Knowledge Domain Visualizations:

Tools to Navigate, Understand, and Internalize the Structure of Scholarly Knowledge and Expertise

An Emerging Field of Information Cartography

Peter A. Hook, J.D., M.S.L.I.S.
Doctoral Student, Indiana University Bloomington
http://ella.slis.indiana.edu/~pahook
Fifth Annual AISTI Mini-Conference
Santa Fe, New Mexico
Wednesday, May 25, 2005

Conclusions:

- Scholarly production and consumption is a complex system that justifies the attention of information scientists to contribute to macro and micro efficiencies in the use and understanding of information.
- KDVs contribute to the effective use and understanding of information.
- KDVs are a form of intellectual capital (in the competitive intelligence sense) and are cognitive scaffolding (in the pedagogical sense).
- There is an emerging field of information cartography.

OVERVIEW

- Introduction
- History of KDVs
- How Created
- How Used
- Educational Aspects
- Metro Map Metaphor
- Predictions

Introduction to KDVs

Definitions

- **Knowledge domain visualizations** (KDVs) are the graphic rendering of bibliometric data designed to provide a global view of a particular domain, its structural details, or its salient characteristics (most cited authors or papers, bursting concepts, etc.).
- The visualized domain may vary in size from large to small.
  - all of knowledge,
  - a grouping of academic subjects (science),
  - a discipline (information science)
  - a single book (Joyce's *Portrait of the Artist as a Young Man.*)
- KDVs are also referred to as domain maps and the process of their creation as domain mapping.

Continuums

- spatially literal ⟷ spatially abstract
- hand generated ⟷ produced by automated means
- primitive ⟷ graphic and cartographic rich

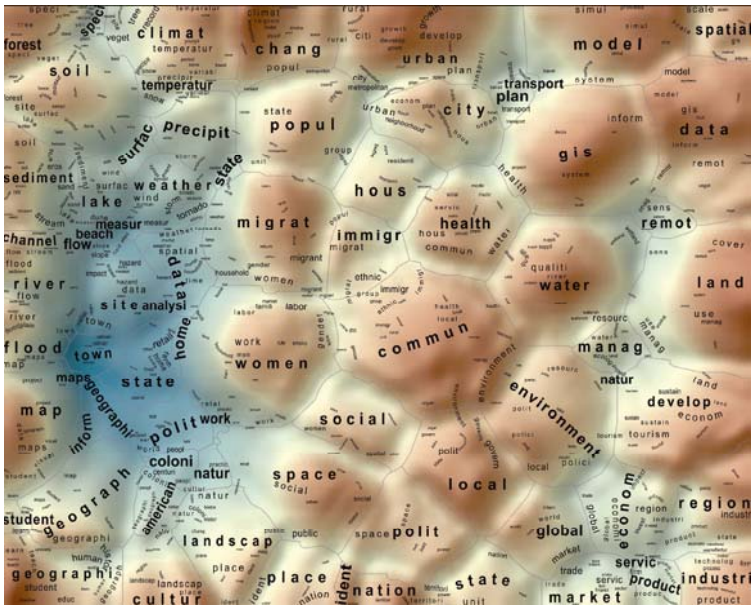
2005

Map of all Science & Social Science

- Each dot is one journal
- Journals group by discipline
- Labeled by hand
- The map is comprised of 7,121 journals from year 2000.

Boyack, K.W., Klavans, R., & Börner, K. (2005, in press). Mapping the backbone of science. *Scientometrics.*

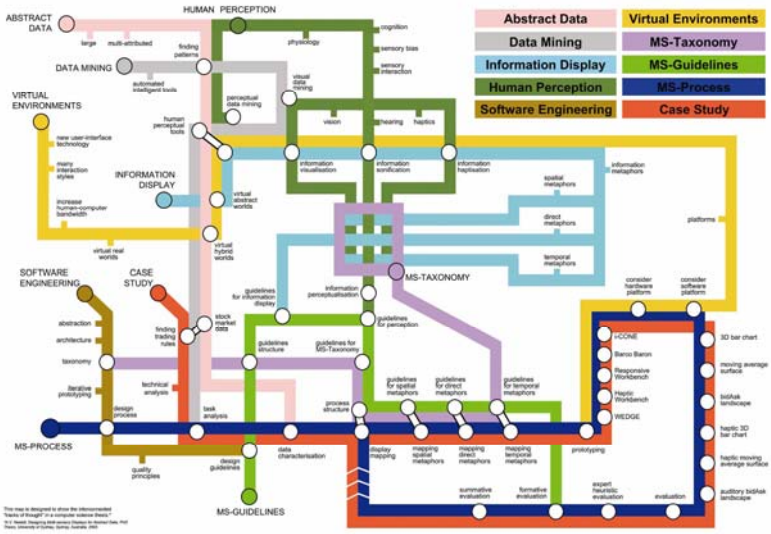
2005



Information space of 22,000 conference abstracts submitted to the AAG annual meeting from 1993-2002.

See: Skupin, A., and Hagelman, R. (2005) Visualizing Demographic Trajectories with Self-Organizing Maps. *GeoInformatica*. 9(2).

2004



Intersecting threads of thought of Keith Nesbitt's doctoral dissertation.

Employing the metro map metaphor.

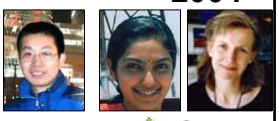
Nesbitt (2004). "Getting to more Abstract Places using the Metro Map Metaphor," IV'04, 8th International Conference on Information Visualisation London,2004.

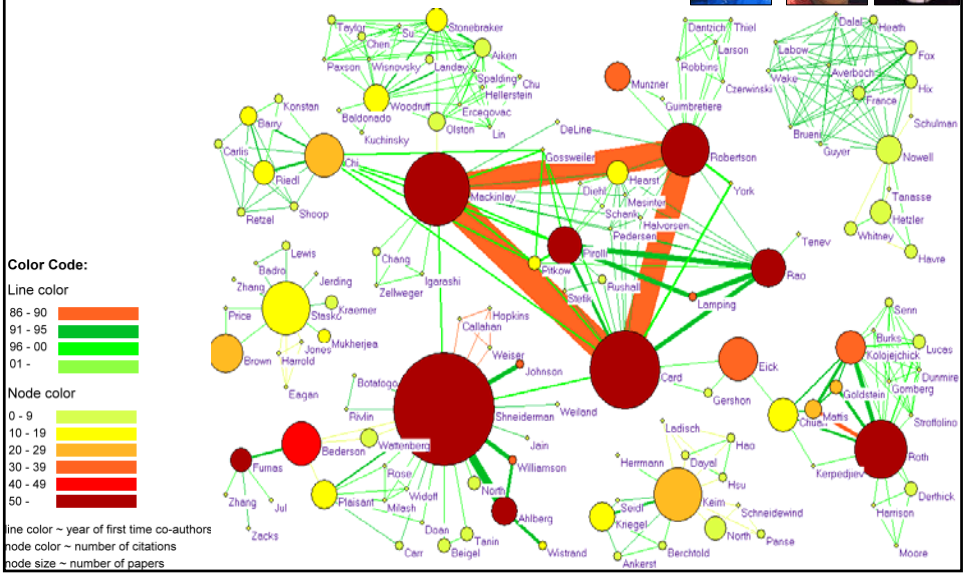
Visualization of Growing Co-Author Networks

2004

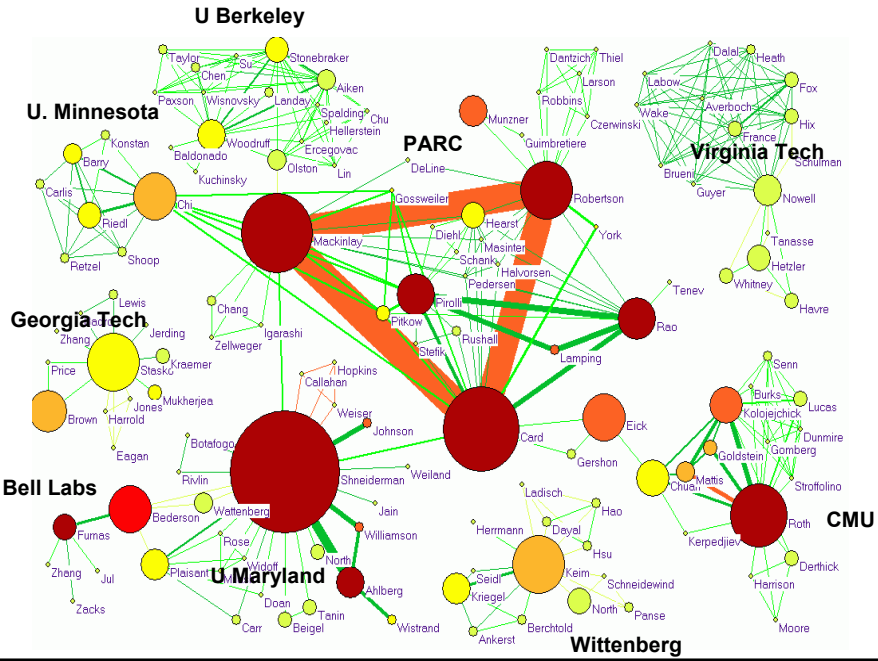
Won 1st prize at the IEEE InfoVis Contest

(Ke, Visvanath & Börner, 2004)





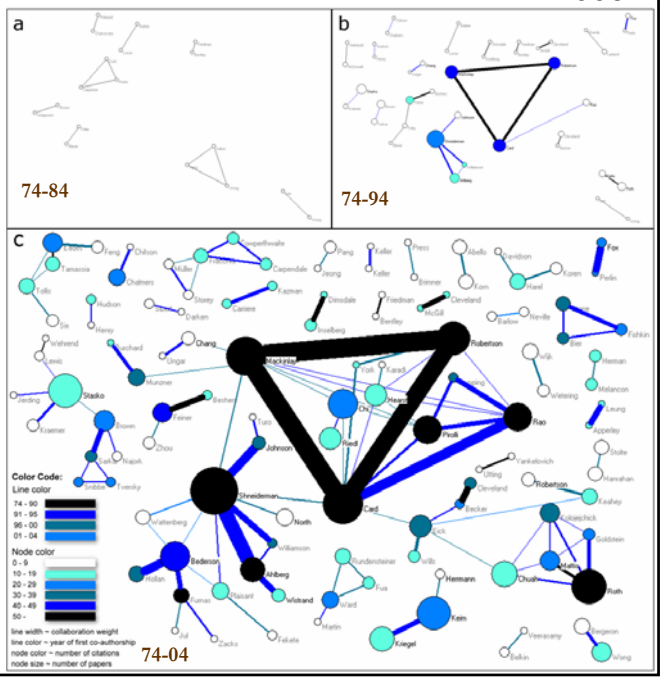
After Stuart Card, IEEE InfoVis Keynote, 2004.



Visual Encoding

2005

- Nodes represent authors
- Edges denote co-authorship relations
- Node area size reflects the number of single-author and co-authored papers published in the respective time period.
- Node color indicates the cumulative number of citations received by an author.
- Edge color reflects the year in which the co-authorship was started.
- Edge width corresponds to the impact weight.



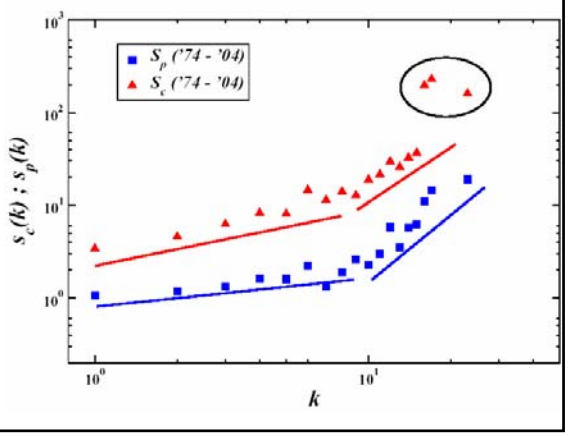
Börner, Katy, Dall'Asta, Luca, Ke, Weimao and Vespignani, Alessandro. (April 2005) Studying the Emerging Global Brain: Analyzing and Visualizing the Impact of Co-Authorship Teams. *Complexity*, special issue on *Understanding Complex Systems*.

Benefits of Co-Authoring

Publication strength S_p and the citation strength S_c of authors versus the degree of authors (number of co-authors) for the 74-04 time slice.
Solid lines are a guide to the eye indicating the presence of two different regimes as a function of the co-authorship degree *k*.

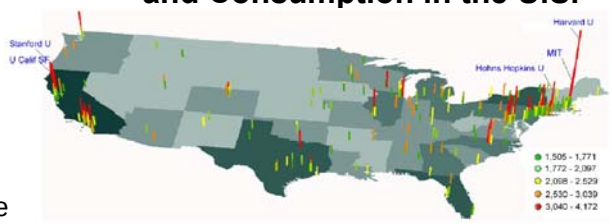
Discussion:
Two definite slopes.
Impact and productivity grow faster for authors with a large number of co-authorships.

The three high degree nodes represent S._K._Card, J._D._Mackinlay, and B._Shneiderman.



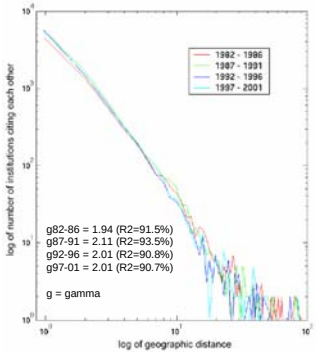
2005

Spatio-Temporal Information Production and Consumption in the U.S.



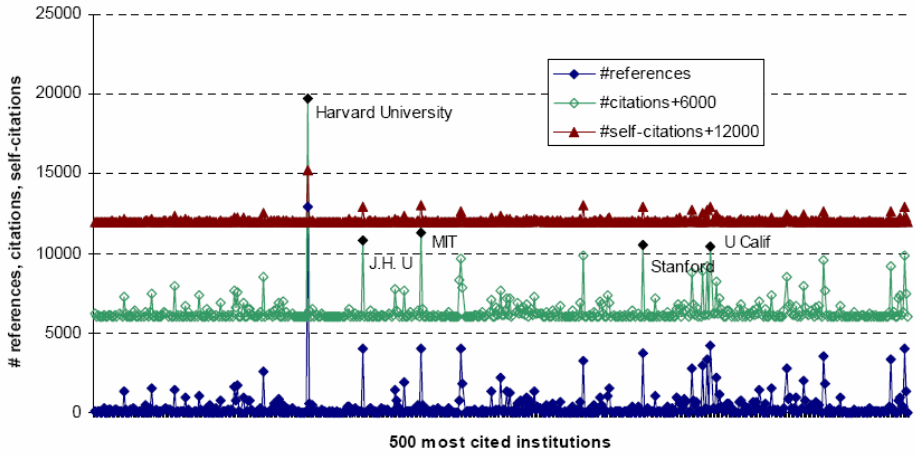
- Dataset: all PNAS papers from 1982-2001 (dominated by research in biology)
- 47K papers, 19K unique authors, 3K institutions
- Each paper was assigned the zip code location of its first author
- Dataset was parsed to determine the 500 top cited (most qualitatively productive) institutions.

- Harvard University (13,763 citations)
- MIT (5,261 citations)
- Johns Hopkins (4,848 citations)
- Stanford (4,546 citations)
- University of California San Francisco (4,471 citations)
- All totals exclude self citation

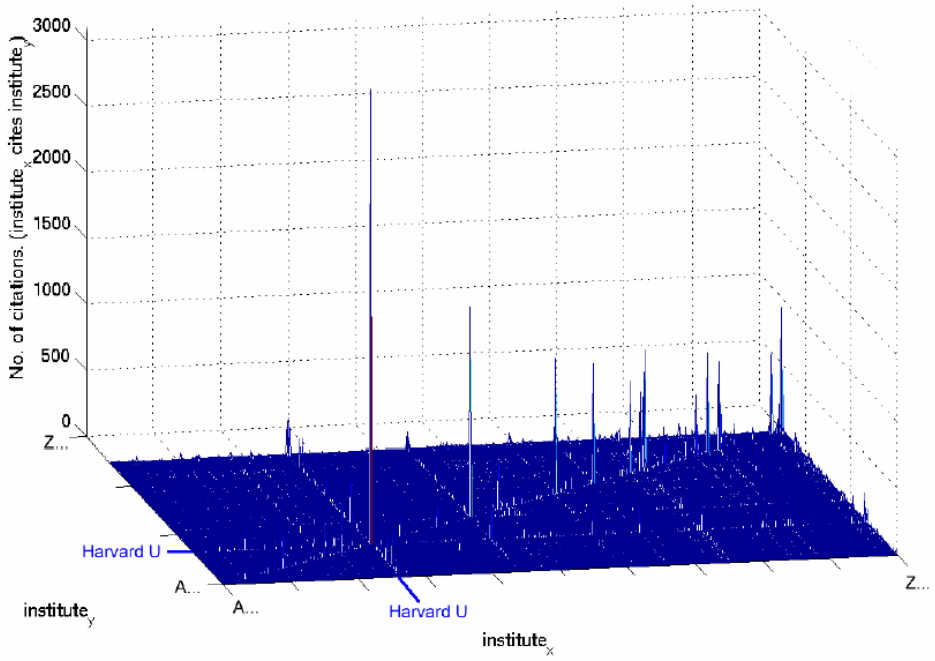


Börner, Katy & Penumarthy, Shashikant. (in press) Spatio-Temporal Information Production and Consumption of Major U.S. Research Institutions. Accepted at the 10th International Conference of the International Society for Scientometrics and Informetrics, Stockholm, Sweden, July 24-28.

Relevant Metrics

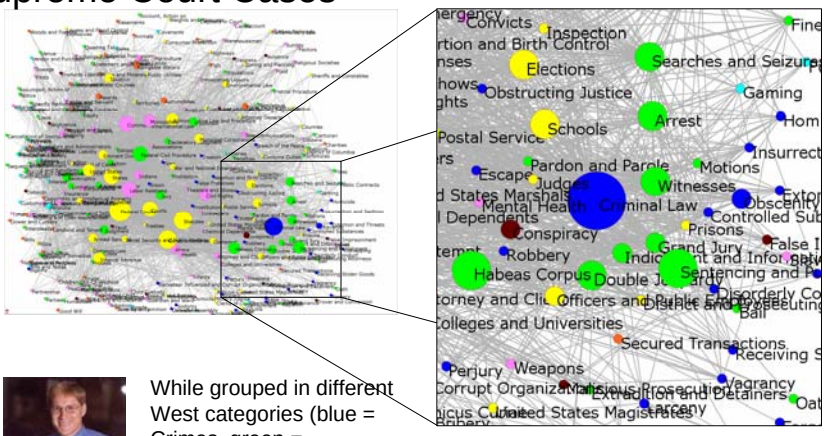


- **References** = institution cites other institutions (Consumes Information)
- **Citations** = institution is cited by other institutions (Produces Information (of utility))
- Methodology can determine the net producers and consumers of information.



Co-Occurrence of West Topics in Supreme Court Cases

2004



While grouped in different West categories (blue = Crimes, green = Remedies), it appears that Criminal Law more closely relates to some Remedies topics than Crimes topics.

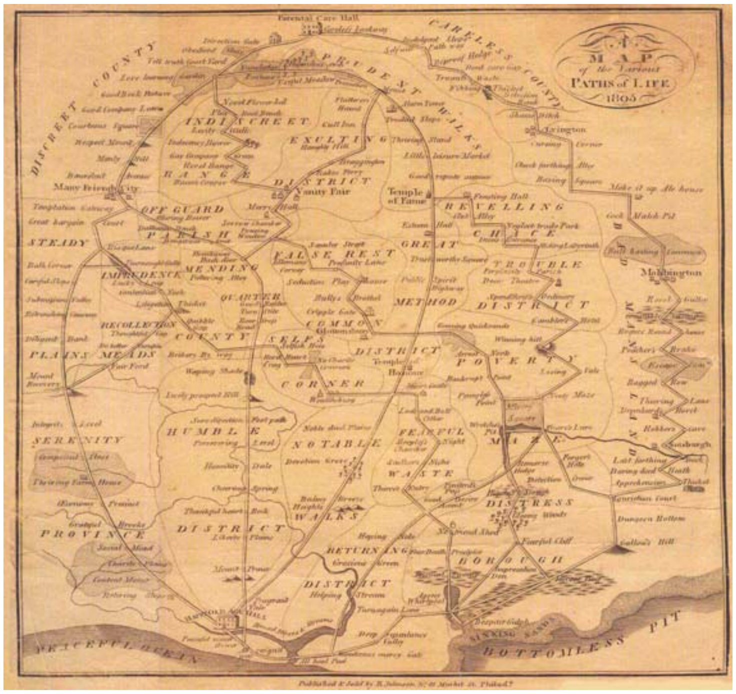
Color Code:

1. Persons
2. Property
3. Contracts
4. Torts
5. Crimes
6. Remedies
7. Government

Node Size ~ Number of times topic appears in the dataset

Node Color ~ West Category

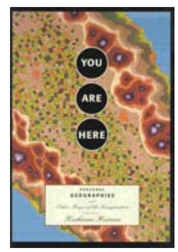
History of KDVs

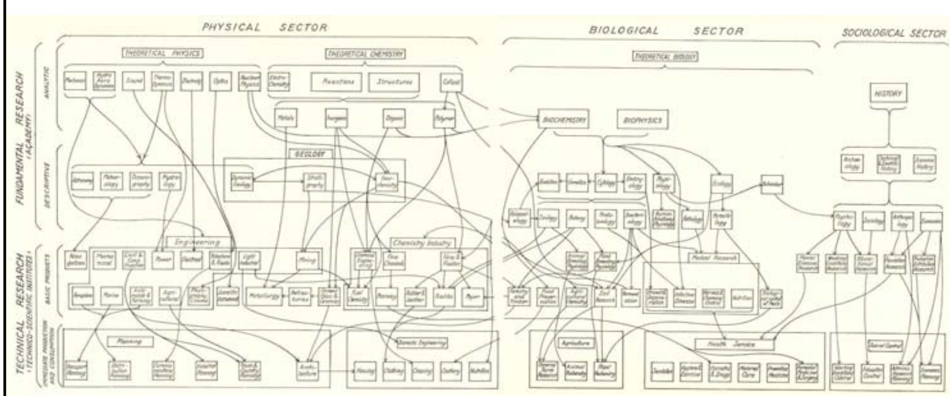


1794

Map of the Various Paths of Life, Benjamin Johnson, Philadelphia, 1805, originally published as a jigsaw puzzle in 1794, Map Collection, Yale University Library.

From: Harmon, Katherine (2004). *You Are Here: Personal Geographies and Other Maps of the Imagination.* New York: Princeton Architectural Press.

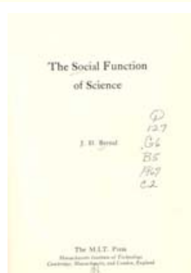


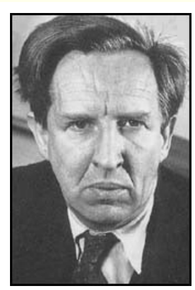


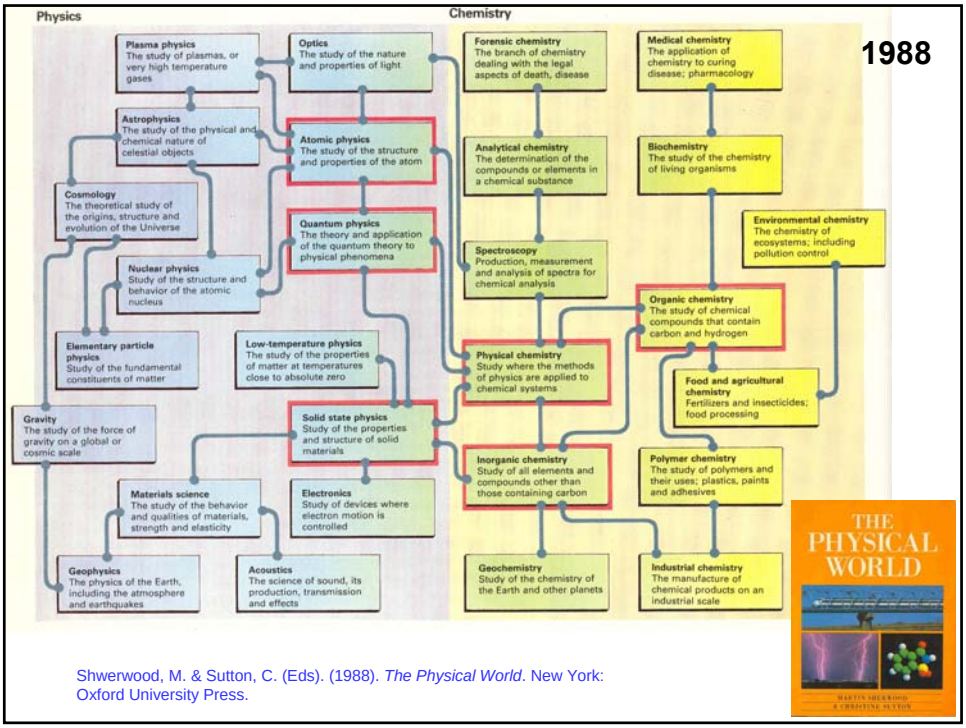
1939

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.







Shwerwood, M. & Sutton, C. (Eds). (1988). *The Physical World*. New York: Oxford University Press.

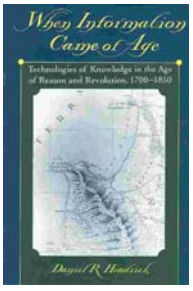
Descriptive → Scientific Cartography

Prior to the 1600's cartography was merely descriptive. Advances in calculating location on the planet made it scientific.

- Triangulation over large distances from a known base measure.
- Using the moons of Jupiter to establish longitude on land.
- Using accurate time pieces to establish longitude at sea.
- Looking below above and below the surface.
- Accurately measuring depth and altitude, and representing them on maps.

The same is now occurring with domain maps.

- We have had descriptive maps for a long time.
- Now they are becoming methodologically rigorous.
- They are becoming scientific.
- This is the involvement of information science and the emerging field of information cartography.



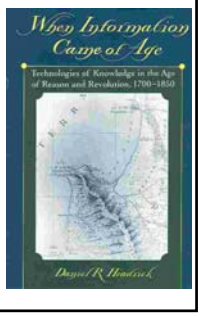
Headrick, Daniel R. (2000). *When Information Came of Age: Technologies of Knowledge in the Age of Reason and Revolution, 1700-1850.* New York: Oxford University Press.

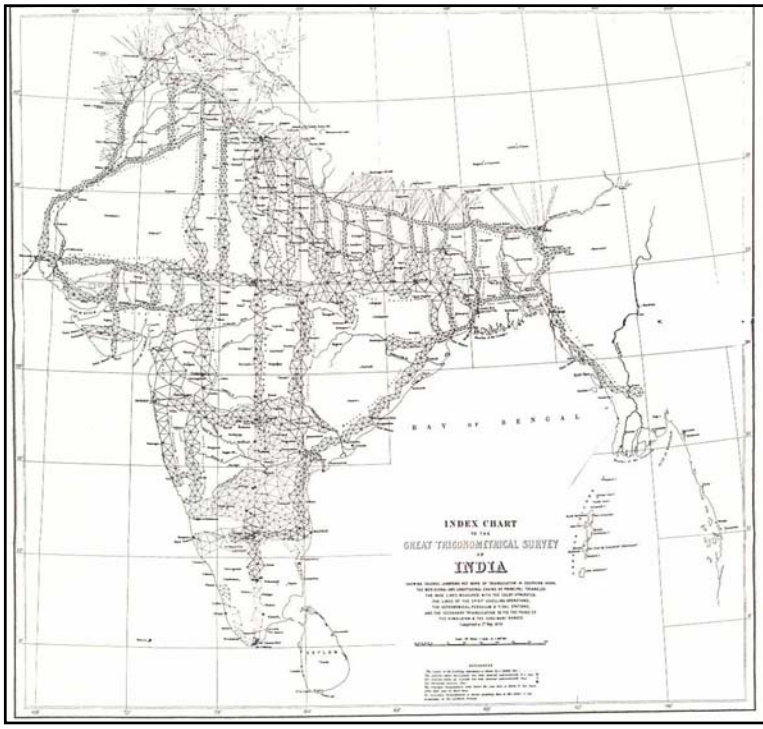
Figure 2. César-François Cassini de Thury, map of France (detail), from trigonometric survey, 1744.

1744

- Rigorous, methodical (scientific) cartography.
- The mapping of France by triangulation.
- César-Francois Cassini de Thury

Headrick, Daniel R. (2000). *When Information Came of Age: Technologies of Knowledge in the Age of Reason and Revolution, 1700-1850.* New York: Oxford University Press.

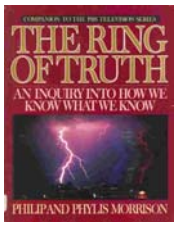


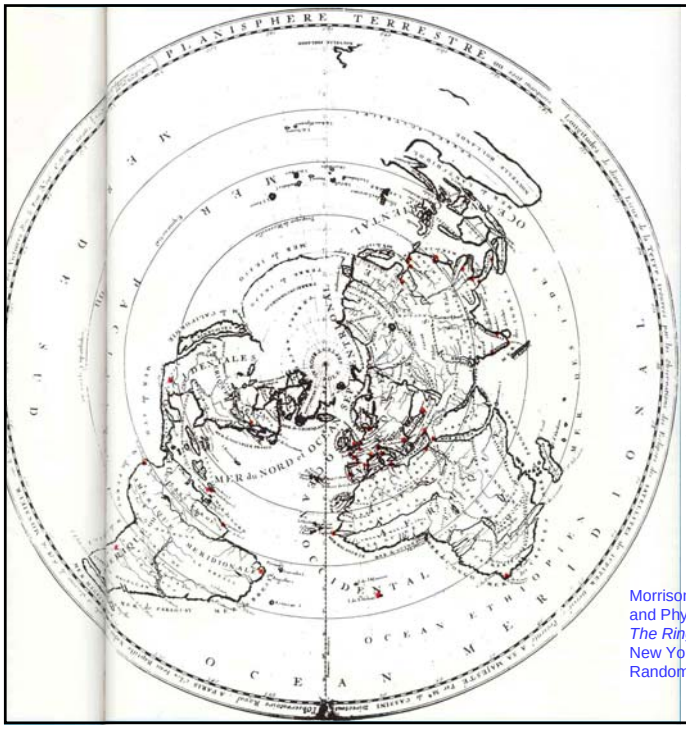


1870

- The mapping of India by triangulation.
- Captain George Everest

Morrison, Philip and Phylis (1987). *The Ring of Truth.* New York: Random House.

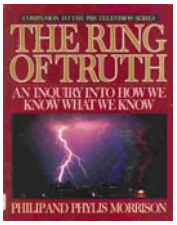




1696

- First Accurate Map of the Earth
- 40 points of accurate longitude
- Based on Moons of Jupiter to compare with local time in Paris.
- Cassini.

Morrison, Philip and Phylis (1987). *The Ring of Truth.* New York: Random House.



1895

1st Otlet and then De Solla Price Contemplate Domain Maps

1895 – Paul Otlet realizes that Melvil Dewey's Decimal Classification could be used to map knowledge domains. Begins work on converting Dewey's system into the more faceted Universal Decimal Classification (UDC).

1918 – Otlet states again that the UDC may be used to create "an immense map of the domains of knowledge" (Otlet, 1918, p. 78).

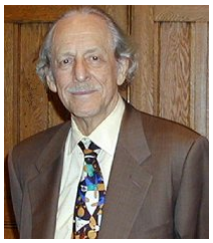
Rayword, B. (1994). Visions of Xanadu: Paul Otlet (1868-1944) and hypertext. JASIS, 45(4), 235-250.

1965 – **Derek De Solla Price** - published contemplation of using bibliometric techniques to create maps of scientific literatures.

De Solla Price, Derek J. (1965). Networks of Scientific Papers *Science*, New Series, Vol. 149, No. 3683. (Jul. 30, 1965), pp. 510-515.

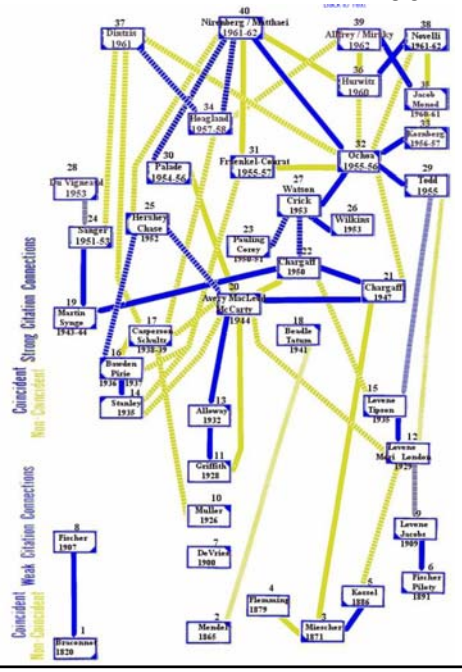
1964

Historiograph of DNA Development
(Garfield, Sher, & Torpie, 1964)
"The Use of Citation Data in Writing the History of Science."
Published by The Institute for Scientific Information, December 1964. Report of research for Air Force Office of Scientific Research under contract F49(638)-1256.



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

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



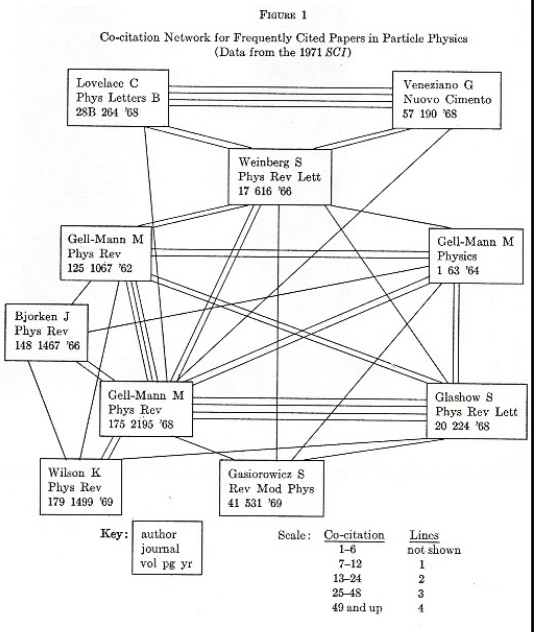
1973

Co – Discoveries / Implementation

Using Co-Citation to create domain maps.

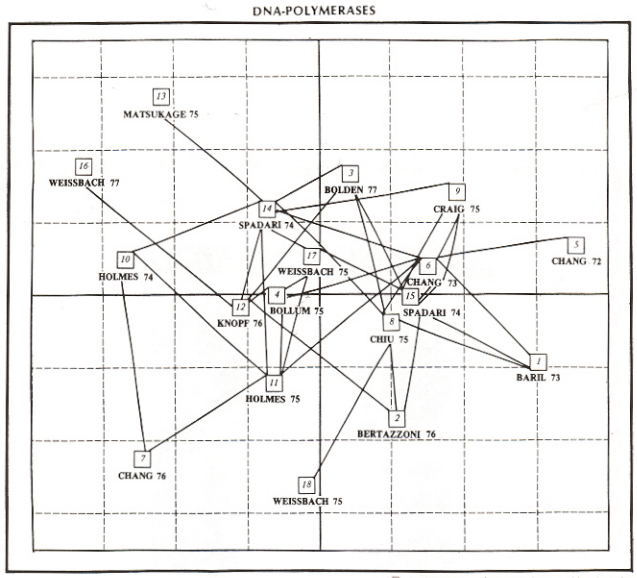
• Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. JASIS, 24, 265-269.

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



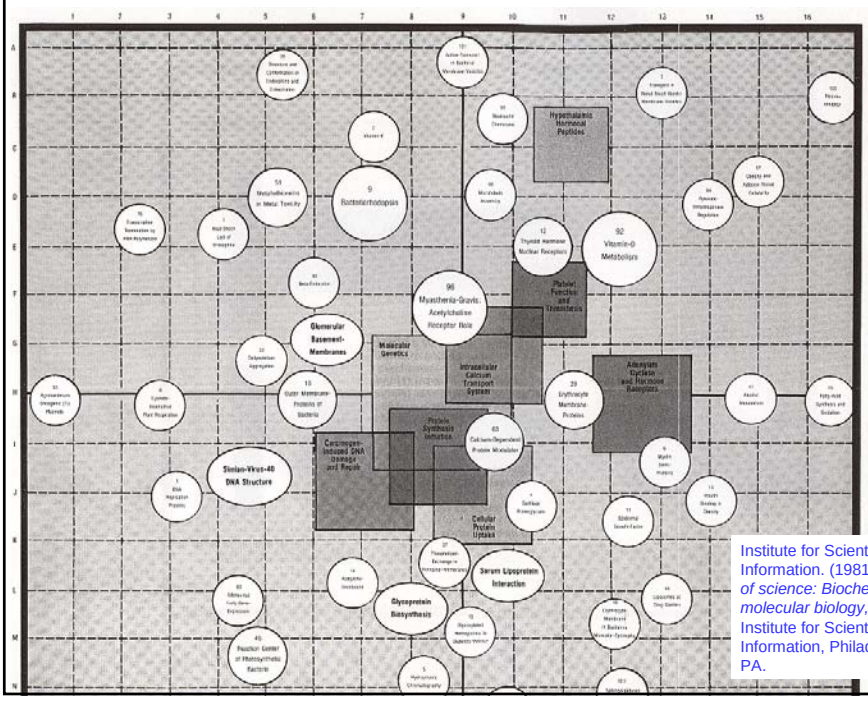
1981

Specialty 27



Institute for Scientific Information. (1981). *ISI atlas of science: Biochemistry and molecular biology, 1978/80*, Institute for Scientific Information, Philadelphia, PA.

1981



Institute for Scientific Information. (1981). *ISI atlas of science: Biochemistry and molecular biology, 1978/80*, Institute for Scientific Information, Philadelphia, PA.

1998

White, Howard & McCain, Katherine. (**1998**). Visualizing a Discipline: An Author Co-citation Analysis of Information Science, 1972-1995, J. Am. Soc. Information Science, 49:4, 327-356.

- IS Dichotomy – Analysts (Bibliometricians) & Retrievalists
- Domain Mapping is a form of Artificial Intelligence

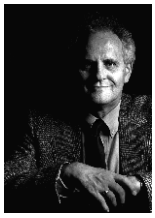


TABLE 1. Journals used to define information science.

Information science
Annual Review of Information Science and Technology
Information Processing & Management (and *Information Storage & Retrieval*)
Journal of the American Society for Information Science
Journal of Documentation
Journal of Information Science
Library & Information Science Research (and *Library Research*)
Proceedings of the American Society for Information Science (and *Proceedings of the ASIS Annual Meeting*)
Scientometrics
Library automation
Electronic Library
Information Technology and Libraries (and *Journal of Library Automation*)
Library Resources & Technical Services
Program—Automated Library and Information Systems

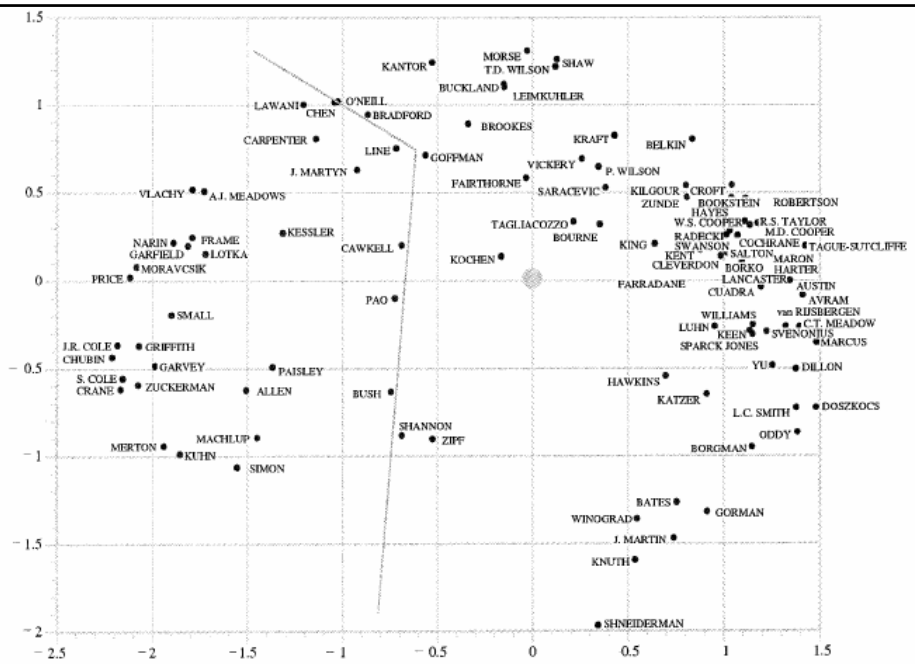
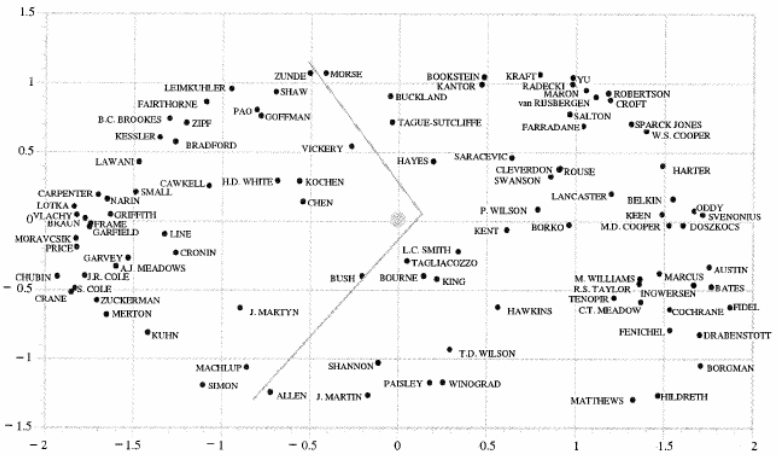


FIG. 2. Top 100 authors in information science, 1972–1979.



Left Side

Entering (3): Braun, Cronin, White

Exiting (0)

Shifted from right (10): Bradford, Brookes, Fairthorne, Goffman, Kochen, Leimkuhler, Shaw, Vickery, Zipf, Zunde

Right Side

Entering (7): Drabenstott, Fenichel, Fidel, Hildreth, Ingwersen, Matthews, Tenopir

Exiting (10): Avram, Cuadra, Dillon, Gorman, Katzer, Kilgour, Knuth, Luhn, O'Neill, Shneiderman

Shifted from left (2): Allen, Paisley

FIG. 3. Top 100 authors in information science, 1980–1987.

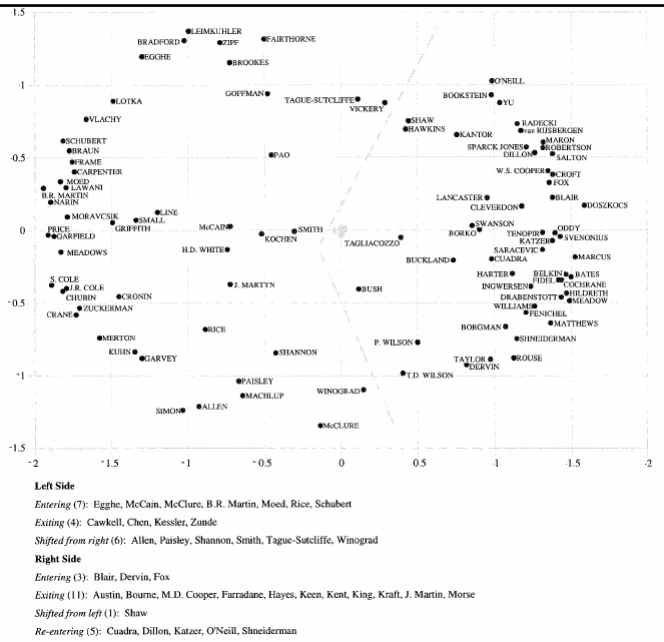


FIG. 4. Top 100 authors in information science, 1988–1995.

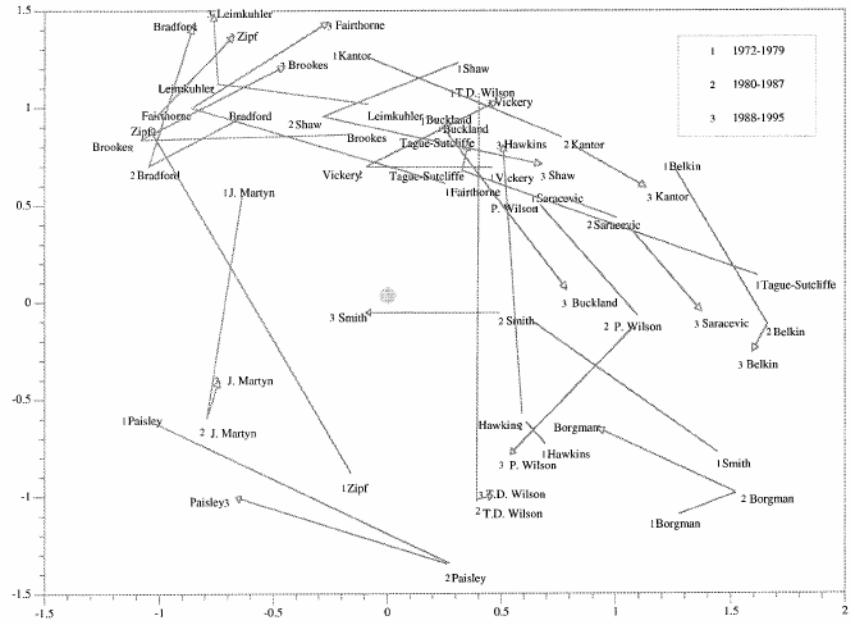


FIG. 5. Shifts in 19 authors' citation images.

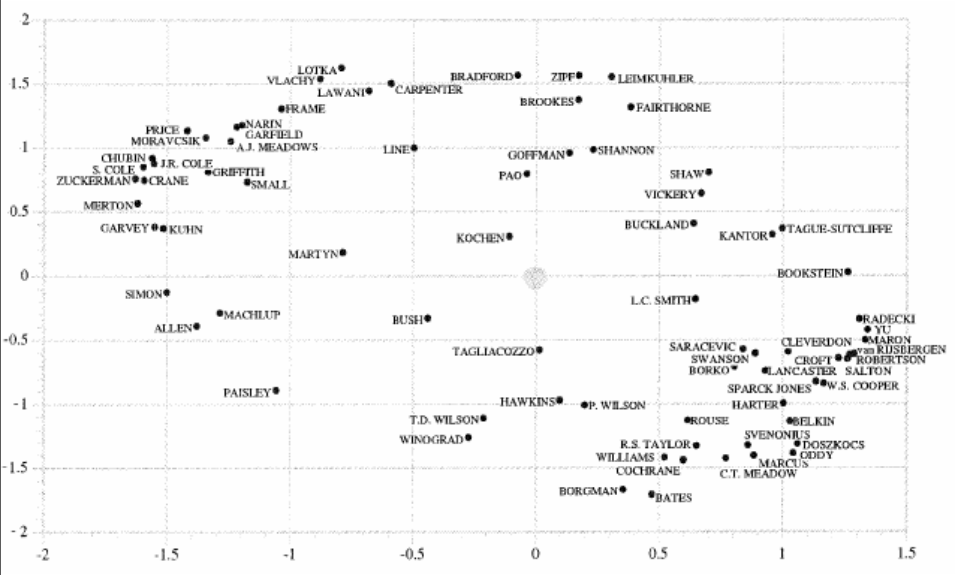


FIG. 6. INDSCAL map of 75 ‘‘canonical’’ information science authors.

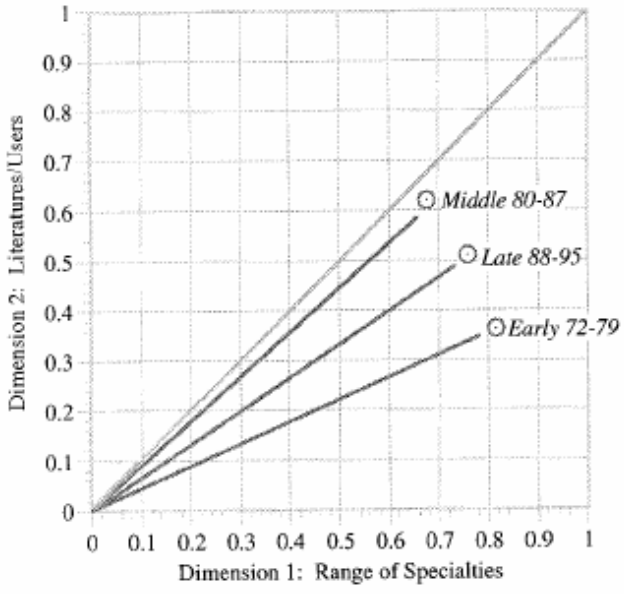


FIG. 7. Changes in importance of axes over three periods.

1999

Small, H. (**1999**). Visualizing Science by Citation Mapping, J. Am. Soc. Information Science, 50:9, 799-813.

1985

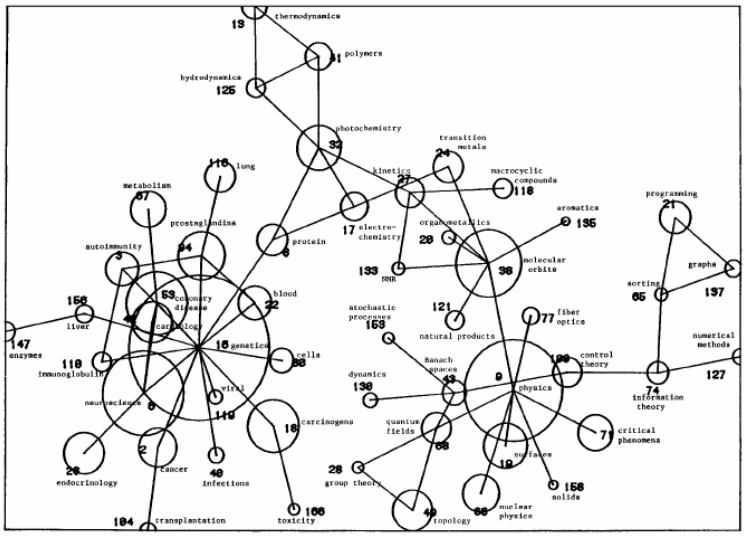


FIG. 1. An early co-citation map of science showing the major disciplines of the natural sciences: Biology, chemistry, and physics. The data are from a five-level co-citation analysis. Multidimensional scaling was used to position macro-clusters (from: Small & Garfield, 1985).

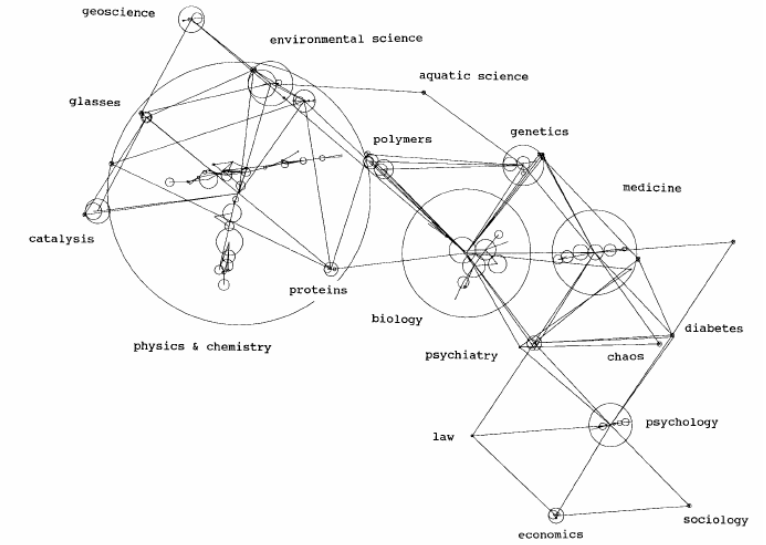


FIG. 3. The map is a two-dimensional computer visualization of 36,720 documents comprising a multidisciplinary sample selected by a fractional citation counting method, covering the years 1981–1995. The circles represent 35 fourth-level clusters, whose internal structure can be explored by zooming into the object. The configuration of third-level clusters is also visible within each fourth-level object. The sizes of the circles reflect the spread of lower level objects contained. Lines represent strong co-citation links among the clusters. Subject matter ranges from social science on the lower right to physical science on the left. Labels are based on a frequency analysis of article titles and journal category names.

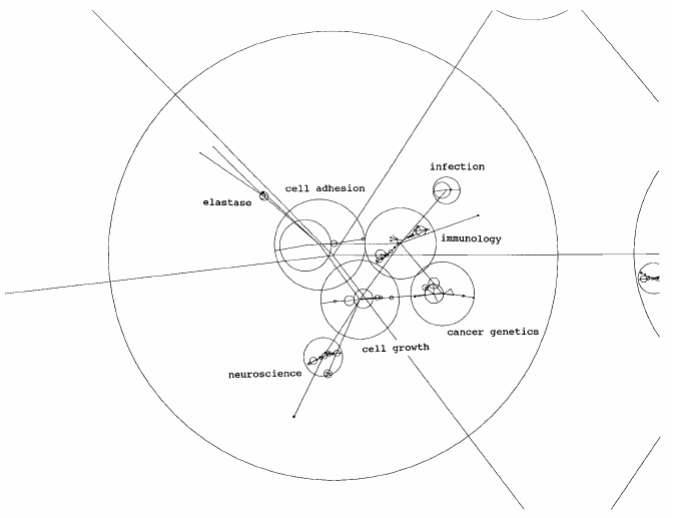


FIG. 4. A detailed view of the large biology circle at the center of Figure 3, showing the level-three clusters within it as well as the level-two objects within each level three cluster. Topics range from immunology to neuroscience.

1999

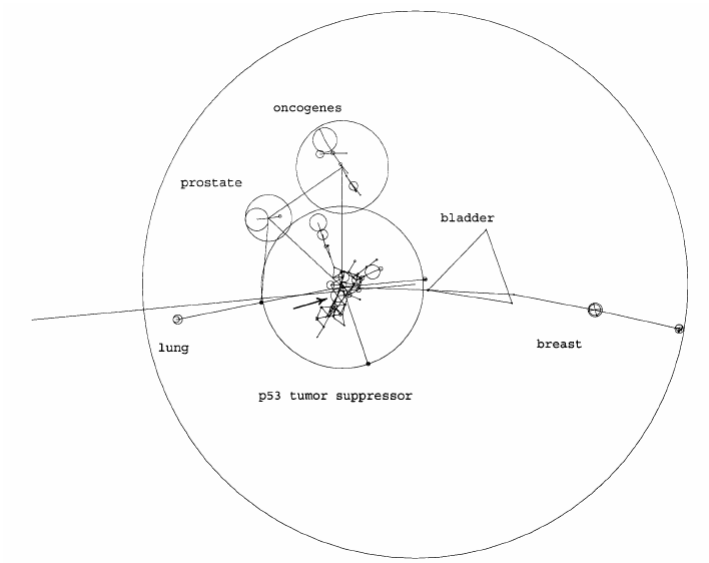


FIG. 5. A detailed view of the level-three cancer genetics cluster on Figure 4, showing the level-two objects contained as well as the level-one objects within them. Clusters dealing with various forms of cancer are arranged around the central cluster on the p53 tumor suppressor protein.

1999

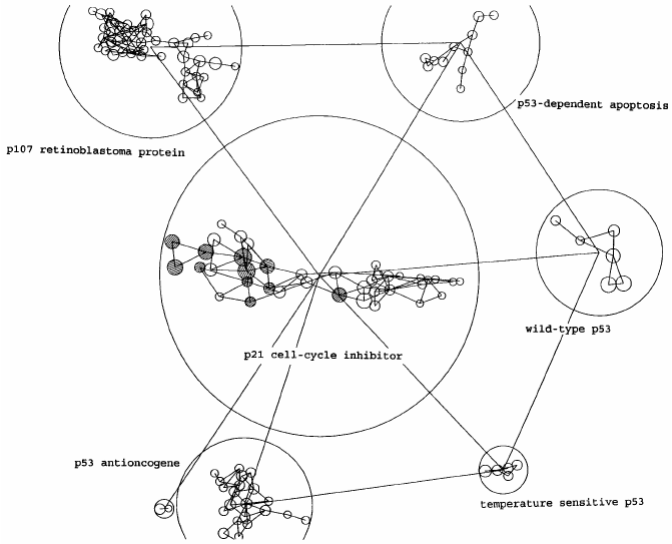


FIG. 6. A detailed view of a group of first-level clusters whose location on Figure 5 is indicated by an arrow within the p53 tumor-suppressor region. The largest level-one cluster is on the p21 protein containing 39 documents. Documents more than two years older than the base year, 1995, are shaded. Unshaded documents are from the two most recent years, 1994 and 1995, making this a “hot” cluster.

How KDVs Are Created

Process Flow for Visualizing KDs

DATA EXTRACTION	UNIT OF ANALYSIS	MEASURES	LAYOUT (often one code does both similarity and ordination steps)		DISPLAY
			SIMILARITY	ORDINATION	

SEARCHES
- ISI
- INSPEC
- Eng Index
- Medline
- ResearchIndex
- Patents
- etc.

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 (words/terms) incl. Singular Value Decomp (SVD)

CORRELATION (if desired)
- Pearson's R on any of above

DIMENSIONALITY REDUCTION
- Eigenvector/ Eigenvalue solutions
- Factor Analysis (FA) and Principal Components Analysis (PCA)
- Multi-dimensional scaling (MDS)
- LSA
- Pathfinder networks (PFNet)
- Self-organizing maps (SOM) includes SOM_ET-maps, etc.

CLUSTER ANALYSIS

SCALAR
- Triangulation
- Force-directed placement (FDP)

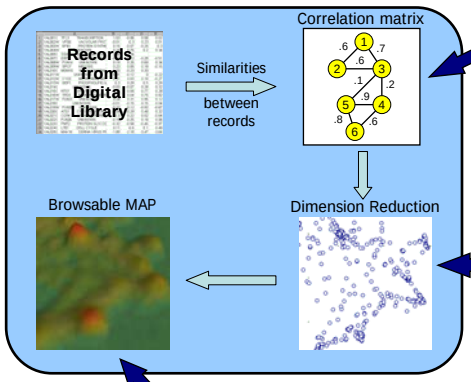
INTERACTION
- Browse
- Pan
- Zoom
- Filter
- Query
- Detail on demand

ANALYSIS



Börner, K., Chen, C., & Boyack, K.W. (2003). Visualizing Knowledge Domains. In *Annual Review of Information Science and Technology*, 37 (B. Cronin, ed.), Information Today, Medford, NJ, pp. 179-255.

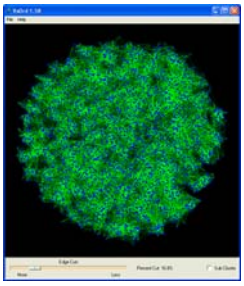
Process Used by Boyack



Common index values such as Cosine

$N_{ij} / \text{sqrt}(N_i N_j)$

VxOrd



VxOrd: Ordination Algorithm

- Force-directed placement
  - Each object tries to minimize an energy equation using a solution space exploration algorithm

$$E_{x,y} = \left[\sum_{i=0}^{n} \left(w_i * l_i^{\,2} \right) \right] + D_{x,y}$$

n = number of edges connected to node

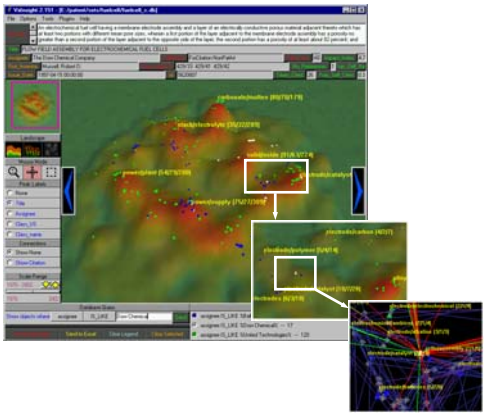
w_i = weight of edge i

l_i = euclidean length of edge i

$D_{x,y}$ = density of objects at/near coordinate x,y

VxInsight – Knowledge Visualization

- Displays graph structures using an intuitive terrain metaphor or as scatterplot
- Enables analysts to navigate and explore graph structures at multiple levels of detail through drill-down
- Can show multiple types of associations or linkages

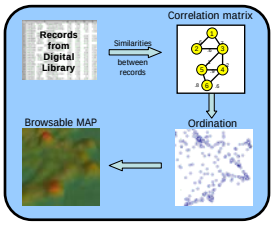


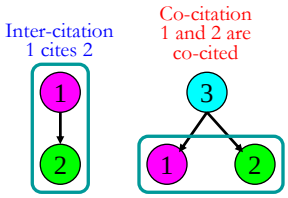
Strategy

- Develop and **validate** process, methods, and algorithms at small scale (~10k objects)
  - Macro-model
  - Using ISI citation data, create disciplinary maps of science using journals (~7000 titles)
  - Validate using the known journal categorization structure
- Employ validated process, methods, and algorithms at larger scale (~1M objects)
  - Micro-model
  - Create paper-level (~1M annually) maps of science from ISI citation data
  - Validate detailed maps at local structural levels where possible
  - Calculate indicators and metrics at the cluster or community level

Macro-model Process

- Identify individual journals
- Calculate similarity between journals from inter-citation data and co-citation data
- Use VxOrd to determine coordinates for each journal
- Generate cluster assignments (k-means)
- Validate against ISI journal category assignments





Algorithms – Similarity Measures

IC-Raw

$$RAW_{i,j} = RAW_{j,i} = C_{i,j} + C_{j,i} \ ,$$

IC-Cosine

$$COS_{i,j} = COS_{j,i} = \frac{\left(RAW_{i,j}\right)}{\sqrt{\sum_{k=1}^{n} C_{i,k} \sum_{k=1}^{n} C_{j,k}}} \ ,$$

IC-Jaccard

$$JAC_{i,j} = JAC_{j,i} = \frac{\left(RAW_{i,j}\right)}{\sum_{k=1}^{n} C_{i,k} + \sum_{k=1}^{n} C_{j,k} - \left(RAW_{i,j}\right)} \ ,$$

IC-Pearson

$$r_{i,j} = \frac{\sum_{k=1}^{n} \left(RAW_{i,k} - \overline{RAW_i}\right)\left(RAW_{j,k} - \overline{RAW_j}\right)}{\sqrt{\sum_{k=1}^{n} \left(RAW_{i,k} - \overline{RAW_i}\right)^2 \sum_{k=1}^{n} \left(RAW_{j,k} - \overline{RAW_j}\right)^2}} \ ,$$

where $\overline{RAW_i} = \frac{1}{n} \sum_{k=1}^{n} RAW_{i,k}, \quad k \neq i \ ,$

IC-RFavg

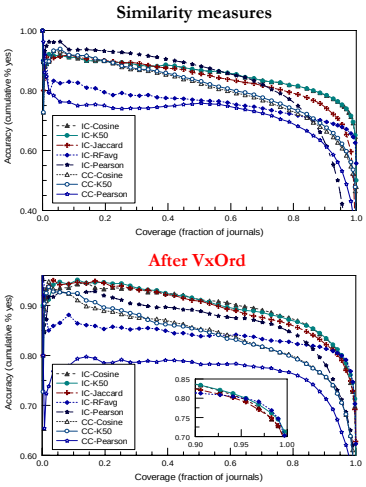
$$RFA_{i,j} = RFA_{j,i} = \left(RF_{i,j} + RF_{j,i}\right) / 2 \ ,$$

where $RF_{i,j} = 10^6 * C_{i,j} \Big/ \left(N_j \sum_{k=1}^{n} C_{i,k} \right) \ .$

Boyack, K.W., Klavans, R., & Börner, K., (2005, in press). Mapping the backbone of science. *Scientometrics*.

Macro-model: Local Accuracy

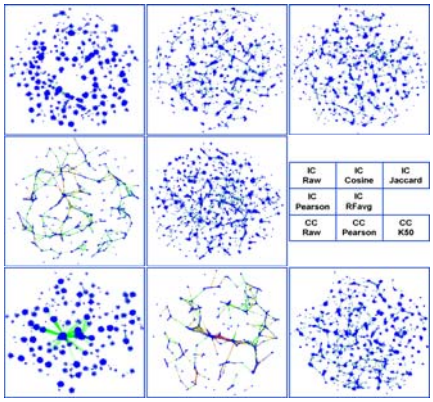
- For each similarity measure, journal pairs were assigned a 1/0 binary score if they were IN/OUT of the same ISI category
- Accuracy vs. coverage curves were generated for each similarity measure
- For each similarity measure, distances (in the VxOrd layouts) between journal pairs were calculated
- Accuracy vs. coverage curves were generated for each re-estimated (distance) similarity measure
- Results after running through VxOrd were more accurate than the raw measures
- Inter-citation measures are best



Klavans, R., & Boyack, K.W. (2005, in press). Identifying a better measure of relatedness for mapping science. *Journal of the American Society for Information Science and Technology.*

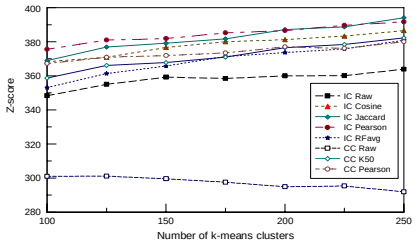
Macro-model: Different Similarity Metrics

- ISI file year 2000, SCIE and SSCI
- Ten different similarity metrics
  - 6 Inter-citation (raw counts, cosine, modified cosine, Jaccard, RF, Pearson)
  - 4 Co-citation (raw counts, cosine, modified cosine, Pearson)
- Inter-citation gives structure based on current citing patterns
- Co-citation gives structure based on how science is currently used



Macro-model: Regional Accuracy

- For each similarity measure, the VxOrd layout was subjected to k-means clustering using different numbers of clusters
- Resulting cluster/category memberships were compared to actual category memberships using entropy/mutual information method
- Increasing Z-score indicates increasing distance from a random solution
- Most similarity measures are within several percent of each other



Boyack, K.W., Klavans, R., & Börner, K., (2005, in press). Mapping the backbone of science. *Scientometrics*.

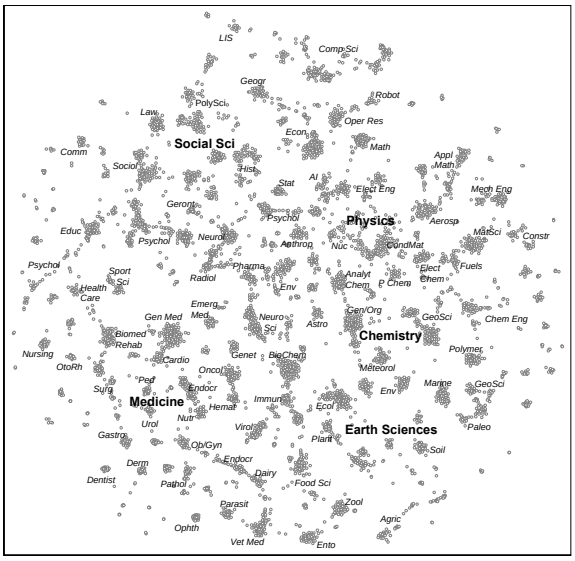
Computing Mutual Information

- Use method of Gibbons and Roth (Genome Research v. 12, pp. 1574-1581, 2002)
- K-means clustering (MATLAB) for each graph layout
  - 8 different similarity measures
  - 3 different k-means runs at 100, 125, 150, 175, 200, 225, 250 clusters
- Quality metric (mutual information) calculated as
  - MI(X,Y) = H(X) + H(Y) – H(X,Y)
  - where H = - $\sum$ *Pi* log2 *Pi*
  - *Pi* are the probabilities of each [cluster, category] combination
  - X (known ISI category assignments), Y (k-means cluster assignments)
- Z-score (indicates distance from randomness, Z=0=random)
  - Z = (MIreal – MIrandom)/ Srandom
  - MIrandom and Srandom vary with number of clusters, calculated from 5000 random solutions

Macro-model: "Best" Map

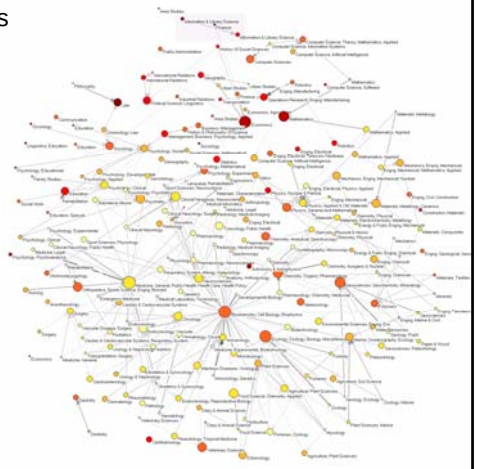
- Each dot is one journal
- Journals group by discipline
- Labeled by hand
- Generated using the *IC-Jaccard* similarity measure.
- The map is comprised of 7,121 journals from year 2000.
- Large font size labels identify major areas of science.
- Small labels denote the disciplinary topics of nearby large clusters of journals.

Boyack, K.W., Klavans, R., & Börner, K. (2005, in press). Mapping the backbone of science. *Scientometrics.*

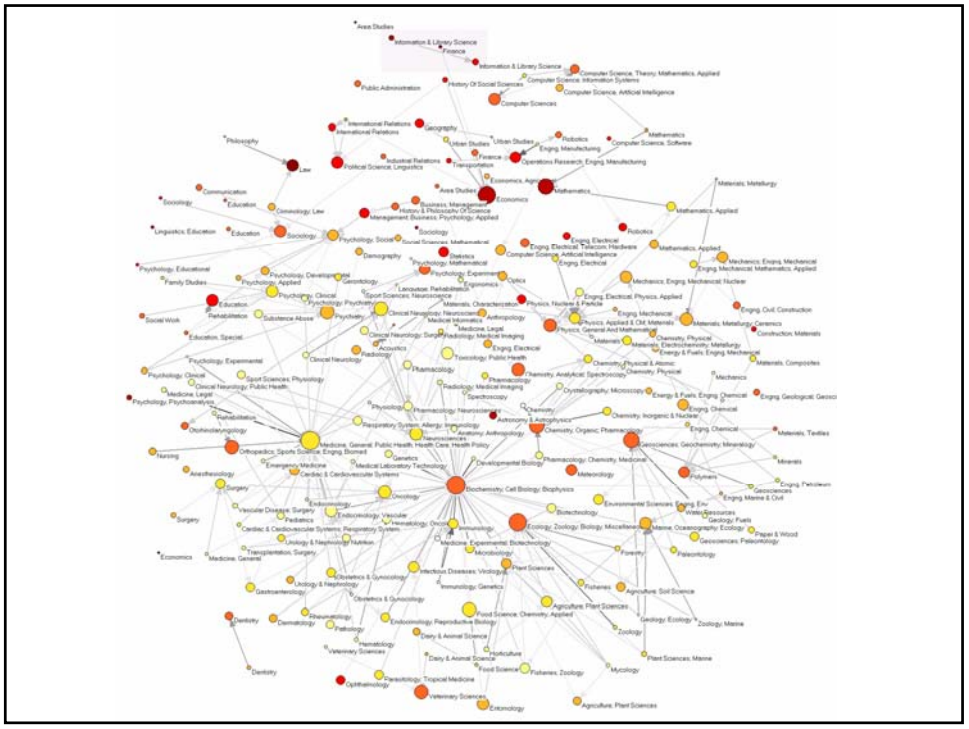


Macro-model: Structural Map

- Clusters of journals denote 212 disciplines (7000 journals).
- Labeled with their dominant ISI category names.
- Circle sizes (area) denote the number of journals in each cluster.
- Circle color depicts the independence of each cluster, with darker colors depicting greater independence.
- Lines denote strongest relationships between disciplines (citing cluster gives more than 7.5% of its total citations to the cited cluster).
- Enables disciplinary diffusion studies.
- Enables comparison of institutions by discipline.

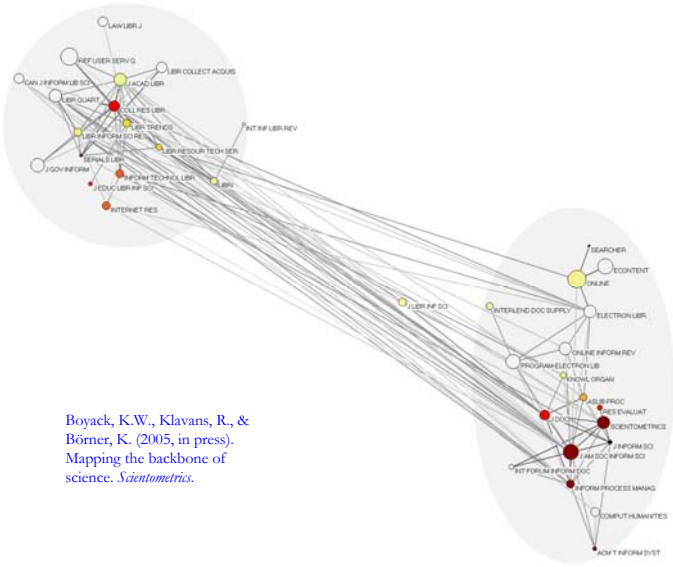


Boyack, K.W., Klavans, R., & Börner, K. (2005, in press). Mapping the backbone of science. *Scientometrics.*



Macro-model: Detail

- Clusters of journals denote disciplines
- Lines denote strongest relationships between journals



Boyack, K.W., Klavans, R., & Börner, K. (2005, in press). Mapping the backbone of science. *Scientometrics.*

What Comes Next?

(1) Further Refinements

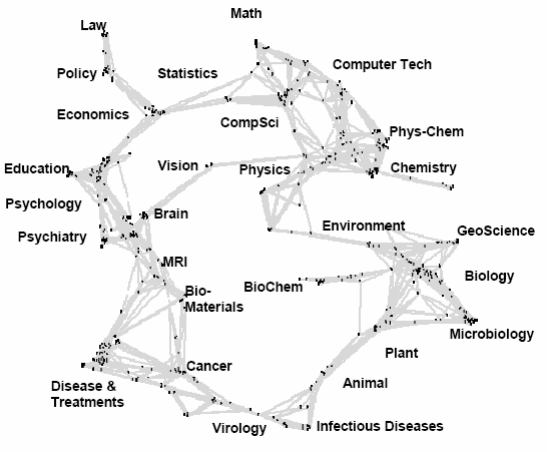
(2) Different Visualizations

(3) Time series to capture the evolution of disciplines

(4) Larger Datasets – Incorporation of patent and grant funding data

(5) A new era in information cartography

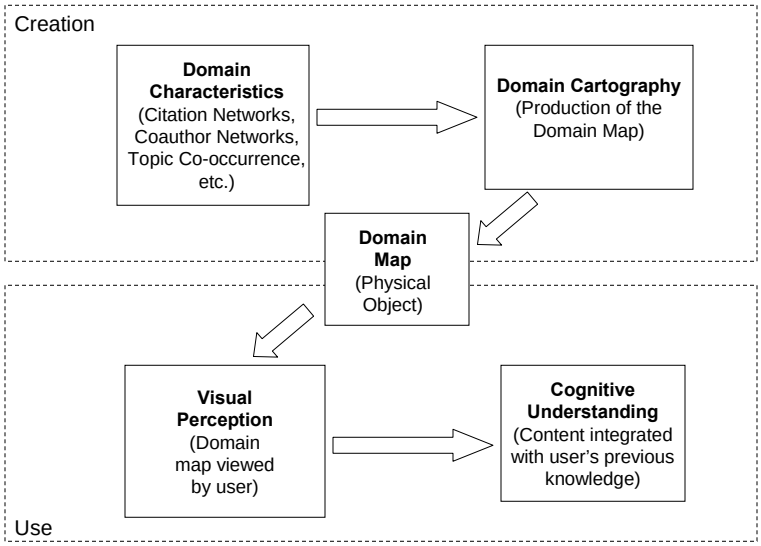
(6) Widespread educational uses of knowledge domain maps.



Uses combined SCIE/SSCI data from 2002.
http://vw.indiana.edu/aag05/slides/boyack.pdf

How KDVs Are Used

Process of KDV Creation and Usage



Domain Visualizations Are Used For …

		QUESTIONS RELATED TO: Fields and paradigms	Communities and networks	Research performance or competitive advantage	Commonly used algorithms
UNIT OF ANALYSIS	Authors		Social structure, intellectual structure, some dynamics	Use network characteristics as indicators	Social network packages, MDS, factor analysis, Pathfinder networks
	Documents	Field structure, dynamics, paradigm development		Use field mapping with indicators	Co-citation, co-term, vector space, LSA, PCA, various clustering methods
	Journals	Science structure, dynamics, classification, diffusion between fields			Co-citation, intercitation
	Words		Cognitive structure, dynamics		Vector space, LSA, LDA (20)
	Indicators and metrics			Comparisons of fields, institutions, countries, etc., input-output	Counts, correlations

Boyack, K.W. (2004). Mapping Knowledge Domains: Characterizing PNAS. *Proceedings of the National Academy of Sciences of the US*, 101(S1), 5192-5199.

Aside: Citation Mapping Comes of Age

- PNAS online interface now generates a citation map for some of its articles.

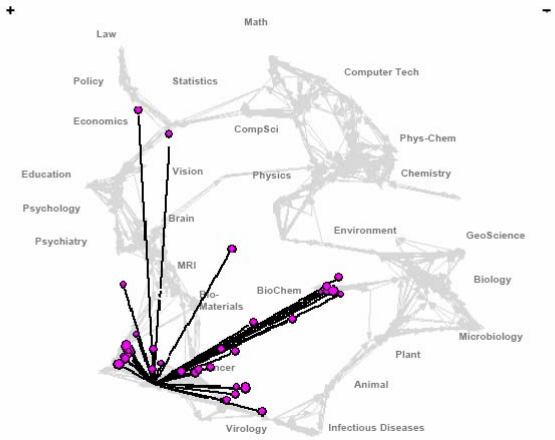
Boyack, K.W. (2004). Mapping Knowledge Domains: Characterizing PNAS. *Proceedings of the National Academy of Sciences of the US*, 101(S1), 5192-5199.

Goals of Sandia Science Mapping Project

- Create maps of science with indicators of innovation, risk, and impact at the research community level
- Enable better R&D through:
  - Identification and evaluation of current work in a global context
  - Identification of highly-ranked communities in areas related to current work
  - Identification and evaluation of proposed work in a global context
  - Identification of research entry points (or potential collaborators) and emerging applications in our areas of focus
  - Identification of opportunity and vulnerability using institutional comparisons
  - Better understanding of the innovation process and better anticipation of future trends(?)

Identifying Opportunities/Threats

(36 Research Communities that will impact GI Research…
that GI Researchers are least likely to be aware of)



http://vw.indiana.edu/aag05/slides/boyack.pdf

Educational Aspects of KDVs

Bridging the Practices of Two Communities

- Domain Maps are widely used in the field of education.
- Called Concept Maps.

(Also called: mind maps, pattern notes, brain patterns, spider maps, networks, semantic maps, semantic networks, and semantic webs.)

- Differ from domain maps produced by information scientists as to scale and method of production.
- Each community has something to offer the other.

Current Practice in Education

- "[L]earning best begins with a big picture, a schema, a holistic cognitive structure, which should be included in the lesson material[.]"
- "[T]he designers' task is to develop a big picture and emphasize it[.]" (West et al., (1991). *Instructional Design: Implications from Cognitive Science.* Englewood Cliffs, New Jersey: Prentice Hall, p. 58)."
- This is done with concept maps.

Benefits of the Big Picture

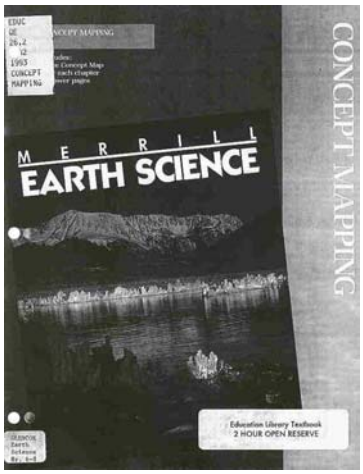
- Provides a structure or scaffolding that students may use to organize the details of a particular subject.
- Information is better assimilated with the student's existing knowledge.
- Visualization enhances recall.
- Makes explicit the connections between conceptual subparts and how they are related to the whole.
- Helps to signal to the student which concepts are most important to learn.

Dual Coding Theory of Memory

- Humans store textual and visual information in different areas of the brain. (Paivio, 1987; Kulhavey & Stock, 1994).
- Visual/spatial images trigger memory of textual elements, and text triggers memory of images and locations.

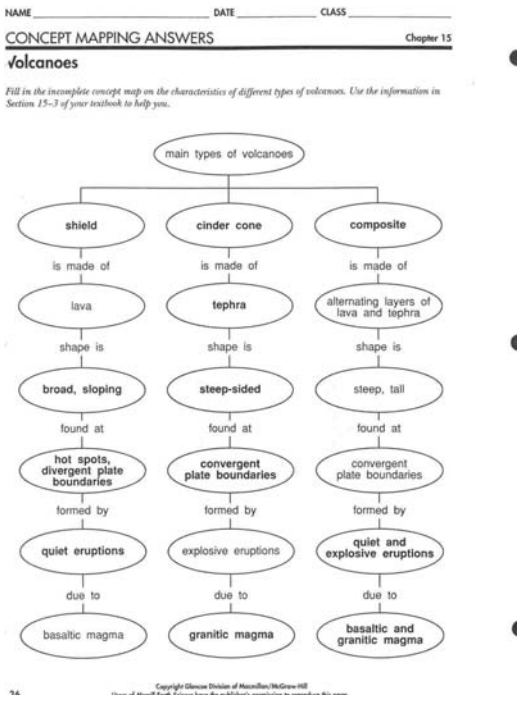
Semantic Network Theory of Learning

- Human memory is organized into networks consisting of interlinked nodes.
- Nodes are concepts or individual words.
- The interlinking of nodes forms knowledge structures or schemas.
- Learning is the process of building new knowledge structures by acquiring new nodes.
- When learners form links between new and existing knowledge, the new knowledge is integrated and comprehended.



GRADES 6-8

Feather, Ralph M. Jr., Snyder, Susan Leach & Hesser, Dale T. (1993). Concept Mapping, workbook to accompany, Merrill Earth Science. Lake Forest, Illinois: Glencoe.



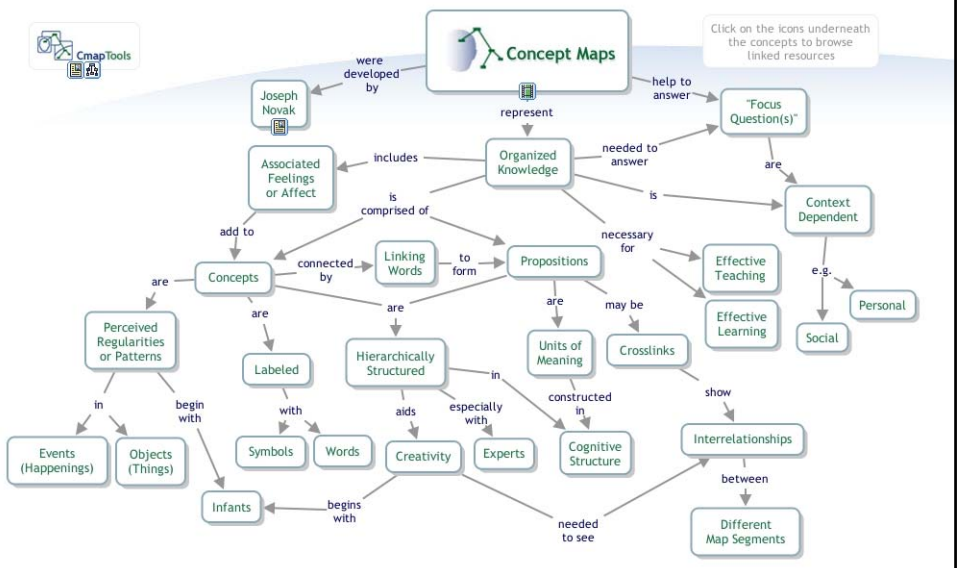
NAME _____ DATE _____ CLASS _____

CONCEPT MAPPING ANSWERS Chapter 15

Volcanoes

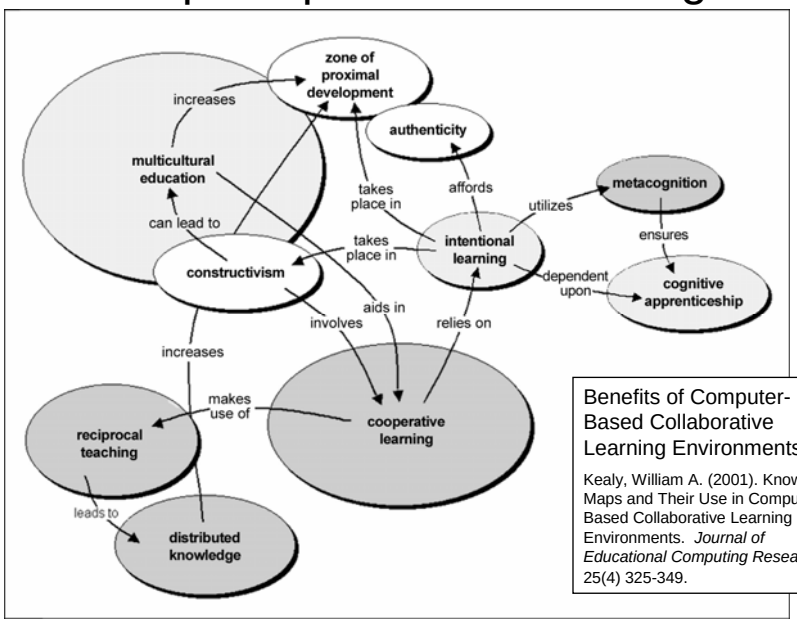
Fill in the incomplete concept map on the characteristics of different types of volcanoes. Use the information in Section 15–3 of your textbook to help you.

Concept Map Produced by Cmap Tools



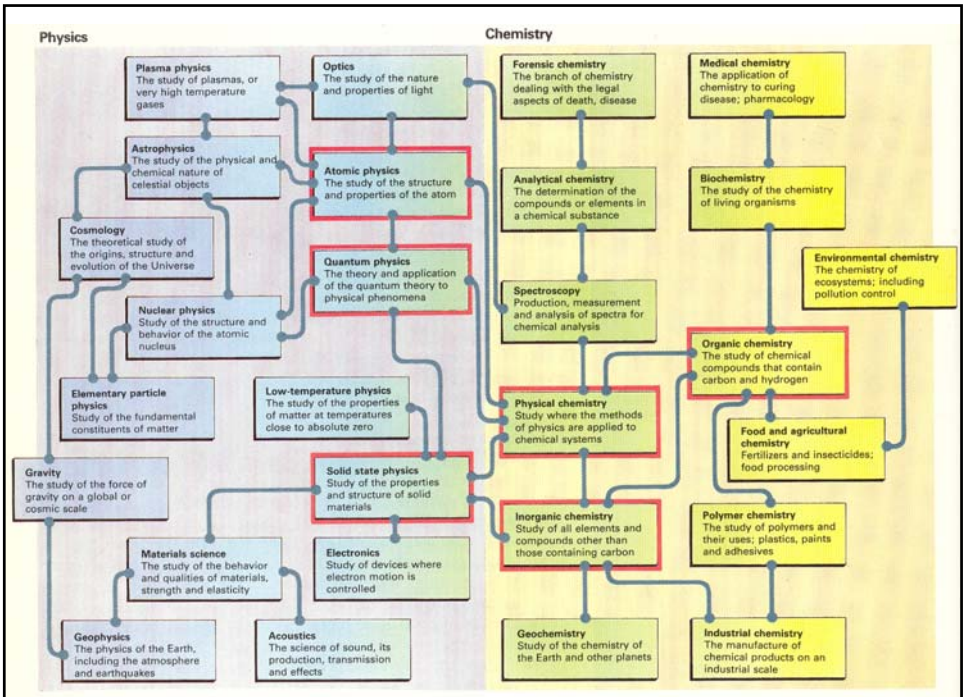
Created by Joseph Novak and rendered with CMapTools. http://cmap.ihmc.us/

Concept Map Created With Rigor



Benefits of Computer-Based Collaborative Learning Environments

Kealy, William A. (2001). Knowledge Maps and Their Use in Computer-Based Collaborative Learning Environments. *Journal of Educational Computing Research.* 25(4) 325-349.



Sherwood, Martin & Sutton, Christine, eds. (1988). *The Physical World.* New York: Oxford University Press.

Elements of Concept Maps

(1) **shapes or nodes**--representing core elements of a concept

(2) **connectors or links** between the shapes or nodes

(3) **connecting words**--that describe how two nodes are related, and

(4) **patterns**--such as a hierarchical or circular ordering of the nodes

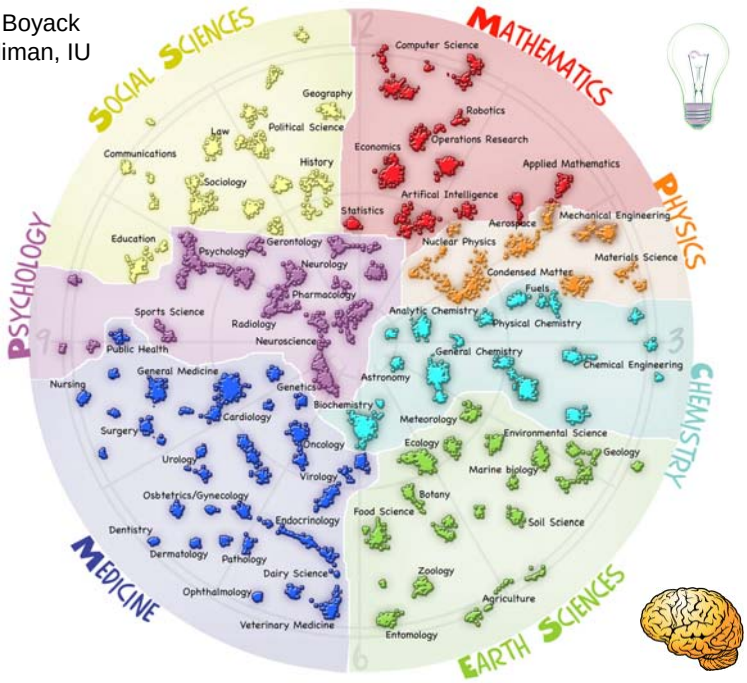
Education → KDVis

- Educational community places more emphasis on explicit connections between nodes and labels for the type of connection.
- Use of action phrases such as “gives rise to,” and “causes.”
- Explicitly labeled connections, linkages, or notion of causality are largely missing from KDVs.
- Also often missing are clearly defined, bounded regions for cognitive chunking and assimilation.

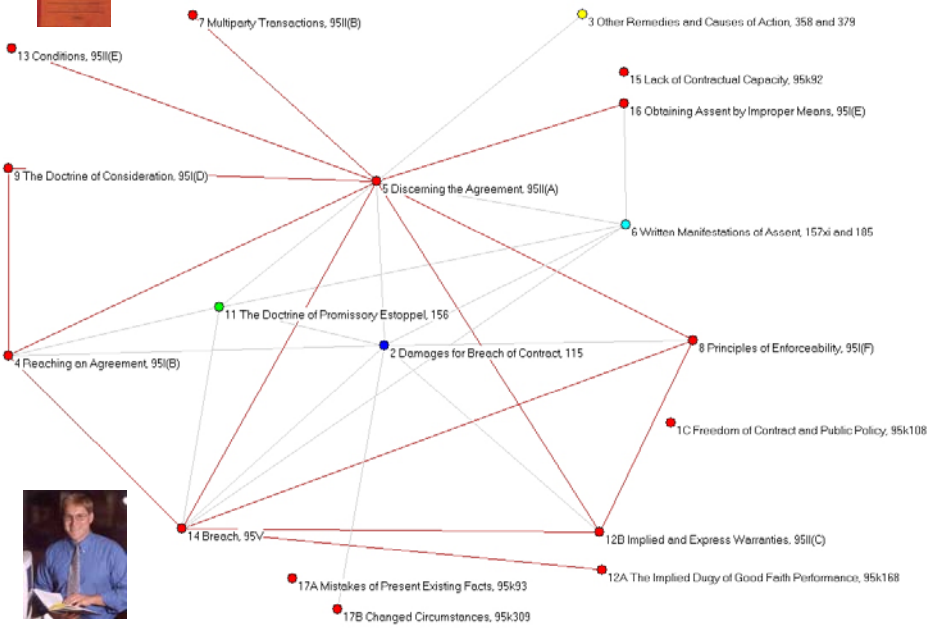
KDVis → Education

- Automated techniques for data harvesting, data processing, and information visualization.
- Larger scale, really ‘Big Picture,’ domain maps (all of science).
- Can market our contribution as concept maps, an educational tool with which educators are already familiar.

Modified from Boyack et al. by Ian Aliman, IU



Domain Map Substrate (Contracts Law)



Metro Map Metaphor

Spatial Representations of Domains

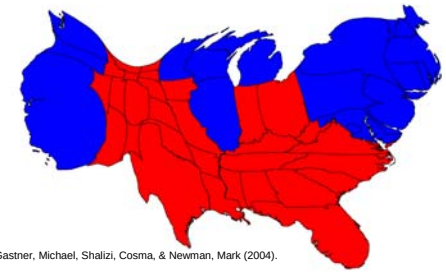
Composed of Two Parts:

- Spatial Substrate
- Thematic Overlay

True for:

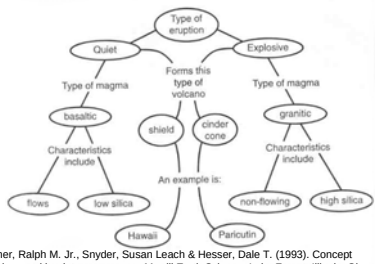
- Thematic Maps (cartography)
- Concept Maps (education, psychology, wayfinding)
- KDVs (information science)
- Metro Map Metaphor Information Spaces (information science)

Thematic Map



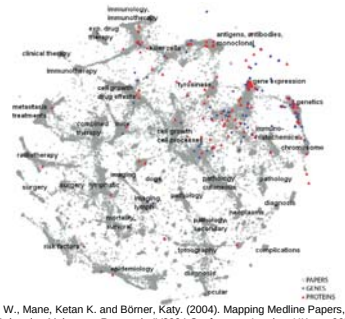
Gastner, Michael, Shalizi, Cosma, & Newman, Mark (2004).

Concept Map



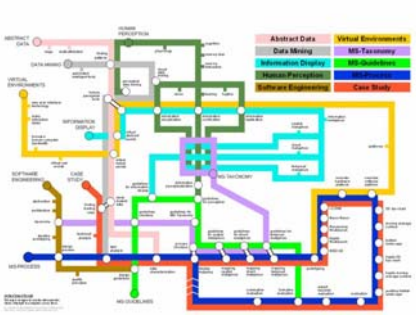
Feather, Ralph M. Jr., Snyder, Susan Leach & Hesser, Dale T. (1993). Concept Mapping, workbook to accompany, Merrill Earth Science. Lake Forest, Illinois: Glencoe.

KDVis



Boyack, Kevin W., Mane, Ketan K. and Börner, Katy. (2004). Mapping Medline Papers, Genes, and Proteins Related to Melanoma Research. IV2004 Conference, London, UK, pp. 965-971

Metro Map Metaphor



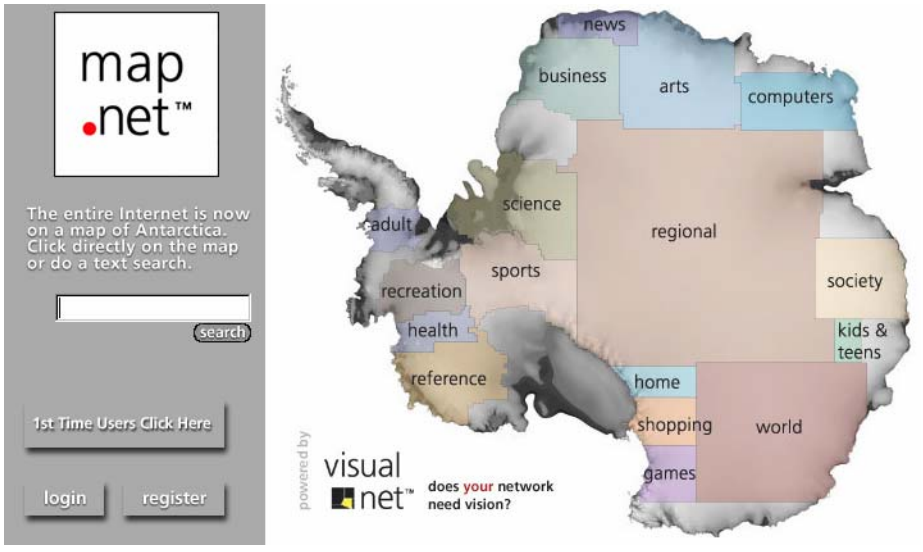
Nesbitt, Keith V. (2004). Getting to more Abstract Places using the Metro Map Metaphor. (IV'04). IEEE.

Comparison of the Spatial Representations of Information

Map Type	Substrate			Overlay
	Usage of nodes & edges	Rigorous spatial layout	Labeled connections (causality, equivalence, similarity, etc.)	
Thematic Maps	rarely	yes	no	yes
Concept Maps	yes	very rare	yes	yes
Knowledge Domain Visualizations	yes	yes	rarely	yes
Metro Map Information Spaces	yes	potentially	no	yes

Iconographic Substrates

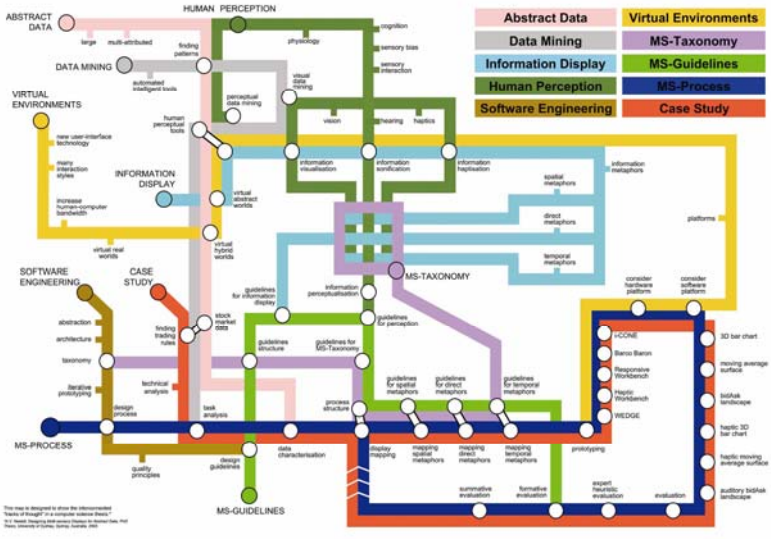
- Information space superimposed onto actual and familiar spaces.
- Analogous to the memory palace device for facilitating recall.
- Associations with the actual place can amplify the meaning of the information space and heighten memory.



DC BLOGS

http://www.reenhead.com/map/metroblogmap.html

2004



Intersecting threads of thought of Keith Nesbitt's doctoral dissertation.

Employing the metro map metaphor.

Nesbitt (2004). "Getting to more Abstract Places using the Metro Map Metaphor," IV'04, 8th International Conference on Information Visualisation London,2004.

Popularity of Metro Map Metaphor

- Explicit Linkages – Does not rely on the distance similarity metaphor.
- Highlights connectivity
- Familiar device to many people.
- Beck – Abstracted the information space from reality.
- Modern trend – Impose abstraction on reality.

Predictions

Six Phases

- in the use, adoption, and implementation of KDVs.
- The first two have already occurred.
- The remaining four are prospective—compelled by the cognitive benefits of KDVs.


Hook, Peter A. and Börner, Katy. (in press) Educational Knowledge Domain Visualizations: Tools to Navigate, Understand, and Internalize the Structure of Scholarly Knowledge and Expertise. In Amanda Spink and Charles Cole (eds.) New Directions in Cognitive Information Retrieval. Springer-Verlag. http://ella.slis.indiana.edu/~pahook/product/05-educ-kdvis.pdf

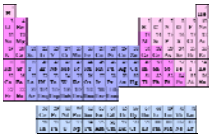
Phases 1 & 2

(1) Bibliometricians realized that they could use bibliographic datasets and techniques such as author co-occurrence to provide maps of a particular discipline.

  – These early maps were graphically simple and painstakingly created by hand.

(2) The implementation of automated techniques for data harvesting, processing, and information visualization.

  – This has facilitated the mapping of larger domains.



Phases 3 & 4

(3) Domain maps will become widely known outside of information Science.

- BBC Article -- Whitehouse, D. (2004). *Scientists Seek 'Map of Science.'* BBC News. Wednesday, 7 April, 2004. Available at: http://news.bbc.co.uk/1/hi/sci/tech/3608385.stm
- They will become popular with educators and will be used to enhance classroom pedagogy.
- They will be as common as regular wall maps and the periodic table of the elements.

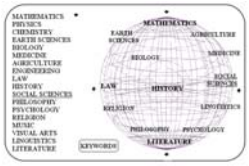
(4) The widespread use of domain maps will lead to steps to harmonize and better preserve the scholarly data from which they are created.

- This might include unique author identification numbers,
- better and standardized citation practices,
- and repositories containing information such as all of the sources cited in books.

Phases 5 & 6

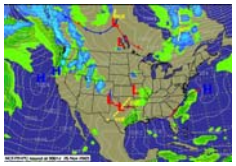
(5) Domain maps will routinely be used as one of the access options to digital libraries and (OPAC's).

- Even if a user chooses to do a keyword search, domain maps will be dynamically displayed unobtrusively in the background, subtly conveying to the user the intellectual landscape of the particular domain and the specific neighborhood of the user's search interest.



(6) Dynamic domain maps will capture and portray the diffusion of information.

- This diffusion of knowledge may occur from one author to the next, one journal to another, or among scientific disciplines, etc.
- Domain maps will have predictive elements that will forecast and model the spread of knowledge.
- They will be used widely for science forecasts in a similar fashion as today's weather forecast maps.



The End

Slides Available:

- http://ella.slis.indiana.edu/~pahook

Special thanks to:

Katy Börner

Kevin Boyack