Netanomics

CASOS IDeaS Carnegie Mellon University

High Dimensional Networks and ORA

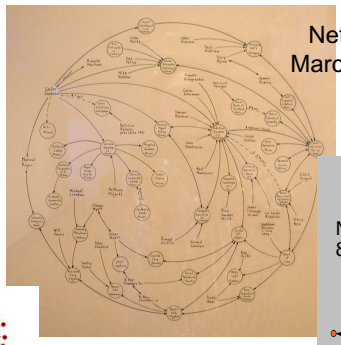
Dr. Kathleen M. Carley
kathleen.carley@cs.cmu.edu

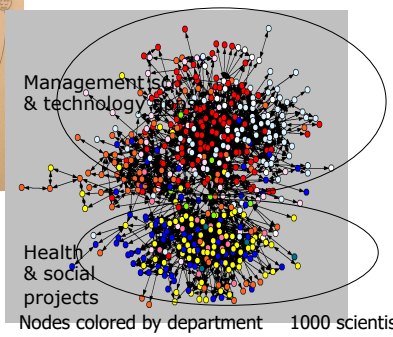
1

Networks!

Fotosearch 2009

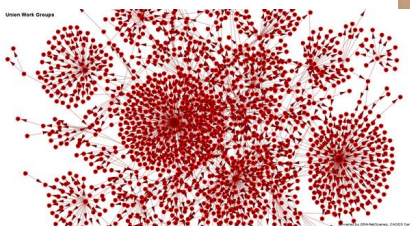
Network Art
Marc Lombardi





Nodes colored by department 1000 scientists

Thanks to Steve Borgatti 2004



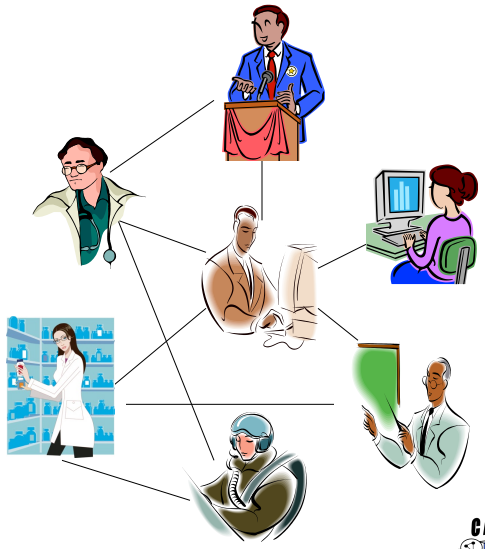
Benghazi Twitter Network 2012 – Kathleen M. Carley et al.

2

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.



Netanomics

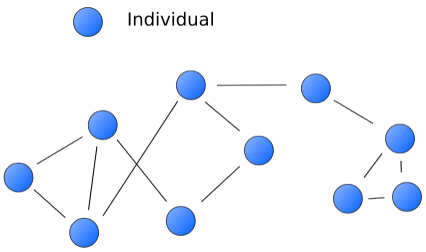
3

Carnegie Mellon University

3

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.



Netanomics

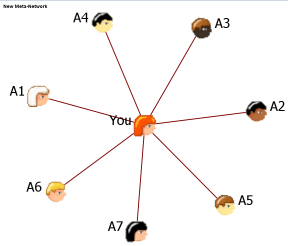
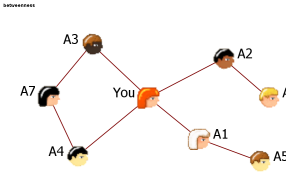
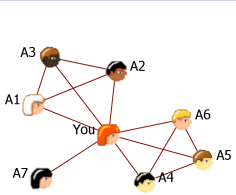
1/28/21

4

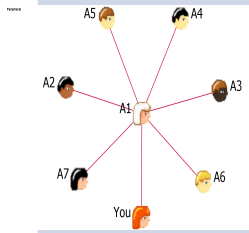
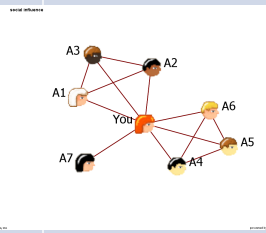
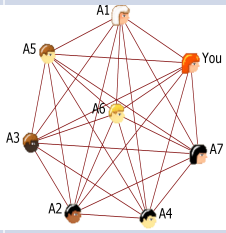
Carnegie Mellon University

4

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

Two Ways to Think About Networks

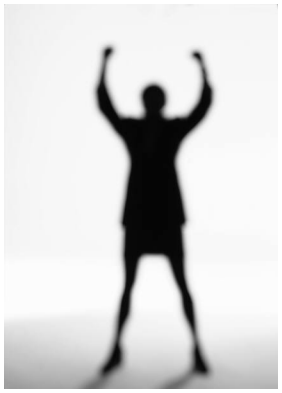
What it means for the individual

- Your position in the network impacts:
  - Promotions
  - Access to information and resources
  - Health
  - Beliefs and voting behavior
  - Etc.

What it means for the organization

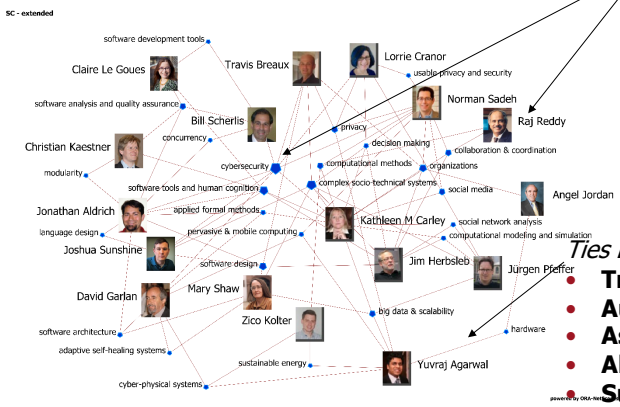
- The network topology of the group impacts:
  - Efficiency
  - Resiliency
  - Speed with which information diffuses
  - Performance caps
  - Points of vulnerability

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

Networks are Ubiquitous



Nodes

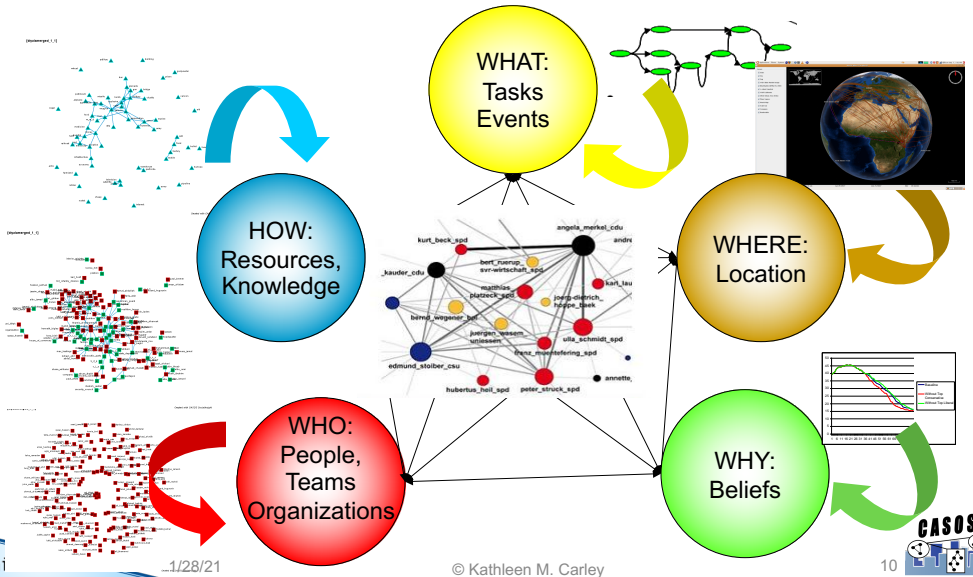
- **People**
- **Topics**
- **Units of action**
- **Countries**
- **Hashtags**
- **Departments**
- **Resources**
- **Ideas or Skills**
- **Events**

Ties Between Nodes (links)

- **Transfer of resources**
- **Authority lines**
- **Association or affiliation**
- **Alliance**
- **Substitution**
- **Proximity**
- **What do you do**
- **Who do you like**

9

Networks form Meta-Network



10

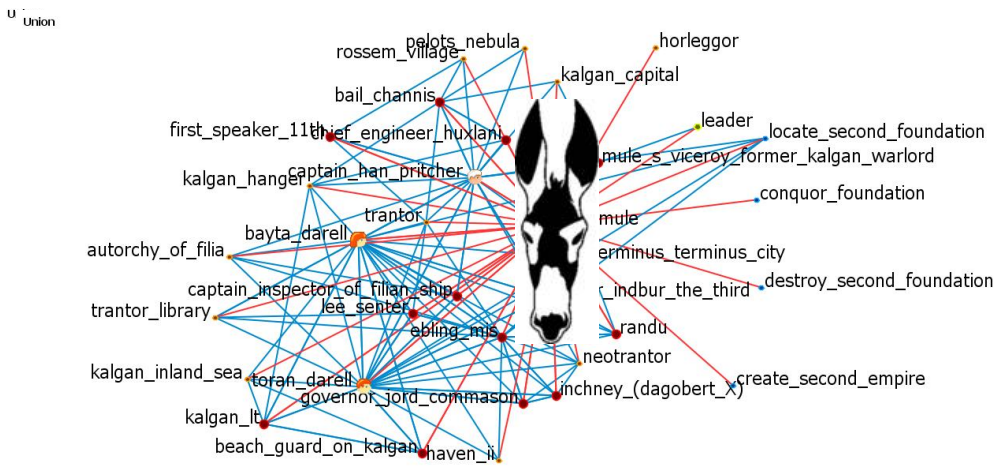
High Dimensional Networks

- Critical for Understanding and Predicting
- Have led to new insights
- Require new methods
  - From ego networks to sphere of influence
  - From centrality to weighted multi-mode centrality
  - From density to complexity
  - New metrics for
    - Overlap/Uniqueness
    - Performance
    - Cognitive demand
    - Workload
  - New Ensemble and MVMC methods for clustering
- Initial versions of all new methods are in ORA-PRO and many in ORA-LITE

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

From Ego Networks to Spheres of Influence



powered by ORA, CASOS Center @ CMU

Netanomics 1/28/21 © Kathleen M. Carley 13 CASOS IDeaS Carnegie Mellon University

13

Key Points About Key Actor Metrics

- There is an analogous graph level metric for all node level metrics
- Shortest path metrics have poor scale properties
- Local versus global influence
  - "atrophication of influence"
- Metrics are influenced by **size** and **density**
- Metrics may or may not take weighted links into account
- Theoretically - Social capital, homophilly and power underlie key actor metrics

Netanomics 1/28/21 © Kathleen M. Carley 14 CASOS IDeaS Carnegie Mellon University

14

Size

Social Network

- Number of nodes (people) in the network
- As size increases
  - *Density decreases*
  - Clustering increases
- Reflects network boundary
- Should always be included as a covariate

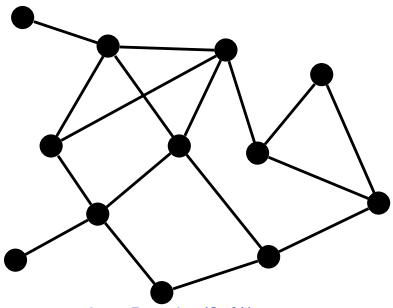
Meta-Network

- Number of nodes of all node types
- As size increases
  - Density decreases
  - Clustering increases
  - Compartmentalization/specialization/siloing increases
  - Should always be included as a covariate
    - Or include size of each square network

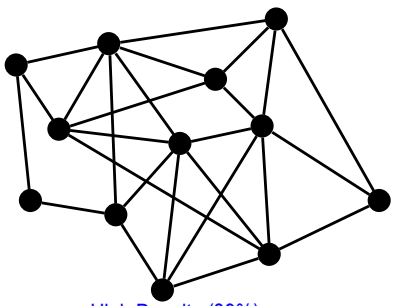
15

Density

- Number of ties for a single relation, expressed as percentage of the number of ordered/unordered pairs
- Number of ties / Number of possible ties
- If number of nodes = N and number of ties is M, then M/(N*(N-1)) if directed and M/((N*(N-1))/2) if undirected



Low Density (25%)
Avg. Dist. = 2.27



High Density (39%)
Avg. Dist. = 1.76

16

Complexity

- Number of ties across all relations, expressed as percentage of the number of ordered/unordered pairs for same mode and multi-mode pairs
- Number of ties / Number of possible ties
- If number of nodes across all modes = N and number of ties is M, then M/(N*(N-1)) if directed and M/((N*(N-1))/2) if undirected
- Undirected may be different for square and rectangular networks

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

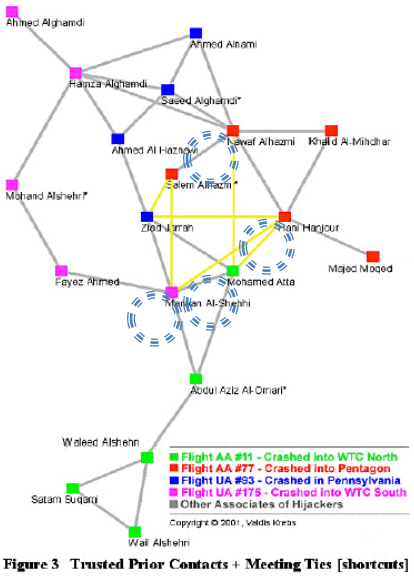
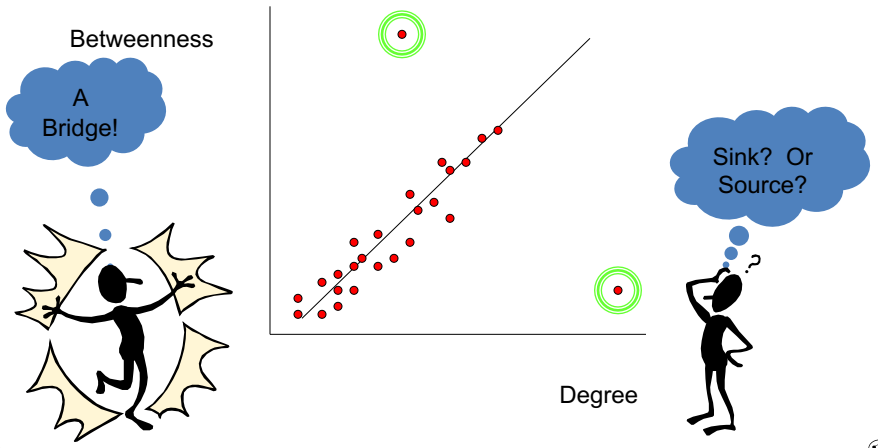


Figure 3 Trusted Prior Contacts + Meeting Ties [shortcuts]

Standard Social Network Measures

Moving Beyond Single Measures

Issue: Centrality Measures are highly correlated



Netanomics

19

Carnegie Mellon University

19

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
- Performance
  - Classification

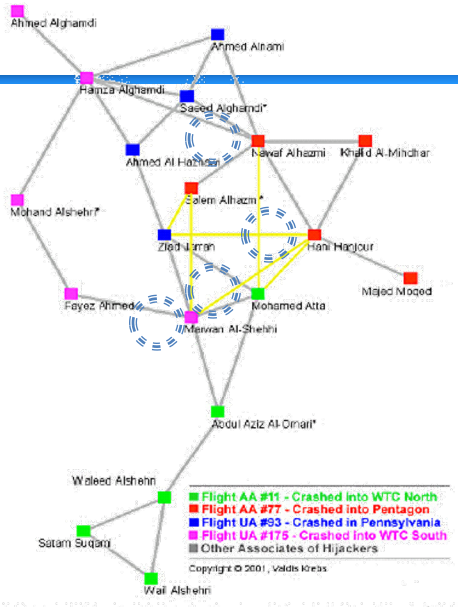


Figure 3 Trusted Prior Contacts + Meeting Ties [abstracts]

Netanomics

1/28/21

20

Carnegie Mellon University

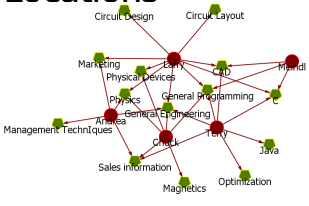
20

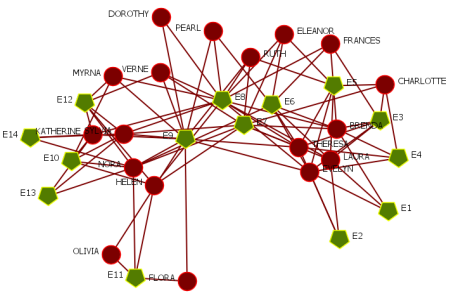
Metrics come in Modes!

- 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

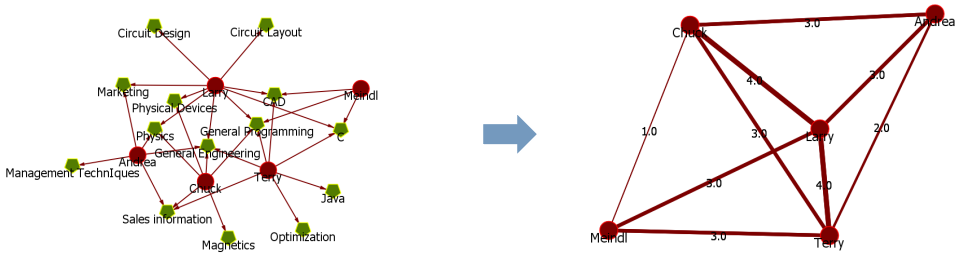
Two-Mode Data

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





Approach: Folding Networks



Person	Avg. Shared Knowledge
Larry	3.50
Terry	3.00
Chuck	2.75
Andrea	2.00
Meindl	1.75

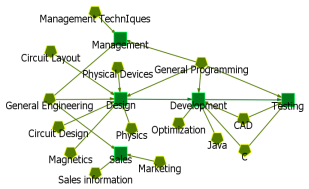
Faust (1997), Borgatti and Everett (1997)

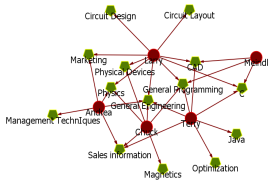
Netanomics

CASOS IDeaS Carnegie Mellon University

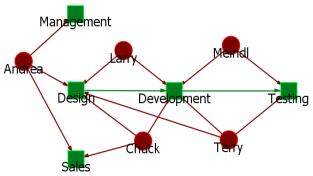
Approach: Meta-Networks

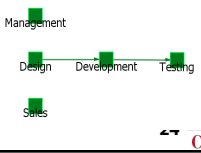
Meta-Matrix entities	People	Knowledge/ Resources	Events/ Tasks	Groups/ Organizations
People	Social network	Knowledge Network/ Resource Network	Attendance Network/ Assignment Network	Membership network
Knowledge/Resources		Information Network/ Substitution Network	Needs network	Organizational capability
Events/Tasks			Temporal Ordering/ Task Flow/ Precedence	Institutional support or attack
Organizations				Interorganizational network





Krackhardt & Carley (1998)
Carley (2002)





Netanomics

CASOS IDeaS Carnegie Mellon University

Many Specific Two-Mode Metrics

- Current ORA measures:

columnDegreeCentrality, inDegreeCentrality, outDegreeCentrality, rowDegreeCentrality, columnCount, rowCount, edgeCount, capability, knowledgeLoad, resourceLoad, density, rowBreadth, columnBreadth., columnDegreeCentralization, inDegreeCentralization, outDegreeCentralization, rowDegreeCentralization, knowledgeDiversity, resourceDiversity, relativeCognitiveSimilarity, cognitiveSimilarity, relativeSimilarity, correlationSimilarity, relativeCognitiveDistinctiveness, cognitiveDistinctiveness, correlationDistinctiveness, relativeCognitiveResemblance, cognitiveResemblance, correlationResemblance, relativeCognitiveExpertise, cognitiveExpertise, relativeExpertise, correlationExpertise, knowledgeExclusivity, resourceExclusivity, taskExclusivity, exclusivityComplete, exclusivity, columnRedundancy, rowRedundancy, knowledgeRedundancy, accessRedundancy, resourceRedundancy, assignmentRedundancy, knowledgeAccessIndex, resourceAccessIndex,

- Classification: Four concept groups of measures
- Node level + dyad level + network level metrics
- [Knowledge] for any kind of affiliation (events, …)

Netanomics

25 CASOS IDeaS Carnegie Mellon University

25

4. Specialization

Identify agents that have either exclusive or redundant connections

- Exclusivity
  - Node: Exclusively connected to [knowledge]
  - *Ashworth and Carley (2006)*

$$\mathbf{x}_i = \sum_{j=1}^{|K|} [AK(i,j) \cdot \exp(1 - \sum AK(:,j))]$$

- Redundancy
  - Network: different agents sharing the same knowledge.
  - Carley (2002)
- Access
  - Node: Critical connections to [knowledge]
  - Ashworth and Carley (2006)

Netanomics

26 CASOS IDeaS Carnegie Mellon University

26

Illustration of Specialization: Redundancy

- Modes: Agents x Tasks
- Average number of redundant agents assigned to tasks. An agent is redundant if there is already an agent assigned to the task.
- Redundancy occurs only when more than one agent is assigned to a task. Define the assignment redundancy for task j as follows:,

$$d_j = \max\{0, sum(AT(:,j)) - 1\} \; 1 \le j \le |T|$$

- Then Assignment Redundancy = $\left(\sum_{j=1}^{|T|} d_j\right) / |T|$

Illustration of Control: Cognitive Demand

The cognitive effort the individual has to do on average

- How many people do you interact with **CENTRALITY**
- How many tasks do you do
- How much knowledge do you have
- How much knowledge is needed to do the tasks
- How many people do you need to interact with to do the tasks
- How many other tasks and so people depend on you
- How many other tasks and so people do you depend on

K-Shell Decomposition

- Given an undirected graph G=(V,E), k-shell decomposition works in a series of steps iteratively:
- k=1: we start by removing all nodes with degree 1 and the associated edges; assign these nodes to 1-shell
- k=2: we remove all nodes of (remaining) degrees of 2 or less, and the associated edges; assign these nodes to 2-shell
- ….
- k-shell: we remove all nodes of (remaining) degrees of k or less, and the associated edges; these nodes are k-shell nodes
- …
- The process stops when no nodes are left. The last k is kmax

- k-core: the graph formed by the nodes that have not been removed at step k
- k-crust: the graph formed by all the nodes in k'-shells, k'=1, …,k

Netanomics

CASOS IDeaS Carnegie Mellon University

29

K-Shell Decomposition: k-Shell Index & Network Structures

- Besides its degree, now each node is also assigned a k-shell index
  - denote by shell(v) for a node v; let deg(v) denotes v's degree
  - give us another "bivariate" (or "multivariate") distribution <deg(v),shell(v)>
- Some simple facts:
  - shell(v) <= deg(v) for all v; and clearly if deg(v)=1, shell(v)=1
  - a high degree node may have low k-shell index: for any v w/ arbitrary deg(v)>1, its k-shell index can be as low as 2
  - for v, if the largest degree of its neighbor is d, then shell(v)<=d+1
  - If v is part of s-clique (and thus deg(v)>=s), then shell(v) >= s.
- Connected components in 1-crust: singleton nodes and isolated edges
- Connected components in 2-crust: stars and stars connected via a path
- ……

Netanomics

CASOS IDeaS Carnegie Mellon University

30

CASOS

Summary on Metrics

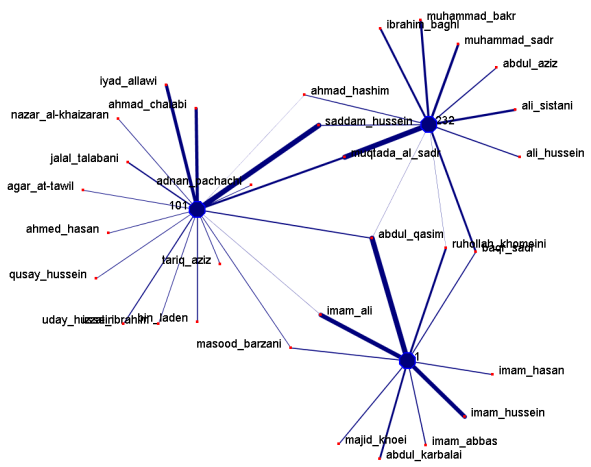
- Level
  - Node level
  - Dyad level
  - Graph level
- Node level
  - Direct
    - E.g. degree
  - Path based
    - E.g. betweenness
  - Iterative
    - E.g. page rank
- Graph level
  - Cohesive
    - E.g. density
  - Spread
    - E.g. characteristic path length
  - Lumpiness
    - E.g. clustering coefficient
  - Min, max, mean, std. dev of node level metrics
- 2 (and n) mode metrics
  - Folding
  - Meta-networks

Netanomics

31 CASOS IDeaS Carnegie Mellon University

31

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

Netanomics

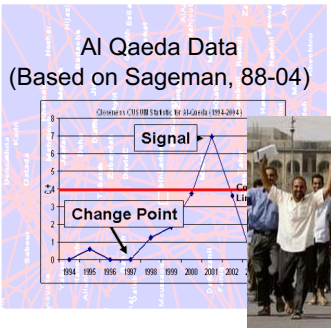
32 CASOS IDeaS Carnegie Mellon University

32

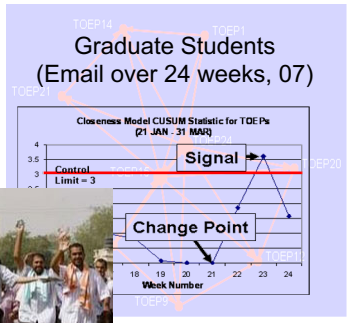
Change Detection - Objective

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)

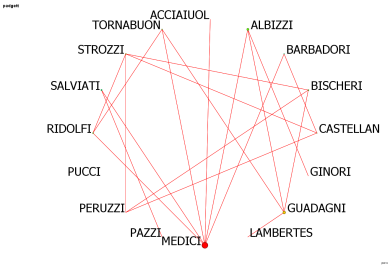


Graduate Students
(Email over 24 weeks, 07)

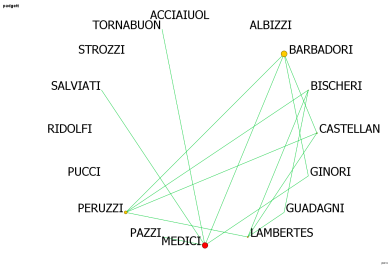


Compare and Contrast

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

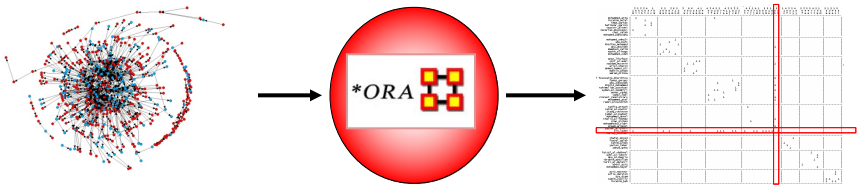


- Banking ties
- Nodes sized by betweenness and colored by degree

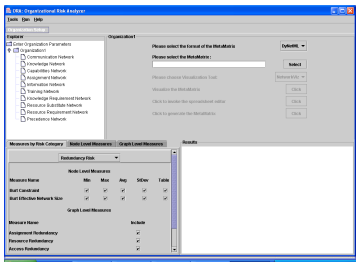


Use Tiles in ORA visualizer to Show Differences

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

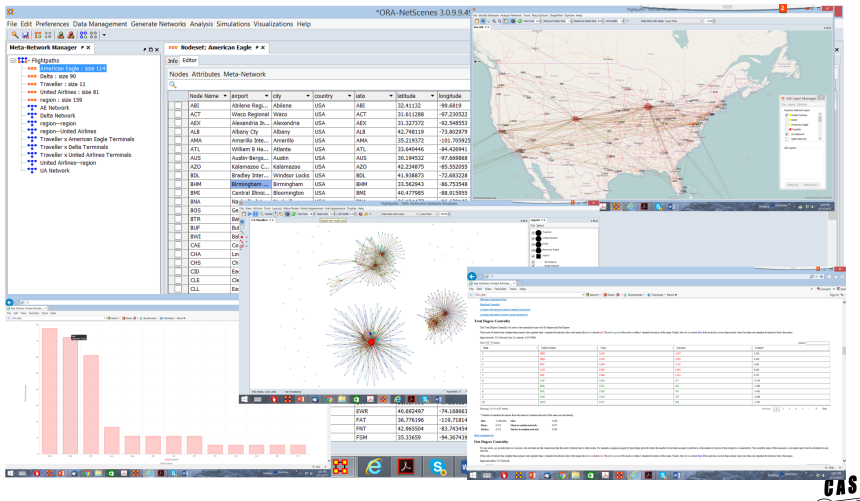
35

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

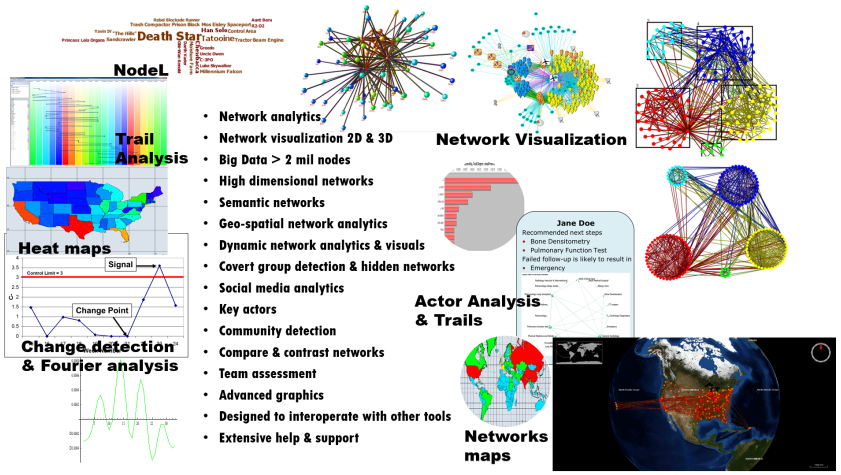
36

ORA in Operation



37

ORA Visualization

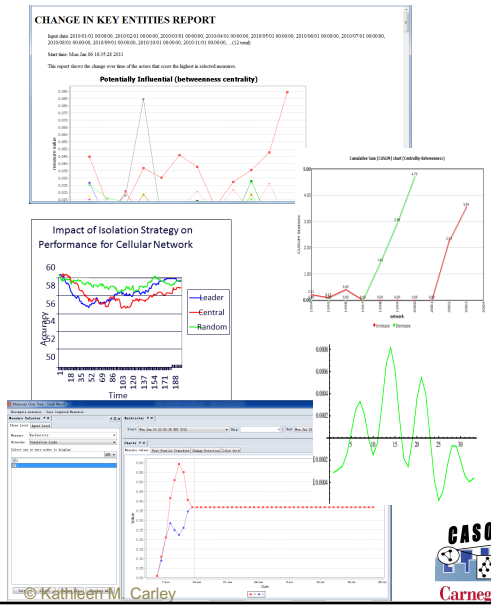


- Network analytics
- Network visualization 2D & 3D
- Big Data > 2 mil nodes
- High dimensional networks
- Semantic networks
- Geo-spatial network analytics
- Dynamic network analytics & visuals
- Covert group detection & hidden networks
- Social media analytics
- Key actors
- Community detection
- Compare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

38

ORA does Dynamic Network Analysis

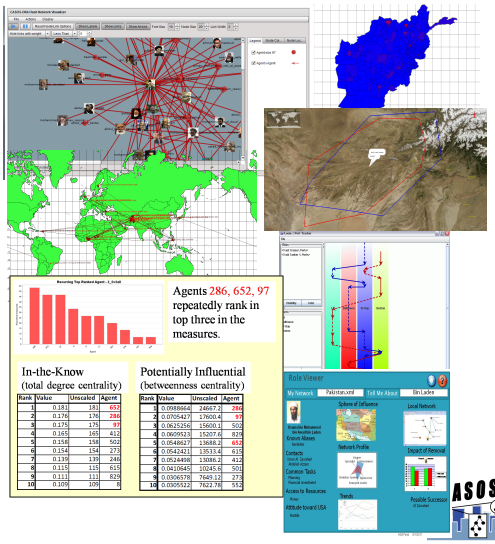
- Trends for metrics
- Trends for nodes
- Through time visualization
- Comparative statistics for two time periods
- Forensic change detection
- Fourier analysis for pattern of life assessment and anomaly detection
- Immediate impact assessment for network change
- Near term forecasting for network change



39

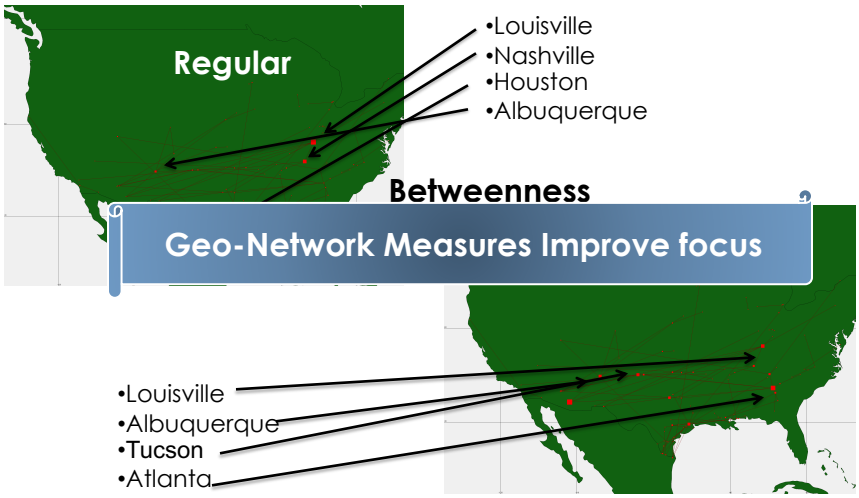
ORA does Geo-Enabled Dynamic Network Analysis

- 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



40

Using Constraints to Set Strength of Links has Practical Consequences



41

Metrics and Grouping Capabilities

- Over 200 metrics
- Key node
  - Centralities
  - Loads
  - Redundancies
- One-mode, two-mode and meta-network metrics

- Illustrative Grouping algorithms
  - Concor
  - Newman-Girvan
  - Fog
  - Johnson-hierarchical
  - Louvaine
  - Leiden
  - Spectral
  - MDS
  - LDA
  - LSA
  - Ensemble
  - MVMC

42

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