Centralities - Measures and Reports

Geoff Dobson
gdobson@andrew.cmu.edu

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

Carnegie Mellon

Agenda

- Guiding Questions
- Review of Centrality Measures
- ORA Walk Through 1
- ORA Walk Through 2
- ORA Walk Through 3

Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

Guiding Questions

- Who is most important to this operation?
- Is Fred a connector or a broker?
- Which shipping port is most vulnerable?
- Which doctor should receive the referral bonus?

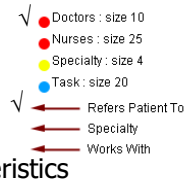
CASOS

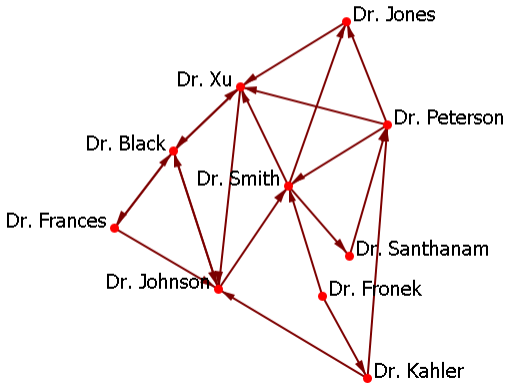
 3

Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

Guiding Questions

- Question?
- Please list all the Drs. to whom you have referred patients in the past six months.





1. Characteristics
   - One Way
   - One mode
   - One type of link
     - Binary
     - Directed

CASOS

 4

CASOS

Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

Steps in a Structural Analysis

- Collect network data.
  - Connections among people, knowledge, resources, events …
- Enter data into ORA.
- Visualize.
- Generate Report.
- If multiple networks create combined measures.
- If needed look at some measures in more depth.
- Possibly drop isolates and pendants
- Check interpretations.

 5

Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

Ego-centric

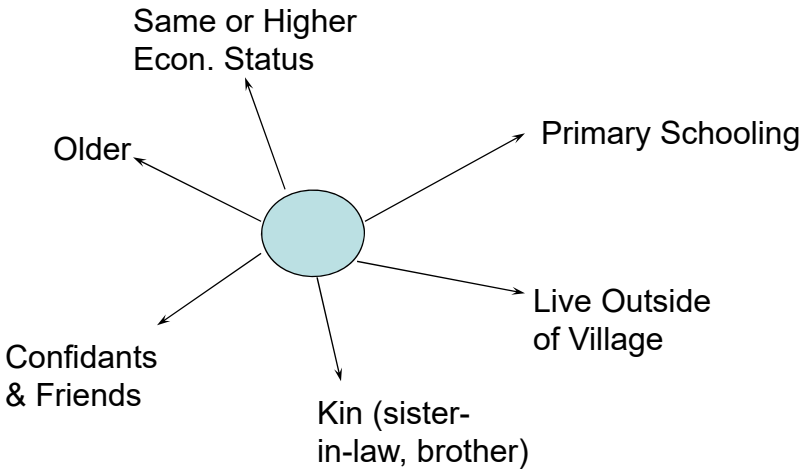
- Please tell me, within the last six months, the first names or initials of up to 5 people you talk to most often about important matters?

Respondent					
1 2 3 4 5	Name _____				
1 2 3 4 5	No Yes	Name _____			
1 2 3 4 5	No Yes	No Yes	Name _____		
1 2 3 4 5	No Yes	No Yes	No Yes	Name _____	
1 2 3 4 5	No Yes	No Yes	No Yes	No Yes	Name __

 6

Carnegie Mellon ISR institute for SOFTWARE RESEARCH

Ego-centric Network Data are Node-based



CASOS

Carnegie Mellon ISR institute for SOFTWARE RESEARCH

Critical Personnel

- Individual whose absence will dramatically alter performance of organization
  - Only person who can do a task
  - Only person with certain organizationally critical knowledge
  - Person who spreads good morale
  - Person who is only one with critical knowledge
  - Only person who knows key people
- Examples
  - Lead scientist
  - Long term administrative assistant
  - One person lab/technician

CASOS

 8

CASOS

Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

Critical Personnel

- Direct identifiers
  - The key entities: e.g., degree centrality
  - The exclusivities: e.g., task
  - The integrators: e.g., simmelian ties
  - The loads: e.g., workload and cognitive demand
- Indirect
  - Those who have access to, can influence, those who are critical

 9

Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

Node Level SNA Metrics Common Used: Centralities

- Degree Centrality
  - Node with the most connections
- Betweenness Centrality
  - Node in the most best paths
- Eigenvector Centrality
  - Node connected well-known people
- Closeness Centrality
  - Node that is closest to all other nodes

 10

Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

Critical Actors
What Nodes Matter?

Degree & Eigenvector	Betweenness	Closeness
0.417 Mohamed Atta	0.334 Nawaf Alhazm	0.571 Mohamed Atta
In the know power	Connects groups	Rapid access

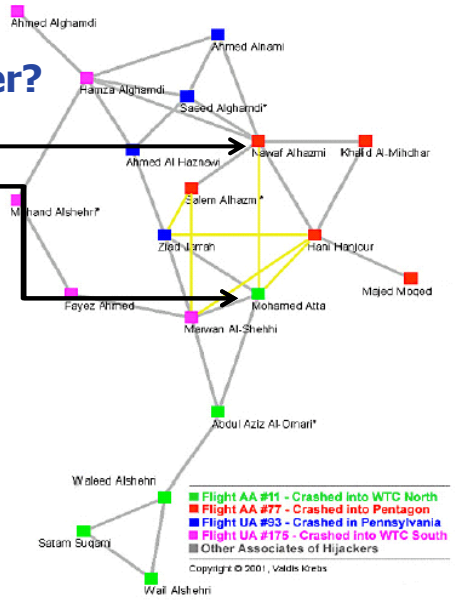


Figure 3 Trusted Prior Contacts + Meeting Ties [shortcuts]

CASOS

 11

Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

Simple SNA Measures

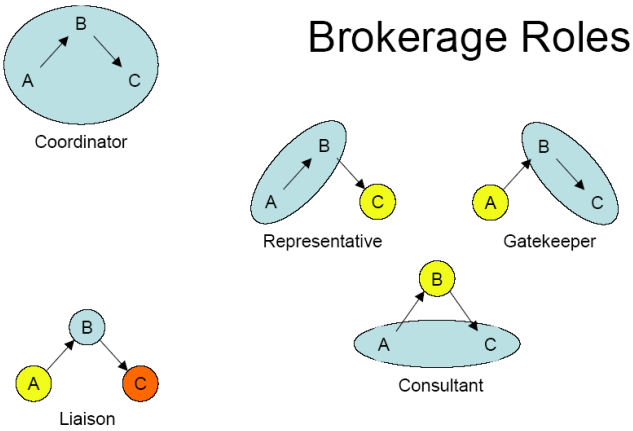
Measure	Definition	Meaning	Usage
Degree Centrality	Node with the most connections	In the know	Identifying sources for intel; Reducing information flow
Betweenness	Node in the most best paths Needs symmetric data	Connects groups	Typically has political influence, but may be too constrained to act
Eigenvector centrality	Node most connected to other highly connected nodes	Strong social capital	Identifying those who can mobilize others
Closeness	Node that is closest to all other nodes	Rapid access to all information	Identifying sources to acquire/transmit information

CASOS

12

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

Brokerage



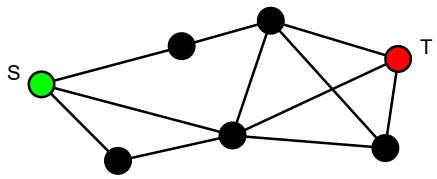
CASOS

13

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

Connectivity

- Line connectivity λ(s,t) is the minimum number of lines that must be removed to disconnect s from t
- Node connectivity κ(s,t) is minimum number of nodes that must be removed to disconnect s from t



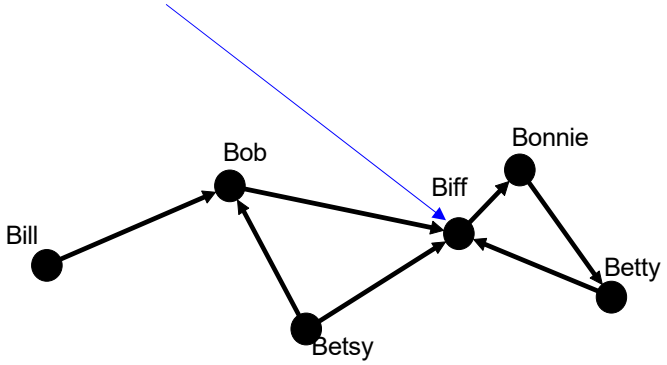
CASOS

14

Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

Cutpoints

- Nodes which, if deleted, would disconnect net



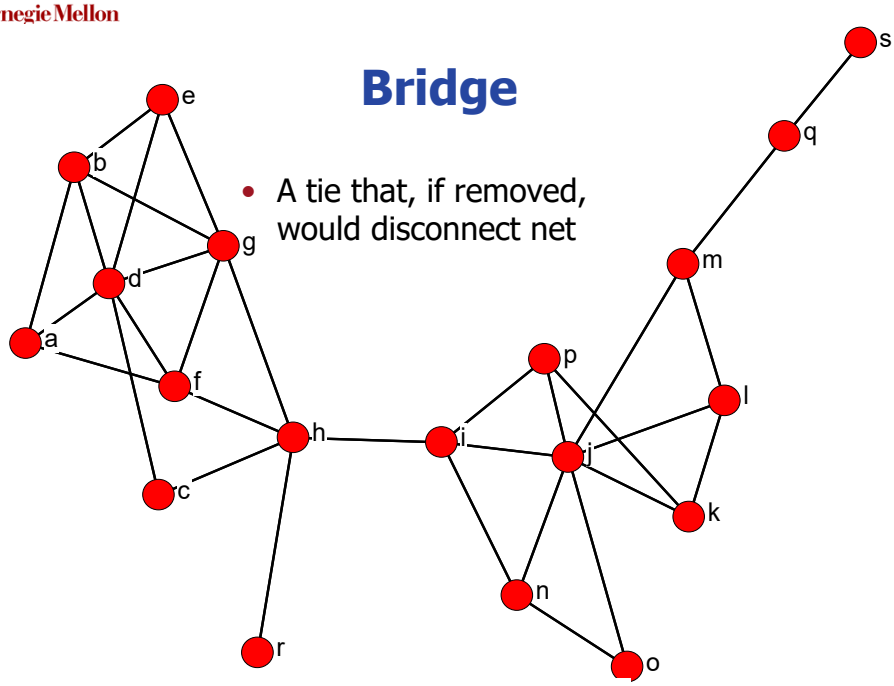
CASOS

15

Carnegie Mellon
ISR

Bridge

- A tie that, if removed, would disconnect net

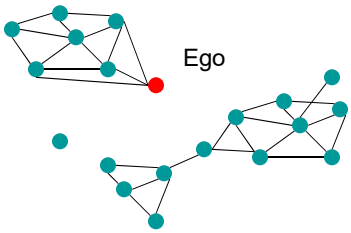


16

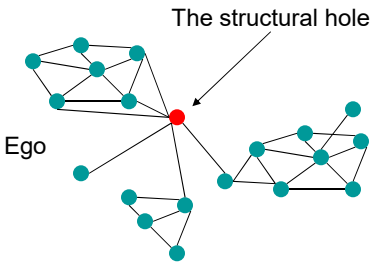
Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

Structural Holes

Local Betweenness



Few structural holes

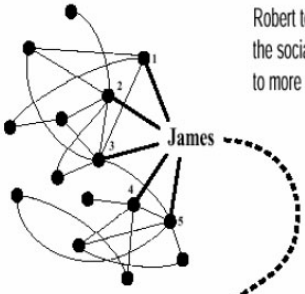


Many structural holes

CASOS

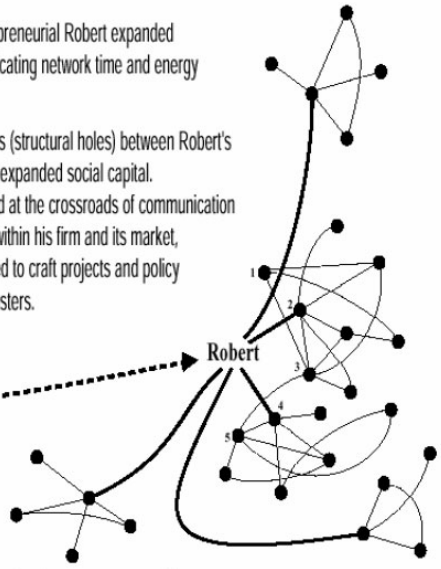
 17

Structural Holes



Robert took over James' job. Entrepreneurial Robert expanded the social capital of the job by reallocating network time and energy to more diverse contacts.

It is the weak connections (structural holes) between Robert's contacts that provide his expanded social capital. Robert is more positioned at the crossroads of communication between social clusters within his firm and its market, and so is better positioned to craft projects and policy that add value across clusters.



Research shows that people like Robert, better positioned for entrepreneurial opportunity, are the key to integrating across functions and across the people of increasingly diverse backgrounds in today's flatter organizations. In research comparisons between managers like James and Robert, it is the people like Robert who get promoted faster, earn higher compensation, receive better performance evaluations, and perform more successfully on teams.

Slide from Ron Burt

Carnegie Mellon ISR institute for SOFTWARE RESEARCH

Alternative Measures

- Page rank
  - Iteratively Weighted degree centrality
  - Page rank of node i is PR(pi)
  - D is the dampening factor
  - L(pj) is the outdegree of node j

$$PR(p_i) = \frac{1-d}{N} + d \sum_{p_j \in M(p_i)} \frac{PR(p_j)}{L(p_j)}$$

- Authority Centrality –
  - A node is authority-central to the extent that its in-links are from nodes that have many out-links.
  - The authority of a node are calculated by: $a(i) = \sum_{(j,i)\in E} h(j)$,
  - Where a(x) is the authority of node x ∈ {i, j}

Carnegie Mellon ISR institute for SOFTWARE RESEARCH

Clustering Coefficient

(Carley et al., 2013: 845-847)

NODE LEVEL –
Measures the degree of clustering in a network by averaging the clustering coefficient of each node, which is defined as the density of the node's ego network.

Let **G=(V, E)** be the graph representation of a square network.
Define for each node **v∈V** its Clustering Coefficient $\mathbf{cc_v}$:
let $\mathbf{G_v}$ **= ego network of entity v**

Then **Clustering Coefficient for entity: entity v =** $\mathbf{CC_v}$ **= density (**$\mathbf{G_v}$**)**

GRAPH LEVEL -

Then **Clustering Coefficient for graph: graph =** $\frac{\sum_{v\in V} CC_v}{|V|}$

Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

Bonacich

- Bonacich's power measure corresponds to the notion that the power of a vertex is recursively defined by the sum of the power of its alters.
  - The nature of the recursion involved is then controlled by the power exponent: positive values imply that vertices become more powerful as their alters become more powerful (as occurs in cooperative relations), while negative values imply that vertices become more powerful only as their alters become *weaker* (as occurs in competitive or antagonistic relations).
  - The magnitude of the exponent indicates the tendency of the effect to decay across long walks; higher magnitudes imply slower decay.
  - One interesting feature of this measure is its relative instability to changes in exponent magnitude (particularly in the negative case). If your theory motivates use of this measure, you should be very careful to choose a decay parameter on a non-ad hoc basis.
- Bonacich, P. (1987). "Power and Centrality: A Family of Measures.'" *American Journal of Sociology*, 92, 1170-1182.

CASOS

 21

Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

Bonacich's power centrality measure

$$c_i(\alpha,\beta) = \alpha(I - \beta A)^{-1} A\vec{1}$$

- *beta* is an attenuation parameter (set here by exponent)
- *A* is the graph adjacency matrix.
- *alpha* acts as a scaling parameter, and is set here (following Bonacich (1987)) such that the sum of squared scores is equal to the number of vertices.
  - This allows 1 to be used as a reference value for the ``middle'' of the centrality range.
- When *beta->1/lambda_A1* (the reciprocal of the largest eigenvalue of *A*), this is to within a constant multiple of the familiar eigenvector centrality score; for other values of *β*, the behavior of the measure is quite different.
- *β* gives positive and negative weight to even and odd walks, respectively, as can be seen from the series expansion *C_BP(alpha,beta) = alpha sum(beta^k A^(k+1) 1, k in 0..infinity)* which converges so long as */beta/<1/lambda_A1*.
- The magnitude of *beta* controls the influence of distant actors on ego's centrality score, with larger magnitudes indicating slower rates of decay.
- High rates imply a greater sensitivity to edge effects.

CASOS

 22

CASOS

Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

ORA Walkthrough 1

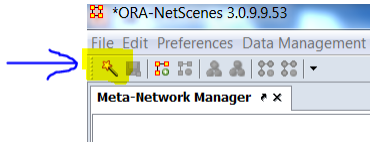
- We'll use Sampson data to get hands on training with Centrality Measures
- Sampson Data: http://bit.ly/1NSa067 (Also on Thumb Drive)

 23

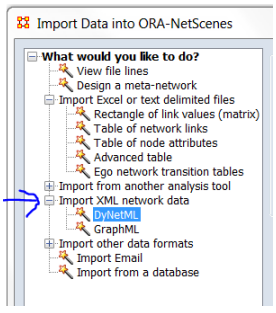
Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

ORA Walkthrough 1

1. Click Import Wizard Button



2. Select Import XML Network Data -> DynetML

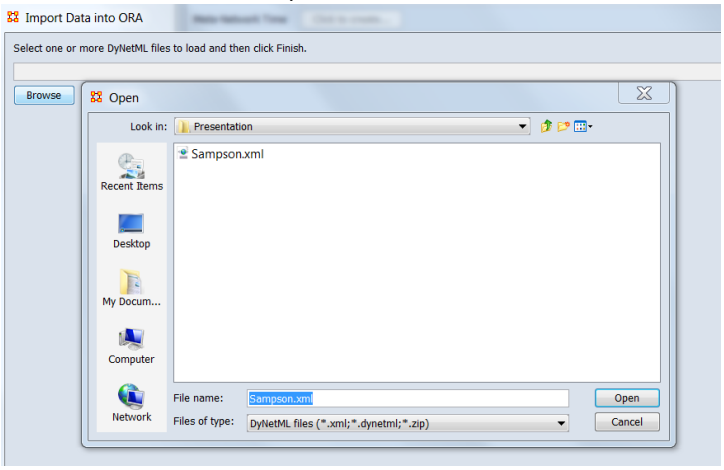


 24

Carnegie Mellon ISR institute for SOFTWARE RESEARCH

ORA Walkthrough 1

3. Click Browse, and select the sampson data from thumb drive

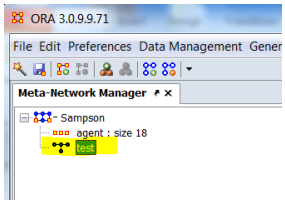


 25

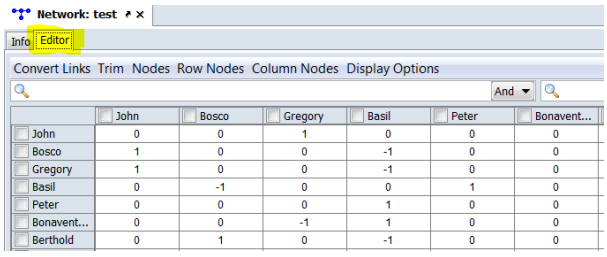
Carnegie Mellon ISR institute for SOFTWARE RESEARCH

ORA Walkthrough 1

4. Click on test network



5. Click on Editor, to manipulate the data



	John	Bosco	Gregory	Basil	Peter	Bonavent...
John	0	0	1	0	0	0
Bosco	1	0	0	-1	0	0
Gregory	1	0	0	-1	0	0
Basil	0	-1	0	0	1	0
Peter	0	0	0	1	0	0
Bonavent...	0	0	-1	1	0	0
Berthold	0	1	0	-1	0	0

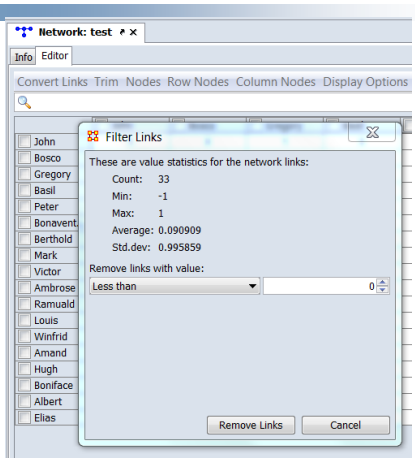
 26

Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

ORA Walkthrough 1

6. Click Trim -> Remove links by value
7. Select Less than, 0
8. Click Remove Links

Now we only have links where one monk liked the other



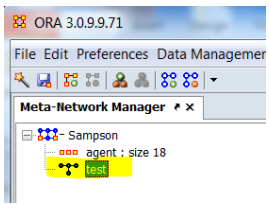
CASOS

 27

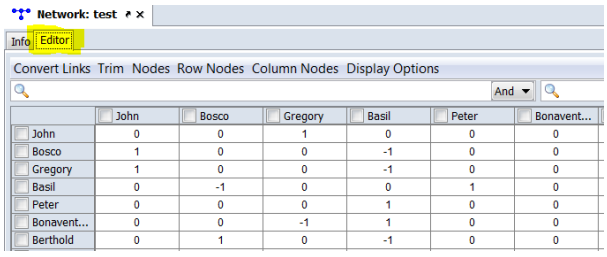
Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

ORA Walkthrough 1

9. Click on test network



10. Click on Editor, to manipulate the data



CASOS

 28

CASOS

Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

ORA Walkthrough 1

11. Click Convert Links -> Binarize link values

Now we're sure that all links have value 1, completing our data cleaning

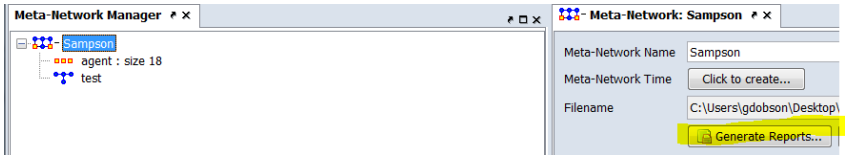
CASOS

 29

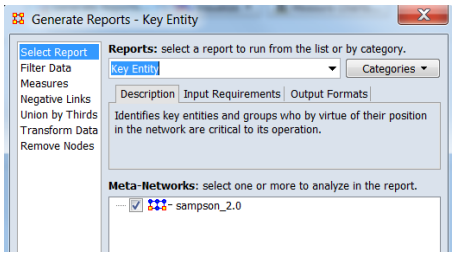
Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

ORA Walkthrough 1

12. Select Sampson Meta Network, Click Generate Reports



13. Select Key Entity Report



CASOS

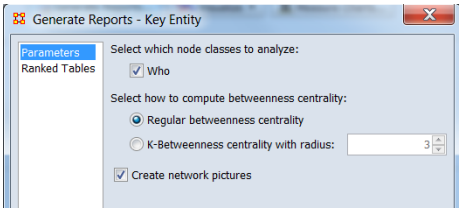
 30

CASOS

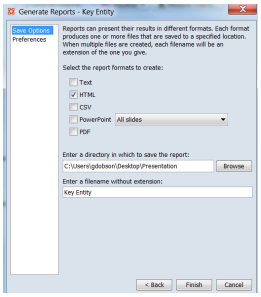
Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

ORA Walkthrough 1

14. Leave the Defaults



15. Leave the Defaults, click Finish



CASOS

 31

Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

ORA Walkthrough 1

16. Click on Key Agent - Who Analysis

KEY ENTITY REPORT

Input data: Sampson

Start time: Mon Jun 4 21:42:26 2018

Data Description

Table of Contents

Key Agent - Who Analysis

Performance Indicators - measures performance of the organization(s) as a whole

Produced by ORA, a joint product of the CASOS center at Carnegie Mellon University and Netanomics

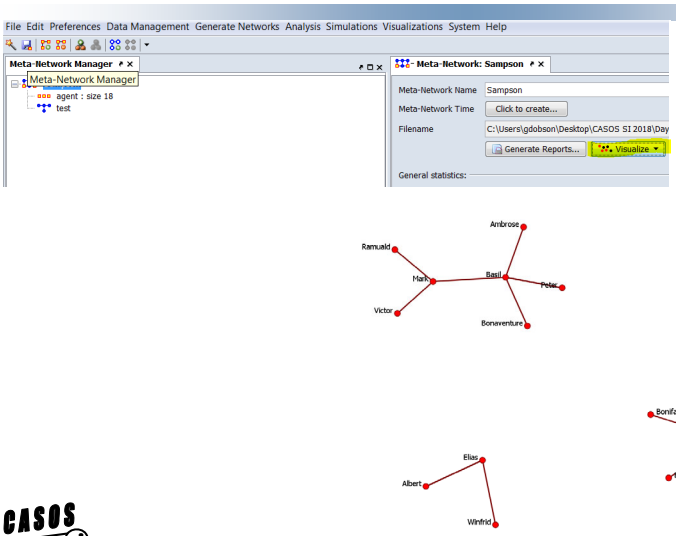
CASOS

 32

CASOS

Carnegie Mellon
institute for SOFTWARE RESEARCH

ORA Walkthrough 1



 33

Carnegie Mellon
institute for SOFTWARE RESEARCH

ORA Walkthrough 1

Eigenvector Centrality

Leaders of strong cliques are individuals or organizations who are collected to others that are themselves highly connected to each other. In other words, if you have a clique then the individual most connected to others in the clique and other cliques, is the leader of the clique. Individuals or organizations who are connected to many otherwise isolated individuals or organizations will have a much lower score in this measure then those that are connected to groups that have many connections themselves. The scientific name of this measure is eigenvector centrality and it is calculated on agent by agent or organization by organization 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).

Input network: test (size: 18, density: 0.0588235)

Show 10 entries Search:

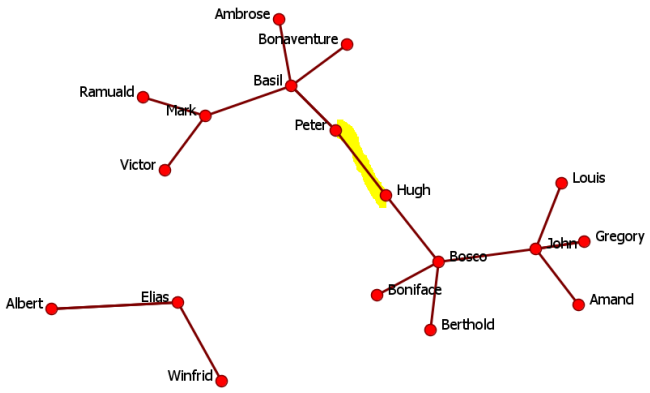
Rank	Agent	Value	Unscaled	Context*
1	Bosco	3.551874e-01	2.511554e-01	3.749497e-01
2	John	3.551874e-01	2.511554e-01	3.749497e-01
3	Basil	3.453620e-01	2.442078e-01	3.457115e-01
4	Mark	2.749860e-01	1.944444e-01	1.362862e-01
5	Elias	1.666667e-01	1.178511e-01	-1.860506e-01
6	Ambrose	1.587632e-01	1.122626e-01	-2.095697e-01
7	Bonaventure	1.587632e-01	1.122626e-01	-2.095697e-01
8	Peter	1.587632e-01	1.122626e-01	-2.095697e-01
9	Amand	1.542431e-01	1.090664e-01	-2.230206e-01
10	Berthold	1.542431e-01	1.090664e-01	-2.230206e-01

 34

Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

ORA Walkthrough 1

16. Open Visualizer, add link between Peter and Hugh



Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

ORA Walkthrough 1

17. Re-Run Generate Reports -> Key Entity -> Save as “likes add peter hugh” - observe “Eigenvector Centrality” rankings

Show 10 entries Search:

Rank	Agent	Value	Unscaled	Context*
1	Bosco	3.551874e-01	2.511554e-01	3.749497e-01
2	John	3.551874e-01	2.511554e-01	3.749497e-01
3	Basil	3.453620e-01	2.442078e-01	3.457115e-01
4	Mark	2.749860e-01	1.944444e-01	1.362862e-01
5	Elias	1.666667e-01	1.178511e-01	-1.860506e-01
6	Ambrose	1.587632e-01	1.122626e-01	-2.095697e-01
7	Bonaventure	1.587632e-01	1.122626e-01	-2.095697e-01
8	Peter	1.587632e-01	1.122626e-01	-2.095697e-01
9	Amand	1.542431e-01	1.090664e-01	-2.230206e-01
10	Berthold	1.542431e-01	1.090664e-01	-2.230206e-01

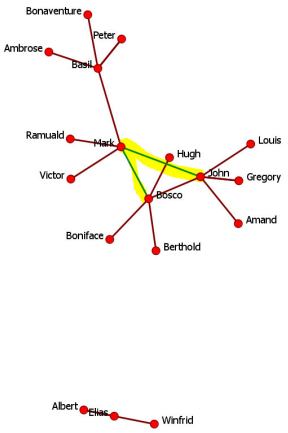
Show 10 entries Search:

Rank	Agent	Value	Unscaled	Context*
1	Bosco	5.963698e-01	4.216971e-01	1.063982e+00
2	John	5.355839e-01	3.787150e-01	8.831583e-01
3	Hugh	3.794598e-01	2.683186e-01	4.187265e-01
4	Basil	3.458540e-01	2.445557e-01	3.187574e-01
5	Peter	3.052566e-01	2.158490e-01	1.979898e-01
6	Berthold	2.509890e-01	1.774760e-01	3.655693e-02
7	Boniface	2.509890e-01	1.774760e-01	3.655693e-02
8	Amand	2.254066e-01	1.593865e-01	-3.954473e-02
9	Gregory	2.254066e-01	1.593865e-01	-3.954473e-02
10	Louis	2.254066e-01	1.593865e-01	-3.954473e-02

Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

ORA Walkthrough 1

18. Open Visualizer, add link from Mark to Bosco and John to Mark



Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

ORA Walkthrough 1

20. Re-Run Generate Reports -> Key Entity -> Save as “likes add mark” - observe “Eigenvector Centrality” rankings

Show 10 entries Search:

Rank	Agent	Value	Unscaled	Context*
1	Bosco	5.963698e-01	4.216971e-01	1.035323e+00
2	John	5.355839e-01	3.787150e-01	8.545626e-01
3	Hugh	3.794598e-01	2.683186e-01	3.902934e-01
4	Basil	3.458540e-01	2.445557e-01	2.903593e-01
5	Peter	3.052566e-01	2.158490e-01	1.696340e 01
6	Berthold	2.509890e-01	1.774760e-01	8.257561e-03
7	Boniface	2.509890e-01	1.774760e-01	8.257561e-03
8	Amand	2.254066e-01	1.593865e-01	-6.781746e-02
9	Gregory	2.254066e-01	1.593865e-01	-6.781746e-02
10	Louis	2.254066e-01	1.593865e-01	-6.781746e-02

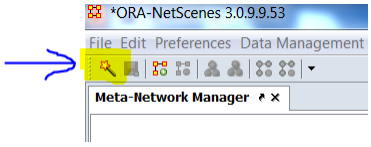
Show 10 entries Search:

Rank	Agent	Value	Unscaled	Context*
1	Mark	5.519503e-01	3.902878e-01	9.033718e-01
2	Bosco	5.318373e-01	3.760658e-01	8.427256e-01
3	John	5.140632e-01	3.634976e-01	7.891317e-01
4	Basil	2.968666e-01	2.099164e-01	1.342238e-01
5	Hugh	2.278233e-01	1.610954e-01	-7.396071e-02
6	Ramuald	1.791029e-01	1.266449e-01	-2.208662e-01
7	Victor	1.791029e-01	1.266449e-01	-2.208662e-01
8	Berthold	1.725764e-01	1.220299e-01	-2.405453e-01
9	Boniface	1.725764e-01	1.220299e-01	-2.405453e-01
10	Peter	1.702571e-01	1.203900e-01	-2.475386e-01

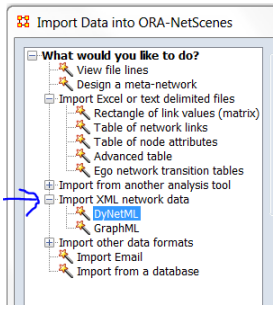
Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

ORA Walkthrough 2

1. Click Import Wizard Button



2. Select Import XML Network Data -> DynetML



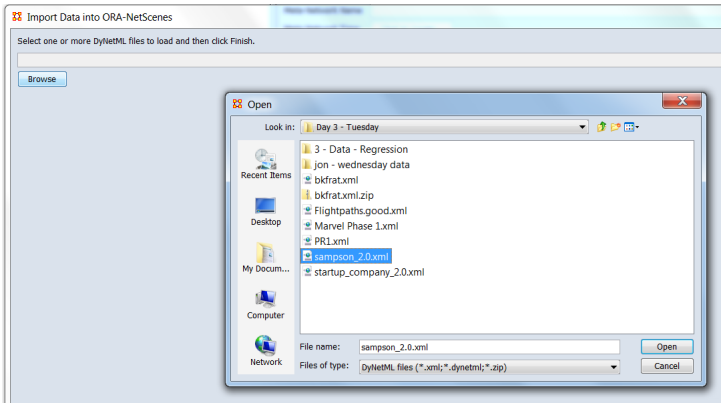
CASOS

 39

Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

ORA Walkthrough 2

3. Click Browse, and select the sampson data from thumb drive



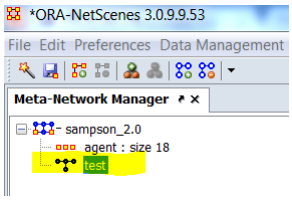
CASOS

 40

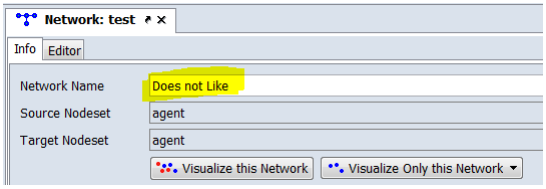
Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

ORA Walkthrough 2

4. Click on test network



5. Change the name to “Does not Like”

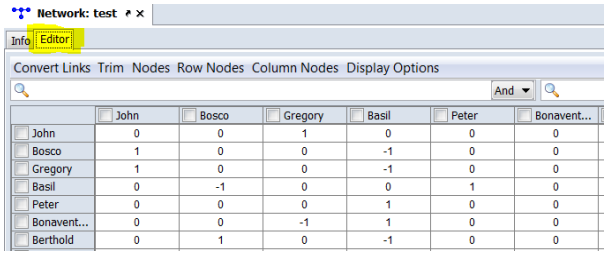


 41

Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

ORA Walkthrough 2

6. Click on Editor, to manipulate the data



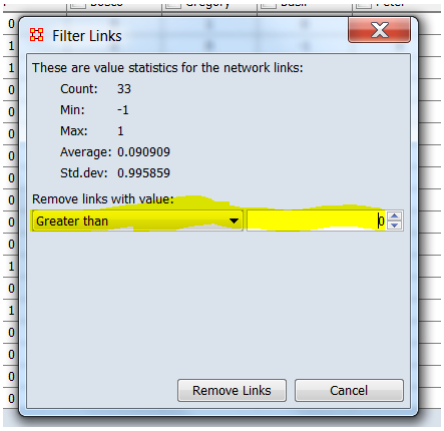
 42

Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

ORA Walkthrough 2

7. Click Trim -> Remove links by value
8. Select Greater than, 0
9. Click Remove Links

Now we only have links where one monk didn't like the other



CASOS

 43

Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

ORA Walkthrough 2

10. Click Convert Links -> Binarize link values

Now we're sure that all links have value 1, completing our data cleaning

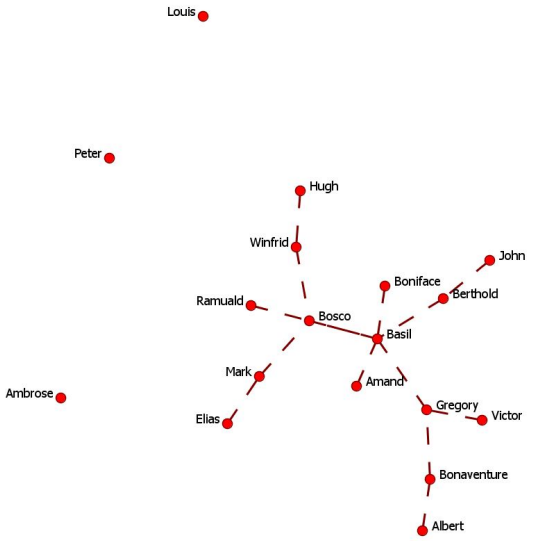
CASOS

 44

CASOS

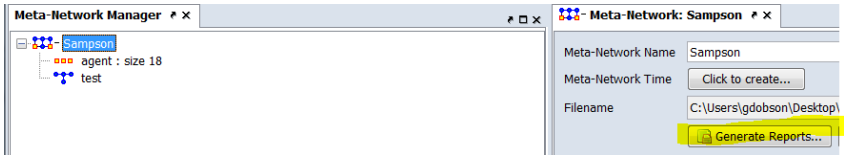
ORA Walkthrough 2

11. Visualize

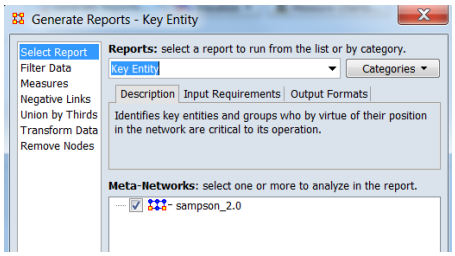


ORA Walkthrough 2

12. Select Sampson Meta Network, Click Generate Reports



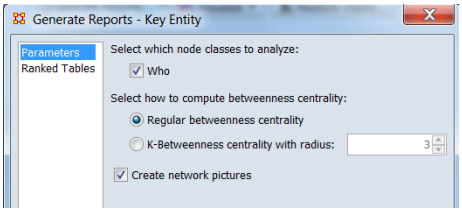
13. Select Key Entity Report



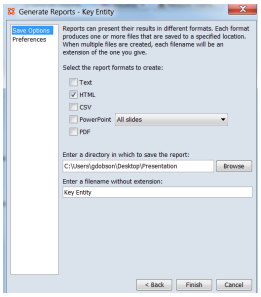
Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

ORA Walkthrough 2

14. Leave the Defaults



15. Leave the Defaults, click Finish



CASOS

 47

Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

ORA Walkthrough 2

16. Click on Key Agent - Who Analysis

KEY ENTITY REPORT

Input data: Sampson

Start time: Mon Jun 4 21:42:26 2018

Data Description

Table of Contents

Key Agent - Who Analysis

Performance Indicators - measures performance of the organization(s) as a whole

Produced by ORA, a joint product of the CASOS center at Carnegie Mellon University and Netanomics

CASOS

 48

CASOS

Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

ORA Walkthrough 2

Total Degree Centrality

Individuals or organizations who are 'in the know' are those w
the know' are identified by degree centrality in the relevant so
is total degree centrality and it is calculated on the agent by ag

If the node of interest has a higher than normal value (greater
Finally, the row is colored blue if the node has a lower than nc

Input network: Does not Like (size: 18, density: 0.0490196)

Show 10 entries

Rank	Agent
1	Basil
2	Bosco
3	Gregory
4	Berthold
5	Bonaventure
6	Mark
7	Winfrid
8	Albert
9	Amand
10	Boniface

Basil and Bosco most disliked…John not disliked!

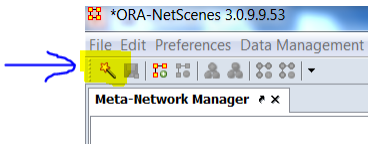
CASOS

 49

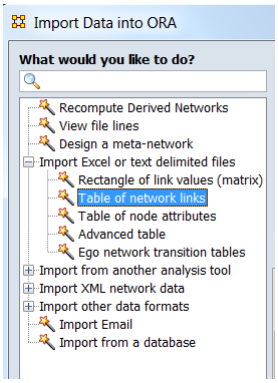
Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

ORA Walkthrough 3

1. Click Import Wizard Button



2. Click Table of Network Links



CASOS

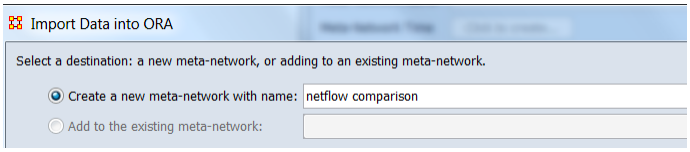
 50

CASOS

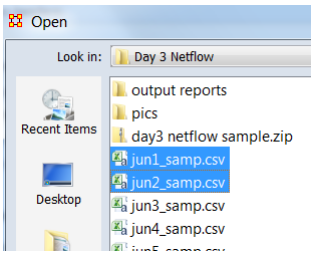
Carnegie Mellon
institute for SOFTWARE RESEARCH

ORA Walkthrough 3

3. Create new meta-network, named “Netflow comparison”



4. Browse, select BOTH of jun1_samp.csv and jun2_samp.csv



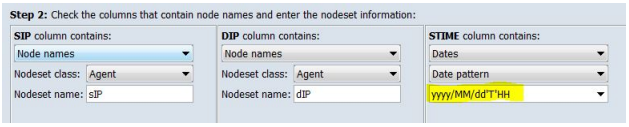
CASOS

 51

Carnegie Mellon
institute for SOFTWARE RESEARCH

ORA Walkthrough 3

5. Set the Step 2 options like so:



6. Set the Step 3 options like so, and then “Next”:

CASOS

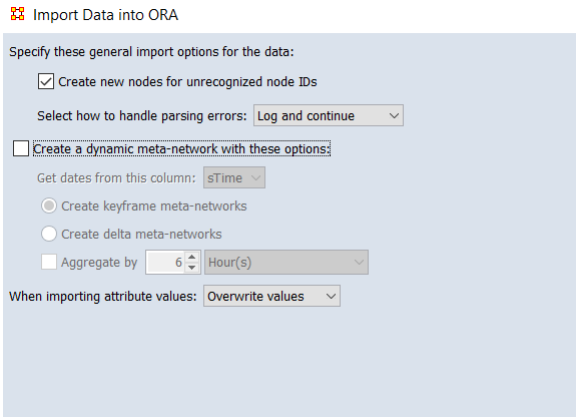
 52

CASOS

Carnegie Mellon
institute for SOFTWARE RESEARCH

ORA Walkthrough 3

7. Uncheck “Create a dynamic meta-network” and click “Finish”

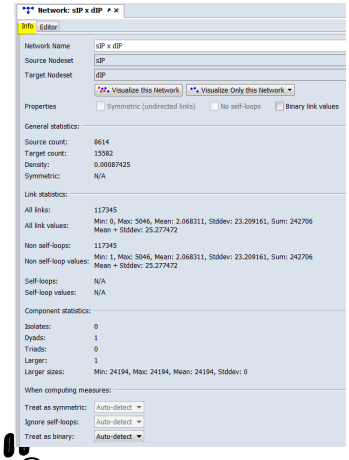


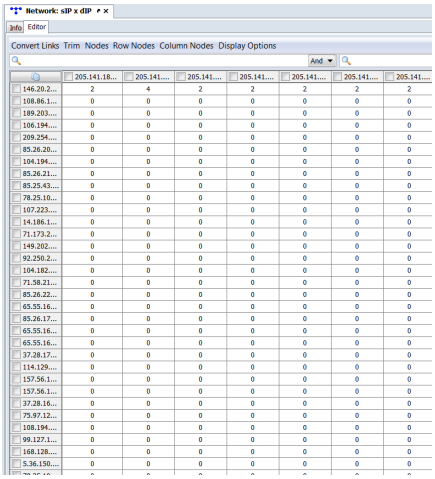
 53

Carnegie Mellon
institute for SOFTWARE RESEARCH

ORA Walkthrough 3

8. Inspect data in the Info and Editor:



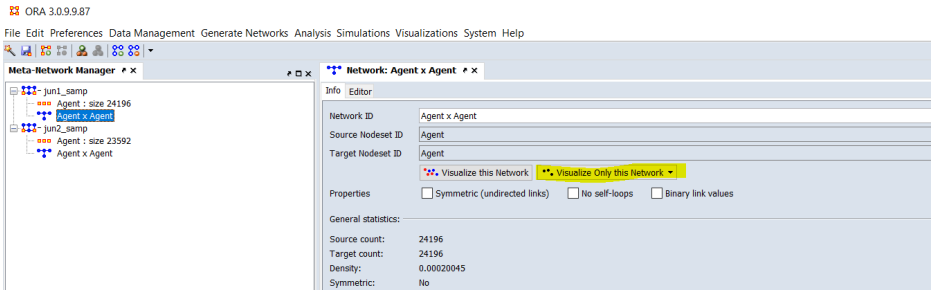


 54

Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

ORA Walkthrough 3

9. Visualize the network (one of them), select “Load Normally”

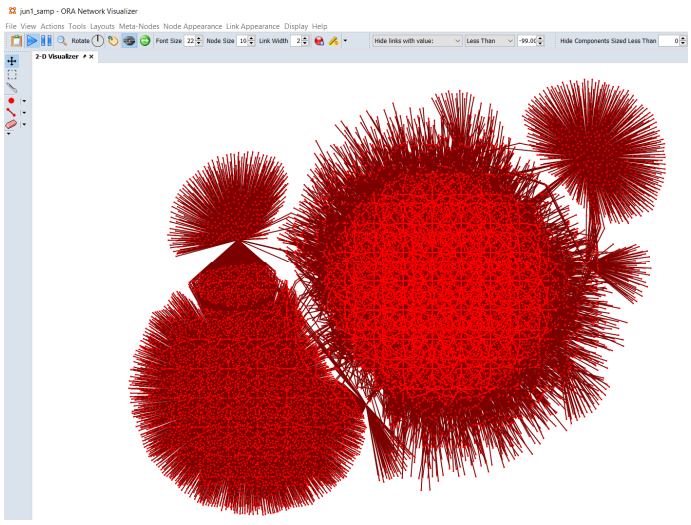


CASOS

 55

Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

ORA Walkthrough 3



CASOS

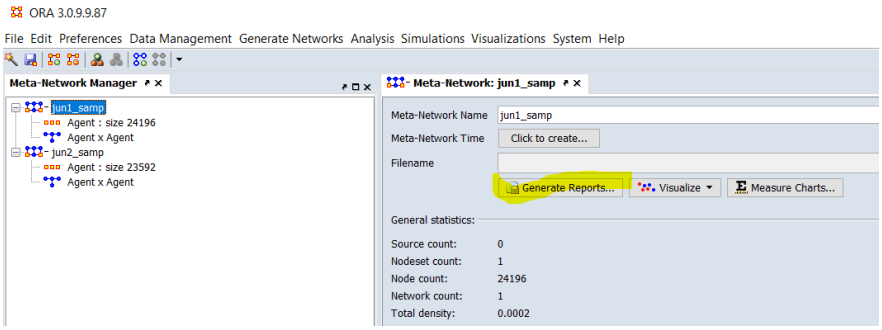
 56

CASOS

Carnegie Mellon
institute for SOFTWARE RESEARCH

ORA Walkthrough 3

10. Select jun1_samp network and Click “Generate Reports”

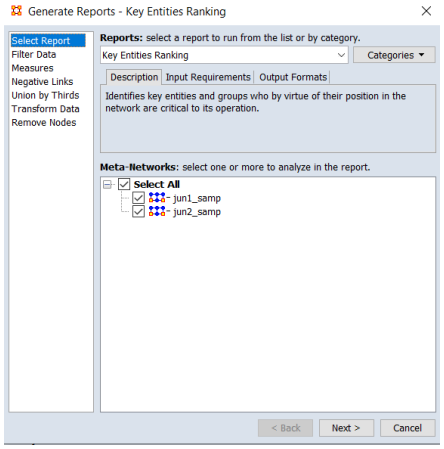


 57

Carnegie Mellon
institute for SOFTWARE RESEARCH

ORA Walkthrough 3

11. Select “Key Entities Rankings” and select both jun1_samp and jun2_samp and click “Next”

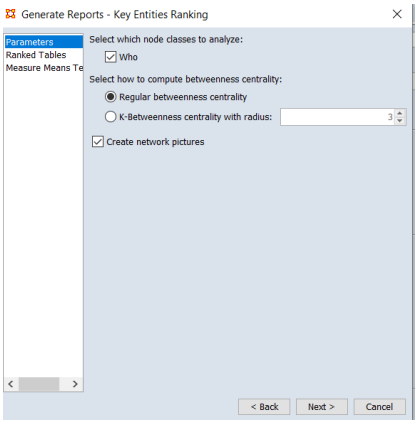


 58

Carnegie Mellon
institute for SOFTWARE RESEARCH

ORA Walkthrough 3

12. Leave defaults and click “Next”

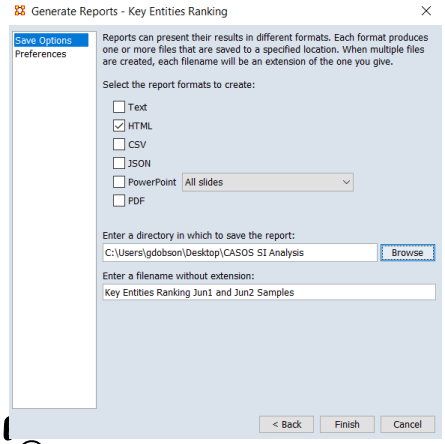


 59

Carnegie Mellon
institute for SOFTWARE RESEARCH

ORA Walkthrough 3

13. Select a folder and name of report, click Finish and browse report



 60