Carnegie Mellon

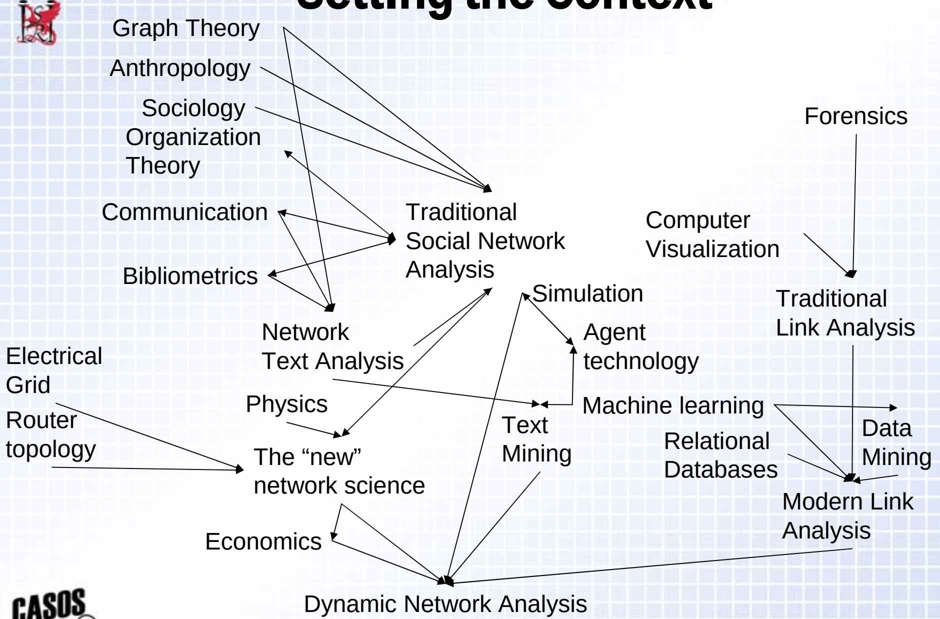
From Social Network Analysis to Dynamic Network Analysis

Kathleen M. Carley
Carnegie Mellon University

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

Carnegie Mellon

Setting the Context



So What is Different?

	SNA	DNA	Old LA	New LA
Node types	1 (2) people	N	N	N
Link types	1 (…) SR	M	M	M
Analytic metric	Yes	Yes	No	Some
Elite Id	Yes	Yes	No	No
Pattern Id	No	No	Yes	Yes
Change	Qual	Yes	Qual	Future = past
Social Intuition	Yes	Yes	Yes	No
Stat. Intuition	No	Yes	No	Yes
Graph Intuition	Yes	Yes	No	No

Programs

- Analysis
  - UCINET is a program for analyzing social network data – available from analytic technologies
  - ORA – is a program for analyzing organizational risk using network data – available from Carley, built in visualization
  - NetStat – stand alone c and r modules for various network modules – available from Carley at legba.casos.ri.cmu.edu
- Network Visualization
  - Social Insight
  - Krackplot
  - Netdraw - available from analytic technologies - Integrated with ucinet
  - Pajek
  - Mage
  - Visualizer

Carnegie Mellon

Characteristics of Traditional Social Network Analysis

- Small networks
  - 15-45 nodes
- Nearly complete information
  - For all pairs know who talks to whom
- Binary data
  - A connection is there or it is not
- One mode or two mode data
  - One type of node with one type of link
  - Two mode data – single matrix with two nodes
- One or two types of relations
  - Often analyzed separately
- Single point in time
- Many measures and algorithms don't scale, don't have
- information content for very large networks

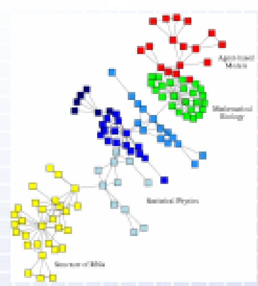
CASOS

Carnegie Mellon

Example of High School Friendships

High school dating: Data drawn from Peter S. Bearman, James Moody,and Katherine Stovel, Chains of affection: The structure of adolescent romantic and sexual networks, preprint, Department of Sociology, Columbia University (2002).

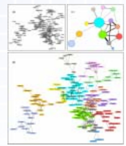
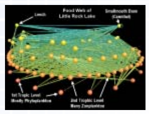
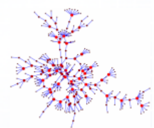
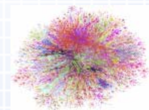
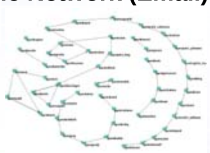
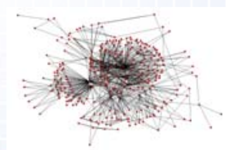
Interdisciplinary collaborations, karate club: M. Girvan and M. E. J. Newman, Community structure in social and biological networks, *Proc. Natl. Acad. Sci. USA* 99, 8271-8276 (2002).



CASOS

Carnegie Mellon

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**
		

CASOS

Carnegie Mellon

Terminology

- Entity class – the type of nodes
- Nodes (entities)
- Links – connections, ties among nodes, relations
- Network – set of links among nodes s.t. nodes may be drawn from one or more entity classes and links may be of one or more classes
- Multi-plex – types of links
  - A multi-plex data set has multiple relations among nodes of the same mode/ entity class, 2 or more types of links
  - Most SNA data sets of single plex
  - Traditional Link analysis uses multi-plex data
- Mode – types of nodes (number of entity classes)
  - Traditional SNA uses 1 mode data (sometimes 2)
  - Traditional Link analysis uses multi-mode
- Multimode
  - 2 or more types of nodes
- Node set – a collection of nodes that group together for some reason

CASOS

Carnegie Mellon

Terminology cont.

- "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

CASOS

Carnegie Mellon

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

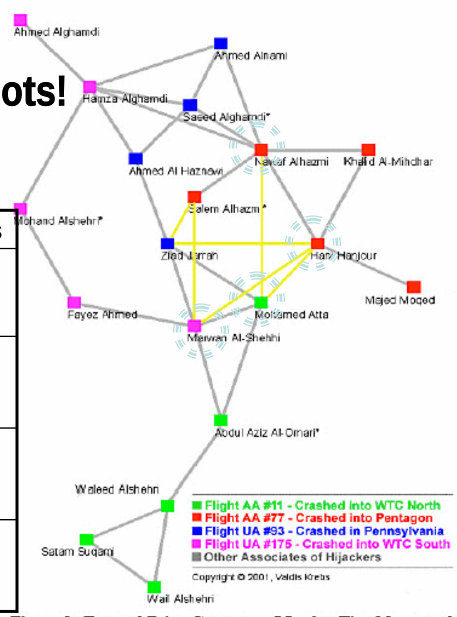


Figure 3 Trusted Prior Contacts + Meeting Ties [shortcuts]

CASOS

Carnegie Mellon

Simple SNA Measures

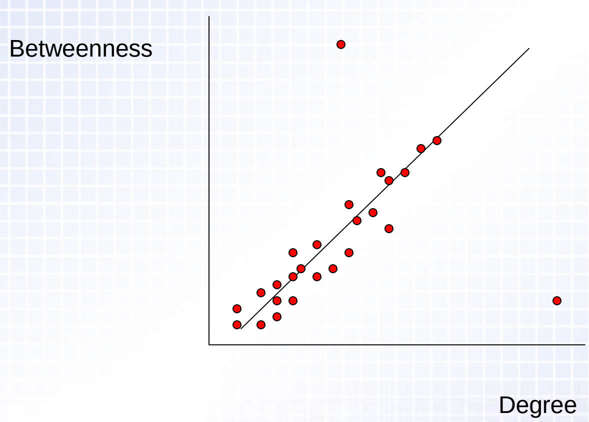
Measure	Definition	Meaning	Usage
Centrality	Node with the most connections	In the know	Identifying sources for intel; Reducing information flow
Betweenness	Node in the most best paths Requires symmetric data	Connects groups	Typically has political influence, but may be too constrained to act
Closeness	Node that is closest to all other nodes	Rapid access to all information	Identifying sources to acquire/transmit information
Betweenness - Closeness	High in betweenness but not closeness	Connects disconnected groups	Reduction in activity

CASOS

Carnegie Mellon

Issue: Centrality Measures are highly correlated

Moving Beyond Single Measures



CASOS

Carnegie Mellon

Getting Results on Key Actors - ORA

- Standard Social Network Measures Report
- Intel Report
  - Selected measures for top nodes
- All Measures Report
  - All measures for all nodes
- User defined
  - Use measures manager to get only the measures you want for all nodes
- Saving data
  - Save as csv and put in excel and process
  - Save as DynetML for future visualization and analysis
  - Save as txt for integrating in word reports

CASOS

Carnegie Mellon

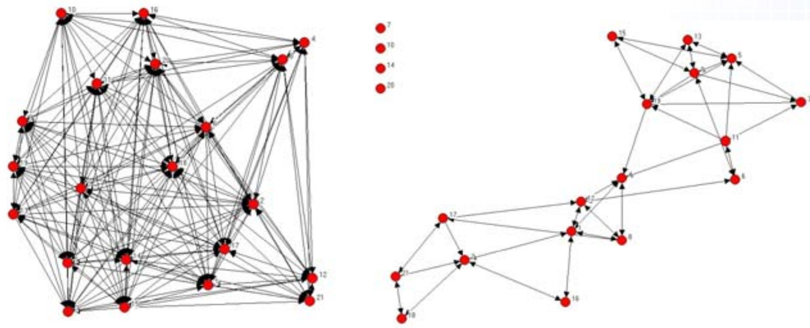
Comparing Networks

- For each measure change or difference can be measured as the percentage difference
  - Percentage difference in old and new structure measure for each previous measure - 100 * (new - old)/old
- Change or difference can be measured for entire network
  - Hamming metric = number of link changes (adds or drops) to convert one structure to another
- Change or difference can be measured as correlation
  - MRQAP comparison of networks

CASOS

Carnegie Mellon

Advice & Friendship



CASOS

Carnegie Mellon

Comparison of Advice and Friendship

		KRACKAD	KRACKFR
1	Mean	0.660	0.143
2	Std Dev	0.474	0.350
3	Sum	277.000	60.000
4	Variance	0.225	0.122
5	SSQ	277.000	60.000
6	MCSSQ	94.312	51.429
7	Euc Norm	16.643	7.746
8	Minimum	0.000	0.000
9	Maximum	1.000	1.000
10	N of Obs	420.000	420.000

CASOS

Carnegie Mellon

Full Matrix Comparison of Advice and Friendship

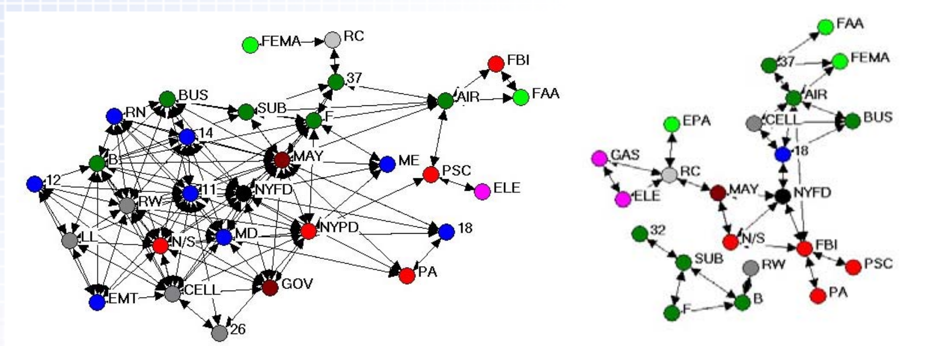
	Value	Sig	P(large)	P(small)
Pearson Corr	0.164	0.002	0.002	0.999
Simple Matching	0.440	0.002	0.002	0.999
Hamming Distance	235	0.002	0.999	0.002

CASOS

Carnegie Mellon

Comparison is at the Heart of Over Time Analysis

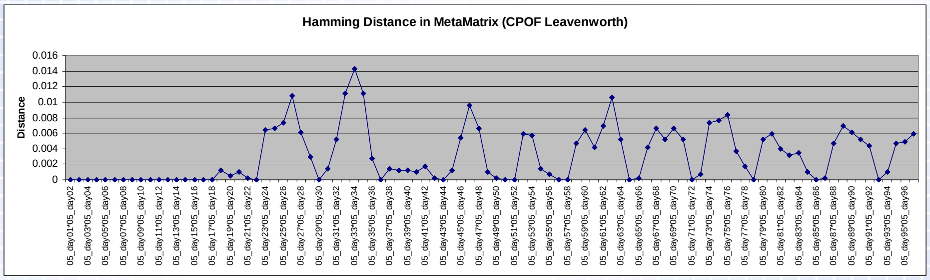
- NYC response network – 9/11/2001
- NYC response network - 9/19/2001



CASOS

Carnegie Mellon

Organizational EKG



CASOS

June 13, 2006 19

Carnegie Mellon

Meta-Matrix Approach to Organizational Representation

	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*

CASOS

June 13, 2006 20

Carnegie Mellon

Meta-Network (Meta-Matrix)

- Complex systems can be represented as a meta-network linking actors, knowledge, resources, tasks …
  - A variety of assessments can be made given this meta-network - such as
    - Vulnerabilities
    - Shared situation awareness
    - Performance and adaptability
    - Performance

	Actors	Knowledge /Resources	Tasks
Actors	**Social Network**	**Knowledge Network**	**Attendance Network**
Know /Reso		**Information Network**	**Needs Network**
Tasks			**Precedence**

Variable

Fixed

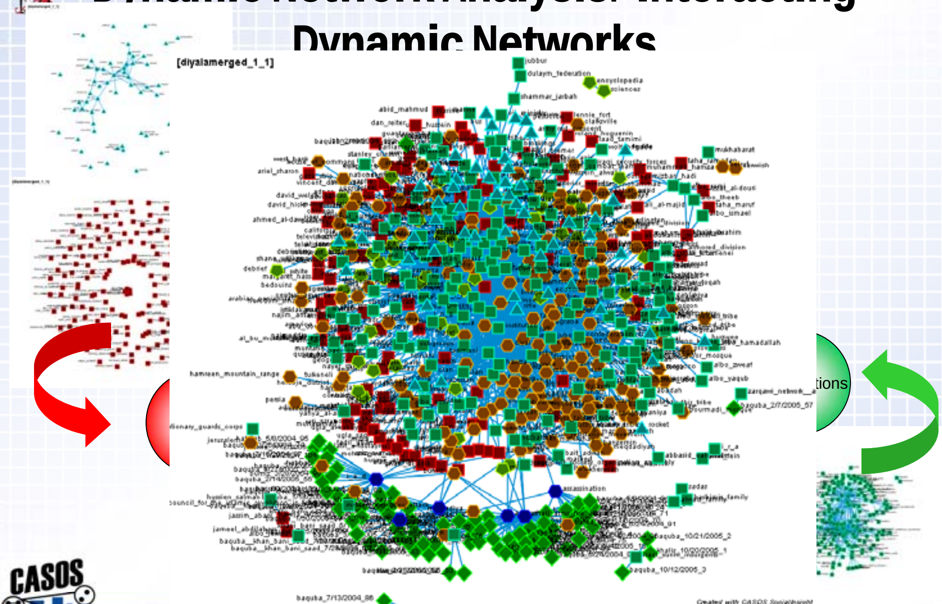
- Portions of this meta-network are more easily changed in the short term while others are relatively fixed being dependent on the current state of science, technology

CASOS

Carnegie Mellon

FOUO

Dynamic Network Analysis: Interacting Dynamic Networks



CASOS

Carnegie Mellon

Meta-Matrix & Measures, cont.

Many measures Few measures

	Individual	Resource	Task
Individual	*Size *Level *Span of control *Network Density *Conductivity *Degree Centralization *Betweenness *Connectivity *Efficiency *Least Upper Boundedness	*Consensus *Resource Specialization *Access Redundancy	*Workload *Assignment *Complexity * Assignment redundancy * Task exclusivity index
Resources		*Size *Network Density *Substitutes	*Needs Complexity
Task			*Size *Network Density *Longest Path

CASOS

Carnegie Mellon

Traditional SNA Measures can be used for the square matrices

- Communication net
  - Agents by agents
- Substitution net
  - Resources by resources
- Information net
  - Knowledge by knowledge
- Precedence net
  - Task by task
- Inter-organizational net
  - Organization by organization

ORA

UCINET

NetStat

CASOS

Carnegie Mellon

Traditional 2-Mode measures can be used for off-diagonal, rectangular matrices

- Clustering
- Complexity (density)

ORA

UCINET

NetStat

CASOS

Carnegie Mellon

New Measures

- Developed for each type of two mode network
  - Knowledge networks
  - Resource access
  - Assignment network
- Redundancy
  - Task level
    - Is the assignment to tasks redundant
  - Resource level
    - Is the access to resources redundant
    - Is there redundancy is resources for tasks
  - Personnel level
    - Is the access to others redundant
- Exclusivities
  - Task exclusivity
  - Knowledge exclusivity
  - Access to individuals

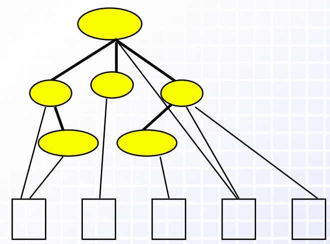
ORA

CASOS

Carnegie Mellon

Resource Access Structure
Resource Access Redundancy

- Who has access to what
- Similar to who knows what
  - Who knows what



Resources

Redundancy = 1.4

- **Resource access Redundancy**
  - **average number of people that have access to the same resources**

ORA

CASOS

June 13, 2006 27

Carnegie Mellon

Assignment Network
Assignment Redundancy

- 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:,

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

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

ORA

CASOS

June 13, 2006 28

Knowledge Exclusivity Index

- Detects agents who have singular knowledge.
- The Knowledge Exclusivity Index (KEI) for agent i is defined as follows:

$$\sum_{j=1}^{|K|} AK(i,j) * e^{(1-sum(AK(:,j)))}$$

- The values are then normalized to be in [0,1] by dividing by the maximum KEI value.

ORA

Task Exclusivity Index

- Detects agents who exclusively perform tasks.
- The Task Exclusivity Index (TEI) for agent i is defined as follows:

$$\sum_{j=1}^{|T|} AT(i,j) * e^{(1-sum(AT(:,j)))}$$

- The values are then normalized to be in [0,1] by dividing by the maximum TEI value.

ORA

Carnegie Mellon

Additional Specialized Measures Exist Particularly Ones Using Multiple Matrices

- Performance
  - Diffusion
  - Accuracy
- Loads
  - Cognitive demand
  - Workload
  - Potential Work Load
- Congruency – fit
  - Communication
  - Knowledge
  - Resource
- Need for Negotiation
- Under Supply

ORA

CASOS

June 13, 2006 31

Carnegie Mellon

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

ORA

OFFICE OF NAVAL RESEARCH

June 13, 2006 32

Carnegie Mellon

Resource Congruence

- Resource Congruence measures to what degree agents have resources when and only when it is needful to complete their assigned tasks.
- This measure captures the similarity between what resources are assigned to tasks via agents, and what resources are required to do tasks
- Input: AR : binary - variable; AT : binary - variable; RT : binary - permanent

$$\Re \in [0,1]$$

- We compare the resources assigned to tasks via agents (RAT) and task resource requirements (RT)
  - let RAT = (RA*AT), then make it binary
  - let d = normalized Hamming distance between RAT and RT
  - Then Resource Congruence = 1 - d

ORA

CASOS

Carnegie Mellon

Communication Congruence

- Measures to what extent communication between agents is limited only to what is needed to complete tasks. Higher congruence occurs when agents don't communicate if the tasks don't require it, and do when tasks require it.
- Input: A : binary – variable; AT : binary – variable; AR – binary – variable; RT : binary – permanent; T : binary – permanent.
- Perfect congruence is agents reciprocally communicate exactly when one or more holds:
  - if i is assigned to a task s and j is assigned to a task t and s directly precedes task t (handoff)
  - if i is assigned to a task s and j is also assigned to s (co-assignment)
  - if i is assigned to a task s and j is not, and there is a resource r to which agents assigned to s have no access but j does (negotiation to get needed resource).

ORA

CASOS

Carnegie Mellon

Communication Congruence (cont.)

- The three cases are computed as follows:
  - let H = AT*T*TA (handoff)
  - let C = AT*TA (co-assignment)
  - let N = AT*Z*RA, where Z(t,r)=(TA*AR - TR)(t,r) < 0 (negotiation)
- Then let Q(i,j) = [(H+C+N) + (H+C+N)’](i,j) > 0
  - transpose enforces reciprocal communication
- let d = Hamming distance between Q and A
  - degree communication differs from what is needed to complete tasks
- Communication Congruence = 1 - d

ORA

CASOS

Carnegie Mellon

Estimating Networks

- Most work on estimating social networks
- To do this – ask – why would characteristics enable or inhibit interaction
- Relies on converting two mode networks to one mode networks
  - Shared attribute approach
  - Relative bases for interaction
    - Knowledge based
    - Resource based
    - Task based

CASOS

Carnegie Mellon

Convert 2 Mode Network to Single Mode

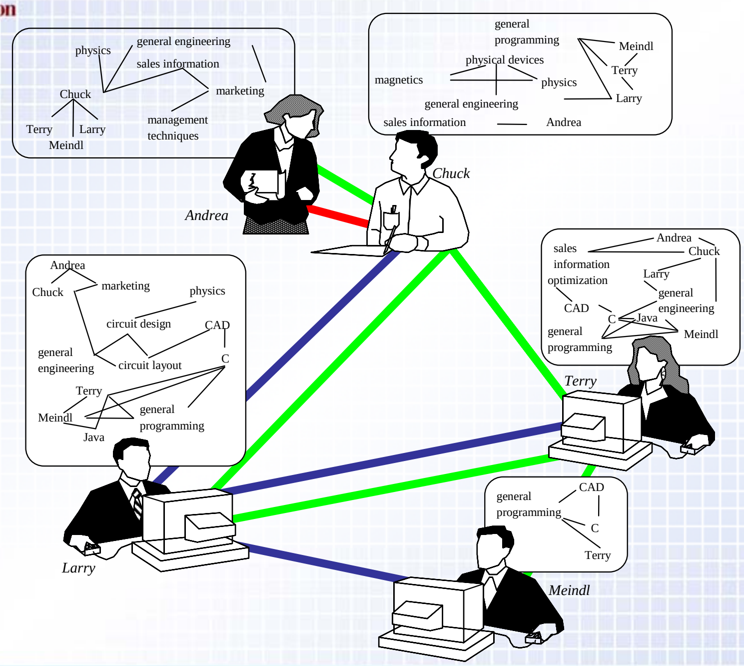
- For any two mode network A
- C= A*A'
- Then if value > mean Cij = 1 else 0
- Alternative, if value > 0 Cij = 1 else 0
- This is similar to shared attributes

Note: Some researchers argue that this is not a "real" social network

CASOS

June 13, 2006 Copyright © 2006 Kathleen M. Carley, CASOS, ISRI, SCS, CMU 37

Carnegie Mellon



CASOS

June 13, 2006 Copyright © 2006 Kathleen M. Carley, CASOS, ISRI, SCS, CMU 38

Carnegie Mellon

Measuring Knowledge Networks - Node Level

Meindel's Information Network

```
General
  programming
C
CAD
Terry
```

Knowledge network

```
ACLTM
11100 Physics
11110 General engineering
11010 Sales information
10100 Marketing
10000 Management techniques
01111 General programming
01100 Physical devices
01000 Magnetics
00010 Optimization
00111 C
00010 Java
00111 CAD
00100 Circuit design
00100 Circuit layout
10110 Chuck
11101 Terry
11010 Larry
11110 Meindl
01110 Andrea
```

Communication Network

```
 ACLTM
A01000
C10110
L01010
T01101
M00010
```

There is a 1 in the knowledge network if person knows that concept, idea, has access to that resource, data, database, or has that skill

CASOS

Carnegie Mellon

Generating Expected Communication Network

- The expected communication network can be generated from the knowledge network
- The "expected" communication is the similarity between what two people know
  - If data is not binary use correlation
  - If data is binary use match
- Then using relative similarity or relative expertise or both convert matrix to binary form
- Example (using RS)

Knowledge network

A	C	T	L	M
1	1	0	1	1
0	0	0	0	0
1	1	0	0	0
0	0	1	1	0

	A	C	T	L	M
A	2	2	0	1	1
C	2	2	0	1	1
T	0	0	1	1	0
L	1	1	1	2	1
M	1	1	0	1	1

Expected communication network

	A	C	T	L	M
A	0	.3	0	.15	.15
C	.3	0	0	.15	.15
T	0	0	0	.5	0
L	.15	.15	.15	0	.15
M	1	1	0	1	0

CASOS

Need for Communicative Ease Relative Similarity

- Shared knowledge / amount of knowledge shared with everyone
- AK_{ik} is knowledge network

I = max number of people
K = max number of ideas

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

ORA

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

Carnegie Mellon

Comparison of Actual and Expected Communication Networks - Node Level

Expected Communication Network

```
 ACLTM
A06651
C60762
L67074
T56703
M12430
```

Actual Communication Network

```
 ACLTM
A01000
C10110
L01010
T01101
M00010
```

```
 ACLTM
A01110
C10110
L01010
T01100
M00110
```

```
Difference
 ACLTM
A==--=
C=====
L=====
T====+
M==-==
```

CASOS

Carnegie Mellon

Need to Know Relative Expertise

- Relative expertise = how much i thinks j knows that i does not know divided by how much i thinks all others know that i does not know
- AKik is knowledge network
- Expected interaction based on relative expertise

If $AK_{ik} = 0$ then $X_{jk} = AK_{jk}$
Else $X_{jk} = 0$

$$RE_i = \frac{\sum_{k=0}^{K} (X_{jk})}{\sum_{j=0}^{I} \sum_{k=0}^{K} (X_{jk})}$$

I = max number of people
K = max number of ideas

$$\text{Cutoff} = \sum_{i=0}^{I} RE_i \;/\; (I * (I - 1))$$

If $RE_i \geq$ Cutoff the Expected interaction = 1 else 0

ORA

CASOS

Relative Expertise - Why

- Expertise: individuals tend to interact with those whom they believe to have information that they need
  - Information gathering
  - Desire to achieve
  - Desire for increase in power
  - Information as power
- Relative: individuals judge expertise relative to others
  - There is a comparison group
  - There is a generalized other