Basic Lessons in ORA and AutoMap 2011

Kathleen M. Carley and Dave Columbus,

June 13, 2011
CMU-ISR-11-109R

Institute of Software Research
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Center for the Computational Analysis of Social and Organization Systems
CASOS technical report

This report/document supersedes CMU-ISR-09-117
"ORA and AutoMap Lessons", June 2009

i

Abstract

ORA is a network analysis tool that detects risks or vulnerabilities of an organization's design structure. The design structure of an organization is the relationship among its personnel, knowledge, resources, and tasks entities. These entities and relationships are represented by the Meta-Matrix. Measures that take as input a Meta-Matrix are used to analyze the structural properties of an organization for potential risk. ORA contains over 100 measures which are categorized by which type of risk they detect. Measures are also organized by input requirements and by output. ORA generates formatted reports viewable on screen or in log files, and reads and writes networks in multiple data formats to be interoperable with existing network analysis packages. In addition, it has tools for graphically visualizing Meta-Matrix data and for optimizing a network's design structure. ORA uses a Java interface for ease of use, and a C++ computational backend. The current version ORA1.2 software is available on the CASOS website: http://www.casos.ece.cmu.edu/projects/ORA/index.html.

Table of Contents

ORA Lessons

Below are links to lessons on how to perform many common dynamic network analysis tasks as well as using ORA in general:

An Overview of the ORA GUI panels

Meta-Networks: From the beginning

Running a Key Entity Report

Using The Over-Time Viewer

Working with The Network Converter

Change Detection

Working With Attributes

ORA Overview
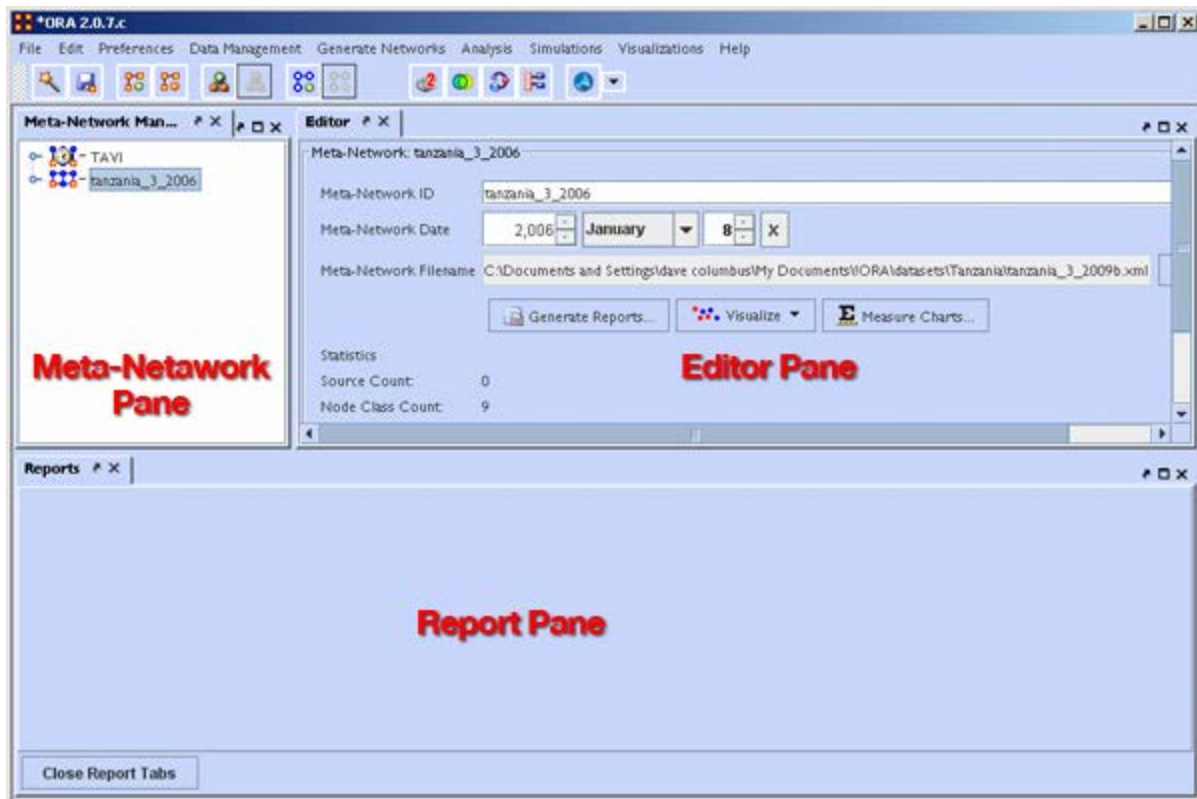
The **Main Menu** runs across the top of the window, and the rest of the screen is divided into three panels.

Below the menu is the **Toolbar**. ORA starts with buttons to add and remove Meta-Networks, NodeClasses and Networks. Using the inverted triangle you can add or delete buttons from this strip.

The Meta-Network Pane displays a list of the Meta-Networks that has been loaded into ORA, its components and sub-networks. To view information about a particular meta-network or sub-network, single-click on it.

The Editor Pane displays information about the meta-network or sub-network highlighted in Panel 1.

The Report Pane displays information about the results of analyses run on your meta-network.

Loading a meta-network into ORA

You have three options for opening your meta-network in ORA:

- From the main menu select **File ⇒ Open Meta-Network**. A small window titled **Help: Opening meta-network** pops up. Your data should already be in the correct format, so click **[OK]**.

- Next, a file chooser titled **Open** pops up. Browse to the location of your dataset, **single-click on it**, and **click the Open button** in the bottom right corner.

- **Click** the first icon in the row directly below the menu bar. It looks like a half-open file folder. A window titled **Import Data into ORA** will pop up. **Click** the Select Files button. A file chooser titled **Open** pops up. Browse to the location of your dataset, single-click on it, and click the Open button in the bottom right corner. You will be returned to the **Import Data into ORA** window. Click the Finish button in the bottom right corner.

- **In panel 2**, click the **Load button** on the far right. A window titled **Import Data into ORA** pops up. Click the Select Files button. A file chooser titled **Open** pops up. Browse to the location of your dataset, single-click on it, and click the Open button in the bottom right corner. You will be returned to the **Import Data into ORA** window. Click the Finish button in the bottom right corner.

A list of your meta-network and sub-categories in Panel 1, as well as the location and basic statistics of your meta-network in Panel 2 will appear. ORA is ready to begin analyzing the meta-network.

The Visualizer

To visualize a meta-network, first make sure the meta-network you wish to visualize is highlighted in Panel 1. If it is not highlighted, single-click on it.

In Panel 2, click the Visualize this Meta-Network button. A new window will pop up displaying your meta-network graphically. A small legend also appears.

Below is a screen capture showing the visualization of the **agent x event** graph from the Stargate Summit Meta-Network.
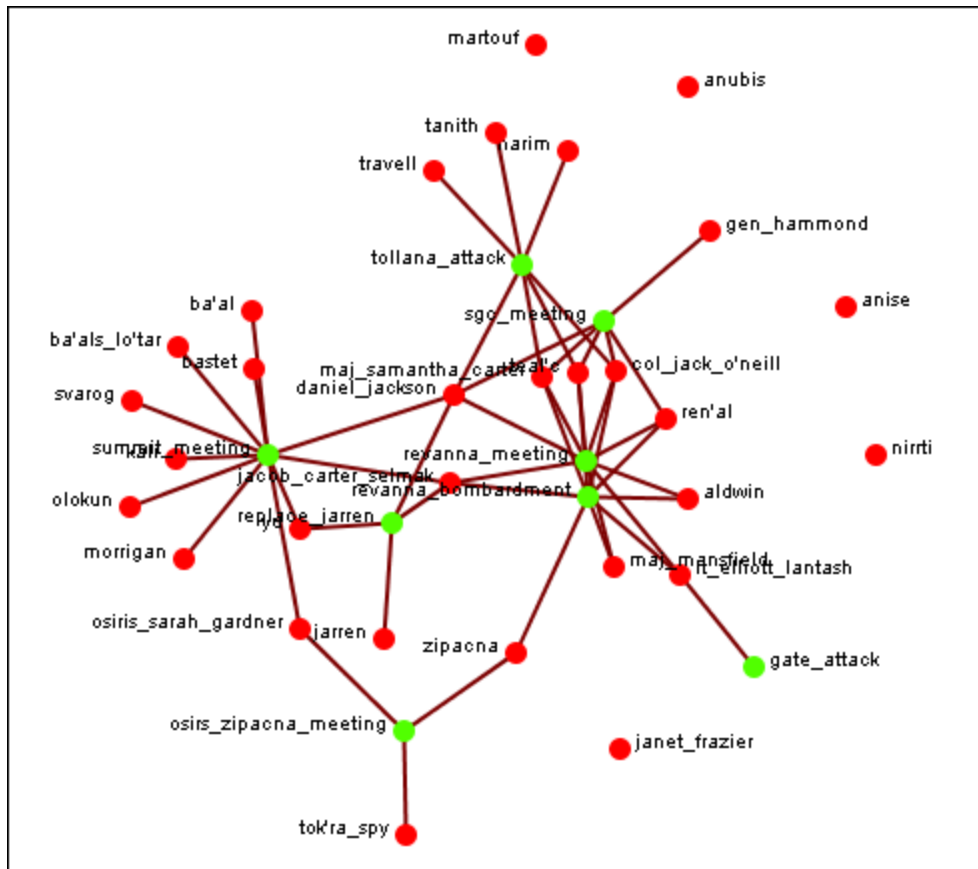
The visualization displays the connections between all items in the **Stargate Summit Meta-Network**. The red circles represent the agents, the green circles represent individual events in the scenario, and the lines represent the connections between the agents and the events.

You'll notice there are agents with no connections to any events. They represent nodes that were connected to the scenario but not to the particular events of the Summit timeline.

Creating a New Meta-Network

These lessons will give you the information in order to create a **Meta-Network** from data you have on-hand.

Lesson - 101

This section describes the process of looking over your data and deciding what is relevant for inclusion into a Meta-Network. It takes the examples I used for the Stargate dataset.

A step by step run through of creating the Meta-Network from working with Excel, exporting data to csv files and finally building a complete Meta-Network.

lessons - 301+

Lessons dealing with importing data from other programs.

Examine Your Data

General Thoughts

When you first open up ORA you need to construct a Meta-Network. But the question is, how do you go about this task?

First, you need to understand what constitutes a Meta-network. The basic building blocks of all Meta-Networks is a **Node***. Similar nodes are combined to build **Node Sets***. Node Sets are combined to make **Networks***. Multiple networks are combined to create a **Meta-network.***.

It is up to you on **HOW** to construct the individual Node Sets and Networks and how to combine the Networks into a Meta-network. This means you need to gather as much information regarding the scenario as possible **(who, what, when, where, and how)**

Knowing as much about the scenario as possible is important. Information about the people involved, where these people were seen, what items they had available and used, what they knew, what tasks they performed, what resources are needed to perform those tasks, etc. The more relevant information you collect the better your final Meta-network. will be.

*For the purpose of creating the **Stargate Meta-network.** the two-episode story arc **(Summit / Last Stand)** was chosen as the basis for all the nodes, node sets, and networks. But even though the information was clearly on the tv screen, many intuitive decisions had to be made.*

*For the **Julius Caesar Meta-network.** both the play and spark notes were used as reference. As with the **Stargate Meta-network.** many decisions were made about how to create data sets and what connections would be required.*

I found that for the help file using a set of data where you **know all the answers** helps. But that won't always be the case. If you already know what the results are supposed to be then it's easy to build a correct Meta-Network.

But the most important part of this is it helps you to learn what to look for. You'll initially make mistakes but if you instantly jump into creation mode then, well, since you're not sure how to build this in the first place then how can you be sure your Meta-Networks are built correctly in the first place?

*Initially the **Stargate Meta-network**. was built with both the **allies and the bad Guys**. This appeared to be the best way and it looked good displayed in the visualizer. But it was discovered that having the **bad Guys** interfered with getting proper information from the reports. It wasn't until after the **bad guys** were removed that reports began **to make more sense**. What looks good in the beginning **ISN'T** necessarily your best option.*

What's in a Node Class

There are **six** major classes: **Agents, Location, Events, Tasks, Knowledge, and Resources**.. Nearly everything will fall into one of these six class. ORA contains the ability to create new Node Classes but in most cases that's entirely unnecessary because most of ORA measures are designed to work with most common six classes mentioned above.

*With **Stargate** the agents were easy. The Stargate personnel, the Tok'ra, and some assorted others. The locations also were easy using scenes form the show. It was the events, knowledge, resources, and tasks which became slightly more difficult to discern. There were multiple **false starts** (missing some events, unnecessary locations, misplaced knowledge/resources) before I came to a proper conclusion. Initially I created extra Node sets (like faction & groups). This did nothing but clutter up the Meta-network. and make finding information much harder.*

*The **Julius Caesar** Agent NodeSet was also easy using the **cast of characters** but it too had it's construction problems. Julius Caesar had limited resources so knowledge and resources were essentially combined into one NodeSet called **knowledge**.*

Node sets can be either **binary (0 or 1)** or **numbered (negative #, 0 or positive #)**. A Meta-network. can contain a combination of the two.

When building Node Classes and Networks (and then onward to Meta-networks) you'll discover that many networks are are very apparent while others were almost invisible. You'll have some large networks while others may be extremely small. But all networks are necessary to produce the complete Meta-network.

And the depth of the Node Classes can be extended by affixing **attributes*** to them.

*Stargate uses attributes to give extra information such as an agent's **gender or time of death** or the **order in which events happened***

*Use the **Julius Caesar** attributes **allegiance and persuasion** as a control mechanism to show the strength which various agents use these attributes.*

Agents (Who)

It's usually best to start out by constructing the **Agent NodeSet**. This helps reveal many facets of the rest of the Meta-network. Gather all possible information about the **Whos** involved. Agents are the persons involved Stargates' **daniel_jackson & jacob_carter_selmac** and over at Julius Caesar **Brutus and 1st Commoner**.

*The **Stargate NodeSet** contains a mere 16 agents. Originally I plotted both the **allies and bad guys** but found there was too much clutter and unnecessary nodes. Removing the bad guys helped streamline the NodeSet.*

*The **Julius Caesar NodeSet** came in with more than twice as many. Unlike the **Stargate NodeSet** all of the characters were necessary. Everybody was plotting against everybody else.*

After you've collected all the agents involved you can proceed to creating your first network. Figure out which agents had any contact the situation and place a **"1"** in the corresponding cell. Depending on the particular circumstances you will have to establish criteria for an association. An **Agent x Agent** network creates a **Square Network** meaning the names in the rows are identical to the columns.

*The **Stargate (and Julius Caesar) Meta-networks** were fairly easy to construct agent Node sets and agent by agent Networks. Real life data doesn't always come with such an **easy to see** diagram. After a Meta-network. is constructed it very likely that new information will surface which will change the face of the current Meta-network. and, thusly, the results of any reports.*

There are three modes of thought in creating connections. **1)** The most common would be to say if one agent knows another agent then place a 1 in the cell. **2)** Next would be to use the preceding criteria but to weight the ties where one person could have a **higher regard** for a person more than there are reciprocal feelings. **3)** The third would be to use a hierarchical style showing who reports to who in a organizational tree situation. It's important that the person creating the Meta-network. makes a call whether there is and and how strong a tie will be.

>All three of these steps (unfortunately) were gone through in the creation of the **Stargate Meta-network.** Initially, using only binary data, the logic of the data set was very fuzzy as to the exact nature of many of the connections Many of the **allies** knew many of the **bad guys** on a first name basis. Next weighted ties were tried with higher numbers for stronger connections but the results that should have been - just weren't. It wasn't until a hierarchical style showing who reported to who was used did the results begin to ring true.

The **Julius Caesar Meta-network.** didn't have the same problems. It was apparent from the start that a **weighted person-by-person** network was the key.

As you learn more about Meta-networks it becomes easier to see which mode you need to use.

You might not create a Meta-network. correctly the first time **(or the second or the third time)** if you rush into it. Best to take a little time and review the information before heading straight into the creation process. Sometimes **too much information** will find it's way into your Meta-network.

*On the first run for the **Stargate Meta-network.** both the allies and the bad guys were placed in the network figuring both were needed for proper interaction. And it worked find until the reports had some **bad guys** popping up where they actually interfered with the results. So those nodes were removed and run the reports a second (or third time). This become a waste of time and effort. This solution was to **create the data set with only the allies**. This then gave more reliable information in the reports.*

A little time spent on reviewing the data and any reports will great improve you chances of getting it **somewhat** right in the first place. It will also help in your deciding the best way to build your Networks.

*Originally, when both allies and bad guys were used, **weighted ties** were used to describe the connections between various nodes. Each connection between the nodes on the same side were individually weighted depending on how they felt toward that other node and weights between opposing sides were always less than **"1"**.*

Originally the Network was done with binary ties but was changed to weighted ties.

Remember, you'll always find **something** that you think needs tweaked as more and more information is uncovered.

Locations (Where)

Next create a **NodeSet** of all the **locations** involved in the scenario. Locations are places like Stargate's **earth or revanna** or Julius Caesar's **battlefield or senate**.

*For the **Stargate Meta-network.** it was a simple matter of following the episode and marking down who was in what scene. Not quite a simple when you've got less than perfect data.*

*Ditto for **Julius Caesar** If a character was in a scene location then a **"1"** was placed in that character's matrix cell showing he was there.*

Create a list of all possible locations and use your best guess as to whether that location is important. Remember, just because people are seen in a location doesn't mean it's important to the Meta-network. You may have a location you think is important initially but come to find that nothing happened of interest there (or visa versa).

*For the **Stargate Meta-network.** it was difficult to focus in on the exact **locations** necessary. The first time only the major locations were used, but it was difficult defining specific tasks. Then all locations were added but it ended up have many isolates and pendants that became unnecessary.*

*For the **Julius Caesar Meta-network.** there was the scene on the battlefield and various sub-scenes in individual battle-tents. It was decided to combine all the battle-tent scenes together. Individual **tasks** would be used to sort out the Meta-network.*

Events (When)

Creating an **Events Node sets** is a little more difficult than for either agents or locations. Agents and locations are **tangible items** whereas events are **intangible**. They also depend on the interpretation of the individual creating the NodeSet. Events can include items like Stargate's **summit_meeting and gate_attack** or Julius Caesar's **Brutus speaks to citizens and Funeral of Julius Caesar**.

*Many of the events in **Summit scenario** were easy (the meetings, the summit, the attack). But others were a little harder to discern. Initially there were a mere **eight** events. But shortly after it was found to be inadequate. After a quick re-adjustment he number of events doubled.*

When plotting events it is sometimes necessary to show **what order** they occurred. What happened first, second, third, and so on. This helps discern which agents are needed in which parts of the scenario.

*In the **Stargate Meta-network.** the event sequence shows the beginning event was the **tollana_attack**. This led to the **sgc_meeting** and onto the **revanna_meeting**. There the sequence split into two branches eventually both converging in on **gate_attack**.*

To create an ordered **event x event** network start with the row of earliest event. Place a checkmark in the column for the next event. Find the row for the column you just check marked. Follow that to the column with the next event. Proceed in

this manner till you have all your events. You may, as happened in the Stargate Network, that two events sprout from a single event. They may, or may not. converge back to an ending.

Tasks (How)

Tasks are things that are done by the agents such as Stargate's **replace_jarren and poison_summit** and Julius Caesar's **Kill Caesar and Attend Senate**.

A task network can be set up to show the order that tasks need to be done in order for the scenario to be played out. For example, *procuring materials to build something* is always necessary before *building something* but not necessarily before *planning*. This is usually done with a directed graph **(i.e. a link from procuring to building but not the other way around)**.

*As in the example of **Events** above certain tasks need to be completed before others can be started. **Stargate** uses a **binary, directed** network showing a pathway of which tasks **are required** to be completed before the next task can be started. **create_poison**, **infiltrate_yuWorld, and create_memory_drug** all needed done before **present_plan_to_sgc**. But none of this could have taken place without **get_reole_chemical** leading to the **create_memory_drug**.*

Tasks are exclusively attached to agents. You can't have a **location, event, etc. performing a task**. But this is not to say that you can't have a task attached to a location as it might necessarily need to be done at a certain location.

***daniel_jackson** could only perform the task **poison_summit** at the location **hasaraSystem_spaceStation** as that's where the event **summit_meeting** took place. He could not have done **poison_summit** on **revanna** or **yu_homeworld**.*

Knowledge (What)

Many times more than one person knows the same thing. And other times there is a single person who has some vital knowledge. This is a way to help find out if one person is completely indispensable or whether there are other pathways to accomplishing the goal.

In the **Stargate Meta-network.** *there are some **knowledge** which many agents have. But there are also some **knowledge** that only one person has. And if that person is taken out of the picture then the plan fails to materialize.*

*Another situation is two people both having knowledge but it's much more important to one than the other. Both **daniel_jackson and jacob_carter_selmak** use spying but it's clear that **jacob_carter_selmak** is better at it than **daniel_jackson**.*

Resources (What)

Resources can be very similar to **knowledge** and many times these two overlap. It's not necessary to always have both.

*The **Stargate Meta-network.** is was very necessary to separate the two as sometimes it was necessary to have **knowledge regarding a resource** before you could actually use that resource. Before Daniel could use the **symbiote poison** or the **memory-altering drug** he was required to have the knowledge of how the devices operated.*

*In the **Julius Caesar Meta-network.** there was no real reason to separate **knowledge and resources**. The resource NodeSet was simply not used for this data set as it would have been redundant. There is no such constraint on **learning how a dagger worked** in Julius Caesar.*

Most of the **Measures** that can be run on **knowledge** have a corresponding measure that runs on **resources**. So some Meta-Network can eliminate one or the other.

Networks

After creating all your **Node sets** you then move on to creating **Networks**. Every Meta-network. will require a slightly different set of **Node sets and Networks**. Below are the **Stargate** and the **Julius Caesar** Node sets The Node sets in **bold/italic** are **not** contained in the other Meta-network.

*The **Stargate NodeSet** consisted of: agent x agent, agent x event, agent x knowledge, agent x location, **agent x resource**, agent, x tasks, event x event, **event x resource**, knowledge x tasks, tasks x event, tasks x tasks*

*The **Julius Caesar NodeSet** consisted of: agent x agent, agent x event, agent x knowledge, agent x location, agent x tasks, event x event, knowledge x tasks, **location x location**, task x event, task x task.*

Notice that the two are almost identical but **Stargate** contains **Networks** for the **Resource Node sets** The **Julius Caesar** used a **location x location** network.

The big difference was the construction of the **agent x agent** network. Julius Caesar was a weighted matrix with connections to everyone an agent knew. The Stargate **agent x agent** network used a hierarchical method showing who reported to who.

Another difference was the lack of a **resource nodeset** in Julius Caesar. There was no need as all that was placed under knowledge. Stargate, on the other-hand required separate knowledge and resources as it was sometimes necessary to have a certain knowledge before using a certain resource.

The two data sets are very similar, most of the major Node sets are contained in both. But the circumstances for each are unique which requires a slightly different set of Node sets

Excel and CSV

You learned from **Meta-Network 101** about how to survey your data, the different forms it can take, and how to make it a fit for ORA. In this lesson we'll go through the procedure of creating a Meta-Network using the Stargate data set in some practical examples.

Your first Node Set

Your first Node set will more than likely be an **Agent Node set** since all other node sets are either referred to/or by the agent node set Review you data and make a list of the people involved.

*After watching the **Summit/Last Stand** episodes (a couple times) and reading various reviews on the storyline I came up with 16 agent nodes. My first thought was to only use the people involved in the assassination plot **(The SGC and the Tok'ra)** but found that two of the **Gou'ald** had vital interactions with the the team members.*

What program should be used to create the information?

You need a program that can create **.csv** files that your system can read. **Microsoft Excel is an excellent choice** as it has good copy/paste features, multiple spreadsheets in one files, and known compatibility with ORA.
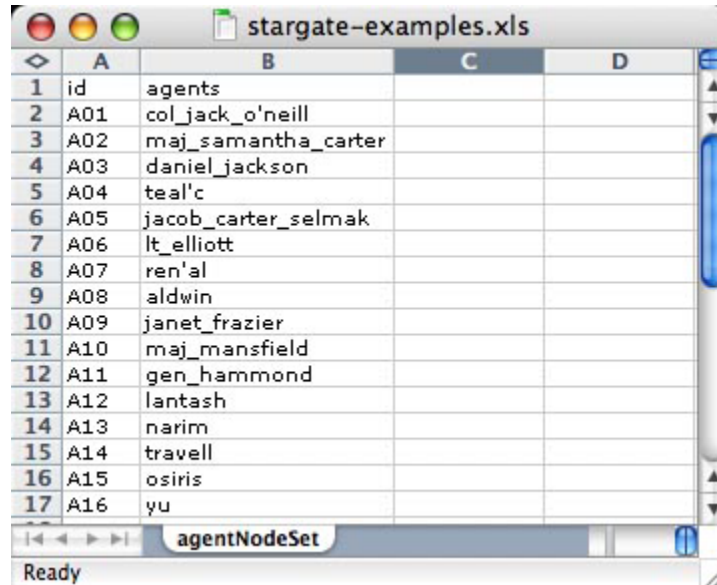
The examples in the rest of this lesson will be using Excel in OSX.

The agent Node set was created with IDs **A01, A02, A03,... A16** and the names of the agents were put in a column called **agent**. The order of the agents does not

matter as long as you keep the relationship between IDs and agents consistent throughout your work. The IDs are what allows ORA to match various files.

Column A contains my IDs A01, A02,… A16 and **column B** contains the names of all the agents in my Agent Node set

*NOTE: I use **A01** instead of **A1** preferring to prefix the zero to the numbers less than ten. This assists when doing a sort within the editor. If you think that you will have a Node set with more than 100 nodes then use A001, A002, …*

Your first Network

Now that the agent Node set is done the next item is to create an **agent x agent** Network. As described in **Meta-Network 101** just how your agents interact with one another is subjective to the type of Network you have.

Insert a new spreadsheet in Excel and place the agent IDs running down in column A starting with row 2 and the same set of agent IDS in row 1 starting with column B. Then place your **link** values for the **agent by agent** connexions in this matrix.

*The **Stargate** example was done with a **hierarchical** method showing the chain of command. Nodes **A02**, **A03**, **& A04** represent **maj_samantha_carter**, **daniel_jackson, & teal'c** who all report to **col_jack_o'neill (node A01)**, leader of **SG-1**. jack_o'neill reports directly to **gen_hammond (node A05)**. daniel_jackson, & teal'c, both being civilians, also report to gen_hammond and thusly have two links in the chain of command. maj_samantha_carter, being military, directly reports to her team leader, col_jack_o'neill.*

	A01	A02	A03	A04	A05	A06	A07	A08	A09	A10	A11	A12	A13	A14	A15	A16
A01	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
A02	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A03	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
A04	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
A05	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0
A06	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0
A07	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A08	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
A09	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
A10	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
A11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A12	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0
A13	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
A14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
A16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0

The rest of the Node sets and Networks

Next you have to decide what other Node sets you need…

*I created Stargate Node sets for **events:** important events during the course of the scenario, **knowledge:** important information necessary for the completion of the mission, **resource:** material objects important to the mission, **location:** where particular events took place, and **tasks:** specific jobs for the completion of the mission.*

…and how these Node sets need to interact with each other.

*For Stargate_summit I decided I needed: **agent x agent, agent x event, agent x knowledge, agent x resource, agent x task, event x event, event x resource, knowledge x task, resource x task, task x event, and task x task**.*

Each Meta-Network will require different handling and may require more or less networks.

Then, as done in the agent x agent network, decide how the various nodes interact with each other for each Network. When finished you will have one spreadsheet from which you can create all your **.csv** files.

NOTE: *I keep all my spreadsheets in one file for easy access. One set for the **Node sets** and the other set for the **Networks**.*
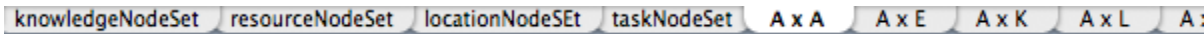

Saving as .CSV files

The next step is to convert these Excel file to a format that ORA can import, namely the .csv format. This is easily done with Excel's **Save As...** function. Select the Excel spreadsheet you want to convert to CSV. From the Excel main menu select **File ⇒ Save As...**. Navigate to the directory to save your file, from the drop down menu select the **csv** format, and hit **OK**. Opening up this file in a text editor will reveal it's new format.
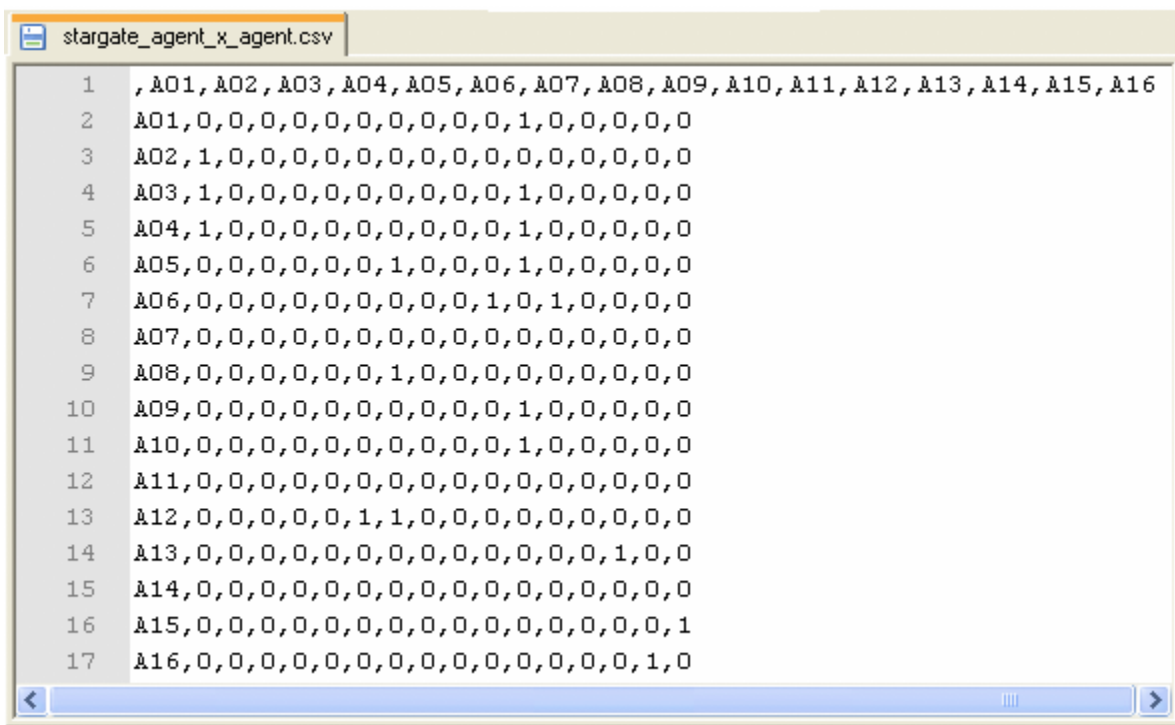
You can now convert all your Excel spreadsheets to CSV files in preparation for importing into ORA.
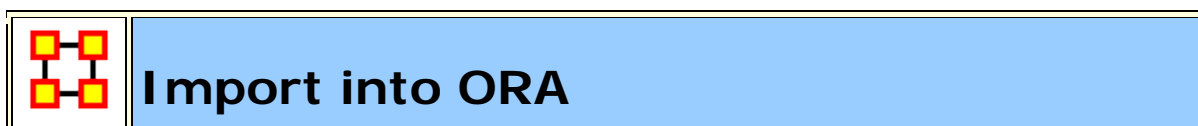
Import into ORA

Now that we've got the Excel spreadsheets created and converted to **.csv** files it's time to learn about **importing a CSV file into ORA**.

The Datasets

Below is the file we will be importing. It denotes **who reports to who** in the SG-1 unit and is called **sg1_agent_x_agent.csv**. This, and the other Stargate files, can be found in the ORA program folder **sample data > Stargate**.
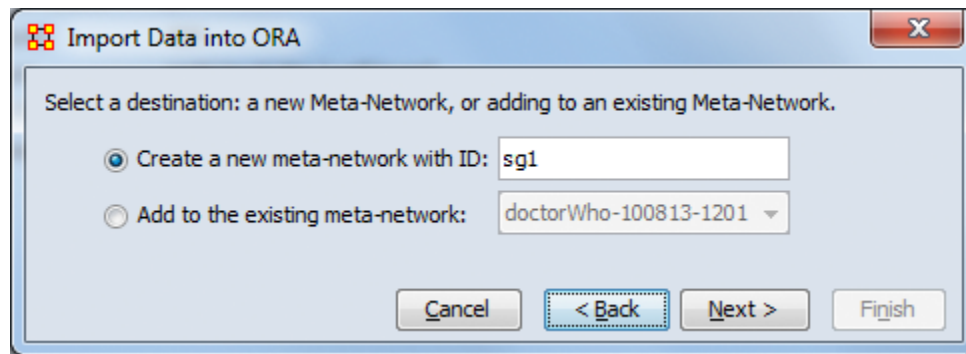
| sg1_agent_x_agent.csv					
	A01	A02	A03	A04	A05
A01	0	0	0	0	1
A02	1	0	0	0	0
A03	1	0	0	0	1
A04	1	0	0	0	1
A05	0	0	0	0	0

Starting a new Meta-Network

From the main menu of ORA, select **File > Data Import Wizard** or hit **Ctrl-W**. The Data Import dialog asks **What would you like to do?**. For this example we will be importing a network comprised of links so select **Rectangle of link weights**. Then click **[Next]**.
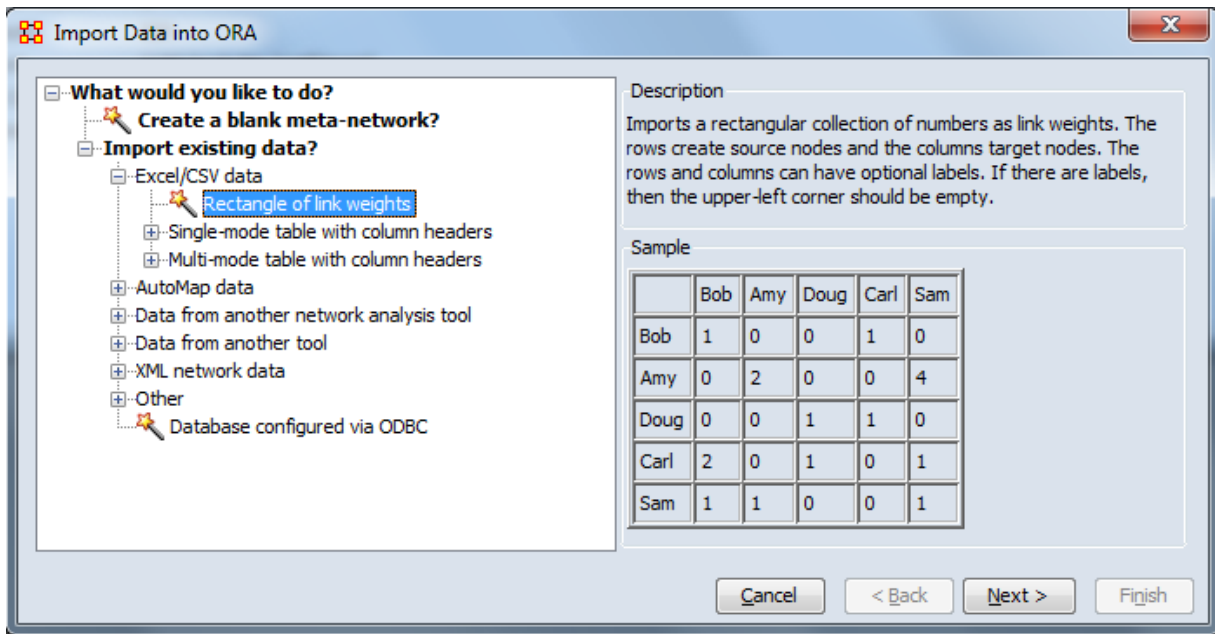
In the next dialog box you will want to select **Create a meta-network with ID:**. This allows you to import a **.csv** file in the form of a **matrix** with labels on the X and Y axis (i.e. an **agent by agent** graph or **resource by event** and link data in the body). Give your new meta-network a name. I named mine **sg1**. When you are done, click **[Next]**.

Click the **[Browse]** button and navigate to the **ORA > sample data > Stargate > csv** directory (mine were in a different directory). Select the **sg1_agent_x_agent.csv** file. In the Source type and Target type drop down menus select the types of nodes you are using. In this case both will be **agent**. In the **Network ID:** field, name the network **(typically for this it would be agent x agent)**. When done click **[Finish]**.
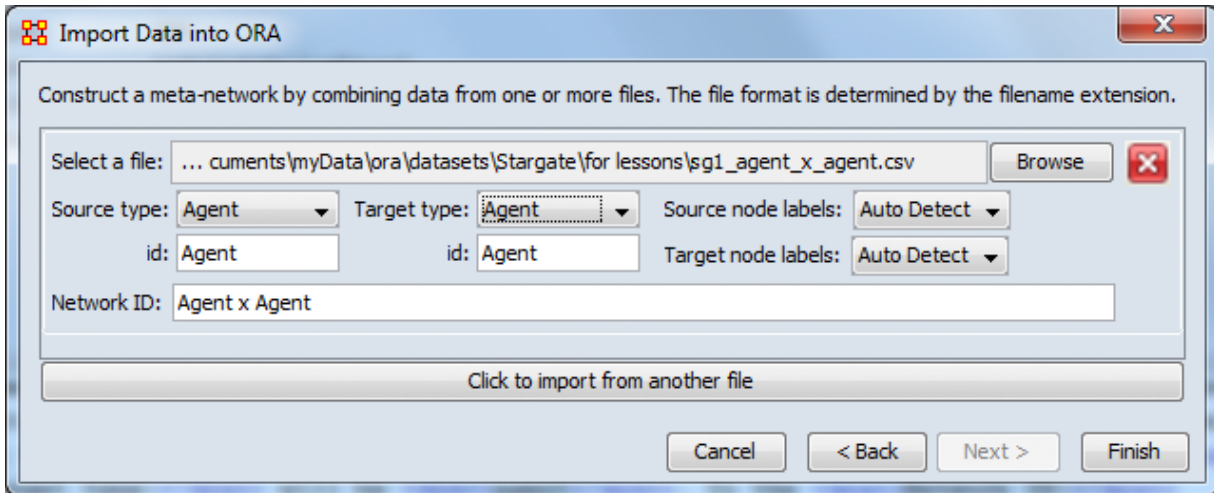
Looking in the Meta-Network panel you'll see the stargate Meta-Network. Clicking the [+] will expand it to reveal the NodeClass and Network created. Highlighting either the **Agent: size 5** or the **Agent x Agent** line and selecting the **Editor** tab will reveal the imported data.
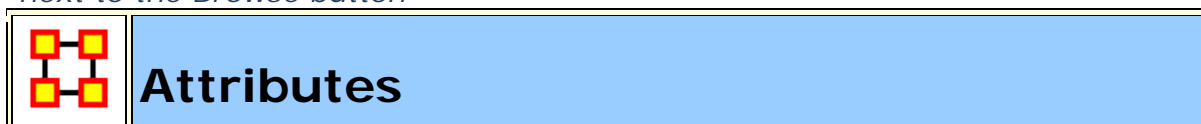

Adding to the new Meta-Network

You could, if you had multiple networks, import them all in at the same time. For each new Network file to import, click the **[Click to import from another file]** button and repeat the above procedure.

NOTE: If at any time you want to remove a Network from the list select the **[X]** *next to the Browse button*

Attributes

The **SG-1 NodeSet** has been created when the network was imported. But without naming the individual nodes and without any attributes for those nodes it's very limited in it's information. **Attributes** help define nodes as individuals by designating what is special about each node.

The Datasets

sg1_agent_attributes.csv contains some base information got the five nodes of our basic network. With it we will set the nodes title and give it information on one attribute, gender. This, and the other Stargate files, can be found in the ORA program folder **sample data > Stargate**.

sg1_agent_attributes.csv		
id	title	gender
A01	col_jack_o_neill	male
A02	maj_samantha_carter	female
A03	dr_daniel_jackson	male
A04	teal_c	male
A05	gen_hammond	male

The Special Attribute, "Title"

All attributes give information about the node they're attached to. But the **title** attribute is special. After importing the **title attribute** to a NodeSet it will lend this title to all Networks that use this NodeSet.

*NOTE: In the **SG-1 Nodeset** the title attribute consists of the **character's names**. This is the standard operating procedure but exactly what you use as the title is entirely up to you.*

Currently the NodeSet and Networks created only contain IDs which are not very easy for discerning **who is who**.

Below is the SG-1 **agent NodeSet** (top) and the **agent x agent Network** (bottom). Truly this needs more definition

Adding Attributes

Once again data needs to be imported to give it some meaning. But in this example the option, **Node Attributes** will be used as we want to import attribute data for the NodeSet, not Network linking data.

First we need a duplicate of the sg1 Meta-Network for this example. Highlight the Meta-Network then either select **Edit > Copy** or press **Ctrl-C** to copy it. Then either select **Edit > Paste** or press **Ctrl-P** to paste another copy. Highlight the bottom Meta-Network. In the Editor panel select the Meta-Netowrk ID textbox. Rename this Meta-Network **stargate** as we are going to import different information into each.
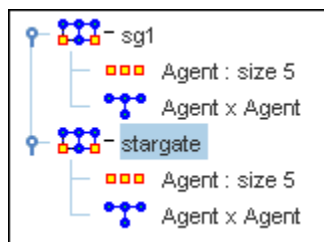
Find the **stargate_agentNodeSet.csv**. This contains all the attributes for all the nodes. In the first one we only want attributes for the five original nodes. In the second we'll allow ORA to populate our Meta-Network with any new nodes it finds.

Highlight the **sg1** Meta-Network. Then click **Import existing data? > Excel/CSV data > Single-mode table with column headers > Node Attributes**. The click **[Next]**
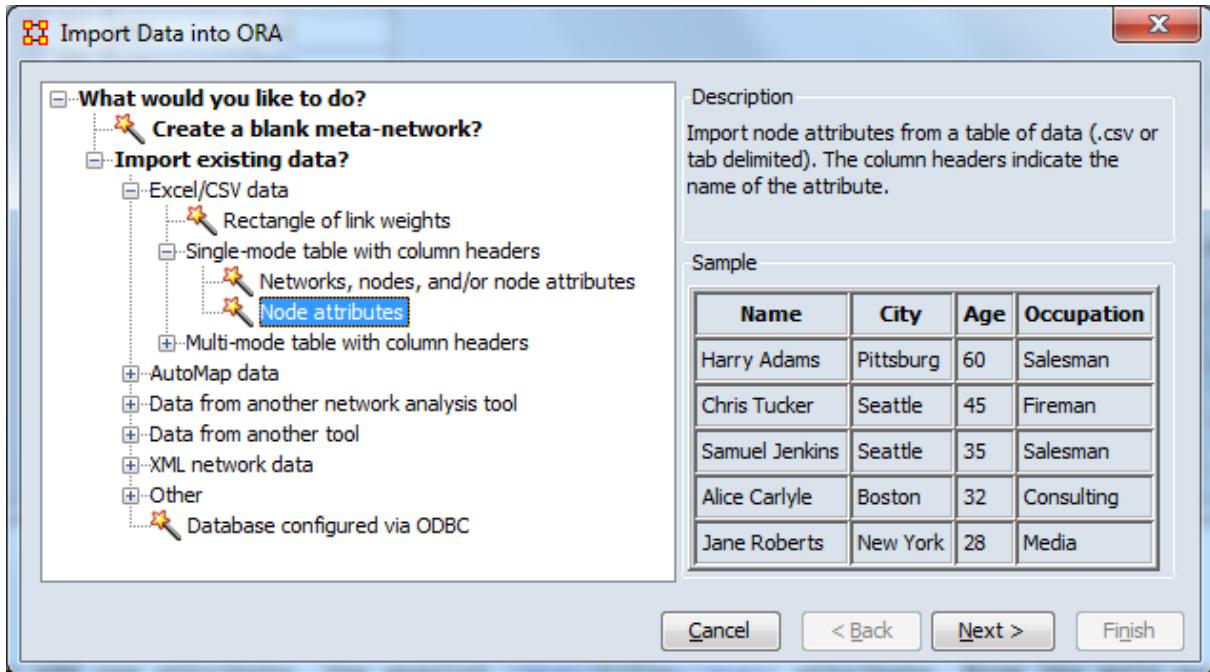
With the basic SG-1 NodeClass and Network already loaded we will now add attributes to the NodeClass. Place a check mark next to the sg1 dataset. Then click the **[Next]** button.
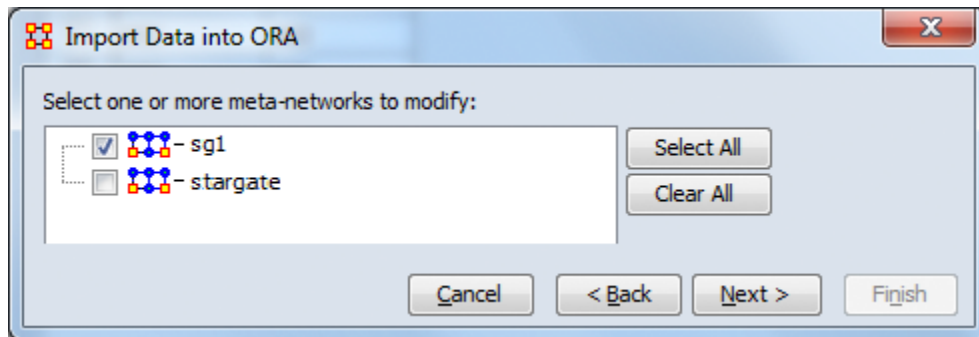
Below is the main dialog box to inform ORA **what and how to import the new attributes**. After the image is a description of each of the numbered steps.
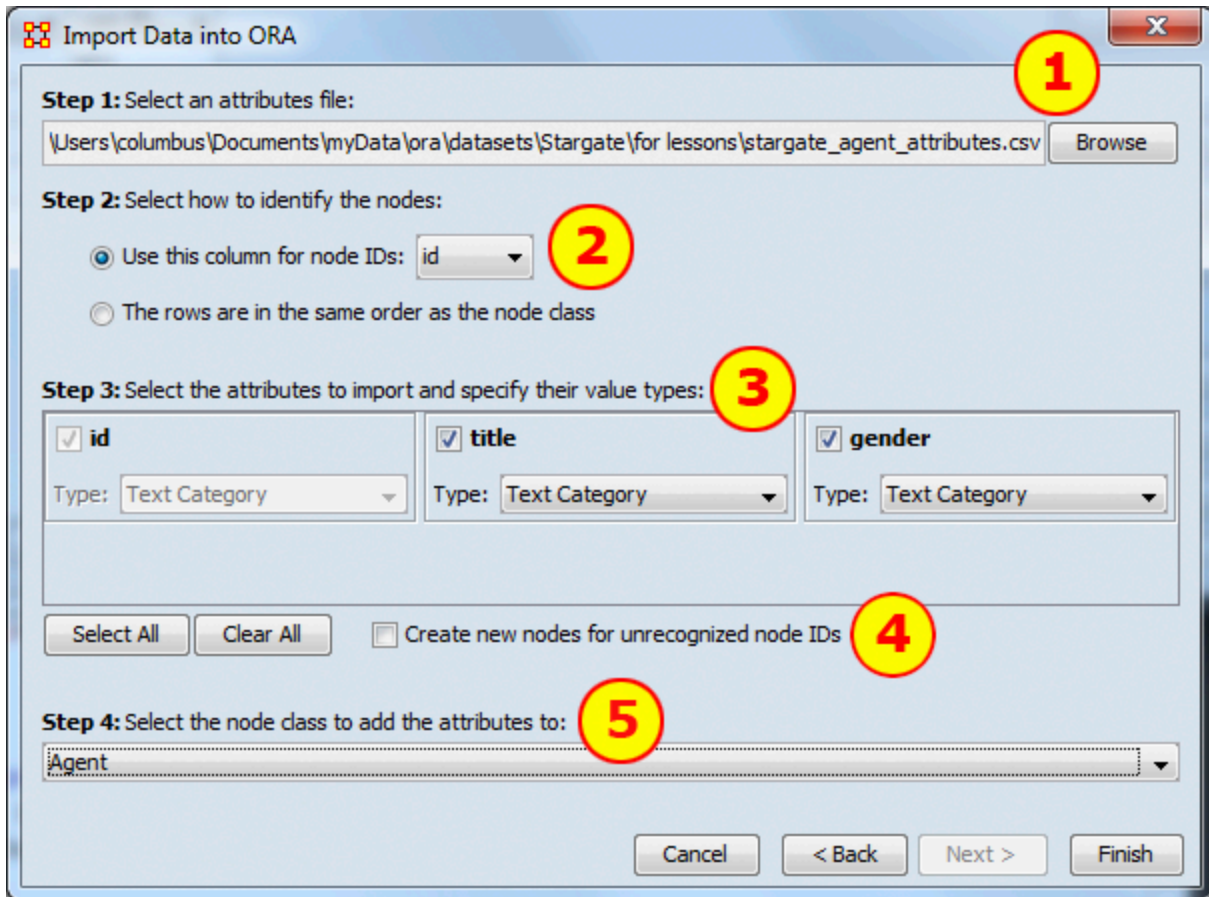
- **1** With the **[Browse]** button you load the file containing the attributes to add. Once a file is selected, click the **[Next]** button

- **2** Selects how to identify the nodes. If your file contains a column with node IDs use the top selection. Then select the column from the drop down menu. If your attribute file contains information in the same order as the nodes are already in then you can use the second option.

NOTE : If you select the second option, and your nodes are in a differnt order than your attributes, the ORA will place the attributes in sequential order disregarding whether the IDs match or not.

- **3** For each column a box will appear allowing you to chose which attributes to import. Place a check mark next to each attribute you want and give it a Type: **Text Category, Text, URI, Number Category, Number, or Date**. An attribute without a check mark will not be imported.

- **4** In this example leave this unchecked as we only want attributes for the first five nodes.

- **5** Select the NodeClass to add the attributes. This drop down will contain multiple entries if your Meta-Network contains more than one NodeClass.

When finished entering information on this screen, click the **[Finish]** button. The NodeSet now contains names that match up to the IDs from the imported file.

	Nod...	Nod...	ally	gender	group	name	position	race	servitude
☐	A01	A01	yes	male	sg-1	col_jack_o'...	sg-1_leader	tauri	us_military
☐	A02	A02	yes	female	sg-1	maj_samant...		tauri	us_military
☐	A03	A03	yes	male	sg-1	daniel_jack...		tauri	sgc
☐	A04	A04	yes	male	sg-1	teal'c		jaffa	sgc
☐	A05	A05	yes	male	sgc	gen_hammo...	sgc_comma...	tauri	us_military

Now repeat the proces but this time select the **stargate** Meta-Network. Be sure to place a check mark in the check box **Create new nodes for unrecognized node IDs**. This will tell ORA that if it finds ndoe IDs which are not in the current Meta-Network then add them and their attributes.

	Nod	Nod	ally	dies	dies_when	gender	group	host	name	position	race	race_2	servitude	symbiote
☐	A01	A01	yes			male	sg-1		col_jack_o'	sg-1_leader	tauri		us_military	
☐	A02	A02	yes			female	sg-1		maj_samant...		tauri		us_military	
☐	A03	A03	yes			male	sg-1		daniel_jack...		tauri		sgc	
☐	A04	A04	yes			male	sg-1		teal'c		jaffa		sgc	
☐	A05	A05	yes			male	sgc		gen_hammo...	sgc_comma...	tauri		us_military	
☐	A06	A06	yes	dies	setting_off_...	male	sg-17	t_elliott	t_elliott		tok'ra	human	us_military	lantash
☐	A07	A07	yes	dies	revanna_att...	female	tok'ra_council	unknown	ren'al	tok-ra_coun...	tok'ra	human	tok'ra	ren'al
☐	A08	A08	yes	dies	revanna_att...	male		unknown	aldwin		tok'ra	human	tok'ra	aldwin
☐	A09	A09	yes			female	sgc_medical		janet_frazier	sgc_medica...	tauri		sgc	
☐	A10	A10	yes	dies	revanna_att...	male	sg-17		maj_mansfi...	sg-17_leader	tauri		us_military	
☐	A11	A11	yes			male	tok'ra/earth...	jacob_carter	jacob_carte...	tok'ra/earth...	tok'ra	human	tok'ra	selmak
☐	A12	A12	yes	dies	setting_off_...	male		t_elliott	lantash		tok'ra		tok'ra	lantash
☐	A13	A13	yes	dead		male	tollan_council		narim		tollan		tollan_council	
☐	A14	A14	yes	dead		female	tollan_council		travell	tollan_coun...	tollan		tollan_council	
☐	A15	A15	no			female	system_lords	sarah_gard...	osiris	system_lord	gou'ald	human	anubis	osris
☐	A16	A16	no			male	system_lords	unknown	yu	system_lord	gou'ald	human	system_lords	yu

NOTE : *One thing to notice is there were more attributes added in the second procedure. When importing attributes into* **sg1** *ORA ignored any attributes which contained no values for any of the nodes in the Meta-Network. If the value of an attributes was* **null** *for all the nodes then it was not added.*

The remaining NodeClasses and Networks

With any remaining NodeClasses and Networks for your dataset you can repeat the previous two lessons: first importing a Network of links then importing the attributes for that nodeClass.

Modifying a Meta-Network

At this point the **Meta-Network** has been created and it contains the NodeSets and Networks. And each of the NodeSets contains titles and attributes in which to

better define the NodeSets and Networks. But as with any dataset there will always be changes.

Most certainly, as time goes on, NodeSets will require extra attributes or even new nodes. Possibly even new NodeSets will need to be created and worked into the Meta-Network. Or maybe a current set of attributes needs to be changed. What follows are the procedures to accomplish those tasks.

Adding a New Node to a Meta-Network

In this section we'll add a new node to the agent NodeSet using the **Create new nodes from unrecognized IDs** checkbox. It's for the situation that the data being imported contains nodes that are not in the current NodeSet.

The new .csv file we're using contains one new set of information, the 17th agent to add to the agent NodeSet. Upon importing the new node we will have a 17th node with the **title** attribute.

id,title,position,symbiote,host,dies_when,dies,ally,group,servitude,gender,race,race-2

A17,zipacna,,zipacna,unknown,,,no,anubis,anubis,male,gou'ald,human

It was created in Excel (the same way the other NodeClasses and Networks were) and exported out to a new .csv file called **stargate_zipacna.csv**.

The easiest way to bring in a new node is to use the **Attribute Import button**.
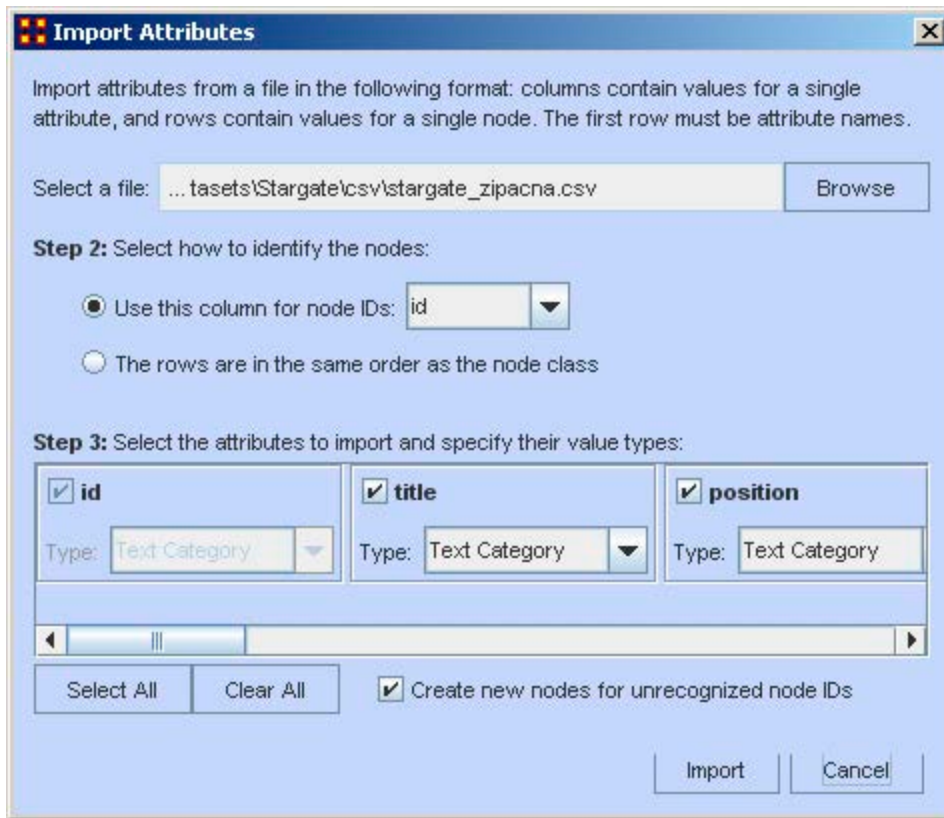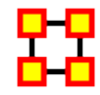
1. Use the **Browse** button to navigate to the directory containing the new .csv file. It will show up in the textbox to the left.

2. Check the proper column to use for the node ID and from the drop down menu select the type of node.

3. Select the **Attribute(s)** you want to import.

4. And most importantly, make sure the box **Create new nodes for unrecognized IDs** contains a checkmark in order to allow ORA to create new nodes. Without this checkmark ORA will ignore any nodeIDs that aren't already in the NodeSet.

When you review your agent NodeSet in the editor you will see a new entry at the bottom, the 17th node with **zipacna** in the **Entity Title** attribute.

	Node ID	Node Title	ally	dies
	<set filte... ▼	<set filte... ▼	<set filte... ▼	<set filte... ▼
☐	A01	col_jack_o'...	yes	
☐	A02	maj_samant...	yes	
☐	A03	daniel_jack...	yes	
☐	A04	teal'c	yes	
☐	A05	gen_hammo...	yes	
☐	A06	lt_elliott	yes	dies
☐	A07	ren'al	yes	dies
☐	A08	aldwin	yes	dies
☐	A09	janet_frazier	yes	
☐	A10	maj_mansfi...	yes	dies
☐	A11	jacob_carte...	yes	
☐	A12	lantash	yes	dies
☑	A13	narim	yes	dead
☐	A14	travell	yes	dead
☐	A15	osiris	no	
☐	A16	yu	no	
☐	A17	zipacna	no	
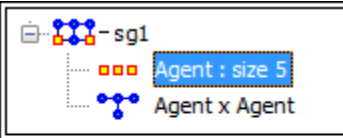
The new node, and all their attributes, have been imported into the Meta-Network.

Working with SubSets

There will be times when you have a SubSet of a larger Meta-Network and you'll want to import data from a larger, main dataset. And you want to do this **without** creating new nodes. What follows are the procedures for accomplishing those goals.

We start off with the SubSet of **SG-1 team and the SGC commanding officer**, a Meta-Network of five agents.

	Node ID	Node Title
	<set filt... ▼	<set filt... ▼
☐	A01	col_jack_o'...
☐	A02	maj_samant...
☐	A03	daniel_jack...
☐	A04	teal'c
☐	A05	gen_hammo...

sg1
  Agent : size 5
  Agent x Agent

The dataset to use has **17 agents with all their attributes**.

26

A	B
id	title
A01	col_jack_o'neill
A02	maj_samantha_carter
A03	daniel_jackson
A04	teal'c
A05	gen_hammond
A06	lt_elliott
A07	ren'al
A08	aldwin
A09	janet_frazier
A10	maj_mansfield
A11	jacob_carter_selmak
A12	lantash
A13	narim
A14	travell
A15	osiris
A16	yu
A17	zipacna

Now, as before, navigate to the directory with your csv dataset file and import it. Select the column with your node ID, select the Attribute tab and create new Attributes using the **[New]** button, and finally, the most important part, remove the checkmark from the box next to **Create new nodes for unrecognized IDs**. Its the removing of this checkmark that will prevent ORA from creating extra nodes and only import data for the nodes in the selected NodeSet.

Import Data into ORA

Select a file: \\.PSF\xChange\ORA\csv\stargate_agentNodeSet.csv   [Browse]

Step 1: Check the columns that contain node IDs and enter the node class information:

☑ **ID**
Type: Agent ▼
ID: Agent
☐ Make unique

☐ **AGENT**
Type: <Select...> ▼
ID:
☐ Make unique

☐ **TITLE**
Type: <Select...> ▼
ID:
☐ Make unique

☐ **POSITION**
Type: <Select.. ▼
ID:
☐ Make unique

Step 2: Define networks and attributes based on the columns:

Networks | Networks and Labels | Networks combined IDs | **Attributes**

Node IDs	Attribute Values	Value Type	Attribute Name
id ▼	position ▼	Continuous ▼	position
id ▼	gender ▼	Categorical ▼	gender
id ▼	leadership ▼	Continuous ▼	leadership
id ▼	race ▼	Continuous ▼	race

[New] [Clear]

☐ Create new nodes for unrecognized node IDs

[Cancel] [Back] [Next] [Finish]

| Entity ID	Entity Title	position	leadership	gender	race
<set filt... ▼	<set filt... ▼	<set filt... ▼	<set filt... ▼	<set filt... ▼	<set filt... ▼
A01	col_jack_...		sg-1	male	tauri
A02	maj_sam...			female	tauri
A03	daniel_jac...			male	tauri
A04	teal'c			male	jaffa
A11	gen_ham...	commander	sgc	male	tauri

– sg1
  Agent : size 5

The result is maintaining a NodeSet with five nodes but attribute data imported only for the five nodes in the NodeSet. No new nodes will be created in this process.

Attribute Columns

There are three functions that can done with Attributes: 1) Edit and Attribute name; 2) edit Attribute type, and; 3) Delete an Attribute.

All three are accomplished by **Right-clicking** on the Attribute Name to bring up the contextual menu.
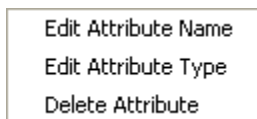
Edit Attribute Name : Used for renaming an existing Attribute column.

Edit Attribute Type : Attributes can be tagged with two types: **Categorical** or **Continuous**.

Delete Attribute : Removes an Attribute column from a NodeSet.

Replacing an Attribute Column

*NOTE: This method is suggested as ORA has the capability of retaining multiple values. If you want to completely replace the values then **deletion and replacement** of the column is suggested.*

For most **small** changes you can type in a new attribute value or two directly into the Editor. But say your NodeSet contained quite a few nodes, and the majority of the values needed changed. If might be easier to create a new .csv file and import the values in just as in the previous lesson. In the example below it was decided to completely replace all the **skill_needed** with a new set of values.

Go to the ORA main interface and select the NodeSet. Then select with the attribute column you want to change. **Right-click** on the attribute name which brings up the contextual menu. Select **Delete Attribute** to remove it from the NodeSet.

29

After deleting a column then use the procedure for importing Attributes from **Lesson-203** to import a new attribute information. Remember, you'll want to import into an **Existing** Meta-Network as in the example above.

skill_needed
<set filter>
symbiote_poison_formula
pilot
tok'ra_tech
memory_drug_formula
tok'ra_tech
tok'ra_tech

The **skill_needed** attribute has been imported into the NodeSet with the revised values.

Updating Your Data Files

After importing data from the .csv files into ORA and then altering the data in ORA this somewhat makes your original data files out-of-date.

What follows are the procedures to update your original .csv and Excel files with any changes that have been made recently.

Saving Your Network Data

From the main interface highlight the Network you want to save and **Right-Click**. This will bring up a contextual menu. Select **Save Network**.
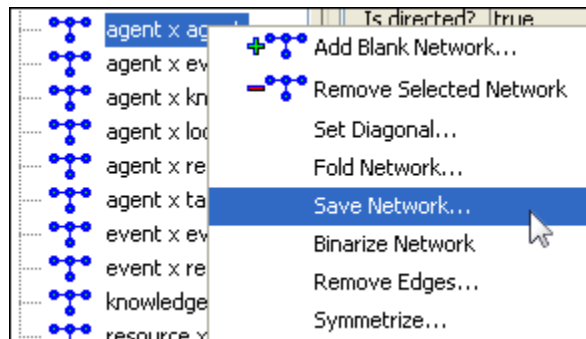
This will bring up a **Save** dialog box. First, navigate to the directory you want to save your new Network file. Next, select the format **(DyNetML, UCINET, DL, or CSV)** to save the file. Give it a filename and select **[Save]**.
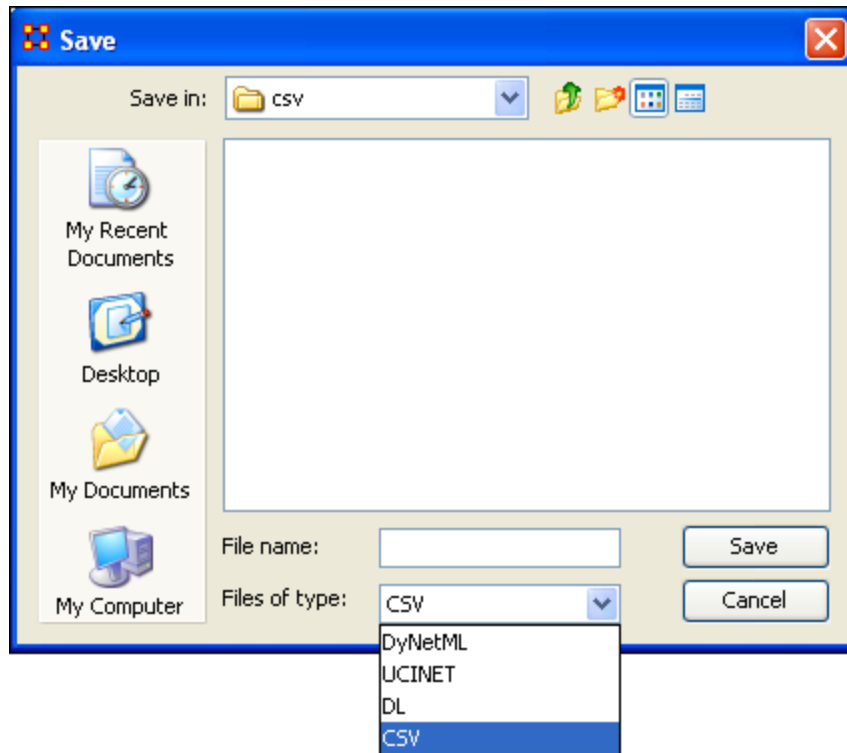
Saving to each of the **four** formats creates **five** files.

REMEMBER: The UCINET uses both .##d and .##h extensions and both are needed to build a Meta-Network.

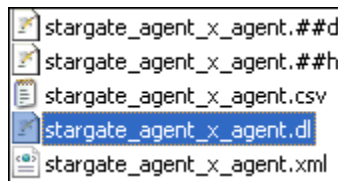*NOTE: Since this **Save** box does not display the files present in the directory it is **Highly Suggested** that you save to a new directory and verify which previous file you want to replace. More information on these file formats can be found **HERE***

Saving Your Attribute Data

The above mentioned procedure works well for saving Networks but there are also the attributes in the NodeSets to think about. All the data is can be saved from **File ⇒ Save Meta-Network** and the **File ⇒ Save Meta-Network…**. But this saves the data into **one xml file**. What if you what to save the data into individual files, the same as when you imported it.

The quickest way is using the **Copy / Paste** method. Values copied directly from the **Main Interface Editor** can be pasted into a **Excel spreadsheet**.

31

In the ORA editor, highlight the cells you need to copy. Use either **Ctrl-C** or **Edit ⇒ Copy Editor Values**.

Entity ID	Entity Title	position	symbiote	dies_when	ally	ho
<set filt... ⌄	<set filt... ⌄	<set filt... ⌄	<set filt... ⌄	<set filt... ⌄	<set filt... ⌄	<s∈
A01	col_jack_...	sg-1_leader			yes	
A02	maj_sam...				yes	
A03	daniel_jac...				yes	
A04	teal'c				yes	
A05	jacob_cart...	tok'ra/eart...	selmak		yes	jac

Then switch to your **Excel** spreadsheet. Highlight the cell that corresponds to the **top/left** cell you started with in ORA and use either **Ctrl-V** or **Edit ⇒ Paste**. Entire NodeSets can be done in this fashion.

◇	A	B	C	D	E	F	G
1	id	title	position	symbiote	dies_when	ally	dies
2	A01	col_jack_o'neill	sg-1_leader			yes	
3	A02	maj_samantha_carter				yes	
4	A03	daniel_jackson				yes	
5	A04	teal'c				yes	
6	A05	jacob_carter_selmak					

*IMPORTANT NOTE: Before you **Copy/Paste** in this manner make sure your columns are in the same order. This operation will **NOT** match up data as it's a **blind copy/paste** operation.*

You can then backup all of your NodeSets in the procedure and then re-save it to your .csv files.

Key Entity Report

The Reports

A **Key Entity Report** creates the following reports: central actors, organizations, central knowledge, resources, central events, tasks, central locations, and performance indicators.

The Key Entity Report can either run its measures over the entire network or be refined or filtered to run them over only a select portion of the network.

More detailed information regarding the basic workings of the Key Entity Report screens can be found **Reports Section**.

Running a Key Entity Report:

We'll be using the network **Stargate: Summit** contains all the agents involved in the scenario, friend and foe. The first report run will contain all the agents, friend and foe.

- From the main menu select **Analysis ⇒ Generate Reports** which brings up the **Generate Reports** dialog box.

- At the top of the window is a dropdown menu titled **Select a Report**. Pull down the menu by clicking the small inverted arrow icon to its right, and select **Key Entity**.

- In the box titled **Select one or more meta matrices:** select the network to run the report on.

- At the bottom of the box is the **Transform options**.

  - **General Mathematical Transformation :** Contains various formulas for manipulating a network.

  - **Remove Entities :** Remove node(s) from a network before running the measures.

  - **Partition entities into separate meta matrices by attribute value :** Create new networks based on your choice of nodes and attributes.
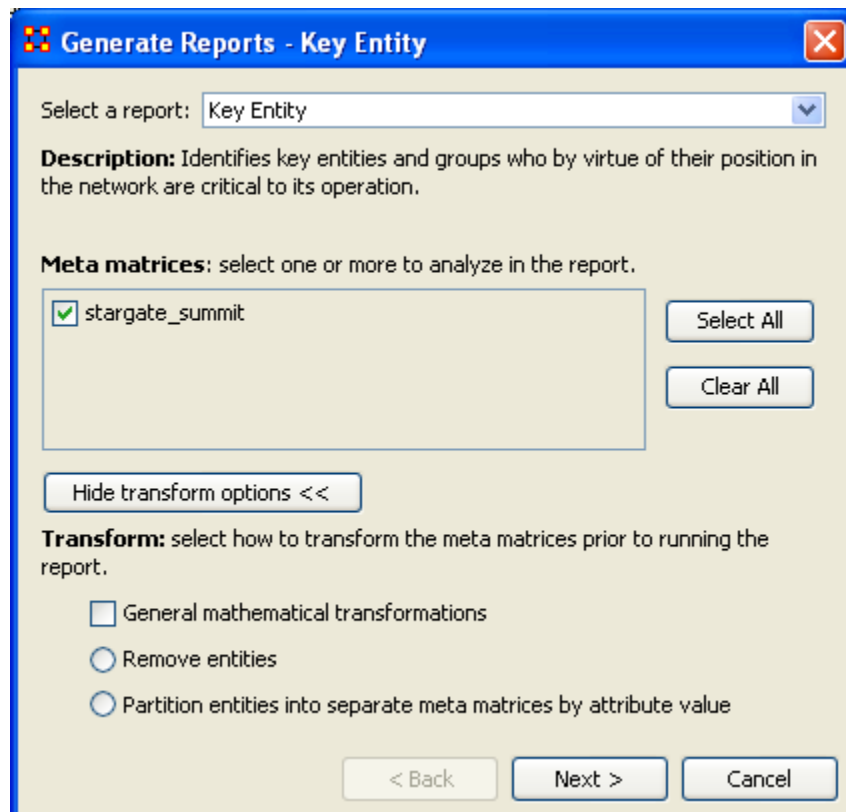
Select the **[Next >]** button at the bottom of the dialog box.

The next dialog box is for selecting **the number of ranked nodes**, **critical attribute and values** and also for selecting **which reports** you want to view. I am only interested in the top five nodes in the network and place a **5** in the top box. I don't want to point out any particular nodes, only the friends, so I'll leave the textbox blank. I want all the reports run so I place a checkmark in all four boxes.

Click the **[Next >]** button.

The last dialog box is for selecting what type of format for the final reports and the directory to place them. **Text** will create reports that will appear in **Panel 3** of the interface, **HTML*** creates reports that will open up your browser, and **CSV*** creates files that are usable in ORA. The bottom section allows you to navigate to a directory where you want to save your files.

Remove Nodes from a Key Entity Report:

Next we'll look at removing nodes from the report. The network **Stargate: Summit** contains both the friends and foes in the scenario. But say you only want to see only how the **friendly agents** interacted. Being that there's an **ally** attribute this will make it simple to eliminate the **foe agents**.

Follow the same procedure as above except click in the **Remove entities** radio button.
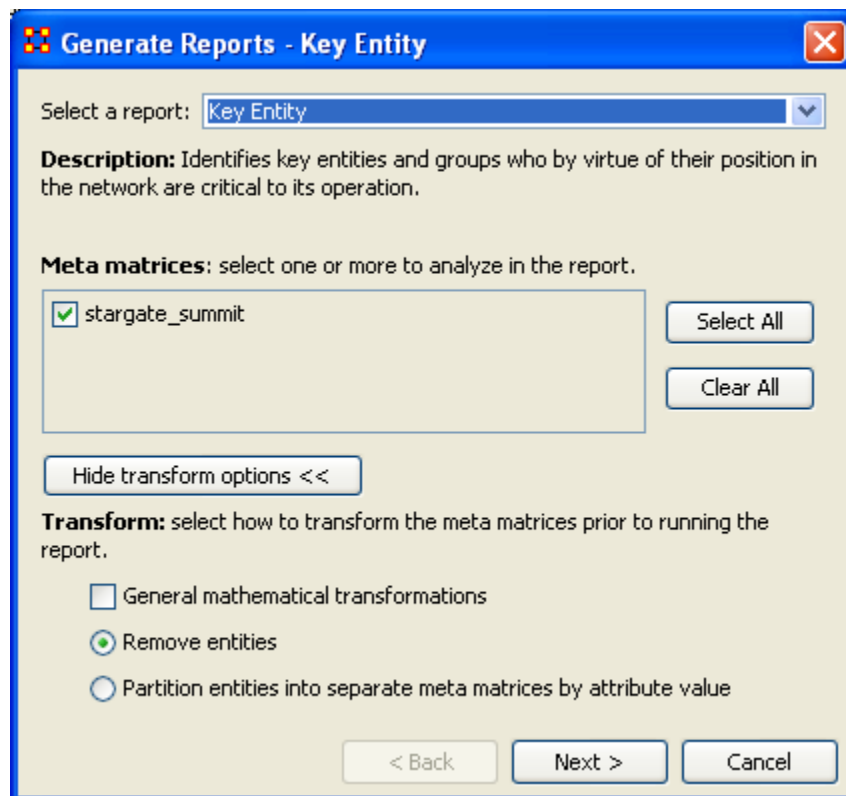
In this example we're going to eliminate a portion of the network before we run the measures. Select the **Remove entities** button. Then select the **[Next >]** button at the bottom of the dialog box.

In this dialog box you can select exactly which nodes to remove.
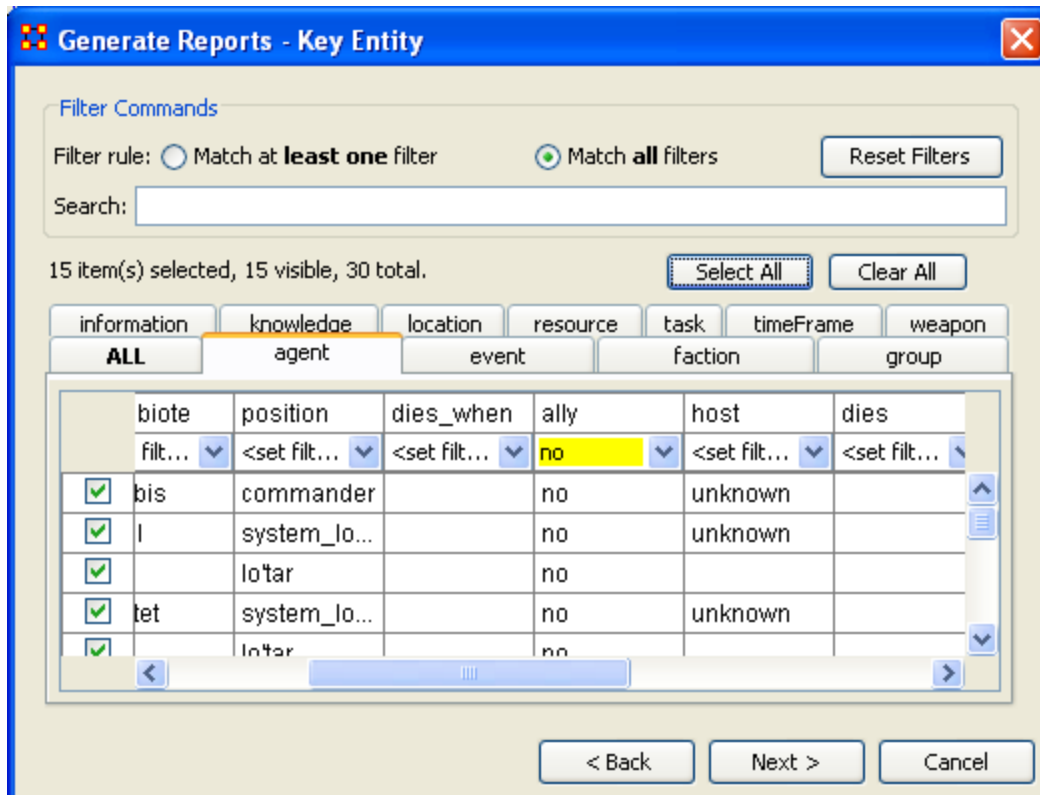
The top section, **Filter Commands**, allows you to create a **single filter** or compound them and force it to **Match all filters**.

The bottom sections contains tabs which further allow you to refine your filter. There is a tab for each nodeset in your network.

I only want to run the measures on the **Allies** so I select the **agent** tab, scroll over to the **ally** tab, and in the dropdown menu I select **no**. This will call up all the **non-ally** agents. I then click the **[Select All]** button which places a checkmark in each box effectively removing them from the network.

Click the **[Next >]** button.
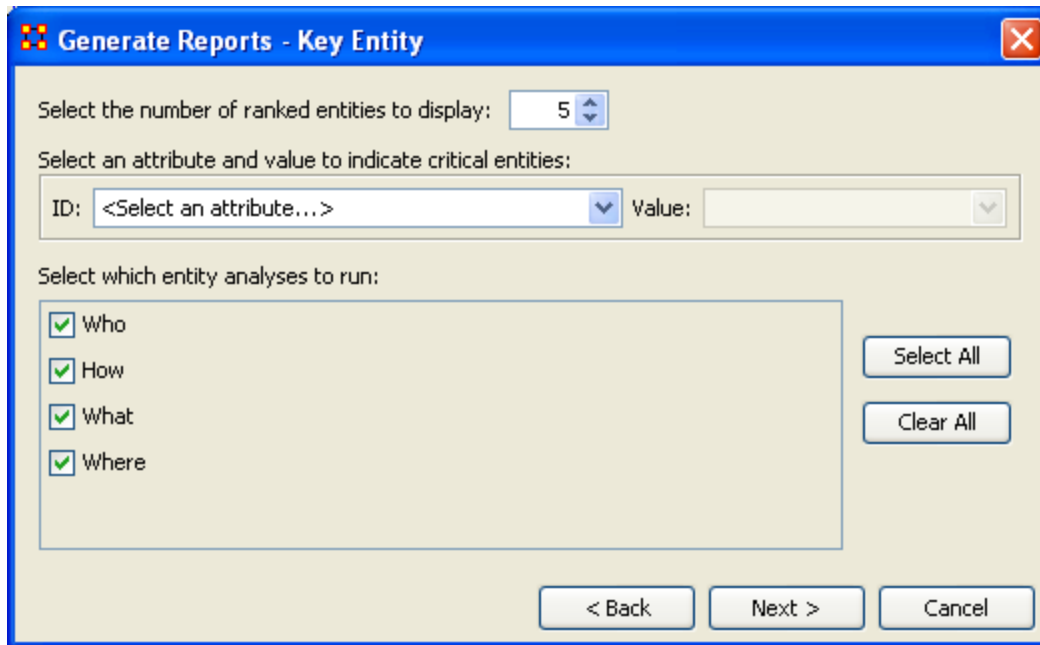
The next dialog box is for selecting **the number of ranked nodes**, **critical attribute and values** and also for selecting **which reports** you want to view. I am only interested in the top five nodes in the network and place a **5** in the top box. I don't want to point out any particular attributes so I'll leave the textbox blank.

Click the **[Next >]** button.
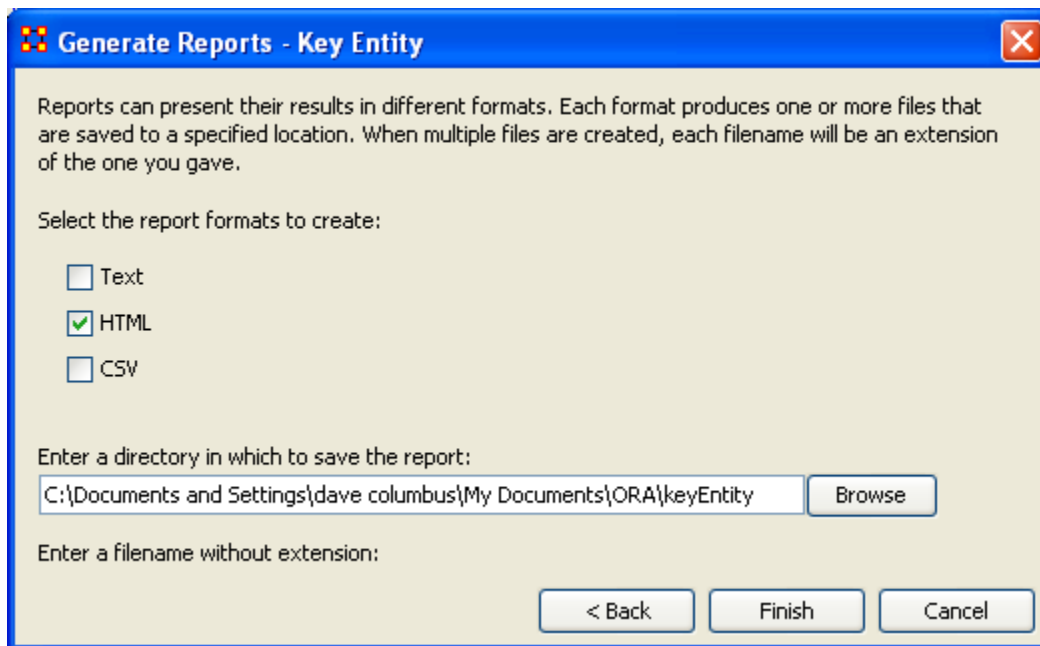
The last dialog box is for selecting what type of format for the final reports and the directory to place them. **Text** will create reports that will appear in **Panel 3** of the

interface, **HTML*** creates reports that will open up your browser, and **CSV*** creates files that are usable in ORA. The bottom section allows you to navigate to a directory where you want to save all your files.

Comparison of the two reports

As the reports are rather large, let's compare a small portion of cognitive effort from the reports.

| All agents			Allies-only		
Rank	Value	Agent	Rank	Value	Agent
1	0.340911	jacob_carter_selmak	1	0.354626	jacob_carter_selmak
2	0.257479	ren'al	2	0.284762	daniel_jackson
3	0.248333	daniel_jackson	3	0.284218	ren'al
4	0.205698	olokun	4	0.166667	maj_samantha_carter
5	0.205698	morrigan	5	0.164218	teal'c

The left-hand column shows all nodes, including the non-allies **olokun and morrigan**. But after eliminating the non-allies two friendly agents take their place. You'll also notice that **daniel_jackson and ren'al** swapped places. With various links removed, after eliminating the non-allies, the order of the remaining nodes can shift. It should also be noticed that the rating for jacob_carter_selmak has increased elevating his position as Emergent Leader.

Over-Time Analysis

Performing a View Network Over-Time Analysis

What follows are procedures to perform a **View Network Over Time** and **View Measures Over Time** Analysis.

To either requires multiple network datasets representative of your network over different time periods. The goal is to analyze how a network changes across multiple time periods. The time period is determined by the intervals of collected network data and presented as multiple Meta-Networks, which are loaded into ORA

in sequential order. The steps below provide detailed instruction on how type of analysis.

To perform a **View Network Over-Time Analysis**, you must load Meta-Networks that relate to different time captures of your network data, which need to be in the form of multiple Meta-Networks.

In the following example, we will use network time captures of the **1998 Embassy Bombing in Tanzania** Meta-Network data set and compare the same data set as we knew it in 1997 and 1996 and 1998. We will then compare all the Meta-Networks in the Network Over-Time Viewer to gain a visual representation of how this network changed between 1996 and 1998.

Note: For our purposes, we have changed data in this example and have renamed the Meta-Network for three different years. This will denote different network variables that have changed over time.

You will see in the left window pane under input dataset three Meta-Networks, 1998, 1997, and 1996, have been loaded.

From the menu bar: File > Open Meta Network
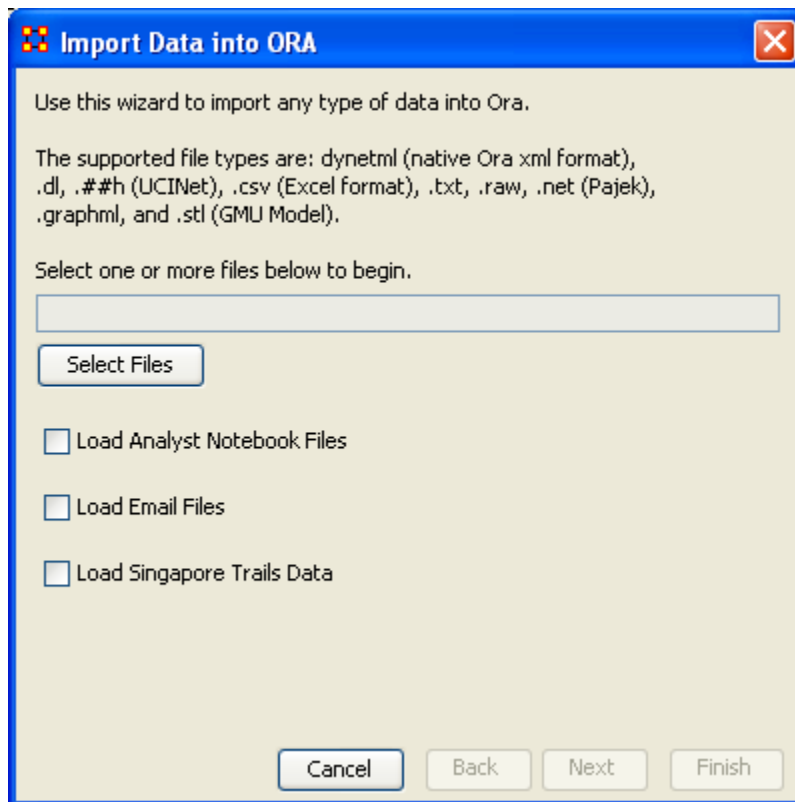
Choose [Select Files]
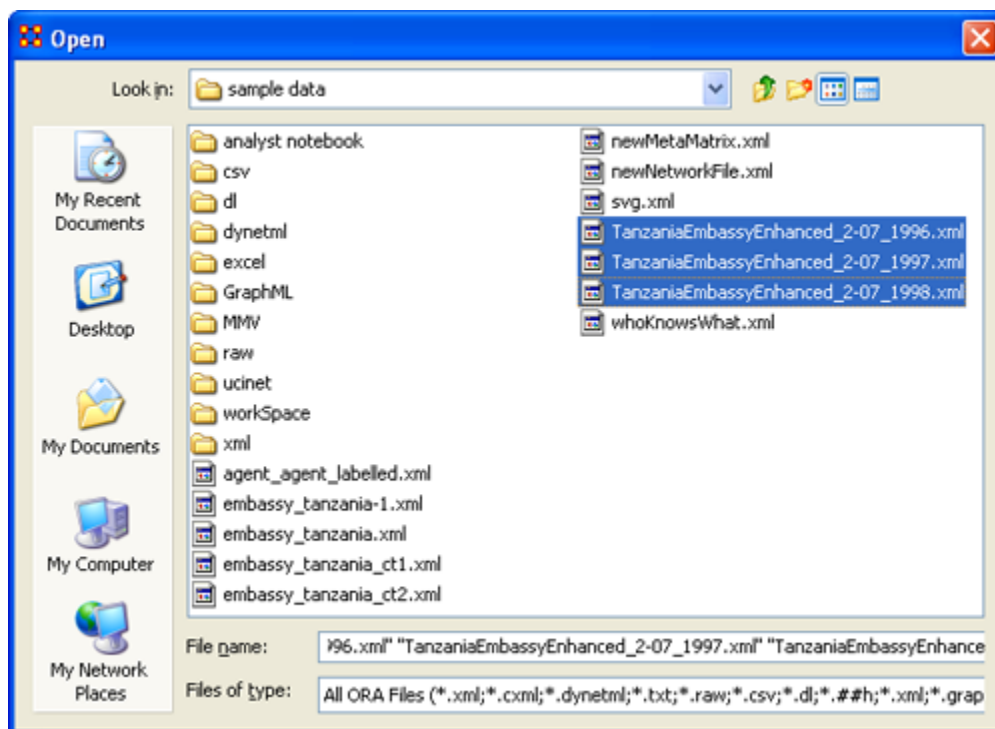
You will see all of our example data sets, which will load simultaneously into ORA, buy highlighting and selected all the files.

Highlight all files you analyze in Over-Time Viewer. Press [Enter].
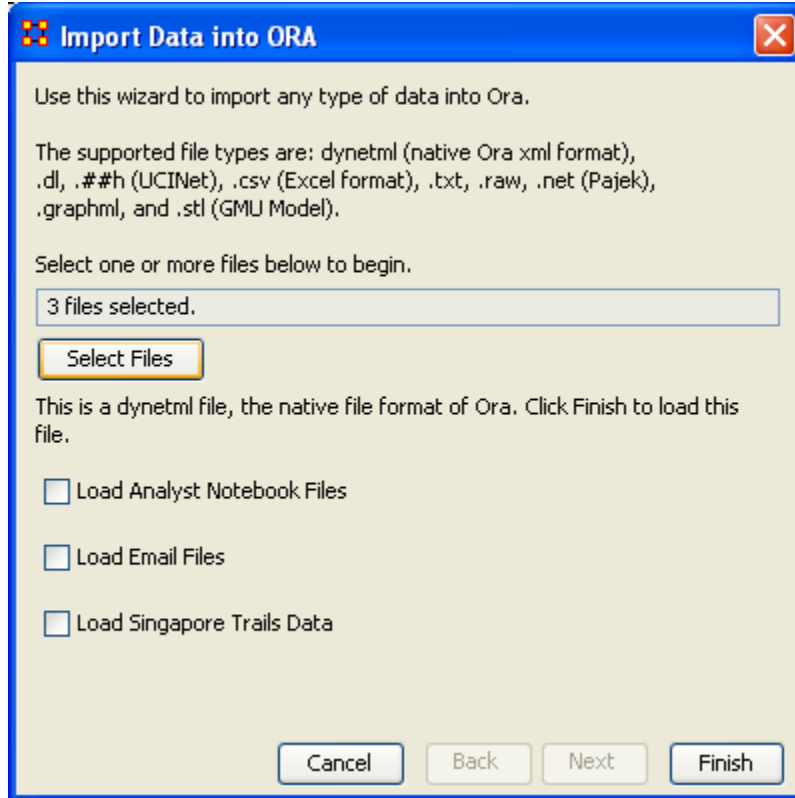
The import data window will appear (below). This will confirm the files you wish to analyze.
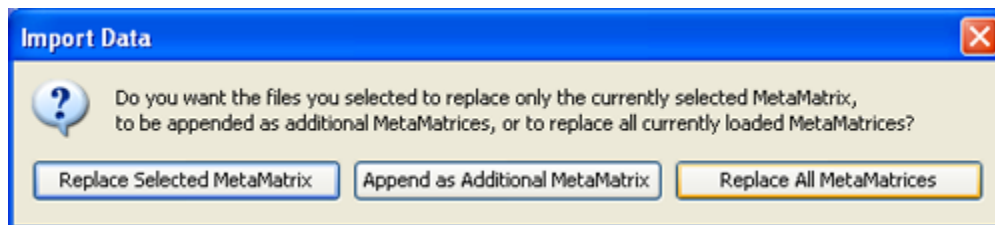
Select [Finish].

All three Meta-Networks should display in Panel 1.
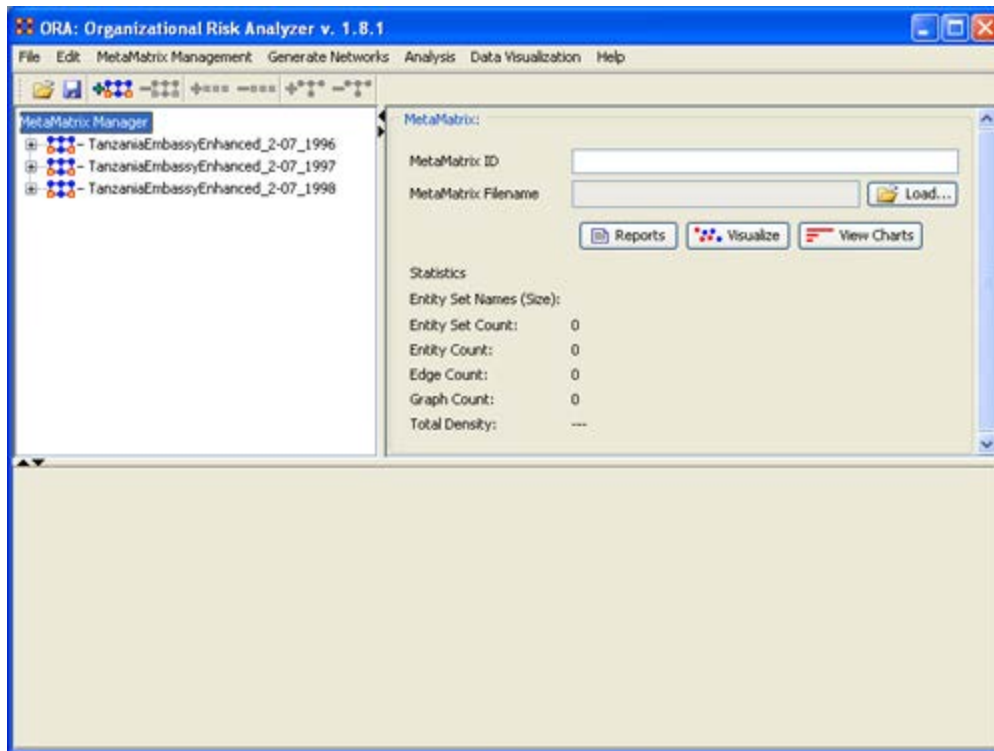
Performing the Over-Time Analysis

Now that we have time captures of our network data represented and loaded into ORA in the form of three Meta-Networks relating to **1996, 1997 and 1998**, we can perform the Over-Time Analysis.

From the Menu Bar > Data Visualization > View Network Over-Time
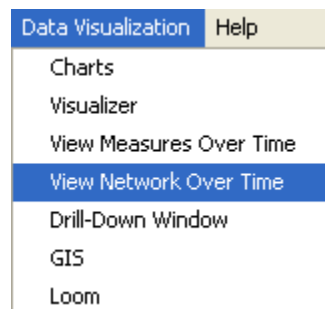
The ORA Visualizer will render all of the loaded Meta-Networks.

Please take special note of the Time Viewer pop-up Window. This window displays a slider, which will allow you to quickly change from the first Meta-Network time capture of **1996, to 1997 and 1998** respectively.

By moving the slider from one Meta-Network to the next, you will gain a visual perspective on how the network has changed over Time.

The series of screen shots below display how the visualized data changes from one slider to position to the next. This allows you a picture of how the network changes Over-Time.

Please note: Depending on the complexity of the network and the extent to which relationships (ties) have changed, the visualization of those changes may be subtle.

The screen shots below, however, will emphasize the extent to which the role of the **WTC Event** entity has changed from **1996 to 1998** in relation to the network of which it is part of namely the Tanzania data set. Below these screen captures, are a series of detail screen captures focusing on the WTC Event node.

Example Slider Position 1
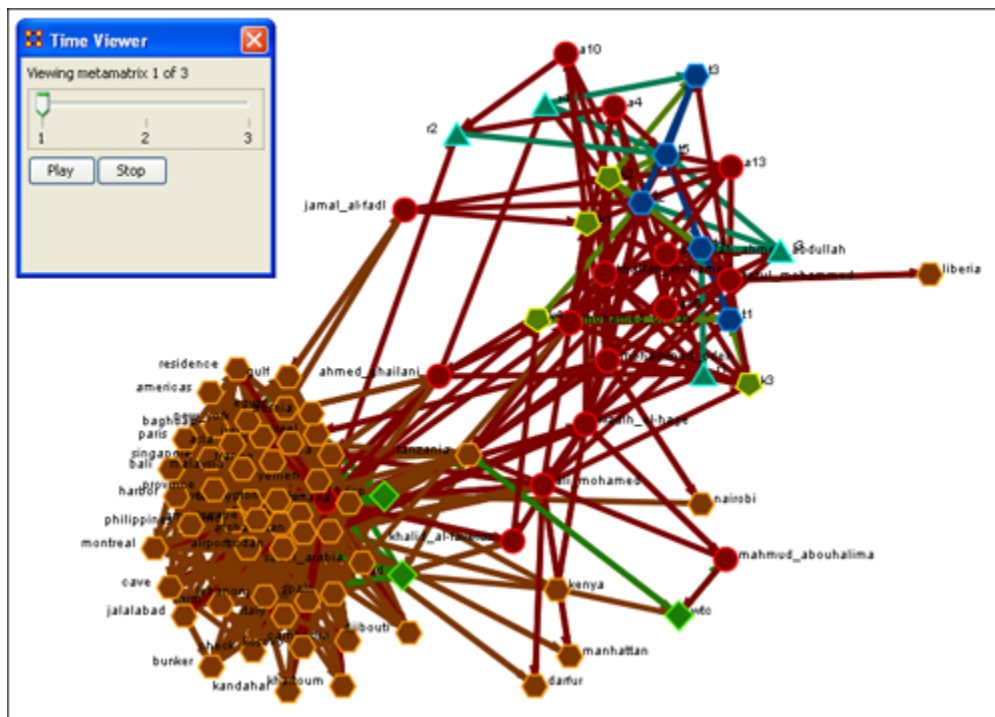
Example Slider Position 2
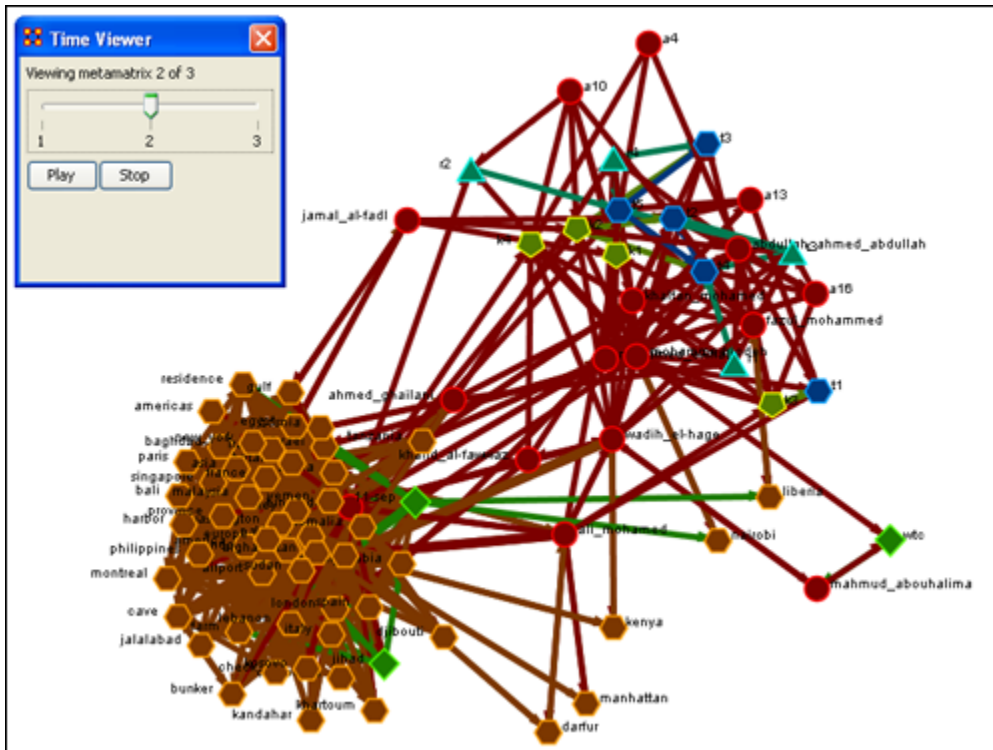
Example Slider Position 3
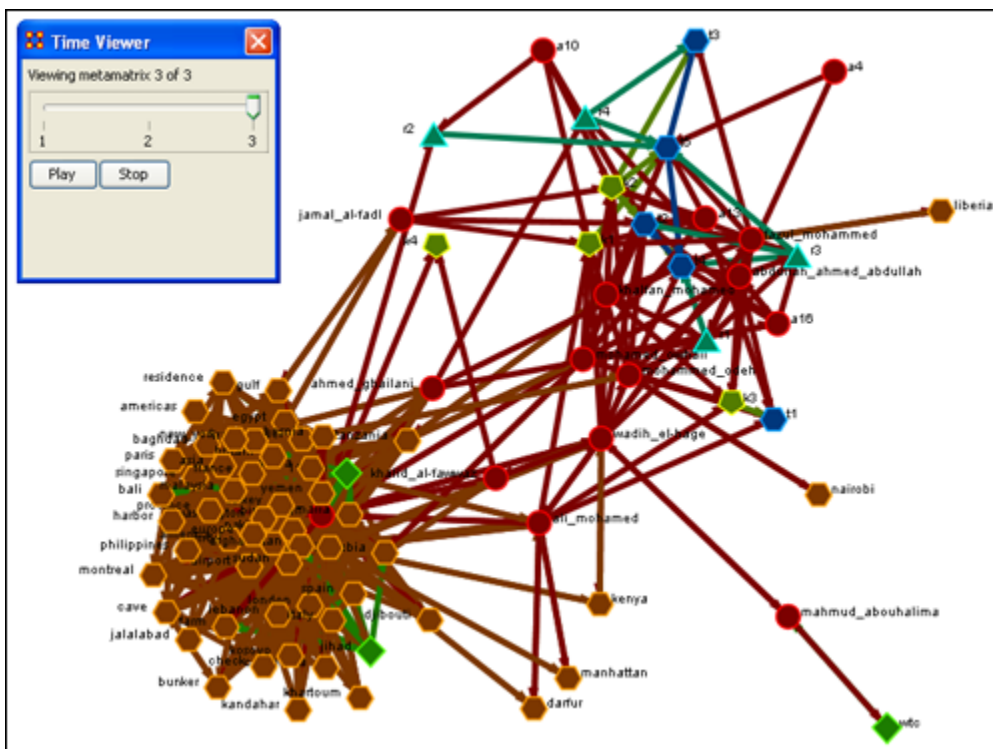
Detail of WTC Event Node when visualized in the Over-Time Viewer as part of the 1996, 1997 and 1998 Tanzania Data Sets.

WTC Event Node: Detail 1 - 1996
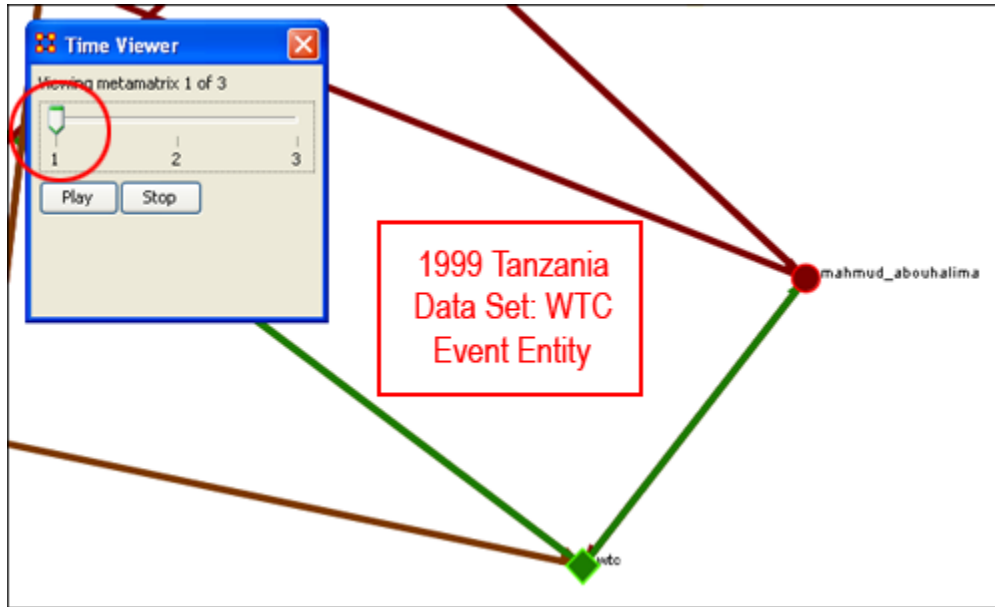
WTC Event Node: Detail 2 - 1997

Notice who how the relationships of this node to the rest of the network has changed from 1996 to 1997.
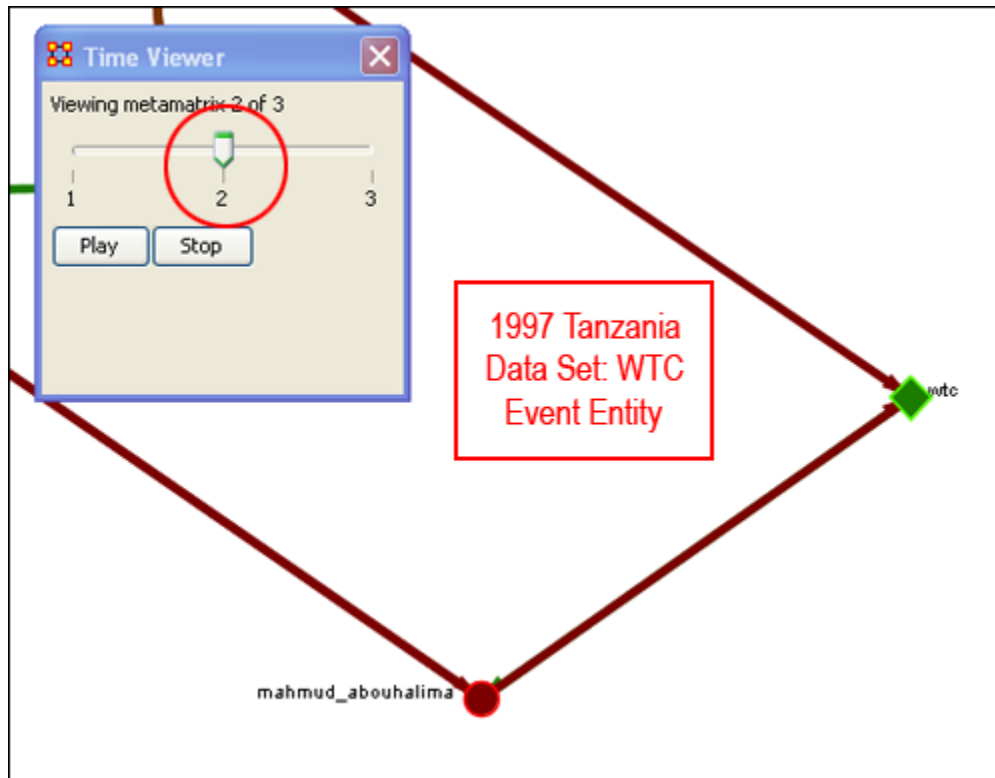
Notice that the connections (links) have changed from 1996 to 1997.
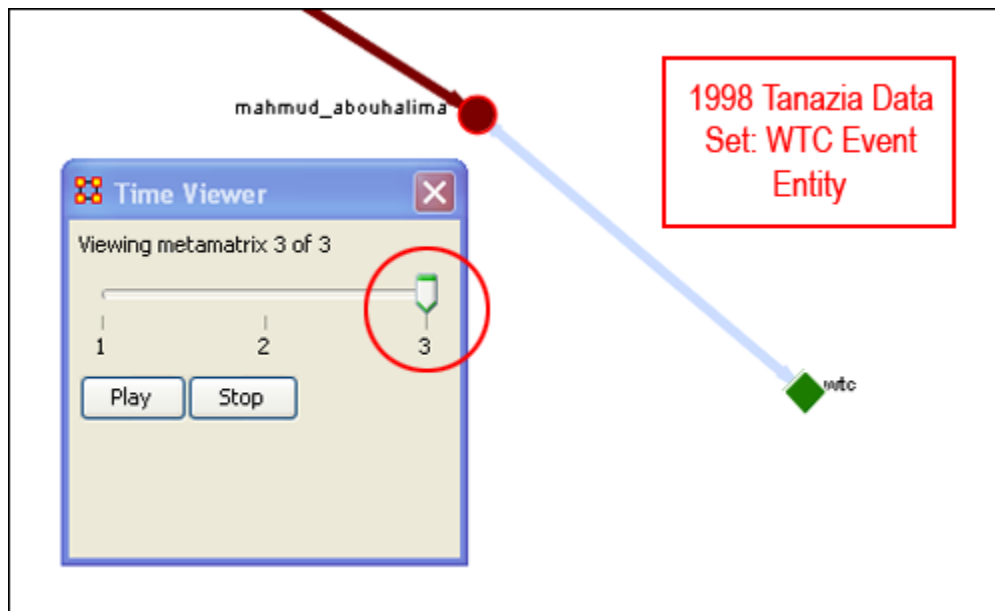
As you can see, as we progress through the data sets pertaining to **1996, 1997 and 1998**, the event node WTC lost direct **ties** to other parts of the network. In the 1996 data set, **WTC had 3 ties to other entities in the Tanzania Embassy network; In 1997, only 2; Finally, in 1998, the WTC event node had only 1 link to Tanzania Network**. This may indicate that the WTC node may have lost importance from one year to the next.

Summary of Lesson

At this point, you should be able to complete the following tasks to perform an Over-Time Network Analysis.

- Load multiple Time Data Captures in the form of Meta-Networks, which are representative of your network (Meta-Network) over time.

- Use the Over-Time slider window tool to move from one time capture to the next.

- Gain a visual perspective of how your overall network has changed from one time capture to the next.

Performing the View Measures Over-Time Analysis

ORA provides an alternative method to analyzing your network data **(in the form of Meta-Networks captured over time)** than the only in the Visualizer.

By running a **View Measures Over Time analysis** you can visualize how your network changes in relation to any given measure in the form of charts, which take you from one time point to the next.

Below is a series of screen captures and instruction on how to access the View Measures Over Time feature:

From the main menu bar: Data Visualization > View Measures Over Time
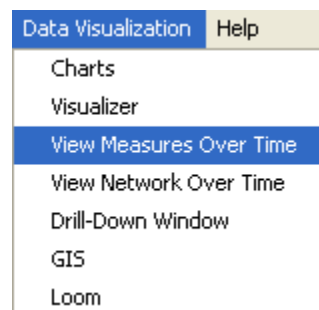
Select View Measures Over Time.
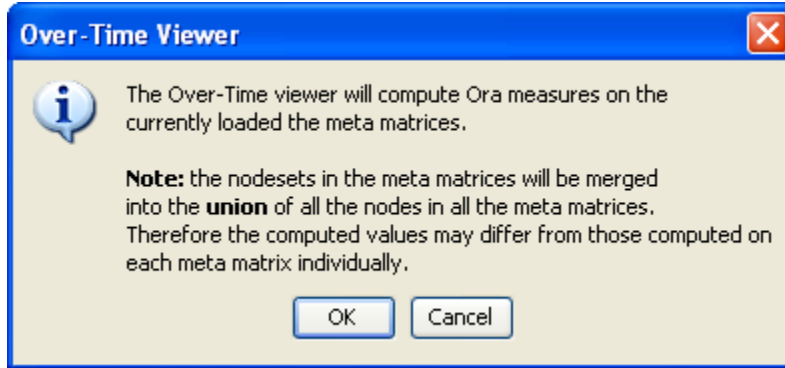
Select [OK]

The **View Measure Over Time** Window will appear. You will see that this window is primarily organized into three sections.

Area 1: is a slider where you can control various inputs that affect the aggregation level of your Meta-Networks.

Area 2: is a tabbed index allowing you to select the graph level you desire and the individual entity sets you are interested in analyzing.

Area 3: displays the actual chart based on the measures you selected in the tabbed index and the position of the aggregation slider.
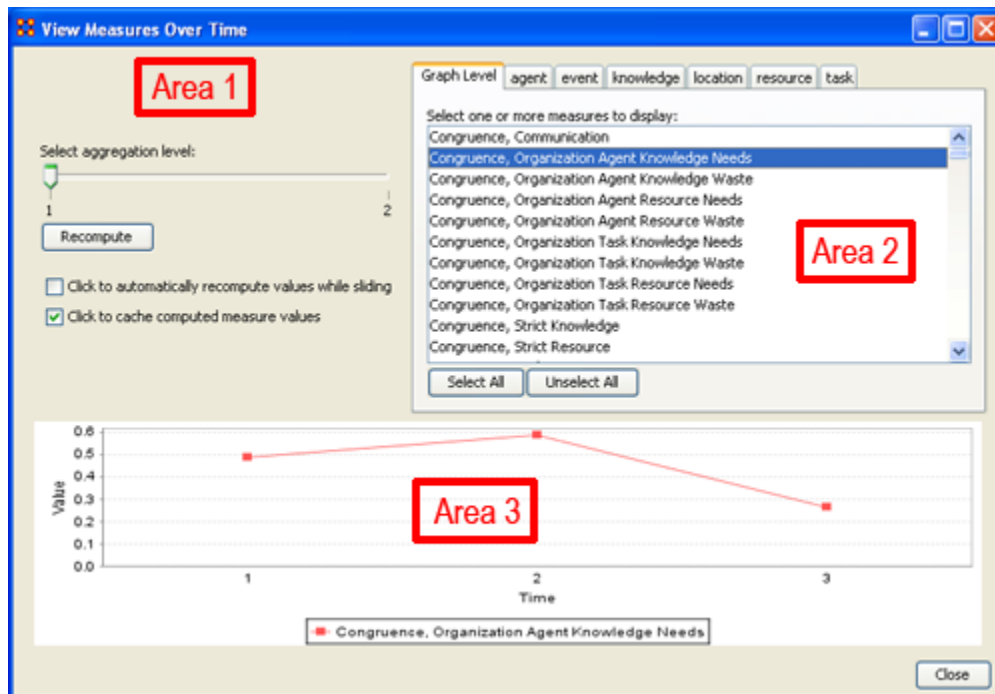
Interpreting The Results After Performing View Measures Over-Time Analysis

In the example above, we have loaded all three of our network time captures into the View Measures Over Time Analysis tool. The result of our analysis, as we have indicated before, will be displayed as a chart in the bottom section of the ORA interface in Panel 3. But, what does this tell us? To find out, we will examine the Graph Level measure **Knowledge Waste.** From the tabbed measures window:

Graph tab > Congruence, Organization Measure, Knowledge Waste.

The result is displayed below.
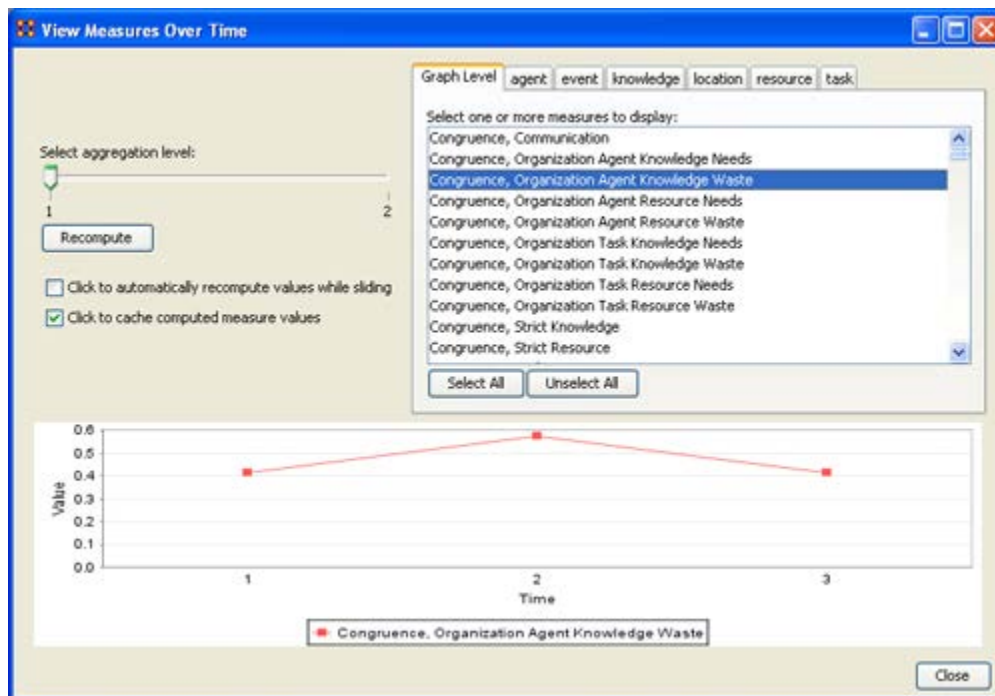
We can see the chart displays an axis scale correlating to our networks over time. The graph level measure of **Knowledge Waste** was selected using the **View Measures Over Time** tool and the result is displayed in the chart area of the window.

We can see that the graph level measure of Knowledge Waste held a value slightly above **40** in the first Meta-Network, 1996, displayed with **1** on the Time axis. We can deduce from the chart that in 1997 something affected our measure of Knowledge Waste to increase the value to nearly **60**. This corresponds to the time scale axis of **2**, which represents 1997, our second year of network data. In the final year, 1998, our measure of Knowledge Waste declined to the value in 1996. This should be visually evident with its relative line graph position.

It is clear that the value for the Knowledge Waste Measure increased from the first Meta-Network (**1996)** to the second **(1997)** and then retreated back to its original

(1996) level. It would be up to you to determine what may have affected that measure for it to undergo this progression and regression.

Summary of Lesson

At this point, you should be able to complete the following tasks to perform a View Measures Over-Time Analysis:

- Load multiple data files into ORA **(in the form of Meta-Networks)** representative of your network across certain periods of time.

- Select an aggregation level and choose a graph level or entity entity set to analyze.

- Analyze the data output in the charting tool located in the bottom section of the View Measures Over Time window.

- See how your network may have changed from one time capture to the next for any given graph level or entity node measure.

- Use this visual comprehension to draw conclusions about how your network changed across multiple time periods.

Working with Attributes

- **What is an attribute?**
- An **attribute** is a quality, an added description, a differentiation of some sort, of a Who, What, Where, How and When and Whatever. It also indicates the presences, absence or strength of a particular connection. As an example, you might create a network of people (Whos) who work at certain buildings (Locations). The **attributes** of those people could range anywhere from their first and last name, their employee identification number, their job grade in the corporation, their salary, their home phone number, education, political affiliation and the list goes on. Moreover, we might say that you have a connection to both your friend and father. However, we might give the connection to your father the attribute of **family** whereas your friend might be attributed **personal**. The key is to know that attributes can be both differentiate nodes or ties or both.
- It follows that it would be helpful to view a network of Whos (and or any of the other node classes) and see how those Whos are grouped according to any particular attribute you are interested in analyzing. ORA has the capability to do just that and provides several ways of doing so. We will explore several methods of locating groups namely by using ORA's report tool and in the Visualizer.
- **>> Go to Using attributes to find groups in ORA Reports**
- **>> Go to using attributes to find groups in the ORA Visualizer**

Using Attributes to find groups in ORA Reports

To **create and assess groups by attribute** use the **Locate Subgroups report** accessible in ORA Reports at the following path:

Main Menu Bar > Analysis > Generate Reports

1. Select the **Locate Subgroups report** (see screen shot below for more detail)

2. Select the **grouping algorithm "Attribute"**

3. Choose the **attribute of interest**

4. Click **Next**

5. Make sure to select Create block diagram at the top

By default a new **meta-network** is created that contains the networks used in the grouping.

In addition, there is a new node class that contains the group nodes, which in this case are just the **attribute value** nodes. Also two **block model group x group networks** are created.

To access this tool, follow the steps below:

Above is an image of the Locate Subgroups report tool in ORA.

The Locate Subgroups report tool is divided into three sections labeled "Step 1" "Step 2" and "Step 3".

In **"Step 1"** the tool can be applied to the entire Meta-Network by selecting the check box. Individual networks can be selected by scrolling down in the "Step 1" box and selecting the specific network level to which to apply the Locate Subgroups tool (i.e., agent x agent, agent x knowledge, et al.)

In **"Step 2"** a particular type of grouping algorithm can be selected (*Concor, Newman, Johnson, et al.*). Use the scroll bar to locate the grouping algorithm to select.

In **"Step 3"** You can select the specific grouping algorithm parameters for the grouping algorithm selected in "Step 2" (i.e., If the "Johnson" grouping algorithm was selected, use the panel tabs in "Step 3" to select "similarity" or "dissimilarity", number of groups to create, et al.).

After all required input is entered in Steps 1, 2, and 3, select "Next" to run the Locate Subgroups report.

Above is an image of the algorithm output box.

The algorithm output box allows you specify the type of algorithm output type you may need. You can choose from the following types of output:

- Group Membership Matrix

- Block Diagram

- Hierarchical clustering diagram

- Dendogram

- Attribute value distributions

Once you select the appropriate algorithm output type, select **[Next]** to choose report model format. You will have the following choices available:

- Text

- HTML

- CSV

Finally, choose a file path to save your report name the report accordingly without specifying an extension. You can use the file chooser or enter the path manually to specify a location to save the report.

You will also see that your locate subgroups output will have generated a Meta-Network, which will be loaded into ORA's Meta-Network Manager. You can save out this new Meta-Network, run measures, generate reports or apply any of ORA's functions and tools otherwise available in performing an analysis of a Meta-Network

Using Attributes to find groups in the Visualizer

What follows are instructions for finding various groups that may exist in your Meta-Network by using attributes and the ORA Visualizer too. To follow along with this lesson, please load the SG1 data set included with your installation.

Step 1: Open the Stargate Meta-Network.
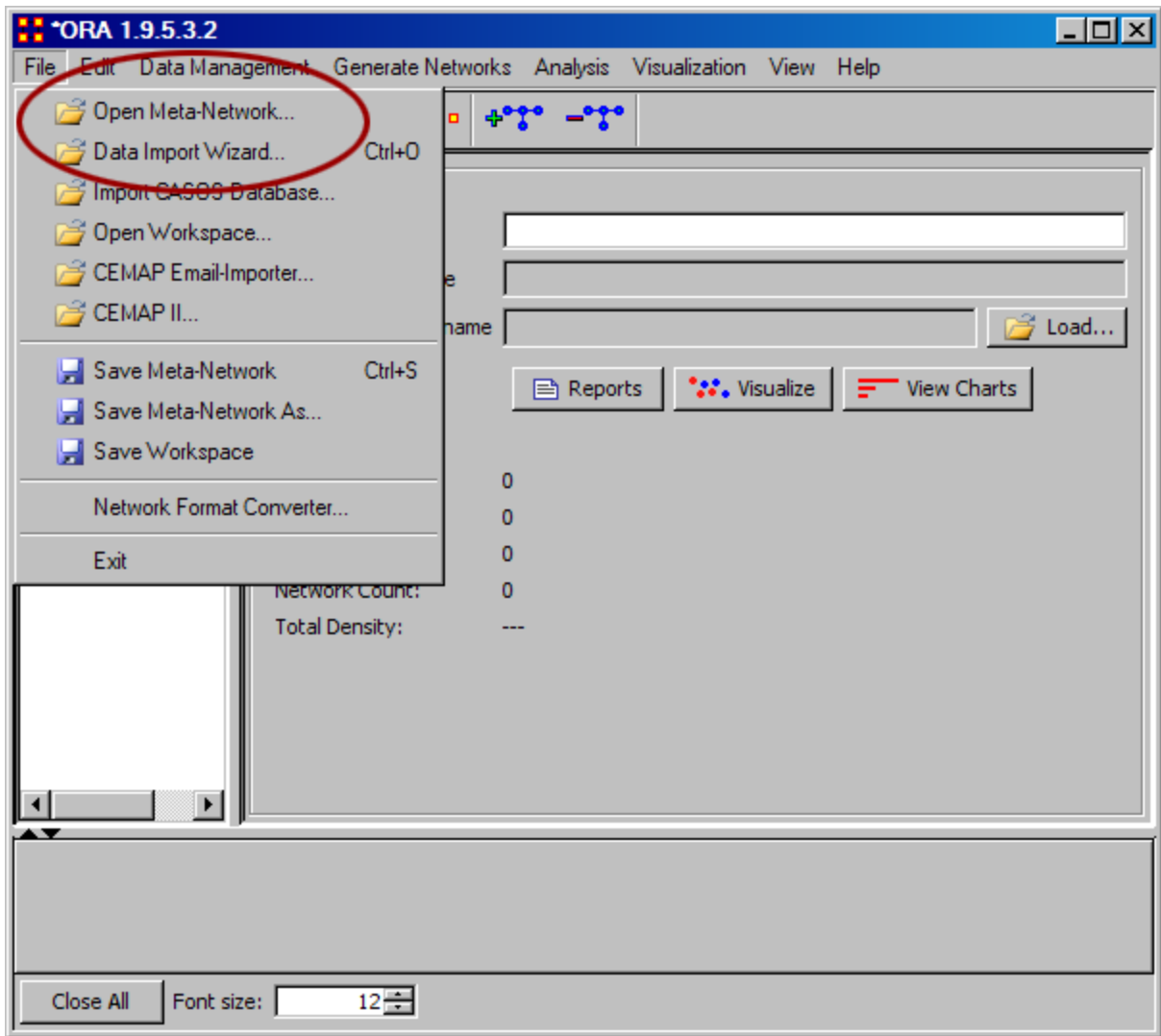
Step 2: Select [Visualizer] button.
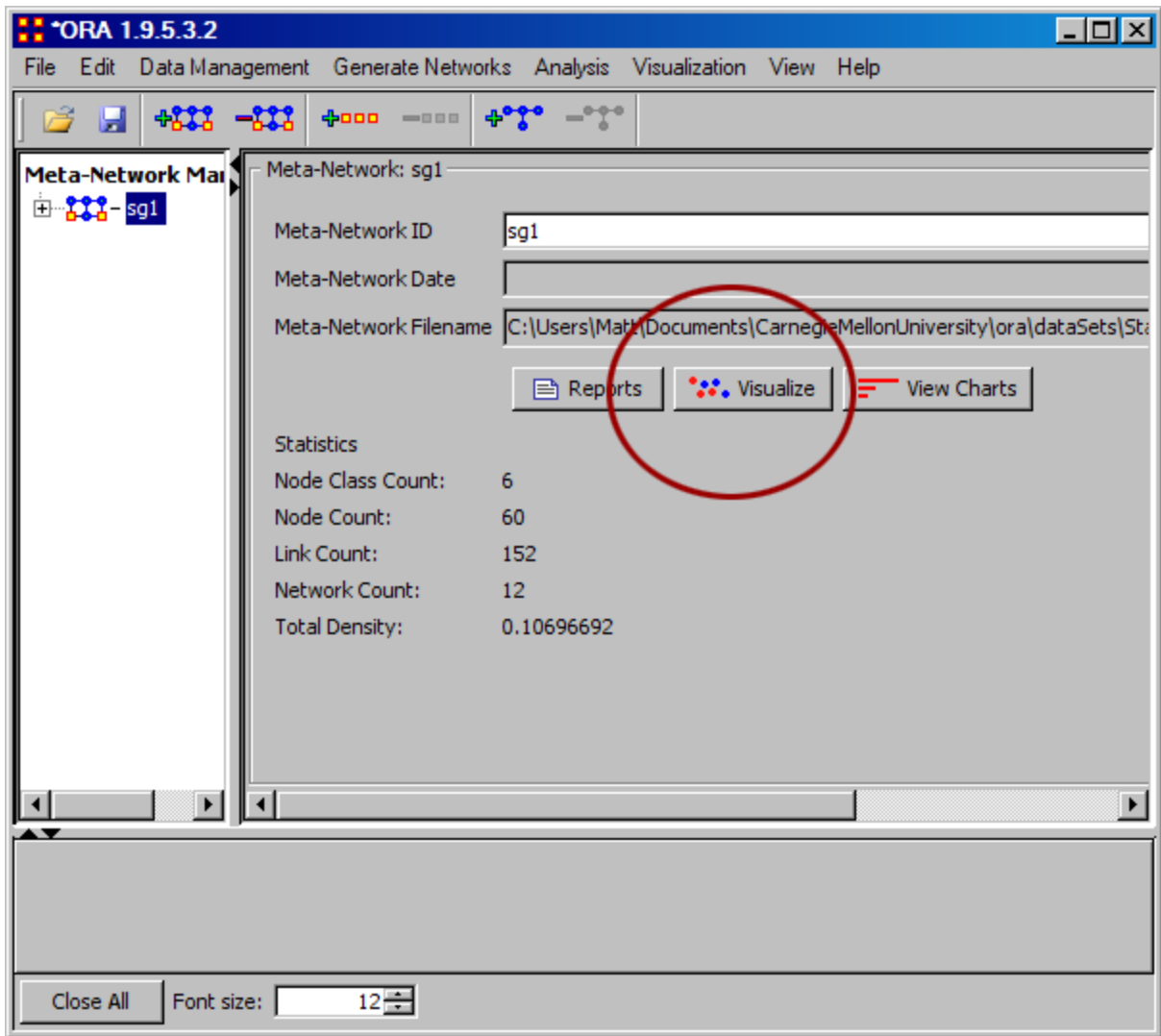
Step 3: In The Visualization, select "Attribute / Measure" Layout.
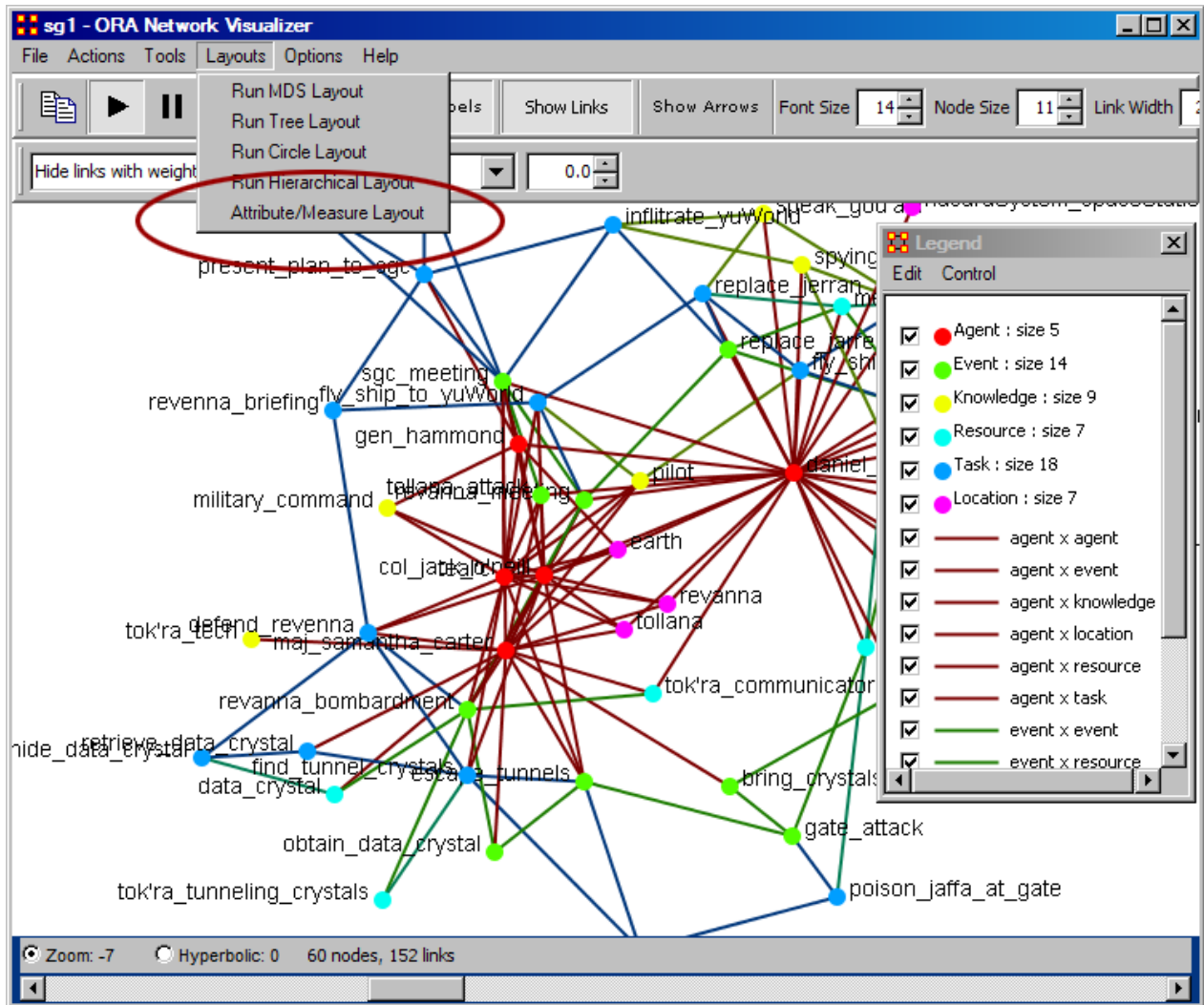
Step 4: Next (screen shot below) you will see the "Attribute / Measure" Layout Dialog box appear. In this box, select the attribute for an X and Y coordinate. You can use the save button to come back to these settings later. After choosing your attributes, click "Run Layout."
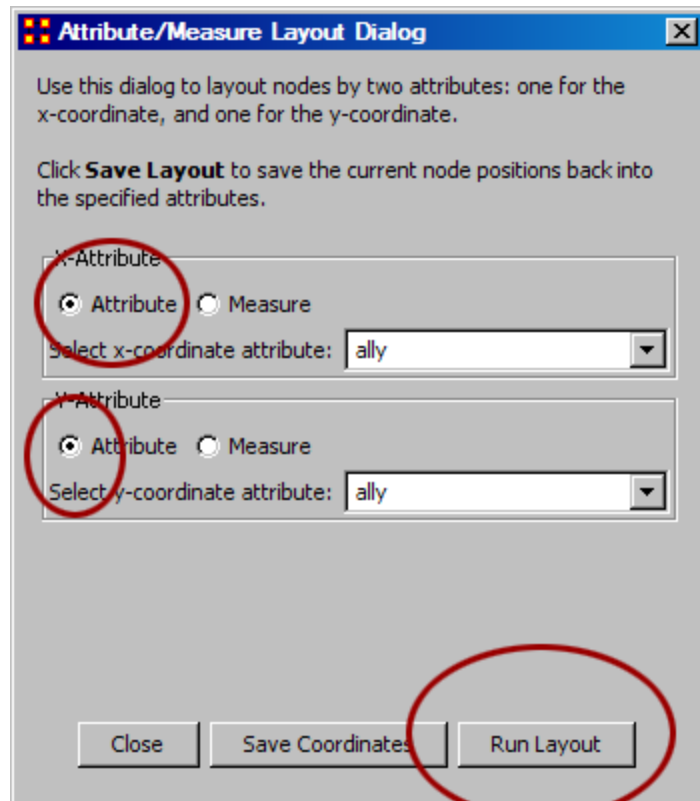
Step 5: You can experiment with the X and Y values by using either attributes or measures. In the final result shown below, we choose to use the X value as an attribute and the Y values as a measure.
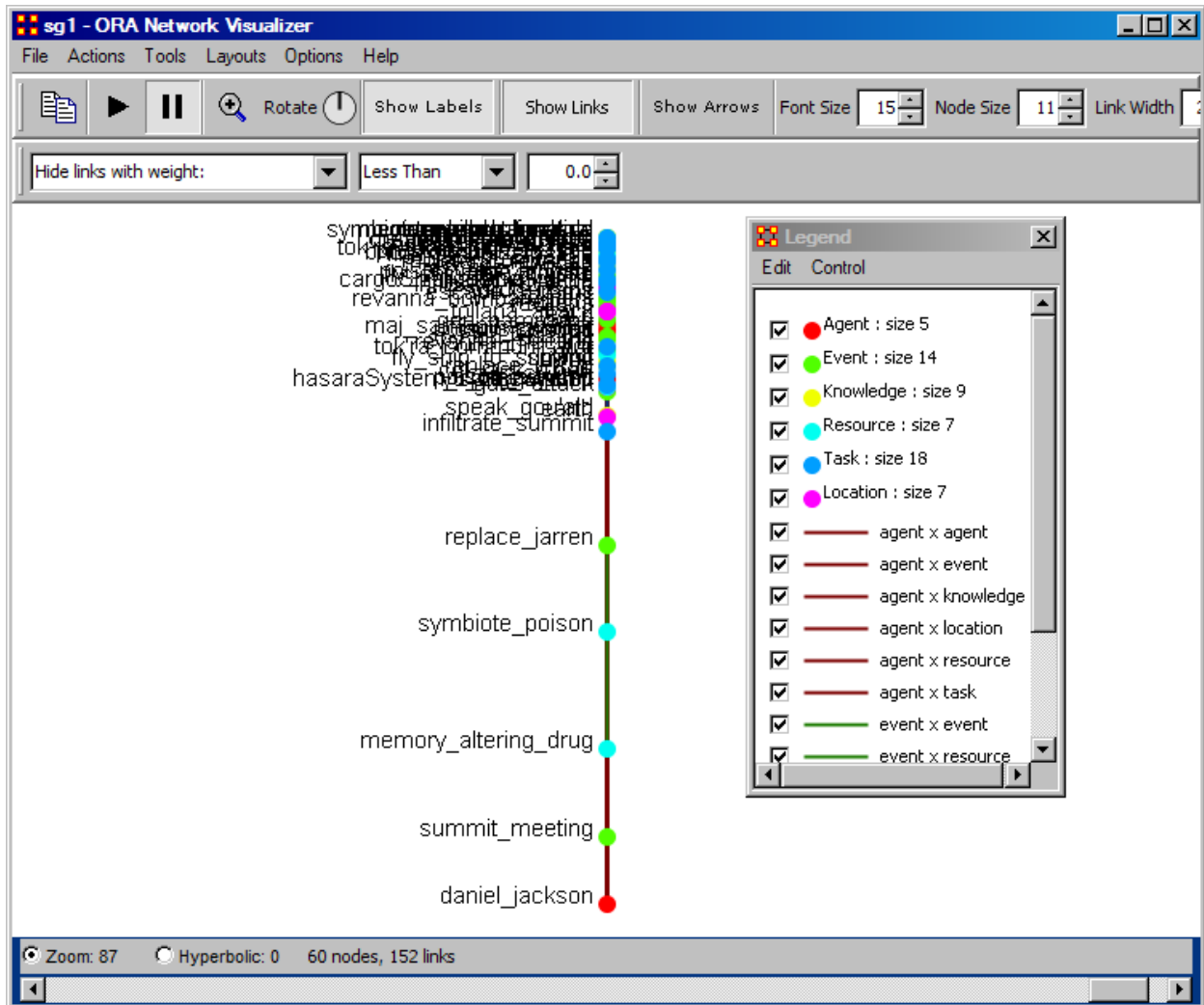
Working with Loom Clustering

Loom allows you to view nodes over time and which locations these nodes visited.

◦ 🔳 – Star Wars Episode IV: A New Hope, part 1	
◦ 🔳 – Star Wars Episode IV: A New Hope, part 2	
◦ 🔳 – Star Wars Episode IV: A New Hope, part 3	
◦ 🔳 – Star Wars Episode IV: A New Hope, part 4	
◦ 🔳 – Star Wars Episode IV: A New Hope, part 5	
◦ 🔳 – Star Wars Episode IV: A New Hope, part 6	
◦ 🔳 – Star Wars Episode IV: A New Hope, part 7	
◦ 🔳 – Star Wars Episode IV: A New Hope, part 8	
◦ 🔳 – Star Wars Episode IV: A New Hope, part 9	
◦ 🔳 – Star Wars Episode IV: A New Hope, part 10	
◦ 🔳 – Star Wars Episode IV: A New Hope, part 11	
◦ 🔳 – Star Wars Episode IV: A New Hope, part 12	
◦ 🔳 – Star Wars Episode IV: A New Hope, part 13	
◦ 🔳 – Star Wars Episode IV: A New Hope, part 14	
◦ 🔳 – Star Wars Episode IV: A New Hope, part 15	
◦ 🔳 – Star Wars Episode IV: A New Hope, part 16	
◦ 🔳 – Star Wars Episode IV: A New Hope, part 17	
◦ 🔳 – Star Wars Episode IV: A New Hope, part 18	
◦ 🔳 – Star Wars Episode IV: A New Hope, part 19	
◦ 🔳 – Star Wars Episode IV: A New Hope, part 20	
◦ 🔳 – Star Wars Episode IV: A New Hope, part 21	
◦ 🔳 – Star Wars Episode IV: A New Hope, part 22	
◦ 🔳 – Star Wars Episode IV: A New Hope, part 23	
◦ 🔳 – Star Wars Episode IV: A New Hope, part 24	

Loading the Star Wars dataset

For this example we will be using the Star Wars dataset included in the sample folder. You can find it in the following folder: **ORA > sample data > Loom data > Star Wars**. There will be 24 xml files. From the main menu select **File > Open Meta-Network...**. Navigate to the folder specified above. Select the first file then Shift-Click the last file to select them all. Then select [**Open**]. This will display them in the Meta-Network panel.

Trail pattern length determines which clusters will be shown.

To open this data in Loom make sure all Meta-Networks are selected. From the drop down menu select **Visualizations > View Trails**. You will be asked to chose a network. For this dataset chose **Characters x Location**. On the left will be two panes. On top is a list of Characters and below that is a list of locations.

After opening the dataset in Loom all the characters will appear in the left/top pane and will start with no nodes checked. Below are the locations and the colors which will signify them on the Loom.
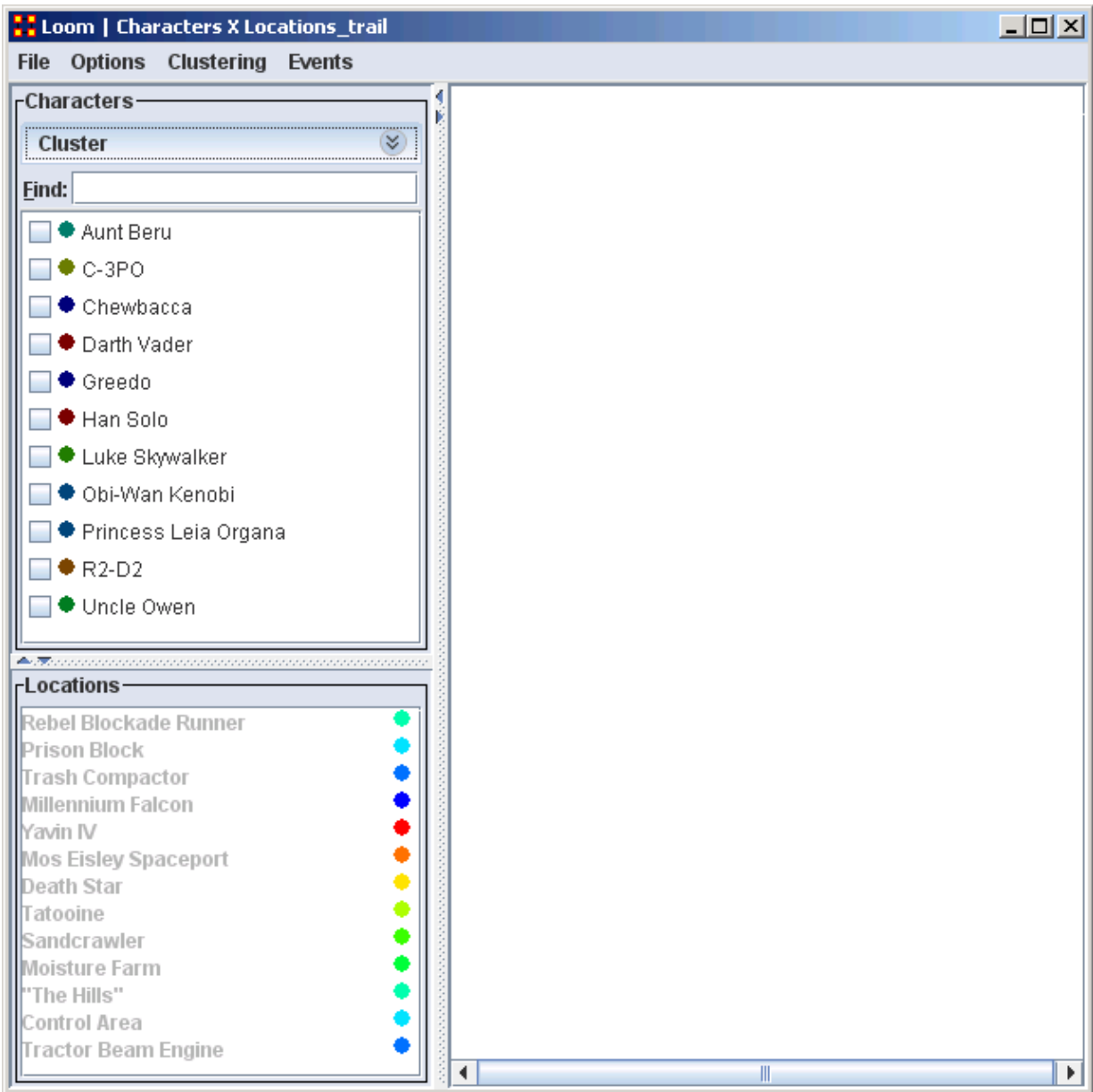
Trail pattern length

Trail pattern length displays clusters of nodes which have the same general trail. The number is the minimum number of points they all need to have in common.

Uncle Owen and Aunt Beru are in [4] timeslices. They will continue to be clustered until the **Trail pattern length** reaches [5]. At that point they no longer meet the conditions and are no longer clustered.
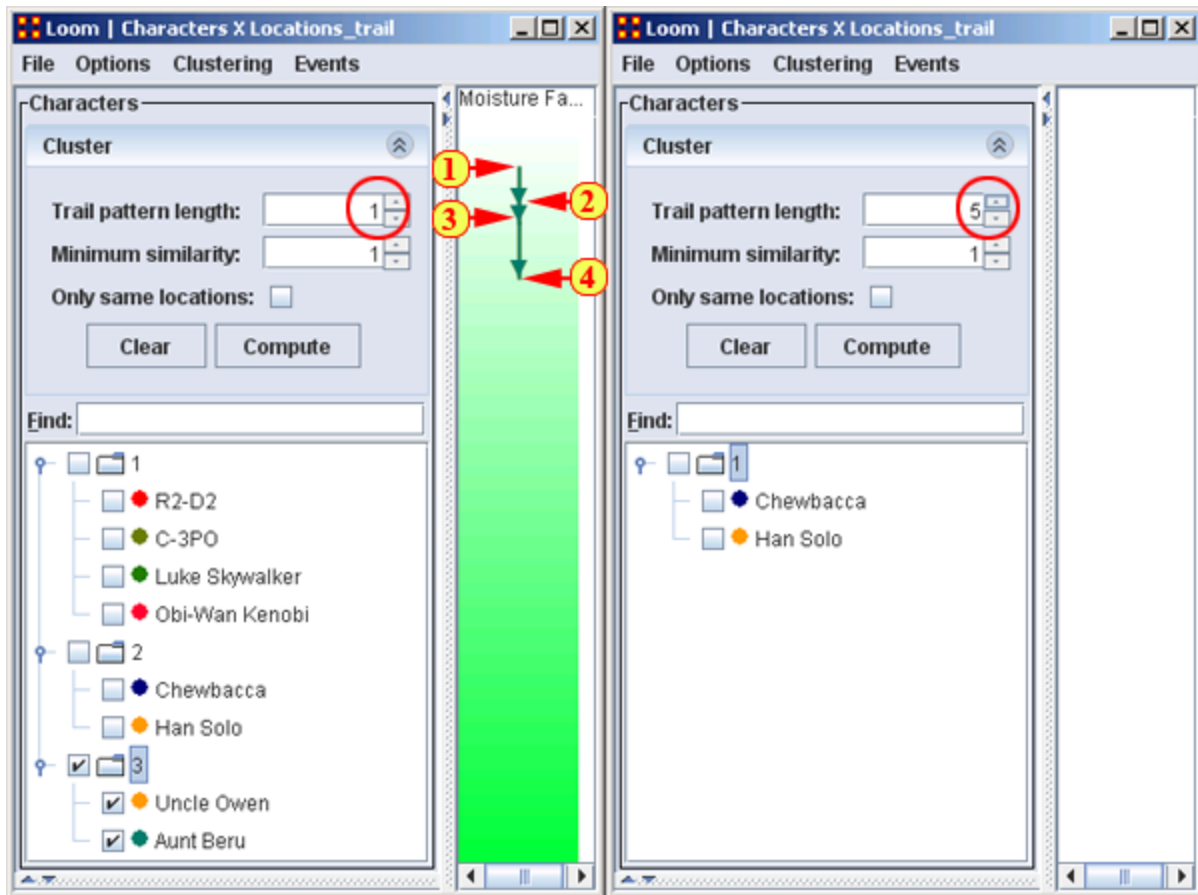
Minimum Similarity
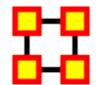
This is the minimum number of [continuous] points that nodes need to have in common to be clustered. Whenever a node falls under this number it is dropped from the cluster.

Here **R2-D2**, **C-3PO**, **Luke Skywalker**, and **Obi-Wan Kenobi** are clustered. When the Minimum Similarity reaches four Obi-Wan drops out of the cluster when he leaves for a different location.
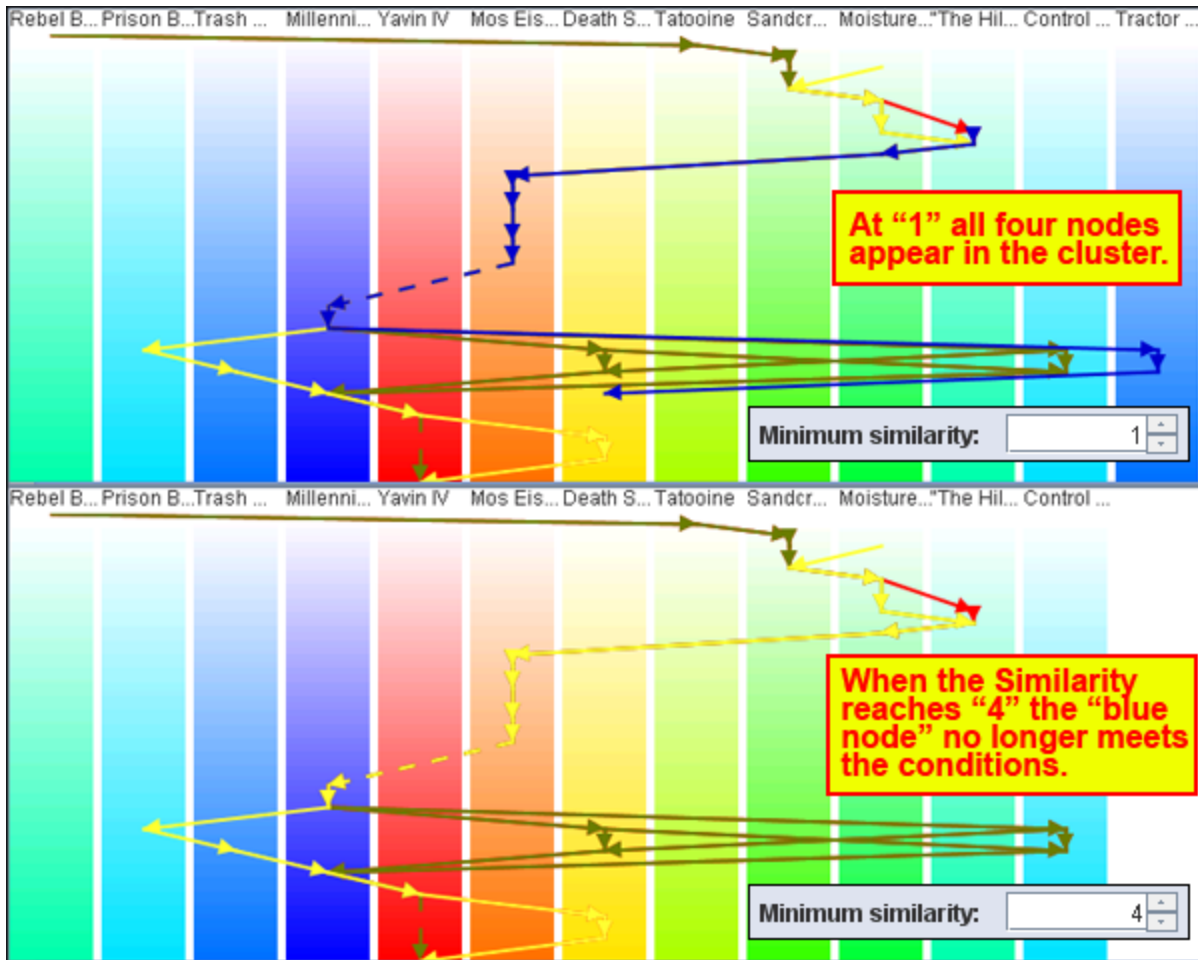
Only Same Location

With this check box checked only nodes with **all the same locations** will be clustered. The Star Wars dataset has two clusters where nodes were in all the same locations for all timeslices **Uncle Owen and Aunt Beru** and **Chewbacca and Han Solo**. The display below shows the route of **Han Solo and Chewbacca**.
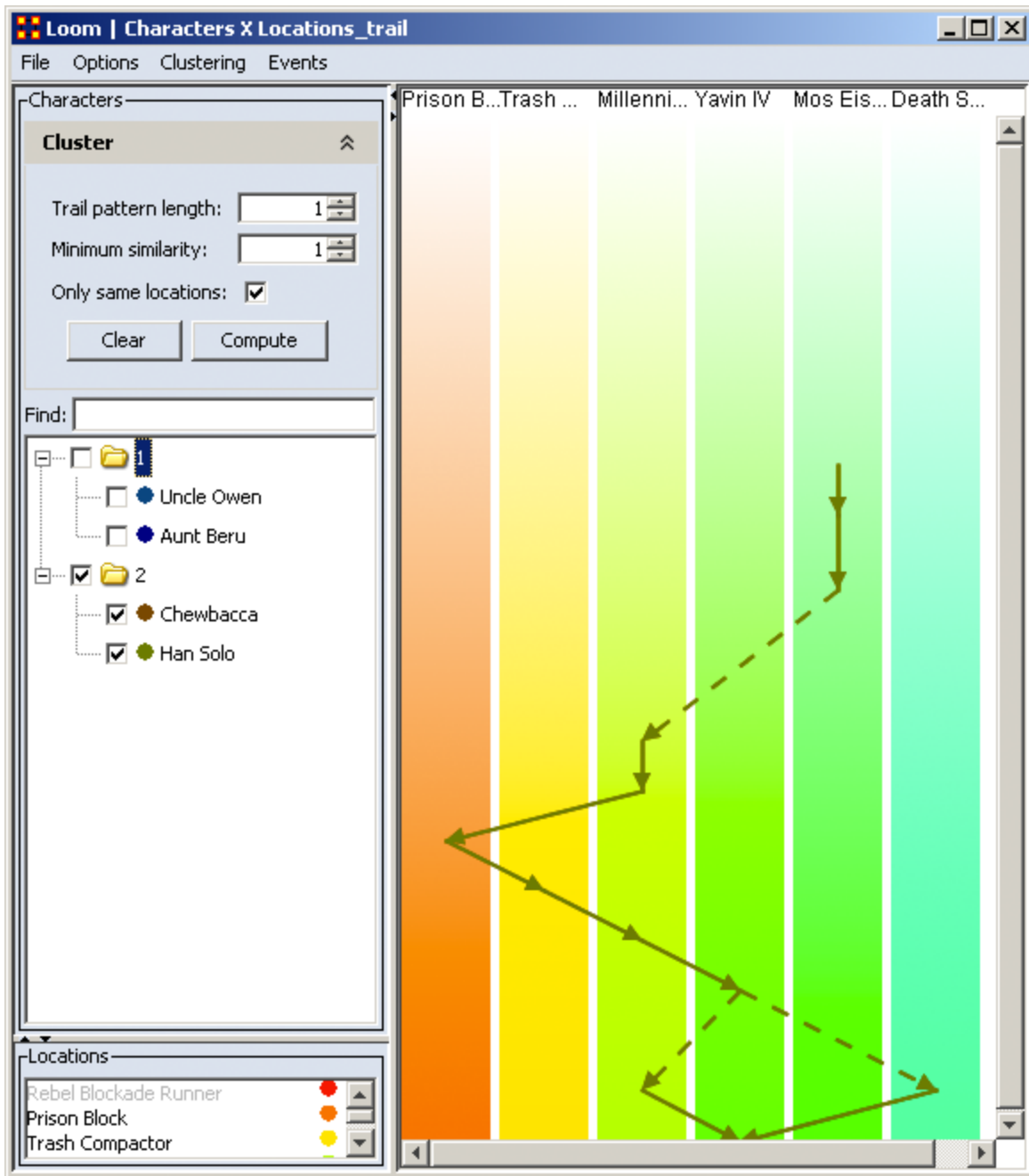
Change Detection (Over-Time Viewer)

Description

The **View Measures Over-Time** is used to find out how quickly you can identify changes in a network to exploit subtle organization change, And also to facilitate improved command and control subject to a specifid risk of false alarm. Various

measures of a network are taken at different points in time. You need to identify a) that a change **did** occur and b) **when** did that change occur.

NOTE : *It is important to note that this is not predicting change. But rather it is detecting that a change has occurred quickly and be able to make some inference about the actual time of change.*

Example of Change Detection

Change Detection may offer you a tool to operate inside the normal decision cycle. The following figure represents some measure of interest over time. It could be the revenue of a company, the combat power of an enemy, or a measure of interest from a social network. When do you conclude from this measure that a change may have occurred? Let's assume that by conventional methods you can detect a change in organizational behavior as of **today**, the vertical line. This time point might be too late to take preventative or mitigating action. In other words, this could be the point of inevitable bankruptcy for the company, or the successful culmination of a terrorist attack. Identifying that a change occurred by **time period E** might allow the analyst to respond to the change before it is too late; get inside the decision cycle.
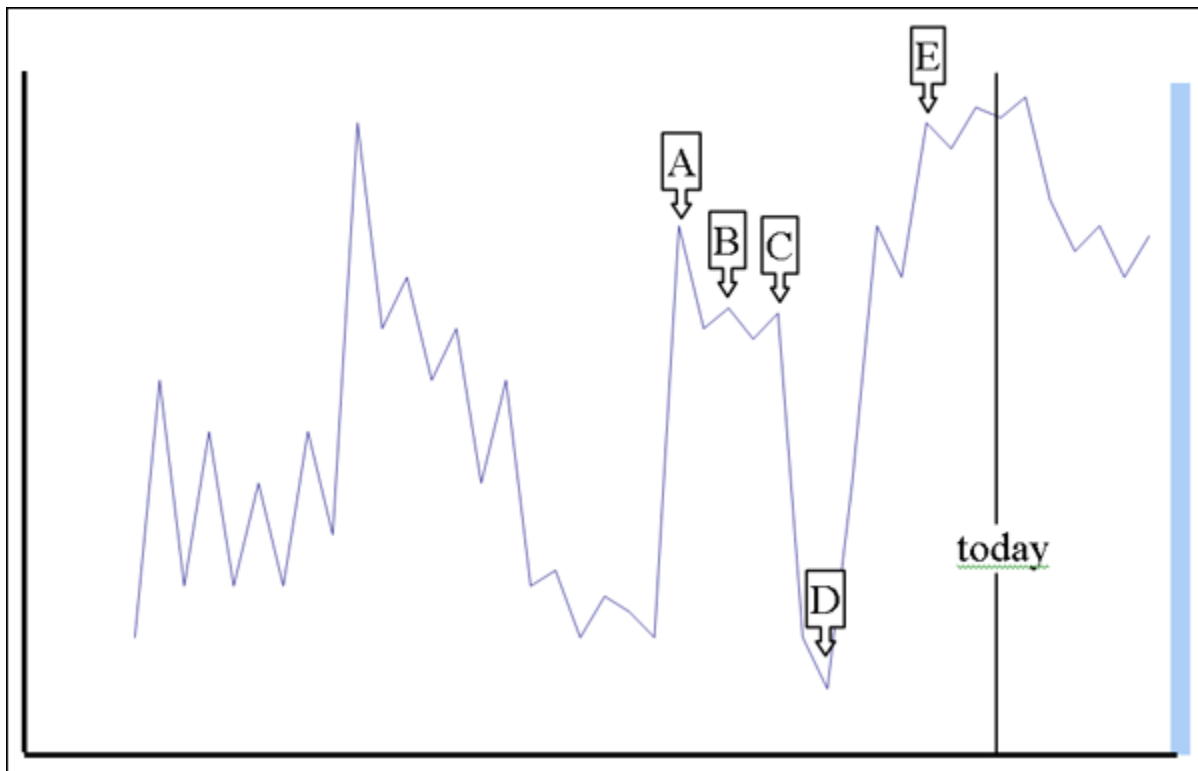
Change detection is more challenging than it may seem at first. The **sudden change in the measure between time D and time E**, however, may look very similar to the **peak at time A**. Furthermore, if you assert that a change in fact occurs at time A, there may exist a large amount of time periods to investigate for

the cause of any change. If you can identify more likely points in time where change may have occurred, you can reduce the costs in terms of time and resources to search for the potential causes of change. Identifying the likely time that a change may have occurred is called **Change Point Identification**.

Another problem you face is detecting the change as quickly as possible after the change occurred. Can you improve the ability to get inside the decision cycle by detecting the change at time D, or even better at time B? **This is called Change Detection**.

NOTE : *It is important that you use at least 10 time periods for an accurate simulation.*

The Three Procedures

The Shewhart X-bar Control Chart

This is simple and easy to implement and provides a Proof of Concept. If calculates a samle average of observations and measures of a process. The decision interval indicates whether the observation is in the tail of the disbtribution.
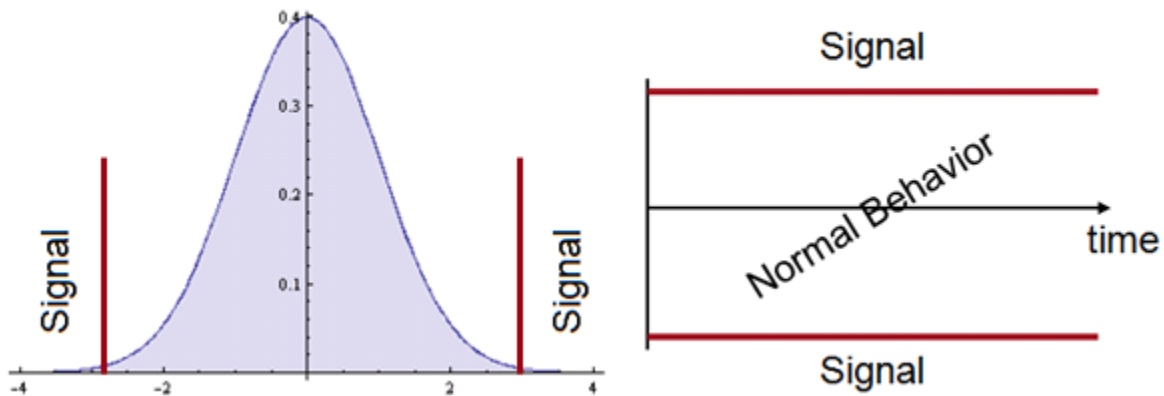
Cumulative Sum (CUSUM)

Other statistical schemes increase power (faster detection, given the same probability for false alarm). The CUSUM considers previous values of the observed process. It can take a long time to detect small changes. Intuitively, we can see change earlier. How to detect it?
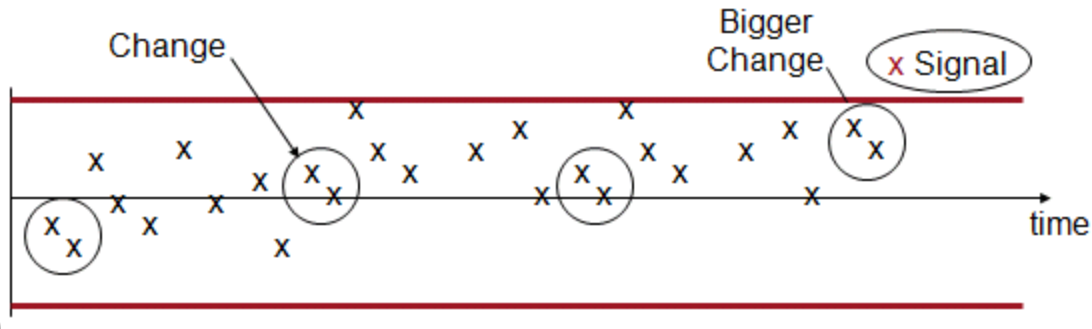
Cumulative Sum is good at detecting small changes in mean over time and has good buildt-in change point detection.

The **Load data and use CUSUM option**. CUSUM has an additional parameter. This is equivalent to ? in the equation presented earlier. It is an optimality parameter for the CUSUM approach that will make this approach the uniformly most powerful for a change in the social network measure of ?. As a rule of thumb this is usually set to 1.

There are three ways to set the risk of a False Alarm.

1. The user can specify the decision interval.

2. The user can specify a probability as in the Shewhart approach.

3. The user can specify the average number of in-control networks expected before a false alarm.

This is due to the nested conditional probability inherent in the CUSUM.

The CUSUM signals a change, when the statistic exceeds the decision interval. The most likely point in time that the change actually occurred is interpreted as the point that the CUSUM was last equal to 0. In this case, the time period is 19, which is one time point before the actual change. This is where an analyst would begin to look for explanations of change in the group. There are two separate charts plotted. One is for increases (red) in the social network measure, and the other is for decreases (blue) in the measure.

Exponetially Weighted Moving Average

This chart requires two items.

- A definition of what normal behavior is.

- What is the risk of false alarm. (This is a trade-off between speed and mistakes).

Change Detection GUI
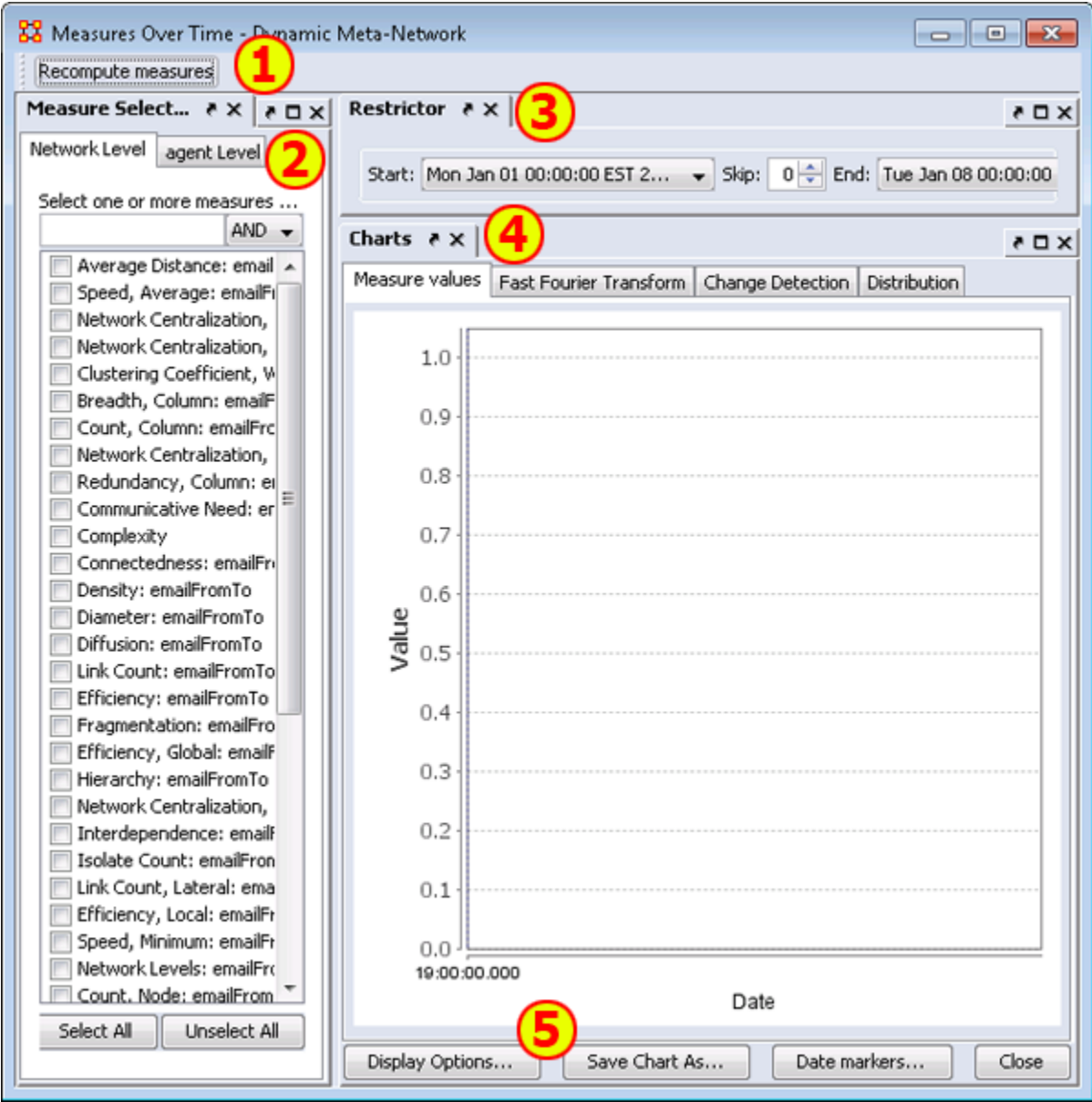
1) Recompute measures button :

2) Measure Selector pane : In the left pane you can choose the particular measures that you want to analyze in either **Network or agent level**.
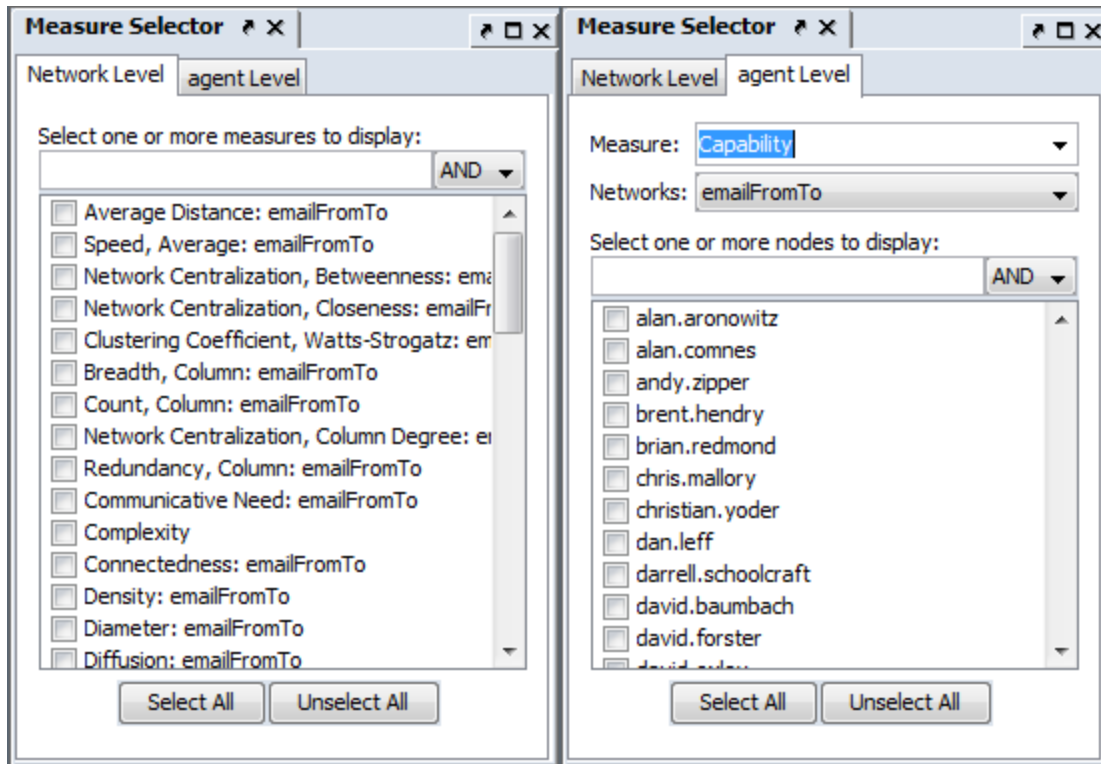
3) Restrictor pane : Allows you to set the dates which the Over-Time Viewer will work and also if you want to set a **skip** interval.

4) Chart pane : Area for displaying the final results. This area contains four tabs:

Measures Values :
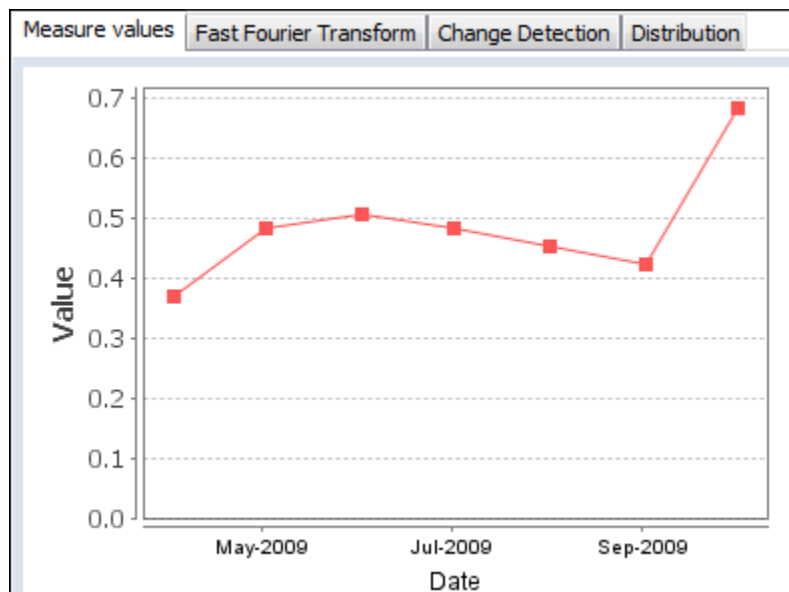
Fast Fourier :
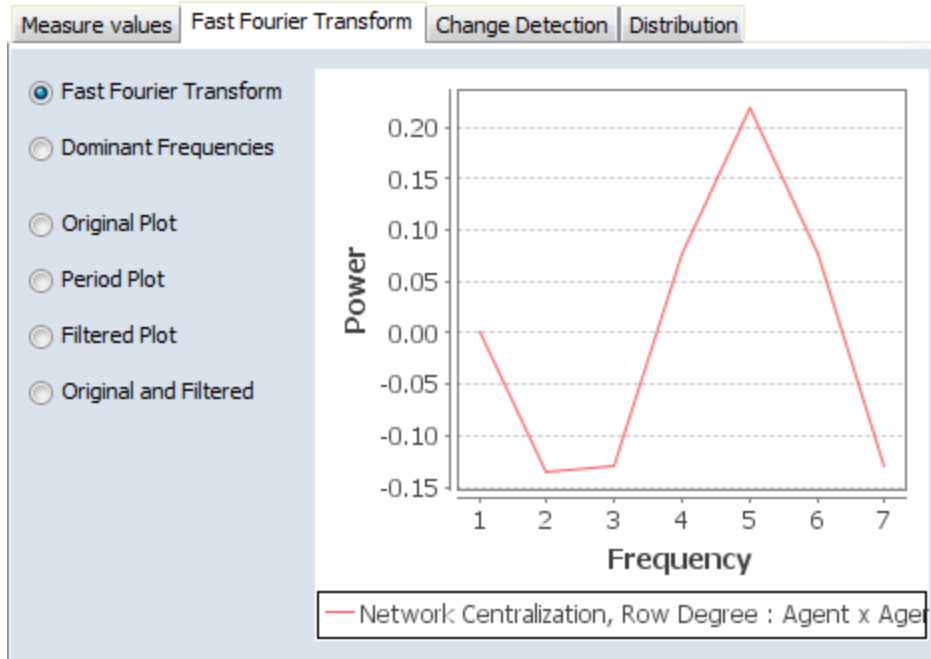
Change Detection : Used to quickly determine **that** a changed has occurred. Find the **change point** to detemine the exact date.

Distribution :

	Network C...
2009/03/0...	
2009/04/0...	
2009/05/0...	
2009/06/0...	
2009/07/0...	
2009/08/0...	
2009/09/0...	
2009/10/0...	
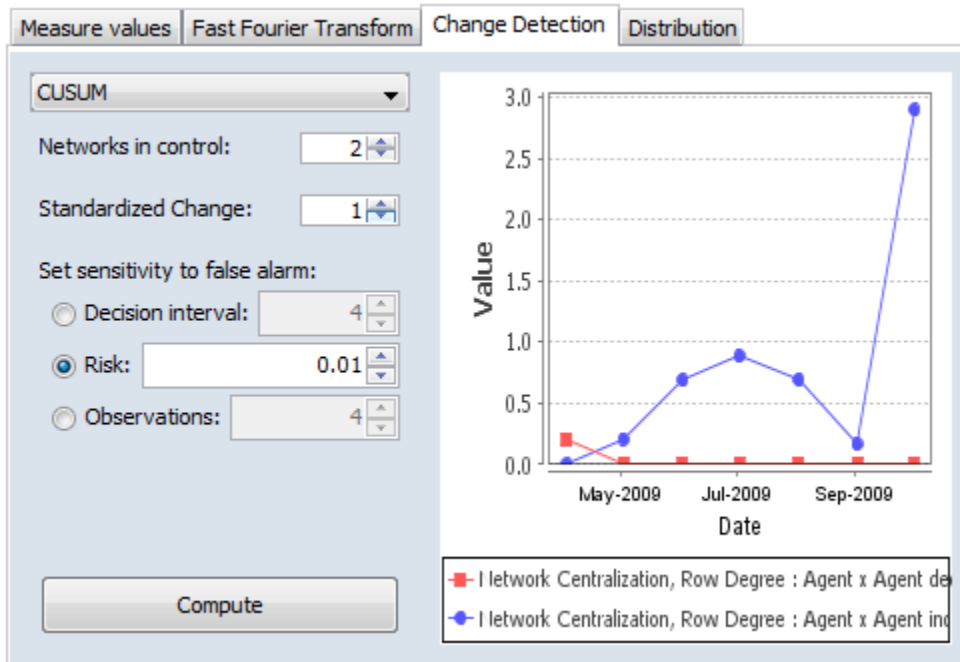
Tabs: Measure values | Fast Fourier Transform | Change Detection | **Distribution**
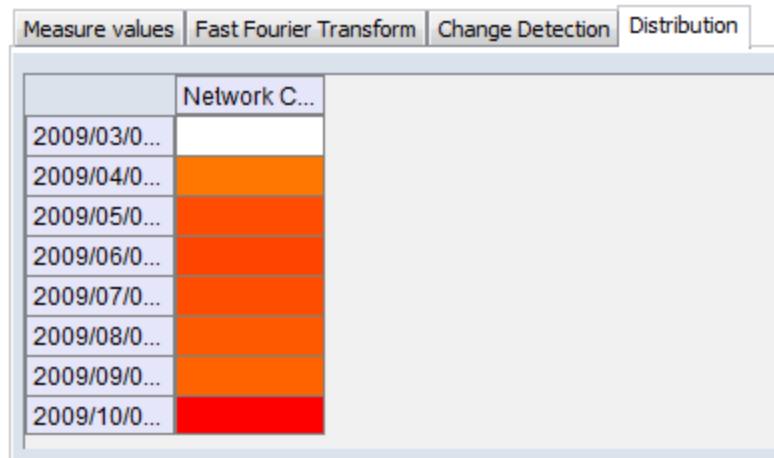
5) Function Buttons :

- **Display Options :**

- **Save Chart As... :**

- **Date markers... :**

- **Close :**
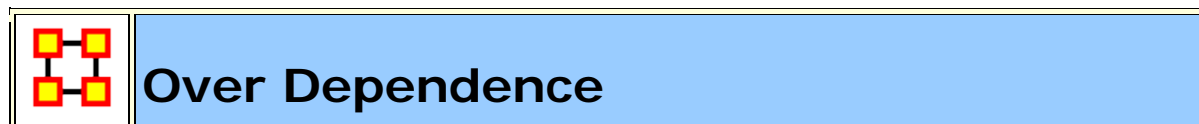
Over Dependence

Over Time Dependence

One major obstacle to the study of network dynamics is **periodicity or over-time dependence** in longitudinal network data. For example, if we define a social network link as an agent sending an email to another, we have continuous time stamped data. Intuitively, we can imagine that individuals are more likely to email each other at certain times of the day, days of the week, etc. If the individuals in the network are students, then their email traffic might follow the school's academic calendar. Seasonal trends in data are common in a variety of other applications as well. When these periodic changes occur in the relationships that define social network links, social network change detection methods are more likely to signal a false positive.

A False Positive : *This occurs when the social network change detection method indicates that a change in the network may have occurred, when in fact there has been no change.*

To illustrate, assume that you are monitoring the density of the network for change in hourly intervals. The density of the network measured for the interval between 3 A.M. and 4 A.M. might be significantly less than the network measured from 3 P.M. to 4 P.M. because most of the people in the network are

asleep and not communicating between 3 A.M. and 4 A.M. This behavior is to be expected, however, and it is not desireable for the change detection algorithm to signal a potential change at this point. Rather, it would be ideal to control for this phenomenon by accounting for the time periodicity in the density measure. Only then can real change be identified quickly in a background of noise.

Periodicity can occur in many kinds of longitudinal data. Organizations may experience periodicity as a result of scheduled events, such as a weekly meeting or monthly social event. Social networks collected on college students are likely to have periodicity driven by both the semester schedule and academic year. Even the weather may introduce periodicity in social network data, as people are more or less likely to email, or interact face-to-face.

Spectral Analysis provides a framework to understand periodicity. Spectral analysis is mathematical tool used to analyze functions or signals in the frequency domain as opposed to the time domain. If we look at some measure of a social group over time, we are conducting analysis in the time domain. The frequency domain allows us to investigate how much of the given measure lies within each frequency band over a range of frequencies.

For example, the figure below shows a notional measure on some made-up group in the time domain. It can be seen that the measure is larger at points B and D corresponding to the middle of the week. The measure is smaller at points A, C, and E.
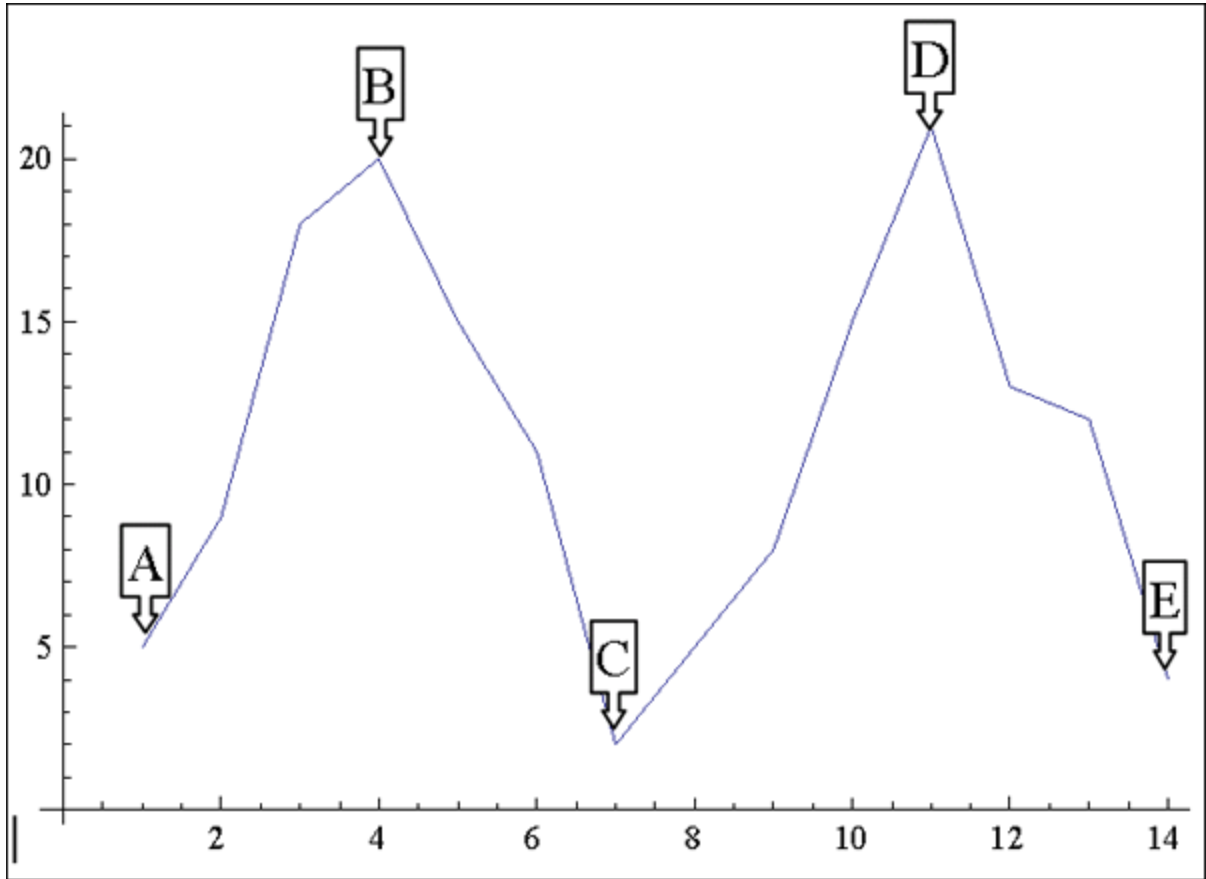
If that signal is converted to the frequency domain as shown in the figure bwlow, you can see how much of the measure lies within certain frequency bands. The negative spike corresponds to 7 days, which is the weekly periodicity in the notional signal. The actual frequency signal only runs to a value of 8 on the x-axis. The frequency domain signal after a value of 8 is a mirror image, or harmonic of the actual frequency signal.
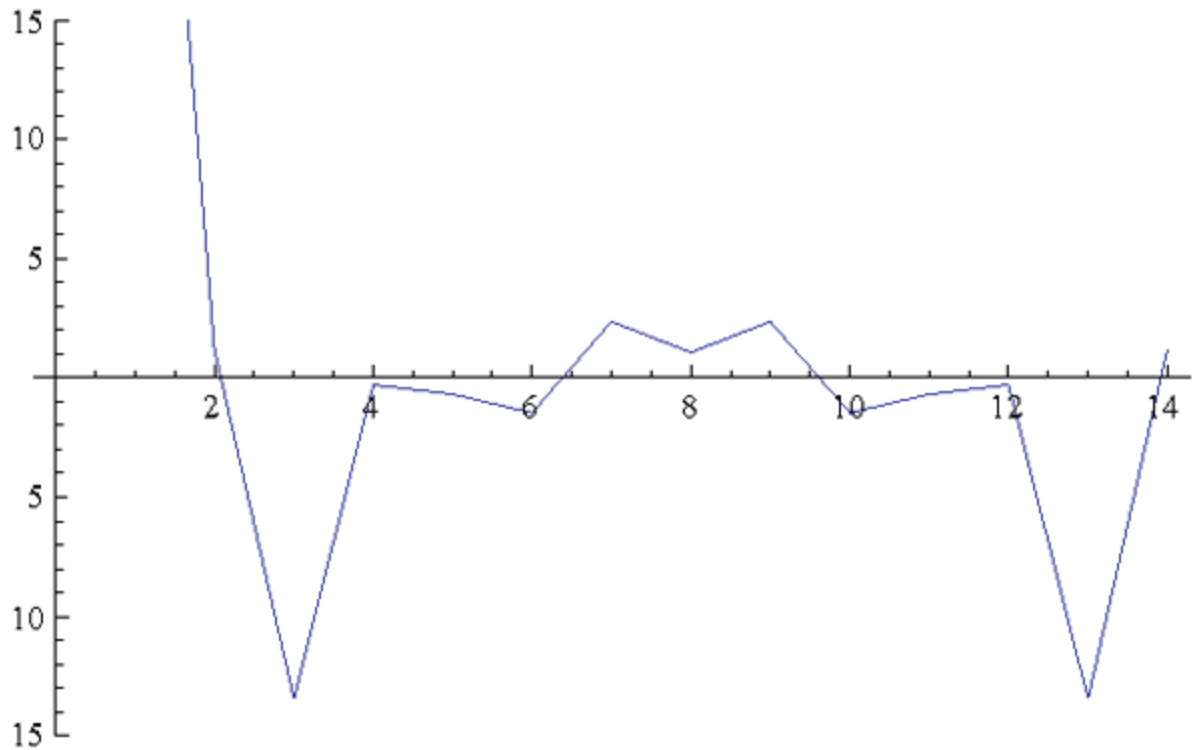
The frequency domain representation of a signal also includes the phase shift that must be applied to a summation of sine functions to reconstruct the original over-time signal. In other words, we can combine daily, weekly, monthly, semester, and annual periodicity to recover the expected signal over-time due to periodicity. For example, the next figures represent monthly, weekly, and sub-weekly periodicities. If these signals are added together, meaning that the observed social network exhibits all three of these periodic behaviors, the resulting signal is shown in **Sum of the Signals**.

Monthly Period · Weekly Period · Sub-Weekly Period

Sum of the Signals
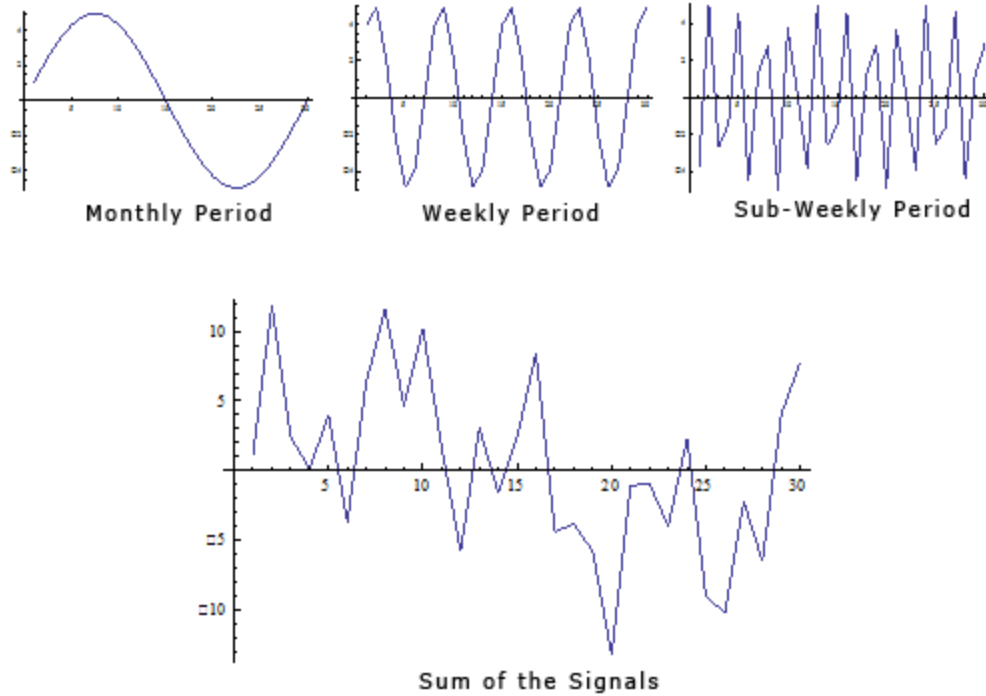
If the periodicity in the signal is not accounted for, it appears that there may be a change in behavior around time period 20, where the signal is negatively spiked. In reality, this behavior is caused by periodicity. If you transform the signal to the frequency domain you can see the weekly periodicity at point B and the sub-weekly periodicity at point A.
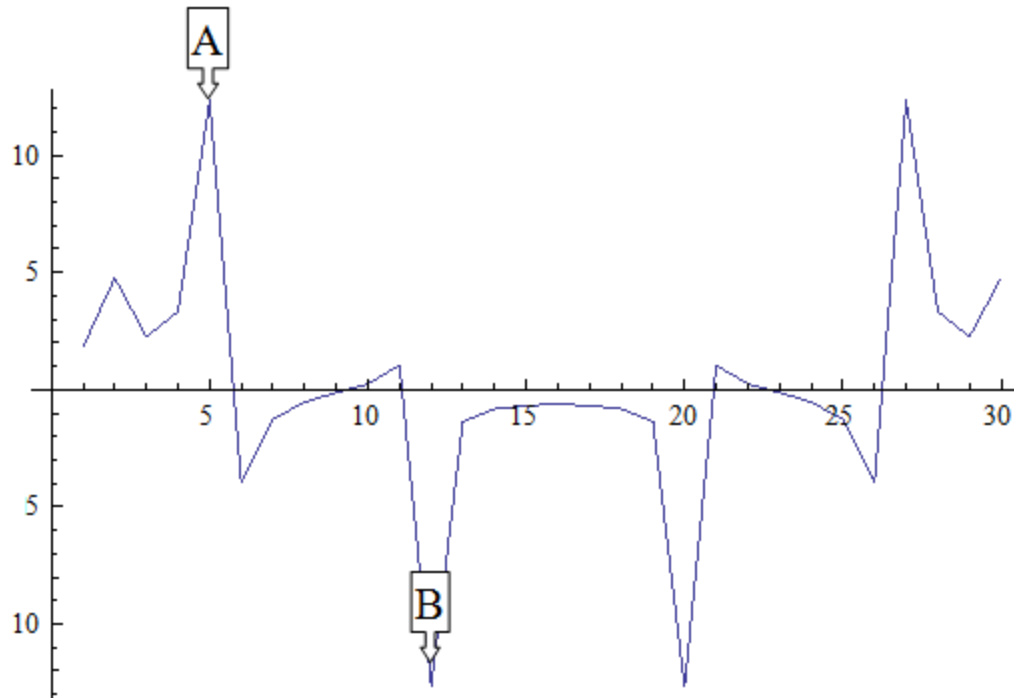
Transformation to Frequency Domain

Spectral Analysis applied to social network measures over time will identify periodicity in the network. The next procedure will transform an over time network measure from the time domain to the frequency domain using a Fourier transform. Then the significant periodicity will be identified in the over-time network and present two methods for handling the periodicity.

The over time dependence analysis is accessed by selecting the **Fast Fourier Transform** tab in the Over Time Viewer. This displays the frequency plot of the data.

You can use the Over Time Viewer to help determine which frequencies are significant.

The [**Dominant Frequencies radio button**], displays only the statistically significant frequencies.

The [**Fourier Transform radio button** uses the normal distribution in order to transform data from the time domain to the frequency domain. Therefore, the normal distribution is an appropriate distribution to fit to the frequencies plotted in the frequency plot. All frequencies that are within two standard deviations of the mean are then set equal to zero for the dominant frequency plot, revealing only the dominant frequencies.

NOTE : *A dominant frequency is a potential source of periodicity, as opposed to random noise in the over time signal.*

The analyst will often want to transform the statistically significant frequencies from the frequency domain back into the time domain so that he/she can make better sense of them. To do this, the analyst must select the radio button on the lower left called Period Plot.

The [**Period Plot radio button**] shows the analyst the expected periodicity in the over-time data. In the example, you can see weekly periodicity. The peaks and valleys in the period plot occur approximately every 7 days. At this point, the analyst may wish to merge the daily data into weekly networks. This would average out the effects of weekends and evenings that are likely to affect the properties of daily networks. Another approach is to simply look at the networks departure from what is expected.

The [**Filtered Plot radio button**] will create an over-time plot of how the measure deviates from what is expected, based on the periodicity of the measure. You can also plot the filtered measure with the original measure to see the difference as shown below.

Procedure

Your Dataset
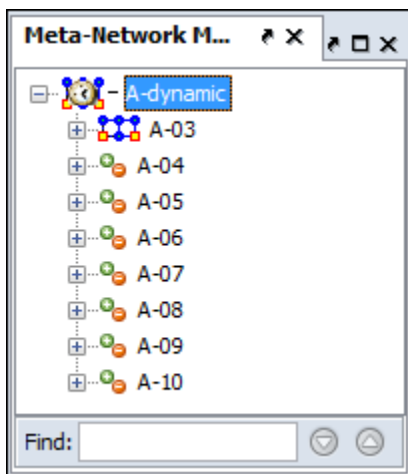
Load your dataset which you need to run the Change Detection on into ORA. This should be a dataset with multiple time periods. Measures will be run on all the time periods and this will determine **IF** a change has happened.

Highlight the Meta-Network.

Start Measures Over Time Function

From the Editor's main menu select **Visualizations > View Measures Over Time...** to start the process. In the dialog box select **Centrality measures** radio button. The measures for Change Detection you are mostly concerned

75

about are the Closeness and Betweenness measures. But this will not always be the case.

Finally click [**Compute**].

Set Parameters For Trigger

False Alarm vs. Fast Detection : You need to determine which type detection you are looking for. This will help you set the trigger point. Where you set the risk bar has an effect on triggering alarms. Set it too low and it will trigger an alarm quite frequently possibly giving too many false alarms. Set it too high and you might not see the trigger until it's too late.

Example

I select **Network Centralization, Row Betweenness : A-03**

Next I click the **Change Detection** tab abd select **CUSUM** from the drop down menu.

I then set the **Risk** to **0.17** and click [**Compute**].
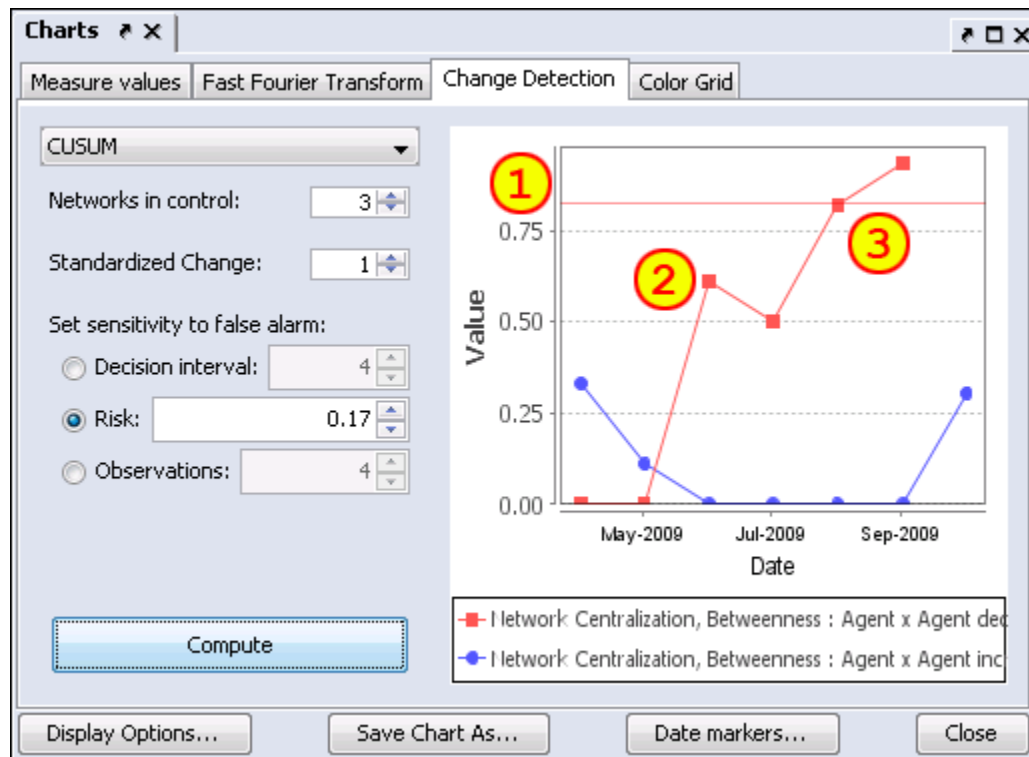
Setting the Risk to 0.17 sets the trigger line in the chart **(1)**.

The datapoint at **(2)** is a major spike but has not set off the alarm yet.

After a slight decrease the next jump at datapoint **(3)** is still under the trigger point but very close. and the following datapoint sets off the alarm **Sep-2009**. Tracing backwards the point of change is **Jul-2009** when the upward trend began.

NOTE : There are two separate charts plotted. One is for increases (red) in the social network measures, and the other is for decreases (blue) in the measure.

NOTE : When a change signals in both increasing and decreasing directions, it is likey a false alarm.

NOTE : Once a signal occurs the CUSUM is no longer plotted over time.

Early Detection/False Alarm
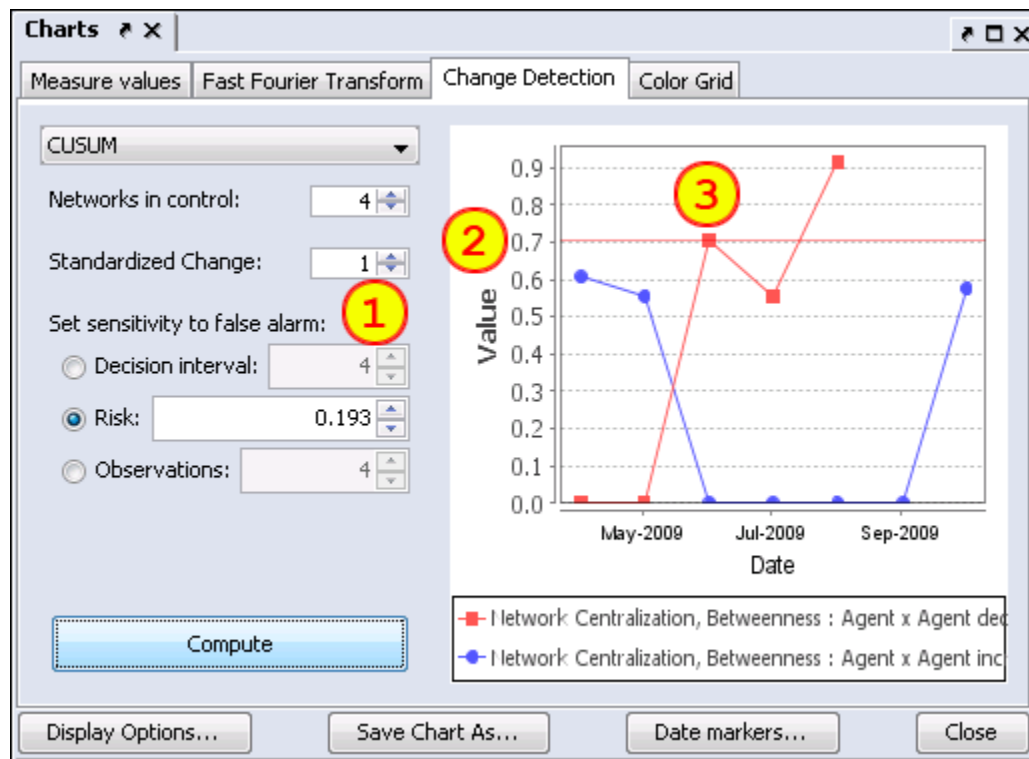
Setting parameters to different values can make changes in the alarm rate. Changes in the parameters **(1)** are **Networks in control** increase to 4 and **Risk** increases to 0.193. This lowers the alarm trigger point to 0.7 **(2)**.

(3) shows that the alarm is triggered in **Aug-2009**, a month earlier than the previous example although the upward trend is also **Jul-2009**.

Had the Risk been risen any more the alarm would have been triggered in **Jun-2009** and the changed point would be taken as **May-2009**, a full two months earlier than the previous example.

The question would be is **May-2009** the true change point or is it merely a false alarm given the parameters set to detect it.

AutoMap - Lessons Starter

Lessons Starter

1. **Your First GUI Run**
2. **Text Encoding**
3. **Using a Concept List**
4. **Data Collection**
5. **Using a Delete List**
6. **Using a General Thesaurus**
7. **Compare Concept Lists**
8. **Compare Thesauri**
9. **Remove Items**

First Run with the GUI

Description

This is an overview of your first run of AutoMap3. It is not an all-inclusive view of AutoMap but will cover the, more or less, necessary basics.

Procedure

1. Create Workspace

A good starting point is creating a project directory, a place where all your input files and output files will reside. This helps prevent files from getting lost. One suggestion is to create a top level project directory and create input and output directories within that directory. I also create a directory to place all support files such as Delete Lists and all the Thesauri.
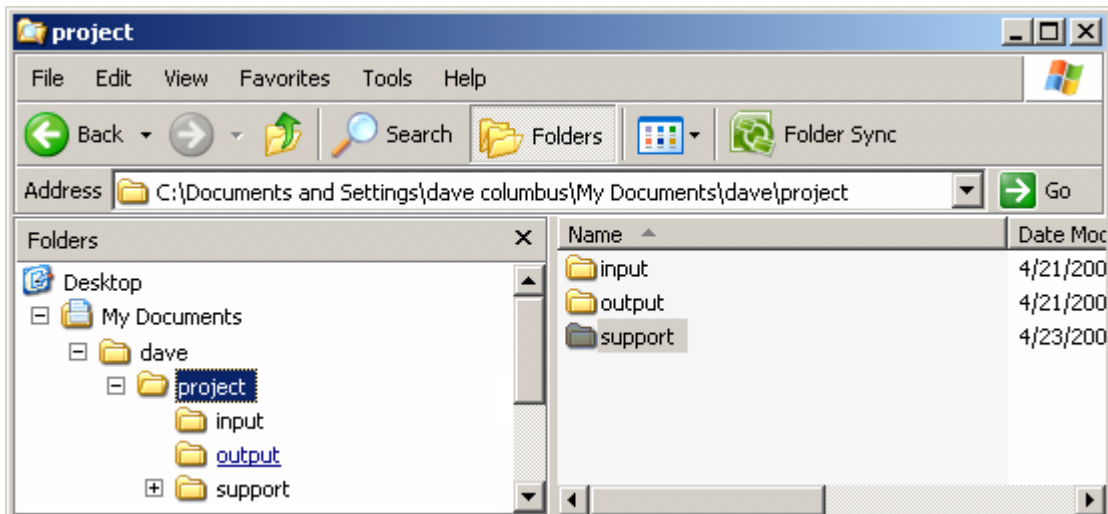
Place all your input files into the **input** directory. Place any Delete Lists, Thesauri and other files to be used in the **support** directory. Direct all your results to the **output** directory.

2. Preparing Your Files

Files can come from any source. But to use in AutoMap3 they must be .txt files. Files in a **Word** format or **html** files will not be accepted into AutoMap. Any of these formats must be re-saved as a **.txt** format. This could be as simple as re-saving the file in the correct format to doing a copy-and-paste of the text into a text editor.

NOTE : *AutoMap accepts a variety of text encodings but they must be in a .txt format.*

3. Load your files into AutoMap

Once the text files are in the correct format they can be loaded into AutoMap. From the Pull Down Menu select *File = Select Input Directory* and navigate to the directory where you placed your text files. The first of these files will appear in the main window.

usCitizen.txt

```
John Smith lives in the USA. He is a United States citizen.
```
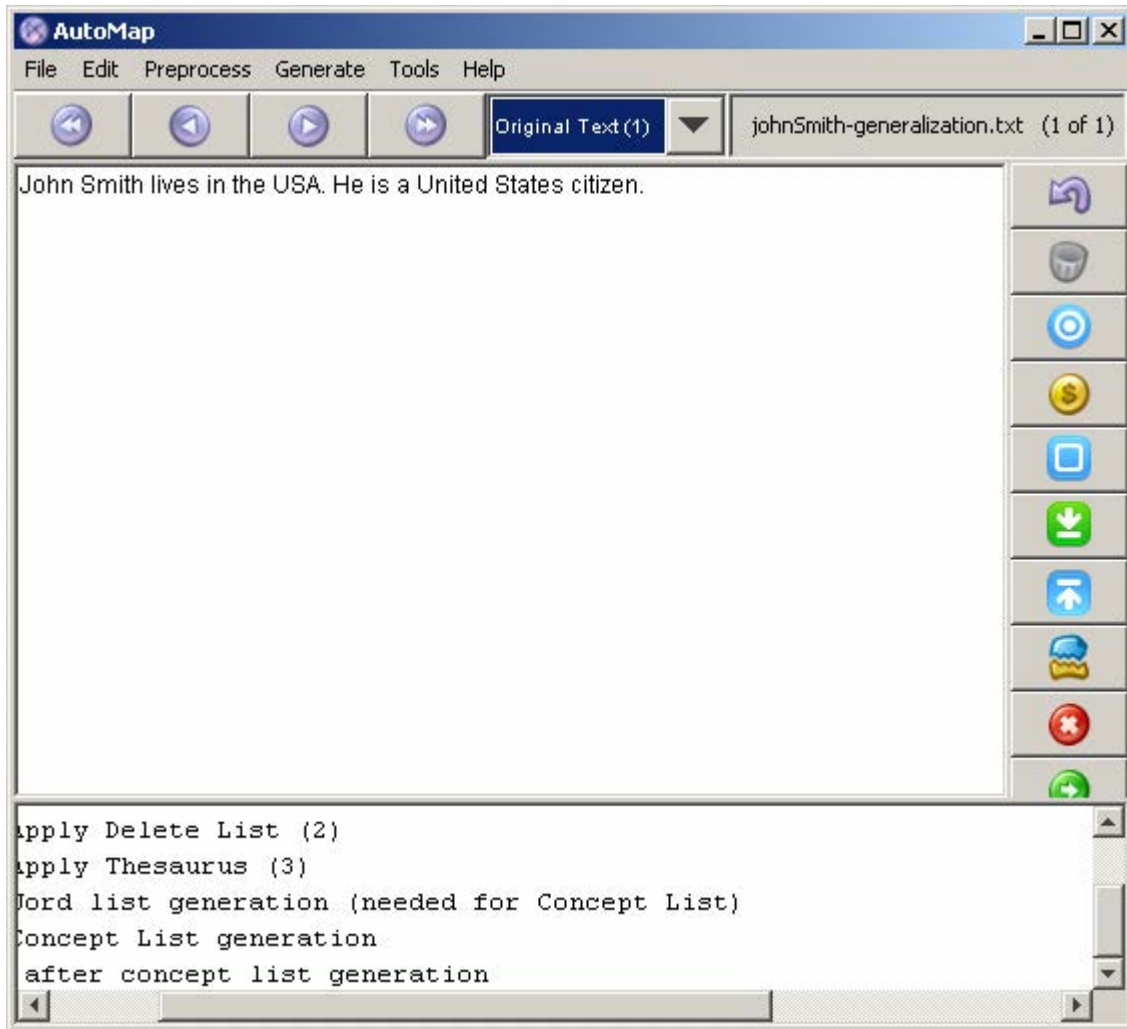
4. Decide on the Preprocessing Functions to use

Functions from the **Preprocessing Menu** affect all loaded text. None of these functions create any output files (though some require externally created files to work). These functions **remove excess concepts** (e.g. Remove Punctuation, Numbers, extra white spaces, or Symbols) or **modify concept names** (e.g. Thesauri for creating key_concepts) for easier generation and post-processing functions.

5. Creating a Generalization Thesaurus

Many people, places and things can be known by a variety of different formats of their names. A General Thesauri helps consolidate these various names under one unifying term. Below is an example. The format is **concept,key_concept**. Concept can be one or more words but key_concept must be one single word which can use the underscore.

usCitizenGenThes.csv
John Smith,John_Smith
United States,United_States_of_America
USA,United_States_of_America

NOTE : *The Thesaurus changes both the 2nd and 3rd concepts to the key_concept* ***United_States_of_America****.*

Create this file in a text editor or spreadsheet program and save it as a **.csv** file.

6. Apply a Generalization Thesaurus

From the Pull Down Menu select ***Preprocess => Apply Generalization Thesauri*** and navigate to your Generalization Thesaurus. At the Adjacency dialog box select **Rhetorical**. After applying the text will change to reflect the application of the thesaurus.

Text after General Thesauri applied

John_Smith lives in the **United_States_of_America**. He is a **United_States_of_America** citizen.

7. Edit your General Thesauri

After applying a Thesaurus the list can be altered by selecting from the Pull Down Menu ***Tools => Thesauri Editor***.

8. Delete Extra/Unneeded Concepts

Texts usually have many extra concepts, or noise, that are not relevant to the semantic connections. A Delete List removes those extra words. From the Pull Down Menu select ***Preprocess => Apply Delete List*** and navigate to the Delete List you want to use. The text in the main will change to reflect the application of the Delete List.

usCitizenDeleteList.txt
in
the
is
a

Text after Delete List with rhetorical adjacency

John Smith lives **xxx xxx** USA. He **xxx xxx** United States citizen.

9. View and Alter Delete Lists

After creating Delete List you may want to make changes to it with the Delete List Editor. From the Pull Down Menu select ***Tools => Delete List Editor***. From this

tool you can add or remove concepts from a Delete List. From the Pull Down Menu select *File => Save as Delete List* and either replace the old file or save it as a new file.

*NOTE : If you decide to make changes to the Delete List then **Undo** the applied Delete List and reapply the new one.*

10. Generate Output

After preprocessing the text it's time to produce output from them. The Pull Down Menu *Generate* contains functions that tell AutoMap to write output for the function selected; Concept Lists, Semantic List, Parts of Speech, and other useful functions.

Each function outputs files that can be examined for analysis and used to further process the text files.

11. Example of Output for Concept List

From the Pull Down Menu select *Generate => Concept List*. AutoMap will ask for the directory to save the concept list.

```
concept,frequency,relative_frequency,gram_type,tf-idf
"He","1","0.5","single","0.0"
"John_Smith","1","0.5","single","0.0"
"United_States_of_America","2","1.0","single","0.0"
"citizen","1","0.5","single","0.0"
"lives","1","0.5","single","0.0"
```

23 SEP 09

Encoding Lesson

Encoding Problems

The first you need to know is when you use the **Select Input Directory** AutoMap expects to find files in the standard **UTF-8** format. If the files in the directory are encoded differently the text in the display will not show up correctly.

Because the text you want to analyze could possibly come from a variety of sources there's no assurance that it's in the UTF-8 format. Word files, web pages, emails, or whatever else you can find can have a variety of encodings. And sometimes when you import text you find it's **NOT** in the form you thought it would be.

Those empty little boxes

Occasionally, when importing text, you will get empty boxes instead of the some specialty characters. This is due to differences in the encoding schemes. What is the problem? It's a simple explanation and a simple fix.

```
There are two kinds of UTF-8, with and without the BOM (Byte
Order Mark) at the beginning of the file. Microsoft products
require the BOM in order to recognize UTF-8, and the UTF-8
they produce has a BOM. Most other products produce UTF-8
without a BOM.

Technically no BOM is required for UTF-8, but Microsoft has
adopted the convention of using its presence to distinguish
that encoding from the OS default.
```

So without that marker some of your text may become mismarked.

Cut-and-Paste problems

Your first document may have started out with a proper UTF-8 format. But as you begin cutting and pasting material from other sources you may be adding oher formats. Anytime text is pasted into another document it retains it's encoding.

Smart Quotes

Smart quotes are not considered proper quote characters. The standard value for the straight quote character is **34**; there is only one character for both the beginning and ending quotes. The smart, or curly, quotes are actually two entirely different characters, **"** (left curly quote mark) and **"** (right curly quote mark).

The original designers of the ASCII character set did not define a standard method for identifying properly curved quotation marks, so computers have had a problem properly exchanging quotation marks ever since.

The Solution

First try AutoMap's **Text Import** function and allow AutoMap to try to detect the sources encoding. This function is designed to convert text into standard UTF-8 format.

If the program doesn't work, you can also use a third party text-editor to convert the data. Both Microsoft Word and the free Notepad++ text editor (http://notepad-plus.sourceforge.net/uk/site.htm) support this functionality. Simply open the file in your editor and do a **Save As...** a .txt file. Microsoft's Notepad saves files as single-byte ANSI (ASCII) by default.

Foreign Characters Sets

Foreign characters sets are an entirely different matter. These will require you to have the proper font installed on your computer as each font can possibly be encoded differently.

Using a Concept List

Description

Concept Lists (the frequency of concepts in one file) and Union Concept Lists (the frequency of concepts throughout all loaded files) can be used to work with concepts from one or more text files. It lists the frequency, relative frequency, and gram type. When working with multiple files you'll find the Union Concept List useful.

NOTE : *The number of unique concepts considers each concept only once. The number of total concepts considers repetitions of concepts.*

Concept List Procedure

1. Select a text file(s) to use

Place your text file(s) in an empty directory. Load the file(s) by selecting from the Pull Down Menu *File => Import Text Files*.

theBoy-4.txt:

```
See the boy named Dave. He has 2 balls. 1 ball is red. 1 ball
is blue.
```

milkAndCookies.txt:

```
Dave wants milk and cookies. He drives to the store. He then
buys milk and cookies.
```

2. Create Concept List
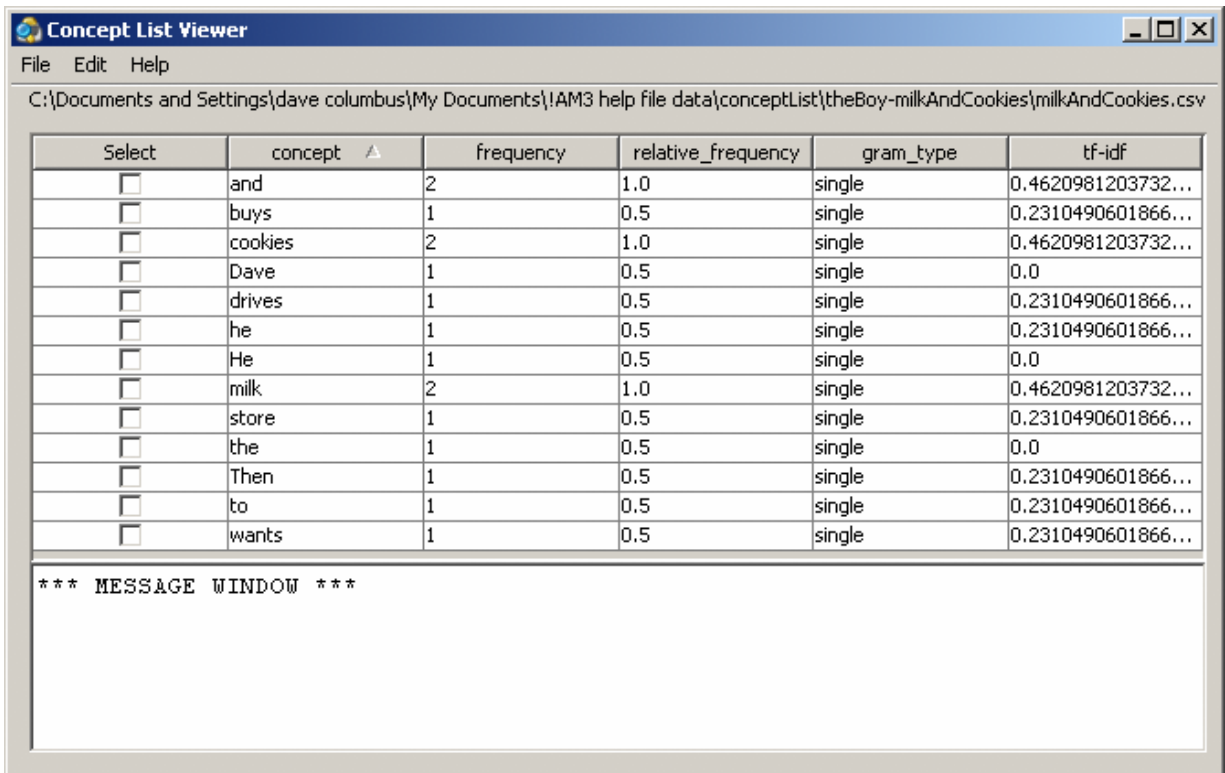
From the Pull Down Menu select *Generate => Concept List => Concept List (per text)*. Navigate to where you want to save the file(s) and click **Select**.

3. Decide if you need Union Concept List

After specifying the directory for the Concept List(s) AutoMap will ask if you want to create a Union Concept List. Unless you know you will not need it, click **Yes**.

4. Review a Concept List in the Concept List Viewer

The Concept List(s) and Union Concept List can be viewed using **Concept List Viewer**. From the Pull Down Menu select *Tools => Concept List Viewer*. From the Viewer Pull Down Menu select *File => Open File* and navigate to the location of the concept list to view.

NOTE : *If you load a Union Concept List the right-most column has the header* **relative_percentage** *denoting the frequency of a concept occurring in all text files.*

Concept List Viewer functions

Sorting a Concept List

A Concept List can be sorted by clicking on any of the headers. This will sort the list in an ascending order. Clicking the same header again will reverse the sorting order to descending. Which header is being used will be denoted by a small triangle to the left of the header name.

NOTE : *The small triangle to the right of the header will tell you which header is used for sorting and whether it's in ascending upward facing arrow or descending downward facing arrow order.*

Creating a Delete List

A Delete List can be created from the Concept List Viewer. Place a check mark in the **Selected** column of each concept to include in a Delete List. From the Pull Down Menu select *File => Save as Delete List*. AutoMap will prompt you to select a location to save the Delete List. Give it a unique filename and click **Open**.

NOTE : *The Delete List can be saved in either the .txt or a .csv format.*

This new Delete File can now be loaded and applied to your texts.

Selecting Concepts

In the Viewer *Edit* menu the viewer gives you options for selecting/unselecting multiple concepts.

> **Select All :** Places a check mark in every check box.

> **Select None :** Clears all check marks from the concept list.

> **Select Minimum Threshold :** The number input will select concepts with frequencies equal to or greater than the threshold to be selected.

> **Select Maximum Threshold :** The number input will select concepts with frequencies equal to or lesser than the threshold to be selected.

> **Find :** will highlight the input text, if found, with a cyan background.

> **Reset Colors :** removes all highlighting.

23 SEP 09

Data Collection

Description

AutoMap is designed to extract, analyze, and interpret relational data (also known as network data) from unstructured, natural language text data.

Relation Extraction Sources

The source of your daya can be anything: books, television, newspapers, blogs, emails, internet sites. AutoMap will extract the data and sort it into relational data which can be further analyzed in ORA.

Method

The first thing to do is identify the problem/goal.

Next all/some of the concepts need identified in the texts and the links between them (binary, typed, directed, weighted) can be defined.

Now this data can be represented as relational data (graph or list).

Then the data can be analyzed.

And finally the results can be interpreted.

How is network data collected?

Interviews, Automated (web-based surveys).

Person	Albert	Betty	Charlie
Albert	0	1	0
Betty	0	0	1
Charlie	0	1	0

Data collection is more of an approximation via Network Text Analysis as most real-world networks and sequential data are not iid (independent and identically distrbuted). Network data is a concise represntation of what's in the text data - Is it not the truth, only an approximation.

22 JUL 09

Using a Delete List

Description

Delete Lists allow you to remove **non-content** bearing conjunctions, articles and other noise from texts. It also allows you to delete concepts that you just don't care about for analysis purposes.

Delete Lists can be created internally in AutoMap or externally in a text editor or spreadsheet. They are a preprocessing stage done before doing any output.

NOTE : *Whether you apply the Delete List(s) before or after applying a Thesauri will depend on your exact circumstances. AutoMap allows for applying multiple Delete Lists is that is a necessity, one before and one after.*

Delete List Procedure

1. Select a text file(s) to use

Place the text to use in an empty directory. Below is an example text.

tedInUSA.txt

    Ted lives in the United States of America. He lives on a dairy
    farm. He considers it a good life. Would he ever consider
    leaving?

2. Create a Concept List

To create a Delete List it helps to know the frequency of the concepts in the files. From the Pull Down Menu select *Generate => Concept List*. Save the file in your output directory.

3. Create a Delete List

A Delete List can be created within AutoMap using the Concept List Viewer. From the Pull Down Menu select *Tools => Concept List Viewer*. Navigate to the directory containing the Concept List file and select a file. In the Viewer place a check mark in the Selected column next to the concepts to include. From the View Pull Down Menu select **File => Save as Delete List**. Save the file in your support directory. The Delete List created can be viewed in the Delete List Viewer by selecting *Tools => Delete List Viewer*.

A Delete List can also be created in Excel. Load the Concept List in Excel and sort by the frequency column. Create a new column and label it Delete List. Place an **X** next to all concepts to include in the Delete List. Sort the spreadsheet by the Delete List column. Copy all the rows containing an X in the Delete List column. Create a new sheet and paste these rows into it. Delete the column with the **X**s. Save this file as a .txt file.

NOTE : For large concept lists review the top 100 entries and add concepts to the Delete List items. Resort primarily by the Delete List column and secondarily by the concept column. Review the top 100 again. Repeat this process until the top 100 entries are of interest. review the rest of the list for other unneeded or unwanted concepts. Save this list as a .txt file.

TIP : Create a cut-off limit (e.g. a word needs to be used at least three times. Concepts used less than that are placed on the Delete List.

tedInUSADeleteList.txt
```
in
the
of
he
on
a
it
```

4. Apply a Delete List

From the Pull Down Menu select *Preprocess => Apply Delete List*. Navigate to the directory where your delete list is stored.

5. Adjacency Option

AutoMap will ask what type of adjacency you what to use. The **Adjacency option** determines whether AutoMap will replace deleted concepts with a placeholder or not.

- **Direct Adjacency :** Removes concepts in the text that match concepts specified in the delete list and causes the remaining concepts to become adjacent.

- **Rhetorical Adjacency :** Removes concepts in the text that match concepts specified in the delete list and replaces them with **(xxx)**. The placeholders retain the original distances of the deleted concepts. This is helpful for visual analysis.

6. The newly pre-processed texts can be viewed in the main window.

Delete List with Rhetorical Adjacency

```
Ted lives xxx xxx United States xxx America. He lives xxx xxx
dairy farm. He considers xxx xxx good life. Would he ever
consider leaving?
```

Delete List with Direct Adjacency

```
Ted lives United States America. He lives dairy farm. He
considers good life. Would he ever consider leaving?
```

NOTE : *If using Direct Adjacency the concepts are* ***NOT*** *replaced with anything. The concepts are moved next to the ones before and after. For more information on Delete Lists see the Content section.*

NOTE : *If you need to remove the Delete List it can be* ***Un-applied*** *using the* ***Undo*** *function under the Pull Down Menu* ***Preprocess*** *or from the quick launch buttons.*

Other Delete List Functions

Multiple Delete Lists

Multiple delete lists can be applied to the same text in any order the user wants. They can be viewed in order using the Pull Down Menu.

NOTE : *This is useful if you have multiple specialty Delete Lists.*

Modifying a Delete List

After a Delete list is created you can modify it using the **Delete List Editor**. From the Pull Down Menu select *Tools => Delete List Editor*. From the View Pull Down Menu select *Open File* and navigate to the Delete List to view.
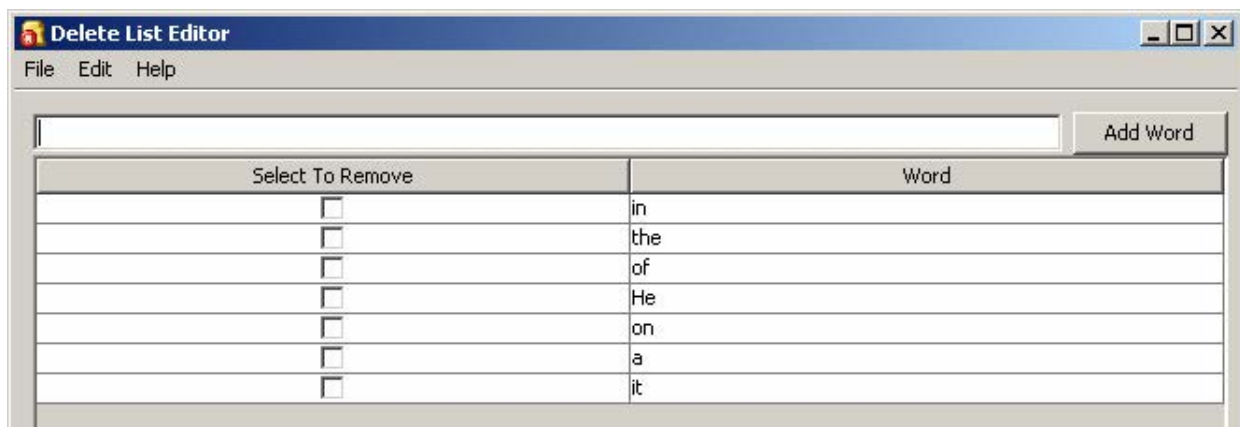
From this window you can:

- **Add Concepts:** In the textbox above the list type **one** new concept then click **[Add Word]**. Your new concept will be added to the list. Repeat until you've added all concepts necessary. The next time you save the Delete List it will be saved without the checked concepts.

- **Remove Concepts :** Click in the check box next to the concept to remove in the **Select to Remove** column.. The next time you save the Delete List it will be saved without the checked concepts.

90

> ***NOTE :*** *No concepts are added or deleted until you actually save the file.*

- **Create New Delete List :** From the viewer Pull Down Menu select ***File => Save as Delete List***. AutoMap will prompt you to select a directory and give the file a new file name.

   > ***NOTE :*** *Make sure to give the file the **.txt** extension.*

Save text(s) after Delete List

After applying a Delete you can save your texts by selecting from the Pull Down Menu ***File => Save Preprocess Files***. This step must be done before any other preprocessing as this option saves the texts at the highest level of preprocessing.

23 SEP 09

Using a Generalization Thesaurus
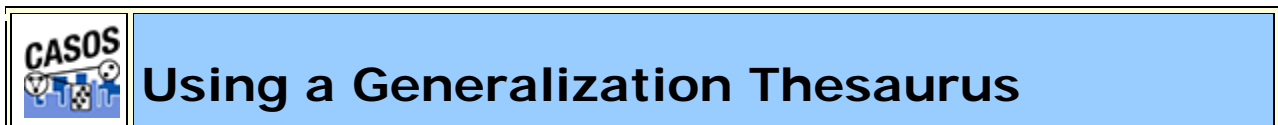
Description

Thesauri are generally used to take multiple concepts, in different forms, and compile them under one **key concept**. If this is not done then the same concept could be listed many times as individual concepts.

Thesauri Procedure

1. Select your text file(s)

Copy these files into a text editor and save them as **johnSmith.txt** and **countryThesauri.csv**. Place the johnSmith.txt in a folder by itself. Place countryThesauri.csv in an accessible folder other than where johnSmith.txt resides.

usCitizen.txt

```
John Smith lives in the USA. He is a United States citizen.
```

2. Prepare your Thesauri

For this simple file the thesauri is short. Larger texts could easily have hundreds of entries. This can be done in either a text editor or a spreadsheet.

usCitizenGenThes-Ext.csv
```
The United States of America,United_States
United states,United_States
John Smith,John_Smith
America,United_States
USA,United_States
```

3. Loading the files

From the main menu select *File => Import Text Files*. Navigate to the directory where you placed johnSmith.txt and click **Select**. The file will appear in the main window.

4. Applying the thesauri

From the main menu select *Preprocess => Apply Generalization Thesauri*. Navigate to the directory where you placed countryThesauri.csv and click **Select**. The thesauri will be applied and a new text will appear in the window with the thesauri substitutions. **Apply Thesauri (3)** will now appear in the dropdown menu denoting AutoMap has applied the thesauri.

Text after thesauri applied

```
John_Smith lives in the United_States. He is a United_States
citizen.
```

Thesauri Editor

AutoMap contains a **Thesauri Editor** making it easy to revise your thesauri files. From the dropdown menu select *Tools => Thesauri Editor*.

From the Thesauri Editor menu select *File => Open File*. Navigate to the directory with your thesauri and click **[Okay]**. From the Thesauri Editor you can make changes to the thesauri.

Questions regarding Thesauri

Different Thesauri for different purposes

You might initially think it's necessary to keep all the thesauri entries in one place. Just easier to find everything. But it essence it would make it easier to keep track by splitting up the thesauri into smaller files, each with it's own purpose. Below are some of the specialty thesauri used in this lesson.

You might have a general countries thesauri that is always used to fix multiple ways of posting a country. The U.S. can be listed in a multitude of ways:

```
U.S., America, United States, The United States of America
```

A country thesauri could have all these permutations listed and convert them all to the same concept say, `United_States`. So instead of four individual concepts you would have one.

```
U.S.,United_States
America,United_States
United states,United_States
```

`The United States of America,United_States`

If your project concentrates on a particular field you could have a thesauri that contains names of organizations, resources, or people's names which appear with regularity.

Analysts working on a similar subject every day would need the same names, places, and resource for each run. A special thesauri could take care of that and would be easier to maintain as a single file.

Is the thesaurus case-specific?

NO! If there are two or more entries for the same concept the first thesaurus entry will be used for all replacements.

`HE,Tom`
`He,Dick`
`he,Harry`

Every instance of **HE, He, and he** will be replaced with **Tom** without comparing their case.

Running the Delete List Before or After the Thesauri

You may also find it necessary to run a Delete List on your files. Whether you apply a delete list before or after applying a generalization thesaurus will depend on your set of files. A longer discussion can be found in the Content section under **Process Sequencing**.

Using of one large thesauri vs. multiple smaller thesauri

This will be a personal choice of the user. Multiple thesauri have the advantage of easier editing but the downside is needing to apply multiple thesauri to the same set of files.

NOTE : *When doing multiple runs on the same data some analysts prefer to maintain a single thesaurus.*

NOTE : *The Generalization Thesaurus is NOT case sensitive to what it finds in the text. United States, United states, and united States are all considered the same bi-gram and would be replaced with the same entry.*

23 SEP 09

Compare Concept Lists

Description

Compares two Concept Lists and displays concepts that appear in either both files (white background) or in a single file (red or green background). The viewer is called from the pull-down menu by selecting *Tools => Concept List Viewer*.

Load
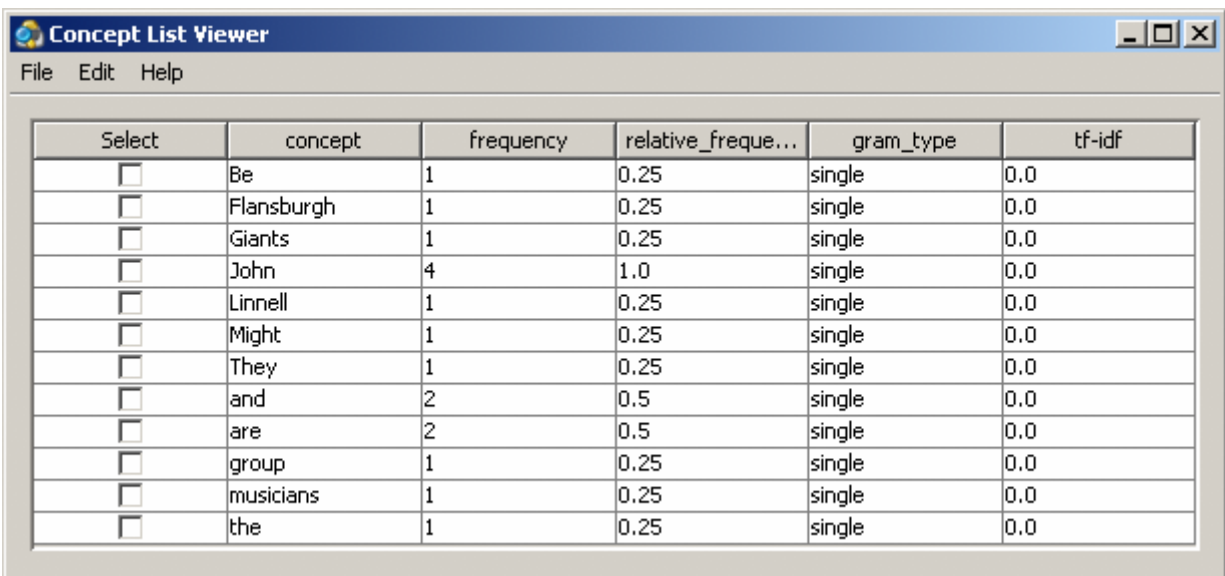
This example will use a Concept List created with a raw file and a Concept List created after applying a thesaurus. This will be used to demonstrate how the viewer detects frequency differences.

tmbg.txt

```
John Flansburgh and John Linnell are musicians. John and John
are the group They Might Be Giants.
```

The tmbg's file is loaded into AutoMap with *File => Import Text Files* and a Concept List is created *Generate => Concept List => Concept List (Per Text)*. This Concept List is then loaded into AutoMap with *Tools => Concept List Viewer*.

*NOTE : Notice the concept **John** has a frequency of **4**.*

Close the Concept List Viewer.

Go to the directory where your Concept List is saved and rename it tmbg-noThes.csv. This will prevent AutoMap from overwritting the file.

Apply the generalization thesaurus tmbg-genThes.csv file with *Preprocess => Apply Generalization Thesauri*.

tmbg-genThes.csv
```
John Flansburgh,John_Flansburgh
John Linnell,John_Linnell
They Might Be Giants,They_Might_Be_Giants
```

Create a Concept List with *Generate => Concept List => Concept List (Per Text)*. This Concept List, named tmbg.csv, will be used as the compare file.

Comparing the two files

Start the Concept List Viewer with *Tools => Concept List Viewer*. Load the first Concept List with *File => Open File* and navigate to the directory containing the tmbg-noThes.csv (the file without the thesaurus applied). Next we will compare this with the second file. From the pull-down menu select *Edit => Compare File* and navigate to the direcoty containing tmbg.csv (the file with the thesaurus applied).

The concepts with white backgrounds are found in both files. The concepts with red backgrounds are found only in the original file. The concepts in green are found only in the newly compared file.

The cells with the yellow backgrounds are concepts found in both files (e.g. concepts on white backgrounds) but they have different values. Notice on the image below the arrow hovering over the frequency value for **John**. In the original file John had a value of **4**. The tooltip displays a value of **2** which is the value in the compared file.

After making a comparison AutoMap will display the statistics of the comparison in the Message Log Window. It will display added (green), deleted (red), and altered (yellow).

08 SEP 09

Compare Thesauri Files

Description

The **Compare File** functions is implemented in many of the tools: **Delete List Editor, Thesauri Editor, Concept List Viewer, and Table Viewer**. It will compare the currently loaded file with a second file on disk and display if an entry is contained in both files (white background), originally loaded file (red background), or the compared file (green background).

NOTE : *The example uses thesauri files but the color coding is identical across all the included tools.*

The Files

Below are the files used in this lesson. They can be copied and saved and used for your examples. Be sure to save them in the .csv format.

abc.csv	xyz.csv	a.csv	az.csv	a-caps.csv
a,alpha b,beta c,charlie	x,x-ray y-yankee z,zulu	a,alpha	a,alpha z,zulu	a,ALPHA

The Comparisons

Open a file

Start the Thesaurus Editor from the pull-down menus *Tools => Thesauri Editor*. From the Editor pull-down menu open the file **abc.csv**.

Select	To	From
☐	a	alpha
☐	b	beta
☐	c	charlie

Comparing a sub-set file

From the editor pull-down menu select *File => Compare Files*. Select the file **a.csv**. This file contains a single concept from the original **abc.csv** file.

Select	To	From
☐	a	alpha
☐	b	beta
☐	c	charlie

The line **a,alpha** in white shows this concept is found in both files. The next two lines **b,beta; c,charlie** show these concepts are found in the original file but not in the compared file.

Compare to an exclusive file

From the pull-down menu select *File => Compare Files* and select the file **xyz.csv**. The **xyz.csv** file contains no concepts that are found in the **abc.csv** file.

Select	To	From
☐	a	alpha
☐	b	beta
☐	c	charlie
☐	x	x-ray
☐	y	yankee
☐	z	zulu

The green lines for **x,x-ray; y,yankee; z,zulu** show these concepts are only found in the new file. The red lines for **a,alpha; b,beta; c,charlie** show these concepts were found in the original file but are absent from the compared file.

A Super-set

From the pull-down menu select *File => Compare Files* and select the file **abcxyz.csv**. The compared file is the super-set of the original file which contains all the original concepts plus some unique concepts.

Select	To	From
☐	a	alpha
☐	b	beta
☐	c	charlie
☐	x	x-ray
☐	y	yankee
☐	z	zulu

The white lines **a,alpha; b,beta; c,charlie** show these concepts are contained in both files. The green lines **x,x-ray, y,yankee; z,zulu** show these concepts are found only in the compared file.

Different values

From the pull-down menu select *File => Compare Files* and select the file **a-caps.csv**. The compared file contains only one similar term to the original file but the key_concept is different.

There will be times when the same concept has different values in each file. This is defined by a yellow background. The value will display what is in the original file. By hovering over the cell it will display in a tooltip the value contained in the compared file.

Select	To	From
☐	a	alpha
☐	b	beta
☐	c	charlie ALPHA

A Mixed Thesaurus

The most common comparison will be a mixture of terms that appear, or are missing from, each file.

From the pull-down menu select *File => Compare Files* and select the file **az.csv**. This file contains all the concepts from the **abc.csv** plus three additional concepts.

Select	To	From
☐	a	alpha
☐	b	beta
☐	c	charlie
☐	z	zulu

The white line for **a,alpha** shows this concept was found in both the original and compared files. The red lines **b,beta; c,charlie** show these concepts were found in the original file but not the compared file. The green line **z,zulu** shows this concept exists only in the compared file.

21 AUG 09

CASOS Remove Items

Description

Removing unwanted, or unneeded, items is important to reduce the amount of data to analyze. This include white space, punctuation, numbers and symbols.

Remove Extra White Spaces

In many texts there are extra white spaces inserted between words or before and after punctuation. This is partly a holdover from the days of the typewriter of double spacing after a period. But with proporional fonts it's now an unnecessary practice. AutoMap finds instances of multiple spaces and replaces them a single space. Below is a text with varying number of white spaces between words.

spaces.txt

```
one space. two  spaces. three   spaces. four    spaces.
```

After AutoMap removes extra white spaces it's much easier to read.

```
one space. two spaces. three spaces. four spaces.
```

Remove Punctuation

Punctuation is mainly for use in making sure the reader understands how the words are expressed. During analysis they are somewhat unnecessary as the words themselves are more important. The Remove Punctuation function removes the following punctuation from the text: **.,:;' "()!?-**. AutoMap will remove the punctuation and either close up the sapce between or insert a white space as a placemarker.

punctuation.txt

```
"English" is hard (so very hard)!?! What's with all these
commas (,), semi-colons (;), and colons (:).
```

Removing Punctuation and inserting white space

```
 English  is hard  so very hard    What s with all these
commas       semi colons       and colons
```

Removing Punctuation and NOT inserting white space

```
English is hard so very hard Whats with all these commas
 semicolons  and colons
```

Remove Symbols

Symbols are parts of language which are not concepts but assist in understanding the language. Occasionally these symbols need to be removed to make semantic networks and meta-networks better understood.

Removing the Default List of Symbols

AutoMap has a default list of symbols that can be removed:
~`@#$%^&*_+={}[]\|/<>.

NOTE: This option is an all-or-nothing function.

symbols.txt

```
As he emailed {bob@jewelry.com} he knew the $200.00*
|+shipping| on [http://jewelry.com\~necklace] would = a ^50%
was a <`bargain>. And his #1 girl & mom deserved the best.
```

Removing Symbols and inserting white space

```
As he emailed  bob jewelry.com  he knew the  200.00
 shipping  on  http:  jewelry.com  necklace  would   a  50
 was a   bargain . And his  1 girl   mom deserved the best.
```

Removing Symbols and NOT inserting white space

```
As he emailed bobjewelry.com he knew the 200.00 shipping on
http:jewelry.comnecklace would  a 50 was a bargain. And his 1
girl  mom deserved the best.
```

Removing a User Set of Symbols

The second option for removing symbols is to define the list you want removed. The list consists of one line with all the symbols to remove together with no spaces between the entries.

symbols.txt

```
As he emailed {bob@jewelry.com} he knew the $200.00*
|+shipping| on [http://jewelry.com\~necklace] would = a ^50%
was a <`bargain>. And his #1 girl & mom deserved the best.
```

Removing Set of Symbols containing {}[]#

```
As he emailed bob@jewelry.com he knew the $200.00* |+shipping|
on http://jewelry.com\~necklace would = a ^50% was a
<`bargain>. And his 1 girl & mom deserved the best.
```

Remove Numbers

Removing numbers will remove not only numbers as individual concepts but also removes numbers embedded within concepts.

Remove Options

There are two options for removing numbers.

1. Replacing the number(s) with a space
2. Removing the number(s) and closing the distance between the letters before and after.

Examples

Remove numbers as individual concepts.

buckleMyShoe.txt

1, 2, buckle my shoe! 3, 4, shut the door

Text after RemoveNumber:

, , buckle my shoe! , , shut the door.

Numbers within other concepts and closing up distance.

c3pO.txt

C3PO was a robot in the movie Star Wars.

Text after RemoveNumber:

CPO was a robot in the movie Star Wars.

Numbers within other concepts and inserting white space.

c3pO.txt

C3PO was a robot in the movie Star Wars.

Text after RemoveNumber:

C PO was a robot in the movie Star Wars.

Script Runner

Scripts can be created in AutoMap and saved or directly in the ScriptRunner GUI.

Import a text file in AutoMap. For this lesson I'll be using Julius Caesar's Act 4 **act 1.txt**. Now either press the **Quick Launch Delete List button** or select **PreProcess > Apply Delete List** and apply a Delete List. For the Delete List I'll use **juliusCaesar_DL.txt**.

When AutoMap is finished processing select from the menu **File > Save Script File**. Navigate to a directory to save your script.

Start the ScriptRunner by selecting from the menu **Tools > ScriptRunner**.

NOTE : If you don't have the newset .NET installed you will get a .NET Framework error. Just press OK - it will start.

First you need to create a temp directory to store processing results. It also contains all the files ScriptRunner generates. The temp directory differs from the Output directory as the temp directory is where AM3 stores it's work file. Output directories are created by AM3 on an individual basis for each type of output generated.

From the main menu select **File > Load Script File**. Navigate to the directory where the file just created was saved, select it, and click OK.

Click the Parameters tab and review the procedures from the loaded script. You will see one line containing the parameters of the file to use:

> **tempWorkSpace :** Displays the pathway for ScriptRunner's work area.

> **textDirection :** LT, RT, LB, RB

> **textDirectory :** The pathway to the file(s) loaded.

> **textEncoding :** Displays how ScriptRunner will load the file. In our example this is **autoDetect**

Click the **PreProcessing** tab. You will see the one process performed from AutoMap, Apply Delete List. In this simple script this is the only item under the PreProcessing tab.

> **adjacency :** Value can be **r** - rhetoric or **d** - direct.

> **deleteListLocation :** Display the pathway for the Delete List. This pathway shows where the Delete List resides when we created the script in AutoMap.

> **inputDirectory :** Allows you to specify a different directory to use to for your input files. Useful if you want to test something on a subset of files. When

using a script created in AutoMap this does not have to be filled in. ScriptRunner will use the directory used by AutoMap.

NOTE : *If you designate an input directory that doesn't exist the script will fail and you will see the dialog box with the little bomb.*

outputDirectory : Allows you to change the final destination for output files written. When using a script created in AutoMap this does not have to be filled in. ScriptRunner will use the directory used by AutoMap.

NOTE : *If creating a script in ScriptRunner then you will have to specify these two directories else you might not know where your files are being written. AM3 won't create directories for you. It will only browse. This makes you decide where you want to store the results.*

For some reason you may want to either remind yourself of some fact about this Delete list or tell a new user something they may need to know. You can **insert a comment or print a message**.

Display Message : The message typed in the text box will be displayed in a dialog box when the script is run. Primarily used to send a message to the user.

Insert Comment : Usually used to annotate the script. Primarily used to remind the creator of the script the purpose of any particular action.

Insert a Comment does not display anything on screen. It is for you to annotate what you did. When other people open your script they won't have to guess why you chose to do a certain action. It's also helpful when you come back to a script to jog your memory as to your reasons for how it was constructed.

When creating more complicated scripts you can **drag and drop** any of the buttons to the canvas to add items. You can also delete any item by clicking the **Red X**. You can also modify the order of the items however you want.
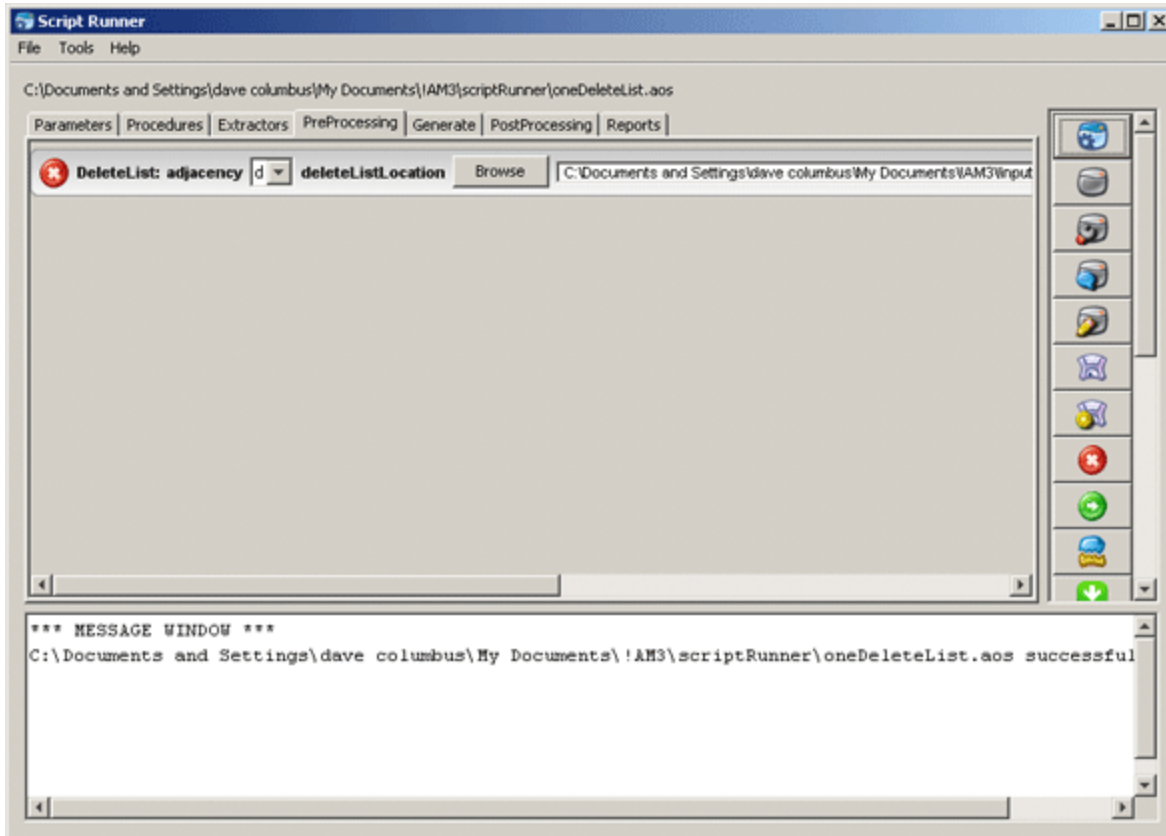
Above shows how to place a message that will be displayed in a dialog box for the user. You can type whatever you want displayed in the text box.

NOTE : *Yes could also drag-and-drop a second Delete List as duplicate processes can be executed twice in a row. This is usually done when you have multiple Delete Lists for specific purposes.*

Running External Programs

Used to extend ScriptRunner's usage by running other applications (i.e. the NotePad application). Primarily used if you need to adjust some data or result before allowing the script to proceed.

Add New Programs

Fills in the pathway to the program and also inserts a new button at the bottom of the quick launch pane.

Working with Large Thesauri

Description

When working with a large amount of documents, each one being large themselves, you will find your thesauri growing increasing large. There are several items which you should be aware of in this circumstance.

The Order of the Thesaurus

The order of the thesaurus entries can have an impact on your results. Better results will be obtained if the thesaurus is in a descending length order.

The **Sort Thesaurus** function can assist in ordering your thesauri. From the pull-down menu select *Procedures => Sort Thesaurus*. Navigate to the thesaurus you want to sort. Then give it a new name, and if you want, a new location and save it.

The Sort Thesaurus functions sorts by number of words in an entry. The order of entries with the same number of words will not change.

Conflicts in Data

A problem can occur when you have more than one agent or location with the same name. In the beginning the text may specifically refer to **John Smith** and **John Doe** but afterwards you might find that a person is only referred to as **John**. AutoMap will use the first occurrence of John it finds when making substitutions.

AutoMap processes a thesaurus from top to bottom. This is why it's important to sort a thesaurus by the length of the entries.

Sequence of Operations

    Mohammed abd al-Sha'bai

This name contains both punctuation and symbols. Depending on the order of the preprocessing you could end up with any of the following three:

Mohammed_abd_al_Sha_bai
Mohammed_abd_al-Sha'bai
Mohammed_abd_alShabai

How General Is Too General?

If you start with a 2,000 page text set, you will probably be unable to read all of the text prior to starting the processing. One of the problems you may encounter in such situations relates to common names and terms. For example, a preliminary review of the text may reveal that the documents contain an individual named **Joe**. Given that this is very common name, the data may contain several other individuals named **Joe**. Should this occur the program will incorrectly process all Joes as if they were the first Joe. Users need to exhibit caution before including common names and common terms in their thesauri.

If your text contains two names, `Joe Smith and Joe Jones` and a thesaurus entry for the name `Joe` then AutoMap will substitute all Joes for the first entry in the thesaurus.

Using the ThesauriContentOnly option

You create a Meta-Network (Carley, 200) with the one-grams **dog, cow, and farm**. If you are going to use the **UseThesauriContentOnly** option then those three terms need to be in your General Thesaurus also. If they are not in the thesaurus then they will be eliminated from the output and the Meta-Network will not see them to tag them.

Positive or Negative Links

AutoMap creates links between nodes. But AutoMap does not differentiate between positive and negative links. After processing that is up to the analyst.

```
The U.S. lacks formal diplomatic relations with North Korea,
Bhutan, and Cuba, but has close relationships with Canada and
the U.K.
```

Using a large window will create a fully connected graph with all links the same type. It will also make connections for all the countries involved. The reality is that there should be positive links to Canada and the U.K. and negative links to North Korea, Bhutan, and Cuba.

Using a small window size would create a link set that is wrong:

```
US => North Korea => Bhutan => Cuba => Canada => UK
```

right:

```
negative links: US => North Korea; US => Bhutan; US => Cuba
```

```
positive links: US => Canada; US => UK
```

Extracting a Semantic Network

Description

Text files have connections but they are sometimes difficult to see. You can use AutoMap and process them to create semantic networks which can be viewed in ORA.

This lesson details processing text files in AutoMap to extract a Semantic Network, how to view it in ORA. Other lessons will detail specific reports that can prove useful.

What is a Semantic Network?

Semantic networks are knowledge representation schemes involving nodes and links between nodes. It is a way of representing relationships between concepts. The nodes represent concepts and the links represent relations between nodes. The links are directed and labeled; thus, a semantic network is a directed graph.

Procedure

This lesson will use the file: **JC_summary-1.txt**.

Load text document into AutoMap

Place all the text files for conversion into a single folder. From the Pull Down Menu select *File => Import Text Files*. The first text will be displayed in the main window and the filename will appear in the Filename Box. Using the File Navigation Buttons you can navigate through the loaded files.

Build a General Thesauri

Many people, places and things are made up of two or more words. For example Julius Caesar, Brutus's House, status of Caesar. Before producing any files usable in ORA it's necessary to combine these multi-word concepts into **key concepts**.

NOTE : Some concepts include the definite article in their name and should be included in the thesaurus.

If you have no previous thesaurus then one will need to be created from scratch. This will require going through the text files and finding those multi-word concepts and creating a list of key concepts. The format for this is `multi word concept,key_concept`.

NOTE : *Be sure **NOT** to leave any spaces before or after the comma.*

Below is part of the Generalization Thesaurus that is used for this lesson. It contains concepts from the Julius Caesar text.

juliusCaesar-GenThes.csv
```
Ides of March,Ides_of_March
Julius Caesar,Julius_Caesar
Julius Caesar's,Julius_Caesar
Julius Caesar's status,statue_of_Julius_Caesar
kill Caesar,kill_Caesar
kills herself,commit_suicide
king,emperor
letter,forged_letters
```

Apply a General Thesauri

After the thesaurus is created it is time to apply it to the text. From the Pull Down Menu select ***Preprocess => Apply Generalization Thesauri***. Navigate to the directory where the thesauri is saved and click **[Select]**. Next a dialog box will appears asking if you want to use **Thesaurus Content Only**. Leave the response as **No**.

See *Content => Thesaurus Content Only* *for more information.*

Notes about Thesaurus Building:
1. In large texts there may be multiple person with the same first name.

2. The definite article in the concept like **the USDA** would be placed in the Thesaurus instead of being deleted in the Delete List.

Create the Concept Lists

Next we need to create a Delete List. One way is to first create a Concept List and use this to help in creating a Delete List. The frequency attribute will assist in finding unneeded and unwanted terms.

From the main menu select ***Generate => Concept List => Concept List (Per Text)***. Navigate to the directory to save the files and click **[Select]**. AutoMap will ask if you want to create a **Union Concept List**. Click **[No]** as you only have one file loaded.

NOTE : *With multiple files loaded you would select **Generate => Concept List => Concept List (Union Only)**. This creates one list for all files currently loaded.*

Build a Delete List

Open the Concept List Viewer by selecting ***Tools => Concept List Viewer***. From the viewer menu select ***File => Open File***. Now navigate to the directory containing the newly created Concept List and click **[OK]**.

Click the header **Frequency**. This will sort the concepts by the number of occurrences in the file(s). To build a Delete List place a check mark in the **Selected** column for all the concepts you wish to place in the Delete List. When you are finished select *File => Save Delete List*. Navigate to the folder you want to save the Delete List file. Close the Viewer.

Apply a Delete List

From the main menu select *Preprocess => Apply Delete List*. Navigate to the directory with your newly created file and click **[OK]**. You will be asked whether you want Rhetorical (replaces deleted concepts with a placeholder **xxx**) or Direct (removes the concept entirely) adjacency. For this lesson I choose rhetorical.

*NOTE :* *The placeholder* ***xxx*** *will not output to the DyNetML file as a concept.*

Create a DyNetML file

Now it's time to generate the DyNetML. From the Pull Down Menu select *Generate => Concept List => Concept Network (Per Text)* for separate DyNetML files or *Generate => Concept List => Concept Network (Union Only)* to create one file with concepts from all files. AutoMap will output **XML file(s)** usable directly in ORA. You will directed to select the destination folder for these file(s).

NOTE : *When processing multiple files and selecting the* ***Per Text*** *function AutoMap will ask if you want to create a Union of all Semantic Network files.*

The DyNetML file(s) will contain one NodeClass of Concepts. After loading into ORA Nodes can be separated into individual NodeClasses and links can be created to form Networks.

First Run with the Script
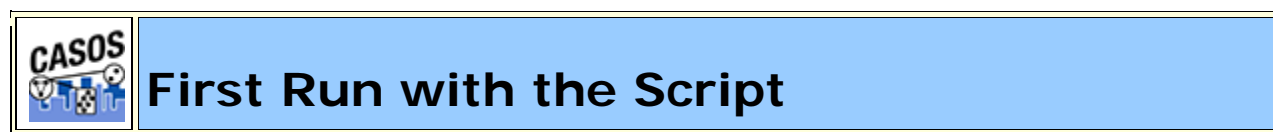
Description

All of AutoMap's functions can be accessed through the script. The two required files are the AM3Script (The AutoMap program) and a .config file (designed by the user). Additional files could include Delete Lists, Thesauri, or other list files necessary by the program.

Create a Workspace

A good starting point is creating a project directory, a place where all your input (your text files), output (files AutoMap writes), and support (required files by certain functions) files will reside. This helps prevent files from getting lost. One

suggestion is to create a top level project directory then create input, output, and support directories within that directory.

```
C:\My Documents\dave\project\input
C:\My Documents\dave\project\output
C:\My Documents\dave\project\support
```

The .config file

Find the blank .config file in the AutoMap directory and make a copy. Rename this to something regarding your project. Open it in your text editor to begin editing the file. The blank .config file will appear as below.

```xml
<?xml version="1.0" encoding="UTF-8"?>
<Script>
<Settings>
   AutoMap textDirectory="" tempWorkspace="" textEncoding=""/>
</Settings>

<Utilities>
<PreProcessing>

</PreProcessing>
<Processing>

</Processing>
<PostProcessing>

</PostProcessing>
</Utilities>
</Script>
```

Initial Setup

The first thing to do is tell AutoMap where your input files are and where you want the output files to be written.

```xml
<Settings>
   AutoMap textDirectory="C:\My Documents\dave\project\input"
   tempWorkspace="C:\My Documents\dave\project\output"
   textEncoding="" />
</Settings>
```

PreProcessing Functions

Now decide which functions of AutoMap you need to run on your files. These are divided into three areas: **Preprocessing, Processing, and PostProcessing**. Review the documentation on the various functions to decide which functions you need to run on your text.

A Generalization Thesaurus

Usually a Generalization Thesaurus is the first file to create. This can be done in either a text editor or spreadsheet. Create a list of single/multi word concepts from the text and the key concepts they should be translated to.

In a **text editor** create each pair on a single line separated by a comma. Make sure to **NOT** leave a space between the comma and the two items.

```
United States of America,United_States_of_America
```

Save this file as a **.csv** file.

In a **spreadsheet program** place the single/multi word concept in the first column and the key concept in the second column.

A	B
United States of America	United_States_of_America

Save this file as a **.csv** file.

In your project .config file in the Preprocessing section insert the command for applying a Generalization Thesaurus. Place the pathway to the newly created thesaurus in the **thesauriLocation** parameter and choose whether to use the **thesauriContentOnly** option.

*NOTE : thesauriContentOnly is set to **y** (put only concepts from the thesaurus in the output file) or **n** (use all concepts form the text files).*

```
<PreProcessing>
   <Generalization thesauriLocation="C:\My
   Documents\dave\project\support\thesauri.csv"
   useThesauriContentOnly="y" />
</PreProcessing>
```

Save the file.

A Delete List

After all the key concepts have been identified it's time to find the **unneeded and unwanted** concepts. A Delete List removes these concepts and reduces the overall number of concepts to analyze. The procedure for applying a Delete List is similar to applying a thesaurus.

In a **text editor** create a list of concepts to be removed from the text. Each line should contain only one concept which consists of a single word. There should be no extra spaces or punctuation included.

```
and
the
but
```

Save this file as a **.csv** file.

In a **spreadsheet program** place each concept to delete in a single cell in the first column

A
and
the
but

Save this file as a **.csv** file.

In your project .config file in the Preprocessing section insert the command for applying a Delete List. Put the pathway to the newly created Delete List in the **deleteListLocation** parameter and choose whether to use the **saveTexts** option.

```
<PreProcessing>
   <Generalization thesauriLocation="C:\My
   Documents\dave\project\support\thesauri.csv"
   useThesauriContentOnly="y" />
   <DeleteList adjacency="r" deleteListLocation="C:\My
   Documents\dave\project\support\deleteList.txt" saveTexts="y"/>
</PreProcessing>
```

Save the file.

Other Preprocessing Functions

Any number of the preprocessing functions can be included in the script file in whatever order you need them. Insert the commands within the **<Preprocessing** tags in the order you need them performed.

NOTE : am3script will perform these function in the order they are placed in the script. Make sure you know what order you need to perform each task.

Processing Functions

Next thing to consider are the steps to take after the preprocessing is finished. These include most of the functions that output files and are based on the text after preprocessing. To run a Processing function the command is placed between the **<Processing>** tags.

Processing functions with no parameters

Processing functions which take no parameters include **Anaphora, ConceptList, UnionConceptList, and NGramExtraction**. Placing these tags between the **<Processing>** tags automatically performs these functions in the order that order.

```
<Processing>
   <ConceptList />
   <UnionConceptList />
</Processing>
```

Processing functions with parameters

All other Processing functions require parameters to set for successful completion. This is either the location of a support file or parameters necessary to complete the function. The parameters follow the tag and are separated by a space and the value is enclosed in quotes.

```
<Processing>
   <SemanticNetworkList directional="U" resetNumber="1"
   textUnit="S" windowsize="5" />
</Processing>
```

If a tag requires a file to work then the pathway needs to be placed in the location parameter. An incorrect pathway will cause the AutoMap function to fail.

```
<Processing>
   <Meta-Network thesauriLocation="C:\My
   Documents\dave\project\support\thesauri.csv" />
</Processing>
```

PostProcessing Functions

The last step is to determine which, if any, postprocessing functions are needed. These are used to alter the DyNetML before running in ORA. They include **addAttribute, addAttribute3Col, and UnionDynetml**. The addAttribute and addAttribute3Col both take an external file with a list of attributes to add. The UnionDynetml requires the **unionType** to create either a **s** semantic or **m** Meta-Network (Carley, 2002) output file.

```
<PostProcessing>
   <addAttribute attributeFile="C:\My
   Documents\dave\project\support\attribute.csv" />
   <unionDynetml unionType="s" />
</PostProcessing>
```

Output files will be written to the designated output directory.

Running the Script

After all the Preprocessing and Processing tags are completed the script can now be run.

Open a Command Window from the Start menu and navigate to the directory containing AM3Script. Move your **.config file** to this directory. At the command prompt type **am3script project.config**. Am3script will execute using the .config specified. Output files will be found in the directory specified in the tempWorkspace parameter.

```
<Settings>
   AutoMap textDirectory="" tempWorkspace="" textEncoding=""/>
</Settings>
```

NOTE : Be sure to leave a space between am3script and the name of your config file.