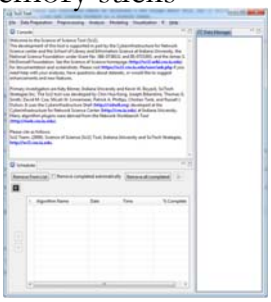
Before starting…

A copy of these slides and all other materials you will need during this tutorial can be obtained from any of the DVDs or USB memory sticks being passed around right now



- Please register, download, and decompress Sci² from http://sci2.cns.iu.edu
  - If you have any problems, ask


David Joseph

- Please try opening `PostScript test - chessboard.ps`
  - You should see

Sci² Manual: http://sci2.wiki.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

NWB/Sci²: A Tool for Science of Science Research and Practice


David M. Coe and Joseph Biberstine
Cyberinfrastructure for Network Science Center
School of Library and Information Science
Indiana University, Bloomington, IN
http://cns.iu.edu

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

Tuesday, June 19, 2011 • 09:00 – 12:00

Tutorial Overview

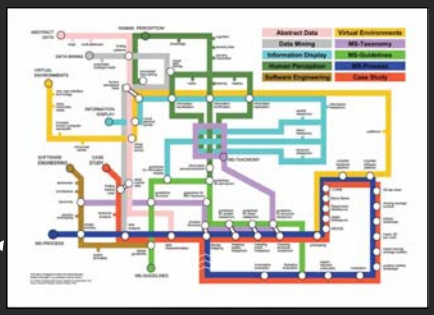
- Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools
- Sci² Tool Basics
  - Download and run the Sci² Tool
  - Walkthrough: Load, analyze, and visualize a network
  - Walkthrough: Analyzing the publications of four prominent network science researchers
    - Load and clean a dataset; extract networks from raw data
    - Calculate basic statistics and analyses of the network
    - Visualize the results
- Sci² Tool – Advanced Topics
  - Walkthrough: Visualizing temporal data for NSF projects
  - Walkthrough: Locating data on a geographic map
  - Walkthrough: Examining an evolving network
  - Interacting with the statistical toolkit R and the network visualization package Gephi
  - Sci² tool visualizations
    - Bipartite networks
    - Map of Science
- Outlook and Q&A
- Adjourn

Tutorial Overview

- Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools
- Sci² Tool Basics
- Sci² Tool – Advanced Topics
- Outlook and Q&A
- Adjourn

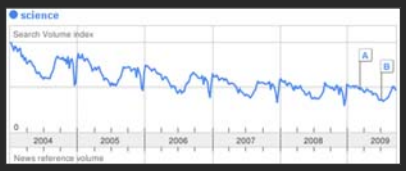
Take terabytes of data

Plug-and-Play Macroscopes



Find your way

Find collaborators, friends



Identify trends

Macroscopes

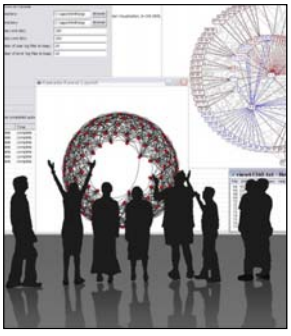
Decision making in science, industry, and politics, as well as in daily life, requires that we make sense of datasets representing the structure and dynamics of complex systems

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 making things larger or smaller, **macroscopes let us observe what is too great, slow, or complex for the human eye and mind to notice and comprehend**

Microscopes

Telescopes



Macroscopes

Plug-and-Play Macroscopes

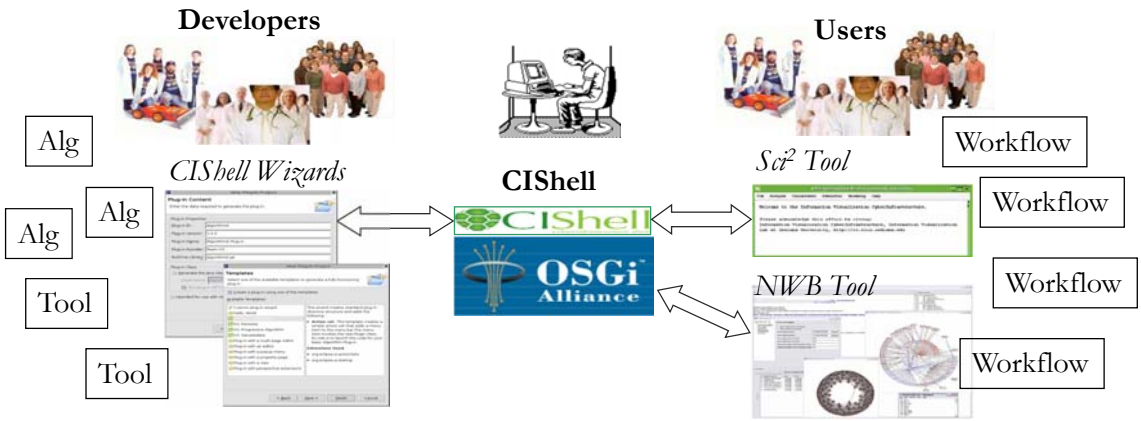
While microscopes and telescopes are physical instruments, macroscopes are **continuously changing bundles of software plugins**

Macroscopes make it easy to

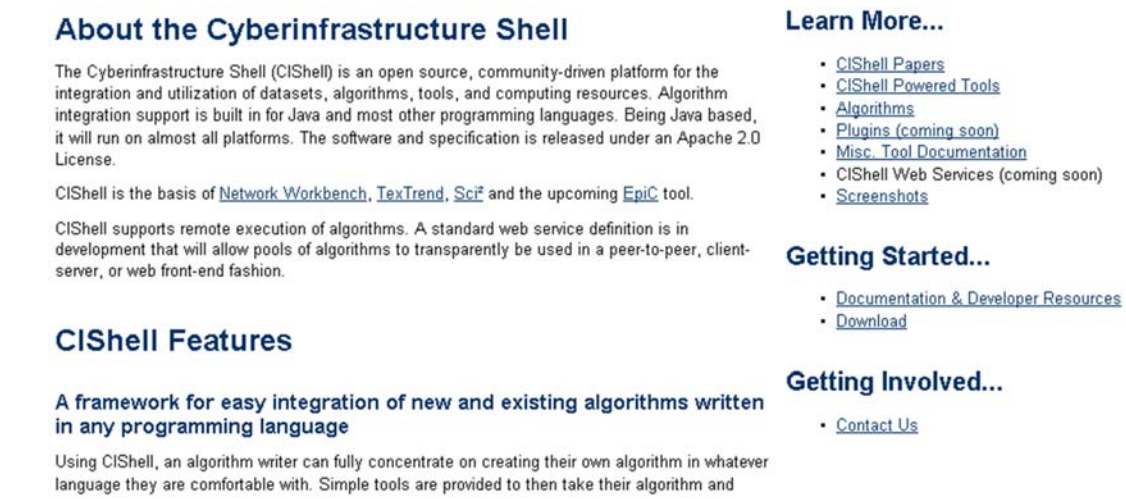
- Select and combine not only domain-specific algorithms and bridges to existing tools but also to meet the cross-cutting infrastructural requirements needed for a scientifically rigorous cyberinfrastructure
- Put together plugins to create customized tools
  - Share plugins via email, flash drives, or online
    - Simply drop plugins into the tool they appear in the menu, ready to use
    - Sharing algorithm components, tools, or novel interfaces becomes as easy as sharing images on Flickr or videos on YouTube

OSGi & Cyberinfrastructure Shell (CIShell)

- CIShell (http://cishell.org) is an open source software specification for the integration and utilization of datasets, algorithms, and tools
- It extends the Open Services Gateway Initiative (OSGi) (http://osgi.org), a standardized, modularized service platform
  - Widely used in industry for over 10 years
- CIShell provides "sockets" into which algorithms, tools, and datasets can be plugged using a wizard-driven process



CIShell – Integrate New Algorithms



About the Cyberinfrastructure Shell

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

CIShell is the basis of Network Workbench, TexTrend, 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

OSGi/CIShell Adoption

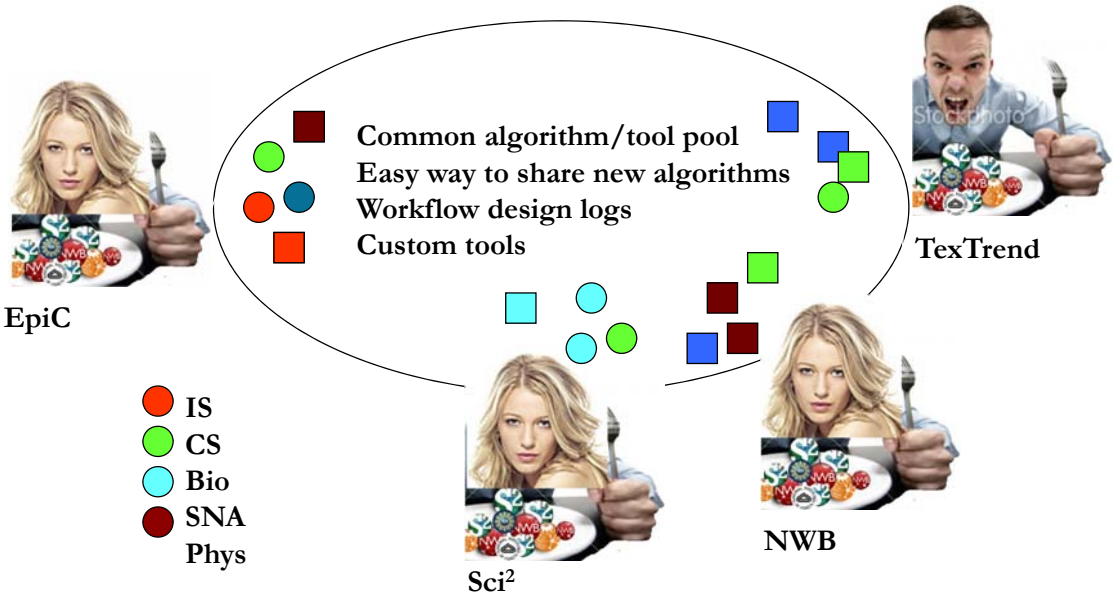
USA

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

OSGi/CIShell-Powered Tools Support Algorithm Sharing



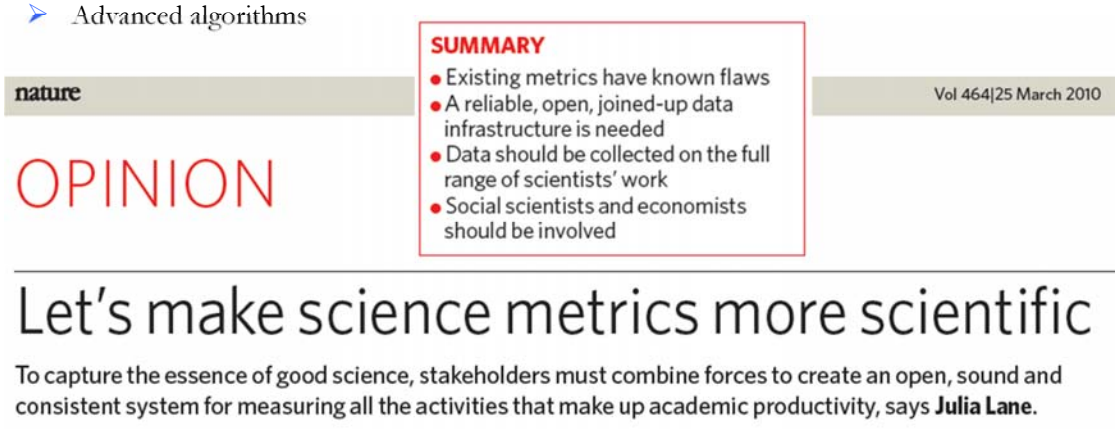
Tutorial Overview

- Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools
- Sci² Tool Basics
  - Download and run the Sci² Tool
  - Walkthrough: Load, analyze, and visualize a network
  - Walkthrough: Analyzing the publications of four prominent network science researchers
    - Load and clean a dataset; extract networks from raw data
    - Calculate basic statistics and analyses of the network
    - Visualize the results
- Sci² Tool – Advanced Topics
- Outlook and Q&A
- Adjourn

Science of Science (Sci^2) Tool

http://sci2.cns.iu.edu

- Built on CIShell/OSGi
- Explicitly designed for science of science research and practice
- Well-documented
- Easy to use
- Empowers average users to run common studies and expert users to perform novel research
- Advanced algorithms
- Effective visualizations
- Carefully designed and documented common workflows
- Full logging and data history for perfect replication of studies
- Free and open source software
  - Anyone can review and extend the code, or use it for commercial purposes



nature

Vol 464|25 March 2010

OPINION

SUMMARY

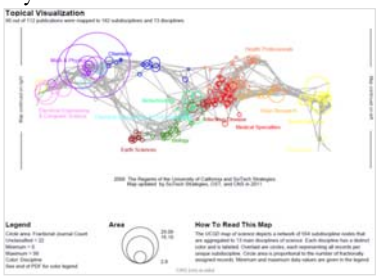
- Existing metrics have known flaws
- A reliable, open, joined-up data infrastructure is needed
- Data should be collected on the full range of scientists' work
- Social scientists and economists should be involved

Let's make science metrics more scientific

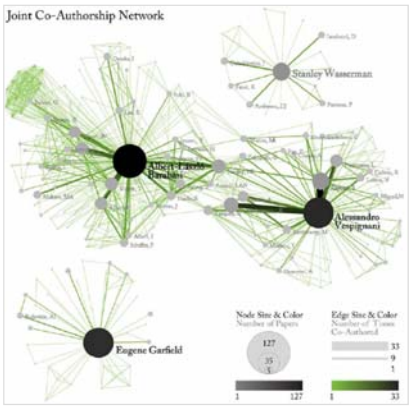
To capture the essence of good science, stakeholders must combine forces to create an open, sound and consistent system for measuring all the activities that make up academic productivity, says **Julia Lane**.

Sci^2 Tool

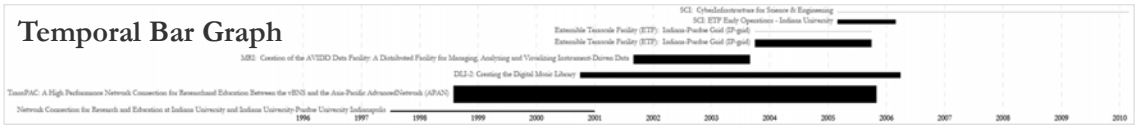
An OSGi/CIShell-powered tool with NWB plugins and many new scientometric and visualization plugins



Map of Science

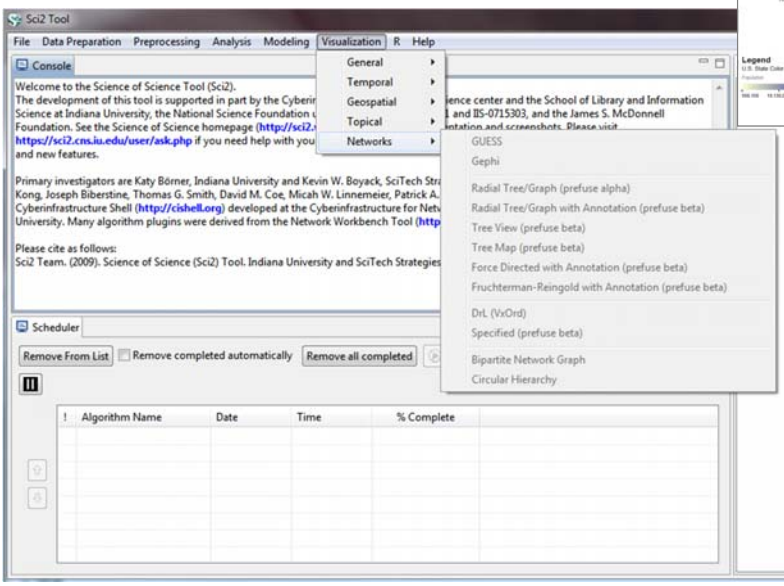


GUESS Network Visualization

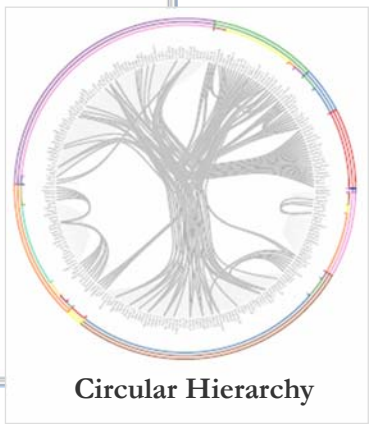


Börner, Katy, Huang, Weixia (Bonnie), Linnemeier, Micah, Duhon, Russell Jackson, Phillips, Patrick, Ma, Nianli, Zoss, Angela, Guo, Hanning & Price, Mark. (2009). Rete-Netzwerk-Red: Analyzing and Visualizing Scholarly Networks Using the Scholarly Database and the Network Workbench Tool. Proceedings of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

Sci² Tool Visualizations

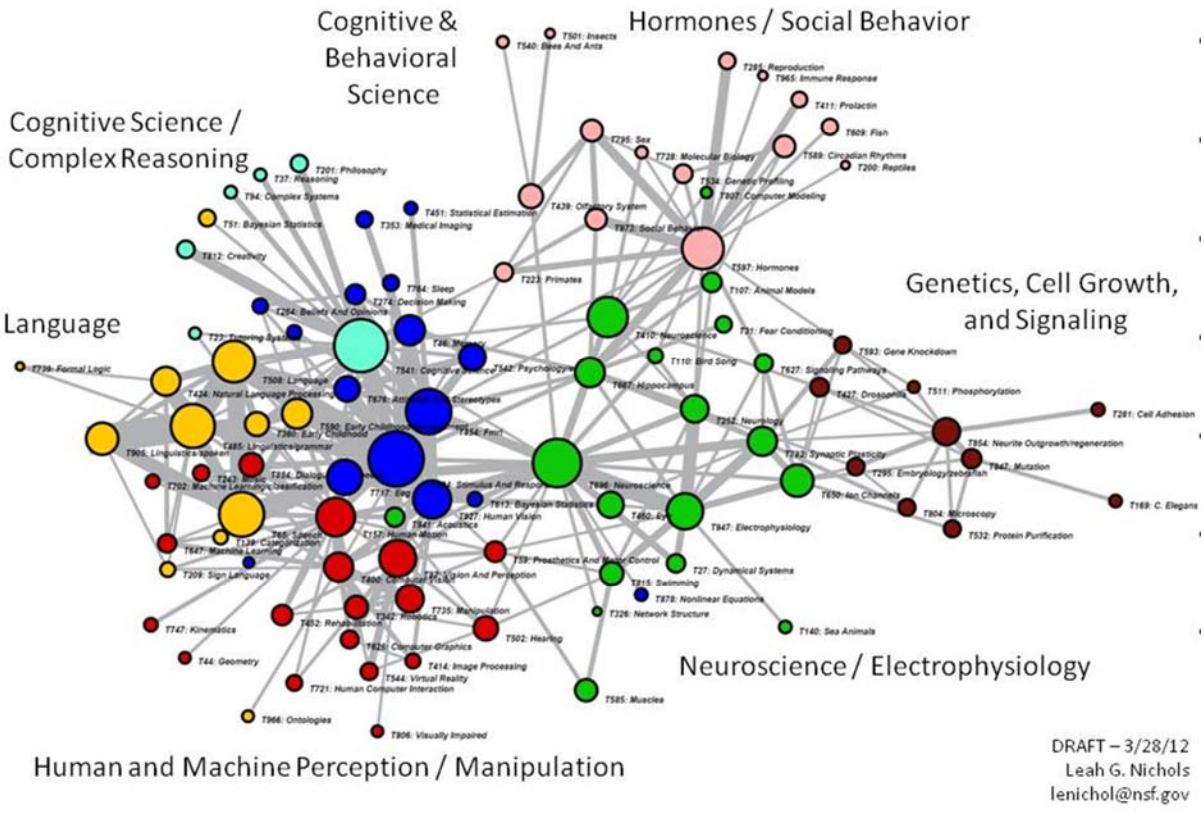


Choropleth Map



Circular Hierarchy

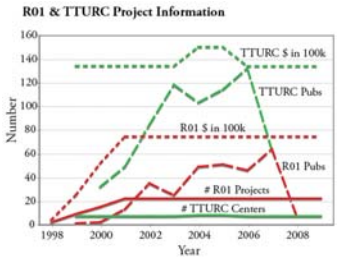
Cognitive and Neuroscience at the NSF: 2007-2011

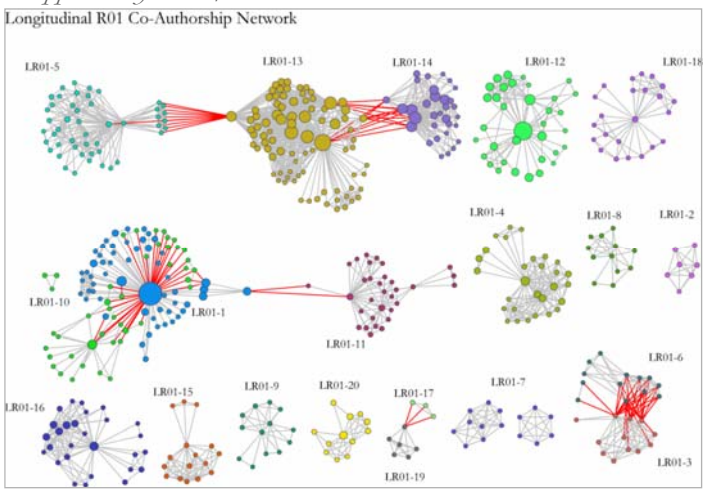


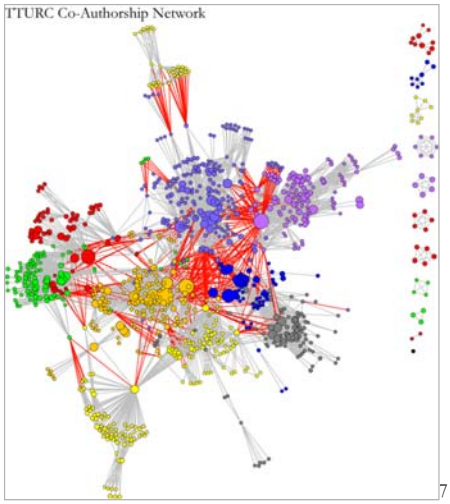
Sci² Tool Usage at National Institutes of Health

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







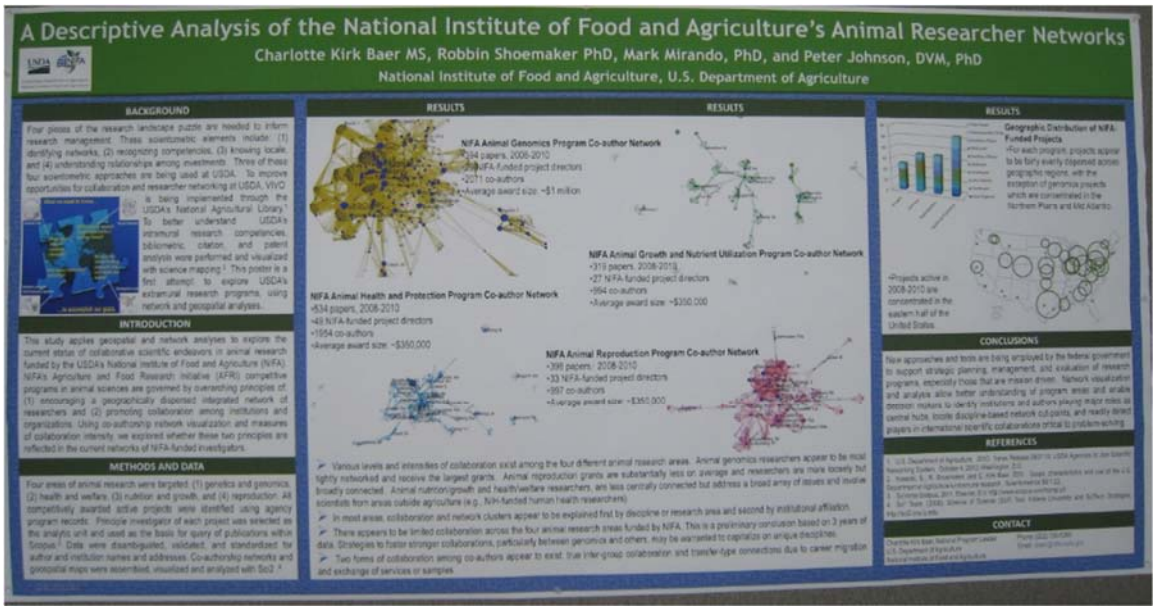
7

Sci² Tool Usage at National Institutes of Health

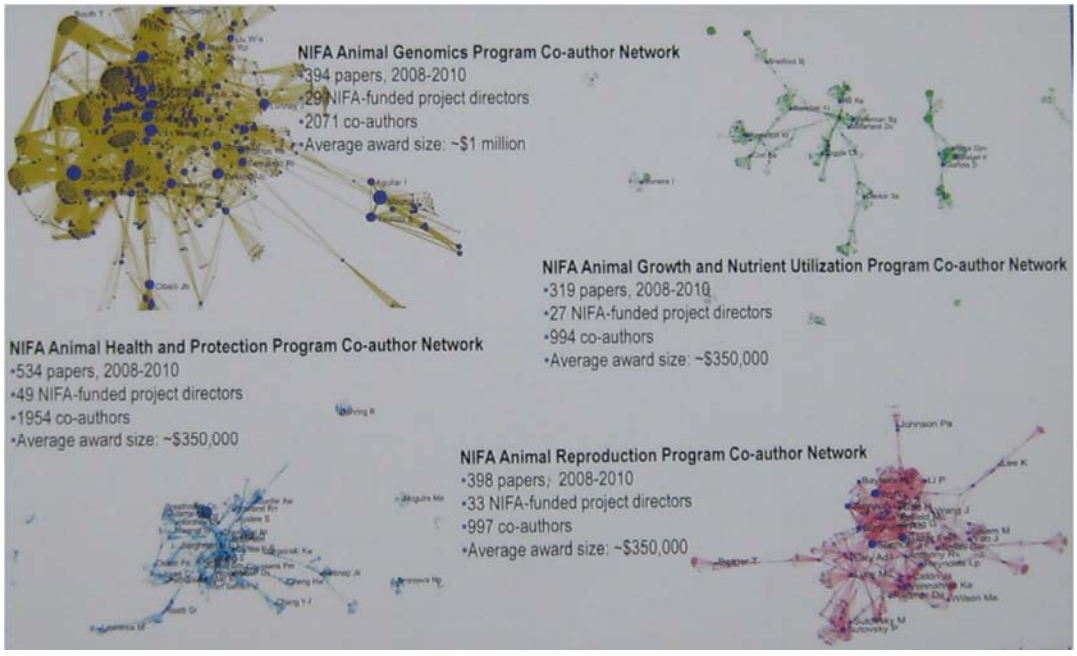
Sci² Tool now supports Web services and serves as a visual interface to publically available NIH RePORT Expenditure and Results. RePORTER/ RePORTER data provided by NIH.

Sci² Tool Usage at US Department of Agriculture

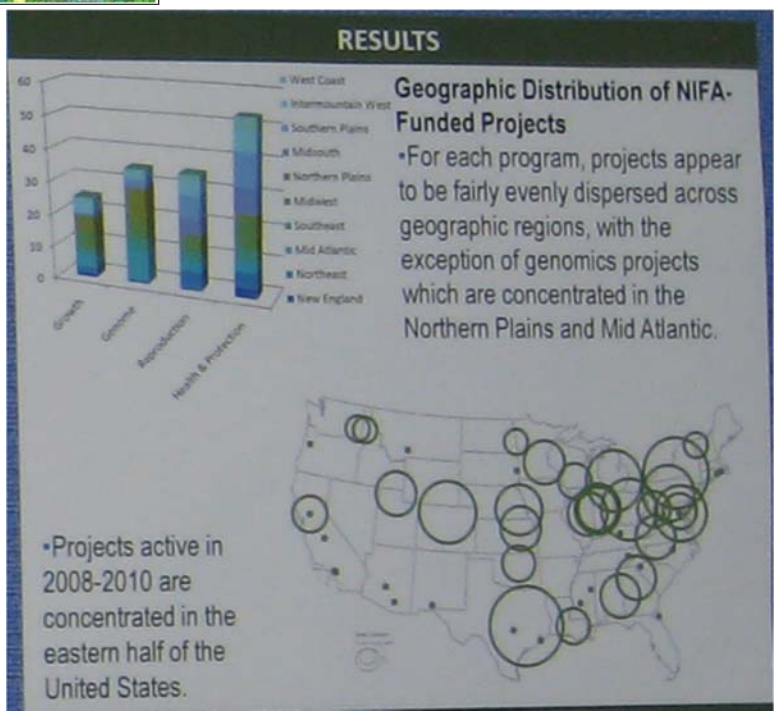
First time portrait of intramural research conducted by the U.S. Department of Agriculture (USDA) presented at the VIVO Conference 2012.



Sci² Tool Usage at US Department of Agriculture



Sci² Tool Usage at US Department of Agriculture



Sci² Tool Usage at James S. McDonnell Foundation

How did cognitive neuroscience of attention emerge from neurobiology and psychology, 1980–2005? Author co-citation analysis and 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.

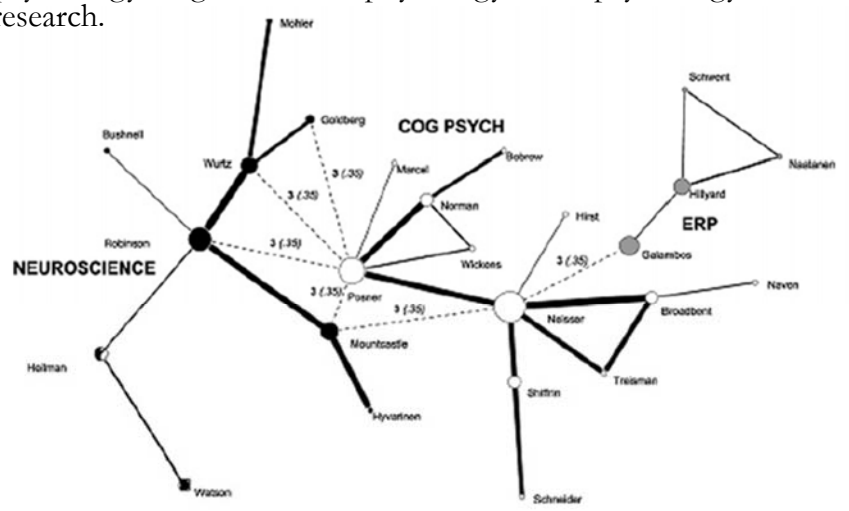


Fig. 1 In the 1980 net, neuroscience (*black* nodes and *black–white* nodes) and cognitive psychology (*white* nodes) develop as clusters with high internal co-citation rates. ERP (*grey* nodes) develops later in net construction. These clusters are connected by secondary edges at very low levels of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

Sci² Tool Usage at James S. McDonnell Foundation

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

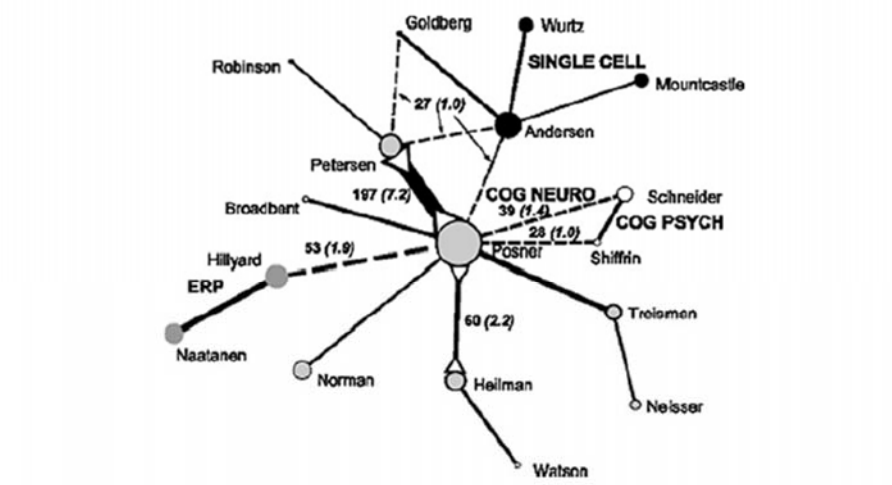


Fig. 5 The strongest link in the 1995 net is a primary edge linking Posner and Petersen. ERP and single cell neurophysiology are linked to cognitive neuroscience cluster by secondary edges

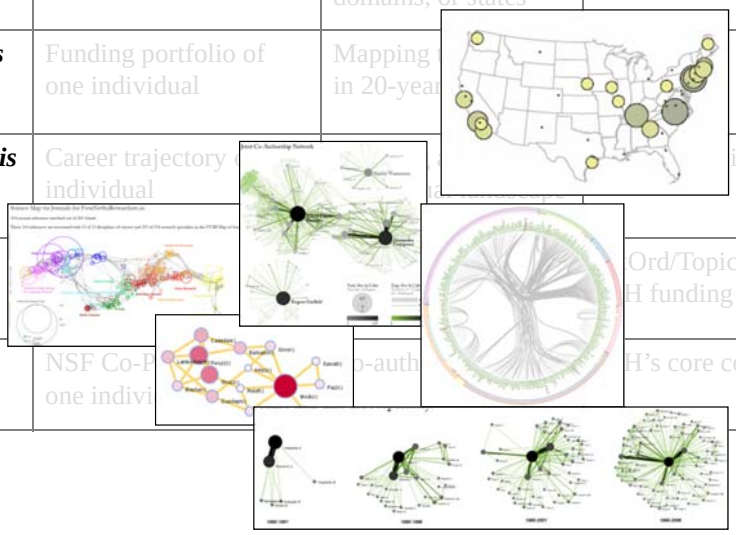
Bruer, John T. (2010). Can we talk? How the cognitive neuroscience of attention emerged from neurobiology and psychology, 1980.2005. Scientometrics, 83(3), 751-764. http://ivl.cns.iu.edu/km/tools/2010-bruer-scientometrics.pdf

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 NS… SA, all of sci…
Temporal Analysis (When)	Funding portfolio of one individual	…ic bursts …f PNAS	113 Years of P… Research
Geospatial Analysis (Where)	Career trajectory of one individual	Mapping a … intellectual l…	PNAS …
Topical Analysis (What)	…s.	…research	VxOrd/Topic r… NIH funding
Network Analysis (With Whom?)	NSF …twork of one…	C…	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	…	…iations
Topical Analysis (What)			…Ord/Topic maps of …H funding
Network Analysis (With Whom?)	NSF Co-P… one indivi…	…o-auth…	…H's core competency



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

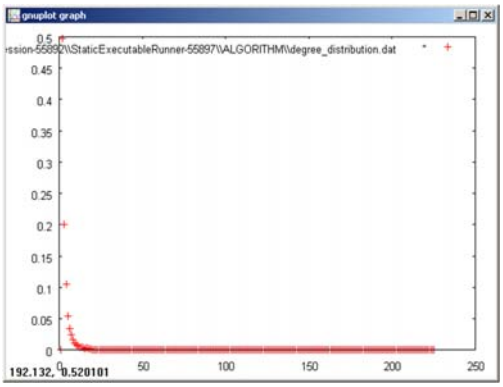
- GraphML (*.xml or *.graphml)
- Pajek .MAT (*.mat)
- Pajek .NET (*.net)
- NWB (*.nwb)
- XGMML (*.xml)
- CSV (*.csv)

Image Formats

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

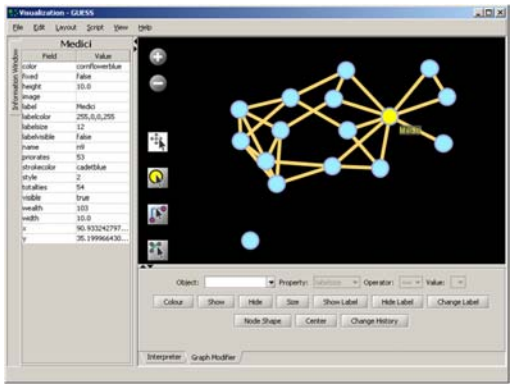
Formats are documented at http://sci2.wiki.cns.iu.edu/display/SCI2TUTORIAL/2.3+Data+Formats.

Sci² Tool – Supported Tools



Gnuplot

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



GUESS

exploratory data analysis and visualization tool for graphs and networks.

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

Sci² Tool – Supported Tools

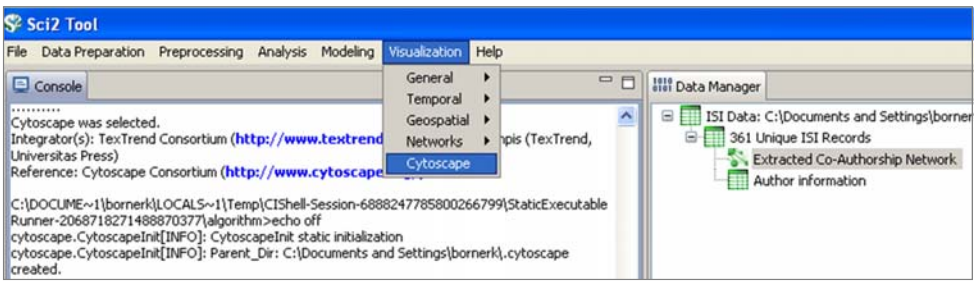
Adding more layout algorithms and network visualization interactivity

via Cytoscape http://www.cytoscape.org.

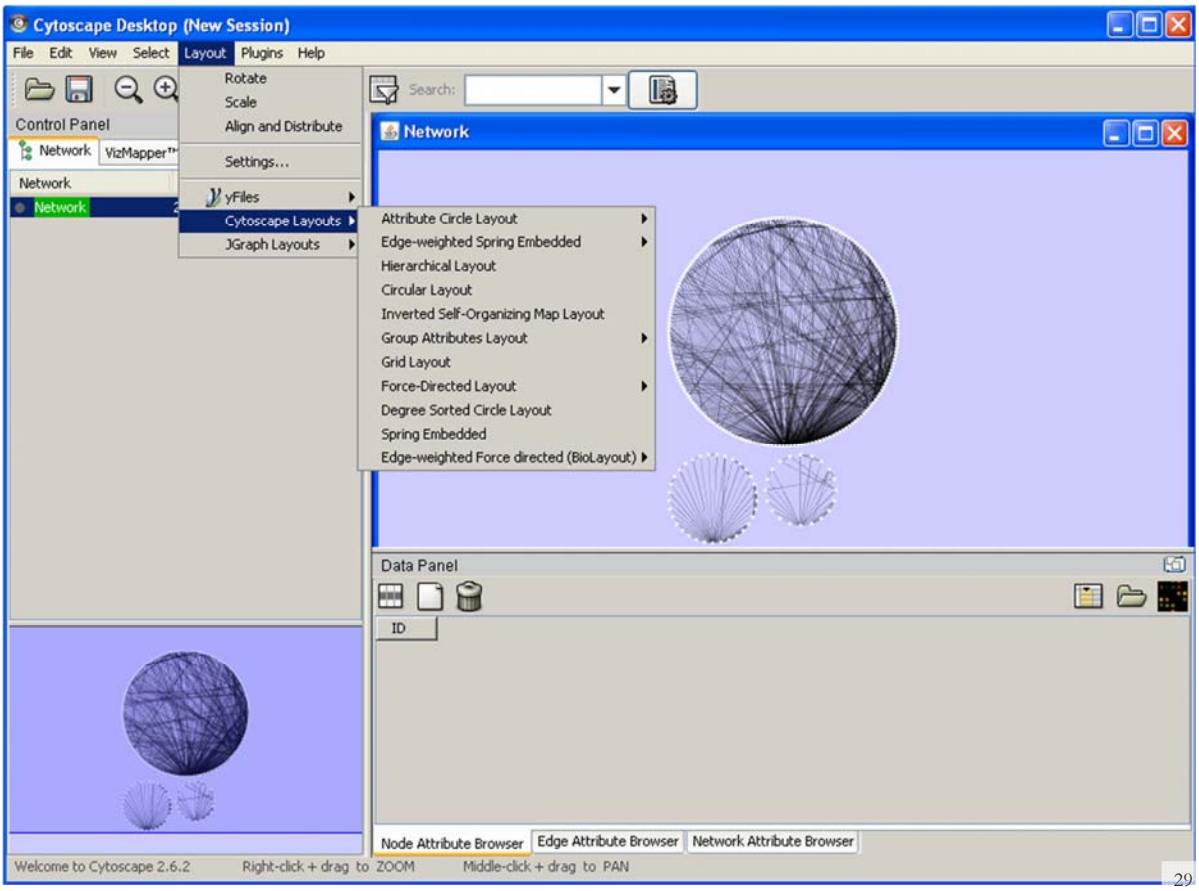
Simply add *org.textrend.visualization.cytoscape_0.0.3.jar* into your /plugin directory.

Restart Sci² Tool

Cytoscape now shows in the Visualization Menu



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



29

Tutorial Overview

- Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools
- **Sci² Tool Basics**
  - **Download and run the Sci² Tool**
  - Walkthrough: Load, analyze, and visualize a network
  - Walkthrough: Analyzing the publications of four prominent network science researchers
    - Load and clean a dataset; extract networks from raw data
    - Calculate basic statistics and analyses of the network
    - Visualize the results
- Sci² Tool – Advanced Topics
- Outlook and Q&A
- Adjourn

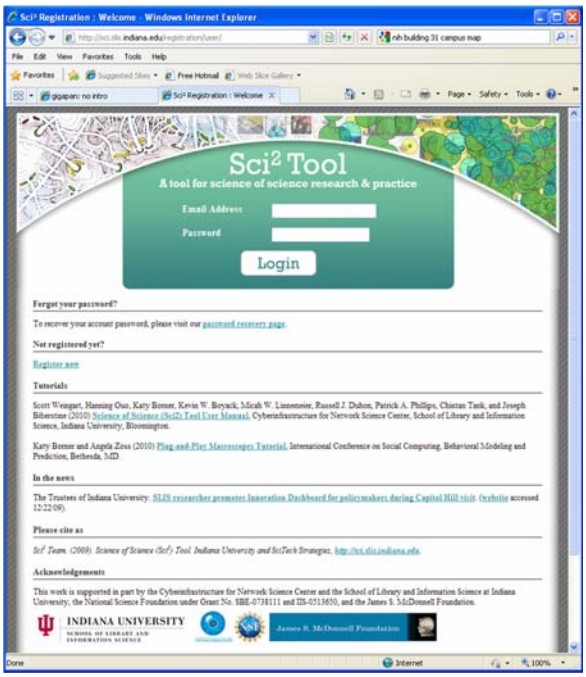
Sci²: Download, Install, and Run

Sci² v1.0 Alpha

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

Sci²: Download, Install, and Run

Sci² v1.0 alpha

- Supports ASCII UTF-8 characters
- Web-based Yahoo! and offline geocoders
- New visualizations for Temporal, Topical, Geographical, and Bipartite data
- Customizable stop word lists
- New home page, wiki-based tutorial
- Reader for Google Scholar
- Gephi and R support
- Bug fixes, streamlined workflows

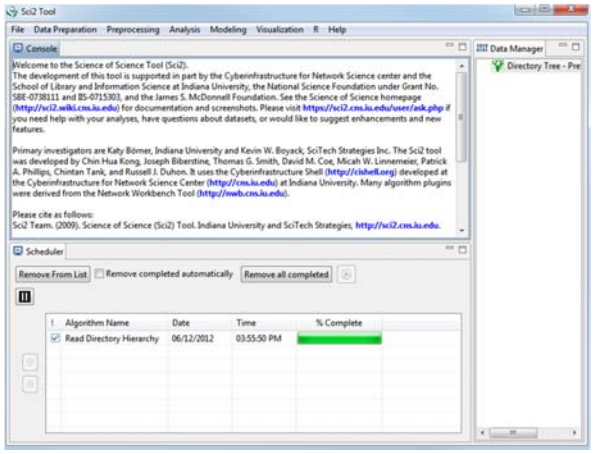
Sci² runs on Windows, Mac, and Linux.
Decompress the archive and run **`sci2.exe`**

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

- Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools
- Sci² Tool Basics
  - Download and run the Sci² Tool
  - **Walkthrough: Load, analyze, and visualize a network**
  - Walkthrough: Analyzing the publications of four prominent network science researchers
    - Load and clean a dataset; extract networks from raw data
    - Calculate basic statistics and analyses of the network
    - Visualize the results
- Sci² Tool – Advanced Topics
- Outlook and Q&A
- Adjourn

Padgett's Floren
visualize family

Florentine families related through busi
ties such as loans, credits and joint part

Node attributes

- Wealth: Each family's net wealth in
- Priorates: The number of seats on t
- Totalities: Number of business/mar
  families.

Edge attributes:

- Marriage T/F
- Business T/F

"Substantively, the data include familie
political control of the city of Florence
dominant in this struggle: one revolved
around the powerful Strozzis."

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

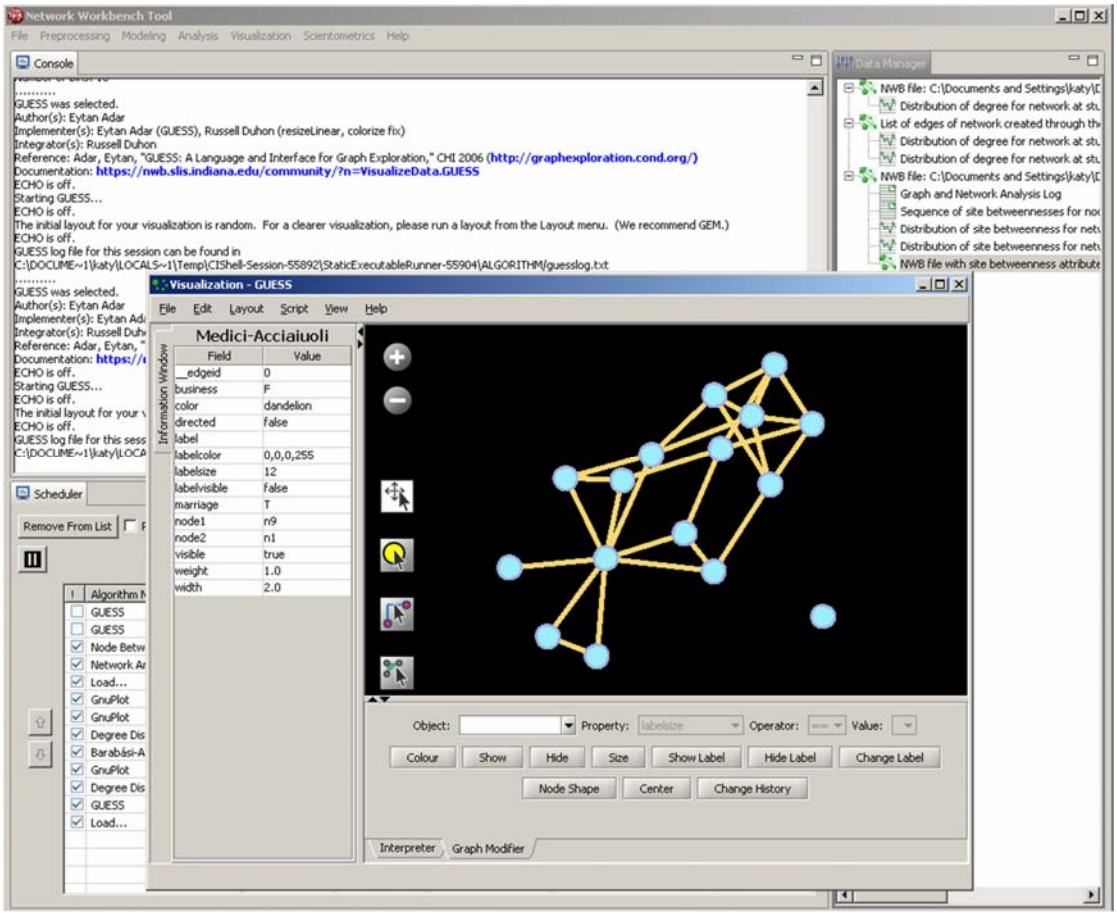
```
*Nodes
id*int label*string wealth*int
totalities*int priorates*int
1 "Acciaiuoli" 10 2 53
2 "Albizzi" 36 3 65
3 "Barbadori" 55 14 0
4 "Bischeri" 44 9 12
5 "Castellani" 20 18 22
6 "Ginori" 32 9 0
7 "Guadagni" 8 14 21
8 "Lamberteschi" 42 14 0
9 "Medici" 103 54 53
10 "Pazzi" 48 7 0
11 "Peruzzi" 49 32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

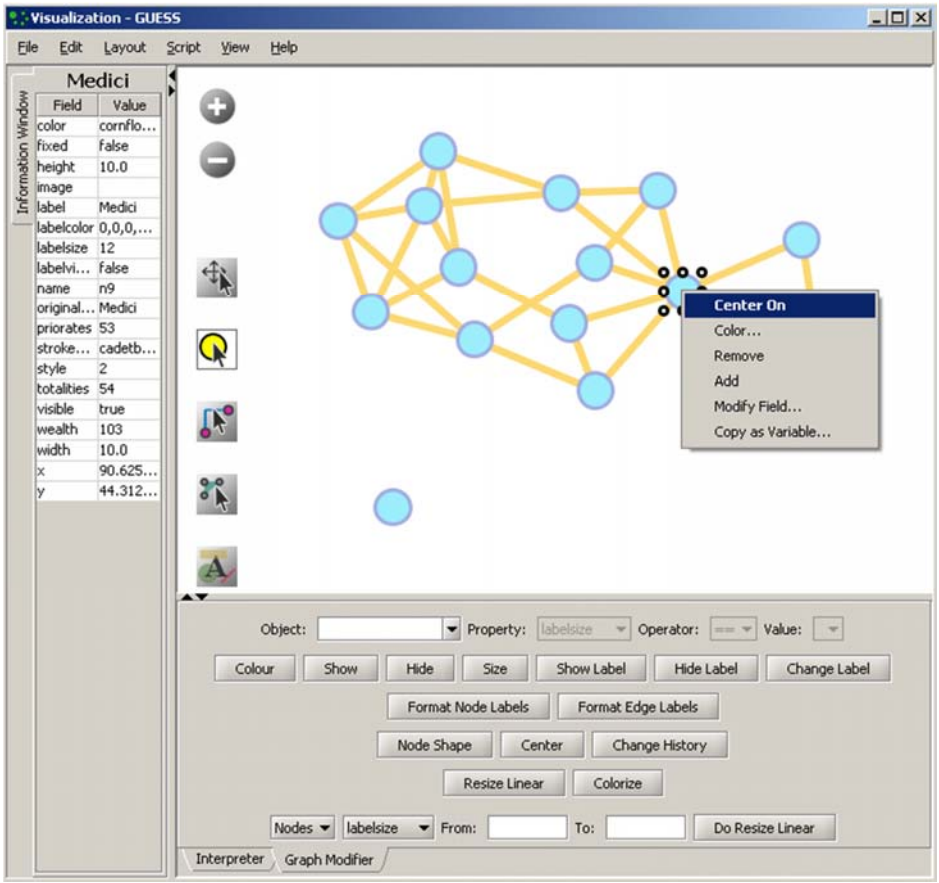

Padgett's Florentine Families – Load, compute basic network properties & view in GUESS

- Load ***yoursci2directory***`/sampledata/socialscience/florentine.nwb`
- Run `Analysis > Networks > 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
    Did not detect any numeric attributes.
    This network does not seem to be a valued network.
Average degree: 3.375
This graph is not weakly connected.
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 > Networks > GUESS` to open GUESS with file loaded
- Apply `Layout > GEM`





Pan:

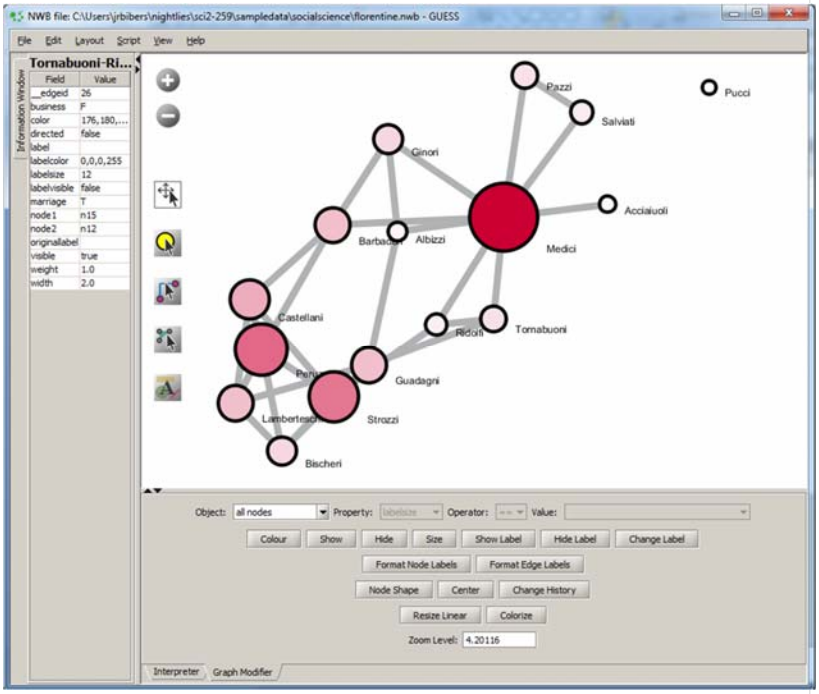
"grab" the background by holding left-click and moving your mouse.

Zoom:

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

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

Right click to modify Color, etc.



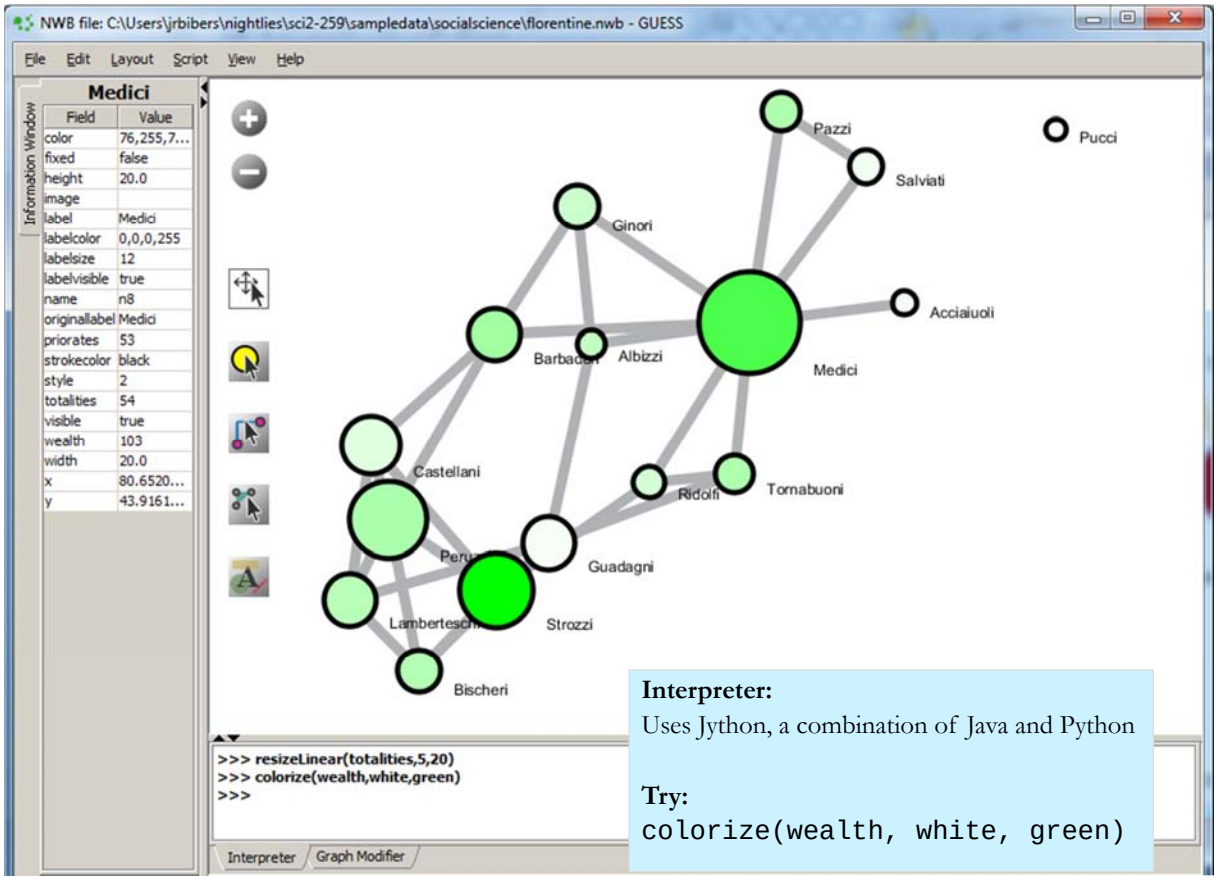
Graph Modifier:

Select `all nodes` in the Object drop-down menu and click `Show Label` button.

Select `Resize Linear > Nodes > totalities` drop-down menu, then type `5` and `20` into the `From` and `To` value box separately. Then select `Do Resize Linear`.

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

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



Interpreter:

Uses Jython, a combination of Java and Python

Try:

`colorize(wealth, white, green)`

Tutorial Overview

- Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools
- Sci² Tool Basics
  - Download and run the Sci² Tool
  - Walkthrough: Load, analyze, and visualize a network
  - **Walkthrough: Analyzing the publications of four prominent network science researchers**
    - **Load and clean a dataset; extract networks from raw data**
    - **Calculate basic statistics and analyses of the network**
    - **Visualize the results**
- Sci² Tool – Advanced Topics
- Outlook and Q&A
- Adjourn

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)

Data Acquisition from Web of Science

In December 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

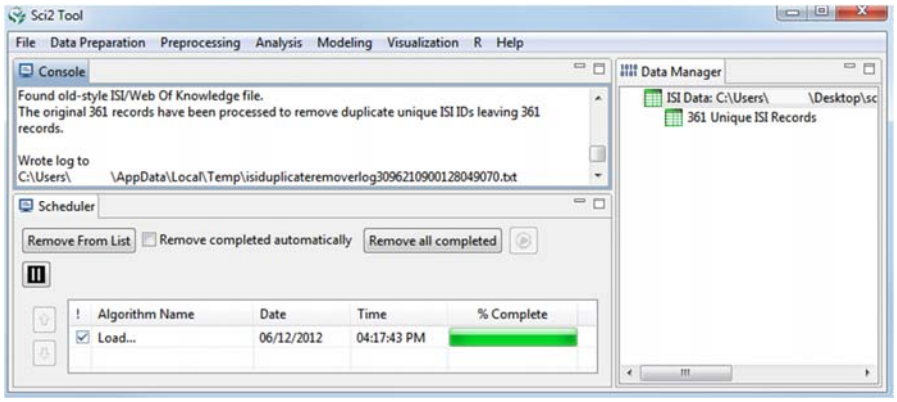
Comparison of Counts

No books and other non-WoS publications are covered.

Researcher		Age	Total # Cites	Total # Papers	H-Index
Eugene Garfield		82	1525	672	31
Stanley Wasserman			122	35	17
Alessandro Vespignani		42	451	101	33
Albert-László Barabási	(Dec 2007)	40	2218	126	47
	(Dec 2008)	41	16920	159	52
	(April 2011)	44	30102	201	68

Extract Co-Author Network

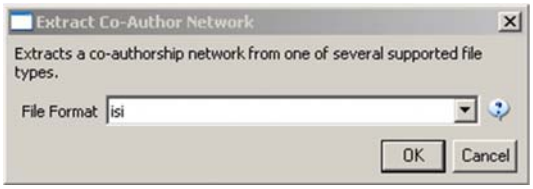
Load
YourSci2Directory`/sampledata/scientometrics/isi/FourNetSciResearchers.isi`
using `File > Load…`
A table of 361 unique records will be loaded into the Data Manager
Duplicates are removed and a log file is created



Extract Co-Author Network

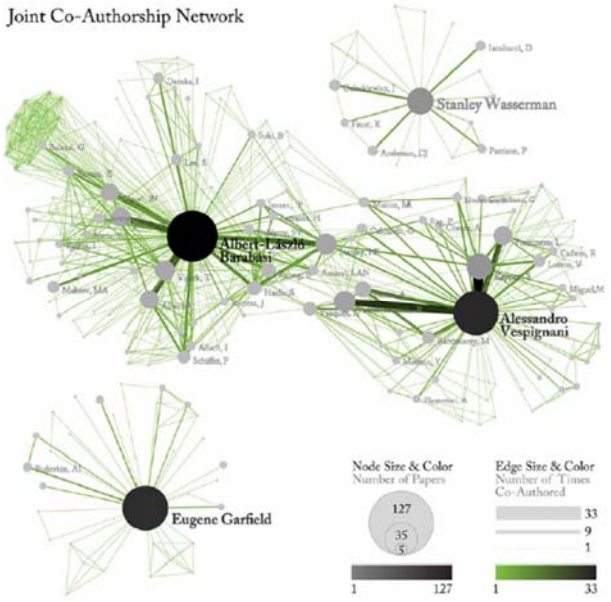
(see section 5.1.4.2 on correcting duplicate/misspelled author names)

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



- The result is an undirected but weighted network of co-authors in the Data Manager.
- Run ***`Analysis > Networks > Network Analysis Toolkit (NAT)`*** to calculate basic properties: the network has 247 nodes and 891 edges.
- Use ***`Analysis > Networks > 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`***

Co-Author Network of all Four NetSci Researchers

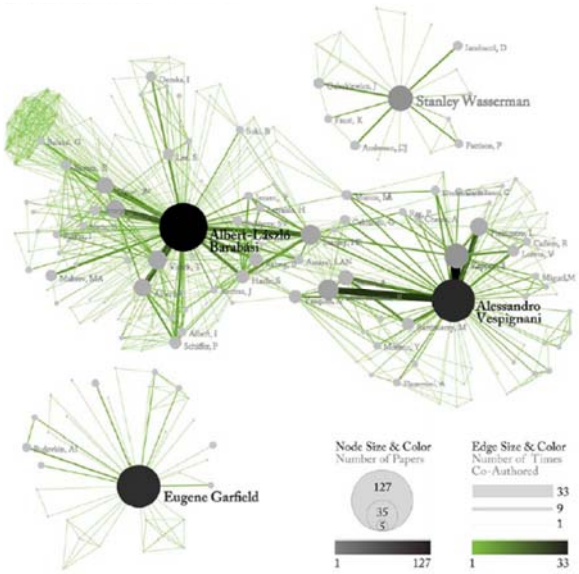


Co-Author Network of all Four NetSci Researchers

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

Calculate node degrees in Sci^2

An image editor can be used to add legends



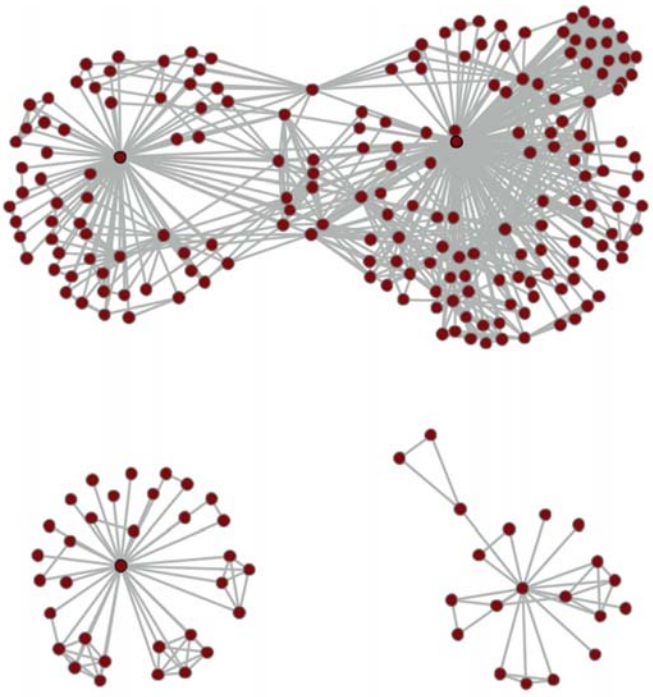
Network Visualization: Node Layout

Load... was selected.
Found old-style ISI/Web Of Knowledge file.
The original 361 records have been processed to remove duplicate unique ISI IDs leaving 361 records.

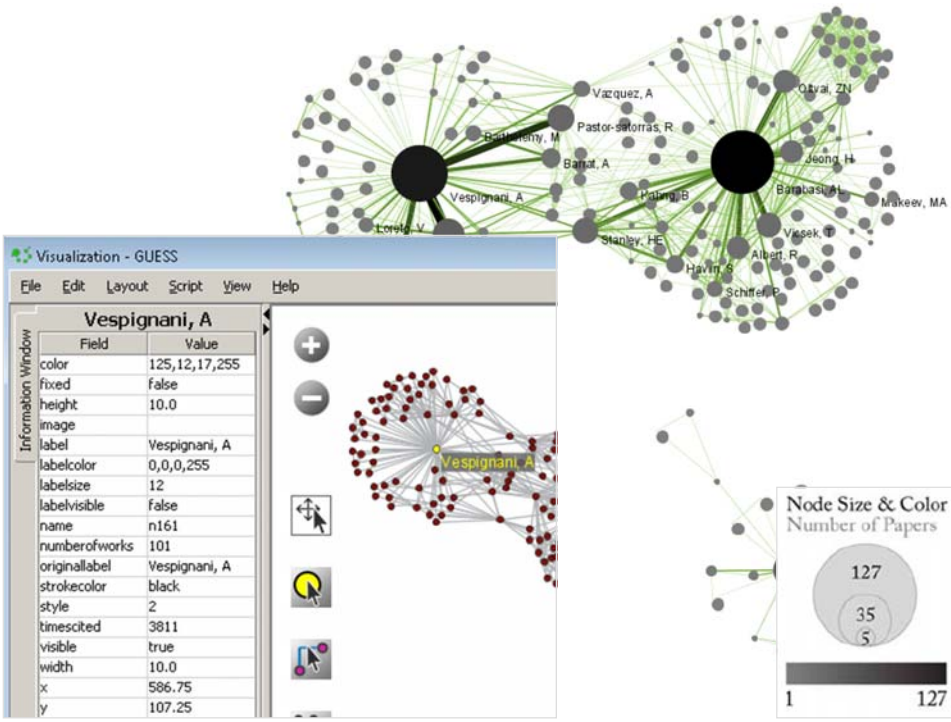
.....
Extract Co-Author Network was selected.
Input Parameters:
File Format: isi

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

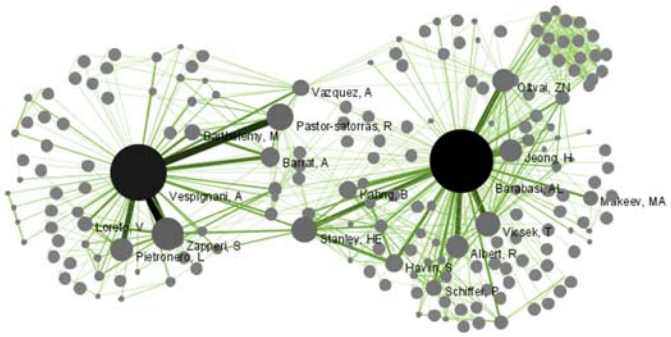
.....
GUESS was selected.



Network Visualization: Color/Size Coding by Data Attribute Values



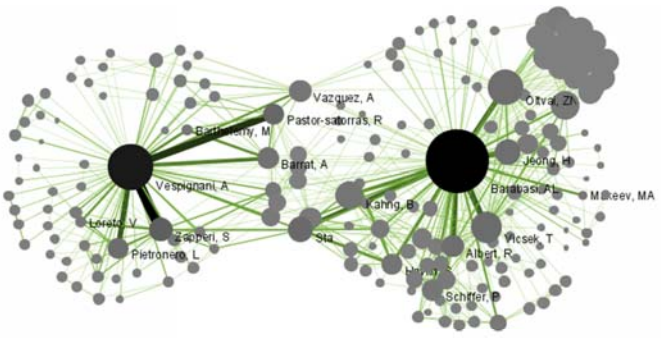
Network Visualization: Giant Component



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

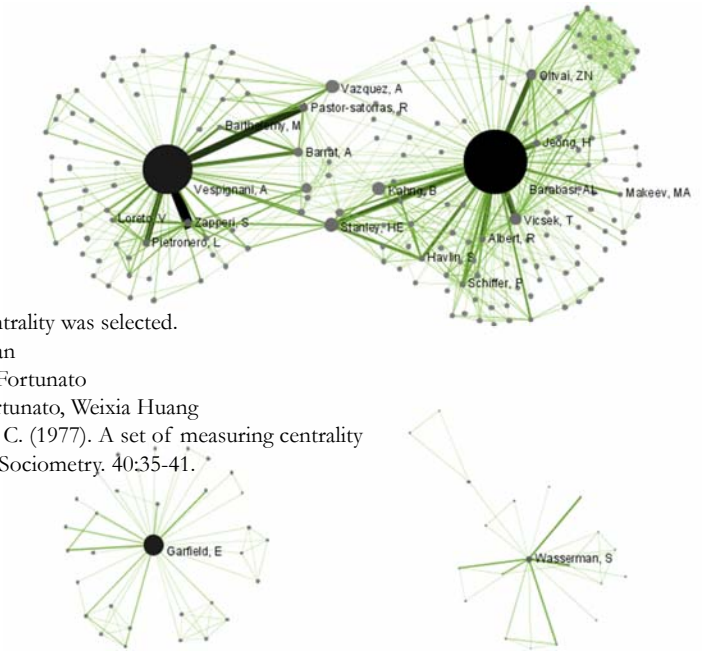
Input Parameters:
Number of top clusters: 10
3 clusters found, generating graphs for the top 3 clusters.
.....

Network Visualization: Color/Size Coding by Degree



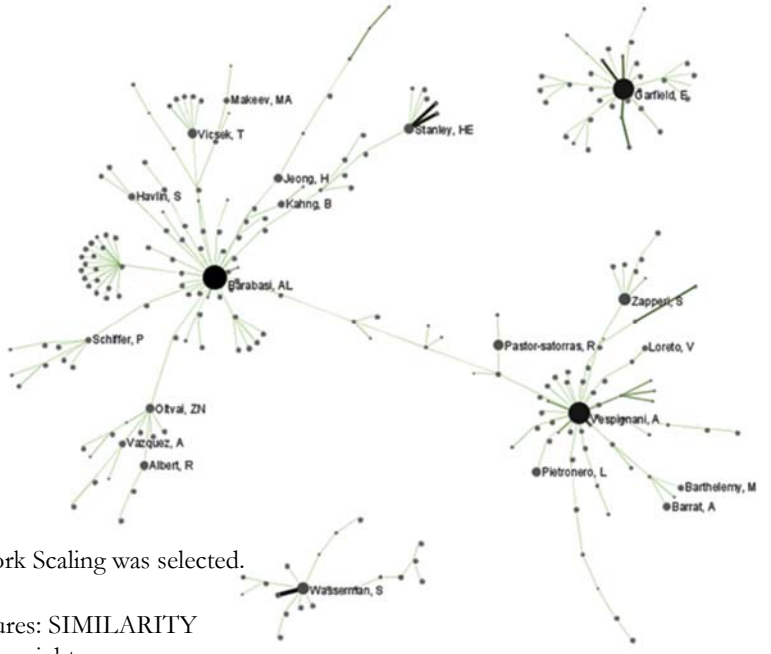
.....
Node Degree was selected.
Documentation:
https://nwb.slis.indiana.edu/community/?n=AnalyzeData.NodeDegree
.....

Network Visualization:
Color/Size Coding by Betweenness Centrality



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

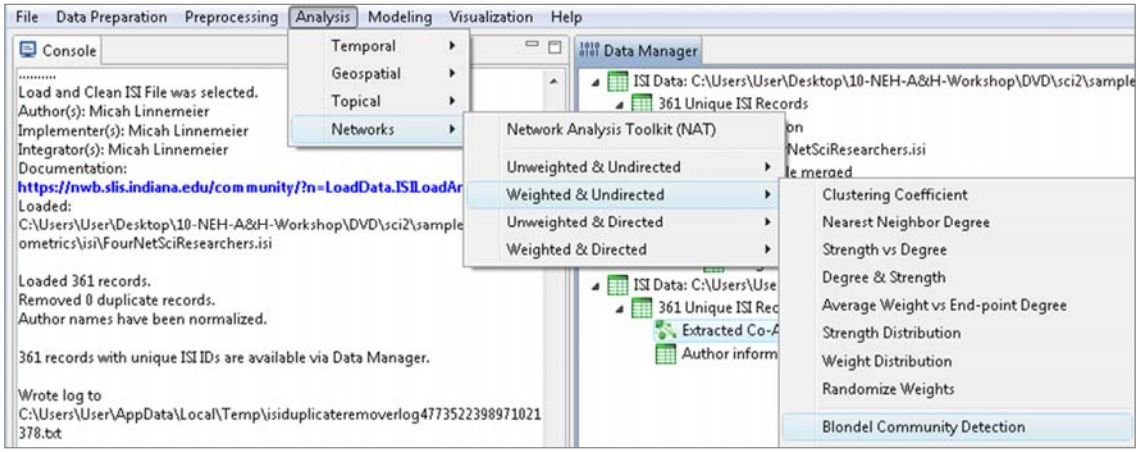
Network Visualization:
Reduced Network After Pathfinder Network Scaling



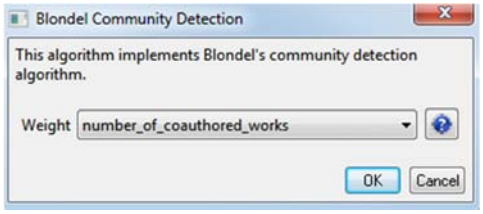
.....
MST-Pathfinder Network Scaling was selected.
Input Parameters:
Weight Attribute measures: SIMILARITY
Edge Weight Attribute: weight
.....

Network Visualization: Circular Hierarchy Visualization

Select Co-Author Network and run Blondel Community detection:

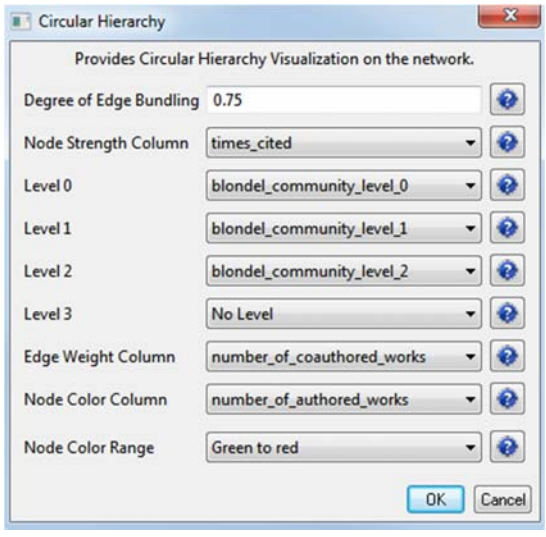


With parameter values



Network Visualization: Circular Hierarchy Visualization

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



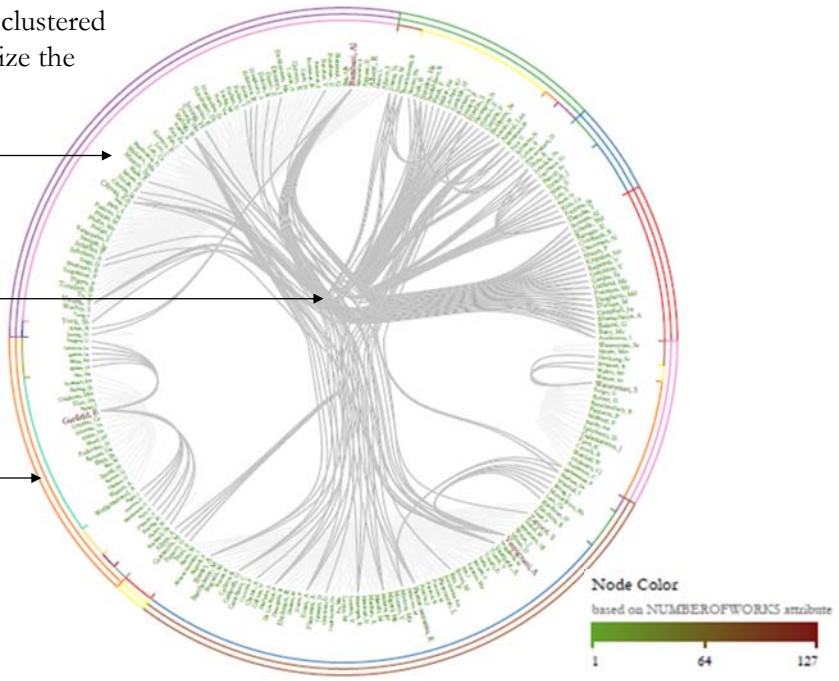
Network Visualization: Circular Hierarchy Visualization

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

Node labels, e.g., author names.

Network structure using edge bundling.

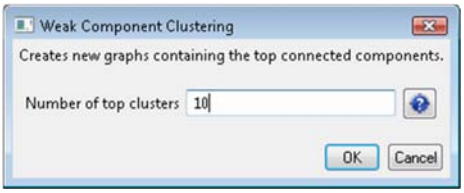
Color coded cluster hierarchy according to Blondel community detection algorithm.

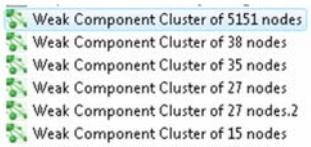


Note:
Header/footer info, legend, and more meaningful color coding are under development.

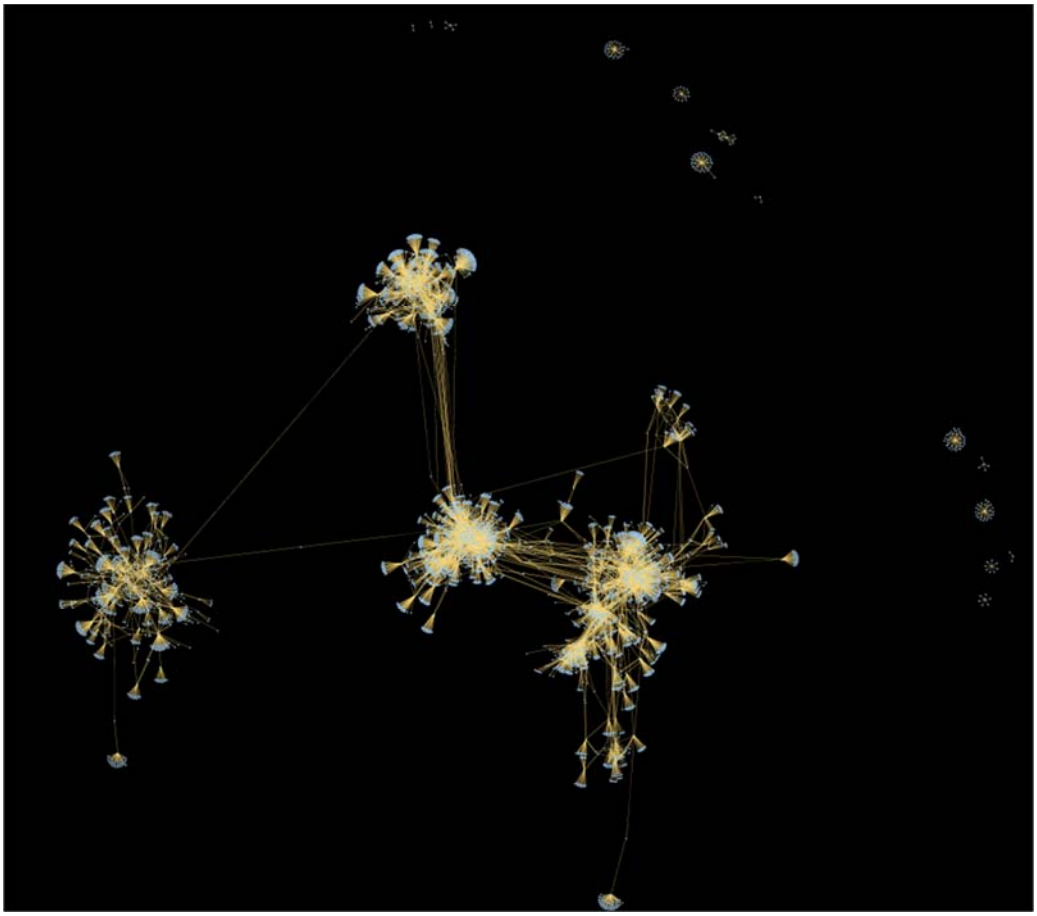
Paper-Citation Network Layout

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





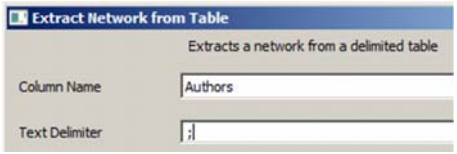
to identify top-10 largest components. The largest (giant) component has 5,151 nodes.

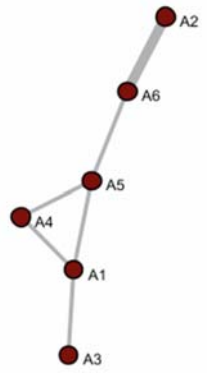


General Network Extraction: Weighted, Undirected Co-Occurrence Network

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

Author co-occurrence network



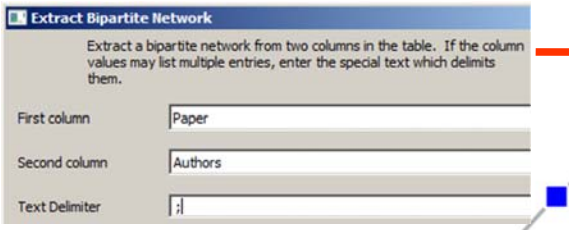


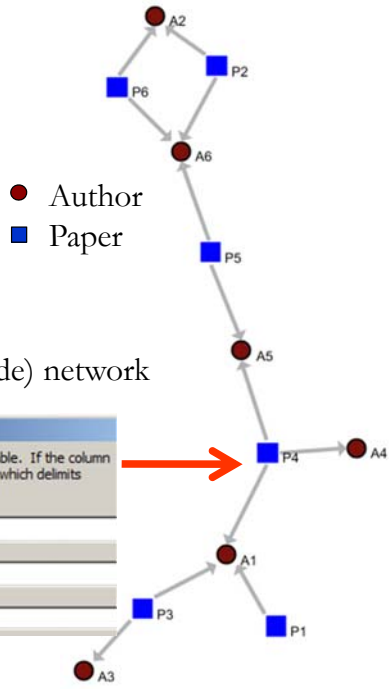

```
*Vertices 6
1 A1
2 A6
3 A2
4 A3
5 A5
6 A4
*Edges 6
2 3 2
1 4 1
1 5 1
5 6 1
1 6 1
2 5 1
```

General Network Extraction: Unweighted, Directed Bipartite Network

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

Paper-author bipartite (2-mode) network

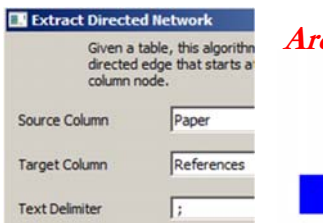




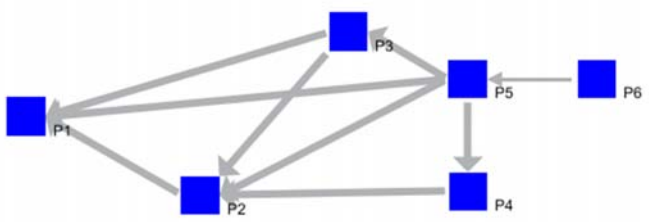

```
*Vertices 12
1 P1 bipartitetype "Paper"
2 A1 bipartitetype "Authors"
3 P2 bipartitetype "Paper"
4 A2 bipartitetype "Authors"
5 A6 bipartitetype "Authors"
6 P3 bipartitetype "Paper"
7 A3 bipartitetype "Authors"
8 P4 bipartitetype "Paper"
9 A4 bipartitetype "Authors"
10 A5 bipartitetype "Authors"
11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

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



Arcs from papers to references

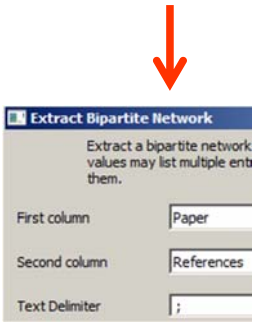


1970 1980 1990 1995 2000

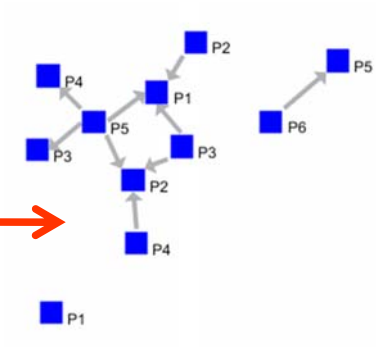
```
*Vertices 6
1 P1
2 P2
3 P3
4 P4
5 P5
6 P6
*Arcs
2 1
3 1
3 2
4 2
5 4
5 3
5 1
5 2
6 5
```

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



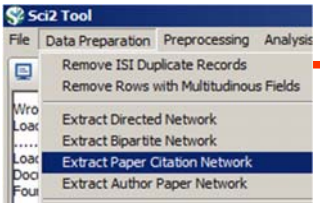
Mistake!



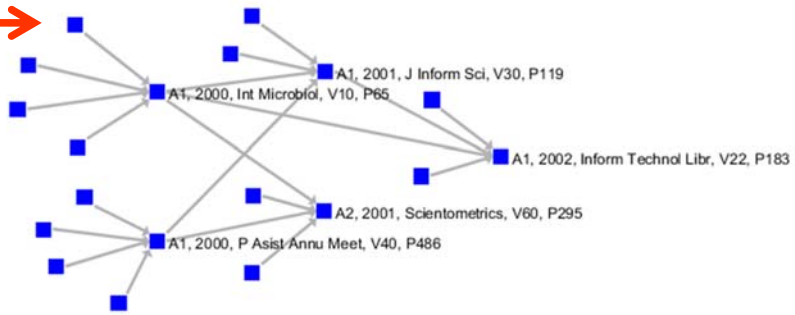

```
*Vertices 11
1 P1 bipartitetype "Paper"
2 P2 bipartitetype "Paper"
3 P1 bipartitetype "References"
4 P3 bipartitetype "Paper"
5 P2 bipartitetype "References"
6 P4 bipartitetype "Paper"
7 P5 bipartitetype "Paper"
8 P4 bipartitetype "References"
9 P3 bipartitetype "References"
10 P6 bipartitetype "Paper"
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|BROI	2000	T1	A1, 2000, INT MICROBIOL, V10, P65
3	A1	BENSMAN SJ, 1999, LIBR RESOUR TECH SER, V42, P147|BROI	2000	T2	A1, 2000, P ASIST ANNU MEET, V40, P486
4	A2|A3	GARFIELD E, 1985, ESSAYS INFORMATION S, V8, P403|GILBE	2001	T3	A2, 2001, SCIENTOMETRICS, V60, P295
5	A1	ASIMOV A, 1963, GENETIC CODE|LEDERBERG J, 1972, NATUI	2001	T4	A1, 2001, J INFORM SCI, V30, P119
6	A1|A2	AVERY OT, 1944, J EXP MED, V79, P137|SMALL H, 1985, J INF	2002	T5	A1, 2002, INFORM TECHNOL LIBR, V22, P183



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



2000 *2001* *2002*

Tutorial Overview

- Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools
- Sci² Tool Basics
- **Sci² Tool – Advanced Topics**
  - Walkthrough: Visualizing temporal data for NSF projects
  - Walkthrough: Locating data on a geographic map
  - Walkthrough: Examining an evolving network
  - Interacting with the statistical toolkit R and the network visualization package Gephi
  - Sci² tool visualizations
    - Bipartite networks
    - Map of Science
- Outlook and Q&A
- Adjourn

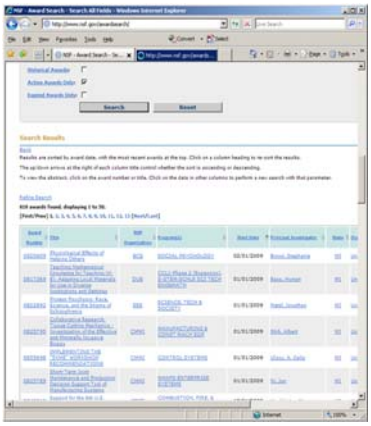
Tutorial Overview

- Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools
- Sci² Tool Basics
- **Sci² Tool – Advanced Topics**
  - **Walkthrough: Visualizing temporal data for NSF projects**
  - Walkthrough: Locating data on a geographic map
  - Walkthrough: Examining an evolving network
  - Interacting with the statistical toolkit R and the network visualization package Gephi
  - Sci² tool visualizations
    - Bipartite networks
    - Map of Science
- Outlook and Q&A
- Adjourn

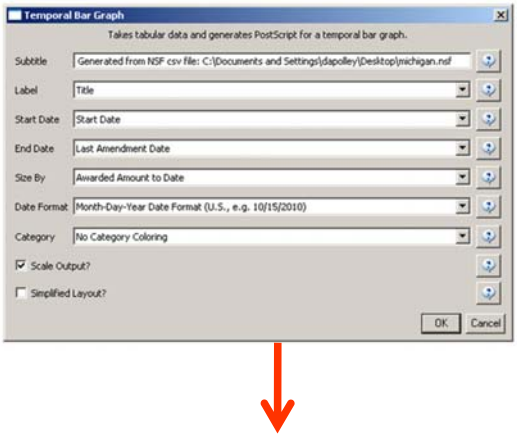
Temporal bar graph for NSF projects

See 5.2.1 Funding Profiles of Three Universities (NSF Data)

Download NSF data



Visualize as Temporal Bar Graph



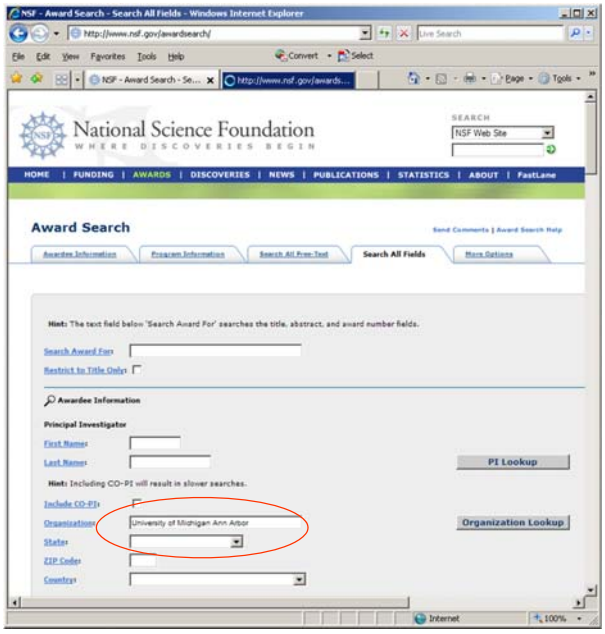
Area size equals numerical value, e.g., award amount.

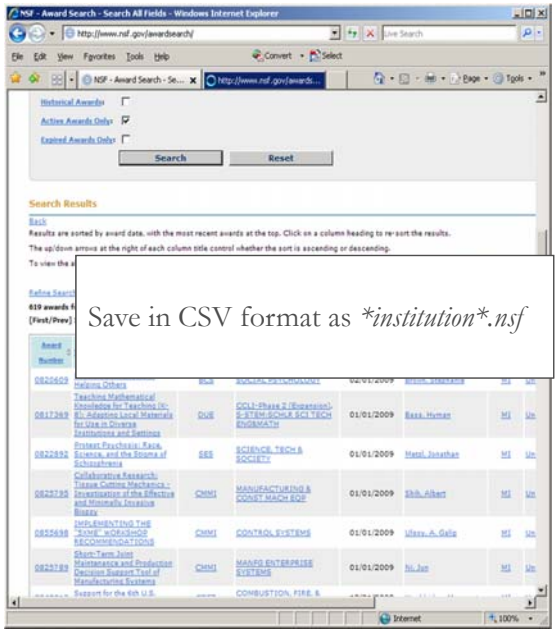
Text

Start date End date

Temporal Bar Graph of NSF projects

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



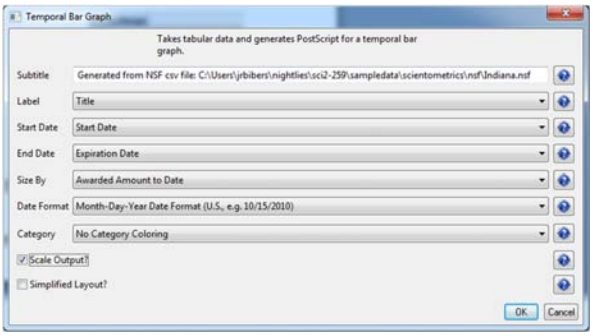


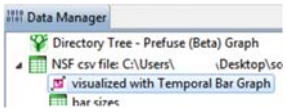
Save in CSV format as **institution*.nsf*

Temporal Bar 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/Indiana.nsf`

Run `Visualization > Temporal > Temporal Bar Graph` using parameters:



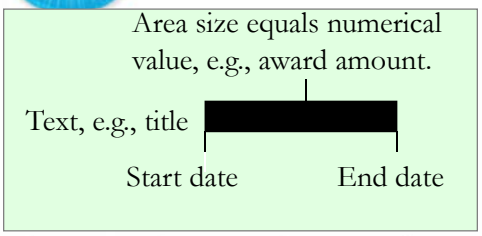


Save `visualized with Temporal Bar Graph` as PS or EPS file

Convert into PDF and view

Zoom to see details in visualizations of large datasets, e.g., all NSF awards ever made

Temporal Bar Graph of NSF projects



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)

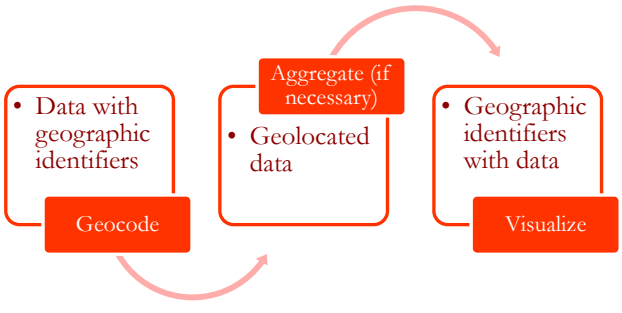
5.2.3 Biomedical Funding Profile of NSF (NSF Data)

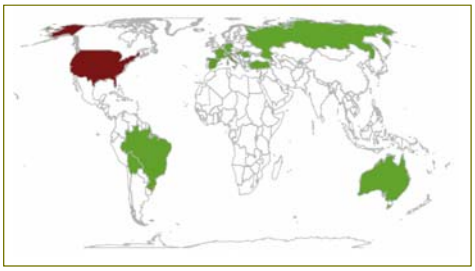
Tutorial Overview

- Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools
- Sci² Tool Basics
- **Sci² Tool – Advanced Topics**
  - Walkthrough: Visualizing temporal data for NSF projects
  - **Walkthrough: Locating data on a geographic map**
  - Walkthrough: Examining an evolving network
  - Interacting with the statistical toolkit R and the network visualization package Gephi
  - Sci² tool visualizations
    - Bipartite networks
    - Map of Science
- Outlook and Q&A
- Adjourn

Geocoding and Geospatial Maps

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





Region names + numeric data
(Choropleth Map)

Geocoordinates + numeric data
(Proportional Symbol Map)

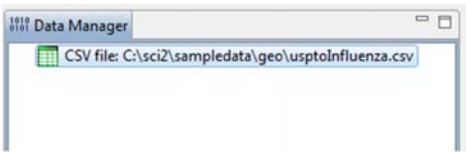
Load File with Address and Times Cited Fields

Run `File > Load…`

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

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

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



Yahoo Geocoder

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.

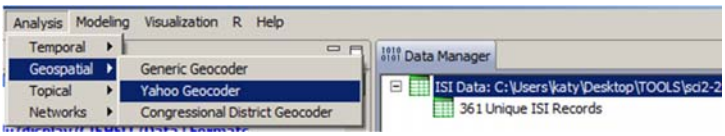
Applications

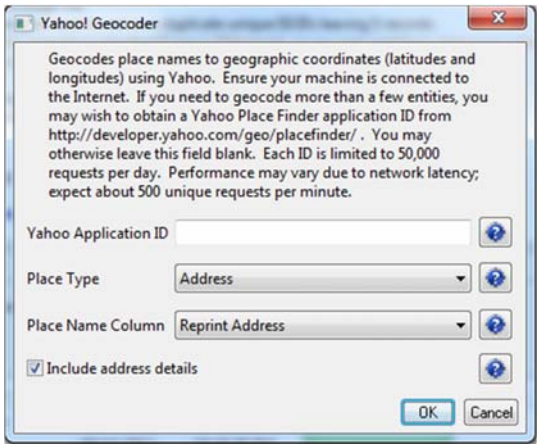
The plugin is useful for scientists who would like to visualize their data on a geographical map (geomap). User can obtain the geographical coordinates (Latitude and Longitude values) and feed them to the visualization plugin.

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

Using Yahoo! Geocoder

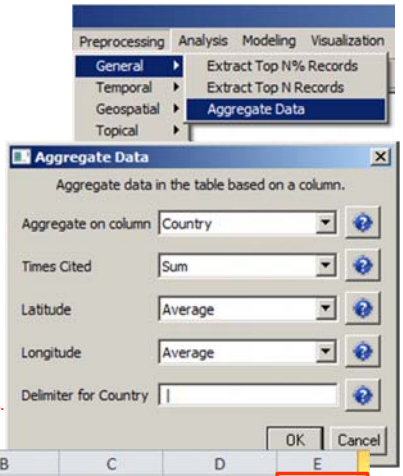
Run `Analysis > Geospatial > Yahoo Geocoder`





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

Aggregate by Country

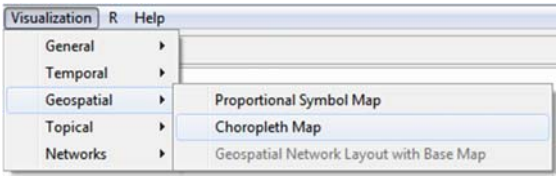


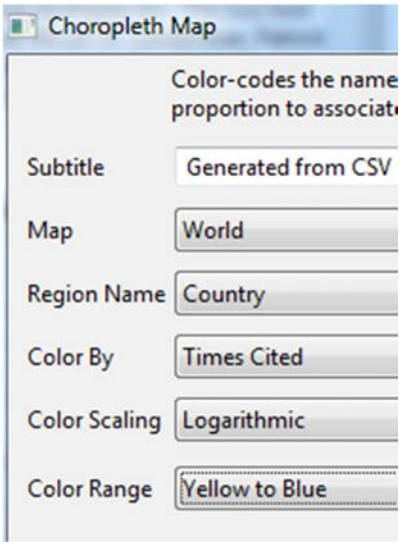
Aggregate Data was selected.
Implementer(s): Chintan Tank
Documentation: **http://wiki.cns.iu.edu/display/CISHELL/Aggregate+Data**
Input Parameters:
Aggregate on column: Country
Delimiter for Country: |
Longitude: AVERAGE
Latitude: AVERAGE
Times Cited: SUM
Aggregated by '': All rows of Latitude column were skipped due to no non-null, non-empty values.
Aggregated by '': All rows of Longitude column were skipped due to no non-null, non-empty values.
Frequency of unique "Country" values added to "Count" column.

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

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

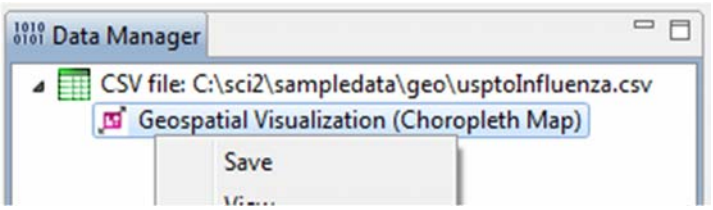
Choropleth Map





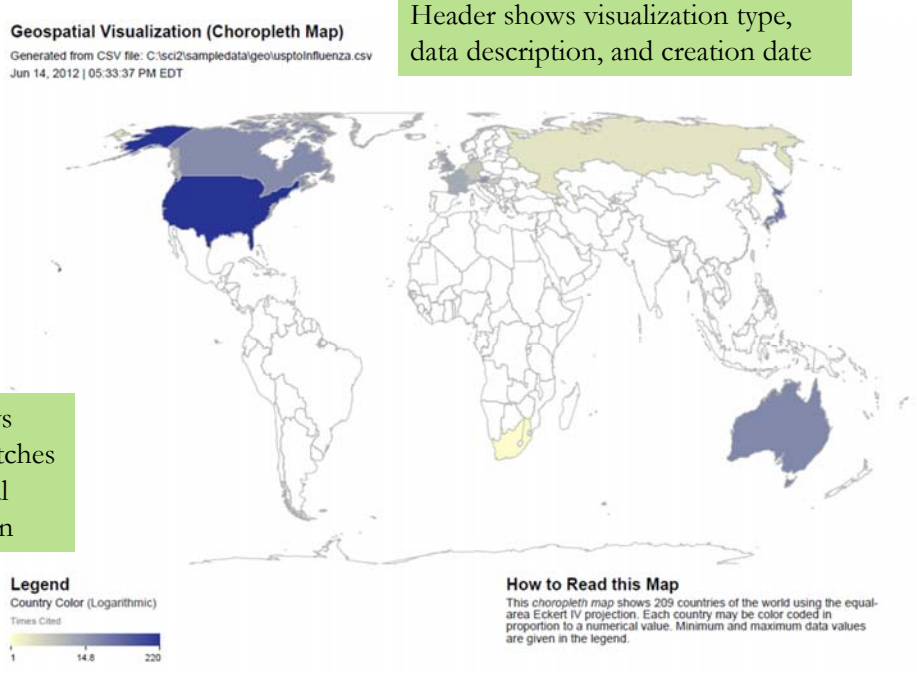
Right-click and **Save** map

Open this PostScript file to visualize



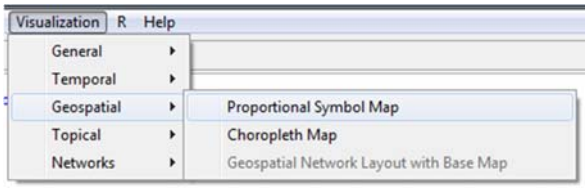
Interpreting the Choropleth Map

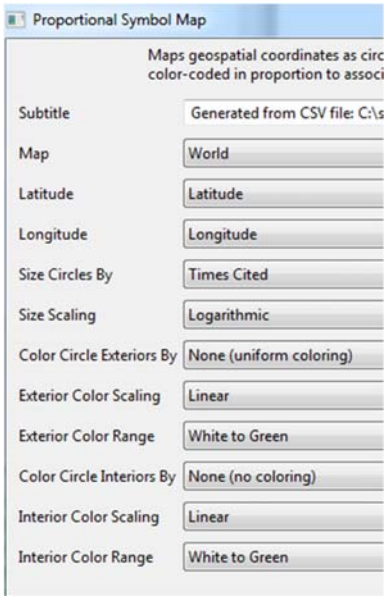
Header shows visualization type, data description, and creation date



Legend shows how data matches up with visual representation

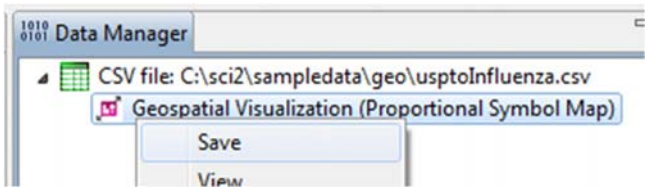
Proportional Symbol Map



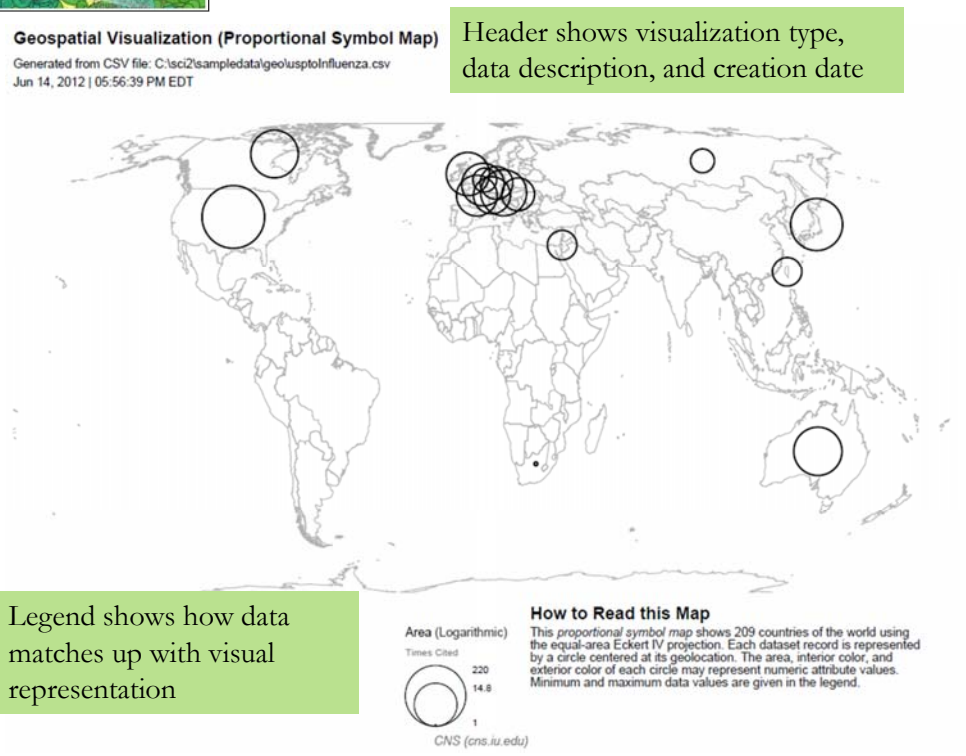


Right-click and **Save** map

Open this PostScript file to visualize



Interpreting the Proportional Symbol Map

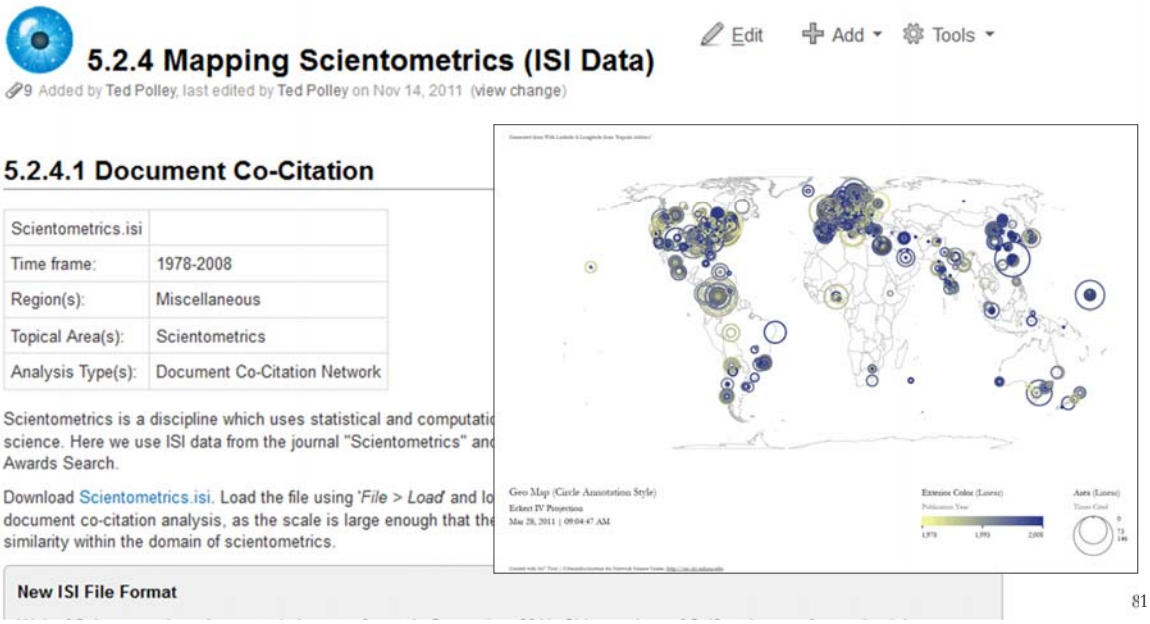


Header shows visualization type, data description, and creation date

Legend shows how data matches up with visual representation

Sci² Manual – Geospatial Workflows for Scientometrics

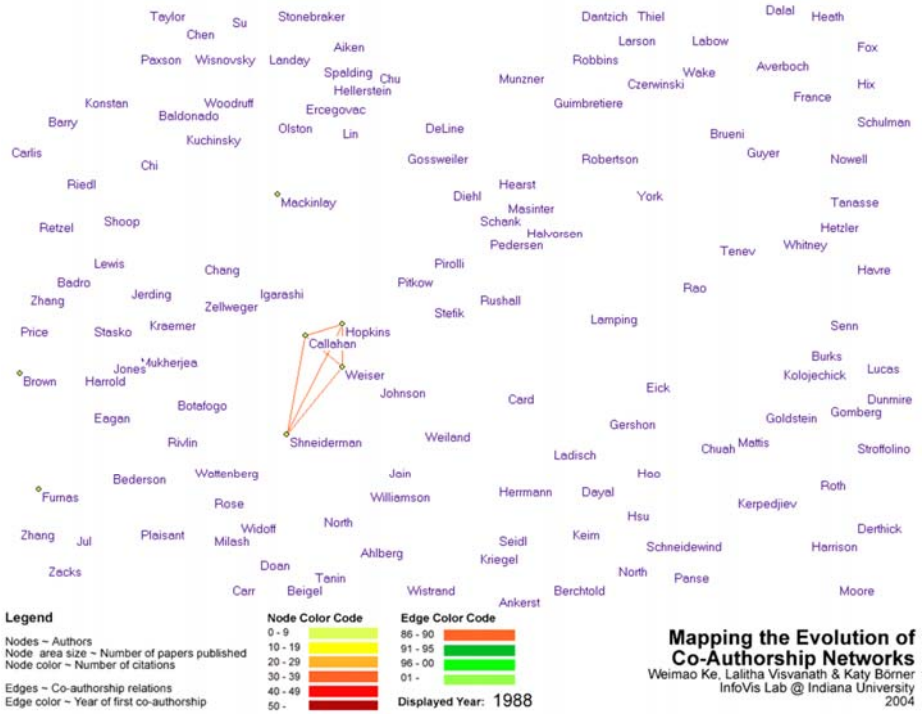
For more information on creating geospatial visualizations, see Sci² Manual Section 5.2.4.1.
http://wiki.cns.iu.edu/display/SCI2TUTORIAL/5.2.4+Mapping+Scientometrics+%28ISI+Data%29



Tutorial Overview

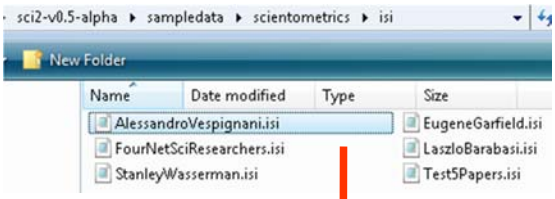
- Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools
- Sci² Tool Basics
- **Sci² Tool – Advanced Topics**
  - Walkthrough: Visualizing temporal data for NSF projects
  - Walkthrough: Locating data on a geographic map
  - **Walkthrough: Examining an evolving network**
  - Interacting with the statistical toolkit R and the network visualization package Gephi
  - Sci² tool visualizations
    - Bipartite networks
    - Map of Science
- Outlook and Q&A
- Adjourn

Sci² Demo II:
Evolving collaboration networks



Evolving Collaboration Networks

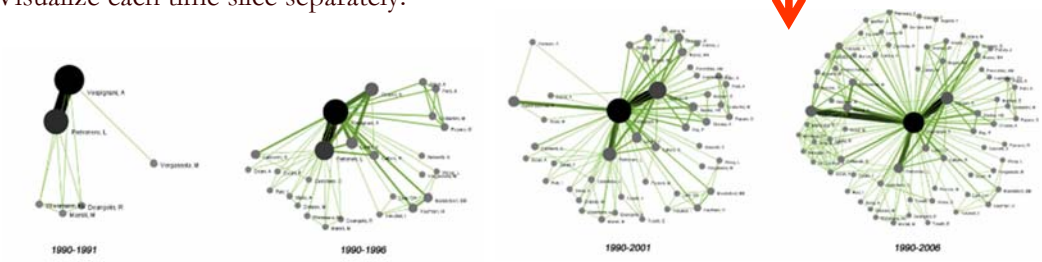
Load ISI formatted file



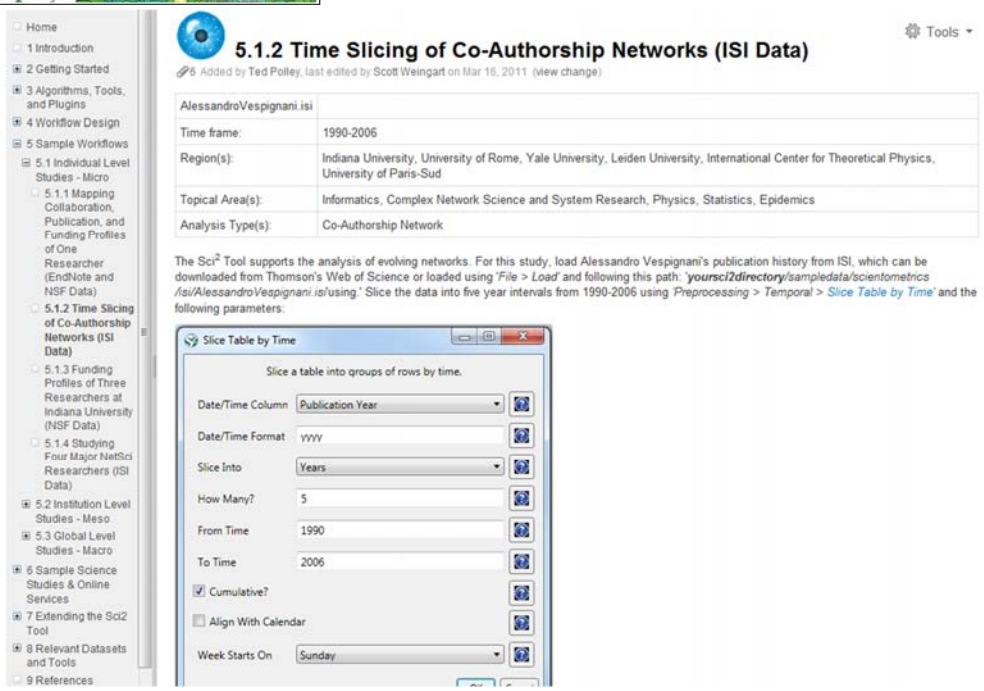
As CSV, file looks like:

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

Visualize each time slice separately:

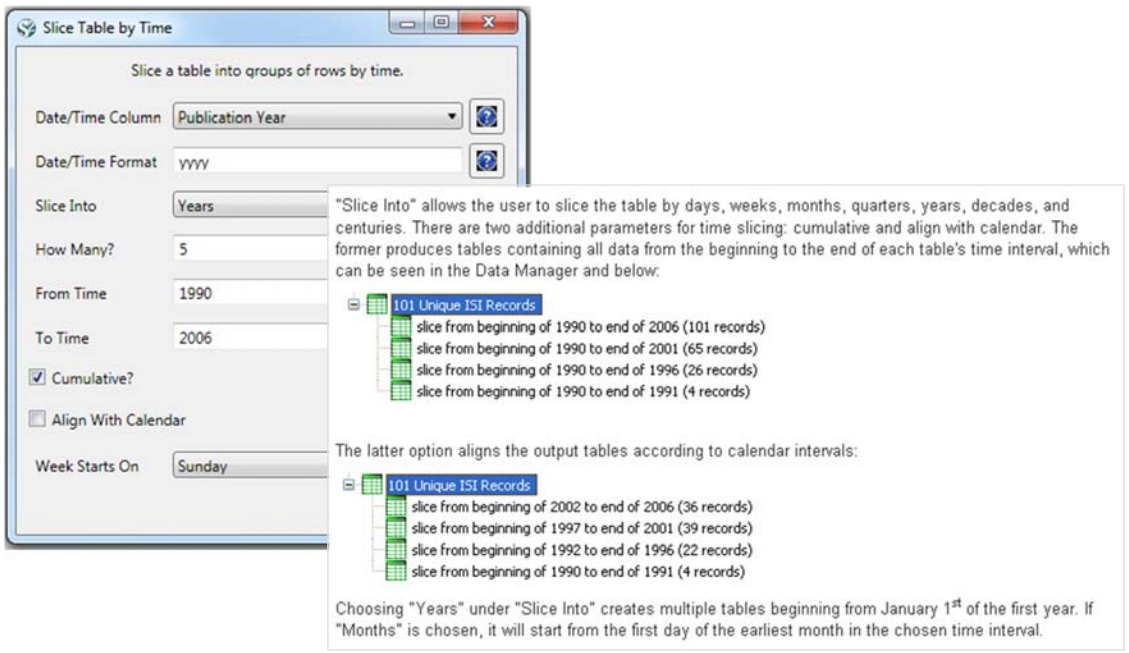


Relevant Sci² Manual Entry



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

Slice Table by Time



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

Slice Table by Time

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

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

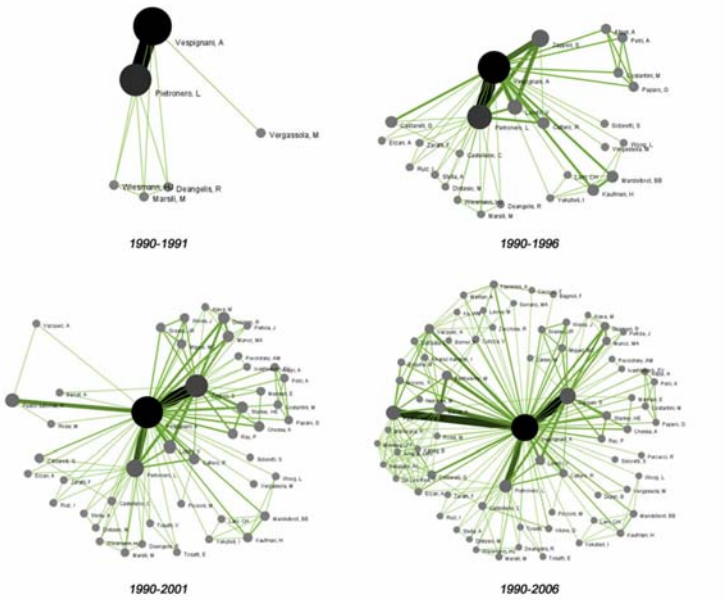
This algorithm uses the Joda Time library extensively, which provides significantly improved capabilities compared to the default Java algorithms for dates and times.

Visualize Each Network, Keep Node Positions

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

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

Visualize Each Network, Keep Node Positions

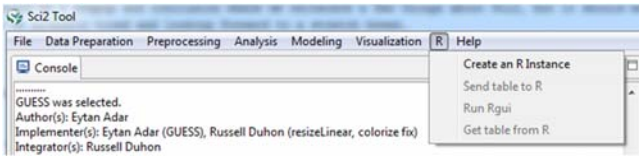


http://sci2.wiki.cns.iu.edu/5.1.2+Time+Slicing+of+Co-Authorship+Networks+(ISI+Data)

Tutorial Overview

- Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools
- Sci² Tool Basics
- **Sci² Tool – Advanced Topics**
  - Walkthrough: Visualizing temporal data for NSF projects
  - Walkthrough: Locating data on a geographic map
  - Walkthrough: Examining an evolving network
  - **Interacting with the statistical toolkit R and the network visualization package Gephi**
  - Sci² tool visualizations
    - Bipartite networks
    - Map of Science
- Outlook and Q&A
- Adjourn

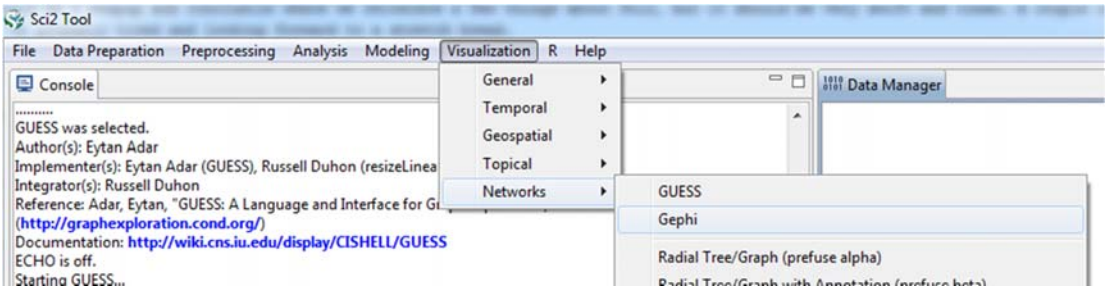
R-Bridge



- Run '*R* > *Create an R Instance*'. You must set the 'R Executable Directory' parameter to 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* > *Send Table to 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* > *Get Table From R'*. Choose the name of the variable from the dropdown list.

Gephi Bridge

- Gephi must already be installed on the system for this bridge plugin to work.
- Select any network file, then choose *Visualization* > *Networks* > *Gephi*
- The selected network file will be opened in Gephi using the *Import report*



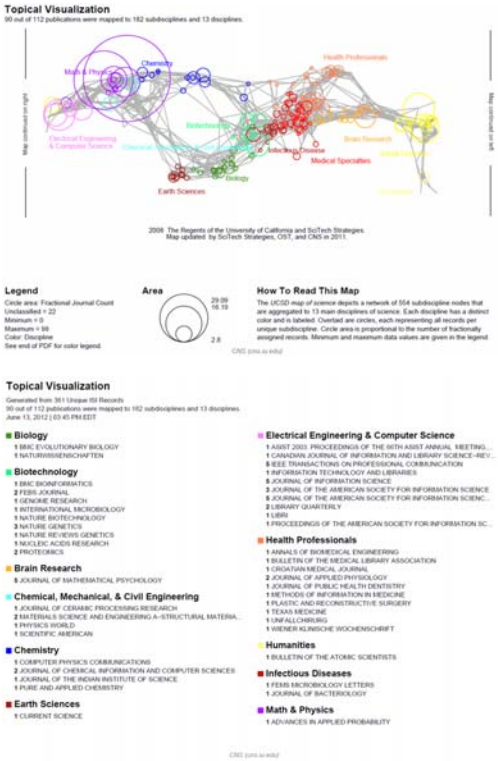
Tutorial Overview

- Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools
- Sci² Tool Basics
- **Sci² Tool – Advanced Topics**
  - Walkthrough: Visualizing temporal data for NSF projects
  - Walkthrough: Locating data on a geographic map
  - Walkthrough: Examining an evolving network
  - Interacting with the statistical toolkit R and the network visualization package Gephi
  - **Sci² tool visualizations**
    - **Bipartite networks**
    - **Map of Science**
- Outlook and Q&A
- Adjourn

Map of Science

http://wiki.cns.iu.edu/display/CISHELL/Map+of+Science+via+Journals

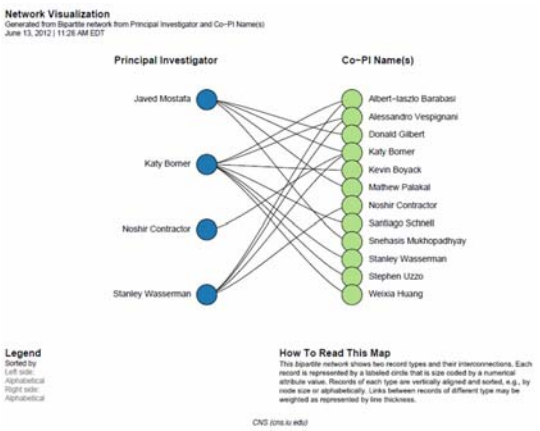
- The **Map of Science** is a visual representation of 554 sub-disciplines within 13 disciplines of science and their relationships to one another
- A set of journals can be overlaid onto this base map, showing the areas of science encompassed
  - Each circle size represents the (weighted) proportion of the journal set mapped to that sub-discipline
- Detailed breakdown of each discipline



Bipartite Network Graph

http://wiki.cns.iu.edu/display/CISHELL/Bipartite+Network+Graph

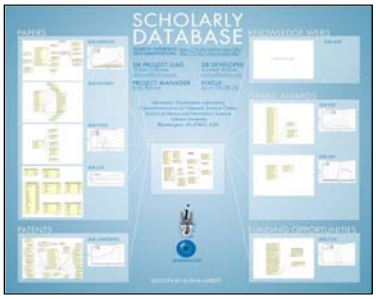
- A bipartite network graph can be visualized using this plugin.
- Supports edge and node weights
- Node ordering can be sorted independently



Tutorial Overview

- Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools
- Sci² Tool Basics
- Sci² Tool – Advanced Topics
- **Outlook and Q&A**
- Adjourn

Computational Scientometrics Cyberinfrastructures



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

James S. McDonnell Foundation

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

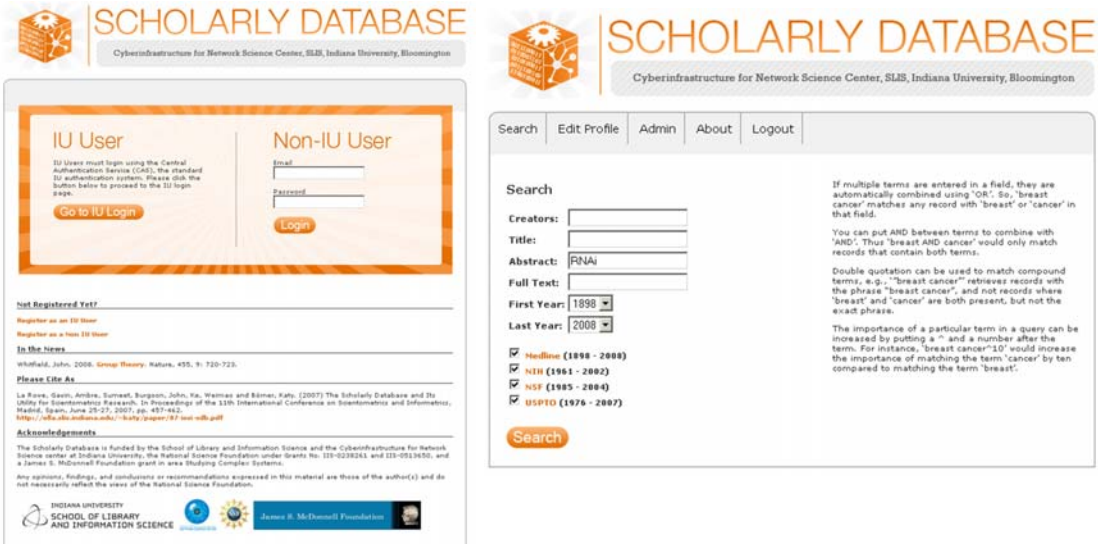
Epidemics Tool & Marketplace
Forthcoming

OSGi Alliance CIShell cyberinfrastructure shell

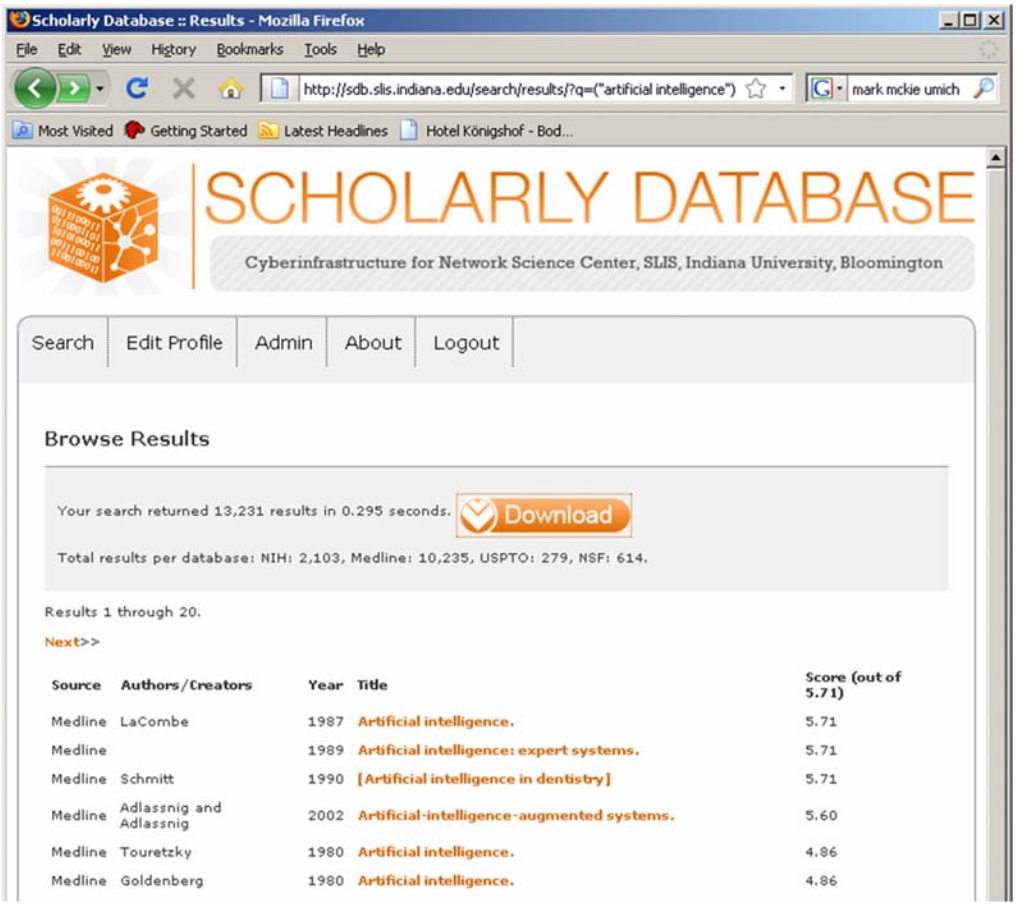
SDB | SCHOLARLY DATABASE

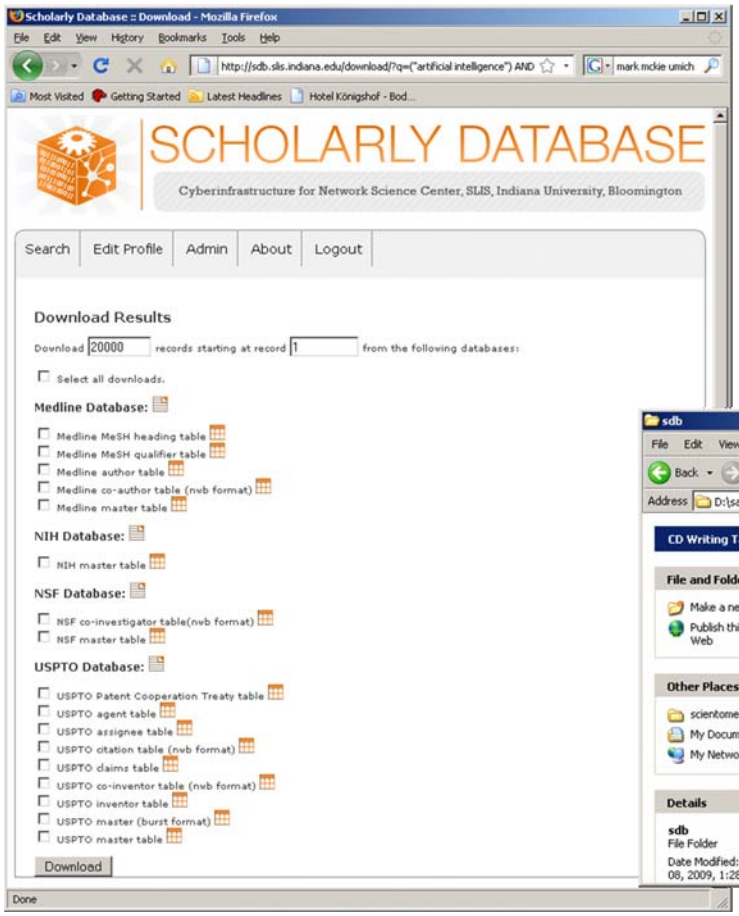
Scholarly Database at Indiana University
http://sdb.wiki.cns.iu.edu

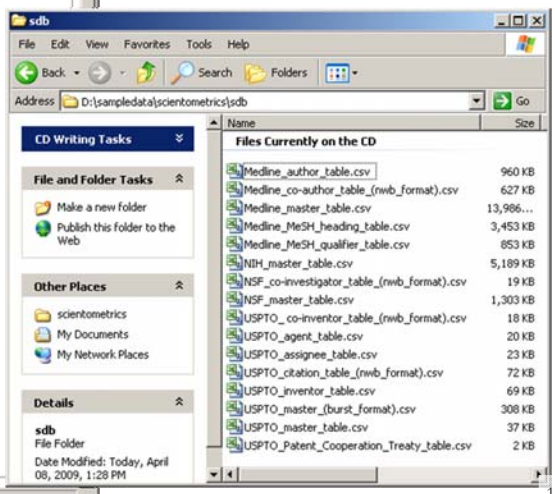
Supports federated search of 25 million publication, patent, grant records.
Results can be downloaded as data dump and (evolving) co-author, paper-citation networks.



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





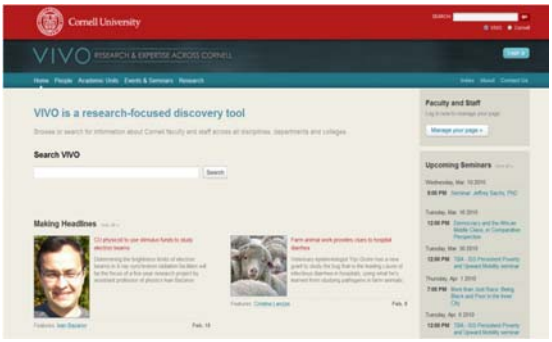


Download automatically extracted networks:
- Co-author
- Co-investigator
- Co-inventor
- Patent-citation

Download tables suitable for burst analysis

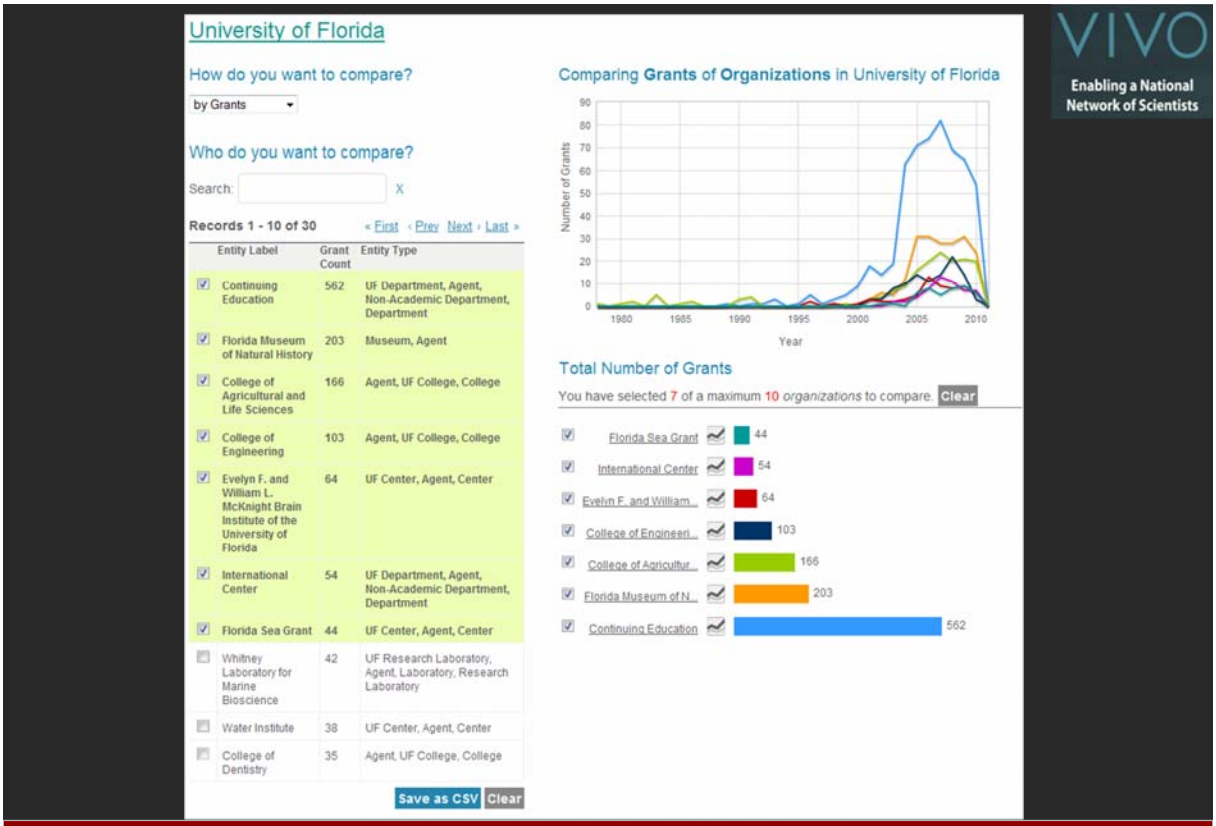
VIVO: A Semantic Approach to Creating a National Network of Researchers (http://vivoweb.org)

- 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.

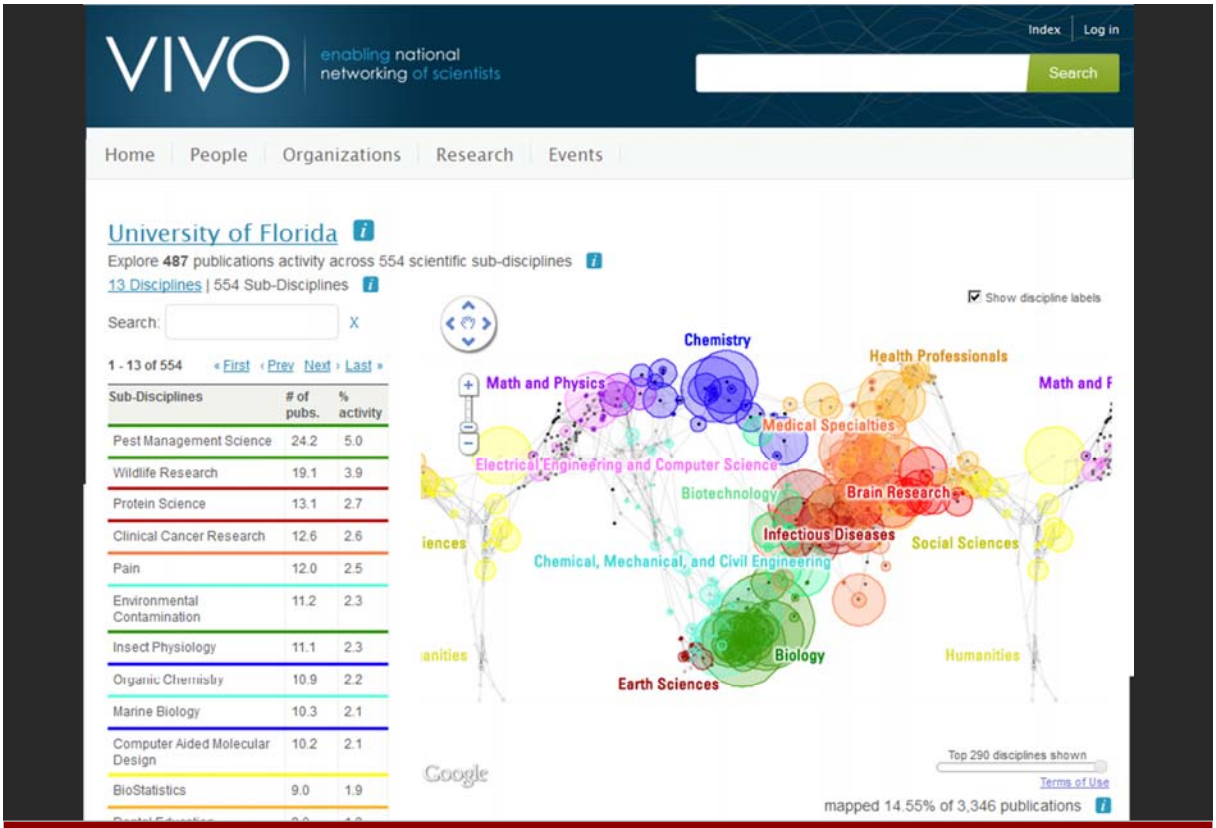


Funded by $12 million NIH award.

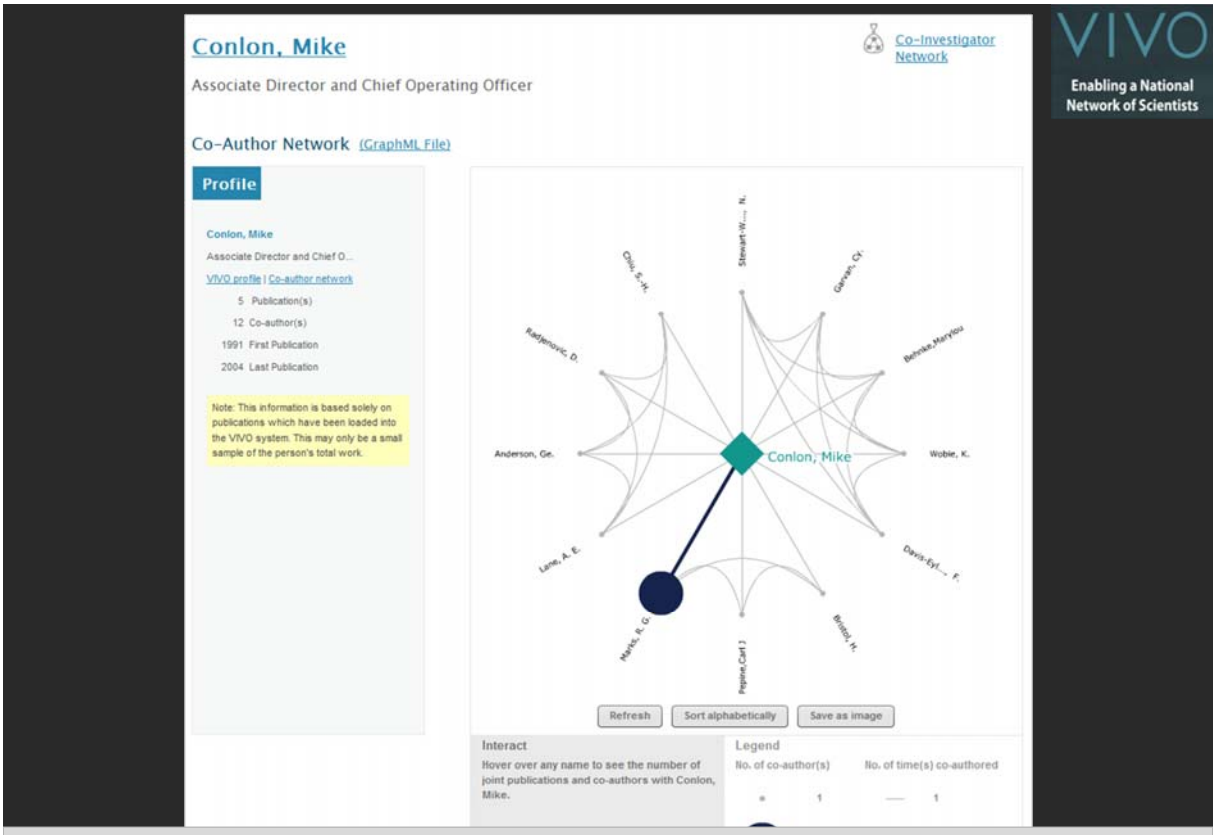
Cornell University: Dean Krafft (Cornell PI), Manolo Bevia, Jim Blake, Nick Cappadona, Brian Caruso, Jon Corson-Rikert, Elly Cramer, Medha Devare, John Fereira, Brian Lowe, Stella Mitchell, Holly Mistlebauer, Anup Sawant, Christopher Westling, Rebecca Younes. **University of Florida:** Mike Conlon (VIVO 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



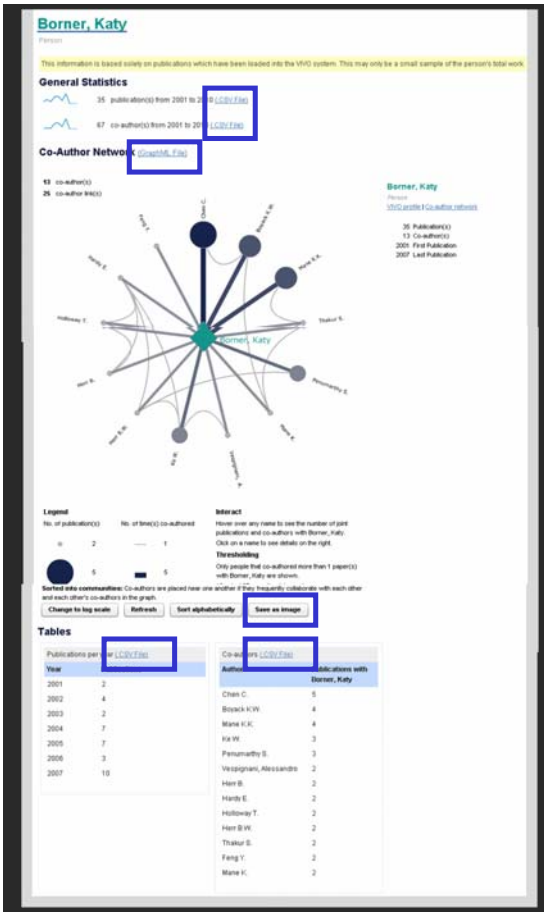
Topical Analysis (What) Science map overlays will show where a person, department, or university publishes most in the world of science. (in work)



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

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

VIVO Enabling a National Network of Scientists

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

VIVO Enabling a National Network of Scientists

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

v

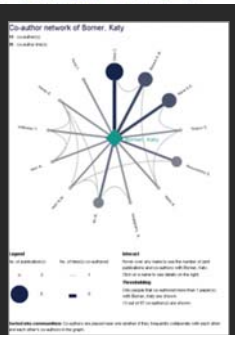
80 co-author(s) from 2001 to 2010 (.CSV File)

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)

```
<?xml version="1.0" encoding="UTF-8"?>
<graphml xmlns="http://graphml.graphdrawing.org/xmlns"
 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
 xsi:schemaLocation="http://graphml.graphdrawing.org/xmlns
 http://graphml.graphdrawing.org/xmlns/1.0/graphml.xsd">
<key id="label" for="node" attr.name="label" attr.type="string" />
<key id="number_of_authored_works" for="node" attr.name="number_of_authored_works" attr.type="int" />
<key id="num_unknown_publication" for="node" attr.name="num_unknown_publication" attr.type="int" />
<key id="num_latest_publication" for="node" attr.name="num_latest_publication" attr.type="int" />
<key id="latest_publication" for="node" attr.name="latest_publication" attr.type="int" />
<key id="profile_url" for="node" attr.name="profile_url" attr.type="string" />
```

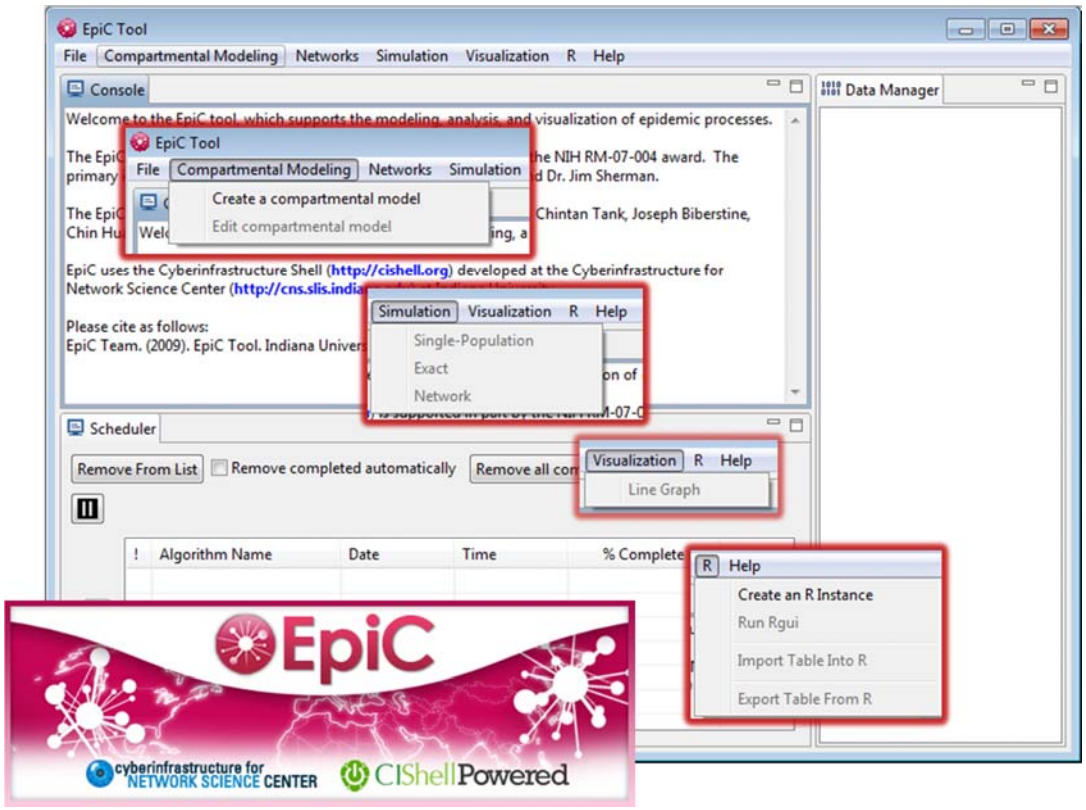
Save as Image (.PNG file)



Publications per year (.CSV File), see top file.

Co-authors (.CSV File)

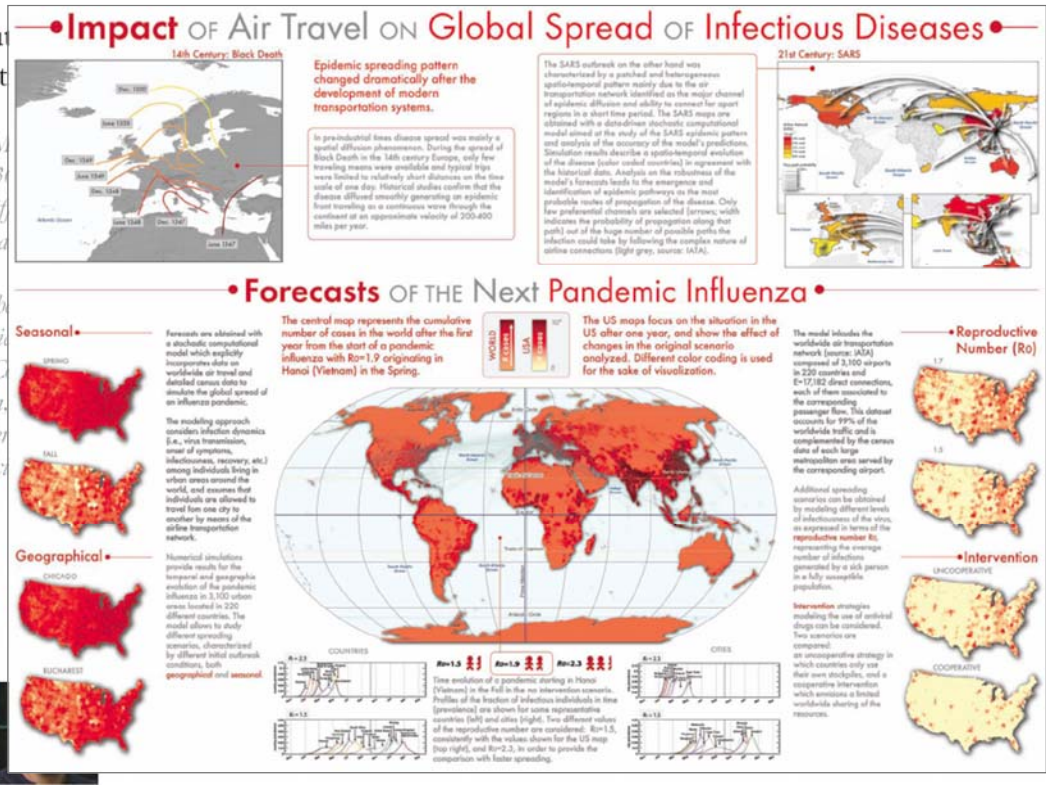
Co-Author	Count
Andrienko G.	1
Andrienko N.	1
Ben-Miled Z.	1
Blackwell A.	1
Boyack K.W.	4
Bozicevic M.	1
Brodbeck D.	1
Burkhard R.A.	1
Chen C.	5

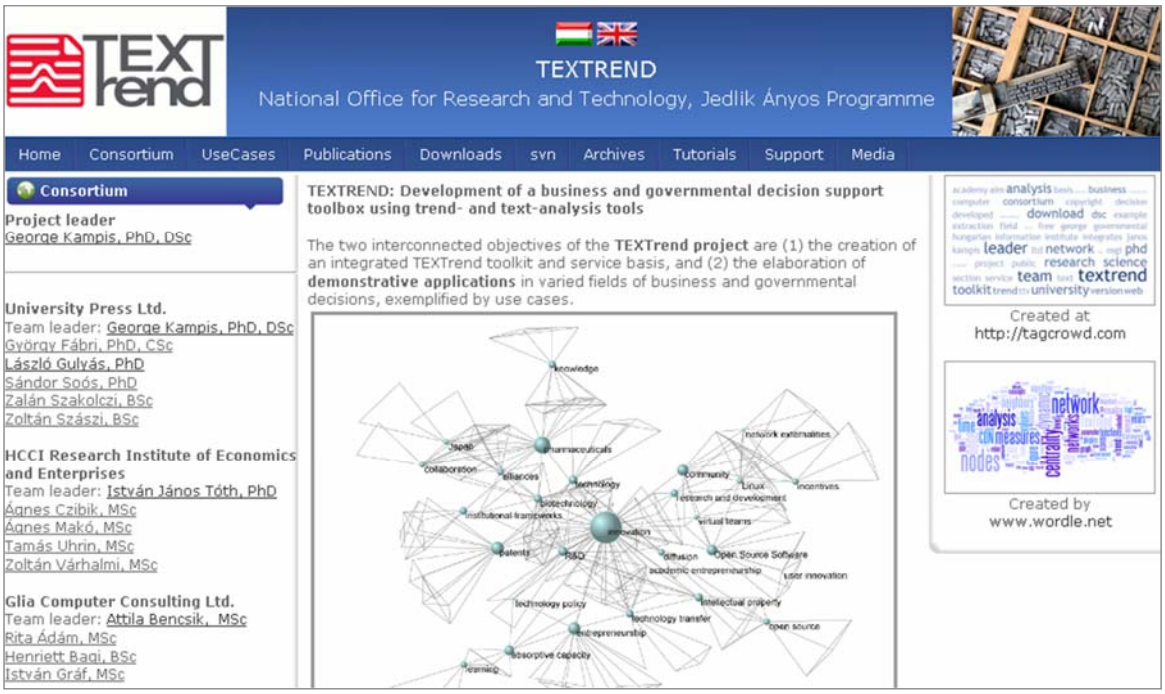


Comput
Forecast

Epidemic M
Biologie, 33
Reaction-dif
A.Vespigna

Modeling th
of Pandemi
Case and C
V. Colizza,
A.-J. Valle
PloS-Medic



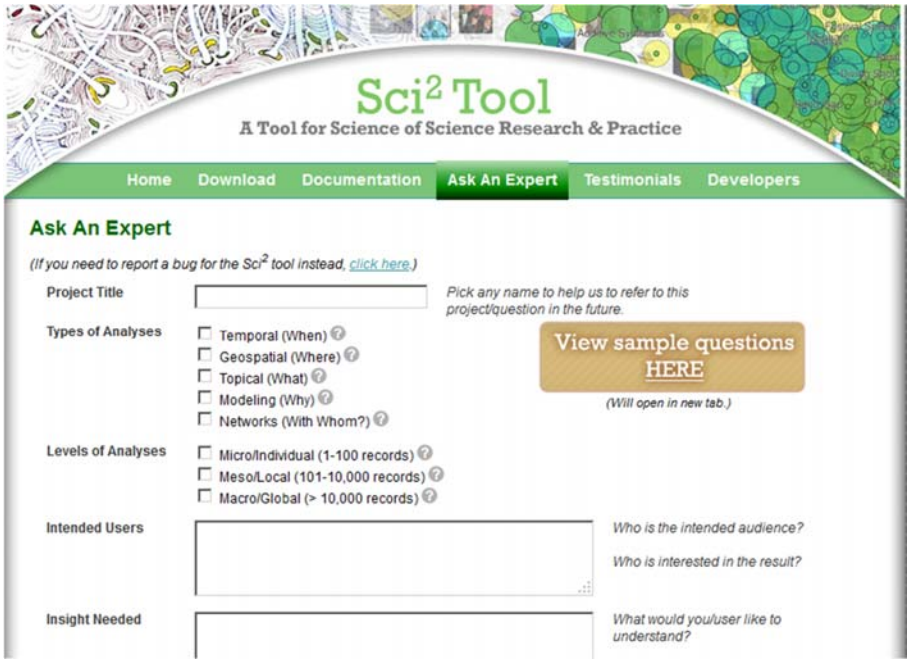


TEXTrend adds WEKA, Wordij, CFinder, and more.

See the latest versions of TEXTrend Toolkit modules at

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

Need Help? Ask an Expert!



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

Tutorial Overview

- Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools
- Sci² Tool Basics
  - Download and run the Sci² Tool
  - Walkthrough: Load, analyze, and visualize a network
  - Walkthrough: Analyzing the publications of four prominent network science researchers
    - Load and clean a dataset; extract networks from raw data
    - Calculate basic statistics and analyses of the network
    - Visualize the results
- Sci² Tool – Advanced Topics
  - Walkthrough: Visualizing temporal data for NSF projects
  - Walkthrough: Locating data on a geographic map
  - Walkthrough: Examining an evolving network
  - Interacting with the statistical toolkit R and the network visualization package Gephi
  - Sci² tool visualizations
    - Bipartite networks
    - Map of Science
- Outlook and Q&A
- Adjourn

Thank you!

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