

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

Dr. Katy Börner and Dr. Monika Herzig
Indiana University, Bloomington, Indiana, USA
<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

4th Annual International Science of Team Science Conference
Evanston, IL

Thursday June 27, 2013 • 8:30am-1pm



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

- These slides
<http://ivl.slis.indiana.edu/km/pres/2013-borner-sci2tutorial-scits.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>
- Make sure you have Java 1.6 (32-bit suffices) or higher installed or download from <http://www.java.com/en/download>. To check your Java version, open a terminal and run 'java -version'.
- Some visualizations are saved as Postscript files. A free Postscript to PDF viewer is at <http://ps2pdf.com> and a free PDF Viewer at <http://www.adobe.com/products/reader.html>.





Tutorial Overview

8:30a Welcome and Overview of Tutorial and Attendees

8:45a 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**

10-10:30a Networking Break

- **Topical Analysis: Visualize research profiles**
- **Network Analysis: Co-occurrence networks and bimodal networks**
- **Network Analysis: Evolving collaboration networks**

12:30p IVMOOC

12:45p Outlook and Q&A

1:00p Adjourn

3



Tutorial Overview

8:30a Welcome and Overview of Tutorial and Attendees

8:45a 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**

10-10:30a Networking Break

- **Topical Analysis: Visualize research profiles**
- **Network Analysis: Co-occurrence networks and bimodal networks**
- **Network Analysis: Evolving collaboration networks**

12:30p IVMOOC

12:45p Outlook and Q&A

1:00p Adjourn

4



Using the Sci2 Tool to Visualize Tutorial Registrants

Use *File > Read* to load cleaned *SciTS-Workshop-attendee-info.csv* with 17 records

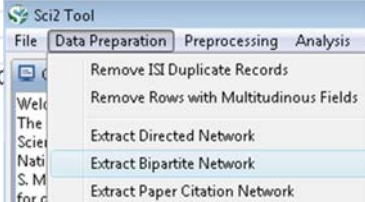
	A	B
1	Last Name	Team Science Role(s)
2	Aydinoglu	Team Science Investigator/Scholar
3	Chien	Team Science Investigator/Scholar
4	Chung	Other
5	Davidson	Other
6	Gunn	Team Science Investigator/Scholar; Agency Program Officer; Research Administrator
7	Hayes	Other
8	Hong	Team Science Practitioner/Leader
9	Huang	Other
10	Kennedy	Team Science Investigator/Scholar
11	Konkol	Other
12	Nodine	Team Science Practitioner/Leader
13	Starren	Other
14	Sugihara	Research Development Professional
15	Wagner	Team Science Investigator/Scholar
16	Whetsell	Team Science Investigator/Scholar
17	Williams	Research Development Professional
18	Yamamoto	Research Administrator

5



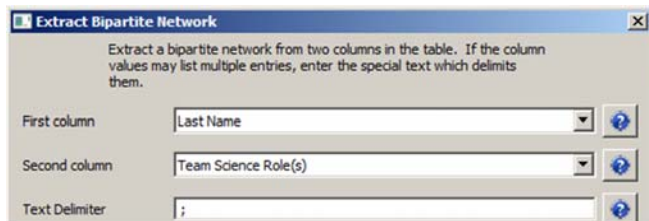
Using the Sci2 Tool to Visualize Tutorial Registrants

	A	B
1	Last Name	Team Science Role(s)
2	Aydinoglu	Team Science Investigator/Scholar
3	Chien	Team Science Investigator/Scholar
4	Chung	Other
5	Davidson	Other
6	Gunn	Team Science Investigator/Scholar; Agency Program Officer; Research Administrator



Run *Data Preparation > Extract Bipartite Network*

With parameter values:



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: Last Name

Second column: Team Science Role(s)

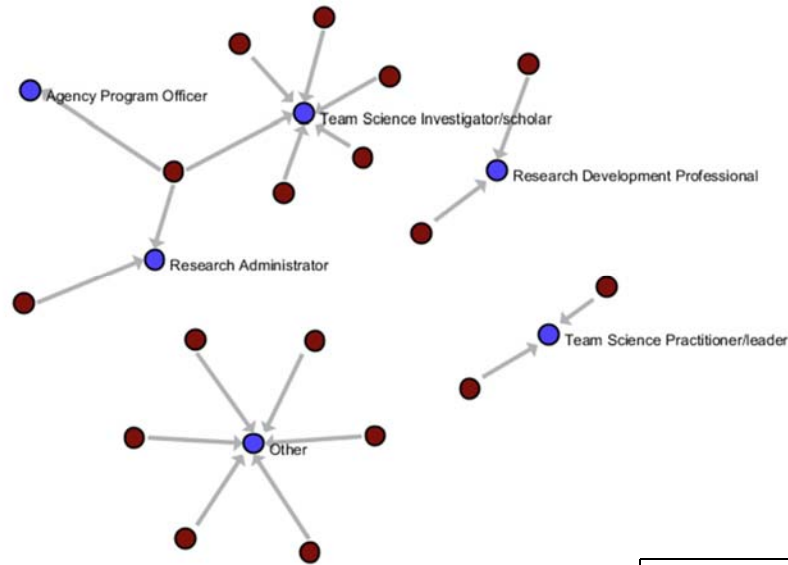
Text Delimiter: ;

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

6

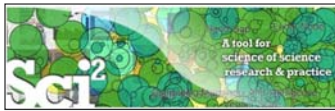


Sci2 Tool – Visualize Tutorial Attendees: Bi-partite Name-Role Network

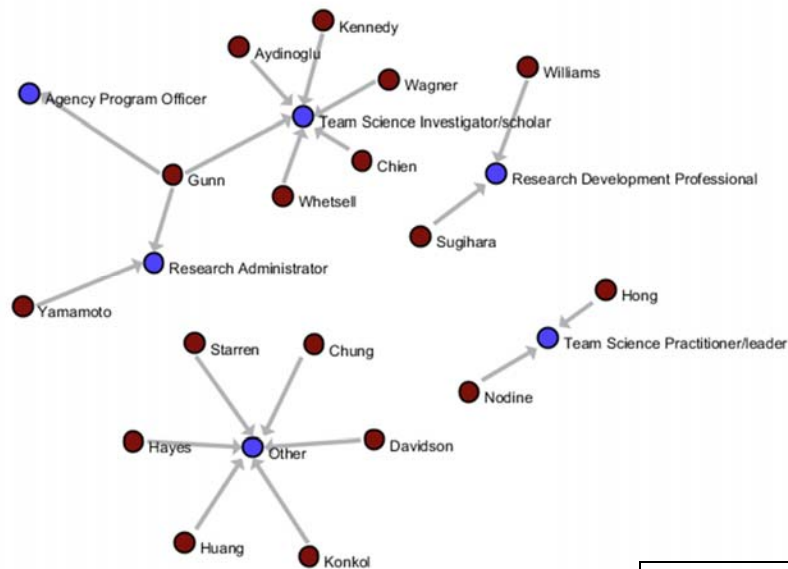


- Role
- Name

7



Sci2 Tool – Visualize Tutorial Attendees: Bi-partite Name-Role Network



- Role
- Name

8



Tutorial Overview

8:30a Welcome and Overview of Tutorial and Attendees

8:45a 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

10-10:30a Networking Break

- Topical Analysis: Visualize research profiles
- Network Analysis: Co-occurrence networks and bimodal networks
- Network Analysis: Evolving collaboration networks

12:30p IVMOOC

12:45p Outlook and Q&A

1:00p Adjourn

9



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**.

10



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

11



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>

12

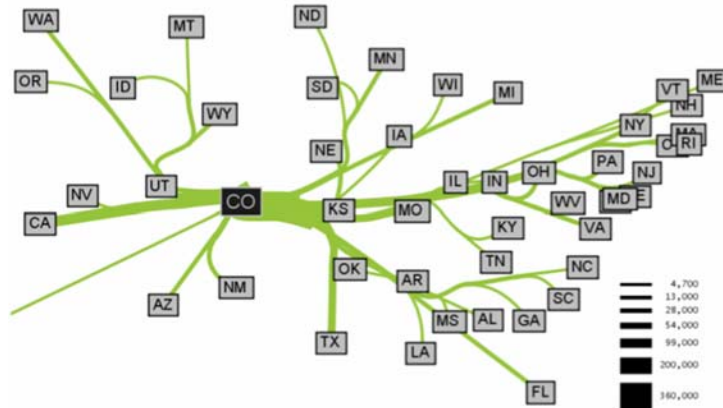


Sci2 Tool v1.1 Alpha (planned for August 2013)

New Features

- Twitter, Facebook, and Flickr readers
- Bing Geocoder
- Flow map visualization, see below
- Comprehensive online documentation

Bug fixes



13

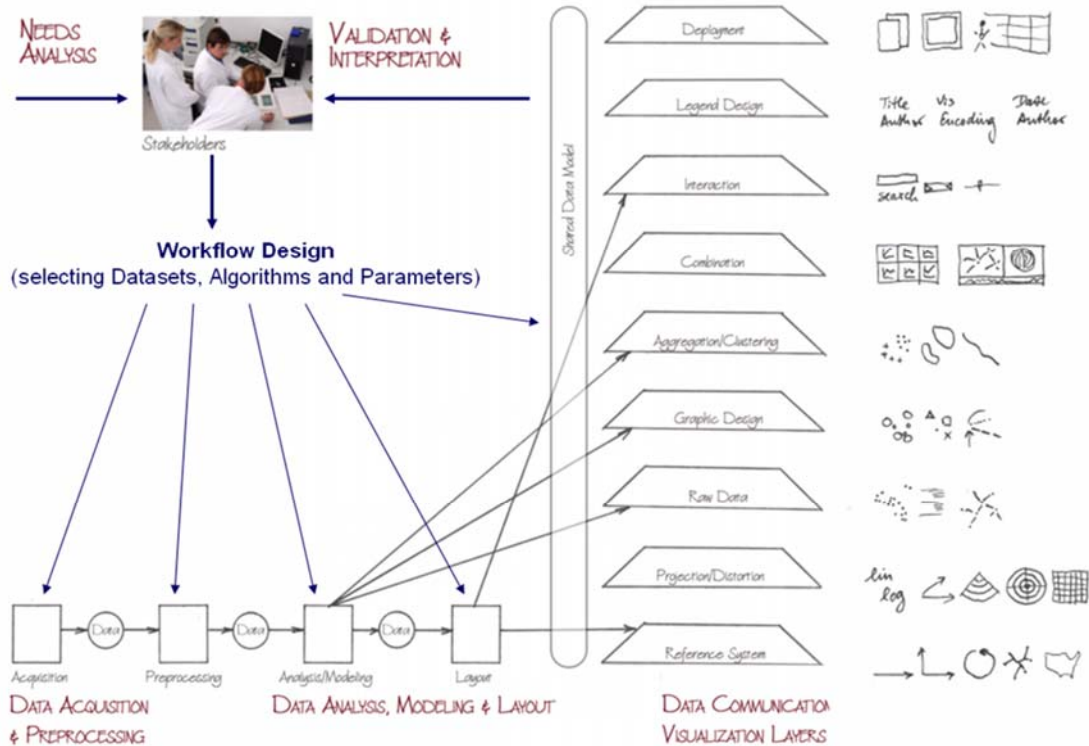


Type of Analysis vs. Level of Analysis

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

14

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



Börner, Katy (2010) *Atlas of Science*. MIT Press. 15



Sci² Tool – Supported Data Formats

Input:

Network Formats

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

Scientometric Formats

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

Other Formats

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

Output:

Network File Formats

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

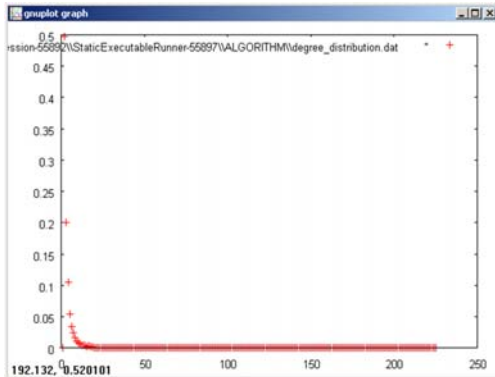
Image Formats

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

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

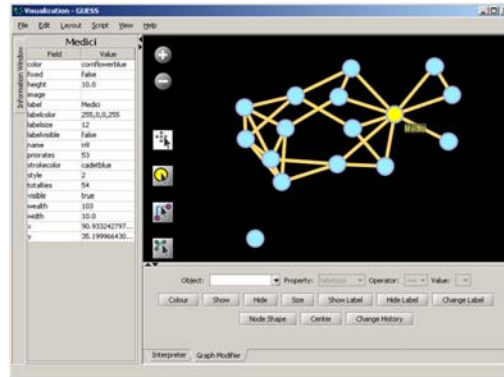


Sci² Tool – Supported Tools



Gnuplot

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



GUESS

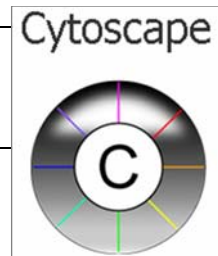
exploratory data analysis and visualization tool
for graphs and networks.

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

17



Sci² Tool – Supported Tools

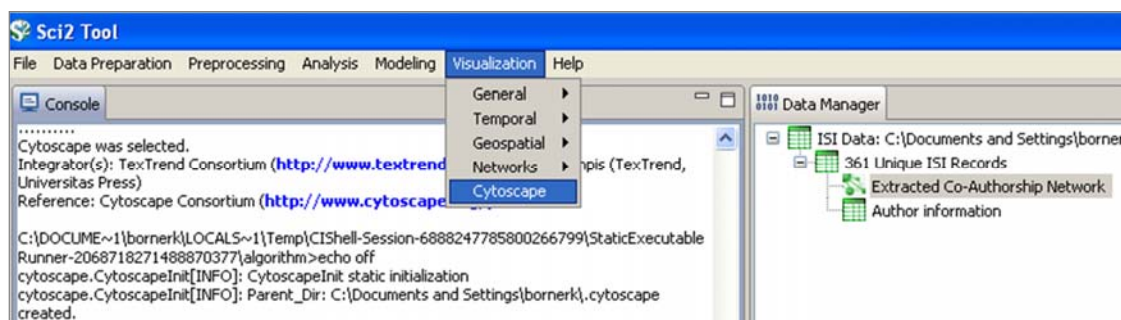


Adding more layout algorithms and network visualization interactivity
via Cytoscape <http://www.cytoscape.org>.

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

Restart Sci² Tool

Cytoscape now shows in the Visualization Menu



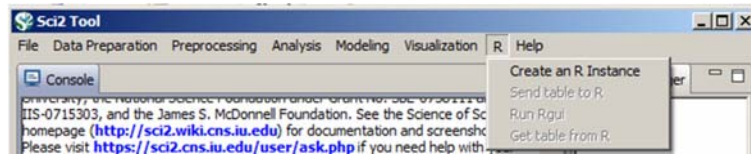
Select a network in Data Manager, run Cytoscape and the tool will start with this
network loaded.

18

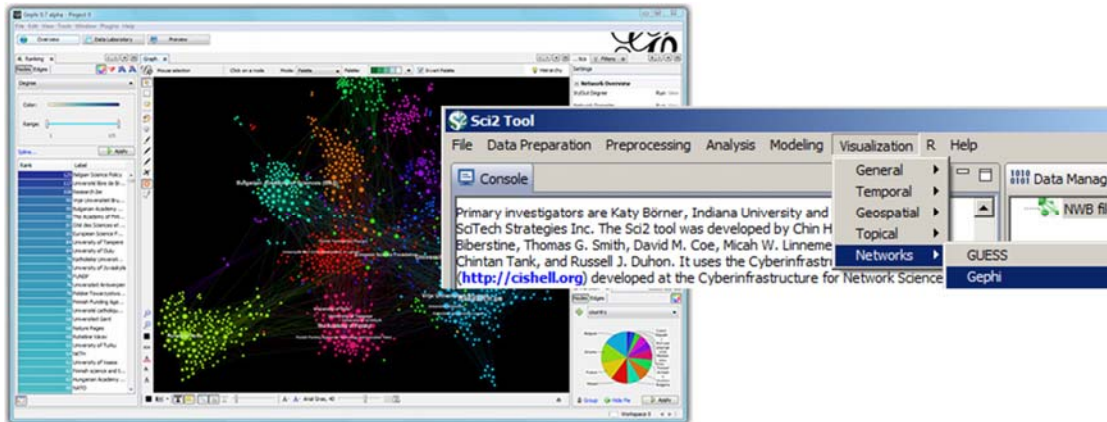


Sci² Tool – Bridged Tools

R statistical tool bridging



Gephi visualization tool bridging



19



Sci² Tool: Download, Install, and Run

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.

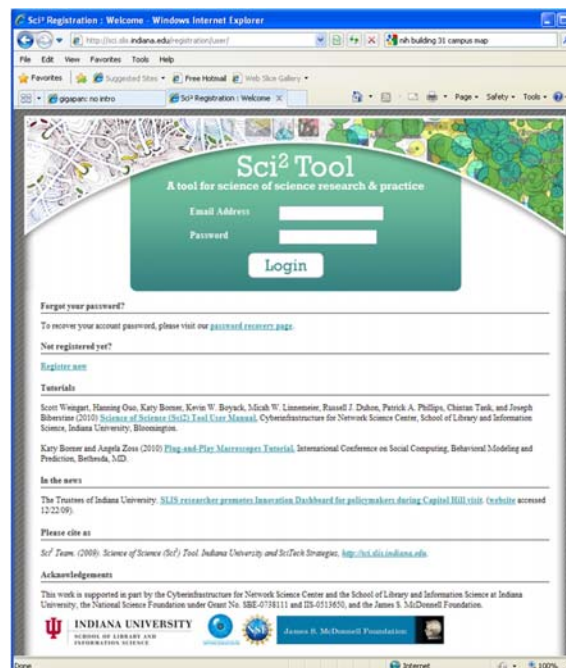
Run /sci2/sci2.exe

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>



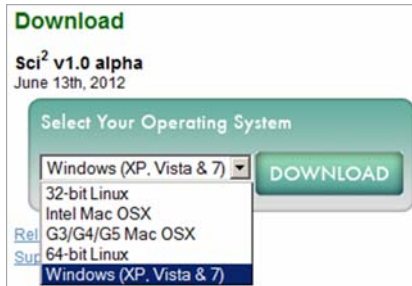
20



Sci² Tool: Download, Install, and Run

Sci² Tool v1.0 Alpha (June 13, 2012)

Sci² Tool runs on Windows, Mac, and Linux.



Name	Size
sci2-N-1.0.0.201206130117NGT-linux.gtk.x86_64.zip	128,231 KB
sci2-N-1.0.0.201206130117NGT-macosx.carbon.x86.zip	131,160 KB
sci2-N-1.0.0.201206130117NGT-win32.win32.x86.zip	131,373 KB
Adobe-PDF-Reader.exe	1,013 KB
test.ps	2 KB

Unzip.

Run `/sci2/sci2.exe`



21

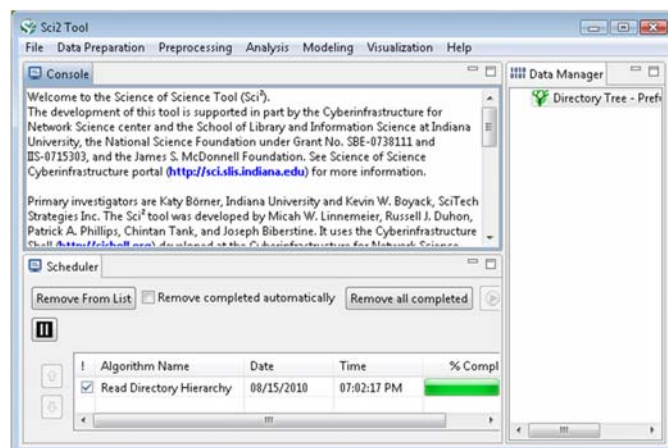


Sci² Tool Interface Components

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>

22



Tutorial Overview

8:30a Welcome and Overview of Tutorial and Attendees

8:45a 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

10-10:30a Networking Break

- Topical Analysis: Visualize research profiles
- Network Analysis: Co-occurrence networks and bimodal networks
- Network Analysis: Evolving collaboration networks

12:30p IVMOOC

12:45p Outlook and Q&A

1:00p Adjourn

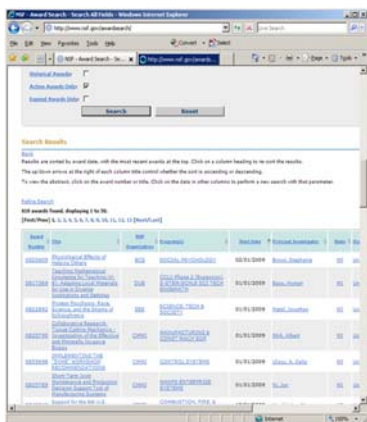
23



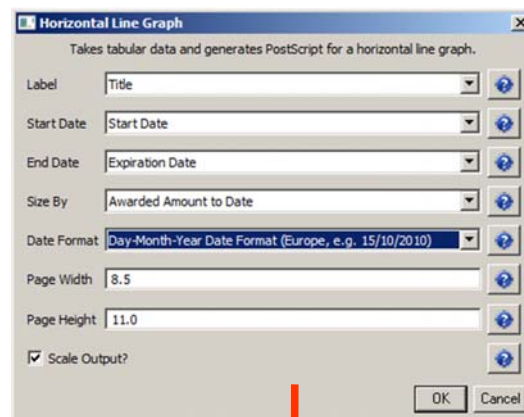
Horizontal line graph of NSF projects

See [5.2.1 Funding Profiles of Three Universities \(NSF Data\)](#)

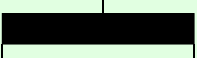
Download NSF data





Visualize as Horizontal Line Graph



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

Text 

Start date  End date 



24



Horizontal line graph of NSF projects

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

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

Award Number	Title	Agency	Program	Start Date	PI Name	PI Affiliation
0822603	Indiana Cyber Teaching Mathematical Knowledge for Teachers (I-CyT) - An Advanced Local Interest in Use in Schools	NSF	DOL-Phase 2 (Expansion)-STEM SOURCE TECH EDUCATION	01/01/2009	Barb Hunter	ISI
0817265	Advanced Process Technology for the Chemical Industry	NSF	MANUFACTURING & CONSTRUCTION	01/01/2009	Mark J. Heule	ISI
0822852	Collaborative Research: Tissue Culture Structures - Investigation of the Effects and Molecular Mechanisms	NSF	SCIENCE, TECH & SOCIETY	01/01/2009	Hani Jazayeri	ISI
0822735	Implementing the "Home" Workforce	NSF	MANUFACTURING & CONSTRUCTION	01/01/2009	Shih Albert	ISI
0822638	Implementing the "Home" Workforce	NSF	CONTROL SYSTEMS	01/01/2009	Ulises A. Gels	ISI
0822739	Support for the 4th U.S.	NSF	COMBUSTION, FIRE & ...	01/01/2009	Hi-Jae	ISI

25



Temporal bar graph of NSF projects

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

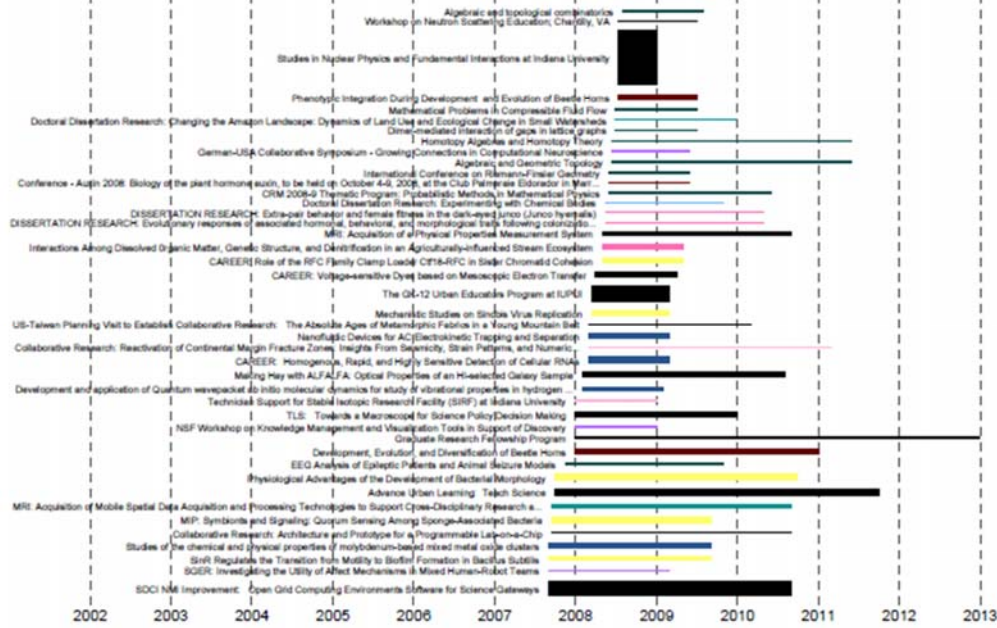
Run *'Visualization > Temporal > Temporal Bar Graph'* using parameters:

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.

26

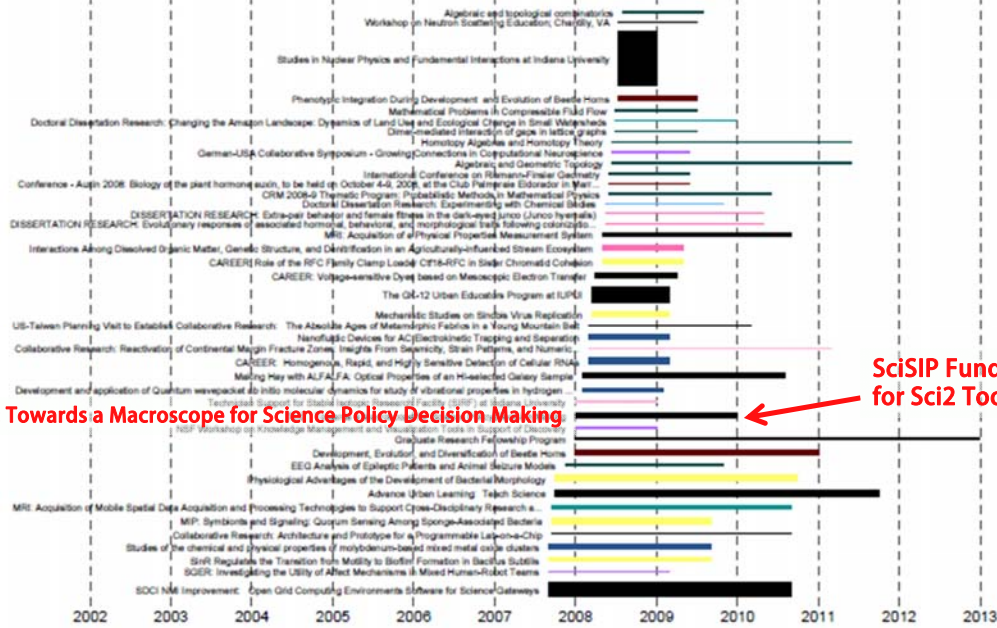
Temporal Visualization

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



Temporal Visualization

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

Temporal Bar Graph

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

Subtitle:

Label:

Start Date:

End Date:

Size By:

Date Format:

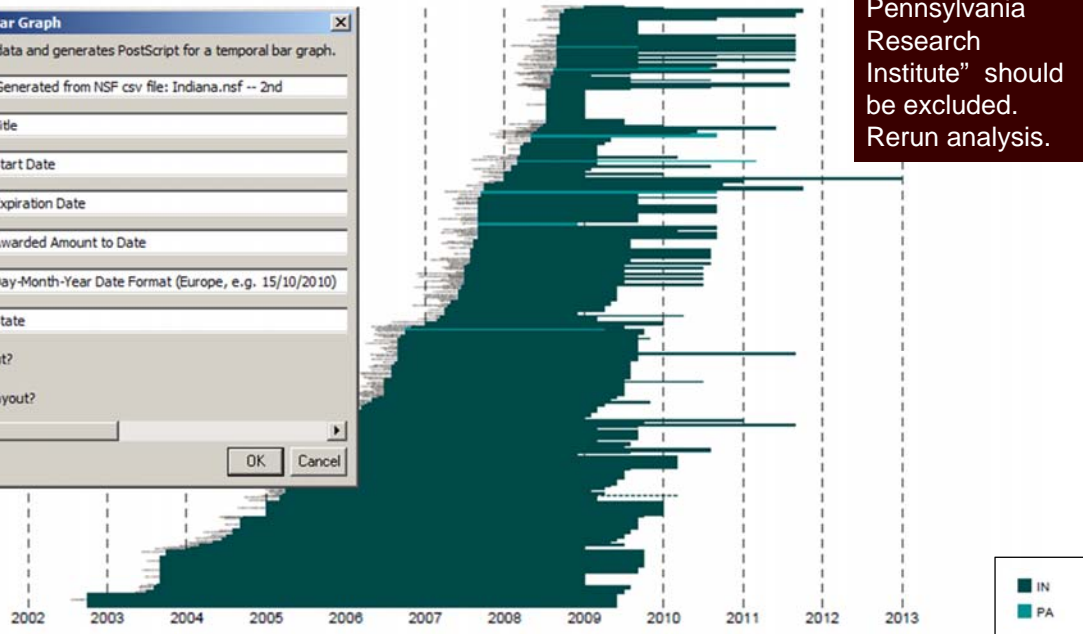
Category:

Scale Output?

Simplified Layout?

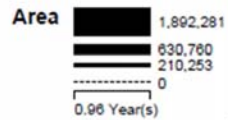
OK Cancel

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



Legend

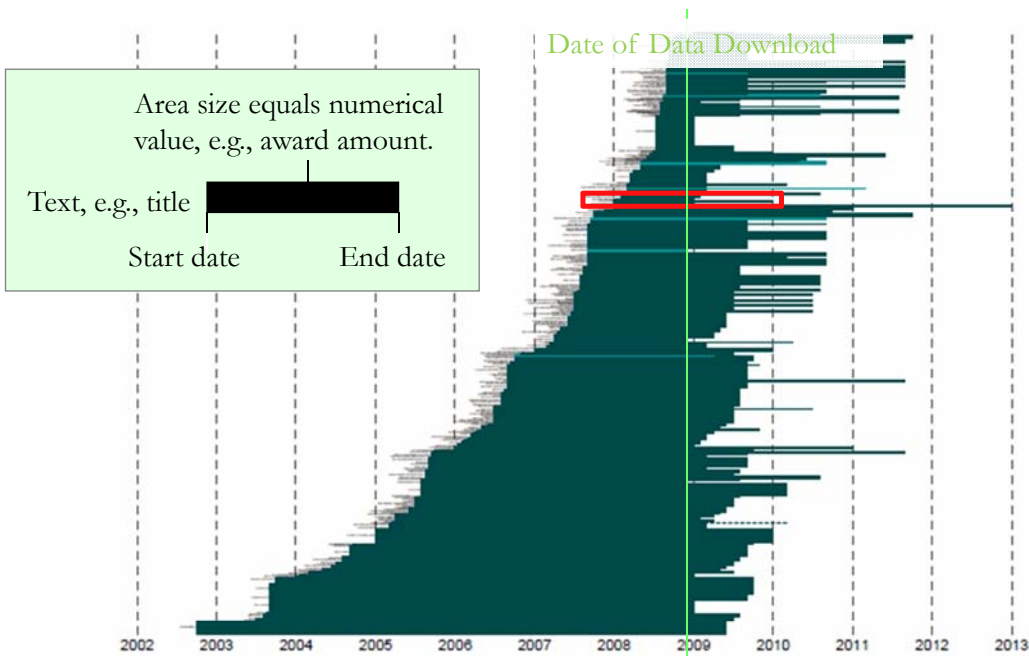
Area size: Awarded Amount to Date
Minimum = 0
Maximum = 6,402,330
Text label: Title
Color: State
See end of PDF for color legend.



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.

GNS (cns.iu.edu)



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

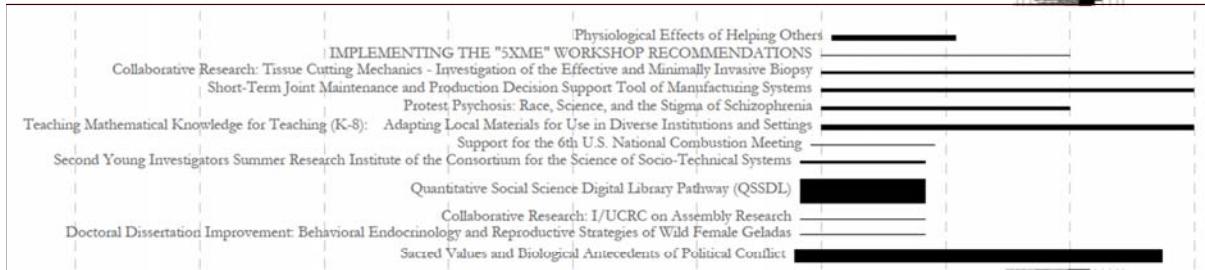
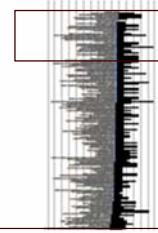
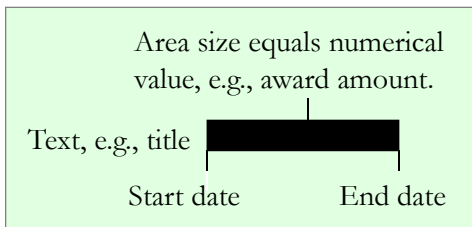
TLS: Towards a Macroscopic for Science Policy Decision Making

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

Graduate Research Fellowship Program



Temporal bar graph of NSF projects



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 \(NSF Data\)](#)



31



Tutorial Overview

8:30a Welcome and Overview of Tutorial and Attendees

8:45a 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

10-10:30a Networking Break

- Topical Analysis: Visualize research profiles
- Network Analysis: Co-occurrence networks and bimodal networks
- Network Analysis: Evolving collaboration networks

12:30p IVMOOC

12:45p Outlook and Q&A

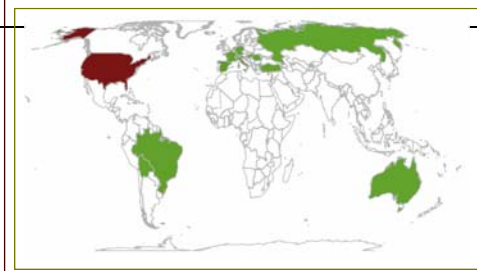
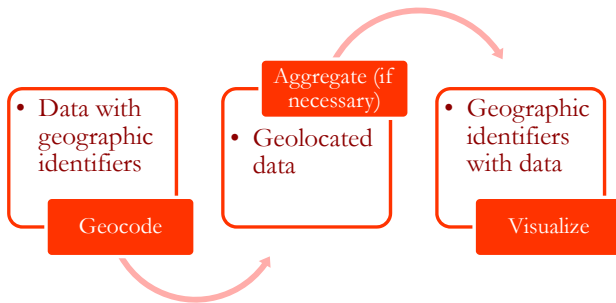
1:00p Adjourn

32

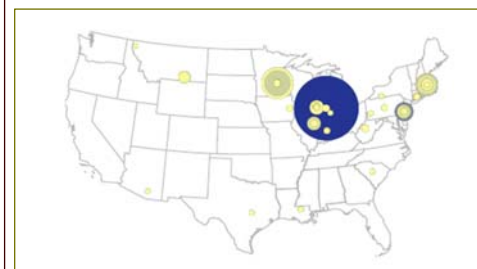


Geocoding and Geospatial Maps

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

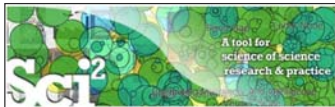


Region names + numeric data
(Choropleth Map)



Geocoordinates + numeric data
(Proportional Symbol Map)

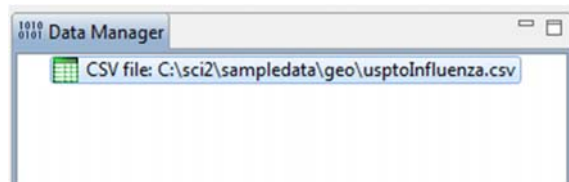
33



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



34



Yahoo Geocoder



Edit Add Tools ▾

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

Description

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

Pros & Cons

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

Applications

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

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

35



Yahoo Geocoder



Edit Share Add Tools ▾

Added by [Mayur Masrani](#), last edited by [Mayur Masrani](#) on Apr 24, 2013

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 Bing geocoder service. Internet access must be available during geocoding.

Pros & Cons

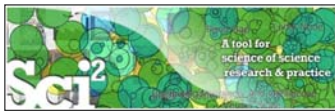
1. The performance is slower than the [Geocoder](#) and may vary due to the network latency since the queries are requested through internet service.
2. Bing Geocoder supports address geocoding with international coverage which is not supported by [Geocoder](#).
3. To use Bing Geocoder, user has to obtain an API Keys from [Bing Maps](#). Save your api keys and provide it when requested by the Bing Geocoder. Since each api key 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 (see [Geospatial Visualization](#)). User can obtain the geographical coordinates (Latitude and Longitude values) and feed them to the visualization plugin.

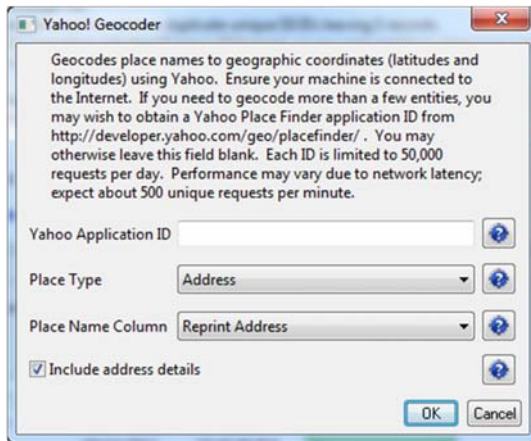
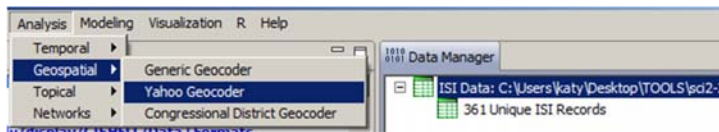
<http://wiki.cns.iu.edu/display/CISHELL/Bing+Geocoder>

36



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/>

37



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

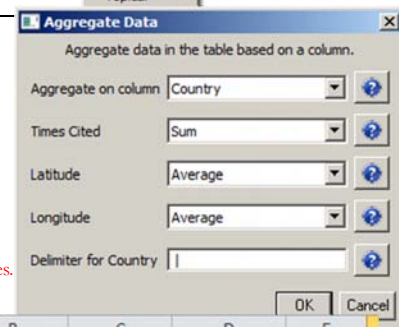
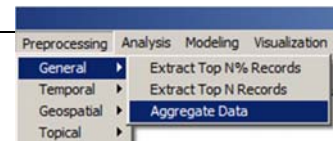
Latitude: AVERAGE

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.

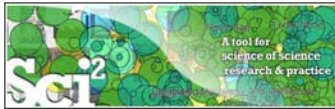


	A	B	C	D
1	Times Cited	Latitude	Longitude	Country
2	7	42.02946091	-87.68838501	United States
3	0			
4	0			
5	2	42.34999466	-71.08765411	United States
6	14	41.70074844	-86.23918915	United States
7	15	41.70074844	-86.23918915	United States
8	29	41.89422607	-87.61901855	United States
9	32	41.70074844	-86.23918915	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

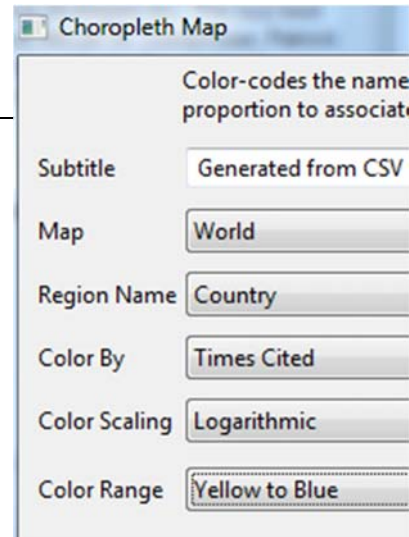
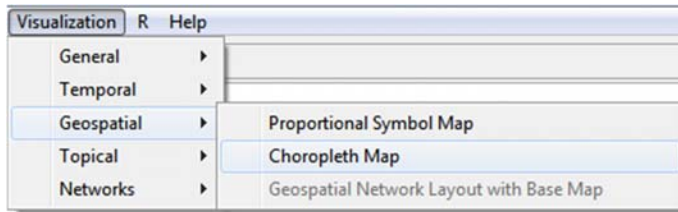


	A	B	C	D	E
1	Times Cited	Latitude	Longitude	Country	Count
2	14680	[41.10645f]	[-82.45309f]	United States	194
3	1802				57
4	398	[47.506226f]	[19.06482f]	Hungary	14
5	101	[37.25198f]	[127.08451f]	South Korea	4
6	18	[32.08439f]	[34.81297f]	Israel	1
7	57	[46.768517f]	[23.585135f]	Romania	2
8	55	[47.06615f]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

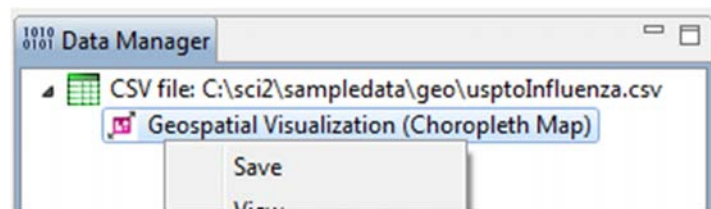
38



Choropleth Map



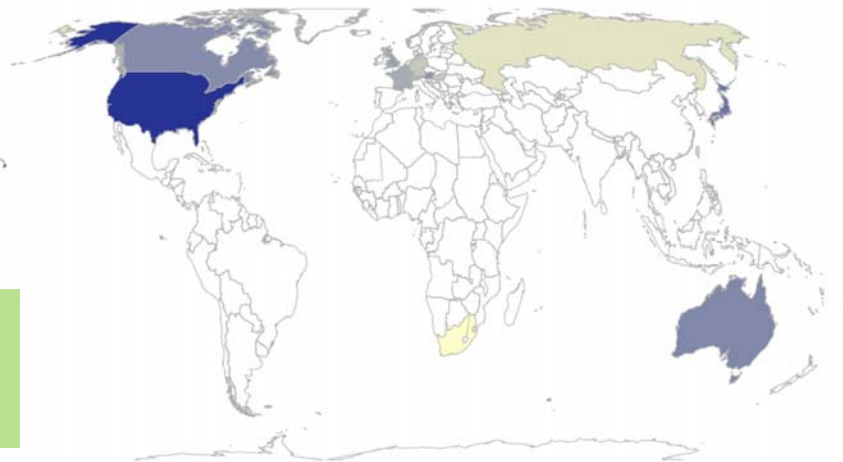
Right-click and **Save** map as PostScript file. Use PostScript Viewer or convert to pdf to view.



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



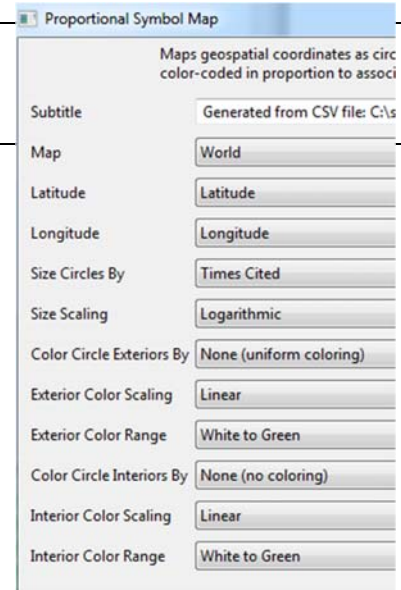
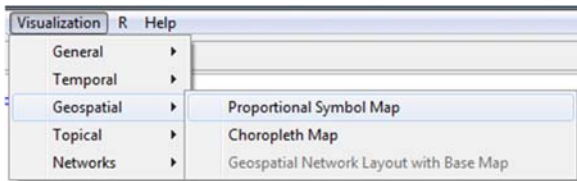
Legend shows how data matches up with visual representation



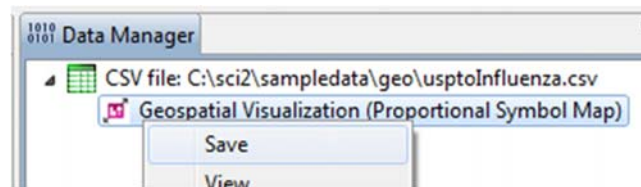
How to Read this Map
This choropleth map shows 209 countries of the world using the equal-area Eckert IV projection. Each country may be color coded in proportion to a numerical value. Minimum and maximum data values are given in the legend.



Proportional Symbol Map



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



41



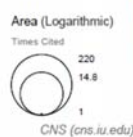
Reading the Proportional Symbol Map

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



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.

42



Relevant Sci2 Manual entry



5.2.4 Mapping Scientometrics (ISI Data)

Edit Add Tools

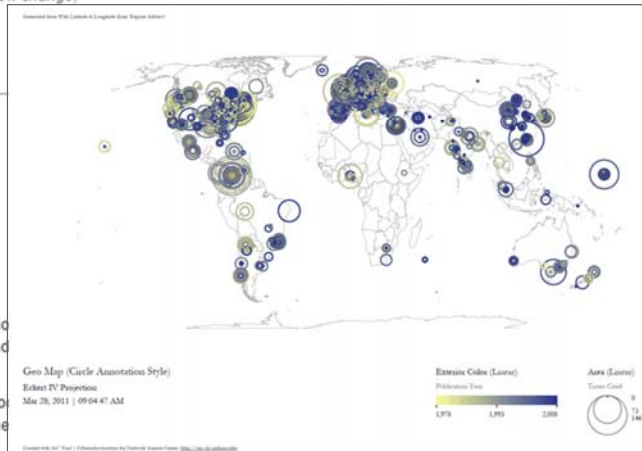
9 Added by Ted Polley, last edited by Ted Polley on Nov 14, 2011 (view change)

5.2.4.1 Document Co-Citation

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

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

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



New ISI File Format

Web of Science made a change to their output format in September, 2011. Older versions of Sci2 tool may refuse to load these new files, with an error like "Invalid ISI format file selected."

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

43



Tutorial Overview

8:30a Welcome and Overview of Tutorial and Attendees

8:45a 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

10-10:30a Networking Break

- Topical Analysis: Visualize research profiles
- Network Analysis: Co-occurrence networks and bimodal networks
- Network Analysis: Evolving collaboration networks

12:30p IVMOOC

12:45p Outlook and Q&A

1:00p Adjourn

44

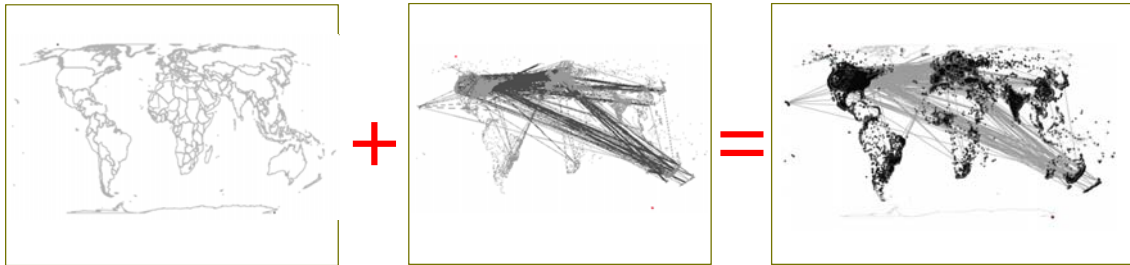


Geomap with Gephi Network Overlay

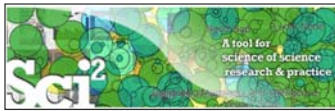
See 4.7.6 on <http://sci2.wiki.cns.in.edu>

File with geolocations and linkage info, e.g., an isi bibliography 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-occurrences
See sample /geo/LaszloBarabasiGeo.net with co-occurrence of “Research Addresses” and full counting of TC per geolocation.

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



45



Relevant Sci2 Manual entry

Dashboard > Sci2 Manual > ... > 4 Workflow Design > 4.7 Geospatial Analysis (Where) Browse Log In Search Confluence

Search

- Home
- 1 Introduction
- 2 Getting Started
- 3 Algorithms, Tools, and Plugins
- 4 Workflow Design
 - 4.1 Overview
 - 4.2 Data Acquisition and Preparation
 - 4.3 Database Loading and Manipulation
 - 4.4 Summaries and Table Extractions
 - 4.5 Statistical Analysis and Profiling
 - 4.6 Temporal Analysis (When)
 - 4.7 Geospatial Analysis (Where)
 - 4.8 Topical Analysis (What)
 - 4.9 Network Analysis (With Whom?)
 - 4.10 Modeling (Why?)
- 5 Sample Workflows
- 6 Sample Science Studies & Online Services
- 7 Extending the Sci2 Tool
- 8 Relevant Datasets and Tools
- 9 References

4.7.6 Using Gephi to Render Networks Overlaid on Geo Maps

Loading and Saving Geovisualization Files in Sci2

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

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

Geo Maps (network template)

Creates a world map, and processes the input network so that latitude

Map: Countries

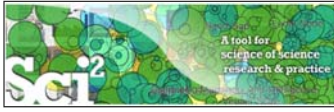
Latitude: ypos

Longitude: xpos

OK Cancel

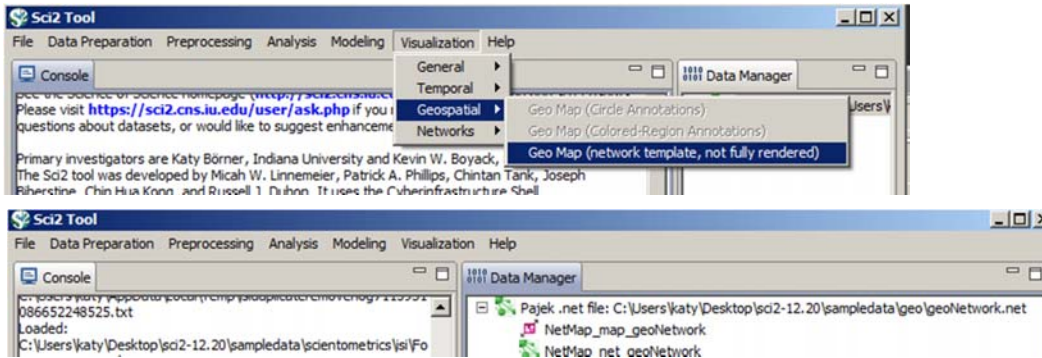
<http://sci2.wiki.cns.in.edu/display/SCI2TUTORIAL/4.7+Geospatial+Analysis+%28Where%29#4.7GeospatialAnalysis%28Where%29-4.7.6UsingGephitoRenderNetworksOverlaidonGeoMaps>

46



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.

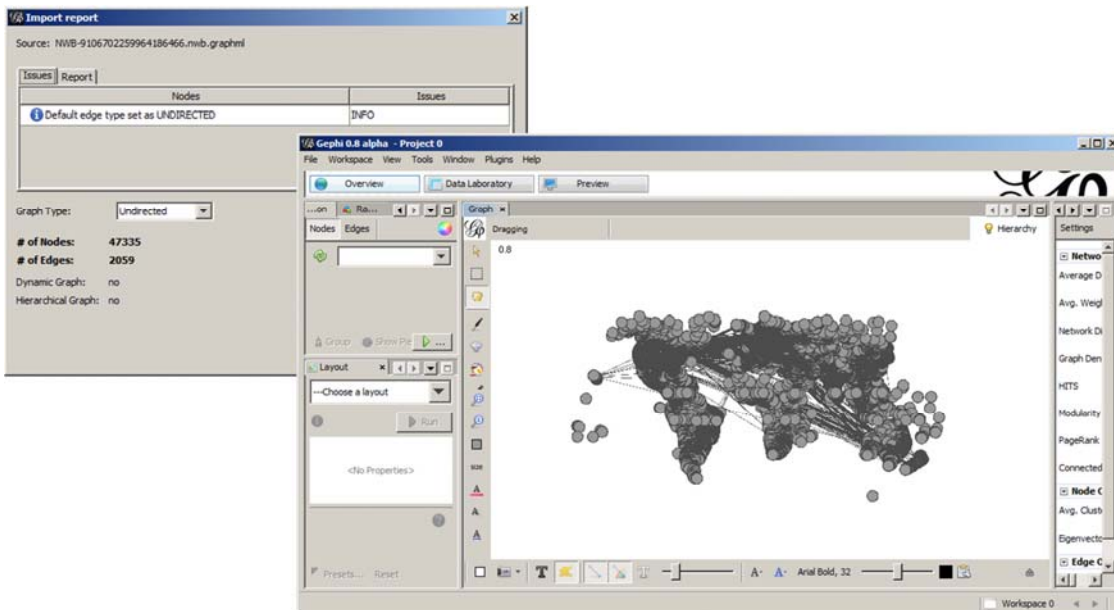


47



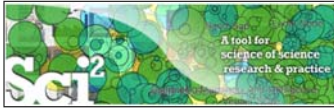
Use Gephi to Generate Network Layout

Start gephi. Use *New Project > Open a graph file* to read .graphml file that Sci2 generated.



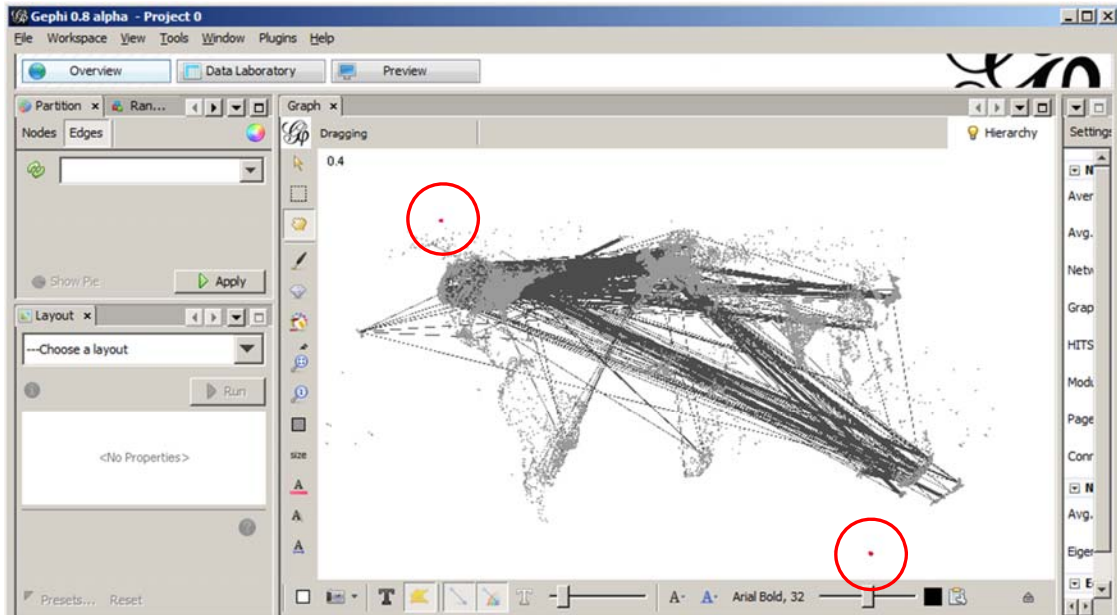
Follow instructions in online tutorial on **Manipulating the Network File in Gephi**

48

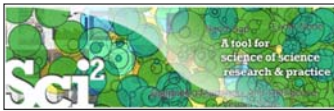


Use Gephi to Generate Network Layout

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*.

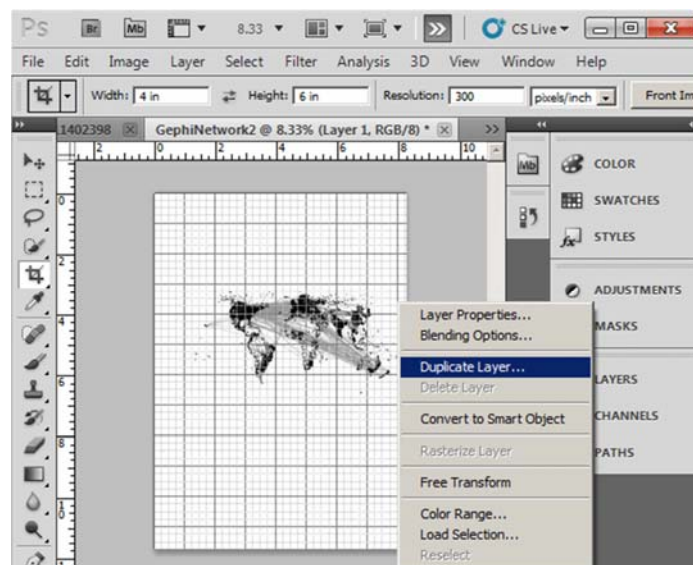


49

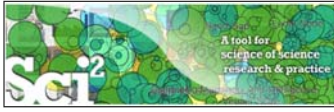


Use Photoshop to Overlay Network on Geomap

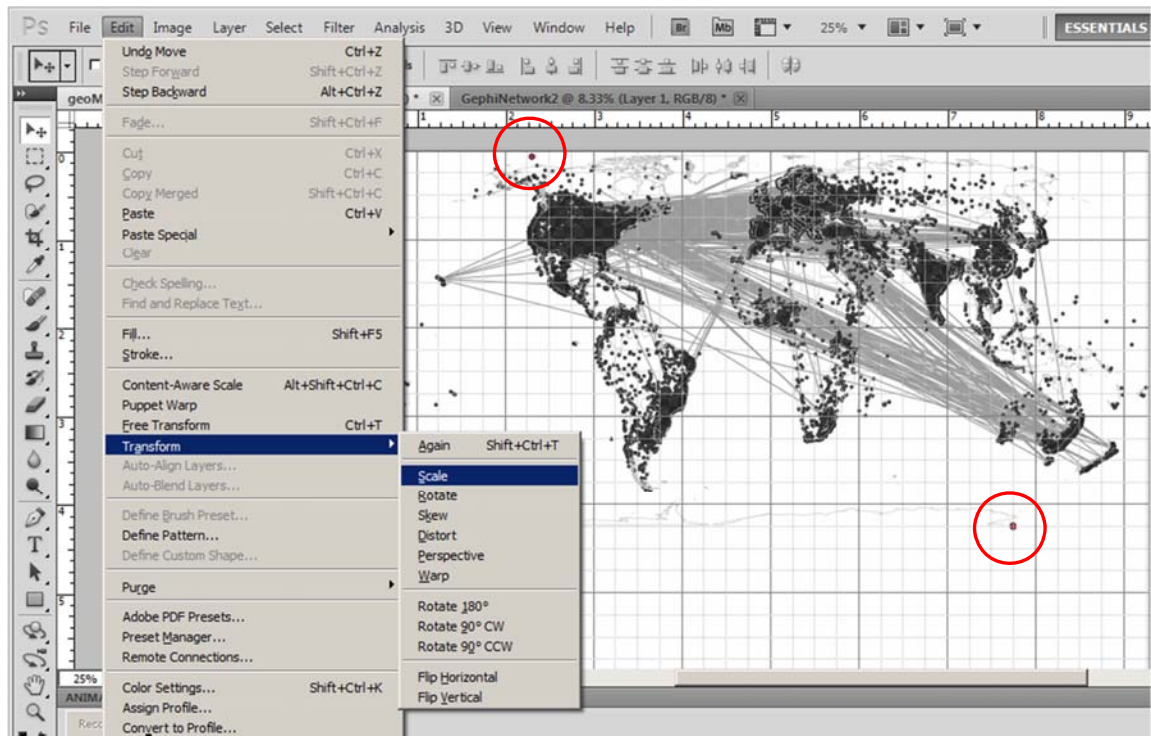
Load geomap and network files into Photoshop. Select ‘network’ layer and use ‘Right click, Duplicate Layer’ to copy network over to ‘geomap’ file as a second layer. Use *Edit > Transform > Scale* and align using the “Near Antarctica” anchor nodes, see instructions in online tutorial on **Creating the Visualization in Photoshop**.



50

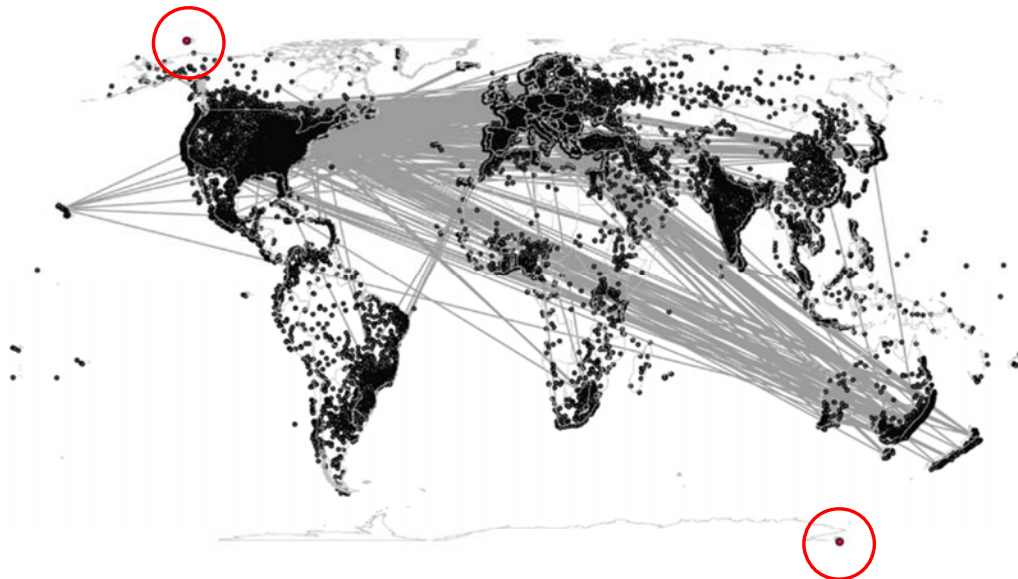


Use Photoshop to Overlay Network on Geomap



51

Delete anchor nodes and save in preferred format.

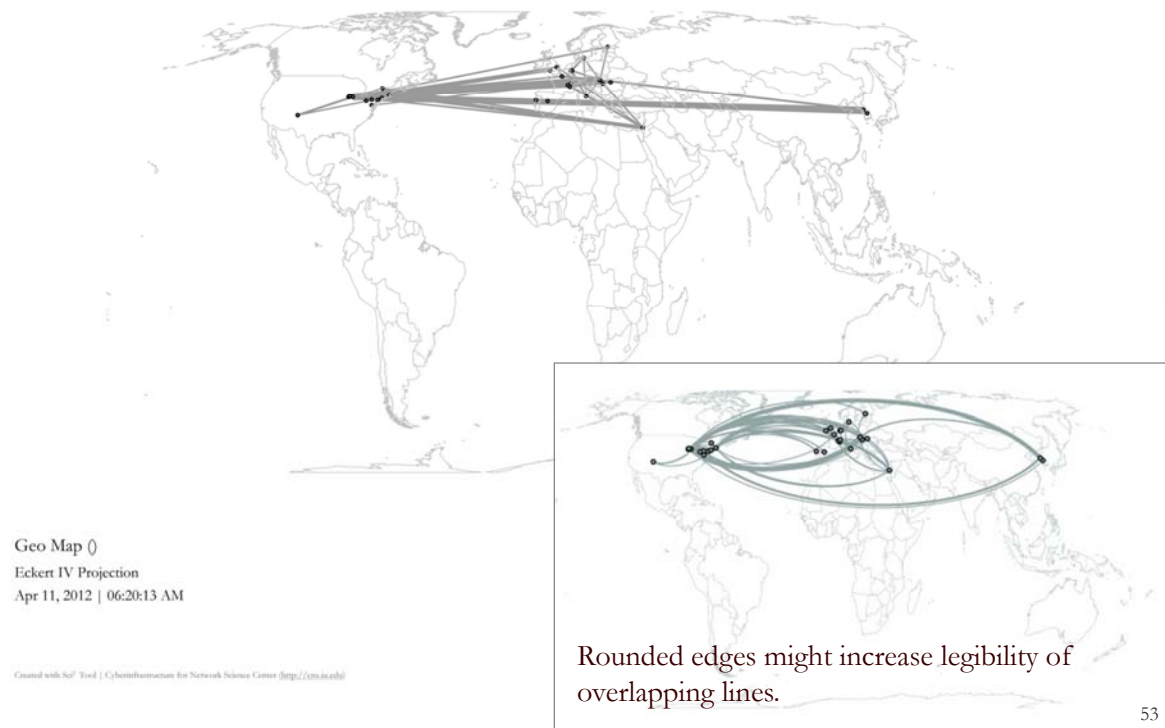


Geo Map ()
Eckert IV Projection
Apr 06, 2012 | 03:19:51 AM

52

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

4.7.6 Using Gephi to Render Networks Overlaid on GeoMaps



Tutorial Overview

8:30a Welcome and Overview of Tutorial and Attendees

8:45a 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

10-10:30a Networking Break

- **Topical Analysis: Visualize research profiles**
- Network Analysis: Co-occurrence networks and bimodal networks
- Network Analysis: Evolving collaboration networks

12:30p IVMOOC

12:45p Outlook and Q&A

1:00p Adjourn

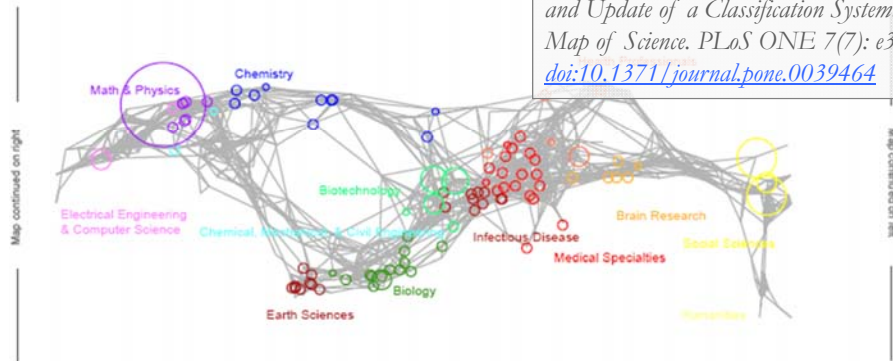


Topical Analysis: Research Profiles

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



2009 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



How To Read This Map

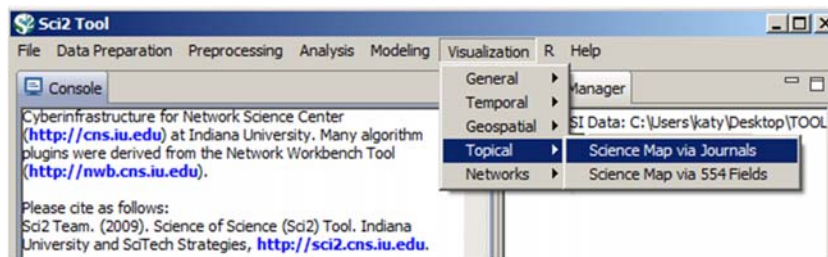
The UCSD map of science depicts a network of 554 subdiscipline nodes that are aggregated to 13 main disciplines of science. Each discipline has a distinct color and is labeled. Overlaid are circles, each representing all records per unique subdiscipline. Circle area is proportional to the number of fractionally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

CNS (cns.iu.edu)



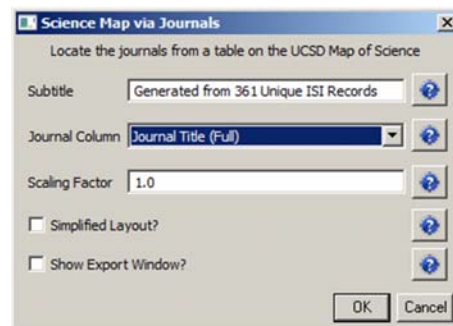
Research Profiles—Publication Data

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



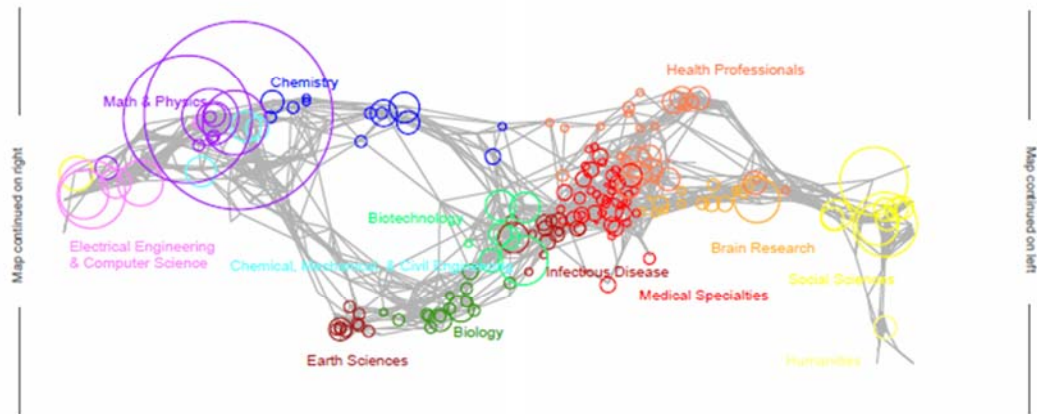
Run `'Visualization > Topical > Science Map via Journals'` using parameters given to the right.

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



Topical Visualization

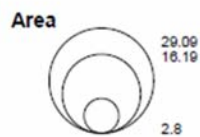
Generated from 361 Unique ISI Records
90 out of 112 publications were mapped to 182 subdisciplines and 13 disciplines.
June 24, 2012 | 04:04 PM EDT



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 = 22
Minimum = 0
Maximum = 98
Color: Discipline
See end of PDF for color legend.



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

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

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

CNS (cns.iu.edu)

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

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
3 CURRENT CONTENTS/LIFE SCIENCES
1 FEDERATION PROCEEDINGS
5 FRACTALS-AN INTERDISCIPLINARY JOURNAL ON THE COMPLEX GE...
1 FRONTIERS OF LIBRARIANSHIP-SYRACUSE UNIVERSITY

CNS (cns.iu.edu)



Research Profiles—Existing Classifications

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	C	D	E	F	G
KNOWLEDGE AREA	NO. Projects	USDA Staff Years	STATE APPR	TOTAL FUNDS	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

Science Map via 554 Fields (Circle Annotations) [X]

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?

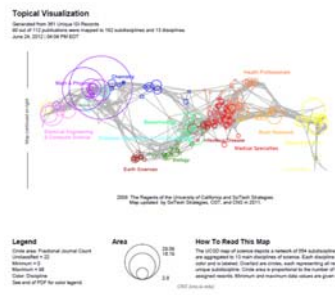
Run *Visualization > Topical > Science Map via 554 Fields* using parameters given to the right.

Postscript file will appear in *Data Manager*.

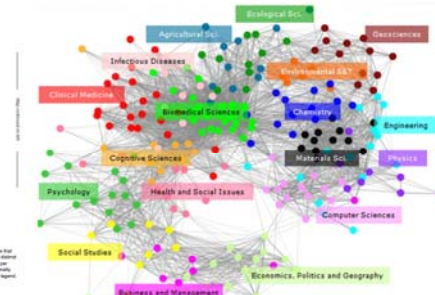
Save and open with a Postscript Viewer.



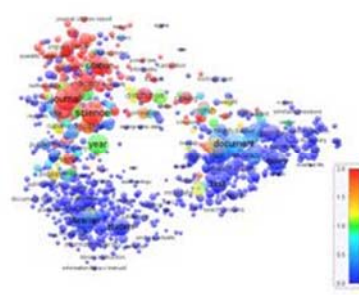
Align Science Basemaps using the Sci2 Tool



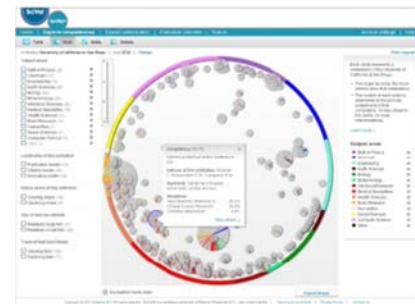
UCSD Map



Loet et al science maps ISI categories



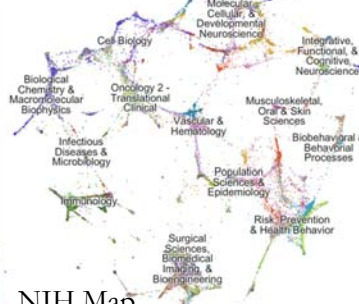
<http://vosviewer.com>



Elsevier's SciVal Map



Science-Metrix.com



NIH Map
<https://app.nihmaps.org>

61



Tutorial Overview

8:30a Welcome and Overview of Tutorial and Attendees

8:45a 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

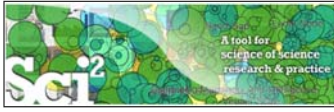
10-10:30a Networking Break

- Topical Analysis: Visualize research profiles
- **Network Analysis: Co-occurrence networks and bimodal networks**
- Network Analysis: Evolving collaboration networks

12:30p IVMOOC

12:45p Outlook and Q&A

1:00p 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



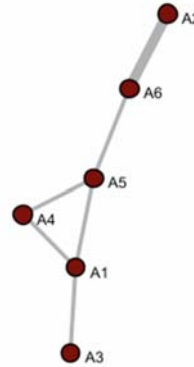
Author co-occurrence network

Extract Network from Table

Extracts a network from a delimited table

Column Name:

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

63



General Network Extraction: Unweighted, Directed Bipartite Network

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



Paper-author bipartite (2-mode) network

Extract Bipartite Network

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

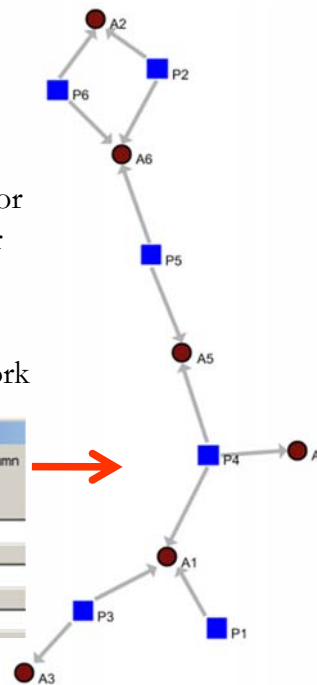
First column:

Second column:

Text Delimiter:

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

● Author
■ Paper



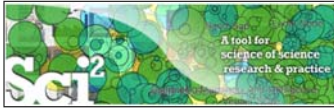
*Vertices 12

- 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

64



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

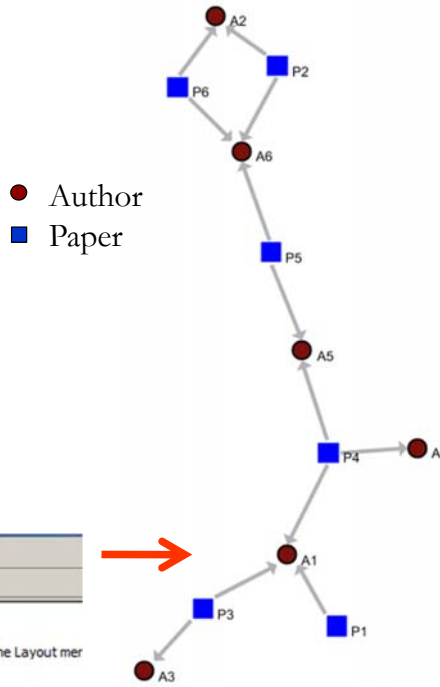
Geospatial

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

65



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

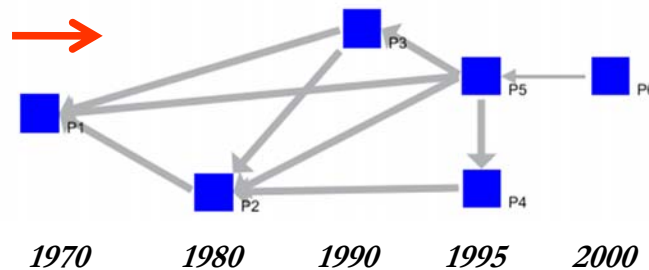
Given a table, this algorithm extracts a directed edge that starts at a column node.

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

66



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

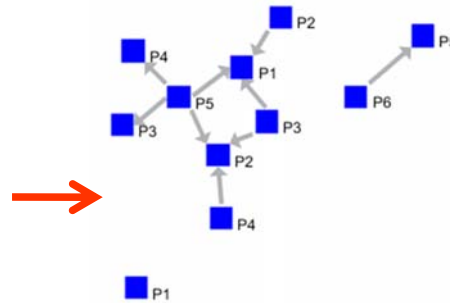
Extract Bipartite Network

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

67



ISI Paper-Citation Network Extraction

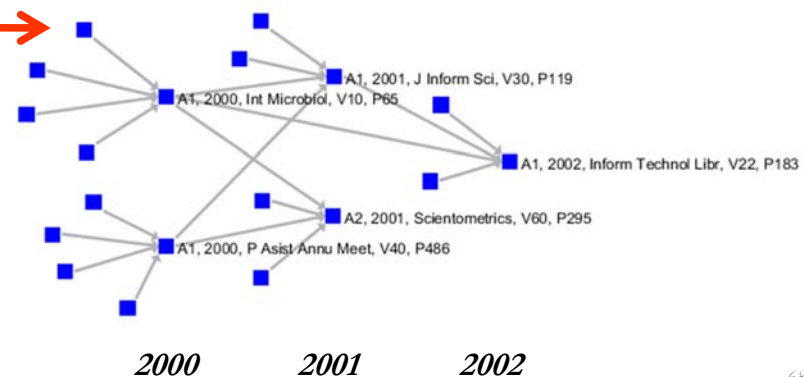
	A	B	C	D	E
1	Authors	Cited References	Publication Year	Title	Cite Me As
2	A1 A2	BENSMAN SJ, 1998, LIBR RESOUR TECH SER, V42, P147 BRO	2000	T1	A1, 2000, INT MICROBIOL, V10, P65
3	A1	BENSMAN SJ, 1999, LIBR RESOUR TECH SER, V42, P147 BRO	2000	T2	A1, 2000, P ASIST ANNU MEET, V40, P486
4	A2 A3	GARFIELD E, 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*



68



Tutorial Overview

8:30a Welcome and Overview of Tutorial and Attendees

8:45a 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

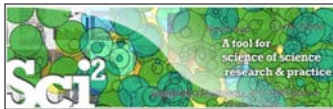
10-10:30a Networking Break

- Topical Analysis: Visualize research profiles
- Network Analysis: Co-occurrence networks and bimodal networks
- Network Analysis: Evolving collaboration networks

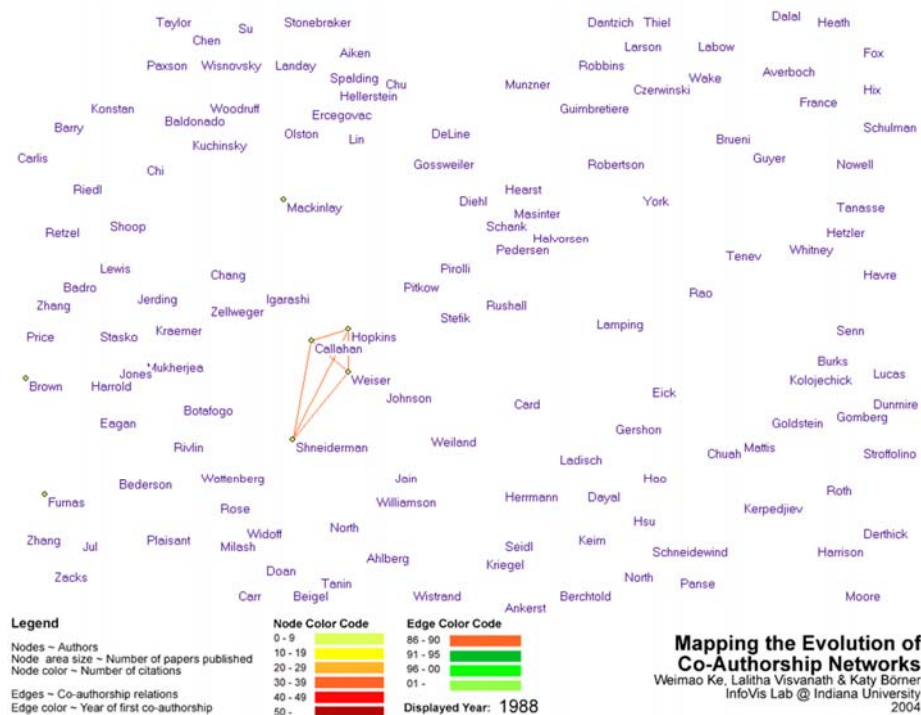
12:30p IVMOOC

12:45p Outlook and Q&A

1:00p Adjourn



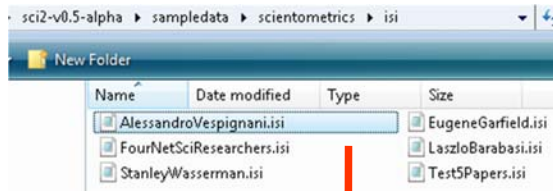
Evolving collaboration networks





Evolving Collaboration Networks

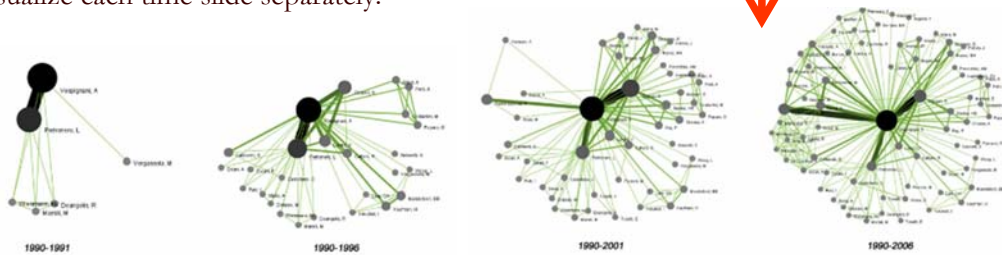
Load isi formatted file



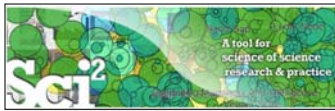
As csv, file looks like:

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

Visualize each time slide separately:



71



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

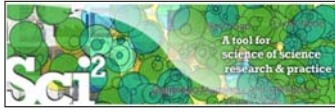
Cumulative?

Align With Calendar

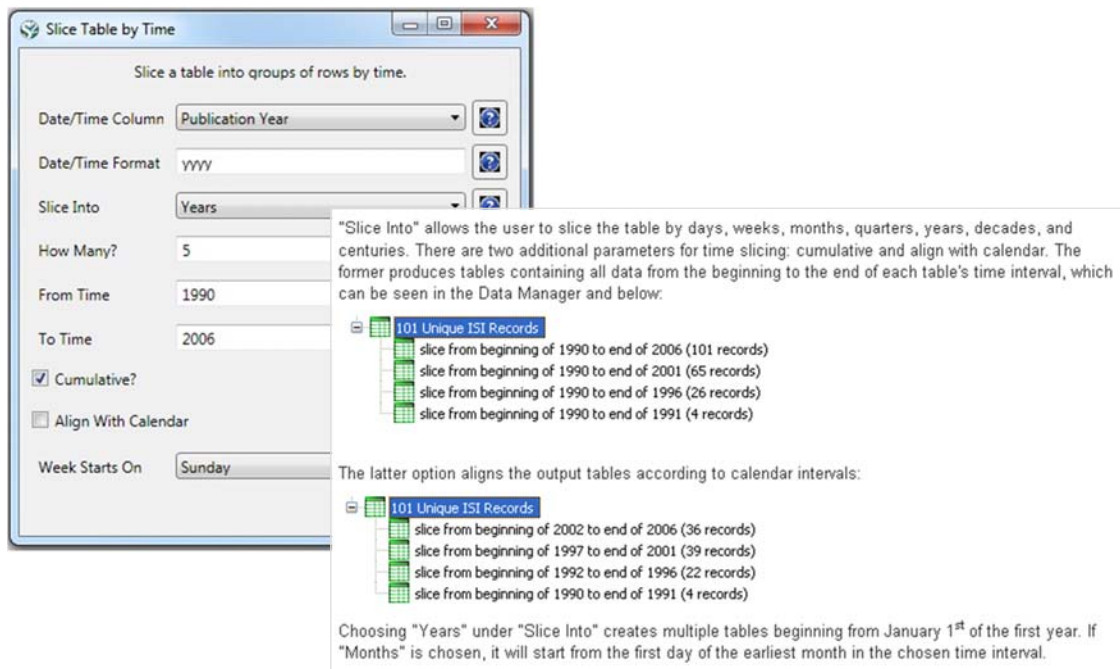
Week Starts On: Sunday

[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))

72



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))

73



Visualize Each Network, Keep Node Positions

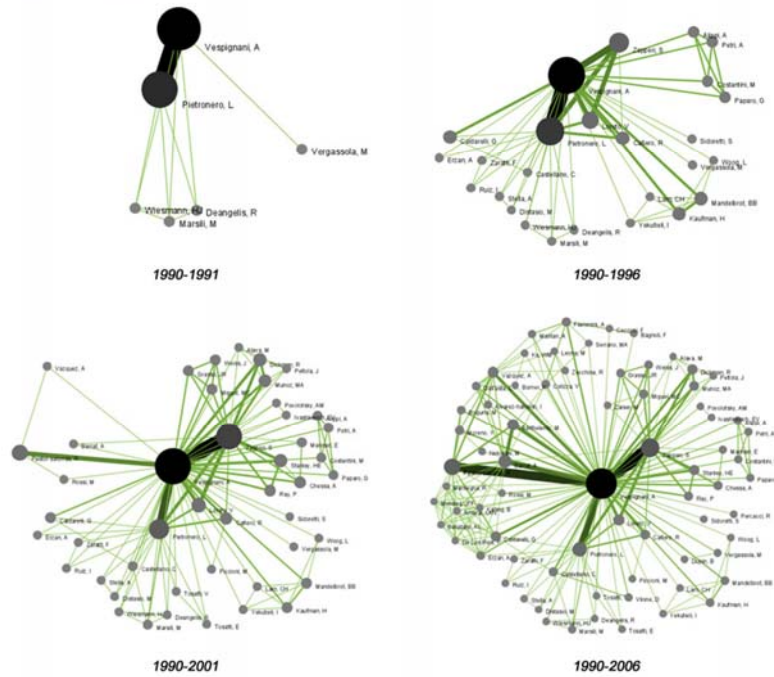
1. To see the evolution of Vespignani's co-authorship network over time, check *'cumulative'*.
2. Extract co-authorship networks one at a time for each sliced time table using *'Data Preparation > Extract Co-Author Network'*, making sure to select "ISI" from the pop-up window during the extraction.
3.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

74



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))

75



Relevant CShell plugin

CShell Slice Table by Time

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

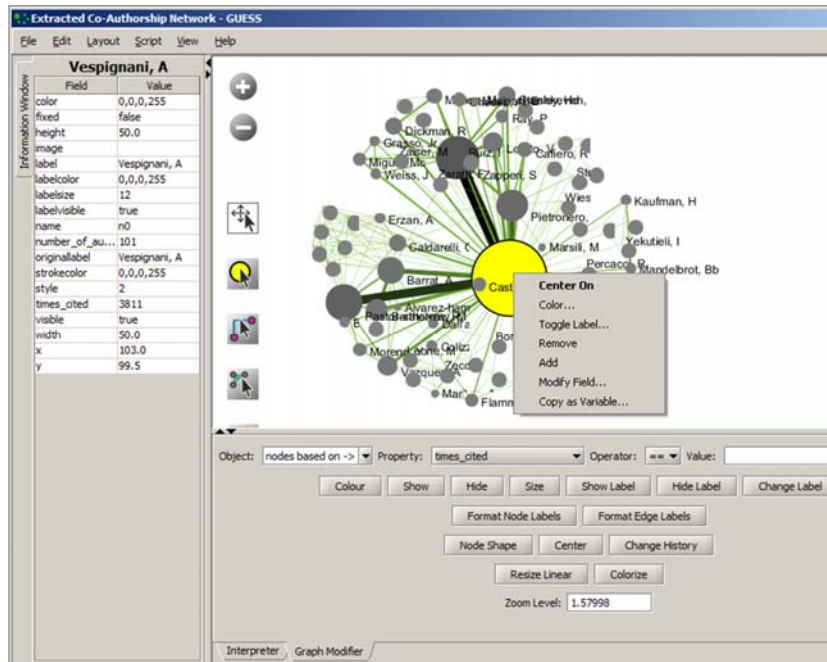
This algorithm uses the Joda Time library extensively, which provides significantly improved capabilities compared to the default Java algorithms for dates and times.

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

76



Network Visualization with GUESS




Pan:

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

Zoom:

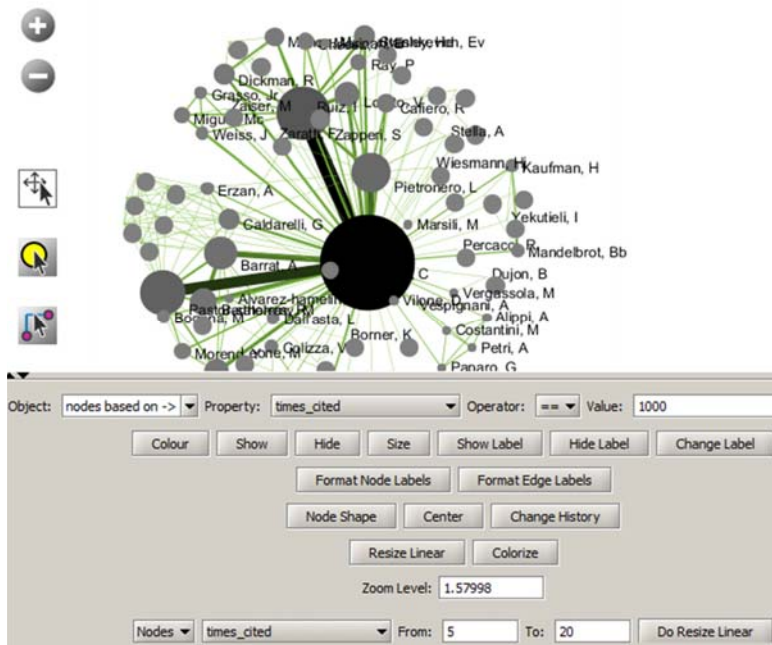
Using scroll wheel, press the “+” and “-” buttons in the upper-left hand corner, or right-click and move the mouse left or right. Center graph by selecting ‘View -> Center’.

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

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



Network Visualization with GUESS



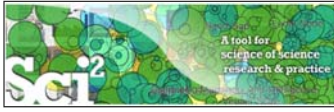
Graph Modifier:

Select “all nodes” in the Object drop-down menu and click ‘Show Label’ button.

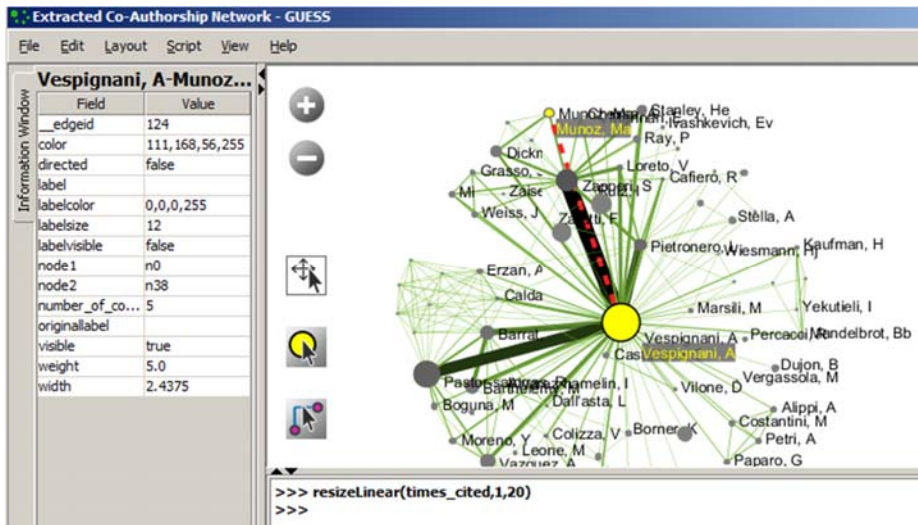
Select ‘Resize Linear > Nodes > times_cited’ drop-down menu, then type “5” and “20” into the “From” and “To” Value box separately. Then select ‘Do Resize Linear’.

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.’



Network Visualization with GUESS



Interpreter uses Jython a combination of Java and Python.

Try

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

79



Tutorial Overview

8:30a Welcome and Overview of Tutorial and Attendees

8:45a 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

10-10:30a Networking Break

- Topical Analysis: Visualize research profiles
- Network Analysis: Co-occurrence networks and bimodal networks
- Network Analysis: Evolving collaboration networks

12:30p IVMOOC

12:45p Outlook and Q&A

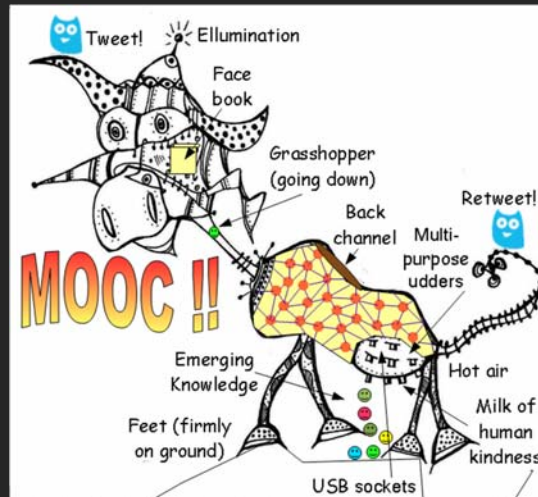
1:00p Adjourn

80

MOOCs

In 2012, Google hosted three massive open online courses (MOOCs) collectively reaching over 400,000 registrants.

By the end of 2013 more than 250 courses will be run using the Google, Coursera, Udacity, EdX, and other platforms.



81

Information Visualization MOOC

INDIANA UNIVERSITY CNS



Overview

This course provides an overview about the state of the art in information visualization. It teaches the process of producing effective visualizations that take the needs of users into account.

Among other topics, the course covers:

- Data analysis algorithms that enable extraction of patterns and trends in data
- Major temporal, geospatial, topical, and network visualization techniques
- Discussions of systems that drive research and development.

Please watch the introduction video to get better acquainted with the course.

Everybody who registers gains free access to the Scholarly Database (26 million paper, patent, and grant records) and the Sci2 Tool (100+ algorithms and tools).

Katy Börner, Ph.D.
Indiana University

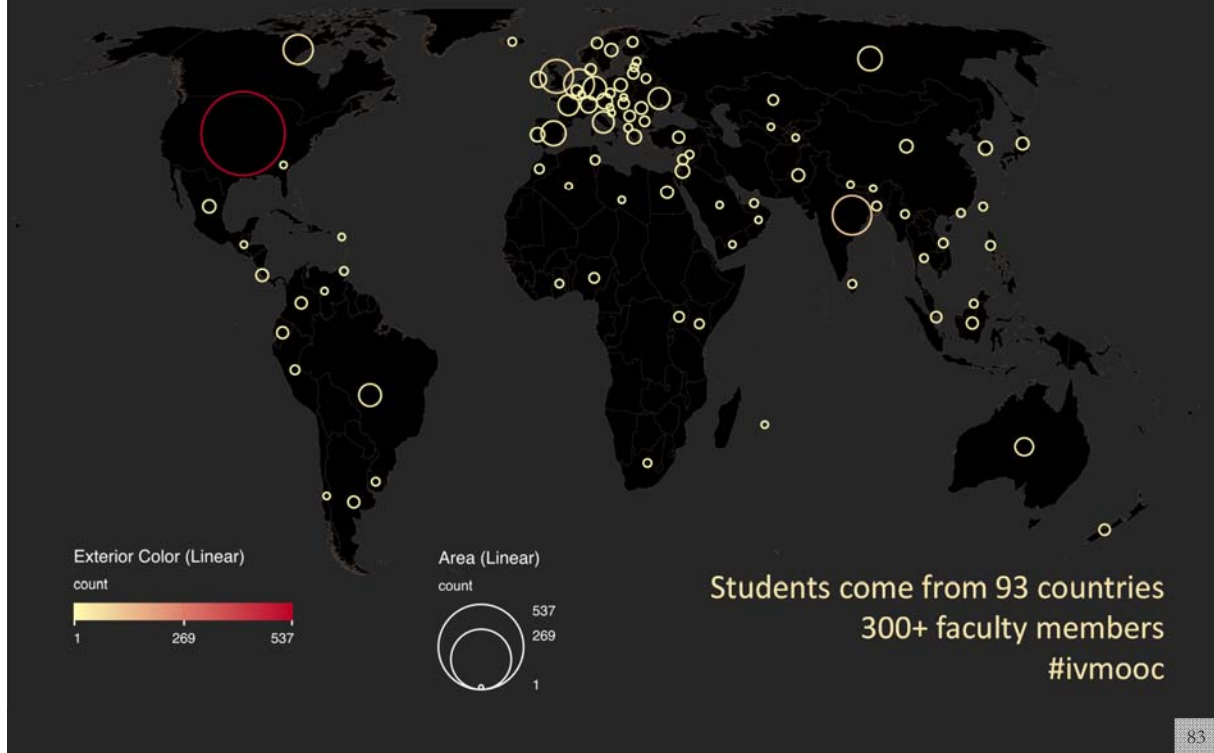


Go To The Course

ivmooc.cns.iu.edu

82

The Information Visualization MOOC
ivmooc.cns.iu.edu



Instructors

Katy Börner – Theory Parts

Instructor, Professor at SLIS



David E. Polley – Hands-on Parts

CNS Staff, Research Assistant with MIS/MLS
Teaches & Tests Sci2 Tool



Scott B. Weingart – Client Work

Assistant Instructor, SLIS PhD student



Course Schedule

Course started on January 22, 2013

- **Session 1** – Workflow design and visualization framework
- **Session 2** – “When:” Temporal Data
- **Session 3** – “Where:” Geospatial Data
- **Session 4** – “What:” Topical Data

Mid-Term

Students work in teams with clients.

- **Session 5** – “With Whom:” Trees
- **Session 6** – “With Whom:” Networks
- **Session 7** – Dynamic Visualizations and Deployment

Final Exam

85

Unit Structure

The course and each unit has three components:

Theory: Videos and Slides

Self-Assessment (not graded)

Hands-on: Videos and Slides & Wiki pages with workflows

Homework (not graded)

Client Work: Using Drupal Forum (graded)

86

Grading

All students are asked to create a personal profile to support working in teams.

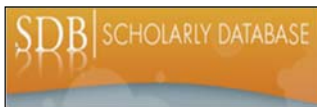


Final grade is based on Midterm (**30%**), Final (**40%**), Client Project (**30%**).

- Weekly self-assessments are not graded.
- Homework is graded automatically.
- Midterm and Final test materials from theory and hands-on sessions are graded automatically.
- Client work is peer-reviewed via online forum.

All students that receive more than **80%** of all available points get an official certificate/badge.

87



Scholarly Database at Indiana University

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

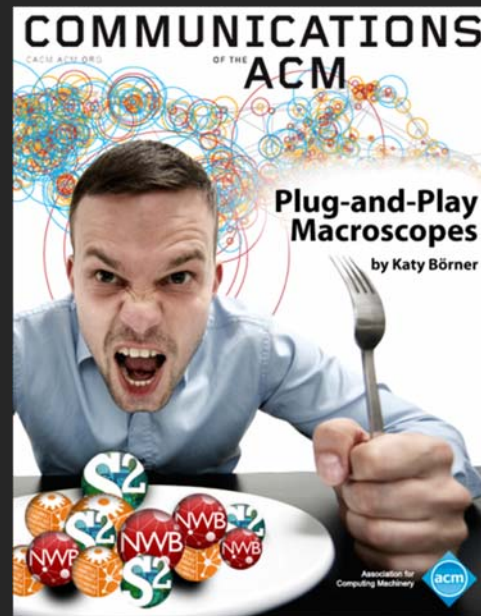
Supports federated search of 26 million publication, patent, clinical trials, and grant records. Results can be downloaded as data dump and (evolving) co-author, paper-citation networks.

The image shows two screenshots of the Scholarly Database website. The left screenshot is the login page, featuring fields for 'Email' and 'Password' for 'Non-IU User' and a 'Go to IU Login' button for 'IU User'. Below the login fields are links for 'Not Registered? Test?', 'Register as an IU User', and 'Register as a Non-IU User'. There is also a 'In the News' section with a citation for Wolfald, John (2006) and a 'Please Cite As' section with a citation for La Rosa, Sean, Ambler, Sumner, Burgess, John, Ye, Weisman and Böhm, Kath. (2007). The right screenshot shows the search interface with a search bar, filters for 'First Year' (1898) and 'Last Year' (2008), and checkboxes for 'Headline (1898 - 2008)', 'NIH (1961 - 2002)', 'NSF (1985 - 2004)', and 'USPTO (1976 - 2007)'. A 'Search' button is at the bottom.

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





88

Plug-and-Play Macroscopes
cishell.org





Börner, Katy. (March 2011). Plug-and-Play Macroscopes. *Communications of the ACM*, 54(3), 60-69. <http://www.scivee.tv/node/27704>

Clients

Information Visualization MOOC    

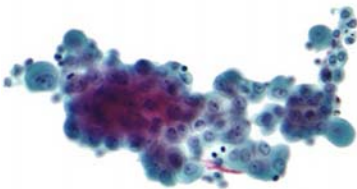
List of Clients

ISIS  **Project Title:** Isis: 100 Years
Client Name: Jay Malone
Project goal/scientific or practical value: A visual representation Isis' contributors and locales over the past 100 years. Isis is the journal of the History of Science Society. This representation will provide a dynamic picture of how scholarship in the history of science has shifted over the past century.
Information on dataset(s) to be used: Citation information, author locale, and issue number for Isis publications.
Relevant publications, websites, etc: <http://www.press.uchicago.edu/ucp/journals/journal/isis.html>
Conditions under which students can publish results and/or add project results to their resume: Client would like to approve results.

ycib  **Project Title:** e-Xploration
Client Name: Luyi
Project goal/scientific or practical value: e-Xploration is an agent-based model for the ethnographic observation and the registry, analysis, and interpretation of social practices in virtual communities for intervention in the development of collaboration and cooperation. This project will analyze the interactions between subjects and objects in a platform collaborative community called OYCIB, a project based on e-Xploration (e-crick.net).
Information on dataset(s) to be used: I can provide a data base in .graphml format for the students. The file .graphml contains the interactions between subjects and objects in a platform collaborative community called OYCIB. In the level of practice, it is not necessary that students know agent-based models for using the database. But, in another level, for example: the collaborate level for the OYCIB development, it is necessary to have basic knowledge in AMS or MAS and another competences like PHP and MySQL.
Relevant publications, websites, etc: <http://www.e-crick.net/logs>
Conditions under which students can publish results and/or add project results to their resume: If any person or institution use my dataset or another info about eXploration (e-crick.net, oycib.net), I need to approve the results and appear as co-author.

http://ivmooc.cns.iu.edu/ivmooc_clientprojects.html

Diogo Carmo



Mesothelioma

Main title topics in Medline papers

Mesothelioma is a rare form of cancer that develops from transformed cells originating in the mesothelium, the protective lining that covers many of the internal organs of the body. It is usually caused by exposure to asbestos.

The most common anatomical site for the development of mesothelioma is the pleura (the outer lining of the lungs and internal chest wall), but it can also arise in the peritoneum (the lining of the abdominal cavity), and the pericardium (the sac that surrounds the heart) or the tunica vaginalis (a sac that surrounds the testis).

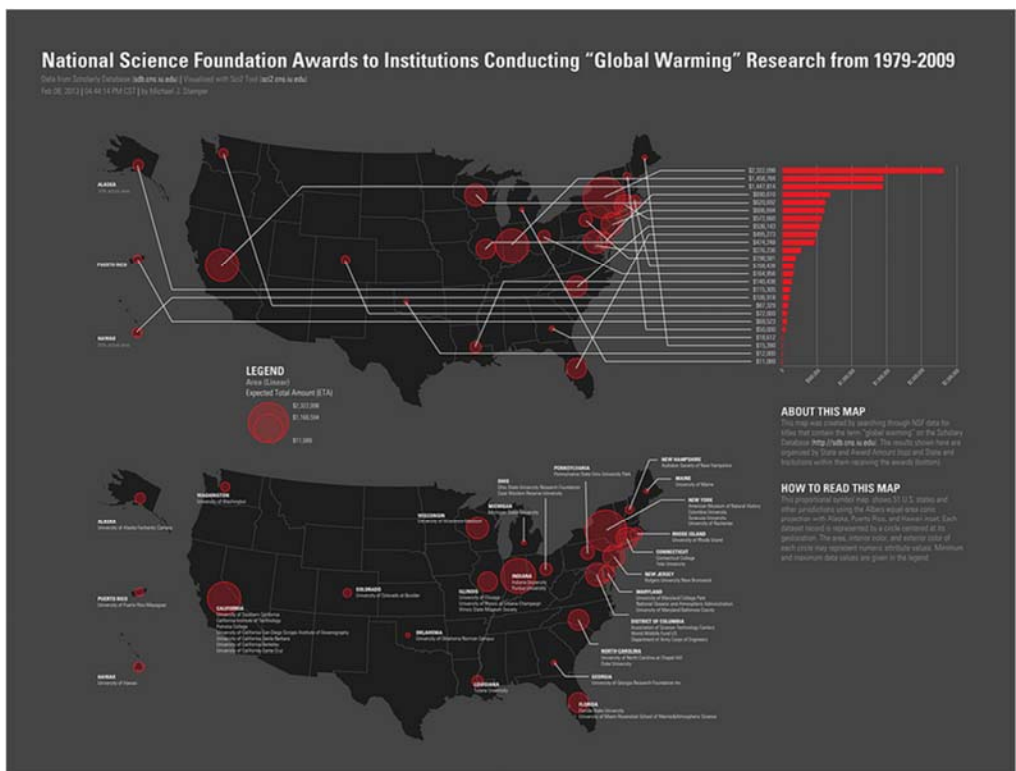
Most people who develop mesothelioma have worked in jobs where they inhaled asbestos, or were exposed to asbestos dust and fibers in other ways. It has also been suggested that smoking cigarettes if a family member who worked with asbestos increases their risk for developing mesothelioma. Unlike lung cancer, there seems to be no association between mesothelioma and tobacco smoking, but smoking greatly increases the risk of other asbestos-induced cancers. Some people who were exposed to asbestos have collected damages for asbestos-related disease, including mesothelioma. Compensation via asbestos funds or class action lawsuits is an important issue in law practices regarding mesothelioma.

MALIGNANT PLEURAL CYSTIC BENIGN DIAGNOSIS



How To Read This Map
This map shows the geographic distribution of mesothelioma cases each year as a horizontal bar with a specific start and end date and a label on its left side. The area of each bar encodes the number of cases and target magnitude - in the space identified in the labels (target words were summed).

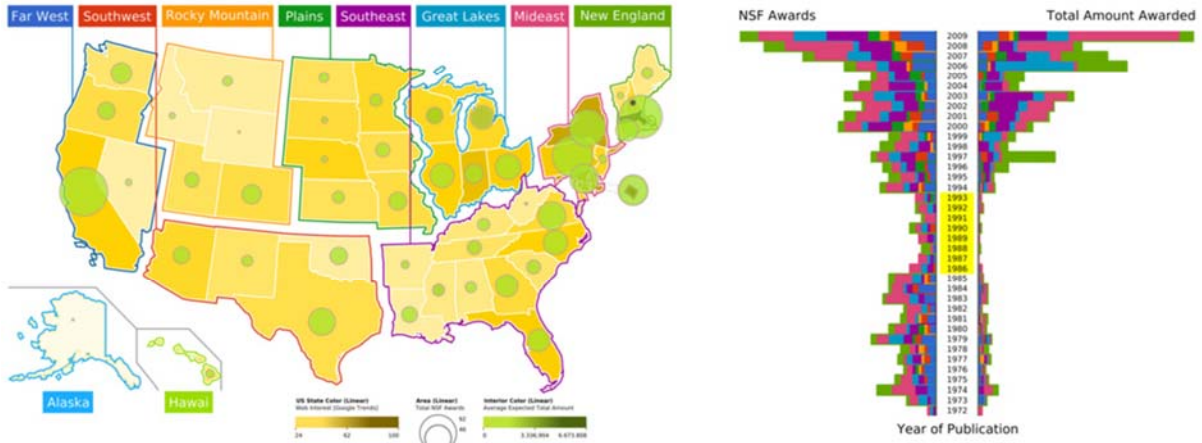
Author: Diogo Carmo <http://dx.doi.org/10.1371/journal.pone.0020001>; Visualization software: Sci2 Tools, (2009) Science of Science (Sci2) Tool, Indiana University and SciTech Strategies, <http://sci2.ucis.edu/>; (BioSart) Medline Papers, as available in SciSearch Database, <http://pubs.nlm.nih.gov/>; Text and Images: Wikipedia/Mesothelioma article, available at <http://en.wikipedia.org/wiki/Mesothelioma>; From: Tomarney, by Bernd Mohr © 2001 - All Rights Reserved. This font family is licensed under available at <http://pubs.nlm.nih.gov/>



mjstamper ivmooc

Innovation & Entrepreneurship

NSF Funding Across the US, from 1972 to 2009, and Current Web Interest

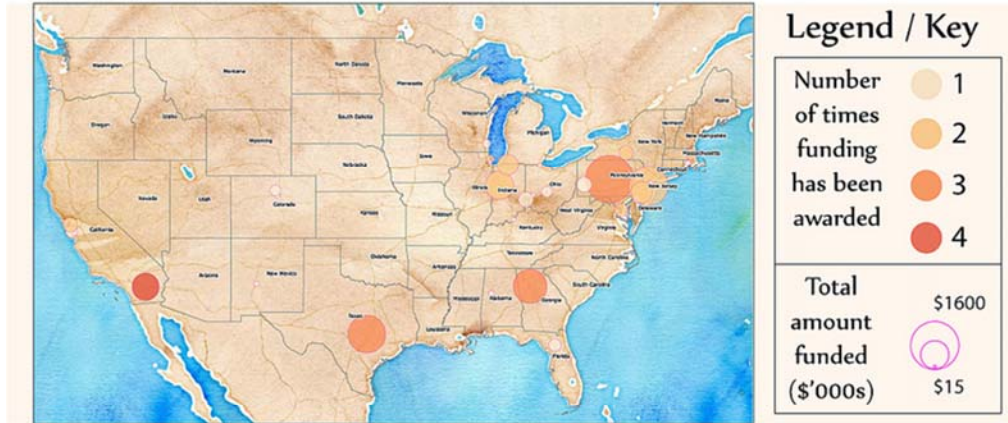


Author: Diogo Carmo <http://diogo.carmo@unh.edu> | Visualization software: Gird Team (2009), Science of Science (2012) Tool: Indiana University and TuTech Strategies, <http://tutec.com/usa/>; Google Sheets was used to produce the bar graphs | Dataset: National Science Foundation (NSF) Awards, as available in Scholarly Database <http://dx.doi.org/10.1016/j.enecon.2010.08.001>; Time + innovation (2010) entrepreneurship; Google Trends, for innovation (2010) entrepreneurship | Photo: Tomaric, by Bernd Montag (2010) | All Rights Reserved. This font family is licensed and is available at <http://fontbundles.net/>; and, Digitally by Bitstream, Inc. © 2010. All Rights Reserved. This font family is available at <http://fontbundles.net/>

[Diogo Carmo](#)

NSF Funding - Graphene Projects 2004-2010

#ivmooc Week 3 homework @jonopatterson

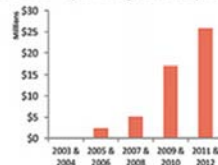


What is Graphene?

Graphene is a two dimensional material consisting of a single layer of carbon atoms arranged in a honeycomb or chicken wire structure. It is the thinnest material known and yet is also one of the strongest. It conducts electricity as efficiently as copper and outperforms all other materials as a conductor of heat. Graphene is almost completely transparent, yet so dense that even the smallest atom helium cannot pass through it.

Originally thought to be unstable in its free state it proved to be quite the opposite when isolated by Andre Geim and Konstantin Novoselov at the University of Manchester in 2003. The results of this work, which were published in 2004, heralded a new dawn in the study of two dimensional materials and of graphene in particular.

Total NSF Spending on Graphene



How to read this map

The map shows NSF funding awards to US institutions. Each circle corresponds to an institution. The depth of colour represents the number of times funding was awarded to the same institution for different projects. The circle size indicates the amount awarded in \$'000s.

Sources: NSF; Manchester University
Open Street Mapping by CC. Styled by Stamen

[JonoPatterson](#)

Visualizing IVMOOC Data

Empowering Teachers: How to make sense of the activities of thousands of students? How to guide them?

Empowering Students: How to navigate learning materials and develop successful learning collaborations across disciplines and time zones?

Empowering MOOC Platform Designers: What technology helps and what hurts?

Research: What teaching and learning works in a MOOC?

95



Tutorial Overview

8:30a Welcome and Overview of Tutorial and Attendees

8:45a 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

10-10:30a Networking Break

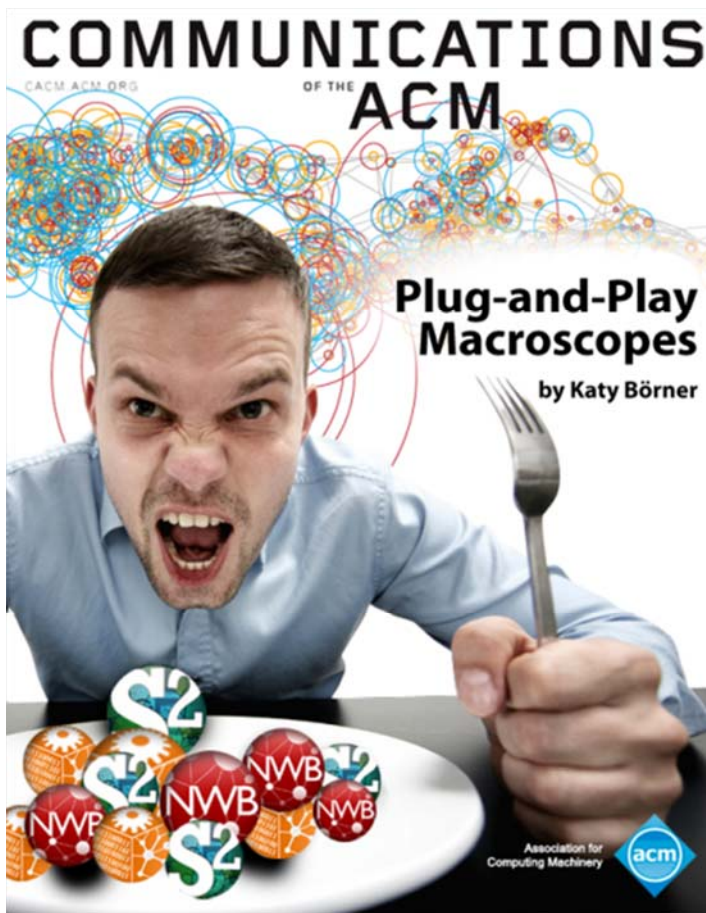
- Topical Analysis: Visualize research profiles
- Network Analysis: Co-occurrence networks and bimodal networks
- Network Analysis: Evolving collaboration networks

12:30p IVMOOC

12:45p Outlook and Q&A

1:00p Adjourn

96



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>

97



OSGi/CIShell Adoption

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

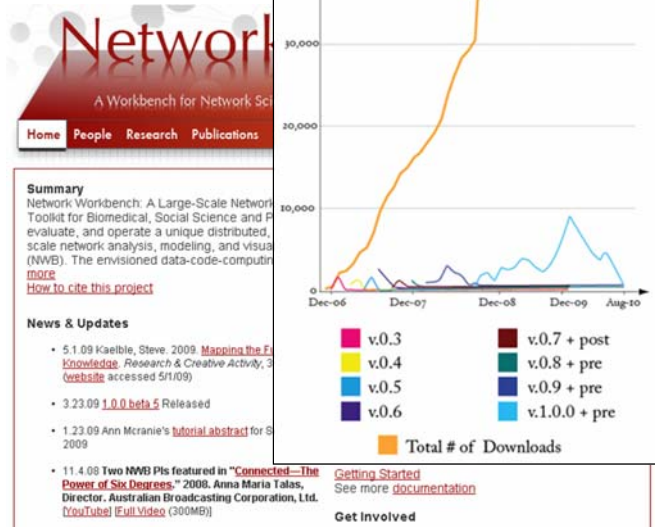
- USA**
- *Cytoscape* (<http://cytoscape.org>) Led by Trey Ideker at the University of California, San Diego is an open source bioinformatics software platform for visualizing molecular interaction networks and integrating these interactions with gene expression profiles and other state data (Shannon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

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.

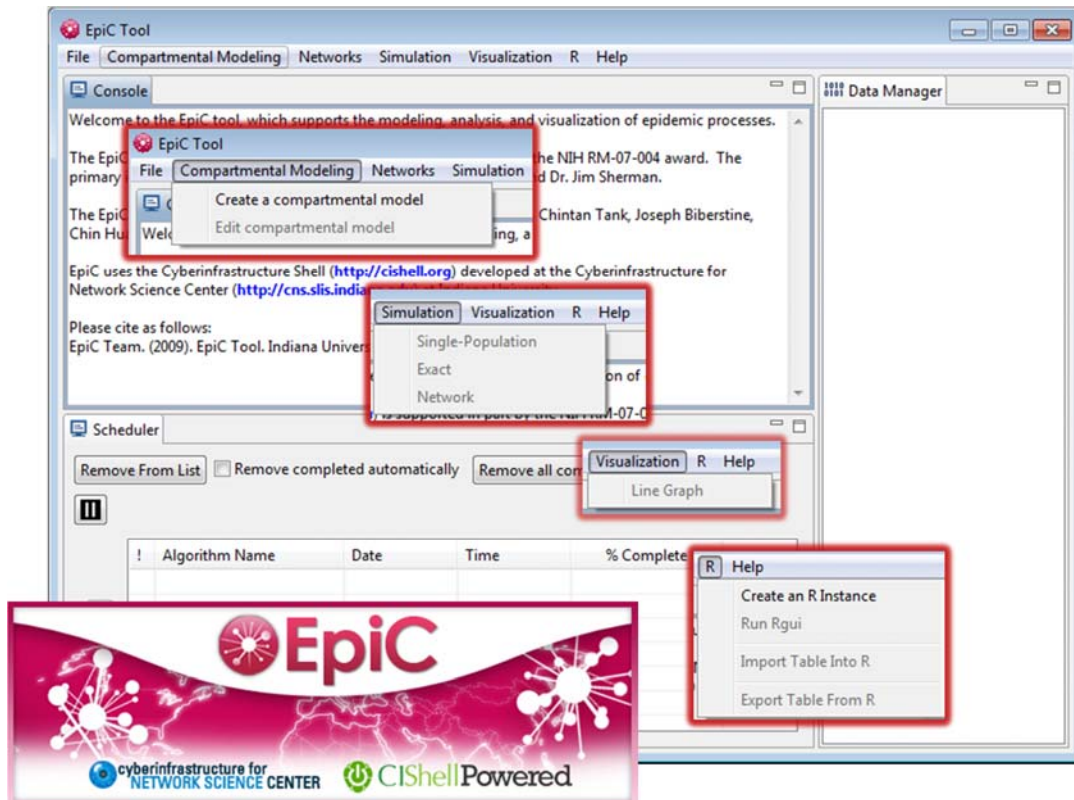
More than 50 of these plugins can be applied or were specifically designed for S&T studies.

It has been downloaded more than 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.

99



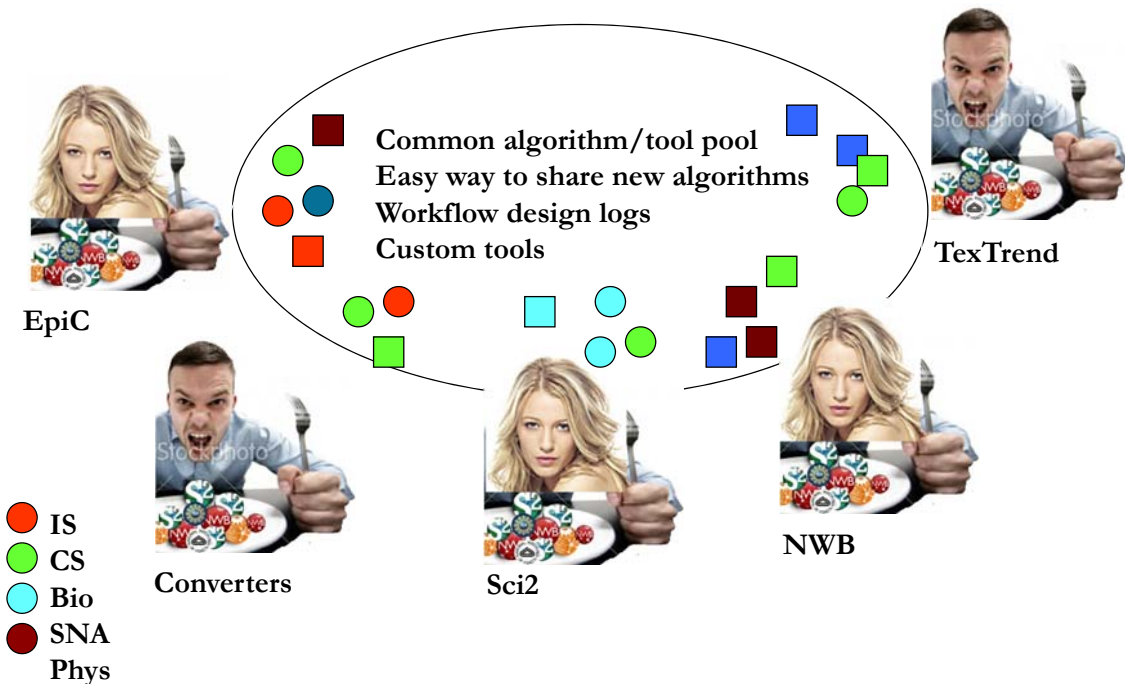
TEXTrend adds R bridge, WEKA, Wordij, CFinder, and more.

See the latest versions of TEXTrend Toolkit modules at

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



OSGi/CIShell-Powered Tools Support Algorithm Sharing





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>

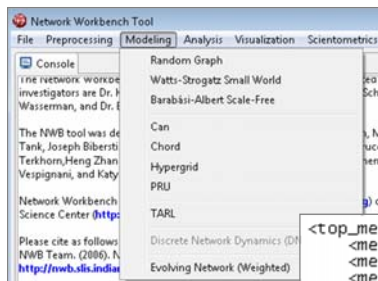
Additional Sci2 Plugins are at <http://sci2.wiki.cns.iu.edu/3.2+Additional+Plugins>

103



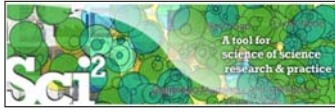
CIShell – Customize Menu

- 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>
```

104



Need Help? Ask an Expert!

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

105

If you enjoyed the tutorial, please thank Noshir Contractor,
Brian Uzzi, and Latonia Trimuel

Q & A

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

Bug reports and all comments are welcome.

CNS Cyberinfrastructure for Network Science Center

search.cns.iu.edu Search

About Us Research Development Teaching Outreach Videos News & Events Connect With Us

We work closely with clients to provide custom-made data, visualization, and software solutions

Research
 Publication: Global Multi-Level Analysis of the 'Scientific Food Web'

Latest News
 Katy Börner attends DASER at the National Academies in Washington, D.C.

Development
 Behind the scenes of the design and development of AcademyScope

Outreach
 See some of the most fascinating data visualizations in the world.

Upcoming Events

JUL 24	Robert Light attends Joint Conference on Digital Libraries 2013
07.24	Adam Simpson & Chin Hua Kong attend DevCon 5
08.11	Places & Spaces poster exhibit goes on display in Chile
08.13	Robert Light & Chin Hua Kong attend 2013 VIVO Conference

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>