

# Science of Science Research and Tools

## Tutorial #01 of 12

**Dr. Katy Börner**

Cyberinfrastructure for Network Science Center, Director  
Information Visualization Laboratory, Director  
School of Library and Information Science  
Indiana University, Bloomington, IN  
<http://info.slis.indiana.edu/~katy>

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

Invited by Robin M. Wagner, Ph.D., M.S.  
Chief Reporting Branch, Division of Information Services  
Office of Research Information Systems, Office of Extramural Research  
Office of the Director, National Institutes of Health

*Suite 4090, 6705 Rockledge Drive, Bethesda, MD 20892  
10a-noon, July 6, 2010*



## 12 Tutorials in 12 Days at NIH—Overview

1. Science of Science Research **1st Week**
2. Information Visualization
3. CShell Powered Tools: Network Workbench and Science of Science Tool
4. Temporal Analysis—Burst Detection **2nd Week**
5. Geospatial Analysis and Mapping
6. Topical Analysis & Mapping
7. Tree Analysis and Visualization **3rd Week**
8. Network Analysis
9. Large Network Analysis
10. Using the Scholarly Database at IU **4th Week**
11. VIVO National Researcher Networking
12. Future Developments

## Week 1: The Basics

3



### 12 Tutorials in 12 Days at NIH—Overview

#### **[#01] Science of Science Research**

- Brief History (Atlas timeline)
- Micro, Meso, Macro Studies
- Workflow Design
- Sample Studies / Mapping Science Exhibit
- Validation
- Promising Research Directions

#### **Recommended Reading**

- Börner, Katy, Chen, Chaomei, and Boyack, Kevin. (2003). Visualizing Knowledge Domains. In Blaise Cronin (Ed.), *ARIST*, Medford, NJ: Information Today, Inc./American Society for Information Science and Technology, Volume 37, Chapter 5, pp. 179-255. <http://ivl.slis.indiana.edu/km/pub/2003-borner-arist.pdf>
- Shiffrin, Richard M. and Börner, Katy (Eds.) (2004). Mapping Knowledge Domains. *Proceedings of the National Academy of Sciences of the United States of America*, 101(Suppl\_1). [http://www.pnas.org/content/vol101/suppl\\_1](http://www.pnas.org/content/vol101/suppl_1)
- Börner, Katy (2010) Atlas of Science. MIT Press. <http://scimaps.org/atlas>

4



## 12 Tutorials in 12 Days at NIH—Overview

### **[#02] Information Visualization**

- Introduction
- Designing Effective Visualizations
- Visualization Layers
- Visual Languages
- Promising Research Directions

### **Recommended Reading**

- Information Visualization class at Indiana University, <http://ella.slis.indiana.edu/~katy/S637-S10>
- Edward R. Tufte (1990) [Envisioning Information](#). Graphics Press.
- Edward R. Tufte (1992) [The Visual Display of Quantitative Information](#). Graphics Press.
- Edward R. Tufte (1997) [Visual Explanations: Images and Quantities, Evidence and Narrative](#). Graphics Press.
- Colin Ware (1999) [Information Visualization: Perception for Design](#), Morgan Kaufmann Publishers.

5



## 12 Tutorials in 12 Days at NIH—Overview

### **[#03] CIShell Powered Tools: Network Workbench and Science of Science Tool**

- Using a Million Minds to Build Custom Tools
- Open Service Gateway Initiative (OSGi)
- Cyberinfrastructure Shell (CIShell)
- **Network Workbench (NWB) Tool**
- **Science of Science (Sci2) Tool**
- Adding Plugins to CIShell Powered Tools
- Promising Research Directions

### **Recommended Reading**

- Herr, Bruce W., Huang, Weixia, Penumarthy, Shashikant, Börner, Katy . (2007) Designing Highly Flexible and Usable Cyberinfrastructures for Convergence. In William S. Bainbridge and Mihail C. Roco (Eds.) *Progress in Convergence – Technologies for Human Wellbeing. Annals of the New York Academy of Sciences*, Boston, MA, volume 1093, pp. 161-179. <http://cishell.org/papers/06-cishell.pdf>
- Cyberinfrastructure Shell home page, <http://cishell.org>.
- Network Workbench (NWB) Tool home page, <http://nwb.slis.indiana.edu>
- Science of Science (Sci2) Tool home page, <http://sci.slis.indiana.edu/sci2>

6

## General Considerations

7



### User Needs Driven

This tutorial aims to address the needs of the *Reporting Branch* at NIH as well as other tutorial participants. Your input and feedback is welcome and appreciated. The structure and format of the tutorials might change in response to your suggestions.

**Please complete “Questionnaire #1”.**

**Input**





## Hands-on

This tutorial teaches (peer reviewed) theory as well as practically useful knowledge. There will be many demonstrations and interactive in-class exercises.

**Demo**

**Exercise**

I will use NIH (internal) data and real-world relevant workflows as time permits.

We might modify/implement new plugins to make the existing tools relevant for NIH research and praxis.

I will invite feedback regularly to optimize tools, workflows, documentation, and tutorials.

Implementing and testing new software, getting familiar with NIH internal data, designing and optimizing custom workflows, developing documentation and tutorials takes time. I have exactly 12 work days.

The tools, workflows, documentation, tutorial will not be perfect.

**You are highly valued beta testers.**

**Please do report software bugs and problems to me and I will try to resolve them.**

9



## Documentation

**Input**

All 12 tutorials will be audio recorded.

The 12 audio files together with the (revised) slides will be made available online.

Your questionnaire responses, comments, and suggestions will NOT be made available to others.

**Distribute “General Questionnaire” after each of the 12 tutorials.**

**Responses will be collected at the end of each day.**

10

## [#01] Science of Science Research

- Brief History (Atlas timeline)
- Micro, Meso, Macro Studies
- Workflow Design
- Sample Studies / Mapping Science Exhibit
- Validation
- Promising Research Directions



Thanks go to Kevin Boyack for making his slides available.

### Early Maps of the World

VERSUS

### Early Maps of Science

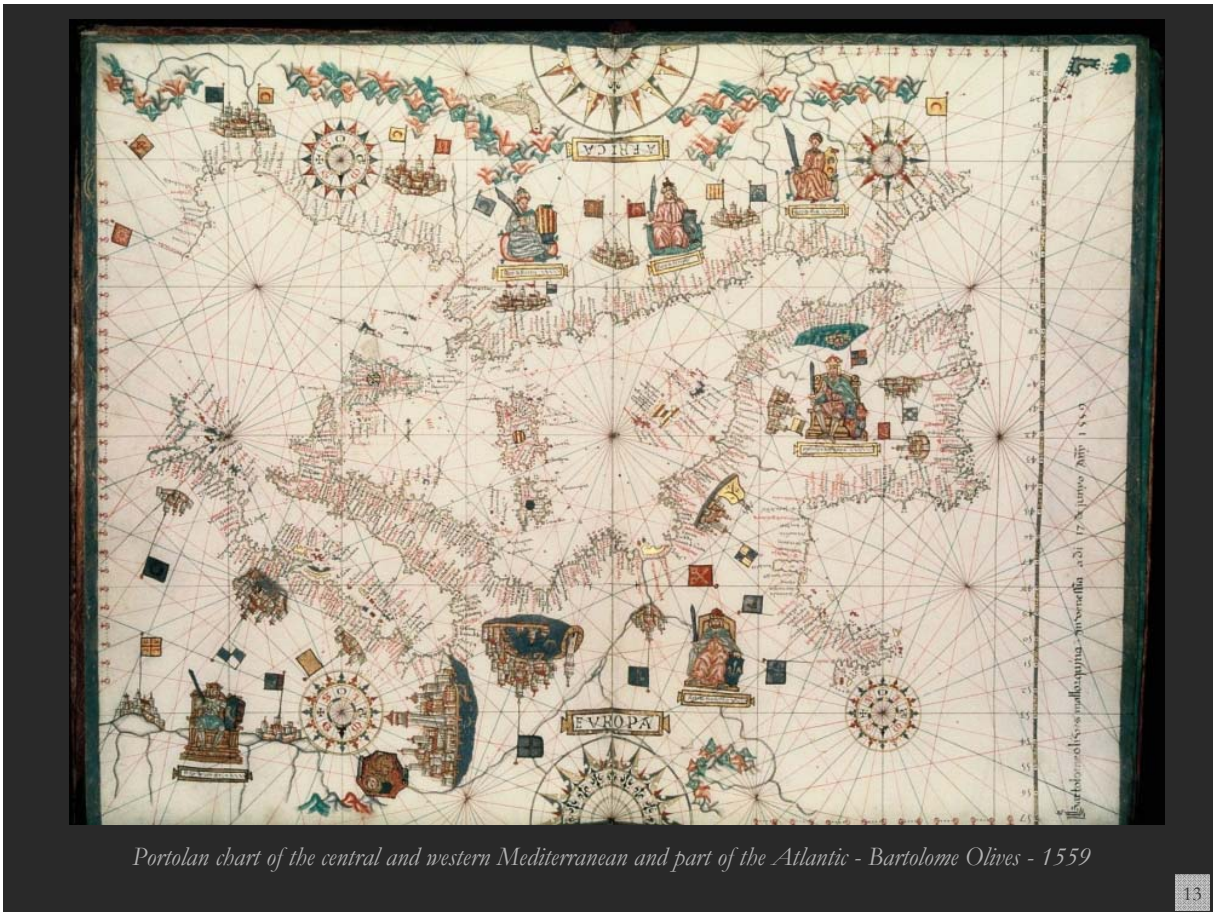


3D  
Physically-based  
Accuracy is measurable  
Trade-offs have more to do with granularity  
2-D projections are very accurate at local levels  
Centuries of experience  
**Geo-maps can be a template for other data**

n-D  
Abstract space  
Accuracy is difficult  
Trade-offs indirectly affect accuracy  
2-D projections neglect a great deal of data  
Decades of experience  
**Science maps can be a template for other data**

*Kevin W. Boyack, UCGIS Summer Meeting, June, 2009*









## Milestones of Mapping Science



1934

2007

Timeline on GigaPan: <http://gigapan.org/viewGigapan.php?id=25917>

Börner, Katy. (2010). *Atlas of Science: Visualizing What We Know*.

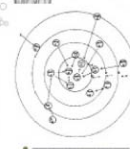
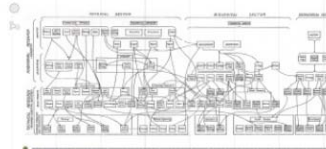
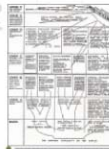
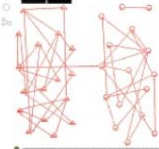
MIT Press.

# 1930-1960

Factor Analysis (FA)  
Thurstone

Cluster Analysis  
Troy

Information Theory  
Shannon



Sociometry  
Moreno

Lady Bountiful  
Lundberg & Steele

World Encyclopedia  
Wells

Map of Science  
Bernal

Target Sociogram  
Northrup

Natural S&T Chart  
Ellingham

1930

1935

1940

1945

1950

1955

Scope

- Individual
- Local
- Global
- Mixture

Layout

- Manual
- Algorithmic

Type

- Temporal
- Semantic
- Geographic
- Network
- Mixture

Exhibit Map

Traité de Documentation: Le Livre sur le Livre - Théorie et Pratique  
Oster  
Who Shall Survive?  
Moreno

World Brains  
Wells  
The Social Function of Science  
Bernal

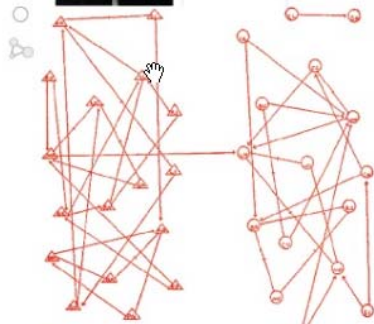
Documentation  
Braudel  
De Mechanisering van het Wereldbeeld  
Dijkshoorn

Putting Science in Its F  
Geographers of Science  
Livingstone

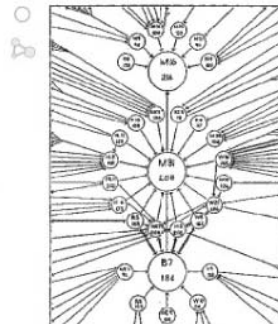
1930

1955

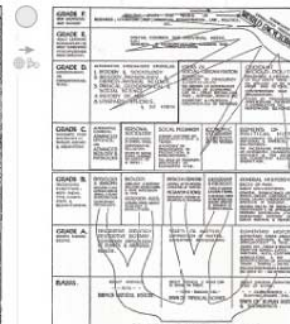
17



Sociometry  
Moreno



Lady Bountiful  
Lundberg & Steele



World Encyclopedia  
Wells

1930

1935

1940

Scope

- Individual
- Local
- Global
- Mixture

Layout

- Manual
- Algorithmic

Type

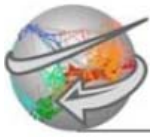
- Temporal
- Semantic
- Geographic
- Network
- Mixture

Exhibit Map

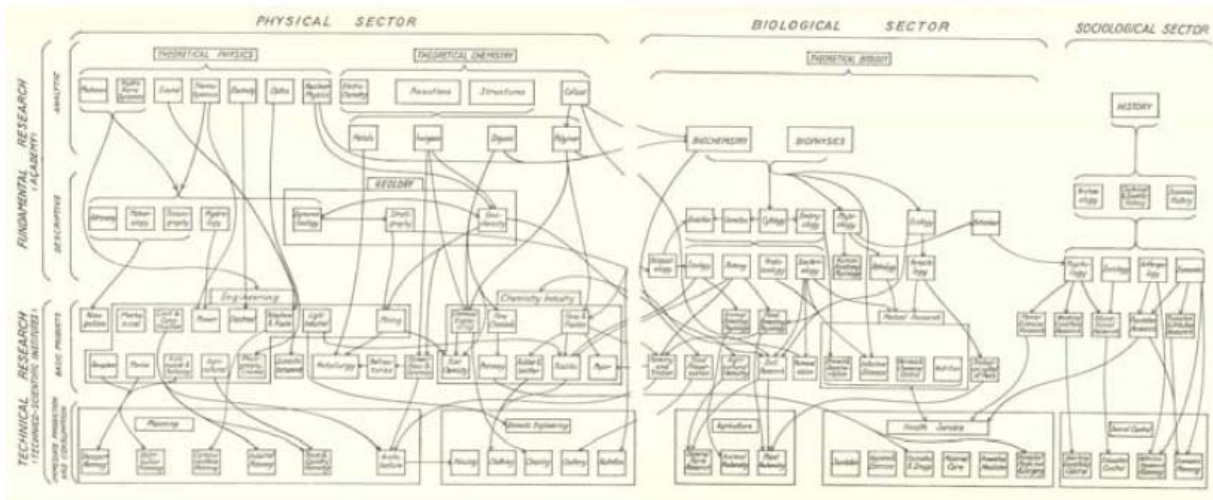
Zoom into one map and legend

18





## Bernal, 1939



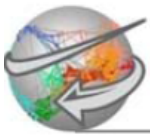
John D. Bernal was a world renowned physicist, a historian of science, and a sociologist of science. He is considered to have produced one of the first 'maps' of science.

Bernal, J.D. (1939). *The Social Function of Science*. London: Routledge & Kegan Ltd.

The Social Function  
of Science

J. D. BERNAL  
1939

19



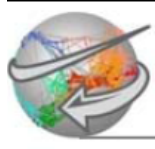
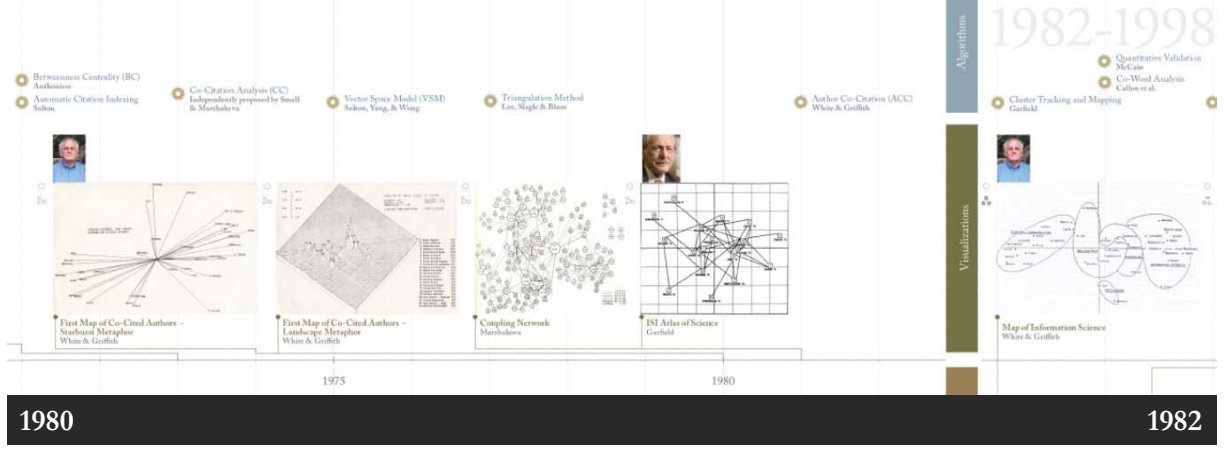
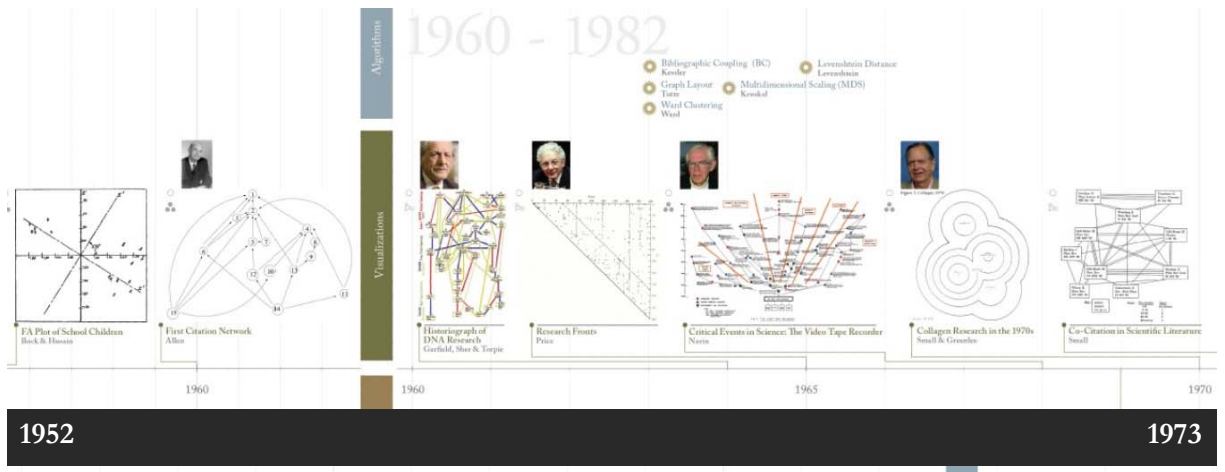
## Ellingham, 1948

Ellingham's "Relations Between the Branches of Natural Science and Technology" with an overlay of "Abstracts or Groups of Abstracts Covering A Very Wide Field" (1948)



H. J. T. Ellingham (1948). "Divisions of Natural Science and Technology," Royal Society Scientific Information Conference, 21 June to 2 July 1948, London: Burlington House,

20



# Garfield, 1964

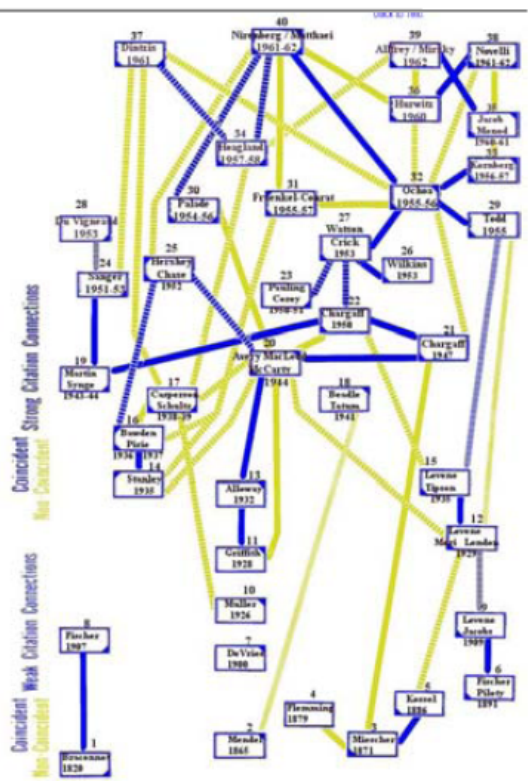
## Historiograph of DNA Development

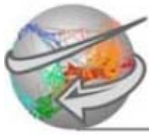


**Eugene Garfield,** recent photo. Creator of the ISI Web of Science citation database.

<http://www.garfield.library.upenn.edu/>

Garfield, Sher, & Torpie (1964). "The Use of Citation Data in Writing the History of Science." Air Force Office of Scientific Research under contract F49(638)-1256.





# Small, 1973

Using co-citation to create domain maps

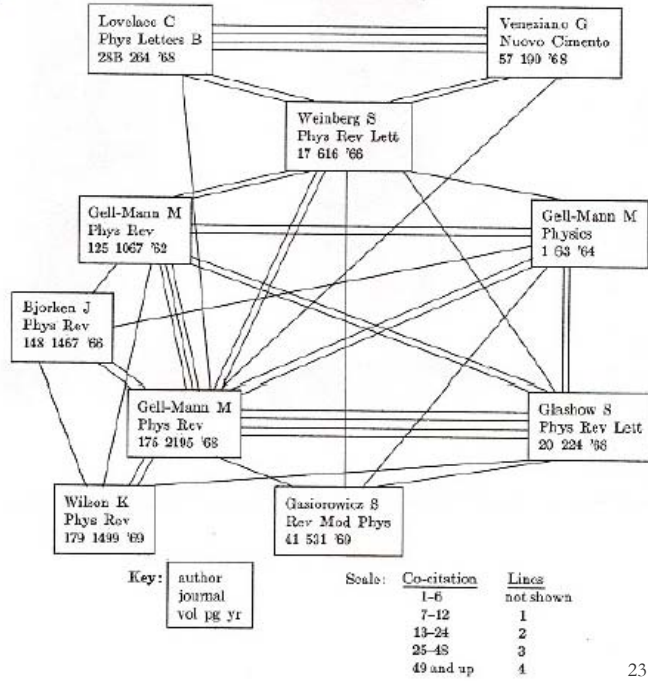


**Henry Small.**  
Head of research  
at ISI, now  
Thomson Reuters  
Scientific.

Small, H. (1973). "Co-citation in the scientific literature: A new measure of the relationship between two documents." *JASIS*, 24, 265-269.

Marshakova, I.V. (1973). "A system of document connections based on references." *Scientific and Technical Information Serial of VINITI*, 6, 3-8.

FIGURE 1  
Co-citation Network for Frequently Cited Papers in Particle Physics  
(Data from the 1971 *SCJ*)



1987

- Spring Graph Layout: Eades
- Self-Organizing Map (SOM): Kohonen
- Journal Co-Citation Mapping: McCaa
- Tree Map Layout: Johnson & Shneiderman
- Frischerman Ringed Graph Layout: Frischerman & Ringold
- Kamada-Kawai Graph Layout: Kamada & Kawai
- Identifying Scientific Frontiers: Garfield & Small
- Pathfinder Network Scaling (PFNet): Schreyer
- Latent Semantic Analysis (LSA): Deerwester et al.

NoteCards: Hilde, Moran & Trigg at Xerox PARC  
Specialties in Sociology: Eades  
SOM of Newsgroup Postings: Kohonen  
Butterfly Citation Browser: Mackinley, Cool & Row at Xerox Research  
SOMET Map: H. Chen et al.  
Concept Map: Novak

1997

1998

- Automatic Citation Indexing: Giles, Bolacker & Lawrence
- PageRank: Brin & Page
- Hubs & Authorities: Kleinberg
- Fixing Complementary Literatures: Swanson & Southamer
- Combined Linkage: Small
- Longitudinal Coupling: Small

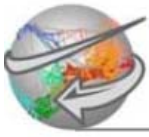
Dewey Tree Map: Shneiderman  
MDS Map of Information Science: White & McCaa  
Map of the Market: Waterberg  
Collaborative StarWalker: C. Chen et al.  
1996 Map of Science: Small

1999

1997

1999



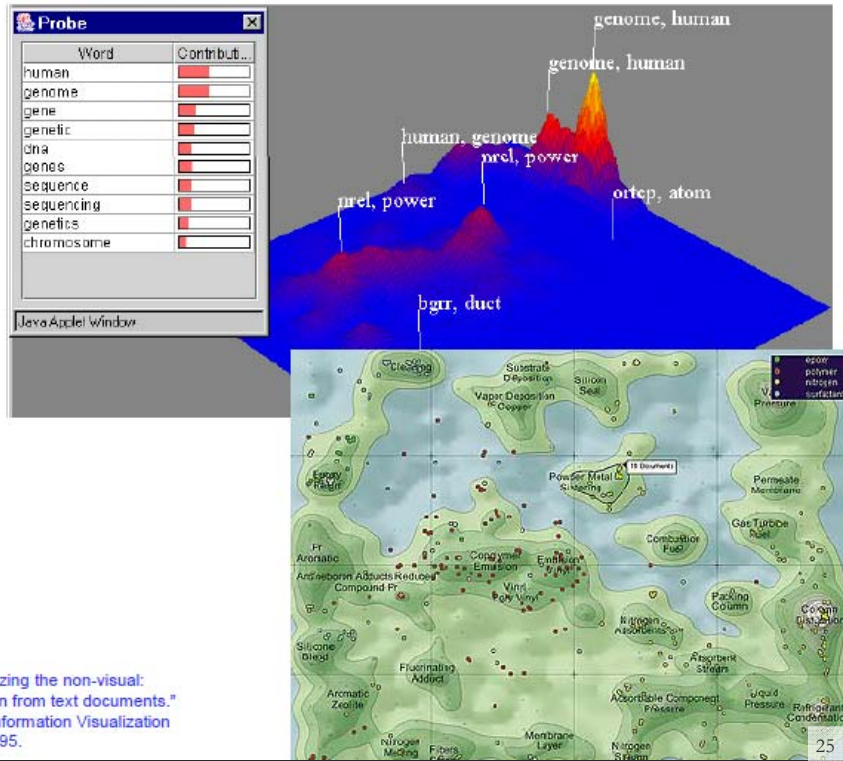


# SPIRE, Themescape, 1995

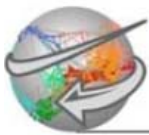
Pacific Northwest Labs introduces a mapping tool based on text

2.5-D representation of intensity of "themes" using topography

Later spinoff of same technology used in patent analysis products



Wise, Thomas, Pennock, et al. (1995). "Visualizing the non-visual: Spatial analysis and interaction with information from text documents." Paper presented at the IEEE Symposium on Information Visualization '95, Atlanta, Georgia, USA., October 30-31, 1995.

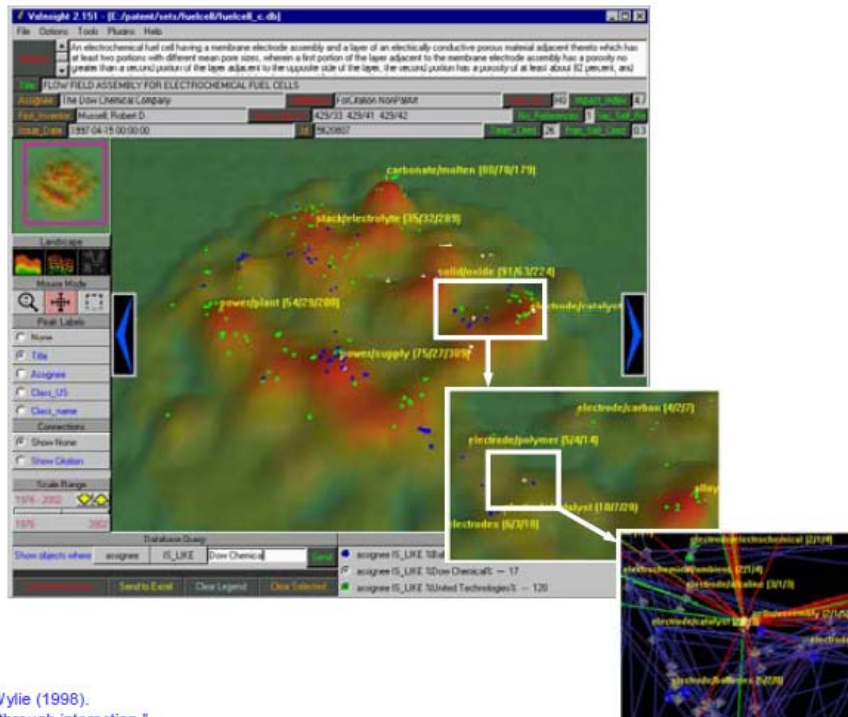


# VxInsight, 1998

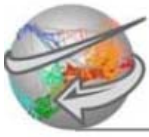
Sandia National Labs introduces an interactive browsing tool for exploring "maps"

Primarily for exploring citation-based maps, but ultimately used in science studies and genomics

Zoom, pan, query, etc. capabilities



Davidson, Hendrickson, Johnson, Meyers & Wylie (1998). "Knowledge mining with VxInsight: Discovery through interaction." *Journal of Intelligent Information Systems*, 11(3), 259-285.



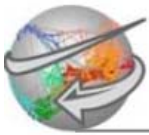
# Author Co-citation Map, White, 1998

120 highly-cited authors in Information Science

Layout using multi-dimensional scaling



White & McCain (1998) "Visualizing a discipline: An author co-citation analysis of information science 1972-1995." *JASIS* 49(4), 327-356.



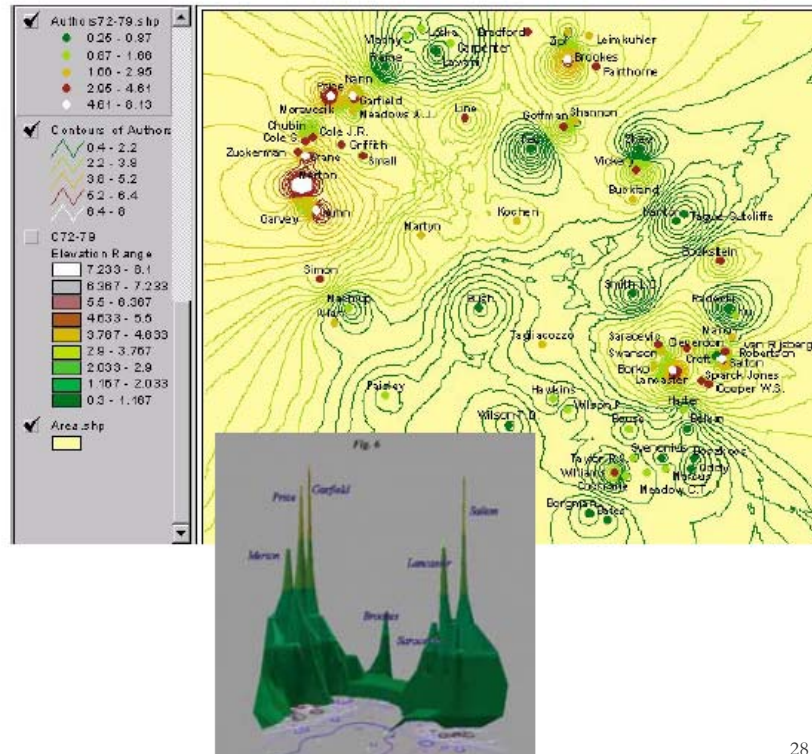
# Old, 2001

Utilizing spatial information systems for non-spatial data analysis

Authors in Information Science

Topography added

3D representation also





2000

Algorithms

Visualizations

2000

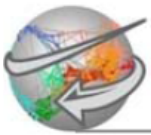
1999

2000

2001

2001

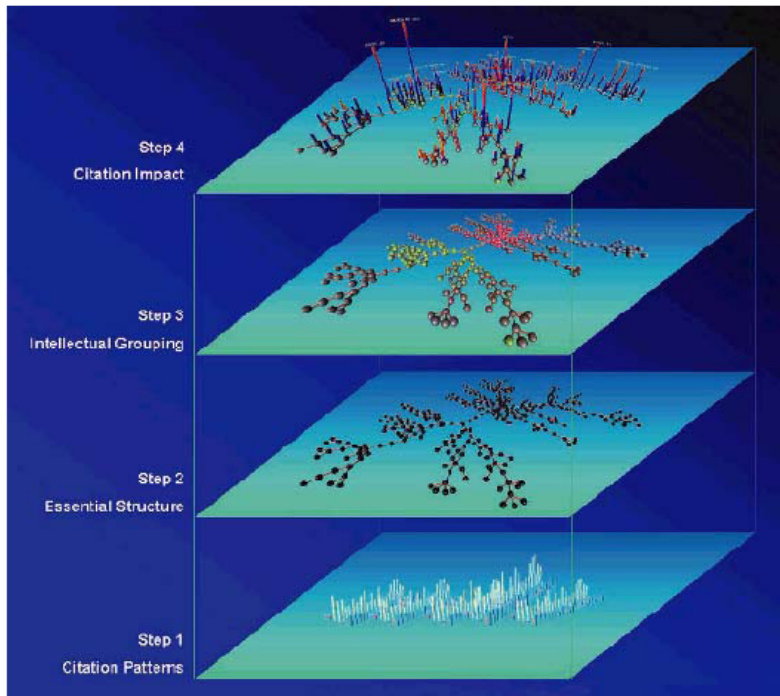
29

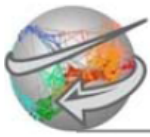


## Chen, 2001

Software: CiteSpace I

Four-step procedure for visualizing intellectual structures using co-cited documents

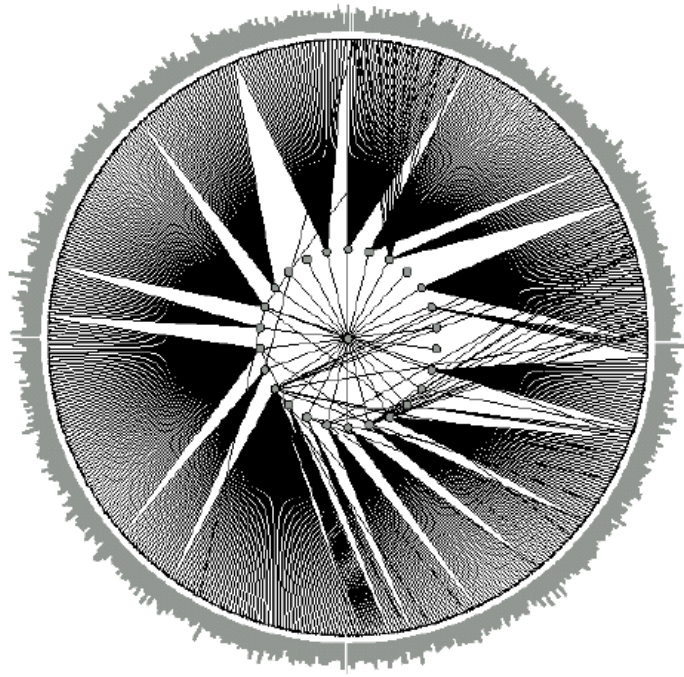




# Newman, 2001

Physicist bringing new tools to the problem

2-generation co-authorship graph of Mark himself (center node)



Newman, M.E.J. (2001). "Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality." *Physical Review E*, 64, paper # 016132.

2002		2003				
Algorithms						
Visualizations	<p>Cartographic Map of Conference Abstracts Shogan</p>	<p>Author Link &amp; Concept Link Lin, Whit, &amp; Boryshkowsky</p>	<p>Scholarly Genealogies Lerner</p>	<p>Geography of Science Berry</p>	<p>Linking Papers and Funding Boyak &amp; Borner</p>	
Tools	<p>SOCNET Haimson &amp; Dejn</p>	<p>JUNG O'Mahain et al.</p>	<p>Information Visualization Cyberinfrastructure Baumgartner et al.</p>	<p>Social Social Network Image Scientist Bender-JeMall &amp; McFarland</p>	<p>RefViz Overviews of Literature Search Results Thomas IS ResearchSoft</p>	<p>Tulp Software Auber</p>
Books		<ul style="list-style-type: none"> <li>Theories of Communication Networks</li> <li>Moag &amp; Contractor</li> <li>Six Degrees</li> <li>Watts</li> </ul>				
	2002				2003	



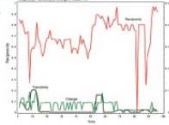


# 2005

- New-Evaluation Spring Embedders  
Kobourov & Wampler
- Flow Map Layout  
Phan et al.
- Acknowledgment Indexing  
Cormier et al.



Critical Paths and Trajectories of Individuals  
Bender & Mell



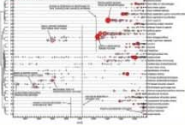
Longitudinal Social Network Metrics  
Moudry, McFarland & Bender-Del-Mell



Journal Flow Map of Data by  
Author  
Cormier et al.



Treemap View of 2004 Usenet Remotives  
Smith & Fisher



CrossMap of Anthrax Research  
Morris

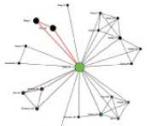


Backbone of Science  
Berach, Klauer & Birner

Algorithms

Visualizations

# 2005



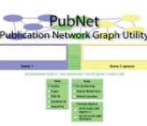
Perflow Visualization API  
Hess, Cuel & Landon



GUESS Graph Exploration System  
Adar



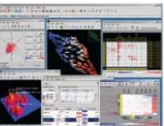
Author-Name-Disambiguation  
Author-ity  
Farruk et al.



PubNet  
Douglas et al.



CINI Researchers Link Viewer  
Jehou et al.



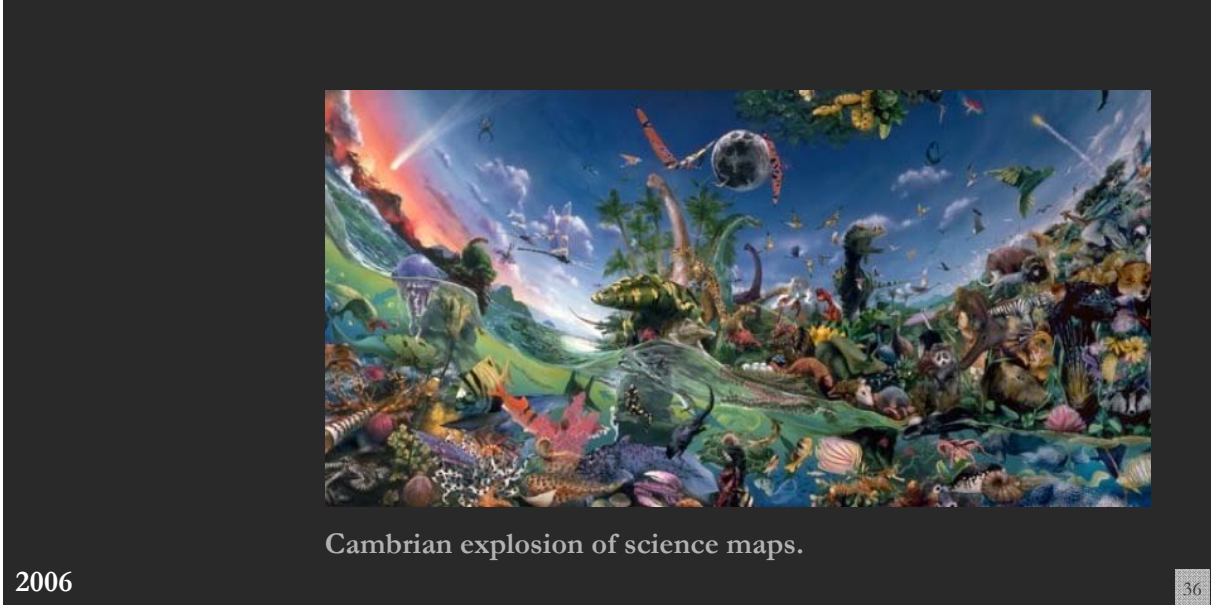
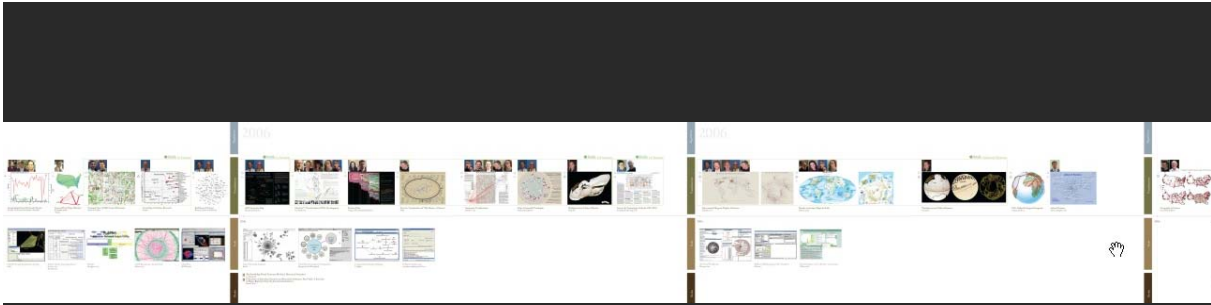
OmniVis  
Bui-Wadoux

- Models and Methods in Social Network Analysis  
Carrington & Wasserman (Eds.)
- The Hand of Science: Academic Writing and its Rewards  
Cramer
- Measurement and Statistics on Science and Technology: 1920  
to the Present (Knowledge Studies in the History of Science,  
Technology, and Medicine)  
Gottis

Tools

Books

# 2005



Cambrian explosion of science maps.

# 2006

New work is built on existing work. Each of the examples below cites a series of works that developed in a progressive fashion, as one born from the other:

- Garfield's original historiography of DNA research (1962); his long-term development of HistCite (first published in 2004); and his exhibit map (2006), which incorporates a re-rendering of the 1962 historiography and the application of HistCite.
- White et al.'s pioneering *Maps of Co-Cited Authors* (1982), *Map of Information Science* (1998), and the interactive AuthorLink (2002).
- Tobler's early works on the visualization of flow, his Flow Mapper tool (1987), and the tool's application in geospatial and network journal data (2005).
- Shneiderman's introduction of treemap layouts (1992, their utilization in the *Dewey Map* (1992), H. Chen's *ET Map* (1995), and later Wattenberg's *Map of the Market* (1989) and Smith et al.'s *Usenet visualizations* (2005).
- White and McCain's *Map of Information Science* (1998) and Old's GIS rendering of same (2001).
- C. Chen's *Collaborative Information Spaces* (1999), *Multi-Layer Science Maps* (2001), *Mapping Scientific Frontiers* (2004), and *Mapping the Universe* (2007); and his continuous development of CiteSpace for trend analysis (2004).
- Batty et al.'s work on the geography of science (2003 and 2006).
- Moody et al.'s studies of contour sociograms (2004) and longitudinal social network movies (2005).
- Boyack and Klavan's work toward a base map of science followed by the creation of a series of maps (2005–2007).

Over time, former tools are subsumed by new tools, software APIs, and libraries.

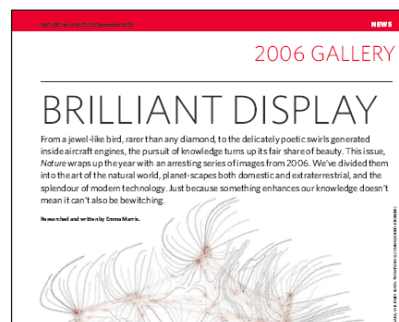
Examples include the *Information Visualization Cyberinfrastructure* (2003), Fekete's *The InfoVis Toolkit* (2004), and the *Network Workbench* (2006). Mashups also emerge, such as Herr et al.'s *Interactive Google Map of 2006 Society for Neuroscience Abstracts*.

## Computational Scientometrics References

Börner, Katy, Chen, Chaomei, and Boyack, Kevin. (2003). **Visualizing Knowledge Domains.** In Blaise Cronin (Ed.), *ARIST*, Medford, NJ: Information Today, Inc./American Society for Information Science and Technology, Volume 37, Chapter 5, pp. 179-255. <http://ivl.slis.indiana.edu/km/pub/2003-borner-arist.pdf>



Shiffrin, Richard M. and Börner, Katy (Eds.) (2004). **Mapping Knowledge Domains.** *Proceedings of the National Academy of Sciences of the United States of America*, 101(Suppl\_1). [http://www.pnas.org/content/vol101/suppl\\_1/](http://www.pnas.org/content/vol101/suppl_1/)



Börner, Katy, Sanyal, Soma and Vespignani, Alessandro (2007). **Network Science.** In Blaise Cronin (Ed.), *ARIST*, Information Today, Inc./American Society for Information Science and Technology, Medford, NJ, Volume 41, Chapter 12, pp. 537-607. <http://ivl.slis.indiana.edu/km/pub/2007-borner-arist.pdf>



Börner, Katy (2010) *Atlas of Science*. MIT Press. <http://scimaps.org/atlas>







## Type of Analysis vs. Scale of Level of Analysis

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

41



## There are many more questions than answers: First results from a questionnaire study on insights needed by science policy makers

*Priority scale of 1-5, with 1=urgent to 5=nice to know*

### Priority Questions

#### Temporal Analysis

- 1 funding trends in individual institutes, all NIH, all funding / Topical – to examine NIH scientific topic area broadly and in detail
- 1 Topical/temporal – how are the current structures of scientific/translational/clinical research changing, what are the emerging areas, and how are the submitted applications different from awarded grants in these areas.
- 2 What new biomedical fields of research are emerging, and 1) is NIH currently funding such research, 2) are there enough trained scientists to address these new research fields, and 3) where is the emerging fields research being conducted (are there geographic clusters)?
- 2 Temporal patterns of distribution / Temporal – examine scientific trends
- 3 What are the prevailing trends in topics receiving funding across NIH? By specific institute?
- 3 Meso vs global (topical/temporal) – how does NIH funding relate to funding from other agencies/countries

#### Geospatial Analysis

- 1 Diffusion of knowledge globally
- 5 Have there been any changes in degree of international collaboration in the biomedical sciences?

#### Topic Analysis

- 1 What NIH Funds / How do we identify emerging concept / Are there emerging areas of opportunity to which NIH should direct more support?
- 1 How are NIH research findings being used by partners, health providers and the public?
- 2 How do we identify gaps in knowledge?
- 2 How can we characterize (or categorize) the research that NIH supports? AND How do these areas of investment compare to public health needs?

#### Network Analysis

- 2 How can we quickly understand the current network of nodule and collaboration? What information will we need to do so?
- 4 Have our efforts to encourage interdisciplinary research been effective? And which strategies have been the most effective?
- ? Identify instances of knowledge transfer within and across research networks
- ? Network approaches to measuring or detecting innovation? E.g. publication or concept that disturbs the stability of a network.

42



## Type of Analysis vs. Scale of Level of Analysis

	<b>Micro/Individual</b> (1-100 records)	<b>Meso/Local</b> (101-10,000 records)	<b>Macro/Global</b> (10,000 < records)
<b>Statistical Analysis/Profiling</b>	Individual person and their expertise profiles	Larger labs, centers, universities, research states	All of NSF SA, all of sci
<b>Temporal Analysis (When)</b>	Funding portfolio of one individual	Public bursts of PNAS	113 Years of P Research
<b>Geospatial Analysis (Where)</b>	Career trajectory of one individual	Mapping a s intellectual l	PNAS
<b>Topical Analysis (What)</b>		flows in research	VxOrd/Topic NIH funding
<b>Network Analysis (With Whom?)</b>	NSF one work of	NIH's	NIH's cy

43



## Type of Analysis vs. Scale of Level of Analysis

	<b>Micro/Individual</b> (1-100 records)	<b>Meso/Local</b> (101-10,000 records)	<b>Macro/Global</b> (10,000 < records)
<b>Statistical Analysis</b>	Individual person and	Larger labs, centers, research states	All of NSF SA, all of sci
<b>Temporal Analysis (When)</b>		Public bursts of PNAS	113 Years of P Research
<b>Geospatial Analysis (Where)</b>		Mapping a s intellectual l	PNAS
<b>Topical Analysis (What)</b>		flows in research	VxOrd/Topic NIH funding
<b>Network Analysis (With Whom?)</b>	NSF one work of	NIH's	NIH's cy

Common analysis types are

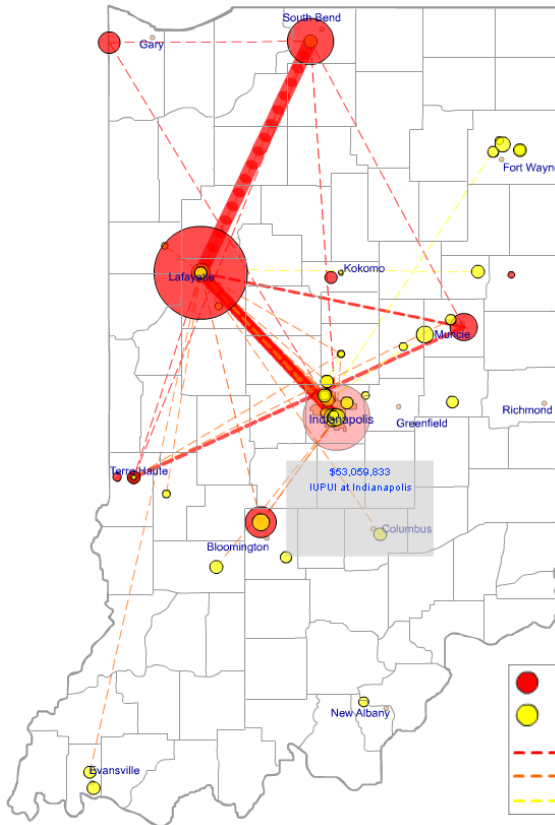
- Temporal
- Geospatial
- Topical
- Network

or combinations thereof.

*The data used determines the scope of the analysis.*

*We also list the main analysis goal.*

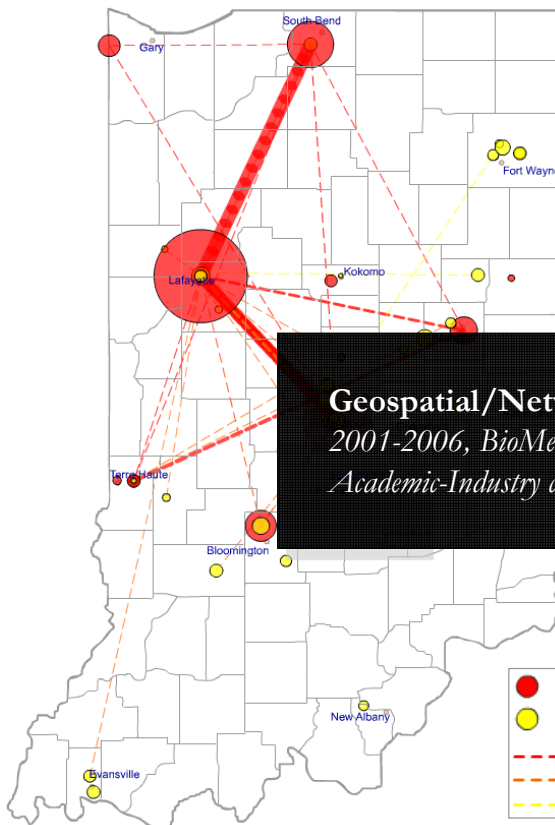
44



## Mapping Indiana's Intellectual Space

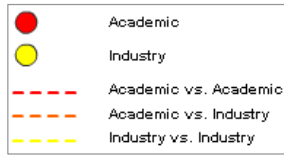
Identify

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



## Mapping Indiana's Intellectual Space

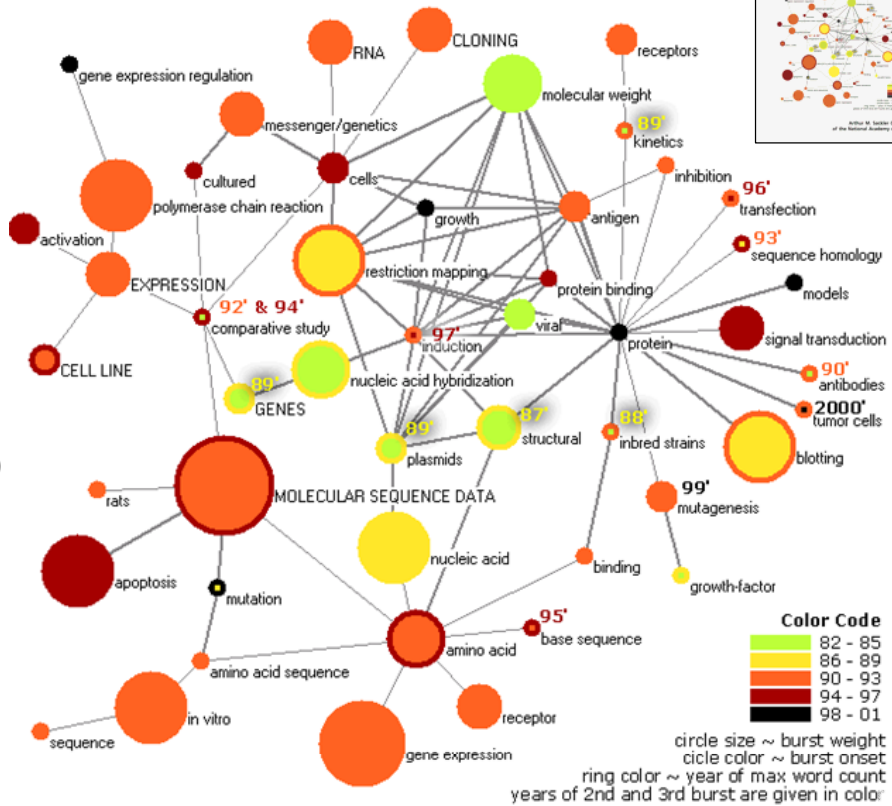
**Geospatial/Network Analysis**  
 2001-2006, BioMed, IN Scope  
*Academic-Industry collaborations and knowledge diffusion*



## 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.

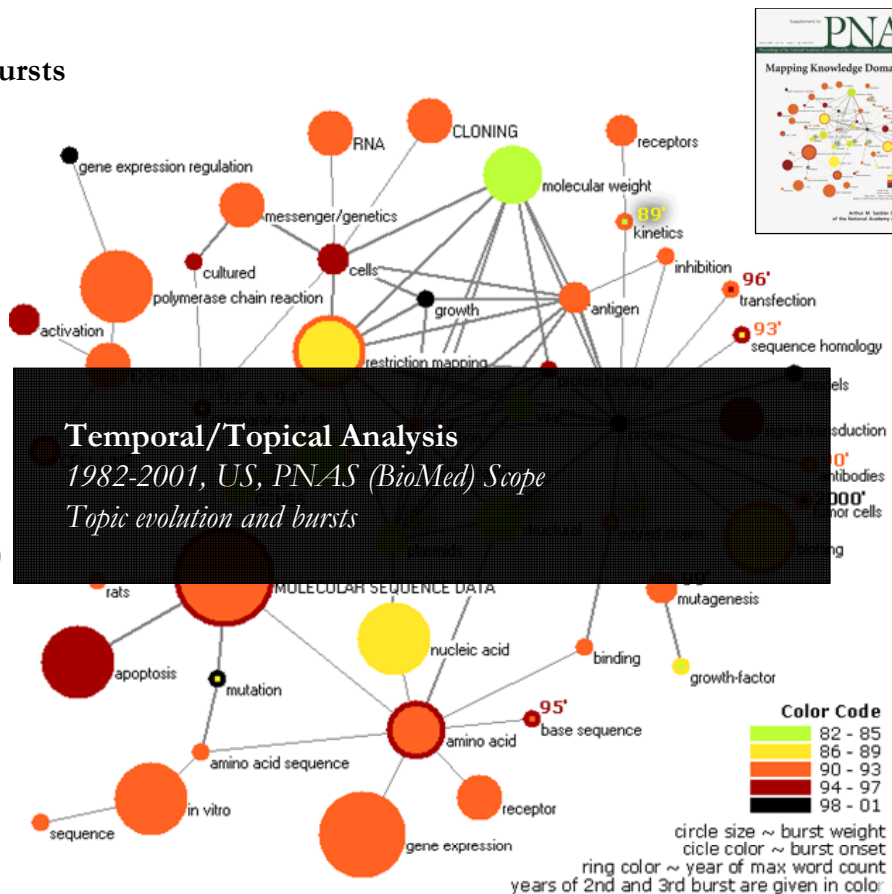
*Mane & Börner. (2004)  
PNAS, 101(Suppl. 1):  
5287-5290.*



## 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.*





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

### Research Institutions

Börner, Katy, Penumarthy, Shashikant, Meiss, Mark and Ke, Weimao. (2006)

*Mapping the Diffusion of Scholarly Knowledge Among Major U.S. Research*

*Institutions. Scientometrics. 68(3), pp. 415-426*

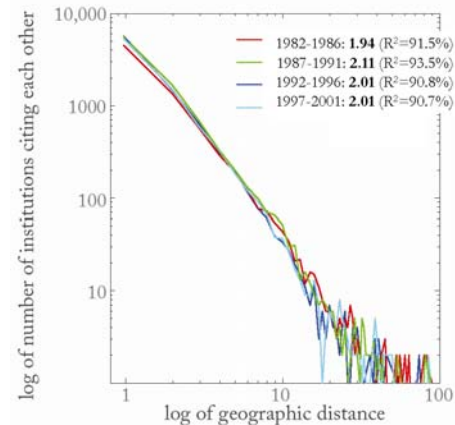
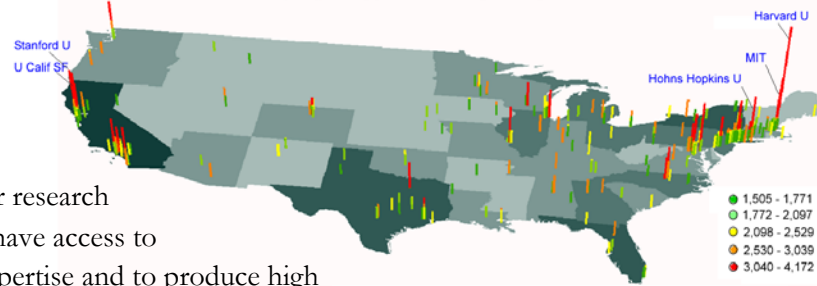


#### Research questions:

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

#### 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.



49

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

### Research Institutions

Börner, Katy, Penumarthy, Shashikant, Meiss, Mark and Ke, Weimao. (2006)

*Mapping the Diffusion of Scholarly Knowledge Among Major U.S. Research*

*Institutions. Scientometrics. 68(3), pp. 415-426*



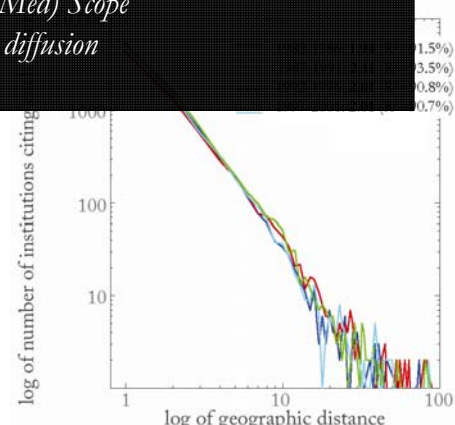
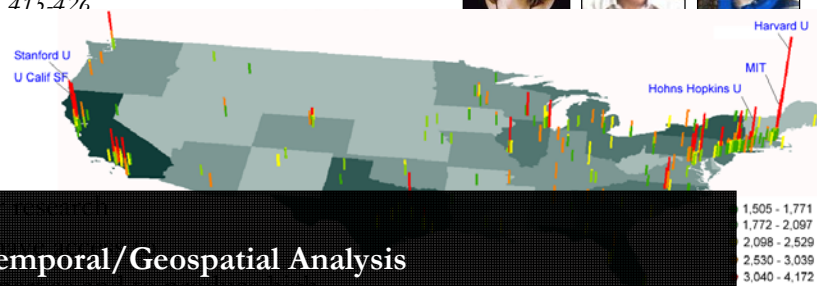
#### Research questions:

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



50

## Research Collaborations by the Chinese Academy of Sciences

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



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.

51

## 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*

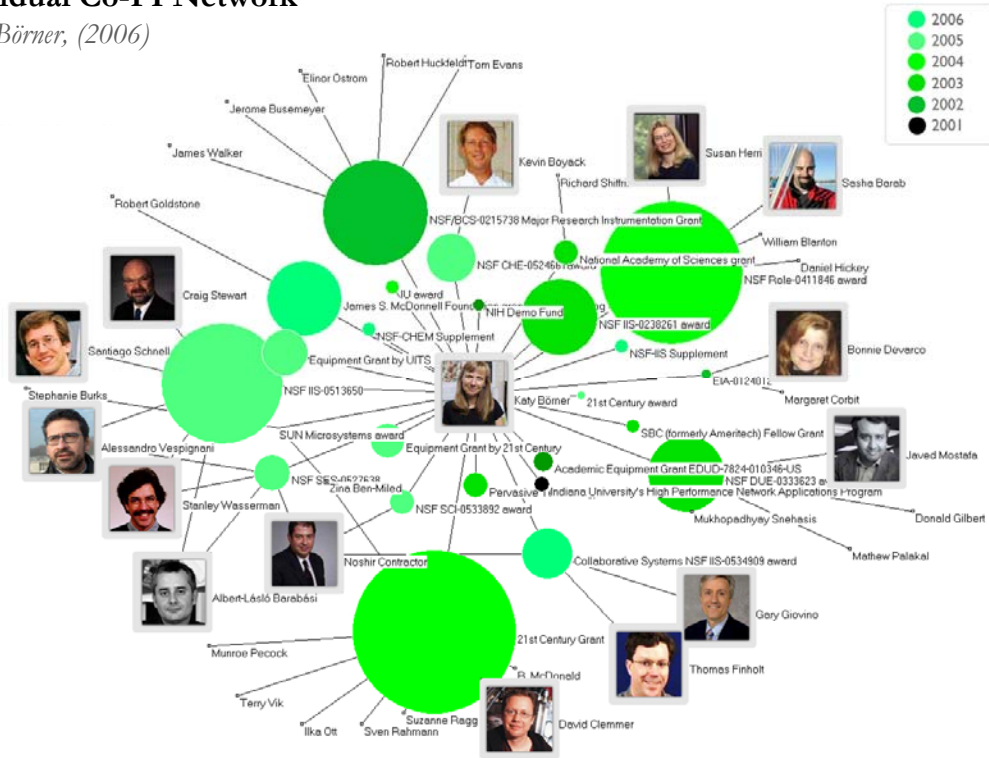
*Collaboration and knowledge diffusion via co-author networks*

This map highlights the research 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.

52

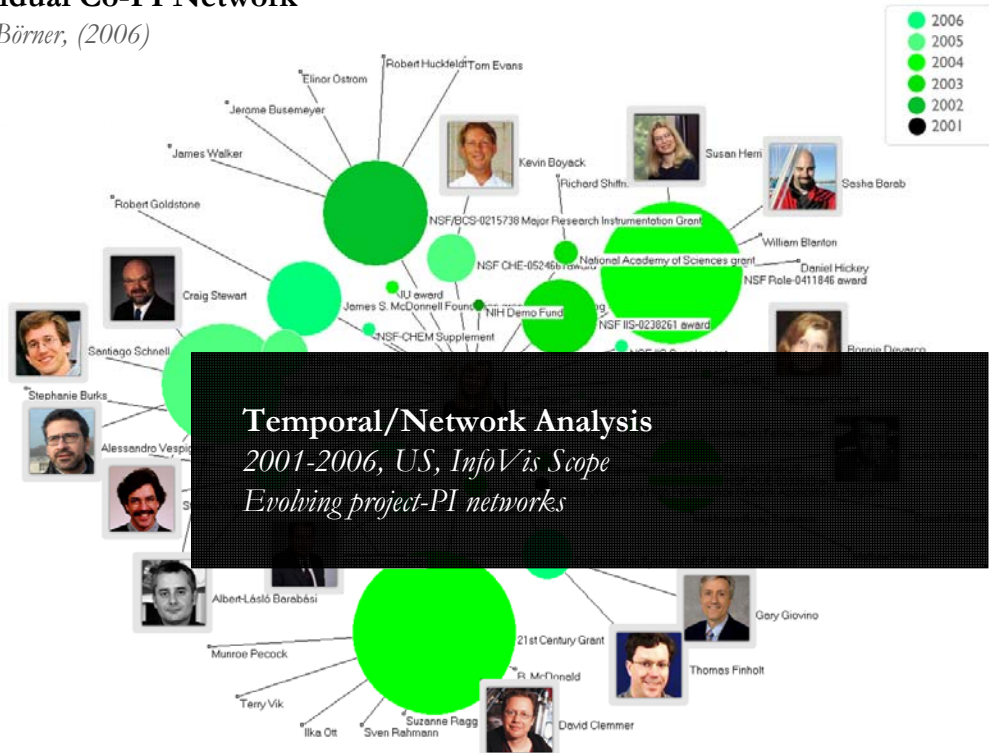
# Individual Co-PI Network

Ke & Börner, (2006)



# Individual Co-PI Network

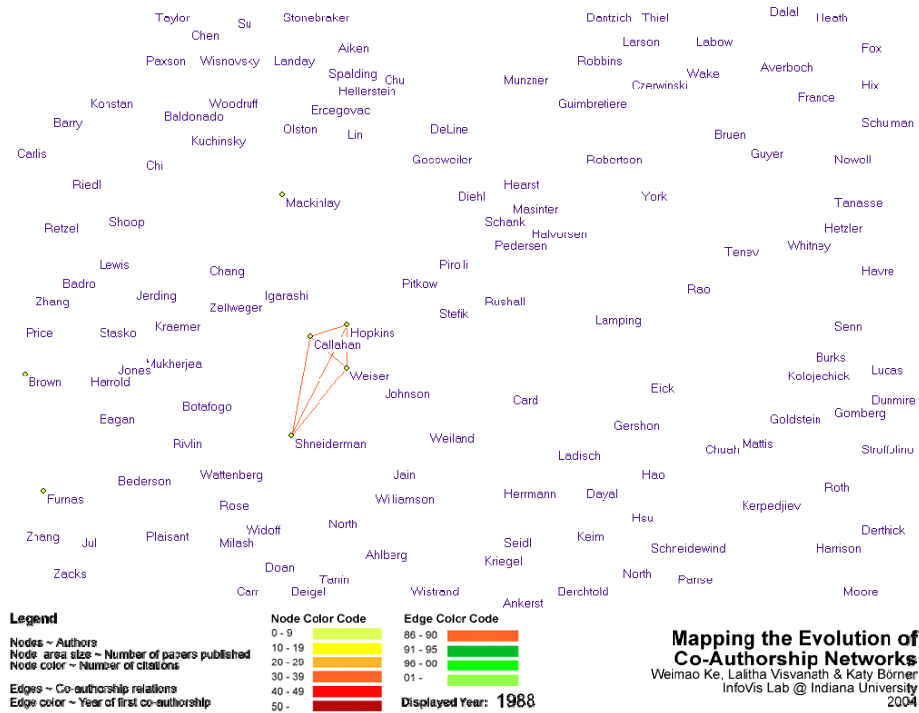
Ke & Börner, (2006)





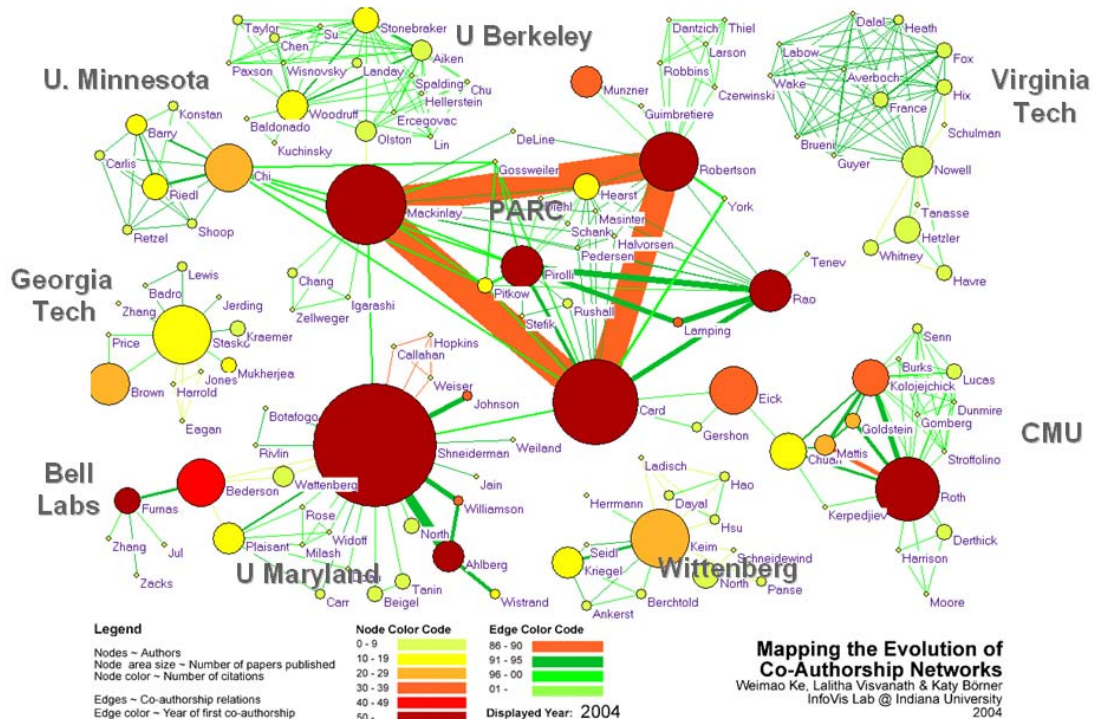
# Mapping the Evolution of Co-Authorship Networks

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



# Mapping the Evolution of Co-Authorship Networks

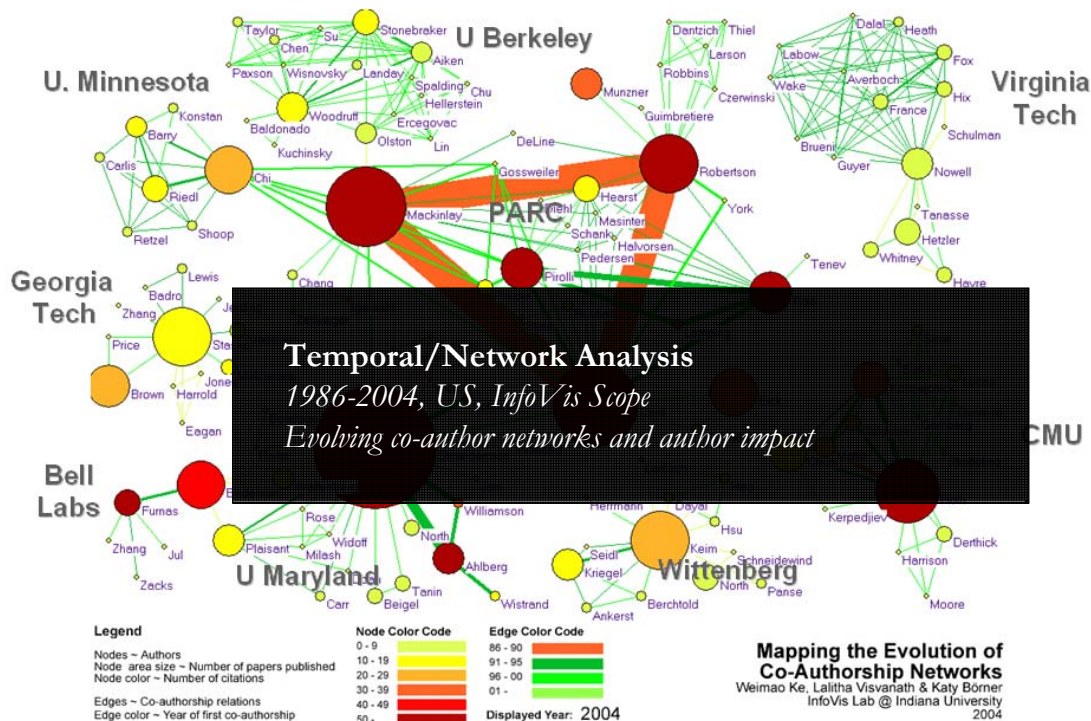
Ke, Visvanath & Börner, (2004) Won 1st price at the IEEE InfoVis Contest.





# Mapping the Evolution of Co-Authorship Networks

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



57

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

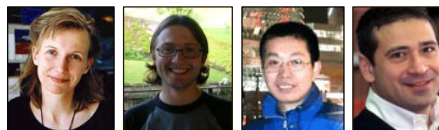
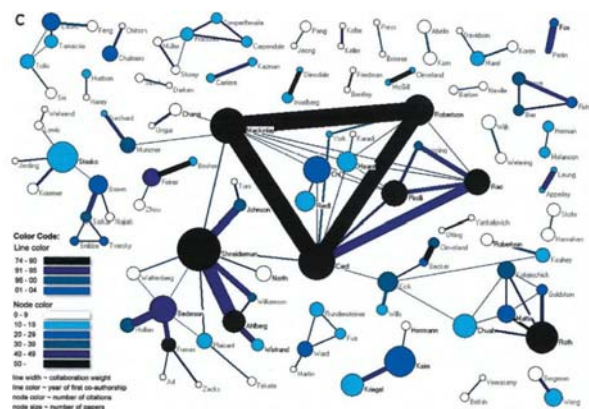
Börner, Dall'Asta, Ke & Vespignani (2005) *Complexity*, 10(4):58-67.

### Research question:

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

### Contributions:

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



58

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

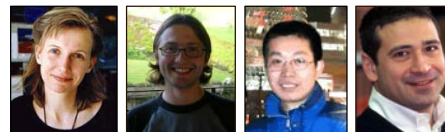
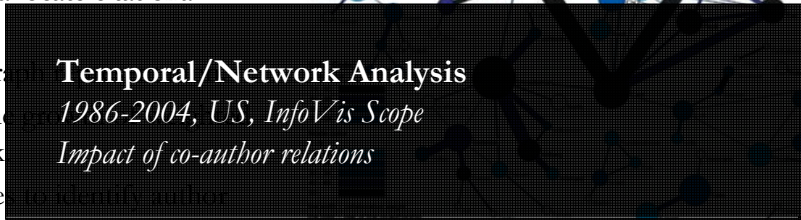
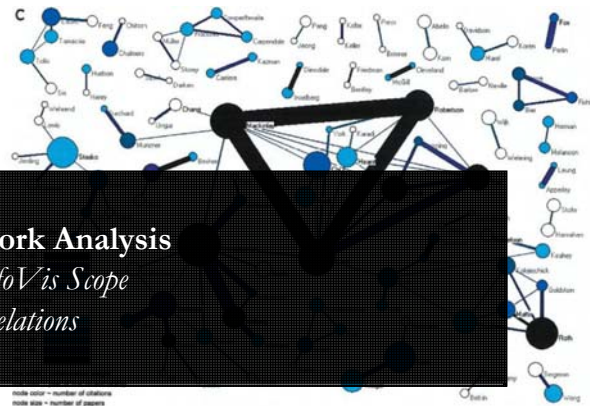
Börner, Dall'Asta, Ke & Vespignani (2005) *Complexity*, 10(4):58-67.

## Research question:

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

## Contributions:

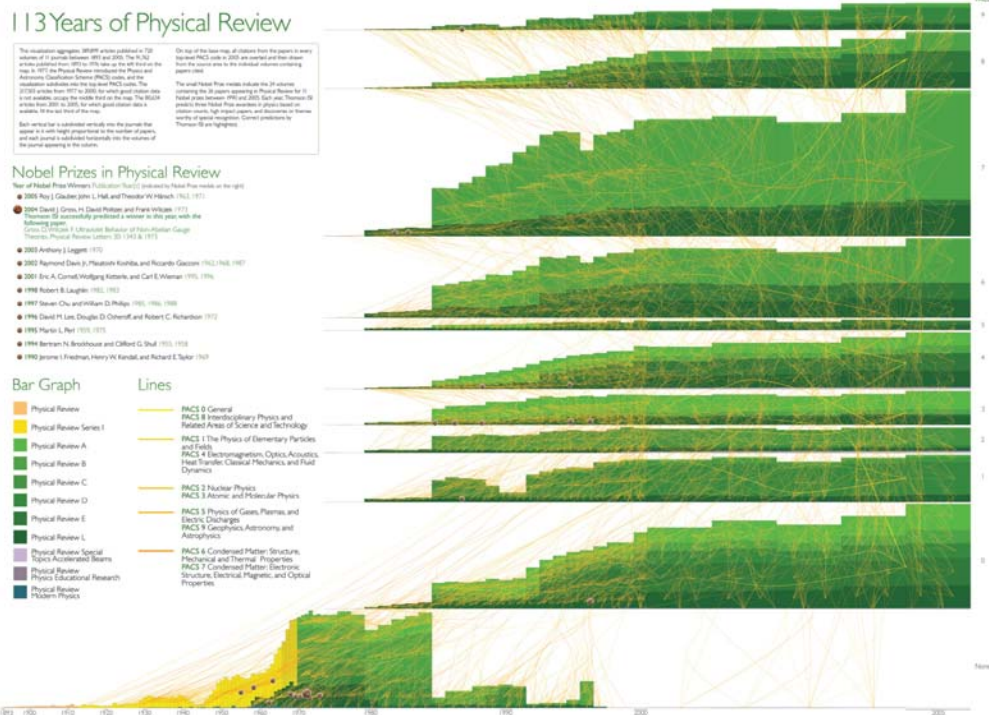
- New approach to allocate citational credit.
- Novel weighted graph.
- Visualization of the co-author network.
- Centrality measure of impact.
- Global statistical analysis of paper production and citations in correlation with co-authorship team size over time.
- Local, author-centered entropy measure.



# 113 Years of Physical Review

[http://scimaps.org/dev/map\\_detail.php?map\\_id=171](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*)

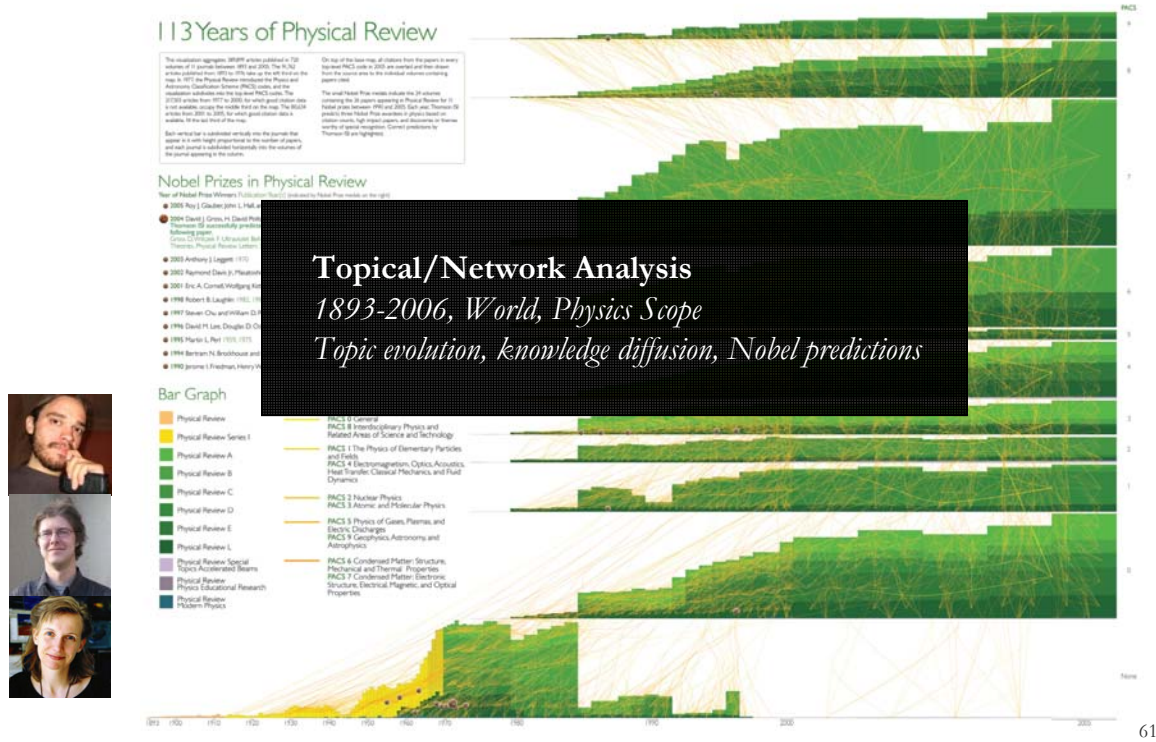




# 113 Years of Physical Review

[http://scimaps.org/dev/map\\_detail.php?map\\_id=171](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)

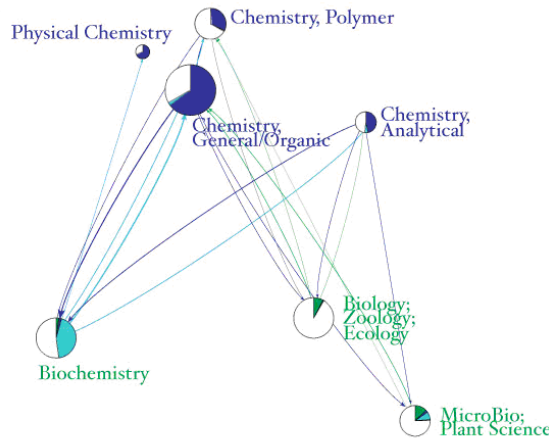


# Topical Composition and Knowledge Flow Patterns in Chemistry Research for 1974 and 2004

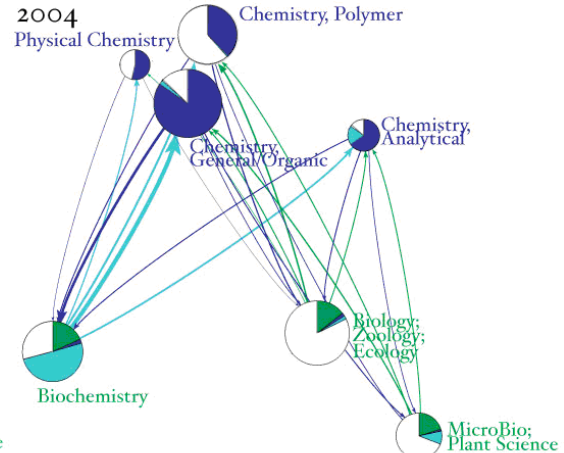
Kevin W. Boyack, Katy Börner, & Richard Klavans (2007)

## Chemistry - Biology Interface

1974



2004



Number of papers by cluster



Fraction of papers by cluster



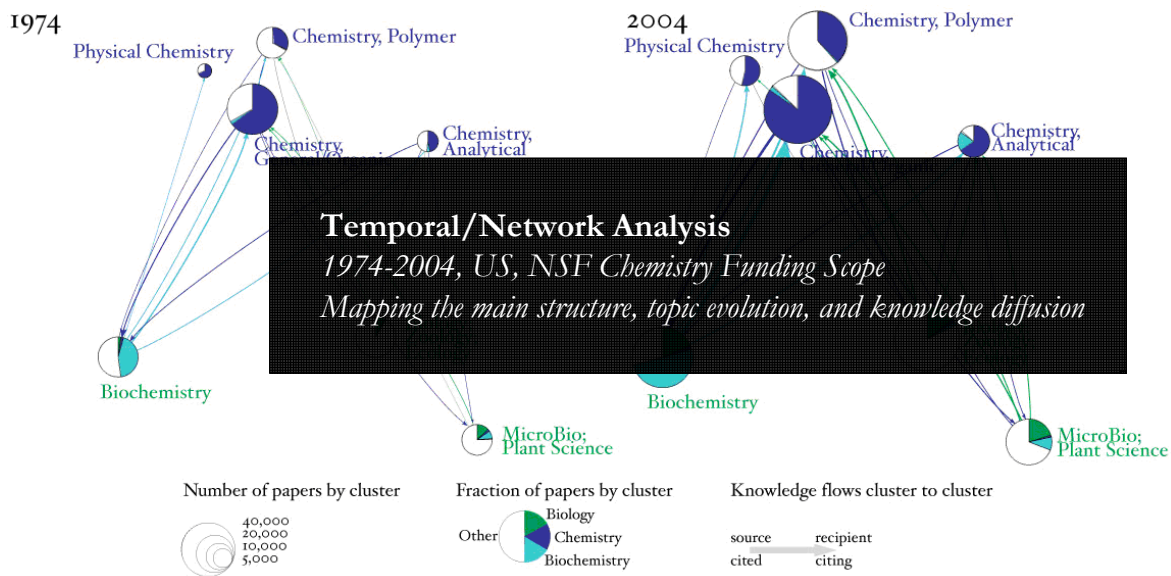
Knowledge flows cluster to cluster



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

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

## Chemistry - Biology Interface

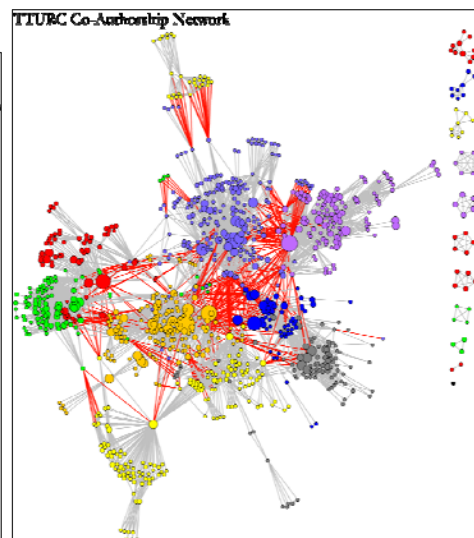
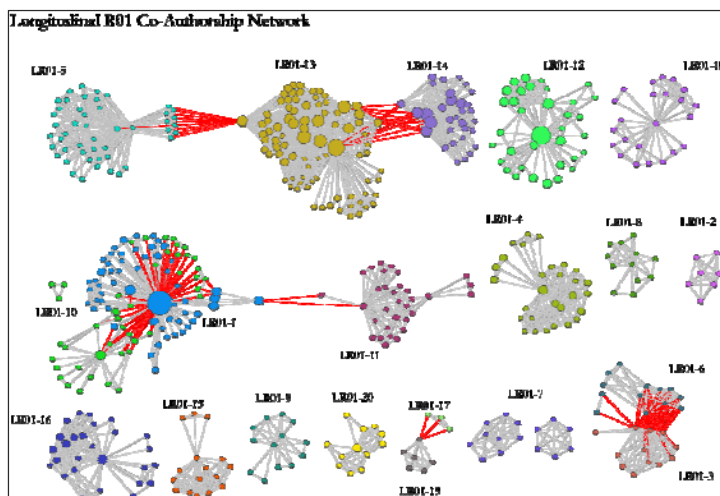
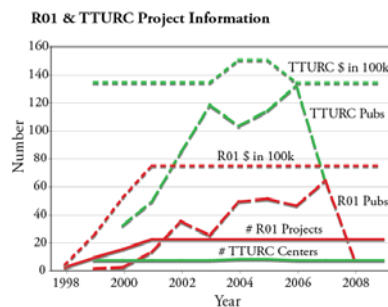


63

## 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.

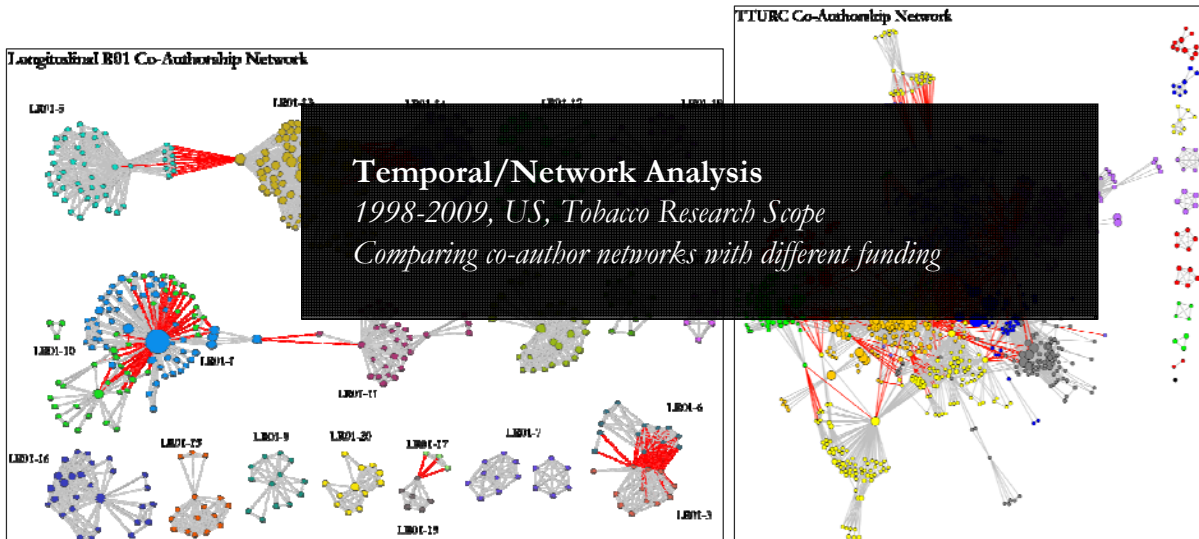
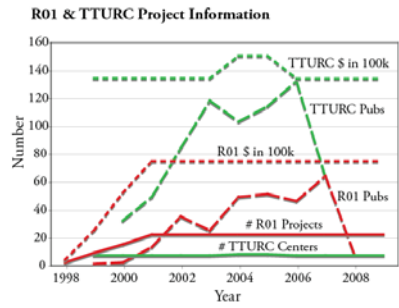


64

# 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

**(c) Details**

For each PDF file: Overlay of all matched journal references on 554 scientific fields (nodes) in UCSD Map of Science

Circle size denotes # references

Colors and names of science areas that are cited

Alphabetic listing of cited journals and # of times cited

**(d) Top-10 Most Similar**

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

# 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**

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

**For each PDF file: Overlay of all matched journal references on 554 scientific fields (nodes) in UCSD Map of Science**

**Circle size denotes # references**

**Colors and names of science areas that are cited**

**Alphabetic listing of cited journals and # of times cited**

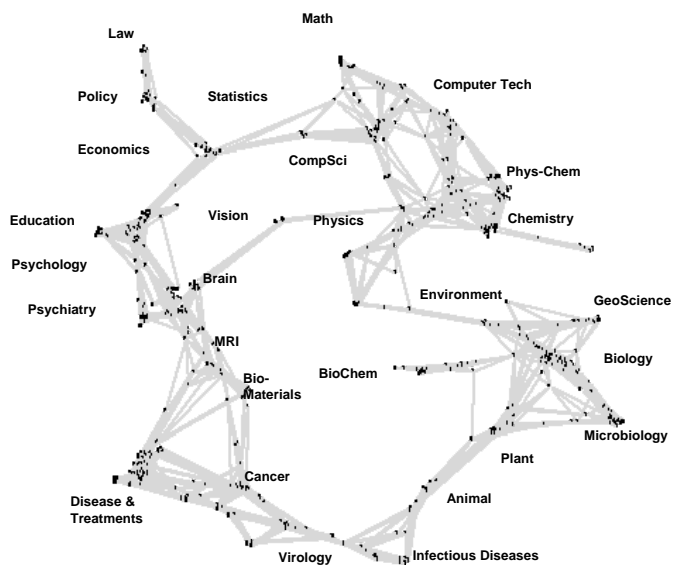
**Top-n most similar PDF files identified based on journal name co-occurrences**  
 The similarity of each PDF file to itself is 1

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

## Latest 'Base Map' of Science

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

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

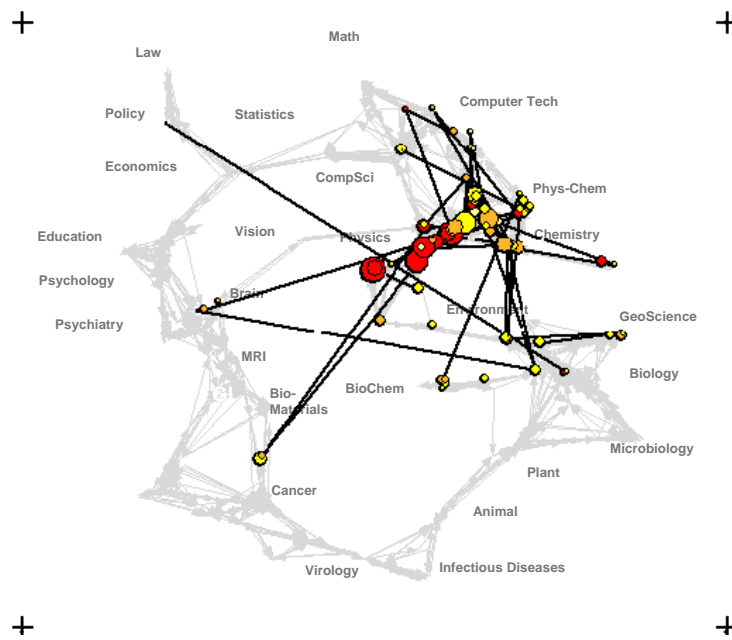




## Science map applications: Identifying core competency

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

### Funding patterns of the US Department of Energy (DOE)

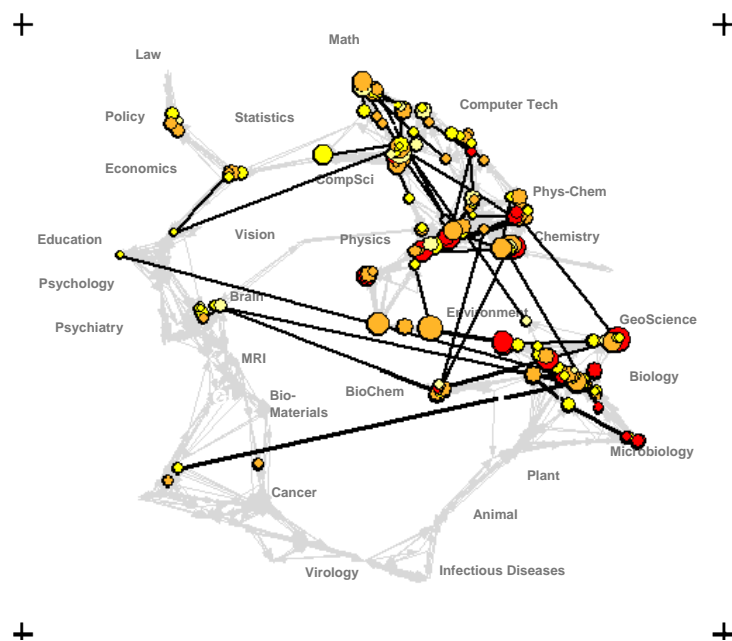


69

## Science map applications: Identifying core competency

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

### Funding Patterns of the National Science Foundation (NSF)

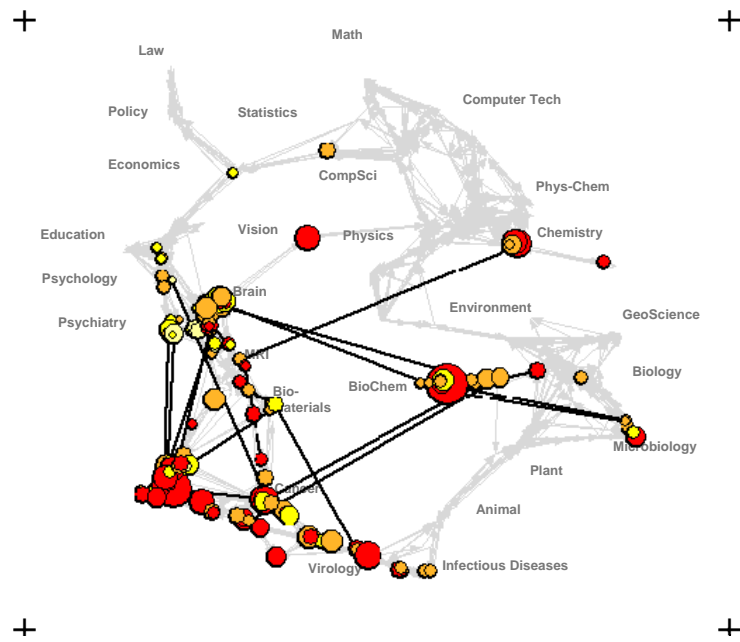


70

## Science map applications: Identifying core competency

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

### Funding Patterns of the National Institutes of Health (NIH)

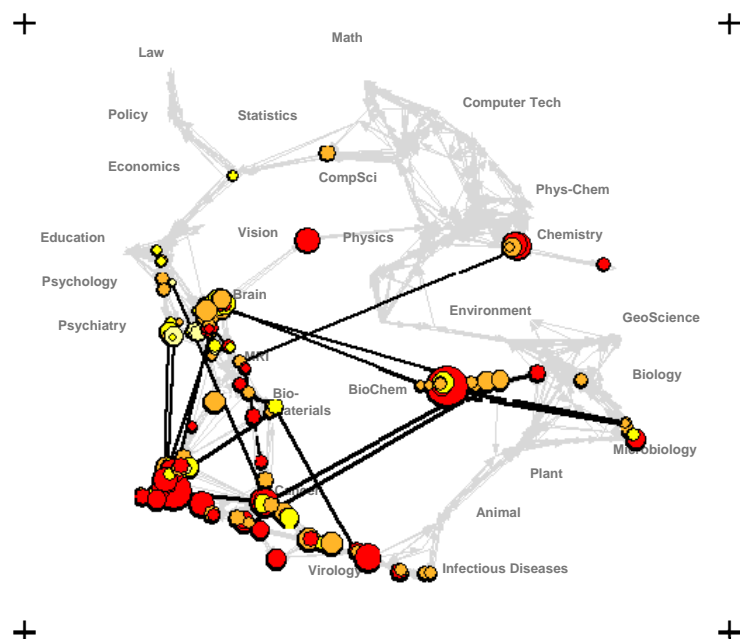


71

## Science map applications: Identifying core competency

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

### Funding Patterns of the National Institutes of Health (NIH)



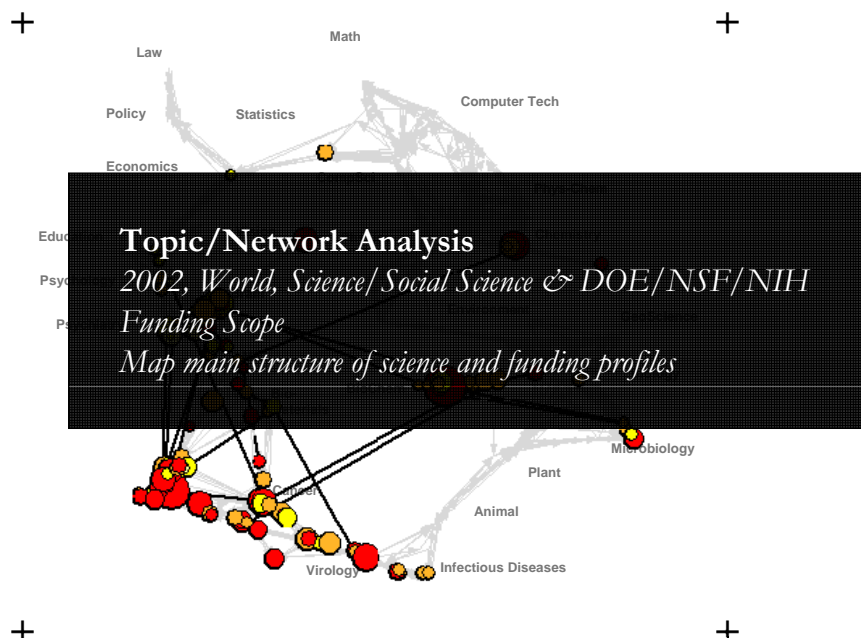
72



## Science map applications: Identifying core competency

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

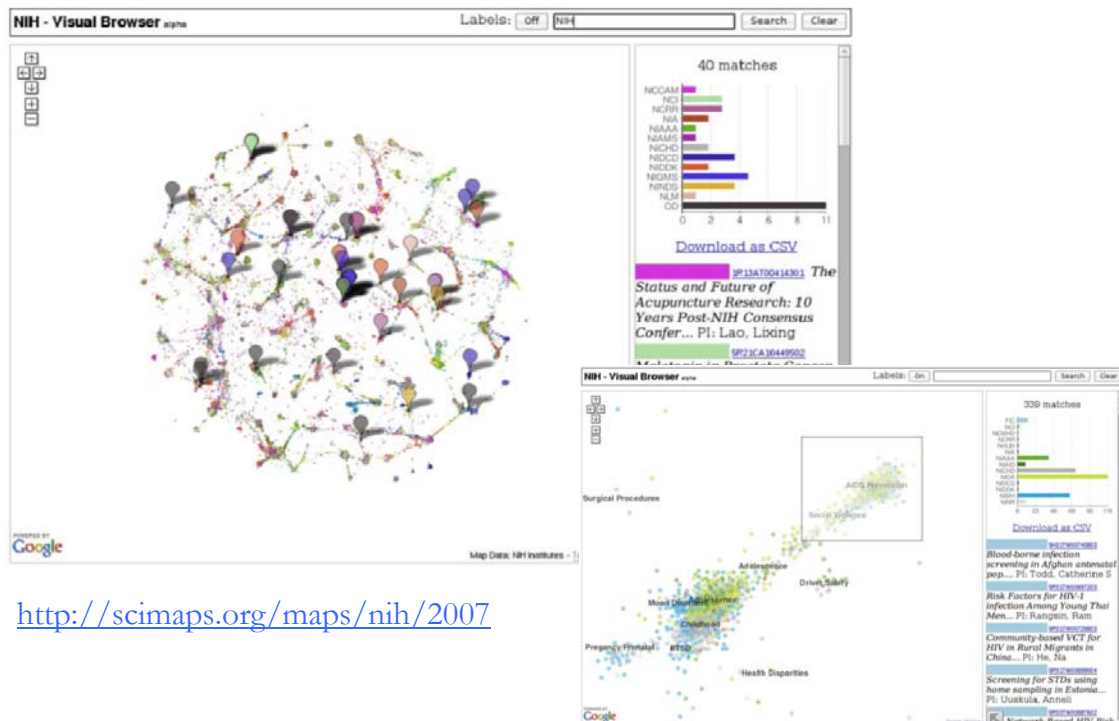
### Funding Patterns of the National Institutes of Health (NIH)



73

## Interactive Science Map of NIH Funding

Herr II, Bruce W., Talley, Edmund M, Burns, Gully APC, Newman, David & La Rowe, Gavin. (2009).

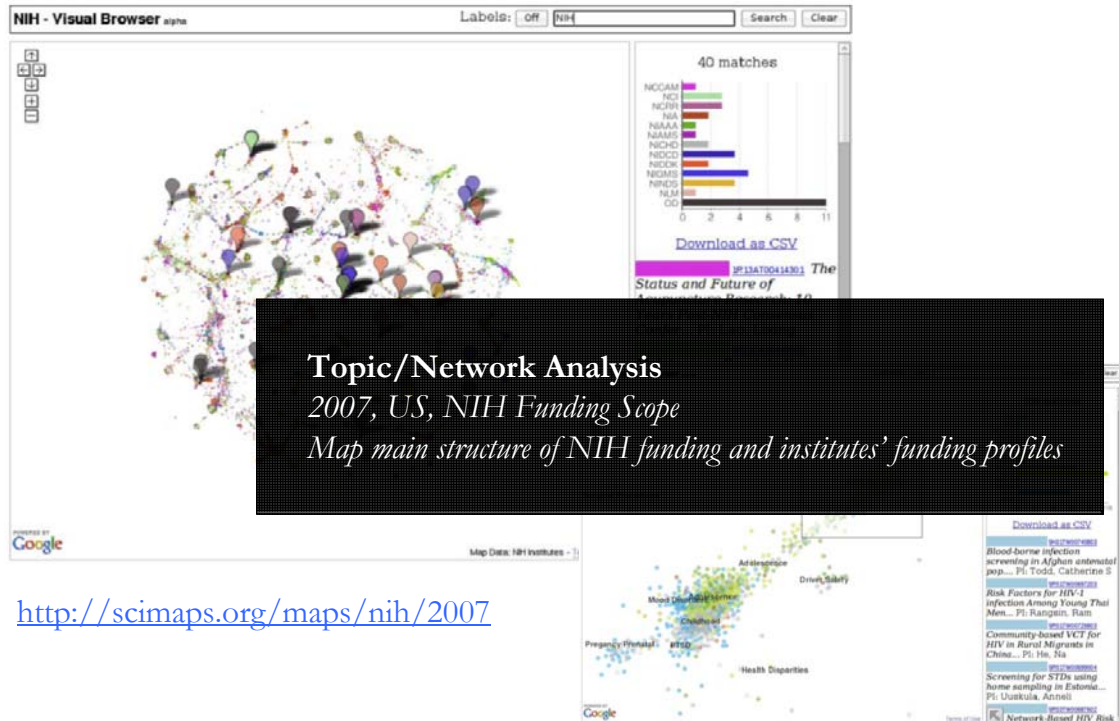


<http://scimaps.org/maps/nih/2007>

74

## Interactive Science Map of NIH Funding

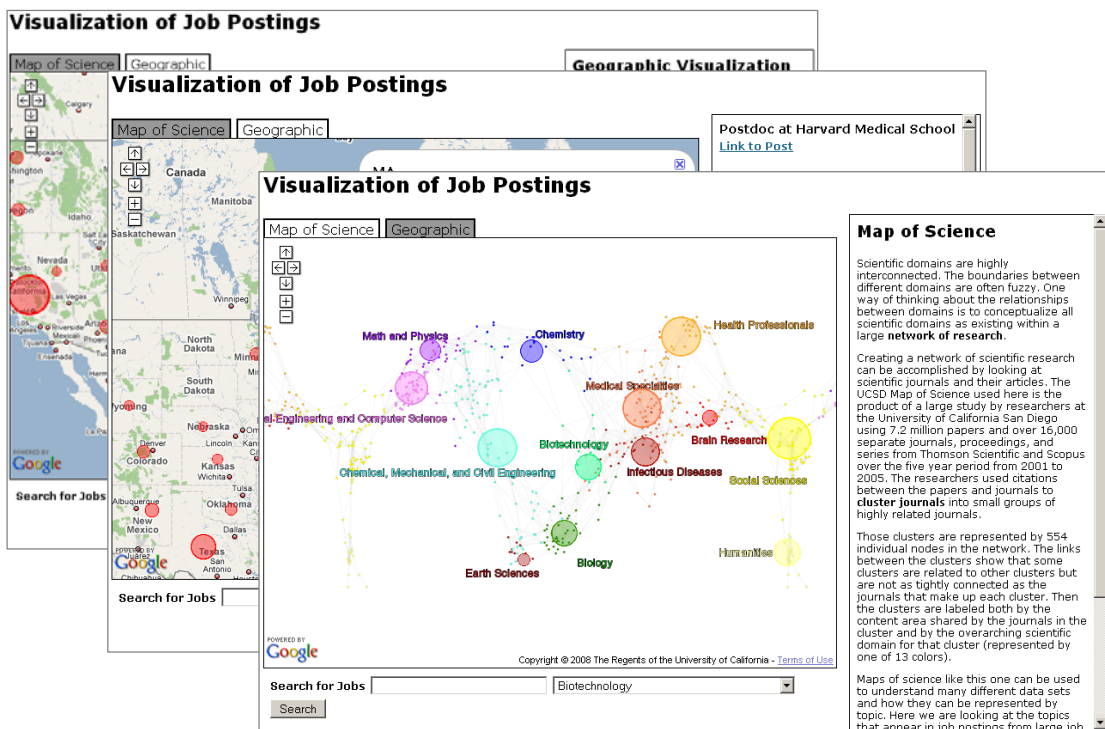
Herr II, Bruce W., Talley, Edmund M, Burns, Gully APC, Newman, David & La Rowe, Gavin. (2009).



75

## Interactive World and Science Map of S&T Jobs

Angela Zoss, Michael Connover, Katy Börner (2010).



76

# Interactive World and Science Map of S&T Jobs

Angela Zoss, Michael Connover, Katy Börner (2010).

**Visualization of Job Postings**

Map of Science | Geographic

**Visualization of Job Postings**

Map of Science | Geographic

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

**Visualization of Job Postings**

Map of Science | Geographic

**Map of Science**  
Scientific domains are highly

**Geospatial/Topic Analysis**  
*2008/2009, World, 100 Job RSS feeds*  
*Map evolving job market in real time*

Earth Sciences | Biology | Humanities

POWERED BY Google

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

Search for Jobs

Search for Jobs

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 job postings from large job

77

## [#01] Science of Science Research

- Brief History (Atlas timeline)
- Micro, Meso, Macro Studies
- Workflow Design
- Sample Studies / Mapping Science Exhibit
- Validation
- Promising Research Directions

78





## General Process of Analyzing and Mapping Science

(1) Data Extraction	(2) Unit of Analysis	(3) Measures	Layout (often one code does both similarity and ordination steps)		(6) Display
			(4) Similarity	(5) Ordination	
<b>Searches</b> <i>WoS</i> <i>Scopus</i> <i>Google Scholar</i> <i>MEDLINE</i> <i>Patents</i> <i>Funding</i>  <b>Broadening</b> <i>By citation</i> <i>By terms</i>	<b>Common Choices</b> <i>Journal</i> <i>Document</i> <i>Author</i> <i>Term</i>	<b>Counts/ Frequencies</b> <i>Attributes (e.g. terms)</i> <i>Author citations</i> <i>Co-citations</i> <i>By year</i>  <b>Thresholds</b> <i>By counts</i>	<b>Scalar (unit by unit matrix)</b> <i>Direct citation</i> <i>Co-citation</i> <i>Combined linkage</i> <i>Co-word/co-term</i> <i>Co-classification</i>  <b>Vector (unit by attribute matrix)</b> <i>Vector Space Model (words/terms)</i> <i>Latent Semantic Analysis (LSA)</i> <i>Singular Value Decomposition (SVD)</i>  <b>Correlation (if desired)</b> <i>Pearson's R on any of above</i>	<b>Dimensionality Reduction</b> <i>Eigenvector/Eigenvalue Solutions</i> <i>Factor Analysis (FA)</i> <i>Principal Components Analysis (PCA)</i> <i>Multi-Dimensional Scaling (MDS)</i> <i>Pathfinder Networks (PFNet)</i> <i>Self-Organizing Maps (SOM)</i> <i>Topics Model</i>  <b>Cluster Analysis</b> <i>Partition</i> <i>Hierarchical</i>  <b>Spatial Placement</b> <i>Triangulation</i> <i>Force-Directed Placement (FDP)</i>	<b>Interaction</b> <i>Browse</i> <i>Pan</i> <i>Zoom</i> <i>Filter</i> <i>Query</i> <i>Detail on demand</i>  <b>Analysis &amp; Interpretation</b>

Börner, Katy, Chen, Chaomei, and Boyack, Kevin. (2003) *Visualizing Knowledge Domains*. In Blaise Cronin (Ed.), *Annual Review of Information Science & Technology, Volume 37*, Medford, NJ: Information Today, Inc./ American Society for Information Science and Technology, chapter 5, pp. 179-255.



## NWB Tool Interface Components

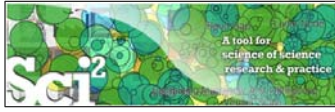
**Console** displays data operations (save, load, view, etc.) and algorithm input parameters, selection, & acknowledgements as well as error reporting.

**Scheduler** lists what algorithms you've used and displays algorithm progress.

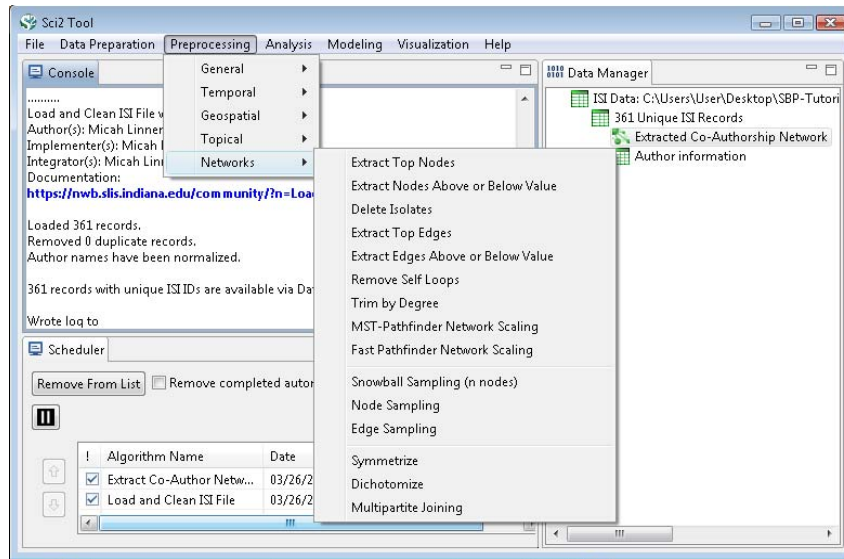
**Data Manager** keeps track of all datasets that are available for algorithmic visualization or manipulation.

Algorithm Name	Date	Time	% Complete

- Table
- Matrix
- Plot
- Text
- GUESS
- Tree
- Network



## Science of Science (Sci<sup>2</sup>) Tool Interface



### Acknowledgments

This work is supported in part by the Cyberinfrastructure for Network Science center and the School of Library and Information Science at Indiana University, the National Science Foundation under Grant No. SBE-0738111 and IIS-0513650, and the James S. McDonnell Foundation.

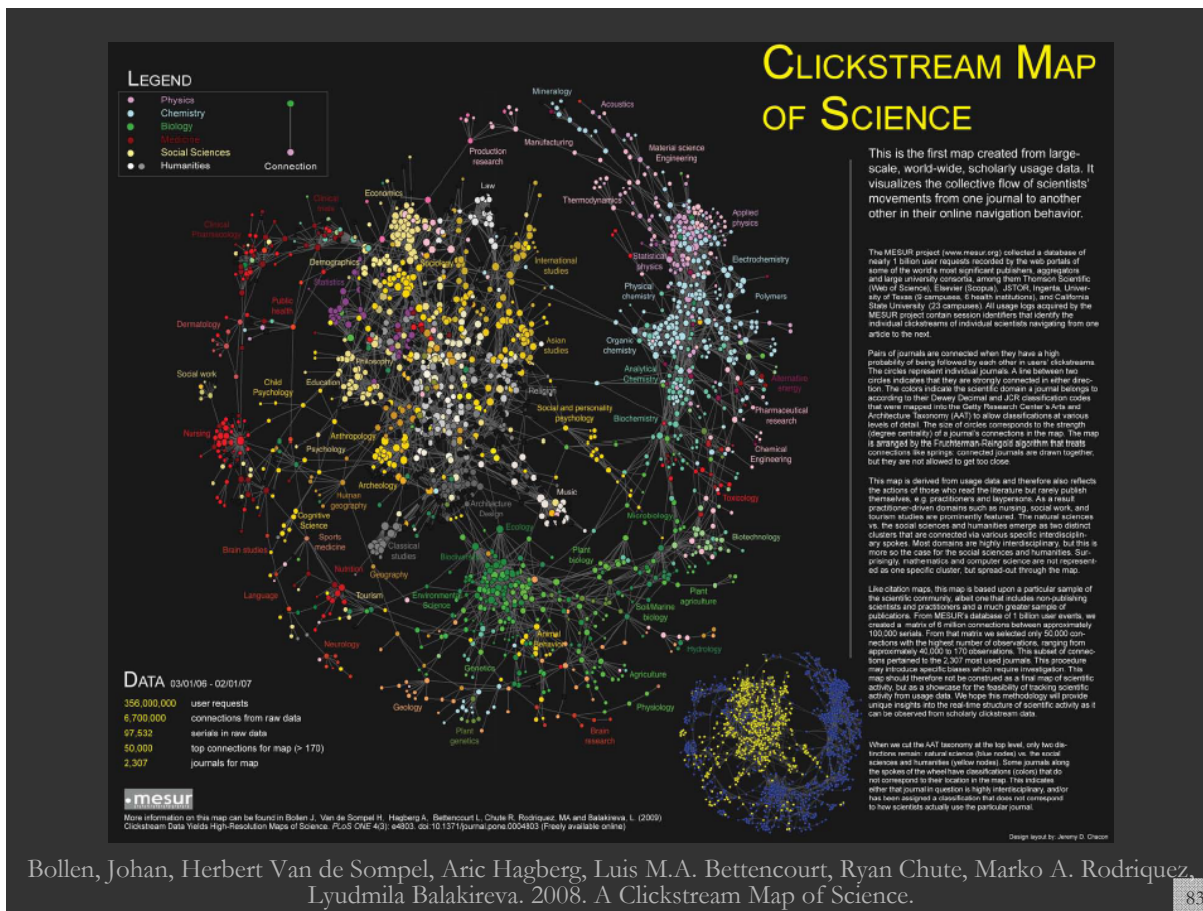


81

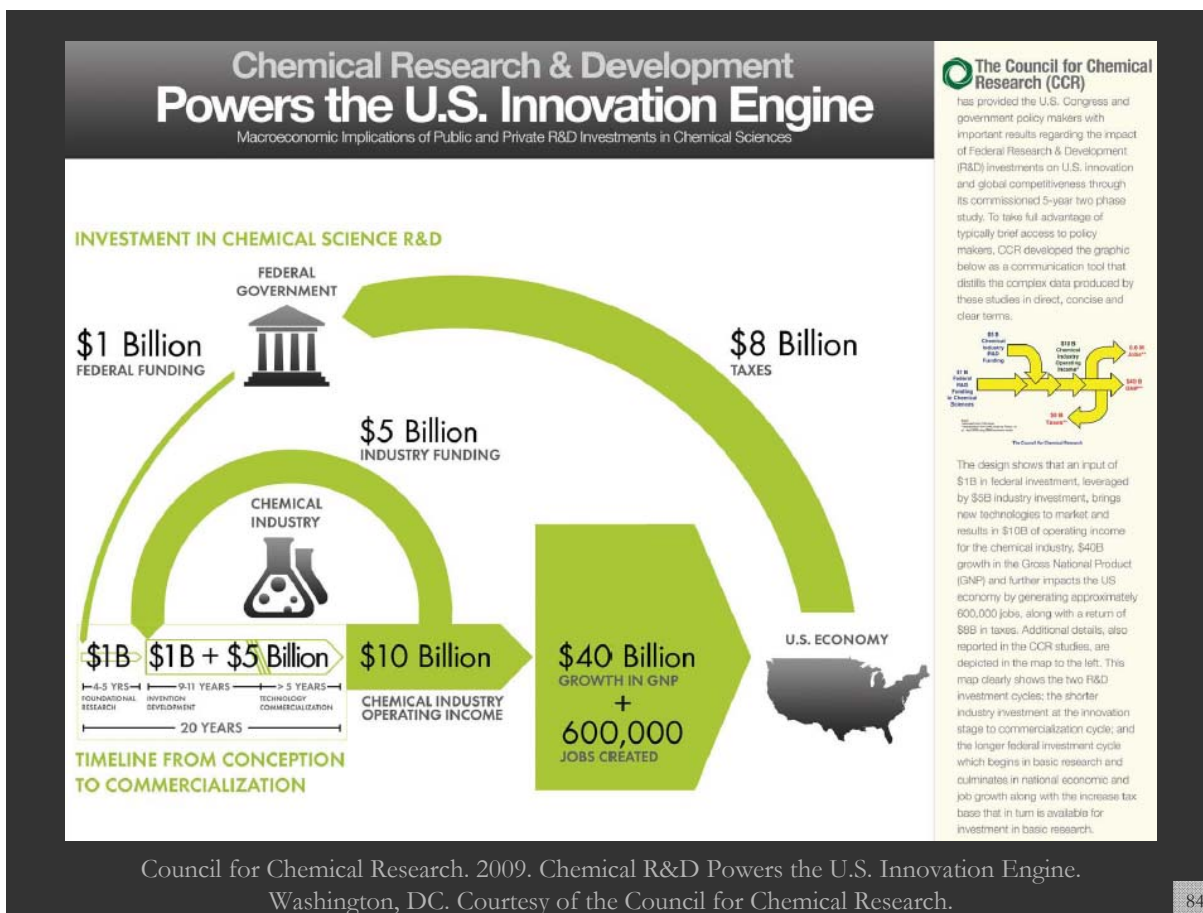
## [#01] Science of Science Research

- Brief History (Atlas timeline)
- Micro, Meso, Macro Studies
- Workflow Design
- Sample Studies / Mapping Science Exhibit
- Validation
- Promising Research Directions

82



Bollen, Johan, Herbert Van de Sompel, Aric Hagberg, Luis M.A. Bettencourt, Ryan Chute, Marko A. Rodriguez, Lyudmila Balakireva. 2008. A Clickstream Map of Science. 83



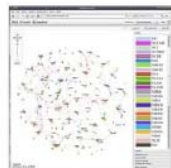


# A Topic Map of NIH Grants 2007

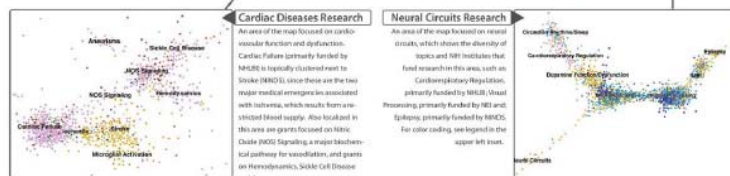
ChalkLabs UCI UCLIVNE

Bruce W. Herr II (ChalkLabs & IU), Gully Burns (IS), David Newman (UCI), Edmund Talley (NIH)

The National Institutes of Health (NIH) is organized as a multitude of Institutes and Centers whose missions are primarily focused on distinct diseases. However, disease etiologies and therapies blur scientific boundaries, and thus there is tremendous overlap in the kinds of research funded by each institute. This creates a daunting landscape for decisions on research directions, funding allocations, and policy formulations. Shown here is devised an interactive topic map for navigating this landscape, online at [www.nih.gov/td](http://www.nih.gov/td). Institute abbreviations can be found at [www.nih.gov/td](http://www.nih.gov/td).



Topic modeling, a statistical technique that automatically learns semantic categories, was applied to assess projects in terms used by researchers to describe their work, without the biases of keywords or subject headings. Grant similarities were derived from their topic mixtures, and grants were then clustered on a two-dimensional map using a force-directed simulated annealing algorithm. This analysis creates an interactive environment for assessing grant relevance to research categories and to NIH Institutes in which grants are localized.



## National Cancer Institute (NCI)

- TOP 10 TOPICS
1. Oncology Clinical Trials
  2. Cancer Treatment
  3. Cancer Therapy
  4. Carcinogenesis
  5. Risk Factor Analysis
  6. Cancer Chemotherapy
  7. Metastasis
  8. Leukemia
  9. Radiation/Oncology
  10. Cancer Chemoprevention

## National Institute of General Medical Sciences (NIGMS)

- TOP 10 TOPICS
1. Bioactive Organic Synthesis
  2. X-ray Crystallography
  3. Protein-NMR
  4. Computational Models
  5. Yeast Biology
  6. Microbiology
  7. Enzymatic Mechanisms
  8. Protein Complexes
  9. Invertebrate/Zebrafish Genomics
  10. Cell Division

## National Heart, Lung, and Blood Institute (NHLBI)

- TOP 10 TOPICS
1. Cardiac Failure
  2. Pulmonary Injury
  3. Genetic Linkage Analysis
  4. Cardiovascular Disease
  5. Atherosclerosis
  6. Hemostasis
  7. Blood Pressure
  8. Asthma/Allergy/Allergic Disease
  9. Gene Association
  10. Lipoproteins

## National Institute of Mental Health (NIMH)

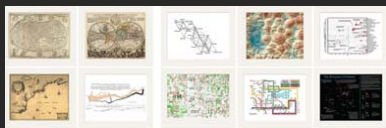
- TOP 10 TOPICS
1. Mood Disorders
  2. Schizophrenia
  3. Behavioral/Intervention Studies
  4. Alcohol Abuse
  5. Depression
  6. Cognitive-Behavior Therapy
  7. ADD/Prevention
  8. Genetic Linkage Analysis
  9. Adolescence
  10. Childhood

Herr II, Bruce W., Gully Burns, David Newman, Edmund Talley. 2007. A Topic Map of NIH Grants 2007. Bloomington, IN.

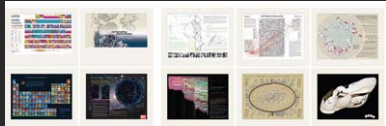
## Mapping Science Exhibit – 10 Iterations in 10 years

<http://scimaps.org>

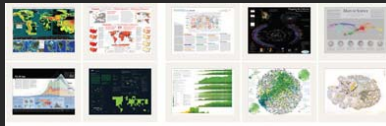
### The Power of Maps (2005)



### The Power of Reference Systems (2006)



### The Power of Forecasts (2007)



### Science Maps for Economic Decision Makers (2008)



### Science Maps for Science Policy Makers (2009)



### Science Maps for Scholars (2010)

### Science Maps as Visual Interfaces to Digital Libraries (2011)

### Science Maps for Kids (2012)

### Science Forecasts (2013)

### How to Lie with Science Maps (2014)

Exhibit has been shown in 72 venues on four continents. Currently at

- NSF, 10th Floor, 4201 Wilson Boulevard, Arlington, VA
- Marston Science Library, University of Florida, Gainesville, FL
- Center of Advanced European Studies and Research, Bonn, Germany
- Science Train, Germany.





Debut of 5<sup>th</sup> 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>



Science Maps in “Expedition Zukunft” science train visiting 62 cities in 7 months  
12 coaches, 300 m long

Opening was on April 23<sup>rd</sup>, 2009 by German Chancellor Merkel

<http://www.expedition-zukunft.de>





This is the only mockup in this slide show.  
Everything else is available today.

## [#01] Science of Science Research

- Brief History (Atlas timeline)
- Micro, Meso, Macro Studies
- Workflow Design
- Sample Studies / Mapping Science Exhibit
- Validation – out of time, please see Tutorial #6
- Promising Research Directions



## [#01] Science of Science Research

- Brief History (Atlas timeline)
- Micro, Meso, Macro Studies
- Workflow Design
- Sample Studies / Mapping Science Exhibit
- Validation
- Promising Research Directions

91



### Promising Research Directions

- Advancing the theoretical foundations of *Science of Science Research* via replicable studies, see NSF SciSIP program.
- Development of high quality, high coverage open datasets, e.g., MEDLINE, RePORTER, Scholarly Database, VIVO.
- Design and implementation of algorithms and open tools that address the specific needs, views, values of specific user groups:
  - Congress will want ROI, peer reviewed published discoveries/\$ million/year.
  - Scientists want to do cutting edge science.
  - NIH wants health improvements.
  - General public wants to benefit from tax money paid research.

Input

92

