



Rete-Netzwerk-Red: Analyzing and Visualizing Scholarly Networks Using the Network Workbench Tool

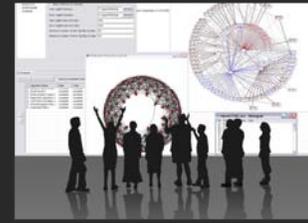
Dr. Katy Börner

Cyberinfrastructure for Network Science Center, Director
Information Visualization Laboratory, Director
School of Library and Information Science
Indiana University, Bloomington, IN
katy@indiana.edu

Co-Authors:

Bonnie (Weixia) Huang, Micah Linnemeier, Russell J. Duhon,
Patrick Phillips, Ninali Ma, Angela Zoss, Hanning Guo,
Mark A. Price

July 17, 2009, ISSI, Rio de Janeiro, Brazil



Overview

What cyberinfrastructure is required to measure, model, analyze, and communicate scholarly data and ultimately scientific progress?

This talk presents our efforts to create a science of science cyberinfrastructure that supports:

- Data access and federation via the Scholarly Database, <http://sdb.slis.indiana.edu>,
- Data preprocessing, modeling, analysis, and visualization using plug-and-play cyberinfrastructures such as the Network Workbench, <http://nwb.slis.indiana.edu>, and
- Communication of science to a general audience via the *Mapping Science* exhibit at <http://scimaps.org>.

This talk should be particularly interesting for those interested to

- Map their very own domain of research,
- Test and compare data federation, mining, visualization algorithms on large scale datasets,
- Use advanced network science algorithms in their own research.



Overview

1. Needs Analysis
2. Conceptualizations of Science
3. Scholarly Database (SDB)
4. Network Workbench (NWB) Tool
- 5. Exemplary Analyses and Visualizations using SDB/NWB**
6. Mapping Science Exhibit



1. Needs Analysis

A total of 34 science policy makers and researchers at university campus level (8), program officer level (12), and division director level at national, state, and private foundations (10) as well as science policy makers from Europe and Asia (4) were interviewed between Feb. 8th, 2008 and Oct. 2nd, 2008.

Each interview comprised a 40 min, audio-taped, informal discussion on specific information needs, datasets and tools currently used, and information on what a 'dream tool' might look and feel like. There is also a pre-interview questionnaire to acquire demographics and a post-interview questionnaire to get input on priorities.

Data compilation is in progress, should be completed in July 2009, and will be submitted as a journal paper. Some data excerpts are given here.

In the Post-Questionnaire Subjects were asked:

“What are initial thoughts regarding the utility of science of science studies for improving decision making? How would access to datasets and tool speed up and increase the quality of your work?”

Excerpts of answers:

- Two areas have great potential: Understanding S&T as a dynamic system, means to display, visualize and manipulate large interrelated amounts of data in maps that allow better intuitive understanding.
- Look for new areas of research to encourage growth/broader impacts of research-- how to assess/ transformative science--what scientific results transformed the field or created a new field/ finding panelists/reviews/ how much to invested until a plateau in knowledge generation is reached/how to define programs in the division.
- Scientometrics as cartography of the evolution of scientific practice that no single actor (even Nobel Laureates) can have. Databases provide a macro-view of the whole of scientific field and its structure. This is needed to make rational decision at the level of countries/states/provinces/regions.
- Understanding where funded scientists are positioned in the global map of science.
- Self-knowledge about effects of funding/ self-knowledge about how to improve funding schemes.
- Ability to see connections between people and ideas, integrate research findings, metadata, clustering career measurement, workforce models, impact (economic/social) on society-interactions between levels of science; lab, institution, agency, Fed Budget, public interests.
- It would be valuable to have tools that would allow one automatically to generate co-citation, co-authorship maps...I am particularly interested in network dynamics.

- It would enable more quantitative decision making in place of an "impression-based" system, and provide a way to track trends, which is not done now.
- When NSF started SciSIP, I was skeptical, but I am more disposed to the idea behind it now although I still don't have a clear idea what scientific metrics will be....how they will apply across disciplines and whether it's really possible to predict with any accuracy the consequences of any particular decision of a grant award.
- SoS potentially useful to policymakers by providing qualitative and quantitative data on the impacts of science toward government policy goals...ideally these studies would enable policy makers to make better decisions for linking science to progress toward policy goals.
- Tracking faculty's work over time to determine what factors get in the way of productivity and which enhance, e.g. course-releases to allow more time--does this really work or do people who want to achieve do so in spite of barriers.
- I'm not sure that this has relevance to my decision-making. There is a huge need for more reliable data about my organization and similar ones, but that seems distinct from data and tools to study science.
- It would assist me enormously.
- Help to give precedents that would rationalize decisions--help to assess research outside one's major area. Ways of assessing innovation, ways of assessing interactions (among researchers, across areas, outside academia).
- It would allow me to answer questions from members of congress provide visual presentations of data for them.
- Very positive step--could fill important need in understanding innovation systems and organizations.



2. Conceptualizations of Science

See Special Issue of *Journal of Informetrics*, 3(3), Jan 2009.

Science of Science: Conceptualizations and Models of Science

Guest Editors: Katy Börner, Indiana University & Andrea Scharnhorst, Royal Netherlands Academy of Arts and Sciences

This special issue of the journal *Informetrics* aims to improve our understanding of the structure and evolution of science by reviewing and advancing existing conceptualizations and models of scholarly activity.

Existing conceptualizations and models of science have been created by scholars from very different disciplines and backgrounds. They have the form of

- philosophical concepts (Bernal, Kuhn, Popper),
 - (utopian) stories (Wells, Lem),
 - visual drawings (Otlet),
 - empirical measurements (Price, Garfield), or
 - mathematical theories (Goffman, Yablonski)
- among others.

It is our belief that a theoretically grounded and practically useful shared conceptualization of science can provide the intellectual framework to interlink and puzzle together the hundreds of science models in existence today. This is analogous to how meteorologists or seismologists integrate rather different local weather models or seismic hazard predictions into a global coherent model that has higher predictive value and broader coverage. With this issue we aim to start an interdisciplinary discourse towards a science of science models.

The design of such a conceptualization requires the identification of the

- Boundaries of the system or object.
- Basic building blocks of science, e.g., units of analysis or key actors.
- Interactions of building blocks, e.g., via coupled networks.
- Basic mechanisms of growth and change.

Editorial is available at <http://inl.slis.indiana.edu/km/pub/2009-borner-scharnhorst-joj-sos-intro.pdf>



3. Scholarly Database

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



Nianli Ma

“From Data Silos to Wind Chimes”

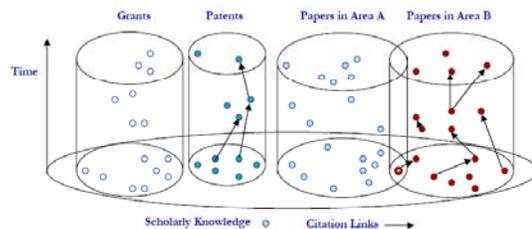
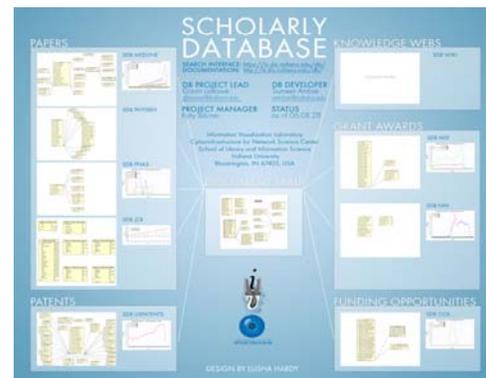


Figure 1: The interoperability and cross linkage problem. Many but not all of today's scholarly datasets, e.g., papers, patents, grants, are stored and made available so that 'vertical' citation linkages can be traversed. There are very few instances in which datasets of different origin and/or type are 'horizontally' interlinked.



- Create public databases that any scholar can use. Share the burden of data cleaning and federation.
- Interlink creators, data, software/tools, publications, patents, funding, etc.

La Rowe, Gavin, Ambre, Sumeet, Burgoon, John, Ke, Weimao and Börner, Katy. (2007) *The Scholarly Database and Its Utility for Scientometrics Research*. In *Proceedings of the 11th International Conference on Scientometrics and Informetrics*, Madrid, Spain, June 25-27, 2007, pp. 457-462. <http://ella.slis.indiana.edu/~katy/paper/07-issi-sdb.pdf>



3. Scholarly Database: # Records & Years Covered

Datasets available via the Scholarly Database (* internally)

Dataset	# Records	Years Covered	Updated	Restricted Access
Medline	17,764,826	1898-2008	Yes	
PhysRev	398,005	1893-2006		Yes
PNAS	16,167	1997-2002		Yes
JCR	59,078	1974, 1979, 1984, 1989 1994-2004		Yes
USPTO	3,710,952	1976-2008	Yes*	
NSF	174,835	1985-2002	Yes*	
NIH	1,043,804	1961-2002	Yes*	
Total	23,167,642	1893-2006	4	3

Aim for comprehensive time, geospatial, and topic coverage.



3. Scholarly Database: Web Interface

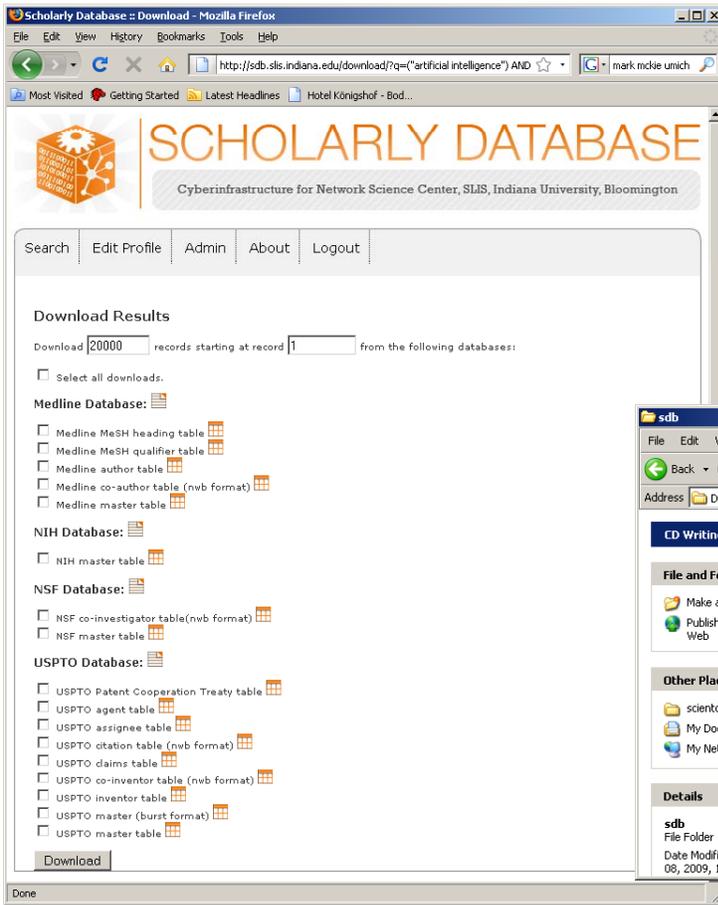
Anybody can register for free to search the about 23 million records and download results as data dumps.

Currently the system has over 120 registered users from academia, industry, and government from over 60 institutions and four continents.

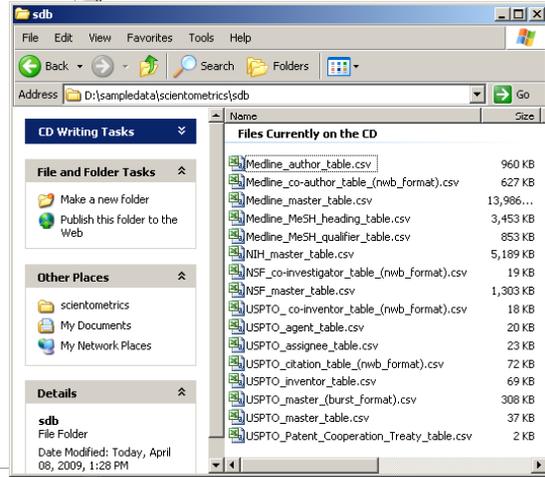
The screenshot displays two browser windows from Mozilla Firefox. The left window shows the search interface with the URL <http://sdb.sls.indiana.edu/search/>. The search box contains the text "artificial intelligence". Below the search box, there are filters for "First Year" (1865) and "Last Year" (2008). A list of databases is checked: Medline (1865 - 2008), NIH (1961 - 2002), NSF (1985 - 2004), and USPTO (1976 - 2008). A "Search" button is visible at the bottom.

The right window shows the "Browse Results" page for the search query "artificial intelligence". The URL is <http://sdb.sls.indiana.edu/search/results?q='artificial intelligence'>. It indicates that the search returned 13,233 results in 0.295 seconds. A "Download" button is present. Below this, it shows "Total results per database: NIH: 2,103, Medline: 10,235, USPTO: 279, NSF: 614." The results list shows "Results 1 through 20." with a "Next>>" link. The table below lists the first few results:

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



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



4. Scientometrics Filling of Network Workbench Tool

will ultimately be 'packaged' as a SciPolicy' tool

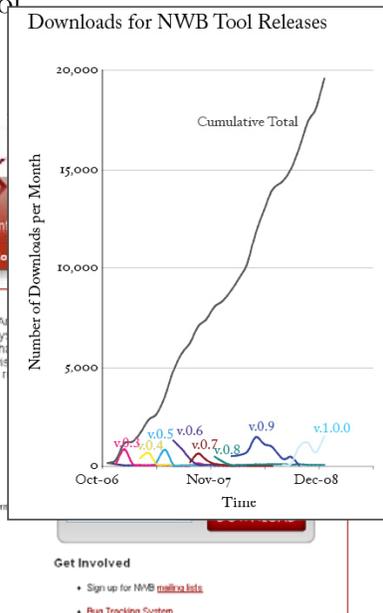
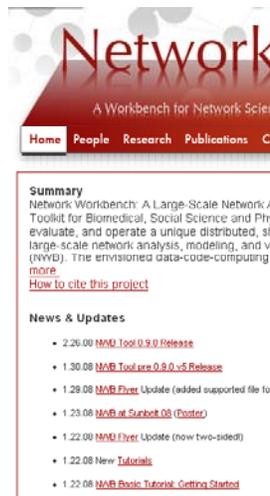
<http://nwb.slis.indiana.edu/>

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

In Feb. 2009, the tool provides more 100 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 19,000 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.

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, Mark Price, 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>



4. NWB Tool: Supported Data Formats

Personal Bibliographies

- Bibtex (.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, Bibtex, 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)

4. NWB Tool: Algorithms (July 1st, 2008)

See <https://nwb.slis.indiana.edu/community> and handout for details.

Preprocessing <small>Edit</small>	Analysis <small>Edit</small>	Visualization <small>Edit</small>
Remove Nodes Extract Top Nodes Extract Nodes Above or Below Val Delete High Degree Nodes Delete Random Nodes Delete Isoletes	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 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	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)
Modeling <small>Edit</small> General Random Graph Watts-Strogatz Small World Barabási-Albert Scale-Free Structured CAN Chord Unstructured Hypergrid PRU Other TARL Discrete Network Dynamics	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 Correla Unnamed Category? Page Rank Based on local graph structure Dyad Reciprocity² Arc Reciprocity² Adjacency Transitivity² Based on components	Scientometrics <small>Edit</small> 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 Detect Duplicate Nodes Remove Rows With Multitudinous Fields²



4. NWB Tool: Output Formats

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)

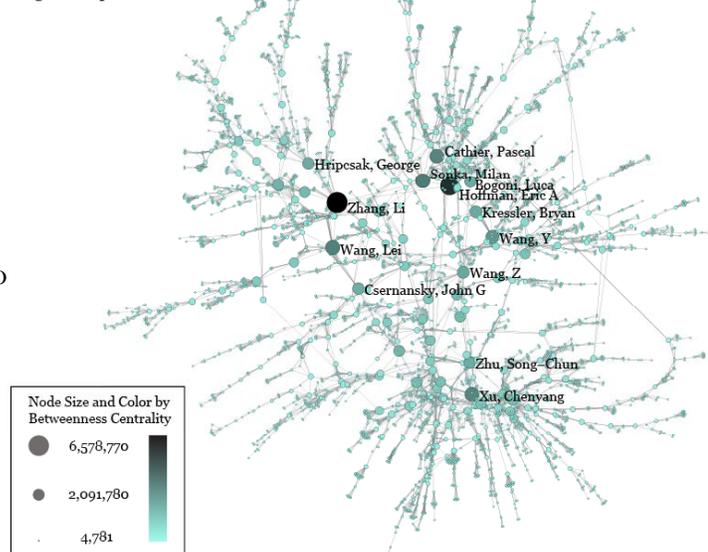
GUESS

- Supports export of images into common image file formats.

Horizontal Bar Graphs

- saves out raster and ps files.

Medline Co-authorship Network
Largest Component





5. 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.



5. 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.



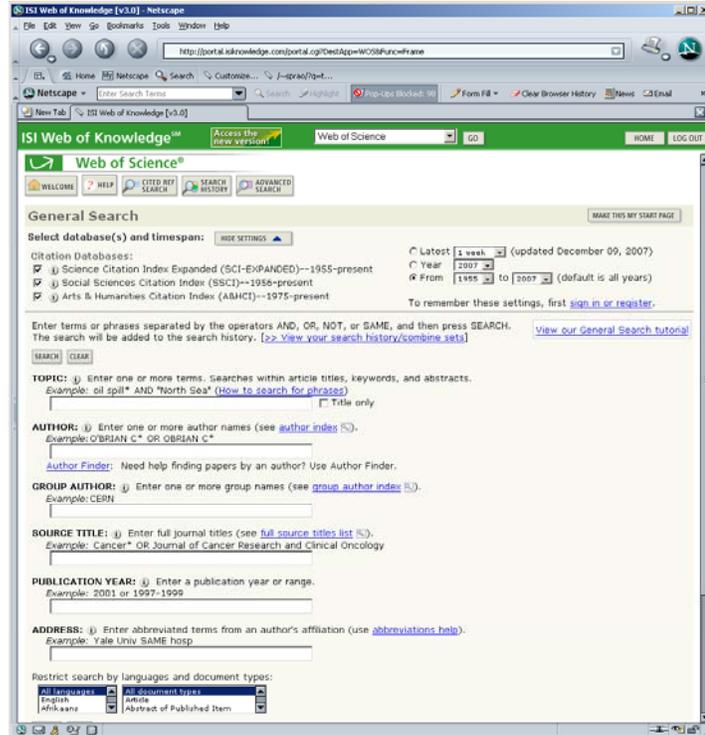
Data Acquisition from Web of Science

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.

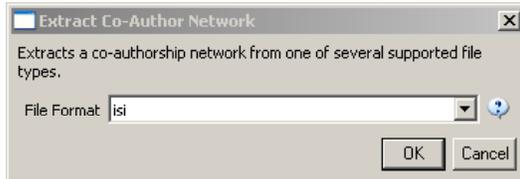
	Age	Total # Cites	Total # Papers	H-Index
Eugene Garfield	82	1,525	672	31
Stanley Wasserman		122	35	17
Alessandro Vespignani	42	451	101	33
Albert-László Barabási	40	2,218	126	47 <i>(Dec 2007)</i>
	41	16,920	159	52 <i>(Dec 2008)</i>



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:

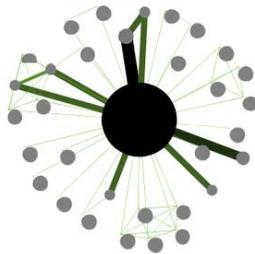


The result is an undirected network of co-authors in the Data Manager. It has 247 nodes and 891 edges.

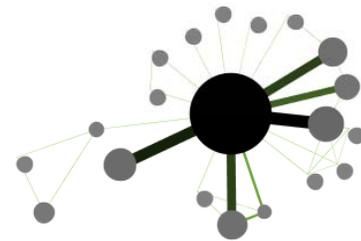
To view the complete network, select the network and run *'Visualization > GUESS > GEM'*. Run *Script > Run Script... . And select Script folder > GUESS > co-author-nw.py*.



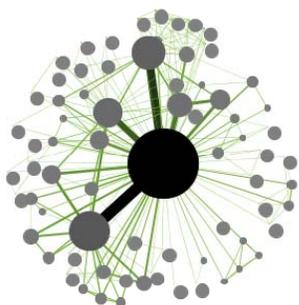
Comparison of Co-Author Networks



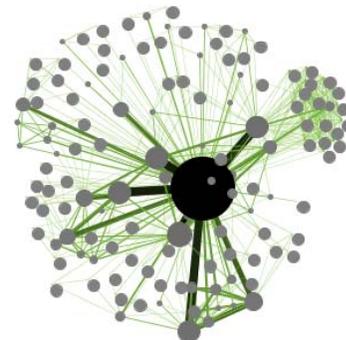
Eugene Garfield



Stanley Wasserman



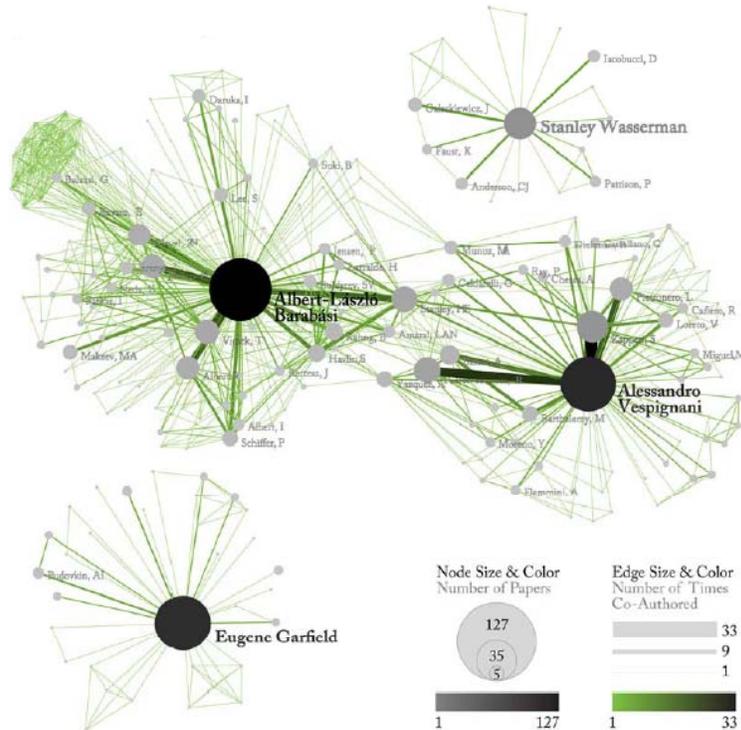
Alessandro Vespignani



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

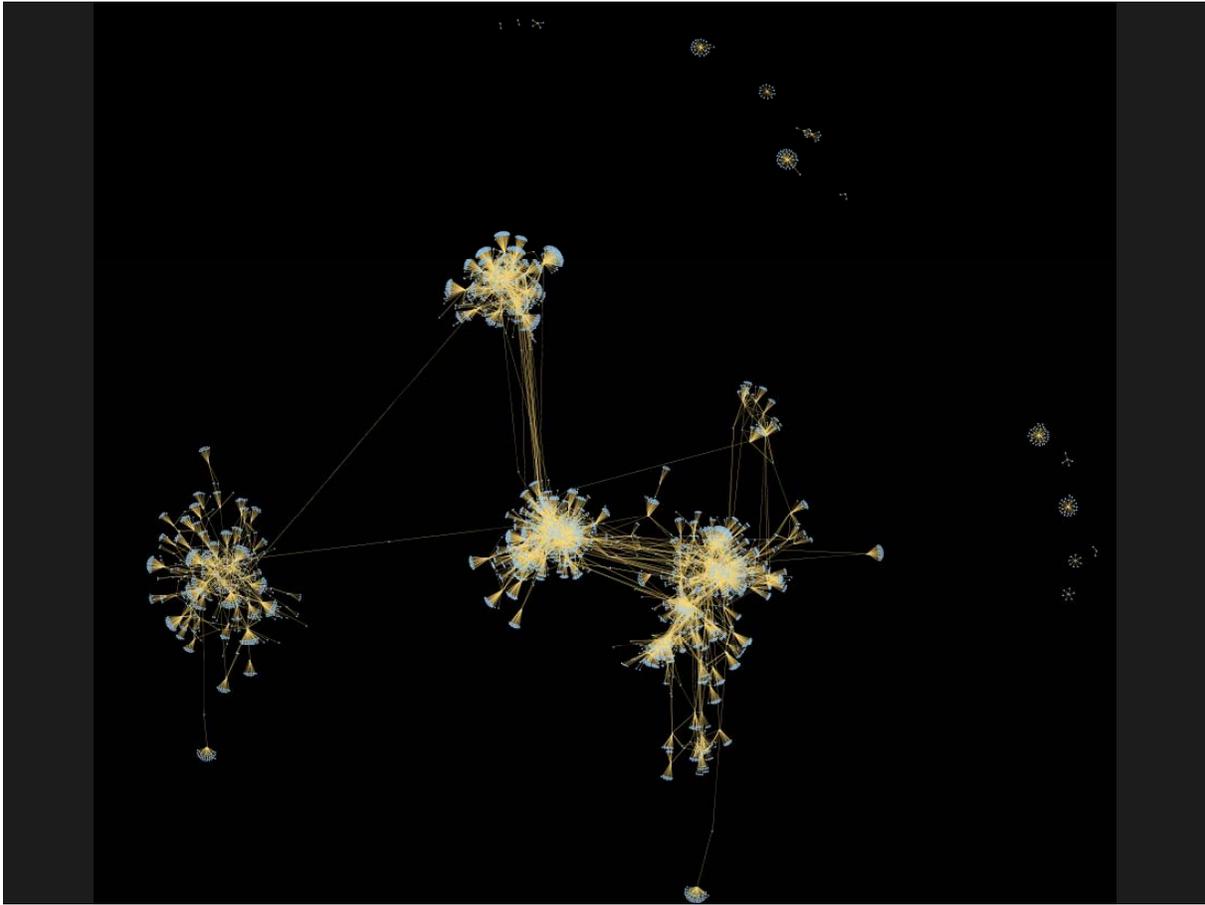
Text Delimiter: |

Aggregate Function File: C:\Documents and Settings\kaly\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*.



5. 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 screenshot shows the NSF Award Search interface. On the left, the search criteria are: Principal Investigator First Name: geoffrey, Last Name: fox, Include CO-PI: checked. On the right, a list of awards is displayed, including titles like 'Research in Computer Science and Computational Physics' and 'Applications of Parallel Supercomputing to Astrophysical N-body Calculations'. A 'Save in CSV format as *name*.nsf' button is highlighted in the export options.



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:

Extract Network from Table

Extracts a network from a delimited table

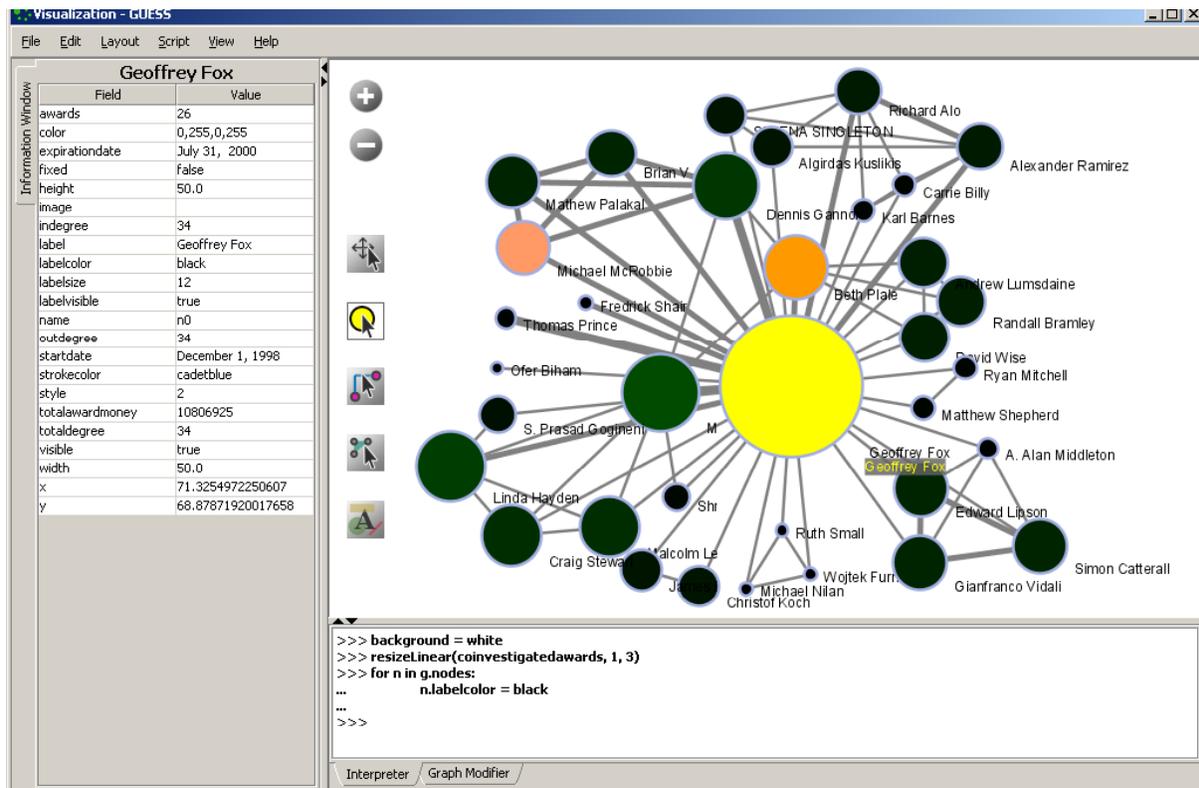
Column Name: All Investigators

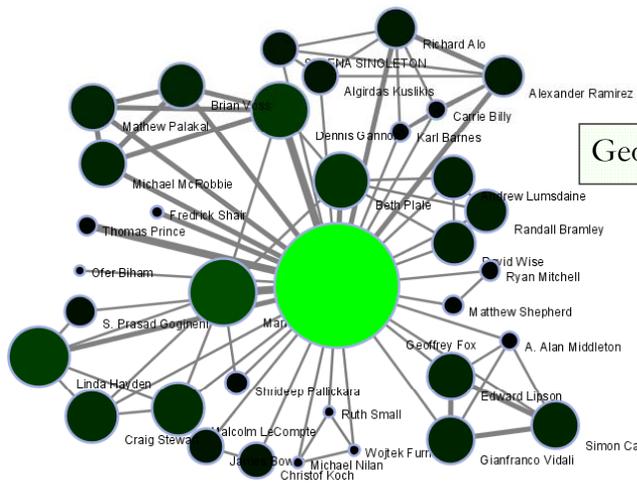
Text Delimiter: |

Aggregation Function File: C:\Documents and Settings\katy\Desktop\nwb-scipolicy\sampladata\scientometrics\properties\nsfCoPI.properties

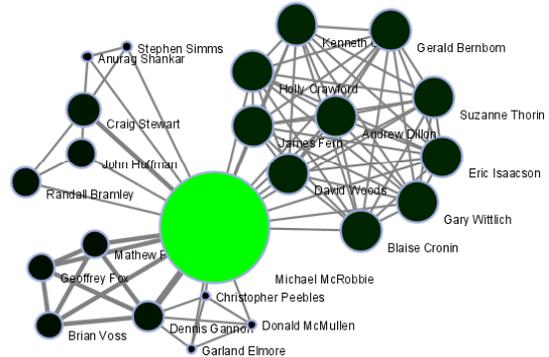
OK Cancel

- 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.

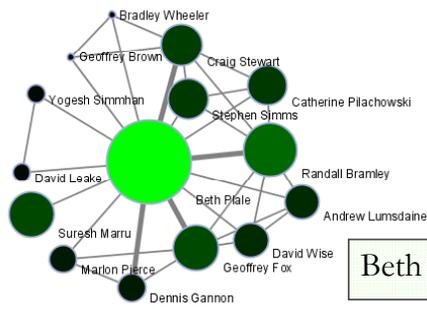




Geoffrey Fox



Michael McRobbie



Beth Plale

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



5. 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 screenshot displays the NSF Awards Search website in a Windows Internet Explorer browser. The page is titled "NSF - Award Search - Search All Fields" and features the NSF logo and navigation menu. The "Award Search" section is active, showing a search form with a "Search Award For" field and a "Restrict to Title Only" checkbox. The "Awardee Information" section includes fields for "Principal Investigator" (First Name, Last Name), "Include CO-PI" (checked), "Organization" (University of Michigan Ann Arbor), "State", "ZIP Code", and "Country".

The search results are displayed in a table with columns for Award Number, Title, DUE, CMMI, and Date. The results are sorted by award date, with the most recent awards at the top. A "Save in CSV format as *institution*.nsf" button is visible above the table.

Award Number	Title	DUE	CMMI	Date	PI
0820609	Physiology				
0817369	Teaching of Mathematical Knowledge for Teaching (K-12): Adapting Local Materials for Use in Diverse Institutions and Settings	DUE		01/01/2009	Bass, Hyma
0822892	Protest Psychosis: Race, Science, and the Stigma of Schizophrenia	SES		01/01/2009	Metel, Jonathan
0825795	Collaborative Research: Tissue Cutting Mechanics - Investigation of the Effective and Minimally Invasive Biology	CMMI		01/01/2009	Shih, Albert
0855698	IMPLEMENTING THE "5X1ME" WORKSHOP RECOMMENDATIONS	CMMI		01/01/2009	Ulsoy, A. G.
0825789	Short-Term Joint Maintenance and Production Decision Support Tool of Manufacturing Systems	CMMI		01/01/2009	Ni, Jun
0820609	Support for the 6th U.S.				

Active NSF Awards on 11/07/2008:

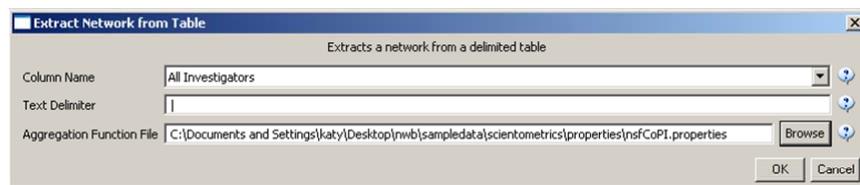
- Indiana University 257
(there is also Indiana University at South Bend Indiana University Foundation, Indiana University Northwest, Indiana University-Purdue University at Fort Wayne, Indiana University-Purdue University at Indianapolis, Indiana University-Purdue University School of Medicine)
- Cornell 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)

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'*.

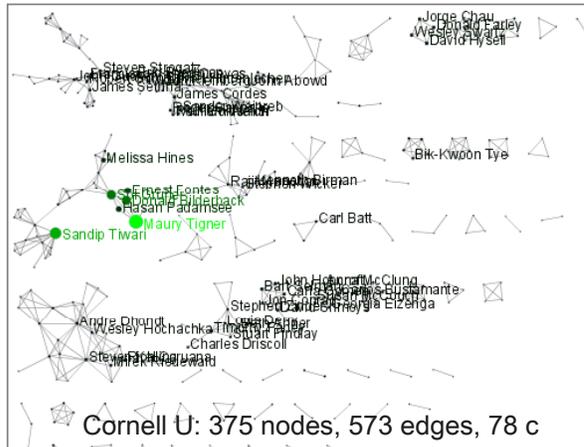
Select *'Visualization > GUESS'* to visualize. Run *'co-PI-nw.py'* script.



Indiana U: 223 nodes, 312 edges, 52 components



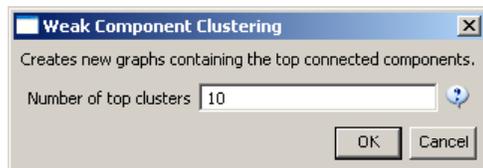
U of Michigan: 497 nodes, 672 edges, 117 c



Cornell U: 375 nodes, 573 edges, 78 c

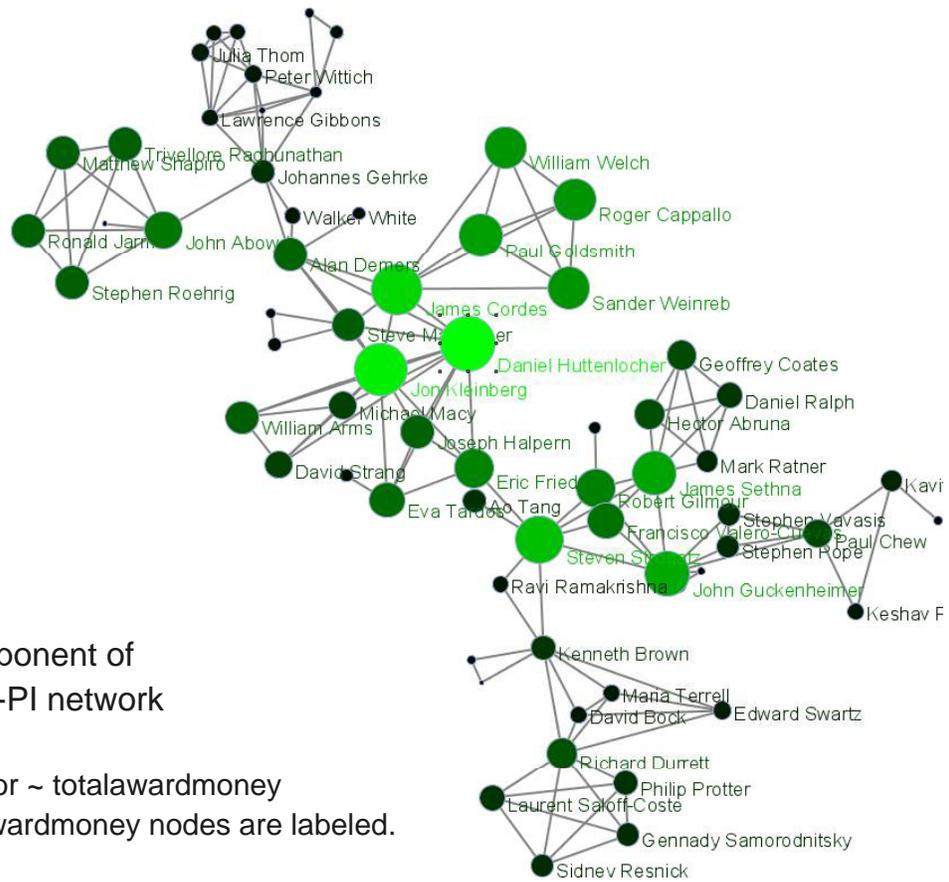
Extract Giant Component

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: 23,35,0921
 Paul Berman: 23,350,921
 Roberto Merlin: 23,350,921
 Robert Schoeni: 21,991,140
 Wei-Jun Jean Yeung: 21,991,140



5. 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.

SCHOLARLY DATABASE
Cyberinfrastructure for Network Science Center, SLIS, Indiana University, Bloomington

Search | Edit Profile | About | Logout

Search

Creators:
Title:
Abstract:
All Text: "artificial intelligence"
First Year: 1898
Last Year: 2008

Medline (1898 - 2008)
 NIH (1961 - 2002)
 NSF (1985 - 2004)
 USPTO (1976 - 2008)

Search

Browse Results

Your search returned 13,225 results in 0.162 seconds.

Total results per database: NIH: 2,103, Medline: 10,229, USPTO: 279, NSF: 614.

Results 1 through 20.

Next>>

Source	Authors/Creators	Year	Title
Medline	LaCombe	1987	Artificial intelligence.
Medline		1989	Artificial intelligence: expert systems.
Medline	Schmitt	1990	[Artificial intelligence in dentistry]
Medline	Adlassinig and Adlassinig	2002	Artificial-intelligence-augmented systems.

Download Results

Select All Sample File Data Dictionary

Medline Database:

Medline master table
 Medline author table
 Medline MeSH heading table
 Medline MeSH qualifier table
 Medline co-author table (nwb format)

NIH Database:

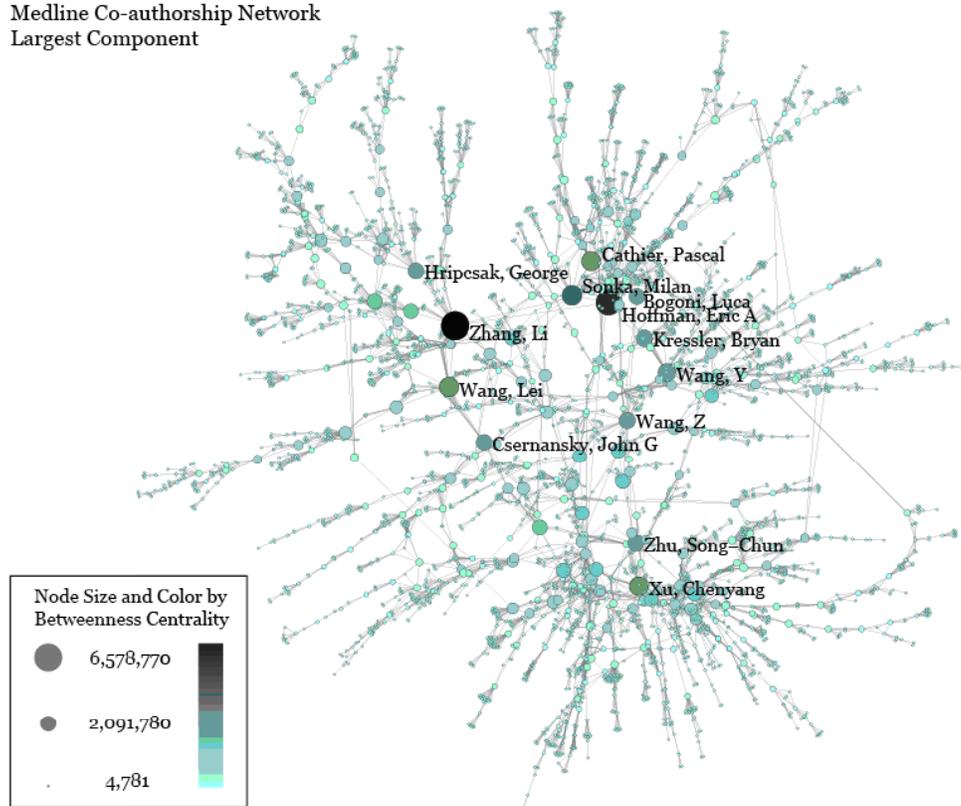
NIH master table

NSF Database:

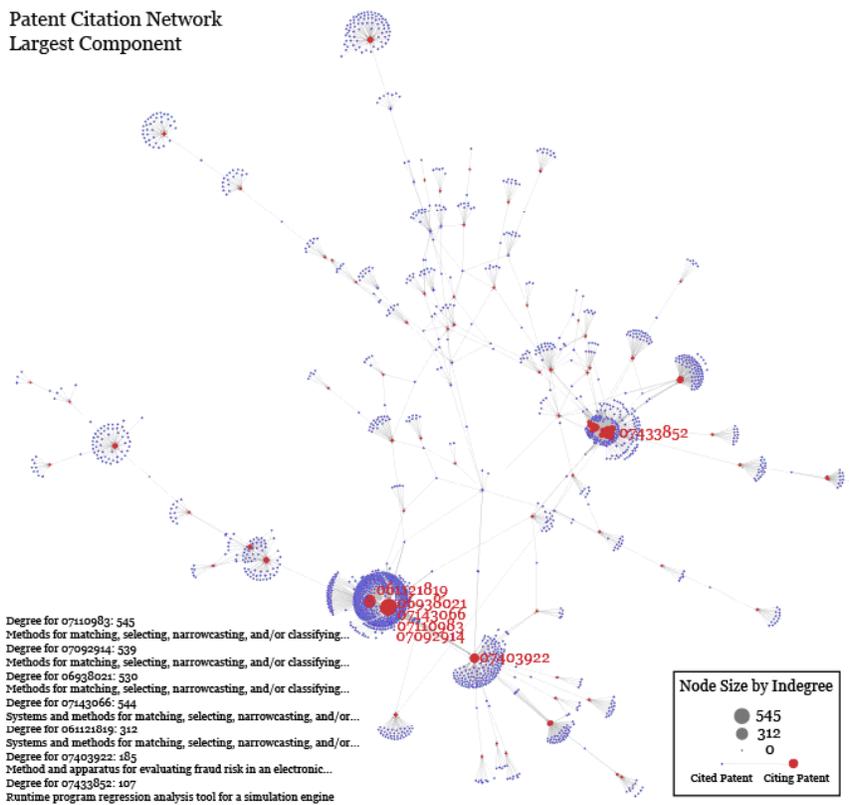
NSF master table
 NSF co-investigator table (nwb format)

Download

Medline Co-authorship Network
Largest Component



Patent Citation Network
Largest Component



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



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, 30 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>

Mapping Science Exhibit – 10 Iterations in 10 years

<http://scimaps.org/>



The Power of Maps (2005)



Science Maps for Economic Decision Makers (2008)



The Power of Reference Systems (2006)



Science Maps for Science Policy Makers (2009)

Science Maps for Scholars (2010)

Science Maps as Visual Interfaces to Digital Libraries (2011)

Science Maps for Kids (2012)

Science Forecasts (2013)

The Power of Forecasts (2007)



How to Lie with Science Maps (2014)



Exhibit has been shown in 52 venues on four continents. Also at

- NSF, 10th Floor, 4201 Wilson Boulevard, Arlington, VA.
- Chinese Academy of Sciences, China, May 17-Nov. 15, 2008.
- University of Alberta, Edmonton, Canada, Nov 10-Jan 31, 2009
- Center of Advanced European Studies and Research, Bonn, Germany, Dec. 11-19, 2008.



Debut of 5th Iteration of Mapping Science Exhibit at MEDIA X on May 18, 2009
at Wallenberg Hall, Stanford University

<http://mediax.stanford.edu>

<http://scaleindependentthought.typepad.com/photos/scimaps>

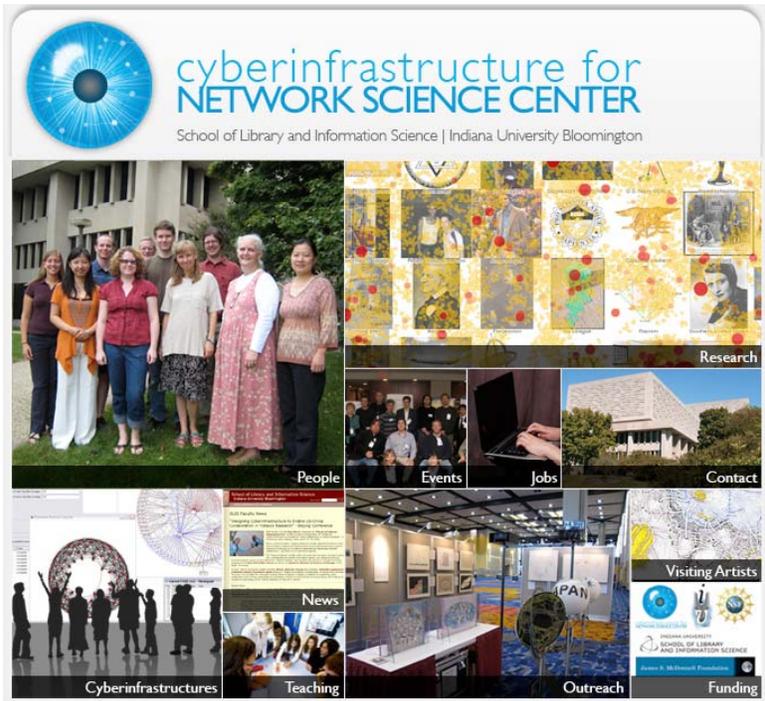


Science Maps in “Expedition Zukunft” science train visiting 62 cities in 7 months
12 coaches, 300 m long
Opened on April 23rd, 2009 by German Chancellor Merkel
<http://www.expedition-zukunft.de>

This is the only mockup in this slide show.

Everything else is available today.





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