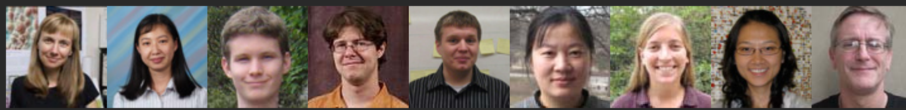


Plug-and-Play Macroscopes

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

Visualization for Collective, Connective & Distributed Intelligence
Dynamic Knowledge Networks ~ Synthetic Minds
Stanford University, CA: August 12, 2009

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 Fickr 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.

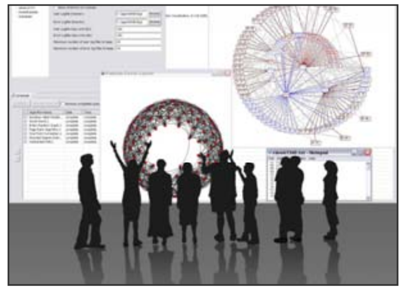
Static Instrument -> Evolving Cyberinfrastructure (CI): The importance of hardware instruments that are rather static and expensive decreases relative to software infrastructures that are highly flexible and continuously evolving according to the needs of different sciences. Some of the most successful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

Microscopes, Telescopes, and Macrocopes



Just as the **microscope** empowered our naked eyes to see cells, microbes, and viruses thereby advancing the progress of biology and medicine or the **telescope** opened our minds to the immensity of the cosmos and has prepared mankind for the conquest of space, **macroscopes** promise to help us cope with another infinite: the infinitely complex. Macroscopes give us a 'vision of the whole' and help us 'synthesize'. They let us detect patterns, trends, outliers, and access details in the landscape of science. Instead of making things larger or smaller, macroscopes let us observe what is at once too great, too slow, or too complex for our eyes.

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.

Network Workbench Tool

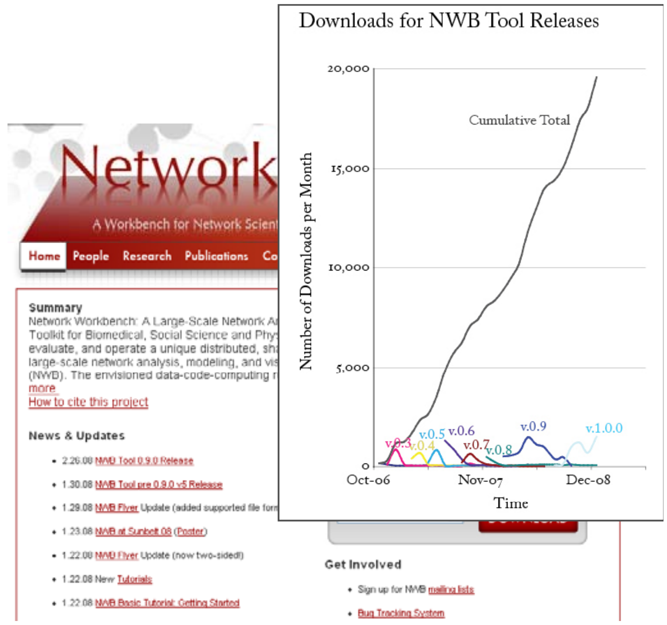
http://nwb.slis.indiana.edu

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 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.

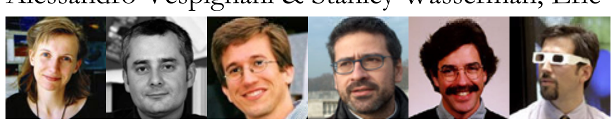
It has been downloaded more than 30,000 times since Dec. 2006.



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, Mihail C. (Eds.), Progress in Convergence - Technologies for Human Wellbeing (Vol. 1093, pp. 161-179), Annals of the New York Academy of Sciences, Boston, MA.

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

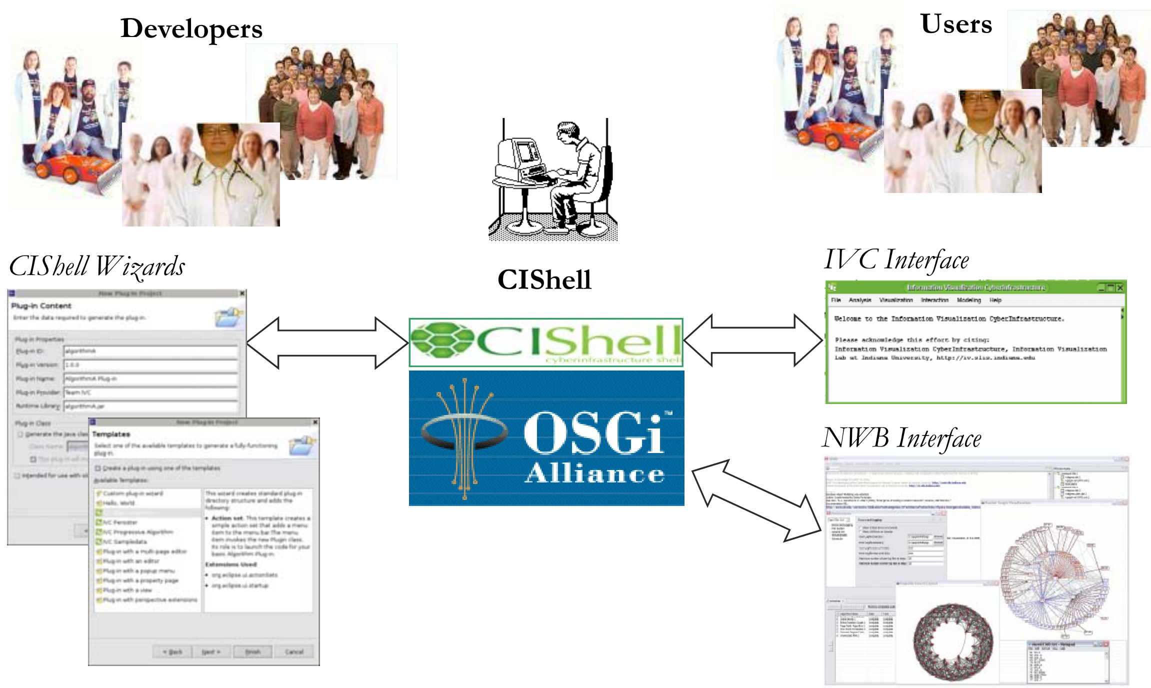
Goal: Develop a large-scale network analysis, modeling and visualization toolkit for physics, biomedical, and social science research.

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

Duration: Sept. 2005 - Aug. 2009

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

Serving Non-CS Algorithm Developers & Users



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)

NWB Tool: Algorithms (July 1st, 2008)

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

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 Degree?
  - 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 Toolkit?
- **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-Core?
    - Annotate K-Coreness?
- **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 Correlations?
  - Unnamed Category?
    - Page Rank
  - Based on local graph structure #2
    - Dyad Reciprocity?
    - Arc Reciprocity?

tion Edit

- **Tools**
  - GUESS
  - GnuPlot?
- **Predefined Positions Layout**
  - DrL (VxOrd)
  - Pre-defined Positions (prefuse beta)?
- **Move**
  - Circular
- **Tree Layouts**
  - Radial Tree (prefuse alpha)
  - Radial Tree with Annotations (prefuse beta)?
  - Tree Map
  - Tree View
  - Balloon Graph (prefuse alpha)?
- **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)?
  - LaNet (k-Core Decomposition)

etrics Edit

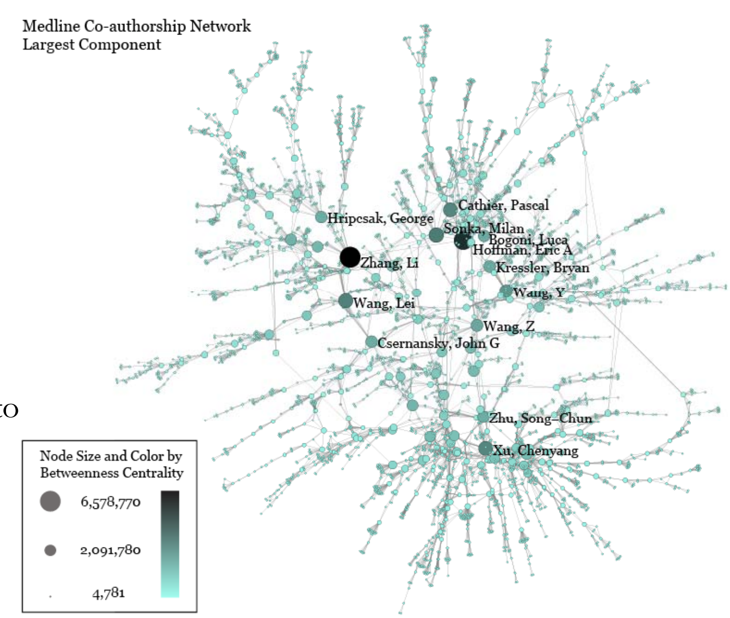
- **Extract Network From Table**
  - Extract Co-Authorship Network
  - Extract Co-Occurrence Network From Table?
  - Extract Directed Network From Table?
- **Extract Network From Another Network**
  - Extract Bibliographic Coupling Similarity Network
  - Extract Co-Citation Similarity Network?
- **Cleaning**
  - Remove ISI Duplicate Records

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.



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.

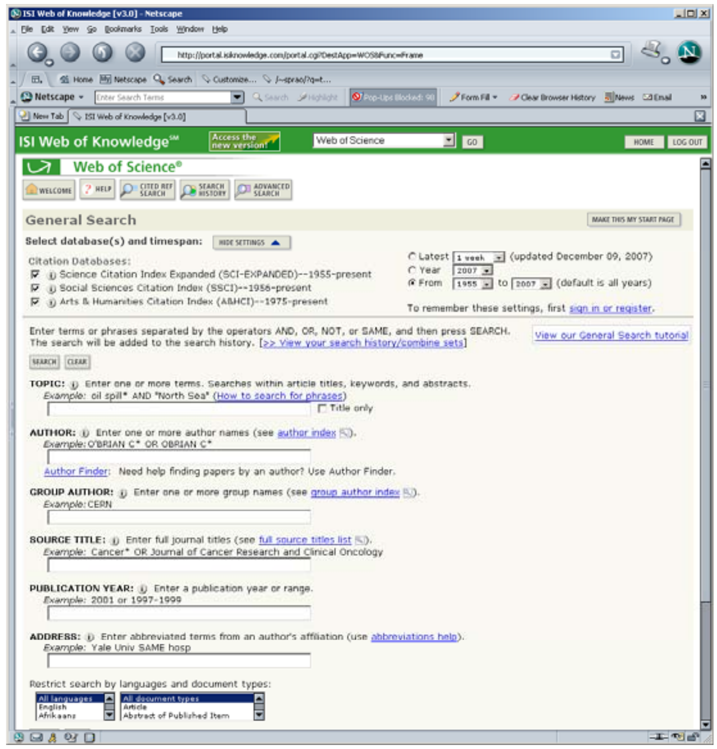
Scientific Field Level

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

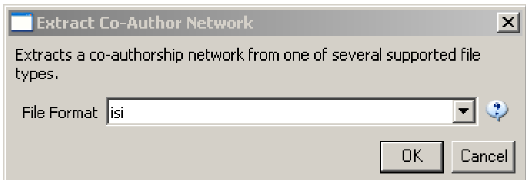
No books and other non-WoS publications are covered.

	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	*(Dec 2007)*
	41	16,920	159	52	*(Dec 2008)*

Extract Co-Author Network

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

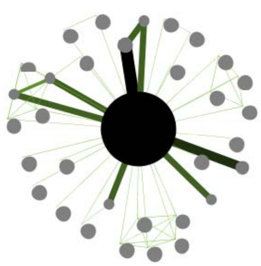
To extract the co-author network, select the *'361 Unique ISI Records'* table and run *'Scientometrics > Extract Co-Author Network'* using isi file format:



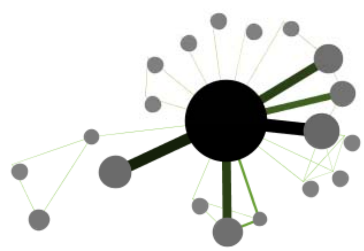
The result is an undirected network of co-authors in the Data Manager. It has 247 nodes and 891 edges.

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*.

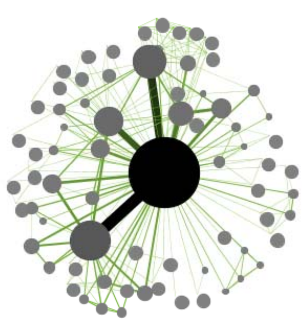
Comparison of Co-Author Networks



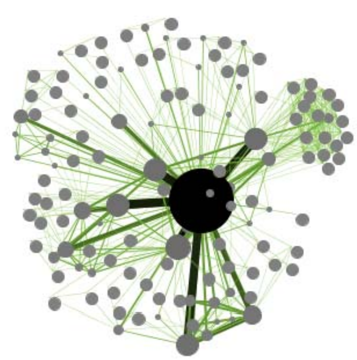
Eugene Garfield



Stanley Wasserman

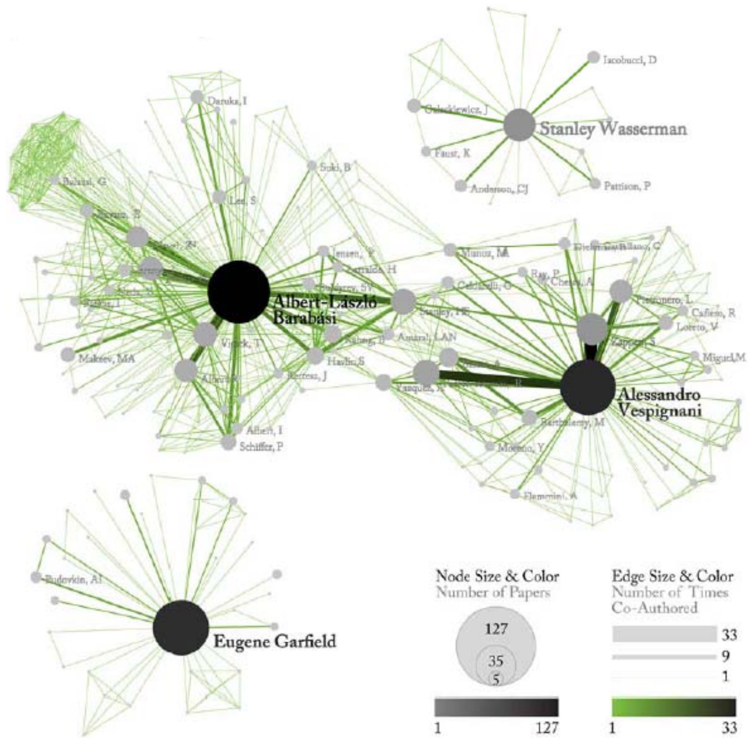


Alessandro Vespignani



Albert-László Barabási

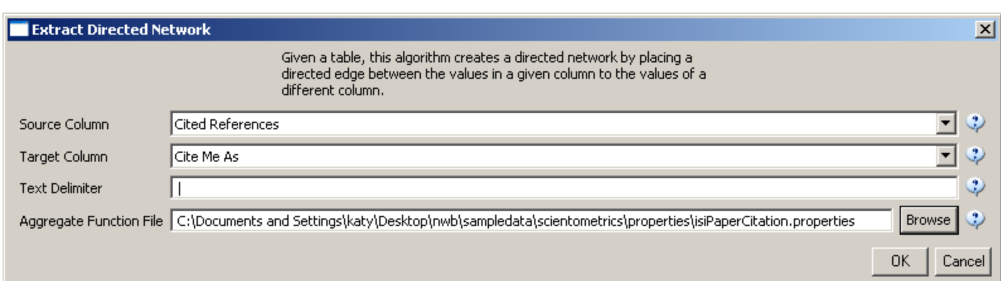
Joint Co-Author Network of all Four NetsSci Researchers



Paper-Citation Network Layout

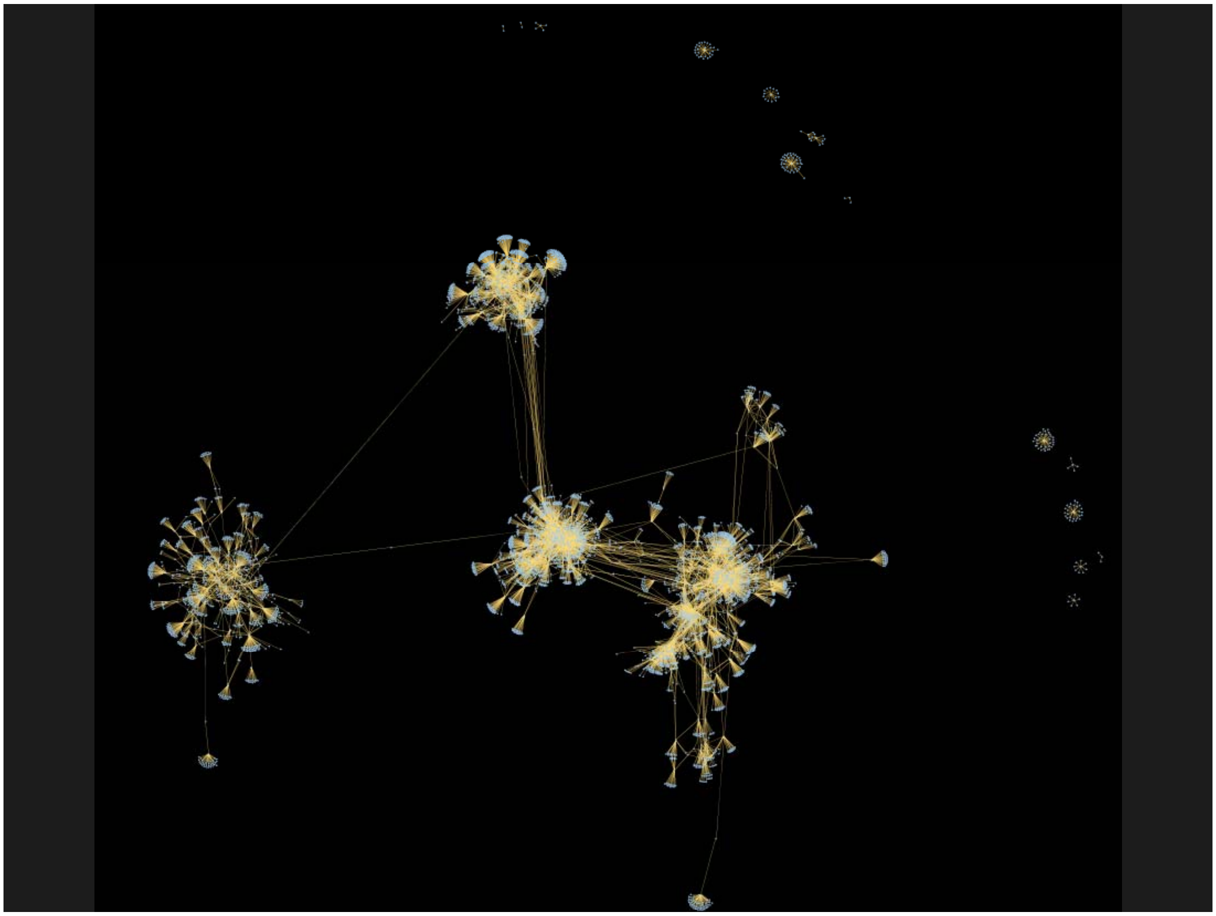
Load *'*yournwbdirectory*/sampledata/scientometrics/isi/FourNetSciResearchers.isi'* using *'File > Load and Clean ISI File'*.

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



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

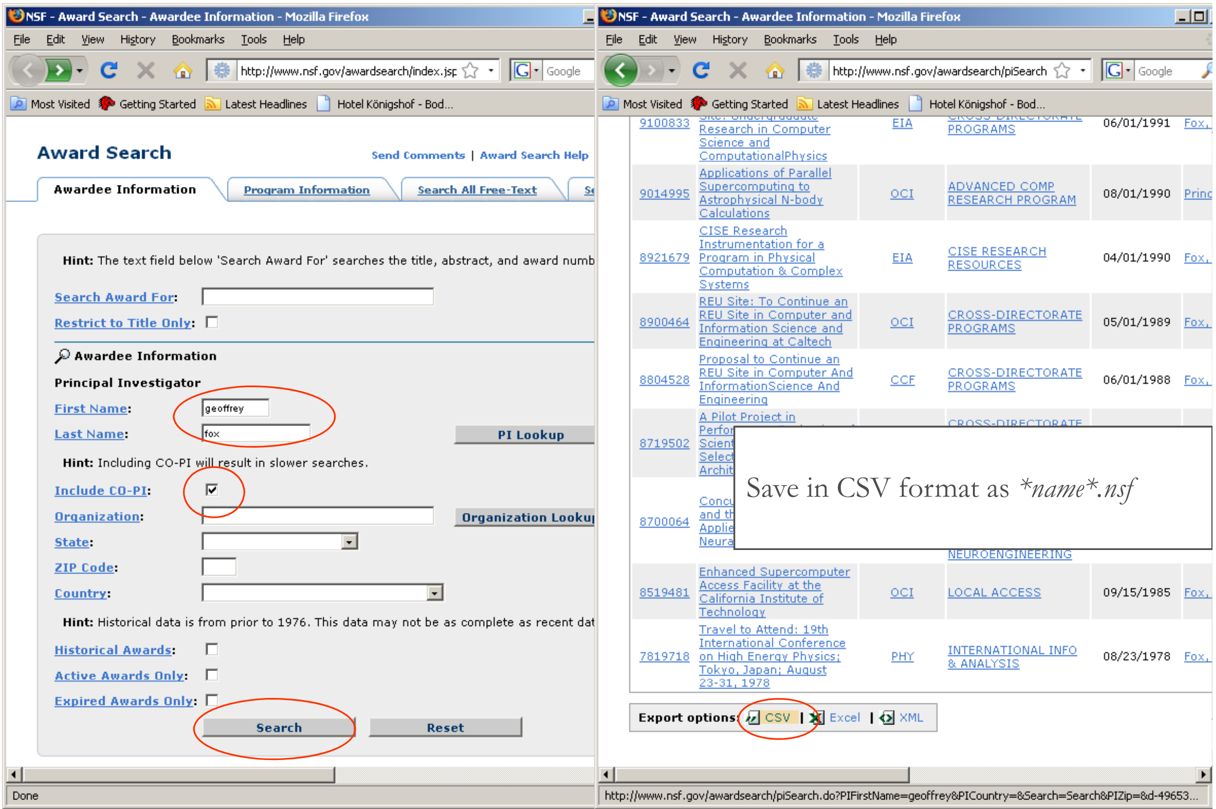
Institution Level

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



NSF Awards Search Results

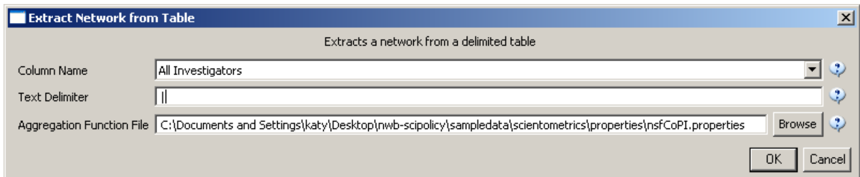
Name	# Awards	First A. Starts	Total Amount to Date
Geoffrey Fox	27	Aug 1978	12,196,260
Michael McRobbie	8	July 1997	19,611,178
Beth Plale	10	Aug 2005	7,224,522

Disclaimer:

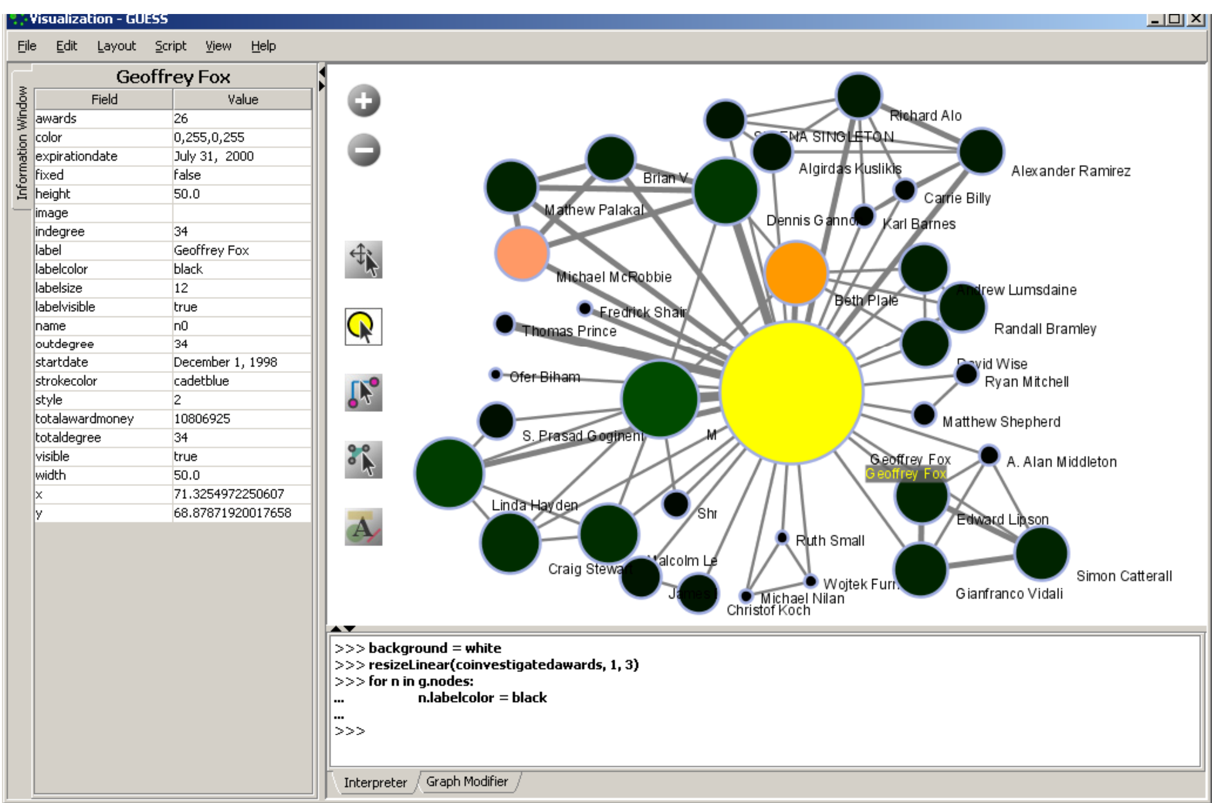
Only NSF funding, no funding in which they were senior personnel, only as good as NSF's internal record keeping and unique person ID. If there are 'collaborative' awards then only their portion of the project (award) will be included.

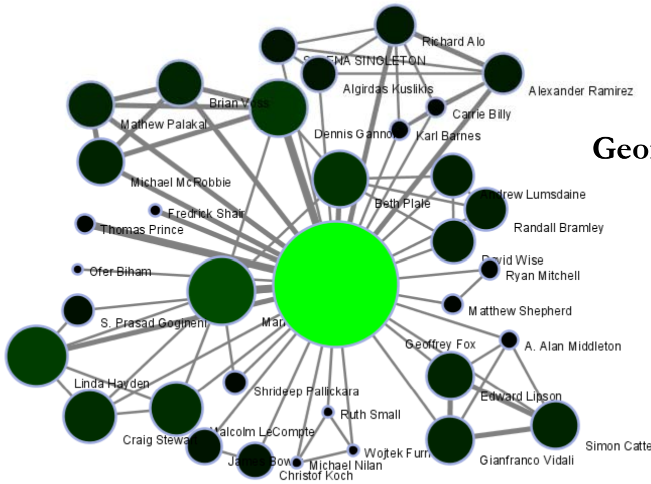
Using NWB to Extract Co-PI Networks

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

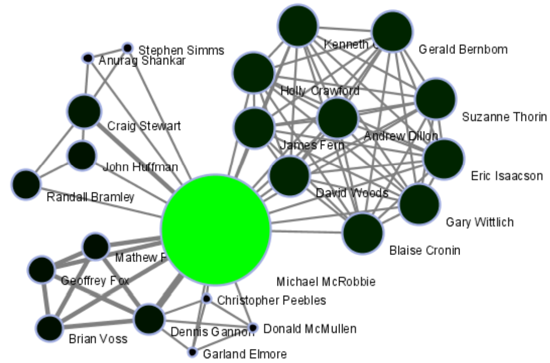


- Select *“Extracted Network ..”* and run *‘Analysis > Network Analysis Toolkit (NAT)’*
- Remove unconnected nodes via *‘Preprocessing > Delete Isolates’*.
- *‘Visualization > GUESS’*, layout with GEM
- Run ‘co-PI-nw.py’ GUESS script to color/size code.

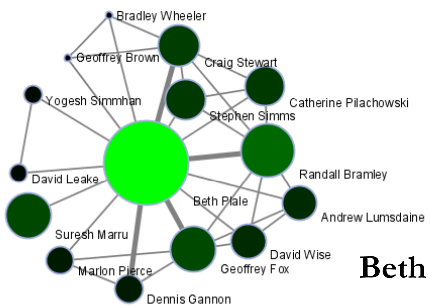




Geoffrey Fox



Michael McRobbie



Beth Plale

Geoffrey Fox

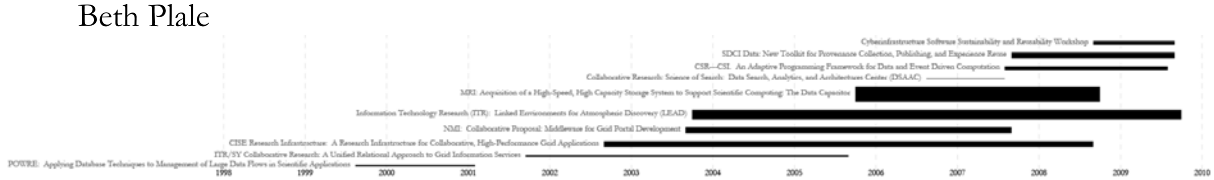
Last Expiration date

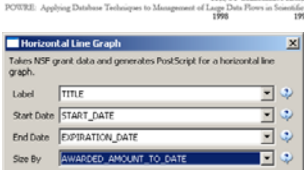
July 10

Michael McRobbie

Feb 10

Beth Plale





Sept 09

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

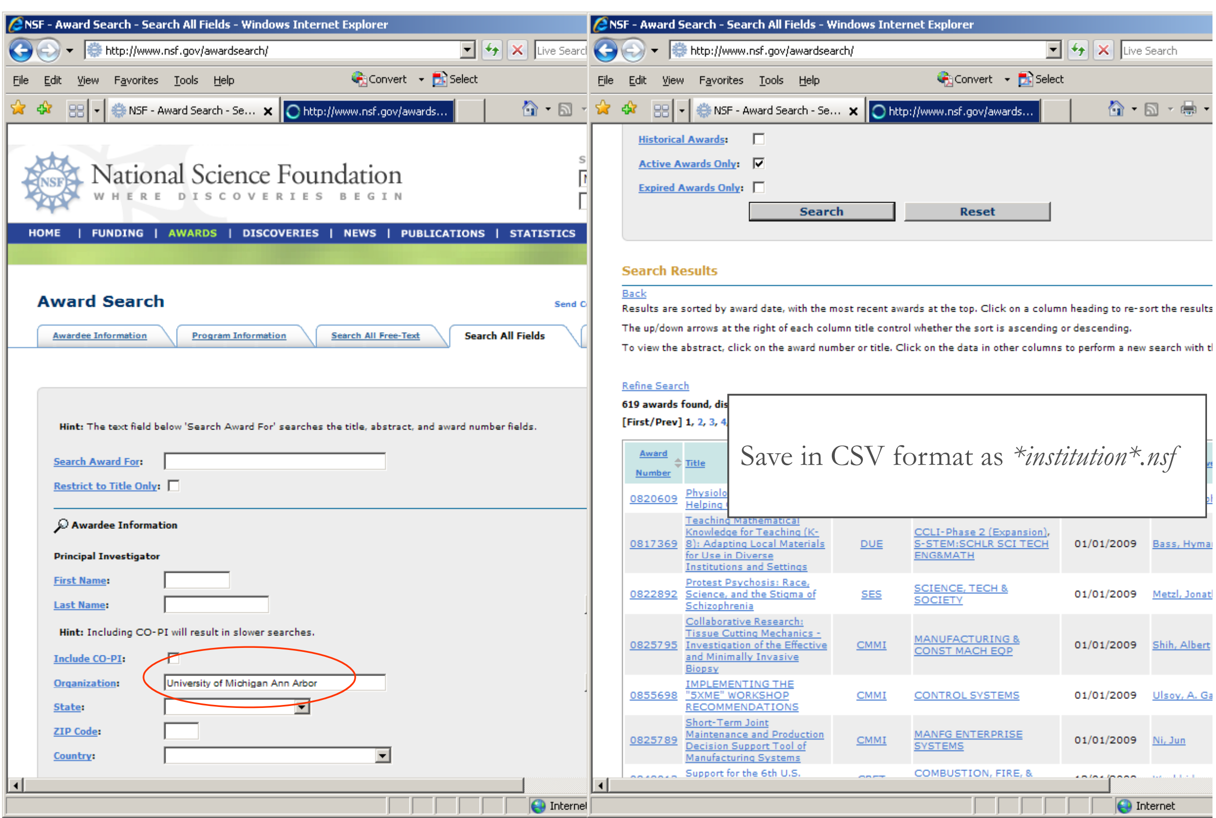
Institution Level

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



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)

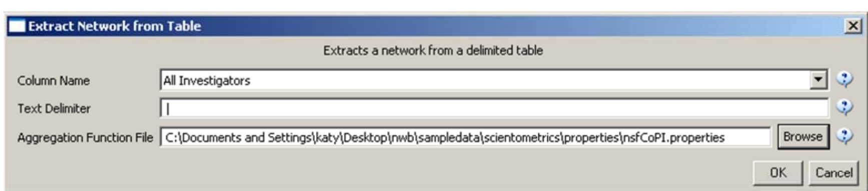
Active NSF Awards on 09/10/2009:

- 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

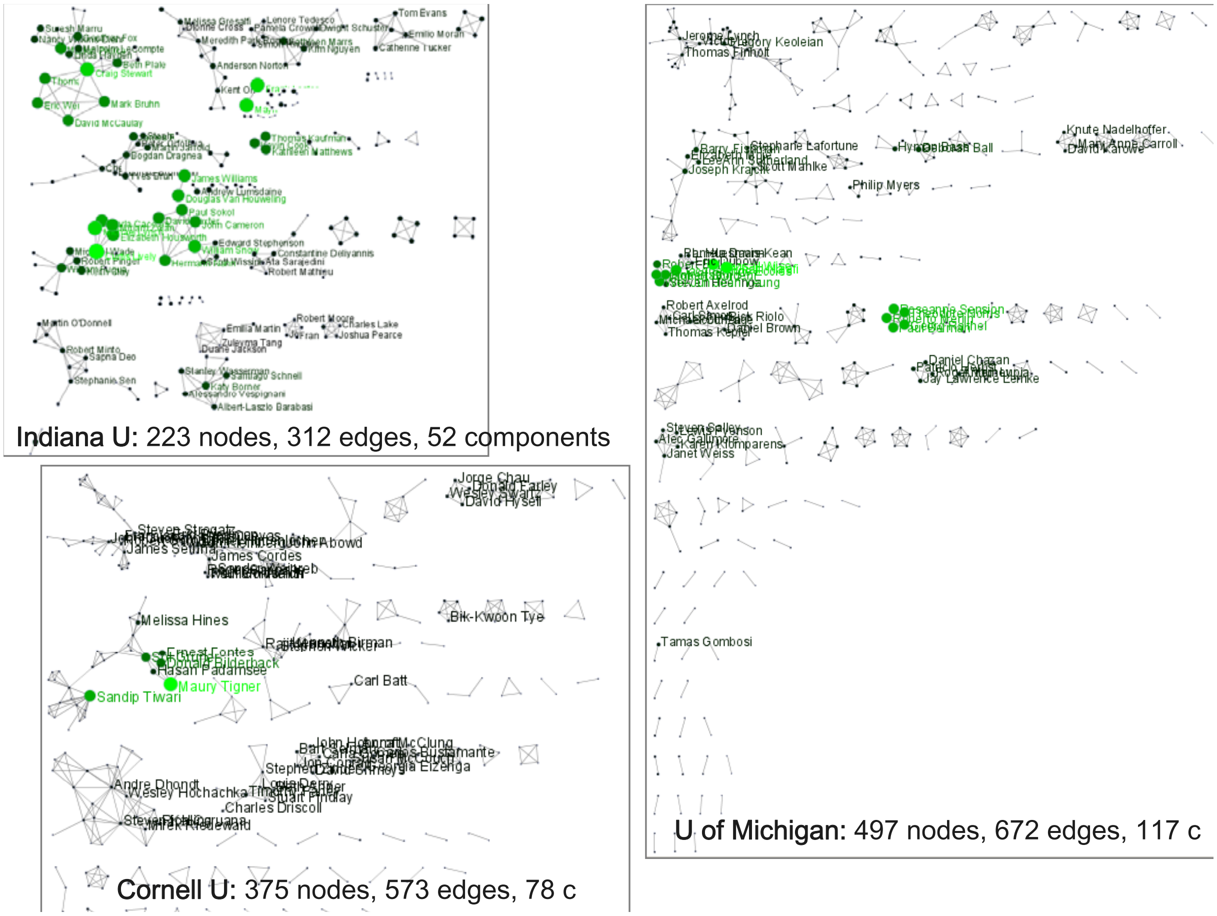
Load NSF data, selecting the loaded dataset in the Data Manager window, run *'Scientometrics > Extract Co-Occurrence Network'* using parameters:



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.

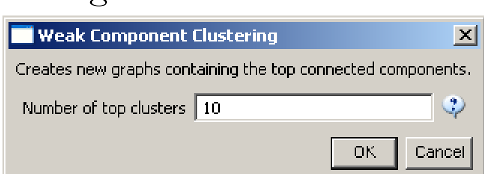
Running the *'Analysis > Network Analysis Toolkit (NAT)'* reveals that the number of nodes and edges but also of isolate nodes that can be removed running *'Preprocessing > Delete Isolates'*.

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



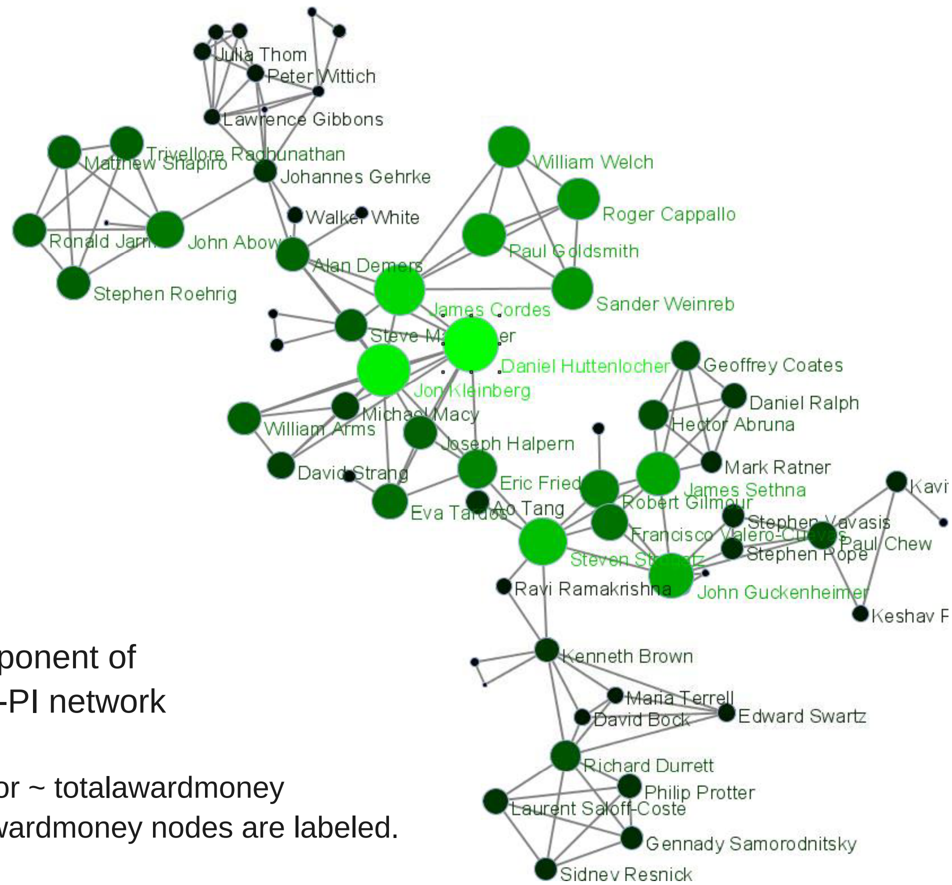
Extract Giant Component

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.

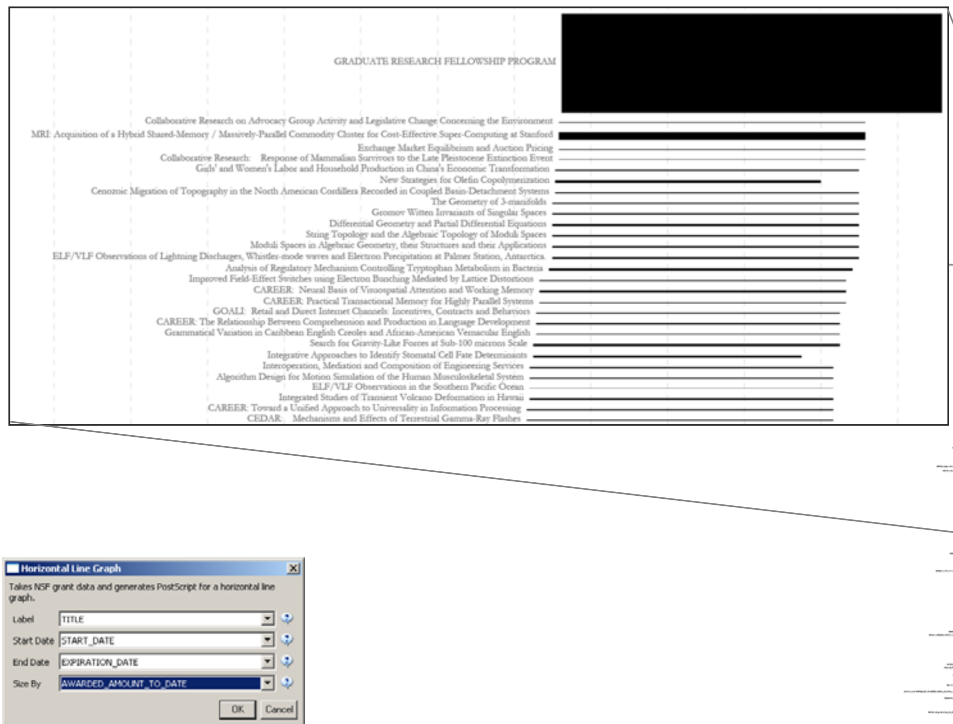
Top-10 Investigators by Total Award Money

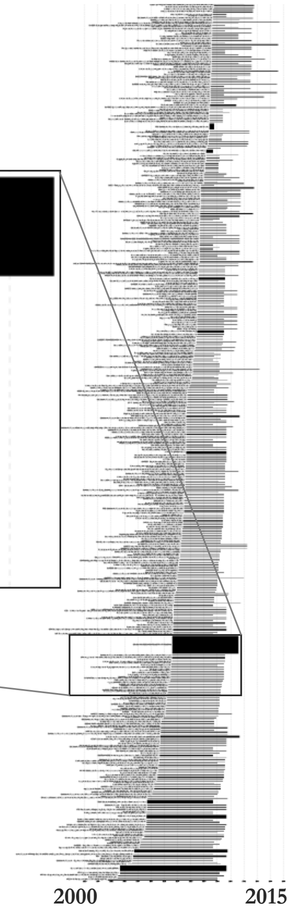
```
for i in range(0, 10):
    print str(nodesbytotalawardmoney[i].label) + ": " +
    str(nodesbytotalawardmoney[i].totalawardmoney)
```

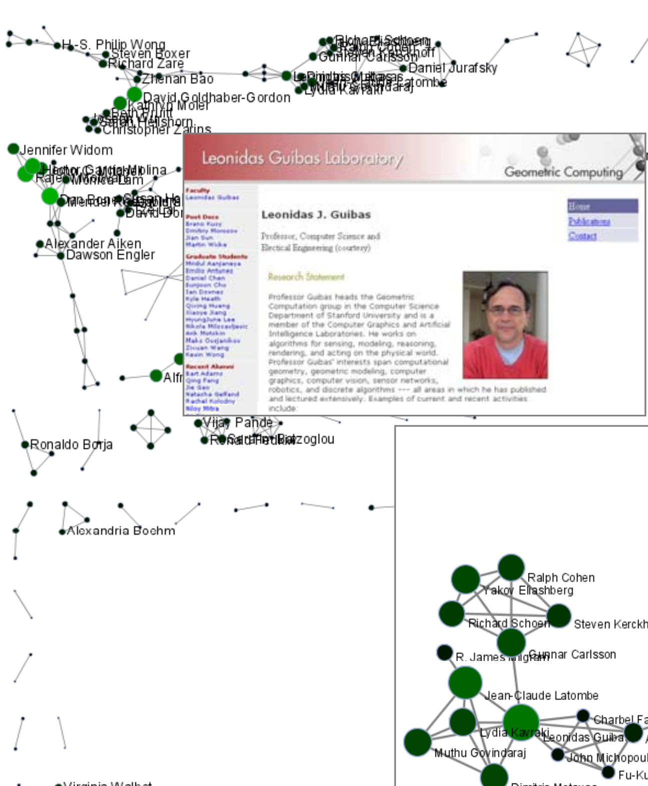
Indiana University		Cornell University		Michigan University	
Curtis Lively:	7,436,828	Maury Tigner:	107,216,976	Khalil Najafi:	32,541,158
Frank Lester:	6,402,330	Sandip Tiwari:	72,094,578	Kensall Wise:	32,164,404
Maynard Thompson:	6,402,330	Sol Gruner:	48,469,991	Jacquelynne Eccles:	25,890,711
Michael Lynch:	6,361,796	Donald Bilderback:	47,360,053	Georg Raithel:	23,832,421
Craig Stewart:	6,216,352	Ernest Fontes:	29,380,053	Roseanne Sension:	23,812,921
William Snow:	5,434,796	Hasan Padamsee:	18,292,000	Theodore Norris:	23,35,0921
Douglas V. Houweling:	5,068,122	Melissa Hines:	13,099,545	Paul Berman:	23,350,921
James Williams:	5,068,122	Daniel Huttenlocher:	7,614,326	Roberto Merlin:	23,350,921
Miriam Zolan:	5,000,627	Timothy Fahey:	7,223,112	Robert Schoeni:	21,991,140
Carla Caceres:	5,000,627	Jon Kleinberg:	7,165,507	Wei-Jun Jean Yeung:	21,991,140

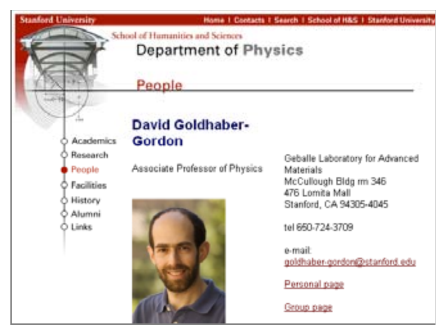
Stanford University

429 active NSF awards on 09/10/2009

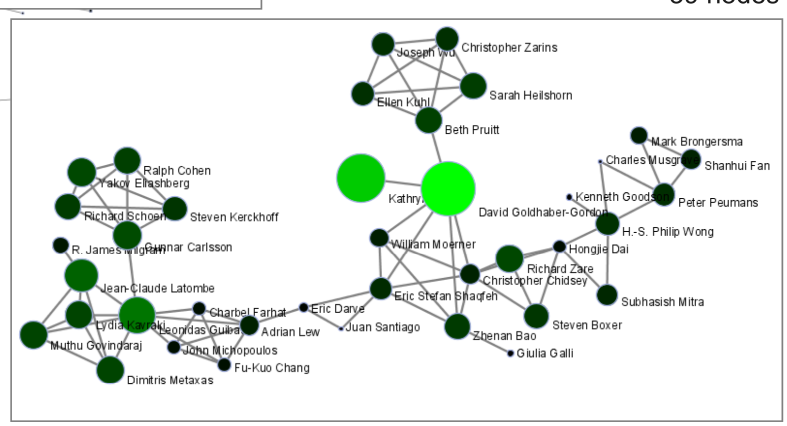








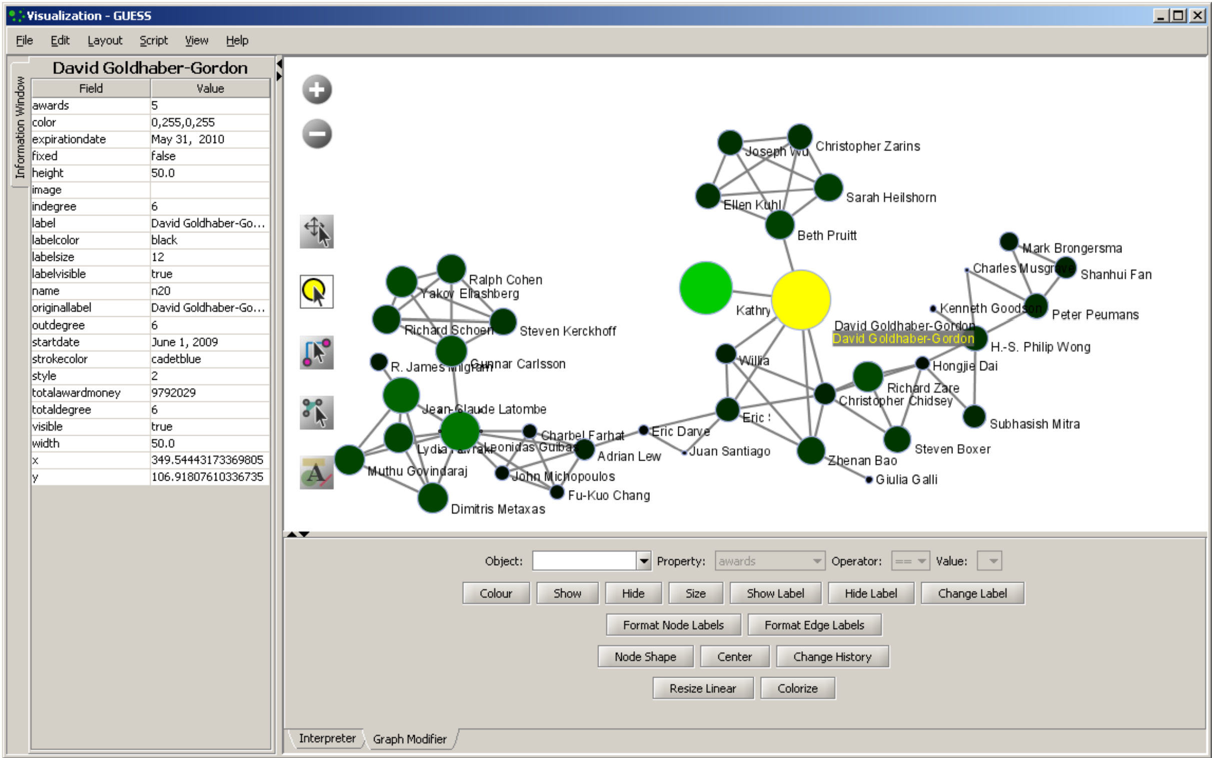
Largest component
39 nodes



Stanford U:

218 nodes, 285 edges, 49 components

157 isolate nodes were deleted

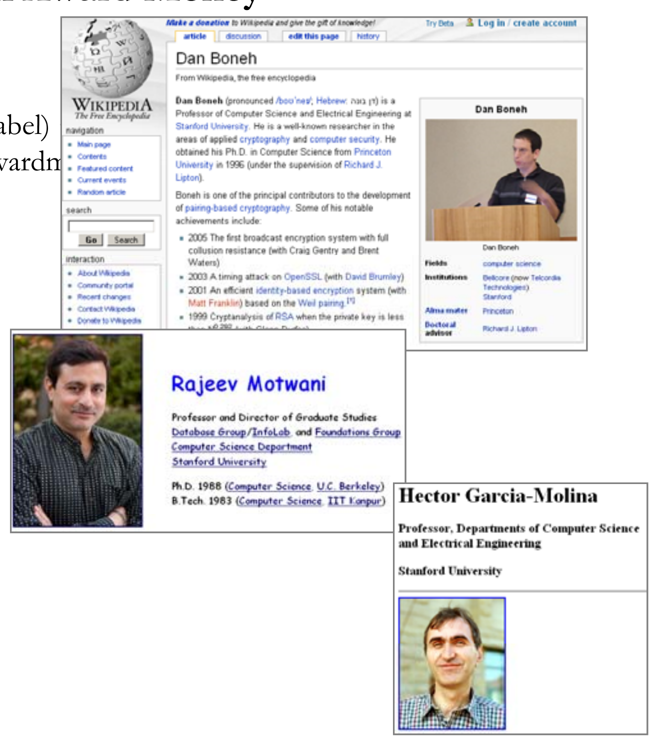


Top-10 Investigators by Total Award Money

```
for i in range(0, 10):
    print str(nodesbytotalawardmoney[i].label)
    str(nodesbytotalawardmoney[i].totalawardm
```

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



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.

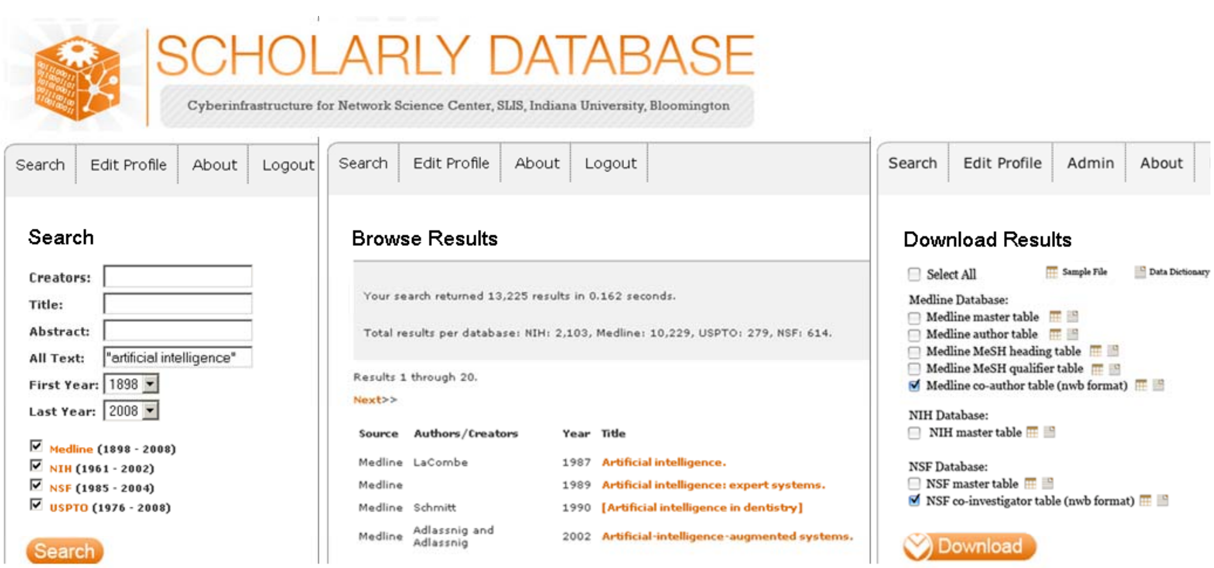
B. Loading NSF datasets with currently active NSF funding for 3 researchers at Indiana U

Institution Level

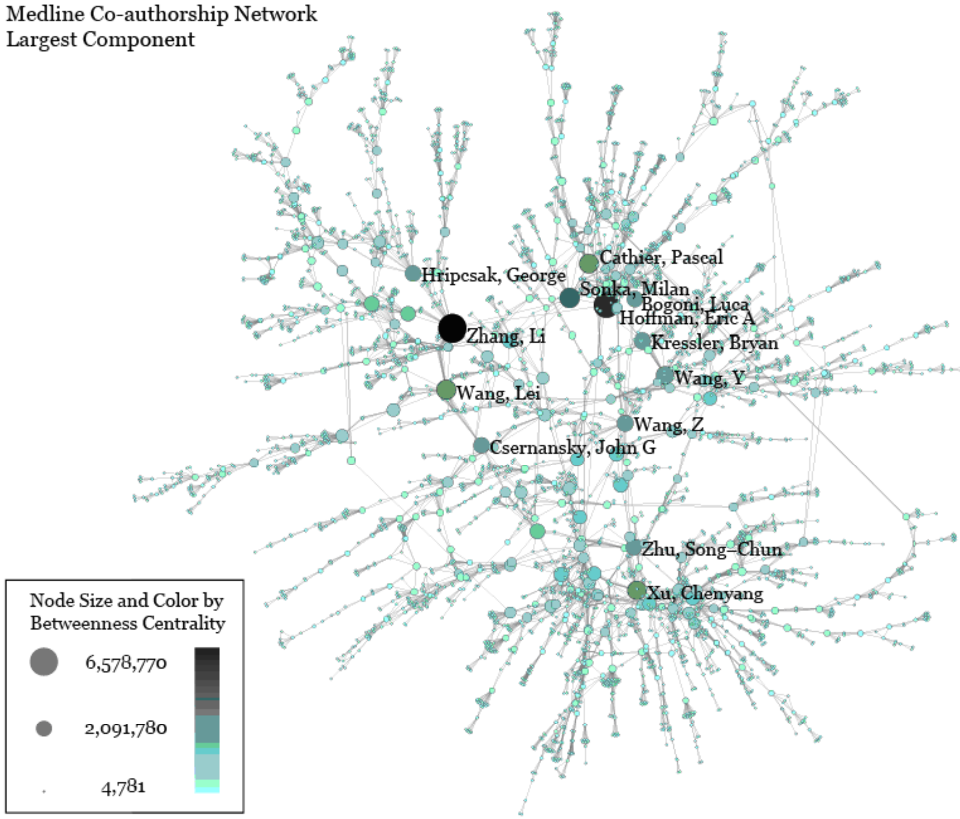
C. Indiana U, Cornell U, and Michigan U, extracting, and comparing Co-PI networks.

Scientific Field Level

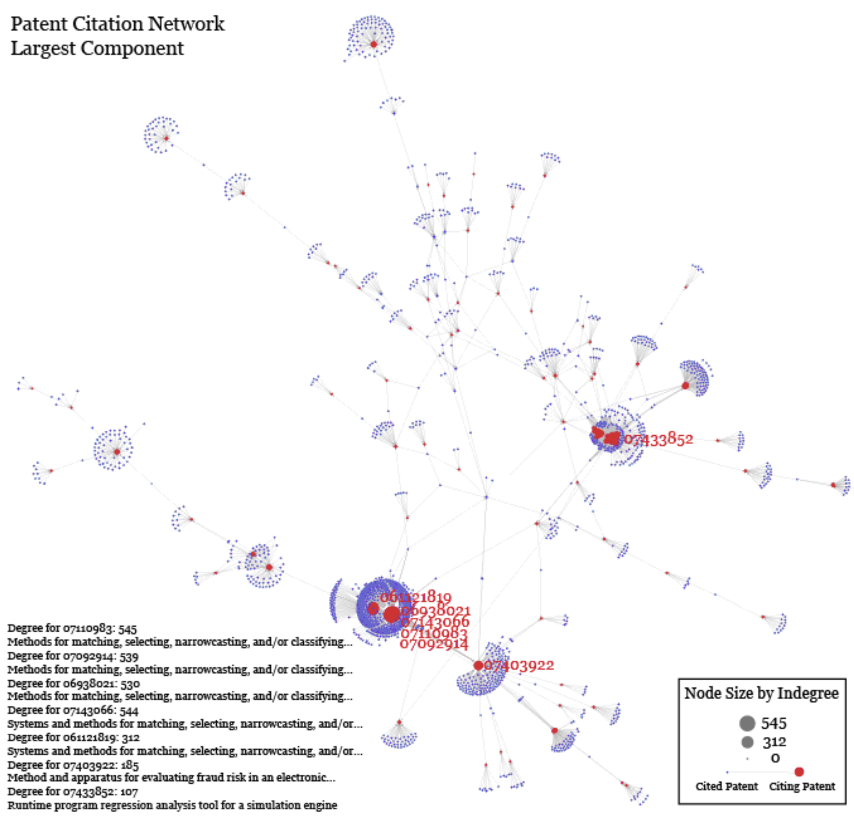
D. Extracting co-author networks, patent-citation networks, and detecting bursts in SDB data.



Medline Co-authorship Network
Largest Component



Patent Citation Network
Largest Component



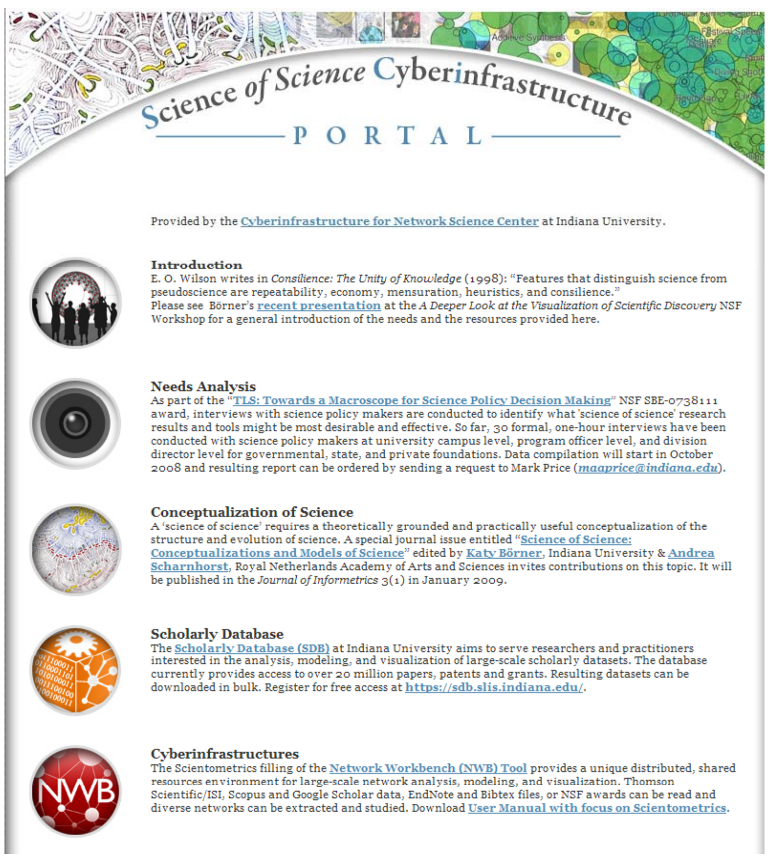
Top-10 burst terms from abstracts of the AI search results.

Medline				
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

NIH				
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

NSF				
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

USPTO				
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
PORTAL

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 Macroscope 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 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

Macrosope Outlook

CIShell/OSGi is at the core of different CIs and a total of 169 unique plugins are used in the
Information Visualization (http://iv.slis.indiana.edu),
Network Science (http://nwb.slis.indiana.edu),
Science Policy (http://sci.slis.indiana.edu), and
Epidemics (http://epic.slis.indiana.edu) research communities.

Most interestingly, a number of other projects recently adopted OSGi and one adopted CIShell:

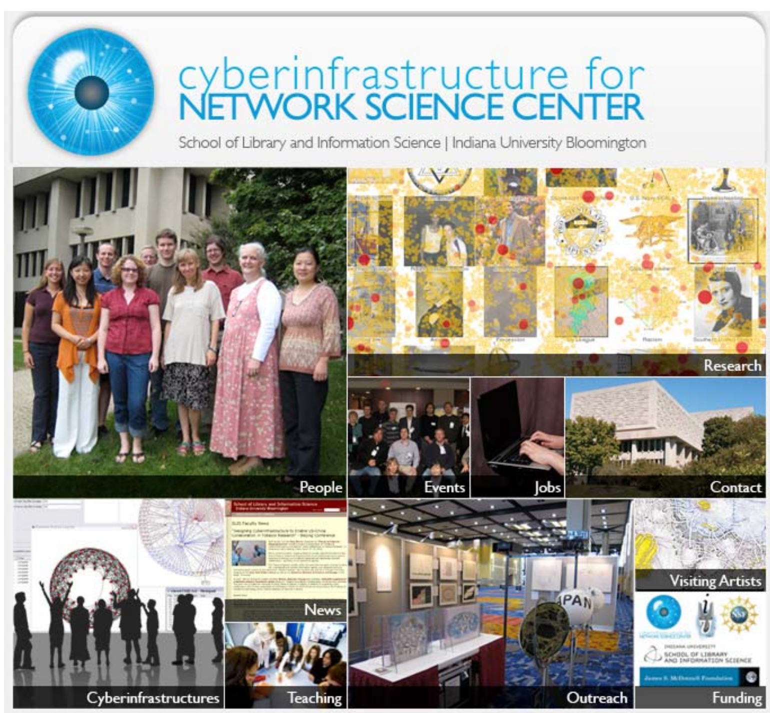
Cytoscape (http://www.cytoscape.org) lead by Trey Ideker, UCSD is an open source bioinformatics software platform for visualizing molecular interaction networks and integrating these interactions with gene expression profiles and other state data (Shannon et al., 2002).

Taverna Workbench (http://taverna.sourceforge.net) lead by Carol Goble, University of Manchester, UK is a free software tool for designing and executing workflows (Hull et al., 2006). Taverna allows users to integrate many different software tools, including over 30,000 web services.

MAEviz (https://wiki.ncsa.uiuc.edu/display/MAE/Home) managed by Shawn Hampton, NCSA is an open-source, extensible software platform which supports seismic risk assessment based on the Mid-America Earthquake (MAE) Center research.

TEXTrend (http://www.textrend.org) lead by George Kampis, Eötvös University, Hungary develops a framework for the easy and flexible integration, configuration, and extension of plugin-based components in support of natural language processing (NLP), classification/mining, and graph algorithms for the analysis of business and governmental text corpuses with an inherently temporal component.

As the functionality of OSGi-based software frameworks improves and the number and diversity of dataset and algorithm plugins increases, the capabilities of custom tools or macroscopes will expand.



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