# Mapping Science ~ History and Future


**Dr. Katy Börner**
Cyberinfrastructure for Network Science Center, Director
Information Visualization Laboratory, Director
School of Library and Information Science
Indiana University, Bloomington, IN
[katy@indiana.edu](mailto:katy@indiana.edu)


**Several slides were taken from a talk by Kevin W. Boyack for the UCGIS Summer Meeting, June, 2009.**

*Visualization for Collective, Connective & Distributed Intelligence*
*Dynamic Knowledge Networks ~ Synthetic Minds*
*Stanford University, CA: August 12, 2009*

## Early Maps of the World VERSUS Early Maps of Science



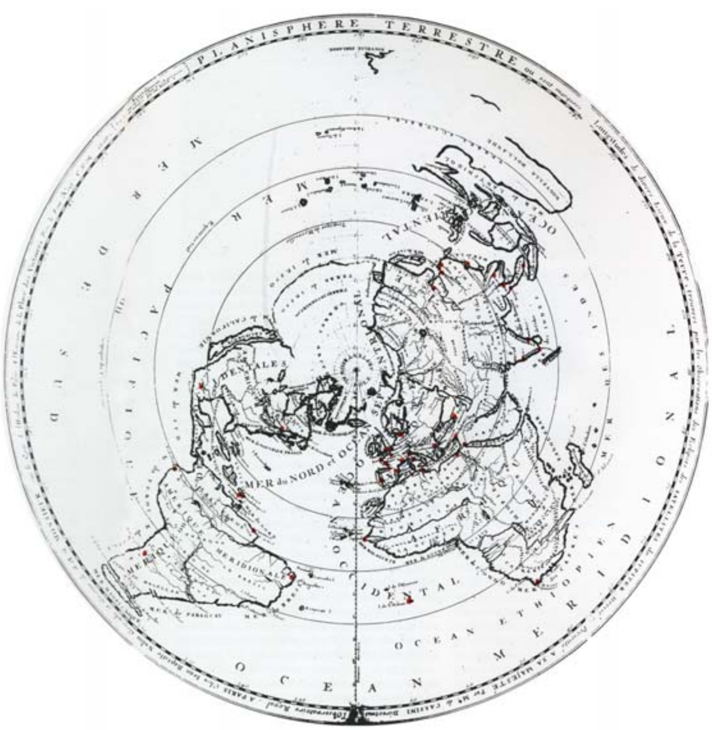
| Early Maps of the World | Early Maps of Science |
| --- | --- |
| 3D | n-D |
| Physically-based | Abstract space |
| Accuracy is measurable | Accuracy is difficult |
| Trade-offs have more to do with granularity | Trade-offs indirectly affect accuracy |
| 2-D projections are very accurate at local levels | 2-D projections neglect a great deal of data |
| Centuries of experience | Decades of experience |
| **Geo-maps can be a template for other data** | **Science maps can be a template for other data** |

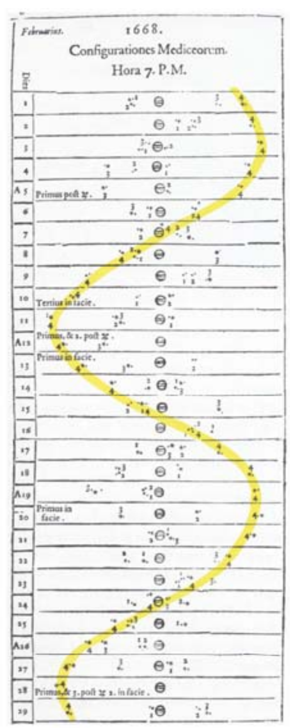
*Kevin W. Boyack, UCGIS Summer Meeting, June, 2009*

## Towards a World Map

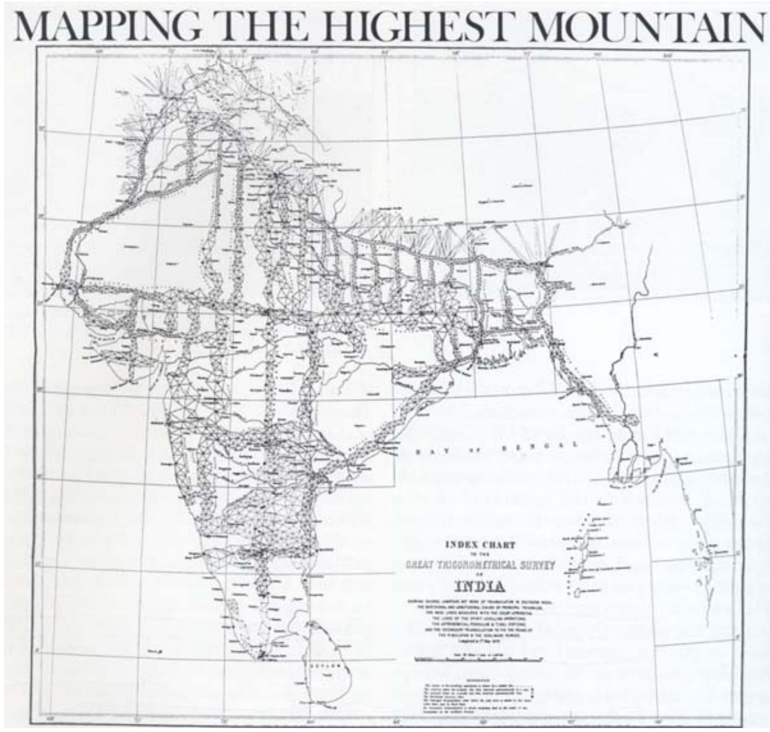

*Portolan chart of the central and western Mediterranean and part of the Atlantic - Bartolome Olives - 1559*


*Islandia - Abraham Ortelius (1527-1598) - 1606*




*In **1696,** the first accurate map (shown below left) of the Earth was drawn by César-François Cassini de Thury based on 40 points (given in red) of accurate latitude and longitude. The north-south position (latitude) of any point on Earth could be determined via star paths. To measure the east-west position (longitude), exact time measurement was essential: one minute of uncertainty implied a 10-mile margin of error in location. Inspired by Galileo's work, the mapmakers used the planet Jupiter as a "clock in the sky." They carefully recorded the motions of Jupiter's moons (see Cassini's 1668 table of the eclipses of Jupiter's moons below).*




*In **1744**, Cassini's team started to map France in a rigorous fashion using triangulation. In the late 1700s, the world's first national land survey of France was completed. In 1870, Captain George Everest embarked to map India by triangulation. For generations, a vast network of repeating sightline triangles was meticulously measured and recorded (see map below). What resembles a pattern of eyelashes on the northern border represents the sightlines to stations built above treetops. While analyzing the triangles in the calculating offices of Calcutta, the mapmakers discovered the highest peak in the world: Mount Everest*

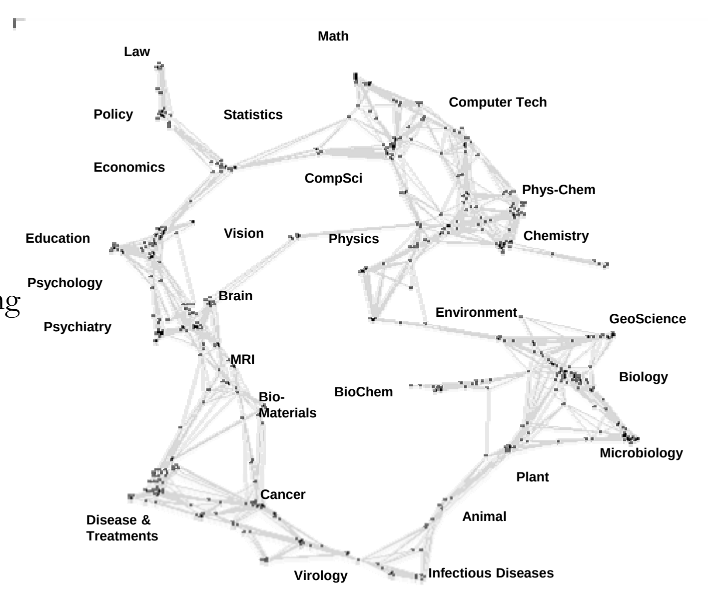

*A New Map of the Whole World with Trade Winds According to the Latest and Most Exact Observations - Herman Moll - 1736*

# Towards a Map of all Sciences

## 2002 '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.*

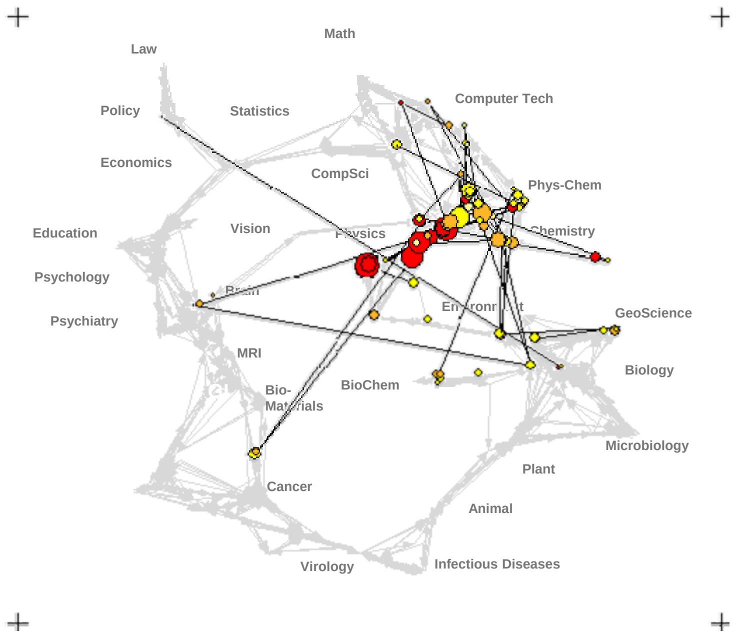
- Uses combined SCI/SSCI from 2002
  - 1.07M papers, 24.5M references, 7,300 journals
  - Bibliographic coupling of papers, aggregated to journals
- Initial ordination and clustering of journals gave 671 clusters
- Coupling counts were reaggregated at the journal cluster level to calculate the
  - (x,y)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

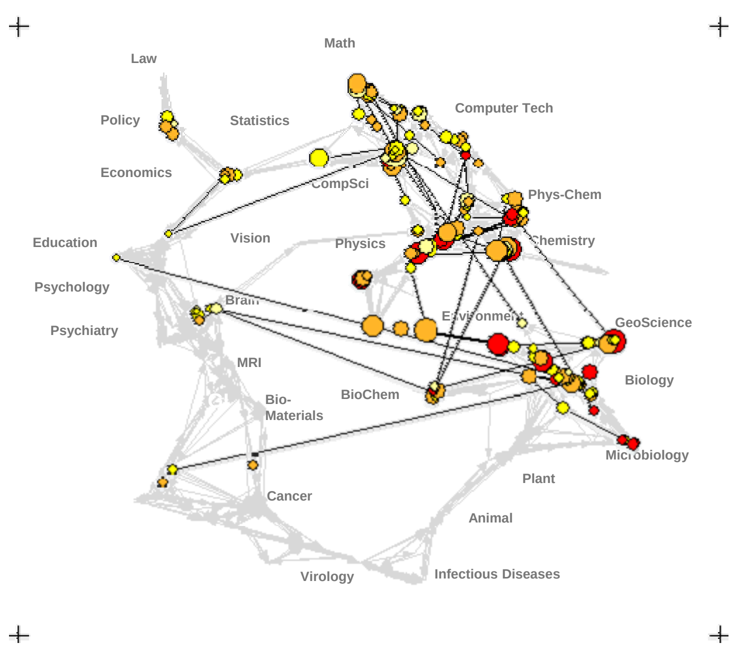
Funding patterns of the US Department of Energy (DOE)




**Science map applications: Identifying core competency**

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

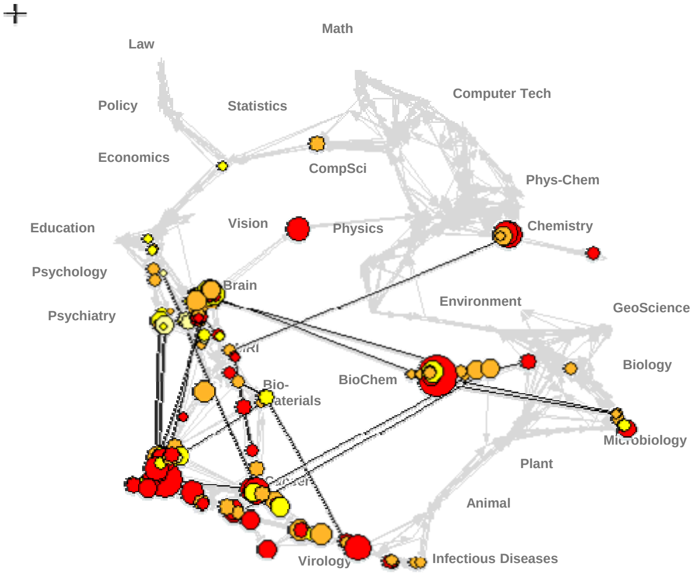
Funding Patterns of the National Science Foundation (NSF)



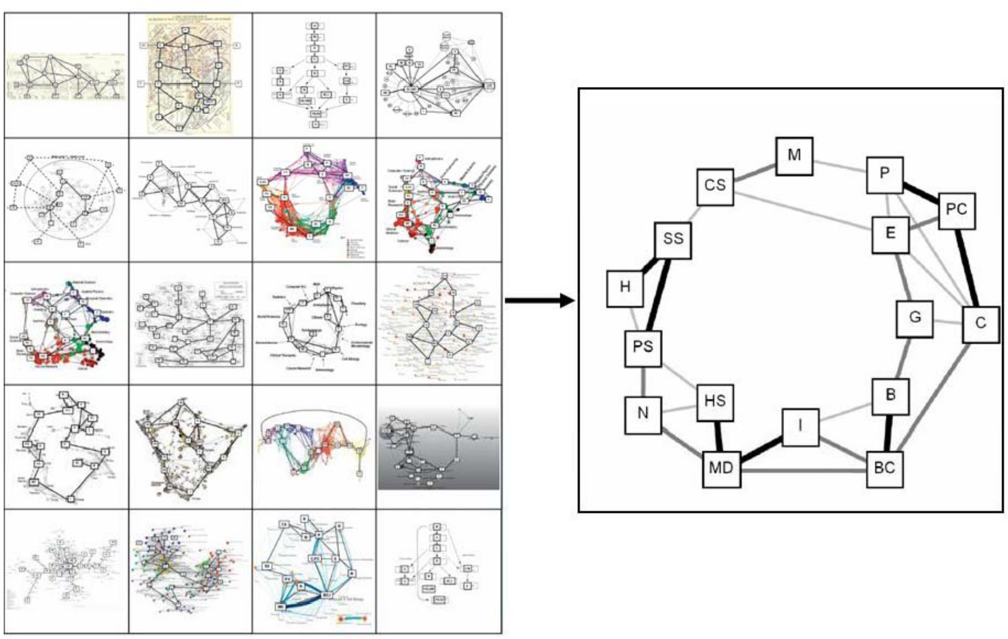
## Science map applications: Identifying core competency

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

Funding Patterns of the National Institutes of Health (NIH)




## Towards a Consensus Map of Science

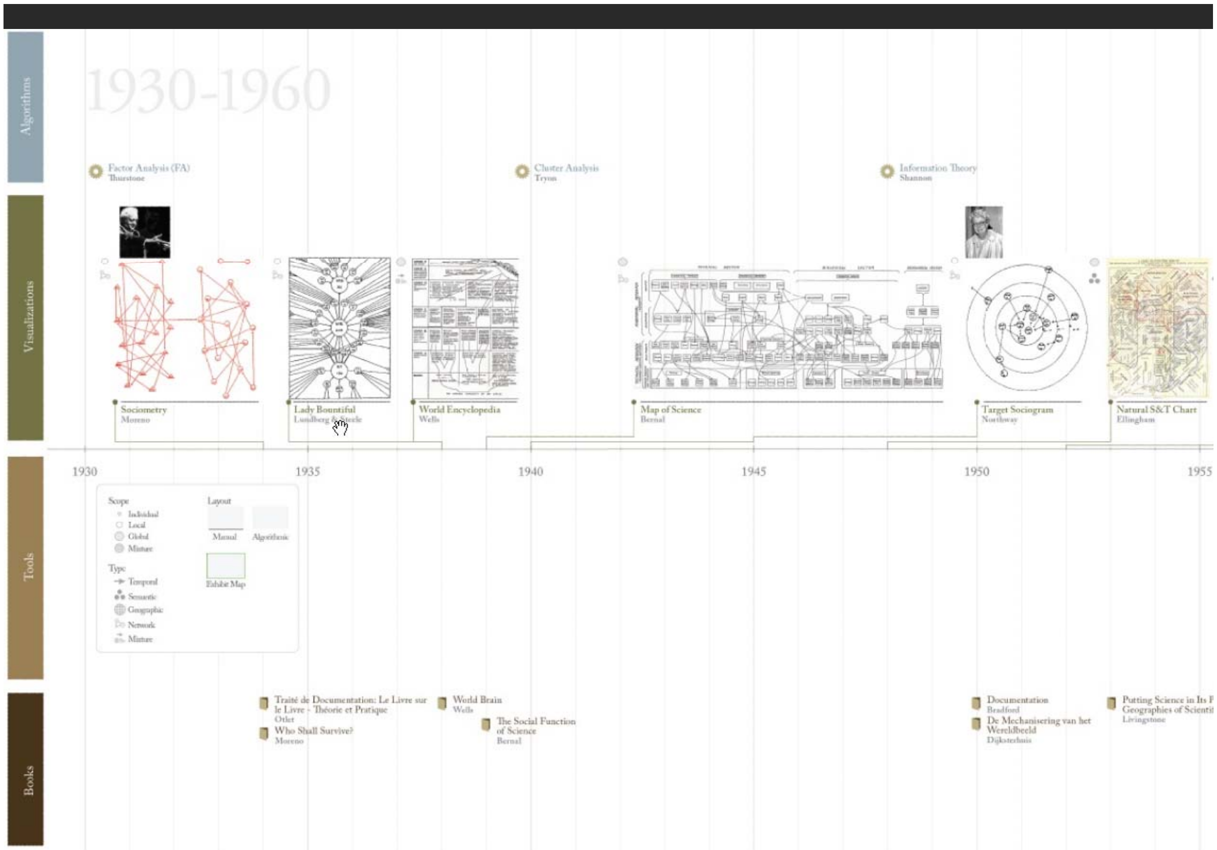



Klavans, R., & Boyack, K. W. (2009). "Toward a consensus map of science." *Journal of the American Society for Information Science and Technology* 60(3), 455-476.

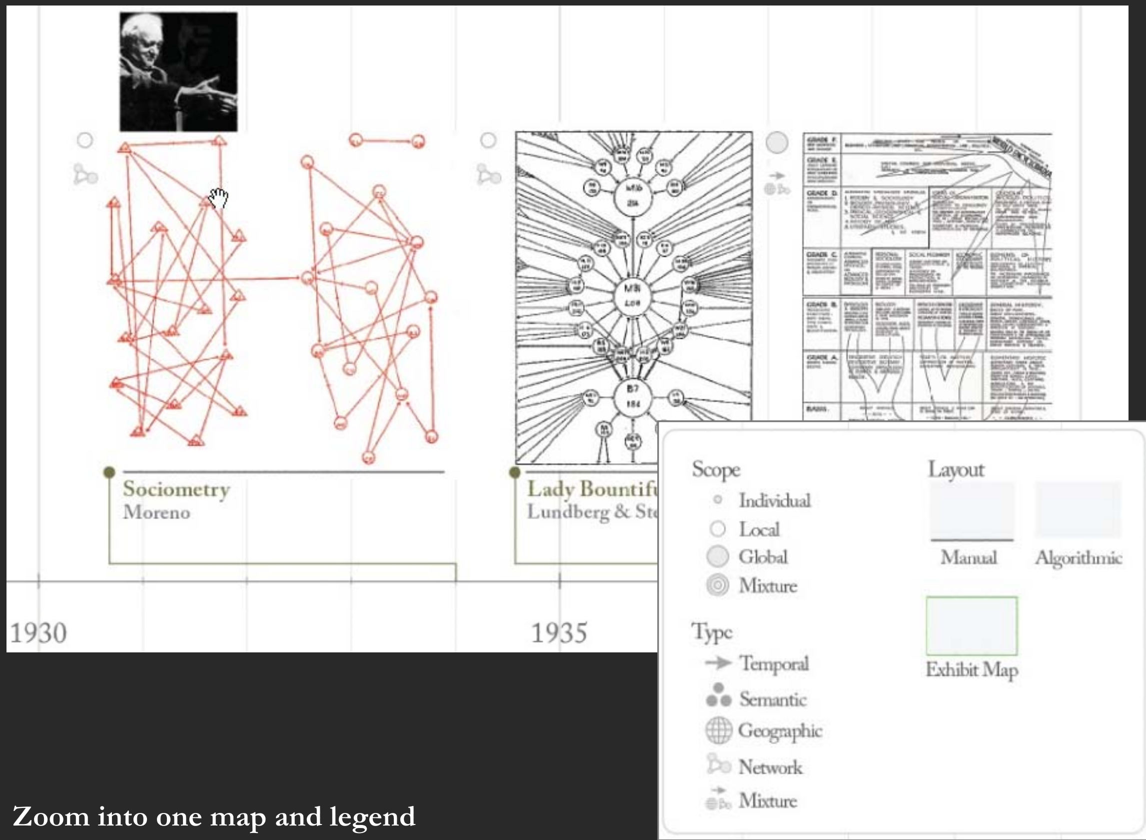
# Milestones of Mapping Science

1934 2007

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

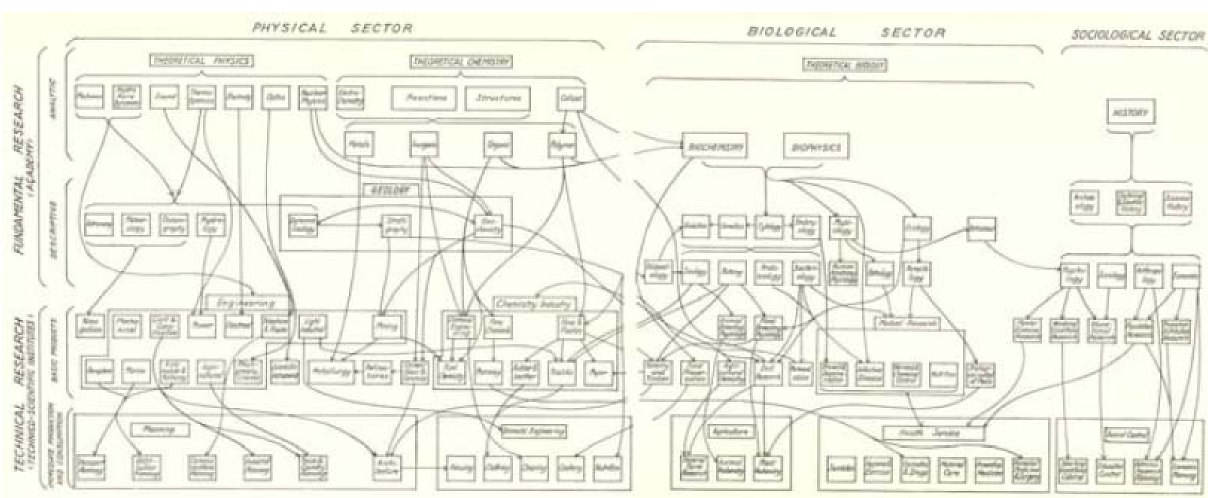



1930 1955




Zoom into one map and legend

# Bernal, 1939



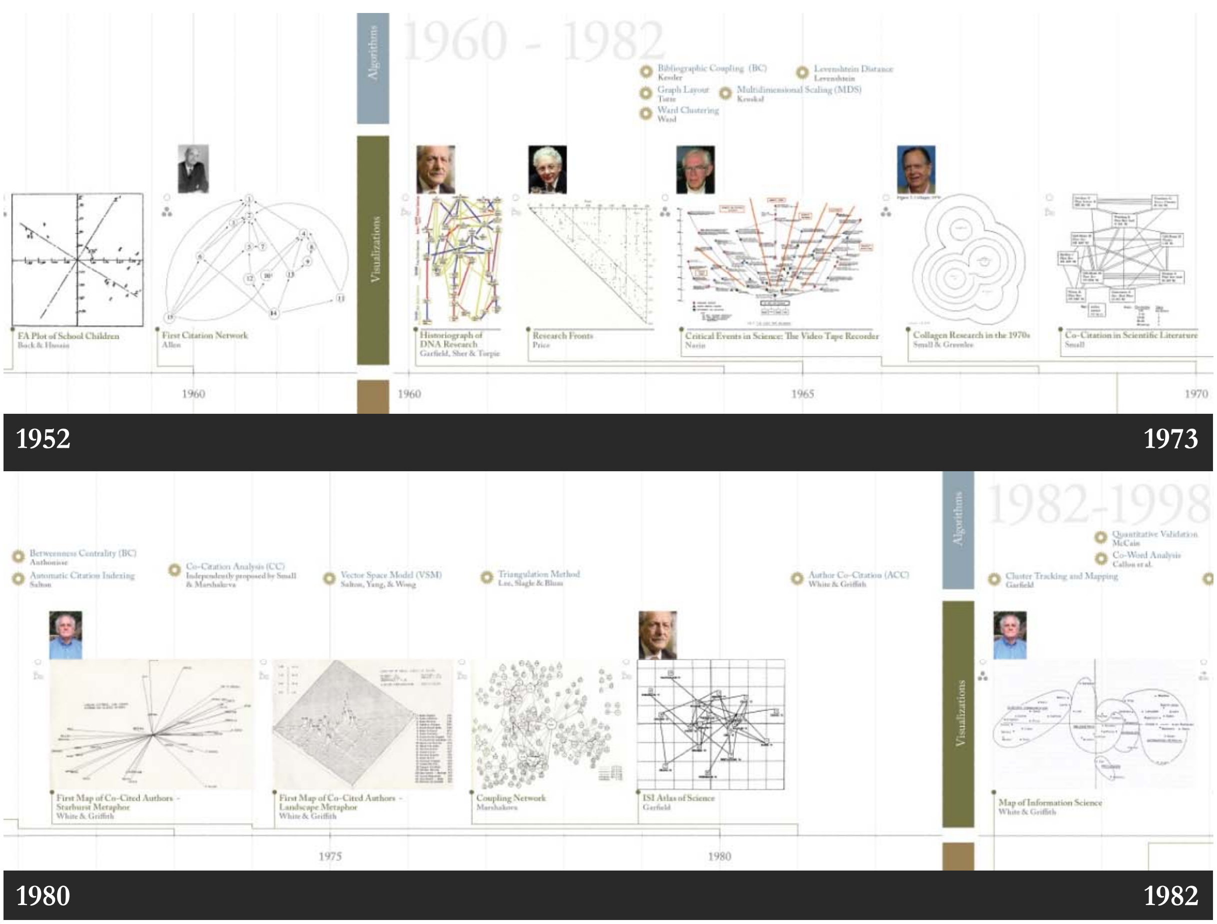

John D. Bernal was a world renowned physicist, a historian of science, and a sociologist of science. He is considered to have produced one of the first 'maps' of science.


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)

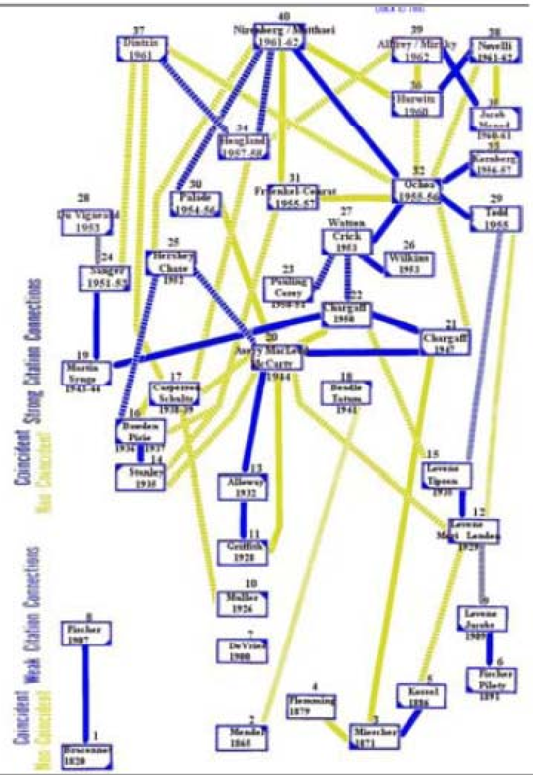

H. J. T. Ellingham (1948). "Divisions of Natural Science and Technology," Royal Society Scientific Information Conference, 21 June to 2 July 1948, London: Burlington House,



# 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.

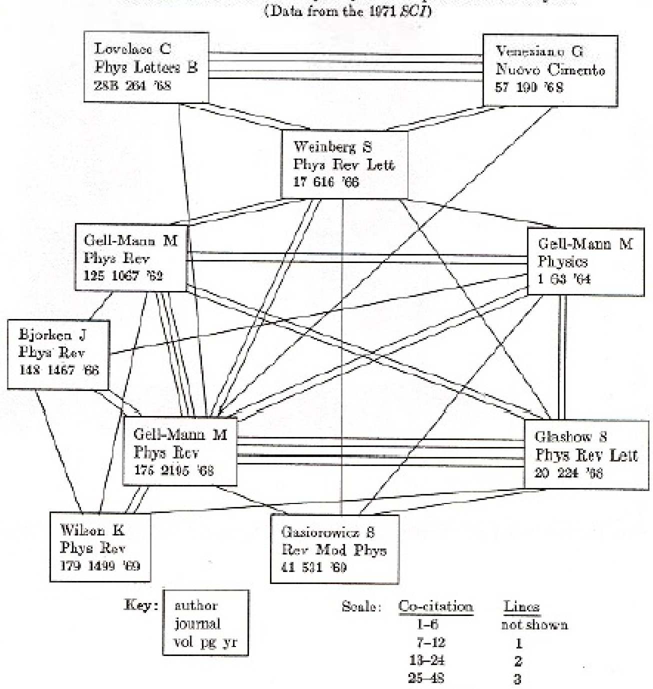
# Small, 1973

Using co-citation to create domain maps


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

FIGURE 1

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



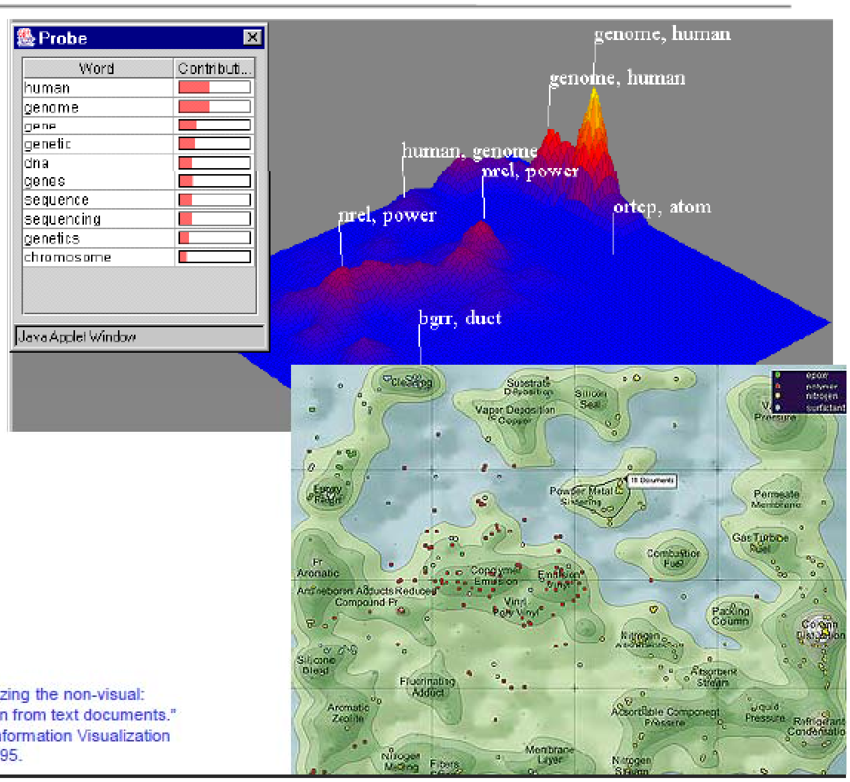

Small, H. (1973). "Co-citation in the scientific literature: A new measure of the relationship between two documents." *JASIS*, 24, 265-269.

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


# 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

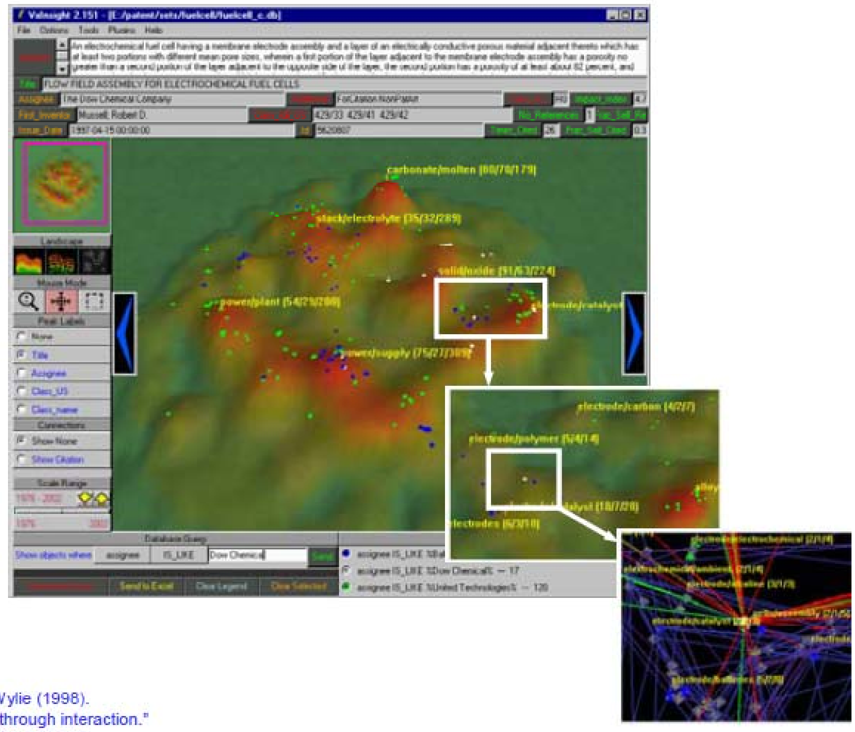
Wise, Thomas, Pennock, et al. (1995). "Visualizing the non-visual: Spatial analysis and interaction with information from text documents." Paper presented at the IEEE Symposium on Information Visualization '95, Atlanta, Georgia, USA., October 30-31, 1995.

# 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

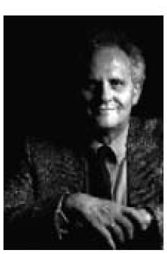


Davidson, Hendrickson, Johnson, Meyers & Wylie (1998). "Knowledge mining with VxInsight: Discovery through interaction." *Journal of Intelligent Information Systems*, 11(3), 259-285.

# 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.

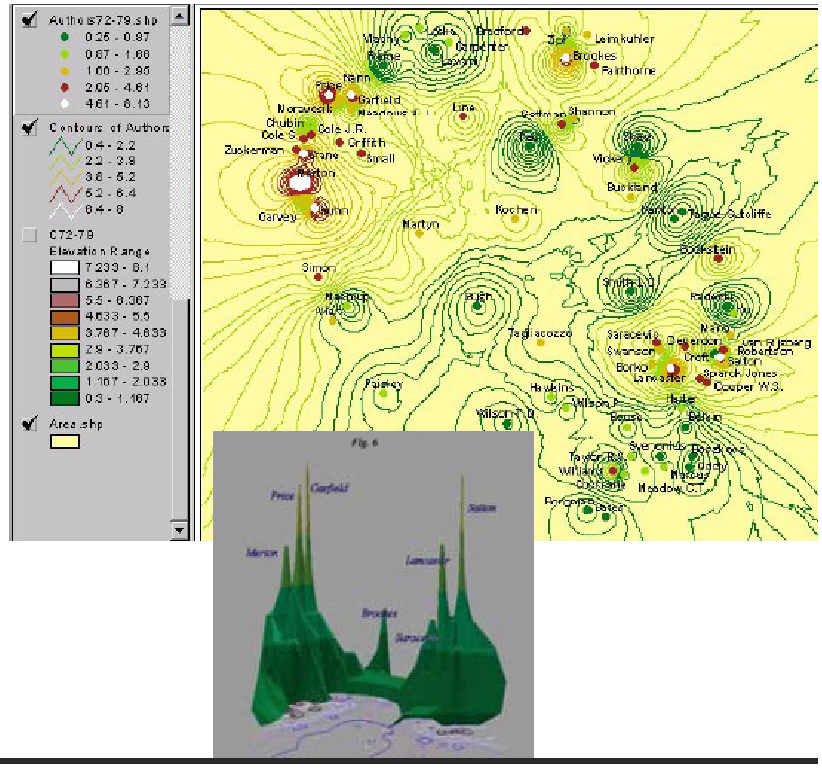
# Old, 2001

Utilizing spatial information systems for non-spatial data analysis

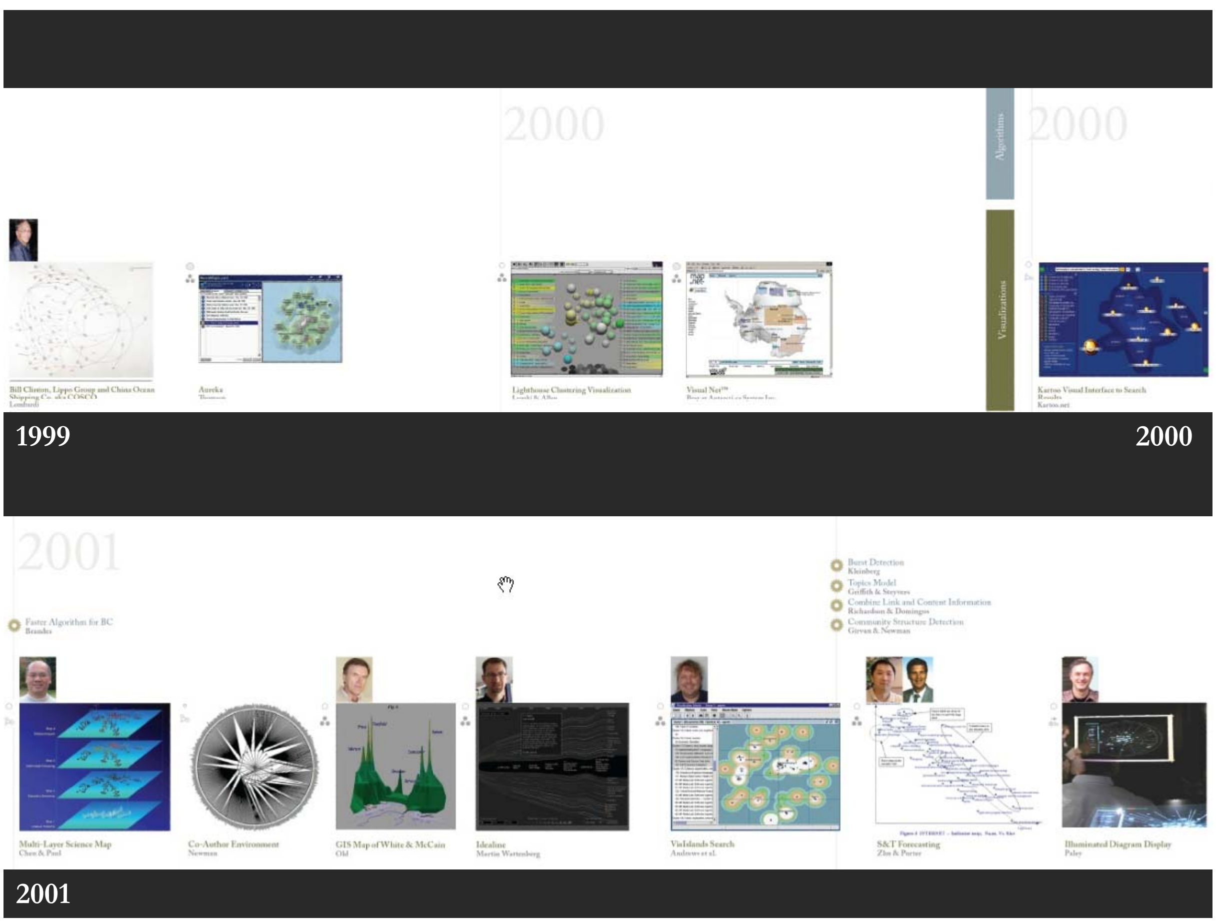
Authors in Information Science

Topography added

3D representation also



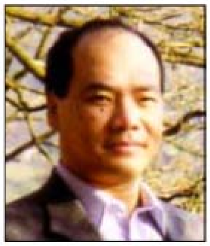

http://php.indiana.edu/~jold/SLIS/L710/L710.htm

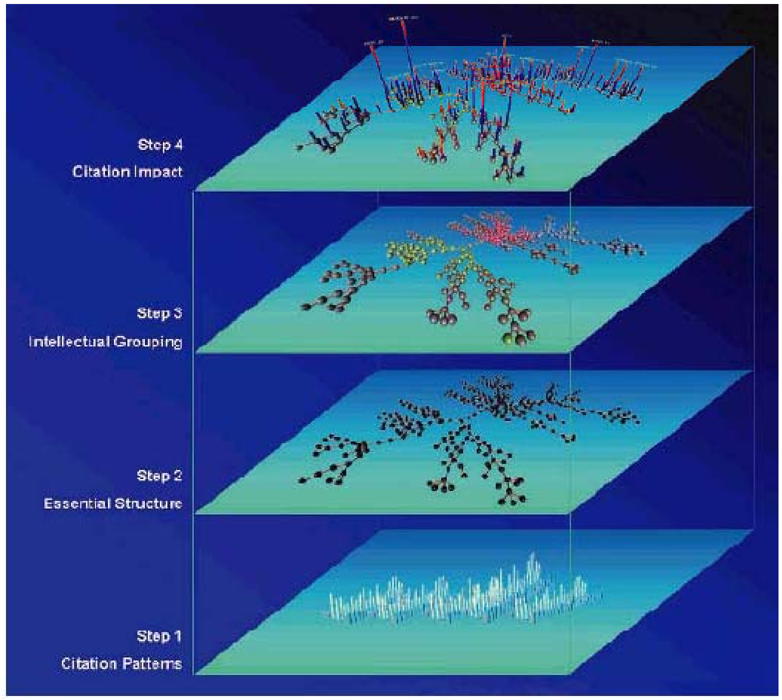


# Chen, 2001

Software: CiteSpace I

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



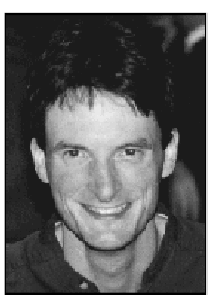


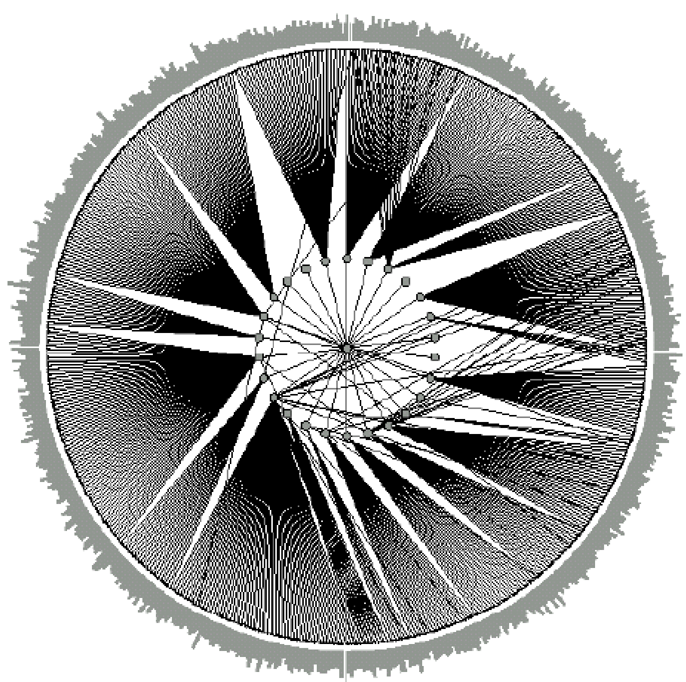

Chen & Paul (2001). "Visualizing a knowledge domain's intellectual structure." *Computer*, 34(3), 65-71.

# Newman, 2001

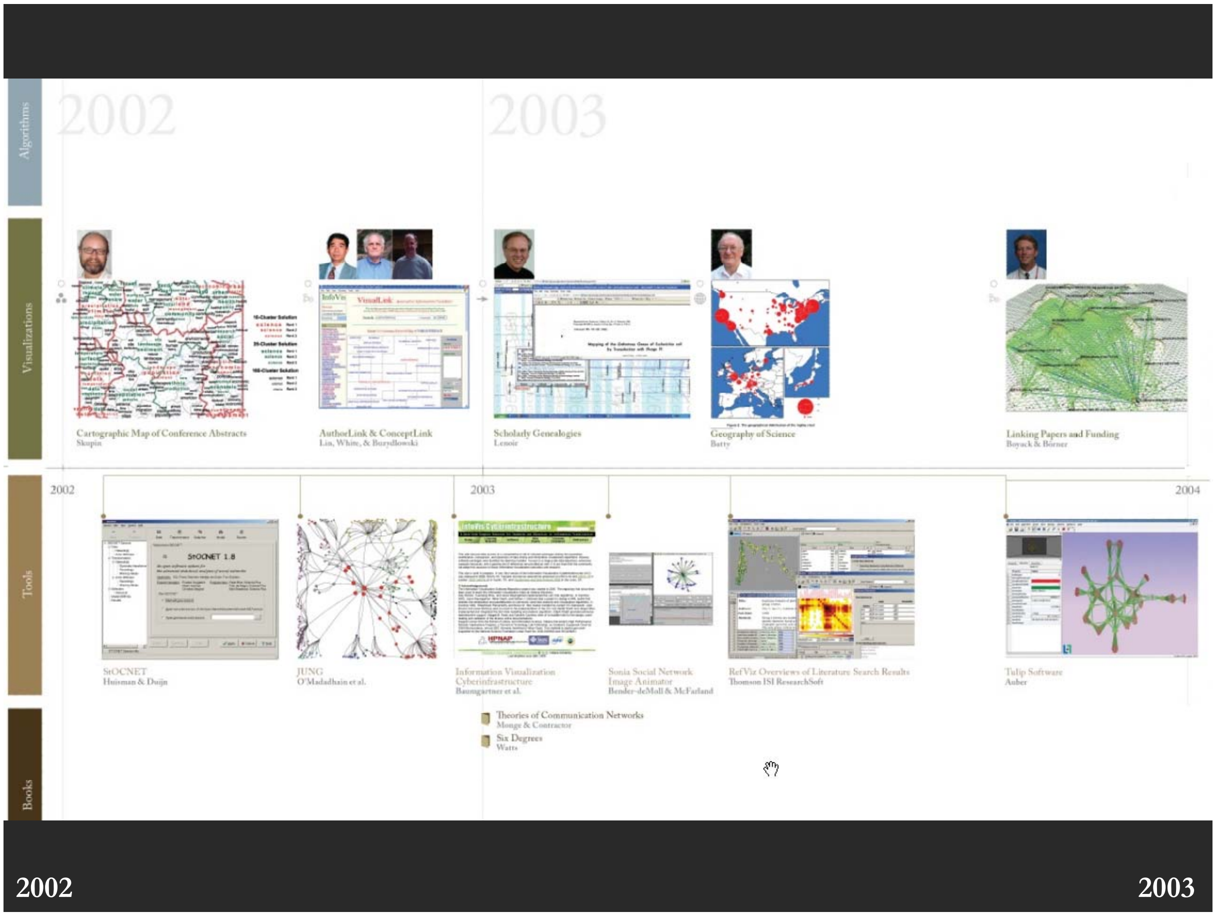
Physicist bringing new tools to the problem

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





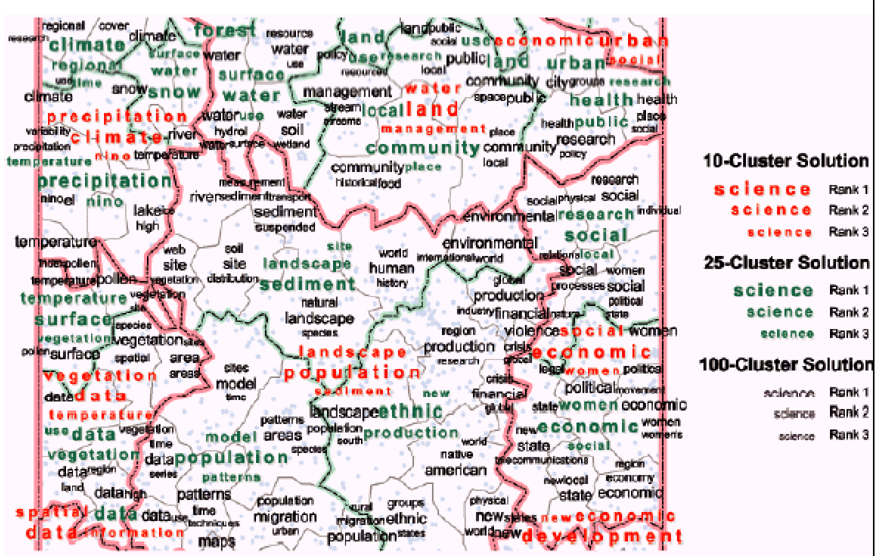
Newman, M.E.J. (2001). "Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality." *Physical Review E*, 64, paper # 016132.



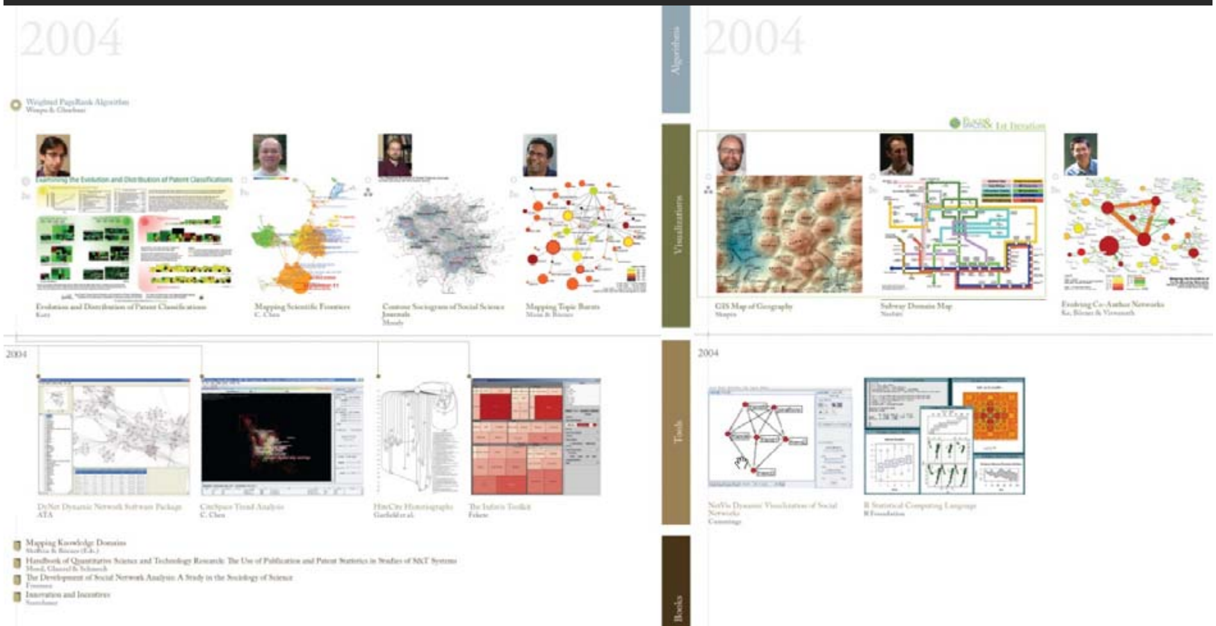

# Skupin, 2002

Applied self-organizing map to conference abstracts

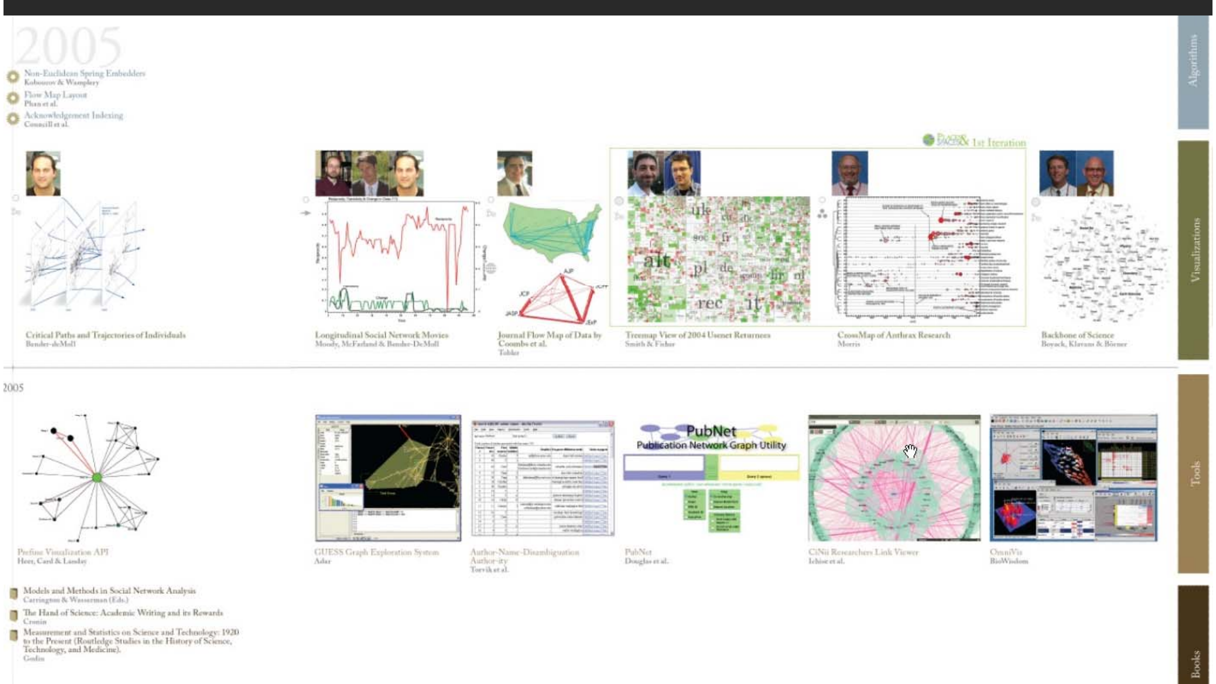
Applied cartographic principles and techniques to the resulting SOM




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


2004
Algorithms
Visualizations
Tools
Books

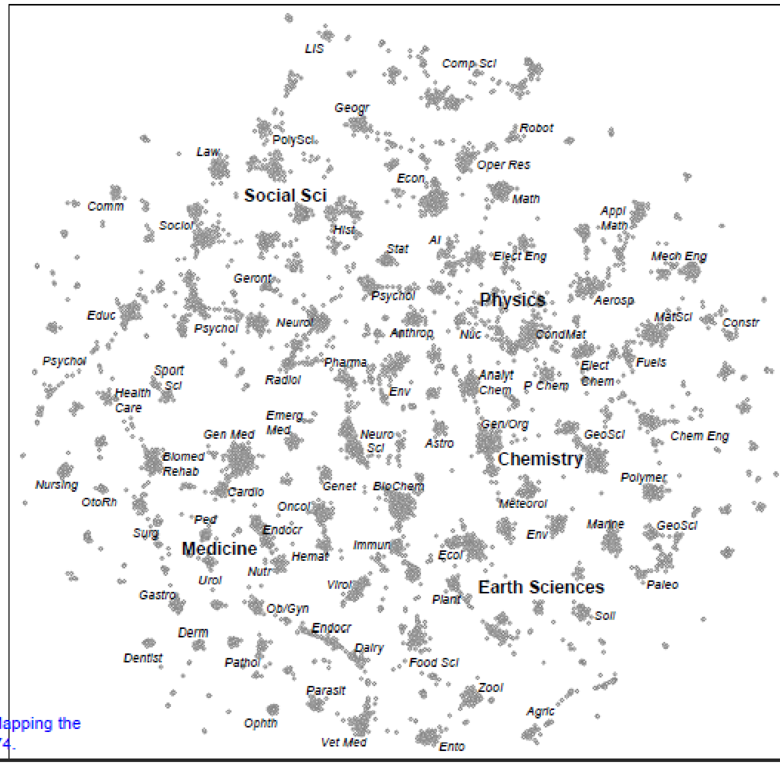
**2004**


2005
Algorithms
Visualizations
Tools
Books
PubNet
Publication Network Graph Utility

**2005**

# All of Science: Boyack et al., 2003

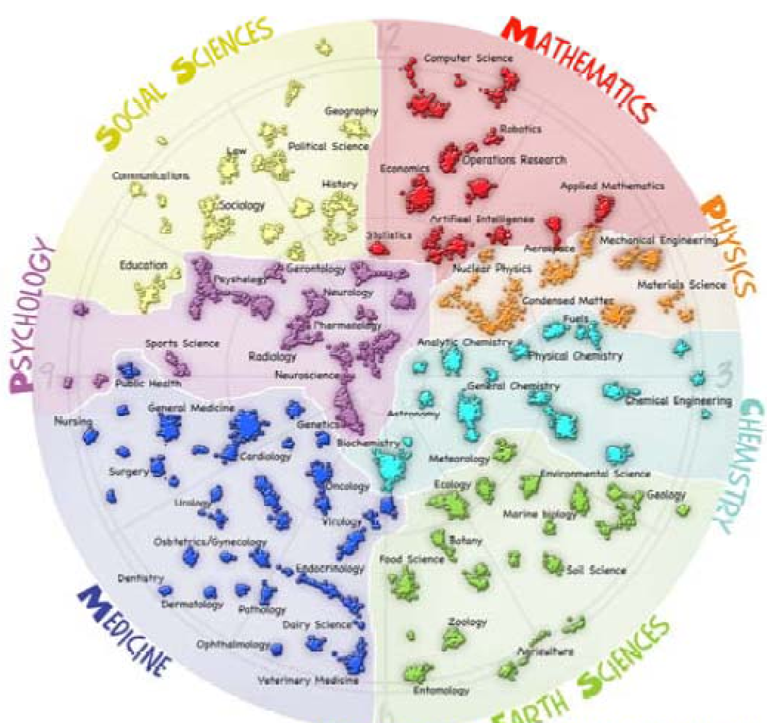
- Map of over 7,000 journals
- Over 200 clusters of journals based on inter-citation statistics
- Graph layout techniques to position clusters
- Inaccurate locations: it appears there is a center to science, but it is an artifact of the graph layout




Boyack, K. W., Klavans, R., & Börner, K. (2005). "Mapping the backbone of science." *Scientometrics* 64(3), 351-374.

# All of Science: Boyack et al., 2003

- Map of over 7,000 journals
- Over 200 clusters of journals based on inter-citation statistics
- Graph layout techniques to position clusters
- Inaccurate locations: it appears there is a center to science, but it is an artifact of the graph layout




Modified from Boyack et al. by Ian Aliman, Indiana University.

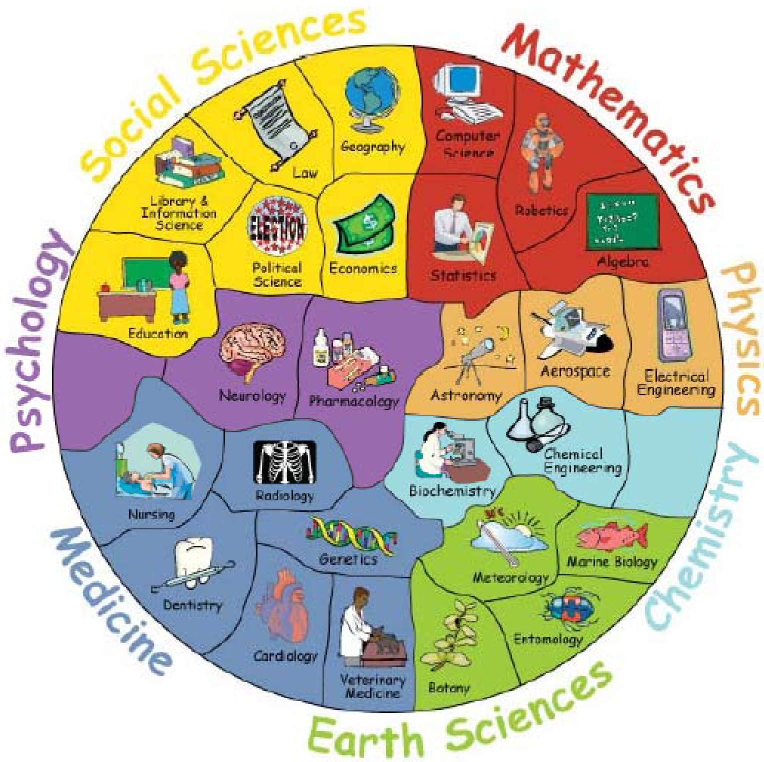
# All of Science: Boyack et al., 2003

Map of over 7,000 journals

Over 200 clusters of journals based on inter-citation statistics

Graph layout techniques to position clusters

Inaccurate locations: it appears there is a center to science, but it is an artifact of the graph layout




**Cambrian explosion** ~ seemingly rapid appearance of most major groups of complex animals around 530 million years ago.

**2006**

# All of Science: Klavans, 2007

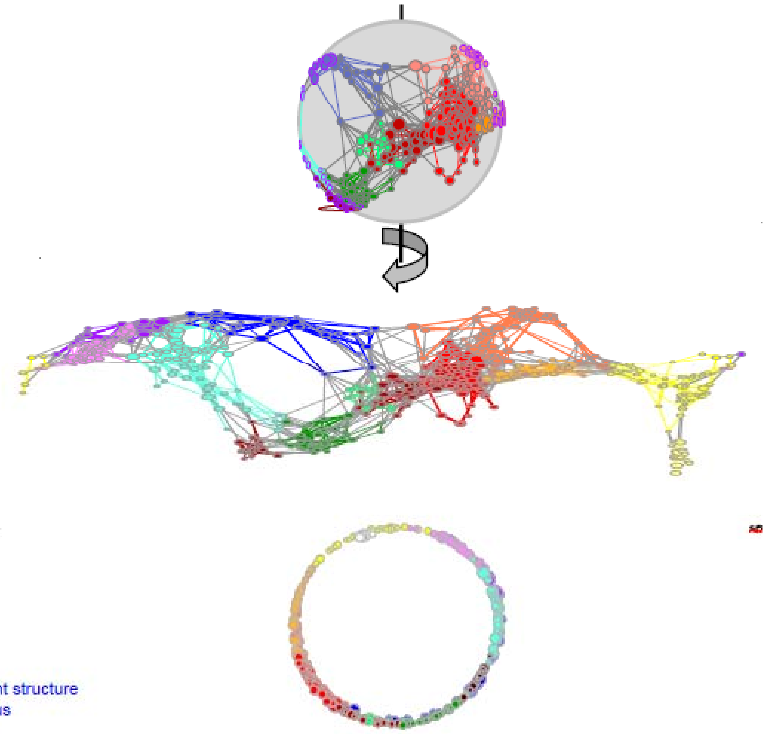
Map of over 16,000 journals and proceedings from ISI and Scopus databases (2001-2005 fileyears)

Two levels of clustering

Spherical layout of 554 clusters

Mercator projection of spherical layout

View from "south pole" is nearly circular

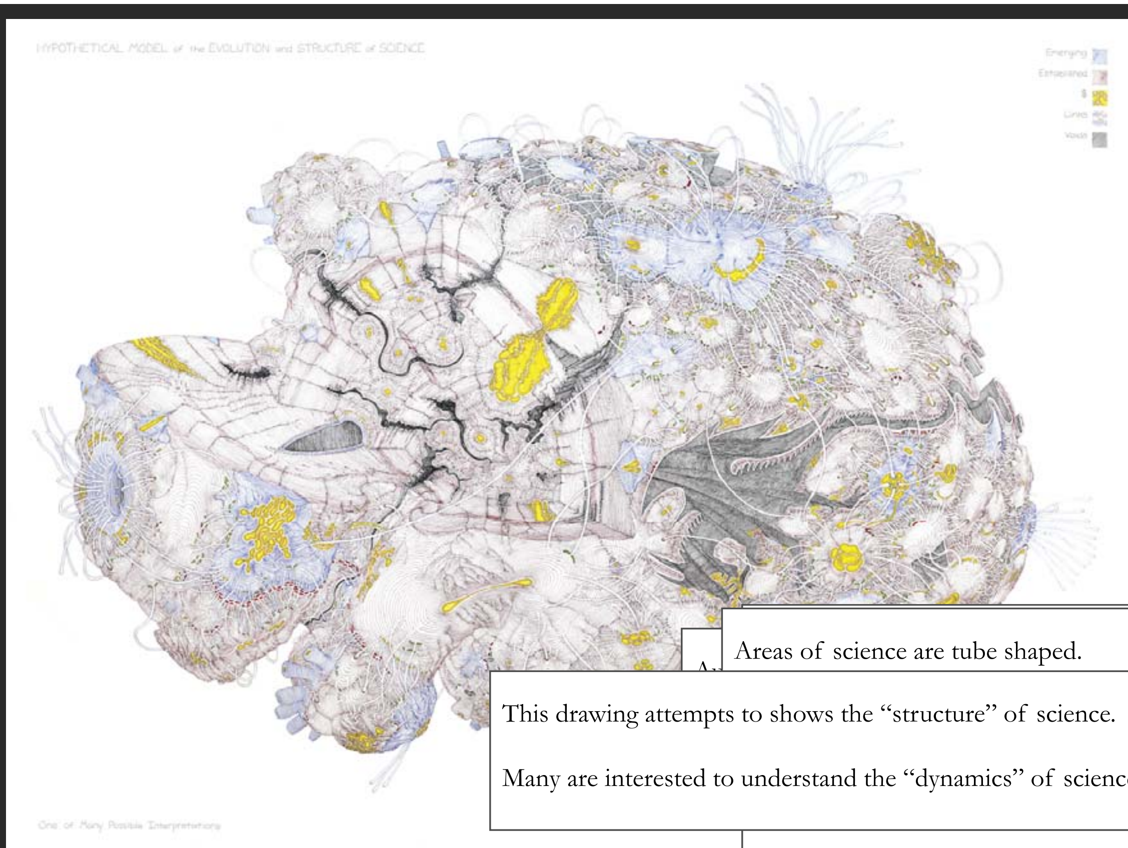


Klavans, R. & Boyack, K. W. (2007). "Is there a convergent structure of science? A comparison of maps from the ISI and Scopus databases." *ISSI 2007.*

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*.



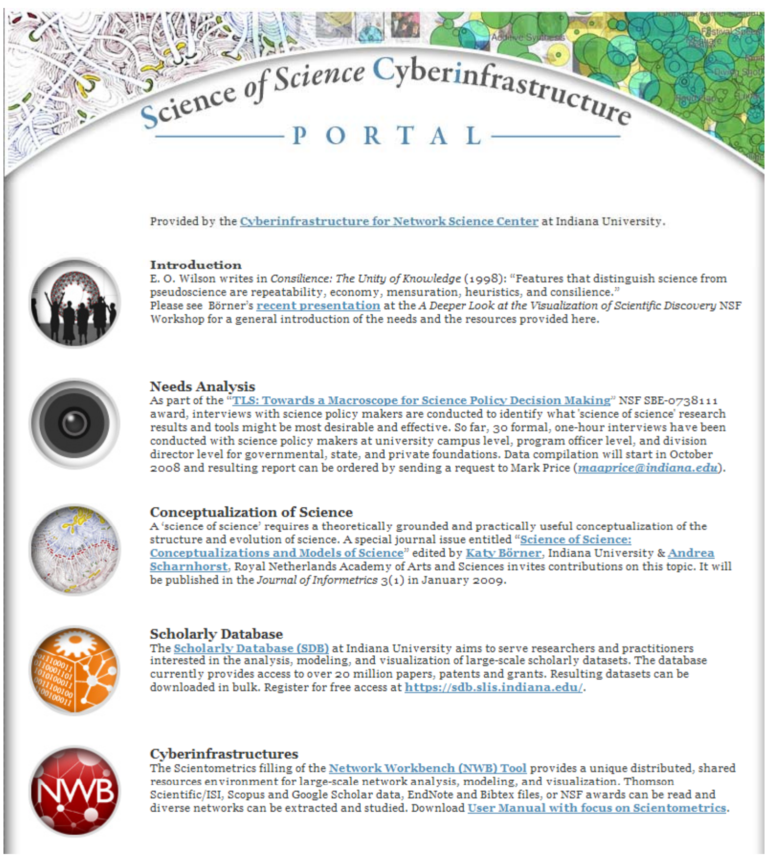
Areas of science are tube shaped.

This drawing attempts to shows the "structure" of science.

Many are interested to understand the "dynamics" of science.

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





Science of Science Cyberinfrastructure

PORTAL

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

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

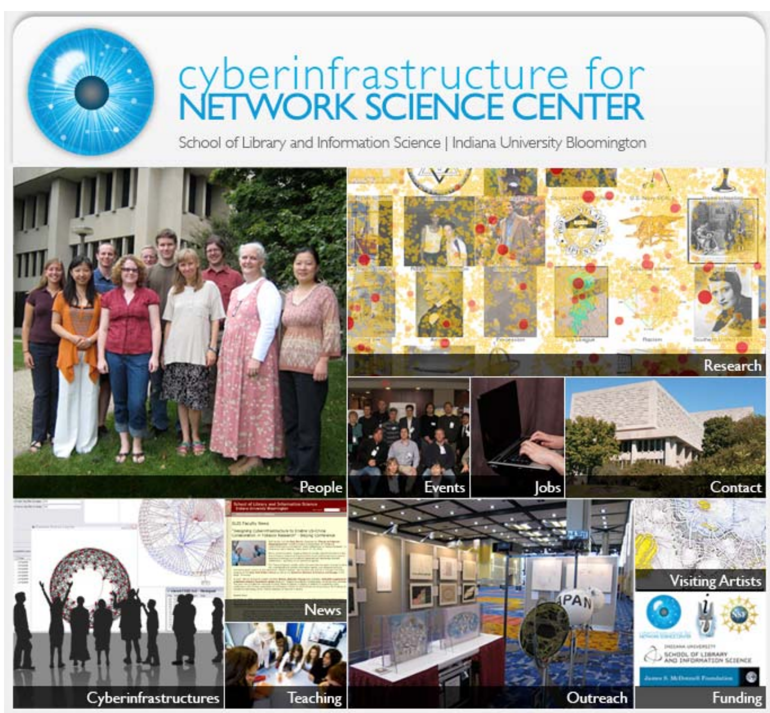
**Needs Analysis**
As part of the "TLS: Towards a Macroscope for Science Policy Decision Making" NSF SBE-0738111 award, interviews with science policy makers are conducted to identify what 'science of science' research results and tools might be most desirable and effective. So far, 30 formal, one-hour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

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

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

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

http://sci.slis.indiana.edu


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