

# "Sci2 Tool: Temporal, Geospatial, Topical, and Network Analysis and Visualization" Tutorial for Digital Arts and Humanities Scholars



**Dr. Katy Börner**

CNS & IVL, SLIS, Indiana University, Bloomington, IN, USA  
Data Archiving and Networked Services (DANS) and e-Humanities group,  
Royal Netherlands Academy of Arts and Sciences (KNAW), The Netherlands  
<http://cns.iu.edu>

**Albert Meroño Peñuela (Meertens) & Chin Hua Kong, Joseph R. Biberstine (CNS, IU)**  
provide technical assistance.

With special thanks to Kevin W. Boyack, Chin Hua Kong, Micah Linnemeier, Russell J. Duhon, Patrick Phillips, Chintan Tank, Thomas Smith, Nianli Ma, 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**

*Meertens Institute, Amsterdam, The Netherlands  
Thursday Feb 16, 2012 • 12:30-16:30*



## Software, Datasets, Plugins, and Documentation

- These slides  
<http://ivl.slis.indiana.edu/km/pres/2012-borner-sci2tutorial-meertens.pdf>
- Sci2 Tool Manual v0.5.1 Alpha  
<http://sci2.wiki.cns.iu.edu>
- Sci2 Tool v0.5.1 Alpha (May 4, 2011)  
<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.



## Tutorial Overview

**12:30p Welcome and Overview of Tutorial and Attendees**

**12:45p Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools**

**1:00p Sci2 Tool Basics**

- Download and run the Sci2 Tool
- Load, analyze, and visualize family and business networks
- 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

*2:30p Break*

**3:00p Sci2 Tool Novel Functionality**

- Yahoo! Geocoder
- Evolving collaboration networks
- R-Bridge

**4:00p Outlook and Q&A**

*4:30 Adjourn*

3



## Tutorial Overview

**12:30p Welcome and Overview of Tutorial and Attendees**

**12:45p Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools**

**1:00p Sci2 Tool Basics**

- Download and run the Sci2 Tool
- Load, analyze, and visualize family and business networks
- 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

*2:30p Break*

**3:00p Sci2 Tool Novel Functionality**

- Yahoo! Geocoder
- Evolving collaboration networks
- R-Bridge

**4:00p Outlook and Q&A**

*4:30 Adjourn*

4



## Using the Sci2 Tool to Visualize Tutorial Registrants

Use *File > Read* to load cleaned *Sci2TutorialRegistrants.csv*

	A	B
1	<b>Affiliation</b>	<b>Background</b>
2	RU	Geography, Science & Technology Studies
3	EUR	Cultural History
4	EUR	Sociology
5	Huygens	Historical Disciplines
6	Meertens	Computational Musicology
7	Leiden U	Religious Studies, Semitic Languages
8	UVA	Sociology

<b>Affiliation</b>	<b>#</b>
CWI	1
CWTS	1
DANS	1
e-Humanities Group	1
EUR	3
Huygens	2
Leiden U	2
Meertens	3
Open Knowledge Foundation Network	1
RU	1
Utwente	1
UU	1
UVA	2
VU	6
<b>Grand Total</b>	<b>26</b>

Total of 26 (max #) registrants on Oct 10, 2011.  
Main affiliations are shown to the right.

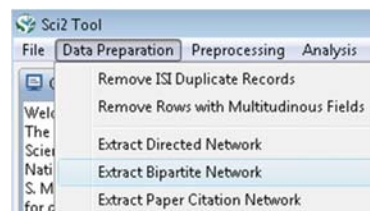
5



## Using the Sci2 Tool to Visualize Tutorial Registrants

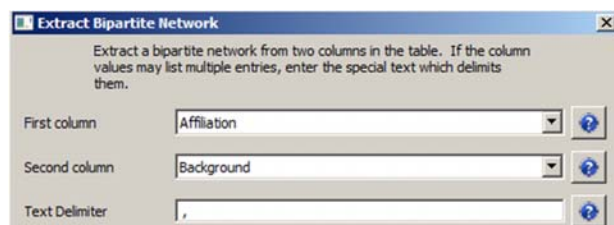
Use *File > Read* to load cleaned *Sci2TutorialRegistrants.csv*

	A	B
1	<b>Affiliation</b>	<b>Background</b>
2	RU	Geography, Science & Technology Studies
3	EUR	Cultural History
4	EUR	Sociology
5	Huygens	Historical Disciplines
6	Meertens	Computational Musicology
7	Leiden U	Religious Studies, Semitic Languages
8	UVA	Sociology



Run *Data Preparation > Extract Bipartite Network*

With parameter values:

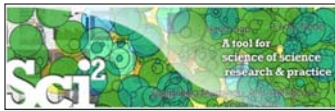
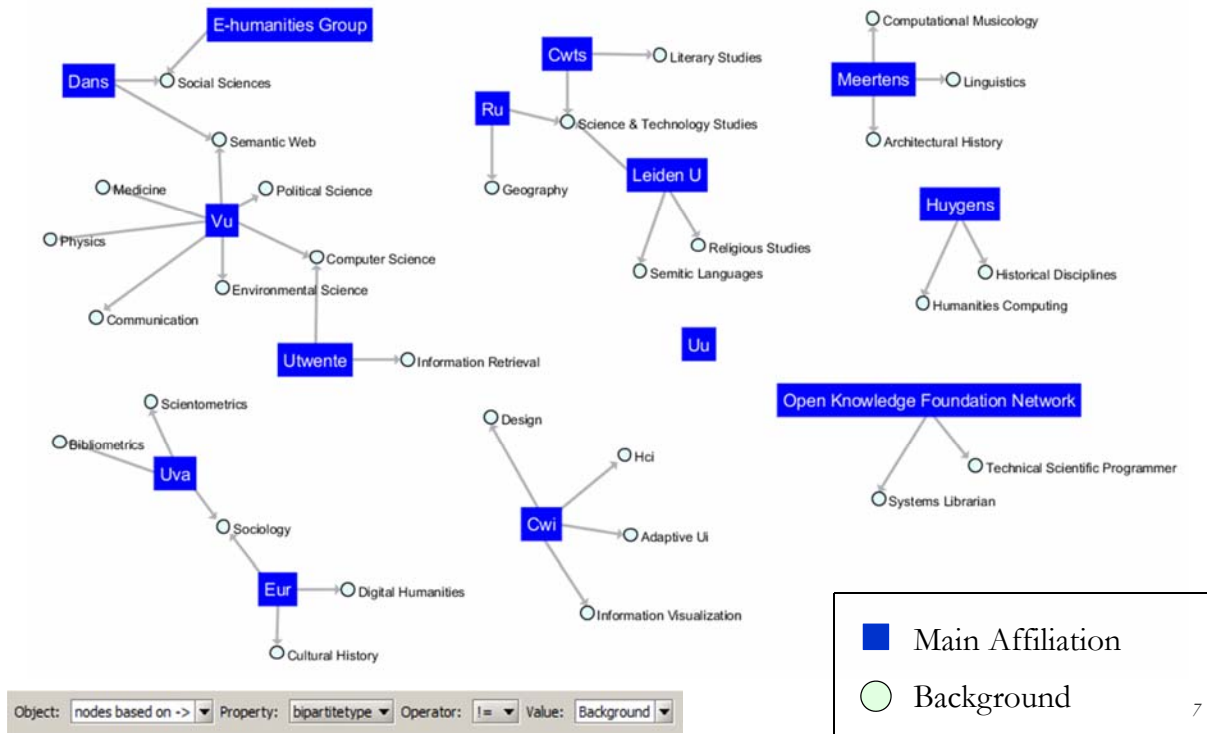


Visualize resulting *Bipartite network from Affiliation to Background* using *Visualization > Network > GUESS* and *Layout > GEM*, *Layout > Bin Pack*

6



## Sci2 Tool – Visualize Tutorial Attendees: Bi-partite Affiliation-Background Network



## Wordle.net of “Interested to Learn” Response



Learn
Get to know new tools to attach to our university repository software
A basic understanding of the possibilities of the Sc2 tool in particular in e-humanities research
Principles behind the development and use of Sci2, so that I will be able to decide whether they (probably the principles, rather than the program itself), are also applicable to my own field of Digital Philology.
which research questions can be addressed (advantages and disadvantages) - insights in how to prepare data for analysis - examples on how visualizations can be interpreted
How to interface the tool with other systems, technical details of the software. How we can collaborate with Sci2 and BibServer/BibSoup



## Tutorial Overview

12:30p Welcome and Overview of Tutorial and Attendees

12:45p Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools

1:00p Sci2 Tool Basics

- Download and run the Sci2 Tool
- Load, analyze, and visualize family and business networks
- 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

2:30p Break

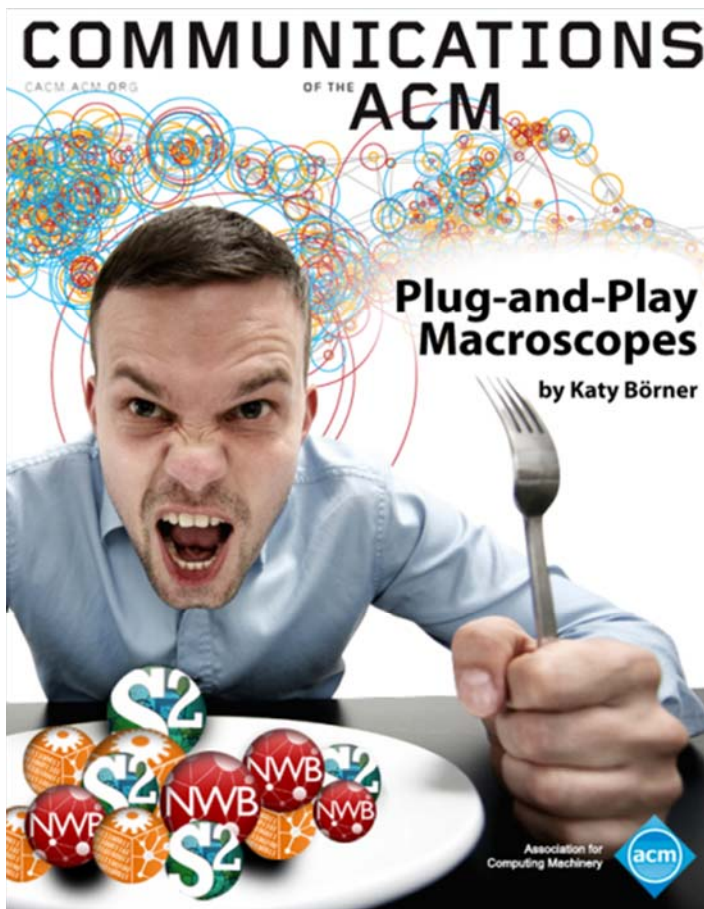
3:00p Sci2 Tool Novel Functionality

- Yahoo! Geocoder
- Evolving collaboration networks
- R-Bridge

4:00p Outlook and Q&A

4:30 Adjourn

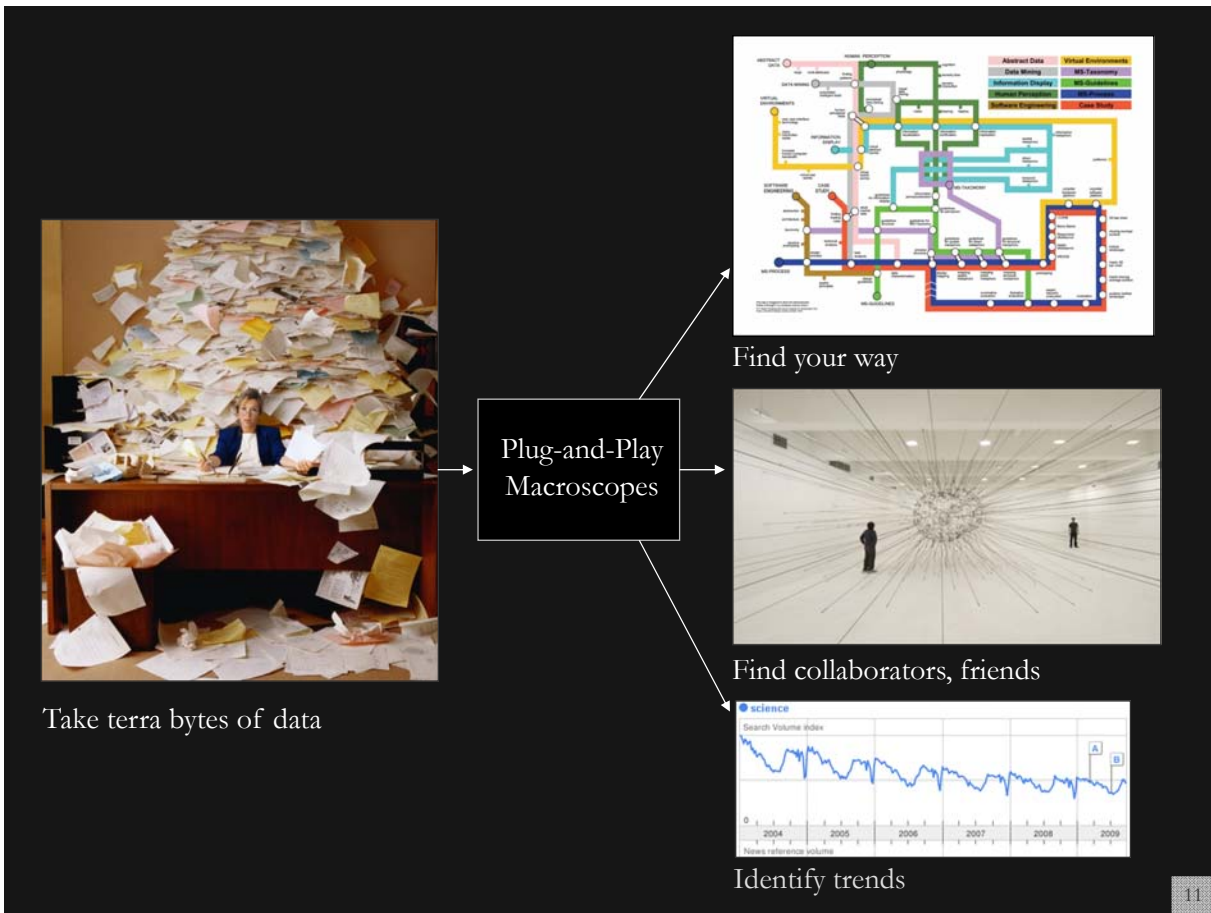
9



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>

10



## Macroscopes

Decision making in science, industry, and politics, as well as in daily life, requires that we make sense of data sets representing the structure and dynamics of complex systems. Macroscopes provide a “vision of the whole,” helping us “synthesize” the related elements and enabling us to detect patterns, trends, and outliers while granting access to myriad details. Rather than make things larger or smaller, **macroscopes let us observe what is at once too great, slow, or complex for the human eye and mind to notice and comprehend.**



Microscopes



Telescopes



Macroscopes

## Plug-and-Play Macroscopes

While microscopes and telescopes are **physical instruments**, macroscopes resemble **continuously changing bundles of software plug-ins**. Macroscopes make it easy to select and combine algorithm and tool plug-ins but also interface plug-ins, workflow support, logging, scheduling, and other plug-ins needed for scientifically rigorous yet effective work.

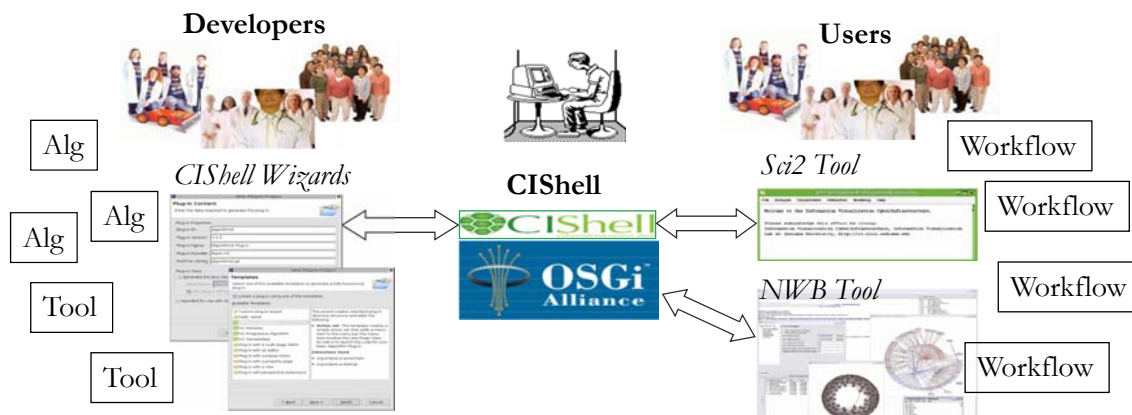
They make it easy to share plug-ins via email, flash drives, or online. To use new plugins, simply copy the files into the plug-in directory, and they appear in the tool menu ready for use. No restart of the tool is necessary. **Sharing algorithm components, tools, or novel interfaces becomes as easy as sharing images on Flickr or videos on YouTube. Assembling custom tools is as quick as compiling your custom music collection.**

13

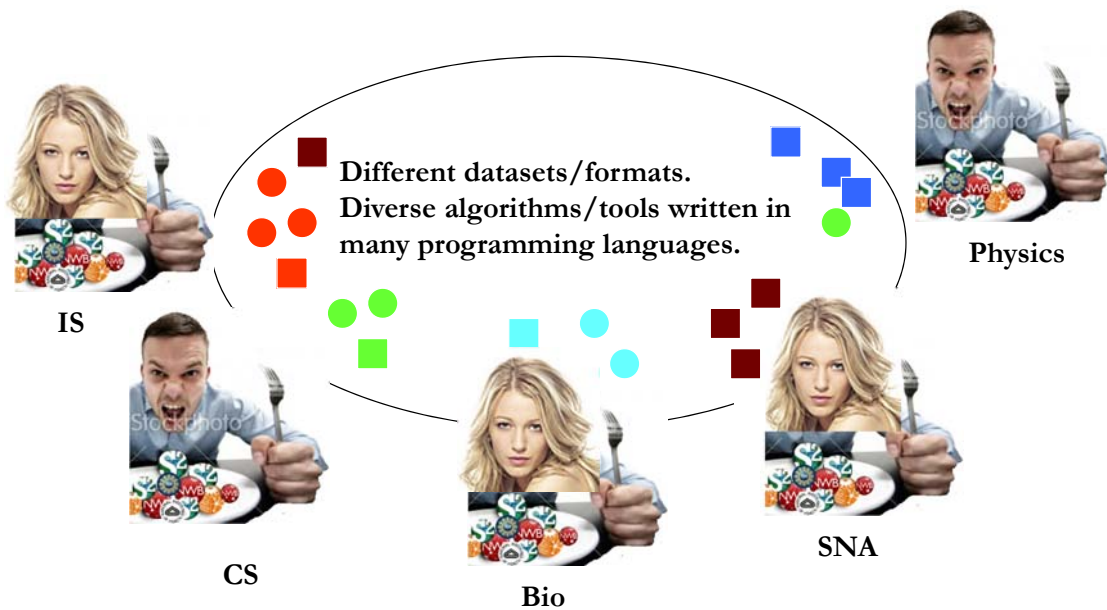


### OSGi & Cyberinfrastructure Shell (CIShell)

- CIShell (<http://cishell.org>) is an open source software specification for the integration and utilization of datasets, algorithms, and tools.
- It extends the Open Services Gateway Initiative (OSGi) (<http://osgi.org>), a standardized, component oriented, computing environment for networked services widely used in industry since more than 10 years.
- Specifically, CIShell provides “sockets” into which existing and new datasets, algorithms, and tools can be plugged using a wizard-driven process.



14



15



## Tutorial Overview

12:30p Welcome and Overview of Tutorial and Attendees

12:45p Plug-and-Play Macrosopes, OSGi/CIShell Powered Tools

**1:00p Sci2 Tool Basics**

- **Download and run the Sci2 Tool**
- Load, analyze, and visualize family and business networks
- 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

*2:30p Break*

3:00p Sci2 Tool Novel Functionality

- Yahoo! Geocoder
- Evolving collaboration networks
- R-Bridge

4:00p Outlook and Q&A

*4:30 Adjourn*

16





## 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

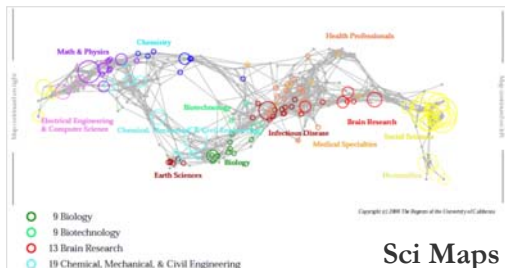
To capture the essence of good science, stakeholders must combine forces to create an open, sound and consistent system for measuring all the activities that make up academic productivity, says **Julia Lane**.

17

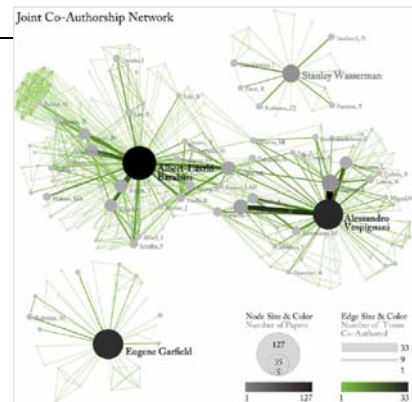


## Sci2 Tool – “Open Code for S&T Assessment”

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

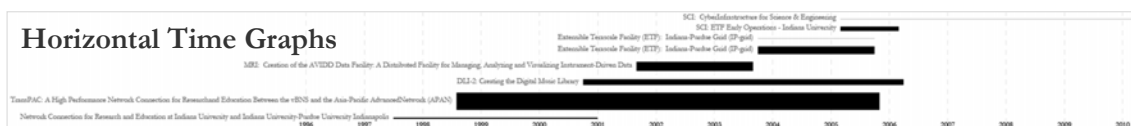


Sci Maps



GUESS Network Vis

### Horizontal Time Graphs

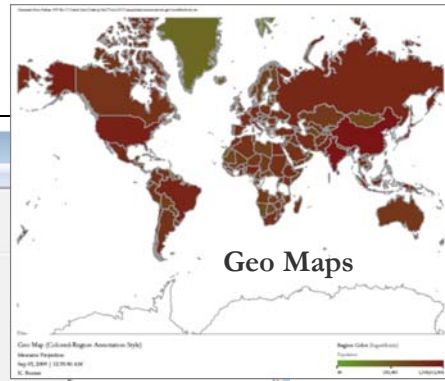


Börner, Katy, Huang, Weixia (Bonnie), Linnemeier, Micab, Dubon, Russell Jackson, Phillips, Patrick, Ma, Nianli, Zoss, Angela, Guo, Hanning & Price, Mark. (2009). *Reve-Network-Red: Analyzing and Visualizing Scholarly Networks Using the Scholarly Database and the Network Workbench Tool*. *Proceedings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

18



# Sci² Tool Visualizations



Sci² Tool

File Preprocessing Modeling Analysis Visualization Scientometrics Help

Console

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

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

Please cite as follows:  
Sci² Team. (2009). Science of Science Tool. In Proceedings of the 2009 ACM SIGMIS Database on the Science of Science. Indianapolis, IN: ACM.

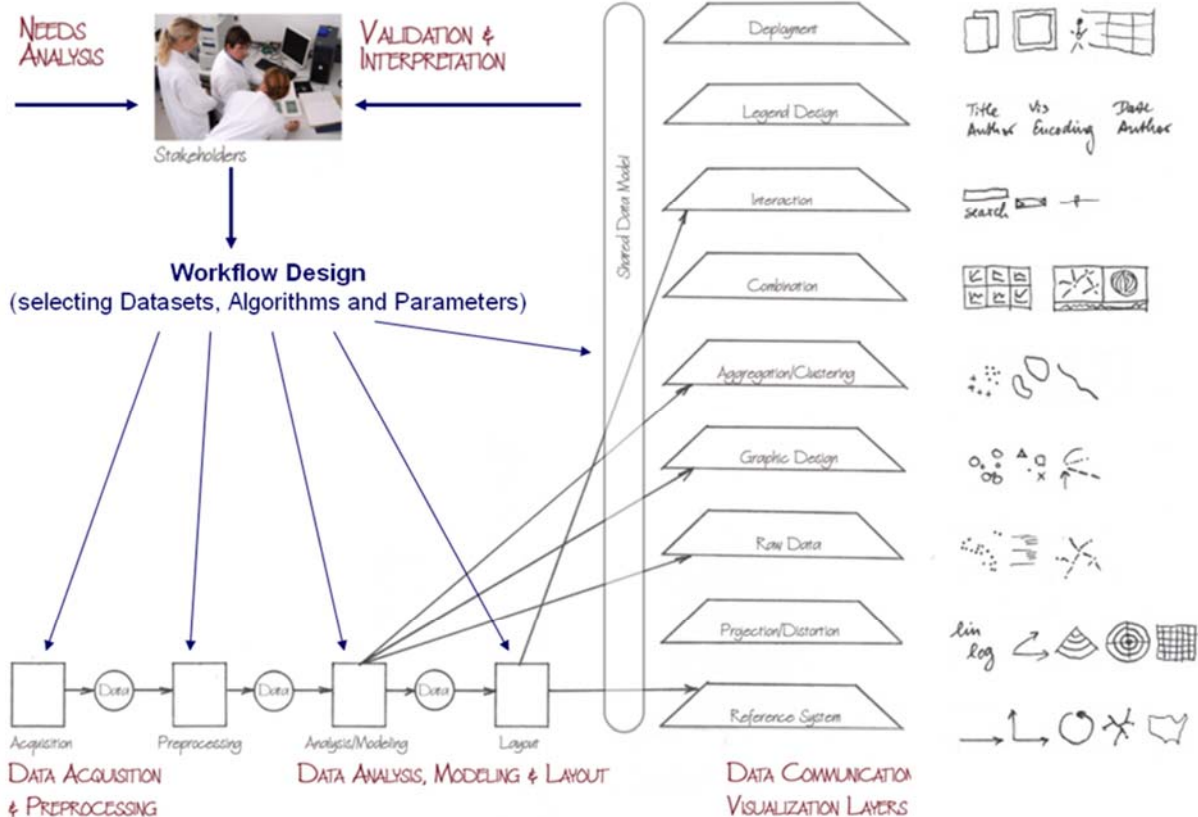
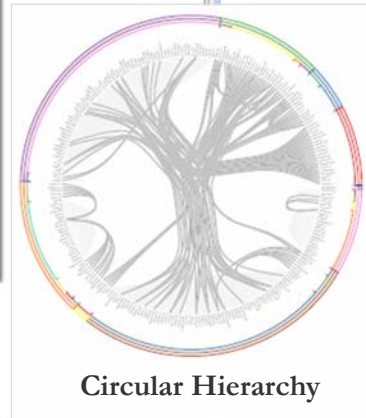
Scheduler

Remove From List  Remove completed

Algorithm Name	Date	Time	% Complete
Extract Co-Author Network	09/03/2009	00:15:20 AM	100%
Load and Clean ISI File	09/03/2009	00:15:05 AM	100%

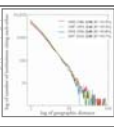

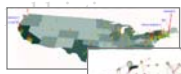
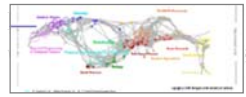
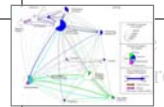
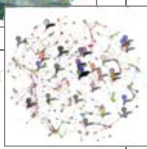

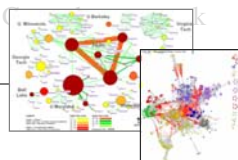
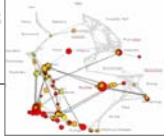
Visualization menu:

- 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




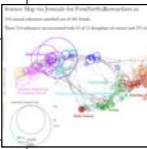
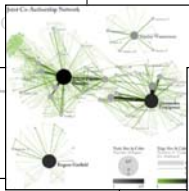
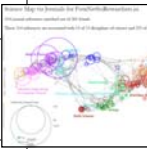
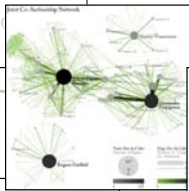

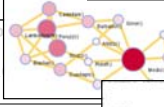
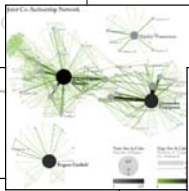
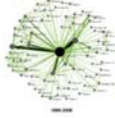


## Type of Analysis vs. Level of Analysis

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



## Type of Analysis vs. Level of Analysis Covered Today:

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



## Sci<sup>2</sup> 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

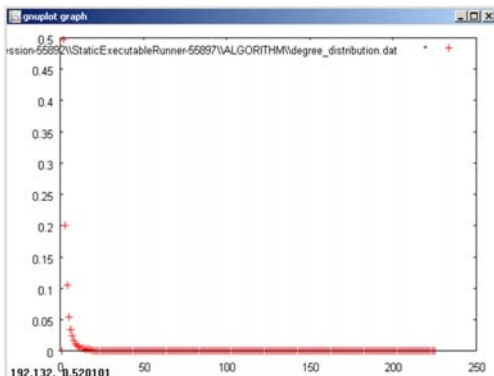
- JPEG (\*.jpg)
- PDF (\*.pdf)
- PostScript (\*.ps)

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

23

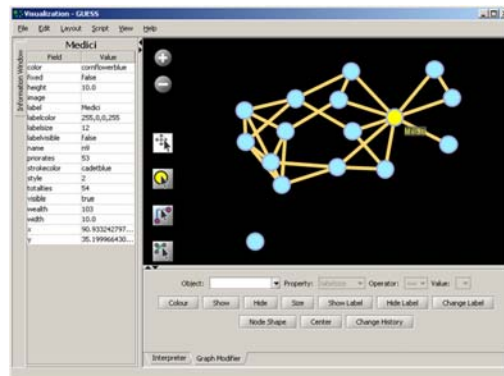


## Sci<sup>2</sup> Tool – Supported Tools



### 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>.

24



## Sci² Tool: Download, Install, and Run

### Sci² Tool v0.5.1 Alpha (May 4, 2011)

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.

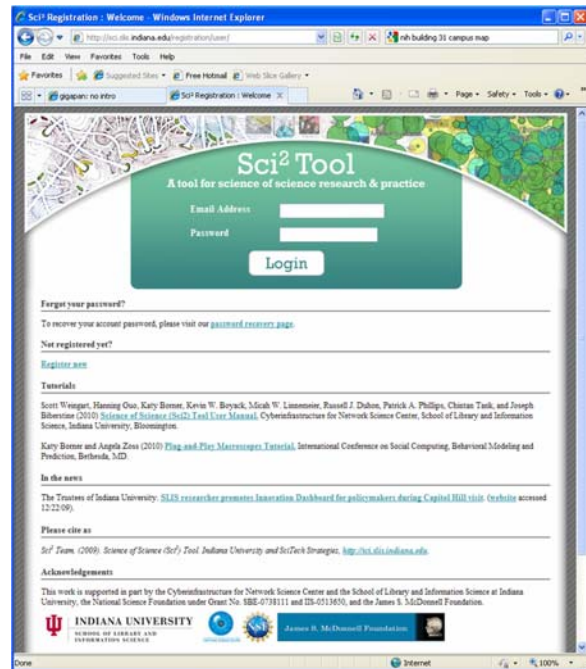
Run /sci2/sci2.exe

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>*



25



## Sci² Tool: Download, Install, and Run

### Sci² Tool v0.5 Alpha (May 4, 2011)

- Supports ASCII UTF-8 characters
- Web-based Yahoo! and desktop Geocoders
- U.S. and World geomapper
- Customizable stop word lists
- Merging of networks
- New home page, wiki-based tutorial
- Bug fixes, streamlined workflows



Sci² Tool runs on Windows, Mac, and Linux.

Unzip.

Run /sci2/sci2.exe



26

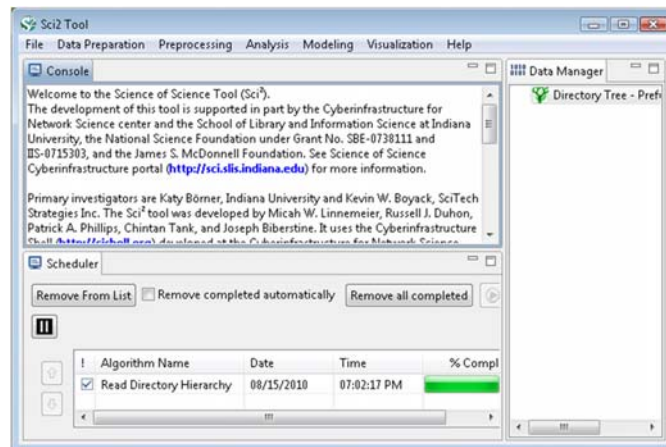


## Sci2 Tool Interface Components

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 an 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>

27



## Tutorial Overview

12:30p Welcome and Overview of Tutorial and Attendees

12:45p Plug-and-Play Macrosopes, OSGi/CIShell Powered Tools

**1:00p Sci2 Tool Basics**

- Download and run the Sci2 Tool
- **Load, analyze, and visualize family and business networks**
- 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

*2:30p Break*

3:00p Sci2 Tool Novel Functionality

- Yahoo! Geocoder
- Evolving collaboration networks
- R-Bridge

4:00p Outlook and Q&A

*4:30 Adjourn*

28



## 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

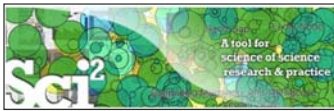
- Wealth: Each family's net wealth in 1427 (in thousands of lira).
- Priorates: The number of seats on the civic council held between 1282-1344.
- Totalities: Number of business/marriage ties in complete dataset of 116 families.

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>



## 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"

```



## 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'.

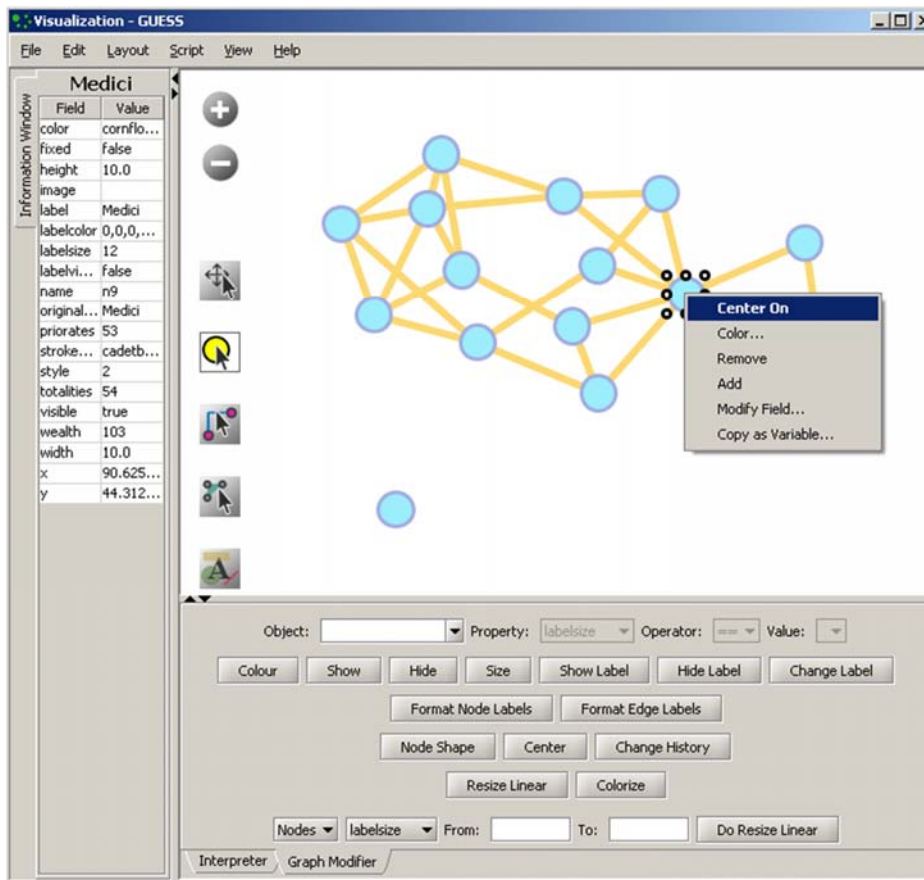
31

The screenshot shows the Network Workbench Tool interface. The main window displays a network graph titled 'Medici-Acciaiuoli'. The graph consists of 16 nodes (blue circles) and 27 edges (yellow lines). The nodes are arranged in a complex, interconnected structure. The interface includes a console window on the left showing GUESS execution logs, a data manager on the right, and a central visualization window with a table of node attributes.

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


32



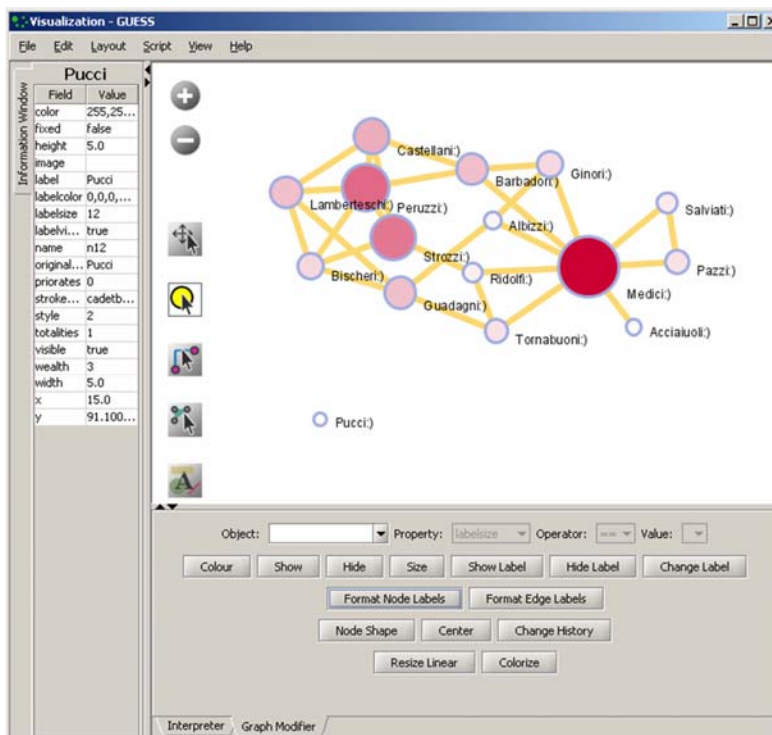


Pan:  
“grab” the background by holding left-click and moving your mouse.

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 node 'Acciaiuoli:)'. It contains a table with the following data:

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

The main window displays a network graph with nodes labeled: Castellani:, Barbadori:, Ginori:, Lamberteschi:, Peruzzi:, Albizzi:, Salviati:, Strozzi:, Ridolfi:, Pazzi:, Medici:, Acciaiuoli:, Guadagni:, Tomabuoni:, and Pucci:. The nodes are connected by yellow lines. The 'Medici:' node is the largest and colored red, while others are smaller and colored white or light blue. A toolbar with various icons is visible on the left side of the graph area.

At the bottom of the window is an 'Interpreter' window with the following code:

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

A callout box next to the interpreter window contains the text:

**Interpreter:**  
Uses Jython a combination of Java and Python.  
  
Try  
colorize(wealth, white, red)

35



## Tutorial Overview

12:30p Welcome and Overview of Tutorial and Attendees

12:45p Plug-and-Play Macrosopes, OSGi/CIShell Powered Tools

**1:00p Sci2 Tool Basics**

- Download and run the Sci2 Tool
- Load, analyze, and visualize family and business networks
- **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

*2:30p Break*

3:00p Sci2 Tool Novel Functionality

- Yahoo! Geocoder
- Evolving collaboration networks
- R-Bridge

4:00p Outlook and Q&A

*4:30 Adjourn*

36



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

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

Thomson Reuter's Web of Knowledge (WoS) is a leading citation database. 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))

37



## 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



38

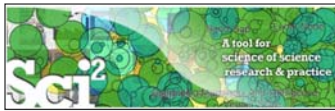


## Comparison of Counts

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 <i>(Dec 2007)</i>
	41	16,920	159	52 <i>(Dec 2008)</i>
	44	30,102	201	68 <i>(April 2011)</i>

39

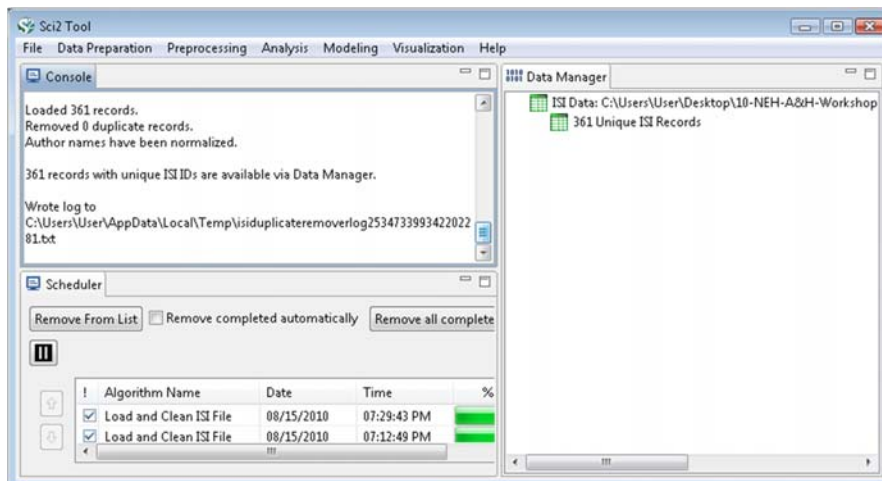


## Extract Co-Author Network

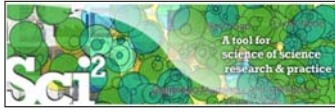
Load *\*yoursci2directory\*/sampledata/scientometrics/isi/FourNetSciResearchers.isi'*  
using *'File > Load ...'*

And file with 361 records appears in the Data Manager.

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



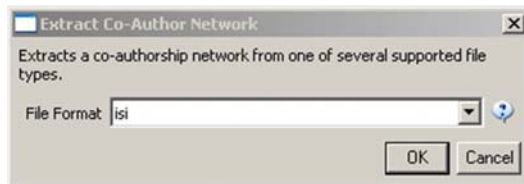
40



## 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.

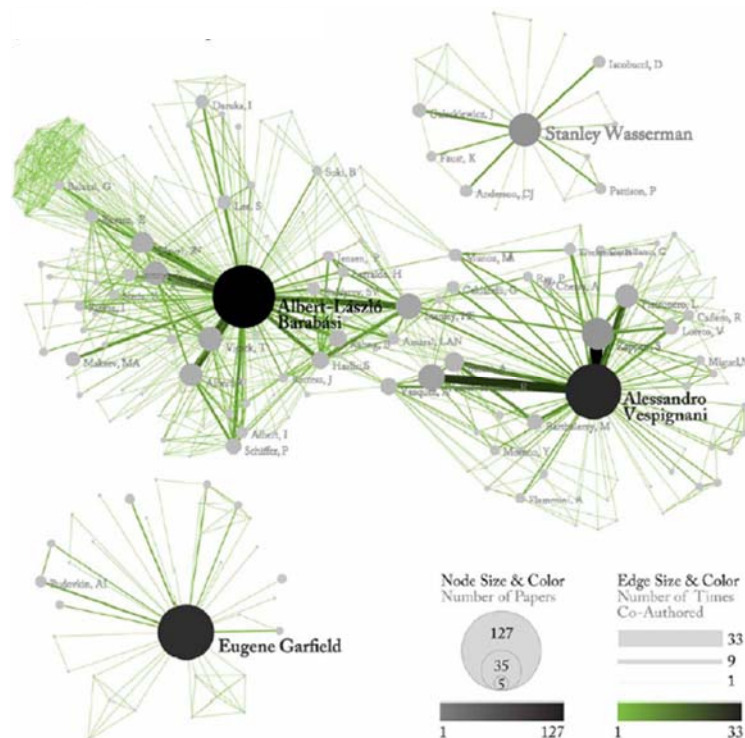
To view the complete network, select the 'Extracted Co-Authorship Network' and run 'Visualization > Networks > GUESS'.

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'.

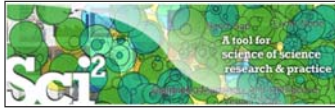
41



## Co-Author Network of all Four NetsSci Researchers



42

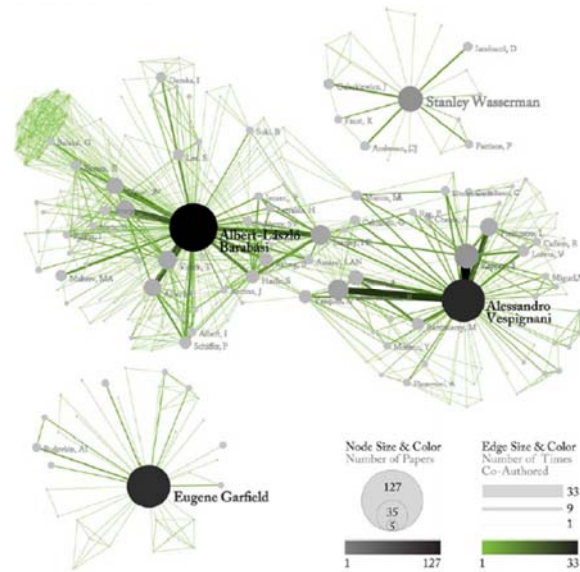


## Co-Author Network of all Four NetsSci Researchers

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

Calculate node degrees in Sci2 Tool.

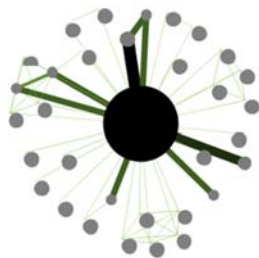
Use a graphic program to add legend.



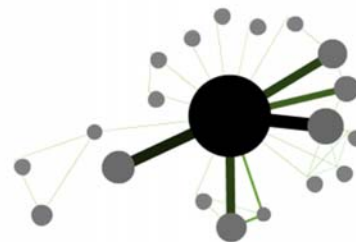
43



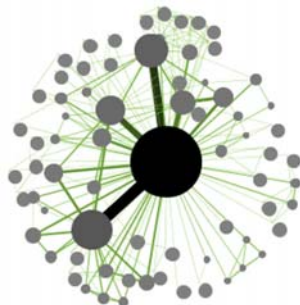
## Individual Co-Author Networks (Read/map 4 files separately)



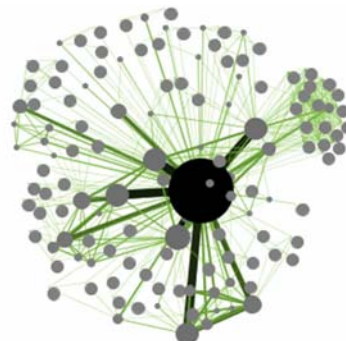
Eugene Garfield



Stanley Wasserman



Alessandro Vespignani



Albert-László Barabási

44

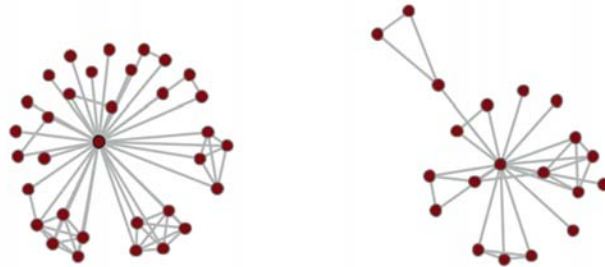
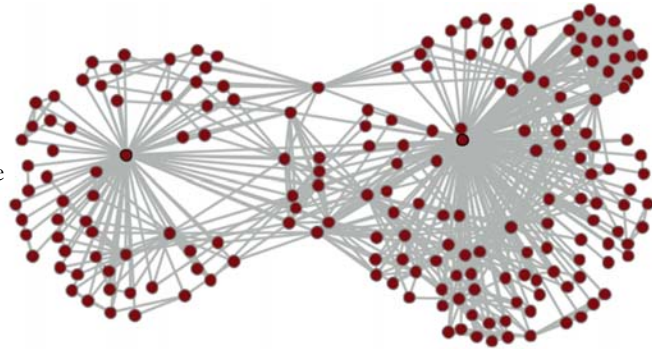


## Network Visualization: Node Layout

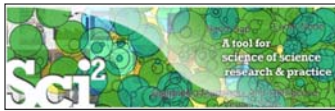
Load and Clean ISI File was selected.  
Loaded 361 records.  
Removed 0 duplicate records.  
Author names have been normalized.  
361 records with unique ISI IDs are available  
via Data Manager.

.....  
Extract Co-Author Network was selected.  
Input Parameters:  
File Format: isi

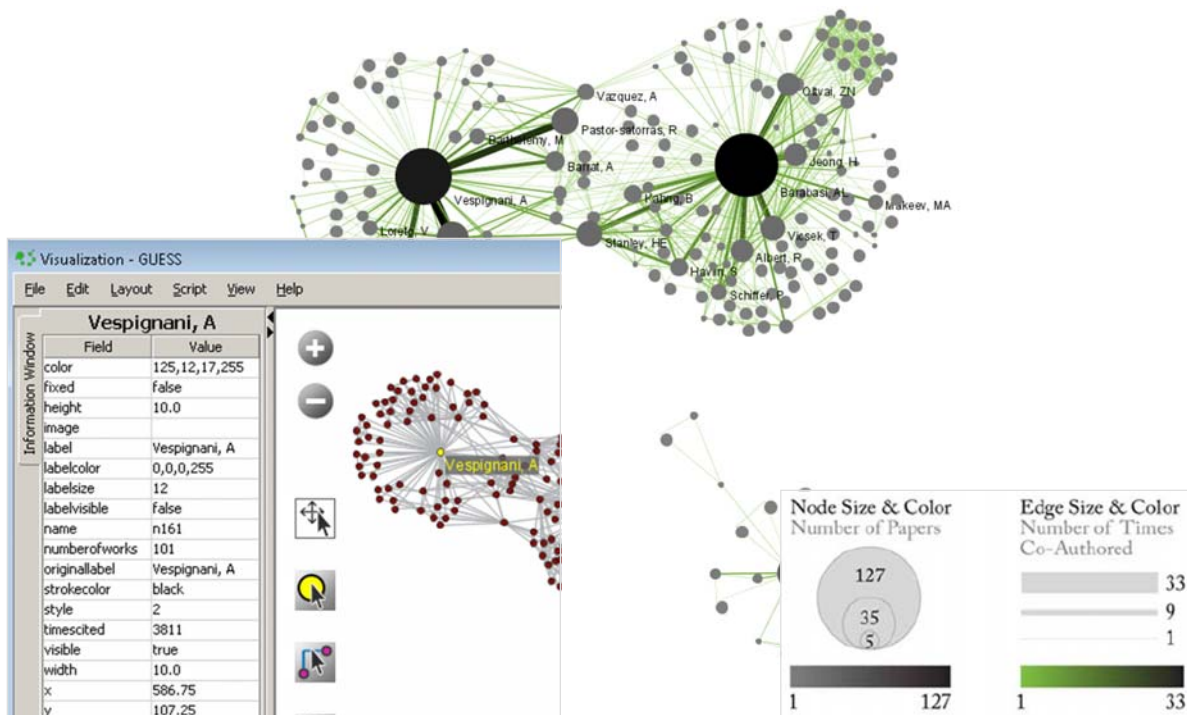
.....  
Network Analysis Toolkit (NAT) was selected.  
Nodes: 247  
Edges: 891  
.....  
GUESS was selected.



45

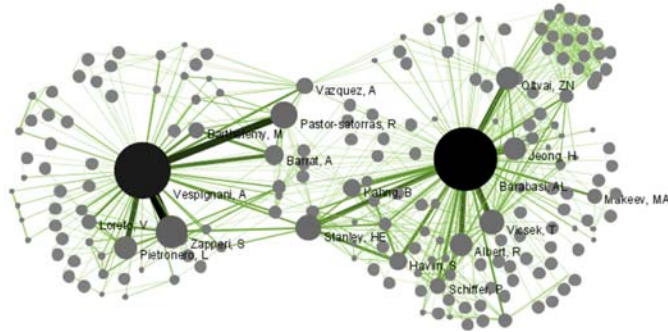


## Network Visualization: Color/Size Coding by Data Attribute Values





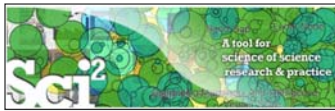
## Network Visualization: Giant Component



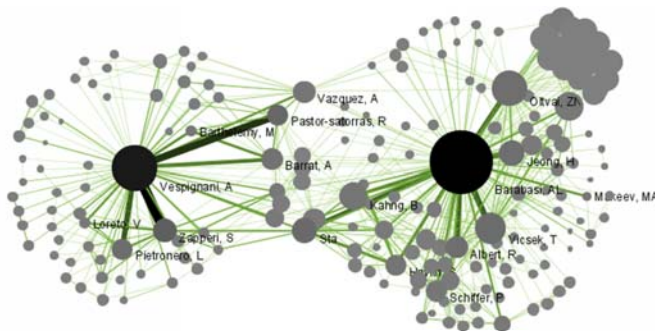
.....  
Weak Component Clustering was selected.  
Implementer(s): Russell Duhon  
Integrator(s): Russell Duhon

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

47



## Network Visualization: Color/Size Coding by Degree



.....  
Node Degree was selected.  
Documentation:  
<https://nwb.slis.indiana.edu/community/?n=AnalyzeData.No deDegree>  
.....

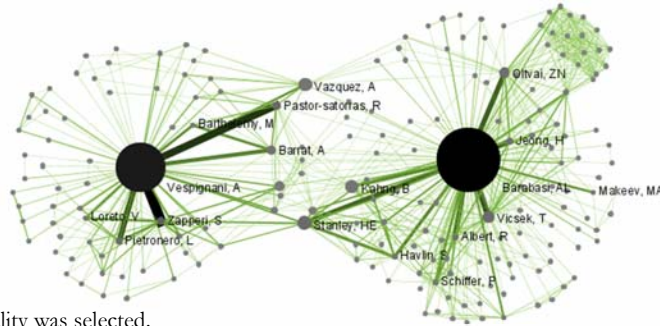


48





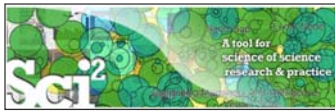
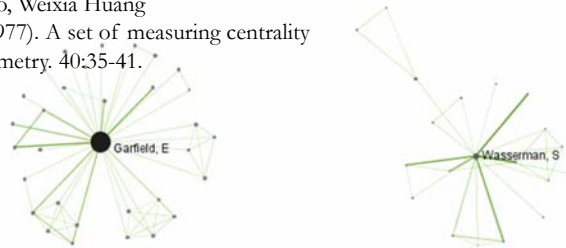
## Network Visualization: Color/Size Coding by Betweenness Centrality



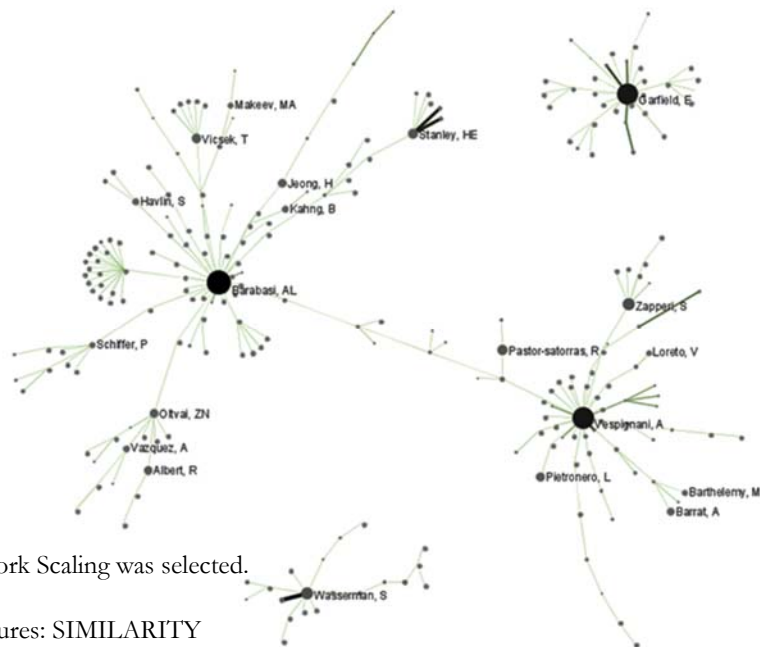
.....  
 Node Betweenness Centrality was selected.  
 Author(s): L. C. Freeman  
 Implementer(s): Santo Fortunato  
 Integrator(s): Santo Fortunato, Weixia Huang  
 Reference: Freeman, L. C. (1977). A set of measuring centrality based on betweenness. Sociometry. 40:35-41.

Input Parameters:  
 Number of bins: 10

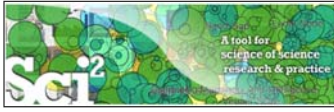
umber of bins: 10  
 .....



## Network Visualization: Reduced Network After Pathfinder Network Scaling

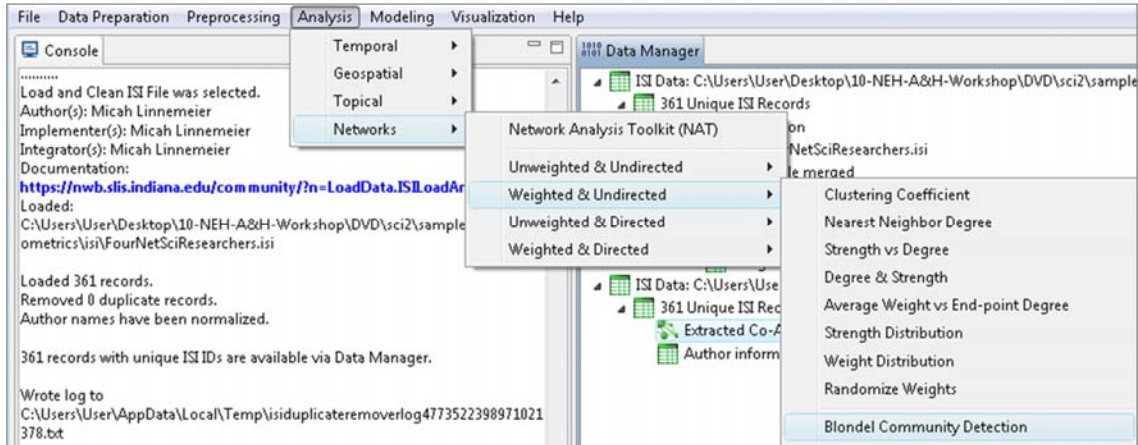


.....  
 MST-Pathfinder Network Scaling was selected.  
 Input Parameters:  
 Weight Attribute measures: SIMILARITY  
 Edge Weight Attribute: weight  
 .....

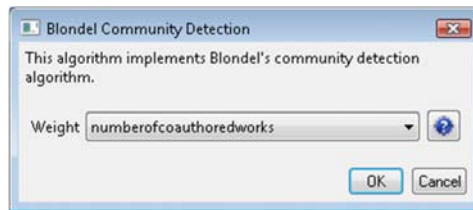


## Network Visualization: Circular Hierarchy Visualization

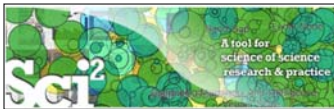
Select Co-Author Network and run Blondel Community detection:



With parameter values

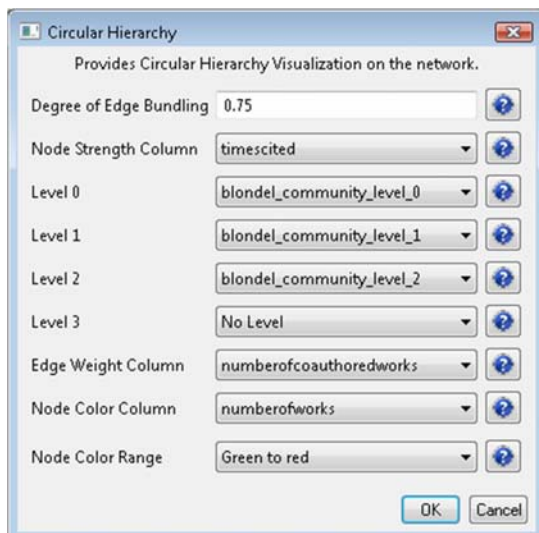


51

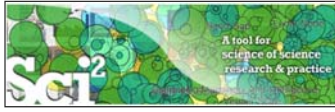


## Network Visualization: Circular Hierarchy Visualization

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



52



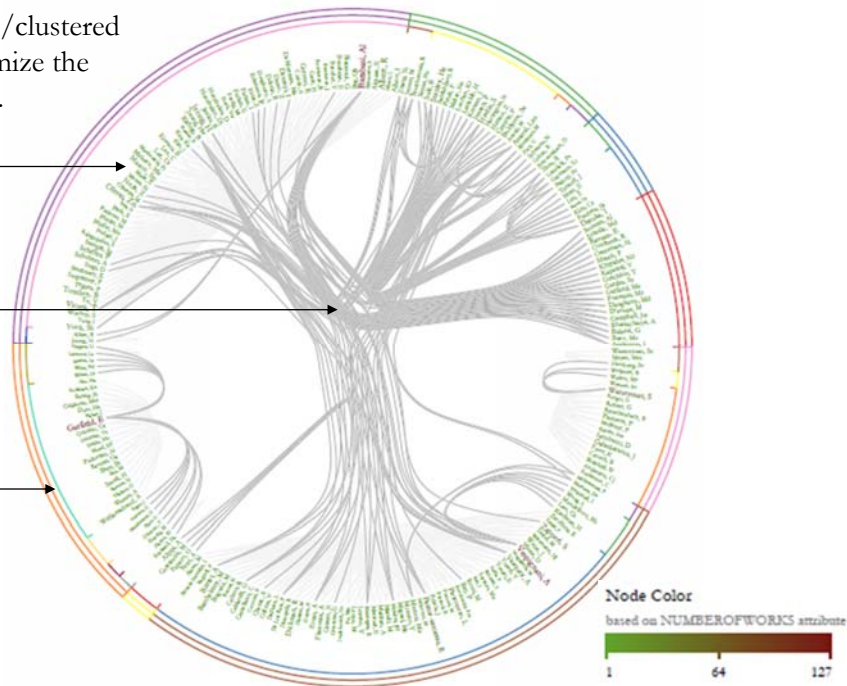
## Network Visualization: Circular Hierarchy Visualization

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.

53



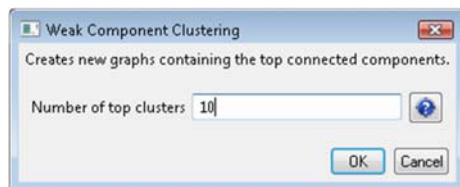
## Paper-Citation Network Layout

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)

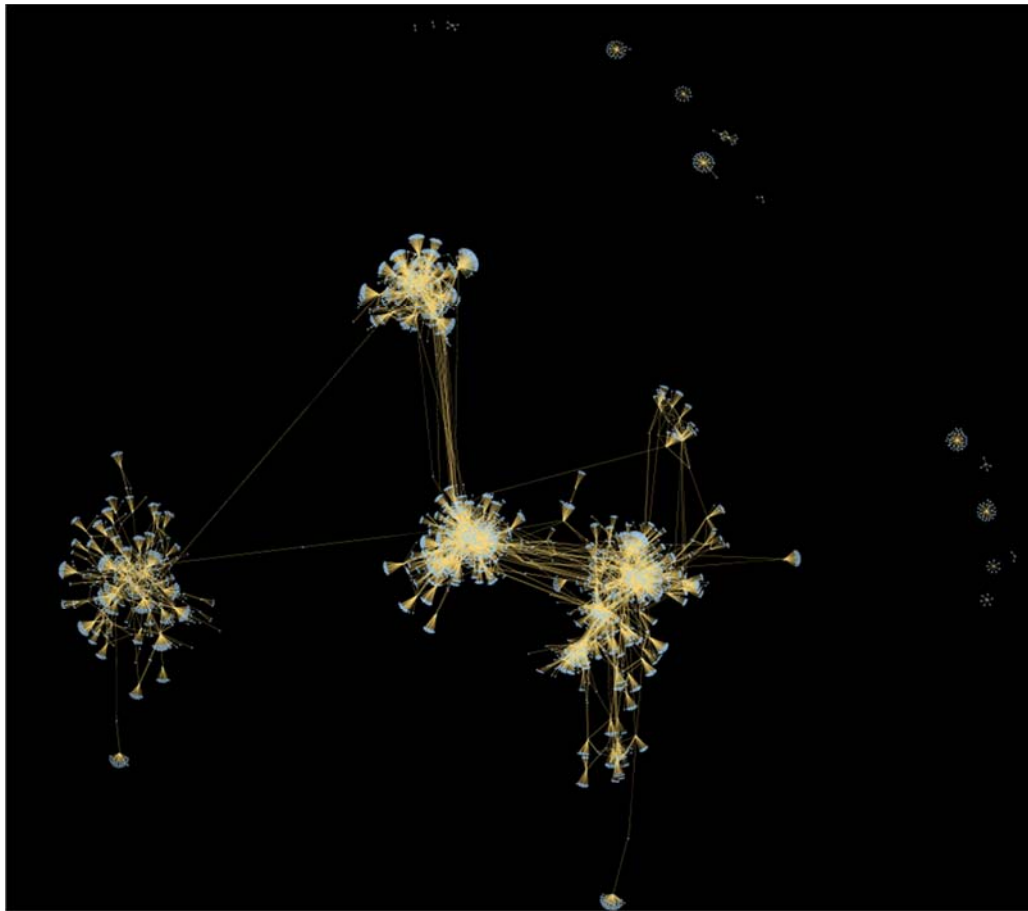
Run *Analysis > Networks > Unweighted and Directed > Weak Component Clustering* with parameters



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*.

54



55



### 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

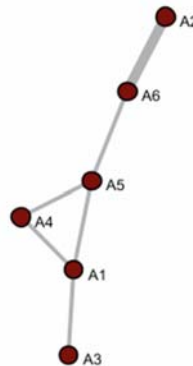


Author co-occurrence network

**Extract Network from Table**  
Extracts a network from a delimited table

Column Name:

Text Delimiter:



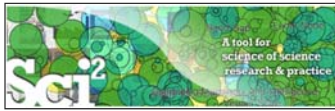
\*Vertices 6

- 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

56



## 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:

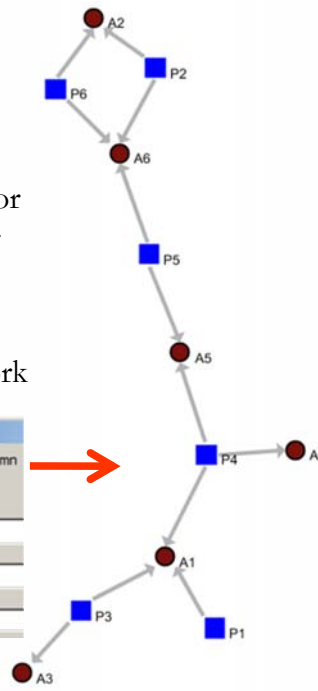
Second column:

Text Delimiter:

---

Object:  Property:  Operator:  Value:

● 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

57



## 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:

Target Column:

Text Delimiter:

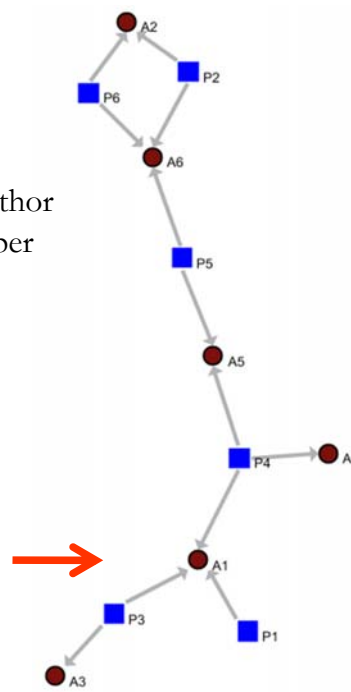
---

Analysis Modeling Visualization R Help

- Temporal
- Geospatial
- Topical
- Networks**
  - Network Analysis Toolkit (NAT)
  - Unweighted & Undirected
  - Weighted & Undirected
  - Unweighted & Directed**

Node Indegree

● Author  
■ Paper



- \*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

58



## 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**

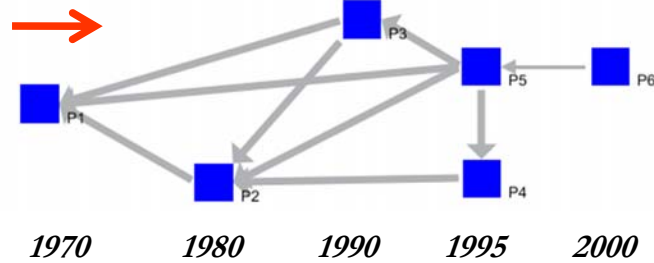
Given a table, this algorithm directed edge that starts at a column node.

Source Column:

Target Column:

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

59



## 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



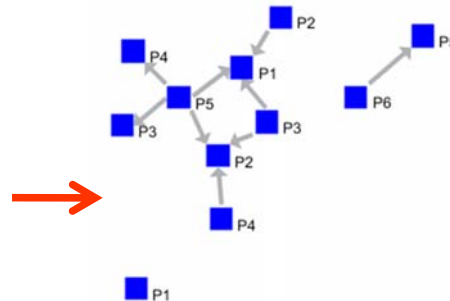
**Extract Bipartite Network**

Extract a bipartite network values may list multiple ent them.

First column:

Second column:

Text Delimiter:



- \*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

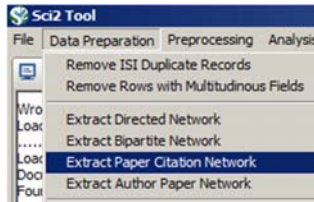
**WRONG!!!**

60

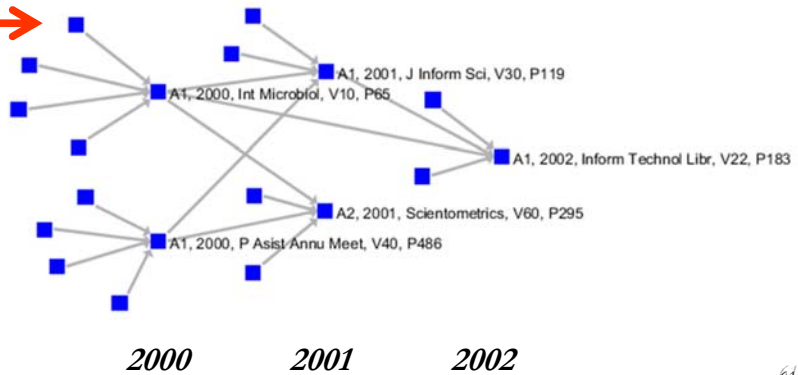


## ISI Paper-Citation Network Extraction

	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 BROI	2000	T1	A1, 2000, INT MICROBIOL, V10, P65
3	A1	BENSMAN SJ, 1999, LIBR RESOUR TECH SER, V42, P147 BROI	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, NATUI	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



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



# Break



## Tutorial Overview

12:30p Welcome and Overview of Tutorial and Attendees

12:45p Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools

1:00p Sci2 Tool Basics

- Download and run the Sci2 Tool
- Load, analyze, and visualize family and business networks
- 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

2:30p Break

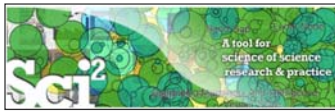
3:00p Sci2 Tool Novel Functionality

- **Yahoo! Geocoder**
- Evolving collaboration networks
- R-Bridge

4:00p Outlook and Q&A

4:30 Adjourn

63



## Yahoo! Geocoding and Geospatial Maps

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

Address Times Cited (TC) → Use Yahoo! Geocoder to identify Latitude, Longitude, Country, ZIP



Aggregate TC over Country



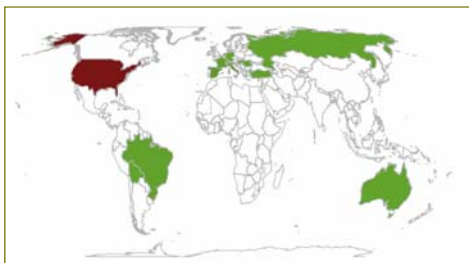
Restrict to US records  
Aggregate TC over ZIP



Use GeoMap (region)



Use GeoMap (circle)



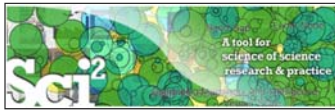
World map colored by sum of TC per country (Choropleth Map)



US Map with circles sized by sum of TC per ZIP code, circle ring colored by count. (Proportional Symbol Map)

64





## Load File with Address and Times Cited Fields

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.

AM	AN	AO	AP	AQ	AR	AS	AT	AU	AV	AW	AX			
1	Publicatio	Publisher	Publisher	Reprint Address		Research	Research	Research	Special	Iss	Subject	Ct	Suppleme	Times Cite
2	2006	BIOMED C	MIDDLESEX	HOUSE, 38	Wuchty, S, Northwestern Univ, NW Inst Complex, Chambers Hall	Northwestern Univ, NW Inst Complex, Ev	Evolutionary	Biolog						7
3	2005	BLACKWEI	9600	GARSINGTON	RD, OXFORD OX4 2DQ, OXON, ENGLAND	Univ Pittsburgh, Dept Pathol, Pittsbu								0
4	2005	BLACKWEI	9600	GARSINGTON	RD, OXFORD OX4 2DQ, OXON, ENGLAND	Univ Notre Dame, Dept Phys, Notre Dam								0
5	2005	NATURE P	MACMILLAN	BUILDING	Barabasi, AL, Harvard Univ, Dana Farber Canc Inst, Ctr Canc Syst Bi	Harvard Univ, Dana Farber Canc Inst, Ctr								2
6	2005	EDP SCIEN	17, AVE DU	HOGGAR,	Macdonald, PJ, Univ Notre Dame, Ctr Network Res, Notre Dame,	Univ Notre Dame, Ctr Network Res, Notre								14
7	2005	NATURE P	MACMILLAN	BUILDING	Oliveira, JG, Univ Notre Dame, Ctr Complex Network Res, Notre	Univ Notre Dame, Ctr Complex Network								15
8	2005	NATL ACA	2101	CONSTITUTION	Olvtai, ZN, Northwestern Univ, Dept Pathol, Chicago, IL 60611 US	Northwestern Univ, Dept Pathol, Chicago								29
9	2005	NATURE P	MACMILLAN	BUILDING	Barabasi, AL, Univ Notre Dame, Ctr Complex Networks Res, Notre	Univ Notre Dame, Ctr Complex Networks								32

2 missing values

65



## Relevant CShell plugin



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

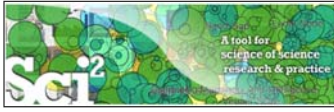
1. The performance is slower than the [Geocoder](#) and may vary due to the network latency since the queries are requested through internet service. The benchmark test geocoded 470 unique locations per minute
2. Yahoo! Geocoder supports address geocoding with international coverage which is not supported 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.

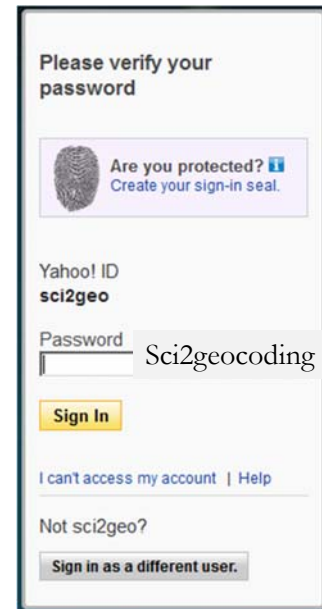
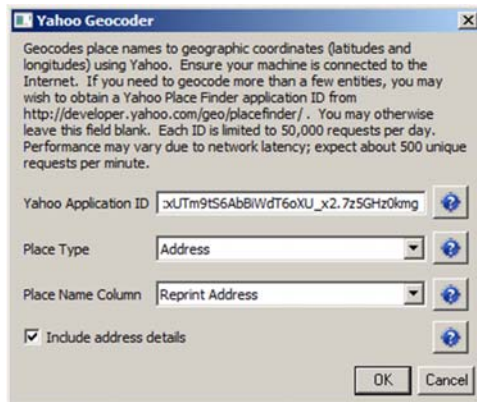
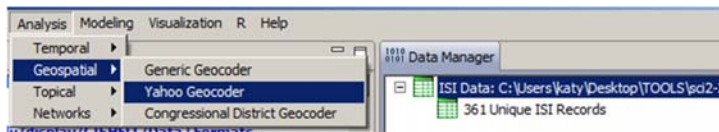
<http://wiki.cns.iu.edu/display/CISHELL/Yahoo+Geocoder>

66



## Geocode Address Using Yahoo! Geocoder

Run *'Analysis > Geospatial > Yahoo Geocoder'*



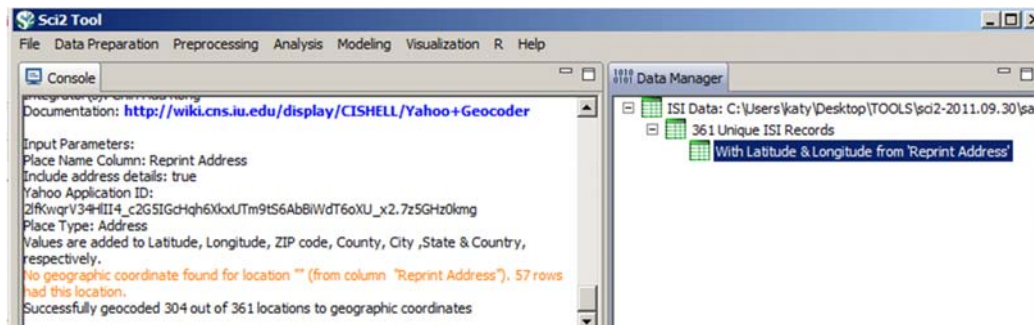
2lfKwqrV34HIII4\_c2G5IGcHqh6XkxUTm9tS6AbBiWdT6oXU\_x2.7z5GHz0kmg

67



## Geocode Address Using Yahoo! Geocoder

Result is a new file in Data Manager

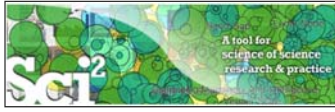


with 7 new columns

BF	BG	BH	BI	BJ	BK	BL
Latitude	Longitude	ZIP code	County	City	State	Country
42.02946091	-87.68838501	60202	Cook Cour	Evanston	Illinois	United States
42.34999466	-71.08765411	2115	Suffolk Cc	Boston	Massachu	United States
41.70074844	-86.23918915	46556	St. Joseph St.	Marys	Indiana	United States
41.70074844	-86.23918915	46556	St. Joseph St.	Marys	Indiana	United States
41.89422607	-87.61901855	60611	Cook Cour	Chicago	Illinois	United States
41.70074844	-86.23918915	46556	St. Joseph St.	Marys	Indiana	United States
41.70074844	-86.23918915	46556	St. Joseph St.	Marys	Indiana	United States
41.70074844	-86.23918915	46556	St. Joseph St.	Marys	Indiana	United States

**2 missing values, there is a total of 57 missing in these 361 records**

68



## Console Messages

Load... was selected.  
 Documentation: <http://wiki.cns.iu.edu/display/CISHELL/Data+Formats>  
 Found old-style ISI/Web Of Knowledge file.  
 Found old-style ISI/Web Of Knowledge file.  
 Found old-style ISI/Web Of Knowledge file.  
 Found old-style ISI/Web Of Knowledge file.

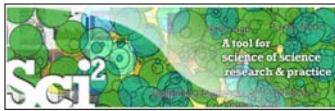
Loaded 361 records.  
 Removed 0 duplicate records.  
 Author names have been normalized.

361 records with unique ISI IDs are available via Data Manager.

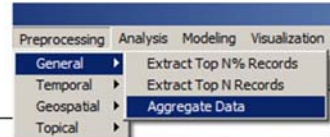
Wrote log to C:\Users\katy\AppData\Local\Temp\isiduplicateremoverlog2749984290663814202.txt  
 Loaded: C:\Users\katy\Desktop\TOOLS\sci2-2011.09.30\sampladata\scientometrics\isi\FourNetSciResearchers.isi  
 .....

Yahoo Geocoder was selected.  
 Implementer(s): Chin Hua Kong  
 Integrator(s): Chin Hua Kong  
 Documentation: <http://wiki.cns.iu.edu/display/CISHELL/Yahoo+Geocoder>

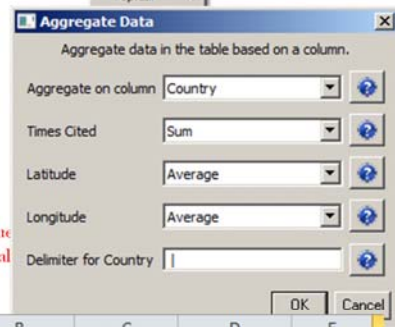
Input Parameters:  
 Place Name Column: Reprint Address  
 Include address details: true  
 Yahoo Application ID: 2lfKwqrV34fHIII4\_c2G5IGcHqh6XkxUTm9tS6AbBiWdT6oXU\_x2.7z5GHZ0kmg  
 Place Type: Address  
 Values are added to Latitude, Longitude, ZIP code, County, City, State & Country, respectively.  
 No geographic coordinate found for location "" (from column "Reprint Address"): 57 rows had this location.  
 Successfully geocoded 304 out of 361 locations to geographic coordinates



## 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

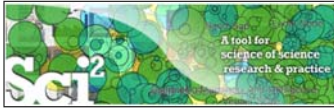


Aggregated by ": All rows of Latitude column were skipped due to no non-null, non-empty value  
 Aggregated by ": All rows of Longitude column were skipped due to no non-null, non-empty value  
 Frequency of unique "Country" values added to "Count" column.

	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



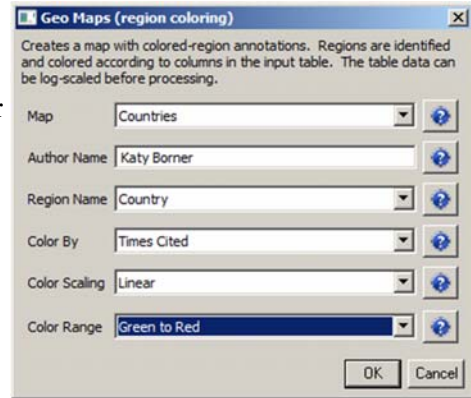
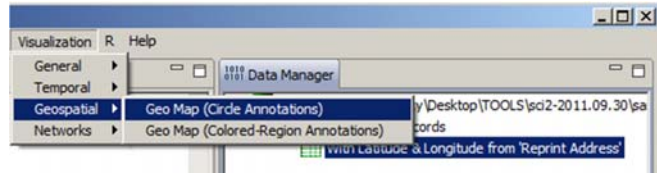
	A	B	C	D	E
1	Times Cited	Latitude	Longitude	Country	Count
2	14680	[41.10645f]	[-82.45309f]	United States	194
3	1802				57
4	398	[47.506226f]	[19.06482f]	Hungary	14
5	101	[37.25198f]	[127.08451f]	South Korea	4
6	18	[32.08439f]	[34.81297f]	Israel	1
7	57	[46.768517f]	[23.585135f]	Romania	2
8	55	[47.06615f]	[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



# Aggregate + Geo Map (Colored-Region Annotations) *Also called 'Choropleth Map'*

Geo Map (Colored-Region Annotations) was selected.  
Author(s): Joseph R. Biberstine  
Implementer(s): Joseph R. Biberstine  
Integrator(s): Joseph R. Biberstine  
Documentation: <http://wiki.cns.iu.edu/display/CISHELL/Geo+Map>

Input Parameters:  
Author Name: Katy Borner  
Color By: Times Cited  
Region Name: Country  
Map: Countries  
Color Range: Green to Red  
Color Scaling: Linear  
Printing PostScript.  
Saved: C:\Users\katy\Desktop\geoMaps5661700855281902370.ps



Save map to Desktop, view with PostScript viewer or convert into pdf to view.

Vector format - can be rendered LARGE.

71

## How to Read the Geo Map

World Map with data overlay.

Listing of map type, author, and parameters used.

Map legend with color coding.

Header and footer with information when this map was created, by whom and using what data set.

Geo Map (Colored-Region Annotation Style)  
Eckert IV Projection  
Feb 15, 2012 | 05:51:44 AM  
Katy Borner

Region Color (Linear)  
Times Cited  
0 7,250 14,500

72



## Relevant Sci2 Manual entry



### 5.2.4 Mapping Scientometrics (ISI Data)

Edit Add Tools

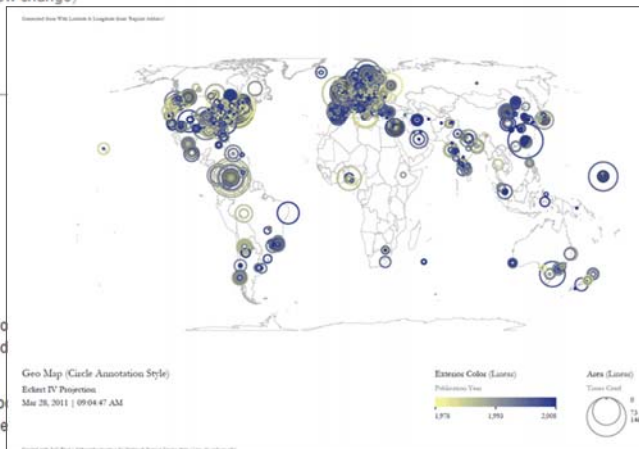
9 Added by Ted Polley, last edited by Ted Polley on Nov 14, 2011 (view change)

#### 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."

<http://wiki.cns.iu.edu/display/SCI2TUTORIAL/5.2.4+Mapping+Scientometrics+%28ISI+Data%29>

73

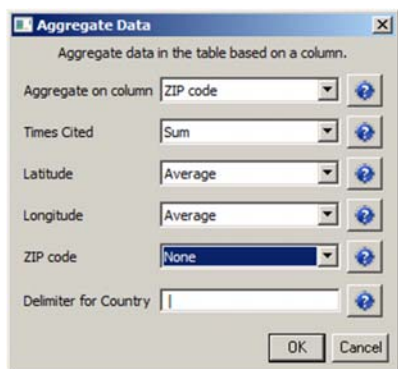


## Create Geo Map (Circle Annotations)

Also called 'Proportional Symbol Map'

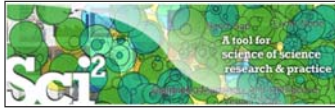
Restrict to 195 US records

Aggregate by ZIP code



	A	B	C	D	E	F
1	Times Cited	latitude	Longitude	ZIP code	Country	Count
2	7	42.02946091	-87.68838501	[60202]	United States	1
3	2	42.34999466	-71.08765411	[2115]	United States	1
4	9438	41.70074844	-86.23918915	[46556]	United States	66
5	316	41.89422607	-87.61901855	[60611]	United States	5
6	2	41.11500168	-85.73377991	[46566]	United States	1
7	7	42.17333984	-77.98394775	[14880]	United States	3
8	34	41.69326782	-86.26696777	[46616]	United States	1
9	222	45.42283363	-107.4489517	[ ]	United States	5
10	2076	44.97523499	-93.23405075	[55455]	United States	2
11	18	40.7978363	-77.86255646	[16802]	United States	1
12	961	42.3462112	-71.10174036	[2215]	United States	16
13	208	42.35863495	-71.05670166	[2109]	United States	5
14	13	39.95228958	-75.16243235	[19107]	United States	3
15	601	39.95987072	-75.19809494	[19104]	United States	40
16	0	48.38810349	-115.5551758	[59923]	United States	1
17	3	32.22143936	-110.9711914	[85701]	United States	1

74



## Create Geo Map (Circle Annotations)

*Also called 'Proportional Symbol Map'*

Geo Map (Circle Annotations) was selected.

Author(s): Joseph R. Biberstine

Implementer(s): Joseph R. Biberstine

Integrator(s): Joseph R. Biberstine

Documentation: <http://wiki.cns.iu.edu/display/CISHELL/Geo+Map>

Input Parameters:

Longitude: Longitude

Size Circles By: Times Cited

Color Circle Exteriors By: Count

Color Circle Interiors By: None (no inner color)

Exterior Color Scaling: Linear

Exterior Color Range: Yellow to Blue

Interior Color Range: Yellow to Blue

Size Scaling: Linear

Map: US States

Author Name:

Interior Color Scaling: Linear

Latitude: Latitude

Printing PostScript..

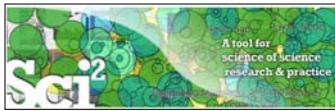
Done.

Saved: C:\Users\katy\Desktop\geoUS.ps

Save ps file, convert to pdf, view.



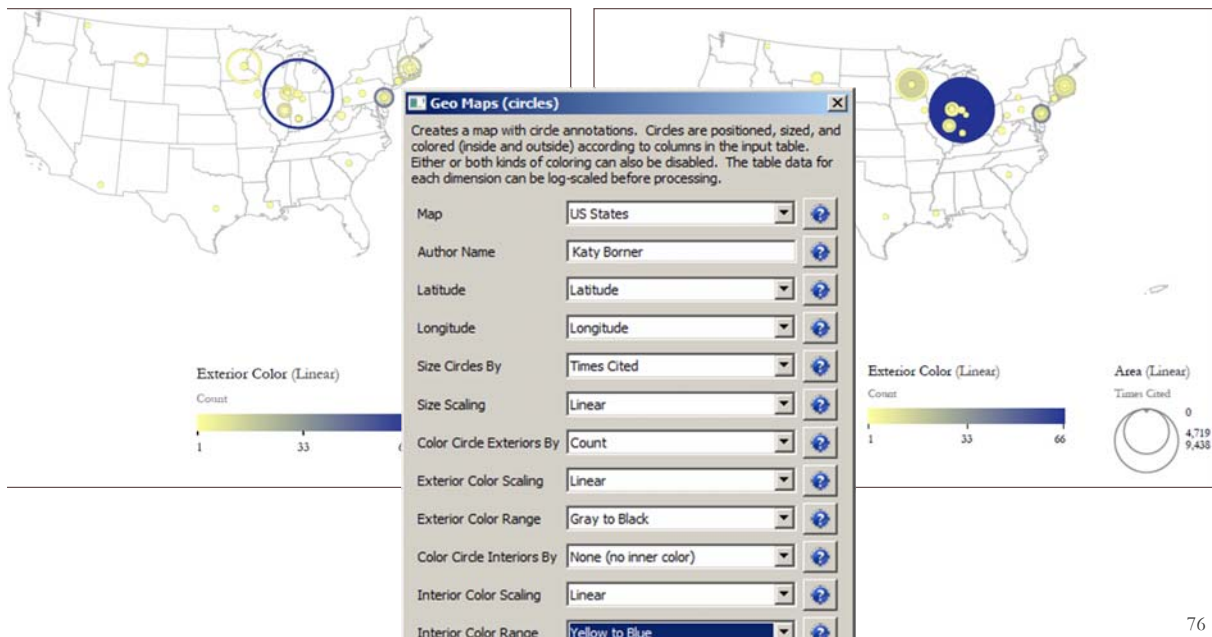
75



## Create Geo Map (Circle Annotations)

*Also called 'Proportional Symbol Map'*

Different parameter values result in different maps and corresponding legends:



76



## Tutorial Overview

12:30p Welcome and Overview of Tutorial and Attendees

12:45p Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools

1:00p Sci2 Tool Basics

- Download and run the Sci2 Tool
- Load, analyze, and visualize family and business networks
- 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

2:30p Break

3:00p Sci2 Tool Novel Functionality

- Yahoo! Geocoder
- Evolving collaboration networks
- R-Bridge

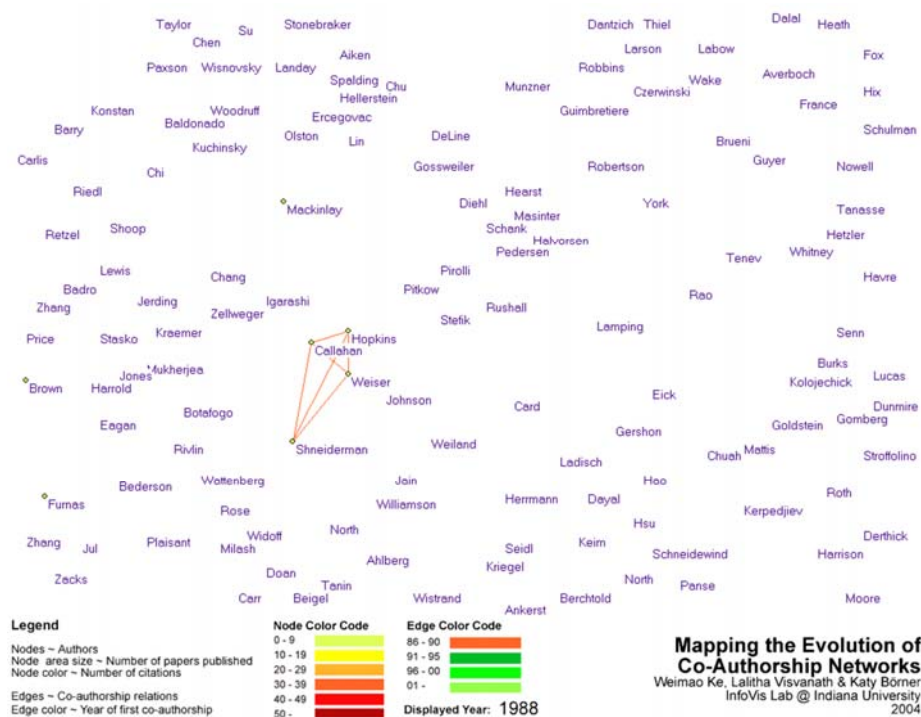
4:00p Outlook and Q&A

4:30 Adjourn

77



## Sci2 Demo II: Evolving collaboration networks

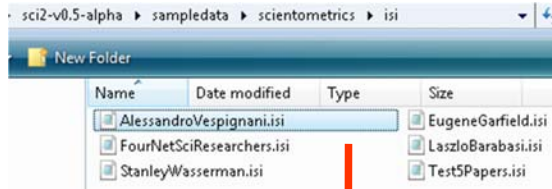


78



## Evolving Collaboration Networks

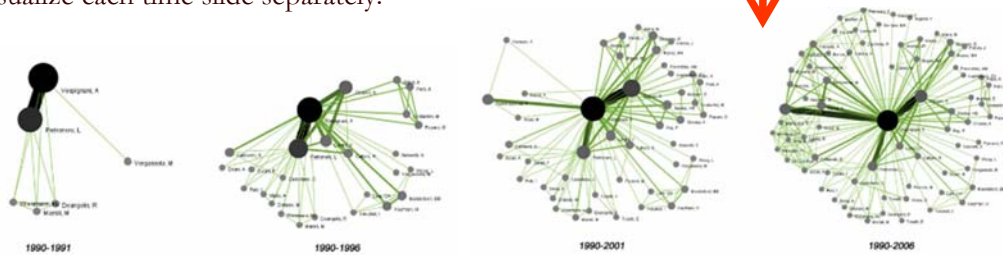
Load isi formatted file



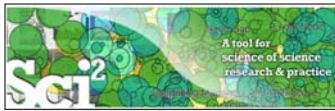
As csv, file looks like:

	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:



79



## 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 Sci<sup>2</sup> 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?

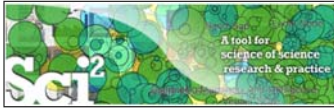
Align With Calendar

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))

80





## 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.

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

Choosing "Years" under "Slice Into" creates multiple tables beginning from January 1<sup>st</sup> 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))

81



## Visualize Each Network, Keep Node Positions

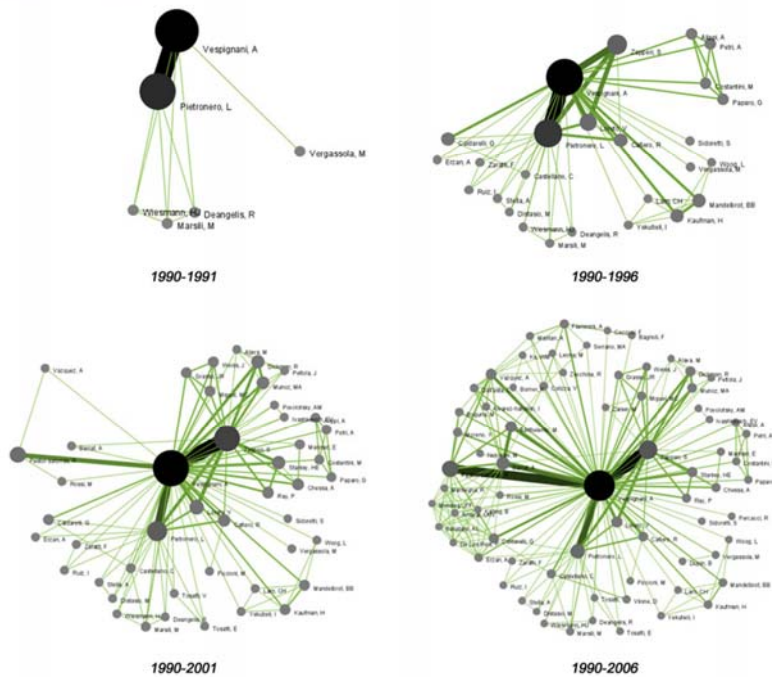
1. To see the evolution of Vespignani's co-authorship network over time, check *'cumulative'*.
2. Extract co-authorship networks one at a time for each sliced time table using *'Data Preparation > Extract Co-Author Network'*, making sure to select "ISI" from the pop-up window during the extraction.
3. 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))

82



## 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))

83



## Relevant CShell plugin



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>

84



## Tutorial Overview

12:30p Welcome and Overview of Tutorial and Attendees

12:45p Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools

1:00p Sci2 Tool Basics

- Download and run the Sci2 Tool
- Load, analyze, and visualize family and business networks
- 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

2:30p Break

3:00p Sci2 Tool Novel Functionality

- Yahoo! Geocoder
- Evolving collaboration networks
- **R-Bridge**

4:00p Outlook and Q&A

4:30 Adjourn

85



## 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.

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.

86



## Tutorial Overview

12:30p Welcome and Overview of Tutorial and Attendees

12:45p Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools

1:00p Sci2 Tool Basics

- Download and run the Sci2 Tool
- Load, analyze, and visualize family and business networks
- 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

2:30p Break

3:00p Sci2 Tool Novel Functionality

- Yahoo! Geocoder
- Evolving collaboration networks
- R-Bridge

4:00p Outlook and Q&A

4:30 Adjourn

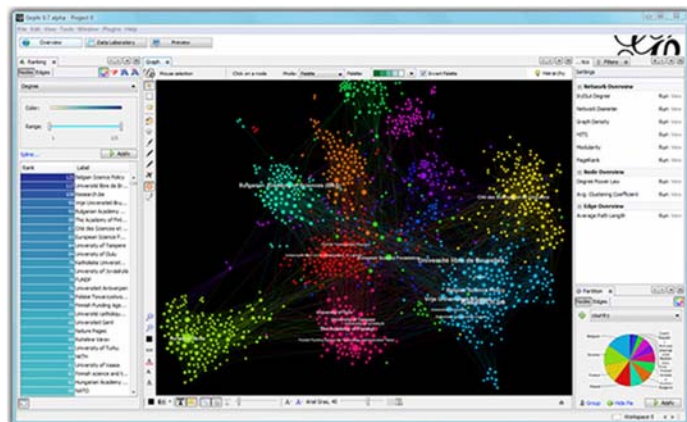
87



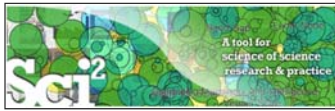
## New Sci2 Release Preview

The next Sci2 release will become available in April 2012. It will add

- New ISI file format reader
- Publication and funding database support
- Improved geographic map legend
- New 10-year Global Map of Science (formerly UCSD Map)
- R-bridge
- Gephi.org graph layout tool



88

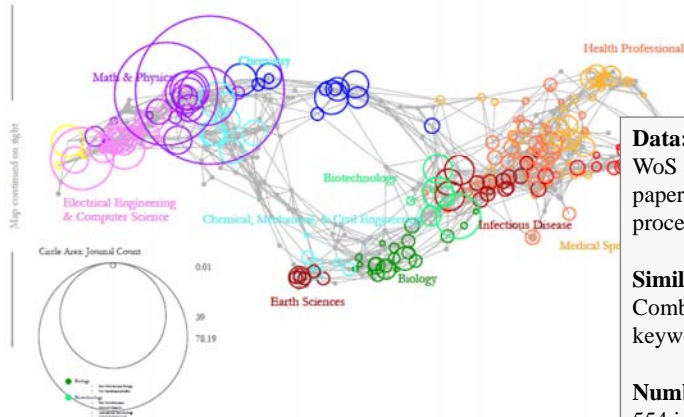


## Topic Mapping: UCSD Science Map

Science Map via Journals for FourNetSciResearchers.isi

314 journal references matched out of 361 found.

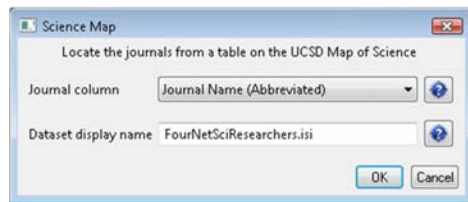
These 314 references are associated with 13 of 13 disciplines of science and 255 of 554 research specialties in the UCSD Map of Science.



**Data:**  
WoS and Scopus for 2001–2005, 7.2 million papers, more than 16,000 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.

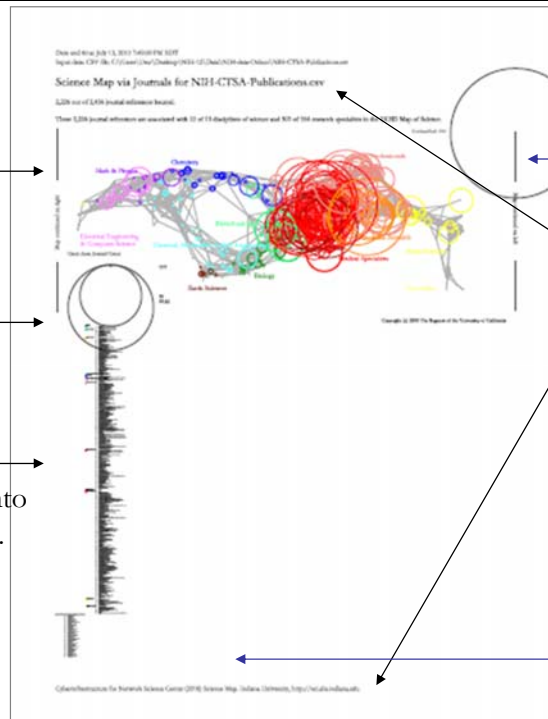


## How to Read the UCSD Map


UCSD Science Map with data overlay.

Map legend of circle area size coding


Listing of all data records organized into UCSD science areas.

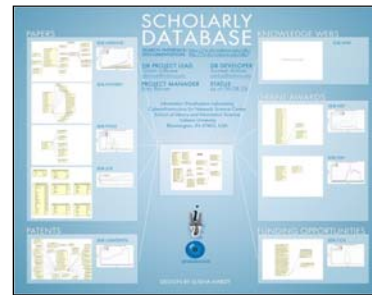



## Computational Scientometrics Cyberinfrastructures

 **Scholarly Database: 25 million scholarly records**  
<http://sdb.slis.indiana.edu>




 **VIVO Research Networking**  
<http://vivoweb.org>

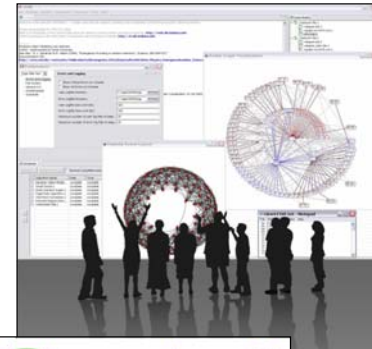


 **Information Visualization Cyberinfrastructure**  
<http://iv.cns.iu.edu>

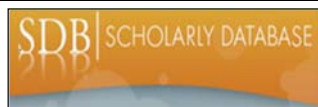
 **Network Workbench Tool & Community Wiki**  
<http://nwb.cns.iu.edu>

 **Science of Science (Sci<sup>2</sup>) Tool**  
<http://sci2.cns.iu.edu>

 **Epidemics Tool & Marketplace**  
**Forthcoming**



91

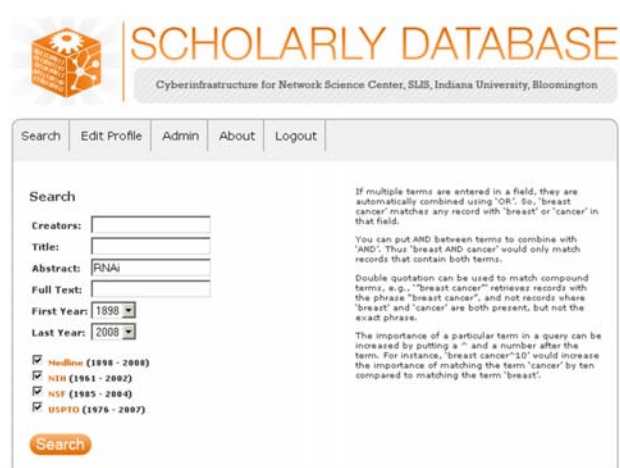


## Scholarly Database at Indiana University

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

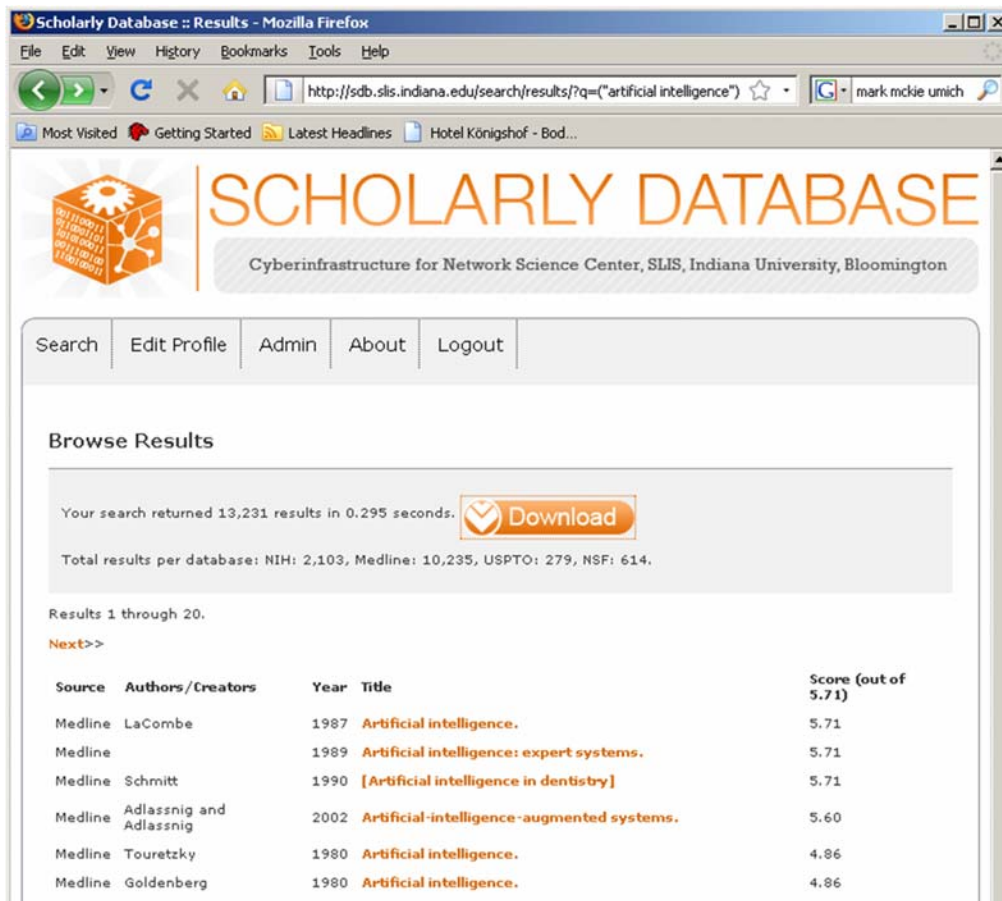
Supports federated search of 25 million publication, patent, grant records.

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

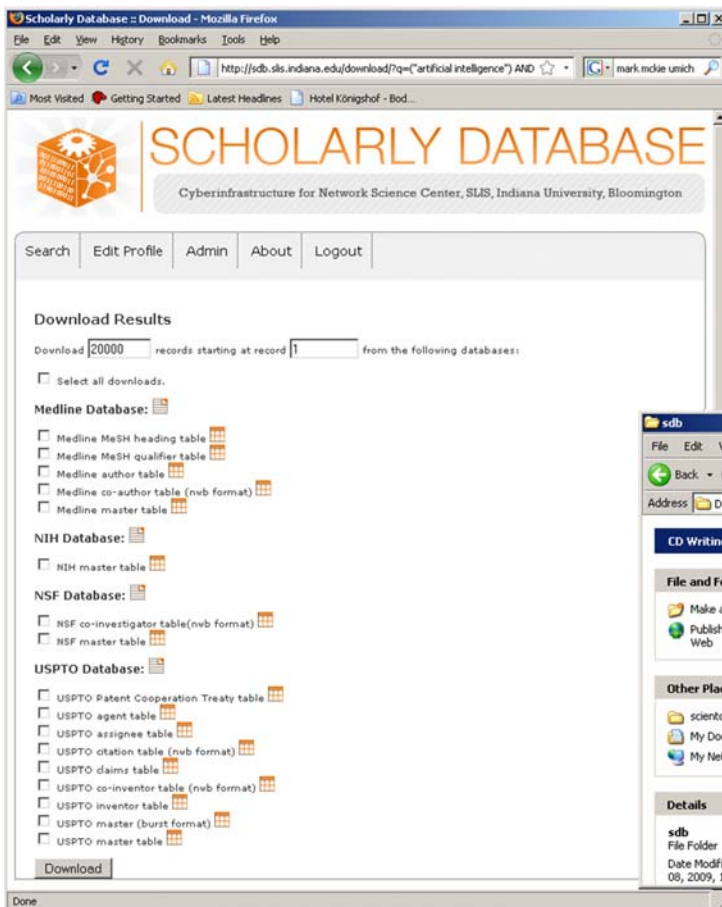


Register for free access at <http://sdb.cns.iu.edu>

92



93

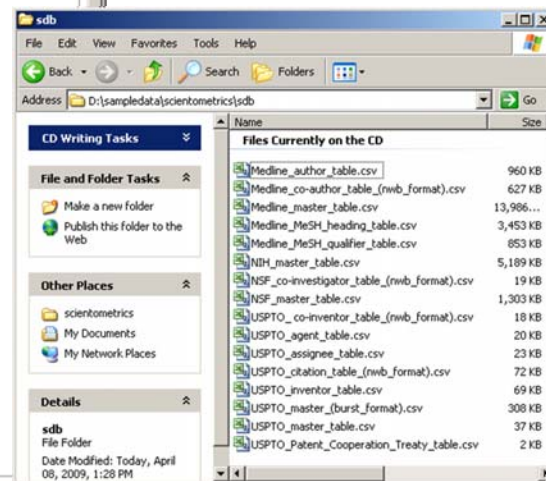


### Since March 2009:

Users can download networks:

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

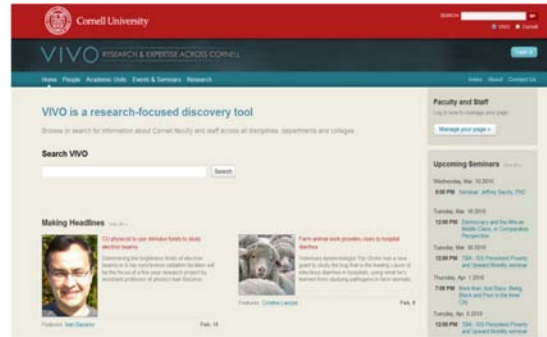
and tables for burst analysis in NWB.



94

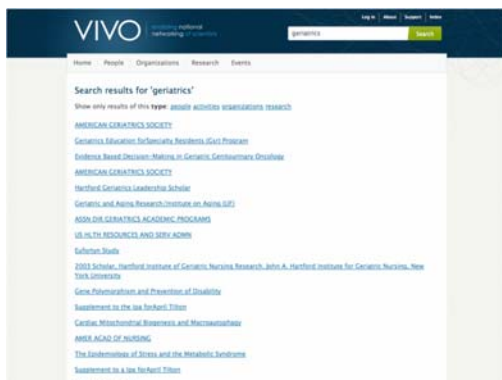
## VIVO: A Semantic Approach to Creating a National Network of Researchers (<http://vivoweb.org>)

- Semantic web application and ontology editor originally developed at Cornell U.
- Integrates research and scholarship info from systems of record across institution(s).
- Facilitates research discovery and cross-disciplinary collaboration.
- Simplify reporting tasks, e.g., generate biosketch, department report.



Funded by \$12 million NIH award.

**Cornell University:** Dean Krafft (Cornell PI), Manolo Bevia, Jim Blake, Nick Cappadona, Brian Caruso, Jon Corson-Rikert, Elly Cramer, Medha Devare, John Ferreira, Brian Lowe, Stella Mitchell, Holly Mistlebauer, Anup Sawant, Christopher Westling, Rebecca Younes. **University of Florida:** Mike Conlon (VIVO and UF PI), Cecilia Botero, Kerry Britt, Erin Brooks, Amy Buhler, Ellie Bushhousen, Chris Case, Valrie Davis, Nita Ferree, Chris Haines, Rae Jesano, Margeaux Johnson, Sara Kreinest, Yang Li, Paula Markes, Sara Russell Gonzalez, Alexander Rockwell, Nancy Schaefer, Michele R. Tennant, George Hack, Chris Barnes, Narayan Raum, Brenda Stevens, Alicia Turner, Stephen Williams. **Indiana University:** Katy Borner (IU PI), William Barnett, Shanshan Chen, Ying Ding, Russell Duhon, Jon Dunn, Micah Linnemeier, Nianli Ma, Robert McDonald, Barbara Ann O'Leary, Mark Ping, Yuyin Sun, Alan Walsh, Brian Wheeler, Angela Zoss. **Ponce School of Medicine:** Richard Noel (Ponce PI), Ricardo Espada, Damaris Torres. **The Scripps Research Institute:** Gerald Joyce (Scripps PI), Greg Dunlap, Catherine Dunn, Brant Kelley, Paula King, Angela Murrell, Barbara Noble, Cary Thomas, Michaelen Trimarchi. **Washington University, St. Louis:** Rakesh Nagarajan (WUSTL PI), Kristi L. Holmes, Sunita B. Koul, Leslie D. McIntosh. **Weill Cornell Medical College:** Curtis Cole (Weill PI), Paul Albert, Victor Brodsky, Adam Cheriff, Oscar Cruz, Dan Dickinson, Chris Huang, Itay Klaz, Peter Michelini, Grace Migliorisi, John Ruffing, Jason Specland, Tru Tran, Jesse Turner, Vinay Varughese.





### University of Florida

How do you want to compare?  
by Grants

Who do you want to compare?  
Search:  X

Records 1 - 10 of 30 < First < Prev Next > Last >

Entity Label	Grant Count	Entity Type
<input checked="" type="checkbox"/> Continuing Education	562	UF Department, Agent, Non-Academic Department, Department
<input checked="" type="checkbox"/> Florida Museum of Natural History	203	Museum, Agent
<input checked="" type="checkbox"/> College of Agricultural and Life Sciences	166	Agent, UF College, College
<input checked="" type="checkbox"/> College of Engineering	103	Agent, UF College, College
<input checked="" type="checkbox"/> Evelyn F. and William L. McKnight Brain Institute of the University of Florida	64	UF Center, Agent, Center
<input checked="" type="checkbox"/> International Center	54	UF Department, Agent, Non-Academic Department, Department
<input checked="" type="checkbox"/> Florida Sea Grant	44	UF Center, Agent, Center
<input type="checkbox"/> Whitney Laboratory for Marine Bioscience	42	UF Research Laboratory, Agent, Laboratory, Research Laboratory
<input type="checkbox"/> Water Institute	38	UF Center, Agent, Center
<input type="checkbox"/> College of Dentistry	35	Agent, UF College, College

[Save as CSV](#) [Clear](#)

### Comparing Grants of Organizations in University of Florida

**Total Number of Grants**  
You have selected 7 of a maximum 10 organizations to compare. [Clear](#)

- Florida Sea Grant 44
- International Center 54
- Evelyn F. and William L. McKnight Brain Institute of the University of Florida 64
- College of Engineering 103
- College of Agricultural and Life Sciences 166
- Florida Museum of Natural History 203
- Continuing Education 562

**Temporal Analysis (When)** Temporal visualizations of the number of papers/funding award at the institution, school, department, and people level

enabling national networking of scientists

Index Log in

Search

Home
People
Organizations
Research
Events

### University of Florida ?

Explore 487 publications activity across 554 scientific sub-disciplines ?

[13 Disciplines](#) | 554 Sub-Disciplines ?

Search:  X

1 - 13 of 554 < First < Prev Next > Last >

Sub-Disciplines	# of pubs.	% activity
Pest Management Science	24.2	5.0
Wildlife Research	19.1	3.9
Protein Science	13.1	2.7
Clinical Cancer Research	12.6	2.6
Pain	12.0	2.5
Environmental Contamination	11.2	2.3
Insect Physiology	11.1	2.3
Organic Chemistry	10.9	2.2
Marine Biology	10.3	2.1
Computer Aided Molecular Design	10.2	2.1
BioStatistics	9.0	1.9

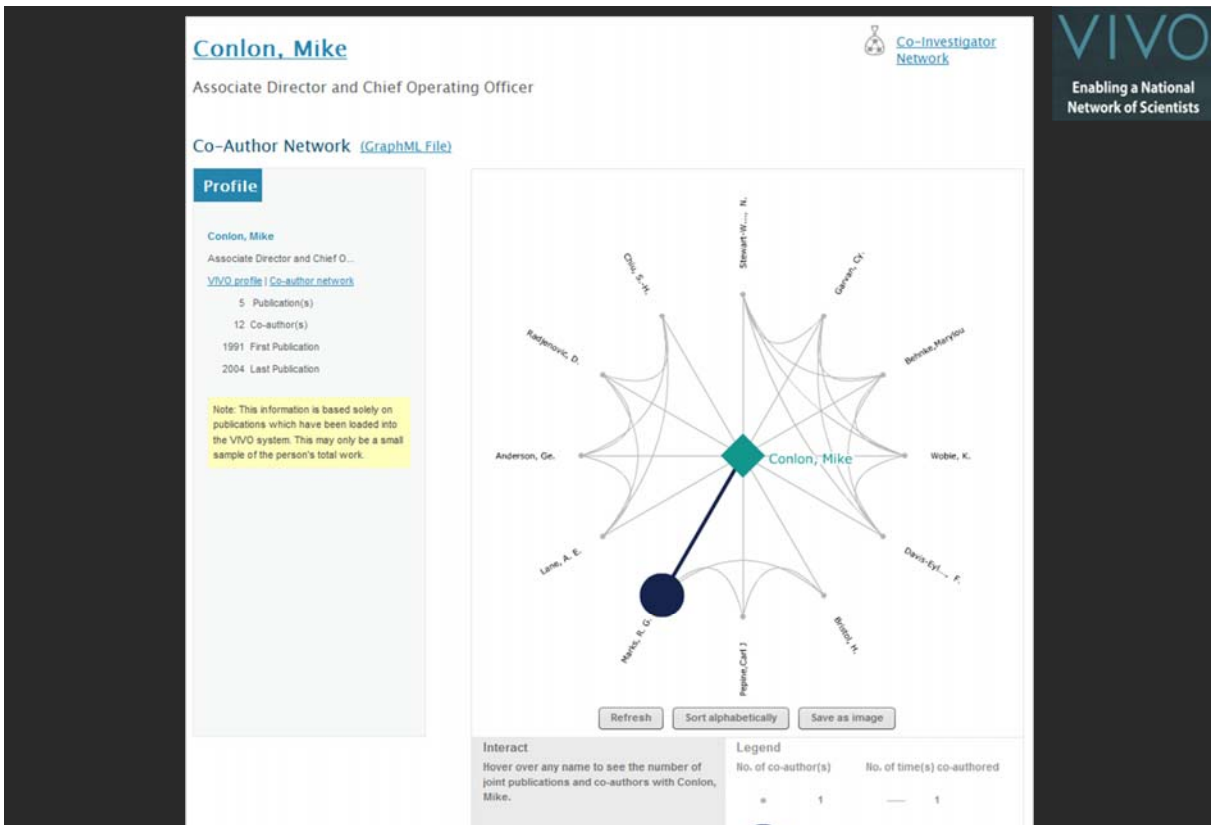
Show discipline labels

Top 290 disciplines shown

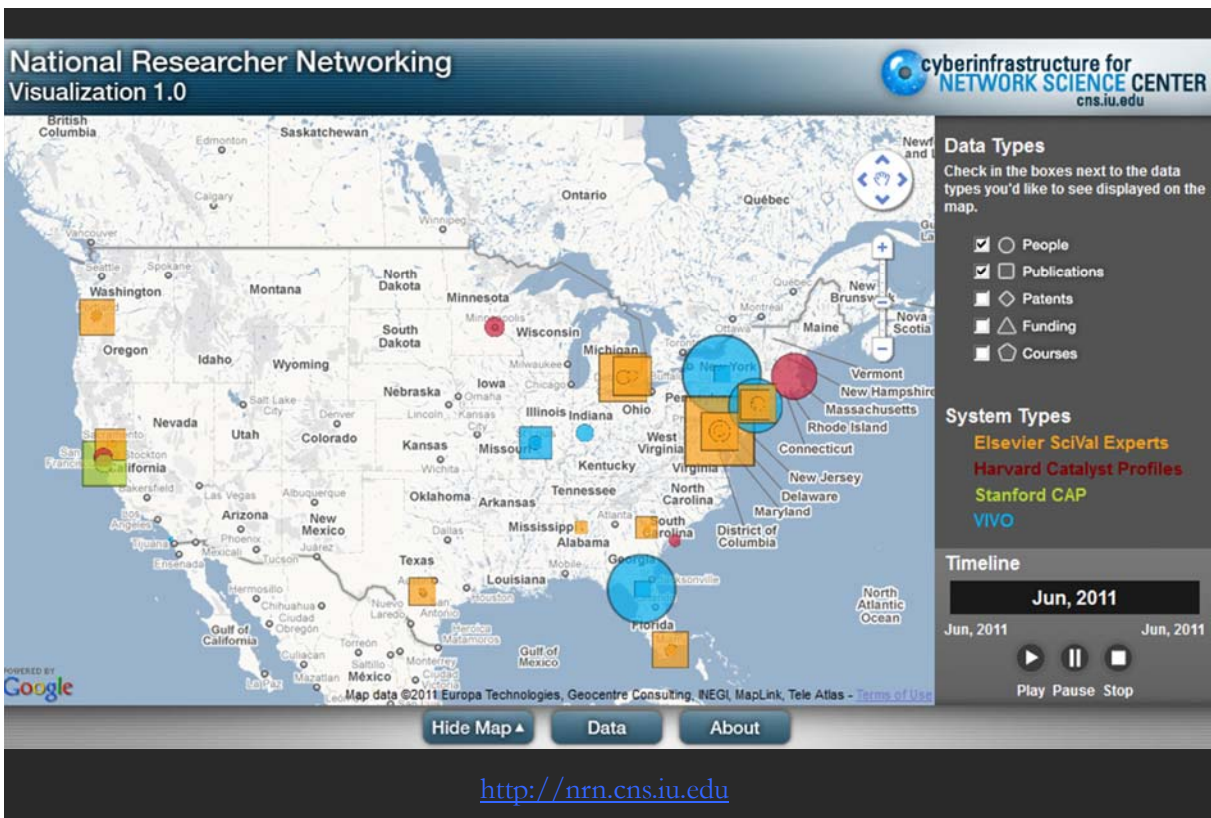
[Terms of Use](#)

mapped 14.55% of 3,346 publications ?

**Topical Analysis (What)** Science map overlays will show where a person, department, or university publishes most in the world of science. (in work)



**Network Analysis (With Whom?)** Who is co-authoring, co-investigating, co-inventing with whom? What teams are most productive in what projects? 99



**Geospatial Analysis (Where)** Where is what science performed by whom? Science is global and needs to be studied globally. 100

**Borner, Katy**  
Person

This information is based solely on publications which have been loaded into the VIVO system. This may only be a small sample of the person's total work.

**General Statistics**

- 35 publication(s) from 2001 to 2010 [\(.CSV File\)](#)
- 87 co-author(s) from 2001 to 2010 [\(.CSV File\)](#)

**Co-Author Network** [\(GraphML File\)](#)

**Legend**

No. of publication(s) | No. of time(s) co-authored

**Interact**

Hover over any name to see the number of past publications and co-author with Borner, Katy. Click on a name to see details on the right.

**Disambiguation**

Only people that co-authored more than 1 paper(s) with Borner, Katy are shown.

**Tables**

Publications per Year [\(.CSV File\)](#)

Year	Count
2001	2
2002	4
2003	2
2004	7
2005	7
2006	3
2007	10
2010	1

Co-author Publications with Borner, Katy [\(.CSV File\)](#)

Author	Count
Chen C.	5
Boyack K.W.	4
Mane K.K.	4
Ka W.	3
Penumarthy S.	3
Vespijnani, Alessandro	2
Hart B.	2
Hart E.	2
Holloway T.	2
Hart S.W.	2
Thakur S.	2
Feng Y.	2
Mane H.	2

## Download Data

### General Statistics

- 36 publication(s) from 2001 to 2010 [\(.CSV File\)](#)
- 80 co-author(s) from 2001 to 2010 [\(.CSV File\)](#)

### Co-Author Network

[\(GraphML File\)](#)

Save as Image (.PNG file)

### Tables

- Publications per year [\(.CSV File\)](#)
- Co-authors [\(.CSV File\)](#)

[http://vivo-netsci.cns.iu.edu/vivo/visualization?uri=http%3A%2F%2Fvivo-trunk.indiana.edu%2Findividual%2FPerson74&vis=person\\_level&render\\_mode=standalone](http://vivo-netsci.cns.iu.edu/vivo/visualization?uri=http%3A%2F%2Fvivo-trunk.indiana.edu%2Findividual%2FPerson74&vis=person_level&render_mode=standalone)

36 publication(s) from 2001 to 2010 [\(.CSV File\)](#)

80 co-author(s) from 2001 to 2010 [\(.CSV File\)](#)

Co-author network [\(GraphML File\)](#)

```

1 <?xml version="1.0" encoding="UTF-8"?>
2 <graphml xmlns="http://graphml.graphdrawing.org/xmlns"
3 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
4 xsi:schemaLocation="http://graphml.graphdrawing.org/xmlns
5 http://graphml.graphdrawing.org/xmlns/1.0/graphml.xsd">
6 <key id="label" for="node" attr.name="label" attr.type="string" />
7 <key id="number_of_authored_works" for="node" attr.name="number_of_authored_works" attr.type="int" />
8 <key id="num_unknown_publication" for="node" attr.name="num_unknown_publication" attr.type="int" />
9 <key id="num_latest_publication" for="node" attr.name="num_latest_publication" attr.type="int" />
10 <key id="latest_publication" for="node" attr.name="latest_publication" attr.type="int" />
11 <key id="profile_url" for="node" attr.name="profile_url" attr.type="string" />

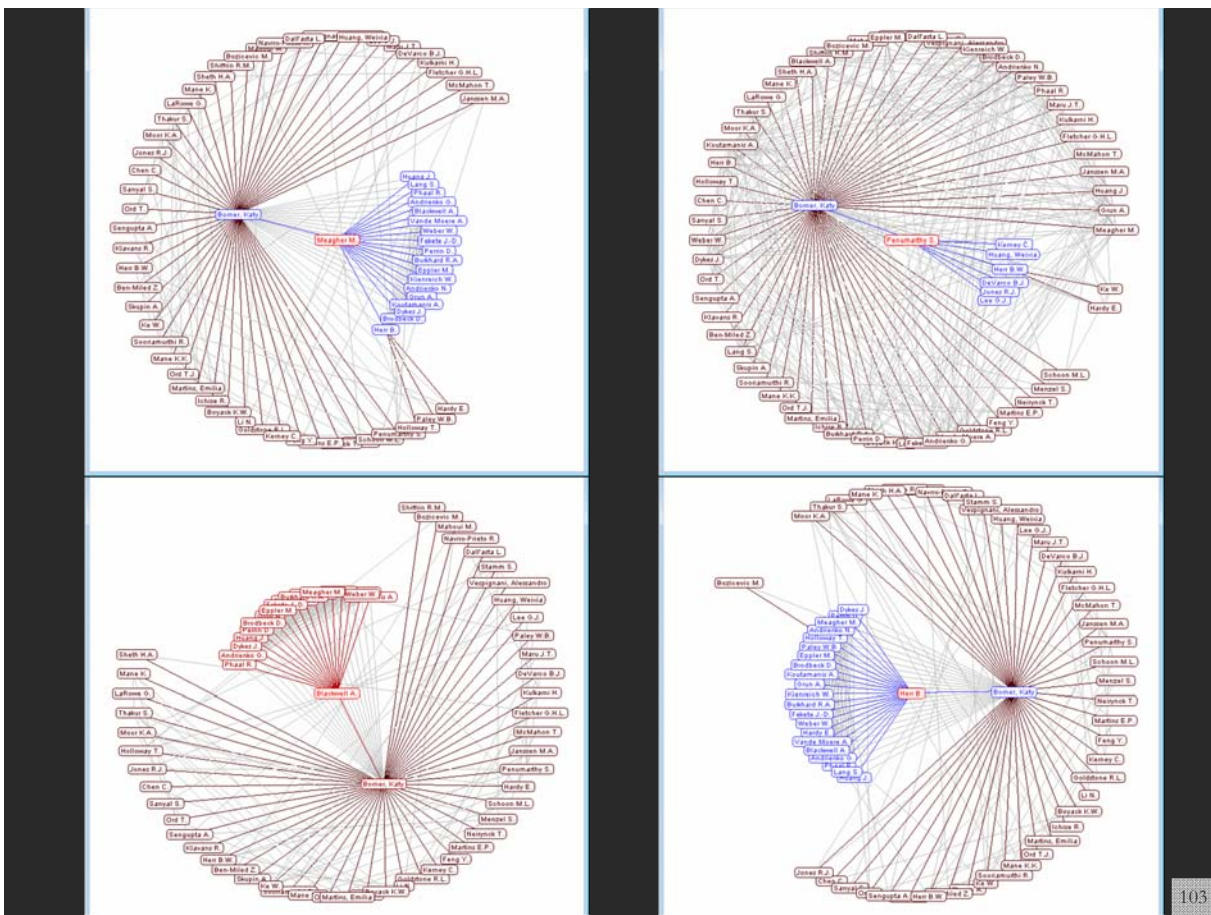
```

Save as Image (.PNG file)

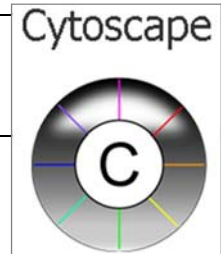
Publications per year [\(.CSV File\)](#), see top file.

Co-authors [\(.CSV File\)](#)

Co-Author	Count
Andrienko G.	1
Andrienko N.	1
Ben-Miled Z.	1
Blackwell A.	1
Boyack K.W.	4
Bozicevic M.	1
Brodbeck D.	1
Burkhard R.A.	1
Chen C.	5



CIShell – Add new Plugins, e.g., Cytoscape

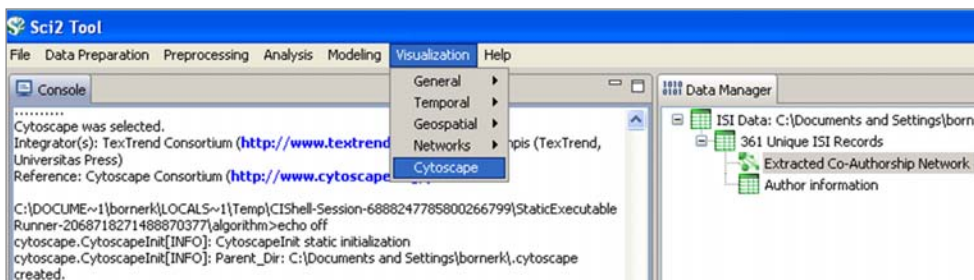


Adding more alyout algorithms and network visualization interactivity via Cytoscape <http://www.cytoscape.org>.

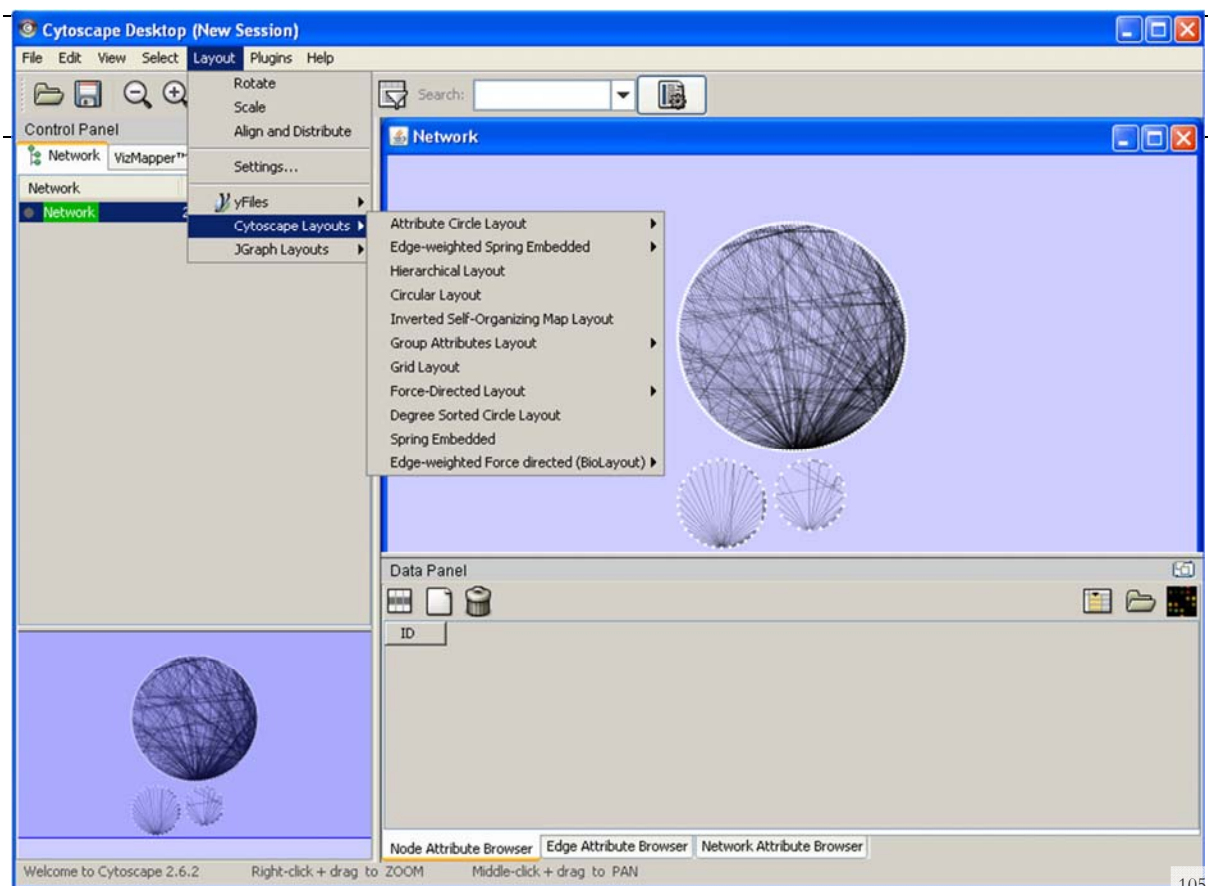
Simply add *org.textrend.visualization.cytoscape\_0.0.3.jar* into your /plugin directory.

Restart Sci2 Tool.

Cytoscape now shows in the Visualization Menu.



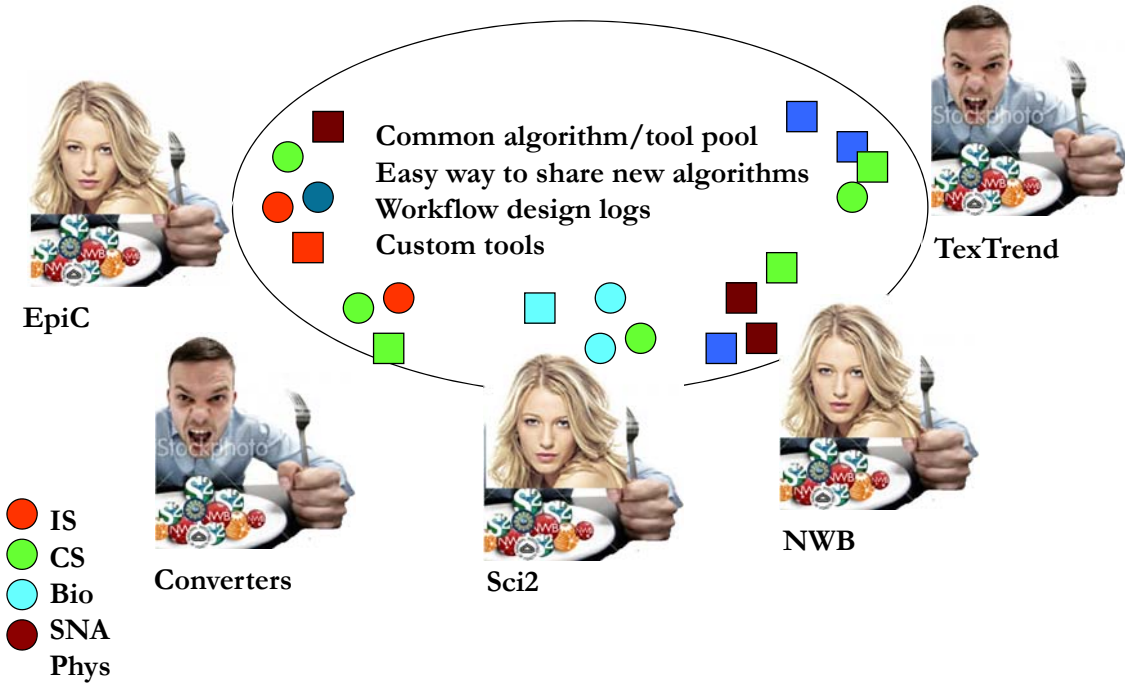
Select a network in Data Manager, run Cytoscape and the tool will start with this network loaded.



## 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 corpuses 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.



107



## Network Workbench Tool

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

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 65,000 times since December 2006.

**Network Workbench**  
 A Workbench for Network Scientists

Home People Research Publications

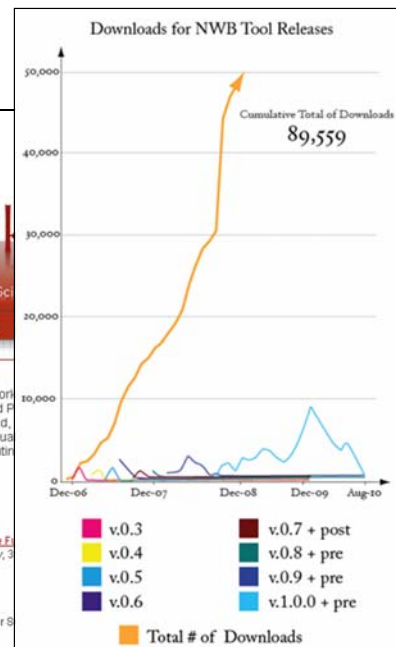
**Summary**  
 Network Workbench: A Large-Scale Network Toolkit for Biomedical, Social Science and Physics. Evaluate, and operate a unique distributed, scale network analysis, modeling, and visualization (NWB). The envisioned data-code-computer more  
[How to cite this project](#)

**News & Updates**

- 5.1.09 Kaelble, Steve. 2009. [Mapping the Future of Knowledge, Research & Creative Activity, 3 \(website\)](#) accessed 5/1/09
- 3.23.09 [1.0.0 beta 5 Released](#)
- 1.23.09 Ann Mcranie's [tutorial abstract](#) for S 2009
- 11.4.08 Two NWB PIs featured in "Connected—The Power of Six Degrees." 2008. Anna Maria Talas, Director. Australian Broadcasting Corporation, Ltd. [\[YouTube\]](#) [\[Full Video\]](#) (300MB)

[Getting Started](#)  
 See more [documentation](#)

**Get Involved**

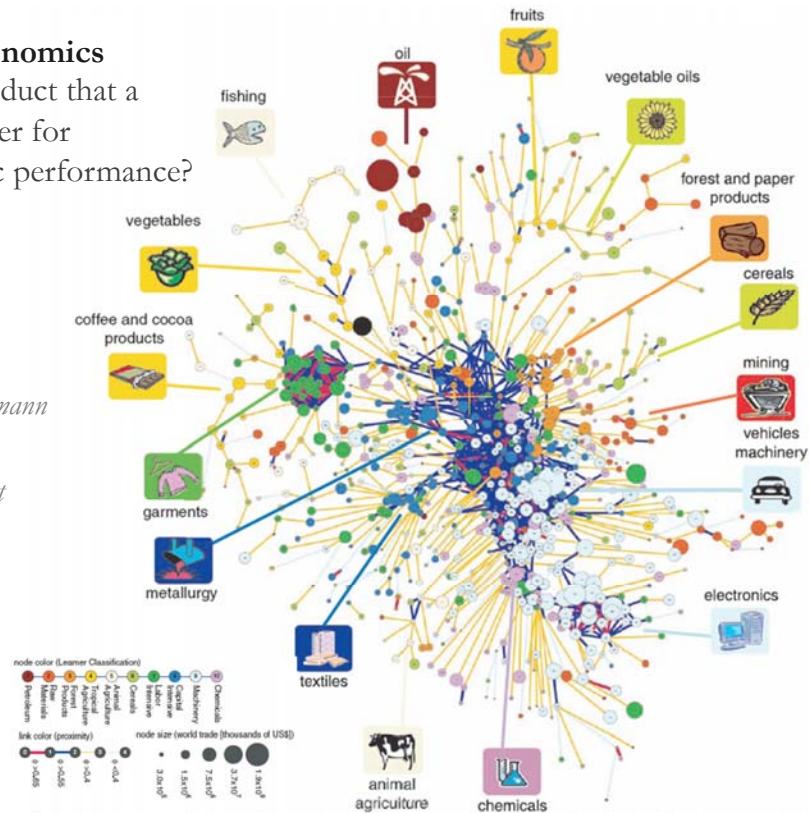


Herr II, Bruce W., Huang, Weixia (Bonnie), Penumarthy, Shashikant & Börner, Katy. (2007). Designing Highly Flexible and Usable Cyberinfrastructures for Convergence. In Bainbridge, William S. & Roco, Mibail C. (Eds.), *Progress in Convergence - Technologies for Human Wellbeing* (Vol. 1093, pp. 161-179),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

## Computational Economics

Does the type of product that a country exports matter for subsequent economic performance?

*C. A. Hidalgo, B. Klinger, A.-L. Barabási, R. Hausmann (2007) The Product Space Conditions the Development of Nations. Science 317, 482 (2007).*



**Fig. 1.** The product space. (A) Hierarchically clustered proximity matrix representing the 775 SITC-4 product classes exported in the 1998–2000 period. (B) Network representation of the product space. Links are color coded with their proximity value. The sizes of the nodes are proportional to world trade, and their colors are chosen according to the classification introduced by Leamer.

## Computational Social Science

Studying large scale social networks such as Wikipedia

*Second Sight: An Emergent Mosaic of Wikipedian Activity, The NewScientist, May 19, 2007*

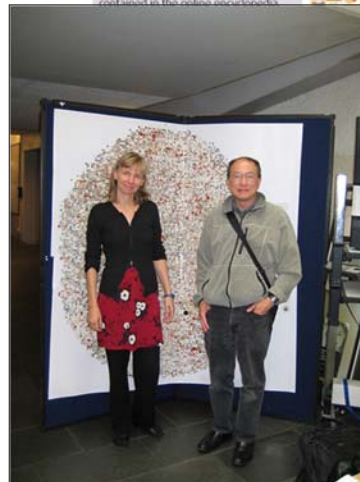


### Second sight

Image: Bruce W. Hest and Todd M. Holloway

#### Power struggle

How do you keep track of the bubbling mass of information that is Wikipedia? This chaotic-looking mosaic is one attempt to show which topics are contained in the online encyclopedia.



...pages at the time of writing include entries on Sheffield Wednesday football club, Mikhail Gorbachev and pigs). The mosaic has been commended in a competition for images that visualise network dynamics, coinciding with this week's International Workshop and Conference on Network Science in Bloomington.



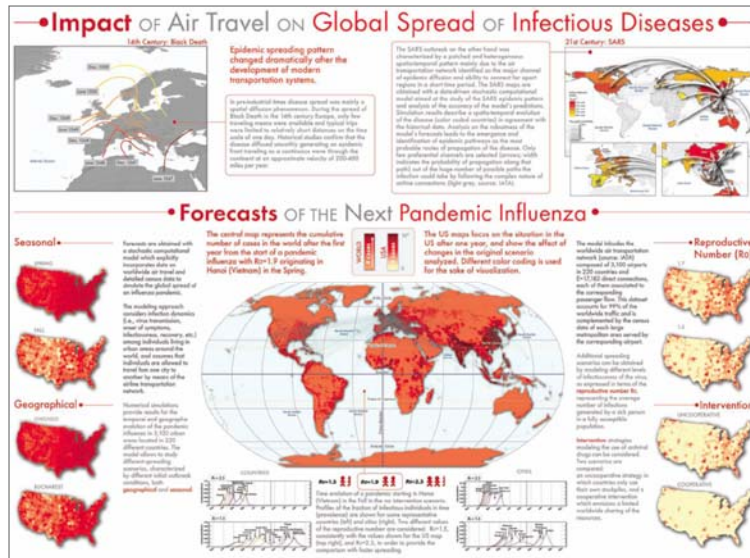
## Computational Epidemics

Forecasting (and preventing the effects of) the next pandemic.

*Epidemic Modeling in Complex realities*, V. Colizza, A. Barrat, M. Barthelemy, A. Vespignani, *Comptes Rendus Biologie*, 330, 364-374 (2007).

*Reaction-diffusion processes and metapopulation models in heterogeneous networks*, V. Colizza, R. Pastor-Satorras, A. Vespignani, *Nature Physics* 3, 276-282 (2007).

*Modeling the Worldwide Spread of Pandemic Influenza: Baseline Case and Containment Interventions*, V. Colizza, A. Barrat, M. Barthelemy, A.-J. Valleron, A. Vespignani, *PLoS-Medicine* 4, e13, 95-110 (2007).



## Computational Proteomics

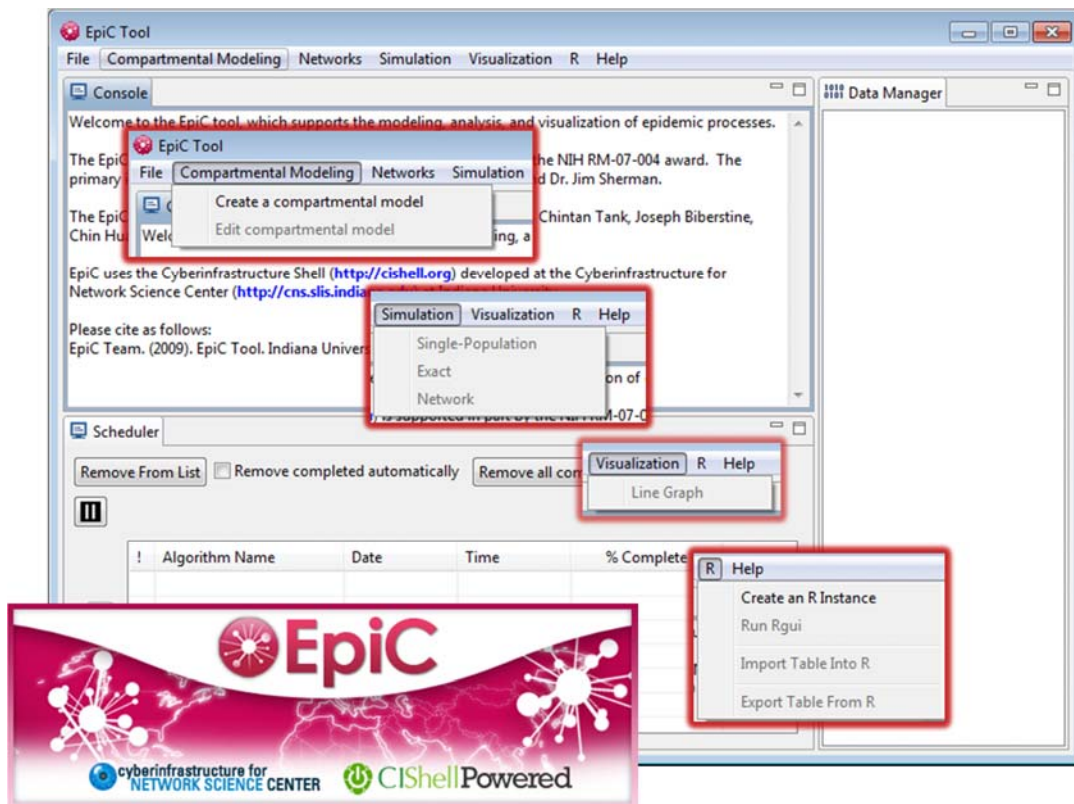
What relationships exist between protein targets of all drugs and all disease-gene products in the human protein-protein interaction network?

*Yildirim, Muhammed A., Kwan-II Goh, Michael E. Cusick, Albert-László Barabási, and Marc Vidal. (2007). Drug-target Network. Nature Biotechnology 25 no. 10: 1119-1126.*



**Figure 2** Drug-target network (DT network). The DT network is generated by using the known associations between FDA-approved drugs and their target proteins. Circles and rectangles correspond to drugs and target proteins, respectively. A link is placed between a drug node and a target node if the protein is a known target of that drug. The area of the drug (protein) node is proportional to the number of targets that the drug has (the number of drugs targeting the protein). Color codes are given in the legend. Drug nodes (circles) are colored according to their Anatomical Therapeutic Chemical Classification, and the target proteins (rectangular boxes) are colored according to their cellular component obtained from the Gene Ontology database.





113

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

114



## 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>

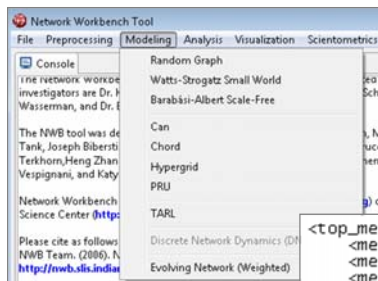
Additional Sci2 Plugins are at <http://sci2.wiki.cns.iu.edu/3.2+Additional+Plugins>

115



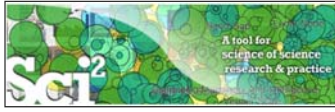
## CIShell – Customize Menu

- 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" standalone="no" ?>
<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>
```

116



Need Help? Ask an Expert!

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

117

If you enjoyed the tutorial, please thank Andrea Scharnhorst, Albert Meroño Peñuela, Jeannette Haagsma, Anja de Haas, Nick Jankowski, Sally Wyatt; US side: Chin Hua Kong, Joseph R. Biberstine, Samantha Hale

There will be a Sci2 Tutorial for *Science Policy Makers* at OECD, Paris on April 13, 2012.

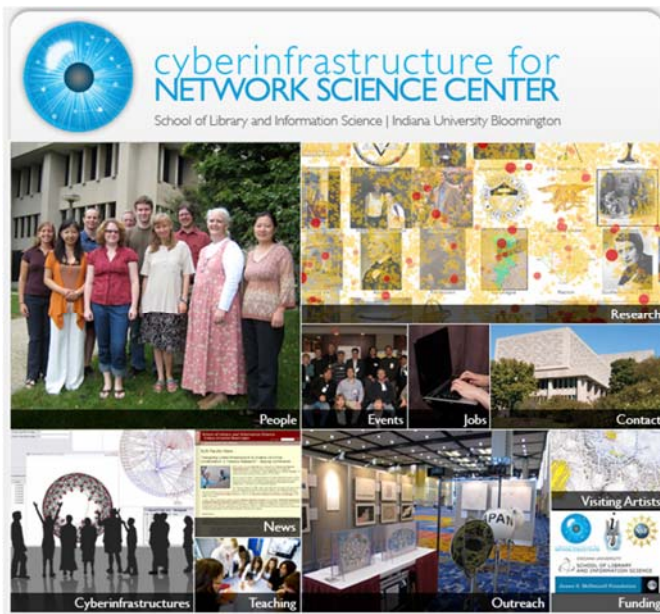
## Q & A

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

\*\*\*

Bug reports and all comments are welcome.

118



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>