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

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

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

Please
- download the Sci2 Tool from http://sci2.cns.iu.edu and
- complete the Pre-Tutorial Questionnaire

Portfolio Analysis Symposium
NIH Natcher Auditorium, Bethesda, MD

Tuesday July 24, 2012 • 2-3:30pm

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

➢ These slides
  http://ivl.slis.indiana.edu/km/pres/2012-borner-sci2tutorial-nih.pdf
➢ Sci2 Tool Manual v0.5.1 Alpha, updated to match v1.0 Alpha tool release
  http://sci2.wiki.cns.iu.edu
➢ Sci2 Tool v1.0 Alpha (June 13, 2012)
  http://sci2.cns.iu.edu

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

Or copy them from the DVD or memory stick.

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

Tutorial Overview

2p Welcome and Overview of Tutorial and Attendees

2:15p Sci2 Tool Hands-on
➤ **Download and run the Sci2 Tool**
➤ **Temporal Analysis: Horizontal line graph of NSF projects**
➤ **Geospatial Analysis: US and world maps**
➤ **Geospatial Analysis: Geomap with network overlays**
➤ **Topical Analysis: Visualize research profiles**
➤ **Network Analysis: Co-occurrence networks and bimodal networks**
➤ **Network Analysis: Evolving collaboration networks**
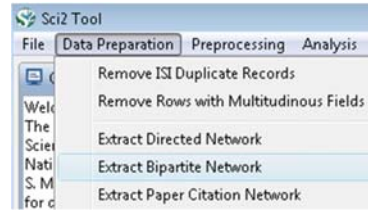
3:15p Outlook and Q&A

3:30p Adjourn

Tutorial Overview

2p Welcome and Overview of Tutorial and Attendees

2:15p Sci2 Tool Hands-on
➤ **Download and run the Sci2 Tool**
➤ **Temporal Analysis: Horizontal line graph of NSF projects**
➤ **Geospatial Analysis: US and world maps**
➤ **Geospatial Analysis: Geomap with network overlays**
➤ **Topical Analysis: Visualize research profiles**
➤ **Network Analysis: Co-occurrence networks and bimodal networks**
➤ **Network Analysis: Evolving collaboration networks**

3:15p Outlook and Q&A

3:30p Adjourn

Use *'File > Read'* to load cleaned *Sci2TutorialRegistrants.csv* with 41 records

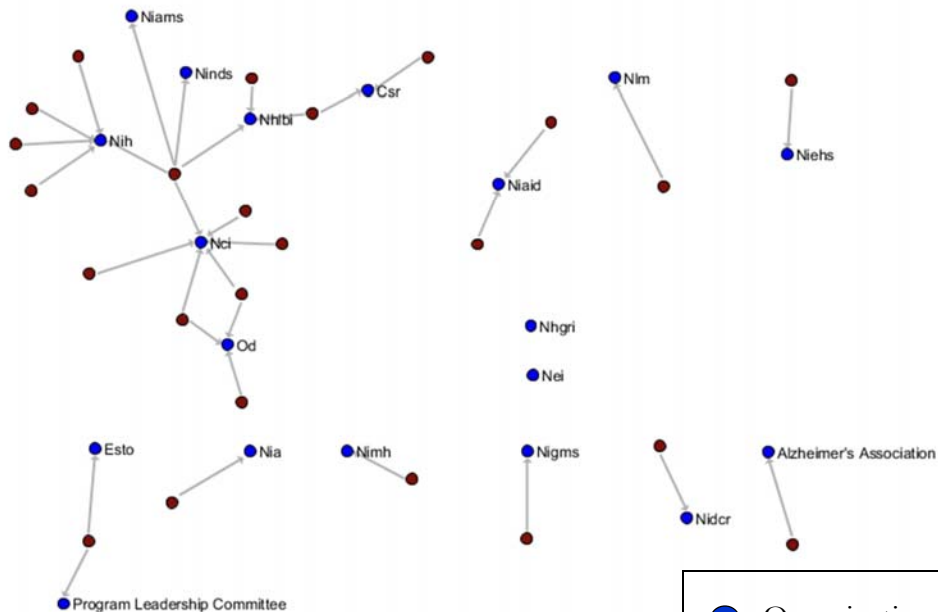| E	F
Organization	**Title**
NIH	Acting Director
NCI	Program Manager
CSR	Director
OD;NCI	Chief, Knowledge Management and Special Projects Branch
NHLBI	Program Director
NIH	Program Director
NIH	

Run *'Data Preparation > Extract Bipartite Network'*

With parameter values:
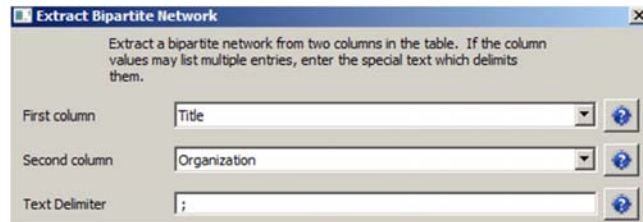
Visualize resulting *'Bipartite network from Affiliation to Background'* using *'Visualization > Network > GUESS'* and *'Layout > GEM', 'Layout > Bin Pack'*
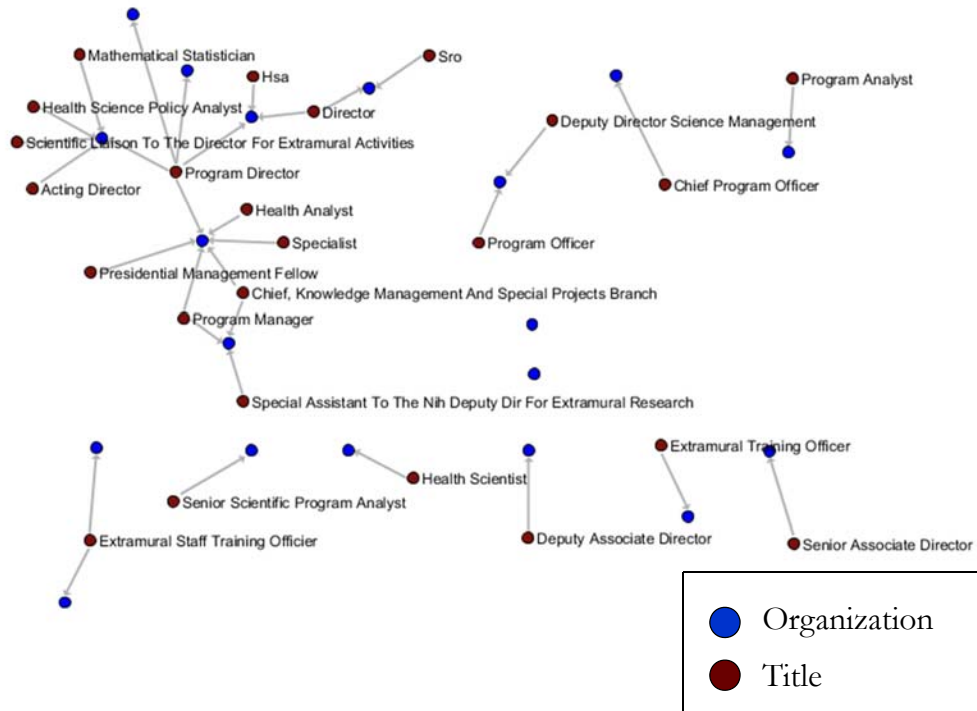
5

● Organization
● Title

6

Mathematical Statistician

Sro

Hsa

Program Analyst

Health Science Policy Analyst

Director

Deputy Director Science Management

Scientific Liaison To The Director For Extramural Activities

Acting Director

Program Director

Chief Program Officer

Health Analyst

Specialist

Program Officer

Presidential Management Fellow

Chief, Knowledge Management And Special Projects Branch

Program Manager

Special Assistant To The Nih Deputy Dir For Extramural Research

Extramural Training Officer

Senior Scientific Program Analyst

Health Scientist

Extramural Staff Training Officier

Deputy Associate Director

Senior Associate Director

- ● Organization
- ● Title

7

Tutorial Overview

2p Welcome and Overview of Tutorial and Attendees

2:15p Sci2 Tool Hands-on
➤ **Download and run the Sci2 Tool**
➤ Temporal Analysis: Horizontal line graph of NSF projects
➤ Geospatial Analysis: US and world maps
➤ Geospatial Analysis: Geomap with network overlays
➤ Topical Analysis: Visualize research profiles
➤ Network Analysis: Co-occurrence networks and bimodal networks
➤ Network Analysis: Evolving collaboration networks

3:15p Outlook and Q&A

3:30p Adjourn

8

Science of Science (Sci2) Tool

- 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.

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

nature

OPINION

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**.

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

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

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

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

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

Major Release

featuring a Web services compatible CIShell v2.0 (http://cishell.org)

New Features

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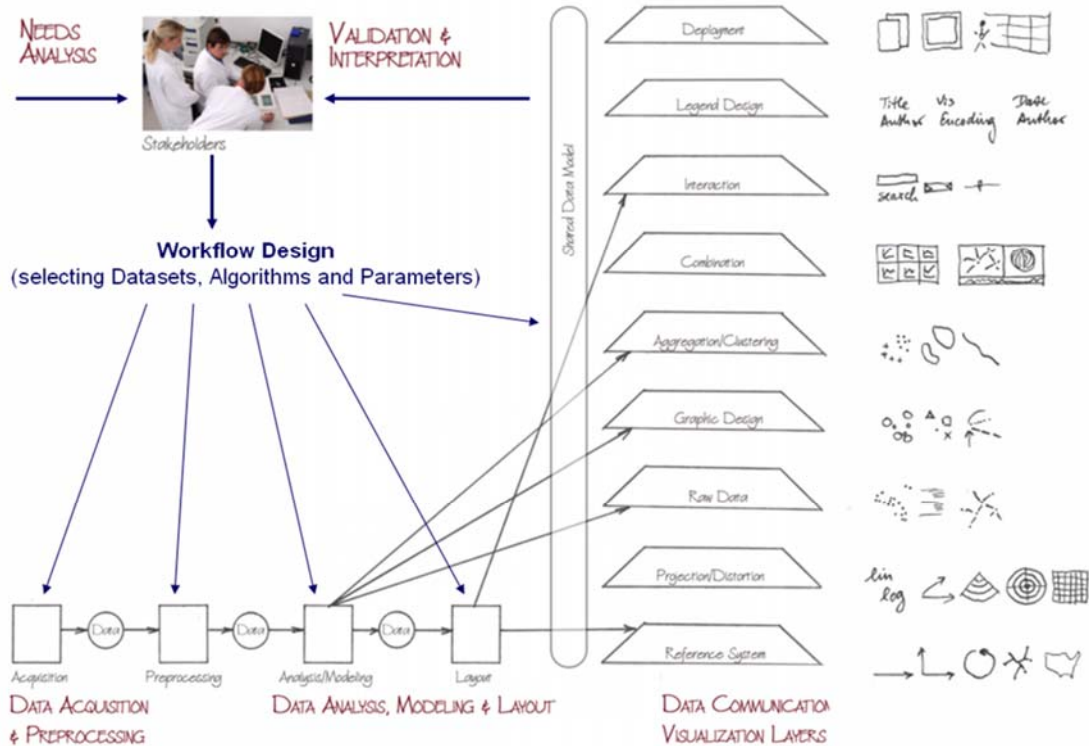- Comprehensive online documentation

Release Note Details

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

11

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 or states	All of NSF, all of USA, all of science.
Temporal Analysis (When)	Funding portfolio of one individual		Physics
Geospatial Analysis (Where)	Career trajectory of individual		
Topical Analysis (What)			Ord/ H funding
Network Analysis (With Whom?)	NSF Co-PI network of one individual	Co-auth	H's core competency

12

Needs-Driven Workflow Design using a modular data acquisition/analysis/ modeling/ visualization pipeline as well as modular visualization layers.

13

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.

14
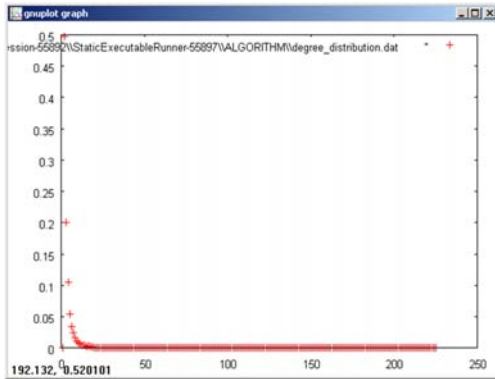
Gnuplot

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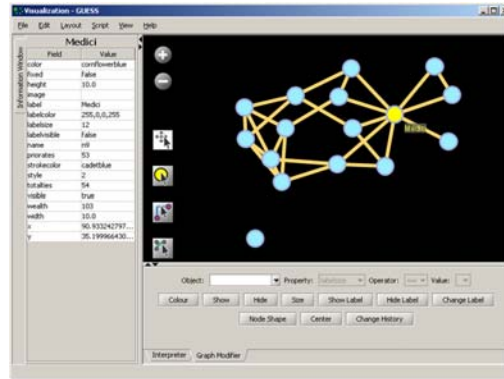
GUESS

exploratory data analysis and visualization tool for graphs and networks.

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

15
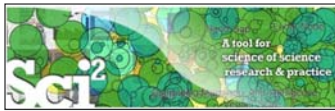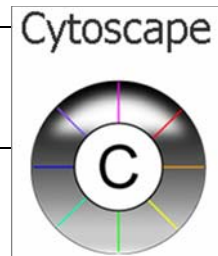
Adding more layout algorithms and network visualization interactivity via Cytoscape http://www.cytoscape.org.

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

Restart Sci² Tool

Cytoscape now shows in the Visualization Menu
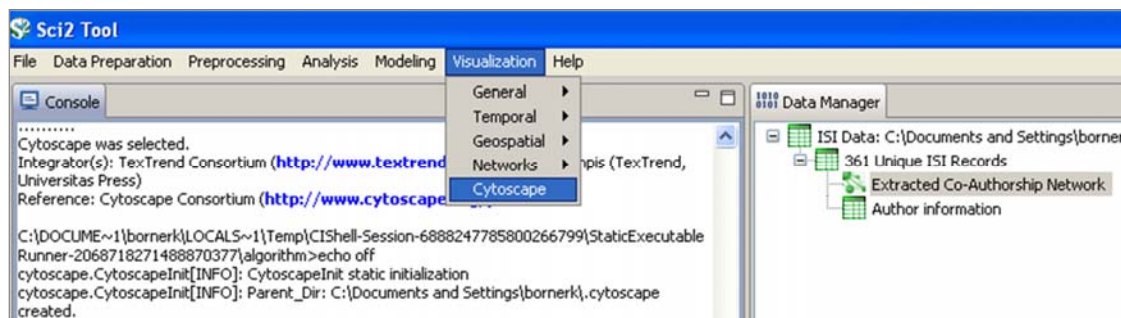
Select a network in Data Manager, run Cytoscape and the tool will start with this network loaded.
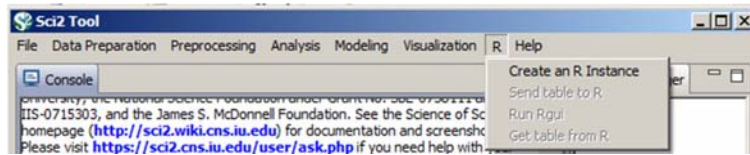
16

R statistical tool bridging

Gephi visualization tool bridging
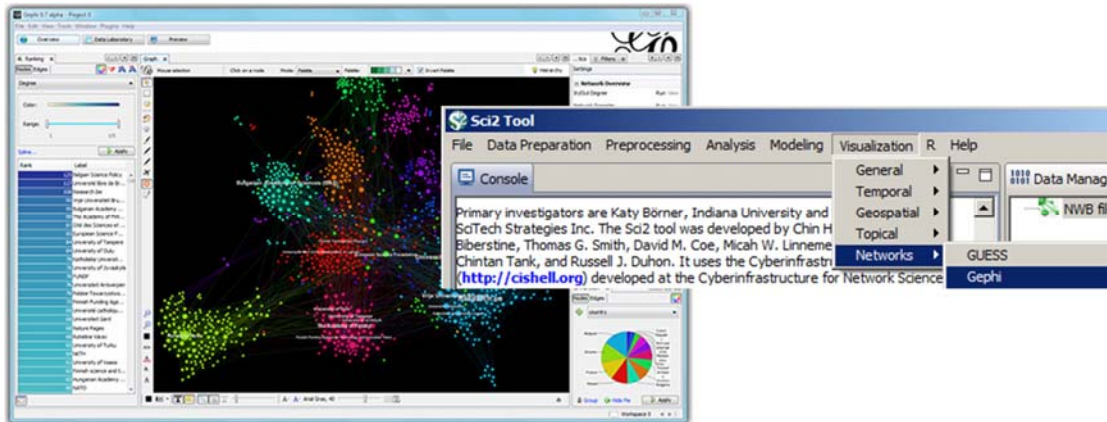
Sci2 Tool v1.0 Alpha (June 13, 2012)

Can be freely downloaded for all major operating systems from
http://sci2.cns.iu.edu

Select your operating system from the pull down menu and download.
Unpack into a /sci2 directory.
Run /sci2/sci2.exe

Sci2 Manual is at
http://sci2.wiki.cns.iu.edu

Cite as

Sci² Team. (2009). Science of Science (Sci²) Tool. Indiana University and SciTech Strategies, http://sci2.cns.iu.edu
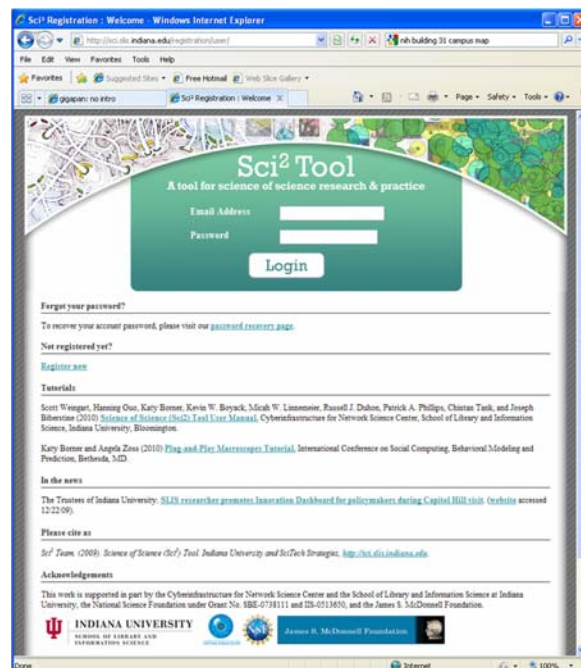
Sci2 Tool v1.0 Alpha (June 13, 2012)

Sci2 Tool runs on Windows, Mac, and Linux.
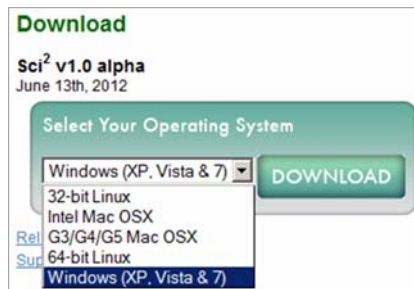
Unzip.

Run /sci2/sci2.exe

Sci2 Tool Interface Components
See also http://sci2.wiki.cns.iu.edu/2.2+User+Interface

Use

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

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

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

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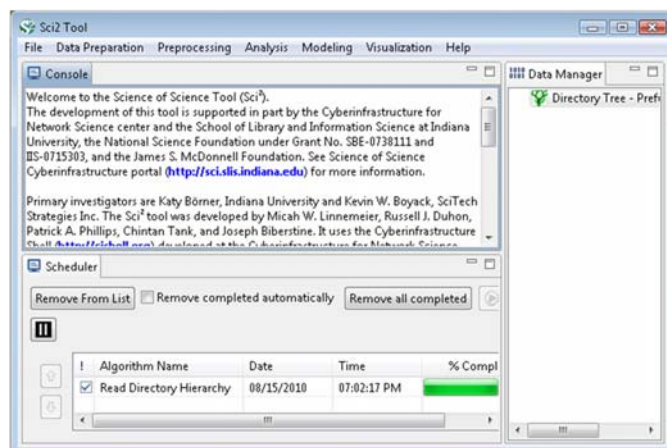
All workflows are recorded into a log file (see /sci2/logs/…), and soon can be re-run for easy replication. If errors occur, they are saved in a error log to ease bug reporting.

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

Tutorial Overview

2p Welcome and Overview of Tutorial and Attendees

2:15p Sci2 Tool Hands-on
➤   Download and run the Sci2 Tool
➤   **Temporal Analysis: Horizontal line graph of NSF projects**
➤   Geospatial Analysis: US and world maps
➤   Geospatial Analysis: Geomap with network overlays
➤   Topical Analysis: Visualize research profiles
➤   Network Analysis: Co-occurrence networks and bimodal networks
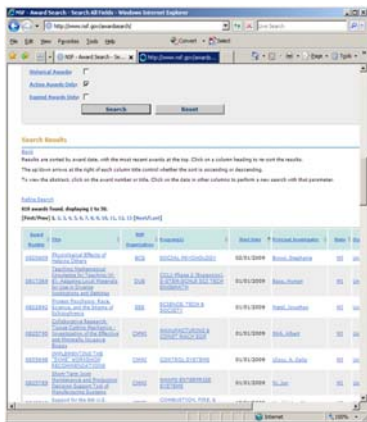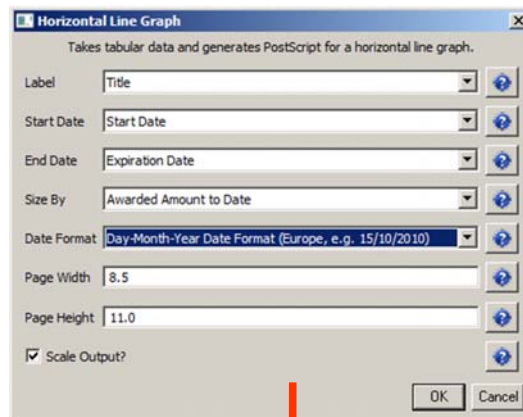➤   Network Analysis: Evolving collaboration networks

3:15p Outlook and Q&A

3:30p Adjourn

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

Download NSF data   Visualize as Horizontal Line Graph

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

Text

Start date   End date
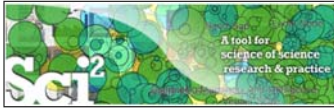
Horizontal line graph of NSF projects

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

Save in CSV format as *institution*.nsf

Temporal bar graph of NSF projects

Download and load a dataset of your choice or load one of the sample data files, e.g., *'sampledata/scientometrics/nsf/Indiana.nsf.'*
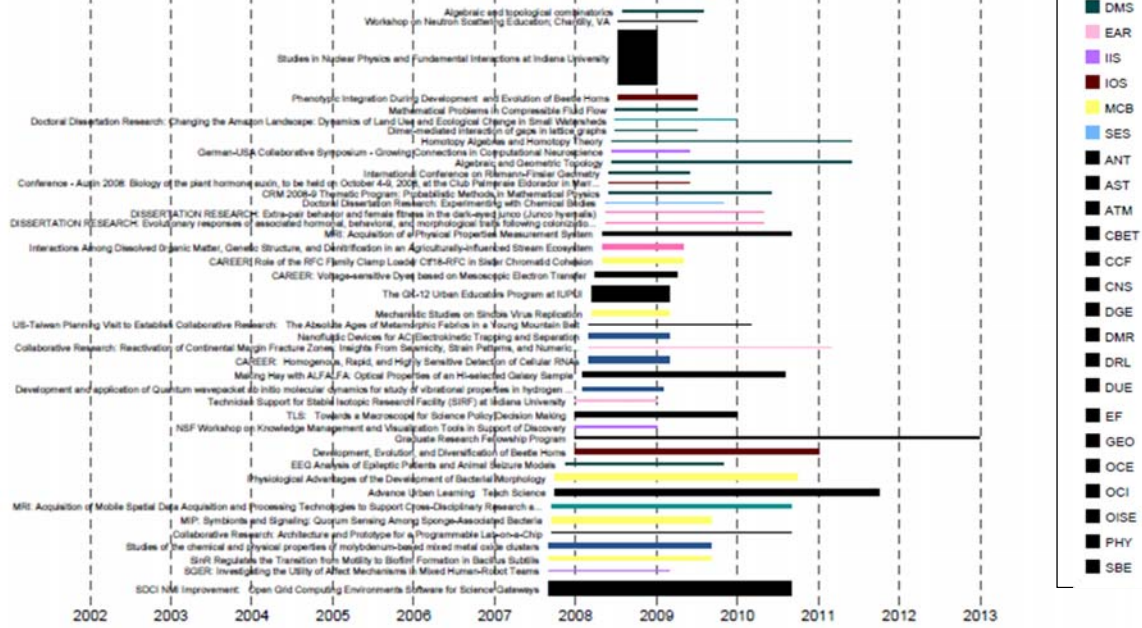
Run *'Visualization > Temporal > Temporal Bar Graph'* using parameters:

Save *'visualized with 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.
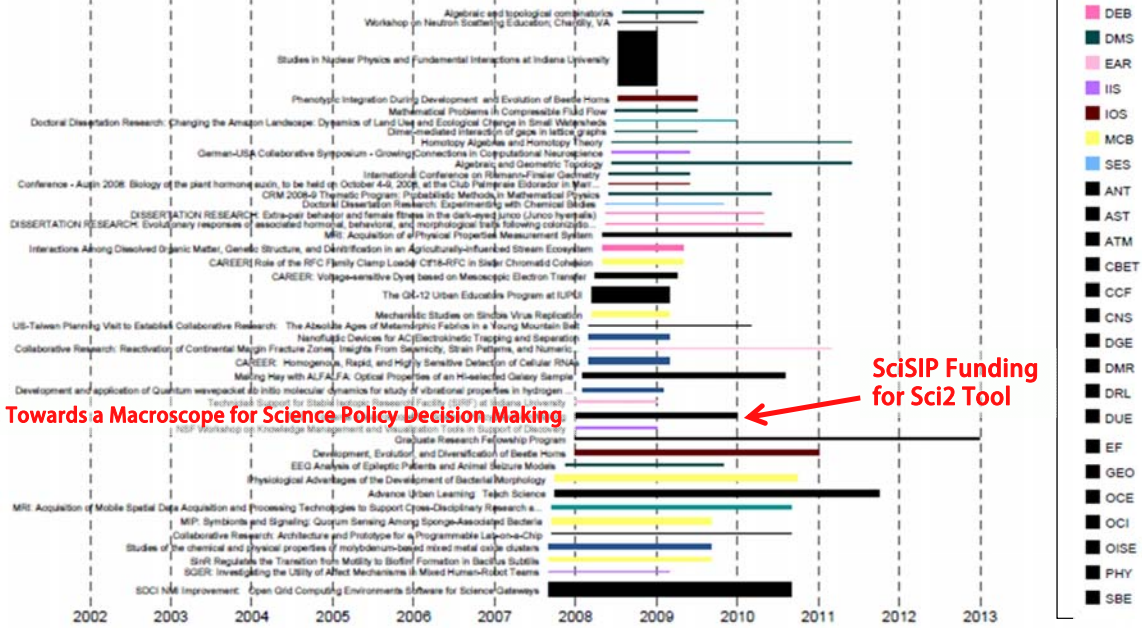
Temporal Visualization
Generated from NSF csv file: Indiana.nsf
July 18, 2012 | 8:46 AM EDT

Temporal Visualization
Generated from NSF csv file: Indiana.nsf
July 18, 2012 | 8:46 AM EDT

SciSIP Funding for Sci2 Tool

TLS: Towards a Macroscope for Science Policy Decision Making

Temporal Visualization
Generated from NSF csv file: Indiana.nsf -- 2nd
July 18, 2012 | 8:56 AM EDT

Seven grants by the "Indiana University of Pennsylvania Research Institute" should be excluded. Rerun analysis.

Temporal Bar Graph
Takes tabular data and generates PostScript for a temporal bar graph.

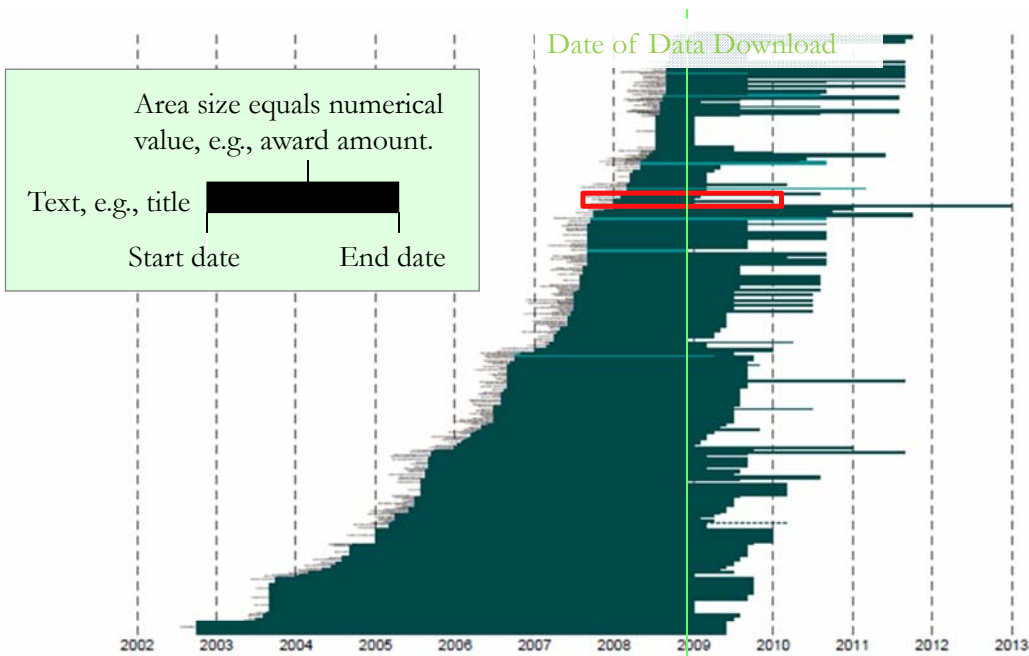|	
Subtitle	Generated from NSF csv file: Indiana.nsf -- 2nd
Label	Title
Start Date	Start Date
End Date	Expiration Date
Size By	Awarded Amount to Date
Date Format	Day-Month-Year Date Format (Europe, e.g. 15/10/2010)
Category	State

☑ Scale Output?
☐ Simplified Layout?

OK   Cancel

Legend
Area size: Awarded Amount to Date
Minimum = 0
Maximum = 6,402,330
Text label: Title
Color: State
See end of PDF for color legend.

Area
1,892,281
630,760
210,253
0
0.96 Year(s)
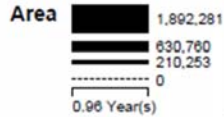
How To Read This Map
This *temporal bar graph* visualization represents each record as a horizontal bar with a specific start and end date and a text label on its left side. The area of each bar encodes a numerical attribute value, e.g., total amount of funding. Bars may be colored to present categorical attribute values of records.

CNS (cns.iu.edu)

■ IN
■ PA

Date of Data Download

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

Text, e.g., title

Start date   End date

Development and application of Quantum wavepacket ab initio molecular dynamics for study of vibrational properties in hydrogen bonded systems

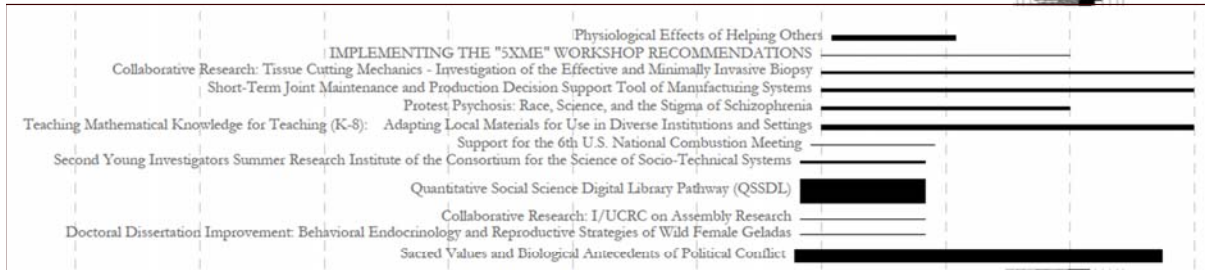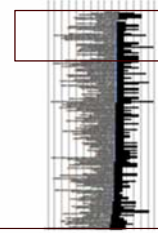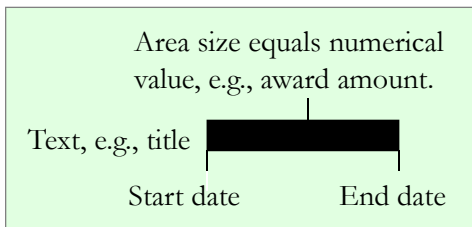Technician Support for Stable Isotopic Research Facility (SIRF) at Indiana University

TLS: Towards a Macroscope for Science Policy Decision Making

TLS: Towards a Macroscope for Science Policy Decision Making

NSF Workshop on Knowledge Management and Visualization Tools in Support of Discovery

Graduate Research Fellowship Program
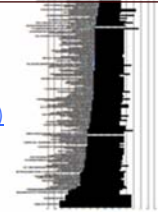
Temporal bar graph of NSF projects

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

Text, e.g., title

Start date   End date

Physiological Effects of Helping Others
IMPLEMENTING THE "5XME" WORKSHOP RECOMMENDATIONS
Collaborative Research: Tissue Cutting Mechanics - Investigation of the Effective and Minimally Invasive Biopsy
Short-Term Joint Maintenance and Production Decision Support Tool of Manufacturing Systems
Protest Psychosis: Race, Science, and the Stigma of Schizophrenia
Teaching Mathematical Knowledge for Teaching (K-8):   Adapting Local Materials for Use in Diverse Institutions and Settings
Support for the 6th U.S. National Combustion Meeting
Second Young Investigators Summer Research Institute of the Consortium for the Science of Socio-Technical Systems
Quantitative Social Science Digital Library Pathway (QSSDL)
Collaborative Research: I/UCRC on Assembly Research
Doctoral Dissertation Improvement: Behavioral Endocrinology and Reproductive Strategies of Wild Female Geladas
Sacred Values and Biological Antecedents of Political Conflict

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)

29

Tutorial Overview

2p Welcome and Overview of Tutorial and Attendees

2:15p Sci2 Tool Hands-on
- ➤ **Download and run the Sci2 Tool**
- ➤ **Temporal Analysis: Horizontal line graph of NSF projects**
- ➤ **Geospatial Analysis: US and world maps**
- ➤ **Geospatial Analysis: Geomap with network overlays**
- ➤ **Topical Analysis: Visualize research profiles**
- ➤ **Network Analysis: Co-occurrence networks and bimodal networks**
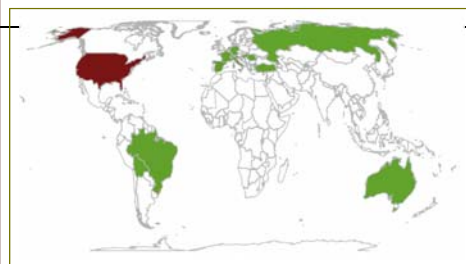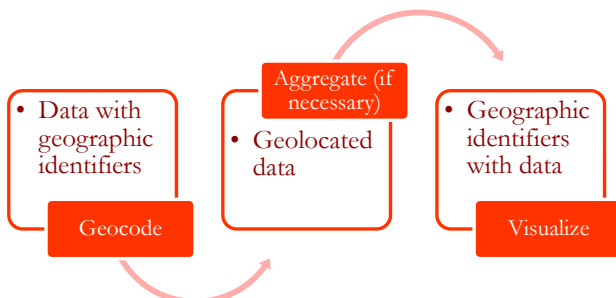- ➤ **Network Analysis: Evolving collaboration networks**
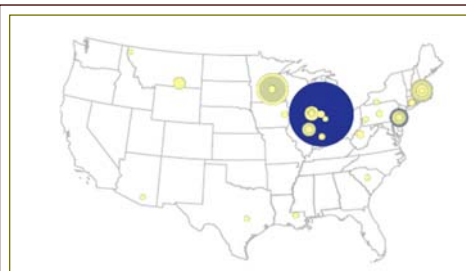
3:15p Outlook and Q&A

3:30p Adjourn

30

Geocoding and Geospatial Maps

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

- Data with geographic identifiers

  Geocode

- Aggregate (if necessary)
- Geolocated data

- Geographic identifiers with data

  Visualize

Region names + numeric data (Choropleth Map)


Geocoordinates + numeric data (Proportional Symbol Map)

Load File with Address and Times Cited Fields

Run '*File > Load…*' and select the sample data table '*sampledata/geo/usptoInfluenza.csv*'
Create a map of influenza patents held by different countries.

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

Data Manager

CSV file: C:\sci2\sampledata\geo\usptoInfluenza.csv

Yahoo Geocoder
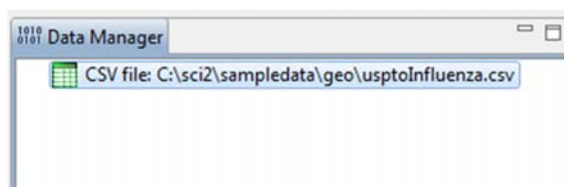
CIShell Yahoo Geocoder   ✎ Edit   ➕ Add ▾   ⚙ Tools ▾

Description

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

Pros & Cons

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

Applications

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

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

33

Using Yahoo! Geocoder

Run '*Analysis > Geospatial > Yahoo Geocoder*'
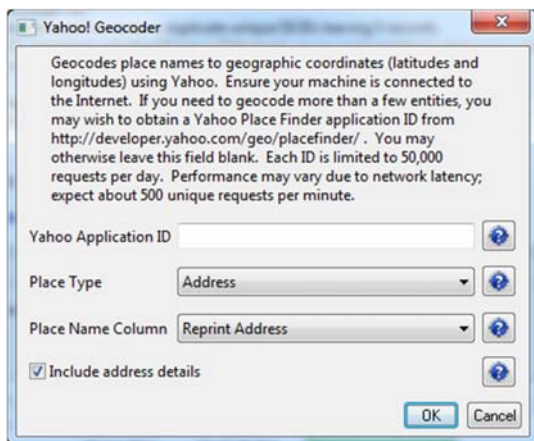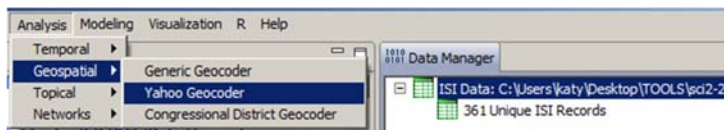
You can leave Application ID blank for trial purposes, but for heavy use, register for your own personal Yahoo!
Application ID, see:
http://developer.yahoo.com/geo/placefinder/

34

Aggregate by Country

Aggregate Data was selected.
Implementer(s): Chintan Tank
Documentation: **http://wiki.cns.iu.edu/display/CISHELL/Aggregate+Data**
Input Parameters:
Aggregate on column: Country
Delimiter for Country: |
Longitude: AVERAGE
Latitude: AVERAGE
Times Cited: SUM
Aggregated by '': All rows of Latitude column were skipped due to no non-null, non-empty values.
Aggregated by '': All rows of Longitude column were skipped due to no non-null, non-empty values.
Frequency of unique "Country" values added to "Count" column.
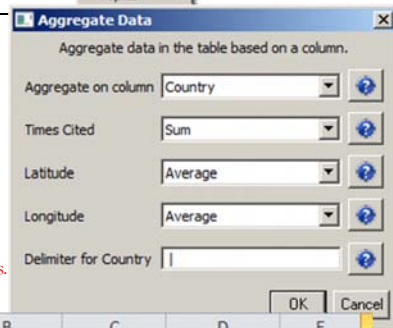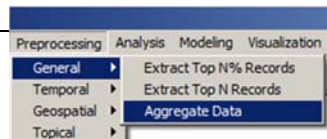
	A	B	C	D
1	Times Cited	Latitude	Longitude	Country
2	7	42.02946091	-87.68838501	United States
3	0			
4	0			
5	2	42.34999466	-71.08765411	United States
6	14	41.70074844	-86.23918915	United States
7	15	41.70074844	-86.23918915	United States
8	29	41.89422607	-87.61901855	United States
9	32	41.70074844	-86.23918915	United States
10	7	41.70074844	-86.23918915	United States
11	5	41.70074844	-86.23918915	United States
12	2	41.11500168	-85.73377991	United States
13	10	47.50622177	19.06481934	Hungary
14	44	41.70074844	-86.23918915	United States
15	0	47.50622559	19.06481934	Hungary
16	19	41.70074844	-86.23918915	United States

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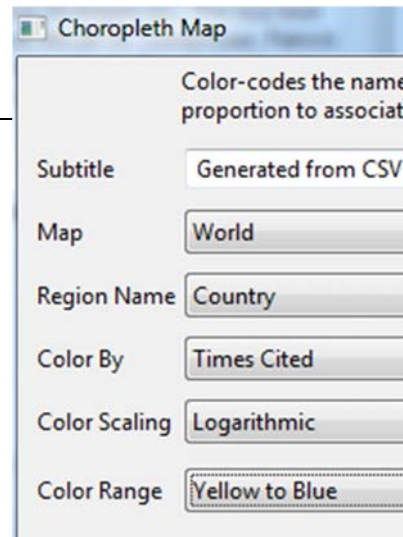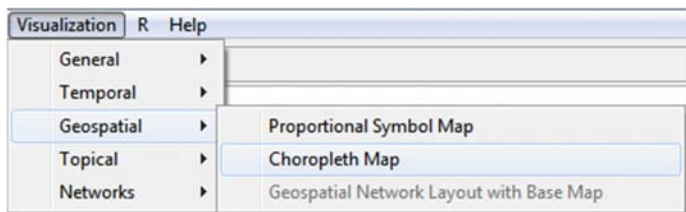
Choropleth Map

Right-click and **Save** map as PostScript file. Use PostScript Viewer or convert to pdf to view.
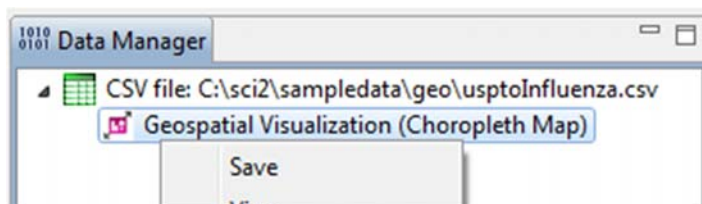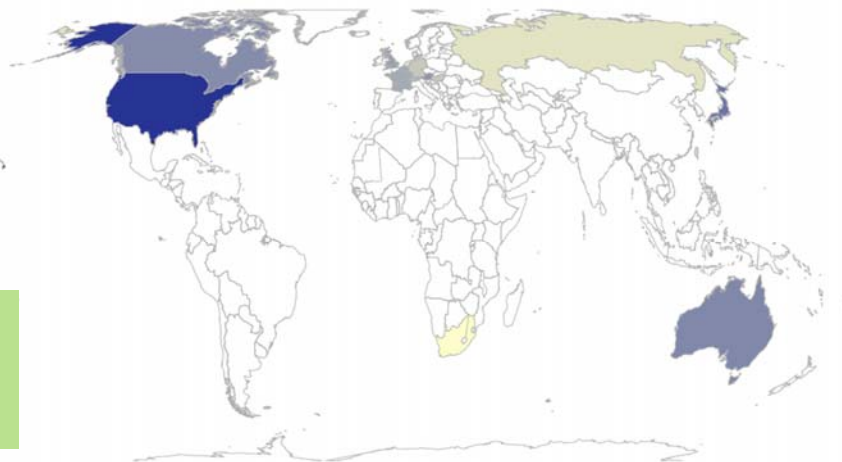
Reading the Choropleth Map

Geospatial Visualization (Choropleth Map)
Generated from CSV file: C:\sci2\sampledata\geo\usptoInfluenza.csv
Jun 14, 2012 | 05:33:37 PM EDT

Header shows visualization type, data description, and creation date

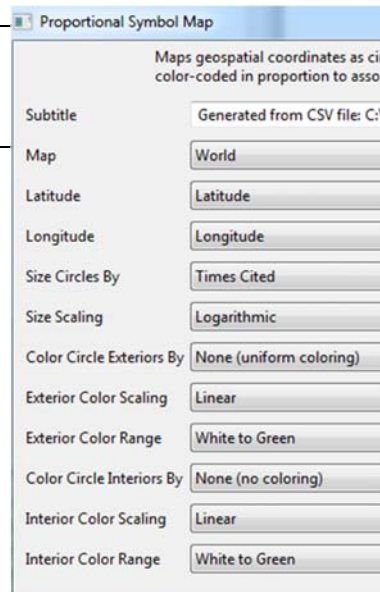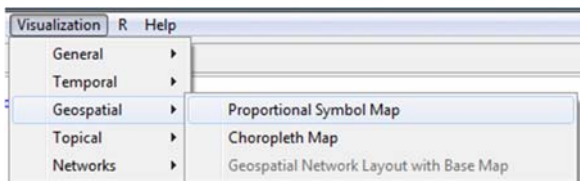Legend shows how data matches up with visual representation

Legend
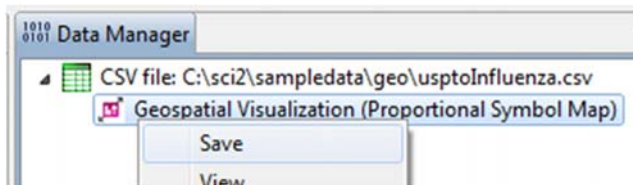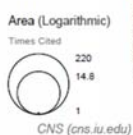Country Color (Logarithmic)
Times Cited

1   14.8   220

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

37

Proportional Symbol Map

Proportional Symbol Map

Maps geospatial coordinates as cir
color-coded in proportion to associ

Subtitle	Generated from CSV file: C:\s
Map	World
Latitude	Latitude
Longitude	Longitude
Size Circles By	Times Cited
Size Scaling	Logarithmic
Color Circle Exteriors By	None (uniform coloring)
Exterior Color Scaling	Linear
Exterior Color Range	White to Green
Color Circle Interiors By	None (no coloring)
Interior Color Scaling	Linear
Interior Color Range	White to Green

Visualization   R   Help

General   ▶
Temporal   ▶
Geospatial   ▶   Proportional Symbol Map
Topical   ▶   Choropleth Map
Networks   ▶   Geospatial Network Layout with Base Map

Right-click and **Save** map as PostScript file. Use PostScript Viewer or convert to pdf to view.

1010
0101 Data Manager

◢ 🔲 CSV file: C:\sci2\sampledata\geo\usptoInfluenza.csv
       📄 Geospatial Visualization (Proportional Symbol Map)
              Save
              View

38

Reading the Proportional Symbol Map

Geospatial Visualization (Proportional Symbol Map)
Generated from CSV file: C:\sci2\sampledata\geo\usptoInfluenza.csv
Jun 14, 2012 | 05:56:39 PM EDT

Header shows visualization type, data description, and creation date

Legend shows how data matches up with visual representation

Area (Logarithmic)
Times Cited
220
14.8
1
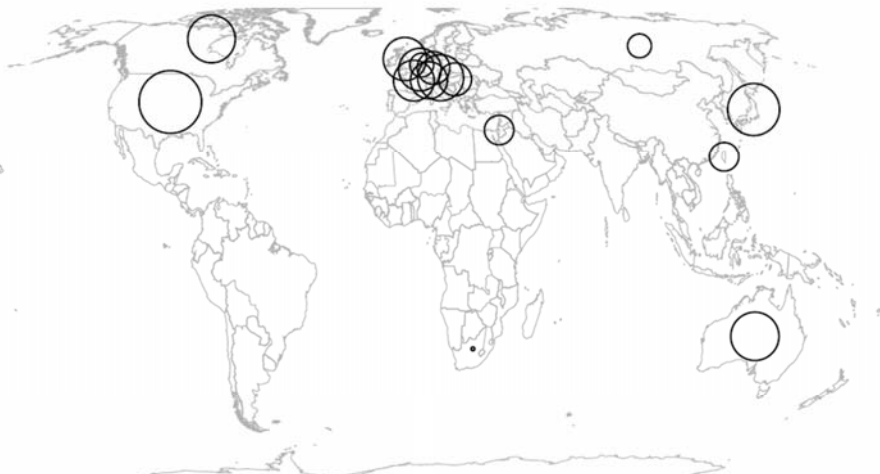
How to Read this Map
This *proportional symbol map* shows 209 countries of the world using the equal-area Eckert IV projection. Each dataset record is represented by a circle centered at its geolocation. The area, interior color, and exterior color of each circle may represent numeric attribute values. Minimum and maximum data values are given in the legend.

CNS (cns.iu.edu)

39

Relevant Sci2 Manual entry

✏ Edit   ➕ Add ▾   ⚙ Tools ▾

5.2.4 Mapping Scientometrics (ISI Data)
📎9 Added by Ted Polley, last edited by Ted Polley on Nov 14, 2011 (view change)

5.2.4.1 Document Co-Citation

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

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

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

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

Extetior Color (Linear)
Publication Year
1978   1993   2008

Area (Linear)
Times Cited
0
73
146
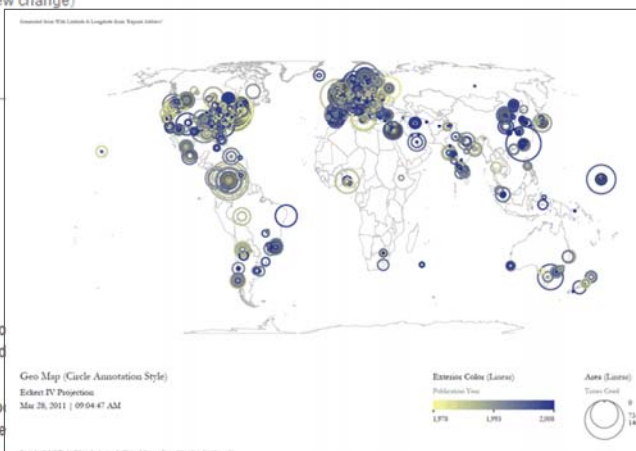
New ISI File Format
Web of Science made a change to their output format in September, 2011. Older versions of Sci2 tool may refuse to load these new files, with an error like "Invalid ISI format file selected."

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

40

Tutorial Overview

2p Welcome and Overview of Tutorial and Attendees

2:15p Sci2 Tool Hands-on
➢ **Download and run the Sci2 Tool**
➢ **Temporal Analysis: Horizontal line graph of NSF projects**
➢ **Geospatial Analysis: US and world maps**
➢ **Geospatial Analysis: Geomap with network overlays**
➢ **Topical Analysis: Visualize research profiles**
➢ **Network Analysis: Co-occurrence networks and bimodal networks**
➢ **Network Analysis: Evolving collaboration networks**

3:15p Outlook and Q&A

3:30p Adjourn

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

File with
geolocations and
linkage info, e.g.,
an isi biblio-
graphy file.

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

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

 + =

Relevant Sci2 Manual entry

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

---

Use Sci2 Tool to Generate Geomap and Network File
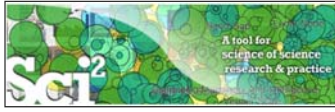
Read prepared .net file and run:

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

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

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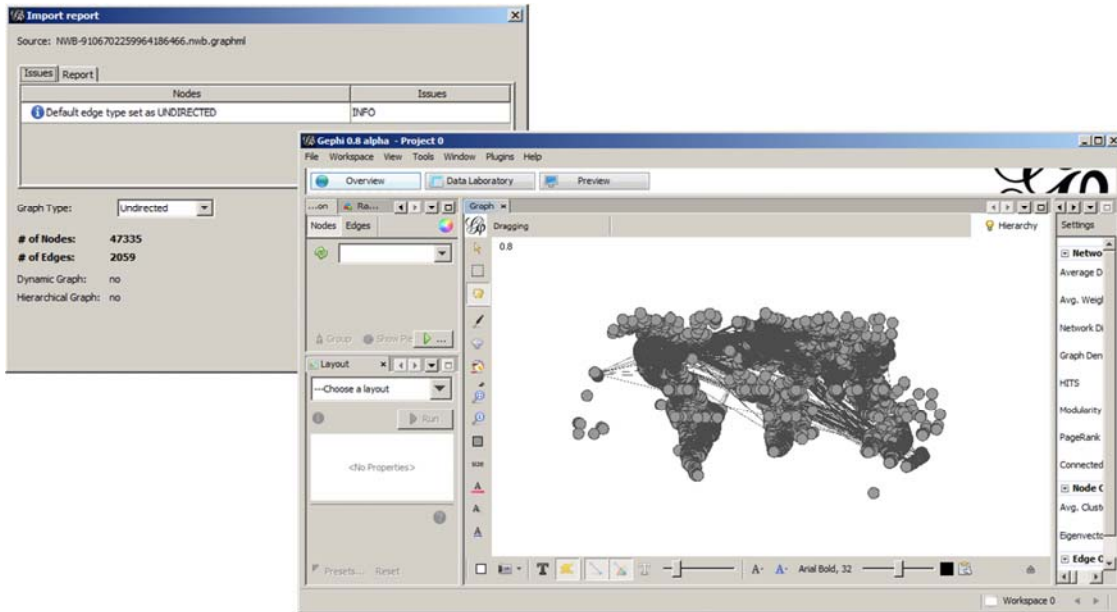
Follow instructions in online tutorial on **Manipulating the Network File in Gephi**

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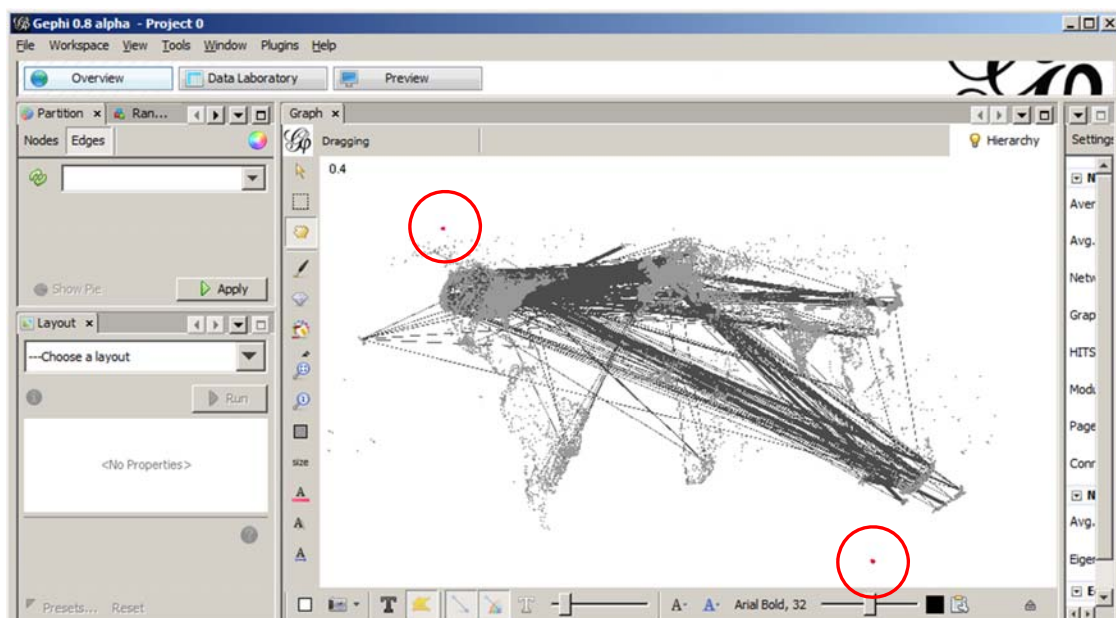
Load geomap and network files into Photoshop. Select 'network' layer an use 'Right click, Duplicate Layer' to copy network over to 'geomap' file as a second layer.
Use Edit > Transform > Scale' and align using the "Near Antarctica" anchor nodes, see instructions in online tutorial on **Creating the Visualization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

47
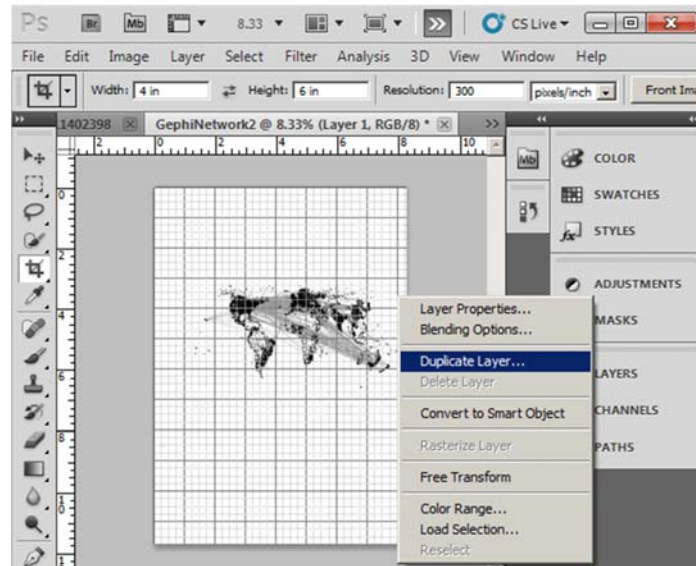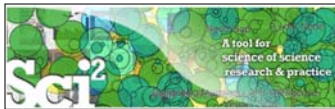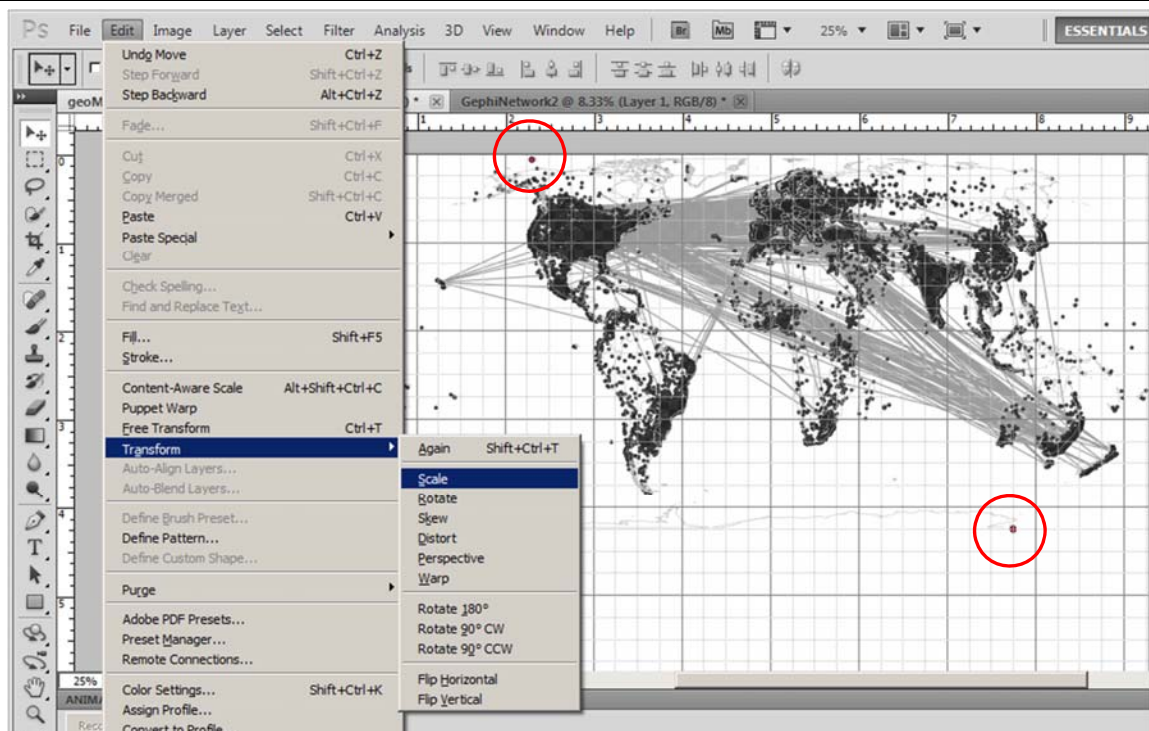
48

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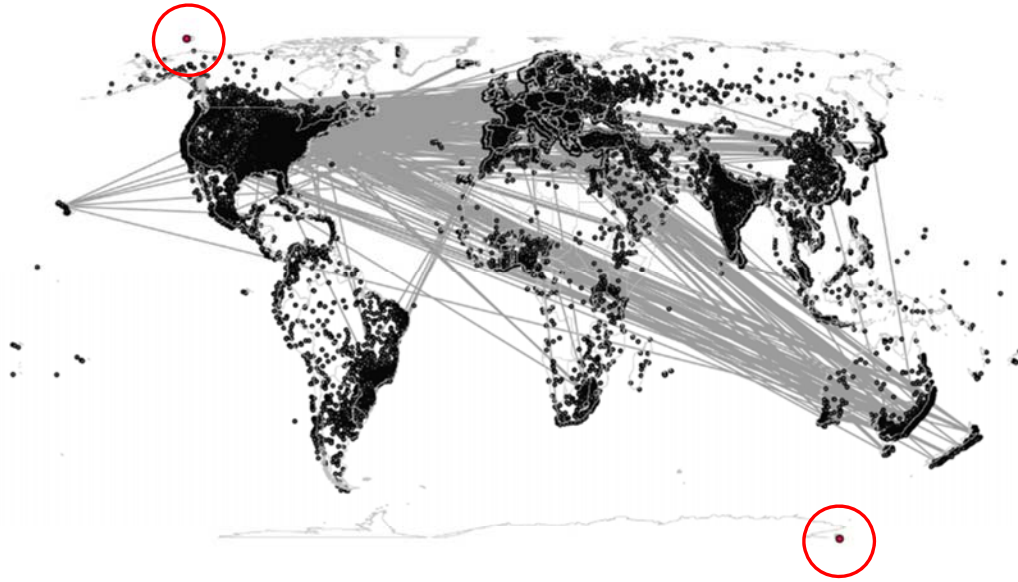
Geo Map ()
Eckert IV Projection
Apr 06, 2012 | 03:19:51 AM

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

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

4.7.6 Using Gephi to Render Networks Overlaid on GeoMaps
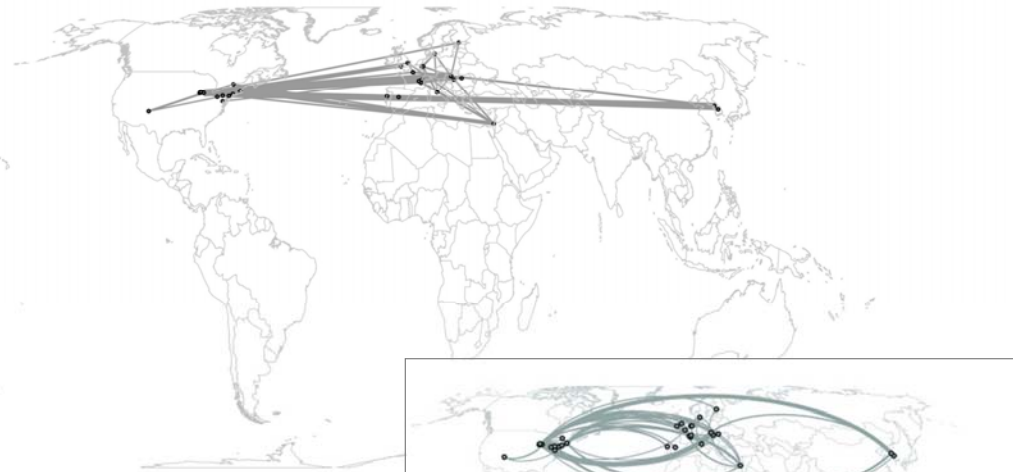
Geo Map ()
Eckert IV Projection
Apr 11, 2012 | 06:20:13 AM

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

Rounded edges might increase legibility of overlapping lines.

Tutorial Overview

2p Welcome and Overview of Tutorial and Attendees

2:15p Sci2 Tool Hands-on
- **Download and run the Sci2 Tool**
- **Temporal Analysis: Horizontal line graph of NSF projects**
- **Geospatial Analysis: US and world maps**
- **Geospatial Analysis: Geomap with network overlays**
- **Topical Analysis: Visualize research profiles**
- **Network Analysis: Co-occurrence networks and bimodal networks**
- **Network Analysis: Evolving collaboration networks**

3:15p Outlook and Q&A

3:30p Adjourn
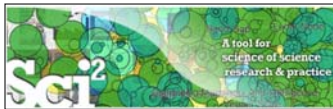
Topical Analysis: Research Profiles

Data: WoS and Scopus paper level data for 2001–2010, about 25,000 separate journals, proceedings, and series.
Similarity Metric: Combination of bibliographic coupling and keyword vectors.
Number of Disciplines: 554 journal clusters further aggregated into 13 main disciplines.
Börner, Katy, Richard Klavans, et al. (2012) Design and Update of a Classification System: The UCSD Map of Science. PLoS ONE 7(7): e39464.
doi:10.1371/journal.pone.0039464

Topical Visualization

Generated from 361 Unique ISI Records of 4 NetSci Researchers
14 out of 109 publications were mapped to 94 subdisciplines and 12 disciplines.
June 05, 2012 | 05:39 PM EDT
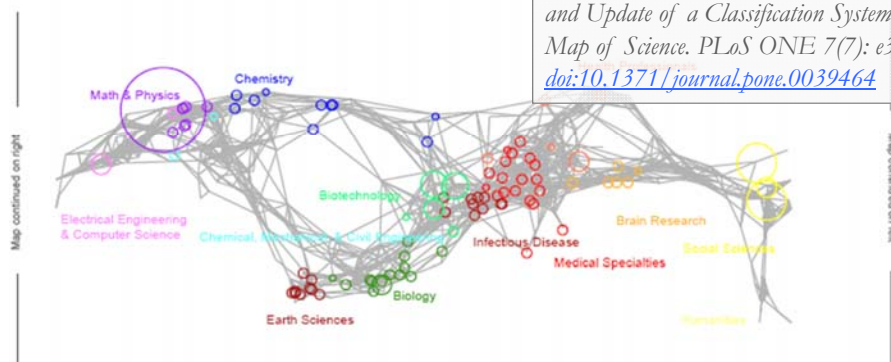
2008 The Regents of the University of California and SciTech Strategies.
Map updated by SciTech Strategies, OST, and CNS in 2011.

Legend
Circle area: Fractional Journal Count
Unclassified = 95
Minimum = 0
Maximum = 25
Color: Discipline
See end of PDF for color legend.

Area
29.09
16.19
2.8

How To Read This Map
The *UCSD* map of science depicts a network of 554 subdiscipline nodes that are aggregated to 13 main disciplines of science. Each discipline has a distinct color and is labeled. Overlaid are circles, each representing all records per unique subdiscipline. Circle area is proportional to the number of fractionally assigned records. Minimum and maximum data values are given in the legend.
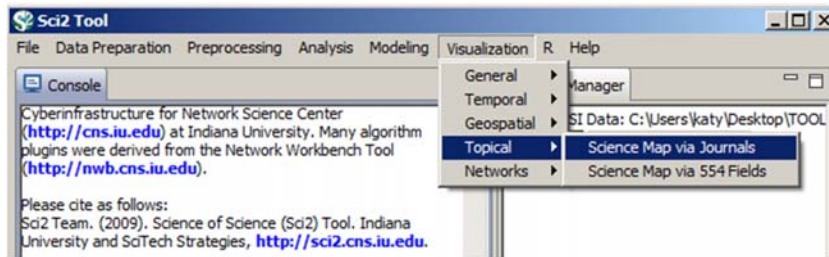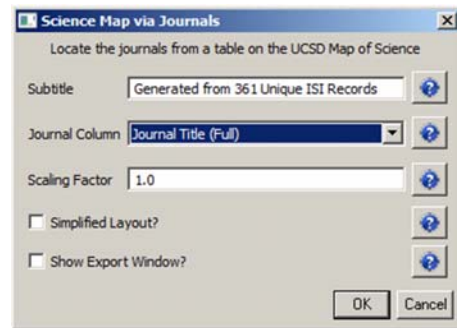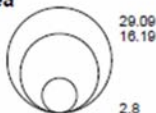
CNS (cns.iu.edu)

Load an ISI (*.isi), Bibtex (*.bib), Endnote Export Format (*.enw), Scopus csv (*.scopus) file such as */sci2/sampledata/scientometrics/isi/FourNetSciResearchers.isi*

Run '*Visualization > Topical > Science Map via Journals*'
using parameters given to the right.

Postscript file will appear in *Data Manager*.
Save and open with a Postscript Viewer.

53

Topical Visualization

Generated from 361 Unique ISI Records
90 out of 112 publications were mapped to 182 subdisciplines and 13 disciplines.
June 24, 2012 | 04:04 PM EDT

Legend

Circle area: Fractional Journal Count
Unclassified = 22
Minimum = 0
Maximum = 98
Color: Discipline
See end of PDF for color legend.

Area

29.09
16.19

2.8

How To Read This Map

The *UCSD map of science* depicts a network of 554 subdiscipline nodes that are aggregated to 13 main disciplines of science. Each discipline has a distinct color and is labeled. Overlaid are circles, each representing all records per unique subdiscipline. Circle area is proportional to the number of fractionally assigned records. Minimum and maximum data values are given in the legend.

CNS (cns.iu.edu)

Topical Visualization

■ Biology
1 BMC EVOLUTIONARY BIOLOGY
1 NATURWISSENSCHAFTEN

■ Biotechnology
1 BMC BIOINFORMATICS
2 FEBS JOURNAL
1 GENOME RESEARCH
1 INTERNATIONAL MICROBIOLOGY
1 NATURE BIOTECHNOLOGY
3 NATURE GENETICS
1 NATURE REVIEWS GENETICS
1 NUCLEIC ACIDS RESEARCH
2 PROTEOMICS

■ Brain Research
5 JOURNAL OF MATHEMATICAL PSYCHOLOGY

■ Chemical, Mechanical, & Civil Engineering
1 JOURNAL OF CERAMIC PROCESSING RESEARCH
2 MATERIALS SCIENCE AND ENGINEERING A–STRUCTURAL MATERIA...
1 PHYSICS WORLD
1 SCIENTIFIC AMERICAN

■ Chemistry
1 COMPUTER PHYSICS COMMUNICATIONS
2 JOURNAL OF CHEMICAL INFORMATION AND COMPUTER SCIENCES
1 JOURNAL OF THE INDIAN INSTITUTE OF SCIENCE
1 PURE AND APPLIED CHEMISTRY

■ Earth Sciences
1 CURRENT SCIENCE

■ Electrical Engineering & Computer Science
1 ASIST 2003: PROCEEDINGS OF THE 66TH ASIST ANNUAL MEETING,...
1 CANADIAN JOURNAL OF INFORMATION AND LIBRARY SCIENCE–REV...
5 IEEE TRANSACTIONS ON PROFESSIONAL COMMUNICATION
1 INFORMATION TECHNOLOGY AND LIBRARIES
5 JOURNAL OF INFORMATION SCIENCE
3 JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE
5 JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENC...
2 LIBRARY QUARTERLY
1 LIBRI
1 PROCEEDINGS OF THE AMERICAN SOCIETY FOR INFORMATION SC...

■ Health Professionals
1 ANNALS OF BIOMEDICAL ENGINEERING
1 BULLETIN OF THE MEDICAL LIBRARY ASSOCIATION
1 CROATIAN MEDICAL JOURNAL
2 JOURNAL OF APPLIED PHYSIOLOGY
1 JOURNAL OF PUBLIC HEALTH DENTISTRY
1 METHODS OF INFORMATION IN MEDICINE
1 PLASTIC AND RECONSTRUCTIVE SURGERY
1 TEXAS MEDICINE
1 UNFALLCHIRURG
1 WIENER KLINISCHE WOCHENSCHRIFT

■ Humanities
1 BULLETIN OF THE ATOMIC SCIENTISTS

■ Infectious Diseases
1 FEMS MICROBIOLOGY LETTERS
1 JOURNAL OF BACTERIOLOGY

■ Math & Physics
1 ADVANCES IN APPLIED PROBABILITY

Topical Visualization

■ Math & Physics
10 APPLIED PHYSICS LETTERS
1 BRAZILIAN JOURNAL OF PHYSICS
3 CHAOS SOLITONS & FRACTALS
1 COMPLEXITY
1 COMPUTATIONAL MATERIALS SCIENCE
11 EUROPEAN PHYSICAL JOURNAL B
12 EUROPHYSICS LETTERS
2 INTERNATIONAL JOURNAL OF MODERN PHYSICS B
6 JOURNAL OF PHYSICS A–MATHEMATICAL AND GENERAL
1 JOURNAL OF STATISTICAL MECHANICS–THEORY AND EXPERIMENT
1 JOURNAL OF STATISTICAL PHYSICS
1 JOURNAL OF THE KOREAN PHYSICAL SOCIETY
1 MATERIALS SCIENCE AND ENGINEERING B–SOLID STATE MATERIAL...
3 NATURE PHYSICS
3 NUCLEAR INSTRUMENTS & METHODS IN PHYSICS RESEARCH SEC...
12 PHYSICA A
5 PHYSICAL REVIEW A
2 PHYSICAL REVIEW B
45 PHYSICAL REVIEW LETTERS
2 REVIEWS OF MODERN PHYSICS

■ Medical Specialties
1 ANNALS OF INTERNAL MEDICINE
1 REVISTA DE INVESTIGACION CLINICA

■ Social Sciences
1 ADMINISTRATIVE SCIENCE QUARTERLY
1 AMERICAN BEHAVIORAL SCIENTIST
1 AMERICAN SOCIOLOGICAL REVIEW
1 ANNALS OF THE AMERICAN ACADEMY OF POLITICAL AND SOCIAL S...
1 ARBOR–CIENCIA PENSAMIENTO Y CULTURA
3 BRITISH JOURNAL OF MATHEMATICAL & STATISTICAL PSYCHOLOGY
1 JOURNAL OF CLASSIFICATION

■ Social Sciences
2 JOURNAL OF MATHEMATICAL SOCIOLOGY
3 JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2 PSYCHOLOGICAL BULLETIN
5 PSYCHOMETRIKA
1 RECHERCHE
5 SCIENTOMETRICS
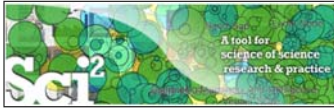1 SOCIAL FORCES
6 SOCIAL NETWORKS
3 SOCIOLOGICAL METHODS & RESEARCH

■ Multiple Categories
1 BRITISH MEDICAL JOURNAL
2 JAMA–JOURNAL OF THE AMERICAN MEDICAL ASSOCIATION
1 JOURNAL OF THEORETICAL BIOLOGY
18 NATURE
44 PHYSICAL REVIEW E
5 PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE ...
6 SCIENCE

■ Unclassified
1 ALGORITHMS AND MODELS FOR THE WEB–GRAPHS, PROCEEDINGS
2 AMERICAN DOCUMENTATION
2 ASIST 2002: PROCEEDINGS OF THE 65TH ASIST ANNUAL MEETING, ...
1 BIOLOGIYA MORYA–MARINE BIOLOGY
1 BULLETIN OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE
1 CHEMIKER–ZEITUNG
3 CHEMTECH
1 COMBINATORIAL AND ALGORITHMIC ASPECTS OF NETWORKING
7 CURRENT COMMENTS
2 CURRENT CONTENTS/LIFE SCIENCES
1 FEDERATION PROCEEDINGS
5 FRACTALS–AN INTERDISCIPLINARY JOURNAL ON THE COMPLEX GE...
1 FRONTIERS OF LIBRARIANSHIP–SYRACUSE UNIVERSITY

In addition to using **journal names** to
- Map career trajectories
- Identify evolving expertise areas
- Compare expertise profiles

Existing classifications can be aligned and used to generate science map overlays.

B KNOWLEDGE AREA	C NO. Projects	D USDA Staff Years	E STATE APPR	F TOTAL FUNDS	G UCSD Map Field Name
101 Appraisal of Soil Resources					315
102 Soil, Plant, Water, Nutrient Relationships					227
103 Management of Saline and Sodic Soils and Salinity					158
104 Protect Soil from Harmful Effects of Natural Elements					120
111 Conservation and Efficient Use of Water					245
112 Watershed Protection and Management					245
121 Management of Range Resources					520
122 Management and Control of Forest and Range Fires					520
123 Management and Sustainability of Forest Resources					231
124 Urban Forestry					231
125 Agroforestry					231
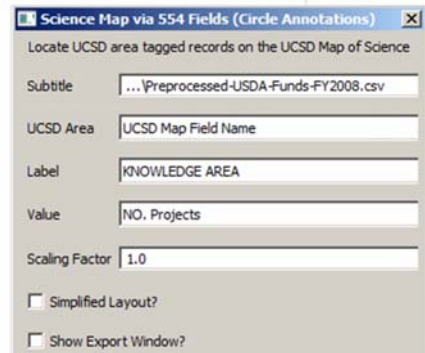
Run *Visualization > Topical > Science Map via 554 Fields*
using parameters given to the right.
Postscript file will appear in *Data Manager*.
Save and open with a Postscript Viewer.

Science Map via 554 Fields (Circle Annotations)

Locate UCSD area tagged records on the UCSD Map of Science

Subtitle	...\Preprocessed-USDA-Funds-FY2008.csv
UCSD Area	UCSD Map Field Name
Label	KNOWLEDGE AREA
Value	NO. Projects
Scaling Factor	1.0

☐ Simplified Layout?

☐ Show Export Window?

57
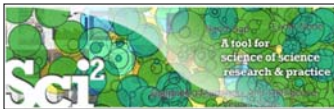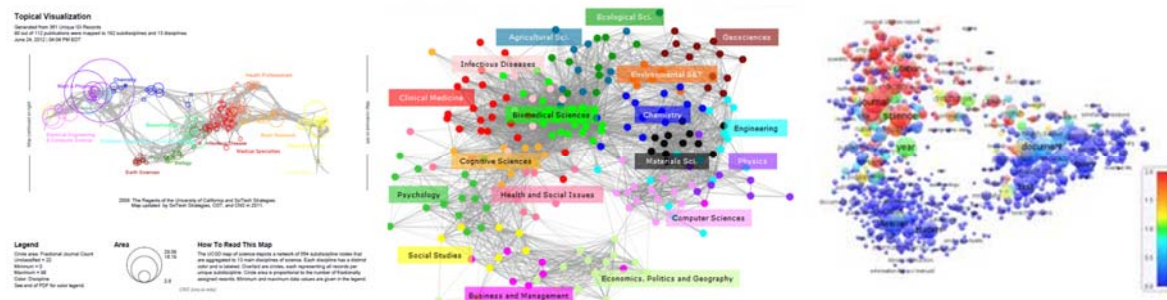
UCSD Map   Loet et al science maps ISI categories   http://vosviewer.com

Elsevier's SciVal Map   Science-Metrix.com   NIH Map (https://app.nihmaps.org)
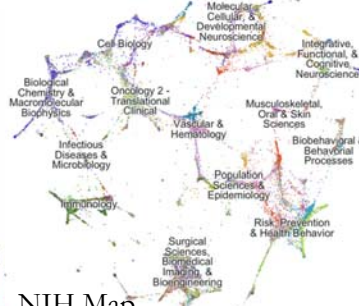
58

Tutorial Overview

2p Welcome and Overview of Tutorial and Attendees

2:15p Sci2 Tool Hands-on
➢ **Download and run the Sci2 Tool**
➢ **Temporal Analysis: Horizontal line graph of NSF projects**
➢ **Geospatial Analysis: US and world maps**
➢ **Geospatial Analysis: Geomap with network overlays**
➢ **Topical Analysis: Visualize research profiles**
➢ **Network Analysis: Co-occurrence networks and bimodal networks**
➢ **Network Analysis: Evolving collaboration networks**

3:15p Outlook and Q&A

3:30p Adjourn

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   Authors
Text Delimiter   ;
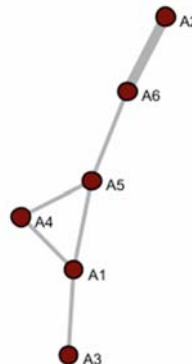
*Vertices 6
1 A1
2 A6
3 A2
4 A3
5 A5
6 A4
*Edges 6
2 3 2
1 4 1
1 5 1
5 6 1
1 6 1
2 5 1

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

● Author
■ Paper

Paper-author bipartite (2-mode) network

Extract Bipartite Network

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

First column | Paper
Second column | Authors
Text Delimiter | ;

Object: nodes based on -> Property: bipartitetype Operator: == Value: Paper
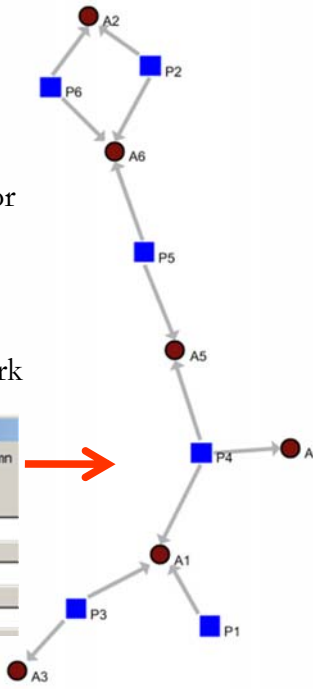
*Vertices 12
1 P1 bipartitetype "Paper"
2 A1 bipartitetype "Authors"
3 P2 bipartitetype "Paper"
4 A2 bipartitetype "Authors"
5 A6 bipartitetype "Authors"
6 P3 bipartitetype "Paper"
7 A3 bipartitetype "Authors"
8 P4 bipartitetype "Paper"
9 A4 bipartitetype "Authors"
10 A5 bipartitetype "Authors"
11 P5 bipartitetype "Paper"
12 P6 bipartitetype "Paper"
*Arcs
1 2
3 4
3 5
6 2
6 7
8 2
8 10
8 9
11 5
11 10
12 4
12 5

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

● Author
■ Paper

Extract Directed Network

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

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

Analysis Modeling Visualization R Help
   Temporal ▸
   Geospatial ▸
   Topical ▸
   Networks ▸   Network Analysis Toolkit (NAT)
...from the Layout mer
und in   Unweighted & Undirected ▸
hell-Session-46639   Weighted & Undirected ▸
                Unweighted & Directed ▸   Node Indegree
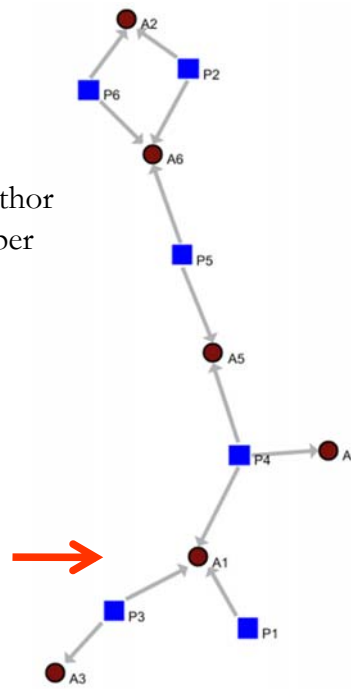
*Vertices 12
1 P1 indegree 0
2 A1 indegree 3
3 P2 indegree 0
4 A2 indegree 2
5 A6 indegree 3
6 P3 indegree 0
7 A3 indegree 1
8 P4 indegree 0
9 A4 indegree 1
10 A5 indegree 2
11 P5 indegree 0
12 P6 indegree 0
*Arcs
1 2
3 4
3 5
6 2
6 7
8 10
8 2
8 9
11 10
11 5
12 4
12 5

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

Extract Directed Network

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

Source Column: Paper

Target Column: References

Text Delimiter: ;

Arcs from papers to references

1970   1980   1990   1995   2000
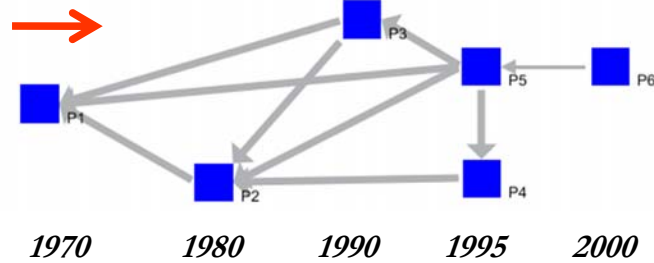
*Vertices 6
1 P1
2 P2
3 P3
4 P4
5 P5
6 P6
*Arcs
2 1
3 1
3 2
4 2
5 4
5 3
5 1
5 2
6 5

63

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

Extract Bipartite Network

Extract a bipartite network
values may list multiple ent
them.

First column: Paper

Second column: References

Text Delimiter: ;

WRONG!!!

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

64

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
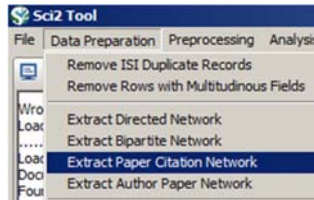
Arcs from references to papers—
in the direction of information flow
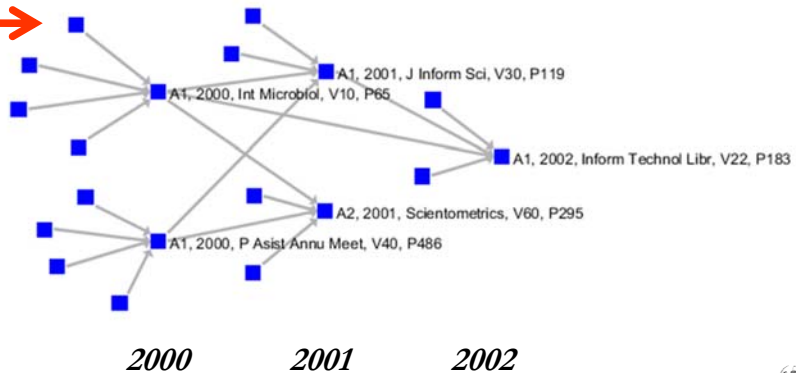
Sci2 Tool

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

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

2000   2001   2002

65

Tutorial Overview

2p Welcome and Overview of Tutorial and Attendees

2:15p Sci2 Tool Hands-on
  ➢ Download and run the Sci2 Tool
  ➢ Temporal Analysis: Horizontal line graph of NSF projects
  ➢ Geospatial Analysis: US and world maps
  ➢ Geospatial Analysis: Geomap with network overlays
  ➢ Topical Analysis: Visualize research profiles
  ➢ Network Analysis: Co-occurrence networks and bimodal networks
  ➢ **Network Analysis: Evolving collaboration networks**

3:15p Outlook and Q&A

3:30p Adjourn

66

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

67
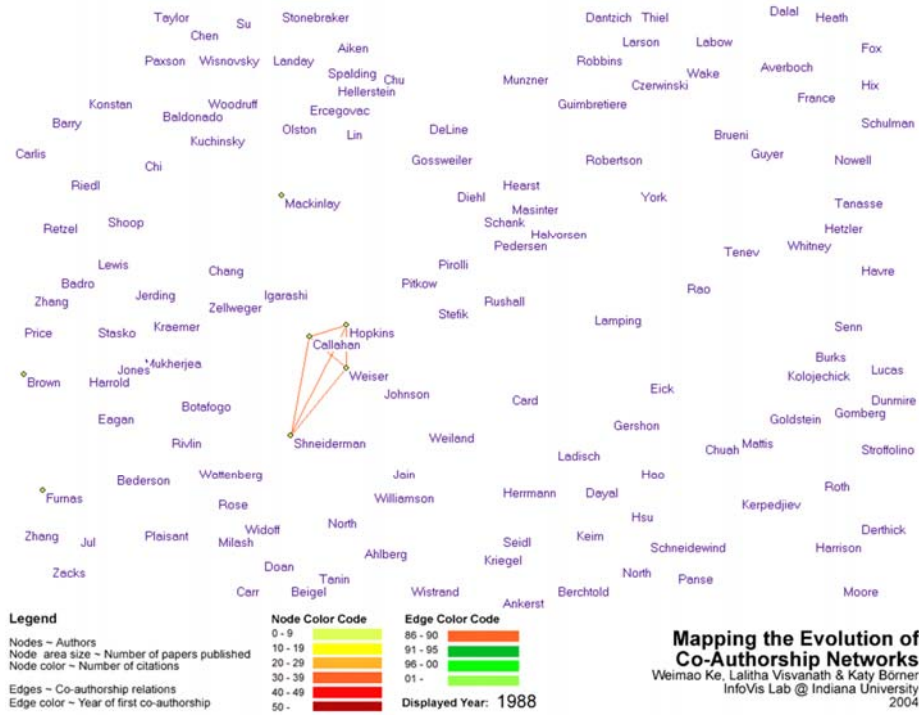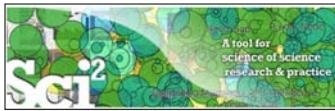
Load isi formatted file

As csv, file looks like:

Visualize each time slide separately:

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

68

Relevant Sci2 Manual entry

Slice Table by Time

Visualize Each Network, Keep Node Positions

1. To see the evolution of Vespignani's co-authorship network over time, check '*cumulative*'.

2. Extract co-authorship networks one at a time for each sliced time table using *'Data Preparation > Extract Co-Author Network'*, making sure to select "ISI" from the pop-up window during the extraction.

3. To view each of the Co-Authorship Networks over time using the same graph layout, begin by clicking on longest slice network (the *'Extracted Co-Authorship Network'* under *'slice from beginning of 1990 to end of 2006 (101 records)'*) in the data manager. Visualize it in GUESS using *'Visualization > Networks > GUESS'*.

4. From here, run *'Layout > GEM'* followed by *'Layout > Bin Pack'*. Run *'Script > Run Script …'* and select '*yoursci2directory/scripts/GUESS/co-author-nw.py*'.

5. In order to save the x, y coordinates of each node and to apply them to the other time slices in GUESS, select *'File > Export Node Positions'* and save the result as '*yoursci2directory/NodePositions.csv*'. Load the remaining three networks in GUESS using the steps described above and for each network visualization, run *'File > Import Node Positions'* and open *'yoursci2directory/NodePositions.csv'*.

6. To match the resulting networks stylistically with the original visualization, run *'Script > Run Script …'* and select *'yoursci2directory/scripts/GUESS/co-author-nw.py', followed by 'Layout > Bin Pack',* for each.

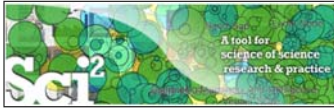*http://sci2.wiki.cns.iu.edu/5.1.2+Time+Slicing+of+Co-Authorship+Networks+(ISI+Data)*

Visualize Each Network, Keep Node Positions

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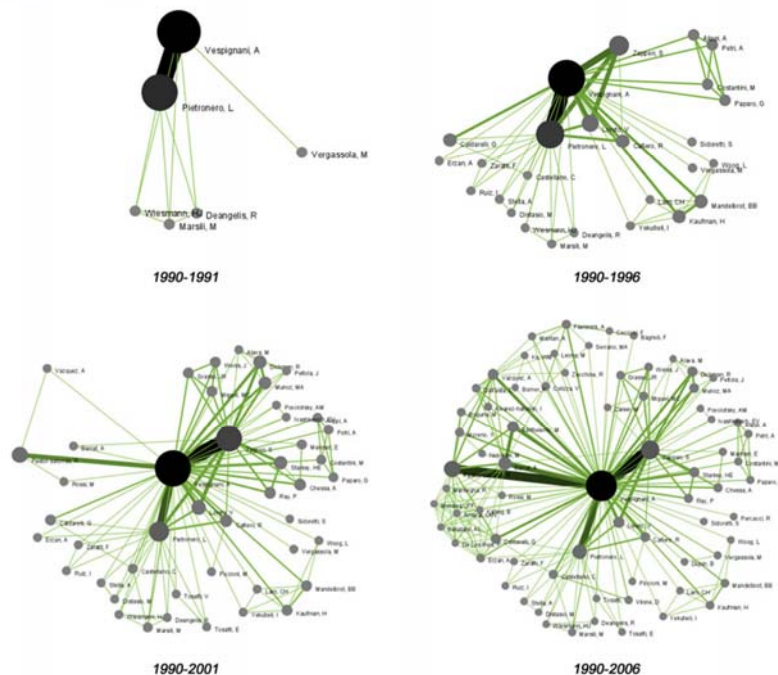
Relevant CIShell plugin

CIShell Slice Table by Time

Added by Aretha Alencar, last edited by Ted Polley on Jan 12, 2011 (view change)

Description

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

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

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

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

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

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

Pros & Cons

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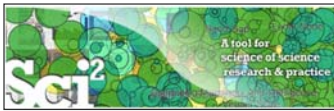
Applications

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

Implementation Details

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

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

Network Visualization with GUESS

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.

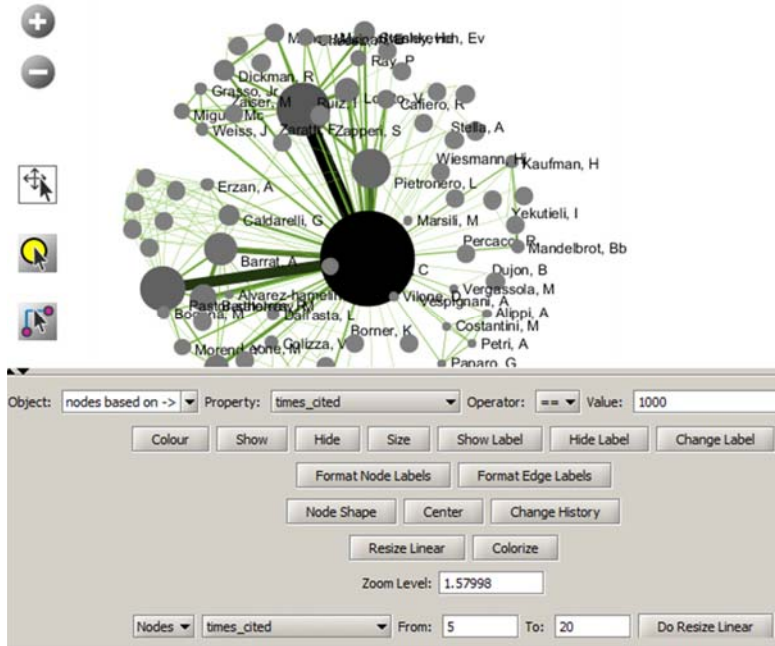Right click node/edge to modify Color, Shape, etc.

Graph Modifier:

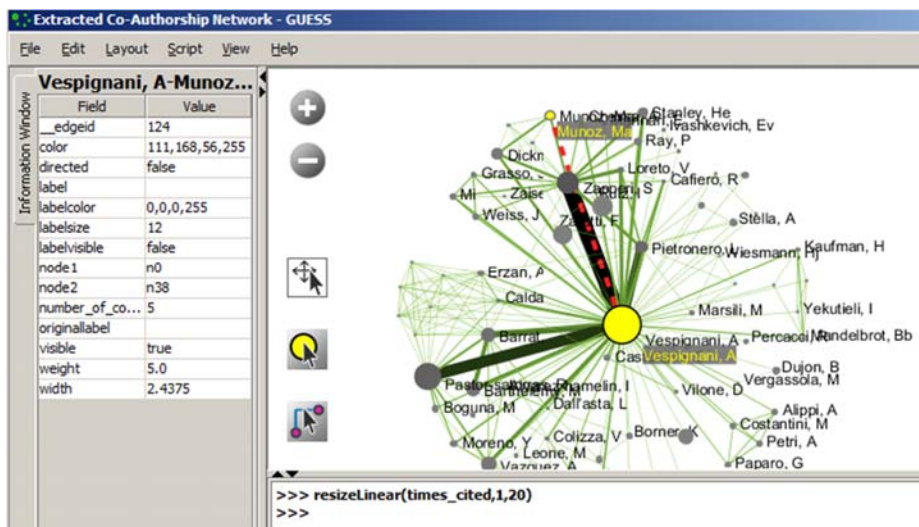Select "all nodes" in the Object drop-down menu and click 'Show Label' button.

Select 'Resize Linear > Nodes > times_cited' drop-down menu, then type "5" and "20" into the From" and To" Value box separately. Then select 'Do Resize Linear'.

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

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

Interpreter uses Jython a combination of Java and Python.

Try
resizeLinear(times_cited,1,20)
colorize(times_cited, white, red)

Tutorial Overview

2p Welcome and Overview of Tutorial and Attendees

2:15p Sci2 Tool Hands-on
- Download and run the Sci2 Tool
- Temporal Analysis: Horizontal line graph of NSF projects
- Geospatial Analysis: US and world maps
- Geospatial Analysis: Geomap with network overlays
- Topical Analysis: Visualize research profiles
- Network Analysis: Co-occurrence networks and bimodal networks
- Network Analysis: Evolving collaboration networks

3:15p Outlook and Q&A

3:30p Adjourn
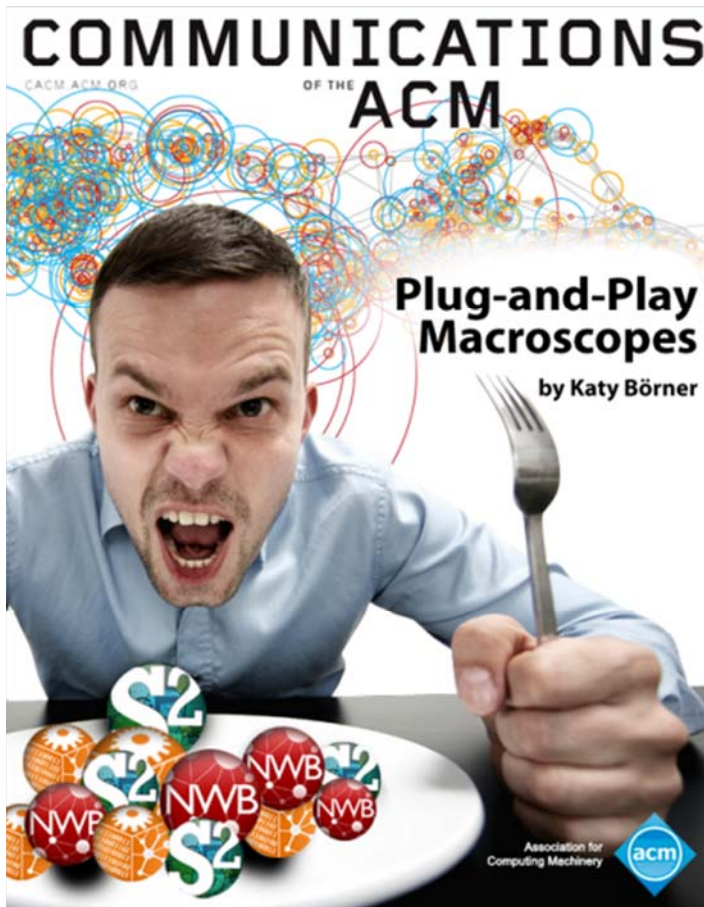
Börner, Katy. (March 2011).

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

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

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

USA

➢ *Cytoscape (http://cytoscape.org)* Led by Trey Ideker at the University of California, San Diego is an open source bioinformatics software platform for visualizing molecular interaction networks and integrating these interactions with gene expression profiles and other state data (Shannon et al., 2002).

➢ *MAEviz (https://wiki.ncsa.uiuc.edu/display/MAE/Home)* Managed by Jong Lee at NCSA is an open-source, extensible software platform which supports seismic risk assessment based on the Mid-America Earthquake (MAE) Center research.

Europe

➢ *Taverna Workbench (http://taverna.org.uk)* Developed by the myGrid team (http://mygrid.org.uk) led by Carol Goble at the University of Manchester, U.K. is a free software tool for designing and executing workflows (Hull et al., 2006). Taverna allows users to integrate many different software tools, including over 30,000 web services.

➢ *TEXTrend (http://textrend.org)* Led by George Kampis at Eötvös Loránd University, Budapest, Hungary supports natural language processing (NLP), classification/mining, and graph algorithms for the analysis of business and governmental text corpuses with an inherently temporal component.

➢ *DynaNets (http://www.dynanets.org)* Coordinated by Peter M.A. Sloot at the University of Amsterdam, The Netherlands develops algorithms to study evolving networks.

➢ *SISOB (http://sisob.lcc.uma.es)* An Observatory for Science in Society Based in Social Models.

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

Network Workbench Tool
http://nwb.slis.indiana.edu

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

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

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

It has been downloaded more than 125,000 times since December 2006.
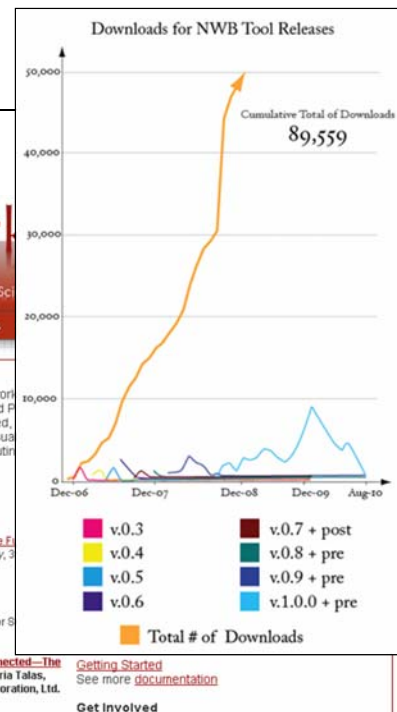
Herr II, Bruce W., Huang, Weixia (Bonnie), Penumarthy, Shashikant & Börner, Katy. (2007). Designing Highly Flexible and Usable Cyberinfrastructures for Convergence. In Bainbridge, William S. & Roco, Mihail C. (Eds.), Progress in Convergence - Technologies for Human Wellbeing (Vol. 1093, pp. 161-179), Annals of the New York Academy of Sciences, Boston, MA.
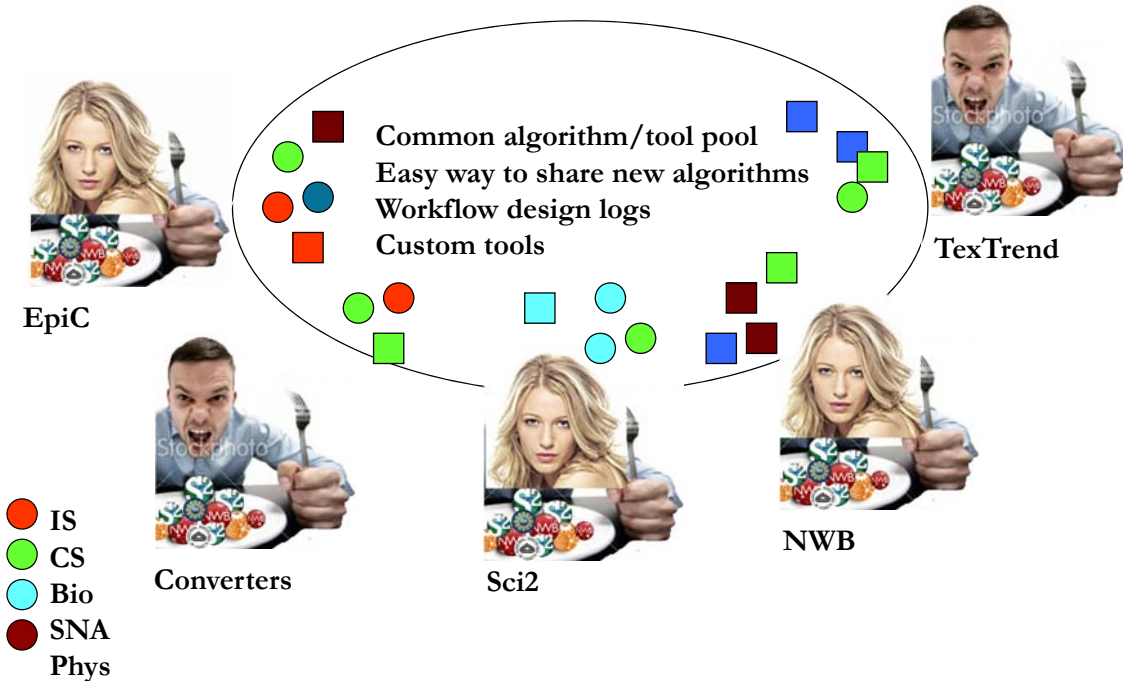
TEXTrend adds R bridge, WEKA, Wordij, CFinder, and more.

See the latest versions of TEXTrend Toolkit modules at

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

EpiC

TexTrend

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

Converters

Sci2

NWB

IS
CS
Bio
SNA
Phys

83

CIShell – Integrate New Algorithms

About the Cyberinfrastructure Shell

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

CIShell is the basis of Network Workbench, TexTrend, Sci² and the upcoming EpiC tool.

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

CIShell Features

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

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

Learn More...

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

Getting Started...

- Documentation & Developer Resources
- Download

Getting Involved...

- Contact Us

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

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

84

> 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
<top_menu name="Modeling">
    <menu pid="edu.iu.nwb.modeling.erdosrandomgraph"/>
    <menu pid="edu.iu.nwb.modeling.smallworld"/>
    <menu pid="edu.iu.nwb.modeling.barabasialbert"/>
    <menu type="break"/>
    <menu pid="edu.iu.iv.modeling.p2p.can.CanAlgorithm"/>
    <menu pid="edu.iu.iv.modeling.p2p.chord.ChordAlgorithm"/>
    <menu pid="edu.iu.iv.modeling.p2p.hypergrid.Hypergrid"/>
    <menu pid="edu.iu.iv.modeling.p2p.pru.PruAlgorithm"/>
    <menu type="break"/>
    <menu pid="edu.iu.iv.modeling.tarl.TarlAlgorithm"/>
    <menu type="break"/>
    <menu pid="edu.iu.nwb.modeling.discretenetworkdynamics.DNDAlgorithm"/>
    <menu type="break"/>
    <menu pid="edu.iu.nwb.modeling.weighted.evolvingnetwork"/>
</top_menu>
```

Need Help? Ask an Expert!

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

If you enjoyed the tutorial, please thank Carole Christian and George Santangelo, NIH as well as Hannah Kim and Mini Saraswati, The Hill Group.
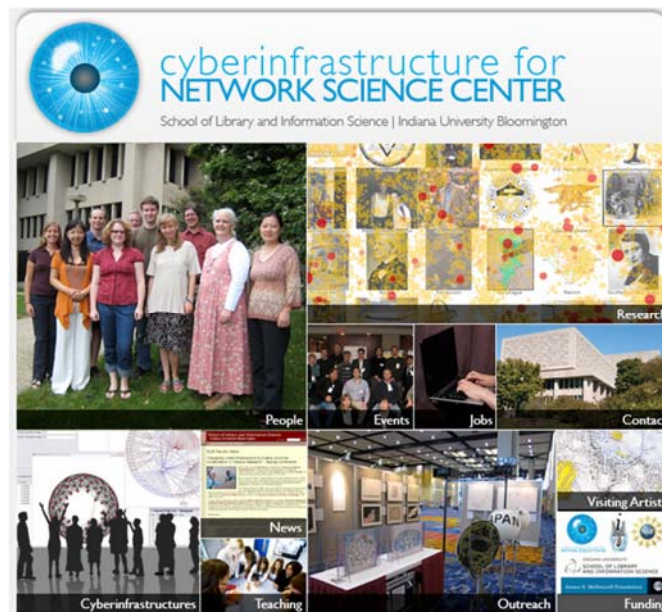
Q & A

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

* * *

Bug reports and all comments are welcome.

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

CNS Facebook: http://www.facebook.com/cnscenter
Mapping Science Exhibit Facebook: http://www.facebook.com/mappingscience