

# ANALYZING AND VISUALIZING CORRESPONDENCE NETWORKS FOR BROWSABLE INTERFACES

**Scott Weingart**

Cyberinfrastructure for Network Science Center  
Information Visualization Laboratory  
Indiana Philosophy Ontology Project

School of Library and Information Science  
Department of History and Philosophy of Science

Indiana University, Bloomington, IN

<http://www.scottbot.net>

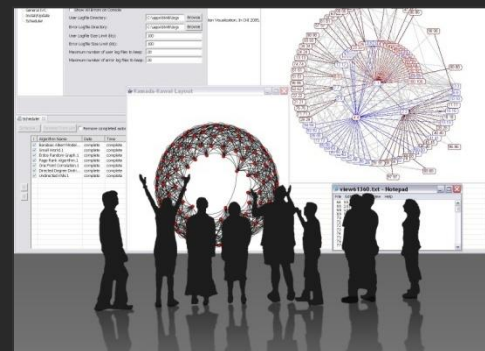
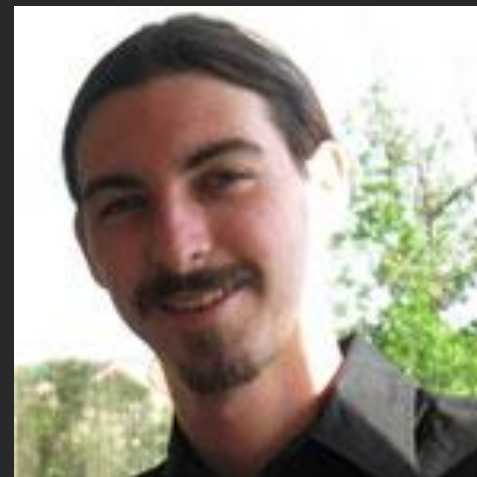
*Thanks to Katy Borner, The Sci2 Team, and Huygens ING.*

*Representing The Republic of Letters, 6/30/2011-7/1/2011*

*Huygens ING – Institute for Dutch History*

*The Hague, The Netherlands*

*10:50-12:20 on July 1, 2011*





## Workshop Overview

**10:55-11:05 The Importance and Dangers of Visualization – Use & Theory**

**11:05-11:20 The Epistolarium – Networks, Topics & Tools**

**11:20-11:25 Computational Modeling**

***11:25-11:35 Move to other room***

**11:35-11:50 Sci2 Tool Basics**

**11:50-12:10 Sci2 Workflow Design: Padgett's Florentine Families - Prepare, load, analyze, and visualize family and business networks from 15th century Florence.**

**12:10-12:20 Q&A and Technical Assistance**



## Workshop Overview

**10:55-11:05 The Importance and Dangers of Visualization – Use & Theory**

11:05-11:20 The Epistolarium – Networks, Topics & Tools

11:20-11:25 Computational Modeling

*11:25-11:35 Move to other room*

11:35-11:50 Sci2 Tool Basics

11:50-12:10 Sci2 Workflow Design: Padgett's Florentine Families - Prepare, load, analyze, and visualize family and business networks from 15th century Florence.

12:10-12:20 Q&A and Technical Assistance

# The Importance and Dangers of Visualization: Use & Theory



## Uses of Visualization

- **Solidifies objects of inquiry**
- **Exploration**
- **Discovery**
- **Trend-spotting**
- **Evidence**
- **Audience Engagement**
- **Engaging public / funding agencies**



## The Importance of Visualization

[Visualization] aim at more than making the invisible visible. [It aspires] to all-at-once-ness, the condensation of laborious, step-by-step procedures in to an immediate *coup d'oeil*... What was a painstaking process of calculation and correlation—for example, in the construction of a table of variables—becomes a flash of intuition. And all-at-once intuition is traditionally the way that angels know, in contrast to the plodding demonstrations of humans.

Descartes's craving for angelic all-at-once-ness emerged forcefully in his mathematics..., compressing the steps of mathematical proof into a single bright flare of insight: "I see the whole thing at once, by intuition."

*Lorraine Daston – On Scientific Observation*



## Warnings

[H]umanists have adopted many applications such as GIS mapping, graphs, and charts for statistical display that were developed in other disciplines... such graphical tools are a kind of intellectual Trojan horse...

Data pass themselves off as mere descriptions of a priori conditions. Rendering *observation* (the act of creating a statistical, empirical, or subjective account or image) as if it were *the same as the phenomena observed* collapses the critical distance between the phenomenal world and its interpretation, undoing the basis of interpretation on which humanistic knowledge production is based... we seem ready and eager to suspend critical judgment in a rush to visualization.

*Johanna Drucker – Humanities Approaches to Graphical Display*



## Warnings

- **Data format limits use, already an act of interpretation.**
- **Statistics is often misused (wield it very carefully).**
- **Interpreting spatial distance as meaningful.**
- **Always include a legend (this presentation breaks that rule).**
- **Accidental legitimization in eyes of public.**





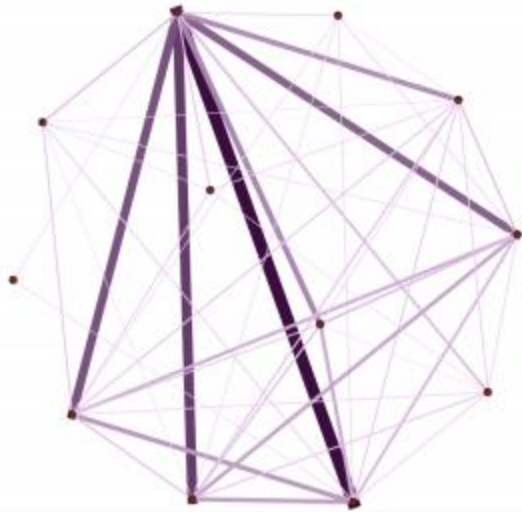
## Character Networks in the 19<sup>th</sup> Century British Novel -Graham Sack

I use computational methods to count the frequency and co-occurrence of a generally ignored sub-class of common words, namely, character names. Character names are often regarded as noise and excluded from authorship and stylistics analysis because they are not consistent across texts. This study makes character names its main object of analysis because the objective is quite different: rather than style or authorship, this study attempts to make inferences about *characterization* and *social form*, two areas about which computational analysis has had comparatively little to say.

# Character Network Sociograms

Figures 20 a, b, & C

The Ambassadors (James)



## General Features:

- Small network (12 characters)
- No isolates
- Very high graph density (71%) and clustering coefficient (85%)
- Low average path length (1.3)
- Low degree inequality (-4.9)
- High proportion of strong ties (28%)

## Conclusions

- Tightly knit social world focused on deep relationships between small set of characters
- Social interaction broadly evenly distributed

Middlemarch (Eliot)



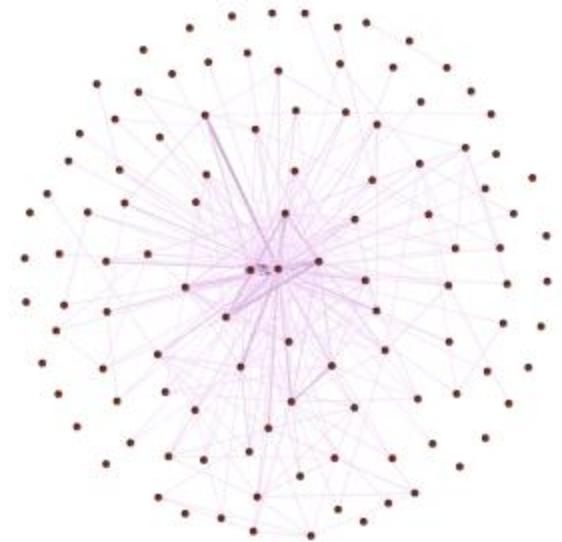
## General Features:

- Large network (99 characters)
- Moderately high % of isolates (17%)
- Low graph density (7%) and clustering coefficient (73%)
- High average path length (2.4)
- Moderate degree inequality (1.9)
- Moderate proportion of strong ties (18%)

## Conclusions

- Large but comparatively integrated social world with deep interaction between core characters

The Pickwick Papers (Dickens)



## General Features:

- Large network (112 characters)
- High proportion of isolates (20%)
- Very low graph density (4%) and clustering coefficient (72%)
- High average path length (2.2)
- High degree inequality (3.0)
- Low proportion of strong ties (13%)

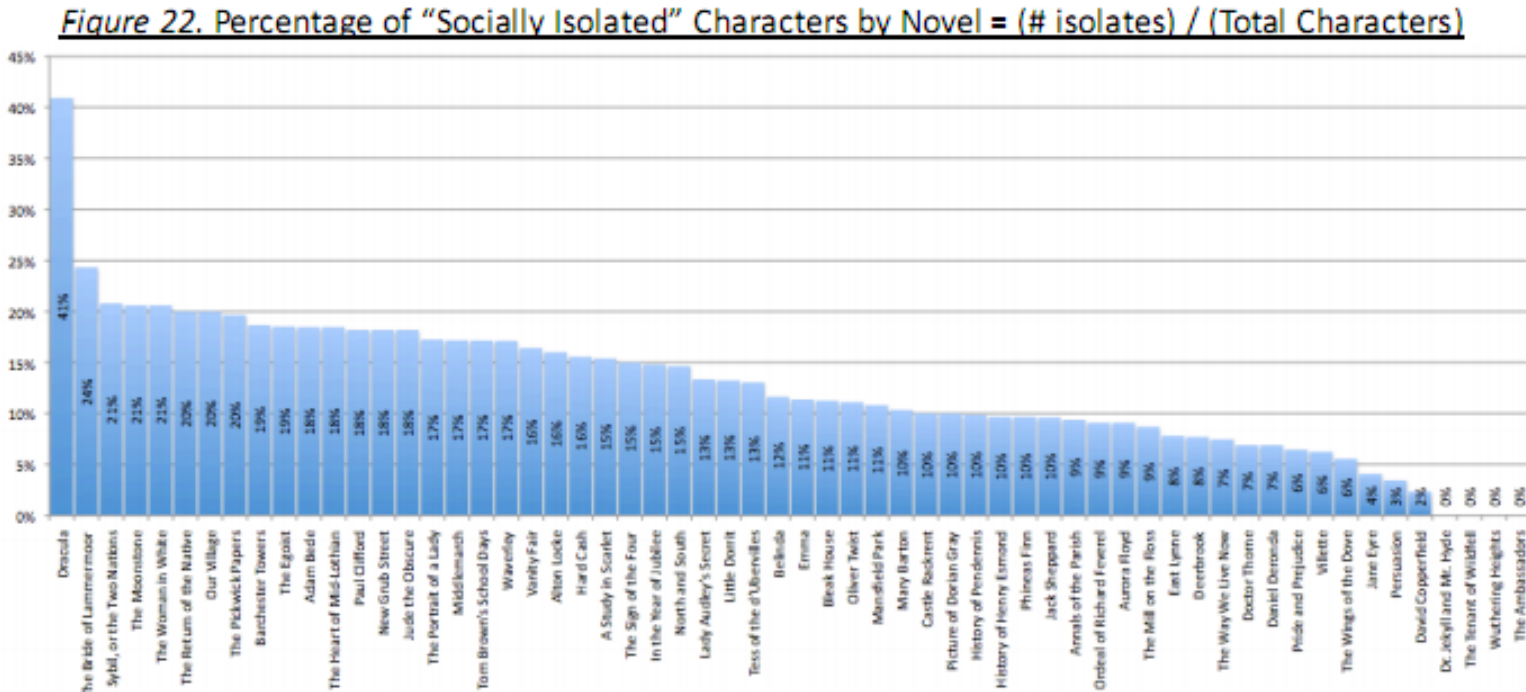
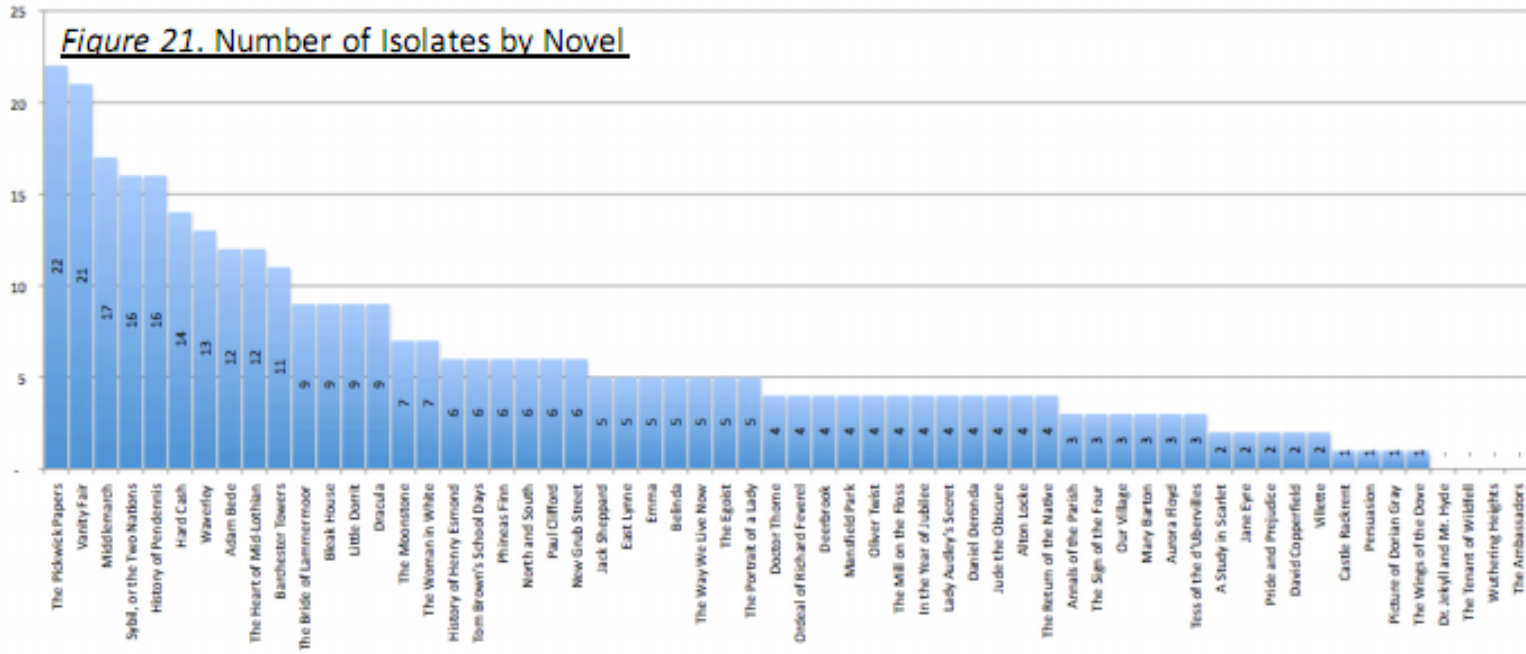
## Conclusions

- Expansive but diffuse social world with passing social interactions and many isolated characters



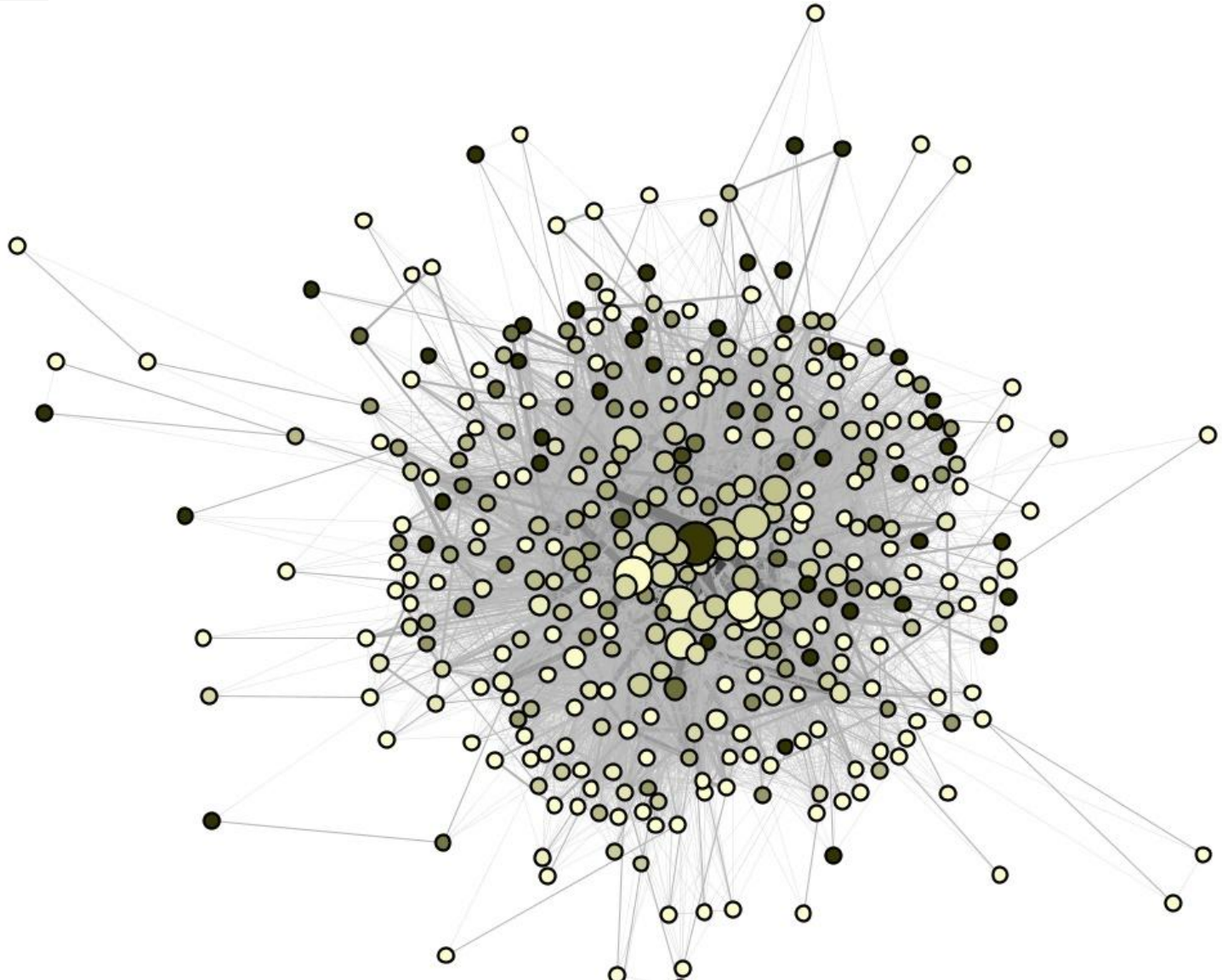
# Character Networks in the 19<sup>th</sup> Century British Novel - Graham Sack

## Social Metrics – By Novel (1/3)





# Word Co-Occurrences in European Fairytales -Jorgensen & Weingart





## Workshop Overview

10:55-11:05 The Importance and Dangers of Visualization – Use & Theory

**11:05-11:20 The Epistolarium – Networks, Topics & Tools**

11:20-11:25 Computational Modeling

*11:25-11:35 Move to other room*

11:35-11:50 Sci2 Tool Basics

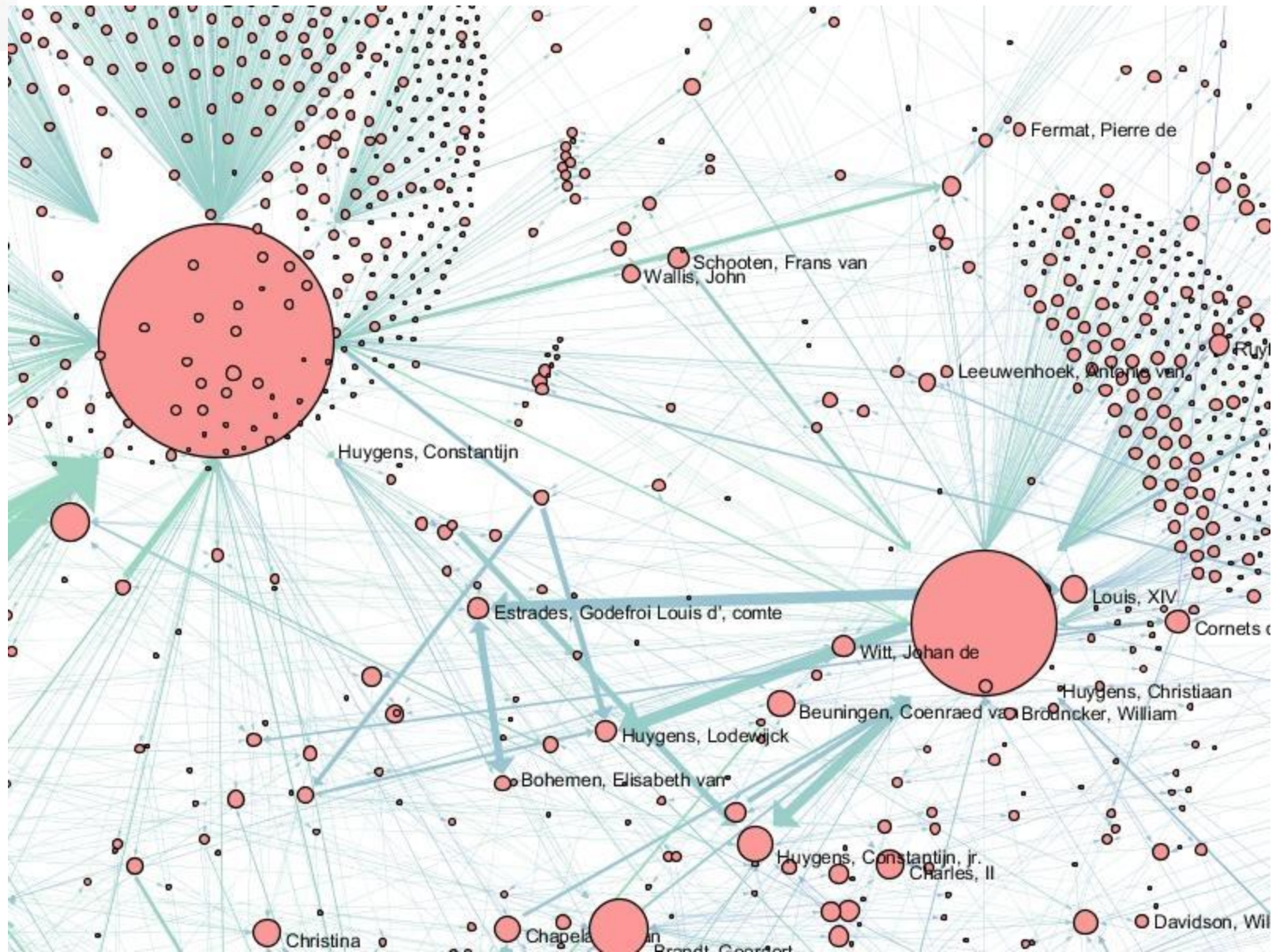
11:50-12:10 Sci2 Workflow Design: Padgett's Florentine Families - Prepare, load, analyze, and visualize family and business networks from 15th century Florence.

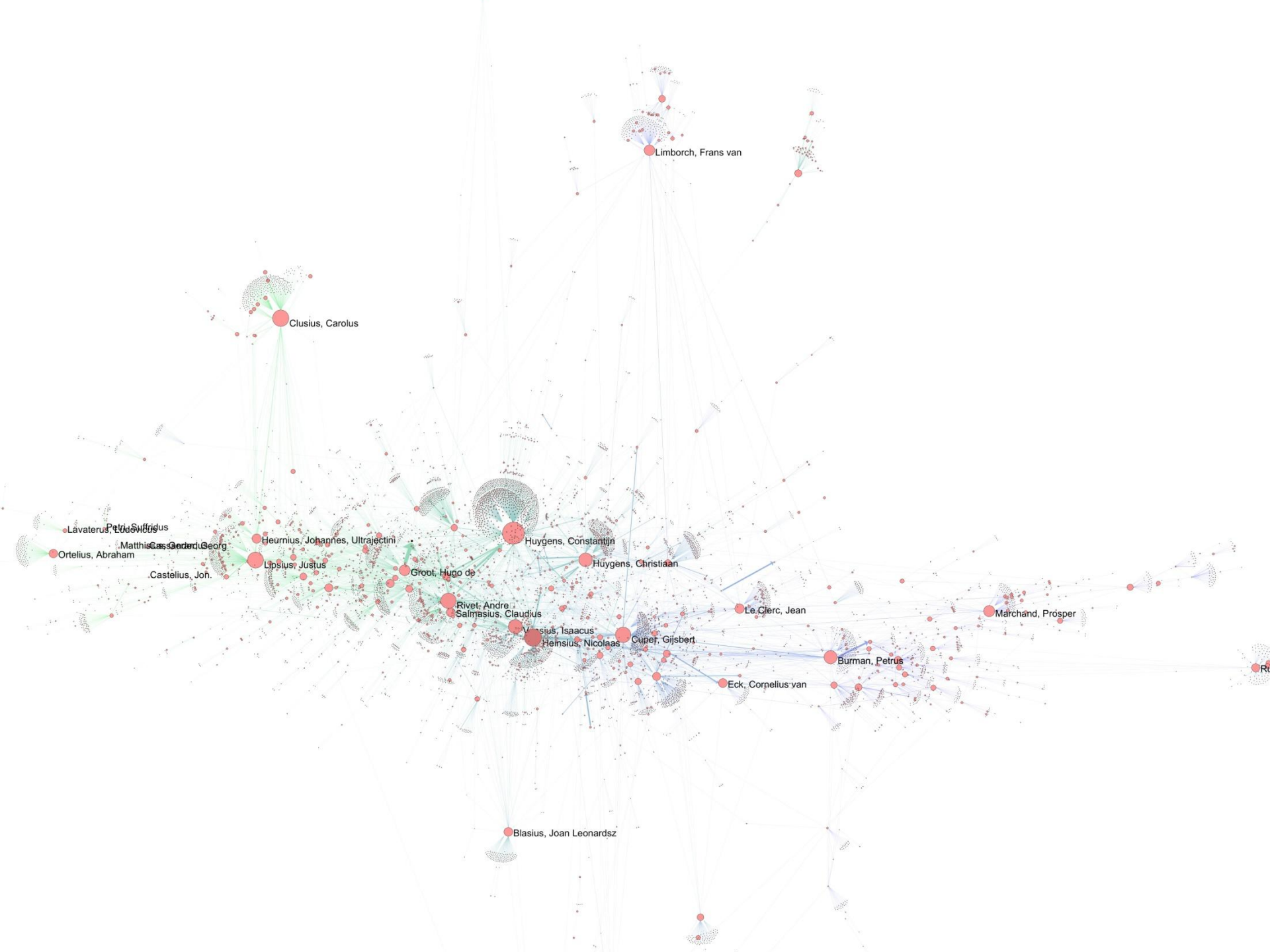
12:10-12:20 Q&A and Technical Assistance

# The Epistolarium – Networks, Topics & Tools



# Networks





Clusius, Carolus

Limborch, Frans van

Lavaterus, Petrus

Ortelius, Abraham

Matthias Caspar, Georg

Castellius, Joh.

Heurnius, Johannes, Ultrajectini

Lipsius, Justus

Groot, Hugo de

Rivet, Andre

Salmasius, Claudius

Wassius, Isaacus

Hensius, Nicolaas

Cuper, Gijsbert

Blasius, Joan Leonardsz

Huygens, Constantijn

Huygens, Christiaan

Le Clerc, Jean

Eck, Corpelius van

Burman, Petrus

Marchand, Prosper

R...





## Networks

Correspondence, Citation, and Co-Citation Networks show no less nor more information than is already available to the researcher, and is subject to the same biases.

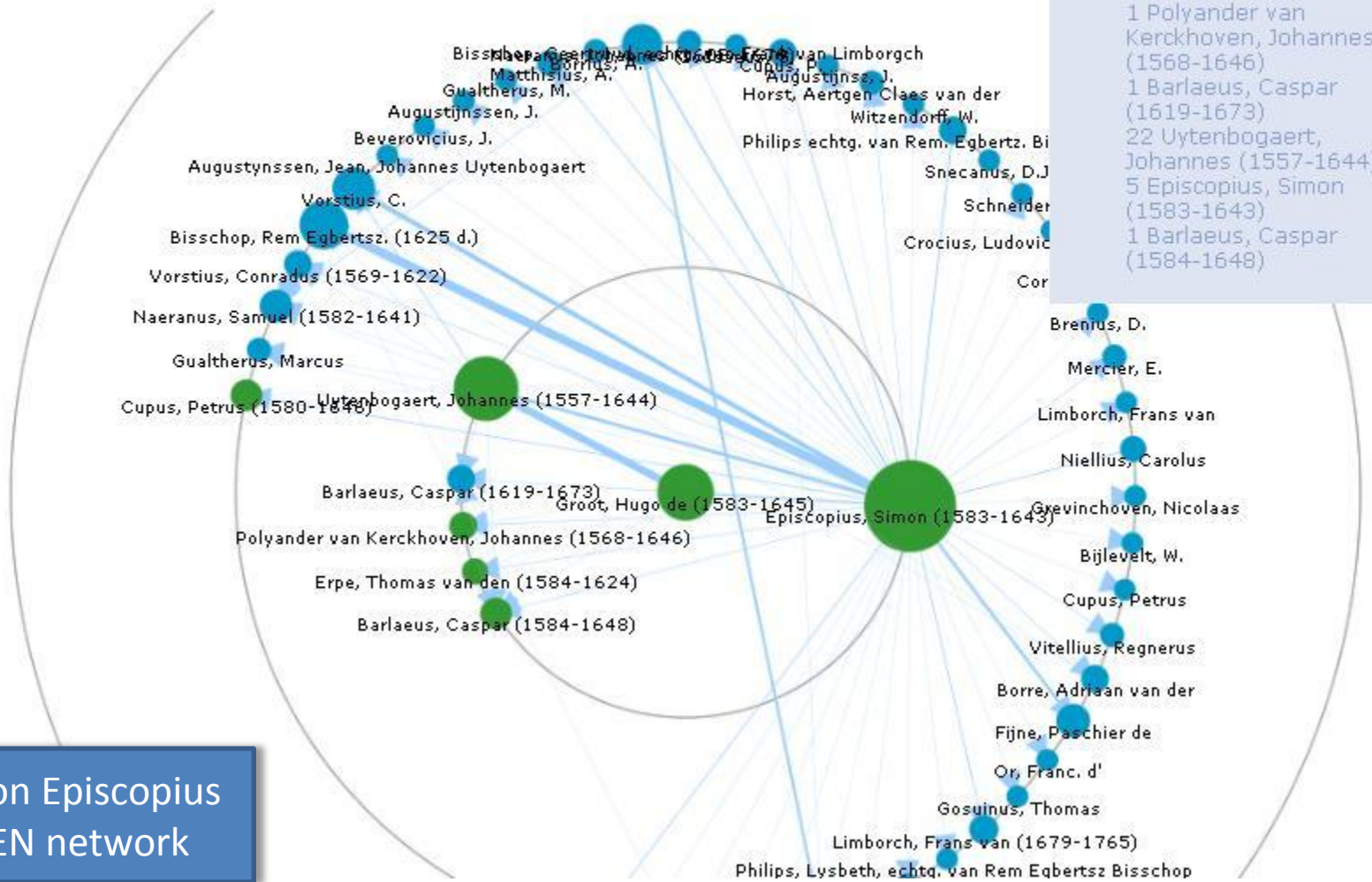
These networks show that same information in a new light, and allows us to ask new sorts of questions, rethink our objects of inquiry, and systematize our methods of large scale analysis and comparison.



# The Epistolarium



Embedding in CEN Network



## Groot, Hugo de (1583-1645)

- 1 Erpe, Thomas van den (1584-1624)
- 1 Polyander van Kerckhoven, Johannes (1568-1646)
- 1 Barlaeus, Caspar (1619-1673)
- 22 Uytenbogaert, Johannes (1557-1644)
- 5 Episcopus, Simon (1583-1643)
- 1 Barlaeus, Caspar (1584-1648)

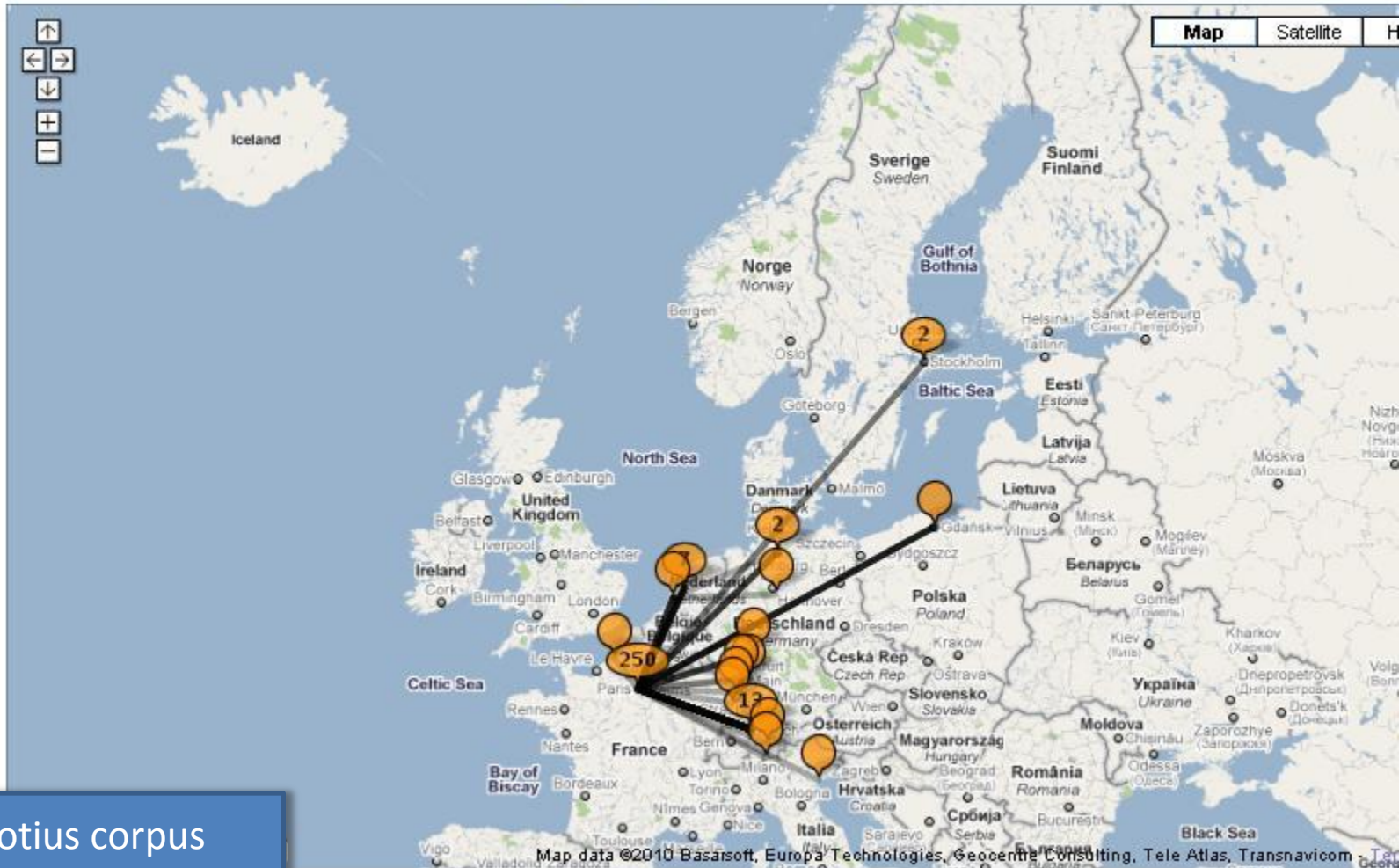
## Simon Episcopus in CEN network



# The Epistolarium

311 Letters filtered from 1098 originally ([Reset All Filters](#))

25 results out of 311 cannot be plotted.



Grotius corpus  
French letters



# The Epistolarium

ckcc

CKCC Epistolarium

**Topics** 1

- 633 letters
- 559 French-diplomacy
- 533 diplomacy
- 295 Breda-siege**
- 219 campaign
- 96 European-diplomacy

**Persons**

- 1 Rabennoord, Carl
- 5 Rivet, André (1572-1651)
- 1 Saige, Guillaume le (-1637)
- 242 Solms, Amalia van (1602-1675)

**Locations**

- 29 Assenede, BE
- 6 Bergen op Zoom (Noordgeest), NL
- 13 Bergen op Zoom, NL
- 59 Breda, NL
- 13 Budberg, DE

**Language**

295 French

Wordcloud Timeline Map

Letters filtered from 2312 originally ([Reset All Filters](#))





## Topics





# The Epistolarium

ckcc

CKCC Epistolarium

## Topics

361 letters	▲
339 mathematics	■
144 astronomy	■
106 salutation	■
74 diplomacy	■
59 geometry	▼

## Persons

49 ?	▲
2 Académie des Sciences (Paris)	■
4 Acta Eruditorum	■
2 Alberghetti, Sigismondo (-1702)	■
1 Albert de Lange	▼

## Locations

795 ?	▲
38 Den Haag, NL	■
9 Dordrecht, NL	■
11 Hoorn (NH), NL	■
1 Leiden, NL	■
4 Leipzig, DE	▼

## Language

797 Latin	■
-----------	---

Wordcloud

Timeline

Map

astronomy barometry books botany dioptrics-telescopes diplomacy Dutch-diplomacy family-news French-Dutch-relations geometry Hugo-and-Jan-Grotius Hugo-and-Willem-Grotius laws-of-motion Leiden-university **letters** **mathematics** military-affairs news-about-scientists peace-treaties poetry religion salutation scholastic-theology theology trigonometry valediction

Chr. Huygens corpus  
Latin letters



# The Epistolarium



CKCC Epistolarium

## Topics

- 21 astronomy
- 20 letters
- 11 salutation
- 2 dioptrics-telescopes
- 2 mathematics
- 1 books

## Persons

- 2 Hesius, Guilielmus (1601-1690)
- 2 Heuraet, Hendrik van (1633-)
- 26 Hevelius, Johannes (1611-1687)**
- 1 Hire, Philippe de la

## Locations

- 26 ?

## Language

- 26 Latin

Wordcloud

Timeline

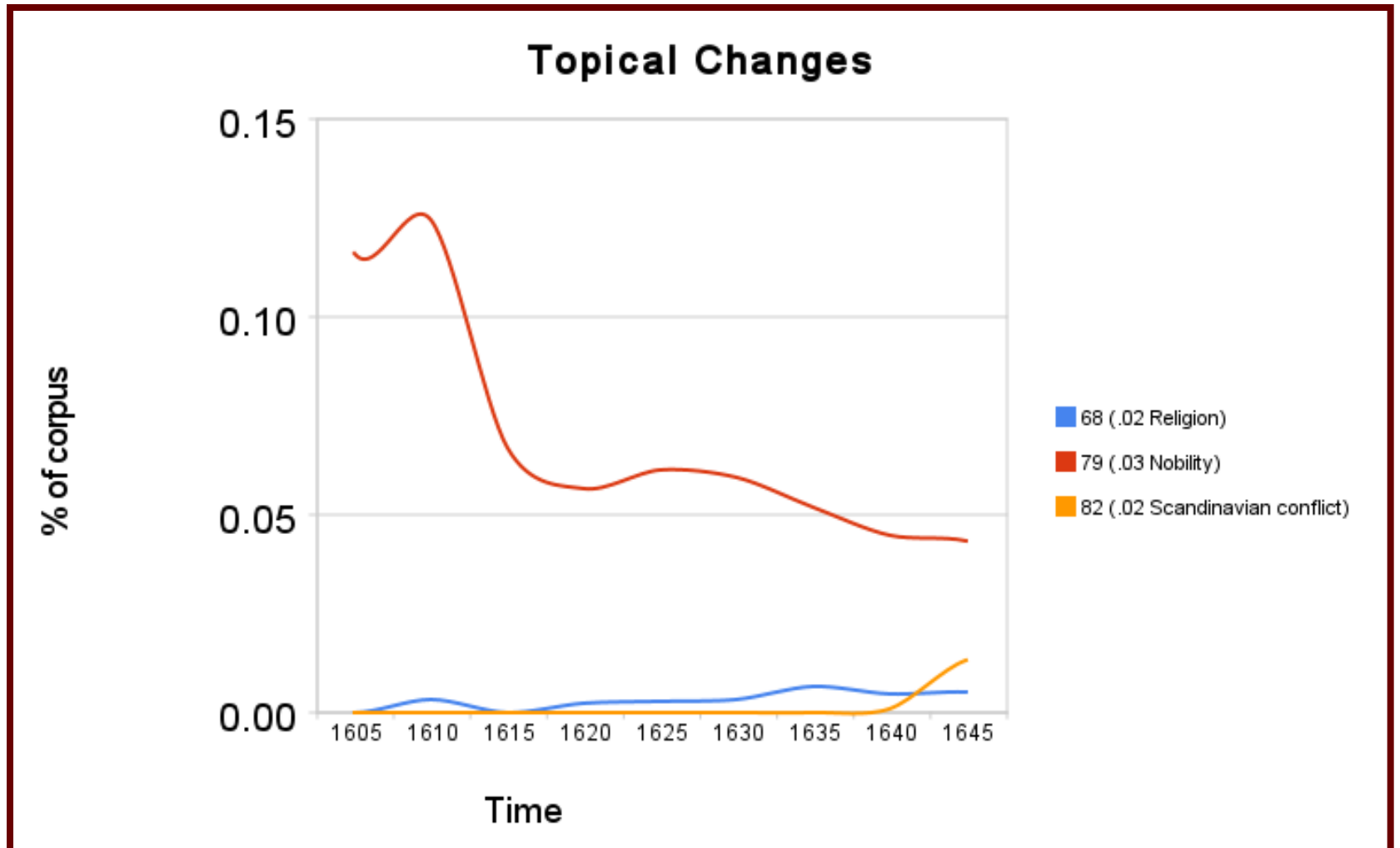
Map

astronomy books dioptrics-telescopes letters mathematics salutation

Chr. Huygens corpus  
Latin letters



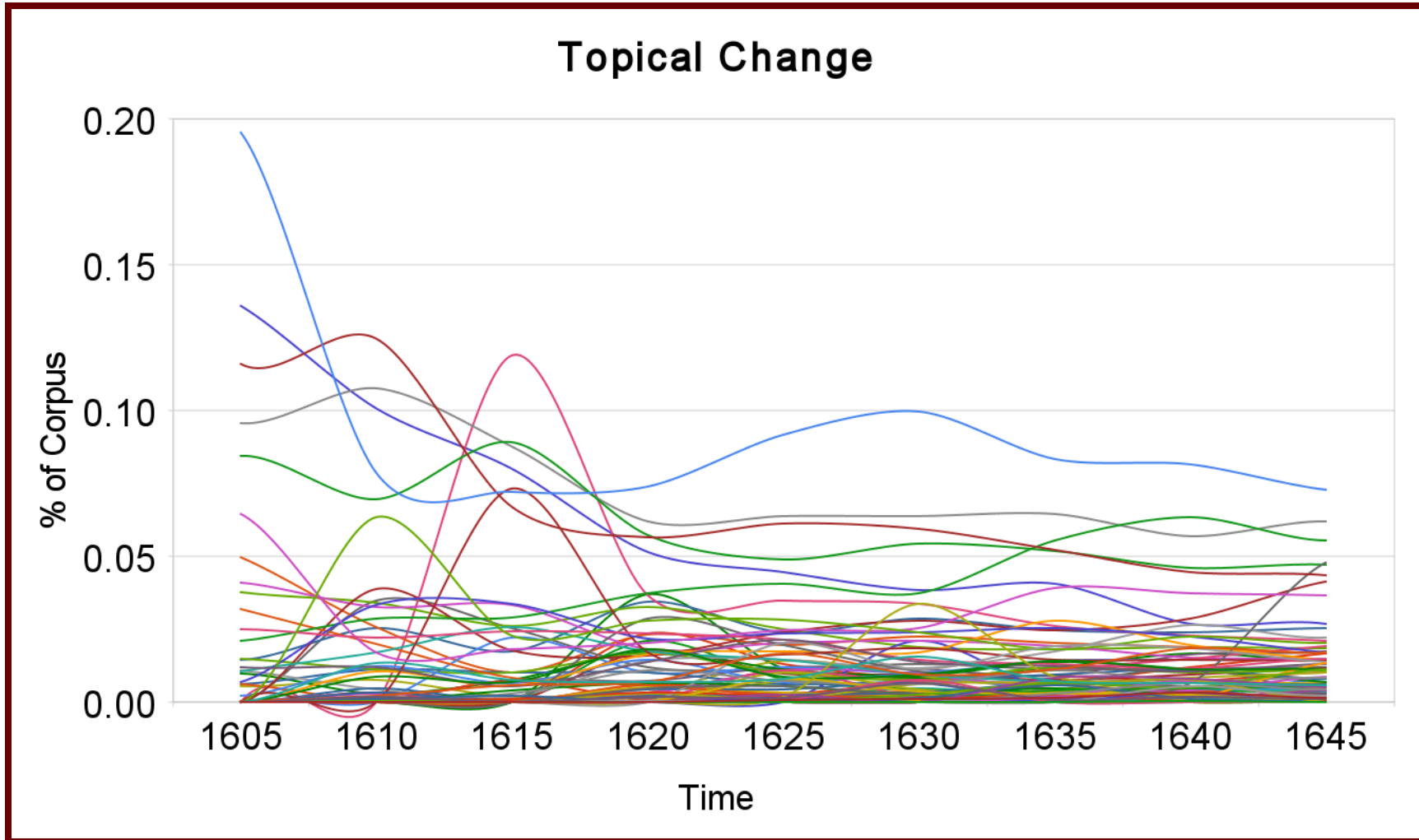
## Topics







## Topics





## Workshop Overview

10:55-11:05 The Importance and Dangers of Visualization – Use & Theory

11:05-11:20 The Epistolarium – Networks, Topics & Tools

**11:20-11:25 Computational Modeling**

*11:25-11:35 Move to other room*

11:35-11:50 Sci2 Tool Basics

11:50-12:10 Sci2 Workflow Design: Padgett's Florentine Families - Prepare, load, analyze, and visualize family and business networks from 15th century Florence.

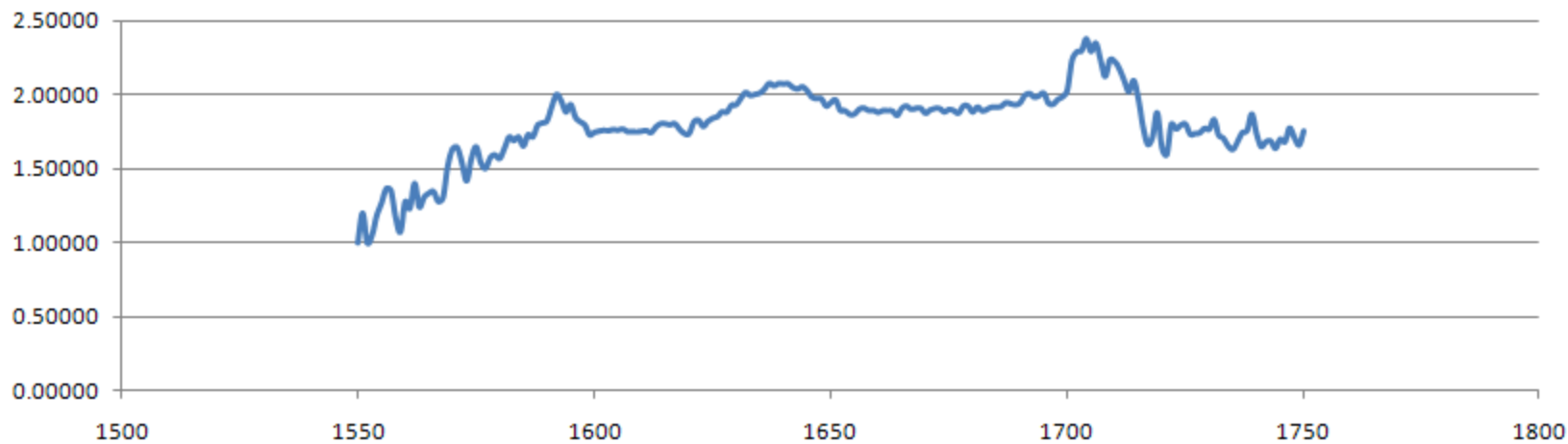
12:10-12:20 Q&A and Technical Assistance

# Computational Modeling

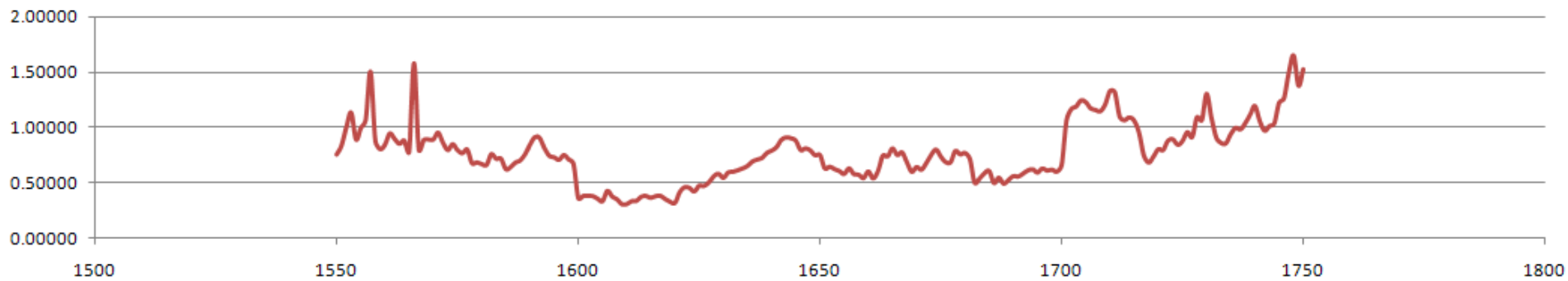


## CEN Statistics

### Average Total Degree



### Weight\_mean



Quality = .8

Quality = .24

Idea 1

Idea 2

Scholar 1

Scholar 2

Scholar 3

Strength: 4

Strength: 1

Strength: 1

Strength: 3

Age: 12, Lifespan: 50

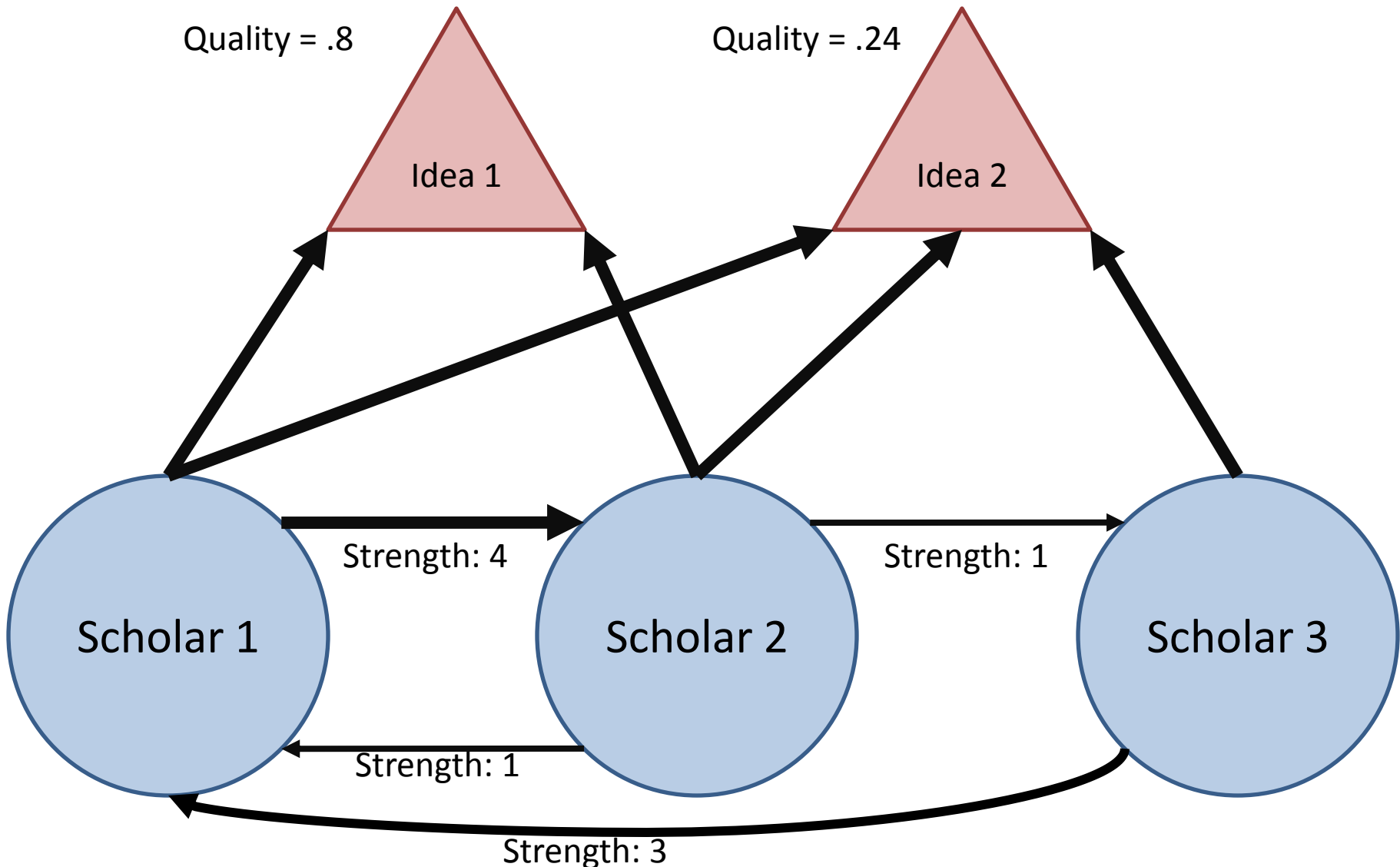
Age: 40, Lifespan: 85

Age: 18, Lifespan: 80

Quality of new ideas: .4 +/- .05

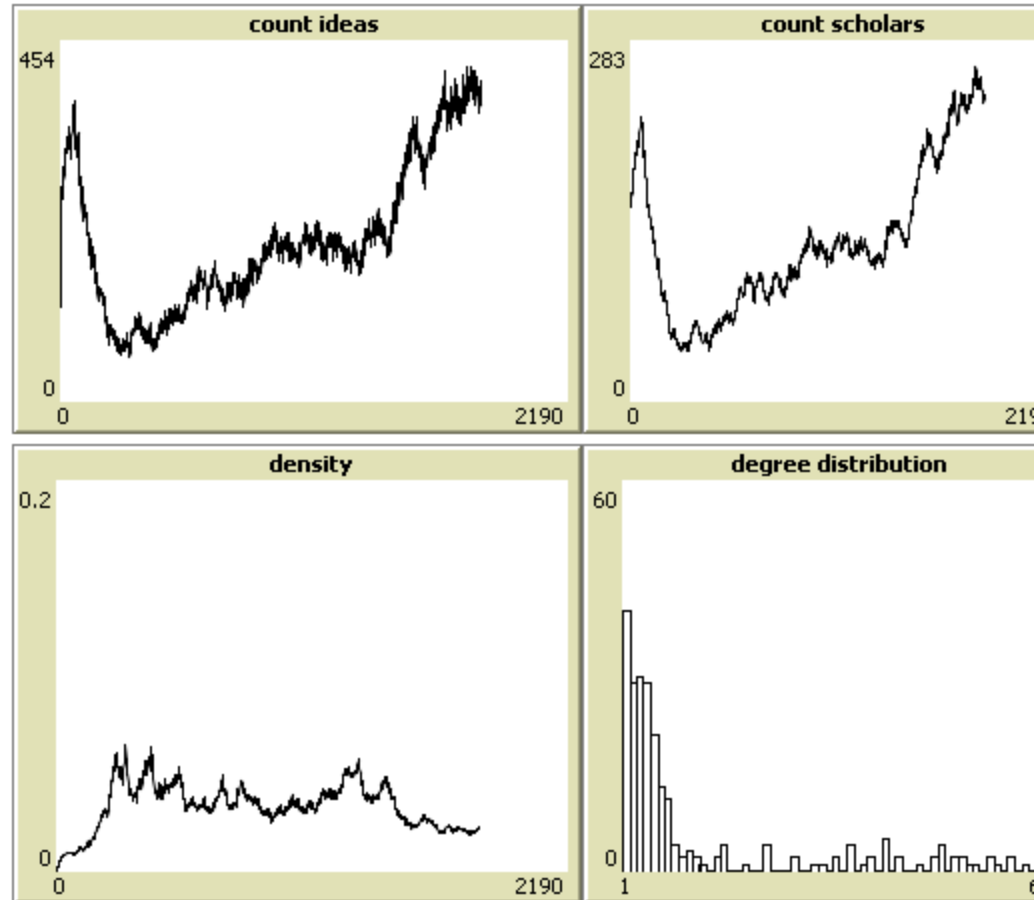
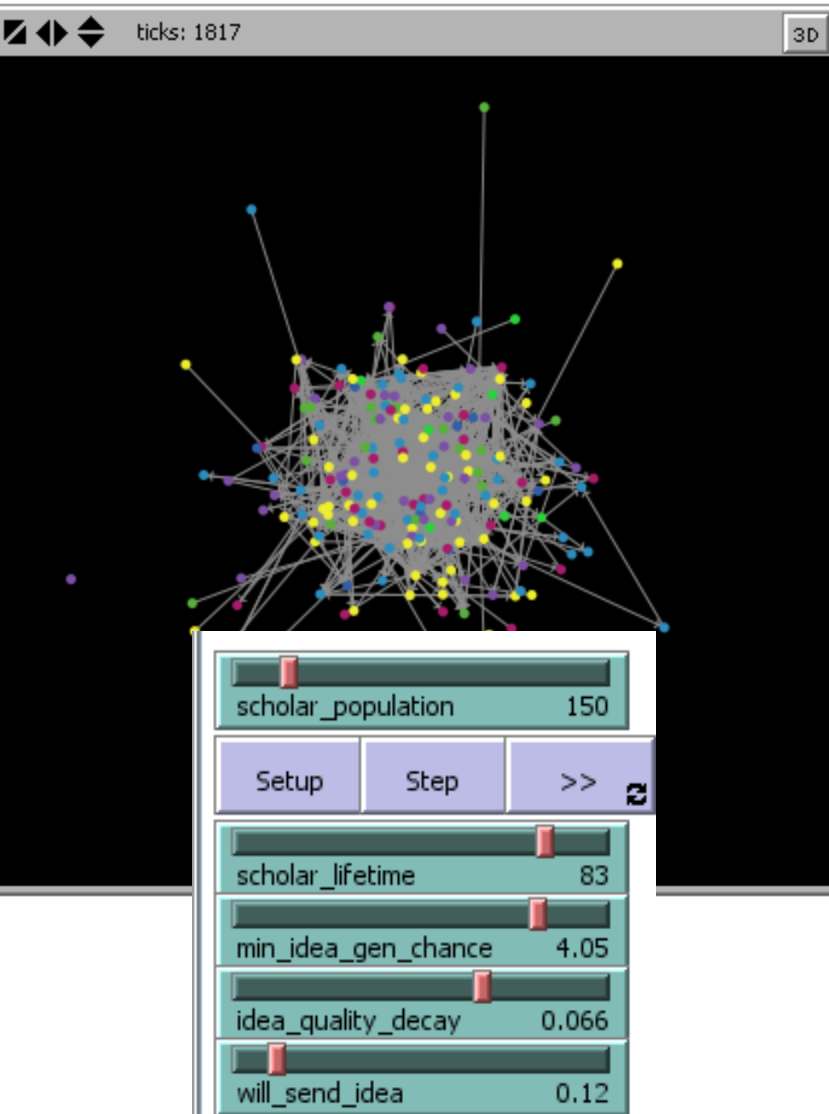
Quality of new ideas: .2 +/- .15

Quality of new ideas: .9 +/- .03





# Modeling the Republic of Letters or *Mmmmm Spaghetti Dinner with Meatballs*





## Workshop Overview

10:55-11:05 The Importance and Dangers of Visualization – Use & Theory

11:05-11:20 The Epistolarium – Networks, Topics & Tools

11:20-11:25 Computational Modeling

*11:25-11:35 Move to other room*

**11:35-11:50 Sci2 Tool Basics**

11:50-12:10 Sci2 Workflow Design: Padgett's Florentine Families - Prepare, load, analyze, and visualize family and business networks from 15th century Florence.

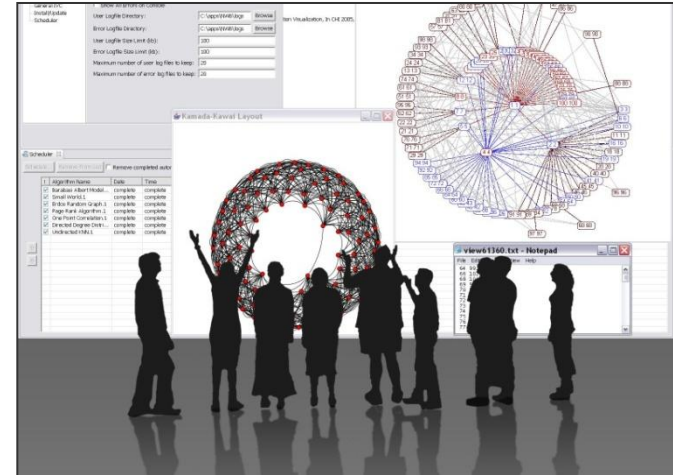
12:10-12:20 Q&A and Technical Assistance

# Sci2 Tool Basics





## Microscopes, Telescopes, and Macroscopes



Just as the **microscope** empowered our naked eyes to see cells, microbes, and viruses thereby advancing the progress of biology and medicine or the **telescope** opened our minds to the immensity of the cosmos and has prepared mankind for the conquest of space, **macroscopes** promise to help us cope with another infinite: the infinitely complex. Macroscopes give us a ‘vision of the whole’ and help us ‘synthesize’. They let us detect patterns, trends, outliers, and access details in the landscape of science. Instead of making things larger or smaller, macroscopes let us observe what is at once too great, too slow, or too complex for our eyes.



## Macroscopic Design

### Custom Tools for Different Scientific Communities

Information Visualization Cyberinfrastructure

<http://iv.slis.indiana.edu>

**Network Workbench Tool** + Community Wiki

<http://nwb.slis.indiana.edu>

**Science of Science (Sci<sup>2</sup>) Tool** and Portal

<http://sci.slis.indiana.edu>

Epidemics Cyberinfrastructure

<http://epic.slis.indiana.edu/>



**180+ Algorithm Plugins and Branded GUIs**

+

**Core Architecture**

Open Services Gateway Initiative (OSGi) Framework.

<http://orgi.org>

Cyberinfrastructure Shell (CIShell)

<http://cishell.org>



# NWB Tool Interface Components

The screenshot shows the Network Workbench Tool interface with the following components and callouts:

- Console:** Displays data operations (save, load, view, etc.) and algorithm input parameters, selection, & acknowledgements as well as error reporting. It contains a welcome message and references to seminal works.
- Scheduler:** Lists what algorithms you've used and displays algorithm progress. It includes buttons for "Remove From List", "Remove all completed", and a play button.
- Data Manager:** Keeps track of all datasets that are available for algorithmic visualization or manipulation. It features a vertical list of icons for different data types: Table, Matrix, Plot, Text, GUESS, Tree, and Network-.

The Scheduler component includes a table with the following structure:

!	Algorithm Name	Date	Time	% Complete

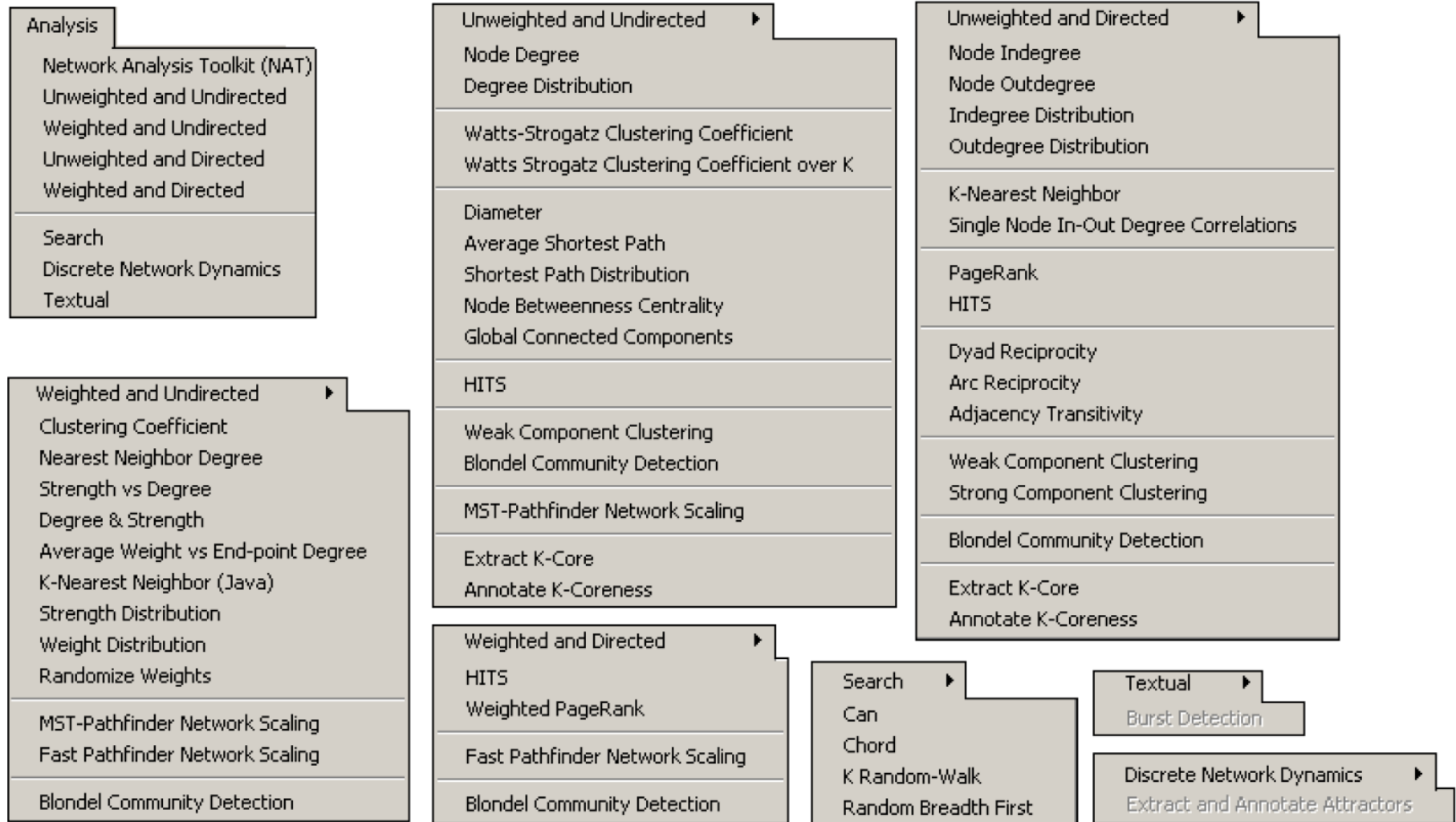
Console shows references to seminal works.  
Workflows are recorded into a log file, and soon can be re-run for easy replication.  
All algorithms are documented online; workflows are given in tutorials.

File	Preprocessing	Modeling	Visualization
Load...	Extract Top Nodes	Random Graph	GUESS
Load and Clean ISI File	Extract Nodes Above or Below Value	Watts-Strogatz Small World	GnuPlot
Read Directory Hierarchy Datasets	Remove Node Attributes	Barabási-Albert Scale-Free	DrL (VxOrd)
Save...	Delete High Degree Nodes	Can	Specified (prefuse beta)
View...	Delete Random Nodes	Chord	Circular (JUNG)
View with...	Delete Isolates	Hypergrid	Radial Tree/Graph (prefuse alpha)
Merge Node and Edge Files	Extract Top Edges	PRU	Radial Tree/Graph with Annotation (prefuse beta)
Split Graph to Node and Edge Files	Extract Edges Above or Below Value	TARL	Tree Map (prefuse beta)
Tests	Remove Edge Attributes	Discrete Network Dynamics (DND)	Tree View (prefuse beta)
Preferences	Remove Self Loops	Evolving Network (Weighted)	Balloon Graph (prefuse alpha)
Exit	Trim by Degree		Force Directed with Annotation (prefuse beta)
	Snowball Sampling (n nodes)		Kamada-Kawai (JUNG)
	Node Sampling		Fruchterman-Reingold (JUNG)
	Edge Sampling		Fruchterman-Reingold with Annotation (prefuse beta)
	Symmetrize		Spring (JUNG)
	Dichotomize		Small World (prefuse alpha)
	Multipartite Joining		Parallel Coordinates (demo)
	Normalize Text		LaNet
	Slice Table by Time		Circular Hierarchy

Börner, Katy, Sanyal, Soma and Vespignani, Alessandro (2007). **Network Science**. In Blaise Cronin (Ed.), *ARIST*, Information Today, Inc./American Society for Information Science and Technology, Medford, NJ, Volume 41, Chapter 12, pp. 537-607.

<http://ivl.slis.indiana.edu/km/pub/2007-borner-arist.pdf>

## Analysis Menu and Submenus



Börner, Katy, Sanyal, Soma and Vespignani, Alessandro (2007). **Network Science**. In Blaise Cronin (Ed.), *ARIST*, Information Today, Inc./American Society for Information Science and Technology, Medford, NJ, Volume 41, Chapter 12, pp. 537-607. <http://ivl.slis.indiana.edu/km/pub/2007-borner-arist.pdf>

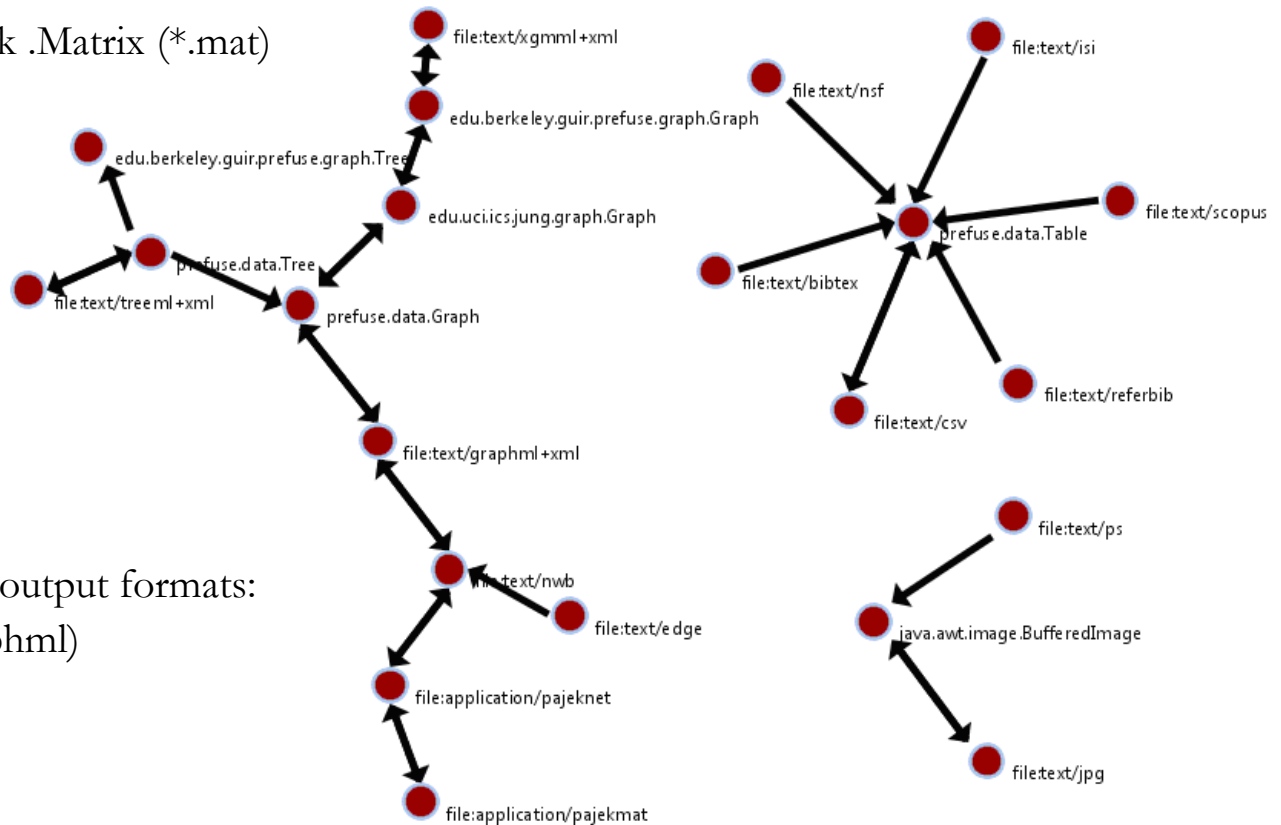
## Supported Data Formats

The NWB tool supports loading the following input file formats:

- GraphML (\*.xml or \*.graphml)
- XGMML (\*.xml)
- Pajek .NET (\*.net) & Pajek .Matrix (\*.mat)
- NWB (\*.nwb)
- TreeML (\*.xml)
- Edge list (\*.edge)
- CSV (\*.csv)
- ISI (\*.isi)
- Scopus (\*.scopus)
- NSF (\*.nsf)
- Bibtext (\*.bib)
- Endnote (\*.enw)

and the following network file output formats:

- GraphML (\*.xml or \*.graphml)
- Pajek .MAT (\*.mat)
- Pajek .NET (\*.net)
- NWB (\*.nwb)
- XGMML (\*.xml)
- CSV (\*.csv)



Formats are documented at <https://nwb.slis.indiana.edu/community/?n=DataFormats.HomePage>.



## File-types

➤ **Excel**

➤ **Database**

➤ **Text**

➤ **CSV**



## Network Formats

	Newton	Oldenburg	Flamsteed
Newton	0	13	38
Oldenburg	24	0	45
Flamsteed	62	7	0

- **Matrix**
- **Adjacency List**
- **Node & Edge List**

Newton	Oldenburg	13
Newton	Flamsteed	38
Oldenburg	Newton	24
Oldenburg	Flamsteed	45
Flamsteed	Newton	62
Flamsteed	Oldenburg	7

Nodes		
1	Newton	
2	Oldenburg	
3	Flamsteed	
Edges		
1	2	13
1	3	38
2	1	24
2	3	45
3	1	62
3	2	7





## NWB Format

### \*Nodes

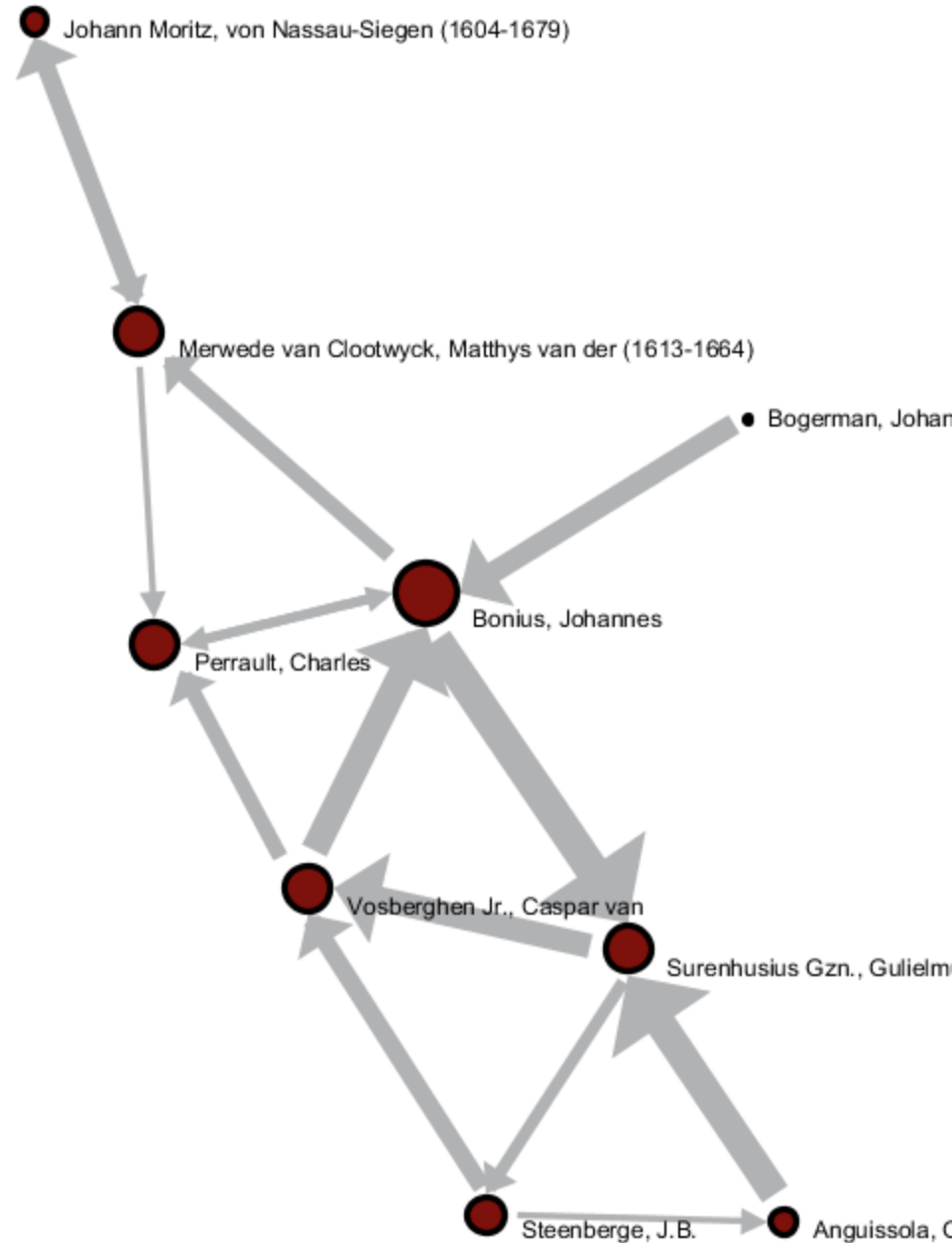
**id\*int label\*string totaldegree\*int**

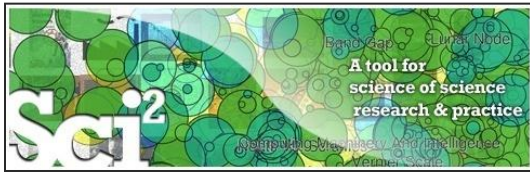
```
16 "Merwede van Clootwyck, Matthys van der (1613-1664)" 1
36 "Perrault, Charles" 1
48 "Bonius, Johannes" 1
67 "Surenhusius Gzn., Gulielmus" 1
99 "Anguissola, Giacomo" 1
126 "Johann Moritz, von Nassau-Siegen (1604-1679)" 6
131 "Steenberge, J.B." 1
133 "Vosberghen Jr., Caspar van" 1
151 "Bogerman, Johannes (1576-1637)" 25
```

### \*DirectedEdges

**source\*int target\*int weight\*float eyear\*int syear\*int**

```
16 36 1 1640 1650
16 126 5 1641 1649
36 48 2 1630 1633
48 16 4 1637 1644
48 67 10 1645 1648
48 36 2 1632 1638
67 133 7 1644 1648
67 131 3 1642 1643
99 67 9 1640 1645
126 16 3 1641 1646
131 133 5 1630 1638
131 99 1 1637 1639
133 36 4 1645 1648
133 48 8 1632 1636
151 48 6 1644 1647
```





## Science of Science (Sci2) Tool

<http://sci.slis.indiana.edu>

- Explicitly designed for SoS research and practice, well documented, easy to use.
- Empowers many to run common studies while making it easy for exports to perform novel research.
- Advanced algorithms, effective visualizations, and many (standard) workflows.
- Supports micro-level documentation and replication of studies.
- Is open source—anybody can review and extend the code, or use it for commercial purposes.

nature

OPINION

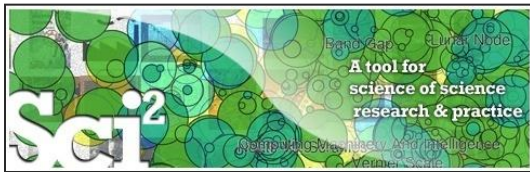
### SUMMARY

- Existing metrics have known flaws
- A reliable, open, joined-up data infrastructure is needed
- Data should be collected on the full range of scientists' work
- Social scientists and economists should be involved

Vol 464|25 March 2010

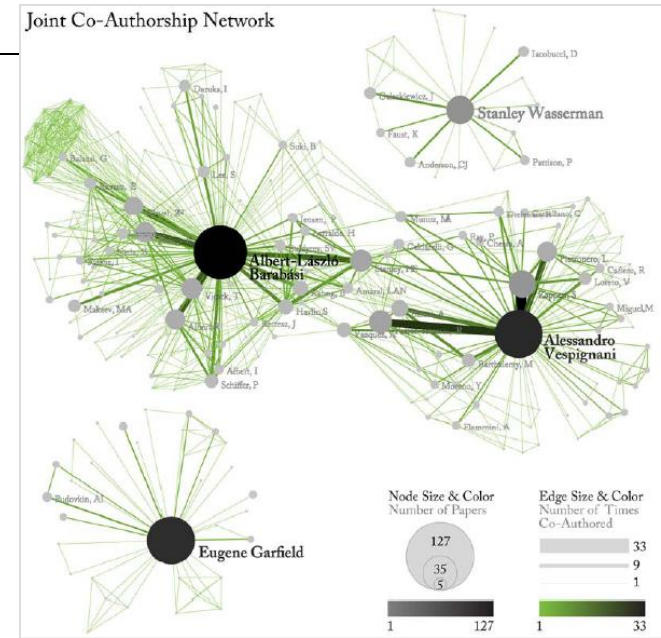
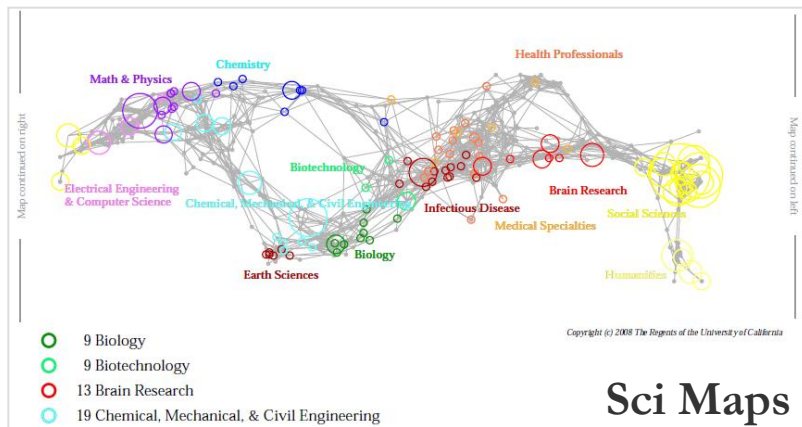
# Let's make science metrics more scientific

To capture the essence of good science, stakeholders must combine forces to create an open, sound and consistent system for measuring all the activities that make up academic productivity, says **Julia Lane**.

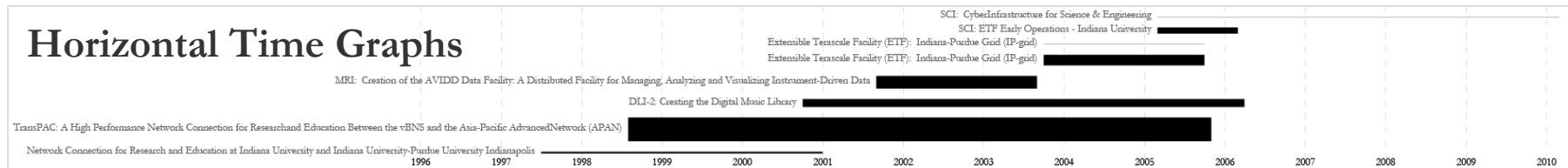


# Sci<sup>2</sup> Tool – “Open Code for S&T Assessment”

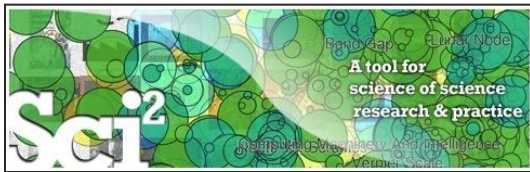
OSGi/CIShell powered tool with NWB plugins and many new scientometrics and visualizations plugins.



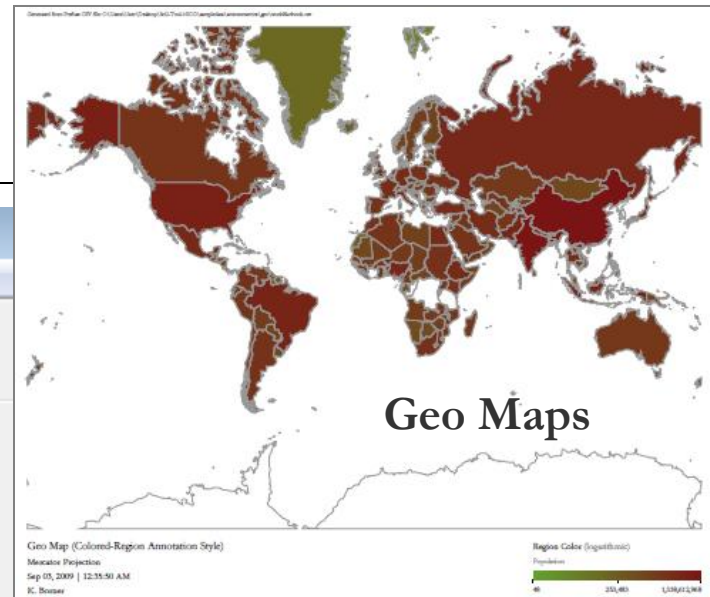
## Horizontal Time Graphs



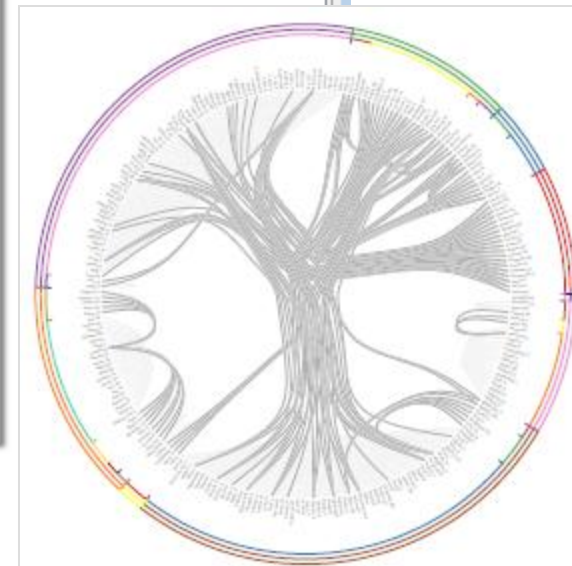
Börner, Katy, Huang, Weixia (Bonnie), Linnemeier, Micah, Dubon, Russell Jackson, Phillips, Patrick, Ma, Nianli, Zoss, Angela, Guo, Hanning & Price, Mark. (2009). *Rete-Netzwerk-Red: Analyzing and Visualizing Scholarly Networks Using the Scholarly Database and the Network Workbench Tool*. *Proceedings of ISSI 2009: 12th International Conference on Scientometrics and Informetrics, Rio de Janeiro, Brazil, July 14-17 . Vol. 2, pp. 619-630.*



# Sci² Tool



Geo Maps



Circular Hierarchy

Sci² Tool

File Preprocessing Modeling Analysis Visualization Scientometrics Help

Console

Welcome to the Science of Science Tool (Sci²). The development of this tool is supported in Network Science center and the School of Li Indiana University, the National Science Foundation and IIS-0715303, and the James S. McDonnell Cyberinfrastructure portal (<http://sci.slis.indiana.edu>)

The primary investigators are Katy Börner, In SciTech Strategies Inc. The Sci² tool was developed by J. Duhon, Patrick A. Phillips, Chintan Tank, a Cyberinfrastructure Shell (<http://cishell.org>) for Network Science Center (<http://cns.slis.indiana.edu>). Many algorithm plugins were derived from the Network Science Center (<http://nwb.slis.indiana.edu>).

Please cite as follows:  
 Sci² Team. (2009). Science of Science Tool. In Sci² Strategies Inc., <http://sci.slis.indiana.edu>.  
 .....

Scheduler

Remove From List  Remove completed

!	Algorithm Name	Date	Time	% Comp
<input checked="" type="checkbox"/>	Extract Co-Author Netw...	09/03/2009	00:15:20 AM	<div style="width: 100%; height: 10px; background-color: green;"></div>
<input checked="" type="checkbox"/>	Load and Clean ISI File	09/03/2009	00:15:05 AM	<div style="width: 100%; height: 10px; background-color: green;"></div>


Visualization menu items:

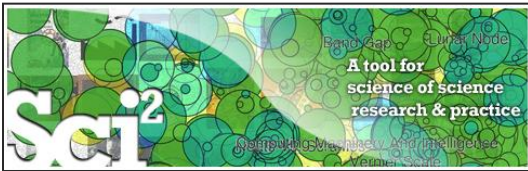
- GUESS
- GnuPlot
- Radial Tree/Graph (prefuse alpha)
- Radial Tree/Graph with Annotation (prefuse beta)
- Tree View (prefuse beta)
- Tree Map (prefuse beta)
- Force Directed with Annotation (prefuse beta)
- Fruchterman-Reingold with Annotation (prefuse beta)
- DrL (VxOrd)
- Specified (prefuse beta)
- Horizontal Line Graph
- Circular Hierarchy
- Geo Map (circle annotations)
- Geo Map (region coloring annotations)
- Image Viewer
- RefMapper

# Contents

Contents.....		
<b>1 Introduction .....</b>		
<b>2 Getting Started .....</b>		
2.1 Download, Install, Uninstall.....		
2.2 User Interface .....		
2.2.1 Menus.....		
2.2.2 Console.....		
2.2.3 Data Manager.....		
2.2.4 Scheduler.....		
2.3 Data Formats .....		
2.4 Saving Visualizations for Publication .....		
2.5 Sample Datasets .....		
<b>3 Algorithm and Tool Plugins .....</b>		
3.1 Sci <sup>2</sup> Tool Plugins.....		
3.2 Load, View, and Save Data.....		
3.3 Memory Allocation .....		
3.3.1 Windows and Linux.....		
3.3.2 Mac.....		
3.4 Memory Limits.....		
<b>4 Workflow Design .....</b>		
4.1 Overview.....		
4.2 Data Acquisition and Preparation.....		
4.2.1 Datasets: Publications .....		
4.2.2 Datasets: Funding.....		
4.2.3 Datasets: Scholarly Database .....		
4.3 Database Loading and Manipulation.....		
4.4 Summaries and Table Extractions.....		
4.5 Statistical Analysis/Profiling.....		
4.6 Temporal Analysis (When).....		
4.6.1 Burst Detection.....		
4.6.2 Slice Table by Time .....		
4.7 Geospatial Analysis (Where).....		
4.8 Topical Analysis (What).....		
4.8.1 Word Co-Occurrence Network .....		
4.9 Network Analysis (With Whom?).....	35	
4.9.1 Network Extraction.....	35	
4.9.2 Compute Basic Network Characteristics.....		
4.9.3 Network Analysis.....		
4.9.4 Network Visualization.....		
4.10 Modeling (Why?).....		
4.10.1 Random Graph Model.....		
4.10.2 Watts-Strogatz Small World.....		
4.10.3 Barabási-Albert Scale Free Model.....		
<b>5 Sample Workflows .....</b>		
5.1 Individual Level Studies - Micro.....		
5.1.1 Mapping Collaboration, Publication and Funding Profiles of One Researcher.....		
5.1.2 Time Slicing of Co-Authorship Networks (ISI Data) .....		
5.1.3 Funding Profiles of Three Researchers at Indiana University (NSF Data).....		
5.1.4 Studying Four Major NetSci Researchers (ISI Data).....		
5.1.5 Studying Four Major NetSci Researchers (ISI Data) using Database .....		
5.2 Institution Level Studies - Meso .....		
5.2.1 Funding Profiles of Three Universities (NSF Data).....		
5.2.2 Funding Profiles of Three Universities (NSF Data) Using Database.....		
5.2.3 Mapping CTSA Centers (NIH RePORTER Data).....		
5.2.4 Biomedical Funding Profile of NSF (NSF Data).....		
5.2.5 Mapping Scientometrics (ISI Data) .....		
5.2.6 Burst Detection in Scientometrics (ISI Data).....		
5.2.7 Mapping the Field of RNAi Research (SDB Data).....		
5.3 Global Level Studies - Macro.....		
5.3.1 Geo USPTO (SDB Data).....		
<b>6 Sample Science Studies &amp; Online Services.....</b>		
6.1 Science Dynamics.....		
6.1.1 Mapping Topics and Topic Bursts in PNAS (2004).....		
6.2 Local Impact-Output / ROI Studies .....		
6.2.1 Indicator-Assisted Evaluation and Funding of Research: Visualizing the Influence of Grants on the Number and Citation Counts of Research Papers (2003).....	87	
6.2.2 Mapping Transdisciplinary Tobacco Use Research Centers Publications (forthcoming).....	88	
6.3 Local and Global Science Studies.....	89	
6.3.1 Mapping the Evolution of Co-Authorship Networks (2004).....	89	
6.3.2 Studying the Emerging Global Brain: Analyzing and Visualizing the Impact of Co-Authorship Teams (2005).....	90	
6.3.4 Mapping the Diffusion of Information Among Major U.S. Research Institutions (2006).....	92	
6.3.5 Research Collaborations by the Chinese Academy of Sciences (2009) .....	93	
6.3.6 Mapping the Structure and Evolution of Chemistry Research (2009).....	94	
6.3.7 Science Map Applications: Identifying Core Competency (2007).....	95	
6.4 Modeling Science.....	96	
6.4.1 113 Years of Physical Review: Using Flow Maps to Show Temporal and Topical Citation (2008).....	96	
6.4.2 The Simultaneous Evolution of Author and Paper Networks (2004).....	97	
6.5 Accuracy Studies.....	98	
6.5.1 Mapping the Backbone of Science (2005).....	98	
6.5.2 Toward a Consensus Map of Science (2009).....	99	
6.6 Databases and Tools.....	100	
6.6.1 The Scholarly Database and Its Utility for Scientometrics Research (2009).....	100	
6.6.2 Reference Mapper.....	101	
6.6.3 Rete-Netzwerk-Red: Analyzing and Visualizing Scholarly Networks Using the Scholarly Database and the Network Workbench Tool (2009).....	102	
6.7 Interactive Online Services .....	103	
6.7.1 The NIH Visual Browser: An Interactive Visualization of Biomedical Research (2009).....	103	
6.7.2 Interactive World and Science Map of S&T Jobs (2010).....	104	
<b>7 Extending the Sci<sup>2</sup> Tool .....</b>	<b>105</b>	
7.1 CShell Basics.....	105	
7.2 Read New Data .....	105	
7.3 Creating and Sharing New Algorithm Plugins.....	105	
7.4 Tools That Use OSGI and/or CShell.....	106	
<b>8 Relevant Datasets and Tools.....</b>	<b>107</b>	
8.1 Datasets.....	107	
8.2 Network Analysis Tools.....	108	
<b>9 References .....</b>	<b>111</b>	

See Sci2 Manual

 2010-03-sci2-manual.pdf

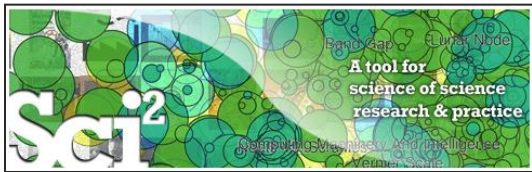


## Studying Four Major NetSci Researchers (ISI Data)

FourNetSciResearchers.isi	
<b>Time frame:</b>	1955-2007
<b>Region(s):</b>	Miscellaneous
<b>Topical Area(s):</b>	Network Science
<b>Analysis Type(s):</b>	Paper Citation Network, Co-Author Network, Bibliographic Coupling Network, Document Co-Citation Network, Word Co-Occurrence Network

Thomson Reuter's Web of Knowledge (WoS) is a leading citation database cataloging over 10,000 journals and over 120,000 conferences. Access it via the "Web of Science" tab at <http://www.isiknowledge.com> (**note:** access to this database requires a paid subscription). Along with Scopus, WoS provides some of the most comprehensive datasets for scientometric analysis.

To find all publications by an author, search for the last name and the first initial followed by an asterisk in the author field.



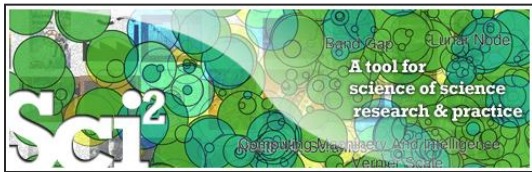
# Data Acquisition from Web of Science

Download all papers by

- Eugene Garfield
- Stanley Wasserman
- Alessandro Vespignani
- Albert-László Barabási

from

- Science Citation Index Expanded (SCI-EXPANDED)  
--1955-present
- Social Sciences Citation Index (SSCI)--1956-present
- Arts & Humanities Citation Index (A&HCI)--1975-present

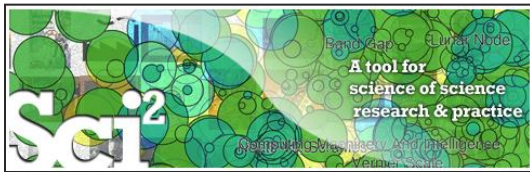


## Comparison of Counts

No books and other non-WoS publications are covered.

	<b>Age</b>	<b>Total # Cites</b>	<b>Total # Papers</b>	<b>H-Index</b>	
Eugene Garfield	82	1,525	672	31	
Stanley Wasserman		122	35	17	
Alessandro Vespignani	42	451	101	33	
Albert-László Barabási	40	2,218	126	47	<i>(Dec 2007)</i>
	41	16,920	159	52	<i>(Dec 2008)</i>





## Extract Co-Author Network

Load *\*yoursci2directory\*/sampledata/scientometrics/isi/FourNetSciResearchers.isi*  
using 'File > Load' and parameters

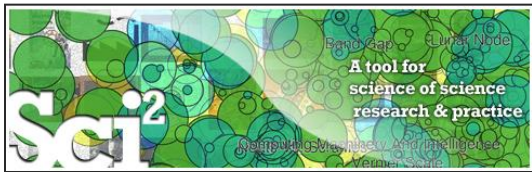
The screenshot shows the Sci2 Tool interface. A 'Load' dialog box is open, asking to select a format for the file. The 'Load as...' dropdown menu is set to 'ISI scholarly format'. The 'Data Manager' window shows the loaded data as '361 Unique ISI Records'. The 'Console' window displays the following text:

```
Loaded 361 records.  
Removed 0 duplicate records.  
Author names have been normalized.  
  
361 records with unique ISI IDs are available via Data Manager.  
  
Wrote log to  
C:\Users\User\AppData\Local\Temp\isiduplicateremoverlog2534733993422022  
81.txt
```

The 'Scheduler' window shows a table of tasks:

!	Algorithm Name	Date	Time	%
<input checked="" type="checkbox"/>	Load and Clean ISI File	08/15/2010	07:29:43 PM	100%
<input checked="" type="checkbox"/>	Load and Clean ISI File	08/15/2010	07:12:49 PM	100%

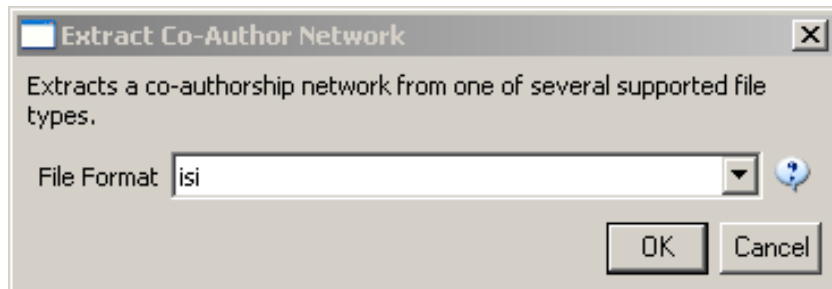
And file with 361 records  
appears in Data Manager.



## Extract Co-Author Network

(see section 5.1.4.2 on correcting duplicate/ misspelled author names)

To extract the co-author network, select the *'361 Unique ISI Records'* table and run *'Data Preparation > Extract Co-Author Network'* using isi file format:



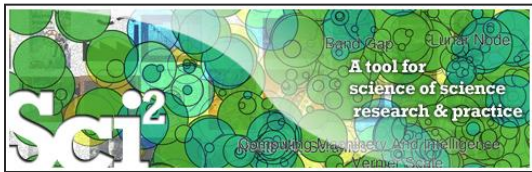
The result is an undirected but weighted network of co-authors in the Data Manager.

Run *'Analysis > Network > Network Analysis Toolkit (NAT)'* to calculate basic properties: the network has 247 nodes and 891 edges.

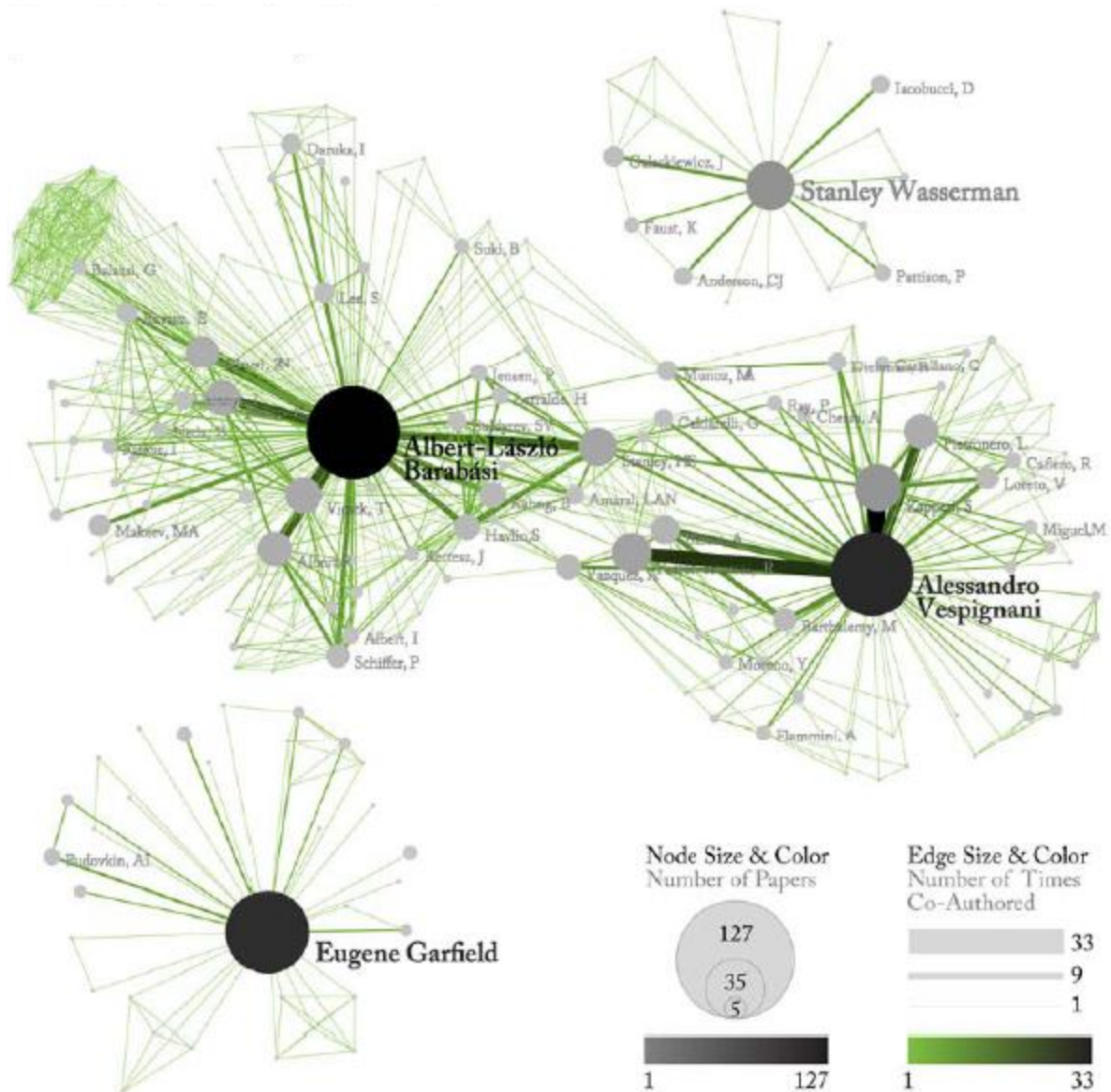
Use *'Analysis > Network > Unweighted and Undirected > Node Degree'* to calculate the number of neighbors for each node.

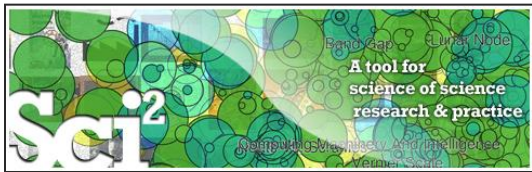
To view the complete network, select the *'Extracted Co-Authorship Network'* and run *'Visualization > Networks > GUESS'*.

Network is loaded with random layout. In GUESS, run *'Layout > GEM'* and *'Layout > Bin Pack'* to improve layout.



# Co-Author Network of all Four NetsSci Researchers





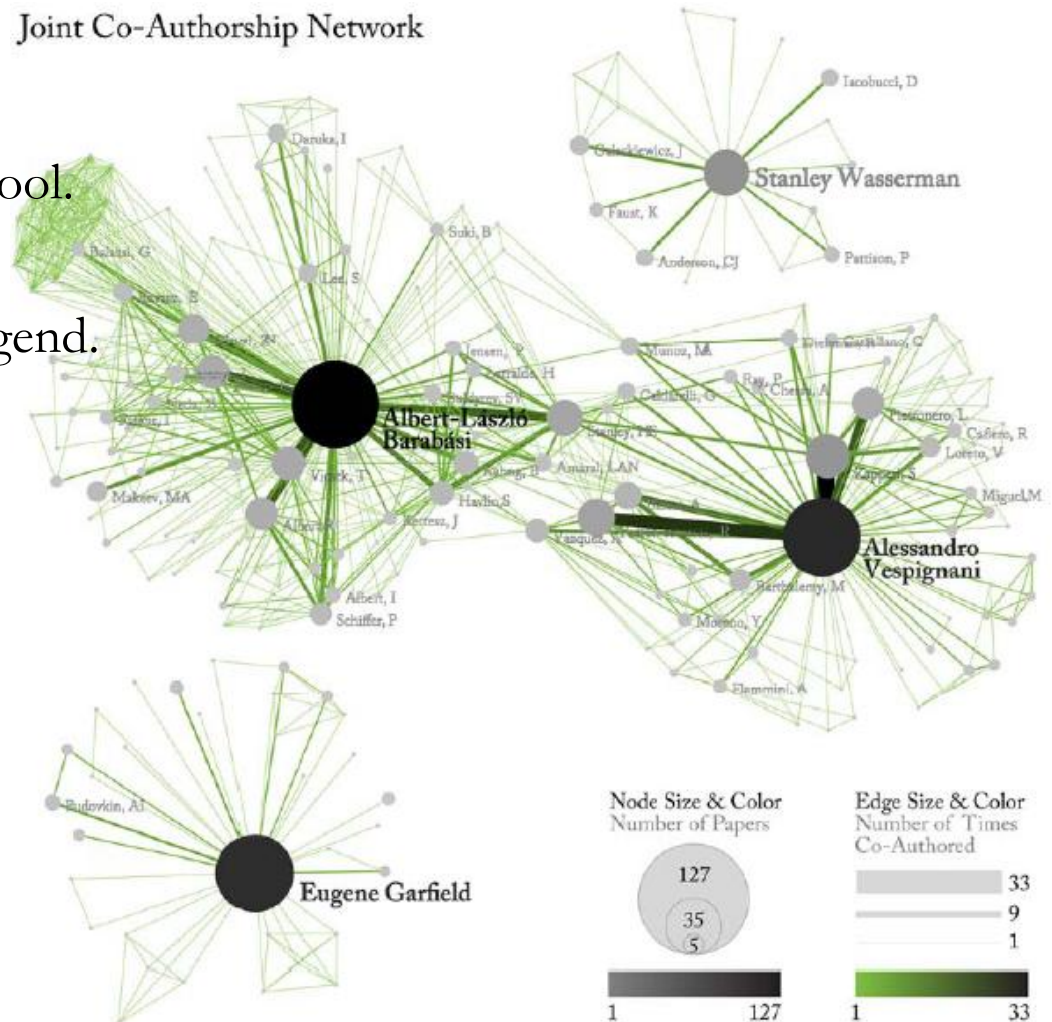
# Co-Author Network of all Four NetsSci Researchers

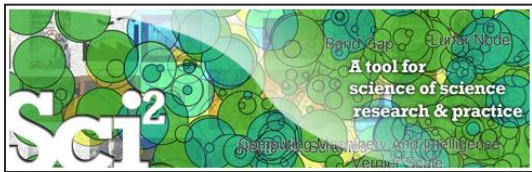
Use the GUESS Graph Modifier to change color and size coding.

Joint Co-Authorship Network

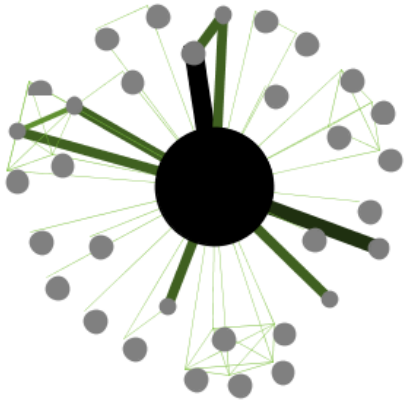
Calculate node degrees in Sci2 Tool.

Use a graphic program to add legend.

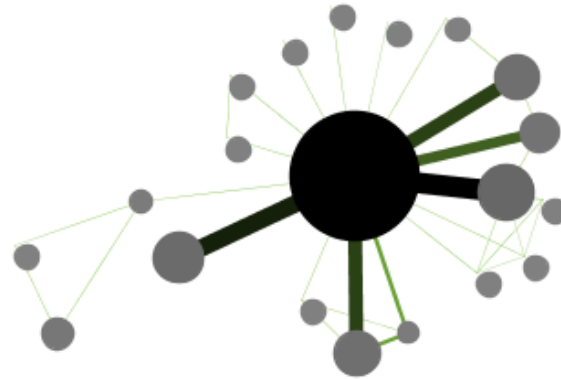




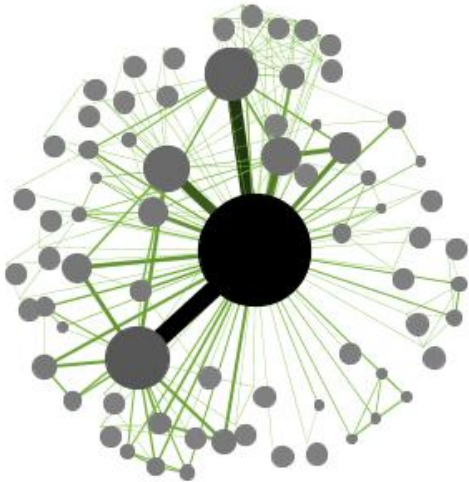
## Individual Co-Author Networks (Read/map 4 files separately)



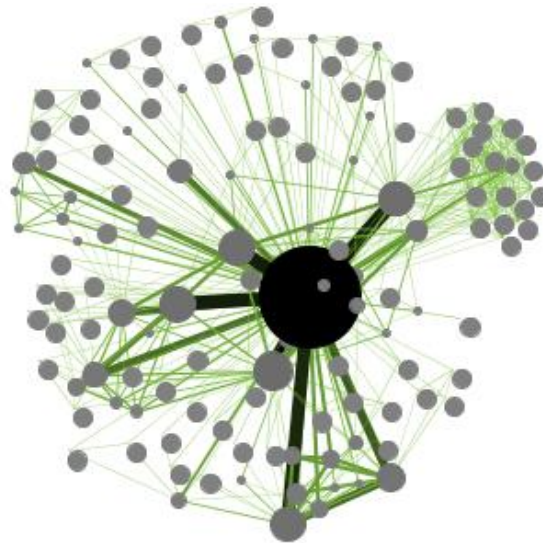
Eugene Garfield



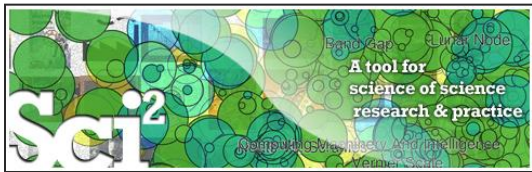
Stanley Wasserman



Alessandro Vespignani



Albert-László Barabási



## Network Visualization: Node Layout

Load and Clean ISI File was selected.  
Loaded 361 records.  
Removed 0 duplicate records.  
Author names have been normalized.  
361 records with unique ISI IDs are available  
via Data Manager.

.....

Extract Co-Author Network was selected.

Input Parameters:

File Format: isi

.....

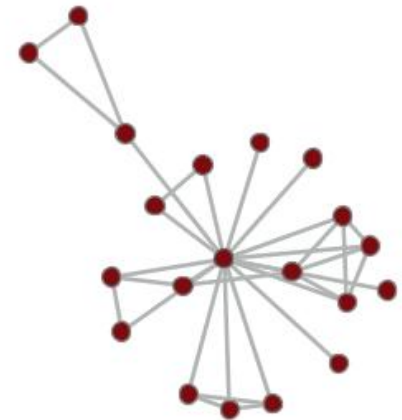
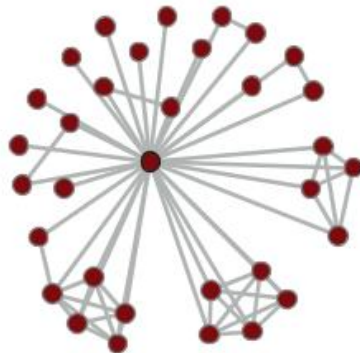
Network Analysis Toolkit (NAT) was selected.

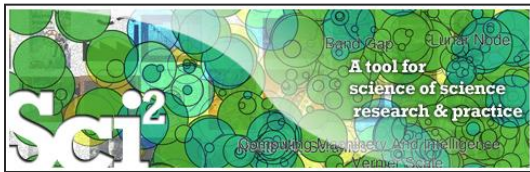
Nodes: 247

Edges: 891

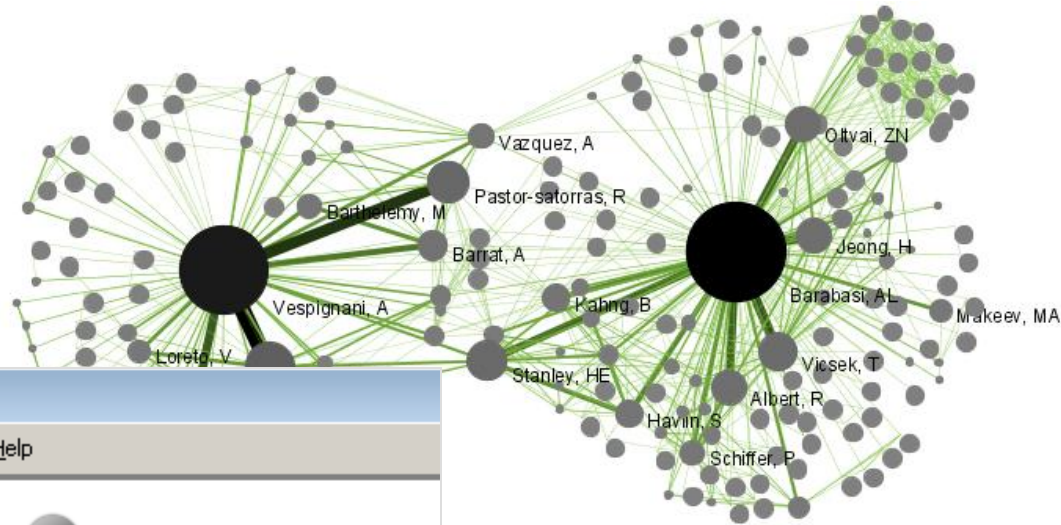
.....

GUESS was selected.





# Network Visualization: Color/Size Coding by Data Attribute Values

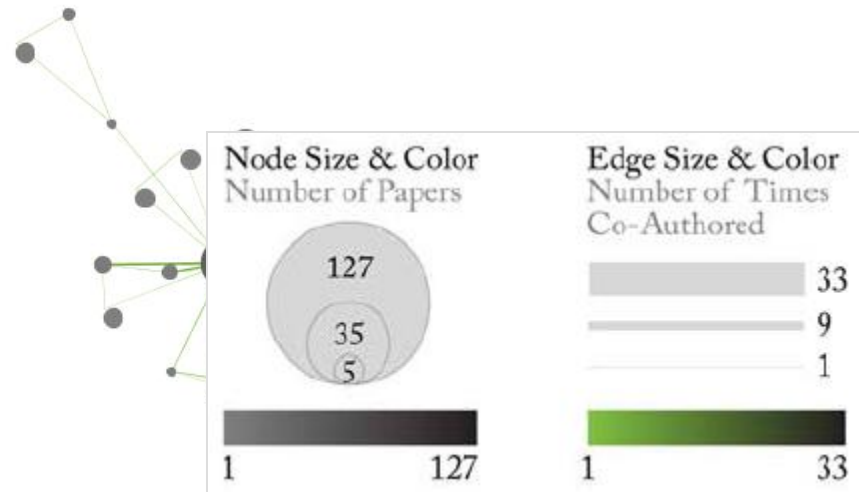


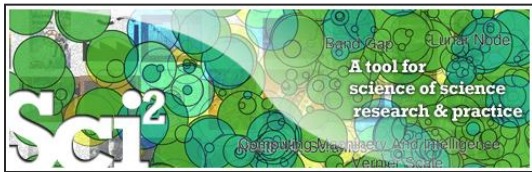
Visualization - GUESS

File Edit Layout Script View Help

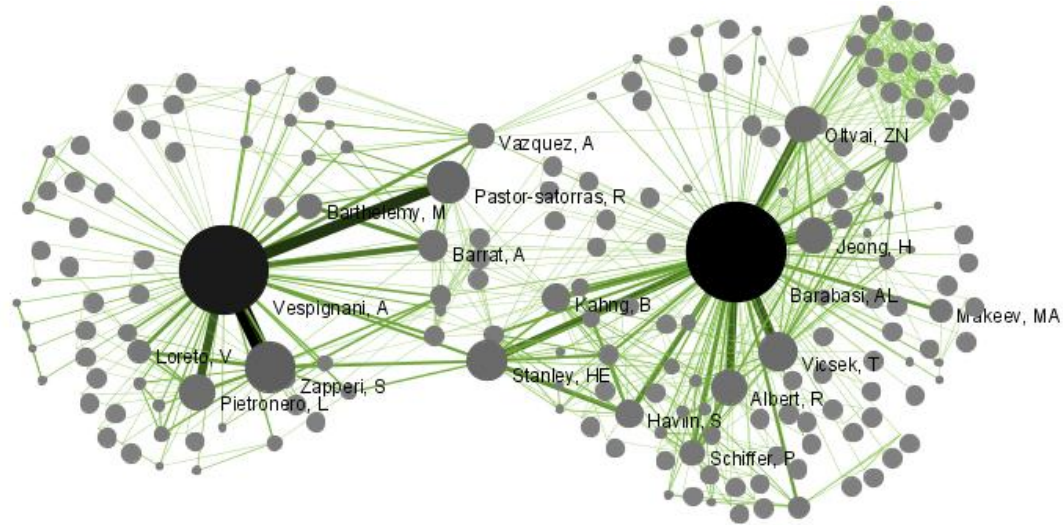
**Vespignani, A**

Field	Value
color	125,12,17,255
fixed	false
height	10.0
image	
label	Vespignani, A
labelcolor	0,0,0,255
labelsize	12
labelvisible	false
name	n161
numberofworks	101
originallabel	Vespignani, A
strokecolor	black
style	2
timescited	3811
visible	true
width	10.0
x	586.75
y	107.25





# Network Visualization: Giant Component

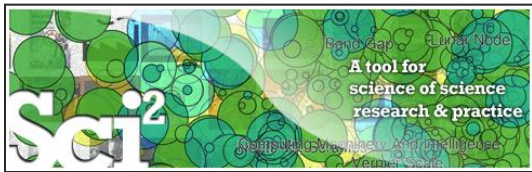


.....  
Weak Component Clustering was selected.  
Implementer(s): Russell Duhon  
Integrator(s): Russell Duhon

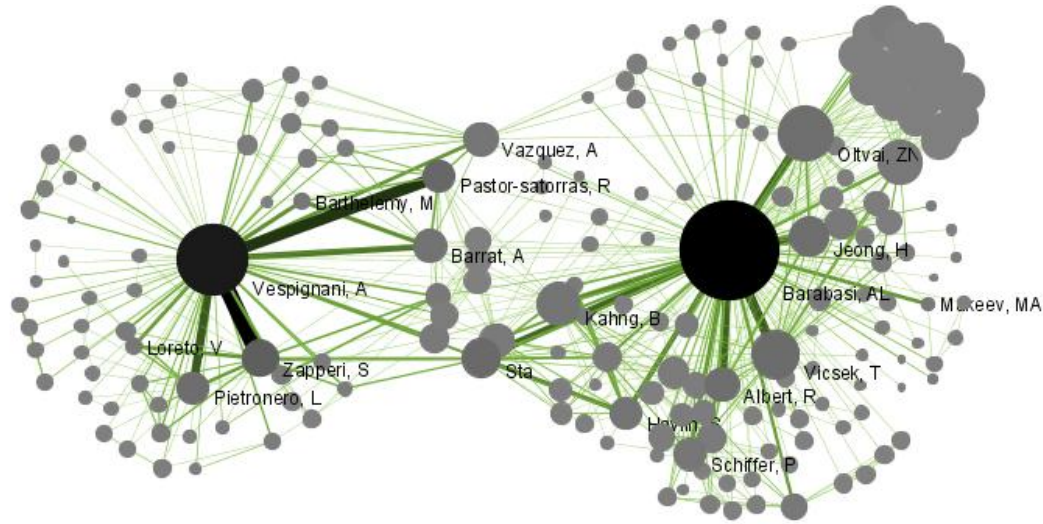
Input Parameters:  
Number of top clusters: 10  
3 clusters found, generating graphs for the top 3 clusters.

.....





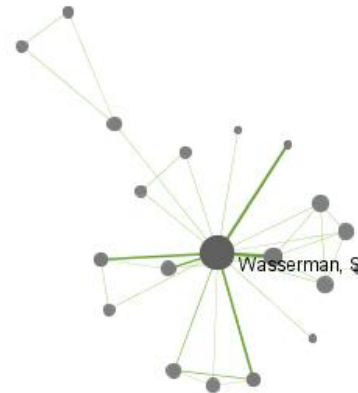
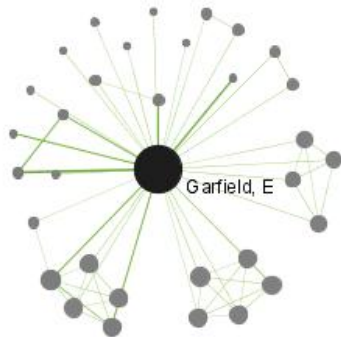
# Network Visualization: Color/Size Coding by Degree

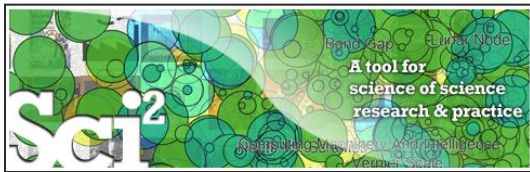


.....  
Node Degree was selected.

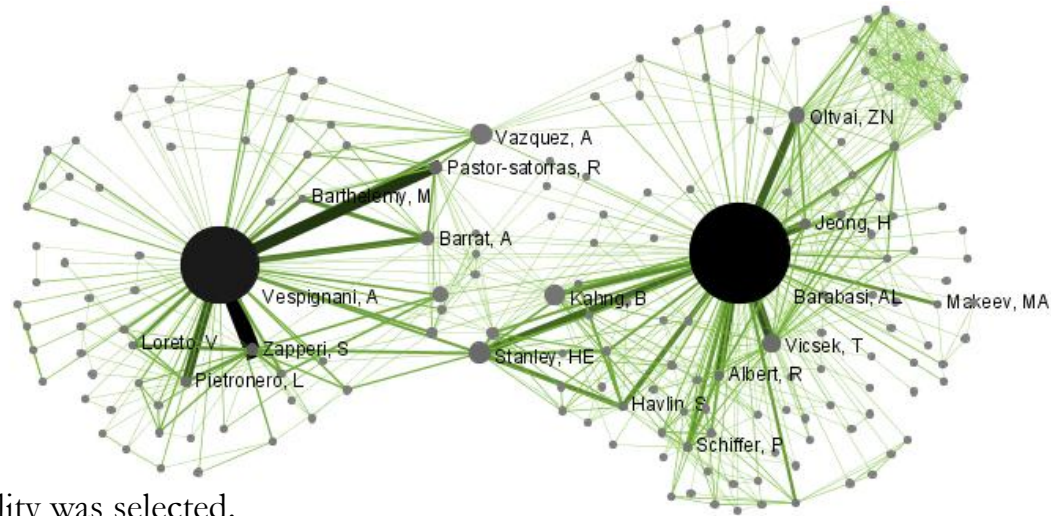
Documentation:  
<https://nwb.slis.indiana.edu/community/?n=AnalyzeData.NoDeDegree>

.....



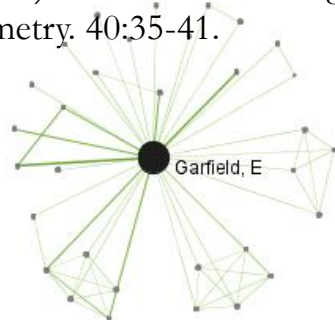


# Network Visualization: Color/Size Coding by Betweenness Centrality



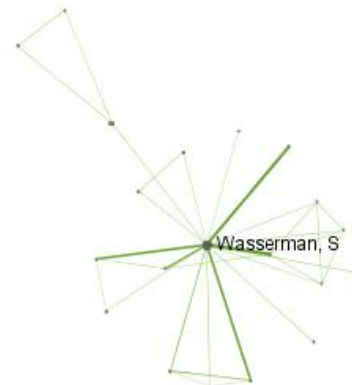
.....  
 Node Betweenness Centrality was selected.  
 Author(s): L. C. Freeman  
 Implementer(s): Santo Fortunato  
 Integrator(s): Santo Fortunato, Weixia Huang  
 Reference: Freeman, L. C. (1977). A set of measuring centrality based on betweenness. *Sociometry*. 40:35-41.

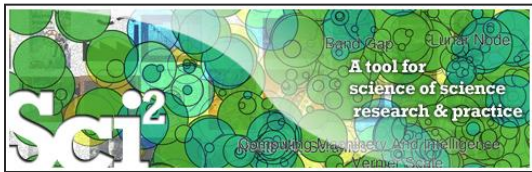
Input Parameters:  
 Number of bins: 10



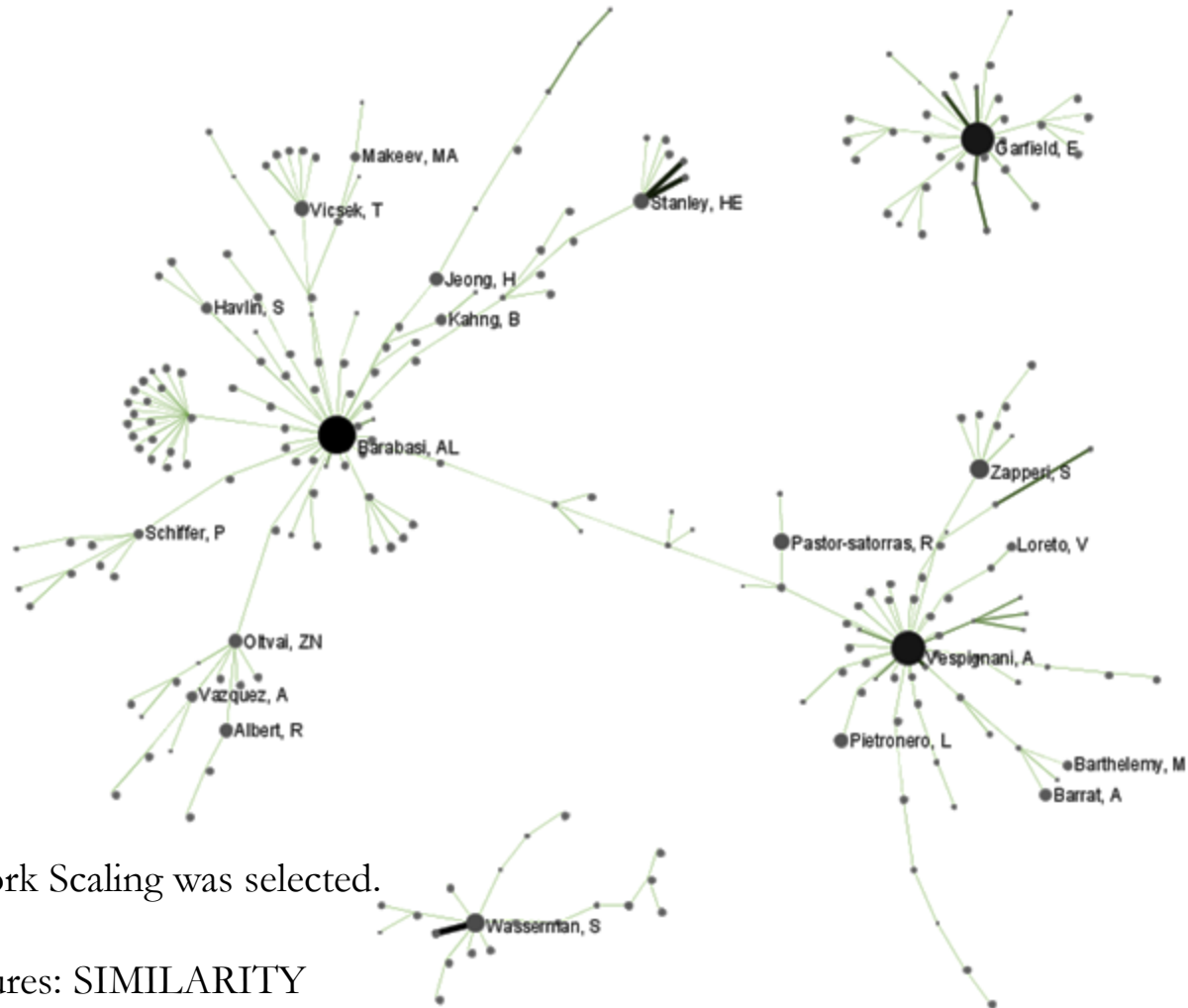
umber of bins: 10

.....





# Network Visualization: Reduced Network After Pathfinder Network Scaling



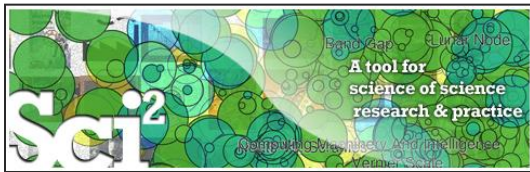
.....  
MST-Pathfinder Network Scaling was selected.

Input Parameters:

Weight Attribute measures: SIMILARITY

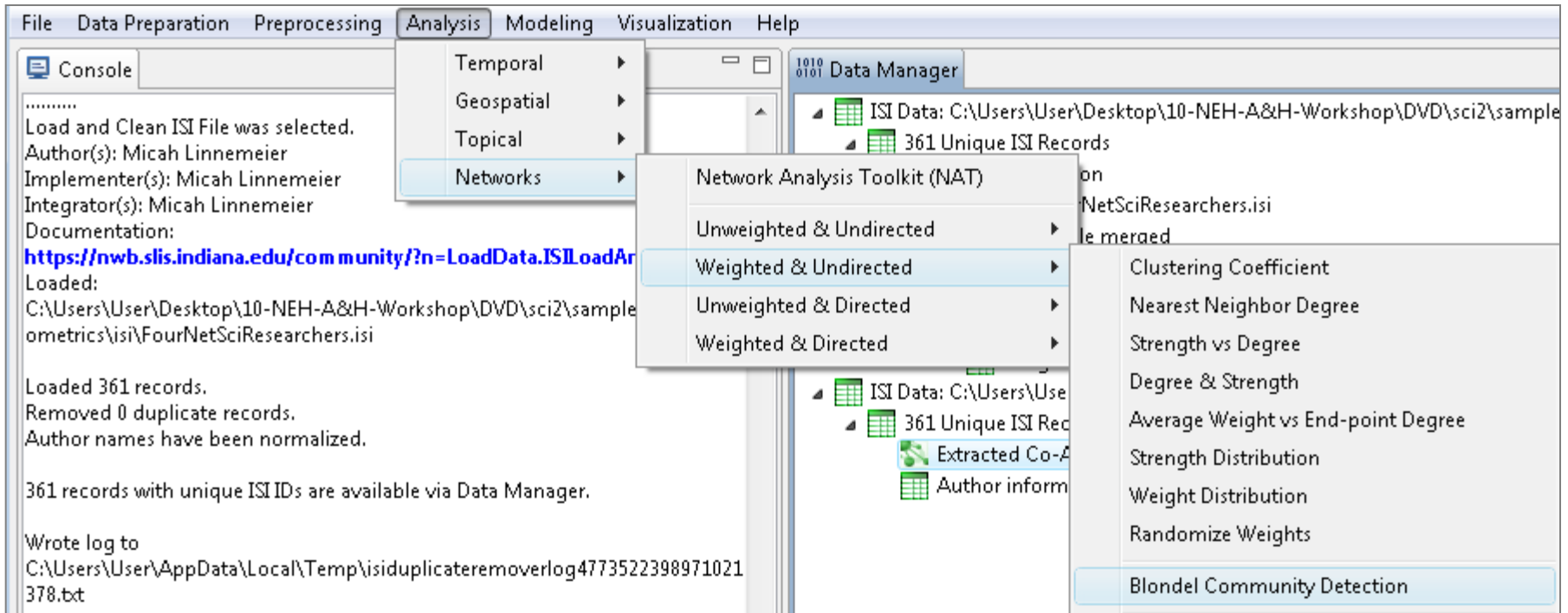
Edge Weight Attribute: weight

.....

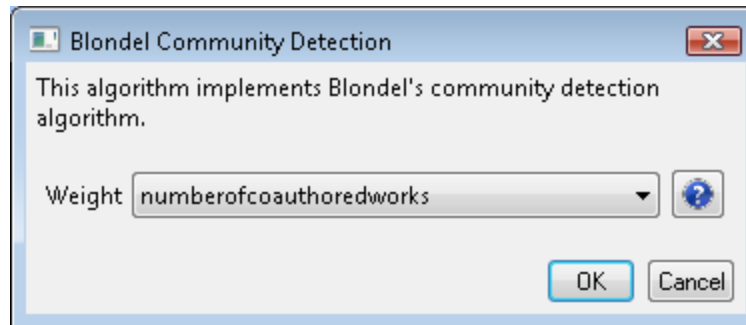


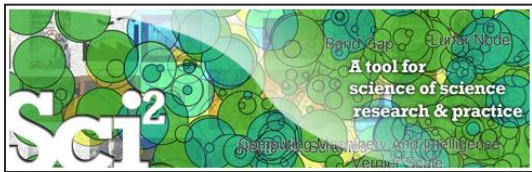
# Network Visualization: Circular Hierarchy Visualization

Select Co-Author Network and run Blondel Community detection:



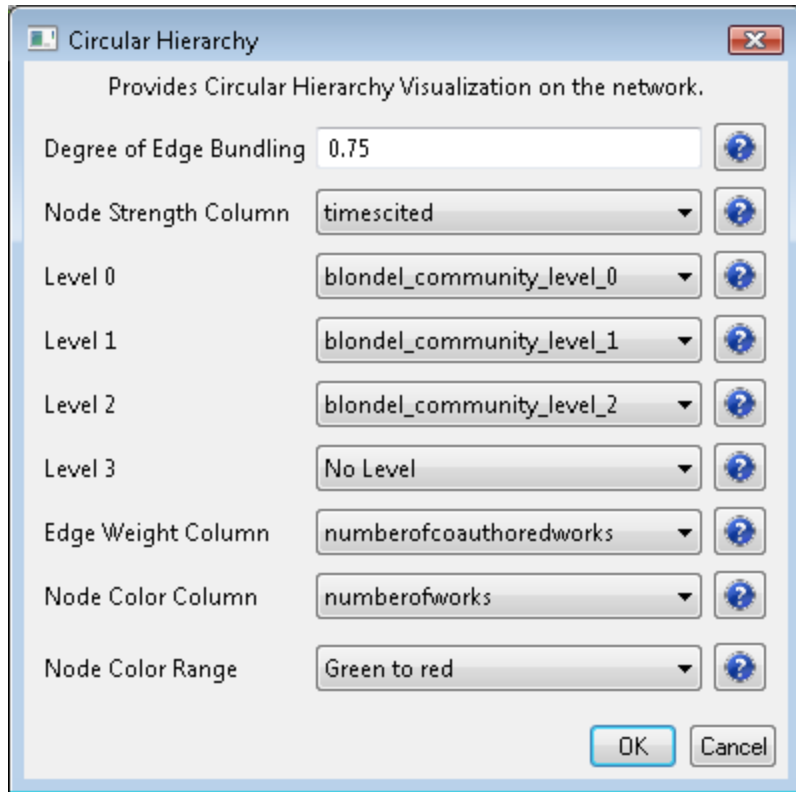
With parameter values

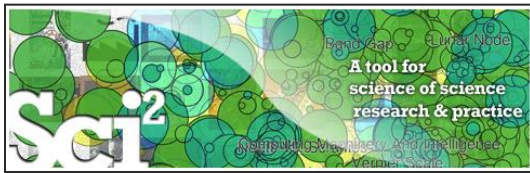




## Network Visualization: Circular Hierarchy Visualization

Visualize resulting file using *Visualization > Networks > Circular Hierarchy* with parameter values





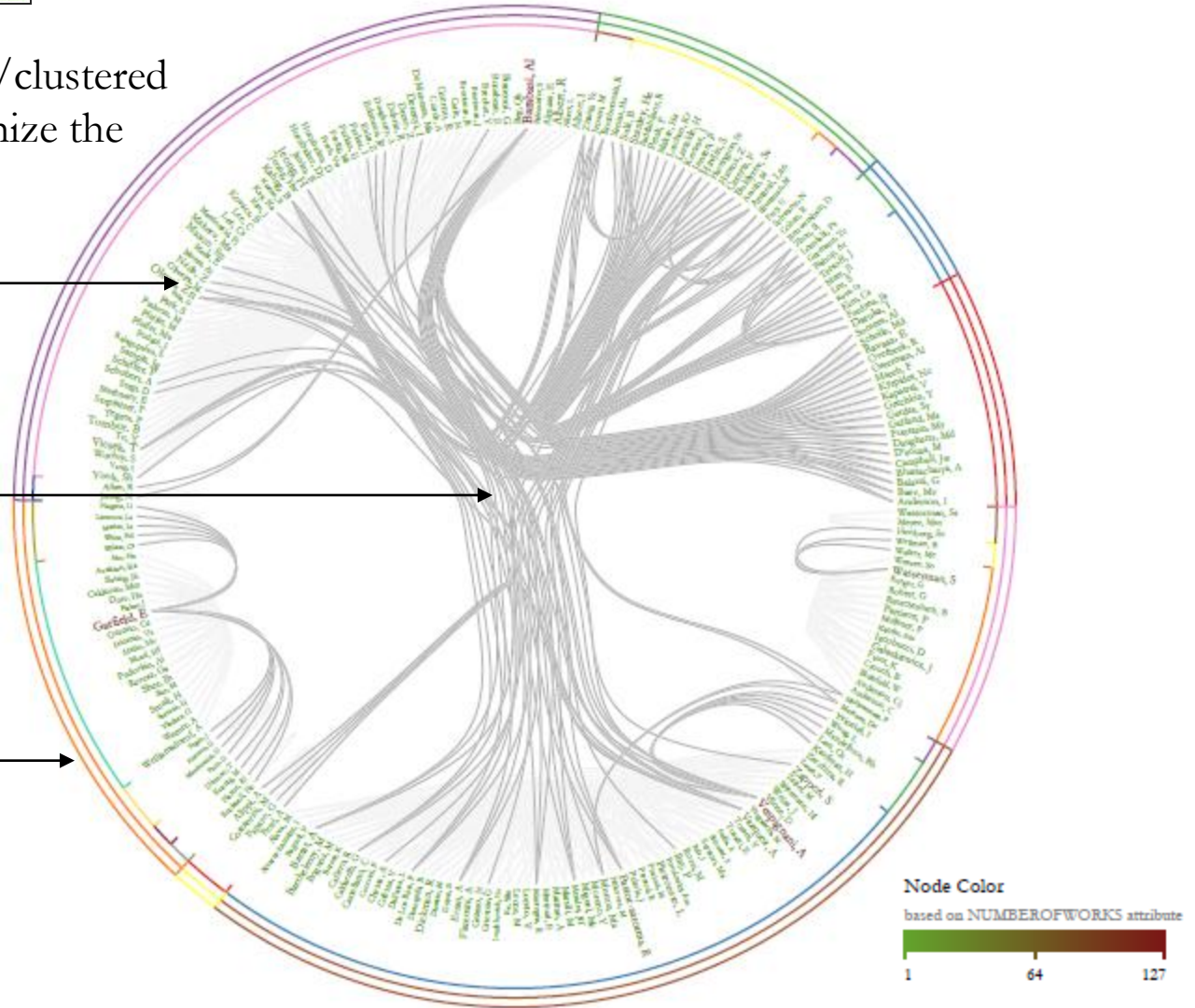
# Network Visualization: Circular Hierarchy Visualization

Nodes that are interlinked/clustered are spatially close to minimize the number of edge crossings.

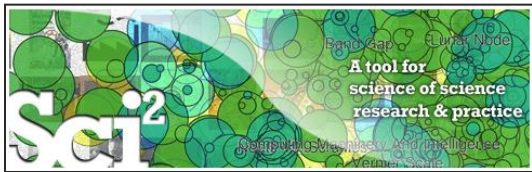
Node labels, e.g., author names.

Network structure using edge bundling.

Color coded cluster hierarchy according to Blondel community detection algorithm.



Note:  
Header/footer info, legend, and more meaningful color coding are under development.



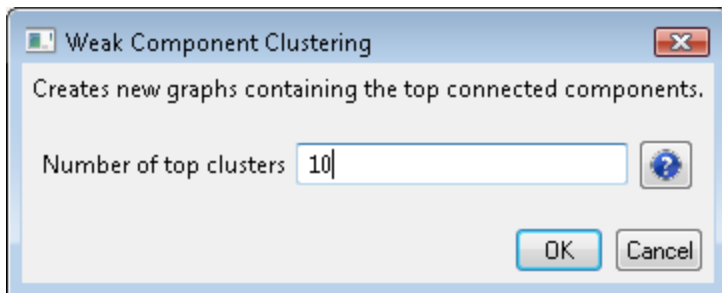
## Paper-Citation Network Layout

To extract the paper-citation network, select the *'361 Unique ISI Records'* table and run *'Data Preparation > Text Files > Extract Paper Citation Network.'*

The result is a unweighted, directed network of papers linked by citations, named *Extracted paper-citation network* in the Data Manager.

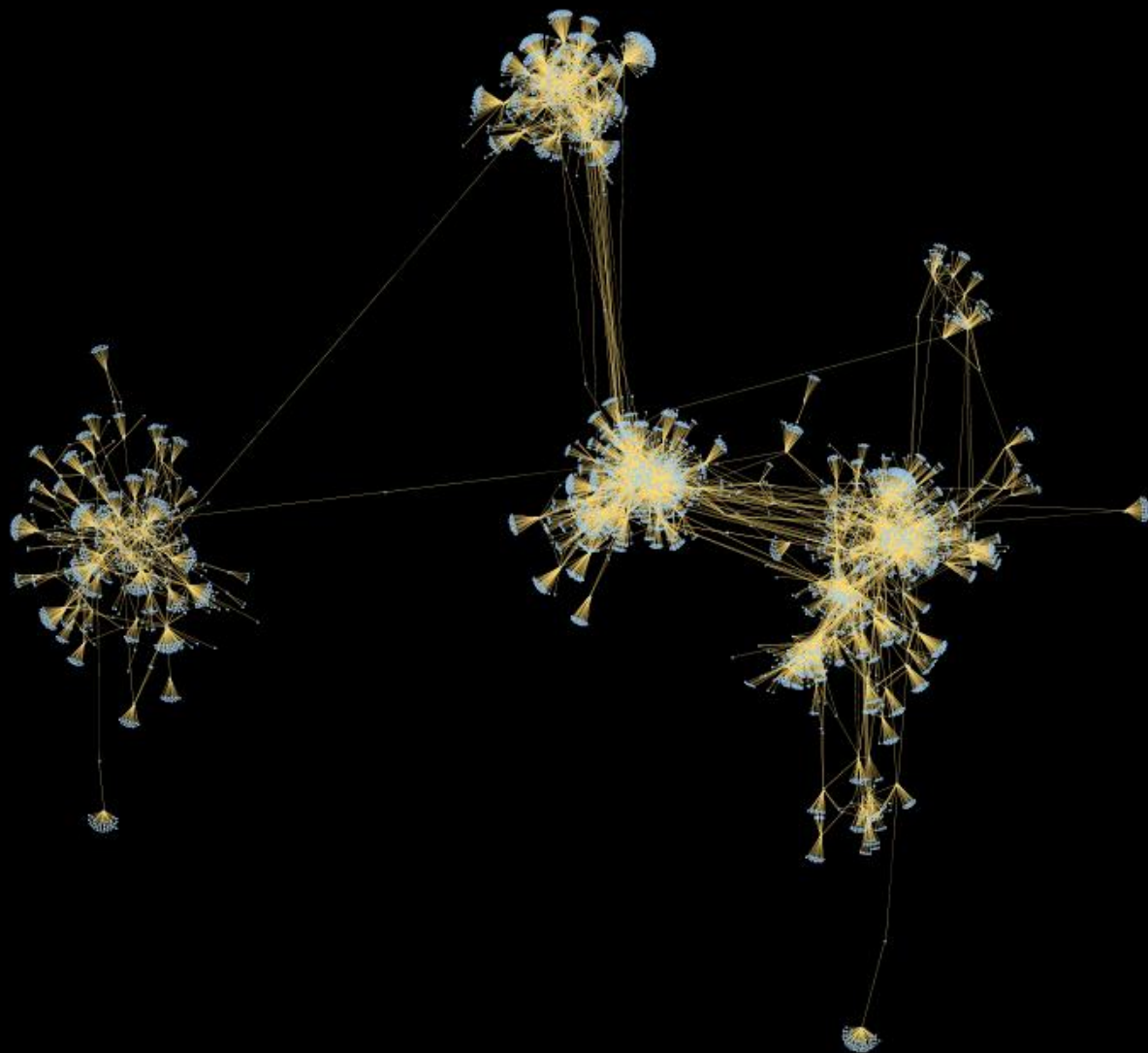
Run *NAT* to calculate that the network has 5,342 nodes and 9,612 edges. There are 15 weakly connected components. (0 isolates)

Run *'Analysis > Networks > Unweighted and Directed > Weak Component Clustering'* with parameters



to identify top-10 largest components. The largest (giant) component has 5,151 nodes.

To view the complete network, select the network and run *'Visualization > GUESS'*.







## Workshop Overview

10:55-11:05 The Importance and Dangers of Visualization – Use & Theory

11:05-11:20 The Epistolarium – Networks, Topics & Tools

11:20-11:25 Computational Modeling

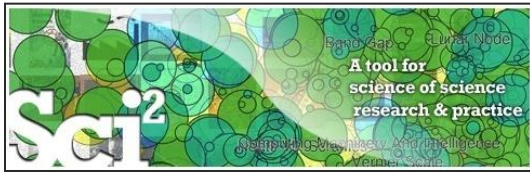
*11:25-11:35 Move to other room*

11:35-11:50 Sci2 Tool Basics

**11:50-12:10 Sci2 Workflow Design: Padgett's Florentine Families - Prepare, load, analyze, and visualize family and business networks from 15th century Florence.**

12:10-12:20 Q&A and Technical Assistance

**Workflow Design:  
Padgett's Florentine Families**



## Padgett's Florentine Families - Compute Basic Network Properties & View in GUESS

- Load *\*yoursci2directory\*/sampledata/socialscience/florentine.nwb*
- Run 'Analysis > Network Analysis Toolkit (NAT)' to get basic properties.

This graph claims to be undirected.

Nodes: 16

Isolated nodes: 1

Node attributes present: label, wealth, totalities, priorates

Edges: 27

No self loops were discovered.

No parallel edges were discovered.

Edge attributes:

Nonnumeric attributes:

Example value

marriag...T

busines...F

Average degree: 3.375

There are 2 weakly connected components. (1 isolates)

The largest connected component consists of 15 nodes.

Did not calculate strong connectedness because this graph was not directed.

Density (disregarding weights): 0.225

- **Optional:** Run 'Analysis > Unweighted & Undirected > Node Betweenness Centrality' with default parameters.
- Select network and run 'Visualization > GUESS' to open GUESS with file loaded.
- Apply 'Layout > GEM'.
- Open NWB File

Console

```

.....
GUESS was selected.
Author(s): Eytan Adar
Implementer(s): Eytan Adar (GUESS), Russell Duhon (resizeLinear, colorize fix)
Integrator(s): Russell Duhon
Reference: Adar, Eytan, "GUESS: A Language and Interface for Graph Exploration," CHI 2006 (http://graphexploration.cond.org/)
Documentation: https://nwb.slis.indiana.edu/community/?n=VisualizeData.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 Layout menu. (We recommend GEM.)
ECHO is off.
GUESS log file for this session can be found in
C:\DOCUME~1\katy\LOCALS~1\Temp\CIShell-Session-55892\StaticExecutableRunner-55904\ALGORITHM\guesslog.txt
.....
GUESS was selected.
Author(s): Eytan Adar
Implementer(s): Eytan Adar
Integrator(s): Russell Duhon
Reference: Adar, Eytan, "
Documentation: https://nwb.slis.indiana.edu/community/?n=VisualizeData.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 Layout menu. (We recommend GEM.)
ECHO is off.
GUESS log file for this session can be found in
C:\DOCUME~1\katy\LOCALS~1\Temp\CIShell-Session-55892\StaticExecutableRunner-55904\ALGORITHM\guesslog.txt
.....

```

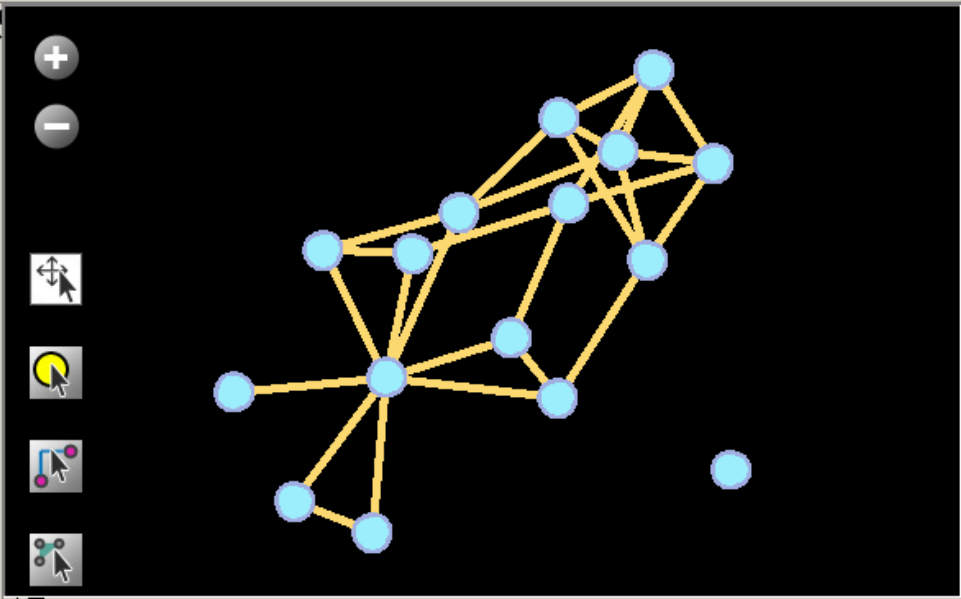
Data Manager

- NWB file: C:\Documents and Settings\katy\Documents\Medici-Acciaiuoli.nwb
- Distribution of degree for network at st...
- List of edges of network created through th...
- Distribution of degree for network at st...
- Distribution of degree for network at st...
- NWB file: C:\Documents and Settings\katy\Documents\Medici-Acciaiuoli.nwb
- Graph and Network Analysis Log
- Sequence of site betweennesses for netwo...
- Distribution of site betweenness for netwo...
- Distribution of site betweenness for netwo...
- NWB file with site betweenness attribute...

Visualization - GUESS

Medici-Acciaiuoli

Field	Value
__edgeid	0
business	F
color	dandelion
directed	false
label	
labelcolor	0,0,0,255
labelsize	12
labelvisible	false
marriage	T
node1	n9
node2	n1
visible	true
weight	1.0
width	2.0



Object: [ ] Property: labelsize Operator: == Value: [ ]

Colour Show Hide Size Show Label Hide Label Change Label

Node Shape Center Change History

Interpreter Graph Modifier

Scheduler

Remove From List



- | !                                   | Algorithm Name |
|-------------------------------------|----------------|
| <input type="checkbox"/>            | GUESS          |
| <input type="checkbox"/>            | GUESS          |
| <input type="checkbox"/>            | Node Betw      |
| <input checked="" type="checkbox"/> | Network Ar     |
| <input checked="" type="checkbox"/> | Load...        |
| <input checked="" type="checkbox"/> | GnuPlot        |
| <input checked="" type="checkbox"/> | Degree Dis     |
| <input checked="" type="checkbox"/> | Barabási-A     |
| <input checked="" type="checkbox"/> | GnuPlot        |
| <input checked="" type="checkbox"/> | Degree Dis     |
| <input checked="" type="checkbox"/> | GUESS          |
| <input checked="" type="checkbox"/> | Load...        |

Visualization - GUESS

File Edit Layout Script View Help

Information Window

Field	Value
color	cornflo...
fixed	false
height	10.0
image	
label	Medici
labelcolor	0,0,0,...
labelsize	12
labelvi...	false
name	n9
original...	Medici
priorates	53
stroke...	cadetb...
style	2
totalities	54
visible	true
wealth	103
width	10.0
x	90.625...
y	44.312...

Object:  Property:  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

Nodes  From:  To:  Do Resize Linear


Interpreter Graph Modifier

Pan:

“grab” the background by holding left-click and moving your mouse.

Zoom:

Using scroll wheel, press the “+” and “-” buttons in the upper-left hand corner, or right-click and move the mouse left or right. Center graph by selecting ‘View -> Center’.

Select  to select/move single nodes. Hold down ‘Shift’ to select multiple.

Right click to modify Color, etc.

Visualization - GUESS

File Edit Layout Script View Help

**Pucci**

Field	Value
color	255,25...
fixed	false
height	5.0
image	
label	Pucci
labelcolor	0,0,0,...
labelsize	12
labelvi...	true
name	n12
original...	Pucci
priorates	0
stroke...	cadetb...
style	2
totalities	1
visible	true
wealth	3
width	5.0
x	15.0
y	91.100...

Object:  Property:  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

Interpreter Graph Modifier

## Graph Modifier:

Select “all nodes” in the Object drop-down menu and click ‘Show Label’ button.

Select ‘Resize Linear > Nodes > totalities’ drop-down menu, then type “5” and “20” into the “From” and “To” Value box separately. Then select ‘Do Resize Linear’.

Select ‘Colorize > Nodes > totalities’, then select white and enter (204,0,51) in the pop-up color boxes on in the “From” and “To” buttons.

Select “Format Node Labels”, replace default text {originallabel} with your own label in the pop-up box ‘Enter a formatting string for node labels.’

Visualization - GUESS

File Edit Layout Script View Help

Information Window

**Acciaiuoli:)**

Field	Value
color	255,24...
fixed	false
height	5.6635...
image	
label	Acciaiu...
labelcolor	0,0,0,...
labelsize	12
labelvi...	true
name	n1
original...	Acciaiuoli
priorates	53
stroke...	cadetb...
style	2
totalities	2
visible	true
wealth	10
width	5.6635...
x	112.01...
y	70.315...

```

>>> resizeLinear(totalities,5,20)
>>> colorize(wealth,white,red)
>>>

```

Interpreter Graph Modifier

### Interpreter:

Uses Jython a combination of Java and Python.

Try

`colorize(wealth, white, red)`

`resizeLinear(sitebetweenness, 5, 25)`



## Workshop Overview

10:55-11:05 The Importance and Dangers of Visualization – Use & Theory

11:05-11:20 The Epistolarium – Networks, Topics & Tools

11:20-11:25 Computational Modeling

*11:25-11:35 Move to other room*

11:35-11:50 Sci2 Tool Basics

11:50-12:10 Sci2 Workflow Design: Padgett's Florentine Families - Prepare, load, analyze, and visualize family and business networks from 15th century Florence.

**12:10-12:20 Q&A and Technical Assistance**