



# From Raw Source Data to Simulation

Annual Tools/Computational Approaches/Methods Conference

March 19, 2008

*USAF Behavioral Influences Analysis (BIA) Center*

Terrill L. Frantz

Carnegie Mellon University

[terrill@cs.cmu.edu](mailto:terrill@cs.cmu.edu)

Center for Computational Analysis of Social and Organizational Systems

[www.casos.cs.cmu.edu/](http://www.casos.cs.cmu.edu/)



# Goal

## *Develop a basic understanding of Dynamic Network Analysis (DNA):*

- Its Worldview
- What it can be used for
- Basic terminology and measures
- Analysis and interpretation
- Available software tools



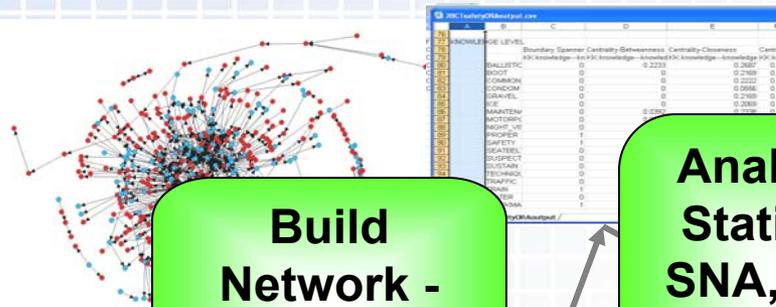
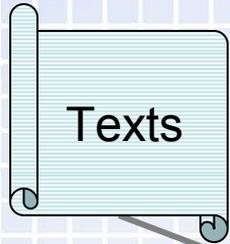
# Gameplan

## *Part I:*

- Introduction to various concepts of DNA & CASOS tools

## *Part II:*

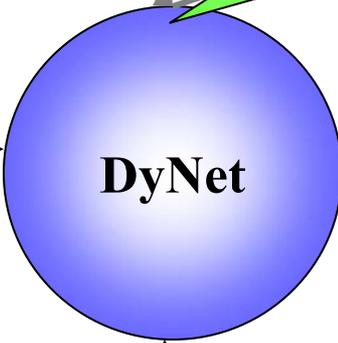
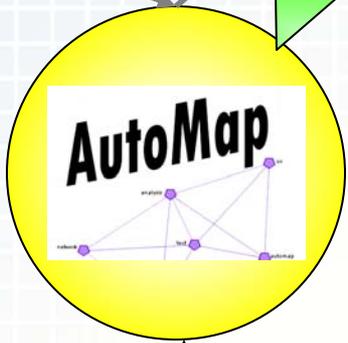
- Operationalize the techniques using CASOS tools:  
Raw data to simulation



Build Network - Text Mining

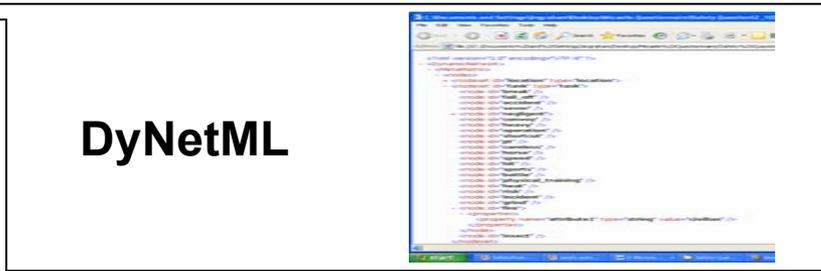
Analyze - Statistics SNA, DNA, Link Analysis

Assess Change, What if Analysis - Multi-agent DNA



**Meta-Network**

Node ID	Node Name	Node Type	Node Role	Node Status	Node Location	Node Contact	Node Description
1	...	...	...	...	...	...	...
2	...	...	...	...	...	...	...
3	...	...	...	...	...	...	...
4	...	...	...	...	...	...	...
5	...	...	...	...	...	...	...
6	...	...	...	...	...	...	...
7	...	...	...	...	...	...	...
8	...	...	...	...	...	...	...
9	...	...	...	...	...	...	...
10	...	...	...	...	...	...	...





# Importing Raw Data into ORA



# ORA

**ORA: Organizational Risk Analyzer v. 1.9.4.4**

File Edit Data Management Generate Networks Analysis Visualization Help

Meta-Network Manager

- embassy
  - agent : size 16
  - knowledge : size 4
  - resource : size 4
  - task : size 5
  - Agent x Agent
  - Agent x Knowledge
  - Agent x Resource
  - Agent x Task
  - Knowledge x Task
  - Resource x Task
  - Task x Task

Meta-Network:

Meta-Network ID

Meta-Network Filename

Statistics

Node Class Count:	0
Node Count:	0
Link Count:	0
Network Count:	0
Total Density:	---



# Raw Data is clearly not network-ready

CASOS tools can import data from:

- DynetML
- Excel/CSV
- Other SNA standards
- SQL
- Email
- Unstructured text data (via Automap)



# Loading Data from Excel/CSV

**Import Data into ORA** [X]

**Step 1:** Select a file containing table data:

[Text Field] [Browse]

**Step 2:** Check the columns that contain node IDs and enter the node class information:

[Large Empty Text Area]

**Step 3:** Define networks and attributes based on the columns:

Networks
  Networks and Labels
  Networks combined IDs
  Attributes

Source IDs	Target IDs	Link Weights	Network ID
[Empty Table Body]			

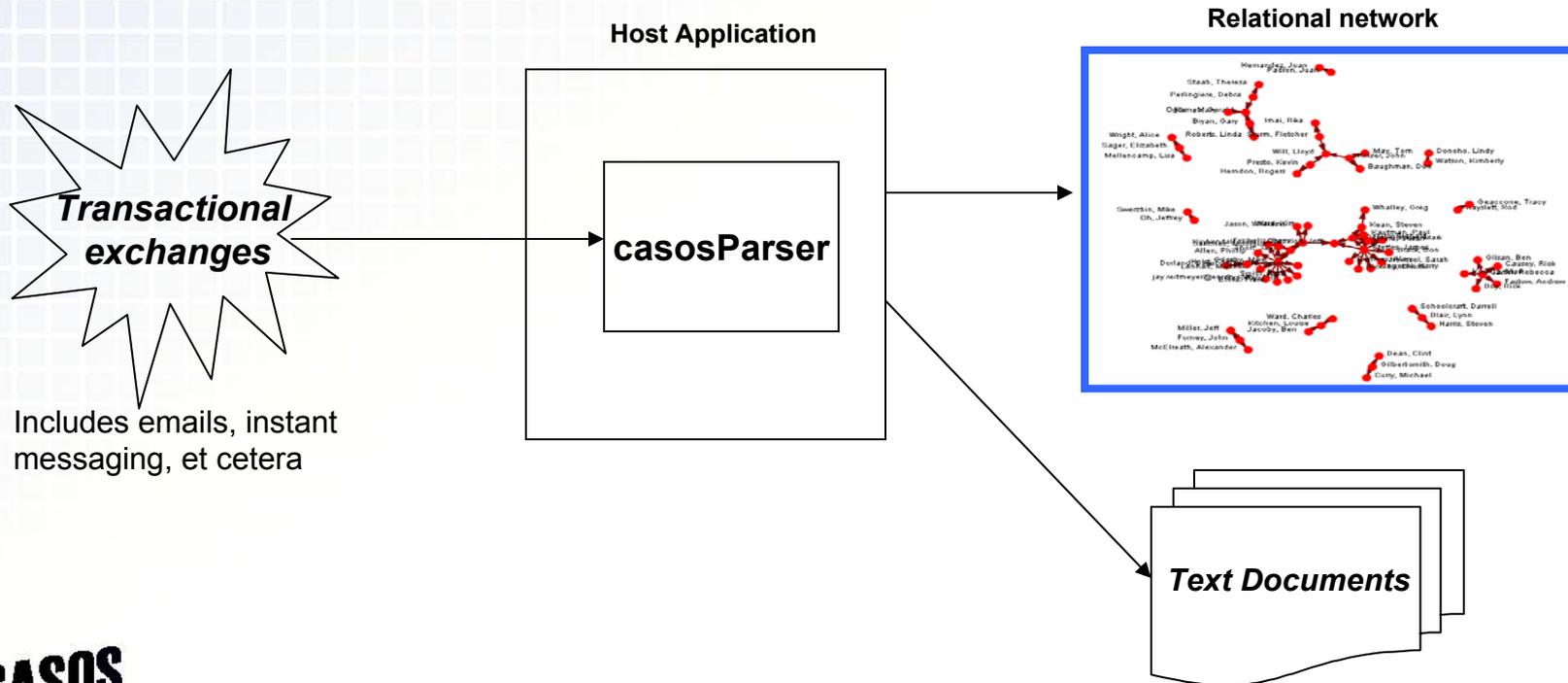
[New] [Clear]

Create new nodes for unrecognized node IDs
 [Load configuration] [Save configuration]



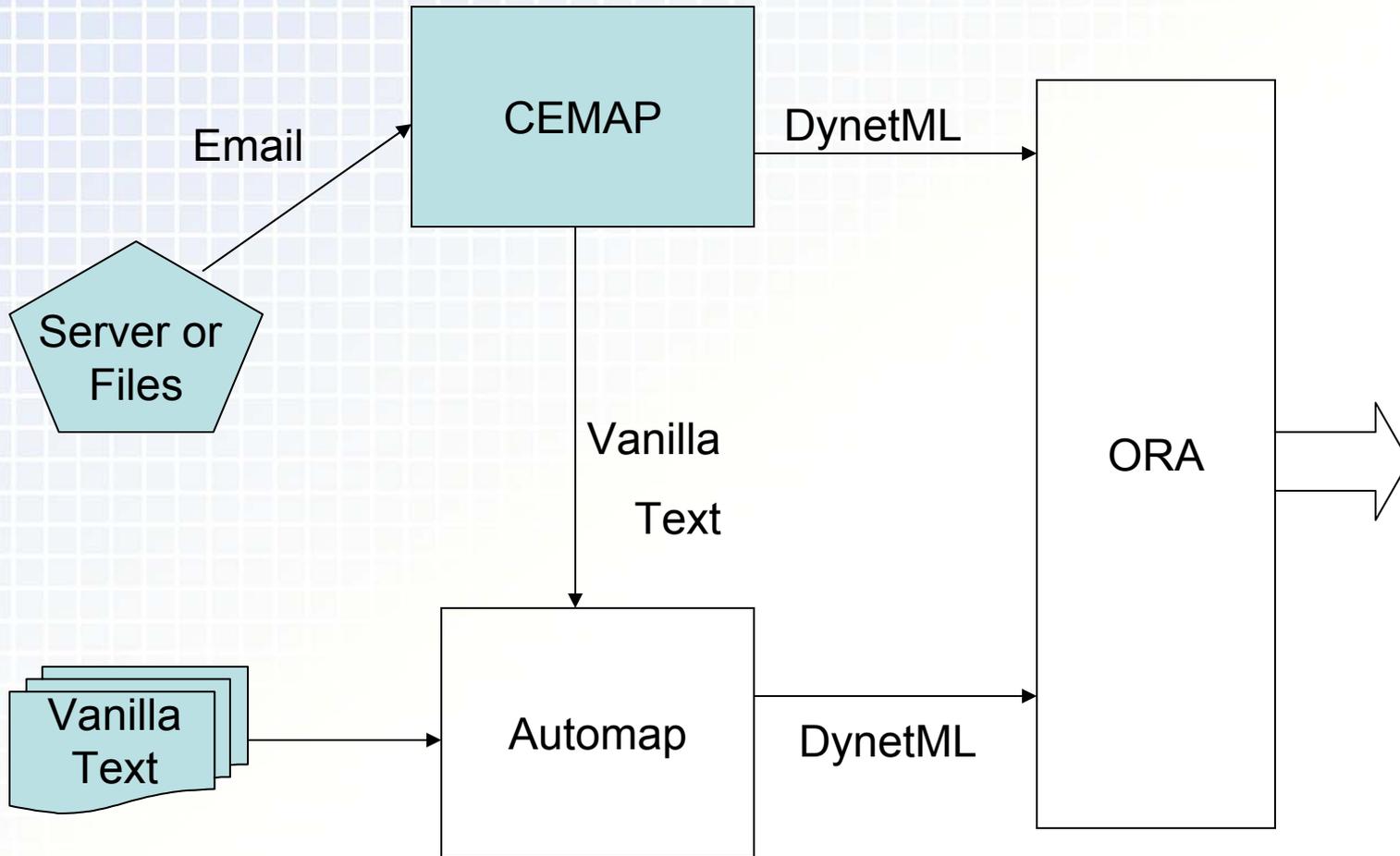
# Loading Email Data

CASOS CEMAP: CASOS-developed software technology that  
 (a) transforms raw, *transactional* text exchanges, into an  
 (b) exploitable *relational* structure and corresponding *document*.





# CEMAP for email processing





# Email Parser

**CASOS CEMAP - Email Analyzer**

Input Style	IMAP
Input File -or- HTTP URL	
Server	imap.srv.cs.cmu.edu
Port	
UserId	terrillFrantz
Password	*****
Mailbox/Folder	INBOX
Select Output Directory for full document bodies	
Select Output Directory for document bodies (new content Only)	
Select Output Directory for document bodies (forwarded content Only)	
Select Output Directory for SubjectLine Bodies	
Select Output Filename for Header Network File	
<input type="checkbox"/> Anonymize Agents	<input type="checkbox"/> De-Dupe Messages
<input type="checkbox"/> SSL (For IMAP and POP3 only)	Start extraction



# CEMAP loads email network directly into ORA

ORA: Organizational Risk Analyzer v. 1.9.4.4

File Edit Data Management Generate Networks Analysis Visualization Help

Meta-Network Manager

- Meta Matrix
  - EAGENT : size 769
  - EMAIL : size 1284
  - SUBJECT : size 1154
  - eAgent2eAgent
  - FROM
  - eAgent2Subject
  - eAgentBCC2Subject
  - eAgentCC2Subject
  - eAgentFROM2Subject
  - eAgentTO2Subject
  - BCC
  - CC
  - TO

Meta-Network:

Meta-Network ID:

Meta-Network Filename:

Statistics

Node Class Count:	0
Node Count:	0
Link Count:	0
Network Count:	0
Total Density:	---



# Visualizing Data in ORA



# Visualize person-to-person network derived from 1,280 Personal Emails

Meta Matrix - ORA Network Visualizer

File Actions Tools Layouts Options Help

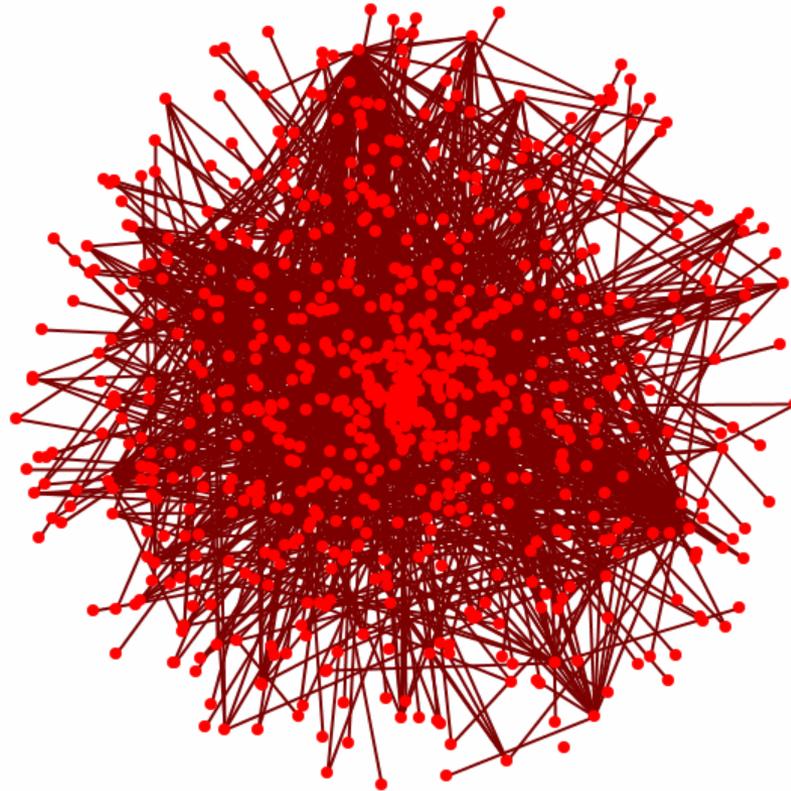
Rotate Show Labels Show Edges Show Arrows Font Size 10 Node Size 8 Link Width 2

Hide links with weight: Less Than 0.0

Legend

Edit Control

- EAGENT : size 769
- eAgent2eAgent



Zoom: -16 Hyperbolic: 0 769 nodes, 1247 links





# Visualize Multi-network Data

embassy - ORA Network Visualizer

File Actions Tools Layouts Options Help

Rotate  Show Labels  Show Edges  Show Arrows

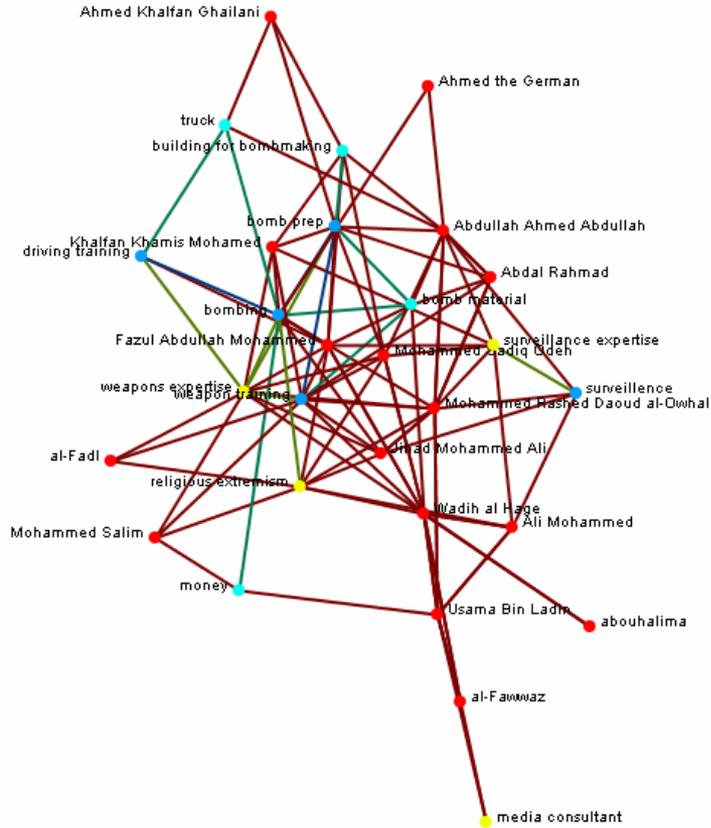
Font Size: 10 Node Size: 8 Link Width: 2

Hide links with weight: Less Than 0.0

Legend

Edit Control

- agent : size 16
- knowledge : size 4
- resource : size 4
- task : size 5
- Agent x Agent
- Agent x Knowledge
- Agent x Resource
- Agent x Task
- Knowledge x Task
- Resource x Task



Zoom: -5 Hyperbolic: 0 29 nodes, 106 links

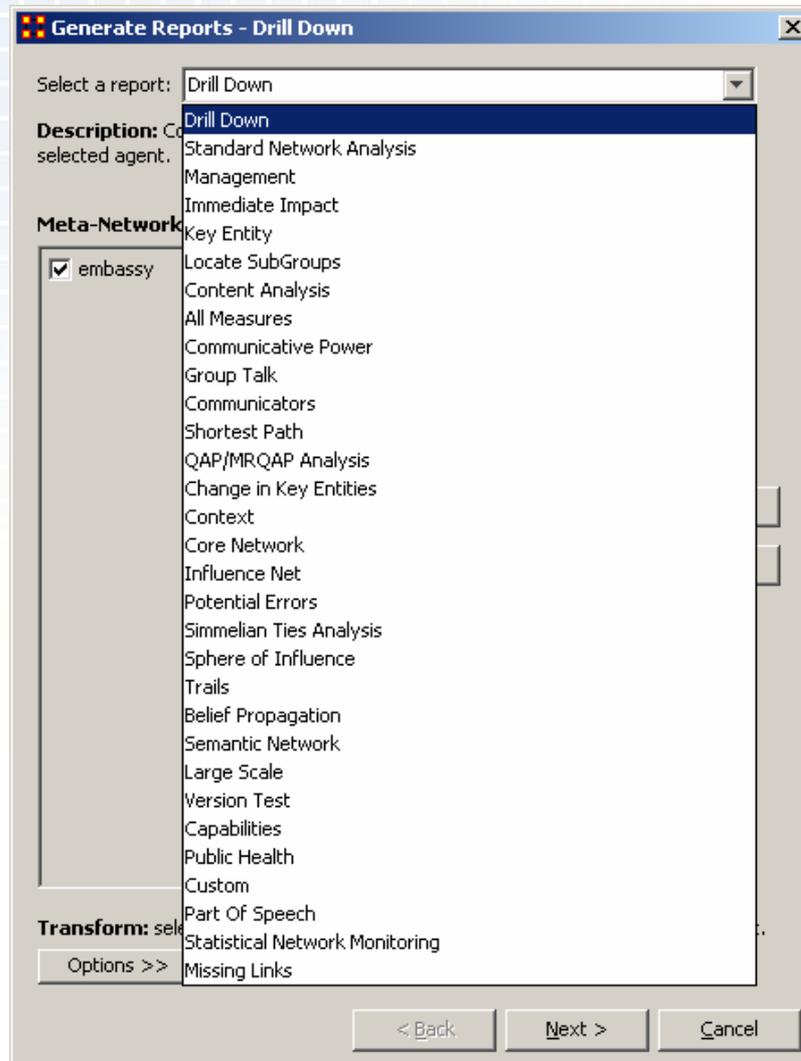




# Creating Reports in ORA



# Dozens of pre-formed reports, or custom





# Report output to main screen ... or browser

ORA: Organizational Risk Analyzer v. 1.9.4.4

File Edit Data Management Generate Networks Analysis Visualization Help

Meta-Network Manager

- embassy
  - agent : size 16
  - knowledge : size 4
  - resource : size 4
  - task : size 5
  - Agent x Agent
  - Agent x Knowledge
  - Agent x Resource
  - Agent x Task
  - Knowledge x Task
  - Resource x Task
  - Task x Task

Meta-Network: embassy

Meta-Network ID: embassy

Meta-Network Filename: C:\Documents and Settings\terrill\My Documents\dataSets\embassy.xml

Reports Visualize View Charts

Statistics

Node Class Count: 4

Node Count: 29

Link Count: 106

Network Count: 7

Total Density: .20866142

MANAGEMENT REPORT

MANAGEMENT REPORT

Input data: embassy

Start time: Tue Mar 18 15:22:29 2008

KEY INDIVIDUALS

This chart shows the Agent that repeatedly rank in the top three in the measures. The value shown is the percentage of measures for which the Agent was ranked in the top three.

Emergent Leader (cognitive demand)

Measures the total amount of cognitive effort expended by each agent to do its tasks.

Input network(s): Agent x Agent

Rank	Value	Agent
1	0.3789	Fazul Abdullah Mohammed
2	0.3411	Khalfan Khamis Mohamed
3	0.2991	Mohammed Rashed Daoud al-
4	0.2803	Jihad Mohammed Ali
5	0.2699	Mohammed Sadik Odah

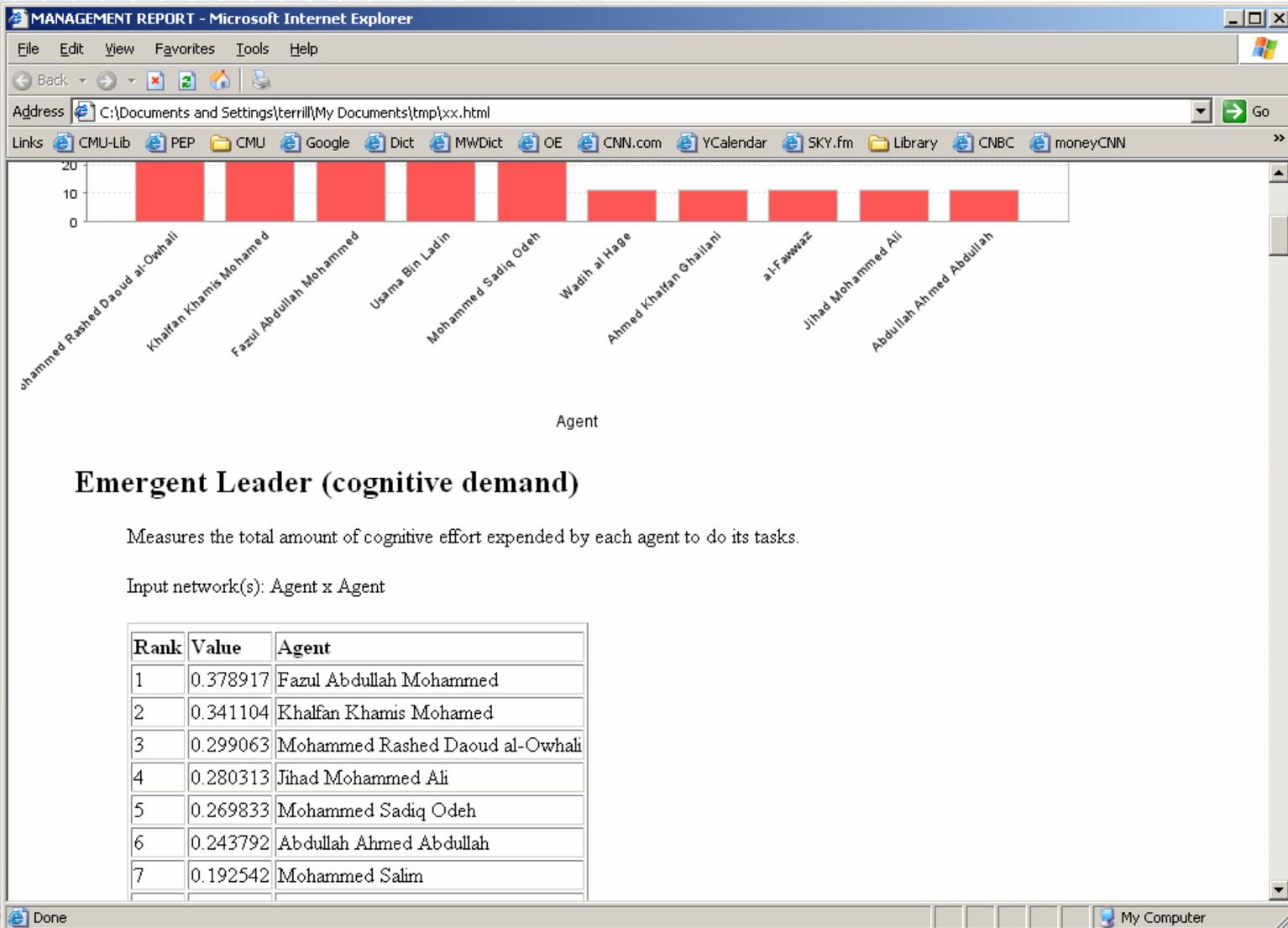


# Management Report





# Management Report (page down)

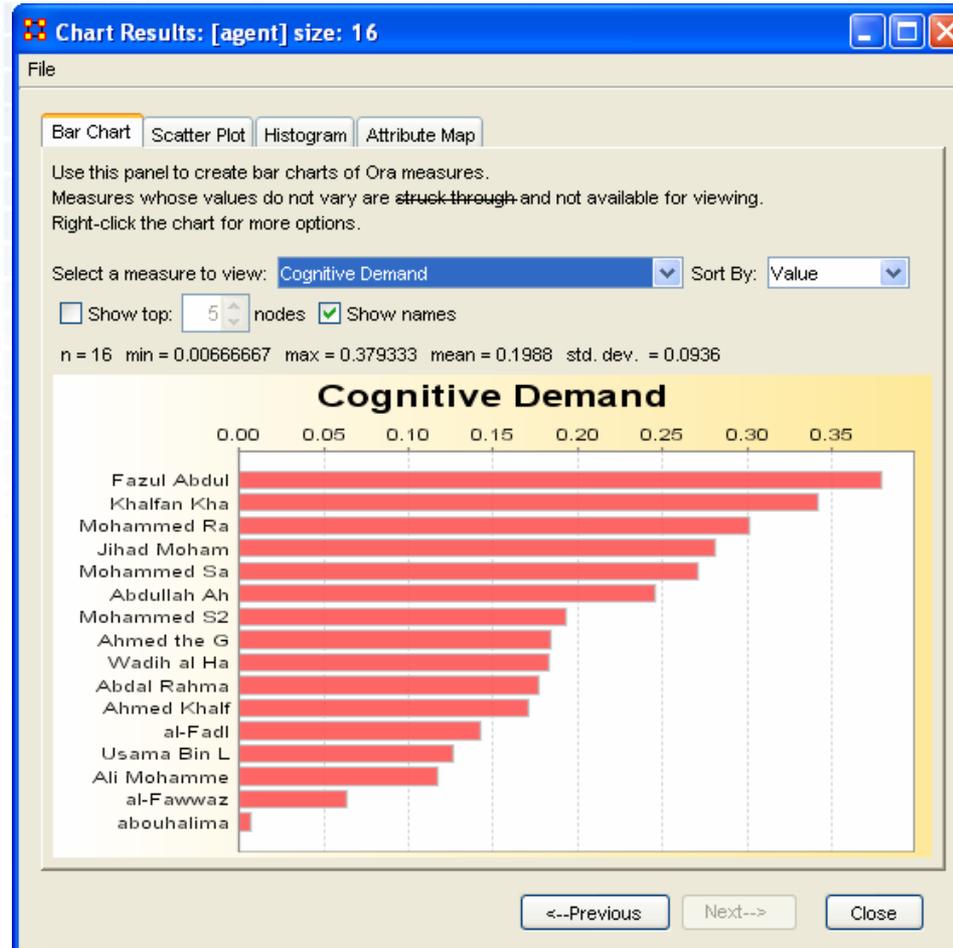




# Charting in ORA

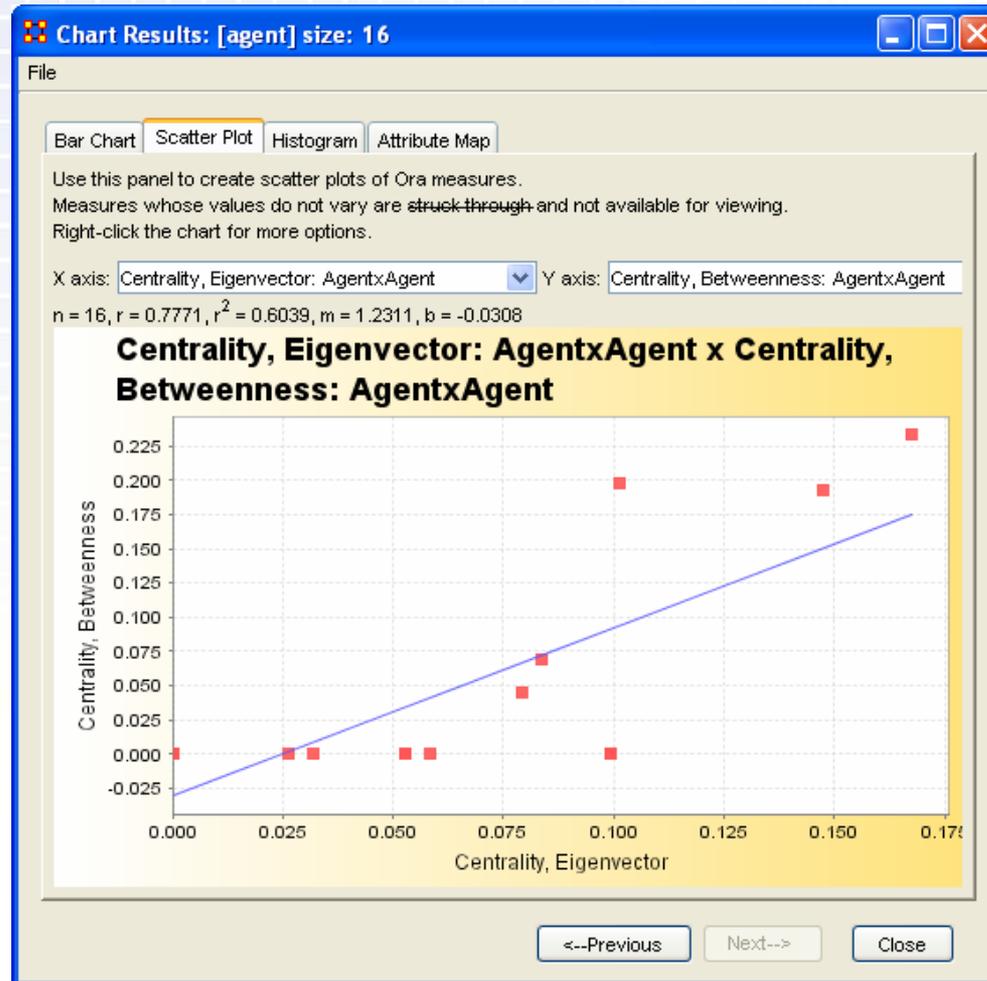


# Chart Tool – Bar Chart



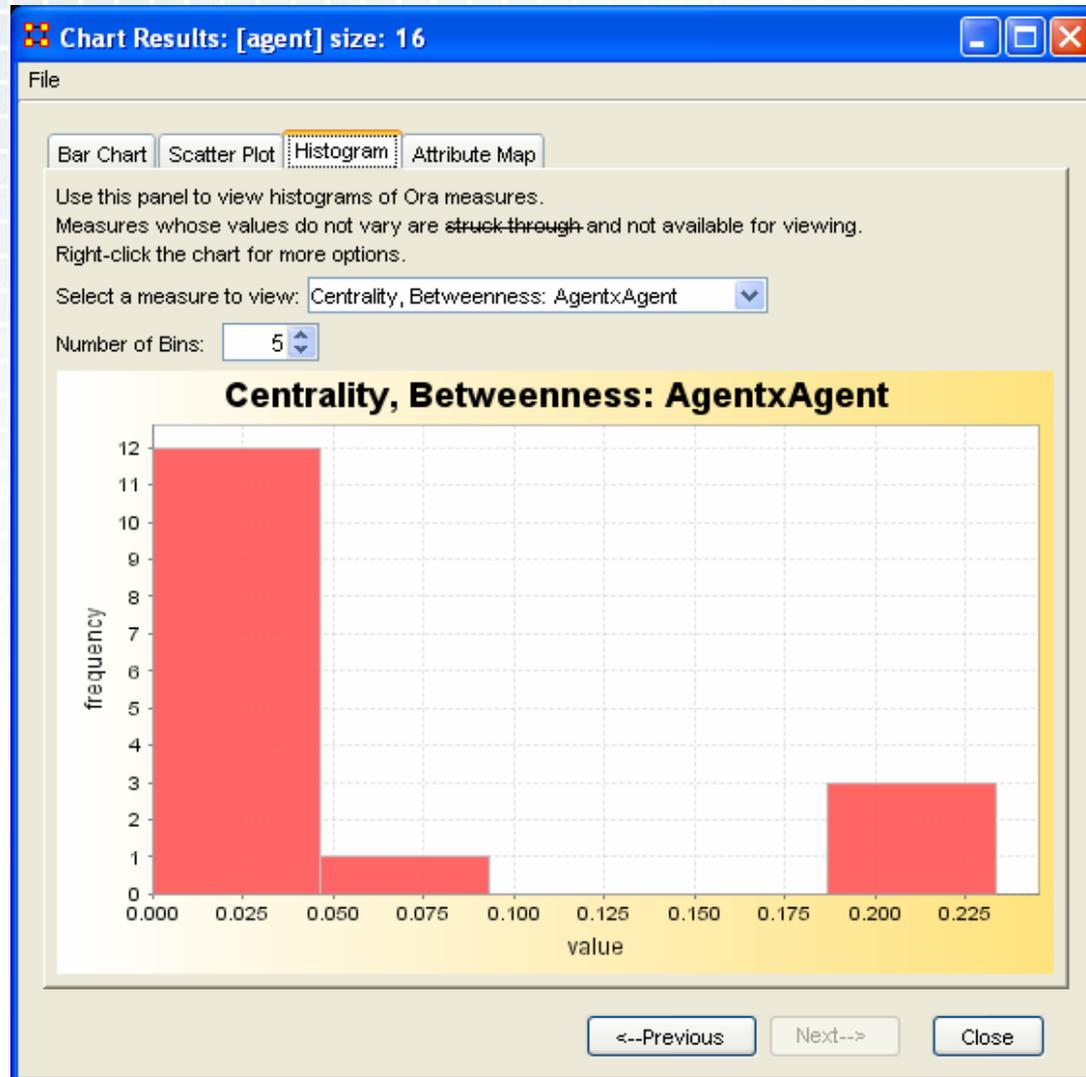


# Chart Tool– Scatter Plot





# Chart Tool– Histogram





# Automap



# Automap: Transform documents into networks

AutoMap-2.6.70

File Run Analysis Tools Help

Go to:  OK File name: C:\Documents and Settings\terrill\My Documents\tmp\robustness.txt

8. Texts after Parts-of-Speech Tagging      9. Texts after Anaphora Resolution  
 6. Texts after Meta-Matrix Thesaurus      7. Texts after Sub-Matrix Selection  
 4. Texts after Deletion                      5. Texts after Generalization  
 1. Original Texts                              2. Texts after Symbol Removal              3. Texts after Stemming

1. Introduction  
 Social network datasets are often incomplete and prone to observation error due to the intricacy of collection-instrument design and the inherent vagueness of human-informant reliability and bias (Stork and Richards, 1992; Feld and Carter, 2002). The error in the observed data may be unintentional or intentional (Albert et al. , 2000; Carley, 2002; Calloway and Morrissey, 1993; Freeman et al. , 1987; Killworth and Bernard, 1976). No matter its nature, the presence of this error is a nontrivial issue (Marsden, 1990; McKnight et al. , 2007) and it raises the question of the impact of the uncertainty, relative to the accuracy of network measures computed from this data. In extensive efforts to ameliorate this problem, researchers have been

1. Concept List   2. Union Concept List   3. Pre-Processing Settings   4. Analysis Settings

Concept	Frequency	In Delete List	Add to Delete List	Translation in Th...
the	853	<input type="checkbox"/>	<input type="checkbox"/>	
of	471	<input type="checkbox"/>	<input type="checkbox"/>	
and	321	<input type="checkbox"/>	<input type="checkbox"/>	
a	223	<input type="checkbox"/>	<input type="checkbox"/>	
in	213	<input type="checkbox"/>	<input type="checkbox"/>	
network	193	<input type="checkbox"/>	<input type="checkbox"/>	
to	166	<input type="checkbox"/>	<input type="checkbox"/>	
is	130	<input type="checkbox"/>	<input type="checkbox"/>	
for	122	<input type="checkbox"/>	<input type="checkbox"/>	
that	109	<input type="checkbox"/>	<input type="checkbox"/>	
topology	84	<input type="checkbox"/>	<input type="checkbox"/>	
we	83	<input type="checkbox"/>	<input type="checkbox"/>	
on	71	<input type="checkbox"/>	<input type="checkbox"/>	
as	67	<input type="checkbox"/>	<input type="checkbox"/>	
data	67	<input type="checkbox"/>	<input type="checkbox"/>	
networks	67	<input type="checkbox"/>	<input type="checkbox"/>	
this	67	<input type="checkbox"/>	<input type="checkbox"/>	

Semantic network of current Text

1. Action Tracer Panel   2. Statistics   3. Network analytic measures

uments and Settings\terrill\My Documents\tmp\robustness.txt

Start | C:\ | A. | F. | r.. | 5:04 PM

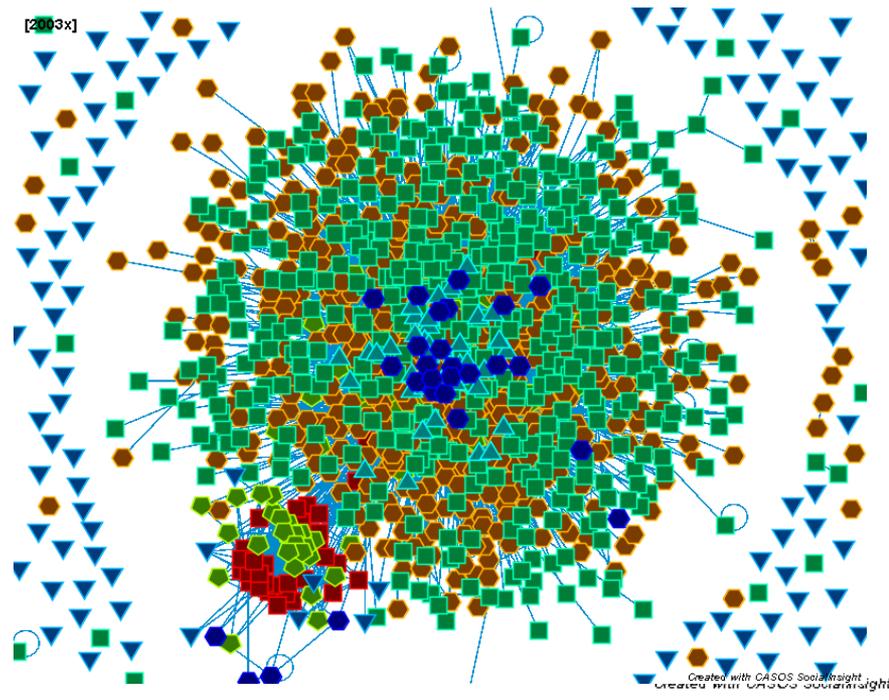
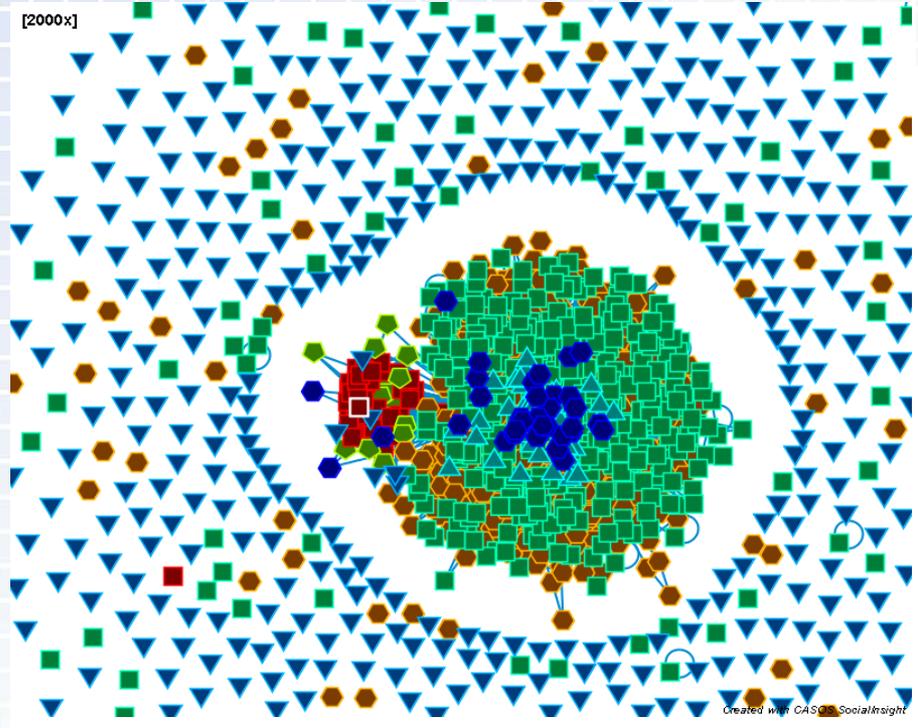




# Networks Over-time & Change Detection

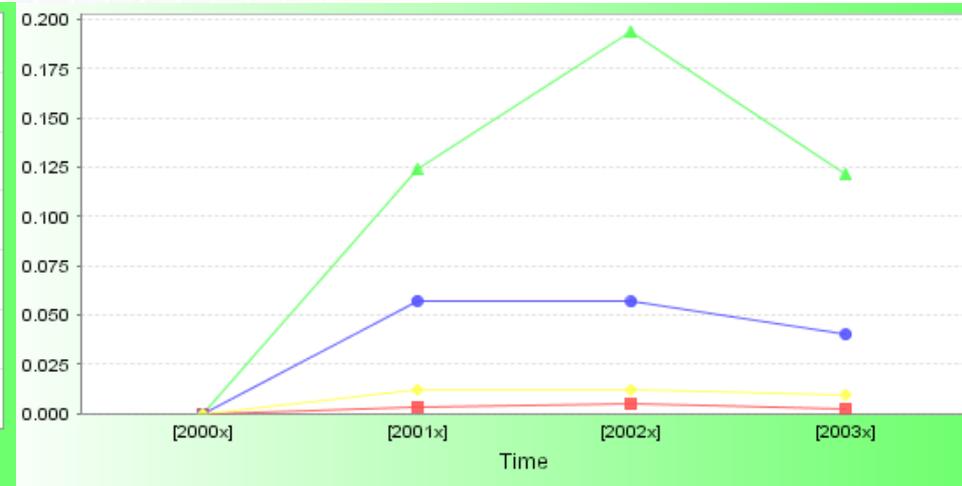
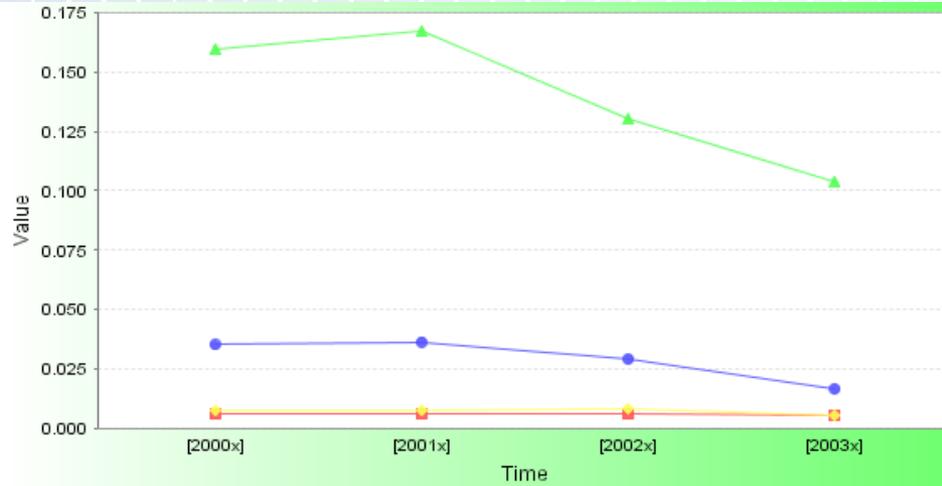


# Change 2000 - 2003



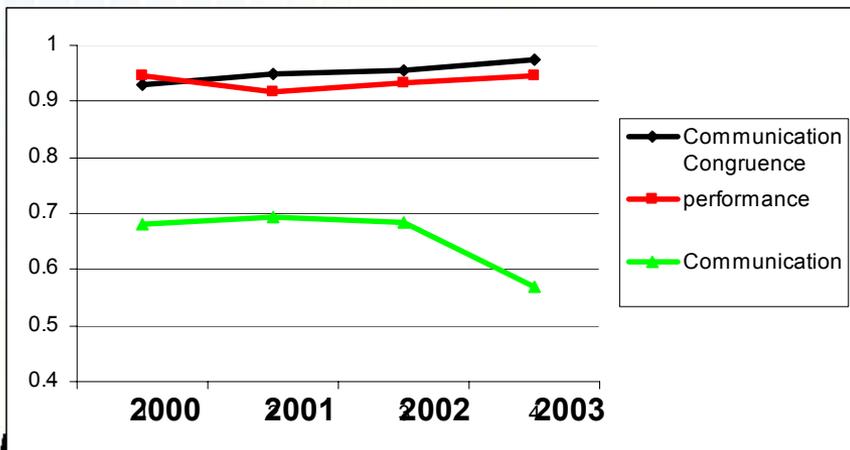


# Change 2000-2003



■ Meta-Matrix Density 
 ● Communication Density 
 ▲ Knowledge Density 
 ◆ Task Density

■ Meta-Matrix Hamming Distance 
 ● Communication Hamming Distance 
 ▲ Knowledge Hamming Distance 
 ◆ Task Hamming Distance





# Simulation: The Near-term Impact Report





# Immediate Impact - Prediction

- *What if ? Remove top 5 emergent leaders*
- Change in performance
  - Anticipated drop – 4% percentage difference
- Change in information diffusion
  - Anticipated increase – 67% percentage difference
- New emergent leaders
  1. 0.0174 said\_mortazavi
  2. 0.0137 kamal\_kharazi
  3. 0.0127 reza\_asefi
  4. 0.0120 morteza\_sarmadi
  5. 0.0100 hashemi\_shahrudi
- Value of “lowest” old emergent leader was .0246

**Immediate  
Impact**



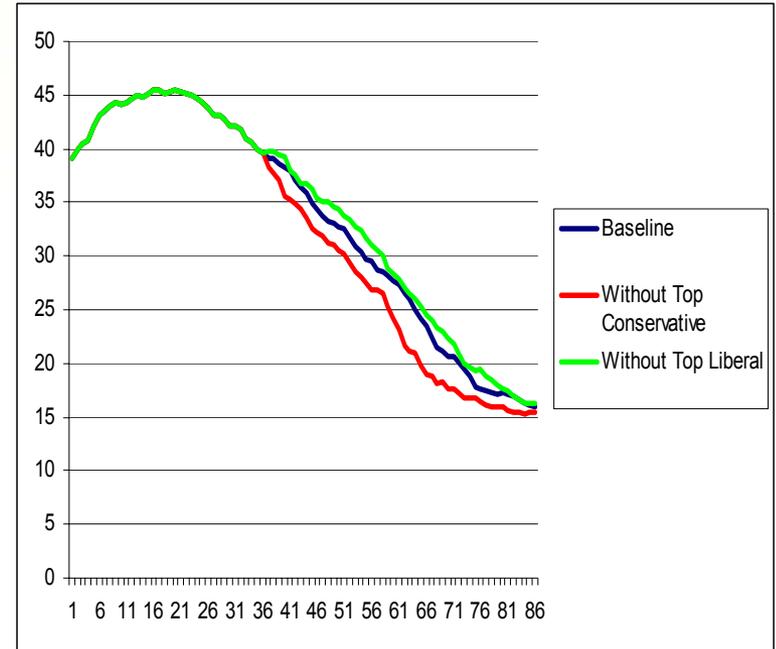
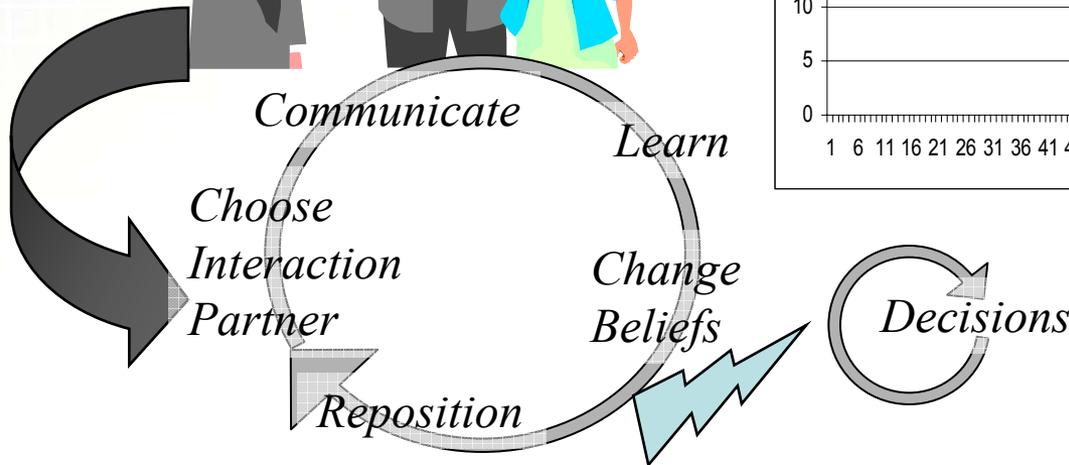
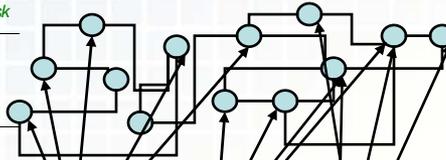
# Who are likely to be new Emergent Leaders?

	<i>Who is Isolated</i>	<i>Baasyir</i>	<i>Bin Laden</i>	<i>High Cognitive Demand</i>	<i>Weak Boundary Spanner</i>
	1	Gokhan	Gokhan	Kandari	Kandari
	2	MaFadli	Al Ha Ghamdi	Aufi	Nawar
	3	Tabarak	Benyaich	Benali	Aufi
	4	Al Ha Ghamdi	Maqbul	HaGhamdi	Jabarah1
	5	Aufi	MaFadli	MOShehri	Ameroude



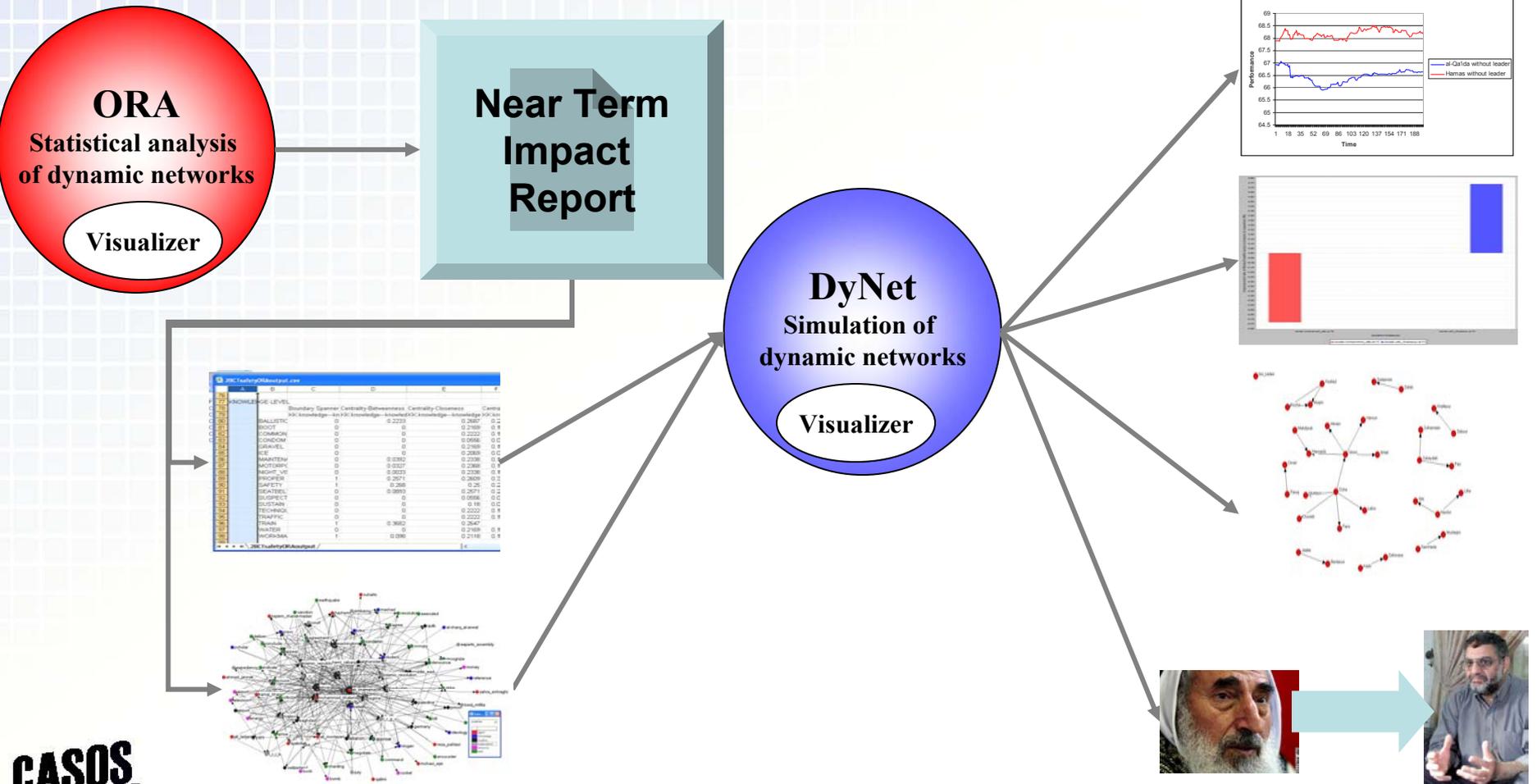
# Change in Reformism – Near Term Prediction

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





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

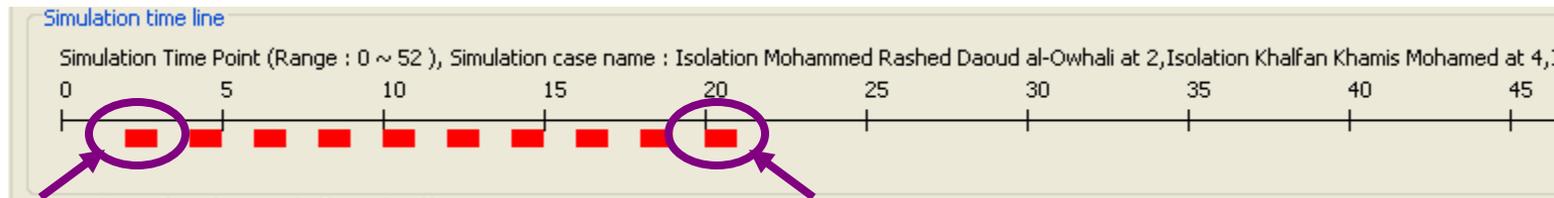


KNOWLEDGE LEVELS	Secondary	Spencer	Centrality	Betweenness	Centrality-Choiness	Centra
BALLISTIC	0	0	0.2607	0.2		
BOOT	0	0	0.2709	0.9		
COMACH	0	0	0.2222	0.9		
CONDOM	0	0	0.0966	0.0		
GHANEL	0	0	0.2169	0.9		
ICE	0	0	0.2908	0.0		
MAKRENU	0	0	0.2328	0.9		
MICROPS	0	0	0.0327	0.2988	0.9	
NOGHT_VIE	0	0	0.2328	0.9		
PROPER	1	0.2571	0.2629	0.3		
SAFETY	1	0.268	0.2629	0.3		
SEATBEL	0	0.0893	0.2671	0.2		
SUSPECT	0	0	0.0866	0.0		
SUSTAIN	0	0	0.18	0.0		
TECHNOG	0	0	0.2222	0.9		
TRAFFIC	0	0.2822	0.2222	0.9		
TRAIL	1	0.2647	0.2647	0.9		
WATER	0	0	0.2169	0.9		
WORLDWAS	1	0.096	0.2118	0.9		



# Input data - Isolation strategy (1)

- To make a what-if scenario, we need a set of agents for exploring isolation strategies
  - Input from ORA
    - Intelligence Report identifies several key nodes (agent, knowledge, resource, organization)
    - Isolate each node from Intelligence Report that is identified by any measures one at a time
    - Isolate all the top nodes for a measure as identified in Intelligence Report and repeat for each measure
  - Input from User
    - Isolate nodes as a user specifies



First intervention, isolate al-Owahali at time-step 2

Last intervention, isolate sadiq-odeh at time-step 20



# Input data - Isolation strategy (2)

- Intelligence report includes the lists of nodes identified by various measures.

Measure	Implication
Cognitive demand	Measures the total amount of effort expended by each agent to do its tasks.
Total degree centrality	The Total Degree Centrality of a node is the normalized sum of its row and column degrees.
clique count	The number of distinct cliques to which each node belongs.
eigenvector centrality	Calculates the principal eigenvector of the network. A node is central to the extent that its neighbors are central.
betweenness centrality	The Betweenness Centrality of node $v$ in a network is defined as: across all node pairs that have a shortest path containing $v$ , the percentage that pass through $v$ .
high betweenness and low degree	The ratio of betweenness to degree centrality; higher scores mean that a node is a potential boundary spanner.
shared situation awareness	A measure of situation awareness between agents.



# Near Term Analysis Result

- The changes of performance measures
  - Knowledge Diffusion
  - Binary Task Accuracy
  - Energy Task Accuracy
  - Belief Level
- The changes of network itself
  - An evolved meta-matrix
- These results shows the near term impact of isolating one agent in organization.



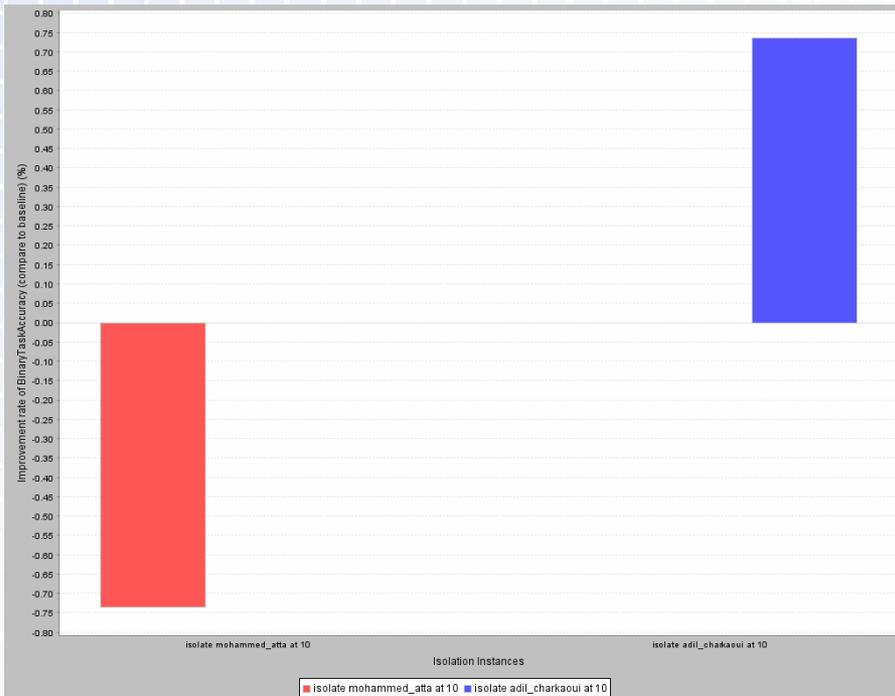
# Destabilization Analysis

- Various aspects in destabilization
  - Performance drops
    - Slower information diffusion
    - Lower task accuracy
    - Worse task execution preparation
  - Evolved network comparisons
    - Skewed knowledge distribution
    - Leader changes and power shifts
    - Fragmentations

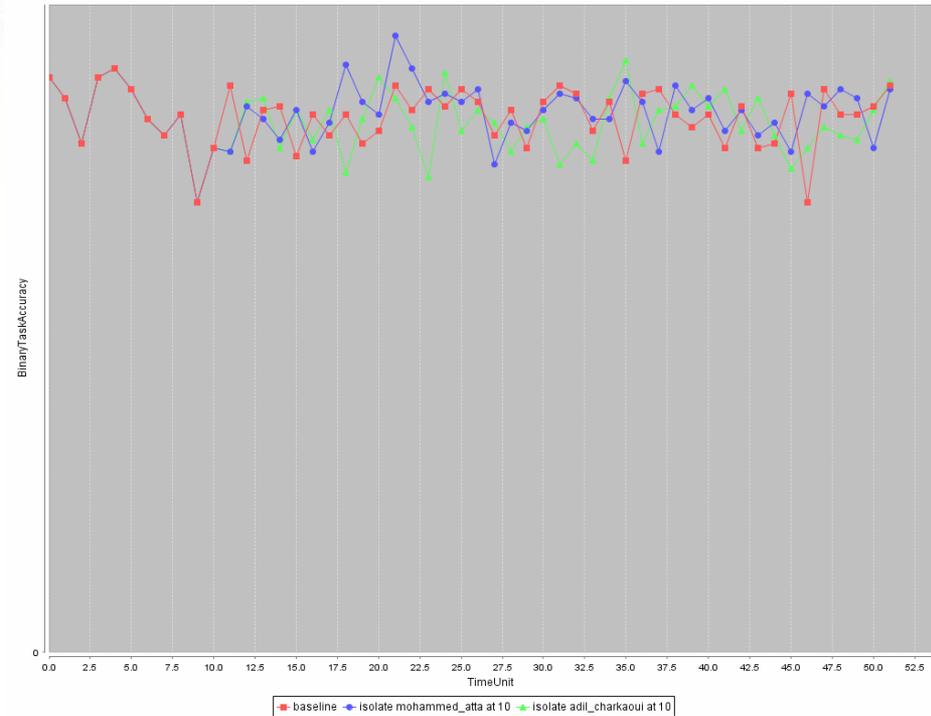


# Impact of isolation

## ORA: Near Term Impact Report



Ending Performance



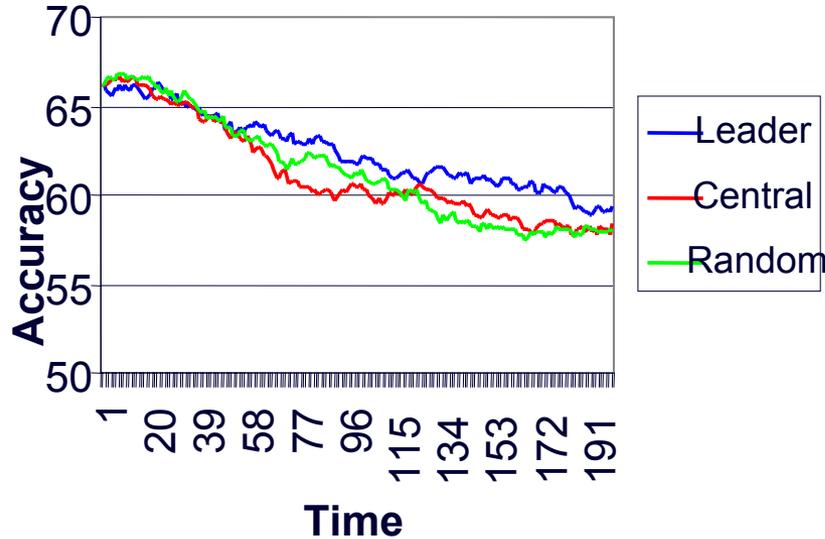
Performance Over Time



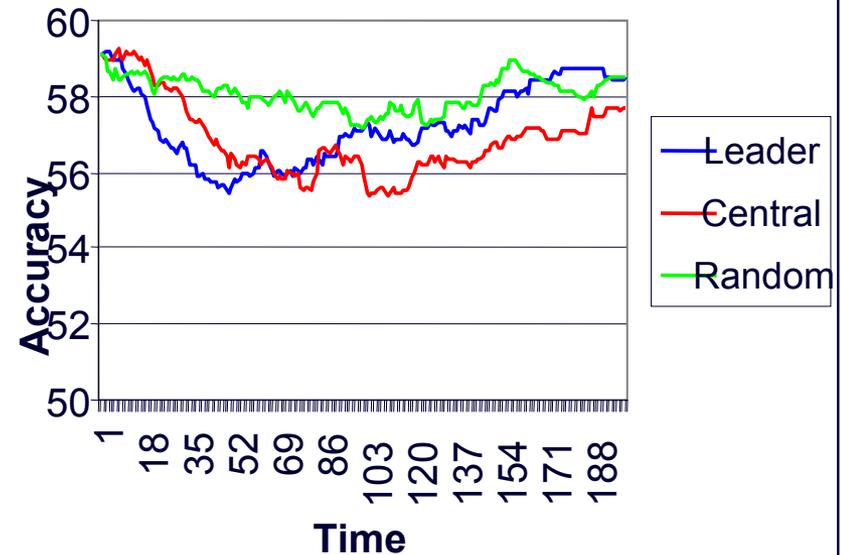


# Impact is exacerbated when forecasting

## Impact of Isolation Strategy on Performance for Random Network

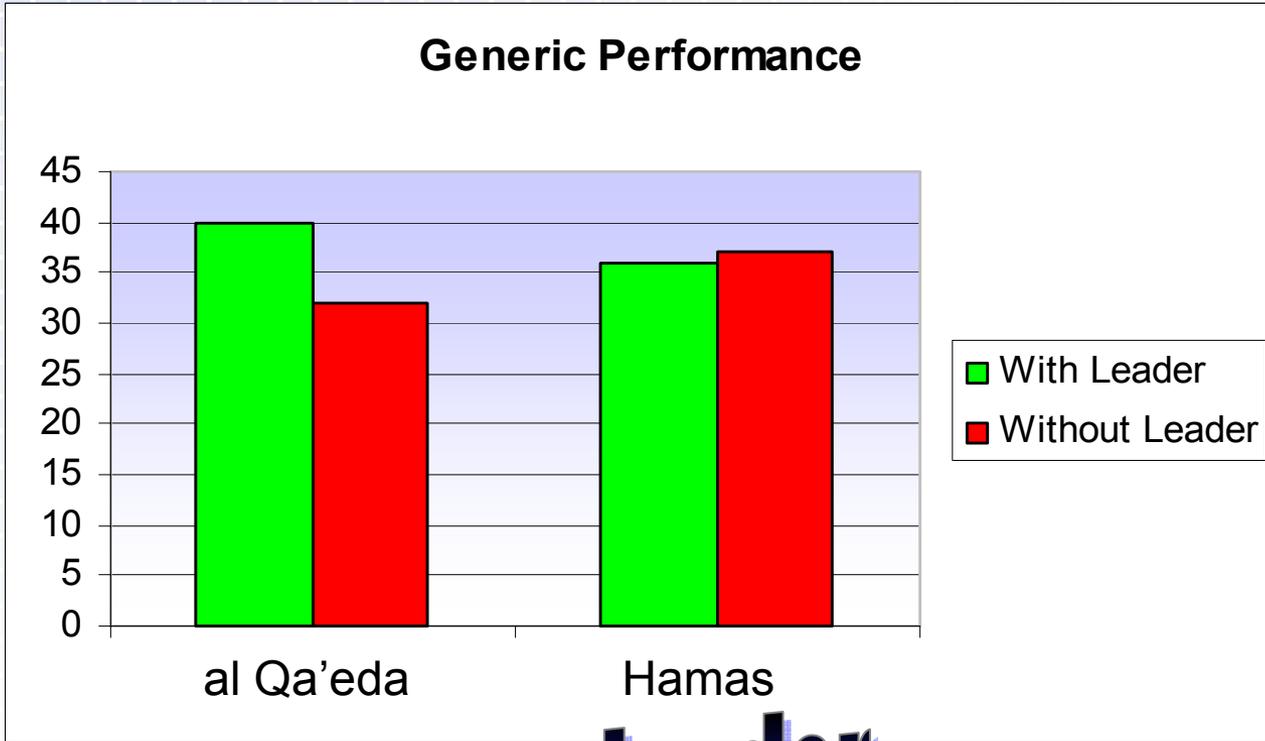


## Impact of Isolation Strategy on Performance for Cellular Network





# What If Analysis



# New Emergent Leader



Bin Laden



al-Zawahiri



Yassin



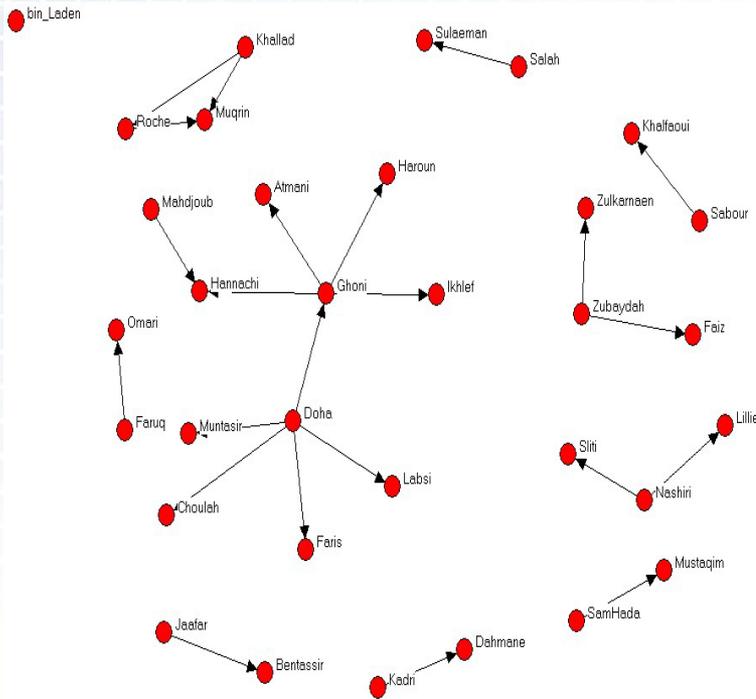
Rantisi





# DyNet Analysis Output

Bin Laden Isolation → New Network Relationships



New Relationships based on bin Laden Isolation

From	To	Value	NormValue
Salah	Sulaeman	0.07378	1
Nashiri	Sliiti	0.016457	0.223052
Ghoni	Haroun	0.012465	0.168949
Zubaydah	Faiz	0.012083	0.163765
Sabour	Khalfaoui	0.011504	0.155918
Ghoni	Ikhlef	0.011457	0.155283
Nashiri	Lillie	0.01078	0.146111
Ghoni	Atmani	0.01067	0.144618
Doha	Ghoni	0.009806	0.132915
Mahdjoub	Hannachi	0.009584	0.129906
Roche	Muqrin	0.009504	0.128819
Ghoni	Hannachi	0.009374	0.12705
Doha	Faris	0.009175	0.124359



# Near-term Impact Report

**Near Term Analysis** [Close] [Maximize] [Minimize]

File Simulation Output Options

---

Input Data

MetaMatrix name

---

Settings

Global parameters

The number of replications - The replications mean repeated simulations with a same simulation setup and different random seeds.

The number of simulated time-points - The number of simulated time points determines the length of each simulation. Dynet will iterate the agent interactions for the number of times

---

Simulation time line

This is a simulation event visualization of a highlighted event in the below list. Simulation Time Point (Range : 0 ~ 5.0 ), Simulation case name : baseline

---

Cases to simulate

**Total 1 runs = 1 simulation cases X 1 replications for each case**

Isolation Case	Isolation Information	Delete	Edit
baseline	baseline		





# Define Intervention Scenarios

**Near Term Analysis**

File Simulation Output Options

Input Data

MetaMatrix name:

Settings

Global parameters

The number of replications - The replications mean repeated simulations with a same simulation setup and different random seeds.

The number of simulated time-points - The number of simulated time points determines the length of each simulation. Dynet will iterate the agent interactions for the number of times

Simulation time line

This is a simulation event visualization of a highlighted event in the below list. Simulation Time Point (Range : 0 ~ 5.0 ), Simulation case name : baseline

Cases to simulate

**Total 2 runs = 2 simulation cases X 1 replications for each case**

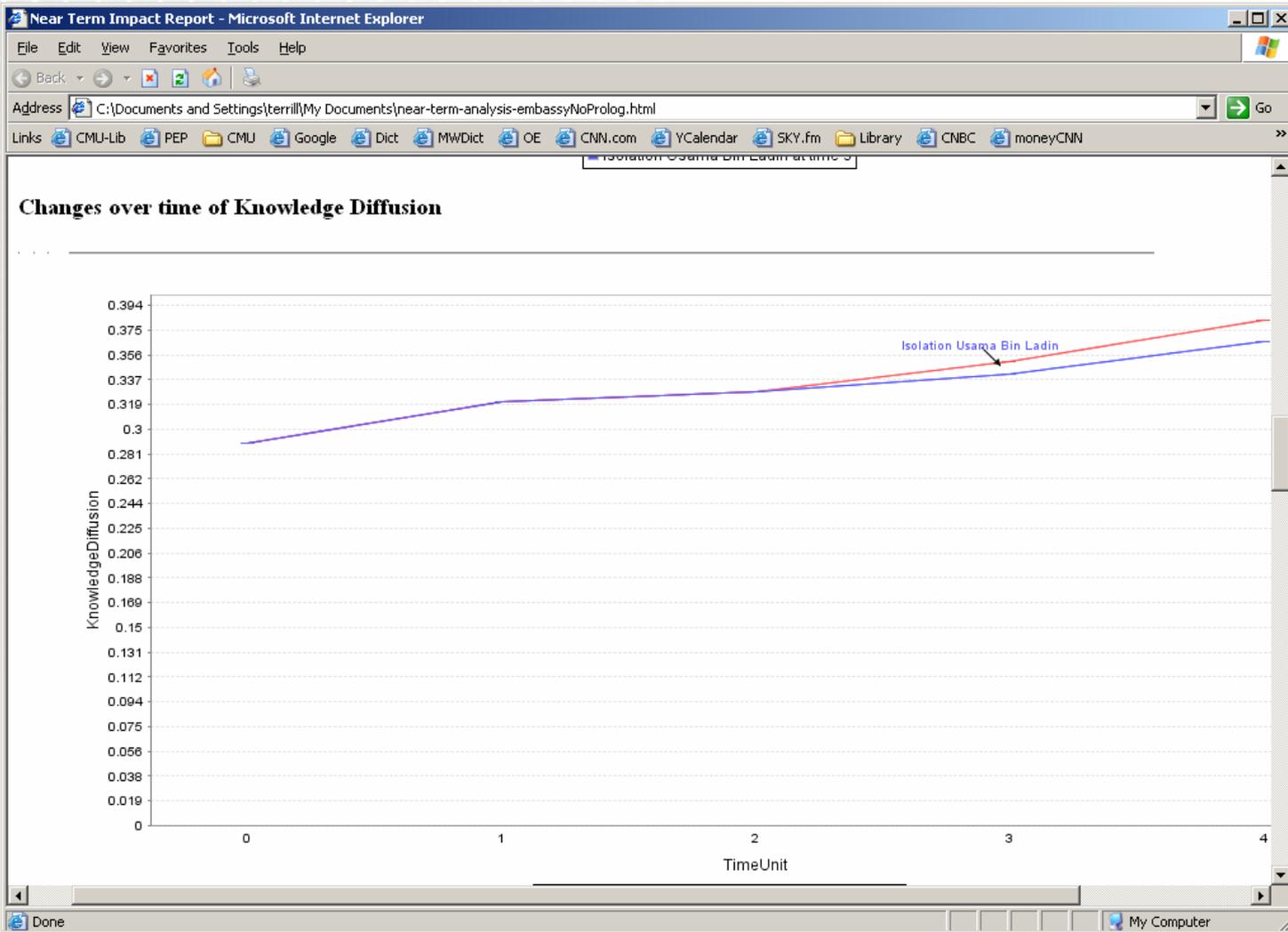
Isolation Case	Isolation Information	Delete	Edit
baseline	baseline		
A7_3	Isolation Usama Bin Ladin at time 3	Delete	Edit

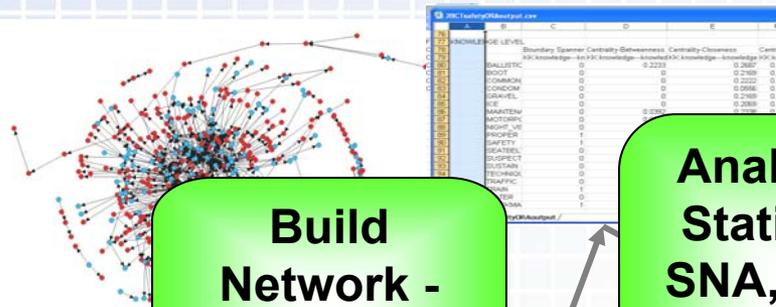
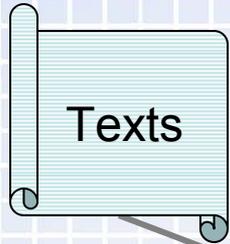
Execute Cancel





# Reports: Scenarios vs. Baseline

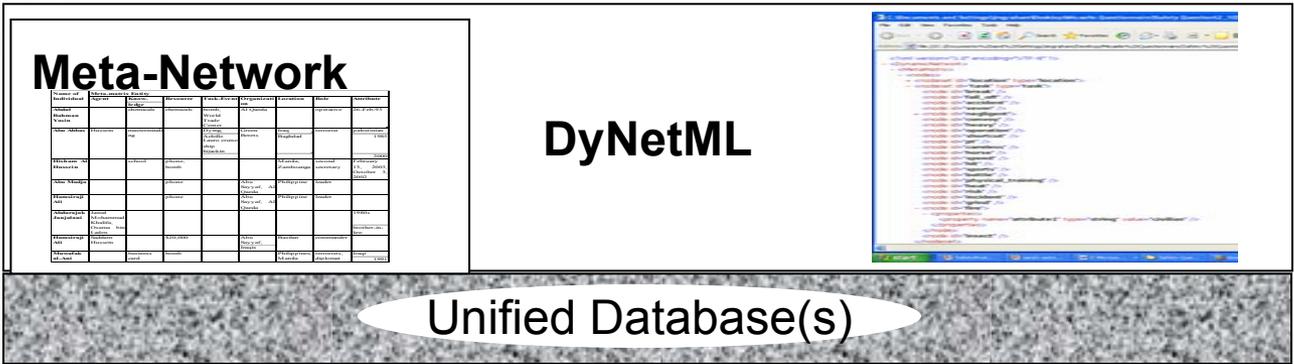
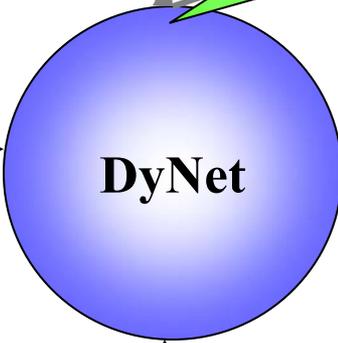
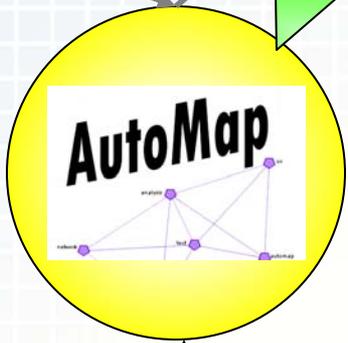




Build Network - Text Mining

Analyze - Statistics SNA, DNA, Link Analysis

Assess Change, What if Analysis - Multi-agent DNA





# Acknowledgements

- This work is part of the Dynamics Networks project at the center for Computational Analysis of Social and Organizational Systems (CASOS) of the School of Computer Science (SCS) at Carnegie Mellon University (CMU). This work is supported in part by the Office of Naval Research (ONR), United States Navy. Additional support was provided by National Science Foundation (NSF) Integrative Graduate Education and Research Traineeship (IGERT) program, NSF 045 2598, NSF 045 2487, and CASOS. MURI: Air Force Office of Scientific Research, FA9550-05-1-0388 for Computational Modeling of Cultural Dimensions in Adversary Organizations. SPAWAR, Network Analysis and Computational Modeling for Combating Terrorist Threats, MOAT Phase II, DNA application to counter-narcotic investigations related to marijuana. The views and proposal contained in this document are those of the author and should not be interpreted as representing the official policies, either expressed or implied, of the Office of Naval Research, the National Science Foundation, or the U.S. government.