"Sci2 Tool: A Tool for Science of Science Research and Practice" Tutorial

Dr. Katy Börner
CNS & IVL, SLIS, Indiana University, Bloomington, Indiana, USA
Royal Netherlands Academy of Arts and Sciences (KNAW), Amsterdam, The Netherlands
http://cns.iu.edu

Albert Meroño Peñuela (KNAW) & 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

OECD, Paris, France
Friday April 13, 2012 • 10:30-15:30

Software, Datasets, Plugins, and Documentation

➢ These slides
http://ivl.slis.indiana.edu/km/pres/2012-borner-sci2tutorial-oecd.pdf
➢ Sci2 Tool Manual v0.5.1 Alpha
http://sci2.wiki.cns.iu.edu
➢ Sci2 Tool v0.5.2 Alpha (April 9, 2012)
http://sci2.cns.iu.edu

➢ Additional Datasets
http://sci2.wiki.cns.iu.edu/2.5+Sample+Datasets
➢ Additional Plugins
http://sci2.wiki.cns.iu.edu/3.2+Additional+Plugins

Or copy them from the DVD or memory stick.

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

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

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

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

Tutorial Overview

10:30a Welcome and Overview of Tutorial and Attendees
10:45a Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools
11:00a Sci2 Tool Basics
➢ **Download and run the Sci2 Tool**
➢ **Load, analyze, and visualize family and business networks**
➢ **Horizontal line graph of NSF projects**
➢ **Studying four major network science researchers**
    - Load and clean a dataset; process raw data into networks
    - Find basic statistics and run various algorithms over the network
    - Visualize as either a circular hierarchy or network
12:30 Lunch Break
1:30 Sci2 Tool Novel Functionality
➢ **Yahoo! Geocoder**
➢ **Evolving collaboration networks**
➢ **R-Bridge**
3:00 Outlook and Q&A
3:30 Adjourn

Tutorial Overview

10:30a Welcome and Overview of Tutorial and Attendees
10:45a Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools
11:00a Sci2 Tool Basics
➤ Download and run the Sci2 Tool
➤ Load, analyze, and visualize family and business networks
➤ Horizontal line graph of NSF projects
➤ Studying four major network science researchers
  - Load and clean a dataset; process raw data into networks
  - Find basic statistics and run various algorithms over the network
  - Visualize as either a circular hierarchy or network
12:30 Lunch Break
1:30 Sci2 Tool Novel Functionality
➤ Yahoo! Geocoder
➤ Evolving collaboration networks
➤ R-Bridge
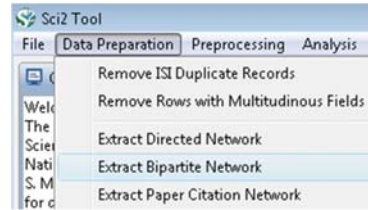3:00 Outlook and Q&A
3:30 Adjourn

5

Using the Sci2 Tool to Visualize Tutorial Registrants

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

Affiliation	Background
OECD	Banking;Finance;Actuarial Science
OECD	Science
OECD	Economics
OECD	Economics
	Biotechnology;International Political
OECD	Economy;Science Policy Studies
OECD	Economics

Total of 49 registrants on April 10, 2012.
Main affiliations are shown to the right,
some unique backgrounds are below.

International
International And Development Economics
International Economics
International Political Economy
Library And Information Science
Library Science

Affiliation	#
BBSRC	4
Belgian Science Policy Office	1
CNRS	2
Cranfield School of Management	1
EDU, PAI	1
Elsevier	1
Eversole Associates	1
German Institute for International Educational Research	1
Karolinska Institutet	2
Karolinska Institutet, University Library	1
MindTracker	1
Nobel Museum	1
OECD	27
Sciences Po Paris	3
Strafitech, ESC Rennes School of Business	1
VU University Amsterdam	1
	49

6

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

Affiliation	Background
OECD	Banking;Finance;Actuarial Science
OECD	Science
OECD	Economics
OECD	Economics
	Biotechnology;International Political Economy;Science Policy Studies
OECD	
OECD	Economics

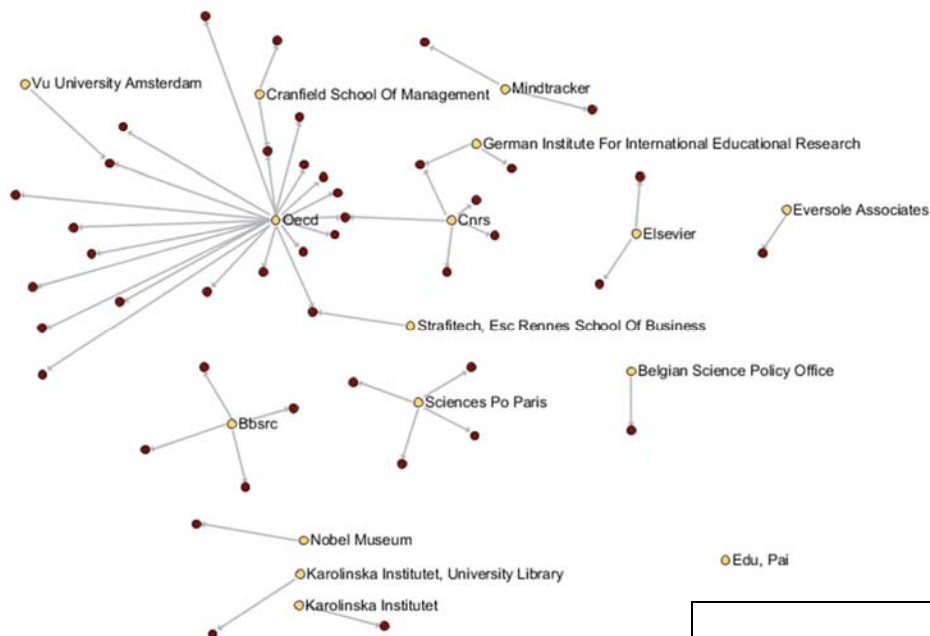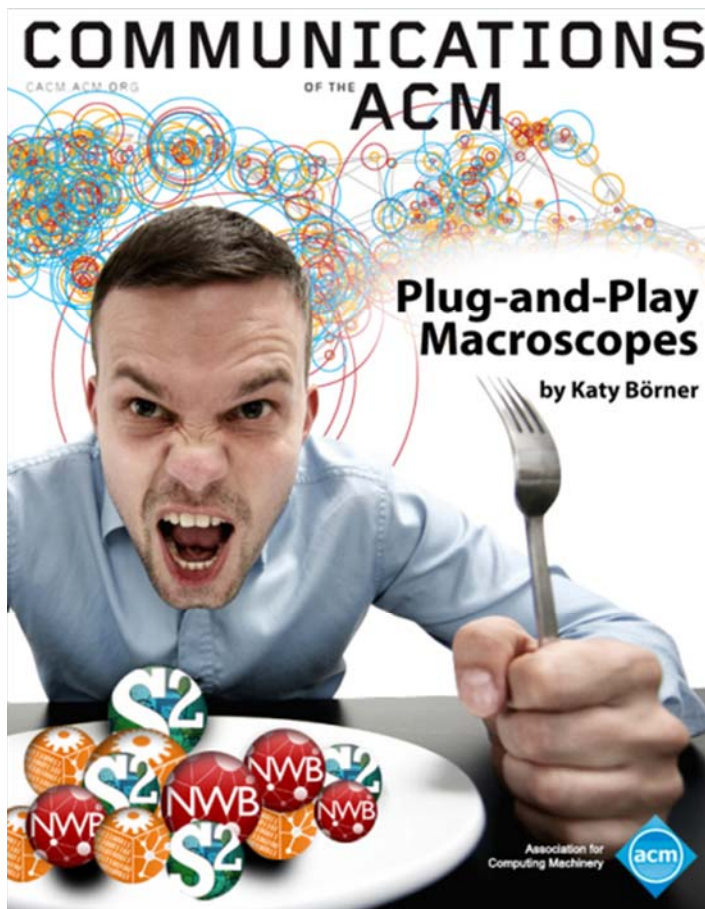Run *'Data Preparation > Extract Bipartite Network'*

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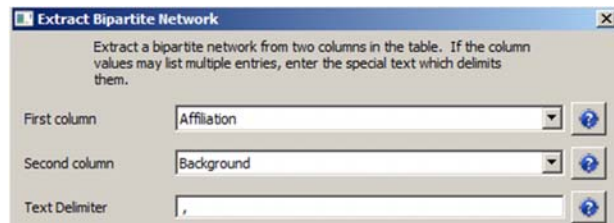
Visualize resulting *'Bipartite network from Affiliation to Background'* using *'Visualization > Network > GUESS'* and *'Layout > GEM'*, *'Layout > Bin Pack'*

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

- 🟡 Affiliation
- 🔴 Background

Object: nodes based on -> | Property: bipartitetype | Operator: != | Value: Background

🟡 Affiliation
🔴 Background

9

Learn

Interest in Graph Theory, Network Theory and in A.-L. Barabási's works.

Its use in analyzing and modeling large-scale longitudinal data.

Applications related to international trade (Hidalgo's Product Space)

What possible to do with it and first practical hands on experience

Overview of possibilities to use tool for research and ways of usage.

10

Tutorial Overview

10:30a Welcome and Overview of Tutorial and Attendees
10:45a Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools
11:00a Sci2 Tool Basics
➢ Download and run the Sci2 Tool
➢ Load, analyze, and visualize family and business networks
➢ Horizontal line graph of NSF projects
➢ Studying four major network science researchers
　　- Load and clean a dataset; process raw data into networks
　　- Find basic statistics and run various algorithms over the network
　　- Visualize as either a circular hierarchy or network
12:30 Lunch Break
1:30 Sci2 Tool Novel Functionality
➢ Yahoo! Geocoder
➢ Evolving collaboration networks
➢ R-Bridge
3:00 Outlook and Q&A
3:30 Adjourn

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

Take terra bytes of data

Plug-and-Play Macroscopes

Find your way

Find collaborators, friends

Identify trends

13
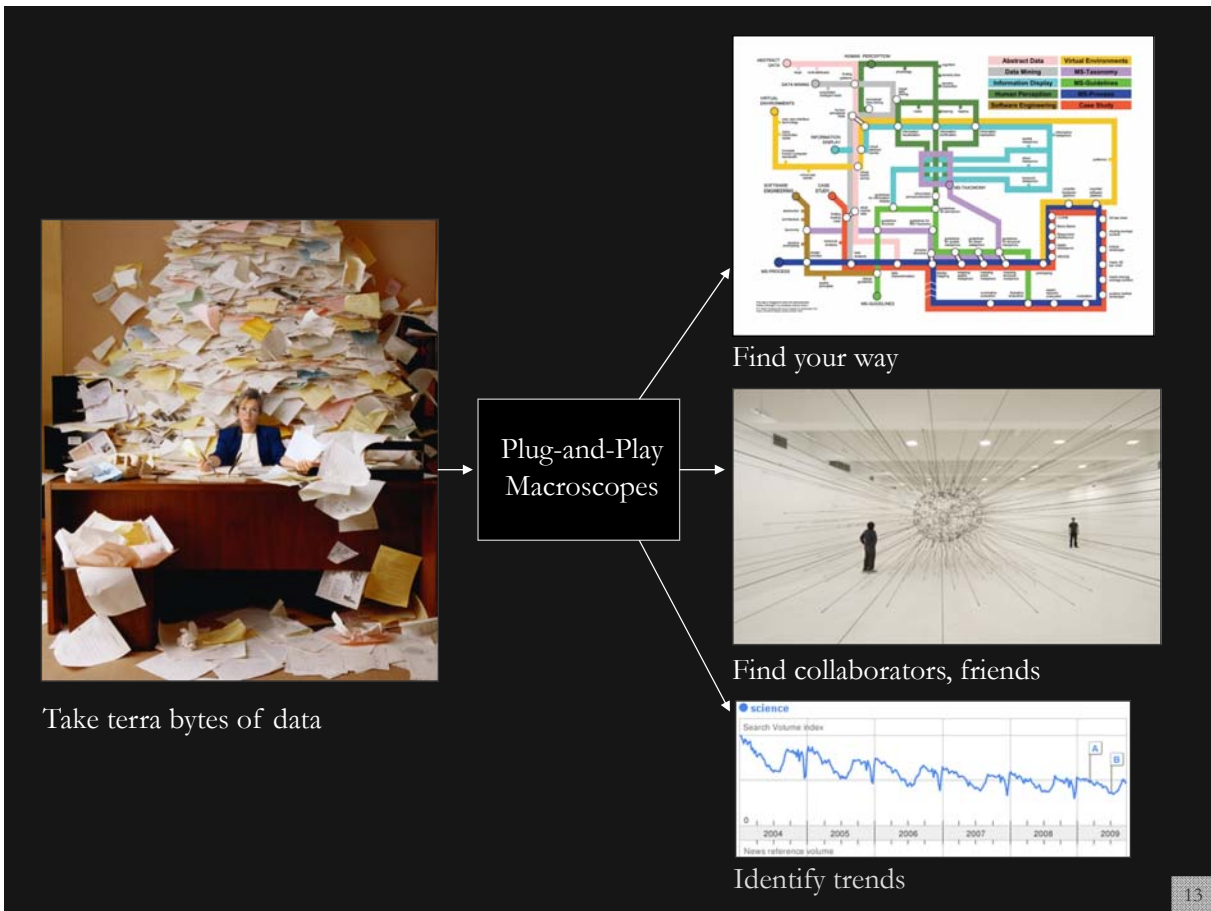
Macroscopes

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**

14

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.**

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.

Different datasets/formats. Diverse algorithms/tools written in many programming languages.

IS

Physics

CS

Bio

SNA
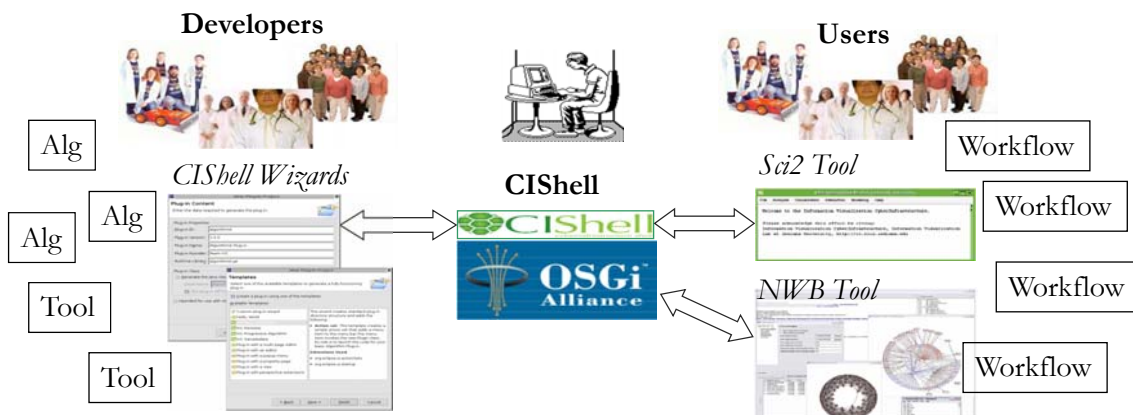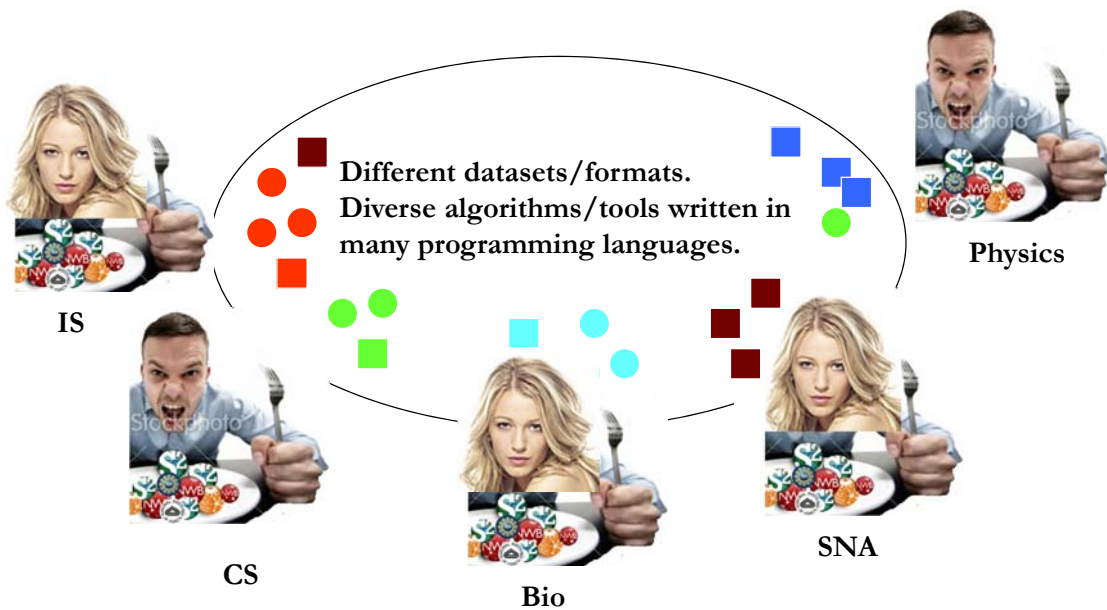
17

Tutorial Overview

10:30a Welcome and Overview of Tutorial and Attendees
10:45a Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools
11:00a Sci2 Tool Basics
➢ **Download and run the Sci2 Tool**
➢ Load, analyze, and visualize family and business networks
➢ Horizontal line graph of NSF projects
➢ Studying four major network science researchers
    - Load and clean a dataset; process raw data into networks
    - Find basic statistics and run various algorithms over the network
    - Visualize as either a circular hierarchy or network
12:30p Lunch Break
1:30p Sci2 Tool Novel Functionality
➢ Yahoo! Geocoder
➢ Evolving collaboration networks
➢ R-Bridge
3:00p Outlook and Q&A
3:30 Adjourn

18

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

Vol 464|25 March 2010

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

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**.

Sci² Tool – "Open Code for S&T Assessment"

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

Joint Co-Authorship Network

Sci Maps

- 9 Biology
- 9 Biotechnology
- 13 Brain Research
- 19 Chemical, Mechanical, & Civil Engineering
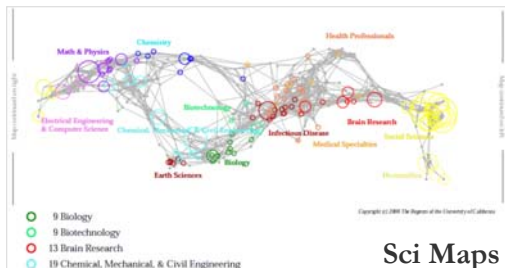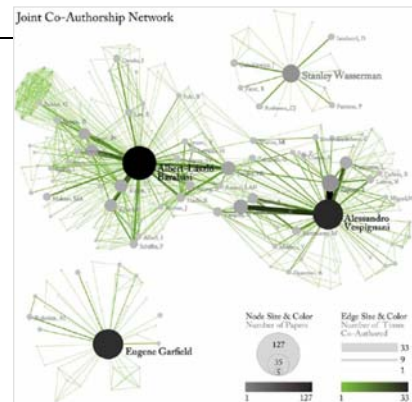
GUESS Network Vis

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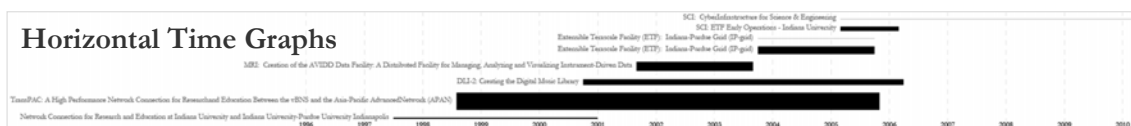
Börner, Katy, Huang, Weixia (Bonnie), Linnemeier, Micah, Duhon, Russell Jackson, Phillips, Patrick, Ma, Nianli, Zoss, Angela, Guo, Hanning & Price, Mark. (2009). Rete-Netzwerk-Red: Analyzing and Visualizing Scholarly Networks Using the Scholarly Database and the Network Workbench Tool. Proceedings of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

Sci² Tool Visualizations
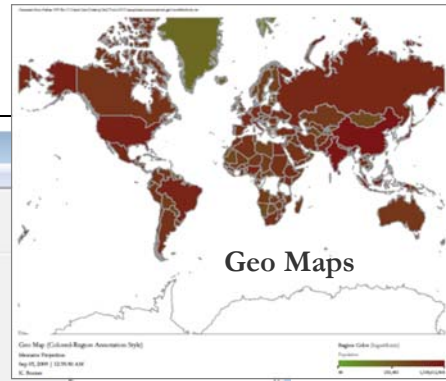
Geo Maps
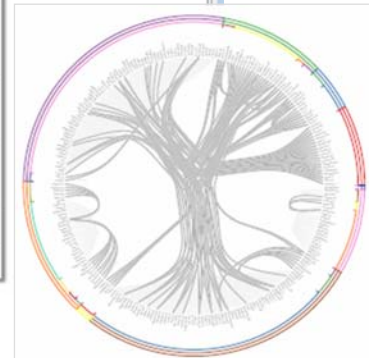
Circular Hierarchy

Sci² Tool Usage at National Science Foundation

Topic co-occurrence network of the 2885 cognitive and neuroscience NSF projects funded between 2007 and 2011. Statistical text mining (Topic model) was used to identify topics from NSF awards and proposals from 2000-2011.
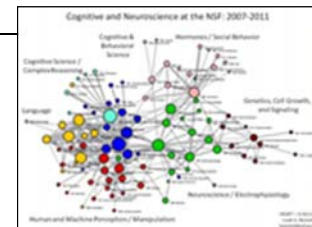
Each award is tagged with up to 4 topics. Lines represent the co-occurrence of the connected topics within an award(s). The nodes are scaled by number of awards (max = 355) and the lines are scaled on number of co-occurrences (max =91). The node colors differentiate the nodes via the level-0 Blondel communities.

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

Cognitive and Neuroscience at the NSF: 2007-2011

Cognitive Science / Complex Reasoning

Cognitive & Behavioral Science

Hormones / Social Behavior

Genetics, Cell Growth, and Signaling

Language

Neuroscience / Electrophysiology

Human and Machine Perception / Manipulation
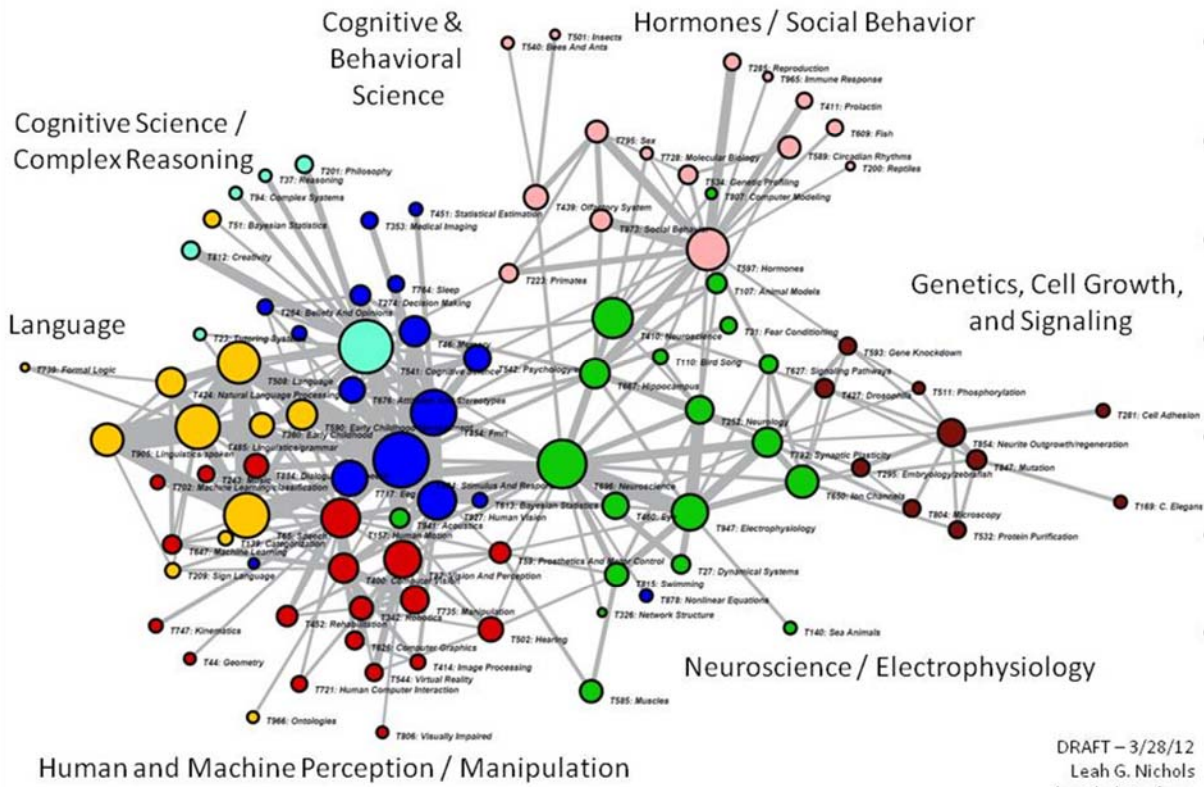
DRAFT – 3/28/12
Leah G. Nichols
lenichol@nsf.gov

Sci² Tool Usage at National Institutes of Health

Mapping Transdisciplinary Tobacco Use Research Centers Publications: Compare R01 investigator based funding with TTURC Center awards in terms of number of publications and evolving co-author networks. *Zoss & Börner, forthcoming.*
Supported by NIH/NCI Contract HHSN261200800812
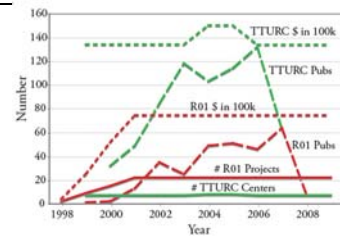
Longitudinal R01 Co-Authorship Network
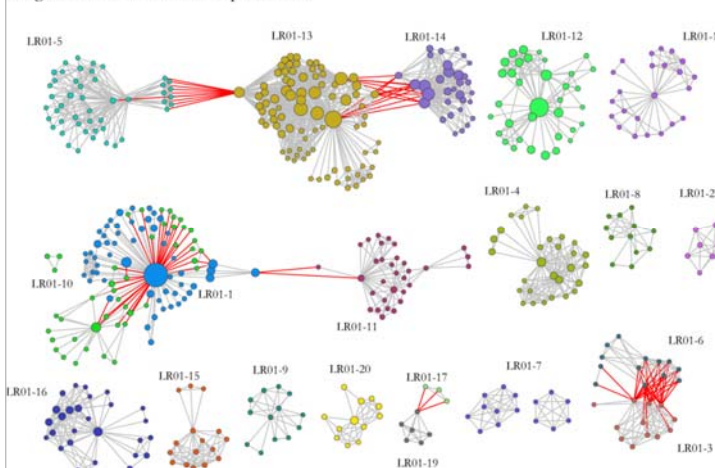
TTURC Co-Authorship Network

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

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

NIFA Animal Genomics Program Co-author Network
•394 papers, 2008-2010
•29 NIFA-funded project directors
•2071 co-authors
•Average award size: ~$1 million

NIFA Animal Growth and Nutrient Utilization Program Co-author Network
•319 papers, 2008-2010
•27 NIFA-funded project directors
•994 co-authors
•Average award size: ~$350,000

NIFA Animal Health and Protection Program Co-author Network
•534 papers, 2008-2010
•49 NIFA-funded project directors
•1954 co-authors
•Average award size: ~$350,000

NIFA Animal Reproduction Program Co-author Network
•398 papers, 2008-2010
•33 NIFA-funded project directors
•997 co-authors
•Average award size: ~$350,000

RESULTS

Geographic Distribution of NIFA-Funded Projects
•For each program, projects appear to be fairly evenly dispersed across geographic regions, with the exception of genomics projects which are concentrated in the Northern Plains and Mid Atlantic.

•Projects active in 2008-2010 are concentrated in the eastern half of the United States.

CONCLUSIONS

New approaches and tools are being employed by the federal government to support strategic planning, management, and evaluation of research programs, especially those that are mission driven. Network visualization and analysis allow better understanding of program areas and enable decision makers to identify institutions and authors playing major roles as central hubs, locate discipline-based network cut-points, and readily detect players in international scientific collaborations critical to problem-solving.

How did cognitive neuroscience of attention emerge from neurobiology and psychology, 1980–2005? Author co-citation analysis and Pfnet is used to **trace prospectively the development of the field from its precursor disciplines**: cognitive psychology, single cell neurophysiology, neuropsychology, and evoked potential research.
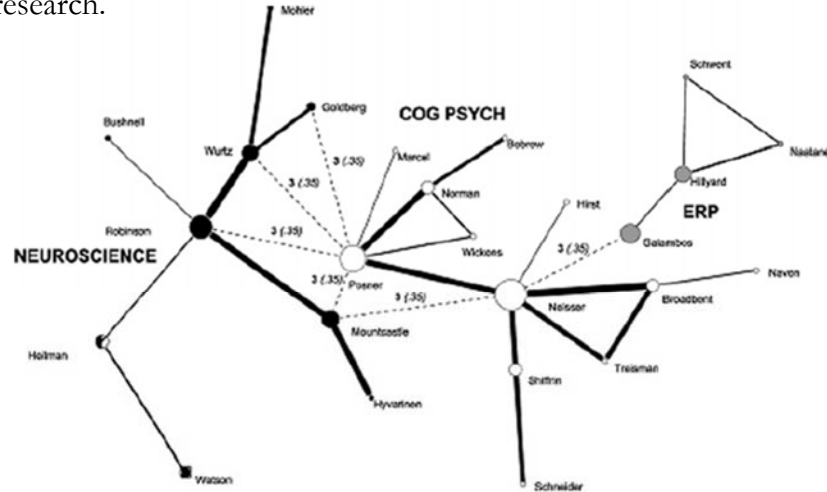
Fig. 1 In the 1980 net, neuroscience (*black* nodes and *black–white* nodes) and cognitive psychology (*white* nodes) develop as clusters with high internal co-citation rates. ERP (*grey* nodes) develops later in net construction. These clusters are connected by secondary edges at very low levels of co-citation

By 1990 a distinct cognitive neuroscience specialty cluster emerges, dominated by authors engaged in brain imaging research.
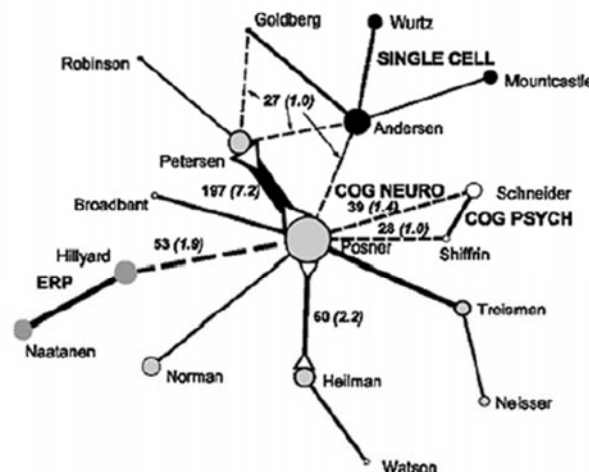
Fig. 5 The strongest link in the 1995 net is a primary edge linking Posner and Petersen. ERP and single cell neurophysiology are linked to cognitive neuroscience cluster by secondary edges

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

---
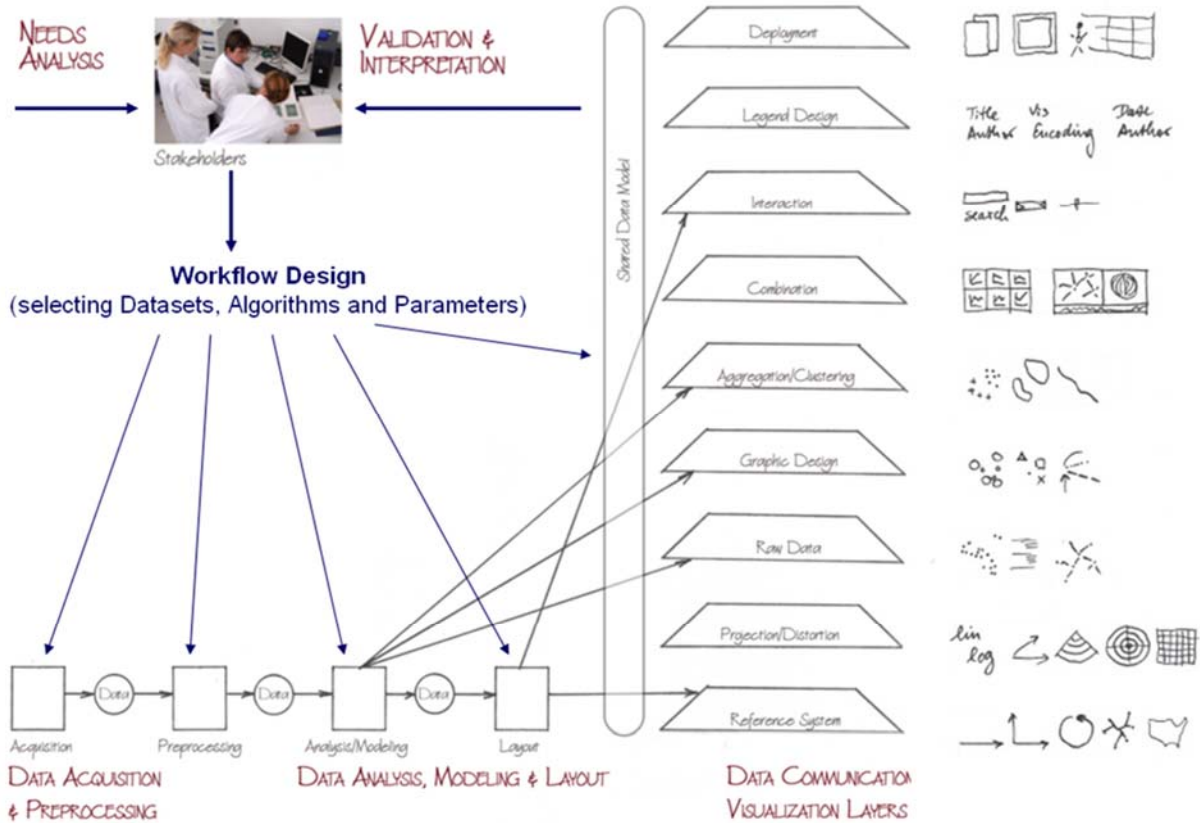
Type of Analysis vs. Level of Analysis

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

Type of Analysis vs. Level of Analysis
Covered Today:

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

33

Sci² Tool – Supported Data Formats

Input:

Network Formats
- GraphML (*.xml or *.graphml)
- XGMML (*.xml)
- Pajek .NET (*.net)
- NWB (*.nwb)

Scientometric Formats
- ISI (*.isi)
- Bibtex (*.bib)
- Endnote Export Format (*.enw)
- Scopus csv (*.scopus)
- NSF csv (*.nsf)

Other Formats
- Pajek Matrix (*.mat)
- TreeML (*.xml)
- Edgelist (*.edge)
- CSV (*.csv)

Output:

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

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

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

34
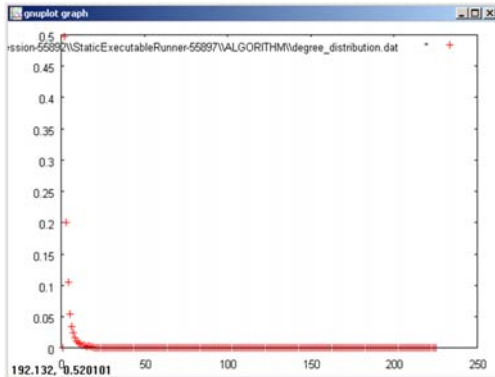
Gnuplot

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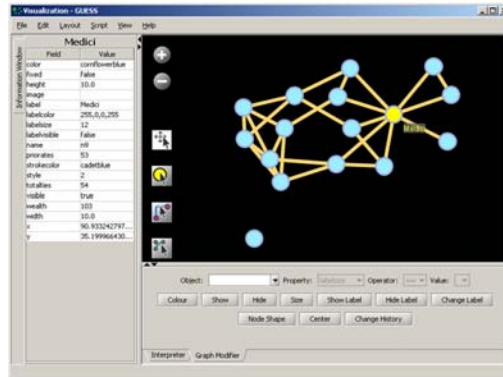
GUESS

exploratory data analysis and visualization tool for graphs and networks.

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

35

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

36
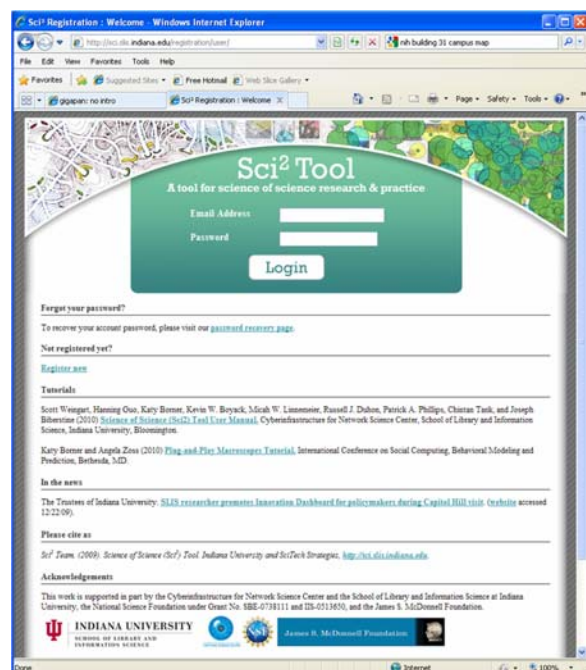
Sci2 Tool v0.5 Alpha (May 4, 2011)

➢ 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-N-1.0.0.201008130505NGT-macosx.carbon.ppc.zip
sci2-N-1.0.0.201008130505NGT-macosx.carbon.x86.zip
sci2-N-1.0.0.201008130505NGT-macosx.cocoa.x86_64.zip
sci2-N-1.0.0.201008130505NGT-win32.win32.x86.zip

Sci2 Tool runs on Windows, Mac, and Linux.

Unzip.

Run /sci2/sci2.exe

Sci2 Tool Interface Components

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

Use

➢ **Menu** to read data, run algorithms.

➢ **Console** to see work log, references to seminal works.

➢ **Data Manager** to select, view, save loaded, simulated, or derived datasets.

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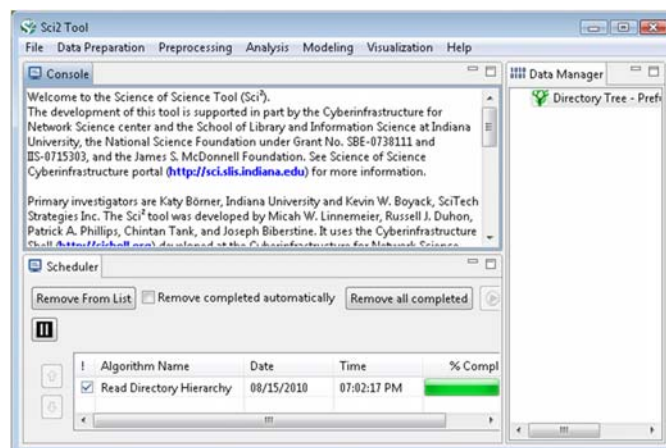
All workflows are recorded into a log file (see /sci2/logs/…), and soon can be re-run for easy replication. If errors occur, they are saved in a error log to ease bug reporting.

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

Tutorial Overview

10:30a Welcome and Overview of Tutorial and Attendees
10:45a Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools
11:00a Sci2 Tool Basics
➢ Download and run the Sci2 Tool
➢ **Load, analyze, and visualize family and business networks**
➢ Horizontal line graph of NSF projects
➢ Studying four major network science researchers
   - Load and clean a dataset; process raw data into networks
   - Find basic statistics and run various algorithms over the network
   - Visualize as either a circular hierarchy or network
12:30 Lunch Break
1:30 Sci2 Tool Novel Functionality
➢ Yahoo! Geocoder
➢ Evolving collaboration networks
➢ R-Bridge
3:00 Outlook and Q&A
3:30 Adjourn

Tutorial Overview

10:30a Welcome and Overview of Tutorial and Attendees
10:45a Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools
11:00a 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
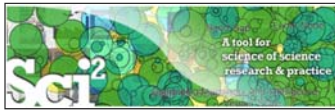   - Visualize as either a circular hierarchy or network
12:30 Lunch Break
1:30 Sci2 Tool Novel Functionality
➢ Yahoo! Geocoder
➢ Evolving collaboration networks
➢ R-Bridge
3:00 Outlook and Q&A
3:30 Adjourn

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 Strozzis."

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

41

```
*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"
```

42

➢ Load **yoursci2directory*/sampledata/socialscience/florentine.nwb*

➢ Run *'Analysis > Network Analysis Toolkit (NAT)'* to get basic properties.

This graph claims to be undirected.

Nodes: 16

Isolated nodes: 1

Node attributes present: label, wealth, totalities, priorates

Edges: 27

No self loops were discovered.

No parallel edges were discovered.

Edge attributes:

Nonnumeric attributes:

Example value

marriag...T

busines...F

Average degree: 3.375

There are 2 weakly connected components. (1 isolates)

The largest connected component consists of 15 nodes.

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

Density (disregarding weights): 0.225

➢ Select network and run *'Visualization > GUESS'* to open GUESS with file loaded.

➢ Apply 'Layout > GEM'.

43

44

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.

45
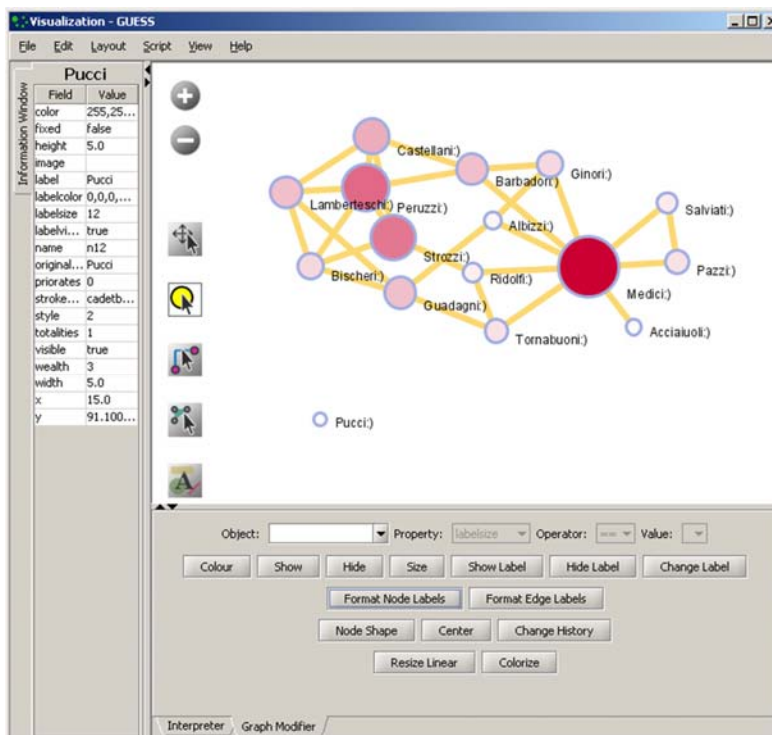
Graph Modifier:
Select "all nodes" in the Object drop-down menu and click 'Show Label' button.

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.'

46

Visualization - GUESS

File Edit Layout Script View Help

Information Window

Acciaiuoli:)

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

Castellani:) Barbadori:) Ginori:) Salviati:)
Lamberteschi:) Peruzzi:) Albizzi:) Pazzi:)
Strozzi:) Medici:)
Bischeri:) Ridolfi:)
Guadagni:) Acciaiuoli:)
Tornabuoni:)

Pucci:)

Interpreter:
Uses Jython a combination of Java and Python.

Try
colorize(wealth, white, red)

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

Interpreter / Graph Modifier

Tutorial Overview

10:30a Welcome and Overview of Tutorial and Attendees
10:45a Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools
11:00a Sci2 Tool Basics
➢ Download and run the Sci2 Tool
➢ Load, analyze, and visualize family and business networks
➢ **Horizontal line graph of NSF projects**
➢ Studying four major network science researchers
   - Load and clean a dataset; process raw data into networks
   - Find basic statistics and run various algorithms over the network
   - Visualize as either a circular hierarchy or network
12:30 Lunch Break
1:30 Sci2 Tool Novel Functionality
➢ Yahoo! Geocoder
➢ Evolving collaboration networks
➢ R-Bridge
3:00 Outlook and Q&A
3:30 Adjourn

Horizontal line graph of NSF projects
See *5.2.1 Funding Profiles of Three Universities (NSF Data)*

Download NSF data
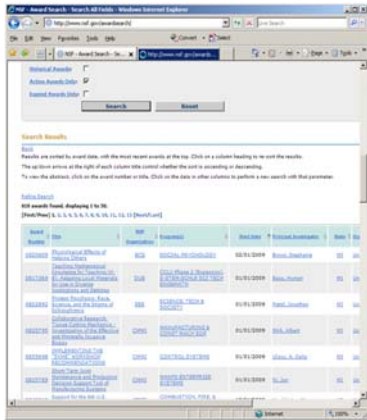
Visualize as Horizontal Line Graph

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

Text

Start date   End date
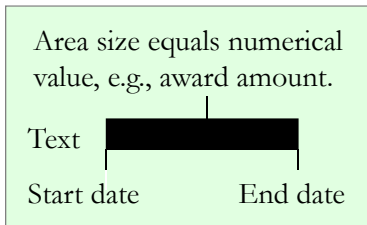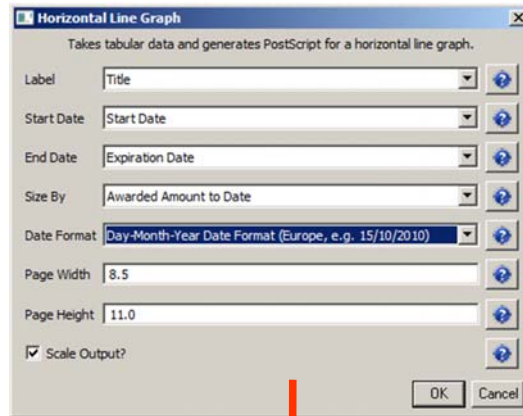
49
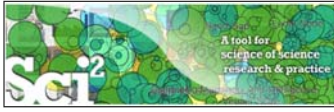
Horizontal line graph of NSF projects

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

Save in CSV format as *institution*.nsf
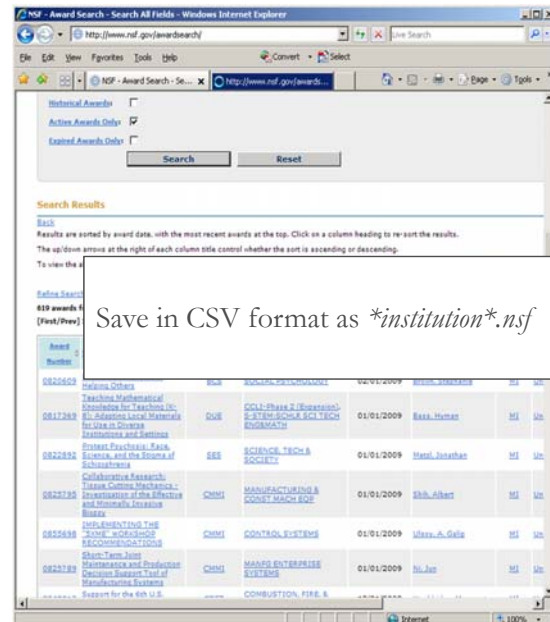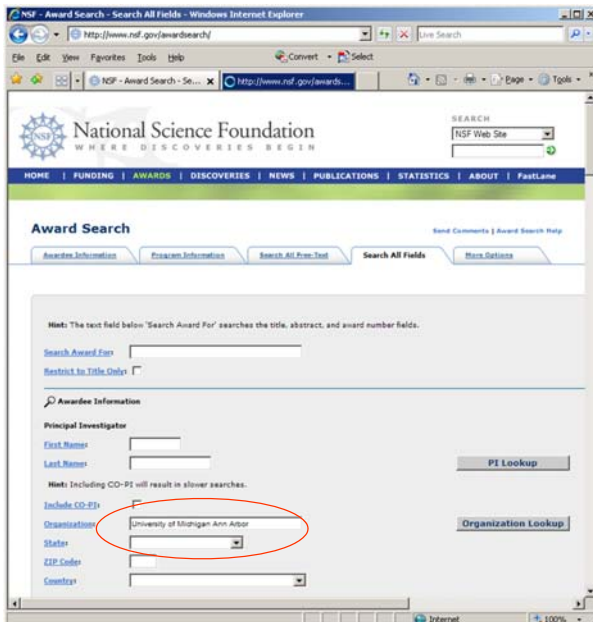
50

Horizontal line graph of NSF projects

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

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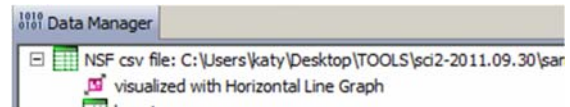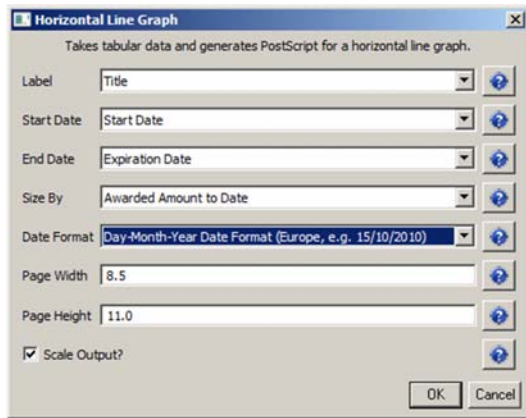
Save *'visualized with Hotizontal Line Graph'* as ps or eps file. Convert into pdf and view. Zoom to see details in visualizations of large datasets, e.g., all NSF awards ever made.

Horizontal line graph of NSF projects

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

Text, e.g., title

Start date   End date
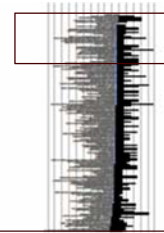
More NSF data workflows can be found in wiki tutorial:

5.1.3 Funding Profiles of Three Researchers at Indiana University (NSF Data)
5.2.1 Funding Profiles of Three Universities (NSF Data)
5.2.3 Biomedical Funding Profile of NSF (NSF Data)

Tutorial Overview

53

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

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

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

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

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

54

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
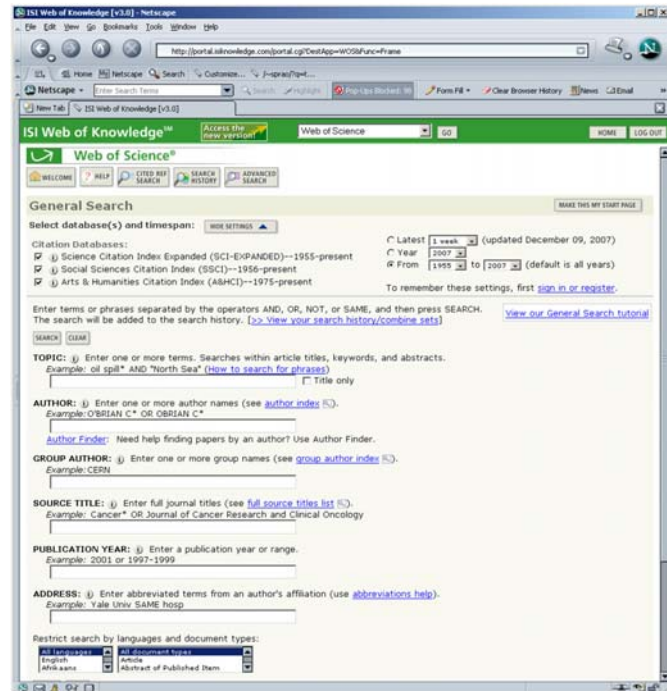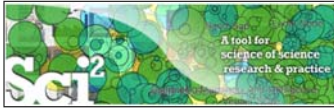
Comparison of Counts
No books and other non-WoS publications are covered.

	Age	Total # Cites	Total # Papers	H-Index	
Eugene Garfield	82	1,525	672	31	
Stanley Wasserman		122	35	17	
Alessandro Vespignani	42	451	101	33	
Albert-László Barabási	40	2,218	126	47	*(Dec 2007)*
	41	16,920	159	52	*(Dec 2008)*
	44	30,102	201	68	*(April 2011)*

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.
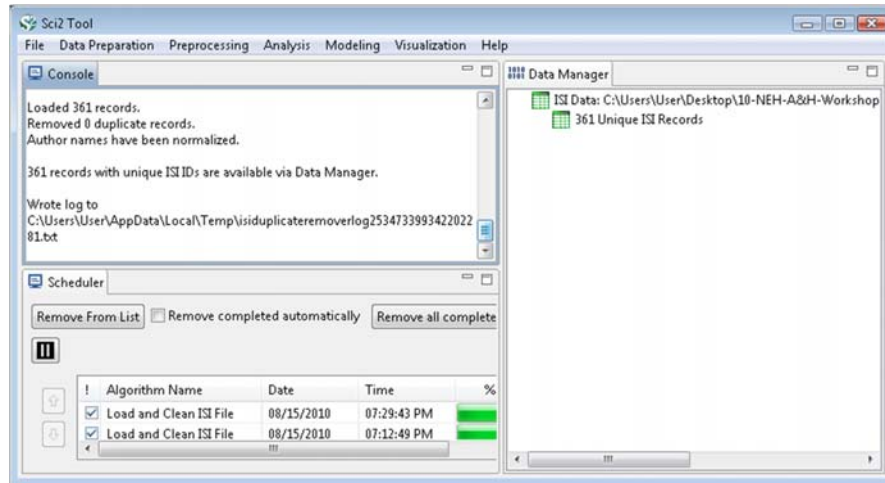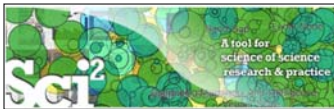
Extract Co-Author Network
(see section 5.1.4.2 on correcting duplicate/misspelled author names)

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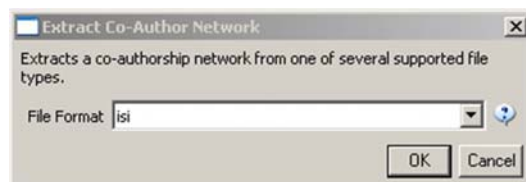
The result is an undirected but weighted network of co-authors in the Data Manager.

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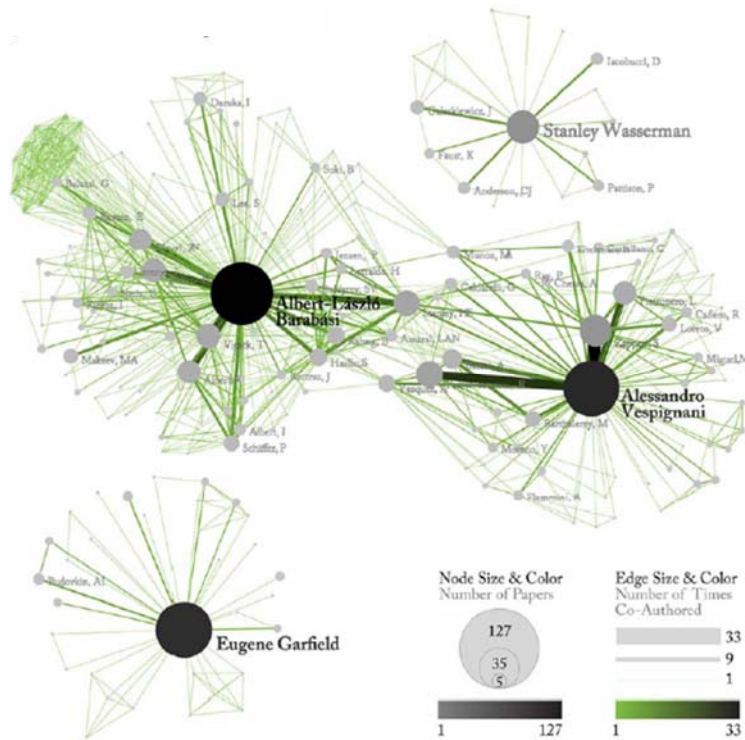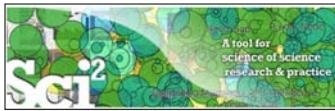
Use the GUESS Graph Modifier to change
color and size coding.

Calculate node degrees in Sci2 Tool.

Use a graphic program to add legend.
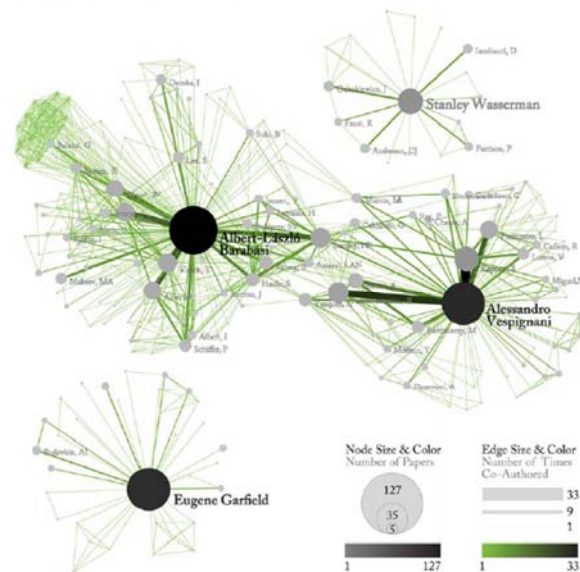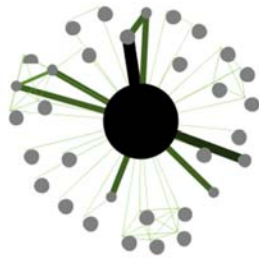
Eugene Garfield
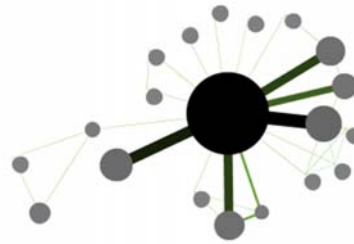
Stanley Wasserman
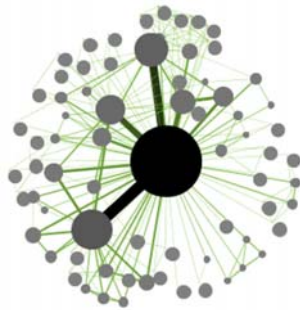
Alessandro Vespignani
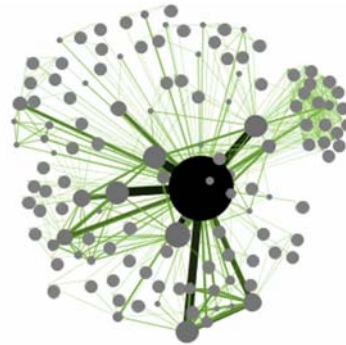
Albert-László Barabási

61

**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.
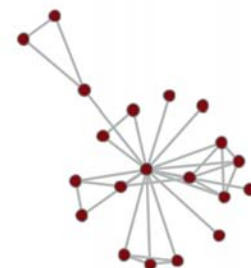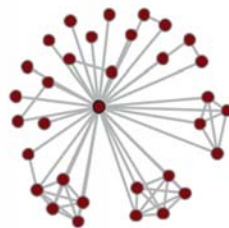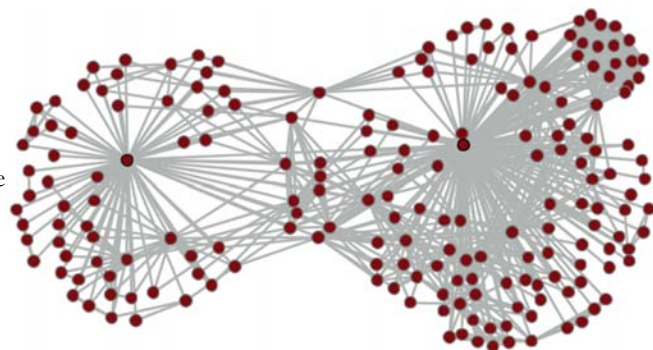
62
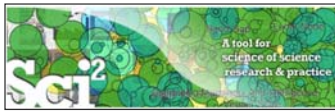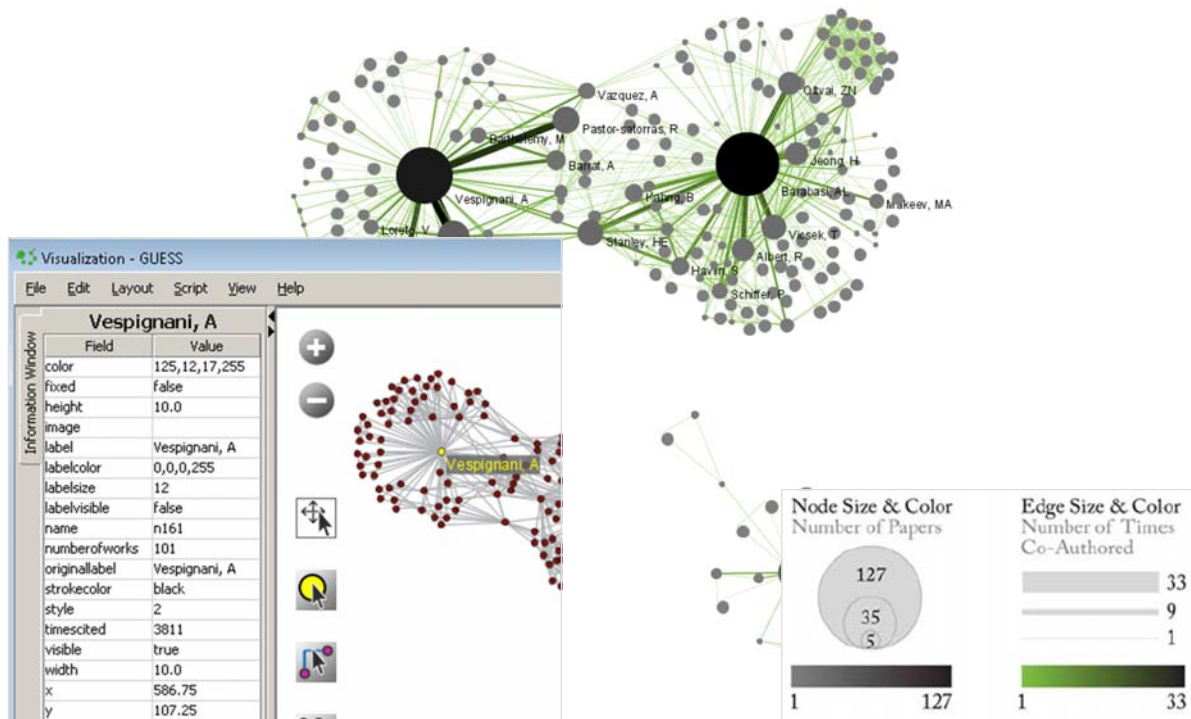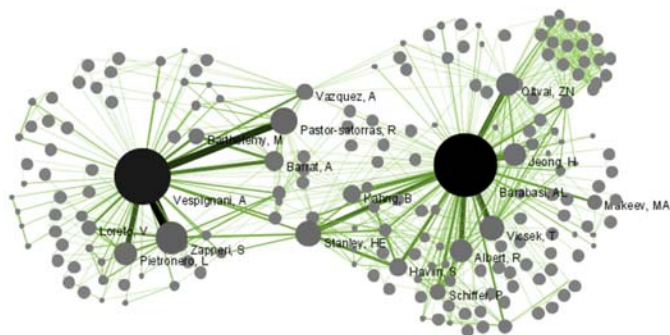
....
Weak Component Clustering was selected.
Implementer(s): Russell Duhon
Integrator(s): Russell Duhon

Input Parameters:
Number of top clusters: 10
3 clusters found, generating graphs for the top 3 clusters.
....
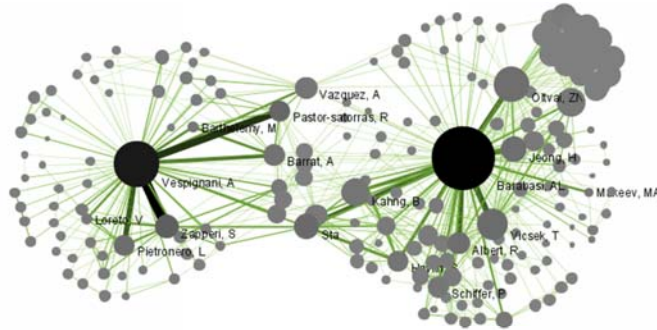
....
Node Degree was selected.
Documentation:
https://nwb.slis.indiana.edu/community/?n=AnalyzeData.No
deDegree
....

65
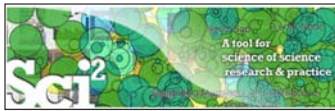
....
Node Betweenness Centrality was selected.
Author(s): L. C. Freeman
Implementer(s): Santo Fortunato
Integrator(s): Santo Fortunato, Weixia Huang
Reference: Freeman, L. C. (1977). A set of measuring centrality
based on betweenness. Sociometry. 40:35-41.

Input Parameters:
Number of bins: 10

umber of bins: 10
....

66
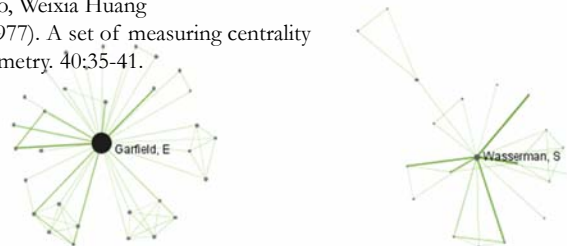
....
MST-Pathfinder Network Scaling was selected.
Input Parameters:
Weight Attribute measures: SIMILARITY
Edge Weight Attribute: weight
....

Network Visualization:
Circular Hierarchy Visualization

Select Co-Author Network and run Blondel Community detection:

With parameter values

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

Circular Hierarchy	
Provides Circular Hierarchy Visualization on the network.	
Degree of Edge Bundling	0.75
Node Strength Column	timescited
Level 0	blondel_community_level_0
Level 1	blondel_community_level_1
Level 2	blondel_community_level_2
Level 3	No Level
Edge Weight Column	numberofcoauthoredworks
Node Color Column	numberofworks
Node Color Range	Green to red

OK   Cancel

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

Node labels, e.g.,
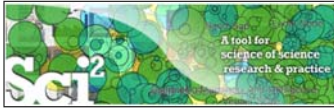author names.

Network structure
using edge bundling.

Color coded cluster
hierarchy according to
Blondel community
detection algorithm.

Node Color
based on NUMBEROFWORKS attribute

1   64   127
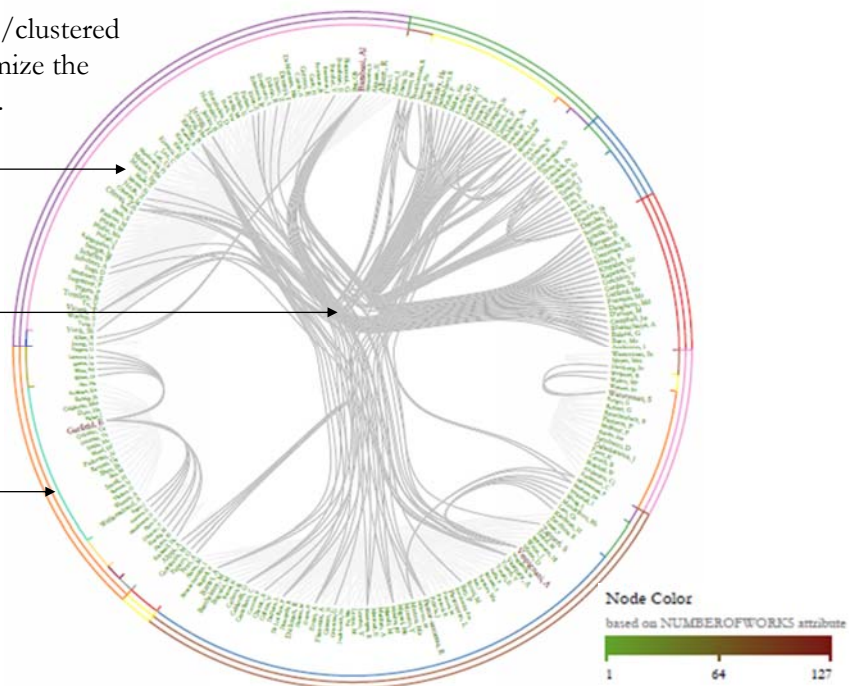
Note:
Header/footer info, legend, and more meaningful color coding are under development.

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)
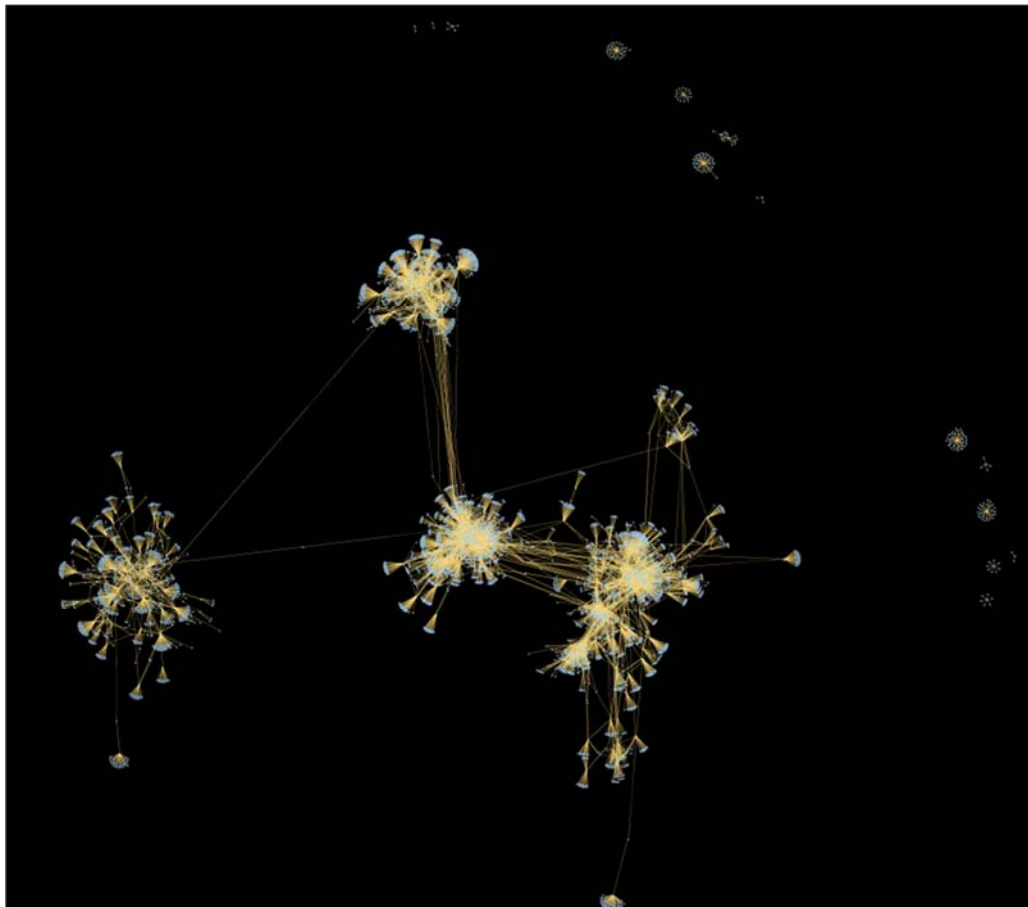
Run *'Analysis > Networks > Unweighted and Directed > Weak Component Clustering'* with parameters
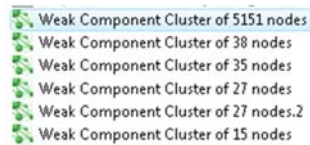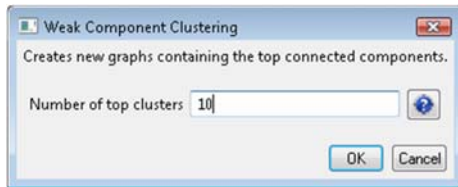
to identify top-10 largest components. The largest (giant) component has 5,151 nodes.

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

71

72

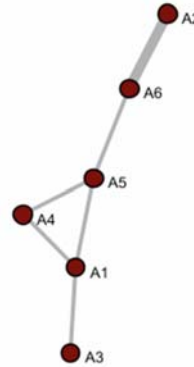|	A	B	C	D
1	Paper	Authors	References	Year
2	P1	A1		1970
3	P2	A2;A6	P1	1980
4	P3	A1;A3	P1;P2	1990
5	P4	A1;A4;A5	P2	1995
6	P5	A5;A6	P1;P2;P3;P4	1995
7	P6	A2;A6	P5	2000

Author co-occurrence network

Extract Network from Table

Extracts a network from a delimited table

| Column Name | Authors |
| 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

73

	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
■ Paper

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.
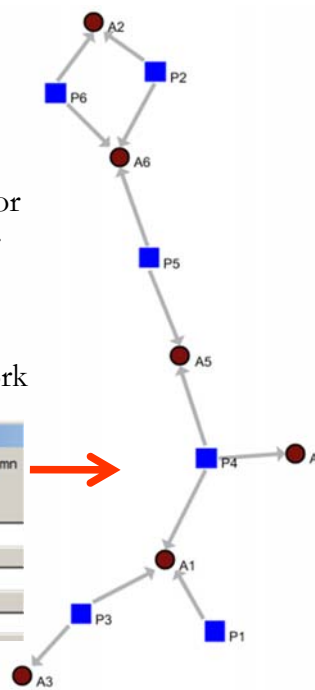
First column	Paper
Second column	Authors
Text Delimiter	;

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

74

General Network Extraction: Unweighted, Directed Network

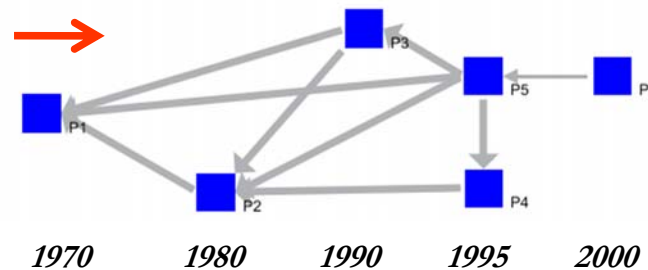|	A	B	C	D
1	Paper	Authors	References	Year
2	P1	A1		1970
3	P2	A2;A6	P1	1980
4	P3	A1;A3	P1;P2	1990
5	P4	A1;A4;A5	P2	1995
6	P5	A5;A6	P1;P2;P3;P4	1995
7	P6	A2;A6	P5	2000

● Author
■ Paper

Extract Directed Network

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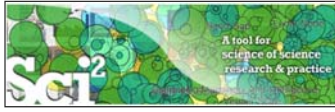column node.

Source Column: Paper
Target Column: Authors
Text Delimiter: ;

Analysis Modeling Visualization R Help
- Temporal ▶
- Geospatial ▶
- Topical ▶
- Networks ▶
  - Network Analysis Toolkit (NAT)
  - Unweighted & Undirected ▶
  - Weighted & Undirected ▶
  - Unweighted & Directed ▶ → Node Indegree

...andom 1 ... om the Layout mer
und in
hell-Session-46639... ...9875114...
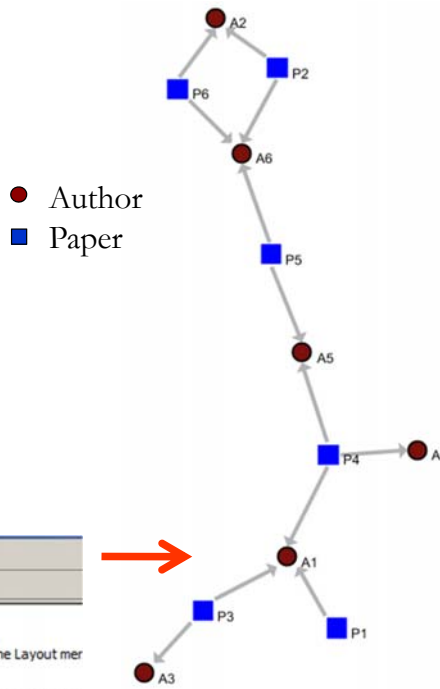
*Vertices 12
1 P1 indegree 0
2 A1 indegree 3
3 P2 indegree 0
4 A2 indegree 2
5 A6 indegree 3
6 P3 indegree 0
7 A3 indegree 1
8 P4 indegree 0
9 A4 indegree 1
10 A5 indegree 2
11 P5 indegree 0
12 P6 indegree 0
*Arcs
1 2
3 4
3 5
6 2
6 7
8 10
8 2
8 9
11 10
11 5
12 4
12 5

General Network Extraction: Unweighted, Directed Paper-Citation Network

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

Extract Directed Network

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

Source Column: Paper
Target Column: References
Text Delimiter: ;

Arcs from papers to references

1970   1980   1990   1995   2000

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

General Network Extraction: Unweighted, Directed Bi-Partite Network

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

Extract Bipartite Network

Extract a bipartite network values may list multiple ent them.

First column: Paper
Second column: References
Text Delimiter: ;
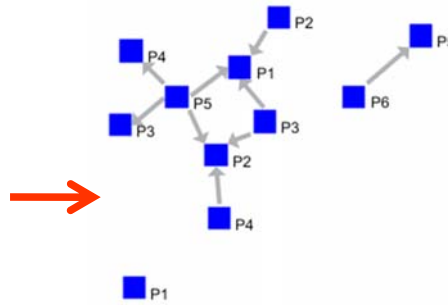
WRONG!!!

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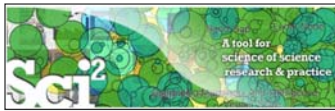
ISI Paper-Citation Network Extraction

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

Sci2 Tool

File Data Preparation Preprocessing Analysis

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

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

A1, 2001, J Inform Sci, V30, P119
A1, 2000, Int Microbiol, V10, P65
A1, 2002, Inform Technol Libr, V22, P183
A2, 2001, Scientometrics, V60, P295
A1, 2000, P Asist Annu Meet, V40, P486

2000   2001   2002

Break

Tutorial Overview

10:30a Welcome and Overview of Tutorial and Attendees
10:45a Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools
11:00a Sci2 Tool Basics
➢ Download and run the Sci2 Tool
➢ Load, analyze, and visualize family and business networks
➢ Horizontal line graph of NSF projects
➢ Studying four major network science researchers
- Load and clean a dataset; process raw data into networks
- Find basic statistics and run various algorithms over the network
- Visualize as either a circular hierarchy or network
12:30 Lunch Break
1:30 Sci2 Tool Novel Functionality
➢ **Yahoo! Geocoder**
➢ Evolving collaboration networks
➢ R-Bridge
3:00 Outlook and Q&A
3:30 Adjourn

Yahoo! Geocoding and Geospatial Maps

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

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

Aggregate TC over Country   Restrict to US records
                                                        Aggregate TC over ZIP

Use GoeMap (region)   Use GeoMap (circle)
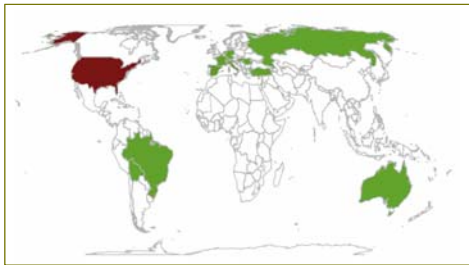
World map colored by sum of TC per country
(Choropleth Map)

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

81

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
	AQ2			fx	Wuchty, S, Northwestern Univ, NW Inst Complex, Chambers Hall,600 Foster St, Evanston, IL 60202 USA.							
1	Publicatio	Publisher	Publisher	Publisher	Reprint Address	research	Research	Research	Special Is:	Subject C	Suppleme	Times Cite
2	2006	BIOMED C	MIDDLESEX HOUSE, 3		Wuchty, S, Northwestern Univ, NW Inst Complex, Chambers Hall	Northwestern Univ, NW Inst Complex, Ev	Evolutionary Biology					7
3	2005	BLACKWEI	9600 GARSINGTON R	D, OXFORD OX4 2DQ,	XON, ENGLAND	Univ Pittsburgh, Dept Pathol, Pittsburgh,	Biochemi:	Suppl. 1				0
4	2005	BLACKWEI	9600 GARSINGTON R	D, OXFORD OX4 2DQ,	XON, ENGLAND	Univ Notre Dame, Dept Phys, Notre Dame	Biochemi:	Suppl. 1				0
5	2005	NATURE P	MACMILLAN BUILDIN	C	Barabasi, AL, Harvard Univ, Dana Farber Canc Inst, Ctr Canc Syst Bi	Harvard Univ, Dana Farber Canc Inst, Ctr C	Physics, Multidiscipl					2
6	2005	EDP SCIEN	17, AVE DU HOGGAR,	I	Macdonald, PJ, Univ Notre Dame, Ctr Network Res, Notre Dame,	Univ Notre Dame, Ctr Network Res, Notre	Physics, Multidiscipl					14
7	2005	NATURE P	MACMILLAN BUILDIN	C	Oliveira, JG, Univ Notre Dame, Ctr Complex Network Res, Notre I	Univ Notre Dame, Ctr Complex Network F	Multidisciplinary Sci					15
8	2005	NATL ACA	2101 CONSTITUTION	A	Oltvai, ZN, Northwestern Univ, Dept Pathol, Chicago, IL 60611 US.	Northwestern Univ, Dept Pathol, Chicago	Multidisciplinary Sci					29
9	2005	NATURE P	MACMILLAN BUILDIN	C	Barabasi, AL, Univ Notre Dame, Ctr Complex Networks Res, Notre	Univ Notre Dame, Ctr Complex Networks	Multidisciplinary Sci					32

2 missing values

82

Relevant CIShell plugin

CIShell Yahoo Geocoder

Edit   Add ▾   Tools ▾

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

Description

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

Pros & Cons

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

Applications

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

http://wiki.cns.iu.edu/display/CISHELL/Yahoo+Geocoder
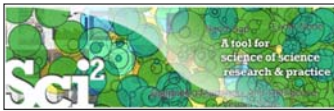
Geocode Address Using Yahoo! Geocoder

Run *'Analysis > Geospatial > Yahoo Geocoder'*
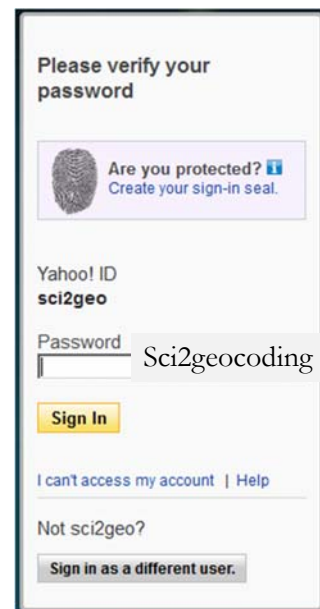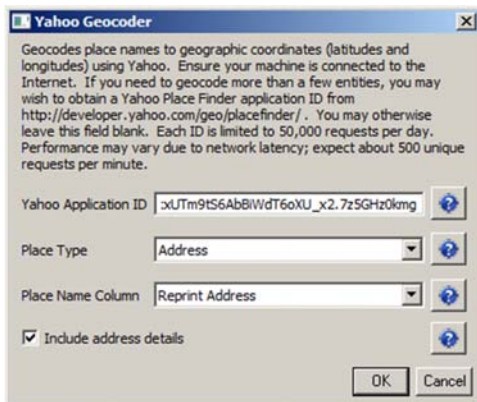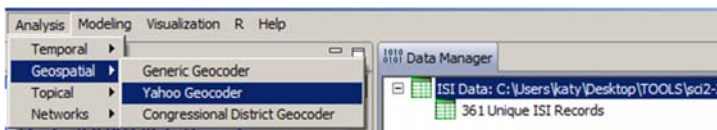
2lfKwqrV34HlII4_c2G5IGcHqh6XkxUTm9tS6AbBiWdT6oXU_x2.7z5GHz0kmg

Geocode Address Using Yahoo! Geocoder

Result is a new file in Data Manager
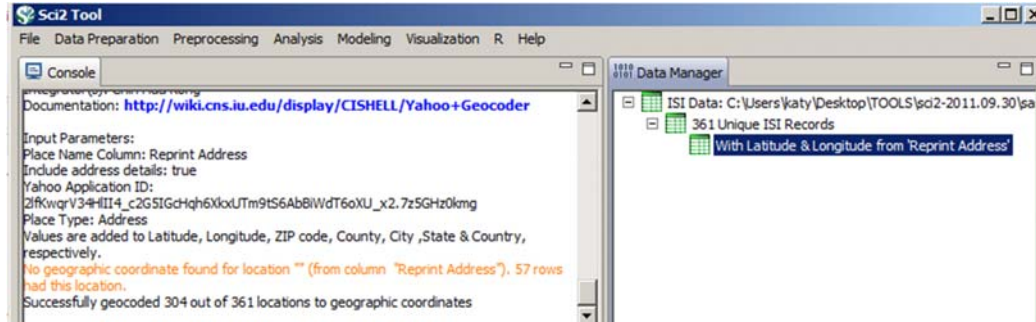
with 7 new columns

BF	BG	BH	BI	BJ	BK	BL
Latitude	Longitude	ZIP code	County	City	State	Country
42.02946091	-87.68838501	60202	Cook Coun	Evanston	Illinois	United States

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

42.34999466	-71.08765411	2115	Suffolk Co	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 Coun	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

---

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\sampledata\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: 2lfKwqrV34HlII4_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
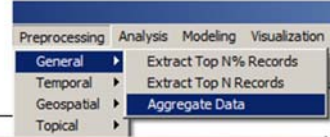
Aggregate Data was selected.
Implementer(s): Chintan Tank
Documentation: **http://wiki.cns.iu.edu/display/CISHELL/Aggregate+Data**
Input Parameters:
Aggregate on column: Country
Delimiter for Country: |
Longitude: AVERAGE
Latitude: AVERAGE
Times Cited: SUM
Aggregated by ": All rows of Latitude column 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 val
Frequency of unique "Country" values added to "Count" column.
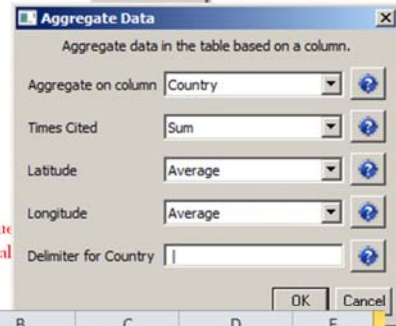
	A	B	C	D
1	Times Cited	Latitude	Longitude	Country
2	7	42.02946091	-87.68838501	United States
3	0			
4	0			
5	2	42.34999466	-71.08765411	United States
6	14	41.70074844	-86.23918915	United States
7	15	41.70074844	-86.23918915	United States
8	29	41.89422607	-87.61901855	United States
9	32	41.70074844	-86.23918915	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

87
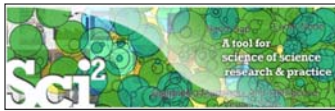
Aggregate + Geo Map (Colored-Region Annotations)
Also called 'Choropleth Map'

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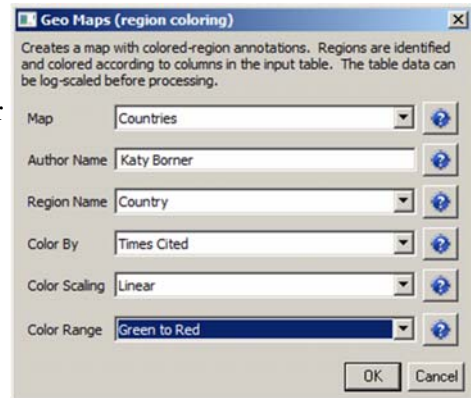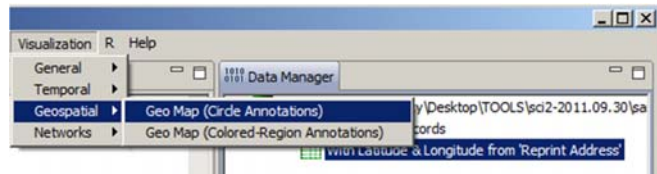
Save map to Desktop, view with PostScript viewer or convert into pdf to view.

Vector format - can be rendered LARGE.

88

How to Read the Geo Map

World Map with data overlay.

Listing of map type, author, and parameters used.

Map legend with color coding.

Generated from Aggregation performed using unique values in 'Country' column.A

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

Created with Sci² Tool | Cyberinfrastructure for Network Science Center (http://cns.iu.edu)

Region Color (Linear)
Times Cited

0   7,340   14,680

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 computatio
science. Here we use ISI data from the journal "Scientometrics" and
Awards Search.

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

Geo Map (Circle Annotation Style)
Eckert IV Projection
Mar 28, 2011 | 09:04:47 AM

Created with Sci² Tool | Cyberinfrastructure for Network Science Center (http://cns.iu.edu)

Extrude Color (Linear)
Publication Year

1978   1993   2008

Area (Linear)
Times Cited
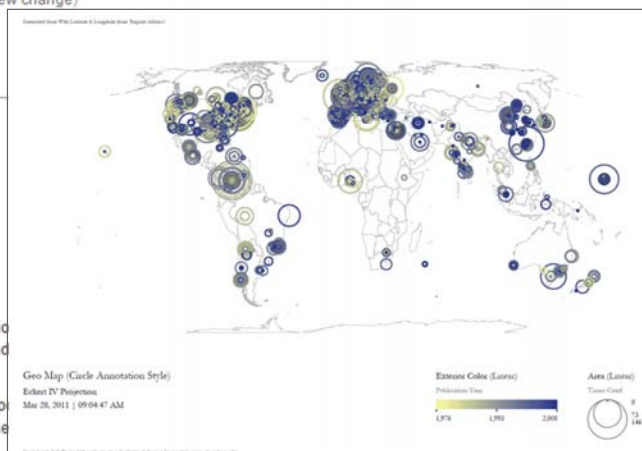
New ISI File Format

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

Restrict to 195 US records

Aggregate by ZIP code

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.

Different parameter values result in different maps and corrsponding legends:

Exterior Color (Linear)
Count

1   33

Geo Maps (circles)

Creates a map with circle annotations. Circles are positioned, sized, and colored (inside and outside) according to columns in the input table. Either or both kinds of coloring can also be disabled. The table data for each dimension can be log-scaled before processing.

Map	US States
Author Name	Katy Borner
Latitude	Latitude
Longitude	Longitude
Size Circles By	Times Cited
Size Scaling	Linear
Color Circle Exteriors By	Count
Exterior Color Scaling	Linear
Exterior Color Range	Gray to Black
Color Circle Interiors By	None (no inner color)
Interior Color Scaling	Linear
Interior Color Range	Yellow to Blue

Exterior Color (Linear)
Count

1   33   66

Area (Linear)
Times Cited

0

4,719
9,438

93

Tutorial Overview

10:30a Welcome and Overview of Tutorial and Attendees
10:45a Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools
11:00a Sci2 Tool Basics
➢   Download and run the Sci2 Tool
➢   Load, analyze, and visualize family and business networks
➢   Horizontal line graph of NSF projects
➢   Studying four major network science researchers
      - Load and clean a dataset; process raw data into networks
      - Find basic statistics and run various algorithms over the network
      - Visualize as either a circular hierarchy or network
12:30 Lunch Break
1:30 Sci2 Tool Novel Functionality
➢   Yahoo! Geocoder
➢   **Evolving collaboration networks**
➢   R-Bridge
3:00 Outlook and Q&A
3:30 Adjourn

94

![Sci2 logo: A tool for science of science research & practice]

**Sci2 Demo II:
Evolving collaboration networks**

Legend

Nodes ~ Authors
Node area size ~ Number of papers published
Node color ~ Number of citations

Edges ~ Co-authorship relations
Edge color ~ Year of first co-authorship

Node Color Code	
0 - 9	
10 - 19	
20 - 29	
30 - 39	
40 - 49	
50 -	

Edge Color Code	
86 - 90	
91 - 95	
96 - 00	
01 -	

Displayed Year: **1988**

**Mapping the Evolution of
Co-Authorship Networks**
Weimao Ke, Lalitha Visvanath & Katy Börner
InfoVis Lab @ Indiana University
2004

![Sci2 logo: A tool for science of science research & practice]

Evolving Collaboration Networks

Load isi formatted file
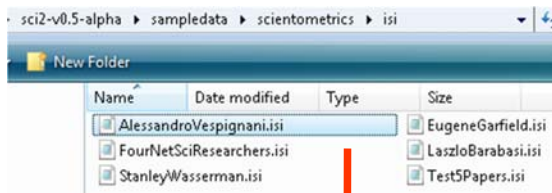
As csv, file looks like:

	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 sp	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:

1990-1991   1990-1996   1990-2001   1990-2006

Relevant Sci2 Manual entry

Slice Table by Time

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

101 Unique ISI Records
- slice from beginning of 1990 to end of 2006 (101 records)
- slice from beginning of 1990 to end of 2001 (65 records)
- slice from beginning of 1990 to end of 1996 (26 records)
- slice from beginning of 1990 to end of 1991 (4 records)

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

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

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

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)
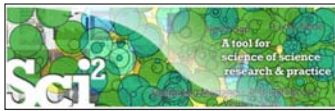
Visualize Each Network, Keep Node Positions
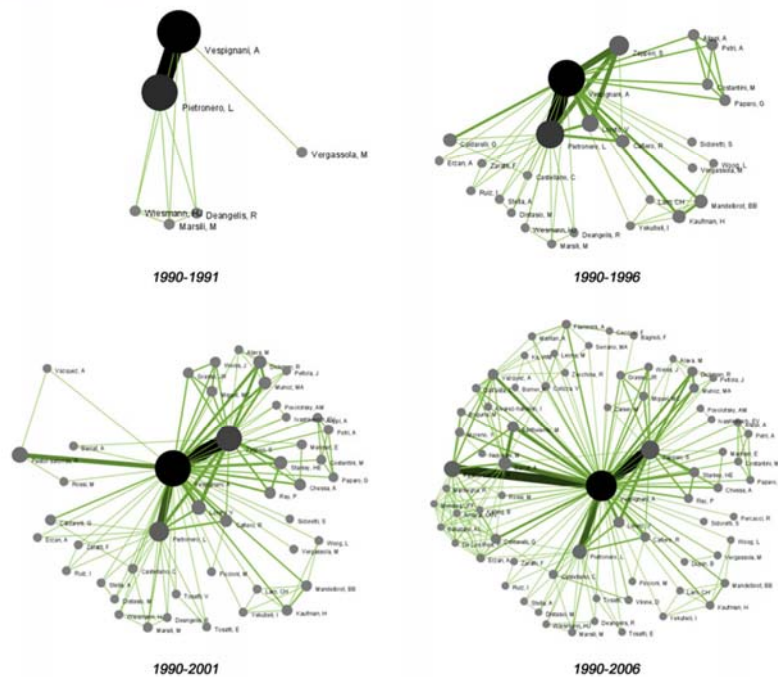
1990-1991   1990-1996

1990-2001   1990-2006

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

Relevant CIShell plugin

CIShell — Slice Table by Time

Tools ▾

Added by Aretha 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

Tutorial Overview

10:30a Welcome and Overview of Tutorial and Attendees
10:45a Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools
11:00a Sci2 Tool Basics
➢ Download and run the Sci2 Tool
➢ Load, analyze, and visualize family and business networks
➢ Horizontal line graph of NSF projects
➢ Studying four major network science researchers
  - Load and clean a dataset; process raw data into networks
  - Find basic statistics and run various algorithms over the network
  - Visualize as either a circular hierarchy or network
12:30 Lunch Break
1:30 Sci2 Tool Novel Functionality
➢ Yahoo! Geocoder
➢ Evolving collaboration networks
➢ **R-Bridge**
3:00 Outlook and Q&A
3:30 Adjourn

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.

Tutorial Overview

10:30a Welcome and Overview of Tutorial and Attendees
10:45a Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools
11:00a Sci2 Tool Basics
➢ **Download and run the Sci2 Tool**
➢ **Load, analyze, and visualize family and business networks**
➢ **Horizontal line graph of NSF projects**
➢ **Studying four major network science researchers**
  - **Load and clean a dataset; process raw data into networks**
  - **Find basic statistics and run various algorithms over the network**
  - **Visualize as either a circular hierarchy or network**
12:30 Lunch Break
1:30 Sci2 Tool Novel Functionality
➢ **Yahoo! Geocoder**
➢ **Evolving collaboration networks**
➢ **R-Bridge**
3:00 Outlook and Q&A
3:30 Adjourn

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
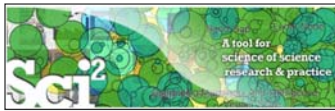
Geomap with Gephi Network Overlay
See 4.7.6 on http://sci2.wiki.cns.iu.edu

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

Read into Sci2 Tool to generate geomap and network file ➡ Layout network in Gephi ➡ Combine geomap and network in Photoshop
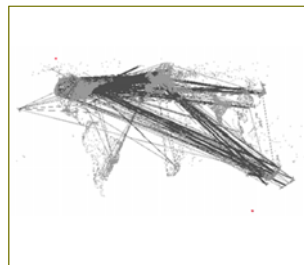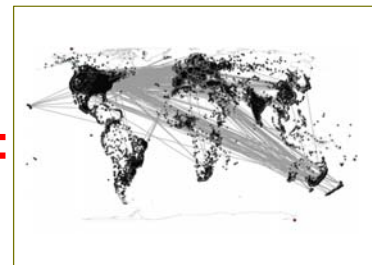
Relevant Sci2 Manual entry

4.7.6 Using Gephi to Render Networks Overlaid on Geo Maps

Loading and Saving Geovisualization Files in Sci2

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

1. Load this network in Sci2.

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

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

---

Use Sci2 Tool to Generate Geomap and Network File

Read prepared .net file and run:

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

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

Start gephi. Use *New Project > Open a graph file* to read .graphml file that Sci2 generated.
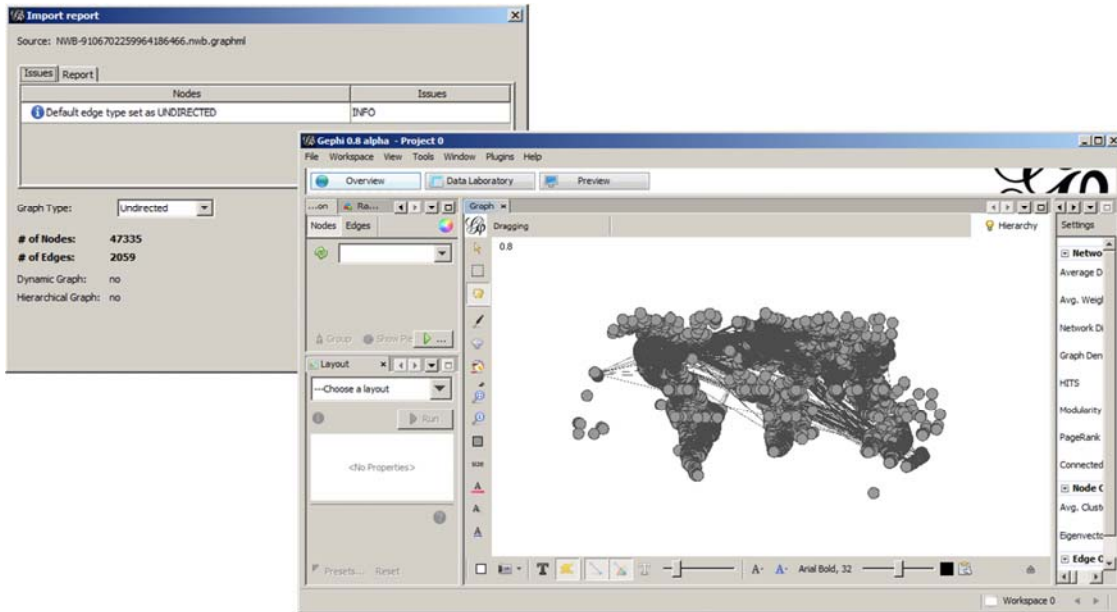
Follow instructions in online tutorial on **Manipulating the Network File in Gephi**

Color or size code the "Near Alaska" and "Near Antarctica" anchor nodes to ease alignment of geomap and network overlay, see instructions in online tutorial on **Manipulating the Network File in Gephi.** Save result using *File > Export > SVG/PDF file.*
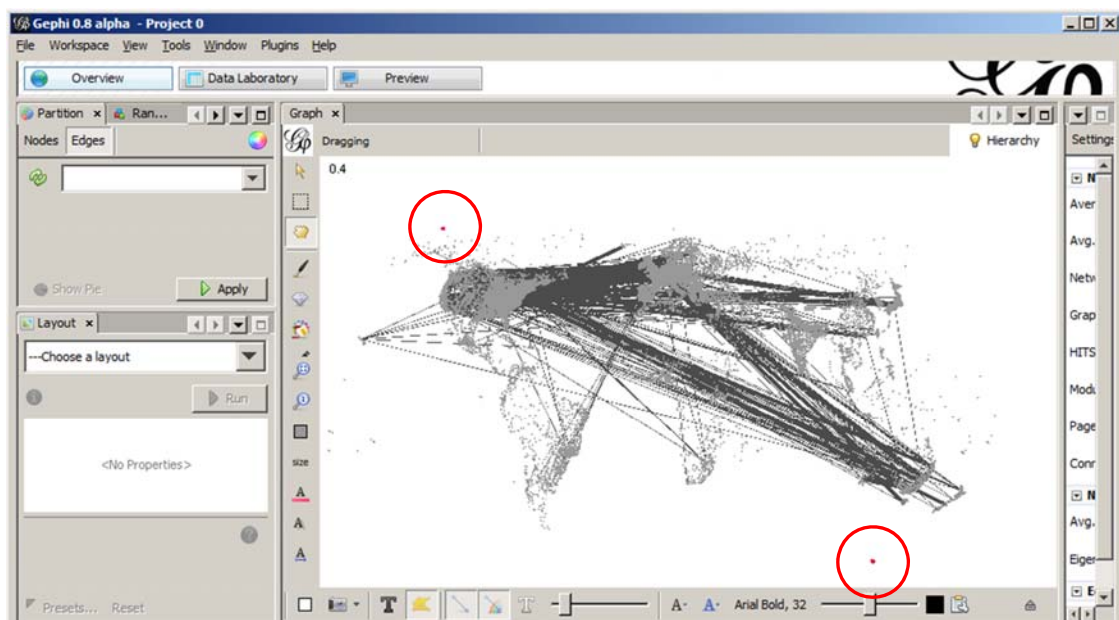
Load geomap and network files into Photoshop. Select 'network' layer an use 'Right click, Duplicate Layer' to copy network over to 'geomap' file as a second layer.
Use Edit > Transform > Scale' and align using the "Near Antarctica" anchor nodes, see instructions in online tutorial on **Creating the Visualization in Photoshop.**

111
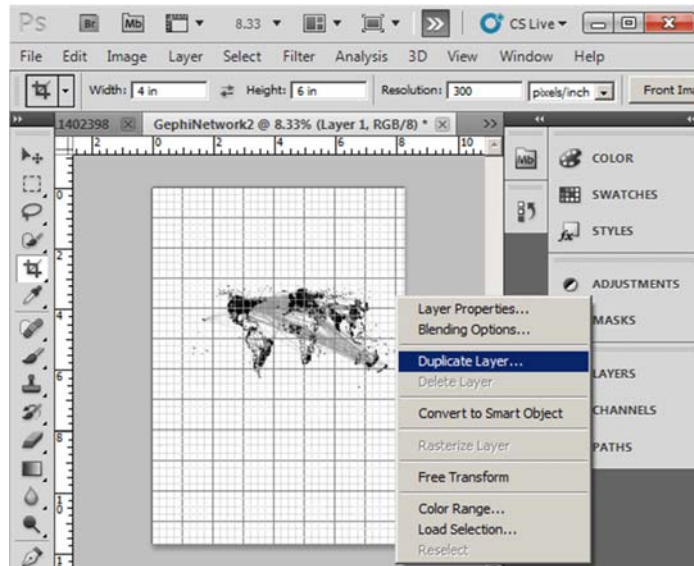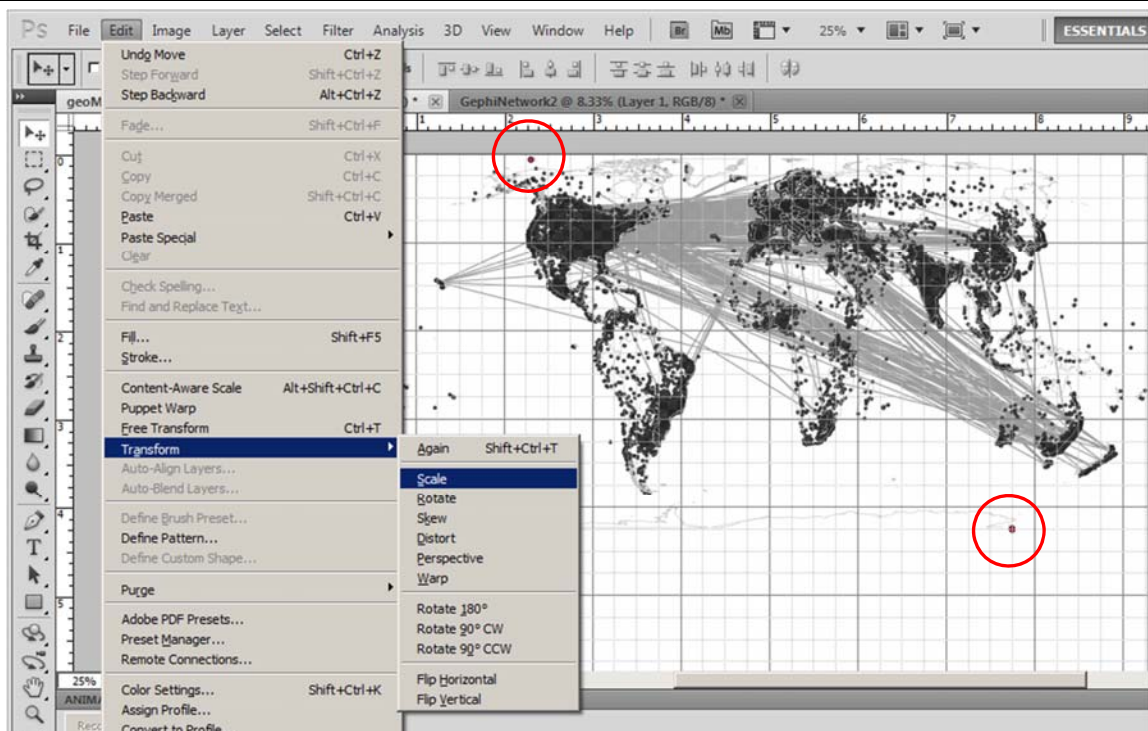
112

Delete anchor nodes and save in preferred format.
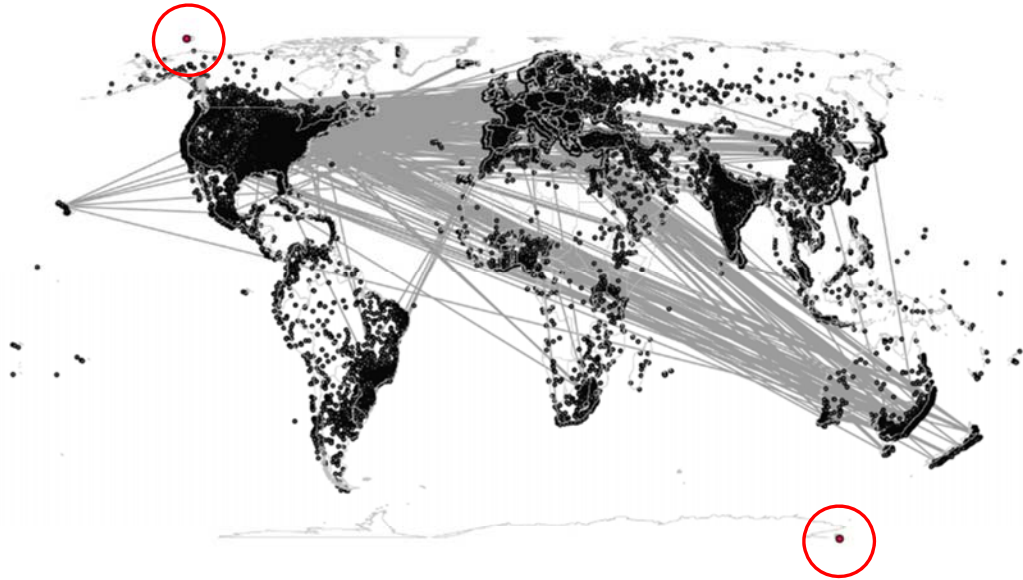
Geo Map ()
Eckert IV Projection
Apr 06, 2012 | 03:19:51 AM

Created with Sci² Tool | Cyberinfrastructure for Network Science Center (http://cns.iu.edu)

Practice these steps using "LaszloBarabasi-collaborations.net" linked from Sci2 wiki:

4.7.6 Using Gephi to Render Networks Overlaid on GeoMaps

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

Created with Sci² Tool | Cyberinfrastructure for Network Science Center (http://cns.iu.edu)

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.
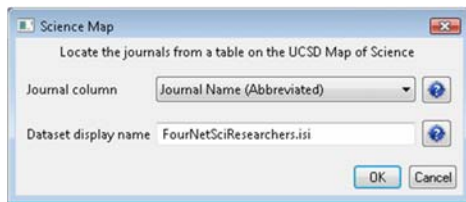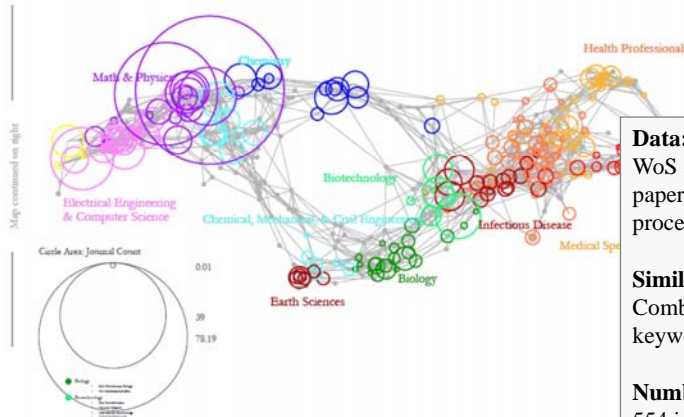
Data:
WoS and Scopus for 2001–2005, 7.2 million papers, more than 16,000 separate journals, proceedings, and series

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.

Science Map
Locate the journals from a table on the UCSD Map of Science
Journal column: Journal Name (Abbreviated)
Dataset display name: FourNetSciResearchers.isi
OK   Cancel

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.
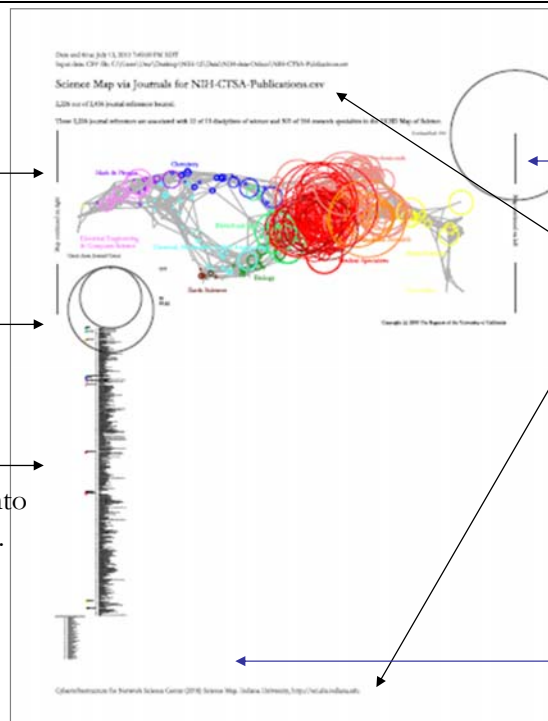
Circle of non-located, e.g., 'Unclassified' records.

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

Listing and circle of non-located, e.g., 'Unclassified' records.

Computational Scientometrics Cyberinfrastructures

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

VIVO Research Networking
http://vivoweb.org
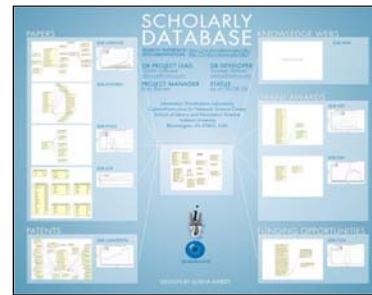
Information Visualization Cyberinfrastructure
http://iv.cns.iu.edu

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

Science of Science (Sci²) Tool
http://sci2.cns.iu.edu

Epidemics Tool & Marketplace
Forthcoming
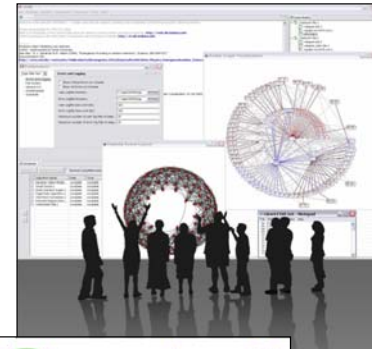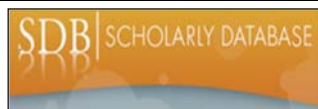
Scholarly Database at Indiana University
http://sdb.wiki.cns.iu.edu

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

Since March 2009:
Users can download networks:
 - Co-author
 - Co-investigator
 - Co-inventor
 - Patent citation
and tables for
burst analysis in NWB.

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.
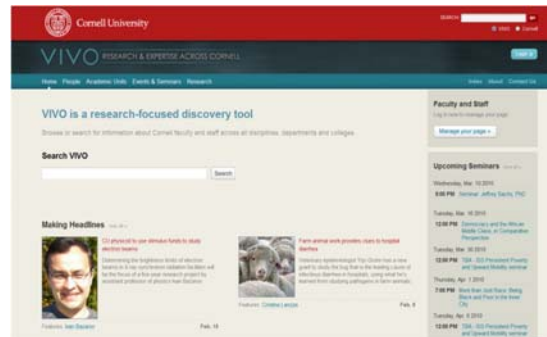
Funded by $12 million NIH award.

Cornell University: Dean Krafft (Cornell PI), Manolo Bevia, Jim Blake, Nick Cappadona, Brian Caruso, Jon Corson-Rikert, Elly Cramer, Medha Devare, John Fereira, 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 Price, 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, Michaeleen 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.

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

123

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

124
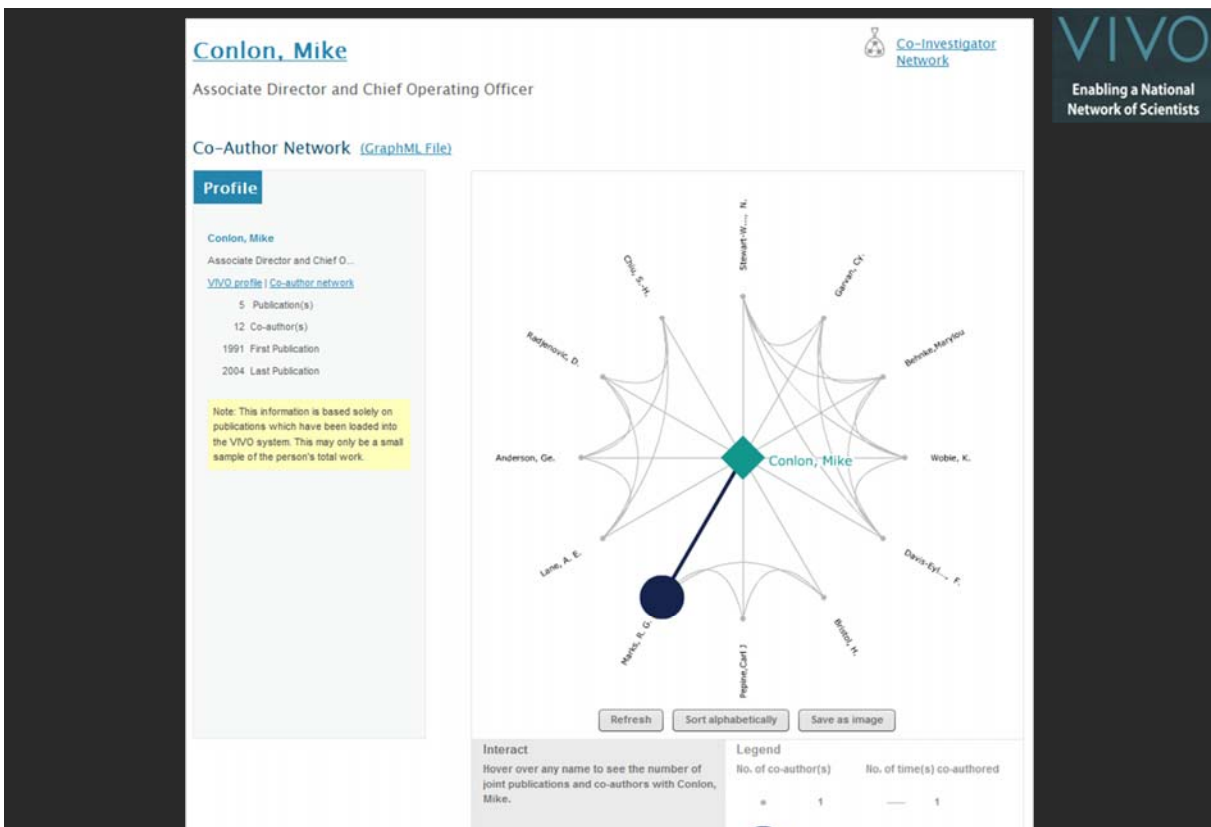
Network Analysis (With Whom?) Who is co-authoring, co-investigating, co-inventing with whom? What teams are most productive in what projects?

http://nrn.cns.iu.edu
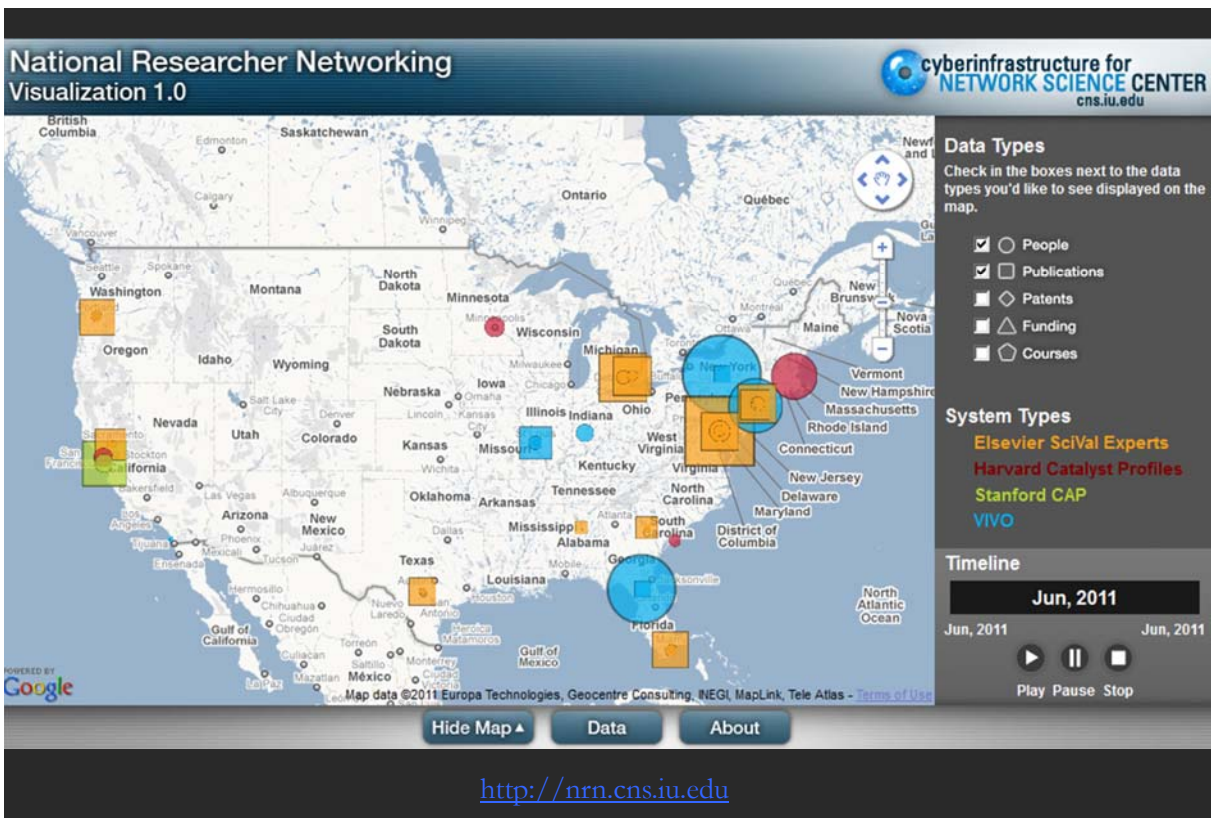
Geospatial Analysis (Where) Where is what science performed by whom? Science is global and needs to be studied globally.
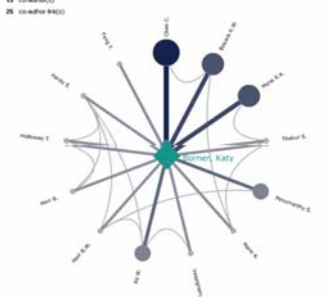
Download Data

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

127

36 publication(s) from 2001 to 2010 (.CSV File)

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

80 co-author(s) from 2001 to 2010 (.CSV File)

Year	Count	Co-Author(s)
2001	1	Chen C.
2002	3	Chen C.; McMahon T.; Feng Y.
2003	2	Chen C.; Boyack K.W.
2004	17	Sengupta A.; Penumarthy S.; Thakur S.; Sooriamurthi R.; Maru J.T.; Shiffrin R.M.; Mane K.; Moor K.A.;

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" />
```
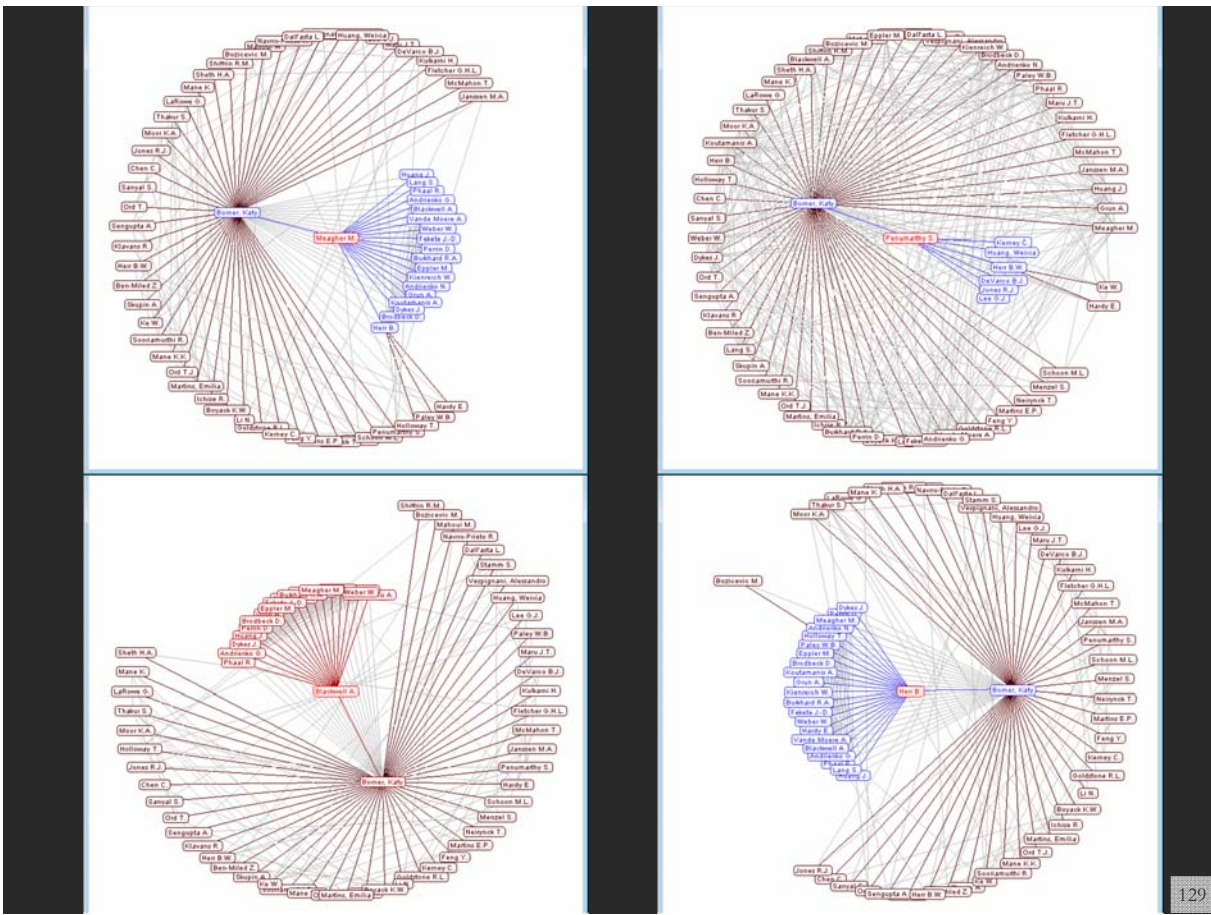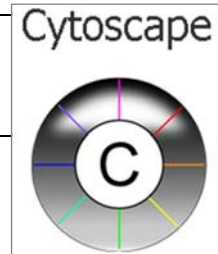
Save as Image (.PNG file)

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

128

129

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

Adding more alyout algorithms and network visualization interactivity via Cytoscape http://www.cytoscape.org.
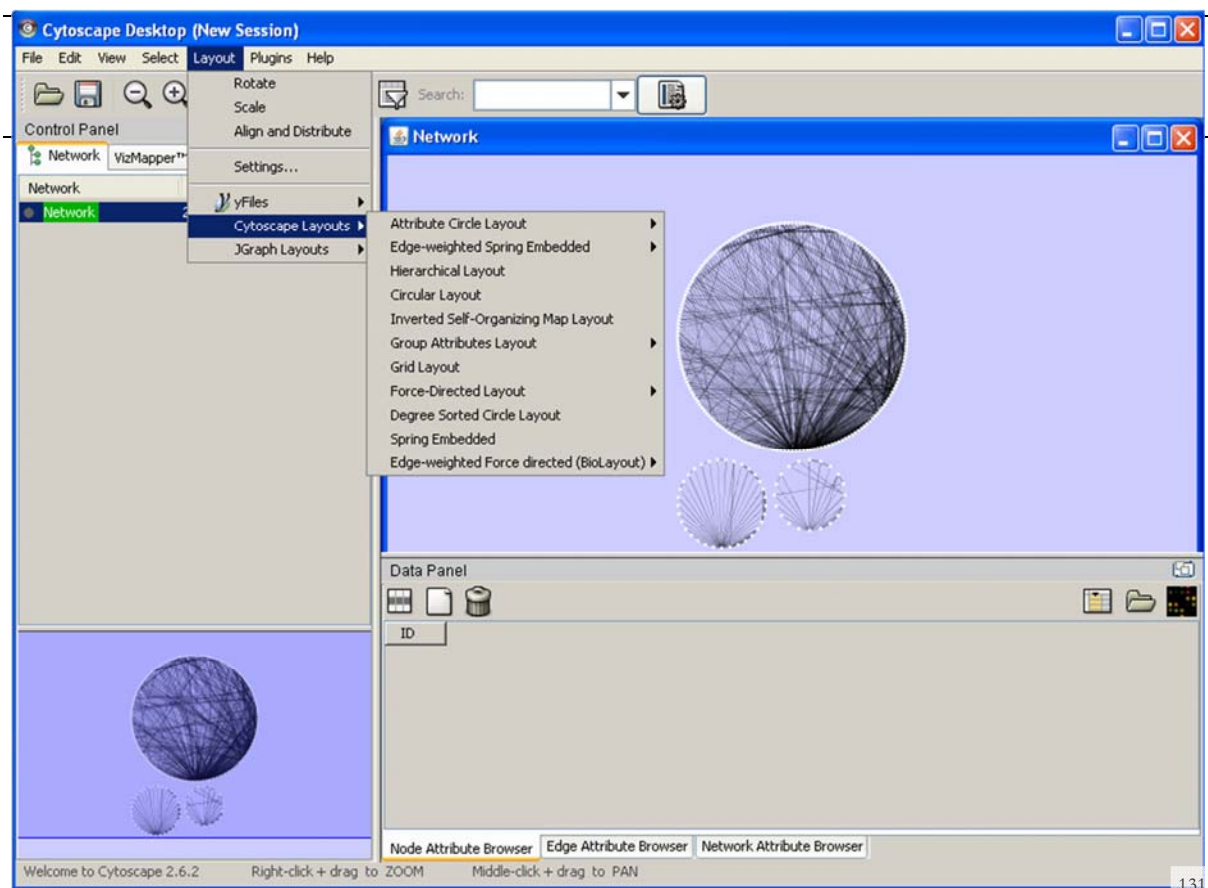
Simply add *org.textrend.visualization.cytoscape_0.0.3.jar* into your /plugin directory.

Restart Sci2 Tool.
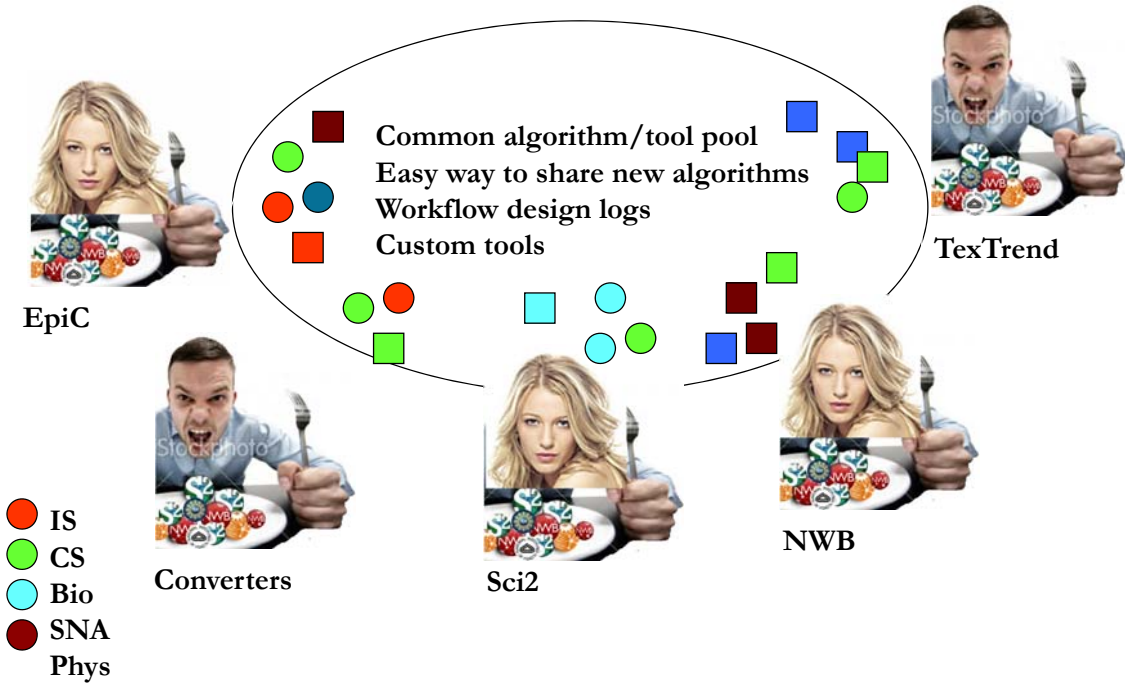
Cytoscape now shows in the Visualization Menu.
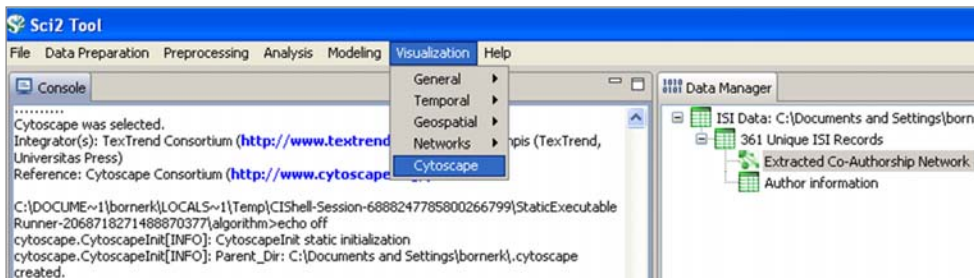
Select a network in Data Manager, run Cytoscape and the tool will start with this network loaded.

130

![CIShell logo - cyberinfrastructure shell]

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://textrend.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.

CIShell
cyberinfrastructure shell

Common algorithm/tool pool
Easy way to share new algorithms
Workflow design logs
Custom tools

EpiC

TexTrend

Converters

Sci2

NWB

- IS
- CS
- Bio
- SNA
- Phys

133

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 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.

Downloads for NWB Tool Releases

Cumulative Total of Downloads
89,559

- v.0.3
- v.0.4
- v.0.5
- v.0.6
- v.0.7 + post
- v.0.8 + pre
- v.0.9 + pre
- v.1.0.0 + pre
- Total # of Downloads
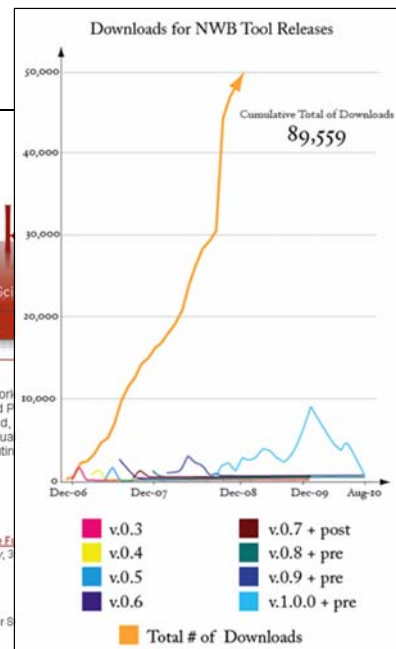
Herr II, Bruce W., Huang, Weixia (Bonnie), Penumarthy, Shashikant & Börner, Katy. (2007). Designing Highly Flexible and Usable Cyberinfrastructures for Convergence. In Bainbridge, William S. & Roco, Mihail C. (Eds.), Progress in Convergence - Technologies for Human Wellbeing (Vol. 1093, pp. 161-179), Annals of the New York Academy of Sciences, Boston, MA.

134

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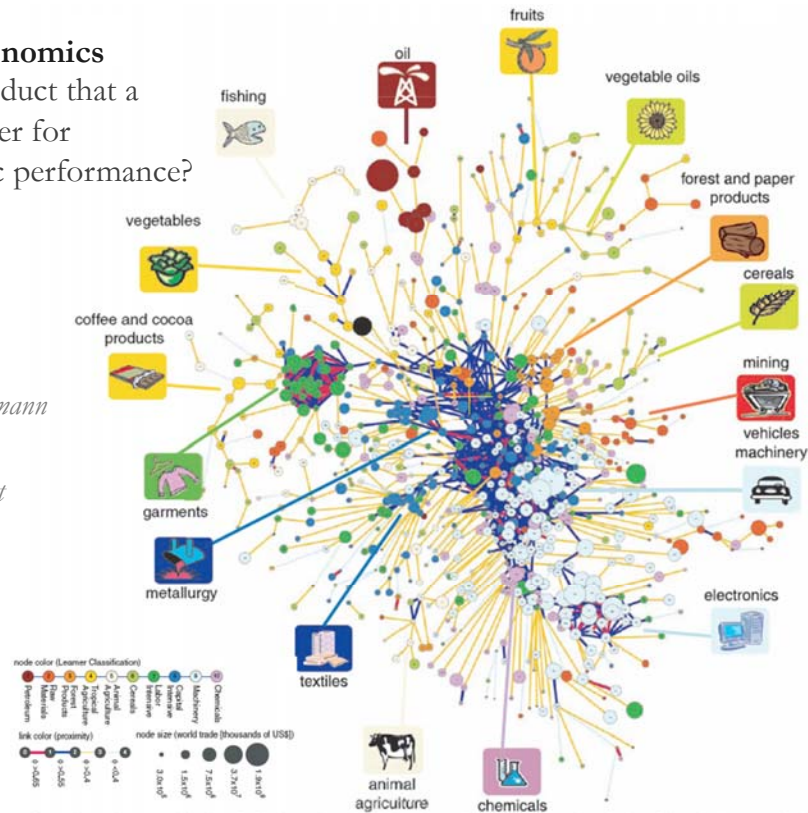
Fig. 1. The product space. **(A)** Hierarchically clustered proximity (φ) matrix representing the 775 SITC-4 product classes exported in the 1998–2000 period. **(B)** Network representation of the product space. Links are color coded with their proximity value. The sizes of the nodes are proportional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

1980

1990

2000

Germany

Japan

ProductSpaceParser {DEU|JPN} 0.5

Computational Social Science
Studying large scale social
networks such as Wikipedia

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

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).
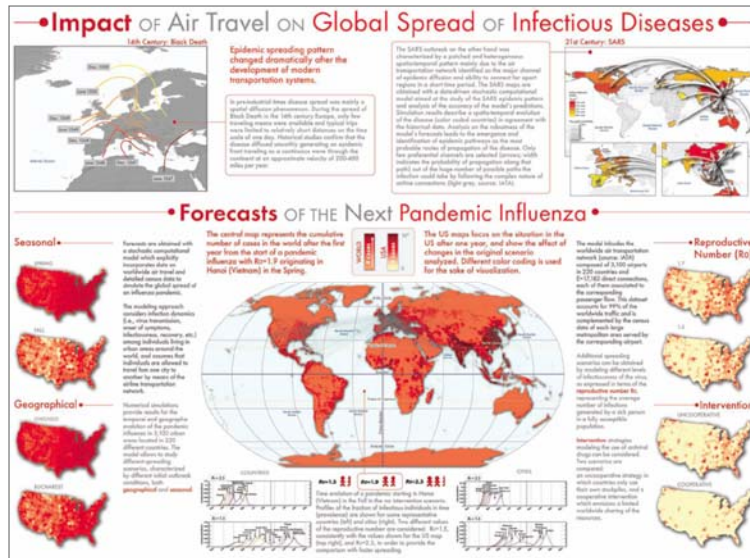

Computational Proteomics

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

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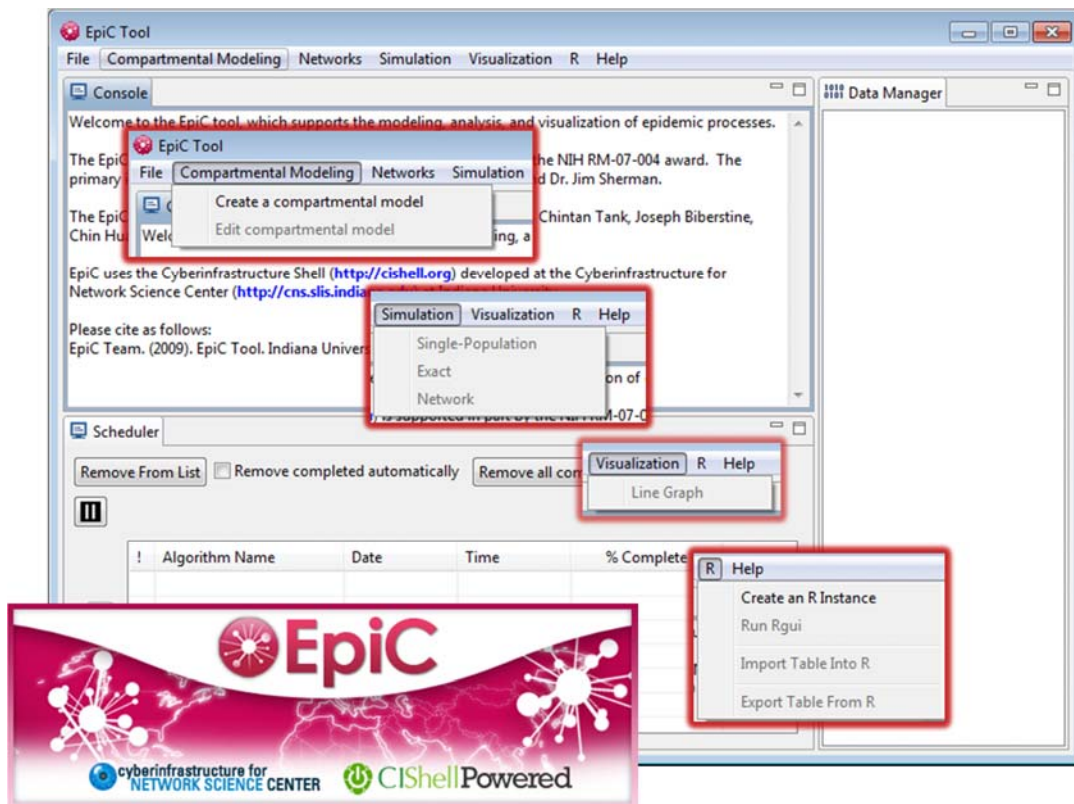
TEXTrend adds R bridge, WEKA, Wordij, CFinder, and more.

See the latest versions of TEXTrend Toolkit modules at

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

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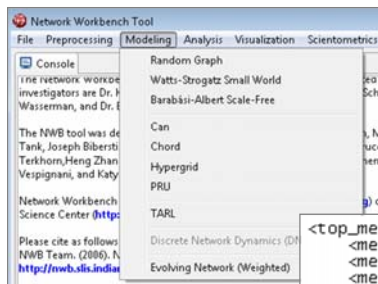
Additional Sci2 Plugins are at http://sci2.wiki.cns.iu.edu/3.2+Additional+Plugins

143

➢ 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:


```
<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.iu.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>
```

144

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

If you enjoyed the tutorial, please thank Andrea Scharnhorst and Albert Meroño Peñuela, KNAW; Fernando GALINDO-RUEDA, Yuko HARAYAMA, and Alessandra COLECCHIA, OECD

US side: Chin Hua Kong, Joseph R. Biberstine, Samantha Hale

Q & A

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

* * *

Bug reports and all comments are welcome.

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