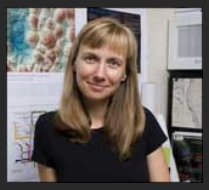
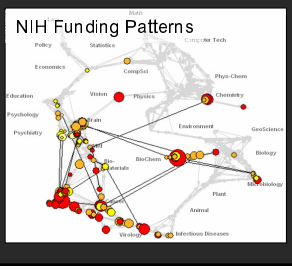
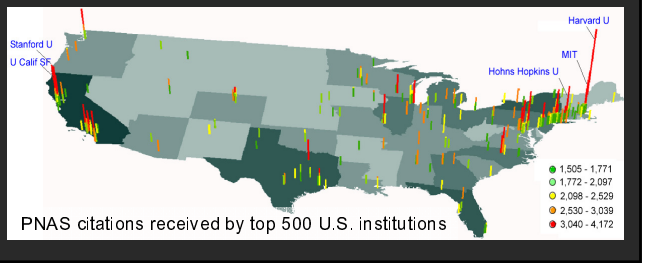
Mapping Science: Opportunities and Challenges



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

Talk at the Cambridge Colloquium on Complexity and Social Networks
Harvard University, Boston, MA.
September 21, 2006





This Talk has Three Parts:

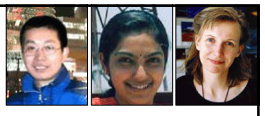
1. Mapping Science: Why?
2. Mapping Science: How?

3. Mapping Science: Applications

This Talk has Three Parts:

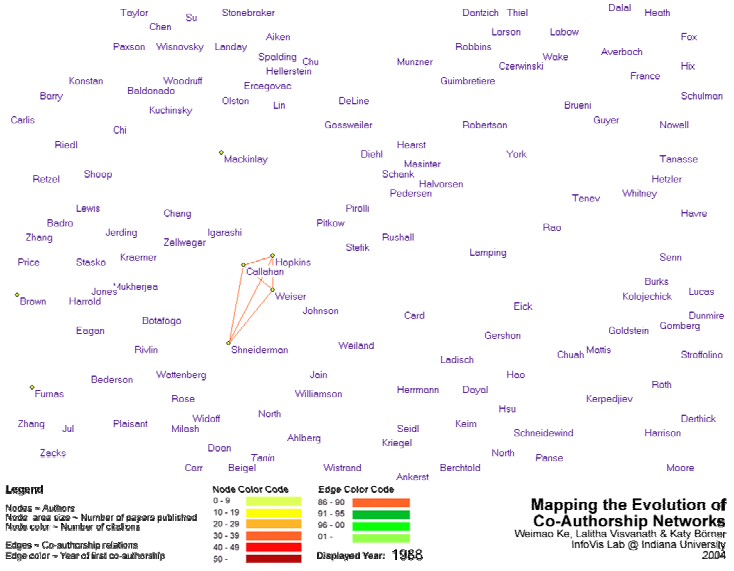
1. **Mapping Science: Why?**
2. Mapping Science: How?

3. Mapping Science: Applications



Mapping the Evolution of Co-Authorship Networks in Information Visualization, 1988 - 2004

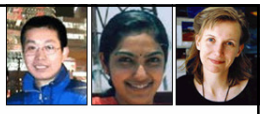
Ke, Visvanath & Börner, (2004) Won 1st price at the IEEE InfoVis Contest.

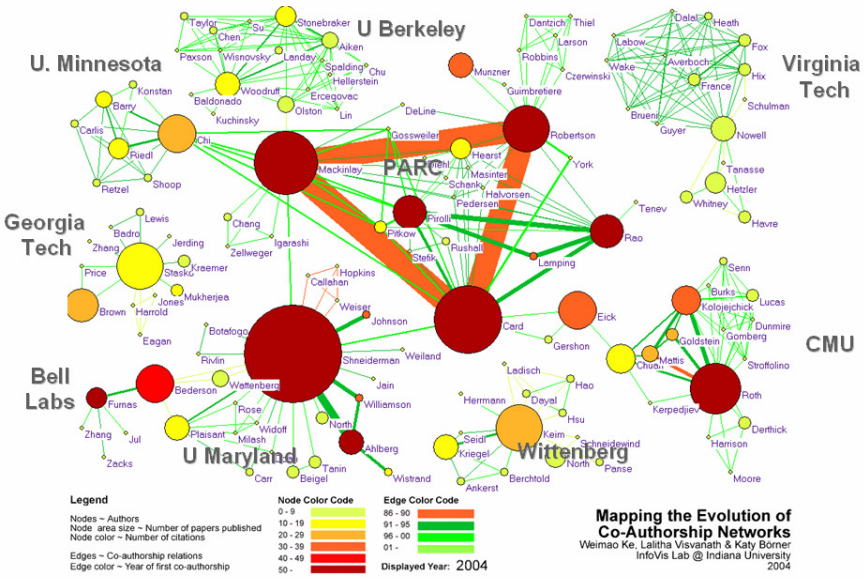


4

Mapping the Evolution of Co-Authorship Networks

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





5

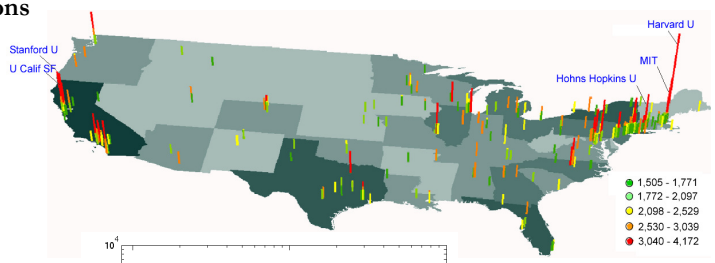
Spatio-Temporal Information Production and Consumption of Major U.S. Research Institutions

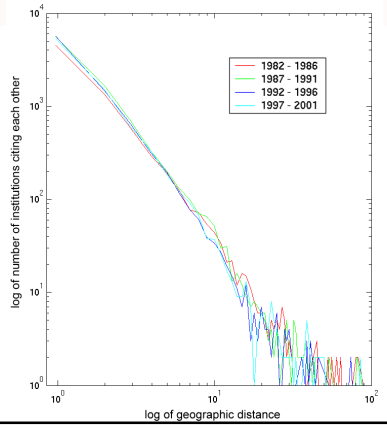
Börner & Penumarthy. (2005)

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.

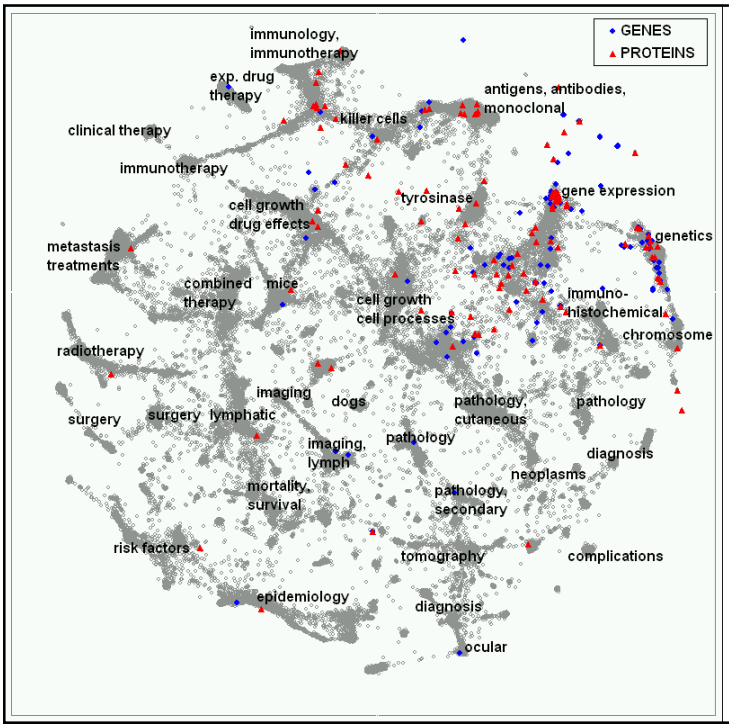
Each institution is assumed to produce and consume information.





$\gamma_{82\text{-}86} = 1.94$ (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%)

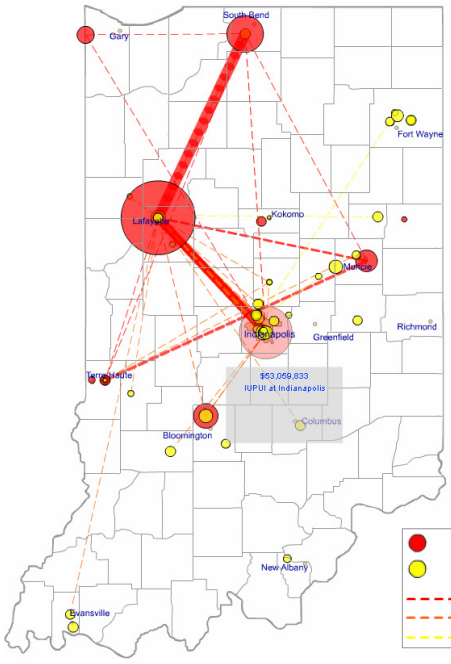
6



Mapping Medline Papers, Genes, and Proteins Related to Melanoma Research

Boyack, Mane & Börner. (2004) IV Conference, pp. 965-971.

7



Mapping Indiana's Intellectual Space

(Ke, Börner & Mei, 2005)

Identify

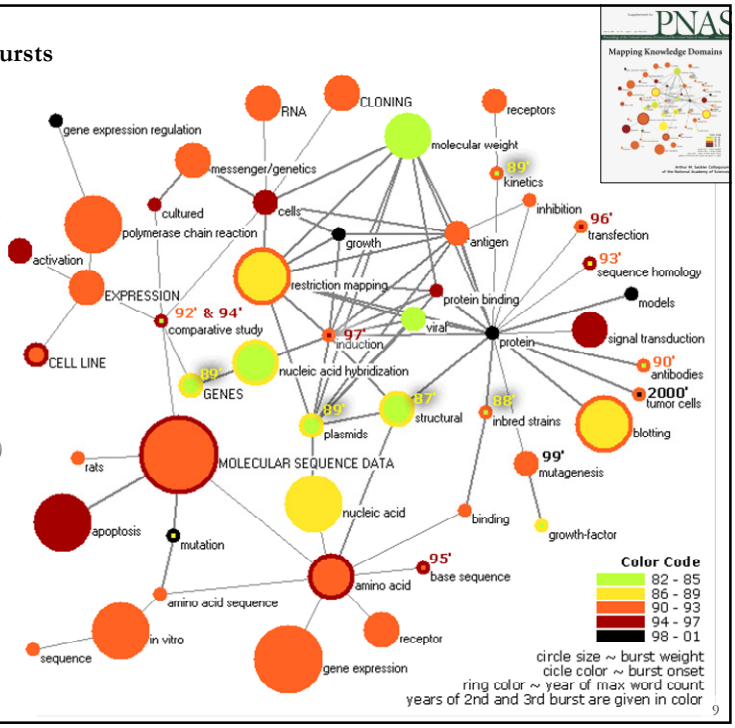
- Pockets of innovation
- Pathways from ideas to products
- Interplay of industry and academia

8

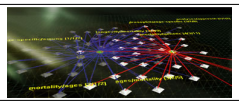
Mapping Topic Bursts

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.



9



Mapping Science: Opportunities

Advantages for Funding Agencies

- Supports monitoring of (long-term) money flow and research developments, evaluation of funding strategies for different programs, decisions on project durations, funding patterns.
- Staff resources can be used for scientific program development, to identify areas for future development, and the stimulation of new research areas.

Advantages for Researchers

- Easy access to research results, relevant funding programs and their success rates, potential collaborators, competitors, related projects/publications **(research push).**
- More time for research and teaching.

Advantages for Industry

- Fast and easy access to major results, experts, etc.
- Can influence the direction of research by entering information on needed technologies **(industry-pull)**.

Advantages for Publishers

- Unique interface to their data.
- Publicly funded development of databases and their interlinkage.

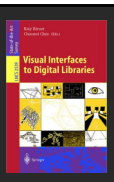
For Society

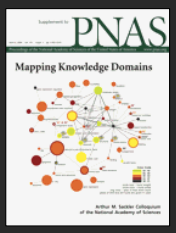
- Dramatically improved access to scientific knowledge and expertise.

10

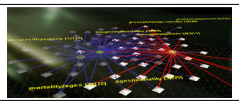
This Talk has Three Parts:

1. Mapping Science: Why?
2. **Mapping Science: How?**
3. Mapping Science: Applications





- *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, 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.*
- *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.*



Work Closely & Listen Carefully to Your 'Clients'

What information needs do

- Funding Agencies
- Researchers
- Industry
- Publishers
- Society

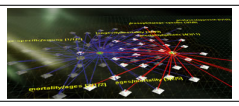
truly have? What would they pay for/use on an every day basis?

In what priority, sequence should information be presented?

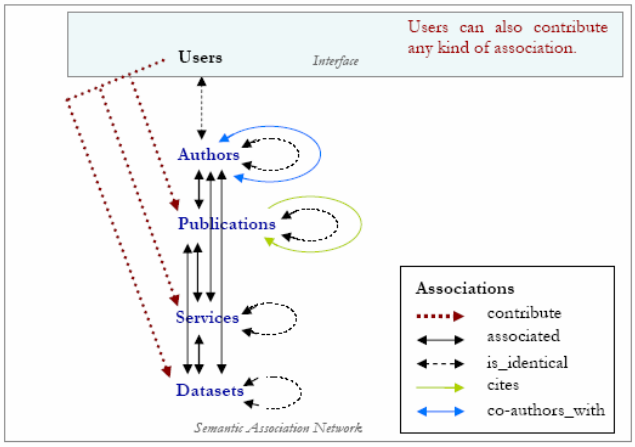
What metaphors, format will work best?

When do global views help or hurt?

12

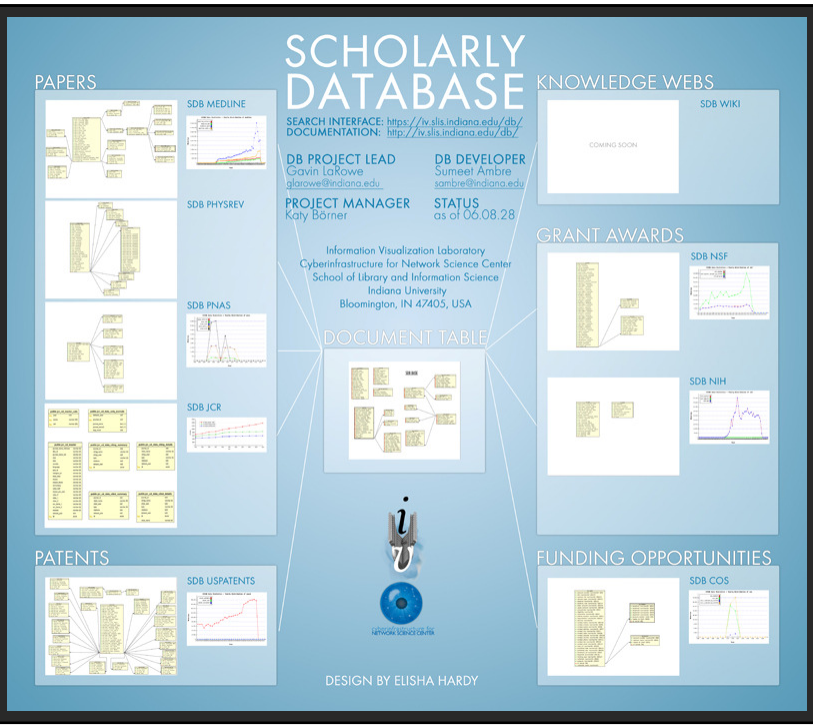


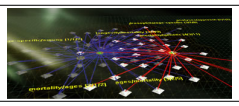
Take High Quality, High Coverage Data
Interlinked it Using Semantic Association Networks



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.

13





Apply the Best Algorithms in Appropriate Sequence

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.

BROADENING
- By citation
- By terms

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

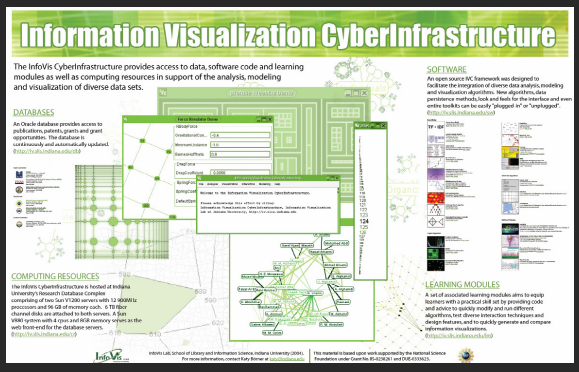
SCALAR
- Triangulation
- Force-directed placement (FDP)

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.

15



Build and Utilize Cyberinfrastructures

Mankind's Scholarly Knowledge Cannot be Processed on One Laptop

CAREER: Visualizing Knowledge Domains. NSF IIS-0238261 award (Katy Börner, $451,000) Sept. 03-Aug. 08.
http://iv.slis.indiana.edu/

NetworkWorkbench
A Workbench for Network Scientists

HPNAP High Performance Network Applications Program

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, Eric Wernert (Senior Personnel), $1,120,926) Sept. 05 - Aug. 08. https://nwb.slis.indiana.edu

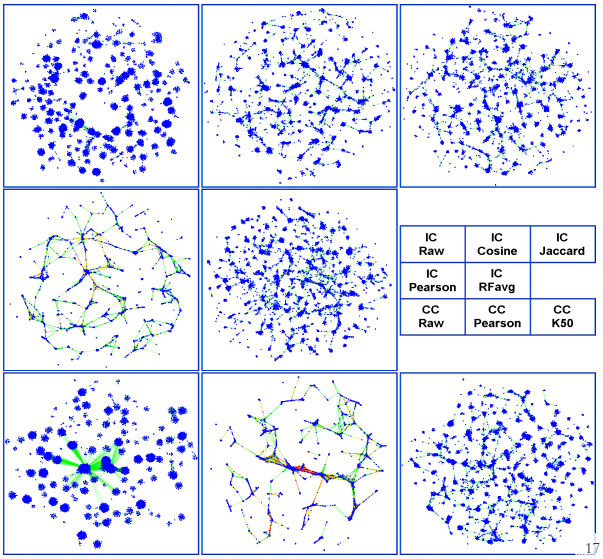
16

Validate Results

Example: 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.

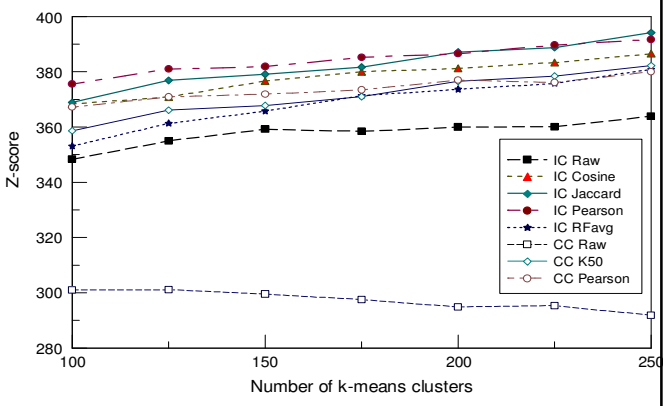


17

Example: Comparison of Similarity Metrics

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.

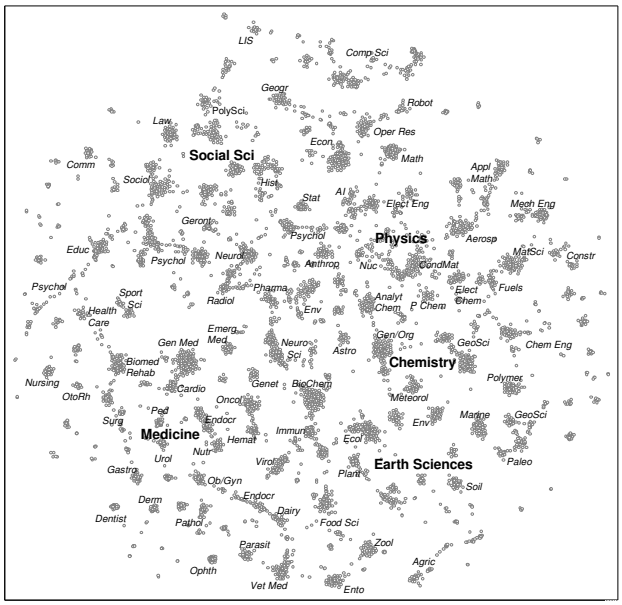


Boyack, Kevin W., Klavans, R. and Börner, Katy. (2005). Mapping the Backbone of Science. Scientometrics. 64(3), 351-374.

18

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.



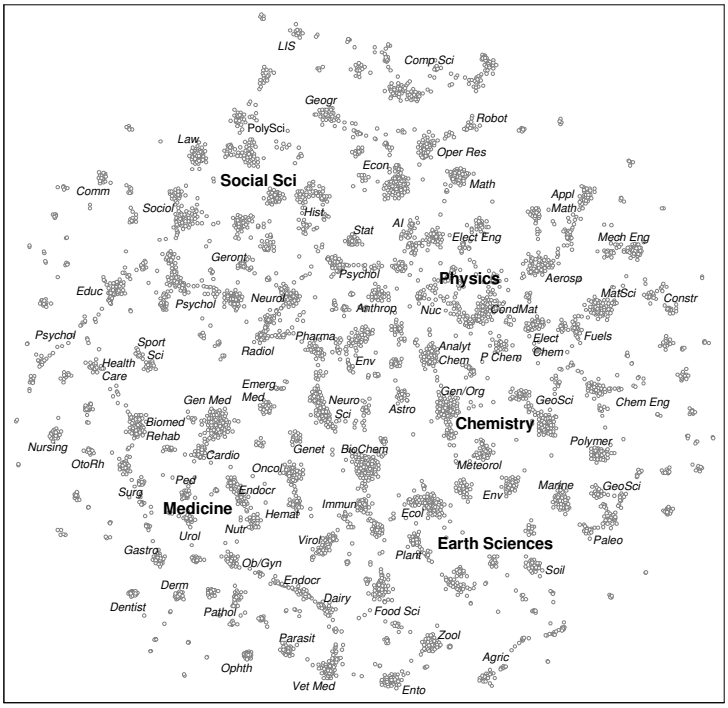
19

This Talk has Three Parts:

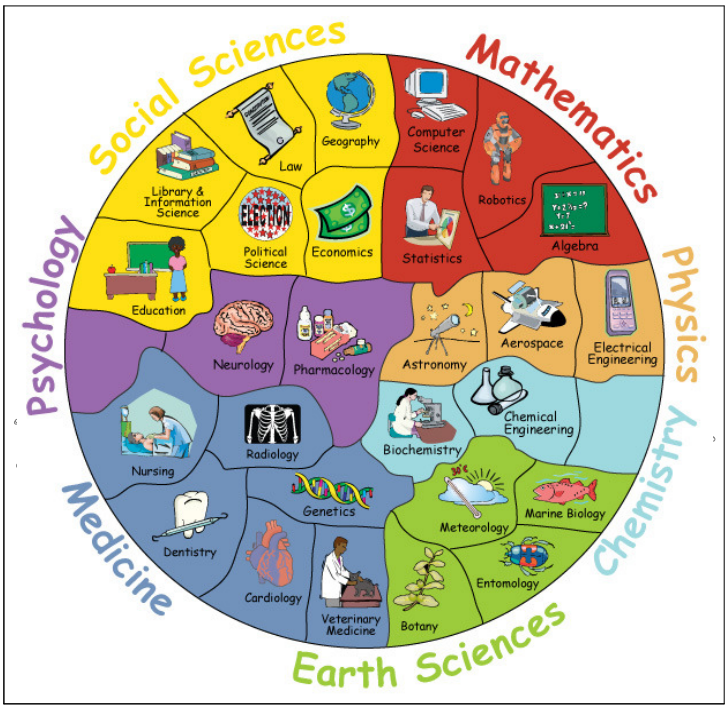
1. Mapping Science: Why?
2. Mapping Science: How?

3. **Mapping Science: Applications**

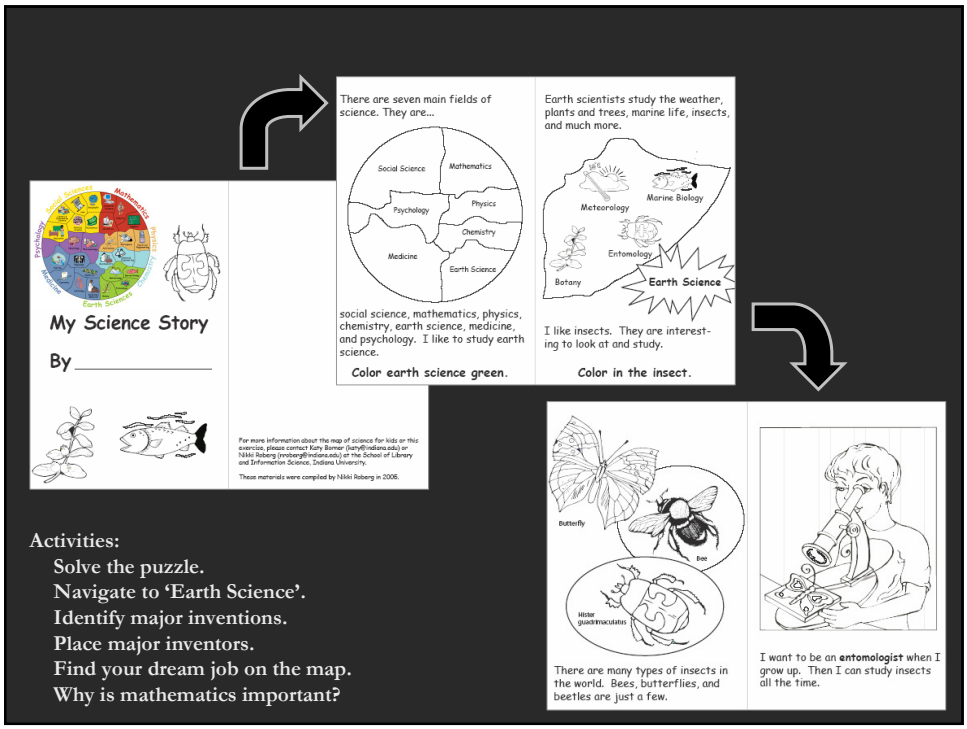
20


LIS
Comp Sci
Geogr
Robot
PolySci
Law
Oper Res
Econ
Social Sci
Math
Comm
Appl Math
Sociol
Hist
Stat
AI
Elect Eng
Mech Eng
Geront
Psychol
Physics
Aerosp
Educ
Psychol
Neurol
MatSci
Constr
Anthrop
Nuc
CondMat
Psychol
Sport Sci
Pharma
Fuels
Radiol
Analyt Chem
Elect Chem
Health Care
Env
P Chem
Emerg Med
Gen/Org
GeoSci
Chem Eng
Gen Med
Neuro Sci
Astro
Biomed
Chemistry
Rehab
Polymer
Nursing
Genet
BioChem
Cardio
Meteorol
OtoRh
Oncol
Marine
GeoSci
Ped
Endocr
Env
Surg
Medicine
Immun
Ecol
Hemat
Nutr
Paleo
Urol
Virol
Earth Sciences
Gastro
Plant
Ob/Gyn
Soil
Endocr
Derm
Dairy
Dentist
Pathol
Food Sci
Parasit
Zool
Agric
Ophth
Vet Med
Ento

21


Social Sciences
Mathematics
Physics
Chemistry
Earth Sciences
Medicine
Psychology
Geography
Law
Computer Science
Robotics
Library & Information Science
Political Science
Economics
Statistics
Algebra
Education
Aerospace
Electrical Engineering
Neurology
Pharmacology
Astronomy
Chemical Engineering
Biochemistry
Radiology
Nursing
Genetics
Meteorology
Marine Biology
Dentistry
Entomology
Cardiology
Veterinary Medicine
Botany

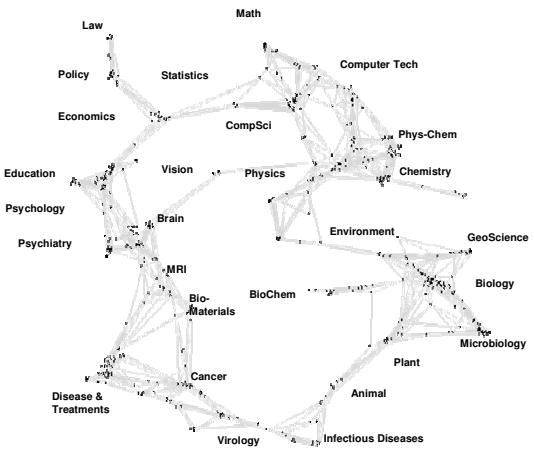
22


My Science Story
By _____
There are seven main fields of science. They are...
Social Science
Mathematics
Psychology
Physics
Chemistry
Medicine
Earth Science
social science, mathematics, physics, chemistry, earth science, medicine, and psychology. I like to study earth science.
Color earth science green.
Earth scientists study the weather, plants and trees, marine life, insects, and much more.
Meteorology
Marine Biology
Entomology
Botany
Earth Science
I like insects. They are interesting to look at and study.
Color in the insect.
Butterfly
Bee
Hister quadrimaculatus
There are many types of insects in the world. Bees, butterflies, and beetles are just a few.
I want to be an entomologist when I grow up. Then I can study insects all the time.
Activities:
Solve the puzzle.
Navigate to 'Earth Science'.
Identify major inventions.
Place major inventors.
Find your dream job on the map.
Why is mathematics important?

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

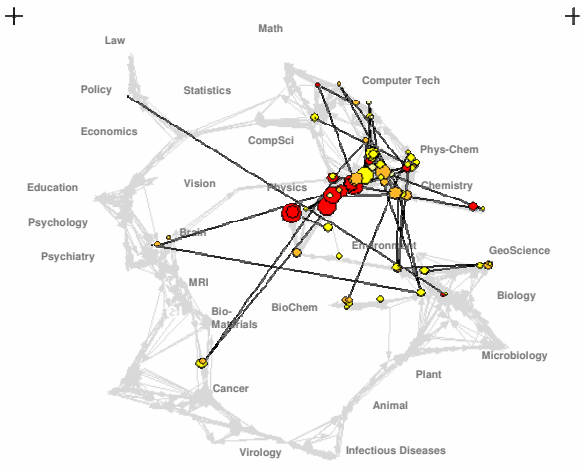


25

Science map applications: Identifying core competency

Kevin W. Boyack & Richard Klavans, unpublished work.

Funding patterns of the US Department of Energy (DOE)

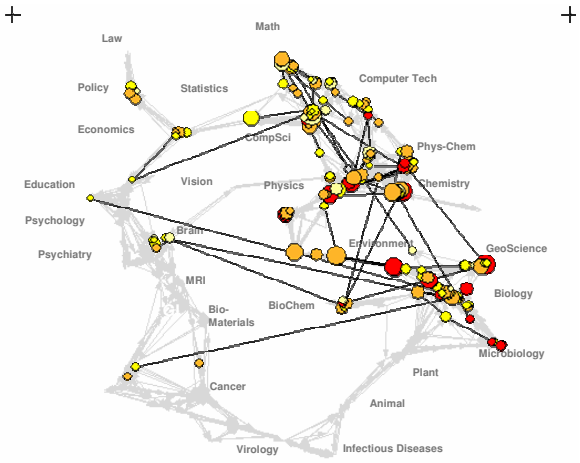


26

Science map applications: Identifying core competency

Kevin W. Boyack & Richard Klavans, unpublished work.

Funding Patterns of the National Science Foundation (NSF)

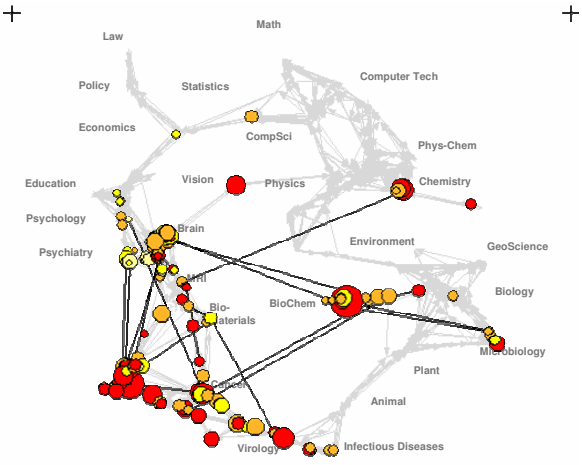


27

Science map applications: Identifying core competency

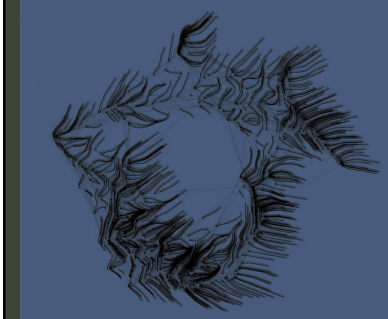
Kevin W. Boyack & Richard Klavans, unpublished work.

Funding Patterns of the National Institutes of Health (NIH)

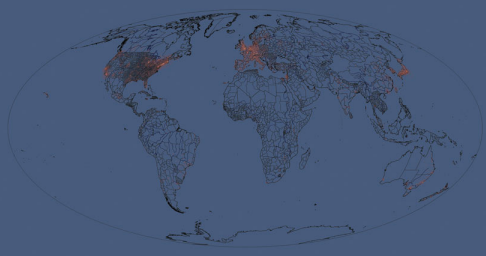


28

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 *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

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.

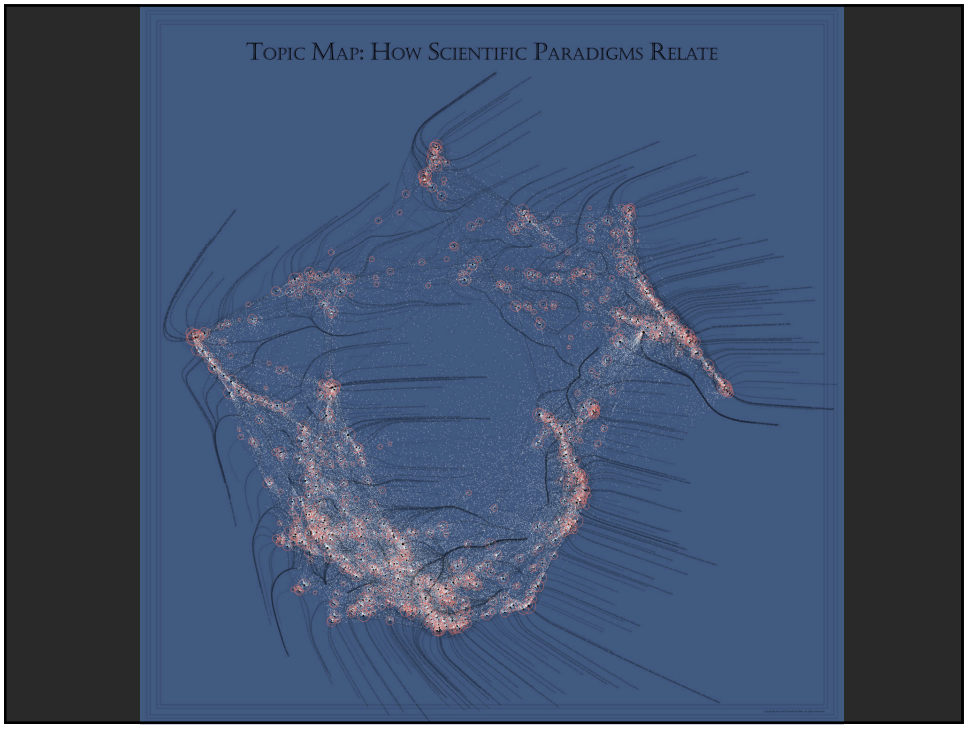
Nanotechnology

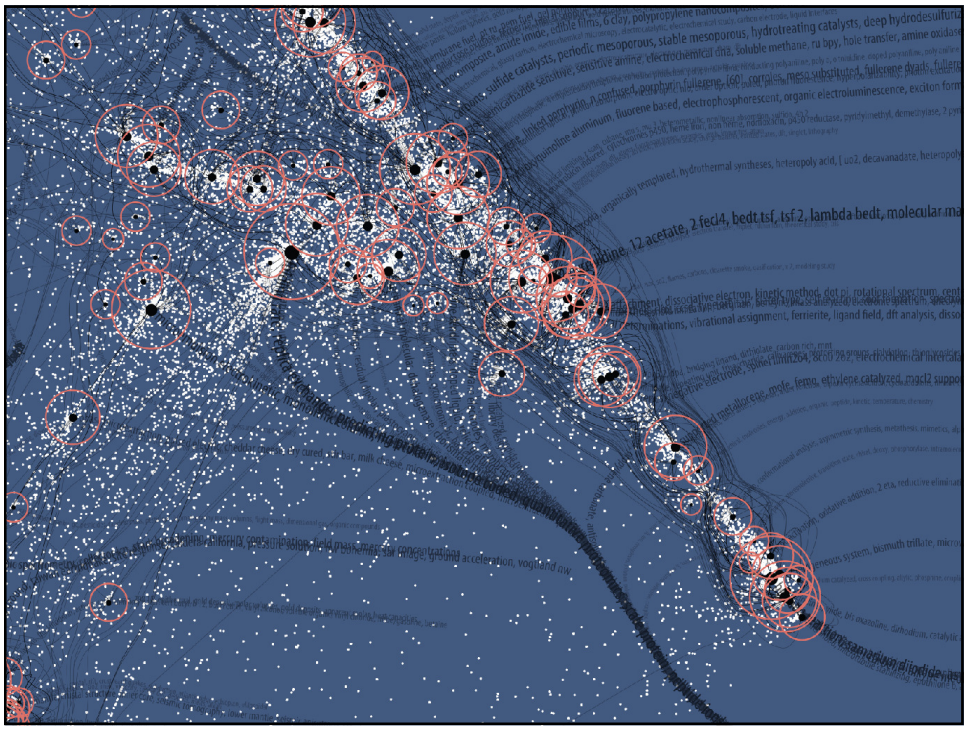
This overlay shows the distribution of nanotechnology within the paradigms of science. The majority of current work in nanotechnology takes places in physics, chemistry, and materials science, at the upper right portion of the map. However, an increasing amount of nanotechnology is being applied in the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

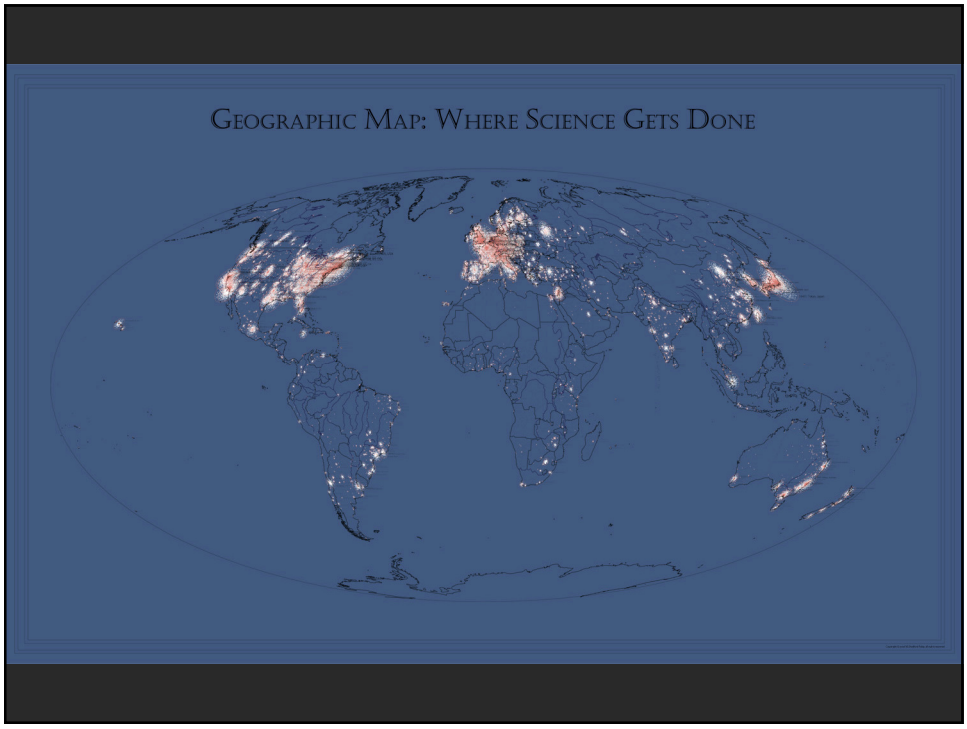
All Topics *Sweep through all 776 scientific paradigms*	**Nanotechnology** *Science on the tiny scale of molecules*	**Francis H. C. CRICK** *Co-discovered DNA's double helix*	**Albert EINSTEIN** *Revitalized physics with Relativity theories*	**Michael E. FISHER** *Models critical phase transitions of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

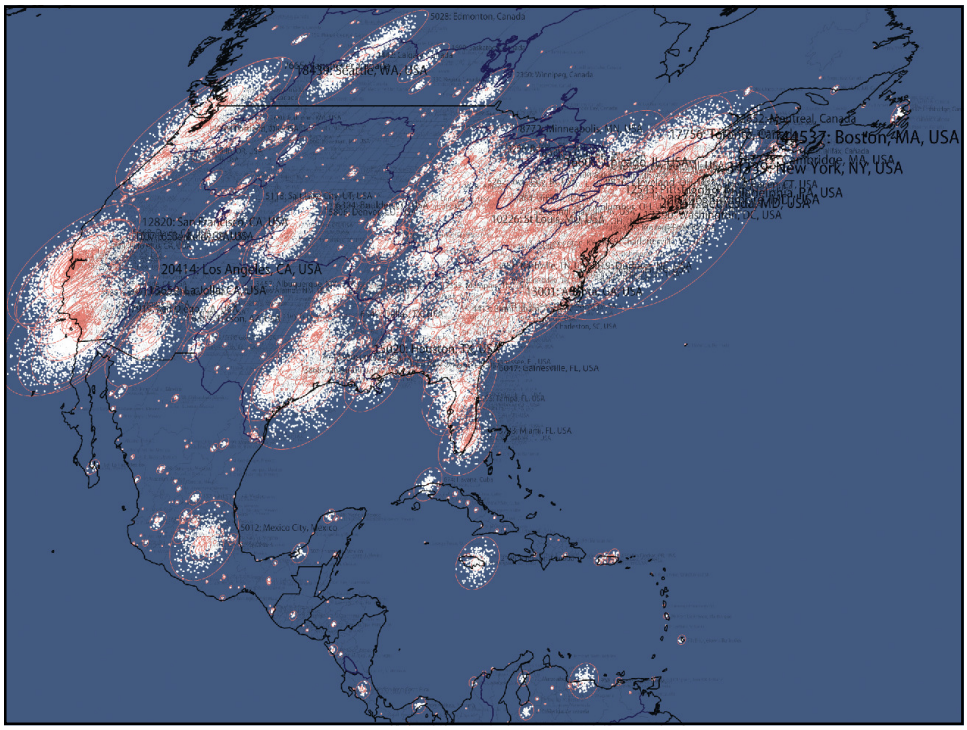
We sweep slowly through adjoining related topics, lighting up the places in the world that study each topic. You may select a subset of the topics that deal with these three interesting subjects by touching it.

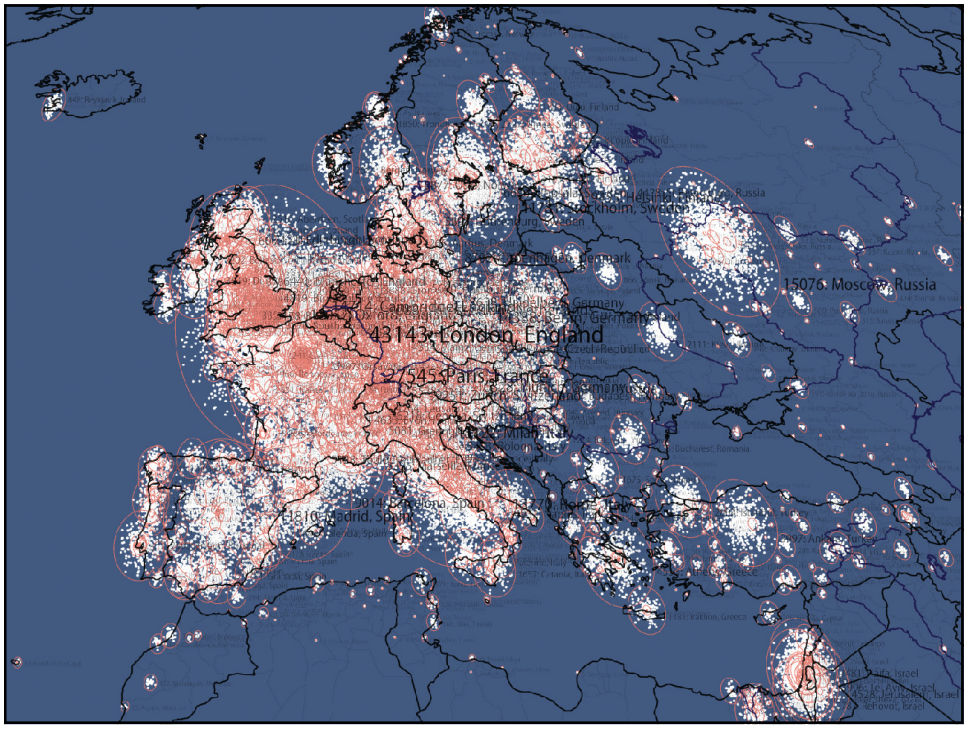
A single person's spreading influence is shown as a series of four snapshots. First, we light only topics and places relating to that person's papers—papers that are still highly cited today. The second lights everything that cites that original work. Note that this first-generation impact extends to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

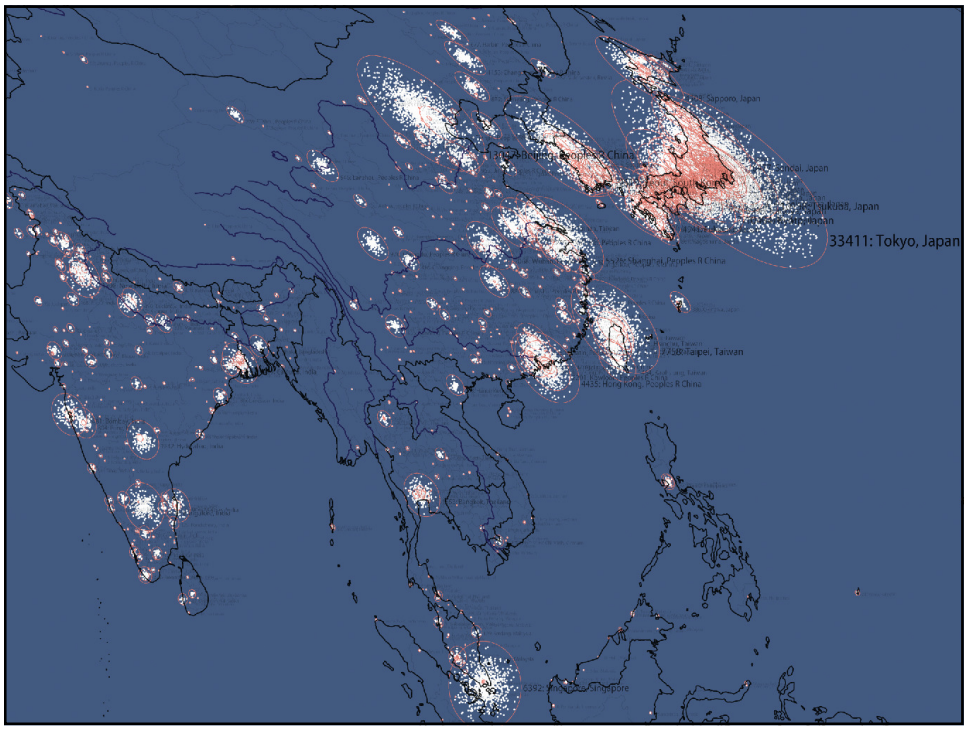

Topic Map: How Scientific Paradigms Relate




Geographic Map: Where Science Gets Done


44537: Boston, MA, USA
New York, NY, USA
20414-Los Angeles, CA, USA

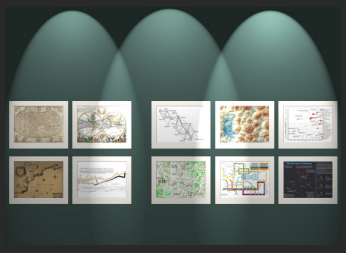

15076: Moscow, Russia
43143: London, England
27545: Paris, France


33411: Tokyo, Japan

"Places & Spaces: Mapping Science"
on display at the NYPL Science, Industry, and Business Library
Madison/34th, New York City
April 3rd - August 31st, 2006.

The Power of Maps

Four Early Maps of Our World
VERSUS
Six Early Maps of Science



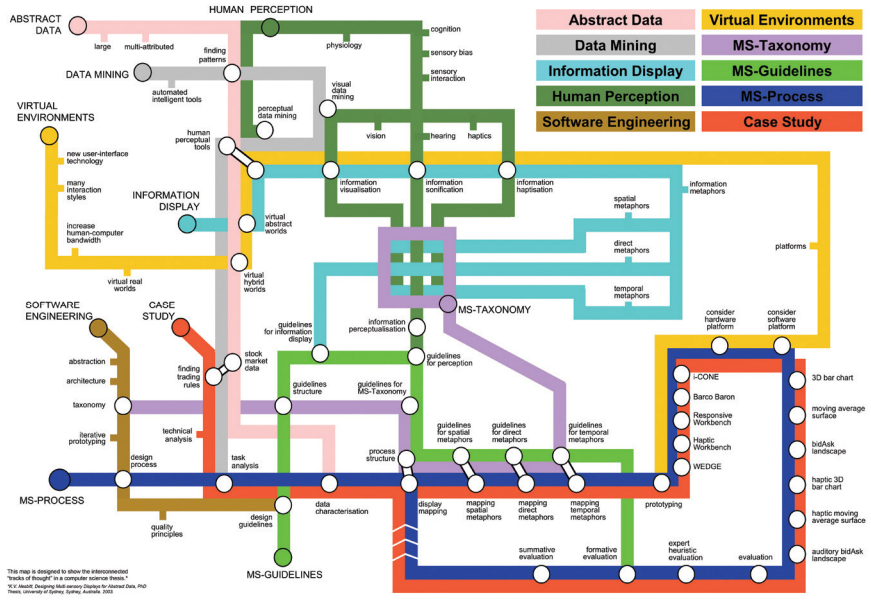
(1st Iteration of Places & Spaces Exhibit - 2005)

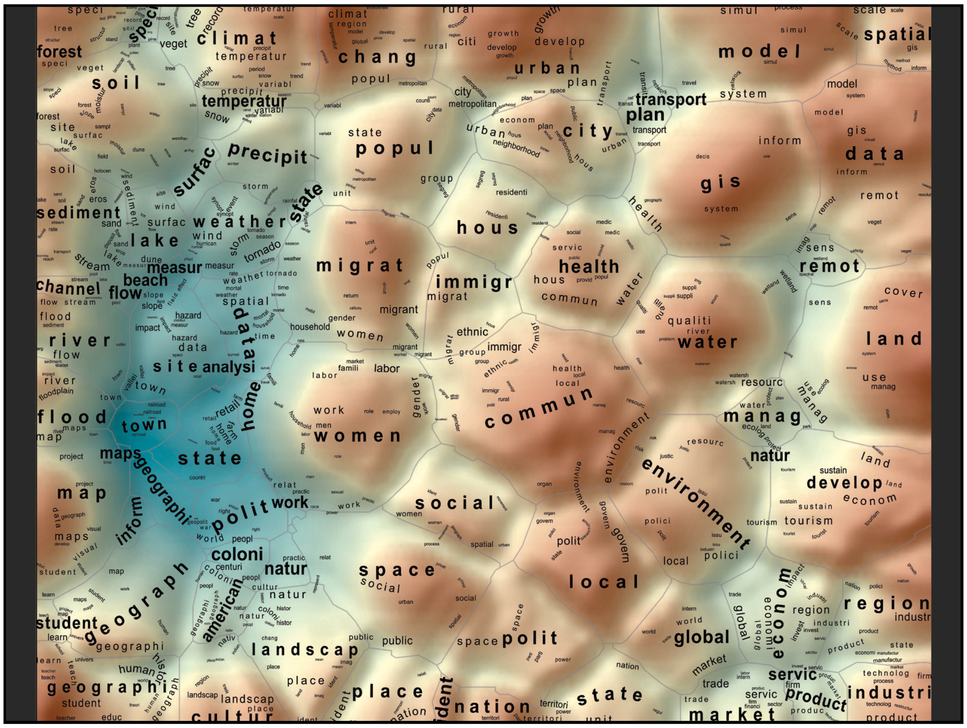

CAVRVS CHORVS VEL IAPIX SIVE ARGESTES
CIRCIVS VEL TRESIAS
SEPTENTRIO VEL APARCTIAS
AQVILO VEL BOREAS
CECIAS APELIOTES
EVROPA
ASIA
CARIA
LIBS INTERIOR
AFFRICA
ARABIA FELIX
MARE INDICVM
ETHIOPIA INTERIOR
MARE
PROPONTIS
AFRICVS VEL LIBS
LIBONOTVS EVROAVSTER
AVSTER VEL NOTVS
EVRONOTVS
ZEPHYRVS VEL ARGESTES


A New Map of the WHOLE WORLD with the Trade winds
According to ye latest and most Exact Obs- ervations By H. Moll Geographer
In this Maps is inserted A View of ye General & Coasting Trade-Winds, Monsoons or ye Shifting Trade-winds Note that ye Arrows among ye Lines shew ye Course of those General & Coasting Winds: and ye Arrows in ye void Spaces shew ye Course of ye Shifting Trade-winds, and ye Abbreviation Sept. &c. Shew ye Times of ye Year when such Winds Blow.
The Signs of the Zodiack The First 6 are Northern, the other Southern Signs
North Pole
South Pole
NORTH AMERICA
SOUTH AMERICA
AMERICA
EUROPE
ASIA
AFRICA
INDIAN OCEAN
New Holland
Printed for Thos Bowles Print and Map seller next ye Chapter House in St Pauls Church yard, and John Bowles Print and Map Seller at the Black Horse in Cornhill London.

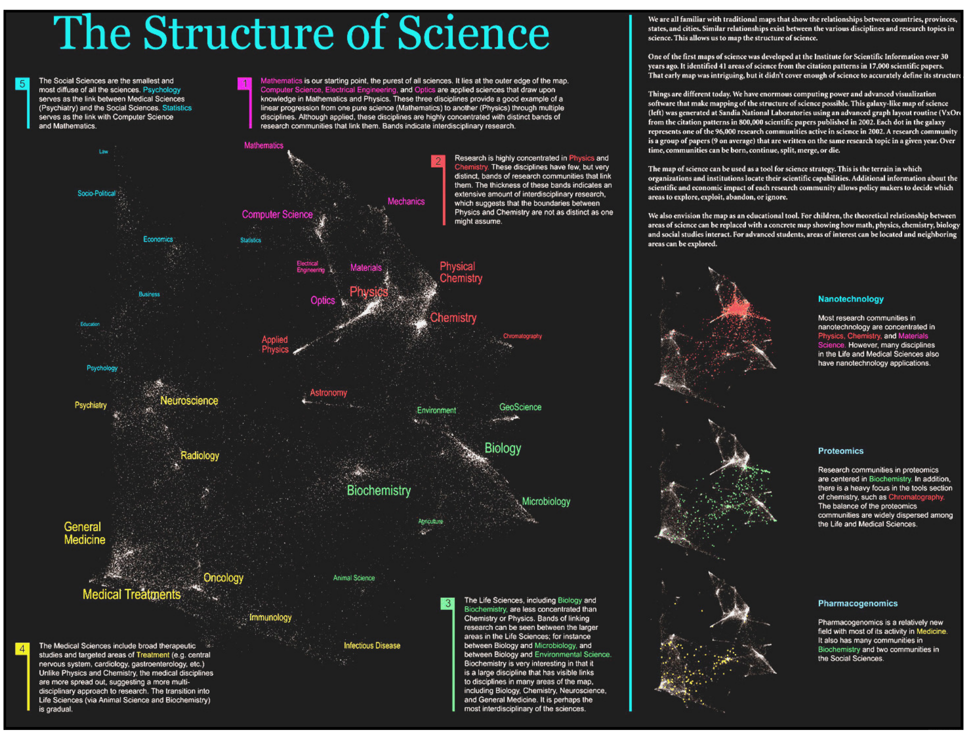
How would a map of science look?

What metaphors would work best?





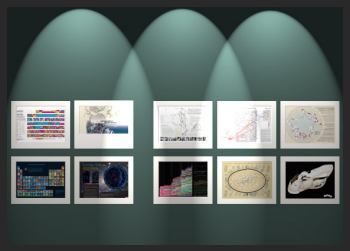
climat
chang
urban
model
spatial
forest
soil
temperatur
precipit
popul
city
plan
transport
data
gis
sediment
weather
lake
surfac
hous
migrat
immigr
health
remot
measur
beach
channel
flow
spatial
river
land
water
site analysi
flood
town
maps
state
home
women
commun
manag
natur
develop
map
environment
geograph
polit
work
social
inform
coloni
natur
space
local
region
student
geograph
landscap
polit
global
servic
industri
geographi
place
nation
state
product
market
cultur



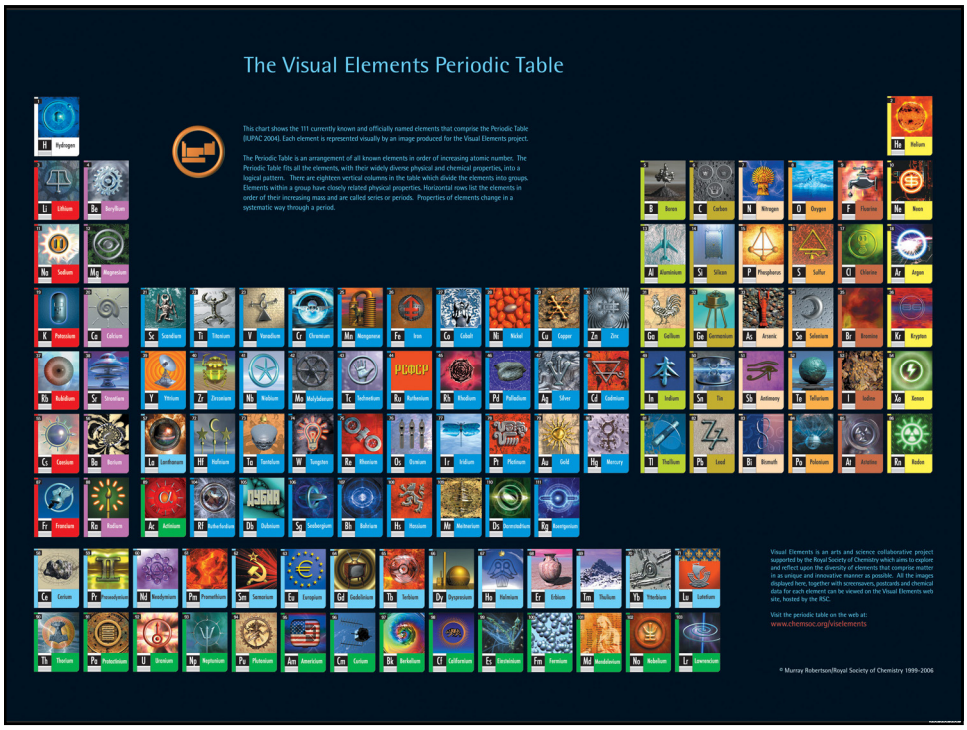
The Structure of Science
5
The Social Sciences are the smallest and most diffuse of all the sciences. Psychology serves as the link between Medical Sciences (Psychiatry) and the Social Sciences. Statistics serves as the link with Computer Science and Mathematics.
1
Mathematics is our starting point, the purest of all sciences. It lies at the outer edge of the map. Computer Science, Electrical Engineering, and Optics 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 boundaries between Physics and Chemistry are not as distinct as one might assume.
Law
Socio-Political
Economics
Business
Mathematics
Computer Science
Statistics
Mechanics
Electrical Engineering
Materials
Physical Chemistry
Optics
Physics
Chemistry
Applied Physics
Chromatography
Psychology
Psychiatry
Neuroscience
Astronomy
Environment
GeoScience
Radiology
Biology
Biochemistry
Microbiology
Agriculture
General Medicine
Oncology
Animal Science
Medical Treatments
Immunology
Infectious Disease
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 multi-disciplinary approach to research. The transition into Life Sciences (via Animal Science and Biochemistry) is gradual.
3
The Life Sciences, including Biology and Biochemistry, are less concentrated than Chemistry or 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 visible links to disciplines in many areas of the map, including Biology, Chemistry, Neuroscience, and General Medicine. It is perhaps the most interdisciplinary of the sciences.
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 30 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-like map of science (left) was generated at Sandia National Laboratories using an advanced graph layout routine (VxOrd from the citation patterns in 800,000 scientific papers published in 2002. Each dot in the galaxy represents one of the 96,000 research communities active in science in 2002. A research community is a group of papers (9 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 interact. 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.

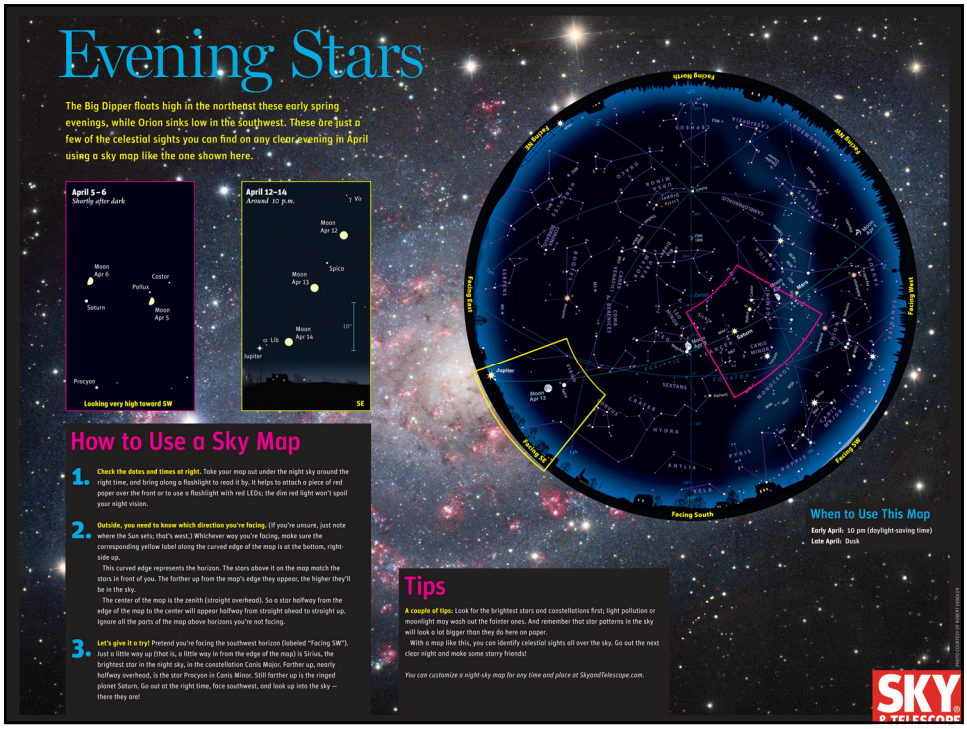
The Power of Reference Systems

Four Existing Reference Systems
VERSUS
Six Potential Reference Systems of Science



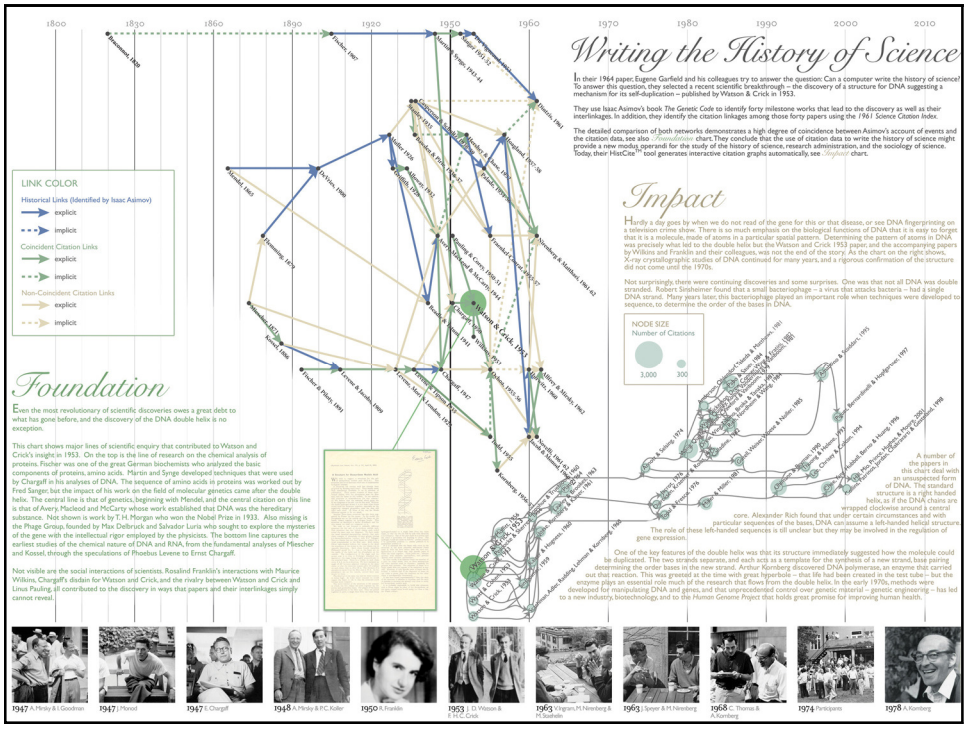
(2nd Iteration of Places & Spaces Exhibit - 2006)

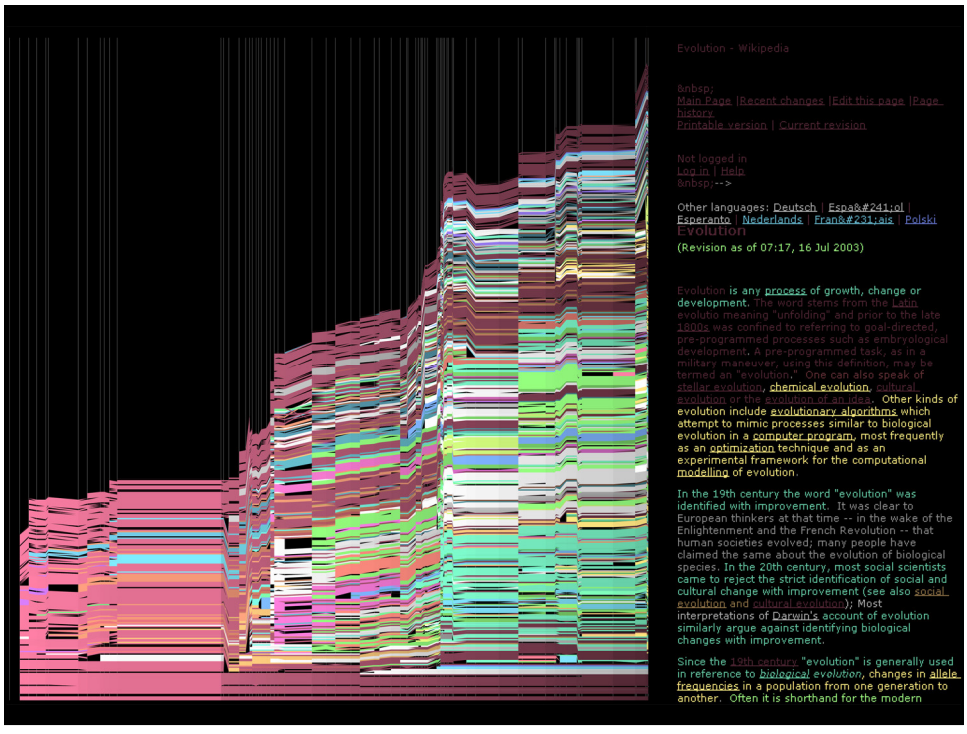


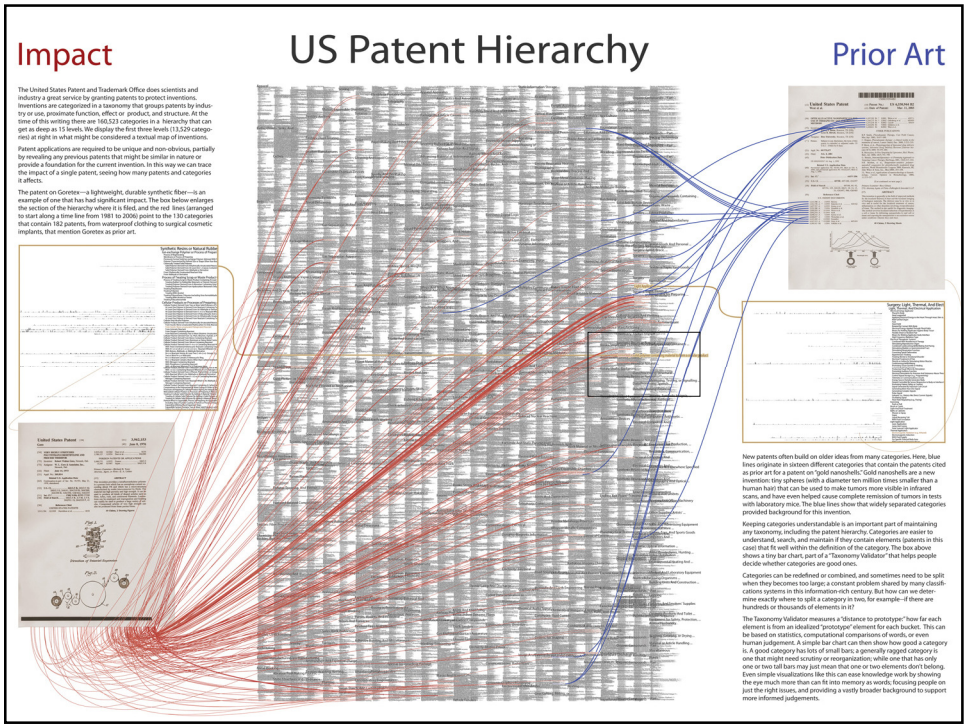


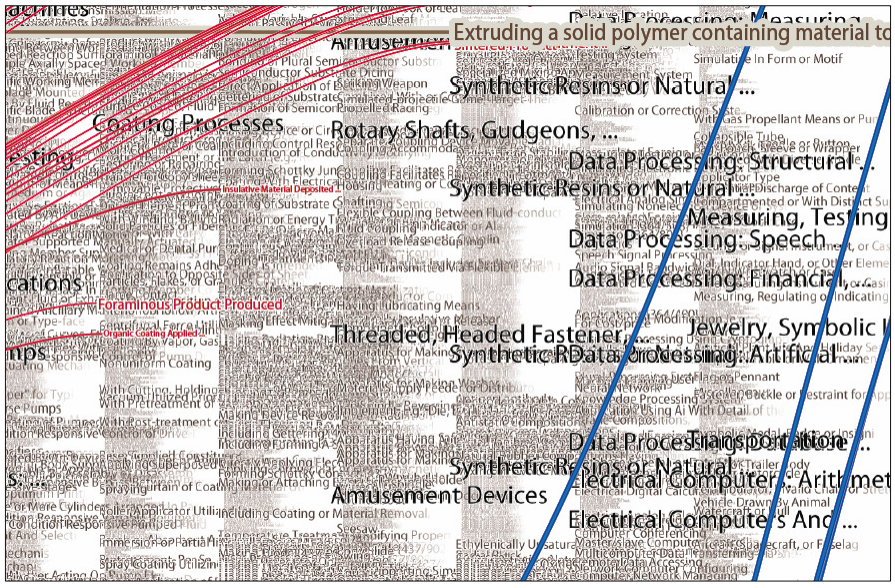
How would a reference system for all of science look?

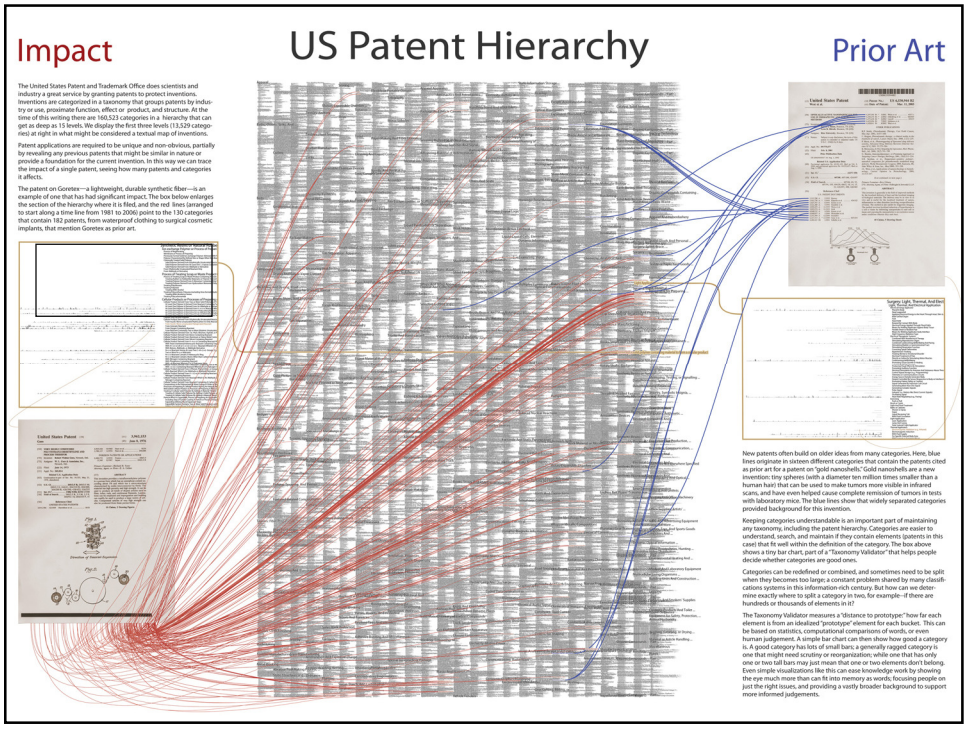
What dimensions would it have?

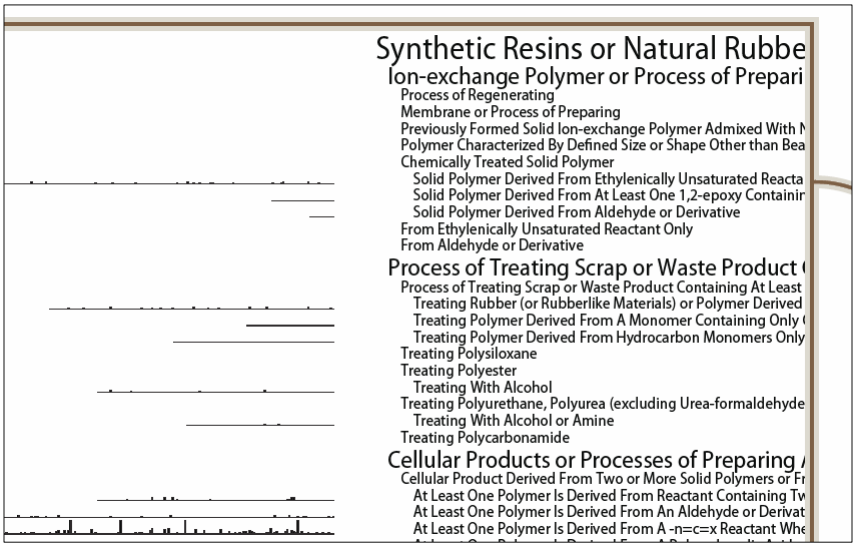

Writing the History of Science
Impact
Foundation
LINK COLOR
Historical Links (Identified by Isaac Asimov)
explicit
implicit
Coincident Citation Links
explicit
implicit
Non-Coincident Citation Links
explicit
implicit
NODE SIZE
Number of Citations
3,000
300
1800
1830
1860
1890
1920
1950
1960
1970
1980
1990
2000
2010


Evolution - Wikipedia
Main Page | Recent changes | Edit this page | Page history
Printable version | Current revision
Not logged in
Log in | Help
Other languages: Deutsch | Español | Esperanto | Nederlands | Français | Polski
Evolution
(Revision as of 07:17, 16 Jul 2003)
Evolution is any process of growth, change or development. The word stems from the Latin evolutio meaning "unfolding" and prior to the late 1800s was confined to referring to goal-directed, pre-programmed processes such as embryological development. A pre-programmed task, as in a military maneuver, using this definition, may be termed an "evolution." One can also speak of stellar evolution, chemical evolution, cultural evolution or the evolution of an idea. Other kinds of evolution include evolutionary algorithms which attempt to mimic processes similar to biological evolution in a computer program, most frequently as an optimization technique and as an experimental framework for the computational modelling of evolution.
In the 19th century the word "evolution" was identified with improvement. It was clear to European thinkers at that time -- in the wake of the Enlightenment and the French Revolution -- that human societies evolved; many people have claimed the same about the evolution of biological species. In the 20th century, most social scientists came to reject the strict identification of social and cultural change with improvement (see also social evolution and cultural evolution); Most interpretations of Darwin's account of evolution similarly argue against identifying biological changes with improvement.
Since the 19th century "evolution" is generally used in reference to biological evolution, changes in allele frequencies in a population from one generation to another. Often it is shorthand for the modern

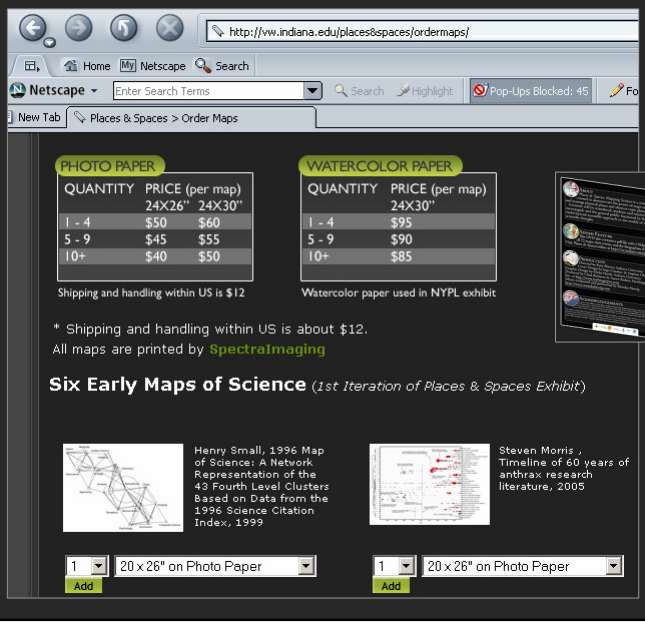

Impact
US Patent Hierarchy
Prior Art
The United States Patent and Trademark Office does scientists and industry a great service by granting patents to protect inventions. Inventions are categorized in a taxonomy that groups patents by industry or use, proximate function, effect or product, and structure. At the time of this writing there are 160,523 categories in a hierarchy that can get as deep as 15 levels. We display the first three levels (13,529 categories) at right in what might be considered a textual map of inventions.
Patent applications are required to be unique and non-obvious, partially by revealing any previous patents that might be similar in nature or provide a foundation for the current invention. In this way we can trace the impact of a single patent, seeing how many patents and categories it affects.
The patent on Goretex—a lightweight, durable synthetic fiber—is an example of one that has had significant impact. The box below enlarges the section of the hierarchy where it is filed, and the red lines (arranged to start along a time line from 1981 to 2006) point to the 130 categories that contain 182 patents, from waterproof clothing to surgical cosmetic implants, that mention Goretex as prior art.
New patents often build on older ideas from many categories. Here, blue lines originate in sixteen different categories that contain the patents cited as prior art for a patent on "gold nanoshells." Gold nanoshells are a new invention: tiny spheres (with a diameter ten million times smaller than a human hair) that can be used to make tumors more visible in infrared scans, and have even helped cause complete remission of tumors in tests with laboratory mice. The blue lines show that widely separated categories provided background for this invention.
Keeping categories understandable is an important part of maintaining any taxonomy, including the patent hierarchy. Categories are easier to understand, search, and maintain if they contain elements (patents in this case) that fit well within the definition of the category. The box above shows a tiny bar chart, part of a "Taxonomy Validator" that helps people decide whether categories are good ones.
Categories can be redefined or combined, and sometimes need to be split when they becomes too large; a constant problem shared by many classifications systems in this information-rich century. But how can we determine exactly where to split a category in two, for example—if there are hundreds or thousands of elements in it?
The Taxonomy Validator measures a "distance to prototype" how far each element is from an idealized "prototype" element for each bucket. This can be based on statistics, computational comparisons of words, or even human judgement. A simple bar chart can then show how good a category is. A good category has lots of small bars; a generally ragged category is one that might need scrutiny or reorganization; while one that has only one or two tall bars may just mean that one or two elements don't belong. Even simple visualizations like this can ease knowledge work by showing the eye much more than can fit into memory as words; focusing people on just the right issues, and providing a vastly broader background to support more informed judgements.


Extruding a solid polymer containing material to
Synthetic Resins or Natural ...
Coating Processes
Rotary Shafts, Gudgeons, ...
Data Processing: Structural ...
Synthetic Resins or Natural ...
Measuring, Testing
Data Processing: Speech
Data Processing: Financial, ...
Jewelry, Symbolic
Threaded, Headed Fastener, ...
Synthetic Resins or Natural ...
Data Processing: Artificial ...
Data Processing: Transportation ...
Synthetic Resins or Natural ...
Electrical Computer: Arithmet
Amusement Devices
Electrical Computers And ...


Impact
US Patent Hierarchy
Prior Art
The United States Patent and Trademark Office does scientists and industry a great service by granting patents to protect inventions. Inventions are categorized in a taxonomy that groups patents by industry or use, proximate function, effect or product, and structure. At the time of this writing there are 160,523 categories in a hierarchy that can get as deep as 15 levels. We display the first three levels (13,529 categories) at right in what might be considered a textual map of inventions.
Patent applications are required to be unique and non-obvious, partially by revealing any previous patents that might be similar in nature or provide a foundation for the current invention. In this way we can trace the impact of a single patent, seeing how many patents and categories it affects.
The patent on Goretex—a lightweight, durable synthetic fiber—is an example of one that has had significant impact. The box below enlarges the section of the hierarchy where it is filed, and the red lines (arranged to start along a time line from 1981 to 2006) point to the 130 categories that contain 182 patents, from waterproof clothing to surgical cosmetic implants, that mention Goretex as prior art.
New patents often build on older ideas from many categories. Here, blue lines originate in sixteen different categories that contain the patents cited as prior art for a patent on "gold nanoshells." Gold nanoshells are a new invention: tiny spheres (with a diameter ten million times smaller than a human hair) that can be used to make tumors more visible in infrared scans, and have even helped cause complete remission of tumors in tests with laboratory mice. The blue lines show that widely separated categories provided background for this invention.
Keeping categories understandable is an important part of maintaining any taxonomy, including the patent hierarchy. Categories are easier to understand, search, and maintain if they contain elements (patents in this case) that fit well within the definition of the category. The box above shows a tiny bar chart, part of a "Taxonomy Validator" that helps people decide whether categories are good ones.
Categories can be redefined or combined, and sometimes need to be split when they becomes too large; a constant problem shared by many classifications systems in this information-rich century. But how can we determine exactly where to split a category in two, for example--if there are hundreds or thousands of elements in it?
The Taxonomy Validator measures a "distance to prototype:" how far each element is from an idealized "prototype" element for each bucket. This can be based on statistics, computational comparisons of words, or even human judgement. A simple bar chart can then show how good a category is. A good category has lots of small bars; a generally ragged category is one that might need scrutiny or reorganization; while one that has only one or two tall bars may just mean that one or two elements don't belong. Even simple visualizations like this can ease knowledge work by showing the eye much more than can fit into memory as words; focusing people on just the right issues, and providing a vastly broader background to support more informed judgements.

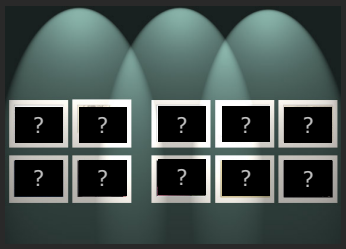

Synthetic Resins or Natural Rubbe
Ion-exchange Polymer or Process of Prepari
Process of Regenerating
Membrane or Process of Preparing
Previously Formed Solid Ion-exchange Polymer Admixed With N
Polymer Characterized By Defined Size or Shape Other than Bea
Chemically Treated Solid Polymer
Solid Polymer Derived From Ethylenically Unsaturated Reacta
Solid Polymer Derived From At Least One 1,2-epoxy Containir
Solid Polymer Derived From Aldehyde or Derivative
From Ethylenically Unsaturated Reactant Only
From Aldehyde or Derivative
Process of Treating Scrap or Waste Product (
Process of Treating Scrap or Waste Product Containing At Least
Treating Rubber (or Rubberlike Materials) or Polymer Derived
Treating Polymer Derived From A Monomer Containing Only (
Treating Polymer Derived From Hydrocarbon Monomers Only
Treating Polysiloxane
Treating Polyester
Treating With Alcohol
Treating Polyurethane, Polyurea (excluding Urea-formaldehyde
Treating With Alcohol or Amine
Treating Polycarbonamide
Cellular Products or Processes of Preparing /
Cellular Product Derived From Two or More Solid Polymers or Fr
At Least One Polymer Is Derived From Reactant Containing Tw
At Least One Polymer Is Derived From An Aldehyde or Derivat
At Least One Polymer Is Derived From A -n=c=x Reactant Whe

All Maps of Science Are for Sale

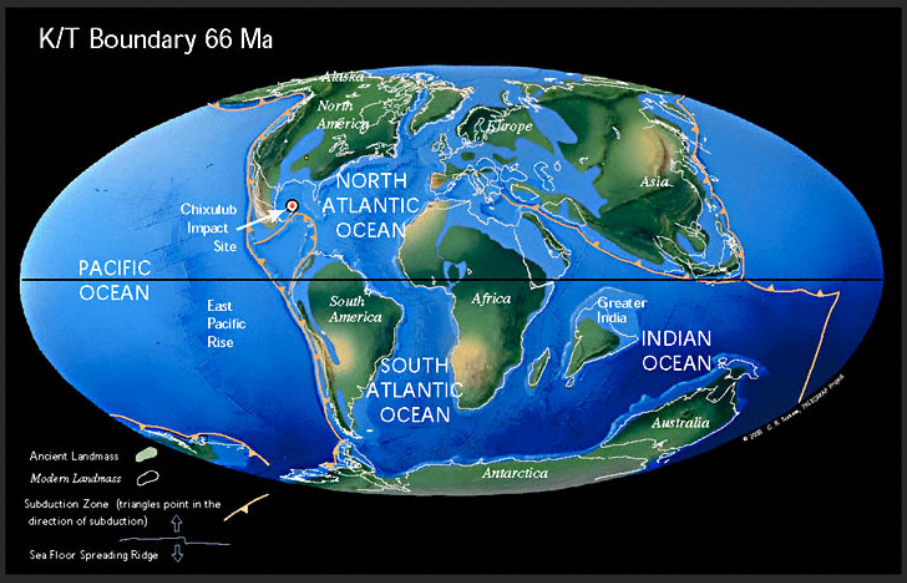


The Power of Forecasts

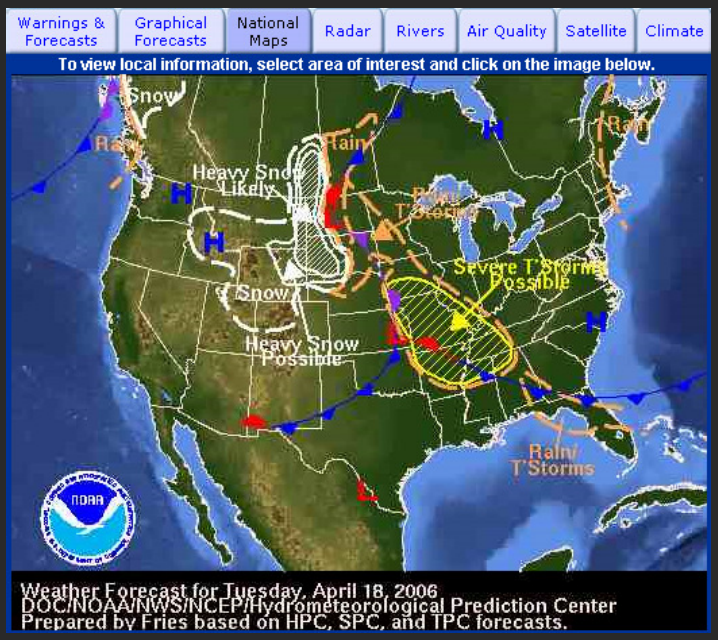
Four Existing Forecasts
VERSUS
Six Potential Science 'Weather' Forecasts



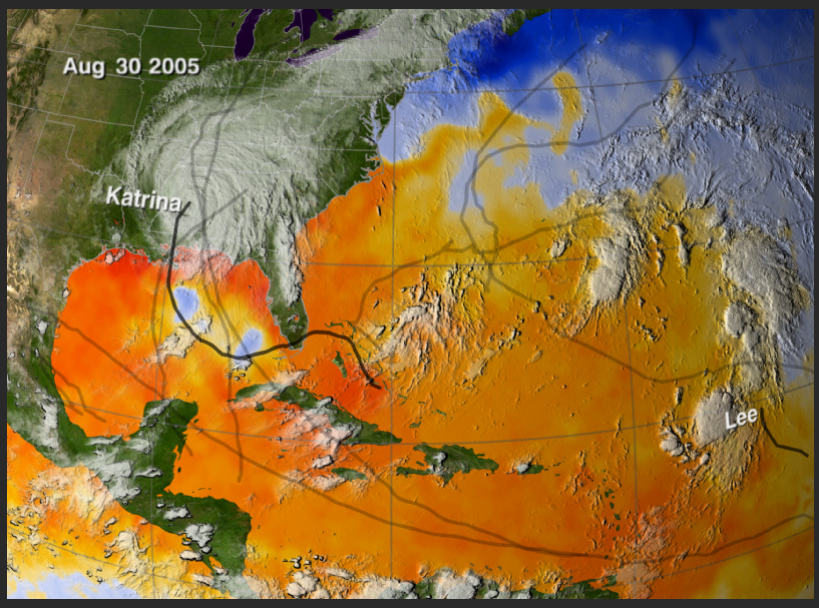
(3rd Iteration of Places & Spaces Exhibit - 2007)



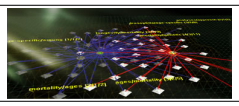
http://www.scotese.com/



http://www.weather.gov



Named Storms, available online at http://svs.gsfc.nasa.gov/vis/a000000/a003200/a003279



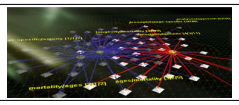
Modeling Science: Opportunities

- Dynamic scien…s, evolving networks, trend…
- Evolution of s…/ skills knowable by in…
- Interplay of co…
- Evolution of fi…
- Interactions an…
- Comparison of…grants, teach the field…
- Impact of publ…ics of fields.
- Diffusion of p…
- How to best c…

Ultimate goal:

Learn how to best…wledge.

61



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*, Special Issue on *Mapping Humanity's Knowledge and Expertise in the Digital Domain.*
- Börner, Katy, Penumarthy, Shashikant, Meiss, Mark and Ke, Weimao. Mapping the Diffusion of Scholarly Knowledge Among Major U.S. Research Institutions. Accepted for *Scientometrics*. Dedicated issue on the *10th International Conference of the International Society for Scientometrics and Informetrics* held in Stockholm.
- Holloway, Todd, Božicevic, Miran and Börner, Katy. Analyzing and Visualizing the Semantic Coverage of Wikipedia and Its Authors. Submitted to *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. Also available as cond-mat/0402380.
- 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. Also available as cond-mat/0311459.

62


Interested to get your own science map?
Contact the map makers!
katy@indiana.edu