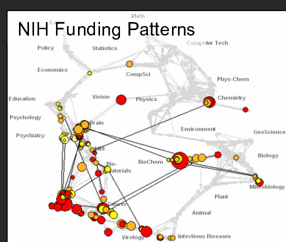
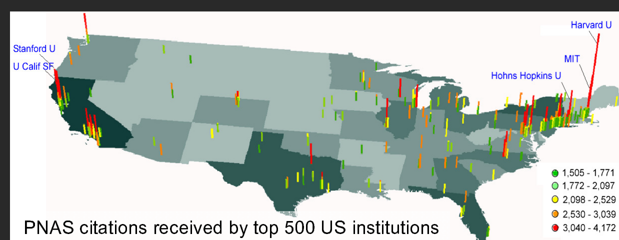
Towards a Macroscope for Decision Support in Science & Technology Policy

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

Workshop on Science of Science Policy: Developing our Understanding of Public Investments in Science
Organisation for Economic Cooperation and Development. Global Science Forum - 15th Meeting
July 12, 2006, Helsinki, Finland.





This Talk has Three Parts:

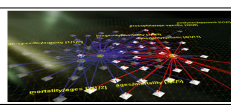
1. What is a Macroscope and why do we need it?
2. Macroscope supported decision making in S&T policy.
3. How to build a Macroscope for S&T policy?

2

This Talk has Three Parts:

1. **What is a Macroscope and why do we need it?**
2. Macroscope supported decision making in S&T policy.
3. How to build a Macroscope for S&T policy?

3



What is a Macroscope?

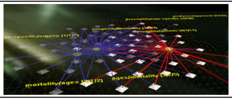
*"The **microscope** has permitted a dizzying plunge into the depths of living matter; it has made possible the discovery of the cell, microbes, and viruses; it has advanced the progress of biology and medicine.*

*The **telescope** has opened the mind to the immensity of the cosmos; it has traced the path of the planets and the stars and has prepared men for the conquest of space.*

*Today we are confronted with another infinite: the infinitely complex. We are confounded by the number and variety of elements, of relationships, of interactions and combinations on which the functions of large systems depend. … We must be able to understand them better in order to guide them better … a new tool is needed by all those … responsible for major decisions in politics, in science, and in industry or are ordinary people as we are. I shall call this instrument the **macroscope** (from macro, great, and skopein, to observe)."*

Quoted from:
Joël de Rosnay (1975) Le macroscope. Vers une vision globale. Editions du Seuil.
English Translation: Robert Edwards (1979) The Macroscope: A new world scientific system. Harper & Row, Publishers, Inc. Available at http://pespmc1.vub.ac.be/macroscope/index.htm.

4



What is a Macroscope? *(continued)*

- It is a tool that gives us a 'vision of the whole' and helps us 'synthesize' but also lets us detect patterns, trends, outliers and access details. *"It is not used to make things larger or smaller but to observe what is at once too great, too slow, and too complex for our eyes."*
- It does not only support holistic, complex systems based science but also analytic, reductionist science.
- It is an instrument that borrows methods and techniques from very different disciplines.
- It requires a cyberinfrastructure (e-science) that provides easy access to huge amounts of data, services, computing resources, and expertise/communities.

- It can be seen as a better kind of calculator and should be paid out of the 'calculator budget'.

Quoted from:
Joël de Rosnay (1975) Le macroscope. Vers une vision globale. Editions du Seuil.
English Translation: Robert Edwards (1979) The Macroscope: A new world scientific system. Harper & Row, Publishers, Inc. Available at http://pespmc1.vub.ac.be/macroscope/index.htm.

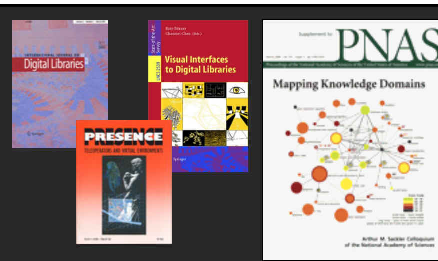
5

This Talk has Three Parts:

1. What is a Macroscope and why do we need it?
2. **Macroscope supported decision making in S&T policy.**
3. How to build a Macroscope for S&T policy?

6

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./American Society for Information Science and Technology, 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).*
- *Börner, Katy, Sanyal, Soma and Vespignani, Alessandro (in press).* ***Network Science.*** *In Blaise Cronin (Ed.), Annual Review of Information Science & Technology, Information Today, Inc./American Society for Information Science and Technology, Medford, NJ.*

- ***Places & Spaces: Mapping Science*** *exhibit, Currently on display at the SIBL branch of the New York Public Library.*

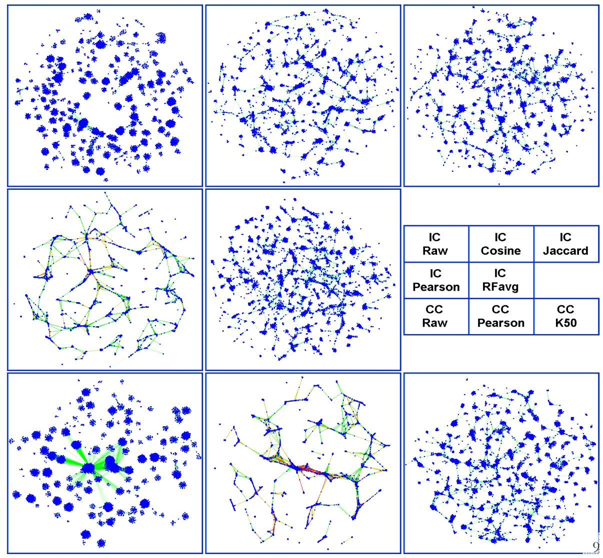
7

Towards a 'Base Map' or 'Reference System' of Science

8

Towards a 'Base Map' of Science

- ISI file year 2000, SCI and SSCI: 7,121 journals.
- Different similarity metrics
  - Inter-citation (raw counts, cosine, modified cosine, Jaccard, RF, Pearson)
  - Co-citation (raw counts, cosine, modified cosine, Pearson)
- Maps were compared based on
  - Regional accuracy,
  - Scalability of the similarity algorithm, and
  - Readability of the layouts.

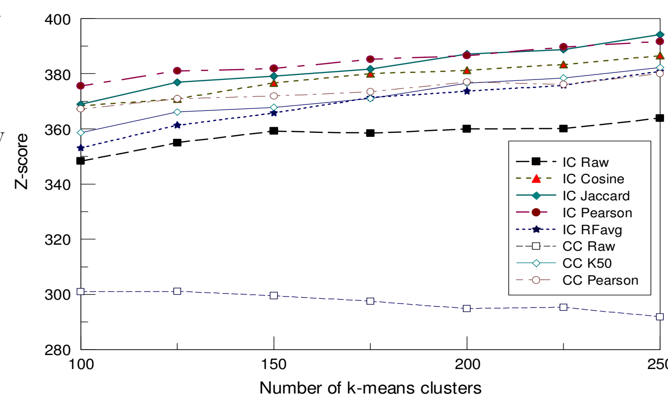


Boyack, Kevin W., Klavans, R. and Börner, Katy. (2005). Mapping the Backbone of Science. Scientometrics. 64(3), 351-374.

9

Selecting the similarity measure with the best regional accuracy

- For each similarity measure, the VxOrd layout was subjected to k-means clustering using different numbers of clusters.
- Resulting cluster/category memberships were compared to actual category memberships using entropy/mutual information method by Gibbons & Roth, 2002.
- Increasing Z-score indicates increasing distance from a random solution.
- Most similarity measures are within several percent of each other.

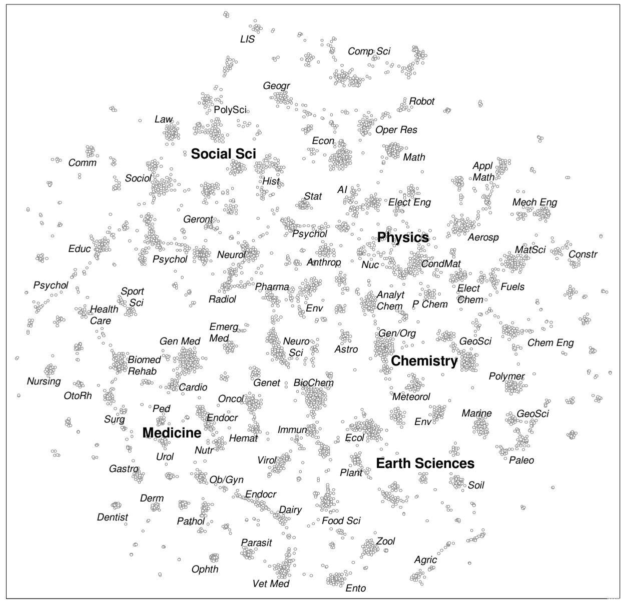


Boyack, Kevin W., Klavans, R. and Börner, Katy. (2005). Mapping the Backbone of Science. Scientometrics. 64(3), 351-374.

10

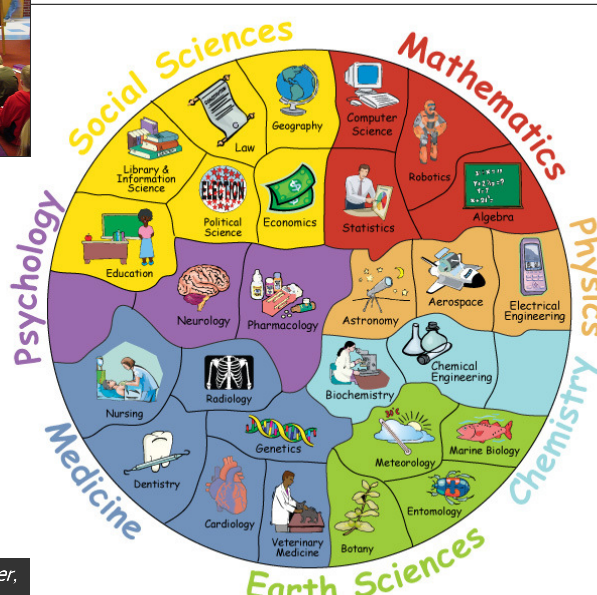
A 'Backbone' Map of Science & Social Science

- The map is comprised of 7,121 journals from year 2000.
- Each dot is one journal
- An *IC-Jaccard* similarity measure was used.
- Journals group by discipline.
- Groups are labeled by hand.
- Large font size labels identify major areas of science.
- Small labels denote the disciplinary topics of nearby large clusters of journals.



Boyack, Kevin W., Klavans, R. and Börner, Katy. (2005). Mapping the Backbone of Science. Scientometrics. 64(3), 351-374.

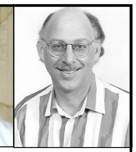
'Backbone' Map Rendered for Kids



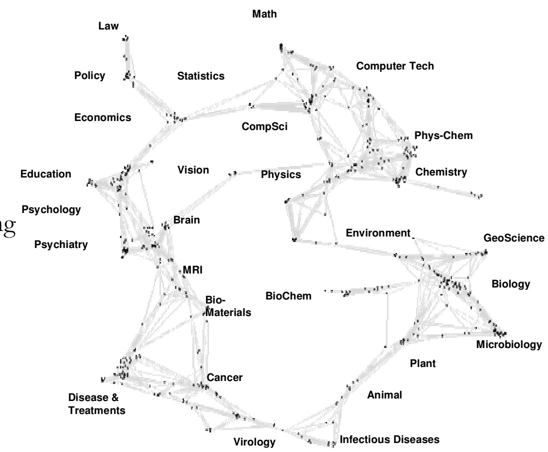
Ian Aliman, Nikki Roberg & Katy Börner, Science Maps for Kids.

Latest 'Base Map' of Science

Kevin W. Boyack & Richard Klavans, unpublished work.



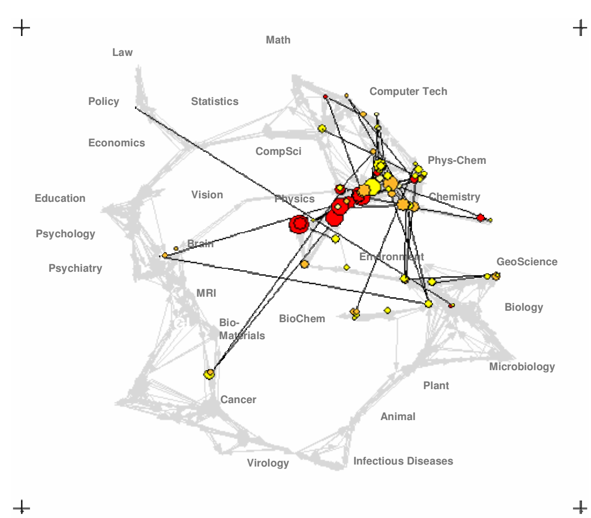
- Uses combined SCI/SSCI from 2002
  - 1.07M papers, 24.5M references, 7,300 journals
  - Bibliographic coupling of papers, aggregated to journals
- Initial ordination and clustering of journals gave 671 clusters
- Coupling counts were reaggregated at the journal cluster level to calculate the
  - (x,y)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



Science map applications: Identifying core competency

Kevin W. Boyack & Richard Klavans, unpublished work.

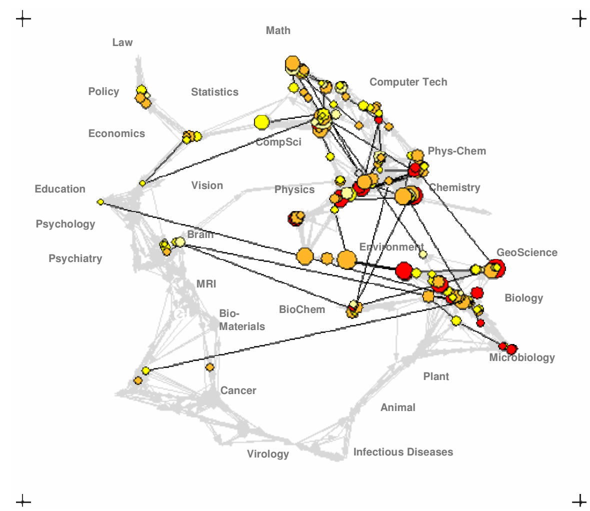
Funding patterns of the US Department of Energy (DOE)



Science map applications: Identifying core competency

Kevin W. Boyack & Richard Klavans, unpublished work.

Funding Patterns of the National Science Foundation (NSF)

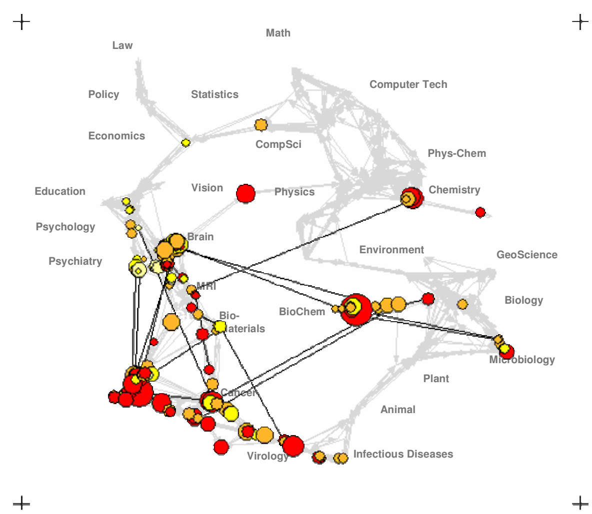


15

Science map applications: Identifying core competency

Kevin W. Boyack & Richard Klavans, unpublished work.

Funding Patterns of the National Institutes of Health (NIH)



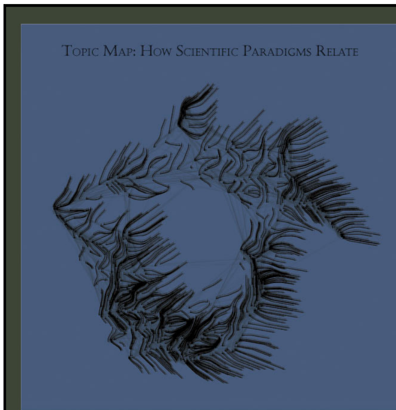
16

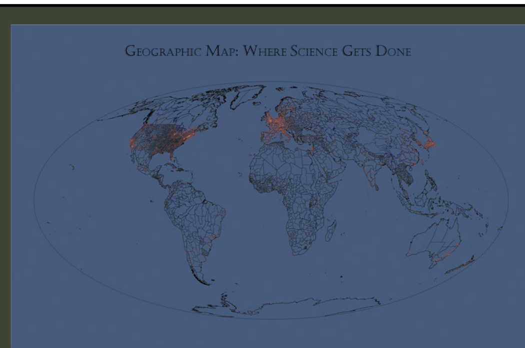
Coupling Science 'Topic Map' & 'Geographic Map'

17

The illuminated diagram maps and installations were created by Kevin W. Boyack (scientometrics and data shaping), John Burgoon (geographic mapmaking), Peter Kennard (system design and programming), Richard Klavans (scientometrics and node layout), W. Bradford Paley (typography, graphics, and interaction design); data courtesy of Thomson ISI; images © 2006.
They are part of Places & Spaces: Mapping Science" on display at the NYPL Science, Industry, and Business Library Madison/34th, New York City, April 3rd - August 31st, 2006.

18





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.

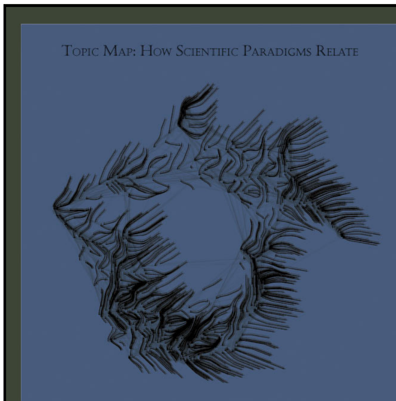
Nanotechnology

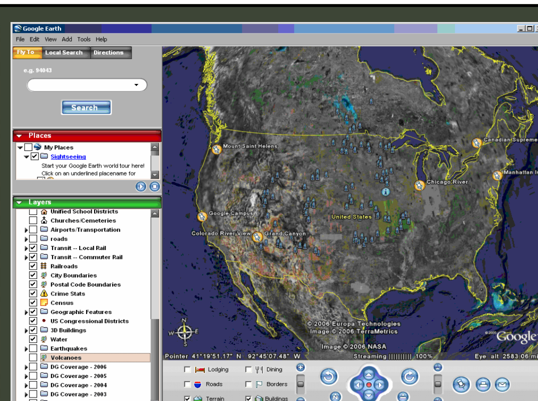
This overlay shows the distribution of nanotechnology within the paradigms of science. The majority of current work in nanotechnology takes places in physics, chemistry, and materials science, at the upper right portion of the map. However, an increasing amount of nanotechnology is being applied in the 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 shapshot lights science that cites the second; and the fourth lights science that cites the third.





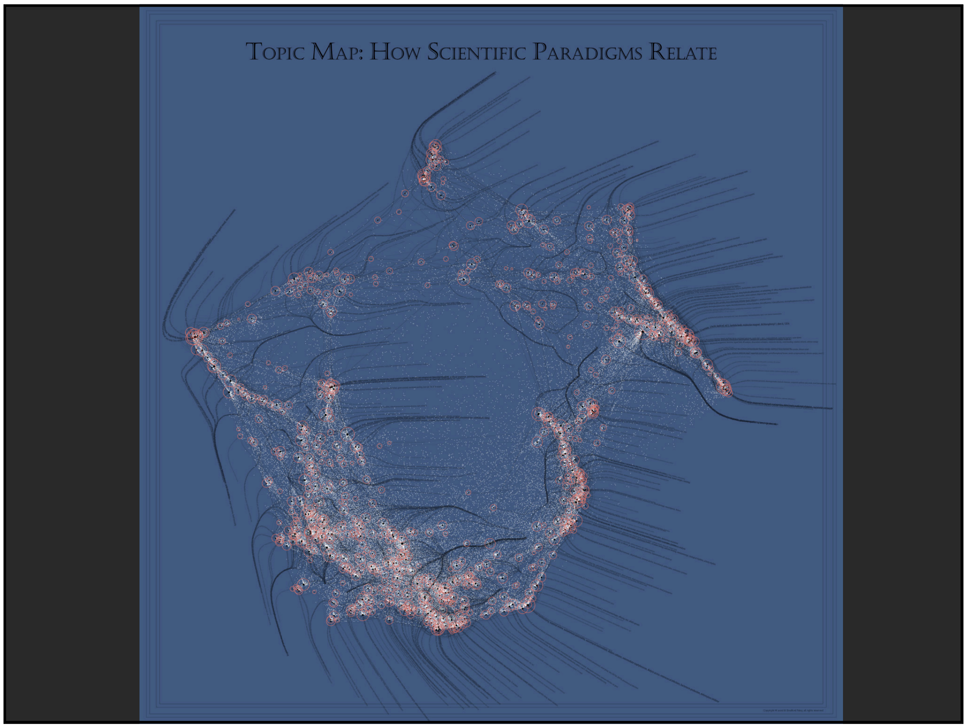
Nanotechnology

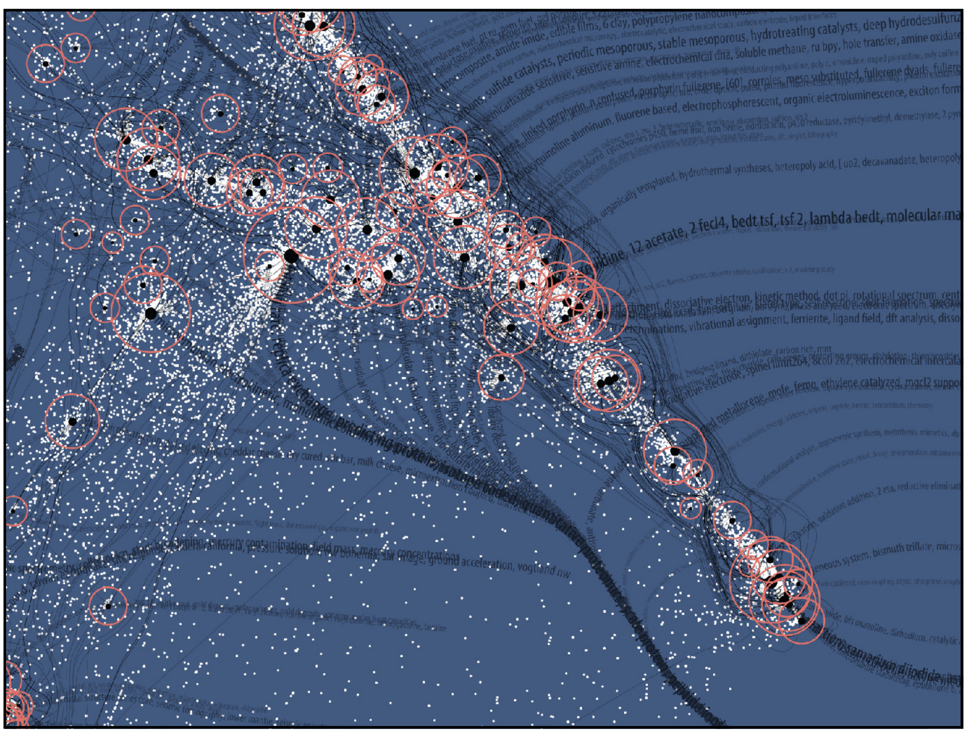
This overlay shows the distribution of nanotechnology within the paradigms of science. The majority of current work in nanotechnology takes places in physics, chemistry, and materials science, at the upper right portion of the map. However, an increasing amount of nanotechnology is being applied in the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

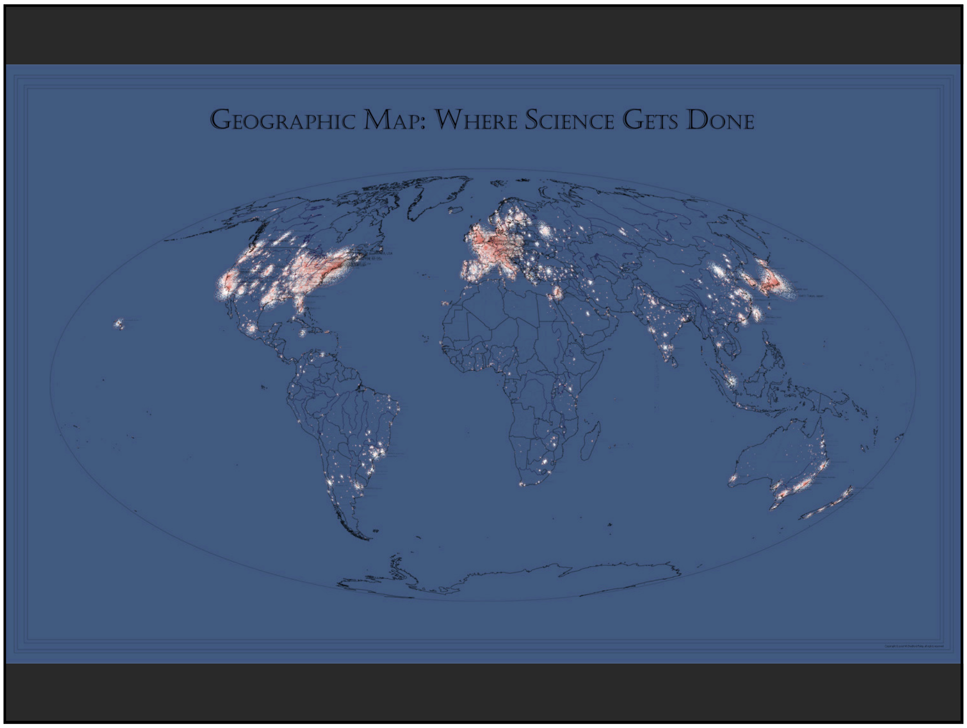
All Topics *Sweep through all 776 scientific paradigms*	**Nanotechnology** *Science on the tiny scale of molecules*	**Francis H. C. CRICK** *Co-discovered DNA's double helix*	**Albert EINSTEIN** *Revitalized physics with Relativity theories*	**Michael E. FISHER** *Models critical phase transitions of 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*

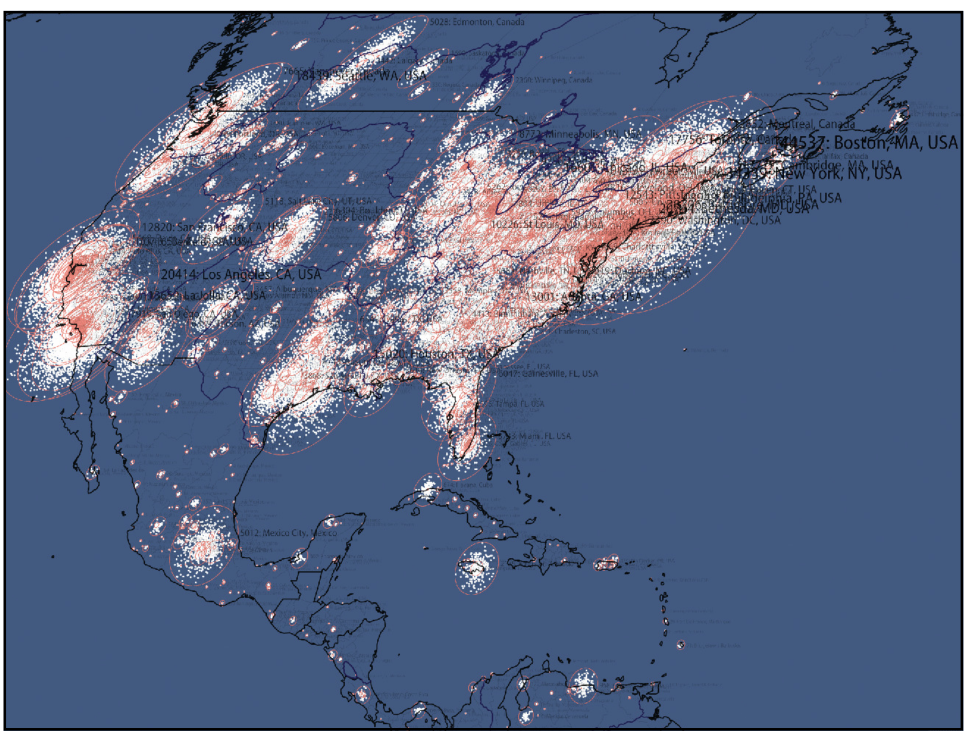
We sweep slowly through adjoining related topics, lighting up the places in the world that study each topic. You may select a subset of the topics that deal with these three interesting subjects by touching it.

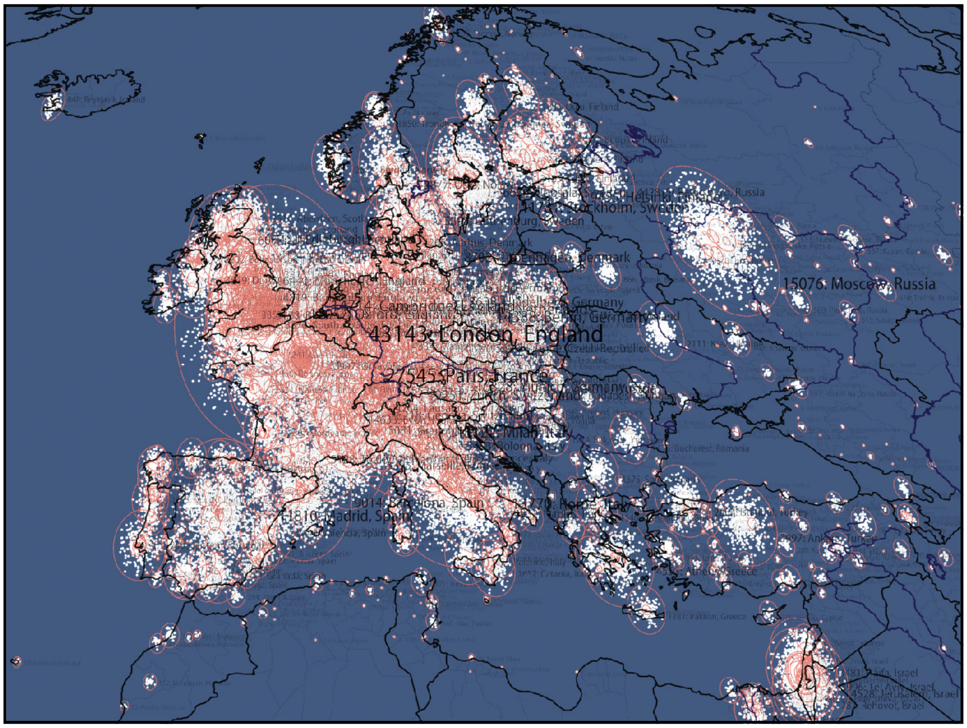
A single person's spreading influence is shown as a series of four snapshots. First, we light only topics and places relating to that person's papers—papers that are still highly cited today. The second lights everything that cites that original work. Note that this first-generation impact extends to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


Topic Map: How Scientific Paradigms Relate

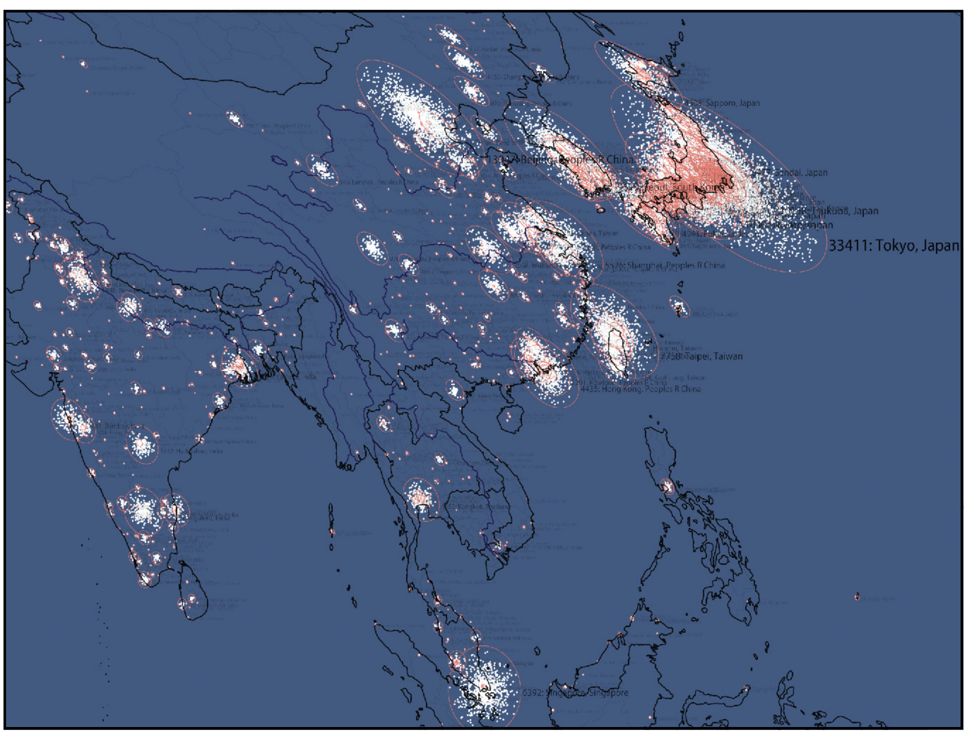



GEOGRAPHIC MAP: WHERE SCIENCE GETS DONE

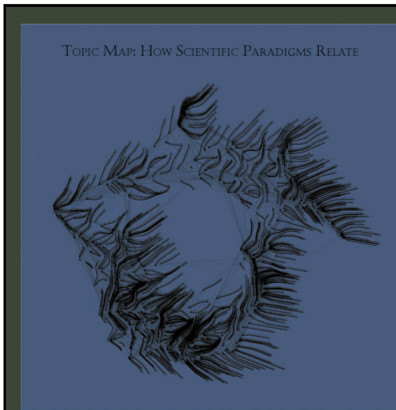

5028: Edmonton, Canada
2300: Winnipeg, Canada
12820: San Francisco, CA, USA
20414: Los Angeles, CA, USA
5118: Salt Lake City, UT, USA
8772: Minneapolis, MN, USA
10226: St Louis, MO, USA
13001: Atlanta, GA, USA
15020: Houston, TX, USA
5017: Gainesville, FL, USA
44537: Boston, MA, USA
Washington, DC, USA
Charleston, SC, USA
5012: Mexico City, Mexico

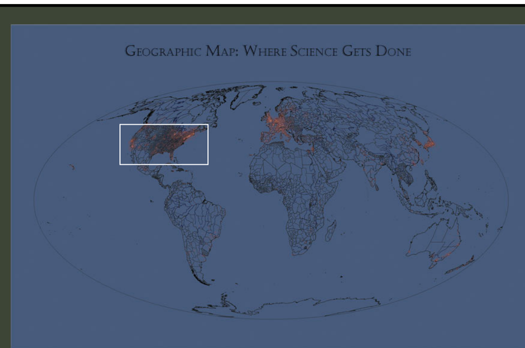


43143: London, England
15076: Moscow, Russia



33411: Tokyo, Japan





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.

Use **Geographic Map** to zoom into U.S.
Request to see **citation patterns** and
the **impact of the Internet on citation behavior**.

27

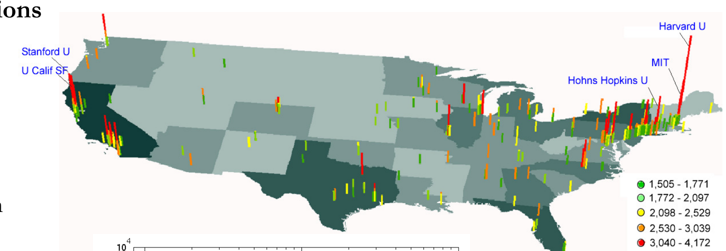
Spatio-Temporal Information Production and Consumption of Major U.S. Research Institutions

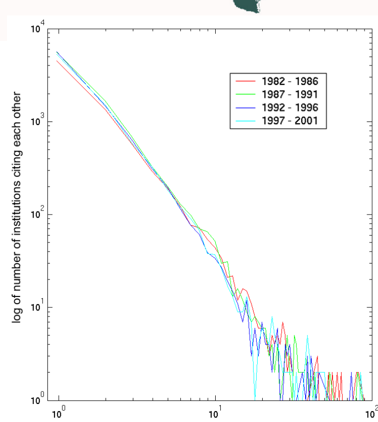
Börner & Penumarthy. (2005)

Does Internet lead to more global citation patterns, i.e., more citation links between papers produced at geographically distant research instructions?

Analysis of top 500 most highly cited U.S. institutions.

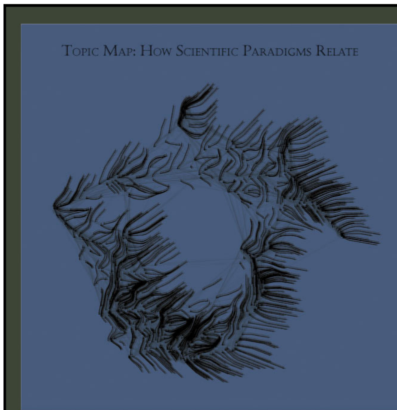
Each institution is assumed to produce and consume information.

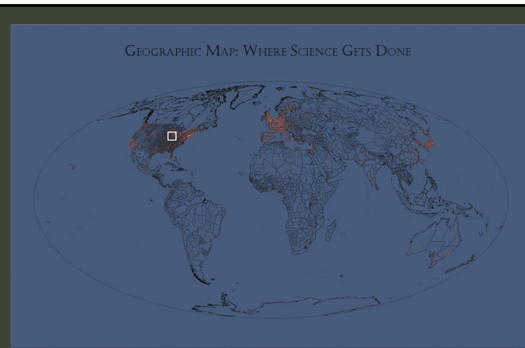




$\gamma_{82\text{-}86} = 1.94$ (R^2=91.5%)
$\gamma_{87\text{-}91} = 2.11$ (R^2=93.5%)
$\gamma_{92\text{-}96} = 2.01$ (R^2=90.8%)
$\gamma_{97\text{-}01} = 2.01$ (R^2=90.7%)

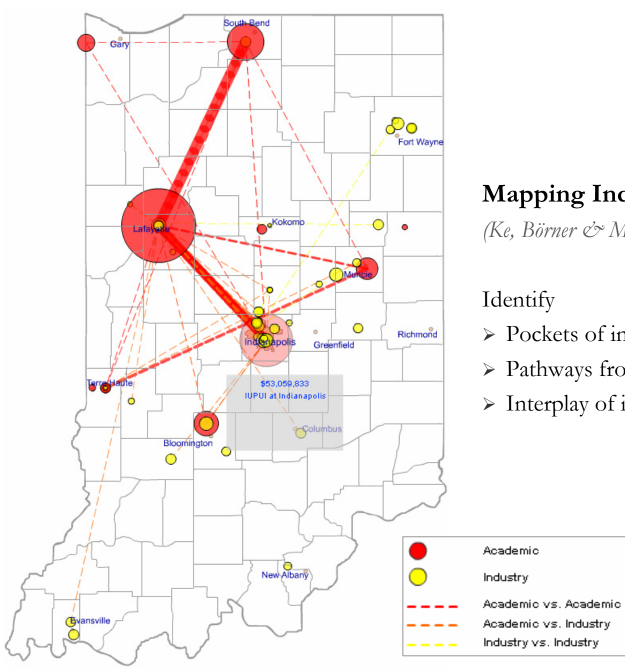
28





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.

Use **Geographic Map** to zoom into Indiana.
Request to see **pockets of innovation, pathways from ideas to products** and the **interplay of industry and academia.**

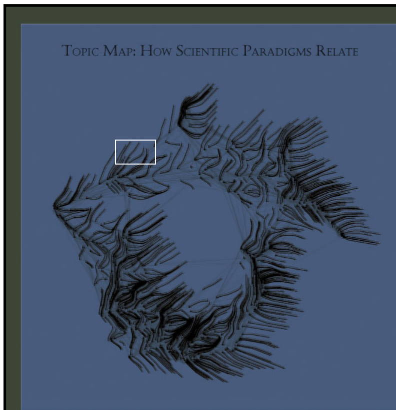


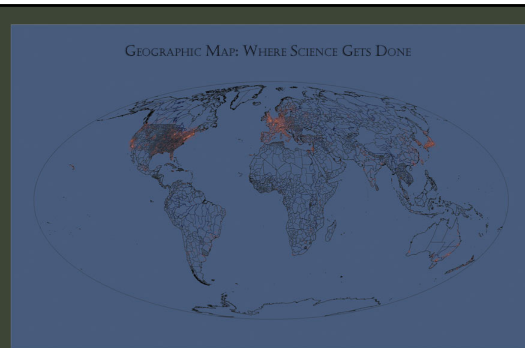
Mapping Indiana's Intellectual Space

(Ke, Börner & Mei, 2005)

Identify

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



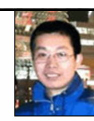


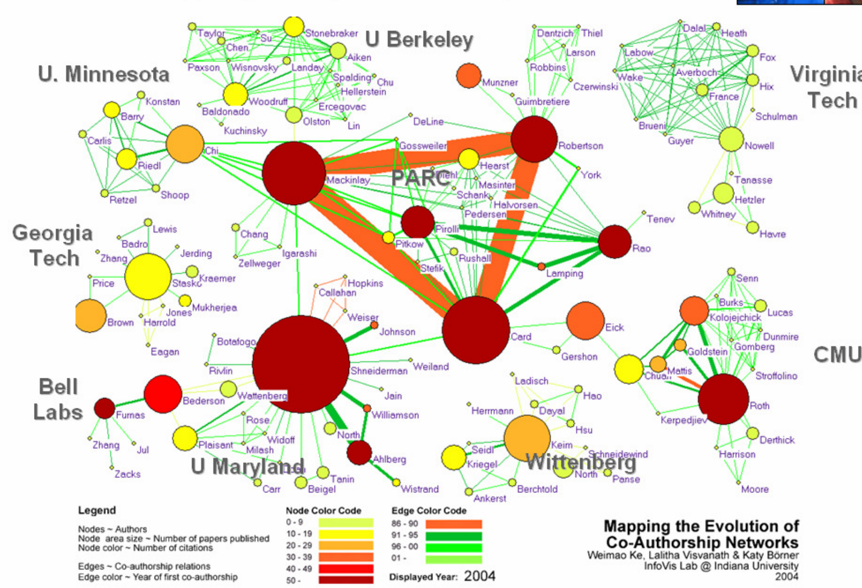
You may run your finger over each of these maps to control the lighting on the other: touching a place on the world map will light up topics studied in that place; touching a paradigm on the topic map will light up the places that study that topic.

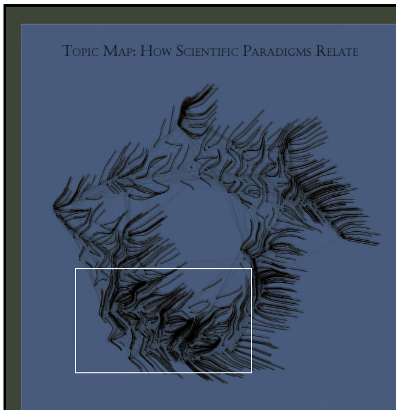
Use **Topic Map** to zoom into Information Visualization
Request to see the **evolution of co-authorship relations** over time.

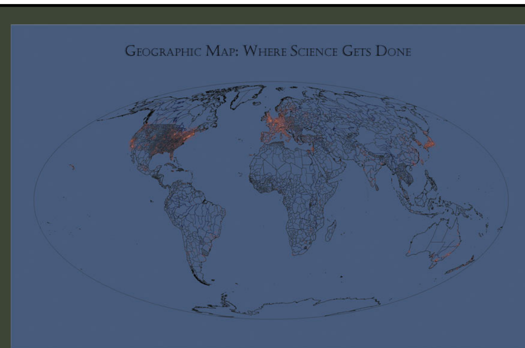
Mapping the Evolution of Co-Authorship Networks

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







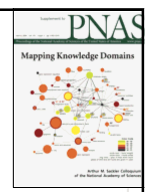


You may run your finger over each of these maps to control the lighting on the other: touching a place on the world map will light up topics studied in that place; touching a paradigm on the topic map will light up the places that study that topic.

Use **Topic Map** to zoom into **Biomedical Research**.
Request to see the **main stream and 'bursting' research topics.**

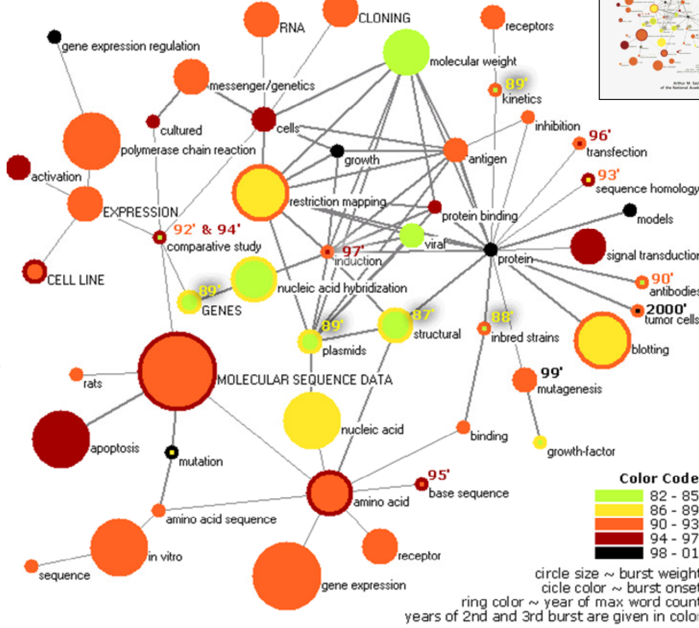
33

Mapping Topic Bursts

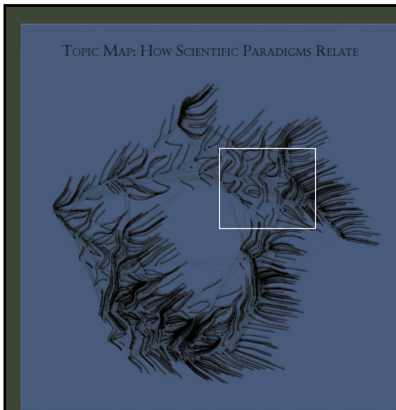


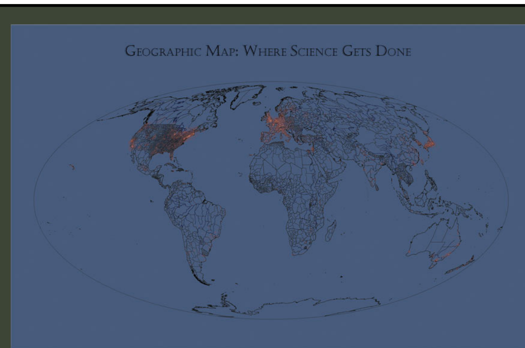
Co-word space of the top 50 highly frequent and bursty words used in the top 10% most highly cited PNAS publications in 1982-2001.

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



34





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.

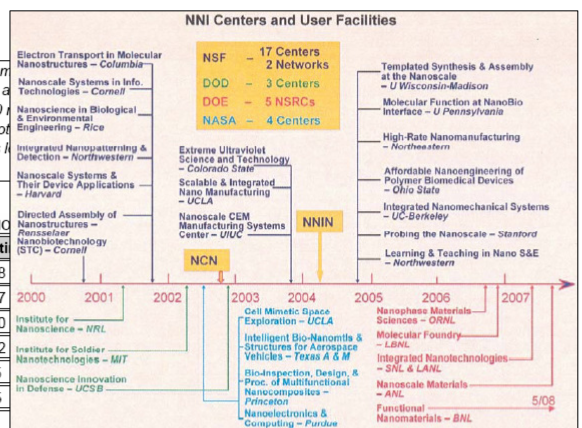
Use **Topic Map** to zoom into **Nanotechnology**.
Request to see **case studies.**

Role and Contribution of the DOE's Nanoscale Science Research Centers. NSRC Report.

"Nanotechnology is the understanding and control of m nanometers, where unique phenomena enable novel a billionth of a meter; a sheet of paper is about 100,000 r nanoscale science, engineering and technology, nanot measuring, modeling, and manipulating matter at this l September 2005).

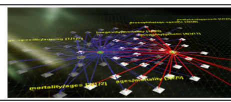
TABLE 3.2. NNI BUDGET, 2004-2006 (DOLLARS IN MILLIO

	2004 Actual	2005 Esti	
NSF	256	338	
DOD*	291	257	
DOE	202	210	
HHS (NIH)	106	142	
DOC (NIST)	77	75	
NASA	47	45	
USDA	2	3	
EPA	5	5	5
HHS (NIOSH)			
DOJ	2		
DHS	1	1	1
TOTAL	**989**	**1,081**	**1,054**



"My biggest worry is we come into the center but carrying pre-conceived notions—people think of moving their personal lab into the new building. We have to do science in a new way." (Scientist, CINT)

Mohrman, Susan A. and Wagner, Caroline S. (July 2006) The Dynamics of Knowledge Creation: A Baseline for the Assessment of the Role and Contribution of the Department of Energy's Nanoscale Science Research Centers, NSRC Final Report. Available at http://www.marshall.usc.edu/ceo/projects/doe/NSRC_final_report.pdf



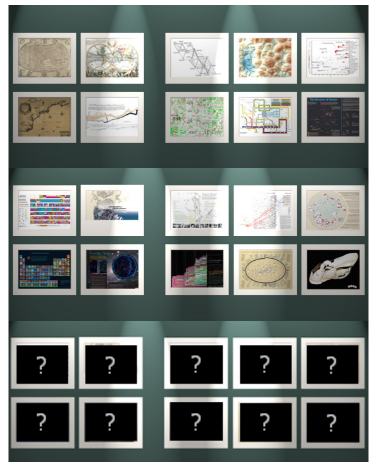
Places & Spaces: Mapping Science

A **10-year** science exhibit designed to demonstrate the power of maps to navigate physical places and abstract topic spaces, see http://vw.indiana.edu/places&spaces.

1st Iteration in 2005
The Power of Maps: Four early maps of our world with vs. six early maps of science.

2nd Iteration in 2006
The Power of Reference Systems: Four existing reference systems vs. six potential reference systems of science.

3rd Iteration in 2007
The Power of Forecasts: Four existing forecasts vs. six potential science 'weather' forecasts.



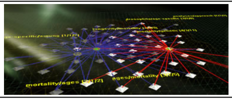
We are looking for display venues.

37

This Talk has Three Parts:

1. What is a Macroscope and why do we need it?
2. Macroscope supported decision making in S&T policy.
3. **How to build a Macroscope for S&T policy?**

38



How to build a Macroscope for S&T policy?

Building a Macroscope requires

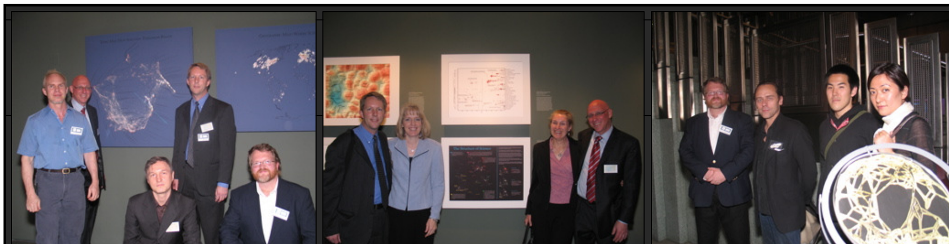
- **Combining the best algorithms,** expertise, and practices world-wide and science-wide.
- Building a powerful **cyberinfrastructure** that can process terabytes of streaming data in a distributed fashion.
- Agreeing on standards and solving privacy and political issues.
- Funding – out of the 'calculator budget'.

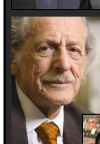
Using a Macroscope requires

- **Access to high quality data**.
- Knowledge and expertise in how to 'tune/focus' the macroscope and how to interpret results.

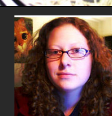
Atkins, D.E., Drogemeier, K.K., Feldman, S.I., Garcia-Molina, H., Klein, M.L., Messerschmitt, D.G., Messian, P., Ostriker, J.P. and Wright, M.H. Revolutionizing science and engineering through cyberinfrastructure: Report of the National Science Foundation blue-ribbon advisory panel on cyberinfrastructure. National Science Foundation, Arlington, VA, 2003. Available at http://www.communitytechnology.org/nsf_ci_report.

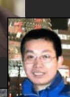
39

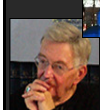


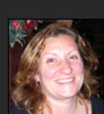


I would like to thank all my collaborators and sponsors. Kevin W. Boyack, Ann B. Carlson, Blaise Cronin, Eugene Garfield, Loet Leydesdorff, Bill Valdez, Caroline Wagner, and Maria Zemankova provided expert advise.



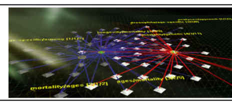






SUN microsystems | SBC | HPNAP High Performance Network Applications Program | Indiana 21st Century Research & Technology Fund | NSF | National Institutes of Health

40

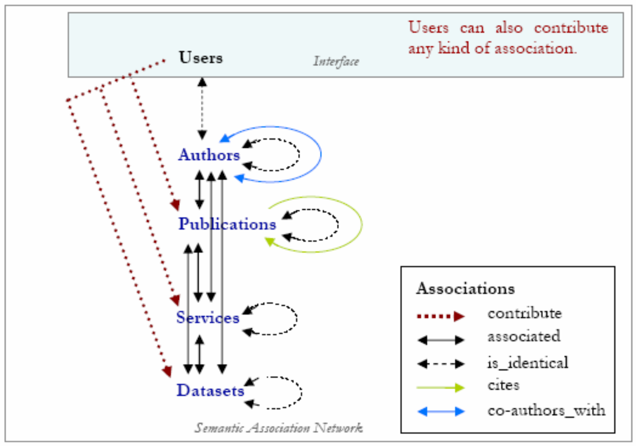


References

- Börner, Katy. Mapping All of Science: How to Collect, Organize and Make Sense of Mankind's Scholarly Knowledge and Expertise. Accepted for *Environment and Planning B.*
- Börner, Katy, Penumarthy, Shashikant, Meiss, Mark and Ke, Weimao. Mapping the Diffusion of Scholarly Knowledge Among Major U.S. Research Institutions. Accepted for *Scientometrics.*
- Holloway, Todd, Božicevic, Miran and Börner, Katy. Analyzing and Visualizing the Semantic Coverage of Wikipedia and Its Authors. Accepted for *Complexity*, Special issue on *Understanding Complex Systems.* Also available as cs.IR/0512085.
- Katy Börner. (2006) Semantic Association Networks: Using Semantic Web Technology to Improve Scholarly Knowledge and Expertise Management. In Vladimir Geroimenko & Chaomei Chen (eds.) *Visualizing the Semantic Web*, Springer Verlag, 2nd Edition, chapter 11, pp. 183-198.
- Boyack, Kevin W., Klavans, R. and Börner, Katy. (2005). Mapping the Backbone of Science. *Scientometrics,* 64(3), 351-374.
- Hook, Peter A. and Börner, Katy. (2005) Educational Knowledge Domain Visualizations: Tools to Navigate, Understand, and Internalize the Structure of Scholarly Knowledge and Expertise. In Amanda Spink and Charles Cole (eds.) *New Directions in Cognitive Information Retrieval.* Springer-Verlag, Netherlands, chapter 5, pp. 187-208.
- Börner, Katy, Dall'Asta, Luca, Ke, Weimao and Vespignani, Alessandro. (April 2005) Studying the Emerging Global Brain: Analyzing and Visualizing the Impact of Co-Authorship Teams. *Complexity*, special issue on *Understanding Complex Systems*, 10(4): pp. 58 - 67. Also available as cond-mat/0502147.
- Ord, Terry J., Martins, Emília P., Thakur, Sidharth, Mane, Ketan K., and Börner, Katy. (2005) Trends in animal behaviour research (1968-2002): Ethoinformatics and mining library databases. *Animal Behaviour*, 69, 1399-1413. Supplementary Material.
- Mane, Ketan K. and Börner, Katy. (2004). Mapping Topics and Topic Bursts in PNAS. *Proceedings of the National Academy of Sciences of the United States of America*, 101(Suppl. 1):5287-5290.
- Börner, Katy, Maru, Jeegar and Goldstone, Robert. (2004). The Simultaneous Evolution of Author and Paper Networks. *Proceedings of the National Academy of Sciences of the United States of America*, 101(Suppl_1):5266-5273.

The End.

Access to High Quality Data – The Utility of Semantic Association Networks



- *Katy Börner. (2006) Semantic Association Networks: Using Semantic Web Technology to Improve Scholarly Knowledge and Expertise Management. In Vladimir Geroimenko & Chaomei Chen (eds.) Visualizing the Semantic Web, Springer Verlag, 2nd Edition, chapter 11, pp. 183-198.*
- *Katy Börner. Mapping All of Science: How to Collect, Organize and Make Sense of Mankind's Scholarly Knowledge and Expertise. Accepted for Environment and Planning B.*

Collect and Integrate Data & Algorithms Using CIShell

CIShell is a plug-in based architecture or 'empty shell' that supports

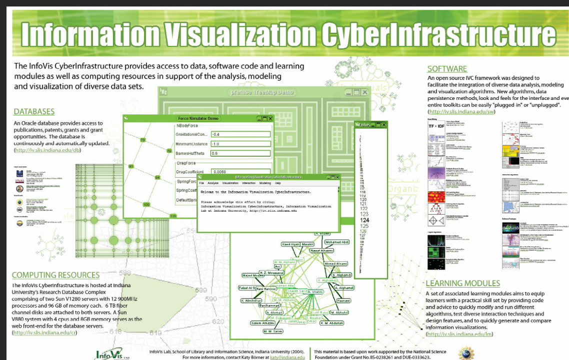
- Easy integration of new datasets and algorithms by algorithm developers and
- Easy usage of algorithms by algorithm users.

Its plug-and-play architecture supports the integration and utilization of diverse

- Datasets, e.g., stored in files, databases, steaming data.
- Algorithms, e.g., data processing, analysis, modeling, visualization.
- Interfaces, e.g., remote services, scripting engines, peer-to-peer clients.
- Services, e.g., workflow support, scheduler.

CIShell is built upon the Open Services Gateway Initiative (OSGi) Framework and available at http://sourceforge.net/projects/cishell.

Weixia Huang, Bruce Herr & Katy Börner (May 2006) CIShell - A Plug-in Based Architecture for the Integration of Algorithms and Data Models. Network Science Workshop and Conference, Bloomington, Indiana. Available online at http://vw.indiana.edu/netsci06/conference/Huang_CIShell.pdf.



Building Powerful Cyberinfrastructures that can process terabytes of data in a distributed fashion.

CAREER: Visualizing Knowledge Domains. NSF IIS-0238261 award (Katy Börner, $440,000) Sept. 03-Aug. 08.
http://iv.slis.indiana.edu/

SEI: Network Workbench: A Large-Scale Network Analysis, Modeling and Visualization Toolkit for Biomedical, Social Science and Physics Research. NSF IIS-0513650 award (Katy Börner, Albert-Laszlo Barabasi, Santiago Schnell, Alessandro Vespignani & Stanley Wasserman, Craig Stewart (Senior Personnel), $1,120,926) Sept. 05 - Aug. 08. http://nwb.slis.indiana.edu

45