

# Computational Scientometrics

*Studying science by scientific means*



**Dr. Katy Börner**

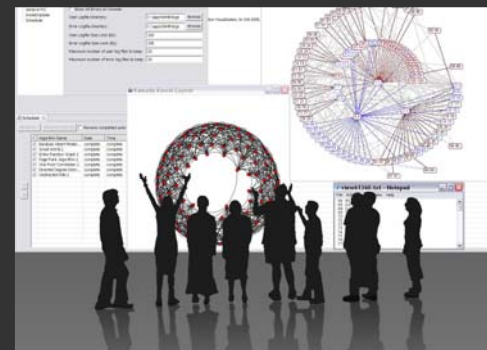
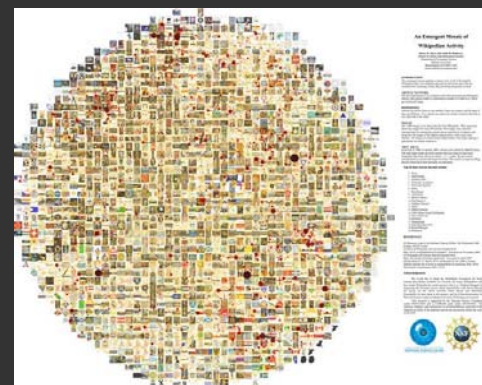
Cyberinfrastructure for Network Science Center, Director  
Information Visualization Laboratory, Director  
School of Library and Information Science  
Indiana University, Bloomington, IN

[katy@indiana.edu](mailto:katy@indiana.edu)

*Designing cyberinfrastructure to enable US-China collaboration in tobacco control and research*

*Beijing, China*

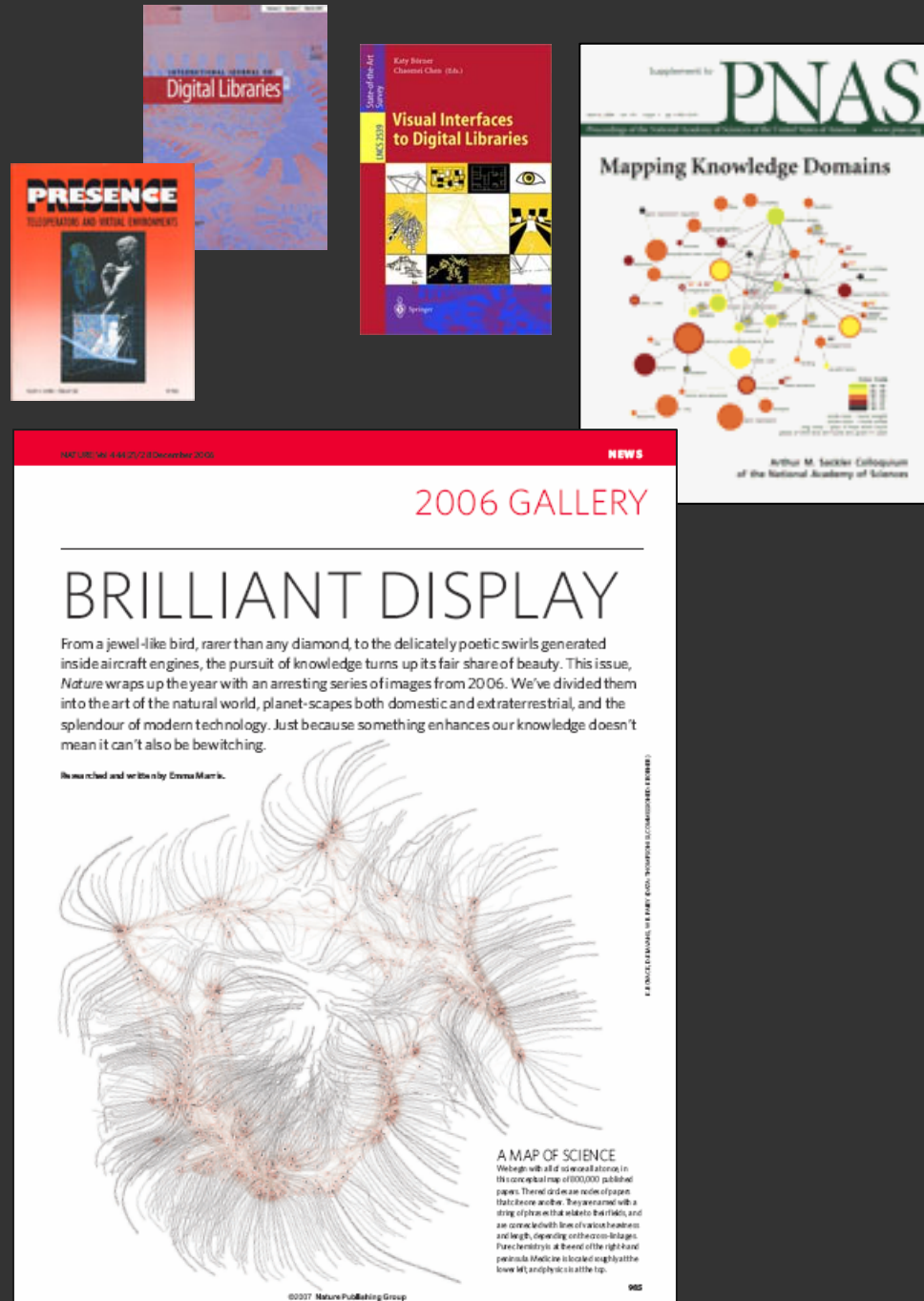
*March 27-29, 2008*



# Computational Scientometrics

*Studying science by scientific means*

- Börner, Katy, Chen, Chaomei, and Boyack, Kevin. (2003). **Visualizing Knowledge Domains.** In Blaise Cronin (Ed.), *Annual Review of Information Science & Technology*, Volume 37, Medford, NJ: Information Today, Inc./ASIST, chapter 5, pp. 179-255.
- 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).
- **Places & Spaces: Mapping Science** exhibit, soon on display at the National Research Council, Ottawa, Canada.  
<http://scimaps.org>.



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

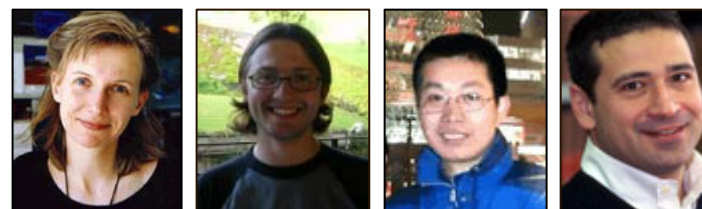
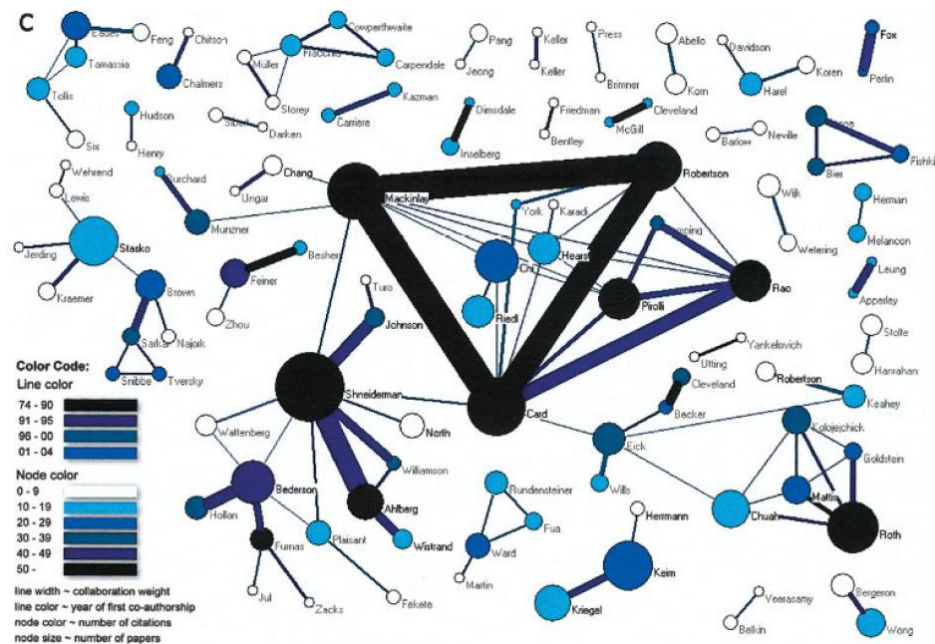
Börner, Dall'Asta, Ke & Vespignani (2005) *Complexity*, 10(4):58-67.

## 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.





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

## Research Institutions

Börner, Katy, Penumathy, Shashikant, Meiss, Mark and Ke, Weimao. (2006)

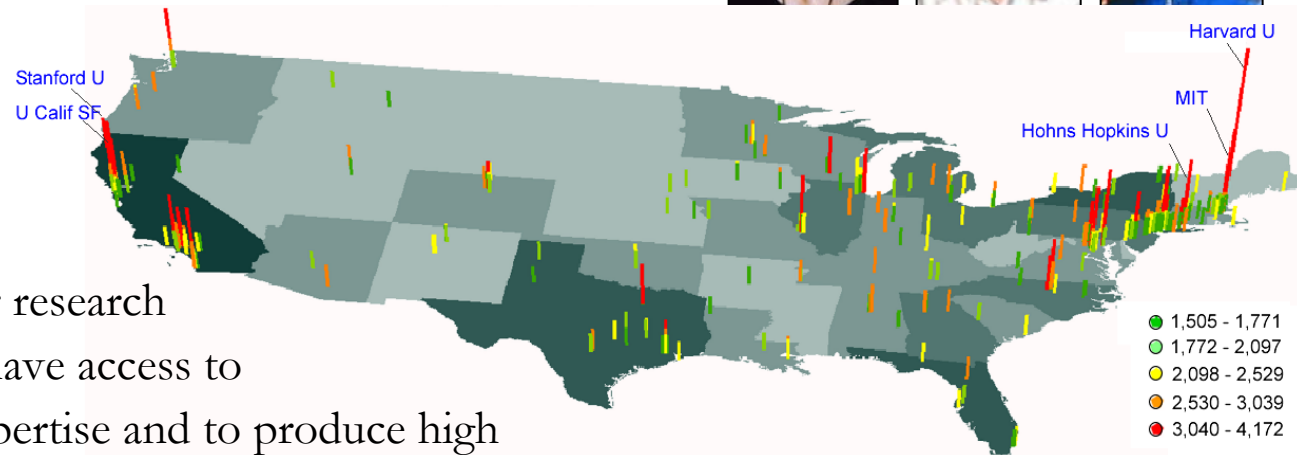
*Mapping the Diffusion of Scholarly Knowledge Among Major U.S. Research*

*Institutions. Scientometrics. 68(3), pp. 415-426.*



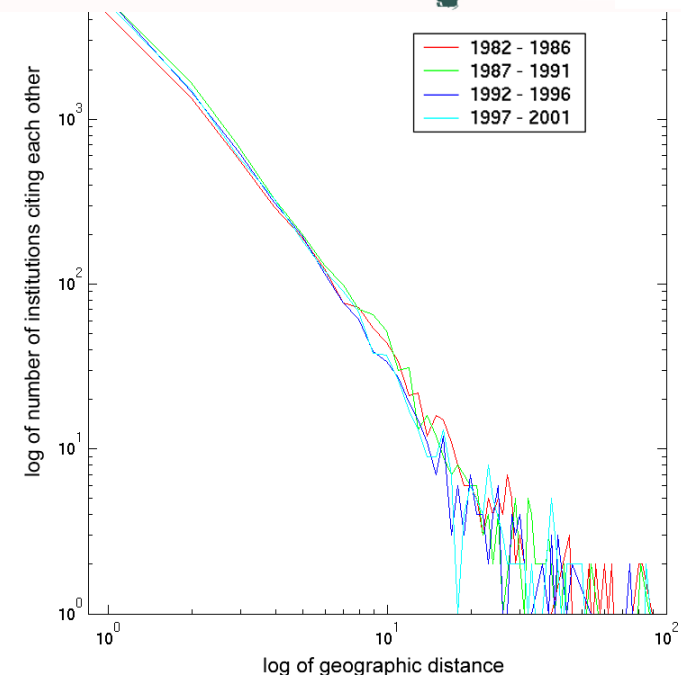
### Research questions:

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.



# Studying large scale social networks such as Wikipedia

## *Vizzards 2007 Entry*

### Second Sight: An Emergent Mosaic of Wikipedian Activity, The NewScientist, May 19, 2007



## Second sight

Image: Bruce W. Herr and Todd M. Holloway

### Power struggle

How do you keep track of the bubbling mass of information that is Wikipedia? This chaotic-looking mosaic is one attempt to show which topics are contained in the online encyclopedia, and those most hotly contested.

It's a mind-boggling task. About 4 million "Wikipedians" have made over 130 million edits, and the English-language version alone contains 1.7 million articles. Every second a new edit is made, and every day 2000 new articles spring up.

To make sense of it all, Bruce Herr and Todd Holloway of Indiana University, Bloomington, created clusters of 300 or so articles that touch on a related topic, such as a religion or a famous person. For each cluster they took one picture from the most popular article and laid them out in a circular grid.

Atop the grid are coloured dots showing how often and how recently each article has been edited. The larger, darker dots mean more intense activity. The list of blitzed articles reveals the idiosyncratic priorities of Wikipedians: Jesus, Adolf Hitler, Nintendo, Hurricane Katrina, Britney Spears and Albert Einstein.

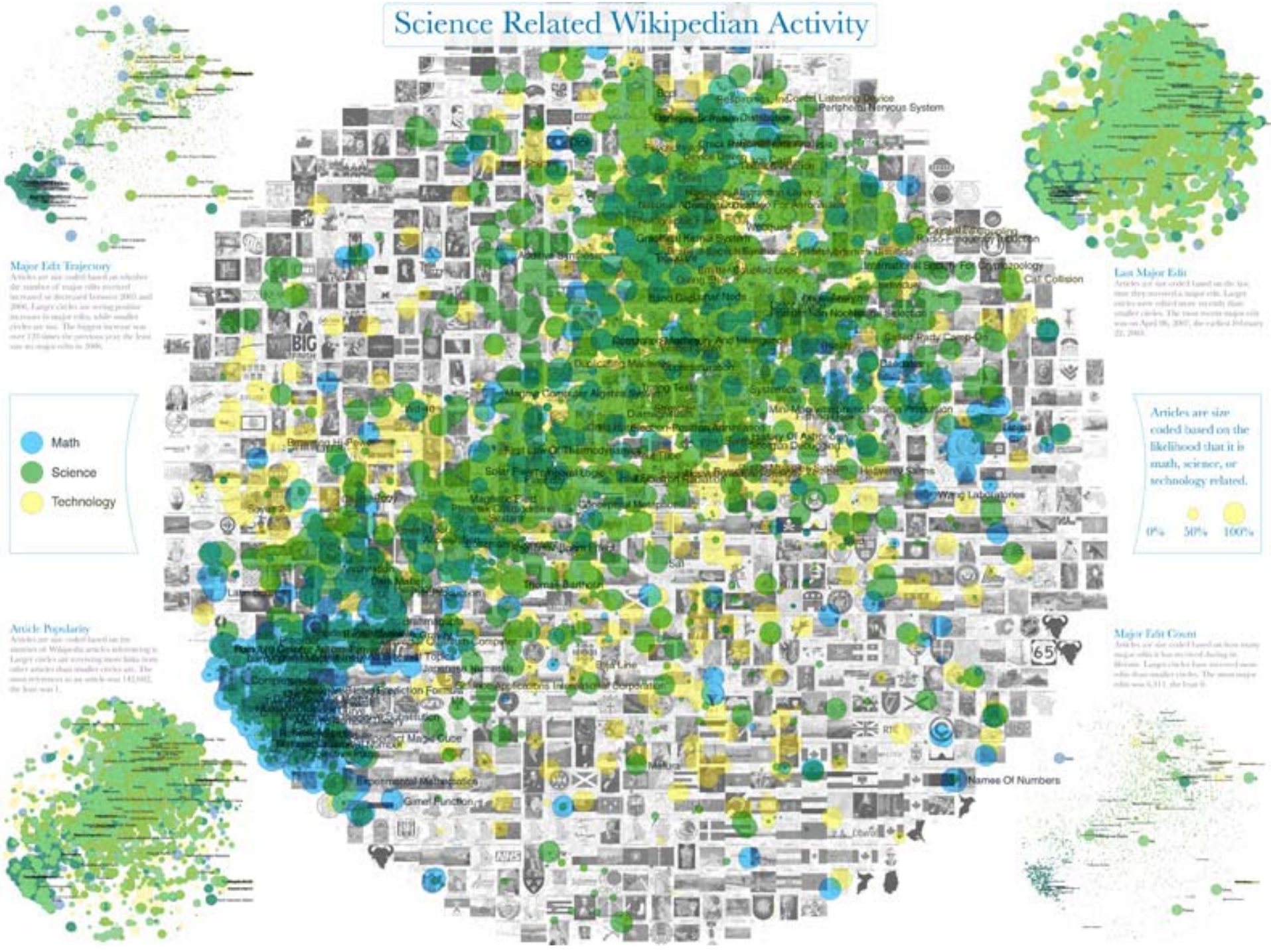
Updating the image in real time would allow Wikipedia's administrators to spot where arguments are taking place, Herr suggests. If rival contributors are repeatedly changing each other's entries, for example, a page could be locked until the mood cools (locked pages at the time of writing include entries on Sheffield Wednesday football club, Mikhail Gorbachev and pigs).

The mosaic has been commended in a competition for images that visualise network dynamics, coinciding with this week's International Workshop and Conference on Network Science in Bloomington.





# Science Related Wikipedia Activity



**Major Edit Trajectory**  
 Articles are size coded based on whether the number of major edits received increased or decreased between 2000 and 2006. Larger circles are moving positive in terms of major edits, while smaller circles are not. The biggest increase was over 120 times the previous year (the first was an major edit in 2000).

**Last Major Edit**  
 Articles are size coded based on the time they received a major edit. Larger circles were edited more recently than smaller circles. The most recent major edit was on April 06, 2007, the earliest February 22, 2003.

● Math  
 ● Science  
 ● Technology

Articles are size coded based on the likelihood that it is math, science, or technology related.

0% 50% 100%

**Article Popularity**  
 Articles are size coded based on the number of Wikipedia articles referencing it. Larger circles are receiving more links from other articles than smaller circles are. The most references to an article was 142,002 (the first was 1).

**Major Edit Count**  
 Articles are size coded based on how many major edits it has received during its lifetime. Larger circles have received more edits than smaller circles. The most major edits was 6,311 (the first 6).



# 113 Years of Physical Review

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

## 113 Years of Physical Review

The visualization aggregates 100,000 articles published in 100 volumes of *Physical Review* between 1915 and 2015. The 10,000 articles published in 2015 are the last to be included in this map. In 1915 the Physical Review combined the *Review* and *Annals of Mathematics* (AMS) under the AMS and the *Annals of Mathematics* was the top-level AMS journal. The 2015 articles from 1915 to 2015 are color-coded according to their subject area in a color key on the right side of the map.

Each article is color-coded according to the journal that appears in its original publication in the number of papers, and each journal is color-coded according to the journal that appears in its original publication.

### Nobel Prizes in Physical Review

Year of Nobel Prize Winner's Publication Year(s) (ordered to show Prize year on the right)

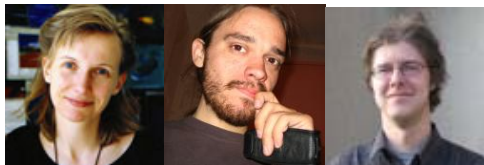
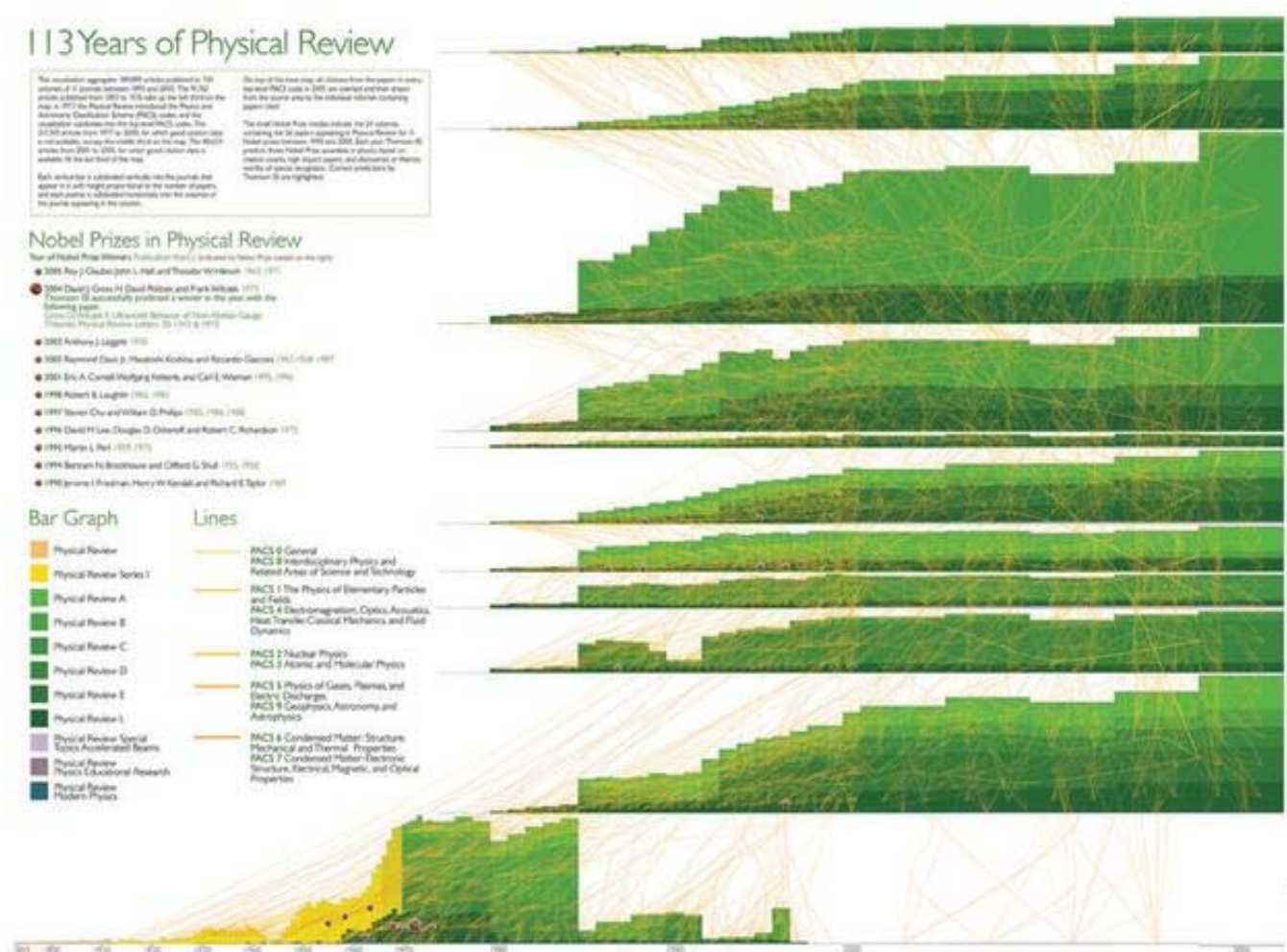
- 2009 Roy J. Glauber (2011, 1961 and Theodor W. Hänsch (1981, 1982)
- 2004 David J. Gross (1975, 1984) and Frank Wilczek (1975, 1984) (Gross and Wilczek predicted a winner in the year with the following year)
- 2002 Anthony J. Leggett (1972)
- 2000 Raymond Davis Jr., Kosuke Kobayashi and Ryoichi Gerasimovic (1972, 1981, 1987)
- 2001 Eric A. Cornell, Wolfgang Ketterle, and Carl E. Wieman (1975, 1996)
- 1998 Robert B. Laughlin (1981, 1982)
- 1997 Steven Chu and William D. Phillips (1975, 1996, 1997)
- 1994 David H. Lee, Douglas D. Osheroff, and Robert C. Richardson (1972)
- 1992 Martin L. Perl (1975, 1992)
- 1990 Steven H. Weinberg and Clifford G. Shaw (1975, 1990)
- 1980 Jerome I. Friedman, Henry W. Kendall, and Richard E. Taylor (1971)

### Bar Graph

- Physical Review
- Physical Review Series I
- Physical Review A
- Physical Review B
- Physical Review C
- Physical Review D
- Physical Review E
- Physical Review S
- Physical Review Special Topics Accelerated Beams
- Physical Review Physics Educational Research
- Physical Review Modern Physics

### Lines

- Physical Review General
- Physical Review Interdisciplinary Physics and Related Areas of Science and Technology
- Physical Review The Physics of Elementary Particles and Fields
- Physical Review Electrodynamics, Optics, Acoustics, Heat Transfer, Classical Mechanics, and Fluid Dynamics
- Physical Review Nuclear Physics
- Physical Review Atomic and Molecular Physics
- Physical Review Physics of Gases, Plasmas, and Electric Discharges
- Physical Review Geophysics, Astronomy, and Astrophysics
- Physical Review Condensed Matter: Structure, Mechanical and Thermal Properties
- Physical Review Condensed Matter: Electronic Structure, Electrical, Magnetic, and Optical Properties





## Cartography of the Physical and the Abstract

An exhibition created for the conference "Mapping Humanity's Knowledge and Expertise in the Digital Domain" at the 2005 Meeting of the American Association of Geographers that is updated regularly with new maps and explanations.

Home Browse Maps Compare & Contrast Maps Connect

Home



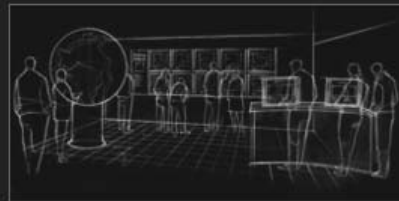
### Exhibit Purpose and Goals

The **Places & Spaces** exhibit has been created to demonstrate the power of maps.

An initial theme of this exhibit is to compare and contrast first maps of our entire planet with the first maps of all of science as we know it.

Come see with your own eyes the extent to which maps can be employed to help make sense of the flood of information we are confronted with and how domain maps can be used to locate complex and beautiful information.

This online part of the exhibit provides links to a selected series of maps and their makers along with detailed explanations of why these maps work. The physical counterpart supports the close inspection of high quality reproductions for display at conferences and education centers. It is meant to inspire cross-disciplinary discussion on how to best track and communicate human activity and scientific progress on a global scale.



## Places & Spaces: Mapping Science

a science exhibit that introduces people to maps of sciences, their makers and users.

<http://scimaps.org>

**Exhibit Curators:** Dr. Katy Börner & Elisha Hardy



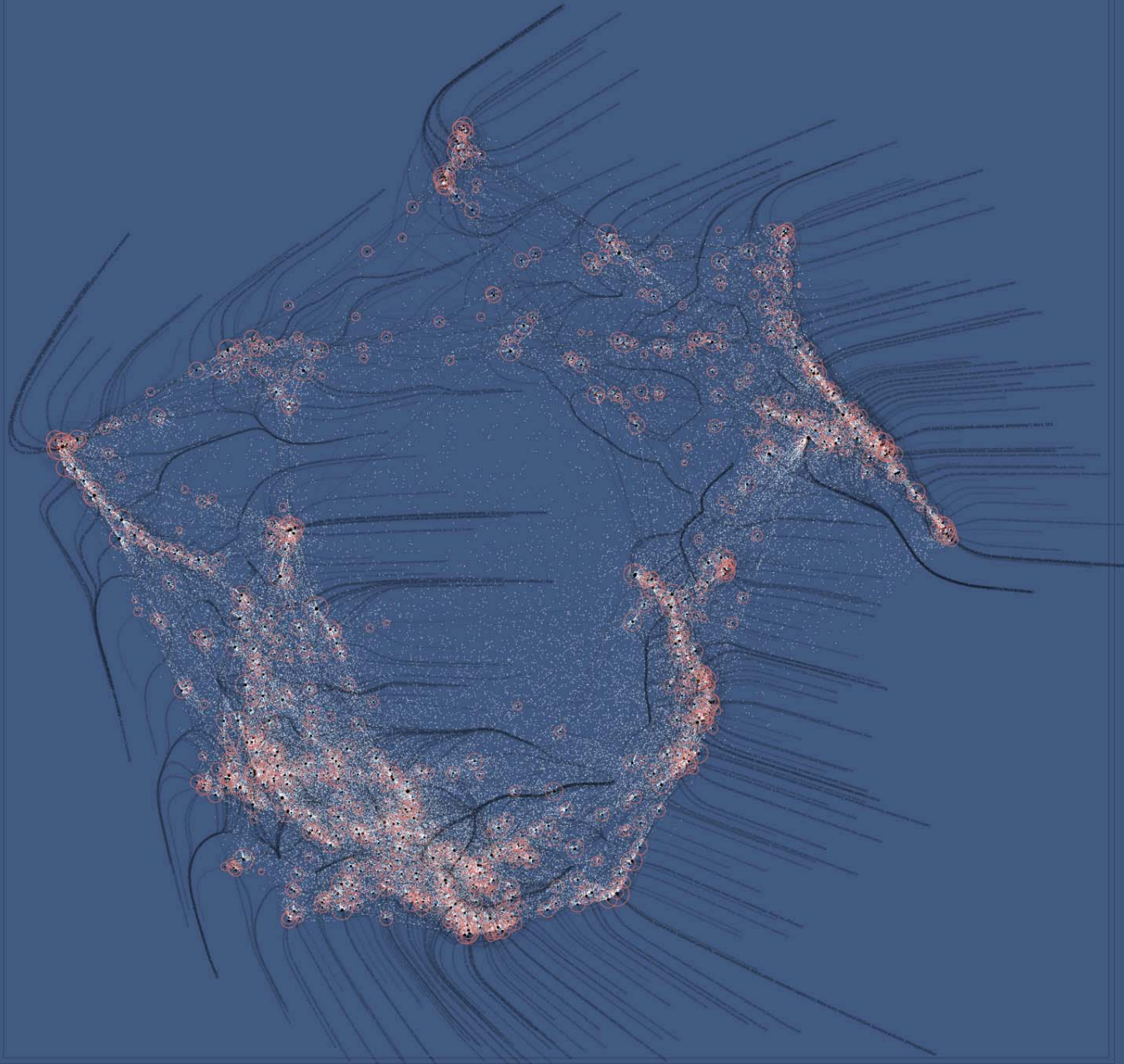


## Illuminated Diagram Display

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

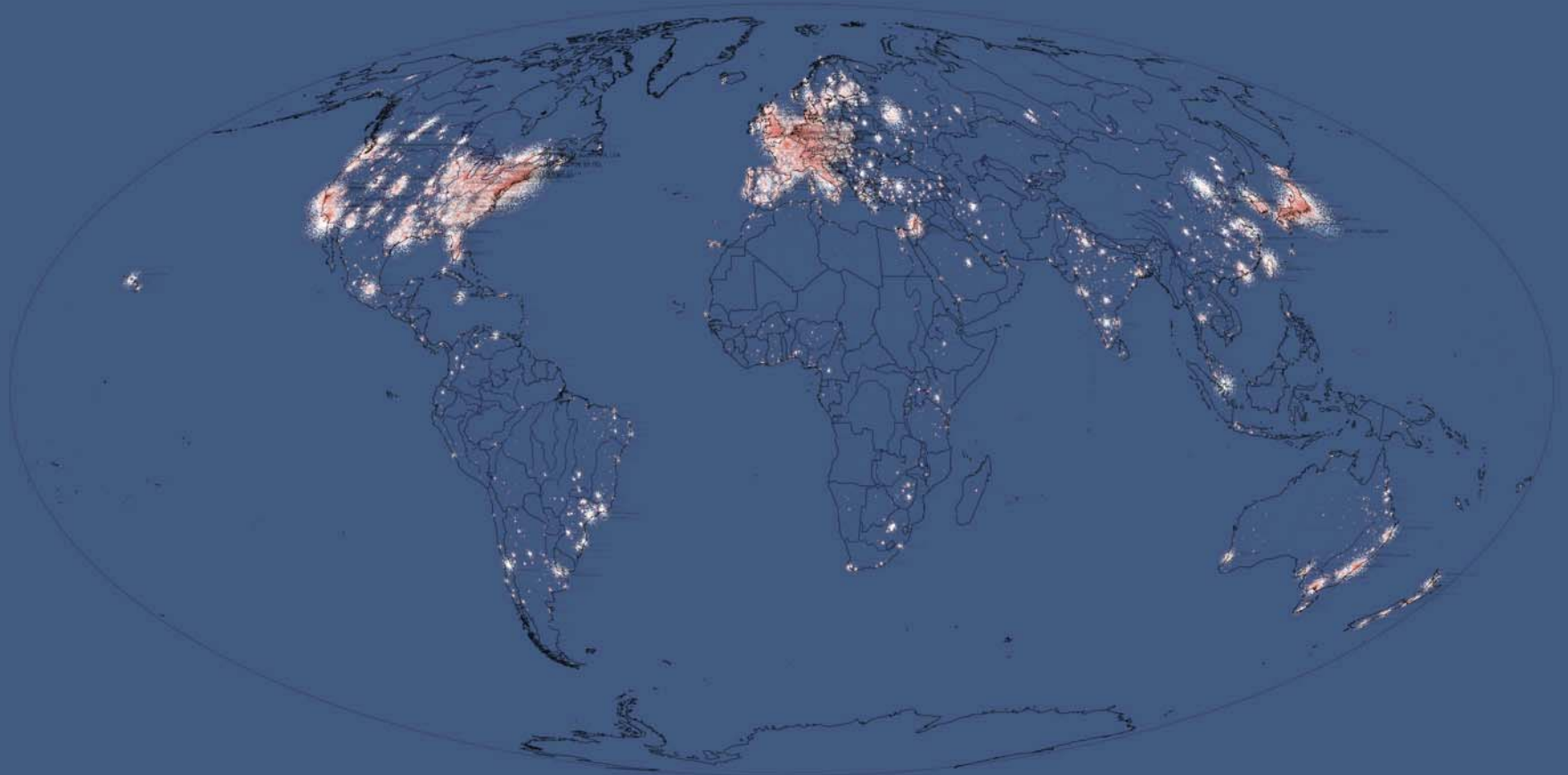


# TOPIC MAP: HOW SCIENTIFIC PARADIGMS RELATE





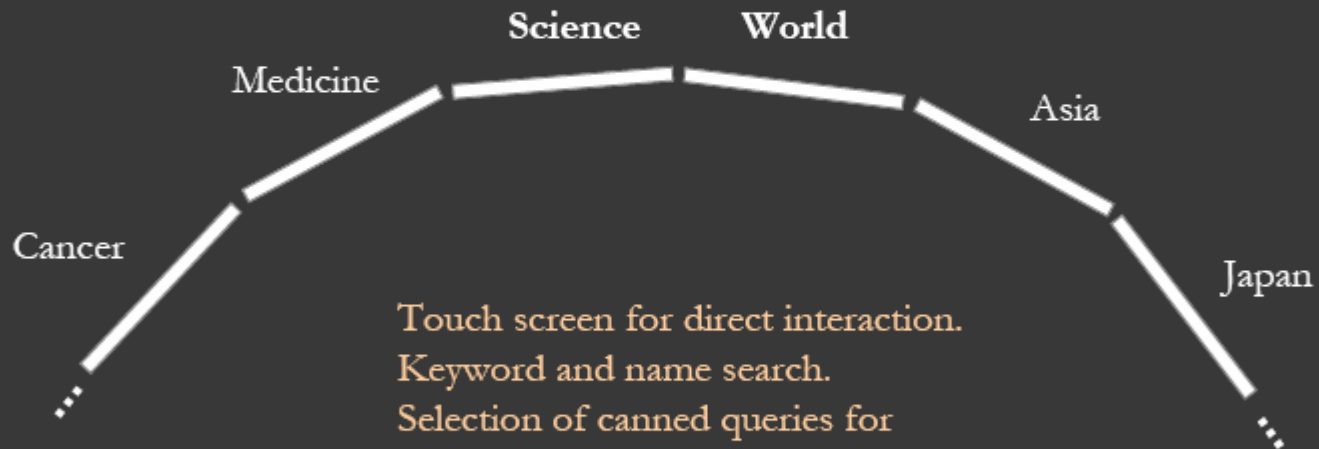
# GEOGRAPHIC MAP: WHERE SCIENCE GETS DONE



## Re-implementation of Illuminated Diagram Software (in progress)

by *Advanced Visualization Lab, Indiana University*

Drives unlimited number of ID screens.



Touch screen for direct interaction.

Keyword and name search.

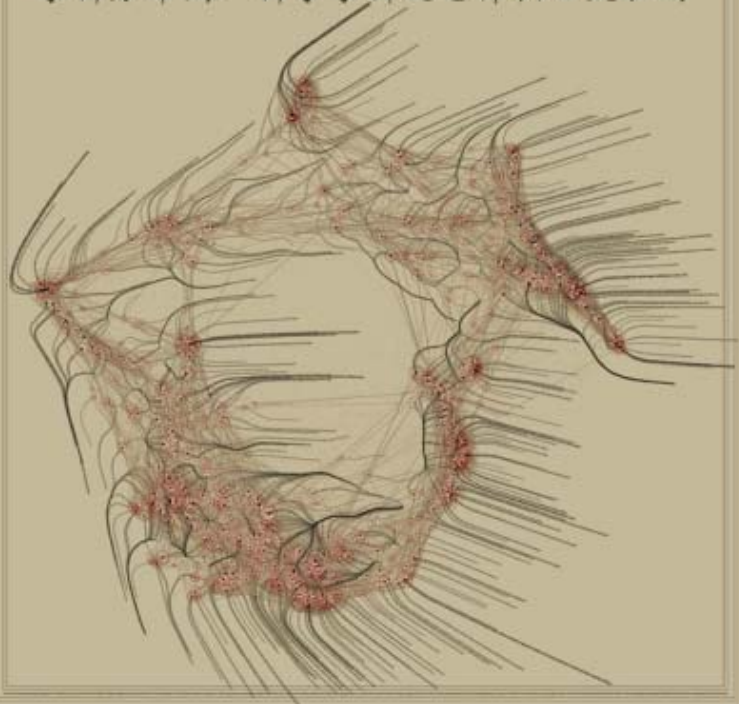
Selection of canned queries for

- interdisciplinary research areas
- famous people
- activity patterns, e.g., bursts, trends, etc.

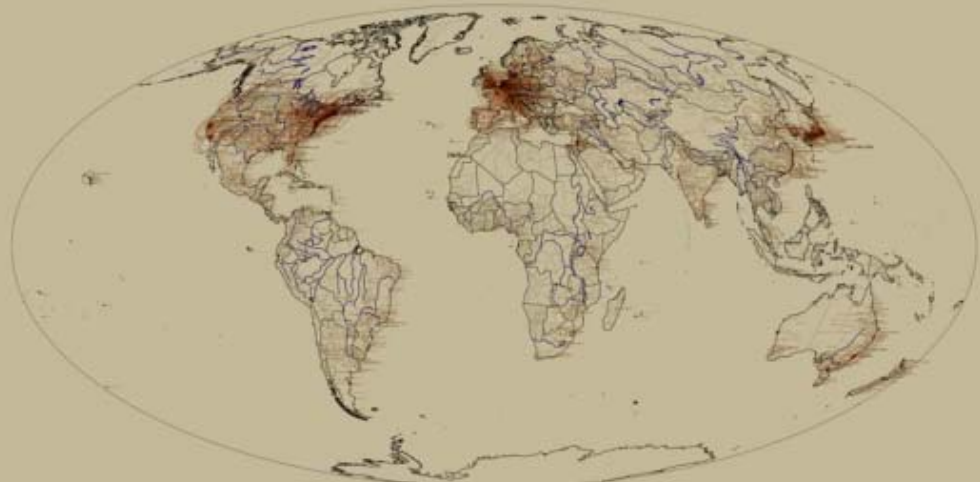




学科分布图：科学学科是怎样相互关联的



世界地图：科学研究在哪里进行着



你可以通过触摸屏在地图上随意指点来改变所到之处的光亮强度。当你触摸世界地图的某一点时，在那个地理位置上的所有研究机构会被点亮，同时在这些研究机构工作的学者的论文所属的学科会在学科分布图上被点亮。而当你触摸学科分布图的某一点时，在那个位置上的科学学科会被点亮，同时从事这些学科研究的研究机构在世界地图上的分布会被点亮。

## 纳米技术

### 探索科学学科的相互关联性

### 探索某个学者的科学著作的影响力的传播

这里显示所有和纳米技术相关的科学学科。纳米技术和科学研究人类在无形的空间里改造世界的的能力。这些空间存在于极其微小以至单个原子的结构中。目前大部分有关纳米的研究主要集中在物理、化学和材料科学领域。它们主要位于学科分布图上半部分的右面。不过，纳米技术在生物学和医药学研究里的应用也越来越多。生物学和医药学位于学科分布图下半部分的右面。

<p><b>所有科学学科</b></p> <p>显示所有776种科学学科</p>	<p><b>纳米技术</b></p> <p>有关微观粒子的科学</p>	<p><b>弗朗西·科里克</b></p> <p>DNA双螺旋状的发现者之一</p>	<p><b>阿尔伯特·爱因斯坦</b></p> <p>用相对论重新激活了物理学</p>	<p><b>迈克尔·费舍尔</b></p> <p>发现了物质转变模式的关键步骤</p>	<p><b>苏珊·费斯克</b></p> <p>研究人的认知是如何产生偏见的</p>
<p><b>可持续性</b></p> <p>一些与人类寄予长期希望相关的科学</p>	<p><b>化学和生物</b></p> <p>化学和生物科学的交叉部分</p>	<p><b>约舒亚·雷德伯格</b></p> <p>细菌遗传机制研究先驱</p>	<p><b>德里克·德索拉·普里斯</b></p> <p>著名的“科学计量学之父”</p>	<p><b>理查德·扎尔</b></p> <p>采用激光化学技术研究分子动态分布</p>	<p><b>关于本次展览</b></p> <p>与此展览相关人员和机构</p>

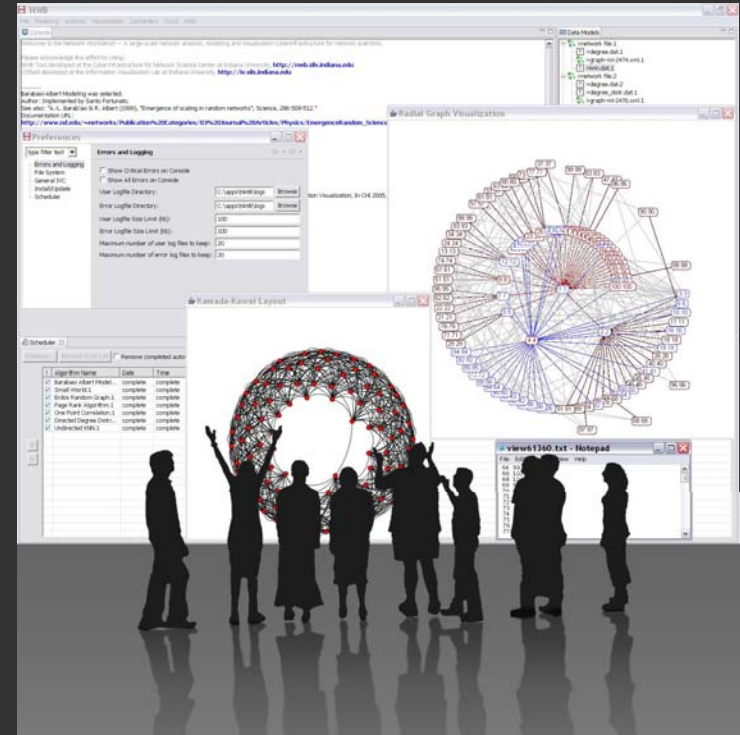
光标缓慢的扫过所有相互关联的科学学科，每一个学科以及从事这方面科学研究的研究机构在世界地图上的位置会被逐一点亮。首先，显示屏会点亮那些产出论文最多、最活跃的科学学科。然后那些小学科或冷门学科会被逐一点亮。

显示屏通过四步来展示某个学者对科学的贡献以及影响力的传播。首先，显示屏点亮该学者所发表的论文所属的学科在学科分布图上的位置以及该学者从事这项研究时所在的研究机构在世界地图上的位置。到目前为止，所有这些论文的引用率仍然很高。第二步，显示屏点亮所有引用在第一步中被点亮的原始论文的论文在学科分布图上的位置以及它们在世界地图上的位置。第三步，显示屏点亮所有引用了在第二步中被点亮的论文的论文在学科分布图上的位置以及它们在世界地图上的位置。第四步，显示屏点亮所有引用了在第三步中被点亮的论文的论文在学科分布图上的位置以及它们在世界地图上的位置。

# Designing Scholarly Marketplaces

## The Cyberinfrastructure for Network Science Center at IU serves

- Scholarly Database of 18 million scholarly records, <https://sdb.slis.indiana.edu>
- Information Visualization Cyberinfrastructure, <http://iv.slis.indiana.edu>
- Network Workbench Tool and Community Wiki, <http://nwb.slis.indiana.edu>



Börner, Katy, Sanyal, Soma & Vespignani, Alessandro. (2007). Network Science. In Cronin, Blaise (Eds.), *Annual Review of Information Science & Technology* (Vol. 41, pp. 537-607), chapter 12, Medford, NJ: Information Today, Inc./ American Society for Information Science and Technology.



Load Data

Select Preferences

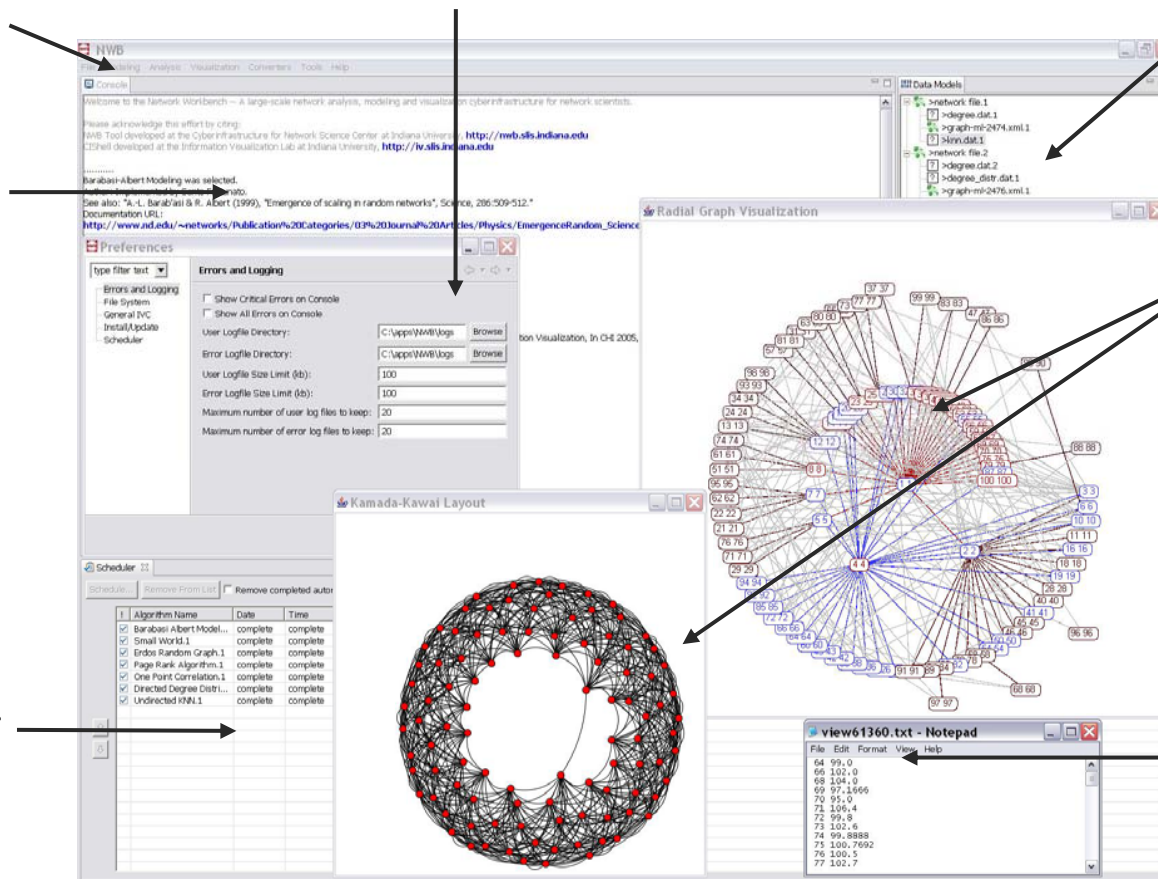
List of Data Models

Console

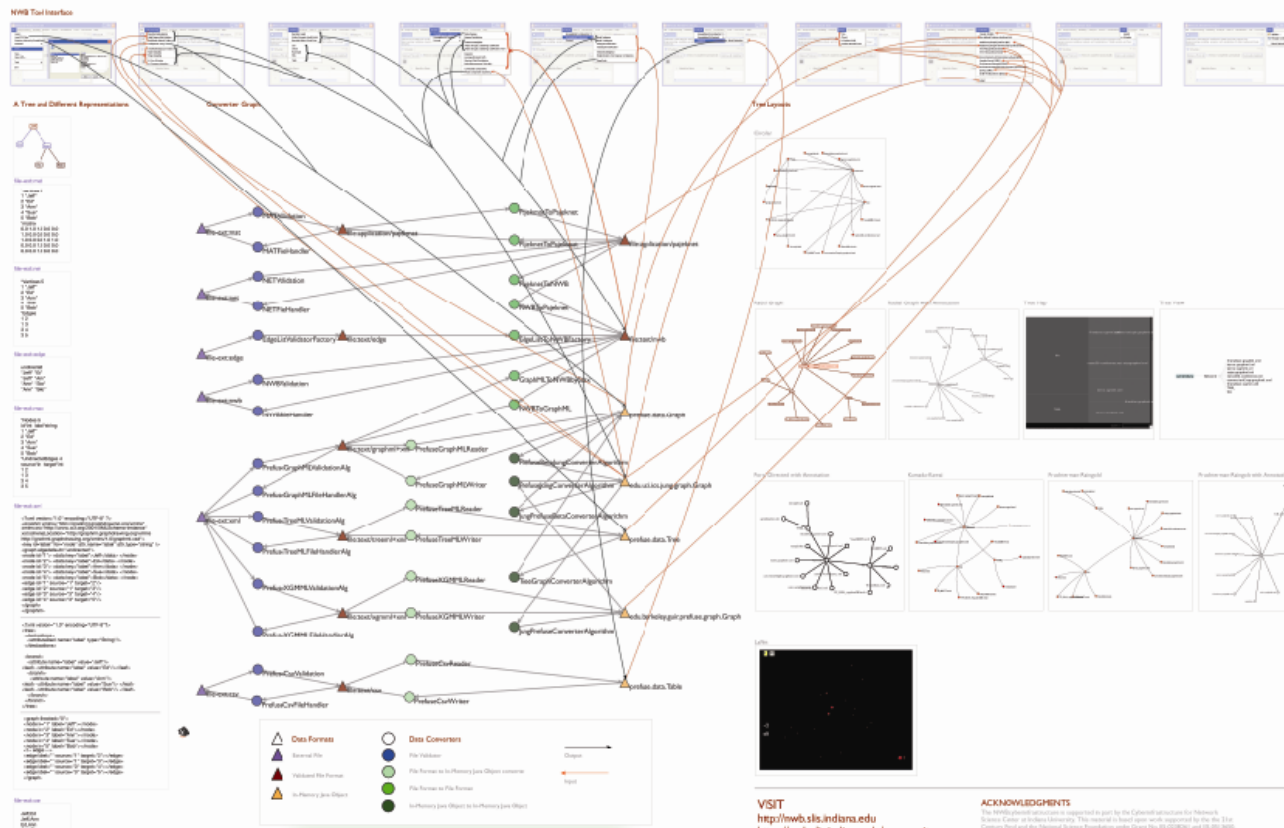
Visualize Data

Scheduler

Open Text Files



# Network Workbench Marketplace: An Ecology of Data Formats, Converters, and Algorithms



**INVESTIGATORS**  
 Dr. Katy Borner  
 Dr. Adam L. Galvani  
 Dr. Santiago Solari  
 Dr. Alexander Hoffmann  
 Dr. Steve Wasserman  
 Dr. Eric A. Wernitz  
 Indiana University

**DEVELOPERS**  
 Bruce Horv  
 Pascal Dubre  
 Timothy Kelley  
 Frank Water-Lorenz  
 Santa Fortunato  
 Sanjiv Kumar  
 Ranga Srikumar  
 Vamsi K. Ramesh  
 Indiana University

**PROJECT MANAGER**  
 Vamsi K. Ramesh  
 vkramesh@indiana.edu

**Network Workbench: A Large-Scale Network Analysis, Modeling and Visualization System for Biological, Social Science and Physics Research.**

The project will design, develop, and operate a unique distributed, shared resources environment for large-scale network analysis, modeling, and visualization, named Network Workbench (NWB). The environment data-computing resources environment will provide a one-stop online portal for researchers, educators, and practitioners interested in the study of biological, social and behavioral science, physics, and other networks.

The NWB will support network science research across scientific boundaries. Users of the NWB will have online access to super network datasets or can upload their own networks. They will be able to perform network analysis with the most effective algorithms available. In addition, they will be able to generate, run, and visualize network models to advance their understanding of the structure and dynamics of particular networks. NWB will provide advanced visualization tools to interactively explore and understand specific networks, as well as their interaction with other types of networks.

A major computer science challenge is the development of an algorithm language framework that supports the easy integration and dissemination of existing and new algorithms and can deal with multiple of network data formats in existence today. Another challenge is the design and implementation of an easy to use menu-based, online portal interface for interactive algorithm selection, data manipulation, user and session management. The NWB will be industrial in diverse research groups and educational settings in biology, social and behavioral science, and physics research. It will be well documented and available in open source for easy duplication and usage at other sites. An annual seminar school and a series of workshops and tutorials are planned to introduce the tool to diverse research communities.

The NWB will provide members of the scientific research community (college professors, physicists, engineers, scientists, social and behavioral scientists, engineers, etc.) with the means to carry out network analysis, modeling, and visualization projects in their own fields. This will result in a direct transfer of knowledge and results from the fields of applied network research to a wider scientific community. Researchers will have access to validated algorithms that in the past have been obtained through time-consuming personal developments of ad hoc computer programs. The NWB is designed to promote and encourage the empirical analysis and model validation of networks, providing an essential acceleration in the development of network science research. Online instructional material will support the use of the NWB in educational settings.

The NWB will provide a unique tool for network science researchers in many disciplines. In effect, NWB can display the knowledge accumulated in network theory and practice across scientific fields and use web-based or any interested researcher, practitioner, or student. The NWB shared resources environment will support and use network science applications and education in biology, social and behavioral science, and large infrastructure analysis, thereby accelerating the rate of scientific discovery.

**NWB Community Wiki**

Home | About | Contact | Search

Recent changes: [List of recent edits]

Users: [List of community members]

Pages: [List of wiki pages]

Categories: [List of categories]

Tools: [List of utility tools]

**VISIT**  
<http://nwb.slis.indiana.edu>  
<https://nwb.slis.indiana.edu/community>  
<http://www.cisshell.org>

**DOWNLOAD: NWB Tool**  
<http://nwb.slis.indiana.edu/software.html>

**ACKNOWLEDGMENTS**

The NWB development is supported in part by the Cyberinfrastructure for Network Science Center at Indiana University. The network is built upon work supported by the The Center for Policy and the Network Science Foundation under Grant No. 05-22030 and 05-22030. Any opinions, findings, and conclusions or recommendations expressed here are those of the authors and do not necessarily reflect the views of the National Science Foundation.





# Education – Learning Modules, NWB User and Developer Workshops

# NetworkWorkbench

A Workbench for Network Scientists

Print | Search:  Go

Custom Fillings / Home Page

- Main**
  - People
  - NWB Tool
  - Update Sites
  - Custom Fillings
- Datasets**
- Algorithms**
  - Related Work
  - FAQ
- Statistics

[DIGG IT!](#)  
[SUBMIT](#)  
[DEL.ICIO.US](#)  
[RSS](#)

## Custom Fillings

Many scientists use a very specific subset of algorithms and datasets in their work. Here, we link to custom fillings designed by different researchers. Descriptions of custom fillings frequently resemble learning modules providing an easy introduction into the working styles of different sciences.

### Physics

[Analysis of Large-Scale Networks](#) by Soma Sanyal.

### Biology

[Analysis of Biological N](#)

### Scientometrics

[Modeling the Co-Evolu](#)  
[Sanyal & Katy Börner,](#)  
[Map Your Bibtex File<sup>2</sup> t](#)  
[Semantic Analysis of S](#)

### Internet Research

[Error and Attack Toler](#)  
[Search Performance o](#)

### Others

[Data Conversion Servi](#)  
[Parallel Coordinate Vist](#)  
Bruce W. Herr II.

Please feel free to make y



Huang with students at the Complex System Summer School in Beijing, China