

The Science of Science (Sci²) Tool and Its Utility for Research

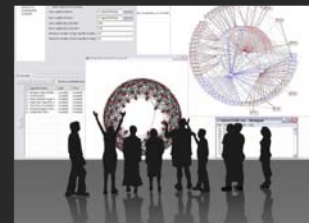
Katy Börner and the Cyberinfrastructure for Network Science Center

School of Library and Information Science
Indiana University, Bloomington, IN

katy@indiana.edu

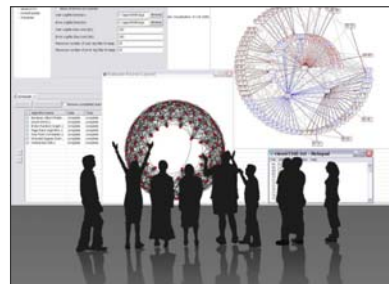
With special thanks to Kevin W. Boyack, Micah Linnemeier, Russell J. Duhon, Patrick Phillips, Joseph Biberstine, Chintan Tank Nianli Ma, Angela M. Zoss, Hanning Guo, Mark A. Price, Scott Weingart

*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 Flickr 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.

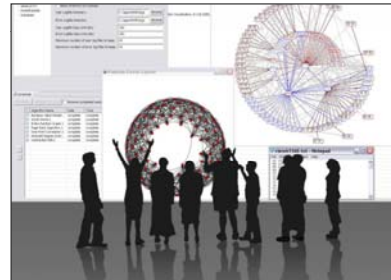
Modularity: The design of software modules with well defined functionality that can be flexibly combined helps reduce costs, makes it possible to have many contribute, and increases flexibility in tool development, augmentation, and customization.

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 Macroscopes



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.

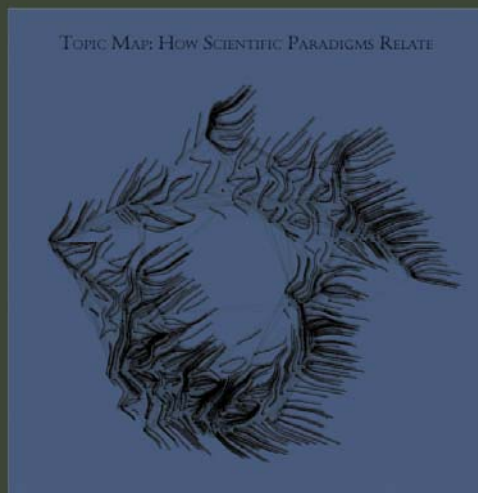
Supported Platforms: If the software is to be used via Web interfaces then Web services need to be implemented. If a majority of domain experts prefers a stand-alone tool running on a specific operating system then a different deployment is necessary.

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.*



You may run your finger over each of these maps to control the lighting on the other: touching a place on the world map will light up topics studied in that place; touching a paradigm on the topic map will light up the places that study that topic.

Nanotechnology

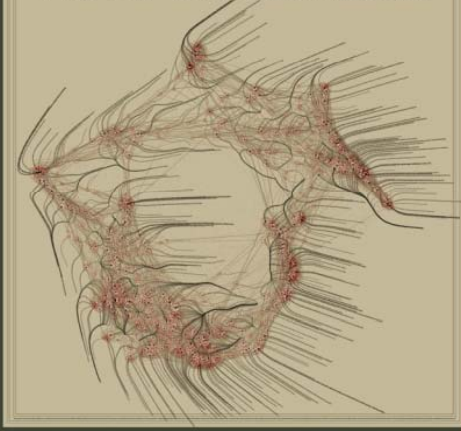
This overlay shows the distribution of nanotechnology within the paradigms of science. The majority of current work in nanotechnology takes place in physics, chemistry, and materials science, at the upper right portion of the map. However, an increasing amount of nanotechnology is being applied in the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

All Topics <i>Sweep through all 376 scientific paradigms</i>	Nanotechnology <i>Science on the tiny scale of molecules</i>	Francis H. C. CRICK <i>Co-discovered DNA's double helix</i>	Albert EINSTEIN <i>Revitalized physics with Relativity theories</i>	Michael E. FISHER <i>Models critical phase transitions of matter</i>	Susan T. FISKE <i>Connects perception and stereotypes</i>
Sustainability <i>The science behind our long-term hopes</i>	Biology & Chemistry <i>The interface between these two vital fields</i>	Joshua LEDERBERG <i>Pioneer in bacterial genetic mechanisms</i>	Derek J. de Solla PRICE <i>Known as the "Father of Scientometrics"</i>	Richard N. ZARE <i>Uses laser chemistry in molecular dynamics</i>	About this display <i>People & organizations that helped create it</i>

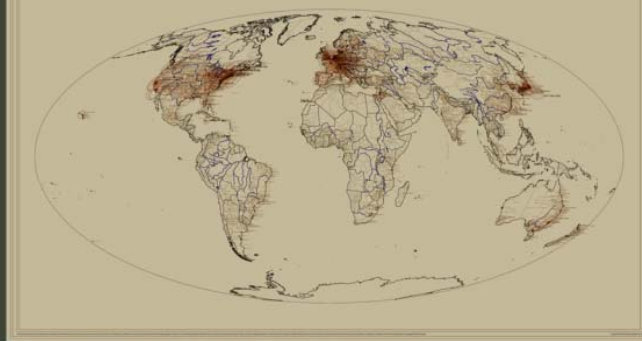
We sweep slowly through adjoining related topics, lighting up the places in the world that study each topic. You may select a subset of the topics that deal with these three interesting subjects by touching it.

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

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 一些与人类寄予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

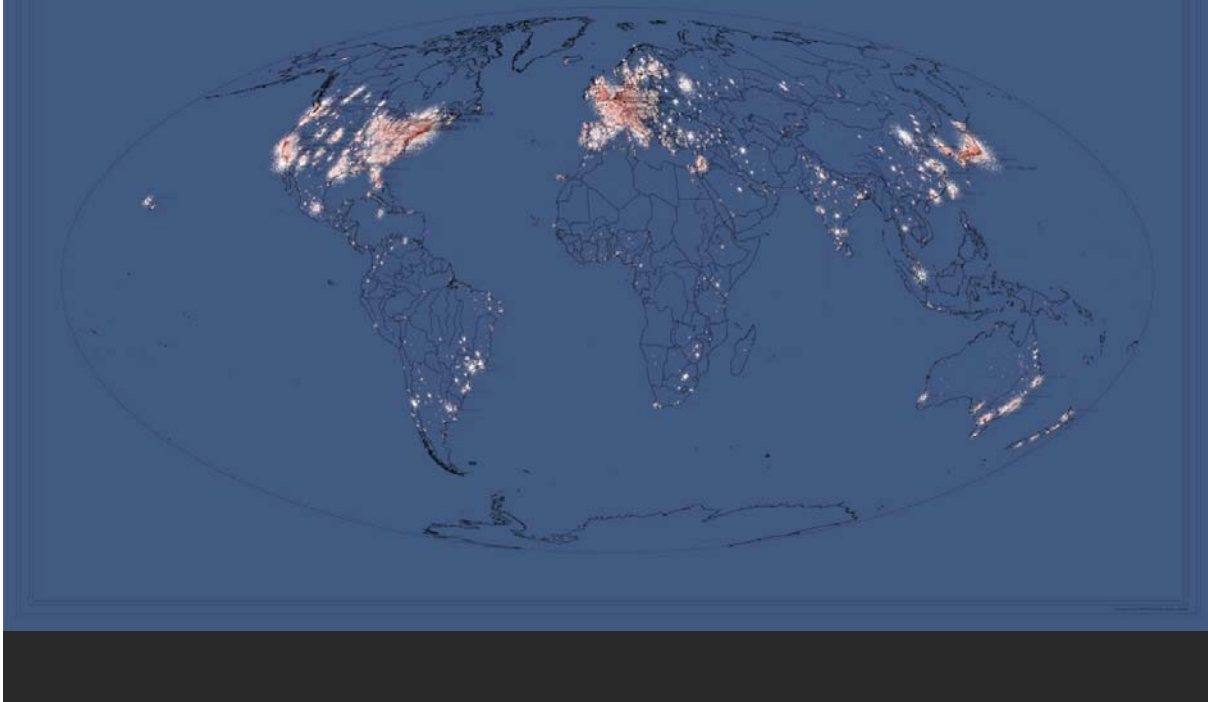
光标缓慢的扫过所有相互关联的科学学科，每一个学科以及从事这方面科学研究的研究机构在世界地图上的位置会被逐一点亮。首先，显示屏会点亮那些产出论文最多、最活跃的科学学科，然后那些小学科或冷门学科会被逐一点亮。

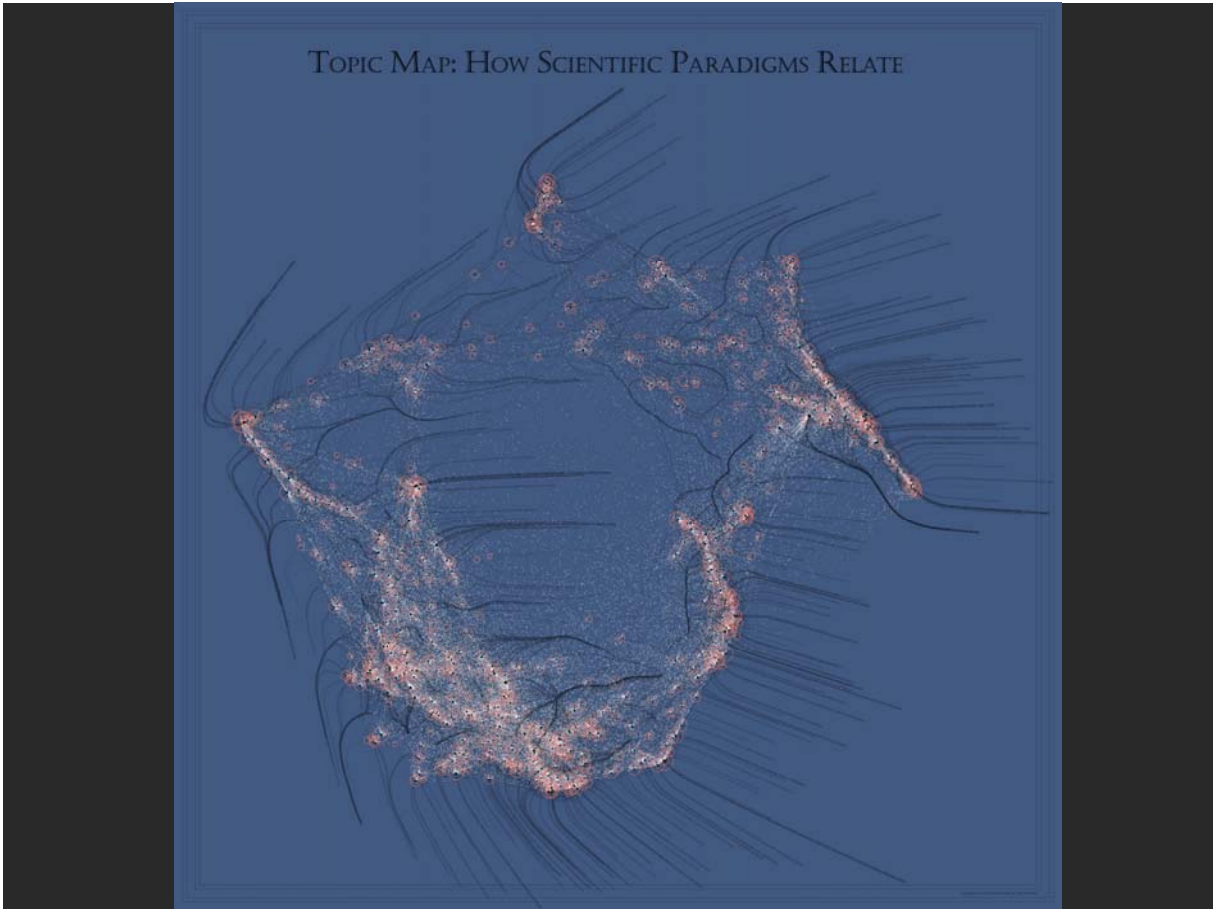
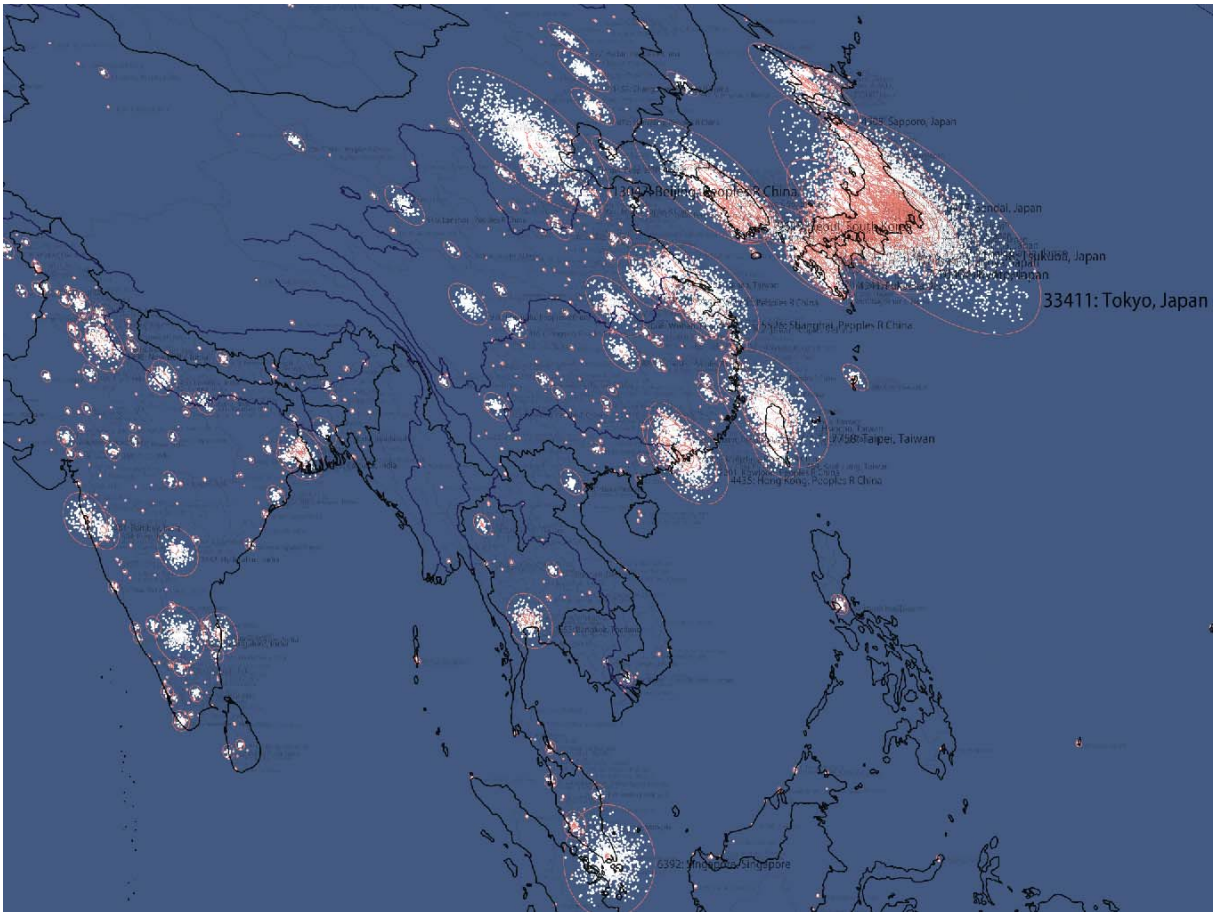
探索某个学者的科学著作的影响力的传播

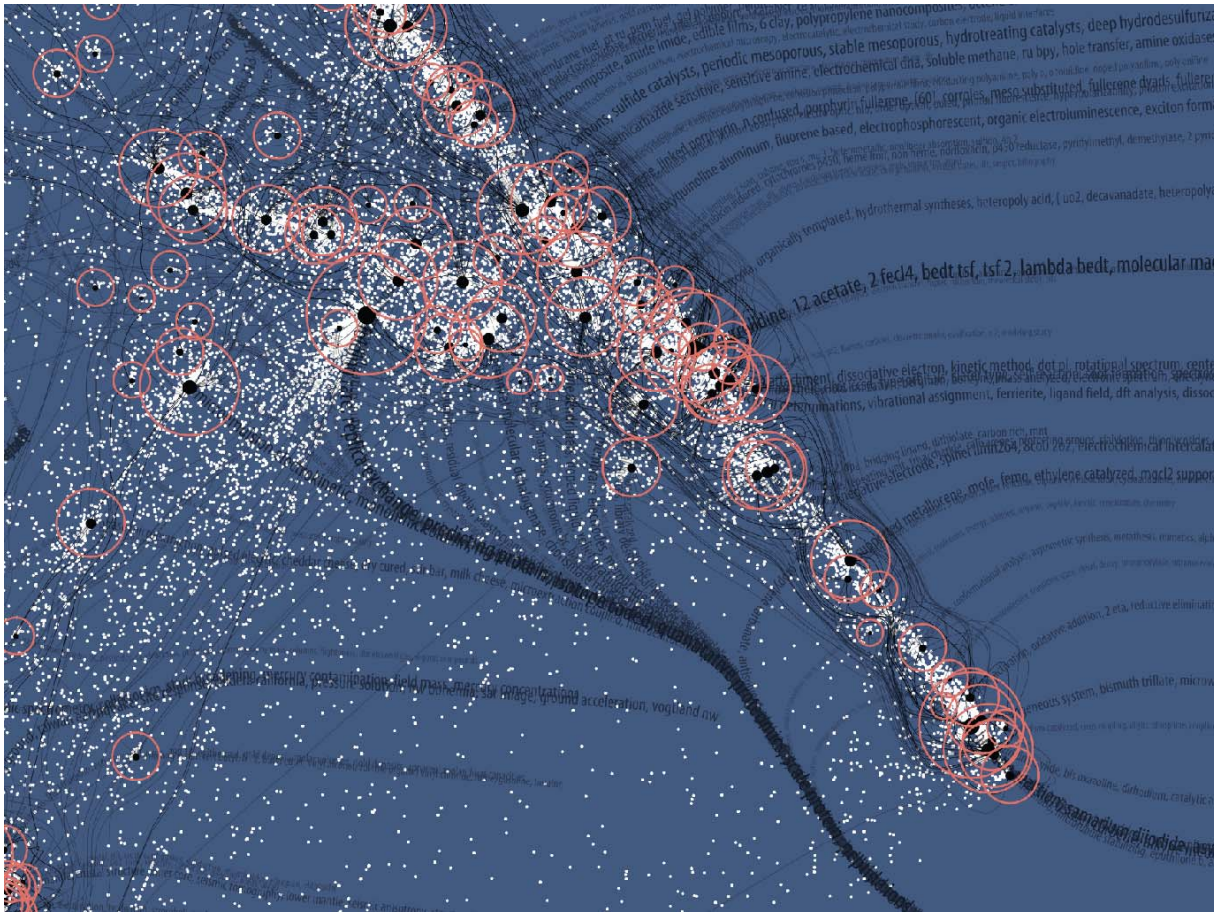
弗郎西·科里克 DNA双螺旋结构的发现者之一	阿尔伯特·爱因斯坦 用相对论重新激活了物理学	迈克尔·费舍尔 发现了物质转变模式的关键步骤	苏珊·费斯克 研究人的认知是如何产生偏见的
约舒亚·雷德伯格 细菌遗传机制研究先驱	德里克·德索拉·普里斯 著名的“科学计量学之父”	理查德·扎尔 采用激光化学技术研究分子动态分布	关于本次展览 与此展览相关人员及机构

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。第四步，显示屏点亮所有引用了在第三步中被点亮的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。

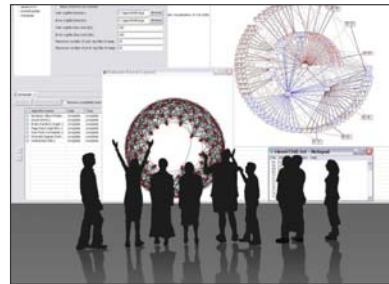
GEOGRAPHIC MAP: WHERE SCIENCE GETS DONE



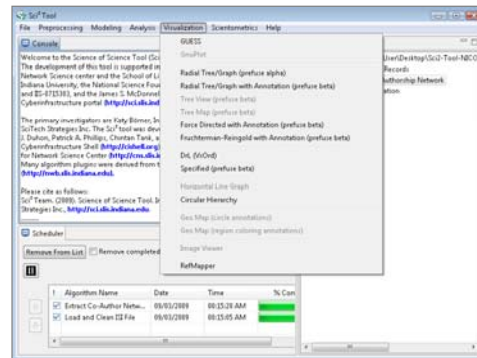




Overview



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>



Information Visualization Cyberinfrastructure

<http://iv.slis.indiana.edu>



Network Workbench Tool + Community Wiki

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



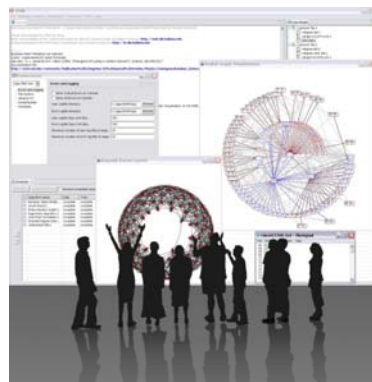
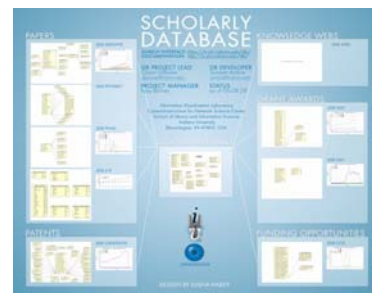
Sci² Tool and Science of Science CI Portal

<http://sci.slis.indiana.edu>



Epidemics Cyberinfrastructure

<http://epic.slis.indiana.edu/>





Macroscopic Design

Network Workbench Tool

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



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

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.



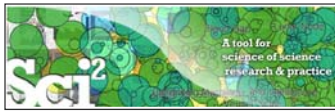
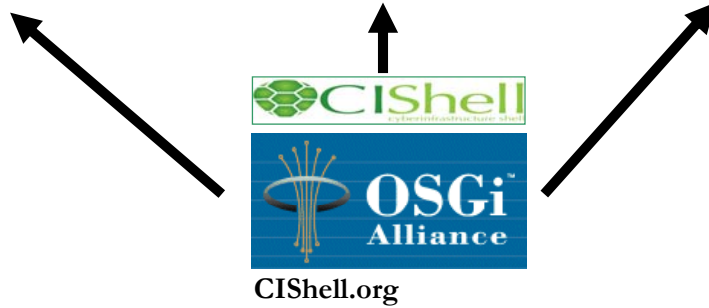
<http://sci.slis.indiana.edu>

Epidemics Research Tool

An open computational infrastructure for epidemics research.



<http://epic.slis.indiana.edu>

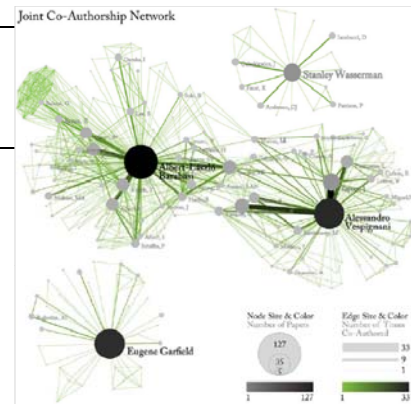
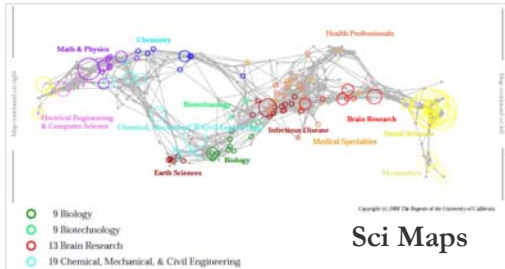


Sci² Tool

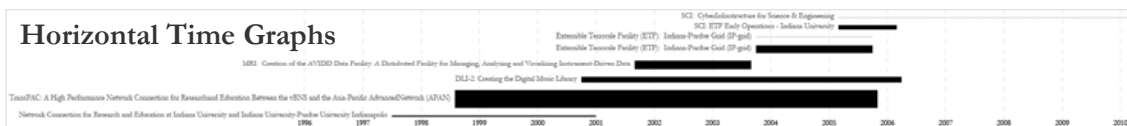
<http://sci.slis.indiana.edu>

“Open Code for S&T Assessment”

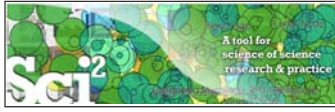
Branded OSGi/CISHell based tool with NWB plugins and many new plugins.



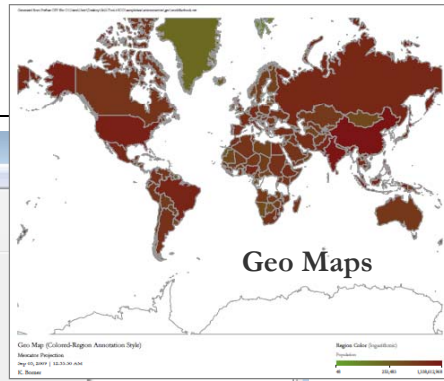
Horizontal Time Graphs



Börner, Katy, Huang, Weixia (Bonnie), Linnemeier, Micah, Dubon, Russell Jackson, Phillips, Patrick, Ma, Nianli, Zoss, Angela, Guo, Hanning & Price, Mark. (2009). *Retz-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



Sci² Tool

File Preprocessing Modeling Analysis Visualization Scientometrics Help

Console

Welcome to the Science of Science Tool (Sci²). The development of this tool is supported in Network Science center and the School of Li Indiana University, the National Science Foundation-0715303, and the James S. McDonnell Cyberinfrastructure portal (<http://sci.slis.indiana.edu>)

The primary investigators are Katy Börner, In SciTech Strategies Inc. The Sci² tool was developed by J. Duhon, Patrick A. Phillips, Chintan Tank, a Cyberinfrastructure Shell (<http://cishell.org>) for Network Science Center (<http://cns.slis.indiana.edu>). Many algorithm plugins were derived from the Network Science Center (<http://nwb.slis.indiana.edu>).

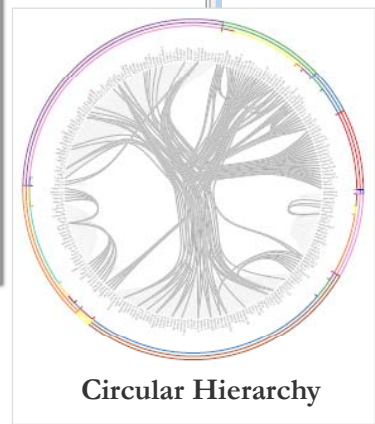
Please cite as follows:
Sci² Team. (2009). Science of Science Tool. In SciTech Strategies Inc., <http://sci.slis.indiana.edu>.

Scheduler

Remove From List Remove completed

Algorithm Name	Date	Time	% Complete
Extract Co-Author Network	09/03/2009	00:15:20 AM	100%
Load and Clean ISI File	09/03/2009	00:15:05 AM	100%

- GUESS
- GnuPlot
- 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 annotations)
- Geo Map (region coloring annotations)
- Image Viewer
- RefMapper

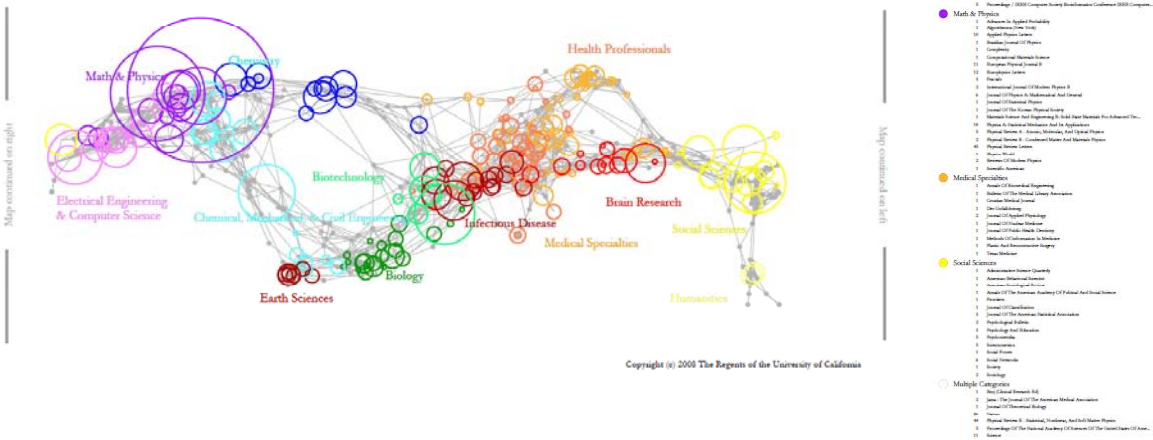


Topic Mapping: UCSD Science Map

Journal locations for FourNetSciResearchers.isi

342 journal references matched out of 361 found.

These 342 references are associated with 13 of 13 disciplines of science and 238 of 554 research specialties in the UCSD Map of Science.





Sci² Tool: Supported Data Formats

Personal Bibliographies

- Bibtext (.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, Bibtext, 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)

23



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

HTTS

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

HTTS

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

24



Sci² Tool: Algorithms cont.

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

<p>-----</p> <p>Extract K-Core</p> <p>Annotate K-Core-ness</p> <p>-----</p> <p>HITS</p> <p>PageRank</p> <p>Weighted & Directed</p> <p>HITS</p> <p>Weighted PageRank</p> <p>Textual</p> <p>Burst Detection</p>	<p>Visualization</p> <p>GnuPlot</p> <p>GUESS</p> <p>Image Viewer</p> <p>-----</p> <p>Radial Tree/Graph (prefuse alpha)</p> <p>Radial Tree/Graph with Annotation (prefuse beta)</p> <p>Tree View (prefuse beta)</p> <p>Tree Map (prefuse beta)</p> <p>Force Directed with Annotation (prefuse beta)</p> <p>Fruchterman-Reingold with Annotation (prefuse beta)</p> <p>-----</p> <p>DrL (VxOrd)</p> <p>Specified (prefuse beta)</p> <p>-----</p> <p>Horizontal Line Graph</p> <p>Circular Hierarchy</p> <p>Geo Map (Circle Annotation Style)</p> <p>Geo Map (Colored-Region Annotation Style)</p>	<p>Scientometrics</p> <p>Remove ISI Duplicate Records</p> <p>Remove Rows with Multitudinous Fields</p> <p>Detect Duplicate Nodes</p> <p>Update Network by Merging Nodes</p> <p>-----</p> <p>Extract Directed Network</p> <p>Extract Paper Citation Network</p> <p>Extract Author Paper Network</p> <p>-----</p> <p>Extract Co-Occurrence Network</p> <p>Extract Word Co-Occurrence Network</p> <p>Extract Co-Author Network</p> <p>Extract Reference Co-Occurrence (Bibliographic Coupling) Network</p> <p>-----</p> <p>Extract Document Co-Citation Network</p>
--	--	---

25



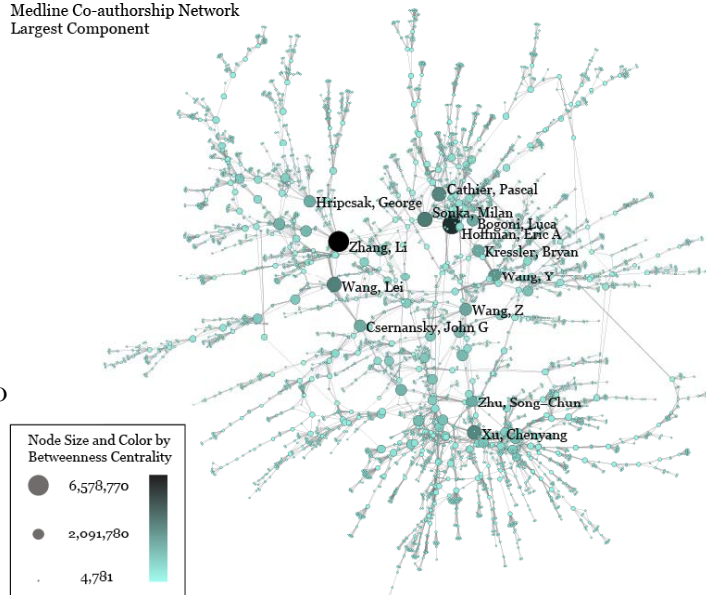
NWB=Sci² Tool: Output Formats

- 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.

Medline Co-authorship Network
Largest Component



26



Sample Study – NSF Funding of STEM

Using NSF Awards Search:

<http://www.nsf.gov/awardsearch>

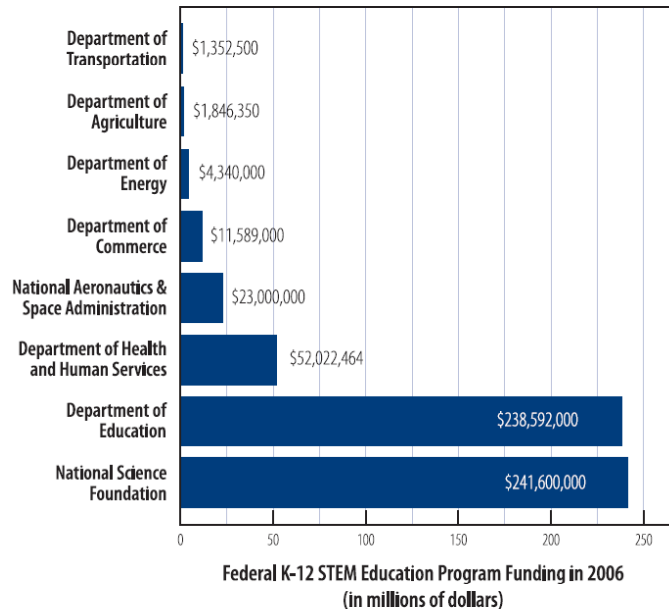
download relevant NSF awards that have “stem” and “education” in title, abstract, and awards.

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

NSF - Award Search - Search All Fields - Mozilla Firefox

http://www.nsf.gov/awardsearch/afSearch.do?SearchType=afSearch&page=4&QueryText="stem"+and

SEARCH
NSF Web Site

HOME | FUNDING | AWARDS | DISCOVERIES | NEWS | PUBLICATIONS | STATISTICS | ABOUT | FastLane

Award Search Send Comments | Award Search Help

Awardee Information | Program Information | Search All Free-Text | Search All Fields | More Options

Hint: The text field below 'Search Award For' searches the title, abstract, and award number fields.

Search Award For: "stem" and "education"

Restrict to Title Only:

Awardee Information
Principal Investigator

Done

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

NSF - Award Search - Search All Fields - Mozilla Firefox

http://www.nsf.gov/awardsearch/afSearch.do?SearchType=afSearch&page=4&QueryText="stem"+anc

Search Results

Results are sorted by award date, with the most recent awards at the top. Click on a column heading to re-sort the results. The up/down arrows at the right of each column title control whether the sort is ascending or descending. To view the abstract, click on the award number or title. Click on the data in other columns to perform a new search with that parameter.

Refine Search

1,340 awards found, displaying 1 to 50.

[First/Prev] 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 [Next/Last]

Award Number	Title	NSF Organization	Program(s)	Start Date	Principal Investigator	State	Organization
0941971	Self-Efficacy and Student Characteristics as Predictors of Success for Supplemental Instruction	DUE	CCLI-Phase 1 (Exploratory), S-STEM: SCHLR SCI	09/01/2010	Feakes, Debra	TX	Texas State U Marcos

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 Molecu	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 (GILC)	3DC	G C: ACT V T I C S I S C I E N C E	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 INSTRUMENTATION	William Smith	DC	AURA/National Optical Astronomy Obser	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.

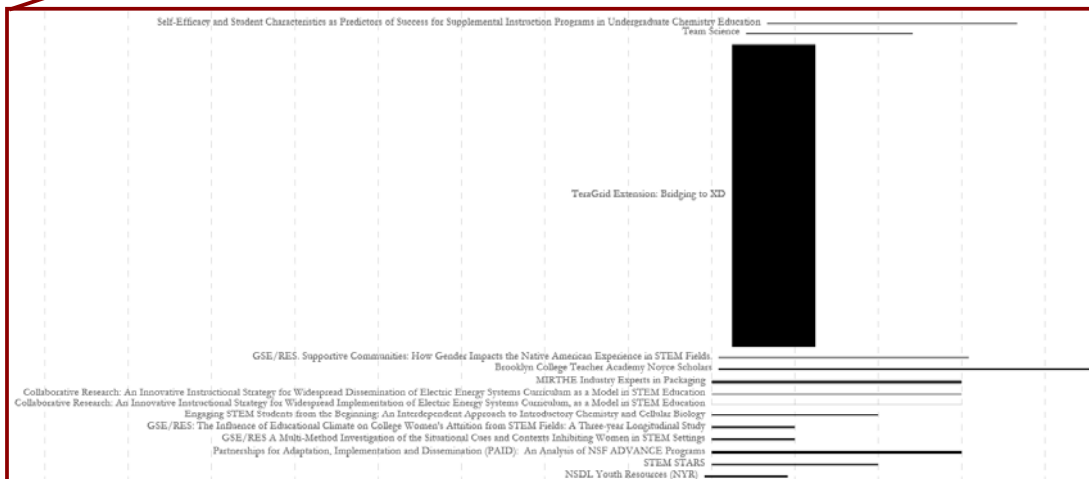
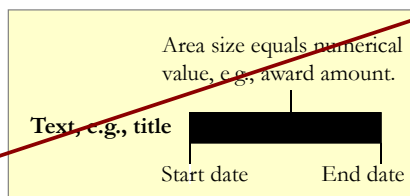
Input Parameters:

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

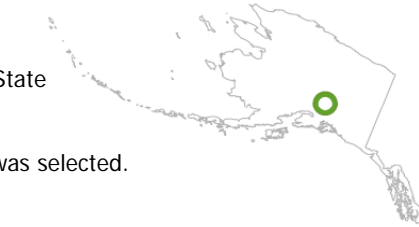
Projection: Albers Equal-Area Conic

Map: US States

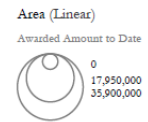
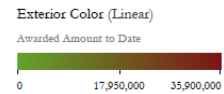
Author Name: K. Borner

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.

.....

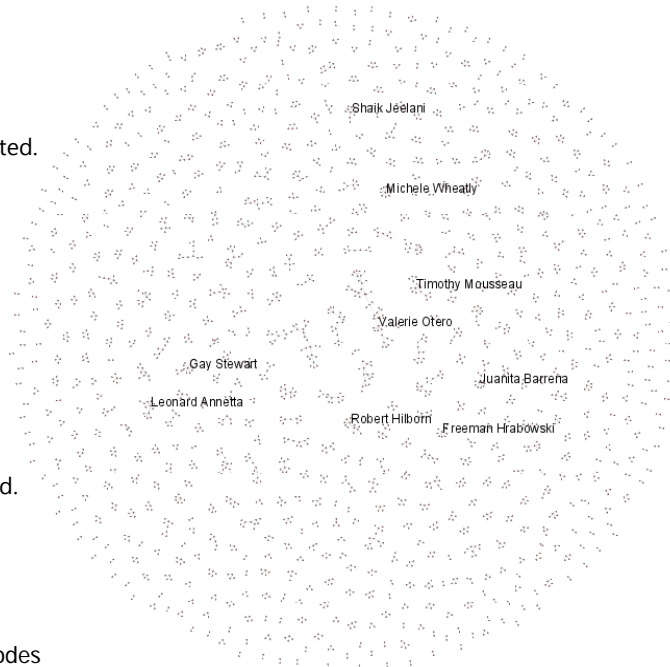
Weak Component Clustering was selected.

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

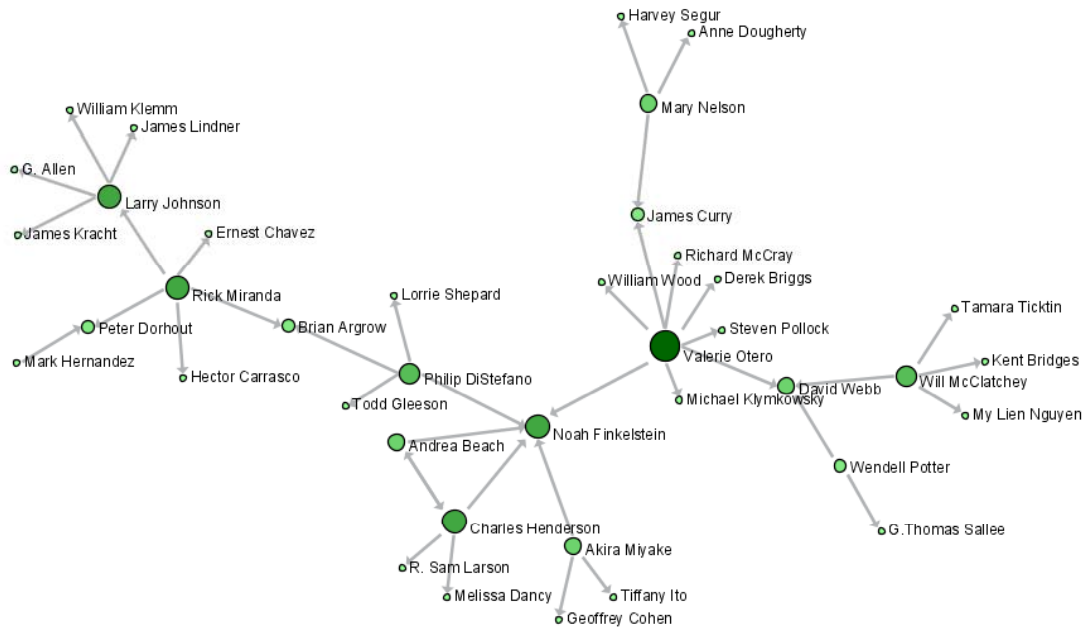


Co-PI Networks – Giant Component

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 selected.

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 components.

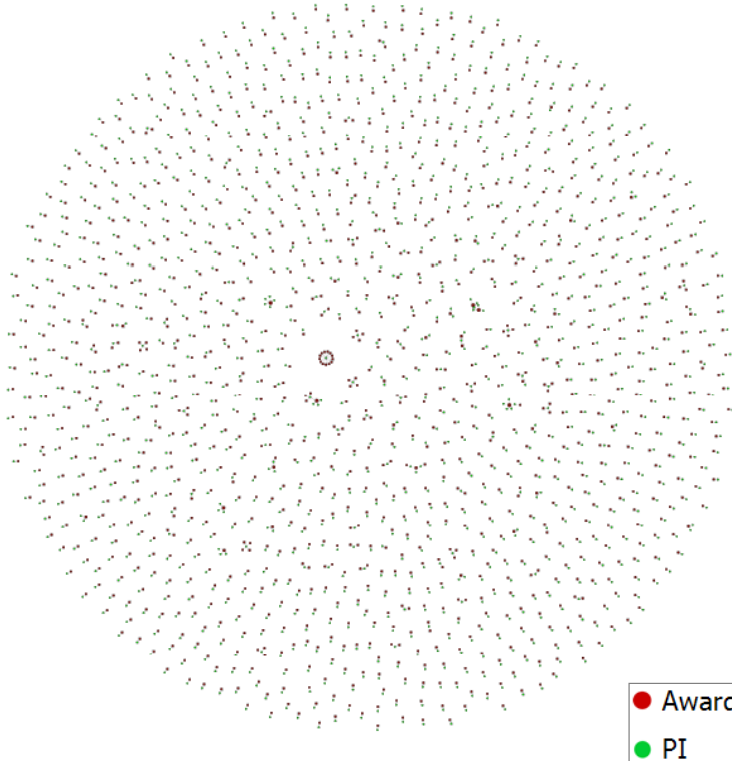
The largest connected component consists of 1144 nodes.

Density (disregarding weights): 0.0002

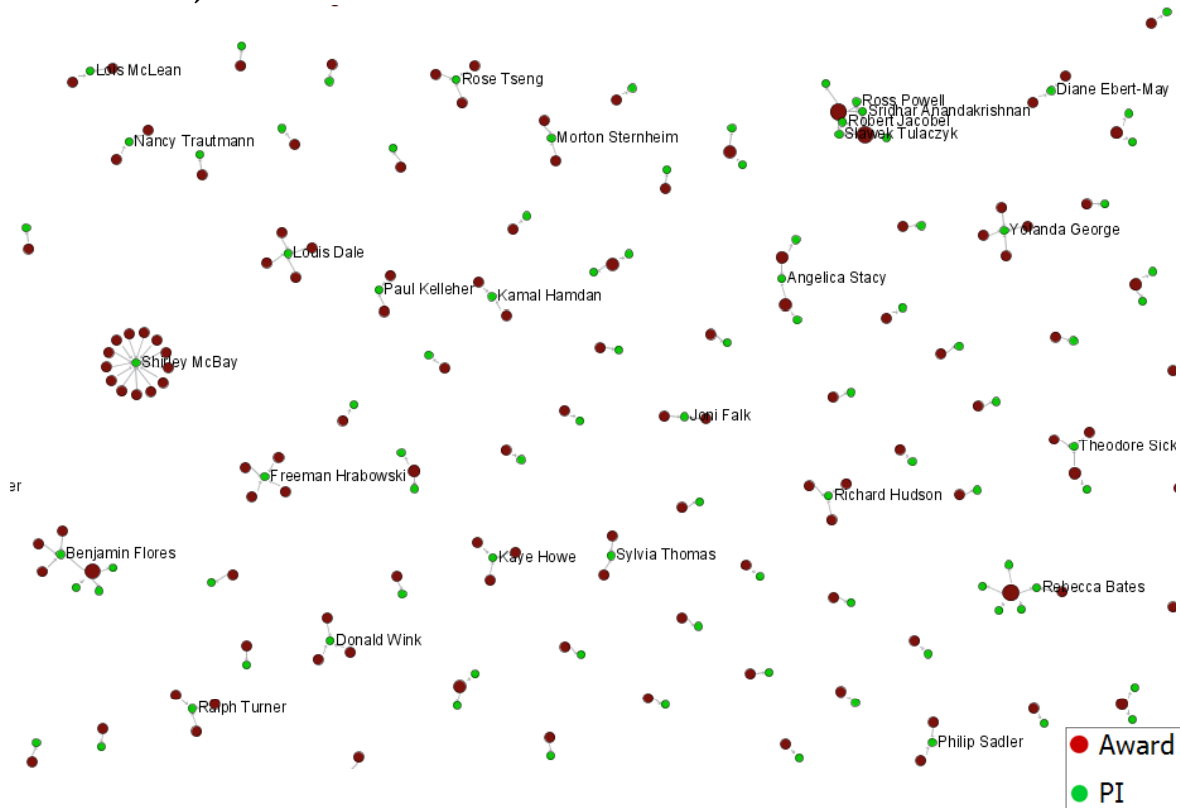
.....
Node Indegree was selected.

.....
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

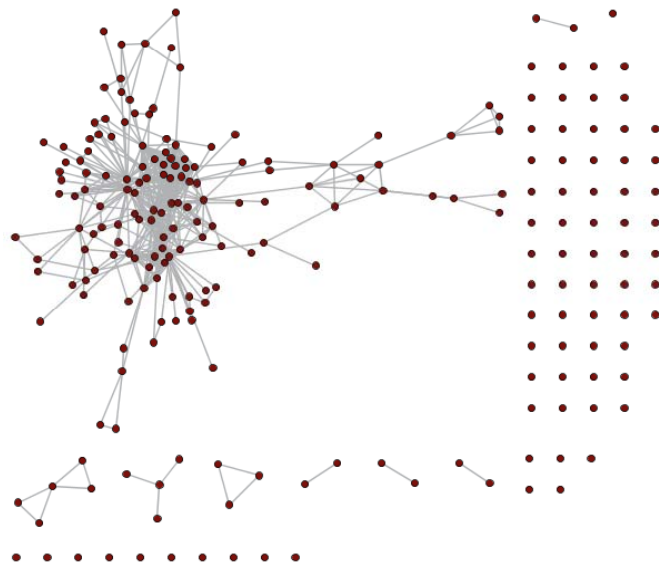
No self loops were discovered.

Average degree: 6.9185

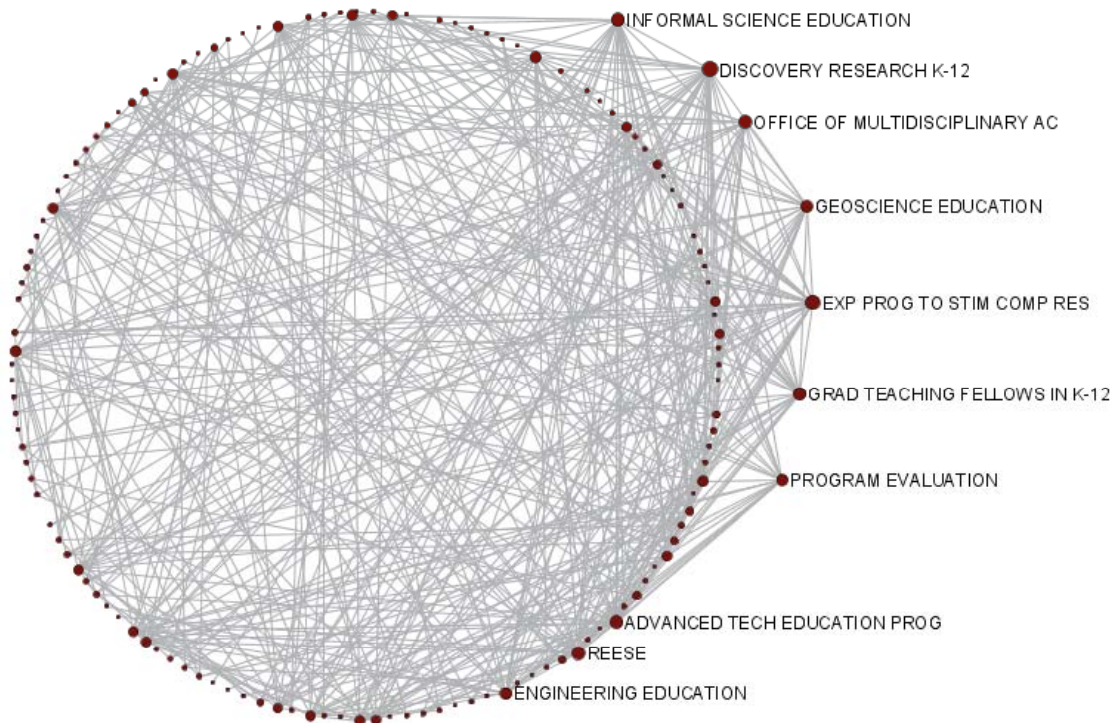
Density (disregarding weights): 0.0516

.....

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

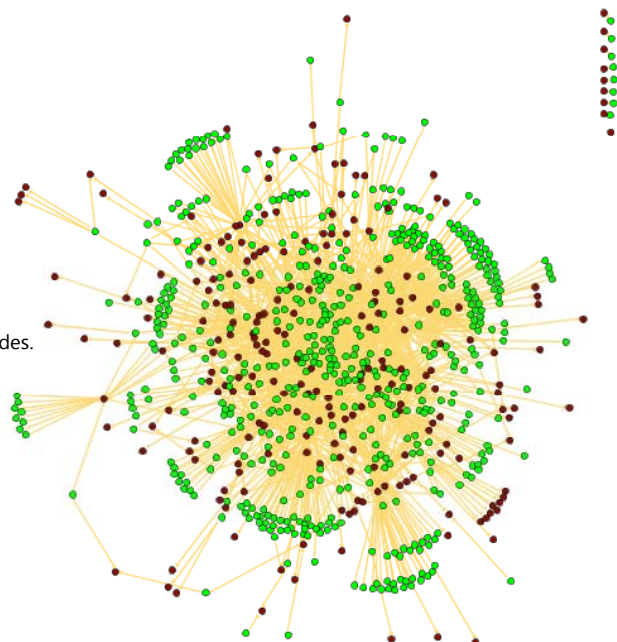
The largest connected component consists of 777 nodes.

Density (disregarding weights): 0.0025

.....
 Node Indegree was selected.

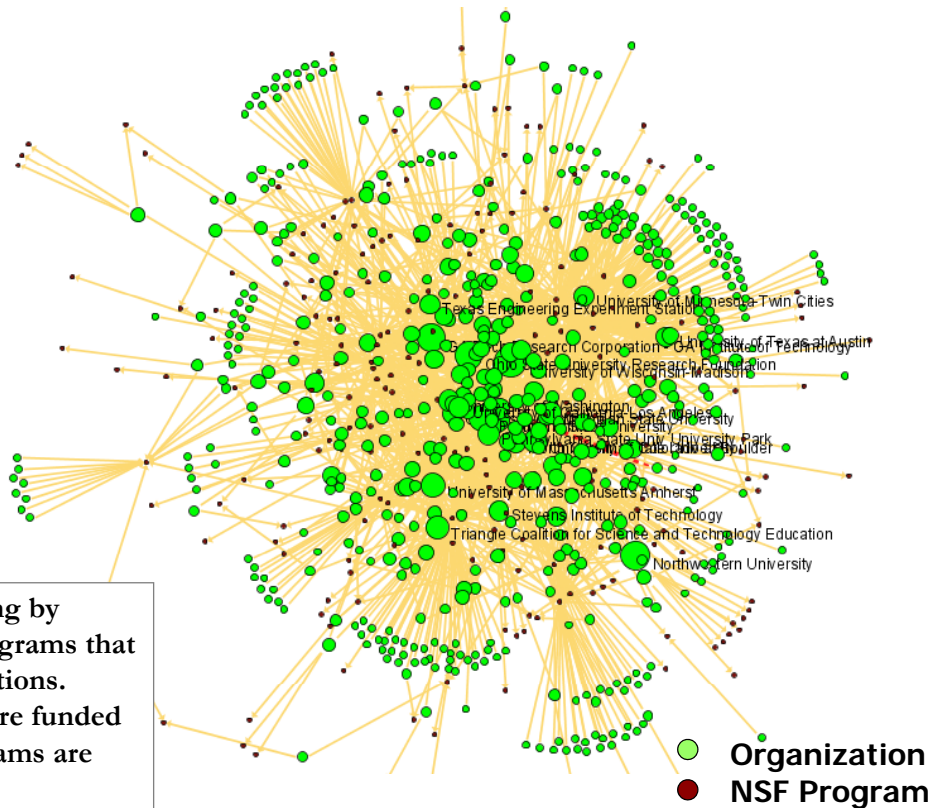
.....
 Node Outdegree was selected.

.....
 GUESS

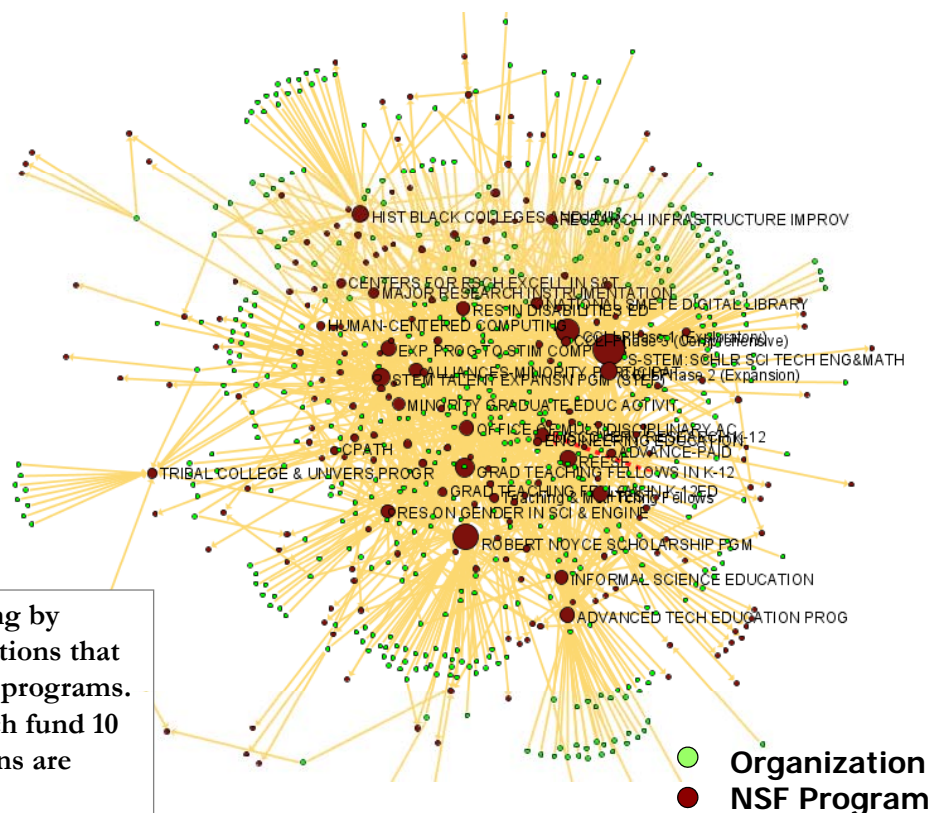


● Organization
 ● NSF Program

What Organizations are funded by what NSF Programs?




What NSF Programs fund how many Organizations?



Science of Science Cyberinfrastructure

P O R T A L


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 Macroscopic 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, 20 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/>.



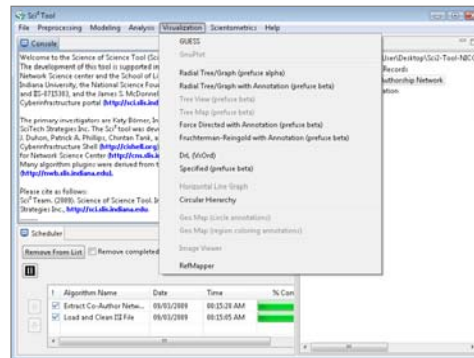
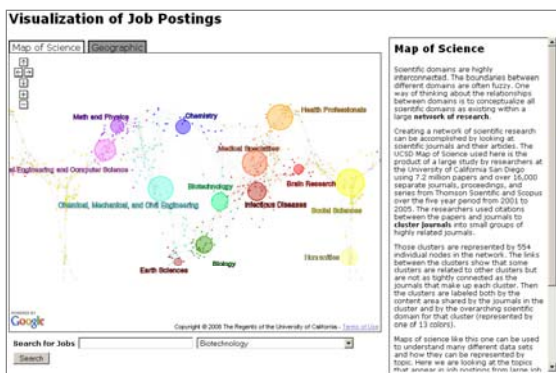
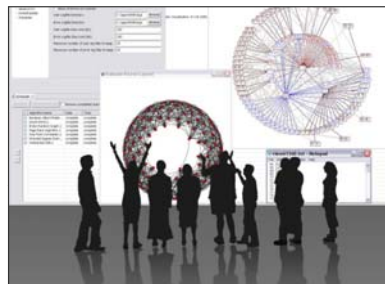
Cyberinfrastructures
The Scientometrics filling of the **Network Workbench (NWB) Tool** provides a unique distributed, shared resources environment for large-scale network analysis, modeling, and visualization. Thomson Scientific/ISI, Scopus and Google Scholar data, EndNote and BibTeX files, or NSF awards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

<http://sci.slis.indiana.edu>



Overview

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





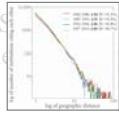
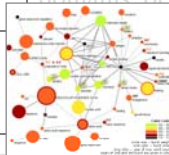



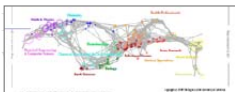



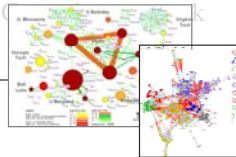
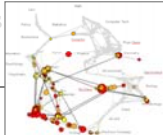
Type of Analysis vs. Level of Analysis

	<i>Micro/Individual</i> (1-100 records)	<i>Meso/Local</i> (101–10,000 records)	<i>Macro/Global</i> (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

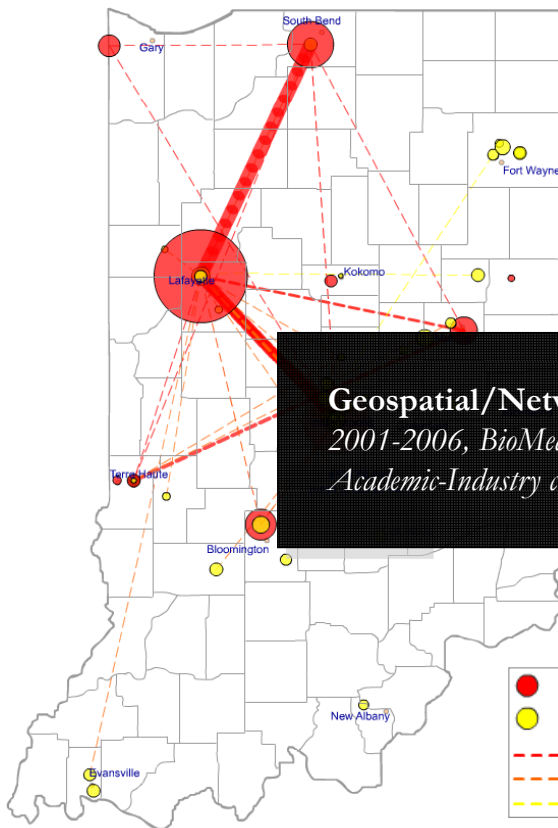
43



Type of Analysis vs. Level of Analysis

	<i>Micro/Individual</i> (1-100 records)	<i>Meso/Local</i> (101–10,000 records)	<i>Macro/Global</i> (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 

44

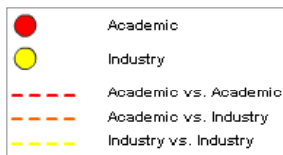


Mapping Indiana's Intellectual Space

Geospatial/Network Analysis

2001-2006, BioMed, IN Scope

Academic-Industry collaborations and knowledge diffusion

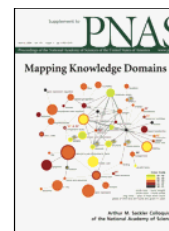
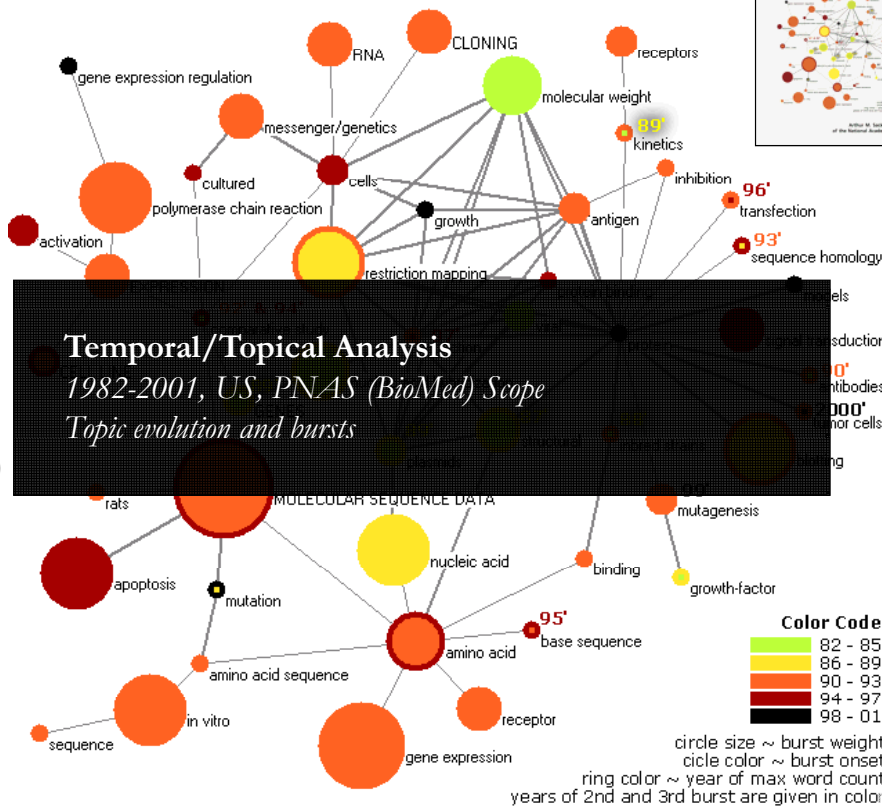


45

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.



46

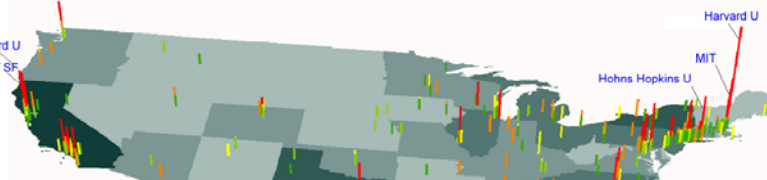
Spatio-Temporal Information Production and Consumption of Major U.S. Research Institutions

Börner, Katy, Penumarthu, 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

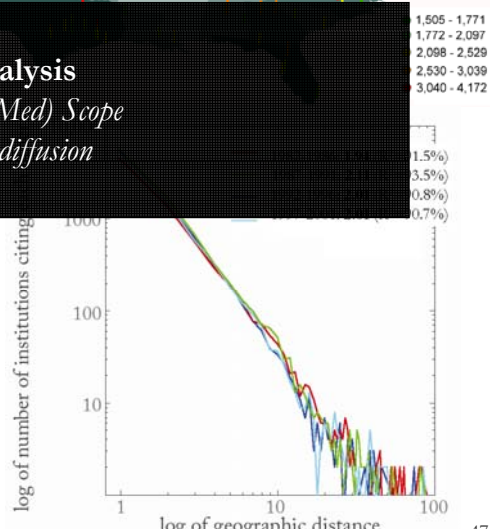


Research questions:

1. Does space still matter in the Internet age?
2. Does one still have to study and work at institutions in order to produce high quality data quality research?
3. Does the Internet change patterns, i.e., more produced at geographically distant research institutions?



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. Dubon, Elisha F. Hardy, Katy Börner, Indiana University, USA

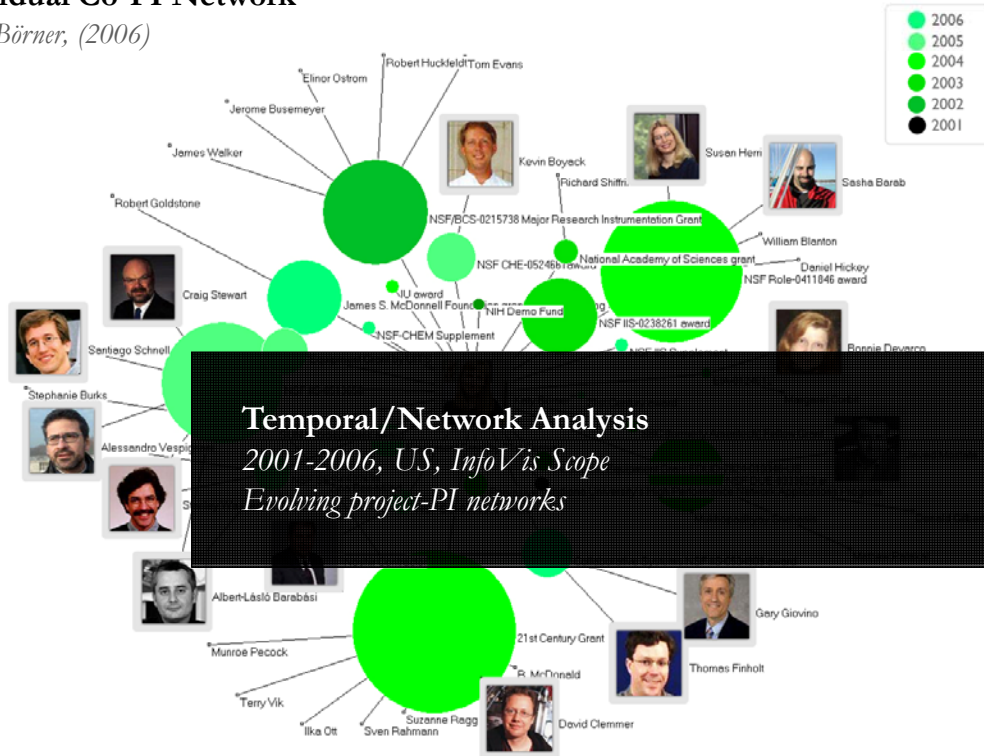


Geospatial Analysis World, Chinese Academy of Science Collaboration and knowledge diffusion via co-author networks

This map highlights the research 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.

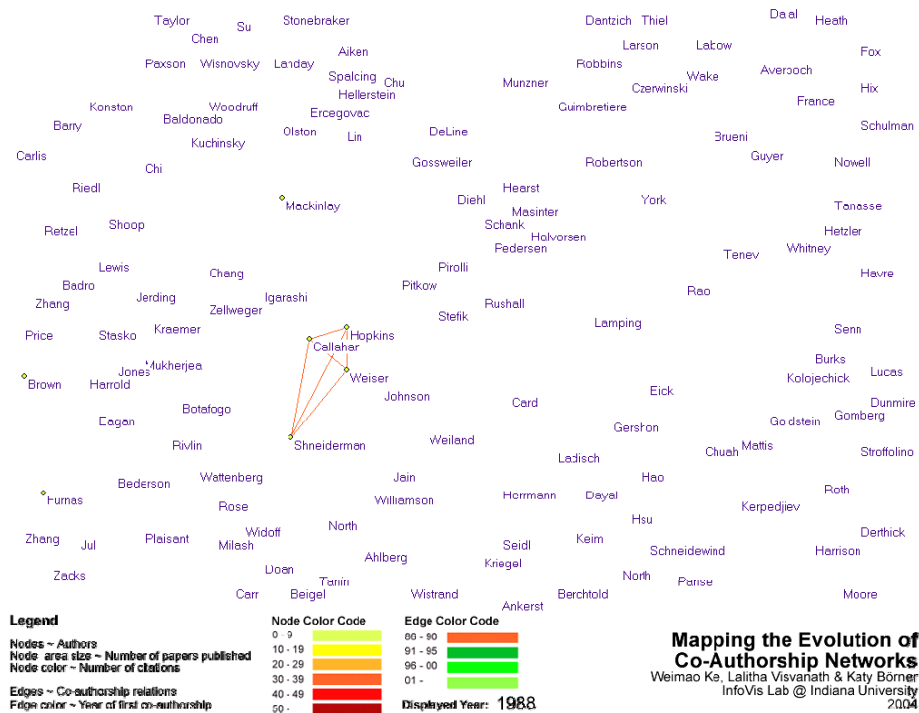
Individual Co-PI Network

Ke & Börner, (2006)



Mapping the Evolution of Co-Authorship Networks

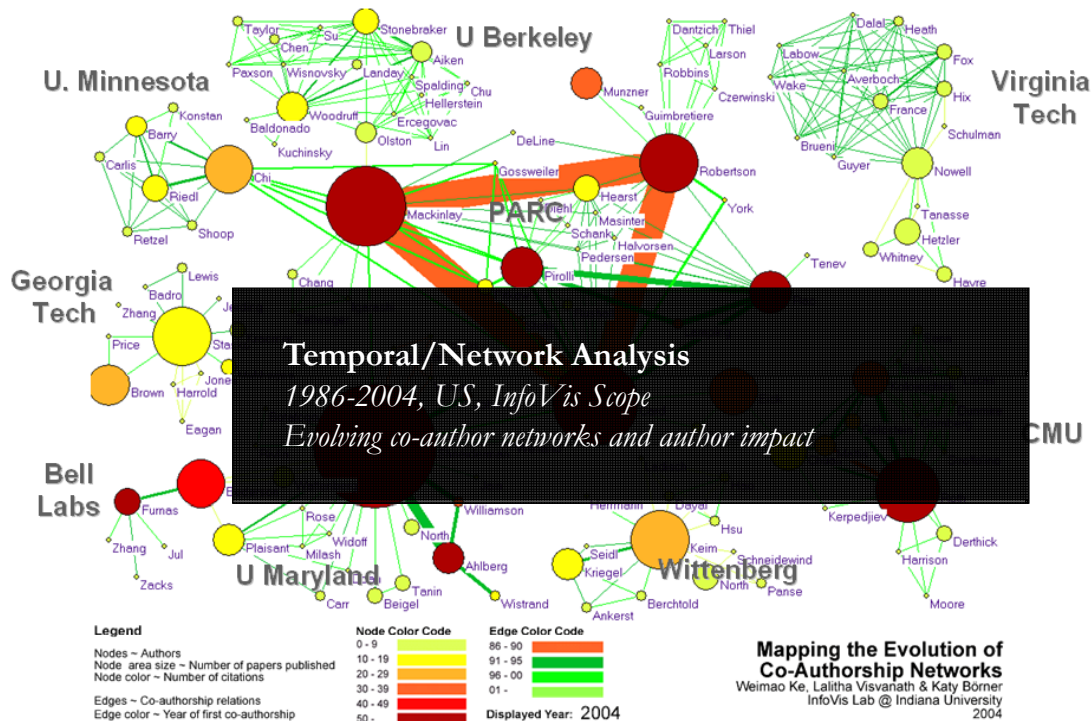
Ke, Visvanath & Börner, (2004) Won 1st price at the IEEE InfoVis Contest.



Mapping the Evolution of Co-Authorship Networks
 Weimao Ke, Lalitha Visvanath & Katy Börner
 InfoVis Lab @ Indiana University
 2004

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

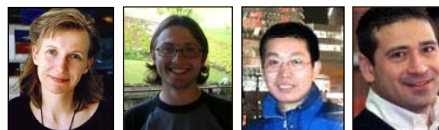
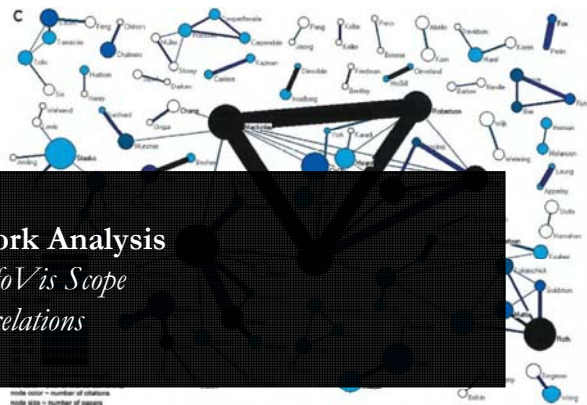
Börner, Dall'Asta, Ke & Vespignani (2005) Complexity, 10(4):58-67.

Research question:

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

Contributions:

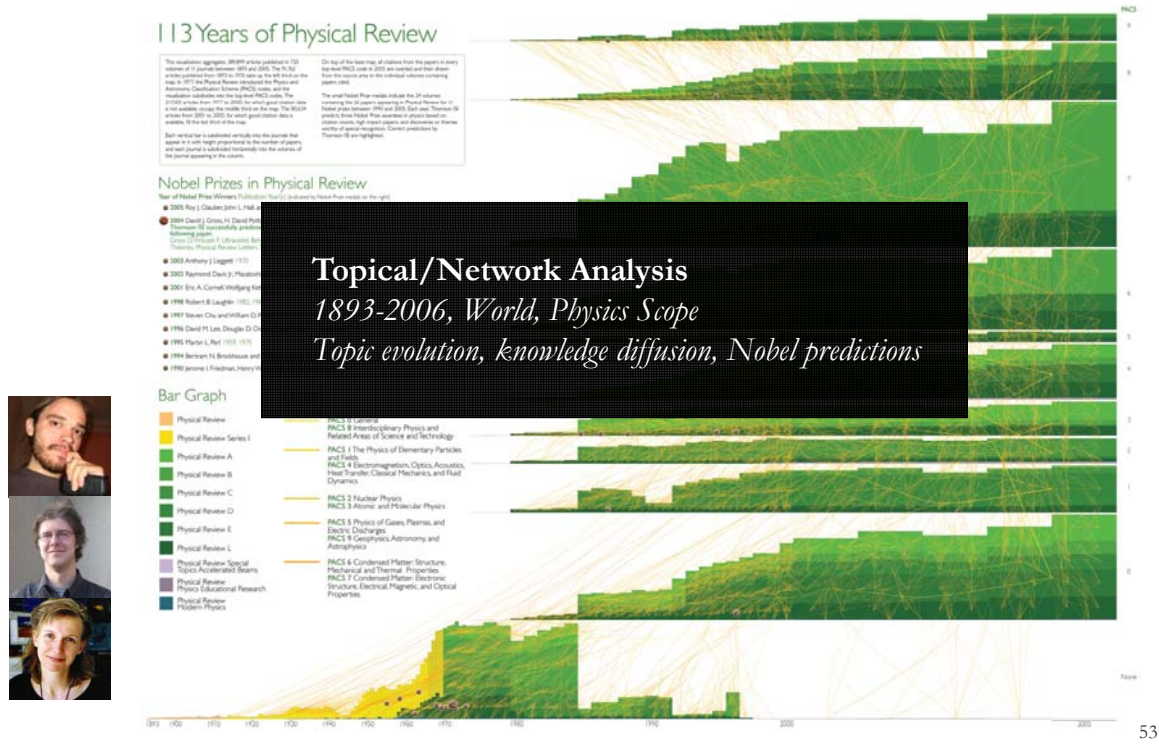
- New approach to allocate citational credit.
- Novel weighted graph
- Visualization of the co-author network
- Centrality measure impact.
- Global statistical analysis of paper production and citations in correlation with co-authorship team size over time.
- Local, author-centered entropy measure.



113 Years of Physical Review

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

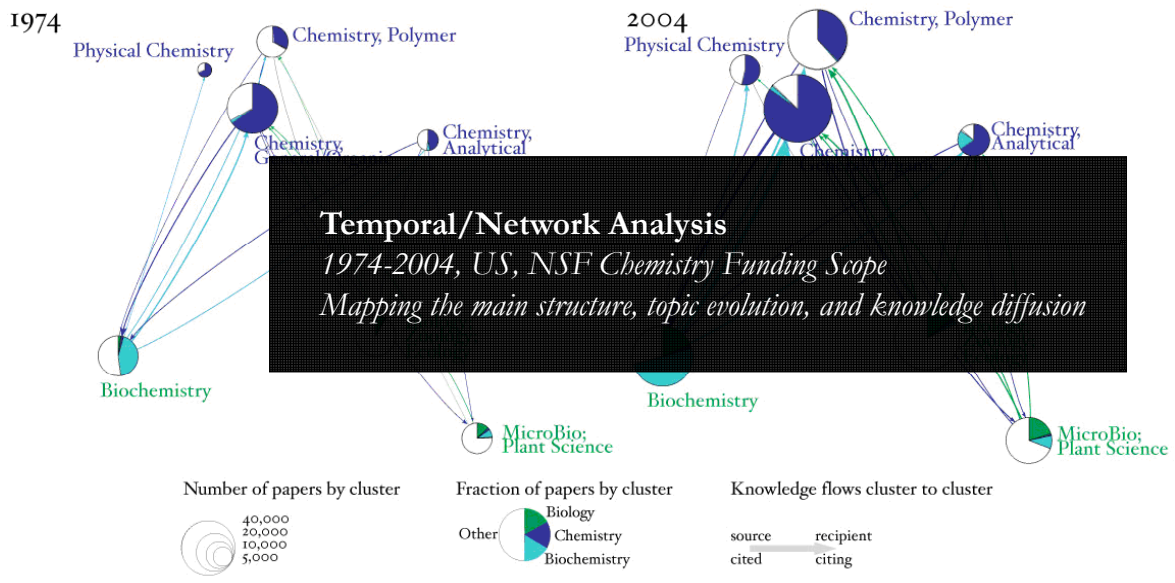
Bruce W. Herr II and Russell Dubon (Data Mining & Visualization), Elisha F. Hardy (Graphic Design), Shashikant Penumarthy (Data Preparation) and Katy Börner (Concept)



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

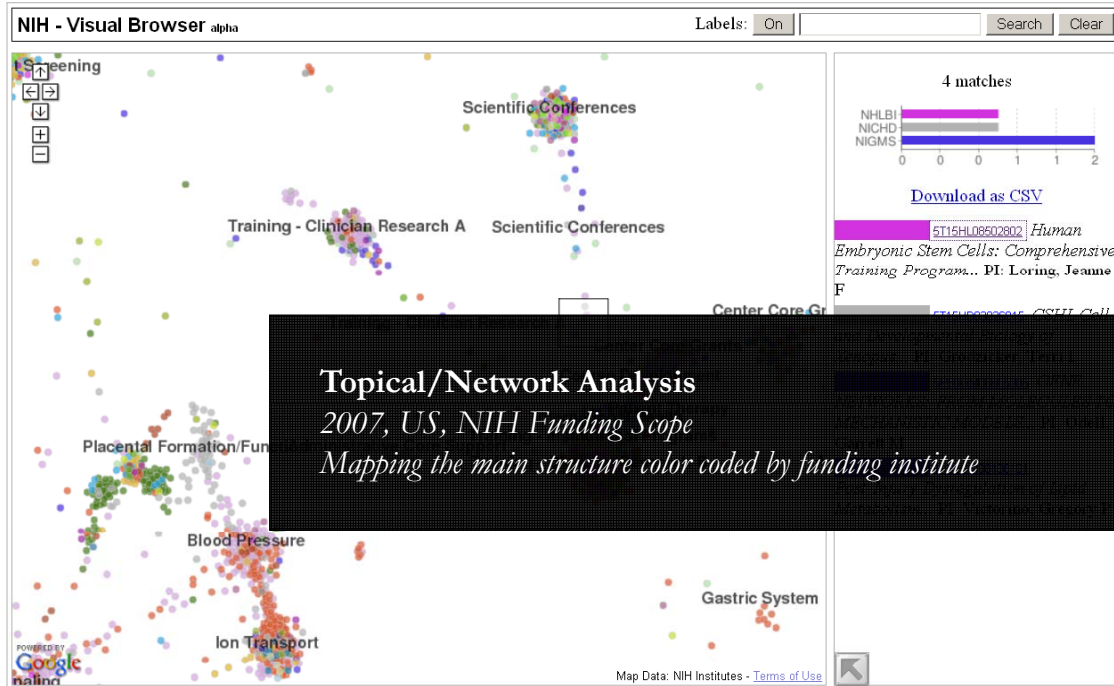
Kevin W. Boyack, Katy Börner, & Richard Klavans (2007)

Chemistry - Biology Interface



Interactive Maps of Science – NIH Funding

Google maps with charts and tables

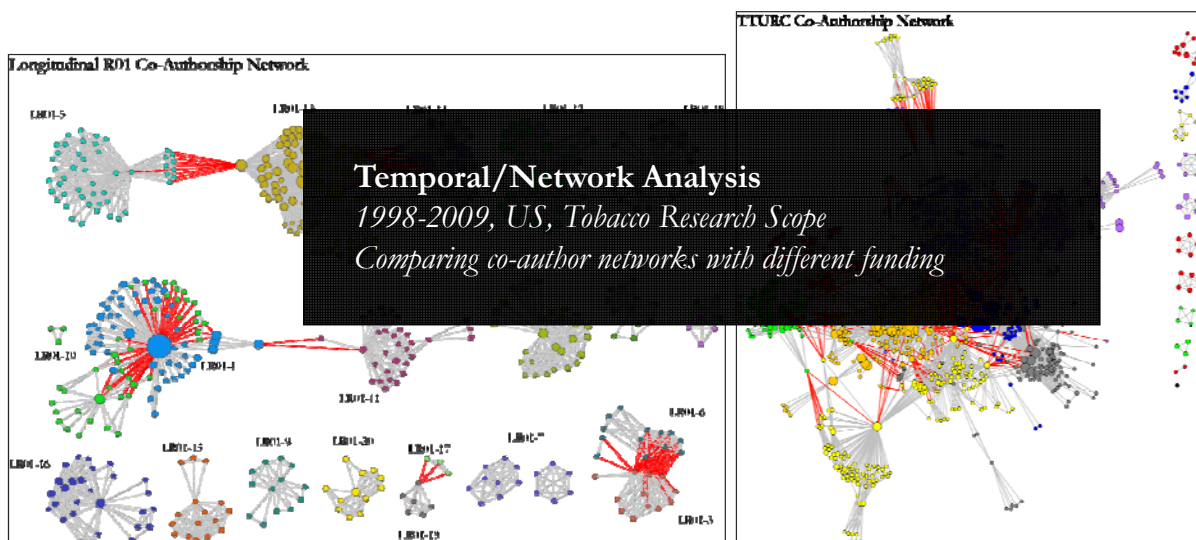
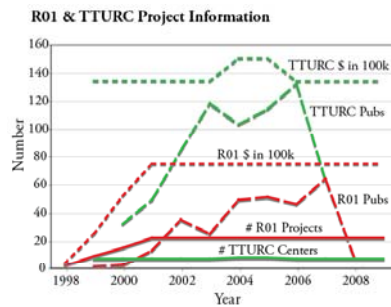


<http://scimaps.org/maps/nih/2007>

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

Dubon & Börner, forthcoming.

(a) Overview

Date and input directory

Basic counts

Overlay of all matched journal references from all PDF files on 554 scientific disciplines (nodes) in UCSD Map of Science

Circle size denotes # references

Listing of all references grouped by 13 science areas

(b) Visual Index

For each PDF file: Basic counts and thumbnail science map

Max 18 per page

Top-n most similar PDF files identified based on journal name co-occurrences. The similarity of each PDF file to itself is 1

Overlay of matched journal references from all above listed PDF files on UCSD Map of Science and grouping by 13 science areas

Topical/Network Analysis
 2009, US, NSF Funding
 Grouping interdisciplinary funding proposals for review



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

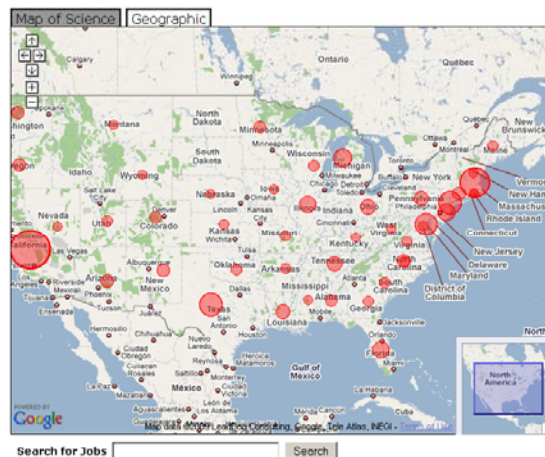
Angela Zoss, Michael Conover

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.

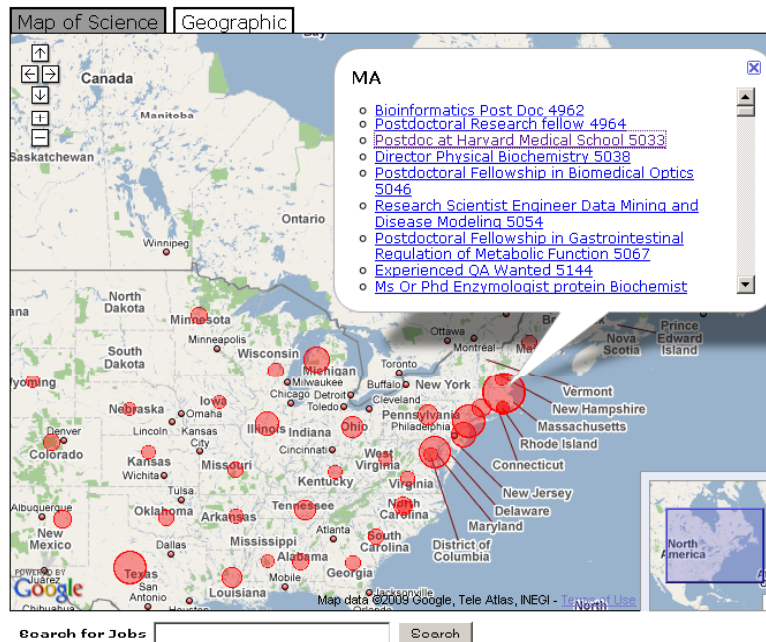
Visualization of Job Postings



Geographic Visualization

Here we have a more traditional view of job postings – a geographic overlay. Featured here are the job postings that list both a city and state in the United States. Feel free to search, zoom, pan, and click on job descriptions.

Visualization of Job Postings



Postdoc at Harvard Medical School
[Link to Post](#)

Harvard Medical School, Massachusetts General Hospital, Gastrointestinal Unit, One Post-doctoral Position Available

We are now looking for an additional post-doctoral fellow who wants to study in the area of cellular and molecular mechanisms during the development of inflammatory bowel disease (ulcerative colitis and Crohn's disease). The successful candidate will be involved in studies on physiological functions of key molecules (including Toll-like receptors and tumor necrosis factor receptors) in colonic epithelial cells/microbial interactions.

Candidates need to have MD, PhD, MD/PhD, or equivalent degree(s) with research training in the field of immunology, pathology, microbiology, biochemistry, and/or molecular biology. Actual starting date will be July or August 2009. An initial appointment will be for 2 years, but the term can be extended depending on the research accomplishment. Salary will be competitive and commensurate with experience.

Massachusetts General Hospital is the third oldest general hospital in the United States and the oldest and largest hospital in New England. The



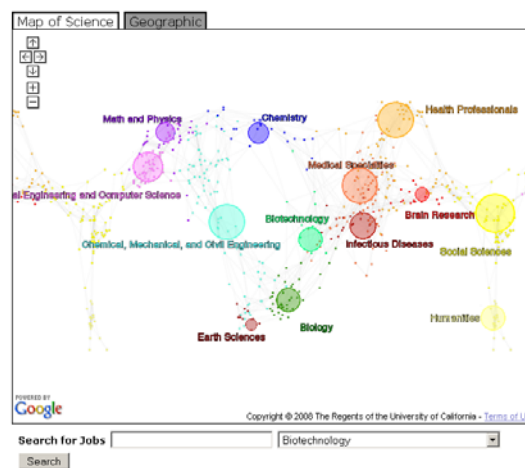
Mapping S&T Job Market Data in Real Time – SciMap

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.

Visualization of Job Postings



Map of Science

Scientific domains are highly interconnected. The boundaries between different domains are often fuzzy. One way of thinking about the relationships between domains is to conceptualize all scientific domains as existing within a large network of research.

Creating a network of scientific research can be accomplished by looking at scientific journals and their articles. 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. The researchers used citations between the papers and journals to cluster journals into small groups of highly related journals.

Those clusters are represented by 554 individual nodes in the network. The links between the clusters show that some clusters are related to other clusters but are not as tightly connected as the journals that make up each cluster. Then the clusters are labeled both by the content area shared by the journals in the cluster and by the overarching scientific domain for that cluster (represented by one of 13 colors).

Maps of science like this one can be used to understand many different data sets and how they can be represented by topic. Here we are looking at the topics that appear in link postings from larva job

Map of Science | Geographic

Brain Research

Gene Therapy: 79 posts

Infectious Diseases

Geospatial and Topical/Network Analysis
 2008-2009, English speaking, Job RSS feeds
 Support interactive search for job postings in geo and topic space.

POWERED BY Google

Copyright © 2008 The Regents of the University of California - [Terms of Use](#)

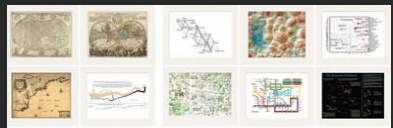
Search for Jobs Any Category Search

Mapping Science Exhibit – 10 Iterations in 10 years

<http://scimaps.org>



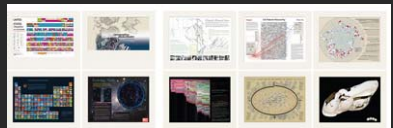
The Power of Maps (2005)



Science Maps for Economic Decision Makers (2008)



The Power of Reference Systems (2006)



- 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)

The Power of Forecasts (2007)



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.





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>

This is the only mockup in this slide show.

Everything else is available today.



