

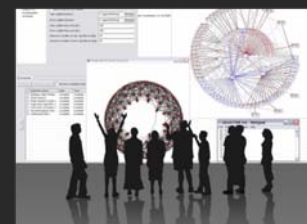
Sci2: A Tool of Science of Science Research and Practice Tutorial

Dr. Katy Börner & Joseph Biberstine
Cyberinfrastructure for Network Science Center
Information Visualization Laboratory
School of Library and Information Science
Indiana University, Bloomington, IN
<http://cns.iu.edu>



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



NSF's Stafford Place II Conference Center in Arlington, Virginia
Monday, Oct 17, 2011 • 8am – 11:30am



Software, Datasets, Plugins, and Documentation

- These slides
<http://ivl.slis.indiana.edu/km/pres/2011-borner-sci2tutorial-nsf.pdf>
- Sci2 Tool Manual v0.5.1 Alpha
<http://sci2.wiki.cns.iu.edu>
- Sci2 Tool v0.5.1 Alpha (May 4, 2011)
<http://sci2.cns.iu.edu>
- Additional Datasets
<http://sci2.wiki.cns.iu.edu/2.5+Sample+Datasets>
- Additional Plugins
<http://sci2.wiki.cns.iu.edu/3.2+Additional+Plugins>



Or copy them from the DVD or memory stick.



Tutorial Overview

8:00a Welcome and Overview of Tutorial and Attendees

8:15a Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools

8:30a 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

10:00 Break

10:15a Sci2 Tool Novel Functionality

- Horizontal line graph of NSF projects
- Geospatial maps aggregated by congressional districts
- Evolving collaboration networks
- R-Bridge

11:00a Outlook and Q&A

11:30 Adjourn

3



Tutorial Overview

8:00a Welcome and Overview of Tutorial and Attendees

8:15a Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools

8:30a 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

10:00 Break

10:15a Sci2 Tool Novel Functionality

- Horizontal line graph of NSF projects
- Geospatial maps aggregated by congressional districts
- Evolving collaboration networks
- R-Bridge

11:00a Outlook and Q&A

11:30 Adjourn

4



Using the Sci2 Tool to Visualize Tutorial Registrants

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

	A	B	C
1	Last Name	Affiliation	Background
2	Di Gangi	Loyola University Maryland	Statistics
3	Cogburn	COTELCO, American University/ Syracuse University	Information Science and Technology
4	Coleman	Duke University	Graduate Student
5	McCain	Drexel University	Life Sciences, Information science, bibliometrics
6	Hewes	NIH	Chemistry
7	Fearon	NIH	Biology
8	Webber	NIH	

Total of 79 registrants on Oct 10, 2011. Affiliations with more than two participants:

Affiliations	#
NSF	26
NIH	16
IDA Science and Technology Policy Institute	3
AAAS	2
AAAS/NSF	2
Discovery Logic, Thomson Reuters	2
DoE	2
NOAA	2
USDA	2

5



Using the Sci2 Tool to Visualize Tutorial Registrants

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

	A	B	C
1	Last Name	Affiliation	Background
2	Di Gangi	Loyola University Maryland	Statistics
3	Cogburn	COTELCO, American University/ Syracuse University	Information Science
4	Coleman	Duke University	Graduate Student
5	McCain	Drexel University	Life Sciences, In
6	Hewes	NIH	Chemistry
7	Fearon	NIH	Biology
8	Webber	NIH	

Run *'Data Preparation > Extract Bipartite Network'*

With parameter values: \longrightarrow

Optional: Calculate Node Degree

Visualize resulting *Bipartite network*

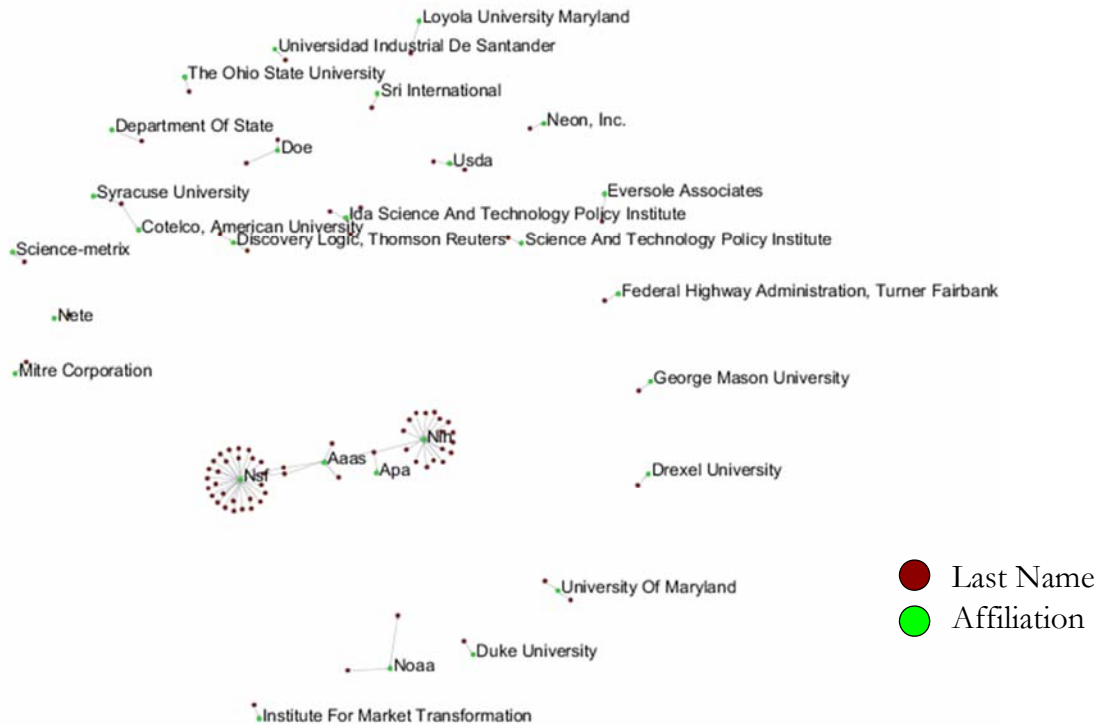
from *Last Name and Org*' using *'Visualization > Network > GUESS'* and

'Layout > GEM, Layout > Bin Pack'

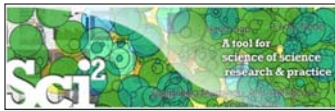
6



Sci2 Tool – Visualize Workshop Attendees: Bi-partite People-Affiliation Network



7



Sci2 Tool – Visualize Workshop Attendees: Bi-partite People-Background Network

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

	A	B	C
1	Last Name	Affiliation	Background
2	Di Gangi	Loyola University Maryland	Statistics
3	Cogburn	COTELCO, American University/ Syracuse University	Information Science
4	Coleman	Duke University	Graduate Student
5	McCain	Drexel University	Life Sciences, In
6	Hewes	NIH	Chemistry
7	Fearon	NIH	Biology
8	Webber	NIH	

Run *'Data Preparation > Extract Bipartite Network'*

With parameter values \longrightarrow

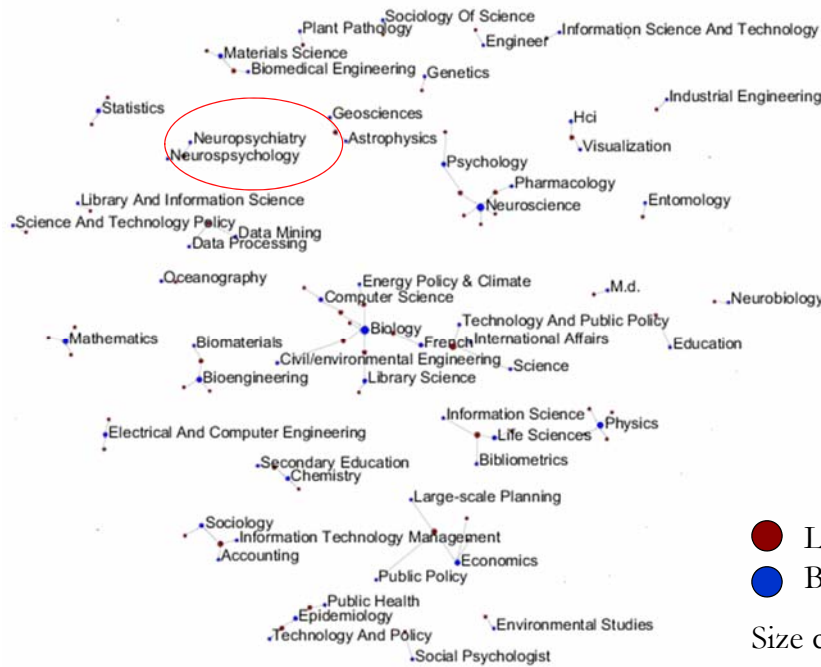
Calculate Node Degree

Then select Bipartite network from Last Name and Background and run *'Visualization > Network > GUESS.'*

8



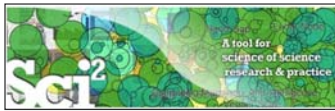
Sci2 Tool – Visualize Workshop Attendees: Bi-partite People-Background Network



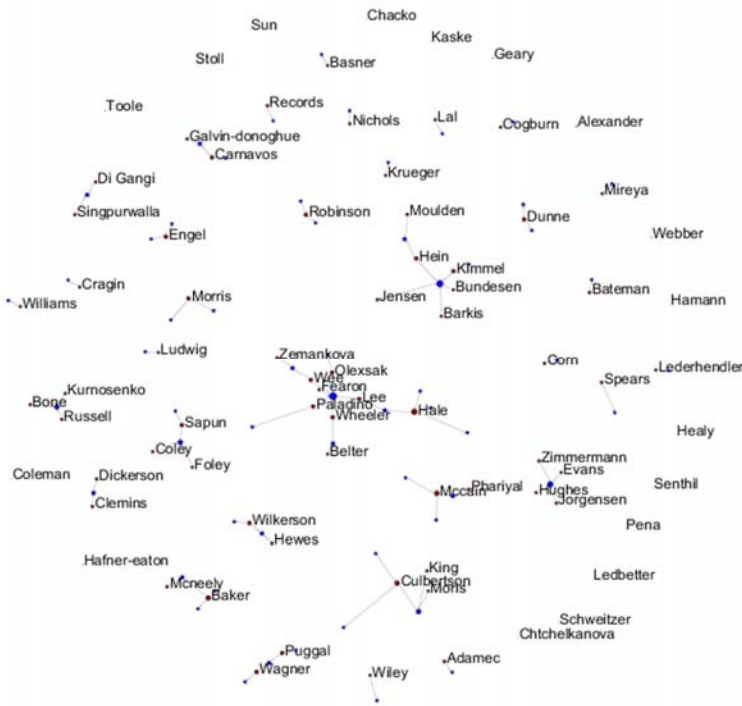
● Last Name
● Background
Size coded by degree.

Color/label via
Size code using

Object: nodes based on -> Property: bipartitetype Operator: == Value: Background
Nodes totaldegree From: 2 To: 20 Do Resize Linear



Sci2 Tool – Visualize Workshop Attendees: Bi-partite People-Background Network



● Last Name
● Background
Size coded by degree.

Color/label via

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



Wordle.net of “Interest to Learn” Response



11



Workshop Overview

8:00a Welcome and Overview of Workshop and Attendees

8:15a Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools

8:30a 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

10:00 Break

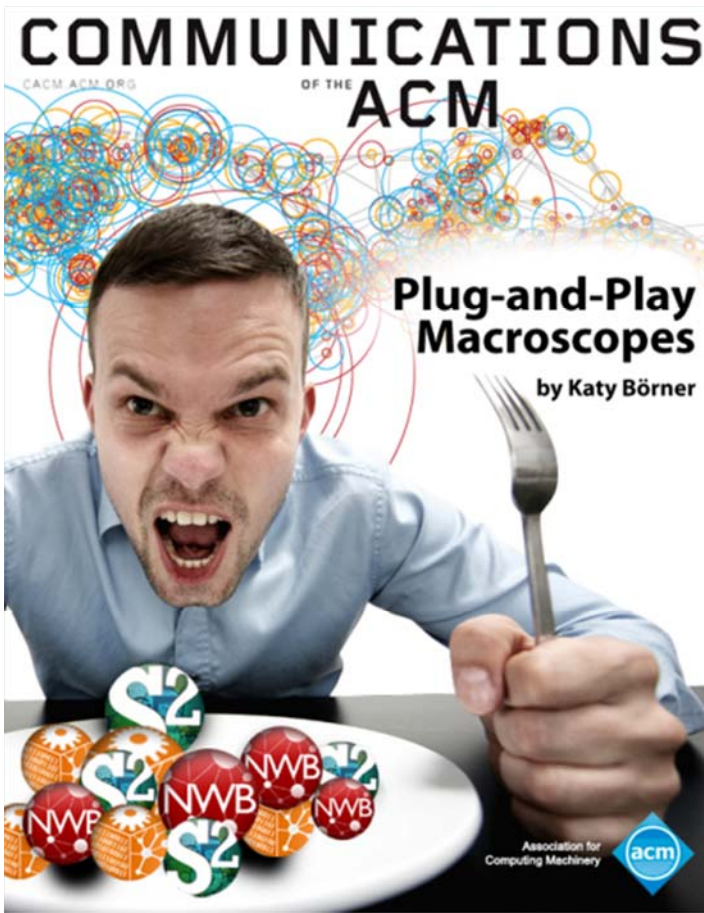
10:15a Sci2 Tool Novel Functionality

- Horizontal line graph of NSF projects
- Geospatial maps aggregated by congressional districts
- Evolving collaboration networks
- R-Bridge

11:00a Outlook and Q&A

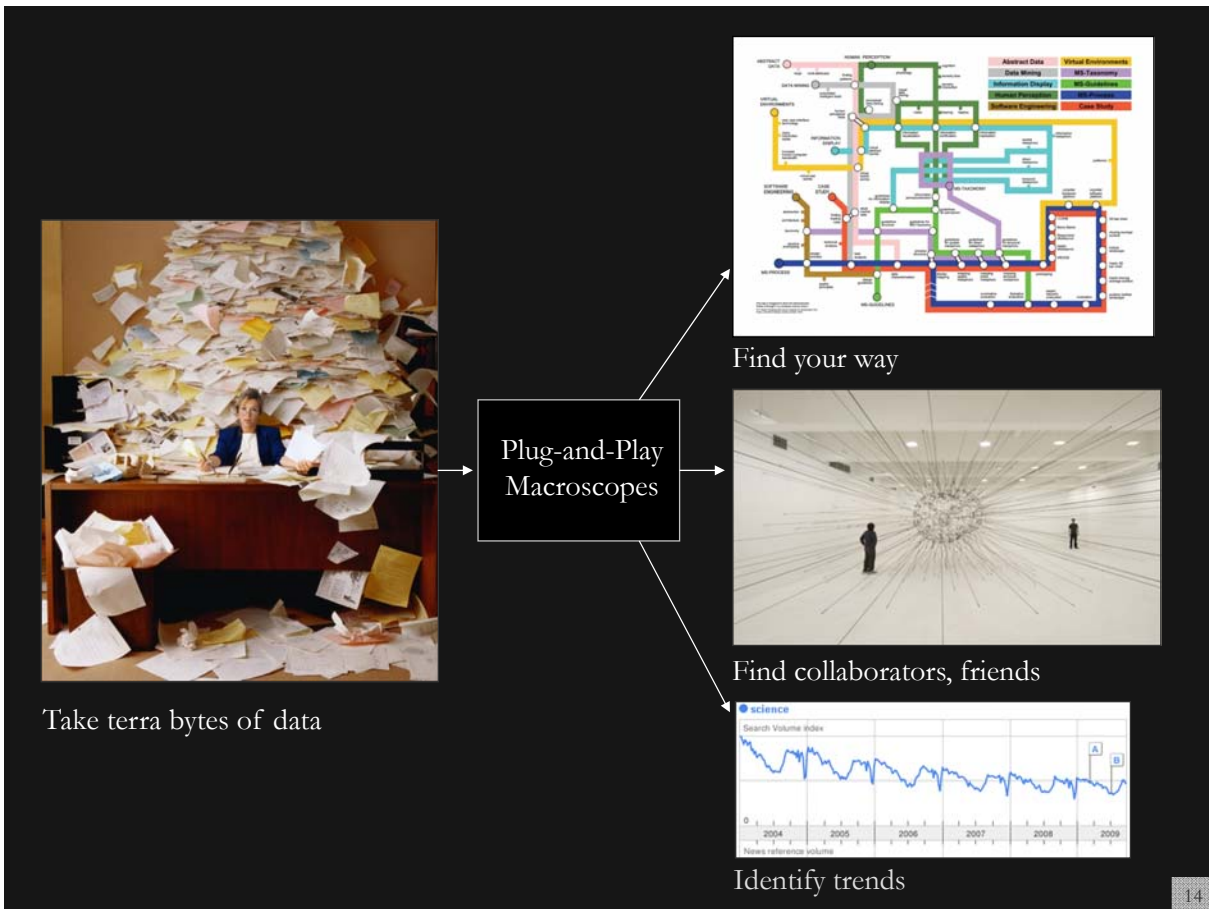
11:30 Adjourn

12



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>



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

15

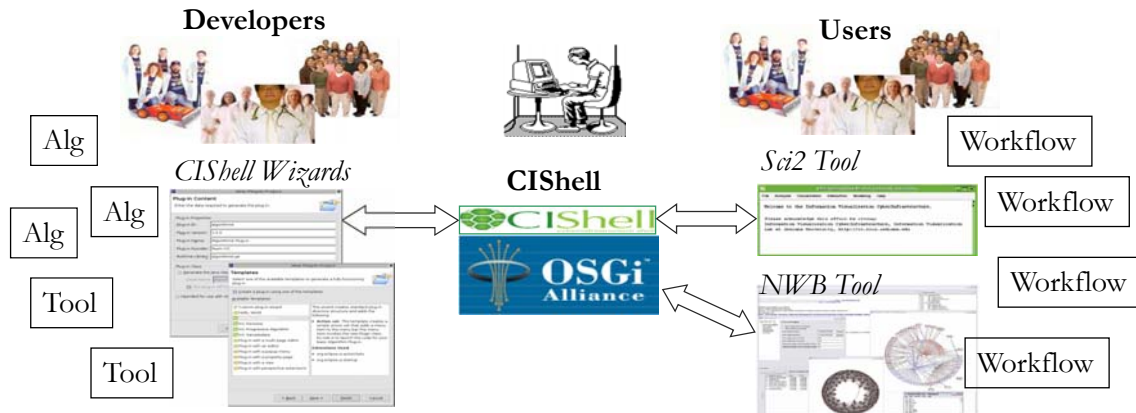
Plug-and-Play Macroscopes

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.

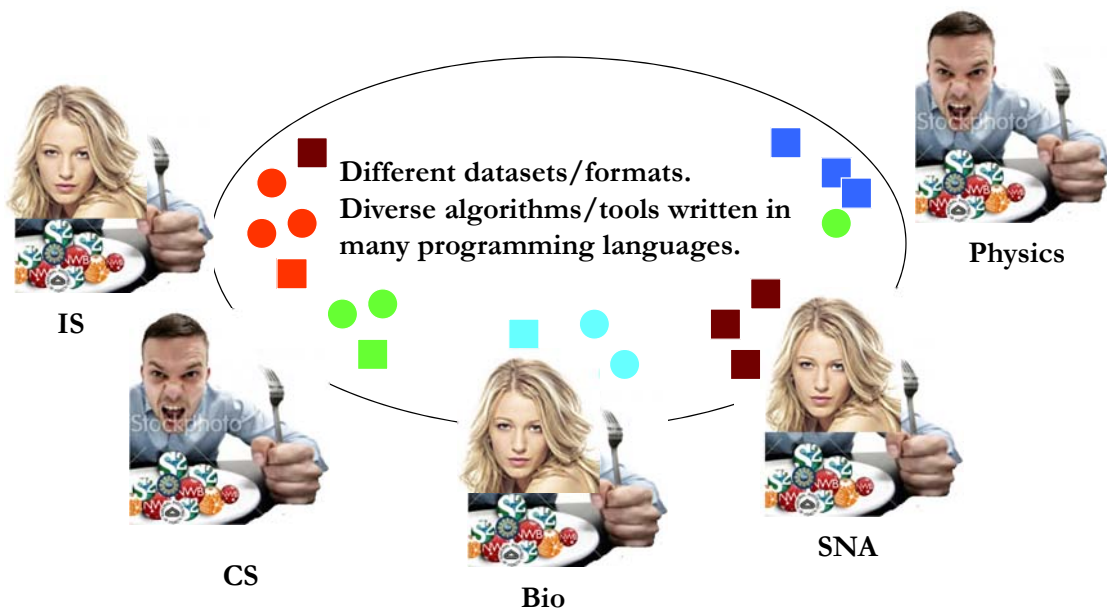
They make it easy to share plug-ins via email, flash drives, or online. To use new plugins, simply copy the files into the plug-in directory, and they appear in the tool menu ready for use. No restart of the tool is necessary. **Sharing algorithm components, tools, or novel interfaces becomes as easy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

16

- 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.



17



18



Tutorial Overview

8:00a Welcome and Overview of Workshop and Attendees

8:15a Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools

8:30a 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

10:00 Break

10:15a Sci2 Tool Novel Functionality

- Horizontal line graph of NSF projects
- Geospatial maps aggregated by congressional districts
- Evolving collaboration networks
- R-Bridge

11:00a Outlook and Q&A

11:30 Adjourn

19



Science of Science (Sci2) Tool

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

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

nature

OPINION

SUMMARY

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

Vol 464|25 March 2010

Let's make science metrics more scientific

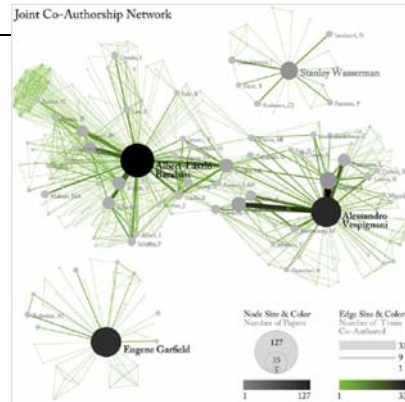
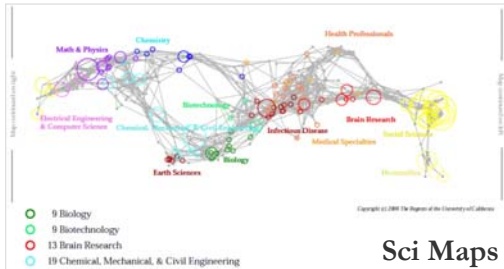
To capture the essence of good science, stakeholders must combine forces to create an open, sound and consistent system for measuring all the activities that make up academic productivity, says **Julia Lane**.

20

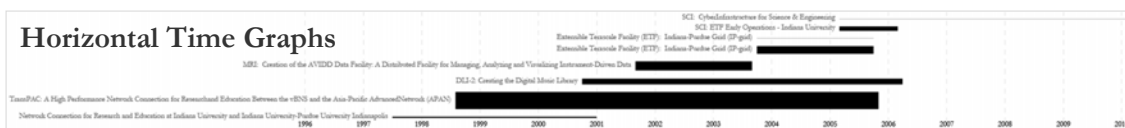


Sci² Tool – “Open Code for S&T Assessment”

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



Horizontal Time Graphs



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

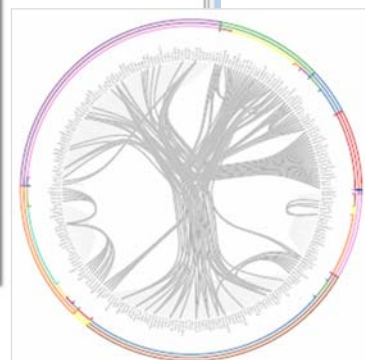
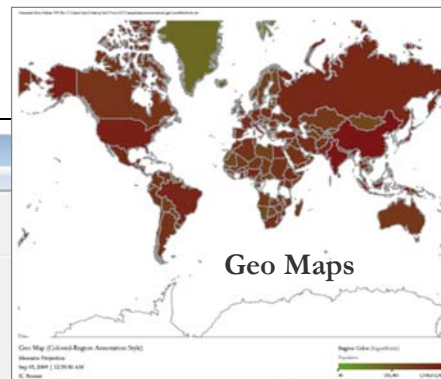


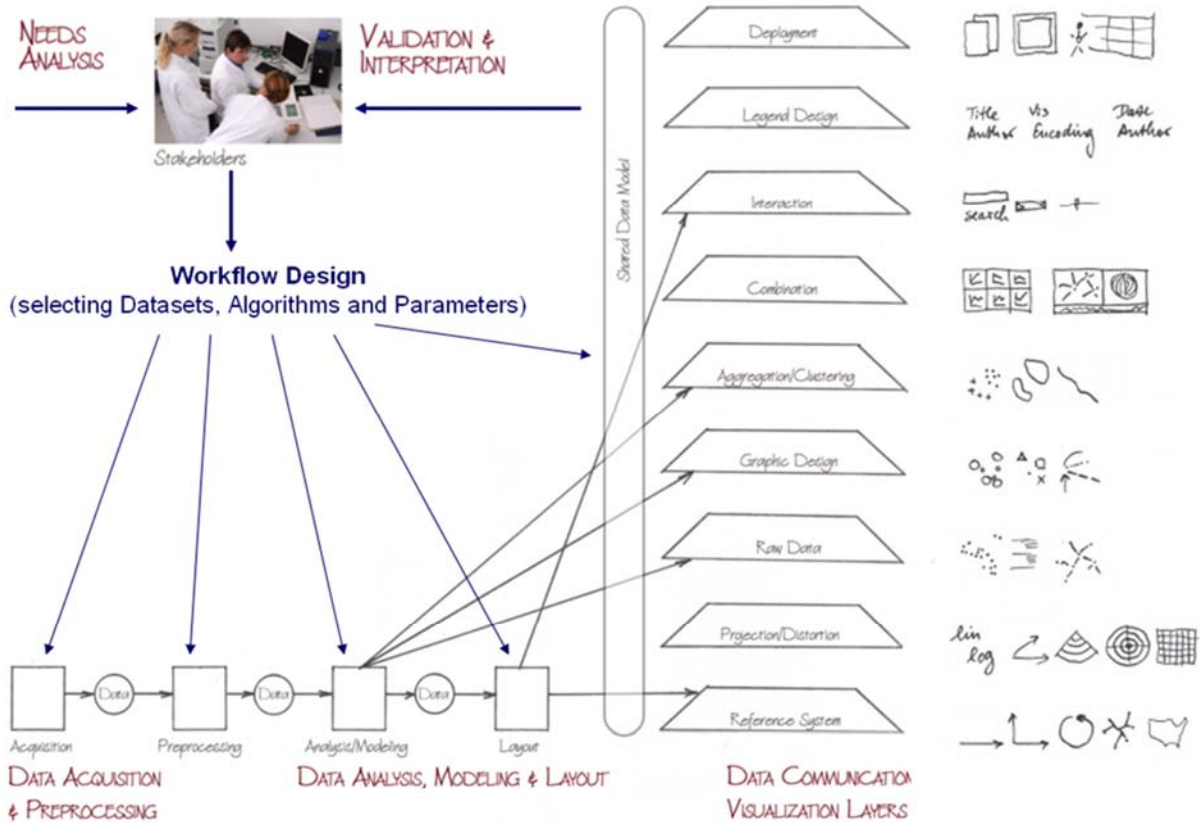
Sci² Tool Visualizations

Visualization Menu:

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

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





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 NSI, SA, all of sci
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	Wrapping a intellectual l	PNAS
Topical Analysis (What)		research	VxOrd/Topic r NIH funding
Network Analysis (With Whom?)	NSI one	Network of	NIH's cy



Type of Analysis vs. Level of Analysis Covered Today:

	<i>Micro/Individual</i> (1-100 records)	<i>Meso/Local</i> (101-10,000 records)	<i>Macro/Global</i> (10,000 < records)
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		Locations
Topical Analysis (What)			Word/Topic maps of funding
Network Analysis (With Whom?)	NSF Co-F one indivi		NSF's core competency

25



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

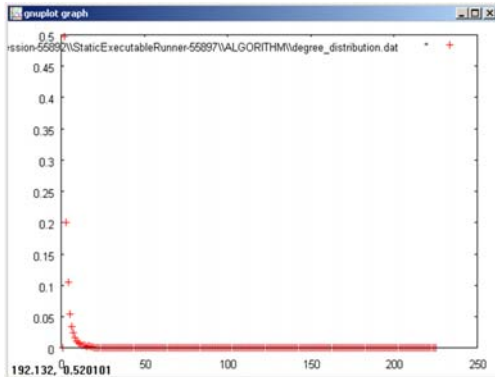
- JPEG (*.jpg)
- PDF (*.pdf)
- PostScript (*.ps)

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

26

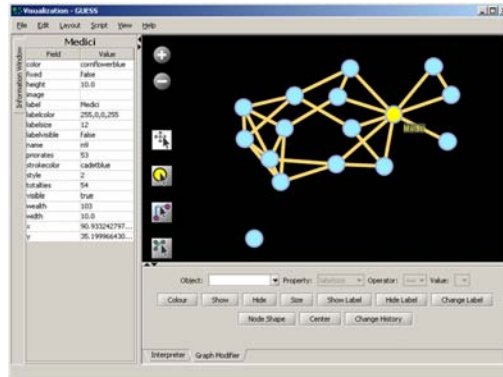


Sci² Tool – Supported Tools



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>.

27



Sci² Tool: Download, Install, and Run

Sci² Tool v0.5 Alpha (April 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.

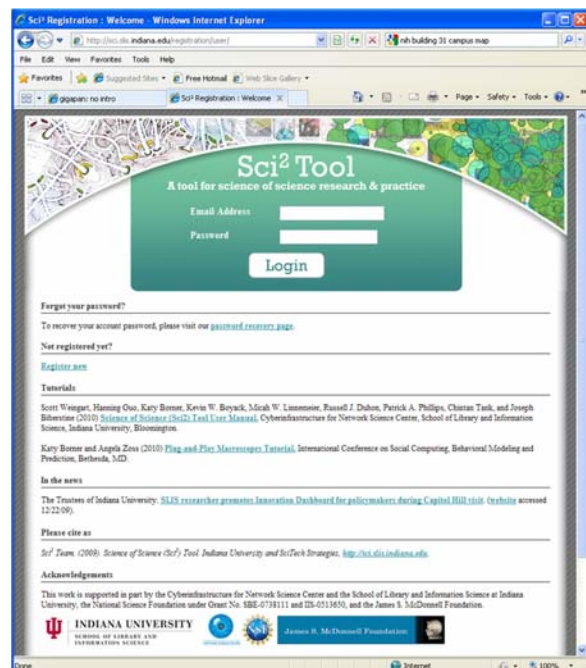
Run /sci2/sci2.exe

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>



28



Sci² Tool: Download, Install, and Run

Sci2 Tool v0.5 Alpha (April 4, 2011)

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



Sci2 Tool runs on Windows, Mac, and Linux.

Unzip.

Run /sci2/sci2.exe



29

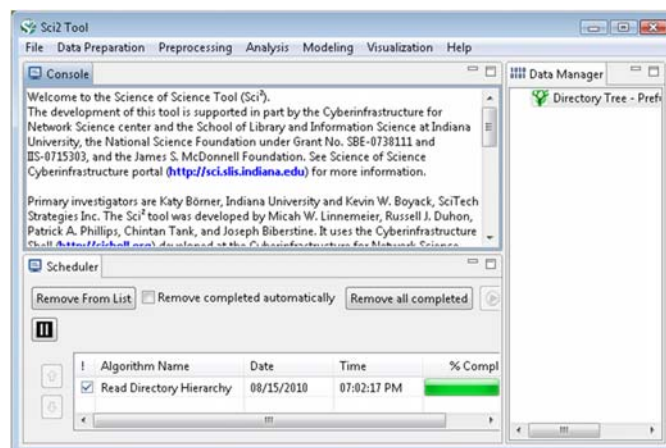


Sci² Tool Interface Components

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>

30



Tutorial Overview

8:00a Welcome and Overview of Workshop and Attendees

8:15a Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools

8:30a 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

10:00 Break

10:15a Sci2 Tool Novel Functionality

- Horizontal line graph of NSF projects
- Geospatial maps aggregated by congressional districts
- Evolving collaboration networks
- R-Bridge

11:00a Outlook and Q&A

11:30 Adjourn

31



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

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

Node attributes

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

Edge attributes:

- Marriage T/F
- Business T/F

“Substantively, the data include families who were locked in a struggle for political control of the city of Florence around 1430. Two factions were dominant in this struggle: one revolved around the infamous Medicis, the other around the powerful Strozziis.”

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

32



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

Florentine families related through business as loans, credits and joint partnerships

Node attributes

- Wealth: Each family's net wealth
- Priorates: The number of seats of
- Totalities: Number of business/

Edge attributes:

- Marriage T/F
- Business T/F

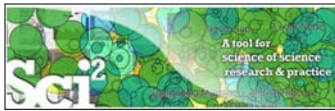
“Substantively, the data include family control of the city of Florence around a struggle: one revolved around the influence of the Strozzi.”

More info is at <http://svitsrv25.epfl>

```
*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"
```

33



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

- Load **yoursci2directory*/sampledata/socialscience/florentine.mvb*
- Run ‘*Analysis > Network Analysis Toolkit (NAT)*’ to get basic properties.

This graph claims to be undirected.

Nodes: 16

Isolated nodes: 1

Node attributes present: label, wealth, totalities, priorates

Edges: 27

No self loops were discovered.

No parallel edges were discovered.

Edge attributes:

Nonnumeric attributes:

Example value

marriage...T

business...F

Average degree: 3.375

There are 2 weakly connected components. (1 isolates)

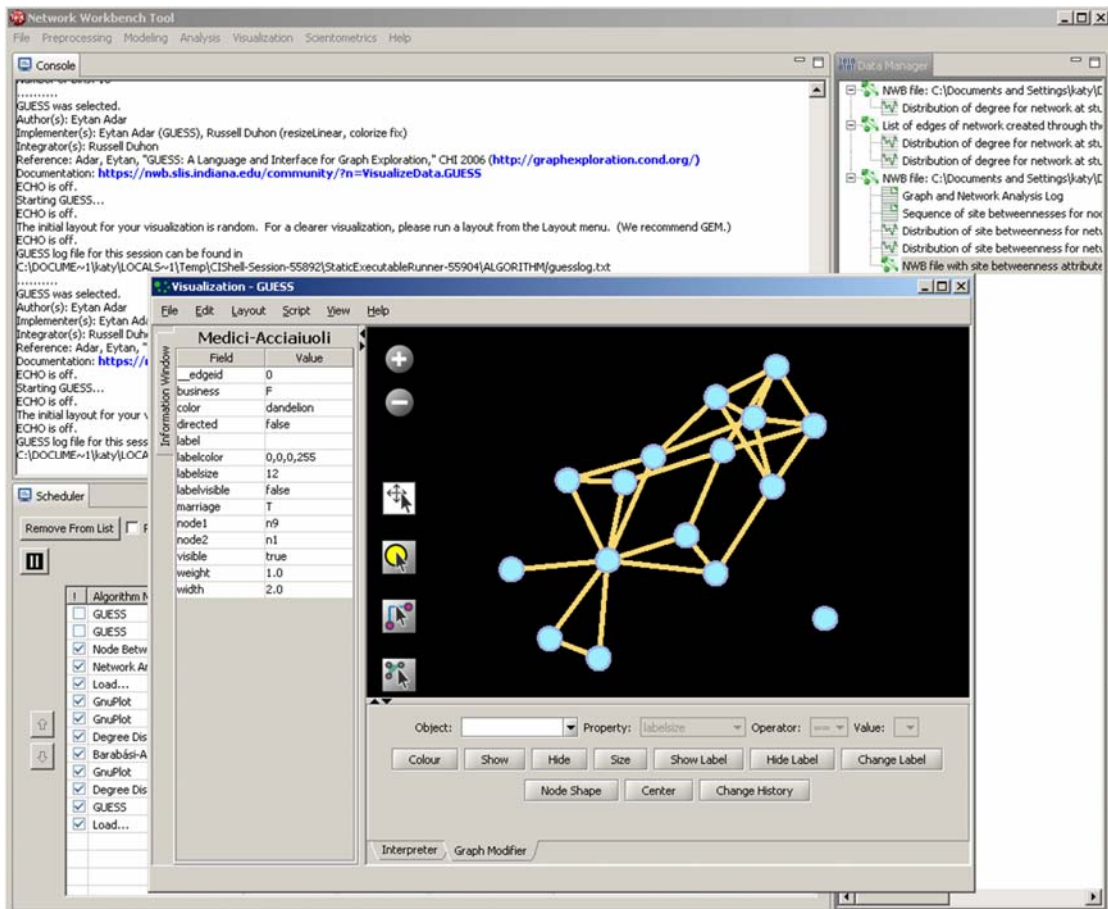
The largest connected component consists of 15 nodes.

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

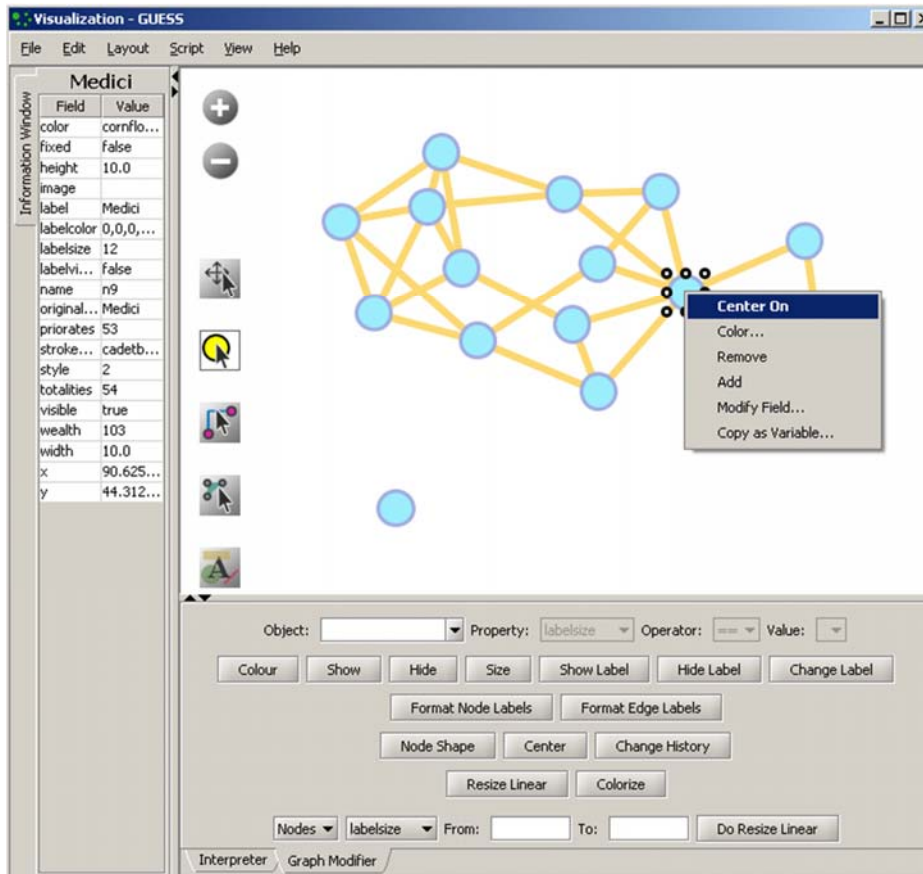
Density (disregarding weights): 0.225

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

34




35



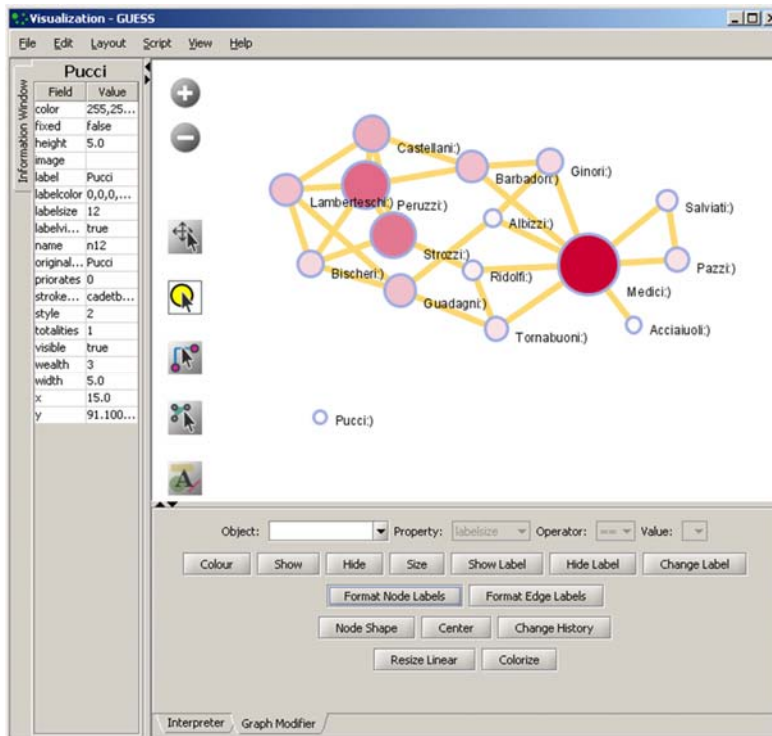
Pan:
"grab" the background by holding left-click and moving your mouse.

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.

36



Graph Modifier:

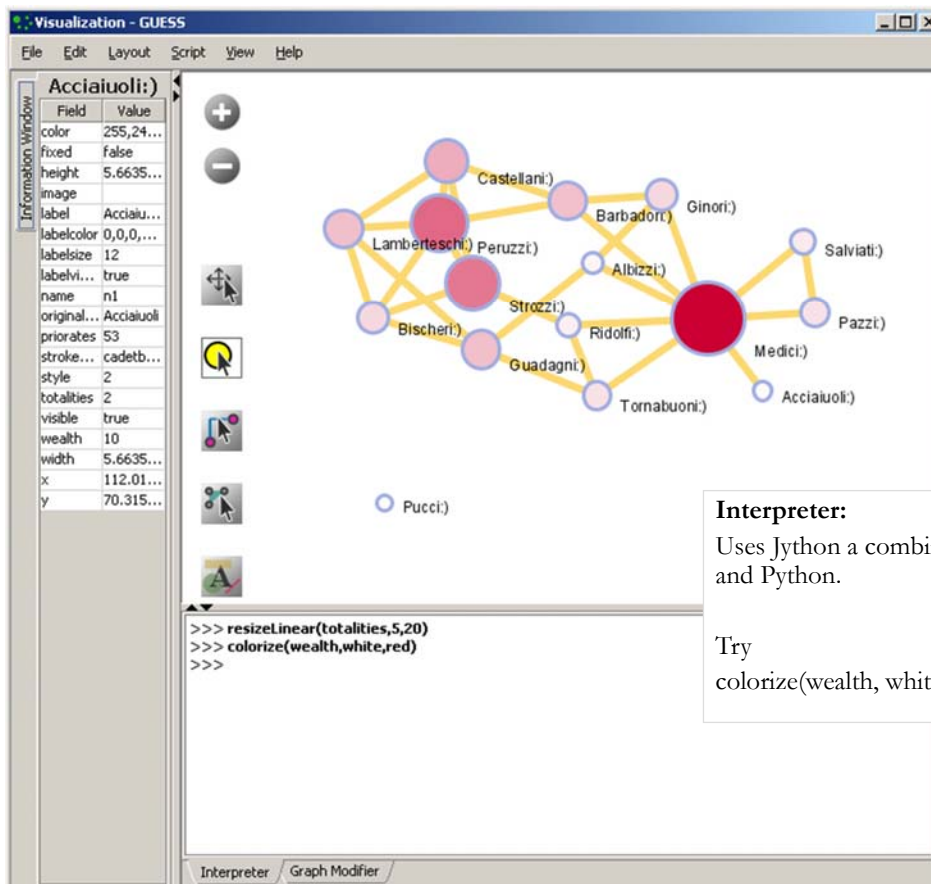
Select "all nodes" in the Object drop-down menu and click 'Show Label' button.

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.'

37



Interpreter:

Uses Jython a combination of Java and Python.

Try
colorize(wealth, white, red)

38



Workshop Overview

8:00a Welcome and Overview of Workshop and Attendees

8:15a Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools

8:30a 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

10:00 Break

10:15a Sci2 Tool Novel Functionality

- Horizontal line graph of NSF projects
- Geospatial maps aggregated by congressional districts
- Evolving collaboration networks
- R-Bridge

11:00a Outlook and Q&A

11:30 Adjourn

39



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

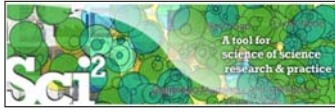
FourNetSciResearchers.isi	
Time frame:	1955-2007
Region(s):	Miscellaneous
Topical Area(s):	Network Science
Analysis Type(s):	Paper Citation Network, Co-Author Network, Bibliographic Coupling Network, Document Co-Citation Network, Word Co-Occurrence Network

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

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

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

40



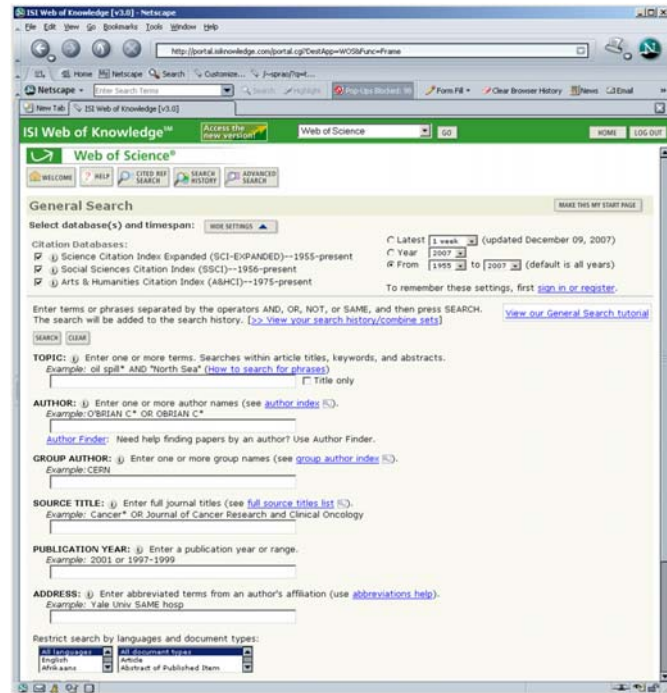
Data Acquisition from Web of Science

In Dec 2007, we downloaded all papers by

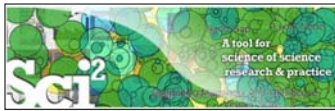
- Eugene Garfield
- Stanley Wasserman
- Alessandro Vespignani
- Albert-László Barabási

from

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



41



Comparison of Counts

No books and other non-WoS publications are covered.

	Age	Total # Cites	Total # Papers	H-Index
Eugene Garfield	82	1,525	672	31
Stanley Wasserman		122	35	17
Alessandro Vespignani	42	451	101	33
Albert-László Barabási	40	2,218	126	47 <i>(Dec 2007)</i>
	41	16,920	159	52 <i>(Dec 2008)</i>
	44	30,102	201	68 <i>(April 11)</i>

42

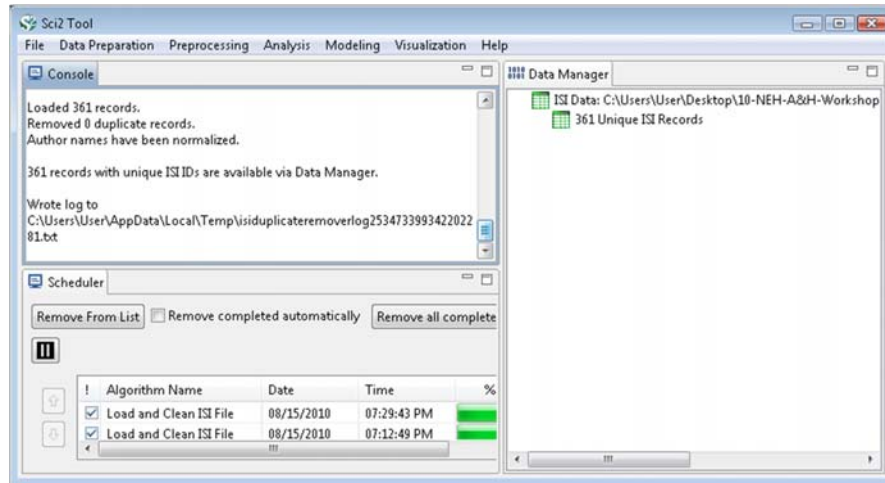


Extract Co-Author Network

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.



43



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.

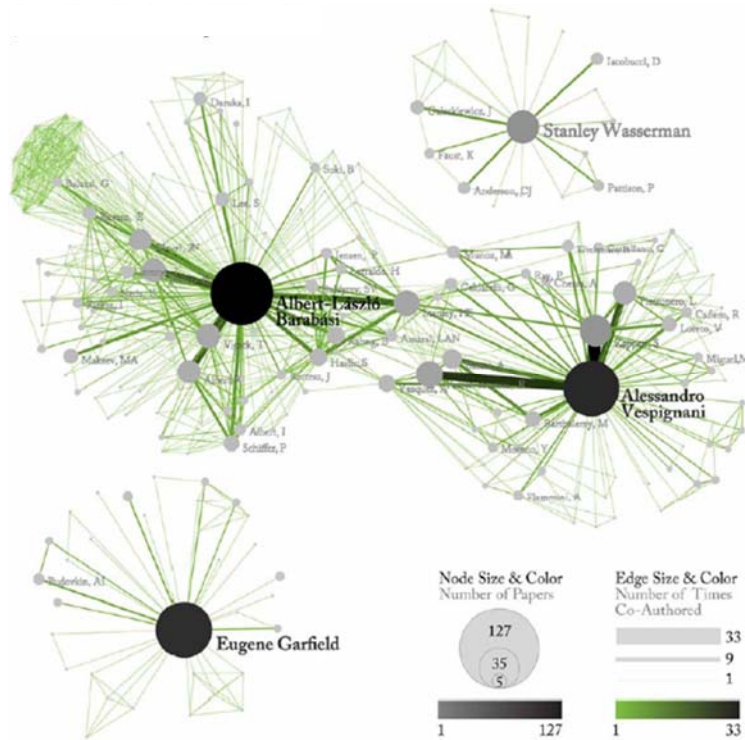
To view the complete network, select the *'Extracted Co-Authorship Network'* and run *'Visualization > Networks > GUESS'*.

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'*.

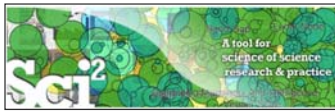
44



Co-Author Network of all Four NetsSci Researchers



45

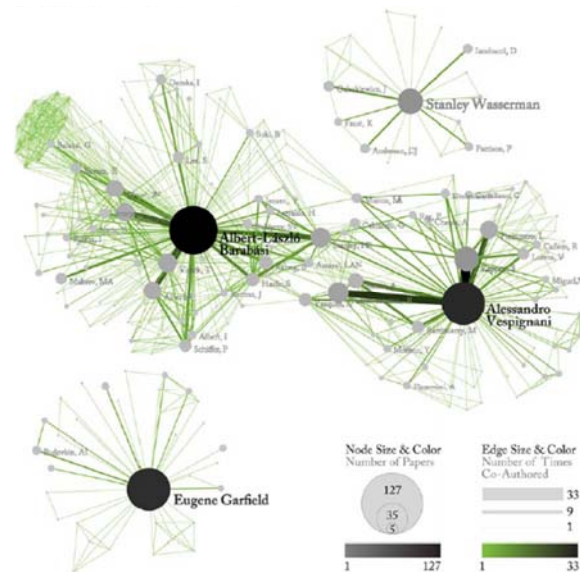


Co-Author Network of all Four NetsSci Researchers

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

Calculate node degrees in Sci2 Tool.

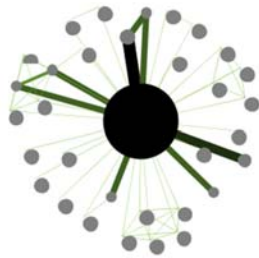
Use a graphic program to add legend.



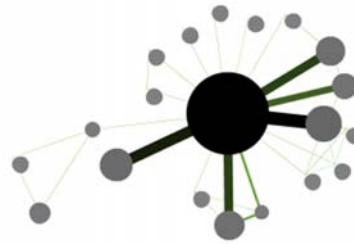
46



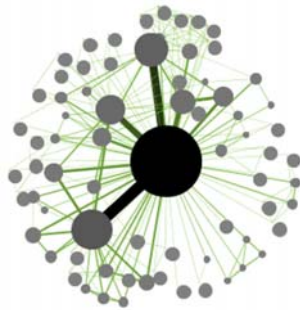
Individual Co-Author Networks (Read/map 4 files separately)



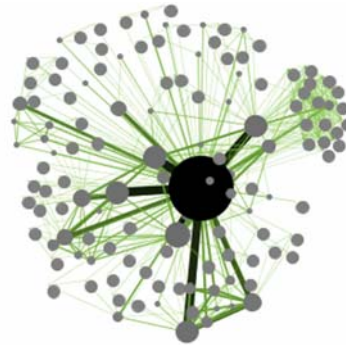
Eugene Garfield



Stanley Wasserman



Alessandro Vespignani



Albert-László Barabási

47



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

.....

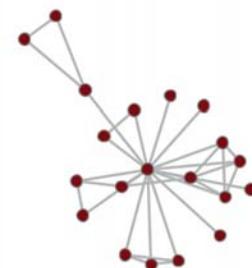
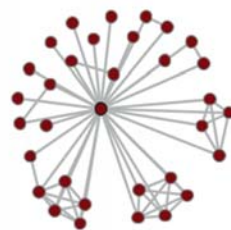
Network Analysis Toolkit (NAT) was selected.

Nodes: 247

Edges: 891

.....

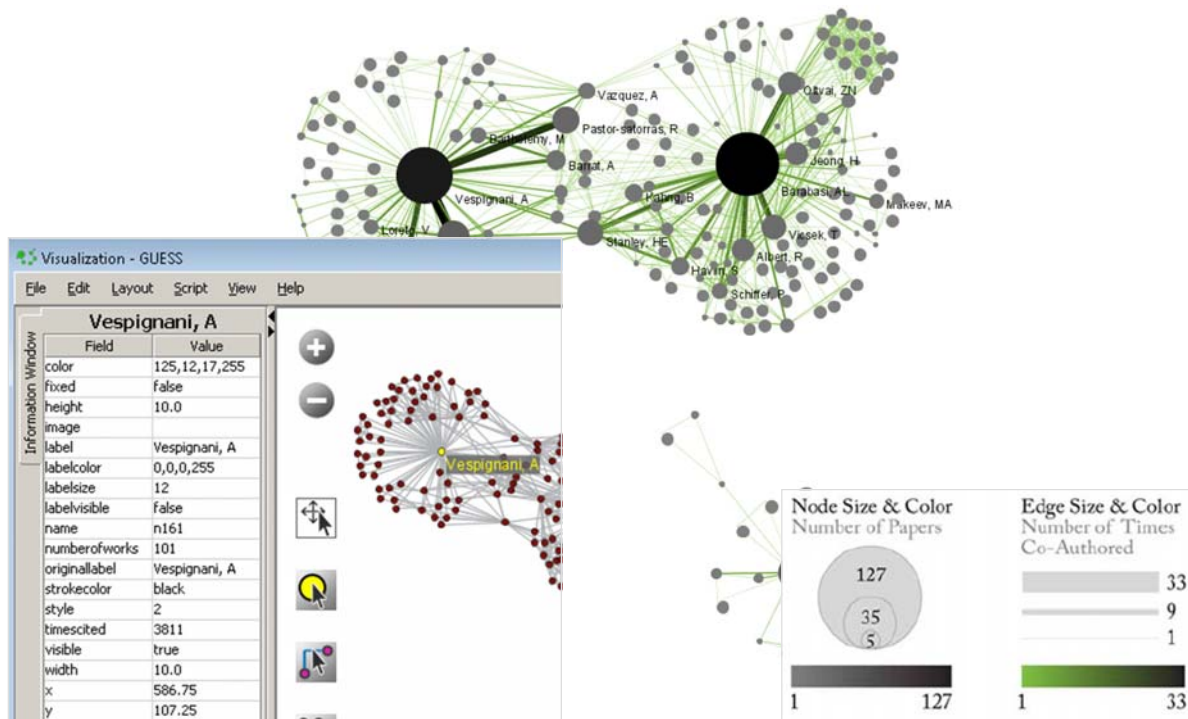
GUESS was selected.



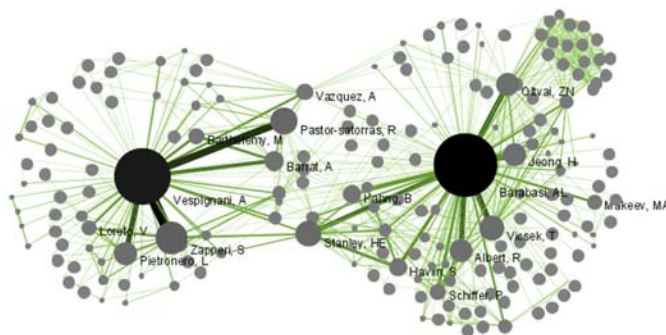
48



Network Visualization: Color/Size Coding by Data Attribute Values



Network Visualization: Giant Component

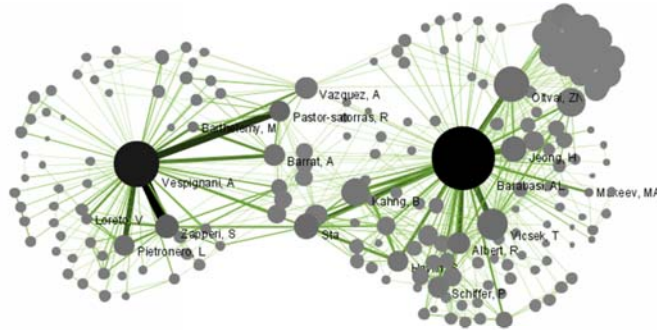


.....
Weak Component Clustering was selected.
Implementer(s): Russell Duhon
Integrator(s): Russell Duhon

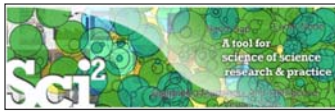
Input Parameters:
Number of top clusters: 10
3 clusters found, generating graphs for the top 3 clusters.
.....



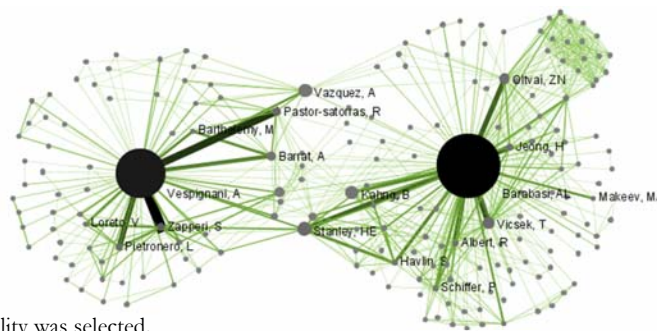
Network Visualization: Color/Size Coding by Degree



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



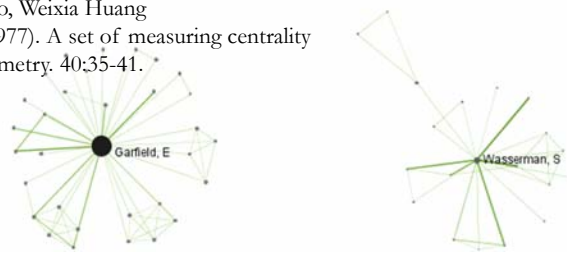
Network Visualization: Color/Size Coding by Betweenness Centrality



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

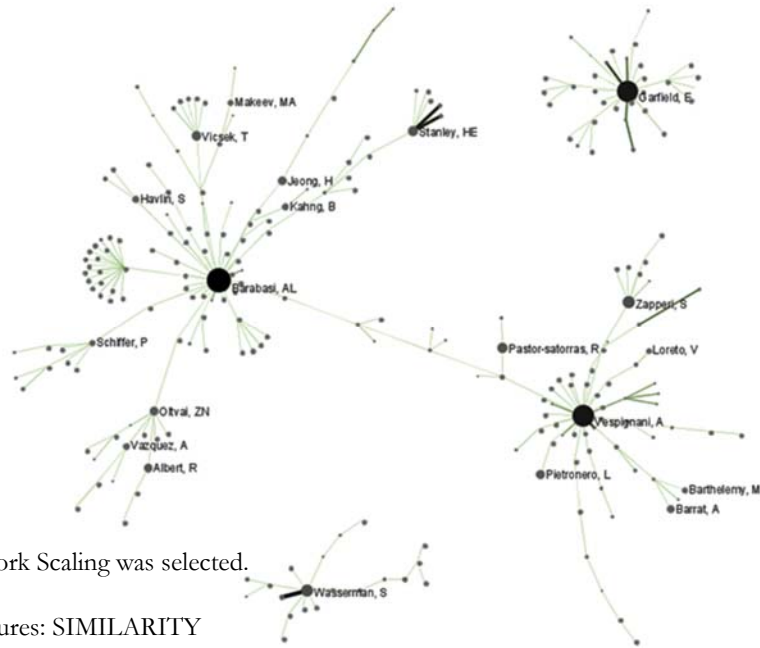
Input Parameters:
Number of bins: 10

umber of bins: 10
.....



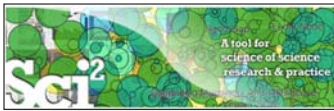


Network Visualization: Reduced Network After Pathfinder Network Scaling



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

53



Network Visualization: Circular Hierarchy Visualization

Select Co-Author Network and run Blondel Community detection:

File Data Preparation Preprocessing Analysis Modeling Visualization Help

Console

.....
Load and Clean ISI File was selected.
Author(s): Micah Linnemeier
Implementer(s): Micah Linnemeier
Integrator(s): Micah Linnemeier
Documentation:
<https://nwb.slis.indiana.edu/community/?n=LoadData.ISI>
Loaded:
C:\Users\User\Desktop\10-NEH-A&H-Workshop\DVD\sci2\sample
ometrics\isi\FourNetSciResearchers.isi
Loaded 361 records.
Removed 0 duplicate records.
Author names have been normalized.
361 records with unique ISI IDs are available via Data Manager.
Wrote log to
C:\Users\User\AppData\Local\Temp\isiduplicateremoverlog4773522398971021
378.txt

Networks

Network Analysis Toolkit (NAT)

- Unweighted & Undirected
- Weighted & Undirected
- Unweighted & Directed
- Weighted & Directed

Blondel Community Detection

With parameter values

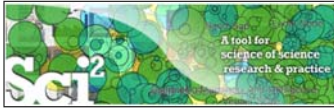
Blondel Community Detection

This algorithm implements Blondel's community detection algorithm.

Weight: numberofcoauthoredworks

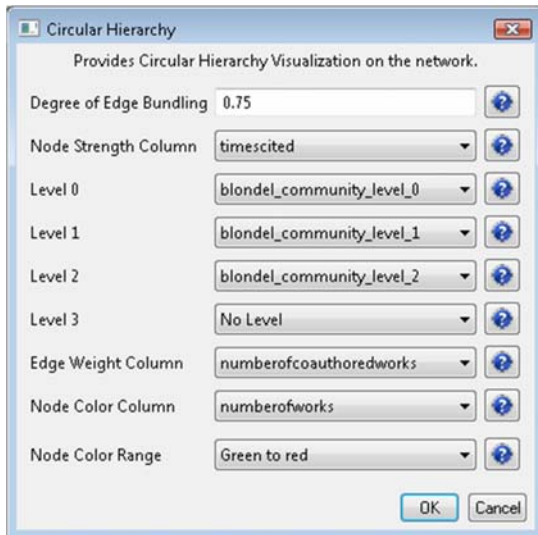
OK Cancel

54



Network Visualization: Circular Hierarchy Visualization

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



55



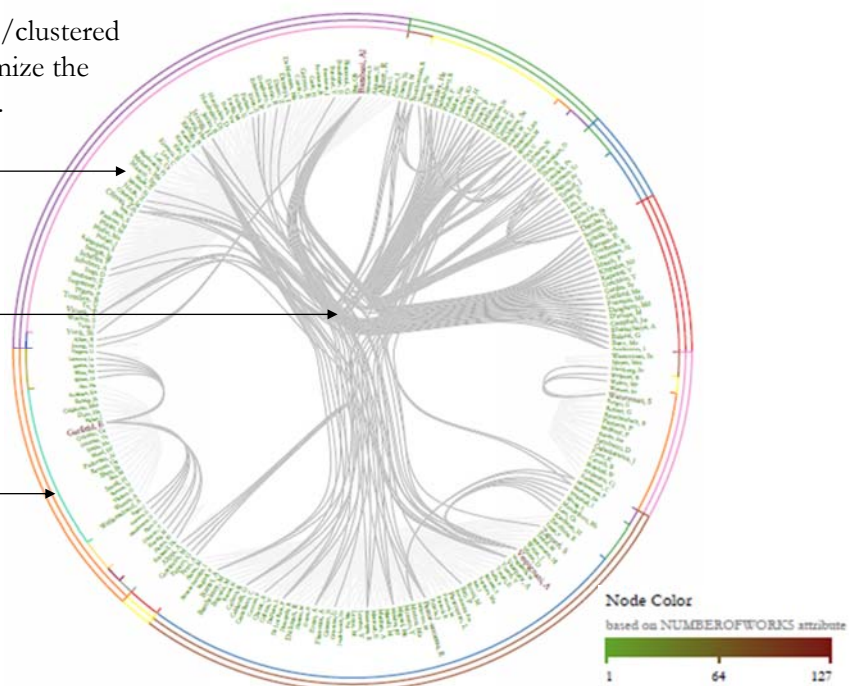
Network Visualization: Circular Hierarchy Visualization

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.

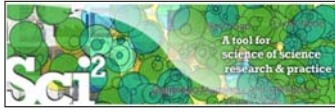
Color coded cluster hierarchy according to Blondel community detection algorithm.



Note:

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

56



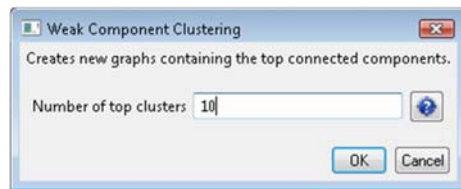
Paper-Citation Network Layout

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

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

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

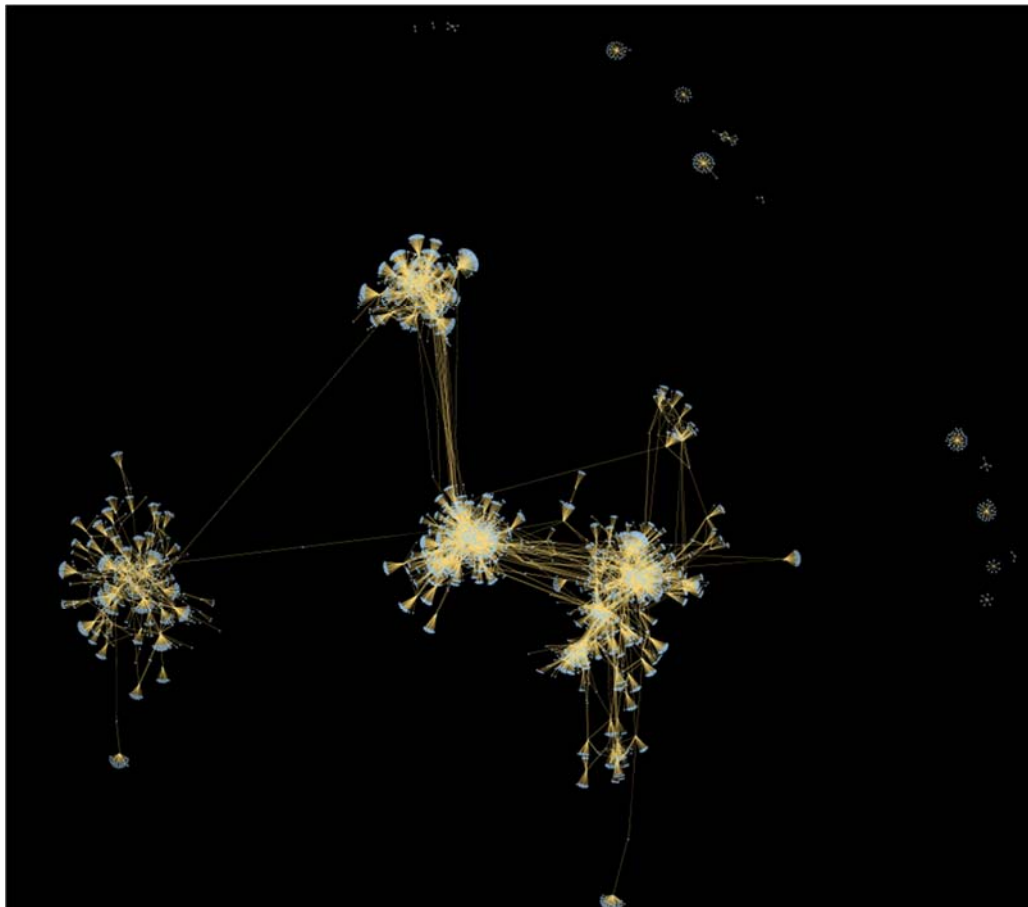
Run *Analysis > Networks > Unweighted and Directed > Weak Component Clustering* with parameters



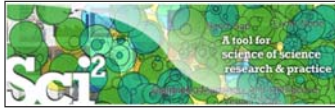
to identify top-10 largest components. The largest (giant) component has 5,151 nodes.

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

57



58

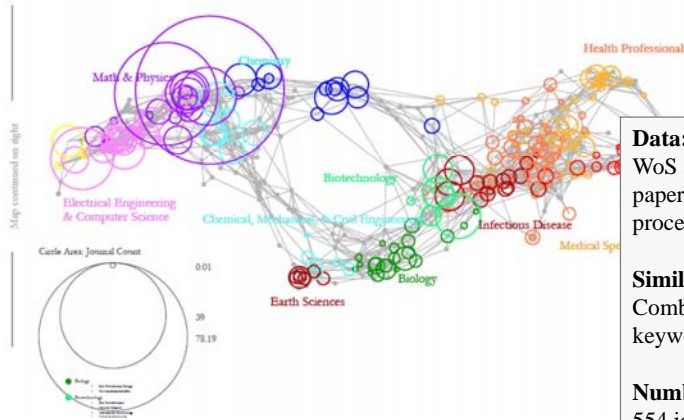


Topic Mapping: UCSD Science Map

Science Map via Journals for FourNetSciResearchers.isi

314 journal references matched out of 361 found.

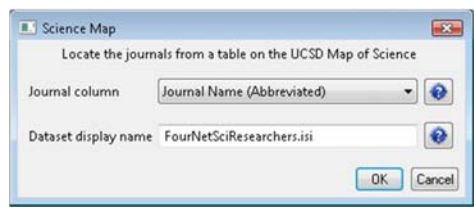
These 314 references are associated with 13 of 13 disciplines of science and 255 of 554 research specialties in the UCSD Map of Science.



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

Similarity Metric:
Combination of bibliographic coupling and keyword vectors

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

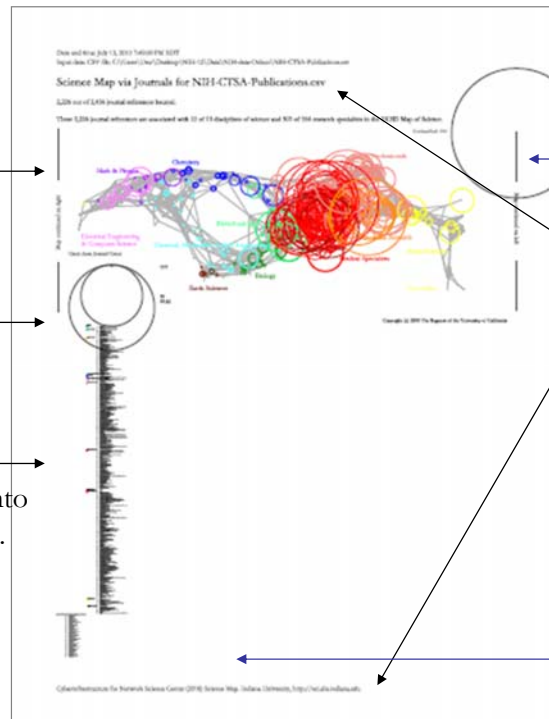


How to Read the UCSD Map

UCSD Science Map with data overlay.

Map legend of circle area size coding

Listing of all data records organized into UCSD science areas.



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.



General Network Extraction: Weighted, Undirected Co-Occurrence Network

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

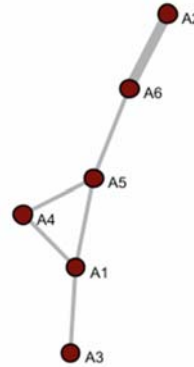


Author co-occurrence network

Extract Network from Table
Extracts a network from a delimited table

Column Name:

Text Delimiter:



*Vertices 6

1 A1

2 A6

3 A2

4 A3

5 A5

6 A4

*Edges 6

2 3 2

1 4 1

1 5 1

5 6 1

1 6 1

2 5 1

61



General Network Extraction: Unweighted, Directed Bipartite Network

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



Paper-author bipartite (2-mode) network

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

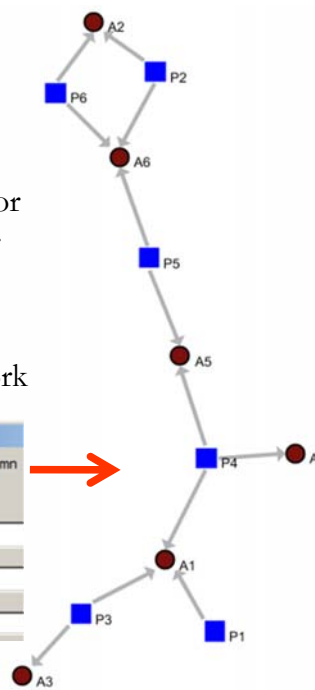
First column:

Second column:

Text Delimiter:



● Author
■ Paper



*Vertices 12

1 P1 bipartitetype "Paper"

2 A1 bipartitetype "Authors"

3 P2 bipartitetype "Paper"

4 A2 bipartitetype "Authors"

5 A6 bipartitetype "Authors"

6 P3 bipartitetype "Paper"

7 A3 bipartitetype "Authors"

8 P4 bipartitetype "Paper"

9 A4 bipartitetype "Authors"

10 A5 bipartitetype "Authors"

11 P5 bipartitetype "Paper"

12 P6 bipartitetype "Paper"

*Arcs

1 2

3 4

3 5

6 2

6 7

8 2

8 10

8 9

11 5

11 10

12 4

12 5

62



General Network Extraction: Unweighted, Directed Network

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

Extract Directed Network

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

Source Column: Paper

Target Column: Authors

Text Delimiter: ;

Analysis Modeling Visualization R Help

Temporal

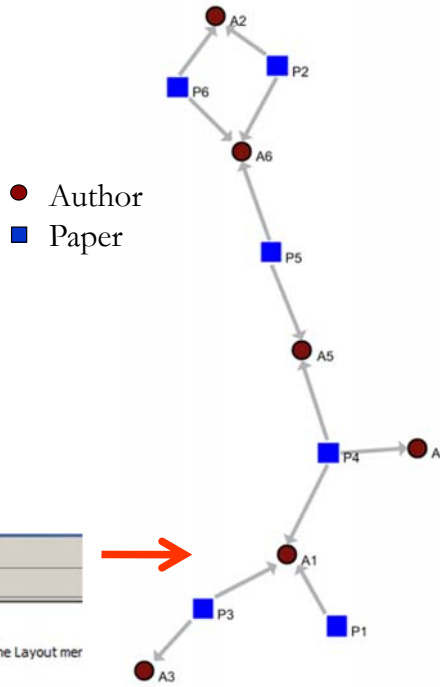
Geospatial

Topical

Networks

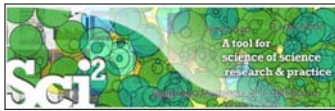
- Network Analysis Toolkit (NAT)
- Unweighted & Undirected
- Weighted & Undirected
- Unweighted & Directed**

Node Indegree



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

63



General Network Extraction: Unweighted, Directed Paper-Citation Network

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

Extract Directed Network

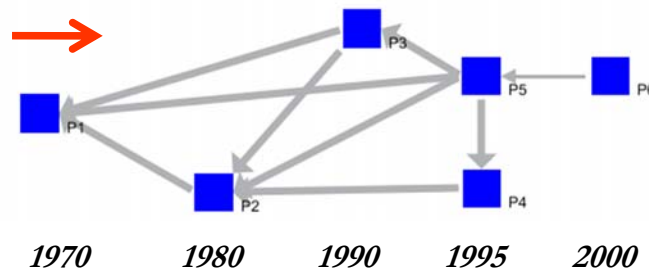
Given a table, this algorithm extracts a directed edge that starts at a column node.

Source Column: Paper

Target Column: References

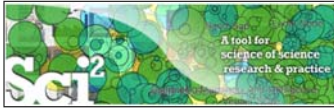
Text Delimiter: ;

Arcs from papers to references



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

64



General Network Extraction: Unweighted, Directed Bi-Partite Network

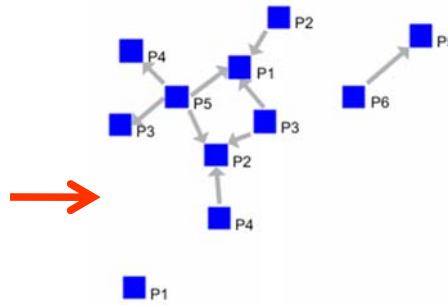
	A	B	C	D
1	Paper	Authors	References	Year
2	P1	A1		1970
3	P2	A2;A6	P1	1980
4	P3	A1;A3	P1;P2	1990
5	P4	A1;A4;A5	P2	1995
6	P5	A5;A6	P1;P2;P3;P4	1995
7	P6	A2;A6	P5	2000

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

First column:

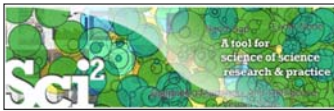
Second column:

Text Delimiter:



WRONG!!!

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



ISI Paper-Citation Network Extraction

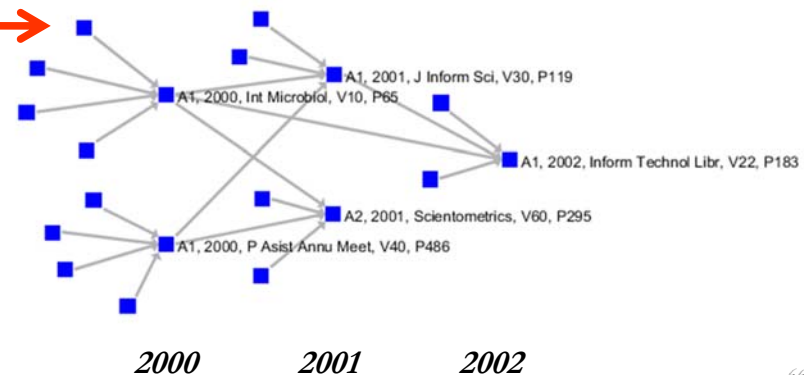
	A	B	C	D	E
1	Authors	Cited References	Publication Year	Title	Cite Me As
2	A1 A2	BENSMAN SJ, 1998, LIBR RESOUR TECH SER, V42, P147 BRO	2000	T1	A1, 2000, INT MICROBIOL, V10, P65
3	A1	BENSMAN SJ, 1999, LIBR RESOUR TECH SER, V42, P147 BRO	2000	T2	A1, 2000, P ASIST ANNU MEET, V40, P486
4	A2 A3	GARFIELD E,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

Sci2 Tool

File | Data Preparation | Preprocessing | Analysis

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

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



Break

67



Tutorial Overview

8:00a Welcome and Overview of Workshop and Attendees

8:15a Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools

8:30a 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

10:00 Break

10:15a Sci2 Tool Novel Functionality

- Horizontal line graph of NSF projects
- Geospatial maps aggregated by congressional districts
- Evolving collaboration networks
- R-Bridge

11:00a Outlook and Q&A

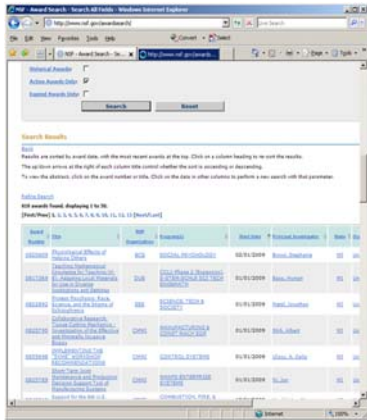
11:30 Adjourn

68

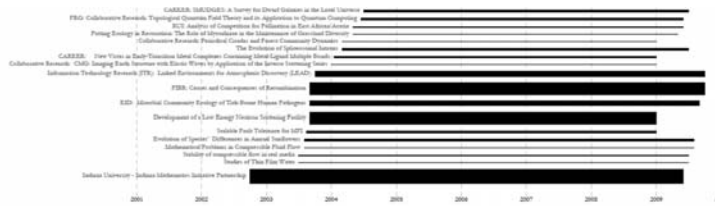
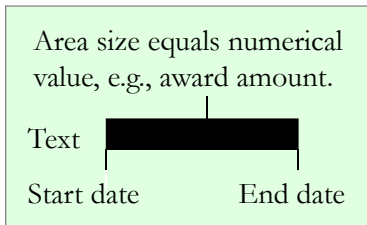
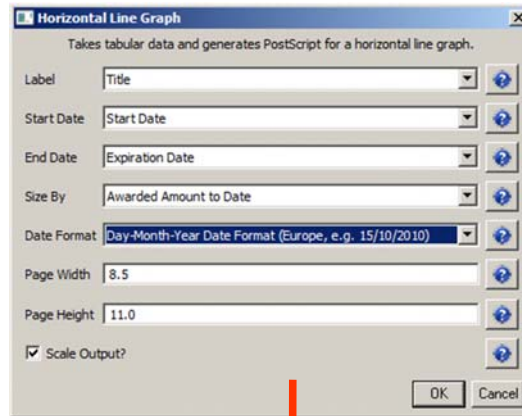


Horizontal line graph of NSF projects

Download NSF data



Visualize as Horizontal Line Graph

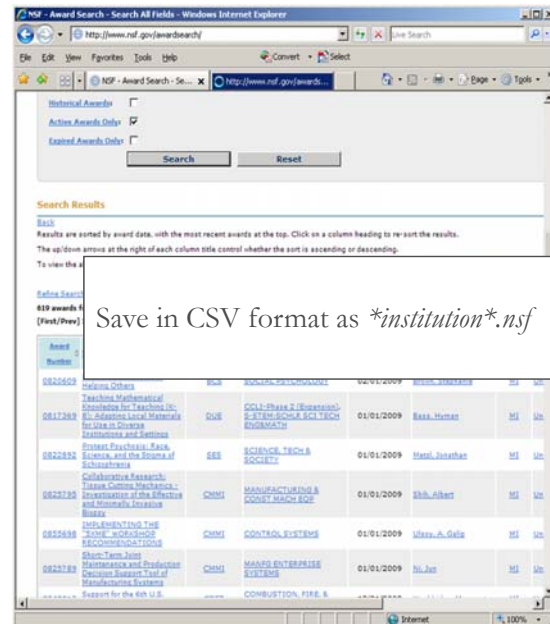
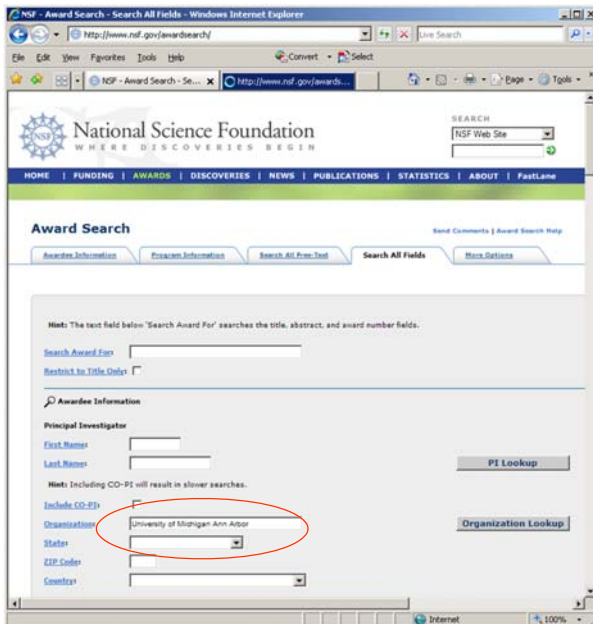


69



Horizontal line graph of NSF projects

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



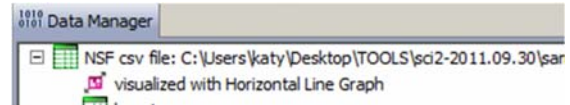
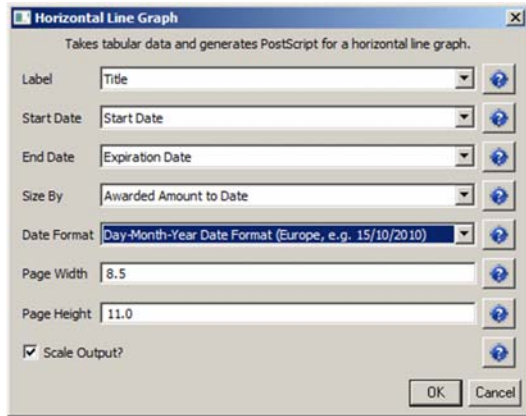
70



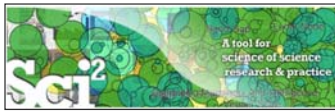
Horizontal line graph of NSF projects

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

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



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

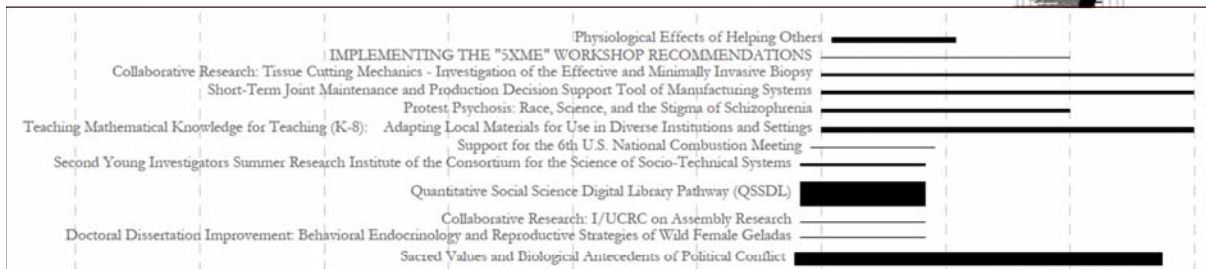
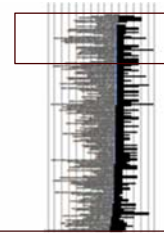


Horizontal line graph of NSF projects

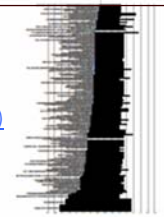
Area size equals numerical value, e.g., award amount.

Text, e.g., title

Start date End date



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





Tutorial Overview

8:00a Welcome and Overview of Workshop and Attendees

8:15a Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools

8:30a 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

10:00 Break

10:15a Sci2 Tool Novel Functionality

- Horizontal line graph of NSF projects
- **Geospatial maps aggregated by congressional districts**
- Evolving collaboration networks
- R-Bridge

11:00a Outlook and Q&A

11:30 Adjourn

73



Geospatial maps with congressional districts

	A
1	Zip code
2	90095
3	4672
4	232980568
5	10032
6	10039242
7	46091500
8	191112434
9	27705
10	981959472
11	10065
12	10065



Identify Congressional District, Latitude, Longitude

	A	B	C	D
1	Zip code	Congressional District	Latitude	Longitude
2	90095	CA-30	34.0735035	-118.6645815
3	4672	ME-02	45.818717	-69.0290345
4	232980568	VA-03	37.270472	-77.0699835



Aggregate/Count identical Congressional Districts

	A	B	C	D
1	Congressional District	Latitude	Longitude	Count
2	CA-30	34.0735035	-118.6645815	4
3	ME-02	45.818717	-69.0290345	2
4	VA-03	37.270472	-77.0699835	1
5	NY-15	40.8341475	-73.9342095	4



74



Relevant Sci2 Manual entry

- Home
- 1 Introduction
- 2 Getting Started
- 3 Algorithms, Tools, and Plugins
- 4 Workflow Design
- 5 Sample Workflows
 - 5.1 Individual Level Studies - Micro
 - 5.2 Institution Level Studies - Meso
 - 5.3 Global Level Studies - Macro
 - 5.3.1 Geo USPTO (SDB Data)
 - 5.3.2 Congressional District Geocoder
- 6 Sample Science Studies & Online Services
- 7 Extending the Sci2 Tool
- 8 Relevant Datasets and Tools
- 9 References
- Appendix 1 Glossary
- Appendix 2 CShell Algorithms
- Appendix 3 Sci2 Release Notes v0.5 alpha



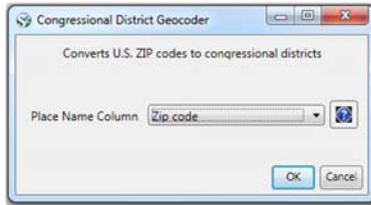
5.3.2 Congressional District Geocoder

14 Added by Scott Weingart, last edited by Ted Polley on Mar 28, 2011 (view change)

Tools ▾

zip code.csv	
Region(s):	United States
Analysis Type(s):	Geospatial Analysis

To visualize Congressional Districts you must first extract that data from a dataset containing either ZIP codes or addresses. You can download the Congressional District Geocoder plugin [here](#). You can load any file that contains 9-digit U.S. ZIP codes to be geocoded. A sample file can be loaded by using 'File > Load' and following this path: 'yoursci2directory/sampledata/geo/zipcode.csv'. Load the file in Standard csv format. Then select the file in the data manager and use 'Analysis > Geospatial > Congressional District Geocoder' with the following parameters:



5-digits ZIP codes with multiple congressional districts, empty entries and invalid ZIP codes that failed to be geocoded will list in warning messages on the console. The output table contains all columns of the input table with three additional columns appended: Congressional district, latitude, and longitude. To view the output table save the file using 'File > Save...' and selecting the desired save location (to view the file in Excel save it as a csv file). Once the file has been saved it can be viewed with your choice of program. Below the file has been opened as a csv file:

	A	B	C	D
1	Zip code	Congressional District	Latitude	Longitude
2	90095 CA-30		34.0735035	-118.6645815

<http://sci2.wiki.cns.in.edu/5.3.2+Congressional+District+Geocoder>

75



Relevant CShell plugin



Congressional District Geocoder

Tools ▾

1 Added by Ted Polley, last edited by Chin Hua Kong on Mar 29, 2011 (view change)

Description

This algorithm converts the given **9-digits U.S. ZIP codes (ZIP+4 codes)** into its congressional districts and geographical coordinates (latitude and longitude). The Benchmark is 50,000 ZIP codes per second. Download the plugin [here](#).

Pros & Cons

1. The algorithm is using a local database mapping with 25MB file size. It will increase the application size dramatically. So it is build as an external plugin
2. For first execution in the same application window, the plugin required 5 seconds to load the database. The consequent execution will not required the pre-loading phase.
3. Since some 5-digits ZIP codes contain multiple districts, the 9-digits ZIP codes is required for the conversion. Warning message will be printed to notice user if the given 5-digits ZIP codes contain multiple districts
4. Congressional district might be varied by each election. The database would need to be maintained and updated relatively.

Applications

This plugin only support U.S. ZIP codes. It convert 9-digits ZIP codes to their belonging congressional district. It is an external plugin since the data size is so large. The dataset is based on the year 2008 election.

<http://cishell.wiki.cns.in.edu/Congressional+District+Geocoder>

76



Console Messages

Load... was selected.

Documentation: <http://wiki.cns.iu.edu/display/CISHELL/Data+Formats>

Loaded: C:\Users\katy\Desktop\NWB-SCI2\sci2-2011.04.04-v0.5a\sampladata\geo\zip code.csv

.....

Congressional District Geocoder was selected.

Implementer(s): Chin Hua Kong

Integrator(s): Chin Hua Kong

Documentation: <https://nwb.cns.iu.edu/community/?n=SampleData.CongressionalDistrictGeocoder>

Input Parameters:

Place Name Column: Zip code

District values added to Congressional District, Latitude and Longitude respectively.

There are 2 rows with "33612" ZIP code, which could not been given a congressional district.

There are 1 rows with "2472" ZIP code, which could not been given a congressional district.

There are 3 rows with "10016" ZIP code, which could not been given a congressional district.

There are 1 rows with "11203" ZIP code, which could not been given a congressional district.

There are 1 rows with "60637" ZIP code, which could not been given a congressional district.

There are 1 rows with "70118" ZIP code, which could not been given a congressional district.

There are 1 rows with "60612" ZIP code, which could not been given a congressional district.

There are 3 rows with "21205" ZIP code, which could not been given a congressional district.

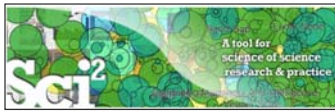
There are 1 rows with "2467" ZIP code, which could not been given a congressional district.

5-digit ZIP codes may often be insufficient, as many zip codes contain multiple congressional districts. 9-digit zip codes may be required. If a zip code was recently created, it may also not be contained in our database.

Successfully converted 86 out of 100 ZIP codes to congressional districts.

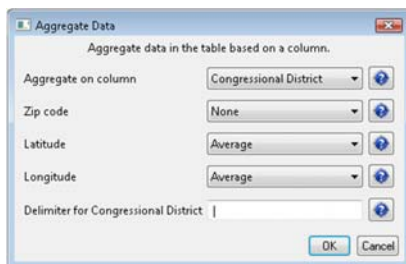
.....

77



Geospatial maps with congressional districts

Run *'Preprocessing > General > Aggregate Data'*
using parameter values



Note: Need lat/long for geomap.

Input Parameters:

Aggregate on column: Congressional District

Longitude: AVERAGE

Latitude: AVERAGE

Delimiter for Congressional District: |

Zip code: NONE

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 "Congressional District" values added to "Count" column.

"Zip code" column has been deleted from the output. Since No aggregation was mentioned for it.

78



Create Geo Map (Circle Annotation)

.....

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: CircleSize

Color Circle Exteriors By: None (no outer color)

Color Circle Interiors By: CircleSize

Exterior Color Scaling: Linear

Exterior Color Range: Yellow to Blue

Interior Color Range: Blue to Red

Size Scaling: Linear

Map: US States

Author Name:

Interior Color Scaling: Linear

Latitude: Latitude

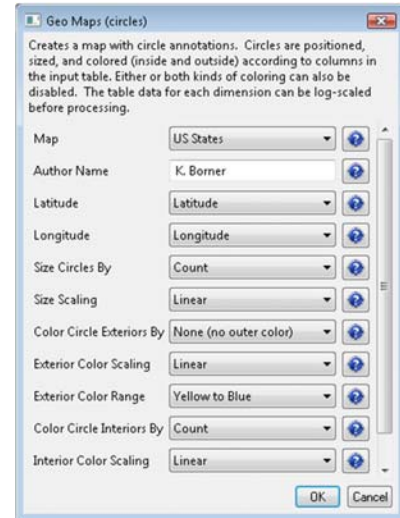
14 rows in the table did not specify all values needed to make a circle; those rows were skipped.

Printing PostScript..

Done.

Saved: C:\Users\katy\Desktop\geoMaps2903082942930990749.ps

Save ps file, convert to pdf, view.



79

How to Read the Geo Map

U.S. Map with data overlay.

Listing of map type, author, and parameters used.



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

Map legend with color coding.

80



Workshop Overview

8:00a Welcome and Overview of Workshop and Attendees

8:15a Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools

8:30a 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

10:00 Break

10:15a Sci2 Tool Novel Functionality

- Horizontal line graph of NSF projects
- Geospatial maps aggregated by congressional districts
- **Evolving collaboration networks**
- R-Bridge

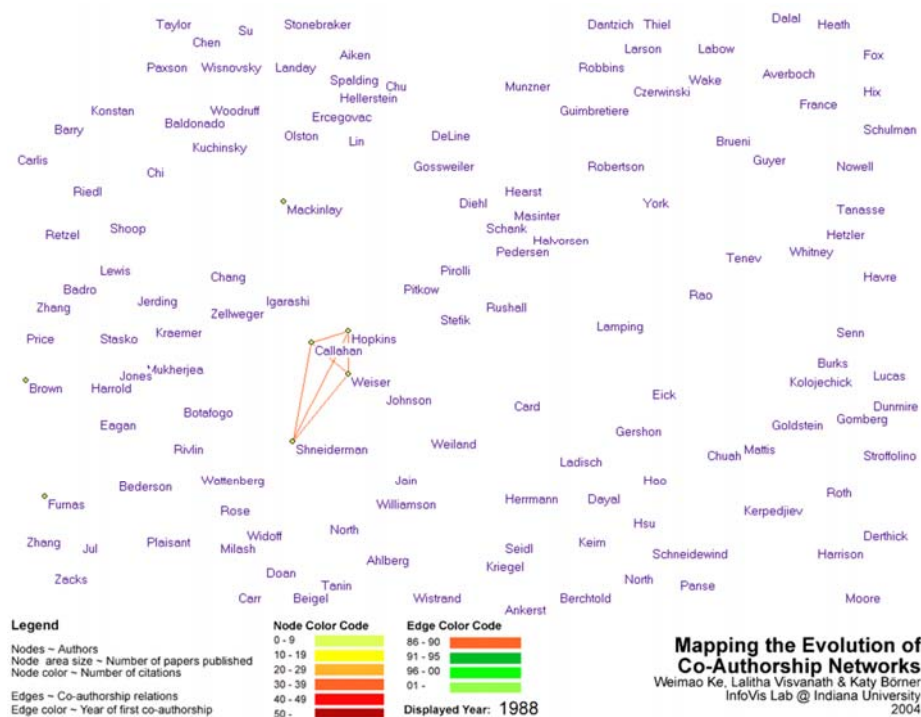
11:00a Outlook and Q&A

11:30 Adjournal

81



Sci2 Demo II: Evolving collaboration networks

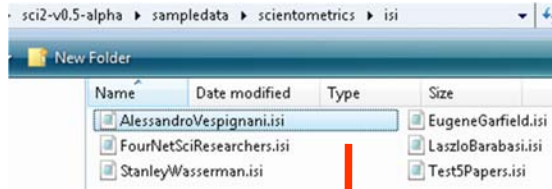


82



Sci2 Demo II: Evolving collaboration networks

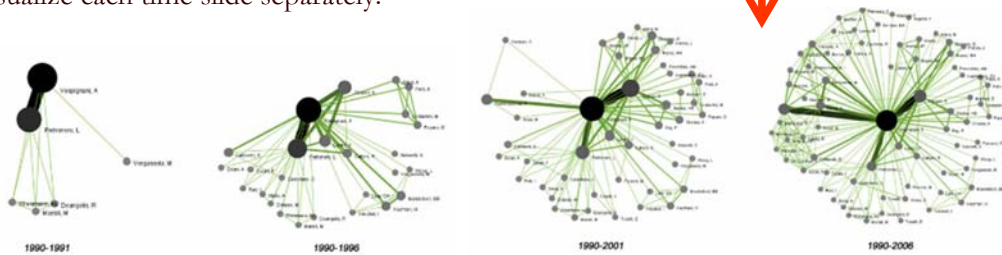
Load isi formatted file



As csv, file looks like:

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

Visualize each time slide separately:



83



Relevant Sci2 Manual entry

- Home
- 1 Introduction
- 2 Getting Started
- 3 Algorithms, Tools, and Plugins
- 4 Workflow Design
- 5 Sample Workflows
 - 5.1 Individual Level Studies - Micro
 - 5.1.1 Mapping Collaboration, Publication, and Funding Profiles of One Researcher (EndNote and NSF Data)
 - 5.1.2 Time Slicing of Co-Authorship Networks (ISI Data)
 - 5.1.3 Funding Profiles of Three Researchers at Indiana University (NSF Data)
 - 5.1.4 Studying Four Major NetSci Researchers (ISI Data)
 - 5.2 Institution Level Studies - Meso
 - 5.3 Global Level Studies - Macro
- 6 Sample Science Studies & Online Services
- 7 Extending the Sci2 Tool
- 8 Relevant Datasets and Tools
- 9 References

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

Tools ▾

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

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

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

Slice Table by Time

Slice a table into groups of rows by time.

Date/Time Column: Publication Year

Date/Time Format: yyyy

Slice Into: Years

How Many?: 5

From Time: 1990

To Time: 2006

Cumulative?

Align With Calendar

Week Starts On: Sunday

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

84



Slice Table by Time

Slice a table into groups of rows by time.

Date/Time Column: Publication Year

Date/Time Format: YYYY

Slice Into: Years

How Many?: 5

From Time: 1990

To Time: 2006

Cumulative?

Align With Calendar

Week Starts On: Sunday

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

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

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

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

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

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

85



Visualize Each Network, Keep Node Positions

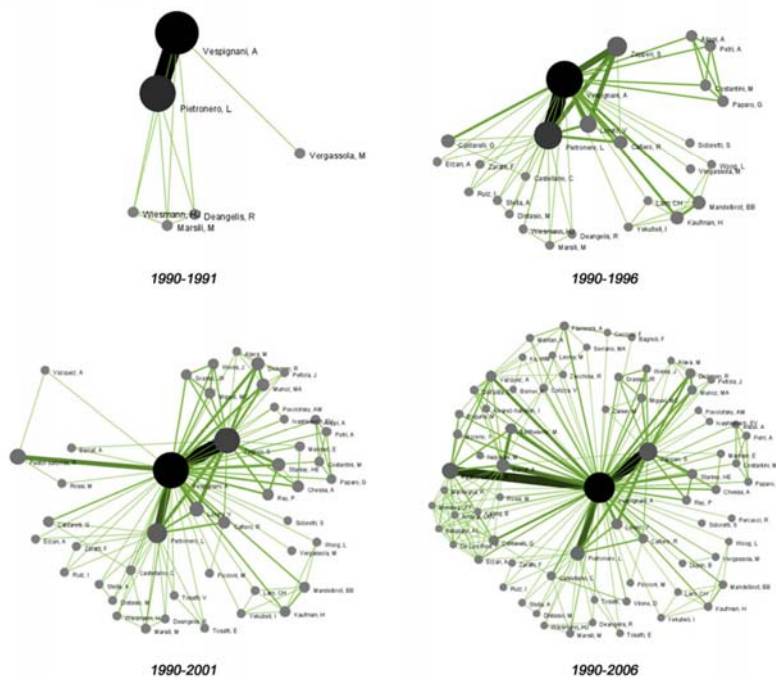
1. To see the evolution of Vespignani's co-authorship network over time, check *'cumulative'*.
2. Extract co-authorship networks one at a time for each sliced time table using *'Data Preparation > Extract Co-Author Network'*, making sure to select "ISI" from the pop-up window during the extraction.
3.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

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

86



Visualize Each Network, Keep Node Positions



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

87



Relevant CShell plugin



Tools ▾

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

Description

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

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

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

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

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

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

Pros & Cons

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

Applications

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

Implementation Details

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

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

88



Workshop Overview

8:00a Welcome and Overview of Workshop and Attendees

8:15a Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools

8:30a 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

10:00 Break

10:15a Sci2 Tool Novel Functionality

- Horizontal line graph of NSF projects
- Geospatial maps aggregated by congressional districts
- Evolving collaboration networks
- **R-Bridge**

11:00a Outlook and Q&A

11:30 Adjourn

89



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.

90



Workshop Overview

8:00a Welcome and Overview of Workshop and Attendees

8:15a Plug-and-Play Macroscopes, OSGi/CIShell Powered Tools

8:30a 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

10:00 Break

10:15a Sci2 Tool Novel Functionality

- Geospatial maps aggregated by congressional districts
- Evolving collaboration networks
- R-Bridge

11:00a Outlook and Q&A

11:30 Adjourn

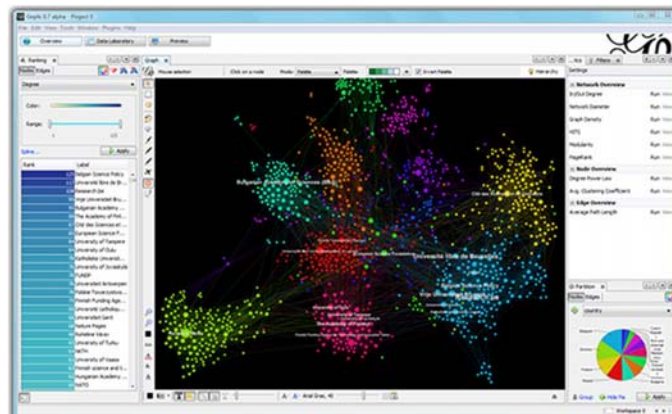
91



New Sci2 Release Preview

The next Sci2 release will become available in Dec 2011. It will add

- New ISI file format reader
- Publication and funding database support
- Improved geographic map legend
- R-bridge
- Gephi.org graph layout tool



92

Computational Scientometrics Cyberinfrastructures



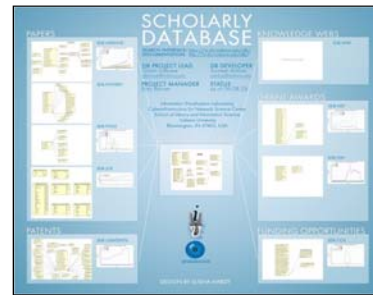
Scholarly Database: 25 million scholarly records

<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>

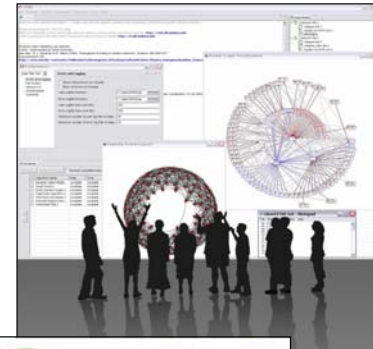


Science of Science (Sci²) Tool

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



Epidemics Tool & Marketplace
Forthcoming



93

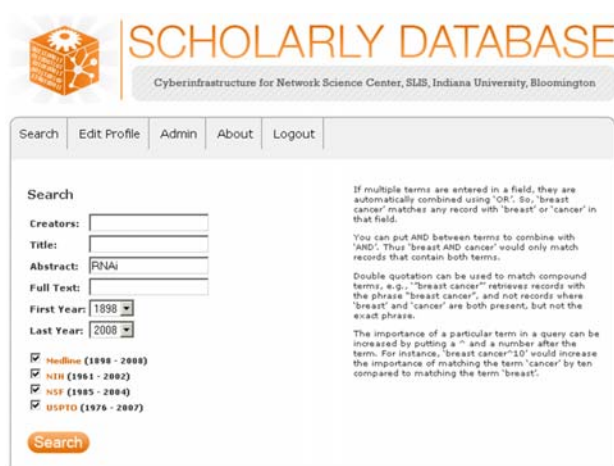


Scholarly Database at Indiana University

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

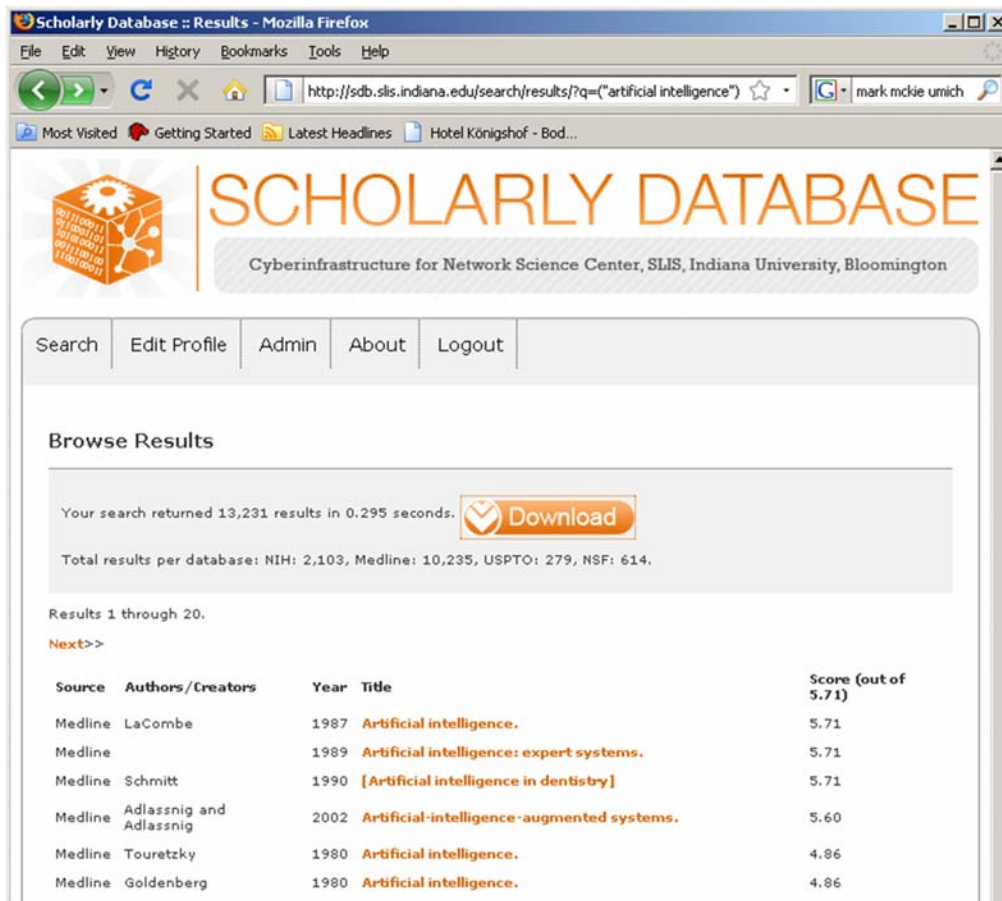
Supports federated search of 25 million publication, patent, grant records.

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

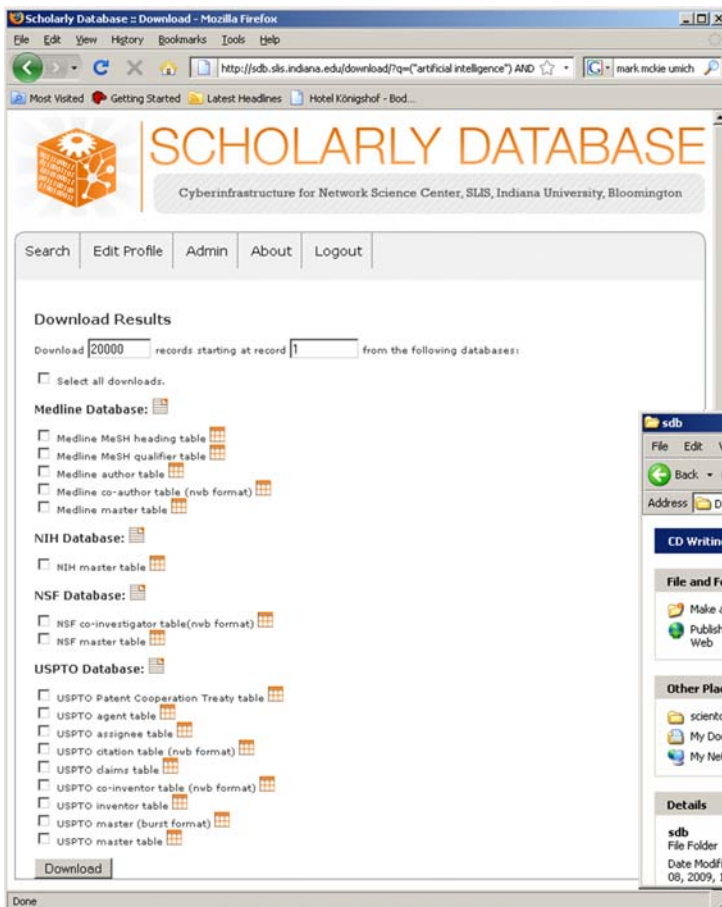


Register for free access at <http://sdb.cns.iu.edu>

94



95

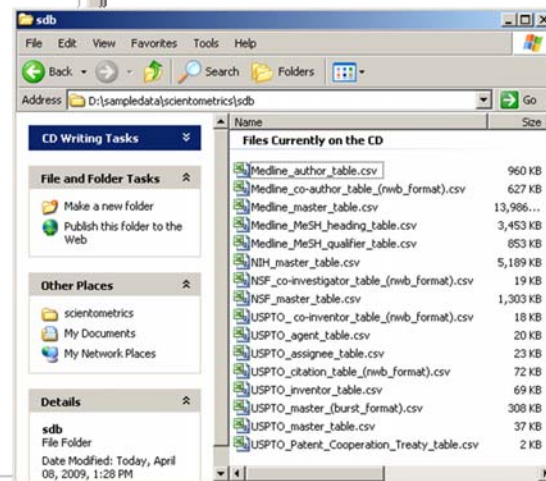


Since March 2009:

Users can download networks:

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

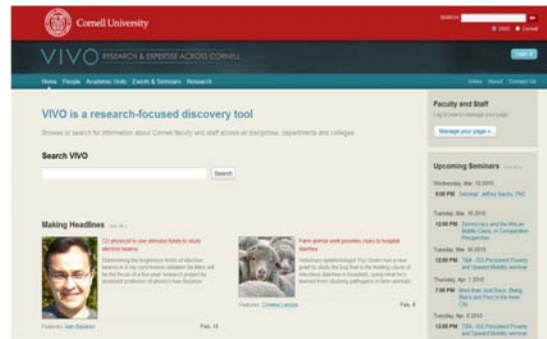
and tables for burst analysis in NWB.



96

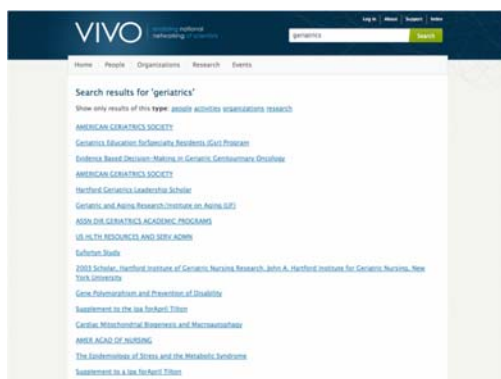
VIVO: A Semantic Approach to Creating a National Network of Researchers (<http://vivoweb.org>)

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



Funded by \$12 million NIH award.

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



University of Florida

How do you want to compare?
by Grants

Who do you want to compare?
Search: X

Records 1 - 10 of 30

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

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

Comparing Grants of Organizations in University of Florida

Total Number of Grants

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

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

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

[Index](#) [Log in](#)

[Search](#)

[Home](#) [People](#) [Organizations](#) [Research](#) [Events](#)

University of Florida

Explore 487 publications activity across 554 scientific sub-disciplines

13 Disciplines | 554 Sub-Disciplines

Search: X

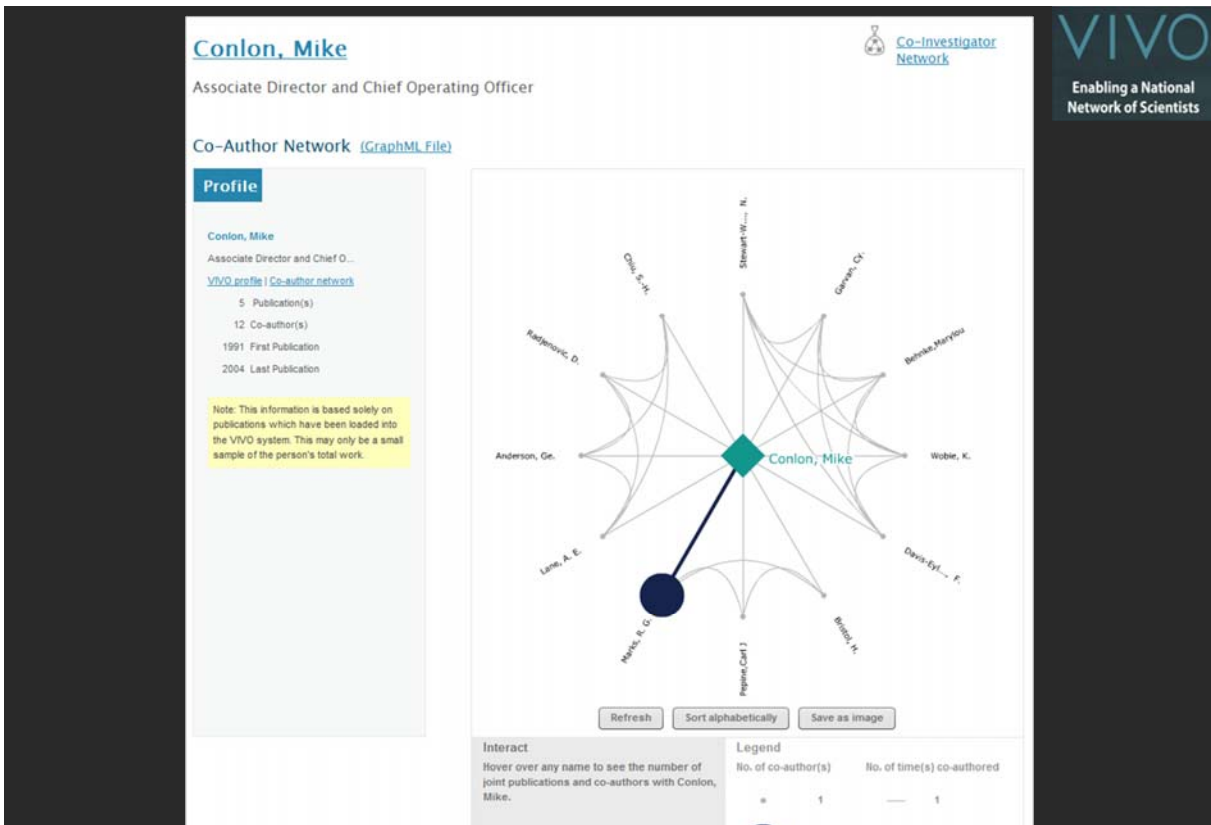
1 - 13 of 554

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

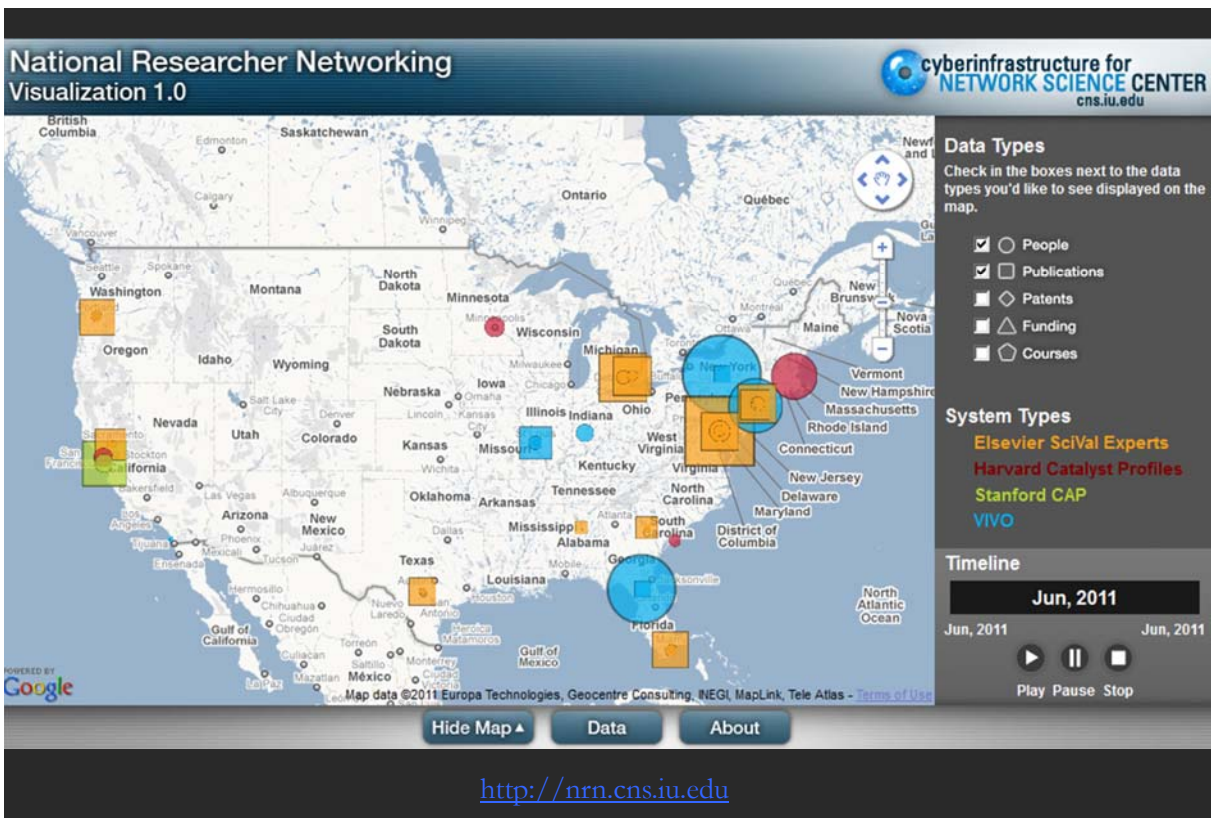
Top 290 disciplines shown

mapped 14.55% of 3,346 publications

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



Network Analysis (With Whom?) Who is co-authoring, co-investigating, co-inventing with whom? What teams are most productive in what projects? 101



Geospatial Analysis (Where) Where is what science performed by whom? Science is global and needs to be studied globally. 102

Borner, Katy
Person

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

General Statistics

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

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

Legend

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

Interact

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

Disambiguation

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

Tables

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

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

Co-author(s) [\(.CSV File\)](#)

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

Download Data

General Statistics

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

Co-Author Network

[\(GraphML File\)](#)

Save as Image (.PNG file)

Tables

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

http://vivo-netsci.cns.iu.edu/vivo/visualization?url=http%3A%2F%2Fvivo-trunk.indiana.edu%2Findividual%2FPerson74&vis=person_level&render_mode=standalone

103

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)

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

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

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.;

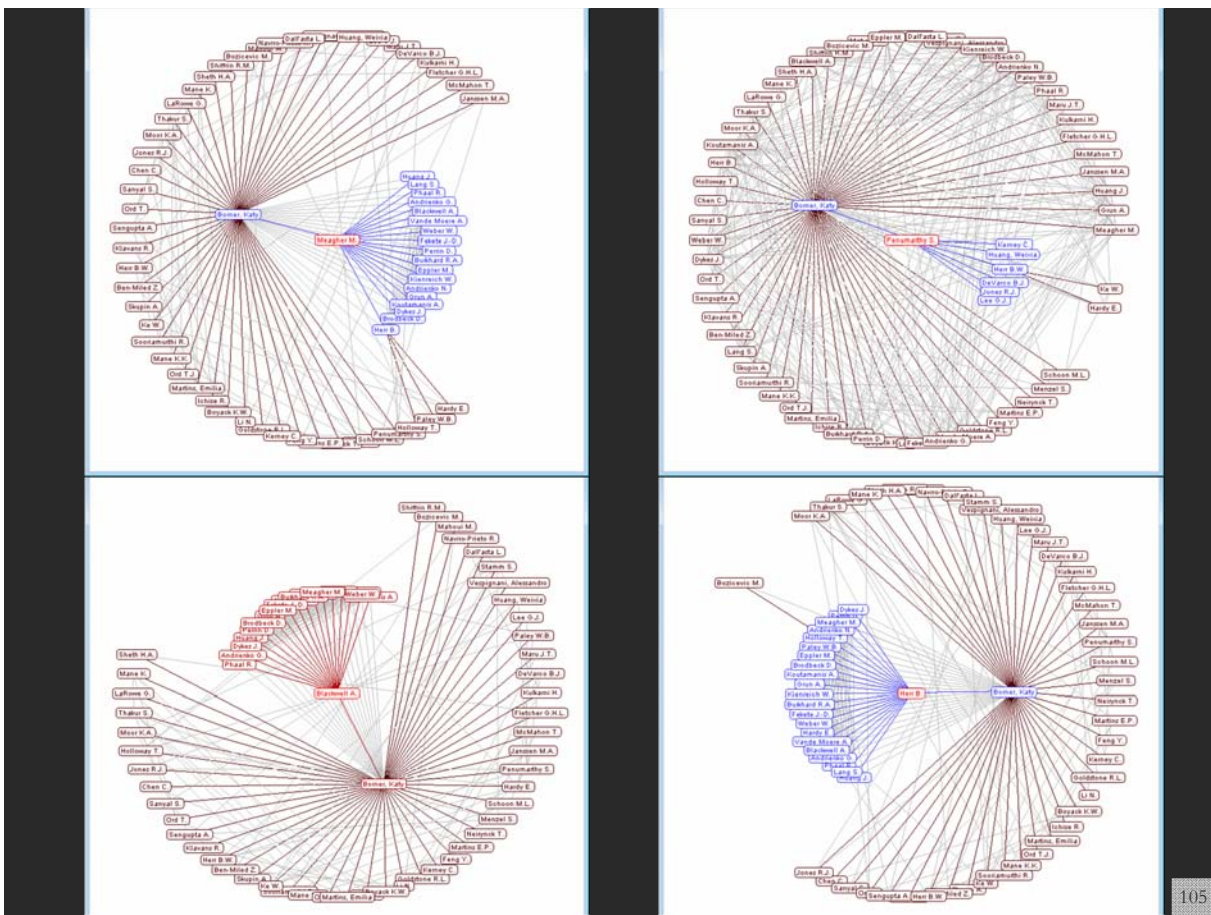
```

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" />

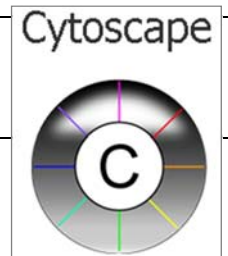
```

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

104



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

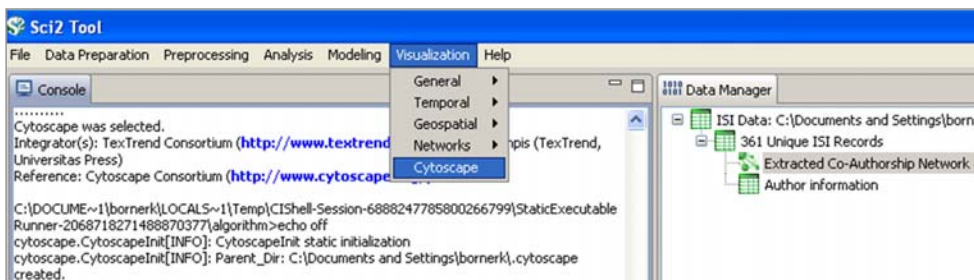


Adding more alyout algorithms and network visualization interactivity via Cytoscape <http://www.cytoscape.org>.

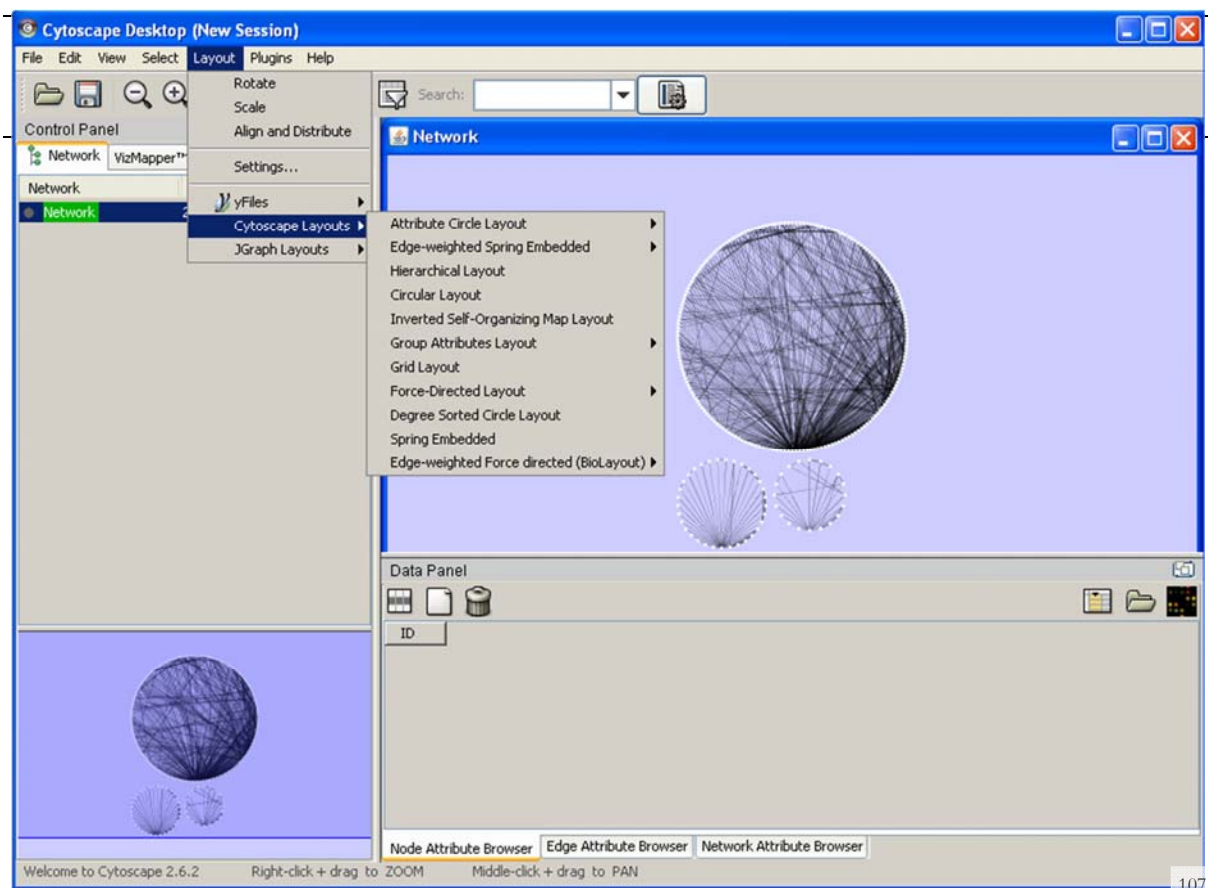
Simply add *org.textrend.visualization.cytoscape_0.0.3.jar* into your /plugin directory.

Restart Sci2 Tool.

Cytoscape now shows in the Visualization Menu.



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



OSGi/CIShell Adoption

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

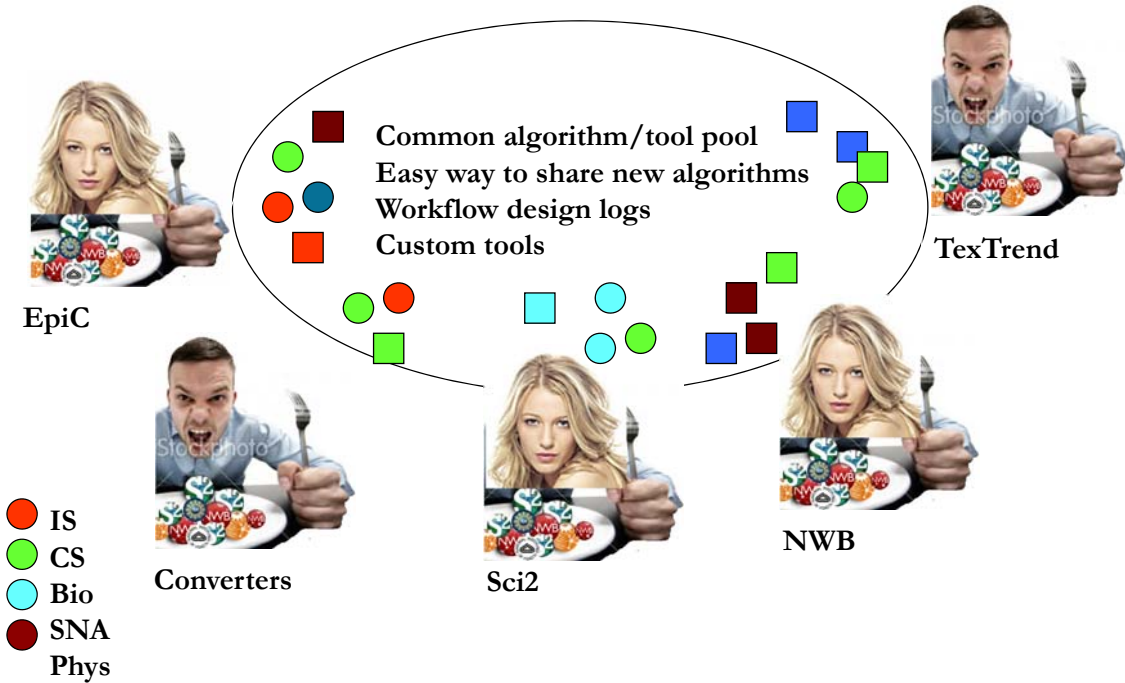
USA

- *Cytoscape* (<http://cytoscape.org>) Led by Trey Ideker at the University of California, San Diego is an open source bioinformatics software platform for visualizing molecular interaction networks and integrating these interactions with gene expression profiles and other state data (Shannon et al., 2002).
- *MAEviz* (<https://wiki.ncsa.uiuc.edu/display/MAE/Home>) Managed by Jong Lee at NCSA is an open-source, extensible software platform which supports seismic risk assessment based on the Mid-America Earthquake (MAE) Center research.

Europe

- *Taverna Workbench* (<http://taverna.org.uk>) Developed by the myGrid team (<http://mygrid.org.uk>) led by Carol Goble at the University of Manchester, U.K. is a free software tool for designing and executing workflows (Hull et al., 2006). Taverna allows users to integrate many different software tools, including over 30,000 web services.
- *TEXTrend* (<http://texttrend.org>) Led by George Kampis at Eötvös Loránd University, Budapest, Hungary supports natural language processing (NLP), classification/mining, and graph algorithms for the analysis of business and governmental text corpuses with an inherently temporal component.
- *DynaNets* (<http://www.dynanets.org>) Coordinated by Peter M.A. Sloot at the University of Amsterdam, The Netherlands develops algorithms to study evolving networks.
- *SISOB* (<http://sisob.lcc.uma.es>) An Observatory for Science in Society Based in Social Models.

As the functionality of OSGi-based software frameworks improves and the number and diversity of dataset and algorithm plugins increases, the capabilities of custom tools will expand.



Network Workbench Tool

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

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

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

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

It has been downloaded more than 65,000 times since December 2006.

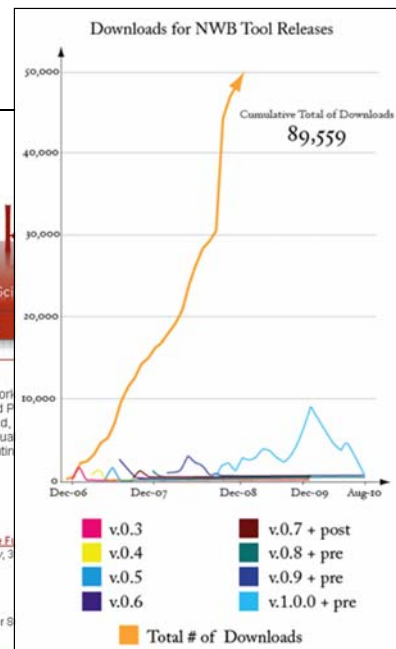
Summary
 Network Workbench: A Large-Scale Network Toolkit for Biomedical, Social Science and Physics. Evaluate, and operate a unique distributed, scale network analysis, modeling, and visualization (NWB). The envisioned data-code-computer more
[How to cite this project](#)

News & Updates

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

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

Get Involved



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

Computational Economics

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

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

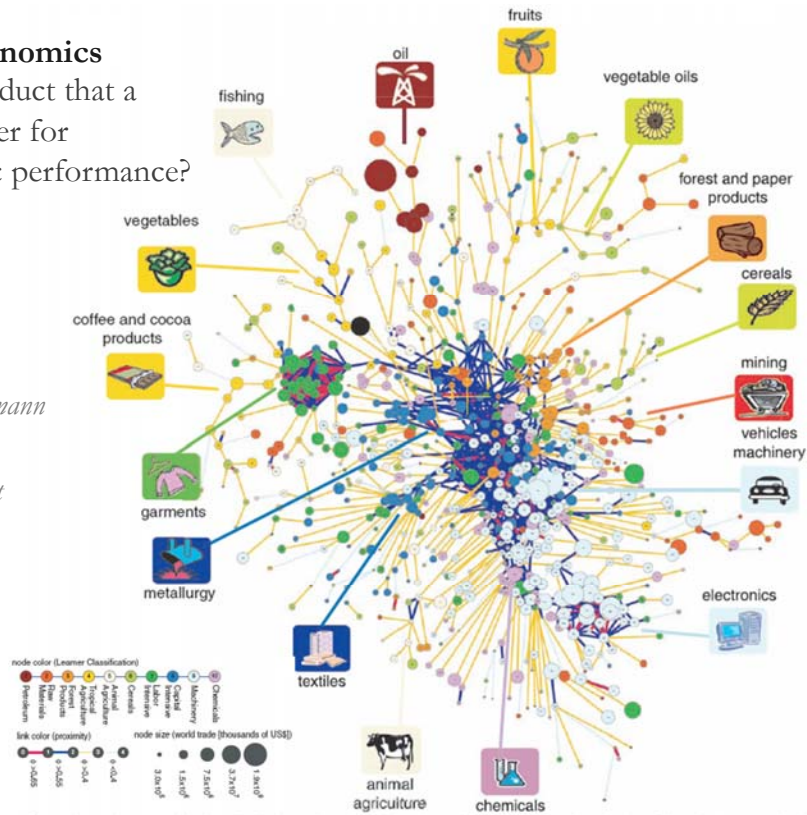


Fig. 1. The product space. (A) Hierarchically clustered proximity matrix representing the 775 SITC-4 product classes exported in the 1998–2000 period. (B) Network representation of the product space. Links are color coded with their proximity value. The sizes of the nodes are proportional to world trade,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

111

Computational Social Science

Studying large scale social networks such as Wikipedia

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



Second sight

Image: Bruce W. Hest and Todd M. Holloway

Power struggle

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



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

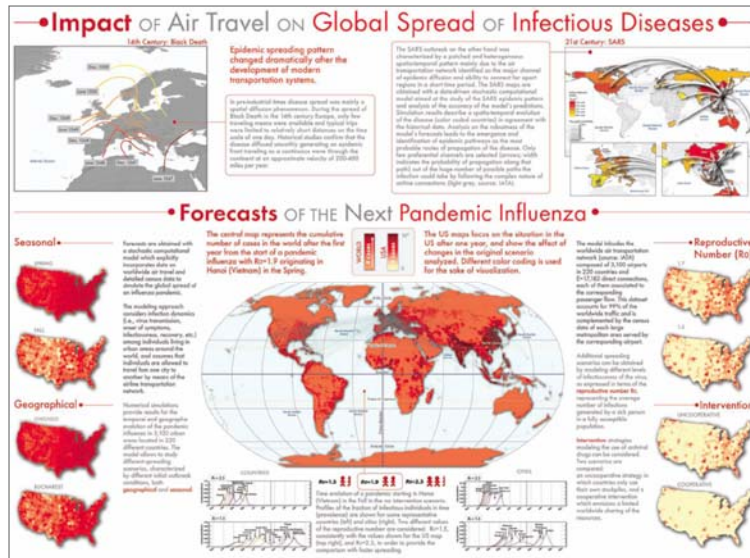
Computational Epidemics

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

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

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

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



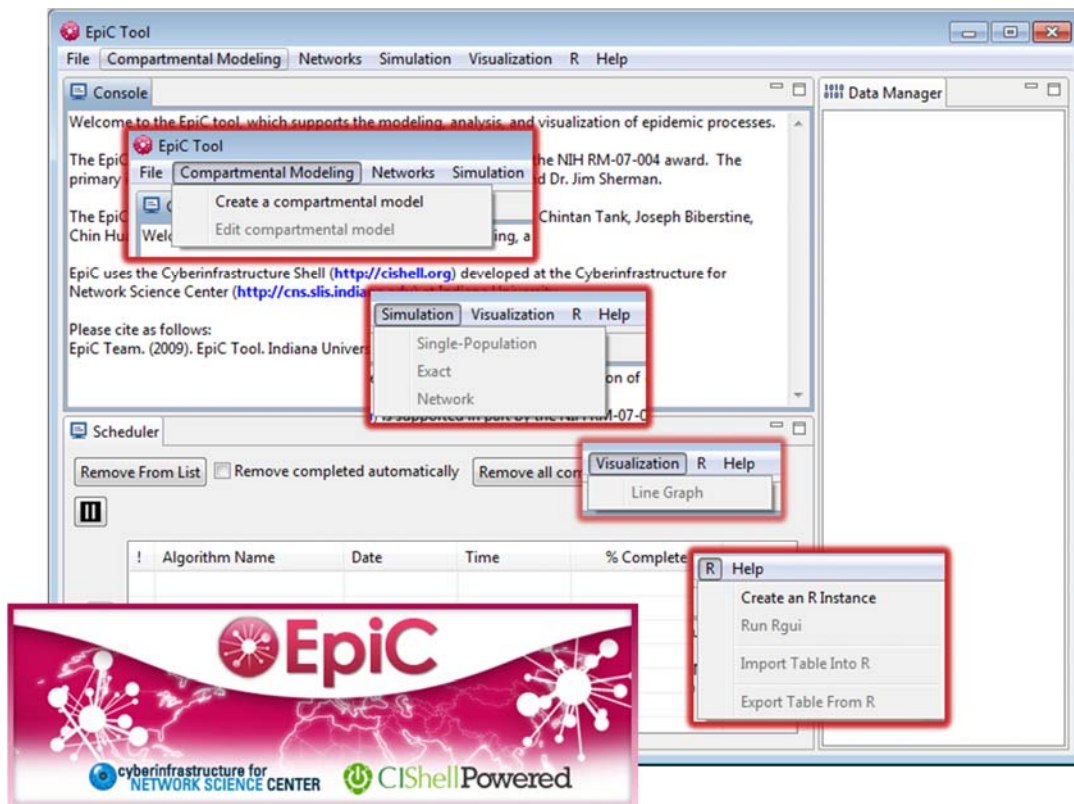
Computational Proteomics

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

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



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



115

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

116



CIShell – Integrate New Algorithms

About the Cyberinfrastructure Shell

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

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

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

CIShell Features

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

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

Learn More...

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

Getting Started...

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

Getting Involved...

- [Contact Us](#)

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

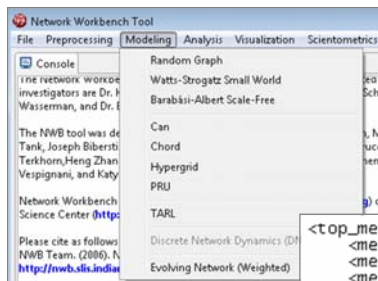
Additional Sci2 Plugins are at <http://sci2.wiki.cns.iu.edu/3.2+Additional+Plugins>

117



CIShell – Customize Menu

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



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

118



Need Help? Ask an Expert!

Sci² Tool
A Tool for Science of Science Research & Practice

Home Download Documentation **Ask An Expert** Testimonials Developers

Ask An Expert

(If you need to report a bug for the Sci² tool instead, [click here.](#))

Project Title *Pick any name to help us to refer to this project/question in the future.*

Types of Analyses

- Temporal (When) ?
- Geospatial (Where) ?
- Topical (What) ?
- Modeling (Why) ?
- Networks (With Whom?) ?

Levels of Analyses

- Micro/Individual (1-100 records) ?
- Meso/Local (101-10,000 records) ?
- Macro/Global (> 10,000 records) ?

Intended Users *Who is the intended audience?*
Who is interested in the result?

Insight Needed *What would you/user like to understand?*

[View sample questions HERE](#)
(Will open in new tab.)

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

119

CTSI Accelerating Science Core—Services Offered

Evaluation & Monitoring: Impact and/or strength analysis for a lab, center (e.g., NSF STCs or NIH CTSA), institution, or region in order to evaluate, plan, or implement research efforts. Relevant data must be provided.

\$2000-\$6000*

Data Compilation: The construction of a custom data set (e.g., all papers, patents, grants for a certain institution or area of research) using the Scholarly Database (<http://sdb.cns.iu.edu>).

\$3000*

Visual Interface to Community Data: Setting up an online interactive interface similar to <http://mapsustain.cns.iu.edu> (relevant data must be provided to the Core).

\$6000*

Accelerating Science Core—Services Offered

VIVO Researcher Networking: Design and implementation of interactive custom VIVO visualizations (see <http://vivoweb.org> for more info on VIVO and <http://vivo.iu.edu/vis/map-of-science/BL-ARSC> for a map of science visualization).

\$5000*

Training and Consulting in Data Mining and Visualization:

Introduction of advanced data mining and visualization tools: 2-hour tutorials or 4-hour tutorials that also feature exemplary analyses of client data. See <http://sci2.cns.iu.edu/user/documentation.php> for sample slides.

\$2000-\$5000*

**Prices vary according to scope of work. Please contact us to request an estimate.*

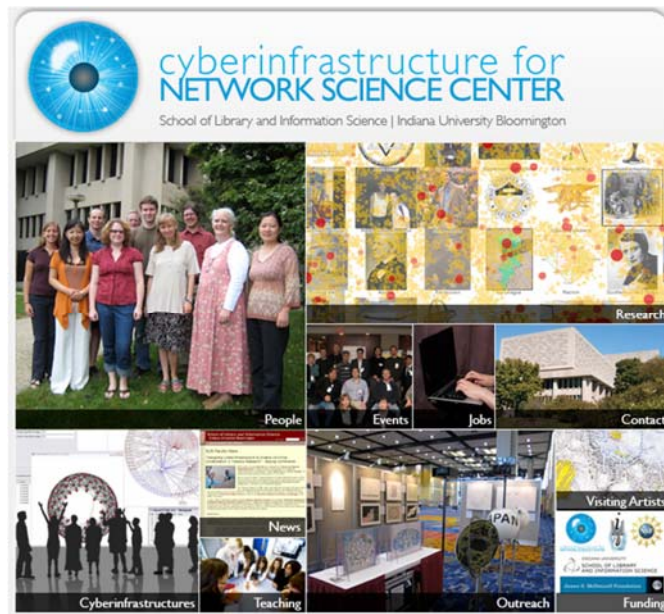
121

Q & A

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

Bug reports and all comments are welcome.

122



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>