

Towards Effective KM Tools

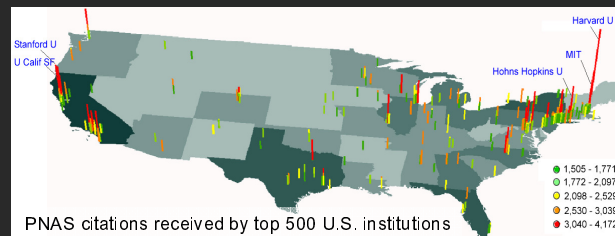
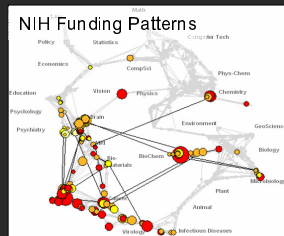
Dr. Katy Börner

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

katy@indiana.edu

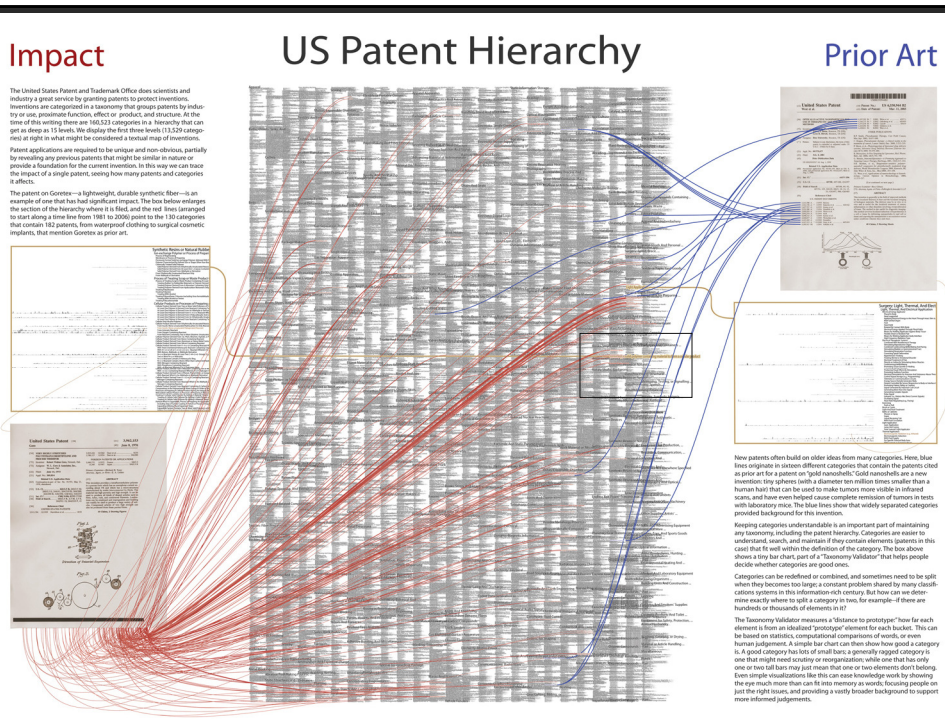


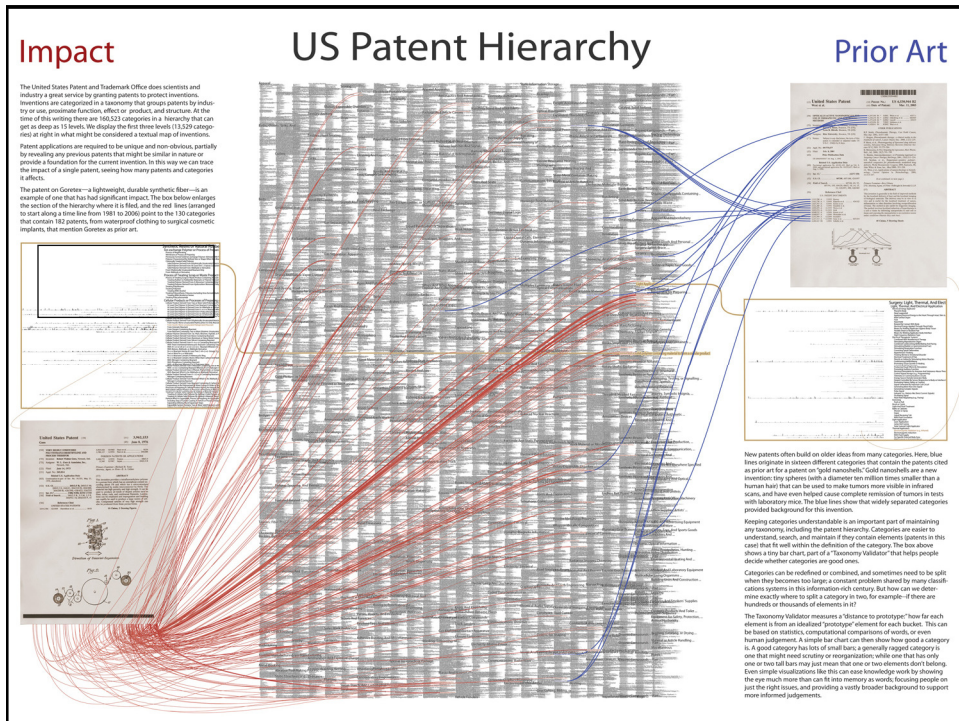
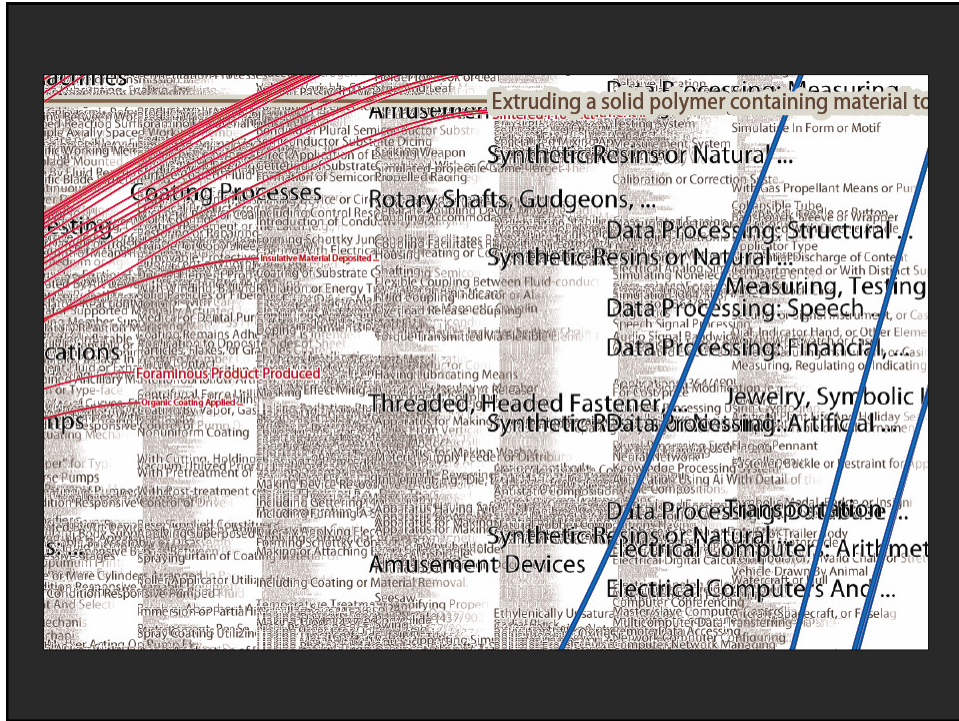
KEBR, December 11-12, 2006

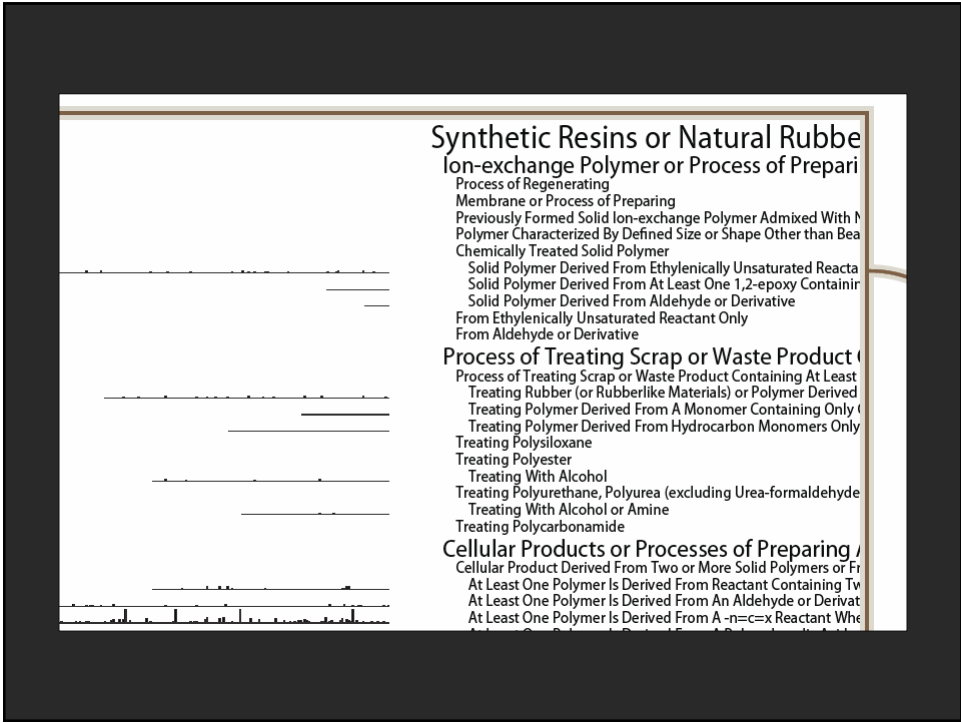


- Taxonomies / Ontologies / Classification Hierarchies
- Mapping Knowledge Spaces such as Science
- Building Data/Software/Resources Bazaars not Cathedrals

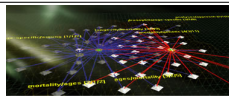
Taxonomies / Ontologies / Classification Hierarchies







Mapping Knowledge Spaces such as Science



Process of Mapping Knowledge Domains

DATA EXTRACTION	UNIT OF ANALYSIS	MEASURES	LAYOUT (often one code does both similarity and ordination steps)		DISPLAY
			SIMILARITY	ORDINATION	
SEARCHES ISI INSPEC Eng Index Medline ResearchIndex Patents etc.	COMMON CHOICES Journal Document Author Term	COUNTS/FREQUENCIES Attributes (e.g. terms) Author citations Co-citations By year THRESHOLDS By counts	SCALAR (unit by unit matrix) Direct citation Co-citation Combined linkage Co-word / co-term Co-classification VECTOR (unit by attribute matrix) Vector space model (words/terms) Latent Semantic Analysis (words/terms) incl. Singular Value Decomp (SVD) CORRELATION (if desired) Pearson's R on any of above	DIMENSIONALITY REDUCTION Eigenvector/ Eigenvalue solutions Factor Analysis (FA) and Principal Components Analysis (PCA) Multi-dimensional scaling (MDS) LSA, Topics Pathfinder networks (PFNet) Self-organizing maps (SOM) includes SOM, ET-maps, etc. CLUSTER ANALYSIS SCALAR Triangulation Force-directed placement (FDP)	INTERACTION Browse Pan Zoom Filter Query Detail on demand ANALYSIS

Börner, Chen & Boyack... (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.

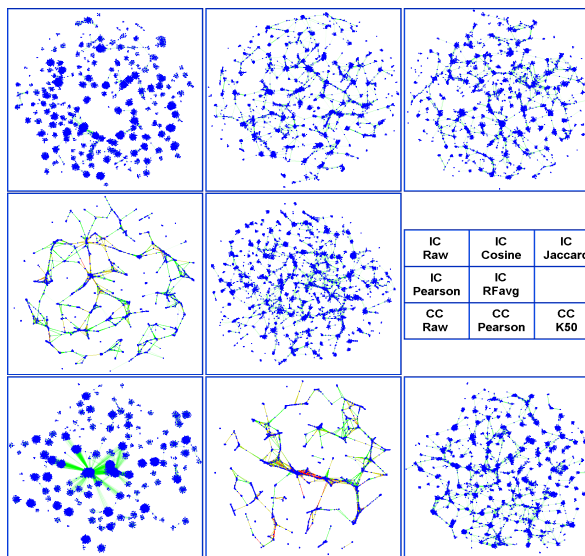


9

Comparison of Similarity Metrics



- 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,
 - the scalability of the similarity algorithm, and
 - the readability of the layouts.

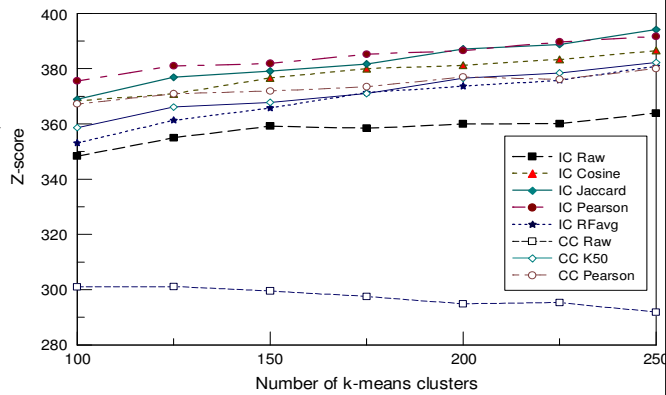


Boyack, Kevin W., Klavans, R. and Börner, Katy. (2005). *Mapping the Backbone of Science*. *Scientometrics*. 64(3), 351-374.

10

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.



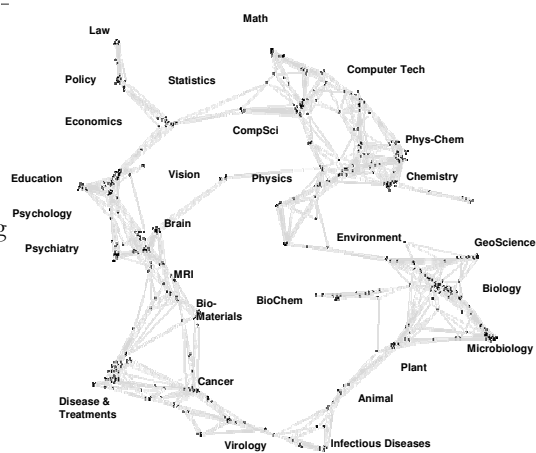
Boyack, Kevin W., Klavans, R. and Börner, Katy. (2005).
Mapping the Backbone of Science. *Scientometrics*. 64(3), 351-374.

11

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

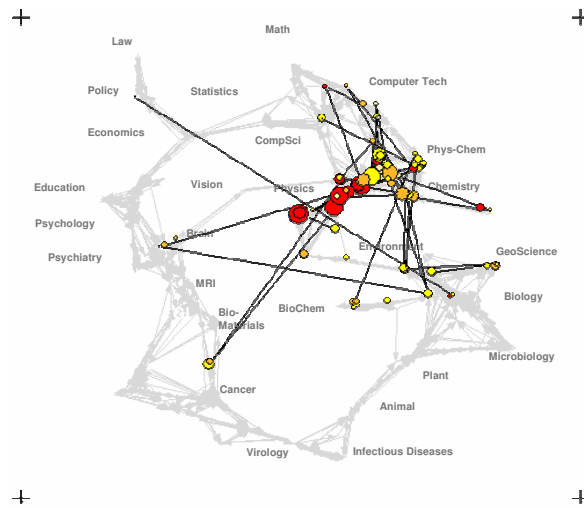


12

Science map applications: Identifying core competency

Kevin W. Boyack & Richard Klavans, unpublished work.

Funding patterns of the US Department of Energy (DOE)

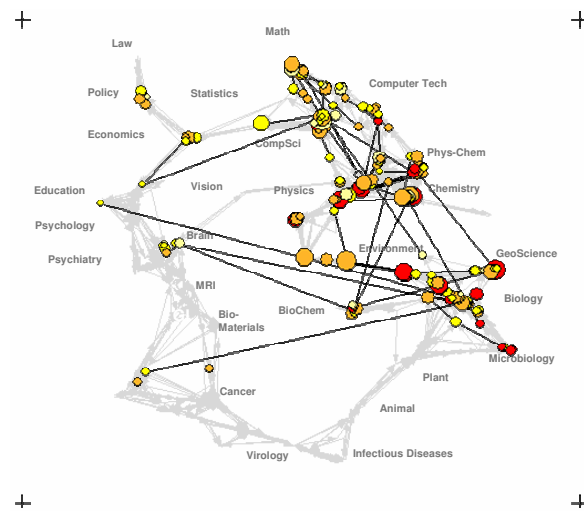


13

Science map applications: Identifying core competency

Kevin W. Boyack & Richard Klavans, unpublished work.

Funding Patterns of the National Science Foundation (NSF)

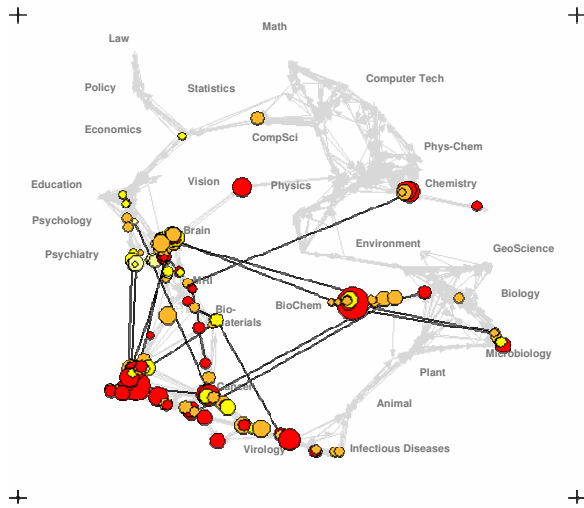


14

Science map applications: Identifying core competency

Kevin W. Boyack & Richard Klavans, unpublished work.

Funding Patterns of the National Institutes of Health (NIH)



15

The Structure of Science

3 The Social Sciences are the smallest and most diffuse of all the sciences. Psychology serves as the link between Medical Sciences (Physiology and the Social Sciences, Statistics serves as the link with Computer Science and Mathematics.

1 Mathematics is our starting point, the purest of all sciences. It lies at the outer edge of the map. Civilian Science, general Engineering, and Space are applied sciences that draw upon knowledge in Mathematics and Physics. These three disciplines provide a good example of a linear progression from one pure science (Mathematics) to another (Physics) through multiple disciplines. Although applied, these disciplines are highly concentrated with distinct bands of research communities that link them. Bands indicate interdisciplinary research.

2 Research is highly concentrated in Physics and Chemistry. These disciplines have few, but very distinct, bands of research communities that link them. The thickness of these bands indicates an extensive amount of interdisciplinary research, which suggests that the transitions between Physics and Chemistry are not as distinct as one might assume.

5 The Life Sciences, including Biology and Biochemistry, are well concentrated from Chemistry of Physics. Bands of linking research can be seen between the larger areas in the Life Sciences, for instance between Biology and Microbiology, and between Biology and Environmental Science. Biochemistry is very interesting in that it is a large discipline that has links into disciplines in many areas of the map, including Biology, Chemistry, Neuroscience, and General Medicine. It is perhaps the most interdisciplinary of the sciences.

4 The Medical Sciences include broad therapeutic studies and targeted areas of treatment (e.g. central nervous system, cardiology, gastroenterology, etc.). Unlike Physics and Chemistry, the medical disciplines are more spread out, suggesting a more multidisciplinary approach to research. The transition into Life Sciences (via Animal Science and Biochemistry) is gradual.

We are all familiar with traditional maps that show the relationships between countries, provinces, states, and cities. Similar relationships exist between the various disciplines and research topics in science. This allows us to map the structure of science.

One of the first maps of science was developed at the Institute for Scientific Information over 40 years ago. It identified 41 areas of science from the citation patterns in 17,000 scientific papers. That early map was intriguing, but it didn't cover enough of science to accurately define its structure.

Things are different today. We have enormous computing power and advanced visualization software that make mapping of the structure of science possible. This galaxy map of science (left) was generated at Sandia National Laboratories using an advanced graph layout routine (VxOrbit) from the citation patterns in 600,000 scientific papers published in 2002. Each dot in the galaxy represents one of the 40,000 research communities active in science in 2002. A research community is a group of papers (on average) that are written on the same research topic in a given year. Over time, communities can be born, continue, split, merge, or die.

The map of science can be used as a tool for science strategy. This is the terrain in which organizations and institutions locate their scientific capabilities. Additional information about the scientific and economic impact of each research community allows policy makers to decide which areas to explore, exploit, abandon, or ignore.

We also envision the map as an educational tool. For children, the theoretical relationship between areas of science can be replaced with a concrete map showing how math, physics, chemistry, biology and social studies intersect. For advanced students, areas of interest can be located and neighboring areas can be explored.

Nanotechnology

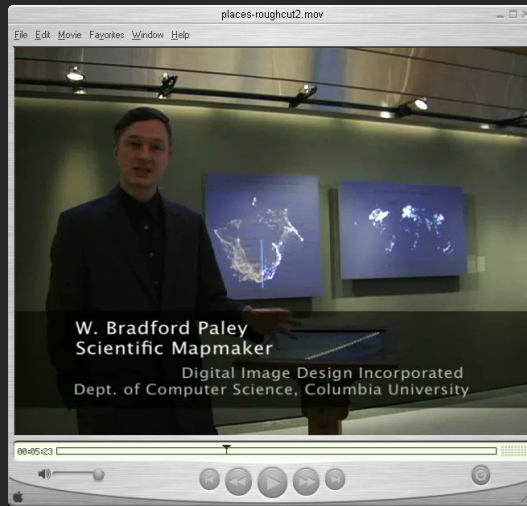
Most research communities in nanotechnology are concentrated in Physics, Chemistry, and Materials Science. However, many disciplines in the Life and Medical Sciences also have nanotechnology applications.

Proteomics

Research communities in proteomics are centered in Biochemistry. In addition, there is a heavy focus in the tools section of chemistry, such as Chromatography. The balance of the proteomics communities are widely dispersed among the Life and Medical Sciences.

Pharmacogenomics

Pharmacogenomics is a relatively new field with most of its activity in Medicine. It also has many communities in Biochemistry and two communities in the Social Sciences.



"Places & Spaces: Mapping Science" Illuminated Diagram Display

Was on display at the NYPL Science, Industry, and Business Library
Madison/34th, New York City, April 3rd - August 31st, 2006.

TOPIC MAP: HOW SCIENTIFIC PARADIGMS RELATE

GEOGRAPHIC MAP: WHERE SCIENCE GETS DONE

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.

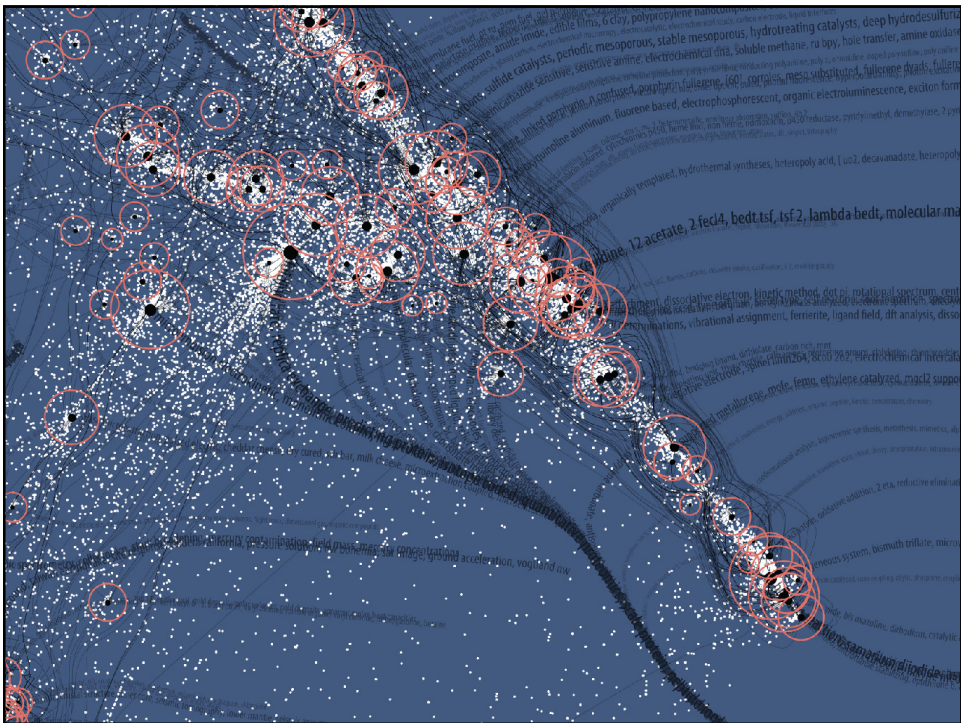
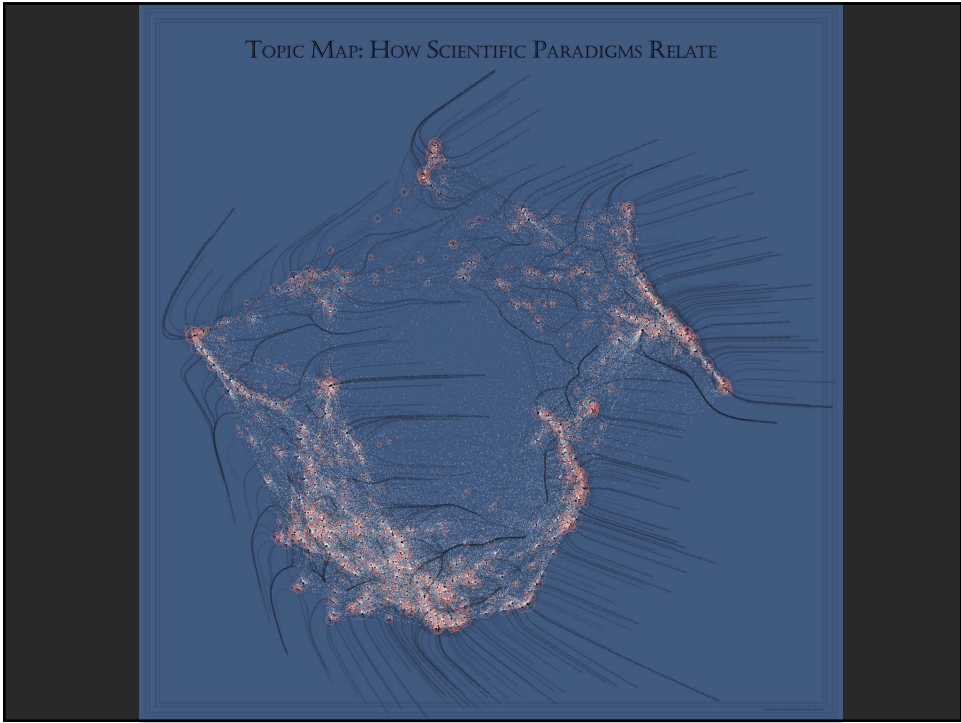
Nanotechnology

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.

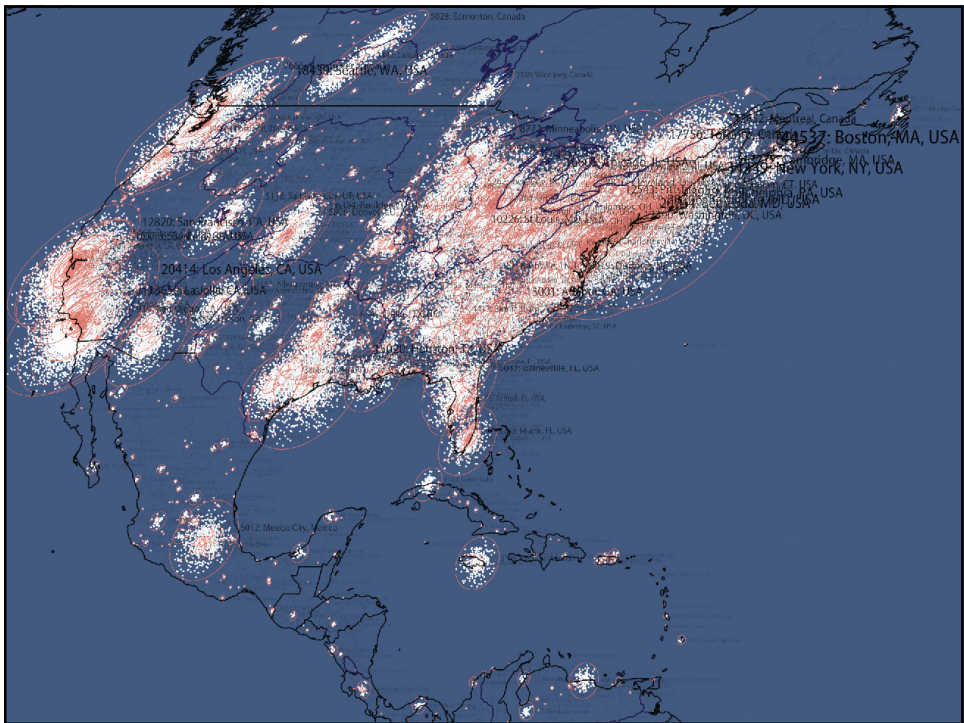
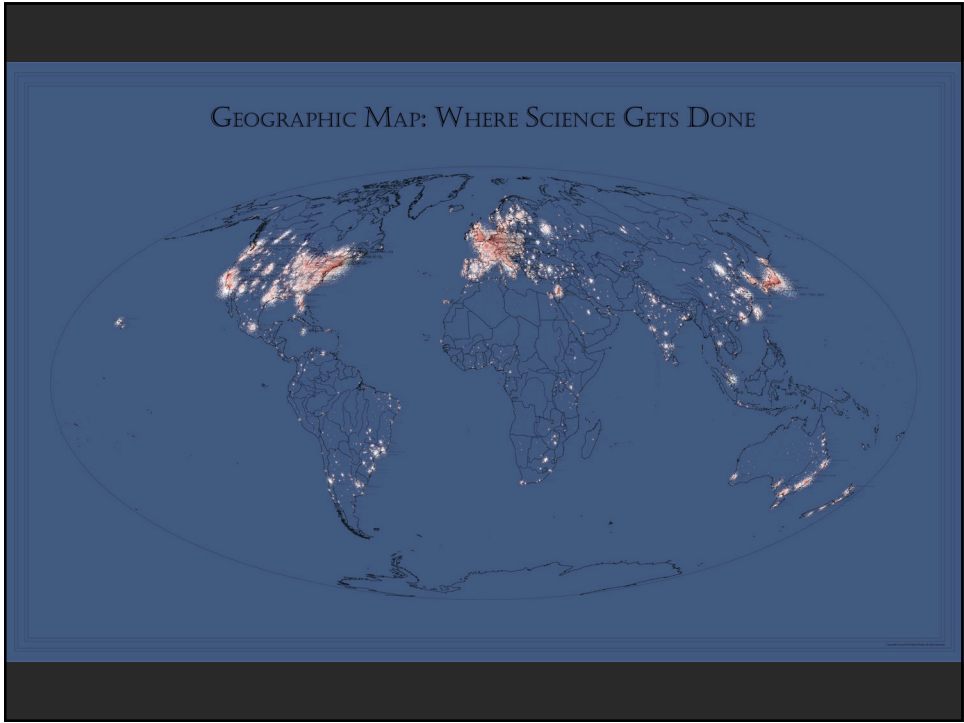
All Topics <i>Sweep through all 776 scientific paradigms</i>	Nanotechnology <i>Science on the tiny scale of molecules</i>	Francis H. C. CRICK <i>Co-discovered DNA's double helix</i>	Albert EINSTEIN <i>Revitalized physics with Relativity theories</i>	Michael E. FISHER <i>Models critical phase transitions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

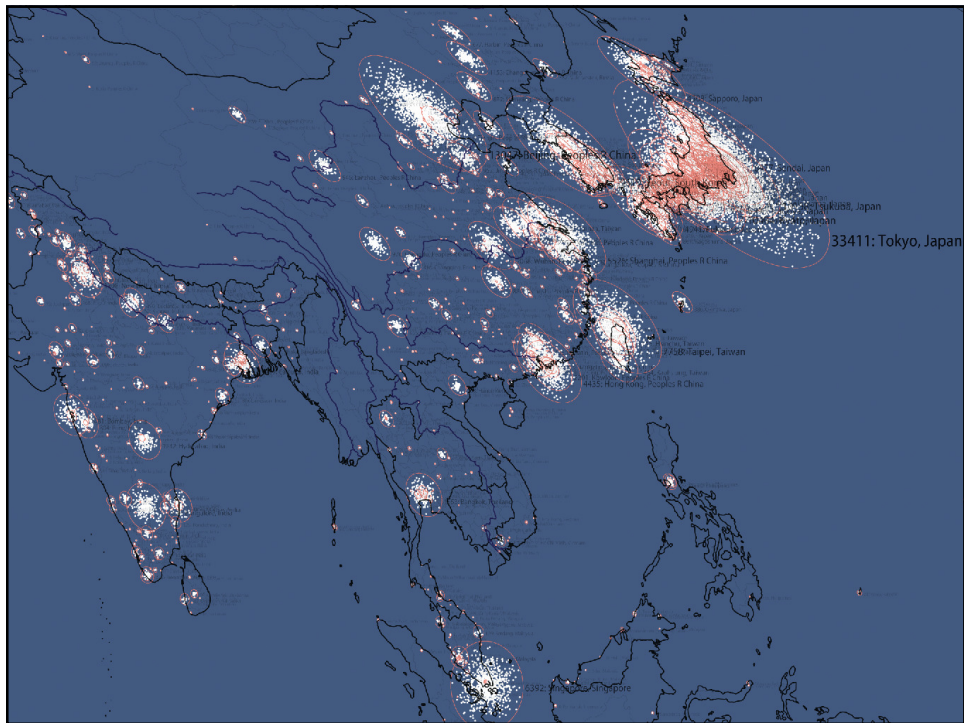
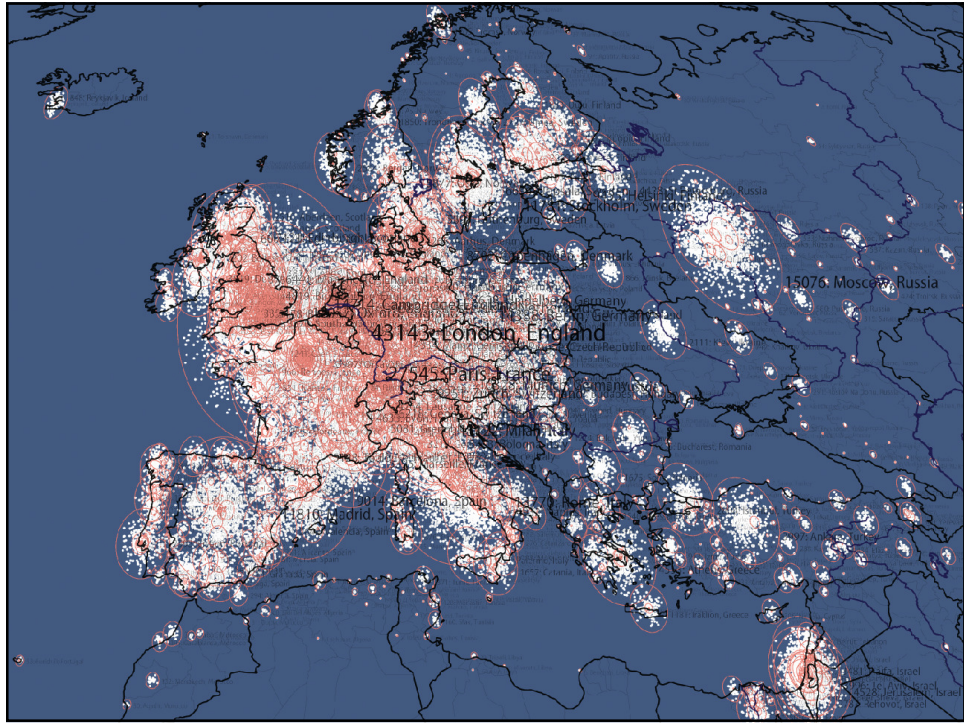
We sweep slowly through adjoining related topics, lighting up the places in the world that study each topic. You may select a subset of the topics that deal with these three interesting subjects by touching it.

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.



GEOGRAPHIC MAP: WHERE SCIENCE GETS DONE







Science Puzzle Map for Kids by Fileve Palmer, Julie Smith, Elisha Hardy and Katy Börner, Indiana University, 2006. (Base map taken from Illuminated Diagram display by Kevin Boyack, Richard Klavans, and W. Bradford Paley.)



For a New Take on Maps Chart Your Way to the New York Hall of Science
Places & Spaces: Mapping Science on view December 9, 2006 – February 25, 2007

Queens, N. Y. – Want to see science from above? Curious to see what impact one single person or invention can have? Keen to find pockets of innovation? Desperate for better tools to manage the flood of information? Or are you simply fascinated by maps? Then visit the *Places & Spaces: Mapping Science* exhibition, which aims to demonstrate the power of maps to navigate and make sense of physical places and abstract topic spaces.

Building Data/Software/Resources Bazaars not Cathedrals

Information Visualization CyberInfrastructure

The InfoVis CyberInfrastructure provides access to data, software code and learning modules as well as computing resources in support of the analysis, modeling and visualization of diverse data sets.

DATABASES

An Open-Database provides access to publications, posters, grants and grant opportunities. The database is continuously and automatically updated.

COMPUTING RESOURCES

The network is distributed across a local and remote University Research Database Complex, comprising of two Open T300 servers with 12 nodes, processors and 48 GB of memory each. A 75-hour shared disk are available to both servers. A team work system with 100 nodes and 100 memory nodes in the work force and 100 for database servers.

SOFTWARE

An open source IVC framework was designed to facilitate the integration of diverse data analysis, modeling and visualization algorithms. New algorithms, data persistence methods, look and feel for the interface and more tools can be easily integrated in the framework.

LEARNING MODULES

A set of associated learning modules aims to equip learners with a practical skill set for providing public and advice to specify, modify and use different algorithms and diverse research techniques and design features, used to quickly generate and compare information visualizations.



Cyberinfrastructure Shell
<http://cisshell.org>

CAREER: Visualizing Knowledge Domains. NSF IIS-0238261 award (Katy Börner, \$451,000) Sept. 03-Aug. 08.
<http://w.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-László Barabási, Santiago Schnell, Alessandro Vespignani & Stanley Wasserman, Eric Wernert (Senior Personnel), \$1,120,926) Sept. 05 - Aug. 08. <http://nwb.slis.indiana.edu>

