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. **CIShell 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

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

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.

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 (Scii2) Tool home page, http://sci.slis.indiana.edu/sci2

General Considerations

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.

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.

[#01] Science of Science Research

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

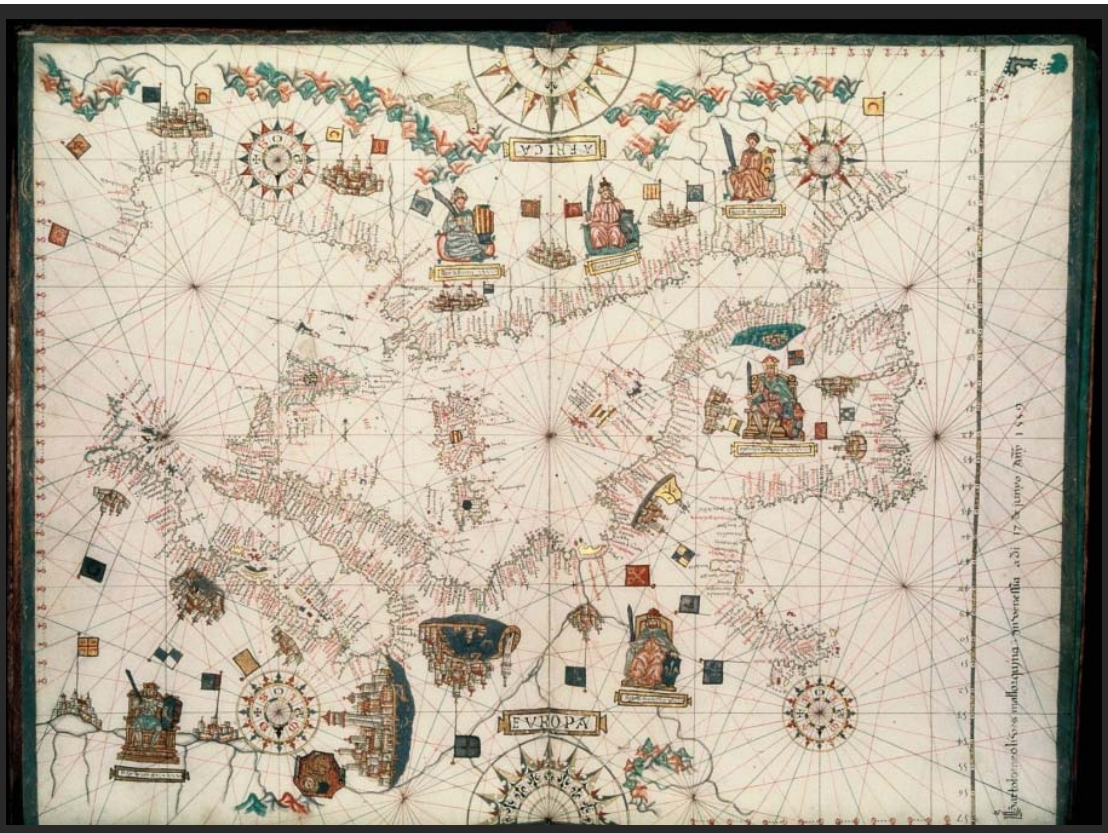
Thanks go to Kevin Boyack for making his slides available.

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

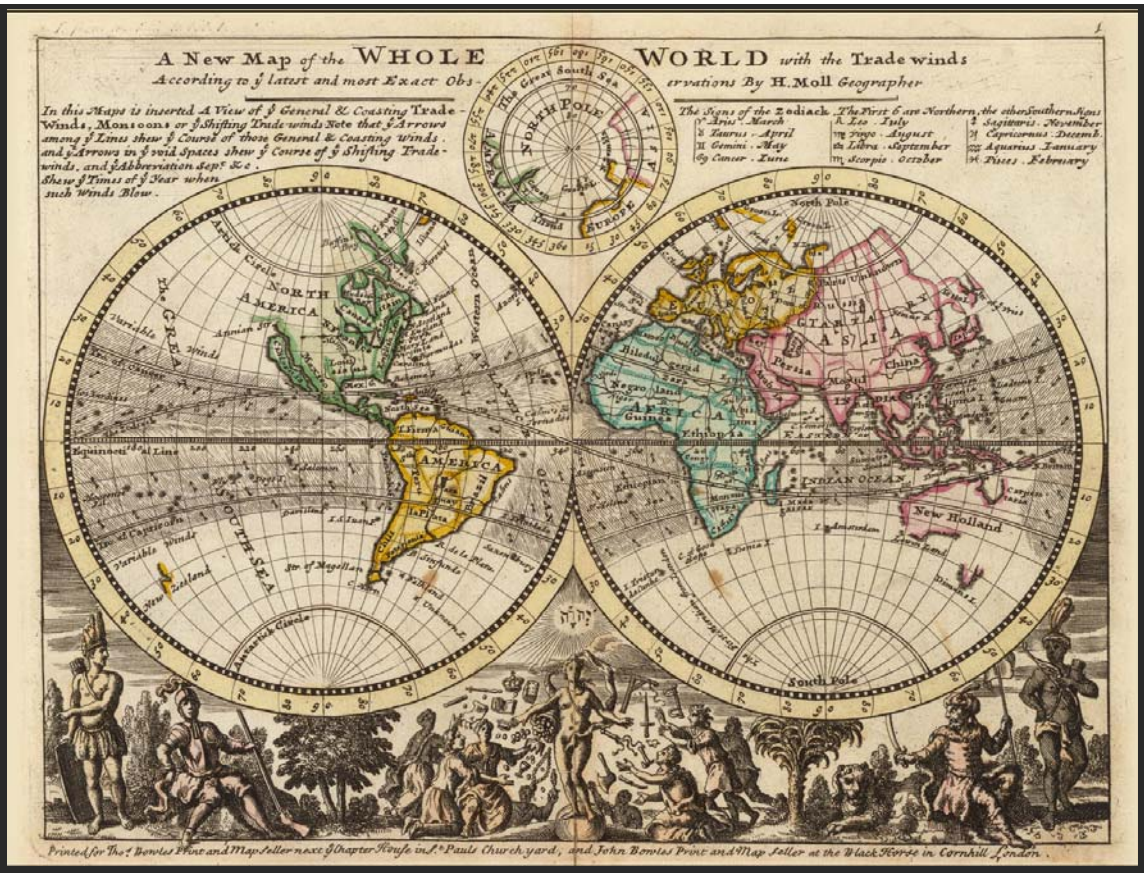


Portolan chart of the central and western Mediterranean and part of the Atlantic - Bartolome Olives - 1559

13

Islandia - Abraham Ortelius (1527-1598) - 1606

14



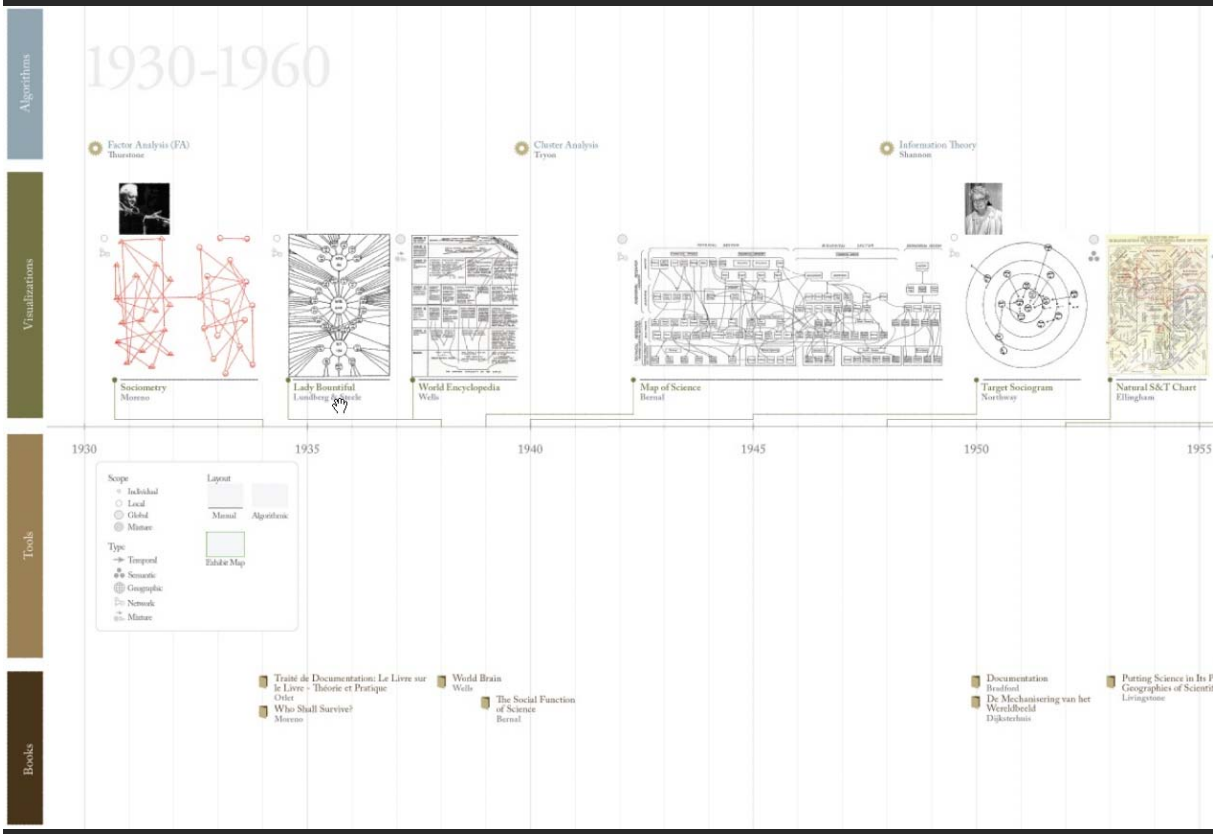
A New Map of the Whole World with Trade Winds According to the Latest and Most Exact Observations - Herman Moll - 1736

Milestones of Mapping Science

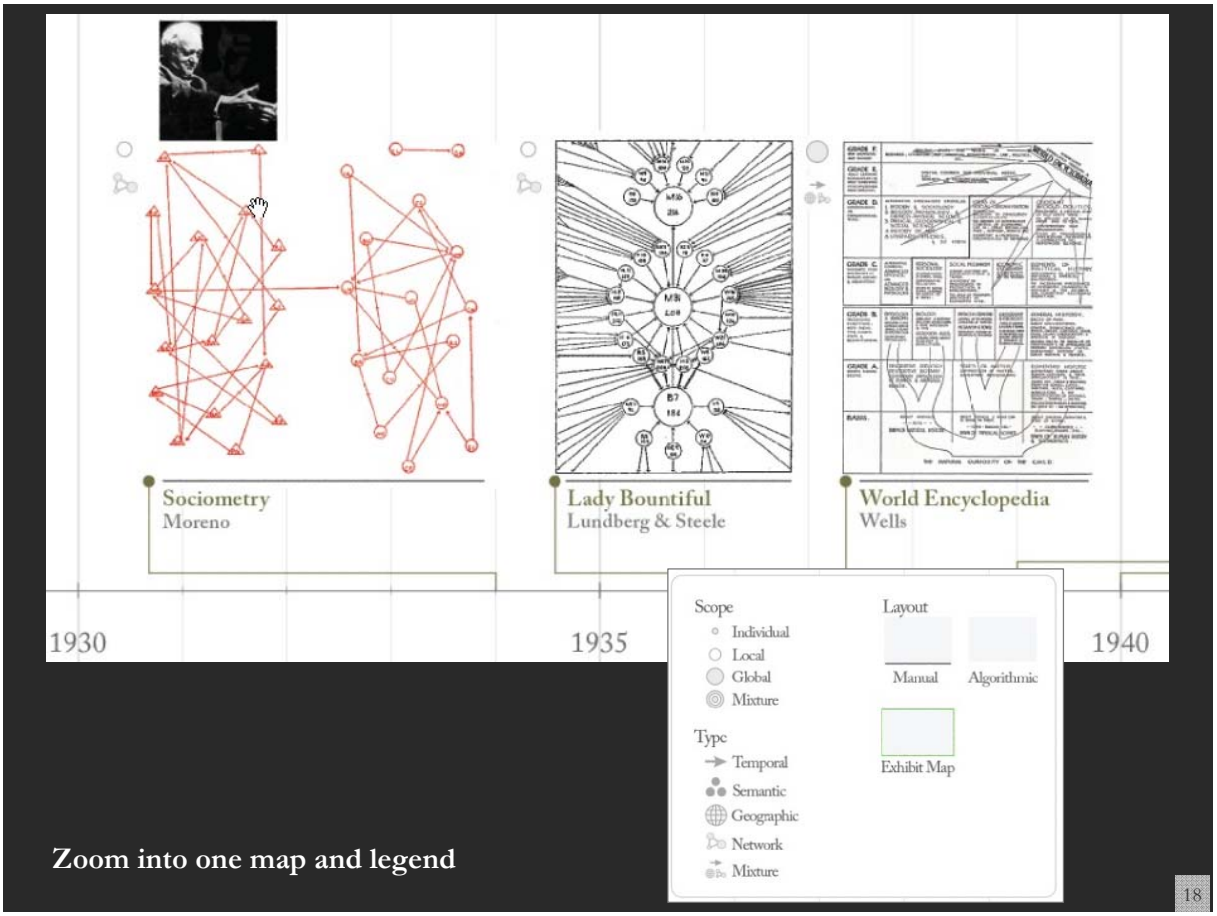
1934 **2007**

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

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

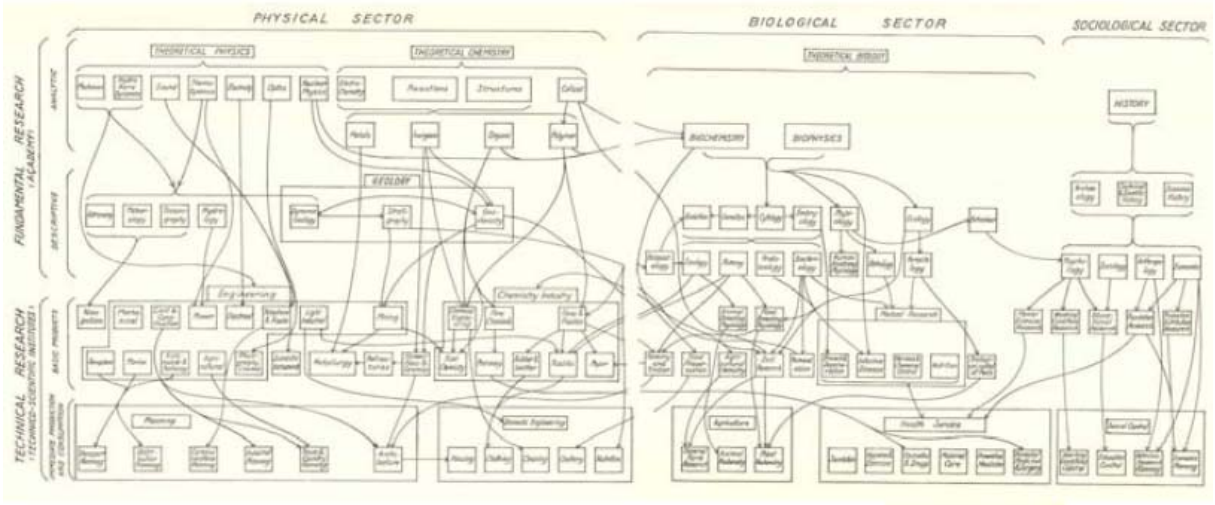


1930 1955



Zoom into one map and legend

Bernal, 1939

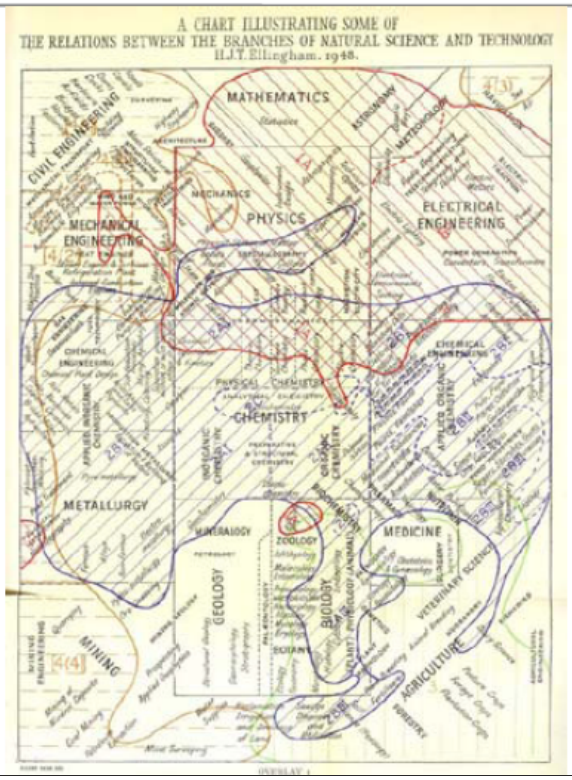


John D. Bernal was a world renowned physicist, a historian of science, and a sociologist of science. He is considered to have produced one of the first 'maps' of science.

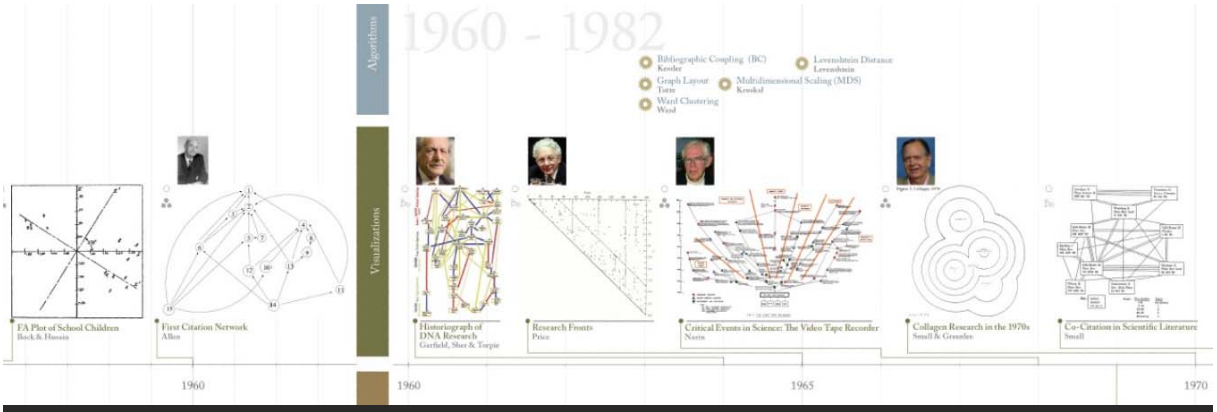
Bernal, J.D. (1939). *The Social Function of Science*. London: Routledge & Kegan Ltd.

Ellingham, 1948

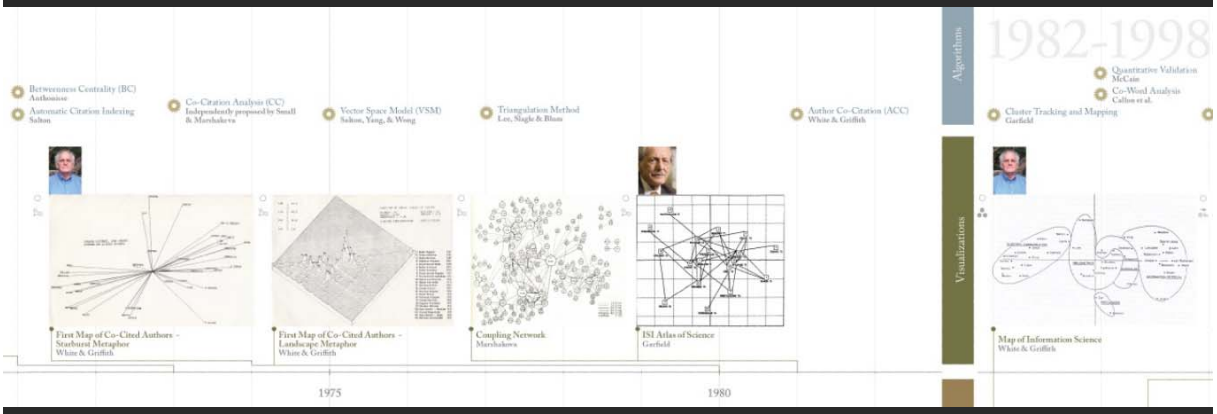
Ellingham's "Relations Between the Branches of Natural Science and Technology" with an overlay of "Abstracts or Groups of Abstracts Covering A Very Wide Field" (1948)



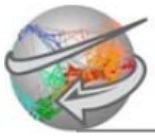
H. J. T. Ellingham (1948). "Divisions of Natural Science and Technology," Royal Society Scientific Information Conference, 21 June to 2 July 1948, London: Burlington House,



1952 1973

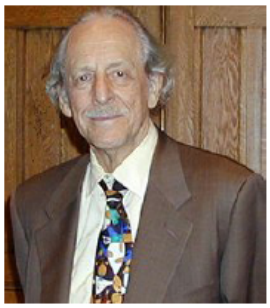


1980 1982



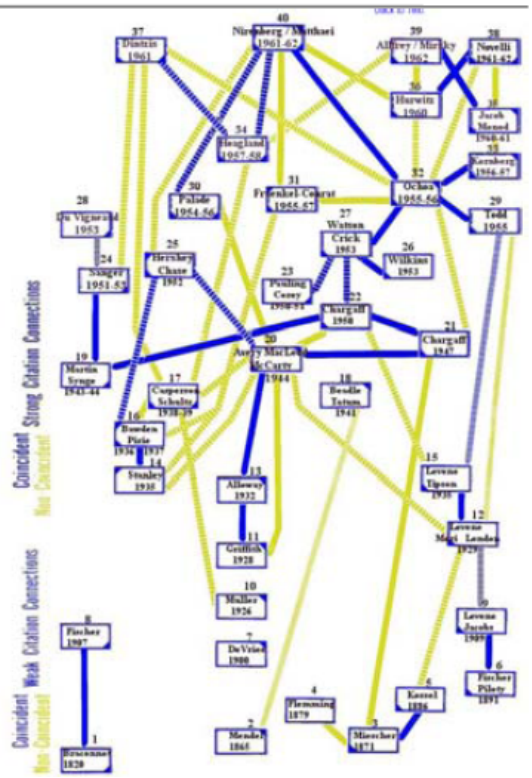
Garfield, 1964

Historiograph of DNA Development



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

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

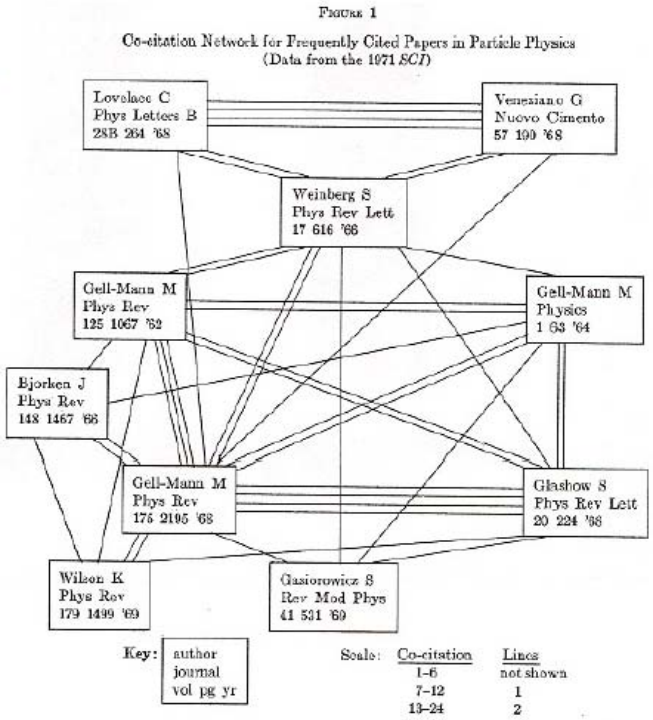


Garfield, Sher, & Torpie (1964). "The Use of Citation Data in Writing the History of Science." Air Force Office of Scientific Research under contract F49(638)-1256.

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.

1987 1997

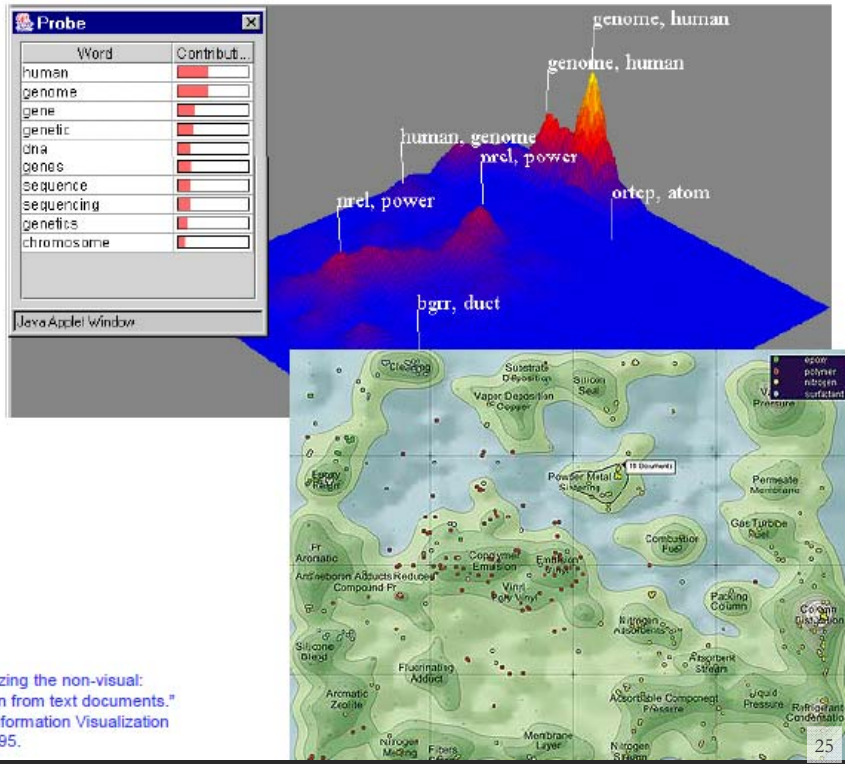
1997 1999

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



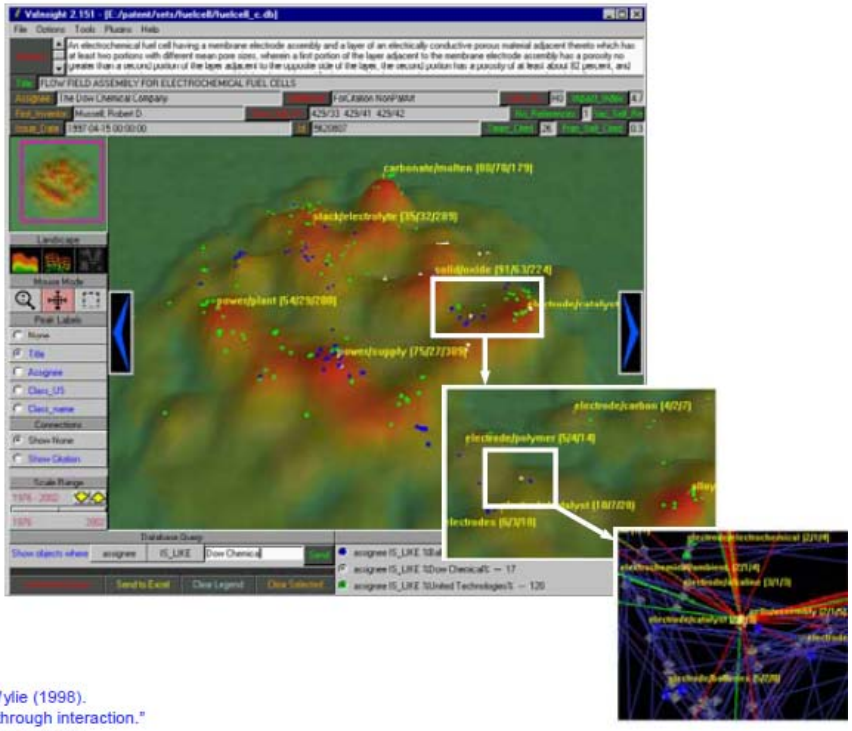
Wise, Thomas, Pennock, et al. (1995). "Visualizing the non-visual: Spatial analysis and interaction with information from text documents." Paper presented at the IEEE Symposium on Information Visualization '95, Atlanta, Georgia, USA., October 30-31, 1995.

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.

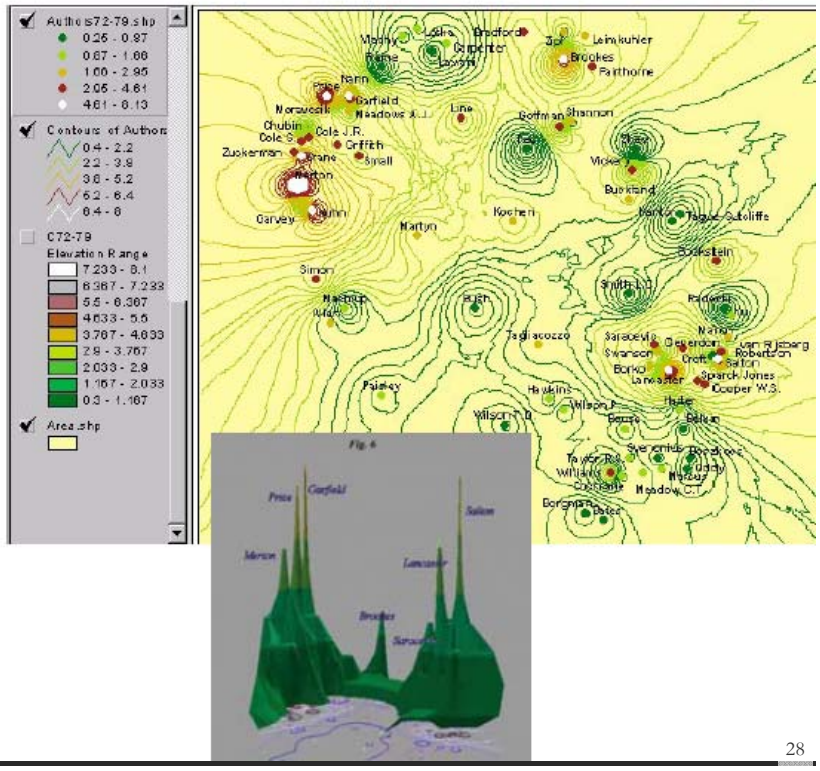
Old, 2001

Utilizing spatial information systems for non-spatial data analysis

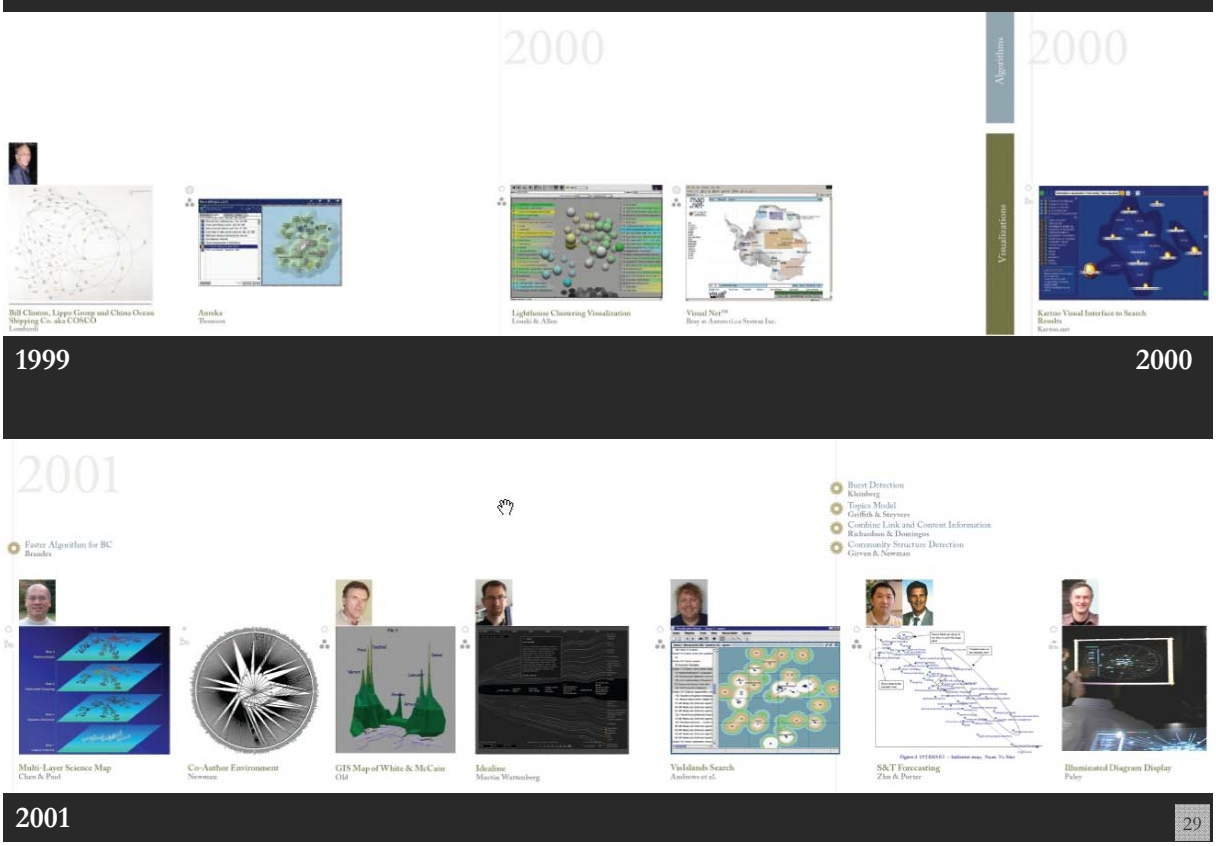
Authors in Information Science

Topography added

3D representation also



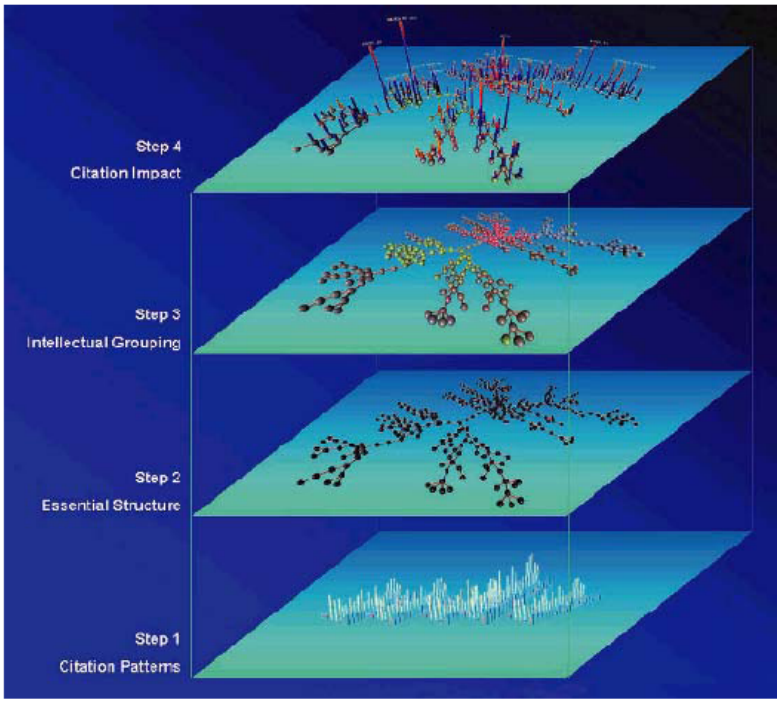
http://php.indiana.edu/~jold/SLIS/L710/L710.htm



Chen, 2001

Software: CiteSpace I

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

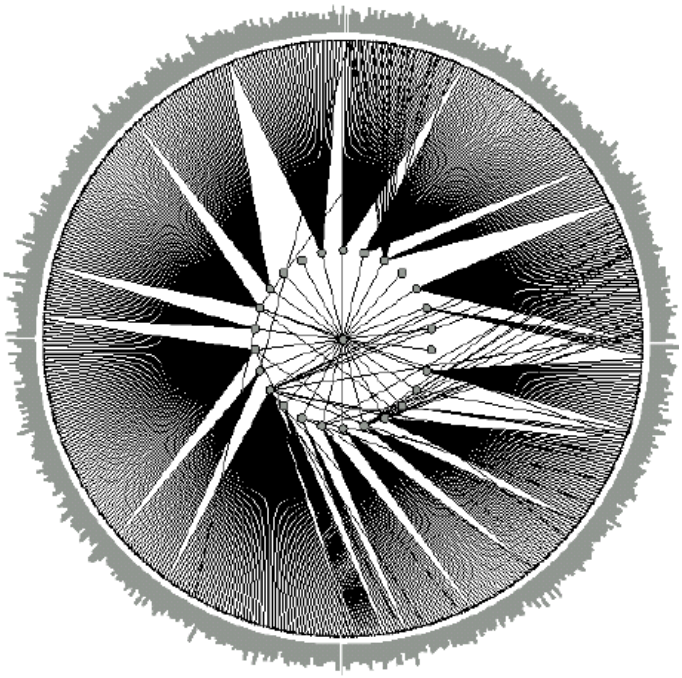


Chen & Paul (2001). "Visualizing a knowledge domain's intellectual structure." *Computer*, 34(3), 65-71.

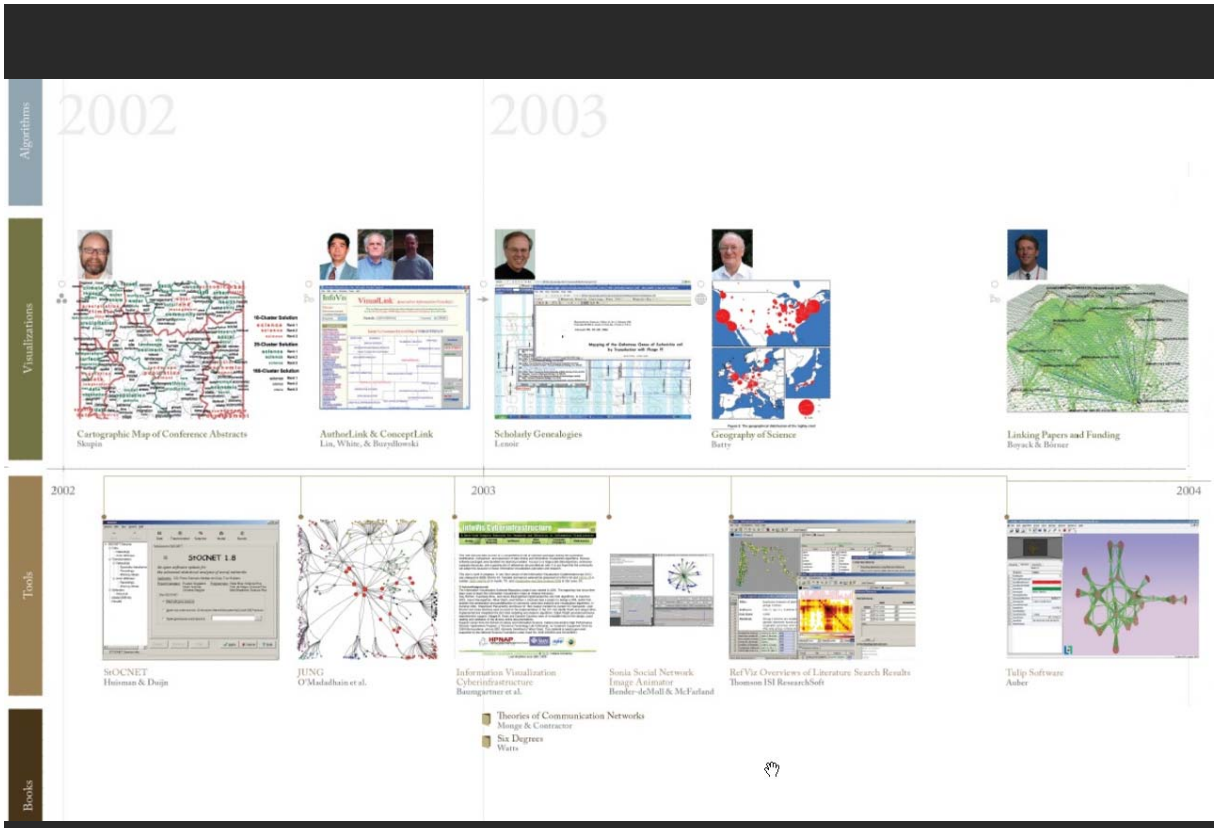
Newman, 2001

Physicist bringing new tools to the problem

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



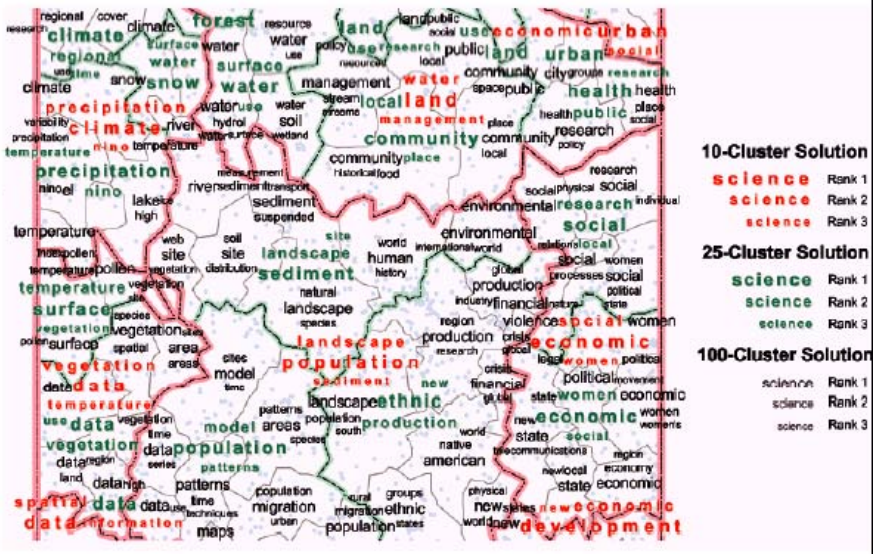
Newman, M.E.J. (2001). "Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality." *Physical Review E*, 64, paper # 016132.



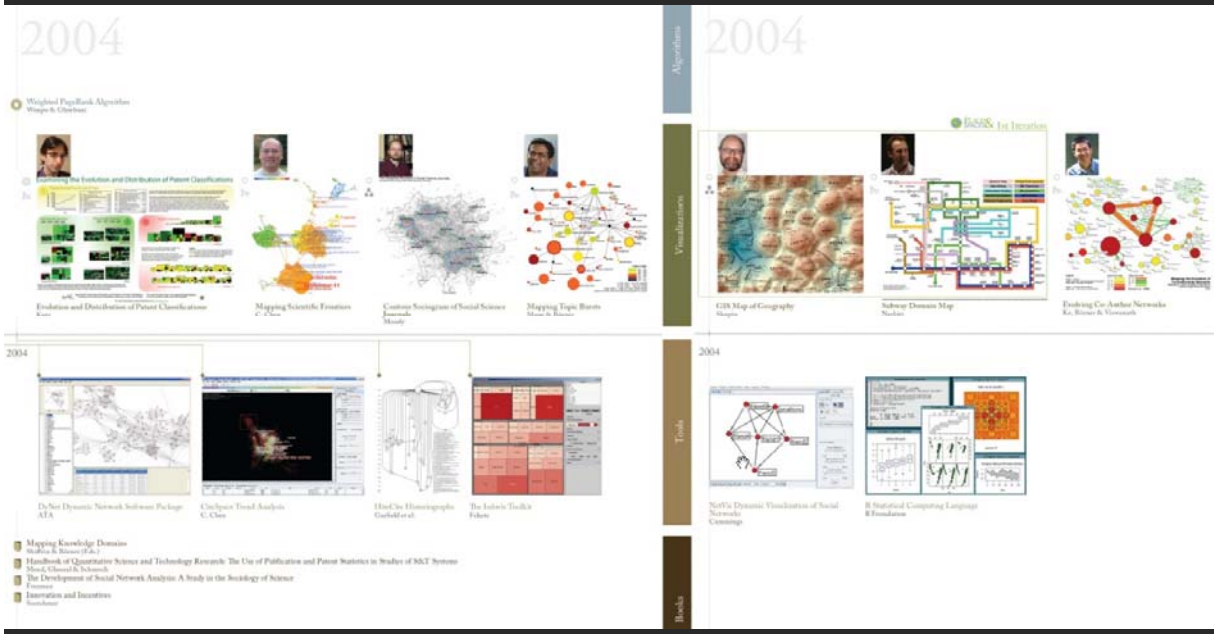
Skupin, 2002

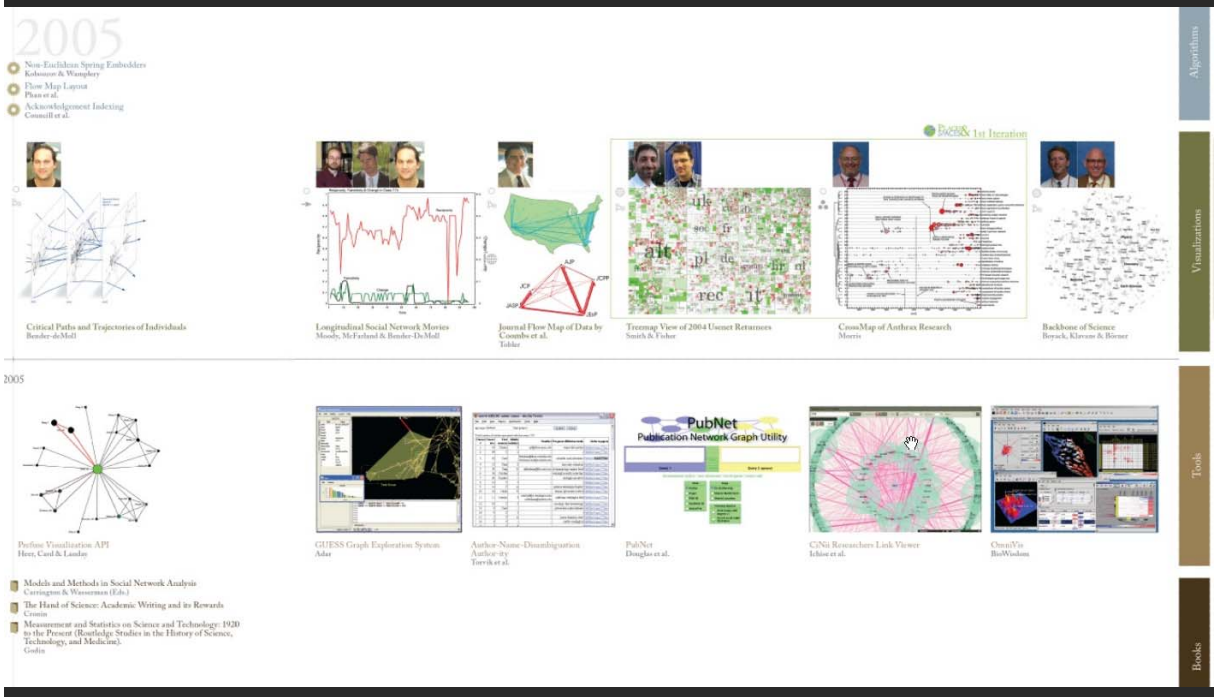
Applied self-organizing map to conference abstracts

Applied cartographic principles and techniques to the resulting SOM

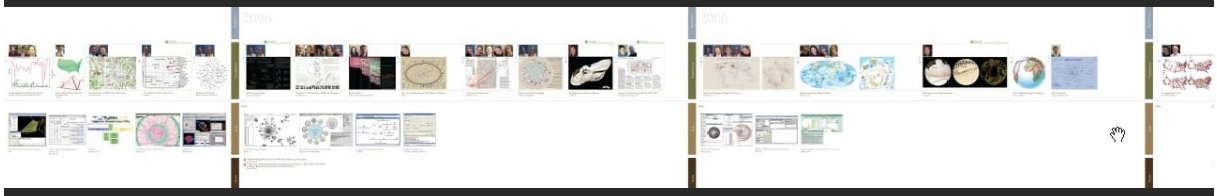


Skupin, A. (2002). "A cartographic approach to visualizing conference abstracts." *IEEE Computer Graphics and Applications*, 22, 50-58.





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*.

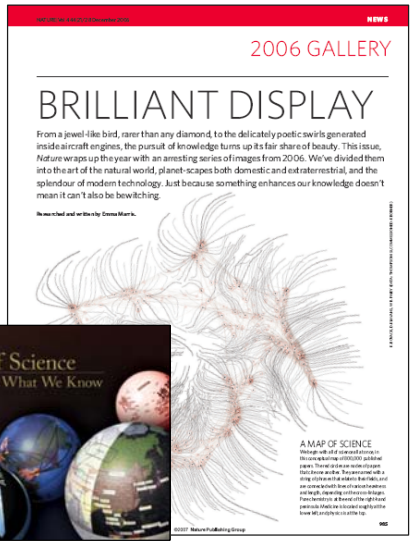
Computational Scientometrics References

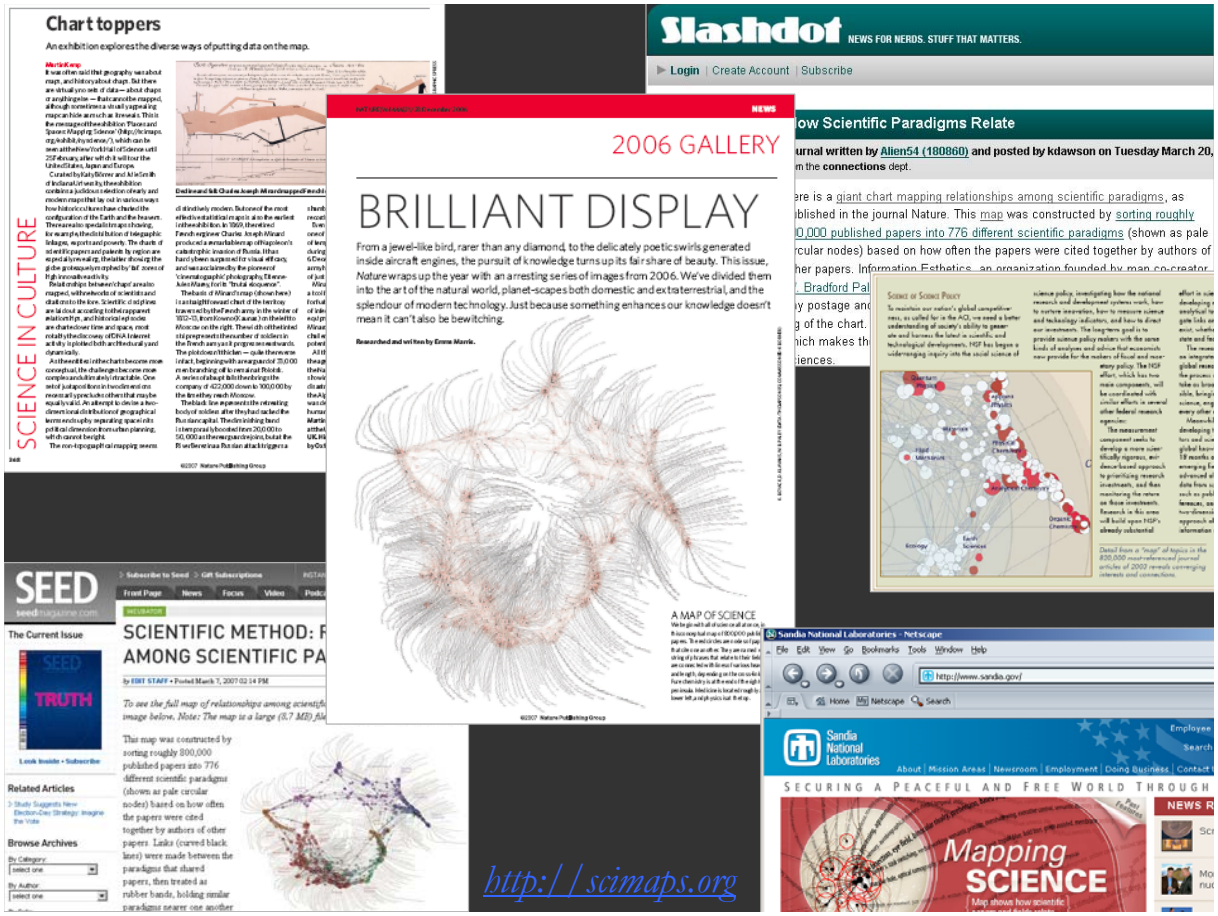
Börner, Katy, Chen, Chaomei, and Boyack, Kevin. (2003). **Visualizing Knowledge Domains.** In Blaise Cronin (Ed.), *ARIST*, Medford, NJ: Information Today, Inc./American Society for Information Science and Technology, Volume 37, Chapter 5, pp. 179-255. http://ivl.slis.indiana.edu/km/pub/2003-borner-arist.pdf

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/

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





[#01] Science of Science Research

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

Type of Analysis vs. Scale of Level of Analysis

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

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.

Type of Analysis vs. Scale of Level of Analysis

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

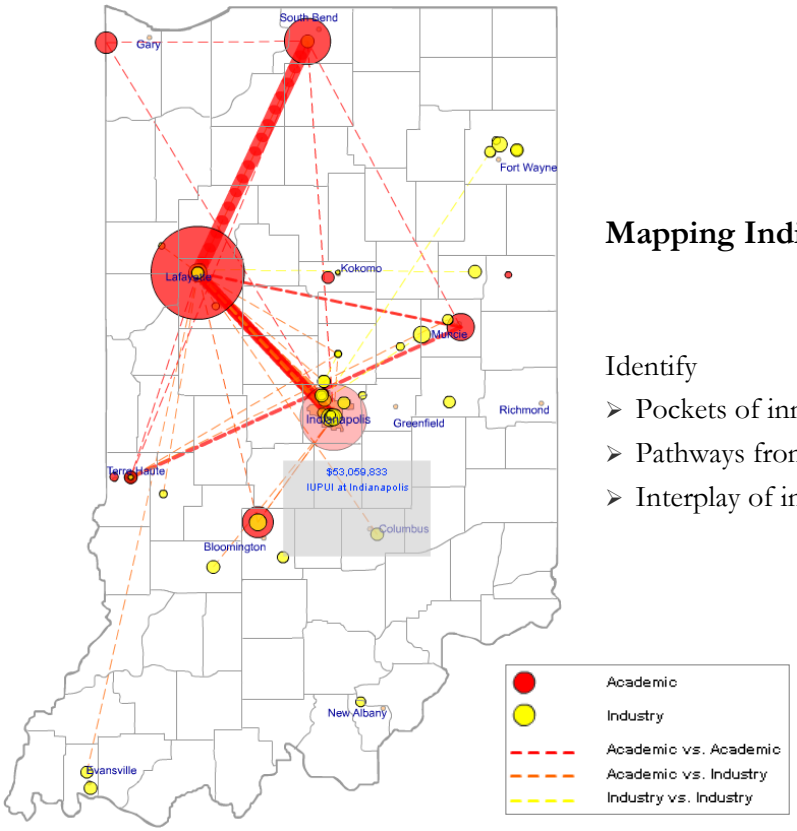
Type of Analysis vs. Scale of Level of Analysis

	Micro/Individual ***(1-100 records)***	***Meso/Local*** ***(101–10,000 records)***	***Macro/Global*** ***(10,000 < records)***
Statistical Analysis/Profiling	Individual person and	Larger labs, centers, search tes	All of NS SA, all of sci
Temporal Analysis (When)		bursts PNAS	113 Years of P Research
Geospatial Analysis (Where)			PNAS
Topical Analysis (What)	s.	ows in research	VxOrd/Topic NIH funding
Network Analysis (With Whom?)	NSF vork of one	k	NIH's cy

Common analysis types are
- Temporal
- Geospatial
- Topical
- Network

or combinations thereof.

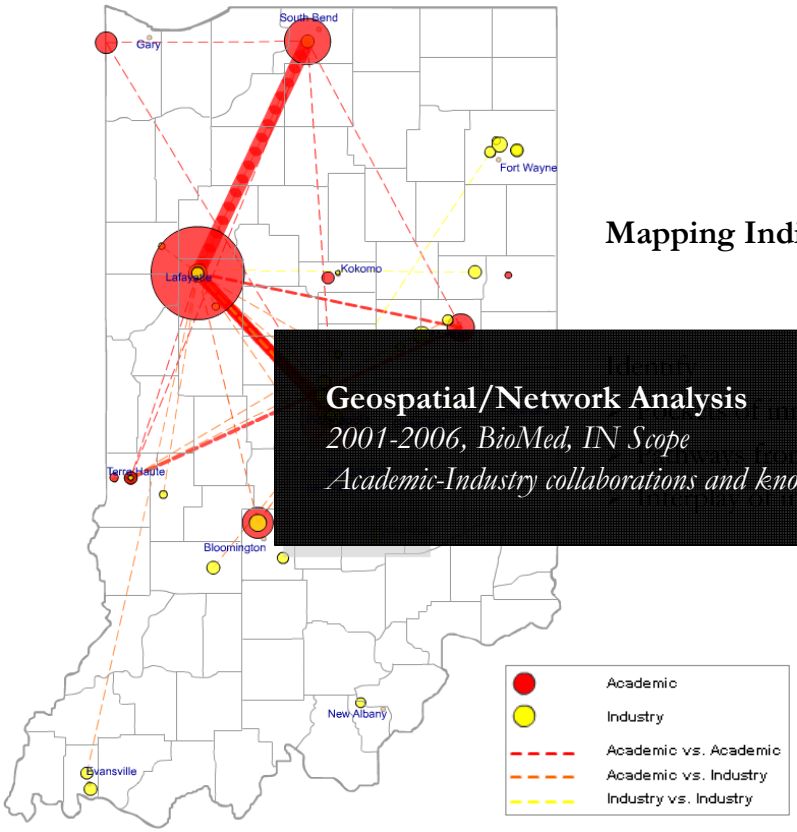
The data used determines the scope of the analysis.
We also list the main analysis goal.



Mapping Indiana's Intellectual Space

Identify

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



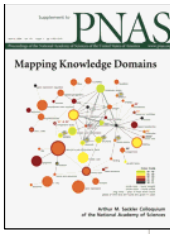
Mapping Indiana's Intellectual Space

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.

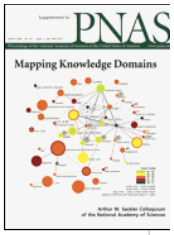
Mane & Börner. (2004) PNAS, 101(Suppl. 1): 5287-5290.



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.

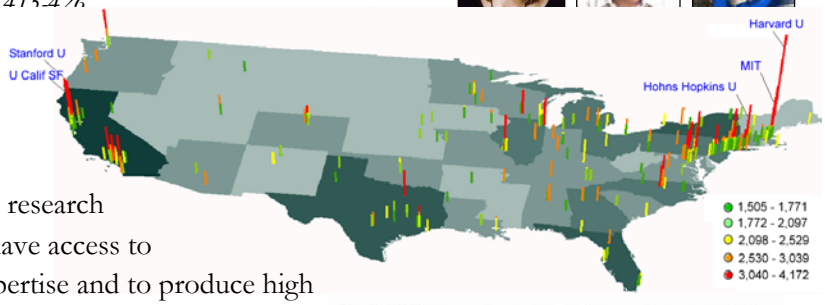


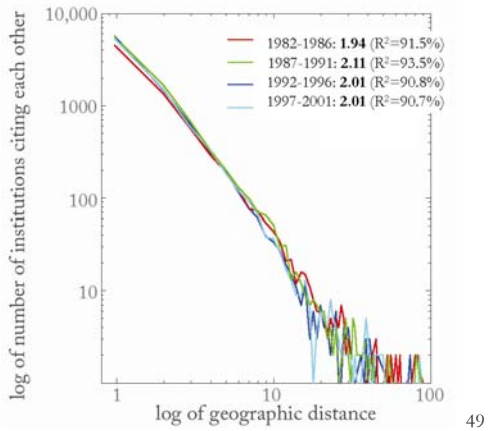
Spatio-Temporal Information Production and Consumption of Major U.S. Research Institutions

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 instructions?





Contributions:

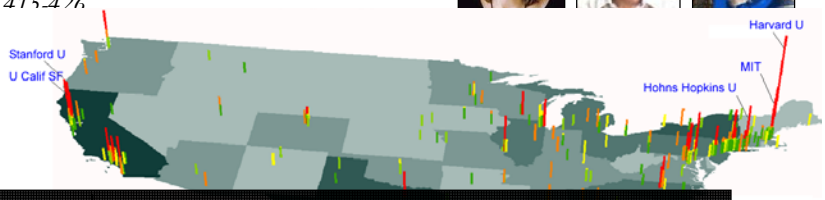
- Answer to Qs 1 + 2 is YES.
- Answer to Qs 3 is NO.
- Novel approach to analyzing the dual role of institutions as information producers and consumers and to study and visualize the diffusion of information among them.

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

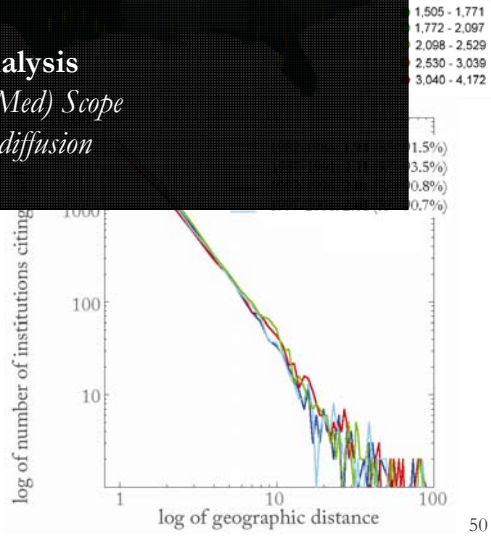
Börner, Katy, Penumarthy, Shashikant, Meiss, Mark and Ke, Weimao. (2006) Mapping the Diffusion of Scholarly Knowledge Among Major U.S. Research Institutions. Scientometrics. 68(3), pp. 415-426.

Research questions:

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



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



Contributions:

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

Research Collaborations by the Chinese Academy of Sciences

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

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.

Research Collaborations by the Chinese Academy of Sciences

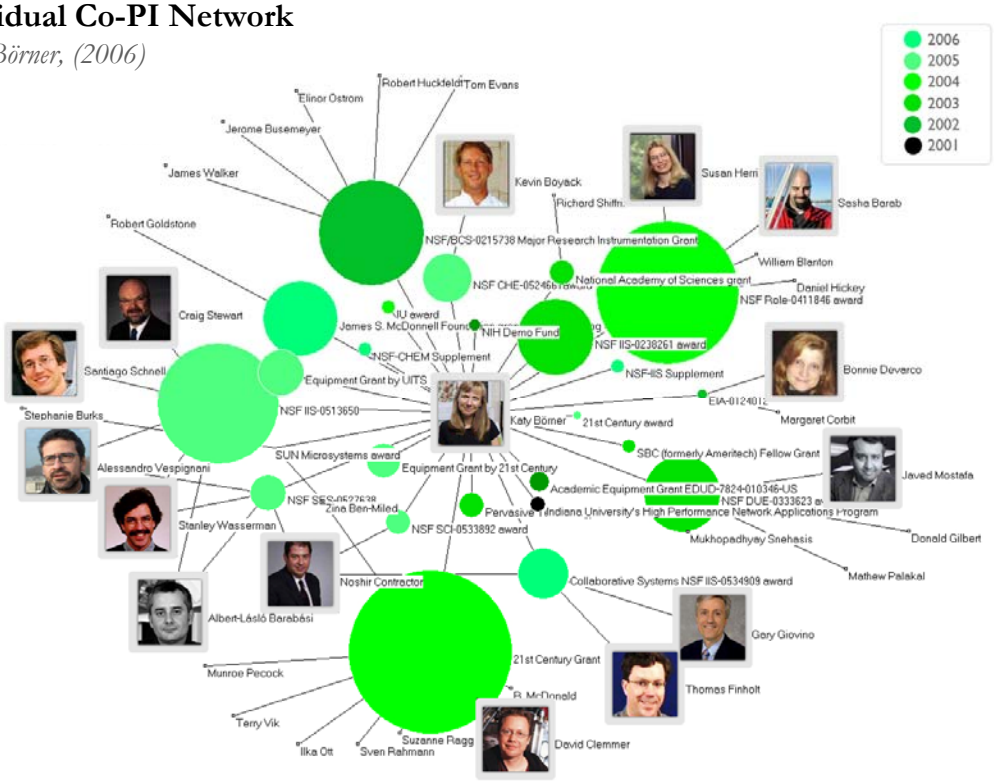
By Weixia (Bonnie) Huang, Russell J. Duhon, Elisha F. Hardy, Katy Börner, Indiana University, USA

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

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

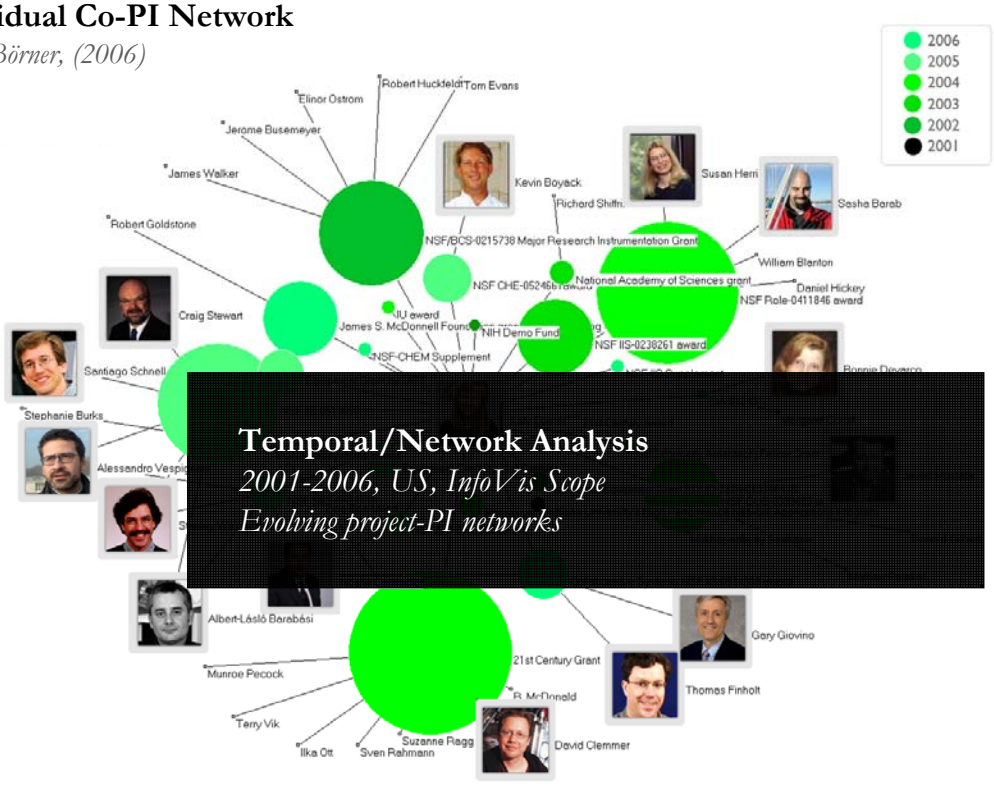
Individual Co-PI Network

Ke & Börner, (2006)



Individual Co-PI Network

Ke & Börner, (2006)



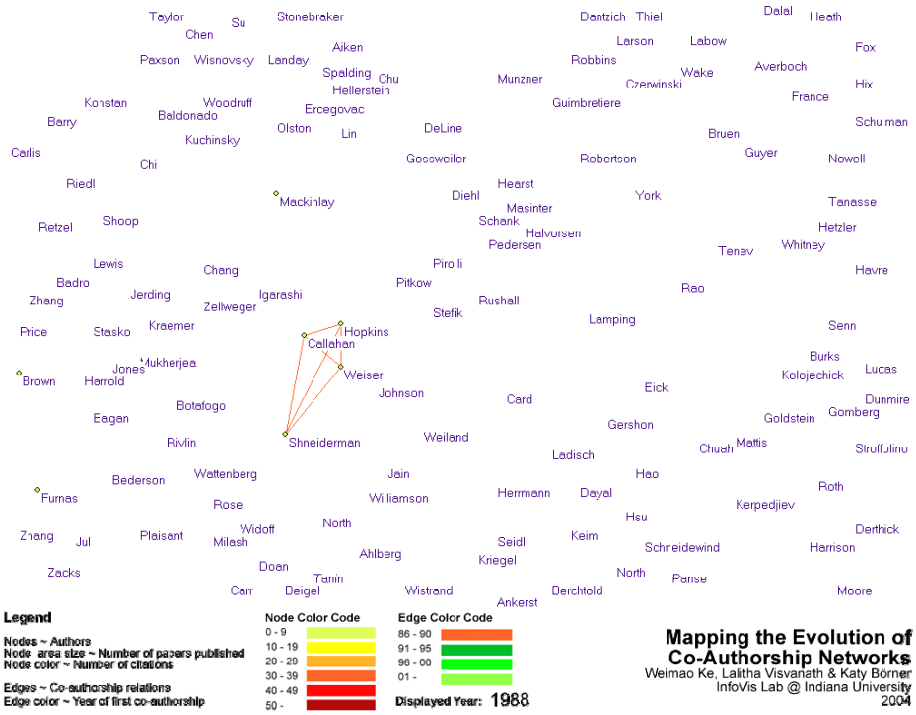
Temporal/Network Analysis

2001-2006, US, InfoVis Scope

Evolving project-PI networks

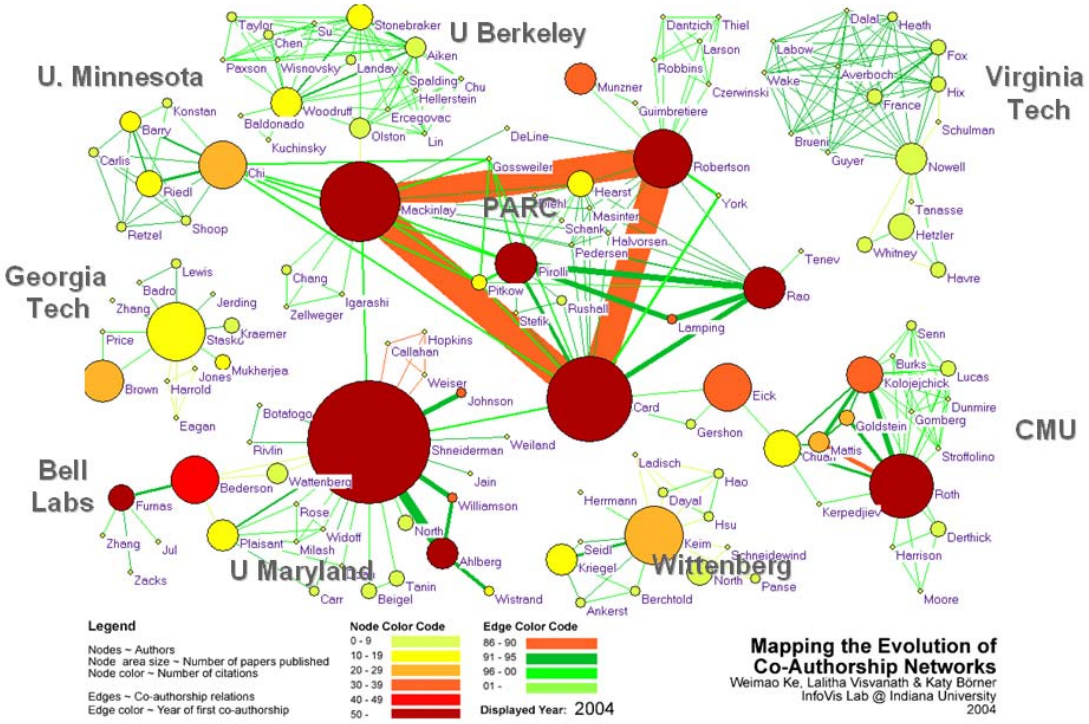
Mapping the Evolution of Co-Authorship Networks

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



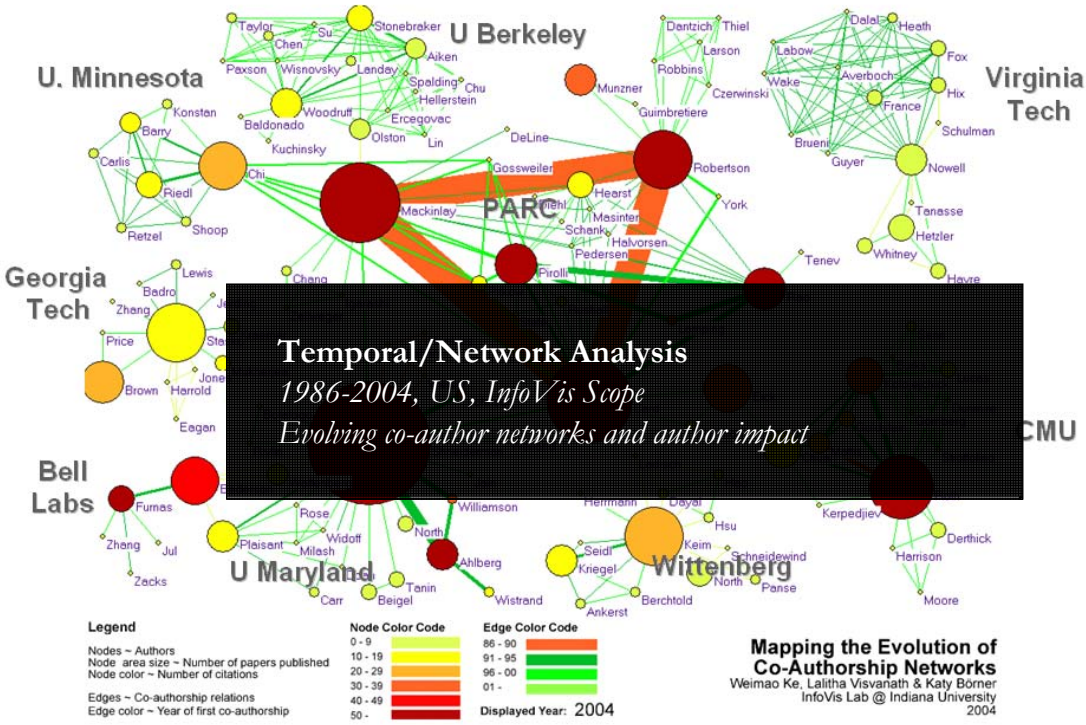
Mapping the Evolution of Co-Authorship Networks

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



Mapping the Evolution of Co-Authorship Networks

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



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

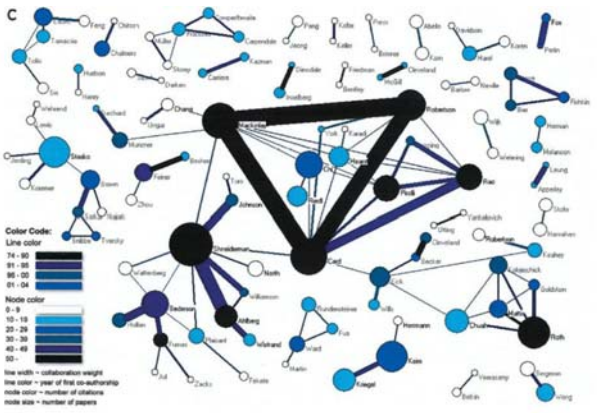
Börner, Dall'Asta, Ke & Vespignani (2005) Complexity, 10(4):58-67.

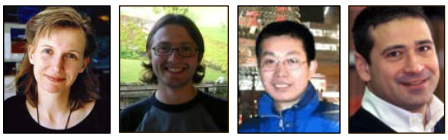
Research question:

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

Contributions:

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





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

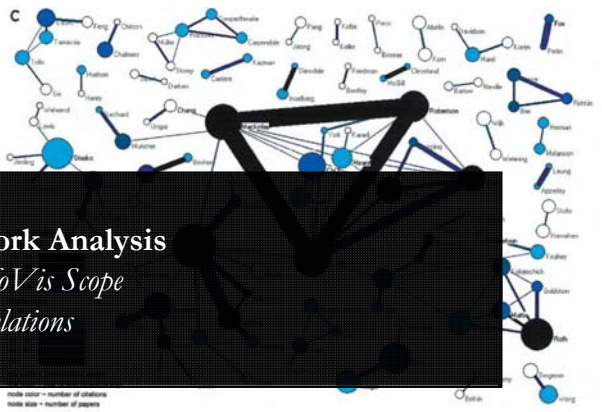
Börner, Dall'Asta, Ke & Vespignani (2005) Complexity, 10(4):58-67.

Research question:

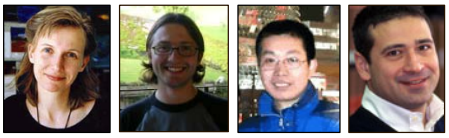
- Is science driven by prolific single experts or by high-impact co-authorship teams?

Contributions:

- New approach to allocate citational credit.
- Novel weighted gr
- Visualization of th
  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.



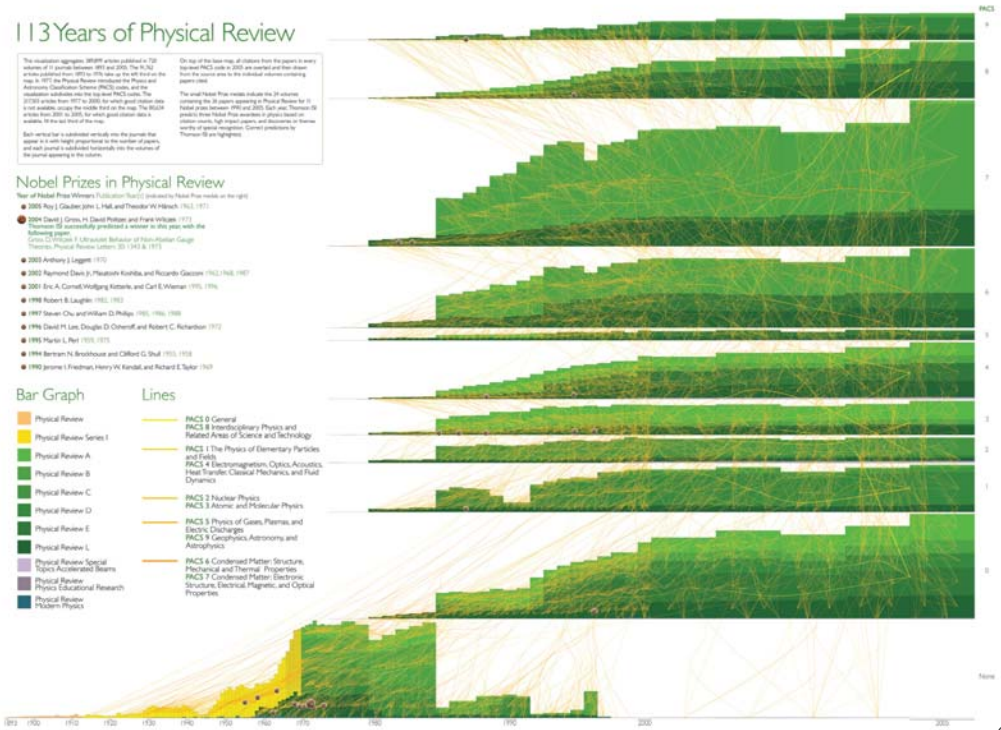
Temporal/Network Analysis
1986-2004, US, InfoVis Scope
Impact of co-author relations



113 Years of Physical Review

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

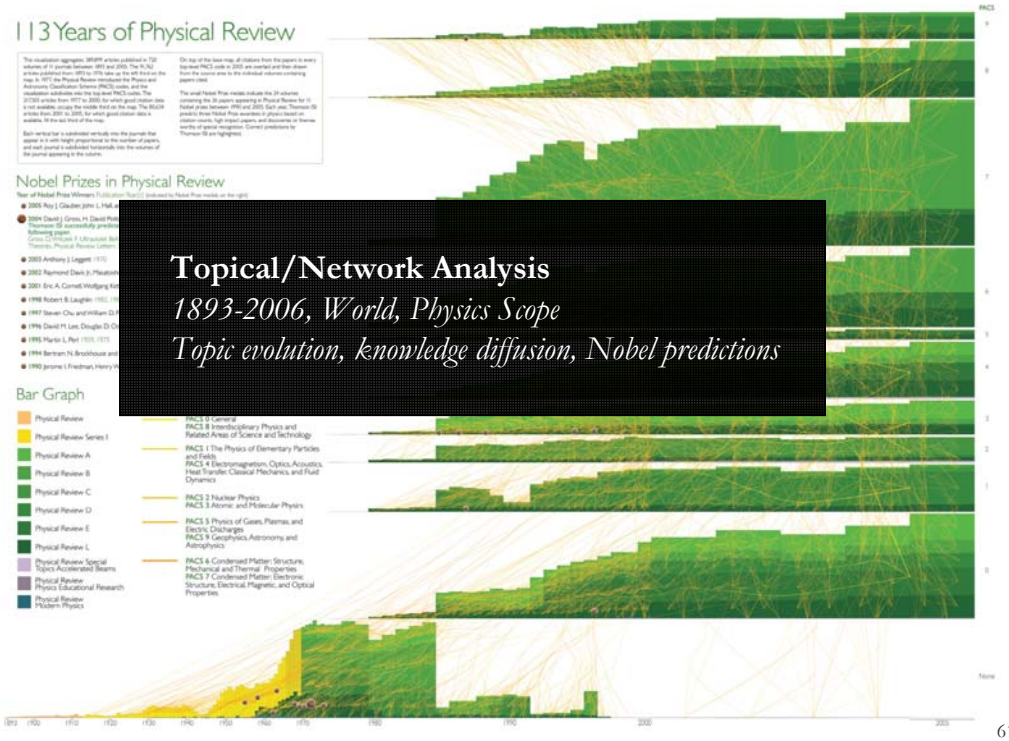
Bruce W. Herr II and Russell Duhon (Data Mining & Visualization), Elisha F. Hardy (Graphic Design), Shashikant Penumarthy (Data Preparation) and Katy Börner (Concept)



113 Years of Physical Review

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

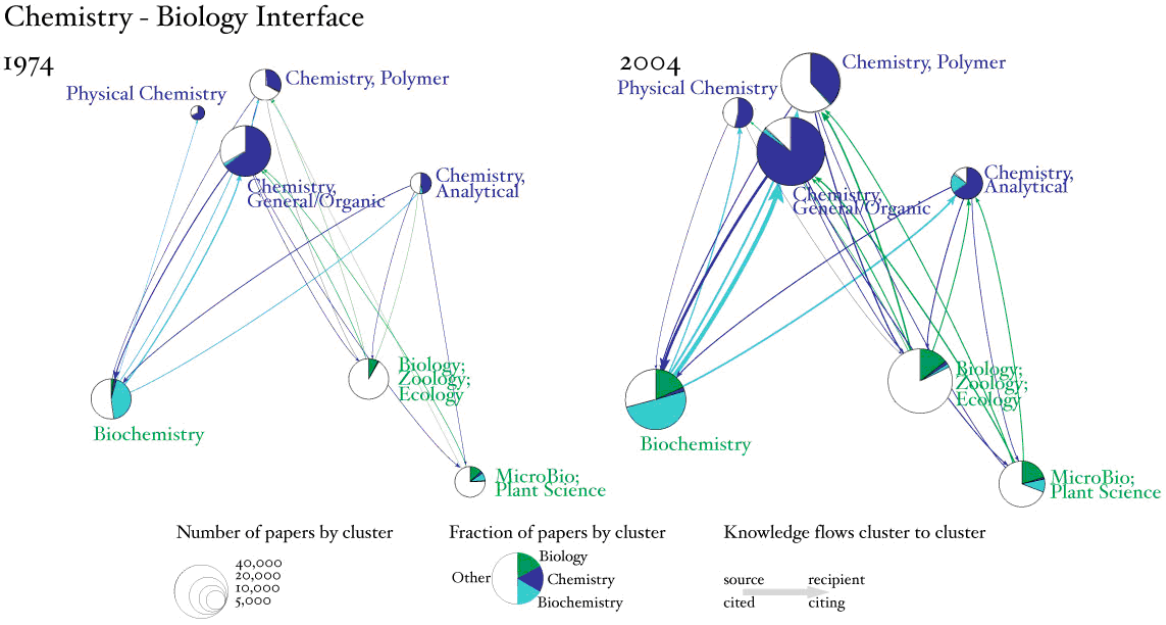
Bruce W. Herr II and Russell Duhon (Data Mining & Visualization), Elisha F. Hardy (Graphic Design), Shashikant Penumarthy (Data Preparation) and Katy Börner (Concept)



Topical/Network Analysis
1893-2006, World, Physics Scope
Topic evolution, knowledge diffusion, Nobel predictions

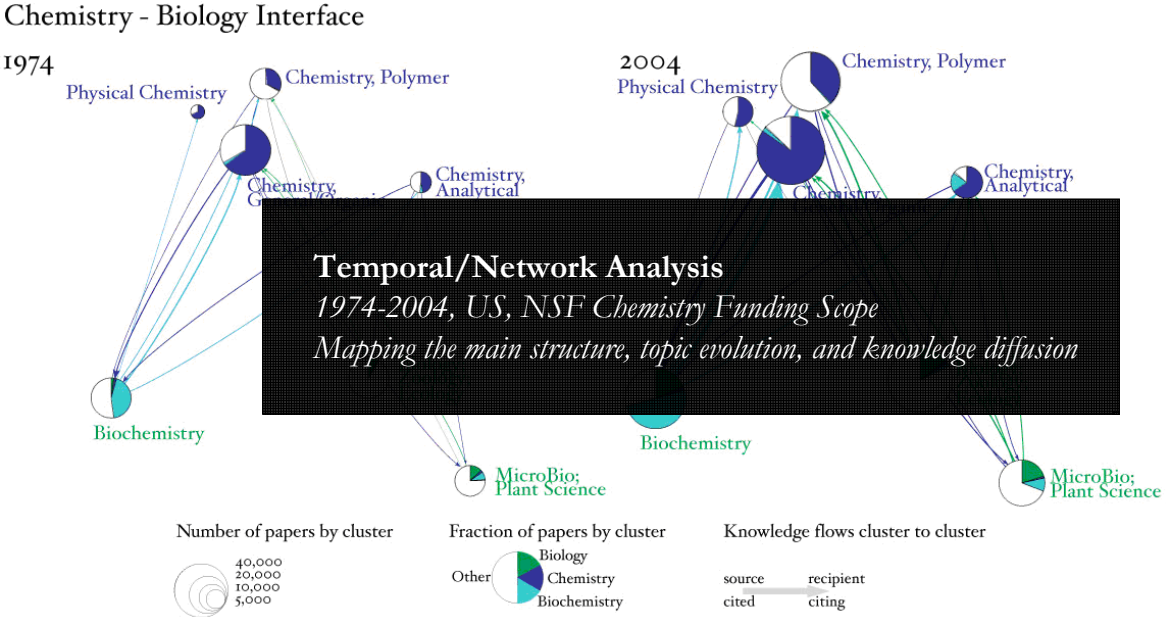
Topical Composition and Knowledge Flow Patterns in Chemistry Research for 1974 and 2004

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



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

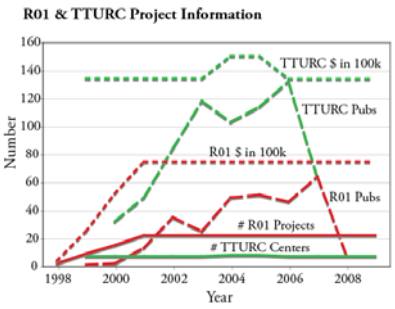
Kevin W. Boyack, Katy Börner, & Richard Klavans (2007)

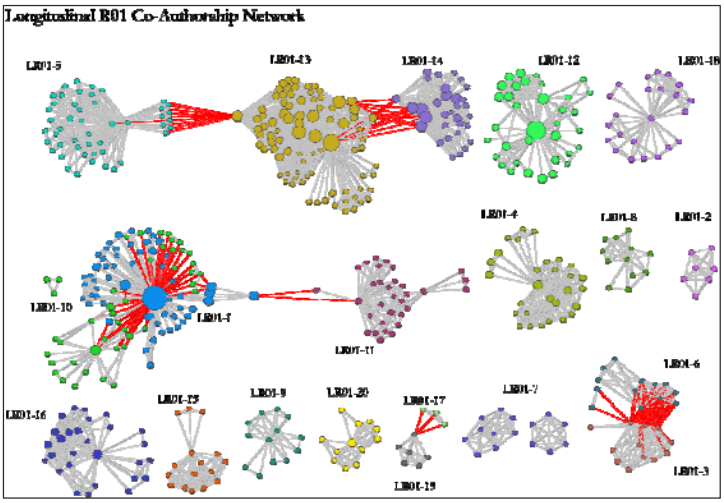


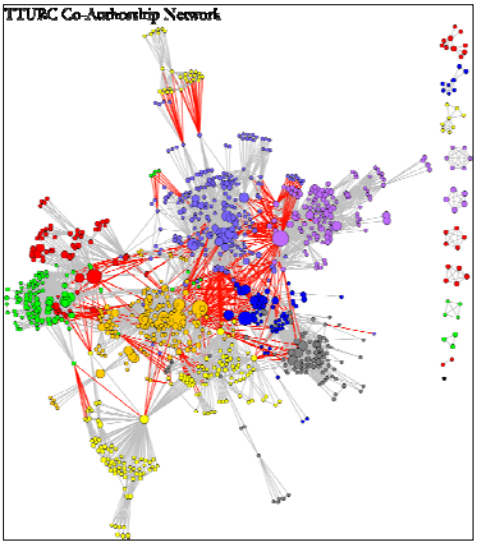
Mapping Transdisciplinary Tobacco Use Research Centers Publications

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

Zoss & Börner, forthcoming.



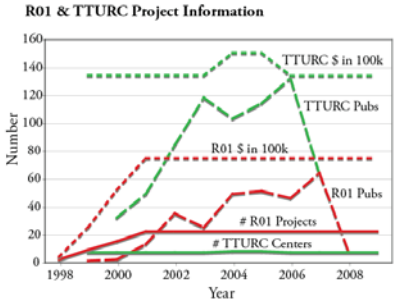


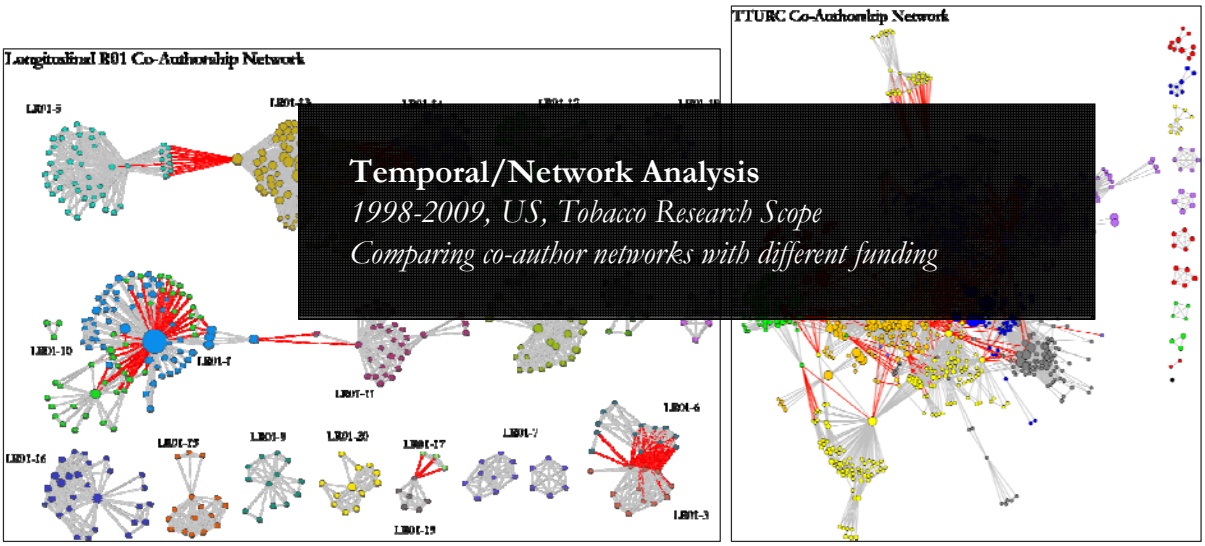


Mapping Transdisciplinary Tobacco Use Research Centers Publications

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

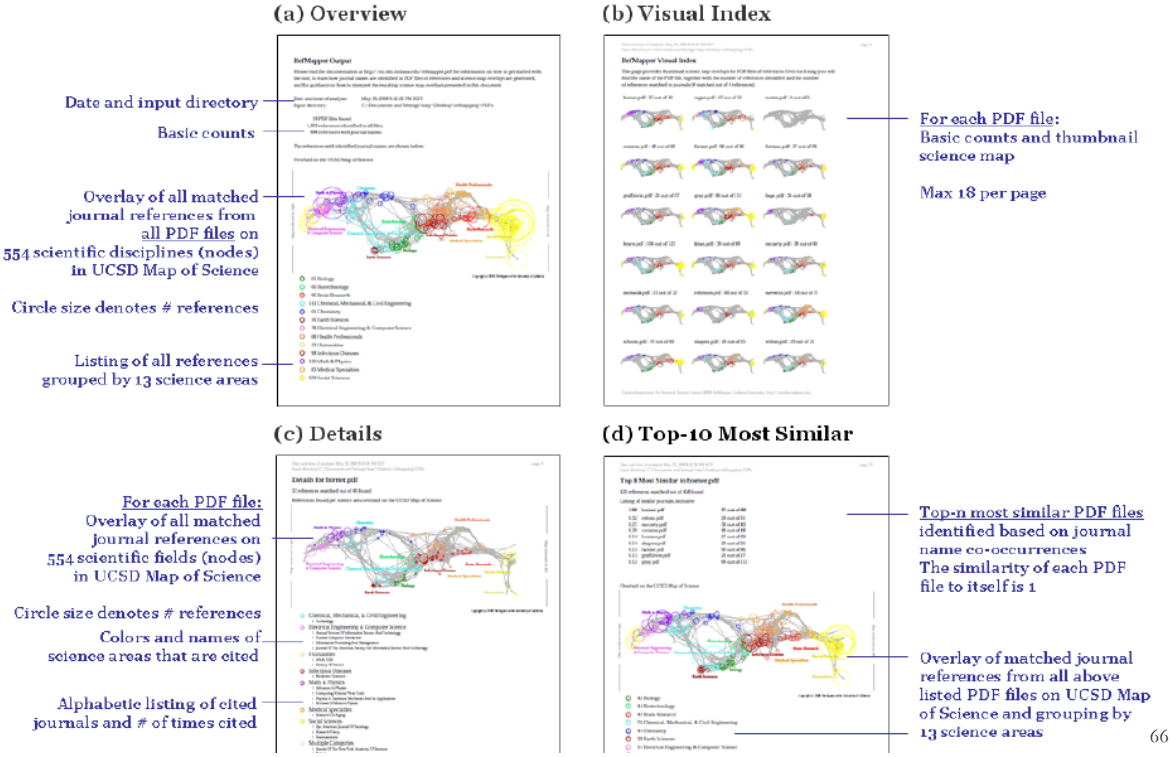
Zoss & Börner, forthcoming.





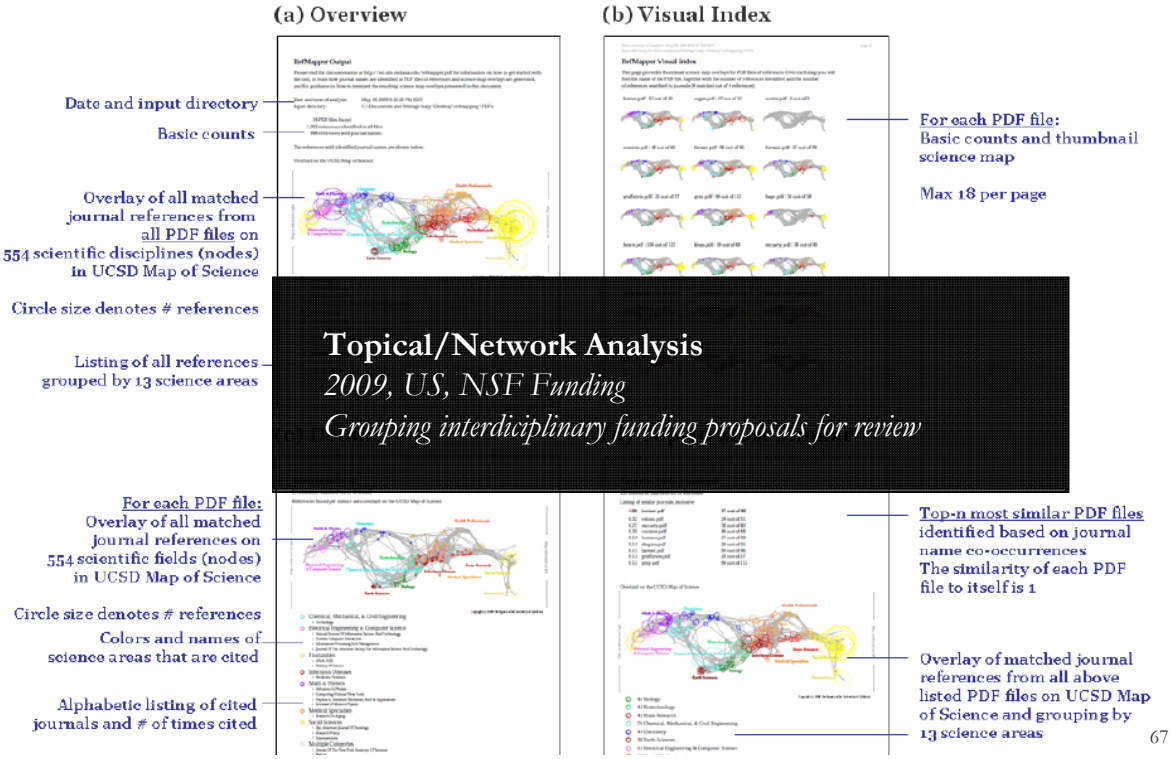
Reference Mapper

Duhon & Börner, forthcoming.



Reference Mapper

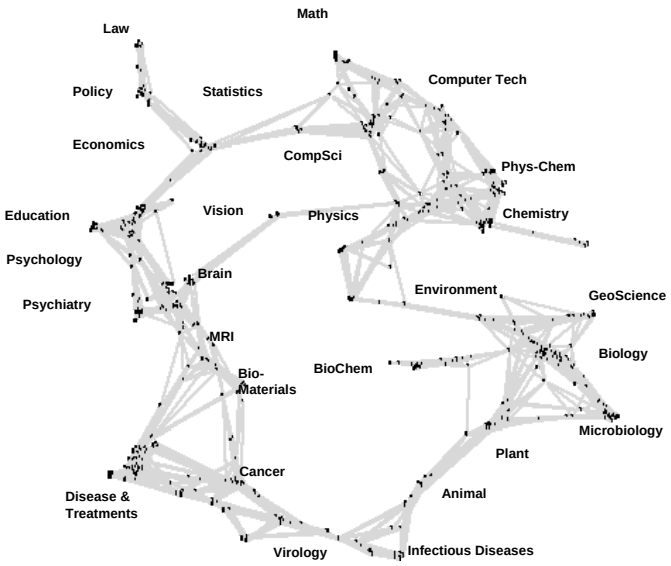
Duhon & Börner, forthcoming.



Latest 'Base Map' of Science

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

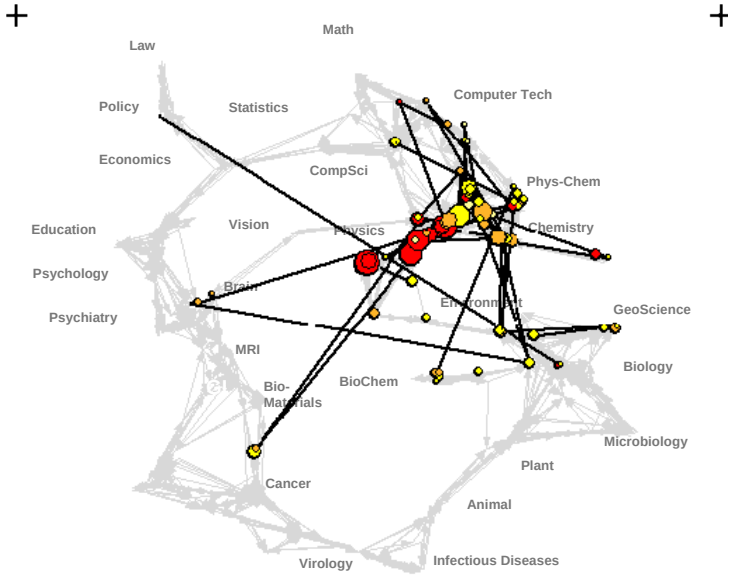
- Uses combined SCI/SSCI from 2002
  - 1.07M papers, 24.5M references, 7,300 journals
  - Bibliographic coupling of papers, aggregated to journals
- Initial ordination and clustering of journals gave 671 clusters
- Coupling counts were reaggregated at the journal cluster level to calculate the
  - (x,y)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



Science map applications: Identifying core competency

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

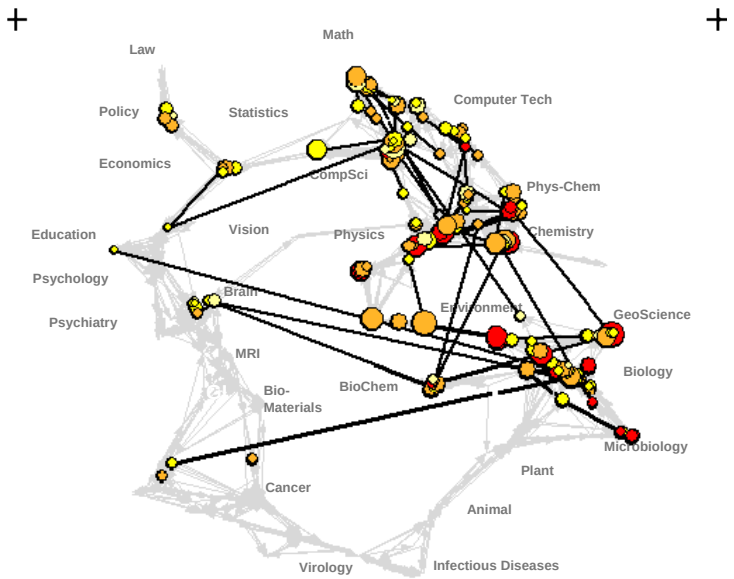
Funding patterns of the US Department of Energy (DOE)



Science map applications: Identifying core competency

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

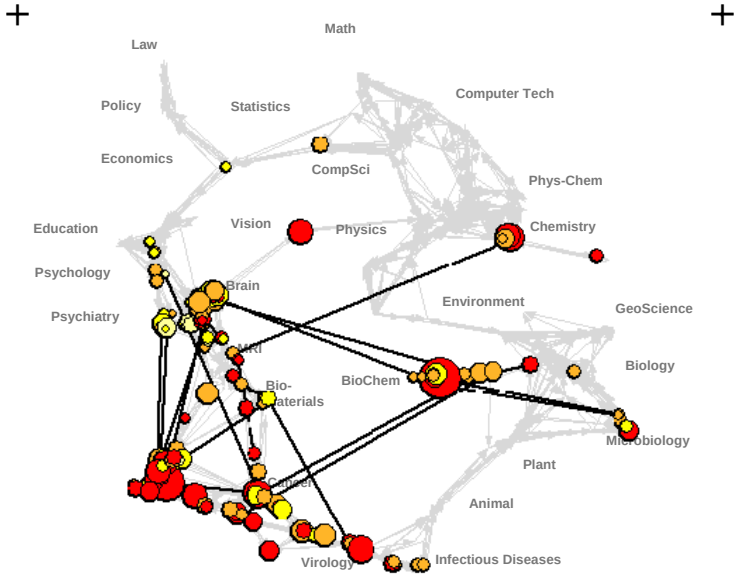
Funding Patterns of the National Science Foundation (NSF)



Science map applications: Identifying core competency

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

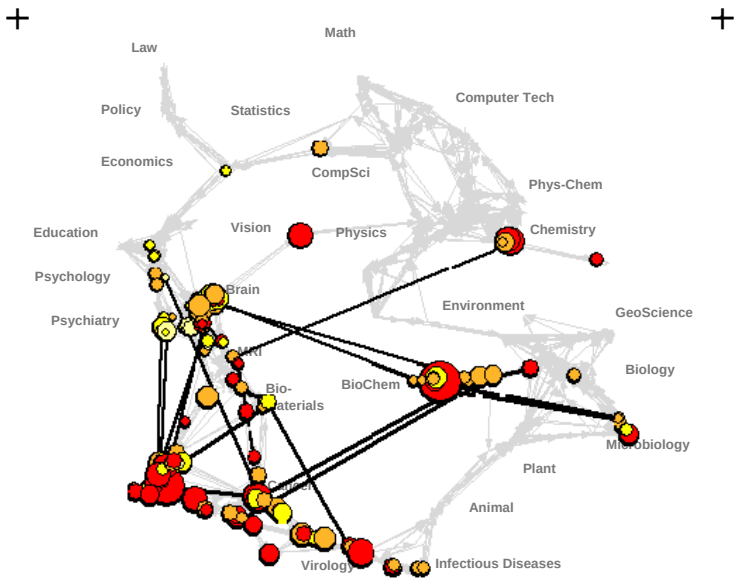
Funding Patterns of the National Institutes of Health (NIH)



Science map applications: Identifying core competency

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

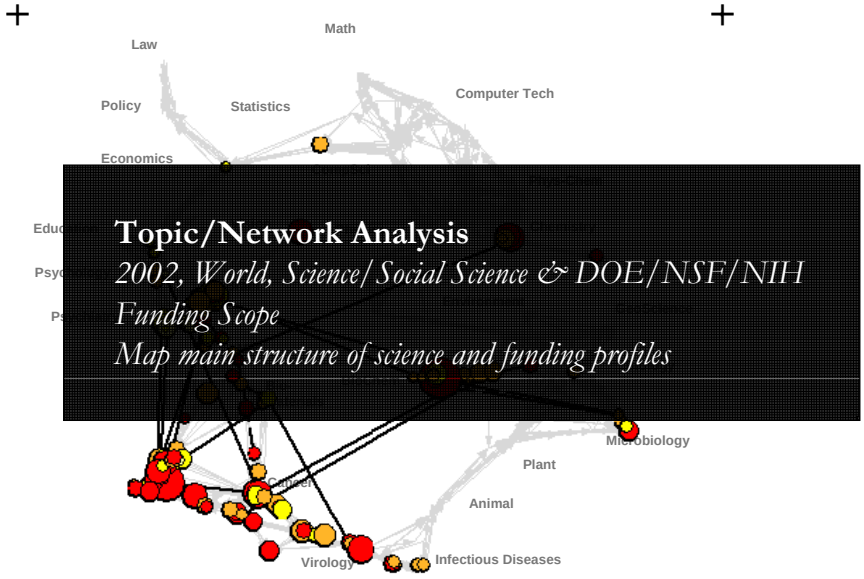
Funding Patterns of the National Institutes of Health (NIH)



Science map applications: Identifying core competency

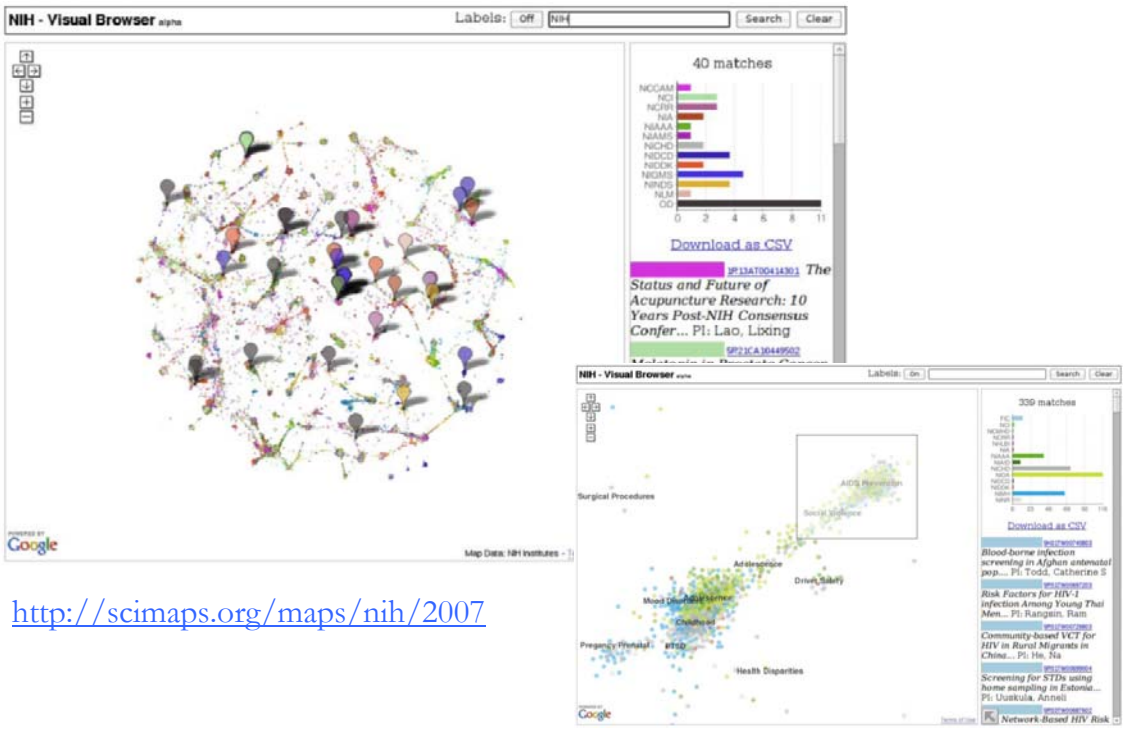
Kevin W. Boyack, Katy Börner, & Richard Klavans (2007).

Funding Patterns of the National Institutes of Health (NIH)



Interactive Science Map of NIH Funding

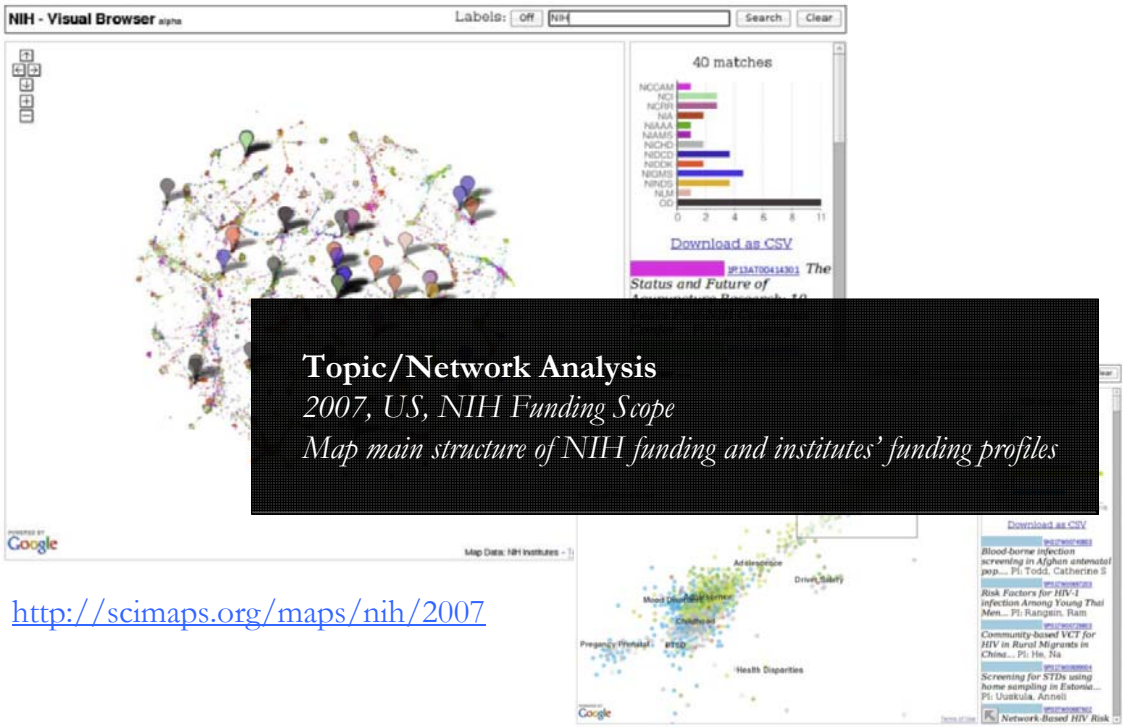
Herr II, Bruce W., Talley, Edmund M, Burns, Gully APC, Newman, David & La Rowe, Gavin. (2009).



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

Interactive Science Map of NIH Funding

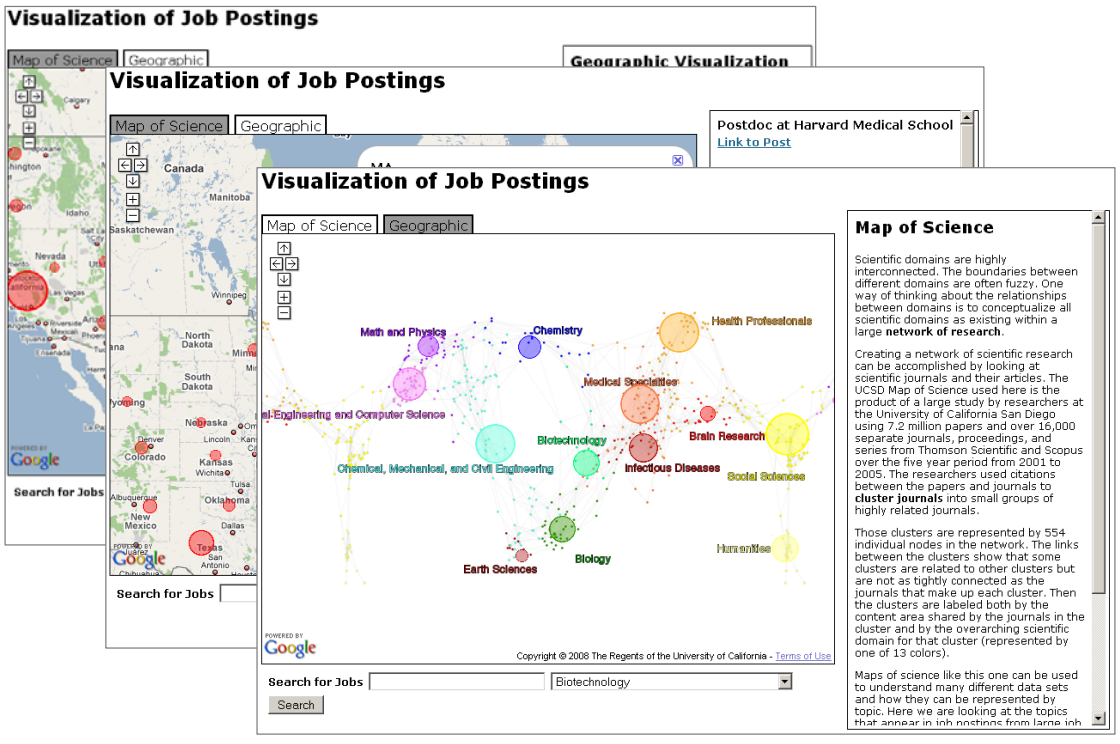
Herr II, Bruce W., Talley, Edmund M, Burns, Gully APC, Newman, David & La Rowe, Gavin. (2009).



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

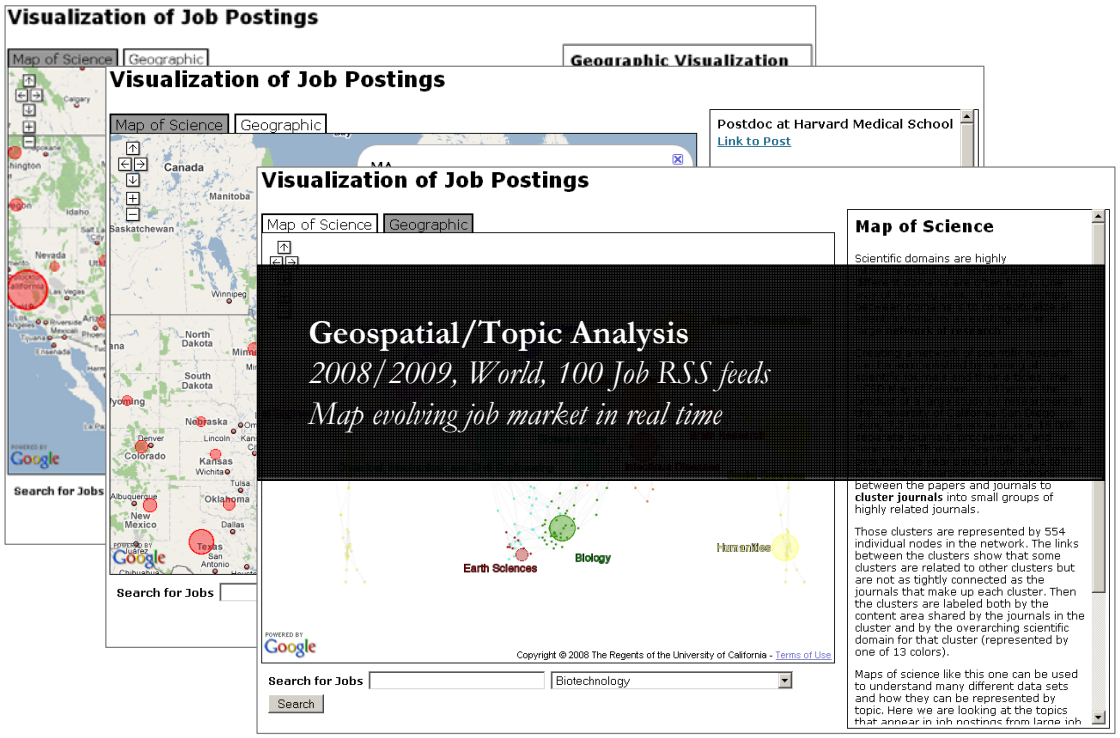
Interactive World and Science Map of S&T Jobs

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



Interactive World and Science Map of S&T Jobs

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



[#01] Science of Science Research

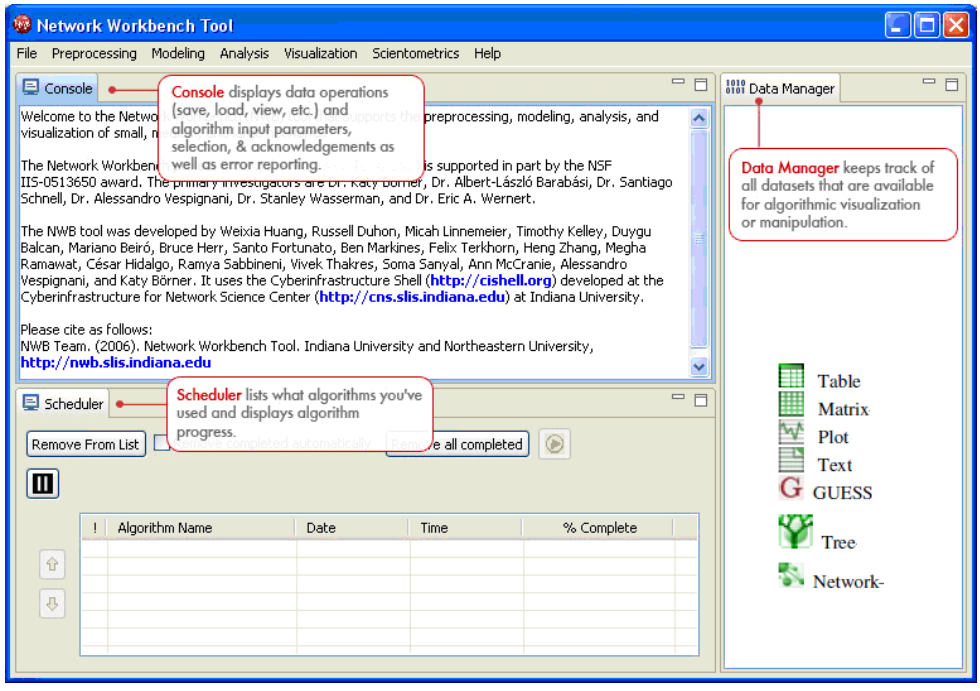
- Brief History (Atlas timeline)
- Micro, Meso, Macro Studies
- **Workflow Design**
- Sample Studies / Mapping Science Exhibit
- Validation
- Promising Research Directions

General Process of Analyzing and Mapping Science

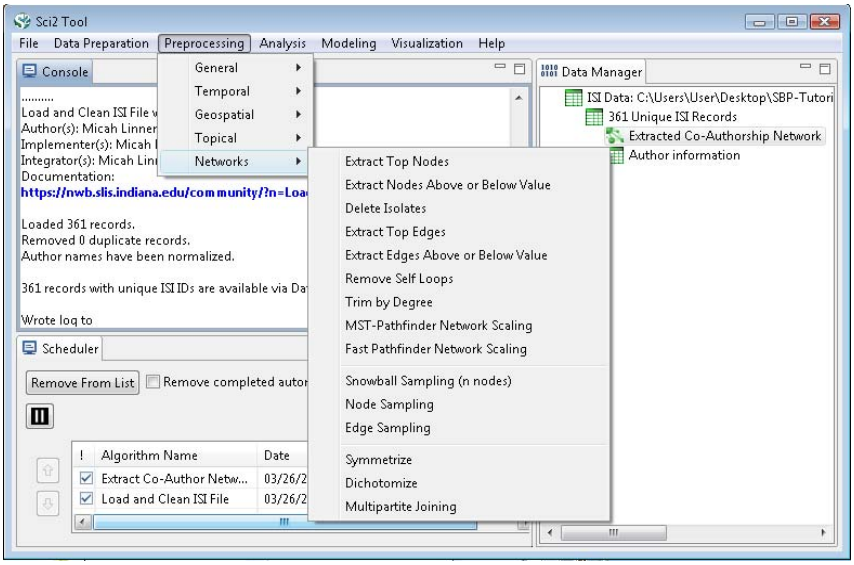
(1) Data Extraction	(2) Unit of Analysis	(3) Measures	Layout (often one code does both similarity and ordination steps)		(6) Display
			(4) Similarity	(5) Ordination	
Searches *WoS* *Scopus* *Google Scholar* *MEDLINE* *Patents* *Funding* **Broadening** *By citation* *By terms*	**Common Choices** *Journal*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

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



Science of Science (Sci²) Tool Interface



Acknowledgments

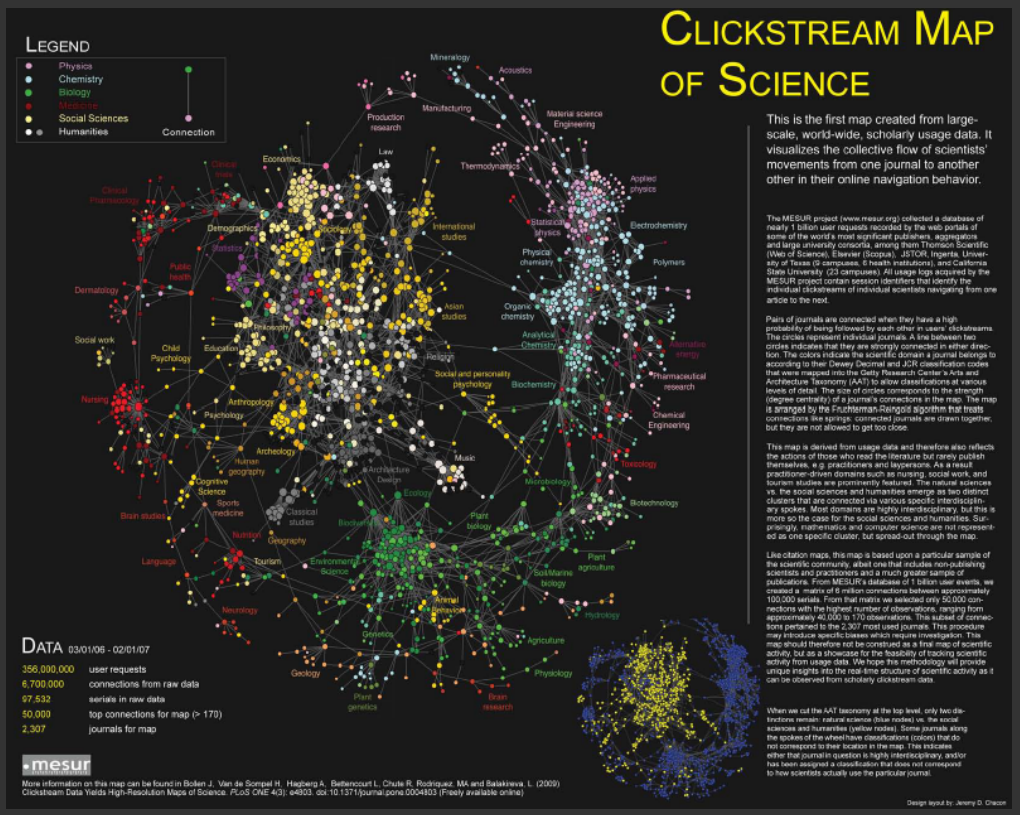
This work is supported in part by the Cyberinfrastructure for Network Science center and the School of Library and Information Science at Indiana University, the National Science Foundation under Grant No. SBE-0738111 and IIS-0513650, and the James S. McDonnell Foundation.

INDIANA UNIVERSITY SCHOOL OF LIBRARY AND INFORMATION SCIENCE

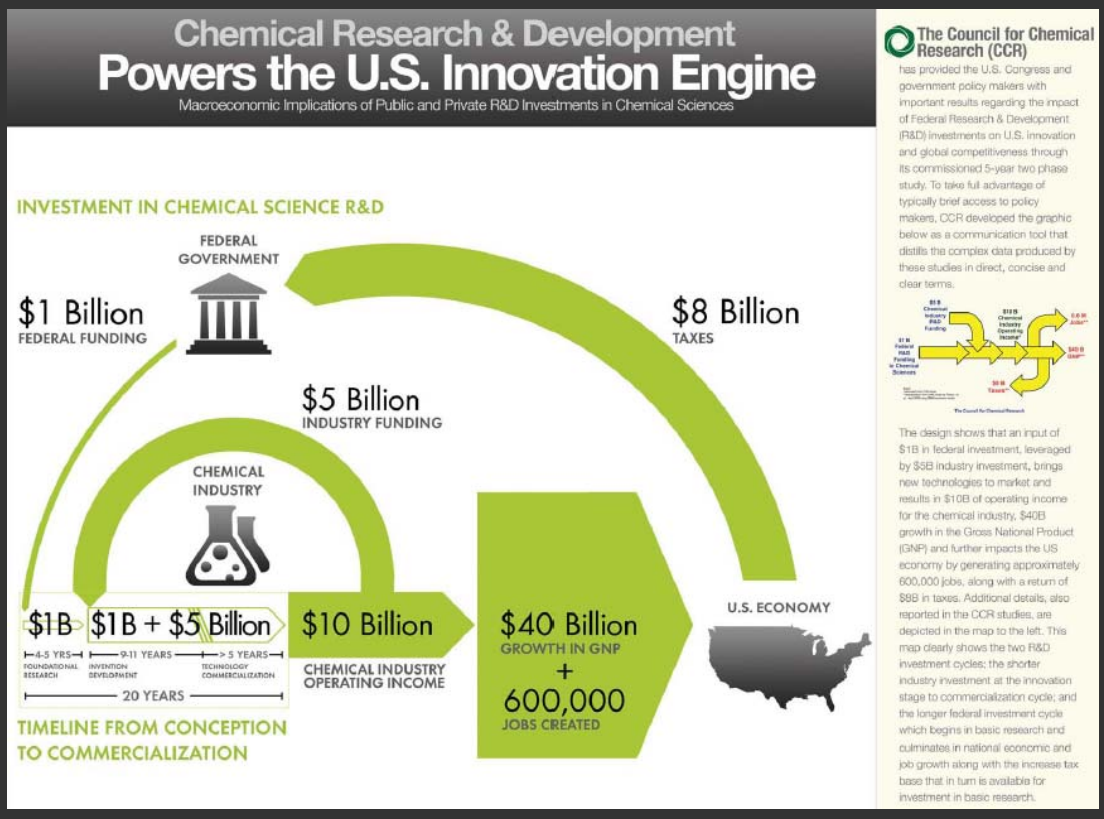
James S. McDonnell Foundation

[#01] Science of Science Research

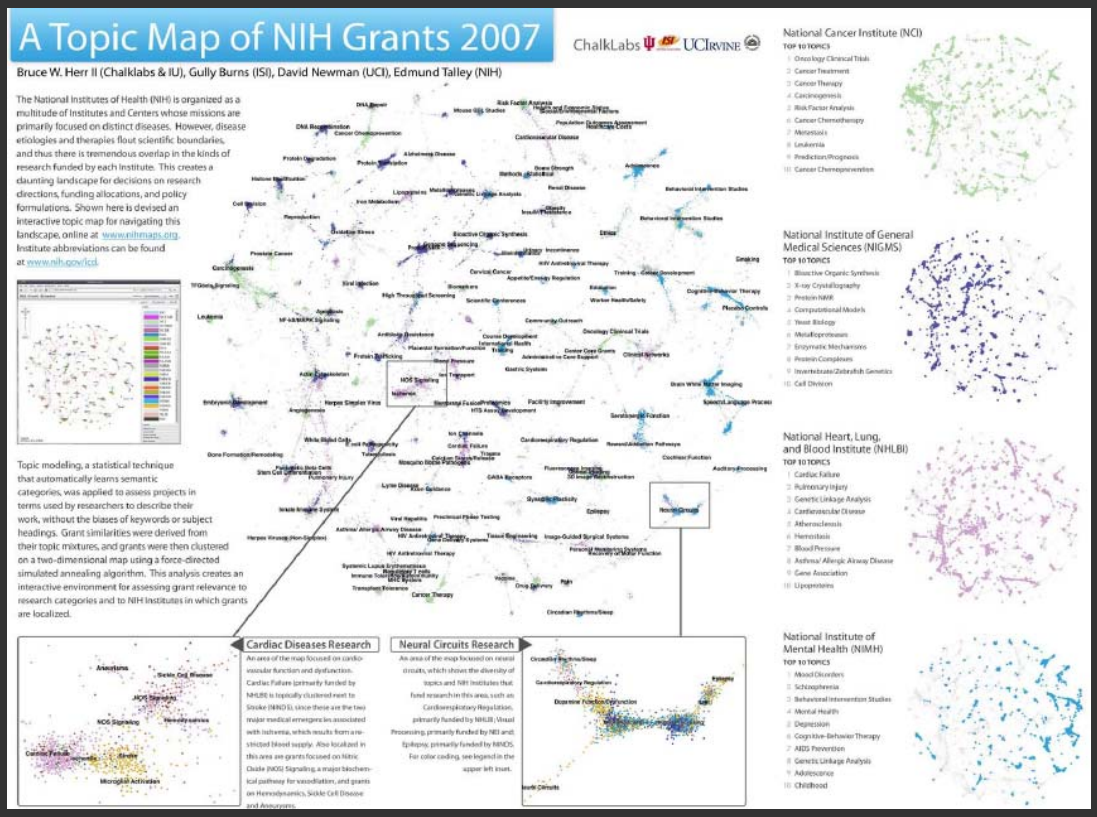
- Brief History (Atlas timeline)
- Micro, Meso, Macro Studies
- Workflow Design
- Sample Studies / Mapping Science Exhibit
- Validation
- Promising Research Directions



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



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

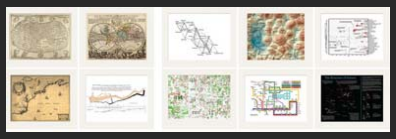


Herr II, Bruce W., Gully Burns, David Newman, Edmund Talley. 2007.
A Topic Map of NIH Grants 2007. Bloomington, IN.

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.

ORDER MAPS

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

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

[#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

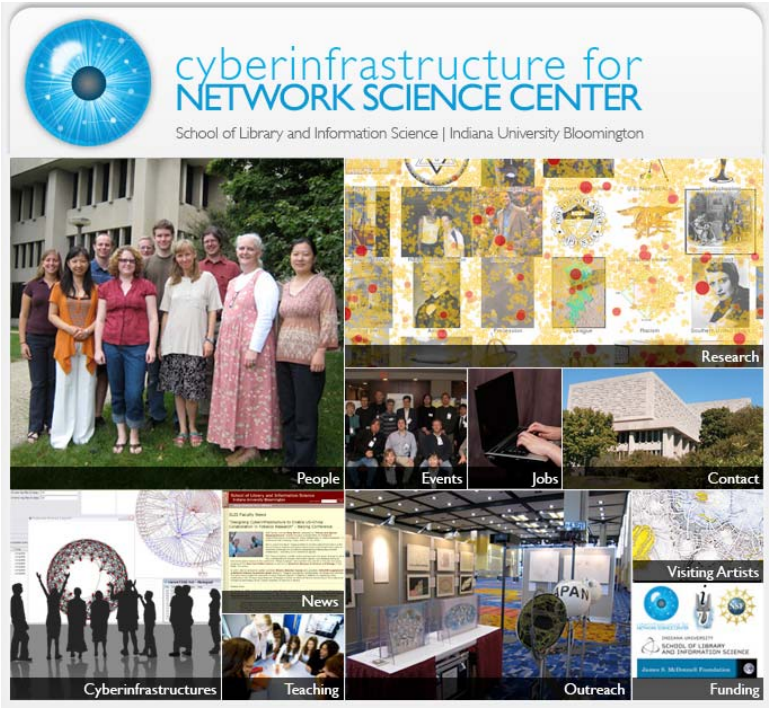
[#01] Science of Science Research

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

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



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