CASOS

Welcome to the CASOS Summer Institute

Prof. Kathleen M. Carley

kathleen.carley@cs.cmu.edu

Carnegie Mellon

**Center for Computational Analysis of
Social and Organizational Systems**
http://www.casos.cs.cmu.edu/

isr institute for SOFTWARE RESEARCH

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

Goal

Learn:
- Gain a basic understanding of network science
  - social network analysis, dynamic network analysis, …
- Gain an understanding of the uses of network analytics
- Learn the basic vocabulary
- Develop an understanding of the fundamental measures.
- Develop an understanding of how to interpret network results
- Gain experience using tools for dynamic network analysis
- Understand network analytic toolchains from text to prediction
- Understand network analytics for social media

NOT about programming, data collection, …

CASOS

CASOS

CASOS

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
    - Near term analysis - ORA (and Construct underneath)
    - Simulation – Construct, ORA Micro-Sim
  - Get the results out!
    - Exporting: CSV, HTML, export to other network tools, pdfs, tiff, …

CASOS

 3

Logistics

- CMU CASOS team
- Get tools on your machine
- On-line
  - Slides
  - Papers
  - Data
  - Tools
- Virtual Reception
- Virtual Poster Session
- Breakouts – see questionnaire each day
- Overall – this has been organized in a "flipped" fashion. All sessions are taped.

CASOS

 4

CASOS

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

Networks
Networks
Networks

Prof. Kathleen M. Carley

kathleen.carley@cs.cmu.edu

Dartmouth 1975
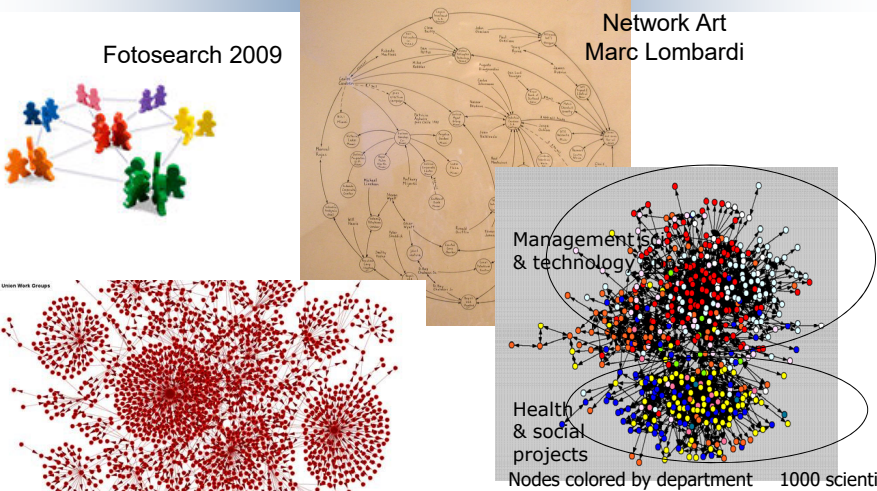
MSSB Social Networks Symp., Hanover, N.H., Sept. 18-21, 1975

6/7/2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU 7

Networks!

Fotosearch 2009

Network Art
Marc Lombardi

Management science
& technology projects

Health
& social
projects

Nodes colored by department 1000 scientists

Thanks to Steve Borgatti 2004

Benghazi Twitter Network 2012 – Kathleen M. Carley et al.

6/7/2020 Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU 8

Social Networks / Network Science
Fast-Growing Scientific Area

- In past 40 years growth has been exponential
- Popular culture
  - Games, plays, television
  - Forbes, Fortune, NY Times

6 Degrees of Separation

- 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

---

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

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

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

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

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, management 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
  - E.g., facebook, twitter
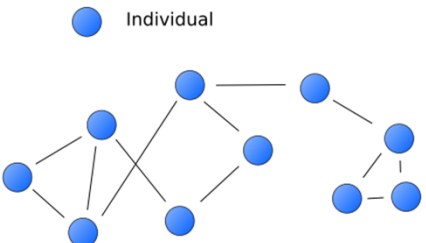
Traditional Social Network Analysis

- 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.
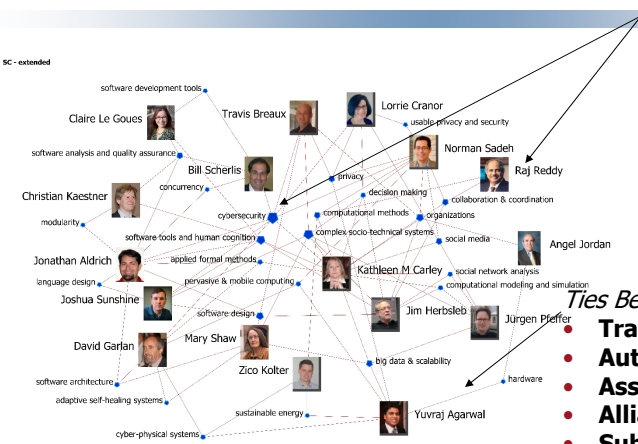
Individual

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**

CASOS

Connect & *Dis*-Connect the Dots!

	Degree	Betweenness	Closeness
1	0.417 Mohamed Atta	0.334 Nawaf Alhazm	0.571 Mohamed Atta
2	0.389 Marwan Al-Shehhi	0.318 Mohamed Atta	0.537 Nawaf Alhazmi
3	0.278 Hani Hanjour	0.227 Hani Hanjour	0.507 Hani Hanjour
4	0.278 Nawaf Alhazmi	0.158 Marwan Al-Shehhi	0.500 Marwan Al-Shehhi

Flight AA #11 - Crashed into WTC North
Flight AA #77 - Crashed into Pentagon
Flight UA #93 - Crashed in Pennsylvania
Flight UA #175 - Crashed into WTC South
Other Associates of Hijackers
Copyright © 2001, Valdis Krebs

Figure 3 Trusted Prior Contacts + Meeting Ties [shortcuts]

Standard Social Network Measures

CASOS

Why is Network Analysis Important

- Networks gives 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

CASOS

CASOS

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

Evolutionary Themes

Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

Small Networks

Big Networks

DNA

Animal Networks

Herds   Neural

Groups   Organizations

Communities   Images

Social

Communication

1940's & before

Ties   People   Semantic   Events   Topics   Inferred
Organizations

Animals

1 Mode   2 Mode   N Mode

Graphs   Statistics   Graph Algorithms

Heuristic

Mass Media   email   Social media

Static   Dynamic

Snow Ball Sampling

Questionnaires

Observation   Sensor

What is Dynamic Network Analysis?

Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

- 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

CASOS

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

CASOS

Key Issues In Assessing Network Data

- Scalability
  - Networks may be large 106 or 107 nodes and sparse
  - Network may be moderate in size but extremely dense
  - And, there are multiple types of nodes
- Robustness
  - Missing data is common
- Flexibility
  - Relations among nodes are flexible
  - Relations among nodes are time variant
  - Relations among nodes vary in strength
- Error
  - Relations among nodes vary in certainty
  - Node membership may be questionable
  - Classes of data may not be discoverable

Typical Mainstream Data Structure

Variables (attributes)

Cases (individuals)

	Age	Gend	Education	Income	State
1001	28	0	< highschool	28,000	PA
1002	35	1	highschool	54,000	MA
1003	26	0	< highschool	26,000	MA
1004	40	1	Bachelors	65,000	MA
1005	24	0	highschool	27,500	PA
1006	55	1	Ph.D.	82,800	PA
1007	31	1	Ph.D.	73,000	MA
1008	M	0	highschool	33,500	PA
…					

Analysis consists of correlating attributes, regression, anova …

The Network Perspective

Standard Statistics	Social Network Analysis	Dynamic Network Analysis
Attributes	Relations	Relations + Attributes
Atomistic	Interdependence	
Actors as independent	Actors constrained and enabled by links	Actors constrained and enabled by links
Actor state matters	Actor state irrelevant	Actor state impacts perception of and use of links

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

Network data is about relations

	Jenny	Martin	Chloe	Ramesh	Lenore	Stephen
Jenny	0	1	1	0	0	0
Martin	1	0	0	1	1	1
Chloe	0	0	0	1	0	0
Ramesh	0	1	1	0	0	1
Lenore	1	1	1	1	0	1
Stephen	0	1	0	1	1	0

Slide 27

Carnegie Mellon
ISr institute for SOFTWARE RESEARCH

Network Data Structures

Adjacency matrices

Friendship

	Jenny	Martin	Chloe	Ramesh
Jenny	-	1	1	0
Martin	1	-	0	1
Chloe	0	0	-	1
Ramesh	0	1	1	-

Shared Concepts

	Jenny	Martin	Chloe	Ramesh
Jenny	-	5	35	21
Martin	5	-	3	24
Chloe	35	3	-	3
Ramesh	21	24	3	-

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

Incidence matrix

	Friendship	Shared Concepts
Jenny-Martin	1	5
Jenny-Chloe	1	35
Jenny-Ramesh	0	21
Martin-Chloe	0	3
Martin-Ramesh	1	24
Chloe-Ramesh	1	3

- The "case" is the dyad or pair of actors
- Values assigned to <u>pairs</u> of actors
- Relations are "attributes" i.e. variables
- Multiple relations are recorded for the same pair of actors

CASOS

Slide 28

Carnegie Mellon
ISr institute for SOFTWARE RESEARCH

Dynamic Network Data

Time 1
Friendship

	Jenny	Martin	Chloe	Ramesh
Jenny	-	1	1	0
Martin	1	-	0	1
Chloe	0	0	-	1
Ramesh	0	1	1	-

Shared Concepts

	Jenny	Martin	Chloe	Ramesh
Jenny	-	5	35	21
Martin	5	-	3	24
Chloe	35	3	-	3
Ramesh	21	24	3	-

Time 2
Friendship

	Jenny	Martin	Chloe	Ramesh
Jenny	-	0	1	1
Martin	1	-	0	0
Chloe	1	0	-	1
Ramesh	1	0	0	-

Shared Concepts

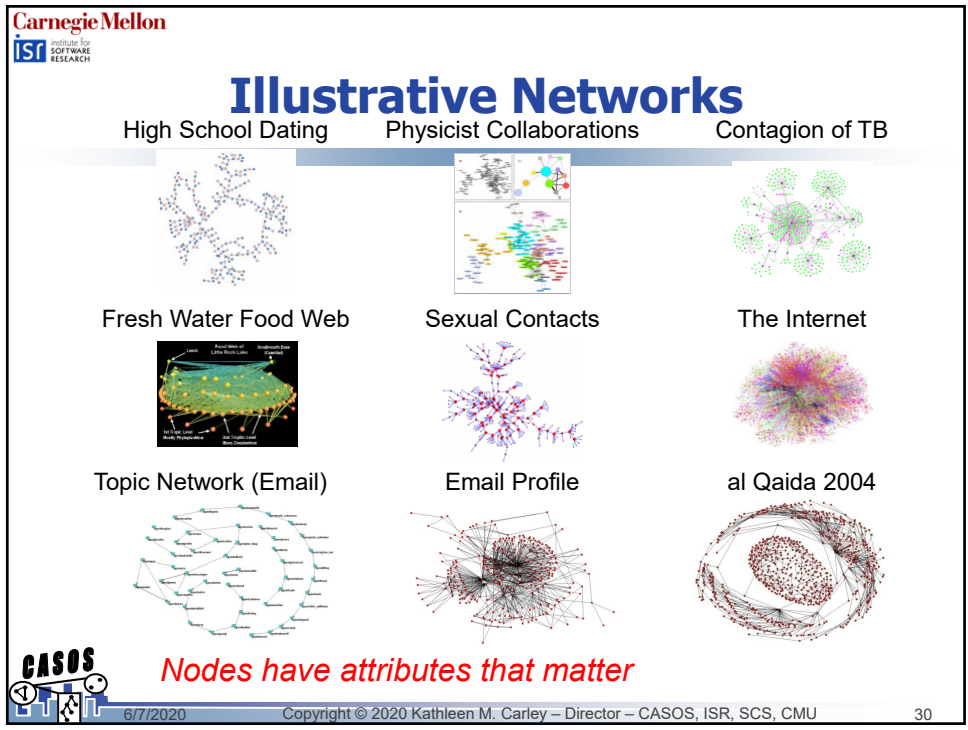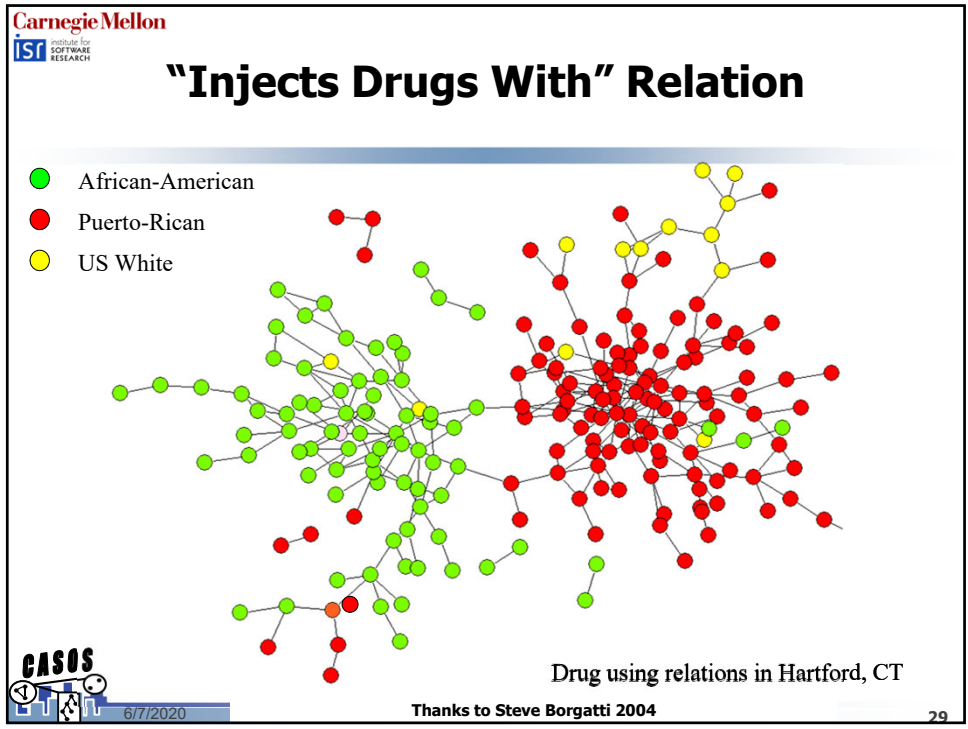|	Jenny	Martin	Chloe	Ramesh
Jenny	-	10	40	30
Martin	10	-	3	24
Chloe	40	3	-	3
Ramesh	30	24	3	-

- 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

CASOS

CASOS

"Injects Drugs With" Relation

- 🟢 African-American
- 🔴 Puerto-Rican
- 🟡 US White

Drug using relations in Hartford, CT

Thanks to Steve Borgatti 2004

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
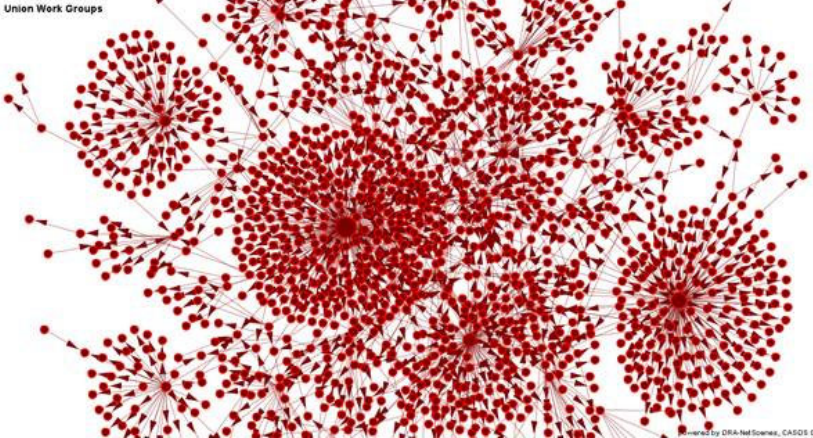
Carnegie Mellon
isr institute for SOFTWARE RESEARCH

Twitter Example - Benghazi

Union Work Groups

Powered by ORA-NetScenes, CASOS Center @ CMU
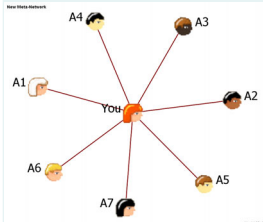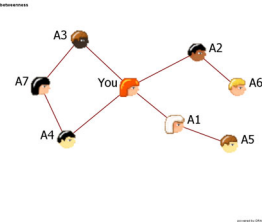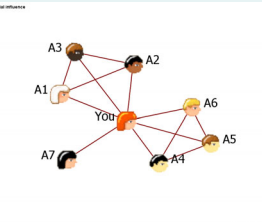
Some Nodes Stand Out

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

Network Effects

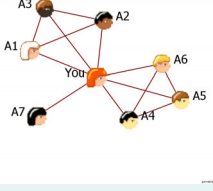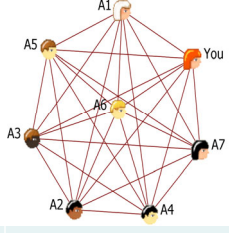| Access	Control of Flows	Influence
The more people you are connected to the more you can know	The more you are on the path between people the more you can control	The more you are connected to others who are connected to each other the more influence they have on you and you on them
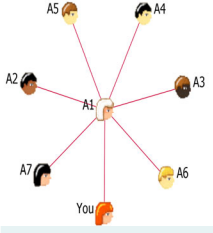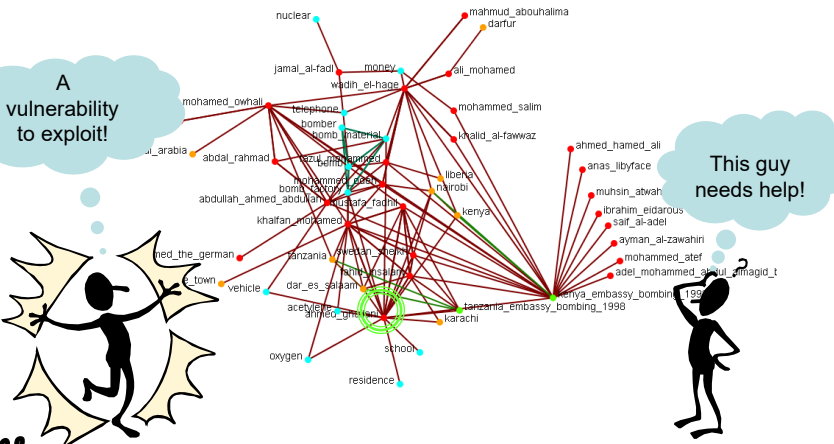
**Influence on You
Network Effects**

Super Spreader	Constraint	Echo Chamber
You are more likely to get information/resources from a super-spreader	The more you are between multiple groups the more constrained your actions	The more you are connected to others who are connected to each other the more influence they have on you

Network Analysis Enables Management and Disruption

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

Non-Lethal Targeting

Low value for Disruption

High value for Information

Numerical values of node importance based on network structure

Max Disruption
Betweenness Centrality

Max Information
Closeness Centrality

Resource Broker
Exclusivity/Specialization

- Target differentiation
- Objective targeting
- Faster
- New insights

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

Organizations are Formal Relations

hierarchy

You

A1   A2

A3   A4   A6   A5

A7

powered by ORA-NetScenes

- 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 …

Organization's have Formal and Informal Structures

Formal

Informal

hierarchy

You
A1
A2
A3
A4
A6
A5
A7

social influence

A3
A2
A1
A6
You
A5
A7
A4

powered by ORA-NetScenes

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

CASOS

Ego Network

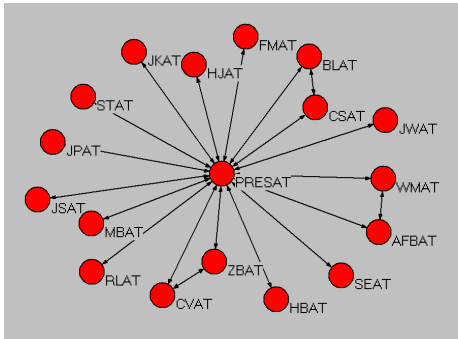- Ego – the node you are examining
- Alters – the nodes directly connected to ego
- The set of ties directly to/from ego to/from an alter
- + (sometimes) the set of ties among ego's alters

6/7/2020   Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU   39
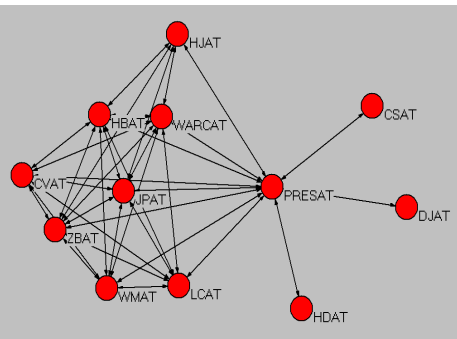
1-mode ego network

Carter Administration meetings

Year 1

Year 4
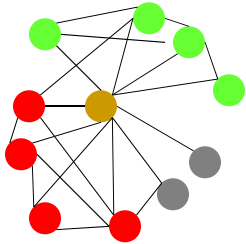
6/7/2020   **Data courtesy of Michael Link**   40
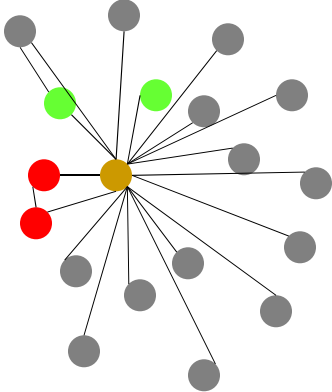
Drug Networks

Normal Person · Cocaine User

Family · Work · Friend
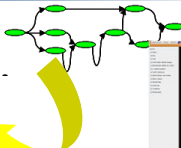
People with Different Roles have Different Networks
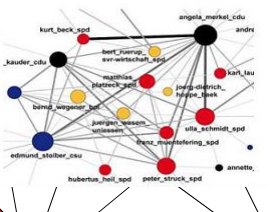
We pull out – the Meta-Network

WHAT: Tasks Events
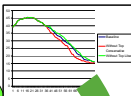
HOW: Resources, Knowledge

WHERE: Location

WHO: People, Teams Organizations

WHY: Beliefs

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

	People	Expertise	Activities	Locations
People	Social Network	Knowledge Network	Assignment Network	Presence Network
Expertise		Information Network	Needs Network	Availability Network
Activities			Precedence Network	Happenings Network
Locations				Border Network

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

CASOS

Carnegie Mellon
institute for SOFTWARE RESEARCH
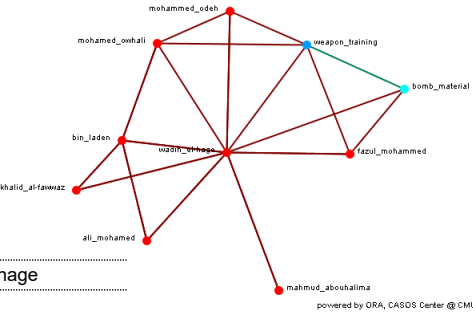
How to Influence

Node Type	Size	Percent
agent	8	+50%
knowledge	2	+50%
location	2	+3%
resource	1	+25%
task	1	+20%

Name	Value
alias	wadih_hage
hostility_level	1
joined_al_qaeda	1989
left_al_qaeda	1998
nationality	lebanese
nationality_relation	hostile
nyi	prosecutor
source_date	2006-05-10
suspected_terrorist	yes

mohammed_odeh
mohamed_owhali
weapon_training
bomb_material
bin_laden
wadih_elhage
fazul_mohammed
khalid_al-fawwaz
ali_mohamed
mahmud_abouhalima
powered by ORA, CASOS Center @ CMU

Most similar other –
jamal al-fadil

CASOS

CASOS

Slide 49

Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

Network Analysis 101

Slide 50

Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

Network Analysis

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

Adjacency Matrix　　　　Diagram

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

Graph　　　　*Network*

CASOS

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

Terminology

- 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

```
0 1 0 1 0          0 1 1 1 0
1 0 0 1 0          1 0 0 1 1          Symmetry achieve by
1 0 0 0 1          1 0 0 1 1          Union of Rij and Rji
0 0 1 0 1          1 1 1 0 1
0 1 1 0 0          0 1 1 1 0
```

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 are pattern that money launderers follow?
  - Use: Locating trails
- Node level
  - E.g., centralities
  - Who has the power?
  - Use: Identifying key actors, events, resources …

CASOS

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.

CASOS

CASOS

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)
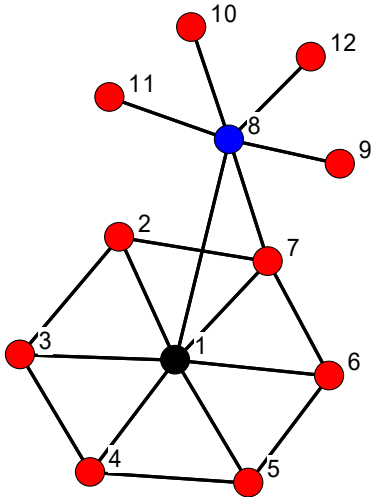
0 2 0 5 0

1 0 0 1 0

1 0 0 0 7

0 0 8 0 1

0 2 4 0 0

Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU   57

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

© Steve Borgatti 2004   58

Slide 59

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 knows 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

```
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
```

Slide 60

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

```
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
```


B's ego Net

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

Network Analysis 201

- Meta-Network
- Applications
- Dynamic Network Analysis

CASOS

CASOS

CASOS

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

Meta Network Approach to Organizational Representation

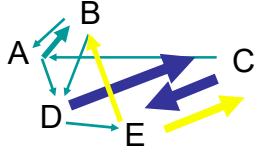|	People	Knowledge	Resources	Tasks	Organizations
People Relation	**Social Network** *Who knows who*	**Knowledge Network** *Who knows what*	**Capabilities Network** *Who has what resource*	**Assignment Network** *Who does what*	**Work Network** *Who works where*
Knowledge Relation		**Information Network** *What informs what*	**Skills Network** *What knowledge is needed to use what resource*	**Needs Network** *What knowledge is needed to do that task*	**Competency Network** *What knowledge is where*
Resources Relation			**Substitution Network** *What resources can be substituted for which*	**Requirements Network** *What resources are needed to do that task*	**Capital Network** *What resources are where*
Tasks Relation				**Precedence Network** *Which tasks must be done before which*	**Market Network** *What tasks are done where*
Organizations Relation					**Inter-Organizational Network** *Which organizations link with which*

6/7/2020   Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU   63

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

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

Flight AA #11 - Crashed Into WTC North
Flight AA #77 - Crashed into Pentagon
Flight UA #93 - Crashed in Pennsylvania
Flight UA #175 - Crashed into WTC South
Other Associates of Hijackers

Copyright © 2001, Valdis Krebs

Figure 3   Trusted Prior Contacts + Meeting Ties [shortcuts]

6/7/2020   Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU   64

CASOS

Carnegie Mellon
IST institute for SOFTWARE RESEARCH

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

CASOS

Carnegie Mellon
IST institute for SOFTWARE RESEARCH

Two-Mode Data

- "Indirect" Connection
  – Co-Publishing, memberships in clubs, boards of companies
  – Events
- Shared …
  – Knowledge, Resources
  – Tasks, Locations

CASOS

Structural Analysis

- Overall Process
- Tools
- Applications

The Overall Network Analysis Process

Get the Data   Analyze & Visualize   Forecast

NetMapper   *ORA   CONSTRUCT

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.
4. Generate Report.
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.

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

Text Mining to Extract Networks

Analyst: Coding Settings

- 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. Context based Sentiment Analysis)
- CUES

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

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

Legend:
- agent (red)
- knowledge (blue)
- location (black)
- organization (gray)
- resource (magenta)
- task (green)

From the Play – Julius Caesar by Shakesphere

Legend:
- Agent : size 48
- Event : size 31
- Knowledge : size 6
- Location : size 8
- Task : size 23
- Agent x Agent
- Agent x Event
- Agent x Knowledge
- Agent x Location
- Agent x Task
- Event x Event
- Knowledge x Task
- Location x location
- Task x Event
- Task x Task
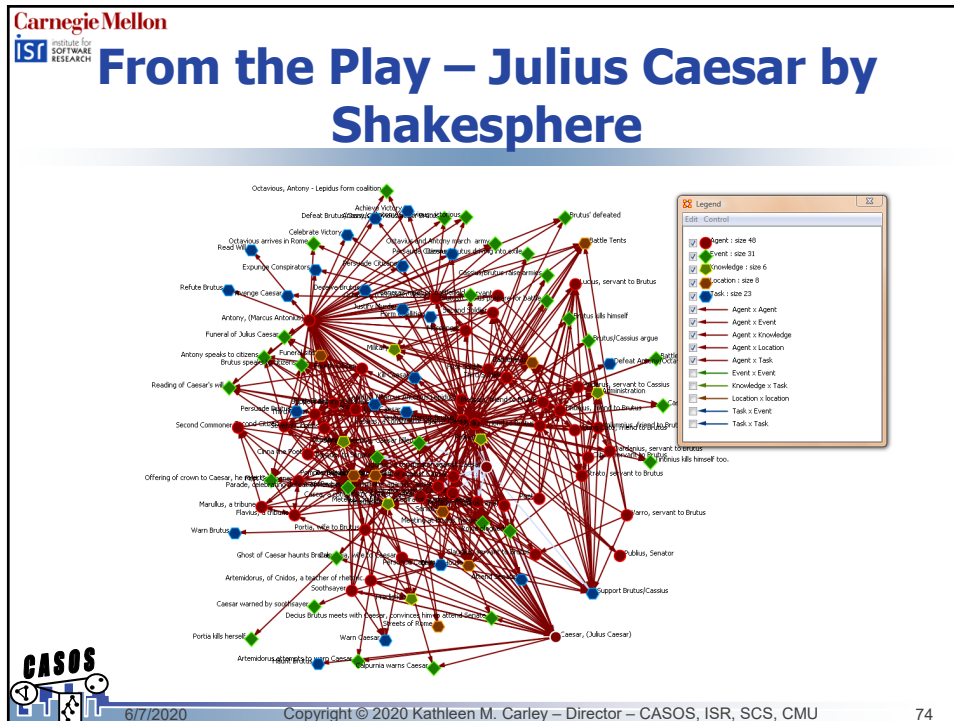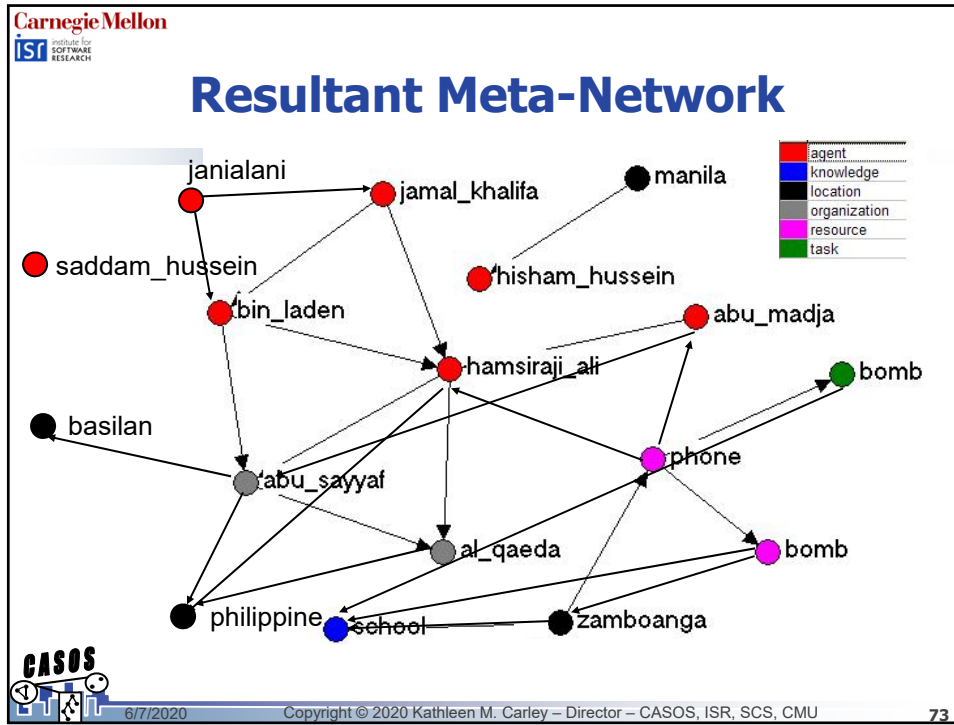
Data to Model – D2M

- 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

CASOS

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

Illustration of Impact of typical Deduplication - Number of texts before and after deduplication applied only to Sudan data		
Sudan	Number before	Number of after
text	32613	18309
concepts	88260	83150
Average frequency per concept	197.417879	88.15785929

Illustration of Impact of Deduplication: Note Deduplication Can Impact Importance of Concepts

Top 10 Concepts Before and After Deduplication - Sudan					
Before			After		
Concept	Count	%	Concept	Count	%
Valencia	1141455	6.55	Valencia	853691	11.65
conflict _task	867688	4.98	conflict	585801	7.99
nanuque	500411	2.87	republic_of_the_sudan	332082	4.53
ampere	448679	2.88	conflict_task	207738	2.83
republic_of_the_sudan	385560	2.21	wilayat_darfur	153976	2.1
wilayat_darfur	344036	1.97	political	113992	1.56
valence_task	178629	1.03	Sudanese	94010	1.28
ner_population	178059	1.02	Khartoum	72409	0.99
faouzi_ben_mohamed_be_ahmed_a	172782	0.99	valence_task	62440	0.85
badou	152547	0.88	environment	60138	0.82

CASOS

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

Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

Illustrative Results for Impact of Specialized Stemmers

Number of concepts after depluralization			
	Original	After	Percentage
Sudan	97492	85758	87.96%
Catnet	24743	22091	89.28%
Singapore	5073	4452	87.76%

Number of texts nouns and verbs before and after depluralization and detensifying				
	Nouns Before	Nouns After	Verbs Before	Verbs After
Sudan	28488	23680	12006	6763
Catnet	7693	6838	4754	3223
Singapore	1677	1445	1213	816

Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

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

Number of Named Entities					
	Texts	People	Organizations	Locations	Events
SNARC	265	33,210	5,523	2,807	120

STEP 5 – Includes Ontological Cross Classification and Thesauri Construction

- Apply standard thesauri and ontological categories
- Applying the standard thesauri coverts 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

Slide 85

Carnegie Mellon
institute for SOFTWARE RESEARCH

Illustrative Results from Named Entity List Before Resolution

conceptFrom	conceptTo	metaOntology
up missions targeting taliban leaders	up_missions_targeting_taliban_leaders	agent
distribute new	distribute_new	agent
main office california office 1899 l street	main_office_california_office_1899_l_street	location
pay afghan	pay_afghan	agent
military commander	military	organization
once u	once_u	agent
peace press washington	peace_press_washington	agent
david katz	david_katz	agent
david kilcullen	david_kilcullen	agent
david lachapelle	david_lachapelle	agent
david lanz	david_lanz	agent

Slide 86

Carnegie Mellon
institute for SOFTWARE RESEARCH

ORA

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

Carnegie Mellon
institute for SOFTWARE RESEARCH

- Network analytics
- Network visualization 2D & 3D
- Big Data > 2 mil nodes
- High dimensional networks
- Semantic networks
- Geo-spatial network analytics
- Dynamic network analytics & visuals
- Covert group detection & hidden networks
- Social media analytics
- Key actors
- Community detection
- Compare & contrast networks
- Team assessment
- Advanced graphics
- Designed to interoperate with other tools
- Extensive help & support

NodeL

Trail Analysis

Heat maps

Change detection & Fourier analysis

Network Visualization

Actor Analysis & Trails

Networks maps

ORA does Dynamic Network Analysis

Carnegie Mellon
institute for SOFTWARE RESEARCH

- 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

Regular
- Louisville
- Nashville
- Houston
- Albuquerque

Betweenness

Geo-Network Measures Improve focus

- Louisville
- Albuquerque
- Tucson
- Atlanta

Metrics and Grouping Capabilities

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

- 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
  - Leiden
  - Spectral
  - MDS
  - LDA
  - LSA

Other Features

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

- 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

Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

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, anonimization and manipulation capabilities
- Companion tools for processing texts and twitter data

Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

Identify: Who are the Key Players? Or Locations, Resources …

Recurring Top Ranked Agent - 2_5v2all

Agents 286, 652, 97 repeatedly rank in top three in the measures.

In-the-Know (total degree centrality)

Rank	Value	Unscaled	Agent
1	0.181	181	652
2	0.176	176	286
3	0.175	175	97
4	0.165	165	412
5	0.158	158	502
6	0.154	154	273
7	0.139	139	246
8	0.115	115	615
9	0.111	111	829
10	0.109	109	8

Potentially Influential (betweenness centrality)

Rank	Value	Unscaled	Agent
1	0.0988664	24667.2	286
2	0.0705427	17600.4	97
3	0.0625256	15600.1	502
4	0.0609523	15207.6	829
5	0.0548627	13688.2	652
6	0.0542421	13533.4	615
7	0.0524498	13086.2	412
8	0.0410645	10245.6	501
9	0.0306578	7649.12	273
10	0.0305522	7622.78	552

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 move slower.

Are Two Critical Actors Linked? Path Finder

tariq_al-fadli

mohammed_atta

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

Real-world Win! – The Context

- TrainingTulsa 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 <u>live</u> arrest and surveillance database

CASOS

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

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

CASOS

CASOS

Two Gangs

Notice – there are two gangs with very different members

6/7/2020   Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU   103

Internal Structure

ILA   GOOD BOYS

Notice – the two gangs have different structures
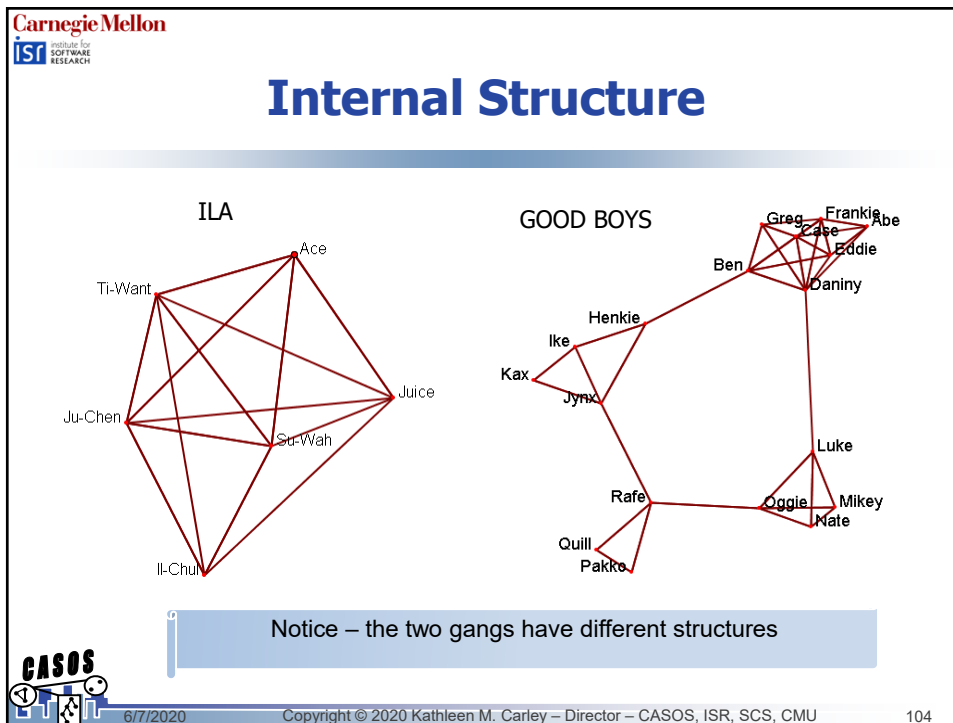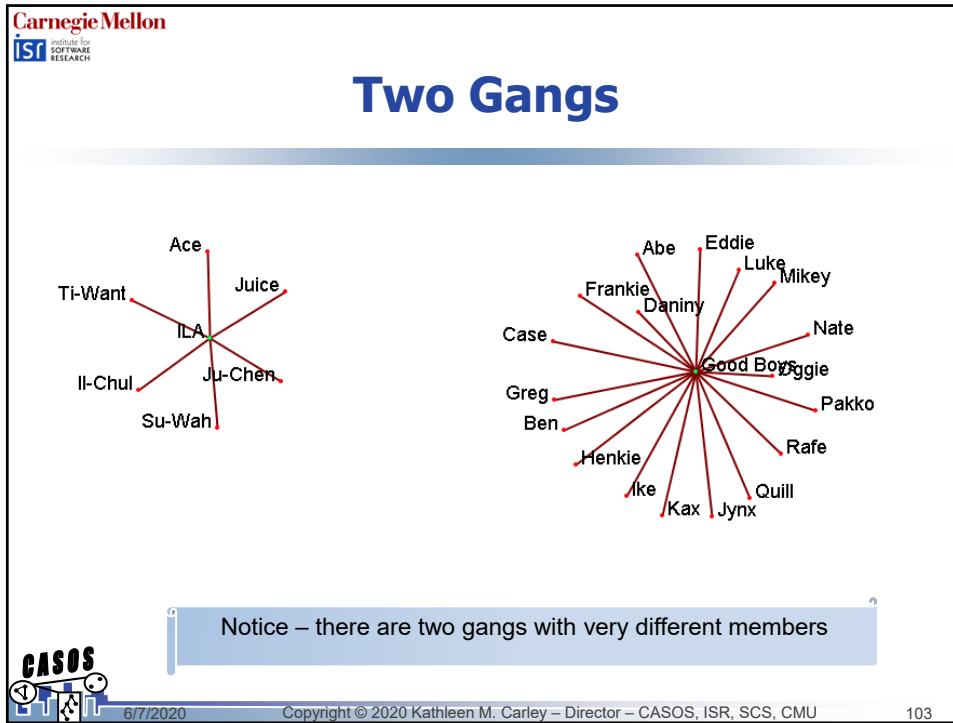
6/7/2020   Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU   104

Are the gangs connected? Yes!

Frankie

Good Boys

ILA — Juice

Greg

The Connector!

Change in Networks

Natural Evolution

Recruitment

Learning Introduction

Isolation

Intervention

Change Detection – ORA
Comparison – ORA
Change in "keyness" – Metric Dynamics - ORA
Immediate Impact – ORA
Near term Impact – ORA-Construct

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

Change Detection
- Objective

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

Graduate Students
(Email over 24 weeks, 07)

Al Qaeda Data
(Based on Sageman, 88-04)

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

Al-Qaeda Application

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

	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
Closeness	0.0027	0.003	0.0028	0.0028	0.0031	0.003	0.0032	0.0034	0.0024	0.0015	0.0004
Z	-0.8729	1.0911	-0.2182	-0.2182	1.7457	1.0911	2.4004	3.7097	-2.8368	-8.7287	-15.9299
C+	0	0.5911	0	0	1.2457	1.8368	3.7372	6.9469	3.6101	0	0
C-	0.3729	0	0	0	0	0	0	0	2.3368	10.5655	25.9955
Signal	0	0	0	0	0	0	0	1			

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
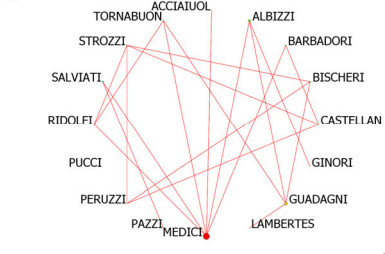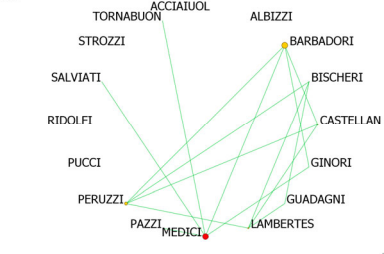
1997 was a critical planning year for Al-Qaeda

CASOS

CASOS

Slide 1:

Carnegie Mellon
institute for SOFTWARE RESEARCH

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

Slide 2:

Carnegie Mellon
institute for SOFTWARE RESEARCH

Simple Comparison – Node Level

File   Edit   View   Favorites   Tools   Help

Google   Search   Share   Bookmarks   Translate   More   Sign In

Total Degree Centrality

Individuals or organizations who are 'in the know' are those who are linked to many others and so, by virtue of their position have access to the ideas, thoughts, beliefs of many others. Individuals who are 'in the know' are identified by degree centrality in the relevant social network. Those who are ranked high on this metrics have more connections to others in the same network. The scientific name of this measure is total degree centrality and it is calculated on the agent by agent matrices.

If the node of interest has a higher than normal value (greater than 1 standard deviation(s) above the mean) the row is colored red. The row is green if the node is within 1 standard deviation of the mean. Finally, the row is colored blue if the node has a lower than normal value (less than one standard deviation(s) below the mean).

Show 10 entries   Search:

	PADGM			PADGB		
Rank	Agent	Value	%Diff to Case 2	Agent	Value	%Diff from Case 1
1	MEDICI	0.400	+16.67%	MEDICI	0.333	-20.00%
2	GUADAGNI	0.267	+50%	BARBADORI	0.267	+50%
3	STROZZI	0.267	+100%	LAMBERTES	0.267	+75%
4	ALBIZZI	0.200	+100%	PERUZZI	0.267	+25.00%
5	BISCHERI	0.200	+0%	BISCHERI	0.200	+0%
6	CASTELLAN	0.200	+0%	CASTELLAN	0.200	+0%
7	PERUZZI	0.200	-33.33%	GINORI	0.133	+50%
8	RIDOLFI	0.200	+100%	GUADAGNI	0.133	-100%
9	TORNABUON	0.200	+66.67%	PAZZI	0.067	+0%
10	BARBADORI	0.133	-100%	SALVIATI	0.067	-100%

Showing 1 to 10 of 16 entries   Previous   1   2   Next

Min:	0	Min:	0
Max:	0.400	Max:	0.333
Mean:	0.167	Mean:	0.125
Std.dev:	0.097	Std.dev	0.113

Back to measure list

CASOS

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

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
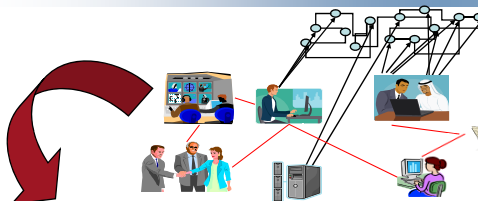
Logics of Adaptation

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

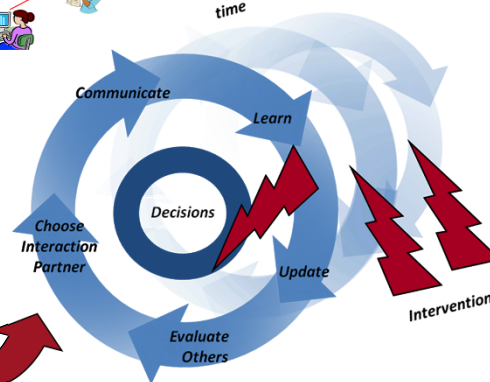6/7/2020   Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU   111

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

Dynamics

	People	Knowledge	Tasks
People Relation	**Social Network** *Who knows who*	**Knowledge Network** *Who knows what*	**Assignment Network** *Who does what*
Knowledge Relation		**Information Network** *What informs what*	**Needs Network** *What knowledge is needed to do that task*
Tasks Relation			**Precedence Network** *Which tasks must be done before which*

time

Communicate

Learn

Decisions

Choose Interaction Partner

Update

Interventions

Evaluate Others

6/7/2020   Copyright © 2020 Kathleen M. Carley – Director – CASOS, ISR, SCS, CMU   112

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

First Impression

Nothing in Common

Rates of Adaptation

	People	Expertise	Activities	Locations
People	*Learning Based*			
Expertise	*Social-Technological Based*			
Activities				
Locations				

Fast

Slow

% Shared

Information

Activity

Beliefs

Time

CASOS

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

Approach to Delivering the Message Matters

Slide 117
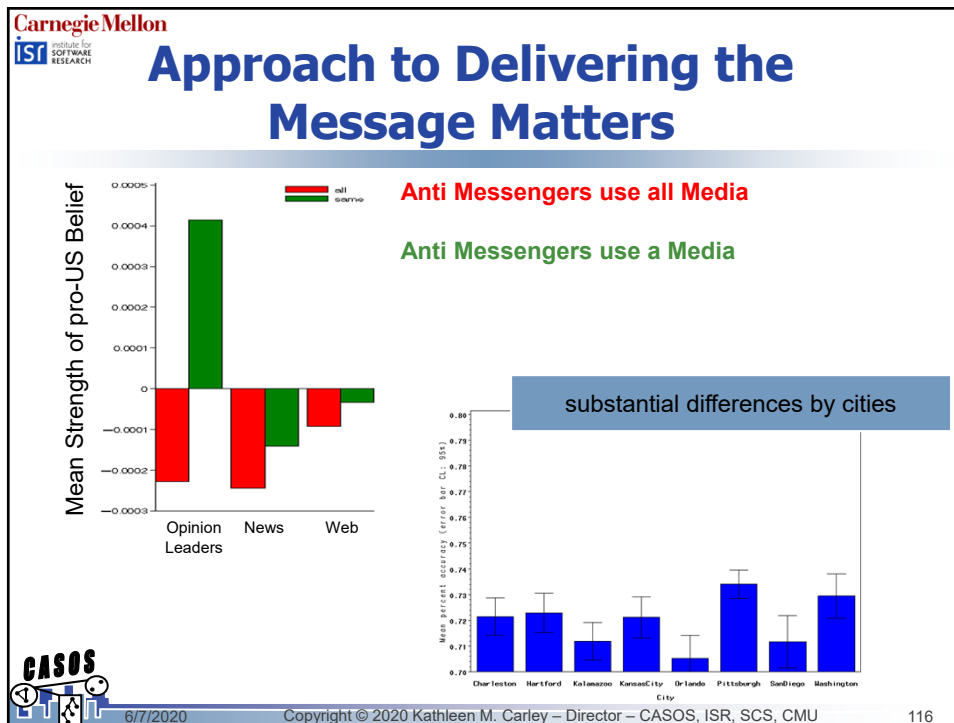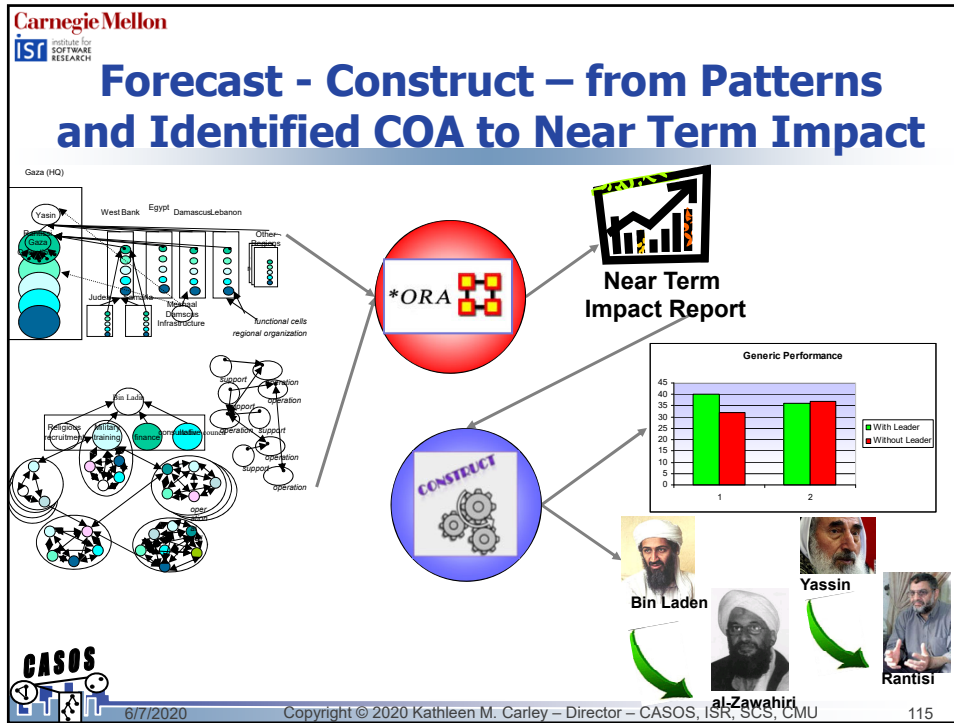
However, edge inference based on meta-network theories is better!

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

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

Type 1

100

● Balance

Random

Exchange

Homophilly

0

0   100
                     Type 2

But – no theory is great – Type 2 twice as likely as type 1

Slide 118

Interaction Style: Need for Communicative Ease - Homophily

- Relative similarity = how much i shares with j divided by how much i shares with all others
- AK_{ik} is knowledge network
  - Knowledge network is agent by knowledge ("facts")
- Expected interaction based on relative similarity

I = max number of agents
K = max number of ideas, facts, pieces of knowledge

$$RS_{ij} \ = \ \frac{\sum_{k=0}^{K}(AK_{ik} * AK_{jk})}{\sum_{j=0}^{I}\sum_{k=0}^{K}(AK_{ik} * AK_{jk})}$$

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

If $RS_{ij} \geq Cutoff$ the Expected interaction = 1
else 0
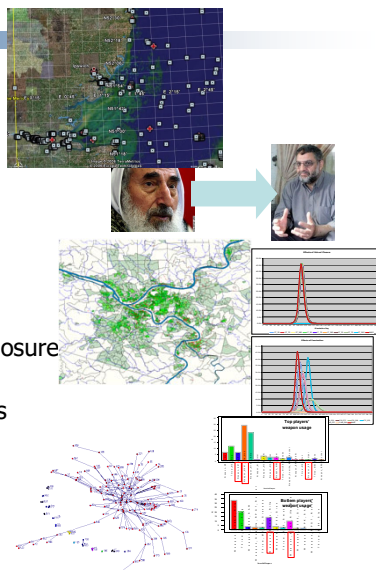
CASOS

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

Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

Tools and Workflows Exist and Are Improving for extracting, analyzing, forecasting, …

Extract Networks from Texts

Analyze and Visualize Networks

Simulate Change in Networks

Data Provider

AutoMap NetMapper

*ORA

CONSTRUCT

IMPORTERS

DE-IDENTIFIER

Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

Let's Go!