

Network Workbench

A Workbench for Network Scientists

Network Workbench Tool For Network Analysis, Modeling, and Visualization

Four-Hour Workshop

Katy Börner and the NWB Team @ IUB

Victor H. Yngve Professor of Information Science

Director of the Cyberinfrastructure for Network Science Center

School of Library and Information Science, Indiana University

10th Street & Jordan Avenue, Wells Library 021

Bloomington, IN. 47405, USA

E-mail: katy@indiana.edu

Network Workbench (<http://nwb.slis.indiana.edu>).

1



The Changing Scientific Landscape

Star Scientist -> Research Teams: In former times, science was driven by key scientists. Today, science is driven by effectively collaborating co-author teams often comprising expertise from multiple disciplines and several geospatial locations (Börner, Dall'Asta, Ke, & Vespignani, 2005; Shneiderman, 2008).

Users -> Contributors: Web 2.0 technologies empower anybody to contribute to Wikipedia and to exchange images and videos via Flickr and YouTube. WikiSpecies, WikiProfessionals, or WikiProteins combine wiki and semantic technology in support of real time community annotation of scientific datasets (Mons et al., 2008).

Cross-disciplinary: The best tools frequently borrow and synergistically combine methods and techniques from different disciplines of science and empower interdisciplinary and/or international teams of researchers, practitioners, or educators to fine-tune and interpret results collectively.

One Specimen -> Data Streams: Microscopes and telescopes were originally used to study one specimen at a time. Today, many researchers must make sense of massive streams of multiple types of data with different formats, dynamics, and origin.

Static Instrument -> Evolving Cyberinfrastructure (CI): The importance of hardware instruments that are rather static and expensive decreases relative to software infrastructures that are highly flexible and continuously evolving according to the needs of different sciences. Some of the most successful services and tools are decentralized increasing scalability and fault tolerance.

Modularity: The design of software modules with well defined functionality that can be flexibly combined helps reduce costs, makes it possible to have many contribute, and increases flexibility in tool development, augmentation, and customization.

Standardization: Adoption of standards speeds up development as existing code can be leveraged. It helps pool resources, supports interoperability, but also eases the migration from research code to production code and hence the transfer of research results into industry applications and products.

Open data and open code: Lets anybody check, improve, or repurpose code and eases the replication of scientific studies.



Desirable Features of Plug-and-Play Macroscopes

Division of Labor: Ideally, labor is divided in a way that the expertise and skills of computer scientists are utilized for the design of standardized, modular, easy to maintain and extend “core architecture”. Dataset and algorithm plugins, i.e., the “filling”, are initially provided by those that care and know most about the data and developed the algorithms: the domain experts.

Ease of Use: As most plugin contributions and usage will come from non-computer scientists it must be possible to contribute, share, and use new plugins without writing one line of code. Wizard-driven integration of new algorithms and data sets by domain experts, sharing via email or online sites, deploying plugins by adding them to the ‘plugin’ directory, and running them via a Menu driven user interfaces (as used in Word processing systems or Web browsers) seems to work well.

Plugin Content and Interfaces: Should a plugin represent one algorithm or an entire tool? What about data converters needed to make the output of one algorithm compatible with the input of the next? Should those be part of the algorithm plugin or should they be packaged separately?

Supported (Central) Data Models: Some tools use a central data model to which all algorithms conform, e.g., Cytoscape, see Related Work section. Other tools support many internal data models and provide an extensive set of data converters, e.g., Network Workbench, see below. The former often speeds up execution and visual rendering while the latter eases the integration of new algorithms. In addition, most tools support an extensive set of input and output formats.

Core vs. Plugins: As will be shown, the “core architecture” and the “plugin filling” can be implemented as sets of plugin bundles. Answers to questions such as: “Should the graphical user interface (GUI), interface menu, scheduler, or data manager be part of the core or its filling?” will depend on the type of tools and services to be delivered.

Supported Platforms: If the software is to be used via Web interfaces then Web services need to be implemented. If a majority of domain experts prefers a stand-alone tool running on a specific operating system then a different deployment is necessary.

NetworkWorkbench
A Workbench for Network Scientists

Project Details

Investigators: Katy Börner, Albert-Laszlo Barabasi, Santiago Schnell, Alessandro Vespignani & Stanley Wasserman, Eric Wernert



Software Team: Lead: Micah Linnemeier
Members: Patrick Phillips, Russell Duhon, Tim Kelley & Ann McCranie
Previous Developers: Weixia (Bonnie) Huang, Bruce Herr, Heng Zhang, Duygu Balcan, Bryan Hook, Ben Markines, Santo Fortunato, Felix Terkhorn, Ramya Sabbineni, Vivek S. Thakre & Cesar Hidalgo



Goal: Develop a large-scale network analysis, modeling and visualization toolkit for physics, biomedical, and social science research.

Amount: \$1,120,926, NSF IIS-0513650 award

Duration: Sept. 2005 - Aug. 2009

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

NWB Advisory Board:

James Hendler (Semantic Web) <http://www.cs.umd.edu/~hendler/>

Jason Leigh (CI) <http://www.evl.uic.edu/spiff/>

Neo Martinez (Biology) <http://online.sfsu.edu/~webhead/>

Michael Macy, Cornell University (Sociology)
<http://www.soc.cornell.edu/faculty/macy.shtml>

Ulrik Brandes (Graph Theory) <http://www.inf.uni-konstanz.de/~brandes/>

Mark Gerstein, Yale University (Bioinformatics) <http://bioinfo.mbb.yale.edu/>

Stephen North (AT&T) <http://public.research.att.com/viewPage.cfm?PageID=81>

Tom Snijders, University of Groningen <http://stat.gamma.rug.nl/snijders/>

Noshir Contractor, Northwestern University <http://www.spcomm.uiuc.edu/nosh/>

**Publications**

- o <http://nwb.slis.indiana.edu/pub.html>

Community Wiki, Tutorials, FAQ

- o <https://nwb.slis.indiana.edu/community>
- o <http://nwb.slis.indiana.edu/doc.html>
- o GUESS Manual <http://guess.wikispot.org/manual>

Software

- o <http://cishell.org>
- o <http://nwb.slis.indiana.edu/download.html>

Developer Resources

- o <http://cns-trac.slis.indiana.edu/trac/nwb>

1. Exemplary Network Science Research by NWB PIs
 - Computational Proteomics
 - Computational Economics
 - Computational Social Science
 - Computational Scientometrics
 - Computational Epidemics
2. NWB Tool Challenges and Opportunities
3. NWB Tool Overview
4. NWB Tool for Scientometrics Research
5. Discussion of Future Work

Computational Proteomics

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

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

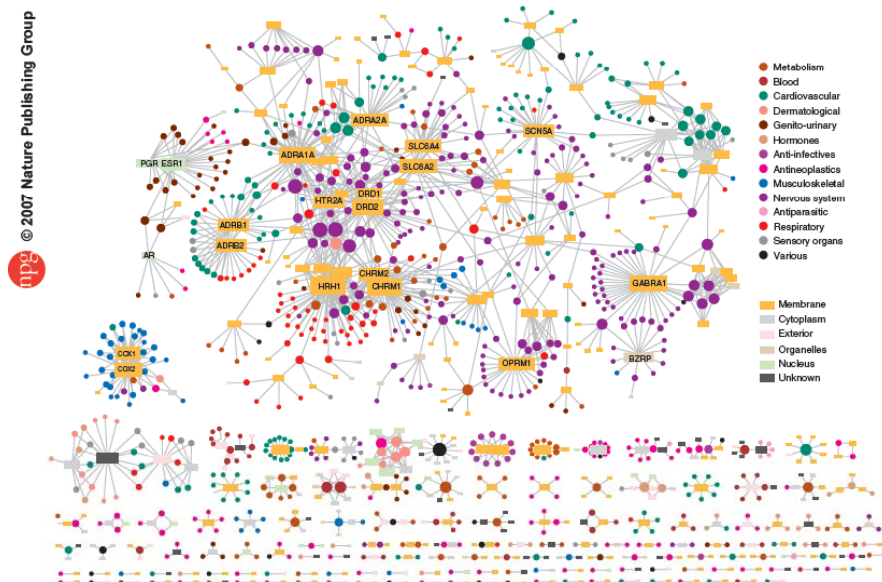


Figure 2 Drug-target network (DT network). The DT network is generated by using the known associations between FDA-approved drugs and their target proteins. Circles and rectangles correspond to drugs and target proteins, respectively. A link is placed between a drug node and a target node if the protein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

Computational Economics

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

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

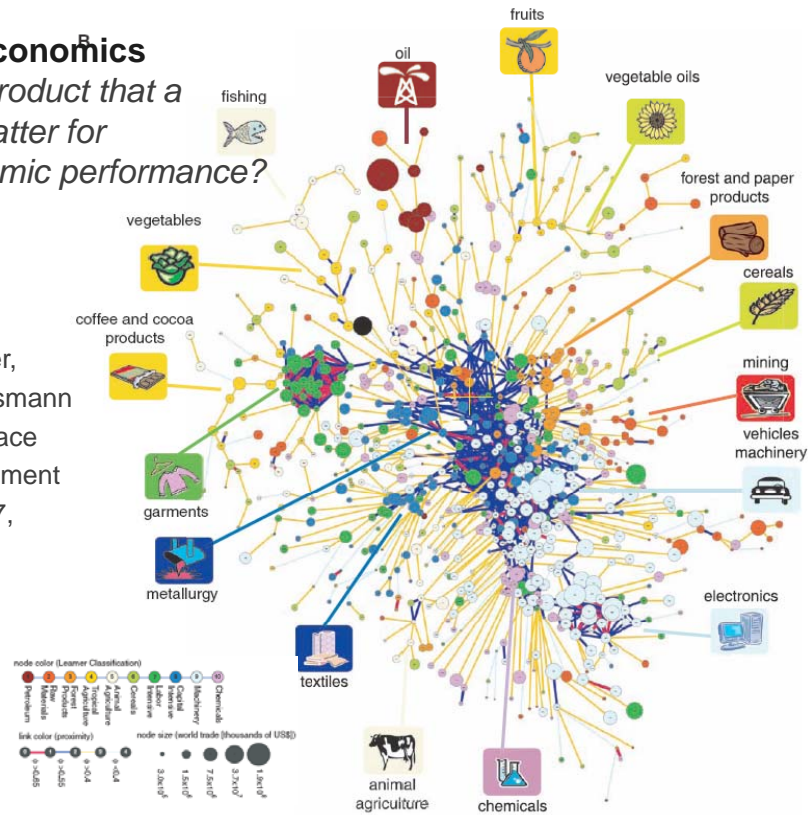


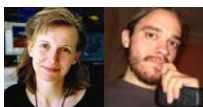
Fig. 1. The product space. (A) Hierarchically clustered proximity matrix representing the 775 SITC-4 product classes exported in the 1998–2000 period. (B) Network representation of the product space. Links are color coded with their proximity value. The sizes of the nodes are proportional to world trade,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

Computational Social Science

Studying large scale social networks such as Wikipedia

Vizzards 2007 Entry

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



Second sight

Image: Bruce W. Herr and Todd M. Holloway

Power struggle

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



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

Computational Scientometrics

113 Years of Physical Review

Bruce W. Herr II and Russell Duhon (Data Mining & Visualization), Elisha F. Hardy (Graphic Design), Shashikant Penumarthi (Data Preparation) and Katy Börner (Concept)

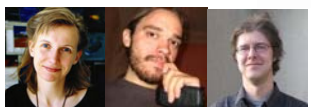
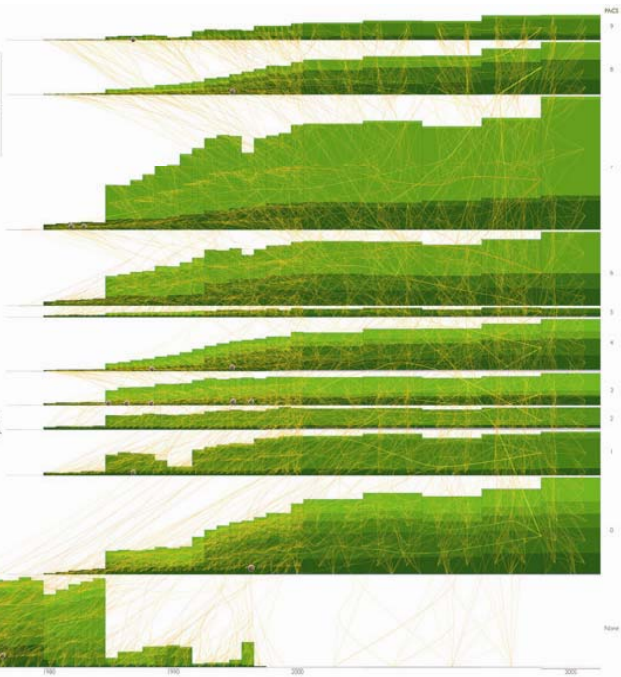
113 Years of Physical Review

The American Physical Society (APS) celebrated its 113th anniversary of founding in 1906 and 2019. The APS is the largest and most influential organization in the field of physics. It is the largest and most influential organization in the field of physics. It is the largest and most influential organization in the field of physics.

Nobel Prizes in Physical Review

- 2005 Roy J. Glauber (John L. Hall and Theodor W. Hänsch)
- 2004 David Gross, H. David Politzer and Frank Wilczek
- 2003 Raymond Davis Jr., Masatoshi Koshiba and Riccardo Giacconi
- 2002 Anthony J. Leggett
- 2001 Eric A. Cornell, Wolfgang Ketterle and Carl E. Wieman
- 1998 Steven Chu and William D. Phillips
- 1994 David H. Lee, Douglas O. Richardson and Robert C. Richardson
- 1993 Martin L. Perl
- 1991 Benjamin H. Brattinhouse and Clifford G. Shull
- 1989 Jerome I. Friedman, Henry W. Kendall and Richard E. Taylor

- Physical Review
- Physical Review Series I
- Physical Review A
- Physical Review B
- Physical Review C
- Physical Review D
- Physical Review E
- Physical Review Letters
- Physical Review Special Topics: Accelerated Beams
- Physical Review Physical Educational Research
- Physical Review Physical Review Physics
- PHYSICS General
- PHYSICS Interdisciplinary Physics and Related Areas of Science and Technology
- PHYSICS The Physics of Elementary Particles
- PHYSICS Electromagnetism, Optics, Acoustics, and Fluid Dynamics
- PHYSICS Nuclear Physics
- PHYSICS Atomic and Molecular Physics
- PHYSICS Physics of Gases, Plasmas, and Electrodynamics
- PHYSICS Geophysics, Astrophysics, and Aerophysics
- PHYSICS Condensed Matter: Structural, Mechanical and Thermal Properties
- PHYSICS Condensed Matter: Electrical, Structural, Electrical, Magnetic, and Critical Properties



Computational Epidemics

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

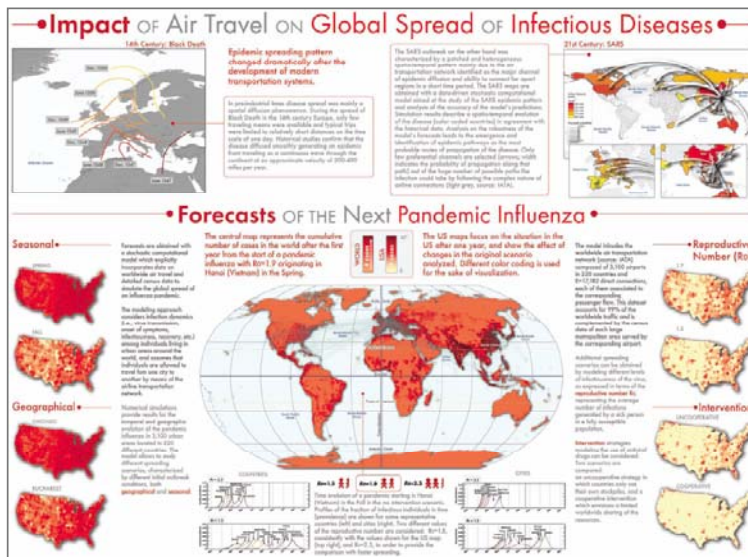
Epidemic Modeling in Complex realities, V. Colizza, A. Barrat, M. Barthelemy, A. Vespignani, Comptes Rendus Biologie, 330, 364-374 (2007).

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

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



Network Workbench (<http://nwb.slis.indiana.edu>).



- Data
 - Different data formats
 - Different data models
- Algorithms
 - Different research purposes (preprocessing, modeling, analysis, visualization, clustering)
 - Different implementations of the same algorithm
 - Different programming languages
 - Algorithm developers/users are not computer scientists
- Different tools (Pajek, UCINet, Guess, Cytoscape, R, ...)
- Different communities, practices, cultures

Network Workbench (NWB) Tool

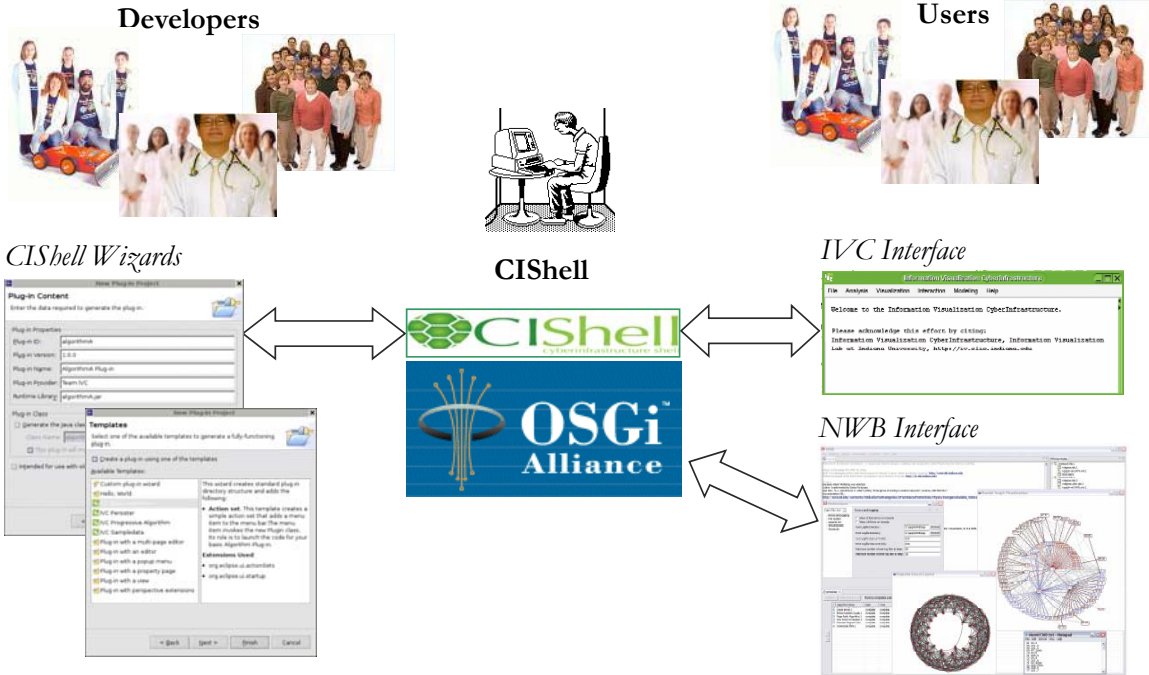
- A network analysis, modeling, and visualization toolkit for physics, biomedical, and social science research.
- Install and run on multiple Operating Systems.
- Supports many file formats.
- Easy integration of new algorithms thanks to CShell/OSGi.

Cyberinfrastructure Shell (CShell)

- An open source, software framework for the integration and utilization of datasets, algorithms, tools, and computing resources.
- Extends OSGi industry standard.



CIShell – Serving Non-CS Algorithm Developers & Users



Network Workbench (<http://nwb.slis.indiana.edu>).

15



CIShell – Builds on OSGi Industry Standard

CIShell is built upon the Open Services Gateway Initiative (OSGi) Framework.

OSGi (<http://www.osgi.org>) is

- A standardized, component oriented, computing environment for networked services.
- Successfully used in the industry from high-end servers to embedded mobile devices since 8 years.
- Alliance members include IBM (Eclipse), Sun, Intel, Oracle, Motorola, NEC and many others.
- Widely adopted in open source realm, especially since Eclipse 3.0 that uses OSGi R4 for its plugin model.

Advantages of Using OSGi

- Any CIShell algorithm is a service that can be used in any OSGi-framework based system.
- Using OSGi, running CIShells/tools can be connected via RPC/RMI supporting peer-to-peer sharing of data, algorithms, and computing power.

Ideally, CIShell becomes a standard for creating OSGi Services for algorithms.

Network Workbench (<http://nwb.slis.indiana.edu>).

16

Network Workbench (NWB) Tool

- A network analysis, modeling, and visualization toolkit for physics, biomedical, and social science research.
- Install and run on multiple Operating Systems.
- Supports many file formats.
- Easy integration of new algorithms thanks to CShell/OSGi.

Cyberinfrastructure Shell (CShell)

- An open source, software framework for the integration and utilization of datasets, algorithms, tools, and computing resources.
- Extends OSGi industry standard.

NWB Community Wiki

- A place for users of the NWB Tool, the Cyberinfrastructure Shell (CShell), or any other CShell-based program to request, obtain, contribute, and share algorithms and datasets.
- All algorithms and datasets that are available via the NWB Tool have been well documented in the Community Wiki.

Network Workbench (<http://nwb.slis.indiana.edu>).

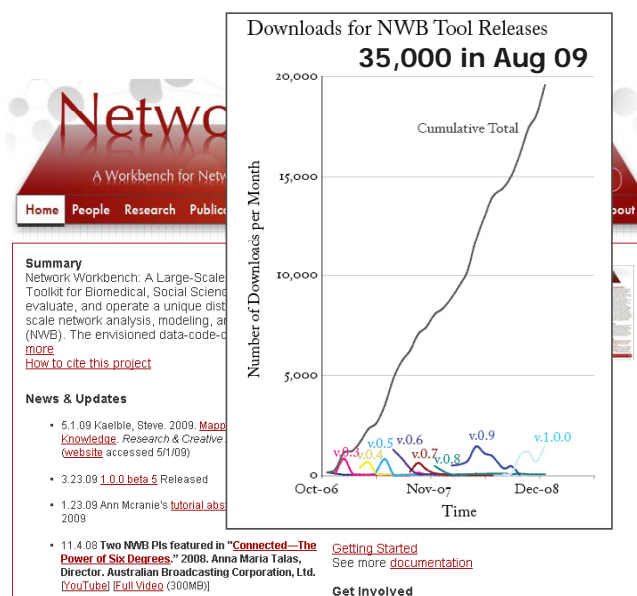
17

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 Aug. 2009, the tool provides more 160 plugins that support the preprocessing, analysis, modeling, and visualization of networks.

More than 40 of these plugins can be applied or were specifically designed for S&T studies.

It has been downloaded more than 35,000 times since Dec. 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.

<https://nwb.slis.indiana.edu/community/>

Network Workbench (<http://nwb.slis.indiana.edu>).

19



NWB Tool: Supported Data Formats

Personal Bibliographies

- Bibtext (.bib)
- Endnote Export Format (.enw)

Data Providers

- Web of Science by Thomson Scientific/Reuters (.isi)
- Scopus by Elsevier (.scopus)
- Google Scholar (access via *Publish or Perish* save as CSV, Bibtext, EndNote)
- Awards Search by National Science Foundation (.nsf)

Scholarly Database (all text files are saved as .csv)

- Medline publications by National Library of Medicine
- NIH funding awards by the National Institutes of Health (NIH)
- NSF funding awards by the National Science Foundation (NSF)
- U.S. patents by the United States Patent and Trademark Office (USPTO)
- Medline papers – NIH Funding

Network Formats

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

Burst Analysis Format

- Burst (.burst)

Other Formats

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

Preprocessing Edit

Remove Nodes

- [Extract Top Nodes](#)
- [Extract Nodes Above or Below Val](#)
- [Delete High Degree Nodes](#)
- [Delete Random Nodes](#)
- [Delete Isolates](#)

Remove Edges

- [Extract Top Edges](#)
- [Extract Edges Above or Below Val](#)
- [Remove Self Loops](#)
- [Trim By Degree²](#)
- [Pathfinder Network Scaling](#)

Sampling

- [Snowball Sampling \(n nodes\)](#)
- [Node Sampling](#)
- [Edge Sampling](#)

Transformations

- [Symmetrize](#)
- [Dichotomize](#)
- [Multipartite Joining](#)

Modeling Edit

General

- [Random Graph](#)
- [Watts-Strogatz Small World](#)
- [Barabási-Albert Scale-Free](#)

Structured

- [CAN](#)
- [Chord](#)

Unstructured

- [Hypergrid](#)
- [PRU](#)

Other

- [TARL](#)
- [Discrete Network Dynamics](#)

Analysis Edit

General Purpose

- [Network Analysis Toolkit²](#)

Unweighted & Undirected

- [Based on degree/](#)
- [Node Degree](#)
- [Node Distribution](#)

Based on clustering

- [k-Nearest Neighbor](#)
- [Watts Strogatz Clustering Coefficient](#)
- [Watts Strogatz Clustering Coefficient Over k](#)

Based on path

- [Diameter](#)
- [Average Shortest Path](#)
- [Shortest Path Distribution](#)
- [Node Betweenness Centrality](#)

Based on components

- [Connected Components](#)
- [Weak Component Clustering](#)

K-Core

- [Extract K-Core²](#)
- [Annotate K-Core²](#)

Unweighted & Directed

Based on degree

- [Node Indegree](#)
- [Node Outdegree](#)
- [Indegree Distribution](#)
- [Outdegree Distribution](#)

Based on local graph structure

- [k-Nearest Neighbor](#)
- [Single Node In-Out Degree Correlations²](#)

Unnamed Category?

- [Page Rank](#)

Based on local graph structure #2

- [Dvad Reciprocity²](#)
- [Arc Reciprocity²](#)

Layout Edit

Tools

- [GUESS](#)
- [GnuPlot²](#)

Predefined Positions Layout

- [DrL \(VxOrd\)](#)
- [Pre-defined Positions \(prefuse beta\)²](#)

Move

- [Circular](#)

Tree Layouts

- [Radial Tree \(prefuse alpha\)](#)
- [Radial Tree with Annotations \(prefuse beta\)²](#)
- [Tree Map](#)
- [Tree View](#)
- [Balloon Graph \(prefuse alpha\)²](#)

Network Layouts

- [Force Directed with Annotation \(prefuse beta\)](#)
- [Kamada-Kawai \(JUNG\)](#)
- [Fruchterman-Reingold \(JUNG\)](#)
- [Fruchterman-Reingold with Annotation \(prefuse beta\)](#)
- [Spring \(JUNG\)](#)
- [Small World \(prefuse alpha\)](#)

Other Layouts

- [Parallel Coordinates \(demo\)²](#)
- [LaNet \(k-Core Decomposition\)](#)

Metrics Edit

Extract Network From Table

- [Extract Co-Authorship Network](#)
- [Extract Co-Occurrence Network From Table²](#)
- [Extract Directed Network From Table²](#)

Extract Network From Another Network

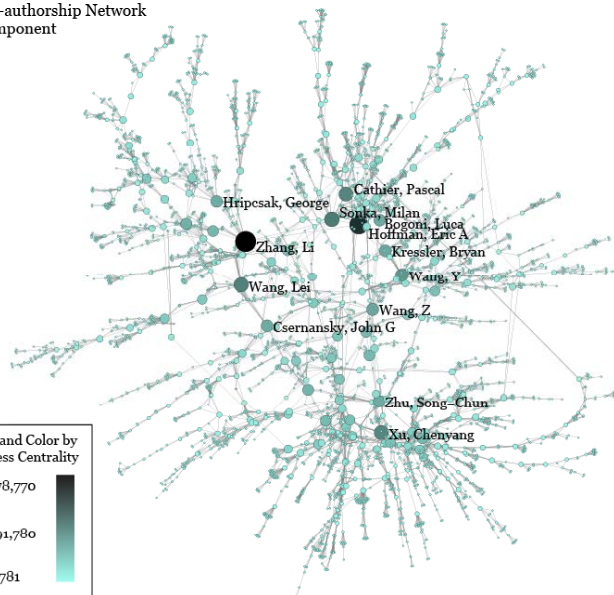
- [Extract Bibliographic Coupling Similarity Network](#)
- [Extract Co-Citation Similarity Network²](#)

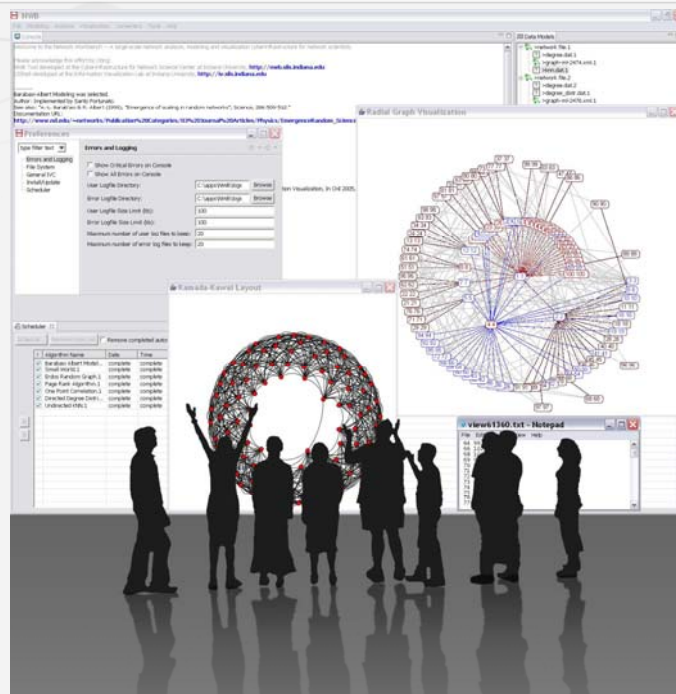
Cleaning

- [Remove ISI Duplicate Records](#)

- NWB tool can be used for data conversion. Supported output formats comprise:
 - CSV (.csv)
 - NWB (.nwb)
 - Pajek (.net)
 - Pajek (.mat)
 - GraphML (.xml or .graphml)
 - XGMML (.xml)
- GUESS
 - Supports export of images into common image file formats.
- Horizontal Bar Graphs
- saves out raster and ps files.

Medline Co-authorship Network
Largest Component





Network Workbench (<http://nwb.slis.indiana.edu>).

23

NWB Tool Overview

1. Download, install, and run.
2. Load, view, convert, save data.
3. Read and visualize a directory hierarchy.
4. Load a network, compute its basic properties, and explore it in GUESS.

24

NWB Tool Overview

1. Download, install, and run.
2. Load, view, convert, save data.
3. Read and visualize a directory hierarchy.
4. Load a network, compute its basic properties, and explore it in GUESS.

Download, install, and run

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

NWB Tool 1.0.0

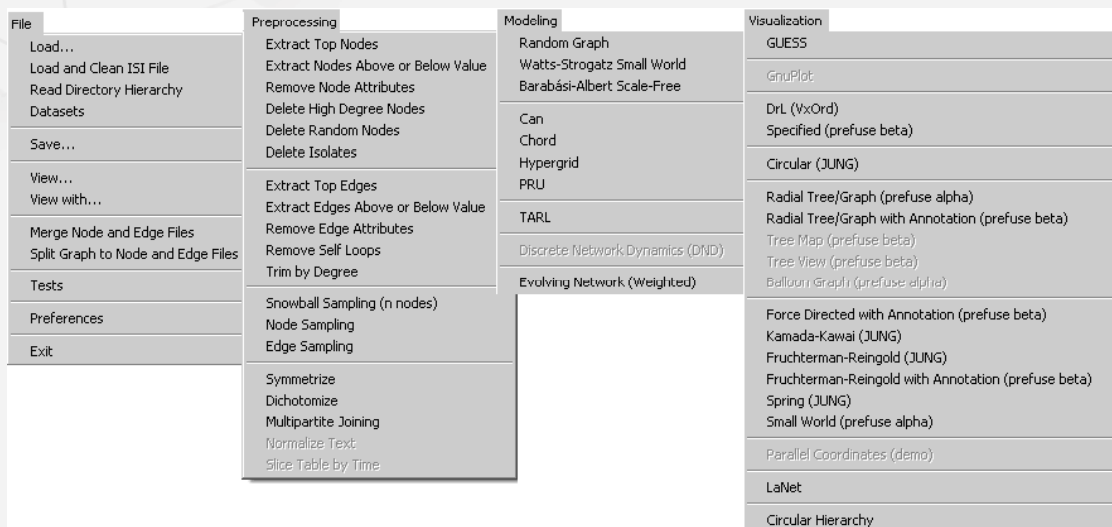
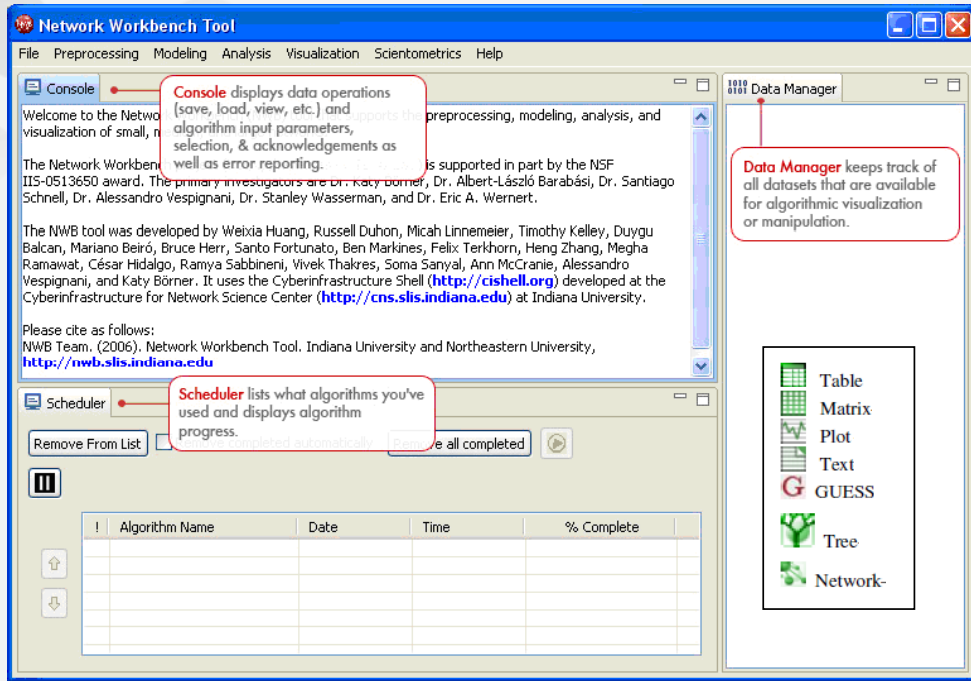
Select your operating system from the pull down menu.

Save as *.jar file.

Install and run.

Session log files are stored in '**yournwbdirectory*/logs*' directory.

The screenshot shows a Mozilla Firefox browser window displaying the Network Workbench website. The page title is "Network Workbench | Download - Mozilla Firefox". The address bar shows the URL "http://nwb.slis.indiana.edu/download.html". The website header includes the Network Workbench logo and navigation links: Home, People, Research, Publications, Community, Download, Documentation, Dev Zone, About. The main content area is titled "Download" and features two download options. The first option is for "NWB Tool 1.0.0 beta 2 (development release)" dated November 08th, 2008, with a note that it is the most up-to-date but not fully tested. It includes a "Select Your Operating System" dropdown menu set to "Windows (XP & Vista)" and a "DOWNLOAD" button. The second option is for "NWB Tool 1.0.0 beta Release" dated September 24th, 2008, with a note to save the download as a jar file. It also has a "Select Your Operating System" dropdown menu set to "Windows (XP & Vista)" and a "DOWNLOAD" button. To the right of the download options is a line graph titled "Downloads for NWB Tool Releases" showing the number of downloads per month from Dec 2004 to Apr 2008. The graph shows a cumulative total of downloads that increases over time, with a significant spike in late 2007. The y-axis is labeled "Number of Downloads per Month" and ranges from 0 to 14,000. The x-axis is labeled "Time" and shows months from Dec 2004 to Apr 2008. A legend indicates "Cumulative Total".



The screenshot displays the 'Analysis' menu in Network Workbench, which is organized into several submenus:

- Analysis**
 - Network Analysis Toolkit (NAT)
 - Unweighted and Undirected
 - Weighted and Undirected
 - Unweighted and Directed
 - Weighted and Directed
 - Search
 - Discrete Network Dynamics
 - Textual
- Unweighted and Undirected**
 - Node Degree
 - Degree Distribution
 - Watts-Strogatz Clustering Coefficient
 - Watts Strogatz Clustering Coefficient over K
 - Diameter
 - Average Shortest Path
 - Shortest Path Distribution
 - Node Betweenness Centrality
 - Global Connected Components
 - HITS
 - Weak Component Clustering
 - Blondel Community Detection
 - MST-Pathfinder Network Scaling
 - Extract K-Core
 - Annotate K-Core
- Weighted and Undirected**
 - Clustering Coefficient
 - Nearest Neighbor Degree
 - Strength vs Degree
 - Degree & Strength
 - Average Weight vs End-point Degree
 - K-Nearest Neighbor (Java)
 - Strength Distribution
 - Weight Distribution
 - Randomize Weights
 - MST-Pathfinder Network Scaling
 - Fast Pathfinder Network Scaling
 - Blondel Community Detection
- Unweighted and Directed**
 - Node Indegree
 - Node Outdegree
 - Indegree Distribution
 - Outdegree Distribution
 - K-Nearest Neighbor
 - Single Node In-Out Degree Correlations
 - PageRank
 - HITS
 - Dyad Reciprocity
 - Arc Reciprocity
 - Adjacency Transitivity
 - Weak Component Clustering
 - Strong Component Clustering
 - Blondel Community Detection
 - Extract K-Core
 - Annotate K-Core
- Discrete Network Dynamics**
 - Extract and Annotate Attractors
- Weighted and Directed**
 - HITS
 - Weighted PageRank
 - Fast Pathfinder Network Scaling
 - Blondel Community Detection
- Search**
 - Can
 - Chord
 - k Random-Walk
 - Random Breadth First
- Textual**
 - Burst Detection

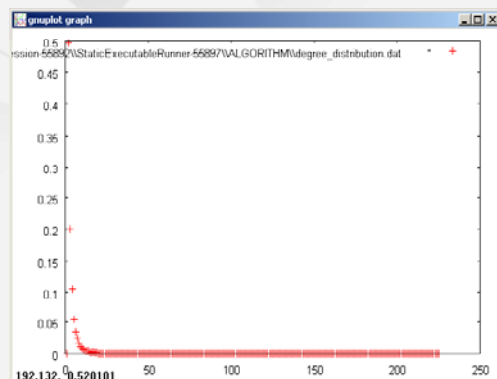
Analysis Menu and Submenus

Network Workbench (<http://nwb.slis.indiana.edu>)

29

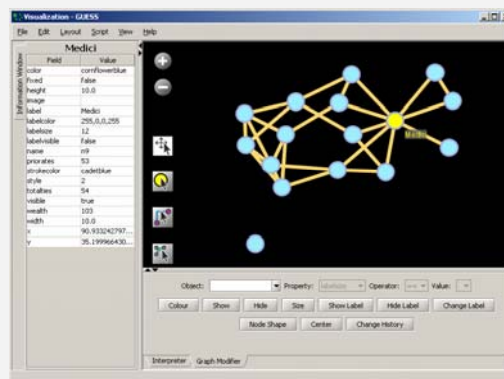
NetworkWorkbench **Integrated Tools**

A Workbench for Network Scientists



Gnuplot

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

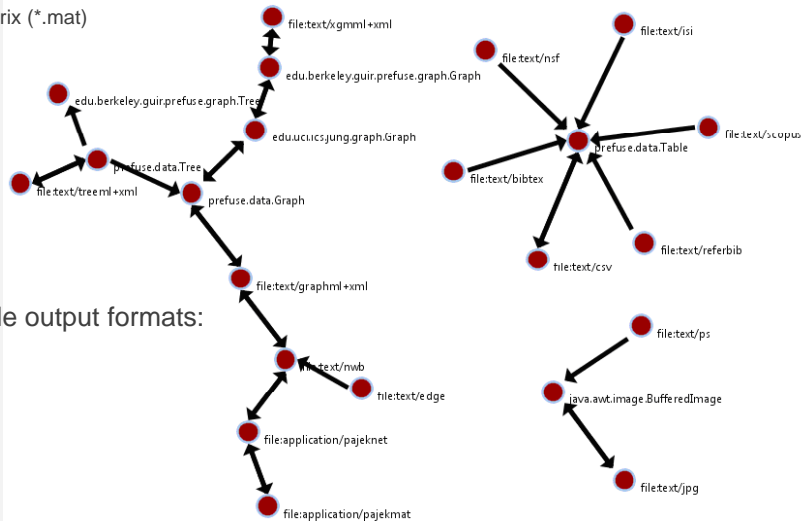


GUESS

exploratory data analysis and visualization tool for graphs and networks.
<https://nwb.slis.indiana.edu/community/?n=VisualizeData.GUESS>

In November 2008, the NWB tool supports loading the following input file formats:

- GraphML (*.xml or *.graphml)
- XGMML (*.xml)
- Pajek .NET (*.net) & Pajek .Matrix (*.mat)
- NWB (*.nwb)
- TreeML (*.xml)
- Edge list (*.edge)
- CSV (*.csv)
- ISI (*.isi)
- Scopus (*.scopus)
- NSF (*.nsf)
- Bibtex (*.bib)
- Endnote (*.enw)



and the following network file output formats:

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

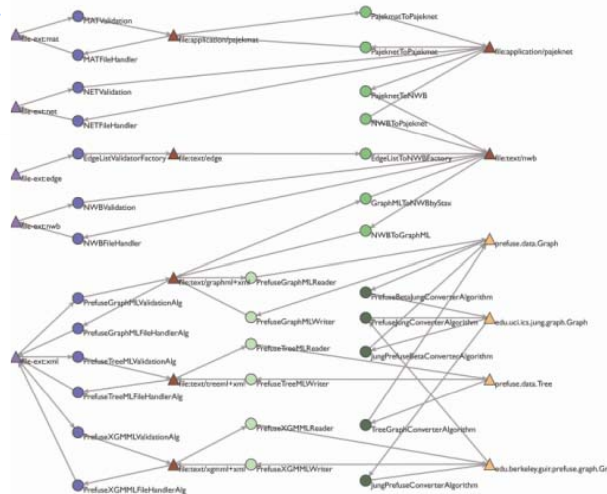
These formats are documented at

<https://nwb.slis.indiana.edu/community/?n=DataFormats.HomePage>.

NWB Ecology of Data Formats and Converters

Not shown are **15** sample datasets, **45** data preprocessing, analysis, modeling and visualization algorithms, **9** services.

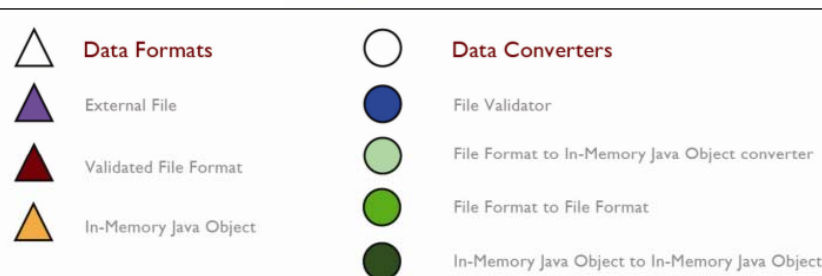
13
Supported data formats



6
Output formats for diverse visualization algorithms

8
Intermediate data formats

Supported by
35
data converters.



VISIT
<http://nwb.slis.indiana.edu>
<https://nwb.slis.indiana.edu/community>
<http://www.cishell.org>

DOWNLOAD: NWB Tool
<http://nwb.slis.indiana.edu/software.html>

The '**yournwbdirectory*/sampledata*' directory provides sample datasets from the biology, network, scientometrics, and social science research domains:

/biology

/network

/scientometrics

/scientometrics/bibtex

/scientometrics/csv

/scientometrics/endnote

/scientometrics/isi

o **FourNetSciResearchers.isi**

/scientometrics/nsf

o **Cornell.nsf**

o **Indiana.nsf**

o **Michigan.nsf**

/scientometrics/scopus

/socialscience

o **florentine.nwb**

The blue ones are used in this tutorial.

The '**yournwbdirectory*/*' directory also contains

/sampledata/scientometrics/properties // Used to extract networks and merge data

- o bibtexCoAuthorship.properties
- o endnoteCoAuthorship.properties
- o isiCoAuthorship.properties
- o isiCoCitation.properties
- o isiPaperCitation.properties
- o mergeBibtexAuthors.properties
- o mergeEndnoteAuthors.properties
- o mergelsiAuthors.properties
- o mergeNsfPIs.properties
- o mergeScopusAuthors.properties
- o nsfCoPI.properties
- o scopusCoAuthorship.properties

/sampledata/scripts/GUESS

// Used to do color/size/shape code networks

- o co-author-nw.py
- o co-PI-nw.py
- o paper-citation-nw.py
- o reference-co-occurrence-nw.py

NWB Tool Overview

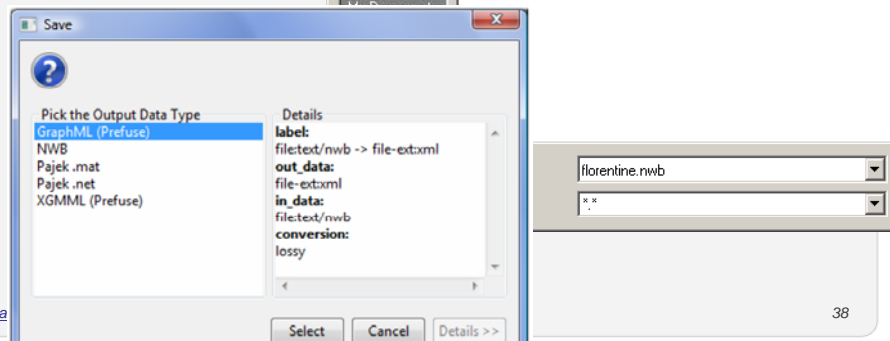
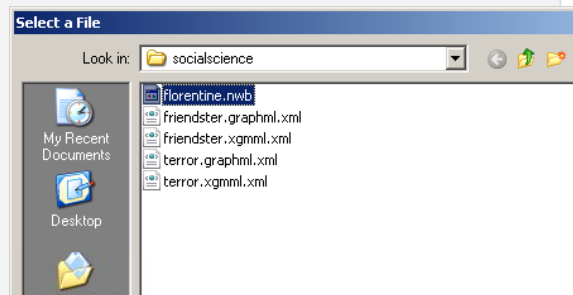
1. Download, install, and run.
2. Load, view, convert, save data.
3. Read and visualize a directory hierarchy.
4. Load a network, compute its basic properties, and explore it in GUESS.

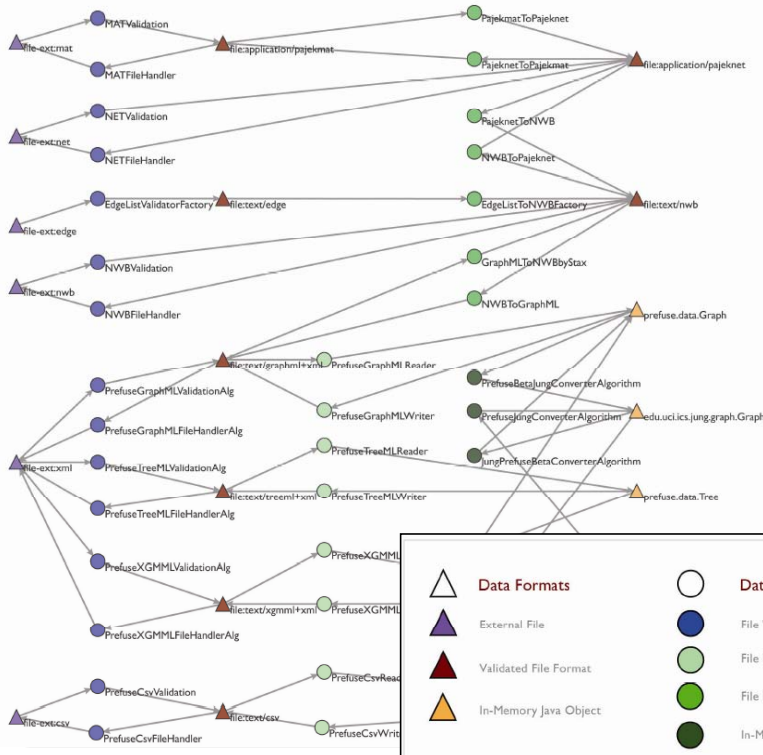
Load, View and Save (Convert) Data

Use 'File > Load File' to load *florentine.nwb* in sample datasets in '**yournwbdirectory*/sampledata/socialscience*'.

The loaded file will appear in the Data Manager window.

Right click loaded file to save, view, rename, or discard.





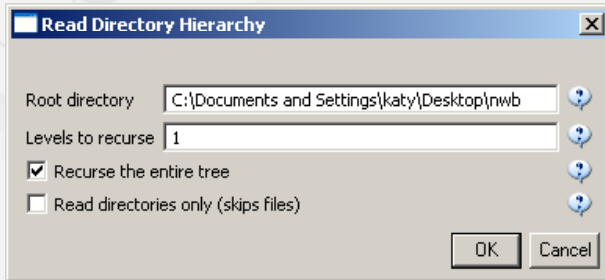
There is no central data format.

Instead, data formats used in different communities and required by the different algorithms are supported.

NWB Tool Overview

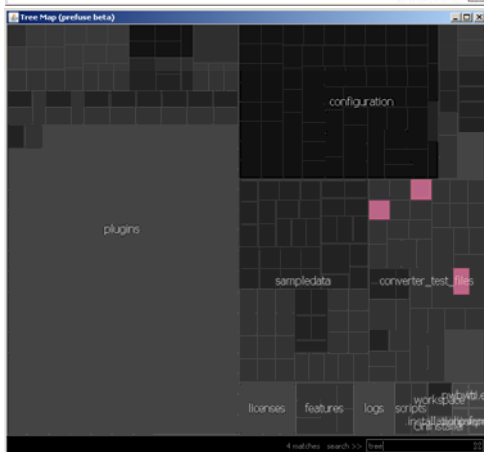
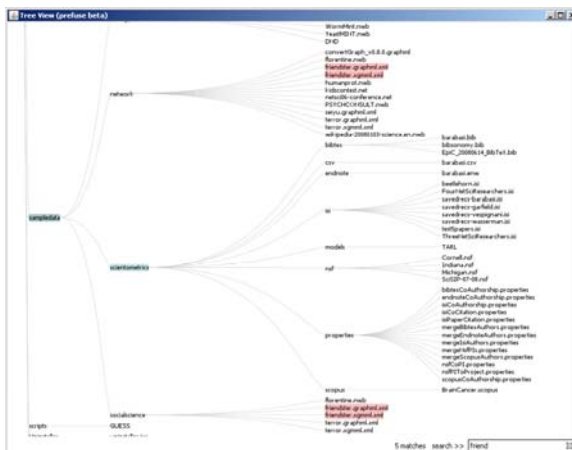
1. Download, install, and run.
2. Load, view, convert, save data.
3. Read and visualize a directory hierarchy.
4. Load a network, compute its basic properties, and explore it in GUESS.

Use 'File > Read Directory Hierarchy' with parameters



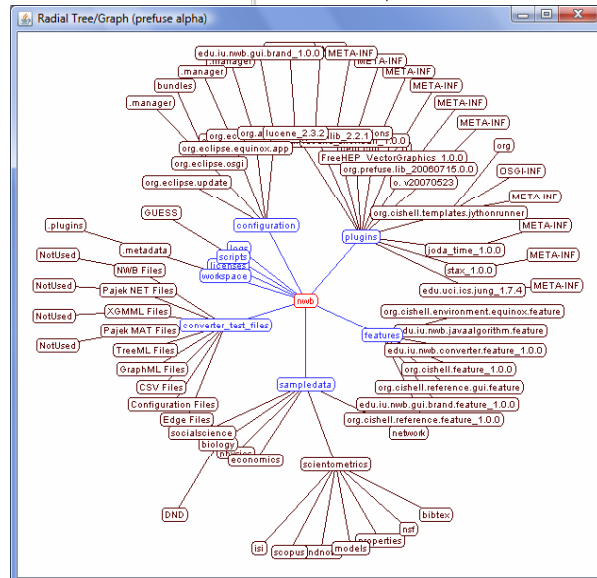
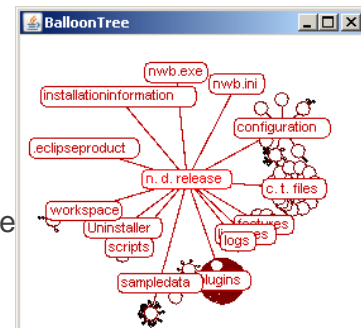
Visualize resulting 'Directory Tree - Prefuse (Beta) Graph' using

- 'Visualization > Tree View (prefuse beta)'
- 'Visualization > Tree Map (prefuse beta)'
- 'Visualization > Balloon Graph (prefuse alpha)'
- 'Visualization > Radial Tree/Graph (prefuse alpha)'



Different views of the /nwb directory hierarchy.

Note the size of the /plugin directory.



NWB Tool Overview

1. Download, install, and run.
2. Load, view, convert, save data.
3. Read and visualize a directory hierarchy.
4. Load a network, compute its basic properties, and explore it in GUESS.

Compute Basic NW Properties & View in GUESS

Select *florentine.nwb* in Data Manager.

- o Run 'Analysis > Network Analysis Toolkit (NAT)' to get basic properties.

```
This graph claims to be undirected.
Nodes: 16
Isolated nodes: 1
Node attributes present: label, wealth, totalties, priorates

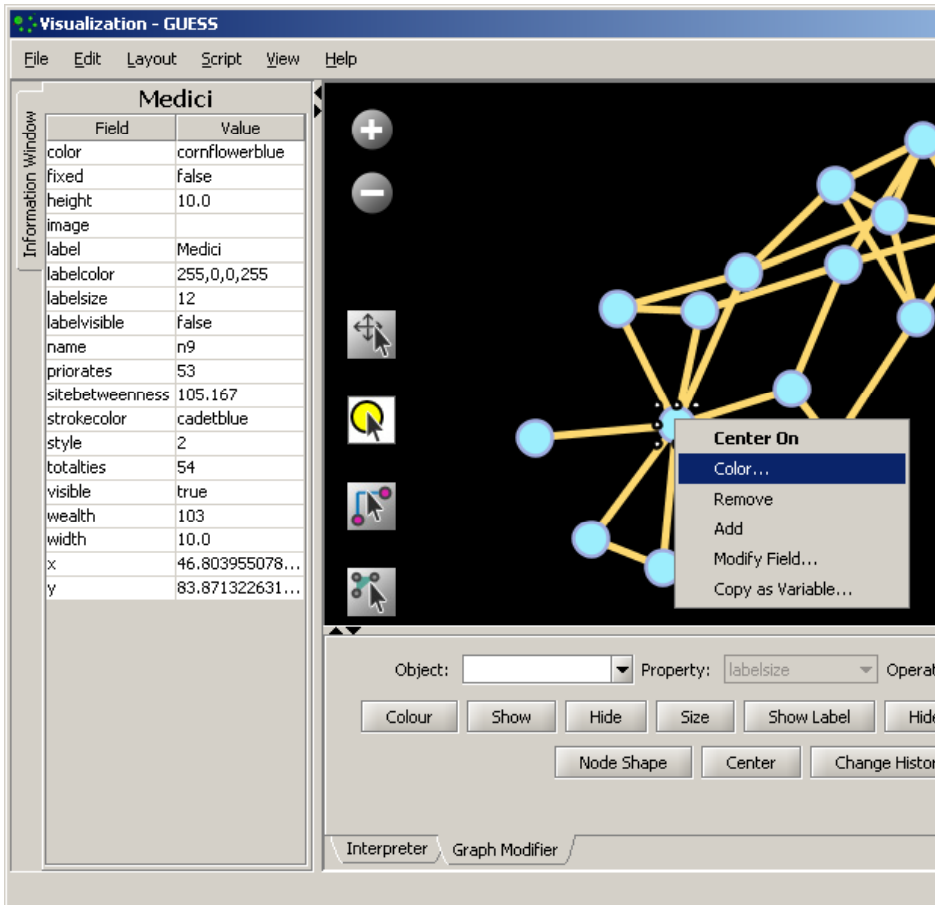
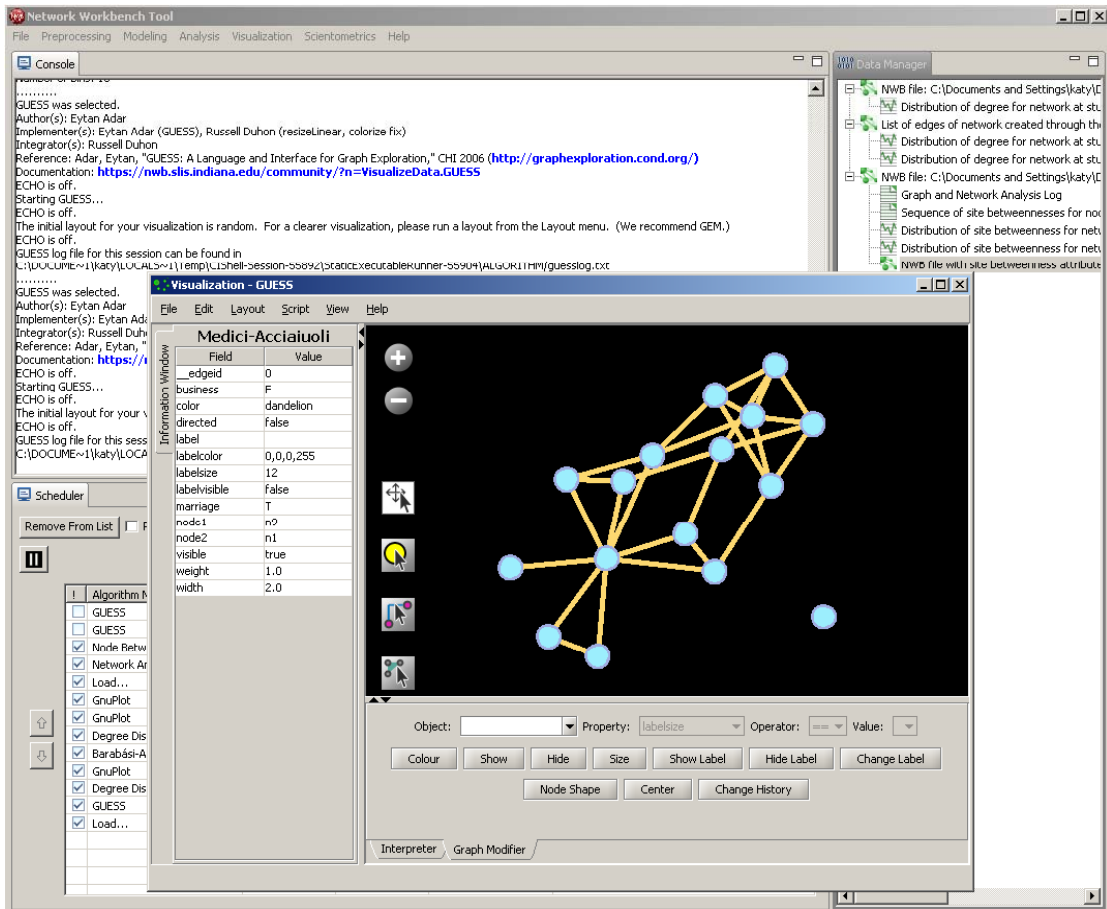
Edges: 27
No self loops were discovered.
No parallel edges were discovered.
Edge attributes:
  Nonnumeric attributes:
    marriag... T Example value
    busines... F

Did not detect any numeric attributes
This network does not seem to be a valued network.

Average degree: 3.375
This graph is not weakly connected.
There are 2 weakly connected components. (1 isolates)
The largest connected component consists of 15 nodes.
Did not calculate strong connectedness because this graph was not directed.


Density (disregarding weights): 0.225
```

- o Optional: Run 'Analysis -> Unweighted & Undirected -> Node Betweenness Centrality' with default parameters.
- o Select network and run 'Visualization > GUESS' to open GUESS with file loaded.
- o Apply 'Layout -> GEM'.

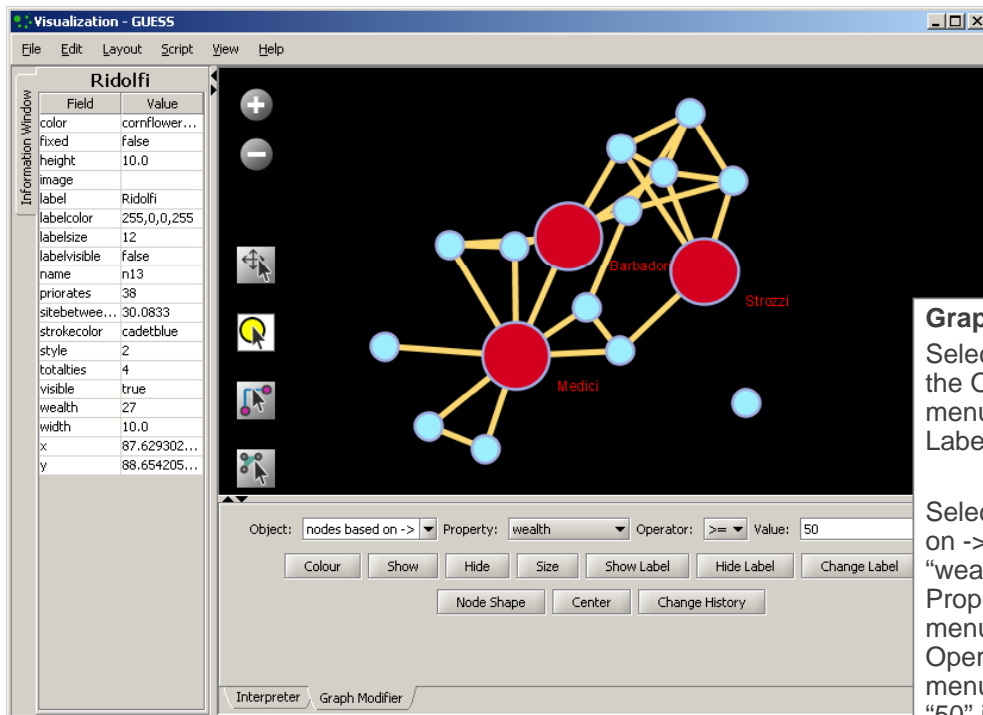


Pan:
"grab" the background by holding left-click and moving your mouse.

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

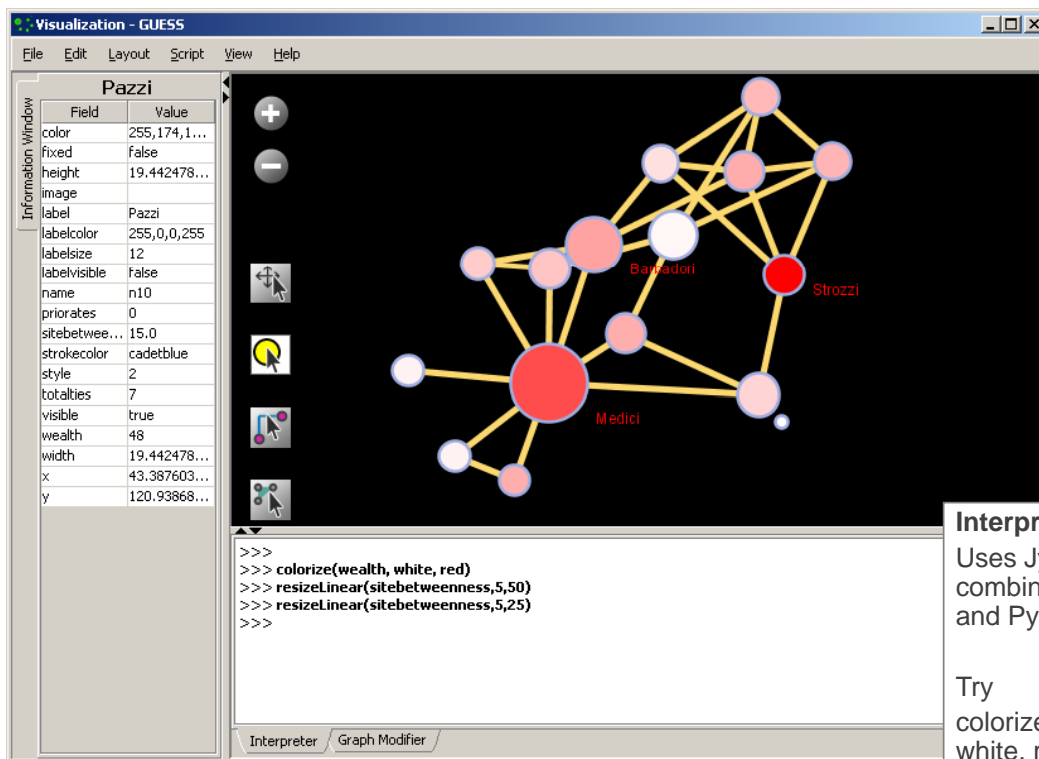
Select  to select/rotate single nodes. Hold down 'Shift' to select multiple.

Right click to modify Color, etc.



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

Select "nodes based on ->", then select "wealth" from the Property drop-down menu, ">=" from the Operator drop-down menu, and finally type "50" into the Value box. Then a color/size/shape code.



Interpreter:
Uses Jython a combination of Java and Python.

Try
colorize(wealth,
white, red)

resizeLinear(sitebet
weenness, 5, 25)



Workflow Design Primer

Modularity at data preprocessing/analysis/modeling level.

Modularity at visualization level:

- 'Data Layers' are used in GIS systems to support the visual layering and coordination of different datasets, e.g., water pipes, streets, electricity lines, etc.
- 'Design Layers' supported by graphic design software such as Photoshop or Dreamweaver enable the separate design and modular composition of design elements.
- 'Visualization Layers' define distinct parts with very specific functionality that collectively define a visualization.

BREAK



Exemplary Analyses and Visualizations

Individual Level

- A. Loading ISI files of major network science researchers, extracting, analyzing and visualizing paper-citation networks and co-author networks.
- B. Loading NSF datasets with currently active NSF funding for 3 researchers at Indiana U

Institution Level

- C. Indiana U, Cornell U, and Michigan U, extracting, and comparing Co-PI networks.

Scientific Field Level

- D. Extracting co-author networks, patent-citation networks, and detecting bursts in SDB data.



Exemplary Analyses and Visualizations

Individual Level

- A. Loading ISI files of major network science researchers, extracting, analyzing and visualizing paper-citation networks and co-author networks.
- B. Loading NSF datasets with currently active NSF funding for 3 researchers at Indiana U

Institution Level

- C. Indiana U, Cornell U, and Michigan U, extracting, and comparing Co-PI networks.

Scientific Field Level

- D. Extracting co-author networks, patent-citation networks, and detecting bursts in SDB data.



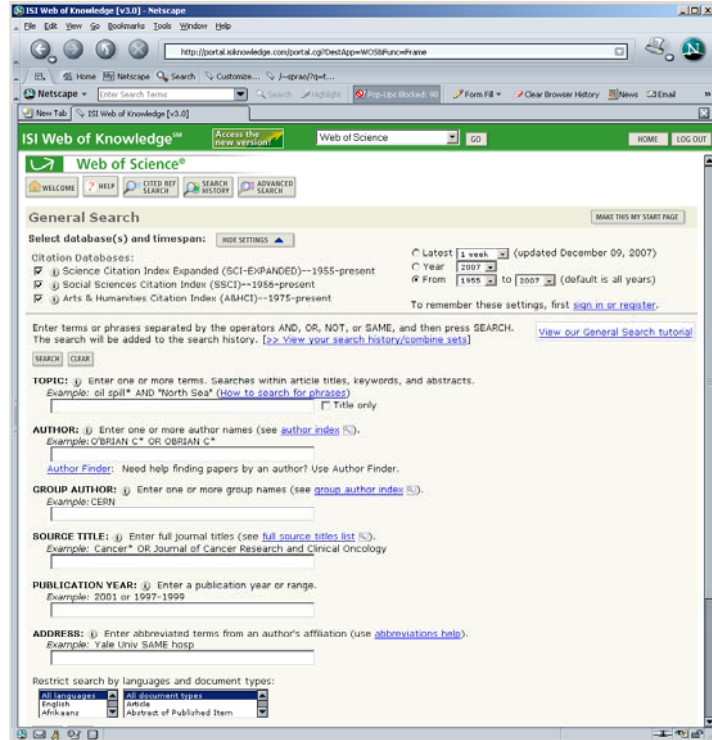
Data Acquisition from Web of Science

Download all papers by

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

from

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



Comparison of Counts

No books and other non-WoS publications are covered.

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

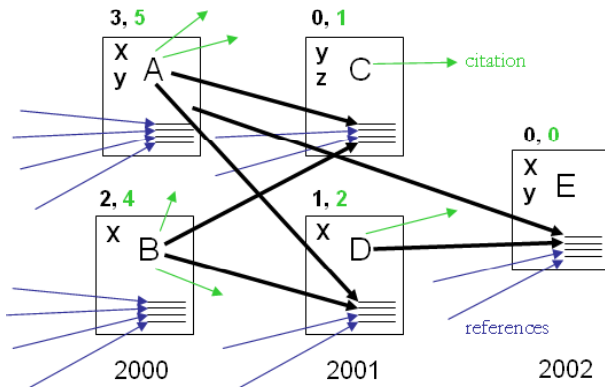


Network Extraction

Sample paper network (left) and four different network types derived from it (right)

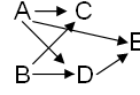
From ISI files, about 30 different networks can be extracted.

Papers A-E written by authors x, y, z over 3 years.
Each paper happens to have 4 references.



Paper-Paper Citation Network

Papers are connected via direct citation links. Arrows represent information flow from older papers to younger papers.



Author-Author (Co-Author) Network

x and y co-author papers A and E together
y and z co-author papers A and E



Document Co-Citation (DCA) Network

A and B are co-cited by C and D
A and D are co-cited by E



Reference Co-Occurrence (Bibliographic Coupling) Network

C and D are bibliographically coupled as they both cite/reference A and B.



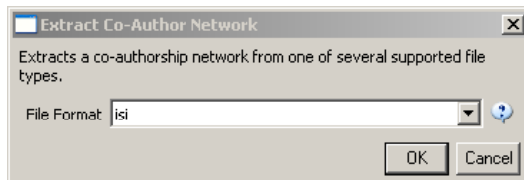
Local citation counts (within this dataset) are given in **black** and global citation counts (ISI times cited) are given in **green** above each paper.



Extract Co-Author Network

Load **yournwbdirectory*/sampledata/scientometrics/isi/FourNetSciResearchers.isi*
using 'File > Load and Clean ISI File'.

To extract the co-author network, select the '361 Unique ISI Records' table and run
'Scientometrics > Extract Co-Author Network' using isi file format:

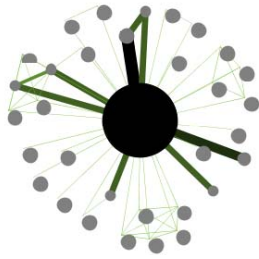


The result is an undirected network of co-authors in the Data Manager. It has 247 nodes and 891 edges.

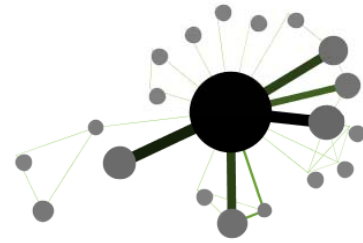
To view the complete network, select the network and run 'Visualization > GUESS > GEM'. Run Script > Run Script... . And select Script folder > GUESS > co-author-nw.py.



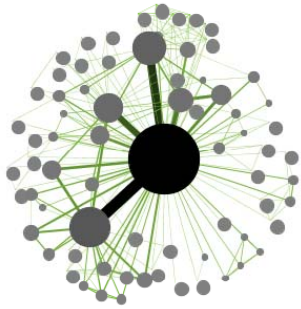
Comparison of Co-Author Networks



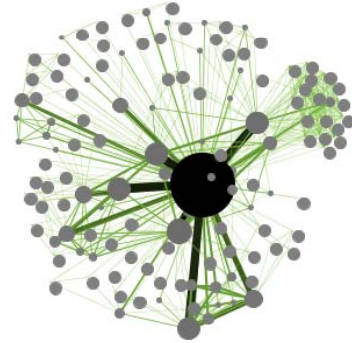
Eugene Garfield



Stanley Wasserman



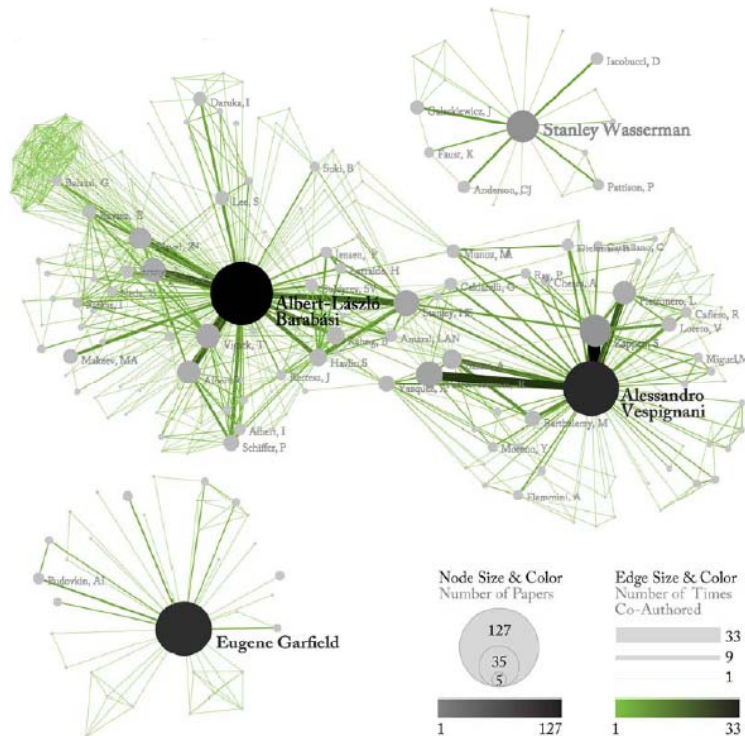
Alessandro Vespignani



Albert-László Barabási



Joint Co-Author Network of all Four NetsSci Researchers





Paper-Citation Network Layout

Load **yournwbdirectory*/sampledata/scientometrics/isi/FourNetSciResearchers.isi* using *'File > Load and Clean ISI File'*.

To extract the paper-citation network, select the *'361 Unique ISI Records'* table and run *'Scientometrics > Extract Directed Network'* using the parameters:

Extract Directed Network

Given a table, this algorithm creates a directed network by placing a directed edge between the values in a given column to the values of a different column.

Source Column: Cited References

Target Column: Cite Me As

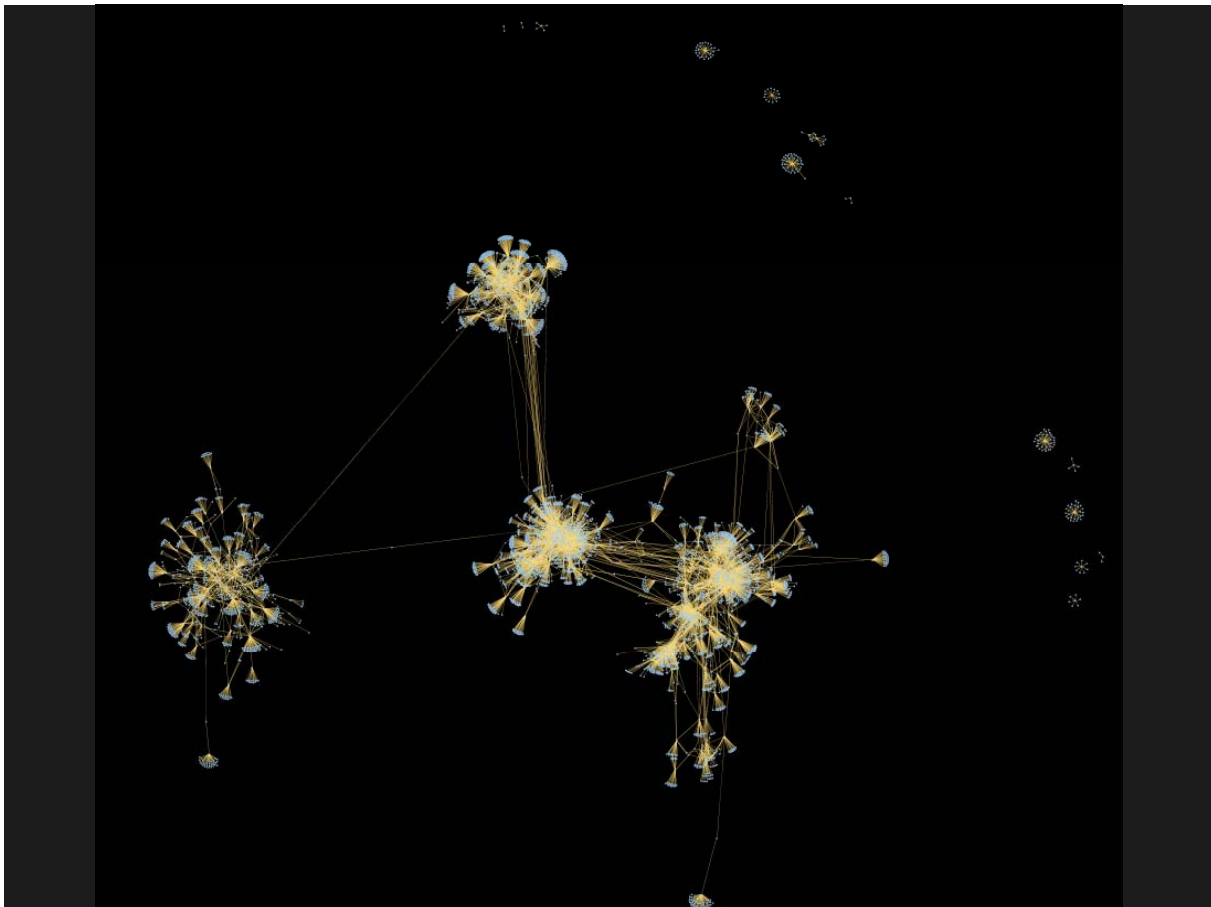
Text Delimiter: |

Aggregate Function File: C:\Documents and Settings\katy\Desktop\nwb\sampledata\scientometrics\properties\isiPaperCitation.properties

OK Cancel

The result is a directed network of paper citations in the Data Manager. It has 5,335 nodes and 9,595 edges.

To view the complete network, select the network and run *'Visualization > GUESS'*. Run *'Script > Run Script ...'* and select *'yournwbdirectory*/script/GUESS/paper-citation-nw.py'*.





Exemplary Analyses and Visualizations

Individual Level

- A. Loading ISI files of major network science researchers, extracting, analyzing and visualizing paper-citation networks and co-author networks.
- B. Loading NSF datasets with currently active NSF funding for 3 researchers at Indiana U

Institution Level

- C. Indiana U, Cornell U, and Michigan U, extracting, and comparing Co-PI networks.

Scientific Field Level

- D. Extracting co-author networks, patent-citation networks, and detecting bursts in SDB data.

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

The image displays two screenshots of the NSF Award Search website. The left screenshot shows the search interface with red circles highlighting the 'Search Award For' field, the 'Include CO-PI' checkbox, and the 'Search' button. The right screenshot shows the search results table with a red circle highlighting the 'Export options' section, specifically the 'CSV' option.

Award Number	Title	Program	Category	Start Date	Link
9100833	Research in Computer Science and Computational Physics	EIA	PROGRAMS	06/01/1991	Fox...
9014995	Applications of Parallel Supercomputing to Astrophysical N-body Calculations	OCI	ADVANCED COMP RESEARCH PROGRAM	08/01/1990	Princ...
8921679	CISE Research Instrumentation for a Program in Physical Computation & Complex Systems	EIA	CISE RESEARCH RESOURCES	04/01/1990	Fox...
8900464	REU Site: To Continue an REU Site in Computer and Information Science and Engineering at Caltech	OCI	CROSS-DIRECTORATE PROGRAMS	05/01/1989	Fox...
8804528	Proposal to Continue an REU Site in Computer And Information Science And Engineering	CCF	CROSS-DIRECTORATE PROGRAMS	06/01/1988	Fox...
8719502	A Pilot Project in Performance Science Select Architect		CROSS-DIRECTORATE		
8700064	Conc and th Applie Neura		NEUROENGINEERING		
8519481	Enhanced Supercomputer Access Facility at the California Institute of Technology	OCI	LOCAL ACCESS	09/15/1985	Fox...
7819718	Travel to Attend: 19th International Conference on High Energy Physics, Tokyo, Japan, August 23-31, 1978	PHY	INTERNATIONAL INFO & ANALYSIS	08/23/1978	Fox...



NSF Awards Search Results

Name	# Awards	First A. Starts	Total Amount to Date
Geoffrey Fox	27	Aug 1978	12,196,260
Michael McRobbie	8	July 1997	19,611,178
Beth Plale	10	Aug 2005	7,224,522

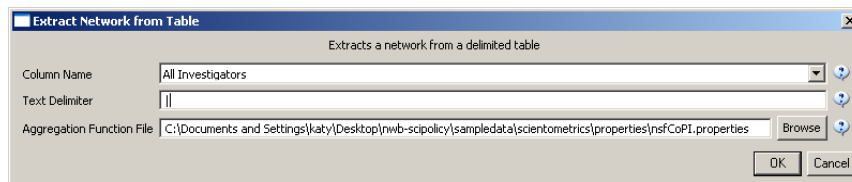
Disclaimer:

Only NSF funding, no funding in which they were senior personnel, only as good as NSF's internal record keeping and unique person ID. If there are 'collaborative' awards then only their portion of the project (award) will be included.

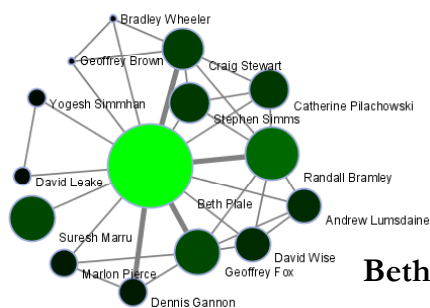
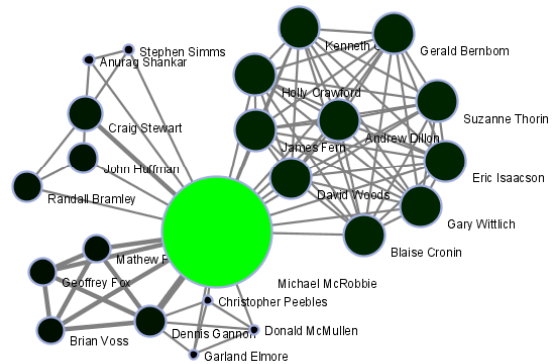
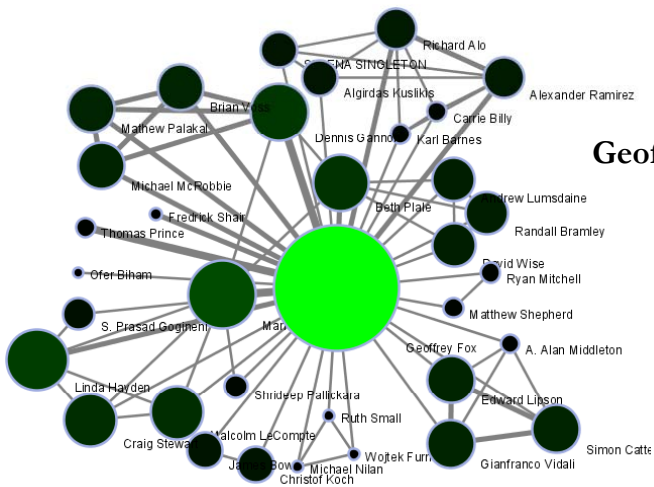
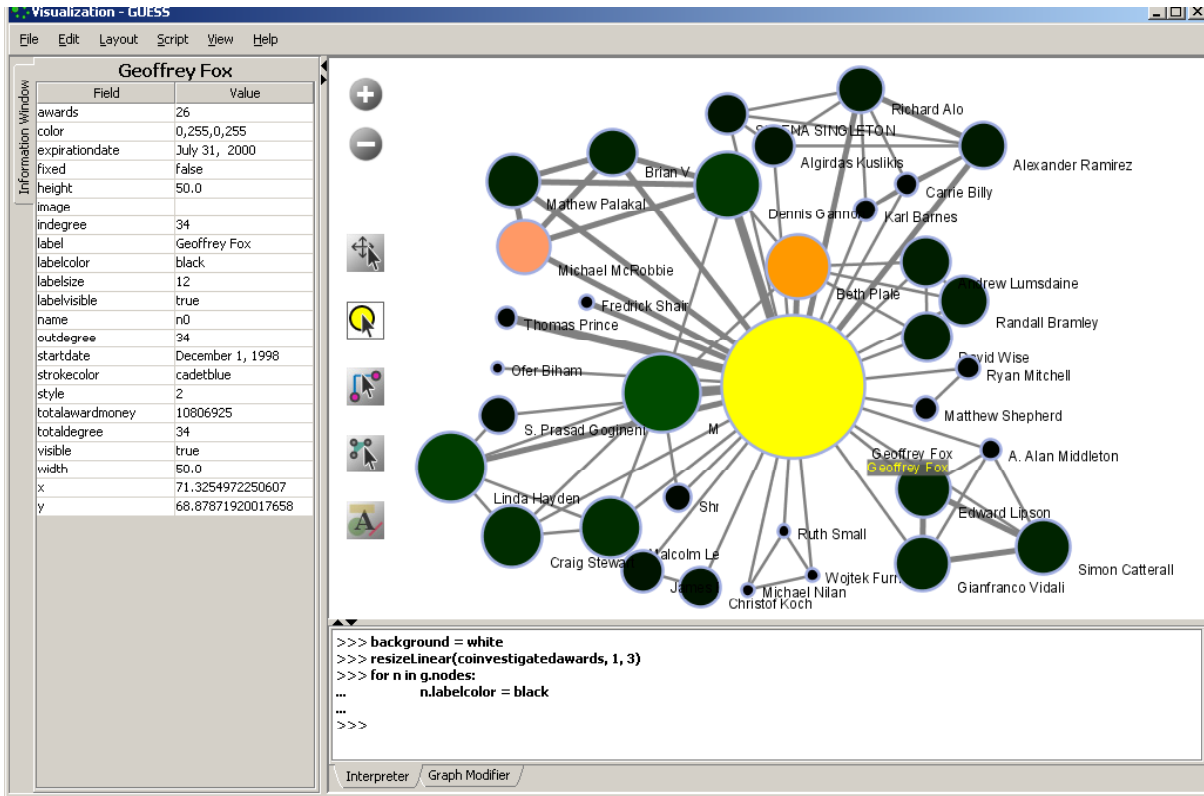


Using NWB to Extract Co-PI Networks

- Load into NWB, open file to count records, compute total award amount.
- Run '*Scientometrics > Extract Co-Occurrence Network*' using parameters:



- Select "*Extracted Network ..*" and run '*Analysis > Network Analysis Toolkit (NAT)*'
- Remove unconnected nodes via '*Preprocessing > Delete Isolates*'.
- '*Visualization > GUESS*', layout with GEM
- Run '*co-PI-nw.py*' GUESS script to color/size code.



Geoffrey Fox

Last Expiration date



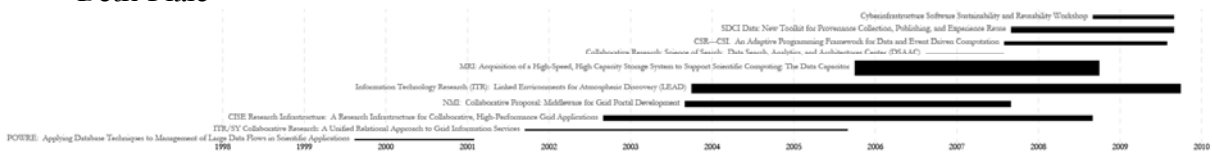
July 10

Michael McRobbie



Feb 10

Beth Plale



Sept 09

Horizontal Line Graph

Takes NSF grant data and generates PostScript for a horizontal line graph.

Label: TITLE

Start Date: START_DATE

End Date: EXPIRATION_DATE

Size By: AWARDED_AMOUNT_TO_DATE



Exemplary Analyses and Visualizations

Individual Level

- A. Loading ISI files of major network science researchers, extracting, analyzing and visualizing paper-citation networks and co-author networks.
- B. Loading NSF datasets with currently active NSF funding for 3 researchers at Indiana U

Institution Level

- C. Indiana U, Cornell U, Michigan U, and Stanford U extracting, and comparing Co-PI networks.

Scientific Field Level

- D. Extracting co-author networks, patent-citation networks, and detecting bursts in SDB data.

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

The screenshot shows the NSF Award Search interface. On the left, the search form is visible with the following details:

- Search Award For: [Empty]
- Restrict to Title Only:
- Awardee Information:
 - Principal Investigator:
 - First Name: [Empty]
 - Last Name: [Empty]
 - Include CO-PI:
 - Organization: University of Michigan Ann Arbor (circled in red)
 - State: [Dropdown]
 - ZIP Code: [Empty]
 - Country: [Dropdown]

On the right, the search results are displayed in a table:

Award Number	Title	Agency	Program	Start Date	PI Name
0820609	PhysJobs Helpings				
0817369	Teaching Mathematical Knowledge for Teaching (K-12): Adapting Local Materials for Use in Diverse Institutions and Settings	DUE	CCLI-Phase 2 (Expansion), S-SYSTEM: SCHLR SCI TECH ENG/MATH	01/01/2009	Bass, Hyma
0822892	Protest Psychosis: Race, Science, and the Stigma of Schizophrenia	SES	SCIENCE, TECH & SOCIETY	01/01/2009	Metal, Jonat
0825795	Collaborative Research: Tissue Culture Mechanics - Investigation of the Effective and Minimally Invasive Biopsy	CMMI	MANUFACTURING & CONST MACH EQP	01/01/2009	Shih, Albert
0855698	IMPLEMENTING THE "SME" WORKSHOP RECOMMENDATIONS	CMMI	CONTROL SYSTEMS	01/01/2009	Ulsov, A. G
0825789	Short-Term Joint Maintenance and Production Decision Support Tool of Manufacturing Systems	CMMI	MANFG ENTERPRISE SYSTEMS	01/01/2009	Ni, Jun
0825789	Support for the 6th U.S.		COMBUSTION, FIRE, &		

A text box is overlaid on the table with the text: "Save in CSV format as *institution*.nsf"

Active NSF Awards on 11/07/2008:

- Indiana University 257
(there is also Indiana University at South Bend Indiana University Foundation, Indiana University Northwest, Indiana University-Purdue University at Fort Wayne, Indiana University-Purdue University at Indianapolis, Indiana University-Purdue University School of Medicine)
- Cornell University 501
(there is also Cornell University – State, Joan and Sanford I. Weill Medical College of Cornell University)
- University of Michigan Ann Arbor 619
(there is also University of Michigan Central Office, University of Michigan Dearborn, University of Michigan Flint, University of Michigan Medical School)

Active NSF Awards on 09/10/2009:

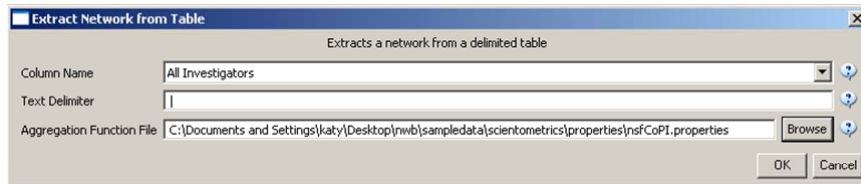
- Stanford University 429

Save files as csv but rename into .nsf.

Or simply use the files saved in **yournwbdirectory*/sampledata/scientometrics/nsf/*.

Extracting Co-PI Networks

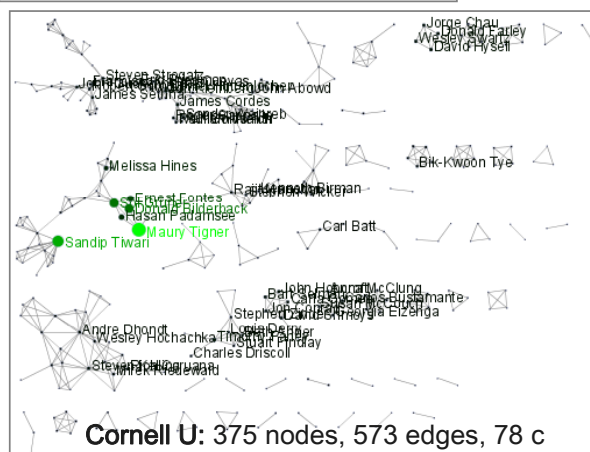
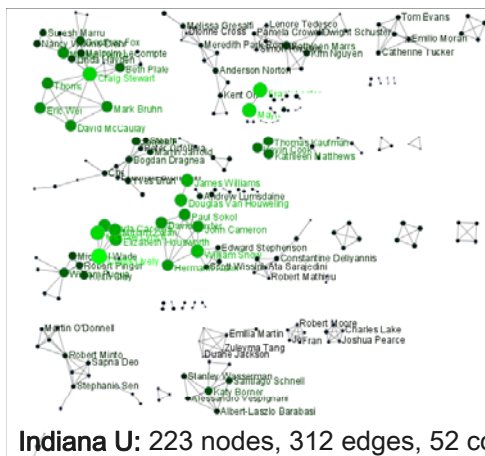
Load NSF data, selecting the loaded dataset in the Data Manager window, run *'Scientometrics > Extract Co-Occurrence Network'* using parameters:



Two derived files will appear in the Data Manager window: the co-PI network and a merge table. In the network, nodes represent investigators and edges denote their co-PI relationships. The merge table can be used to further clean PI names.

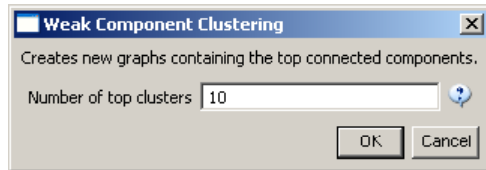
Running the *'Analysis > Network Analysis Toolkit (NAT)'* reveals that the number of nodes and edges but also of isolate nodes that can be removed running *'Preprocessing > Delete Isolates'*.

Select *'Visualization > GUESS'* to visualize. Run 'co-PI-nw.py' script.



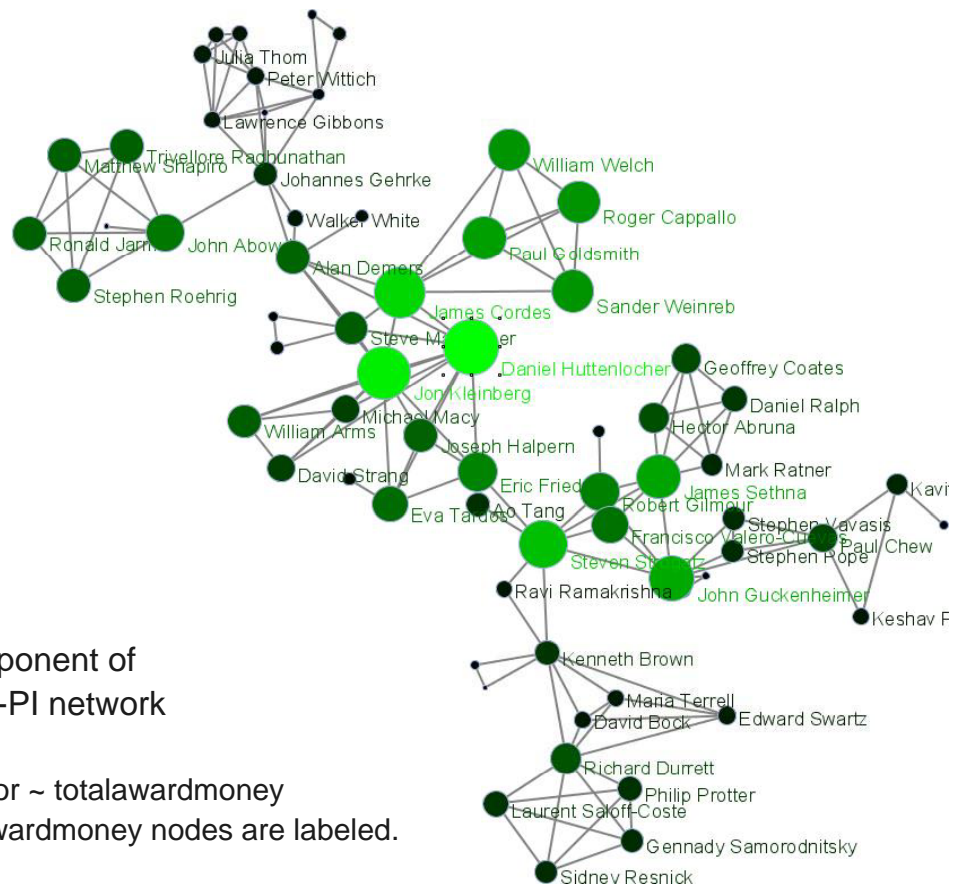
Extract Giant Component

Select network after removing isolates and run *'Analysis > Unweighted and Undirected > Weak Component Clustering'* with parameter



Indiana's largest component has 19 nodes, Cornell's has 67 nodes, Michigan's has 55 nodes.

Visualize Cornell network in GUESS using same .py script and save via *'File > Export Image'* as jpg.



Largest component of
Cornell U co-PI network

Node size/color ~ totalawardmoney
Top-50 totalawardmoney nodes are labeled.

Top-10 Investigators by Total Award Money

for i in range(0, 10):

```
print str(nodesbytotalawardmoney[i].label) + ": " +
str(nodesbytotalawardmoney[i].totalawardmoney)
```

Indiana University

Curtis Lively:	7,436,828
Frank Lester:	6,402,330
Maynard Thompson:	6,402,330
Michael Lynch:	6,361,796
Craig Stewart:	6,216,352
William Snow:	5,434,796
Douglas V. Houweling:	5,068,122
James Williams:	5,068,122
Miriam Zolan:	5,000,627
Carla Caceres:	5,000,627

Cornell University

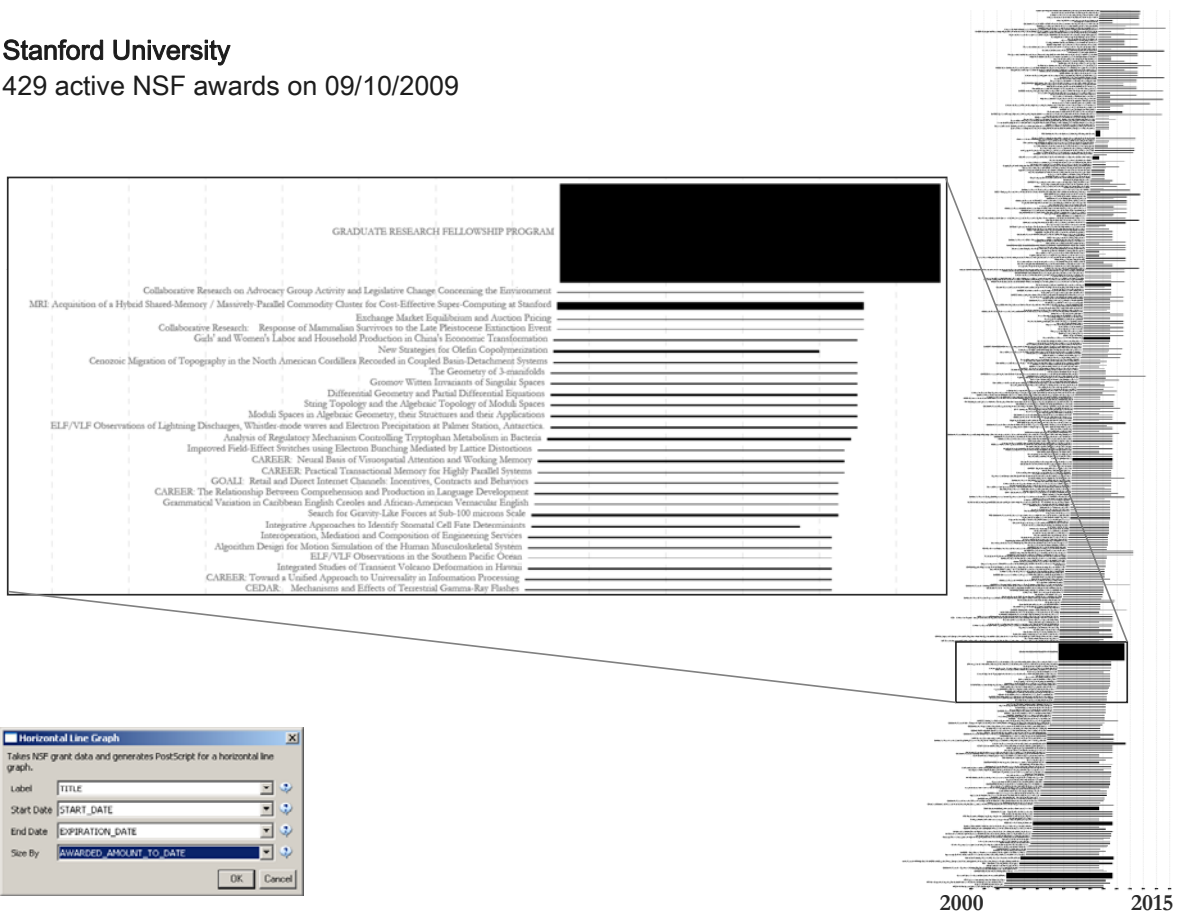
Maury Tigner:	107,216,976
Sandip Tiwari:	72,094,578
Sol Gruner:	48,469,991
Donald Bilderback:	47,360,053
Ernest Fontes:	29,380,053
Hasan Padamsee:	18,292,000
Melissa Hines:	13,099,545
Daniel Huttenlocher:	7,614,326
Timothy Fahey:	7,223,112
Jon Kleinberg:	7,165,507

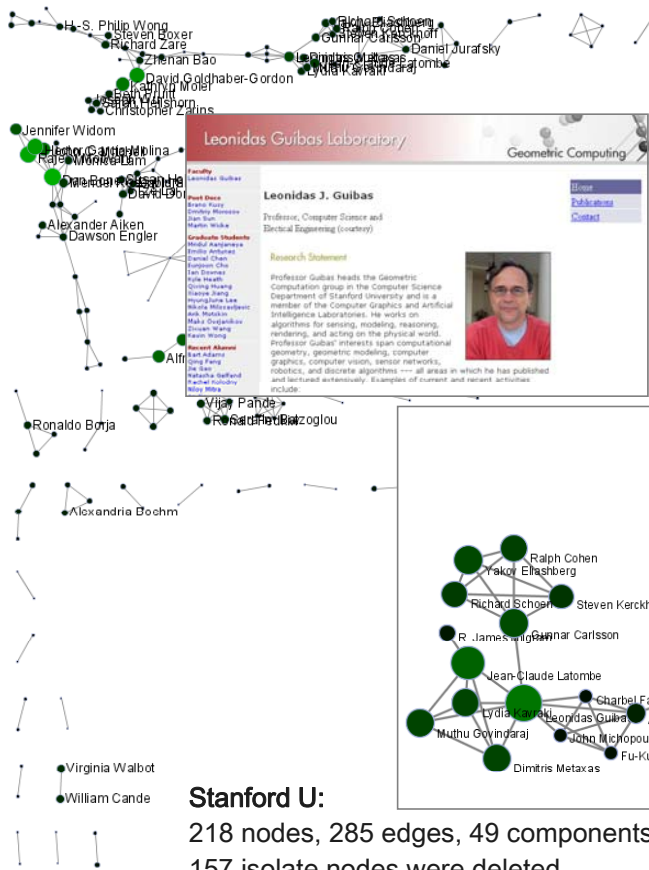
Michigan University

Khalil Najafi:	32,541,158
Kensall Wise:	32,164,404
Jacquelynne Eccles:	25,890,711
Georg Raithel:	23,832,421
Roseanne Sension:	23,812,921
Theodore Norris:	23,350,921
Paul Berman:	23,350,921
Roberto Merlin:	23,350,921
Robert Schoeni:	21,991,140
Wei-Jun Jean Yeung:	21,991,140

Stanford University

429 active NSF awards on 09/10/2009





Stanford U:
 218 nodes, 285 edges, 49 components
 157 isolate nodes were deleted


Stanford University Home | Contacts | Search | School of H&S | Stanford University

School of Humanities and Sciences
Department of Physics

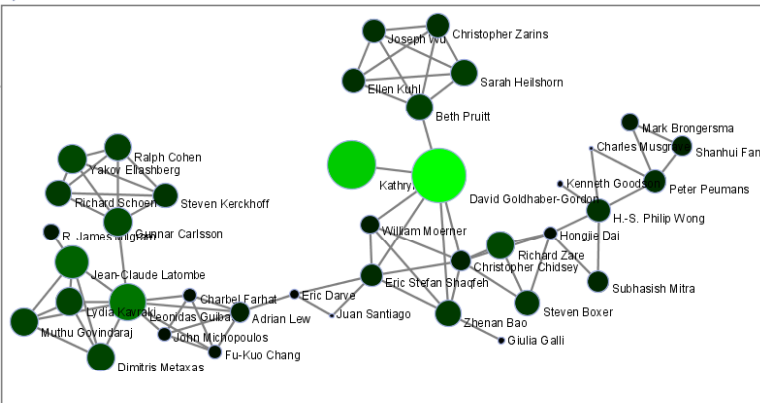
People

David Goldhaber-Gordon
 Associate Professor of Physics

Geballs Laboratory for Advanced Materials
 McCullough Bldg rm 346
 475 Lomita Mall
 Stanford, CA 94305-4045
 tel 650-724-3709
 e-mail goldhaber_gordon@stanford.edu
[Personal page](#)
[Group page](#)



Largest component
 39 nodes



Visualization - GUESS

File Edit Layout Script View Help

David Goldhaber-Gordon

Field	Value
awards	5
color	0,255,0,255
expirationdate	May 31, 2010
fixed	false
height	50.0
image	
indegree	6
label	David Goldhaber-Go...
labelcolor	black
labelsize	12
labelvisible	true
name	n20
originallabel	David Goldhaber-Go...
outdegree	6
startdate	June 1, 2009
strokecolor	cadetblue
style	2
totalawardmoney	9792029
totaldegree	6
visible	true
width	50.0
x	349.54443173369805
y	106.91807610336735

Object: Property: awards Operator: == Value:

Colour Show Hide Size Show Label Hide Label Change Label

Format Node Labels Format Edge Labels

Node Shape Center Change History

Resize Linear Colorize

Interpreter Graph Modifier

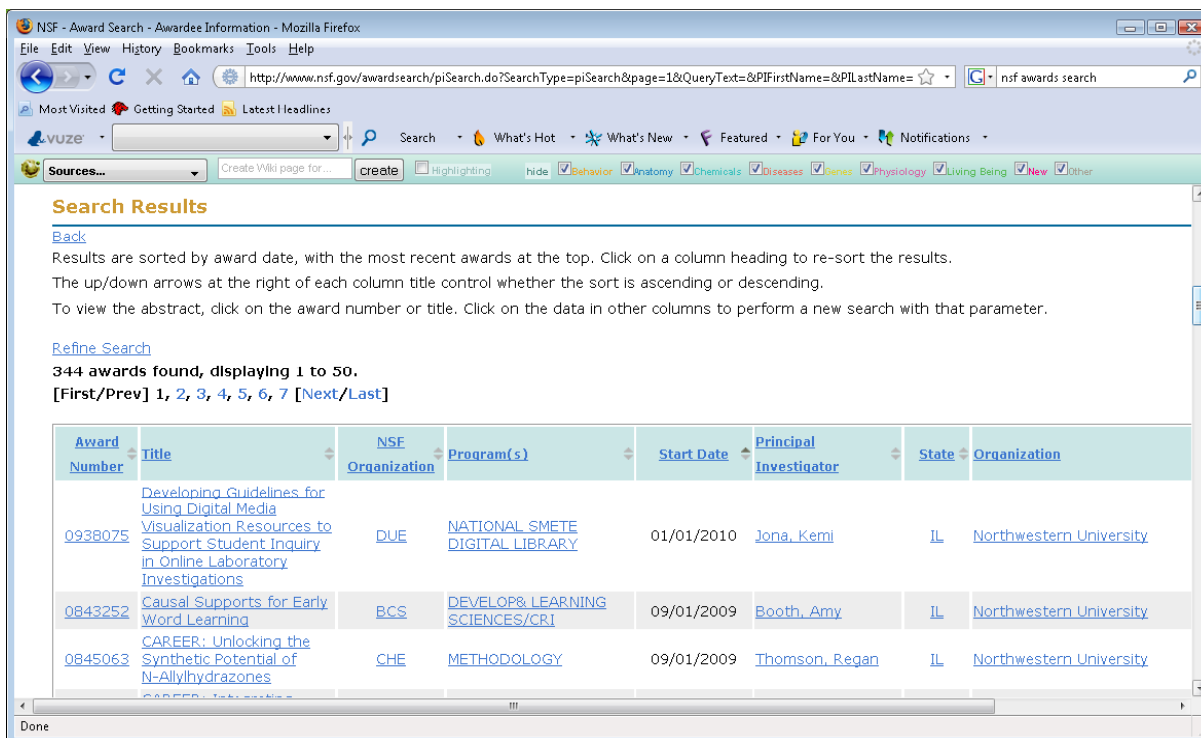
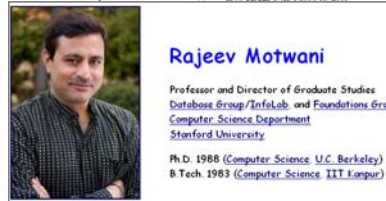
Top-10 Investigators by Total Award Money

for i in range(0, 10):

```
print str(nodesbytotalawardmoney[i].label) + ", "
print str(nodesbytotalawardmoney[i].totalawardmoney)
```

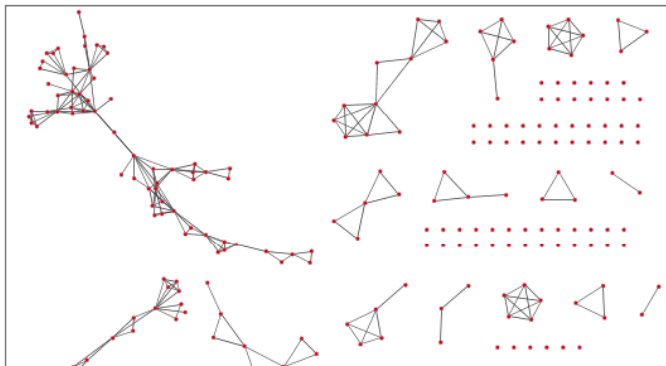
Stanford University

Dan Boneh:	11,837,800
Rajeev Motwani:	11,232,154
Hector Garcia-Molina:	10,577,906
David Goldhaber-Gordon:	9,792,029
Kathryn Moler:	7,870,029
John C. Mitchell:	7,290,668
Alfred Spormann:	6,803,000
Gordon Brown:	6,158,000
Jennifer Widom:	5,661,311



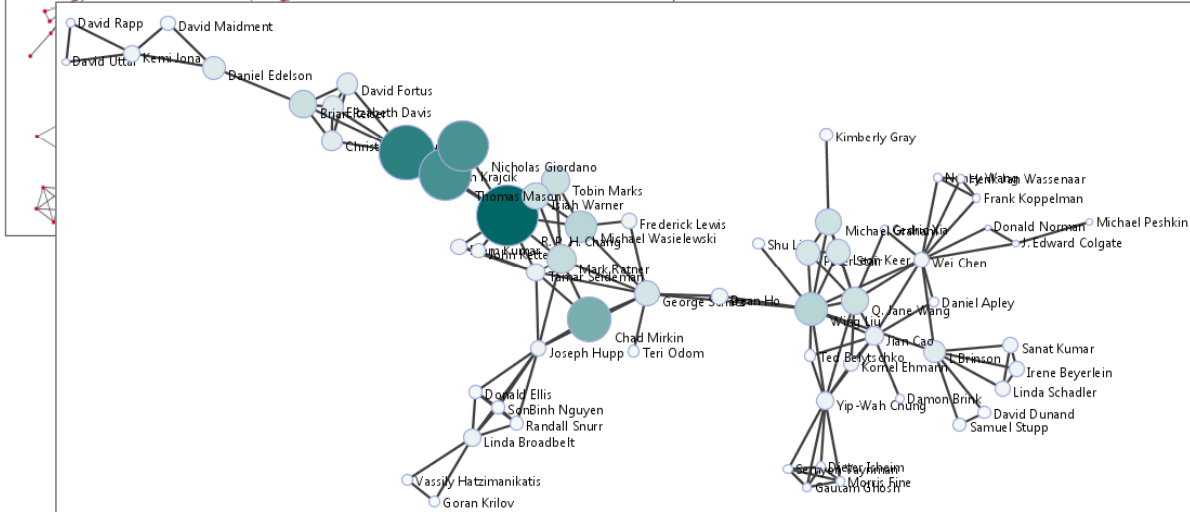
Search for all active NSF awards by Northwestern University on 9/2/2009 via

<http://www.nsf.gov/awardsearch>



Nodes: 323, Edges: 313, Average degree: 1.9, 149 weakly connected components. (107 isolates)

Giant component has 63 nodes, Color and size coding by total award money



3. Exemplary Analyses and Visualizations

Individual Level

- A. Loading ISI files of major network science researchers, extracting, analyzing and visualizing paper-citation networks and co-author networks.
- B. Loading NSF datasets with currently active NSF funding for 3 researchers at Indiana U

Institution Level

- C. Indiana U, Cornell U, and Michigan U, extracting, and comparing Co-PI networks.

Scientific Field Level

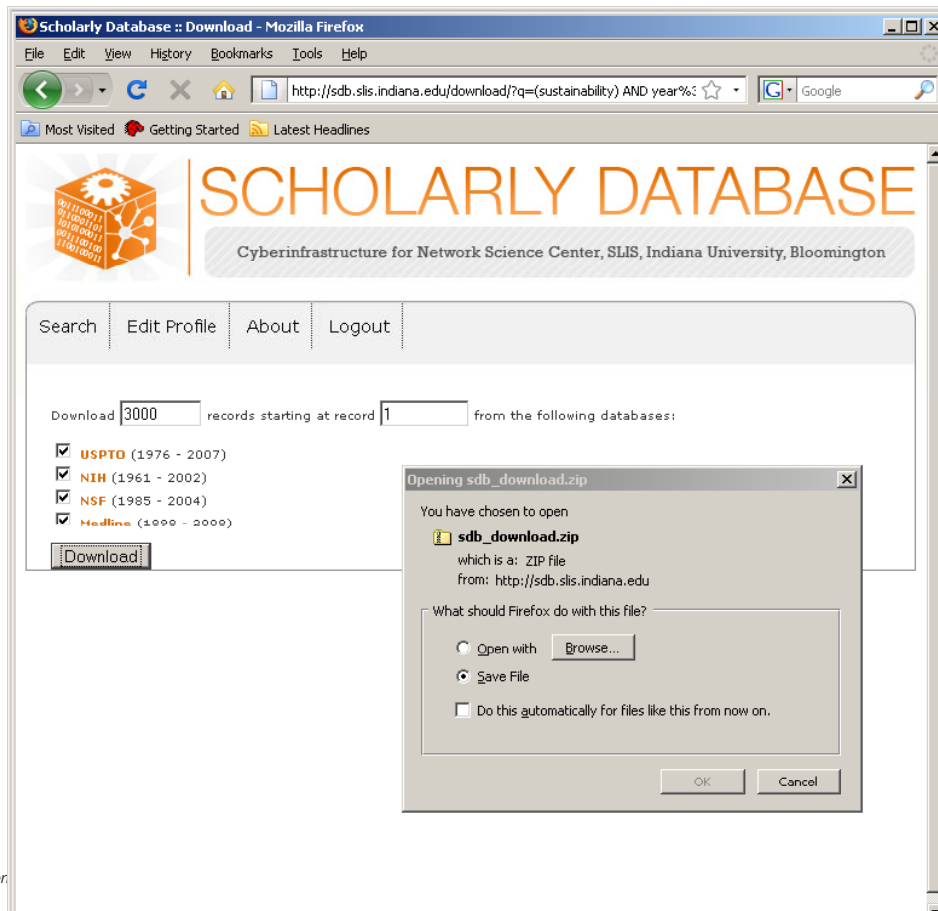
- D. Extracting co-author networks, patent-citation networks, and detecting bursts in SDB data.

Goto: <http://sdb.slis.indiana.edu>

Email: nwb@indiana.edu

Password: nwb

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



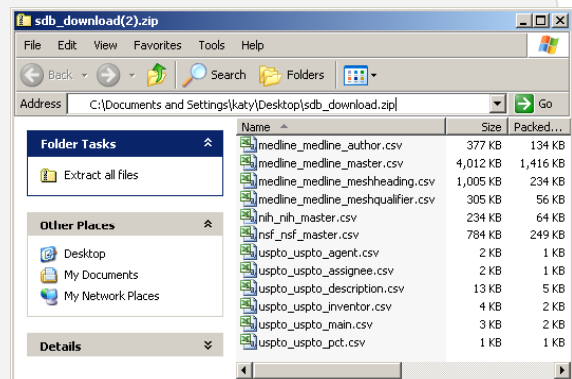
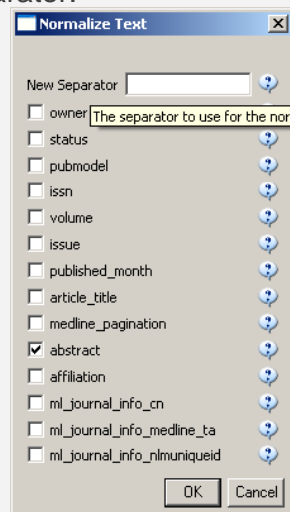
Network Workben

85

NetworkWorkbench Open and Preprocess SDB zip file

Load medline_medline_master.csv to NWB.

Run 'Preprocessing > Normalize Text' with a space as New Separator.



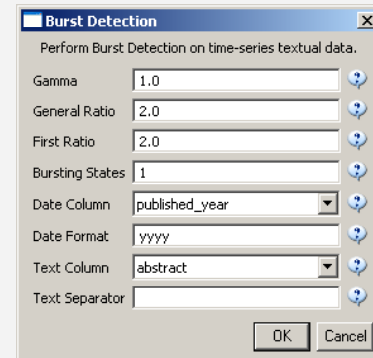
Network Workbench (<http://nwb.slis.indiana.edu>).

86

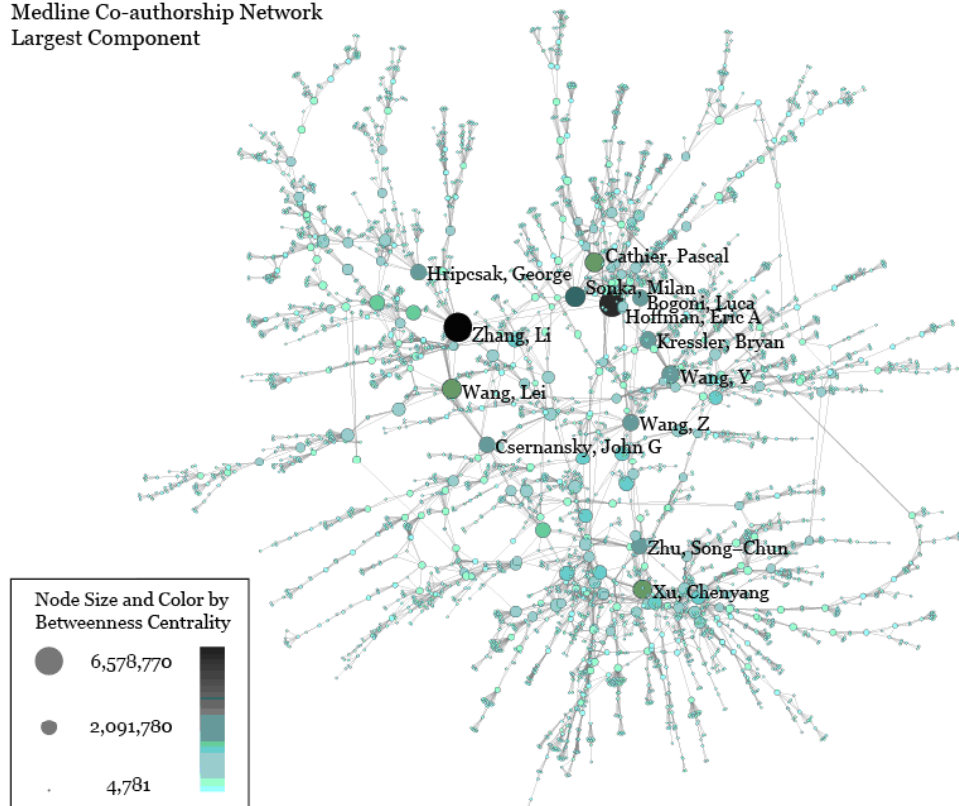
Run 'Analysis > Textual > Burst Detection' with parameters:
and space as a separator.

Sort result by burst weight

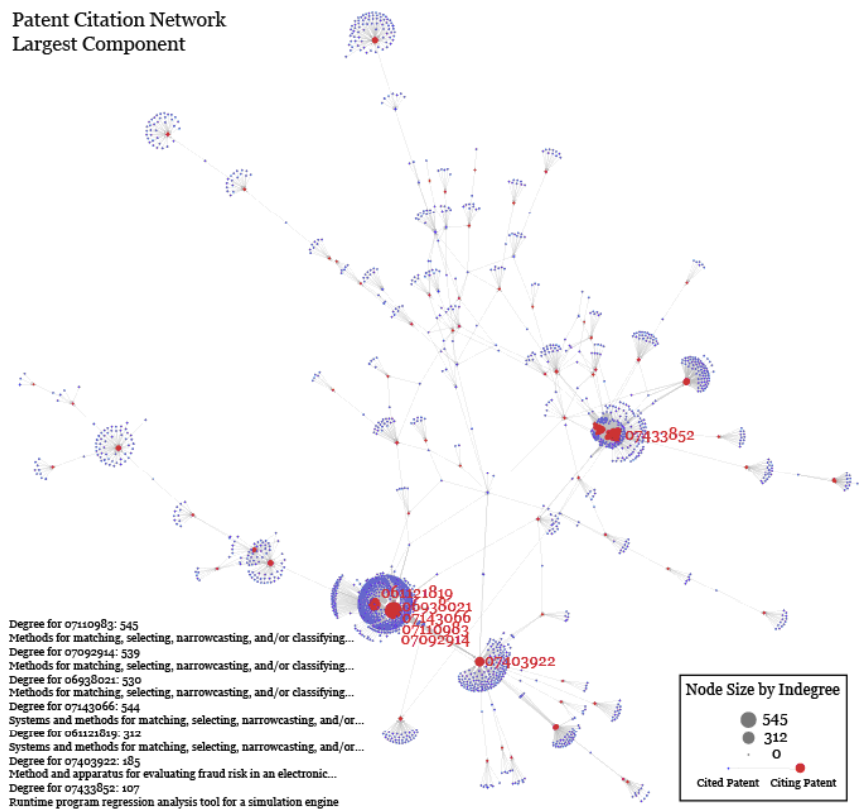
Word	Length	Weight	Strength	Start	End
care	1	Infinity	Infinity	1988	1988
water	1	29.8883	29.8883	2002	2002
countri	10	27.03612	27.03612	1990	1999
protect	1	26.88557	26.88557	2002	2002
farm	1	23.32114	23.32114	2005	2005
villag	2	23.273	40.65081	2008	
crop	2	22.33649	30.42535	2008	
educ	2	22.14556	26.98588	1995	1996
blood	5	22.12166	22.12166	1996	2000



Medline Co-authorship Network
Largest Component



**Patent Citation Network
Largest Component**



Degree for 07110983: 545
 Methods for matching, selecting, narrowcasting, and/or classifying...
 Degree for 07092914: 539
 Methods for matching, selecting, narrowcasting, and/or classifying...
 Degree for 06938021: 539
 Methods for matching, selecting, narrowcasting, and/or classifying...
 Degree for 07430666: 544
 Systems and methods for matching, selecting, narrowcasting, and/or...
 Degree for 061121819: 312
 Systems and methods for matching, selecting, narrowcasting, and/or...
 Degree for 07403922: 185
 Method and apparatus for evaluating fraud risk in an electronic...
 Degree for 07433852: 107
 Runtime program regression analysis tool for a simulation engine

Node Size by Indegree

● 545
 ● 312
 ● 0

— Cited Patent
 — Citing Patent

Top-10 burst terms from abstracts of the AI search results.

<i>Medline</i>				
Word	Length	Weight	Start	End
medical	17	299.7924	1983	1999
knowledge	5	293.9375	1991	1995
knowledge	6	215.2407	1997	2002
expert	13	171.0443	1985	1997
systems	15	170.3306	1985	1999
intelligence	21	123.9794	1981	2001
patient	21	123.9297	1982	2002
care	12	106.5522	1990	2001
registration	5	104.8139	2005	
knowledge-based	16	98.83778	1987	2002

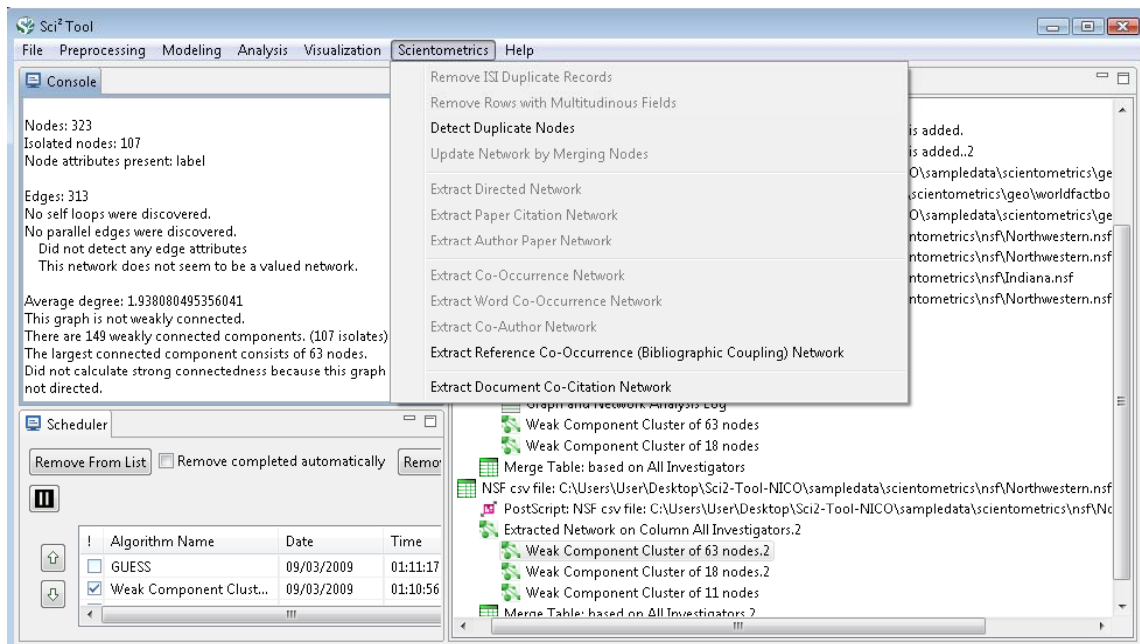
<i>NIH</i>				
Word	Length	Weight	Start	End
Phase	8	117.2205	1993	2000
commercial	9	87.57158	1995	
proposed	9	87.57158	1995	
mass	3	83.36952	1978	1980
protein	1	72.15788	1988	1988
networks	4	71.252	1993	1996
patterns	3	66.44826	1977	1979
being	8	66.29254	1971	1978
reasoning	2	65.68178	1984	1985
expert	4	60.49935	1987	1990

<i>NSF</i>				
Word	Length	Weight	Start	End
their	6	47.05097	1999	
gray	2	28.19808	2000	2001
learning	2	27.40728	1997	1998
human	5	25.4525	2000	
control	2	24.07877	1992	1993
knowledge	1	21.48756	1998	1998
students	1	21.07674	1997	1997
problems	2	20.77133	1998	1999
more	2	19.96109	2000	2001
use	1	19.38503	2001	2001

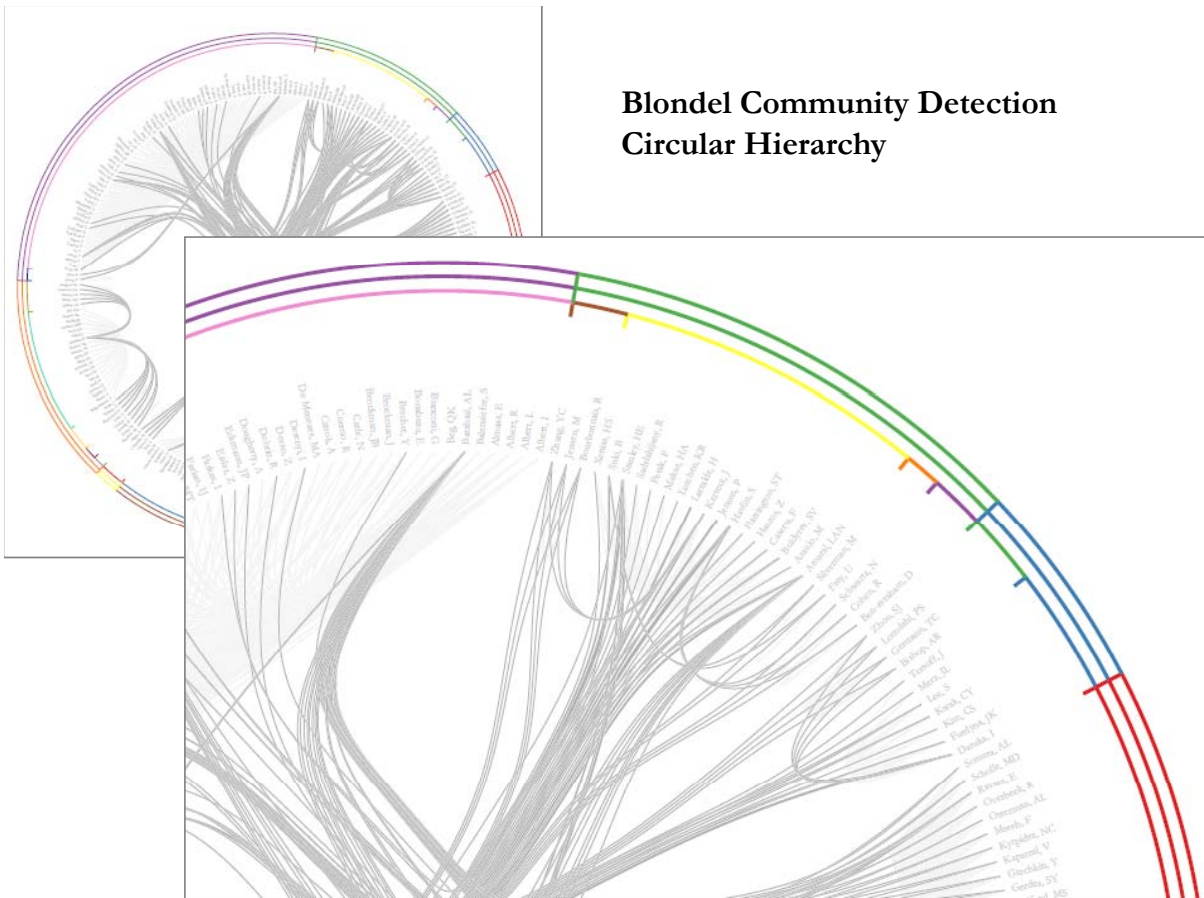
<i>USPTO</i>				
Word	Length	Weight	Start	End
human	3	19.03937321	2004	2006
video	3	15.32736425	1998	2000
disclosed	2	14.06694671	1999	2000
neural	3	13.30105906	2004	2006
"correct"	2	12.4336047	1999	2000
unit	2	12.35745838	2002	2003
material	1	12.08487035	2000	2000
feedback	1	12.07730195	2000	2000
rule	1	12.07730195	2000	2000
elevator	4	11.83351857	1991	1994

Bonus: Sci² Tool

Sci² Tool



Blondel Community Detection Circular Hierarchy



The screenshot shows the Sci2 Tool interface. The console window displays the following input parameters:

- Input Parameters:
- Longitude: Longitude
- Size Circles By: patents
- Color Circle Exteriors By: cited_times
- Color Circle Interiors By: None (no inner color)
- Exterior Color Scaling: linear
- Exterior Color Range: Green to Red
- Interior Color Range: Yellow to Blue
- Size Scaling: linear
- Projection: Mercator
- Map: Countries
- Author Name: Katy Borner
- Interior Color Scaling: linear
- Latitude: Latitude
- Reading shapefile: countries
- Printing PostScript.. Done.

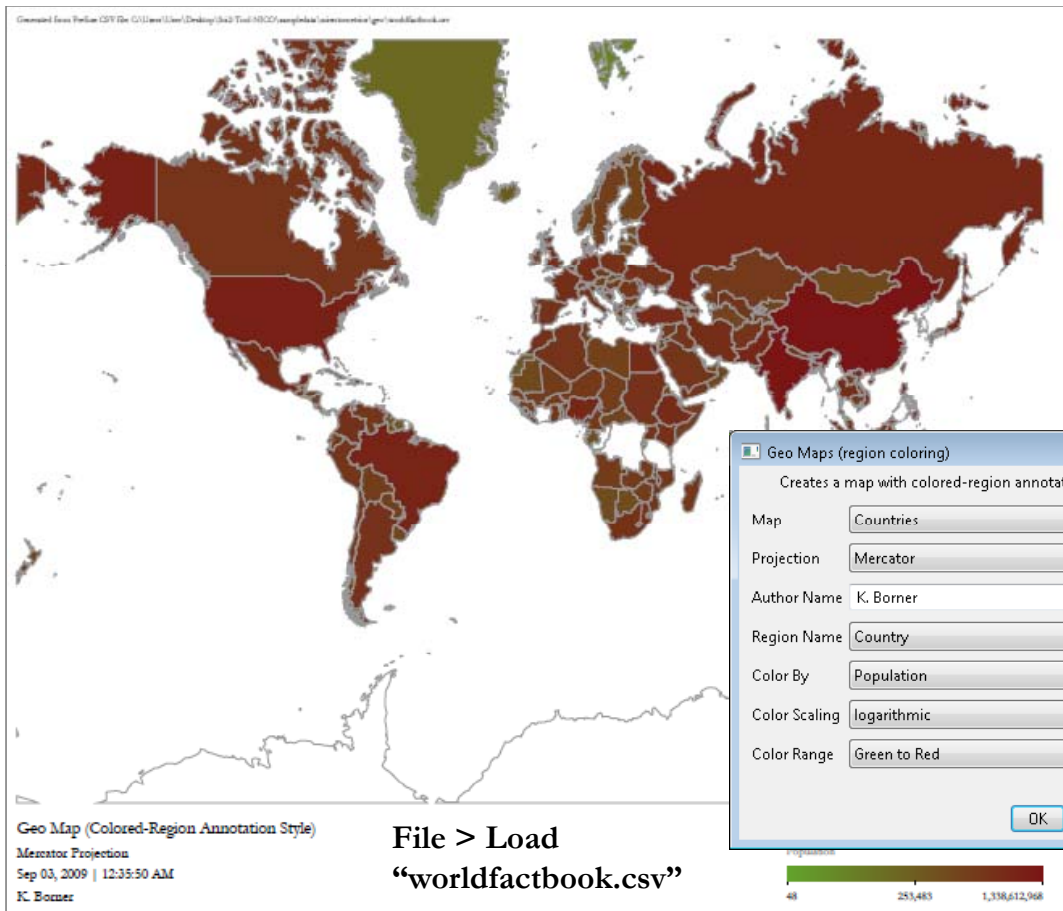
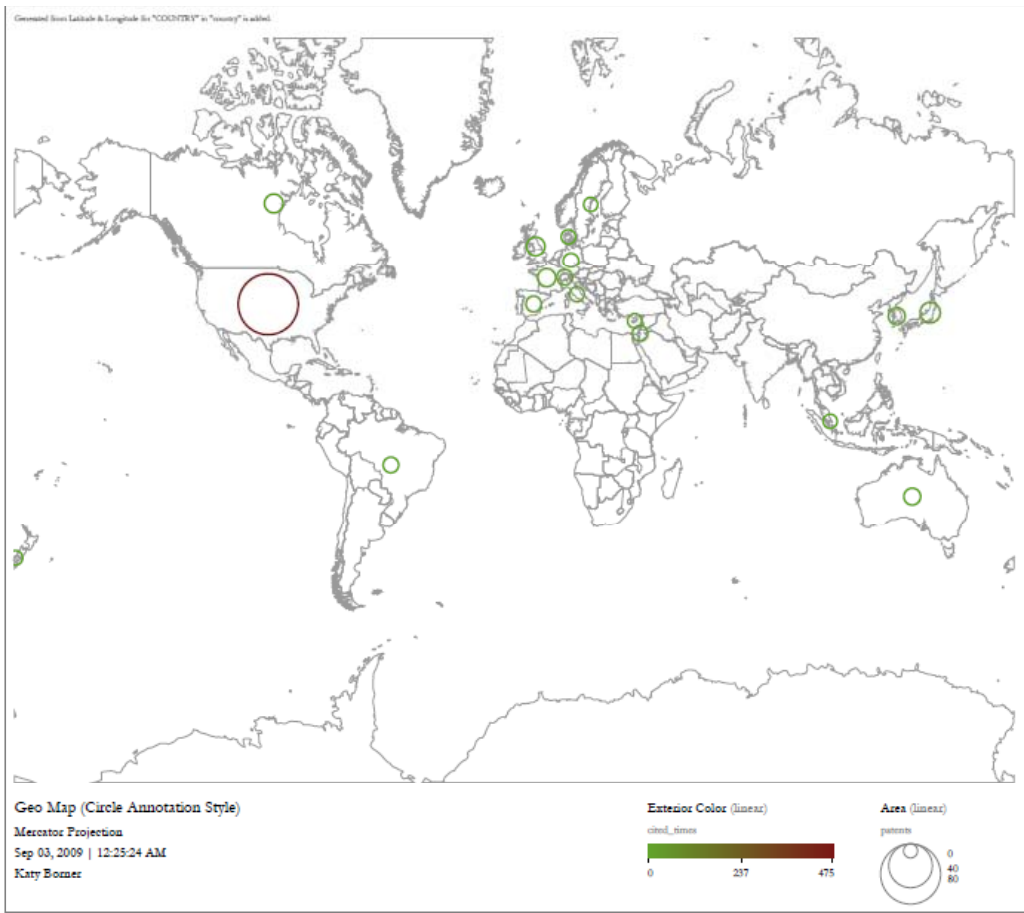
The Data Manager window shows the following data sources:

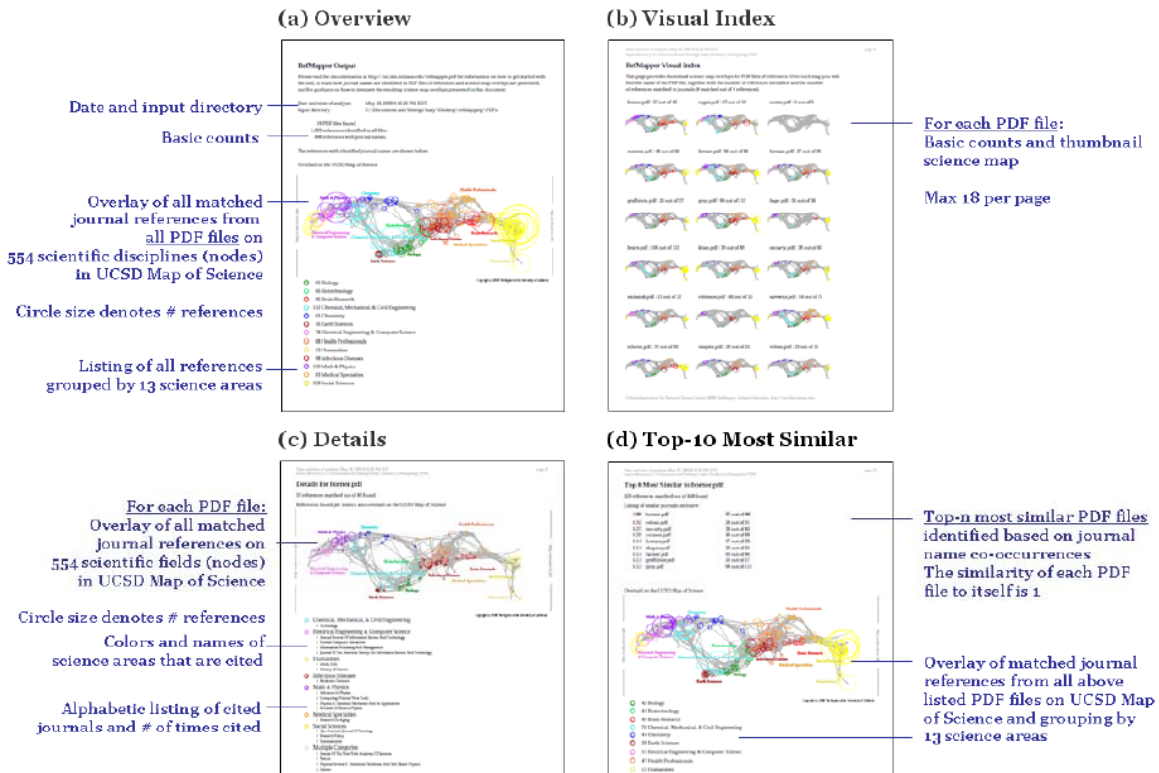
- ISI Data: C:\Users\User\Desktop\Sci2-Tool-NICO\sampladata\scientometrics\isi\FourNetSciResearchers.isi
- 361 Unique ISI Records
- Extracted Co-Authorship Network
- With community attributes
- Circular Hierarchy Viz.ps
- Author information
- Prefuse CSV file: C:\Users\User\Desktop\Sci2-Tool-NICO\sampladata\scientometrics\geo\uspto_geo_data.csv
- Latitude & Longitude for "COUNTRY" in "country" is added.
- PostScript: Latitude & Longitude for "COUNTRY" in "country" is added.

The 'Geo Maps (circles)' dialog box is open, showing the following settings:

- Map: Countries
- Projection: Mercator
- Author Name: Katy Borner
- Latitude: Latitude
- Longitude: Longitude
- Size Circles By: patents
- Size Scaling: linear
- Color Circle Exteriors By: cited_times
- Exterior Color Scaling: linear
- Exterior Color Range: Green to Red

Buttons: OK, Cancel





RefMapper Output

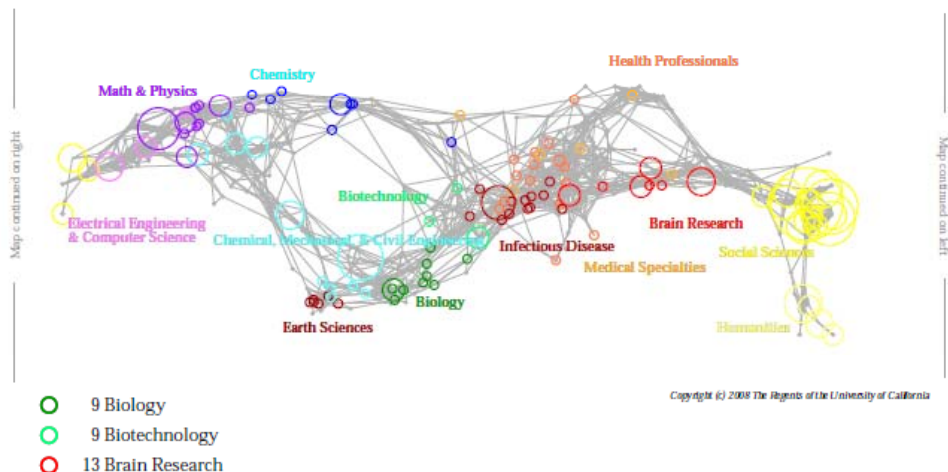
Please read the documentation at <http://sci.slis.indiana.edu/refmapper.pdf> for information on how to get started with the tool, to learn how journal names are identified in PDF files of references and science map overlays are generated, and for guidance on how to interpret the resulting science map overlays presented in this document.


Date and time of analysis: June 5, 2009 10:16:47 AM EDT
 Input directory: C:\Users\User\Desktop\RefMapper\Data\Scisip-2008\Implementing_science_policy

6 PDF files found.
 112 references identified in all files.
 107 references with journal names.

The references with identified journal names are shown below:

Overlaid on the UCSD Map of Science





Science of Science Cyberinfrastructure — P O R T A L —

Provided by the [Cyberinfrastructure for Network Science Center](#) at Indiana University.

Introduction
E. O. Wilson writes in *Consilience: The Unity of Knowledge* (1998): "Features that distinguish science from pseudoscience are repeatability, economy, mensuration, heuristics, and consilience." Please see Börner's [recent presentation](#) at the *A Deeper Look at the Visualization of Scientific Discovery* NSF Workshop for a general introduction of the needs and the resources provided here.

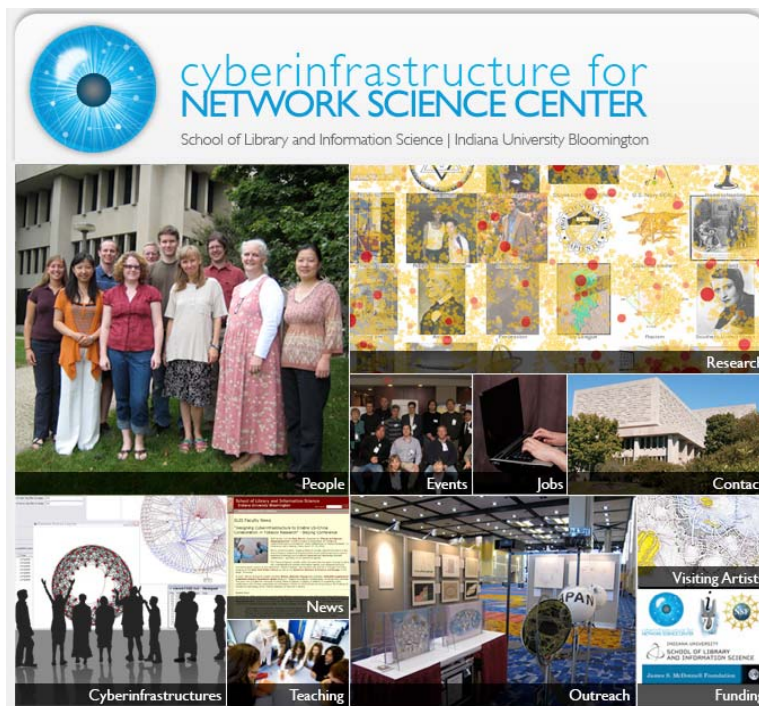
Needs Analysis
As part of the "TLS: Towards a Macroscopic for Science Policy Decision Making" NSF SBE-0738111 award, interviews with science policy makers are conducted to identify what science of science research results and tools might be most desirable and effective. So far, 20 formal, one-hour interviews have been conducted with science policy makers at university campus level, program officer level, and division director level for governmental, state, and private foundations. Data compilation will start in October 2008 and resulting report can be ordered by sending a request to Mark Price (maaprice@indiana.edu).

Conceptualization of Science
A 'science of science' requires a theoretically grounded and practically useful conceptualization of the structure and evolution of science. A special journal issue entitled "[Science of Science: Conceptualizations and Models of Science](#)" edited by [Katy Börner](#), Indiana University & [Andrea Scharnhorst](#), Royal Netherlands Academy of Arts and Sciences invites contributions on this topic. It will be published in the *Journal of Informetrics* 3(1) in January 2009.

Scholarly Database
The [Scholarly Database \(SDB\)](#) at Indiana University aims to serve researchers and practitioners interested in the analysis, modeling, and visualization of large-scale scholarly datasets. The database currently provides access to over 20 million papers, patents and grants. Resulting datasets can be downloaded in bulk. Register for free access at <https://sdb.slis.indiana.edu/>.

Cyberinfrastructures
The Scientometrics filling of the [Network Workbench \(NWB\) Tool](#) provides a unique distributed, shared resources environment for large-scale network analysis, modeling, and visualization. Thomson Scientific/ISI, Scopus and Google Scholar data, EndNote and Bibtext files, or NSF awards can be read and diverse networks can be extracted and studied. Download [User Manual with focus on Scientometrics](#).

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



**cyberinfrastructure for
NETWORK SCIENCE CENTER**
School of Library and Information Science | Indiana University Bloomington

Research

People

Events

Jobs

Contact

News

Cyberinfrastructures

Teaching

Outreach

Visiting Artists

Funding

All papers, maps, cyberinfrastructures, talks, press are linked from <http://cns.slis.indiana.edu>