structured
  – Government and corporate documents
  – Proceedings

• Unstructured
  – Sudan Tribune Review
    • 2003 - 2932
    • 2004 - 6943
    • 2005 - 3828
    • 2006 - 3828
    • 2007 - 5815
    • 2008 - 9266

• Semi-structured
  – UN Reports
  – IDA Study 1796 files

• Structured
  – African gazateer
Steps in a Structural Analysis

1. Collect network data.
   - Connections among people, knowledge, resources, events …
2. Enter data into ORA.
3. Visualize.
5. If multiple networks create combined measures.
6. If needed look at some measures more indepth.
7. Possibly drop isolates and pendants
8. Check interpretations.

Text Mining to Extract Networks

- Network Text Analysis --- Encode links between words in texts and construct network of linked words
- Content Extraction (a.k.a. Content Analysis)
- Semantic Network Extraction (a.k.a. Mental Model Analysis)
- Meta-Network Extraction (a.k.a. Structural Analysis)
- Belief Extraction (a.k.a. Sentiment Analysis)
Example Text Files

Hisham Al Hussein
... the Philippine government booted the second secretary at Iraq's Manila embassy, Hisham Al Hussein, on February 13, 2003, after discovering that the same mobile phone that reached his number on October 3, 2002, six days later rang another cell phone strapped to a bomb at the San Roque Elementary School in Zamboanga.

Abu Madja and Hamsiraji Ali
That mobile phone also registered calls to Abu Madja and Hamsiraji Ali, leaders of Abu Sayyaf, Al Qaeda's Philippine branch.

Abdurajak Janjalani
It was launched in the late 1980s by the late Abdurajak Janjalani, with the help of Jamal Mohammad Khalifa, Osama bin Laden's brother-in-law.

Hamsiraji Ali
... Hamsiraji Ali, an Abu Sayyaf commander on the southern island of Basilan, bragged that his group received almost $20,000 annually from Iraqis close to Saddam Hussein.

Resultant Meta-Network

[Diagram showing connections between entities such as Janjalani, Jamal Khalifa, Hisham Hussein, Abu Madja, Hamsiraji Ali, Saddam Hussein, Bin Laden, Philippine, Basulan, Osama Bin Laden, Al Qaeda, Bomb, Phone, and School.]
From the Play – Julius Caesar by Shakesphere

Data to Model – D2M

- Converting texts to networks is now easier to do following the D2M approach
- Key 1: Each text is separate and a DyNetML will be created from it – use ORA to union and assess
- Key 2: Following D2M you will create a unified named entity and generalization thesauri and use these to process all texts
  - AutoMap contains tools to suggest ontological categories
  - As you develop thesauri – you save and reuse – and so don’t have to re-classify later
STEP 2 – Cleaning the Corpus

- Cleaning of data
  - Removal of navigation, headers and footers (information not pertinent to the article)
  - Remove non textable files
    • E.g. remove maps/scans
  - Converted PDF to .txt
    • Most but not all can be converted with CASOS tool
  - Convert RTF to .txt
    • Currently semi-manual process

- Semi-automatic cleaning of the corpus which is done manually (optional) and automatically
  - Involves the entire corpus as a whole not individual texts
  - Removed word wrapping
  - Run automated cleaning

STEP 3 - Deduplication

- Deduplication
  - Removal of repeated articles
  - Reduces the number of files and allows a more compact analysis
  - Near Miss procedure is best
- Performed Once
- Time depends on number of texts and length
- No deduplication was done on SNARC data as all files unique

<table>
<thead>
<tr>
<th>Illustration of Impact of typical Deduplication - Number of texts before and after deduplication applied only to Sudan data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sudan text</td>
</tr>
<tr>
<td>concepts</td>
</tr>
<tr>
<td>Average frequency per concept</td>
</tr>
</tbody>
</table>
Illustration of Impact of Deduplication: Note Deduplication Can Impact Importance of Concepts

<table>
<thead>
<tr>
<th>Concept</th>
<th>Count</th>
<th>%</th>
<th>Concept</th>
<th>Count</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valencia</td>
<td>1141455</td>
<td>6.55</td>
<td>Valencia</td>
<td>853691</td>
<td>11.65</td>
</tr>
<tr>
<td>conflict_task</td>
<td>867688</td>
<td>4.98</td>
<td>conflict</td>
<td>585801</td>
<td>7.99</td>
</tr>
<tr>
<td>nanuque</td>
<td>500411</td>
<td>2.87</td>
<td>republic_of_the_sudan</td>
<td>332082</td>
<td>4.53</td>
</tr>
<tr>
<td>ampere</td>
<td>448679</td>
<td>2.88</td>
<td>conflict_task</td>
<td>207738</td>
<td>2.83</td>
</tr>
<tr>
<td>republic_of_the_sudan</td>
<td>385560</td>
<td>2.21</td>
<td>wilayat_darfur</td>
<td>153976</td>
<td>2.1</td>
</tr>
<tr>
<td>wilayat_darfur</td>
<td>344036</td>
<td>1.97</td>
<td>political</td>
<td>113992</td>
<td>1.56</td>
</tr>
<tr>
<td>valence_task</td>
<td>178629</td>
<td>1.03</td>
<td>Sudanese</td>
<td>94010</td>
<td>1.28</td>
</tr>
<tr>
<td>ner_population</td>
<td>178059</td>
<td>1.02</td>
<td>Khartoum</td>
<td>72409</td>
<td>0.99</td>
</tr>
<tr>
<td>faouzi_ben_mohamed_be_ahmed_a</td>
<td>172782</td>
<td>0.99</td>
<td>valence_task</td>
<td>62440</td>
<td>0.85</td>
</tr>
<tr>
<td>badou</td>
<td>152547</td>
<td>0.88</td>
<td>environment</td>
<td>60138</td>
<td>0.82</td>
</tr>
</tbody>
</table>

STEP 4 – Automated Text Cleaning

- Removed stand-alone numbers
- Removed extra space
- Fixed common typos
- Removed extra white space
- Expanded contraction and abbreviations
- Removed individual letters not in names
- Converted British to American spelling
- Pronoun resolution
- Converted common hyphenated forms and non-hyphenated to common form e.g. Major-General and Major General
- Generalization using standard plus the named entities
- Removed noise words
- Ran standard ngram conversion

This can be done with AutoMap
Additional Aspects of Automated Text Cleaning

Text preparation: a completely automatic process

- Creation of a thesauri of stemmed and non-stemmed version of nouns and verbs
  - Detensing: Reduce all verbs to their present tense
  - Depluralization: Eliminates the plural form and reduces it to its base form

- Apply an n-gram thesauri to convert multi-words to single concepts

- Delete noise/stop words
  - Prepositions
  - helping verbs
  - verb of being
  - remaining pronouns

Illustrative Results for Impact of Specialized Stemmers

<table>
<thead>
<tr>
<th>Original</th>
<th>After</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sudan</td>
<td>97492</td>
<td>85758</td>
</tr>
<tr>
<td>Catnet</td>
<td>24743</td>
<td>22091</td>
</tr>
<tr>
<td>Singapore</td>
<td>5073</td>
<td>4452</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nouns Before</th>
<th>Nouns After</th>
<th>Verbs Before</th>
<th>Verbs After</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sudan</td>
<td>28488</td>
<td>23680</td>
<td>12006</td>
</tr>
<tr>
<td>Catnet</td>
<td>7693</td>
<td>6838</td>
<td>4754</td>
</tr>
<tr>
<td>Singapore</td>
<td>1677</td>
<td>1445</td>
<td>1213</td>
</tr>
</tbody>
</table>
Step 5 – Named Entity Identification

- Extraction and Identification of named entities:
  - Tag by part of speech
  - Identify proper nouns and n-grams that are proper nouns
- Result: thesauri of named entities
- Prior study showed that human processing time reduced by 80% to 99% by using this approach

<table>
<thead>
<tr>
<th>Number of Named Entities</th>
<th>Texts</th>
<th>People</th>
<th>Organizations</th>
<th>Locations</th>
<th>Events</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNARC</td>
<td>265</td>
<td>33,210</td>
<td>5,523</td>
<td>2,807</td>
<td>120</td>
</tr>
</tbody>
</table>

STEP 5 – Includes Ontological Cross Classification and Thesauri Construction

- Apply standard thesauri and ontological categories
- Applying the standard thesauri converts all multiple-concepts words into a single word/concept already classified
- Ontological classes are suggested using
  - Parts of speech and statistical regularities
Step 6: Named Entity Resolution

- At the same time that the named entities are extracted a meta-network is extracted.
- Named Entity list
  - Contains all concepts/ngrams guessed to be people, organizations, locations, events with best guess
- Meta-network
  - Contains all people, organizations, locations, knowledge, events, activities, resources, beliefs based on existing standard thesauri
  - The specific people, organizations, locations and events in this are viewed as “vetted”
- The vetted list is removed from the named entity list and the vetted class is used
- Humans then go through the remaining named entities to classify those that make sense
- Step 5 and 6 are repeated as needed
- End result is a vetted meta-network

Illustrative Results from Named Entity List Before Resolution

<table>
<thead>
<tr>
<th>conceptFrom</th>
<th>conceptTo</th>
<th>metaOntology</th>
</tr>
</thead>
<tbody>
<tr>
<td>up missions targeting taliban leaders</td>
<td>up_missions_targeting_taliban_leaders</td>
<td>agent</td>
</tr>
<tr>
<td>distribute new</td>
<td>distribute_new</td>
<td>agent</td>
</tr>
<tr>
<td>main office california office 1899 l street</td>
<td>main_office_california_office_1899_l_street</td>
<td>location</td>
</tr>
<tr>
<td>pay afghan</td>
<td>pay_afghan</td>
<td>agent</td>
</tr>
<tr>
<td>military commander</td>
<td>military</td>
<td>organization</td>
</tr>
<tr>
<td>once u</td>
<td>once_u</td>
<td>agent</td>
</tr>
<tr>
<td>peace press washington</td>
<td>peace_press_washington</td>
<td>agent</td>
</tr>
<tr>
<td>david katz</td>
<td>david_katz</td>
<td>agent</td>
</tr>
<tr>
<td>david kilcullen</td>
<td>david_kilcullen</td>
<td>agent</td>
</tr>
<tr>
<td>david lachapelle</td>
<td>david_lachapelle</td>
<td>agent</td>
</tr>
<tr>
<td>david lanz</td>
<td>david_lanz</td>
<td>agent</td>
</tr>
</tbody>
</table>
ORA: a DNA statistical analysis tool for locating patterns and identifying vulnerabilities

- Organized by function not measure; e.g.,
  - Key Entity Report
  - Group Locator Report
- Import/Export tools
- Linkage to mysql
- Visualization components
- Batch, web, thick-client
- Can handle large $10^6$ networks quickly

ORA

- Network analytic and visualization toolkit
- Statistical, graph-theoretic and visual analytics for complex network data
  - Special tools for analyzing and visualizing dynamic networks
  - Special tools for geo-enabled network analysis and visualization
  - Supports 1-mode, 2-mode and meta-network calculations
- Illustrative functions for social networks or link analytic data
  - Ability to identify groups, key nodes, hidden links, critical locations
  - Ability to identify patterns of interest
  - Ability to visualize networks (2D and 3D), visualize change and network dynamics, visualize networks on maps, visualize network attributes by region (e.g., precinct level statistics)
  - Ability to analyze topics, topic groups, and identify changes in topics
- Often referred to just as ORA
ORA in Operation

ORA Visualization

- Network analytics
- Network visualization 1D & 1D
- Big data > 2 mil nodes
- High dimensional networks
- Spatial networks
- Geo-spatial network analytics
- Dynamic network analyses & visualizations
- Crowd group detection & hidden networks
- Social media analytics
- Key actors
- Community detection
- Compare & contrast networks
- Team assessment
- Advanced graphics
- Designed to integrate with other tools
- Extensive help & support

Networks maps
**ORA does Dynamic Network Analysis**

- Trends for metrics
- Trends for nodes
- Through time visualization
- Comparative statistics for two time periods
- Forensic change detection
- Fourier analysis for pattern of life assessment and anomaly detection
- Immediate impact assessment for network change
- Near term forecasting for network change

**ORA does Geo-Enabled Dynamic Network Analysis**

- Complex "meta" networks
  - who, what, where, how, why and when
- Key Features
  - Network visualization
  - Network analytics
    - Key actor, location, idea ...
  - Network dynamics
    - Network comparison
    - Change detection
    - Trend analysis
  - Path identification
  - Link inference
  - Trail analysis and visualization (who was where when)
  - Pattern analysis
    - Extraction of patterns of particular importance in law enforcement such as working cells
    - Node and link attribute assessment
Using Constraints to Set Strength of Links has Practical Consequences

Geo-Network Measures Improve focus

Metrics and Grouping Capabilities

- Over 200 metrics
- Key node
  - Centralities
  - Loads
  - Redundancies
- One-mode, two-mode and meta-network metrics
- Grouping algorithms
  - Concor
  - Newman-Girvan
  - Fog
  - Johnson-hierarchical
  - Louvaine
  - Spectral
  - MDS
  - LDA
  - LSA
Other Features

- Matrix algebra techniques
- Stylized network generation
- Supports editing for
  - Nodes
  - Edges
  - Node attributes
  - Network attributes
- Metrics organized into workflow reports
- Extensive help
- Quick-start guide
- ORA google groups

Interoperability & Data Control

- Data entry
  - Import wizard
  - Direct entry through visualizer
  - Ability to import from many packages
    - E.g. ucinet, i2, mysql, csv, tsv, tweettracker, blogtracker
    - Any csv, tsv, or json file can be imported
  - Export visualization in multiple formats e.g. .png, gif ...
  - Ability to generate maps for ArcGIS, google-earth, NASA Worldwind
  - Data output in HTML and CSV
  - Data editing, cleaning, anonymization and manipulation capabilities
  - Companion tools for processing texts and twitter data
Identify: Who are the Key Players? Or Locations, Resources ...

**In-the-Know (total degree centrality)**

<table>
<thead>
<tr>
<th>Rank</th>
<th>Value</th>
<th>Unnamed Agent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.182</td>
<td>181</td>
</tr>
<tr>
<td>2</td>
<td>0.170</td>
<td>170</td>
</tr>
<tr>
<td>3</td>
<td>0.166</td>
<td>166</td>
</tr>
<tr>
<td>4</td>
<td>0.160</td>
<td>160</td>
</tr>
<tr>
<td>5</td>
<td>0.158</td>
<td>158</td>
</tr>
<tr>
<td>6</td>
<td>0.154</td>
<td>154</td>
</tr>
<tr>
<td>7</td>
<td>0.139</td>
<td>139</td>
</tr>
<tr>
<td>8</td>
<td>0.115</td>
<td>115</td>
</tr>
<tr>
<td>9</td>
<td>0.111</td>
<td>111</td>
</tr>
<tr>
<td>10</td>
<td>0.109</td>
<td>109</td>
</tr>
</tbody>
</table>

**Potentially Influential (betweenness centrality)**

<table>
<thead>
<tr>
<th>Rank</th>
<th>Value</th>
<th>Unnamed Agent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0088606</td>
<td>266.72 986</td>
</tr>
<tr>
<td>2</td>
<td>0.0705427</td>
<td>1700.4 97</td>
</tr>
<tr>
<td>3</td>
<td>0.0625258</td>
<td>1600 95</td>
</tr>
<tr>
<td>4</td>
<td>0.0469338</td>
<td>1500 92</td>
</tr>
<tr>
<td>5</td>
<td>0.0548627</td>
<td>1684.2 652</td>
</tr>
<tr>
<td>6</td>
<td>0.0541431</td>
<td>1913.4 635</td>
</tr>
<tr>
<td>7</td>
<td>0.0524498</td>
<td>1306.2 412</td>
</tr>
<tr>
<td>8</td>
<td>0.0410668</td>
<td>1041.6 505</td>
</tr>
<tr>
<td>9</td>
<td>0.0160678</td>
<td>764.12 273</td>
</tr>
<tr>
<td>10</td>
<td>0.0305132</td>
<td>762.79 552</td>
</tr>
</tbody>
</table>

Drilling down...

*ORA's Key Entity Report shows 3 agents critical to operations.

Narrow our focus from set of interstitial members to small group of leaders.

Assess: How are they organized?

**FOG (Fuzzy Group Clustering) shows suspicious entities organized into 5 groups w/shared members.**

Interstitial members are likely to contain coordinators & leaders.
Illustrative Problems

Gang Interaction
- Two gangs are in the regions
- Recent reports suggest they are working together
- Potential problems
  - Increased drug trafficking
  - Coordinated response to law enforcement activity
  - More dangerous
- What is the connection?
- Who is connecting them?

Targeting
- Which individual or group to isolate to achieve maximal effect
- How to influence
  - Are there important connections
- Who to target (vulnerabilities)
  - What groups or individuals stand out
- What is the immediate effect of a COA
  - On Diffusion, Performance, Leadership

Dynamic Network Analysis Locates Effective Single Strike Target

Commander’s Choice (in the news)

Model Choice

Figure 1: Notional: Terrorist network before isolation of critical actor. Fragmentation: 0 – This is a single unified organization. Diffusion: 0.93 – Messages and resources passed from actor to actor can move quickly.

Figure 2: Notional: Terrorist network after isolation of critical actor. Fragmentation: 0.71 – Now there are many isolated sub-groups. Diffusion: 0.26 – Messages or resources are passed from actor more slowly.
Are Two Critical Actors Linked? Path Finder

\[ \text{lariq_al-fadil} \]

\[ \text{mohammed_atta} \]

Real-world Win! – The Context

- Training Tulsa Police Department Special Investigations Division and Oklahoma Bureau of Narcotics
- Unknown to the CASOS team, the gang unit of a major-city police force had evidence of drug supply-chain that indicated some abnormal cooperation between two dissimilar street-gangs
- Law enforcement had no leads on who was the connection
- The CASOS team was demonstrating ORA
  - to a few officers
  - utilizing a small sample of data from the departments live arrest and surveillance database
Real-world Win! Event!

- During the demonstration of ORA’s pathfinder feature
  - Officers asked the CASOS team to show the path between two gangs
- Voila!
  - Using ORA a human connection between the two gangs was quickly found
  - This human had not been readily-apparent to the officers
  - But the information was buried in their database
- The investigating officer was called into the room to see the newly discovered finding
- This first link noticed proved to be useful!
- In addition
  - Other links have been found in follow-up analyses
  - Gang investigators from his Special Investigations Division are following these additional leads

Two Gangs

Notice – there are two gangs with very different members
Internal Structure

Notice – the two gangs have different structures

Are the gangs connected? Yes!
Change in Networks

- Natural Evolution
- Recruitment
- Learning
- Isolation
- Change Detection – ORA
- Comparison – ORA
- Change in "keyness" – Metric Dynamics - ORA
- Immediate Impact – ORA
- Near term Impact – ORA-Construct

Change Detection - Objective

How can we quickly identify changes in social networks subject to a specified risk of false alarm?

- Al Qaeda Data (Based on Sageman, 88-04)
- Graduate Students (Email over 24 weeks, 07)
Al-Qaeda Application

The control chart signals a change in the network in 2001.

<table>
<thead>
<tr>
<th>Year</th>
<th>Closeness</th>
<th>Z</th>
<th>C+</th>
<th>C-</th>
<th>Signal</th>
</tr>
</thead>
<tbody>
<tr>
<td>1994</td>
<td>0.0027</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1995</td>
<td>0.0003</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1996</td>
<td>0.0028</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1997</td>
<td>0.0031</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1998</td>
<td>0.003</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1999</td>
<td>0.0032</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>0.0034</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2001</td>
<td>0.0024</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2002</td>
<td>0.0015</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2003</td>
<td>0.0004</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Most Likely Estimate of the change point is 1997:
- Re-establish base in Afghanistan
- Bright Star '97 cut short
- Feb '98 Islamic Front
- Embassy bombings in '98

1997 was a critical planning year for Al-Qaeda

Compare and Contrast

- Marriage ties
- Nodes sized by betweenness and colored by degree
- Banking ties
- Nodes sized by betweenness and colored by degree

Use Tiles in ORA visualizer to Show Differences
Simple Comparison – Node Level

Forecasting

The “Holy Grail” is forecasting

Lots of Data: Use it!
- Social Kinetic Theory – that groups, not people, are predictable using statistics
- Tomorrow is not like Today

Logics of Adaptation
- Social Quantum Theory – that the logic of individual adaptation drives social change
- Result: statistically informed adaptive model
Dynamics

Social Dynamics due to Learning

- Implicit link
  - seen together
  - common sources
  - seniority
- Explicit link
  - information exchange
  - learned from each other
  - mentoring

- When meeting a new person
  - Infer expertise based on implicit links
  - Baseline for trust
  - Social shakeout occurs as you move from implicit to explicit links
Rates of Adaptation

<table>
<thead>
<tr>
<th>People</th>
<th>Expertise</th>
<th>Activities</th>
<th>Locations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fast</td>
<td>Learning Based</td>
<td>Social-Technological Based</td>
<td></td>
</tr>
</tbody>
</table>

% Shared

- Information
- Activity
- Beliefs

Time

Forecast - Construct – from Patterns and Identified COA to Near Term Impact

Near Term Impact Report

Generic Performance

Bin Laden
Yassin al-Zawahiri
Rantisi
Meshaal
Damascus
Regional organization
Functional cells
Lebanon
Judea
Samaria
Operation
Operation
Operation
Operation
Operation
Operation
Gaza (HQ)
West Bank
Other regions
Near Term Impact Report

June 2018 Copyright © 2018 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU
Approach to Delivering the Message Matters

Mean Strength of pro-US Belief

Anti Messengers use all Media

Anti Messengers use a Media

substantial differences by cities

However, edge inference based on meta-network theories is better!

- **Type 1:**
  - Predicted no edge but there is an edge

- **Type 2:**
  - Predicted an edge but there is no edge

But – no theory is great – Type 2 twice as likely as type 1
Interaction Style: Need for Communicative Ease - Homophily

- Relative similarity = how much i shares with j divided by how much i shares with all others
- AKik is knowledge network
  - Knowledge network is agent by knowledge (“facts”)
- Expected interaction based on relative similarity

\[
RS_{ij} = \frac{\sum_{k=0}^{K} (AK_{ik} \times AK_{jk})}{\sum_{j=0}^{I} \sum_{k=0}^{K} (AK_{ik} \times AK_{jk})}
\]

Cutoff = \(\sum_{i=0}^{I} RS_{ij} / (I \times (I - 1))\)

If \(RS_{ij} \geq \text{Cutoff}\) the Expected interaction = 1
else 0

Relative Similarity - Why

- Similarity: individuals tend to interact with those whom they deem to be more similar to themselves
  - Comfort
  - Ease of interaction
  - Ease of access
  - Common language
  - More effective for getting information
  - Shared expectations about reciprocity
- Relative: individuals judge similarity relative to others
  - There is a comparison group
  - There is a generalized other
Example Applications

- Public Health – Reorganization
- Service Science – Designing new teams
- Sociology of Science – Citation networks
- Organizational performance – Enron
- Merchant marine vessels – piracy
- Counter-narcotics – marijuana traffic
- Counter-terrorism – Emergent leaders and adaptation
- Pandemic Influenza – Intervention – school closure doesn’t work
- Literature – The structure of plays and movies
- Aviation – Critical airports
- Medicine – Team designs to prevent falls
- Health – Coordinated care
- MMOGs – Many ways to succeed

Tools and Workflows Exist and Are Improving for extracting, analyzing, forecasting, ...
Let’s Go!