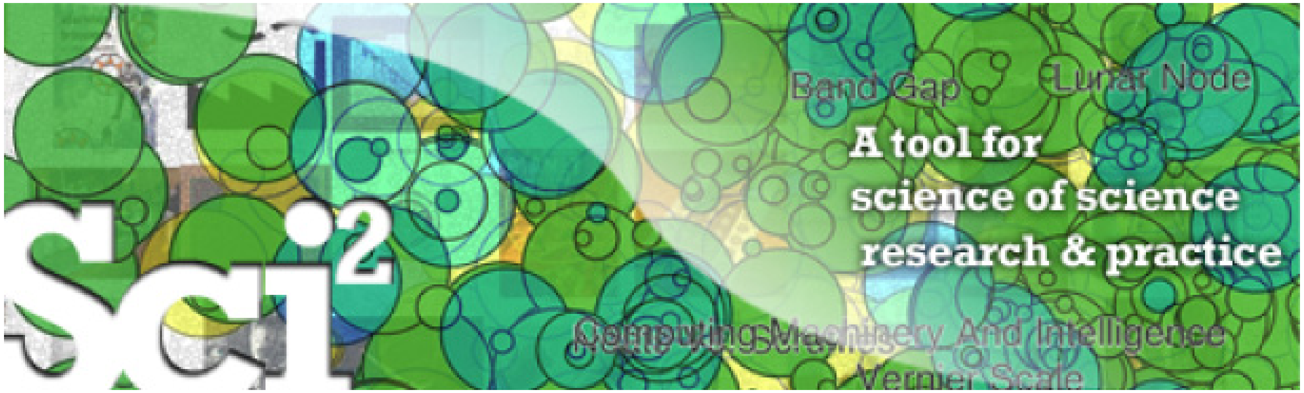
Visualizing Data with the Science of Science (Sci2) Tool



Ted Polley
Research & Editorial Assistant
Cyberinfrastructure for Network Science Center
School of Informatics and Computing
Indiana University Bloomington
http://cns.iu.edu

Presentation Overview

Why should we visualize?

How should we visualize?

Introduction to Sci2

- Introduction
- Macroscopes
- OSGi & Cyberinfrastructure Shell
- Types and levels of analysis
- File formats supported by Sci2
- User Interface
- Supported tools
- Visualizations
- Sci2 Adoption

Brainstorming Session

Break for lunch

Hands-on with Sci2

- Installing Sci2
- Needs Driven Workflow Design
- Introduction to Networks
- Visualizing the Florentine Dataset
- Evolving Co-Authorship Networks
- Visualize Geographic Distribution of Clients
- Visualize Client Brand Networks

Discussion/Questions

Why should we Visualize?

Humans are surprisingly good at:

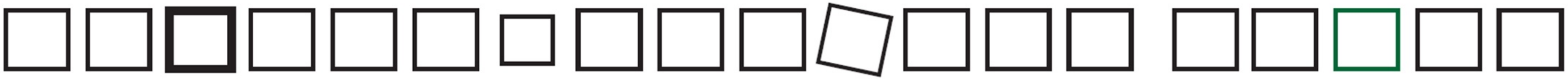
- Pattern matching – especially when it comes to identifying trends, gaps, and outliers. We can use this to predict the future.

- Determining placement, orientation, shape, size, color etc.

Ultimately, visualizations provide access to insight much quicker than simply examining the raw data, or even data that has been statistically analyzed. This insight can lead to informed decision making.

Adapted from Noah Iliinsky's Keynote address at the OCLC Symposium, *Four Pillars of Data Visualization*, at ACRL 2013 in Indianapolis, Indiana.

Why should we Visualize?



How should we Visualize?

Step 1. **Purpose**: identify a clear purpose and focus for the visualization

Step 2. **Content**: obtain the best possible data that suits the purpose

Step 3. **Structure**: identify the best structure for the visualization that suits the purpose

Step 4. **Format**: identity the best way to represent the data so that it is easily digested

Adapted from Noah Iliinsky's Keynote address at the OCLC Symposium, *Four Pillars of Data Visualization*, at ACRL 2013 in Indianapolis, Indiana.

Introduction to Sci2

The Science of Science (Sci2) Tool is an open-source modular toolset originally designed for the study of science. However it has many uses that support temporal, geospatial, topical, and network analysis and visualization of scholarly datasets.

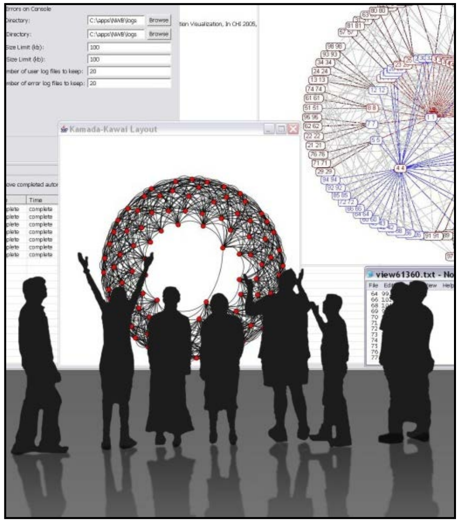
Macroscopes

Decision making in science, industry, and politics, as well as in daily life, requires that we make sense of the massive amounts of data that result from complex systems.

Rather than making things larger or smaller, **macroscopes let us observe what is too great, slow, or complex for us to comprehend or sometimes even notice.**

Microscopes

Telescopes



Macroscopes

Plug-and-Play Macroscopes

While microscopes and telescopes are physical instruments, macroscopes are **continuously changing bundles of software plugins**

Macroscopes make it easy to

- Simply drop plugins into the tool and they appear in the menu, ready to use
- Sharing algorithm components, tools, or novel interfaces becomes as easy as sharing images on Flickr or videos on YouTube

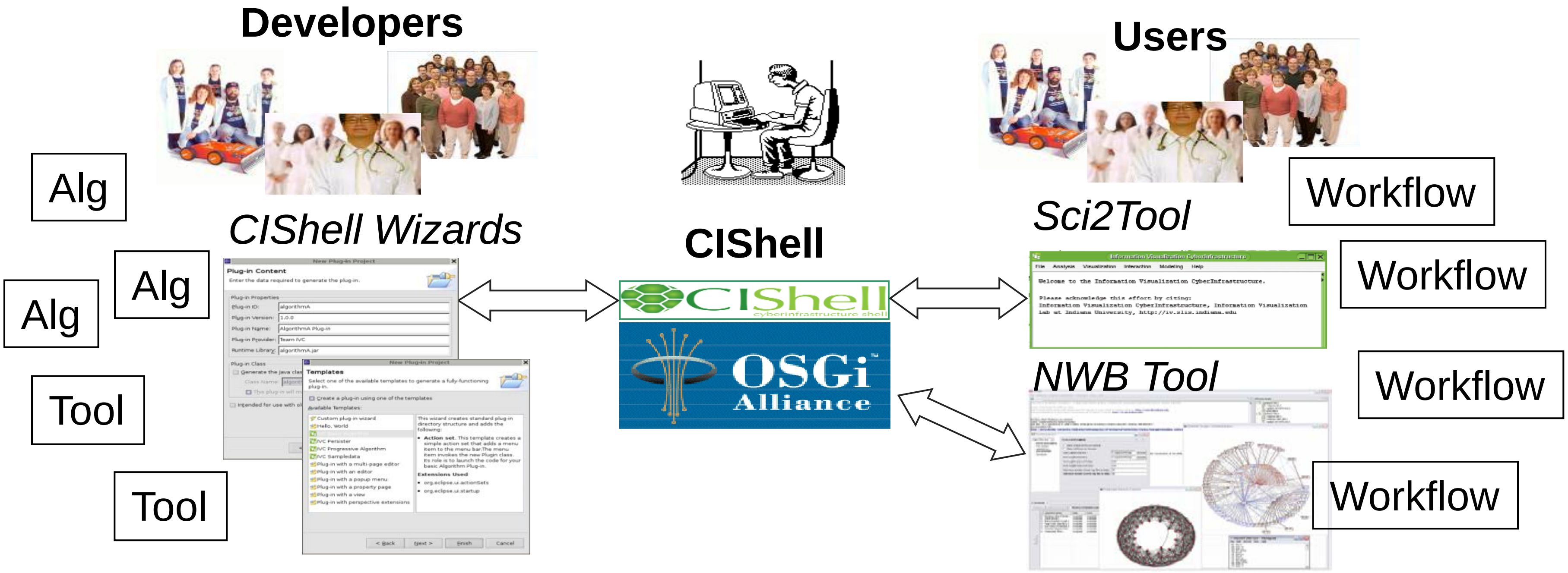
CIShell cyberinfrastructure shell

OSGi & Cyberinfrastructure Shell (CIShell)

- CIShell (http://cishell.org) is an open source software specification for the integration and utilization of datasets, algorithms, and tools
- It extends the Open Services Gateway Initiative (OSGi) (http://osgi.org), a standardized, modularized service platform
- CIShell provides “sockets” into which algorithms, tools, and datasets can be plugged using a wizard-driven process

CIShell cyberinfrastructure shell

OSGi & Cyberinfrastructure Shell (CIShell)



Developers

Alg

Alg

Alg

Tool

Tool

CIShell Wizards

New Plug-in Project

Plug-in Content

Enter the data required to generate the plug-in.

Plug-in Properties

Plug-in ID: algorithmA

Plug-in Version: 1.0.0

Plug-in Name: AlgorithmA Plug-in

Plug-in Provider: Team IVC

Runtime Library: algorithmA.jar

Plug-in Class

□ Generate the java clas

Class Name: algorit

□ This plug-in will m

□ Intended for use with ol

New Plug-in Project

Templates

Select one of the available templates to generate a fully-functioning plug-in.

□ Create a plug-in using one of the templates

Available Templates:

- Custom plug-in wizard
- Hello, World
- IVC Persister
- IVC Progressive Algorithm
- IVC Sampledata
- Plug-in with a multi-page editor
- Plug-in with an editor
- Plug-in with a popup menu
- Plug-in with a property page
- Plug-in with a view
- Plug-in with perspective extensions

This wizard creates standard plug-in directory structure and adds the following:

- **Action set**. This template creates a simple action set that adds a menu item to the menu bar.The menu item invokes the new Plugin class. Its role is to launch the code for your basic Algorithm Plug-in.

Extensions Used

- org.eclipse.ui.actionSets
- org.eclipse.ui.startup

CIShell

CIShell cyberinfrastructure shell

OSGi™ Alliance

Users

Sci2Tool

Information Visualization CyberInfrastructure

```
Welcome to the Information Visualization CyberInfrastructure.

Please acknowledge this effort by citing:
Information Visualization CyberInfrastructure, Information Visualization
Lab at Indiana University, http://iv.slis.indiana.edu
```

NWB Tool

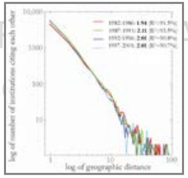
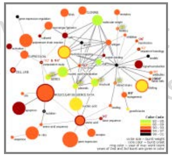
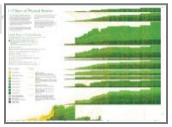
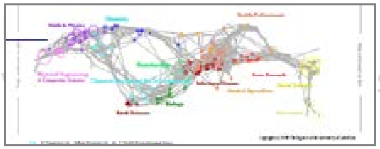
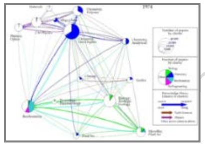
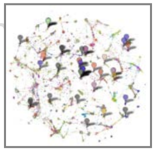
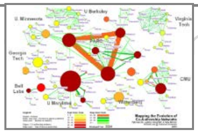
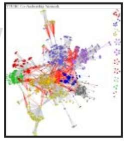
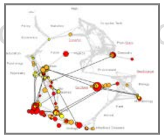
Workflow

Workflow

Workflow

Workflow

Type of Analysis vs. Level of Analysis

	Micro/Individual (1-100 records)	***Meso/Local (101–10,000 records)***	***Macro/Global (10,000 < records)***
Statistical Analysis/Profiling	Individual person and their expertise profiles	Larger labs, centers, universities, research domains, or states	All of NS  A, all of scienc
Temporal Analysis (When)	Funding portfolio of one individual	M  opic bursts i s of PNAS	113 Years of Ph  Research
Geospatial Analysis (Where)	Career trajectory of one individual	Mapping a sta intellectual la pe	PNAS
Topical Analysis (What)	 s.	 flows in research	VxOrd/Topic m  NIH funding
Network Analysis (With Whom?)	NSF C rk of one in	 U. Minnesota U Berkeley Virginia Tech Georgia Tech PARC Bell Labs CMU U Maryland Mapping the Evolution of Co-Authorship Networks netw 	NIH's core c  cy

Sci2 Tool – Supported Data Formats

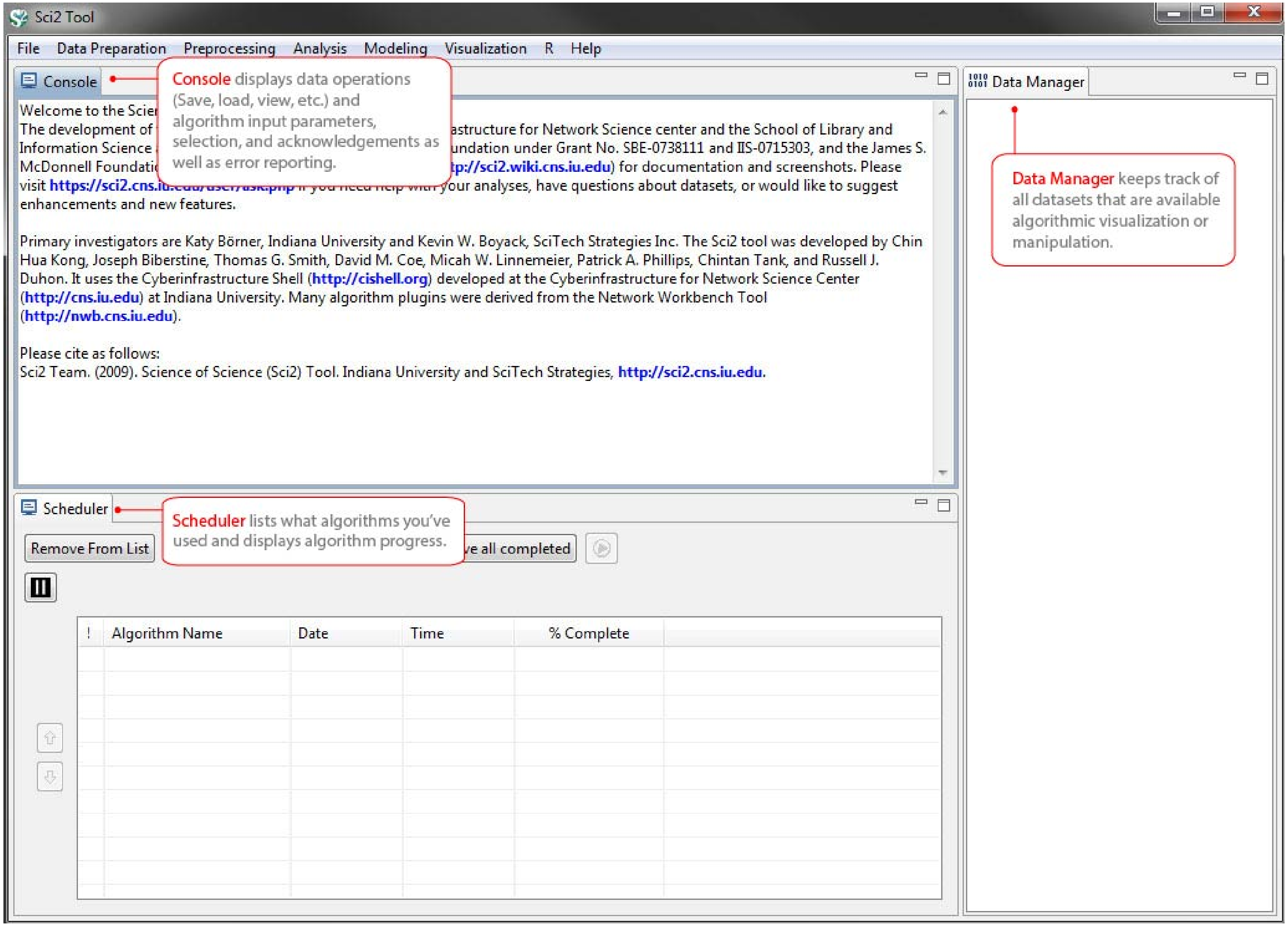
Input:

- Network Formats
- GraphML (*.xml or *.graphml)
- XGMML (*.xml)
- Pajek .NET (*.net)
- NWB (*.nwb)
- Scientometric Formats
- ISI (*.isi)
- Bibtex (*.bib)
- Endnote Export Format (*.enw)
- Scopus csv (*.scopus)
- NSF csv (*.nsf)
- Other Formats
- Pajek Matrix (*.mat)
- TreeML (*.xml)
- Edgelist (*.edge)
- CSV (*.csv)

Output:

- Network File Formats
- GraphML (*.xml or *.graphml)
- Pajek .MAT (*.mat)
- Pajek .NET (*.net)
- NWB (*.nwb)
- XGMML (*.xml)
- CSV (*.csv)
- JPEG (*.jpg)
- PDF (*.pdf)
- PostScript (*.ps)

User Interface



Welcome to the Scie
The development of astructure for Network Science center and the School of Library and
Information Science undation under Grant No. SBE-0738111 and IIS-0715303, and the James S.
McDonnell Foundati tp://sci2.wiki.cns.iu.edu) for documentation and screenshots. Please
visit https://sci2.cns.iu your analyses, have questions about datasets, or would like to suggest
enhancements and new features.

Primary investigators are Katy Börner, Indiana University and Kevin W. Boyack, SciTech Strategies Inc. The Sci2 tool was developed by Chin Hua Kong, Joseph Biberstine, Thomas G. Smith, David M. Coe, Micah W. Linnemeier, Patrick A. Phillips, Chintan Tank, and Russell J. Duhon. It uses the Cyberinfrastructure Shell (http://cishell.org) developed at the Cyberinfrastructure for Network Science Center (http://cns.iu.edu) at Indiana University. Many algorithm plugins were derived from the Network Workbench Tool (http://nwb.cns.iu.edu).

Please cite as follows:
Sci2 Team. (2009). Science of Science (Sci2) Tool. Indiana University and SciTech Strategies, http://sci2.cns.iu.edu.

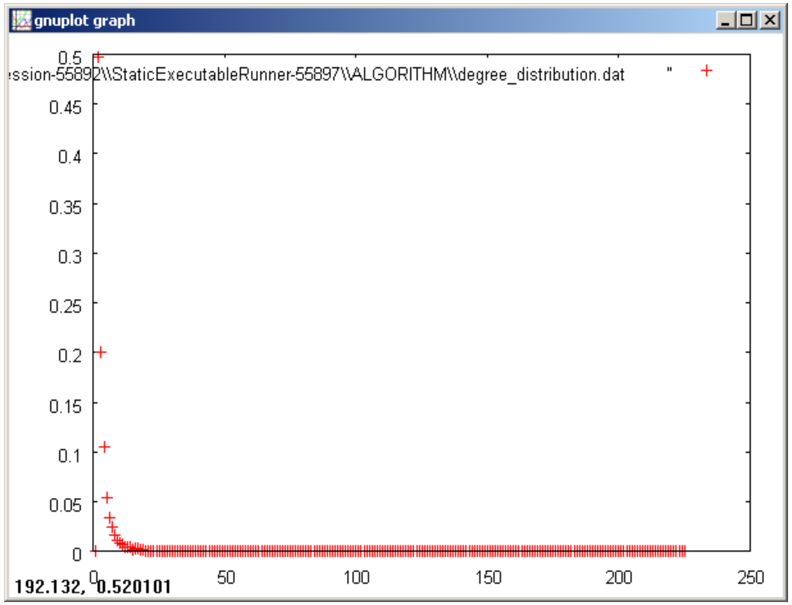
!	Algorithm Name	Date	Time	% Complete

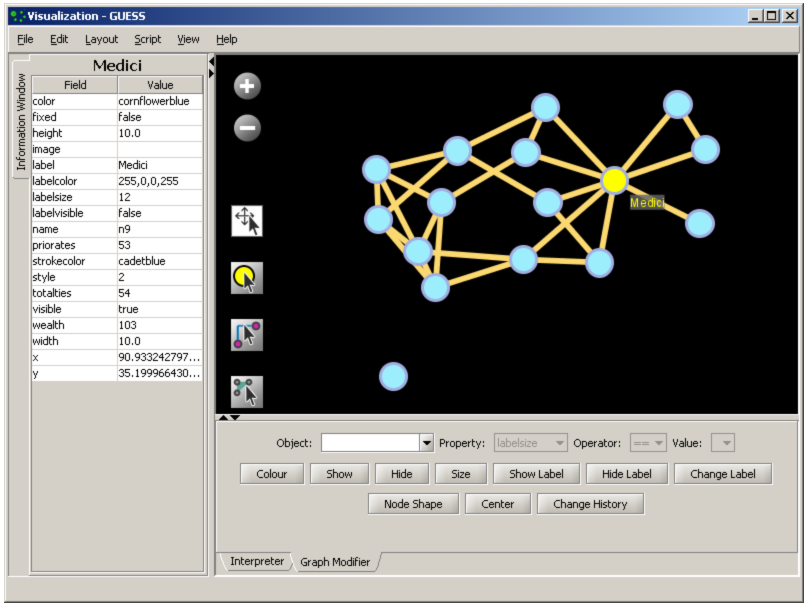
Console displays data operations (Save, load, view, etc.) and algorithm input parameters, selection, and acknowledgements as well as error reporting.

Data Manager keeps track of all datasets that are available algorithmic visualization or manipulation.

Scheduler lists what algorithms you've used and displays algorithm progress.

Supported Tools





Medici

Field	Value
color	cornflowerblue
fixed	false
height	10.0
image	
label	Medici
labelcolor	255,0,0,255
labelsize	12
labelvisible	false
name	n9
priorates	53
strokecolor	cadetblue
style	2
totalties	54
visible	true
wealth	103
width	10.0
x	90.933242797...
y	35.199966430...

Object: Property: labelsize Operator: == Value:

Gnuplot

portable command-line driven interactive data and function plotting utility http://www.gnuplot.info/.

GUESS

exploratory data analysis and visualization tool for graphs and networks.

https://nwb.slis.indiana.edu/community/?n=VisualizeData.GUESS.

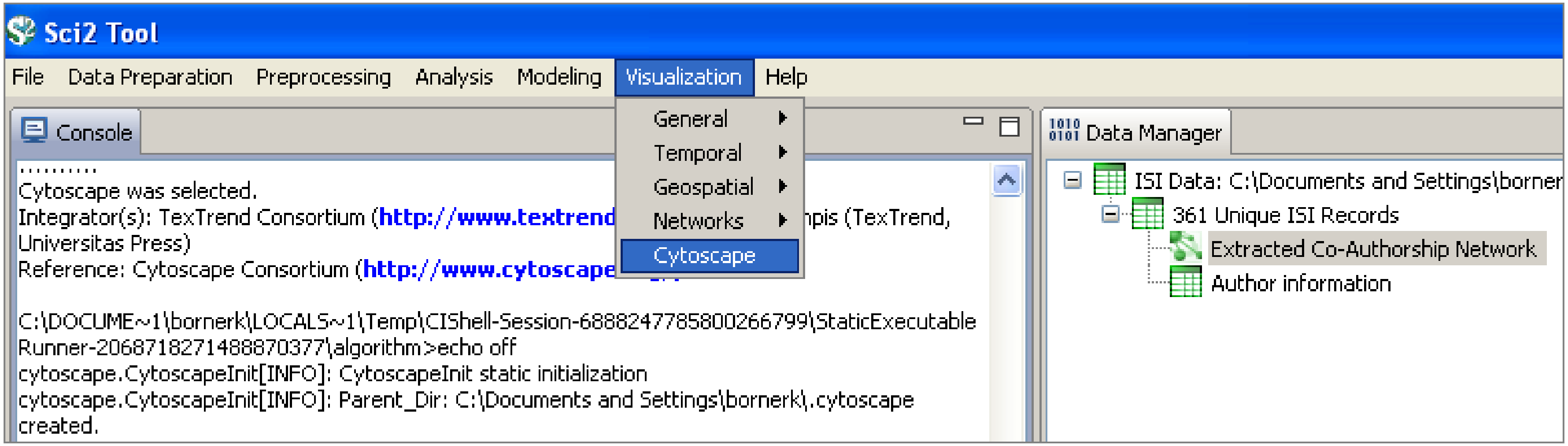
Supported Tools

Adding more layout algorithms and network visualization interactivity via Cytoscape http://www.cytoscape.org.

Simply add *org.textrend.visualization.cytoscape_0.0.3.jar* into your /plugin directory.

Restart Sci^2 Tool

Cytoscape now shows in the Visualization Menu

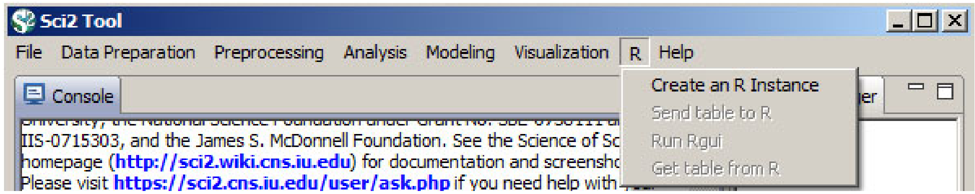


```
..........
Cytoscape was selected.
Integrator(s): TexTrend Consortium (http://www.textrend         pis (TexTrend,
Universitas Press)
Reference: Cytoscape Consortium (http://www.cytoscape

C:\DOCUME~1\bornerk\LOCALS~1\Temp\CIShell-Session-6888247785800266799\StaticExecutable
Runner-2068718271488870377\algorithm>echo off
cytoscape.CytoscapeInit[INFO]: CytoscapeInit static initialization
cytoscape.CytoscapeInit[INFO]: Parent_Dir: C:\Documents and Settings\bornerk\.cytoscape
created.
```

Select a network in Data Manager, run Cytoscape and the tool will start with this network loaded.

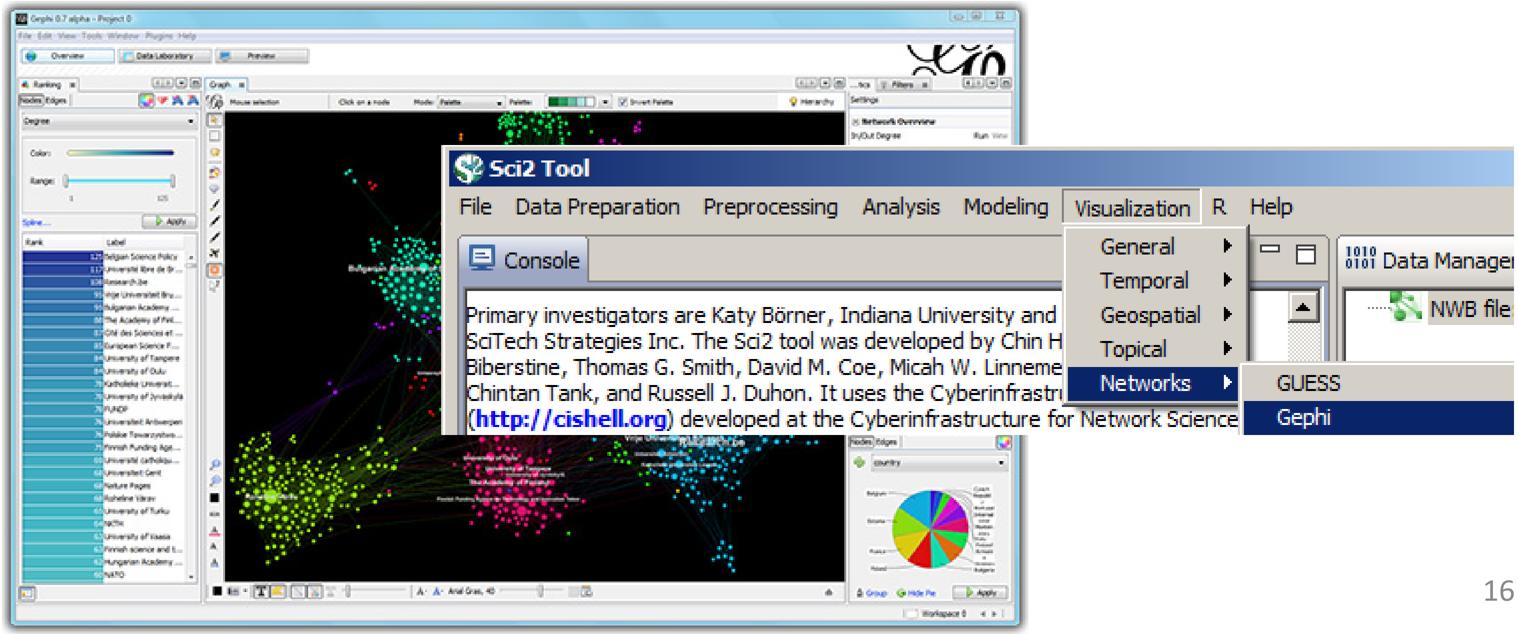
Bridged Tools

R statistical tool



IIS-0715303, and the James S. McDonnell Foundation. See the Science of Sc
homepage (http://sci2.wiki.cns.iu.edu) for documentation and screensho
Please visit https://sci2.cns.iu.edu/user/ask.php if you need help with

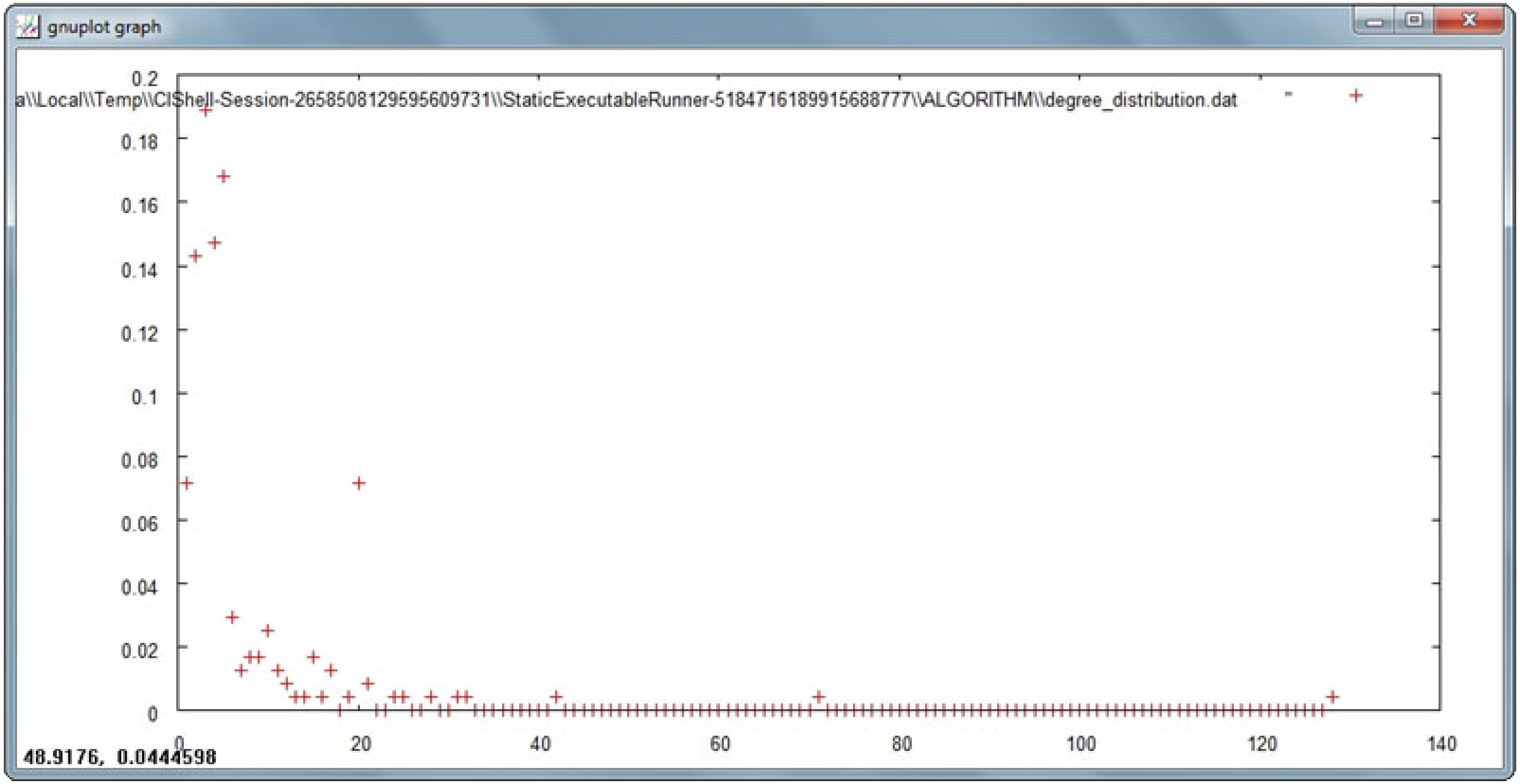
Gephi visualization tool



Primary investigators are Katy Börner, Indiana University and
SciTech Strategies Inc. The Sci2 tool was developed by Chin H
Biberstine, Thomas G. Smith, David M. Coe, Micah W. Linneme
Chintan Tank, and Russell J. Duhon. It uses the Cyberinfrastru
(http://cishell.org) developed at the Cyberinfrastructure for Network Science

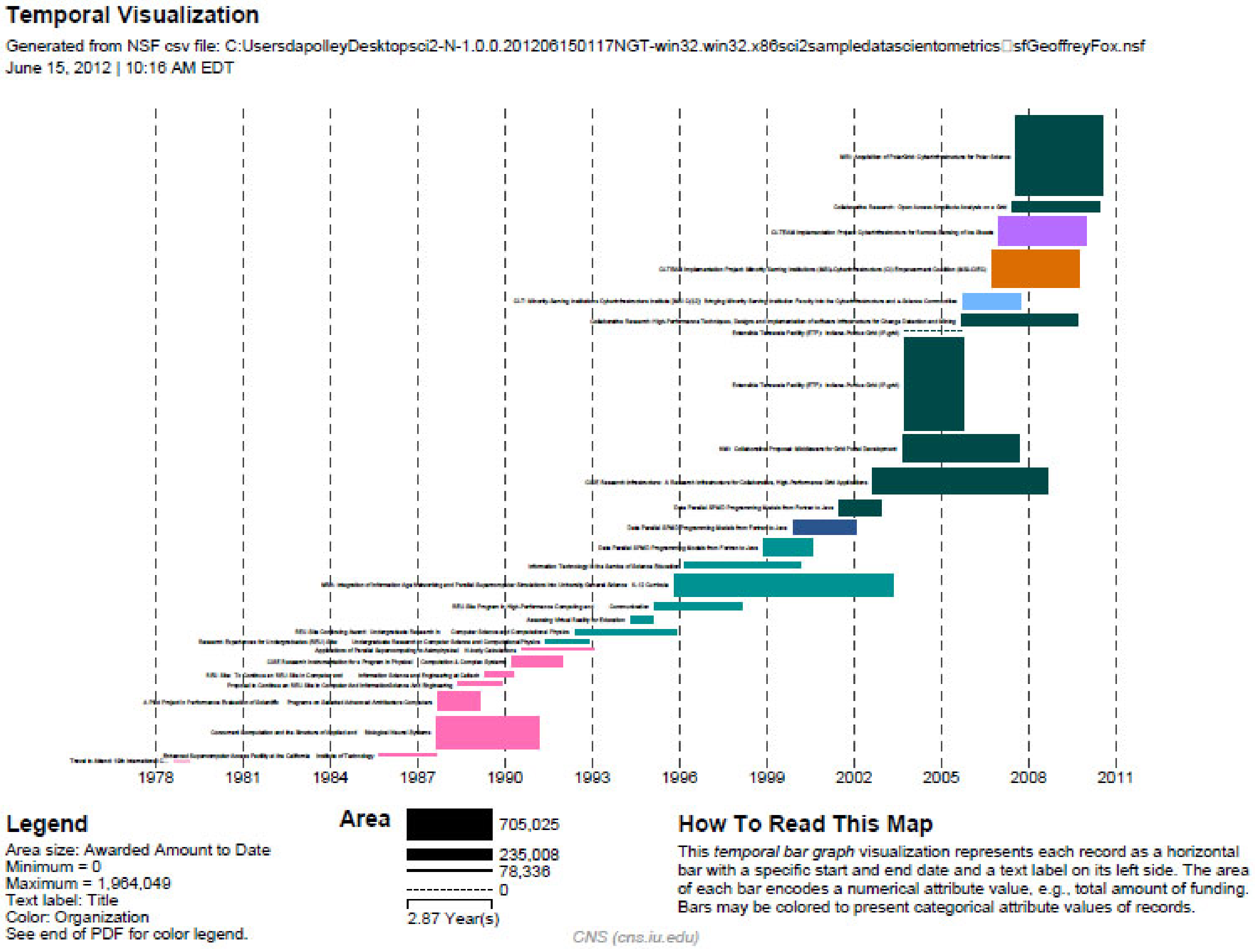
Sci2 Visualizations: *General*

Use GnuPlot to visualize the degree distribution of a co-authorship network extracted from ISI data…


gnuplot graph
a\\Local\\Temp\\CIShell-Session-2658508129595609731\\StaticExecutableRunner-5184716189915688777\\ALGORITHM\\degree_distribution.dat
0.2
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
0
20
40
60
80
100
120
140
48.9176, 0.0444598

Sci2 Tool Visualizations: *Temporal*

Use Temporal Bar Graph to visualize NSF funding profiles over time…



Temporal Visualization

Generated from NSF csv file: C:UsersdapolleyDesktopsci2-N-1.0.0.201206150117NGT-win32.win32.x86sci2sampledatascientometrics□sfGeoffreyFox.nsf
June 15, 2012 | 10:16 AM EDT

Legend

Area size: Awarded Amount to Date
Minimum = 0
Maximum = 1,964,049
Text label: Title
Color: Organization
See end of PDF for color legend.

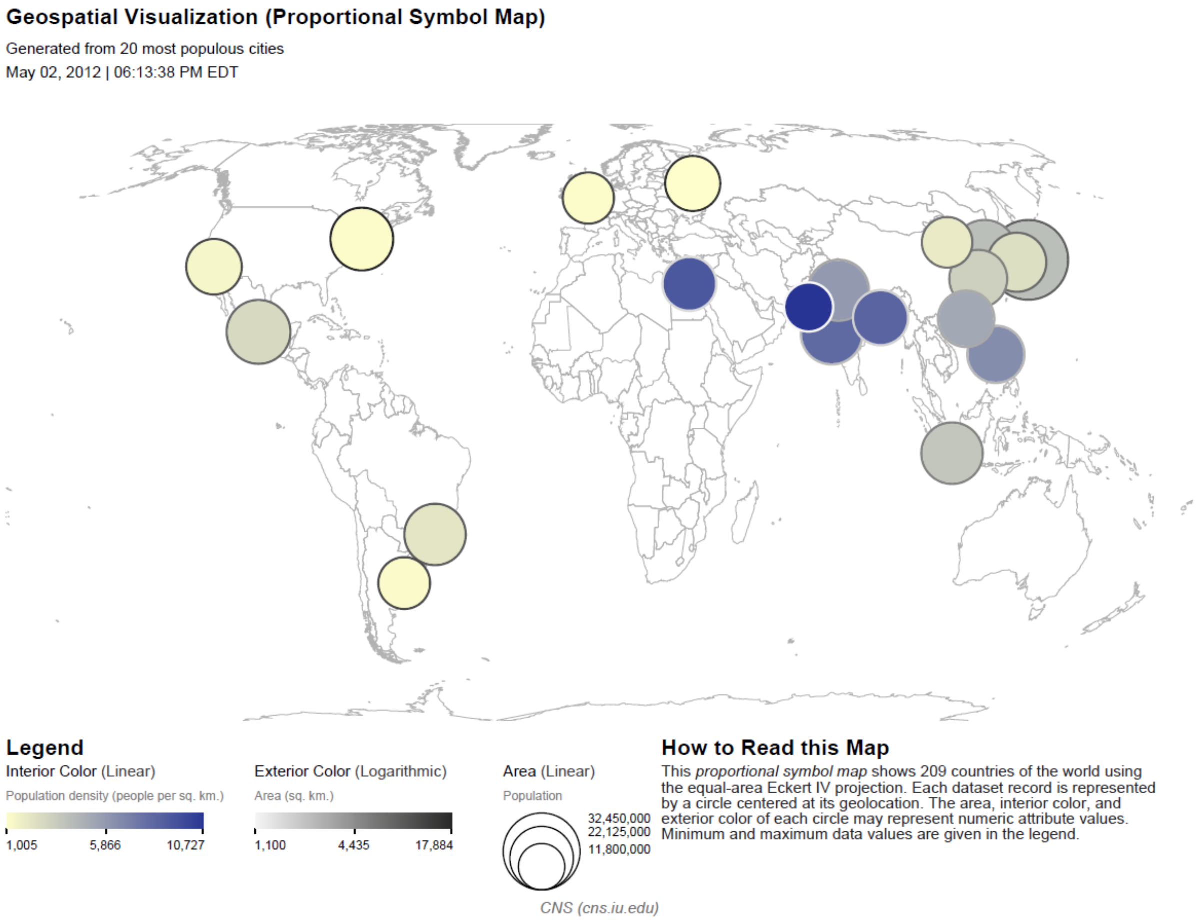
How To Read This Map

This *temporal bar graph* visualization represents each record as a horizontal bar with a specific start and end date and a text label on its left side. The area of each bar encodes a numerical attribute value, e.g., total amount of funding. Bars may be colored to present categorical attribute values of records.

CNS (cns.iu.edu)

Sci2 Tool Visualizations: *Geospatial*

Use the Proportional Symbol Map to size and color symbols proportionally to numeric data, in this case the 20 most populated cities around the world…



Geospatial Visualization (Proportional Symbol Map)

Generated from 20 most populous cities
May 02, 2012 | 06:13:38 PM EDT

Legend

Interior Color (Linear)
Population density (people per sq. km.)

Exterior Color (Logarithmic)
Area (sq. km.)

Area (Linear)
Population

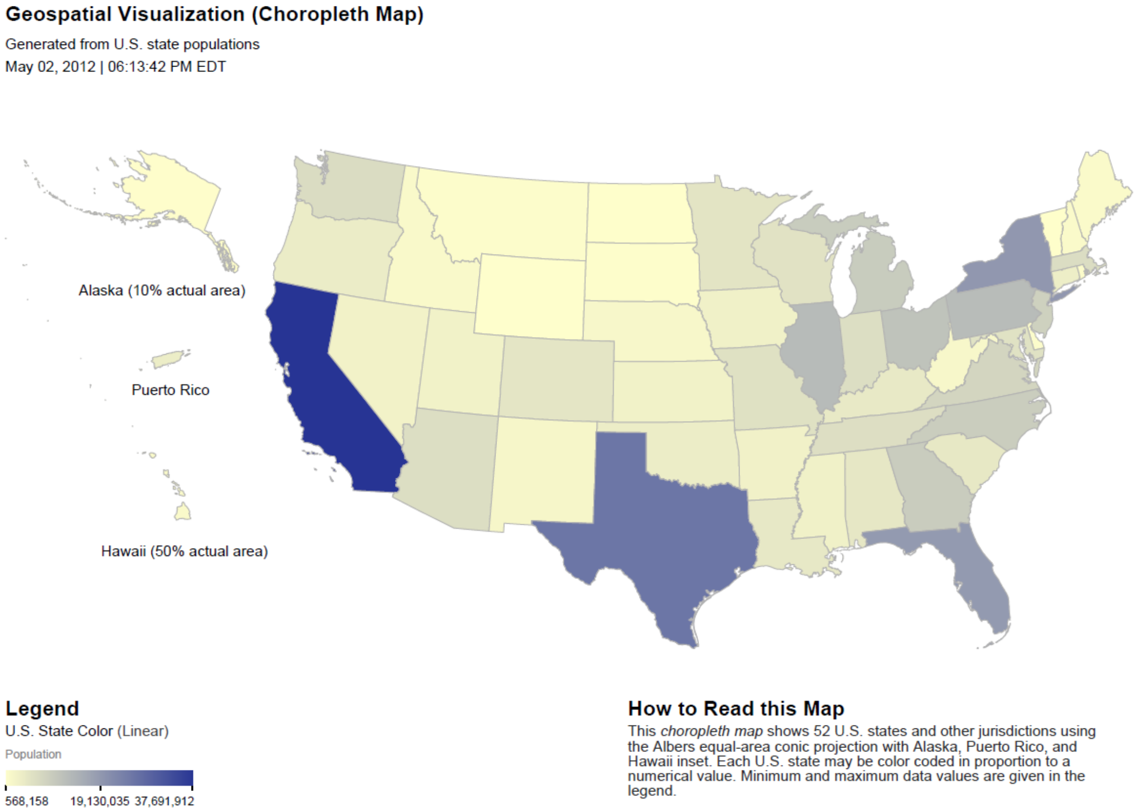
How to Read this Map

This *proportional symbol map* shows 209 countries of the world using the equal-area Eckert IV projection. Each dataset record is represented by a circle centered at its geolocation. The area, interior color, and exterior color of each circle may represent numeric attribute values. Minimum and maximum data values are given in the legend.

CNS (cns.iu.edu)

Sci2 Tool Visualizations: *Geospatial*

Use the Choropleth Map to color regions proportionally to numeric data, in this case the US by state population…



Geospatial Visualization (Choropleth Map)

Generated from U.S. state populations
May 02, 2012 | 06:13:42 PM EDT

Legend

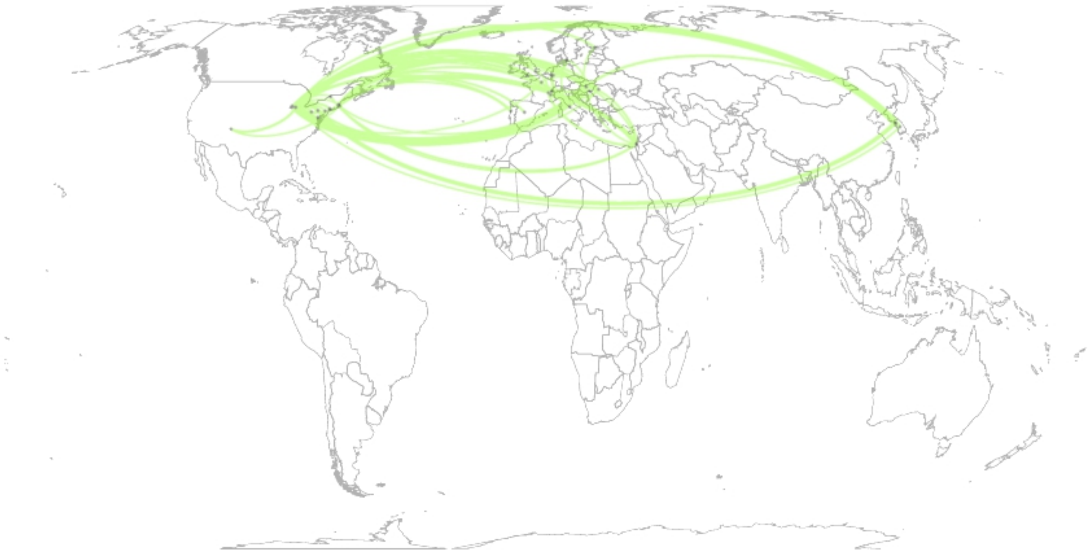
U.S. State Color (Linear)

How to Read this Map

This *choropleth map* shows 52 U.S. states and other jurisdictions using the Albers equal-area conic projection with Alaska, Puerto Rico, and Hawaii inset. Each U.S. state may be color coded in proportion to a numerical value. Minimum and maximum data values are given in the legend.

Sci2 Tool Visualizations: *Geospatial*

Overlay a geospatial network on a base map, in this case Albert-László Barabási and his collaborators…



Geo Map ()

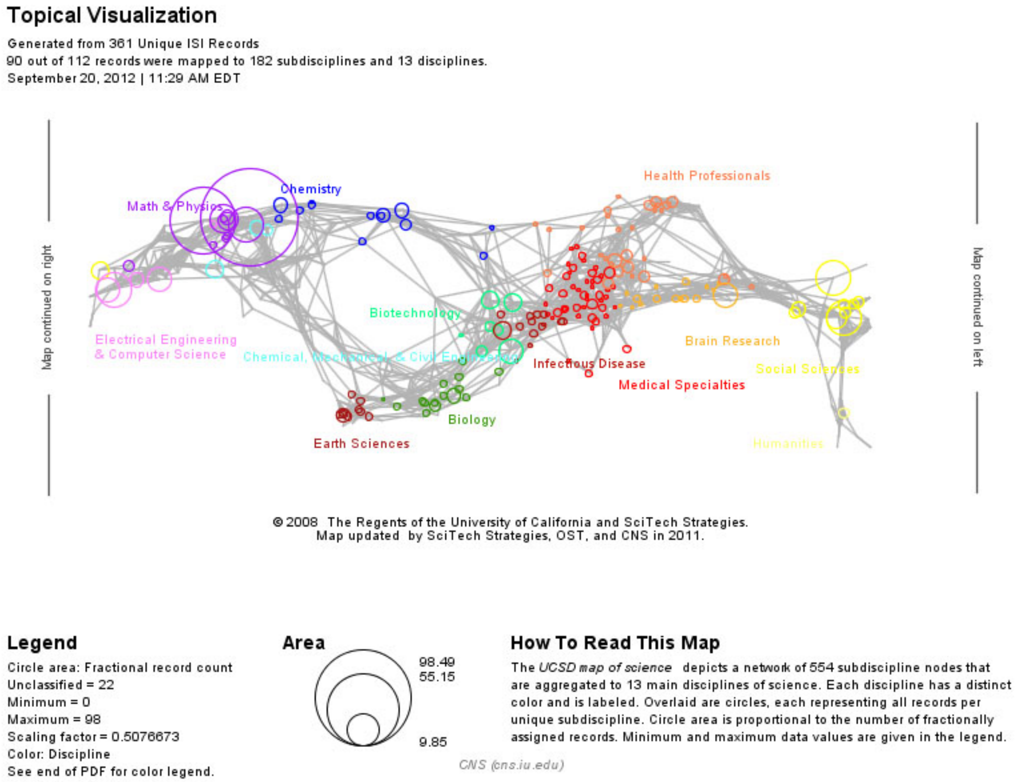
Eckert IV Projection

Apr 19, 2012 | 11:14:48 AM

Created with Sci^2 Tool | Cyberinfrastructure for Network Science Center (http://cns.iu.edu)

Sci2 Tool Visualizations: *Topical*

Use the Map of Science via Journals visualization a network drawn the result of mapping a dataset's journals to the underlying sub-discipline(s) those journals contain…



Topical Visualization

Generated from 361 Unique ISI Records
90 out of 112 records were mapped to 182 subdisciplines and 13 disciplines.
September 20, 2012 | 11:29 AM EDT

Legend

Circle area: Fractional record count
Unclassified = 22
Minimum = 0
Maximum = 98
Scaling factor = 0.5076673
Color: Discipline
See end of PDF for color legend.

Area

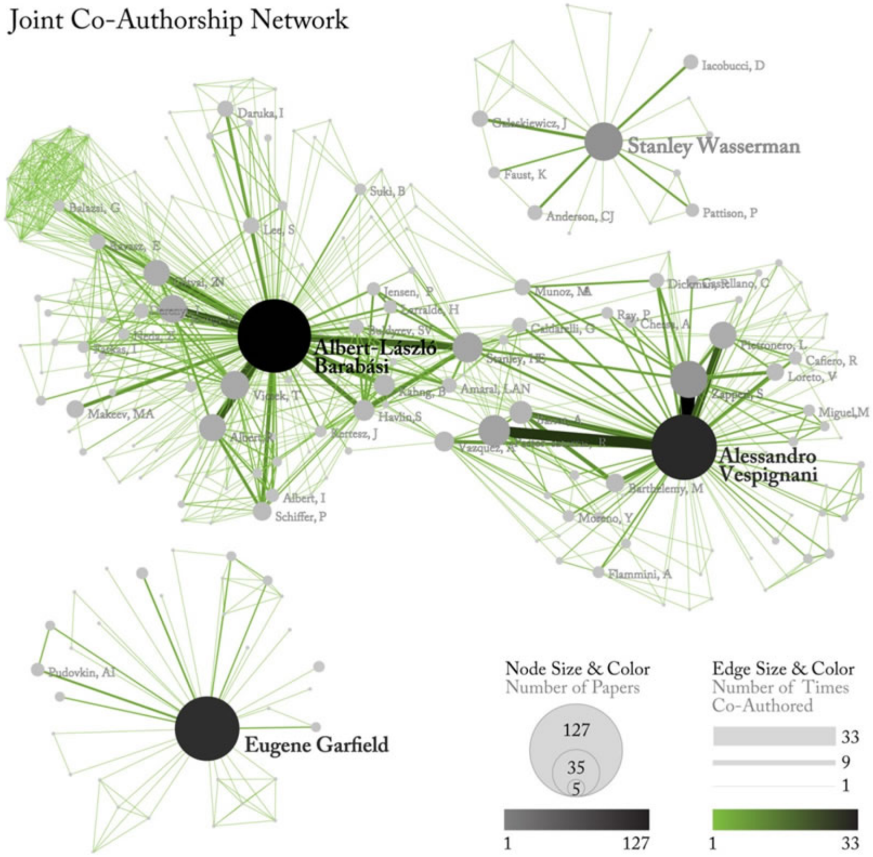
How To Read This Map

The *UCSD map of science* depicts a network of 554 subdiscipline nodes that are aggregated to 13 main disciplines of science. Each discipline has a distinct color and is labeled. Overlaid are circles, each representing all records per unique subdiscipline. Circle area is proportional to the number of fractionally assigned records. Minimum and maximum data values are given in the legend.

CNS (cns.iu.edu)

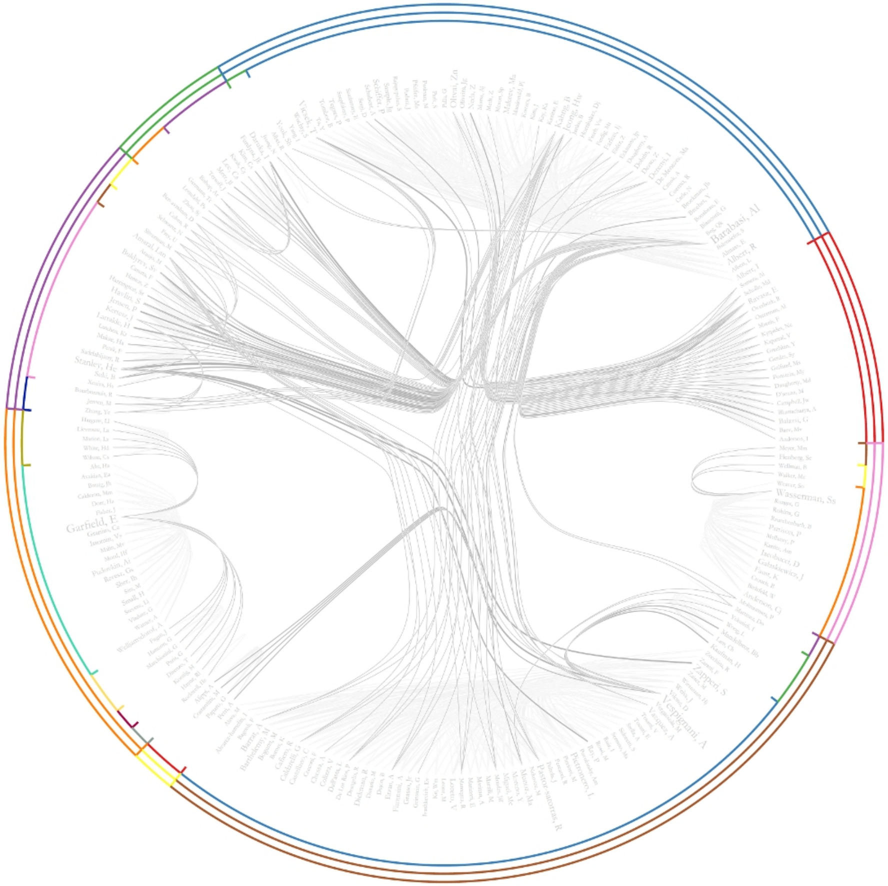
Sci2 Tool Visualizations: *Networks*

Use GUESS to visualize networks, such as this co-authorship network extracted from ISI data…



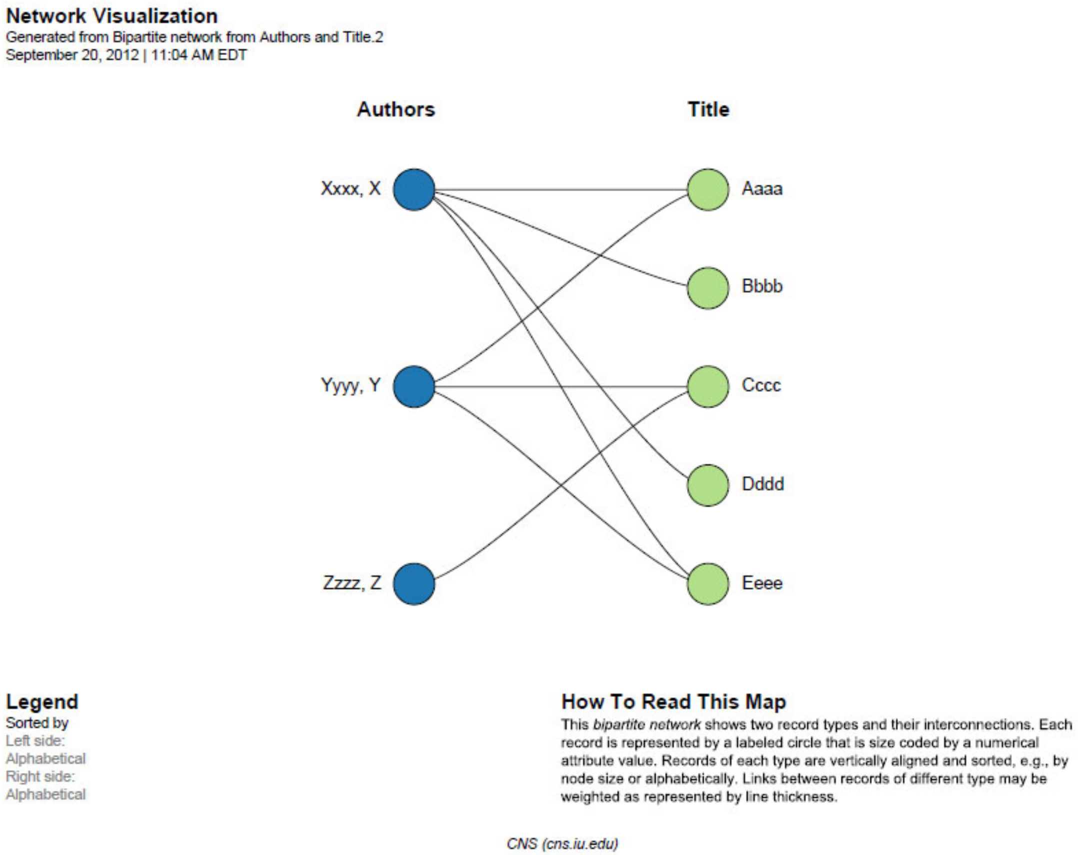
Sci2 Tool Visualizations: *Networks*

Use Circular Hierarchy to visualize networks with community attributes appended…



Sci2 Tool Visualizations: *Networks*

Use the Bipartite Network visualization to create a network of authors and publication titles…



Network Visualization
Generated from Bipartite network from Authors and Title.2
September 20, 2012 | 11:04 AM EDT

Legend
Sorted by
Left side:
Alphabetical
Right side:
Alphabetical

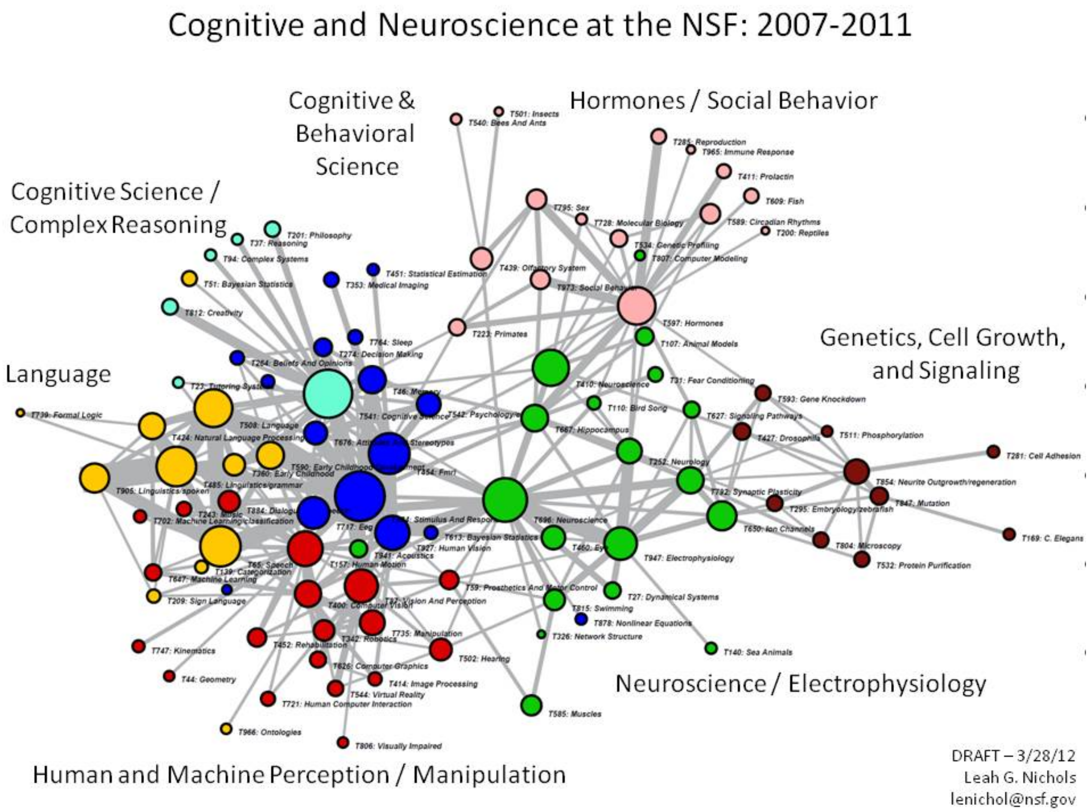
How To Read This Map
This *bipartite network* shows two record types and their interconnections. Each record is represented by a labeled circle that is size coded by a numerical attribute value. Records of each type are vertically aligned and sorted, e.g., by node size or alphabetically. Links between records of different type may be weighted as represented by line thickness.

CNS (cns.iu.edu)

Sci2 Adoption

Topic co-occurrence network of the 2885 cognitive and neuroscience NSF projects funded between 2007 and 2011.

The nodes are labeled based on how the awards were tagged. The nodes are scaled by number of awards (max = 355) with a particular tag and the edges are scaled on number of co-occurrences (max =91) of those tags. The node colors differentiate the different communities of awards, which allows you to identify topic areas.


Cognitive and Neuroscience at the NSF: 2007-2011

DRAFT – 3/28/12
Leah G. Nichols
lenichol@nsf.gov

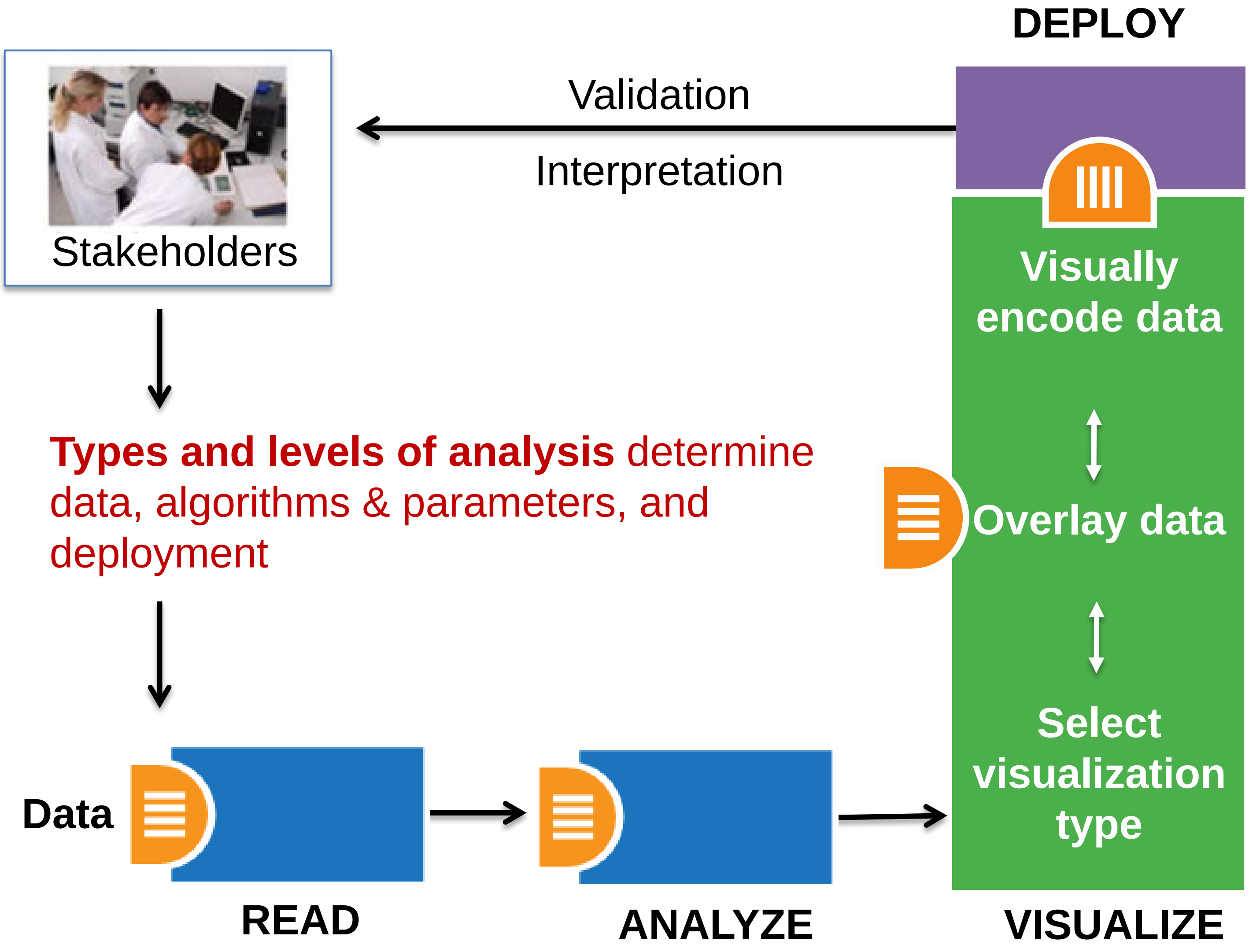
Sci2 Adoption

This is … an **entirely new way of characterizing and understanding the NSF portfolio**. This is in part because this enables **analysis of the content of the awards/proposals independent of the institutional structure.** One can quickly identify ALL of the Cog/Neuro awards throughout the entire NSF portfolio – so it captures research in all of the unexpected institutional places. This method also allows one to **easily identify areas of parallel or potentially collaborative research being funded by different institutional structures** and … to identify potential areas for advancing science by facilitating collaborations.

Leah G. Nichols, NSF

Questions?

Needs-Driven Workflow Design



DEPLOY

Validation

Interpretation

Stakeholders

Types and levels of analysis determine data, algorithms & parameters, and deployment

Data

READ

ANALYZE

VISUALIZE

Visually encode data

Overlay data

Select visualization type

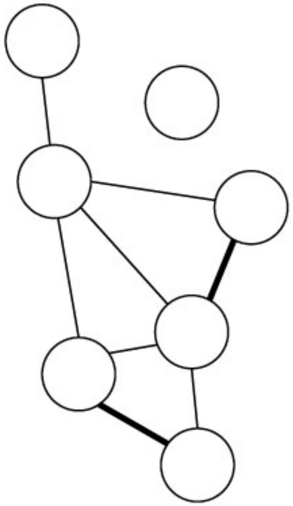
Graphic Variable Types

Modify reference system, add records & links

Visualization Types (reference systems)

Introduction to Networks

Undirected Networks



Nodes:

○

Edges:

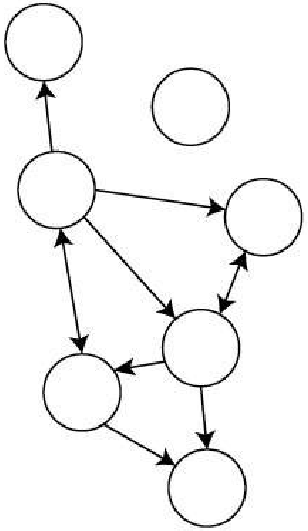
─

Node Degree:
Number of edges connected to nodes

Isolates:
Nodes that are not connected to the rest of the network

Edge Weight:
Demonstrates relative importance of relationships

Directed Networks



Edge Direction:
Directional relationship is represented by arrows

In-Degree:
Number of incoming edges

Out-Degree:
Number of outgoing edges

Visualizing the Florentine Dataset

This example will demonstrate how to visualize data using Sci2. In this workflow we will be working with Padgett's Florentine families dataset which includes 16 different Italian families from the early 15th century. Each family is represented by a node in the network and families are connected by edges that represent either a marriage or business/lending ties. Each node (family) has several attributes: wealth (in thousands of lira), number of priorates (seats on the civic council between 1282-1344), and total ties (total number of business ties and marriages in the dataset).

“Substantively, the data include families who were locked in a struggle for political control of the city of Florence around 1430. Two factions were dominant in this struggle: one revolved around the infamous Medici family, the other around the powerful Strozzis.”

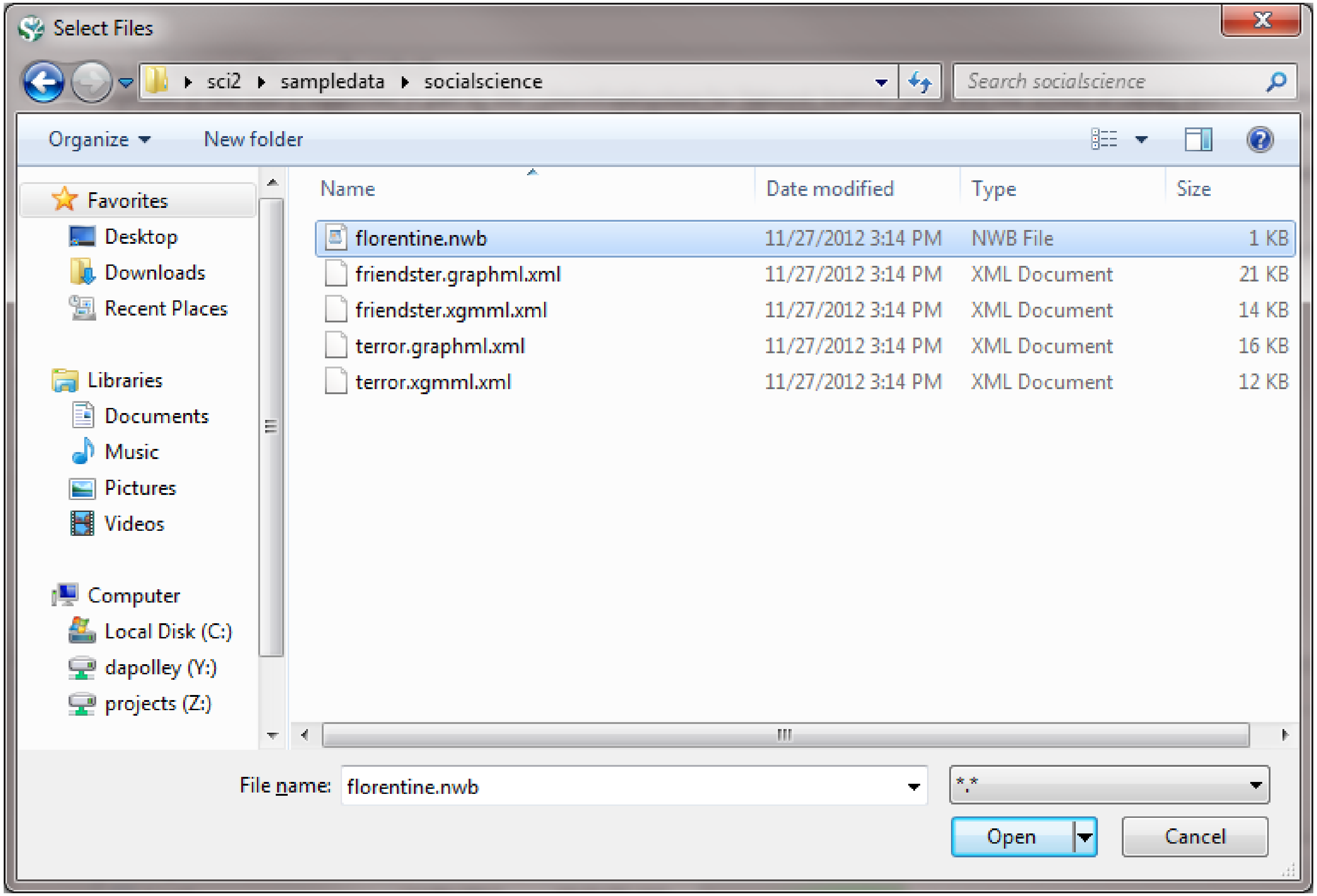
More info at http://svitsrv25.epfl.ch/R-doc/library/ergm/html/florentine.html

Visualizing the Florentine Dataset

```
*Nodes
id*int label*string wealth*int totalities*int priorates*int
1 "Acciaiuoli" 10 2 53
2 "Albizzi" 36 3 65
3 "Barbadori" 55 14 0
4 "Bischeri" 44 9 12
5 "Castellani" 20 18 22
6 "Ginori" 32 9 0
7 "Guadagni" 8 14 21
8 "Lamberteschi" 42 14 0
9 "Medici" 103 54 53
10 "Pazzi" 48 7 0
11 "Peruzzi" 49 32 42
12 "Pucci" 3 1 0
13 "Ridolfi" 27 4 38
14 "Salviati" 10 5 35
15 "Strozzi" 146 29 74
16 "Tornabuoni" 48 7 0
*UndirectedEdges
source*int target*int marriage*string business*string
9 1 "T" "F"
6 2 "T" "F"
7 2 "T" "F"
9 2 "T" "F"
5 3 "T" "T"
```

Visualizing the Florentine Dataset

First, load the florentine.nwb by following *File > Load > **yoursci2directory**/sampledata/scientometrics/endnote/florentine.nwb.*



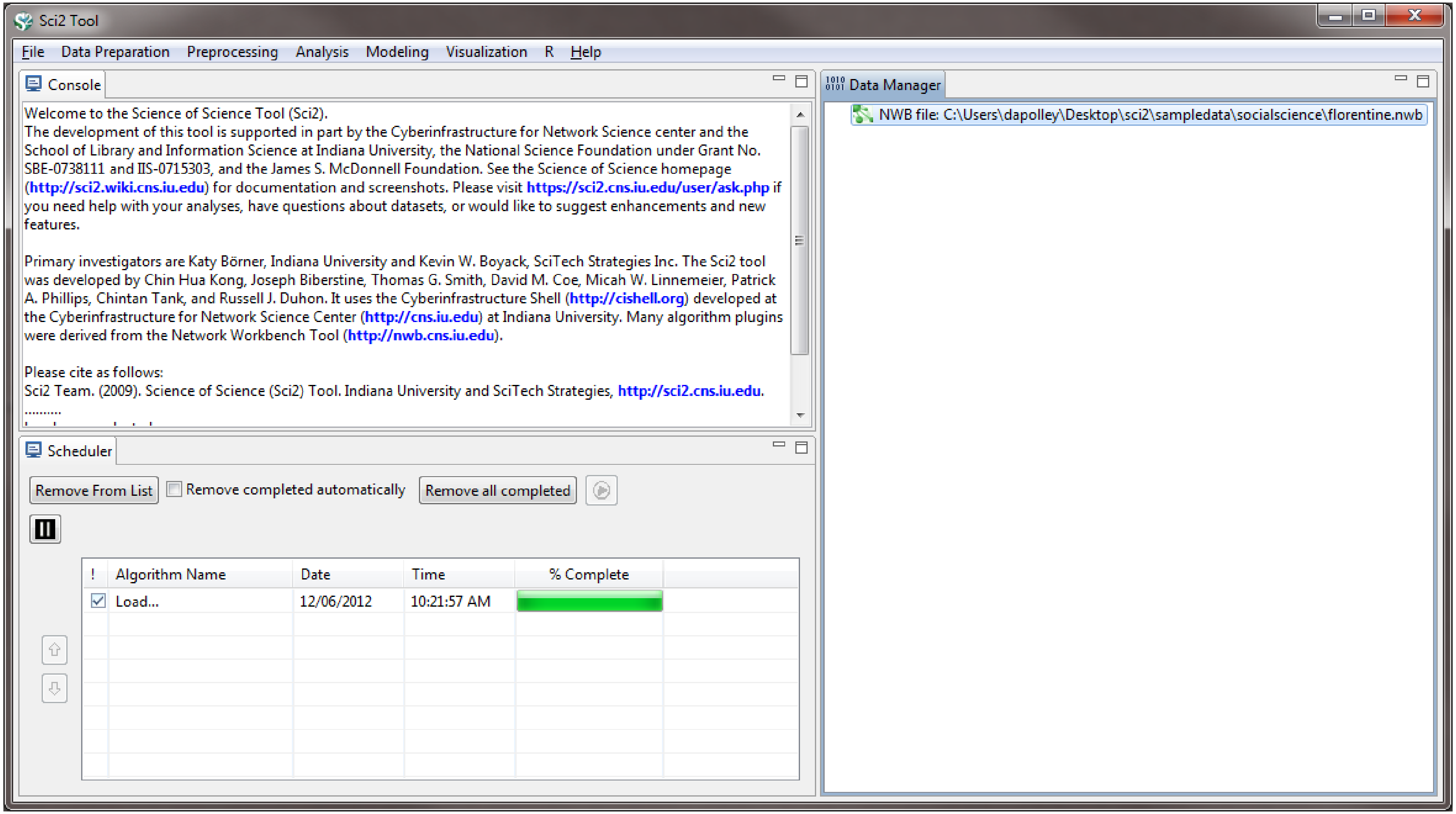
Name	Date modified	Type	Size
florentine.nwb	11/27/2012 3:14 PM	NWB File	1 KB
friendster.graphml.xml	11/27/2012 3:14 PM	XML Document	21 KB
friendster.xgmml.xml	11/27/2012 3:14 PM	XML Document	14 KB
terror.graphml.xml	11/27/2012 3:14 PM	XML Document	16 KB
terror.xgmml.xml	11/27/2012 3:14 PM	XML Document	12 KB

File name: florentine.nwb

.

Visualizing the Florentine Dataset

Once you have loaded the data in Sci2, it will appear in the Data Manager.



Welcome to the Science of Science Tool (Sci2).
The development of this tool is supported in part by the Cyberinfrastructure for Network Science center and the School of Library and Information Science at Indiana University, the National Science Foundation under Grant No. SBE-0738111 and IIS-0715303, and the James S. McDonnell Foundation. See the Science of Science homepage (http://sci2.wiki.cns.iu.edu) for documentation and screenshots. Please visit https://sci2.cns.iu.edu/user/ask.php if you need help with your analyses, have questions about datasets, or would like to suggest enhancements and new features.

Primary investigators are Katy Börner, Indiana University and Kevin W. Boyack, SciTech Strategies Inc. The Sci2 tool was developed by Chin Hua Kong, Joseph Biberstine, Thomas G. Smith, David M. Coe, Micah W. Linnemeier, Patrick A. Phillips, Chintan Tank, and Russell J. Duhon. It uses the Cyberinfrastructure Shell (http://cishell.org) developed at the Cyberinfrastructure for Network Science Center (http://cns.iu.edu) at Indiana University. Many algorithm plugins were derived from the Network Workbench Tool (http://nwb.cns.iu.edu).

Please cite as follows:
Sci2 Team. (2009). Science of Science (Sci2) Tool. Indiana University and SciTech Strategies, http://sci2.cns.iu.edu.
....

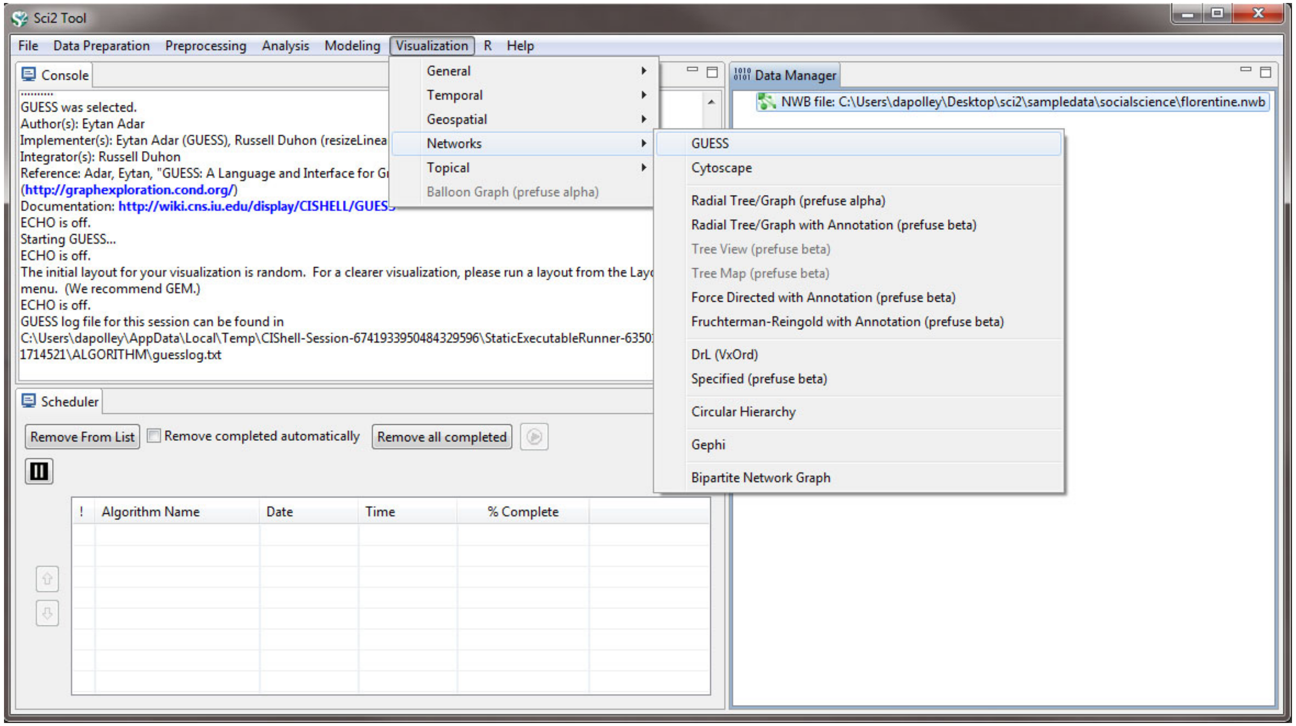
NWB file: C:\Users\dapolley\Desktop\sci2\sampledata\socialscience\florentine.nwb

!	Algorithm Name	Date	Time	% Complete
☑	Load...	12/06/2012	10:21:57 AM	🟩

Visualizing the Florentine Dataset

For this workflow we will skip straight to the visualization step, since the network file that we loaded already has the attributes we are interested in visualizing (wealth, priorates, and totalities). For other datasets, you will likely need to extract networks and run some type of analysis to answer the questions in which you are interested.

To visualize this network select the file from the Data Manager and run *Visualization > Networks > GUESS.*

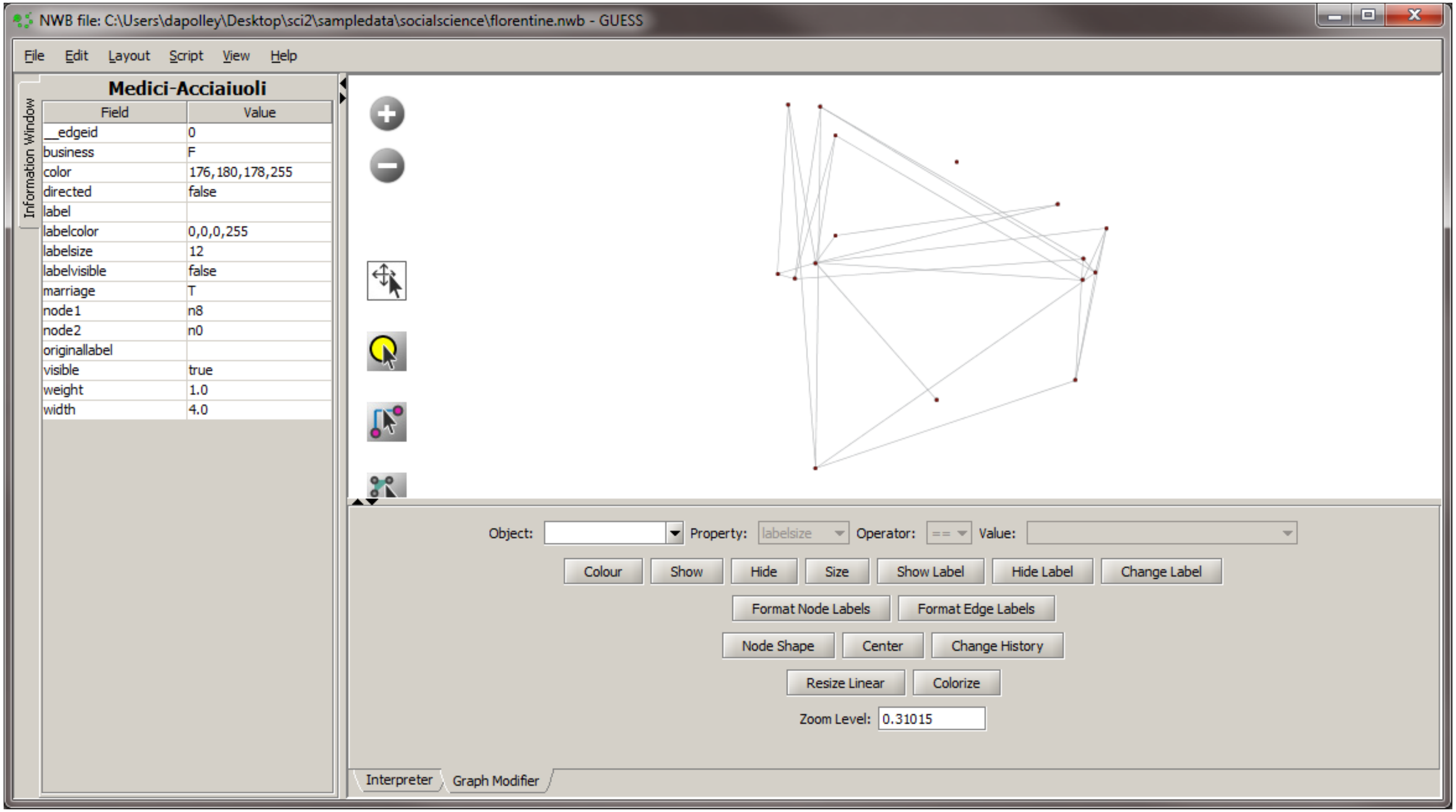


```
..........
GUESS was selected.
Author(s): Eytan Adar
Implementer(s): Eytan Adar (GUESS), Russell Duhon (resizeLinea
Integrator(s): Russell Duhon
Reference: Adar, Eytan, "GUESS: A Language and Interface for Gr
(http://graphexploration.cond.org/)
Documentation: http://wiki.cns.iu.edu/display/CISHELL/GUESS
ECHO is off.
Starting GUESS...
ECHO is off.
The initial layout for your visualization is random.  For a clearer visualization, please run a layout from the Layo
menu.  (We recommend GEM.)
ECHO is off.
GUESS log file for this session can be found in
C:\Users\dapolley\AppData\Local\Temp\CIShell-Session-6741933950484329596\StaticExecutableRunner-6350
1714521\ALGORITHM\guesslog.txt
```

NWB file: C:\Users\dapolley\Desktop\sci2\sampledata\socialscience\florentine.nwb

Visualizing the Florentine Dataset

When the network is loaded in GUESS it will be laid out randomly.



Medici-Acciaiuoli

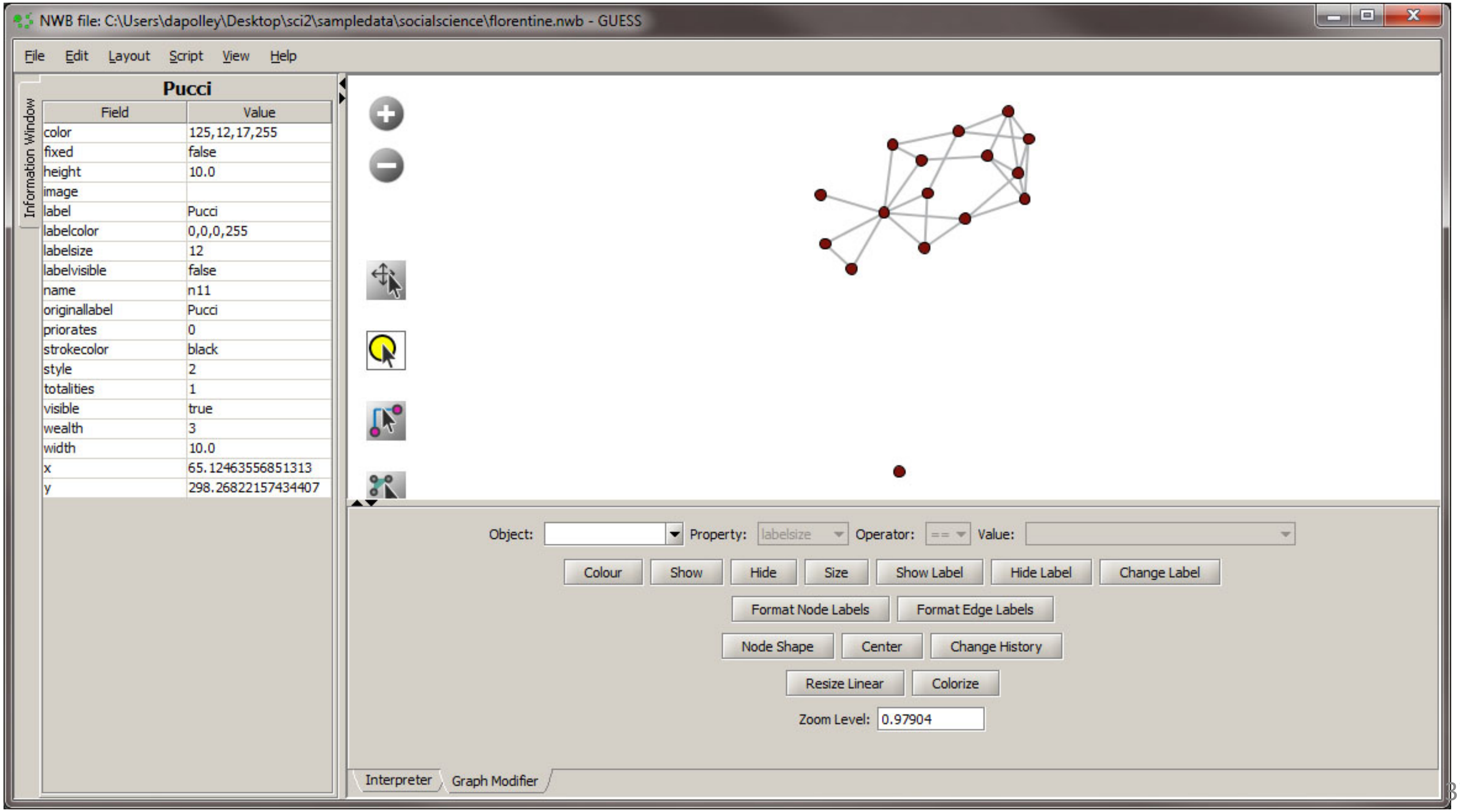
Field	Value
__edgeid	0
business	F
color	176,180,178,255
directed	false
label	
labelcolor	0,0,0,255
labelsize	12
labelvisible	false
marriage	T
node1	n8
node2	n0
originallabel	
visible	true
weight	1.0
width	4.0

Object: Property: labelsize Operator: == Value:

Zoom Level: 0.31015

Visualizing the Florentine Dataset

The first step in enhancing this network visualization is to apply a different layout. For this visualization we will use the GEM layout *Layout > GEM.* You will notice that the GEM layout is random, you can run it multiple times and the network will appear slightly different each time.



NWB file: C:\Users\dapolley\Desktop\sci2\sampledata\socialscience\florentine.nwb - GUESS

Pucci

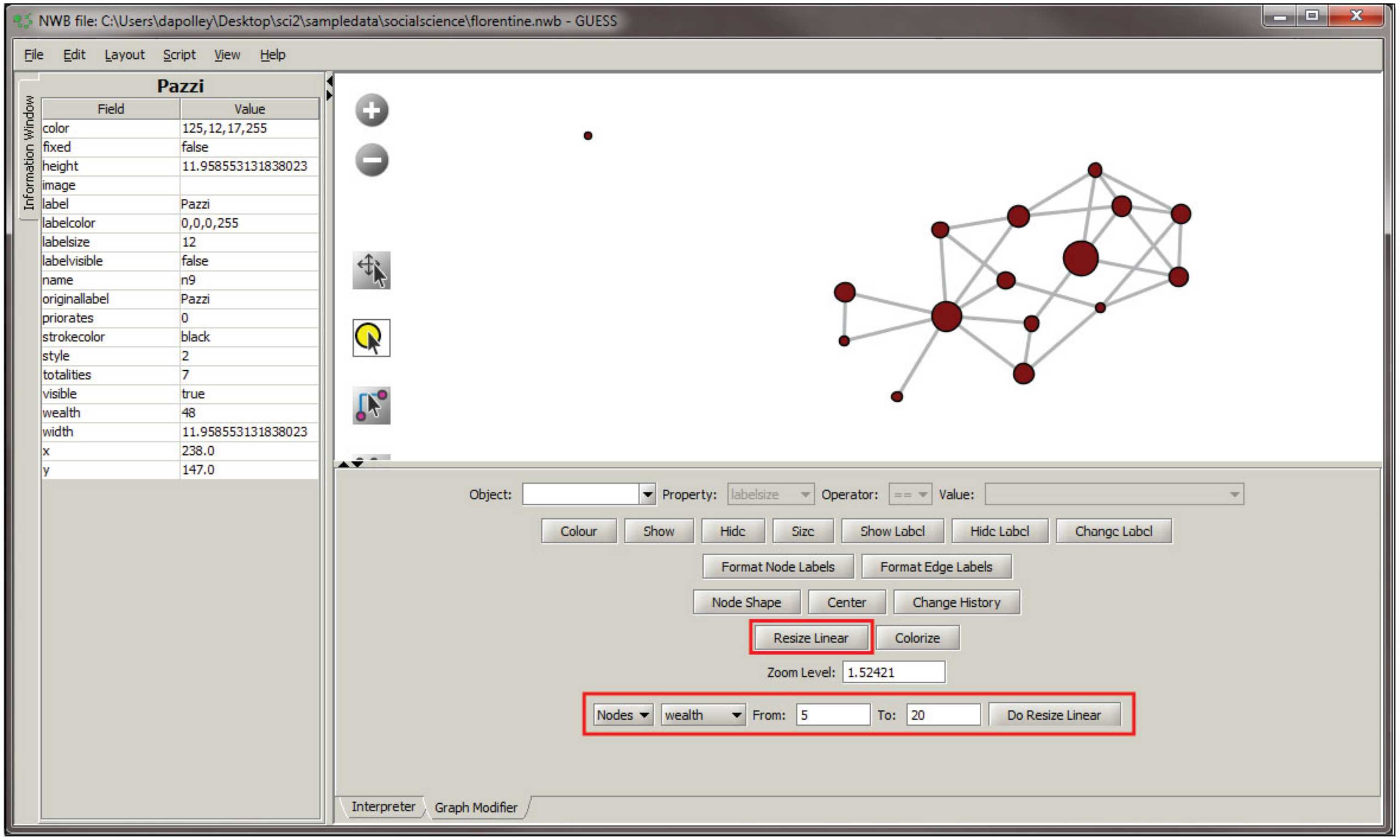
Field	Value
color	125,12,17,255
fixed	false
height	10.0
image	
label	Pucci
labelcolor	0,0,0,255
labelsize	12
labelvisible	false
name	n11
originallabel	Pucci
priorates	0
strokecolor	black
style	2
totalities	1
visible	true
wealth	3
width	10.0
x	65.12463556851313
y	298.26822157434407

Object: Property: labelsize Operator: == Value:

Zoom Level: 0.97904

Visualizing the Florentine Dataset

The next step will be to resize the nodes based on the wealth attribute. To do this resize select the *Resize Linear* button and set the parameters to those shown below.



NWB file: C:\Users\dapolley\Desktop\sci2\sampledata\socialscience\florentine.nwb - GUESS

Pazzi

Field	Value
color	125,12,17,255
fixed	false
height	11.958553131838023
image	
label	Pazzi
labelcolor	0,0,0,255
labelsize	12
labelvisible	false
name	n9
originallabel	Pazzi
priorates	0
strokecolor	black
style	2
totalities	7
visible	true
wealth	48
width	11.958553131838023
x	238.0
y	147.0

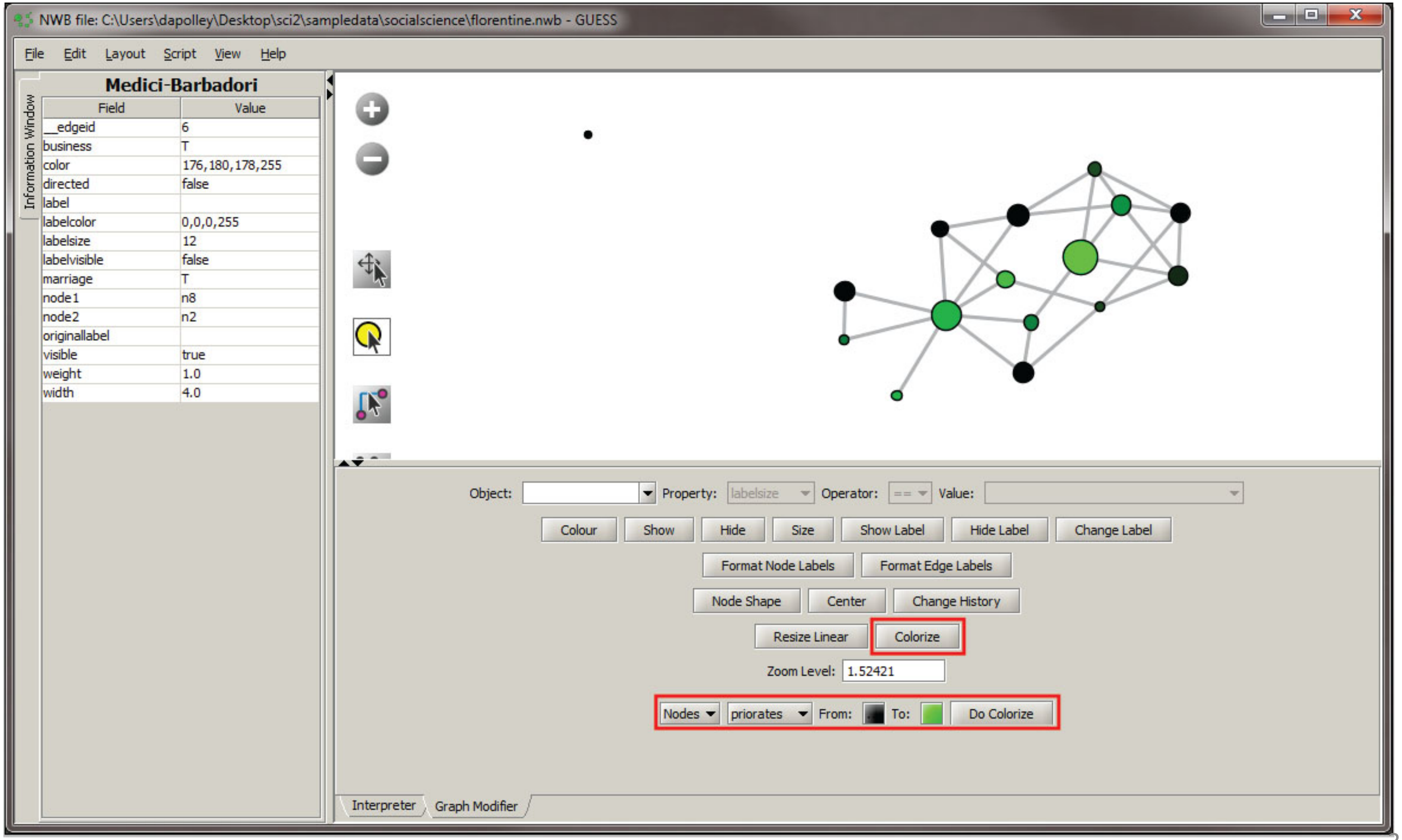
Object: Property: labelsize Operator: == Value:

Zoom Level: 1.52421

Nodes wealth From: 5 To: 20 Do Resize Linear

Visualizing the Florentine Dataset

Next we will colorize the nodes based on priorates to add an additional dimension to this visualization.



NWB file: C:\Users\dapolley\Desktop\sci2\sampledata\socialscience\florentine.nwb - GUESS

Medici-Barbadori

Field	Value
__edgeid	6
business	T
color	176,180,178,255
directed	false
label	
labelcolor	0,0,0,255
labelsize	12
labelvisible	false
marriage	T
node1	n8
node2	n2
originallabel	
visible	true
weight	1.0
width	4.0

Object: Property: labelsize Operator: == Value:

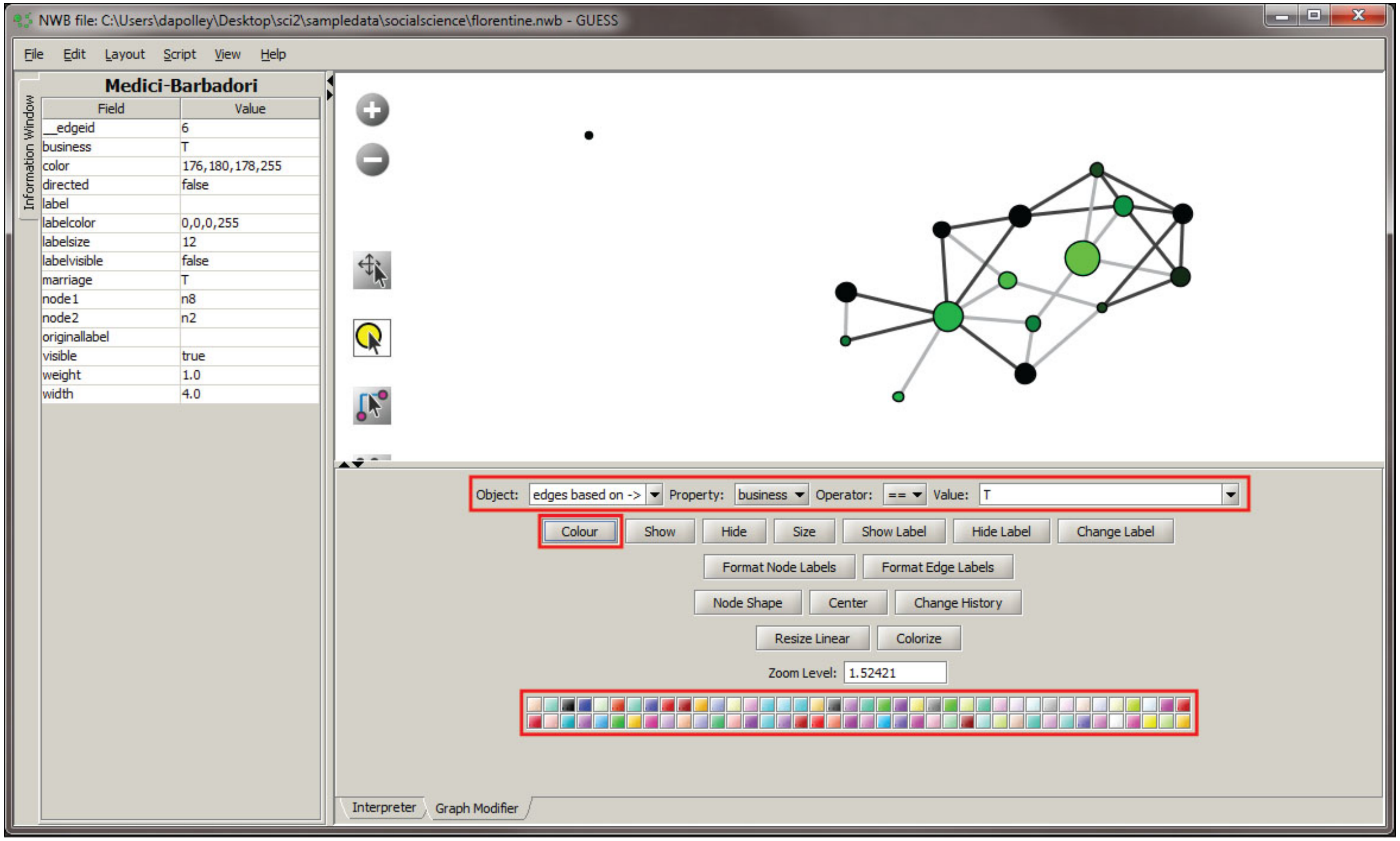
Zoom Level: 1.52421

Nodes priorates From: To: Do Colorize

Cyberinfrastructure for Network Science Center

Visualizing the Florentine Dataset

Next we will color the edges to show the type of relationship between the families. To do this, you will need to select the *Object* edges *based on ->*, set the property to *marriage*, the operator to ==, and the value to *T*. Next, click the *Color* button and you can select the color of your choice from the pallet that will appear at the bottom of the Graph Modifier pane.



NWB file: C:\Users\dapolley\Desktop\sci2\sampledata\socialscience\florentine.nwb - GUESS

Medici-Barbadori

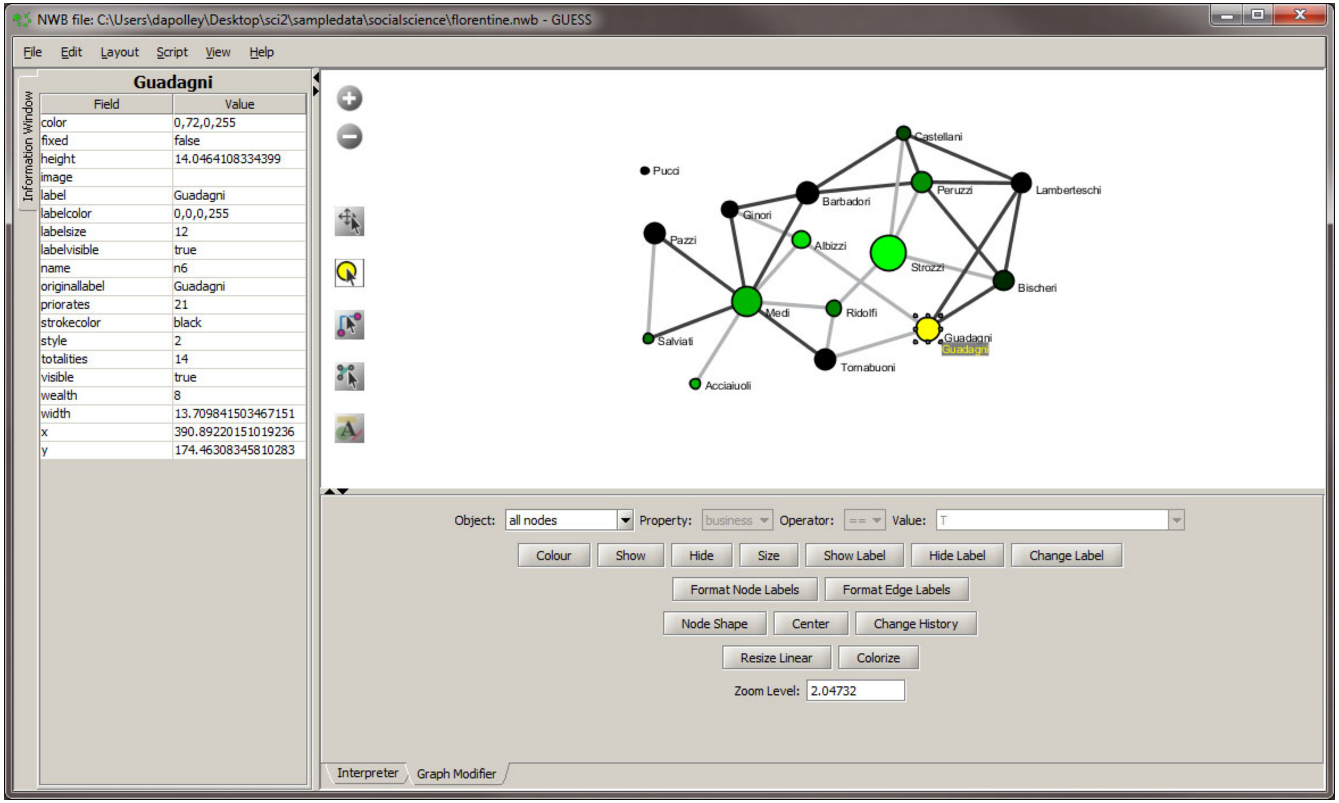
Field	Value
__edgeid	6
business	T
color	176,180,178,255
directed	false
label	
labelcolor	0,0,0,255
labelsize	12
labelvisible	false
marriage	T
node1	n8
node2	n2
originallabel	
visible	true
weight	1.0
width	4.0

Object: edges based on -> Property: business Operator: == Value: T

Zoom Level: 1.52421

Visualizing the Florentine Dataset

You can repeat this process for the *business* property if you want to, or you can leave the edges that represent business ties the default color. In this workflow we will leave them the default color, light gray. The final step is to show all the labels. To do this, you will need to select the "Object" all nodes and the click the *Show Label* button and the labels will appear in the visualization.



NWB file: C:\Users\dapolley\Desktop\sci2\sampledata\socialscience\florentine.nwb - GUESS

Guadagni

Field	Value
color	0,72,0,255
fixed	false
height	14.0464108334399
image	
label	Guadagni
labelcolor	0,0,0,255
labelsize	12
labelvisible	true
name	n6
originallabel	Guadagni
priorates	21
strokecolor	black
style	2
totalities	14
visible	true
wealth	8
width	13.709841503467151
x	390.89220151019236
y	174.46308345810283

Object: all nodes Property: business Operator: == Value: T

Zoom Level: 2.04732

Visualizing the Florentine Dataset

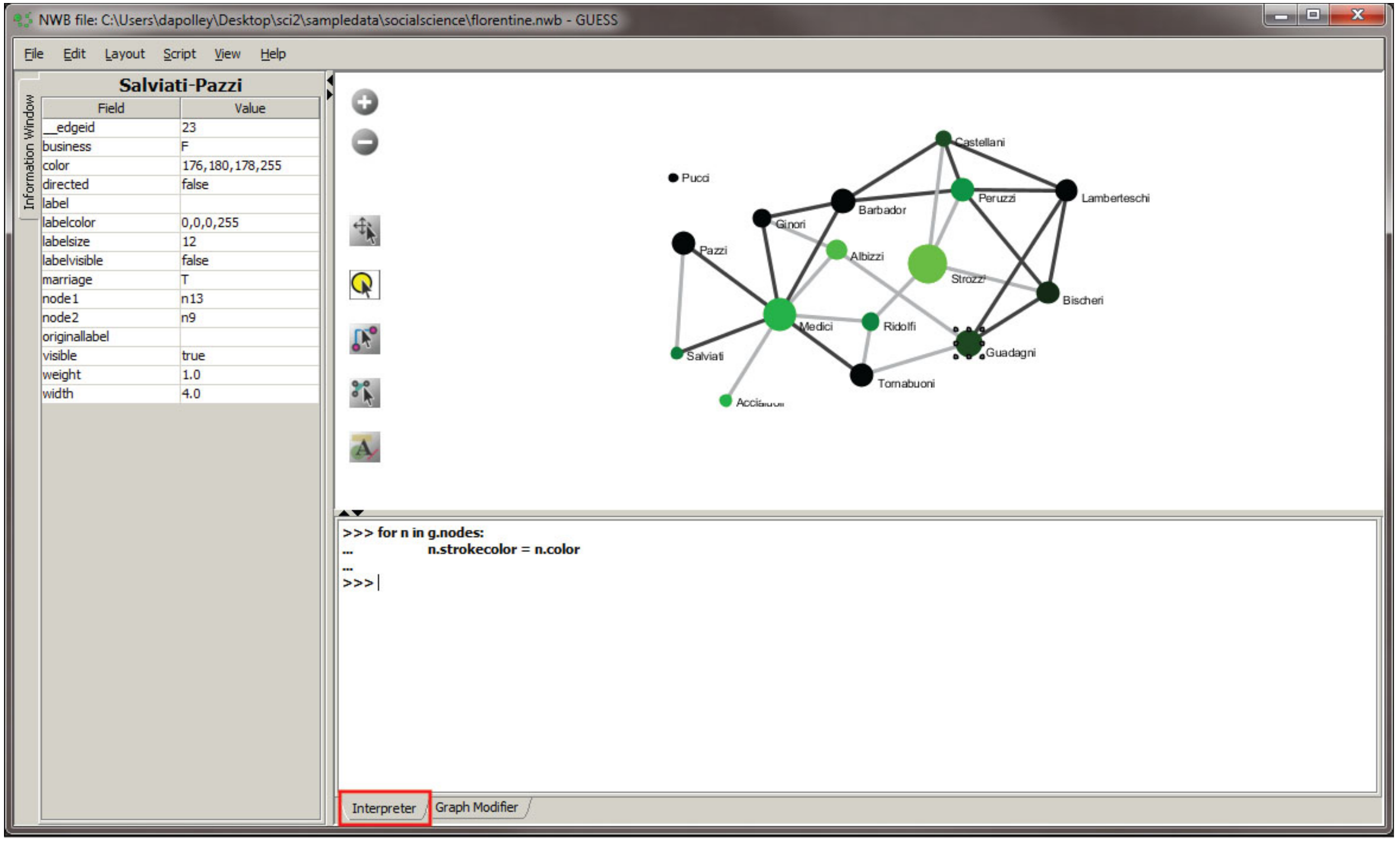
Since the GEM layout is random and all the nodes are spaced more or less evenly apart, you do not have to worry about disrupting the layout. However, other layout algorithms may space the nodes according to specific attributes of the network. Manually moving around nodes in this case would disrupt the layout of the network and distort the meaning of the visualization.

The last thing we want to do to our network is color the border of the nodes the same as the nodes themselves. This is not as crucial for networks with only a few nodes, but as the size of your network increases it can become difficult to read with the thick black lines around every node. To color those the same as the node go to the *Interpreter* tab at the bottom of the GUESS window and type in the following commands:

```
for n in g.nodes:
      n.strokecolor = n.color
```

Visualizing the Florentine Dataset

This code basically tells GUESS that for every node (*n*) in this graph of nodes (*g.nodes)* make the border color of the nodes (*n.strokecolor*) equal to the node color (*n.color)*. After you type the first line you will need to hit the "Tab" key before you start typing the second line of code.



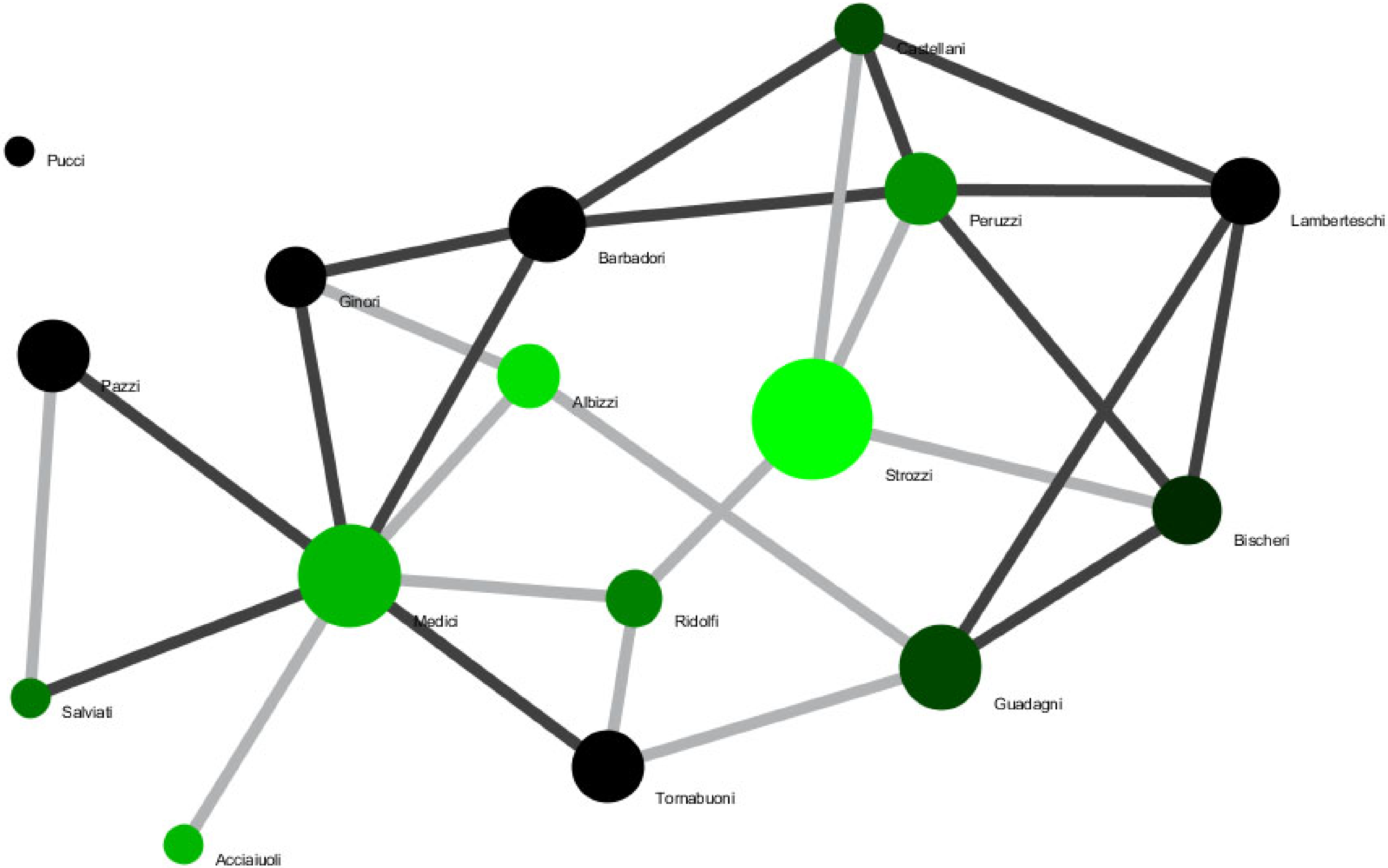
NWB file: C:\Users\dapolley\Desktop\sci2\sampledata\socialscience\florentine.nwb - GUESS

Salviati-Pazzi

Field	Value
__edgeid	23
business	F
color	176,180,178,255
directed	false
label	
labelcolor	0,0,0,255
labelsize	12
labelvisible	false
marriage	T
node1	n13
node2	n9
originallabel	
visible	true
weight	1.0
width	4.0

```
>>> for n in g.nodes:
...             n.strokecolor = n.color
...
>>>
```

Visualizing the Florentine Dataset



Questions?

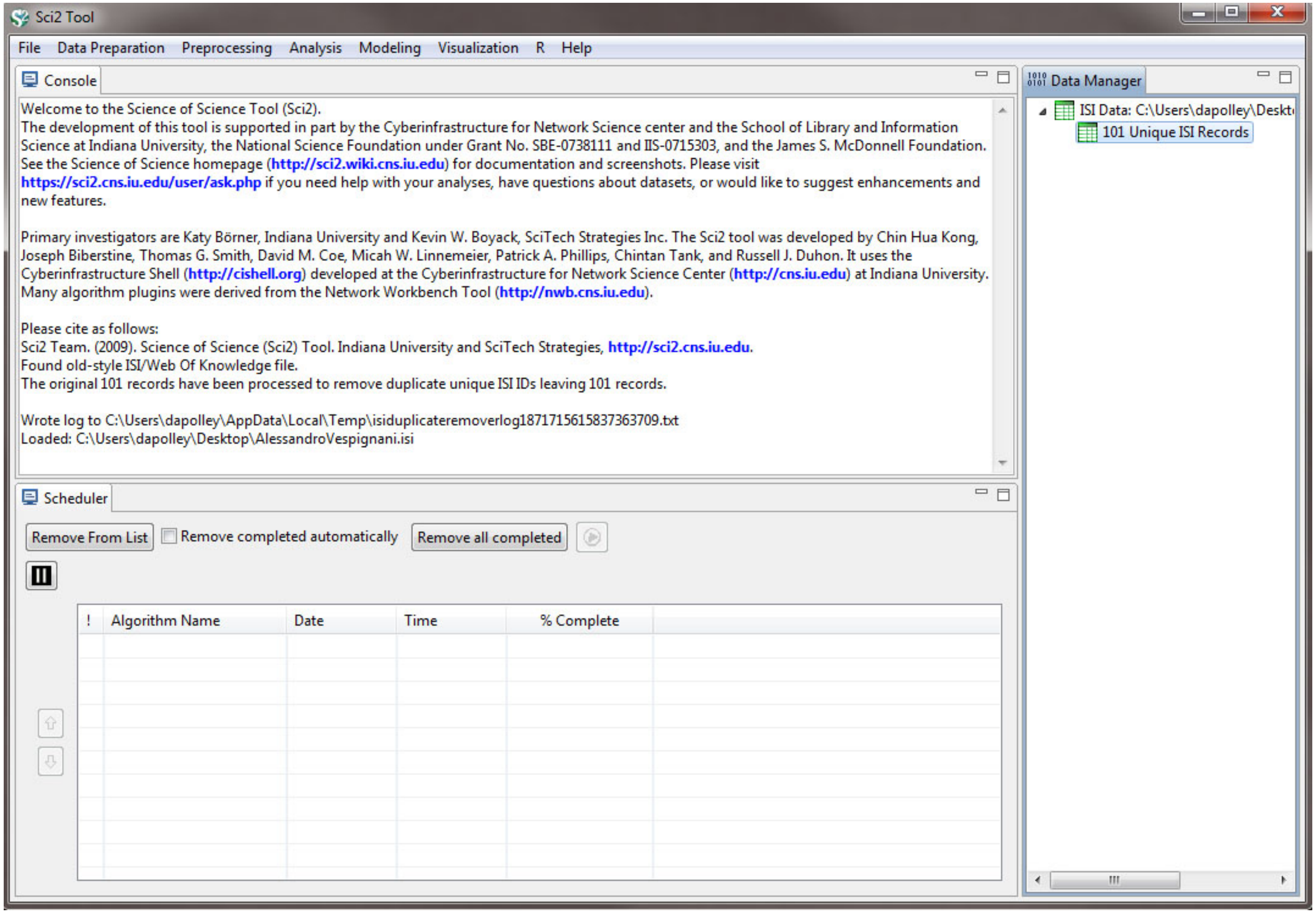
Temporal Analysis: *Evolving Co-Authorship Network*

For this analysis we will be studying the evolution of Alessandro Vespignani’s co-authorship network over time. We will see his network of collaborators grow from 1990 to 2006, giving us a sense of how his scholarly output has grown. Each node in the network will represent an author in the data set and the edges that connect them will be weighted based on how many times they have collaborated.

Evolving Co-Authorship Network

File > Load > AlessandroVespignani.isi

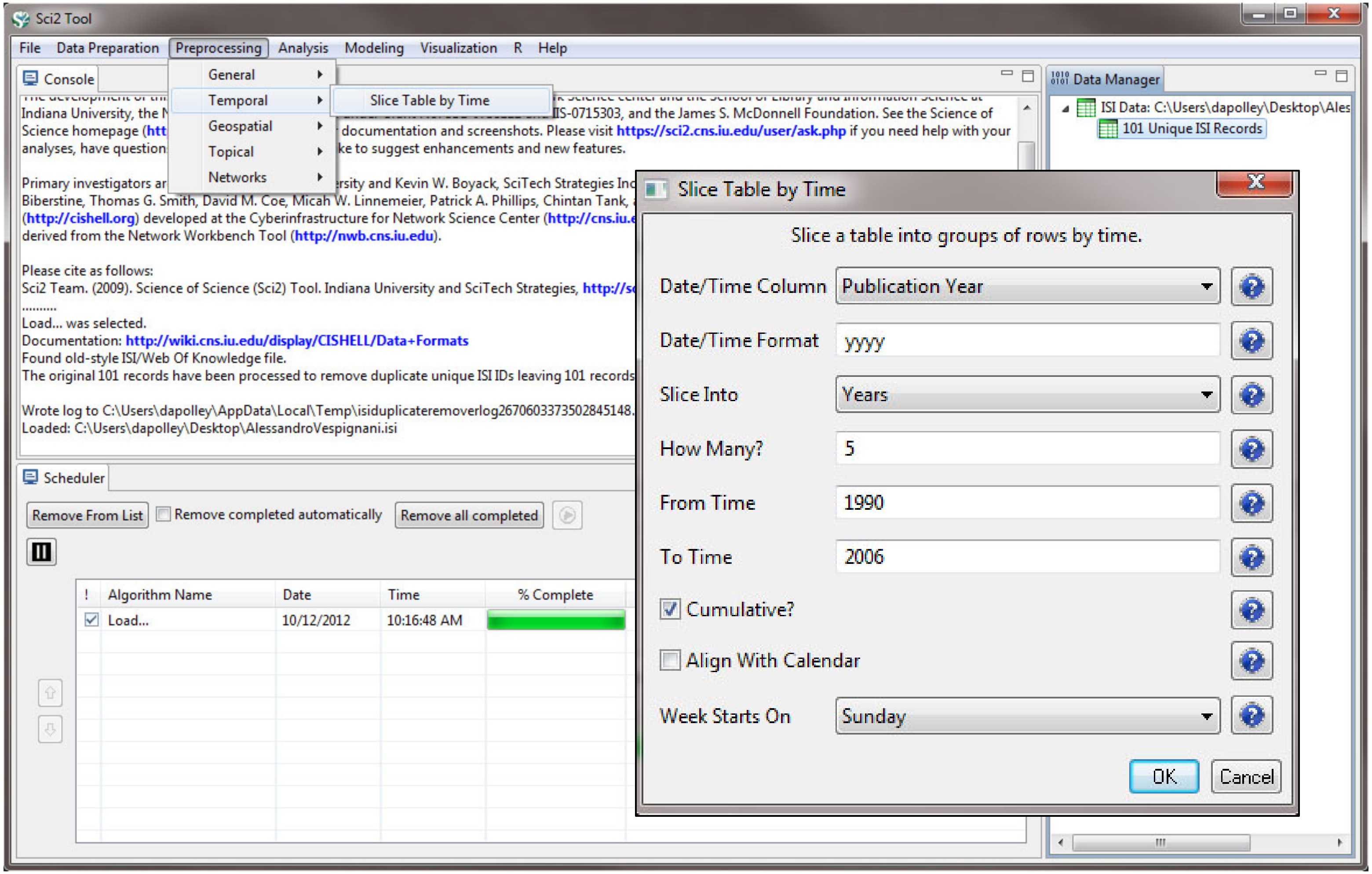
Load this file from the sample data folder you copied from the flash drive.



```
Welcome to the Science of Science Tool (Sci2).
The development of this tool is supported in part by the Cyberinfrastructure for Network Science center and the School of Library and Information
Science at Indiana University, the National Science Foundation under Grant No. SBE-0738111 and IIS-0715303, and the James S. McDonnell Foundation.
See the Science of Science homepage (http://sci2.wiki.cns.iu.edu) for documentation and screenshots. Please visit
https://sci2.cns.iu.edu/user/ask.php if you need help with your analyses, have questions about datasets, or would like to suggest enhancements and
new features.

Primary investigators are Katy Börner, Indiana University and Kevin W. Boyack, SciTech Strategies Inc. The Sci2 tool was developed by Chin Hua Kong,
Joseph Biberstine, Thomas G. Smith, David M. Coe, Micah W. Linnemeier, Patrick A. Phillips, Chintan Tank, and Russell J. Duhon. It uses the
Cyberinfrastructure Shell (http://cishell.org) developed at the Cyberinfrastructure for Network Science Center (http://cns.iu.edu) at Indiana University.
Many algorithm plugins were derived from the Network Workbench Tool (http://nwb.cns.iu.edu).

Please cite as follows:
Sci2 Team. (2009). Science of Science (Sci2) Tool. Indiana University and SciTech Strategies, http://sci2.cns.iu.edu.
Found old-style ISI/Web Of Knowledge file.
The original 101 records have been processed to remove duplicate unique ISI IDs leaving 101 records.

Wrote log to C:\Users\dapolley\AppData\Local\Temp\isiduplicateremoverlog1871715615837363709.txt
Loaded: C:\Users\dapolley\Desktop\AlessandroVespignani.isi
```

Evolving Co-Authorship Network

Select *Preprocessing > Temporal > Slice Table by Time* and choose the parameters shown at the right.



Indiana University, the N … IS-0715303, and the James S. McDonnell Foundation. See the Science of
Science homepage (htt … documentation and screenshots. Please visit https://sci2.cns.iu.edu/user/ask.php if you need help with your
analyses, have question … ke to suggest enhancements and new features.

Primary investigators ar … rsity and Kevin W. Boyack, SciTech Strategies Inc
Biberstine, Thomas G. Smith, David M. Coe, Micah W. Linnemeier, Patrick A. Phillips, Chintan Tank,
(http://cishell.org) developed at the Cyberinfrastructure for Network Science Center (http://cns.iu.e
derived from the Network Workbench Tool (http://nwb.cns.iu.edu).

Please cite as follows:
Sci2 Team. (2009). Science of Science (Sci2) Tool. Indiana University and SciTech Strategies, http://s
....
Load... was selected.
Documentation: http://wiki.cns.iu.edu/display/CISHELL/Data+Formats
Found old-style ISI/Web Of Knowledge file.
The original 101 records have been processed to remove duplicate unique ISI IDs leaving 101 records

Wrote log to C:\Users\dapolley\AppData\Local\Temp\isiduplicateremoverlog2670603373502845148.
Loaded: C:\Users\dapolley\Desktop\AlessandroVespignani.isi

!	Algorithm Name	Date	Time	% Complete
☑	Load...	10/12/2012	10:16:48 AM	🟩

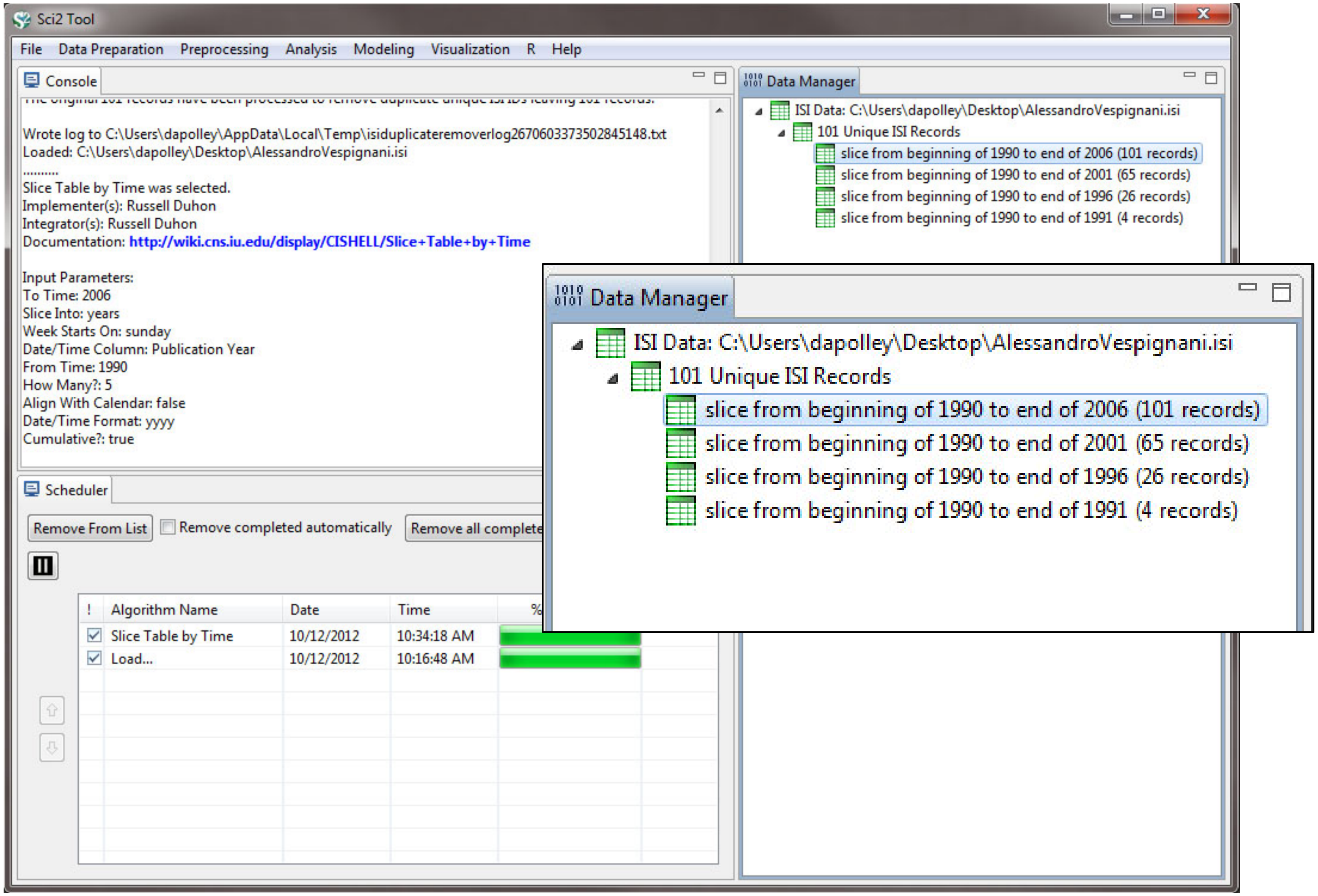
Slice Table by Time

Slice a table into groups of rows by time.

Date/Time Column	Publication Year
Date/Time Format	yyyy
Slice Into	Years
How Many?	5
From Time	1990
To Time	2006
☑ Cumulative?	
☐ Align With Calendar	
Week Starts On	Sunday

Evolving Co-Authorship Network

Now that the algorithm has been run, you will notice the original dataset has been divided into four tables that cumulatively display the evolution of this data.



```
The original 101 records have been processed to remove duplicate unique ISI IDs leaving 101 records.

Wrote log to C:\Users\dapolley\AppData\Local\Temp\isiduplicateremoverlog2670603373502845148.txt
Loaded: C:\Users\dapolley\Desktop\AlessandroVespignani.isi
.........
Slice Table by Time was selected.
Implementer(s): Russell Duhon
Integrator(s): Russell Duhon
Documentation: http://wiki.cns.iu.edu/display/CISHELL/Slice+Table+by+Time

Input Parameters:
To Time: 2006
Slice Into: years
Week Starts On: sunday
Date/Time Column: Publication Year
From Time: 1990
How Many?: 5
Align With Calendar: false
Date/Time Format: yyyy
Cumulative?: true
```

- ISI Data: C:\Users\dapolley\Desktop\AlessandroVespignani.isi
  - 101 Unique ISI Records
    - slice from beginning of 1990 to end of 2006 (101 records)
    - slice from beginning of 1990 to end of 2001 (65 records)
    - slice from beginning of 1990 to end of 1996 (26 records)
    - slice from beginning of 1990 to end of 1991 (4 records)

!	Algorithm Name	Date	Time	%
☑	Slice Table by Time	10/12/2012	10:34:18 AM	🟩
☑	Load...	10/12/2012	10:16:48 AM	🟩

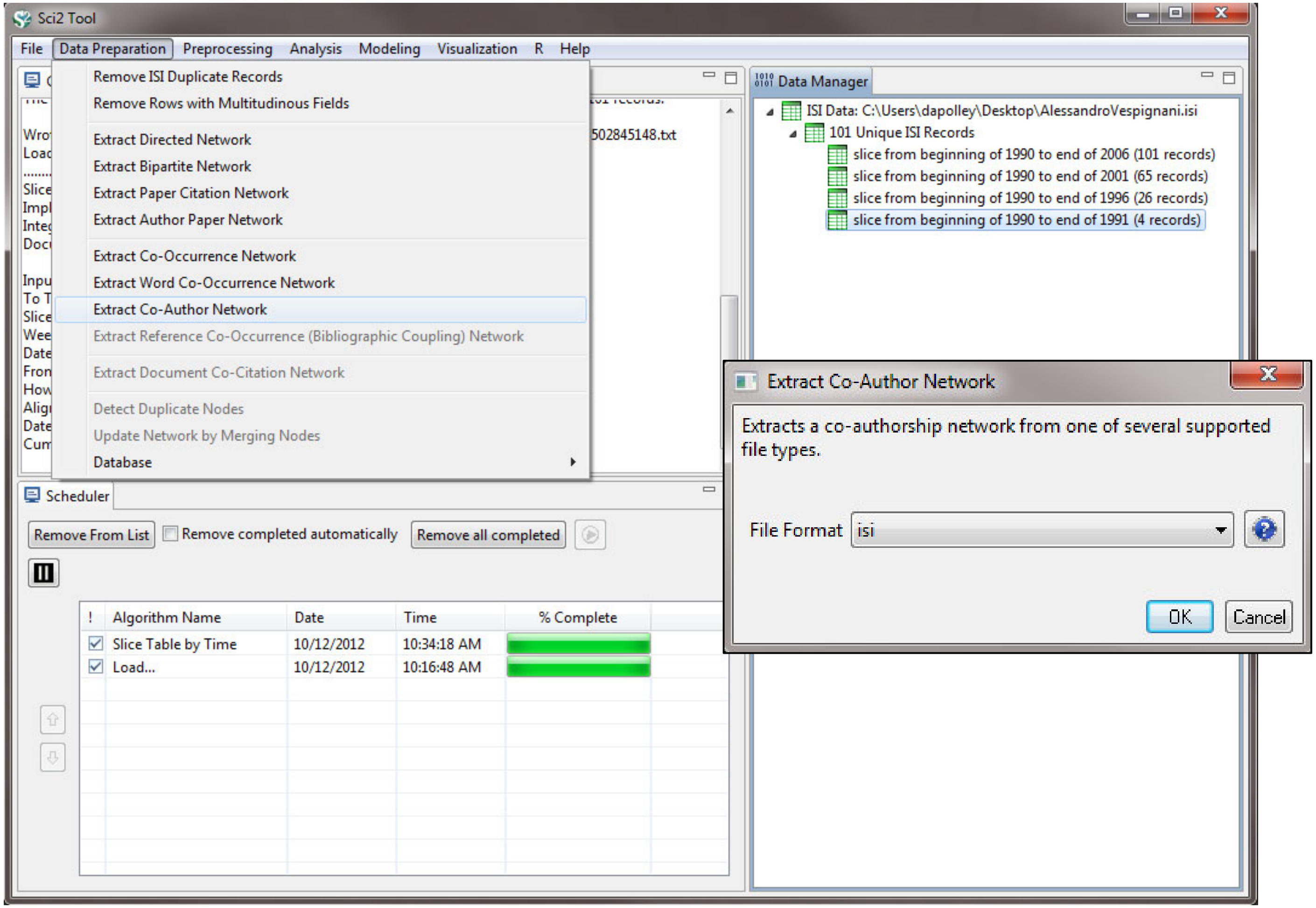
Data Manager

- ISI Data: C:\Users\dapolley\Desktop\AlessandroVespignani.isi
  - 101 Unique ISI Records
    - slice from beginning of 1990 to end of 2006 (101 records)
    - slice from beginning of 1990 to end of 2001 (65 records)
    - slice from beginning of 1990 to end of 1996 (26 records)
    - slice from beginning of 1990 to end of 1991 (4 records)

Evolving Co-Authorship Network

Select the first table and run *Data Preparation > Extract Co-Author Network*

Repeat this step for each of the tables in the Data Manager


Sci2 Tool
File
Data Preparation
Preprocessing
Analysis
Modeling
Visualization
R
Help
Remove ISI Duplicate Records
Remove Rows with Multitudinous Fields
Extract Directed Network
Extract Bipartite Network
Extract Paper Citation Network
Extract Author Paper Network
Extract Co-Occurrence Network
Extract Word Co-Occurrence Network
Extract Co-Author Network
Extract Reference Co-Occurrence (Bibliographic Coupling) Network
Extract Document Co-Citation Network
Detect Duplicate Nodes
Update Network by Merging Nodes
Database
502845148.txt
Data Manager
Scheduler
Remove From List
Remove completed automatically
Remove all completed
Extract Co-Author Network
OK
Cancel

- ISI Data: C:\Users\dapolley\Desktop\AlessandroVespignani.isi
  - 101 Unique ISI Records
    - slice from beginning of 1990 to end of 2006 (101 records)
    - slice from beginning of 1990 to end of 2001 (65 records)
    - slice from beginning of 1990 to end of 1996 (26 records)
    - slice from beginning of 1990 to end of 1991 (4 records)

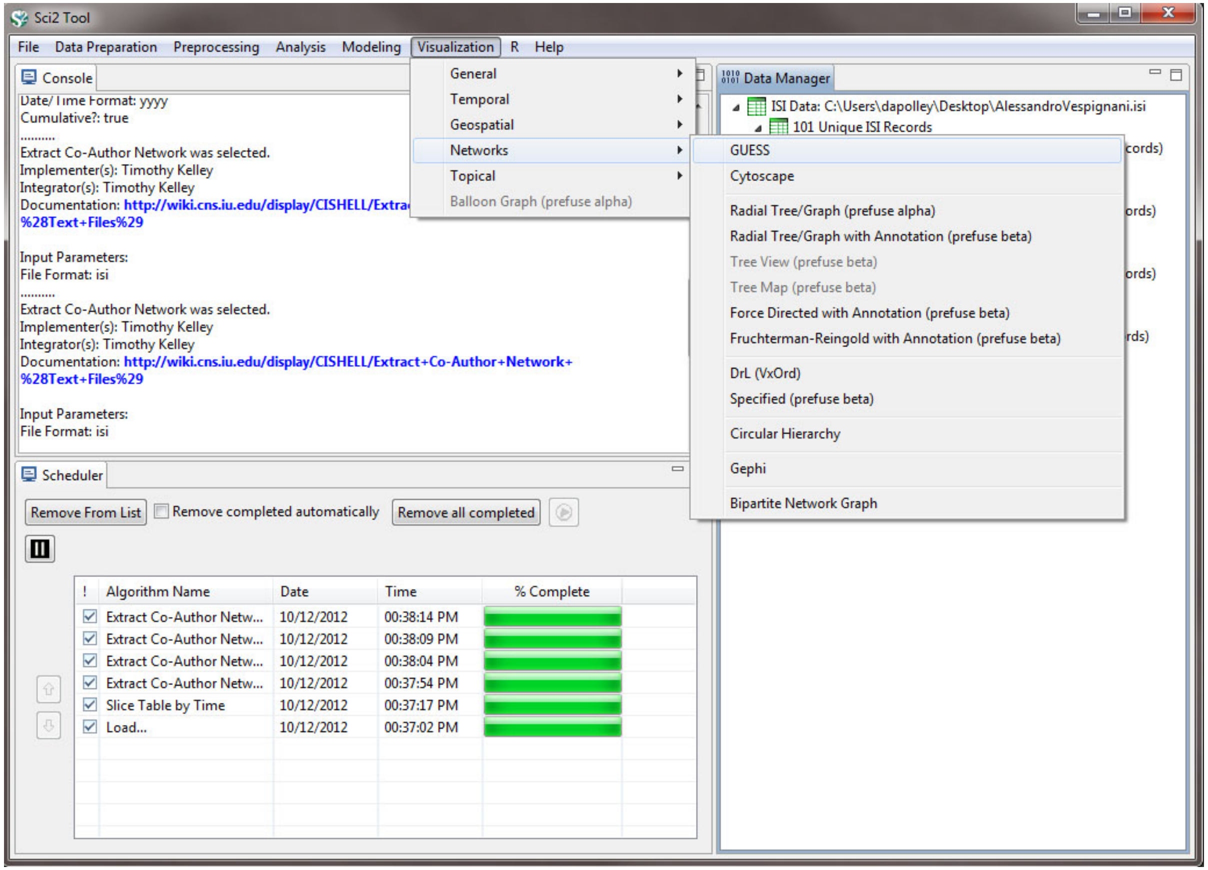
!	Algorithm Name	Date	Time	% Complete
☑	Slice Table by Time	10/12/2012	10:34:18 AM	🟩
☑	Load...	10/12/2012	10:16:48 AM	🟩

Extracts a co-authorship network from one of several supported file types.

File Format: isi

Evolving Co-Authorship Network

Select the first extracted co-author network and run *Visualization > Networks > GUESS* starting with the network that spans 1990 to 2006 because we will export these node positions and use them to layout the other networks.

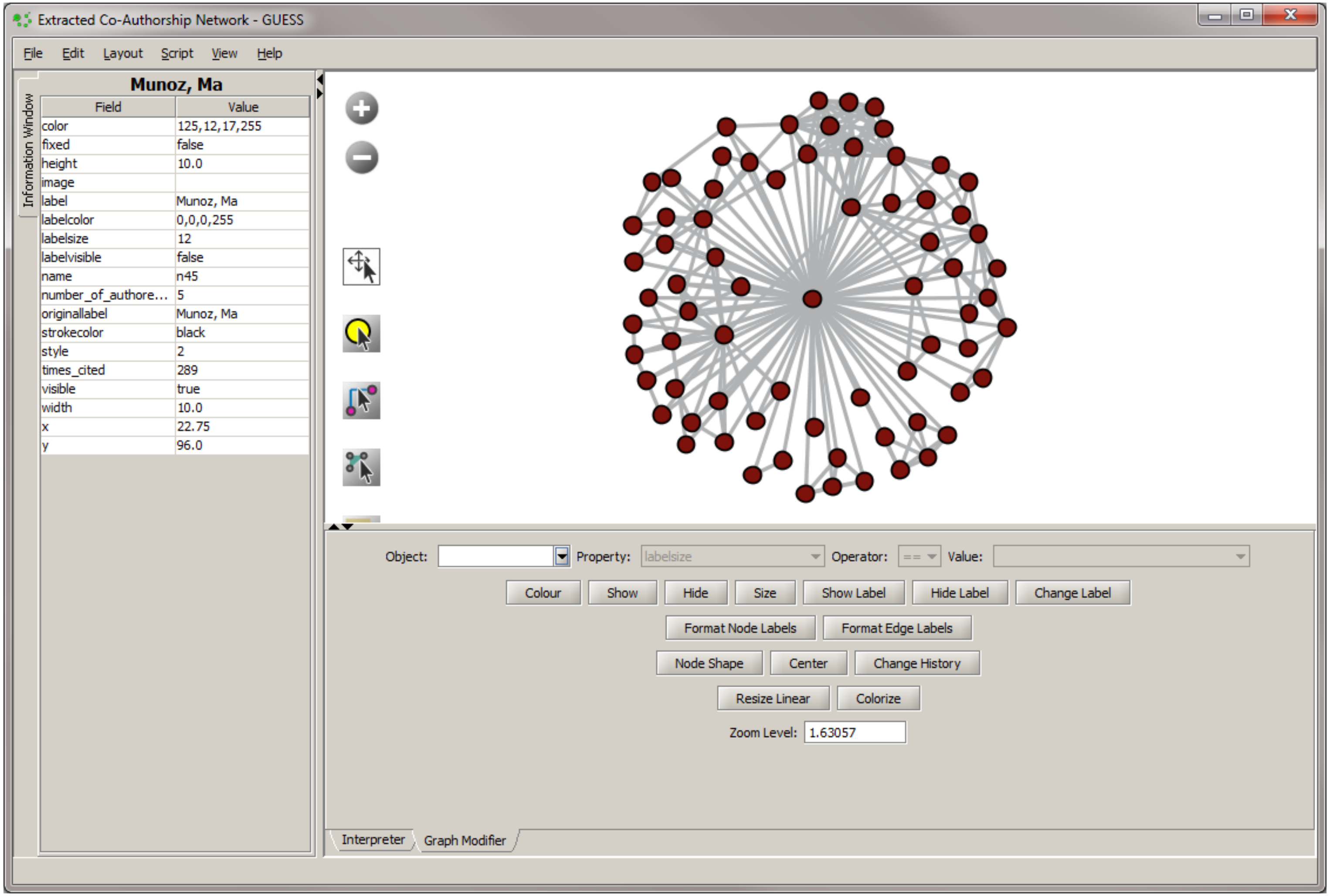


```
Date/Time Format: yyyy
Cumulative?: true
..........
Extract Co-Author Network was selected.
Implementer(s): Timothy Kelley
Integrator(s): Timothy Kelley
Documentation: http://wiki.cns.iu.edu/display/CISHELL/Extra
%28Text+Files%29

Input Parameters:
File Format: isi
..........
Extract Co-Author Network was selected.
Implementer(s): Timothy Kelley
Integrator(s): Timothy Kelley
Documentation: http://wiki.cns.iu.edu/display/CISHELL/Extract+Co-Author+Network+
%28Text+Files%29

Input Parameters:
File Format: isi
```

!	Algorithm Name	Date	Time	% Complete
☑	Extract Co-Author Netw...	10/12/2012	00:38:14 PM	🟩
☑	Extract Co-Author Netw...	10/12/2012	00:38:09 PM	🟩
☑	Extract Co-Author Netw...	10/12/2012	00:38:04 PM	🟩
☑	Extract Co-Author Netw...	10/12/2012	00:37:54 PM	🟩
☑	Slice Table by Time	10/12/2012	00:37:17 PM	🟩
☑	Load...	10/12/2012	00:37:02 PM	🟩

Evolving Co-Authorship Network

The network will be loaded in with a random layout in GUESS
To change the layout select *Layout > Gem*



Extracted Co-Authorship Network - GUESS

Munoz, Ma

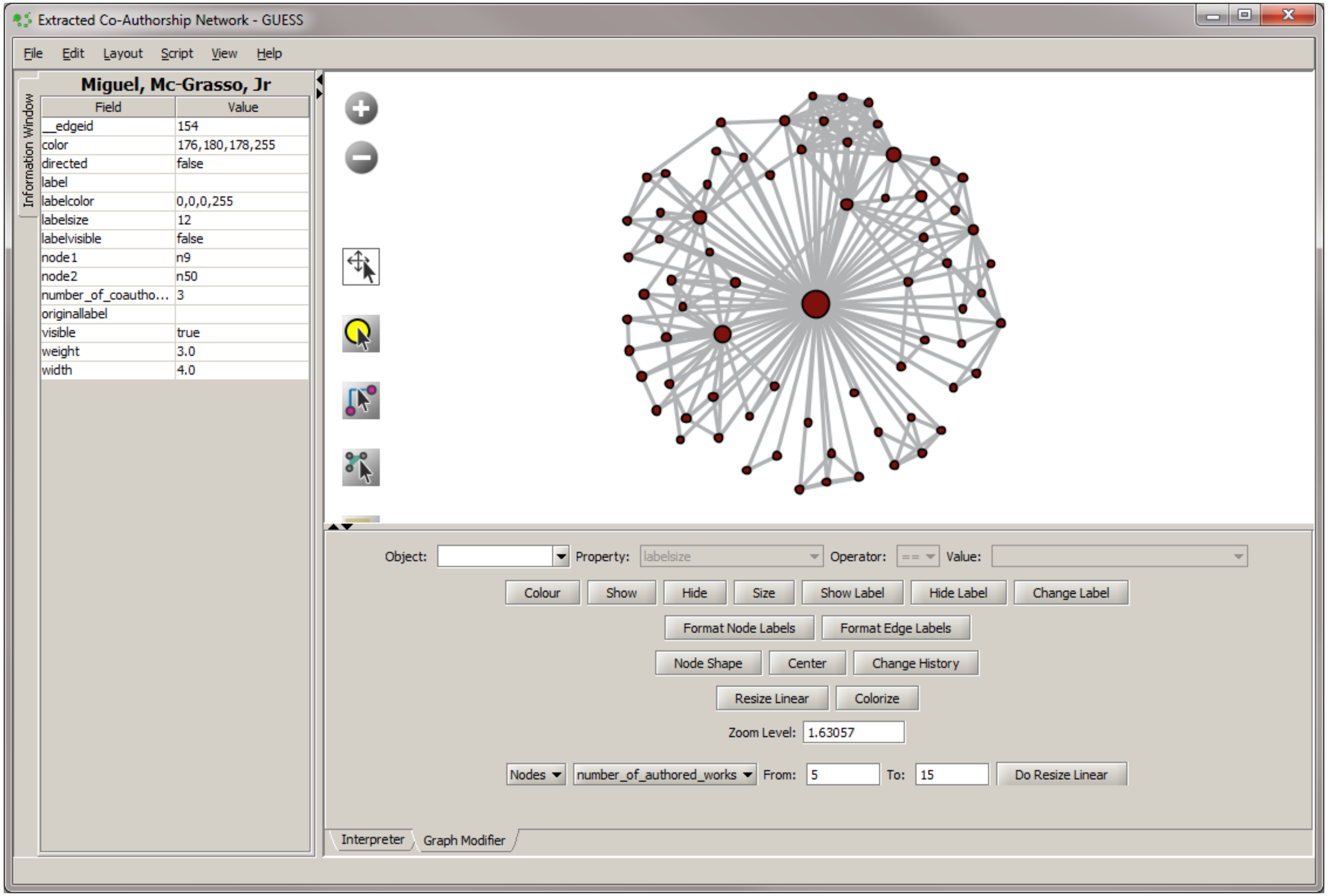
Field	Value
color	125,12,17,255
fixed	false
height	10.0
image	
label	Munoz, Ma
labelcolor	0,0,0,255
labelsize	12
labelvisible	false
name	n45
number_of_authore...	5
originallabel	Munoz, Ma
strokecolor	black
style	2
times_cited	289
visible	true
width	10.0
x	22.75
y	96.0

Object: Property: labelsize Operator: == Value:

Zoom Level: 1.63057

Evolving Co-Authorship Network

Resize the nodes based on *number_of_authored_works*
Set the parameters from 5 to 15 and click *Do Resize Linear*



Extracted Co-Authorship Network - GUESS

Miguel, Mc-Grasso, Jr

Field	Value
__edgeid	154
color	176,180,178,255
directed	false
label	
labelcolor	0,0,0,255
labelsize	12
labelvisible	false
node1	n9
node2	n50
number_of_coautho...	3
originallabel	
visible	true
weight	3.0
width	4.0

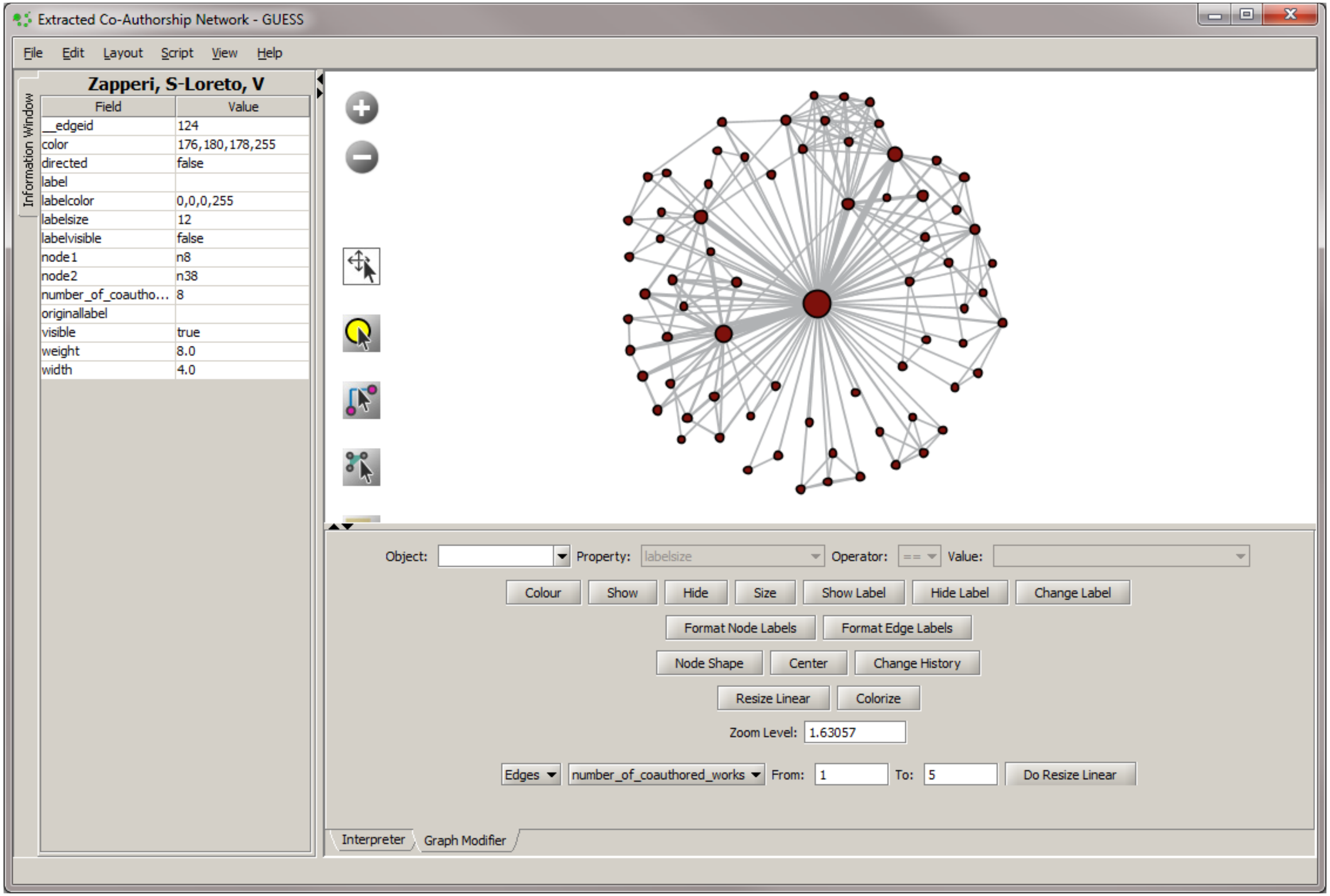
Object: Property: labelsize Operator: == Value:

Zoom Level: 1.63057

Nodes number_of_authored_works From: 5 To: 15 Do Resize Linear

Evolving Co-Authorship Network

Resize the edges based on *number_of_coauthored_works*
Set the parameters from 1 to 5 and click *Do Resize Linear*



Extracted Co-Authorship Network - GUESS

Zapperi, S-Loreto, V

Field	Value
__edgeid	124
color	176,180,178,255
directed	false
label	
labelcolor	0,0,0,255
labelsize	12
labelvisible	false
node1	n8
node2	n38
number_of_coautho...	8
originallabel	
visible	true
weight	8.0
width	4.0

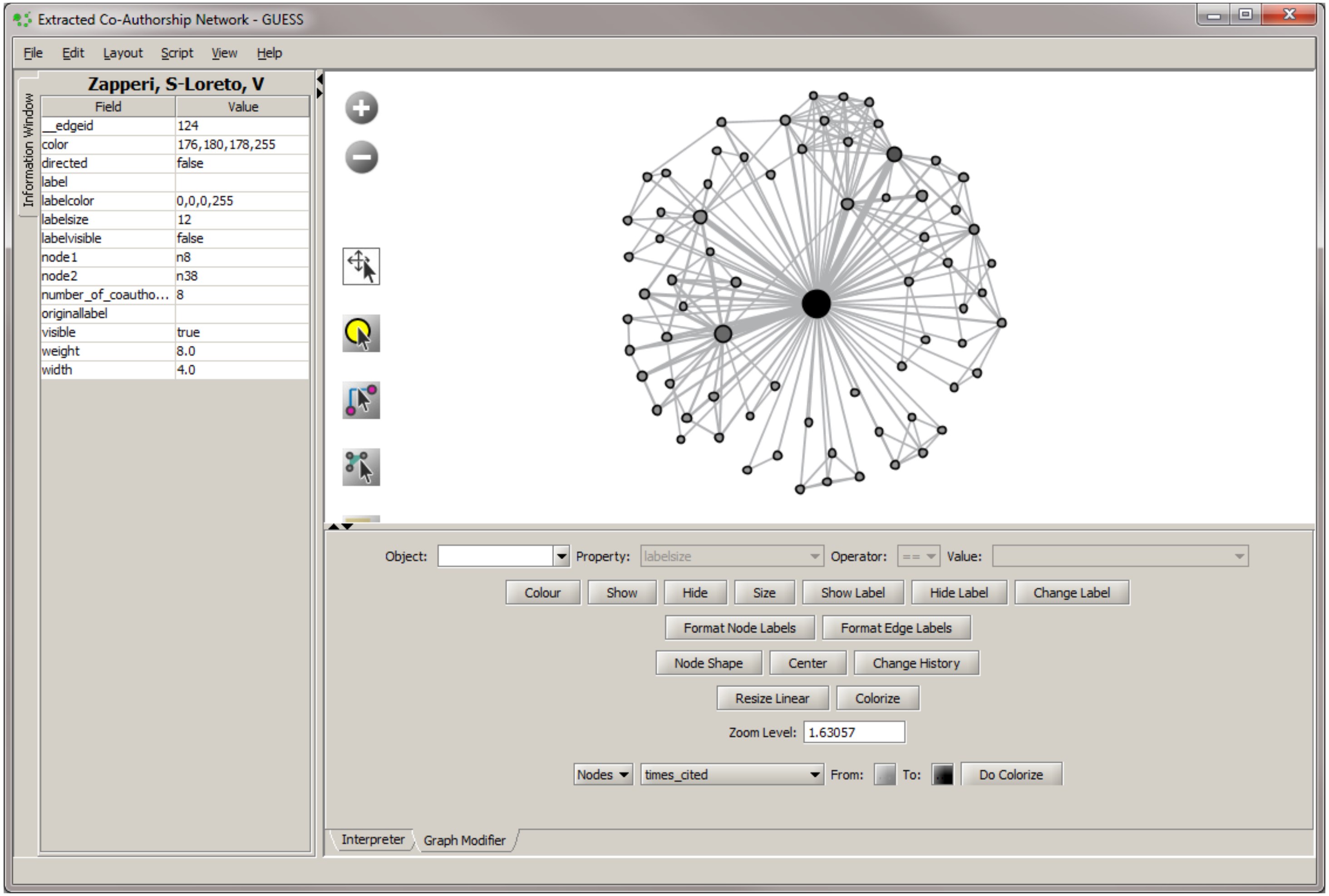
Object: Property: labelsize Operator: == Value:

Zoom Level: 1.63057

Edges number_of_coauthored_works From: 1 To: 5

Evolving Co-Authorship Network

Colorize the nodes based on *times_cited*
Set the parameters from *Gray* to *Black* and click *Do Colorize*



Extracted Co-Authorship Network - GUESS

Zapperi, S-Loreto, V

Field	Value
__edgeid	124
color	176,180,178,255
directed	false
label	
labelcolor	0,0,0,255
labelsize	12
labelvisible	false
node1	n8
node2	n38
number_of_coautho…	8
originallabel	
visible	true
weight	8.0
width	4.0

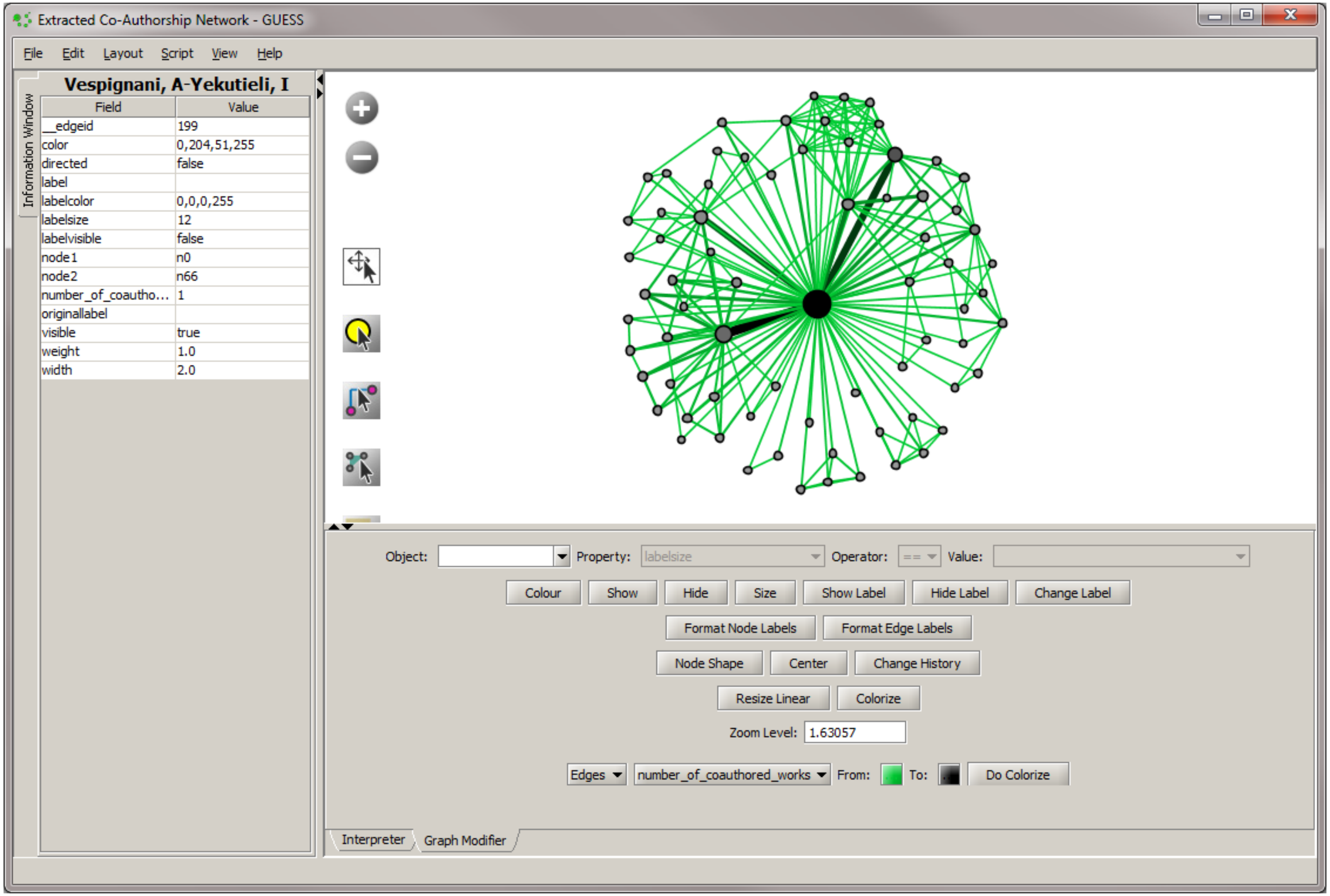
Object: Property: labelsize Operator: == Value:

Zoom Level: 1.63057

Nodes times_cited From: To: Do Colorize

Evolving Co-Authorship Network

Colorize the edges based on *number_of_coauthored_works*
Set the parameters from *Green* to *Black* and click *Do Colorize*



Vespignani, A-Yekutieli, I

Field	Value
__edgeid	199
color	0,204,51,255
directed	false
label	
labelcolor	0,0,0,255
labelsize	12
labelvisible	false
node1	n0
node2	n66
number_of_coautho...	1
originallabel	
visible	true
weight	1.0
width	2.0

Object: Property: labelsize Operator: == Value:

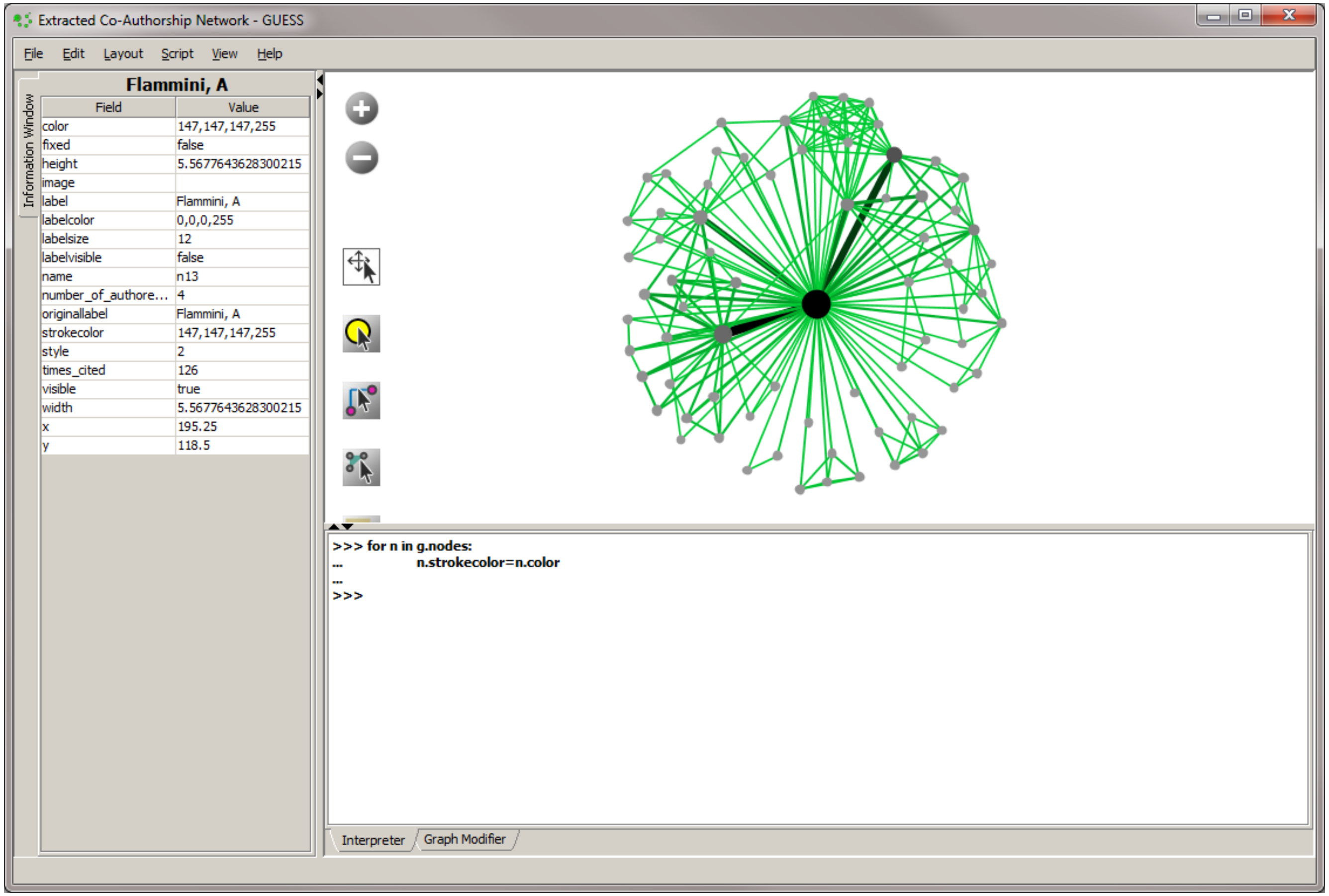
Zoom Level: 1.63057

Edges number_of_coauthored_works From: 🟩 To: ⬛ Do Colorize

Evolving Co-Authorship Network

If you want to remove the borders from the nodes, type the following commands in into the interpreter:

```
for n in g.nodes:
        n.strokecolor=n.color
```



Extracted Co-Authorship Network - GUESS

Flammini, A

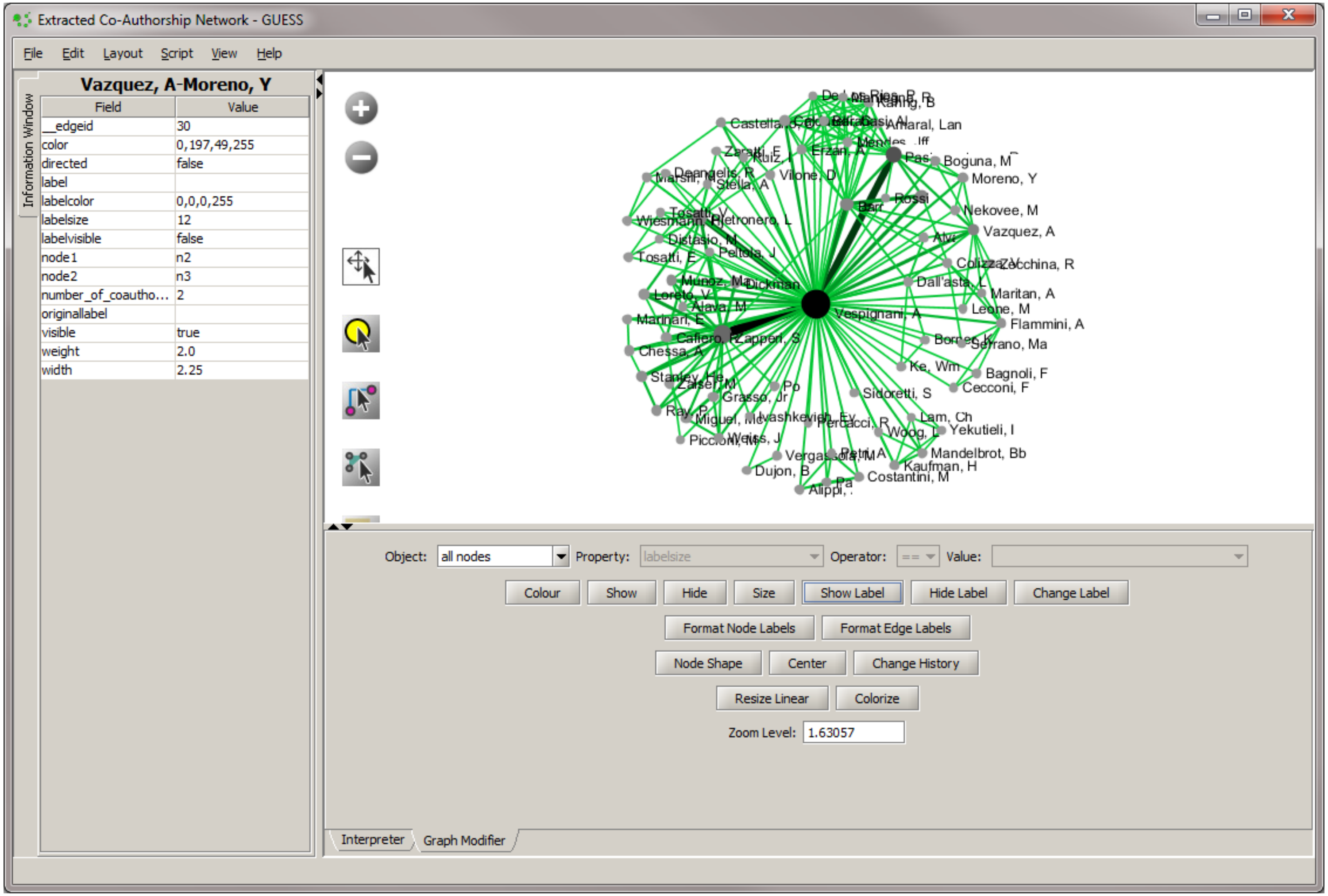
Field	Value
color	147,147,147,255
fixed	false
height	5.5677643628300215
image	
label	Flammini, A
labelcolor	0,0,0,255
labelsize	12
labelvisible	false
name	n13
number_of_authore...	4
originallabel	Flammini, A
strokecolor	147,147,147,255
style	2
times_cited	126
visible	true
width	5.5677643628300215
x	195.25
y	118.5

```
>>> for n in g.nodes:
...             n.strokecolor=n.color
...
>>>
```

CNS Cyberinfrastructure for Network Science Center

Evolving Co-Authorship Network

Finally add the labels to the nodes by selecting object: *all nodes* and then click Show Label



Vazquez, A-Moreno, Y

Field	Value
__edgeid	30
color	0,197,49,255
directed	false
label	
labelcolor	0,0,0,255
labelsize	12
labelvisible	false
node1	n2
node2	n3
number_of_coautho...	2
originallabel	
visible	true
weight	2.0
width	2.25

Object: all nodes Property: labelsize Operator: == Value:

Colour | Show | Hide | Size | Show Label | Hide Label | Change Label

Format Node Labels | Format Edge Labels

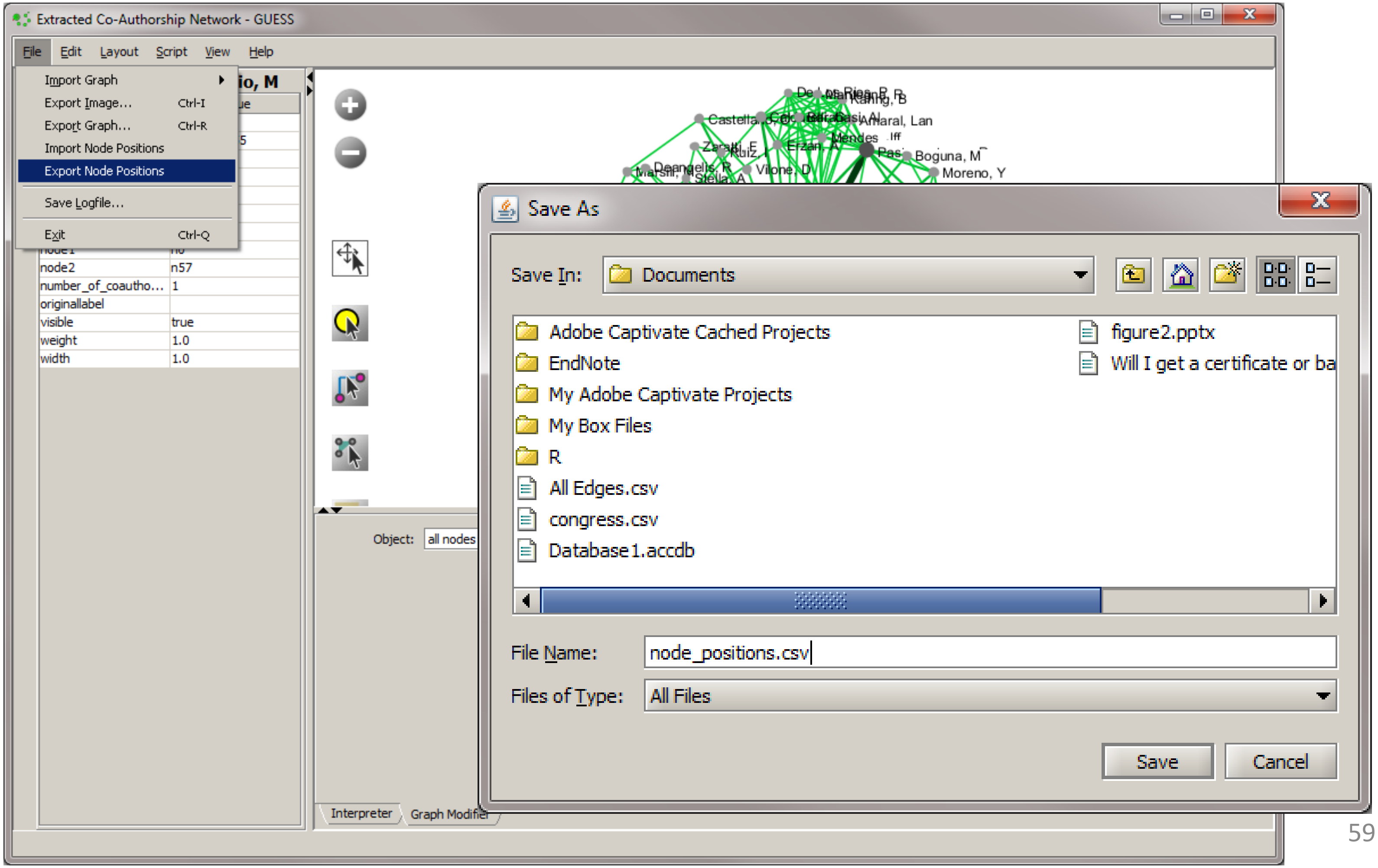
Node Shape | Center | Change History

Resize Linear | Colorize

Zoom Level: 1.63057

Evolving Co-Authorship Network

To save the node positions of the current layout so that the layout is consistent across all time slices select *File > Export Node Positions* and save the file as a CSV file.



node2	n57
number_of_coautho...	1
originallabel	
visible	true
weight	1.0
width	1.0

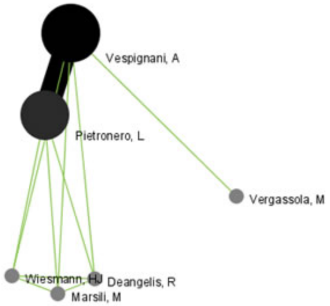
Save In: Documents

File Name: node_positions.csv

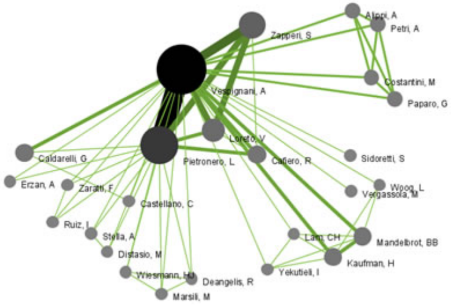
Files of Type: All Files

Evolving Co-Authorship Network

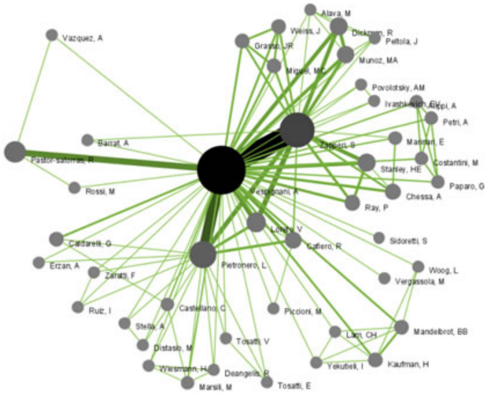
Now when you go to visualize the other three networks you will want to import the node the node positions using *File > Import Node Positions* and the network will be laid out accordingly. When the networks are displayed side-by-side you can see an evolution.



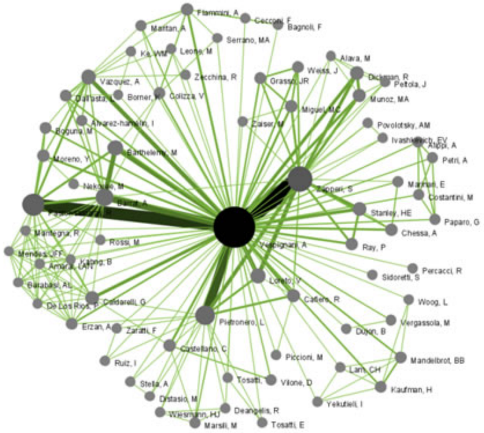
1990-1991



1990-1996



1990-2001



1990-2006

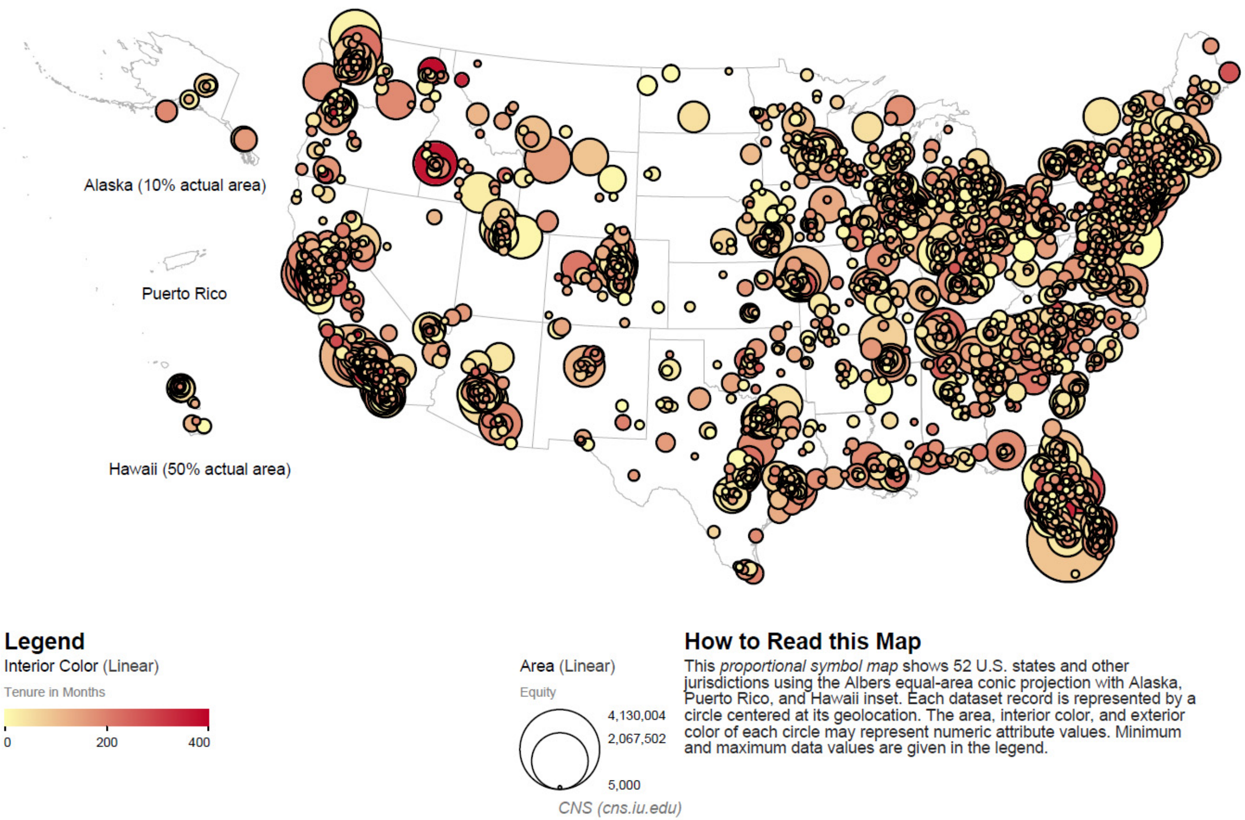
Questions?

Geospatial Analysis: *Proportional Symbol Map of Clients*

For this analysis we will be exploring the geographic distribution of clients and encoding various attributes of the data, such as equity and tenure as a TD Ameritrade client. The sample data has been provided to you prior to the workshop.

Geographic Distribution of Clients

How much is too much?



Geospatial Visualization (Proportional Symbol Map)

Geographic Distribution of Clients
Sep 20, 2013 | 11:01:54 AM EDT

Legend

Interior Color (Linear)

Tenure in Months

Area (Linear)

Equity

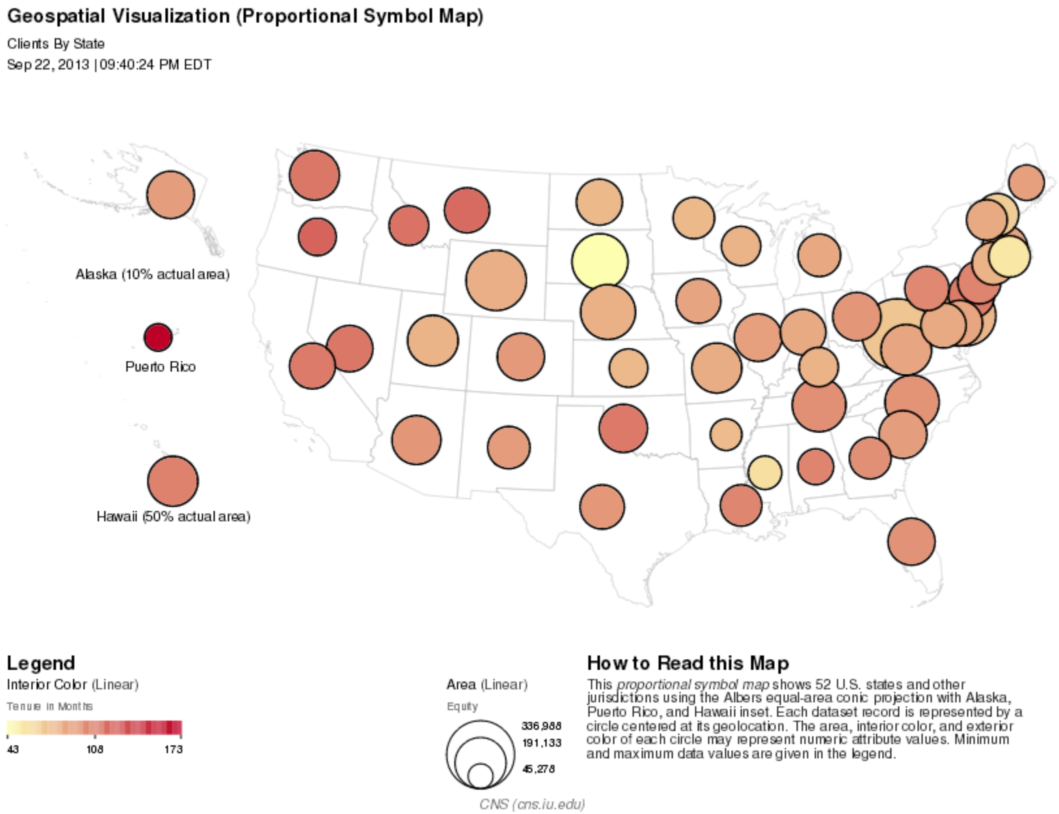
How to Read this Map

This *proportional symbol map* shows 52 U.S. states and other jurisdictions using the Albers equal-area conic projection with Alaska, Puerto Rico, and Hawaii inset. Each dataset record is represented by a circle centered at its geolocation. The area, interior color, and exterior color of each circle may represent numeric attribute values. Minimum and maximum data values are given in the legend.

CNS (cns.iu.edu)

Geographic Distribution of Clients

Aggregate clients by state



Geospatial Visualization (Proportional Symbol Map)

Clients By State
Sep 22, 2013 | 09:40:24 PM EDT

Legend

Interior Color (Linear)

Tenure in Months

Area (Linear)

Equity

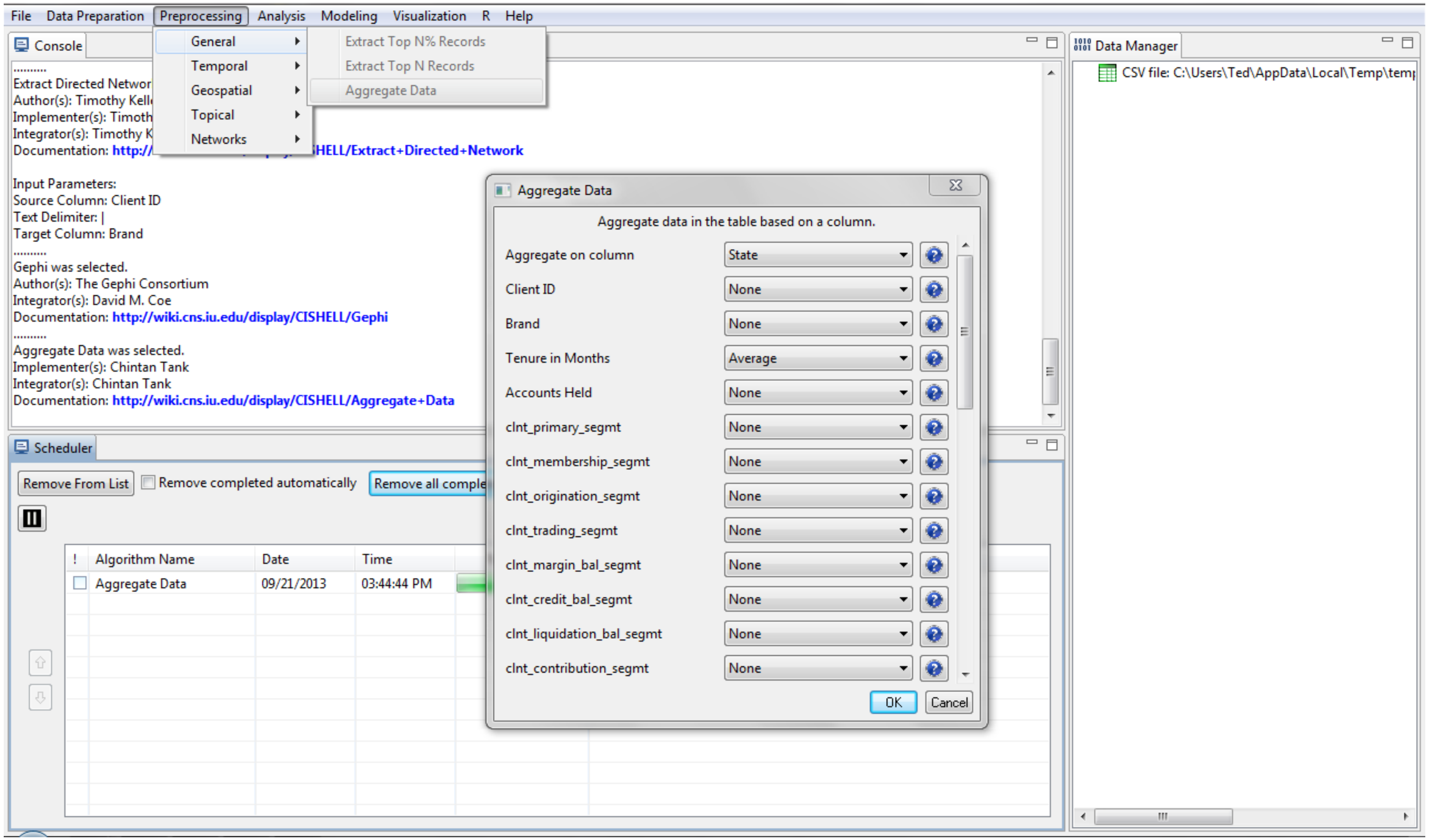
How to Read this Map

This *proportional symbol map* shows 52 U.S. states and other jurisdictions using the Albers equal-area conic projection with Alaska, Puerto Rico, and Hawaii inset. Each dataset record is represented by a circle centered at its geolocation. The area, interior color, and exterior color of each circle may represent numeric attribute values. Minimum and maximum data values are given in the legend.

CNS (cns.iu.edu)

Geographic Distribution of Clients

Preprocessing > General > Aggregate Data and set the Aggregate on Column to State, find the Average Tenure in Months and Equity for the clients in these states.



```
..........
Extract Directed Networ
Author(s): Timothy Kell
Implementer(s): Timoth
Integrator(s): Timothy K
Documentation: http://                      HELL/Extract+Directed+Network

Input Parameters:
Source Column: Client ID
Text Delimiter: |
Target Column: Brand
..........
Gephi was selected.
Author(s): The Gephi Consortium
Integrator(s): David M. Coe
Documentation: http://wiki.cns.iu.edu/display/CISHELL/Gephi
..........
Aggregate Data was selected.
Implementer(s): Chintan Tank
Integrator(s): Chintan Tank
Documentation: http://wiki.cns.iu.edu/display/CISHELL/Aggregate+Data
```

CSV file: C:\Users\Ted\AppData\Local\Temp\temp

!	Algorithm Name	Date	Time
□	Aggregate Data	09/21/2013	03:44:44 PM

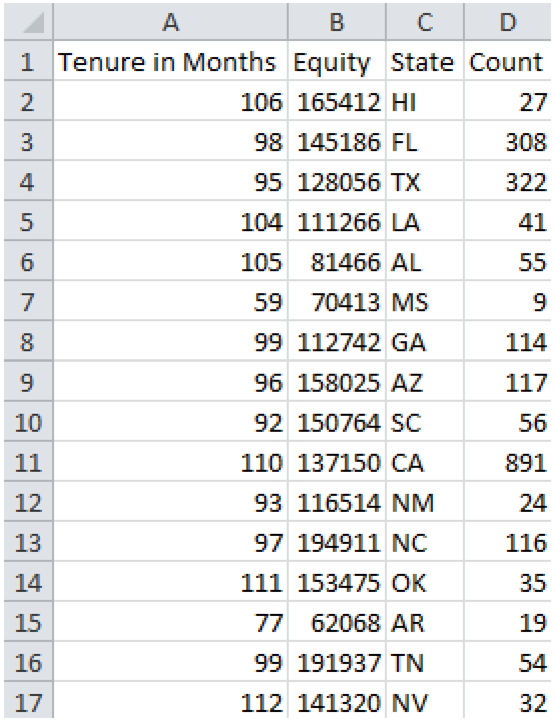
Aggregate Data

Aggregate data in the table based on a column.

Aggregate on column	State
Client ID	None
Brand	None
Tenure in Months	Average
Accounts Held	None
clnt_primary_segmt	None
clnt_membership_segmt	None
clnt_origination_segmt	None
clnt_trading_segmt	None
clnt_margin_bal_segmt	None
clnt_credit_bal_segmt	None
clnt_liquidation_bal_segmt	None
clnt_contribution_segmt	None

Geographic Distribution of Clients

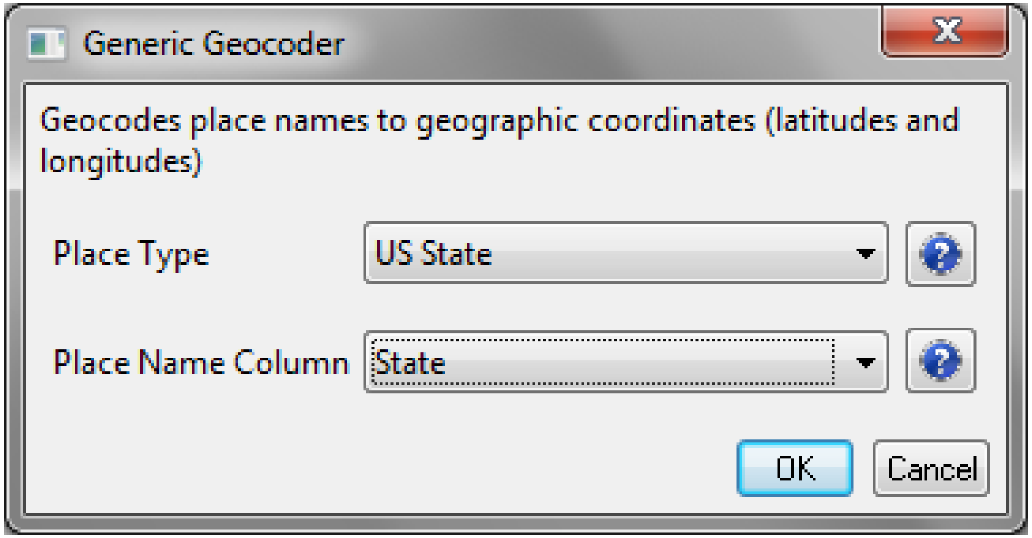
The result is a simplified dataset with the each state present. The *Count* column gives the total number of clients in each state for this dataset. The *Tenure in Months* column gives the average tenure in months for those clients, and the *Equity* column gives the average equity for clients in those states.



	A	B	C	D
1	Tenure in Months	Equity	State	Count
2	106	165412	HI	27
3	98	145186	FL	308
4	95	128056	TX	322
5	104	111266	LA	41
6	105	81466	AL	55
7	59	70413	MS	9
8	99	112742	GA	114
9	96	158025	AZ	117
10	92	150764	SC	56
11	110	137150	CA	891
12	93	116514	NM	24
13	97	194911	NC	116
14	111	153475	OK	35
15	77	62068	AR	19
16	99	191937	TN	54
17	112	141320	NV	32

Geographic Distribution of Clients

Next, select *Analysis* > *Geospatial* > *Generic Geocoder*



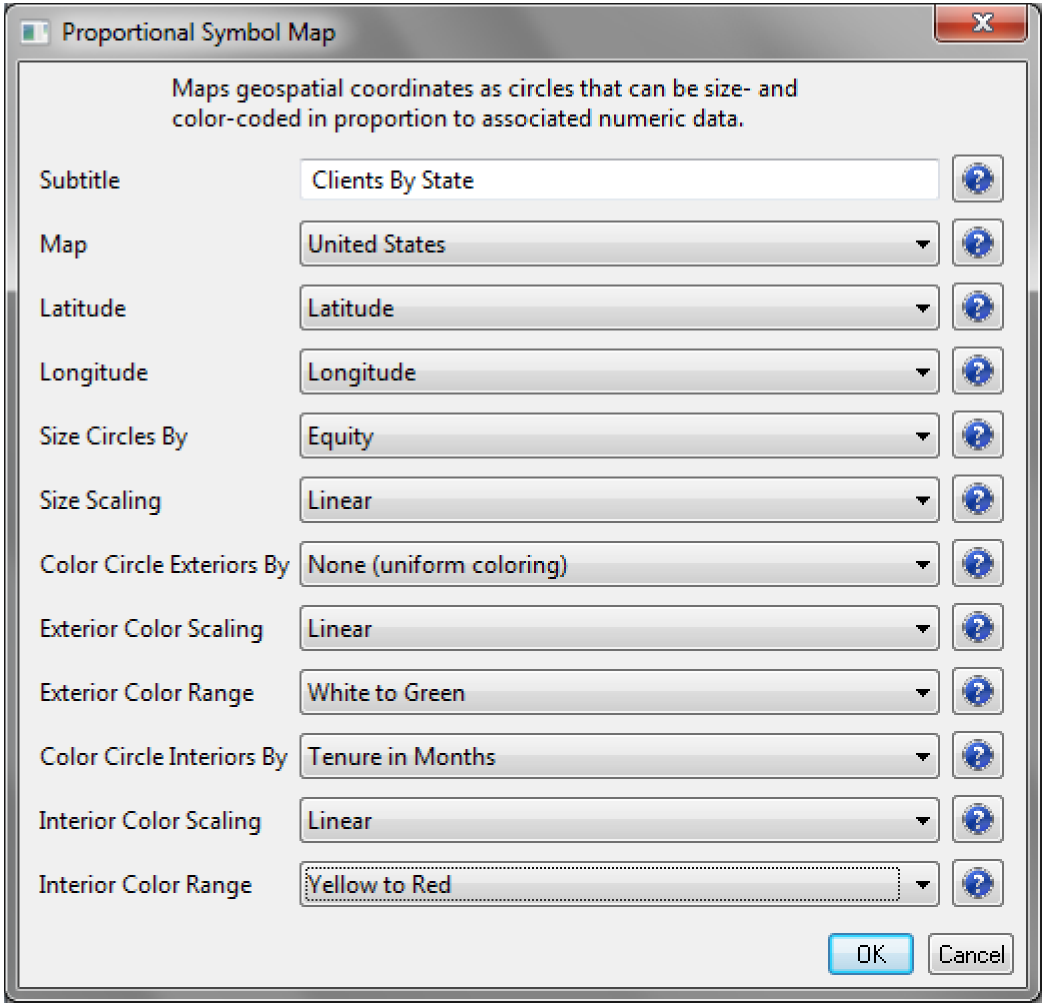
Geocodes place names to geographic coordinates (latitudes and longitudes)

Place Type: US State

Place Name Column: State

Geographic Distribution of Clients

Next, select *Visualization > Geospatial > Proportional Symbol Map*



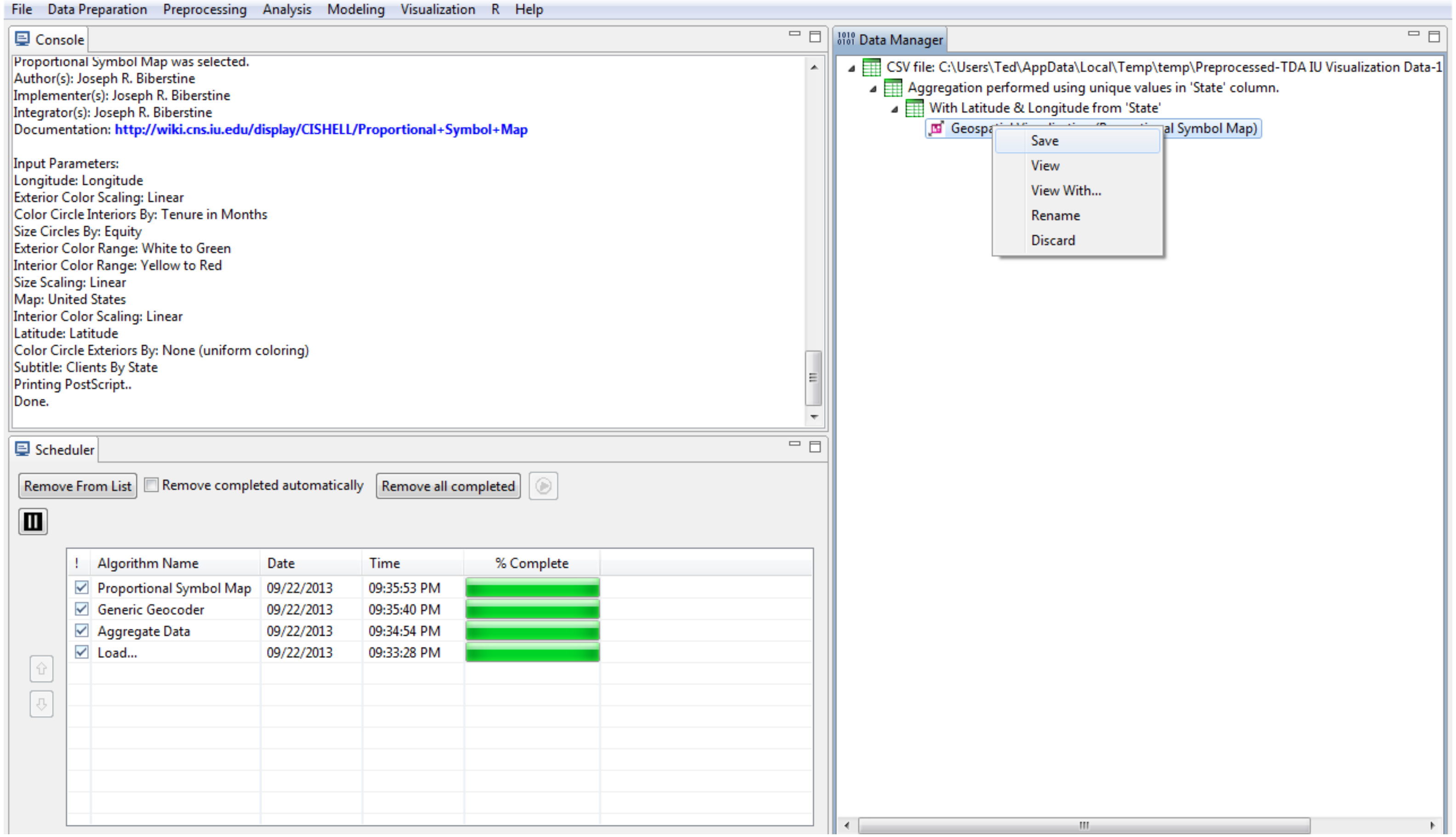
Proportional Symbol Map

Maps geospatial coordinates as circles that can be size- and color-coded in proportion to associated numeric data.

Subtitle	Clients By State
Map	United States
Latitude	Latitude
Longitude	Longitude
Size Circles By	Equity
Size Scaling	Linear
Color Circle Exteriors By	None (uniform coloring)
Exterior Color Scaling	Linear
Exterior Color Range	White to Green
Color Circle Interiors By	Tenure in Months
Interior Color Scaling	Linear
Interior Color Range	Yellow to Red

Geographic Distribution of Clients

The result is a PostScript file in the Data Manager. Right-click on the file and select *Save*.



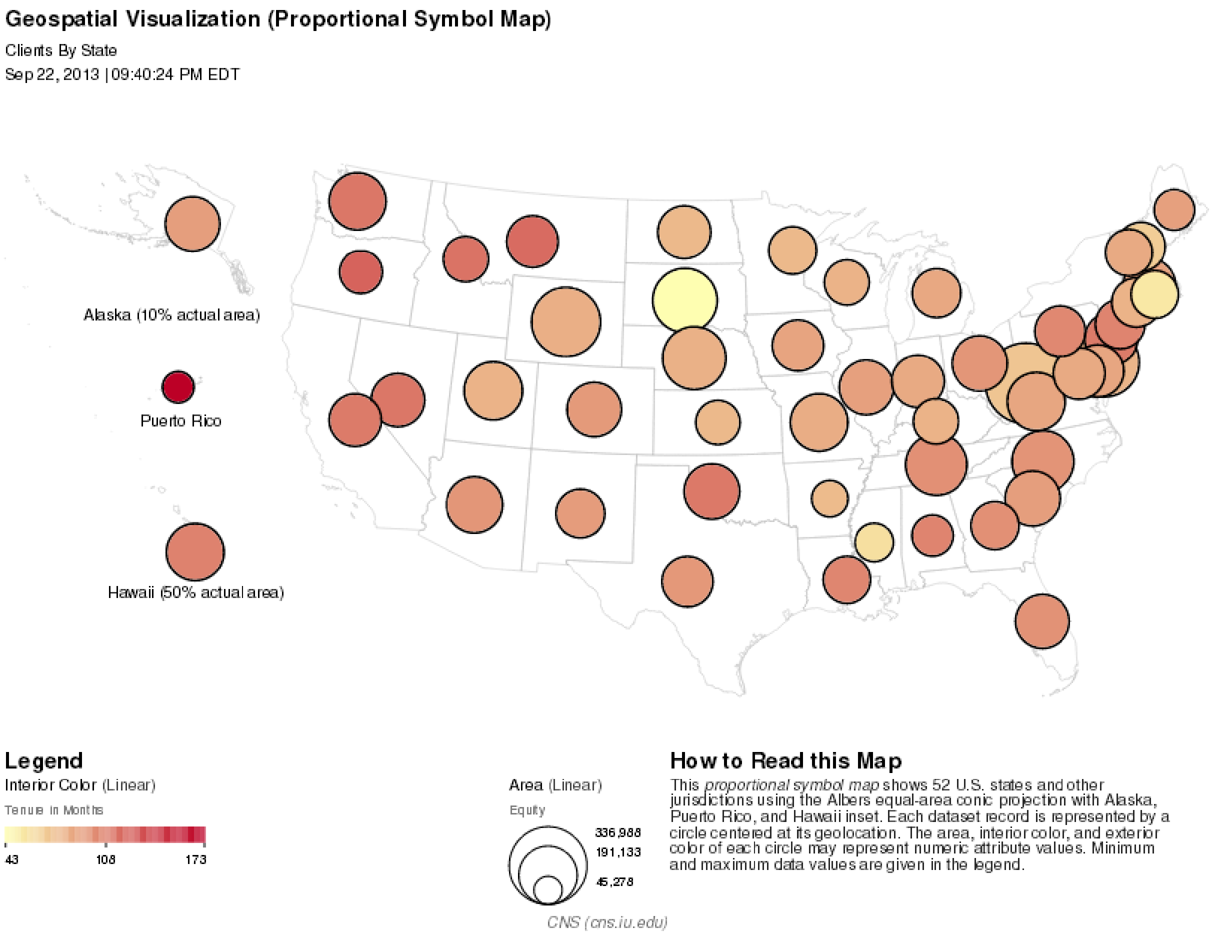
```
Proportional Symbol Map was selected.
Author(s): Joseph R. Biberstine
Implementer(s): Joseph R. Biberstine
Integrator(s): Joseph R. Biberstine
Documentation: http://wiki.cns.iu.edu/display/CISHELL/Proportional+Symbol+Map

Input Parameters:
Longitude: Longitude
Exterior Color Scaling: Linear
Color Circle Interiors By: Tenure in Months
Size Circles By: Equity
Exterior Color Range: White to Green
Interior Color Range: Yellow to Red
Size Scaling: Linear
Map: United States
Interior Color Scaling: Linear
Latitude: Latitude
Color Circle Exteriors By: None (uniform coloring)
Subtitle: Clients By State
Printing PostScript..
Done.
```

!	Algorithm Name	Date	Time	% Complete
☑	Proportional Symbol Map	09/22/2013	09:35:53 PM	🟩
☑	Generic Geocoder	09/22/2013	09:35:40 PM	🟩
☑	Aggregate Data	09/22/2013	09:34:54 PM	🟩
☑	Load...	09/22/2013	09:33:28 PM	🟩

- CSV file: C:\Users\Ted\AppData\Local\Temp\temp\Preprocessed-TDA IU Visualization Data-1
  - Aggregation performed using unique values in 'State' column.
    - With Latitude & Longitude from 'State'
      - Geospa al Symbol Map)

Geographic Distribution of Clients

You will have to convert the PostScript file to a PDF to view it. You can use Adobe Distiller to convert, or an online service, such as http://ps2pdf.com.



Geospatial Visualization (Proportional Symbol Map)

Clients By State

Sep 22, 2013 | 09:40:24 PM EDT

Legend

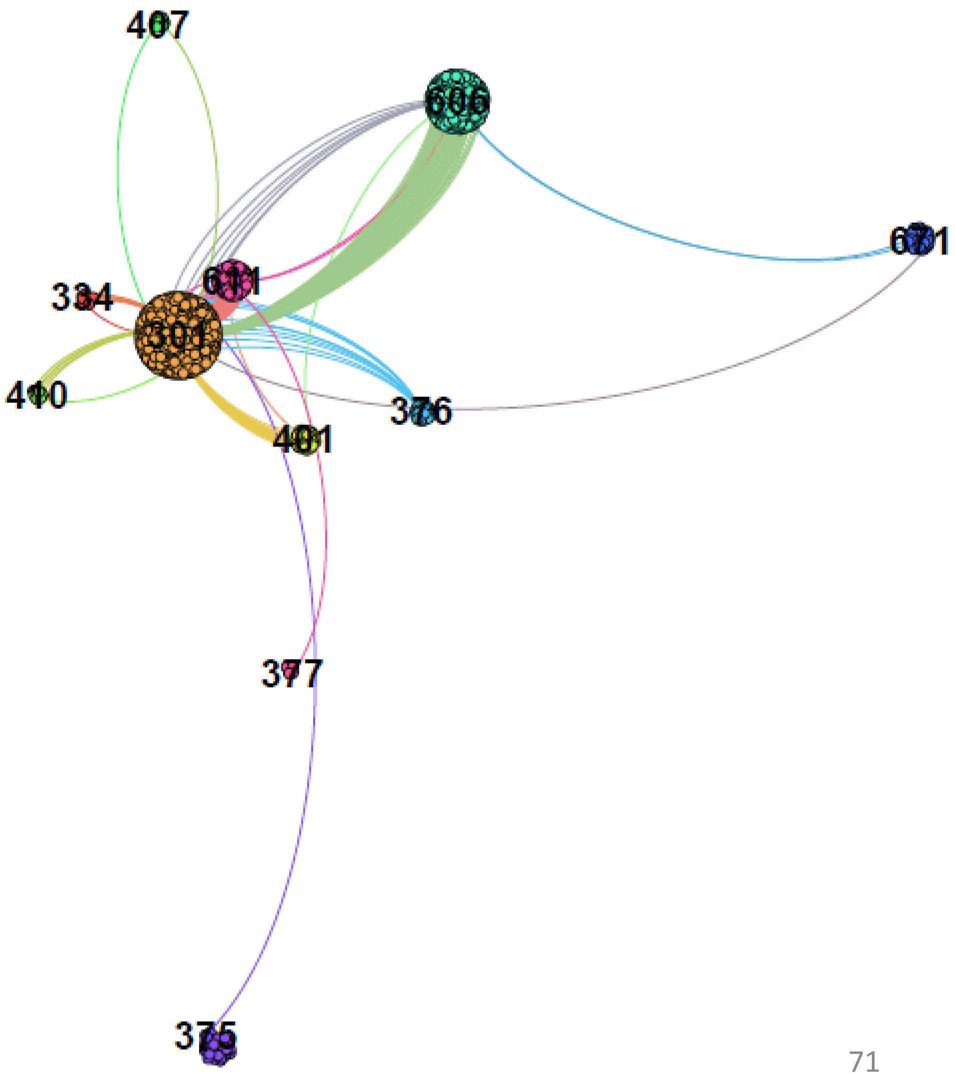
How to Read this Map

This *proportional symbol map* shows 52 U.S. states and other jurisdictions using the Albers equal-area conic projection with Alaska, Puerto Rico, and Hawaii inset. Each dataset record is represented by a circle centered at its geolocation. The area, interior color, and exterior color of each circle may represent numeric attribute values. Minimum and maximum data values are given in the legend.

CNS (cns.iu.edu)

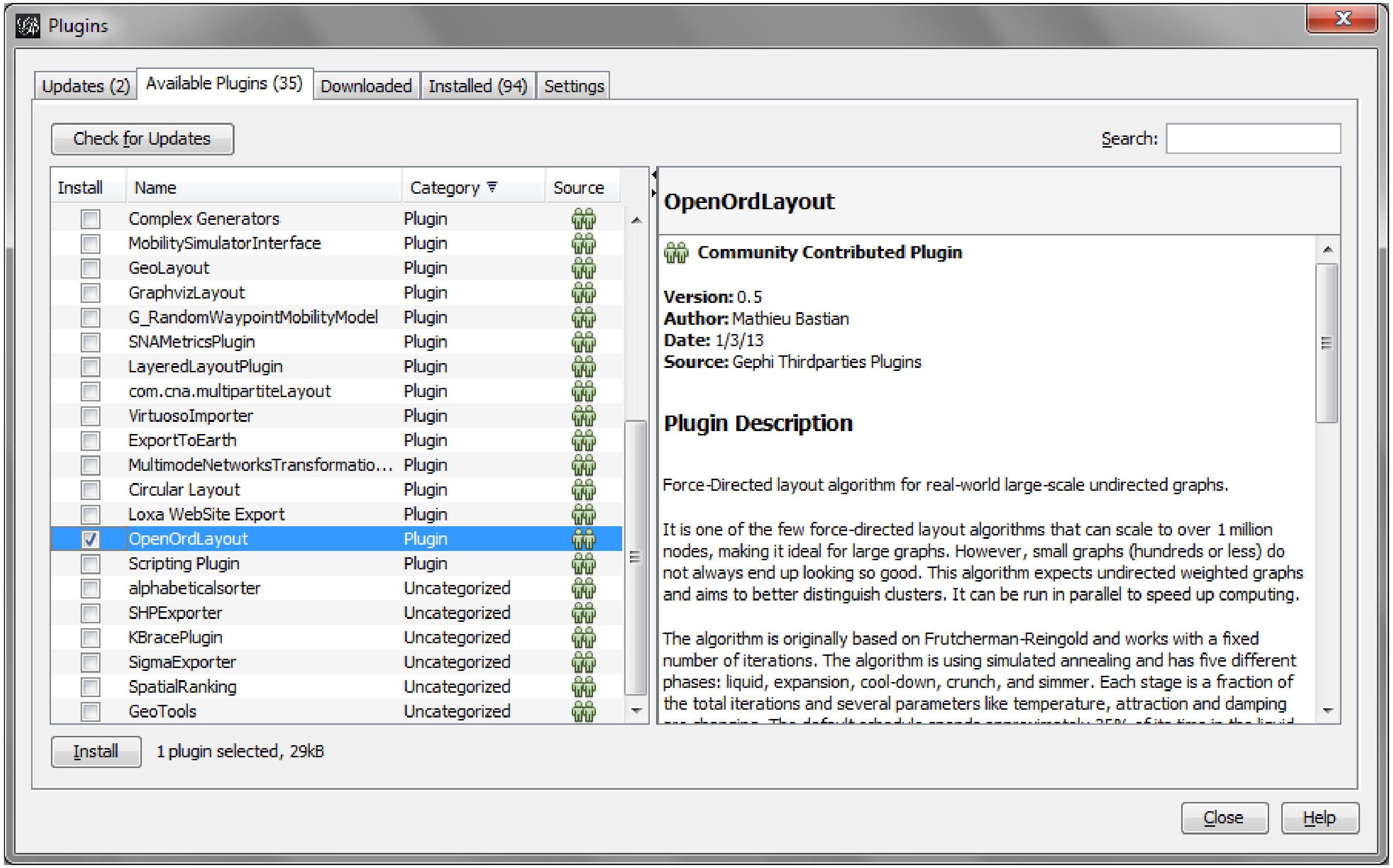
Client Brand Network

Extract a directed network showing the connections between clients and brands.



Client Brand Network

This network visualization requires the OpenOrd Network Layout in Gephi. It is available as a plugin. Simply open Gephi and select, *Tools > Plugins.* Then select the Available Plugins tab and find the OpenOrdLayout plugin. Install this plugin and restart Gephi.



Install	Name	Category	Source
☐	Complex Generators	Plugin	👥
☐	MobilitySimulatorInterface	Plugin	👥
☐	GeoLayout	Plugin	👥
☐	GraphvizLayout	Plugin	👥
☐	G_RandomWaypointMobilityModel	Plugin	👥
☐	SNAMetricsPlugin	Plugin	👥
☐	LayeredLayoutPlugin	Plugin	👥
☐	com.cna.multipartiteLayout	Plugin	👥
☐	VirtuosoImporter	Plugin	👥
☐	ExportToEarth	Plugin	👥
☐	MultimodeNetworksTransformatio…	Plugin	👥
☐	Circular Layout	Plugin	👥
☐	Loxa WebSite Export	Plugin	👥
☑	OpenOrdLayout	Plugin	👥
☐	Scripting Plugin	Plugin	👥
☐	alphabeticalsorter	Uncategorized	👥
☐	SHPExporter	Uncategorized	👥
☐	KBracePlugin	Uncategorized	👥
☐	SigmaExporter	Uncategorized	👥
☐	SpatialRanking	Uncategorized	👥
☐	GeoTools	Uncategorized	👥

1 plugin selected, 29kB

OpenOrdLayout

Community Contributed Plugin

Version: 0.5
Author: Mathieu Bastian
Date: 1/3/13
Source: Gephi Thirdparties Plugins

Plugin Description

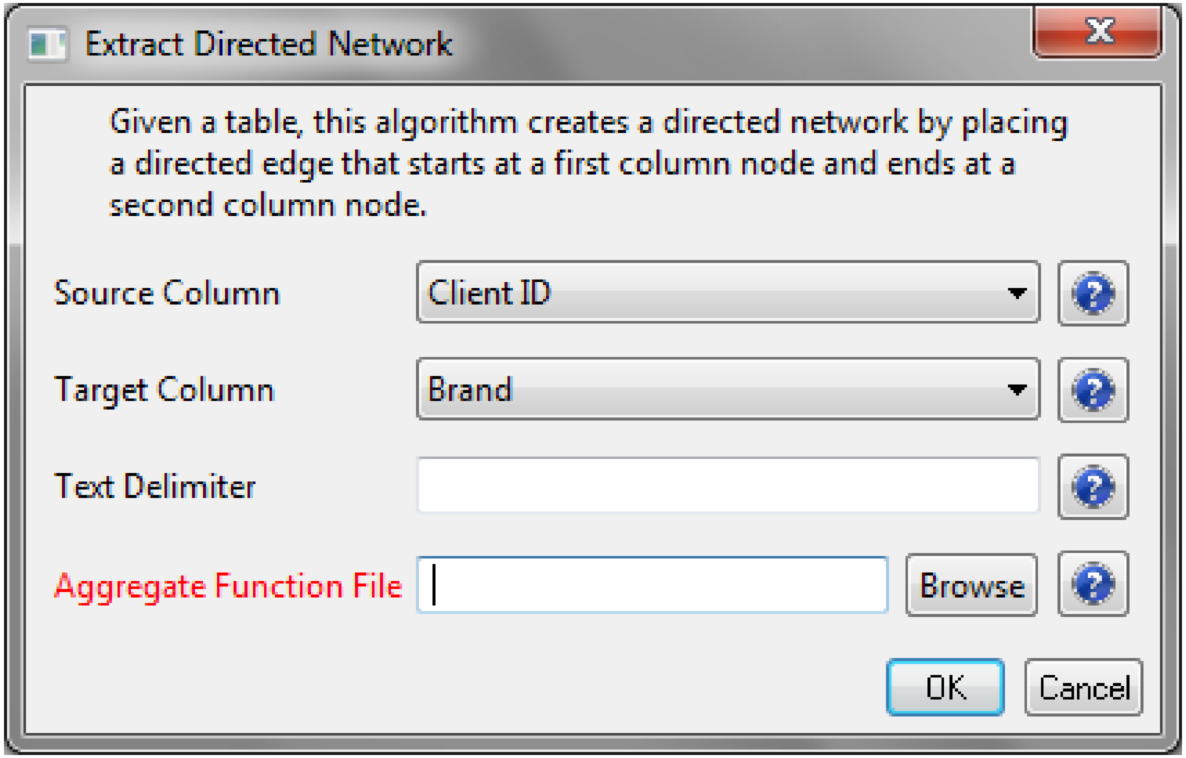
Force-Directed layout algorithm for real-world large-scale undirected graphs.

It is one of the few force-directed layout algorithms that can scale to over 1 million nodes, making it ideal for large graphs. However, small graphs (hundreds or less) do not always end up looking so good. This algorithm expects undirected weighted graphs and aims to better distinguish clusters. It can be run in parallel to speed up computing.

The algorithm is originally based on Frutcherman-Reingold and works with a fixed number of iterations. The algorithm is using simulated annealing and has five different phases: liquid, expansion, cool-down, crunch, and simmer. Each stage is a fraction of the total iterations and several parameters like temperature, attraction and damping

Client Brand Network

Select, *Data Preparation > Extract Directed Network*



Given a table, this algorithm creates a directed network by placing a directed edge that starts at a first column node and ends at a second column node.

Source Column: Client ID

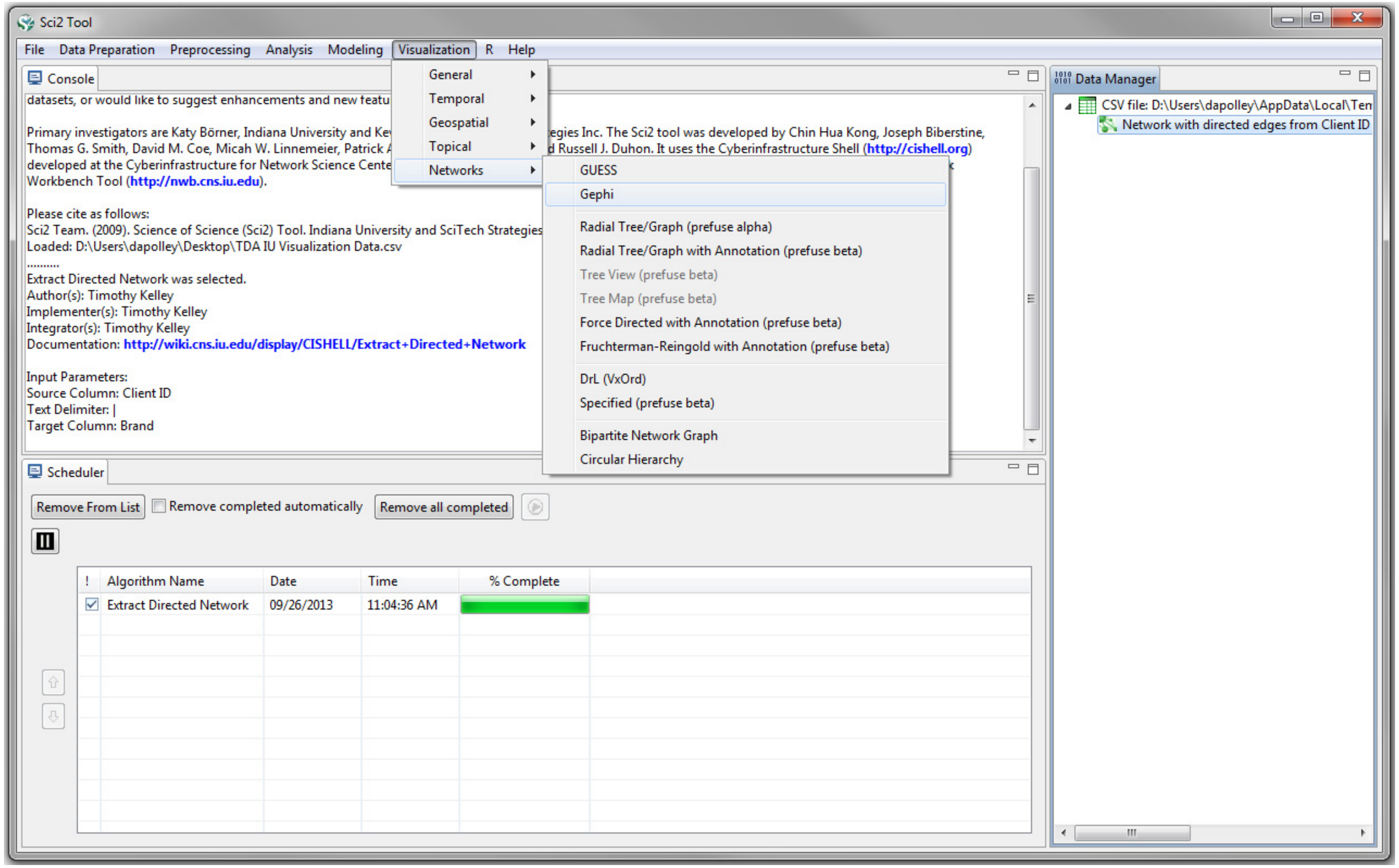
Target Column: Brand

Text Delimiter:

Aggregate Function File:

Client Brand Network

Visualize the network with Gephi, *Visualization> Networks > Gephi*



```
datasets, or would like to suggest enhancements and new featu

Primary investigators are Katy Börner, Indiana University and Ke    egies Inc. The Sci2 tool was developed by Chin Hua Kong, Joseph Biberstine,
Thomas G. Smith, David M. Coe, Micah W. Linnemeier, Patrick A    d Russell J. Duhon. It uses the Cyberinfrastructure Shell (http://cishell.org)
developed at the Cyberinfrastructure for Network Science Cente
Workbench Tool (http://nwb.cns.iu.edu).

Please cite as follows:
Sci2 Team. (2009). Science of Science (Sci2) Tool. Indiana University and SciTech Strategies
Loaded: D:\Users\dapolley\Desktop\TDA IU Visualization Data.csv
..........
Extract Directed Network was selected.
Author(s): Timothy Kelley
Implementer(s): Timothy Kelley
Integrator(s): Timothy Kelley
Documentation: http://wiki.cns.iu.edu/display/CISHELL/Extract+Directed+Network

Input Parameters:
Source Column: Client ID
Text Delimiter: |
Target Column: Brand
```

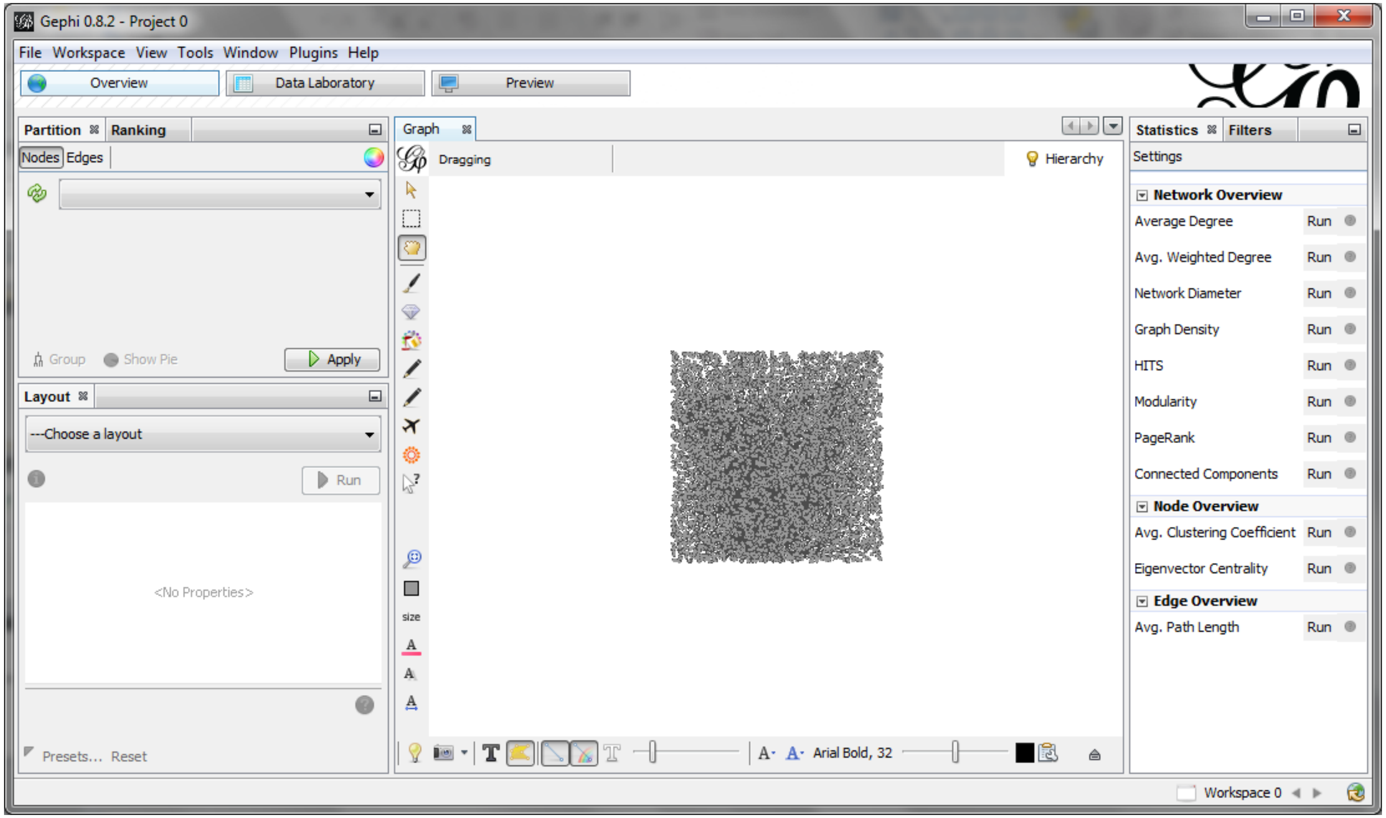
CSV file: D:\Users\dapolley\AppData\Local\Ten
- Network with directed edges from Client ID

□ Remove completed automatically

!	Algorithm Name	Date	Time	% Complete
☑	Extract Directed Network	09/26/2013	11:04:36 AM	🟩

Client Brand Network

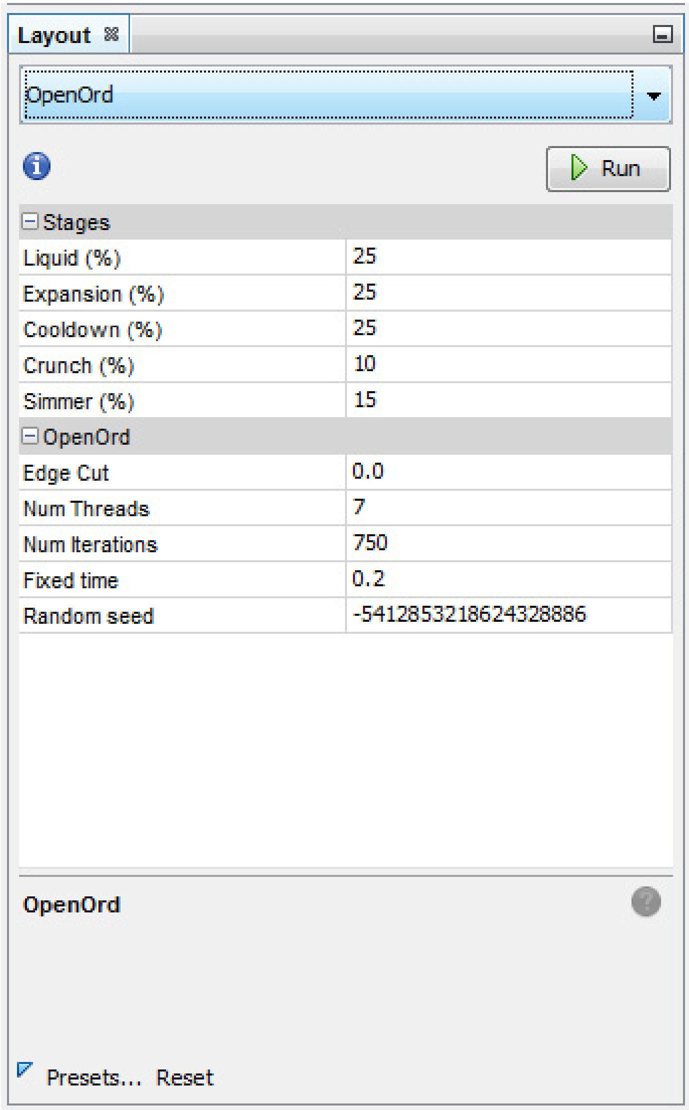
Gephi will layout the nodes randomly



CNS Cyberinfrastructure for Network Science Center

Client Brand Network

Select the OpenOrd layout. This layout is extremely efficient for large networks. Set the Edge Cut parameter to 0.0. This will result in a slightly less clustered result.



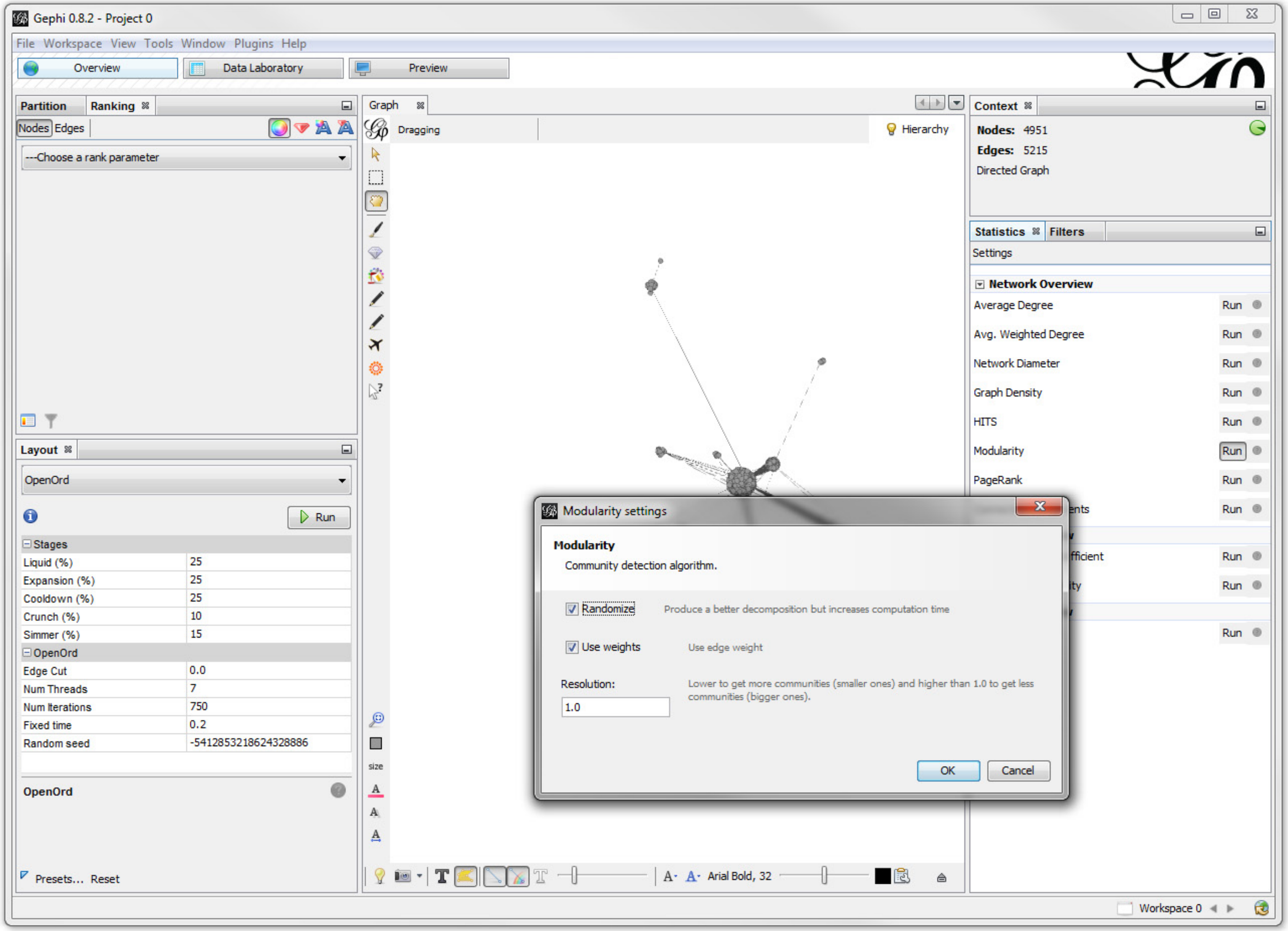
OpenOrd

⊟ Stages	
Liquid (%)	25
Expansion (%)	25
Cooldown (%)	25
Crunch (%)	10
Simmer (%)	15
⊟ OpenOrd	
Edge Cut	0.0
Num Threads	7
Num Iterations	750
Fixed time	0.2
Random seed	-5412853218624328886

OpenOrd

Client Brand Network

In the statistics tab, run Modularity. This identifies communities of clients based on the brands with which they are associated.



Nodes: 4951
Edges: 5215
Directed Graph

Stages	
Liquid (%)	25
Expansion (%)	25
Cooldown (%)	25
Crunch (%)	10
Simmer (%)	15
OpenOrd	
Edge Cut	0.0
Num Threads	7
Num Iterations	750
Fixed time	0.2
Random seed	-5412853218624328886

Modularity settings

Modularity

Community detection algorithm.

☑ Randomize — Produce a better decomposition but increases computation time

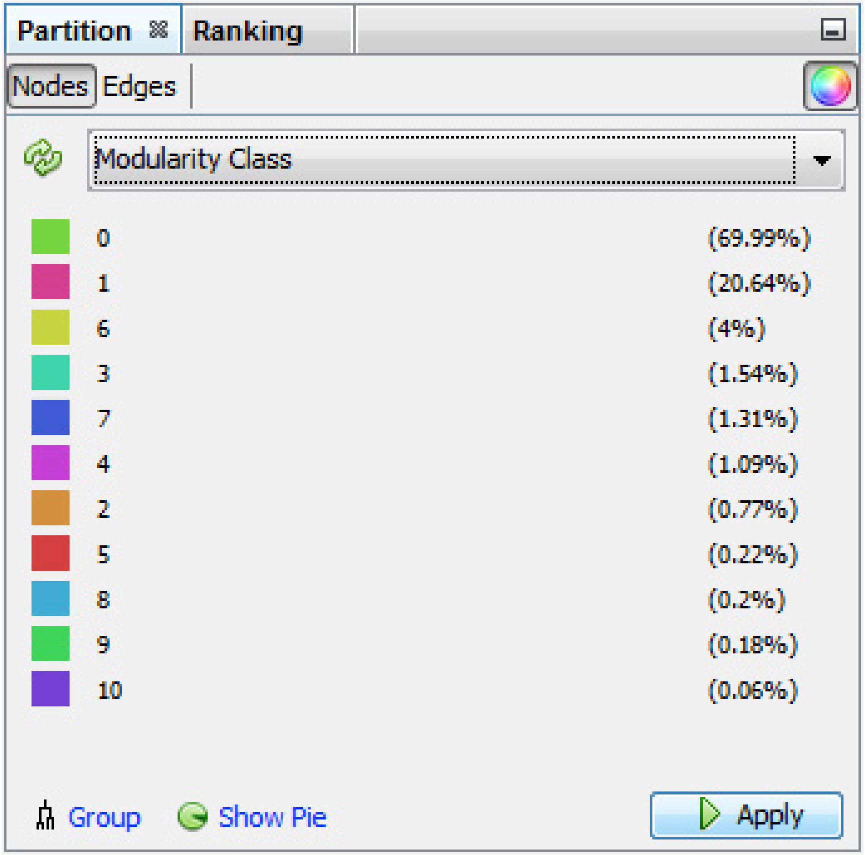
☑ Use weights — Use edge weight

Resolution: 1.0 — Lower to get more communities (smaller ones) and higher than 1.0 to get less communities (bigger ones).

OK Cancel

Client Brand Network

Use the partition tab in the upper right-hand corner. Select Modularity Class. You may have to click refresh. Select Apply and the colors shown will be applied to each modularity class.

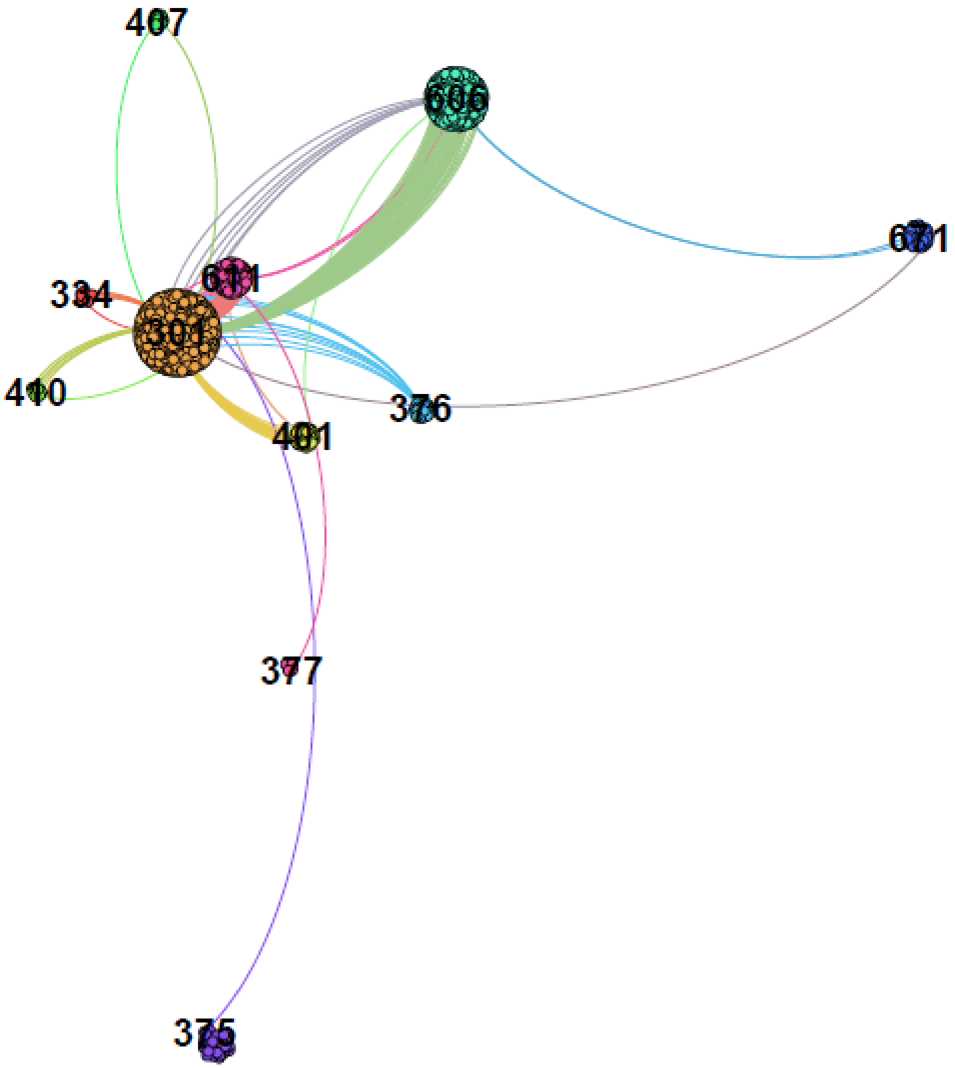


Modularity Class

🟩	0	(69.99%)
🟪	1	(20.64%)
🟨	6	(4%)
🟩	3	(1.54%)
🟦	7	(1.31%)
🟪	4	(1.09%)
🟧	2	(0.77%)
🟥	5	(0.22%)
🟦	8	(0.2%)
🟩	9	(0.18%)
🟪	10	(0.06%)

Client Brand Network

This presents a network where the nodes are distinct clusters based on brand. You can determine brand popularity based on cluster size. Each brand cluster will have a distinct color. There are also a lot of connections between brand clusters.



Brainstorming/Questions?