

“Sci2 Tool: A Tool for Science of Science Research and Practice” Tutorial



Dr. Katy Börner

CNS & IVL, SLIS, Indiana University, Bloomington, Indiana, USA, <http://cns.iu.edu>

With support by Keiko Kayukawa, Katsuhide Fujita, and Junichiro Mori and the Innovation Policy Research Center, The University of Tokyo.

Special thanks go to Kevin W. Boyack, Chin Hua Kong, Micah Linnemeier, Russell J. Duhon, Patrick Phillips, Chintan Tank, Thomas Smith, Nianli Ma, Joseph R. Biberstine, David Coe, Scott Weingart, Hanning Guo, Mark A. Price, Angela M. Zoss, Ted Polley, and Sean Lind.

Please (1) get a name tag, (2) download the Sci2 Tool from <http://sci2.cns.iu.edu> and (3) complete the Pre-Tutorial Questionnaire

The University of Tokyo, Tokyo, Japan
Friday December 14, 2012 • 9am-12:30pm



Software, Datasets, Plugins, and Documentation
also distributed on Memory Stick

- These slides
<http://ivl.slis.indiana.edu/km/pres/2012-borner-sci2tutorial-tokyo.pdf>
- Sci2 Tool Manual v0.5.1 Alpha, updated to match v1.0 Alpha
<http://sci2.wiki.cns.iu.edu>
- Sci2 Tool v1.0 Alpha (June 13, 2012)
<http://sci2.cns.iu.edu>
- Additional Datasets
<http://sci2.wiki.cns.iu.edu/2.5+Sample+Datasets>
- Additional Plugins
<http://sci2.wiki.cns.iu.edu/3.2+Additional+Plugins>



Or copy them from the DVD or memory stick.

Postscript Viewer: Please try opening ‘chessboard.ps.’ You should see





Sci2 Tool v0.5.2 Alpha (Dec 19, 2011)

New Features

- Support new Web of Science format from ISI
- Support network overlay for geographical map
- Support Prefuse's visualizations on Macs OS

Improvements

- Improve memory usage and processing time of Extract top N nodes and Extract top N Edges algorithms
- Unify merging algorithms used by database

Bug fixes

- Fix legend boundary issue in geographical map
- Fix typo error on the output data label
- Fix slice by year algorithm

3



Sci2 Tool v1.0 Alpha (June 13, 2012)

Major Release

featuring a Web services compatible CShell v2.0 (<http://cishell.org>)

New Features

- Google Scholar citation reader
- New visualizations such as
 - geospatial maps
 - science maps
 - bi-modal network layout
- R statistical tool bridging
- Gephi visualization tool bridging
- Comprehensive online documentation

Release Note Details

<http://wiki.cns.iu.edu/display/SCI2TUTORIAL/4.4+Sci2+Release+Notes+v1.0+alpha>

4



Tutorial Overview

9:00am Welcome and Overview of Tutorial and Attendees

9:15am Sci2 Tool Usage and Adoption

10:00am Sci2 Tool Basics

Download and run the Sci2 Tool

Load, analyze, and visualize family and business networks

Horizontal line graph of NSF projects

Studying four major network science researchers

- Load and clean a dataset; process raw data into networks
- Find basic statistics and run various algorithms over the network
- Visualize the network using different layouts

11:00am Break

11:15am Sci2 Tool Novel Functionality

- Yahoo! Geocoder
- Geomap with Gephi network overlay
- Evolving collaboration networks
- R-Bridge
- New Visualizations

12:15pm Outlook and Q&A

12:30pm Adjourn

5



Tutorial Participants

Affiliation

University of Florida
US EPA
Kwansei Gakuin University
Yamagata University
National Institute of Science and Technology Policy
NPG Nature Asia-Pacific
Waseda University
RIETI
University of Nebraska-Lincoln
Elsevier
Hokkaido University
Hitotsubashi University
Center for Aviation Innovation Research, University of Tokyo
University of Tokyo
Central Research Institute of Electric Power Industry
University of Tokyo
University of Tokyo
University of Tokyo
UT, TMI
National Institute for Materials Science
TMI, University of Tokyo

Interested to Learn:

- The basics
- Topical analysis and maps
- Visualization of a large-scale network
- Understanding innovation dynamics
- S&E collaborations between nations
- Patent analysis -> SDB
- Comparison with UCINet, GIS
- Capabilities and limitations of the tool
- Whether a similar tool can be implemented in Japan

6



Tutorial Overview

9:00am Welcome and Overview of Tutorial and Attendees

9:15am Sci2 Tool Usage and Adoption

10:00am Sci2 Tool Basics

Download and run the Sci2 Tool

Load, analyze, and visualize family and business networks

Horizontal line graph of NSF projects

Studying four major network science researchers

- Load and clean a dataset; process raw data into networks
- Find basic statistics and run various algorithms over the network
- Visualize the network using different layouts

11:00am Break

11:15am Sci2 Tool Novel Functionality

- Yahoo! Geocoder
- Geomap with Gephi network overlay
- Evolving collaboration networks
- R-Bridge
- New Visualizations

12:15pm Outlook and Q&A

12:30pm Adjourn

7



Science of Science (Sci2) Tool

<http://sci2.cns.iu.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

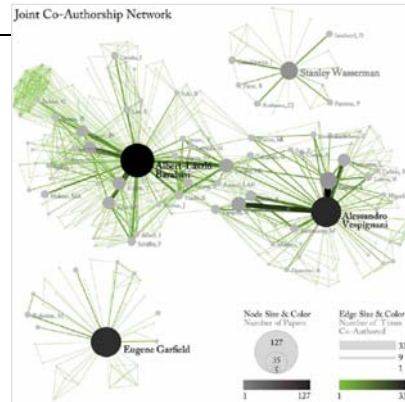
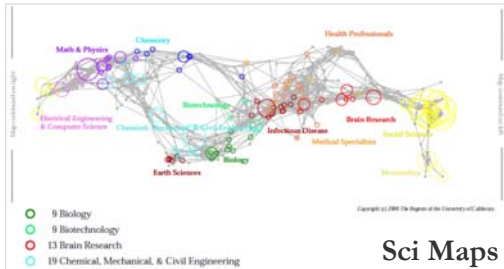
To capture the essence of good science, stakeholders must combine forces to create an open, sound and consistent system for measuring all the activities that make up academic productivity, says **Julia Lane**.

8

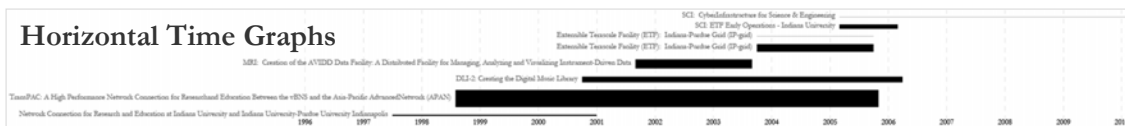


Sci² Tool – “Open Code for S&T Assessment”

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



Horizontal Time Graphs



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

9

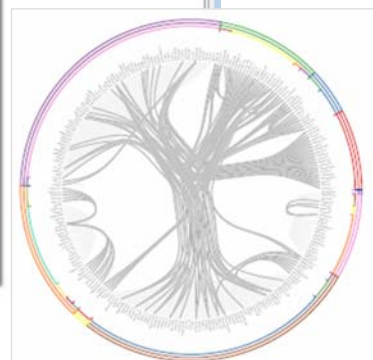
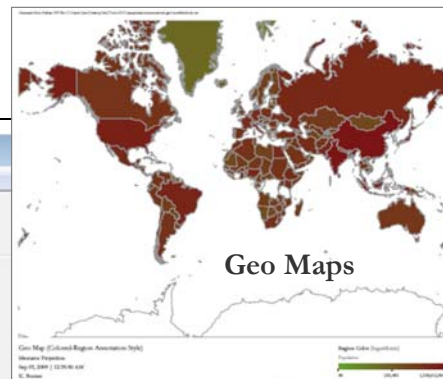


Sci² Tool Visualizations

Visualization Menu:

- 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

!	Algorithm Name	Date	Time	% Con
✓	Extract Co-Author Netw...	09/03/2009	00:15:20 AM	100%
✓	Load and Clean ISI File	09/03/2009	00:15:05 AM	100%

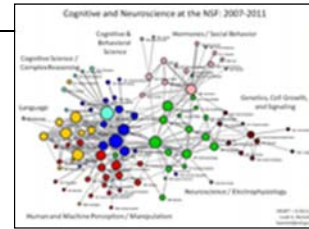


10



Sci² Tool Usage at National Science Foundation

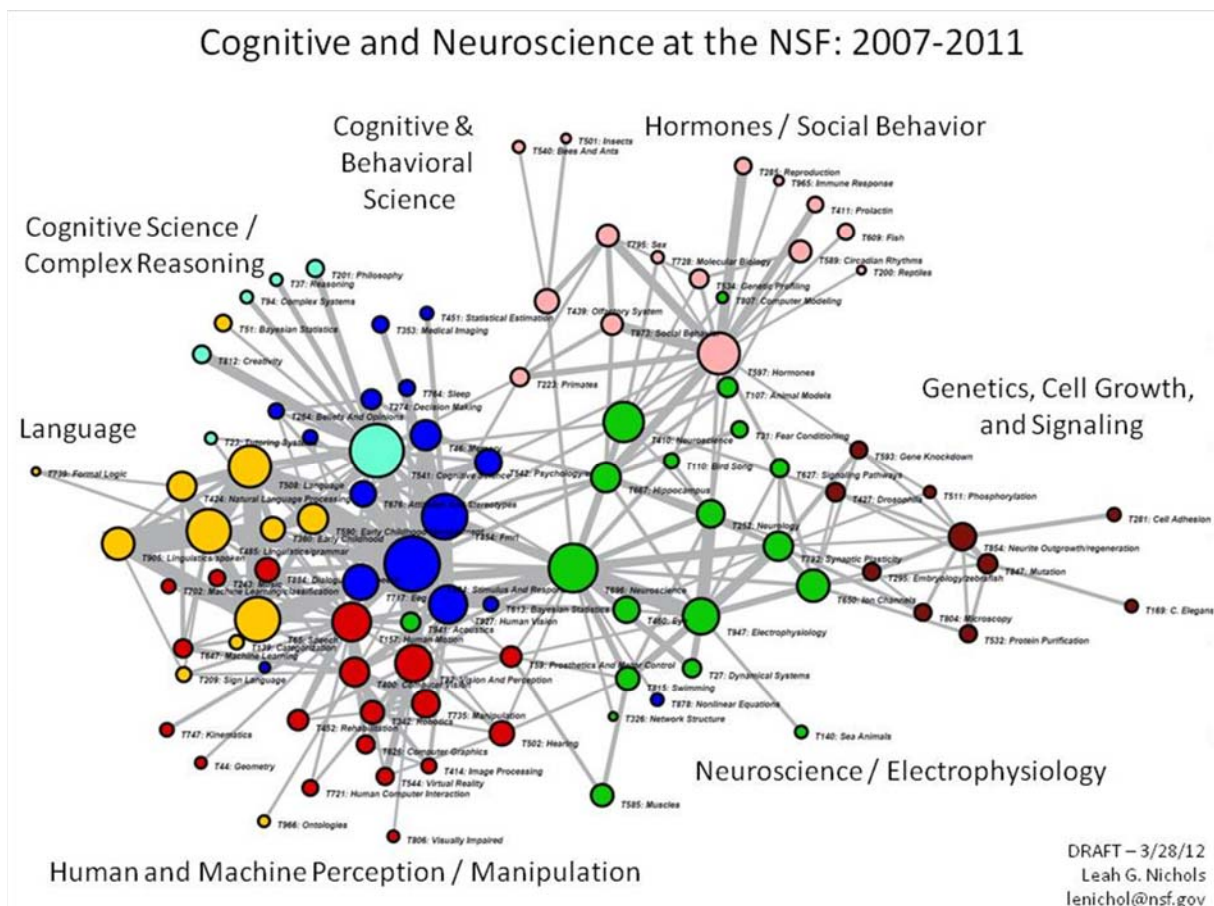
Topic co-occurrence network of the 2885 cognitive and neuroscience NSF projects funded between 2007 and 2011. Statistical text mining (Topic model) was used to identify topics from NSF awards and proposals from 2000-2011.



Each award is tagged with up to 4 topics. Lines represent the co-occurrence of the connected topics within an award(s). The nodes are scaled by number of awards (max = 355) and the lines are scaled on number of co-occurrences (max =91). The node colors differentiate the nodes via the level-0 Blondel communities.

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





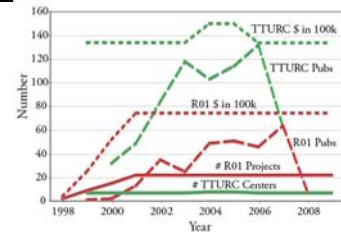
Sci² Tool Usage at National Institutes of Health

Mapping Transdisciplinary Tobacco Use Research Centers Publications: Compare R01 investigator based funding with TTURC Center awards in terms of

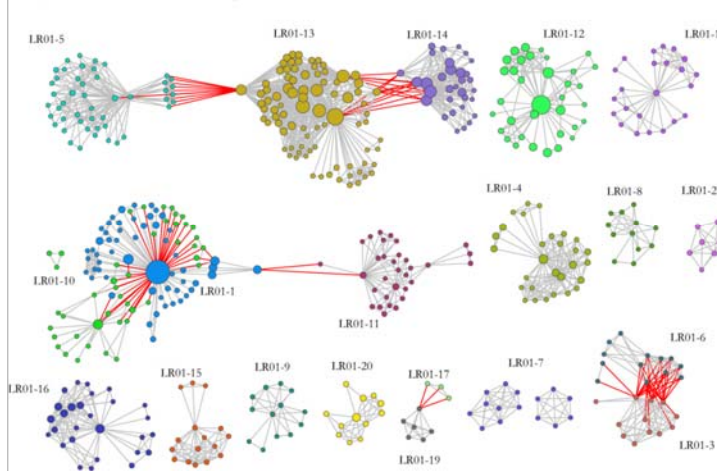
number of publications and evolving co-author networks. *Zoss & Börner, forthcoming.*

Supported by NIH/NCI Contract HHSN261200800812

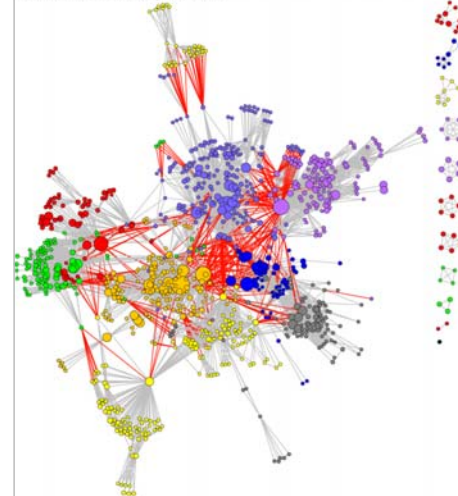
R01 & TTURC Project Information



Longitudinal R01 Co-Authorship Network



TTURC Co-Authorship Network



Sci² Tool Usage at National Institutes of Health

Sci2 Tool now supports Web services and serves as a visual interface to publicly available NIH RePORT Expenditure and Results (RePORTER)/ RePORTER data provided by NIH.

NETE[®] AV
ANALYZER | VISUALIZER

TEMPORAL ANALYSIS GEOSPATIAL ANALYSIS TOPICAL ANALYSIS NETWORK ANALYSIS

"WHERE"^{IS} GEOSPATIAL ANALYSIS

"When" questions are commonly addressed via temporal analyses

"Where" questions often involve the application of geospatial methods

"What" questions require topical analyses

"With whom" questions are often answered via network studies

W H E N
W H E R E
W H A T
W H O M



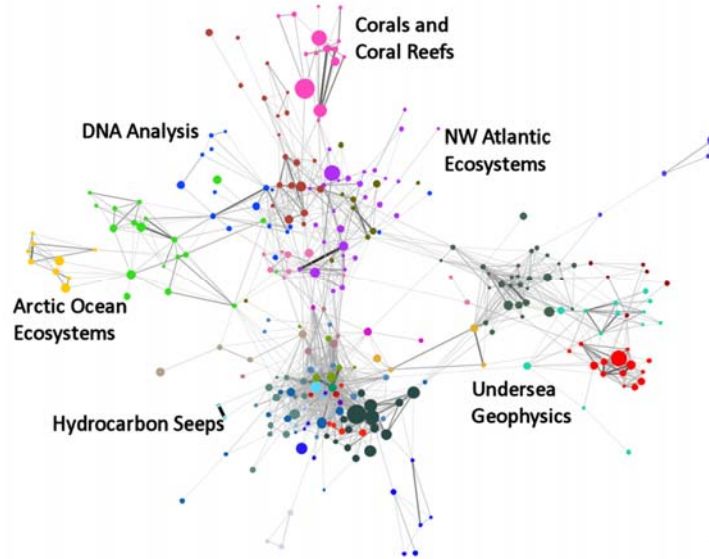
Sci² Tool Usage at the National Oceanic and Atmospheric Administration (NOAA)

Co-author network generated from publications supported by NOAA's Office of Ocean Exploration and Research (OER). Nodes are sized based on the number of publications produced and colored to highlight clustering. Edges are sized and colored based on the number of collaborations between authors.

For details, see "Visualizing Networks of Scientific Research"

by Chris Belter

<http://www.infotoday.com/online/may12/Belter-Visualizing-Networks-of-Scientific-Research.shtml>

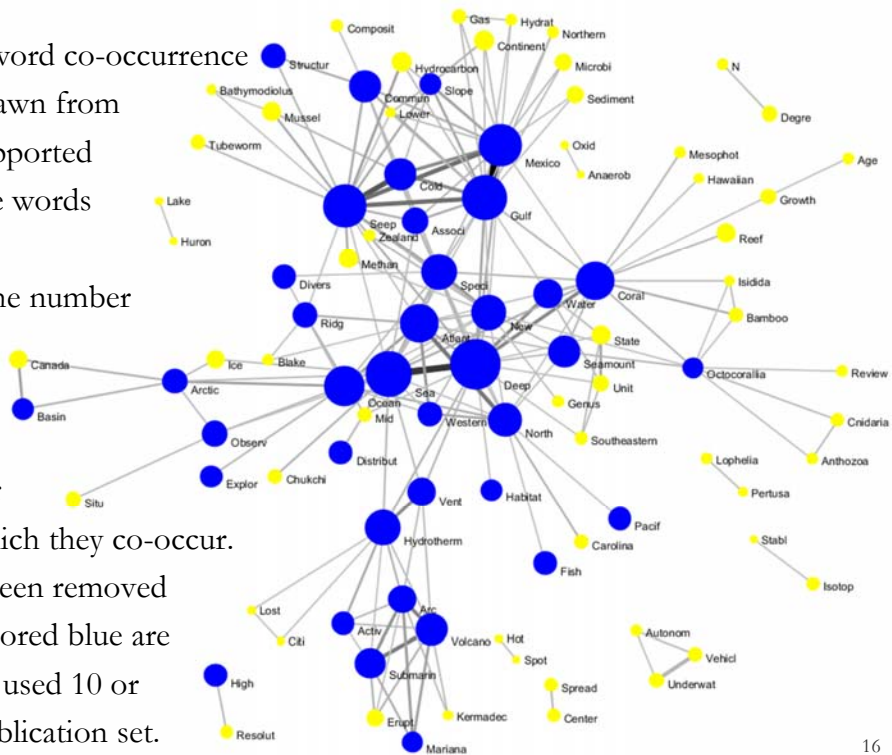


15



Sci² Tool Usage at the National Oceanic and Atmospheric Administration (NOAA)

Central section of a word co-occurrence network of words drawn from the titles of OER-supported journal articles. Some words have been truncated. Nodes are sized by the number of times the word is used and edges are sized and colored based on the number of publications in which they co-occur. Weaker edges have been removed for clarity. Nodes colored blue are those that have been used 10 or more times in the publication set.

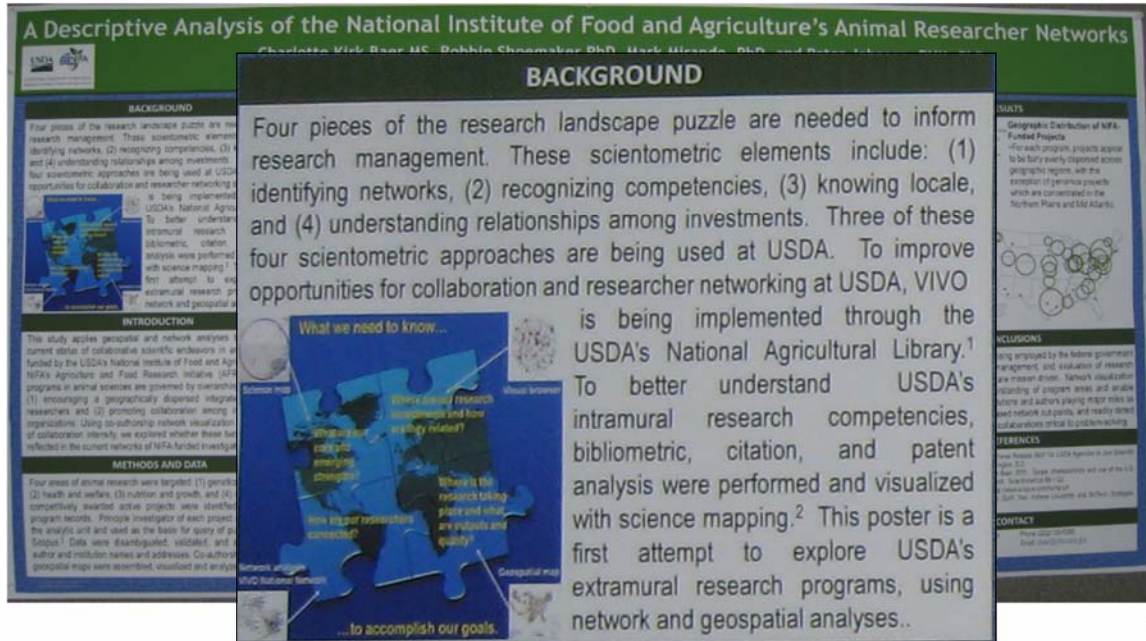


16



Sci² Tool Usage at US Department of Agriculture (USDA)

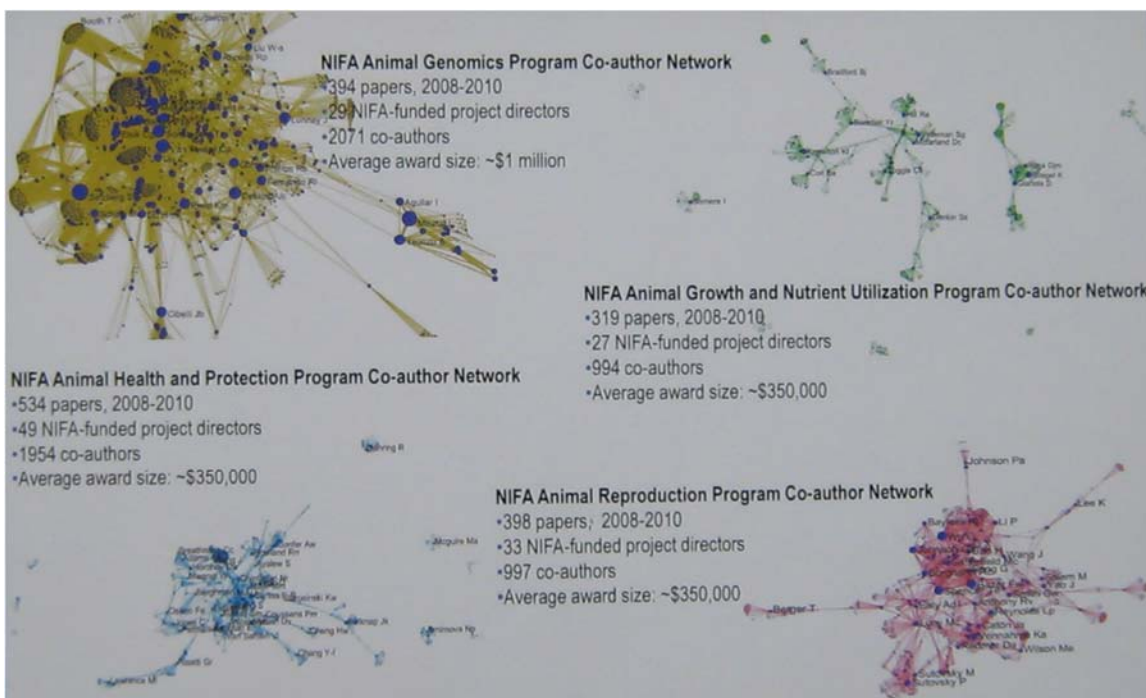
First time portrait of intramural research conducted by the U.S. Department of Agriculture (USDA) presented at the VIVO Conference 2012.



17



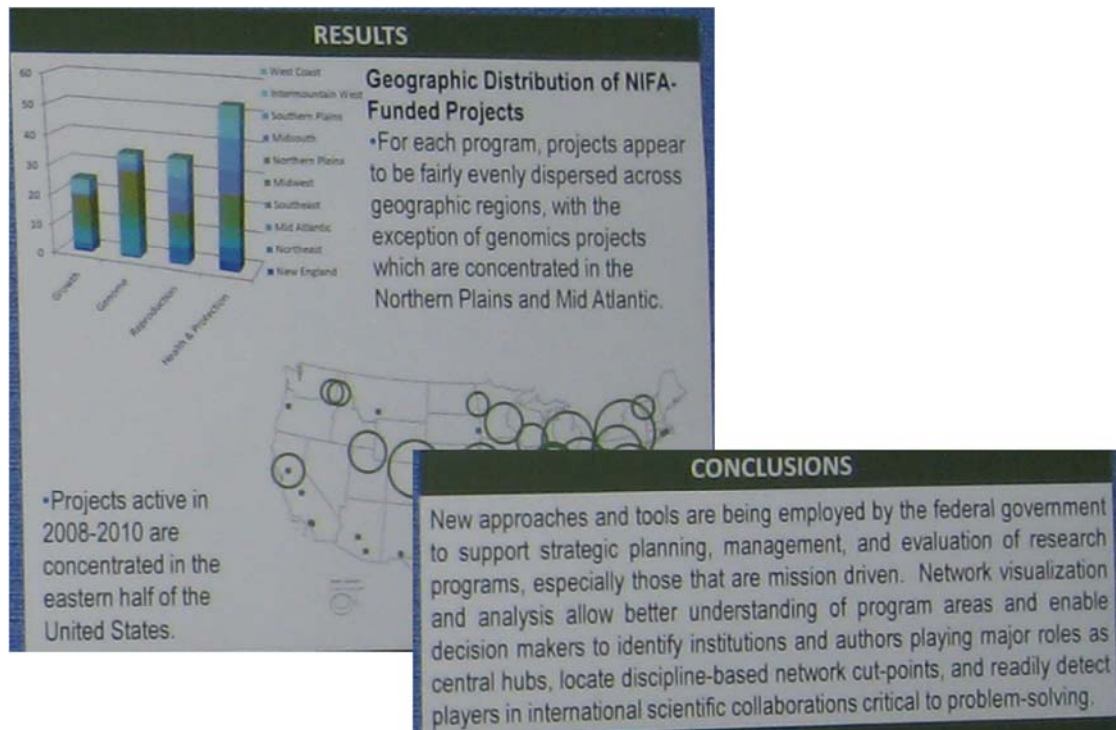
Sci² Tool Usage at US Department of Agriculture



18



Sci² Tool Usage at US Department of Agriculture



19



Sci² Tool Usage at James S. McDonnell Foundation

How did cognitive neuroscience of attention emerge from neurobiology and psychology, 1980–2005? Author co-citation analysis and Pfnets is used to **trace prospectively the development of the field from its precursor disciplines:** cognitive psychology, single cell neurophysiology, neuropsychology, and evoked potential research.

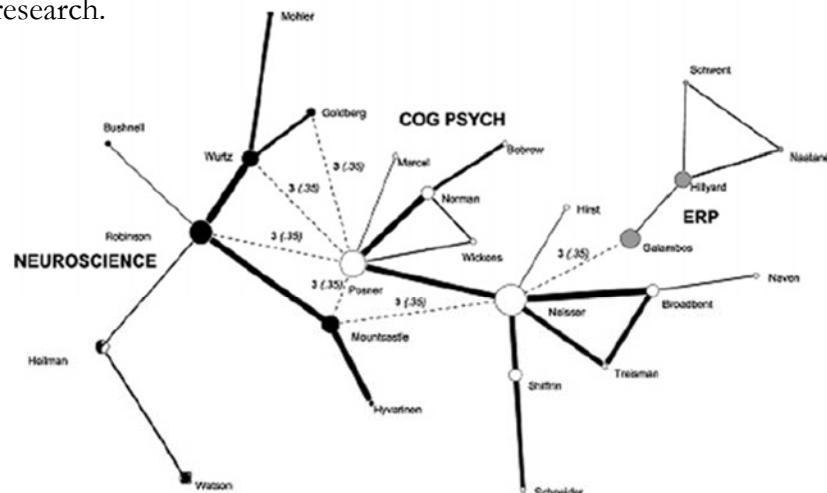


Fig. 1 In the 1980 net, neuroscience (black nodes and black–white nodes) and cognitive psychology (white nodes) develop as clusters with high internal co-citation rates. ERP (grey nodes) develops later in net construction. These clusters are connected by secondary edges at very low levels of co-citation

20



By 1990 a distinct cognitive neuroscience specialty cluster emerges, dominated by authors engaged in brain imaging research.

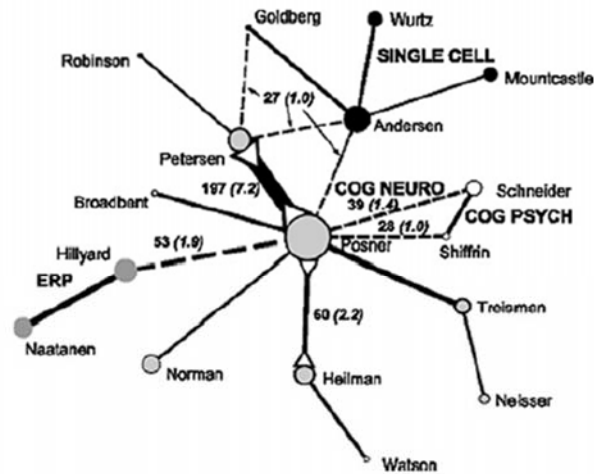


Fig. 5 The strongest link in the 1995 net is a primary edge linking Posner and Petersen. ERP and single cell neurophysiology are linked to cognitive neuroscience cluster by secondary edges

Bruer, John T. (2010). *Can we talk? How the cognitive neuroscience of attention emerged from neurobiology and psychology, 1980.2005.* *Scientometrics*, 83(3), 751-764. <http://inl.cns.iu.edu/km/tools/2010-bruer-scientometrics.pdf>

21



Tutorial Overview

9:00am Welcome and Overview of Tutorial and Attendees

9:15am Sci² Tool Usage and Adoption

10:00am Sci² Tool Basics

Download and run the Sci² Tool

Load, analyze, and visualize family and business networks

Horizontal line graph of NSF projects

Studying four major network science researchers

- Load and clean a dataset; process raw data into networks
- Find basic statistics and run various algorithms over the network
- Visualize the network using different layouts

11:00am *Break*

11:15am Sci² Tool Novel Functionality

- Yahoo! Geocoder
- Geomap with Gephi network overlay
- Evolving collaboration networks
- R-Bridge
- New Visualizations

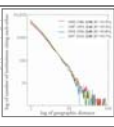
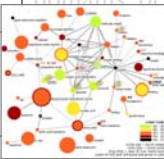


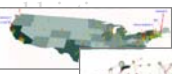
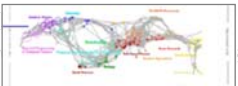
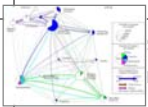


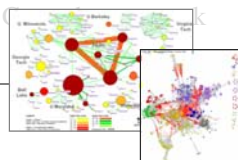
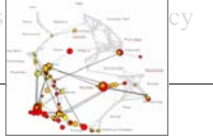
12:15pm Outlook and Q&A

12:30pm *Adjourn*

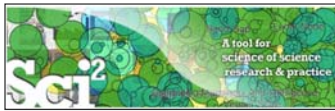
22




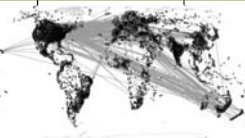

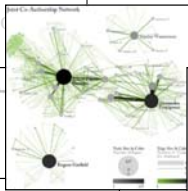
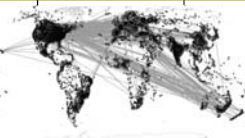

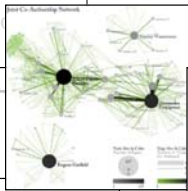

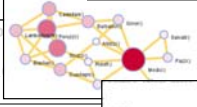

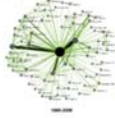
Type of Analysis vs. Level of Analysis

	<i>Micro/Individual</i> (1-100 records)	<i>Meso/Local</i> (101-10,000 records)	<i>Macro/Global</i> (10,000 < records)
Statistical Analysis/Profiling	Individual person and their expertise profiles	Larger labs, centers, universities, research domains, or states	All of NSF, all of USA, all of science. 
Temporal Analysis (When)	Funding portfolio of one individual	Research bursts of PNAS 	113 Years of PNAS Research 
Geospatial Analysis (Where)	Career trajectory of one individual	Mapping a network of intellectual links 	PNAS 
Topical Analysis (What)		Research 	VxOrd/Topic r NIH funding 
Network Analysis (With Whom?)	NSI network of one 	Network 	NIH's network of one 

23



Type of Analysis vs. Level of Analysis

	<i>Micro/Individual</i> (1-100 records)	<i>Meso/Local</i> (101-10,000 records)	<i>Macro/Global</i> (10,000 < records)
Statistical Analysis/Profiling	Individual person and their expertise profiles	Larger labs, centers, universities, research domains, or states	All of NSF, all of USA, all of science.
Temporal Analysis (When)	Funding portfolio of one individual	Mapping in 20-year 	Physics 
Geospatial Analysis (Where)	Career trajectory individual 		
Topical Analysis (What)			Ord/ NIH funding 
Network Analysis (With Whom?)	NSF Co-P one indiv 	-auth 	H's core competency 

24



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)

Image Formats

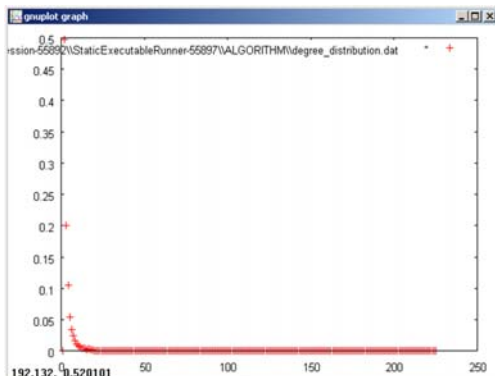
- JPEG (*.jpg)
- PDF (*.pdf)
- PostScript (*.ps)

Formats are documented at <http://sci2.wiki.cns.iu.edu/display/SCI2TUTORIAL/2.3+Data+Formats>.

25

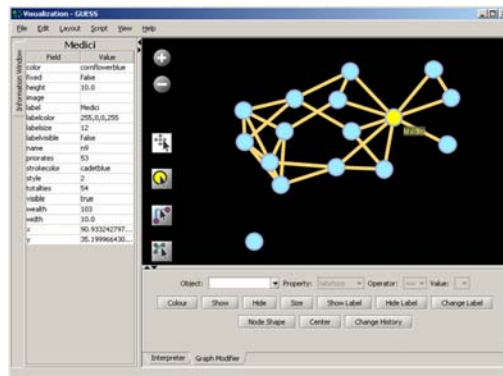


Sci² Tool – Supported Tools



Gnuplot

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



GUESS

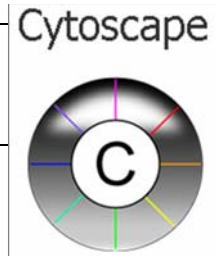
exploratory data analysis and visualization tool
for graphs and networks.

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

26



Sci2 Tool – Supported Tools

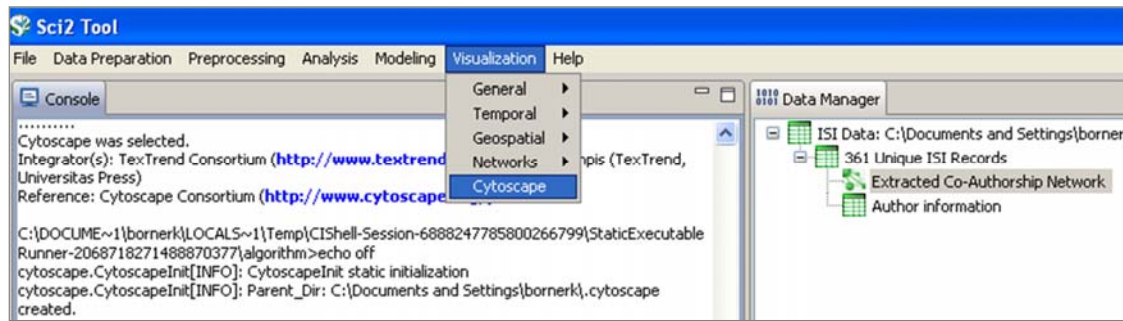


Adding more layout algorithms and network visualization interactivity via Cytoscape <http://www.cytoscape.org>.

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

Restart Sci2 Tool

Cytoscape now shows in the Visualization Menu



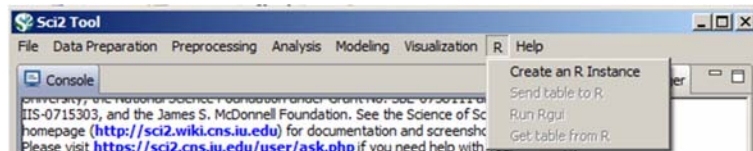
Select a network in Data Manager, run Cytoscape and the tool will start with this network loaded.

27

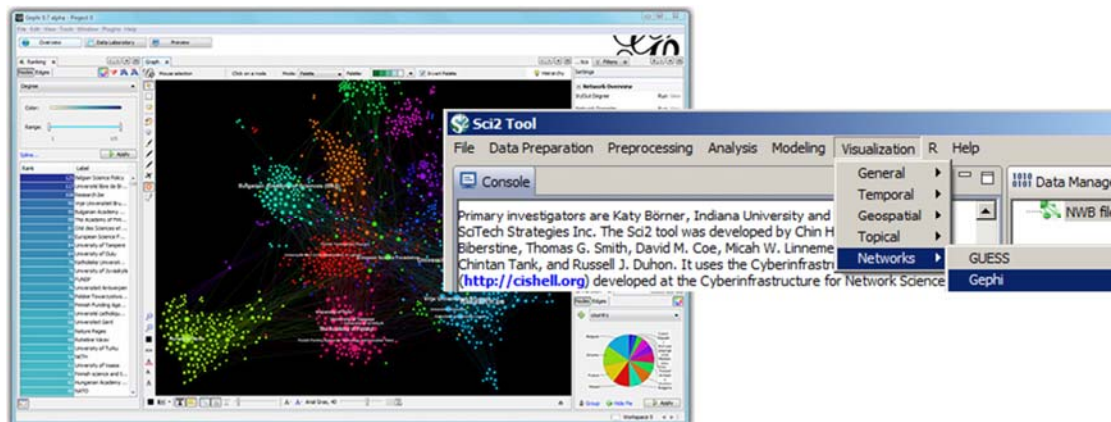


Sci2 Tool – Bridged Tools

R statistical tool bridging



Gephi visualization tool bridging



28



Sci² Tool: Download, Install, and Run

Sci² Tool v1.0 Alpha (June 13, 2012)

Can be freely downloaded for all major operating systems from

<http://sci2.cns.iu.edu>

Select your operating system from the pull down menu and download.

Unpack into a /sci2 directory.

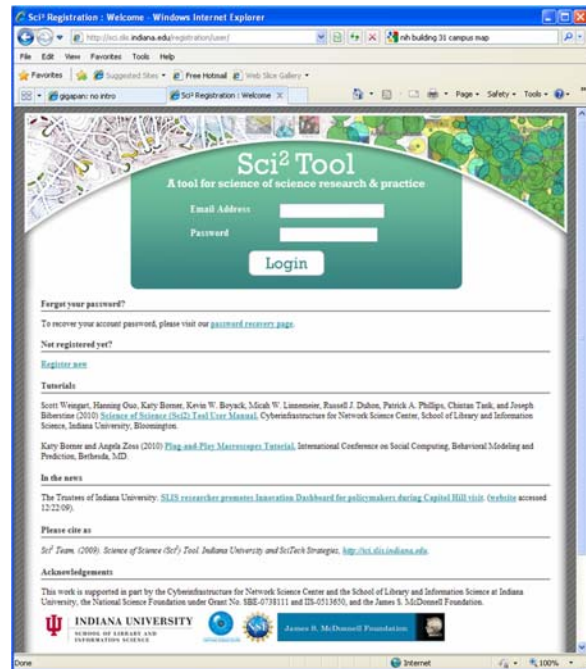
Run /sci2/sci2.exe

Sci² Manual is at

<http://sci2.wiki.cns.iu.edu>

Cite as

Sci² Team. (2009). Science of Science (Sci²) Tool. Indiana University and SciTech Strategies, <http://sci2.cns.iu.edu>



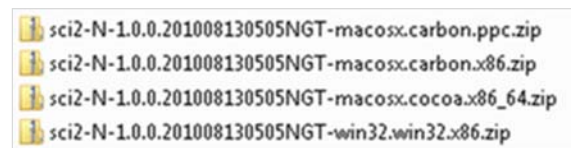
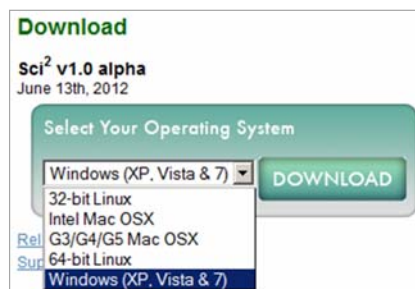
29



Sci² Tool: Download, Install, and Run

Sci² Tool v1.0 Alpha (June 13, 2012)

Sci² Tool runs on Windows, Mac, and Linux.



Unzip.



Run /sci2/sci2.exe

30

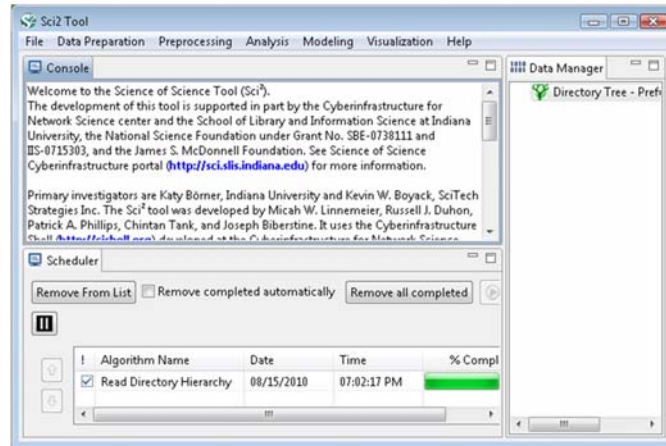


Sci2 Tool Interface Components

See also <http://sci2.wiki.cns.iu.edu/2.2+User+Interface>

Use

- **Menu** to read data, run algorithms.
- **Console** to see work log, references to seminal works.
- **Data Manager** to select, view, save loaded, simulated, or derived datasets.
- **Scheduler** to see status of algorithm execution.



All workflows are recorded into a log file (see /sci2/logs/...), and soon can be re-run for easy replication. If errors occur, they are saved in a error log to ease bug reporting.

All algorithms are documented online; workflows are given in tutorials, see Sci2 Manual at <http://sci2.wiki.cns.iu.edu>

31



Tutorial Overview

9:00am Welcome and Overview of Tutorial and Attendees

9:15am Sci2 Tool Usage and Adoption

10:00am Sci2 Tool Basics

Download and run the Sci2 Tool

Load, analyze, and visualize family and business networks

Horizontal line graph of NSF projects

Studying four major network science researchers

- Load and clean a dataset; process raw data into networks
- Find basic statistics and run various algorithms over the network
- Visualize the network using different layouts

11:00am Break

11:15am Sci2 Tool Novel Functionality

- Yahoo! Geocoder
- Geomap with Gephi network overlay
- Evolving collaboration networks
- R-Bridge
- New Visualizations

12:15pm Outlook and Q&A

12:30pm Adjourn

32



Padgett's Florentine Families – Load, analyze, and visualize family and business networks

Florentine families related through business ties (specifically, recorded financial ties such as loans, credits and joint partnerships) and marriage alliances.

Node attributes

- Wealth: Each family's net wealth in 1427 (in thousands of lira).
- Priorates: The number of seats on the civic council held between 1282-1344.
- Totalities: Number of business/marriage ties in complete dataset of 116 families.

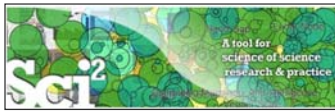
Edge attributes:

- Marriage T/F
- Business T/F

“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 Medicis, the other around the powerful Strozzi.”

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

33



Padgett's Florentine Families – Load, analyze, and visualize family and business networks

Florentine families related through business ties (specifically, recorded financial ties such as loans, credits and joint partnerships) and marriage alliances.

Node attributes

- Wealth: Each family's net wealth in 1427 (in thousands of lira).
- Priorates: The number of seats on the civic council held between 1282-1344.
- Totalities: Number of business/marriage ties in complete dataset of 116 families.

Edge attributes:

- Marriage T/F
- Business T/F

“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 Medicis, the other around the powerful Strozzi.”

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

```
*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"
```

34



Padgett's Florentine Families – Load, compute basic network properties & view in GUESS

- Load **yoursci2directory*/sampledata/socialscience/florentine.mwb*
- 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

marriage...T

business...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

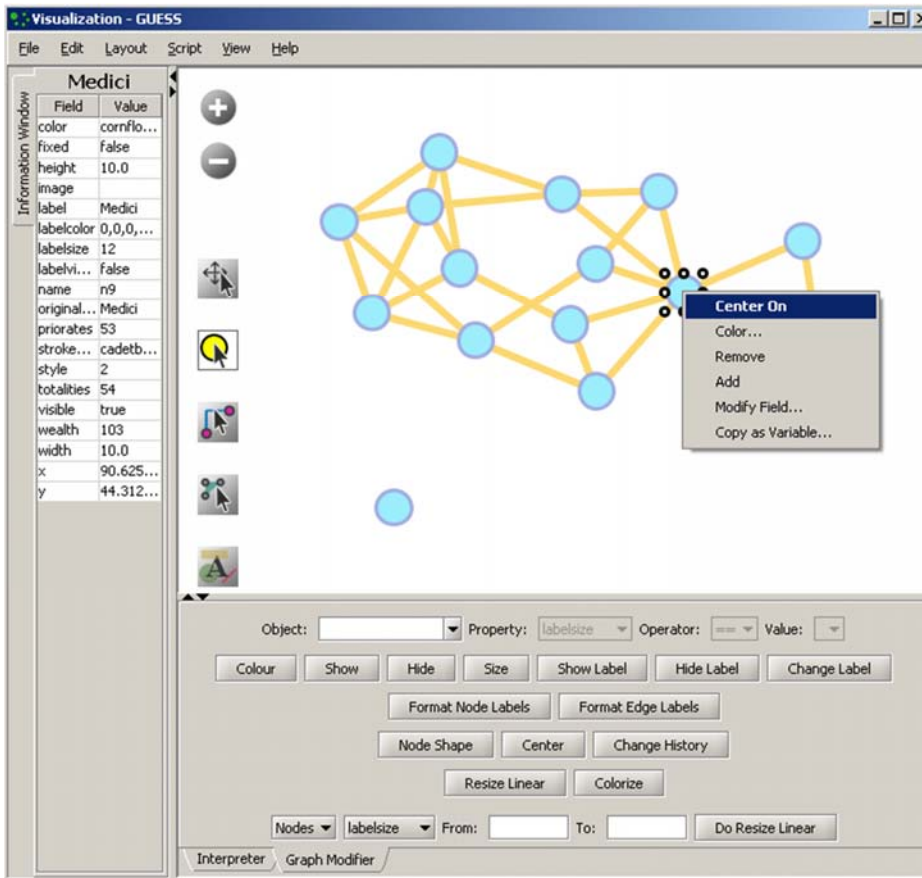
- Select network and run 'Visualization > GUESS' to open GUESS with file loaded.
- Apply 'Layout > GEM'.

35

The screenshot shows the Network Workbench Tool interface. The main window displays a network graph titled 'Medici-Acciaiuoli' with 16 nodes and 27 edges. The nodes are blue circles, and the edges are yellow lines. The graph is undirected and consists of two weakly connected components: a large component of 15 nodes and one isolated node. The interface includes a console window on the left showing GUESS logs, a data manager on the right, and a scheduler at the bottom left. The 'Visualization - GUESS' window is open, showing a table of node attributes for the selected network.


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

36

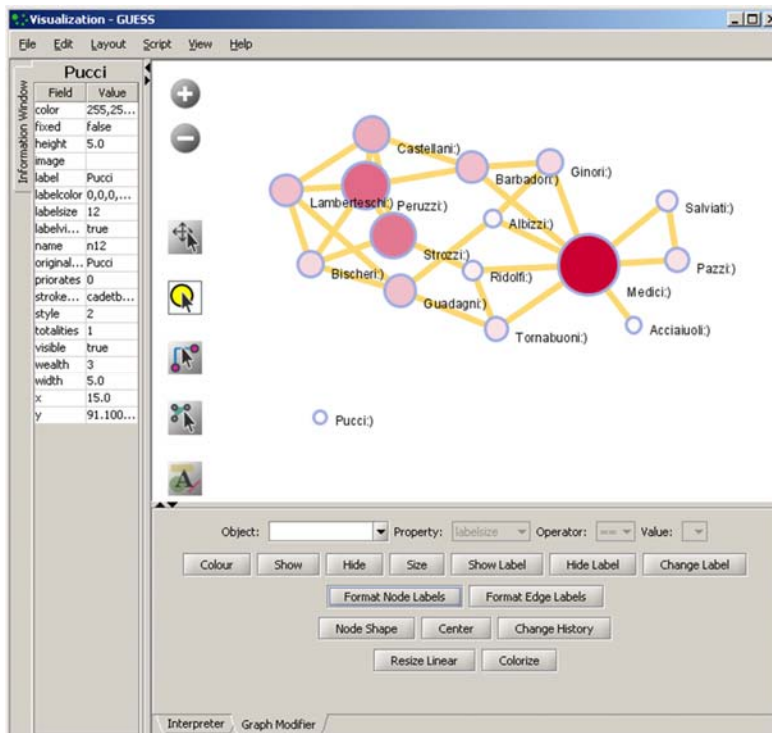


Pan:
 “grab” the background by holding left-click and moving your mouse.

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.



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.’

The screenshot shows the 'Visualization - GUESS' application window. On the left is an 'Information Window' for the 'Acciaiuoli' node, displaying various attributes like color, fixed, height, image, label, labelcolor, labelsize, labelvis, name, original, priorates, stroke, style, totalities, visible, wealth, width, x, and y. The main area displays a network graph with nodes labeled with names like Castellani, Ginori, Salviati, Pazzi, Medici, Acciaiuoli, Tomabuoni, Guadagni, Strozzi, Ridolfi, Pucci, Bischeri, Lamberteschi, Peruzzi, and Barbadori. The nodes are connected by yellow lines. A red tooltip box is overlaid on the graph, containing the text: 'Interpreter: Uses Jython a combination of Java and Python. Try colorize(wealth, white, red)'. At the bottom of the window, there is a code editor with the following commands: `>>> resizeLinear(totalities,5,20)`, `>>> colorize(wealth,white,red)`, and `>>>`. The bottom of the window has tabs for 'Interpreter' and 'Graph Modifier'.

39



Tutorial Overview

9:00am Welcome and Overview of Tutorial and Attendees

9:15am Sci2 Tool Usage and Adoption

10:00am Sci2 Tool Basics

Download and run the Sci2 Tool

Load, analyze, and visualize family and business networks

Horizontal line graph of NSF projects

Studying four major network science researchers

- Load and clean a dataset; process raw data into networks
- Find basic statistics and run various algorithms over the network
- Visualize the network using different layouts

11:00am Break

11:15am Sci2 Tool Novel Functionality

- Yahoo! Geocoder
- Geomap with Gephi network overlay
- Evolving collaboration networks
- R-Bridge
- New Visualizations

12:15pm Outlook and Q&A

12:30pm Adjourn

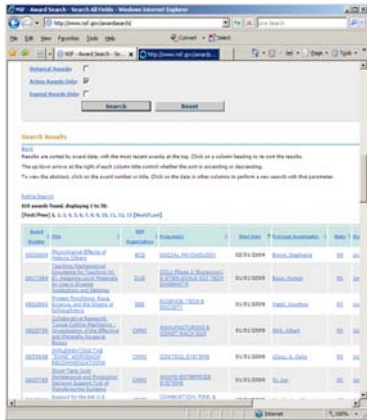
40



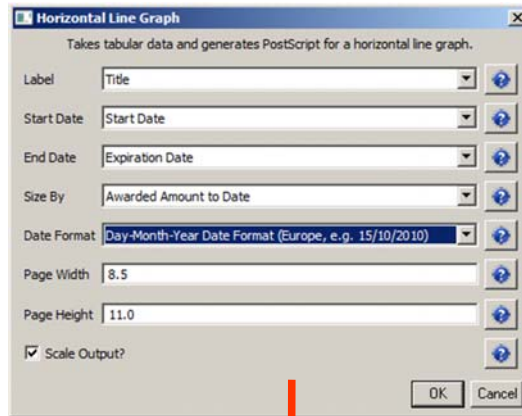
Horizontal line graph of NSF projects

See [5.2.1 Funding Profiles of Three Universities \(NSF Data\)](#)


Download NSF data


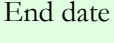


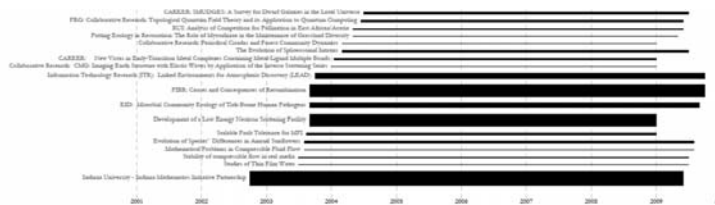
Visualize as Horizontal Line Graph



Area size equals numerical value, e.g., award amount.

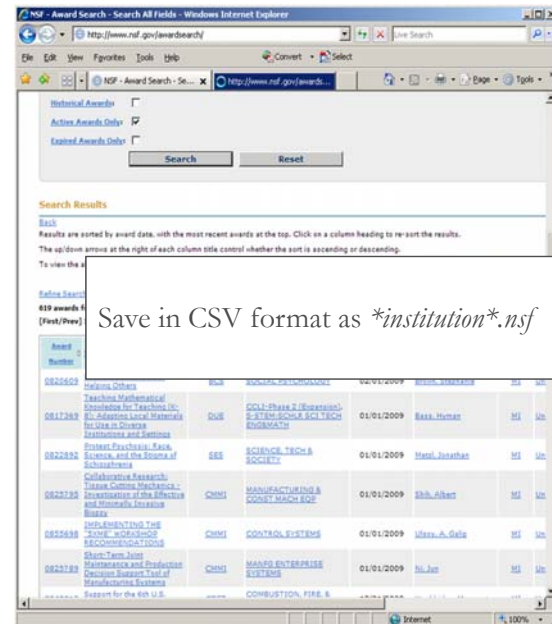
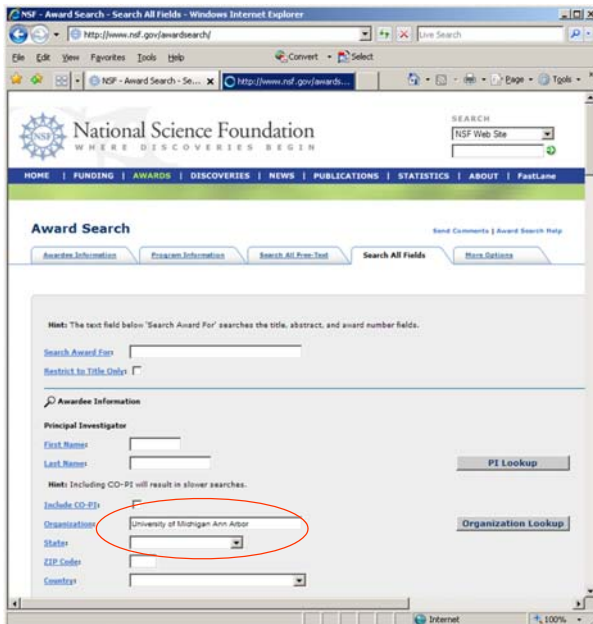
Text 

Start date  End date 



Horizontal line graph of NSF projects

NSF Awards Search via <http://www.nsf.gov/awardsearch>

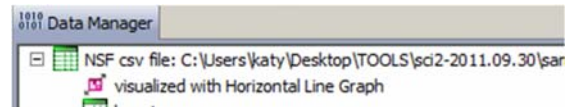
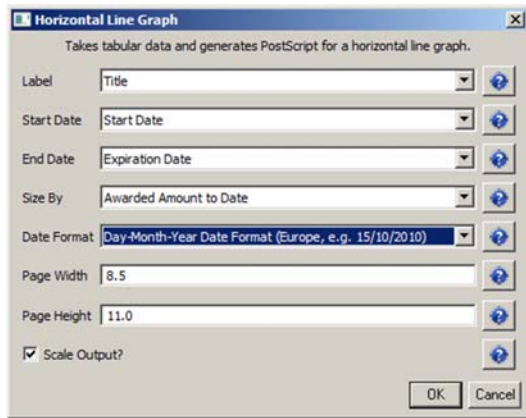




Horizontal line graph of NSF projects

Download and load a dataset of your choice or load one of the sample data files, e.g., *'sampledata/scientometrics/nsf/Michigan.nsf.'*

Run *'Visualization > Temporal > Horizontal Line Graph'* using parameters:



Save *'visualized with Horizontal Line Graph'* as ps or eps file. Convert into pdf and view. Zoom to see details in visualizations of large datasets, e.g., all NSF awards ever made.

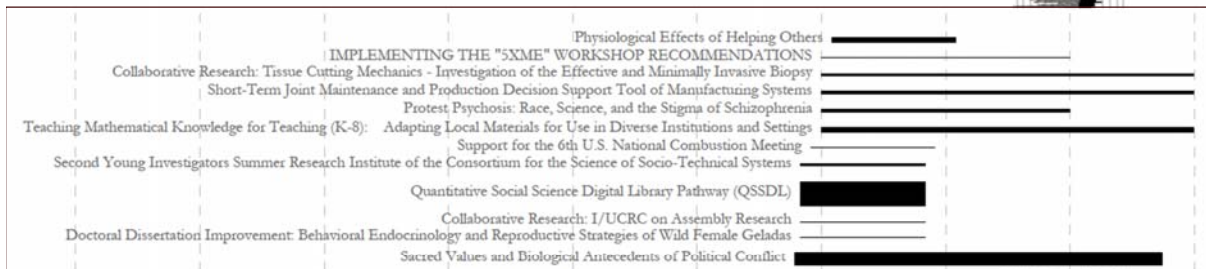
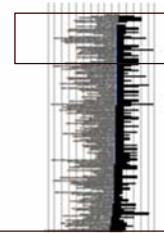


Horizontal line graph of NSF projects

Area size equals numerical value, e.g., award amount.

Text, e.g., title

Start date End date



More NSF data workflows can be found in wiki tutorial:
[5.1.3 Funding Profiles of Three Researchers at Indiana University \(NSF Data\)](#)
[5.2.1 Funding Profiles of Three Universities \(NSF Data\)](#)
[5.2.3 Biomedical Funding Profile of NSF \(NSF Data\)](#)





Tutorial Overview

9:00am Welcome and Overview of Tutorial and Attendees

9:15am Sci2 Tool Usage and Adoption

10:00am Sci2 Tool Basics

Download and run the Sci2 Tool

Load, analyze, and visualize family and business networks

Horizontal line graph of NSF projects

Studying four major network science researchers

- Load and clean a dataset; process raw data into networks
- Find basic statistics and run various algorithms over the network
- Visualize the network using different layouts

11:00am Break

11:15am Sci2 Tool Novel Functionality

- Yahoo! Geocoder
- Geomap with Gephi network overlay
- Evolving collaboration networks
- R-Bridge
- New Visualizations

12:15pm Outlook and Q&A

12:30pm Adjourn

45



Studying Four Major NetSci Researchers (ISI Data) using Database *(section 5.1.4)*

FourNetSciResearchers.isi	
Time frame:	1955-2007
Region(s):	Miscellaneous
Topical Area(s):	Network Science
Analysis Type(s):	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. 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.

[http://sci2.wiki.cns.iu.edu/5.1.4+Studying+Four+Major+NetSci+Researchers+\(ISI+Data\)](http://sci2.wiki.cns.iu.edu/5.1.4+Studying+Four+Major+NetSci+Researchers+(ISI+Data))

46



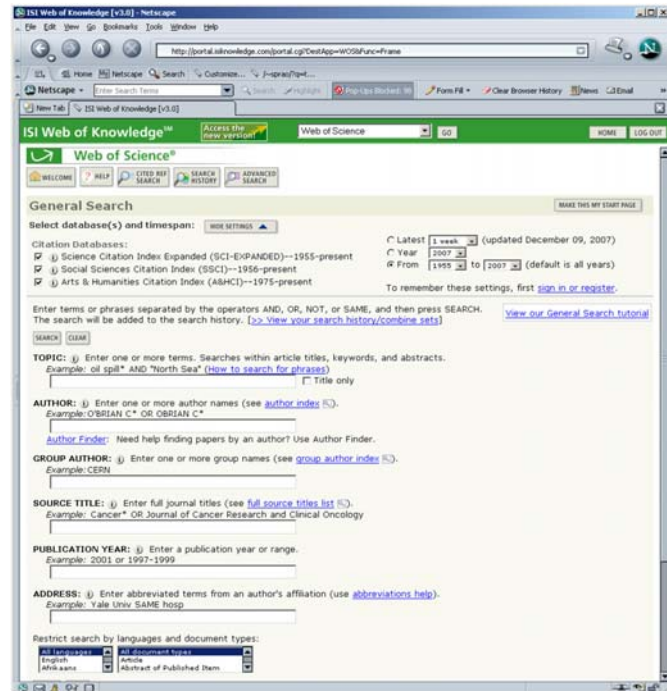
Data Acquisition from Web of Science

In Dec 2007, we downloaded 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



47

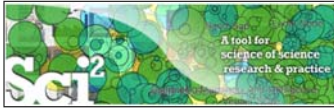


Comparison of Counts

No books and other non-WoS publications are covered.

	Age	Total # Cites	Total # Papers	H-Index
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 (Dec 2007)
	41	16,920	159	52 (Dec 2008)
	44	30,102	201	68 (April 2011)

48

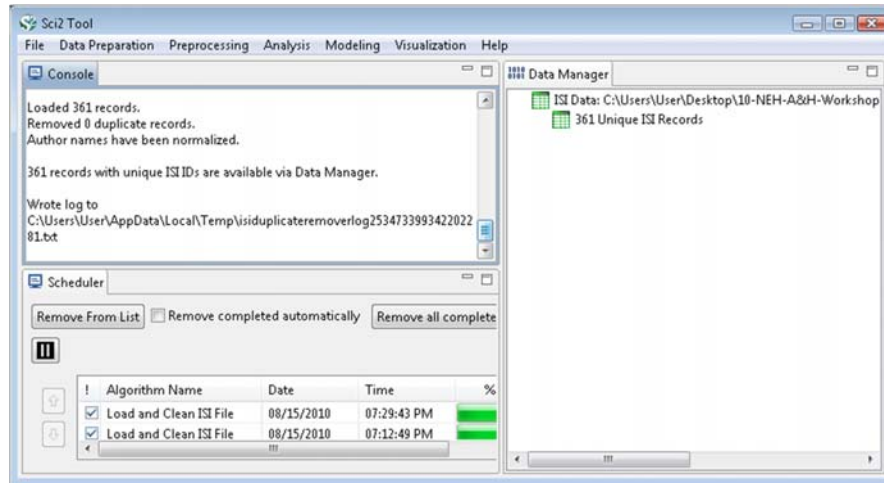


Extract Co-Author Network

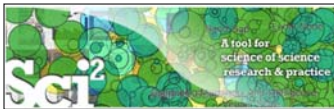
Load **yoursci2directory*/sampledata/scientometrics/isi/FourNetSciResearchers.isi* using *'File > Load ...'*

And file with 361 records appears in the Data Manager.

Duplicates were removed, author names normalized. Log file exists.



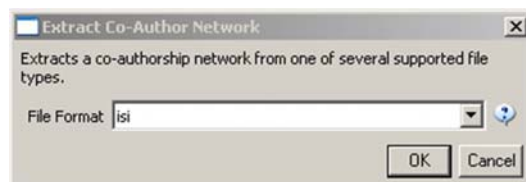
49



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 independent of co-authorship weight.

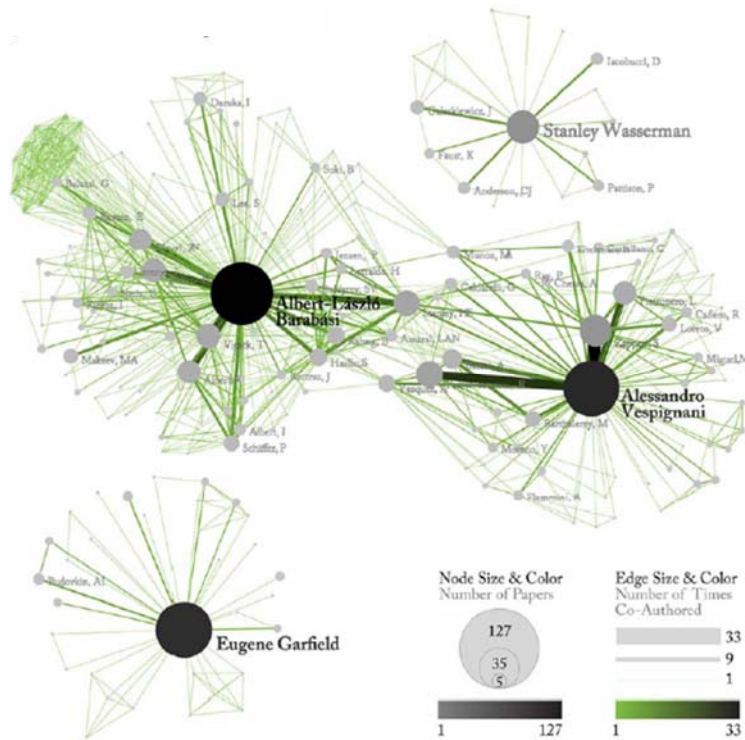
To view the complete network, select the *'Extracted Co-Authorship Network'* and run *'Visualization > Networks > GUESS'*.

Network is loaded with random layout. In GUESS, run *'Layout > GEM'* and *'Layout > Bin Pack'* to improve layout. Run *'Script > Run Script ...'* and select *'yoursci2directory/scripts/GUESS/co-author-nw.py'*.

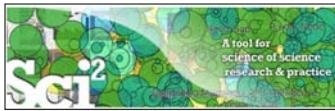
50



Co-Author Network of all Four NetsSci Researchers



51

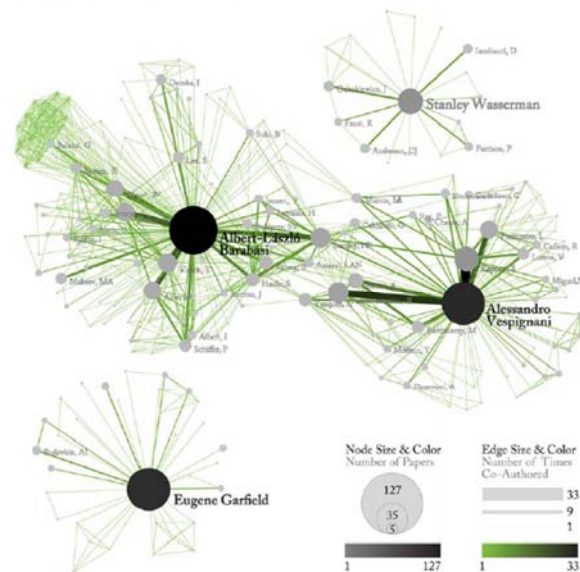


Co-Author Network of all Four NetsSci Researchers

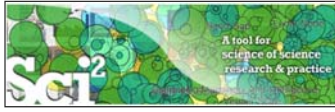
Use the GUESS Graph Modifier to change color and size coding.

Calculate node degrees in Sci2 Tool.

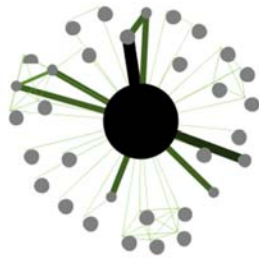
Use a graphic program to add legend.



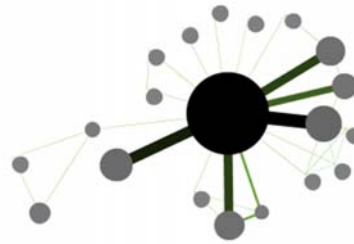
52



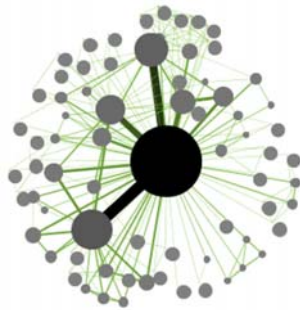
Individual Co-Author Networks (Read/map 4 files separately)



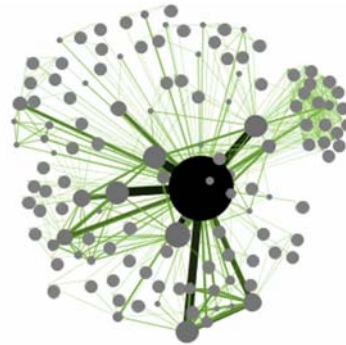
Eugene Garfield



Stanley Wasserman



Alessandro Vespignani



Albert-László Barabási

53



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

.....

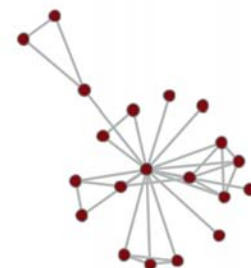
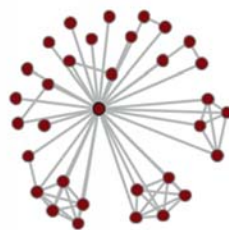
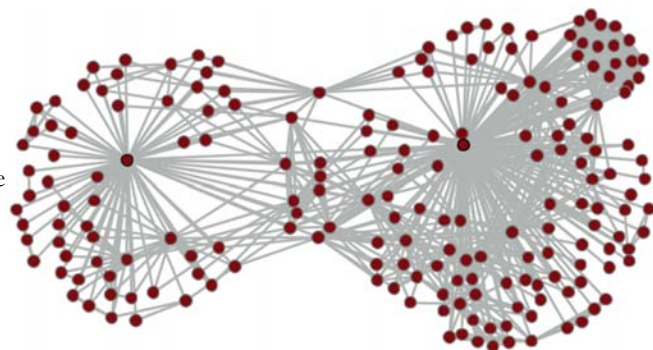
Network Analysis Toolkit (NAT) was selected.

Nodes: 247

Edges: 891

.....

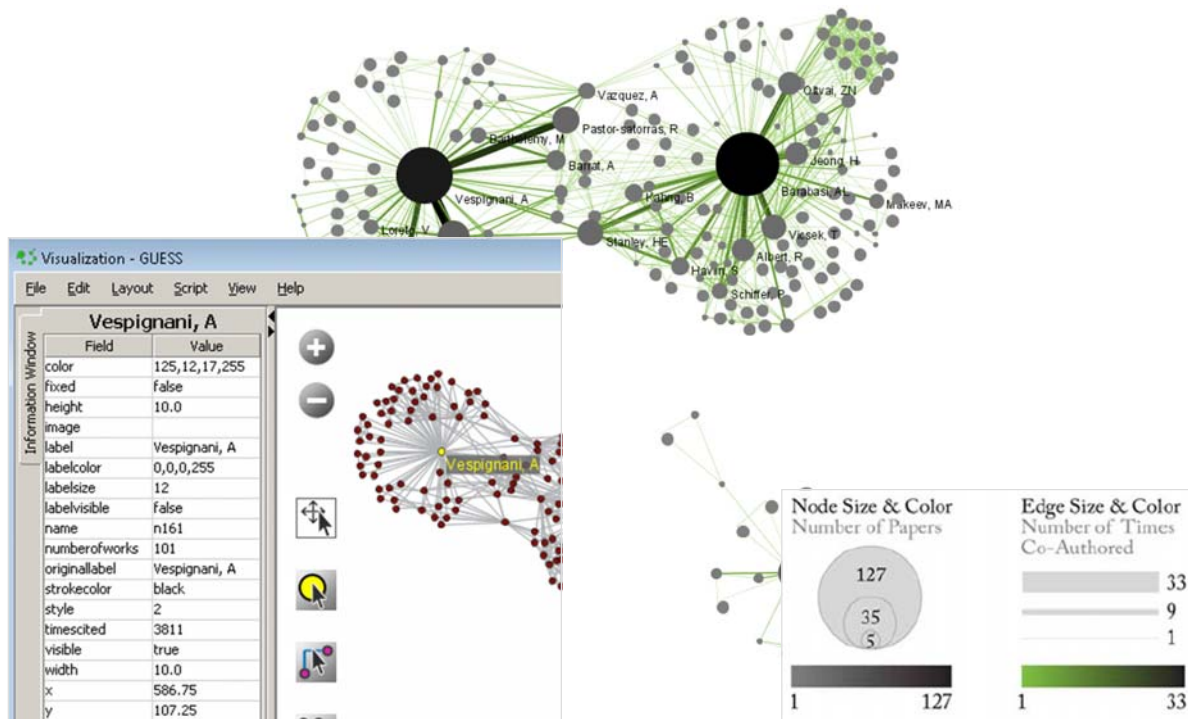
GUESS was selected.



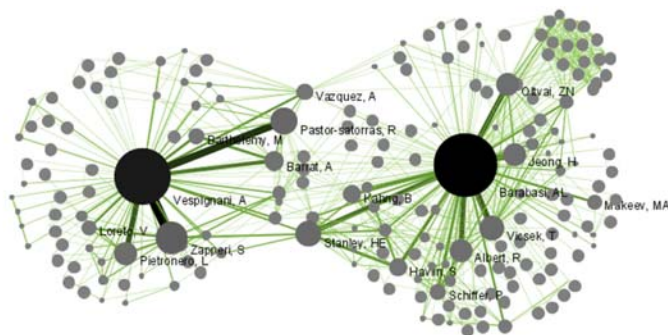
54



Network Visualization: Color/Size Coding by Data Attribute Values

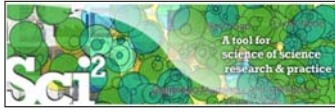


Network Visualization: Giant Component

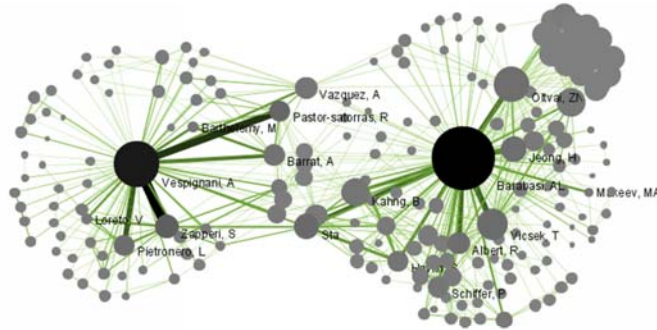


.....
Weak Component Clustering was selected.
Implementer(s): Russell Duhon
Integrator(s): Russell Duhon

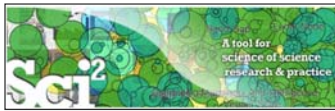
Input Parameters:
Number of top clusters: 10
3 clusters found, generating graphs for the top 3 clusters.
.....



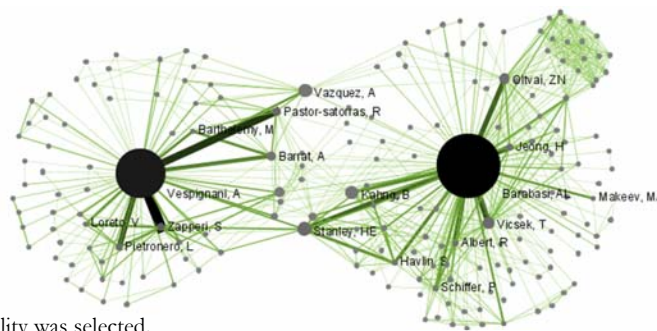
Network Visualization: Color/Size Coding by Degree



.....
Node Degree was selected.
Documentation:
[https://nwb.slis.indiana.edu/community/?n=AnalyzeData.No deDegree](https://nwb.slis.indiana.edu/community/?n=AnalyzeData.No%20deDegree)
.....

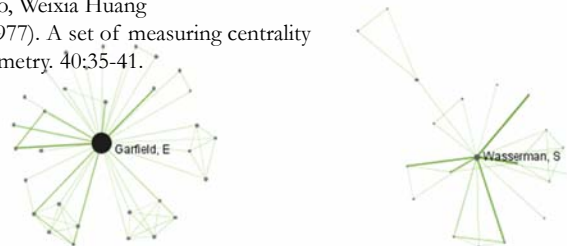


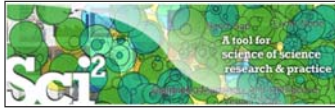
Network Visualization: Color/Size Coding by Betweenness Centrality



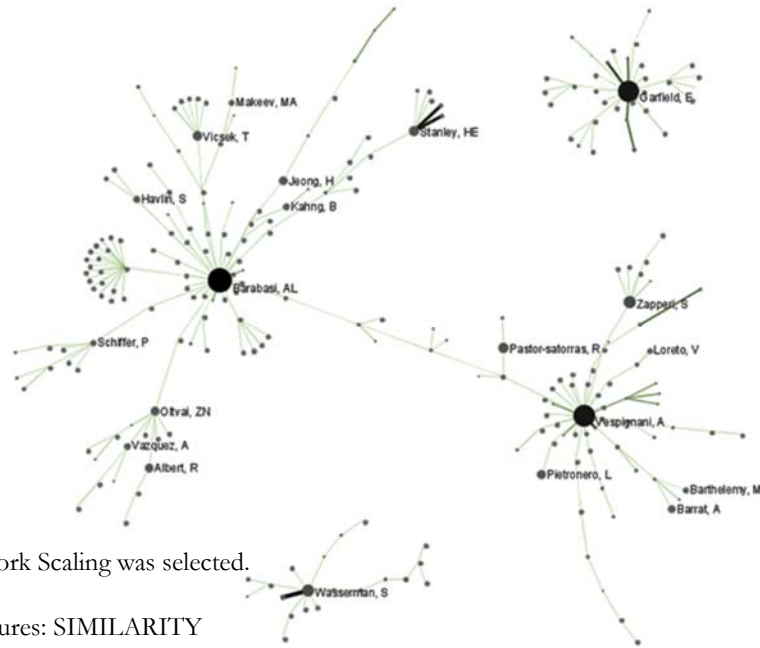
.....
Node Betweenness Centrality was selected.
Author(s): L. C. Freeman
Implementer(s): Santo Fortunato
Integrator(s): Santo Fortunato, Weixia Huang
Reference: Freeman, L. C. (1977). A set of measuring centrality based on betweenness. Sociometry. 40:35-41.

Input Parameters:
Number of bins: 20
Number of bins: 10
.....





Network Visualization: Reduced Network After Pathfinder Network Scaling



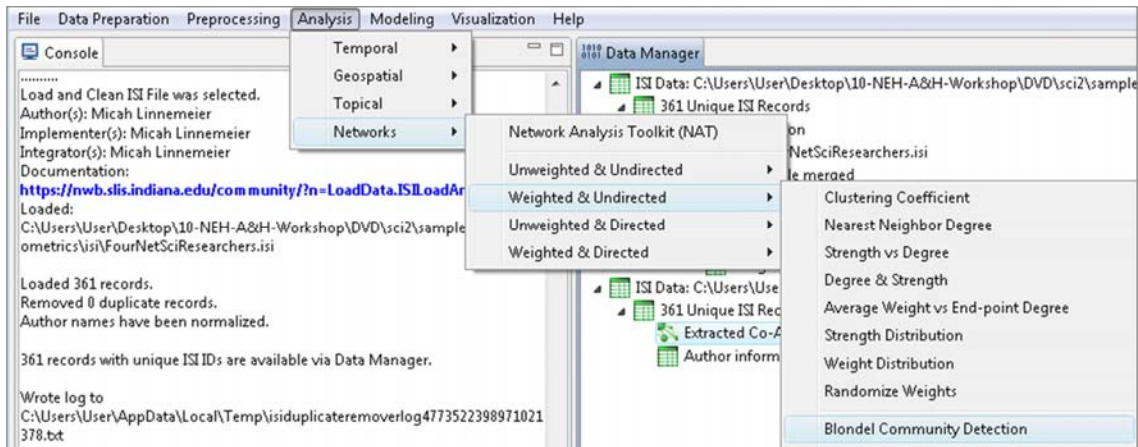
.....
MST-Pathfinder Network Scaling was selected.
Input Parameters:
Weight Attribute measures: SIMILARITY
Edge Weight Attribute: weight
.....

59

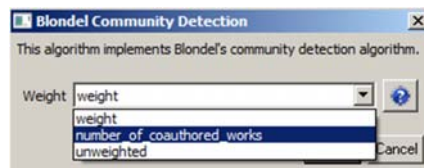


Network Visualization: Circular Hierarchy Visualization

Select Co-Author Network and run Blondel Community detection:



With parameter values

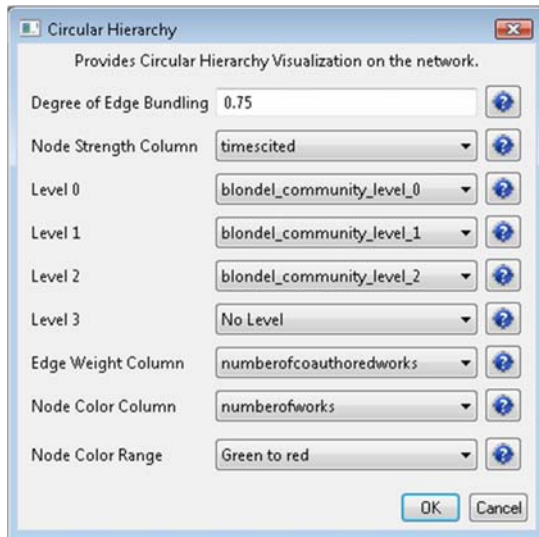


60



Network Visualization: Circular Hierarchy Visualization

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



61



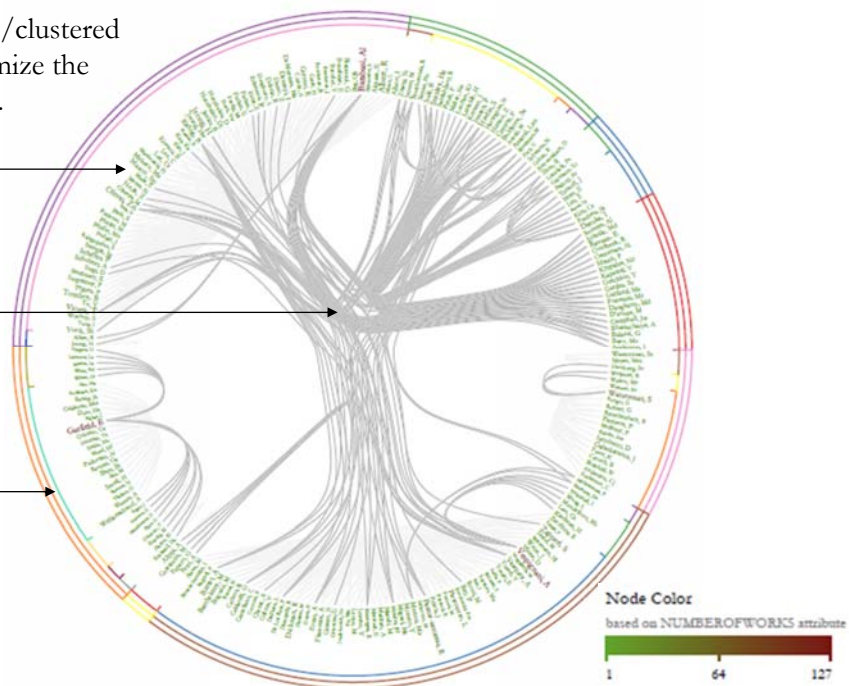
Network Visualization: Circular Hierarchy Visualization

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.

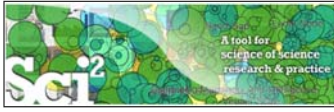
Color coded cluster hierarchy according to Blondel community detection algorithm.



Note:

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

62



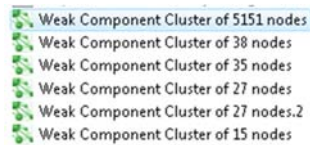
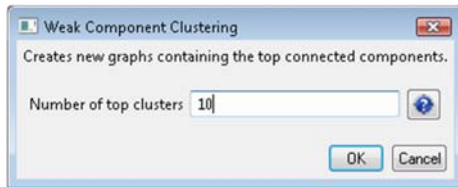
Paper-Citation Network Layout

To extract the paper-citation network, select the '361 Unique ISI Records' table and run *Data Preparation > 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)

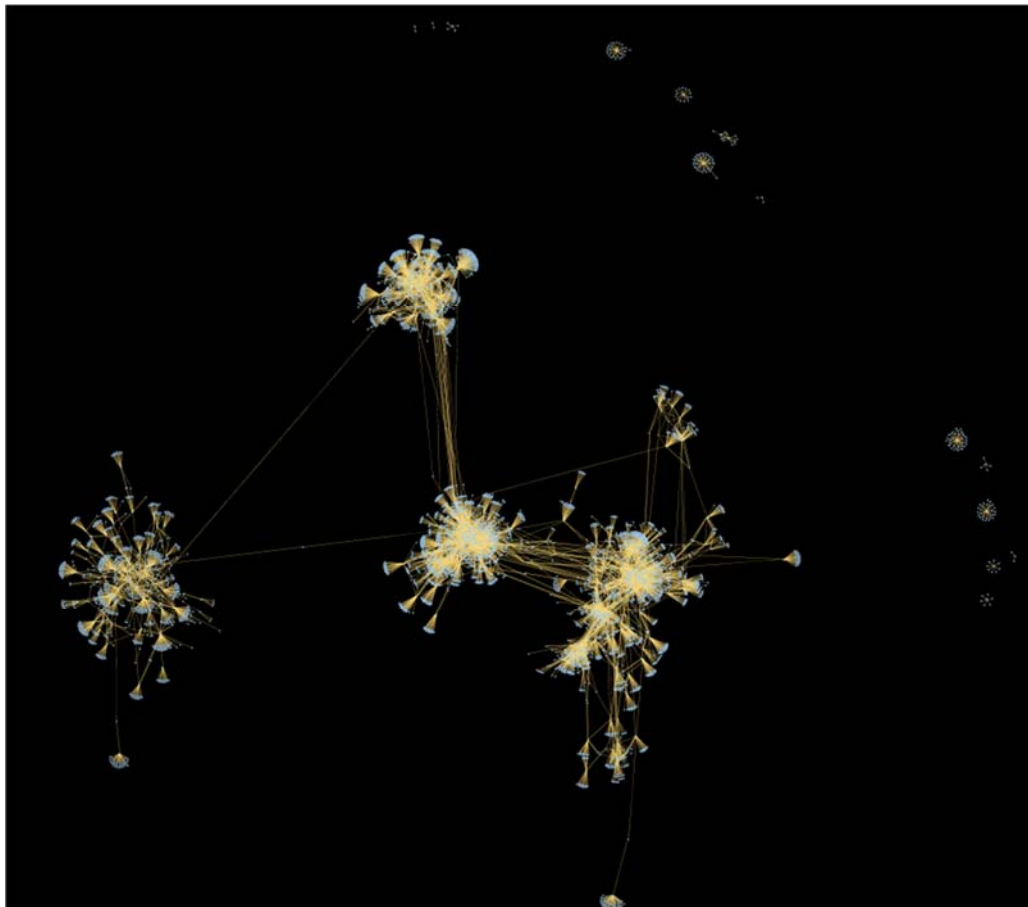
Run *'Analysis > Networks > Unweighted and Directed > Weak Component Clustering'* with parameters



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'*.

63



64



General Network Extraction: Weighted, Undirected Co-Occurrence Network

	A	B	C	D
1	Paper	Authors	References	Year
2	P1	A1		1970
3	P2	A2;A6	P1	1980
4	P3	A1;A3	P1;P2	1990
5	P4	A1;A4;A5	P2	1995
6	P5	A5;A6	P1;P2;P3;P4	1995
7	P6	A2;A6	P5	2000

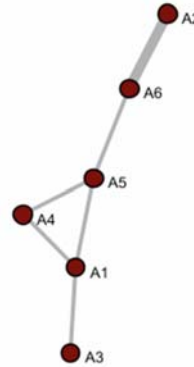


Author co-occurrence network

Extract Network from Table
Extracts a network from a delimited table

Column Name:

Text Delimiter:



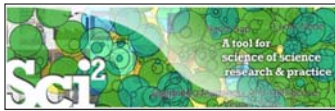
*Vertices 6

- 1 A1
- 2 A6
- 3 A2
- 4 A3
- 5 A5
- 6 A4

*Edges 6

- 2 3 2
- 1 4 1
- 1 5 1
- 5 6 1
- 1 6 1
- 2 5 1

65



General Network Extraction: Unweighted, Directed Bipartite Network

	A	B	C	D
1	Paper	Authors	References	Year
2	P1	A1		1970
3	P2	A2;A6	P1	1980
4	P3	A1;A3	P1;P2	1990
5	P4	A1;A4;A5	P2	1995
6	P5	A5;A6	P1;P2;P3;P4	1995
7	P6	A2;A6	P5	2000



Paper-author bipartite (2-mode) network

Extract Bipartite Network
Extract a bipartite network from two columns in the table. If the column values may list multiple entries, enter the special text which delimits them.

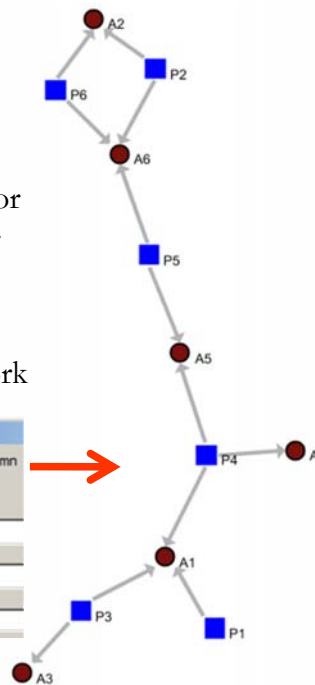
First column:

Second column:

Text Delimiter:

Object: nodes based on -> Property: bipartitetype Operator: == Value: Paper

● Author
■ Paper



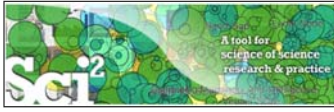
*Vertices 12

- 1 P1 bipartitetype "Paper"
- 2 A1 bipartitetype "Authors"
- 3 P2 bipartitetype "Paper"
- 4 A2 bipartitetype "Authors"
- 5 A6 bipartitetype "Authors"
- 6 P3 bipartitetype "Paper"
- 7 A3 bipartitetype "Authors"
- 8 P4 bipartitetype "Paper"
- 9 A4 bipartitetype "Authors"
- 10 A5 bipartitetype "Authors"
- 11 P5 bipartitetype "Paper"
- 12 P6 bipartitetype "Paper"

*Arcs

- 1 2
- 3 4
- 3 5
- 6 2
- 6 7
- 8 2
- 8 10
- 8 9
- 11 5
- 11 10
- 12 4
- 12 5

66



General Network Extraction: Unweighted, Directed Network

	A	B	C	D
1	Paper	Authors	References	Year
2	P1	A1		1970
3	P2	A2;A6	P1	1980
4	P3	A1;A3	P1;P2	1990
5	P4	A1;A4;A5	P2	1995
6	P5	A5;A6	P1;P2;P3;P4	1995
7	P6	A2;A6	P5	2000

Extract Directed Network

Given a table, this algorithm extracts a directed edge that starts at a column node.

Source Column: Paper

Target Column: Authors

Text Delimiter: ;

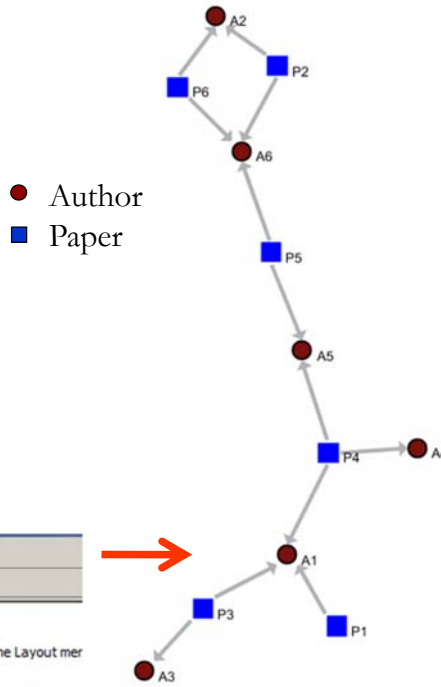
Analysis Modeling Visualization R Help

Temporal Geospatial Topical

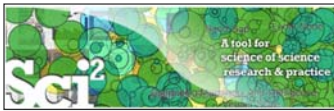
Networks

- Network Analysis Toolkit (NAT)
- Unweighted & Undirected
- Weighted & Undirected
- Unweighted & Directed**

Node Indegree



- *Vertices 12
- 1 P1 indegree 0
 - 2 A1 indegree 3
 - 3 P2 indegree 0
 - 4 A2 indegree 2
 - 5 A6 indegree 3**
 - 6 P3 indegree 0
 - 7 A3 indegree 1
 - 8 P4 indegree 0
 - 9 A4 indegree 1
 - 10 A5 indegree 2
 - 11 P5 indegree 0
 - 12 P6 indegree 0
- *Arcs
- 1 2
 - 3 4
 - 3 5
 - 6 2
 - 6 7
 - 8 10
 - 8 2
 - 8 9
 - 11 10
 - 11 5
 - 12 4
 - 12 5



General Network Extraction: Unweighted, Directed Paper-Citation Network

	A	B	C	D
1	Paper	Authors	References	Year
2	P1	A1		1970
3	P2	A2;A6	P1	1980
4	P3	A1;A3	P1;P2	1990
5	P4	A1;A4;A5	P2	1995
6	P5	A5;A6	P1;P2;P3;P4	1995
7	P6	A2;A6	P5	2000

Extract Directed Network

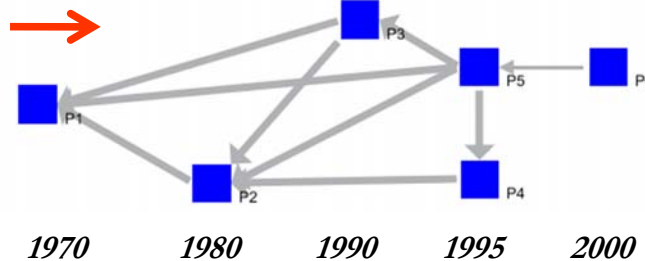
Given a table, this algorithm extracts a directed edge that starts at a column node.

Source Column: Paper

Target Column: References

Text Delimiter: ;

Arcs from papers to references



- *Vertices 6
- 1 P1
 - 2 P2
 - 3 P3
 - 4 P4
 - 5 P5
 - 6 P6
- *Arcs
- 2 1
 - 3 1
 - 3 2
 - 4 2
 - 5 4
 - 5 3
 - 5 1
 - 5 2
 - 6 5



General Network Extraction: Unweighted, Directed Bi-Partite Network

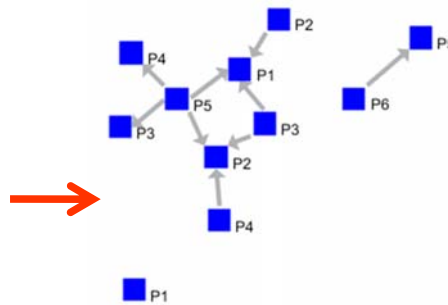
	A	B	C	D
1	Paper	Authors	References	Year
2	P1	A1		1970
3	P2	A2;A6	P1	1980
4	P3	A1;A3	P1;P2	1990
5	P4	A1;A4;A5	P2	1995
6	P5	A5;A6	P1;P2;P3;P4	1995
7	P6	A2;A6	P5	2000

Extract Bipartite Network
Extract a bipartite network values may list multiple ent them.

First column:

Second column:

Text Delimiter:



WRONG!!!

- *Vertices 11
- 1 P1 bipartitetype "Paper"
- 2 P2 bipartitetype "Paper"
- 3 P1 bipartitetype "References"
- 4 P3 bipartitetype "Paper"
- 5 P2 bipartitetype "References"
- 6 P4 bipartitetype "Paper"
- 7 P5 bipartitetype "Paper"
- 8 P4 bipartitetype "References"
- 9 P3 bipartitetype "References"
- 10 P6 bipartitetype "Paper"
- 11 P5 bipartitetype "References"
- *Arcs
- 2 3
- 4 3
- 4 5
- 6 5
- 7 3
- 7 9
- 7 5
- 7 8
- 10 11



ISI Paper-Citation Network Extraction

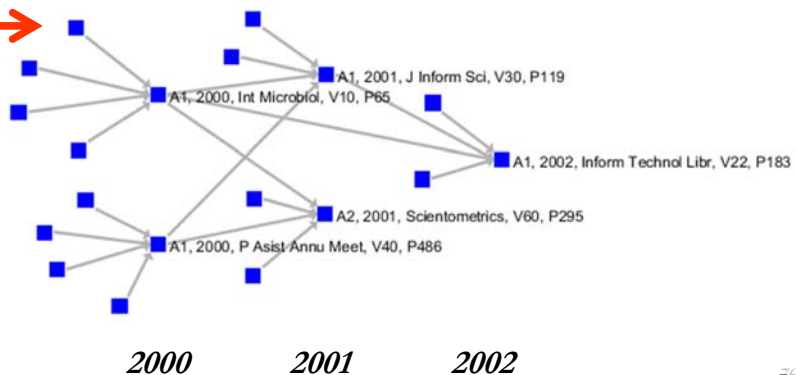
	A	B	C	D	E
1	Authors	Cited References	Publication Year	Title	Cite Me As
2	A1 A2	BENSMAN SJ, 1998, LIBR RESOUR TECH SER, V42, P147 BRO	2000	T1	A1, 2000, INT MICROBIOL, V10, P65
3	A1	BENSMAN SJ, 1999, LIBR RESOUR TECH SER, V42, P147 BRO	2000	T2	A1, 2000, P ASIST ANNU MEET, V40, P486
4	A2 A3	GARFIELD E, 1985, ESSAYS INFORMATION S, V8, P403 GILBE	2001	T3	A2, 2001, SCIENTOMETRICS, V60, P295
5	A1	ASIMOV A, 1963, GENETIC CODE LEDERBERG J, 1972, NATU	2001	T4	A1, 2001, J INFORM SCI, V30, P119
6	A1 A2	AVERY OT, 1944, J EXP MED, V79, P137 SMALL H, 1985, J INF	2002	T5	A1, 2002, INFORM TECHNOL LIBR, V22, P183

Sci2 Tool

File | Data Preparation | Preprocessing | Analysis

- Remove ISI Duplicate Records
- Remove Rows with Multitudinous Fields
- Extract Directed Network
- Extract Bipartite Network
- Extract Paper Citation Network**
- Extract Author Paper Network

*Arcs from references to papers—
in the direction of information flow*



Q&A

71



Tutorial Overview

9:00am Welcome and Overview of Tutorial and Attendees

9:15am Sci2 Tool Usage and Adoption

10:00am Sci2 Tool Basics

Download and run the Sci2 Tool

Load, analyze, and visualize family and business networks

Horizontal line graph of NSF projects

Studying four major network science researchers

- Load and clean a dataset; process raw data into networks
- Find basic statistics and run various algorithms over the network
- Visualize the network using different layouts

11:00am Break

11:15am Sci2 Tool Novel Functionality

- **Yahoo! Geocoder**
- Geomap with Gephi network overlay
- Evolving collaboration networks
- R-Bridge
- New Visualizations

12:15pm Outlook and Q&A

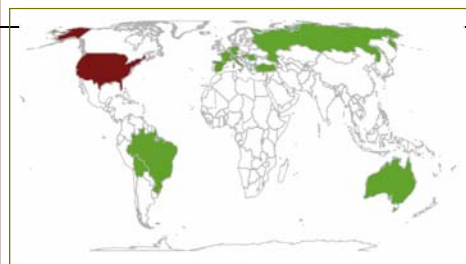
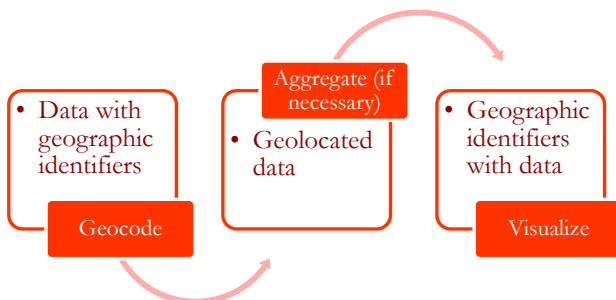
12:30pm Adjourn

72

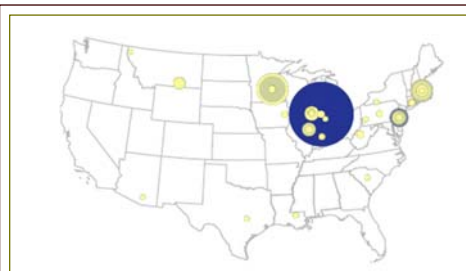


Geocoding and Geospatial Maps

<http://wiki.cns.in.edu/display/CISHELL/Yahoo+Geocoder>



Region names + numeric data
(Choropleth Map)



Geocoordinates + numeric data
(Proportional Symbol Map)

73



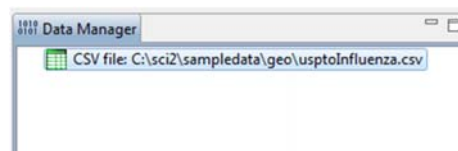
Load File with Address and Times Cited Fields

Run File > Load...

Load the sample data table `sampledata/geo/usptoInfluenza.csv`

Let's create a map showing influenza-related patent activity in the following countries.

	A	B	C	D	E
1	Country	Latitude	Longitude	Patents	Times Cited
2	Hungary	47.16116	19.504959	0.083333333	4
3	Belgium	50.500992	4.47677	3.017857143	11
4	Germany	51.090839	10.45424	4.783333333	4
5	Canada	62.35873	-96.582092	5.539285714	21
6	Russia	59.461479	108.831779	0.266666667	2
7	Austria	47.69651	13.34577	4.2	17
8	Netherlands	52.108089	5.33033	1	2
9	Switzerland	46.813091	8.22414	0.507575758	6
10	Taiwan	23.599751	121.023811	2	3
11	Australia	-24.916201	133.393112	1.617857143	23
12	United States	39.83	-98.58	73.9983889	220
13	France	46.712448	1.71832	2.201165501	9
14	South Africa	-28.483219	24.676991	0.333333333	1
15	Japan	37.487598	139.838287	15.99166667	39
16	Israel	31.389299	35.36124	3.5	3
17	United Kingdom	54.313919	-2.23218	3.85	12



74



Yahoo Geocoder



Edit Add Tools

Added by [Aretha Alencar](#), last edited by [Kavin Kumar Palanisamy](#) on Feb 08, 2012 ([view change](#))

Description

This algorithm converts place names or addresses into Latitude, Longitude co-ordinates. It accepts international addresses, countries, States of United States of America and ZIP codes of United States of America. All co-ordinates are obtained by querying Yahoo! PlaceFinder service. Internet access must be available during geocoding.

Pros & Cons

1. The performance is slower than the [Geocoder](#) and may vary due to the network latency since the queries are requested through internet service. The benchmark test geocoded 470 unique locations per minute
2. Yahoo! Geocoder supports address geocoding with international coverage which is not supported by [Geocoder](#).
3. To use Yahoo! Geocoder, user has to obtain an application id through [Yahoo! registration](#). Save your application id and provide it when requested by the Yahoo! Geocoder. Since each application id is allowed to geocode 50,000 locations per 24 hours, the user is encouraged to test on a small set of data first.

Applications

The plugin is useful for scientists who would like to visualize their data on a geographical map ([geomap](#)). User can obtain the geographical coordinates (Latitude and Longitude values) and feed them to the visualization plugin.

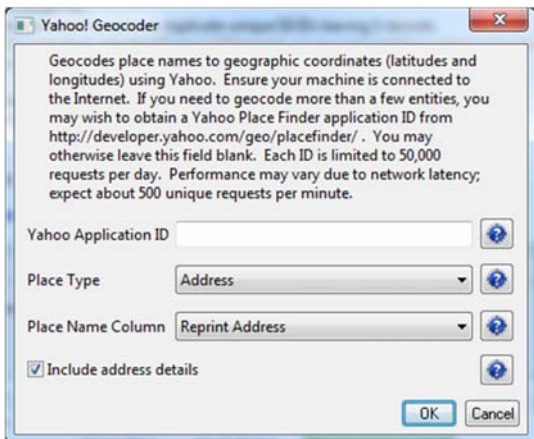
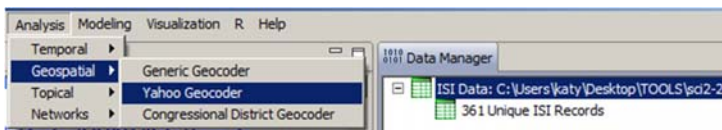
<http://wiki.cns.iu.edu/display/CISHELL/Yahoo+Geocoder>

75



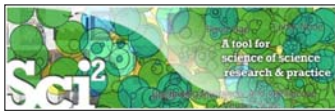
Using Yahoo! Geocoder

Run Analysis > Geospatial > Yahoo Geocoder

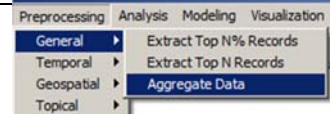


You can leave Application ID blank for trial purposes, but before heavy use, register later for your own personal Yahoo! Application ID, see: <http://developer.yahoo.com/geo/placefinder/>

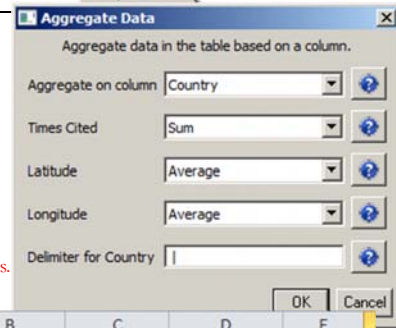
76



Aggregate by Country



Aggregate Data was selected.
 Implementer(s): Chintan Tank
 Documentation: <http://wiki.cns.iu.edu/display/CISHELL/Aggregate+Data>
 Input Parameters:
 Aggregate on column: Country
 Delimiter for Country: |
 Longitude: AVERAGE
 Latitude: AVERAGE
 Times Cited: SUM



Aggregated by ": All rows of Latitude column were skipped due to no non-null, non-empty values.
 Aggregated by ": All rows of Longitude column were skipped due to no non-null, non-empty values.
 Frequency of unique "Country" values added to "Count" column.

	A	B	C	D
1	Times Cited	Latitude	Longitude	Country
2	7	42.02946091	-87.68838501	United States
3	0			
4	0			
5	2	42.34999466	-71.08765411	United States
6	14	41.70074844	-86.23918915	United States
7	15	41.70074844	-86.23918915	United States
8	29	41.89422607	-87.61901855	United States
9	32	41.70074844	-86.23918915	United States
10	7	41.70074844	-86.23918915	United States
11	5	41.70074844	-86.23918915	United States
12	2	41.11500168	-85.73377991	United States
13	10	47.50622177	19.06481934	Hungary
14	44	41.70074844	-86.23918915	United States
15	0	47.50622559	19.06481934	Hungary
16	19	41.70074844	-86.23918915	United States

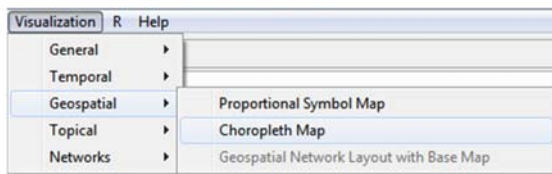
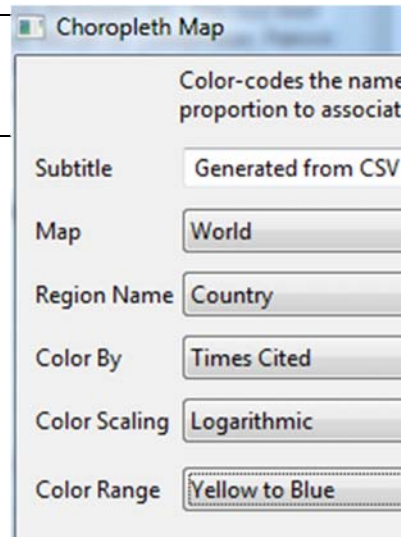


	A	B	C	D	E
1	Times Cited	Latitude	Longitude	Country	Count
2	14680	[41.10645f]	[-82.45309f]	United States	194
3	1802				57
4	398	[47.506226f]	[19.06482f]	Hungary	14
5	101	[37.25198f]	[127.08451f]	South Korea	4
6	18	[32.08439f]	[34.81297f]	Israel	1
7	57	[46.768517f]	[23.585135f]	Romania	2
8	55	[47.06615f]	[7.2015657f]	Switzerland	2
9	455	[47.977184f]	[2.2232702f]	France	12
10	92	[52.15457f]	[4.49463f]	Netherlands	5
11	21	[49.944717f]	[84.528114f]	Russia	2
12	1112	[41.545982f]	[1.7138832f]	Spain	13
13	1381	[43.352654f]	[12.727126f]	Italy	46
14	188	[-22.494667f]	[-45.4818f]	Brazil	3
15	56	[51.24459f]	[10.360385f]	Germany	2
16	0	[-16.49901f]	[-68.14626f]	Bolivia	1

77

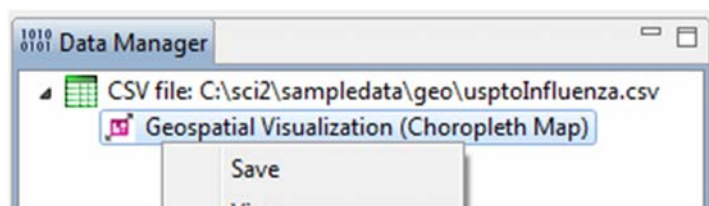


Choropleth Map



Right-click and **Save** map

Open this PostScript file to visualize





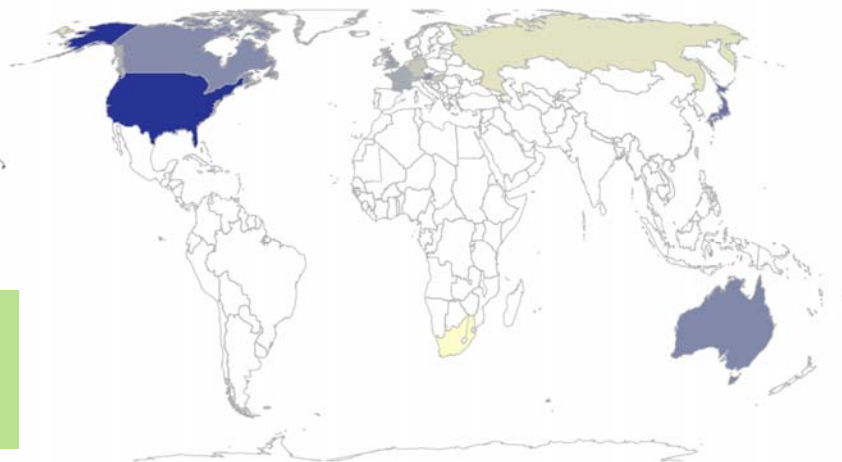
Interpreting the Choropleth Map

Geospatial Visualization (Choropleth Map)

Generated from CSV file: C:\sci2\sampladata\geo\usptoInfluenza.csv
Jun 14, 2012 | 05:33:37 PM EDT

Header shows visualization type, data description, and creation date

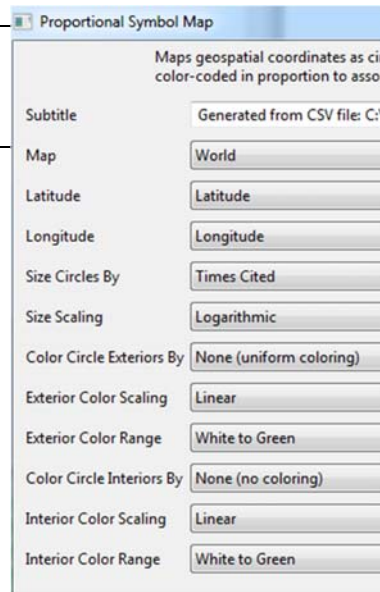
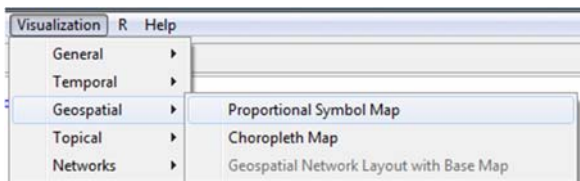
Legend shows how data matches up with visual representation



How to Read this Map
This choropleth map shows 209 countries of the world using the equal-area Eckert IV projection. Each country may be color coded in proportion to a numerical value. Minimum and maximum data values are given in the legend.

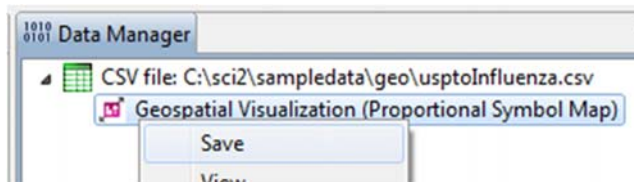


Proportional Symbol Map



Right-click and **Save** map

Open this PostScript file to visualize



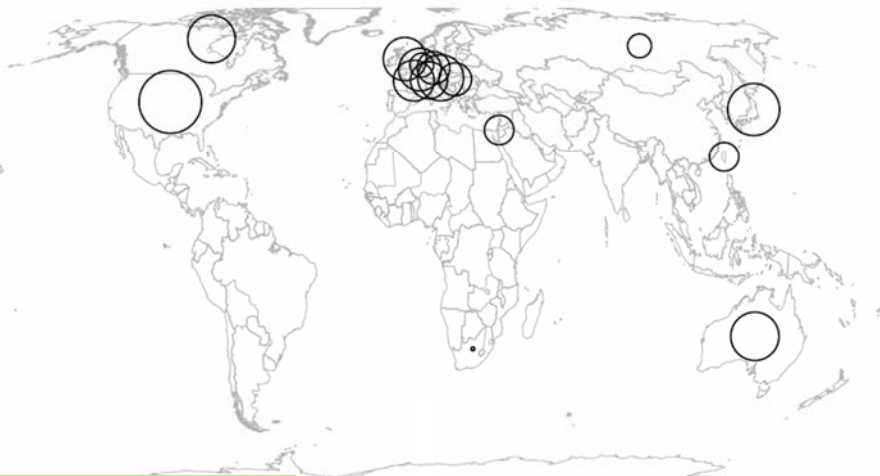


Interpreting the Proportional Symbol Map

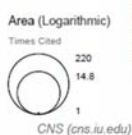
Geospatial Visualization (Proportional Symbol Map)

Generated from CSV file: C:\sci2\sampladata\geol\usptoinfluenza.csv
Jun 14, 2012 | 05:56:39 PM EDT

Header shows visualization type, data description, and creation date



Legend shows how data matches up with visual representation



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.



Relevant Sci2 Manual entry



5.2.4 Mapping Scientometrics (ISI Data)

9 Added by Ted Polley, last edited by Ted Polley on Nov 14, 2011 (view change)

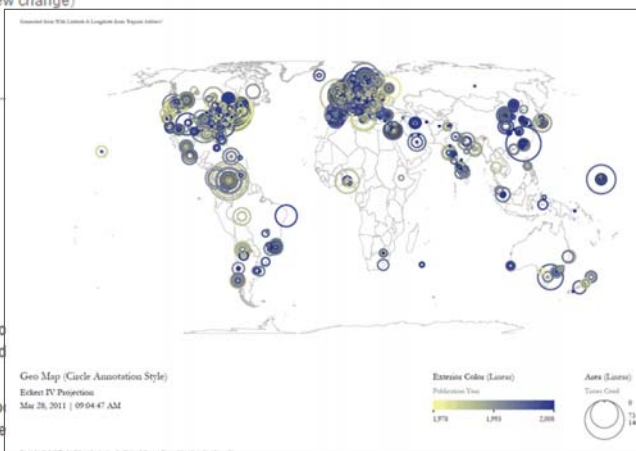
Edit Add Tools

5.2.4.1 Document Co-Citation

Scientometrics.isi	
Time frame:	1978-2008
Region(s):	Miscellaneous
Topical Area(s):	Scientometrics
Analysis Type(s):	Document Co-Citation Network

Scientometrics is a discipline which uses statistical and computational science. Here we use ISI data from the journal "Scientometrics" and Awards Search.

Download [Scientometrics.isi](#). Load the file using 'File > Load' and load document co-citation analysis, as the scale is large enough that the similarity within the domain of scientometrics.



New ISI File Format

Web of Science made a change to their output format in September, 2011. Older versions of Sci2 tool may refuse to load these new files, with an error like "Invalid ISI format file selected."



Tutorial Overview

9:00am Welcome and Overview of Tutorial and Attendees

9:15am Sci2 Tool Usage and Adoption

10:00am Sci2 Tool Basics

Download and run the Sci2 Tool

Load, analyze, and visualize family and business networks

Horizontal line graph of NSF projects

Studying four major network science researchers

- Load and clean a dataset; process raw data into networks
- Find basic statistics and run various algorithms over the network
- Visualize the network using different layouts

11:00am Break

11:15am Sci2 Tool Novel Functionality

- Yahoo! Geocoder
- **Geomap with Gephi network overlay**
- Evolving collaboration networks
- R-Bridge
- New Visualizations

12:15pm Outlook and Q&A

12:30pm Adjourn

83



Geomap with Gephi Network Overlay

See 4.7.6 on <http://sci2.wiki.cns.in.edu>

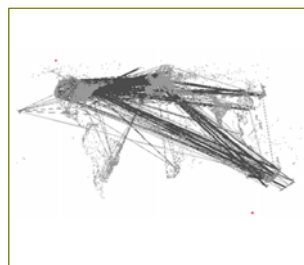
File with geolocations and linkage info, e.g., an isi bibliography file. → Use Yahoo! Geocoder to identify Latitude, Longitude for each geolocation
 Extract attributes per geolocation, e.g., total times cited (TC)
 Extract linkages and their attributes, e.g., number of co-occurrences
 See sample /geo/LaszloBarabasiGeo.net with co-occurrence of “Research Addresses” and full counting of TC per geolocation.



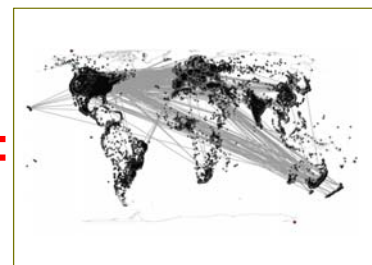
Read into Sci2 Tool to generate geomap and network file → Layout network in Gephi → Combine geomap and network in Photoshop



+



=



84



Relevant Sci2 Manual entry

Dashboard > Sci2 Manual > ... > 4 Workflow Design > 4.7 Geospatial Analysis (Where) Browse Log In Search Confluence

Search

- Home
- 1 Introduction
- 2 Getting Started
- 3 Algorithms, Tools, and Plugins
- 4 Workflow Design
 - 4.1 Overview
 - 4.2 Data Acquisition and Preparation
 - 4.3 Database Loading and Manipulation
 - 4.4 Summaries and Table Extractions
 - 4.5 Statistical Analysis and Profiling
 - 4.6 Temporal Analysis (When)
 - 4.7 Geospatial Analysis (Where)**
 - 4.8 Topical Analysis (What)
 - 4.9 Network Analysis (With Whom?)
 - 4.10 Modeling (Why?)
- 5 Sample Workflows
- 6 Sample Science Studies & Online Services
- 7 Extending the Sci2 Tool
- 8 Relevant Datasets and Tools
- 9 References

4.7.6 Using Gephi to Render Networks Overlaid on Geo Maps

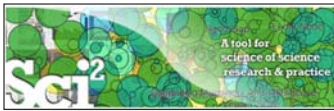
Loading and Saving Geovisualization Files in Sci2

This algorithm allows for the geospatial visualization of network data. The algorithm produces a network file and corresponding blank map. Gephi is used to edit the network produced by Sci2. Once the network has been edited in Gephi it can be exported in a format that will allow it to be overlaid on the map, facilitating visualization of the geospatial data. The following is a brief workflow explaining the process, beginning to end.

1. Load [this](#) network in Sci2.
2. Once the network had been loaded in Sci2 run "Visualization > Geospatial > Geo Map (network template, not fully rendered)" and set the following parameters:

<http://sci2.wiki.cns.iu.edu/display/SCI2TUTORIAL/4.7+Geospatial+Analysis+%28Where%29#4.7GeospatialAnalysis%28Where%29-4.7.6UsingGephitoRenderNetworksOverlaidonGeoMaps>

85



Use Sci2 Tool to Generate Geomap and Network File

Read prepared .net file and run:

Sci2 Tool

File Data Preparation Preprocessing Analysis Modeling Visualization Help

Console

Please visit <https://sci2.cns.iu.edu/user/ask.php> if you have any questions about datasets, or would like to suggest enhancements.

Primary investigators are Katy Börner, Indiana University and Kevin W. Boyack, The Sci2 tool was developed by Micah W. Linnemeier, Patrick A. Phillips, Chintan Tank, Joseph Bheretina, Chin Hui Kwon, and Russell T. Duhon. It runs on the Cyberinfrastructure Shell.

Visualization > Geo Map (network template, not fully rendered)

Console

Loaded: C:\Users\katty\Desktop\sci2-12.20\sampladata\scintometrics\sciFo

Data Manager

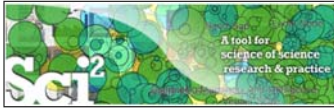
- Pajek .net file: C:\Users\katty\Desktop\sci2-12.20\sampladata\geo\geoNetwork.net
- NetMap_map_geoNetwork
- NetMap net geoNetwork

Save map file as Postscript file and use Adobe or other view to read. It looks like:



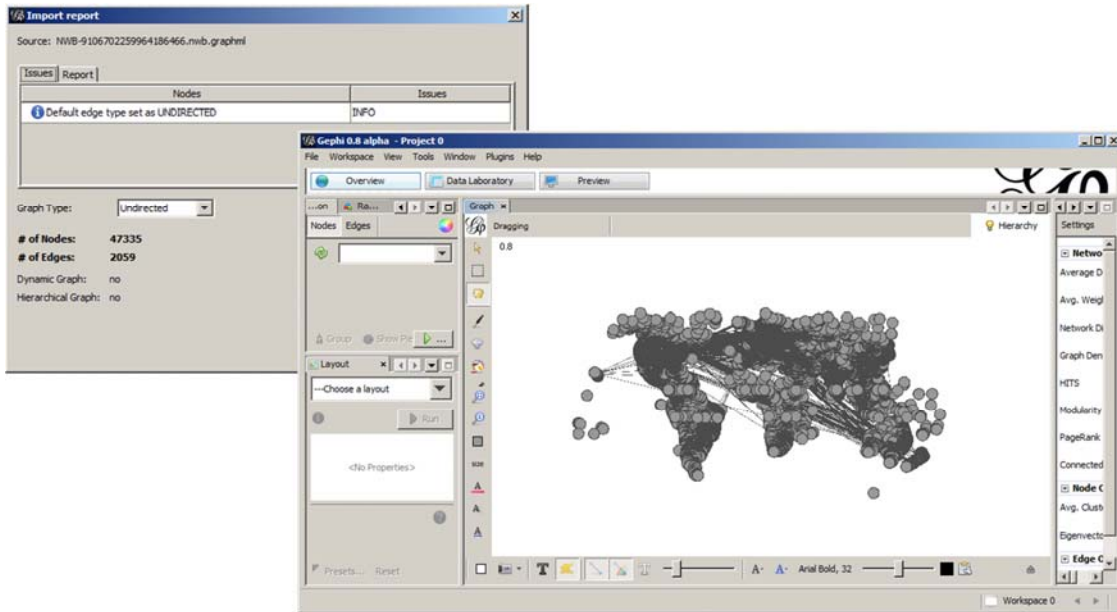
Save .net file as GraphML (Prefuse) and rename to .graphml so that Gephi can read it.

86



Use Gephi to Generate Network Layout

Start gephi. Use *New Project > Open a graph file* to read .graphml file that Sci2 generated.



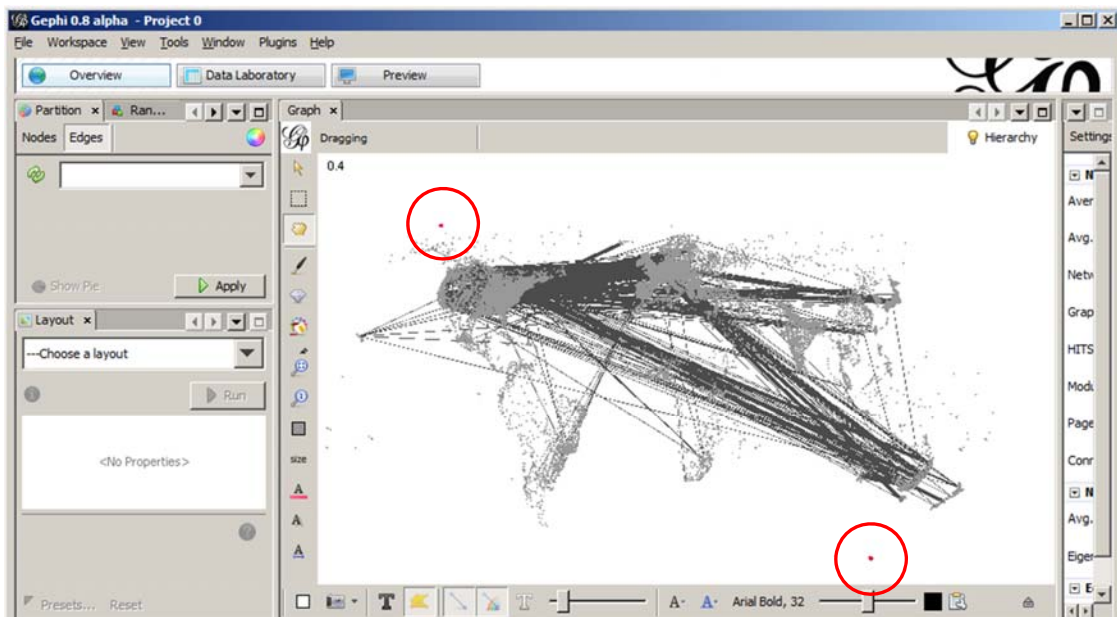
Follow instructions in online tutorial on **Manipulating the Network File in Gephi**

87



Use Gephi to Generate Network Layout

Color or size code the “Near Alaska” and “Near Antarctica” anchor nodes to ease alignment of geomap and network overlay, see instructions in online tutorial on **Manipulating the Network File in Gephi**. Save result using *File > Export > SVG/PDF file*.

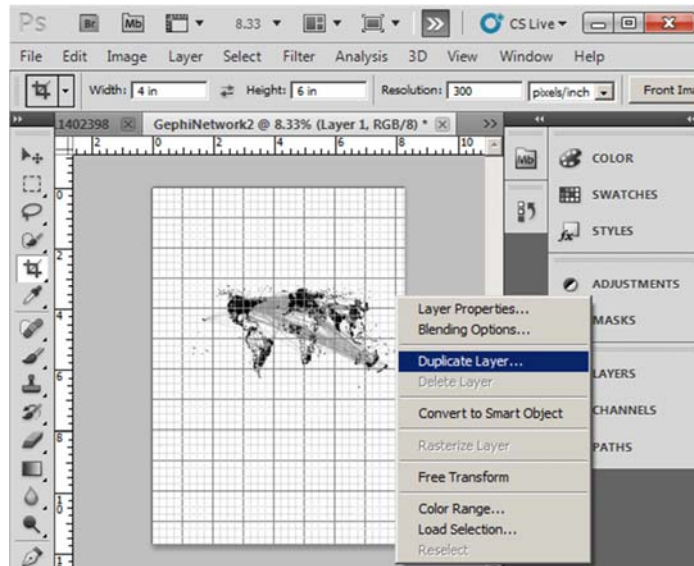


88



Use Photoshop to Overlay Network on Geomap

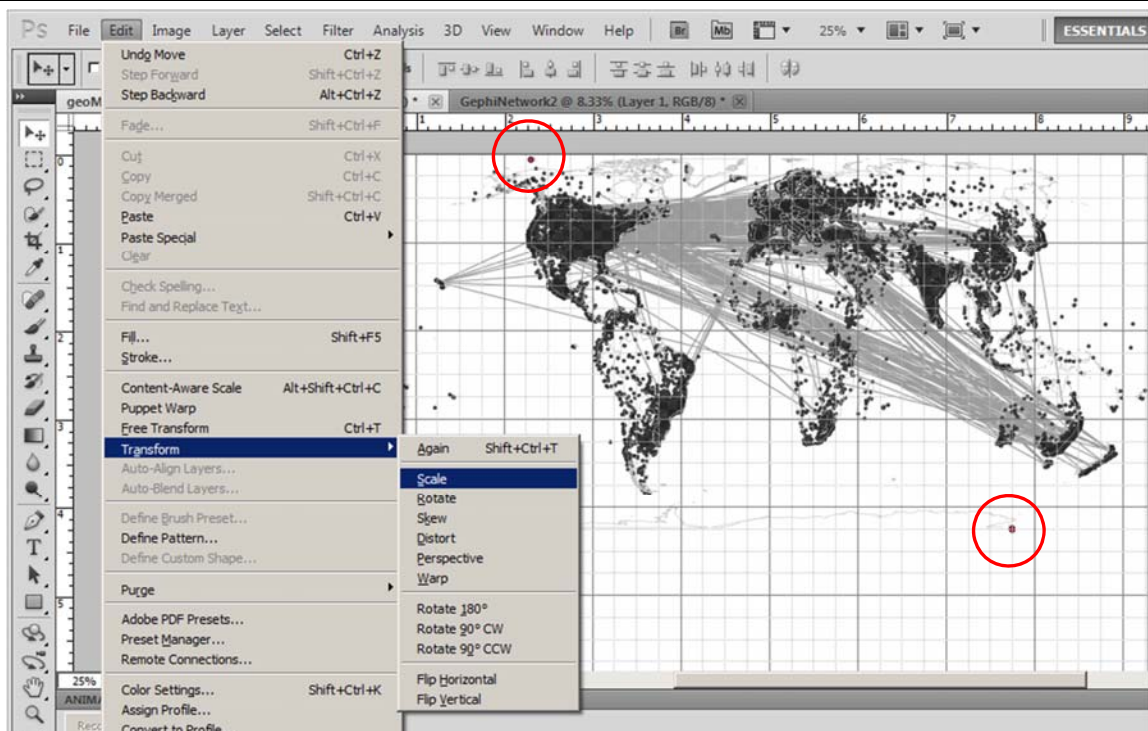
Load geomap and network files into Photoshop. Select 'network' layer and use 'Right click, Duplicate Layer' to copy network over to 'geomap' file as a second layer. Use Edit > Transform > Scale' and align using the "Near Antarctica" anchor nodes, see instructions in online tutorial on **Creating the Visualization in Photoshop**.



89

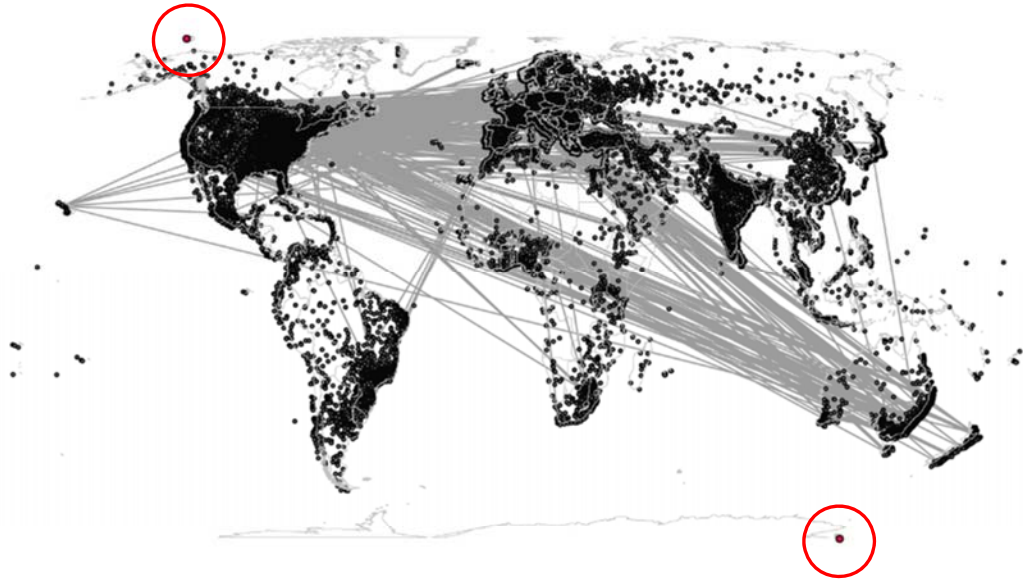


Use Photoshop to Overlay Network on Geomap



90

Delete anchor nodes and save in preferred format.

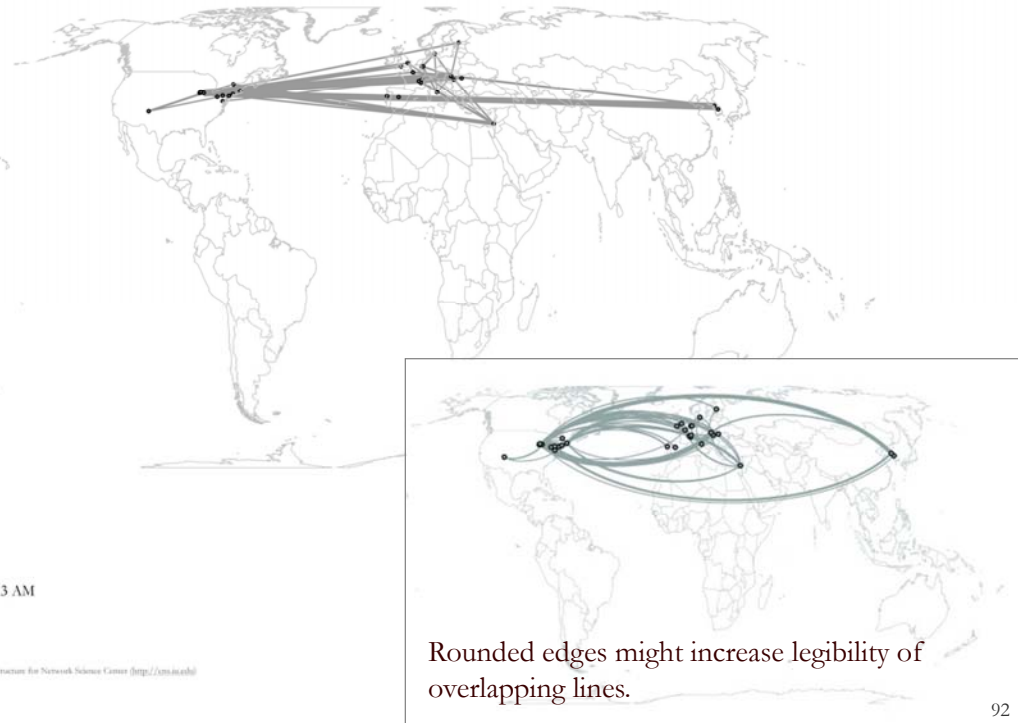


Geo Map ()
Eckert IV Projection
Apr 06, 2012 | 03:19:51 AM

Created with Sci2 Tool | Cyberinfrastructure for Network Science Center (<http://cims.nyu.edu>)

91

Practice these steps using “LaszloBarabasi-collaborations.net” linked from Sci2 wiki:
4.7.6 Using Gephi to Render Networks Overlaid on GeoMaps



Geo Map ()
Eckert IV Projection
Apr 11, 2012 | 06:20:13 AM

Created with Sci2 Tool | Cyberinfrastructure for Network Science Center (<http://cims.nyu.edu>)

Rounded edges might increase legibility of overlapping lines.

92



Tutorial Overview

9:00am Welcome and Overview of Tutorial and Attendees

9:15am Sci2 Tool Usage and Adoption

10:00am Sci2 Tool Basics

Download and run the Sci2 Tool

Load, analyze, and visualize family and business networks

Horizontal line graph of NSF projects

Studying four major network science researchers

- Load and clean a dataset; process raw data into networks
- Find basic statistics and run various algorithms over the network
- Visualize the network using different layouts

11:00am Break

11:15am Sci2 Tool Novel Functionality

- Yahoo! Geocoder
- Geomap with Gephi network overlay
- **Evolving collaboration networks**
- R-Bridge
- New Visualizations

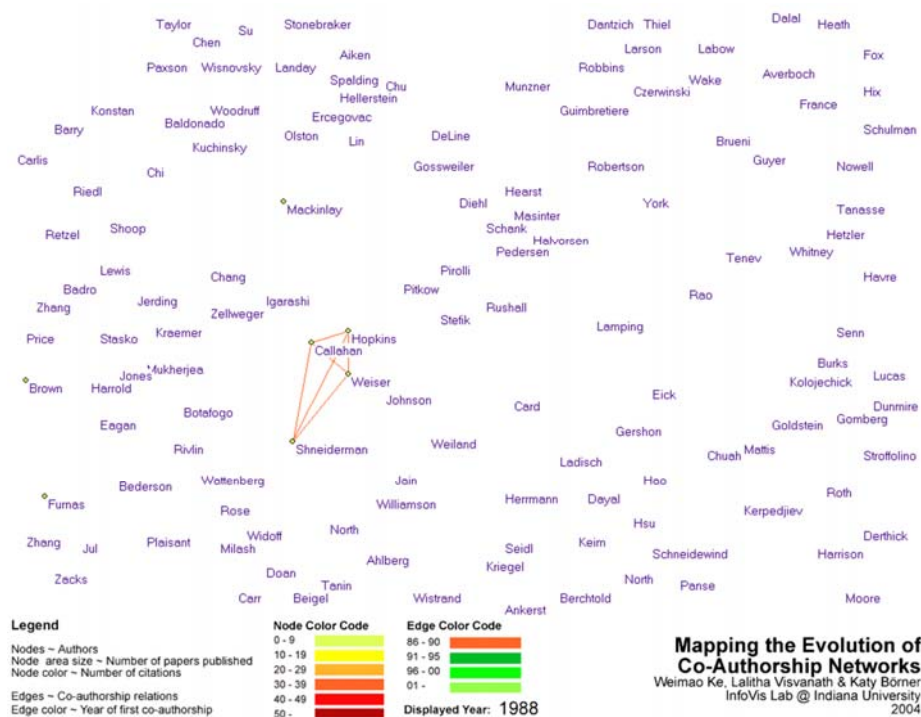
12:15pm Outlook and Q&A

12:30pm Adjourn

93



Sci2 Demo II: Evolving collaboration networks

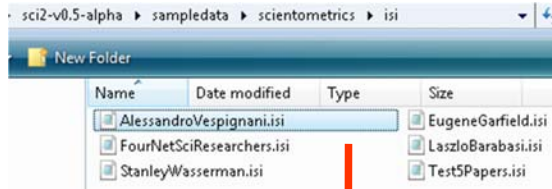


94



Evolving Collaboration Networks

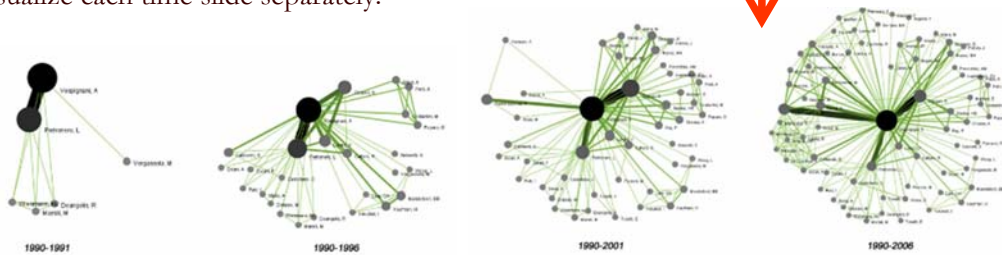
Load isi formatted file



As csv, file looks like:

	A	B	C	D	E	F	G
1	Abstract	Authors	Authors (Full Names)	Beginning	Book Serie	Book Serie	Cited Pate
2	The systematic study of	Colizza, V Barrat, A Barthelemy, M Vespignani, A		2015			
3	Uncovering the hidden r	Colizza, V Flammini, A Serrano, MA Vespignani, A		110			
4	Computer viruses can s	Vespignani, A		135			
5	Mapping the Internet ge	Dall'Asta, L Alvarez-Hamelin, I Barrat, A Vazquez, A Vespignani, A		140			LECTURE NOTES IN

Visualize each time slide separately:



95



Relevant Sci2 Manual entry

- Home
- 1 Introduction
- 2 Getting Started
- 3 Algorithms, Tools, and Plugins
- 4 Workflow Design
- 5 Sample Workflows
 - 5.1 Individual Level Studies - Micro
 - 5.1.1 Mapping Collaboration, Publication, and Funding Profiles of One Researcher (EndNote and NSF Data)
 - 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)
 - 5.2 Institution Level Studies - Meso
 - 5.3 Global Level Studies - Macro
- 6 Sample Science Studies & Online Services
- 7 Extending the Sci2 Tool
- 8 Relevant Datasets and Tools
- 9 References

5.1.2 Time Slicing of Co-Authorship Networks (ISI Data)

Tools ▾

Added by Ted Polley, last edited by Scott Weingart on Mar 16, 2011 (view change)

AlessandroVespignani.isi	
Time frame:	1990-2006
Region(s):	Indiana University, University of Rome, Yale University, Leiden University, International Center for Theoretical Physics, University of Paris-Sud
Topical Area(s):	Informatics, Complex Network Science and System Research, Physics, Statistics, Epidemics
Analysis Type(s):	Co-Authorship Network

The Sci² Tool supports the analysis of evolving networks. For this study, load Alessandro Vespignani's publication history from ISI, which can be downloaded from Thomson's Web of Science or loaded using 'File > Load' and following this path: '[yoursci2directory/sampledata/scientometrics/isi/AlessandroVespignani.isi](#)' using: 'Slice the data into five year intervals from 1990-2006 using 'Preprocessing > Temporal' > 'Slice Table by Time' and the following parameters:

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

[http://sci2.wiki.cns.iu.edu/5.1.2+Time+Slicing+of+Co-Authorship+Networks+\(ISI+Data\)](http://sci2.wiki.cns.iu.edu/5.1.2+Time+Slicing+of+Co-Authorship+Networks+(ISI+Data))

96



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

"Slice Into" allows the user to slice the table by days, weeks, months, quarters, years, decades, and centuries. There are two additional parameters for time slicing: cumulative and align with calendar. The former produces tables containing all data from the beginning to the end of each table's time interval, which can be seen in the Data Manager and below.

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)

The latter option aligns the output tables according to calendar intervals:

101 Unique ISI Records

- slice from beginning of 2002 to end of 2006 (36 records)
- slice from beginning of 1997 to end of 2001 (39 records)
- slice from beginning of 1992 to end of 1996 (22 records)
- slice from beginning of 1990 to end of 1991 (4 records)

Choosing "Years" under "Slice Into" creates multiple tables beginning from January 1st of the first year. If "Months" is chosen, it will start from the first day of the earliest month in the chosen time interval.

[http://sci2.wiki.cns.iu.edu/5.1.2+Time+Slicing+of+Co-Authorship+Networks+\(ISI+Data\)](http://sci2.wiki.cns.iu.edu/5.1.2+Time+Slicing+of+Co-Authorship+Networks+(ISI+Data))

97



Visualize Each Network, Keep Node Positions

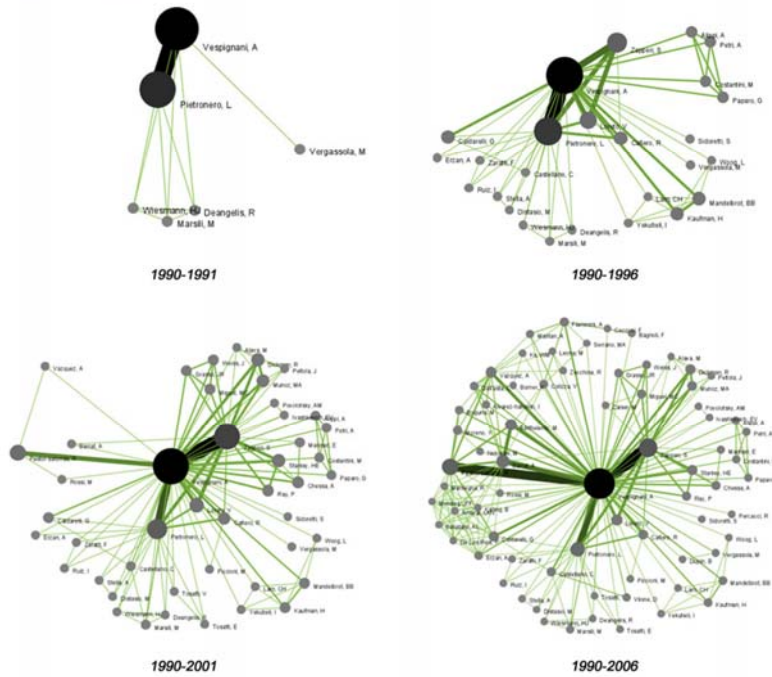
1. To see the evolution of Vespignani's co-authorship network over time, check *'cumulative'*.
2. Extract co-authorship networks one at a time for each sliced time table using *'Data Preparation > Extract Co-Author Network'*, making sure to select "ISI" from the pop-up window during the extraction.
3. To view each of the Co-Authorship Networks over time using the same graph layout, begin by clicking on longest slice network (the *'Extracted Co-Authorship Network'* under *'slice from beginning of 1990 to end of 2006 (101 records)'*) in the data manager. Visualize it in GUESS using *'Visualization > Networks > GUESS'*.
4. From here, run *'Layout > GEM'* followed by *'Layout > Bin Pack'*. Run *'Script > Run Script ...'* and select *'yoursci2directory/scripts/GUESS/co-author-nw.py'*.
5. In order to save the x, y coordinates of each node and to apply them to the other time slices in GUESS, select *'File > Export Node Positions'* and save the result as *'yoursci2directory/NodePositions.csv'*. Load the remaining three networks in GUESS using the steps described above and for each network visualization, run *'File > Import Node Positions'* and open *'yoursci2directory/NodePositions.csv'*.
6. To match the resulting networks stylistically with the original visualization, run *'Script > Run Script ...'* and select *'yoursci2directory/scripts/GUESS/co-author-nw.py'*, followed by *'Layout > Bin Pack'*, for each.

[http://sci2.wiki.cns.iu.edu/5.1.2+Time+Slicing+of+Co-Authorship+Networks+\(ISI+Data\)](http://sci2.wiki.cns.iu.edu/5.1.2+Time+Slicing+of+Co-Authorship+Networks+(ISI+Data))

98

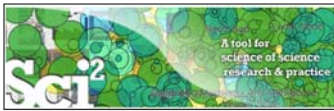


Visualize Each Network, Keep Node Positions



[http://sci2.wiki.cns.iu.edu/5.1.2+Time+Slicing+of+Co-Authorship+Networks+\(ISI+Data\)](http://sci2.wiki.cns.iu.edu/5.1.2+Time+Slicing+of+Co-Authorship+Networks+(ISI+Data))

99



Relevant CShell plugin

CShell Slice Table by Time

Tools ▾

Added by [Artha Alencar](#), last edited by [Ted Polley](#) on Jan 12, 2011 ([view change](#))

Description

Slice Table By Time is an algorithm to chop a table up into new tables, based on a date/time column. It takes the column with the date/time data, a string describing the format of that column, the intervals that the data should be sliced into, whether or not the slices are cumulative, whether or not the slices should be aligned with the calendar, and what day the week is considered to start on (which only matters if the slices are aligned with the calendar) as parameters.

The column to use for date/time values should have a single value for each row of data. It is used by the algorithm to choose which slice(s) the row should end up in. In order to determine what date/time is represented by that row, you must provide the algorithm with a descriptive format, in the second parameter. For instance, a four digit year would be represented by yyyy (the default value). See <http://joda-time.sourceforge.net/api-release/org/joda/time/format/DateTimeFormat.html> for details of all the various formatting options.

The next dropdown has the available intervals to slice the table into. These include milliseconds, seconds, minutes, hours, days, weeks, fortnights, months, quarters, years, decades, and centuries. A future version of the algorithm may include the ability to select how many of these intervals should be grouped together at once.

The checkbox that follows determines if the slices will be cumulative. If the slices are not cumulative, every row in the original table is in one and only one resulting slice. However, if the slices are cumulative, every row in the original table is in the slice it is for and every slice for a period after that.

The checkbox that follows determines if the slices will be aligned with the calendar. For instance, if the first row is for June 7th, 2006 and yearly slices are chosen, then the default behavior will be to have the first slice be from June 7th, 2006 to June 6th, 2007. However, if the slices are aligned with the calendar, the first slice will be from January 1st, 2006 to December 31st, 2006. Alignment does not affect the output for intervals of fortnights, quarters, decades, or milliseconds.

If the slices are aligned with the calendar and are weekly, then the day the week starts is used to determine how they are aligned.

Pros & Cons

The output of the slice algorithm is in separate tables, so a longitudinal analysis will require working with each slice separately, which can be awkward. There will likely be future versions of the time slice algorithm that annotate the original table with the slice the rows belong to.

Applications

When doing longitudinal analysis of data, it can be useful to consider it in chunks, such as to calculate how statistics have changed over time. Alternatively, only a particular time period might be of interest, and this algorithm can extract it from data for a larger time range.

Implementation Details

This algorithm uses the Joda Time library extensively, which provides significantly improved capabilities compared to the default Java algorithms for dates and times.

<http://cishell.wiki.cns.iu.edu/Slice+Table+by+Time>

100



Tutorial Overview

14:15 Welcome and Overview of Tutorial and Attendees

14:30 Sci2 Tool Usage and Adoption

15:00 Sci2 Tool Hands-on

- Download and run the Sci2 Tool
- Load, analyze, and visualize family and business networks
- Horizontal line graph of NSF projects
- Studying four major network science researchers
 - Load and clean a dataset; process raw data into networks
 - Find basic statistics and run various algorithms over the network
 - Visualize as either a circular hierarchy or network

16:00 Sci2 Tool Novel Functionality

- Yahoo! Geocoder
- Geomap with Gephi network overlay
- Evolving collaboration networks
- **R-Bridge**
- New Visualizations

16:30 Outlook and Q&A

16:45 Adjourn

101

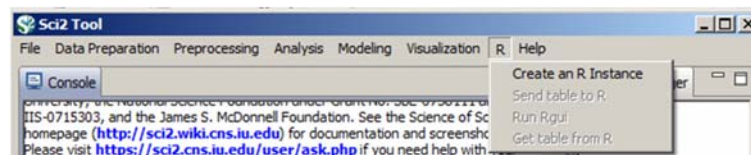


R-Bridge

Download *edu.iu.cns.r_0.0.1.jar* from Additional Plugins wiki page at <http://sci2.wiki.cns.iu.edu/3.2+Additional+Plugins>

Or copy them from the DVD or memory stick.

Available by default starting v1.0.



Run 'R > *Create an R Instance*' parameter should be the path to the directory on your computer that contains Rgui.exe. Results in an 'R Instance' object in the Data Manager.

To send a table from the data manager to an R Instance object, select the table and the R Instance object together then run 'R > *Import Table Into R*'. Select 'R > *Run Rgui*' and the table is available in the R environment using the variable name you specified as a parameter to the Import algorithm.

To pull back data from an R Instance object to the Data Manager, select the R Instance object and run 'R > *Export Table From R*'. Choose the name of the variable from the dropdown list.

102



Tutorial Overview

9:00am Welcome and Overview of Tutorial and Attendees

9:15am Sci2 Tool Usage and Adoption

10:00am Sci2 Tool Basics

Download and run the Sci2 Tool

Load, analyze, and visualize family and business networks

Horizontal line graph of NSF projects

Studying four major network science researchers

- Load and clean a dataset; process raw data into networks
- Find basic statistics and run various algorithms over the network
- Visualize the network using different layouts

11:00am Break

11:15am Sci2 Tool Novel Functionality

- Yahoo! Geocoder
- Geomap with Gephi network overlay
- Evolving collaboration networks
- R-Bridge
- **New Visualizations**

12:15pm Outlook and Q&A

12:30pm Adjourn

103



New Visualizations

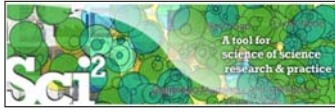
Types

- Geospatial maps: **Choropleth Map** and **Proportional Symbol Map**
- **Science Map**: Based on 25,000 journals or 554 subdisciplines.
- **Bi-Modal Network Layout**

Shared Features

- Uniform layout
- Scalable to extremely large datasets as rendered into PS, PDF files.
- Header information on file mapped and footer information
- Automatic legend generation
- ‘How To Read This Map’ information
- Additional pages for details
- Color coding suitable for black and white printout and color blind users

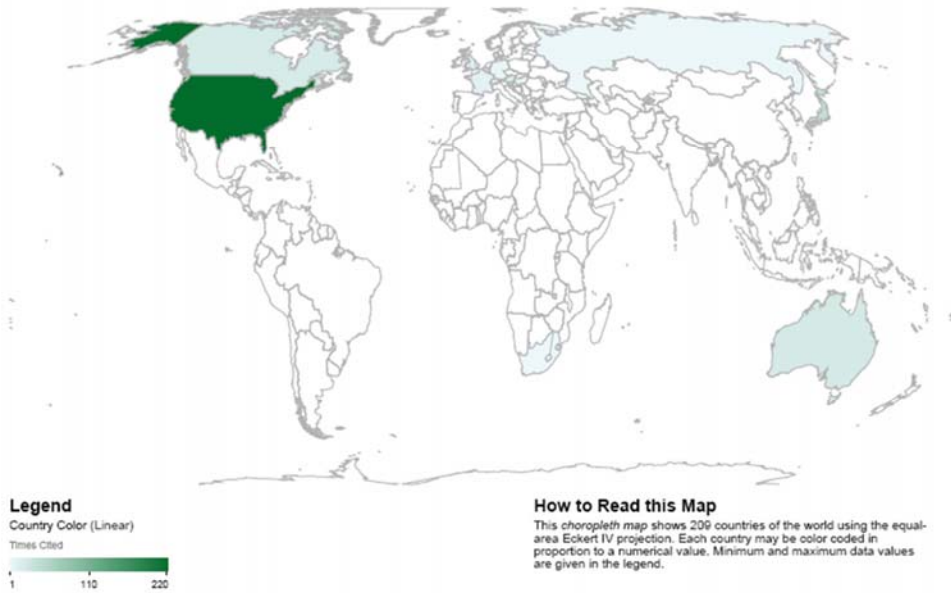
104



New Visualizations

Geospatial Visualization (Choropleth Map)

Generated from CSV file: Preprocessed-usptoInfluenza-8383730930137543104.csv
Jun 05, 2012 | 05:45:00 PM EDT



CNS (cns.iu.edu)

105



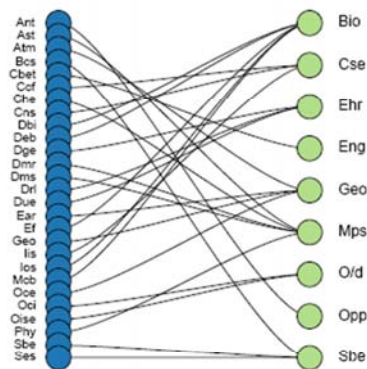
New Visualizations

Network Visualization

Generated from bipartite network from NSF Organization to NSF Directorate
June 5, 2012 | 5:05 PM EDT

NSF Organization

NSF Directorate

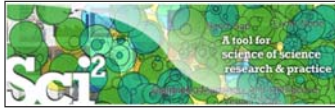


Legend
Sorted by
Left side:
Alphabetical
Right side:
Alphabetical

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)

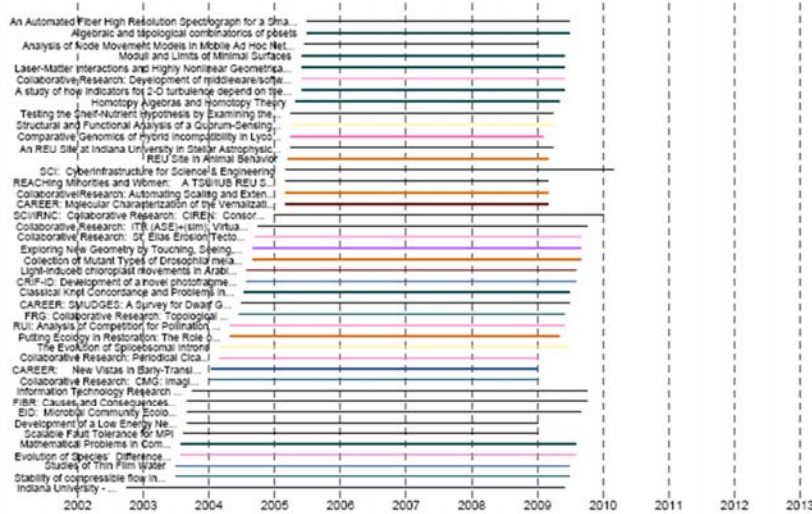
106



New Visualizations

Temporal Visualization

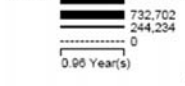
Generated from NSF csv file: C:\Users\katy\Desktop\TOOLS\Sci2-2012.06.04-KNAW\sampled\datascientometrics\isf\indiana.nsf
June 05, 2012 | 4:50 PM EDT



Legend

Area size: Award Number
Minimum = 220,560
Maximum = 852,643
Text label: Title
Color: NSF Organization
See end of PDF for color legend.

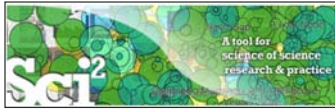
Area



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.

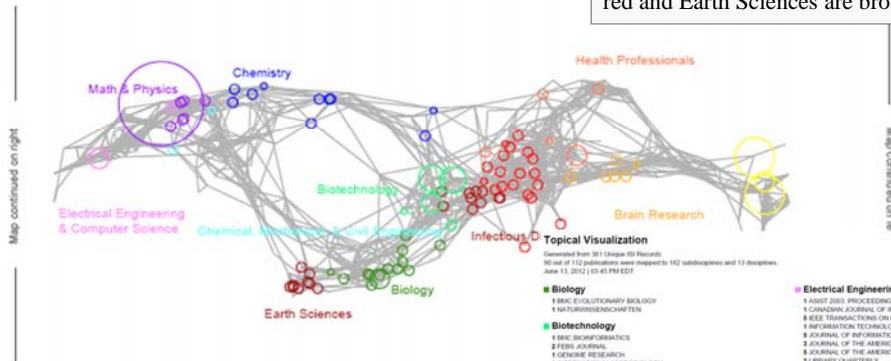
CNS (cns.iu.edu)



New Visualizations

Topical Visualization

Generated from 361 Unique ISI Records of 4 NetSci Researchers
14 out of 109 publications were mapped to 94 subdisciplines and 12 disciplines.
June 05, 2012 | 05:39 PM EDT



Legend

Circle area: Fractional Journal Count
Unclassified = 96
Minimum = 0
Maximum = 25
Color: Discipline
See end of PDF for color legend.

Area



How To Read This Map

The UCSD map are aggregated in color and is labeled unique subdiscipline and assigned record:

CNS (cns.iu.edu)

Data: WoS and Scopus paper level data for 2001–2010, about 25,000 separate journals, proceedings, and series.

Similarity Metric: Combination of bibliographic coupling and keyword vectors.

Number of Disciplines: 554 journal clusters further aggregated into 13 main scientific disciplines that are labeled and color coded in a metaphorical way, e.g., Medicine is blood red and Earth Sciences are brown as soil.



Tutorial Overview

9:00am Welcome and Overview of Tutorial and Attendees

9:15am Sci2 Tool Usage and Adoption

10:00am Sci2 Tool Basics

Download and run the Sci2 Tool

Load, analyze, and visualize family and business networks

Horizontal line graph of NSF projects

Studying four major network science researchers

- Load and clean a dataset; process raw data into networks
- Find basic statistics and run various algorithms over the network
- Visualize the network using different layouts

11:00am Break

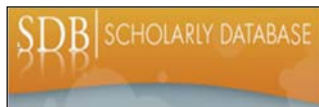
11:15am Sci2 Tool Novel Functionality

- Yahoo! Geocoder
- Geomap with Gephi network overlay
- Evolving collaboration networks
- R-Bridge
- New Visualizations

12:15pm Outlook and Q&A

12:30pm Adjourn

109



Scholarly Database at Indiana University

<http://sdb.wiki.cns.iu.edu>

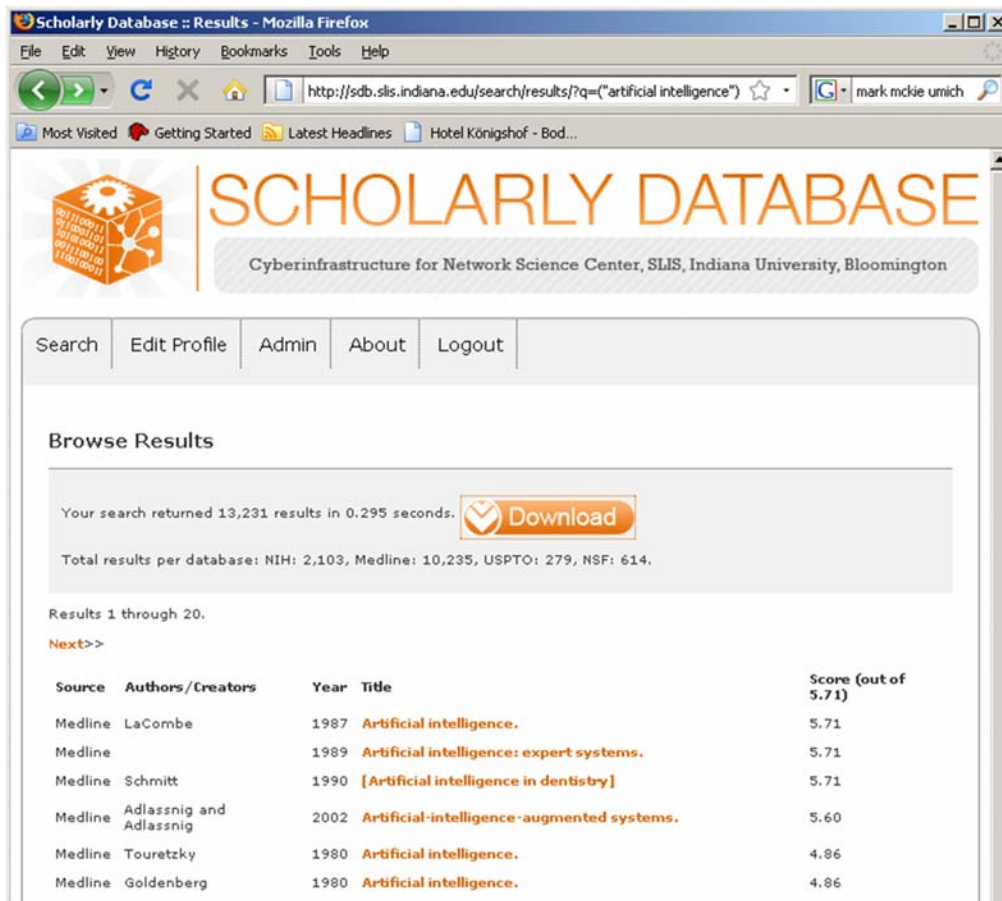
Supports federated search of 25 million publication, patent, grant records.

Results can be downloaded as data dump and (evolving) co-author, paper-citation networks.

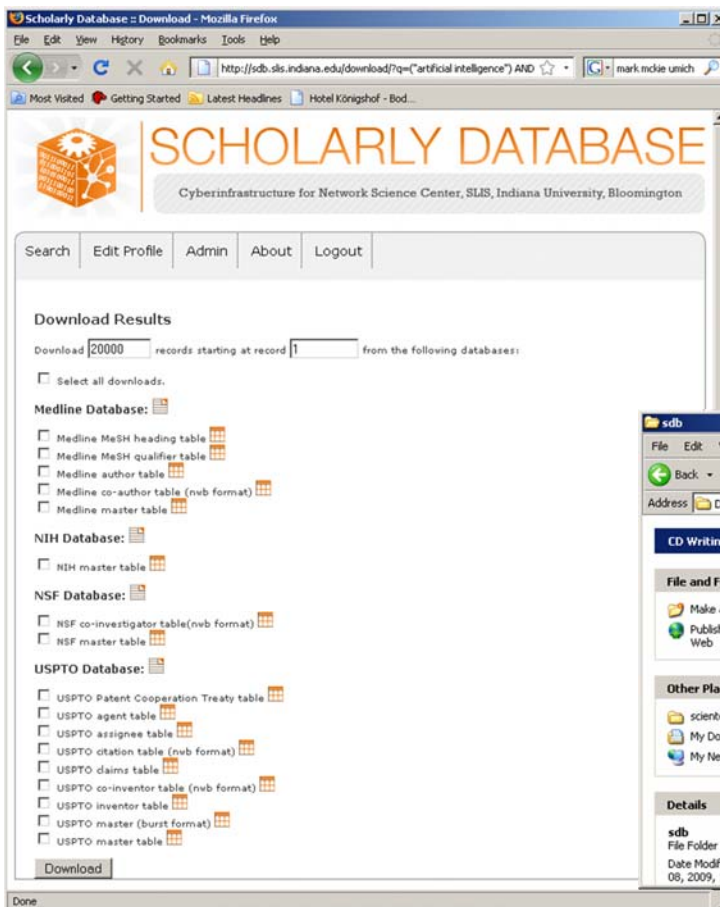
The screenshot displays the Scholarly Database website. On the left is the login page with fields for 'IU User' and 'Non-IU User'. On the right is the search interface with a search bar, filters for 'First Year' and 'Last Year', and a list of search filters including 'Medline (1898 - 2009)', 'NIH (1961 - 2002)', 'NSF (1985 - 2004)', and 'USPTO (1976 - 2007)'. The search button is labeled 'Search'.

Register for free access at <http://sdb.cns.iu.edu>

110



111

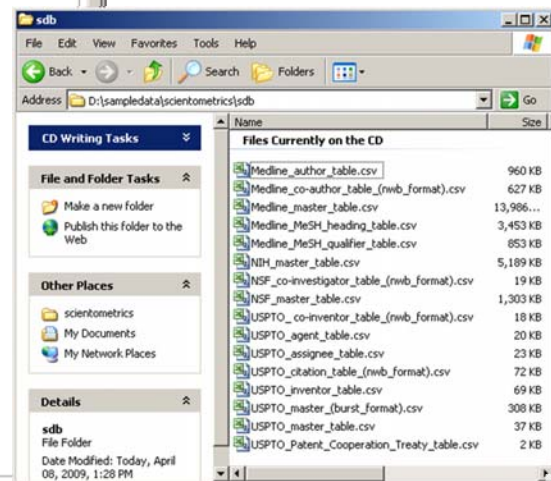


Since March 2009:

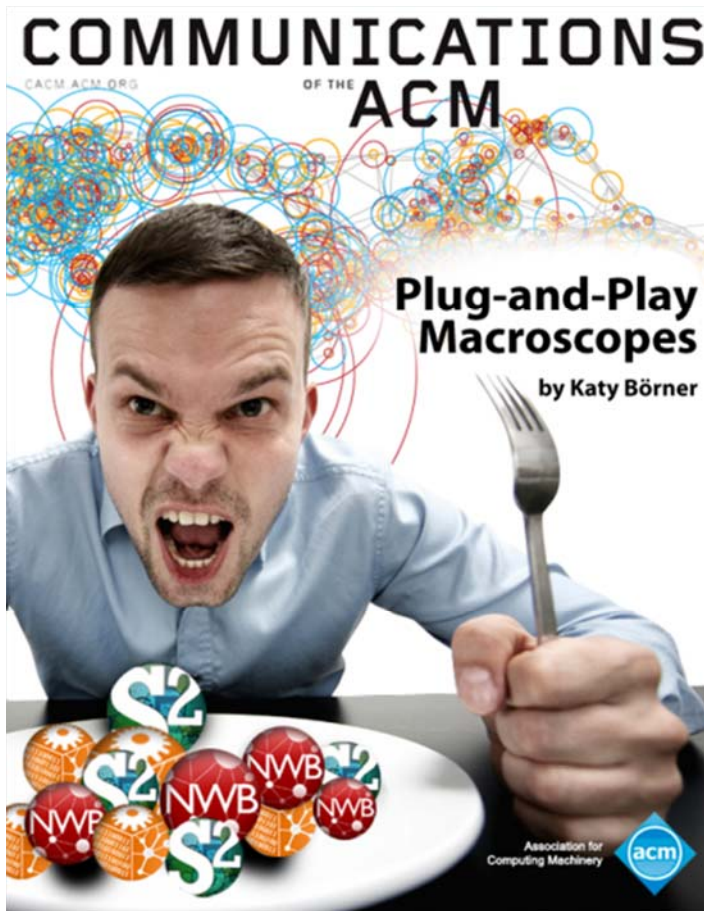
Users can download networks:

- Co-author
- Co-investigator
- Co-inventor
- Patent citation

and tables for burst analysis in NWB.



112



Börner, Katy. (March 2011). Plug-and-Play Macroscopes. *Communications of the ACM*, 54(3), 60-69.

Video and paper are at <http://www.scivee.tv/node/27704>

113



OSGi/CIShell Adoption

A number of other projects recently adopted OSGi and/or CIShell:

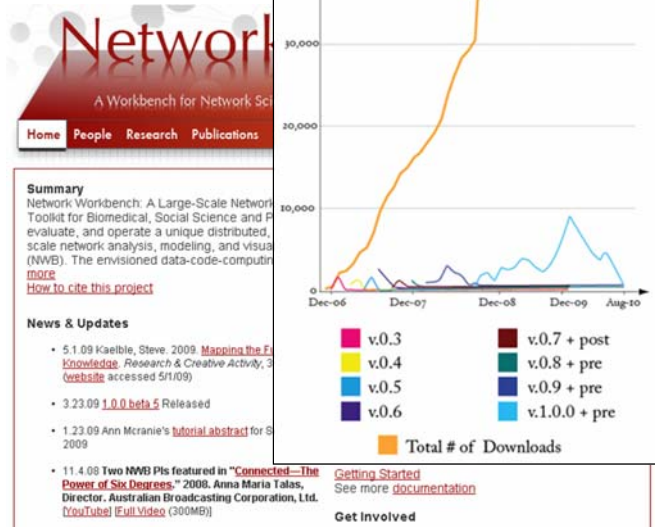
- USA**
- *Cytoscape* (<http://cytoscape.org>) Led by Trey Ideker at the University of California, San Diego is an open source bioinformatics software platform for visualizing molecular interaction networks and integrating these interactions with gene expression profiles and other state data (Shannon et al., 2002).
 - *MAEviz* (<https://wiki.ncsa.uiuc.edu/display/MAE/Home>) Managed by Jong Lee at NCSA is an open-source, extensible software platform which supports seismic risk assessment based on the Mid-America Earthquake (MAE) Center research.
- Europe**
- *Taverna Workbench* (<http://taverna.org.uk>) Developed by the myGrid team (<http://mygrid.org.uk>) led by Carol Goble at the University of Manchester, U.K. is a free software tool for designing and executing workflows (Hull et al., 2006). Taverna allows users to integrate many different software tools, including over 30,000 web services.
 - *TEXTrend* (<http://texttrend.org>) Led by George Kampis at Eötvös Loránd University, Budapest, Hungary supports natural language processing (NLP), classification/mining, and graph algorithms for the analysis of business and governmental text corporuses with an inherently temporal component.
 - *DynaNets* (<http://www.dynanets.org>) Coordinated by Peter M.A. Sloot at the University of Amsterdam, The Netherlands develops algorithms to study evolving networks.
 - *SISOB* (<http://sisob.lcc.uma.es>) An Observatory for Science in Society Based in Social Models.
- As the functionality of OSGi-based software frameworks improves and the number and diversity of dataset and algorithm plugins increases, the capabilities of custom tools will expand.

The Network Workbench (NWB) tool supports researchers, educators, and practitioners interested in the study of biomedical, social and behavioral science, physics, and other networks.

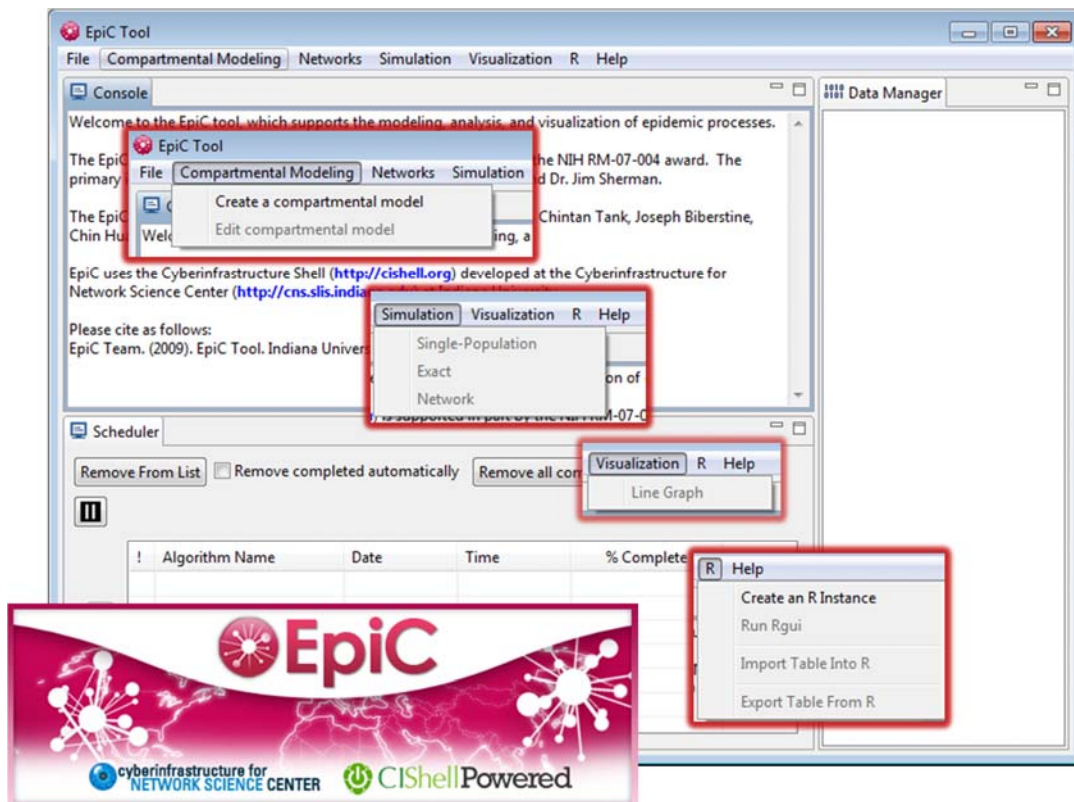
In February 2009, the tool provides more than 169 plugins that support the preprocessing, analysis, modeling, and visualization of networks.

More than 50 of these plugins can be applied or were specifically designed for S&T studies.

It has been downloaded more than 125,000 times since December 2006.



Herr II, Bruce W., Huang, Weixia (Bonnie), Penumarthy, Shashikant & Börner, Katy. (2007). Designing Highly Flexible and Usable Cyberinfrastructures for Convergence. In Bainbridge, William S. & Roco, Mihail C. (Eds.), *Progress in Convergence - Technologies for Human Wellbeing* (Vol. 1093, pp. 161-179), *Annals of the New York Academy of Sciences*, Boston, MA.



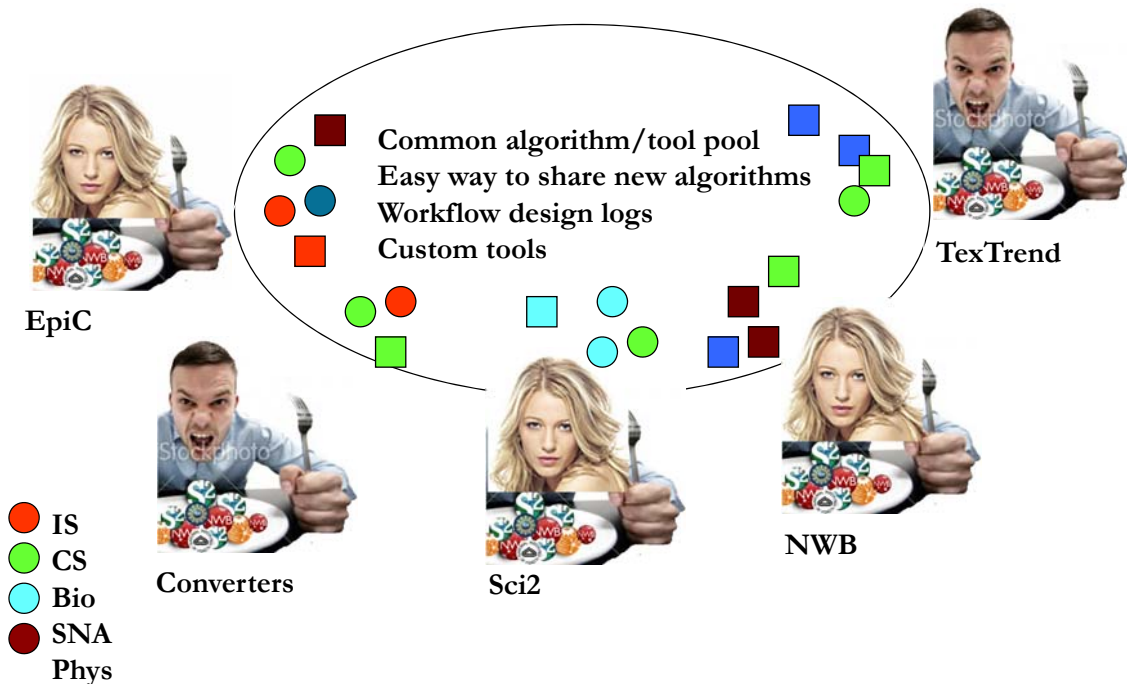
TEXTrend adds R bridge, WEKA, Wordij, CFinder, and more.

See the latest versions of TEXTrend Toolkit modules at

http://textrend.org/index.php?option=com_content&view=article&id=47&Itemid=53



OSGi/CIShell-Powered Tools Support Algorithm Sharing





CIShell – Integrate New Algorithms

About the Cyberinfrastructure Shell

The Cyberinfrastructure Shell (CIShell) is an open source, community-driven platform for the integration and utilization of datasets, algorithms, tools, and computing resources. Algorithm integration support is built in for Java and most other programming languages. Being Java based, it will run on almost all platforms. The software and specification is released under an Apache 2.0 License.

CIShell is the basis of [Network Workbench](#), [TexTrend](#), [SciF](#) and the upcoming [EpiC](#) tool.

CIShell supports remote execution of algorithms. A standard web service definition is in development that will allow pools of algorithms to transparently be used in a peer-to-peer, client-server, or web front-end fashion.

CIShell Features

A framework for easy integration of new and existing algorithms written in any programming language

Using CIShell, an algorithm writer can fully concentrate on creating their own algorithm in whatever language they are comfortable with. Simple tools are provided to then take their algorithm and

Learn More...

- [CIShell Papers](#)
- [CIShell Powered Tools](#)
- [Algorithms](#)
- [Plugins \(coming soon\)](#)
- [Misc. Tool Documentation](#)
- CIShell Web Services (coming soon)
- [Screenshots](#)

Getting Started...

- [Documentation & Developer Resources](#)
- [Download](#)

Getting Involved...

- [Contact Us](#)

CIShell Developer Guide is at <http://cishell.wiki.cns.iu.edu>

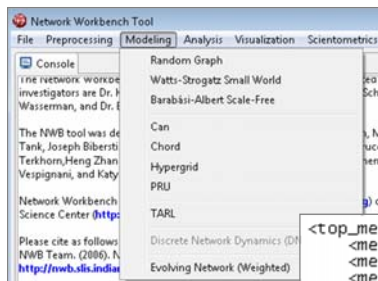
Additional Sci2 Plugins are at <http://sci2.wiki.cns.iu.edu/3.2+Additional+Plugins>

119



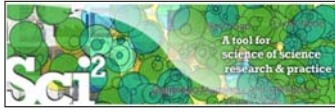
CIShell – Customize Menu

- The file *'yourtooldirectory/configuration/default_menu.xml'* encodes the structure of the menu system.
- In NWB Tool, the Modeling menu (left) is encoded by the following piece of xml code:



```
<?xml version="1.0" encoding="UTF-8" ?>
<top_menu name="Modeling">
  <menu pid="edu.iu.nwb.modeling.erdosrandomgraph"/>
  <menu pid="edu.iu.nwb.modeling.smallworld"/>
  <menu pid="edu.iu.nwb.modeling.barabasiAlbert"/>
  <menu type="break"/>
  <menu pid="edu.iu.iv.modeling.p2p.can.CanAlgorithm"/>
  <menu pid="edu.iu.iv.modeling.p2p.chord.ChordAlgorithm"/>
  <menu pid="edu.id.iv.modeling.p2p.hypergrid.Hypergrid"/>
  <menu pid="edu.iu.iv.modeling.p2p.pru.PruAlgorithm"/>
  <menu type="break"/>
  <menu pid="edu.iu.iv.modeling.tarl.TarlAlgorithm"/>
  <menu type="break"/>
  <menu pid="edu.iu.nwb.modeling.discretenetworkdynamics.DNDAlgorithm"/>
  <menu type="break"/>
  <menu pid="edu.iu.nwb.modeling.weighted.evolvingnetwork"/>
</top_menu>
```

120



Need Help? Ask an Expert!

<https://sci2.cns.iu.edu/user/ask.php>

121



Interested to Learn More? Take the IVMOOC

Information Visualization MOOC

INDIANA UNIVERSITY CNS Cyberinfrastructure for Network Science Center

Home Schedule Announcements My Profile Forum Tools FAQ samueltolemanmills@gmail.com | Logout

Overview

This course provides an overview about the state of the art in information visualization. It teaches the process of producing effective visualizations that take the needs of users into account.

Among other topics, the course covers:

- Data analysis algorithms that enable extraction of relationships in data
- Major visualization and interaction techniques
- Discussions of systems that drive research and development.

A certificate will be issued upon successful completion. Please watch the introduction video to get better acquainted with the course.

Katy Börner, Ph.D.
Indiana University

COMMUNICATIONS
ACM
Plug-and-Play
Macrosopes
by Katy Börner

Börner, Katy. (March 2011). Plug-and-Play Macrosopes. *Communications of the ACM*, 54(3). <http://www.acm.org/node/27704>

Sign Up For The Course

122

If you enjoyed the tutorial, please thank Stefan Hornbostel, Jasmin Schmitz, and Sibylle Glaab at the Institut für Forschungsinformation und Qualitätssicherung in Berlin, Germany

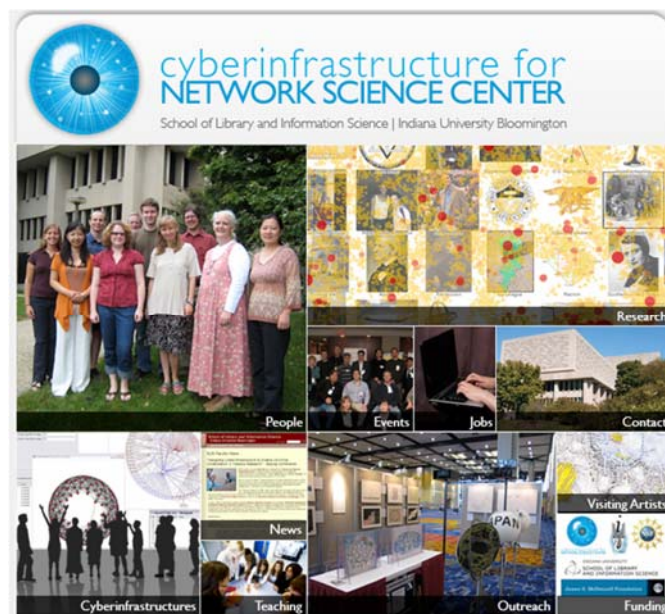
Q & A

Please complete the Post-Tutorial Questionnaire so that we can further improve these tutorials.

*** * ***

Bug reports and all comments are welcome.

123



All papers, maps, tools, talks, press are linked from <http://cns.iu.edu>

CNS Facebook: <http://www.facebook.com/cnscenter>

Mapping Science Exhibit Facebook: <http://www.facebook.com/mappingscience>