

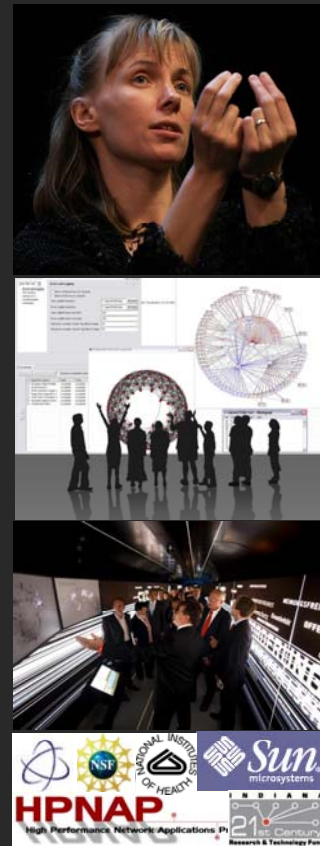
Envisioning Science and Technology

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

*HBS Science-Based Business Initiative Seminar, Boston., MA
May 7, 2010*



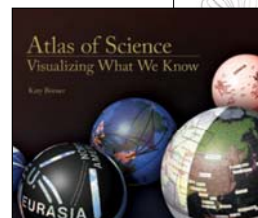
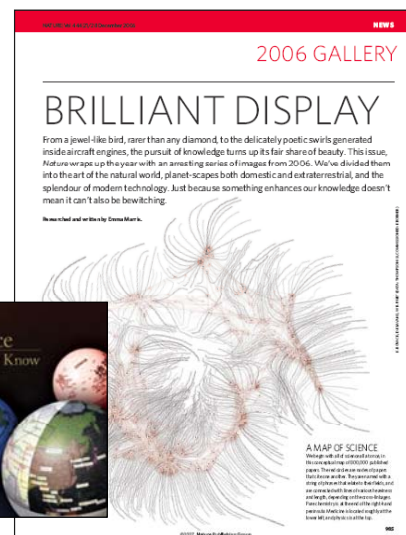
Computational Scientometrics References

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>





Computational Scientometrics CI



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>



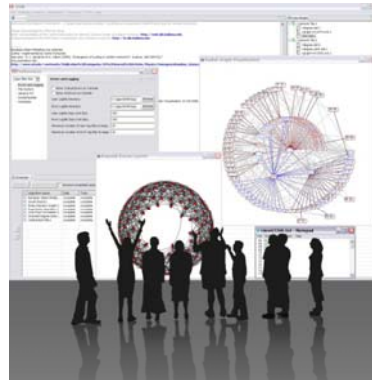
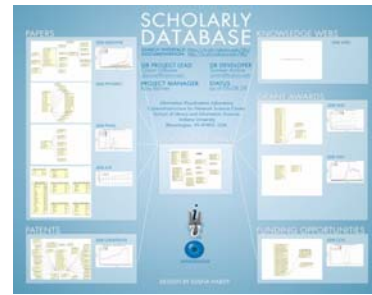
Network Workbench Tool + Community Wiki
<http://nwb.slis.indiana.edu>



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



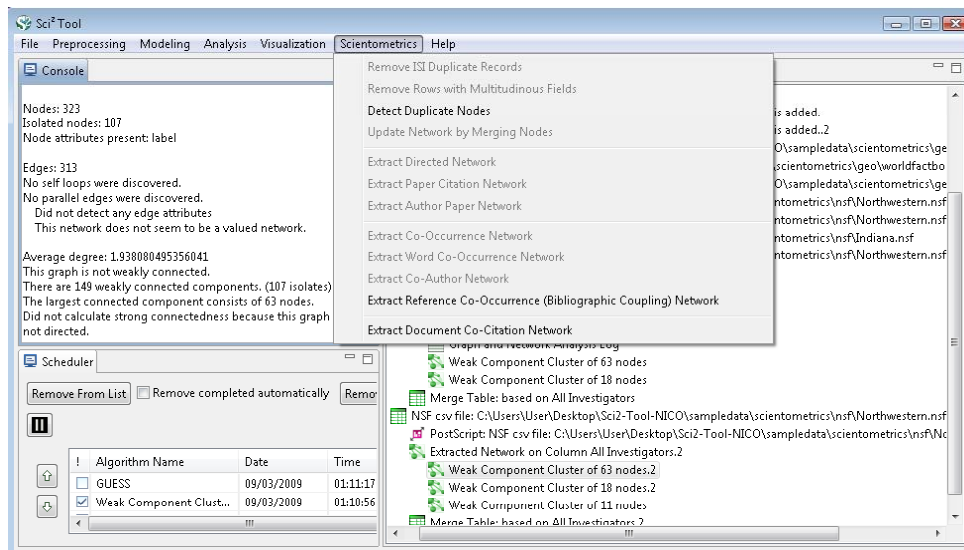
Epidemics Cyberinfrastructure
<http://epic.slis.indiana.edu/>



3



Sci² Tool for Science of Science Research and Practice



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.



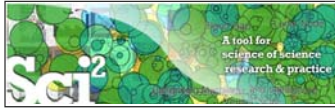
INDIANA UNIVERSITY
SCHOOL OF LIBRARY
AND INFORMATION SCIENCE



James S. McDonnell Foundation



4



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-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

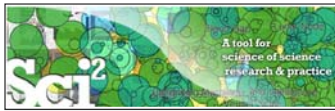
Node Indegree
Node Outdegree
Indegree Distribution
Outdegree Distribution

K-Nearest Neighbor
Single Node in-Out Degree Correlations

Dyad Reciprocity
Arc Reciprocity
Adjacency Transitivity

Weak Component Clustering
Strong Component Clustering

5



Sci² 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

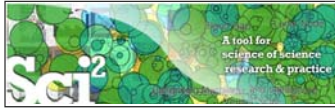
Extract Word Co-Occurrence Network
Extract Co-Author Network
Extract Reference Co-Occurrence
(Bibliographic Coupling) Network

Extract Document Co-Citation Network

* Requires permission from UCSD
All four+ save into Postscript files.

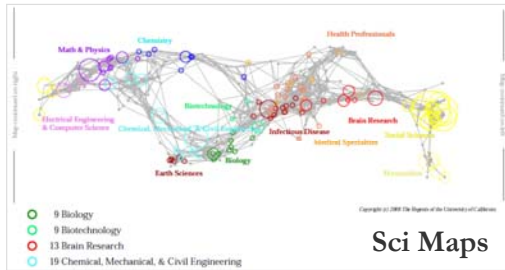
[General Network extraction](#)

6



Sci² Tool

Plugins that render into Postscript files:



Börner, Katy, Huang, Weixia (Bonnie), Linnemeier, Micah, Dubon, Russell Jackson, Phillips, Patrick, Ma, Nianli, Zoss, Angela, Guo, Hanning & Price, Mark. (2009). *Retz-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.*

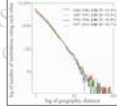


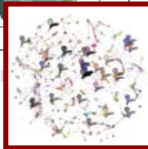



Type of Analysis vs. Level of Analysis

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



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, all of USA, all of science. 
Temporal Analysis (When)	Funding portfolio of one individual	Mapping topic bursts in 20-years of PNAS	113 Years of Physics Research 
Geospatial Analysis (Where)	Career trajectory of one individual	Mapping a states intellectual landscape	PNAS Publications 
Topical Analysis (What)	Base knowledge from which one grant draws.	Knowledge flows in Chemistry research	VxOrd/Topic maps of NIH funding 
Network Analysis (With Whom?)	NSF Co-PI network of one individual	Co-author network	NSF's core competency 

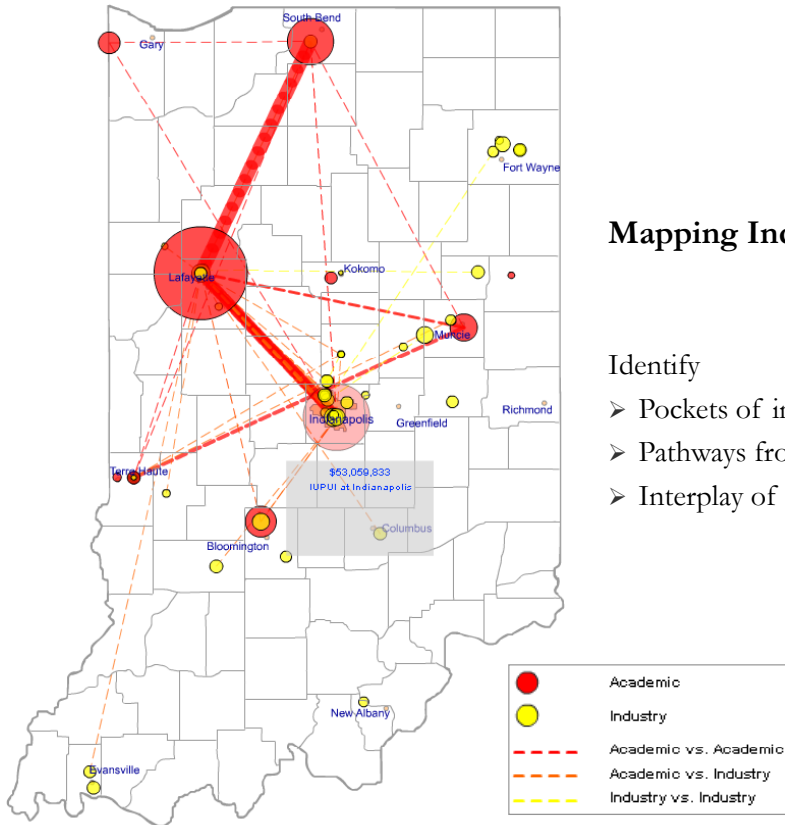
9



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, all of USA, all of science.
Temporal Analysis (When)	Funding portfolio of one individual	Mapping topic bursts in 20-years of PNAS	113 Years of Physics Research
Geospatial Analysis (Where)	Career trajectory of one individual	Mapping a states intellectual landscape	PNAS Publications
Topical Analysis (What)	Base knowledge from which one grant draws.	Knowledge flows in Chemistry research	VxOrd/Topic maps of NIH funding
Network Analysis (With Whom?)	NSF Co-PI network of one individual	Co-author network	NSF's core competency

10



Mapping Indiana's Intellectual Space

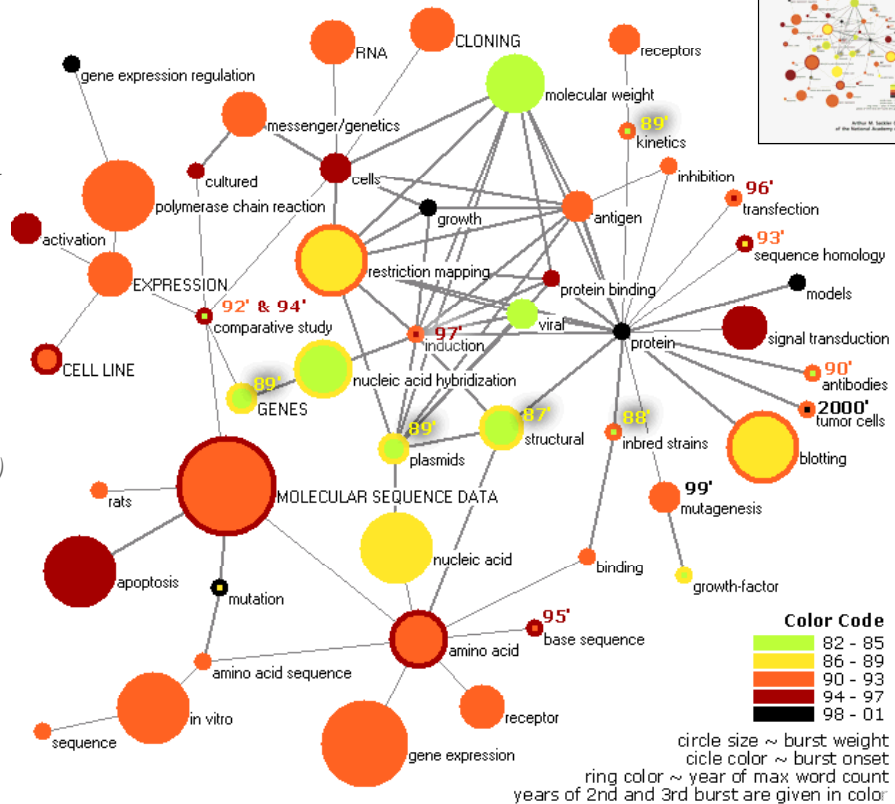
Identify

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

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.



Research Collaborations by the Chinese Academy of Sciences

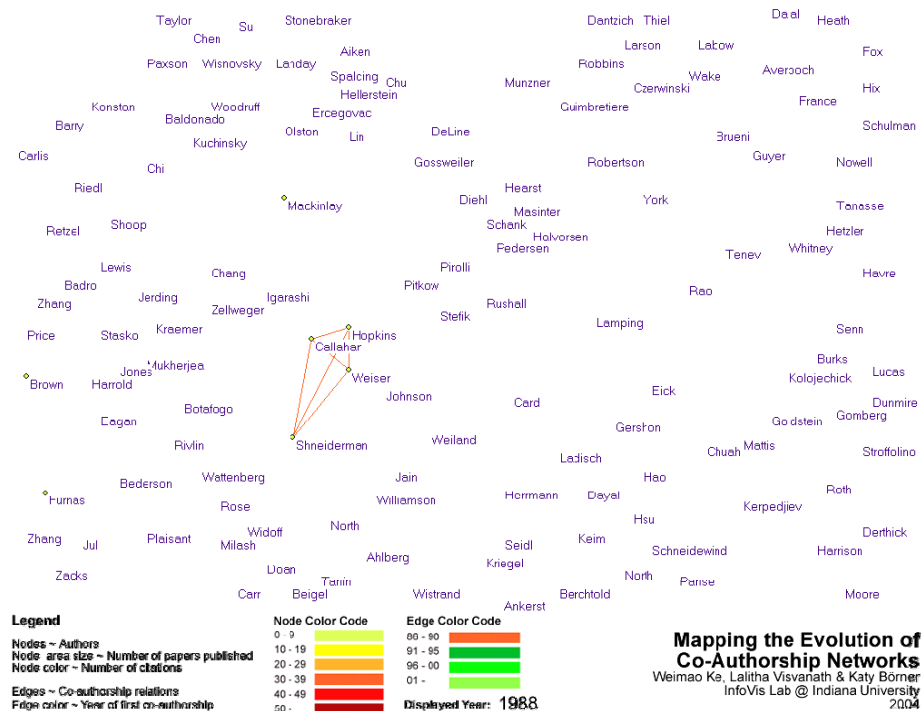
By Weixia (Bonnie) Huang, Russell J. Dubon, Elisha F. Hardy, Katy Börner, Indiana University, USA



This map highlights the research co-authorship 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.

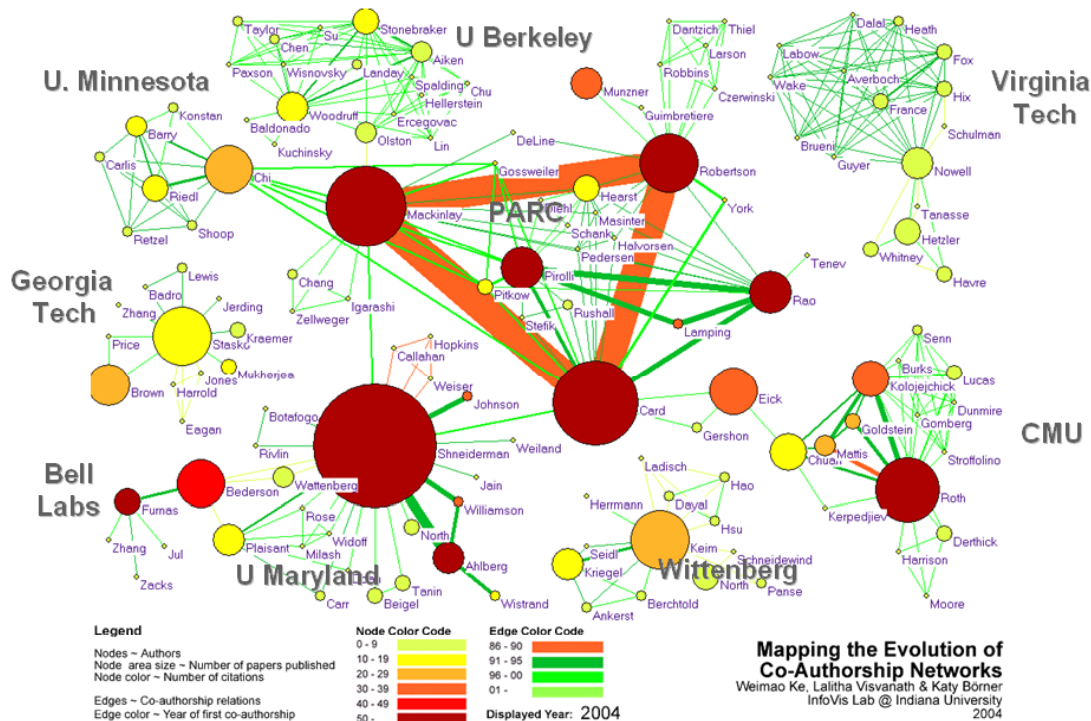
Mapping the Evolution of Co-Authorship Networks

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



Mapping the Evolution of Co-Authorship Networks

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



15

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

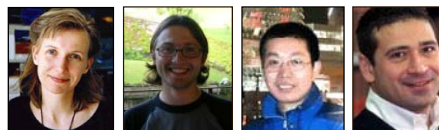
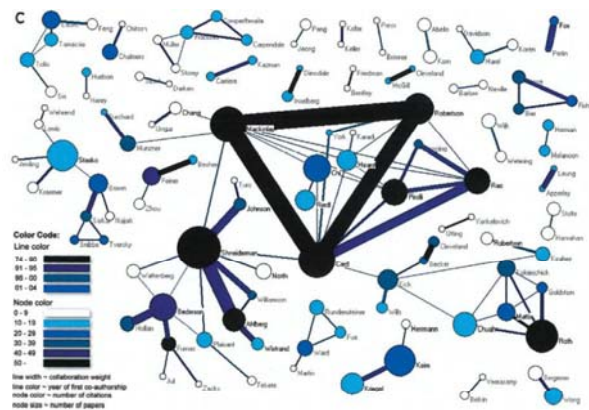
Börner, Dall'Asta, Ke & Vespignani (2005) *Complexity*, 10(4):58-67.

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.



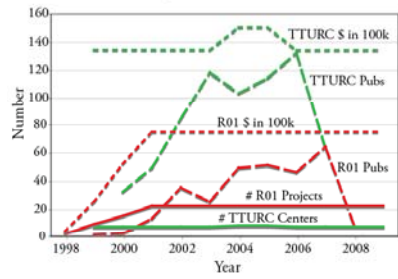
16

Mapping Transdisciplinary Tobacco Use Research Centers Publications

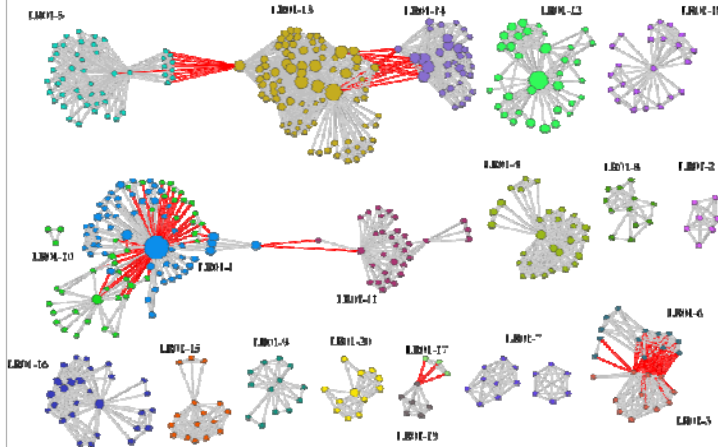
Compare R01 investigator based funding with TTURC Center awards in terms of number of publications and evolving co-author networks.

Zoss & Börner, *forthcoming*.

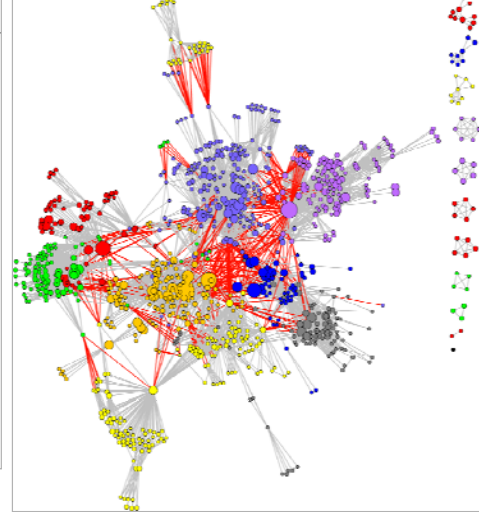
R01 & TTURC Project Information



Longitudinal R01 Co-Authorship Network

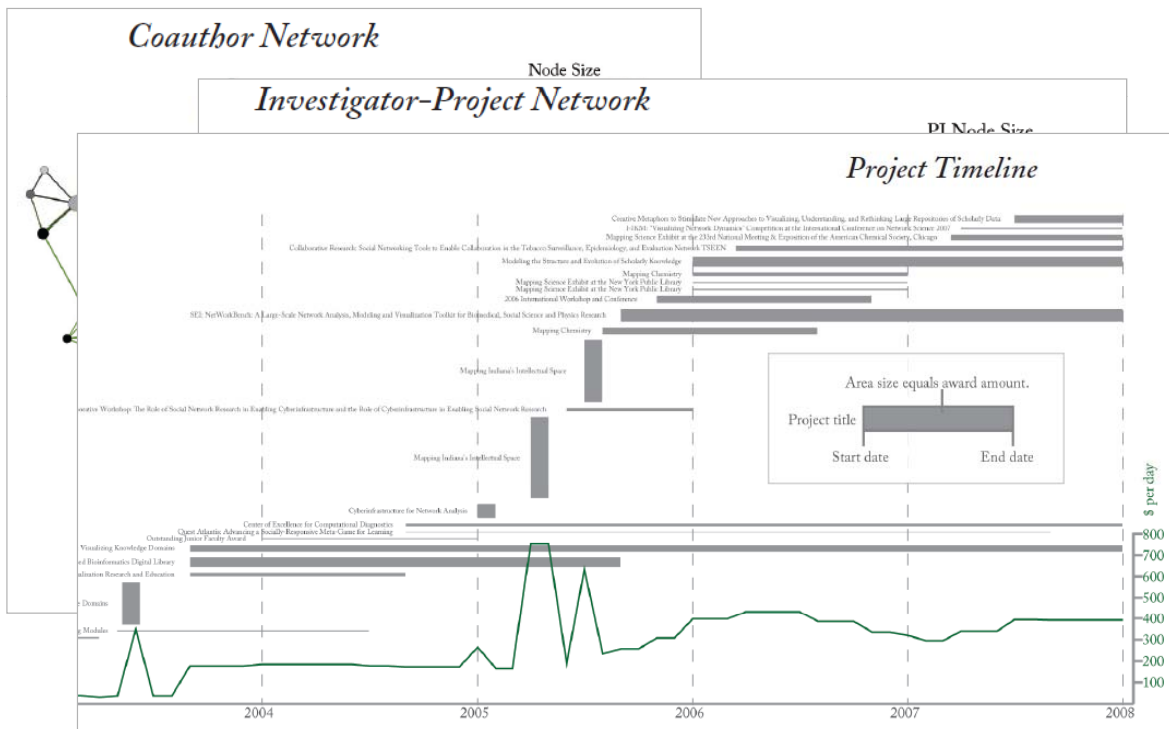


TTURC Co-Authorship Network



Representing, Analyzing, and Visualizing Scholarly Data in Support of Research Management (<http://ivl.slis.indiana.edu>)

Thomas Neirynek and Katy Börner (2007)





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

19

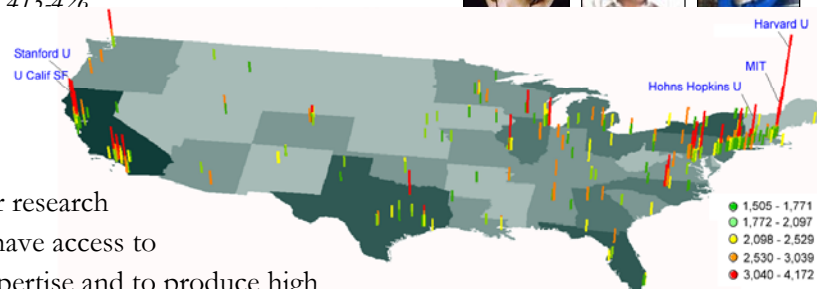
Spatio-Temporal Information Production and Consumption of Major U.S. Research Institutions

Börner, Katy, Penumarthy, 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



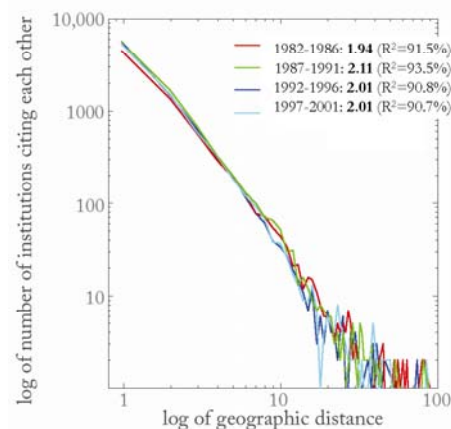
Research questions:

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.

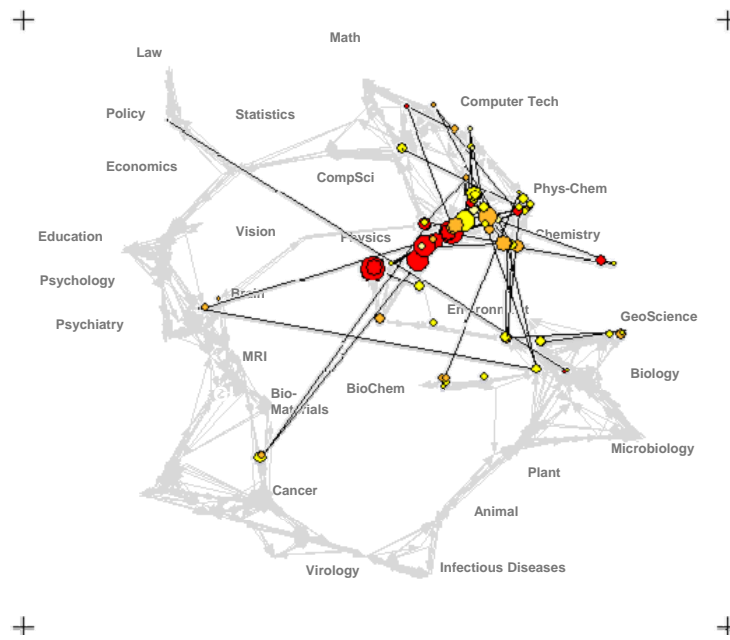


20

Science map applications: Identifying core competency

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

Funding patterns of the US Department of Energy (DOE)

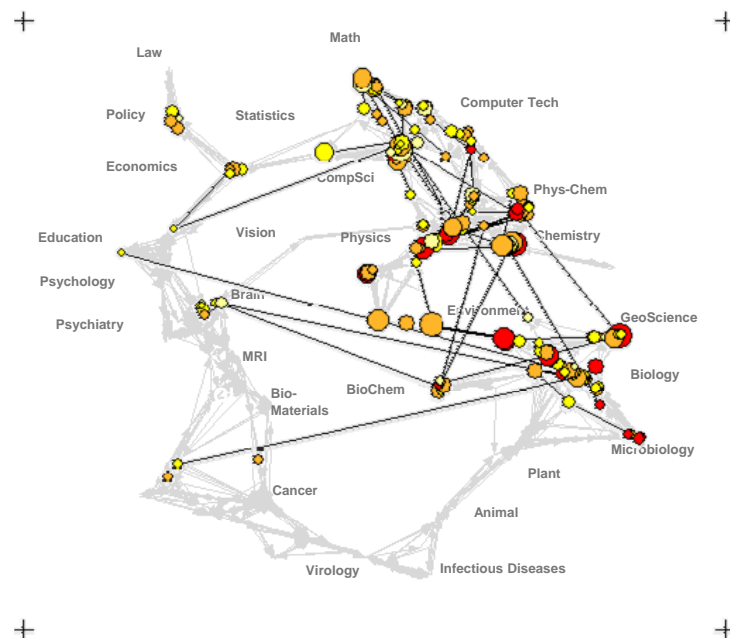


21

Science map applications: Identifying core competency

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

Funding Patterns of the National Science Foundation (NSF)

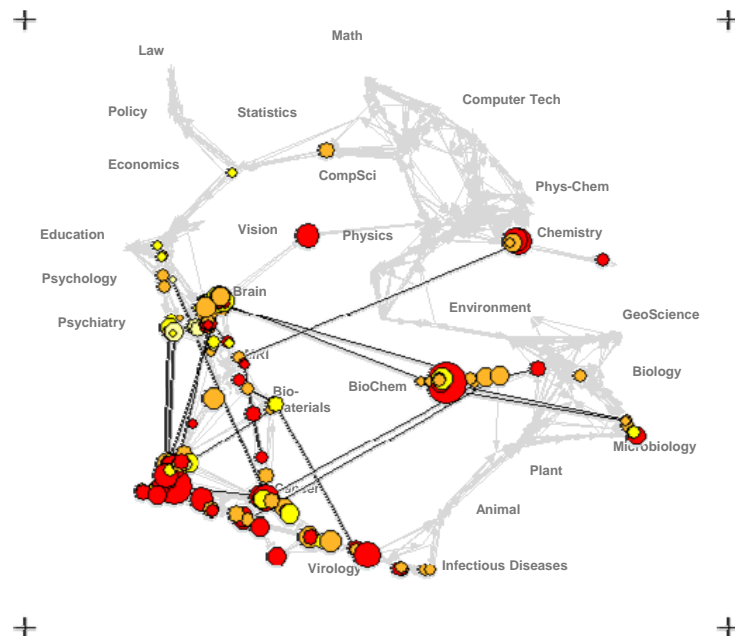


22

Science map applications: Identifying core competency

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

Funding Patterns of the National Institutes of Health (NIH)

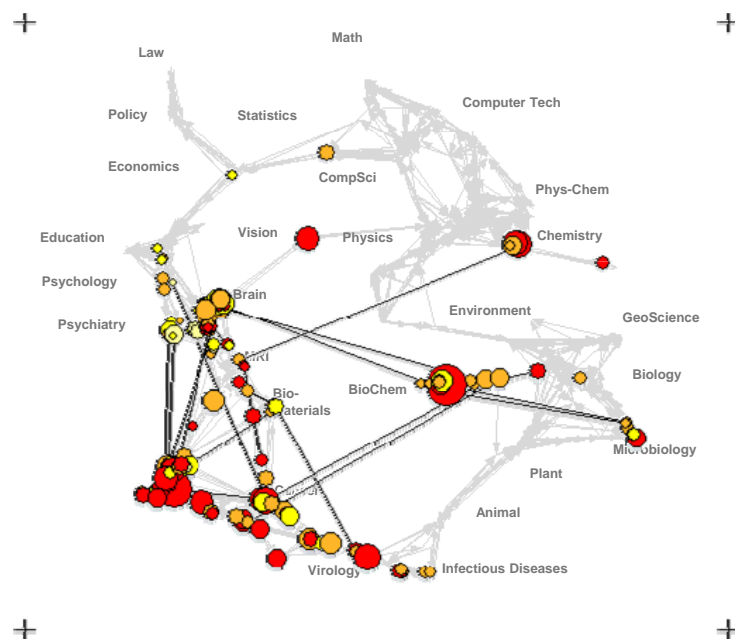


23

Science map applications: Identifying core competency

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

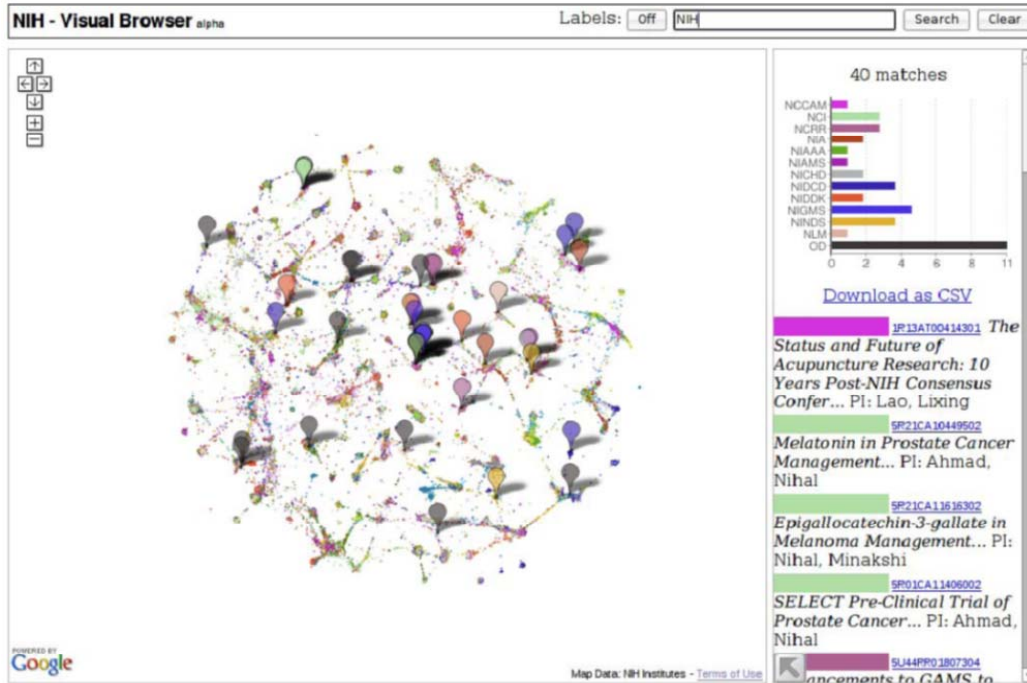
Funding Patterns of the National Institutes of Health (NIH)



24

Interactive Science Map of NIH Funding

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

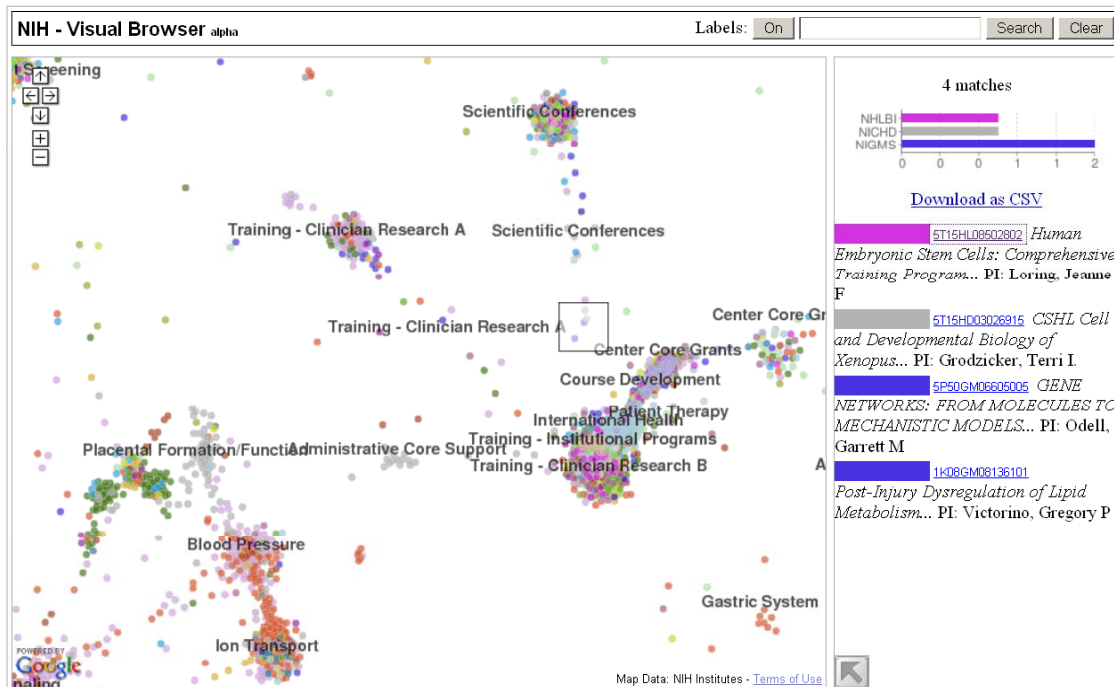


<http://scimaps.org/maps/nih/2007>

25

Interactive Maps of Science – NIH Funding

Google maps with charts and tables



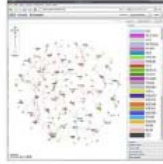
<http://scimaps.org/maps/nih/2007>

A Topic Map of NIH Grants 2007

ChalkLabs UCIrvine

Bruce W. Herr II (ChalkLabs & IU), Gully Burns (ISU), David Newman (UCI), Edmund Talley (NIH)

The National Institutes of Health (NIH) is organized as a multitude of institutes and centers whose missions are primarily focused on distinct diseases. However, disease etiologies and therapies flout scientific boundaries, and thus there is tremendous overlap in the kinds of research funded by each institute. This creates a daunting landscape for decisions on research directions, funding allocations, and policy formulations. Shown here is developed an interactive topic map for navigating this landscape, online at www.nihmap.org. Institute abbreviations can be found at www.nih.gov/ohrt.



Topic modeling, a statistical technique that automatically learns semantic categories, was applied to assess projects in terms used by researchers to describe their work, without the biases of keywords or subject headings. Grant similarities were derived from their topic mixtures, and grants were then clustered on a two-dimensional map using a force-directed simulated annealing algorithm. This analysis creates an interactive environment for assessing grant relevance to research categories and to NIH institutes in which grants are localized.



National Cancer Institute (NCI)

- TOP 10 TOPICS
- 1. Disease: Clinical Trials
 - 2. Cancer Treatment
 - 3. Cancer Therapy
 - 4. Carcinogenesis
 - 5. Risk Factor Analysis
 - 6. Cancer Chemotherapy
 - 7. Metastasis
 - 8. Survival
 - 9. Prognosis/Prognostic
 - 10. Cancer Development

National Institute of General Medical Sciences (NIH/IGMS)

- TOP 10 TOPICS
- 1. Bioactive Organic Synthesis
 - 2. X-ray Crystallography
 - 3. Protein PDB
 - 4. Computational Models
 - 5. Food Safety
 - 6. Metalloproteins
 - 7. Tripartite Mechanisms
 - 8. Protein Complexes
 - 9. Neurobiology/Behavioral Genetics
 - 10. Cell Structure

National Heart, Lung, and Blood Institute (NIH/NHLBI)

- TOP 10 TOPICS
- 1. Cellular Injury
 - 2. Airway Injury
 - 3. Genetic Linkage Analysis
 - 4. Cardiovascular Disease
 - 5. Atherosclerosis
 - 6. Inflammation
 - 7. Blood Pressure
 - 8. Asthma: Airway Airway Disease
 - 9. Gene Expression
 - 10. Hypertension

National Institute of Mental Health (NIH/NIMH)

- TOP 10 TOPICS
- 1. Mood Disorders
 - 2. Schizophrenia
 - 3. Behavioral Neuroscience Studies
 - 4. Mental Health
 - 5. Depression
 - 6. Cognitive Behavioral Therapy
 - 7. AIDS Prevention
 - 8. Genetic Linkage Analysis
 - 9. Addictions
 - 10. Childhood

Herr II, Bruce W., Gully Burns, David Newman, Edmund Talley. 2007. A Topic Map of NIH Grants 2007.

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

Angela Zoss, Michael Connover, Katy Börner (2010)

Visualization of Job Postings

Map of Science

Geographic Visualization

Visualization of Job Postings

Map of Science | Geographic

Map of Science | Geographic

Postdoc at Harvard Medical School
Link to Post

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.

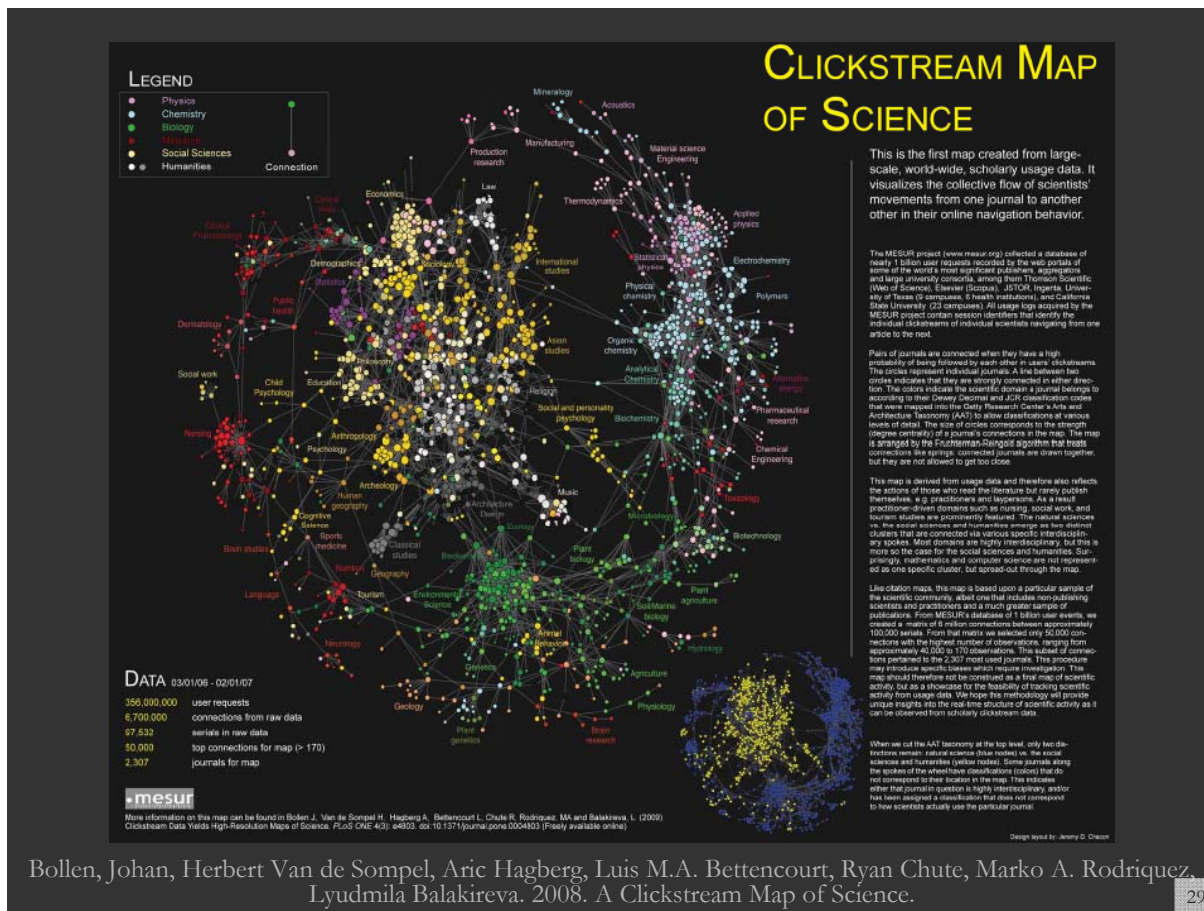
These clusters are represented by 554 individual nodes in the network. The links between the clusters show that some clusters are related to other clusters but are not as tightly connected as the journals that make up each cluster. Then the clusters are labeled both by the content area shared by the 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, high-

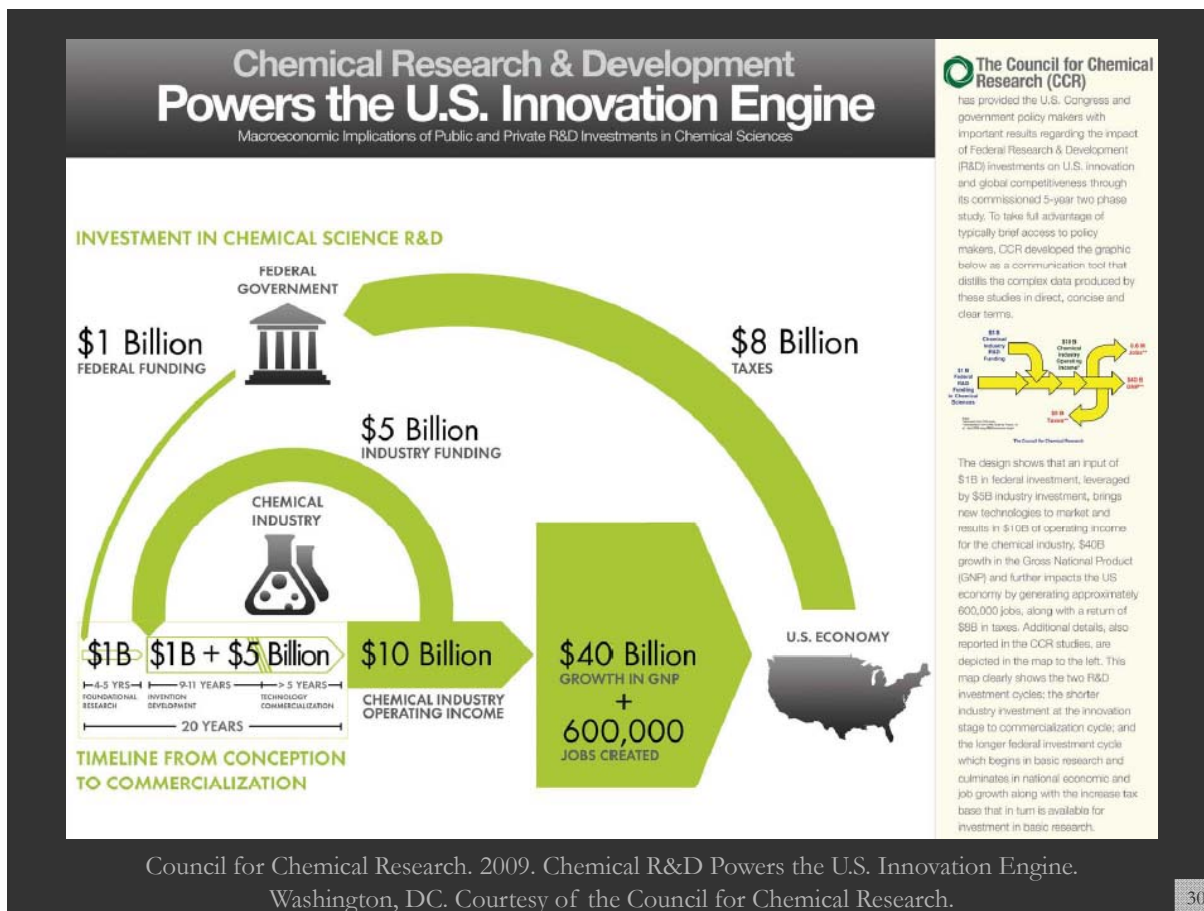
Copyright © 2008 The Regents of the University of California - Terms of Use

Search for Jobs Search

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



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. 29



Council for Chemical Research. 2009. Chemical R&D Powers the U.S. Innovation Engine. Washington, DC. Courtesy of the Council for Chemical Research. 30

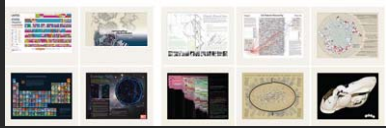
Mapping Science Exhibit – 10 Iterations in 10 years

<http://scimaps.org/>

The Power of Maps (2005)



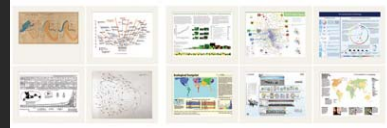
The Power of Reference Systems (2006)



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.



31

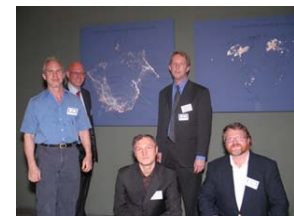
Illuminated Diagram Display

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?

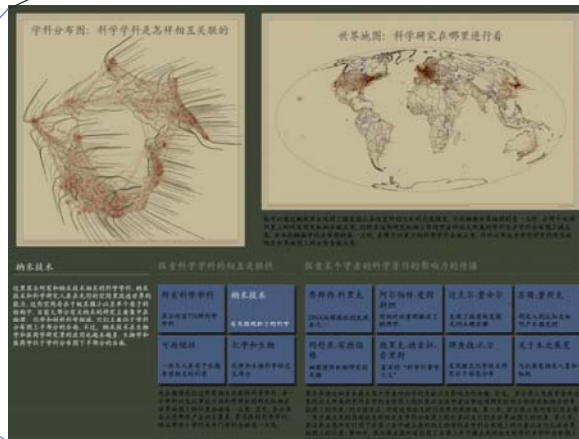


Large-scale, high resolution prints illuminated via projector or screen.

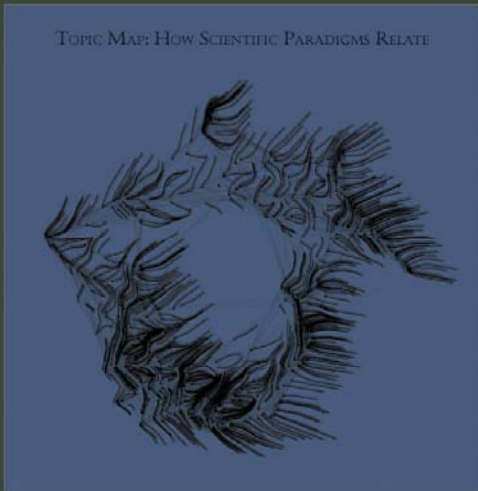
Interactive touch panel.

Contributions:

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



32



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.

All Topics

Sweep through all 776 scientific paradigms

Nanotechnology

Science on the tiny scale of molecules

Francis H. C. CRICK

Co-discovered DNA's double helix

Albert EINSTEIN

Revitalized physics with Relativity theories

Michael E. FISHER

Models critical phase transitions of matter

Susan T. FISKE

Connects perception and stereotypes

Sustainability

The science behind our long-term hopes

Biology & Chemistry

The interface between these two vital fields

Joshua LEDERBERG

Pioneer in bacterial genetic mechanisms

Derek J. de Solla PRICE

Known as the "Father of Scientometrics"

Richard N. ZARE

Uses laser chemistry in molecular dynamics

About this display

People & organizations that helped create it

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.



Debut of 5th 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>

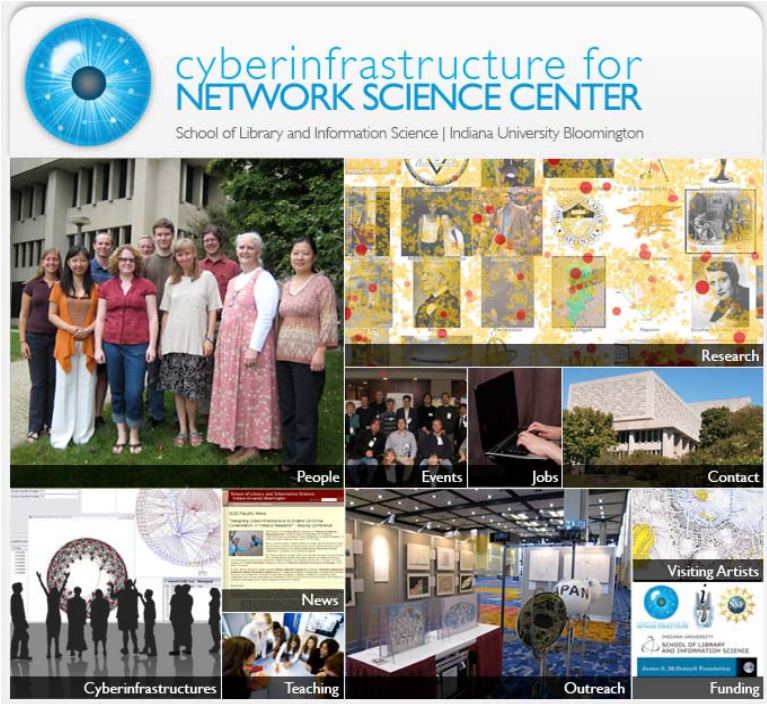


Science Maps in "Expedition Zukunft" science train visiting 62 cities in 7 months
12 coaches, 300 m long
Opening was on April 23rd, 2009 by German Chancellor Merkel
<http://www.expedition-zukunft.de>

This is the only mockup in this slide show.

Everything else is available today.





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