

# Plug-and-Play Microscopes: Custom Tools for Data Analysis, Modeling, and Visualization

**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)



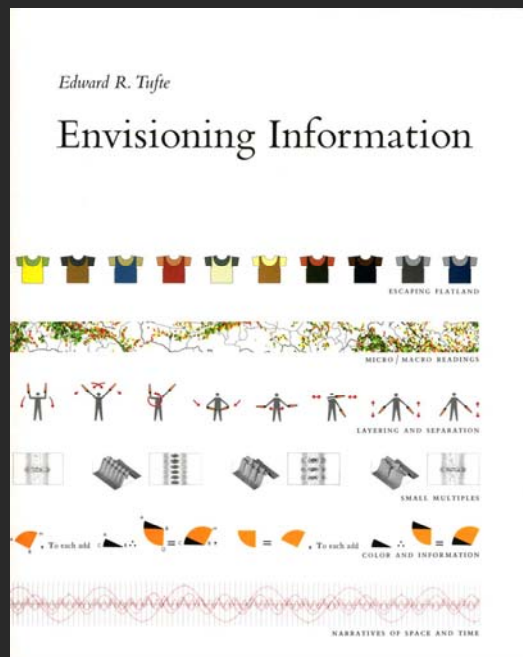
With special thanks to the members at the Cyberinfrastructure for Network Science Center, Mapping Science exhibit map makers and advisory board members, and the VIVO team.

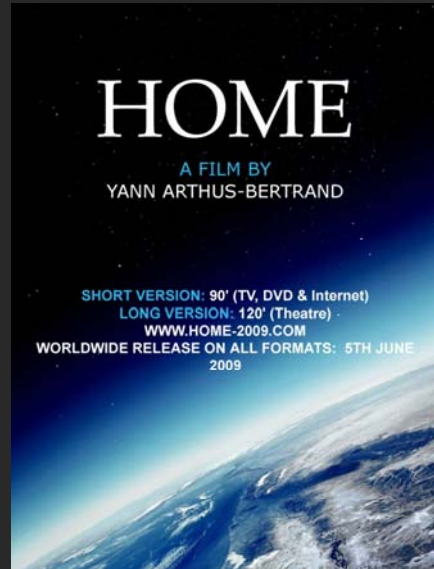
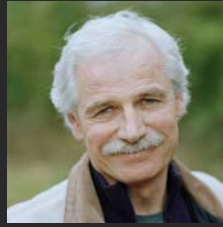
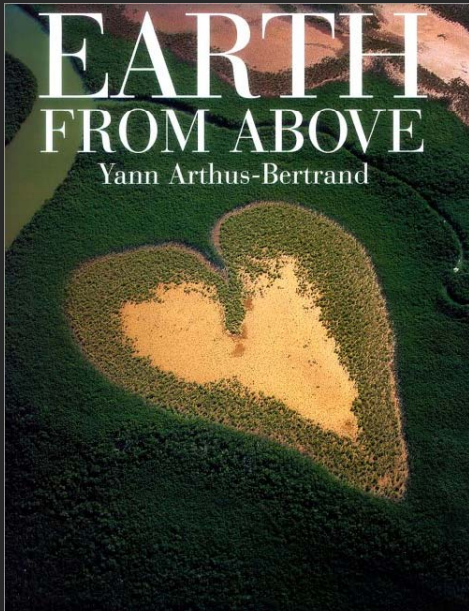
## *THE CHALLENGES OF VISUALISING BIOLOGICAL DATA*

*Workshop run by the UK Biotechnology and Biological Sciences Research Council (BBSRC) and the Arts and Humanities Research Council (AHRC)*

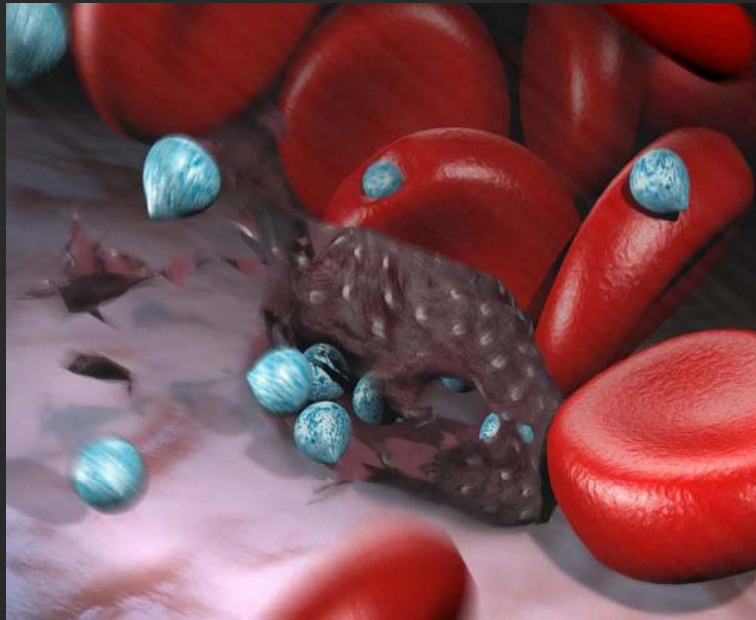
*The Grand by Thistle, Bristol, BS1 2EL*

*November 16-17, 2010*





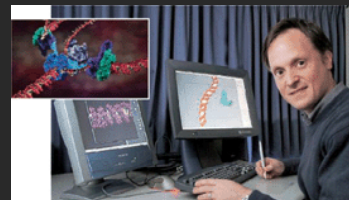
<http://www.home-2009.com>



<http://www.malarialifecycle.com>

The Whole Brain Catalog:

<http://wholebraincatalog.org>

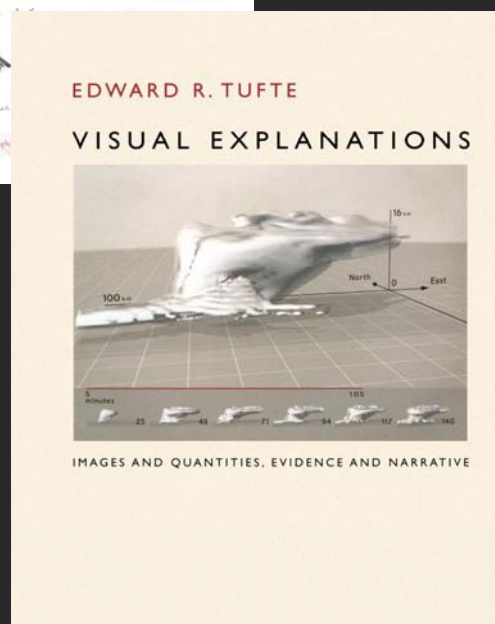
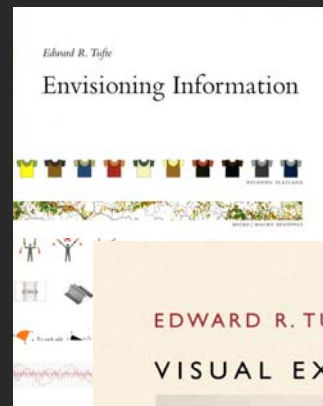
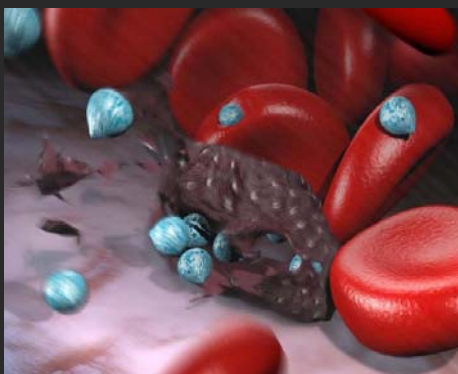
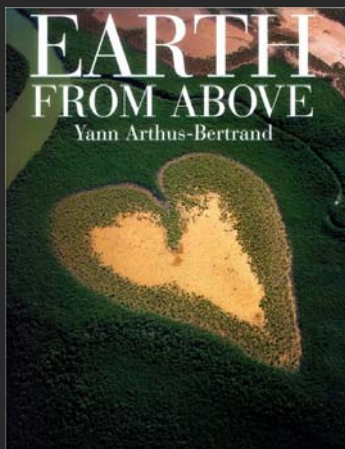


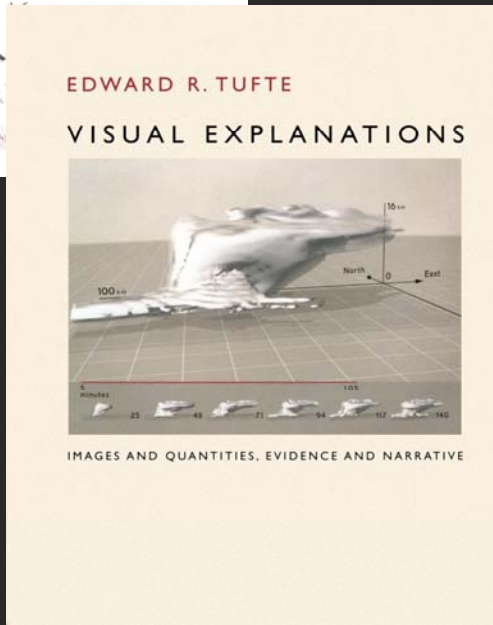
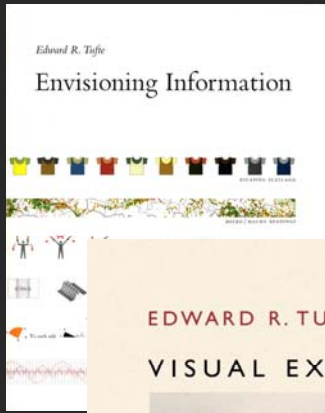
Drew Berry

All three care deeply about

1. Data,
  2. Existing expertise and insight needs, and
  3. Are able to acquire the resources it takes to
- Spent months/years to deeply understand the problem/possible solutions.
  - Render data optimally for the human perceptual-cognitive system – given our current understanding of human perception/cognition and technology.

The result are insightful yet perceptually stunning, intellectually stimulating, and emotion provoking imagery.





Today's massive amounts of streaming data cannot be rendered by hand.

How to use computers to "envision" biomedical science?

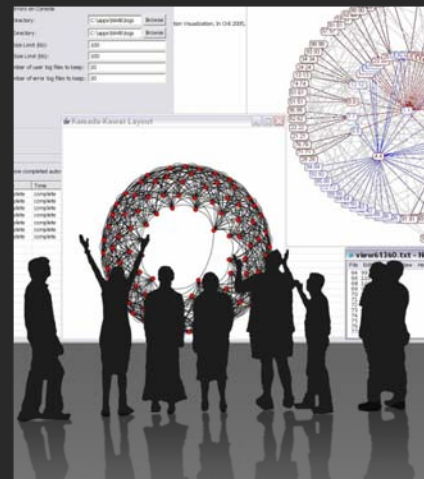
How to combine data mining and visualization algorithms to explore and communicate biomedical science?



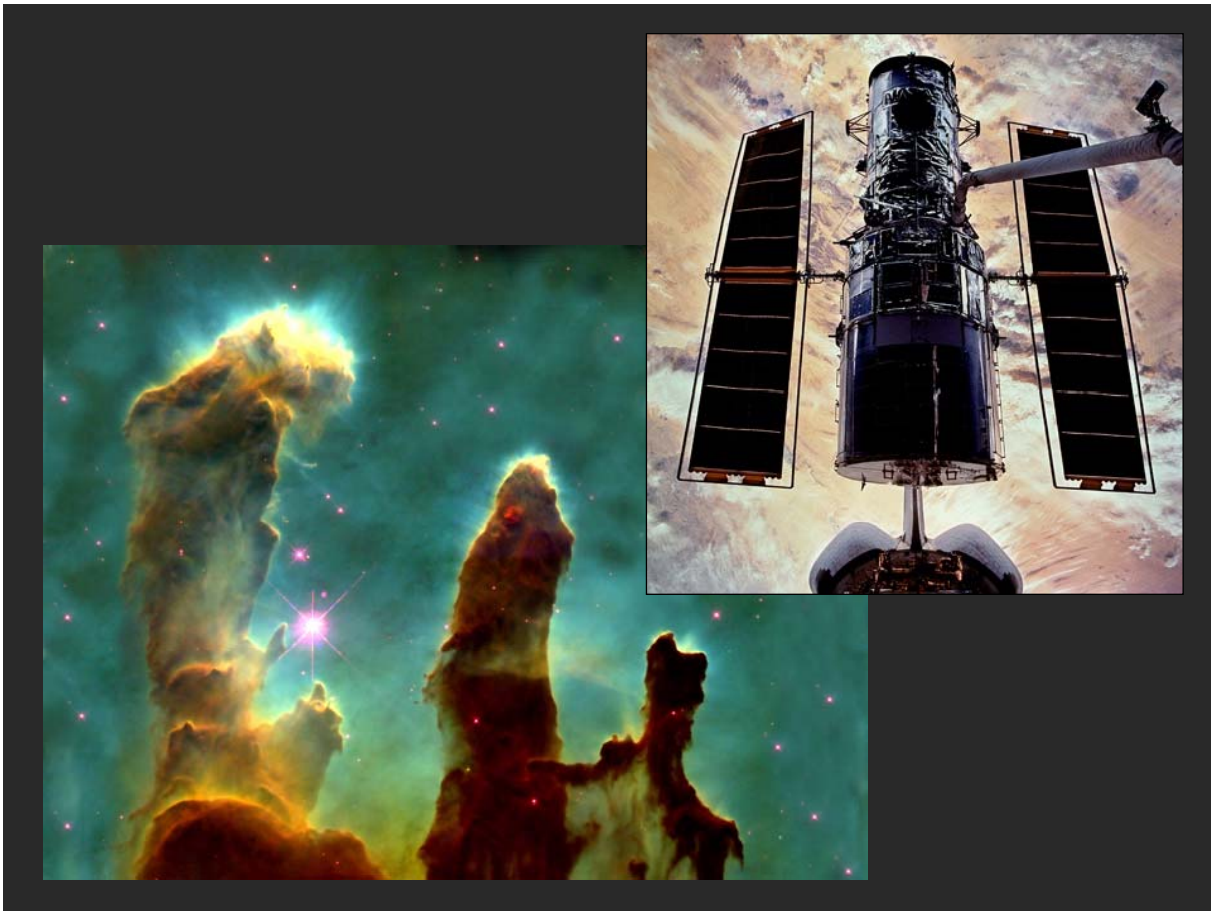
Microscopes



Telescopes

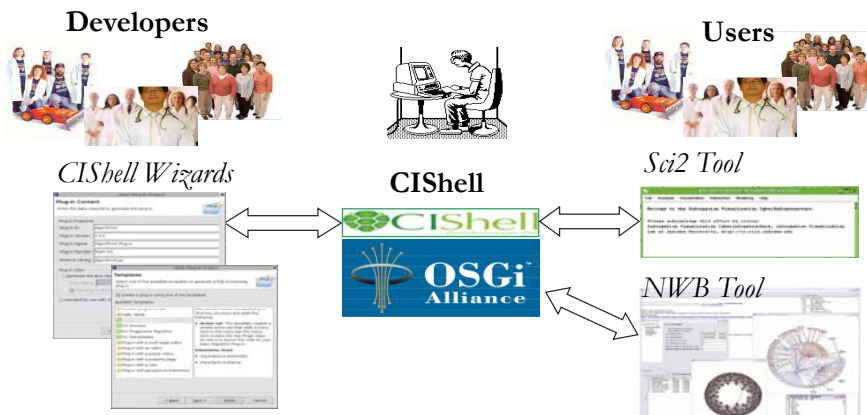


Macrosopes



## Plug-and-Play Macroscopic Design Using OSGi/CIShell

- CIShell (<http://cishell.org>) is an open source software specification for the integration and utilization of datasets, algorithms, and tools.
- It extends the Open Services Gateway Initiative (OSGi) (<http://www.osgi.org>), a standardized, component oriented, computing environment for networked services widely used in industry since more than 10 years.
- Specifically, CIShell provides “sockets” into which existing and new datasets, algorithms, and tools can be plugged using a wizard-driven process.



# Structure of the Remaining Talk

## 1.) Type of Analysis vs. Level of Analysis

Exemplified in Biomedicine

## 2.) Needs-Driven Workflow Design

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

Implementation in different plug-and-play tools/CIs

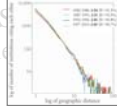

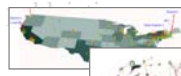
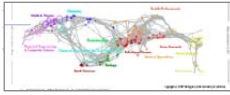
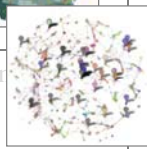

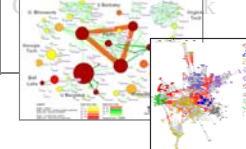



## 1.) Type of Analysis vs. Level of Analysis

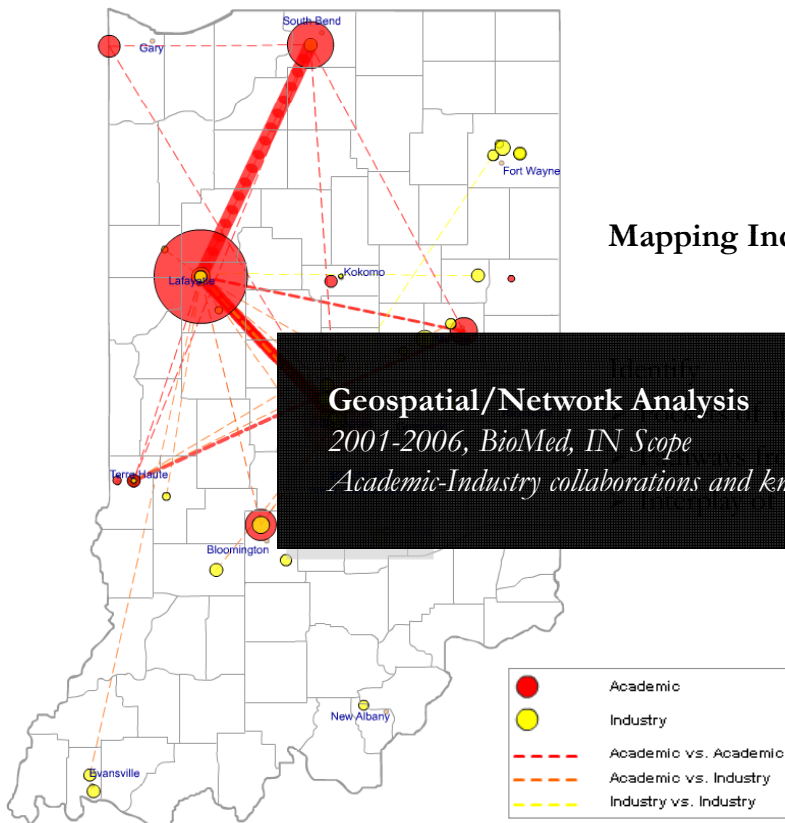
	<i>Micro/Individual</i> (1-100 records)	<i>Meso/Local</i> (101–10,000 records)	<i>Macro/Global</i> (10,000 < records)
<i>Statistical Analysis/Profiling</i>	Individual person and their expertise profiles	Larger labs, centers, universities, research domains, or states	All of NSF, all of USA, all of science.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



## Type of Analysis vs. Level of Analysis

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

13



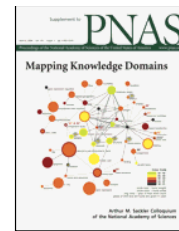
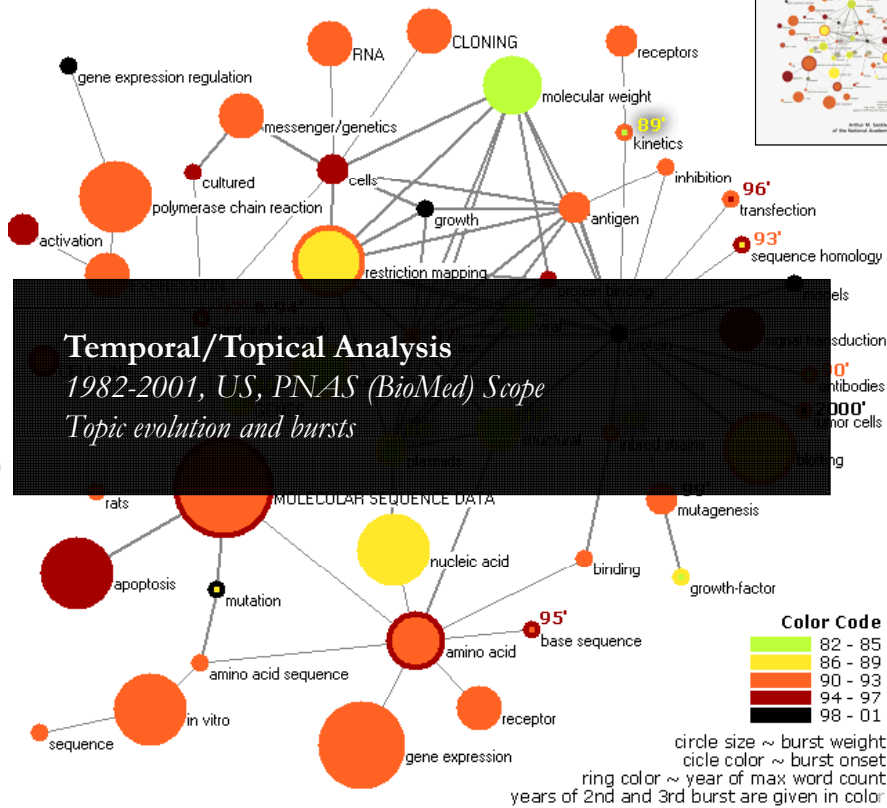
## Mapping Indiana's Intellectual Space

14

## 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.



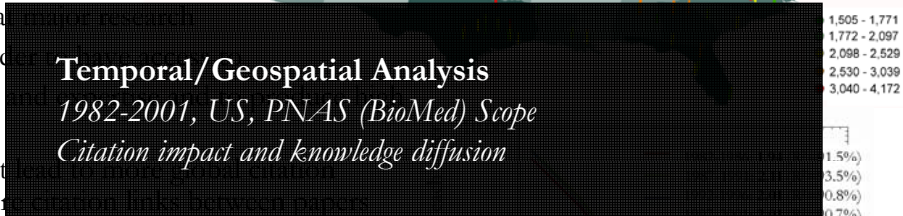
## Spatio-Temporal Information Production and Consumption of Major U.S. Research Institutions

Börner, Katy, Penumarthi, 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



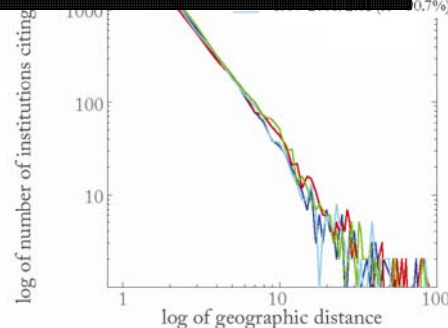
### Research questions:

1. Does space still matter in the Internet age?
2. Does one still have to study and work at institutions in order to produce high quality data quality research?
3. Does the Internet change patterns, i.e., more 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.





# Research Collaborations by the Chinese Academy of Sciences

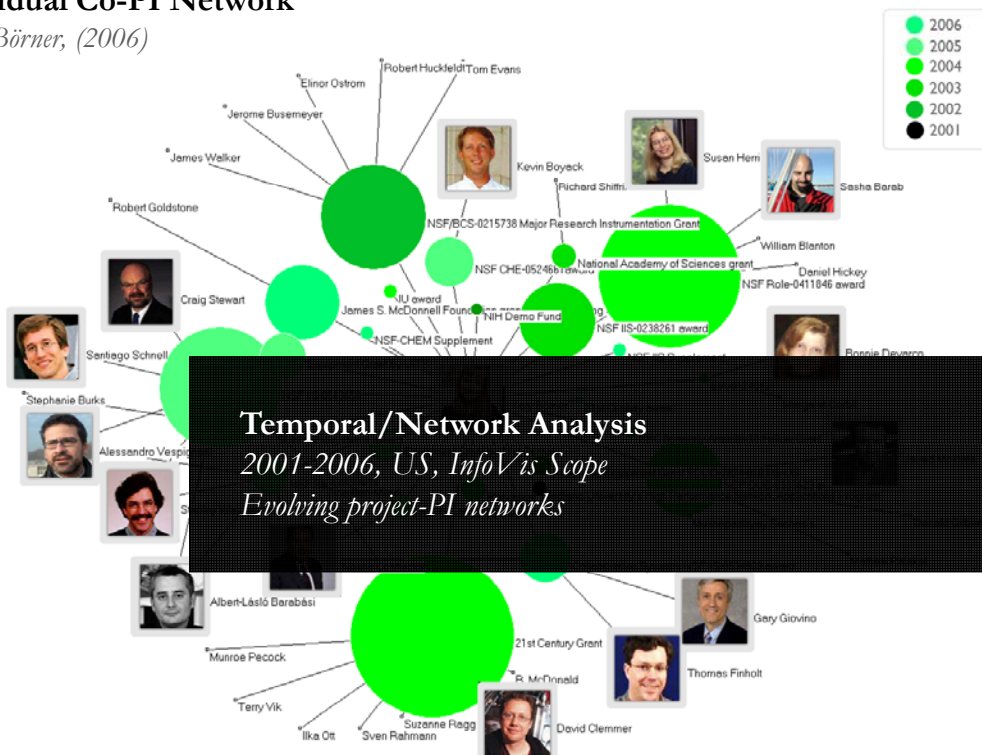
By Weixia (Bonnie) Huang, Russell J. Dubon, Elisha F. Hardy, Katy Börner, Indiana University, USA



This map highlights the research collaborations of the Chinese Academy of Sciences with locations in China and countries around the world. The large geographic map shows the research collaborations of all CAS institutes. Each smaller geographic map shows the research collaborations by the CAS researchers in one province-level administrative division. Collaborations between CAS researchers are not included in the data. On each map, locations are colored on a logarithmic scale by the number of collaborations from red to yellow. The darkest red is 3,395 collaborations by all of CAS with researchers in Beijing. Also, flow lines are drawn from the location of focus to all locations collaborated with. The width of the flow line is linearly proportional to the number of collaborations with the locations it goes to, with the smallest flow lines representing one collaboration and the largest representing differing amounts on each geographic map.

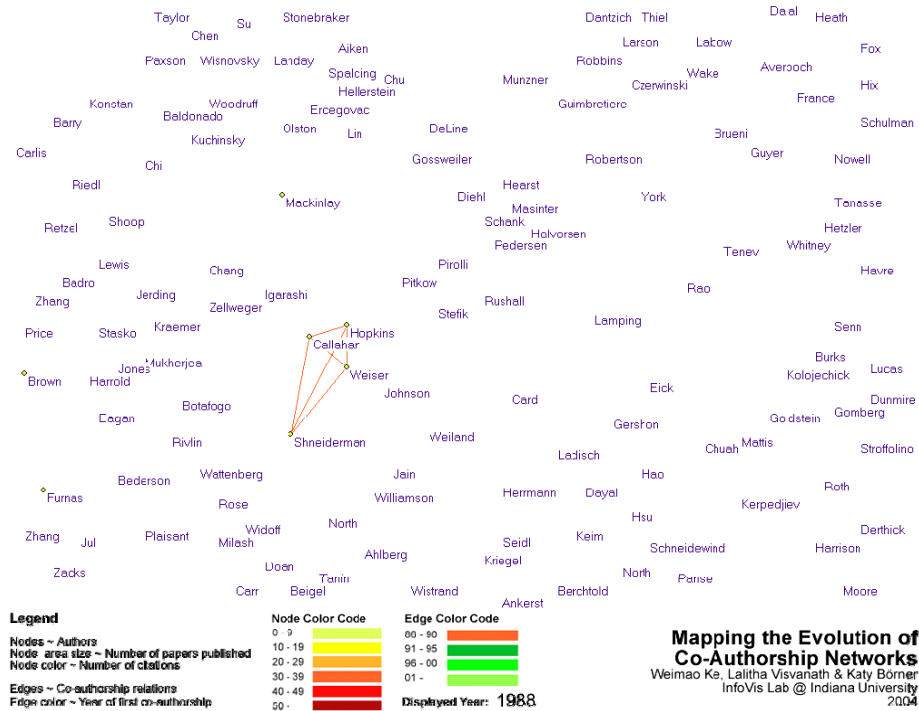
# Individual Co-PI Network

Ke & Börner, (2006)



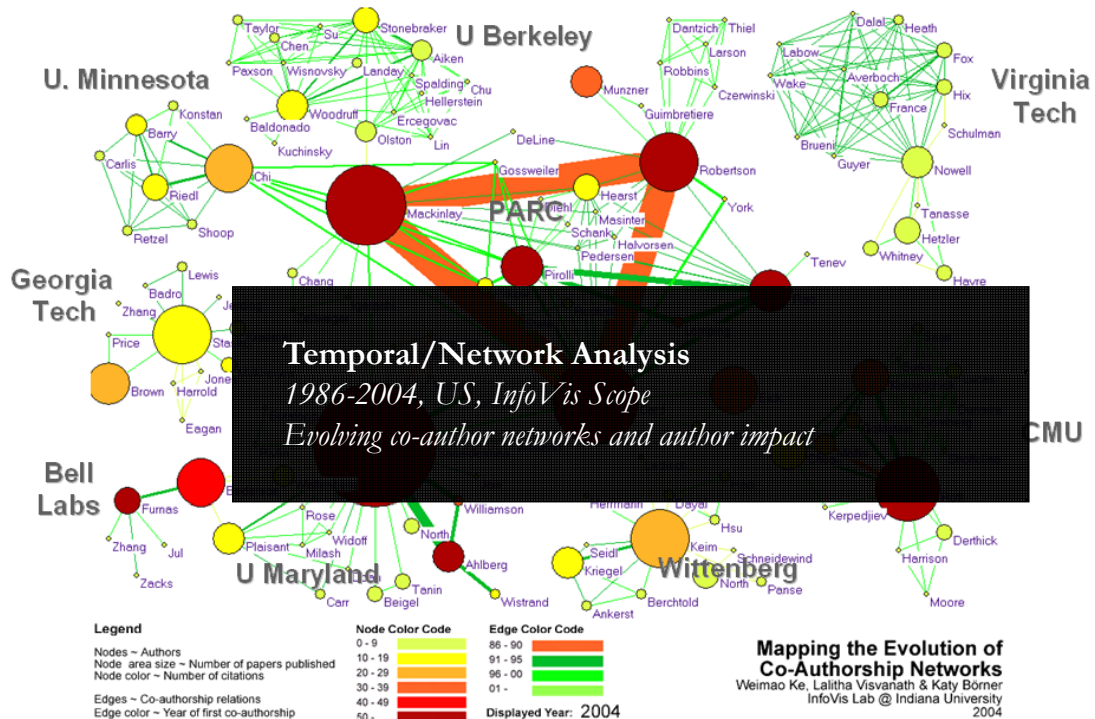
# Mapping the Evolution of Co-Authorship Networks

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

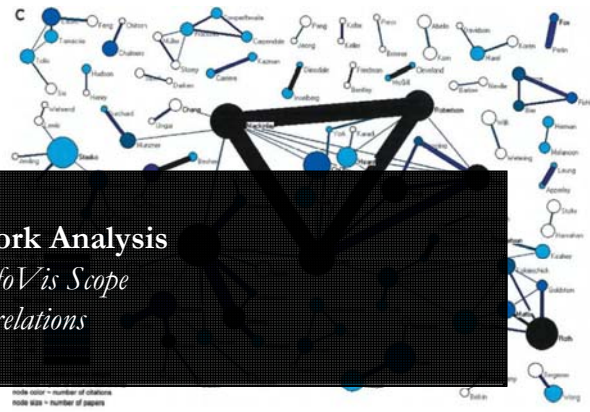
Börner, Dall'Asta, Ke & Vespignani (2005) *Complexity*, 10(4):58-67.

## Research question:

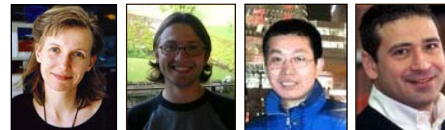
- Is science driven by prolific single experts or by high-impact co-authorship teams?

## Contributions:

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



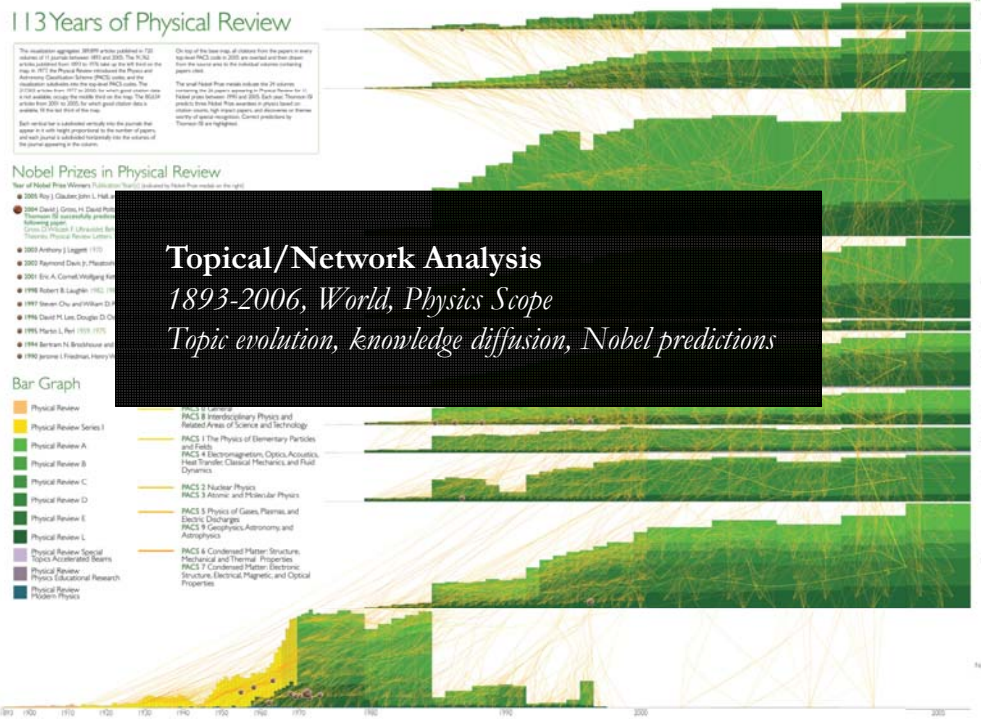
**Temporal/Network Analysis**  
 1986-2004, US, InfoVis Scope  
 Impact of co-author relations



# 113 Years of Physical Review

[http://scimaps.org/dev/map\\_detail.php?map\\_id=171](http://scimaps.org/dev/map_detail.php?map_id=171)

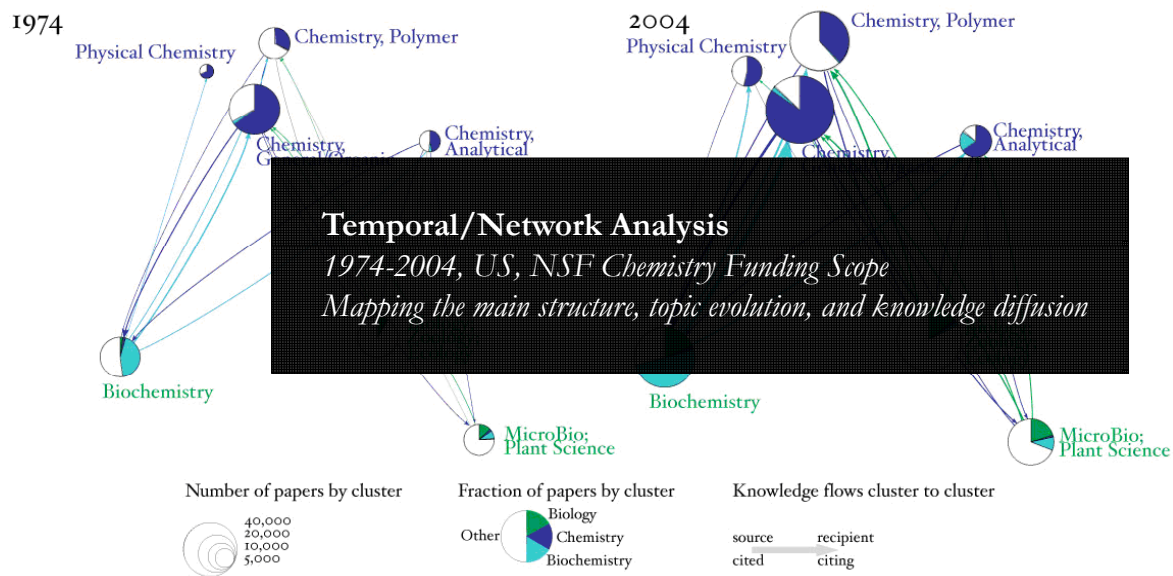
Bruce W. Herr II and Russell Dubon (*Data Mining & Visualization*), Elisha F. Hardy (*Graphic Design*), Shashikant Penumarthy (*Data Preparation*) and Katy Börner (*Concept*)



## Topical Composition and Knowledge Flow Patterns in Chemistry Research for 1974 and 2004

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

### Chemistry - Biology Interface

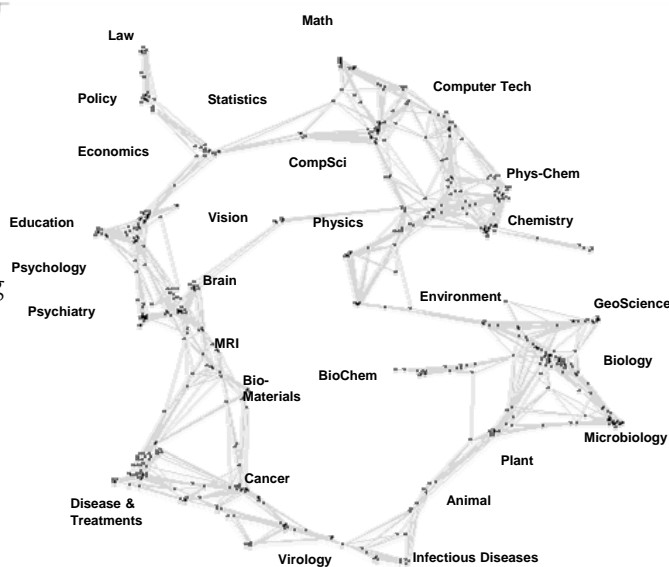


23

### Latest '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

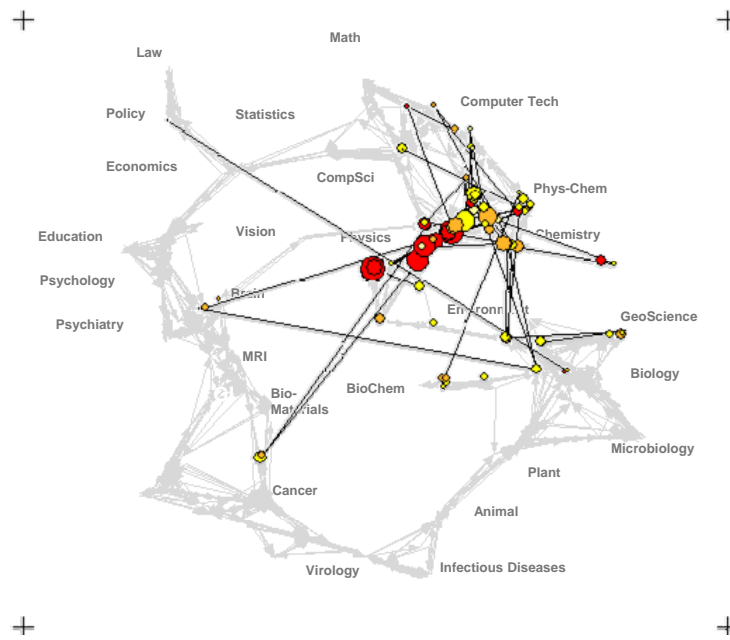


24

## Science map applications: Identifying core competency

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

### Funding patterns of the US Department of Energy (DOE)

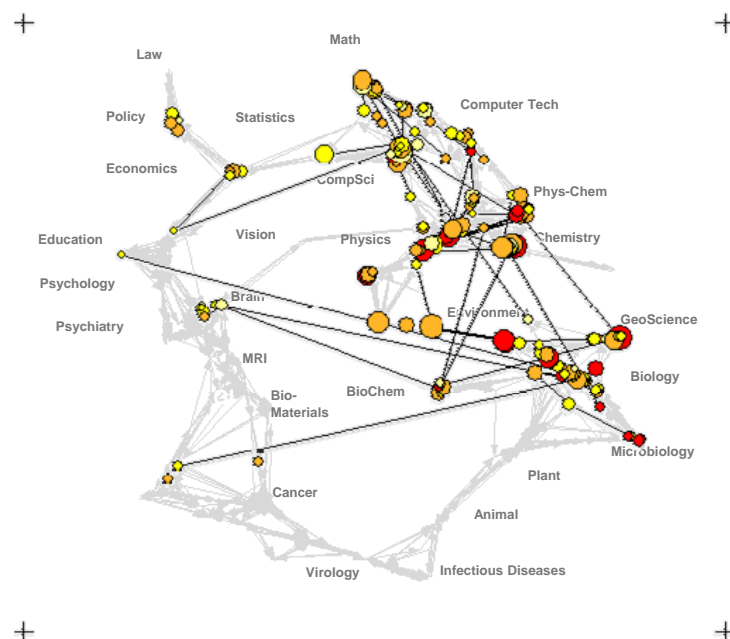


25

## Science map applications: Identifying core competency

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

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

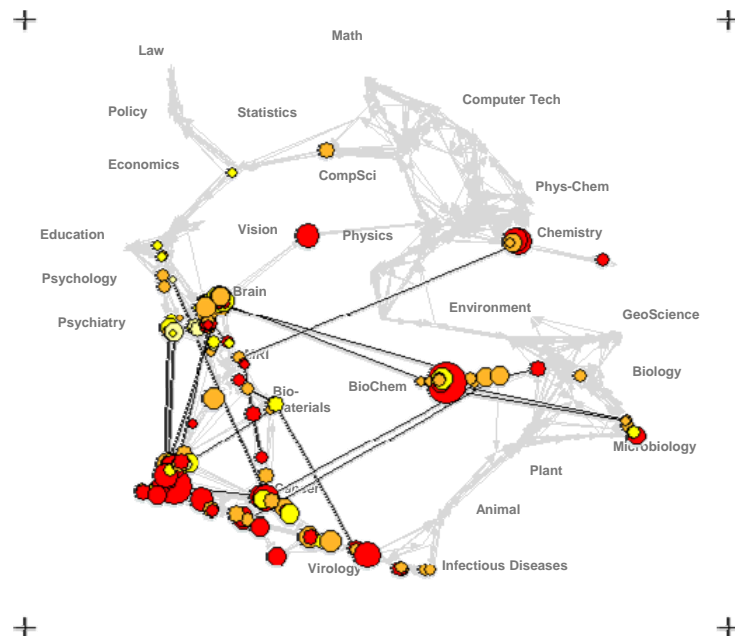


26

## Science map applications: Identifying core competency

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

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

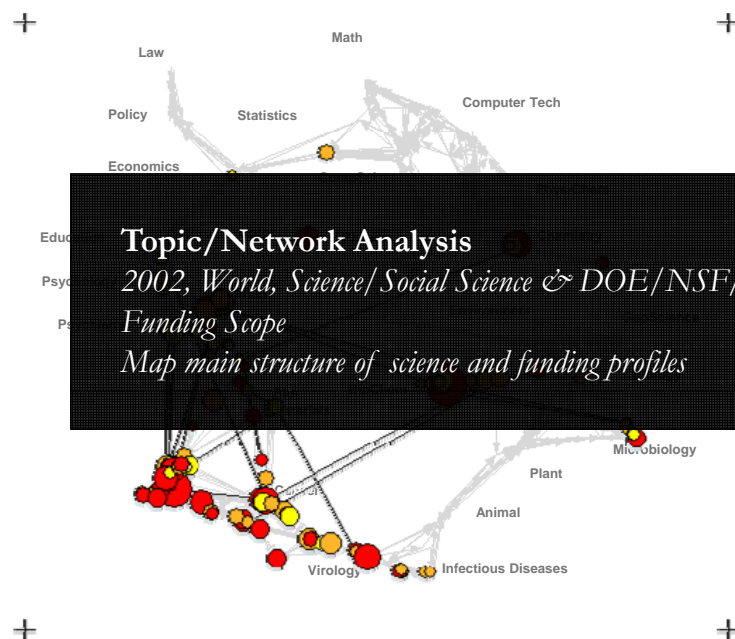


27

## Science map applications: Identifying core competency

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

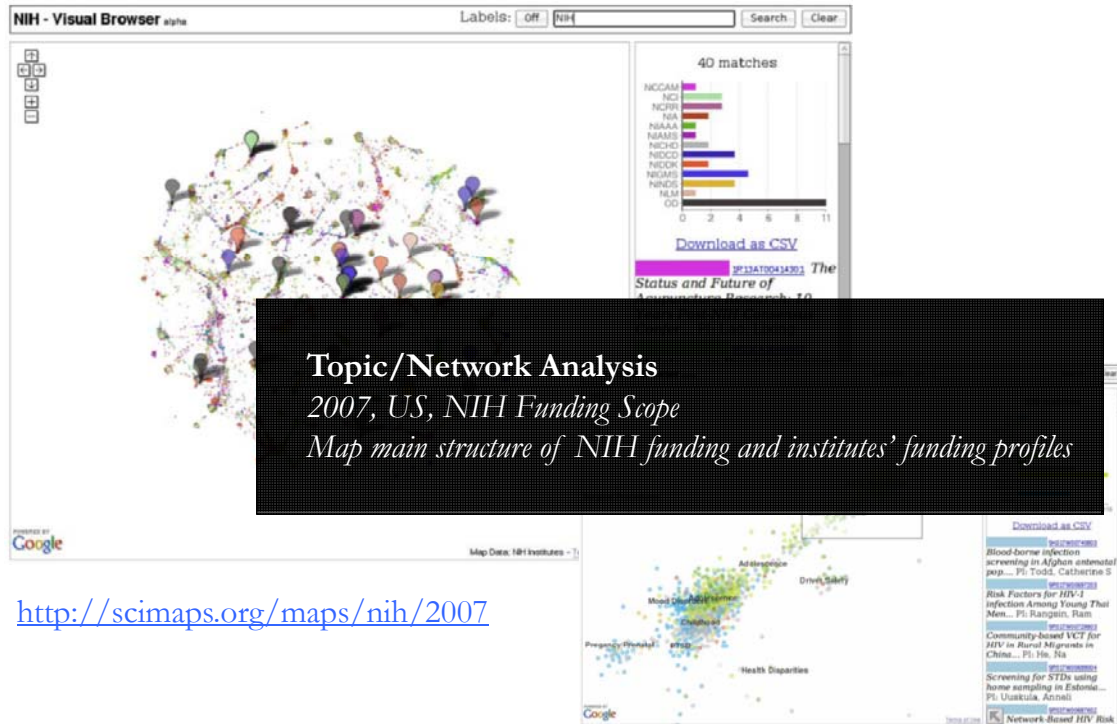
### Funding Patterns of the National Institutes of Health (NIH)



28

## Interactive Science Map of NIH Funding

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



29

## Structure of the Remaining Talk

### 1.) Type of Analysis vs. Level of Analysis

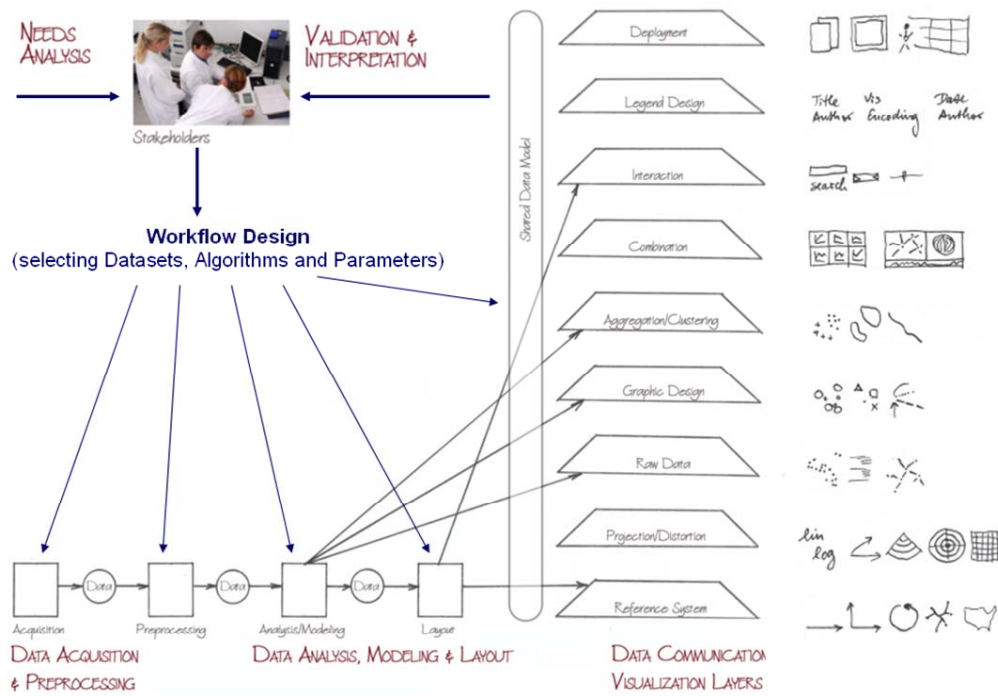
Exemplified in Biomedicine

### 2.) Needs-Driven Workflow Design

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

Implementation in different plug-and-play tools/CIs

2.) **Needs-Driven Workflow Design** using a modular data acquisition/analysis/modeling/ visualization pipeline as well as modular visualization layers.



Börner, Katy (2010) *Atlas of Science*. MIT Press. 31

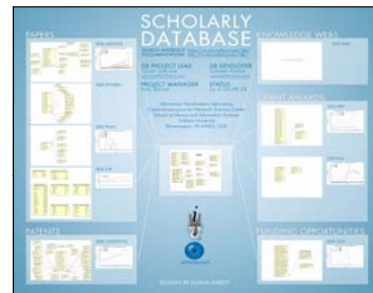
**Computational Scientometrics  
Cyberinfrastructures**



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 (Sci<sup>2</sup>) Tool and CI Portal  
<http://sci.slis.indiana.edu>



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



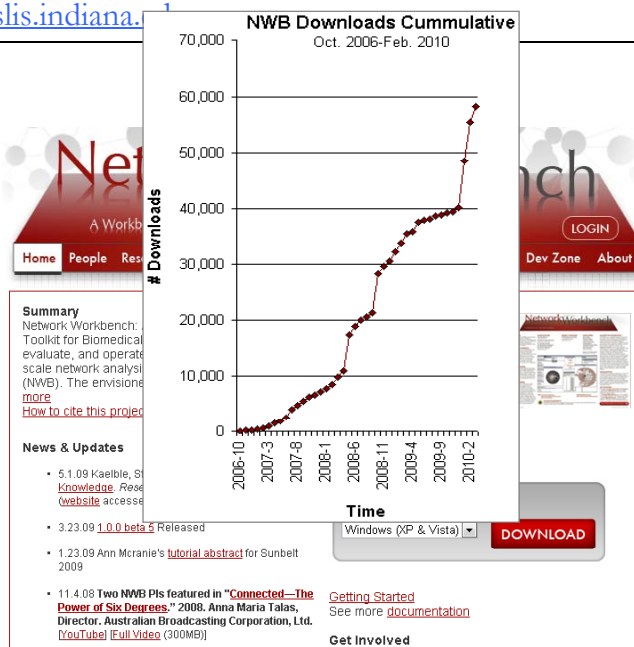


The Network Workbench (NWB) tool supports researchers, educators, and practitioners interested in the study of biomedical, social and behavioral science, physics, and other networks.

In February 2009, the tool provides more than 169 plugins that support the preprocessing, analysis, modeling, and visualization of networks.

**More than 50 of these plugins can be applied or were specifically designed for S&T studies.**

It has been downloaded more than 65,000 times since December 2006.

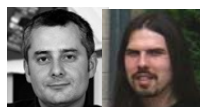


Herr II, Bruce W., Huang, Weixia (Bonnie), Penumarthy, Shashikant & Börner, Katy. (2007). Designing Highly Flexible and Usable Cyberinfrastructures for Convergence. In Bainbridge, William S. & Roco, Mihail C. (Eds.), *Progress in Convergence - Technologies for Human Wellbeing* (Vol. 1093, pp. 161-179), *Annals of the New York Academy of Sciences*, Boston, MA.

## Computational Proteomics

What relationships exist between protein targets of all drugs and all disease-gene products in the human protein-protein interaction network?

Yildirim, Muhammed A., Kwan-II Goh, Michael E. Cusick, Albert-László Barabási, and Marc Vidal. (2007). *Drug-target Network*. *Nature Biotechnology* 25 no. 10: 1119-1126.



**Figure 2** Drug-target network (DT network). The DT network is generated by using the known associations between FDA-approved drugs and their target proteins. Circles and rectangles correspond to drugs and target proteins, respectively. A link is placed between a drug node and a target node if the protein is a known target of that drug. The area of the drug (protein) node is proportional to the number of targets that the drug has (the number of drugs targeting the protein). Color codes are given in the legend. Drug nodes (circles) are colored according to their Anatomical Therapeutic Chemical Classification, and the target proteins (rectangular boxes) are colored according to their cellular component obtained from the Gene Ontology database.

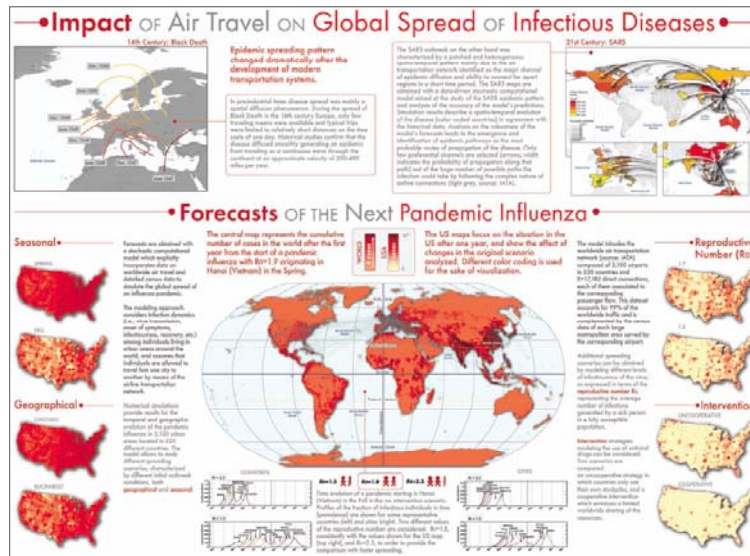
# Computational Epidemics


Forecasting (and preventing the effects of) the next pandemic.

*Epidemic Modeling in Complex realities*, V. Colizza, A. Barrat, M. Barthelemy, A. Vespignani, *Comptes Rendus Biologie*, 330, 364-374 (2007).

*Reaction-diffusion processes and metapopulation models in heterogeneous networks*, V. Colizza, R. Pastor-Satorras, A. Vespignani, *Nature Physics* 3, 276-282 (2007).

*Modeling the Worldwide Spread of Pandemic Influenza: Baseline Case and Containment Interventions*, V. Colizza, A. Barrat, M. Barthelemy, A.-J. Valleron, A. Vespignani, *PLoS-Medicine* 4, e13, 95-110 (2007).





## Sci<sup>2</sup> Tool

A tool for science of science research & practice

Email Address

Password

**Forgot your password?**

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

**Not registered yet?**

[Register now](#)

**Tutorials**

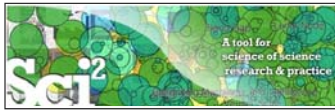
Katy Börner (2010) [Science of Science Research and Tools](#) (12 Tutorials). Reporting Branch, Office of Extramural Research/Office of the Director, National Institutes of Health, Bethesda, MD.

- 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>

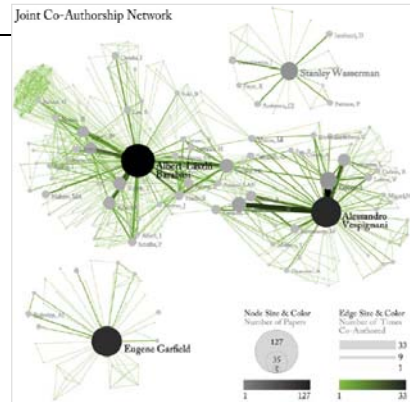
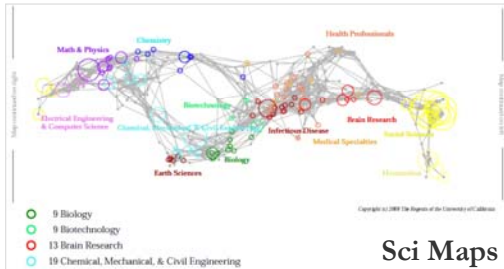
Geetha Senthil (2010) [Multidisciplinary Nature of Work With Reference to PIs and ICs Within a Portfolio](#). PA Group at NIH.

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.



GUESS Network Vis

### Horizontal Time Graphs



Börner, Katy, Huang, Weixia (Bonnie), Linnemeier, Micah, Dubon, Russell Jackson, Phillips, Patrick, Ma, Nianli, Zoss, Angela, Guo, Hanning & Price, Mark. (2009). *Rede-Netzwerk-Red: Analyzing and Visualizing Scholarly Networks Using the Scholarly Database and the Network Workbench Tool. Proceedings of ISSI 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 part by the National Science Foundation, the Network Science center and the School of Li at Indiana University, the National Science Foundation (NSF-0715303), and the James S. McDonnell Cyberscience Center (http://sci.slis.indiana.edu).

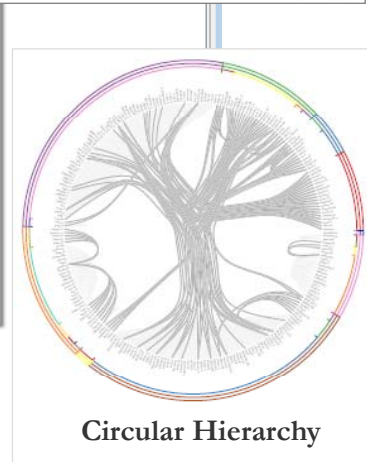
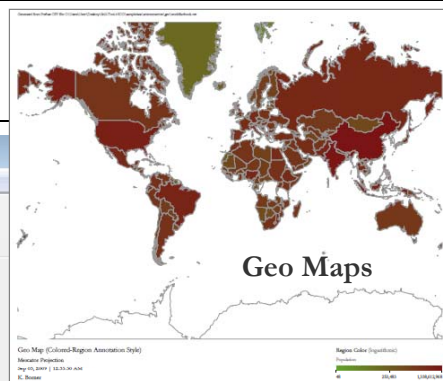
The primary investigators are Katy Börner, in part by the National Science Foundation, the Network Science center and the School of Li at Indiana University, the National Science Foundation (NSF-0715303), and the James S. McDonnell Cyberscience Center (http://sci.slis.indiana.edu).

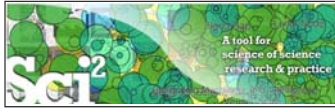
Please cite as follows:  
Sci² Team. (2009). Science of Science Tool. In Proceedings of the 12th International Conference on Scientometrics and Informetrics (ISSI 2009), Rio de Janeiro, Brazil, July 14-17, 2009. Vol. 2, pp. 619-630. http://sci.slis.indiana.edu.

Scheduler

Remove From List  Remove completed

Algorithm Name	Date	Time	% Comp
Extract Co-Author Network	09/03/2009	00:15:20 AM	100%
Load and Clean ISI File	09/03/2009	00:15:05 AM	100%





## Sci<sup>2</sup> 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-Coreeness

-----  
HTTS

### 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

### HTTS

### Unweighted & Directed

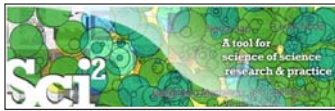
Node Indegree  
Node Outdegree  
Indegree Distribution  
Outdegree Distribution

-----  
K-Nearest Neighbor  
Single Node in-Out Degree Correlations

-----  
Dyad Reciprocity  
Arc Reciprocity  
Adjacency Transitivity

-----  
Weak Component Clustering  
Strong Component Clustering

39



## Sci<sup>2</sup> Tool: Algorithms cont.

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

-----  
Extract K-Core  
Annotate K-Coreeness

-----  
HTTS  
PageRank  
Weighted & Directed  
HTTS  
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:

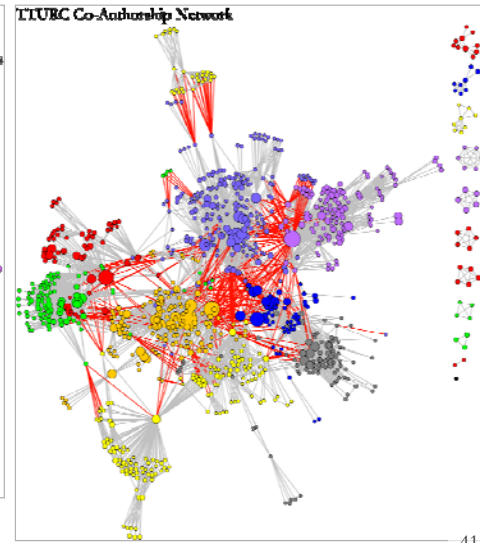
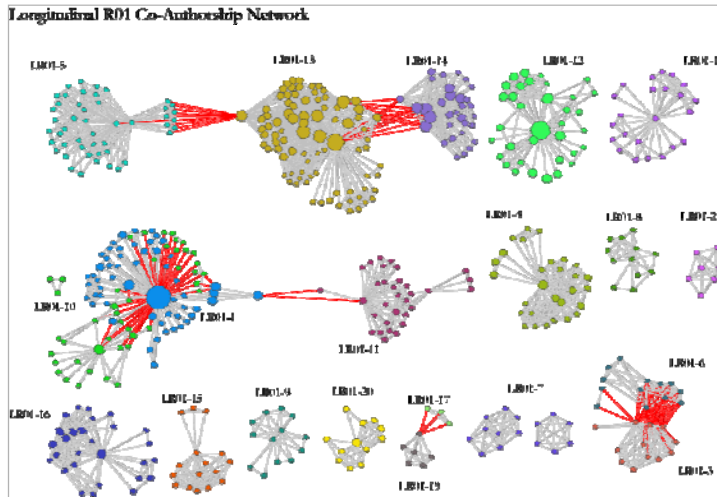
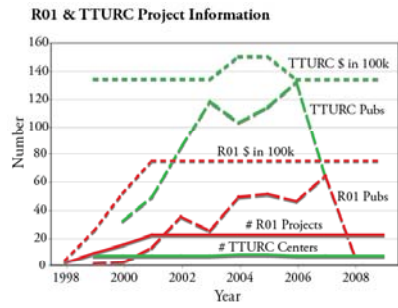
Database support for ISI and NSF data.

40

## 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*.



## Interactive Science Map of NIH Funding

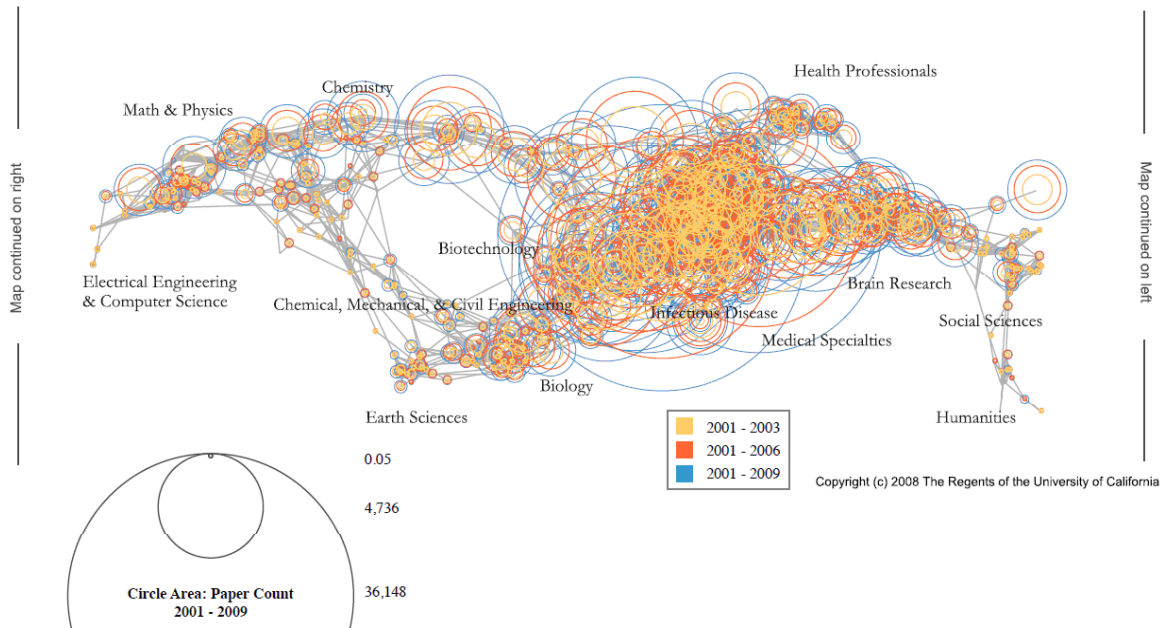
Herr II, Bruce W., Talley, Edmund M, Burns, Gully APC, Newman, David & La Rowe, Gavin. (2009).



<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.



43

# 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 | Geographic | Visualization of Job Postings | Geographic Visualization | Postdoc at Harvard Medical School | Link to Post

Map of Science | Geographic | Visualization of Job Postings

Map of Science | Geographic | 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 postings from large, inh

Map of Science

Math and Physics  
Chemistry  
Health Professionals  
Biotechnology  
Brain Research  
Infectious Disease  
Medical Specialties  
Social Sciences  
Humanities  
Biology  
Earth Sciences  
Chemical, Mechanical, and Civil Engineering

Search for Jobs

Biotechnology

Search

POWERED BY Google

Copyright © 2008 The Regents of the University of California - Terms of Use

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

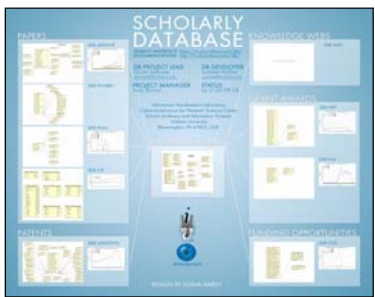
44


# Computational Scientometrics Cyberinfrastructures

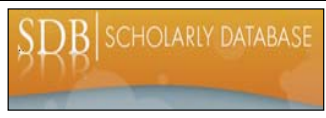
 **Scholarly Database: 25 million scholarly records**  
<http://sdb.slis.indiana.edu>

 James S. McDonnell Foundation

 **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/>



## 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.

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

**IU User**  
 IU users must login using the Central Authentication Service (CAS), the standard IU authentication system. Please click the button below to proceed to the IU login page.

**Non-IU User**  
 Email: \_\_\_\_\_  
 Password: \_\_\_\_\_  
 Login

**Not Registered Yet?**  
 Register as an IU User  
 Register as a Non IU User

**In the News**  
 Whitfield, Julia. 2008. *Emerg. Infect. Diseases*, 4(5), 9: 720-723.

**Please Cite As**  
 La Breve, Sam, Ambler, Susan, Pringle, John, He, Weimin and Bollen, Katy. (2007) The Scholarly Database and Its Utility for Scientometrics Research. In Proceedings of the 11th International Conference on Scientometrics and Information Science, June 29-31, 2007, pp. 493-497.  
<http://sdb.slis.indiana.edu/~kaly/paper/11%20icis%20sdb.pdf>

**Acknowledgments**  
 The Scholarly Database is funded by the School of Library and Information Science and the Cyberinfrastructure for Network Science Center at Indiana University, the National Science Foundation under Grants No. IRI-0338261 and IRI-0513430, and a James S. McDonnell Foundation grant in area: Building Complex Systems.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

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

Search Edit Profile Admin About Logout

**Search**

Creators: \_\_\_\_\_  
 Title: \_\_\_\_\_  
 Abstract:  Full  
 Full Text: \_\_\_\_\_  
 First Year: 1999  
 Last Year: 2009

**Health (1999 - 2009)**  
 **ICIS (1999 - 2007)**  
 **ISI (1999 - 2004)**  
 **ISI/ISI (1979 - 2007)**

Search

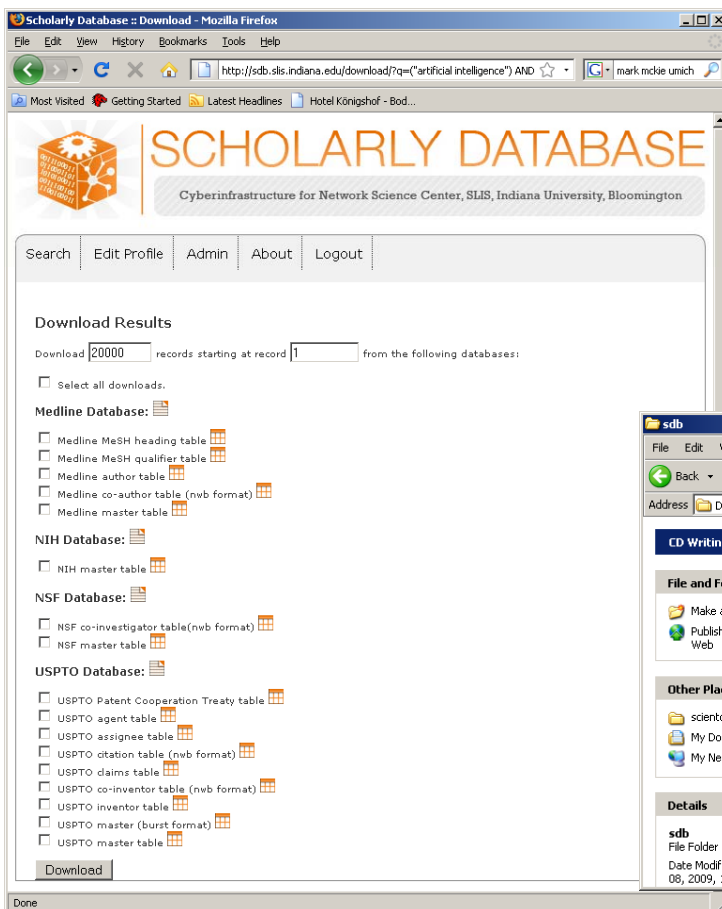
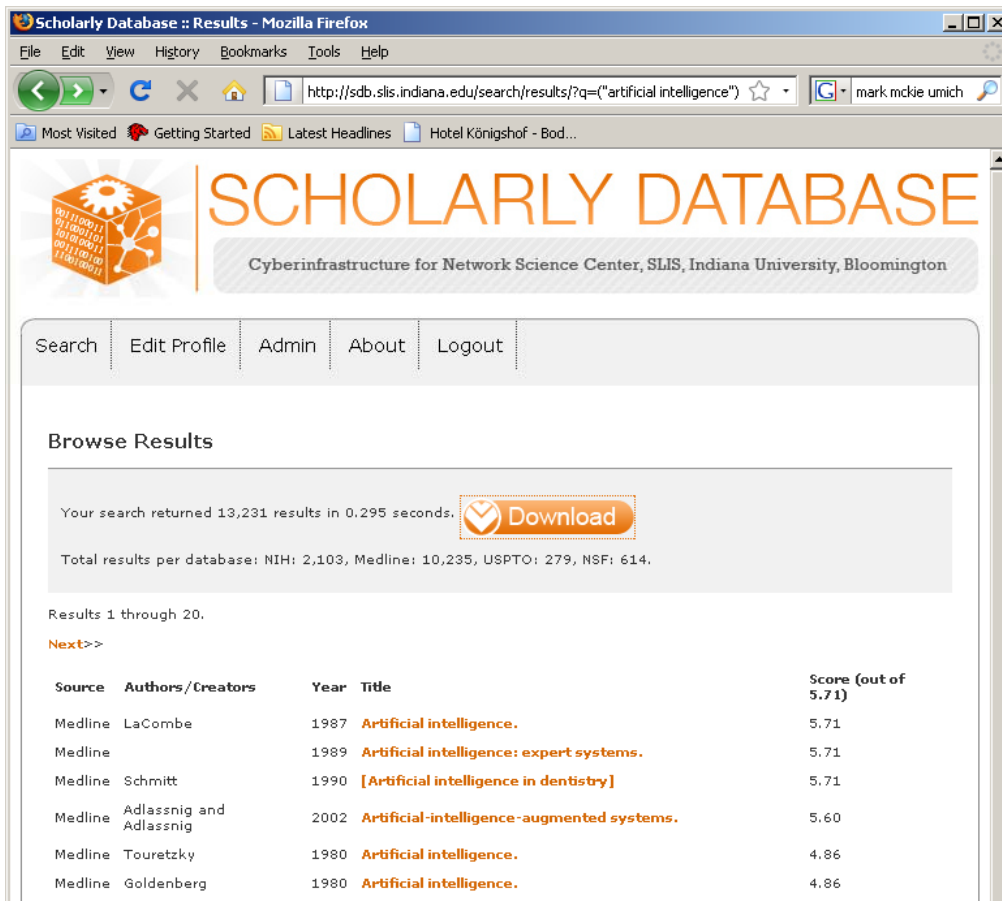
If multiple terms are entered in a field, they are automatically combined using 'OR'. So, 'breast cancer' matches any record with 'breast' or 'cancer' in that field.

You can put AND between terms to combine with 'AND'. Thus 'breast AND cancer' would only match records that contain both terms.

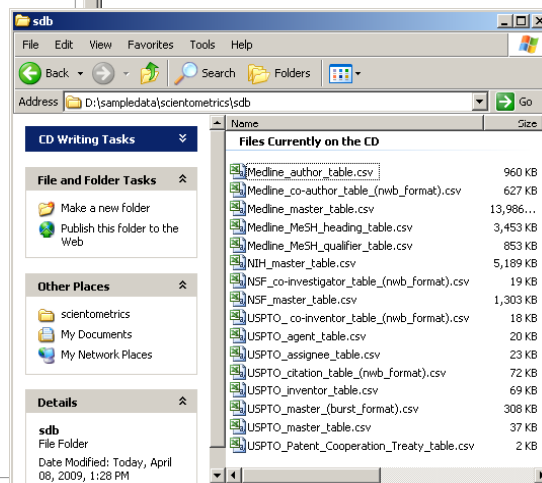
Double quotation can be used to match compound terms, e.g. "breast cancer" returns records with the phrase "breast cancer", and not records where 'breast' and 'cancer' are both present, but not the exact phrase.

The importance of a particular term in a query can be increased by putting a "\*" and a number after the term. For instance, "breast cancer" \*2 would increase the importance of matching the term "cancer" to two compared to matching the term "breast".

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



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



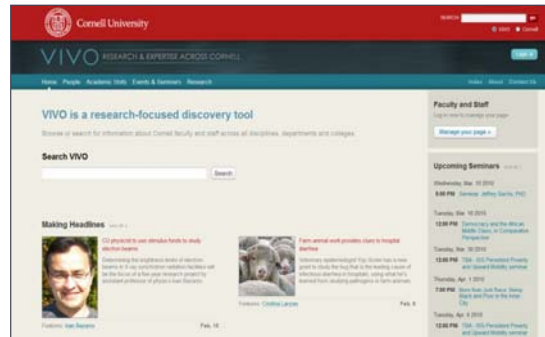


## VIVO: A Semantic Approach to Creating a National Network of Researchers (<http://vivoweb.org>)

- 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 Bushhausen, 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.

## VIVO Users and Needs

- Faculty/Researchers
  - Customize profile created via feeds; find potential collaborators, “people like me”; discovery via high search rankings; info on activity of colleagues...
- Students
  - Create profiles; easily find mentors + collaborators; locate facilities, events, funding opportunities...
- Administrators
  - Quickly find cross-disciplinary expertise (research area; geography); centralize public data from diverse sources; easily repurpose information for consumers; improve faculty collaboration within or across departments and institutions...
- Funding, donor, legislative agencies
  - Discover projects, grants, expertise (e.g. for review panels; targets for funding)...
- General public
  - Find expertise, learn about research in a region/institution...

## VIVO Web Pages

The screenshot shows the Cornell University CALS Research website. At the top, there is a search bar and navigation links for various research areas: New Life Sciences, Environmental Sciences, Land-Grant Mission, Applied Social Sciences, and All CALS Research. Below this is a 'Welcome to CALS Research' section with a sub-header 'A portal to all CALS Research activities. Select the individual academic priority areas in the banner above for a more filtered view.' The main content area is titled 'CALS Research News' and lists recent news items with small images and headlines. The URL <http://research.cals.cornell.edu> is displayed at the bottom of the page.

## VIVO Data Providers & Users

- Eagle-i (“enabling resource discovery” U24 award)
- Federal agencies – NIH (NIH RePORTER), NSF, USDA, ...
- Search Providers – Google, Bing, Yahoo, ...
- Professional Societies – AAAS, ...
- Publishers/vendors – PubMed, Elsevier, Collexis, ISI...
- Semantic Web community – DERI, ...
- Consortia of schools – SURA, CTSA...
- Producers, consumers of semantic web-compliant data

# VIVO & Linked Open Data

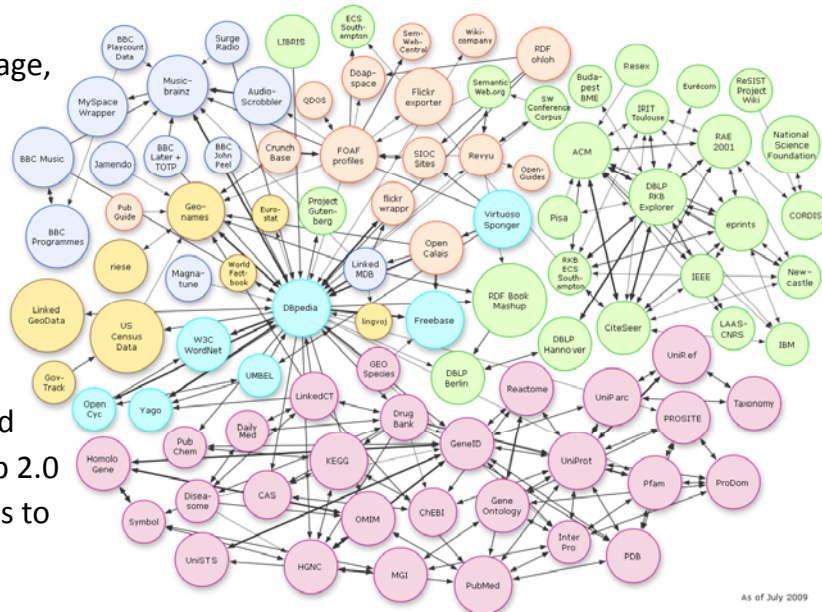
2010 National VIVO Conference August 12&13, NYC

<http://conferences.dce.ufl.edu/vivo>

VIVO makes high coverage, high quality data from systems of record

- available online
- for free, and
- in machine readable format.

VIVO ontology is aligned with many existing Web 2.0 and scholarly ontologies to ease interoperability.



As of July 2009

[http://www4.wiwiw.fu-berlin.de/bizer/pub/lod-datasets\\_2009-07-14\\_colored.png](http://www4.wiwiw.fu-berlin.de/bizer/pub/lod-datasets_2009-07-14_colored.png)

**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  
87 co-author(s) from 2001 to 2010

**Co-Author Network**

13 Co-author(s)  
26 Co-author (INCL)

**Borner, Katy**

35 Publications  
13 Co-author(s)  
2007 First Publication  
2007 Last Publication

**Legend**

No. of publications  
No. of links (co-authored)

**Interact**

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

**The 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.  
13 out of 87 co-author(s) are shown.

**Tables**

Year	Publications per Year	Co-author Publications with Borner, Katy
2001	2	
2002	4	Chen C. 5
2003	2	Stoyck H.W. 4
2004	7	Mans K.K. 4
2005	7	Ha V.V. 3
2006	3	Pencumarty S. 3
2007	10	Venugopalan, Alessandro 2 Hao B. 2 Hardy E. 2 Holloway T. 2 Hao B.P.W. 2 Thurston S. 2 Fang Y. 2 Mans K. 2

## 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%2Fontology%2Fcore%2FPerson72&vis=person\\_level&render\\_mode=standalone](http://vivo-vis.slis.indiana.edu/vivo1/visualization?uri=http%3A%2F%2Fvivo.berkeley.edu%2Fontology%2Fcore%2FPerson72&vis=person_level&render_mode=standalone)

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.; Sooriarmurthi R.; Maru J.T.; Shiffrin R.M.; Mane K.; Moor K.A.;

Co-author network (GraphML File)

```

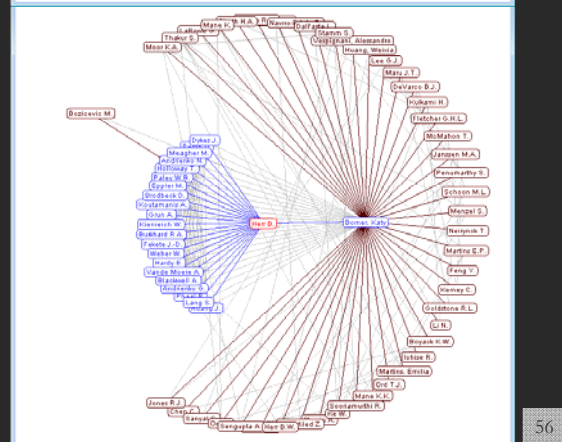
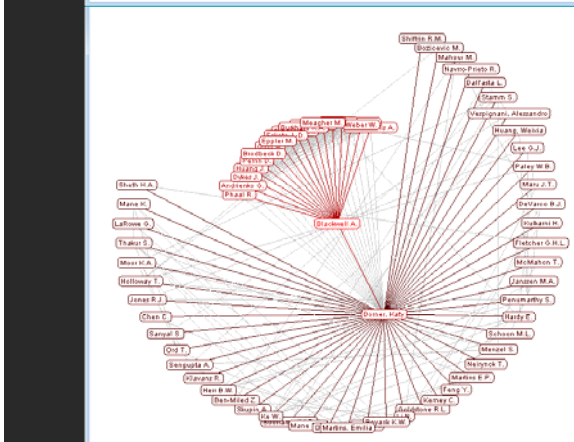
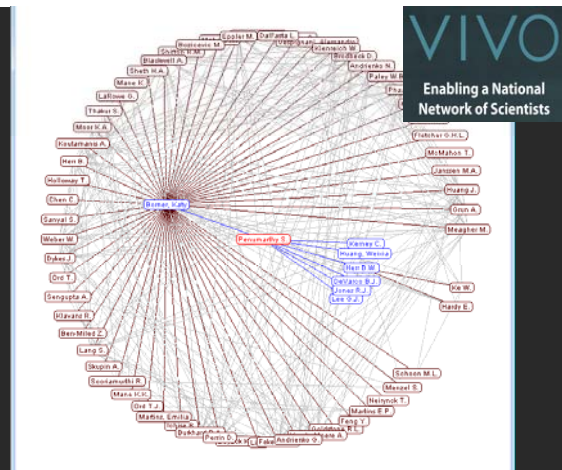
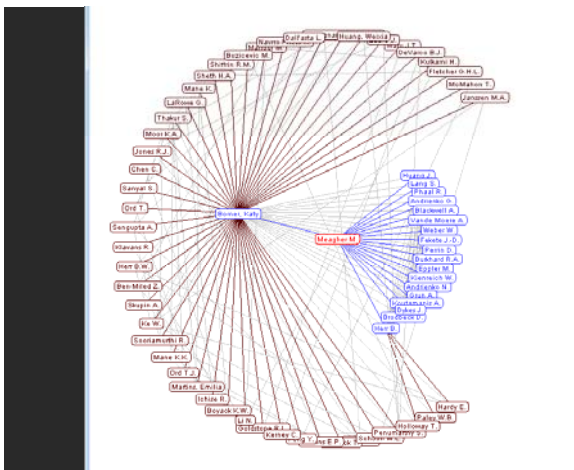
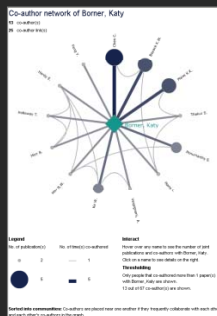
1 <?xml version="1.0" encoding="UTF-8"?>
2 <graphml xmlns="http://graphml.graphdrawing.org/xmlns"
3 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
4 xsi:schemaLocation="http://graphml.graphdrawing.org/xmlns
5 http://graphml.graphdrawing.org/xmlns/1.0/graphml.xsd">
6 <key id="label" for="node" 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.

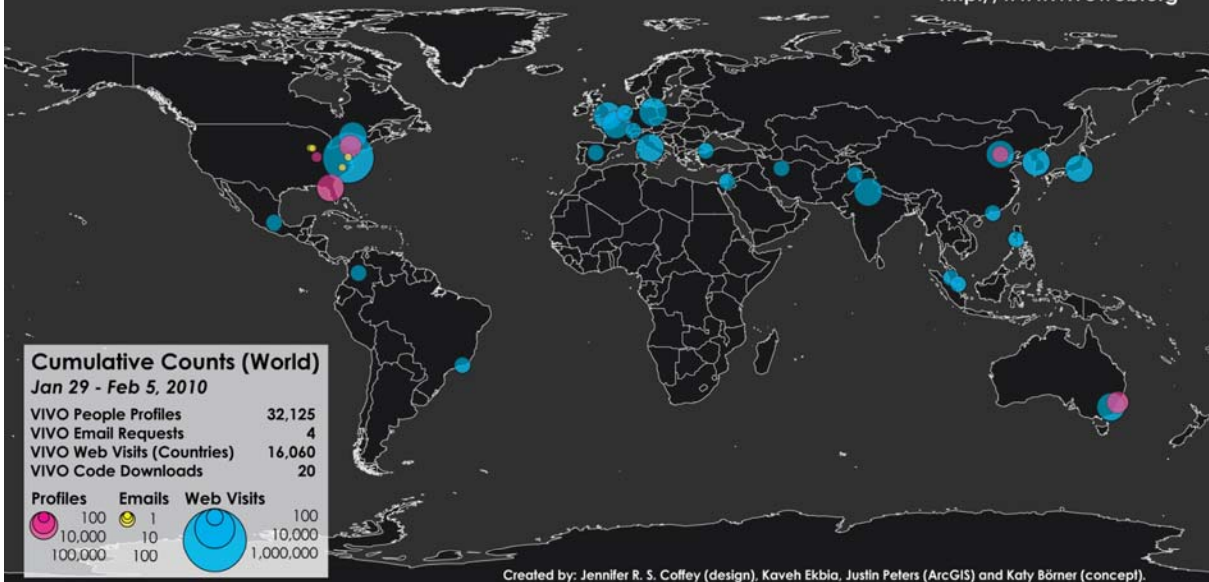
Co-authors (.CSV File)

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



# VIVO Enabling National Networking of Scientists

<http://www.vivoweb.org>

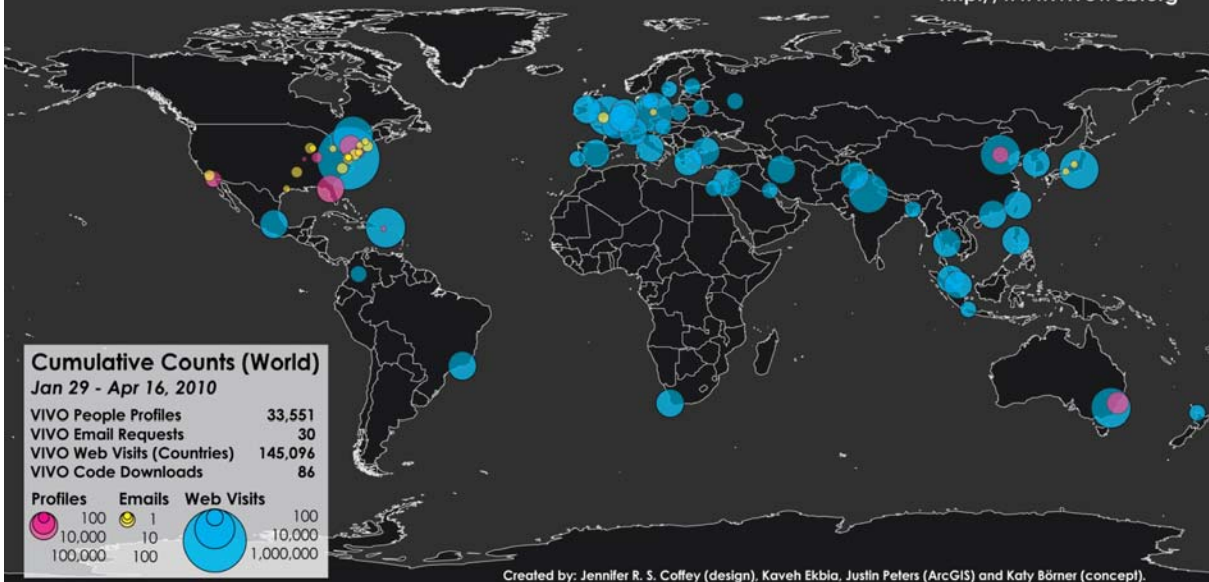


Science is global. World view of VIVO activity.  
Web site visits are aggregated at the country level.

57

# VIVO Enabling National Networking of Scientists

<http://www.vivoweb.org>



Shown are the

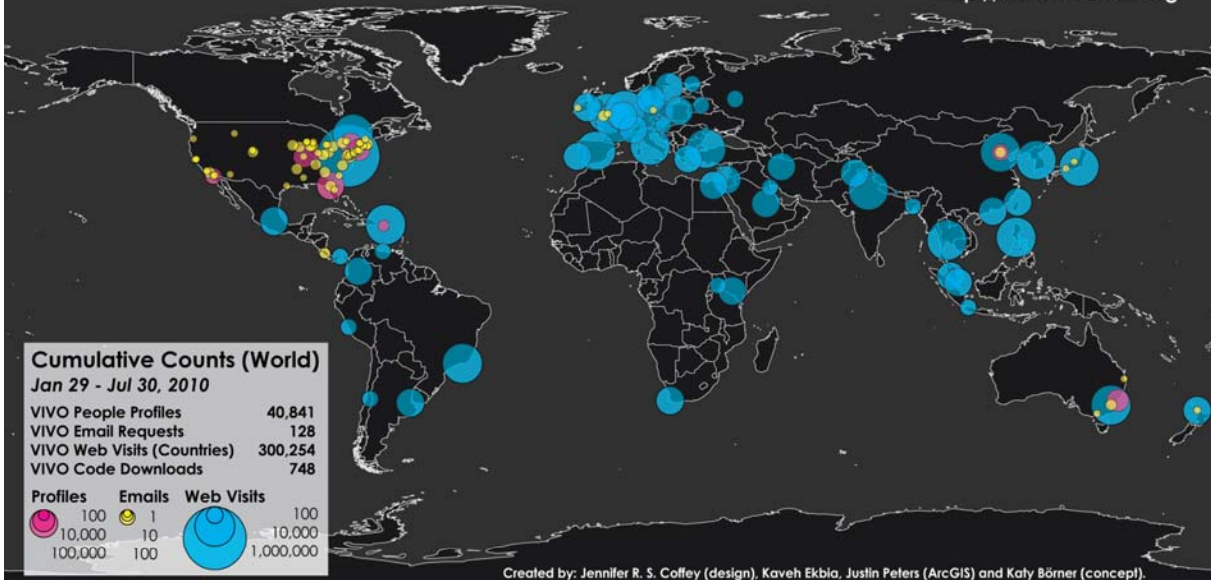
- Number of people profiles in the 7 different VIVO installation sites plus CAS and U Melbourne.
- Email contacts by data and service providers as well as institutions interested to adopt VIVO.
- The number of visitors on <http://vivoweb.org>

Circles are area size coded using a logarithmic scale.

58

# VIVO Enabling National Networking of Scientists

<http://www.vivoweb.org>



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

87 downloads by June 11, 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.

59

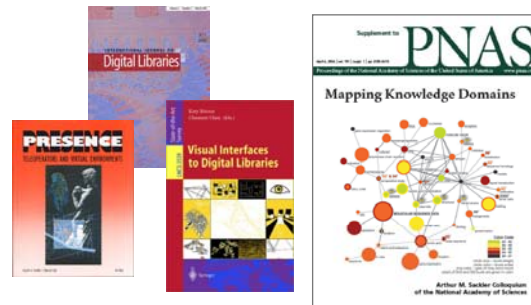
## Computational Scientometrics References

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/](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>



60

## Mapping Science Exhibit – 10 Iterations in 10 years

<http://scimaps.org>

### The Power of Maps (2005)



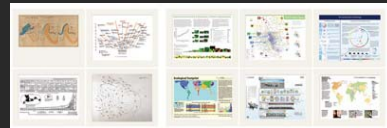
### The Power of Reference Systems (2006)



### The Power of Forecasts (2007)



### Science Maps for Economic Decision Makers (2008)



### Science Maps for Science Policy Makers (2009)



### Science Maps for Scholars (2010)

### Science Maps as Visual Interfaces to Digital Libraries (2011)

### Science Maps for Kids (2012)


### Science Forecasts (2013)


### How to Lie with Science Maps (2014)

Exhibit has been shown in 72 venues on four continents. Currently at  
- NSF, 10th Floor, 4201 Wilson Boulevard, Arlington, VA  
- Marston Science Library, University of Florida, Gainesville, FL  
- Center of Advanced European Studies and Research, Bonn, Germany  
- Science Train, Germany.



61

 **cyberinfrastructure for NETWORK SCIENCE CENTER**  
School of Library and Information Science | Indiana University Bloomington



People Events Jobs Contact  
Cyberinfrastructures Teaching Outreach Visiting Artists Funding

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



Debut of 5<sup>th</sup> Iteration of Mapping Science Exhibit at MEDIA X was on May 18, 2009 at Wallenberg Hall, Stanford University, <http://mediax.stanford.edu>, <http://scaleindependentthought.typepad.com/photos/scimaps>

## Illuminated Diagram Display

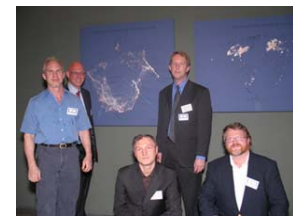
W. Bradford Paley, Kevin W. Boyack, Richard Kalvans, and Katy Börner (2007)  
*Mapping, Illuminating, and Interacting with Science. SIGGRAPH 2007.*

### Questions:

- Who is doing research on what topic and where?
- What is the 'footprint' of interdisciplinary research fields?
- What impact have scientists?

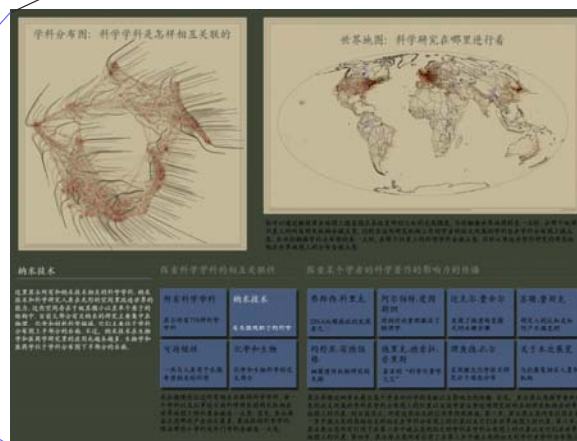
### Contributions:

- Interactive, high resolution interface to access and make sense of data about scholarly activity.

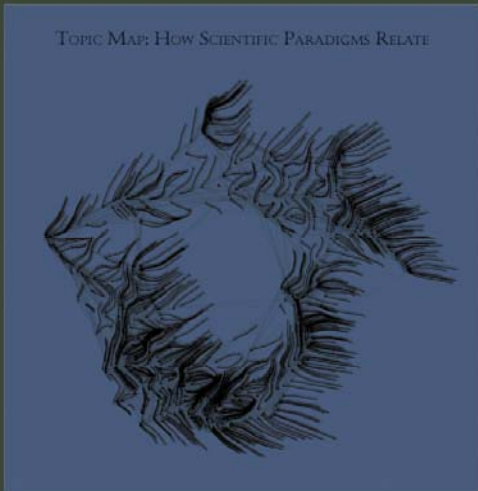


Large-scale, high resolution prints illuminated via projector or screen.

Interactive touch panel.







You may run your finger over each of these maps to control the lighting on the other: touching a place on the world map will light up topics studied in that place; touching a paradigm on the topic map will light up the places that study that topic.

### Nanotechnology

This overlay shows the distribution of nanotechnology within the paradigms of science. The majority of current work in nanotechnology takes place in physics, chemistry, and materials science, at the upper right portion of the map. However, an increasing amount of nanotechnology is being applied in the biological and medical sciences, at the lower right.

<b>All Topics</b> <i>Sweep through all 776 scientific paradigms</i>	<b>Nanotechnology</b> <i>Science on the tiny scale of molecules</i>	<b>Francis H. C. CRICK</b> <i>Co-discovered DNA's double helix</i>	<b>Albert EINSTEIN</b> <i>Revitalized physics with Relativity theories</i>	<b>Michael E. FISHER</b> <i>Models critical phase transitions of matter</i>	<b>Susan T. FISKE</b> <i>Connects perception and stereotypes</i>
<b>Sustainability</b> <i>The science behind our long-term hopes</i>	<b>Biology &amp; Chemistry</b> <i>The interface between these two vital fields</i>	<b>Joshua LEDERBERG</b> <i>Pioneer in bacterial genetic mechanisms</i>	<b>Derek J. de Solla PRICE</b> <i>Known as the "Father of Scientometrics"</i>	<b>Richard N. ZARE</b> <i>Uses laser chemistry in molecular dynamics</i>	<b>About this display</b> <i>People &amp; organizations that helped create it</i>

We sweep slowly through adjoining related topics, lighting up the places in the world that study each topic. You may select a subset of the topics that deal with these three interesting subjects by touching it.

A single person's spreading influence is shown as a series of four snapshots. First, we light only topics and places relating to that person's papers—papers that are still highly cited today. The second lights everything that cites that original work. Note that this first-generation impact extends to far more topics than did the original work. The third snapshot lights science that cites the second, and the fourth lights science that cites the third.

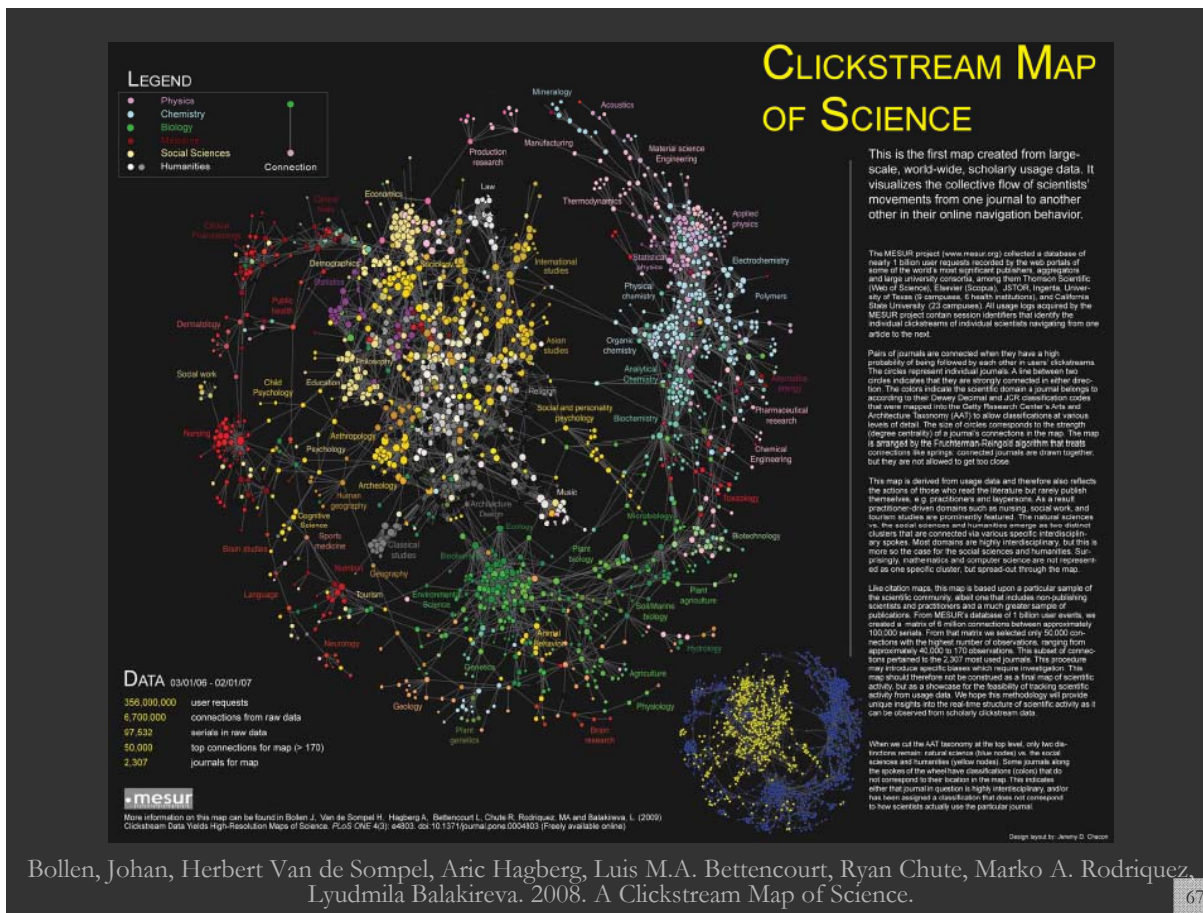


Science Maps in "Expedition Zukunft" science train visiting 62 cities in 7 months

12 coaches, 300 m long

Opening was on April 23<sup>rd</sup>, 2009 by German Chancellor Merkel

<http://www.expedition-zukunft.de>



Bollen, Johan, Herbert Van de Sompel, Aric Hagberg, Luis M.A. Bettencourt, Ryan Chute, Marko A. Rodriguez, Lyudmila Balakireva. 2008. A Clickstream Map of Science. 67

