

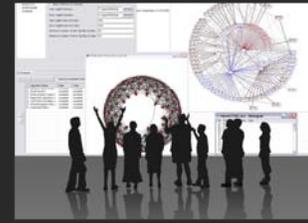


Analyzing and Visualizing Science Using the Scholarly Database and the Network Workbench Tool

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

July 20, 2009, São Paulo, 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.



Scholarly Database

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



Nianli Ma

“From Data Silos to Wind Chimes”

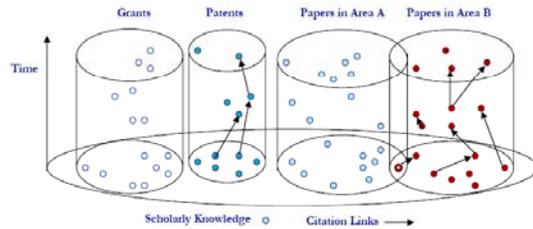
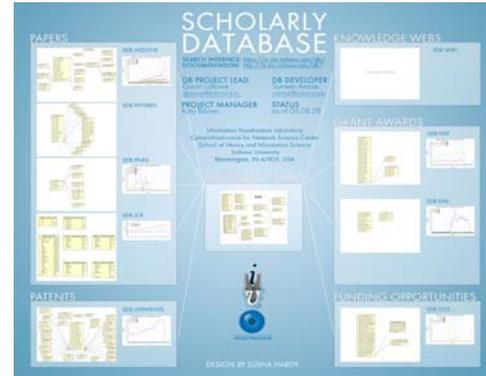


Figure 1: The interoperability and cross linkage problem. Many but not all of today's scholarly datasets, e.g., papers, patents, grants, are stored and made available so that 'vertical' citation linkages can be traversed. There are very few instances in which datasets of different origin and/or type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



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.



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.

Search | Edit Profile | Admin | About | Logout

Search

Creators:

Title:

Abstract:

All Text: "artificial intelligence"

First Year: 1865

Last Year: 2008

Medline (1865 - 2008)

NIH (1961 - 2002)

NSF (1985 - 2004)

USPTO (1976 - 2008)

If multiple terms are entered automatically combined using "AND". Thus 'breast AND cancer' records that contain both terms.

You can put AND between term 'AND'. Thus 'breast AND cancer' records that contain both terms.

Double quotation can be used terms, e.g., "breast cancer" if the phrase "breast cancer" at 'breast' and 'cancer' are both exact phrase.

The importance of a particular increased by putting a "*" and term. For instance, 'breast*' compared to matching the term.

Browse Results

Your search returned 13,231 results in 0.295 seconds.

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

Results 1 through 20.

Next >>

Source	Authors/Creators	Year	Title	Score (out of 5.7)
Medline	LaCombe	1987	Artificial intelligence.	5.71
Medline		1989	Artificial intelligence: expert systems.	5.71
Medline	Schmitt	1990	[Artificial intelligence in dentistry]	5.71
Medline	Adlassnig and Adlassnig	2002	Artificial-intelligence-augmented systems.	5.60

Download Results

Download 20000 records starting at record 1 from the following databases:

Select all downloads.

Medline Database:

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

NIH Database:

- NIH master table

NSF Database:

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

USPTO Database:

- USPTO Patent Cooperation Treaty table
- USPTO agent table
- USPTO assignee table
- USPTO citation table (nwb format)
- USPTO claims table
- USPTO co-inventor table (nwb format)
- USPTO inventor table
- USPTO master (burst format)
- USPTO master table

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

Files Currently on the CD

Name	Size
Medline_author_table.csv	960 KB
Medline_co-author_table_(nwb_format).csv	627 KB
Medline_master_table.csv	13,986...
Medline_MeSH_heading_table.csv	3,453 KB
Medline_MeSH_qualifier_table.csv	853 KB
NIH_master_table.csv	5,189 KB
NSF_co-investigator_table_(nwb_format).csv	19 KB
NSF_master_table.csv	1,303 KB
USPTO_co-inventor_table_(nwb_format).csv	18 KB
USPTO_agent_table.csv	20 KB
USPTO_assignee_table.csv	23 KB
USPTO_citation_table_(nwb_format).csv	72 KB
USPTO_inventor_table.csv	69 KB
USPTO_master_(burst_format).csv	308 KB
USPTO_master_table.csv	37 KB
USPTO_Patent_Cooperation_Treaty_Table.csv	2 KB



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.

Network Workbench
A Workbench for Network Sciences

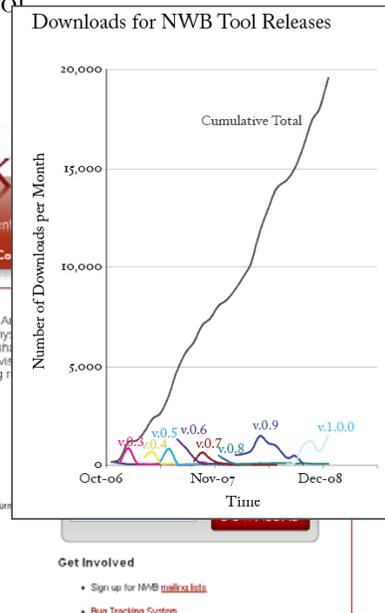
Home People Research Publications Co

Summary
Network Workbench: A Large-Scale Network Analysis Toolkit for Biomedical, Social Science and Physics, and operate a unique distributed, scalable, large-scale network analysis, modeling, and visualization (NWB). The envisioned data-code-computing interface.

How to cite this project

News & Updates

- 2.26.08 [NWB Tool 0.9.0 Release](#)
- 1.30.08 [NWB Tool core 0.9.0 v5 Release](#)
- 1.29.08 [NWB Filter Update](#) (added supported file format)
- 1.23.08 [NWB at Sunbelt 08 \(Poster\)](#)
- 1.22.08 [NWB Filter Update](#) (now two-sided)
- 1.22.08 [New Tutorials](#)
- 1.22.08 [NWB Basic Tutorial: Getting Started](#)



Herr II, Bruce W., Huang, Weixia (Bonnie), Penumathy, 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.

Network Workbench
A Workbench for Network Sciences

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, 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>

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



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)

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 Isolates	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-Coreness²	Tools GUESS GnuPlot² Predefined Positions Layout DrL (VxOrd) Pre-defined Positions (prefuse beta)²
Remove Edges Extract Top Edges Extract Edges Above or Below Val Remove Self Loops Trim By Degree² Pathfinder Network Scaling	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 Correl² Unnamed Category? Page Rank Based on local graph structure Dyad Reciprocity² Arc Reciprocity² Adjacency Transitivity² Based on components	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)
Sampling Snowball Sampling (n nodes) Node Sampling Edge Sampling		Other Layouts Parallel Coordinates (demo)² LaNet (k-Core Decomposition)
Transformations Symmetrize Dichotomize Multipartite Joining		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
Modeling <small>Edit</small> General Random Graph Watts-Strogatz Small World Barabási-Albert Scale-Free		Cleaning Remove ISI Duplicate Records Detect Duplicate Nodes Remove Rows With Multitudinous Fields²
Structured CAN Chord		
Unstructured Hypergrid PRU		
Other TARL Discrete Network Dynamics		



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)

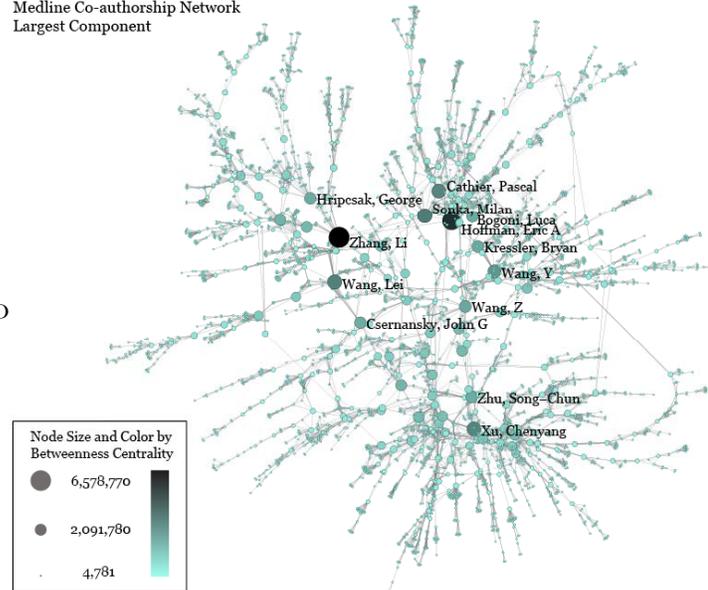
GUESS

- Supports export of images into common image file formats.

Horizontal Bar Graphs

- saves out raster and ps files.

Medline Co-authorship Network
Largest Component



Exemplary Analyses and Visualizations

Individual Level

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

Institution Level

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

Scientific Field Level

- 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

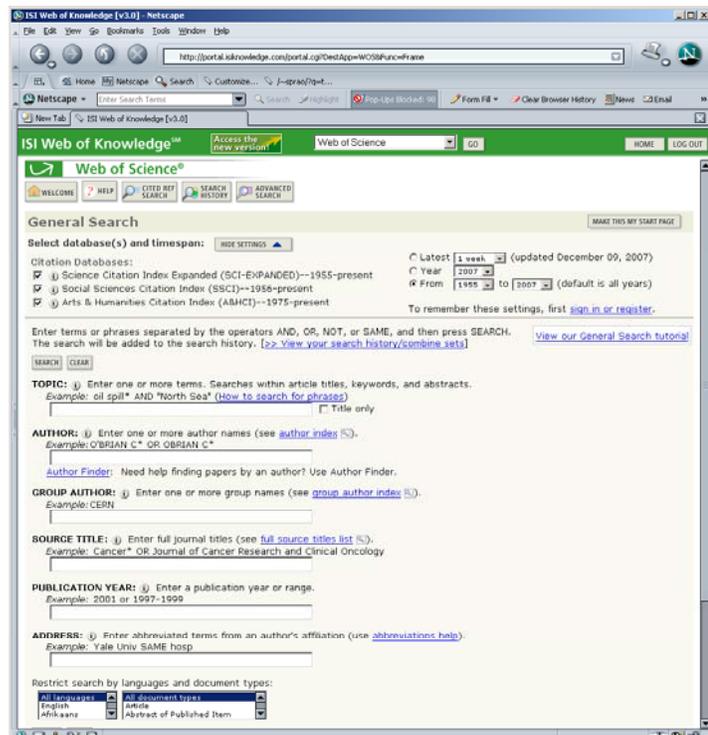
- D. Extracting co-author networks, patent-citation networks, and detecting bursts in SDB data.



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.

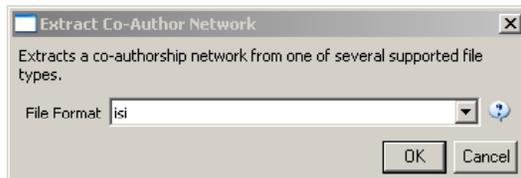
	Age	Total # Cites	Total # Papers	H-Index
Eugene Garfield	82	1,525	672	31
Stanley Wasserman		122	35	17
Alessandro Vespignani	42	451	101	33
Albert-László Barabási	40	2,218	126	47 <i>(Dec 2007)</i>
	41	16,920	159	52 <i>(Dec 2008)</i>



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:

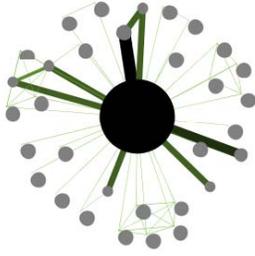


The result is an undirected network of co-authors in the Data Manager. It has 247 nodes and 891 edges.

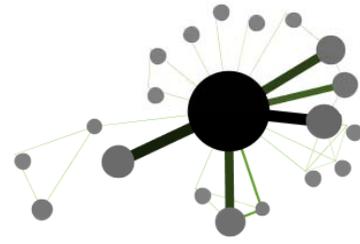
To view the complete network, select the network and run *'Visualization > GUESS > GEM'*. Run *Script > Run Script... . And select Script folder > GUESS > co-author-nw.py*.



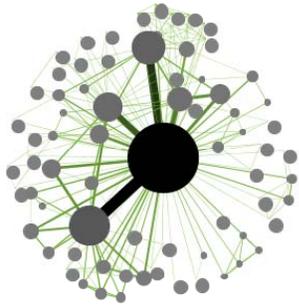
Comparison of Co-Author Networks



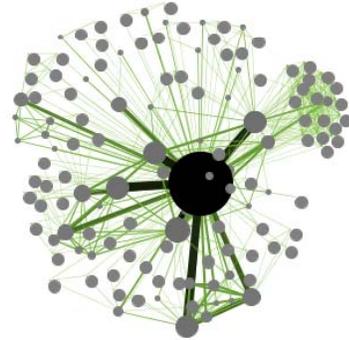
Eugene Garfield



Stanley Wasserman



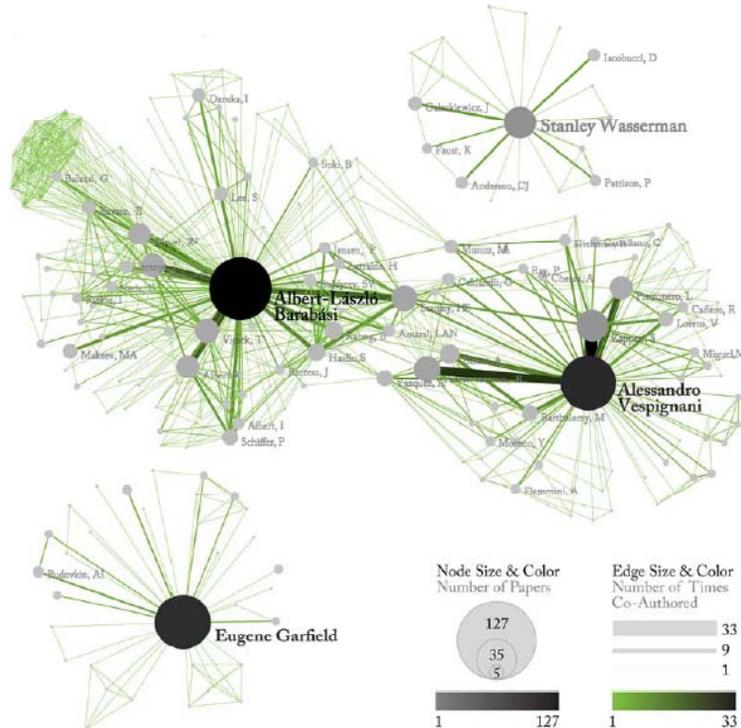
Alessandro Vespignani



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

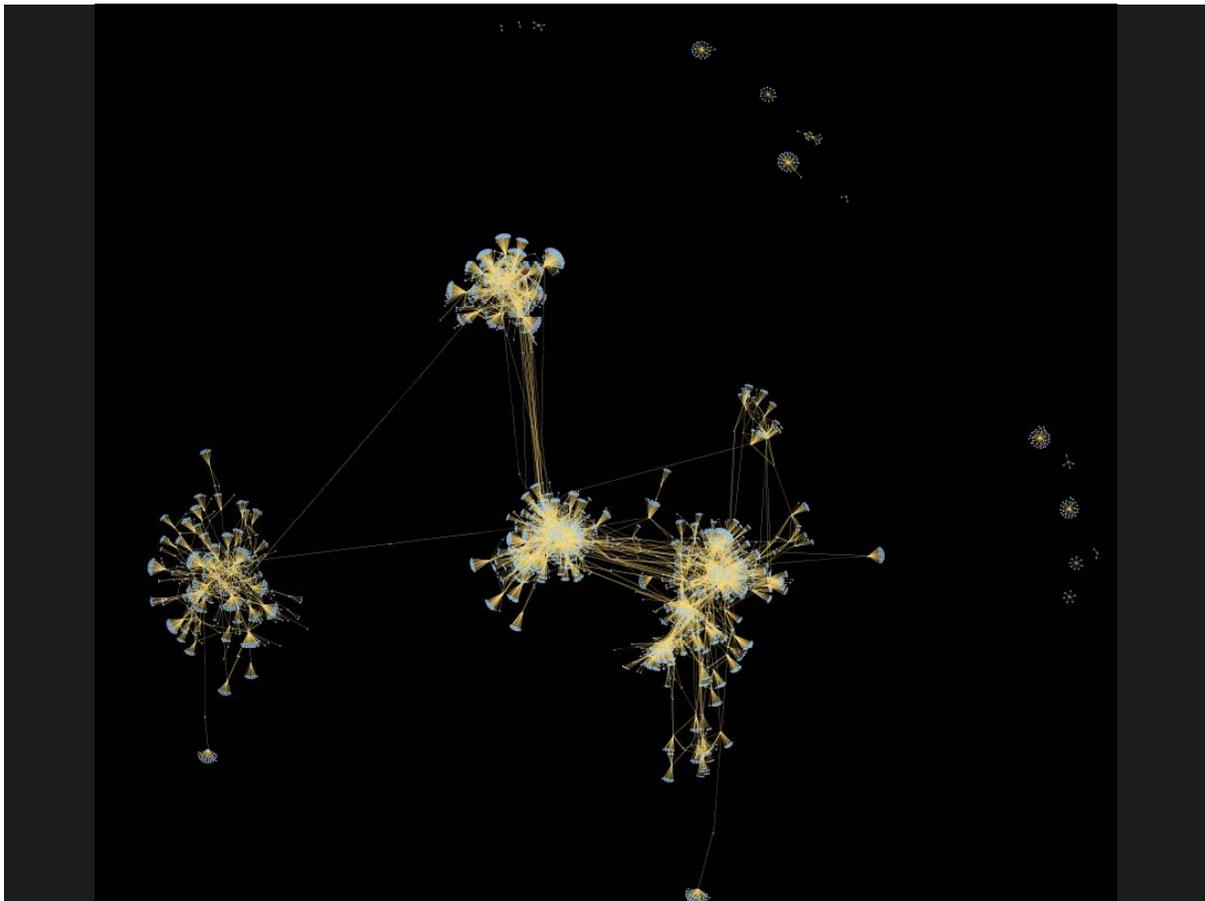
Text Delimiter: |

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 form with the following details:

- Search Award For:** [Empty text box]
- Restrict to Title Only:**
- Awardee Information:**
 - Principal Investigator:**
 - First Name:** geoffrey
 - Last Name:** fox
 - Include CO-PI:**
 - Organization:** [Empty text box]
 - State:** [Dropdown menu]
 - ZIP Code:** [Empty text box]
 - Country:** [Dropdown menu]
- Historical Awards:**
- Active Awards Only:**
- Expired Awards Only:**

- Buttons:** Search, Reset

The right screenshot shows the search results table with the following columns: Award Number, Title, Program, Date, and PI. A text box overlay says "Save in CSV format as *name*.nsf". The "Export options" section has "CSV" selected.

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

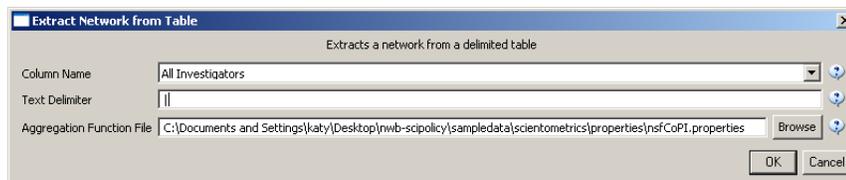
Disclaimer:

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.

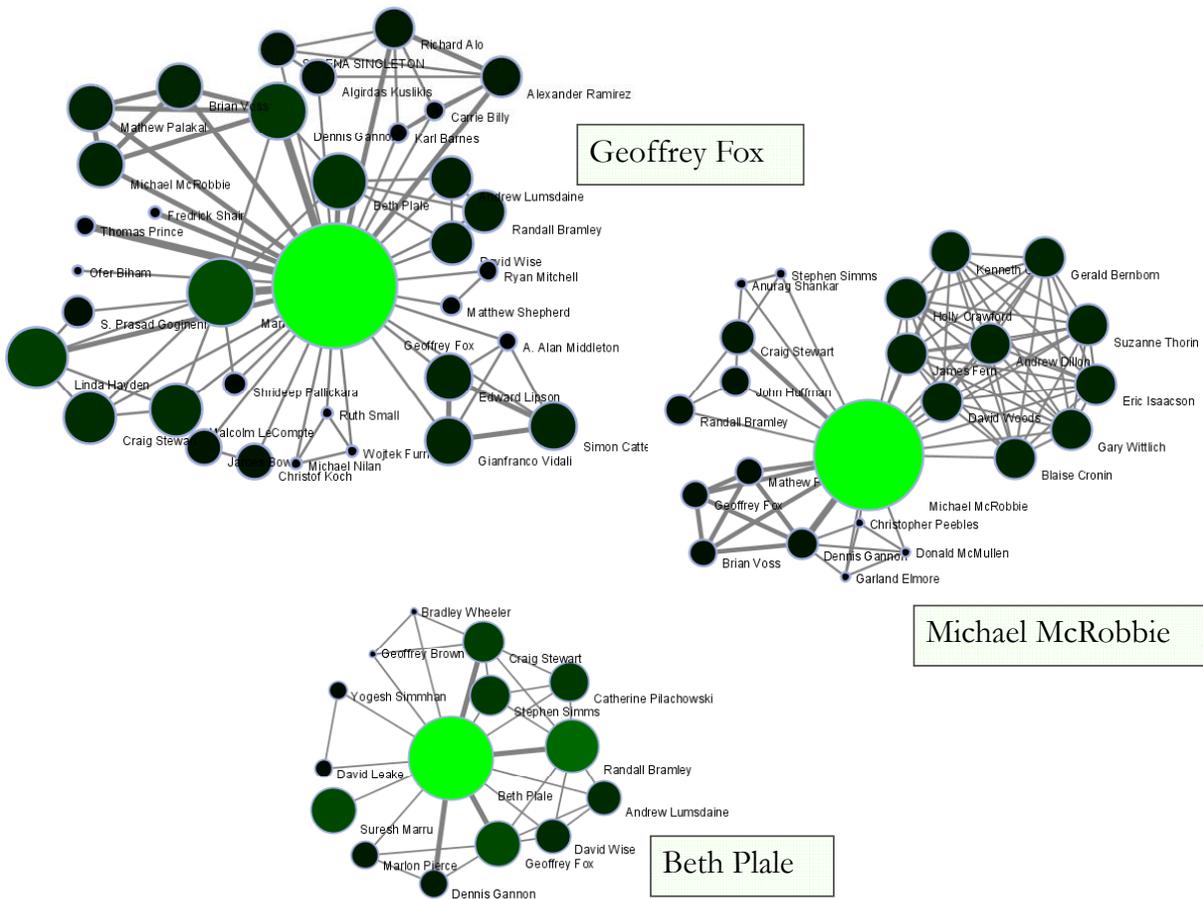
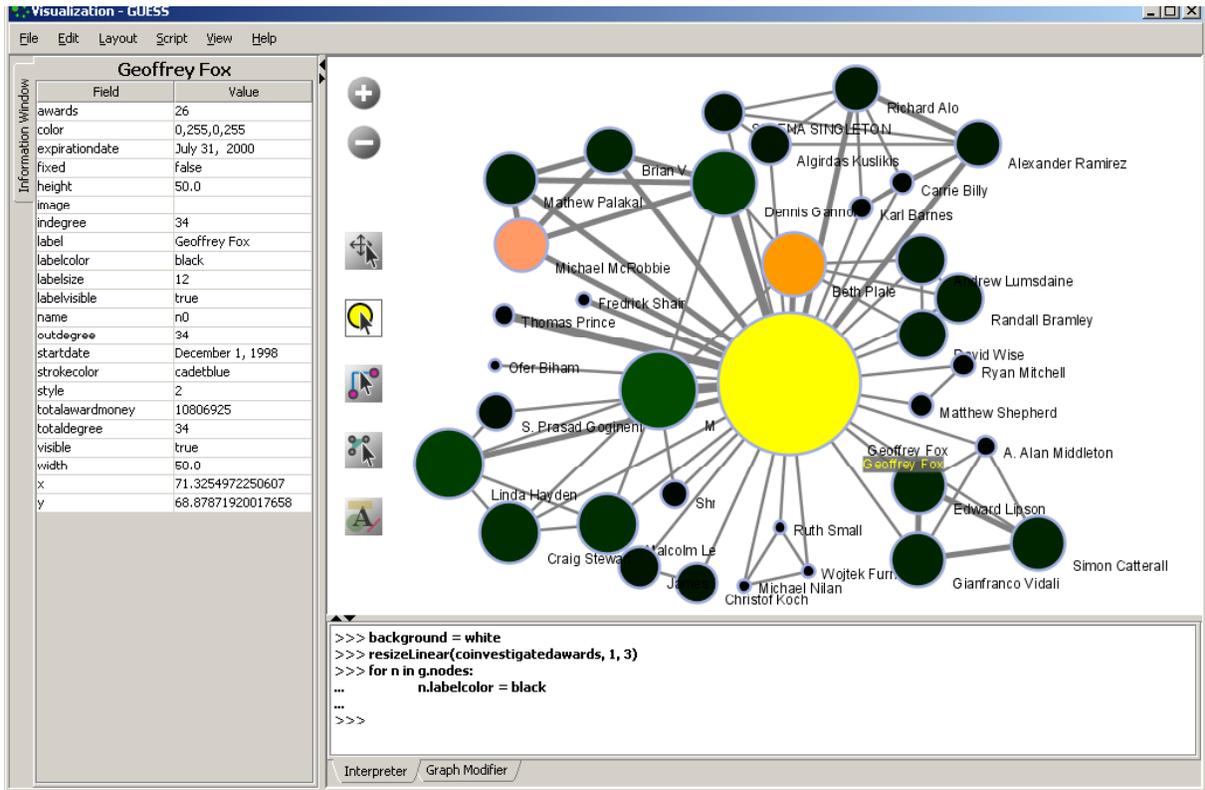


Using NWB to Extract Co-PI Networks

- 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, 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 'Award Search' section includes a search box and filters for 'Award Information', 'Program Information', and 'Search All Fields'. The 'Award Information' filter is expanded, showing 'Principal Investigator' details. The 'Organization' field is set to 'University of Michigan Ann Arbor' and is circled in red. On the right, the search results are displayed in a table. A text box is overlaid on the table with the text 'Save in CSV format as *institution*.nsf'. The table columns include Award Number, Title, Program, Start Date, and PI Name.

Award Number	Title	Program	Start Date	PI Name
0820609	Physical Helping			
0817369	Teaching 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	Metal, Jonat
0825795	Collaborative Research: Tissue Culture Mechanics - Investigation of the Effective and Minimally Invasive Biopsy	CMMI	01/01/2009	Shih, Albert
0855698	IMPLEMENTING THE "SME" WORKSHOP RECOMMENDATIONS	CMMI	01/01/2009	Ulsov, A. G.
0825789	Short-Term Joint Maintenance and Production Decision Support Tool of Manufacturing Systems	CMMI	01/01/2009	Ni, Jun
0825789	Support for the 6th U.S.	COMBUSTION, FIRE, &		

Active NSF Awards on 11/07/2008:

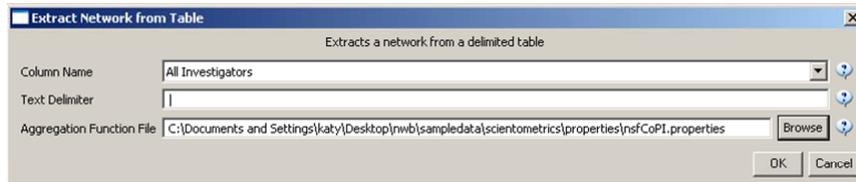
- Indiana University 257
(there is also Indiana University at South Bend Indiana University Foundation, Indiana University Northwest, Indiana University-Purdue University at Fort Wayne, Indiana University-Purdue University at Indianapolis, Indiana University-Purdue University School of Medicine)
- Cornell 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 **yournmbdirectory*/sampledata/scientometrics/nsf/*.

Extracting Co-PI Networks

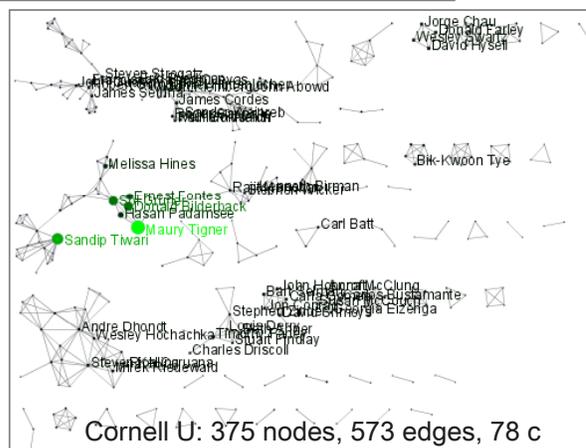
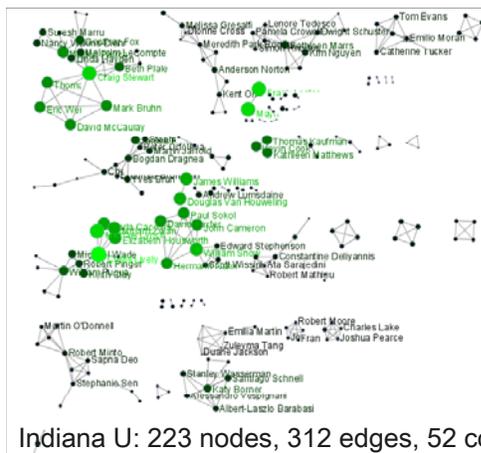
Load NSF data, selecting the loaded dataset in the Data Manager window, run *'Scientometrics > Extract Co-Occurrence Network'* using parameters:



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.



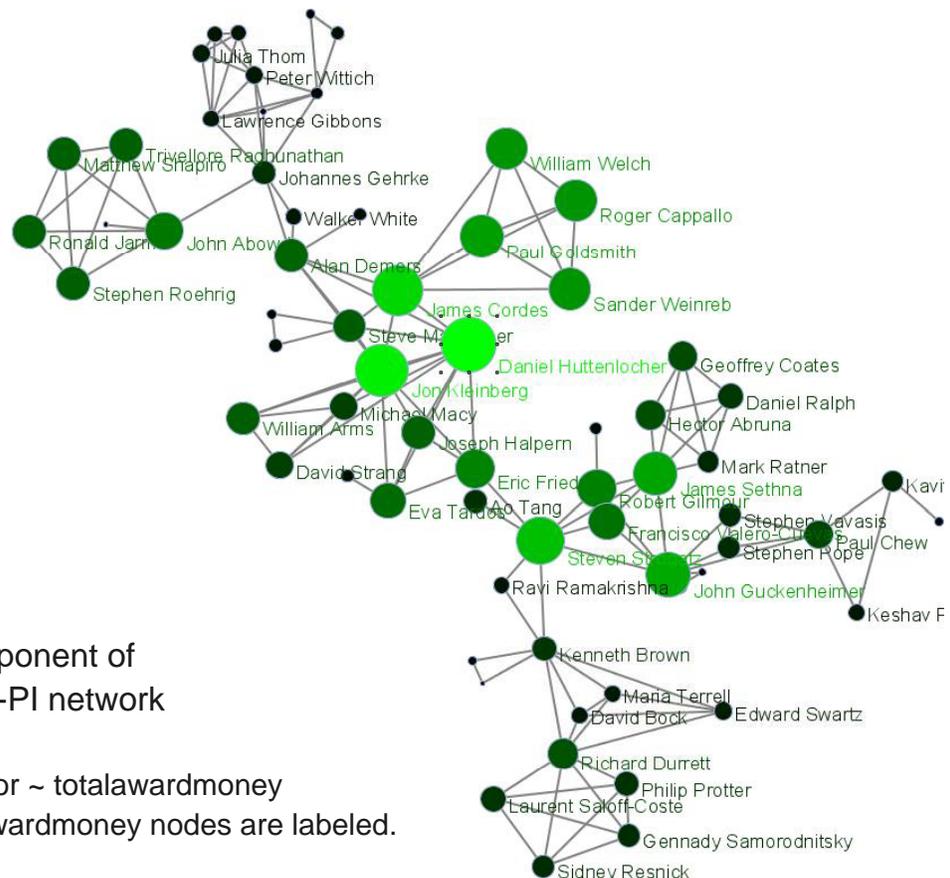
Extract Giant Component

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



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

Search Edit Profile About Logout

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	Adlansnig and Adlansnig	2002	Artificial-intelligence-augmented systems.

Search Edit Profile Admin About

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

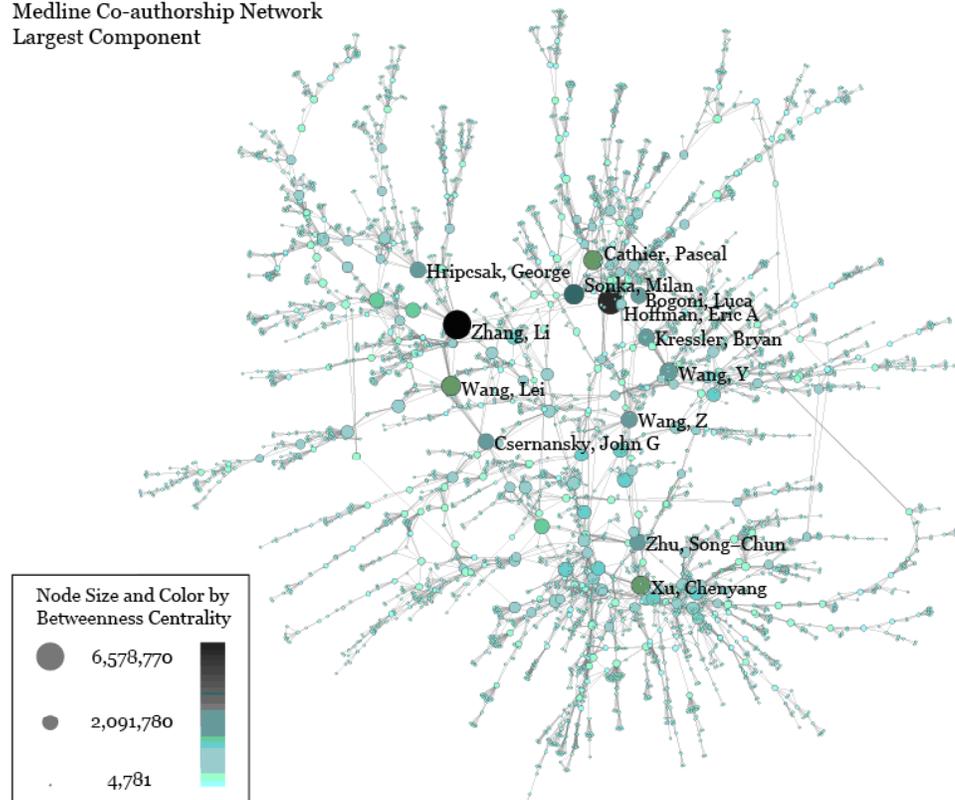
NSF Database:

NSF master table

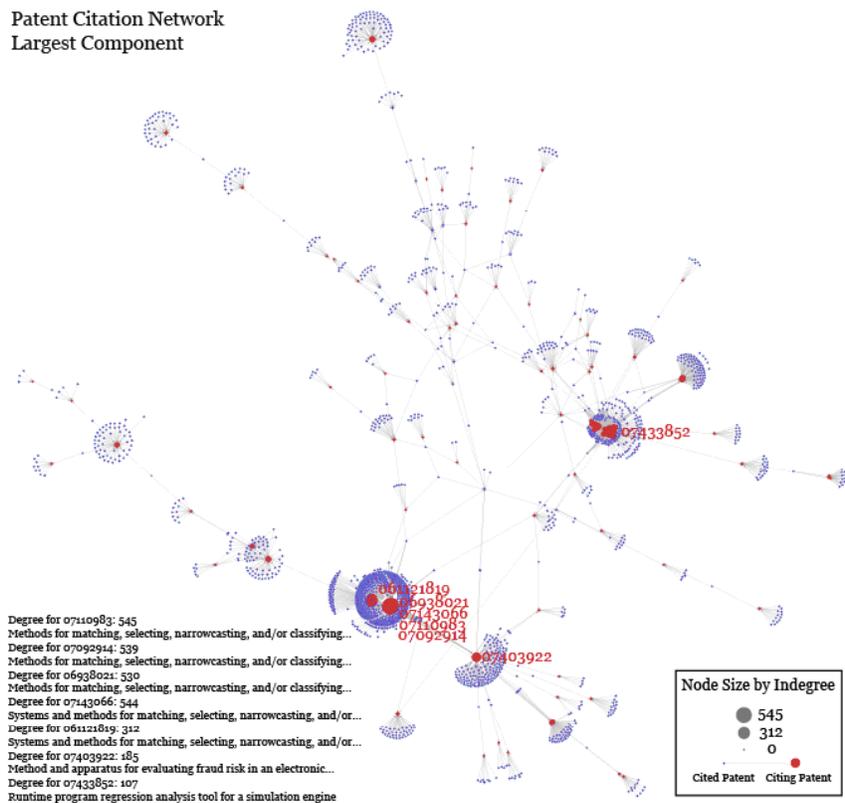
NSF co-investigator table (nwb format)

Download

Medline Co-authorship Network
Largest Component



**Patent Citation Network
Largest Component**



Degree for 0710983: 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 0743066: 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



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 **Macro**scope 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 resource environment for large-scale network analysis, modeling, and visualization. Thomson Scientific/ISI, Scopus and Google Scholar data, EndNote and Bibtex 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/>



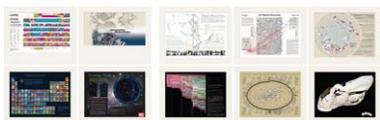
The Power of Maps (2005)



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>



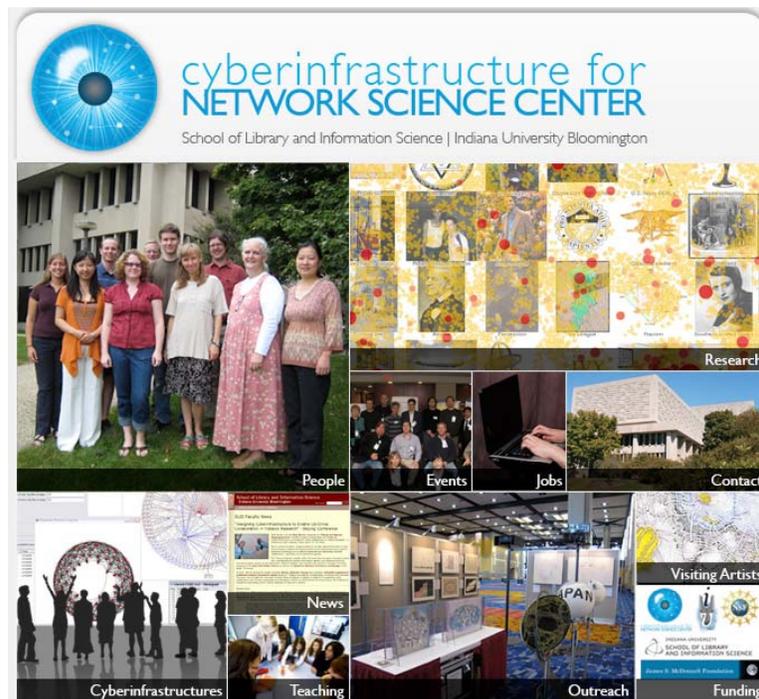
Science Maps in “Expedition Zukunft” science train visiting 62 cities in 7 months
12 coaches, 300 m long

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>