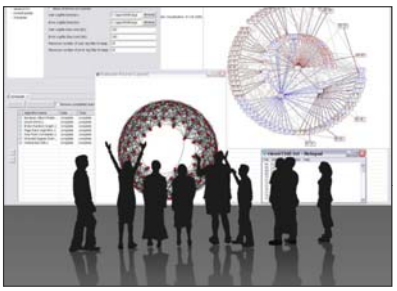
The Science of Science (Sci^2) Tool and Its Utility for Research

Katy Börner and the Cyberinfrastructure for Network Science Center
School of Library and Information Science
Indiana University, Bloomington, IN
katy@indiana.edu

With special thanks to Kevin W. Boyack, Micah Linnemeier, Russell J. Duhon, Patrick Phillips, Joseph Biberstine, Chintan Tank Nianli Ma, Angela M. Zoss, Hanning Guo, Mark A. Price, Scott Weingart

Networks and Complex Systems Talk Series, IUB
December 7th, 2009

Overview



1. Plug-and-Play Macroscopes
2. Sci2 Tool Introduction and Demos
3. Sample Science of Science Studies

The Changing Scientific Landscape

Star Scientist -> Research Teams: In former times, science was driven by key scientists. Today, science is driven by effectively collaborating co-author teams often comprising expertise from multiple disciplines and several geospatial locations (Börner, Dall'Asta, Ke, & Vespignani, 2005; Shneiderman, 2008).

Users -> Contributors: Web 2.0 technologies empower anybody to contribute to Wikipedia and to exchange images and videos via Fickr and YouTube. WikiSpecies, WikiProfessionals, or WikiProteins combine wiki and semantic technology in support of real time community annotation of scientific datasets (Mons et al., 2008).

Cross-disciplinary: The best tools frequently borrow and synergistically combine methods and techniques from different disciplines of science and empower interdisciplinary and/or international teams of researchers, practitioners, or educators to fine-tune and interpret results collectively.

One Specimen -> Data Streams: Microscopes and telescopes were originally used to study one specimen at a time. Today, many researchers must make sense of massive streams of multiple types of data with different formats, dynamics, and origin.

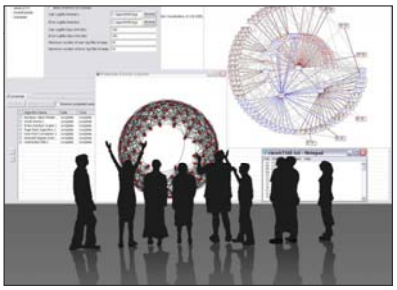
Static Instrument -> Evolving Cyberinfrastructure (CI): The importance of hardware instruments that are rather static and expensive decreases relative to software infrastructures that are highly flexible and continuously evolving according to the needs of different sciences. Some of the most successful services and tools are decentralized increasing scalability and fault tolerance.

Modularity: The design of software modules with well defined functionality that can be flexibly combined helps reduce costs, makes it possible to have many contribute, and increases flexibility in tool development, augmentation, and customization.

Standardization: Adoption of standards speeds up development as existing code can be leveraged. It helps pool resources, supports interoperability, but also eases the migration from research code to production code and hence the transfer of research results into industry applications and products.

Open data and open code: Lets anybody check, improve, or repurpose code and eases the replication of scientific studies.

Microscopes, Telescopes, and Macrocopes



Just as the **microscope** empowered our naked eyes to see cells, microbes, and viruses thereby advancing the progress of biology and medicine or the **telescope** opened our minds to the immensity of the cosmos and has prepared mankind for the conquest of space, **macroscopes** promise to help us cope with another infinite: the infinitely complex. Macroscopes give us a 'vision of the whole' and help us 'synthesize'. They let us detect patterns, trends, outliers, and access details in the landscape of science. Instead of making things larger or smaller, macroscopes let us observe what is at once too great, too slow, or too complex for our eyes.

Desirable Features of Plug-and-Play Macroscopes

Division of Labor: Ideally, labor is divided in a way that the expertise and skills of computer scientists are utilized for the design of standardized, modular, easy to maintain and extend "core architecture". Dataset and algorithm plugins, i.e., the "filling", are initially provided by those that care and know most about the data and developed the algorithms: the domain experts.

Ease of Use: As most plugin contributions and usage will come from non-computer scientists it must be possible to contribute, share, and use new plugins without writing one line of code. Wizard-driven integration of new algorithms and data sets by domain experts, sharing via email or online sites, deploying plugins by adding them to the 'plugin' directory, and running them via a Menu driven user interfaces (as used in Word processing systems or Web browsers) seems to work well.

Plugin Content and Interfaces: Should a plugin represent one algorithm or an entire tool? What about data converters needed to make the output of one algorithm compatible with the input of the next? Should those be part of the algorithm plugin or should they be packaged separately?

Supported (Central) Data Models: Some tools use a central data model to which all algorithms conform, e.g., Cytoscape, see Related Work section. Other tools support many internal data models and provide an extensive set of data converters, e.g., Network Workbench, see below. The former often speeds up execution and visual rendering while the latter eases the integration of new algorithms. In addition, most tools support an extensive set of input and output formats.

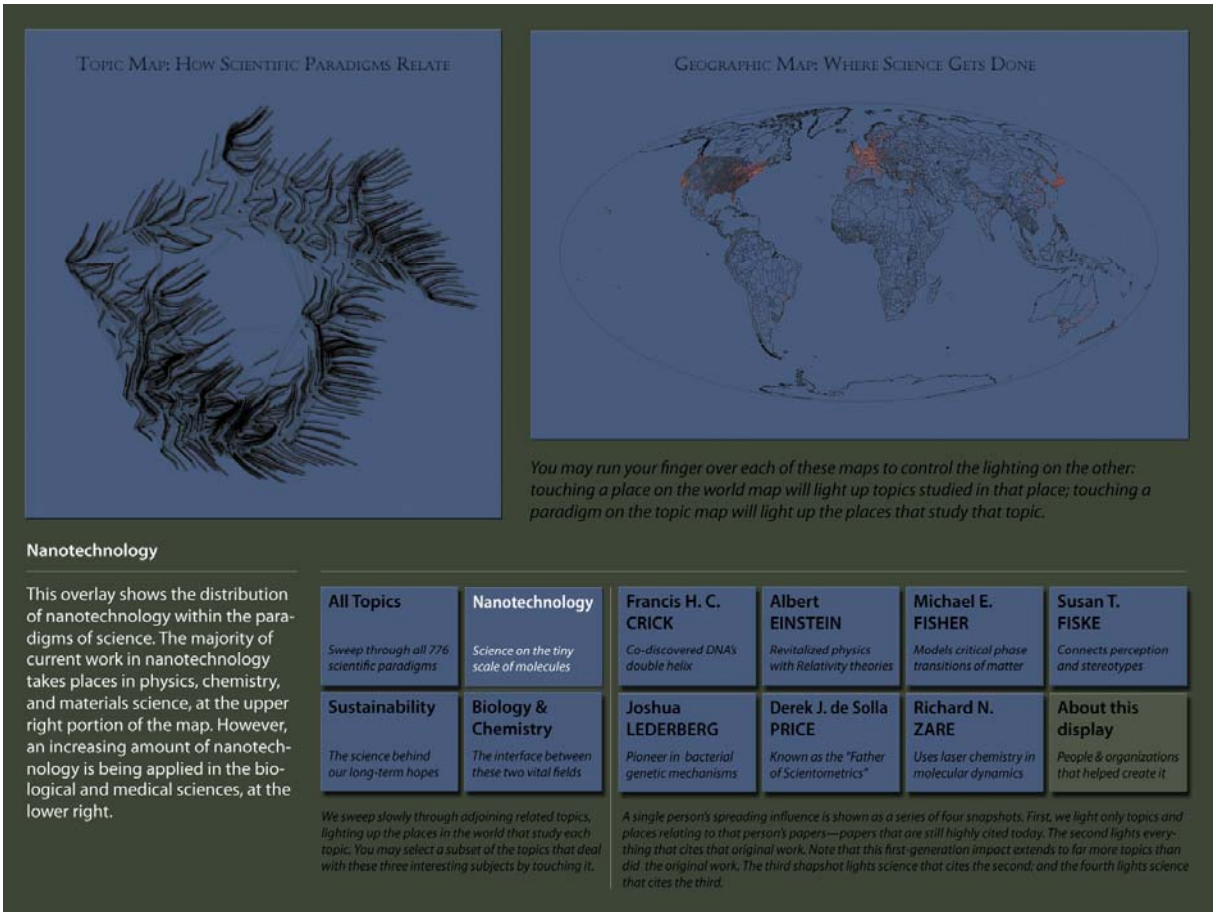
Core vs. Plugins: As will be shown, the "core architecture" and the "plugin filling" can be implemented as sets of plugin bundles. Answers to questions such as: "Should the graphical user interface (GUI), interface menu, scheduler, or data manager be part of the core or its filling?" will depend on the type of tools and services to be delivered.

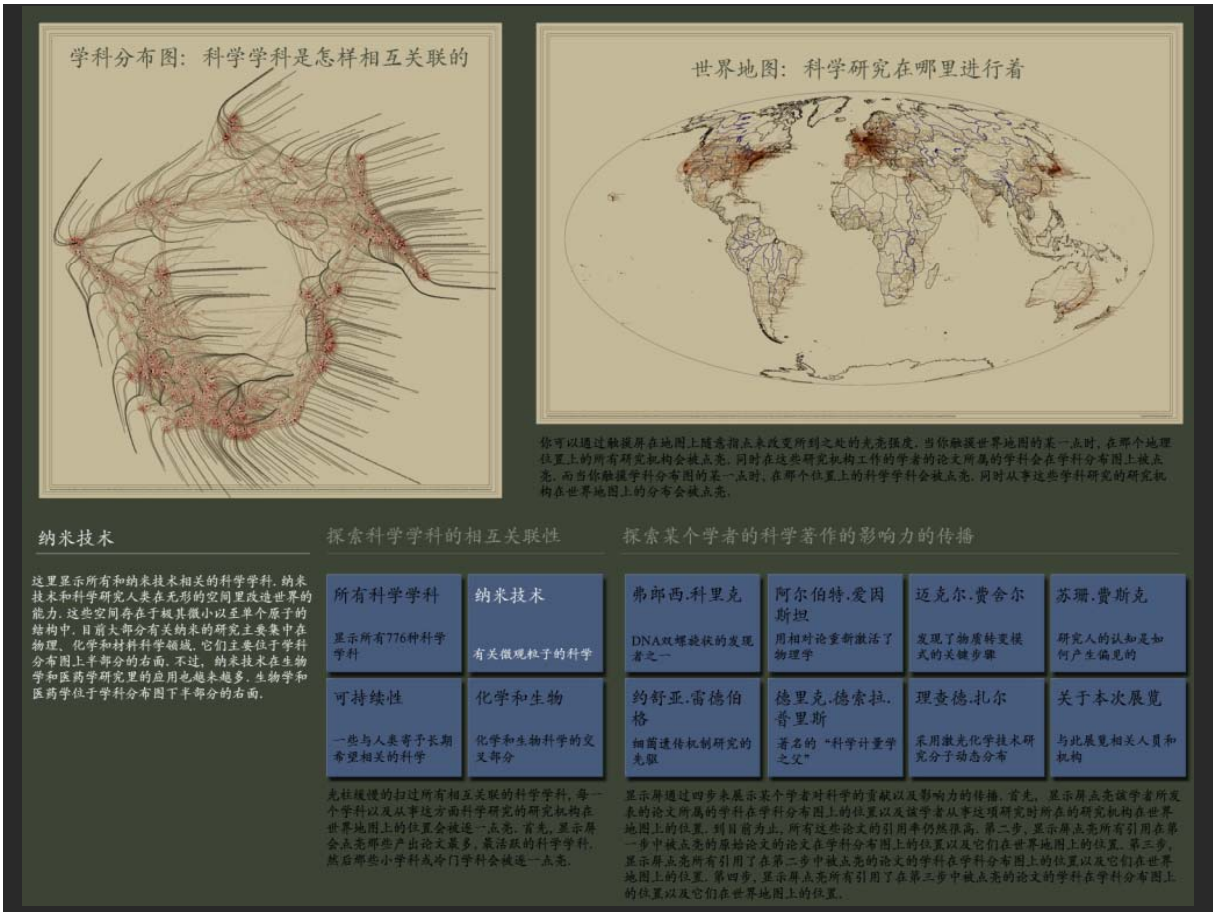
Supported Platforms: If the software is to be used via Web interfaces then Web services need to be implemented. If a majority of domain experts prefers a stand-alone tool running on a specific operating system then a different deployment is necessary.

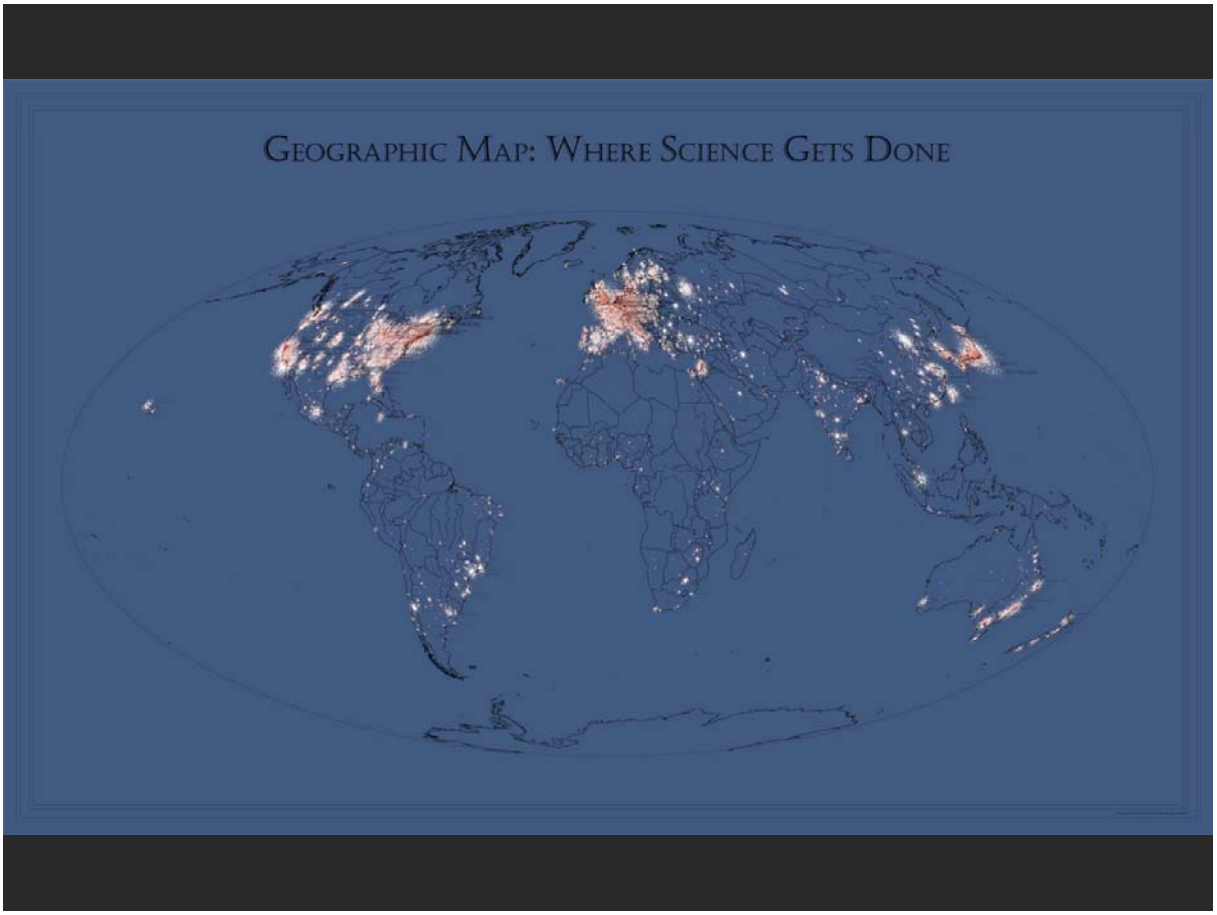
Hubble telescope has cost a total of about $10 billion over its 20 years of operation. Ed Weiler, HST space science chief, computed that this "...equates to about two cents per week per American taxpayer over 20 years" – a true bargain.

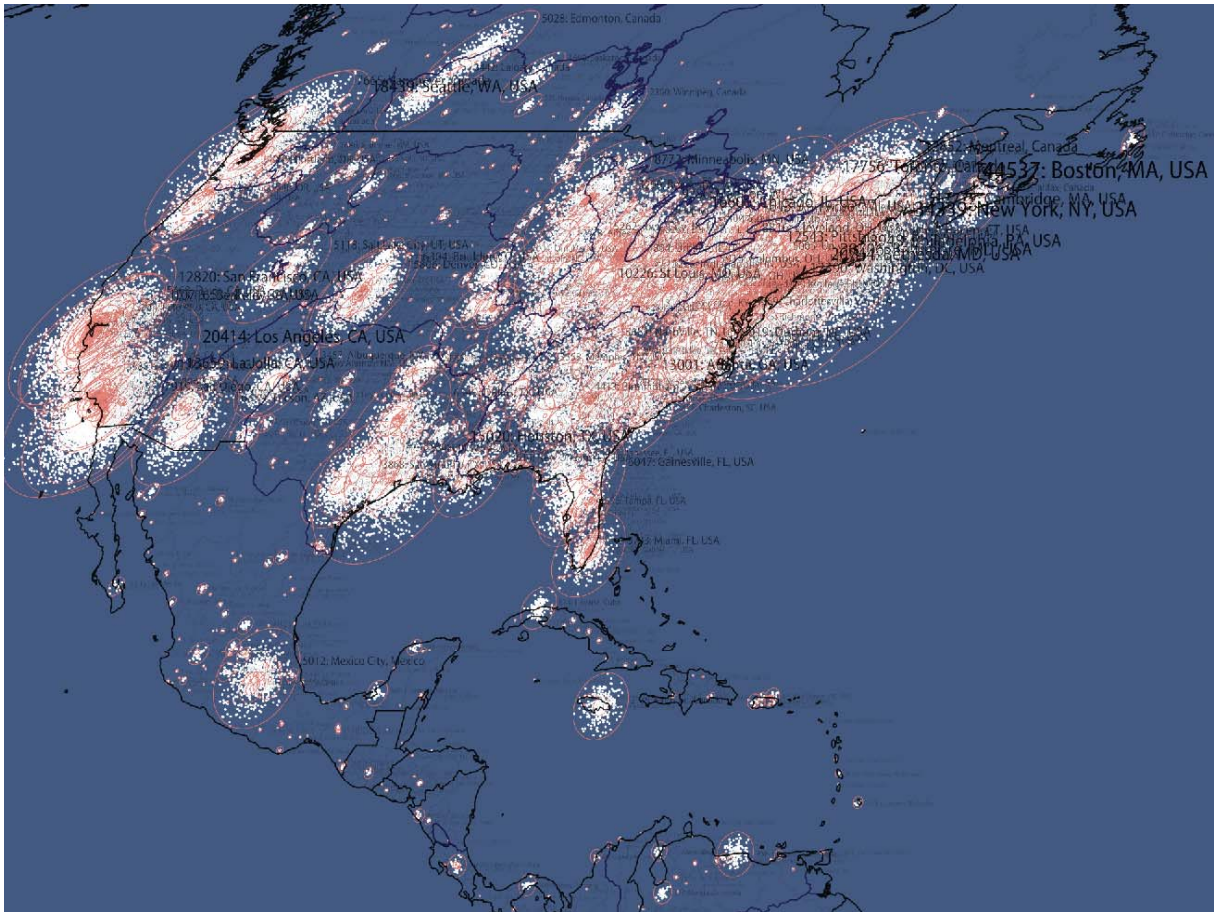
Illuminated Diagram Display

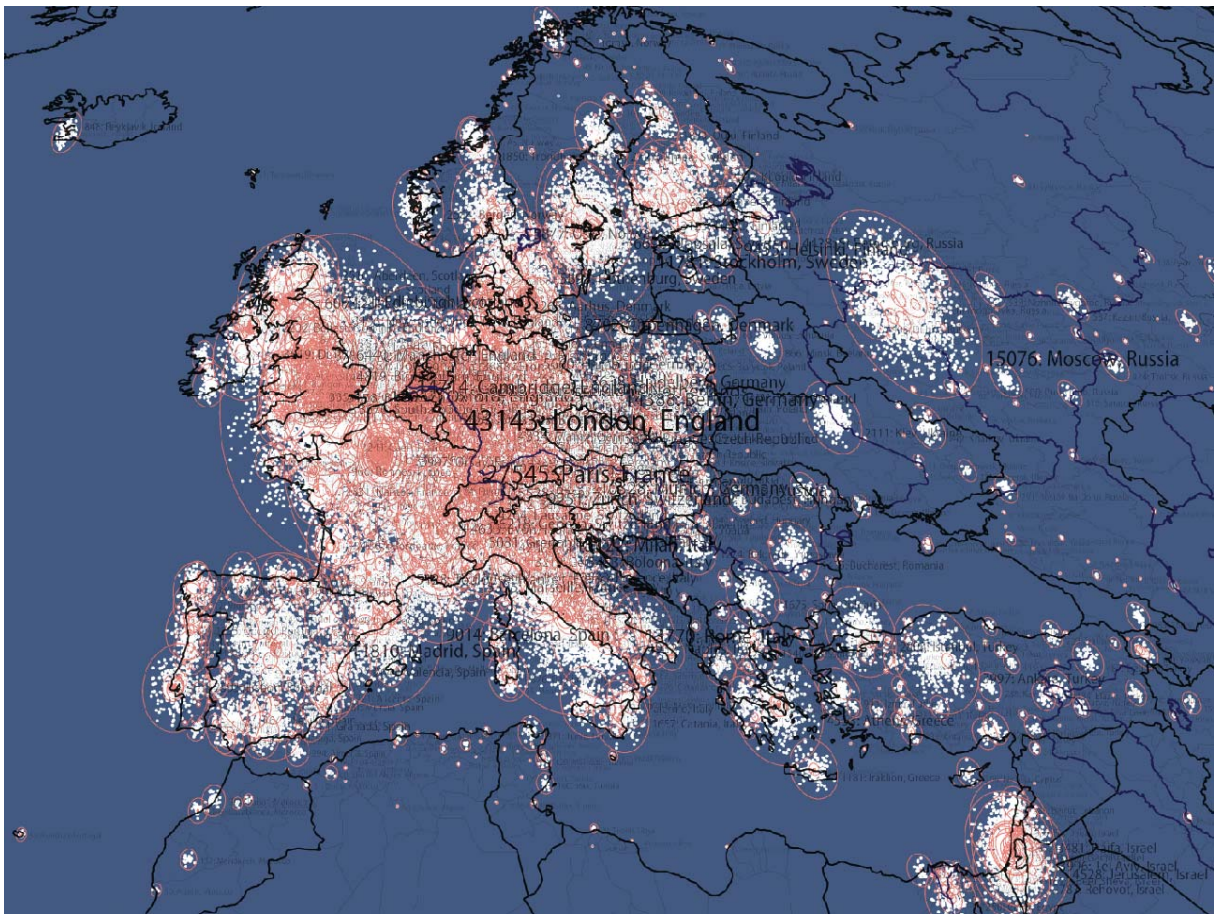
W. Bradford Paley, Kevin W. Boyack, Richard Kalvans, and Katy Börner (2007) Mapping, Illuminating, and Interacting with Science. SIGGRAPH 2007, San Diego, CA.

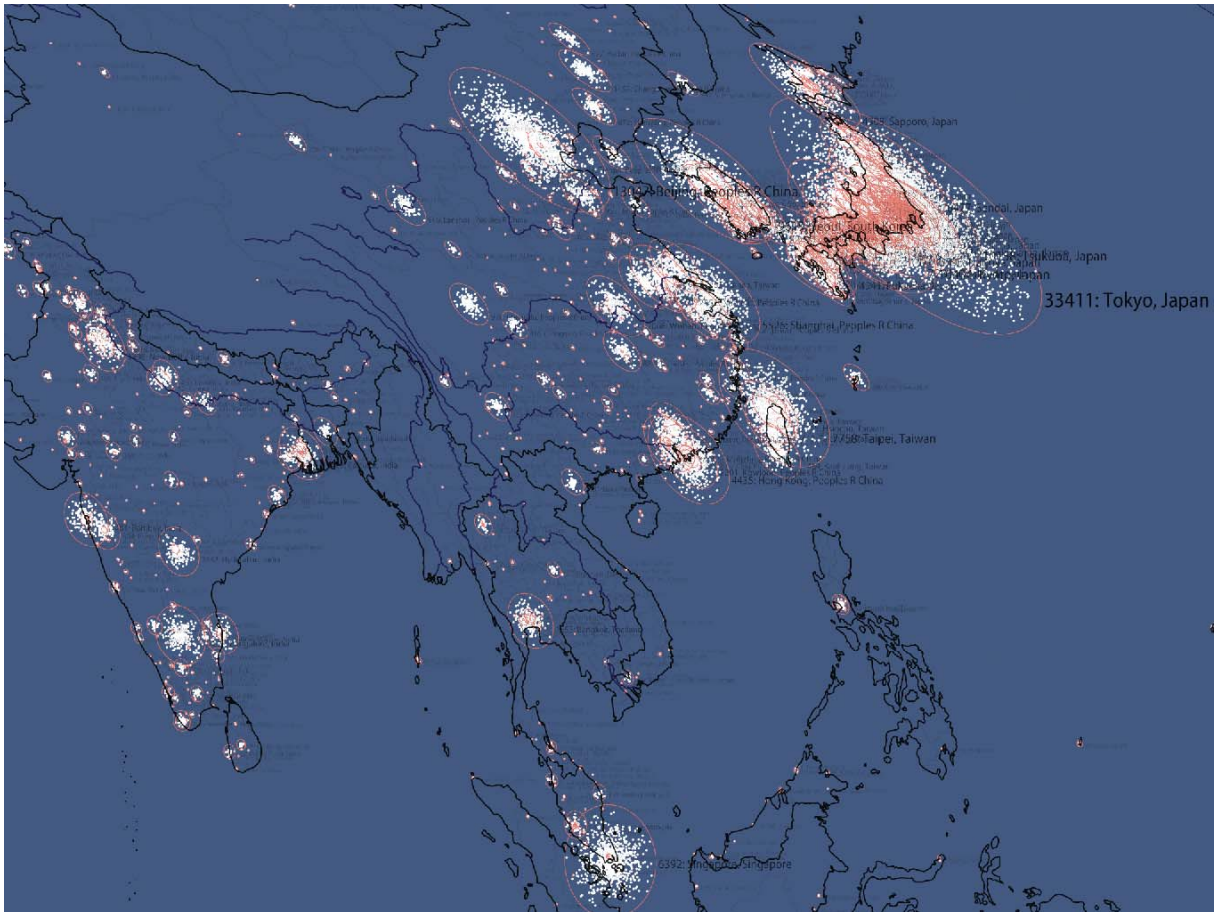



学科分布图：科学学科是怎样相互关联的
世界地图：科学研究在哪里进行着
你可以通过触摸屏在地图上随意指点来改变所到之处的光亮强度。当你触摸世界地图的某一点时，在那个地理位置上的所有研究机构会被点亮。同时在这些研究机构工作的学者的论文所属的学科会在学科分布图上被点亮。而当你触摸学科分布图的某一点时，在那个位置上的科学学科会被点亮。同时从事这些学科研究的研究机构在世界地图上的分布会被点亮。
纳米技术
这里显示所有和纳米技术相关的科学学科。纳米技术和科学研究人类在无形的空间里改造世界的能力。这些空间存在于极其微小以至单个原子的结构中。目前大部分有关纳米的研究主要集中在物理、化学和材料科学领域，它们主要位于学科分布图上半部分的右面。不过，纳米技术在生物学和医药学研究里的应用也越来越多，生物学和医药学位于学科分布图下半部分的右面。
探索科学学科的相互关联性
所有科学学科
显示所有776种科学学科
纳米技术
有关微观粒子的科学
可持续性
一些与人类寄于长期希望相关的科学
化学和生物
化学和生物科学的交叉部分
光柱缓慢的扫过所有相互关联的科学学科，每一个学科以及从事这方面科学研究的研究机构在世界地图上的位置会被逐一点亮。首先，显示屏会点亮那些产出论文最多，最活跃的科学学科，然后那些小学科或冷门学科会被逐一点亮。
探索某个学者的科学著作的影响力的传播
弗郎西.科里克
DNA双螺旋状的发现者之一
阿尔伯特.爱因斯坦
用相对论重新激活了物理学
迈克尔.费舍尔
发现了物质转变模式的关键步骤
苏珊.费斯克
研究人的认知是如何产生偏见的
约舒亚.雷德伯格
细菌遗传机制研究的先驱
德里克.德索拉.普里斯
著名的“科学计量学之父”
理查德.扎尔
采用激光化学技术研究分子动态分布
关于本次展览
与此展览相关人员和机构
显示屏通过四步来展示某个学者对科学的贡献以及影响力的传播。首先，显示屏点亮该学者所发表的论文所属的学科在学科分布图上的位置以及该学者从事这项研究时所在的研究机构在世界地图上的位置。到目前为止，所有这些论文的引用率仍然很高。第二步，显示屏点亮所有引用在第一步中被点亮的原始论文的论文在学科分布图上的位置以及它们在世界地图上的位置。第三步，显示屏点亮所有引用了在第二步中被点亮的论文的学科在学科分布图上的位置以及它们在世界地图上的位置。第四步，显示屏点亮所有引用了在第三步中被点亮的论文的学科在学科分布图上的位置以及它们在世界地图上的位置。

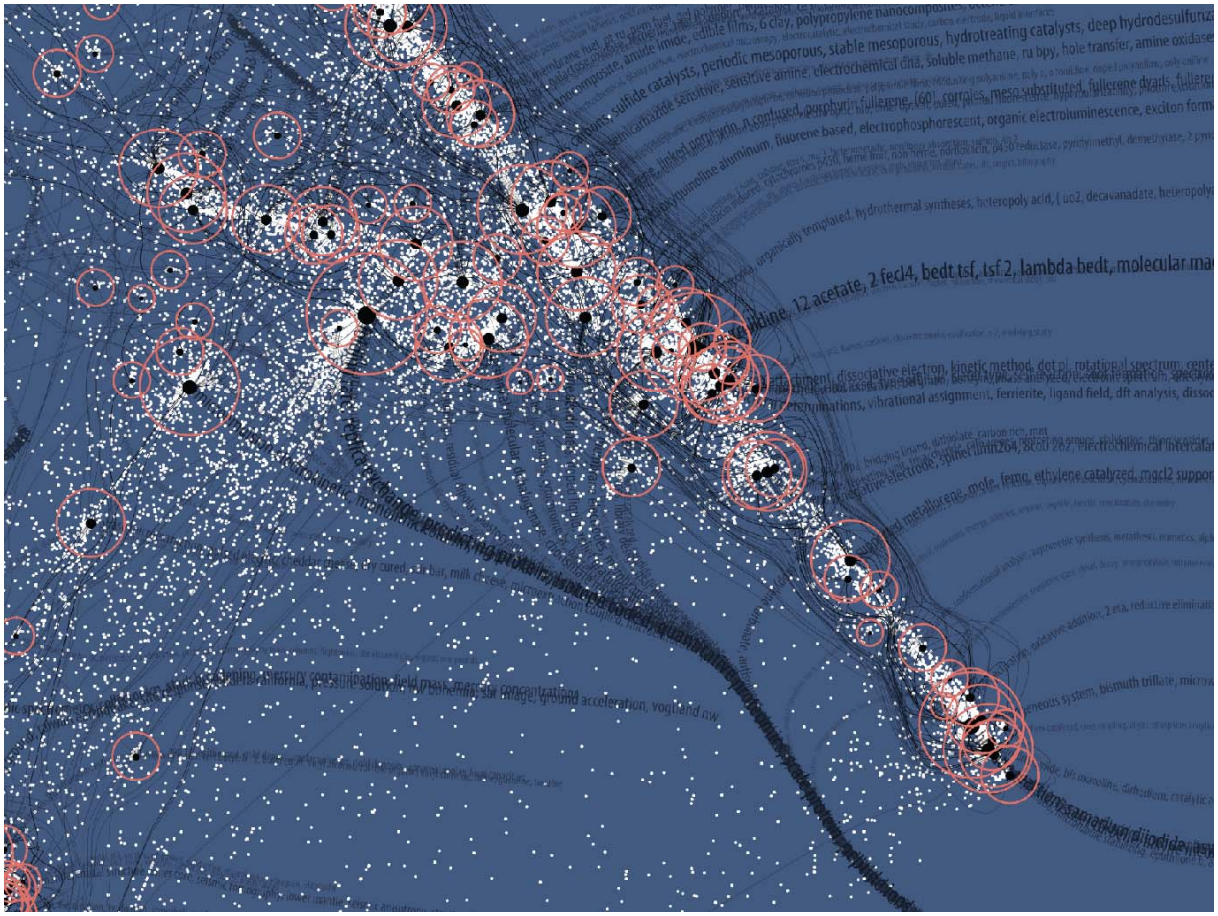

Geographic Map: Where Science Gets Done


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

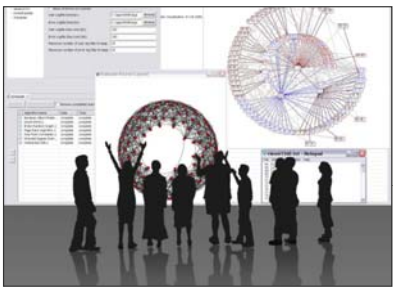

15076: Moscow, Russia
4313: London, England
27545: Paris, France
Rome, Italy
Madrid, Spain


33411: Tokyo, Japan

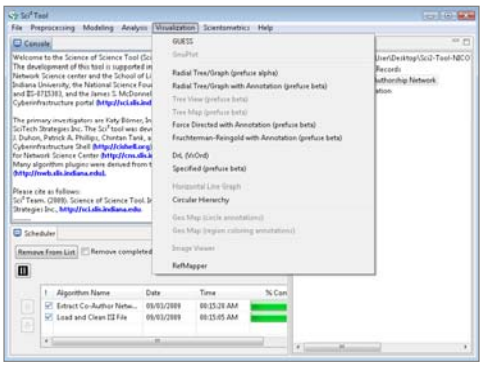

Topic Map: How Scientific Paradigms Relate



Overview



1. Plug-and-Play Macroscopes
2. Sci2 Tool Introduction and Demos
3. Sample Science of Science Studies



Open Data and Open Code
Studying Individual, Local, and Global Flows and Activity Patterns

Design **comprehensive databases** that capture relevant data and **cyberinfrastructures** that can be used to make sense of this data(stream).

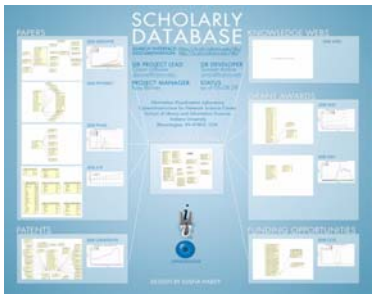
Science studies can be conducted at different levels:

- *local* (individual),
- *meso* (local, e.g., one institute, one funding agency), or
- *global* level (all of science or world wide).

Using

- Statistical Analysis/Profiling
- Temporal Analysis (When)
- Geospatial Analysis (Where)
- Topical Analysis (What)
- Network Analysis (With Whom?)

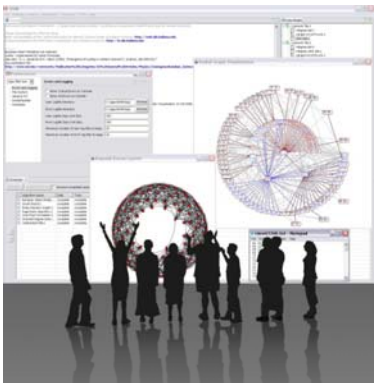
CI for a Science of Science Studies



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

James S. McDonnell Foundation

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



Network Workbench Tool + Community Wiki
http://nwb.slis.indiana.edu

Sci² Tool and Science of Science CI Portal
http://sci.slis.indiana.edu

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

Macroscope Design

Network Workbench Tool
A large-scale network analysis, modeling and visualization toolkit for biomedical, social science and physics research.

Science of Science Tool
for science of science studies at the individual, local, or global level for temporal, geospatial, semantic, or network analysis and vis.

Epidemics Research Tool
An open computational infrastructure for epidemics research.

http://nwb.slis.indiana.edu

http://sci.slis.indiana.edu

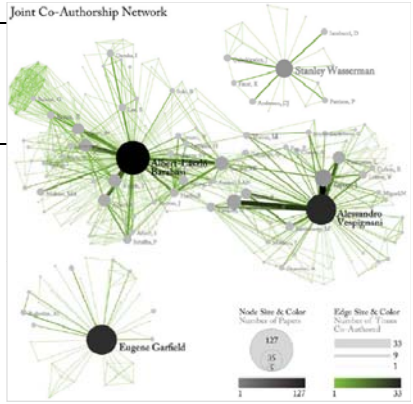
http://epic.slis.indiana.edu

CIShell.org

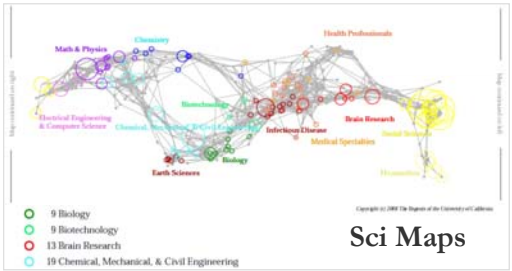
Sci² Tool

http://sci.slis.indiana.edu

"Open Code for S&T Assessment"
Branded OSGi/CIShell based tool with NWB plugins and many new plugins.



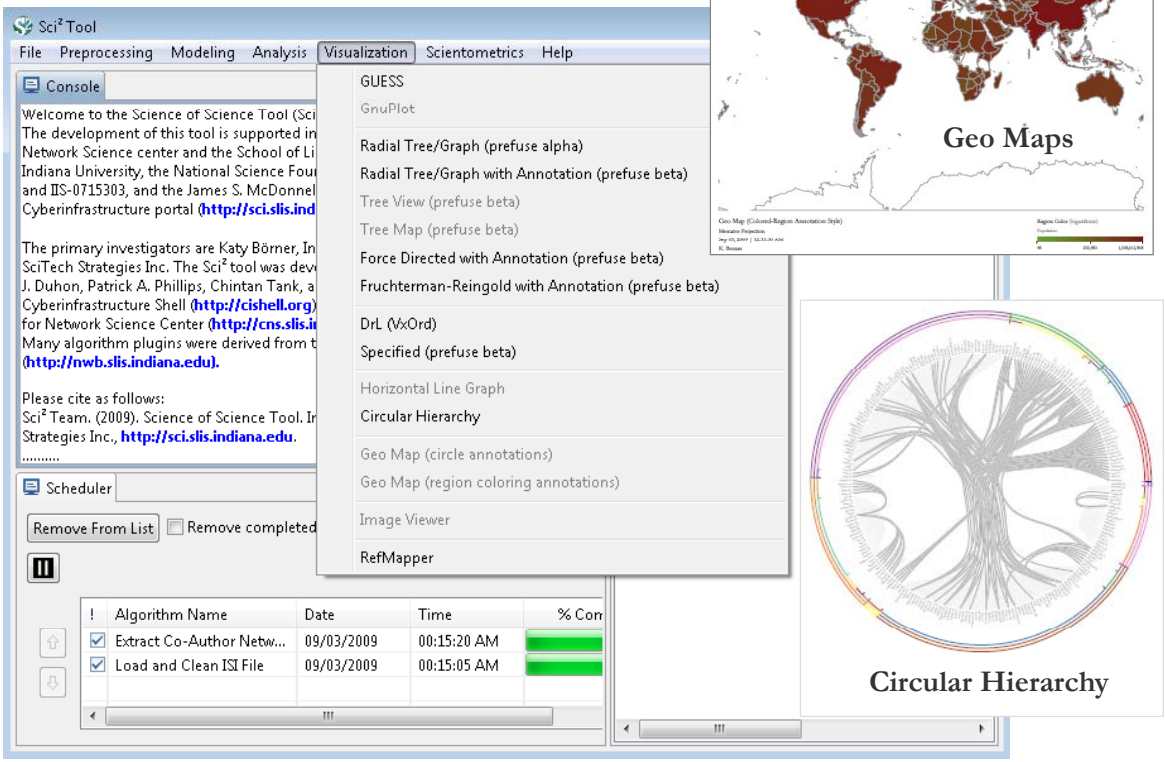
GUESS Network Vis



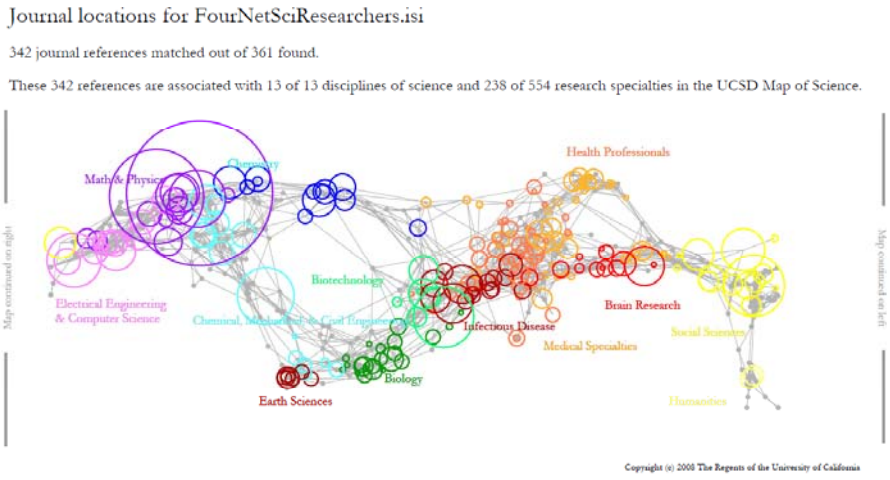
Sci Maps

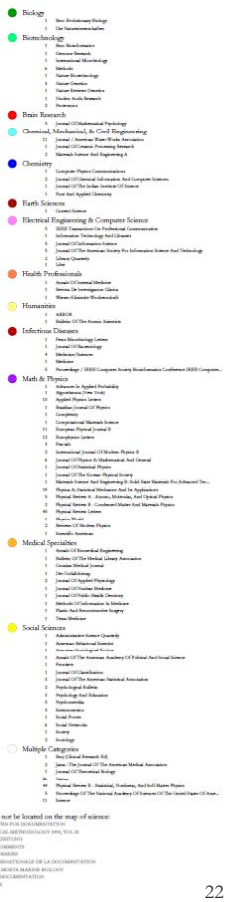
Börner, Katy, Huang, Weixia (Bonnie), Linnemeier, Micah, Duhon, Russell Jackson, Phillips, Patrick, Ma, Nianli, Zoss, Angela, Guo, Hanning & Price, Mark. (2009). Rete-Netzwerk-Red: Analyzing and Visualizing Scholarly Networks Using the Scholarly Database and the Network Workbench Tool. Proceedings of ISSI 2009: 12th International Conference on Scientometrics and Informetrics, Rio de Janeiro, Brazil, July 14-17 . Vol. 2, pp. 619-630.

Sci² Tool



Topic Mapping: UCSD Science Map





Sci² 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)

Sci² Tool: Algorithms

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

Preprocessing
Extract Top N% Records
Extract Top N Records
Normalize Text
Slice Table by Line

Extract Top Nodes
Extract Nodes Above or Below Value
Delete Isolates

Extract top Edges
Extract Edges Above or Below Value
Remove Self Loops
Trim by Degree
MST-Pathfinder Network Scaling
Fast Pathfinder Network Scaling

Snowball Sampling (in nodes)
Node Sampling
Edge Sampling

Symmetrize
Dichotomize
Multipartite Joining

Geocoder

Extract ZIP Code

Modeling
Random Graph
Watts-Strogatz
Small World
Barabási-Albert Scale-Free
TARL

Analysis
Network Analysis Toolkit (NAT)
Unweighted & Undirected
- Node Degree
- Degree Distribution
- ---
- K-Nearest Neighbor (Java)
- Watts-Strogatz Clustering Coefficient
- Watts Strogatz Clustering Coefficient over K
- ---
- Diameter
- Average Shortest Path
- Shortest Path Distribution
- Node Betweenness Centrality
- ---
- Weak Component Clustering
- Global Connected Components
- ---
- Extract K-Core
- Annotate K-Coreness
- ---
- HITS

Weighted & Undirected
- Clustering Coefficient
- Nearest Neighbor Degree
- Strength vs Degree
- Degree & Strength
- Average Weight vs End-point Degree
- Strength Distribution
- Weight Distribution
- Randomize Weights
- ---
- Blondel Community Detection
- ---
- HITS

Unweighted & Directed
- Node Indegree
- Node Outdegree
- Indegree Distribution
- Outdegree Distribution
- ---
- K-Nearest Neighbor
- Single Node in-Out Degree Correlations
- ---
- Dyad Reciprocity
- Arc Reciprocity
- Adjacency Transitivity
- ---
- Weak Component Clustering
- Strong Component Clustering
- ---

Sci² Tool: Algorithms cont.

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

Extract K-Core
Annotate K-Coreness

HITS
PageRank

Weighted & Directed
HITS
Weighted PageRank

Textual

Burst Detection

Visualization

GnuPlot
GUESS
Image Viewer

Radial Tree/Graph (prefuse alpha)
Radial Tree/Graph with Annotation (prefuse beta)
Tree View (prefuse beta)
Tree Map (prefuse beta)
Force Directed with Annotation (prefuse beta)
Fruchterman-Reingold with Annotation (prefuse beta)

DrL (VxOrd)
Specified (prefuse beta)

Horizontal Line Graph
Circular Hierarchy
Geo Map (Circle Annotation Style)
Geo Map (Colored-Region Annotation Style)

Scientometrics

Remove ISI Duplicate Records
Remove Rows with Multitudinous Fields
Detect Duplicate Nodes
Update Network by Merging Nodes

Extract Directed Network
Extract Paper Citation Network
Extract Author Paper Network

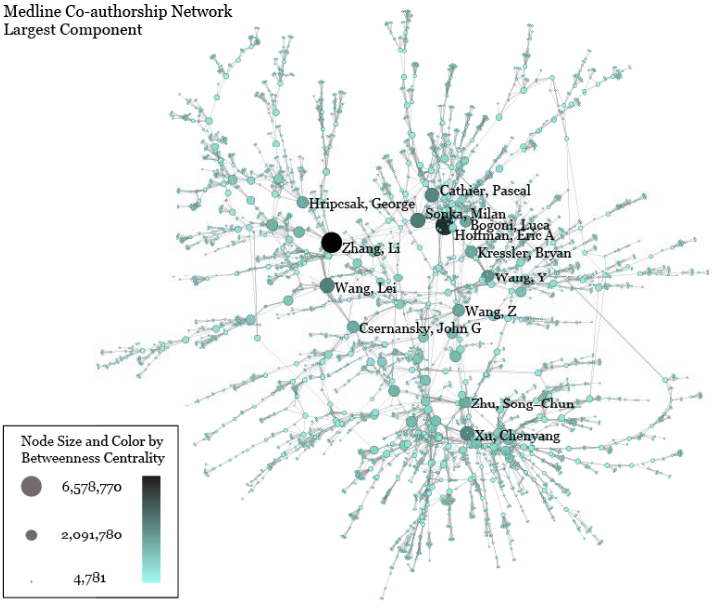
Extract Co-Occurrence Network
Extract Word Co-Occurrence Network
Extract Co-Author Network
Extract Reference Co-Occurrence (Bibliographic Coupling) Network
---
Extract Document Co-Citation Network

NWB=Sci² 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.

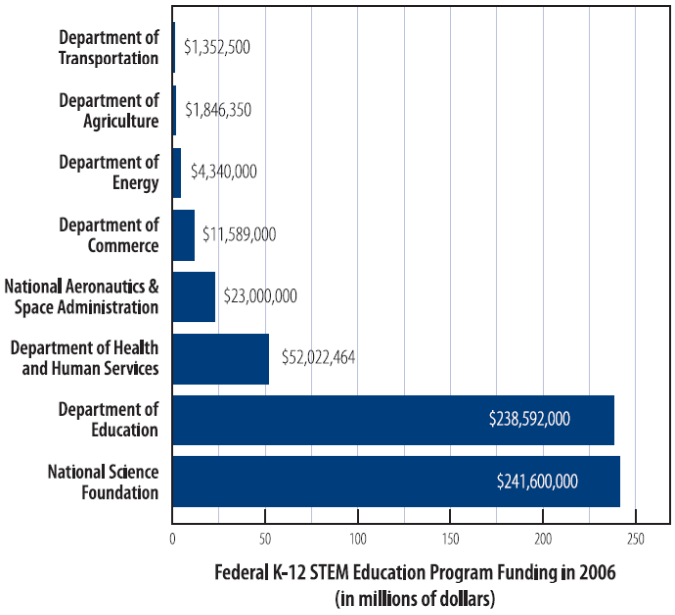


Sample Study – NSF Funding of STEM

Using NSF Awards Search:
http://www.nsf.gov/awardsearch
download relevant NSF awards that have "stem" and "education" in title, abstract, and awards.
Active awards only.

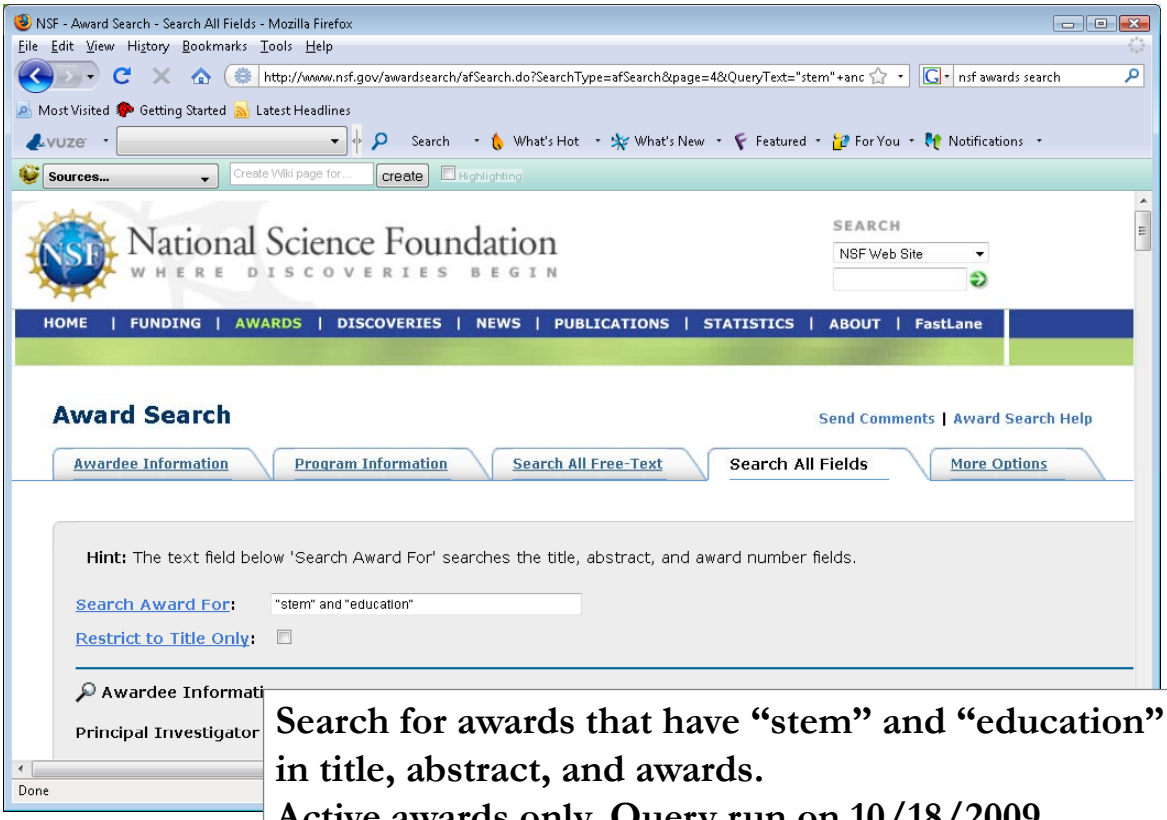
Number of awards: 1,340
Total awarded amount to date: $1,347,802,833

Retrieved on Oct 18, 2009

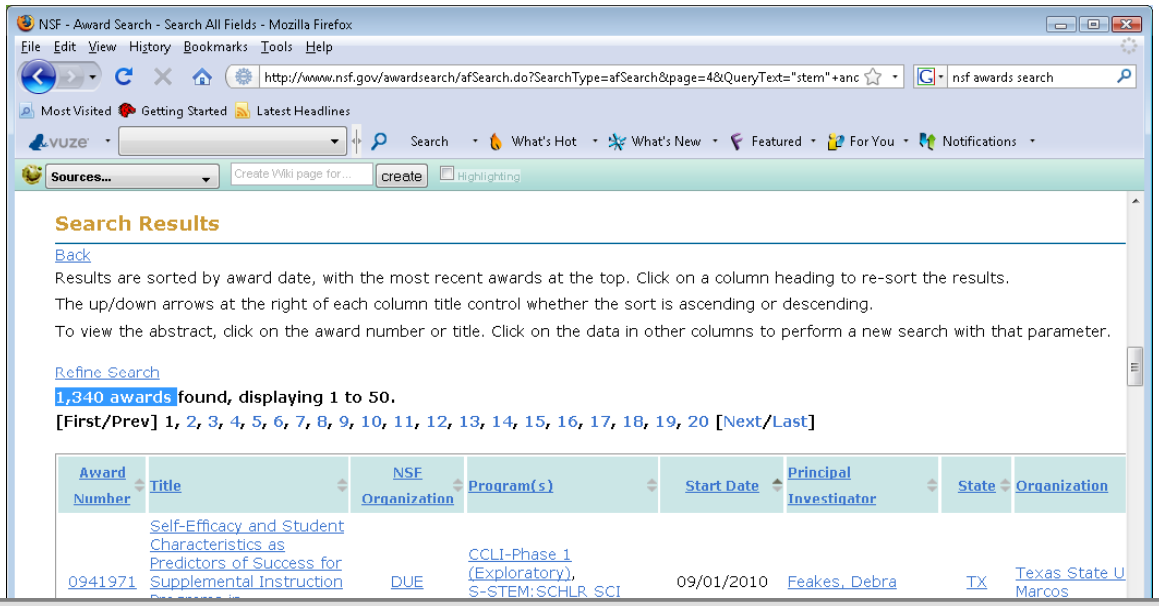


Federal K-12 STEM Education Program Funding in 2006

SOURCE: Department of Education, Report of the Academic Competitiveness Council, 2007



**Search for awards that have "stem" and "education" in title, abstract, and awards.
Active awards only. Query run on 10/18/2009.**



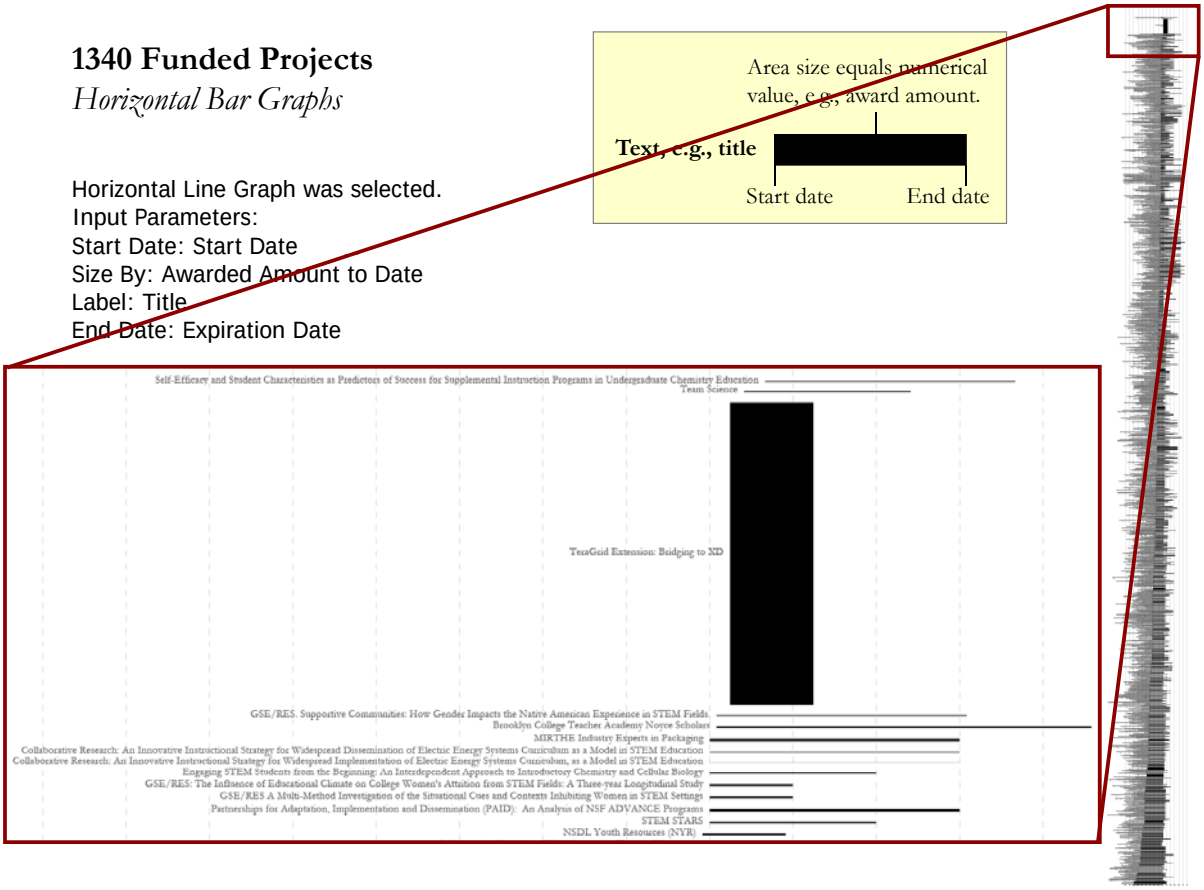
Top-10 Projects with highest Award Amount to Date

Title	NSF Org	Program(s)	PI	State	Organization	$ Awarded to Date
Next Generation Biometrics: Achieving Strength in Molecul	EPS	RESEARCH INFRASTRUCTURE	Paul Hill	WV	Higher Education Policy Commission	10,799,903
UMBC-BCPS STEM Project	DUE	Teaching & Mstr Tchng Fellows||	Anne Spence	MD	University of Maryland Baltimore County	11,410,713
MRSEC: MRSEC on Nanostructured Interfaces	DMR	MATERIALS RSCH SCI & ENG	Juan De Pablo	WI	University of Wisconsin-Madison	11,694,150
Spatial Intelligence and Learning Center (SILC)	SBE	|||SLC ACTIVITIES||SCIENCE O	Nora Newcombe	PA	Temple University	12,570,415
Project Pathways: Opening Routes to Math & Science Suc	DUE	Teaching & Mstr Tchng Fellows||	Marilyn Carlson	AZ	Arizona State University	13,392,614
NCLT: A Center to Develop Nanoscale Science and Engine	DRL	NANO CTR FOR LEARN & TEA	R. P. H. Chang	IL	Northwestern University	15,988,084
Enabling a Giant Segmented Mirror Telescope for the Unite	AST	MID-SCALE INSTRUMENTATIO	William Smith	DC	AURA/National Optical Astronomy Obse	18,000,000
Support of Synchrotron Radiation Center Operations	DMR	MPS DMR INSTRUMENTATION	Martin Cadwallader	WI	University of Wisconsin-Madison	19,526,500
TeraGrid Extension: Bridging to XD	OCI	ETF	Ian Foster	IL	University of Chicago	30,207,358
System-Wide Change for All Learners and Educators	DUE	Teaching & Mstr Tchng Fellows||	Terrence Millar	WI	University of Wisconsin-Madison	35,900,000

1340 Funded Projects

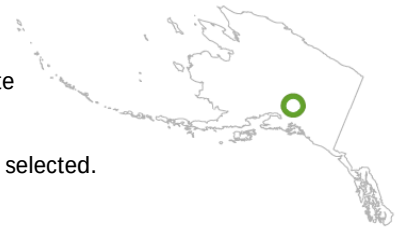
Horizontal Bar Graphs

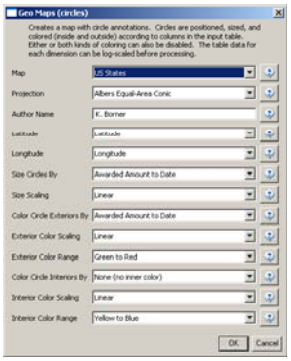
Horizontal Line Graph was selected.
Input Parameters:
Start Date: Start Date
Size By: Awarded Amount to Date
Label: Title
End Date: Expiration Date



1,340 Funded Projects

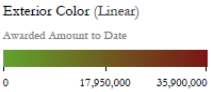
Geographic Maps

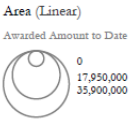




Geocoder was selected.
Input Parameters:
Place Name Column: Organization State
Place Type: STATE
.....
Geo Map (Circle Annotation Style) was selected.
Input Parameters:
Longitude: Longitude
Size Circles By: Awarded Amount to Date
Color Circle Exteriors By: Awarded Amount to Date
Color Circle Interiors By: None (no inner color)
Exterior Color Scaling: Linear
Exterior Color Range: Green to Red
Interior Color Range: Green to Red
Size Scaling: Linear
Projection: Albers Equal-Area Conic
Map: US States
Author Name: K. Borner
Interior Color Scaling: Linear
Latitude: Latitude

Geo Map (Circle Annotation Style)
Albers Equal-Area Conic Projection
Oct 18, 2009 | 11:34:57 AM
K. Borner

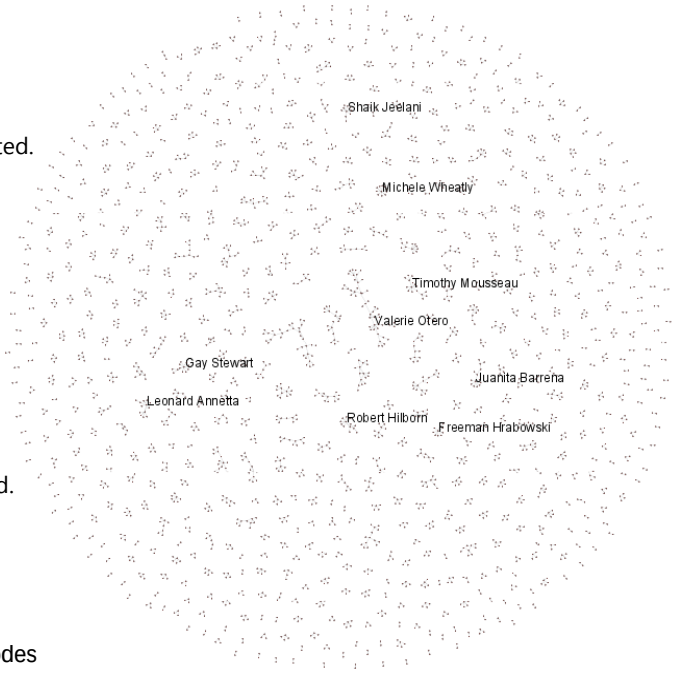




What Co-PI Networks Exist?

Extract Directed Network was selected.
Input Parameters:
Source Column: Principal Investigator
Text Delimiter: |
Target Column: Co-PI Name(s)
.....
Network Analysis Toolkit (NAT) was selected.
Nodes: 3225
Isolated nodes: 276
Edges: 2265
Average total degree: 1.4047
Average in degree: 0.7023
Average out degree: 0.7023
.....
Delete Isolates was selected.
.....
Node Degree was selected.
.....
Weak Component Clustering was selected.
Number of top clusters: 10
722 clusters found, generating graphs
for the top 10 clusters.

Giant component has 39 nodes
Next largest networks have 35, 17, 16 nodes



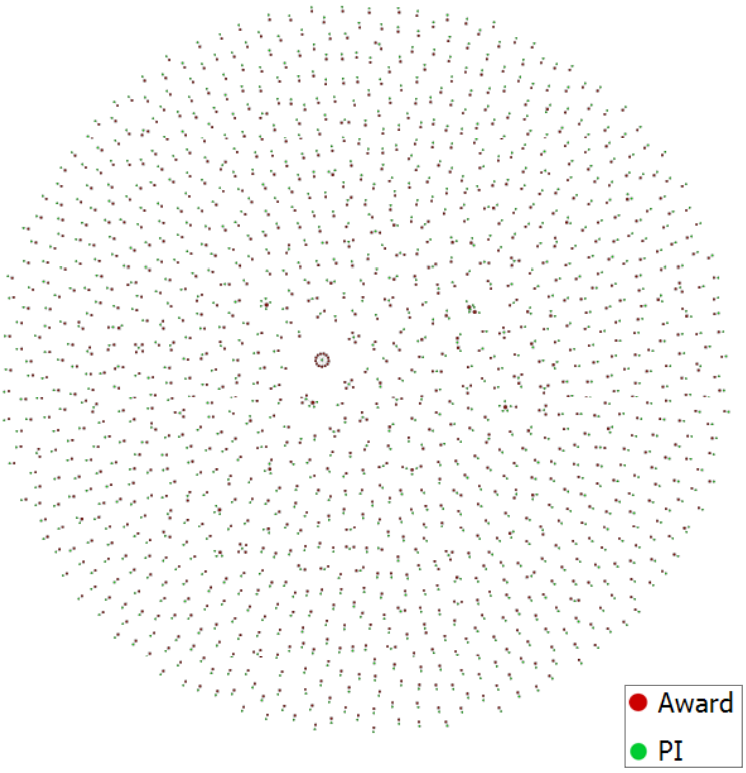
Co-PI Networks – Giant Component

Nodes = investigators
Size and color coded by number of collaborators (degree)

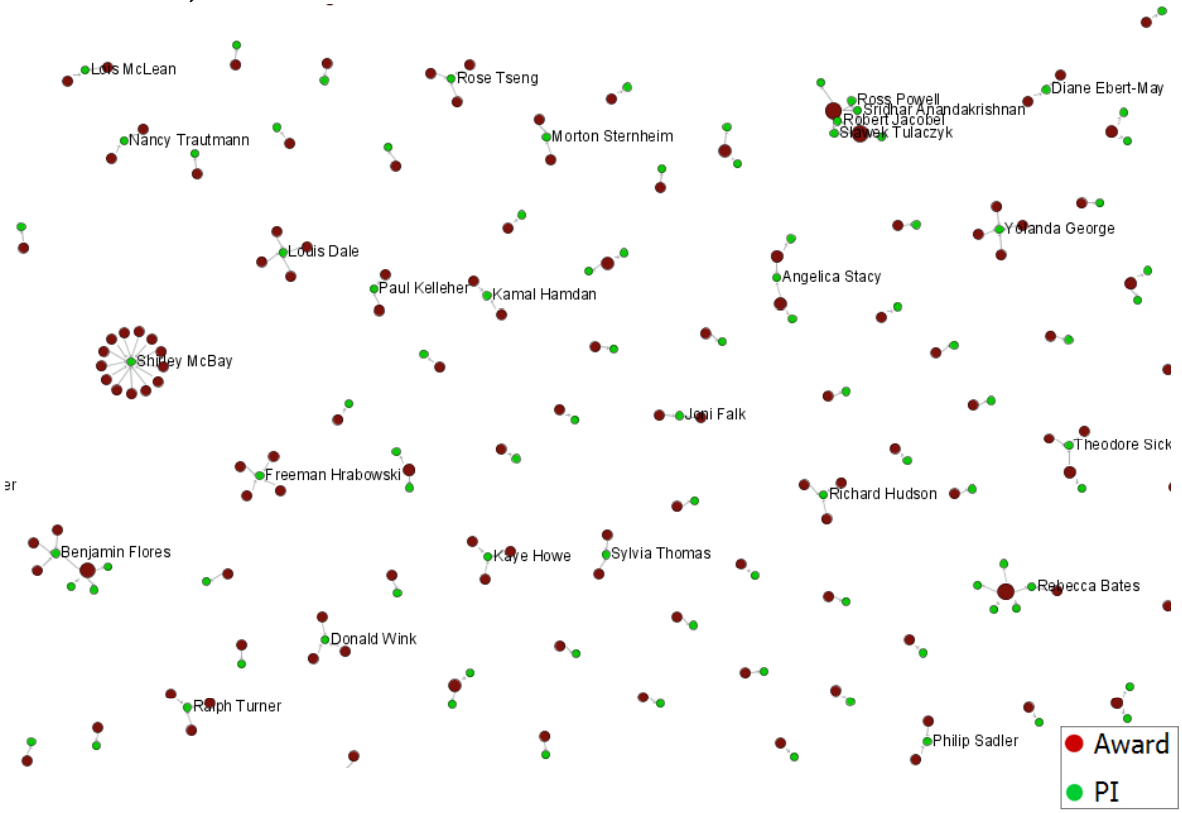
Directed edges from PI to Co-PI

What Projects Fund Which PIs?

Extract Directed Network was selected.
Input Parameters:
Source Column: Title
Text Delimiter: |
Target Column: Principal Investigator
.....
Network Analysis Toolkit (NAT) was selec
Nodes: 2478
Isolated nodes: 0
Edges: 1337
Average total degree: 1.0791
Average in degree: 0.5395
Average out degree: 0.5395
This graph is not weakly connected.
There are 1144 weakly connected compo
The largest connected component consis
Density (disregarding weights): 0.0002
.....
Node Indegree was selected.
.....
Node Outdegree was selected.
.....
GUESS

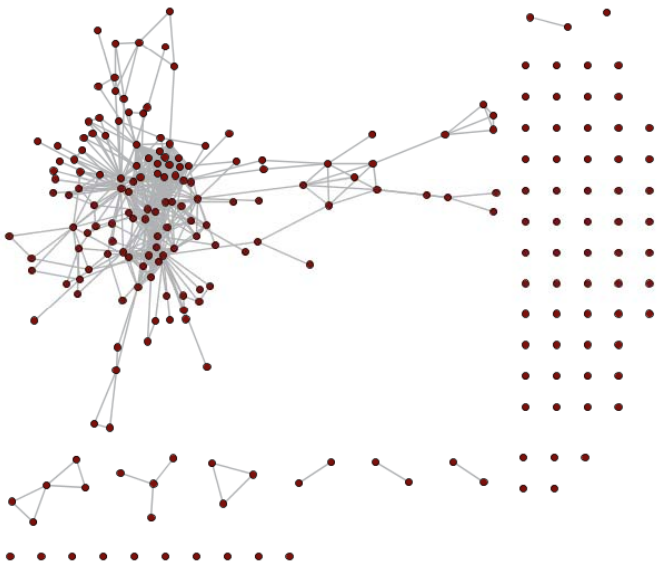


What Projects Fund Which PIs - Zoom

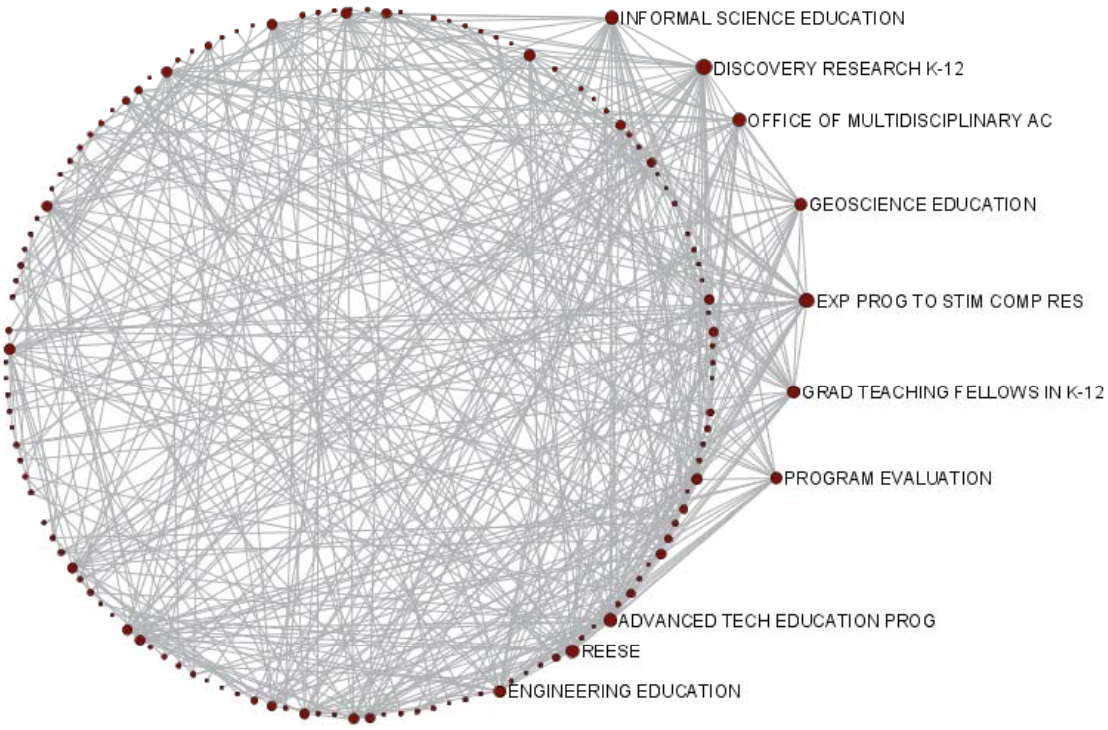


What Programs at NSF are Co-Funding STEM?

Extract Co-Occurrence Network was selected.
Input Parameters:
Text Delimiter: |
Column Name: Program(s)
.....
Node Degree was selected.
.....
Network Analysis Toolkit (NAT) was selected.
Nodes: 226
Isolated nodes: 71
Edges: 483
No self loops were discovered.
Average degree: 4.2743
Density (disregarding weights): 0.019
.....
GUESS
.....
Weak Component Clustering was selected.
79 clusters found
.....
Network Analysis Toolkit (NAT) was selected.
Nodes: 135
Isolated nodes: 0
Edges: 467
No self loops were discovered.
Average degree: 6.9185
Density (disregarding weights): 0.0516
.....
GUESS

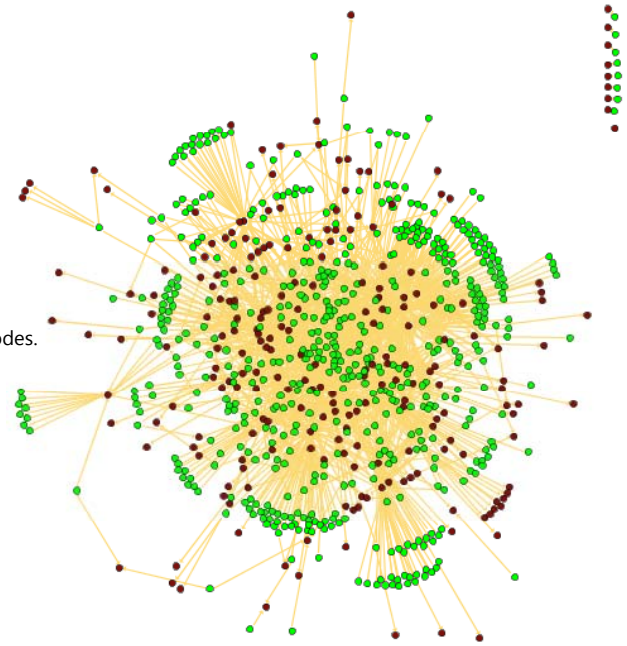


What Programs at NSF are Co-Funding STEM – Giant Component



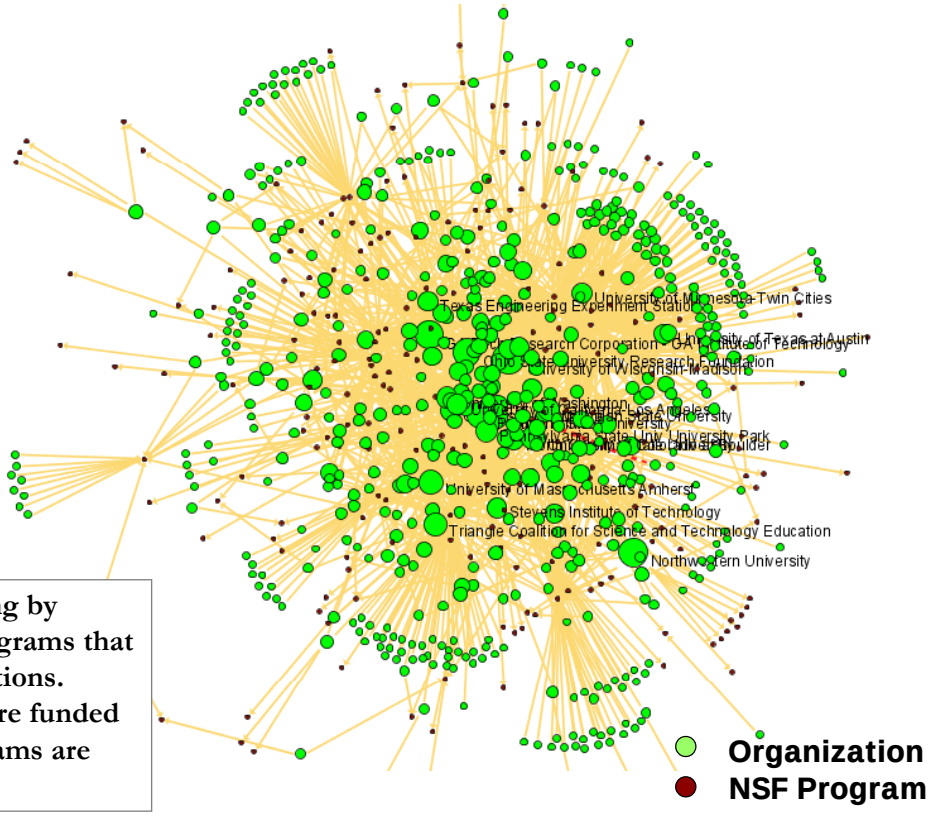
What Organizations are funded by what NSF Programs?

Extract Directed Network was selected.
Input Parameters:
Source Column: Organization
Text Delimiter: |
Target Column: Program(s)
.....
Network Analysis Toolkit (NAT) was selected.
Nodes: 794
Isolated nodes: 1
Edges: 1592
Average total degree: 4.0101
Average in degree: 2.005
Average out degree: 2.005
The largest connected component consists of 777 nodes.
Density (disregarding weights): 0.0025
.....
Node Indegree was selected.
.....
Node Outdegree was selected.
.....
GUESS



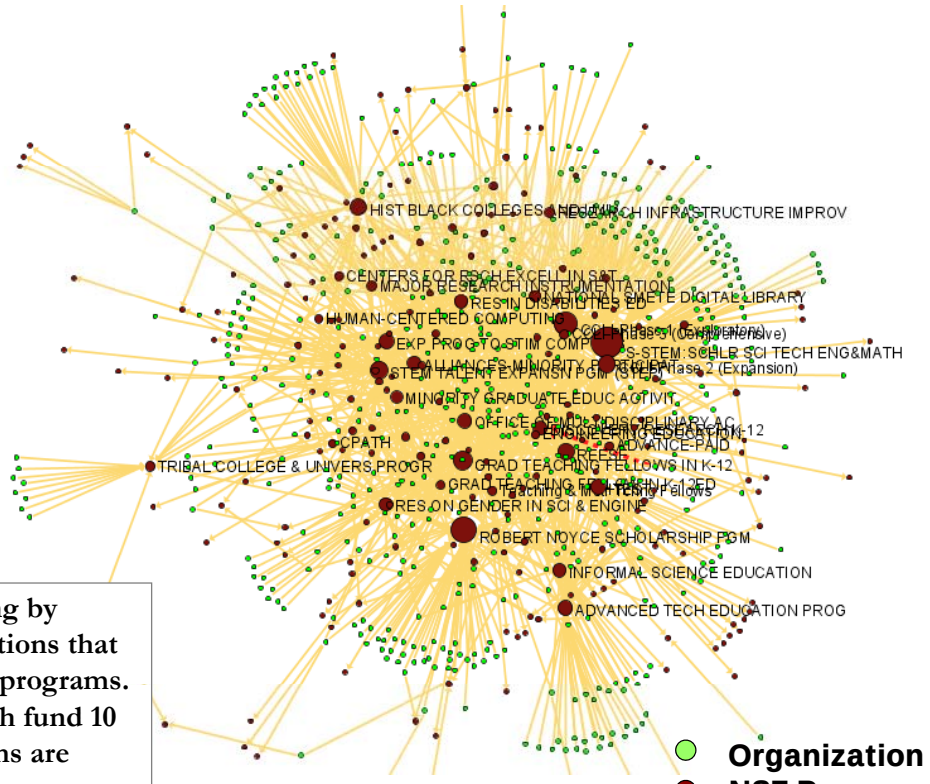
Organization
NSF Program

What Organizations are funded by what NSF Programs?

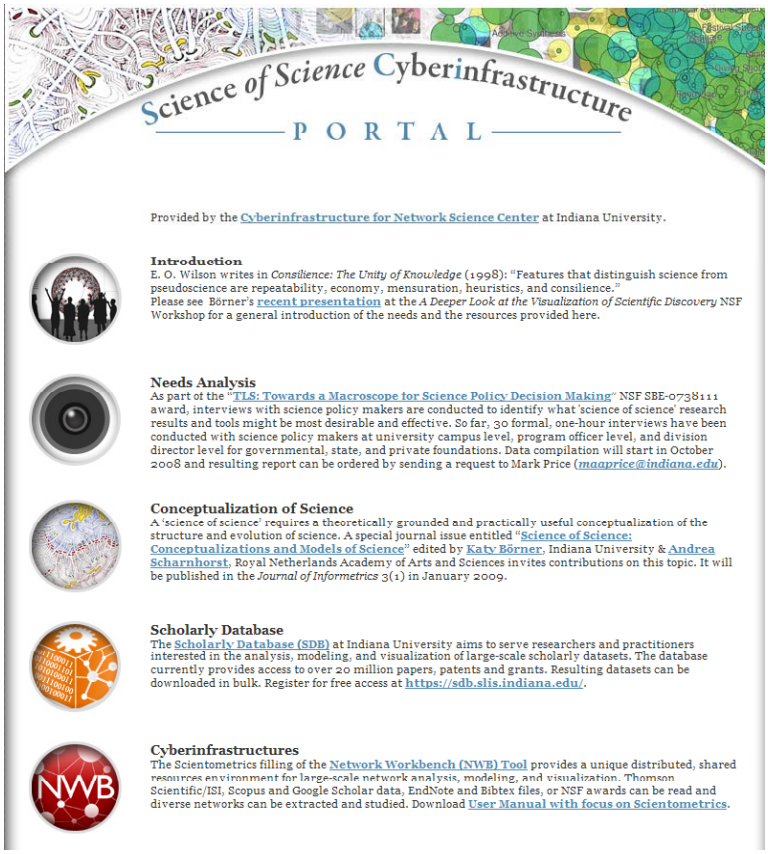


Color and size coding by number of NSF programs that fund these organizations. Institutions which are funded by 10 or more programs are labeled.

What NSF Programs fund how many Organizations?

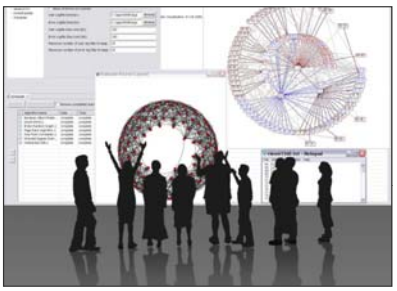


Color and size coding by number of organizations that are funded by these programs. NSF programs which fund 10 or more organizations are labeled.

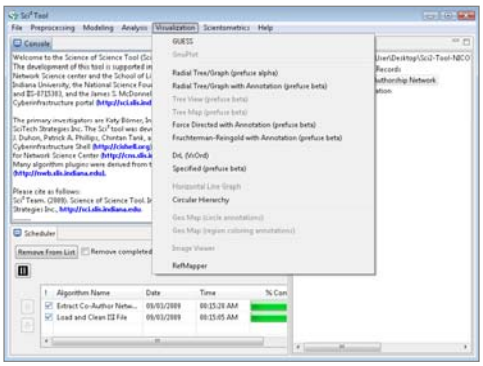


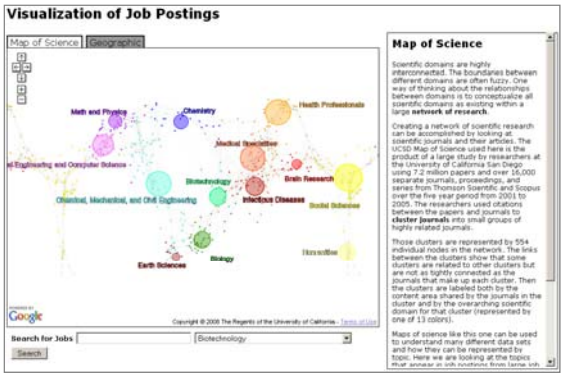
http://sci.slis.indiana.edu

Overview



1. Plug-and-Play Macroscopes
2. Sci2 Tool Introduction and Demos
3. Sample Science of Science Studies



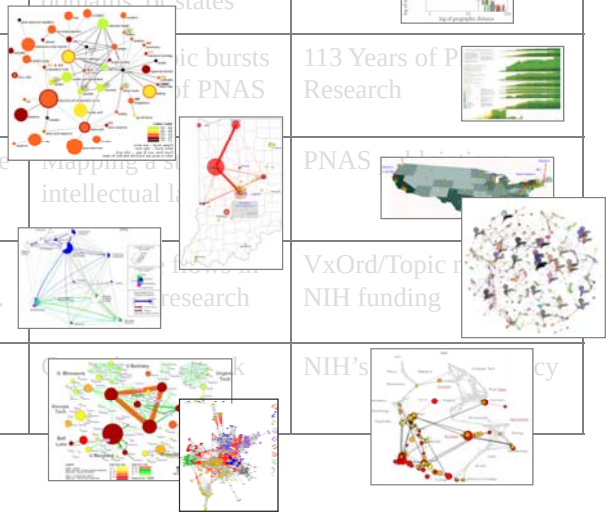


Type of Analysis vs. Level of Analysis

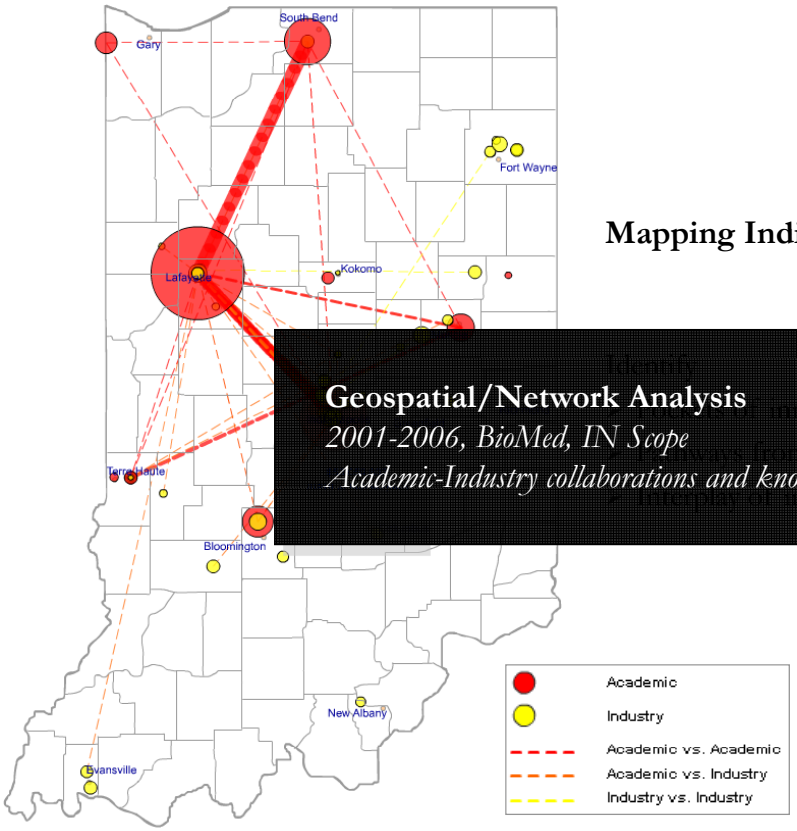
	Micro/Individual (1-100 records)	***Meso/Local (101–10,000 records)***	***Macro/Global (10,000 < records)***
Statistical Analysis/Profiling	Individual person and their expertise profiles	Larger labs, centers, universities, research domains, or states	All of NSF, all of USA, all of science.
Temporal Analysis (When)	Funding portfolio of one individual	Mapping topic bursts in 20-years of PNAS	113 Years of Physics Research
Geospatial Analysis (Where)	Career trajectory of one individual	Mapping a states intellectual landscape	PNAS Publications
Topical Analysis (What)	Base knowledge from which one grant draws.	Knowledge flows in Chemistry research	VxOrd/Topic maps of NIH funding
Network Analysis (With Whom?)	NSF Co-PI network of one individual	Co-author network	NSF's core competency

Type of Analysis vs. Level of Analysis

	Micro/Individual (1-100 records)	***Meso/Local (101–10,000 records)***	***Macro/Global (10,000 < records)***
Statistical Analysis/Profiling	Individual person and their expertise profiles	Larger labs, centers, universities, research domains, or states	All of NS SA, all of sci
Temporal Analysis (When)	Funding portfolio of one individual	ic bursts f PNAS	113 Years of P Research
Geospatial Analysis (Where)	Career trajectory of one individual	Mapping a S intellectual l	PNAS
Topical Analysis (What)	s.	research	VxOrd/Topic m NIH funding
Network Analysis (With Whom?)	NSF vork of one	k	NIH's cy



Mapping Indiana's Intellectual Space

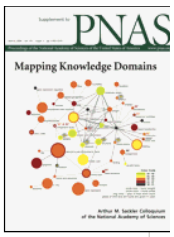


Geospatial/Network Analysis
2001-2006, BioMed, IN Scope
Academic-Industry collaborations and knowledge diffusion

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.



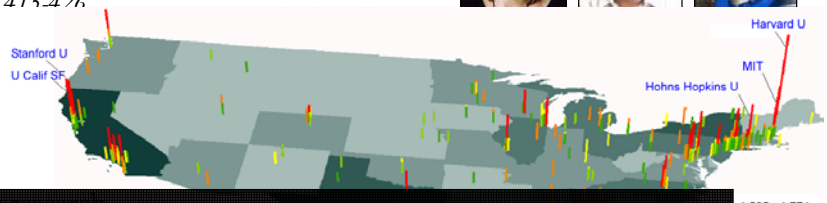
Temporal/Topical Analysis
1982-2001, US, PNAS (BioMed) Scope
Topic evolution and bursts

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

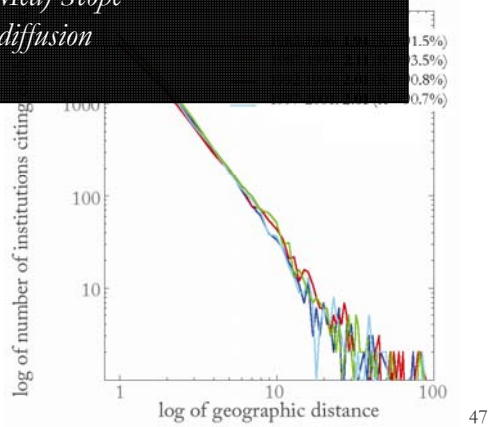
Börner, Katy, Penumarthy, Shashikant, Meiss, Mark and Ke, Weimao. (2006) Mapping the Diffusion of Scholarly Knowledge Among Major U.S. Research Institutions. Scientometrics. 68(3), pp. 415-426.

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 produce high quality data and high quality research?
3. Does the Internet lead to more evenly distributed patterns, i.e., more citation links between papers produced at geographically distant research instructions?



Temporal/Geospatial Analysis
1982-2001, US, PNAS (BioMed) Scope
Citation impact and knowledge diffusion



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.

Research Collaborations by the Chinese Academy of Sciences

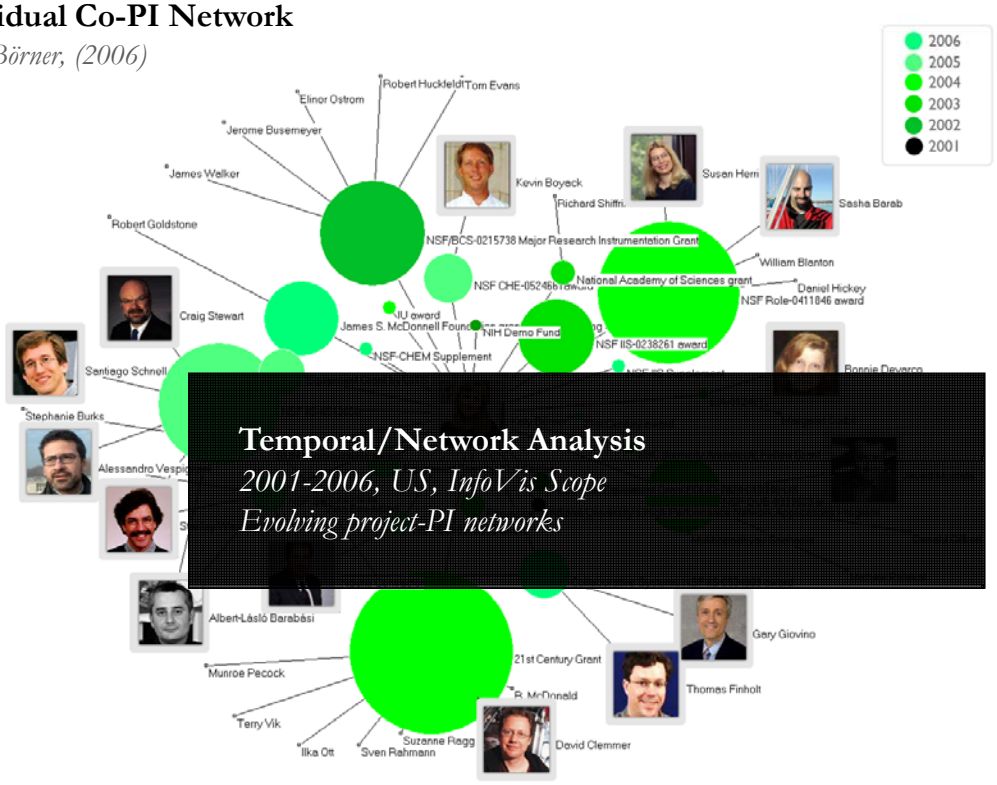
By Weixia (Bonnie) Huang, Russell J. Duhon, Elisha F. Hardy, Katy Börner, Indiana University, USA

Geospatial Analysis
World, Chinese Academy of Science
Collaboration and knowledge diffusion via co-author networks

This map highlights the research co-authorship collaborations of the Chinese Academy of Sciences with locations in China and countries around the world. The large geographic map shows the research collaborations of all CAS institutes. Each smaller geographic map shows the research collaborations by the CAS researchers in one province-level administrative division. Collaborations between CAS researchers are not included in the data. On each map, locations are colored on a logarithmic scale by the number of collaborations from red to yellow. The darkest red is 3,395 collaborations by all of CAS with researchers in Beijing. Also, flow lines are drawn from the location of focus to all locations collaborated with. The width of the flow line is linearly proportional to the number of collaborations with the locations it goes to, with the smallest flow lines representing one collaboration and the largest representing differing amounts on each geographic map.

Individual Co-PI Network

Ke & Börner, (2006)



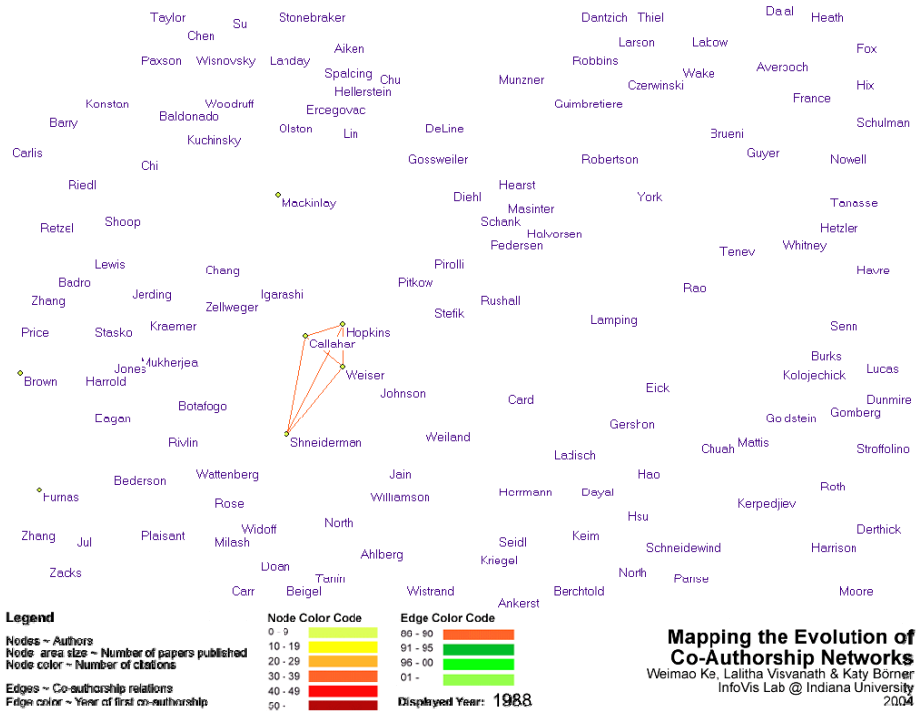
Temporal/Network Analysis

2001-2006, US, InfoVis Scope

Evolving project-PI networks

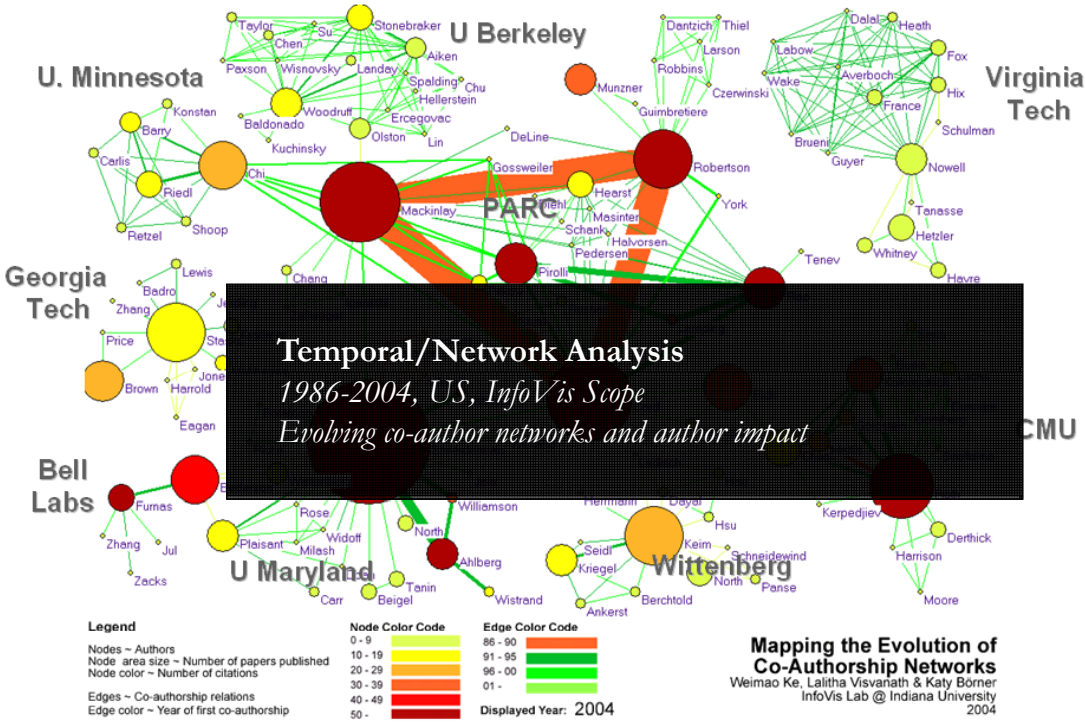
Mapping the Evolution of Co-Authorship Networks

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

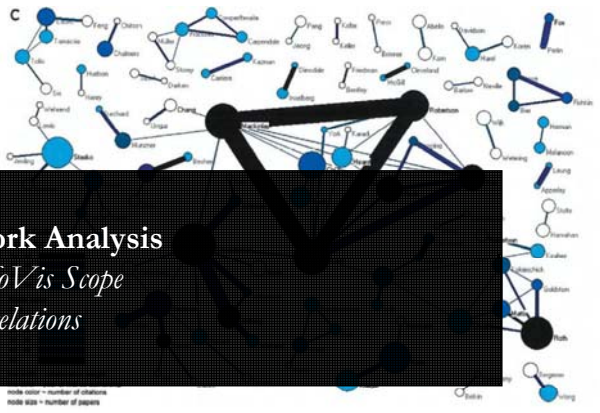


Studying the Emerging Global Brain: Analyzing and Visualizing the Impact of Co-Authorship Teams

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

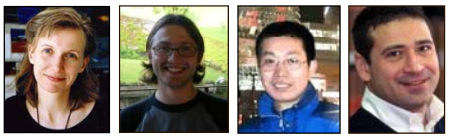
Research question:

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



Contributions:

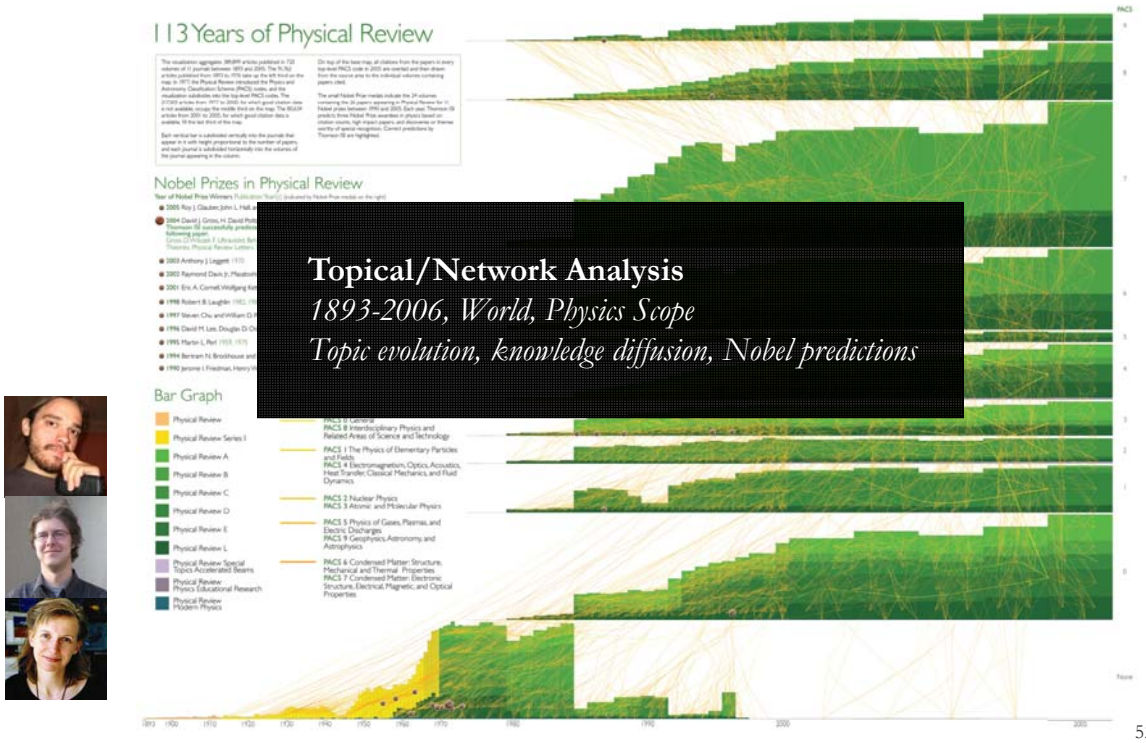
- New approach to allocate citational credit.
- Novel weighted gra
- Visualization of th co-author network
- Centrality measure to identify author impact.
- Global statistical analysis of paper production and citations in correlation with co-authorship team size over time.
- Local, author-centered entropy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



113 Years of Physical Review

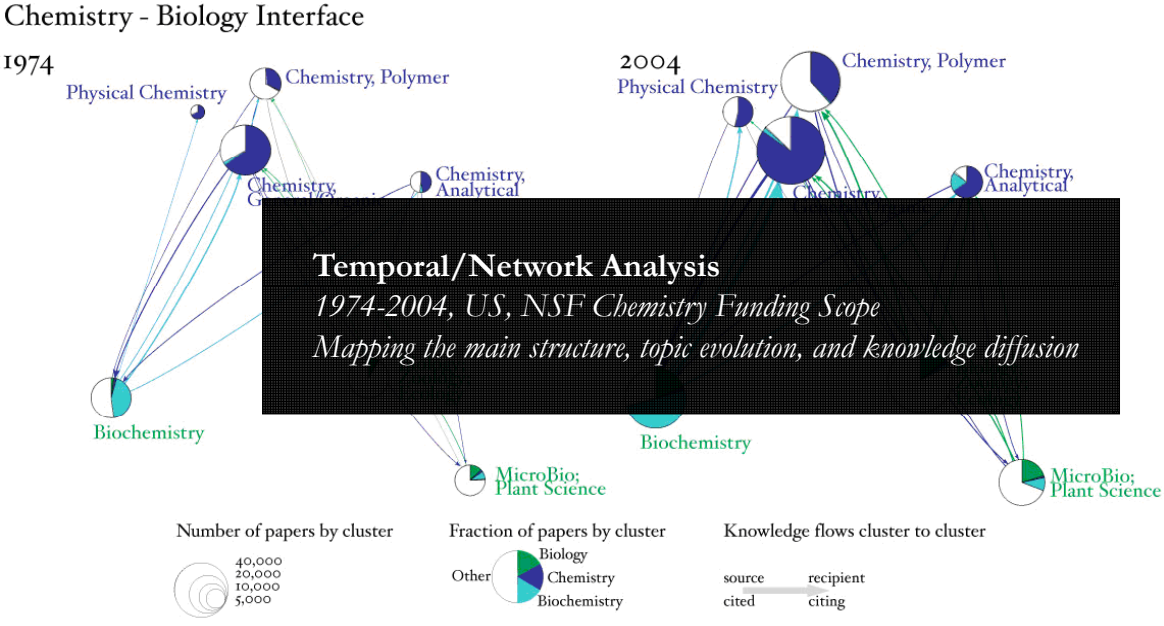
http://scimaps.org/dev/map_detail.php?map_id=171

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



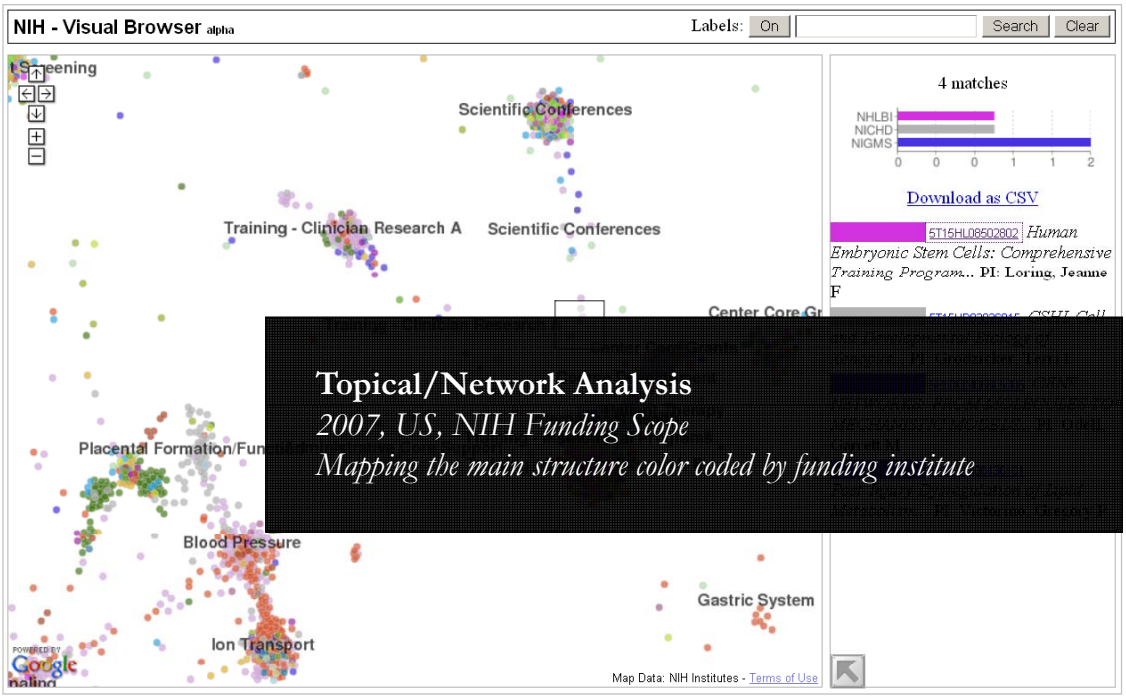
Topical Composition and Knowledge Flow Patterns in Chemistry Research for 1974 and 2004

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



Interactive Maps of Science – NIH Funding

Google maps with charts and tables

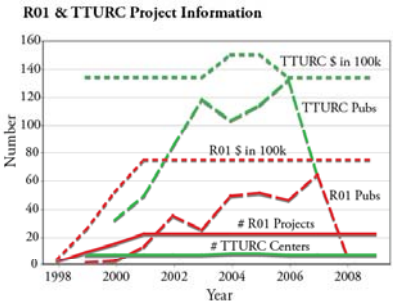


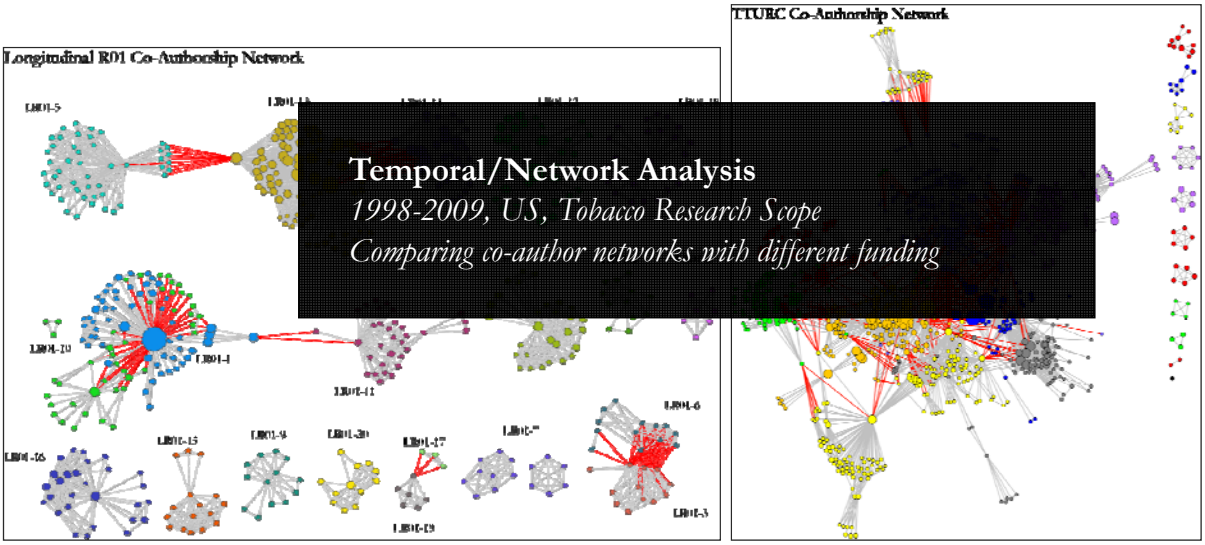
http://scimaps.org/maps/nih/2007

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.

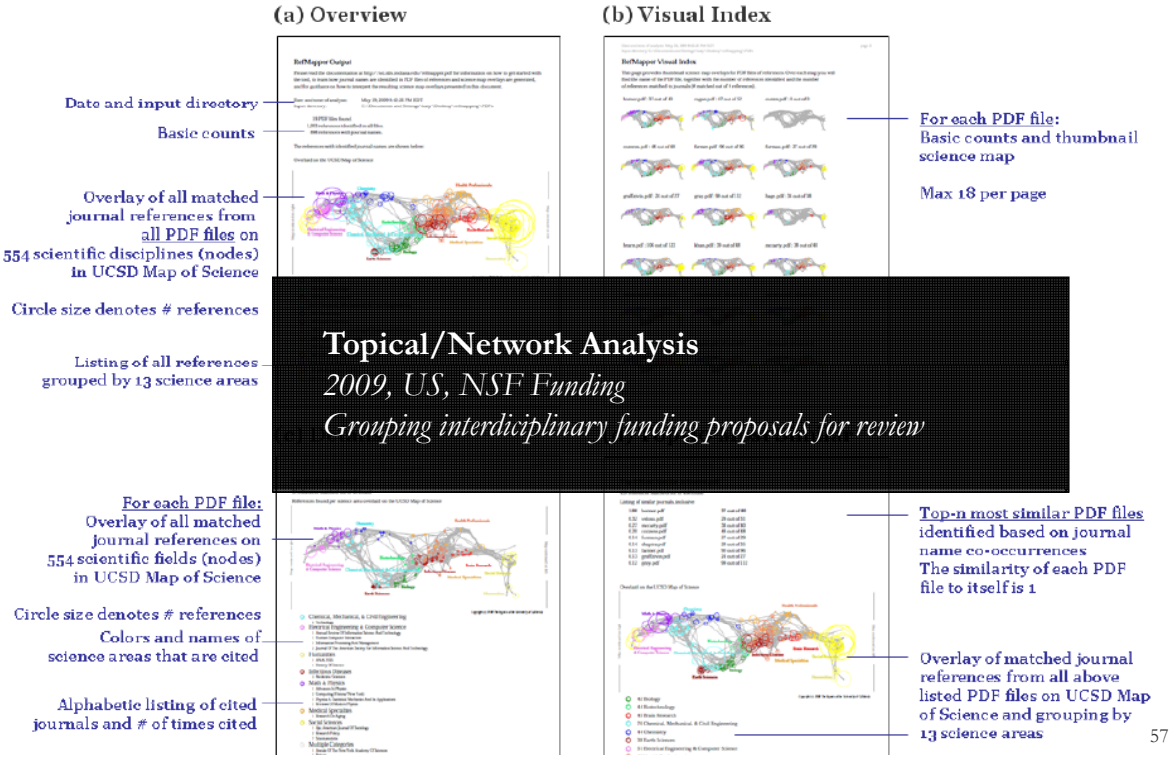
Zoss & Börner, forthcoming.





Reference Mapper

Duhon & Börner, forthcoming.



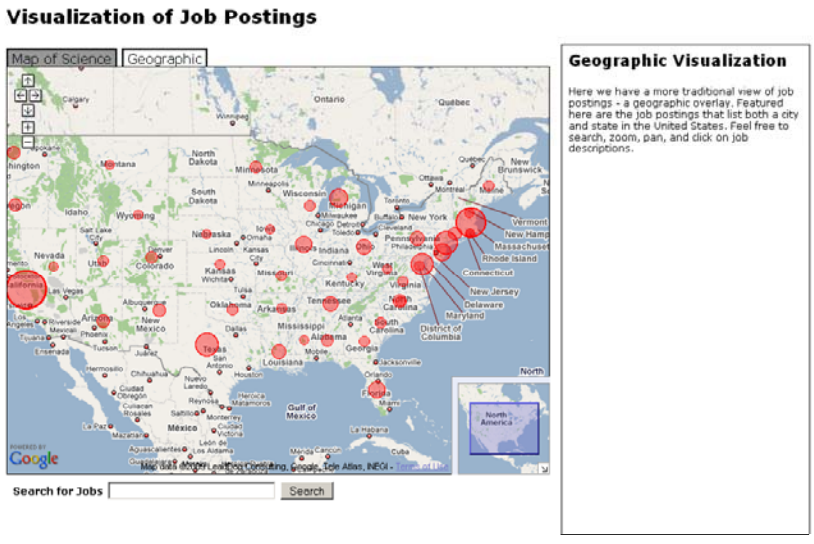
Mapping S&T Job Market Data in Real Time – GeoMap

Angela Zoss, Michael Conover

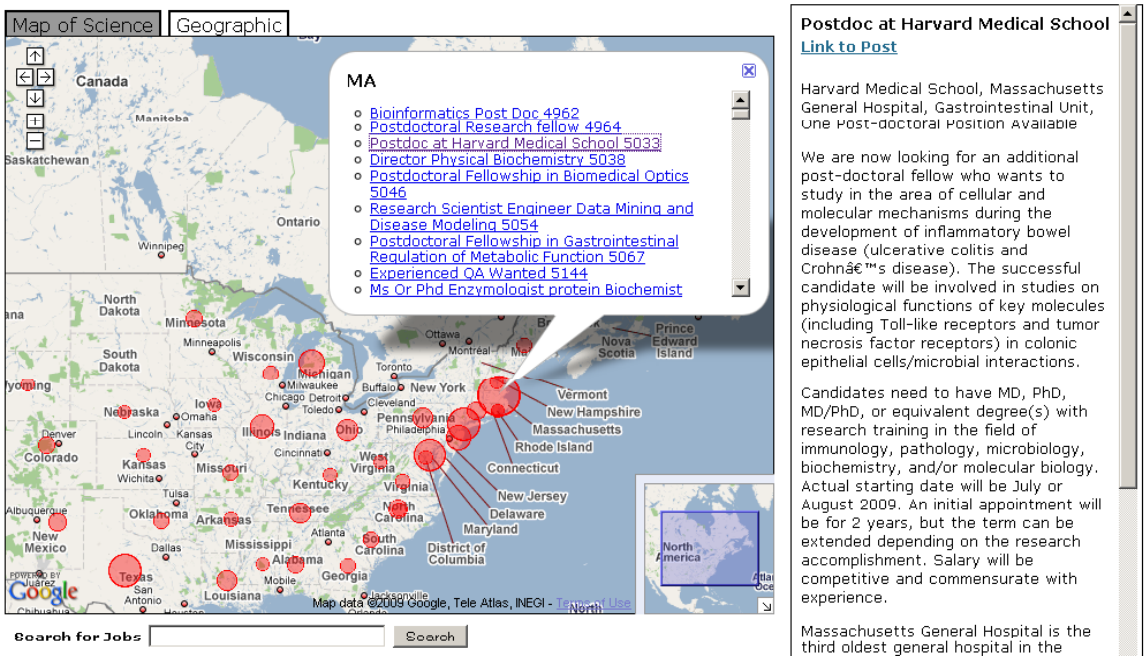
Data

Thousands of full-text, location-specific, time stamped job postings from Nature Jobs and Science Careers RSS feeds. The posts have been parsed and stored in a relational MySQL database.

Jobs have been geolocated on a Google map.



Visualization of Job Postings

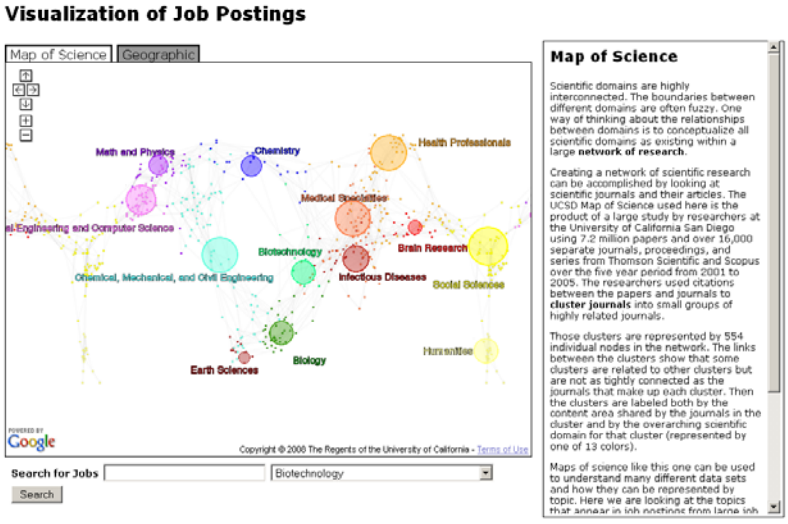


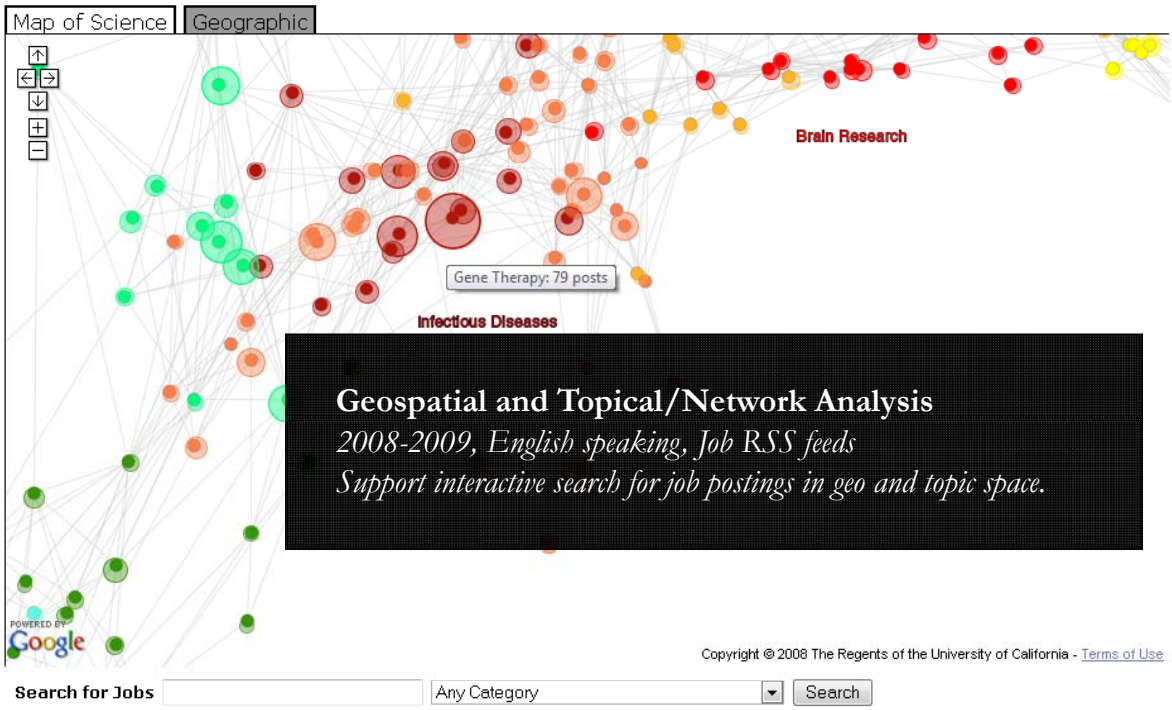
Mapping S&T Job Market Data in Real Time – SciMap

Angela Zoss, Michael Conover

The **UCSD Map of Science** used here is the product of a large study by researchers at the University of California - San Diego using 7.2 million papers and over 16,000 separate journals, proceedings, and series from Thomson Scientific and Scopus over the five year period from 2001 to 2005.

Jobs were associated with nodes in the Map of Science by way of keyword extraction.

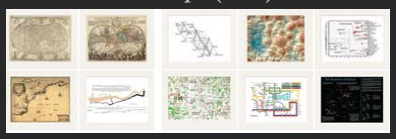




Mapping Science Exhibit – 10 Iterations in 10 years

http://scimaps.org

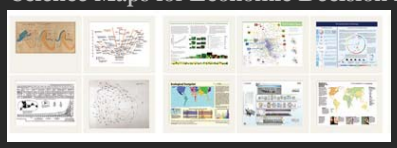
The Power of Maps (2005)



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)

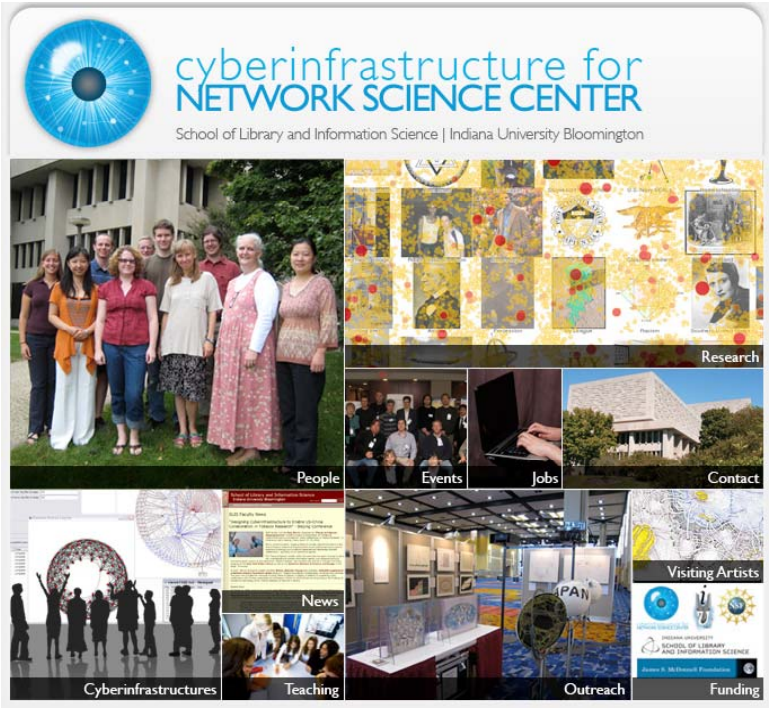
Science Forecasts (2013)

Telling Lies With Science Maps (2014)

Exhibit has been shown in 72 venues on four continents. Currently at
- NSF, 10th Floor, 4201 Wilson Boulevard, Arlington, VA
- Wallenberg Hall, Stanford University, CA
- Center of Advanced European Studies and Research, Bonn, Germany
- Science Train, Germany.

ORDER MAPS

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



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