

Towards a Macroscopic Science Policy Decision Making

Dr. Katy Börner

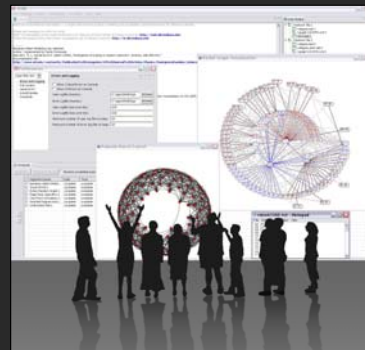
Cyberinfrastructure for Network Science Center, Director
Information Visualization Laboratory, Director
School of Library and Information Science
Indiana University, Bloomington, IN
katy@indiana.edu

Kevin Boyack

SciTech Strategies Inc.
kboyack@mapofscience.com

With special thanks to the members at the Cyberinfrastructure for Network Science Center, Mapping Science exhibit map makers and advisory board members, and the VIVO team.

October 19, 2010
AAAS



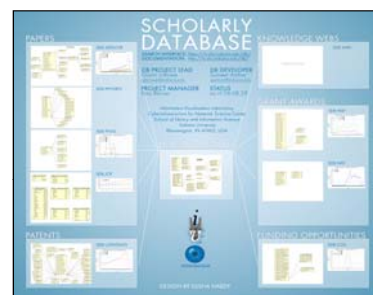
Computational Scientometrics Cyberinfrastructures



Scholarly Database: 25 million scholarly records
<http://sdb.slis.indiana.edu>



VIVO Research Networking
<http://vivoweb.org>



Information Visualization Cyberinfrastructure
<http://iv.slis.indiana.edu>



Network Workbench Tool & Community Wiki
<http://nwb.slis.indiana.edu>



Science of Science (Sci²) Tool and CI Portal
<http://sci.slis.indiana.edu>



Epidemics Cyberinfrastructure
<http://epic.slis.indiana.edu/>



Sci² Tool
A tool for science of science research & practice

Email Address

Password

Login

Forgot your password?
To recover your account password, please visit our [password recovery page](#).

Not registered yet?
[Register now](#)

Tutorials
Katy Börner (2010) Science of Science Research and Tools (12 Tutorials). Reporting Branch, Office of Extramural Research/Office of the Director, National Institutes of Health, Bethesda, MD.

- Tutorial #01: [Science of Science Research](#)
- Tutorial #02: [Network Science / Information Visualization](#)
- Tutorial #03: [CIShell Powered Tools: Network Workbench and Science of Science Tool](#)
- Tutorial #04: [Temporal Analysis—Burst Detection](#)
- Tutorial #05: [Geospatial Analysis and Mapping](#)
- Tutorial #06: [Topical Analysis & Mapping](#)
- Tutorial #07: [Tree Analysis and Visualization](#)
- Tutorial #08: [Network Analysis and Visualization](#)
- Tutorial #09: [Large Network Analysis and Visualization](#)
- Tutorial #10: [Using the Scholarly Database at IU](#)
- Tutorial #11: [VIVO National Researcher Networking](#)
- Tutorial #12: [Future Developments](#)

<http://sci.slis.indiana.edu/sci2>

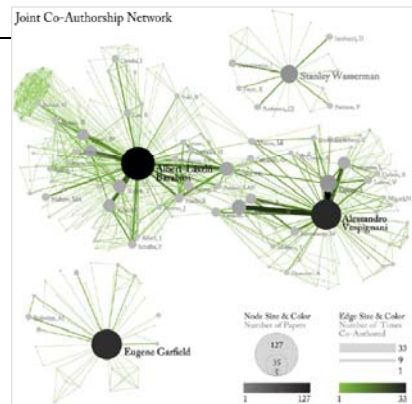
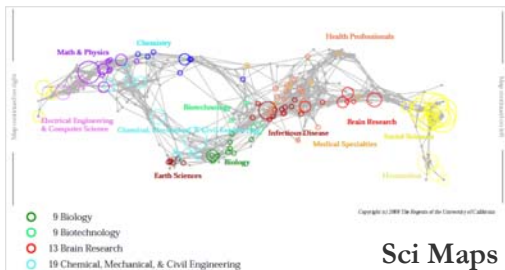
Geetha Senthil (2010). [Multidisciplinary Nature of Work With Reference to PIs and ICs Within a Portfolio](#). PA Group at NIH.

NIH Office of Extramural Research and Katy Börner (2010) [Network Visualizations Using SPIRES Data and the Sci2 Tool](#). Office of Extramural Research at NIH.



Sci² Tool – “Open Code for S&T Assessment”

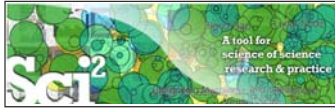
OSGi/CIShell powered tool with NWB plugins and many new scientometrics and visualizations plugins.



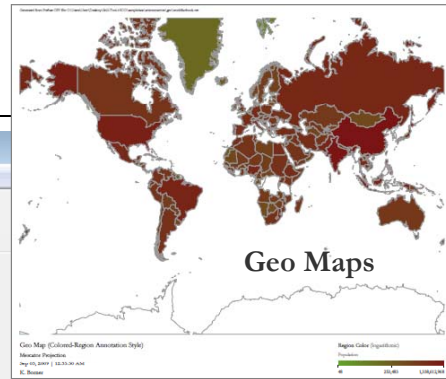
Horizontal Time Graphs



Börner, Katy, Huang, Weixia (Bonnie), Linnemeier, Micah, Dubon, Russell Jackson, Phillips, Patrick, Ma, Nianli, Zoss, Angela, Guo, Hanning & Price, Mark. (2009). *ReTe-Netzwerk-Red: Analyzing and Visualizing Scholarly Networks Using the Scholarly Database and the Network Workbench Tool*. *Proceedings of IS² 2009: 12th International Conference on Scientometrics and Informetrics, Rio de Janeiro, Brazil, July 14-17*. Vol. 2, pp. 619-630.



Sci² Tool



Sci² Tool

File Preprocessing Modeling Analysis Visualization Scientometrics Help

Console

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

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

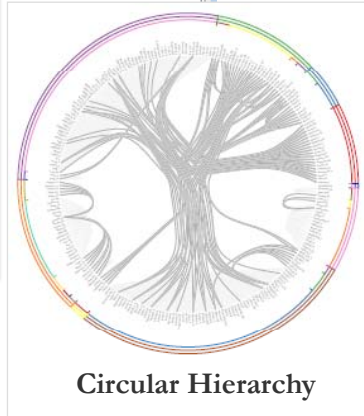
Please cite as follows:
Sci² Team. (2009). Science of Science Tool. In SciTech Strategies Inc., <http://sci.slis.indiana.edu>.

Scheduler

Remove From List Remove completed

!	Algorithm Name	Date	Time	% Comp
<input checked="" type="checkbox"/>	Extract Co-Author Netw...	09/03/2009	00:15:20 AM	100%
<input checked="" type="checkbox"/>	Load and Clean ISI File	09/03/2009	00:15:05 AM	100%

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



Sci² Tool: Algorithms

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

Preprocessing

- Extract Top N% Records
- Extract Top N Records
- Normalize Text
- Slice Table by Line

- Extract Top Nodes
- Extract Nodes Above or Below Value
- Delete Isolates

- Extract top Edges
- Extract Edges Above or Below Value
- Remove Self Loops
- Trim by Degree
- MST-Pathfinder Network Scaling
- Fast Pathfinder Network Scaling

- Snowball Sampling (in nodes)
- Node Sampling
- Edge Sampling

- Symmetrize
- Dichotomize
- Multipartite Joining

- Geocoder
- Extract ZIP Code

Modeling

- Random Graph
- Watts-Strogatz
- Small World
- Barabási-Albert Scale-Free
- TARL

Analysis

- Network Analysis Toolkit (NAT)
- Unweighted & Undirected
 - Node Degree
 - Degree Distribution
- K-Nearest Neighbor (Java)
- Watts-Strogatz Clustering Coefficient
- Watts Strogatz Clustering Coefficient over K
- Diameter
- Average Shortest Path
- Shortest Path Distribution
- Node Betweenness Centrality
- Weak Component Clustering
- Global Connected Components
- Extract K-Core
- Annotate K-Coreness
- HITS

Weighted & Undirected

- Clustering Coefficient
- Nearest Neighbor Degree
- Strength vs Degree
- Degree & Strength
- Average Weight vs End-point Degree
- Strength Distribution
- Weight Distribution
- Randomize Weights

Blondel Community Detection

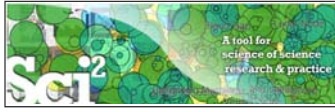
HITS

- Unweighted & Directed
 - Node Indegree
 - Node Outdegree
 - Indegree Distribution
 - Outdegree Distribution

K-Nearest Neighbor Single Node in-Out Degree Correlations

- Dyad Reciprocity
- Arc Reciprocity
- Adjacency Transitivity

- Weak Component Clustering
- Strong Component Clustering



Sci² Tool: Algorithms cont.

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

 Extract K-Core
 Annotate K-Core-ness

HITS
 PageRank
 Weighted & Directed
 HITS
 Weighted PageRank

Textual
 Burst Detection

Visualization

GnuPlot
 GUESS
 Image Viewer

 Radial Tree/Graph (prefuse alpha)
 Radial Tree/Graph with Annotation (prefuse beta)
 Tree View (prefuse beta)
 Tree Map (prefuse beta)
 Force Directed with Annotation (prefuse beta)
 Fruchterman-Reingold with Annotation (prefuse beta)

 DrL (VxOrd)
 Specified (prefuse beta)

Horizontal Line Graph
Circular Hierarchy
Geo Map (Circle Annotation Style)
Geo Map (Colored-Region Annotation Style)
Science Map (Circle Annotation)

Scientometrics

Remove ISI Duplicate Records
 Remove Rows with Multitudinous Fields
 Detect Duplicate Nodes
 Update Network by Merging Nodes

Extract Directed Network

Extract Paper Citation Network
 Extract Author Paper Network

Extract Co-Occurrence Network

Extract Word Co-Occurrence Network
 Extract Co-Author Network
 Extract Reference Co-Occurrence (Bibliographic Coupling) Network

 Extract Document Co-Citation Network

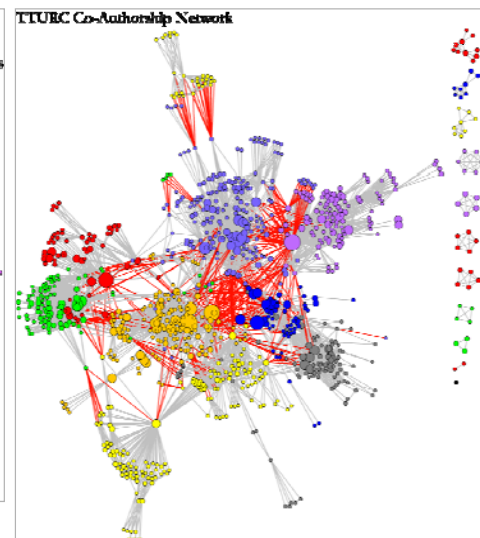
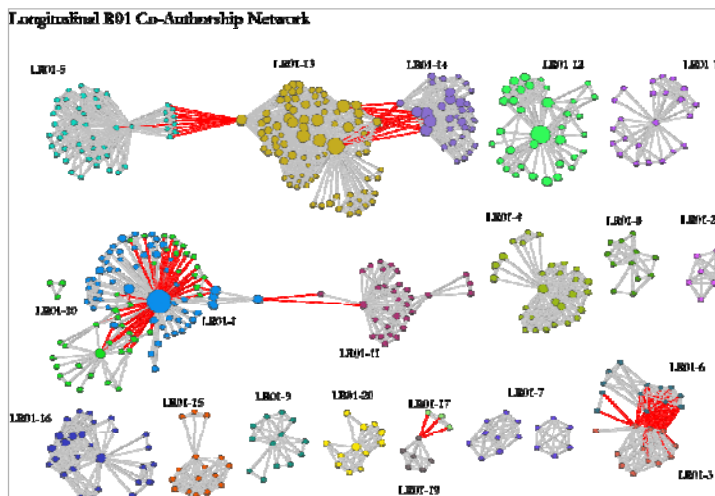
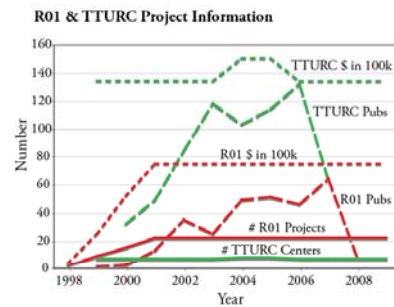
NEW:
 Database support for ISI and NSF data.

7

Mapping Transdisciplinary Tobacco Use Research Centers Publications

Compare R01 investigator based funding with TTURC Center awards in terms of number of publications and evolving co-author networks.

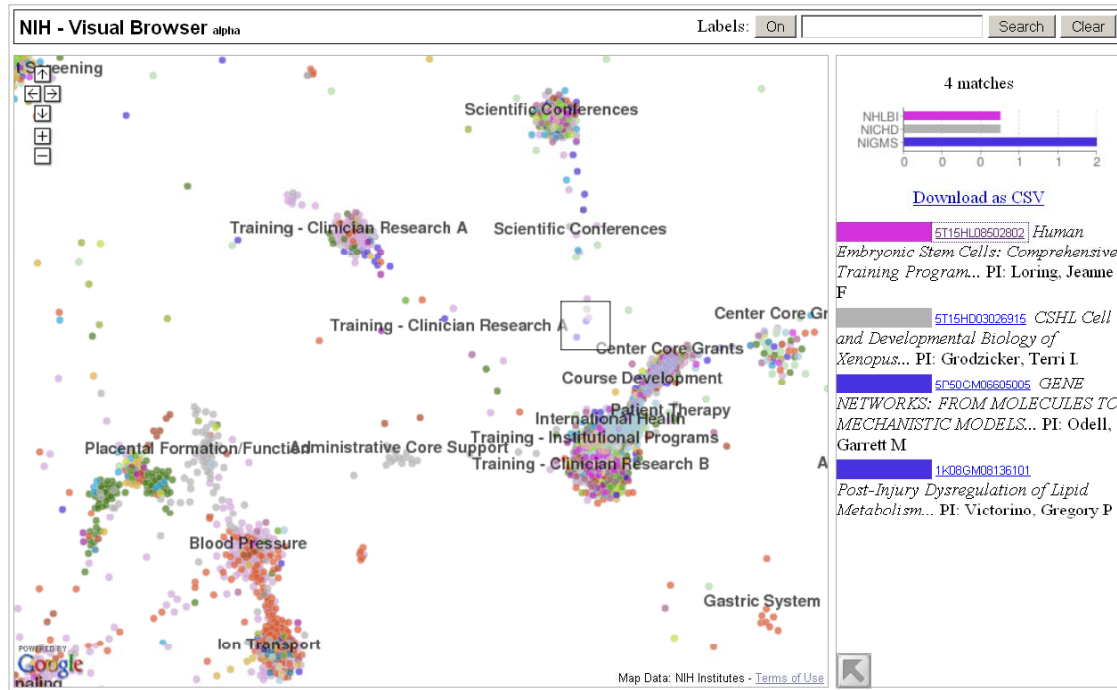
Zoss & Börner, *forthcoming*.



8

Interactive Science Map of NIH Funding

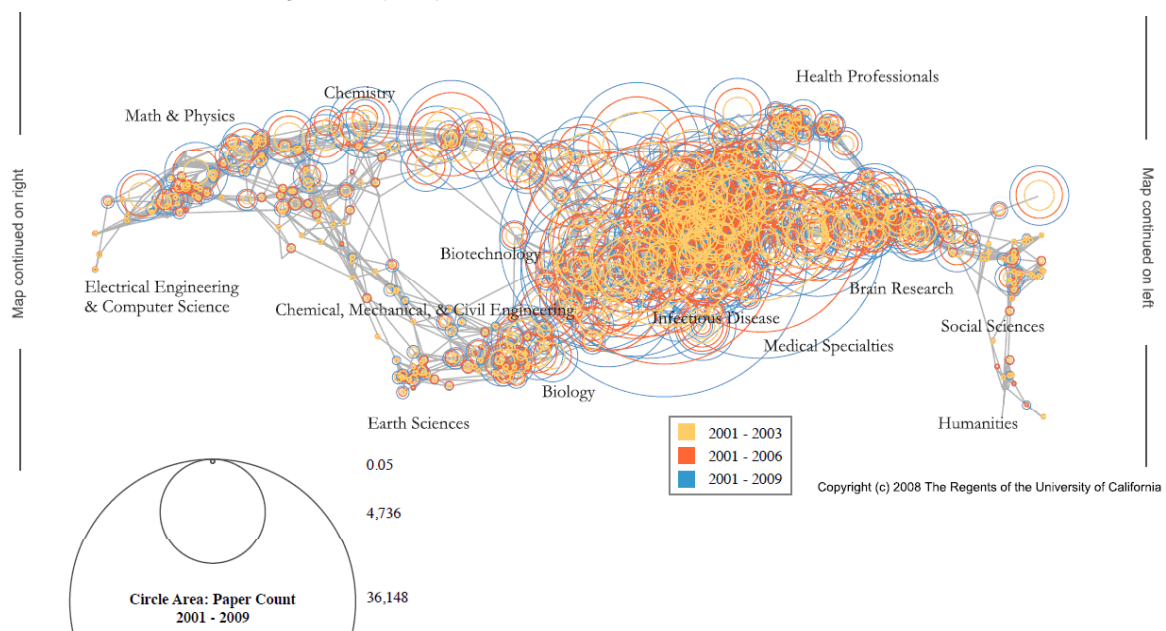
Herr II, Bruce W., Talley, Edmund M, Burns, Gully APC, Newman, David & La Rowe, Gavin. (2009).



<http://scimaps.org/maps/nih/2007>

MEDLINE Publication Output by The National Institutes of Health (NIH) Using Nine Years of ExPORTER Data

Katy Börner, Nianli Ma, Joseph R. Biberstine, Cyberinfrastructure for Network Science Center, SLIS, Indiana University, Robin M. Wagner, Rediet Berhane, Hong Jiang, Susan E. Ivey, Katrina Pearson and Carl McCabe, Reporting Branch, Division of Information Services, Office of Research Information Systems, Office of Extramural Research, Office of the Director, National Institutes of Health (NIH), Bethesda, MD.



Where Are the Academic Jobs? Interactive Exploration of Job Advertisements in Geospatial and Topical Space

Angela Zoss, Michael Conover, Katy Börner (2010)

Visualization of Job Postings

Map of Science

Scientific domains are highly interconnected. The boundaries between different domains are often fuzzy. One way of thinking about the relationships between domains is to conceptualize all scientific domains as existing within a large network of research.

Creating a network of scientific research can be accomplished by looking at scientific journals and their articles. The UCSD Map of Science used here is the product of a large study by researchers at the University of California San Diego using 7.2 million papers and over 16,000 separate journals, proceedings, and series from Thomson Scientific and Scopus over the five year period from 2001 to 2005. The researchers used citations between the papers and journals to cluster journals into small groups of highly related journals.

These clusters are represented by 554 individual nodes in the network. The links between the clusters show that some clusters are related to other clusters but are not as tightly connected as the journals that make up each cluster. Then the clusters are labeled both by the content area shared by the journals in the cluster and by the overarching scientific domain for that cluster (represented by one of 13 colors).

Maps of science like this one can be used to understand many different data sets and how they can be represented by topic. Here we are looking at the topics that appear in job notices from large, job

<http://cns-nd3.slis.indiana.edu/mapjobs/geo>

Type of Analysis vs. Level of Analysis

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

Computational Scientometrics Cyberinfrastructures



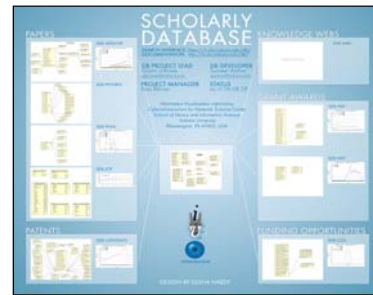
Scholarly Database: 25 million scholarly records

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



VIVO Research Networking

<http://vivoweb.org>



Information Visualization Cyberinfrastructure

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



Network Workbench Tool & Community Wiki

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



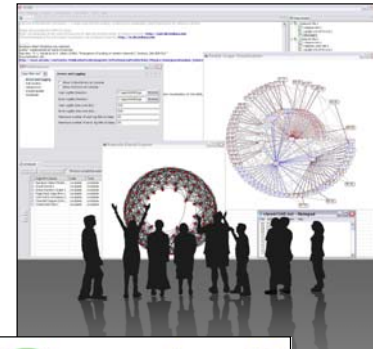
Science of Science (Sci²) Tool and CI Portal

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



Epidemics Cyberinfrastructure

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



13

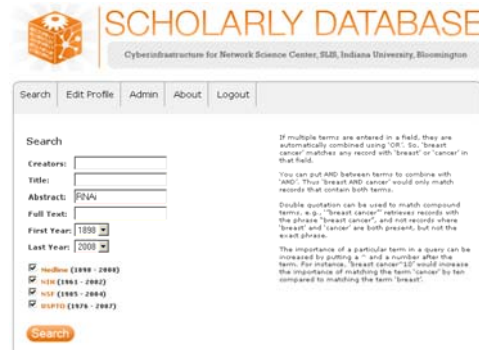


Scholarly Database: Web Interface

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

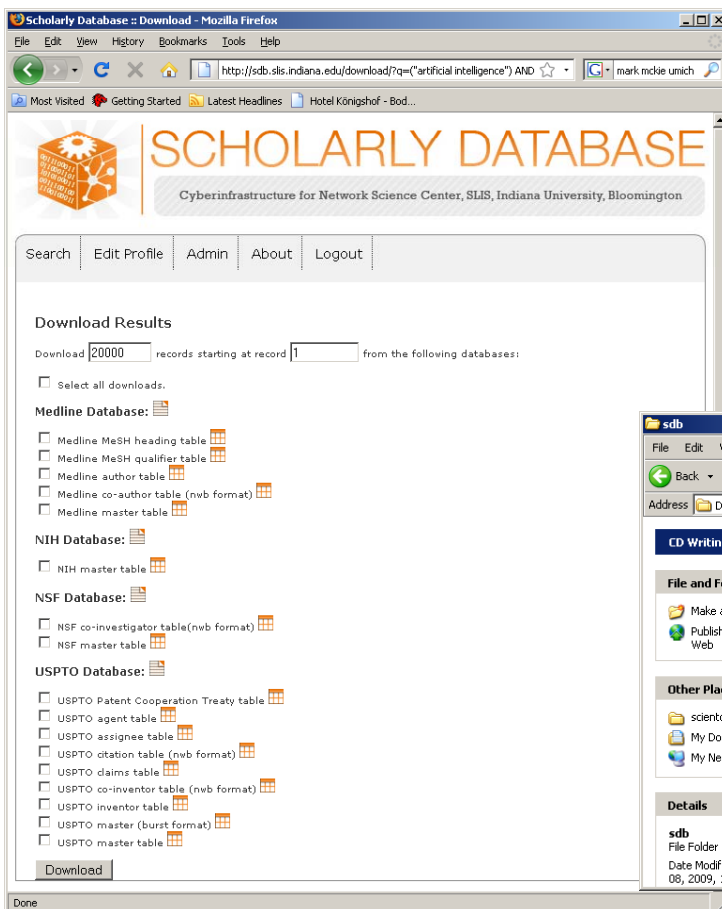
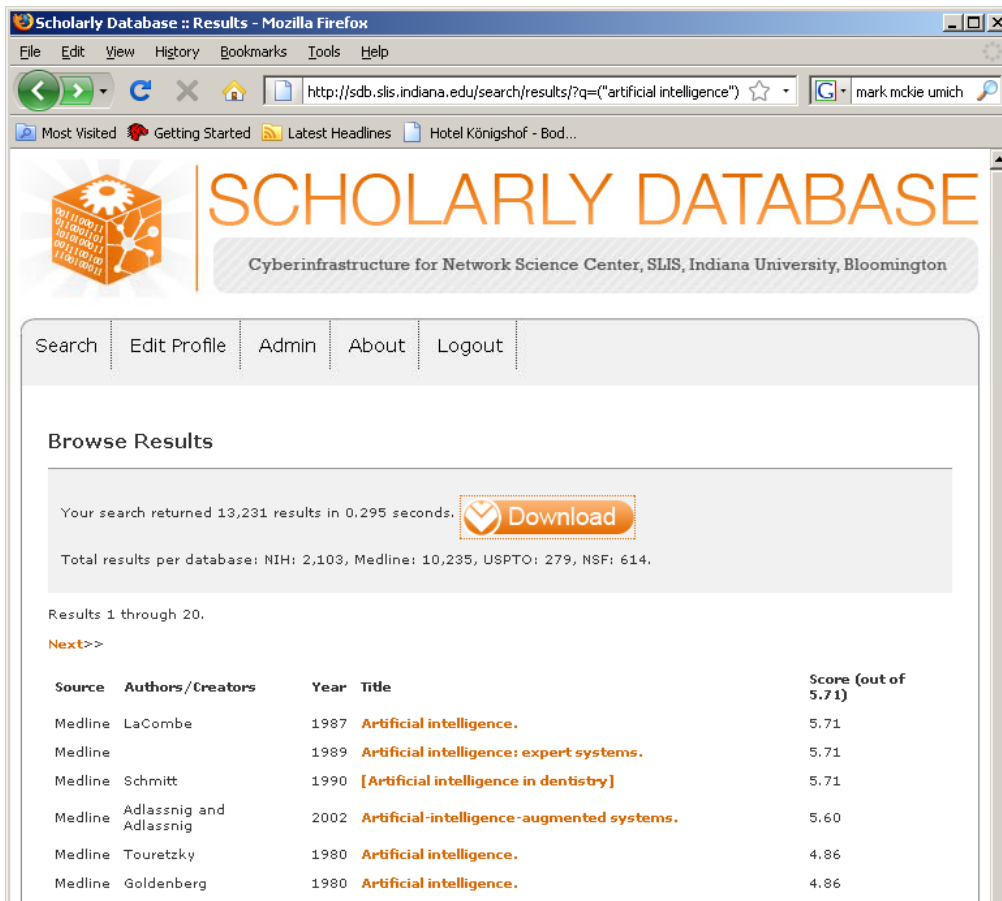
Supports federated search of 25 million publication, patent, grant records.

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

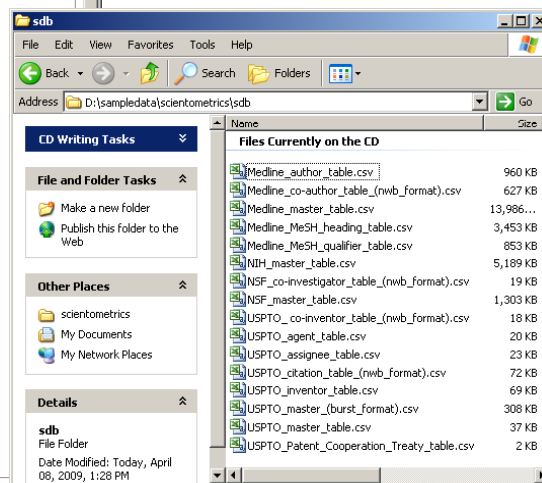


Register for free access at <http://sdb.slis.indiana.edu>

14



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

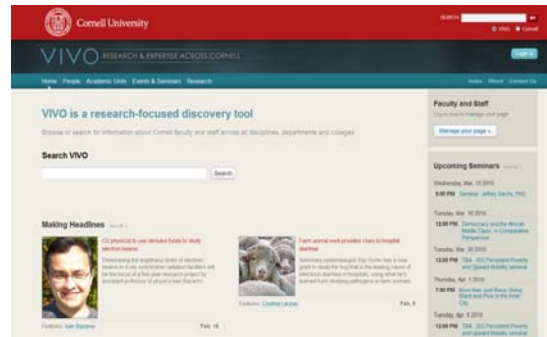


VIVO: A Semantic Approach to Creating a National Network of Researchers (<http://vivoweb.org>)

- Semantic web application and ontology editor originally developed at Cornell U.
- Integrates research and scholarship info from systems of record across institution(s).
- Facilitates research discovery and cross-disciplinary collaboration.

Soon:

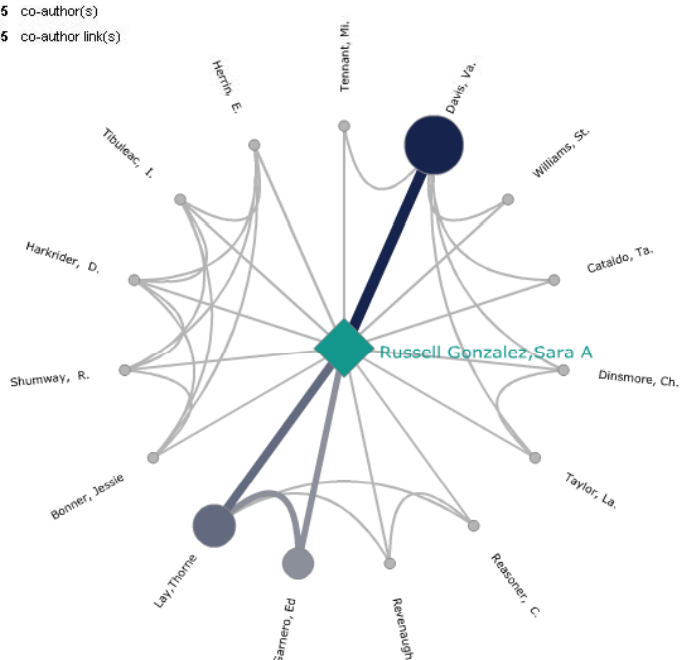
- Simplify reporting tasks, e.g., generate biosketch, department report.



Cornell University: Dean Krafft (Cornell PI), Manolo Bevia, Jim Blake, Nick Cappadona, Brian Caruso, Jon Corson-Rikert, Elly Cramer, Medha Devare, John Ferreira, Brian Lowe, Stella Mitchell, Holly Mistlebauer, Anup Sawant, Christopher Westling, Rebecca Younes. **University of Florida:** Mike Conlon (VIVO and UF PI), Cecilia Botero, Kerry Britt, Erin Brooks, Amy Buhler, Ellie Bushhousen, Chris Case, Valrie Davis, Nita Ferree, Chris Haines, Rae Jesano, Margeaux Johnson, Sara Kreinest, Yang Li, Paula Markes, Sara Russell Gonzalez, Alexander Rockwell, Nancy Schaefer, Michele R. Tennant, George Hack, Chris Barnes, Narayan Raum, Brenda Stevens, Alicia Turner, Stephen Williams. **Indiana University:** Katy Borner (IU PI), William Barnett, Shanshan Chen, Ying Ding, Russell Duhon, Jon Dunn, Micah Linnemeier, Nianli Ma, Robert McDonald, Barbara Ann O'Leary, Mark Price, Yuyin Sun, Alan Walsh, Brian Wheeler, Angela Zoss. **Ponce School of Medicine:** Richard Noel (Ponce PI), Ricardo Espada, Damaris Torres. **The Scripps Research Institute:** Gerald Joyce (Scripps PI), Greg Dunlap, Catherine Dunn, Brant Kelley, Paula King, Angela Murrell, Barbara Noble, Cary Thomas, Michaelen Trimarchi. **Washington University, St. Louis:** Rakesh Nagarajan (WUSTL PI), Kristi L. Holmes, Sunita B. Koul, Leslie D. McIntosh. **Weill Cornell Medical College:** Curtis Cole (Weill PI), Paul Albert, Victor Brodsky, Adam Cheriff, Oscar Cruz, Dan Dickinson, Chris Huang, Itay Klaz, Peter Michelini, Grace Migliorisi, John Ruffing, Jason Specland, Tru Tran, Jesse Turner, Vinay Varughese.

Co-Author Network ([GraphML File](#))

15 co-author(s)
35 co-author link(s)

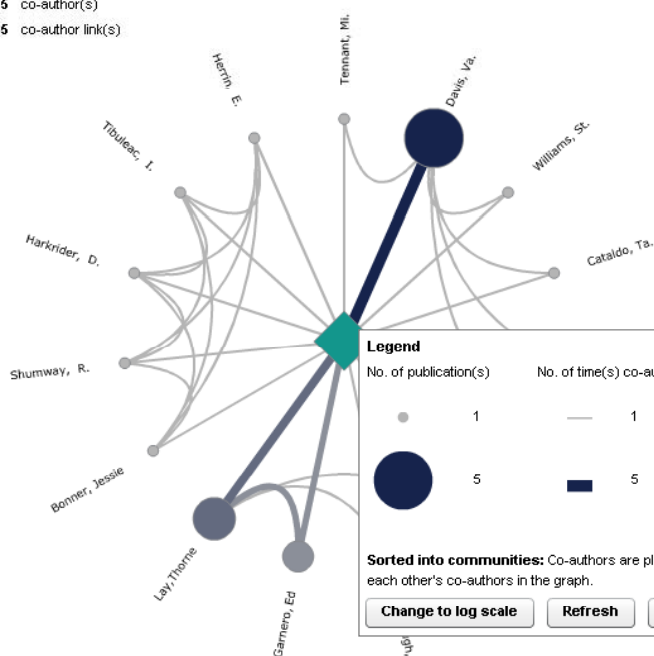


Russell Gonzalez, Sara A
Physical Sciences Librarian
[VIVO profile](#) | [Co-author network](#)

9 Publication(s)
15 Co-author(s)
1998 First Publication
2010 Last Publication

Co-Author Network [\(GraphML File\)](#)

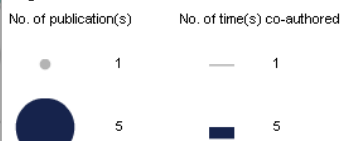
15 co-author(s)
35 co-author link(s)



Russell Gonzalez, Sara A
Physical Sciences Librarian
[VIVO profile](#) | [Co-author network](#)

9 Publication(s)
15 Co-author(s)
1998 First Publication
2010 Last Publication

Legend



Interact

Hover over any name to see the number of joint publications and co-authors with Russell Gonzalez, Sara A.
Click on a name to see details on the right.

Thresholding

Only people that co-authored more than 0 paper(s) with Russell Gonzalez, Sara A. are shown.
0 out of 0 co-author(s) are shown.

Sorted into communities: Co-authors are placed near one another if they frequently collaborate with each other and each other's co-authors in the graph.

[Change to log scale](#) [Refresh](#) [Sort alphabetically](#) [Save as image](#)

19

Borner, Katy

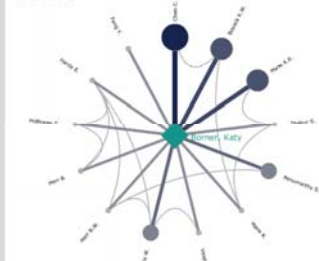
This information is based solely on publications which have been loaded into the VIVO system. This may only be a small sample of the person's total work.

General Statistics

35 publication(s) from 2001 to 2010 [\(.CSV File\)](#)
87 co-author(s) from 2001 to 2010 [\(.CSV File\)](#)

Co-Author Network

13 co-author(s)
28 co-author link(s)



Borner, Katy
Physical Sciences Librarian
35 Publication(s)
13 Co-author(s)
2007 First Publication
2007 Last Publication

Legend



Interact

Hover over any name to see the number of joint publications and co-authors with Borner, Katy.
Click on a name to see details on the right.

Thresholding

Only people that co-authored more than 1 paper(s) with Borner, Katy are shown.
13 out of 87 co-author(s) are shown.

Sorted into communities: Co-authors are placed near one another if they frequently collaborate with each other and each other's co-authors in the graph.

[Change to log scale](#) [Refresh](#) [Sort alphabetically](#) [Save as image](#)

Tables

Publications per Year	Co-author Publications with Borner, Katy
2001	2
2002	4
2003	2
2004	7
2005	7
2006	3
2007	10

Download Data

General Statistics

- 36 publication(s) from 2001 to 2010 [\(.CSV File\)](#)
- 80 co-author(s) from 2001 to 2010 [\(.CSV File\)](#)

Co-Author Network

[\(GraphML File\)](#)

Save as Image (.PNG file)

Tables

- Publications per year [\(.CSV File\)](#)
- Co-authors [\(.CSV File\)](#)

http://vivo-vis.slis.indiana.edu/vivo1/visualization?uri=http%3A%2F%2Fvivo.berkeley.edu%2Ffontology%2Fcore%2Fperson72&vis=person_level&render_mode=standalone

20

36 publication(s) from 2001 to 2010 (.CSV File)

Year	Publications
2001	2
2002	4
2003	2
2004	7
2005	7
2006	3
2007	10
2010	1

80 co-author(s) from 2001 to 2010 (.CSV File)

Year	Count	Co-Author(s)
2001	1	Chen C.
2002	3	Chen C.; McMahon T.; Feng Y.
2003	2	Chen C.; Boyack K.W.
2004	17	Sengupta A.; Penumarthi S.; Thakur S.; Sooriamurthi R.; Maru J.T.; Shiffrin R.M.; Mane K.; Moor K.A.;

Co-author network (GraphML File)

```

1 <?xml version="1.0" encoding="UTF-8"?>
2 <graphml xmlns="http://graphml.graphdrawing.org/xmlns"
3 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
4 xsi:schemaLocation="http://graphml.graphdrawing.org/xmlns
5 http://graphml.graphdrawing.org/xmlns/1.0/graphml.xsd">
6 <key id="label" for="node" attr.name="label" attr.type="string" />
7 <key id="number_of_authored_works" for="node" attr.name="number_of_authored_works" attr.type="int" />
8 <key id="num_unknown_publication" for="node" attr.name="num_unknown_publication" attr.type="int" />
9 <key id="num_latest_publication" for="node" attr.name="num_latest_publication" attr.type="int" />
10 <key id="latest_publication" for="node" attr.name="latest_publication" attr.type="int" />
11 <key id="profile_url" for="node" attr.name="profile_url" attr.type="string" />

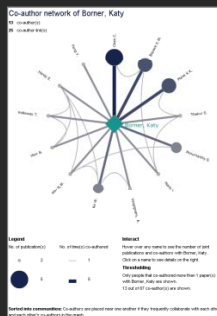
```

Save as Image (.PNG file)

Publications per year (.CSV File), see top file.

Co-authors (.CSV File)

Co-Author	Count
Andrienko G.	1
Andrienko N.	1
Ben-Miled Z.	1
Blackwell A.	1
Boyack K.W.	4
Bozicevic M.	1
Brodbeck D.	1
Burkhard R.A.	1
Chen C.	5



Computational Scientometrics References

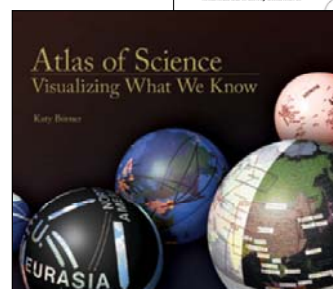
Börner, Katy, Chen, Chaomei, and Boyack, Kevin. (2003). **Visualizing Knowledge Domains**. In Blaise Cronin (Ed.), *ARIST*, Medford, NJ: Information Today, Inc./American Society for Information Science and Technology, Volume 37, Chapter 5, pp. 179-255.
<http://ivl.slis.indiana.edu/km/pub/2003-borner-arist.pdf>



Shiffrin, Richard M. and Börner, Katy (Eds.) (2004). **Mapping Knowledge Domains**. *Proceedings of the National Academy of Sciences of the United States of America*, 101(Suppl_1).
http://www.pnas.org/content/vol101/suppl_1/



Börner, Katy, Sanyal, Soma and Vespignani, Alessandro (2007). **Network Science**. In Blaise Cronin (Ed.), *ARIST*, Information Today, Inc./American Society for Information Science and Technology, Medford, NJ, Volume 41, Chapter 12, pp. 537-607.
<http://ivl.slis.indiana.edu/km/pub/2007-borner-arist.pdf>

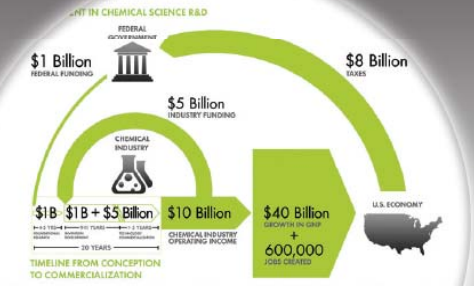
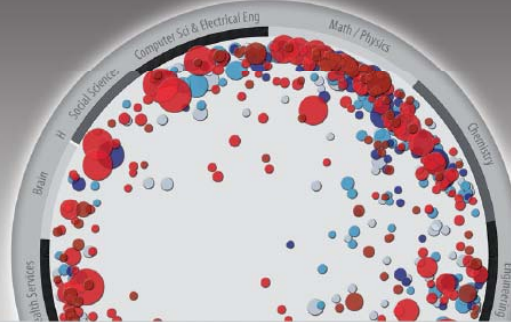


Börner, Katy (2010) *Atlas of Science*. MIT Press.
<http://scimaps.org/atlas>

Mapping Innovation for Congress

CREATING AN **Innovation Dashboard** TO INFORM POLICY MAKERS

A Seminar and Interactive Session | Thursday, December 10, 2009 | Rayburn HOB 2103 | 3:00 p.m. - 4:30 p.m.



U.S. VULNERABILITIES IN SCIENCE

National Strengths are more accurately assessed using a reference-based classification system.

Chemical Research & Development Powers the U.S. Innovation Engine

Macroeconomic Implications of Public and Private R&D Investments in Chemical Science

Have you ever wondered what an "Innovation Dashboard" of your state or congressional district would reveal? Where would the "Pockets of Innovation" lie and what pathways do those innovations take to impact education, health, or profit? Where are the jobs or the innovative education programs? How would you use such a dashboard in your daily work?

This seminar series brings together R&D researchers with a strong background in data analysis and visualization and selected Members of Congress and Congressional Staff. The main goal is to collectively identify features and functionality of an "Innovation Dashboard" for Congress.

AGENDA

- 3:00 p.m. Welcome & Introductions
- 3:10 p.m. Overview of R&D Analyses and Maps
- 3:45 p.m. Open Discussion Forum
- 4:30 p.m. Adjourn


Take ASTRA's Mapping Innovation for Congress Survey—please visit <https://www.surveymonkey.com/s/CX2T1PB>

The first session in this series features Dr. Katy Börner of Indiana University. Dr. Börner will describe different approaches to R&D data analysis and visualization, discuss sample maps from the international Mapping Science exhibit (<http://scimaps.org>) and insights derived from them, as well as near-term developments in the mapping of R&D data. This presentation is followed by a structured but open discussion on desirable features and functionality for "State-Oriented R&D Fact Sheets" and "Innovation Dashboards" and how they would be used to understand, improve, and communicate State and Congressional District innovation metrics.

Please RSVP to the rhoge@comcast.net by December 7, 2009 and confirm your attendance as seating is limited.


Acknowledgments: This work is supported in part by ASTRA, the Cyberinfrastructure for Network Science Center and the School of Library and Information Science at Indiana University, the National Science Foundation (Grants No. 0810730, 0730111, and 0730110), and the James S. McDonnell Foundation.






cyberinfrastructure for NETWORK SCIENCE CENTER


School of Library and Information Science | Indiana University Bloomington




People




Research




Events




Jobs




Contact




News



Teaching



Outreach



Funding

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