

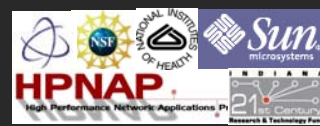
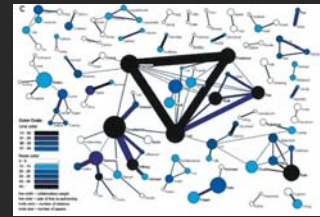
# Designing Insightful (Network) Visualizations of Scholarly Activity

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](mailto:katy@indiana.edu)

*Networks and Network Analysis for the Humanities*  
*Institute for Pure and Applied Mathematics, UCLA, Los Angeles, CA*  
*August 16, 2010*

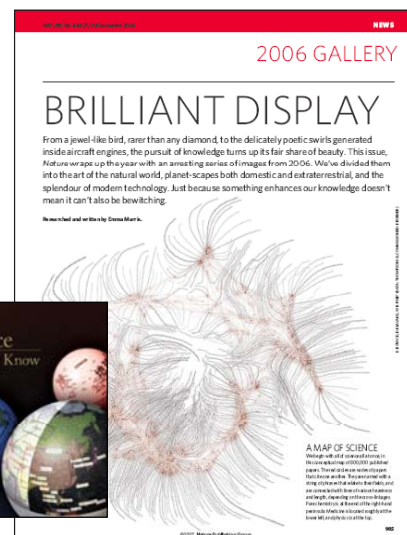


## Computational Scientometrics OR Science of Science Studies

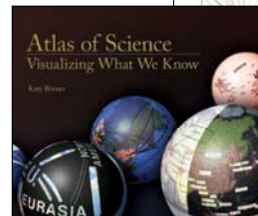
Börner, Katy, Chen, Chaomei, and Boyack, Kevin. (2003).  
**Visualizing Knowledge Domains.** In Blaise Cronin (Ed.), *ARIST*, Medford, NJ: Information Today, Inc./American Society for Information Science and Technology, Volume 37, Chapter 5, pp. 179-255.  
<http://ivl.slis.indiana.edu/km/pub/2003-borner-arist.pdf>



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/](http://www.pnas.org/content/vol101/suppl_1/)

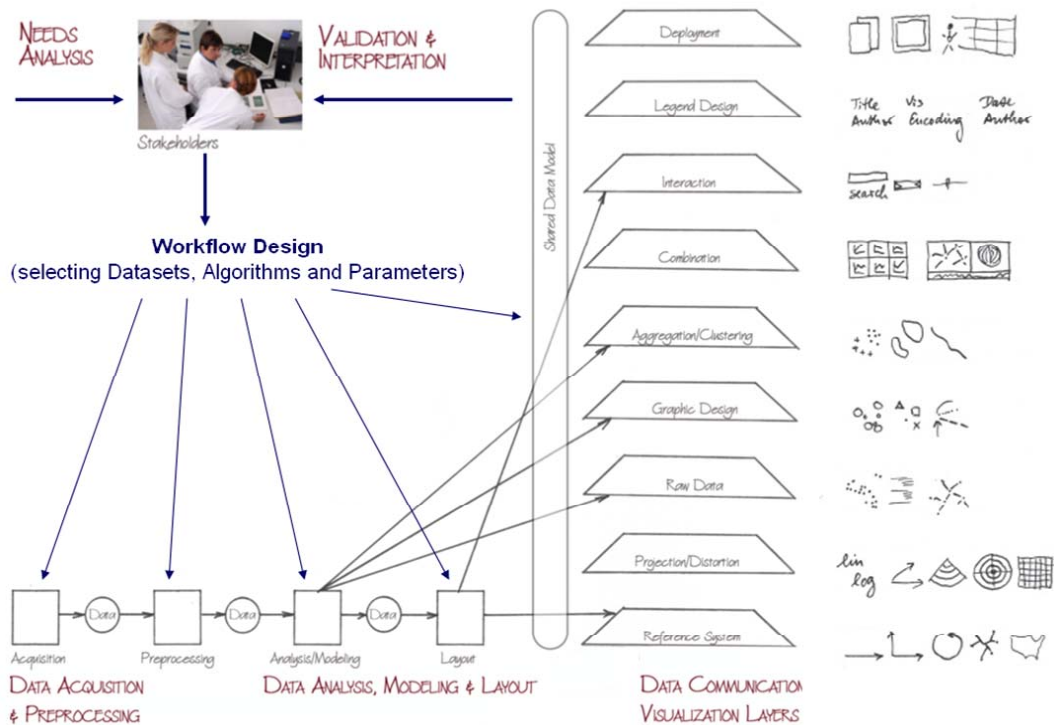


Börner, Katy, Sanyal, Soma and Vespignani, Alessandro (2007). **Network Science.** In Blaise Cronin (Ed.), *ARIST*, Information Today, Inc./American Society for Information Science and Technology, Medford, NJ, Volume 41, Chapter 12, pp. 537-607.  
<http://ivl.slis.indiana.edu/km/pub/2007-borner-arist.pdf>



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

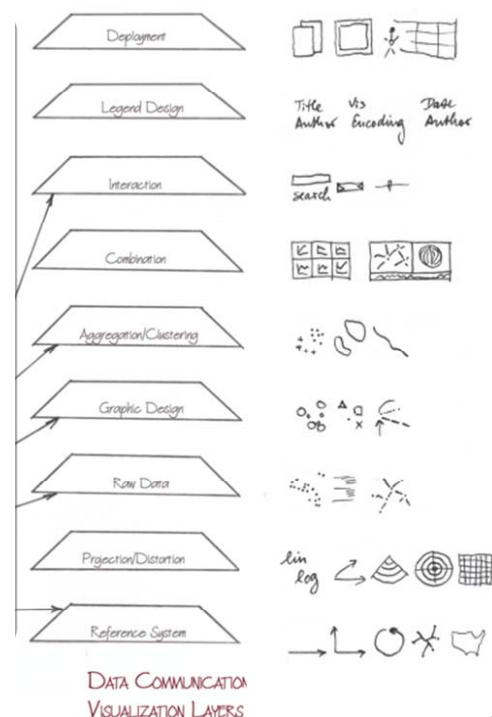
**Visualization Design** using a modular data acquisition/analysis/modeling/visualization pipeline as well as modular visualization layers.



Börner, Katy (2010) *Atlas of Science*. MIT Press. 3

**Visualization Design** using a modular data acquisition/analysis/modeling/visualization pipeline as well as modular visualization layers.

- **Deployment** of results is enabled through paper printouts, online animations, or interactive, three-dimensional, audiovisual environments.
- The **Legend Design** delivers guidance on the purpose, generation, and visual encoding of the data. Mapmakers should proudly sign their visualizations, adding credibility as well as contact information.
- In many cases, it is desirable to **Interact** with the data, that is, to zoom, pan, filter, search, and request details on demand. Selecting a data entity in one view might highlight this entity in other views.
- Sometimes it is beneficial to show multiple simultaneous views of the data, here referred to as **Combination**.
- Frequently, **Aggregation/Clustering** techniques are applied to identify data entities with common attribute values or dense connectivity patterns.
- **Graphic Design** refers to the visual encoding of data attributes using qualities such as size, color, and shape coding of nodes, linkages, or surface areas.
- Placing the **Raw Data** in a reference system reveals spatial patterns.
- **Projections/Distortions** of the reference system help emphasize certain areas or provide focus and context.
- **Reference Systems** organize the space.





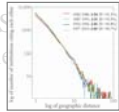
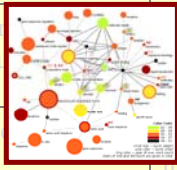




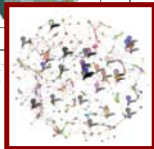

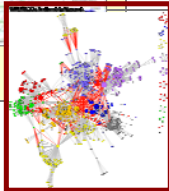

## Type of Analysis vs. Level of Analysis

	<i>Micro/Individual</i> (1-100 records)	<i>Meso/Local</i> (101–10,000 records)	<i>Macro/Global</i> (10,000 < records)
<b>Statistical Analysis/Profiling</b>	Individual person and their expertise profiles	Larger labs, centers, universities, research domains, or states	All of NSF, all of USA, all of science.
<b>Temporal Analysis (When)</b>	Funding portfolio of one individual	Mapping topic bursts in 20-years of PNAS	113 Years of Physics Research
<b>Geospatial Analysis (Where)</b>	Career trajectory of one individual	Mapping a states intellectual landscape	PNAS Publications
<b>Topical Analysis (What)</b>	Base knowledge from which one grant draws.	Knowledge flows in Chemistry research	Topic maps of NIH funding
<b>Network Analysis (With Whom?)</b>	NSF Co-PI network of one individual	Co-author network	NSF's core competency

5



## Type of Analysis vs. Level of Analysis

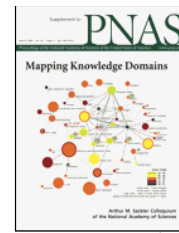
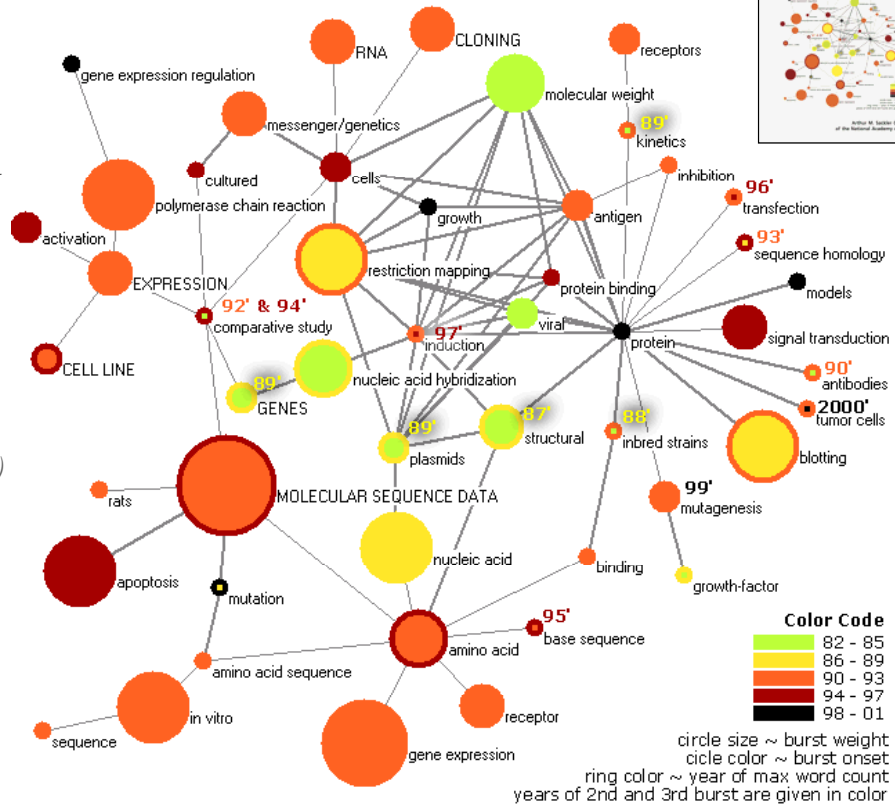
	<i>Micro/Individual</i> (1-100 records)	<i>Meso/Local</i> (101–10,000 records)	<i>Macro/Global</i> (10,000 < records)
<b>Statistical Analysis/Profiling</b>	Individual person and their expertise profiles	Larger labs, centers, universities, research domains, or states	All of NSF, all of USA, all of science. 
<b>Temporal Analysis (When)</b>	Funding portfolio of one individual 	Mapping topic bursts in 20-years 	113 Years of Physics Research 
<b>Geospatial Analysis (Where)</b>	Career trajectory of one individual	Mapping a states intellectual landscape	PNAS Publications 
<b>Topical Analysis (What)</b>	Base knowledge from which one grant draws.	Knowledge flows in Chemistry research 	Topic maps of NIH funding 
<b>Network Analysis (With Whom?)</b>	NSF Co-PI network of one individual 	Co-author network 	NSF's core competency 

6

## Mapping Topic Bursts

Co-word space of the top 50 highly frequent and bursty words used in the top 10% most highly cited PNAS publications in 1982-2001.

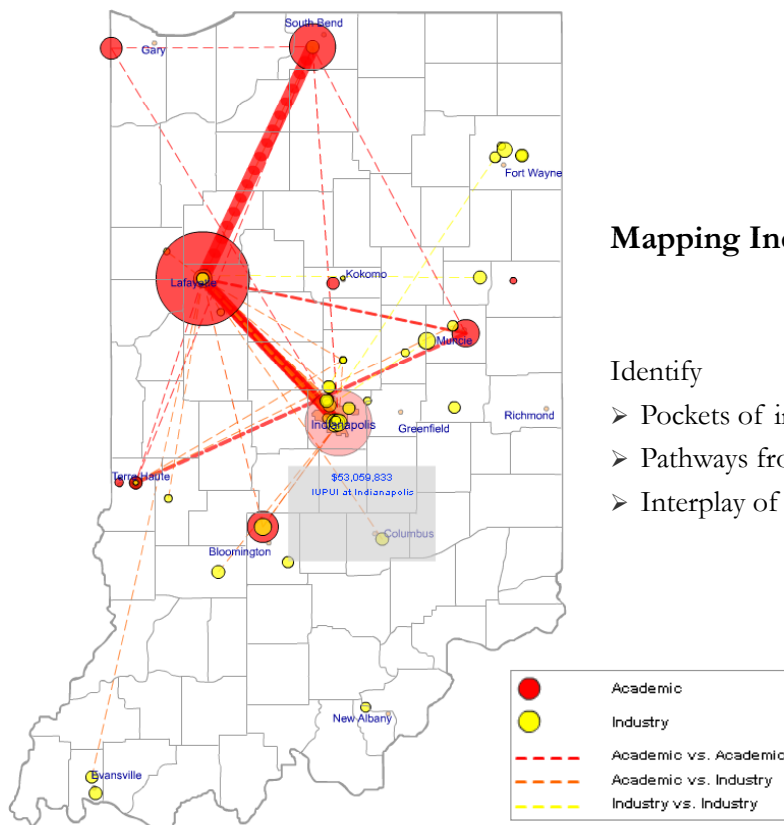
*Mane & Börner. (2004)  
PNAS, 101(Suppl. 1):  
5287-5290.*



## Mapping Indiana's Intellectual Space

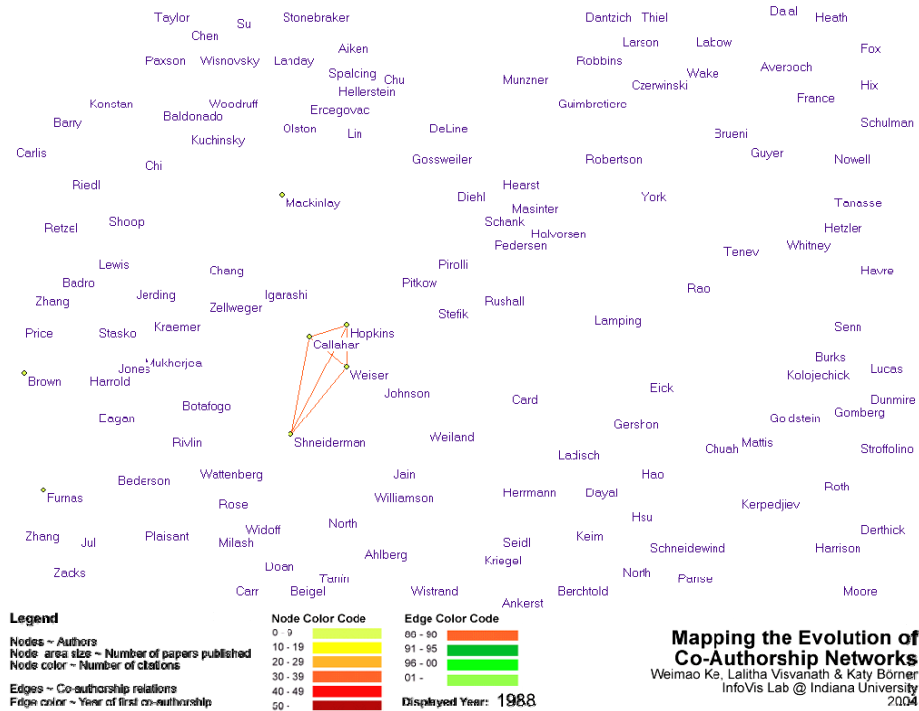
Identify

- Pockets of innovation
- Pathways from ideas to products
- Interplay of industry and academia



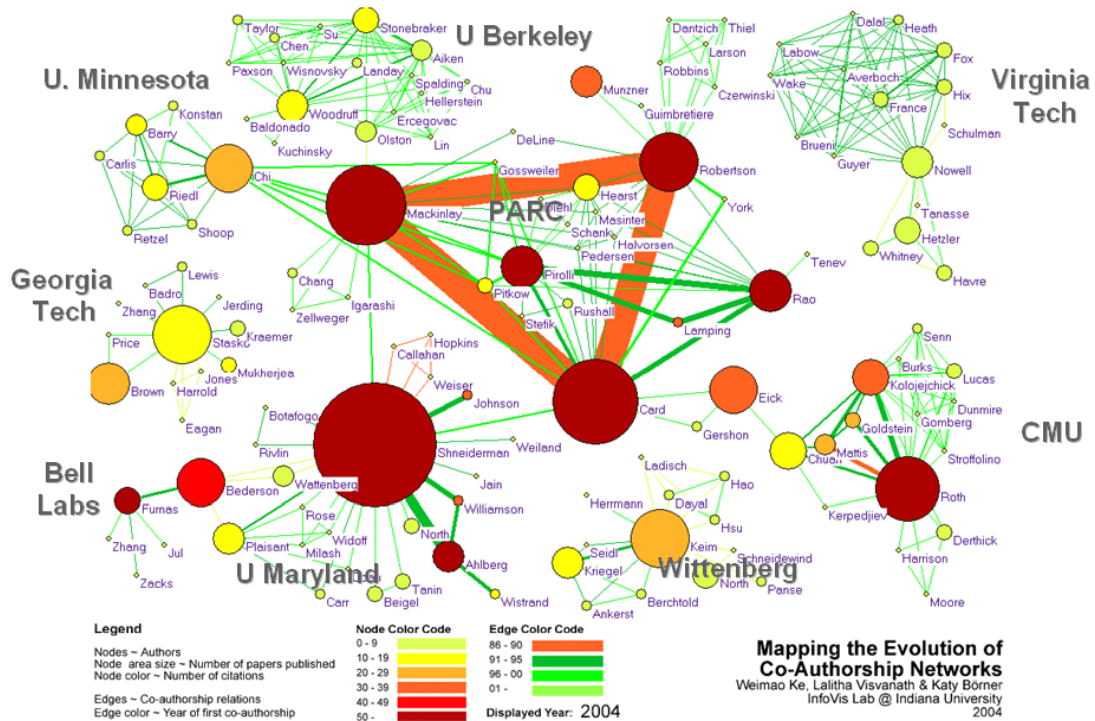
# Mapping the Evolution of Co-Authorship Networks

Ke, Visvanath & Börner, (2004) Won 1st price at the IEEE InfoVis Contest.



# Mapping the Evolution of Co-Authorship Networks

Ke, Visvanath & Börner, (2004) Won 1st price at the IEEE InfoVis Contest.





# Studying the Emerging Global Brain: Analyzing and Visualizing the Impact of Co-Authorship Teams

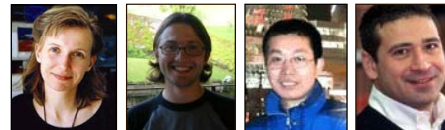
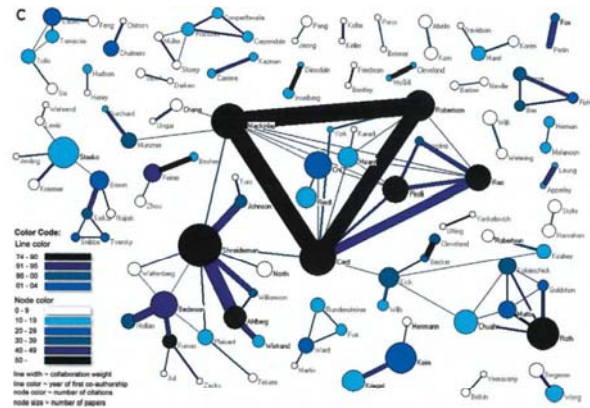
Börner, Dall'Asta, Ke & Vespignani (2005) *Complexity*, 10(4):58-67.

## Research question:

- Is science driven by prolific single experts or by high-impact co-authorship teams?

## Contributions:

- New approach to allocate citational credit.
- Novel weighted graph representation.
- Visualization of the growth of weighted co-author network.
- Centrality measures to identify author impact.
- Global statistical analysis of paper production and citations in correlation with co-authorship team size over time.
- Local, author-centered entropy measure.

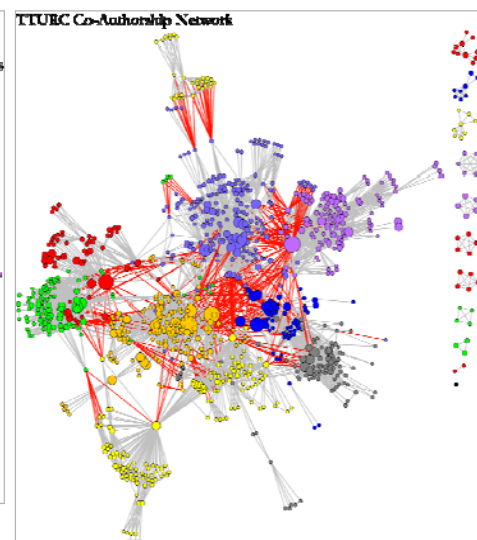
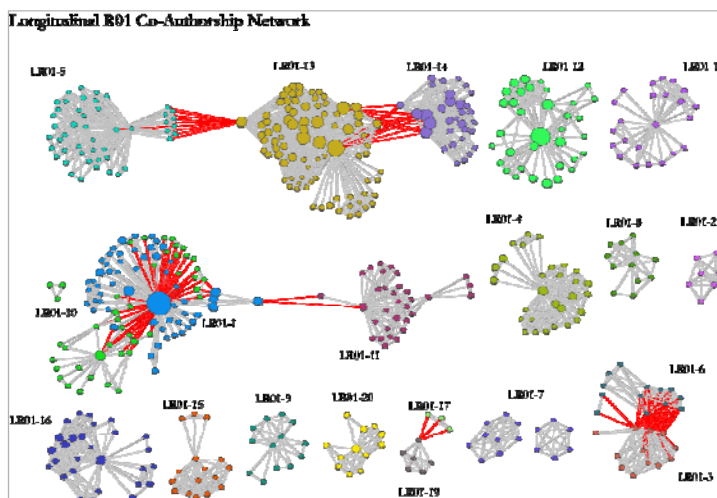
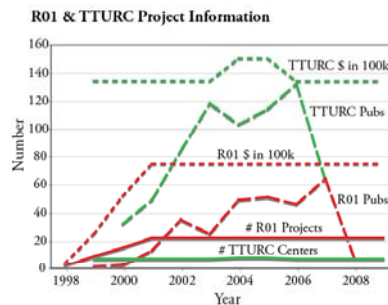


11

# Mapping Transdisciplinary Tobacco Use Research Centers Publications

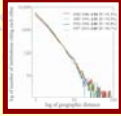






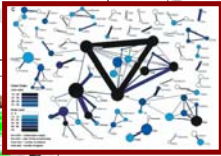
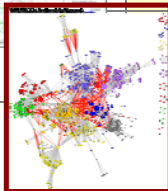


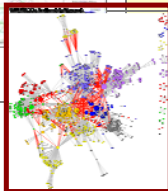

Compare R01 investigator based funding with TTURC Center awards in terms of number of publications and evolving co-author networks.

Zoss & Börner, *forthcoming*.





## 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)
<b>Statistical Analysis/Profiling</b>	Individual person and their expertise profiles	Larger labs, centers, universities, research domains, or states	All of NSF, all of USA, all of science. 
<b>Temporal Analysis (When)</b>	Funding portfolio of one individual 	Mapping topic bursts in 20-years 	113 Years of Physics Research 
<b>Geospatial Analysis (Where)</b>	Career trajectory of one individual 	Mapping a states intellectual landscape 	PNAS Publications 
<b>Topical Analysis (What)</b>	Base knowledge from which one grant draws. 	Knowledge flows in Chemistry research 	Topic maps of NIH funding 
<b>Network Analysis (With Whom?)</b>	NSF Co-PI network of one individual 	Co-author network 	NSF's core competency 

13



## 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)
<b>Statistical Analysis/Profiling</b>	Individual person and their expertise profiles	Larger labs, centers, universities, research domains, or states	All of NSF, all of USA, all of science.
<b>Temporal Analysis (When)</b>	Funding portfolio of one individual	Mapping topic bursts in 20-years of PNAS	113 Years of Physics Research
<b>Geospatial Analysis (Where)</b>	Career trajectory of one individual	Mapping a states intellectual landscape	PNAS Publications
<b>Topical Analysis (What)</b>	Base knowledge from which one grant draws.	Knowledge flows in Chemistry research	Topic maps of NIH funding
<b>Network Analysis (With Whom?)</b>	NSF Co-PI network of one individual	Co-author network	NSF's core competency

14

## Spatio-Temporal Information Production and Consumption of Major U.S. Research Institutions

Börner, Katy, Penumarth, Shashikant, Meiss, Mark and Ke, Weimao. (2006) *Mapping the Diffusion of Scholarly Knowledge Among Major U.S. Research Institutions. Scientometrics. 68(3), pp. 415-426*

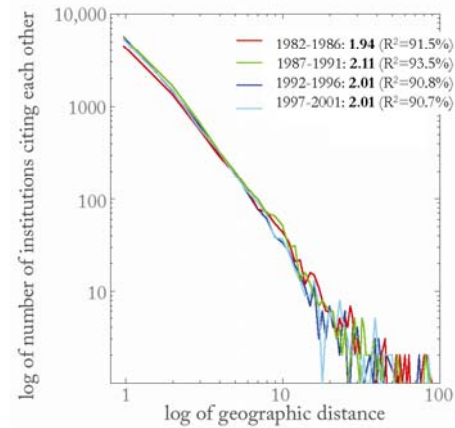
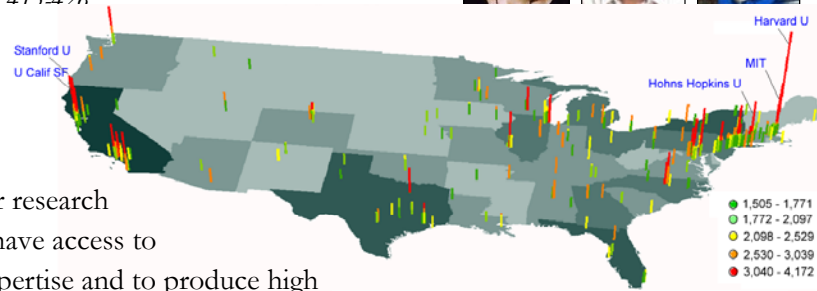


### Research questions:

1. Does space still matter in the Internet age?
2. Does one still have to study and work at major research institutions in order to have access to high quality data and expertise and to produce high quality research?
3. Does the Internet lead to more global citation patterns, i.e., more citation links between papers produced at geographically distant research institutions?

### Contributions:

- Answer to Qs 1 + 2 is YES.
- Answer to Qs 3 is NO.
- Novel approach to analyzing the dual role of institutions as information producers and consumers and to study and visualize the diffusion of information among them.



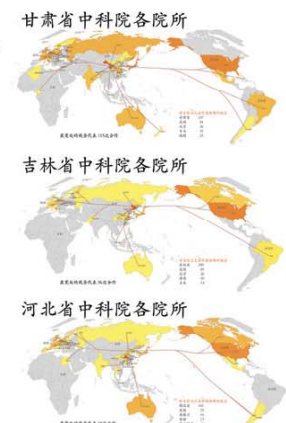
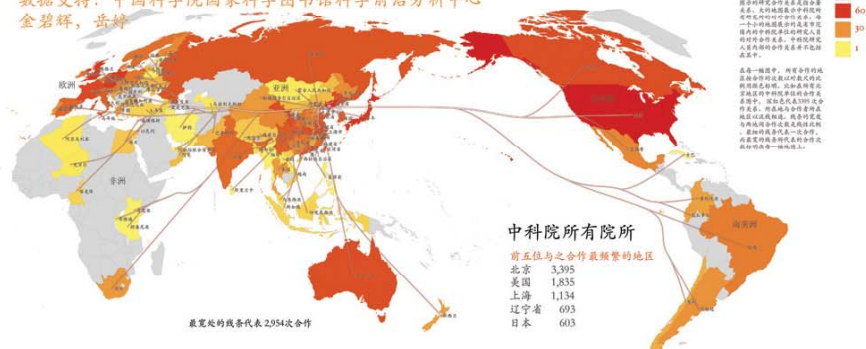
15

## Research Collaborations by the Chinese Academy of Sciences

By Weixia (Bonnie) Huang, Russell J. Dubon, Elisha F. Hardy, Katy Börner, Indiana University, USA

### 中科院与世界各地的研究合作关系

黄维霞, Russell J. Dubon, Elisha F. Hardy, Katy Börner, Indiana University, USA  
 数据支持: 中国科学院国家科学图书馆科学前沿分析中心  
 金碧群, 岳峰



中科院所有院所  
 前五位与之合作最频繁的地区

北京	3,395
美国	1,835
上海	1,134
辽宁省	693
日本	603

Co-authorship collaborations of the Beijing branch of the Chinese Academy of Sciences (left) and six regional branches (3 shown on right) with countries around the world.

Collaborating countries are colored on a logarithmic scale by the number of collaborations from red to yellow. The darkest red denotes 3,395 collaborations by all CAS researchers in Beijing. Flow lines are drawn from the location of focus to all locations collaborated with. Line width is linearly proportional to the number of joint papers.

16



The map of science was constructed by sorting more than 7.2 million journals into disciplines, disciplines represented as nodes, and all of them that share a common literature link (the links between disciplines) are pairs of disciplines that share a common literature. A three-dimensional model was used to determine the position of each discipline on the surface of a sphere based on the linkage between disciplines. The model treats links like rubber bands connecting to bring new disciplines into each other. Pairs of disciplines were then used to wind up the different sides of the map.

# MAPS OF SCIENCE

A visualization of 7.2 million scholarly documents

## Forecasting Large Trends in Science

Calculations were performed using the large robust groupings of disciplines (fields) to determine if any of them were likely to show large scale change in the structure of science over time. Correlation coefficients between fields were calculated for each individual year, 2001-2005. A single regression analysis was conducted for each of these significant changes in these correlation coefficients from year to year.

If the structure of science changed before it was expected, it would suggest some relations between neighboring fields to be broken, and connections between distant fields to be broken. We found the opposite, suggesting that the underlying structure is unstable and likely to change dramatically over the next decade.

We discuss implications from the structure of fields to change and trends in science. Many with other areas from the map of science (including all fields) to be further shown to be likely to become more disjoint. We expect that future maps of science will show changes in clusters corresponding to these observations. Individual will become lighter while the physical sciences will tighten and draw closer to the medical fields.

### Data:

WoS and Scopus for 2001–2005, 7.2 million papers, more than 16,000 separate journals, proceedings, and series

### Similarity Metric:

Combination of bibliographic coupling and keyword vectors

### Number of Disciplines:

554 journal clusters further aggregated into 13 main scientific disciplines that are labeled and color coded in a metaphorical way, e.g., Medicine is blood red and Earth Sciences are brown as soil.



**Technical Sciences & Computer Science (TCS)** is highlighted by the light green color. This field is highlighted because of its central position in the map. It is highlighted because of its central position in the map. It is highlighted because of its central position in the map.



**Engineering** is highlighted by the light green color. This field is highlighted because of its central position in the map. It is highlighted because of its central position in the map. It is highlighted because of its central position in the map.



**Medical Diseases** is highlighted by the dark red color. This field is highlighted because of its central position in the map. It is highlighted because of its central position in the map. It is highlighted because of its central position in the map.



**Medical Specialties** is highlighted by the dark red color. This field is highlighted because of its central position in the map. It is highlighted because of its central position in the map. It is highlighted because of its central position in the map.



**The Health Professionals** is highlighted by the light red color. This field is highlighted because of its central position in the map. It is highlighted because of its central position in the map. It is highlighted because of its central position in the map.



**The Humanities** is highlighted by the white color. This field is highlighted because of its central position in the map. It is highlighted because of its central position in the map. It is highlighted because of its central position in the map.

Richard Klavans and Kevin Boyack. 2007. Maps of Science: Forecasting Large Trends in Science.

## MAPS OF SCIENCE



Overview | Detail | Disciplinary Maps | Competency Maps | Paradigm Maps | Posters

Institutional Strategy: NIH

The following color coding is used for the disciplinary map:

- Math & Physics
- Biotechnology
- Medical Specialties
- Humanities
- Chemistry
- Earth Sciences
- Brain Research
- Health Professionals
- Computer Science & EE
- Biology
- Social Sciences
- Other Engineering
- Infectious Diseases

View all

National Institute of General Med Science

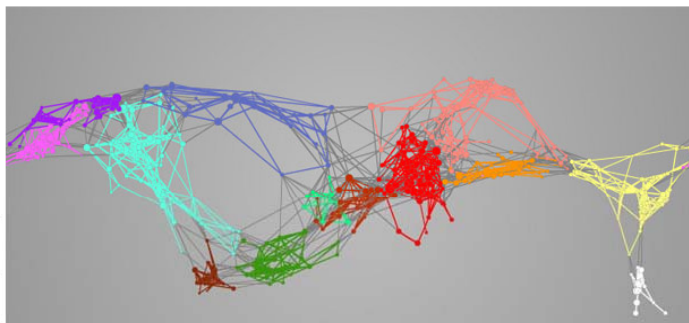
National Institute of Allergy & Inf Disease

Nat. Cancer Institute

Nat. Heart, Lung & Blood Institute

Nat. Inst Diabetes, Dig & Kidney Disease

Nat. Inst of Neuro Disorders & Stroke

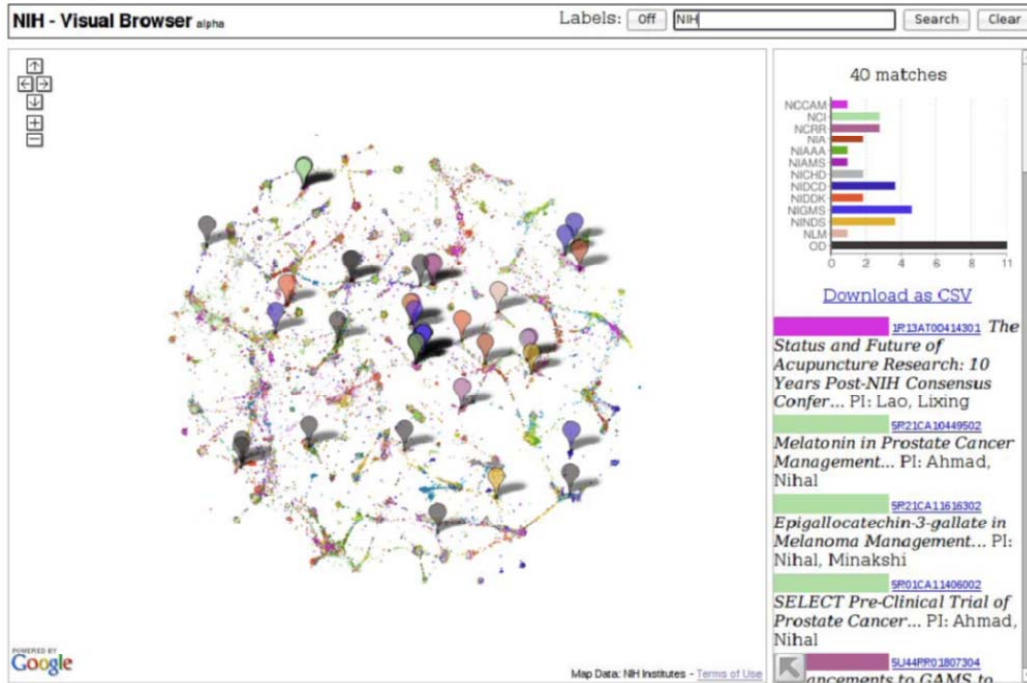


<http://mapofscience.com>



## Interactive Science Map of NIH Funding

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

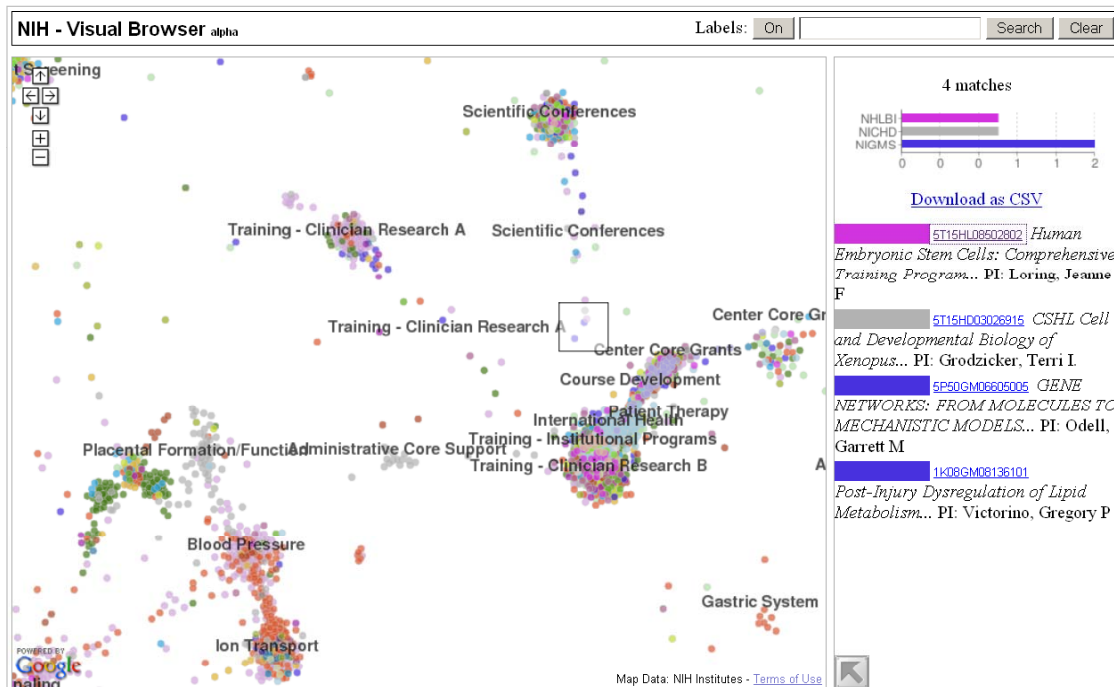


<http://scimaps.org/maps/nih/2007>

21

## Interactive Maps of Science – NIH Funding

Google maps with charts and tables

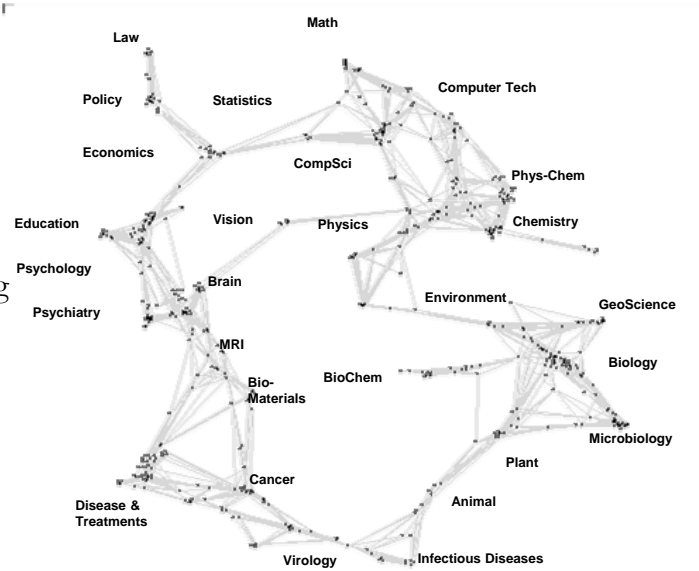


<http://scimaps.org/maps/nih/2007>

## 2002 Base Map of Science

Kevin W. Boyack, Katy Börner, & Richard Klavans (2007). *Mapping the Structure and Evolution of Chemistry Research*. 11th International Conference on Scientometrics and Informetrics. pp. 112-123.

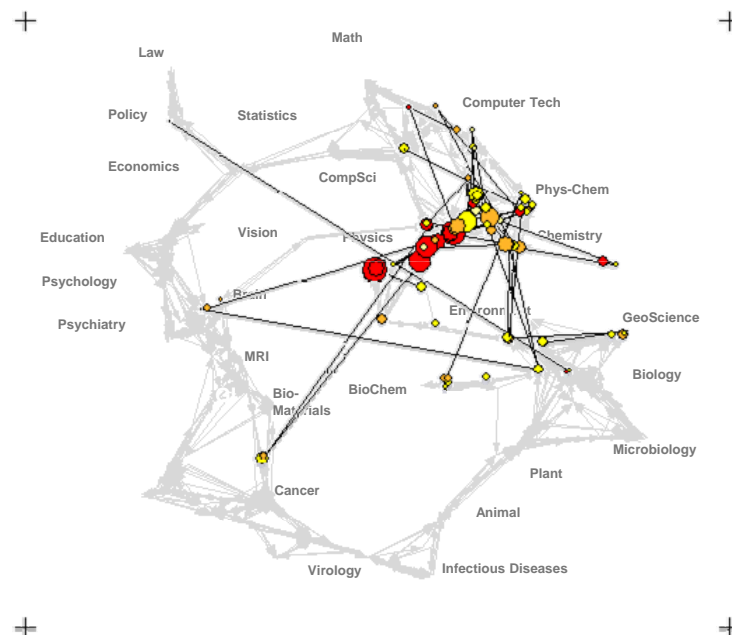
- Uses combined SCI/SSCI from 2002
  - 1.07M papers, 24.5M references, 7,300 journals
  - Bibliographic coupling of papers, aggregated to journals
- Initial ordination and clustering of journals gave 671 clusters
- Coupling counts were reaggregated at the journal cluster level to calculate the
  - (x,y) positions for each journal cluster
  - by association, (x,y) positions for each journal



## Science map applications: Identifying core competency

Kevin W. Boyack, Katy Börner, & Richard Klavans (2007).

Funding patterns of the US Department of Energy (DOE)

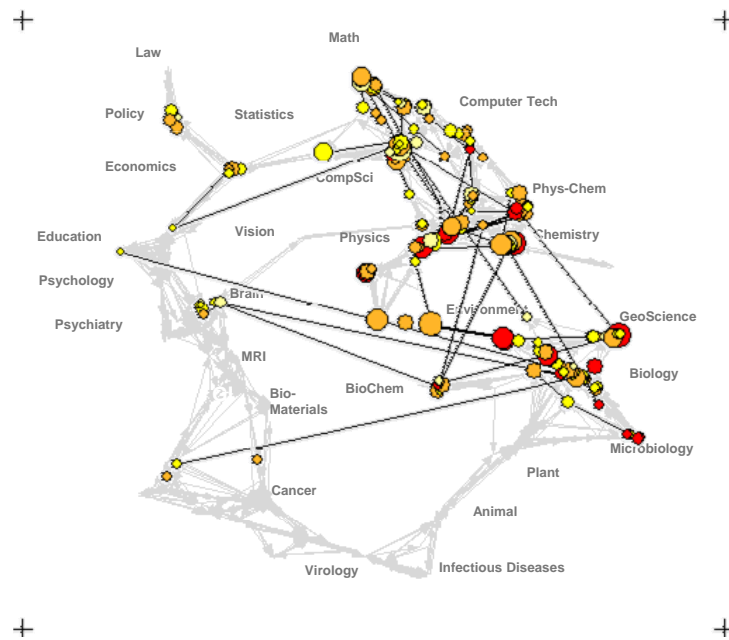




## Science map applications: Identifying core competency

*Kevin W. Boyack, Katy Börner, & Richard Klavans (2007).*

### Funding Patterns of the National Science Foundation (NSF)

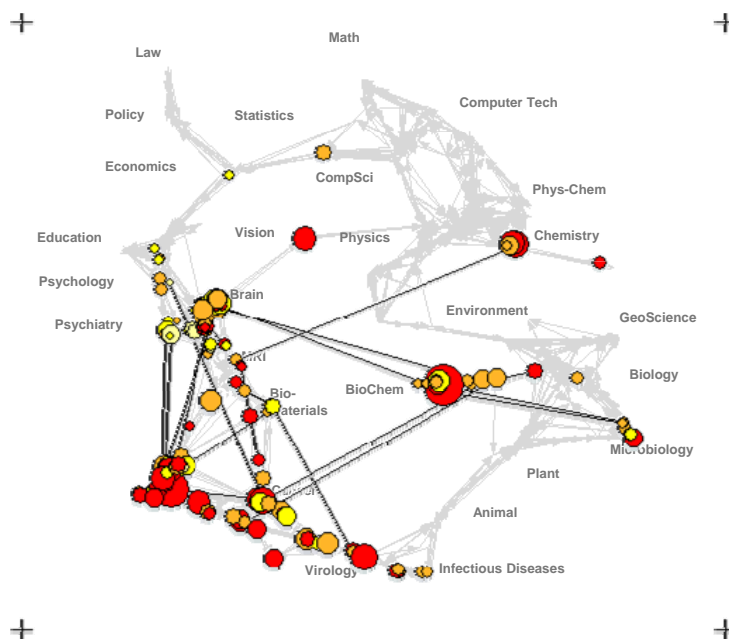


25

## Science map applications: Identifying core competency

*Kevin W. Boyack, Katy Börner, & Richard Klavans (2007).*

### Funding Patterns of the National Institutes of Health (NIH)

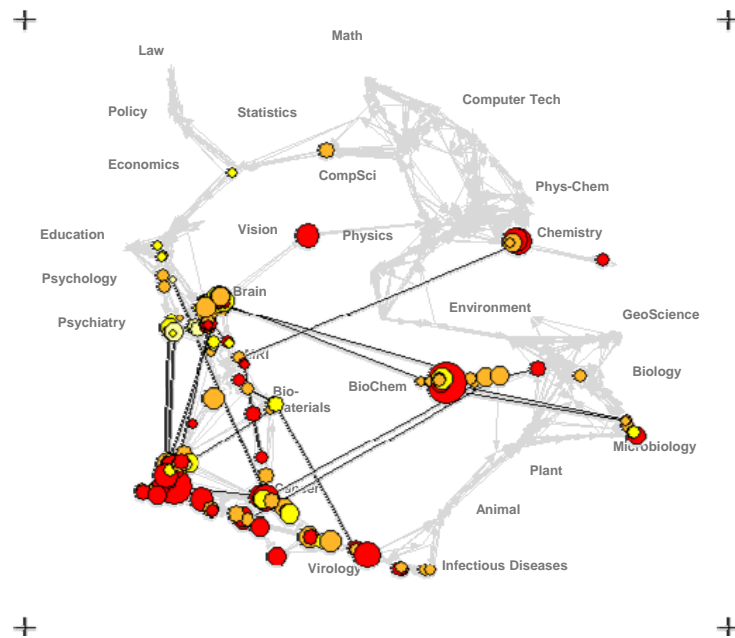


26

## Science map applications: Identifying core competency

Kevin W. Boyack, Katy Börner, & Richard Klavans (2007).

### Funding Patterns of the National Institutes of Health (NIH)



27

## Computational Scientometrics Cyberinfrastructures



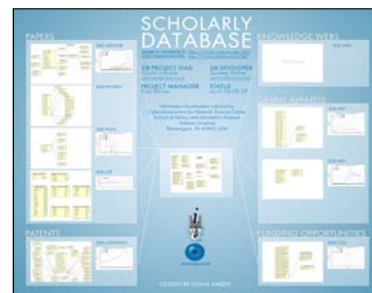
Scholarly Database: 23 million scholarly records

<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>



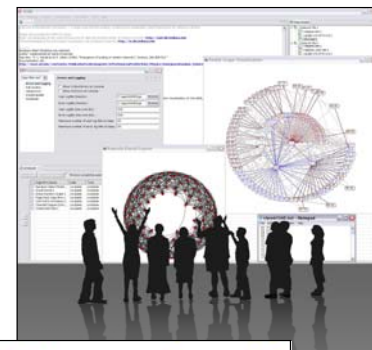
Science of Science (Sci<sup>2</sup>) Tool and CI Portal

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



Epidemics Cyberinfrastructure

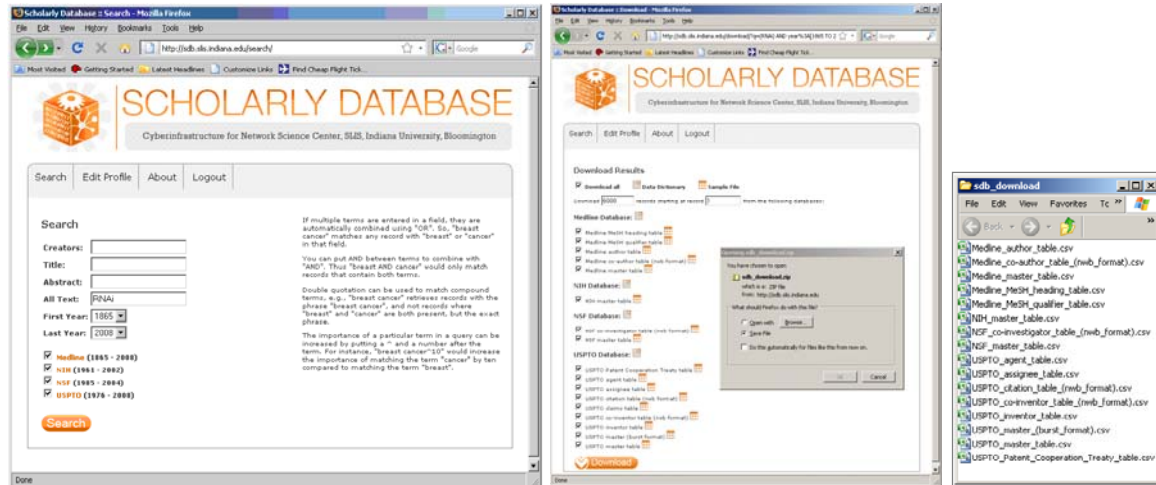
<http://epic.slis.indiana.edu/>



28

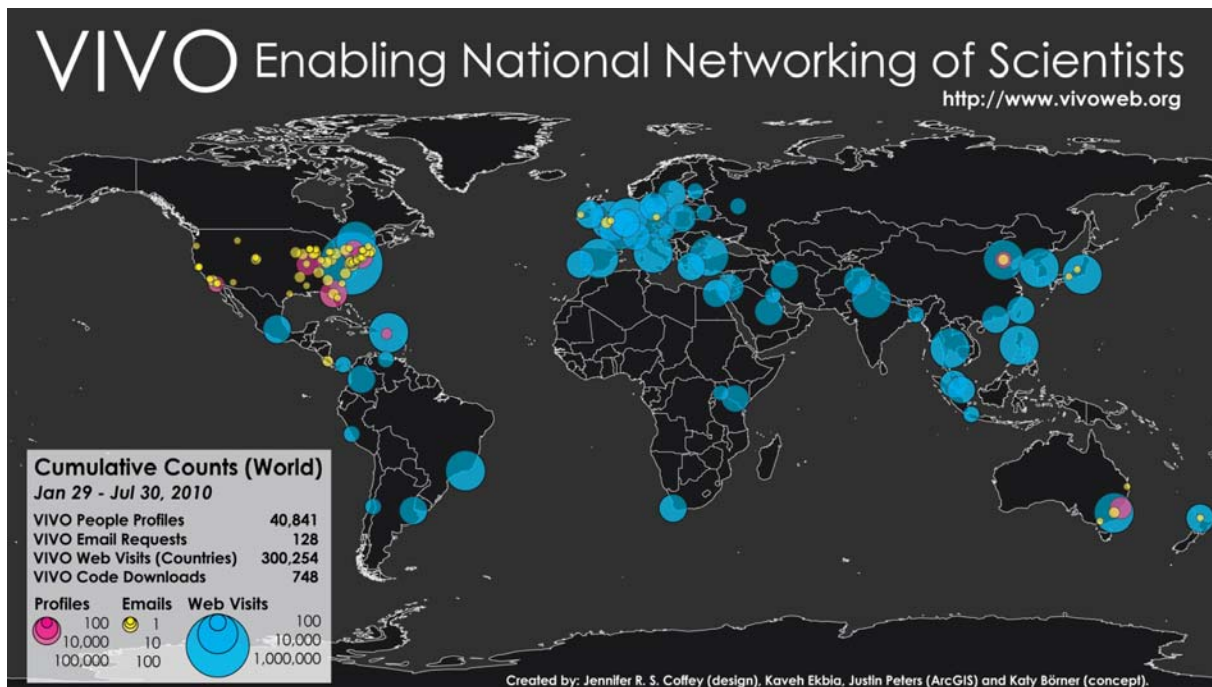


## Scholarly Database (<http://sdb.slis.indiana.edu>)



The **Scholarly Database** at Indiana University provides free access to 23,000,000 papers, patents, and grants. Since March 2009, users can also download networks, e.g., co-author, co-investigator, co-inventor, patent citation, and tables for burst analysis.

29



VIVO is a 2-year, \$12 Million dollar project funded by NIH.

VIVO 1.0 source code was released on April 14, 2010.

The more institutions adopt VIVO, the more high quality data will be available to understand, navigate, manage, utilize, and communicate progress in science and technology.

30



## 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.

It has been downloaded more than 60,000 times since October 2006.

The screenshot shows the NetworkWorkbench website homepage. At the top, there is a navigation menu with links for Home, People, Research, Publications, Community, Download, Documentation, Dev Zone, and About. Below the menu is a 'Summary' section with a 'LOGIN' button. The main content area includes a 'News & Updates' section with several entries, a 'Download 1.0.0 beta 5 Release' section with a 'DOWNLOAD' button, and a 'Get Involved' section.

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.

31



## Sci<sup>2</sup> Tool for Science of Science

Research and Practice (<http://sci.slis.indiana.edu/sci2>)

The screenshot shows the Sci<sup>2</sup> Tool login page. It features a green header with the Sci<sup>2</sup> logo and the text 'A tool for science of science research & practice'. Below the header are input fields for 'Email Address' and 'Password', and a 'Login' button.

### Forgot your password?

To recover your account pas

### Not registered yet?

[Register now](#)

### Tutorials

Scott Weingart, Hanning Guo, Biberstine (2010) *Science of Science*, Indiana University, I

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.

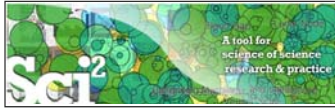
- Tutorial #01: [Science of Science Research](#)
- Tutorial #02: [Network Science / Information Visualization](#)
- Tutorial #03: [CIShell Powered Tools: Network Workbench and Science of Science Tool](#)
- Tutorial #04: [Temporal Analysis—Burst Detection](#)
- Tutorial #05: [Geospatial Analysis and Mapping](#)
- 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](#)

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

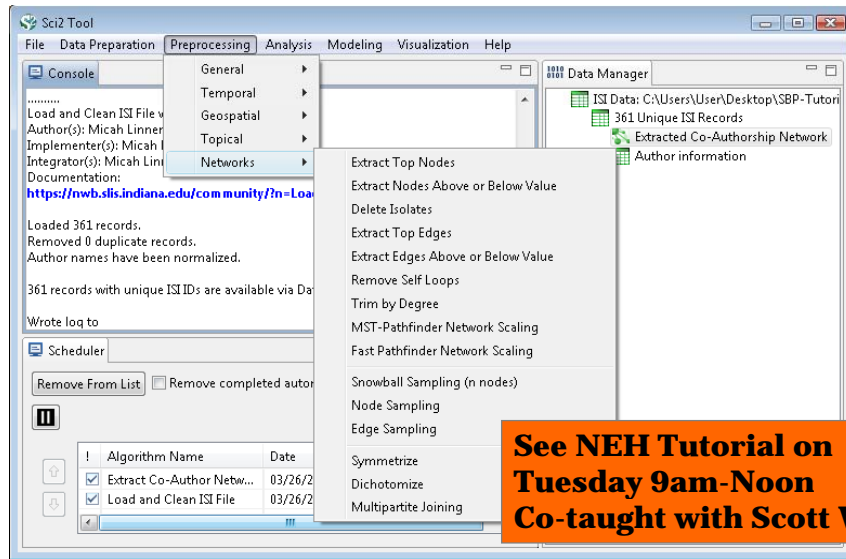
NIH Office of Extramural Research and Katy Börner (2010) [Network Visualizations Using SPIRES Data and the Sci<sup>2</sup> Tool](#). Office of Extramural Research at NIH.

32





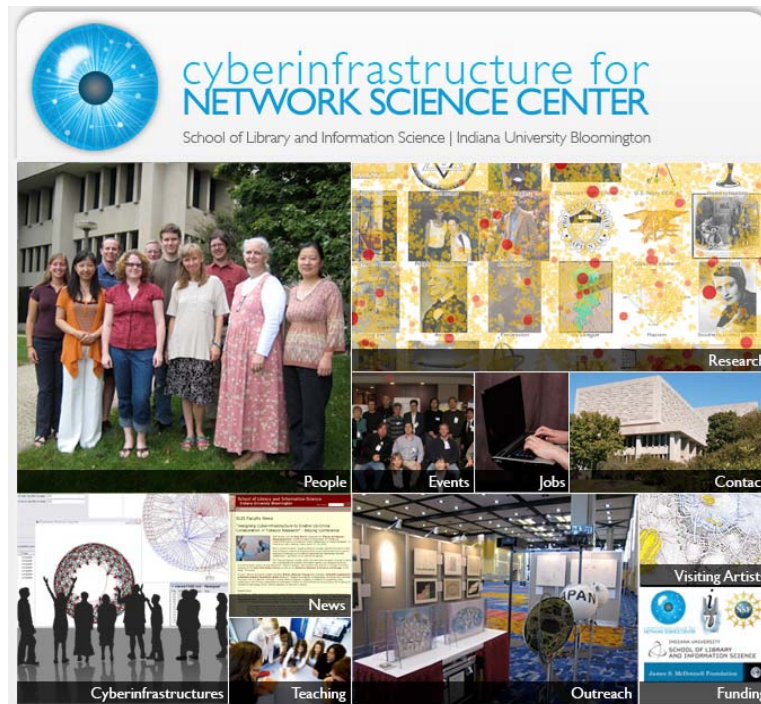
# Sci<sup>2</sup> Tool for Science of Science Research and Practice (<http://sci.slis.indiana.edu/sci2>)



**See NEH Tutorial on Tuesday 9am-Noon Co-taught with Scott Weingart**

### Acknowledgments

This work is supported in part by the Cyberinfrastructure for Network Science center and the School of Library and Information Science at Indiana University, the National Science Foundation under Grant No. SBE-0738111 and IIS-0513650, and the James S. McDonnell Foundation.



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