

Information Visualization Tools

Dr. Katy Börner

Cyberinfrastructure for Network Science Center

Information Visualization Laboratory

School of Library and Information Science

Indiana University, Bloomington, IN

<http://cns.iu.edu>

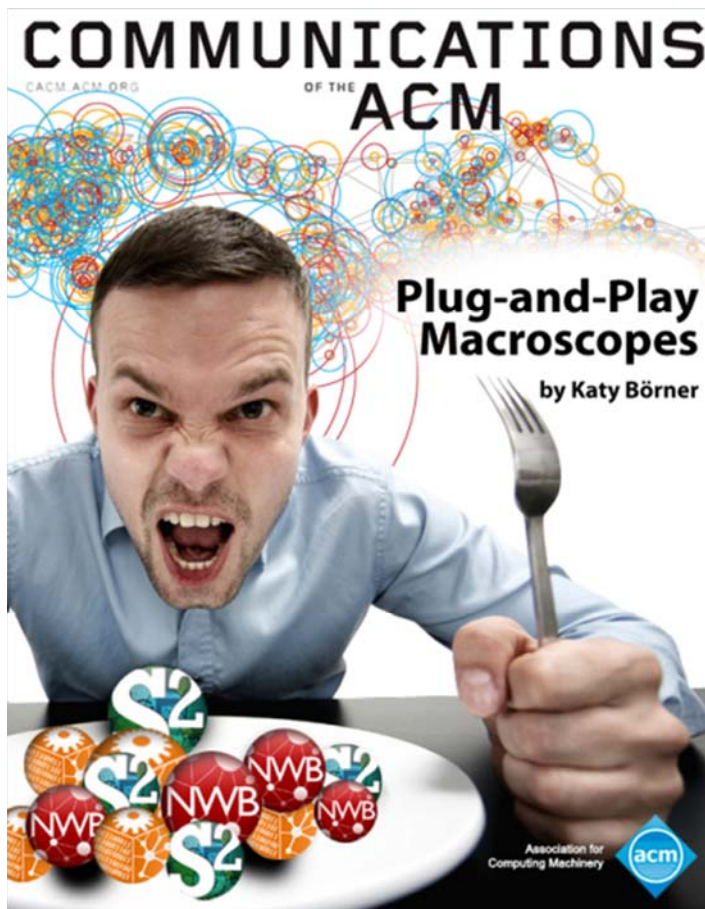


With special thanks to Kevin W. Boyack, Chin Hua Kong, Micah Linnemeier, Russell J. Duhon, Patrick Phillips, Joseph Biberstine, Chintan Tank Nianli Ma, Scott Weingart, Hanning Guo, Mark A. Price, Angela M. Zoss, Ted Polley, and Sean Lind



*Panel Discussion, All School Day
University of North Texas, Denton, TX*

October 1, 2011



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>



Sci² Tool: Download, Install, and Run

Sci² Tool v0.5.1 Alpha (May 4th, 2011)

Can be freely downloaded for all major operating systems from

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

Select your operating system from the pull down menu and download.

Unpack into a /sci2 directory.

Run /sci2/sci2.exe

Sci² Manual is at

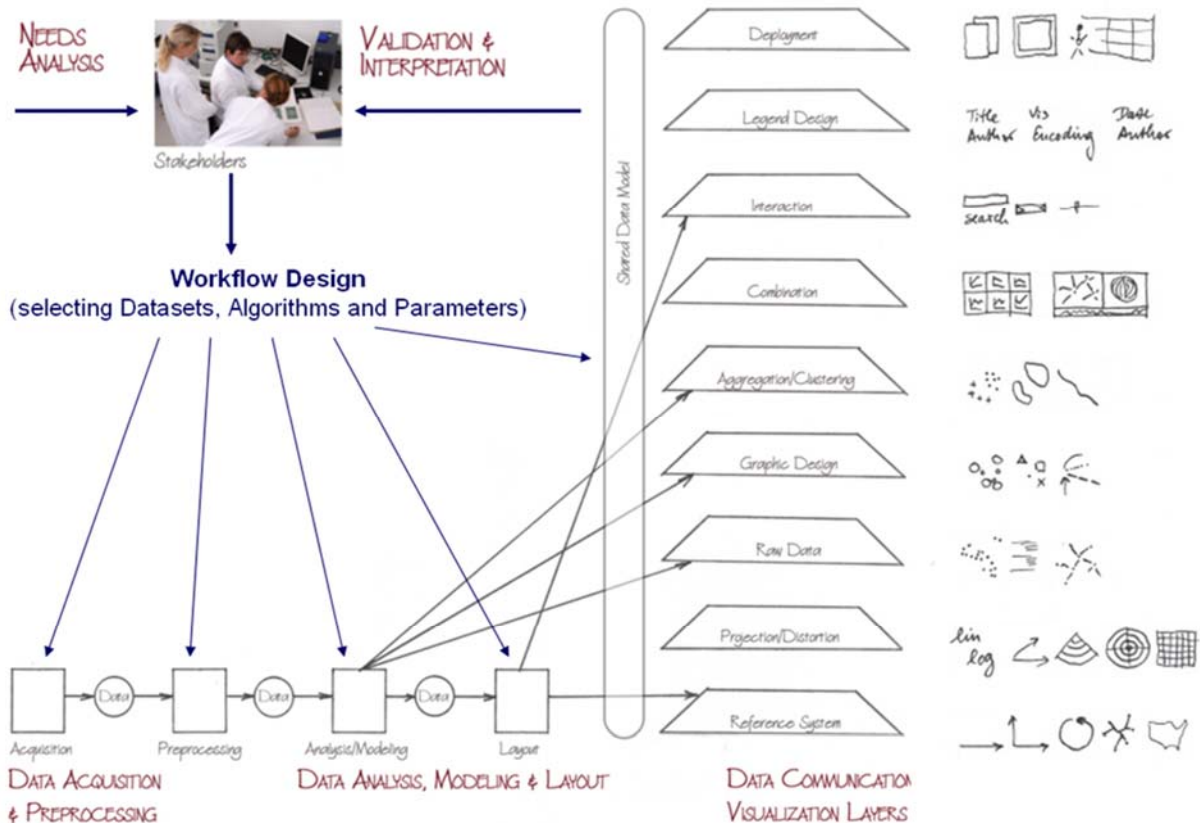
<http://sci2.wiki.cns.iu.edu>



Cite as

Sci² Team. (2009). Science of Science (Sci²) Tool. Indiana University and SciTech Strategies, <http://sci2.cns.iu.edu>

3



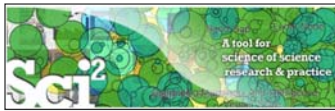
4



Type of Analysis vs. Level of Analysis

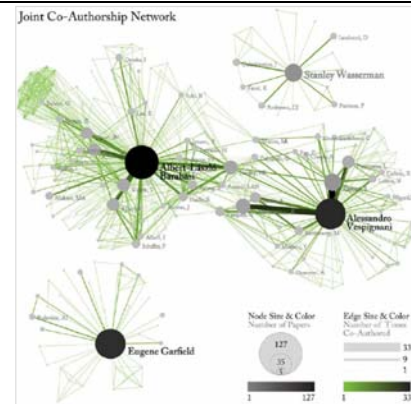
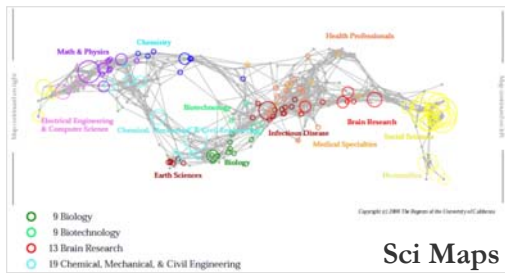
| | <i>Micro/Individual</i> (1-100 records) | <i>Meso/Local</i> (101-10,000 records) | <i>Macro/Global</i> (10,000 < records) |
|---------------------------------------|--|---|---|
| Statistical Analysis/Profiling | Individual person and their expertise profiles | Larger labs, centers, universities, research domains, or states | All of NSF, SA, all of sci |
| Temporal Analysis (When) | Funding portfolio of one individual | Mapping a s... of PNAS | 113 Years of P... Research |
| Geospatial Analysis (Where) | Career trajectory of one individual | Mapping a s... intellectual l... | PNAS |
| Topical Analysis (What) | | | VxOrd/Topic r... NIH funding |
| Network Analysis (With Whom?) | NSI... network of one | | NIH's... cy |

5

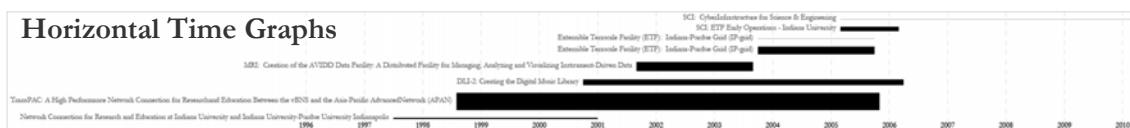


Open Code for Replicable S&T Assessment

OSGi/CIShell powered tool, see <http://cishell.org>
<http://sci2.cns.iu.edu> | <http://sci2.wiki.cns.iu.edu>

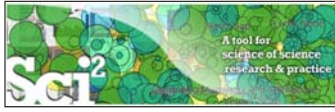


Horizontal Time Graphs

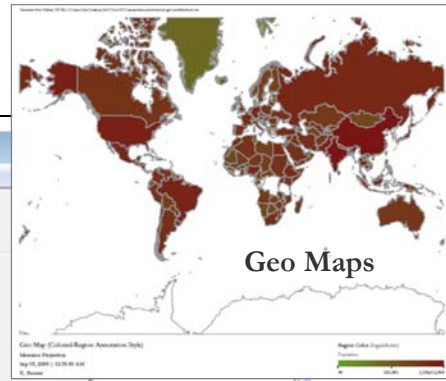


Börner, Katy, Huang, Weixia (Bonnie), Linnemeier, Micab, Dubon, Russell Jackson, Phillips, Patrick, Ma, Nianli, Zoss, Angela, Guo, Hanning & Price, Mark. (2009). *Reti-Netzwerk-Red: Analyzing and Visualizing Scholarly Networks Using the Scholarly Database and the Network Workbench Tool. Proceedings of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

6



Sci² Tool



Sci² Tool

File Preprocessing Modeling Analysis Visualization Scientometrics Help

Console

Welcome to the Science of Science Tool (Sci²). The development of this tool is supported in Network Science center and the School of Li at Indiana University, the National Science Foundation and IIS-0715303, and the James S. McDonnell Cyberinfrastructure portal (<http://sci.slis.indiana.edu>).

The primary investigators are Katy Börner, In SciTech Strategies Inc. The Sci² tool was developed by J. Duhon, Patrick A. Phillips, Chintan Tank, a Cyberinfrastructure Shell (<http://cishell.org>) for Network Science Center (<http://cns.slis.indiana.edu>). Many algorithm plugins were derived from the Network Science Center (<http://nwb.slis.indiana.edu>).

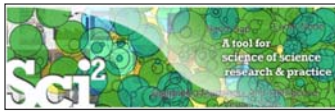
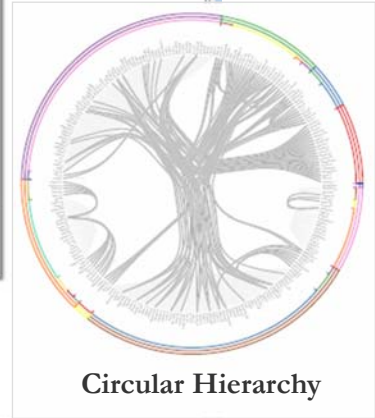
Please cite as follows:
Sci² Team. (2009). Science of Science Tool. In SciTech Strategies Inc., <http://sci.slis.indiana.edu>.

Scheduler

Remove From List Remove completed

| Algorithm Name | Date | Time | % Complete |
|---------------------------|------------|-------------|------------|
| Extract Co-Author Network | 09/03/2009 | 00:15:20 AM | 100% |
| Load and Clean ISI File | 09/03/2009 | 00:15:05 AM | 100% |

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



Network Extraction: Examples

Author co-occurrence network

| | A | B |
|---|-------------|----------|
| 1 | Publication | Authors |
| 2 | Paper1 | A1 |
| 3 | Paper2 | A1;A2;A6 |
| 4 | Paper3 | A1;A3 |
| 5 | Paper4 | A1;A4;A5 |
| 6 | Paper5 | A5;A6 |
| 7 | Paper6 | A1;A2 |

Extract Network from Table

Extracts a network from a delimited table

Column Name: Authors

Text Delimiter: ;

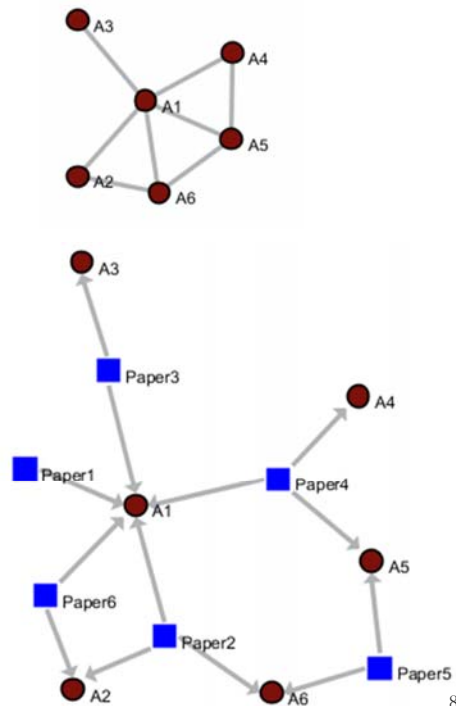
Extract Bipartite Network

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: Publication

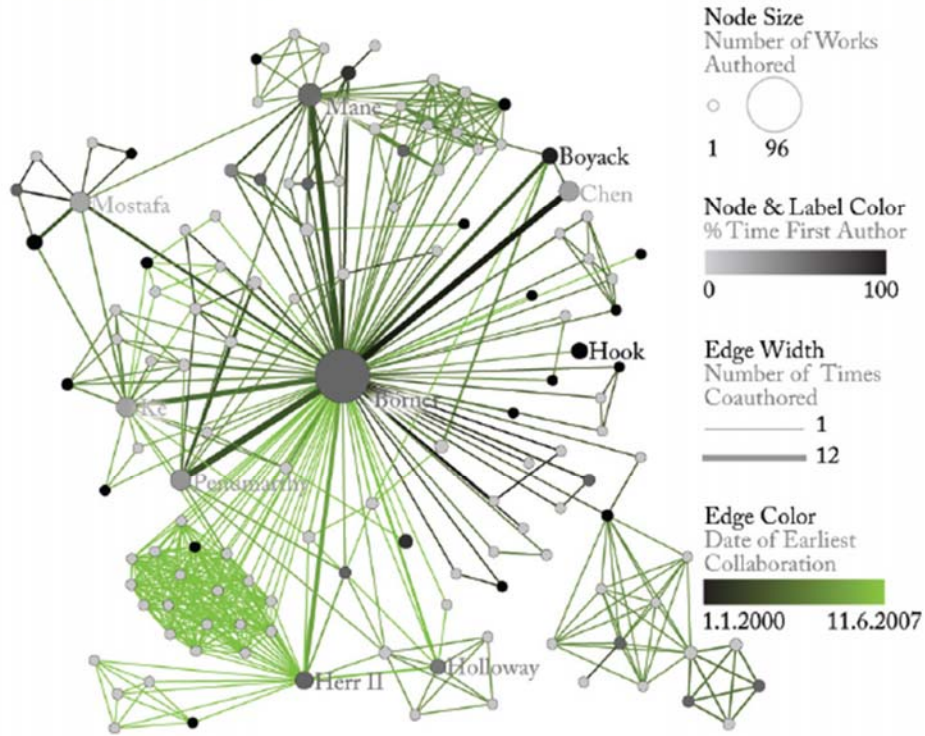
Second column: Authors

Text Delimiter: ;



Paper-author 2-mode network

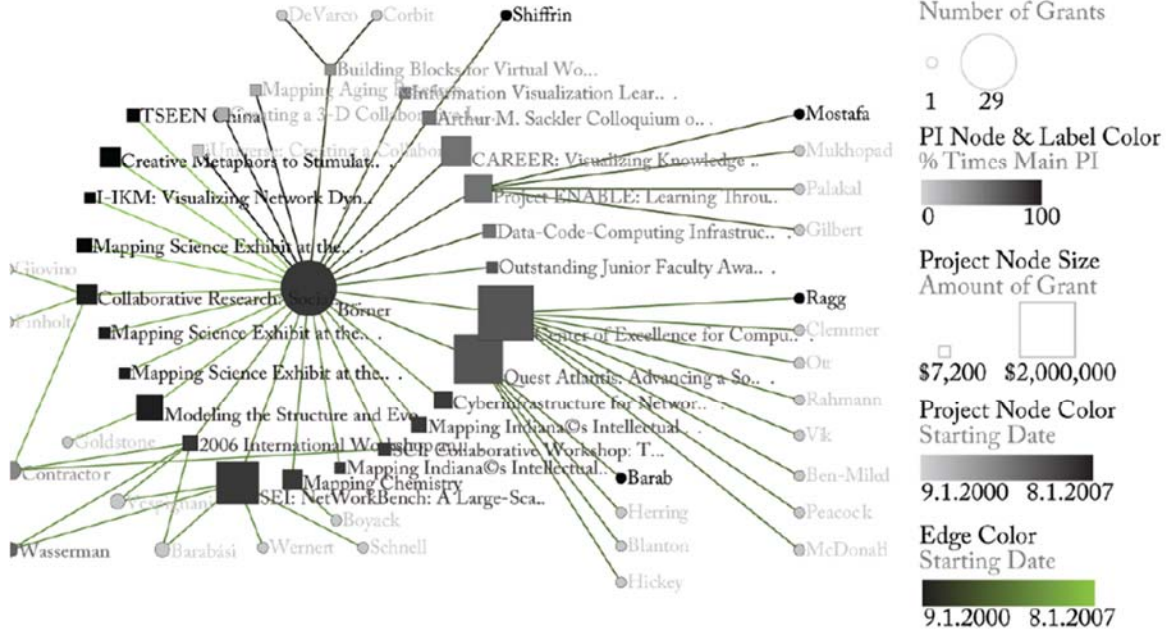
Coauthor Network



Project Timeline



Investigator-Project Network



11

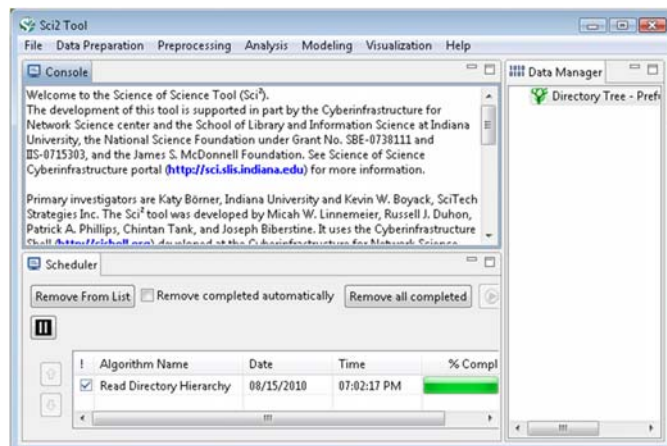


Sci2 Tool Interface Components

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.



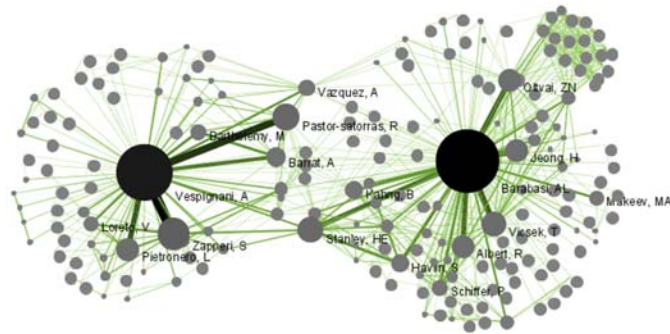
All workflows are recorded into a log file (see /sci2/logs/...), and soon can be re-run for easy replication. If errors occur, they are saved in a error log to ease bug reporting.

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

12



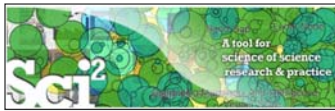
Network Visualization: Giant Component



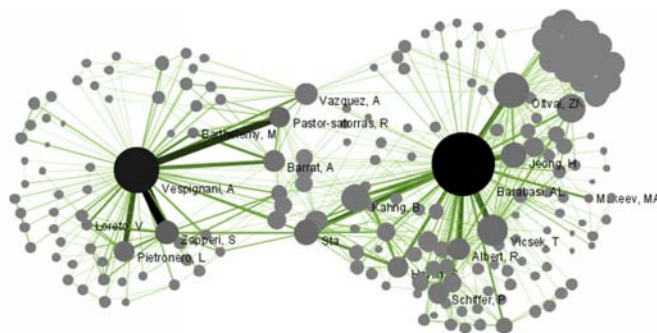
.....
Weak Component Clustering was selected.
Implementer(s): Russell Duhon
Integrator(s): Russell Duhon

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

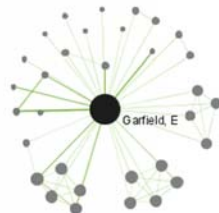
13



Network Visualization: Color/Size Coding by Degree



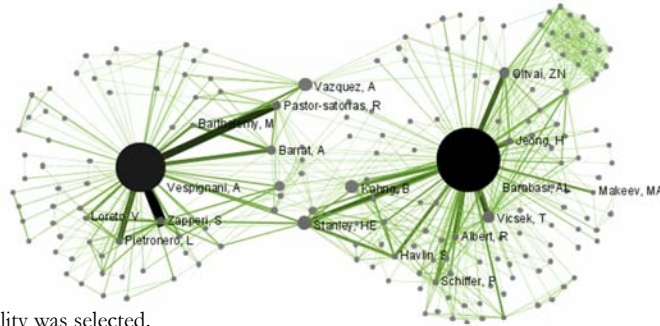
.....
Node Degree was selected.
Documentation:
<https://nwb.slis.indiana.edu/community/?n=AnalyzeData.No deDegree>
.....



14

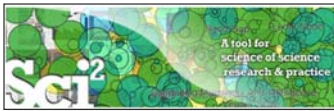
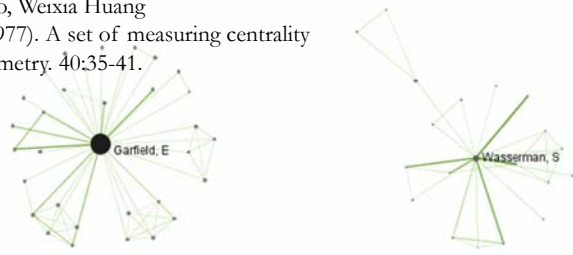


Network Visualization: Color/Size Coding by Betweenness Centrality

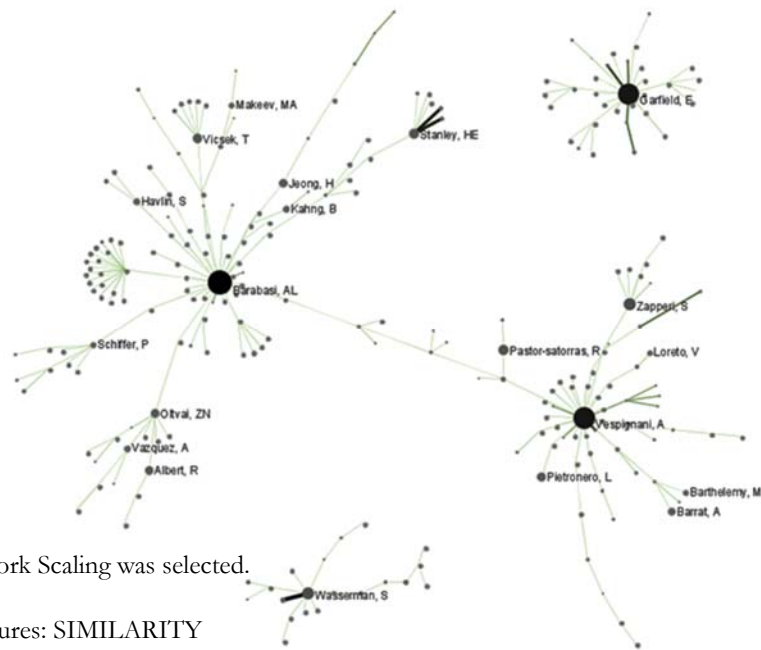


.....
 Node Betweenness Centrality was selected.
 Author(s): L. C. Freeman
 Implementer(s): Santo Fortunato
 Integrator(s): Santo Fortunato, Weixia Huang
 Reference: Freeman, L. C. (1977). A set of measuring centrality based on betweenness. Sociometry. 40:35-41.

Input Parameters:
 Number of bins: 10
 Number of bins: 10



Network Visualization: Reduced Network After Pathfinder Network Scaling

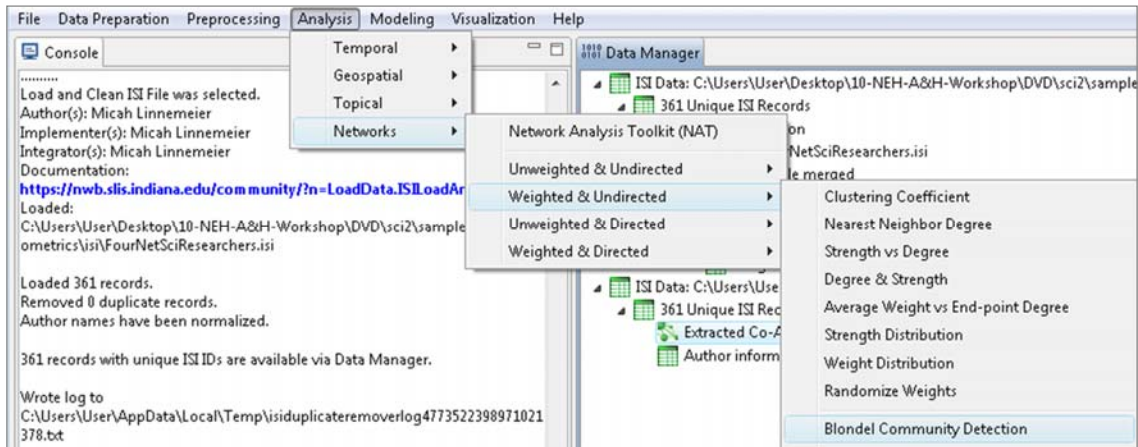


.....
 MST-Pathfinder Network Scaling was selected.
 Input Parameters:
 Weight Attribute measures: SIMILARITY
 Edge Weight Attribute: weight

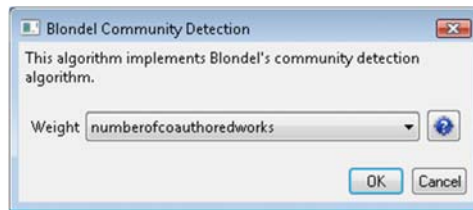


Network Visualization: Circular Hierarchy Visualization

Select Co-Author Network and run Blondel Community detection:



With parameter values



17



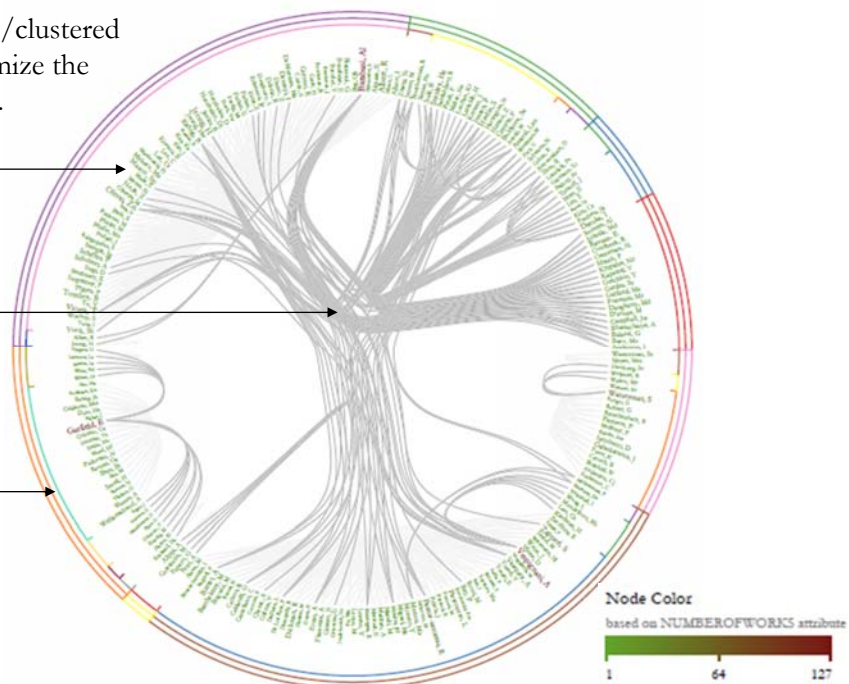
Network Visualization: Circular Hierarchy Visualization

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

Node labels, e.g., author names.

Network structure using edge bundling.

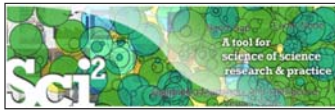
Color coded cluster hierarchy according to Blondel community detection algorithm.



Note:

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

18

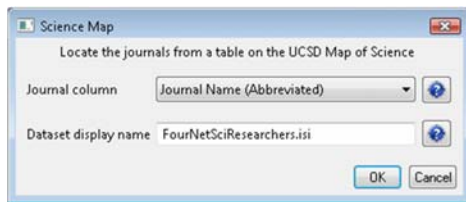
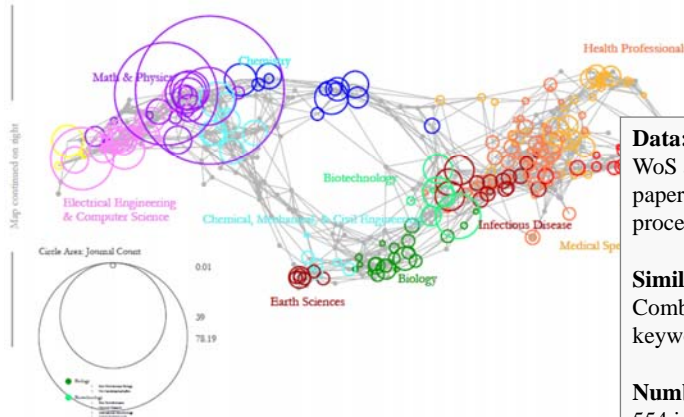


Topic Mapping: UCSD Science Map

Science Map via Journals for FourNetSciResearchers.isi

314 journal references matched out of 361 found.

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



Data:

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

Similarity Metric:

Combination of bibliographic coupling and keyword vectors

Number of Disciplines:

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



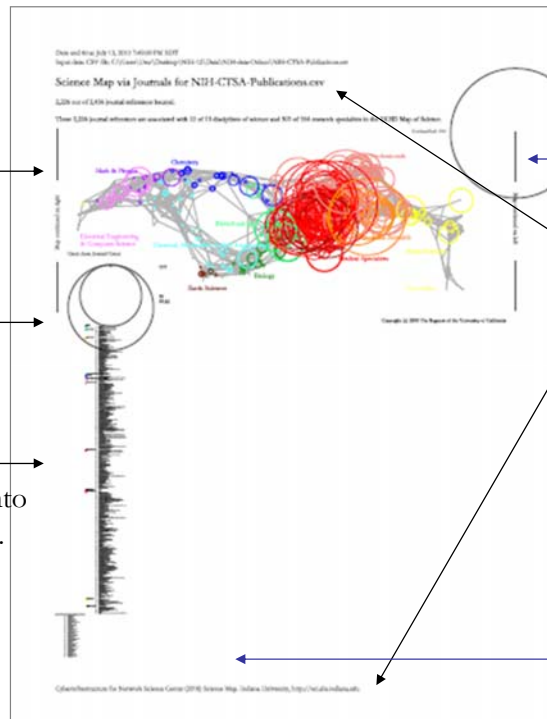
19

How to Read the UCSD Map

UCSD Science Map with data overlay.

Map legend of circle area size coding

Listing of all data records organized into UCSD science areas.



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

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

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

20



Geospatial maps with congressional districts

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



Identify Congressional District, Latitude, Longitude

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



Aggregate/Count identical Congressional Districts

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



21

How to Read the Geo Map

U.S. Map with data overlay.

Listing of map type, author, and parameters used.



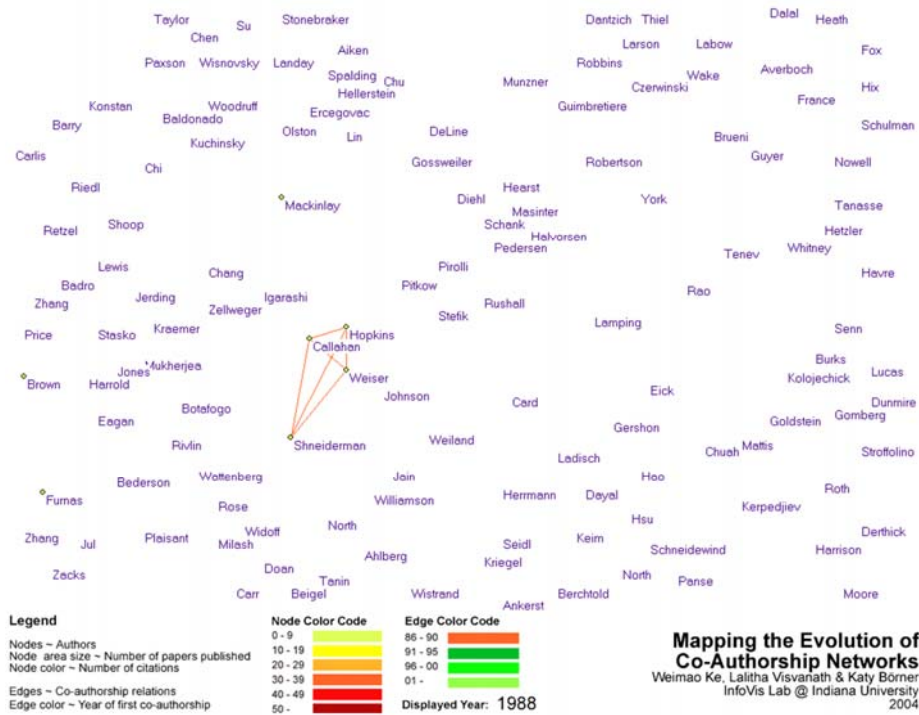
Header and footer with information when this map was created, by whom and using what data set.

Map legend with color coding.

22



Evolving collaboration networks

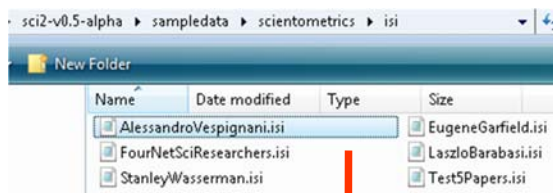


23



Evolving collaboration networks

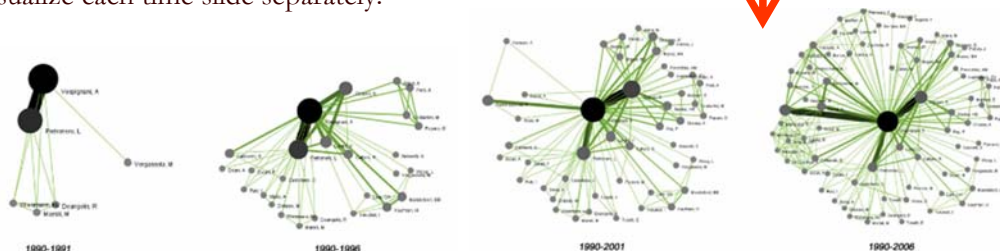
Load isi formatted file



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:



24

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

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

The screenshot shows the Scholarly Database website. On the left is the login page with fields for 'IU User' and 'Non-IU User'. On the right is the search interface with a search bar, filters for 'First Year' and 'Last Year', and checkboxes for 'Medline (1998 - 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>

The screenshot shows the search results page in a Mozilla Firefox browser. The search query is "artificial intelligence". The page displays the following information:

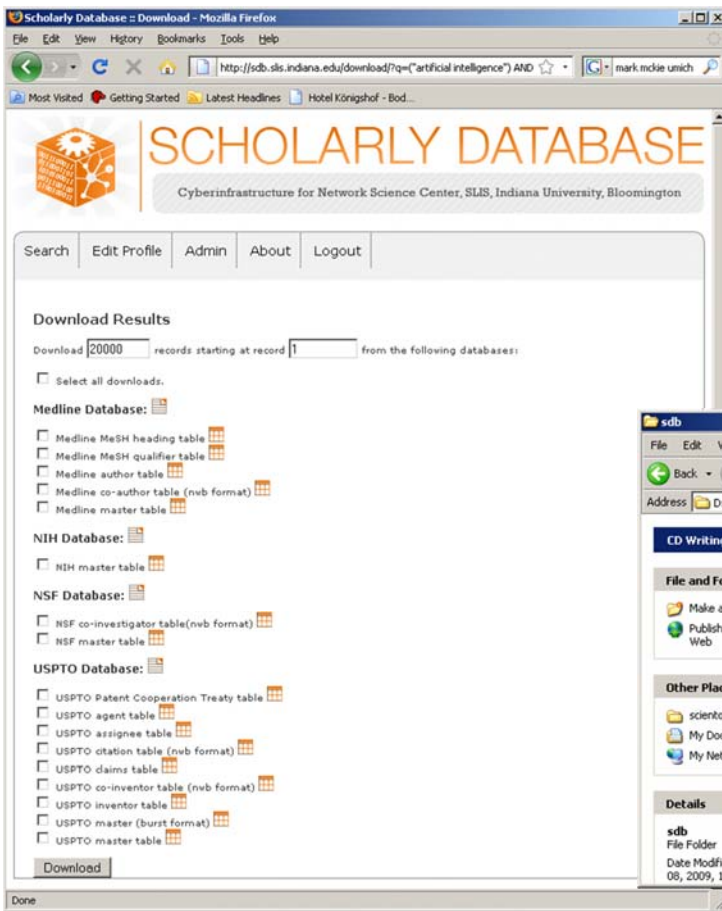
Your search returned 13,231 results in 0.295 seconds. [Download](#)

Total results per database: NIH: 2,103, Medline: 10,235, USPTO: 279, NSF: 614.

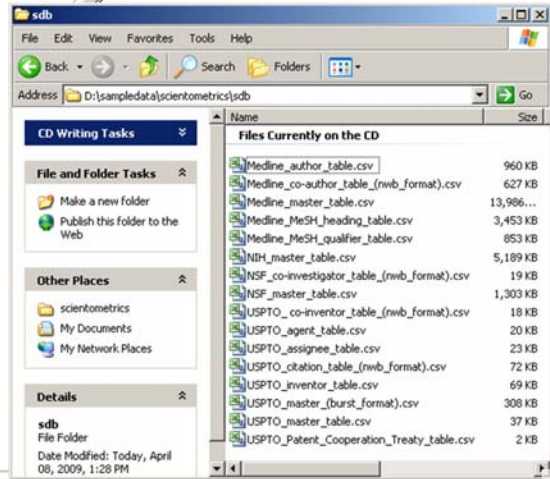
Results 1 through 20.

Next>>

| Source | Authors/Creators | Year | Title | Score (out of 5.71) |
|---------|-------------------------|------|--|---------------------|
| Medline | LaCombe | 1987 | Artificial intelligence. | 5.71 |
| Medline | | 1989 | Artificial intelligence: expert systems. | 5.71 |
| Medline | Schmitt | 1990 | [Artificial intelligence in dentistry] | 5.71 |
| Medline | Adlassnig and Adlassnig | 2002 | Artificial-intelligence-augmented systems. | 5.60 |
| Medline | Touretzky | 1980 | Artificial intelligence. | 4.86 |
| Medline | Goldenberg | 1980 | Artificial intelligence. | 4.86 |



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



27



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](#)

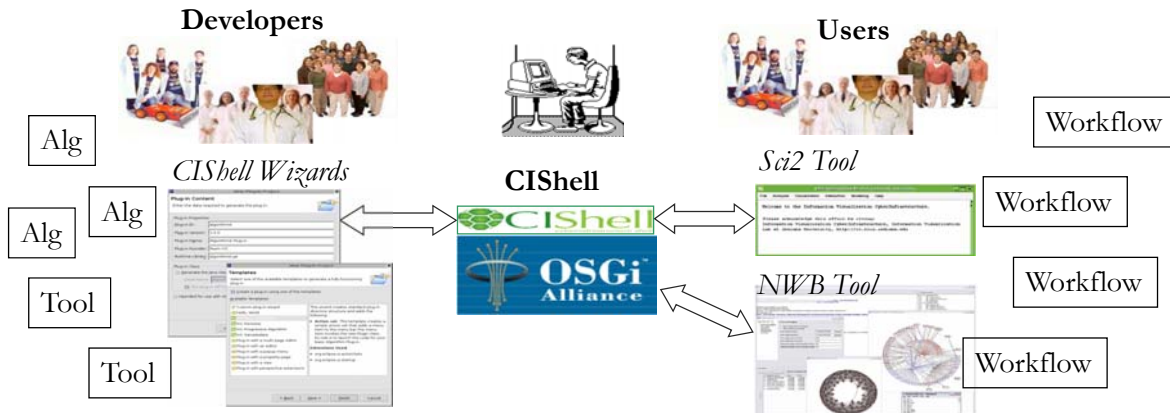
Getting Involved...

- [Contact Us](#)

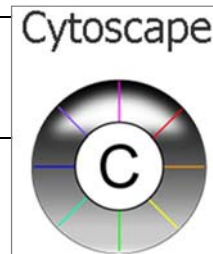
CIShell Developer Guide is at <http://cishell.wiki.cns.iu.edu>

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

- CIShell (<http://cishell.org>) is an open source software specification for the integration and utilization of datasets, algorithms, and tools.
- It extends the Open Services Gateway Initiative (OSGi) (<http://osgi.org>), a standardized, component oriented, computing environment for networked services widely used in industry since more than 10 years.
- Specifically, CIShell provides “sockets” into which existing and new datasets, algorithms, and tools can be plugged using a wizard-driven process.



29

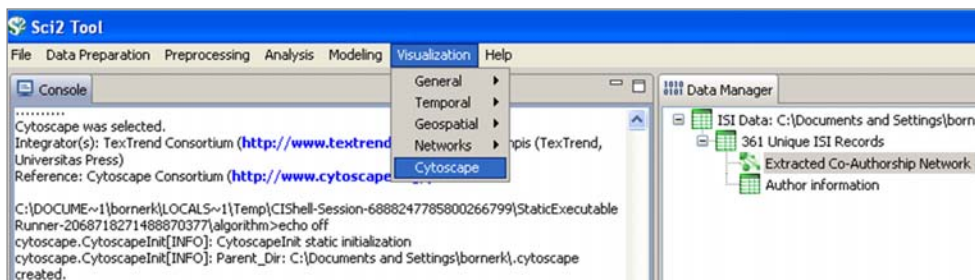


Adding more alyout algorithms and network visualization interactivity via Cytoscape <http://www.cytoscape.org>.

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

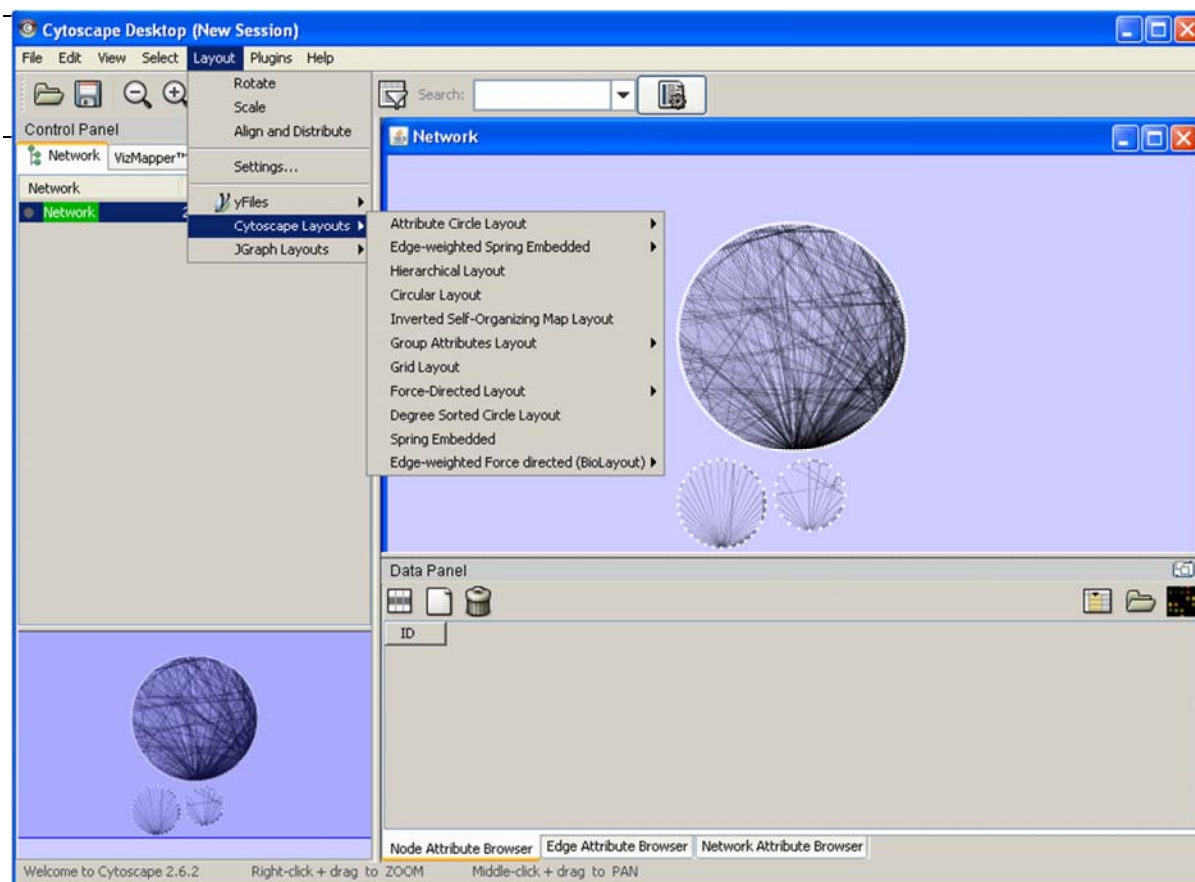
Restart Sci2 Tool.

Cytoscape now shows in the Visualization Menu.



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

30



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 Kamps at Eötvös Loránd University, Budapest, Hungary supports natural language processing (NLP), classification/mining, and graph algorithms for the analysis of business and governmental text corpuses with an inherently temporal component.
 - *DynaNets* (<http://www.dynanets.org>) Coordinated by Peter M.A. Sloot at the University of Amsterdam, The Netherlands develops algorithms to study evolving networks.
 - *SISOB* (<http://sisob.lcc.uma.es>) An Observatory for Science in Society Based in Social Models.
- As the functionality of OSGi-based software frameworks improves and the number and diversity of dataset and algorithm plugins increases, the capabilities of custom tools will expand.

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

33



Network Workbench Tool

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

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

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

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

It has been downloaded more than 100,000 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, Mibail C. (Eds.), *Progress in Convergence - Technologies for Human Wellbeing* (Vol. 1093, pp. 161-179), *Annals of the New York Academy of Sciences*, Boston, MA.

34

