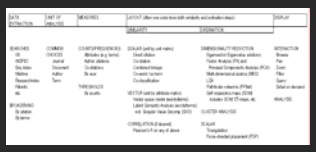
Recommendations for Evaluating Large, Interdisciplinary Research Initiatives



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

Trans-NCI Evaluation Special Interest Group
March 12th, 2008

Desirable Features of Large, Interdisciplinary Research Initiatives/Centers

'Large/major' funding should result in **high quantity and high quality scientific products** such as experts, papers, patents, databases, software tools, Cyberinfrastructures and associated

- intellectual capital
- social capital
- infrastructure capital

that is important to a research community.

They should have an interdisciplinary 'footprint', i.e., **consume and produce** scientific products from many scientific domains.

Education and Outreach are important.

Initiatives/Centers provide critical mass, physical facilities, longevity, stability, visibility that can have a major impact on the growth of a research community.

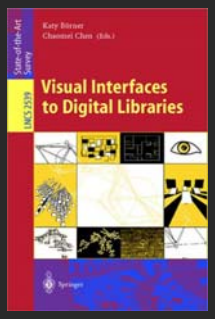
Evaluating Large, Interdisciplinary Research Initiatives/Centers

Data Acquisition
Acquire all (interdisciplinary) scientific products **consumed and produced** *(experts, papers, patents, databases, software tools, cyberinfrastructures, funding)* by the initiative(s)/center(s) in question as well from 'comparison' units.

Data Analysis
Number of consumed/produced products over time.
Geo and topic location of consumed/produced products.
Dynamic features, e.g., burst of activity, (social) network evolution, secondary effects.
Initiatives/Centers should be involved in positive (not negative) feedback cycles.

Result Communication
Top-n lists. Success stories.
Tables of major produced/consumed entities. **Profiles.**
Major produced/consumed entities in their (geo/network) context, e.g., paper citation graphs, funding-papers-PI graphs, evolving scholarly networks., impact on education/public policy.

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, Medford, NJ: Information Today, Inc./American Society for Information Science and Technology, Volume 37, Chapter 5, pp. 179-255.* http://ivl.slis.indiana.edu/km/pub/2003-borner-arist.pdf
- *Shiffrin, Richard M. and Börner, Katy (Eds.) (2004).* ***Mapping Knowledge Domains.*** *Proceedings of the National Academy of Sciences of the United States of America, 101(Suppl_1).* http://www.pnas.org/content/vol101/suppl_1/
- *Börner, Katy, Sanyal, Soma and Vespignani, Alessandro (2007).* ***Network Science.*** *In Blaise Cronin (Ed.), Annual Review of Information Science & Technology, Information Today, Inc./American Society for Information Science and Technology, Medford, NJ, Volume 41, Chapter 12, pp. 537-607.* http://ivl.slis.indiana.edu/km/pub/2007-borner-arist.pdf

- ***Places & Spaces: Mapping Science*** *exhibit, see also http://scimaps.org.*

General Process of Data Acquisition, Analysis and Visualization

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 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, 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.

Analysis of Emergent Research Areas

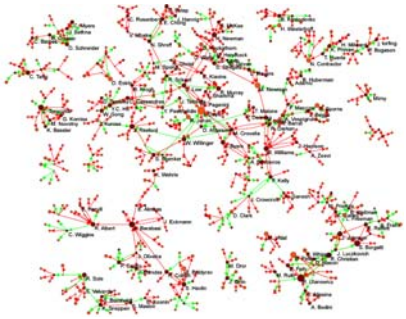
Mapping Network Science

Ke, Weimao & Börner, Katy. (2005). Mapping the Social Network and Expertise of "Network Science" Researchers. Report to the U.S. National Research Council study on Network Science, 88-92, The National Academies Press.

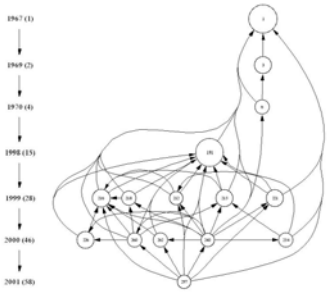
Börner, Katy. (2007). Making Sense of Mankind's Scholarly Knowledge and Expertise: Collecting, Interlinking, and Organizing What We Know and Different Approaches to Mapping (Network) Science. Environment and Planning B: Planning and Design. Vol. 34(5), 808-825, Pion.

Data Acquisition via

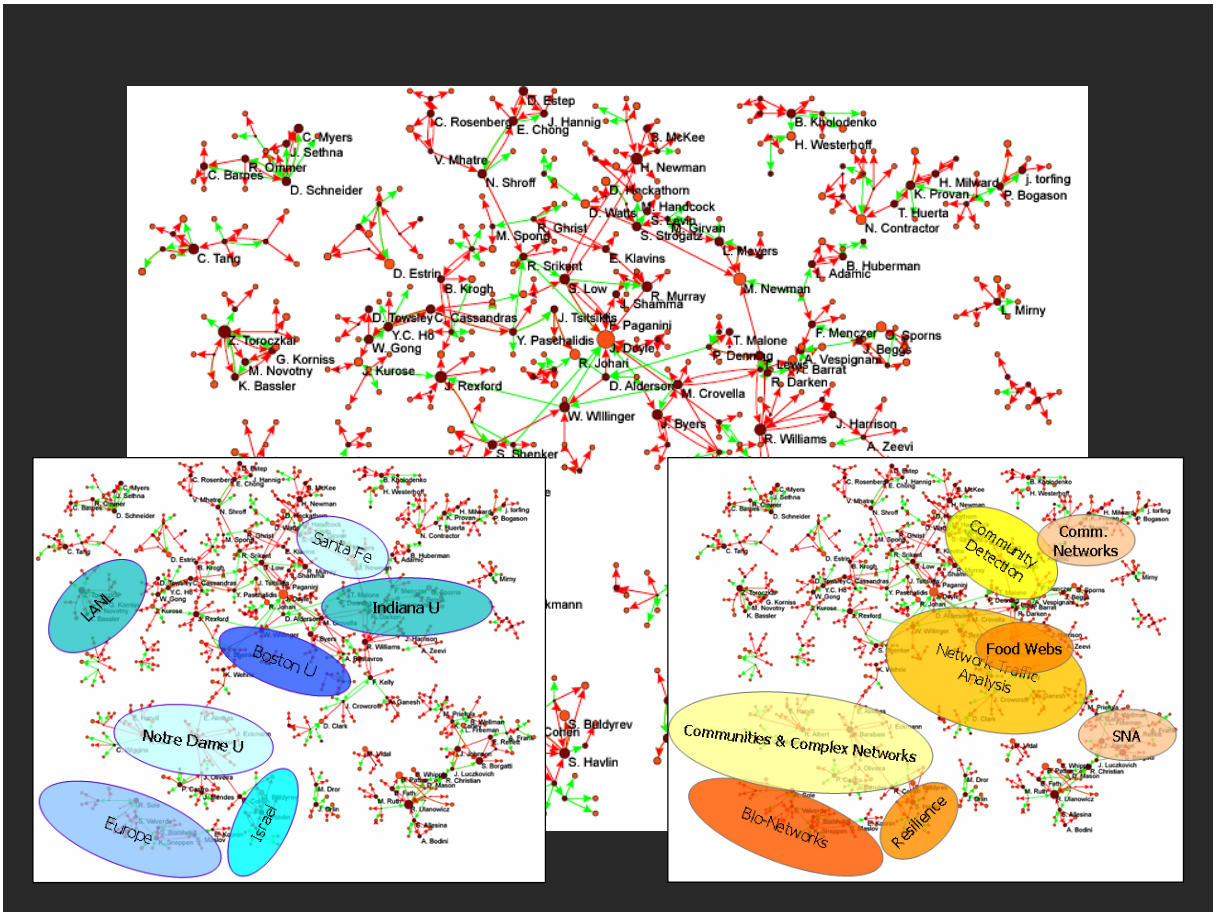
Questionnaires

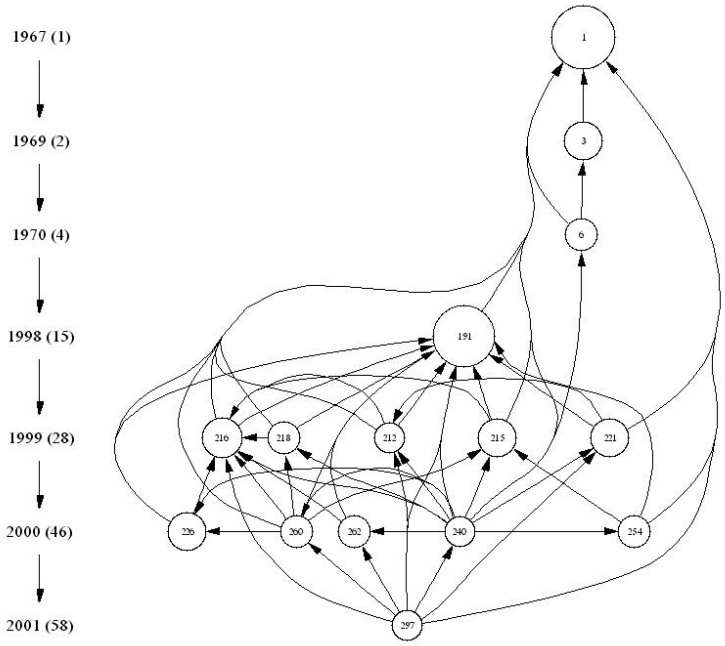


ISI Citation Data



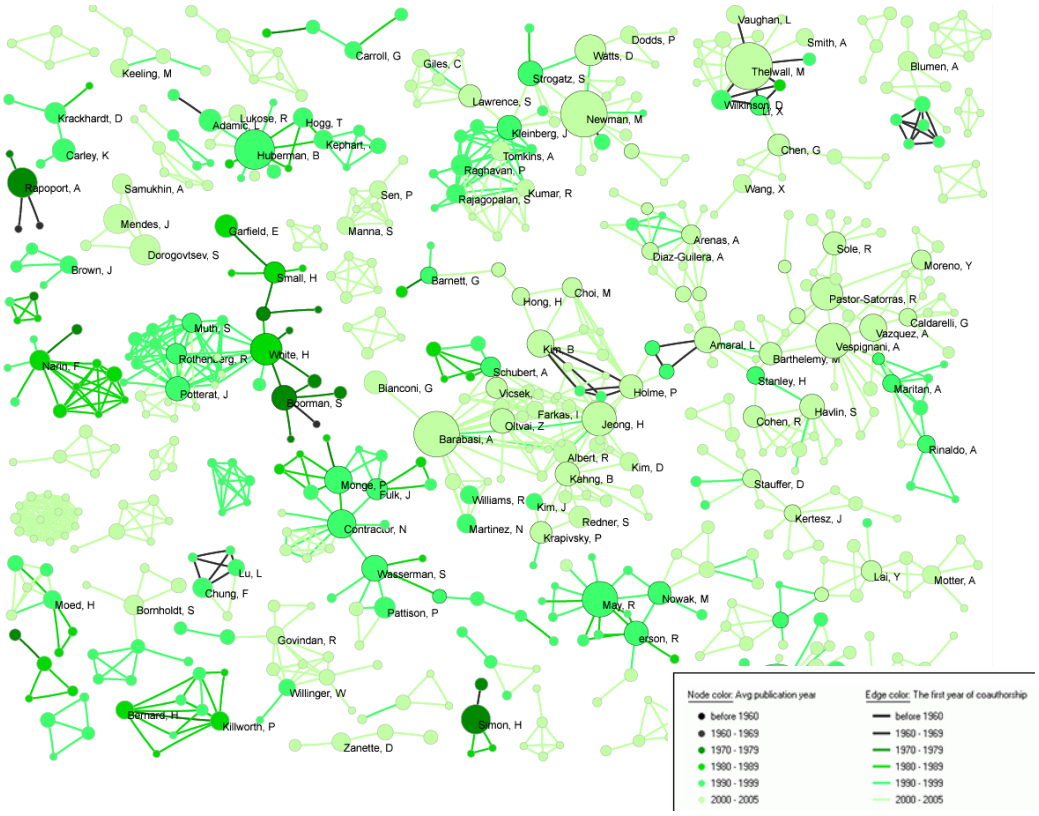
Bibliography Data



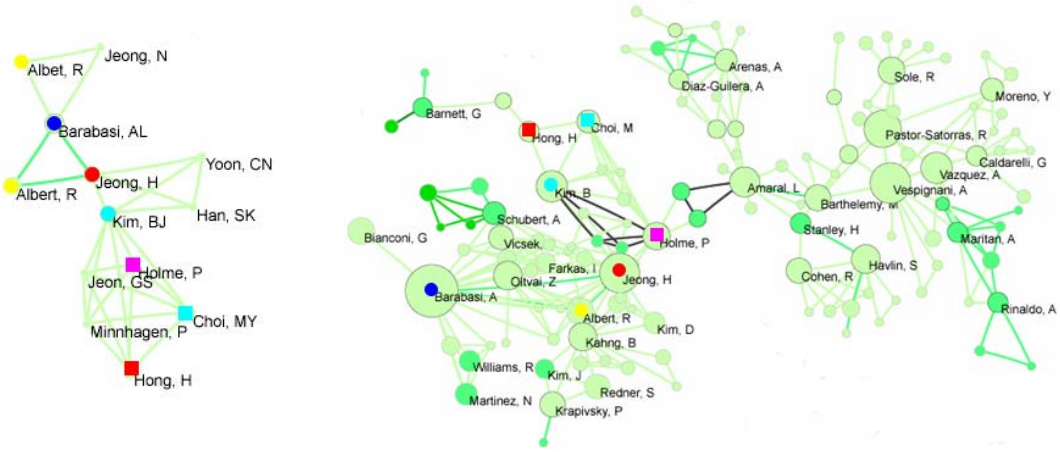


1 1967 MILGRAM S, GCS: 148 LCS: 147
3 1969 TRAVERS J, GCS: 64 LCS: 43
6 1970 KORTE C, GCS: 38 LCS: 31
191 1998 Watts DJ, GCS: 365 LCS: 131
212 1999 Barabasi AL, GCS: 75 LCS: 22
215 1999 Newman MEJ, GCS: 71 LCS: 45
216 1999 Barthelemy M, GCS: 71 LCS: 46
218 1999 Newman MEJ, GCS: 40 LCS: 27
221 1999 Barabasi AL, GCS: 245 LCS: 45
226 2000 Barrat A, GCS: 57 LCS: 39
240 2000 Newman MEJ, GCS: 37 LCS: 22
254 2000 Moore C, GCS: 48 LCS: 25
260 2000 Newman MEJ, GCS: 34 LCS: 26
262 2000 Amaral LAN, GCS: 106 LCS: 28
297 2001 Strogatz SH, GCS: 111 LCS: 23

HistCite paper-citation graph of publications



Components with size larger than three of the co-author network based on bibliography data.

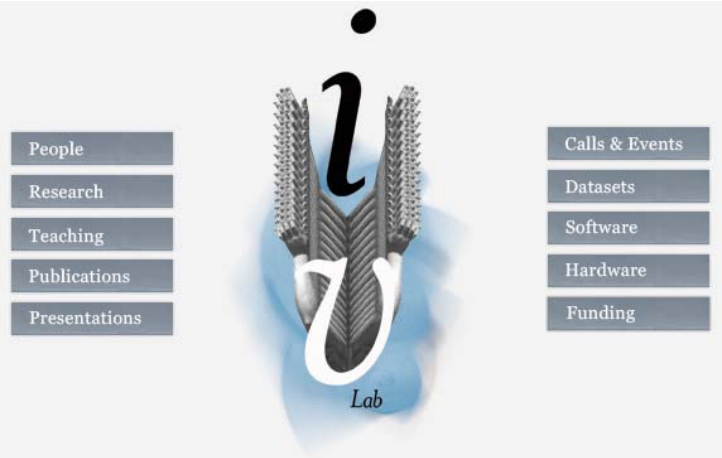


Comparison of the giant components of the co-author network based on citation (left) and bibliography data (right).

Data Acquisition for Comprehensive Analysis

Lab/Center Management System vs. Spacebook and MS Famulus

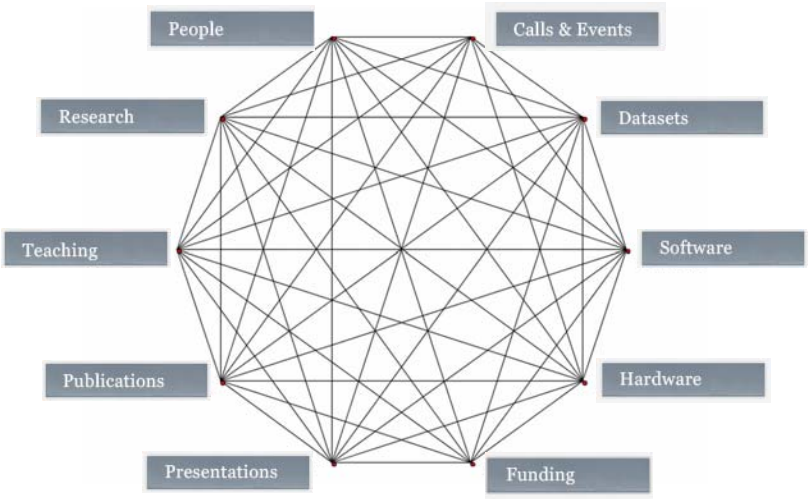
Designed to track, manage, and make use of data relevant for the daily operation of a medium size research team.



http://ivl.slis.indiana.edu

Data Entities and Interlinkages

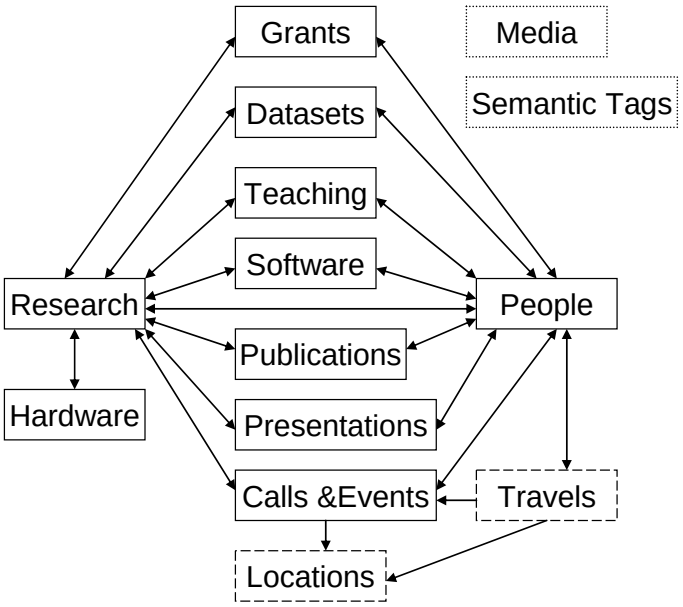
Designed for team leads, members, IT admins but also for external scholars and funding agencies.



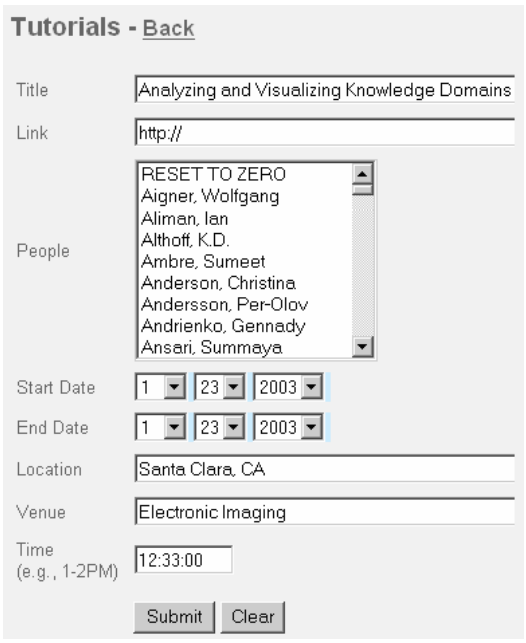
Not covered:
- Queries
- Workflows
- Protocols

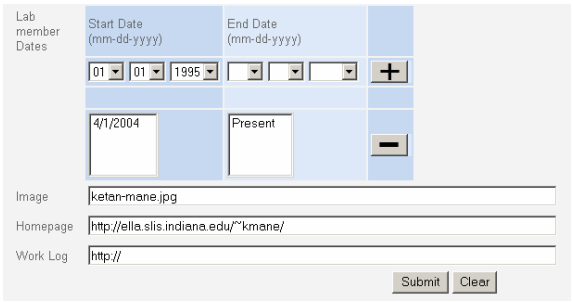
- Comments
- Bookmarks
- Ratings

Simplified representation of the IVL database schema

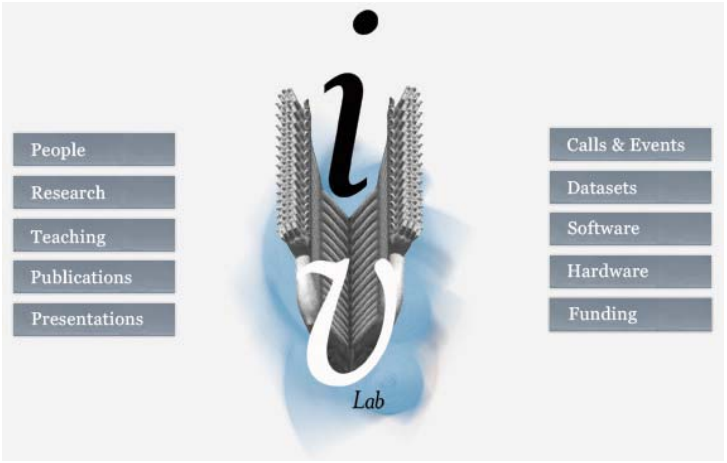


Data Entry



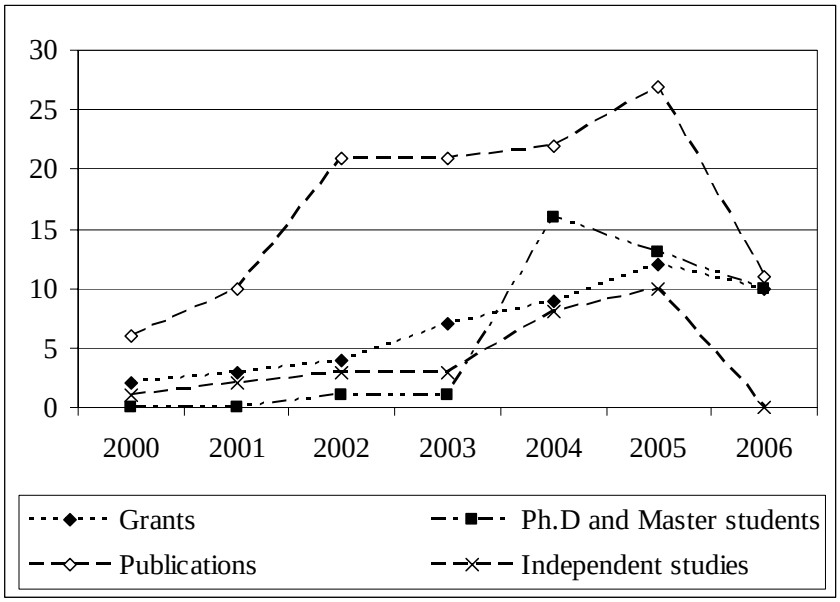


Demo

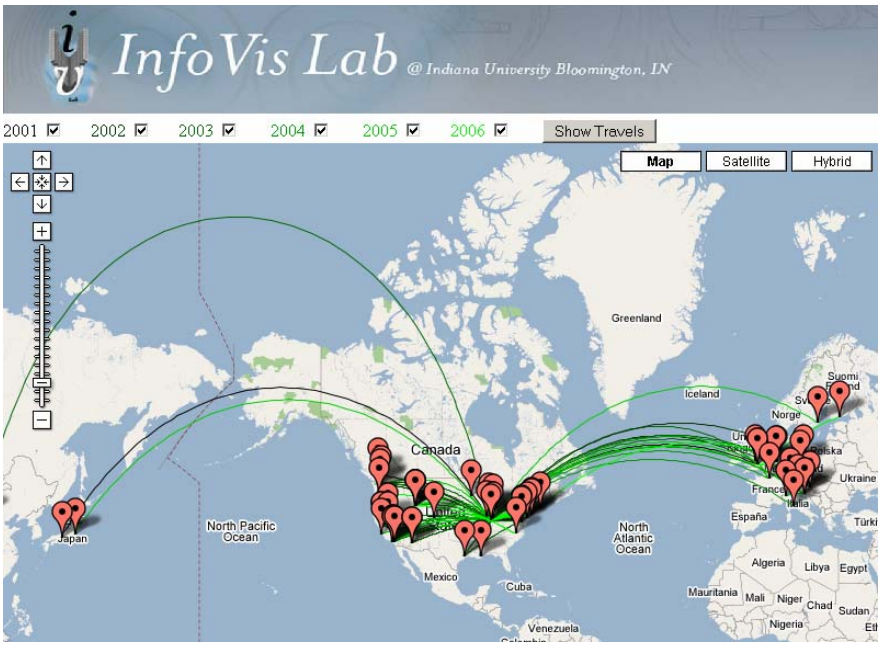


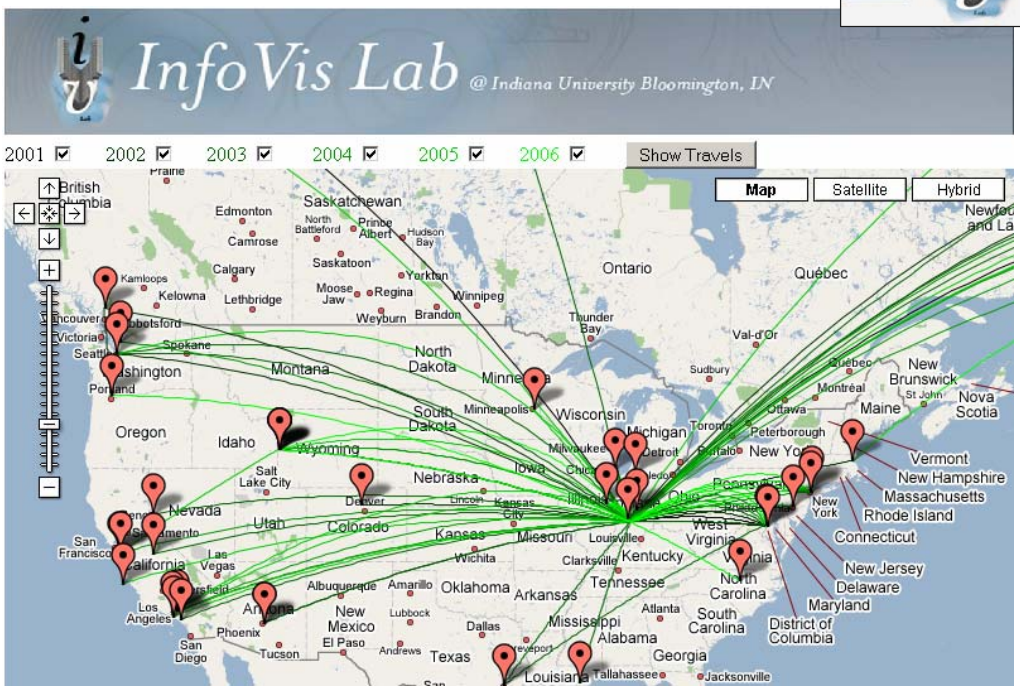
http://ivl.slis.indiana.edu

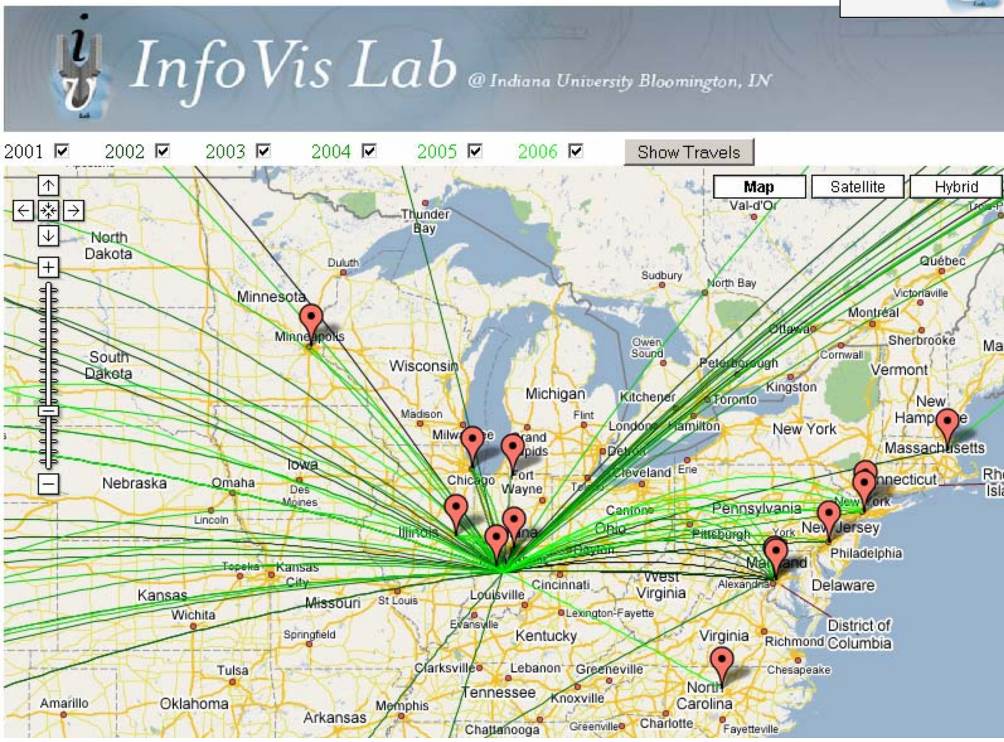
Time series analysis & visualization

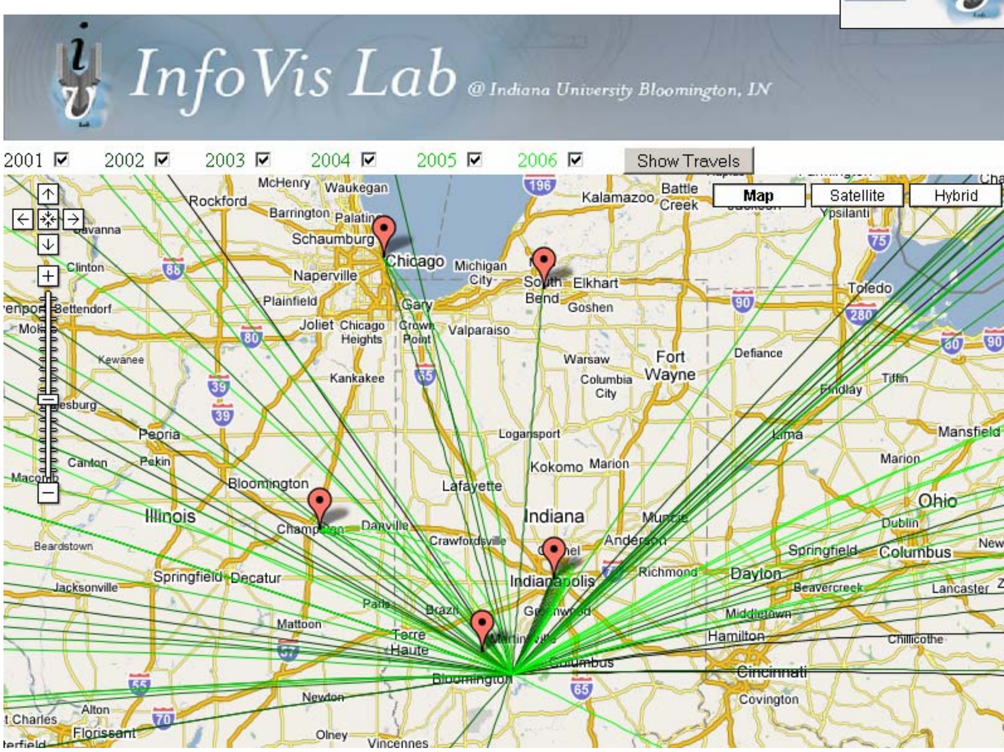


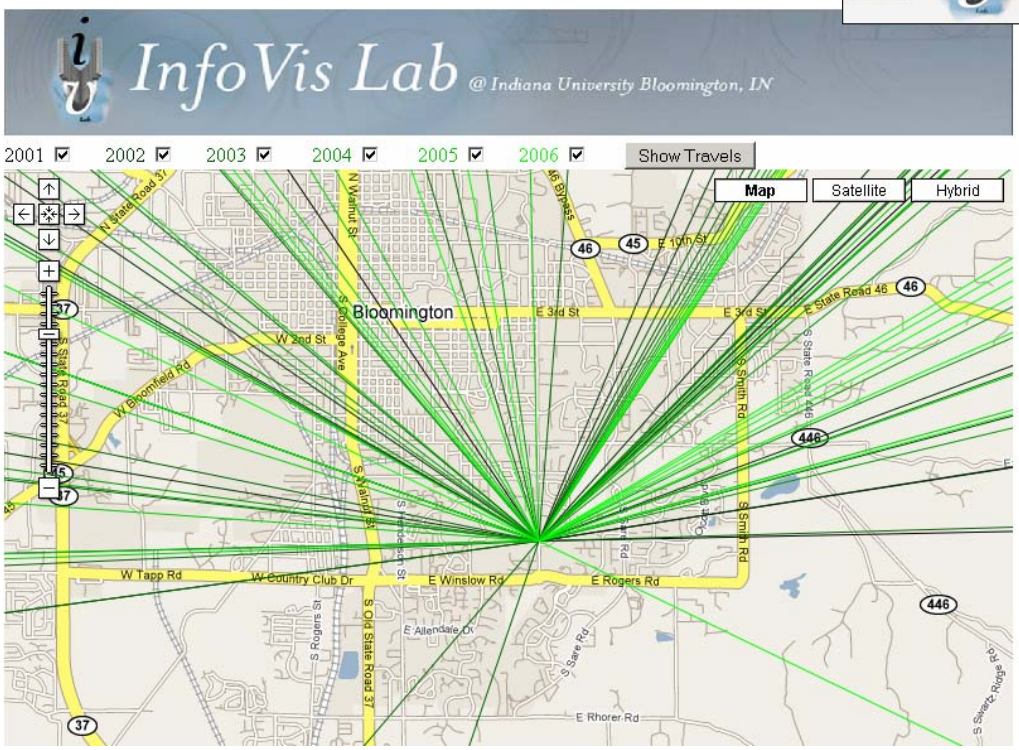
Katy's Travels in 2000-2006

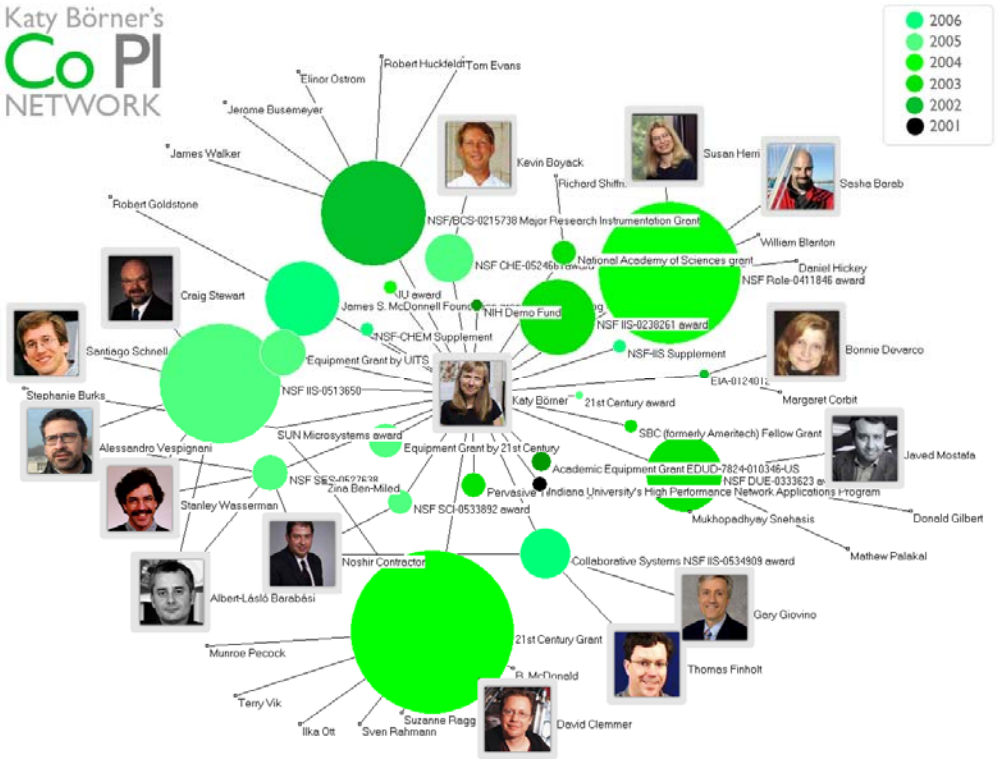





InfoVis Lab
@ Indiana University Bloomington, IN
2001 2002 2003 2004 2005 2006
Show Travels
Map
Satellite
Hybrid
North Dakota
South Dakota
Minnesota
Minneapolis
Wisconsin
Michigan
Iowa
Nebraska
Omaha
Kansas
Wichita
Missouri
Illinois
Chicago
Kentucky
Tennessee
Oklahoma
Arkansas
Ohio
Pennsylvania
New York
Vermont
New Hampshire
Massachusetts
New Jersey
Philadelphia
Maryland
Delaware
West Virginia
Virginia
North Carolina
District of Columbia
Toronto
Ottawa
Montreal
Québec
Cincinnati
Louisville
Memphis
Tulsa
Amarillo
Thunder Bay
Duluth
Sudbury
North Bay
Kingston
Hamilton
London
Detroit
Cleveland
Pittsburgh
Richmond
Charlotte
Knoxville
Lexington-Fayette
St Louis
Springfield
Topeka
Kansas City
Lincoln
Des Moines
Madison
Flint
Kitchener
Chattanooga
Chesapeake
Fayetteville

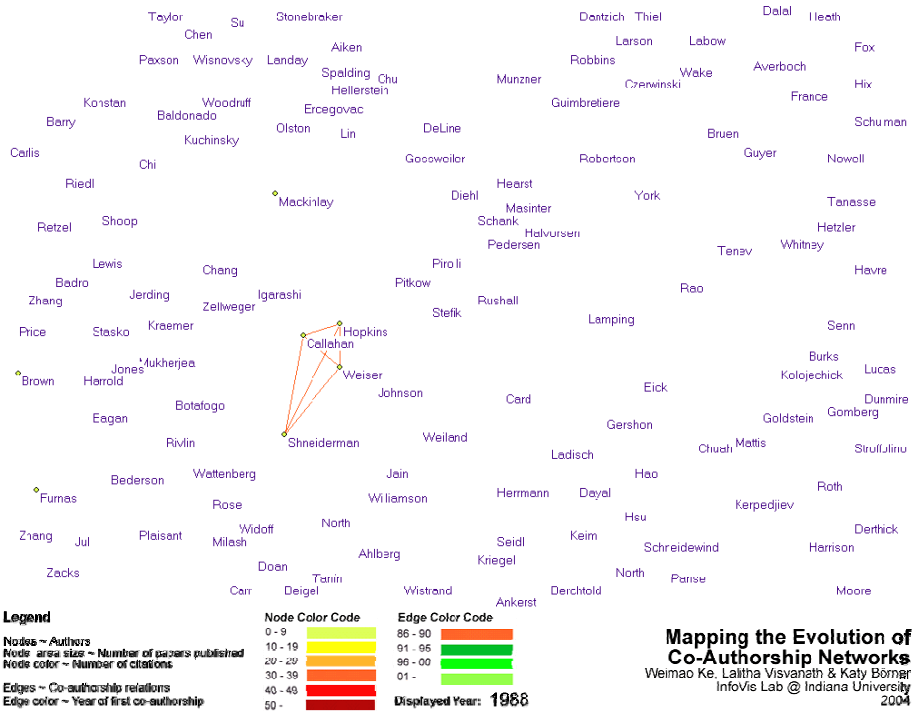

InfoVis Lab
@ Indiana University Bloomington, IN
2001 2002 2003 2004 2005 2006
Show Travels
Map
Satellite
Hybrid
Rockford
McHenry
Waukegan
Schaumburg
Chicago
Naperville
Plainfield
Joliet
Gary
Valparaiso
Michigan City
South Bend
Elkhart
Goshen
Kalamazoo
Battle Creek
Ypsilanti
Toledo
Defiance
Findlay
Tiffin
Fort Wayne
Warsaw
Columbia City
Kankakee
Peoria
Pekin
Canton
Bloomington
Champaign
Danville
Illinois
Springfield
Decatur
Jacksonville
Beardstown
Logansport
Kokomo
Marion
Lafayette
Indiana
Muncie
Anderson
Indianapolis
Richmond
Dayton
Springfield
Columbus
Ohio
Mansfield
Lima
Crawfordsville
Terre Haute
Mattoon
Bloomington
Columbus
Hamilton
Middletown
Cincinnati
Covington
Chillicothe
Lancaster
Beavercreek
Dublin
Newton
Olney
Vincennes
Alton
Florissant
Kewanee
Clinton
Bettendorf
Macomb


InfoVis Lab @ Indiana University Bloomington, IN
2001
2002
2003
2004
2005
2006
Show Travels
Map
Satellite
Hybrid
Bloomington
N State Road 37
N Walnut St
S College Ave
W 2nd St
W Bloomfield Rd
S State Road 37
E 3rd St
E 10th St
46 Bypass
E State Road 46
S State Road 446
S Smith Rd
W Tapp Rd
W Country Club Dr
E Winslow Rd
E Rogers Rd
S Rogers St
S Old State Road 37
E Allendale Dr
S Sare Rd
E Rhorer Rd
S Swartz Ridge Rd
46
45
446
37


Katy Börner's
Co PI
NETWORK
2006
2005
2004
2003
2002
2001
Elinor Ostrom
Robert Huckfeldt
Tom Evans
Jerome Busemeyer
James Walker
Kevin Boyack
Susan Herr
Richard Shiffrin
Sasha Barab
Robert Goldstone
NSF/BCS-0215738 Major Research Instrumentation Grant
William Blanton
NSF CHE-052466
National Academy of Sciences grant
Daniel Hickey
NSF Role-0411846 award
Craig Stewart
JU award
James S. McDonnell Found
NIH Demo Fund
NSF IIS-0238261 award
Santiago Schnell
NSF-CHEM Supplement
NSF-IIS Supplement
Bonnie Deverco
Equipment Grant by UITS
EIA-0124011
Stephanie Burks
NSF IIS-0513650
Katy Börner
21st Century award
Margaret Corbit
SBC (formerly Ameritech) Fellow Grant
SUN Microsystems award
Equipment Grant by 21st Century
Alessandro Vespignani
Javed Mostafa
Academic Equipment Grant EDUD-7824-010346-US
NSF DUE-0333623
Zina Ben-Miled
Pervasive Indiana University's High Performance Network Applications Program
Stanley Wasserman
NSF SCI-0533892 award
Donald Gilbert
Mukhopadhyay Snehasis
Mathew Palakal
Noshir Contractor
Collaborative Systems NSF IIS-0534909 award
Albert-László Barabási
Gary Giovino
21st Century Grant
Munroe Pecock
R McDonald
Thomas Finholt
Terry Vik
Suzanne Ragg
David Clemmer
Ilka Ott
Sven Rahmann

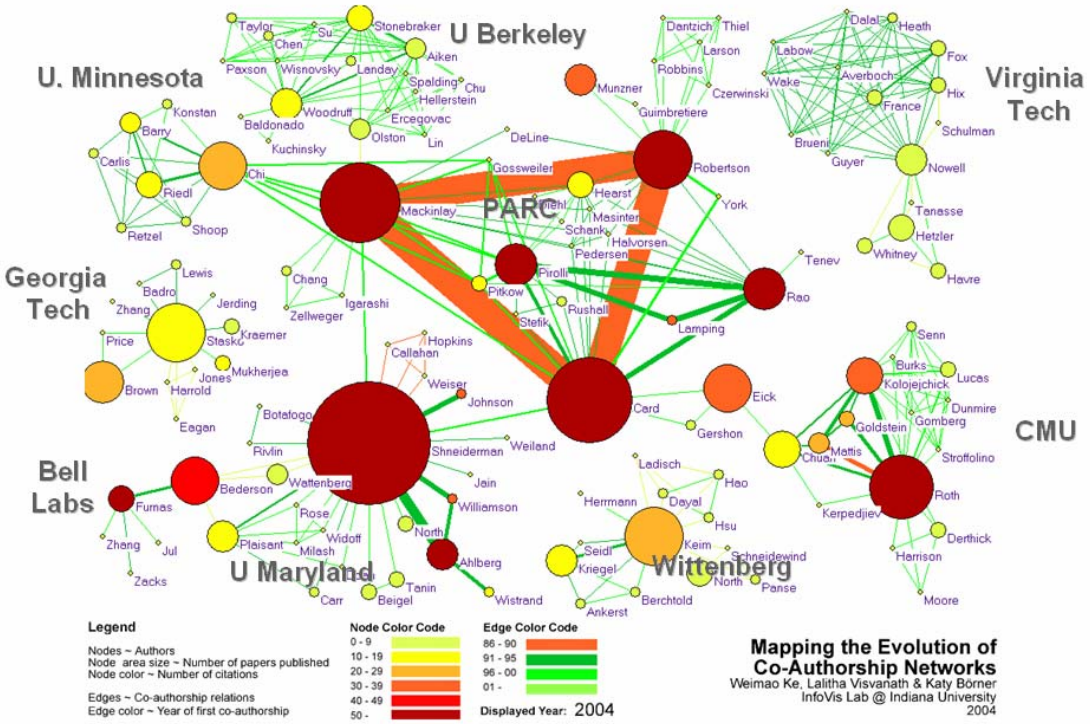
Mapping the Evolution of Co-Authorship Networks

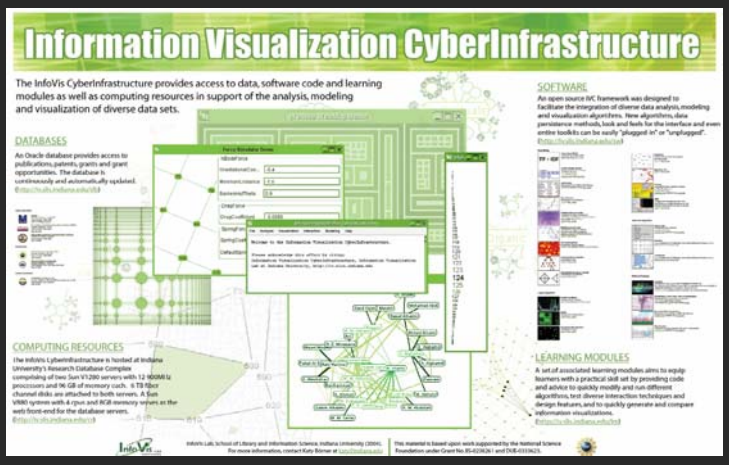
Ke, Visvanath & Börner, (2004) Won 1st price at the IEEE InfoVis Contest.



Mapping the Evolution of Co-Authorship Networks

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





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

CIShell
cyberinfrastructure shell

SDB | SCHOLARLY DATABASE

Scholarly Database
http://sdb.slis.indiana.edu

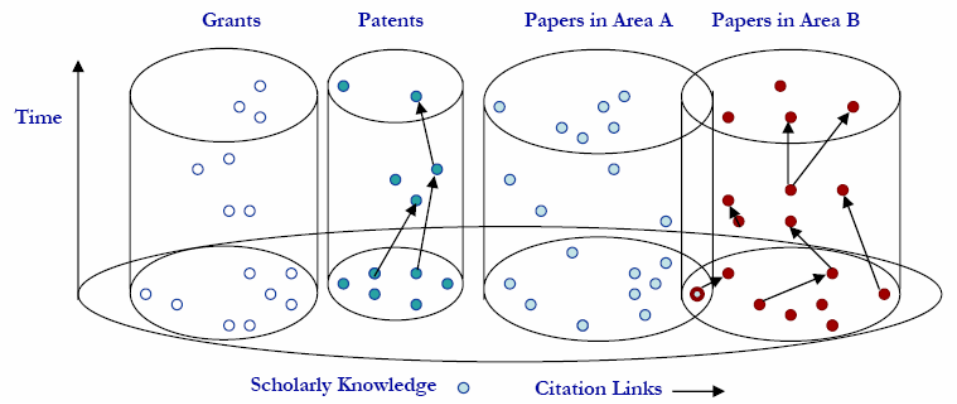
Network Workbench
A Workbench for Network Scientists

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. http://nwb.slis.indiana.edu

27

SDB | SCHOLARLY DATABASE

Challenges - Interlink $ Input & Publication/Patent Citation Output

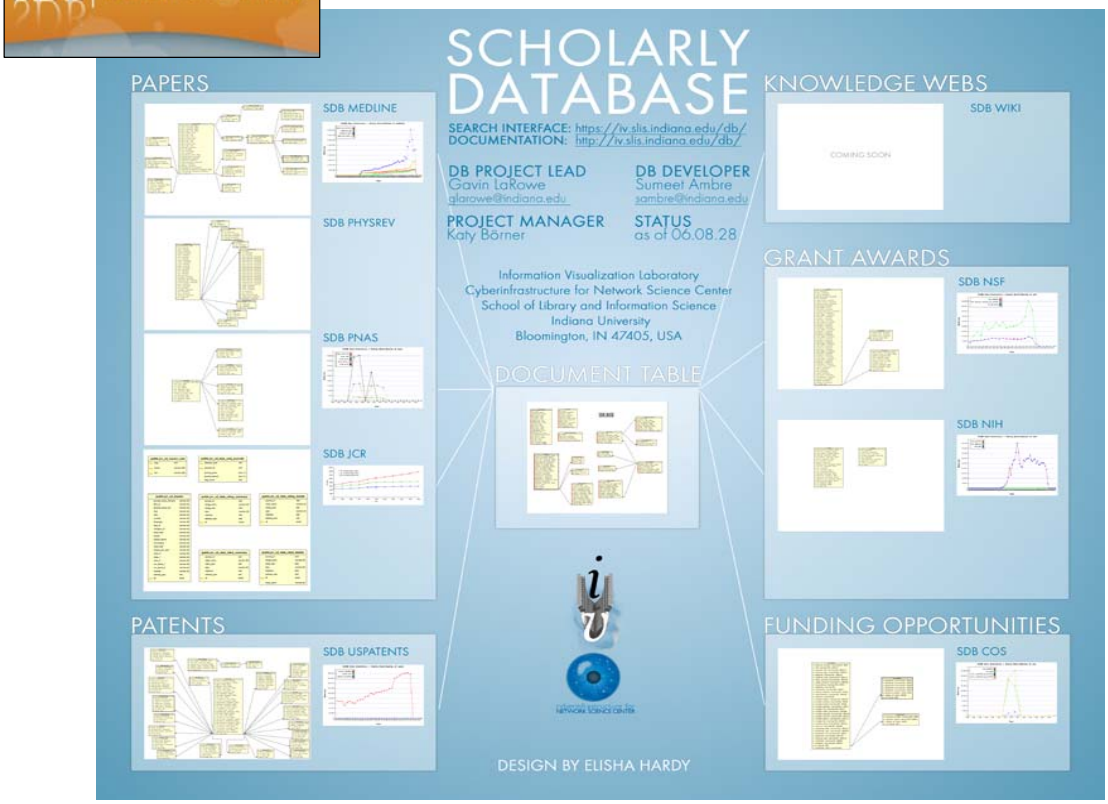


Need to interlink

- Grants and papers/patents.
- Grants/papers/patents and their PIs/authors/inventors, etc.

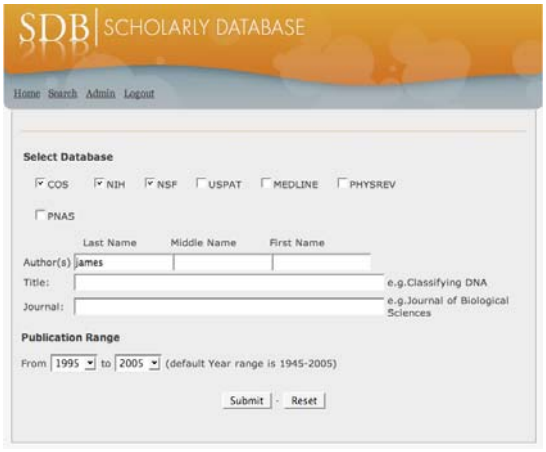
Use resulting networks to

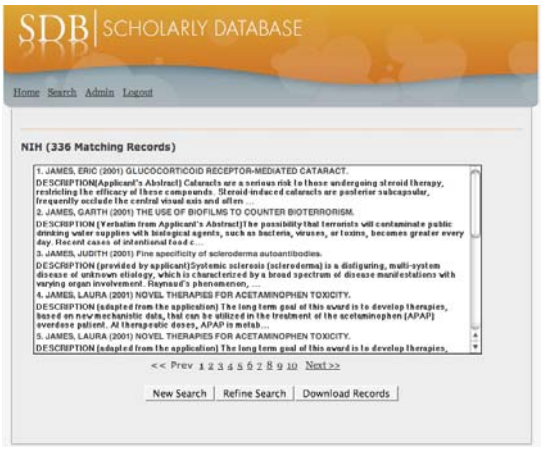
- Count #papers, #citations, etc.
- Determine strength of co-PI/author/inventor relations, etc.



Scholarly Database: Web Interface

Search across publications, patents, grants.
Download records and/or (evolving) co-author, paper-citation networks.





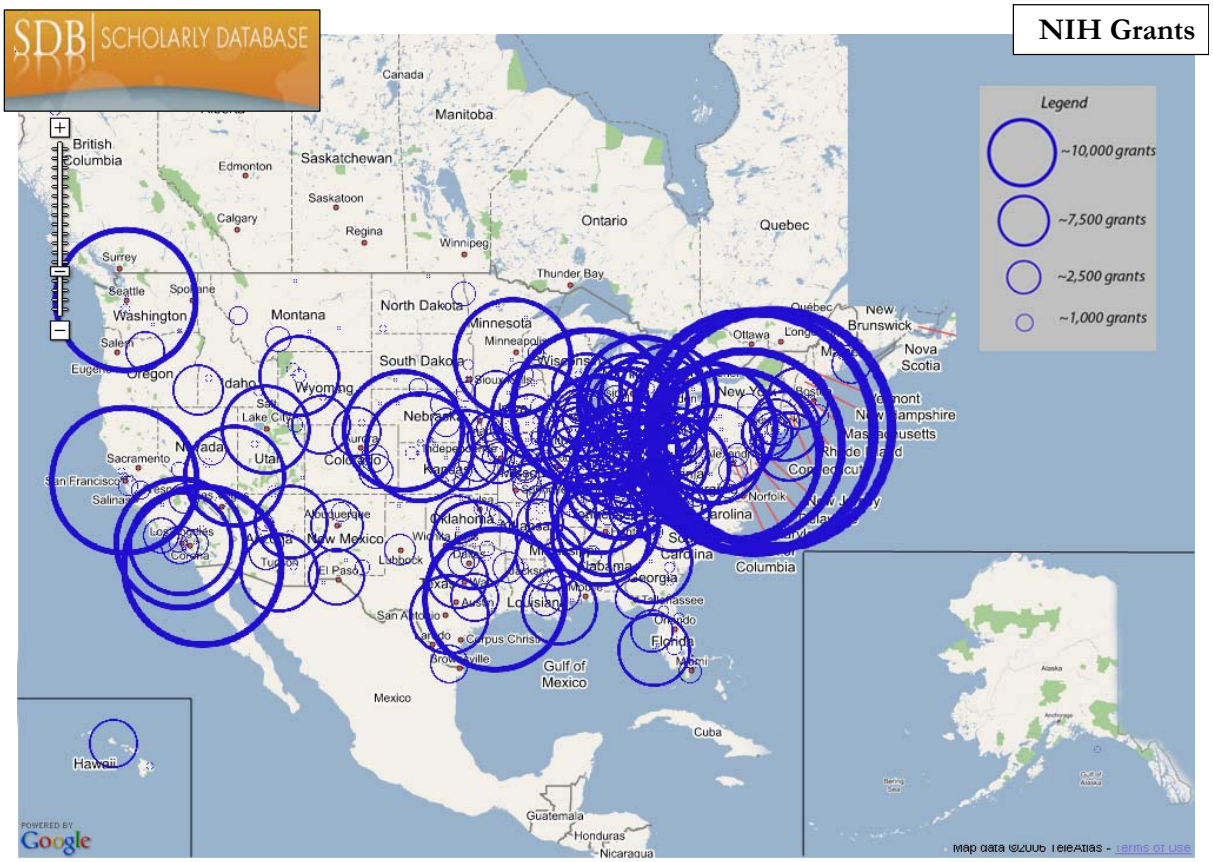
Register for free access at https://sdb.slis.indiana.edu.

Scholarly Database: # Records & Years Covered

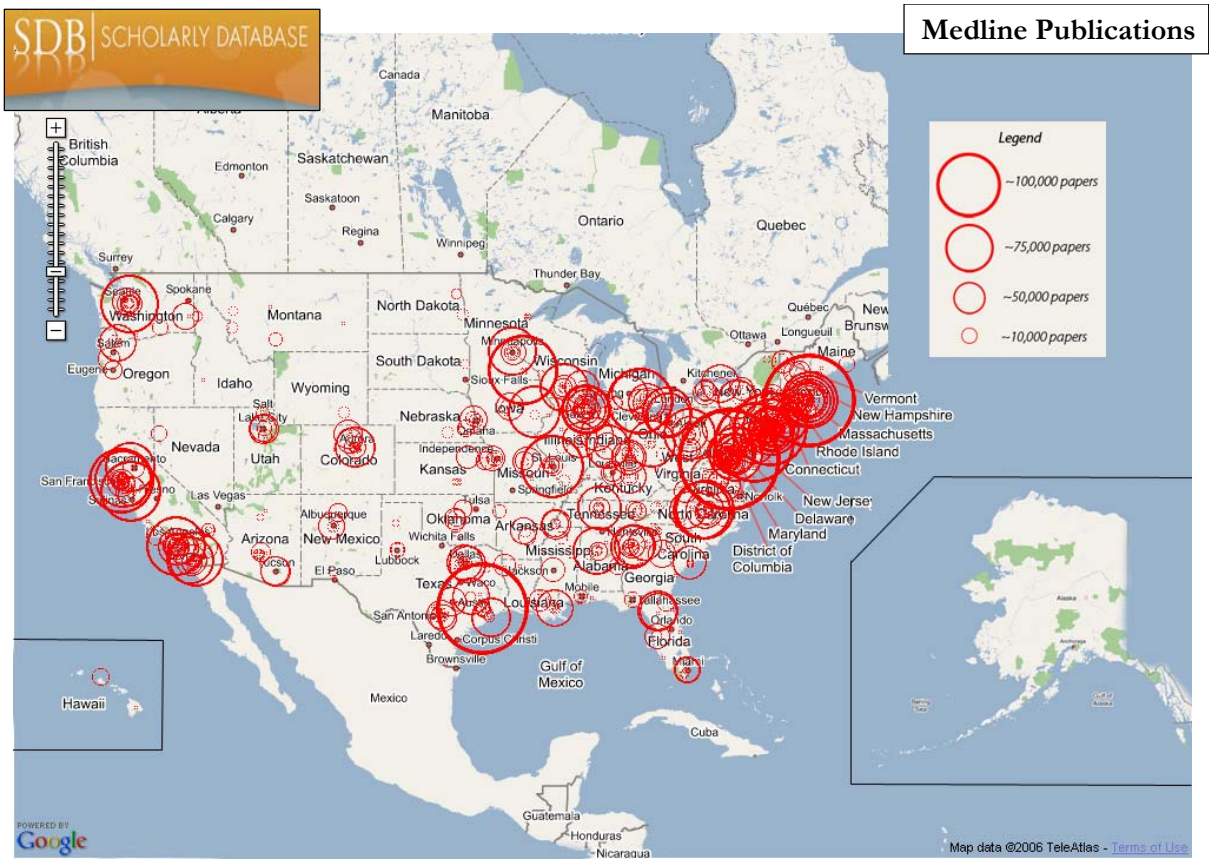
Datasets available via the Scholarly Database (* future feature)

Dataset	# Records	Years Covered	Updated	Restricted Access
Medline	13,149,741	1965-2005	Yes	
PhysRev	398,005	1893-2006		Yes
PNAS	16,167	1997-2002		Yes
JCR	59,078	1974, 1979, 1984, 1989 1994-2004		Yes
USPTO	3,179,930	1976-2004	Yes*	
NSF	174,835	1985-2003	Yes*	
NIH	1,043,804	1972-2002	Yes*	
Total	**18,021,560**	**1893-2006**	**4**	**3**

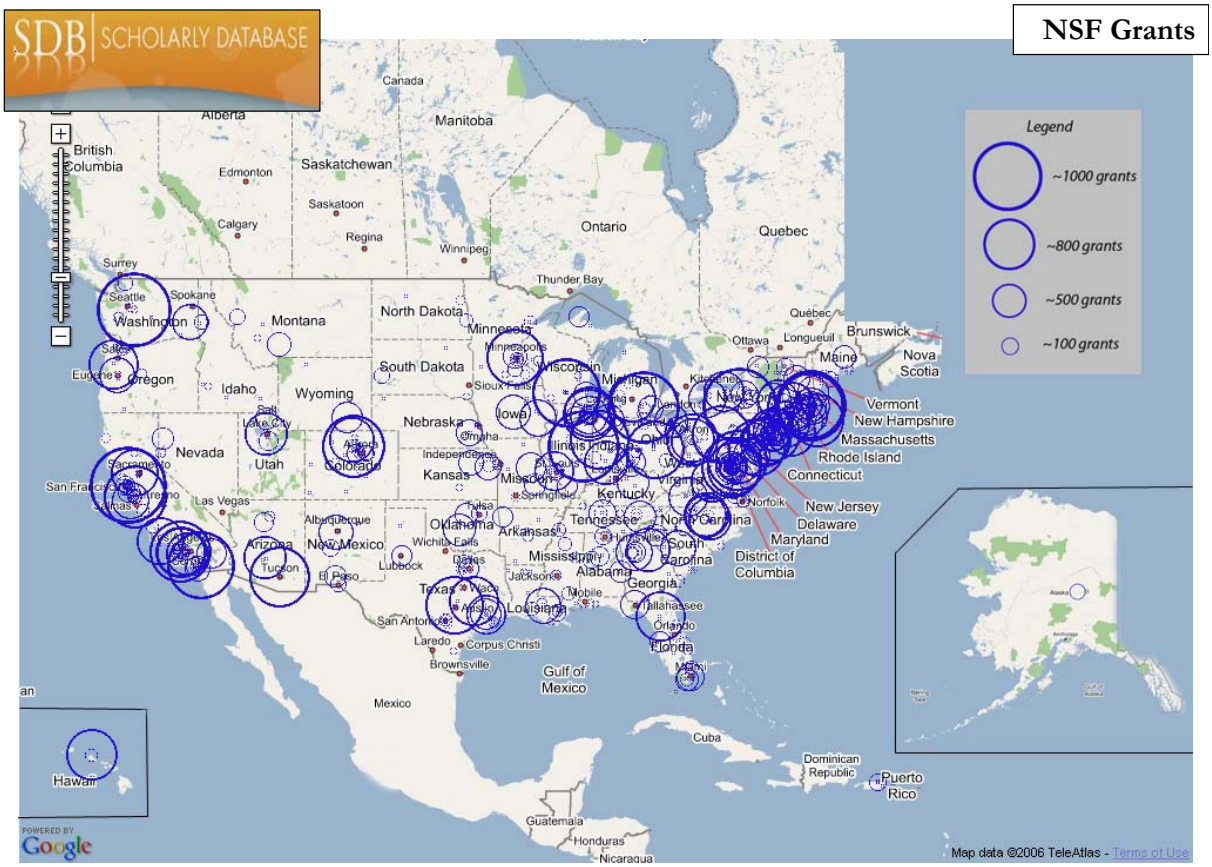
Aim for comprehensive time, geospatial, and topic coverage.

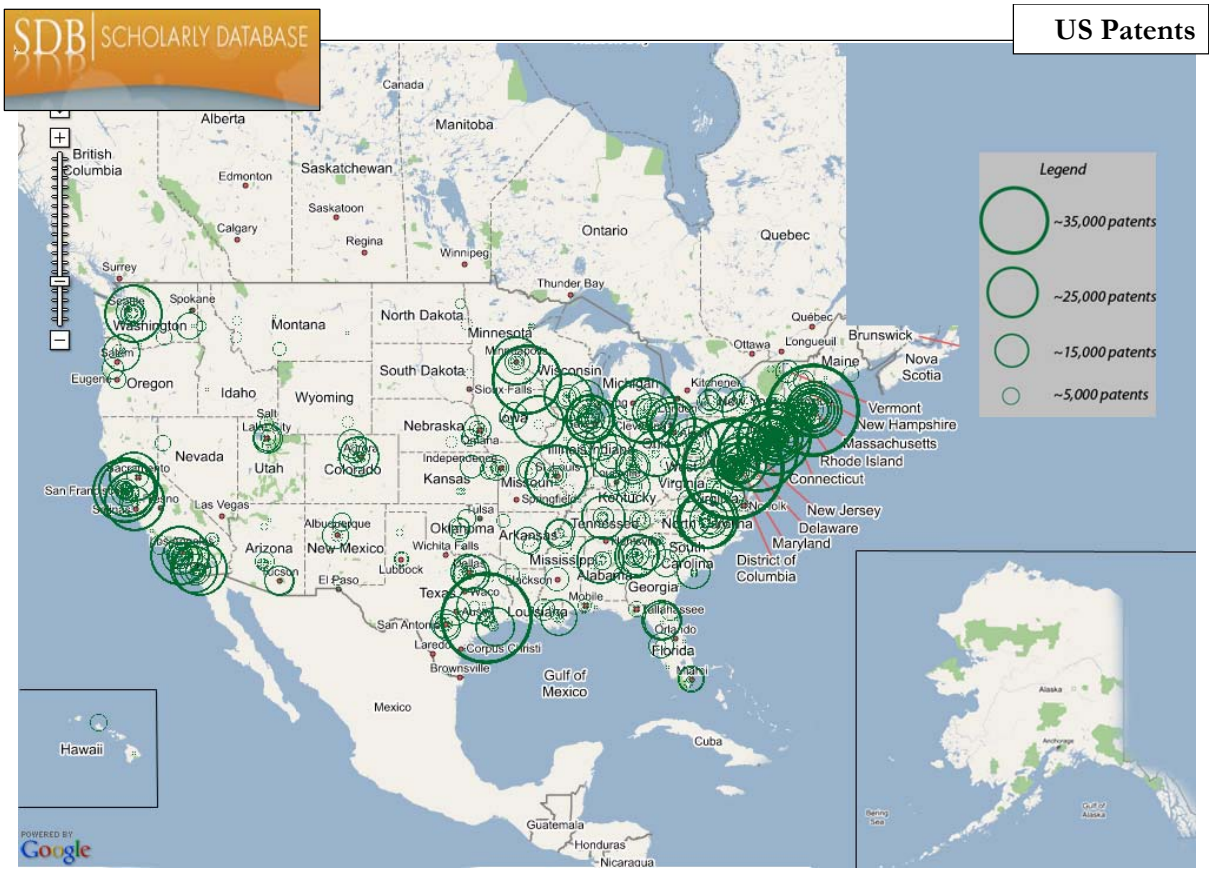


Medline Publications



NSF Grants

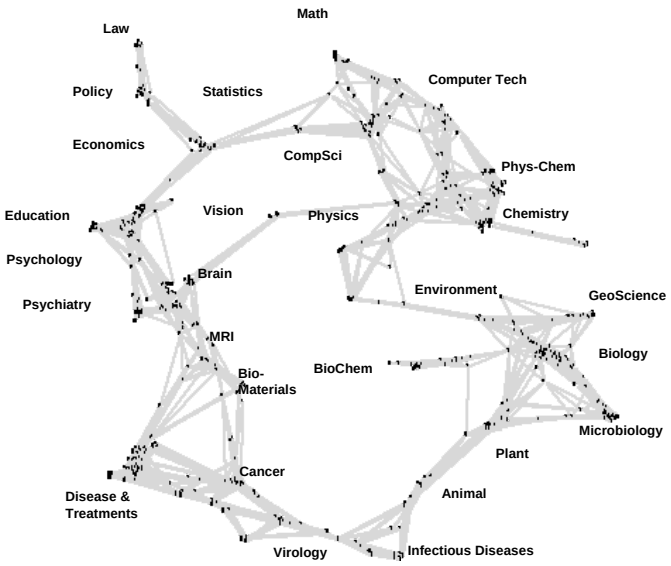




Latest 'Base Map' of Science

Kevin W. Boyack & Richard Klavans, unpublished work.

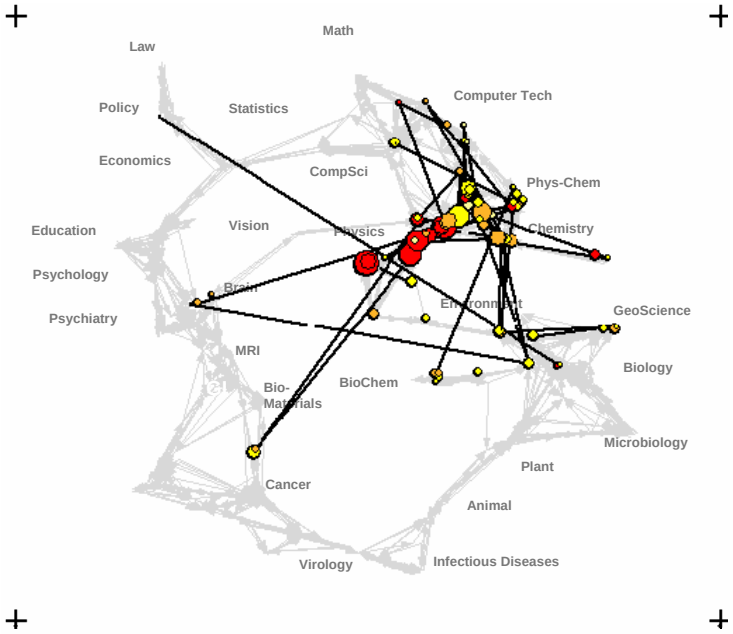
- Uses combined SCI/SSCI from 2002
  - 1.07M papers, 24.5M references, 7,300 journals
  - Bibliographic coupling of papers, aggregated to journals
- Initial ordination and clustering of journals gave 671 clusters
- Coupling counts were reaggregated at the journal cluster level to calculate the
  - (x,y)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

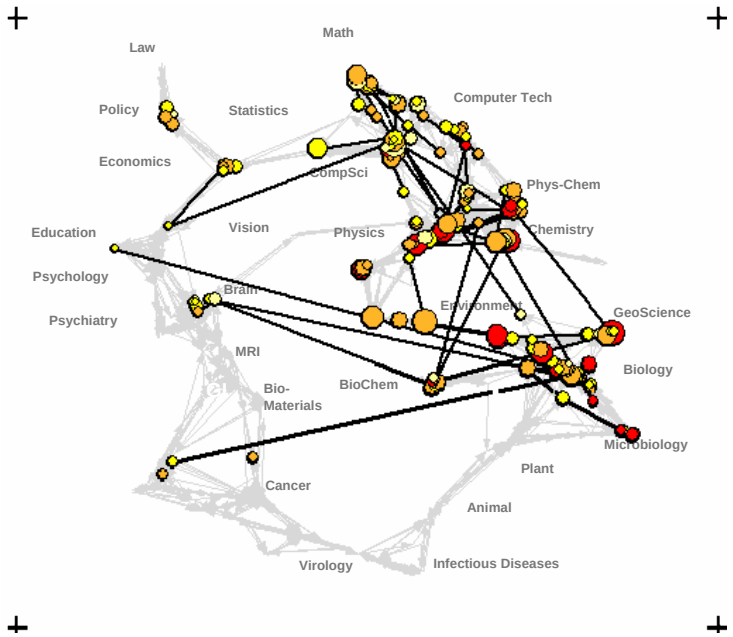
Funding patterns of the US Department of Energy (DOE)



Science map applications: Identifying core competency

Kevin W. Boyack & Richard Klavans, unpublished work.

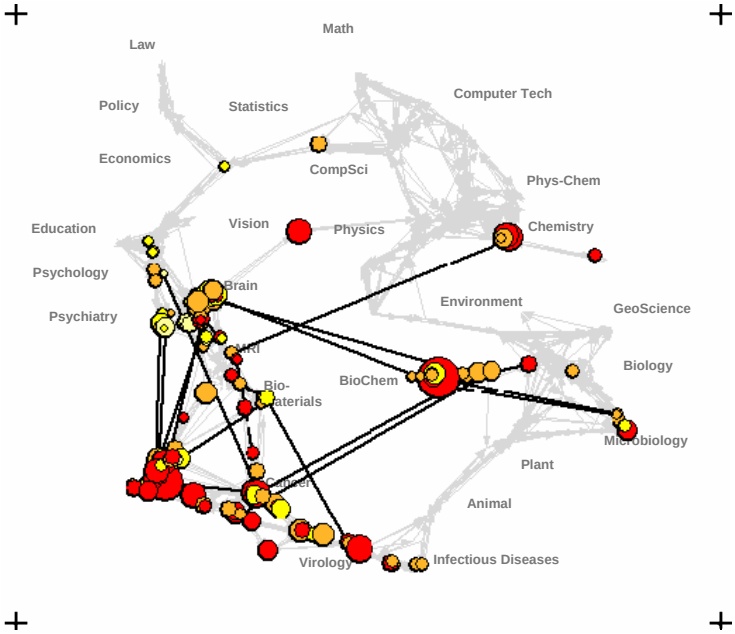
Funding Patterns of the National Science Foundation (NSF)



Science map applications: Identifying core competency

Kevin W. Boyack & Richard Klavans, unpublished work.

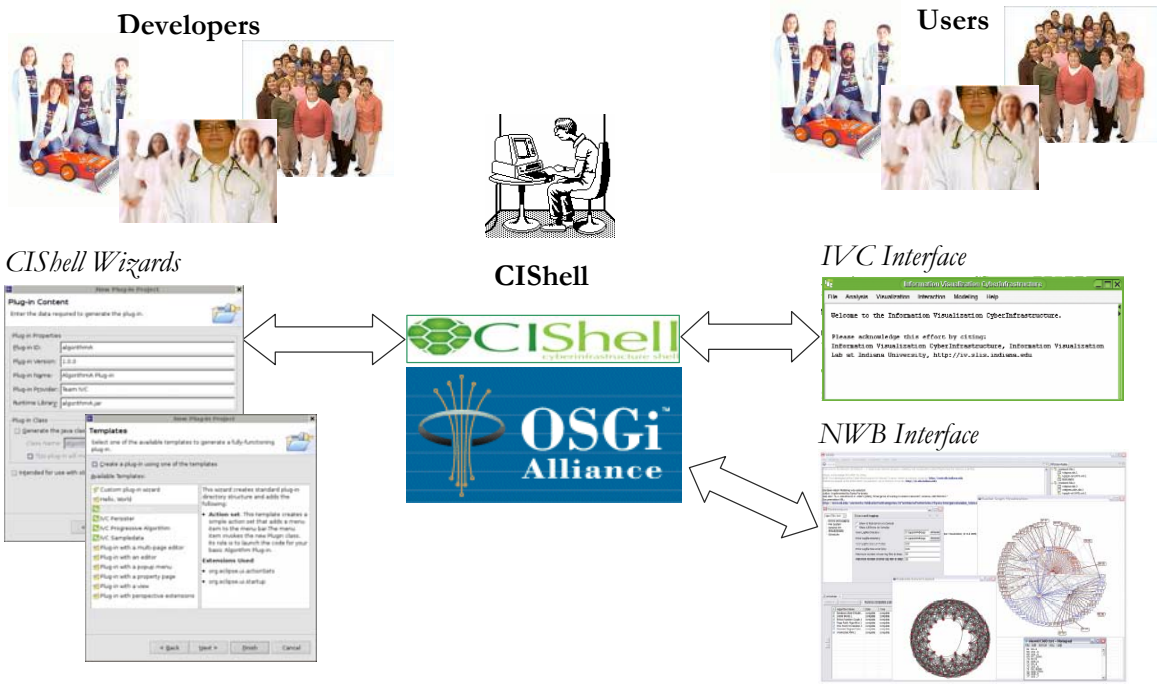
Funding Patterns of the National Institutes of Health (NIH)



Building Market Places not Cathedrals

- 'Software glue' has to interlink datasets and algorithms written in different languages using different data formats.
- The smaller the glue or 'CI Shell', the more likely it can be maintained.

CIShell – Serving Non-CS Algorithm Developers & Users



CIShell – Build on OSGi Industry Standard

CIShell is built upon the Open Services Gateway Initiative (OSGi) Framework.

OSGi (http://www.osgi.org) is

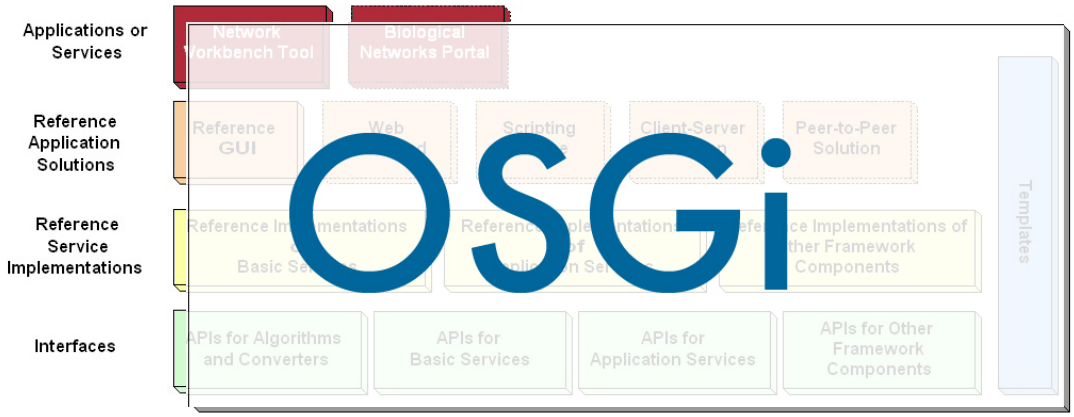
- A standardized, component oriented, computing environment for networked services.
- Successfully used in the industry from high-end servers to embedded mobile devices since 7 years.
- Alliance members include IBM (Eclipse), Sun, Intel, Oracle, Motorola, NEC and many others.
- Widely adopted in open source realm, especially since Eclipse 3.0 that uses OSGi R4 for its plugin model.

Advantages of Using OSGi

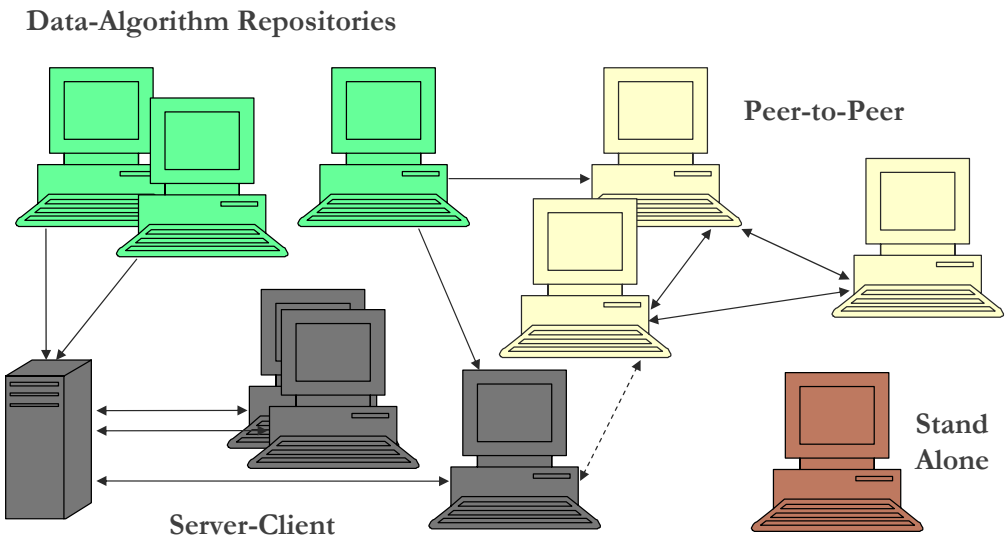
- Any CIShell algorithm is a service that can be used in any OSGi-framework based system.
- Using OSGi, running CIShells/tools can connected via RPC/RMI supporting peer-to-peer sharing of data, algorithms, and computing power.

Ideally, CIShell becomes a standard for creating OSGi Services for algorithms.

CIShell – Layer Cake



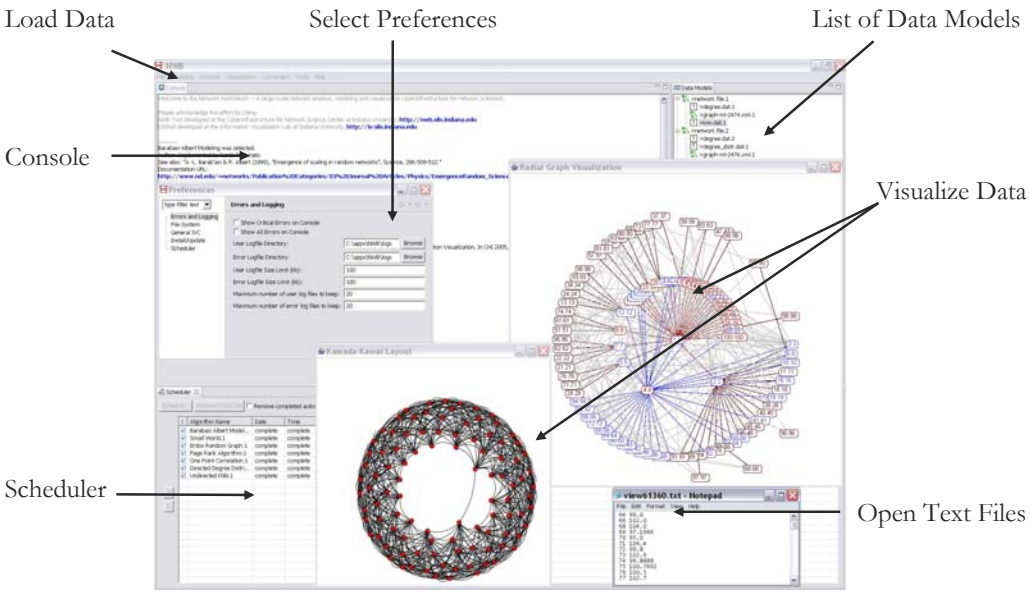
CIShell – Deployment



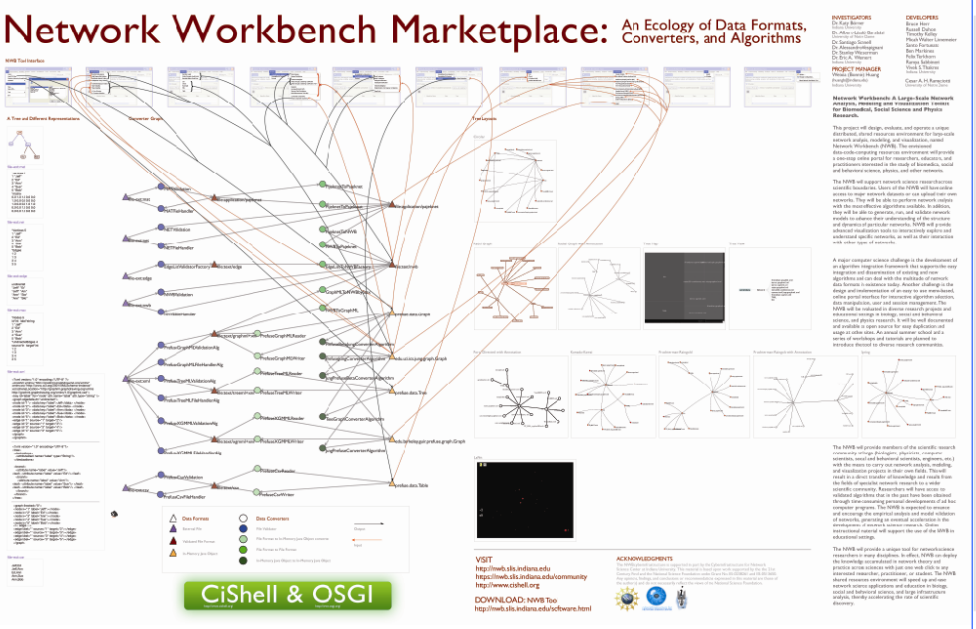
CIShell applications can be deployed as distributed data and algorithm repositories, stand alone applications, peer-to-peer architectures, and server-client architectures.

NWB Tool: Interface Elements

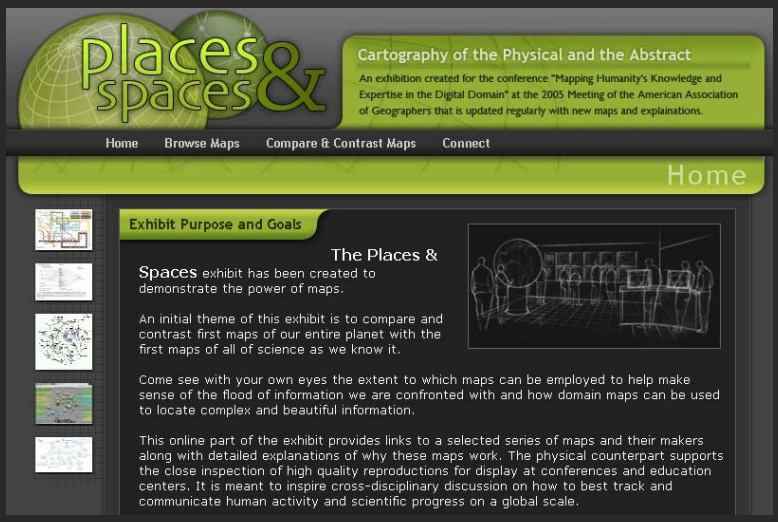
http://nwb.slis.indiana.edu



NWB Community Wiki



https://nwb.slis.indiana.edu/community/

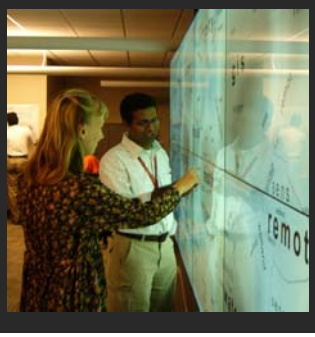


Places & Spaces: Mapping Science

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

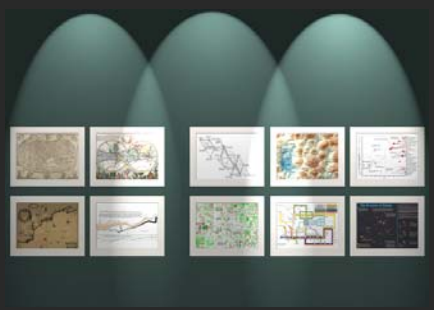
http://scimaps.org.

Exhibit Curators: Dr. Katy Börner & Elisha Hardy

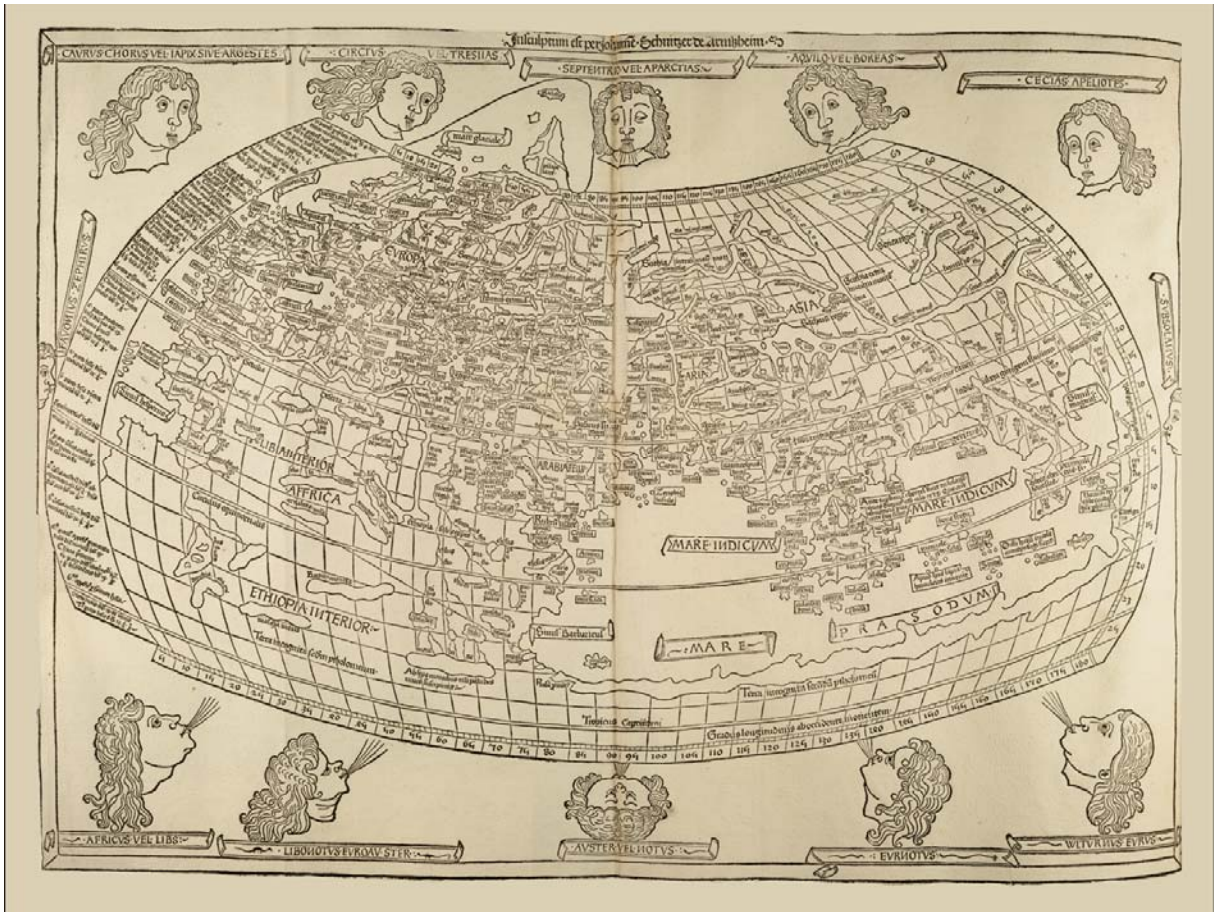


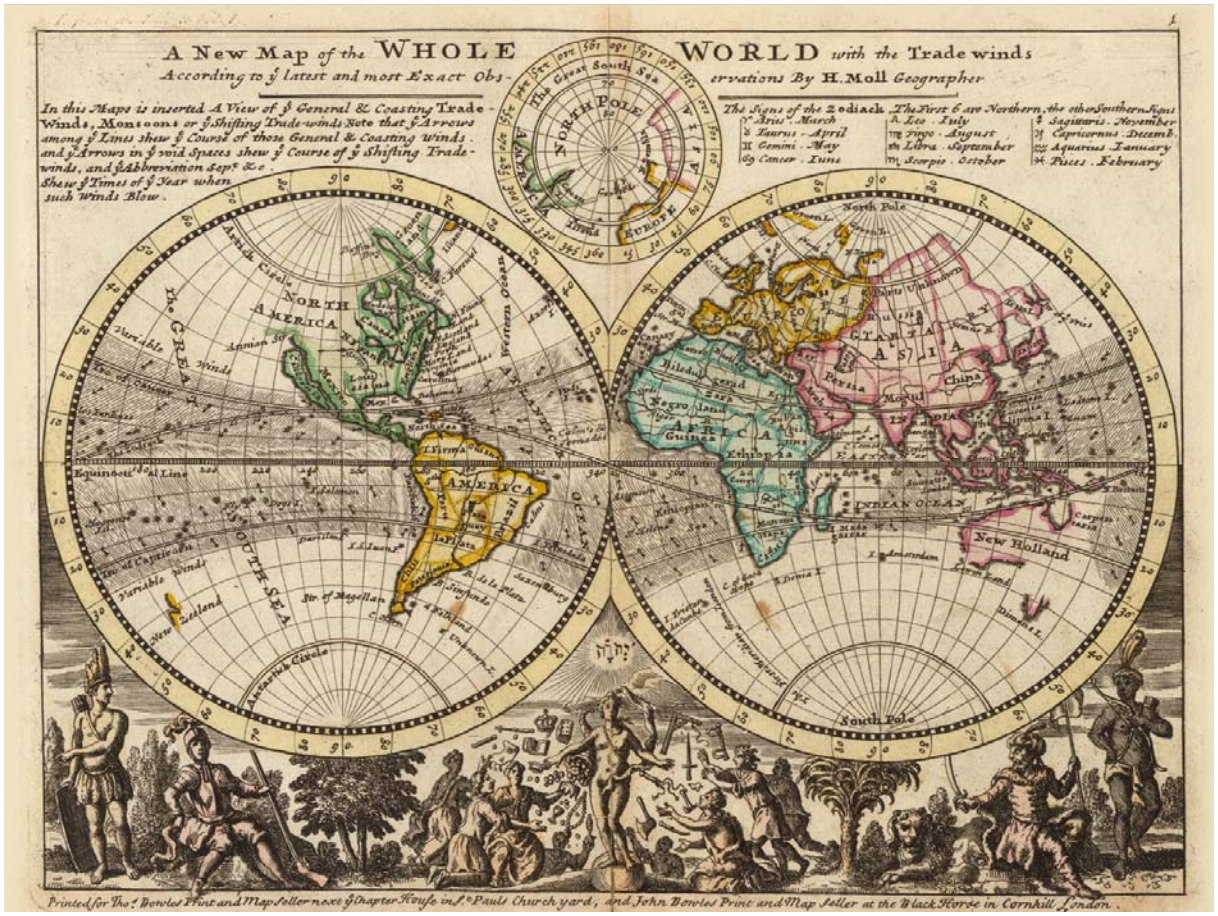
The Power of Maps

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



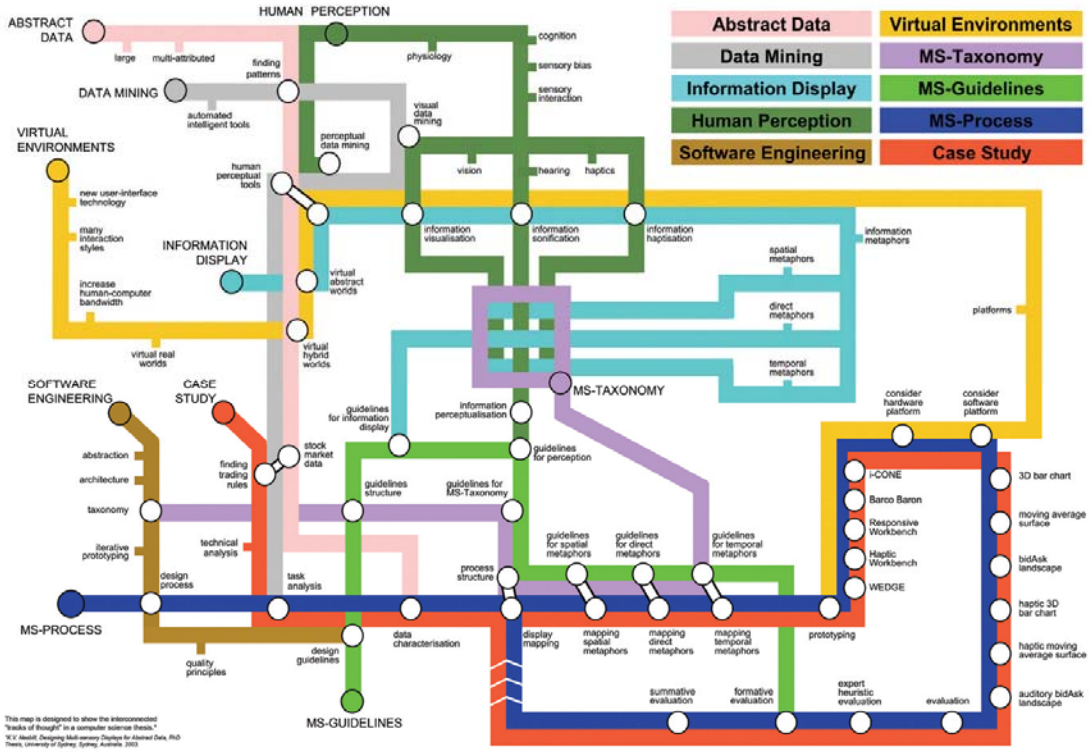
(1st Iteration of Places & Spaces Exhibit - 2005)

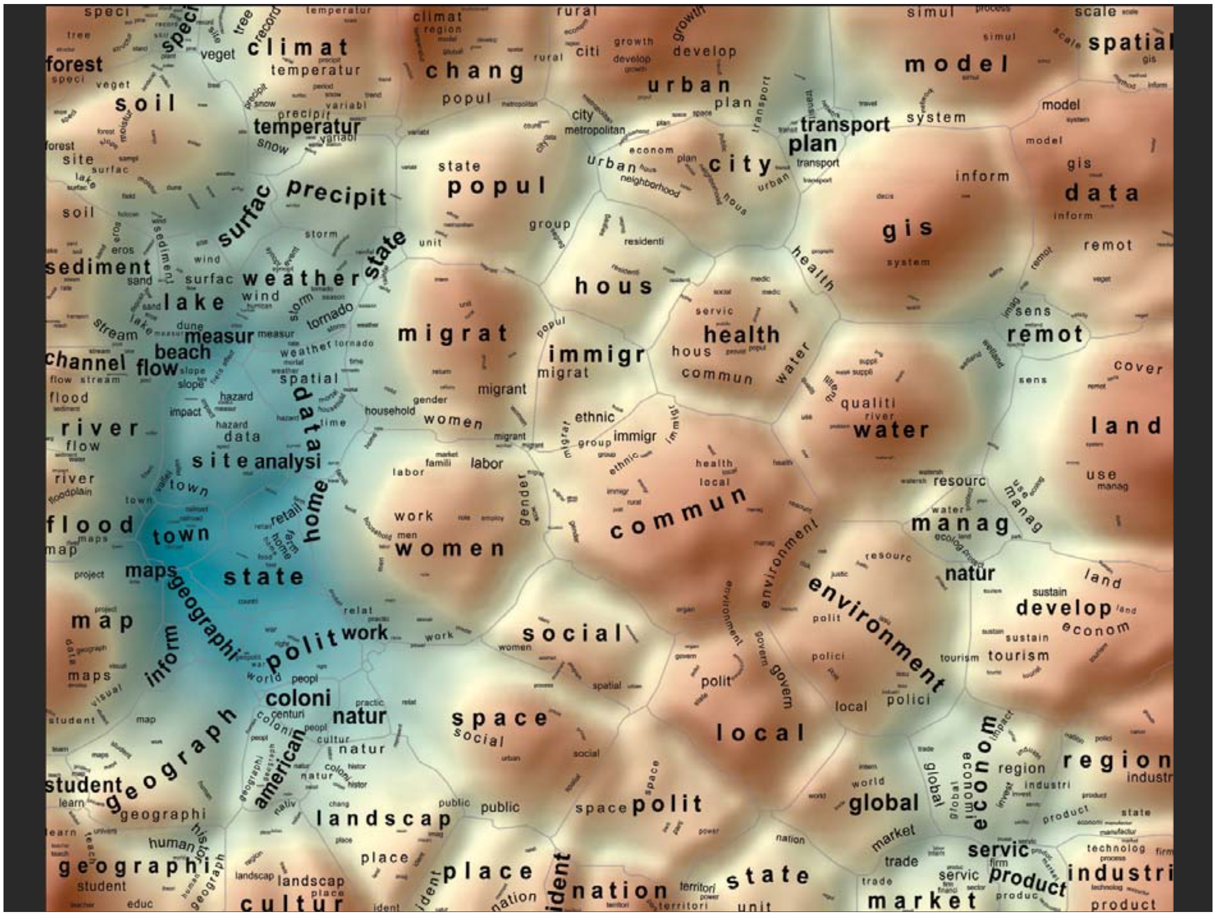

CAVRVS CHORVS VEL IAPIX SIVE ARGESTES
CIRCIVS VEL TRESIAS
SEPTENTRIO VEL APARCTIAS
AQVILO VEL BOREAS
CECIAS APELIOTES
EVROPA
ASIA
AFRICA
ARABIA FELIX
MARE INDICVM
MARE INDICVM
PRASODVM
MARE
ETHIOPIA INTERIOR
AFRICVS VEL LIBS
LIBONOTVS EVRONOTVS
AVSTER VEL NOTVS
EVRONOTVS
WITVRNVS EVRVS


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 mid Spaces shew ye Course of ye Shifting Trade-winds, and ye Abbreviation Sept.r &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
♈ Aries . March
♉ Taurus . April
♊ Gemini . May
♋ Cancer . June
♌ Leo . July
♍ Virgo . August
♎ Libra . September
♏ Scorpio . October
♐ Sagittarius . November
♑ Capricornus . Decemb.
♒ Aquarius . January
♓ Pisces . February
The Great South Sea
NORTH POLE
AMERICA
EUROPE
North Pole
NORTH AMERICA
SOUTH SEA
AFRICA
ASIA
INDIAN OCEAN
New Holland
South Pole
Printed for Tho.s Bowles Print and Map Seller next ye Chapter House in S.t Pauls Church yard, and John Bowles Print and Map Seller at the Black Horse in Cornhill, London.

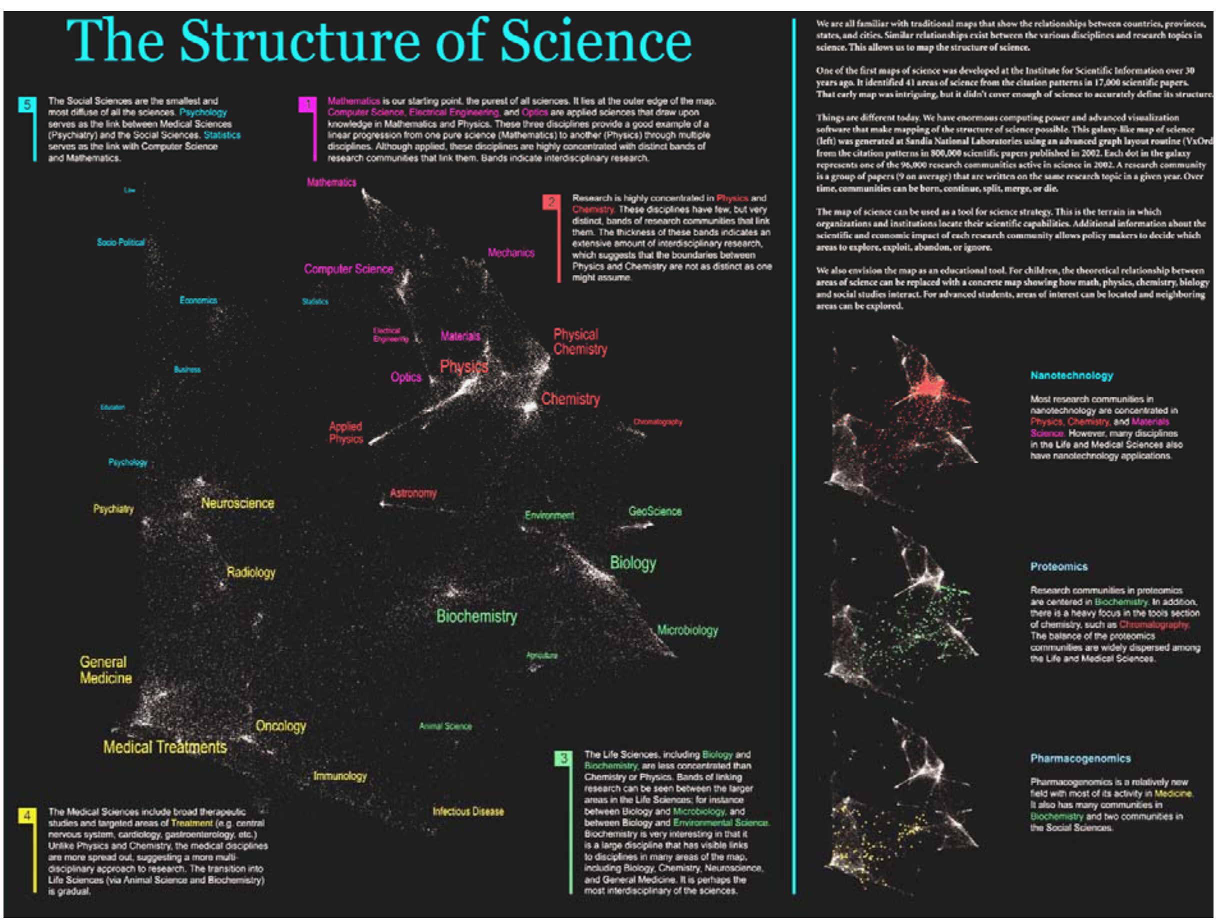
How would a map of science look?

What metaphors would work best?





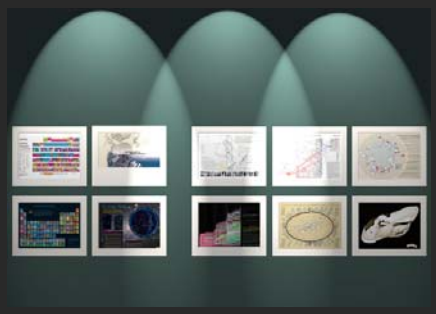
climat
chang
urban
model
spatial
forest
soil
temperatur
precipit
popul
city
transport
plan
data
gis
surfac
sediment
weather
state
migrat
hous
immigr
health
remot
lake
measur
beach
channel
flow
spatial
river
water
land
site analysi
women
commun
manag
flood
town
state
environment
natur
map
maps
geographi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



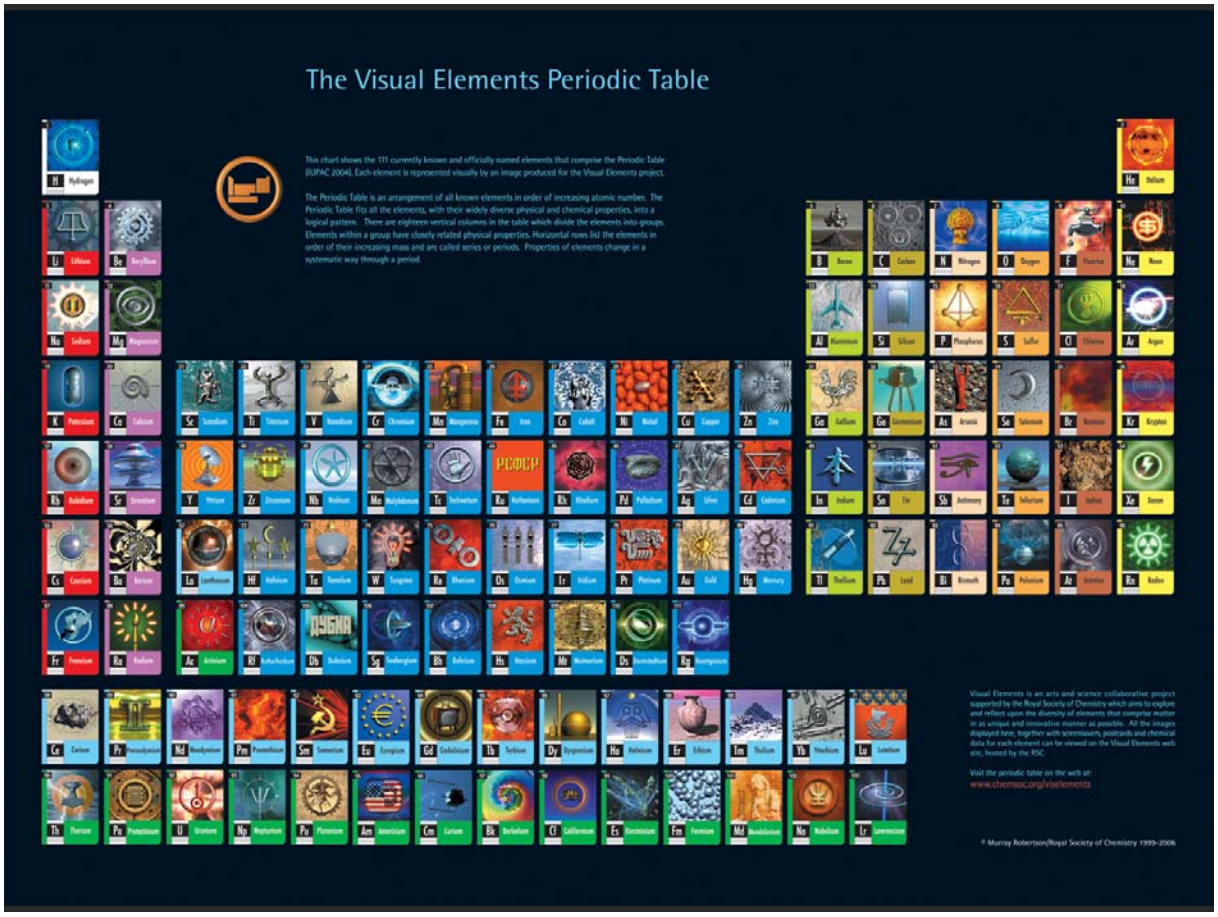
The Structure of Science
5 The Social Sciences are the smallest and most diffuse of all the sciences. Psychology serves as the link between Medical Sciences (Psychiatry) and the Social Sciences. Statistics serves as the link with Computer Science and Mathematics.
1 Mathematics is our starting point, 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
Education
Psychology
Psychiatry
Mathematics
Computer Science
Statistics
Mechanics
Electrical Engineering
Materials
Physics
Optics
Physical Chemistry
Chemistry
Chromatography
Applied Physics
Astronomy
Neuroscience
Environment
GeoScience
Biology
Radiology
Biochemistry
Microbiology
Agriculture
General Medicine
Oncology
Medical Treatments
Animal Science
Immunology
Infectious Disease
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.
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.
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.

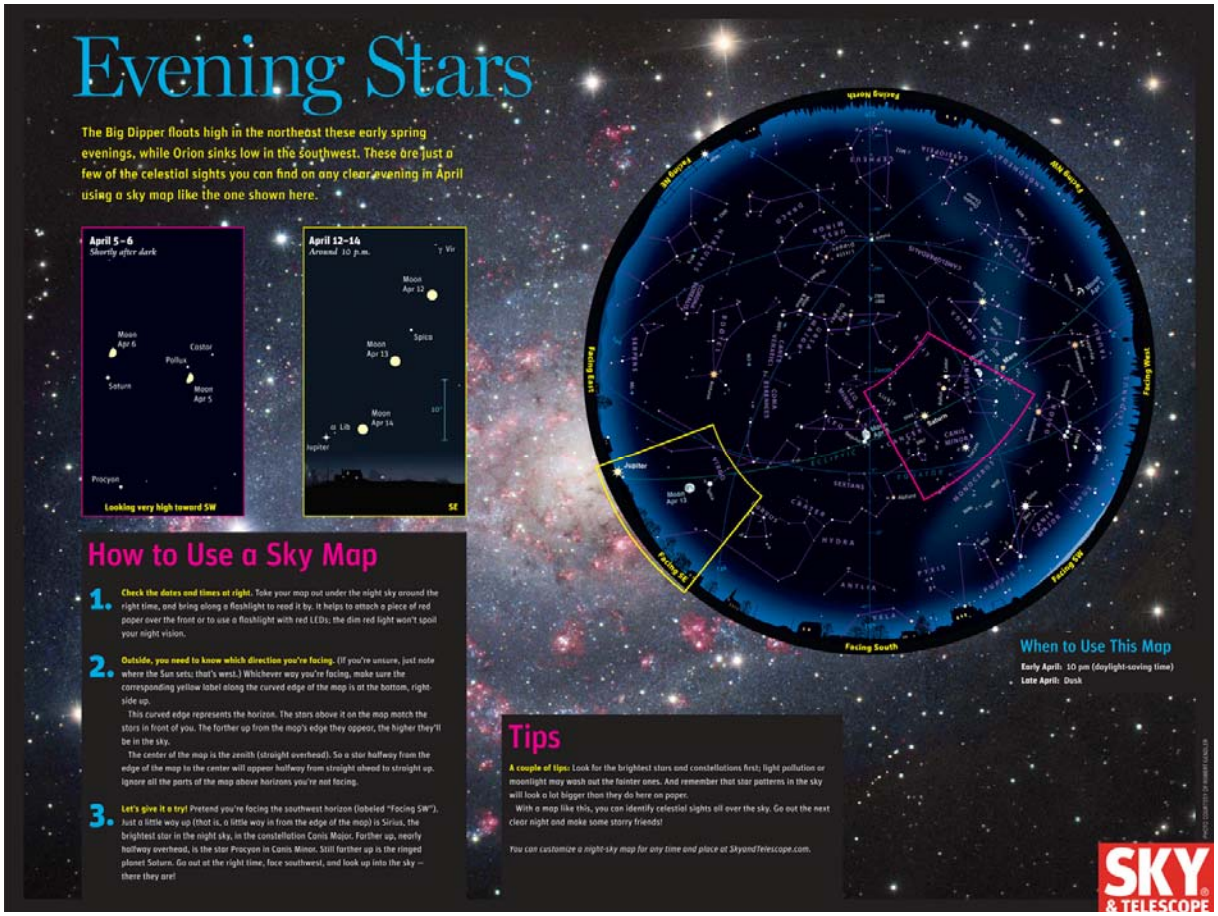
The Power of Reference Systems

Four Existing Reference Systems
VERSUS
Six Potential Reference Systems of Science



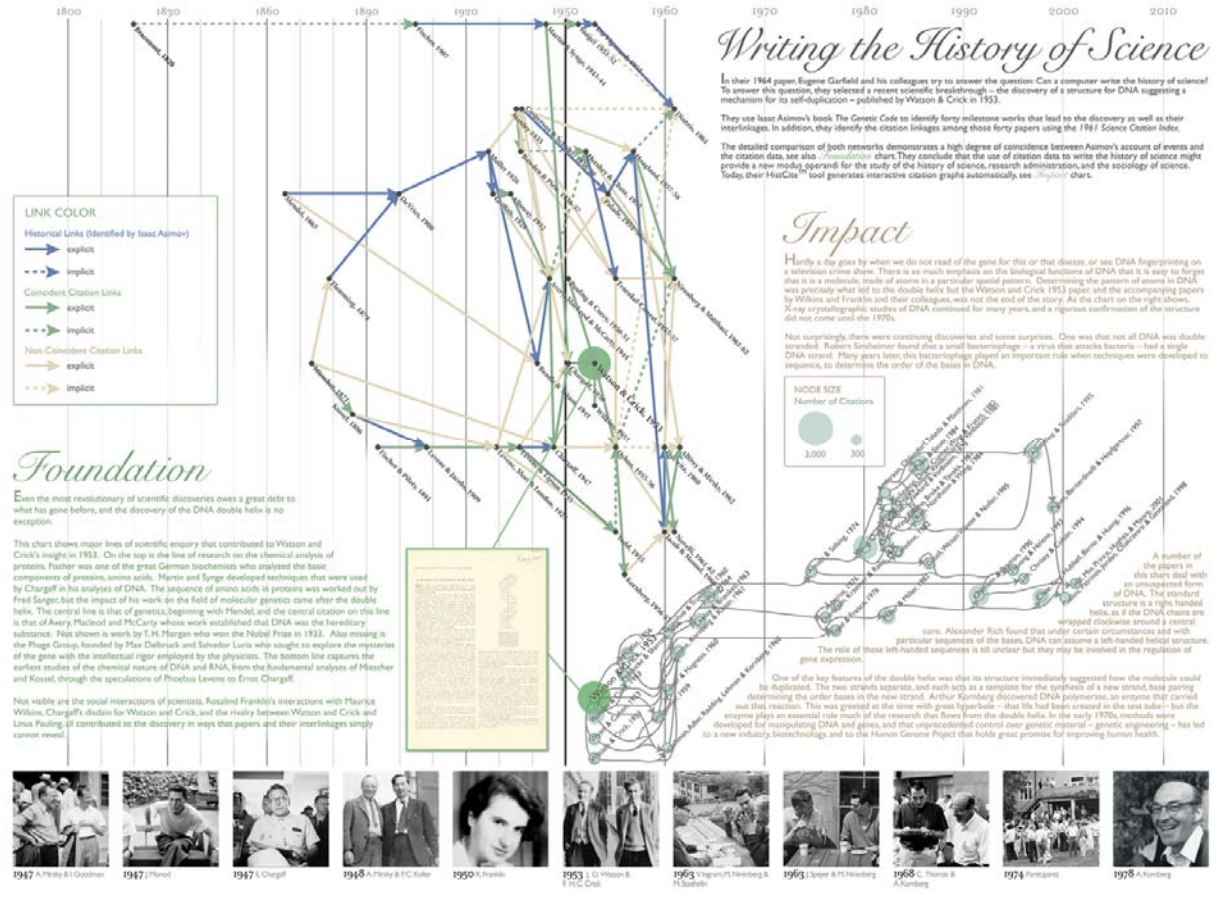
(2nd Iteration of Places & Spaces Exhibit - 2006)

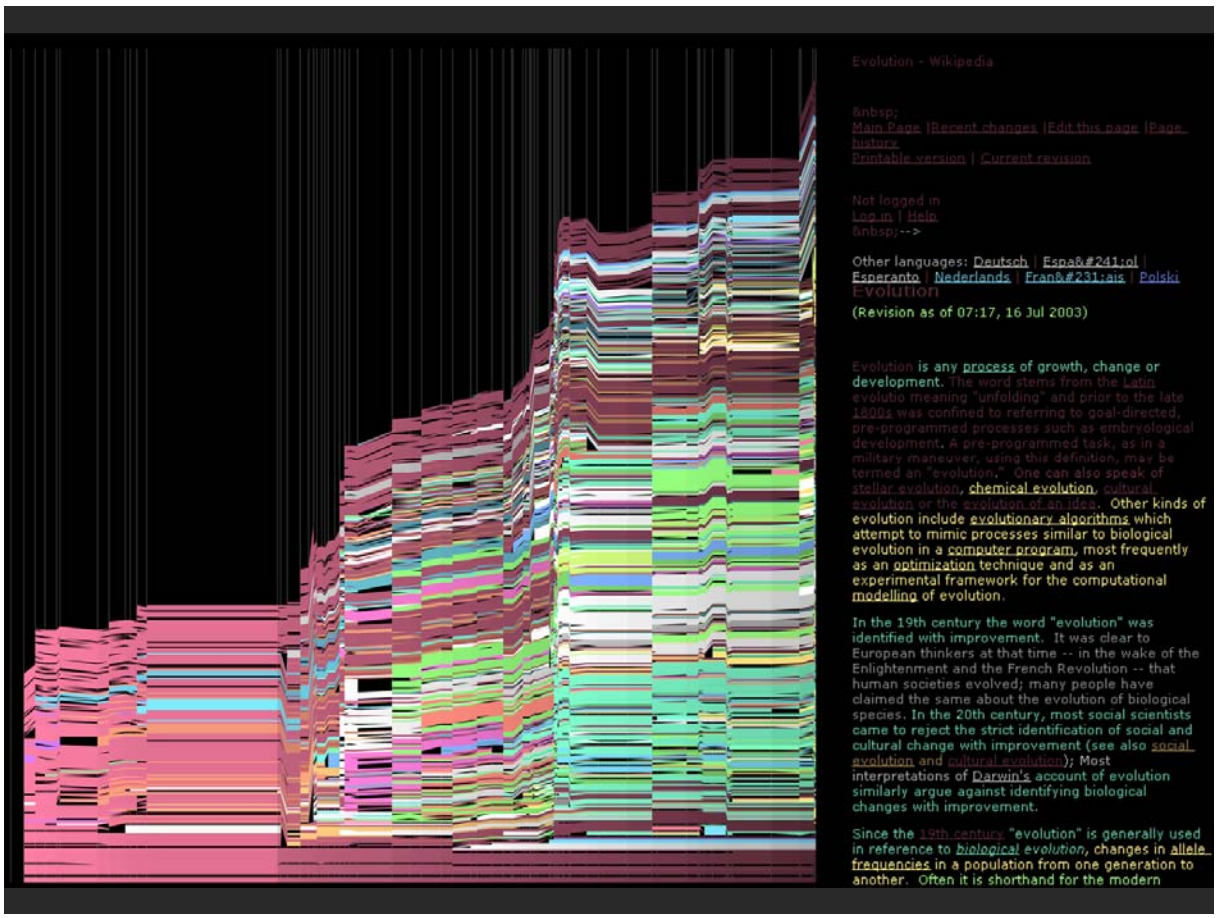




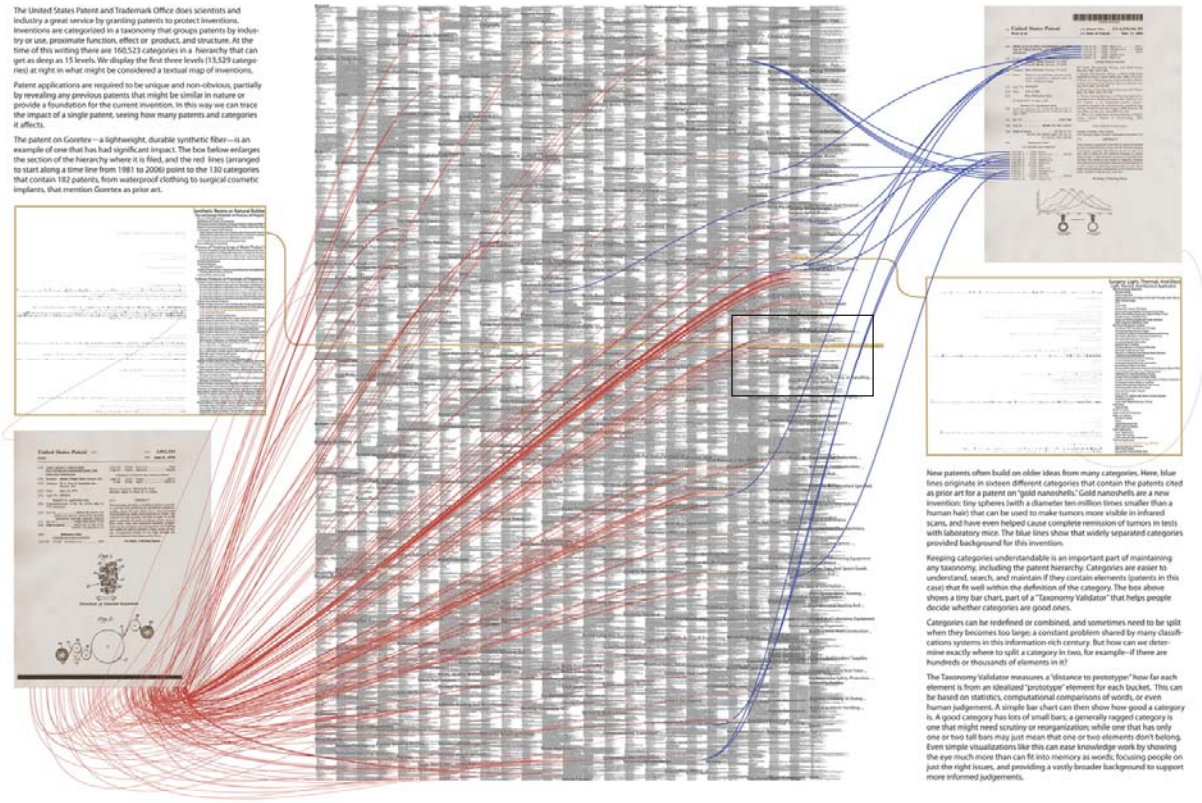
How would a reference system for all of science look?

What dimensions would it have?


Writing the History of Science
In their 1964 paper, Eugene Garfield and his colleagues try to answer the question: Can a computer write the history of science? To answer this question, they selected a recent scientific breakthrough – the discovery of a structure for DNA suggesting a mechanism for its self-duplication – published by Watson & Crick in 1953.
They use Isaac Asimov's book The Genetic Code to identify forty milestone works that lead to the discovery as well as their interlinkages. In addition, they identify the citation linkages among those forty papers using the 1961 Science Citation Index.
The detailed comparison of both networks demonstrates a high degree of coincidence between Asimov's account of events and the citation data, see also Hinduksh chart. They conclude that the use of citation data to write the history of science might provide a new modus operandi for the study of the history of science, research administration, and the sociology of science. Today, their HistCite™ tool generates interactive citation graphs automatically, see Hinduksh chart.
LINK COLOR
Historical Links (Identified by Isaac Asimov)
explicit
implicit
Coincident Citation Links
explicit
implicit
Non-Coincident Citation Links
explicit
implicit
Foundation
Even the most revolutionary of scientific discoveries owes a great debt to what has gone before, and the discovery of the DNA double helix is no exception.
This chart shows major lines of scientific enquiry that contributed to Watson and Crick's insight in 1953. On the top is the line of research on the chemical analysis of proteins. Fischer was one of the great German biochemists who analyzed the basic components of proteins, amino acids. Martin and Synge developed techniques that were used by Chargaff in his analyses of DNA. The sequence of amino acids in proteins was worked out by Fred Sanger, but the impact of his work on the field of molecular genetics came after the double helix. The central line is that of genetics, beginning with Mendel, and the central citation on this line is that of Avery, Macleod and McCarty whose work established that DNA was the hereditary substance. Not shown is work by T.H. Morgan who won the Nobel Prize in 1933. Also missing is the Phage Group, founded by Max Delbrück and Salvador Luria who sought to explore the mysteries of the gene with the intellectual rigor employed by the physicists. The bottom line captures the earliest studies of the chemical nature of DNA and RNA, from the fundamental analyses of Miescher and Kossel, through the speculations of Phoebus Levene to Ernst Chargaff.
Not visible are the social interactions of scientists. Rosalind Franklin's interactions with Maurice Wilkins, Chargaff's disdain for Watson and Crick, and the rivalry between Watson and Crick and Linus Pauling, all contributed to the discovery in ways that papers and their interlinkages simply cannot reveal.
Impact
Hardly a day goes by when we do not read of the gene for this or that disease, or see DNA fingerprinting on a television crime show. There is so much emphasis on the biological functions of DNA that it is easy to forget that it is a molecule, made of atoms in a particular spatial pattern. Determining the pattern of atoms in DNA was precisely what led to the double helix but the Watson and Crick 1953 paper, and the accompanying papers by Wilkins and Franklin and their colleagues, was not the end of the story. As the chart on the right shows, X-ray crystallographic studies of DNA continued for many years, and a rigorous confirmation of the structure did not come until the 1970s.
Not surprisingly, there were continuing discoveries and some surprises. One was that not all DNA was double stranded. Robert Sinsheimer found that a small bacteriophage – a virus that attacks bacteria – had a single DNA strand. Many years later, this bacteriophage played an important role when techniques were developed to sequence, to determine the order of the bases in DNA.
NODE SIZE
Number of Citations
3,000
300
A number of the papers in this chart deal with an unsuspected form of DNA. The standard structure is a right handed helix, as if the DNA chains are wrapped clockwise around a central core. Alexander Rich found that under certain circumstances and with particular sequences of the bases, DNA can assume a left-handed helical structure. The role of these left-handed sequences is still unclear but they may be involved in the regulation of gene expression.
One of the key features of the double helix was that its structure immediately suggested how the molecule could be duplicated. The two strands separate, and each acts as a template for the synthesis of a new strand, base pairing determining the order bases in the new strand. Arthur Kornberg discovered DNA polymerase, an enzyme that carried out that reaction. This was greeted at the time with great hyperbole – that life had been created in the test tube – but the enzyme plays an essential role much of the research that flows from the double helix. In the early 1970s, methods were developed for manipulating DNA and genes, and that unprecedented control over genetic material – genetic engineering – has led to a new industry, biotechnology, and to the Human Genome Project that holds great promise for improving human health.
1947 A. Hershey & I. Goodman
1947 J. Monod
1947 E. Chargaff
1948 A. Hershey & F.C. Koller
1950 R. Franklin
1953 J. D. Watson & F. H. C. Crick
1963 V. Ingram, M. Nirenberg & M. Stahlman
1963 J. Speyer & M. Nirenberg
1968 C. Thomas & A. Kornberg
1974 Participants
1978 A. Kornberg


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
Coating Processes
Rotary Shafts, Gudgeons, ...
Synthetic Resins or Natural ...
Data Processing: Structural ...
Measuring, Testing
Data Processing: Speech
Data Processing: Financial, ...
Foraminous Product Produced
Threaded, Headed Fastener ...
Jewelry, Symbolic ...
Amusement Devices
Electrical Computer: Arithmet
Electrical Compute's And ...

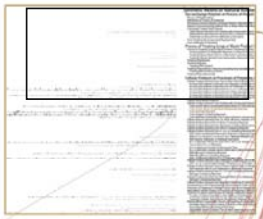
US Patent Hierarchy

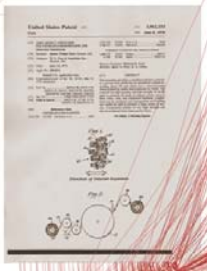
Impact

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.





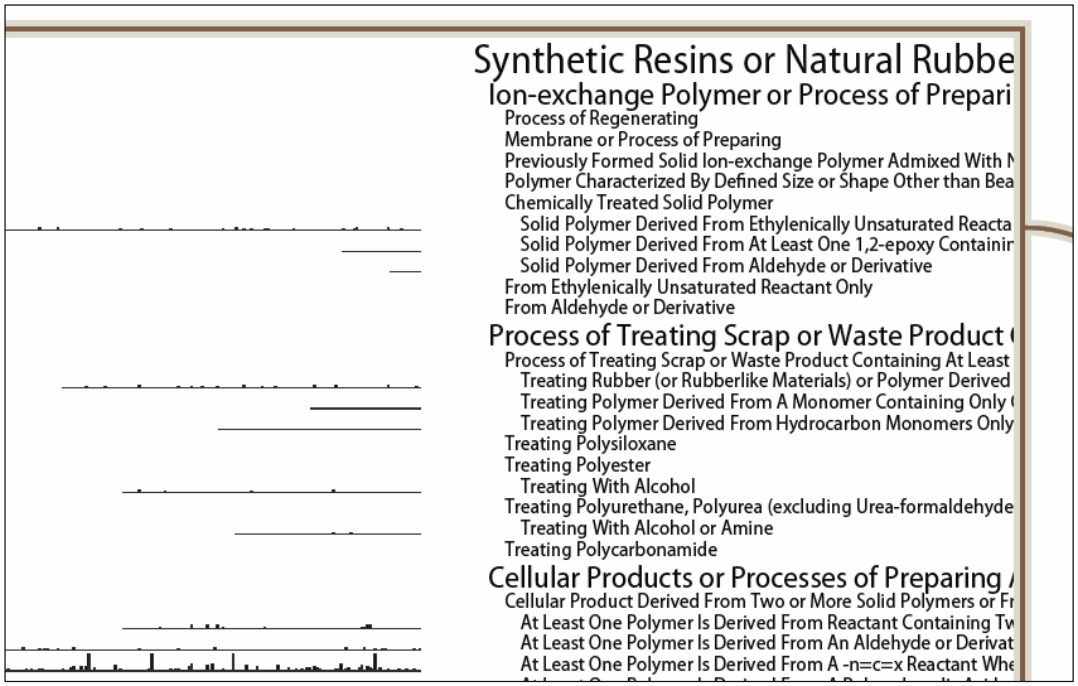
Prior Art

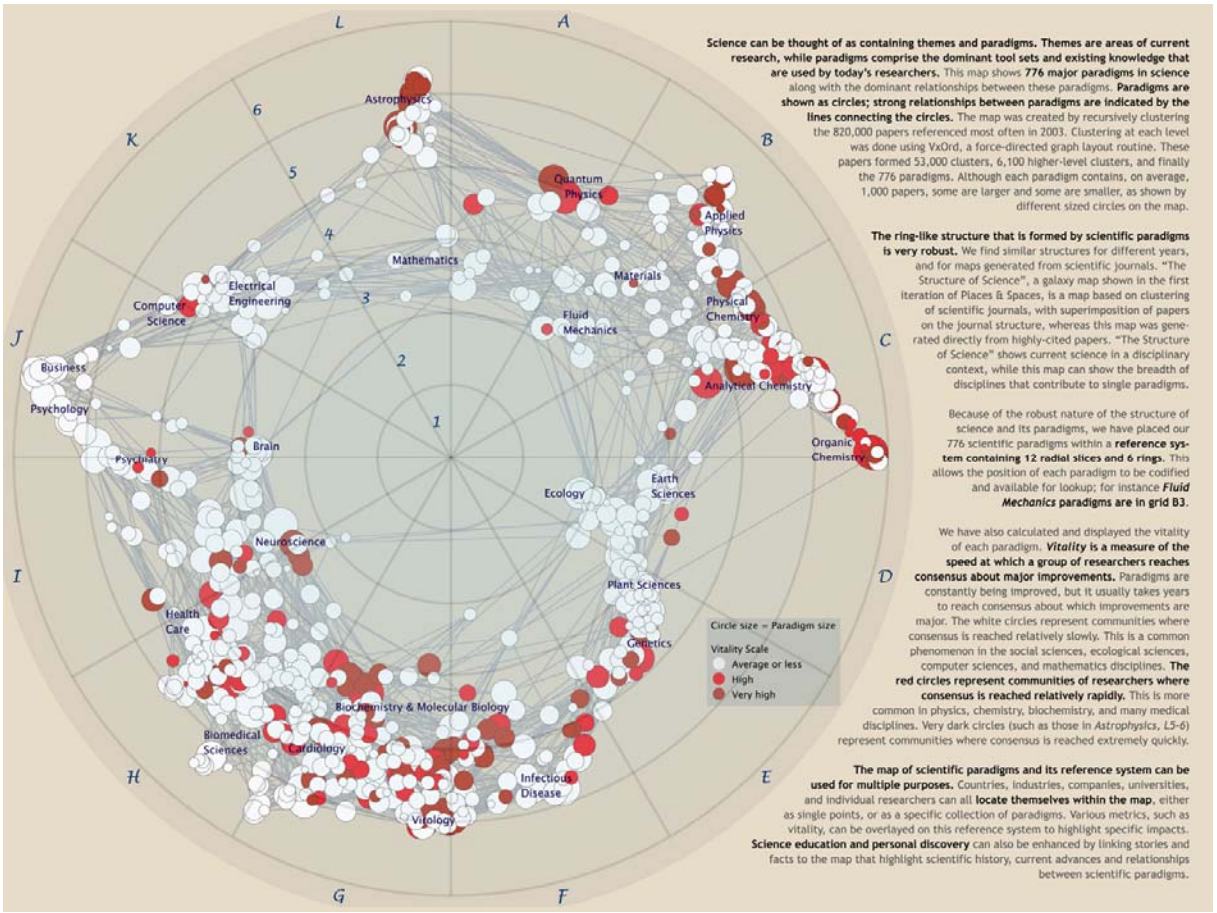
New patents often build on older ideas from many categories. Here, blue lines originate in sixteen different categories that contain the patents cited as prior art for a patent on "gold nanoshells." Gold nanoshells are a new invention: tiny spheres (with a diameter ten-million times smaller than a human 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.





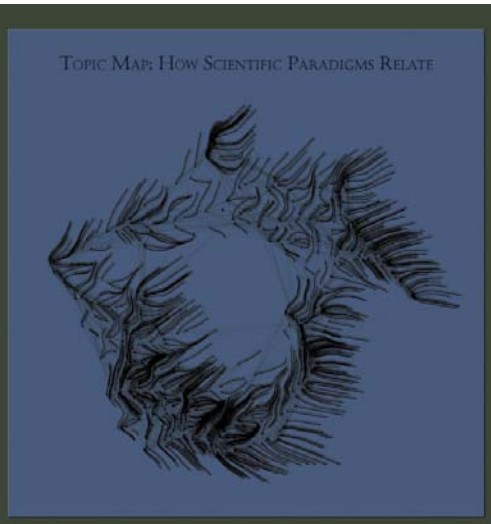
Science can be thought of as containing themes and paradigms. Themes are areas of current research, while paradigms comprise the dominant tool sets and existing knowledge that are used by today's researchers. This map shows **776 major paradigms in science** along with the dominant relationships between these paradigms. **Paradigms are shown as circles; strong relationships between paradigms are indicated by the lines connecting the circles.** The map was created by recursively clustering the 820,000 papers referenced most often in 2003. Clustering at each level was done using VxOrd, a force-directed graph layout routine. These papers formed 53,000 clusters, 6,100 higher-level clusters, and finally the 776 paradigms. Although each paradigm contains, on average, 1,000 papers, some are larger and some are smaller, as shown by different sized circles on the map.

The ring-like structure that is formed by scientific paradigms is very robust. We find similar structures for different years, and for maps generated from scientific journals. "The Structure of Science", a galaxy map shown in the first iteration of Places & Spaces, is a map based on clustering of scientific journals, with superimposition of papers on the journal structure, whereas this map was generated directly from highly-cited papers. "The Structure of Science" shows current science in a disciplinary context, while this map can show the breadth of disciplines that contribute to single paradigms.

Because of the robust nature of the structure of science and its paradigms, we have placed our 776 scientific paradigms within a **reference system containing 12 radial slices and 6 rings.** This allows the position of each paradigm to be codified and available for lookup; for instance ***Fluid Mechanics* paradigms are in grid B3.**

We have also calculated and displayed the vitality of each paradigm. ***Vitality* is a measure of the speed at which a group of researchers reaches consensus about major improvements.** Paradigms are constantly being improved, but it usually takes years to reach consensus about which improvements are major. The white circles represent communities where consensus is reached relatively slowly. This is a common phenomenon in the social sciences, ecological sciences, computer sciences, and mathematics disciplines. **The red circles represent communities of researchers where consensus is reached relatively rapidly.** This is more common in physics, chemistry, biochemistry, and many medical disciplines. Very dark circles (such as those in *Astrophysics, L5-6*) represent communities where consensus is reached extremely quickly.

The map of scientific paradigms and its reference system can be used for multiple purposes. Countries, industries, companies, universities, and individual researchers can all **locate themselves within the map**, either as single points, or as a specific collection of paradigms. Various metrics, such as vitality, can be overlayed on this reference system to highlight specific impacts. **Science education and personal discovery** can also be enhanced by linking stories and facts to the map that highlight scientific history, current advances and relationships between scientific paradigms.



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

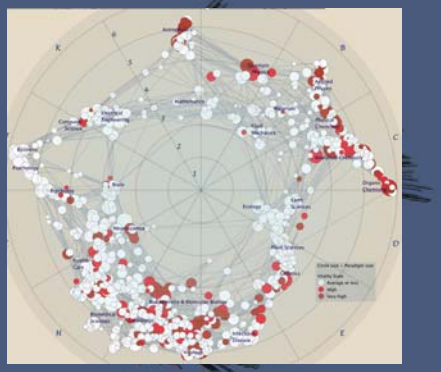
This overlay shows the distribution of nanotechnology within the paradigms of science. The majority of current work in nanotechnology takes places in physics, chemistry, and materials science, at the upper right portion of the map. However, an increasing amount of nanotechnology is being applied in the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

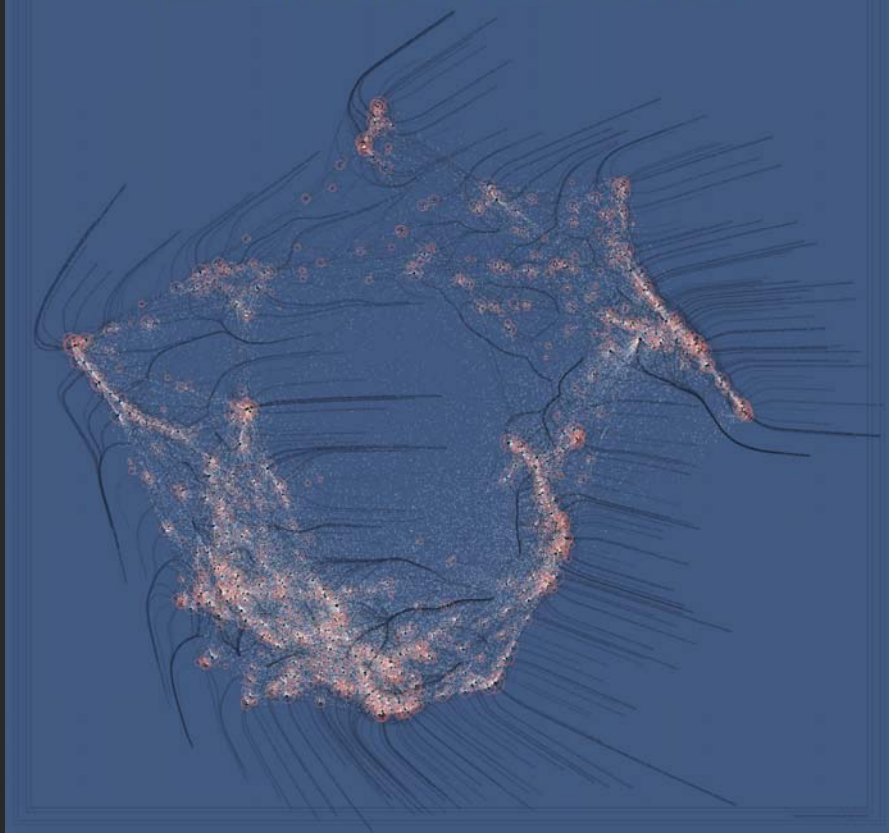
This overlay shows the distribution of nanotechnology within the paradigms of science. The majority of current work in nanotechnology takes places in physics, chemistry, and materials science, at the upper right portion of the map. However, an increasing amount of nanotechnology is being applied in the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

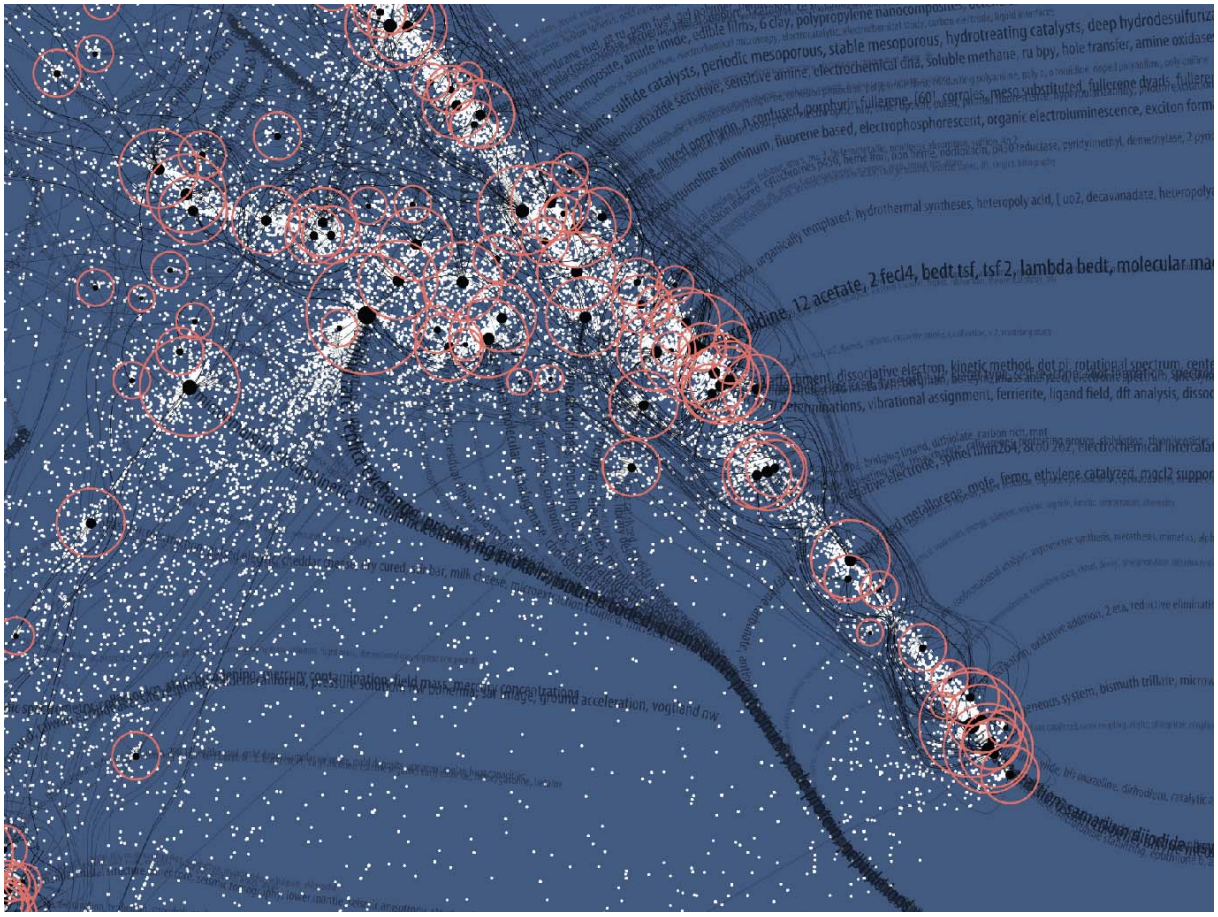
All Topics *Sweep through all 776 scientific paradigms*	**Nanotechnology** *Science on the tiny scale of molecules*	**Francis H. C. CRICK** *Co-discovered DNA's double helix*	**Albert EINSTEIN** *Revitalized physics with Relativity theories*	**Michael E. FISHER** *Models critical phase transitions of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

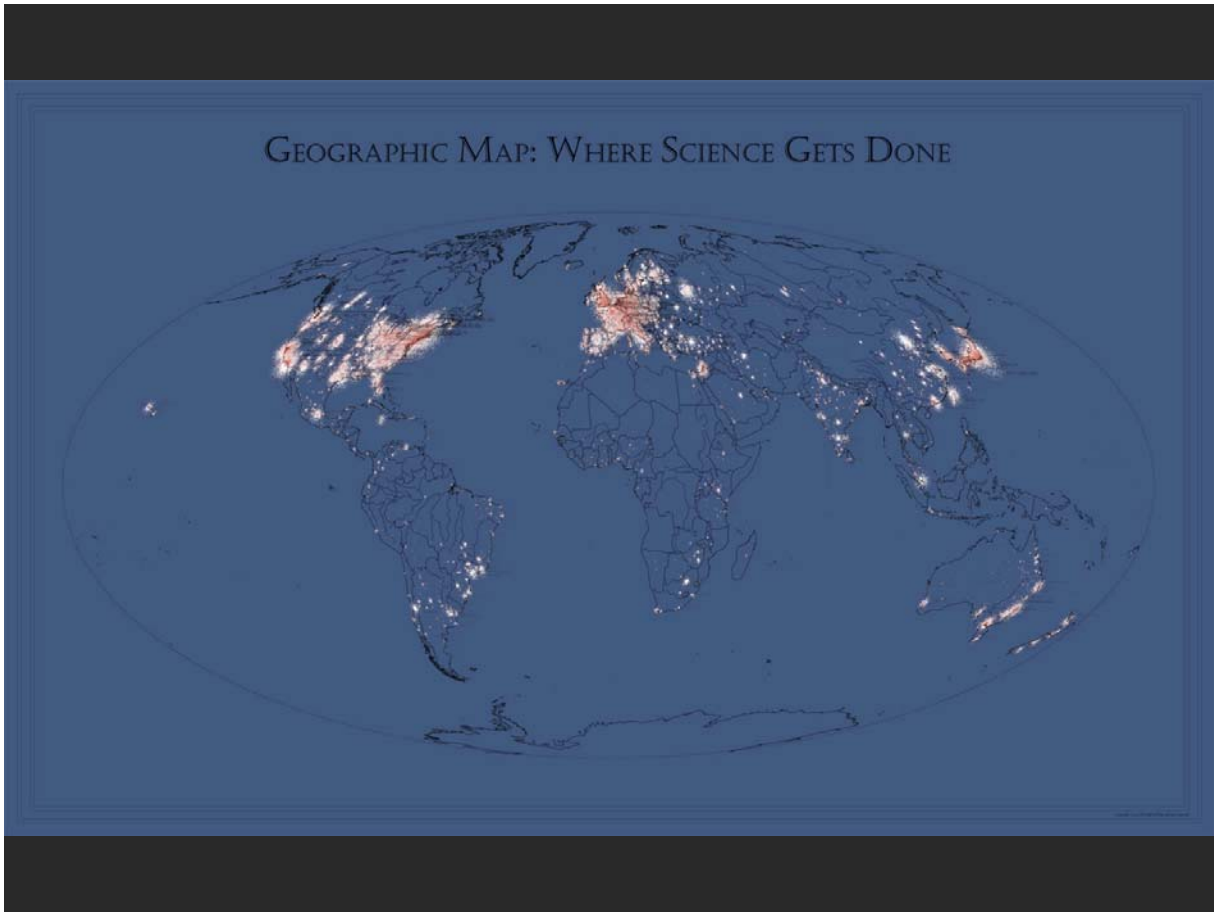
We sweep slowly through adjoining related topics, lighting up the places in the world that study each topic. You may select a subset of the topics that deal with these three interesting subjects by touching it.

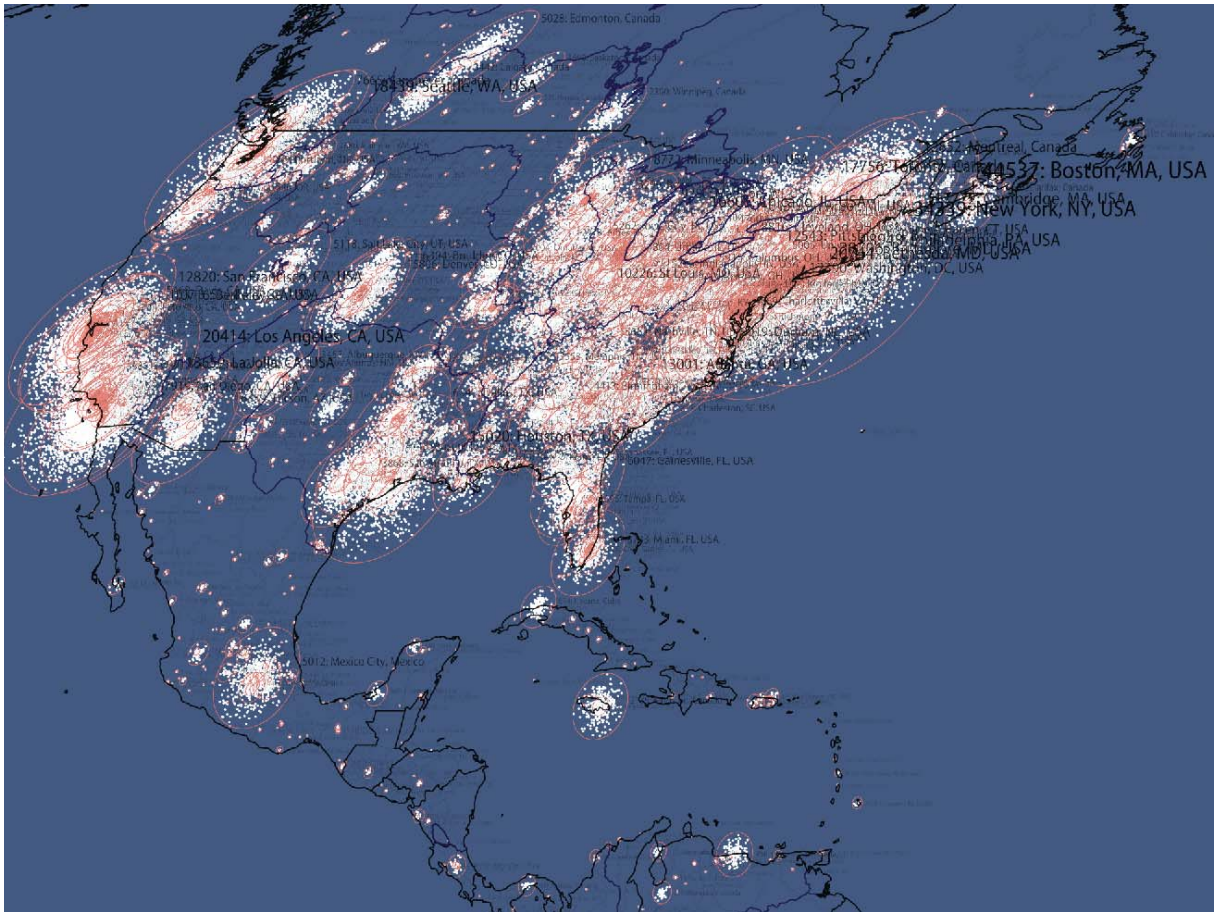
A single person's spreading influence is shown as a series of four snapshots. First, we light only topics and places relating to that person's papers—papers that are still highly cited today. The second lights everything that cites that original work. Note that this first-generation impact extends to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

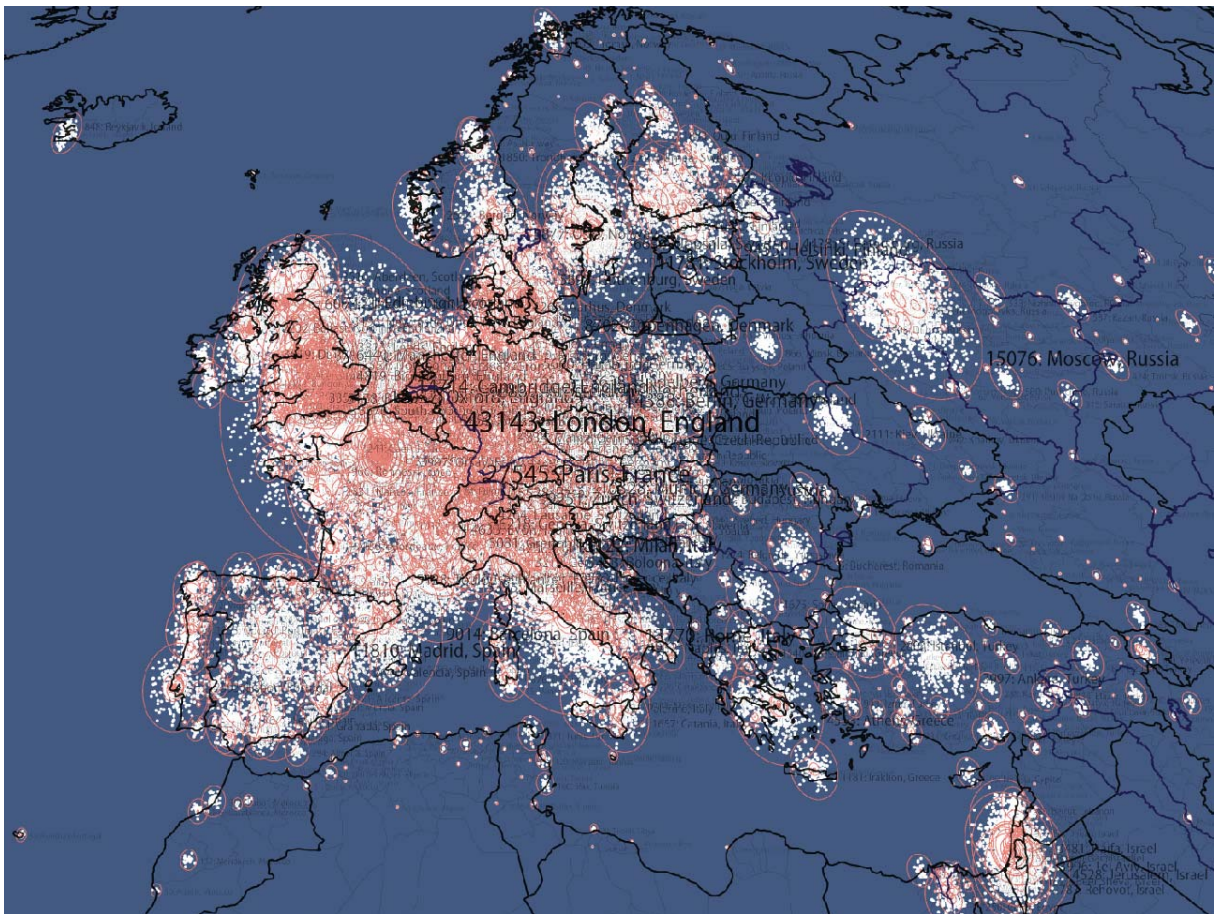
Topic Map: How Scientific Paradigms Relate

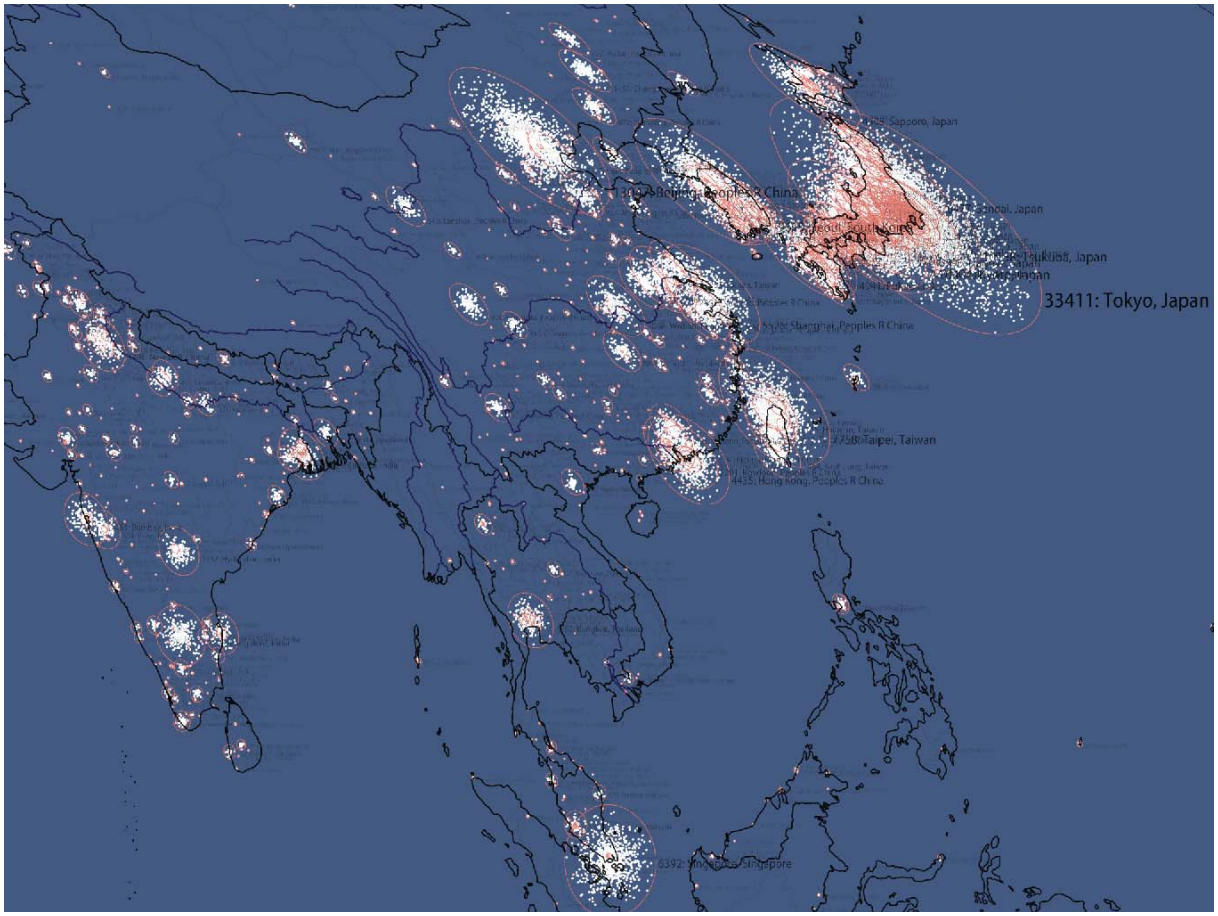


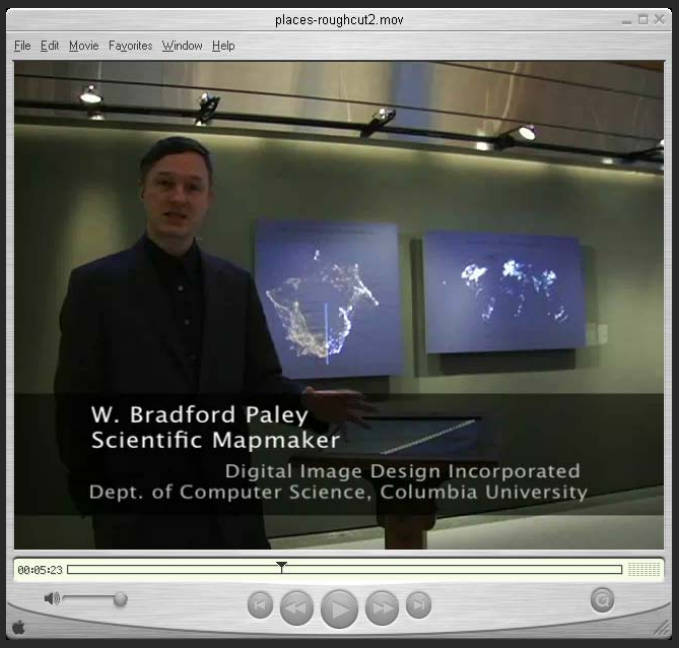



Geographic Map: Where Science Gets Done


5028: Edmonton, Canada
18439: Seattle, WA, USA
4537: Boston, MA, USA
New York, NY, USA
20414: Los Angeles, CA, USA
5012: Mexico City, Mexico


15076: Moscow, Russia
4314: London, England
27545: Paris, France
1810: Madrid, Spain



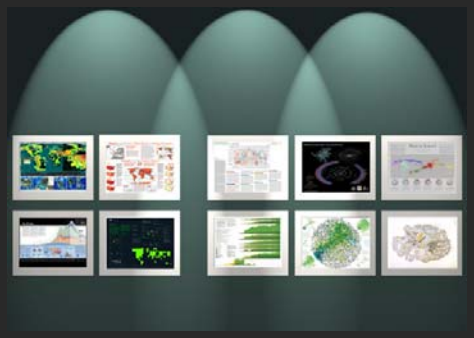


Illuminated Diagram Display

http://www.youtube.com/watch?v=bXABcOABG4E

The Power of Forecasts

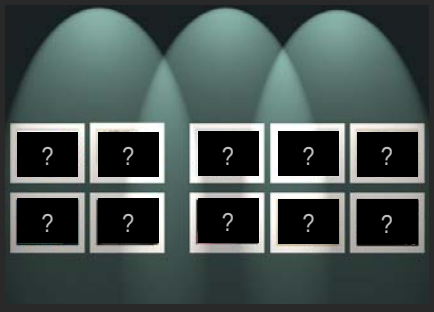
Four Existing Forecasts
VERSUS
Six Potential Science 'Weather' Forecasts



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

Science Maps for Economic Decision Making

Four Existing Maps
VERSUS
Six Science Maps



(4th Iteration of Places & Spaces Exhibit - 2008)

Science Maps in Action

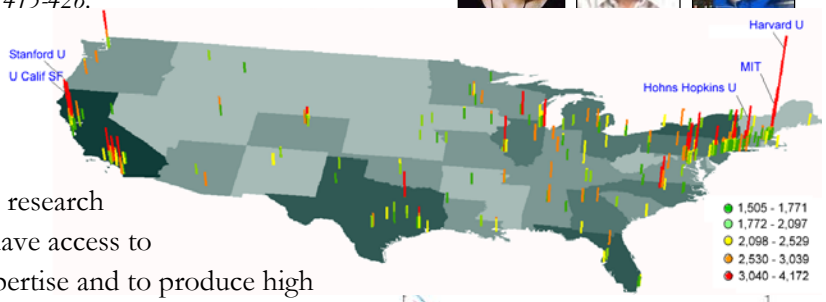
Spatio-Temporal Information Production and Consumption of Major U.S. Research Institutions

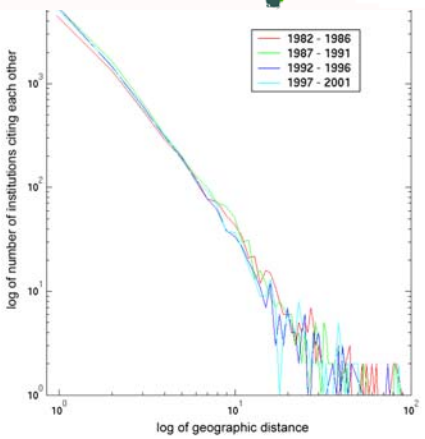


Börner, Katy, Penumarthy, Shashikant, Meiss, Mark and Ke, Weimao. (2006) Mapping the Diffusion of Scholarly Knowledge Among Major U.S. Research Institutions. Scientometrics. 68(3), pp. 415-426.

Research questions:

1. Does space still matter in the Internet age?
2. Does one still have to study and work at major research institutions in order to have access to high quality data and expertise and to produce high quality research?
3. Does the Internet lead to more global citation patterns, i.e., more citation links between papers produced at geographically distant research instructions?





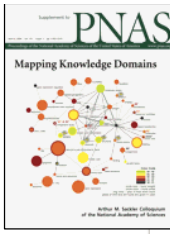
Contributions:

- Answer to Qs 1 + 2 is YES.
- Answer to Qs 3 is NO.
- Novel approach to analyzing the dual role of institutions as information producers and consumers and to study and visualize the diffusion of information among them.

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.

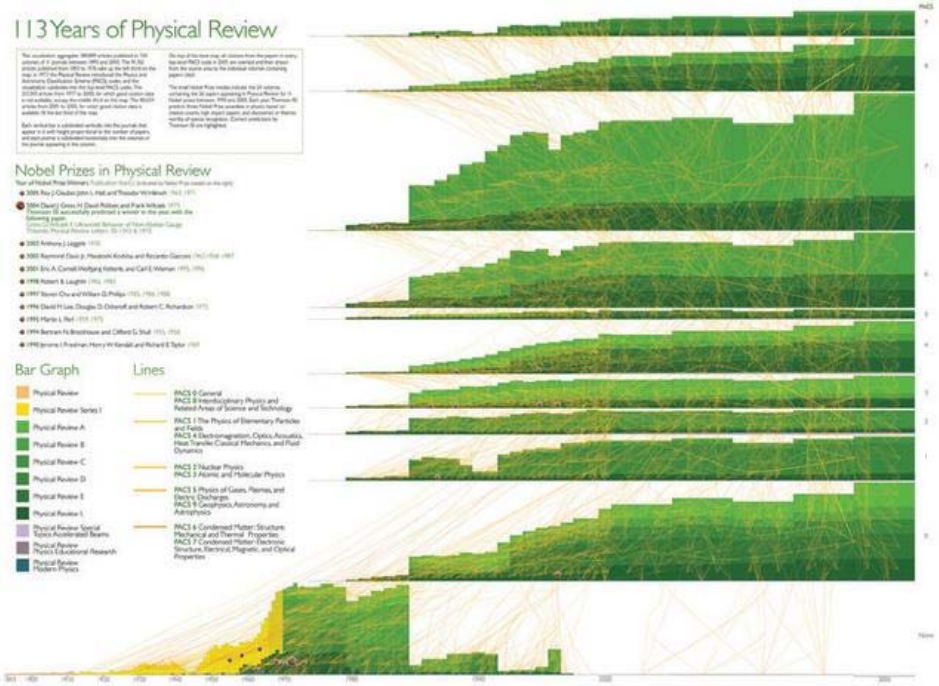
Mane & Börner. (2004) PNAS, 101(Suppl. 1): 5287-5290.



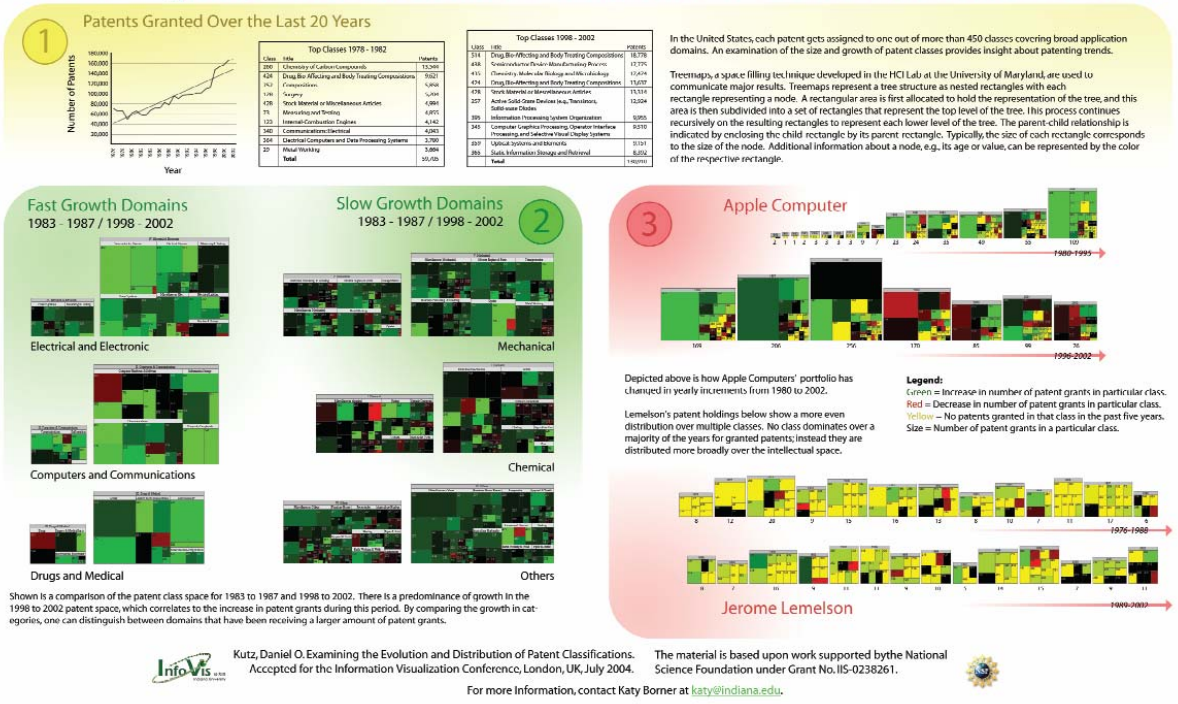
113 Years of Physical Review

http://scimaps.org/dev/map_detail.php?map_id=171

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



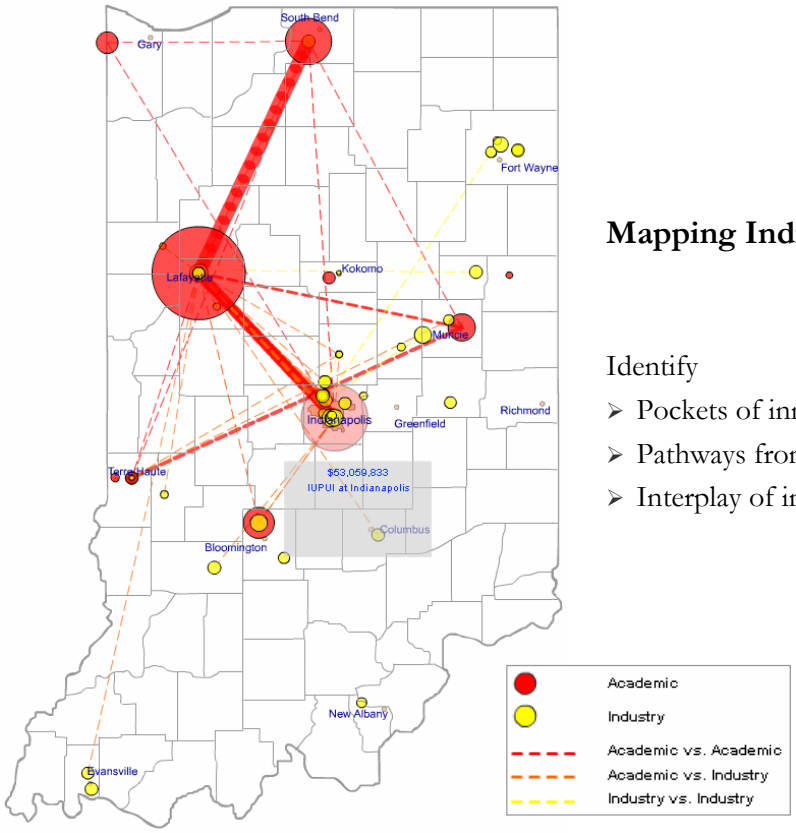
Examining the Evolution and Distribution of Patent Classifications



Kutz, Daniel O. Examining the Evolution and Distribution of Patent Classifications. Accepted for the Information Visualization Conference, London, UK, July 2004.

The material is based upon work supported bythe National Science Foundation under Grant No. IIS-0238261.

For more Information, contact Katy Borner at katy@indiana.edu.



Mapping Indiana's Intellectual Space

Identify

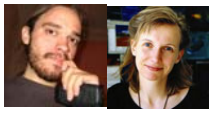
- Pockets of innovation
- Pathways from ideas to products
- Interplay of industry and academia

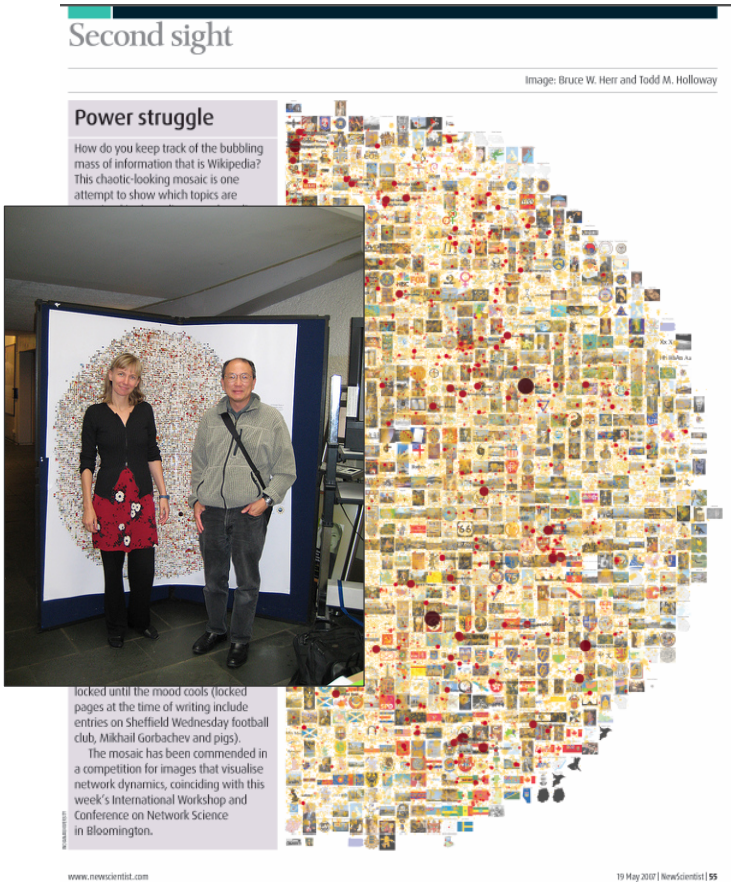
Wikipedian Activity

Studying large scale social networks such as Wikipedia

Vizzards 2007 Entry

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





Second sight

Image: Bruce W. Herr and Todd M. Holloway

Power struggle

How do you keep track of the bubbling mass of information that is Wikipedia? This chaotic-looking mosaic is one attempt to show which topics are

locked until the mood cools (locked pages at the time of writing include entries on Sheffield Wednesday football club, Mikhail Gorbachev and pigs).

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

www.newscientist.com

19 May 2007 | NewScientist | 55

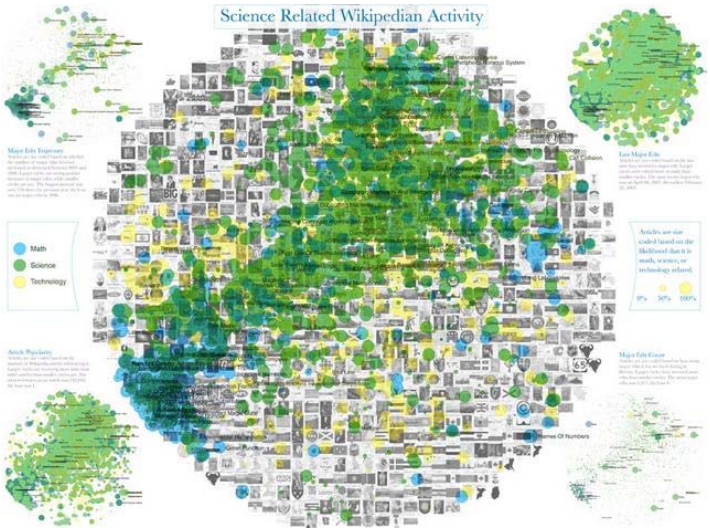
Science Related Wikipedian Activity

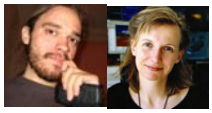
http://scimaps.org/dev/map_detail.php?map_id=165

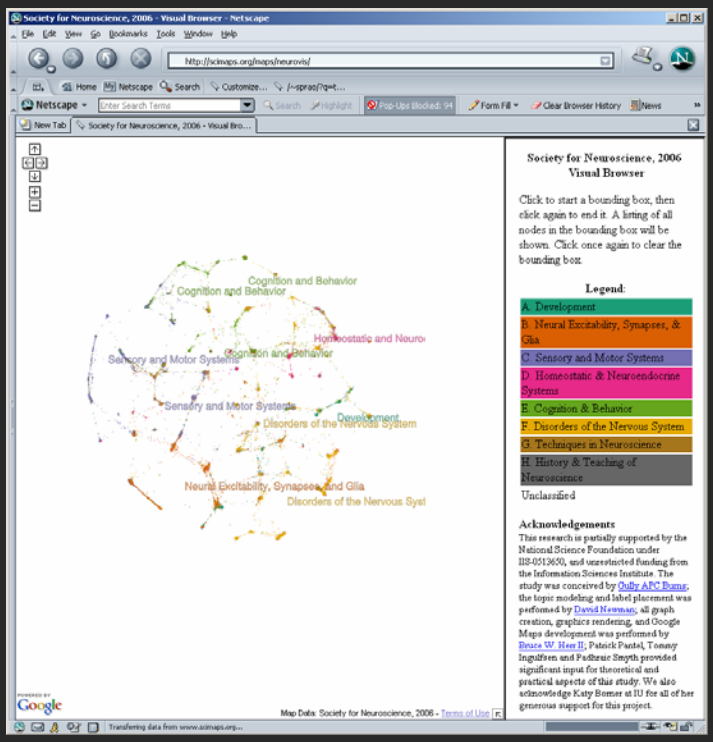
Same base map.

Overlaid are 3,599 math (blue), 6,474 science (green), and 3,164 technology relevant articles (yellow).
All other articles are given in grey.

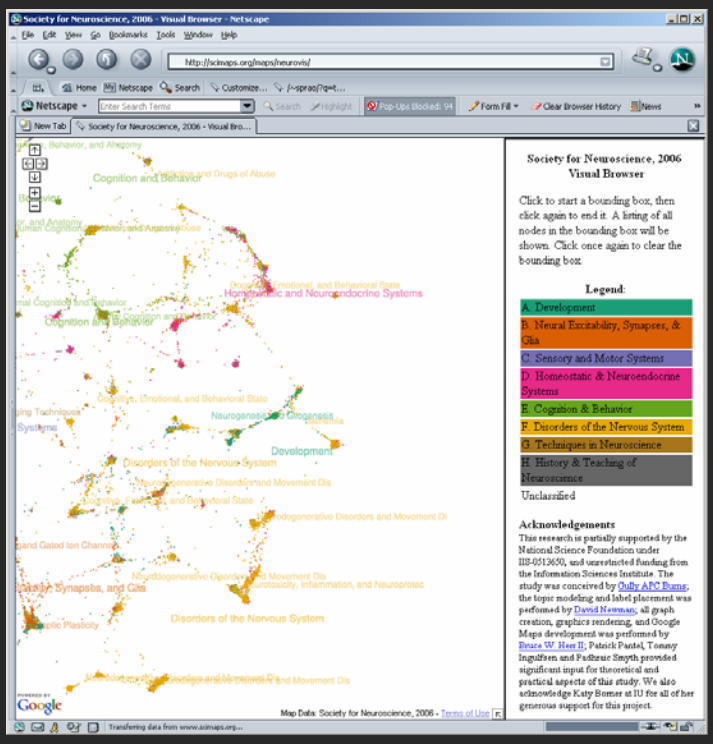
Corners show articles size coded according to
-article edit activity (top left),
- number of major edits (top right),
- number of bursts in edit activity (bottom, right)
- indegree (bottom left).



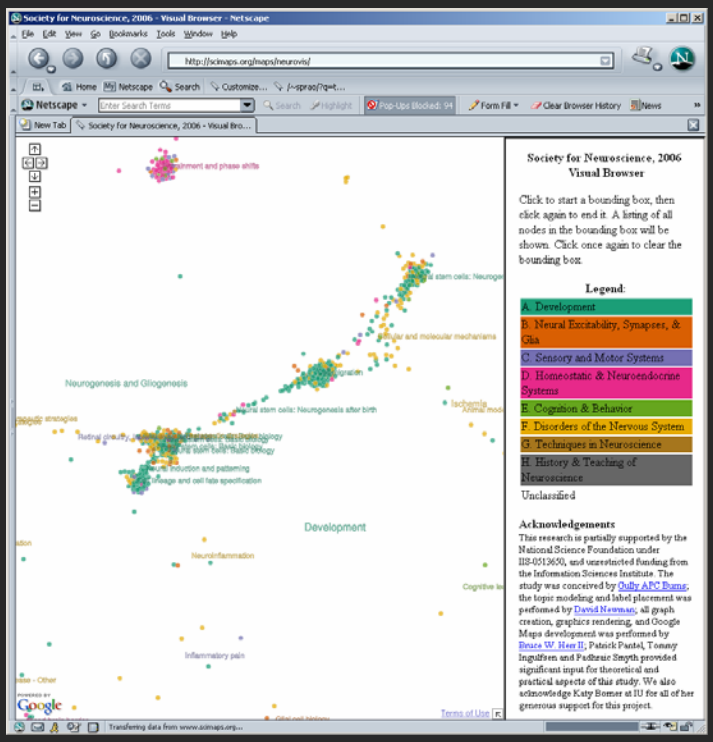




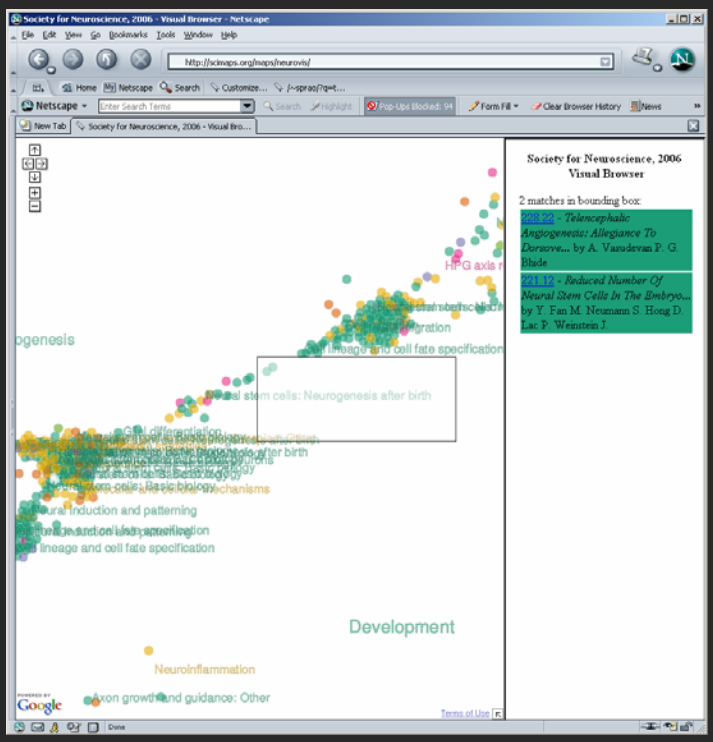
Bruce W. Herr II, Gully Burns (USC), David Newman (UCI), Society for Neuroscience, 2006 Visual Browser, 2007, http://scimaps.org/maps/neurovis/



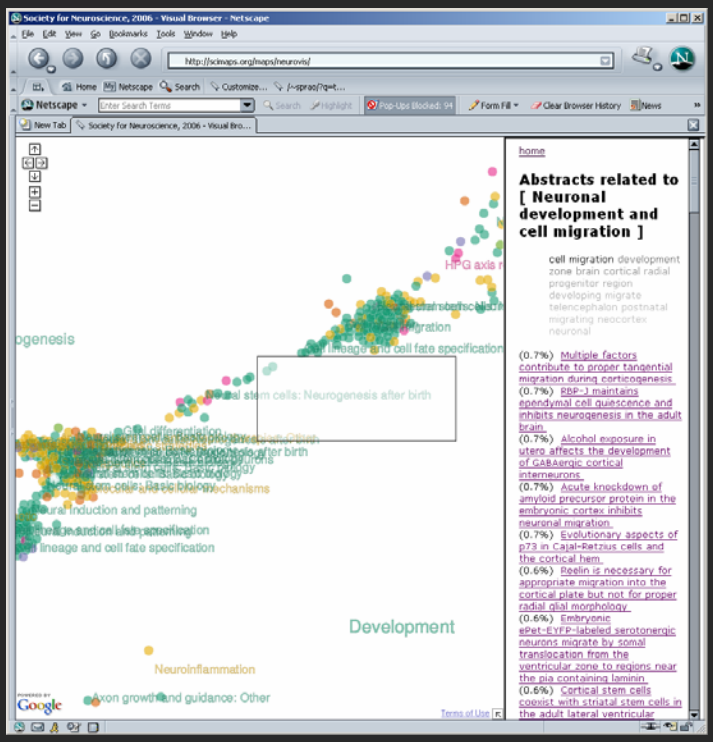
Bruce W. Herr II, Gully Burns (USC), David Newman (UCI), Society for Neuroscience, 2006 Visual Browser, 2007, http://scimaps.org/maps/neurovis/



Bruce W. Herr II, Gully Burns (USC), David Newman (UCI), Society for Neuroscience, 2006 Visual Browser, 2007, http://scimaps.org/maps/neurovis/



Bruce W. Herr II, Gully Burns (USC), David Newman (UCI), Society for Neuroscience, 2006 Visual Browser, 2007, http://scimaps.org/maps/neurovis/



Bruce W. Herr II, Gully Burns (USC), David Newman (UCI), Society for Neuroscience, 2006
Visual Browser, 2007, http://scimaps.org/maps/neurovis/

The End.