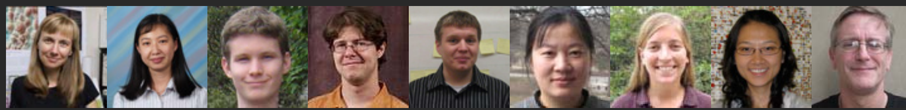
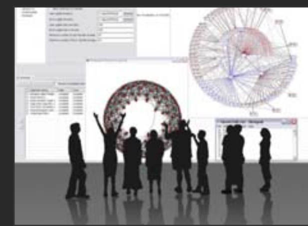


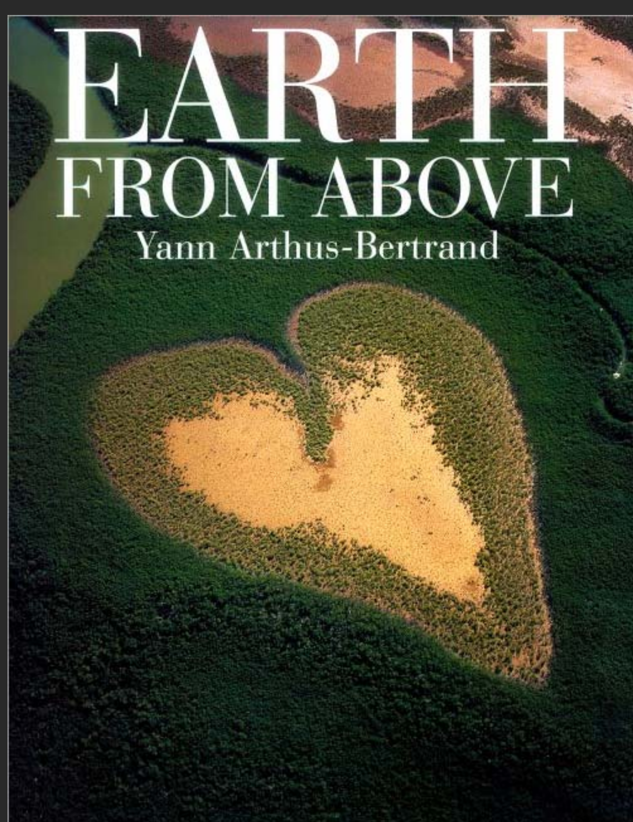


Analyzing and Visualizing the Structure and Evolution of World Wide Science

Angela Zoss, M.S.
Cyberinfrastructure for Network Science Center, Research Assistant
School of Library and Information Science
Indiana University, Bloomington, IN
amzoss@indiana.edu

August 12, 2009 – Information Kinetics: EgoViz
Arteleku, San Sebastián, Spain

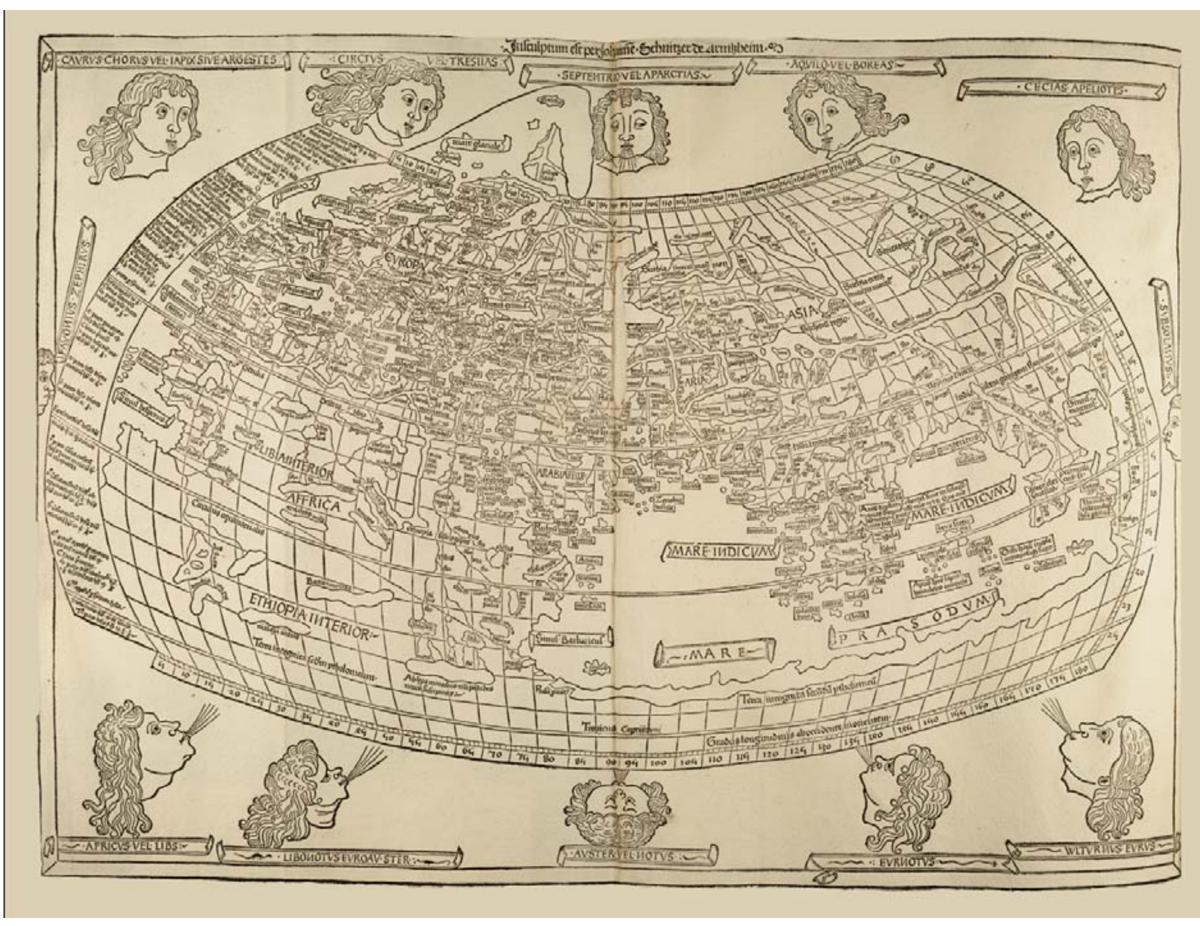
HPNAP



Cartographic maps of physical places have guided mankind's explorations for centuries.

They enabled the discovery of new worlds while also marking territories inhabited by the unknown.

Without maps, we would be lost.



Domain maps of abstract semantic spaces aim to serve today's explorers navigating the world of science.

These maps are generated through a scientific analysis of large-scale scholarly datasets in an effort to connect and make sense of the bits and pieces of knowledge they contain.

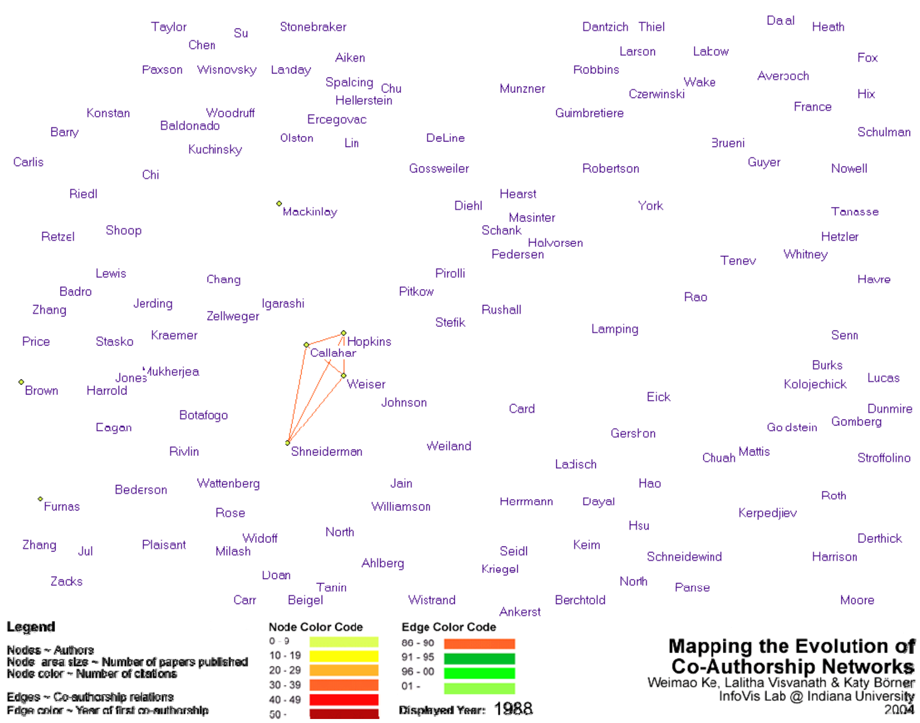
They can be used to identify objectively major research areas, experts, institutions, collections, grants, papers, journals, and ideas in a domain of interest. Science maps can provide overviews of "all-of-science" or of a specific area.

They can show homogeneity vs. heterogeneity, cause and effect, and relative speed. They allow us to track the emergence, evolution, and disappearance of topics and help to identify the most promising areas of research.

Sample Science Studies

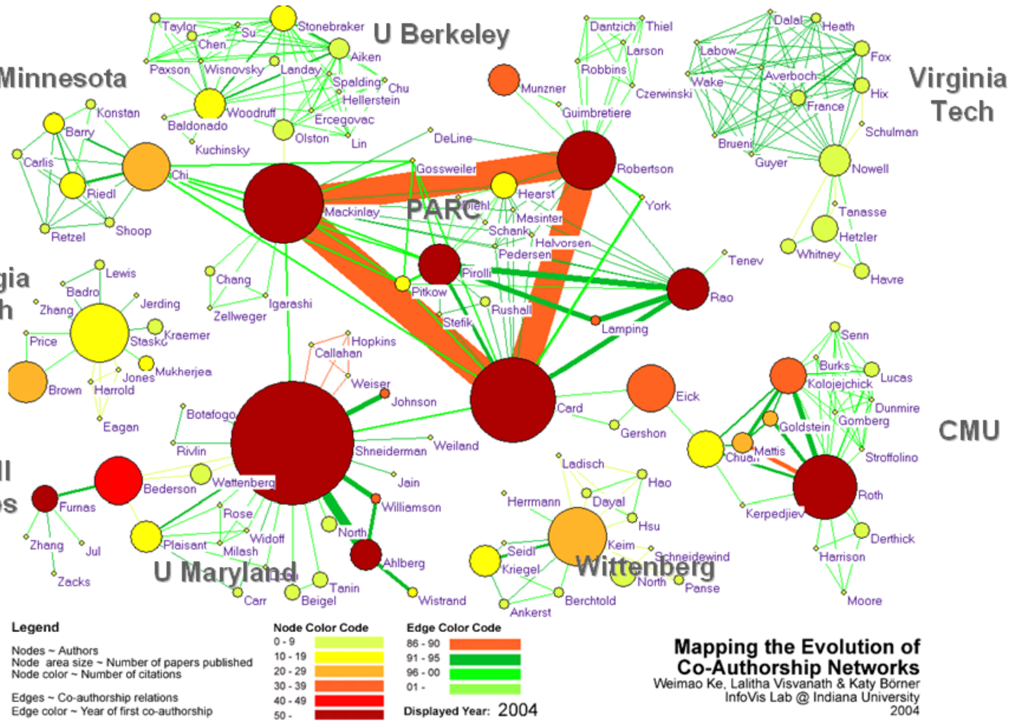
Mapping the Evolution of Co-Authorship Networks

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



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

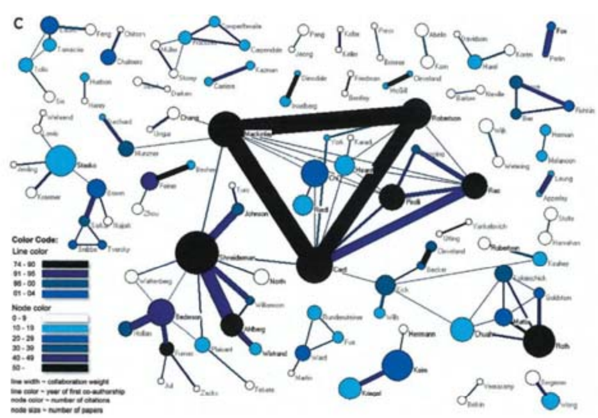
Börner, Dall'Asta, Ke & Vespignani (2005) Complexity, 10(4):58-67.

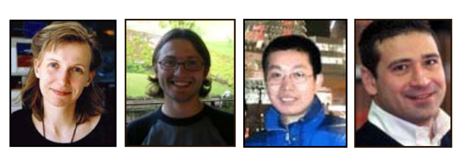
Research question:

- Is science driven by prolific single experts or by high-impact co-authorship teams?

Contributions:

- New approach to allocate citational credit.
- Novel weighted graph representation.
- Visualization of the growth of weighted co-author network.
- Centrality measures to identify author impact.
- Global statistical analysis of paper production and citations in correlation with co-authorship team size over time.
- Local, author-centered entropy measure.



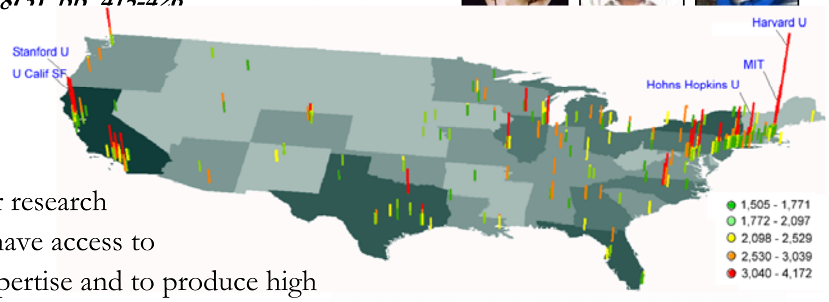


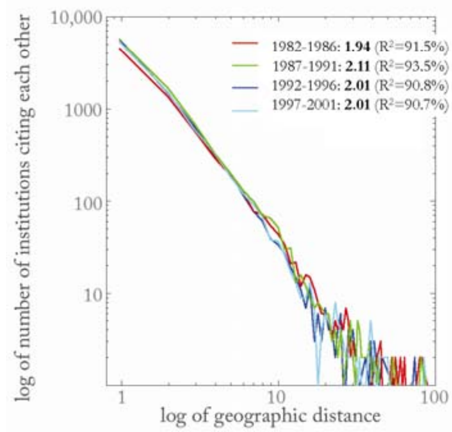
Spatio-Temporal Information Production and Consumption of Major U.S. Research Institutions

Börner, Katy, Penumarthy, Shashikant, Meiss, Mark and Ke, Weimao. (
Mapping the Diffusion of Scholarly Knowledge Among Major U.S. Res
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?

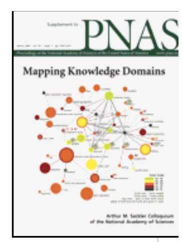




Contributions:

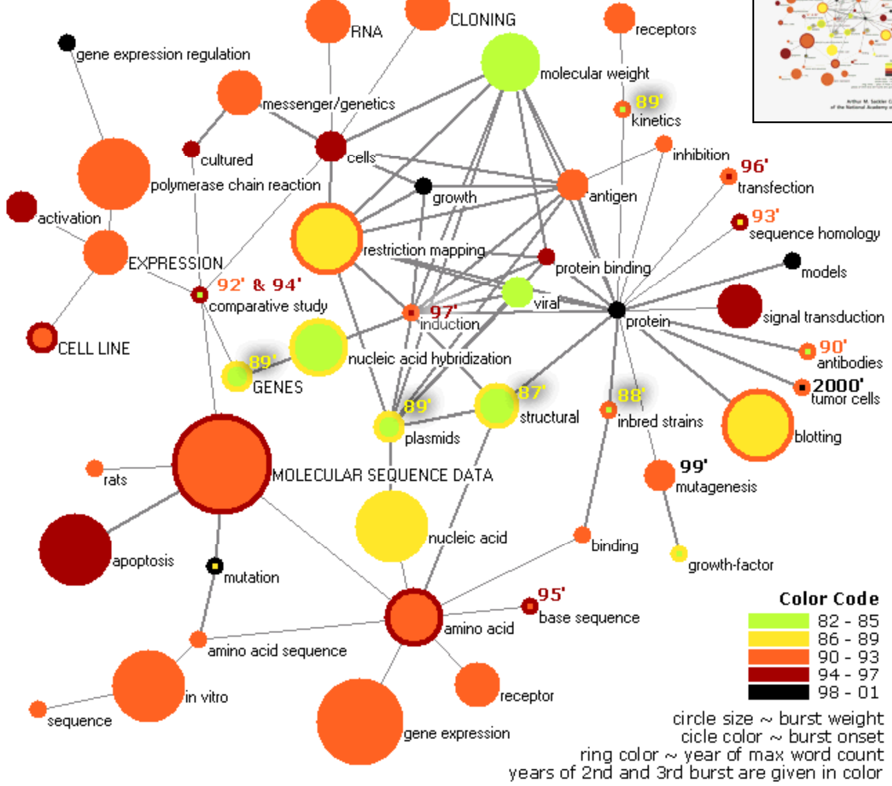
- Answer to Qs 1 + 2 is YES.
- Answer to Qs 3 is NO.
- Novel approach to analyzing the dual role of institutions as information producers and consumers and to study and visualize the diffusion of information among them.

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.

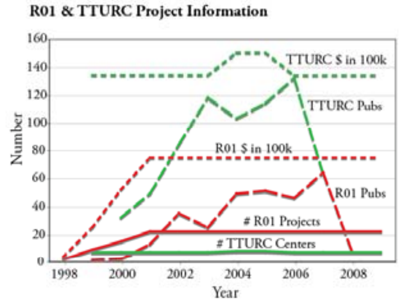
Mane & Börner. (2004) PNAS, 101(Suppl. 1): 5287-5290.



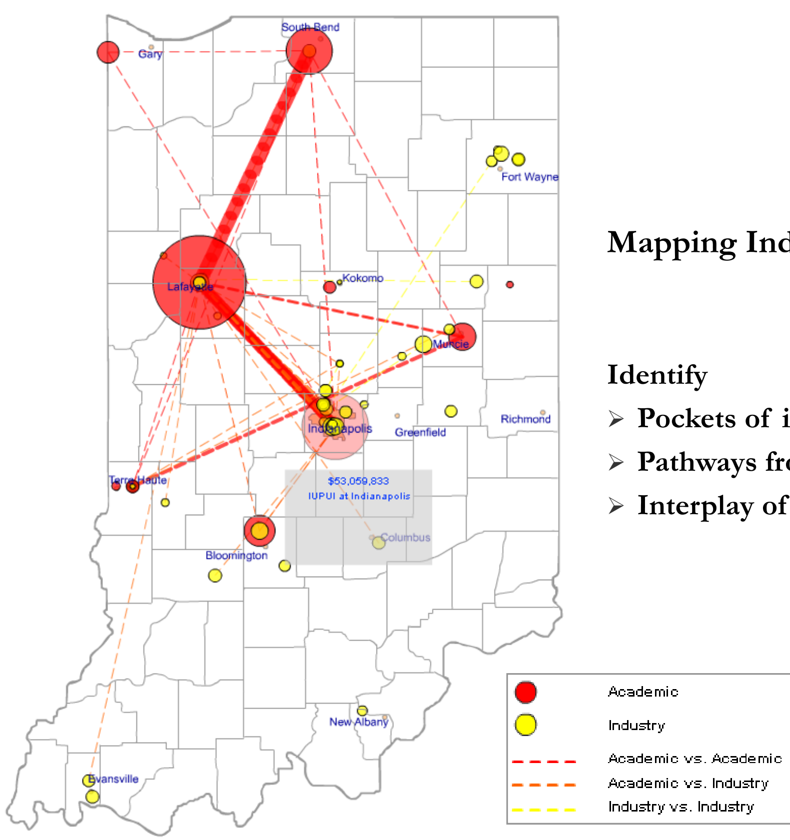
Mapping Transdisciplinary Tobacco Use Research Centers Publications

Compare R01 investigator based funding with TTURC Center awards in terms of number of publications and evolving co-author networks.

Zoss & Börner, forthcoming.

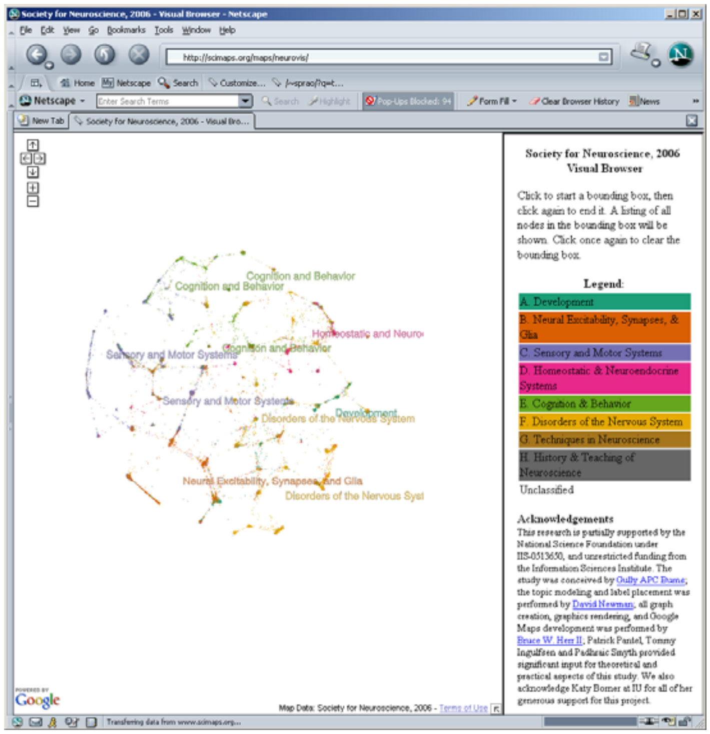


Mapping Indiana's Intellectual Space

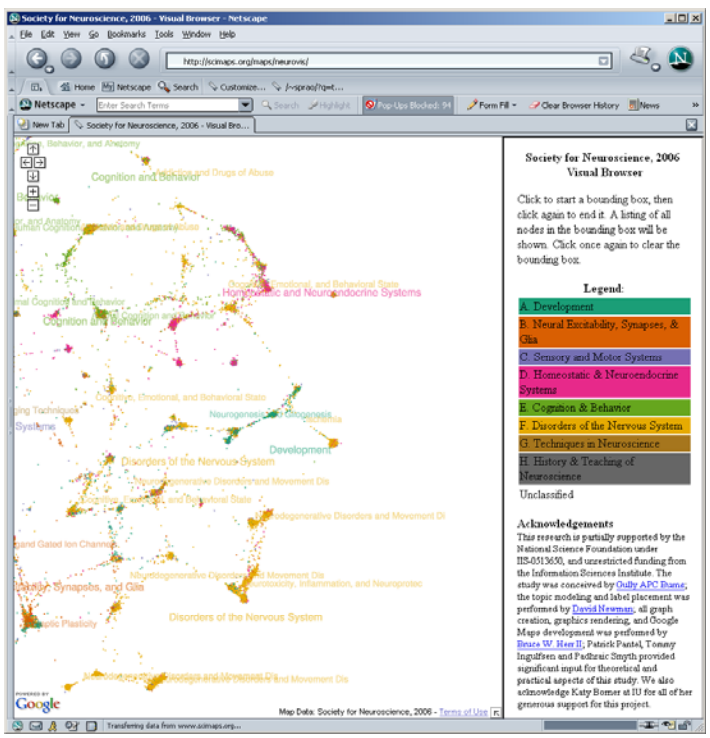


Identify

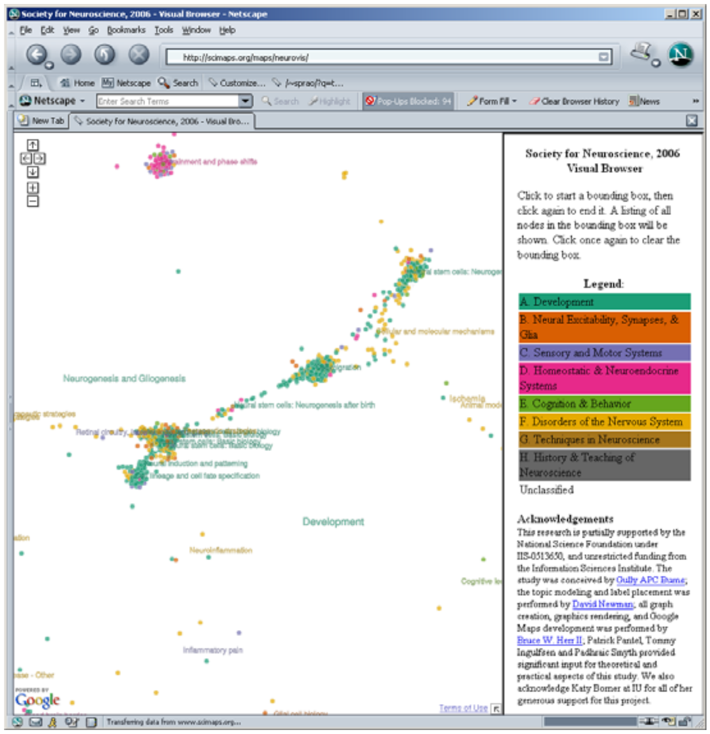
- Pockets of innovation
- Pathways from ideas to products
- Interplay of industry and academia



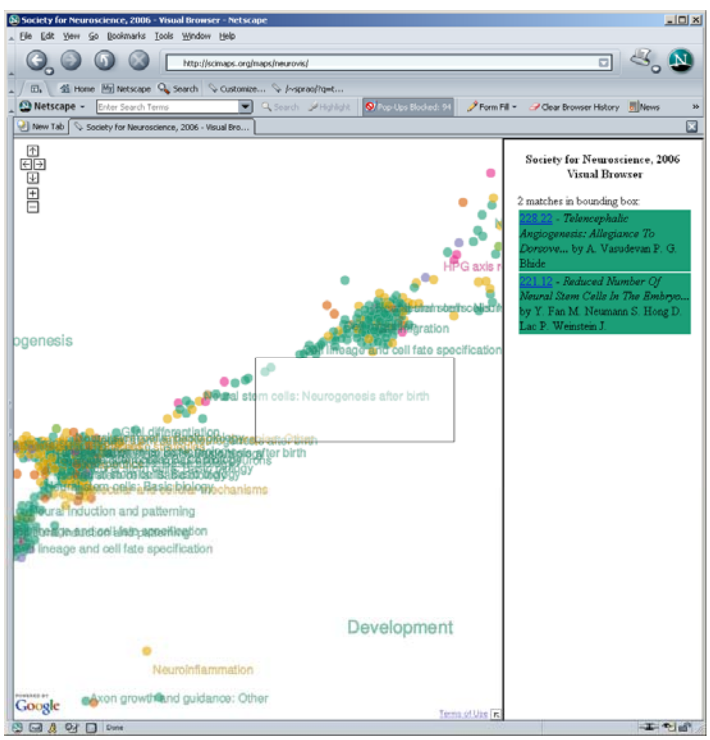
Bruce W. Herr II, Gully Burns (USC), David Newman (UCI), Society for Neuroscience, 2006 Visual Browser, 2007, http://scimaps.org/maps/neurovis/



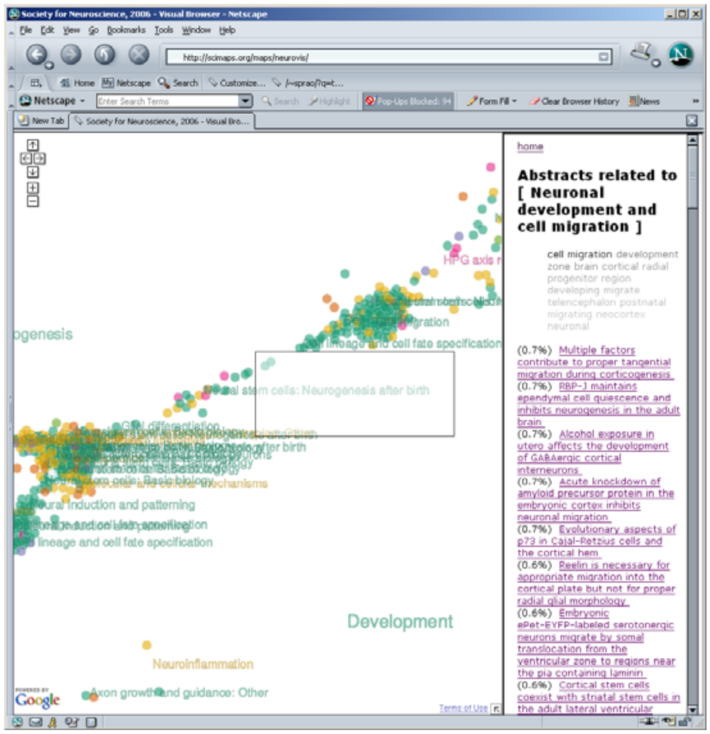
Bruce W. Herr II, Gully Burns (USC), David Newman (UCI), Society for Neuroscience, 2006 Visual Browser, 2007, http://scimaps.org/maps/neurovis/



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/***

Mapping Science Exhibit

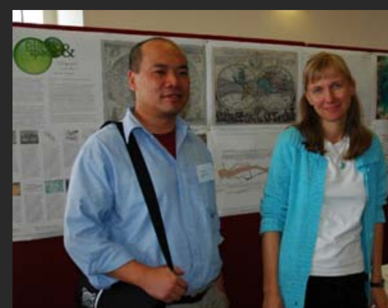
Places & Spaces: Mapping Science

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

Exhibit Curators:

Dr. Katy Börner & Elisha Hardy

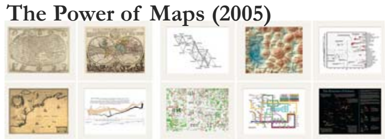
http://scimaps.org



Mapping Science Exhibit – 10 Iterations in 10 years

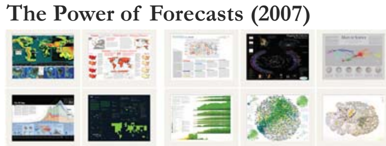
http://scimaps.org/

The Power of Maps (2005)



The Power of Reference Systems (2006)

The Power of Forecasts (2007)



Science Maps for Economic Decision Makers (2008)

Science Maps for Science Policy Makers (2009)

Science Maps for Scholars (2010)

Science Maps as Visual Interfaces to Digital Libraries (2011)

Science Maps for Kids (2012)

Science Forecasts (2013)

How to Lie with Science Maps (2014)

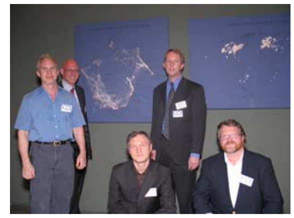
Exhibit has been shown in 52 venues on four continents. Also at

- NSF, 10th Floor, 4201 Wilson Boulevard, Arlington, VA.
- Chinese Academy of Sciences, China, May 17-Nov. 15, 2008.
- University of Alberta, Edmonton, Canada, Nov 10-Jan 31, 2009
- Center of Advanced European Studies and Research, Bonn, Germany, Dec. 11-19, 2008.

ORDER MAPS

Illuminated Diagram Display

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



Large-scale, high resolution prints illuminated via projector or screen.

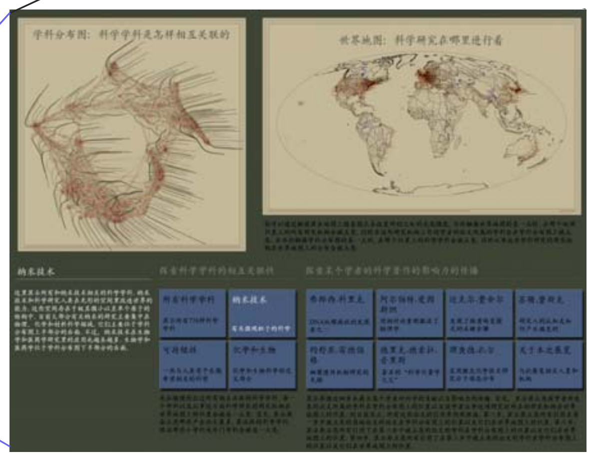
Questions:

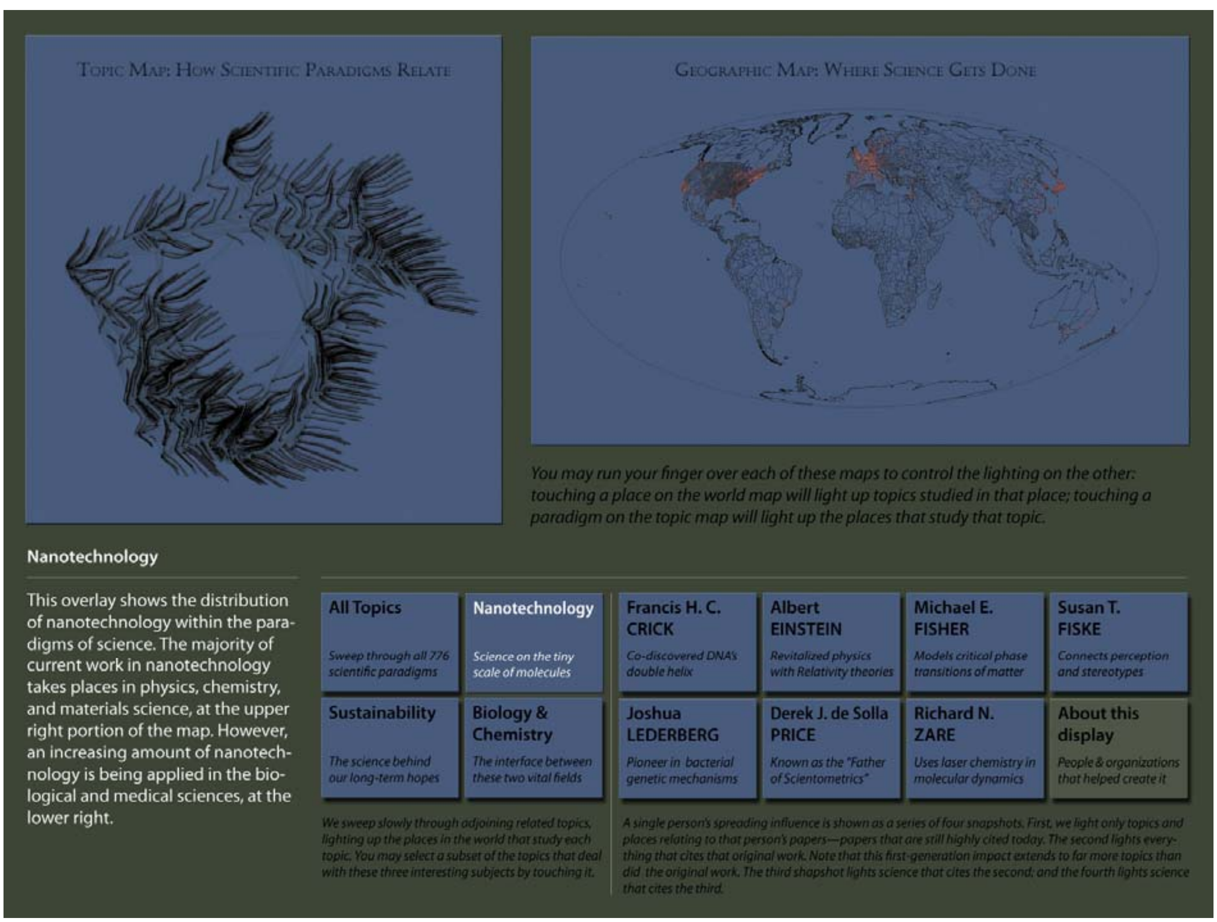
- Who is doing research on what topic and where?
- What is the 'footprint' of interdisciplinary research fields?
- What impact have scientists?

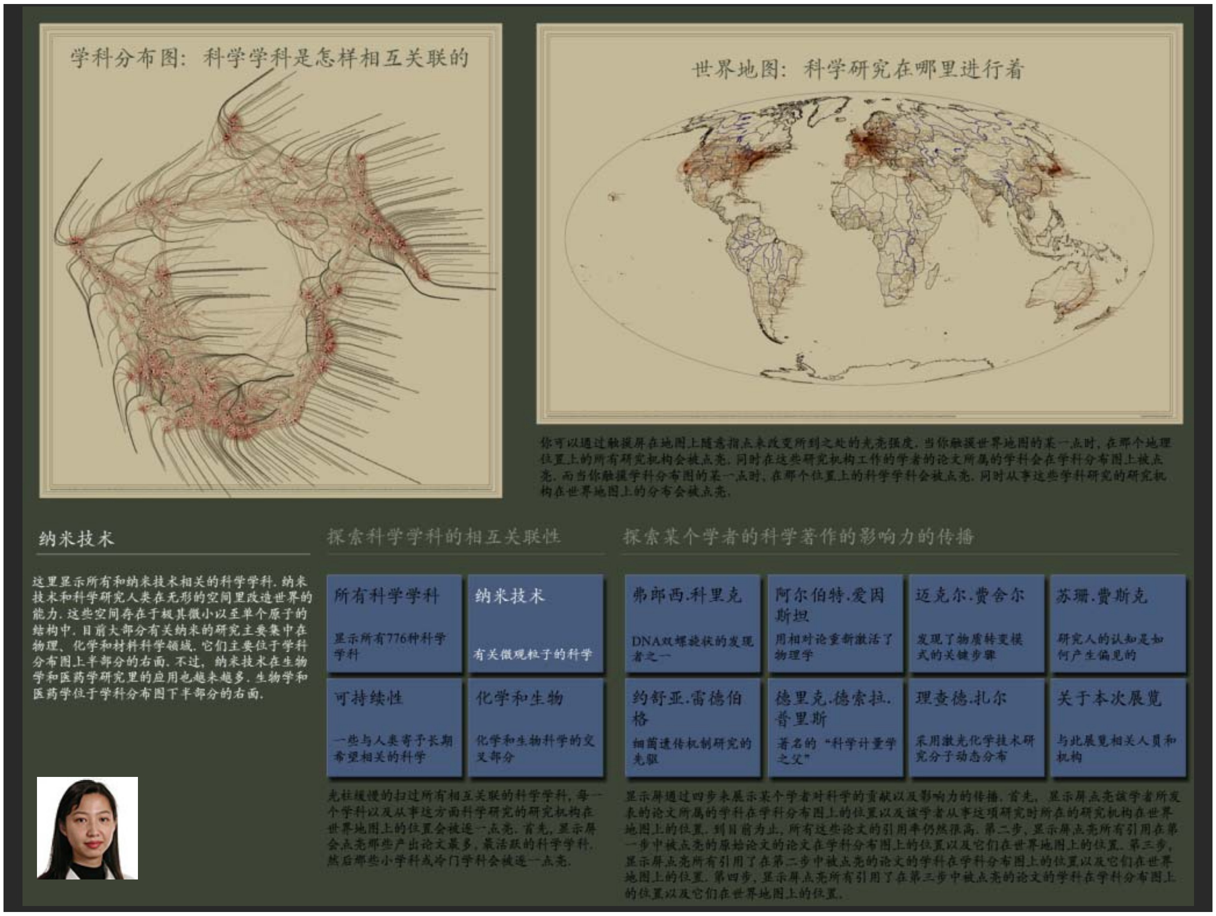
Contributions:

- Interactive, high resolution interface to access and make sense of data about scholarly activity.

Interactive touch panel.







Re-implementation of Illuminated Diagram Software
by Advanced Visualization Lab, Indiana University

Drives unlimited number of ID screens.

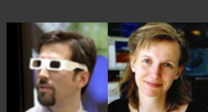
Science World

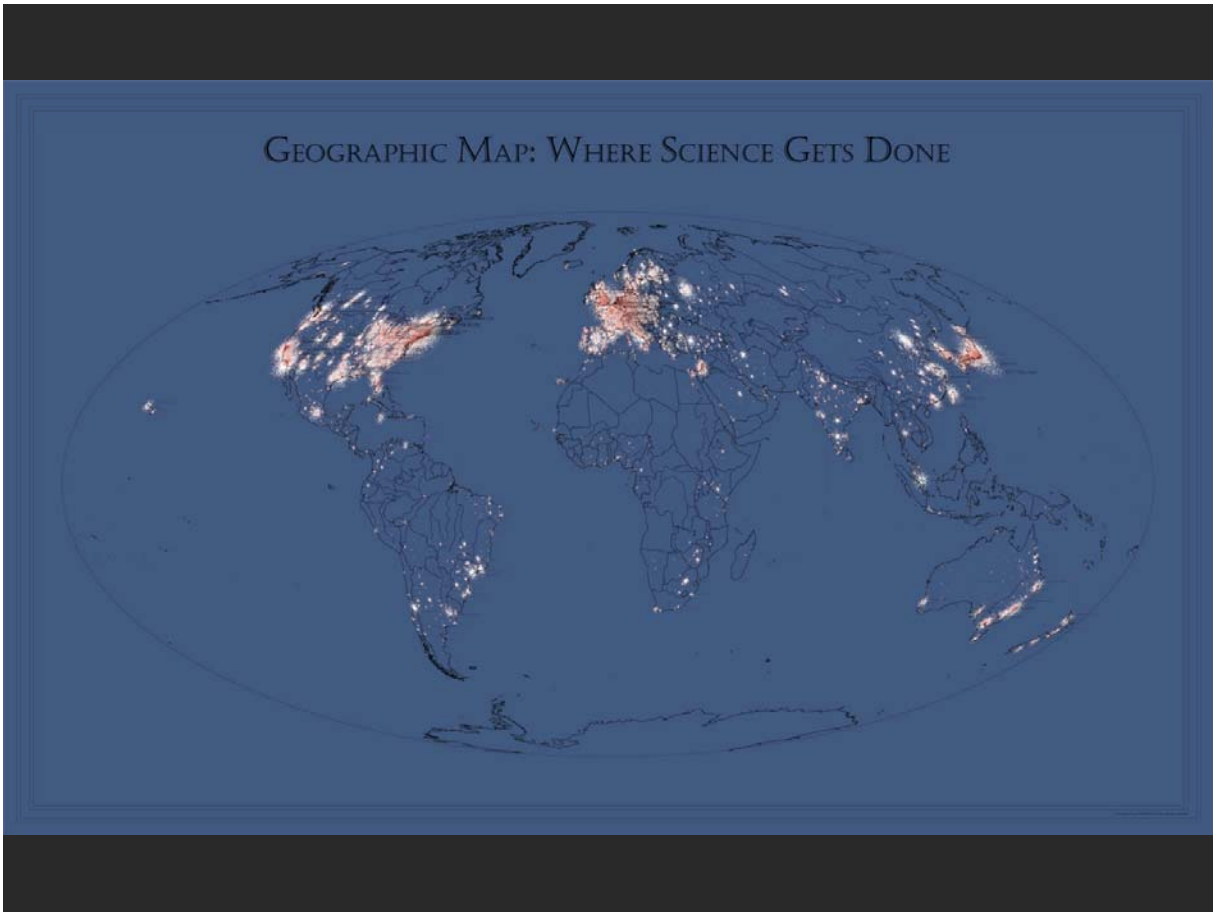
Medicine Asia

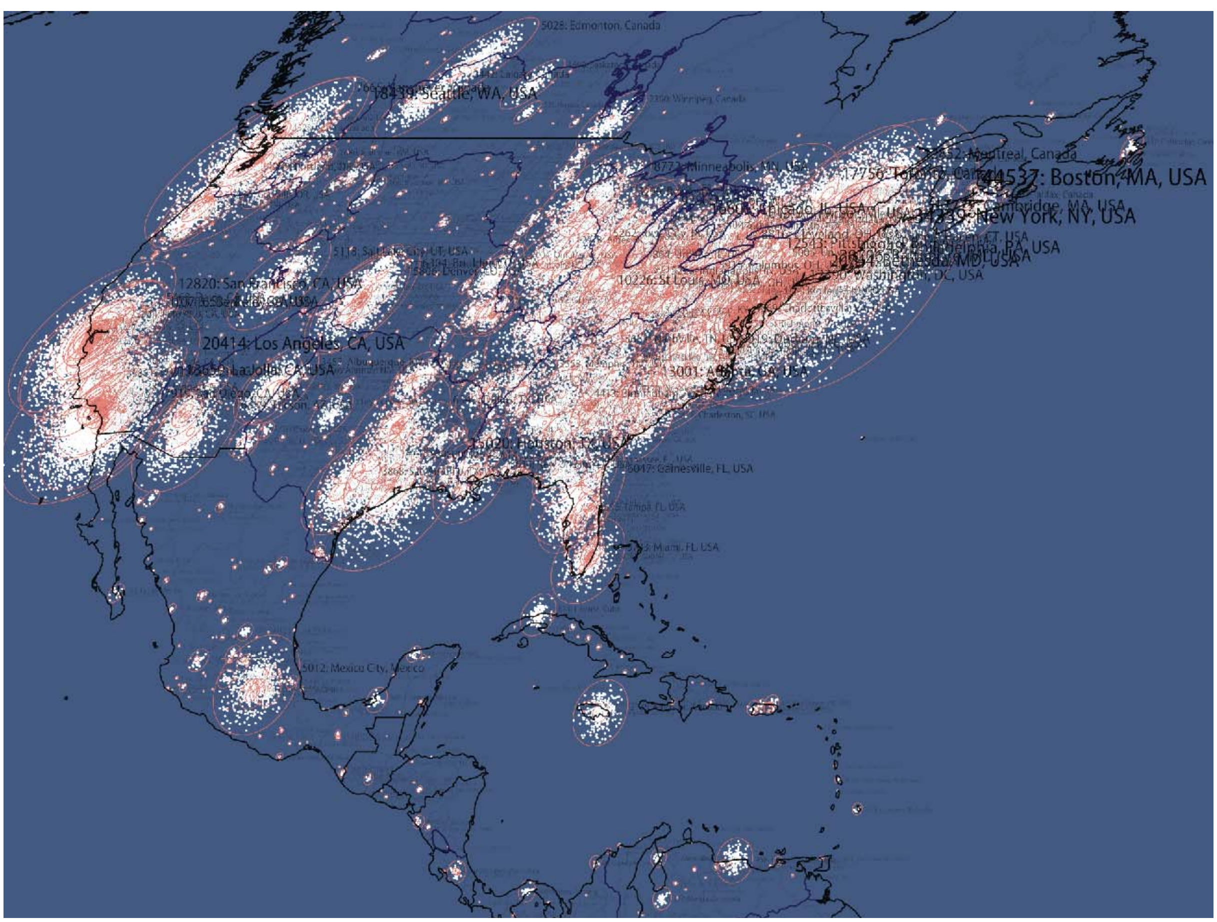
Cancer Japan

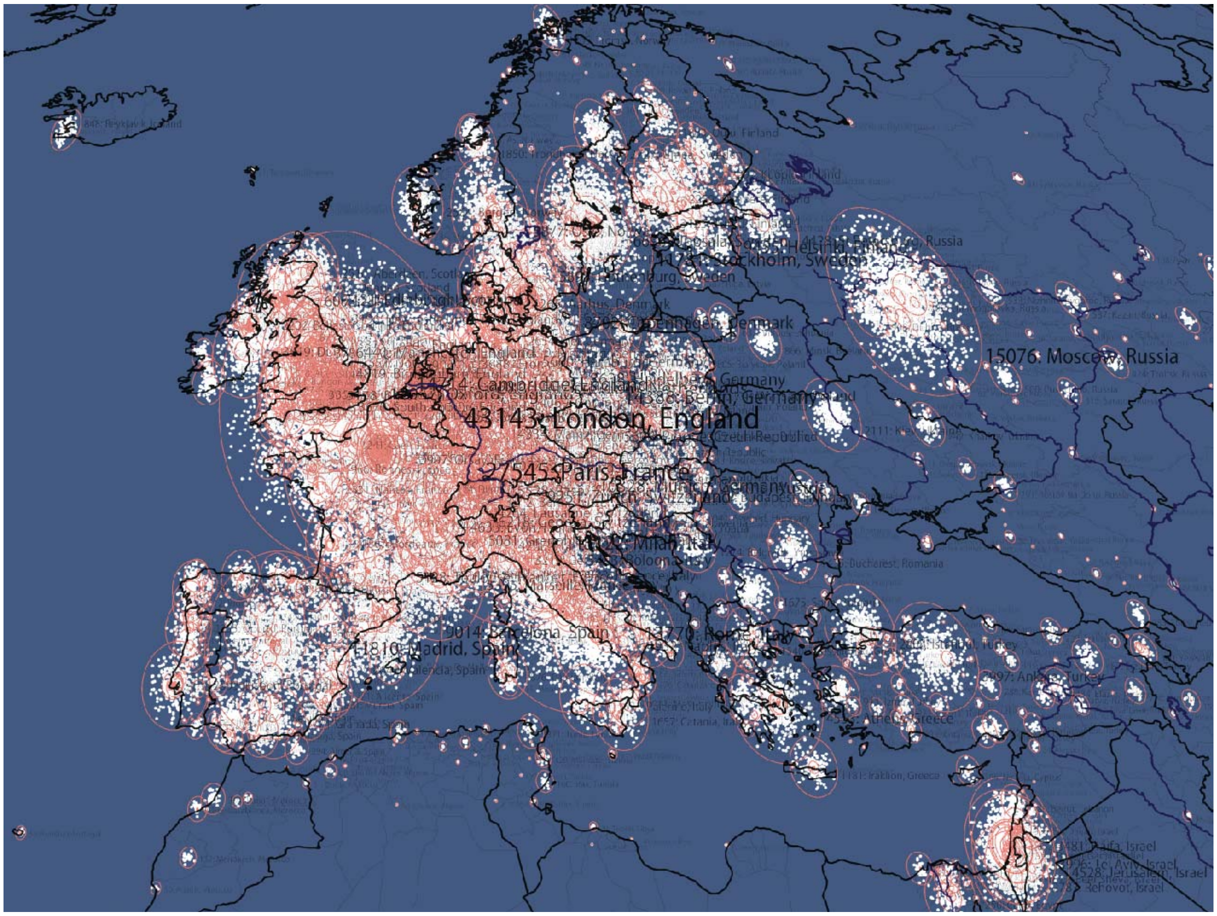
... ...

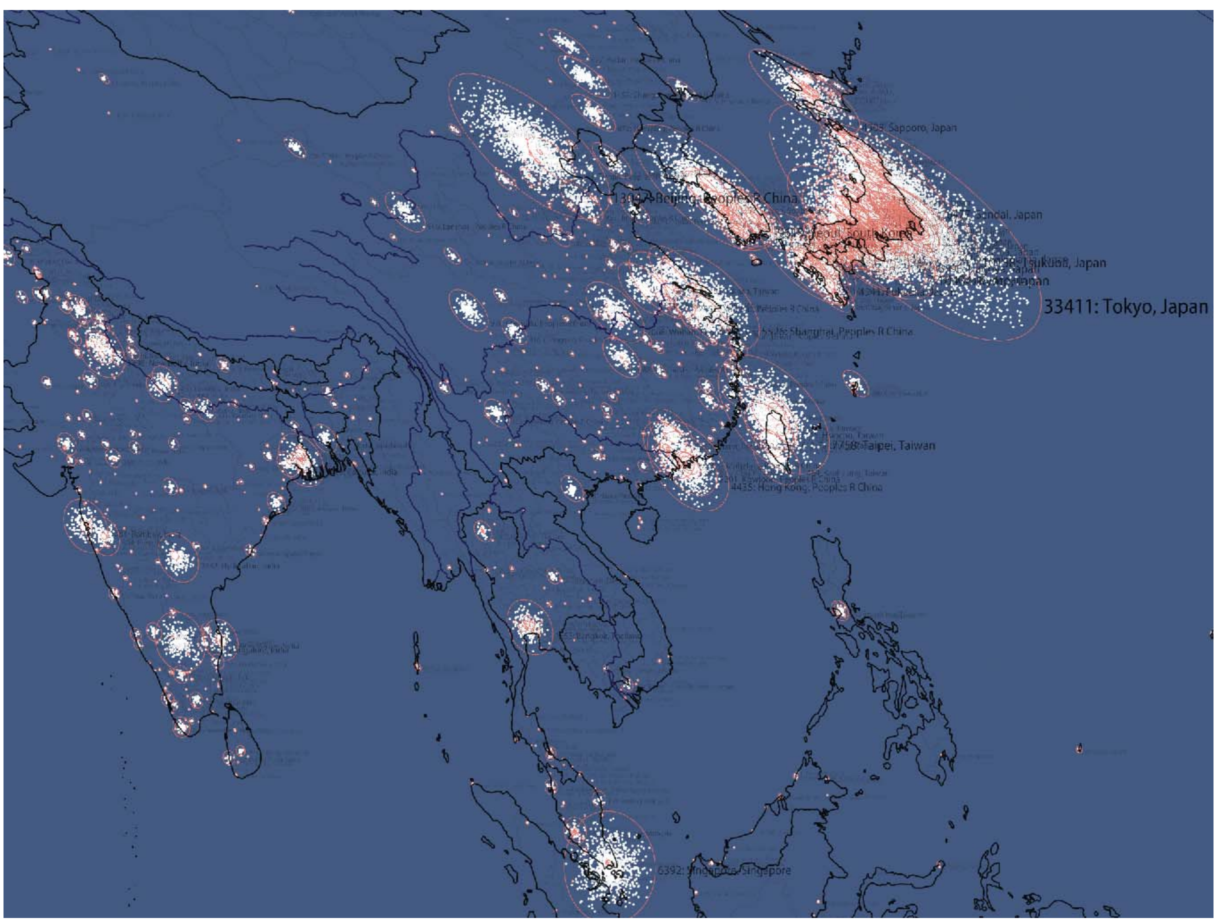
Touch screen for direct interaction.
Keyword and name search.
Selection of canned queries for
- interdisciplinary research areas
- famous people
- activity patterns, e.g., bursts, trends, etc.



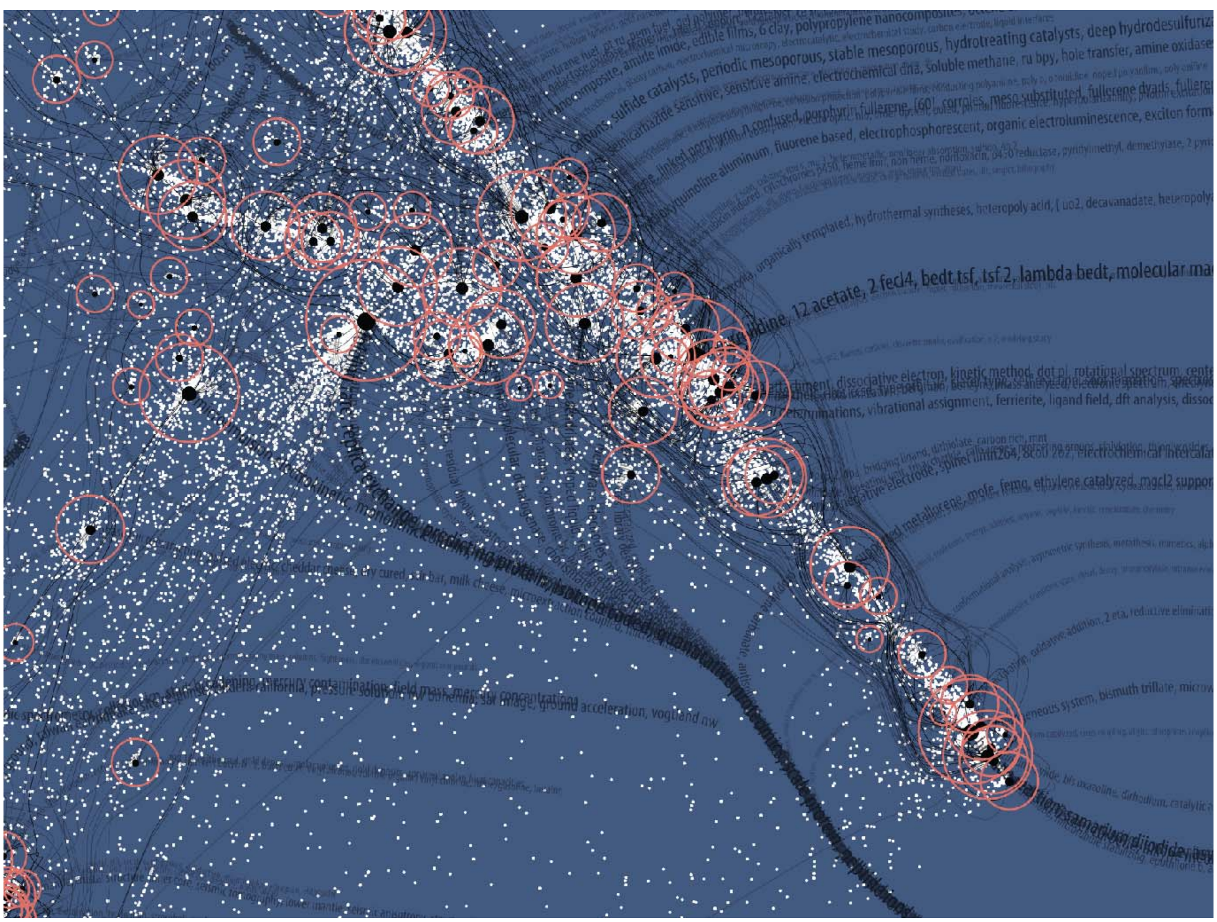

GEOGRAPHIC MAP: WHERE SCIENCE GETS DONE


5028: Edmonton, Canada
18239: Seattle, WA, USA
44537: Boston, MA, USA
New York, NY, USA
12820: San Francisco, CA, USA
20414: Los Angeles, CA, USA
5017: Gainesville, FL, USA
5012: Mexico City, Mexico


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


33411: Tokyo, Japan
Sapporo, Japan
Taipei, Taiwan
Hong Kong, Peoples R China
Singapore

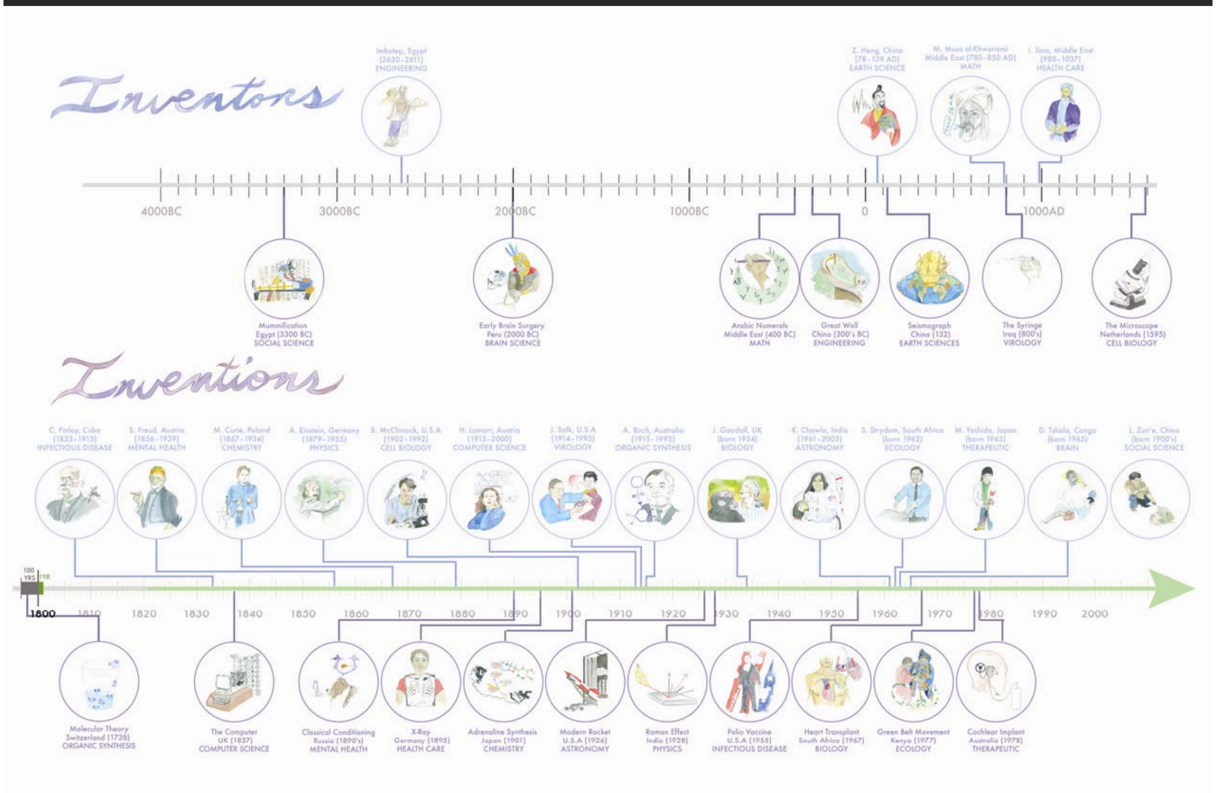

Topic Map: How Scientific Paradigms Relate

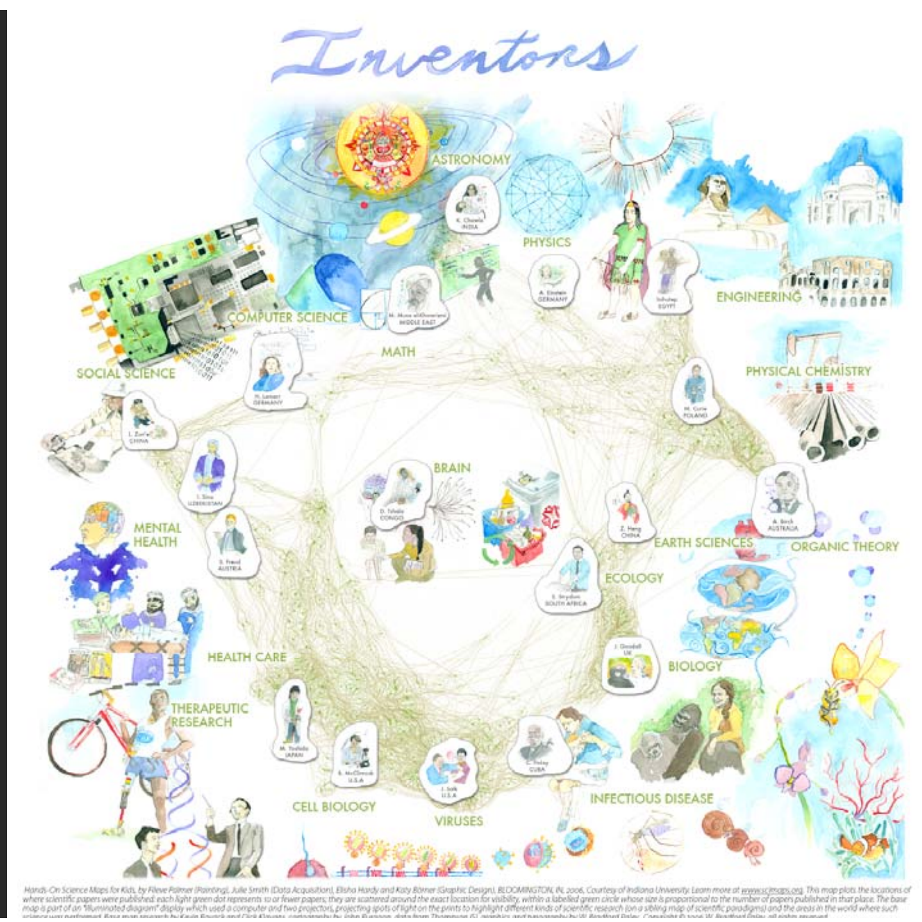


Teaching Children the Structure of Science

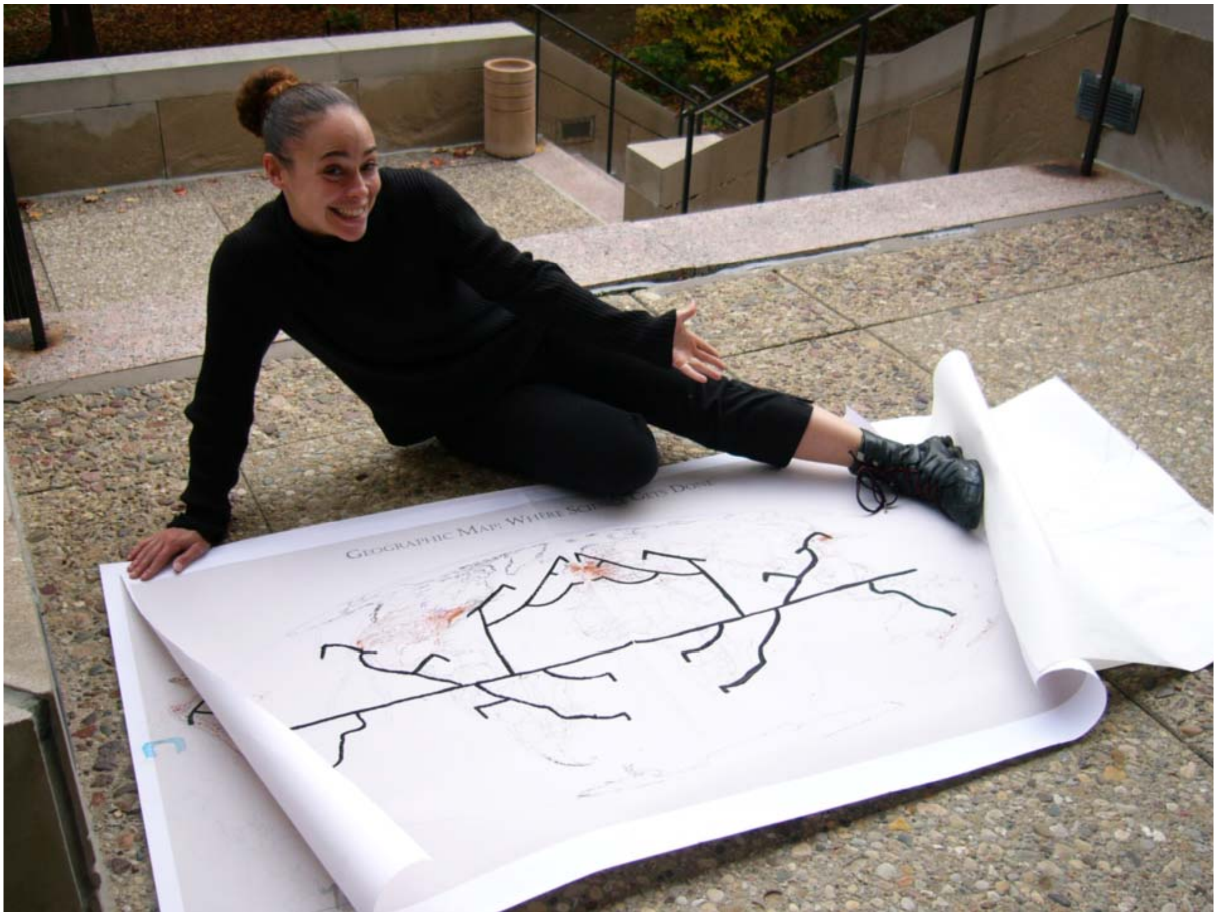
- How can children start to understand the complex interplay of the different sciences?
- How can we teach them to appreciate the very diverse cultures, research approaches, and languages that exist in the different sciences and enable them to ‘speak’ more than one science in order to collaborate across scientific boundaries?
- Last but not least, how can we engage children in the work of real scientists, have them share the excitement of discovery, and allow them to find their own ‘place’ in science?

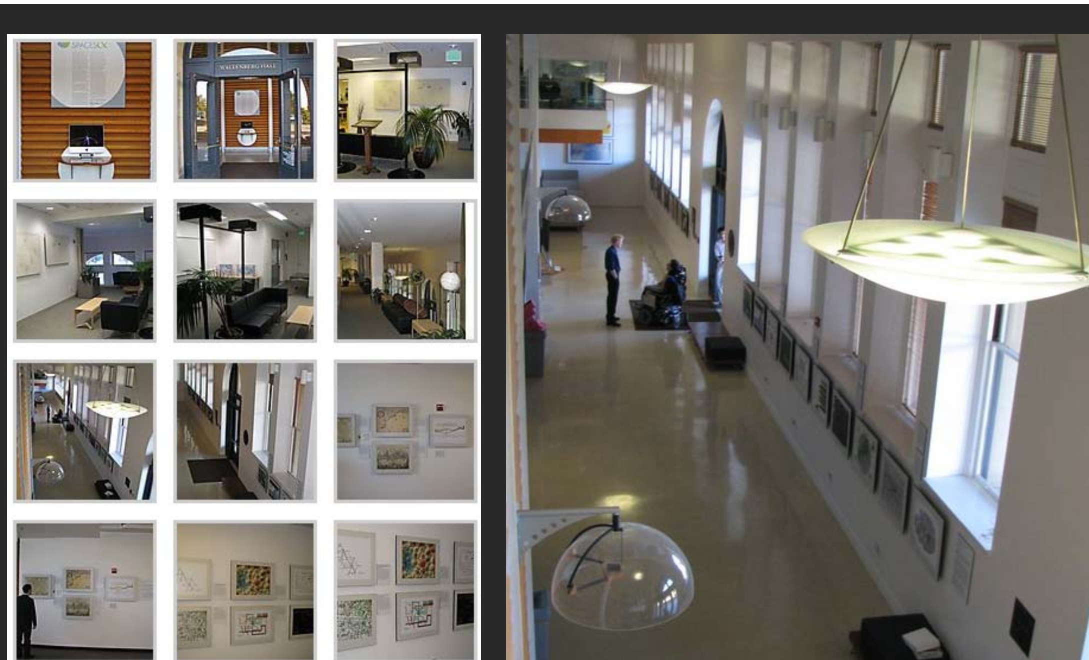
Börner, K., Teaching Children the Structure of Science, Workshop on “Using Maps of Science to teach Science”*, 12th International Conference on Scientometrics and Informetrics (ISSI 2009)*, Rio de Janeiro.


Inventors
Imhotep, Egypt (2630–2611) ENGINEERING
Z. Heng, China (78–139 AD) EARTH SCIENCE
M. Musa al-Khwarizmi Middle East (780–850 AD) MATH
I. Sina, Middle East (980–1037) HEALTH CARE
4000BC
3000BC
2000BC
1000BC
0
1000AD
Mummification Egypt (3300 BC) SOCIAL SCIENCE
Early Brain Surgery Peru (2000 BC) BRAIN SCIENCE
Arabic Numerals Middle East (400 BC) MATH
Great Wall China (300's BC) ENGINEERING
Seismograph China (132) EARTH SCIENCES
The Syringe Iraq (800's) VIROLOGY
The Microscope Netherlands (1595) CELL BIOLOGY
Inventions
C. Finlay, Cuba (1833–1915) INFECTIOUS DISEASE
S. Freud, Austria (1856–1939) MENTAL HEALTH
M. Curie, Poland (1867–1934) CHEMISTRY
A. Einstein, Germany (1879–1955) PHYSICS
B. McClintock, U.S.A (1902–1992) CELL BIOLOGY
H. Lamarr, Austria (1913–2000) COMPUTER SCIENCE
J. Salk, U.S.A (1914–1995) VIROLOGY
A. Birch, Australia (1915–1995) ORGANIC SYNTHESIS
J. Goodall, UK (born 1934) BIOLOGY
K. Chawla, India (1961–2003) ASTRONOMY
S. Strydom, South Africa (born 1962) ECOLOGY
M. Yoshida, Japan (born 1963) THERAPEUTIC
D. Tshala, Congo (born 1965) BRAIN
L. Zun'e, China (born 1900's) SOCIAL SCIENCE
100 YRS
1800
1810
1820
1830
1840
1850
1860
1870
1880
1890
1900
1910
1920
1930
1940
1950
1960
1970
1980
1990
2000
Molecular Theory Switzerland (1738) ORGANIC SYNTHESIS
The Computer UK (1837) COMPUTER SCIENCE
Classical Conditioning Russia (1890's) MENTAL HEALTH
X-Ray Germany (1895) HEALTH CARE
Adrenaline Synthesis Japan (1901) CHEMISTRY
Modern Rocket U.S.A (1926) ASTRONOMY
Raman Effect India (1928) PHYSICS
Polio Vaccine U.S.A (1955) INFECTIOUS DISEASE
Heart Transplant South Africa (1967) BIOLOGY
Green Belt Movement Kenya (1977) ECOLOGY
Cochlear Implant Australia (1978) THERAPEUTIC


Inventors
ASTRONOMY
PHYSICS
ENGINEERING
COMPUTER SCIENCE
MATH
SOCIAL SCIENCE
PHYSICAL CHEMISTRY
BRAIN
MENTAL HEALTH
EARTH SCIENCES
ORGANIC THEORY
ECOLOGY
HEALTH CARE
BIOLOGY
THERAPEUTIC RESEARCH
CELL BIOLOGY
VIRUSES
INFECTIOUS DISEASE
Hands-On Science Maps for Kids, by Fileve Palmer (Painting), Julie Smith (Data Acquisition), Elisha Hardy and Katy Börner (Graphic Design), BLOOMINGTON, IN, 2006. Courtesy of Indiana University. Learn more at www.scimaps.org. This map plots the locations of where scientific papers were published: each light green dot represents 10 or fewer papers; they are scattered around the exact location for visibility, within a labelled green circle whose size is proportional to the number of papers published in that place. The base map is part of an "illuminated diagram" display which used a computer and two projectors, projecting spots of light on the prints to highlight different kinds of scientific research (on a sibling map of scientific paradigms) and the areas in the world where such

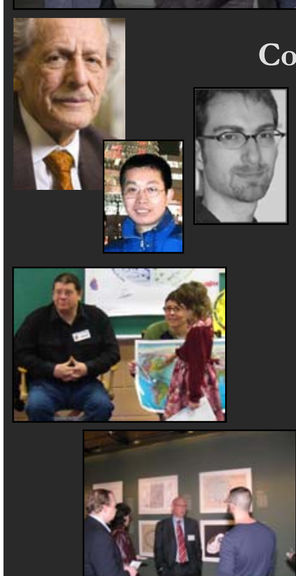
Science Puzzle Map for Kids by Fileve Palmer, Julie Smith, Elisha Hardy and Katy Börner, Indiana University, 2006. (Base map taken from Illuminated Diagram display by Kevin Boyack, Richard Klavans, and W. Bradford Paley.)



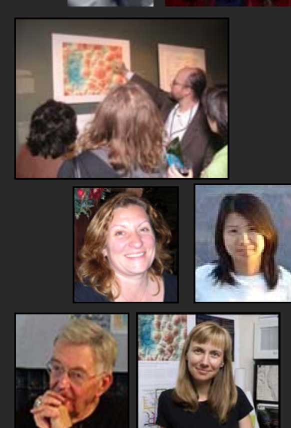


Debut of 5th Iteration of Mapping Science Exhibit at MEDIA X on May 18, 2009
at Wallenberg Hall, Stanford University
http://mediax.stanford.edu
http://scaleindependentthought.typepad.com/photos/scimaps

Science Maps in "Expedition Zukunft" science train visiting 62 cities in 7 months
12 coaches, 300 m long
Opened on April 23rd, 2009 by German Chancellor Merkel
http://www.expedition-zukunft.de

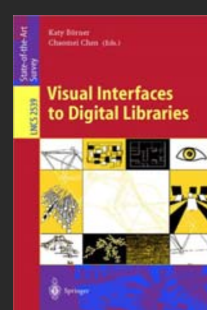


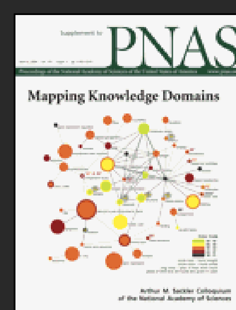
Contact the map makers via the exhibit curators:
Katy Börner (katy@indiana.edu) and Elisha Hardy (efhardy@indiana.edu)



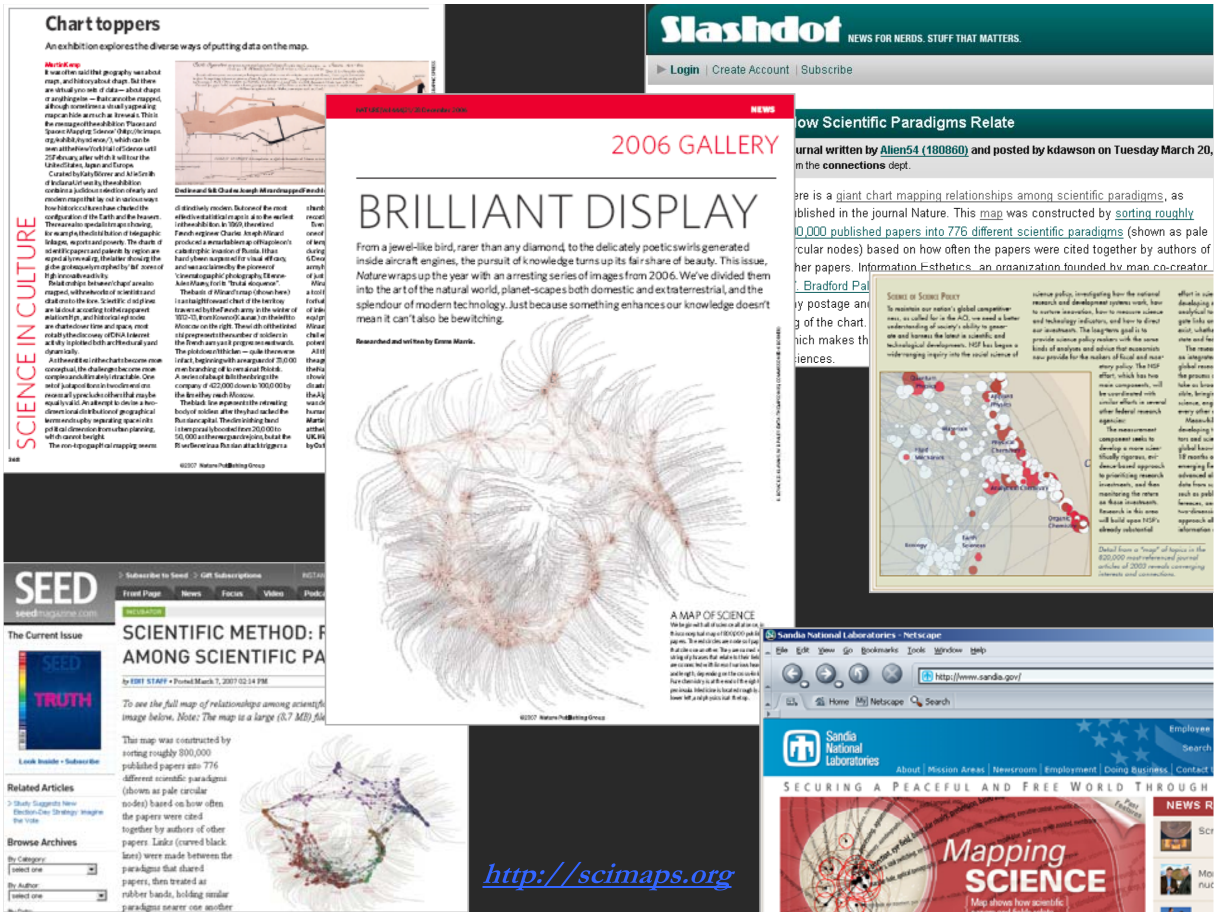
Using Science Maps

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.



Computational Scientometrics 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.

Process of Computational Scientometrics

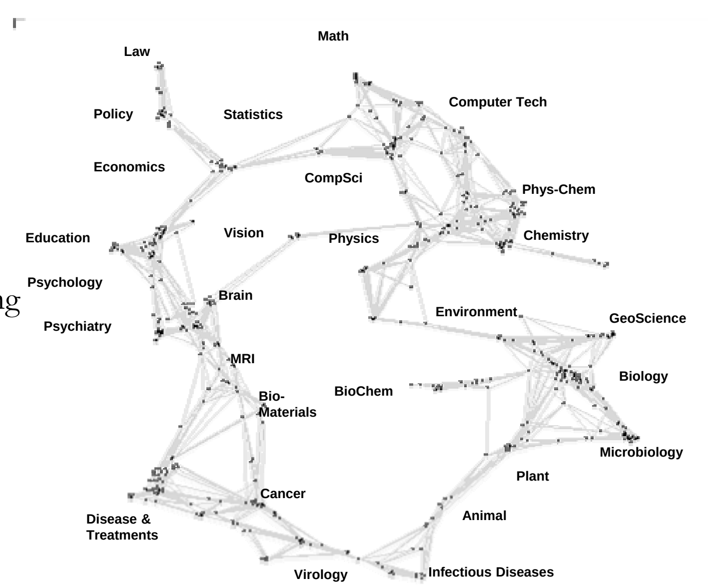
DATA EXTRACTION	UNIT OF ANALYSIS	MEASURES	LAYOUT (often one code does both similarity and ordination steps)		DISPLAY
			SIMILARITY	ORDINATION	
SEARCHES ISI INSPEC Eng Index Medline ResearchIndex Patents etc. BROADENING By citation By terms	COMMON CHOICES Journal Document Author Term	COUNTS/FREQUENCIES 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, 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.

Latest 'Base Map' of Science

Kevin W. Boyack, Katy Börner, & Richard Klavans (2007). Mapping the Structure and Evolution of Chemistry Research. 11th International Conference on Scientometrics and Informetrics. pp. 112-123.

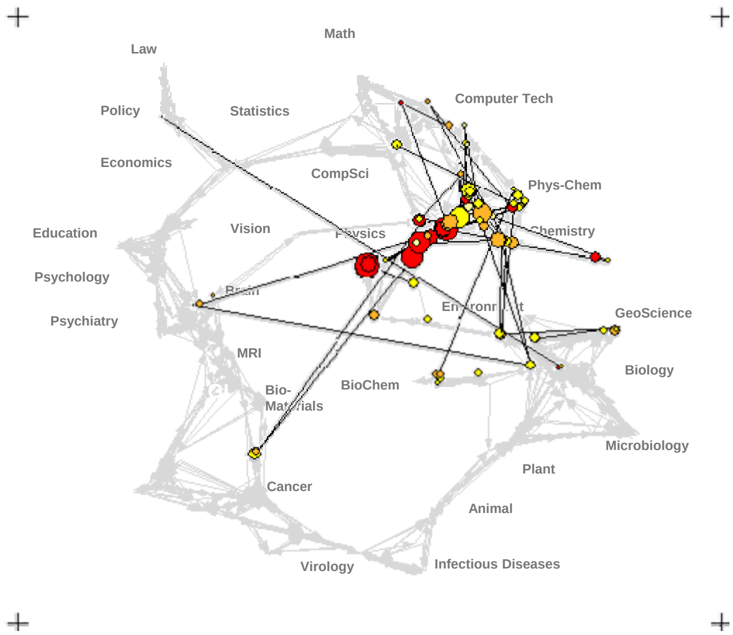
- Uses combined SCI/SSCI from 2002
  - 1.07M papers, 24.5M references, 7,300 journals
  - Bibliographic coupling of papers, aggregated to journals
- Initial ordination and clustering of journals gave 671 clusters
- Coupling counts were reaggregated at the journal cluster level to calculate the
  - (x,y)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



Science map applications: Identifying core competency

Kevin W. Boyack, Katy Börner, & Richard Klavans (2007).

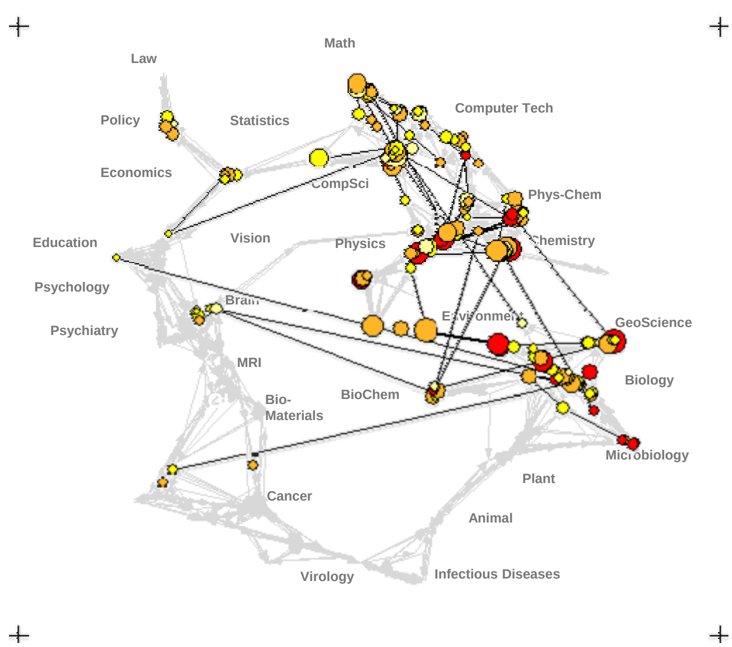
Funding patterns of the US Department of Energy (DOE)



Science map applications: Identifying core competency

Kevin W. Boyack, Katy Börner, & Richard Klavans (2007).

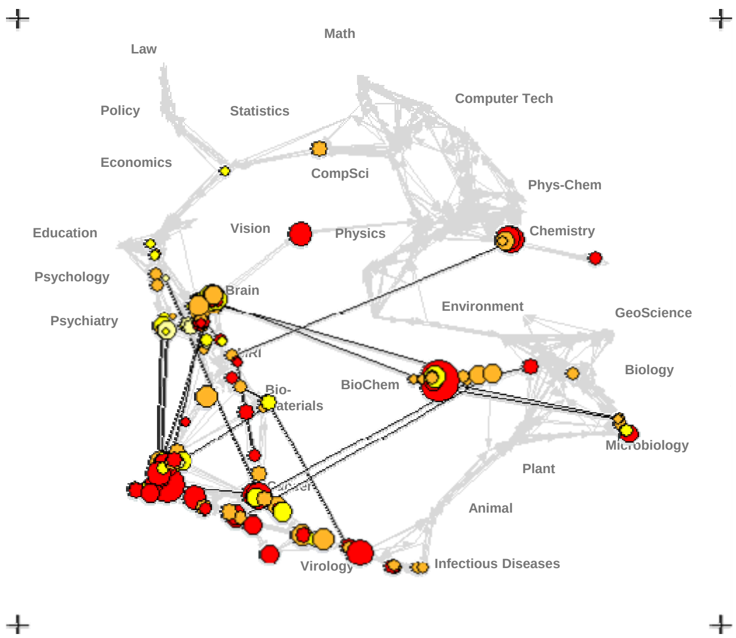
Funding Patterns of the National Science Foundation (NSF)



Science map applications: Identifying core competency

Kevin W. Boyack, Katy Börner, & Richard Klavans (2007).

Funding Patterns of the National Institutes of Health (NIH)



Science of Science Cyberinfrastructure

Overview

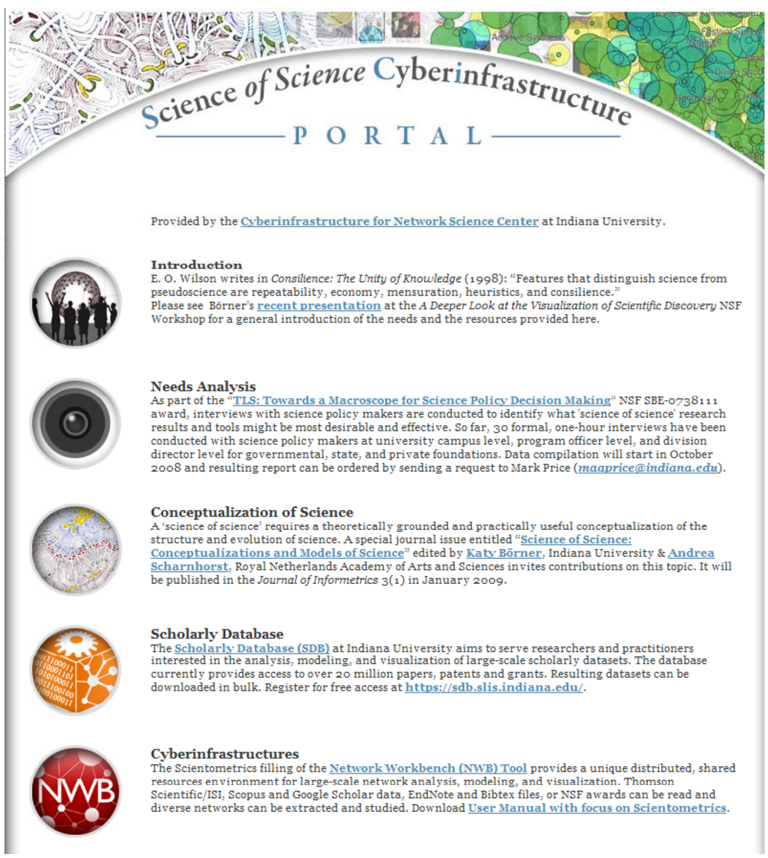
What cyberinfrastructure will be required to measure, model, analyze, and communicate scholarly data and, ultimately, scientific progress?

This talk presents our efforts to create a science of science cyberinfrastructure that supports:

- Data access and federation via the **Scholarly Database**, http://sdb.slis.indiana.edu,
- Data preprocessing, modeling, analysis, and visualization using plug-and-play cyberinfrastructures such as the **Network Workbench**, http://nwb.slis.indiana.edu, and
- Communication of science to a general audience via the **Mapping Science Exhibit** at http://scimaps.org.

The following demos should be particularly interesting for those interested to

- Map their very own domain of research,
- Test and compare data federation, mining, visualization algorithms on large scale datasets,
- Use advanced network science algorithms in their own research.



Science of Science Cyberinfrastructure

PORTAL

Provided by the Cyberinfrastructure for Network Science Center at Indiana University.

Introduction
E. O. Wilson writes in *Consilience: The Unity of Knowledge* (1998): "Features that distinguish science from pseudoscience are repeatability, economy, mensuration, heuristics, and consilience."
Please see Börner's recent presentation at the *A Deeper Look at the Visualization of Scientific Discovery* NSF Workshop for a general introduction of the needs and the resources provided here.

Needs Analysis
As part of the "TLS: Towards a Macroscope for Science Policy Decision Making" NSF SBE-0738111 award, interviews with science policy makers are conducted to identify what 'science of science' research results and tools might be most desirable and effective. So far, 30 formal, one-hour interviews have been conducted with science policy makers at university campus level, program officer level, and division director level for governmental, state, and private foundations. Data compilation will start in October 2008 and resulting report can be ordered by sending a request to Mark Price (*maaprice@indiana.edu*).

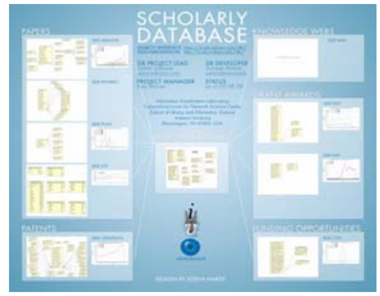
Conceptualization of Science
A 'science of science' requires a theoretically grounded and practically useful conceptualization of the structure and evolution of science. A special journal issue entitled "Science of Science: Conceptualizations and Models of Science" edited by Katy Börner, Indiana University & Andrea Scharnhorst, Royal Netherlands Academy of Arts and Sciences invites contributions on this topic. It will be published in the *Journal of Informetrics* 3(1) in January 2009.

Scholarly Database
The Scholarly Database (SDB) at Indiana University aims to serve researchers and practitioners interested in the analysis, modeling, and visualization of large-scale scholarly datasets. The database currently provides access to over 20 million papers, patents and grants. Resulting datasets can be downloaded in bulk. Register for free access at https://sdb.slis.indiana.edu/.

Cyberinfrastructures
The Scientometrics filling of the Network Workbench (NWB) Tool provides a unique distributed, shared resources environment for large-scale network analysis, modeling, and visualization. Thomson Scientific/ISI, Scopus and Google Scholar data, EndNote and Bibtex files, or NSF awards can be read and diverse networks can be extracted and studied. Download User Manual with focus on Scientometrics.

http://sci.slis.indiana.edu

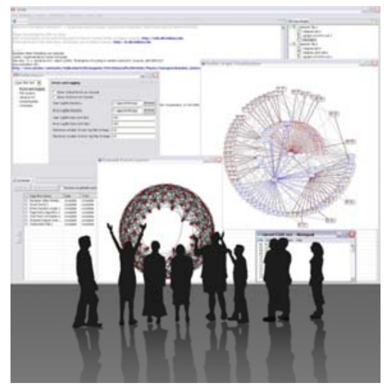
Cyberinfrastructures for a Science of Science



Scholarly Database of 23 million scholarly records
https://sdb.slis.indiana.edu

James S. McDonnell Foundation

Information Visualization Cyberinfrastructure
http://iv.slis.indiana.edu



Network Workbench Tool and Community Wiki
NEW Scientometrics plugins
http://nwb.slis.indiana.edu

Epidemics Cyberinfrastructure
http://epic.slis.indiana.edu/

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

Nianli Ma

"From Data Silos to Wind Chimes"

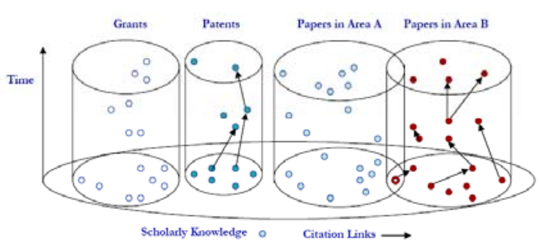
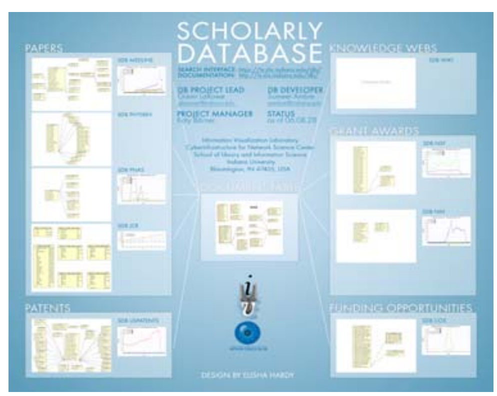


Figure 1: The interoperability and cross linkage problem. Many but not all of today's scholarly datasets, e.g., papers, patents, grants, are stored and made available so that 'vertical' citation linkages can be traversed. There are very few instances in which datasets of different origin and/or type are 'horizontally' interlinked.

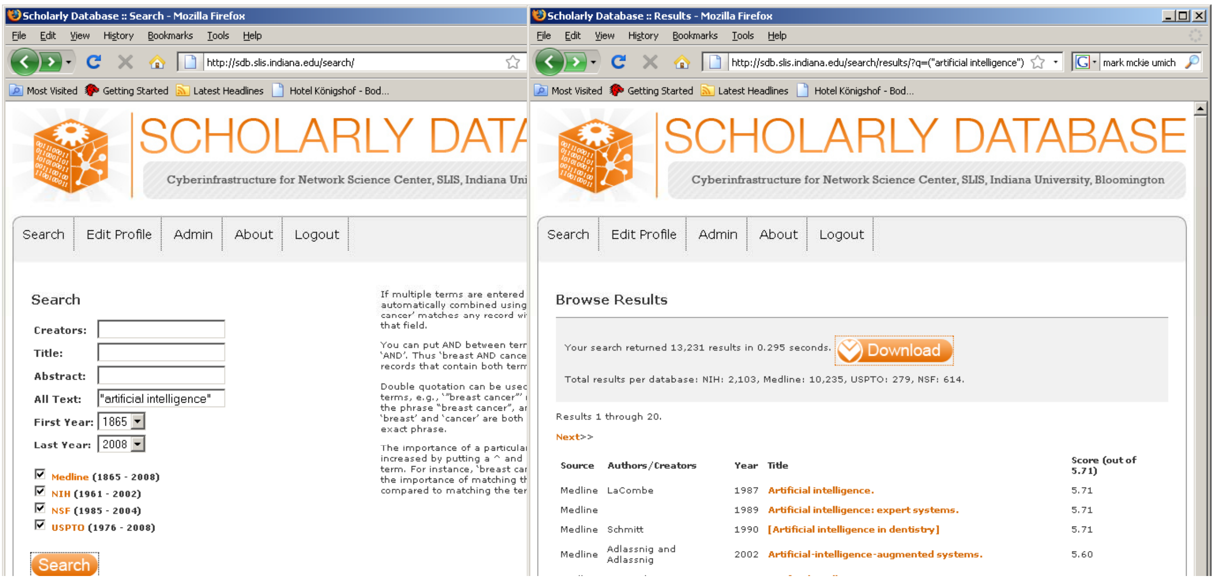


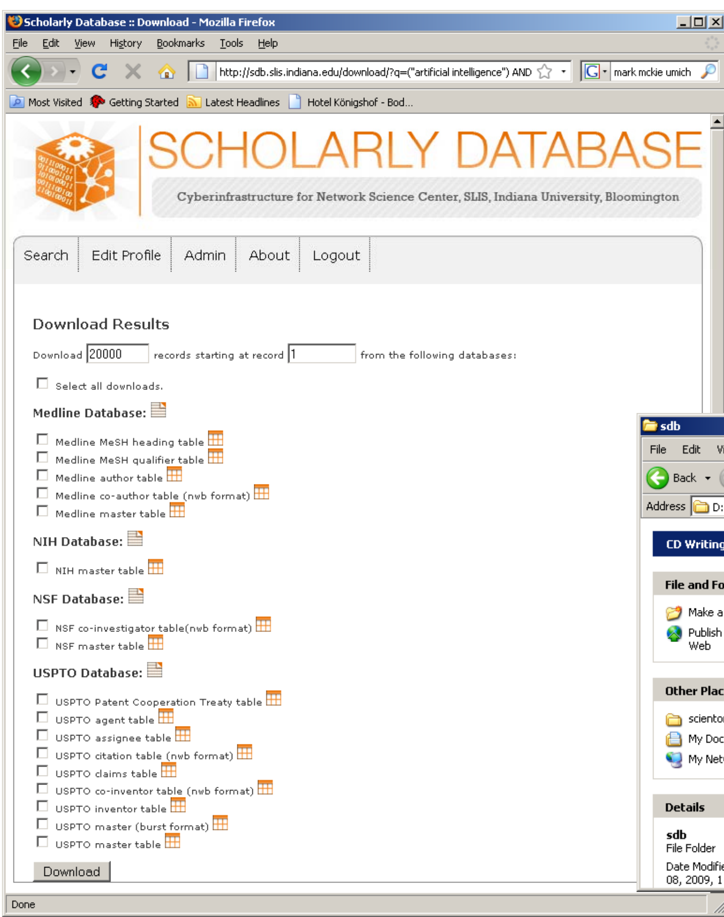
- Create public databases that any scholar can use. Share the burden of data cleaning and federation.
- Interlink creators, data, software/tools, publications, patents, funding, etc.

La Rowe, Gavin, Ambre, Sumeet, Burgoon, John, Ke, Weimao and Börner, Katy. (2007) The Scholarly Database and Its Utility for Scientometrics Research. In Proceedings of the 11th International Conference on Scientometrics and Informetrics, Madrid, Spain, June 25-27, 2007, pp. 457-462. http://ella.slis.indiana.edu/~katy/paper/07-issi-sdb.pdf

Scholarly Database: Web Interface

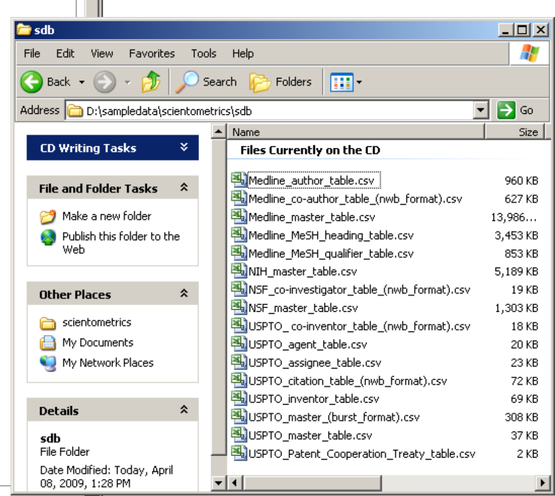
Anybody can register for free to search the about 23 million records and download results as data dumps.
Currently the system has over 120 registered users from academia, industry, and government from over 60 institutions and four continents.





Since March 2009:
Users can download networks:
- Co-author
- Co-investigator
- Co-inventor
- Patent citation

and tables for burst analysis in NWB.



Scholarly Database: # Records & Years Covered

Datasets available via the Scholarly Database (* internally)

Dataset	# Records	Years Covered	Updated	Restricted Access
Medline	17,764,826	1898-2008	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, 710,952	1976-2008	Yes*	
NSF	174,835	1985-2002	Yes*	
NIH	1,043,804	1961-2002	Yes*	
Total	**23,167,642**	**1893-2006**	**4**	**3**

Aim for comprehensive time, geospatial, and topic coverage.

NetworkWorkbench

Investigators: Katy Börner, Albert-Laszlo Barabasi, Santiago Schnell, Alessandro Vespignani & Stanley Wasserman, Eric Wernert

Software Team: Lead: Micah Linnemeier
Members: Patrick Phillips, Russell Duhon, Tim Kelley & Ann McCranie
Previous Developers: Weixia (Bonnie) Huang, Bruce Herr, Heng Zhang, Duygu Balcan, Mark Price, Ben Markines, Santo Fortunato, Felix Terkhorn, Ramya Sabbineni, Vivek S. Thakre & Cesar Hidalgo

Goal: Develop a large-scale network analysis, modeling and visualization toolkit for physics, biomedical, and social science research.

Amount: $1,120,926, NSF IIS-0513650 award

Duration: Sept. 2005 - Aug. 2009

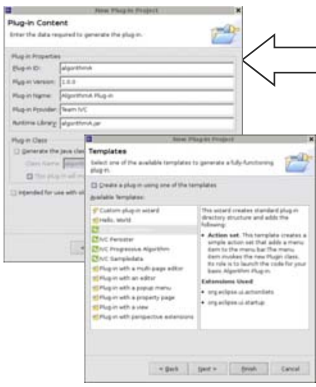
Website: http://nwb.slis.indiana.edu

Serving Non-CS Algorithm Developers & Users

Developers

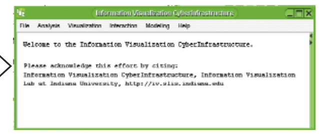
Users

CIShell Wizards

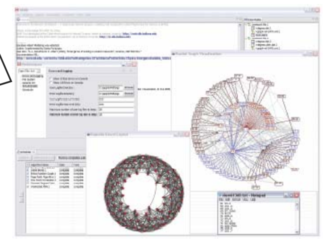


CIShell

IVC Interface



NWB Interface



Katy Börner: Mapping the Structure and Evolution of Science

NWB Tool: Algorithms (July 1st, 2008)

See https://nwb.slis.indiana.edu/community and handout for details.

Algorithms Currently Available

Preprocessing Edit

- **Remove Nodes**
  - Extract Top Nodes
  - Extract Nodes Above or Below Val
  - Delete High Degree Nodes
  - Delete Random Nodes
  - Delete Isolates
- **Remove Edges**
  - Extract Top Edges
  - Extract Edges Above or Below Val
  - Remove Self Loops
  - Trim By Degree[?]
  - Pathfinder Network Scaling
- **Sampling**
  - Snowball Sampling (n nodes)
  - Node Sampling
  - Edge Sampling
- **Transformations**
  - Symmetrize
  - Dichotomize
  - Multipartite Joining

Modeling Edit

- **General**
  - Random Graph
  - Watts-Strogatz Small World
  - Barabási-Albert Scale-Free
- **Structured**
  - CAN
  - Chord
- **Unstructured**
  - Hypergrid
  - PRU
- **Other**

Analysis Edit

- **General Purpose**
  - Network Analysis Toolkit[?]
- **Unweighted & Undirected**
  - **Based on degree/**
    - Node Degree
    - Node Distribution
  - **Based on clustering**
    - k-Nearest Neighbor
    - Watts Strogatz Clustering Coefficie
    - Watts Strogatz Clustering Coefficie
  - **Based on path**
    - Diameter
    - Average Shortest Path
    - Shortest Path Distribution
    - Node Betweenness Centrality
  - **Based on components**
    - Connected Components
    - Weak Component Clustering
  - **K-Core**
    - Extract K-Core[?]
    - Annotate K-Coreness[?]
- **Unweighted & Directed**
  - **Based on degree**
    - Node Indegree
    - Node Outdegree
    - Indegree Distribution
    - Outdegree Distribution
  - **Based on local graph structure**
    - k-Nearest Neighbor
    - Single Node In-Out Degree Correla
  - **Unnamed Category?**
    - Page Rank
  - **Based on local graph structure**

Visualization Edit

- **Tools**
  - GUESS
  - GnuPlot[?]
- **Predefined Positions Layout**
  - DrL (VxOrd)
  - Pre-defined Positions (prefuse beta)[?]
- **Move**
  - Circular
- **Tree Layouts**
  - Radial Tree (prefuse alpha)
  - Radial Tree with Annotations (prefuse beta)[?]
  - Tree Map
  - Tree View
  - Balloon Graph (prefuse alpha)[?]
- **Network Layouts**
  - Force Directed with Annotation (prefuse beta)
  - Kamada-Kawai (JUNG)
  - Fruchterman-Reingold (JUNG)
  - Fruchterman-Reingold with Annotation (prefuse beta)
  - Spring (JUNG)
  - Small World (prefuse alpha)
- **Other Layouts**
  - Parallel Coordinates (demo)[?]
  - LaNet (k-Core Decomposition)

Scientometrics Edit

- **Extract Network From Table**
  - Extract Co-Authorship Network
  - Extract Co-Occurrence Network From Table[?]
  - Extract Directed Network From Table[?]
- **Extract Network From Another Network**
  - Extract Bibliographic Coupling Similarity Network
  - Extract Co-Citation Similarity Network[?]

Scientometrics Filling of Network Workbench Tool

will ultimately be 'packaged' as a SciPolicy' tool.

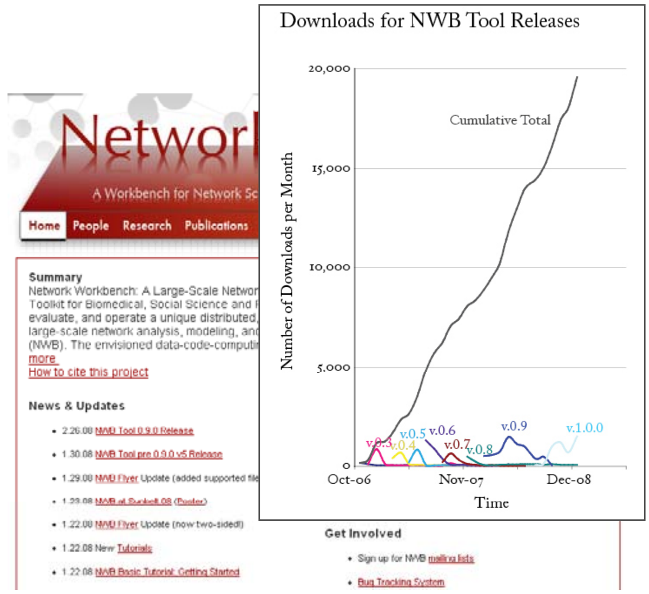
http://nwb.slis.indiana.edu/

The Network Workbench (NWB) tool supports researchers, educators, and practitioners interested in the study of biomedical, social and behavioral science, physics, and other networks.

In Feb. 2009, the tool provides more 100 plugins that support the preprocessing, analysis, modeling, and visualization of networks.

More than 40 of these plugins can be applied or were specifically designed for S&T studies.

It has been downloaded more than 19,000 times since Dec. 2006.



Herr II, Bruce W., Huang, Weixia (Bonnie), Penumarthy, Shashikant & Börner, Katy. (2007). Designing Highly Flexible and Usable Cyberinfrastructures for Convergence. In Bainbridge, William S. & Roco, Mihail C. (Eds.), Progress in Convergence - Technologies for Human Wellbeing (Vol. 1093, pp. 161-179), Annals of the New York Academy of Sciences, Boston, MA.

NWB Tool: Supported Data Formats

Personal Bibliographies

- Bibtex (.bib)
- Endnote Export Format (.enw)

Data Providers

- Web of Science by Thomson Scientific/Reuters (.isi)
- Scopus by Elsevier (.scopus)
- Google Scholar (access via *Publish or Perish* save as CSV, Bibtex, EndNote)
- Awards Search by National Science Foundation (.nsf)

Scholarly Database (all text files are saved as .csv)

- Medline publications by National Library of Medicine
- NIH funding awards by the National Institutes of Health (NIH)
- NSF funding awards by the National Science Foundation (NSF)
- U.S. patents by the United States Patent and Trademark Office (USPTO)
- Medline papers – NIH Funding

Network Formats

- NWB (.nwb)
- Pajek (.net)
- GraphML (.xml or .graphml)
- XGMML (.xml)

Burst Analysis Format

- Burst (.burst)

Other Formats

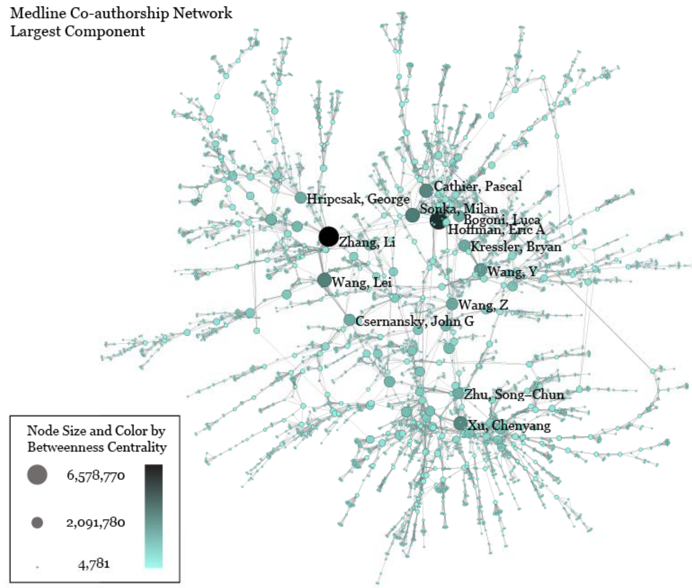
- CSV (.csv)
- Edgelist (.edge)
- Pajek (.mat)
- TreeML (.xml)

NWB Tool: Output Formats

NWB tool can be used for data conversion. Supported output formats comprise:
- CSV (.csv)
- NWB (.nwb)
- Pajek (.net)
- Pajek (.mat)
- GraphML (.xml or .graphml)
- XGMML (.xml)

GUESS
- Supports export of images into common image file formats.

Horizontal Bar Graphs
- saves out raster and ps files.



Exemplary Analyses and Visualizations

Individual Level

A. Loading ISI files of major network science researchers, extracting, analyzing and visualizing paper-citation networks and co-author networks.

B. Loading NSF datasets with currently active NSF funding for 3 researchers at Indiana U

Institution Level

C. Indiana U, Cornell U, and Michigan U, extracting, and comparing Co-PI networks.

Scientific Field Level

D. Extracting co-author networks, patent-citation networks, and detecting bursts in SDB data.

Exemplary Analyses and Visualizations

Individual Level

A. Loading ISI files of major network science researchers, extracting, analyzing and visualizing paper-citation networks and co-author networks.

B. Loading NSF datasets with currently active NSF funding for 3 researchers at Indiana U

Institution Level

C. Indiana U, Cornell U, and Michigan U, extracting, and comparing Co-PI networks.

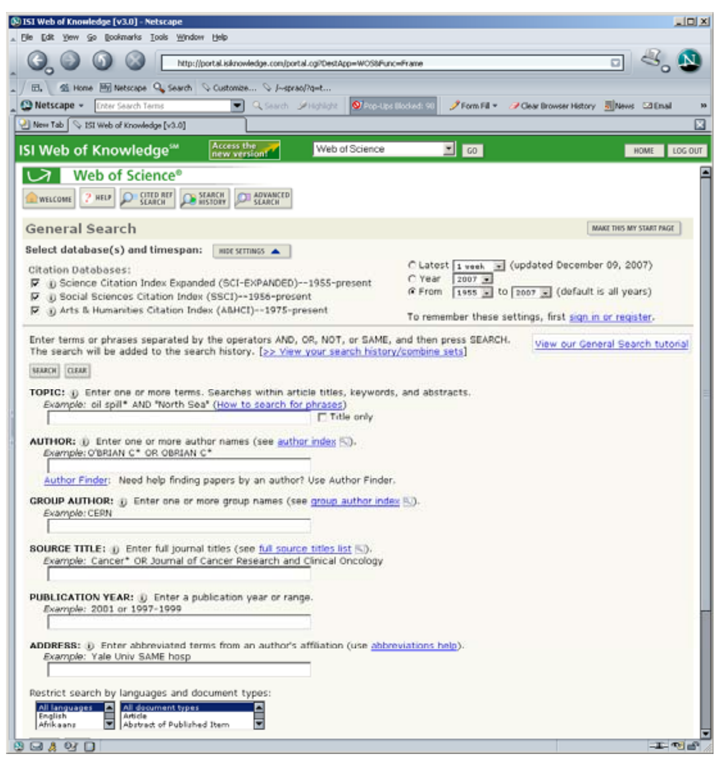
Scientific Field Level

D. Extracting co-author networks, patent-citation networks, and detecting bursts in SDB data.

Data Acquisition from Web of Science

Download all papers by
- Eugene Garfield
- Stanley Wasserman
- Alessandro Vespignani
- Albert-László Barabási

from
- Science Citation Index Expanded (SCI-EXPANDED) --1955-present
- Social Sciences Citation Index (SSCI)--1956-present
- Arts & Humanities Citation Index (A&HCI)--1975-present



Comparison of Counts

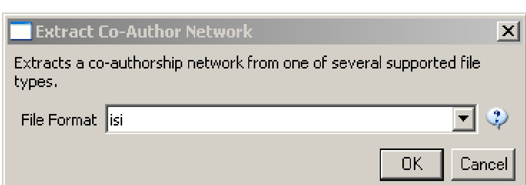
No books and other non-WoS publications are covered.

	Age	Total # Cites	Total # Papers	H-Index
Eugene Garfield	82	1,525	672	31
Stanley Wasserman		122	35	17
Alessandro Vespignani	42	451	101	33
Albert-László Barabási	40	2,218	126	47 *(Dec 2007)*
	41	16,920	159	52 *(Dec 2008)*

Extract Co-Author Network

Load **yournwbdirectory*/sampledata/scientometrics/isi/FourNetSciResearchers.isi'* using *'File > Load and Clean ISI File'*.

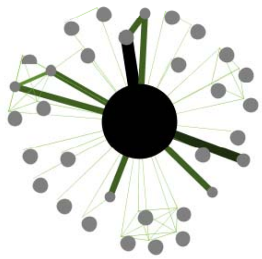
To extract the co-author network, select the *'361 Unique ISI Records'* table and run *'Scientometrics > Extract Co-Author Network'* using isi file format:



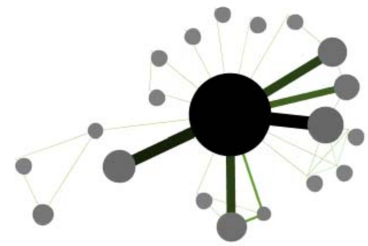
The result is an undirected network of co-authors in the Data Manager. It has 247 nodes and 891 edges.

To view the complete network, select the network and run *'Visualization > GUESS > GEM'*. Run *Script > Run Script… . And select Script folder > GUESS > co-author-nw.py*.

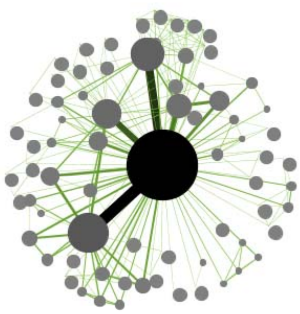
Comparison of Co-Author Networks



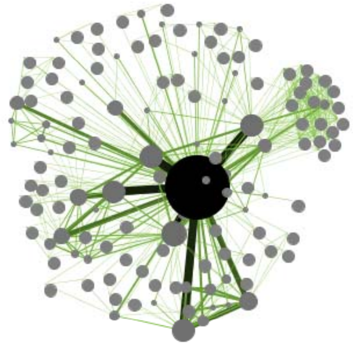
Eugene Garfield



Stanley Wasserman

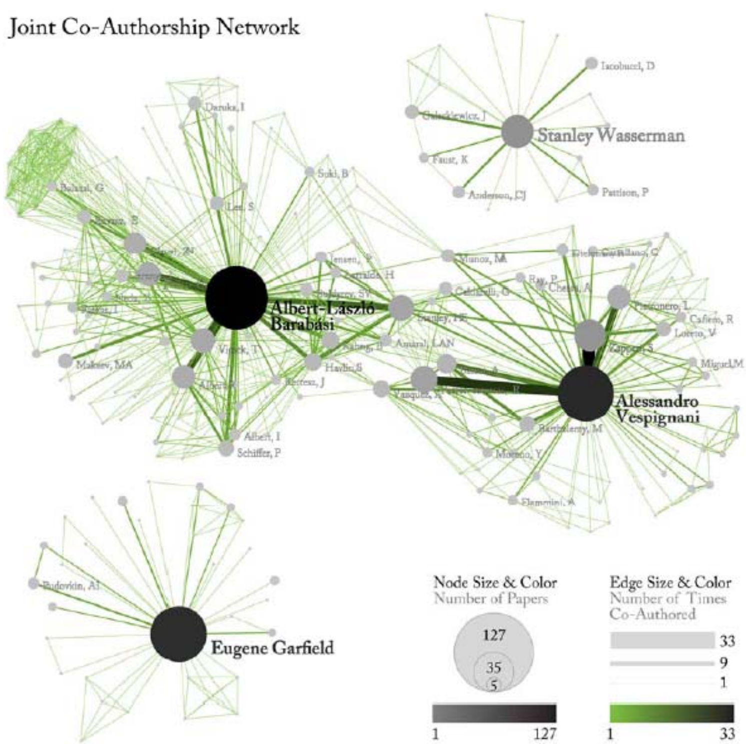


Alessandro Vespignani



Albert-László Barabási

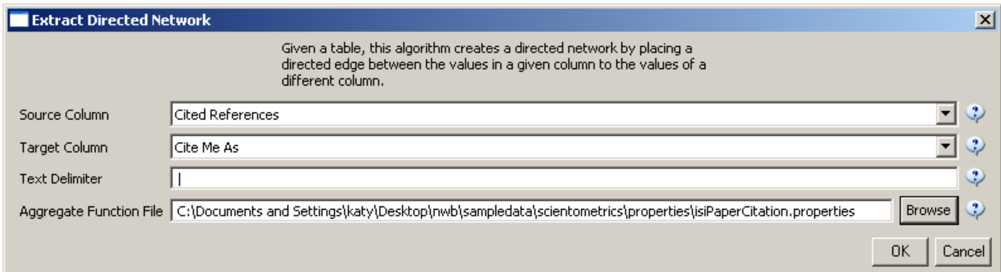
Joint Co-Author Network of all Four NetsSci Researchers



Paper-Citation Network Layout

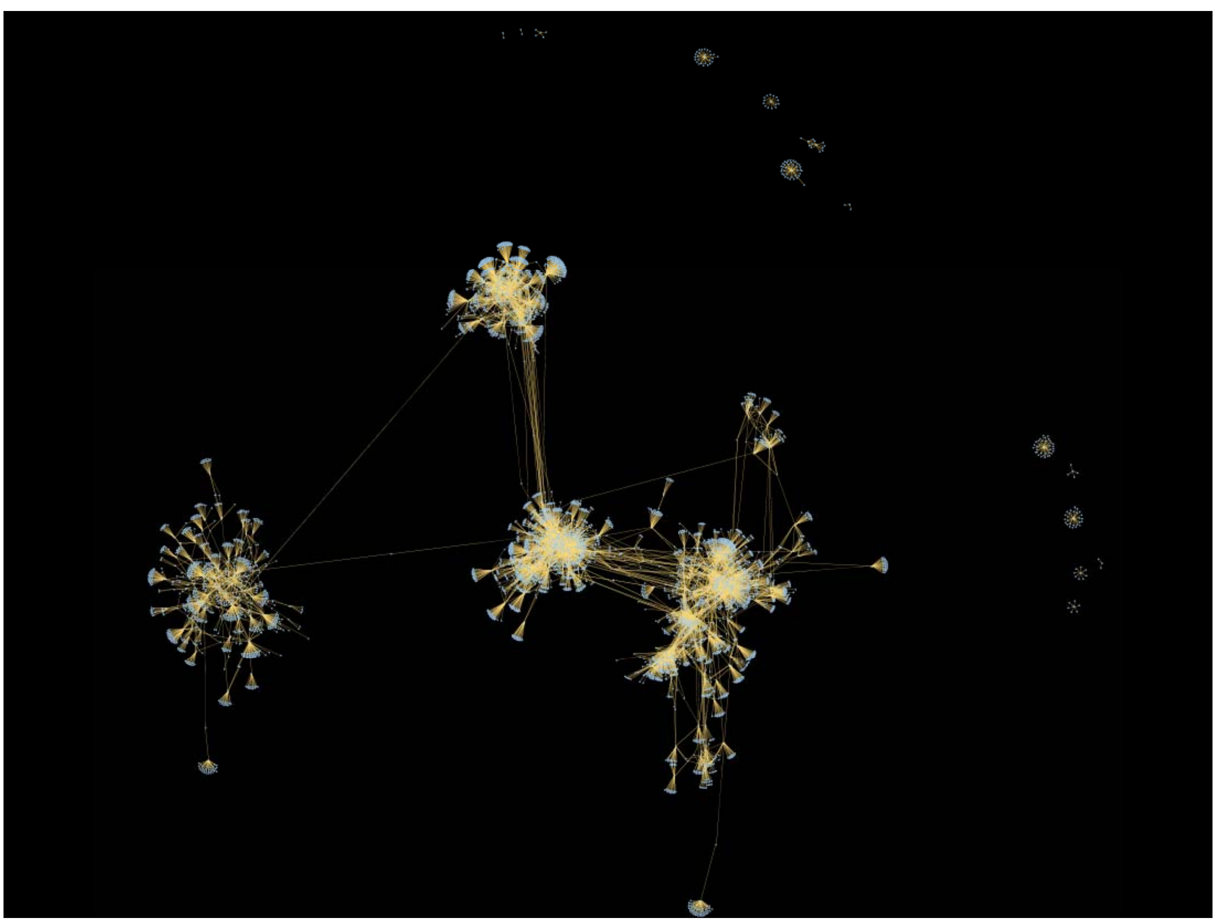
Load *'*yournwbdirectory*/sampledata/scientometrics/isi/FourNetSciResearchers.isi'* using *'File > Load and Clean ISI File'*.

To extract the paper-citation network, select the *'361 Unique ISI Records'* table and run *'Scientometrics > Extract Directed Network'* using the parameters:



The result is a directed network of paper citations in the Data Manager. It has 5,335 nodes and 9,595 edges.

To view the complete network, select the network and run *'Visualization > GUESS'*. Run '*Script > Run Script …' and select 'yournwbdirectory*/script/GUESS/paper-citation-nw.py'*.



Exemplary Analyses and Visualizations

Individual Level

A. Loading ISI files of major network science researchers, extracting, analyzing and visualizing paper-citation networks and co-author networks.

B. Loading NSF datasets with currently active NSF funding for 3 researchers at Indiana U

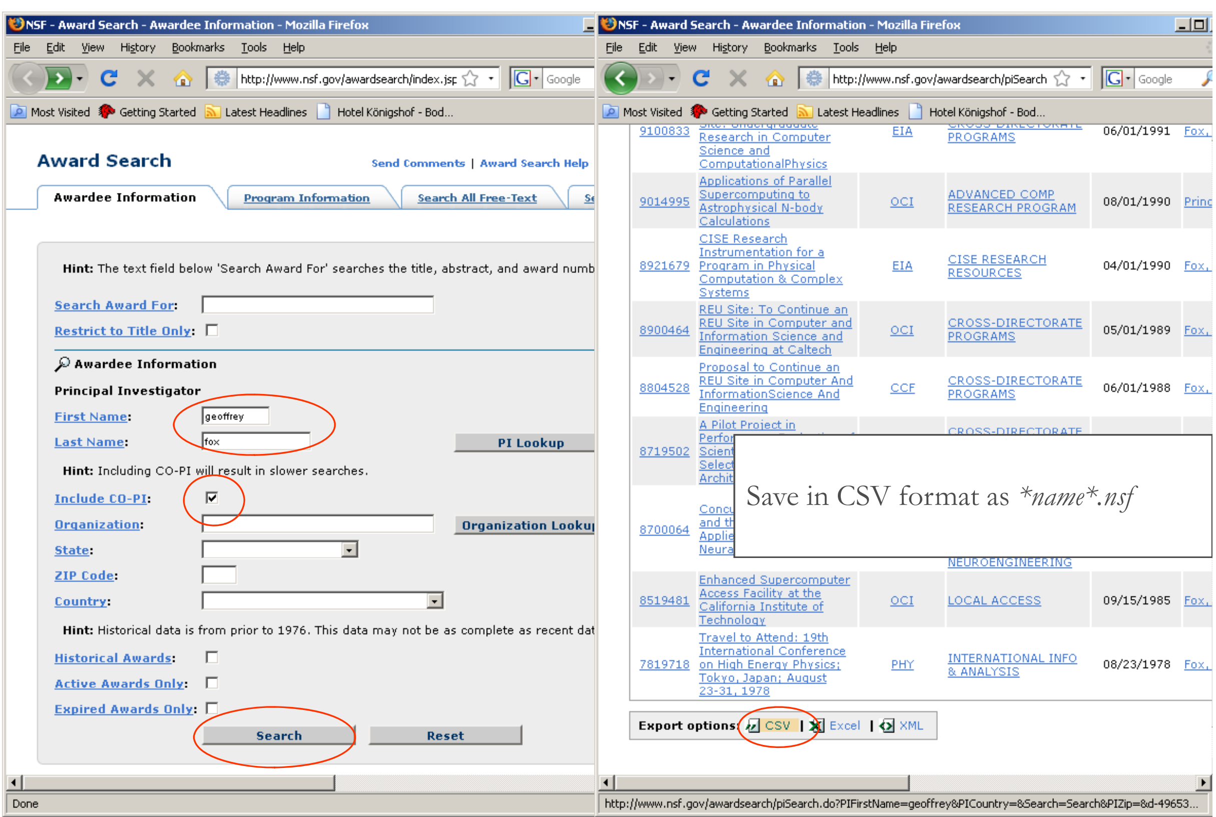
Institution Level

C. Indiana U, Cornell U, and Michigan U, extracting, and comparing Co-PI networks.

Scientific Field Level

D. Extracting co-author networks, patent-citation networks, and detecting bursts in SDB data.

NSF Awards Search via http://www.nsf.gov/awardsearch



NSF Awards Search Results

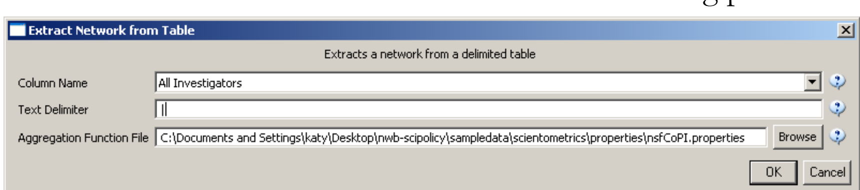
Name	# Awards	First Award Starts	Total Amount to Date
Geoffrey Fox	27	Aug 1978	$12,196,260
Michael McRobbie	8	July 1997	$19,611,178
Beth Plale	10	Aug 2005	$7,224,522

Disclaimer:

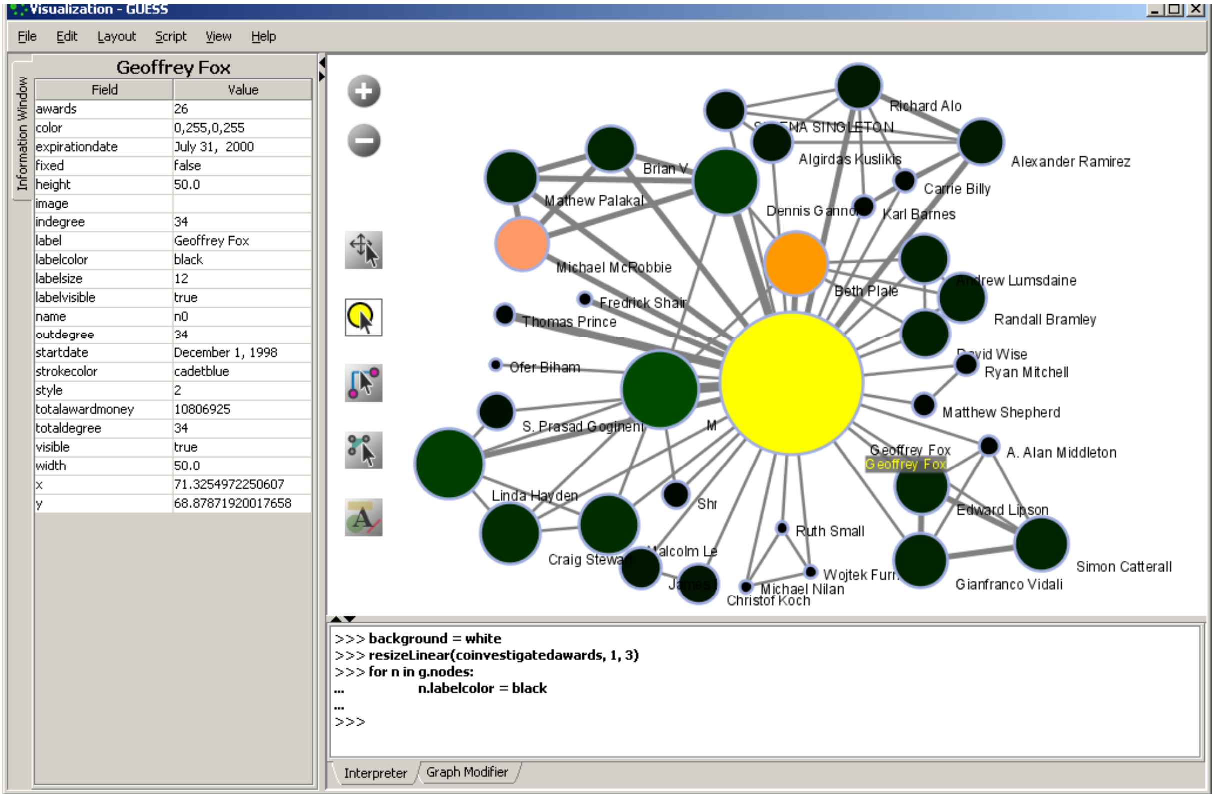
Only NSF funding, no funding in which they were senior personnel, only as good as NSF's internal record keeping and unique person ID. If there are 'collaborative' awards then only their portion of the project (award) will be included.

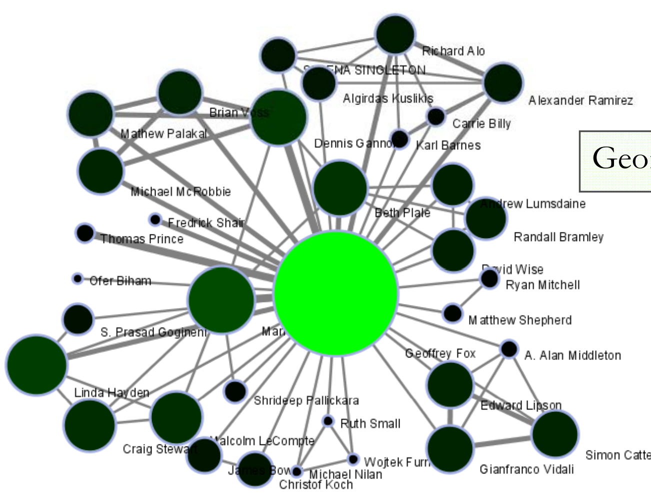
Using NWB to Extract Co-PI Networks

- Load into NWB, open file to count records, compute total award amount.
- Run '*Scientometrics > Extract Co-Occurrence Network*' using parameters:

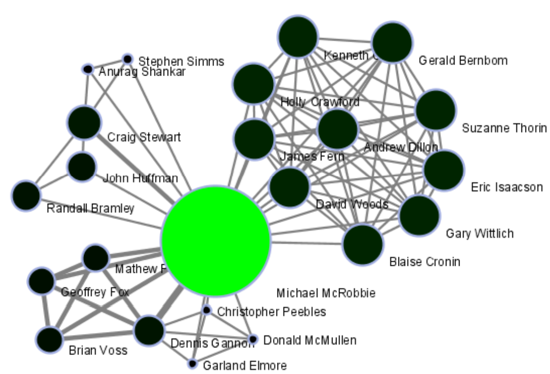


- Select *"Extracted Network .."* and run *'Analysis > Network Analysis Toolkit (NAT)'*
- Remove unconnected nodes via *'Preprocessing > Delete Isolates'*.
- *'Visualization > GUESS'*, layout with GEM
- Run 'co-PI-nw.py' GUESS script to color/size code.

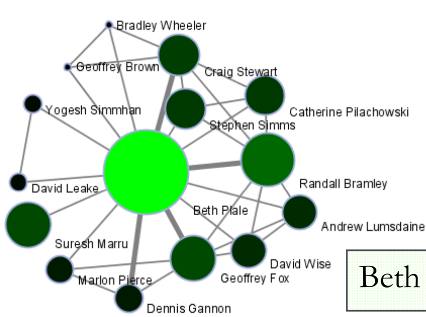




Geoffrey Fox



Michael McRobbie



Beth Plale

Geoffrey Fox

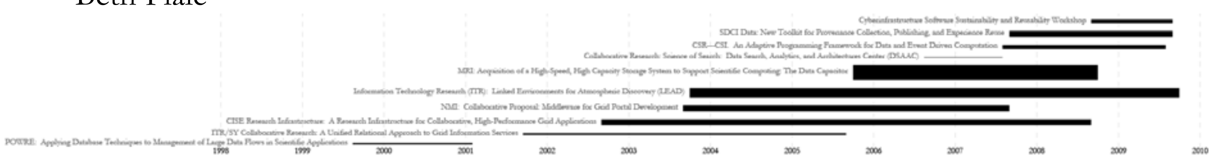
Last Expiration date

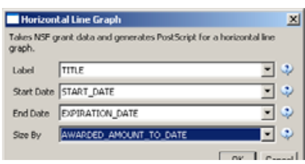
July 10

Michael McRobbie

Feb 10

Beth Plale





Sept 09

Exemplary Analyses and Visualizations

Individual Level

A. Loading ISI files of major network science researchers, extracting, analyzing and visualizing paper-citation networks and co-author networks.

B. Loading NSF datasets with currently active NSF funding for 3 researchers at Indiana U

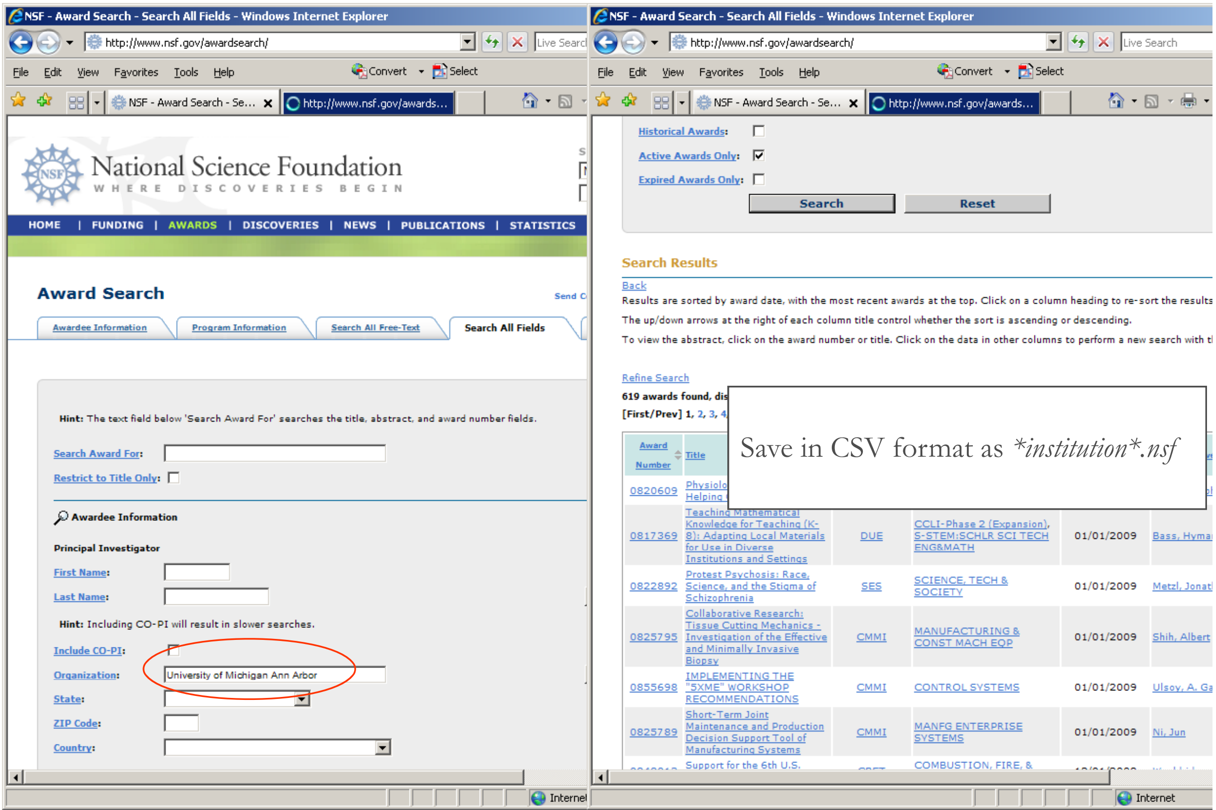
Institution Level

C. Indiana U, Cornell U, and Michigan U, extracting, and comparing Co-PI networks.

Scientific Field Level

D. Extracting co-author networks, patent-citation networks, and detecting bursts in SDB data.

NSF Awards Search via http://www.nsf.gov/awardsearch



Active NSF Awards on 11/07/2008:

- Indiana University 257

  (there is also Indiana University at South Bend Indiana University Foundation, Indiana University Northwest, Indiana University-Purdue University at Fort Wayne, Indiana University-Purdue University at Indianapolis, Indiana University-Purdue University School of Medicine)

- Cornell University 501

  (there is also Cornell University – State, Joan and Sanford I. Weill Medical College of Cornell University)

- University of Michigan Ann Arbor 619

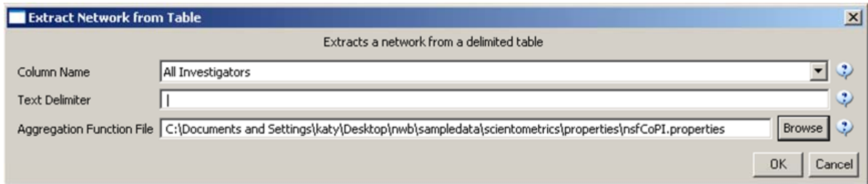
  (there is also University of Michigan Central Office, University of Michigan Dearborn, University of Michigan Flint, University of Michigan Medical School)

Save files as csv but rename into .nsf.

Or simply use the files saved in '*yournwbdirectory*/sampledata/scientometrics/nsf/'.

Extracting Co-PI Networks

Load NSF data, selecting the loaded dataset in the Data Manager window, run *'Scientometrics > Extract Co-Occurrence Network'* using parameters:

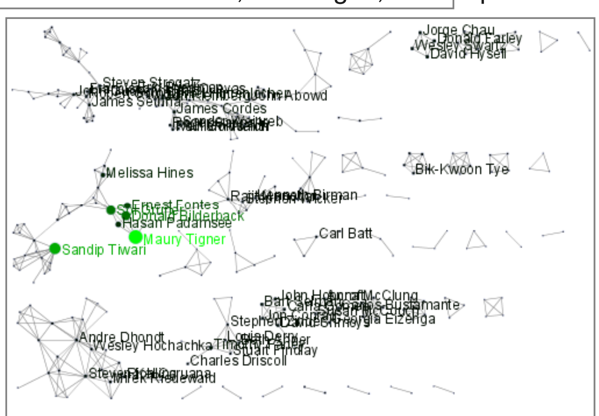


Two derived files will appear in the Data Manager window: the co-PI network and a merge table. In the network, nodes represent investigators and edges denote their co-PI relationships. The merge table can be used to further clean PI names.

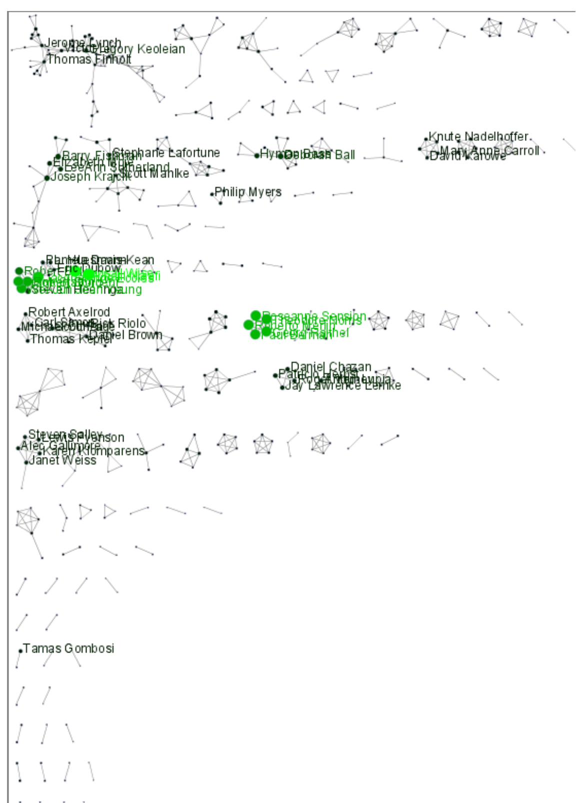
Running the *'Analysis > Network Analysis Toolkit (NAT)'* reveals that the number of nodes and edges but also of isolate nodes that can be removed running *'Preprocessing > Delete Isolates'*.

Select *'Visualization > GUESS'* to visualize. Run 'co-PI-nw.py' script.

Indiana U: 223 nodes, 312 edges, 52 components



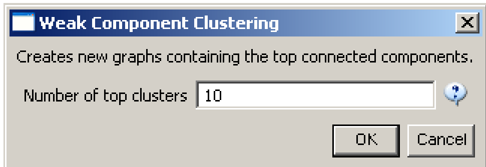
Cornell U: 375 nodes, 573 edges, 78 c



U of Michigan: 497 nodes, 672 edges, 117 c

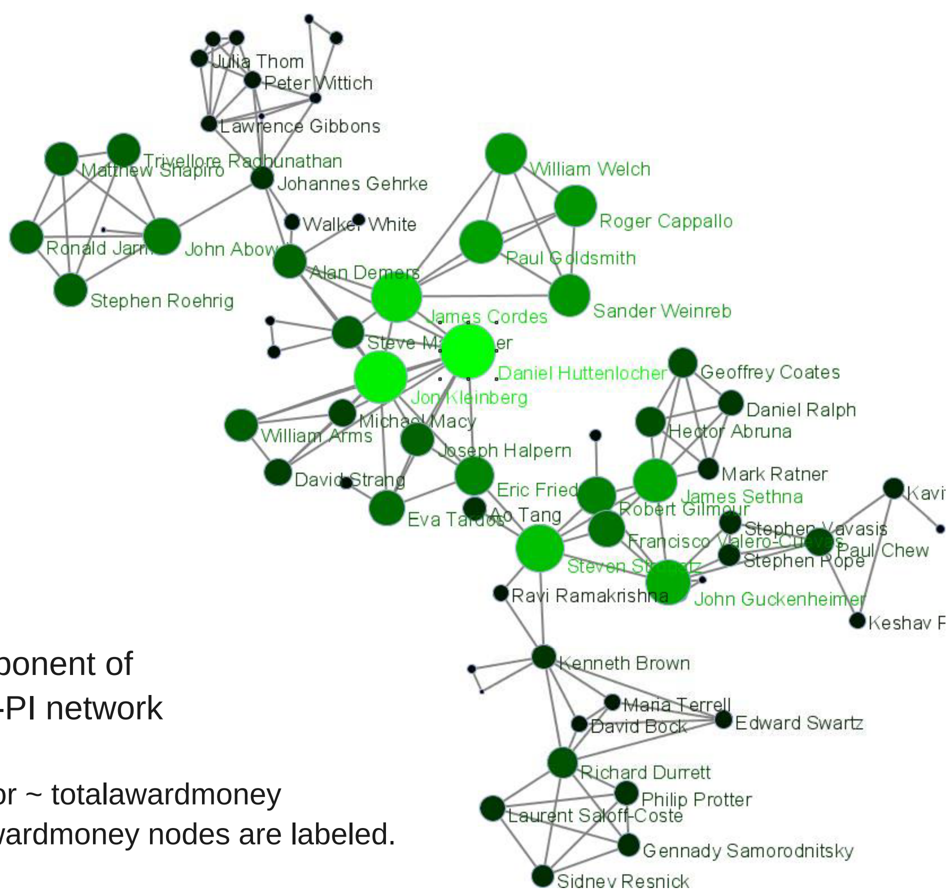
Extract Giant Component

Select network after removing isolates and run *'Analysis > Unweighted and Undirected > Weak Component Clustering'* with parameter



Indiana's largest component has 19 nodes, Cornell's has 67 nodes, Michigan's has 55 nodes.

Visualize Cornell network in GUESS using same .py script and save via *'File > Export Image'* as jpg.



Largest component of
Cornell U co-PI network

Node size/color ~ totalawardmoney
Top-50 totalawardmoney nodes are labeled.

Top-10 Investigators by Total Award Money

```
for i in range(0, 10):
    print str(nodesbytotalawardmoney[i].label) + ": " +
    str(nodesbytotalawardmoney[i].totalawardmoney)
```

Indiana University		Cornell University		Michigan University	
Curtis Lively:	7,436,828	Maury Tigner:	107,216,976	Khalil Najafi:	32,541,158
Frank Lester:	6,402,330	Sandip Tiwari:	72,094,578	Kensall Wise:	32,164,404
Maynard Thompson:	6,402,330	Sol Gruner:	48,469,991	Jacquelynne Eccles:	25,890,711
Michael Lynch:	6,361,796	Donald Bilderback:	47,360,053	Georg Raithel:	23,832,421
Craig Stewart:	6,216,352	Ernest Fontes:	29,380,053	Roseanne Sension:	23,812,921
William Snow:	5,434,796	Hasan Padamsee:	18,292,000	Theodore Norris:	23,35,0921
Douglas V. Houweling:	5,068,122	Melissa Hines:	13,099,545	Paul Berman:	23,350,921
James Williams:	5,068,122	Daniel Huttenlocher:	7,614,326	Roberto Merlin:	23,350,921
Miriam Zolan:	5,000,627	Timothy Fahey:	7,223,112	Robert Schoeni:	21,991,140
Carla Caceres:	5,000,627	Jon Kleinberg:	7,165,507	Wei-Jun Jean Yeung:	21,991,140

Exemplary Analyses and Visualizations

Individual Level

A. Loading ISI files of major network science researchers, extracting, analyzing and visualizing paper-citation networks and co-author networks.

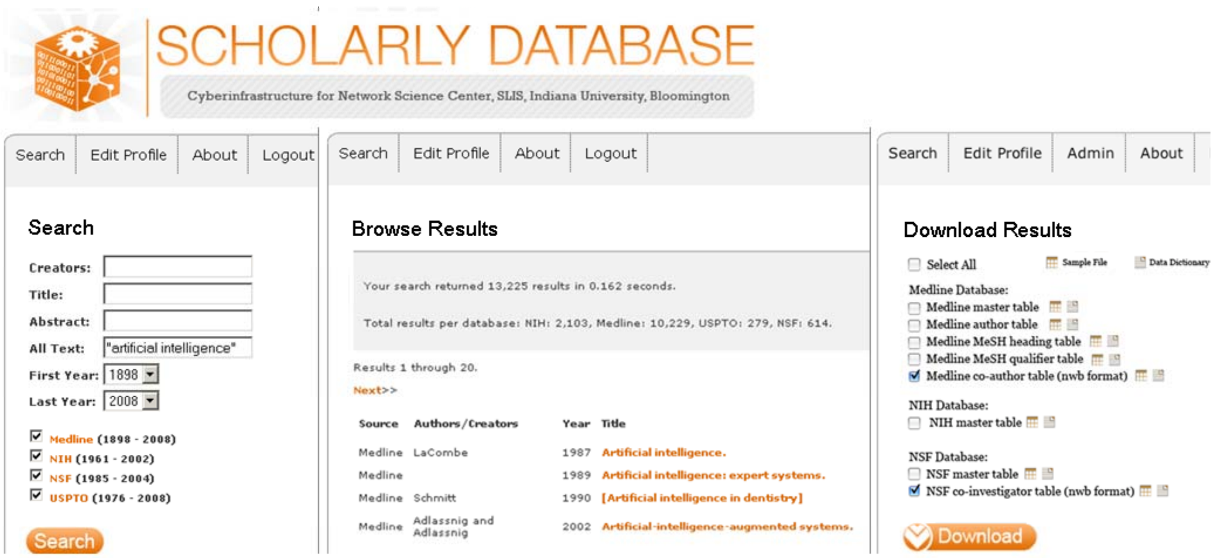
B. Loading NSF datasets with currently active NSF funding for 3 researchers at Indiana U

Institution Level

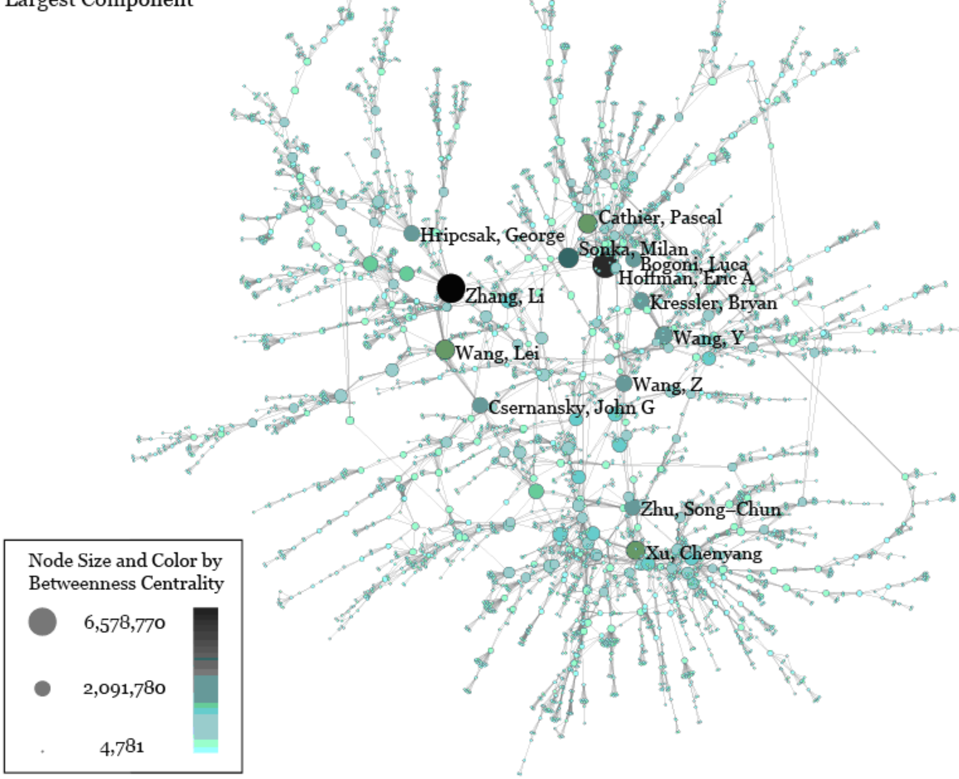
C. Indiana U, Cornell U, and Michigan U, extracting, and comparing Co-PI networks.

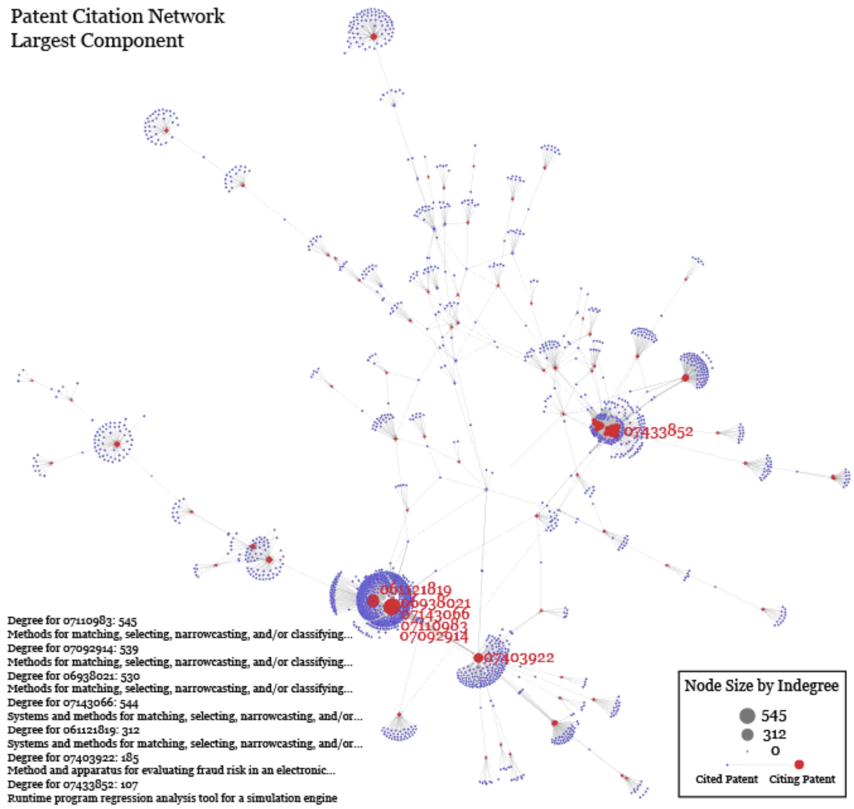
Scientific Field Level

D. Extracting co-author networks, patent-citation networks, and detecting bursts in SDB data.



Medline Co-authorship Network
Largest Component





Top-10 burst terms from abstracts of the AI search results.

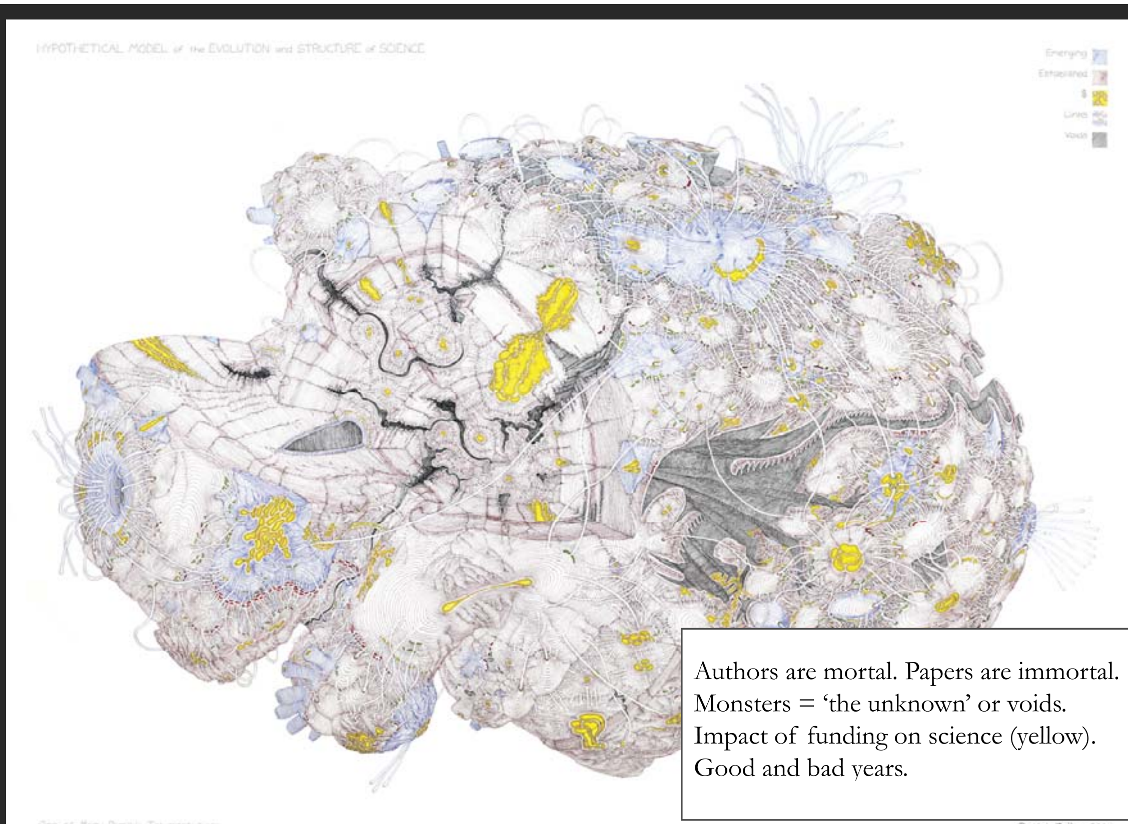
Medline				
Word	**Length**	**Weight**	**Start**	**End**
medical	17	299.7924	1983	1999
knowledge	5	293.9375	1991	1995
knowledge	6	215.2407	1997	2002
expert	13	171.0443	1985	1997
systems	15	170.3306	1985	1999
intelligence	21	123.9794	1981	2001
patient	21	123.9297	1982	2002
care	12	106.5522	1990	2001
registration	5	104.8139	2005	
knowledge-based	16	98.83778	1987	2002

NIH				
Word	**Length**	**Weight**	**Start**	**End**
Phase	8	117.2205	1993	2000
commercial	9	87.57158	1995	
proposed	9	87.57158	1995	
mass	3	83.36952	1978	1980
protein	1	72.15788	1988	1988
networks	4	71.252	1993	1996
patterns	3	66.44826	1977	1979
being	8	66.29254	1971	1978
reasoning	2	65.68178	1984	1985
expert	4	60.49935	1987	1990

NSF				
Word	**Length**	**Weight**	**Start**	**End**
their	6	47.05097	1999	
gray	2	28.19808	2000	2001
learning	2	27.40728	1997	1998
human	5	25.4525	2000	
control	2	24.07877	1992	1993
knowledge	1	21.48756	1998	1998
students	1	21.07674	1997	1997
problems	2	20.77133	1998	1999
more	2	19.96109	2000	2001
use	1	19.38503	2001	2001

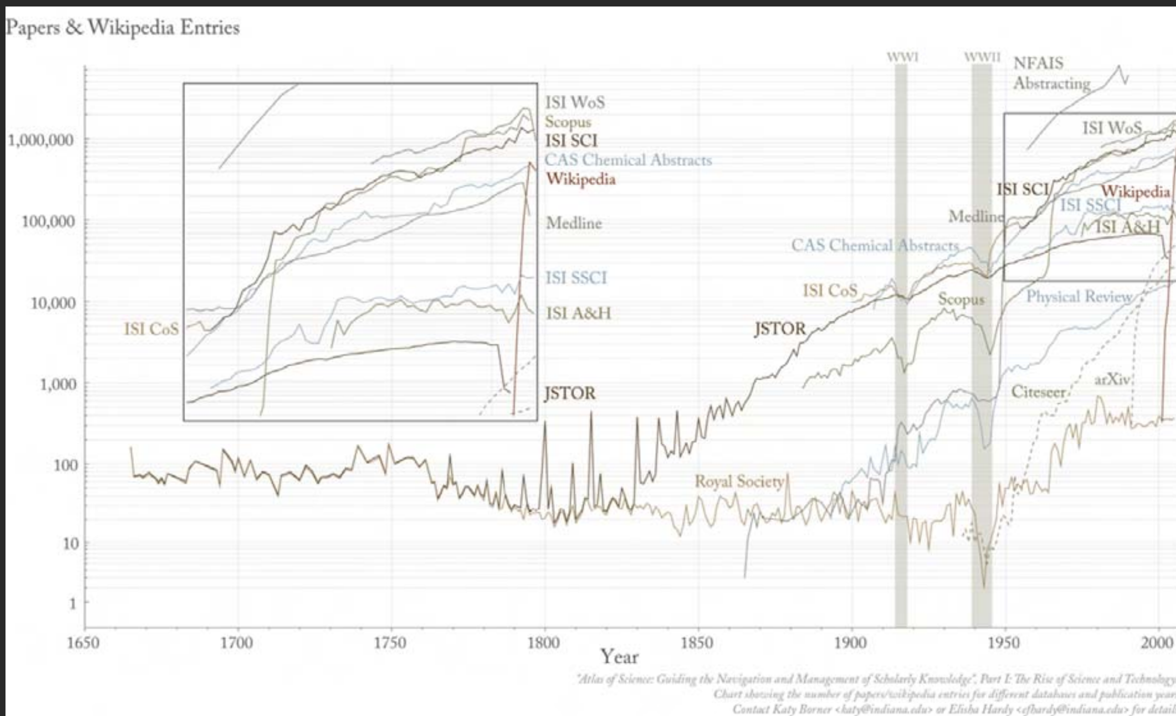
USPTO				
Word	**Length**	**Weight**	**Start**	**End**
human	3	19.03937321	2004	2006
video	3	15.32736425	1998	2000
disclosed	2	14.06694671	1999	2000
neural	3	13.30105906	2004	2006
"correct"	2	12.4336047	1999	2000
unit	2	12.35745838	2002	2003
material	1	12.08487035	2000	2000
feedback	1	12.07730195	2000	2000
rule	1	12.07730195	2000	2000
elevator	4	11.83351857	1991	1994

Conceptualizing Science

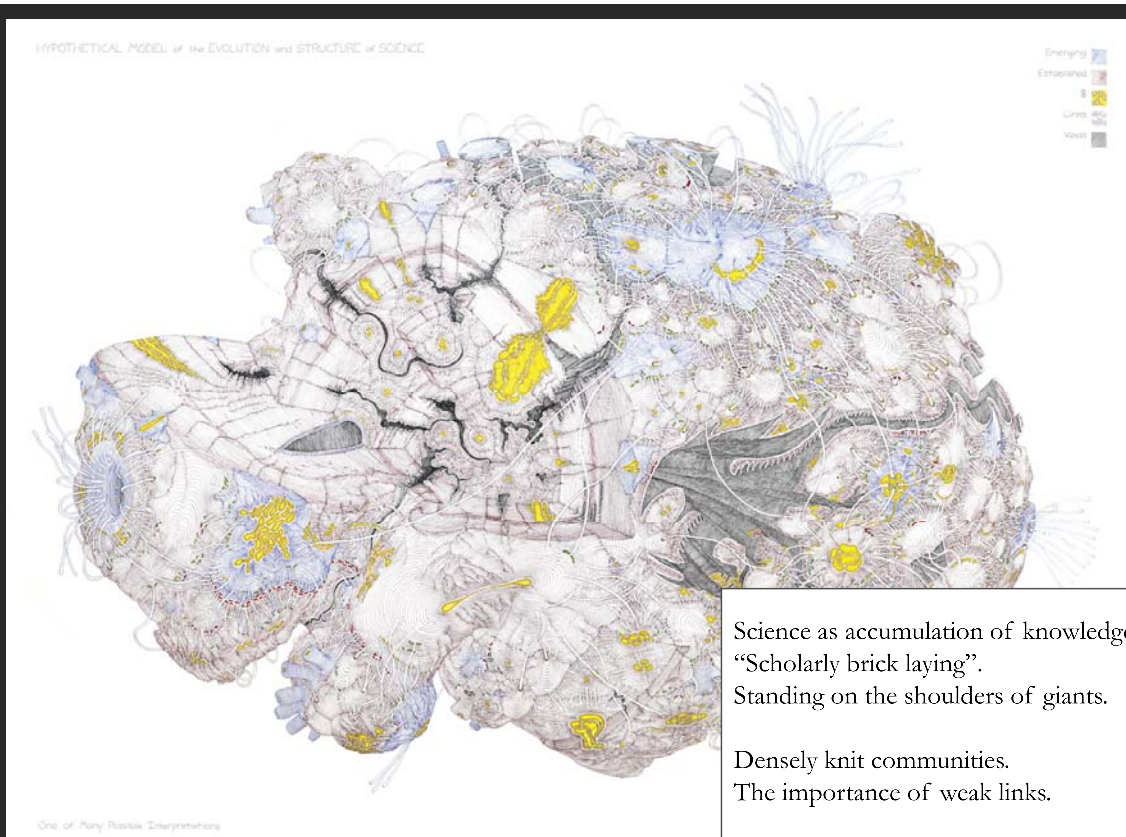


Authors are mortal. Papers are immortal.
Monsters = ‘the unknown’ or voids.
Impact of funding on science (yellow).
Good and bad years.

Hypothetical Model of the Evolution of Science - Daniel Zeller - 2007



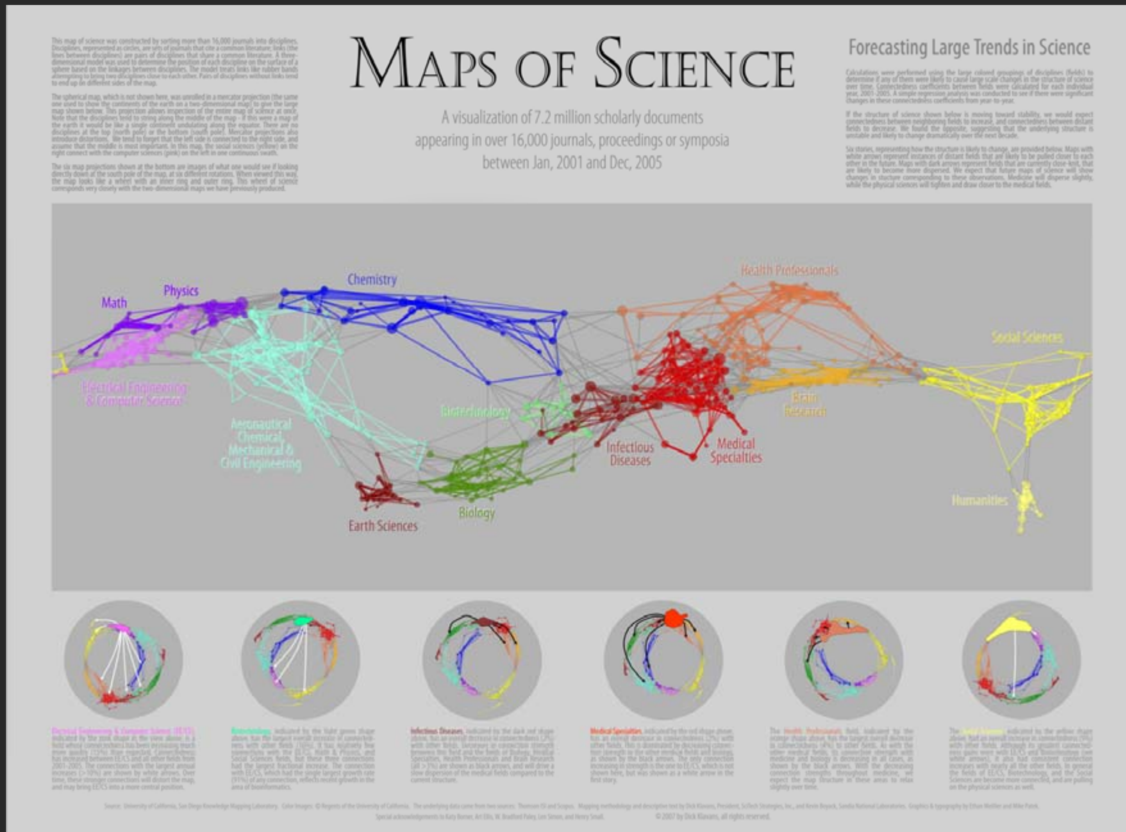
Atlas of Science - Katy Borner - 2010



Science as accumulation of knowledge.
"Scholarly brick laying".
Standing on the shoulders of giants.

Densely knit communities.
The importance of weak links.

Hypothetical Model of the Evolution of Science - Daniel Zeller - 2007

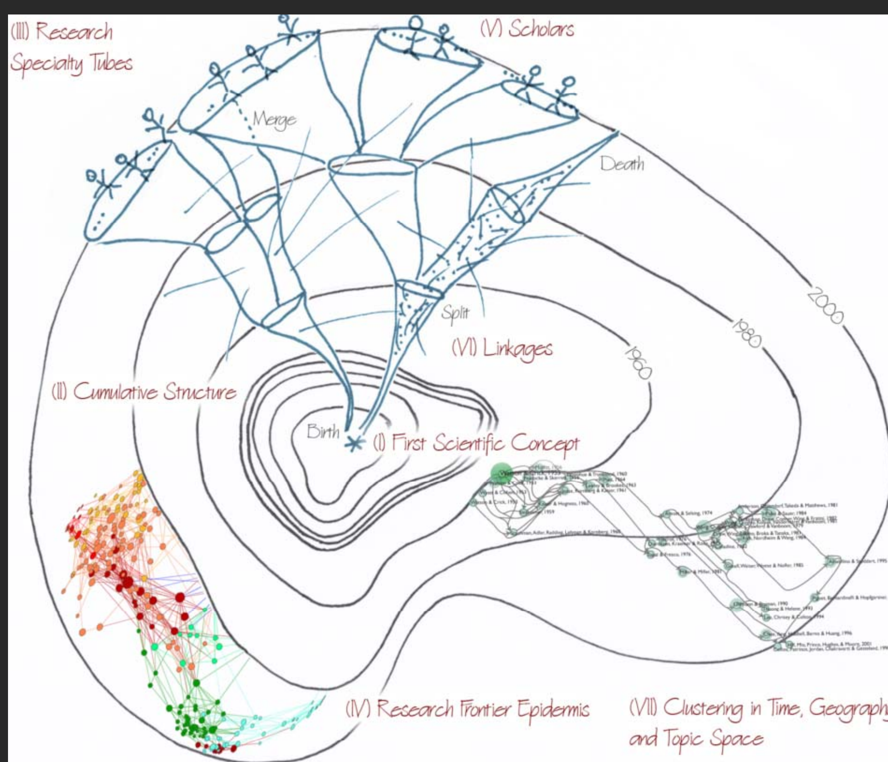


Maps of Science: Forecasting Large Trends in Science - Richard Klavans, Kevin Boyack - 2007

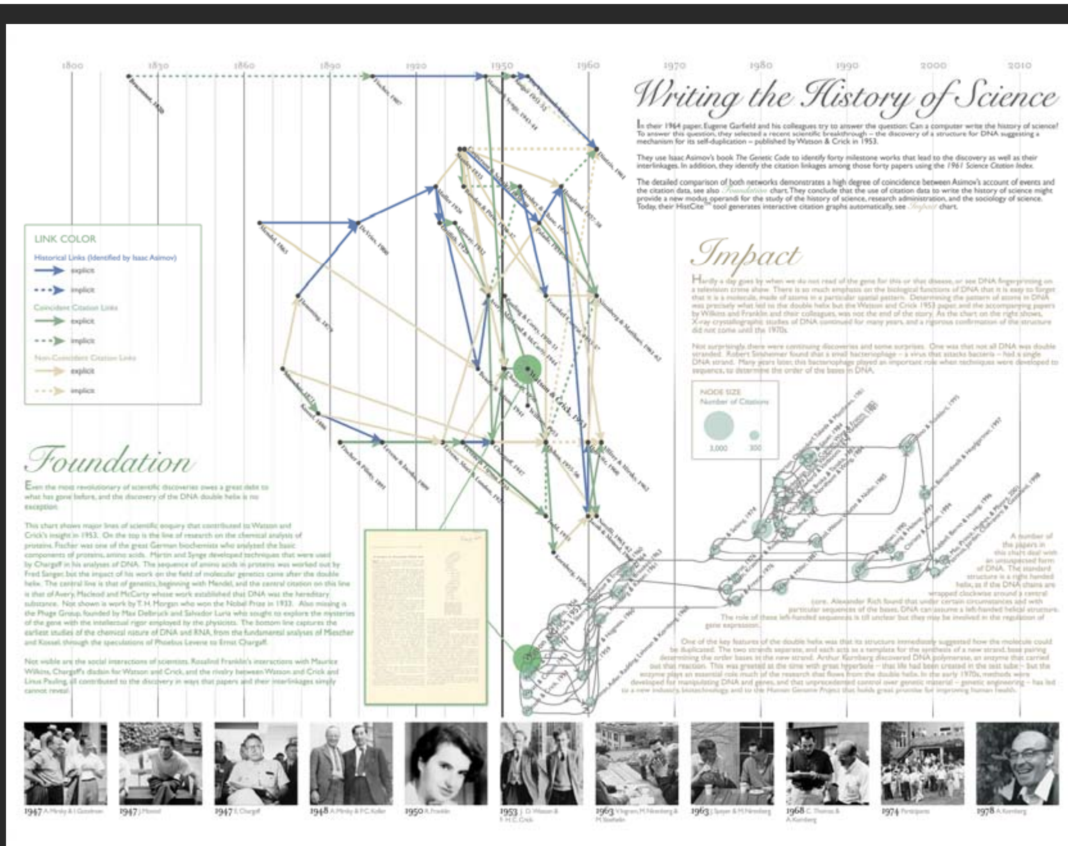
In Terms of Geography - Andre Skupin - 2005

Areas of science are tube shaped.

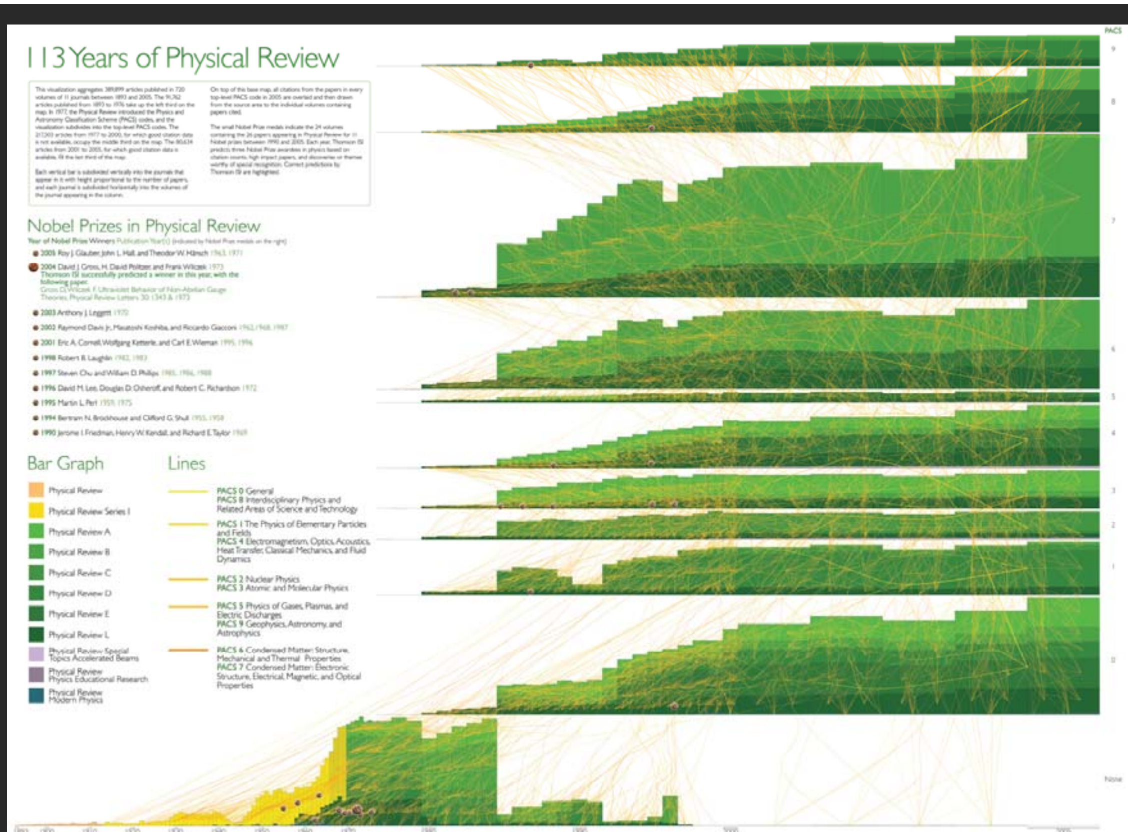
Hypothetical Model of the Evolution of Science - Daniel Zeller - 2007



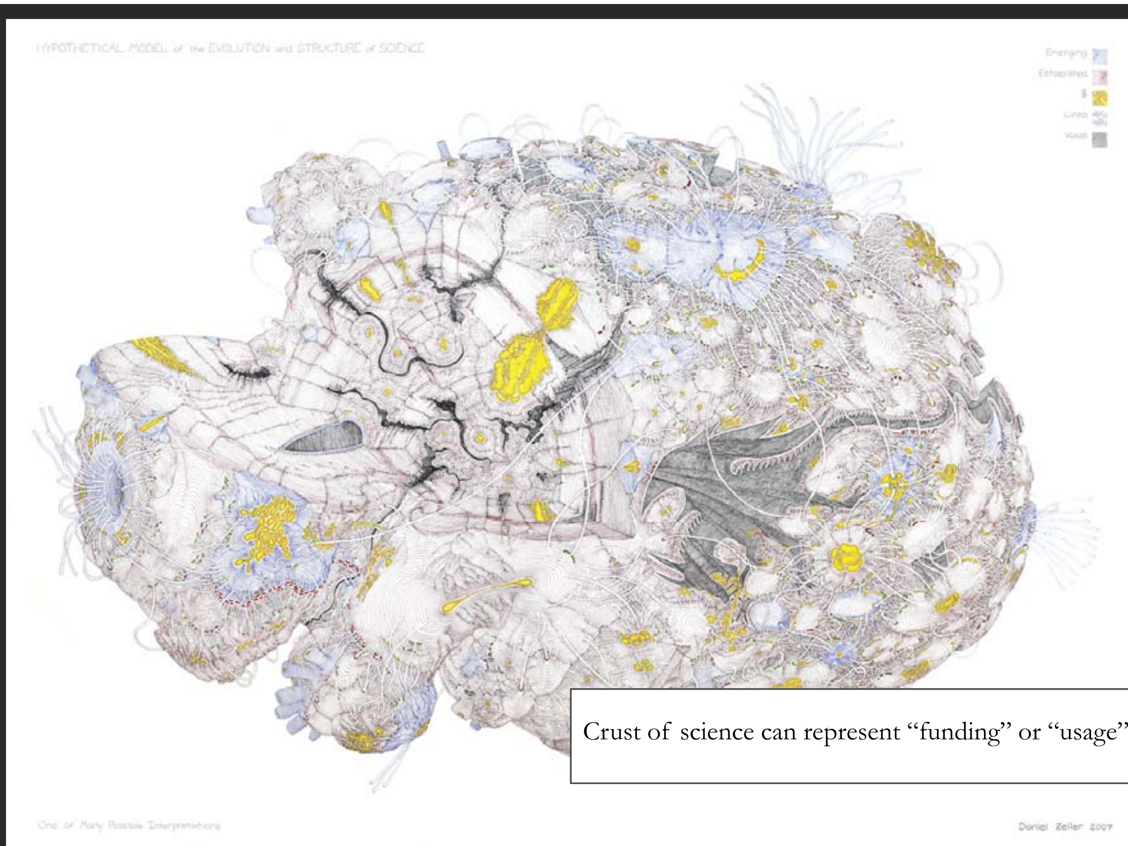
Atlas of Science - Katy Borner - 2010



HistCite[TM] Visualization of DNA Development - Eugene Garfield, Elisha Hardy, Katy Borner, Ludmila Pollock, Jan Witkowski- 2006

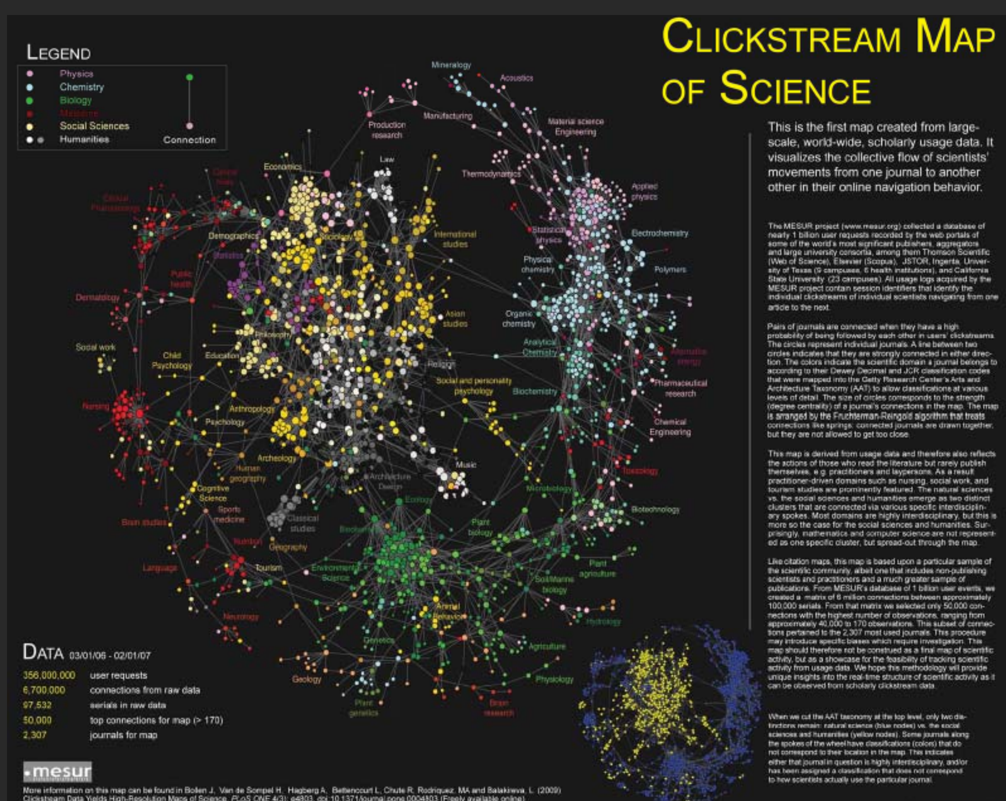


114 Years of Physical Review - Bruce W. Herr II, Russell Duhon, Katy Borner, Elisha Hardy, Shashikant Penumarthy - 2007

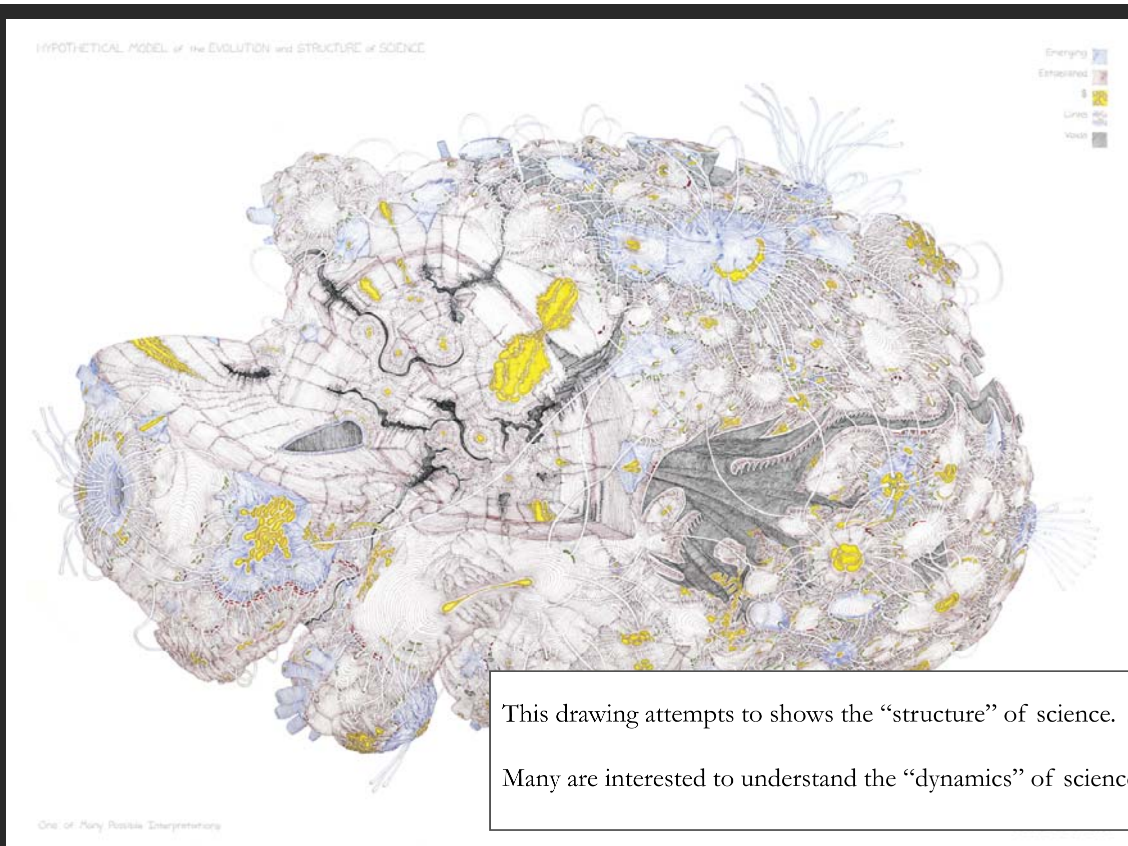


Crust of science can represent "funding" or "usage".

Hypothetical Model of the Evolution of Science - Daniel Zeller - 2007



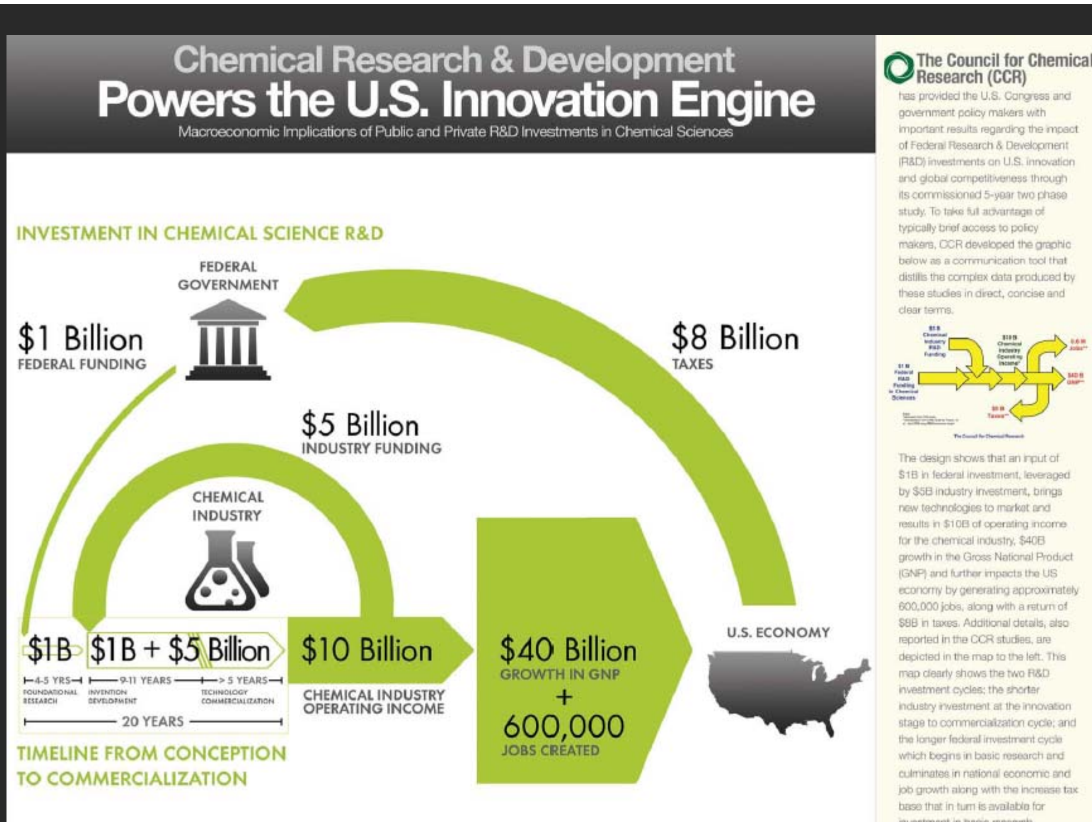
A Clickstream Map of Science – Bollen, Johan, Herbert Van de Sompel, Aric Hagberg, Luis M.A. Bettencourt, Ryan Chute, Marko A. Rodriquez, Lyudmila Balakireva - 2008



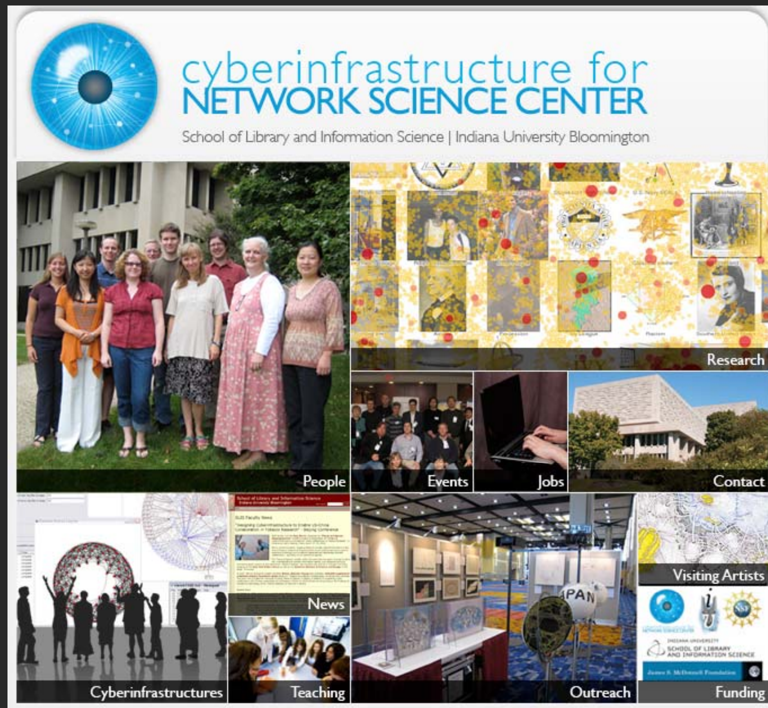
This drawing attempts to shows the “structure” of science.

Many are interested to understand the “dynamics” of science.

Hypothetical Model of the Evolution of Science - Daniel Zeller - 2007



Council for Chemical Research - Chemical R&D Powers the U.S. Innovation Engine. Washington, DC. Courtesy of the Council for Chemical Research - 2009



All papers, maps, cyberinfrastructures, talks, press are linked from http://cns.slis.indiana.edu

The End.