

Mapping Knowledge Domains



Katy Börner

School of Library and Information Science

INDIANA UNIVERSITY

BLOOMINGTON

katy@indiana.edu

Talk at IU's Technology Transfer Office
Indianapolis, IN, July 12th, 2005.

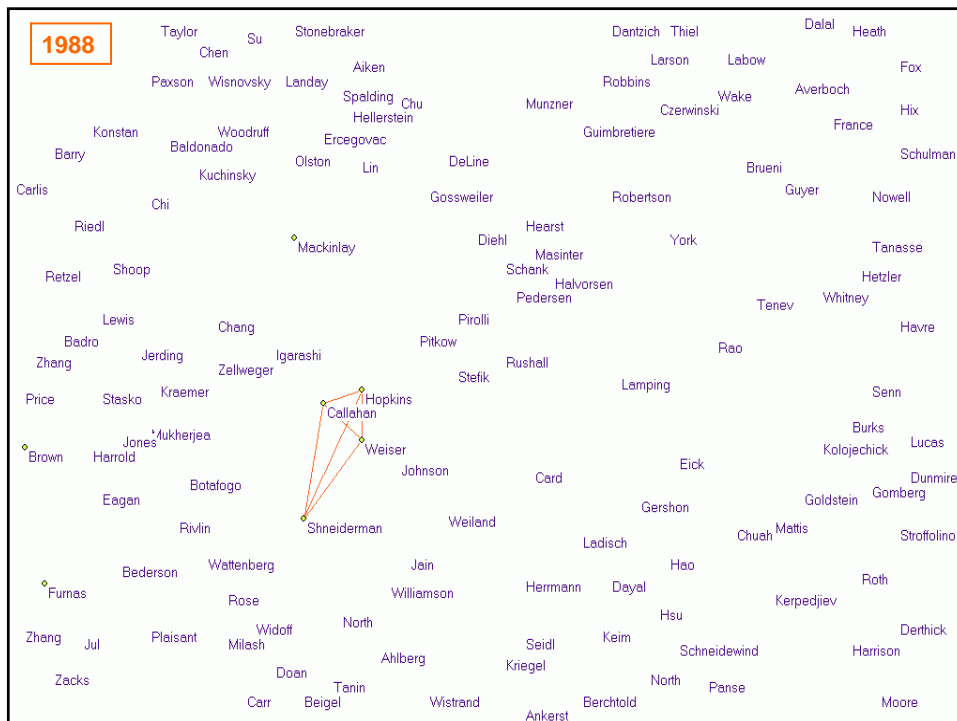
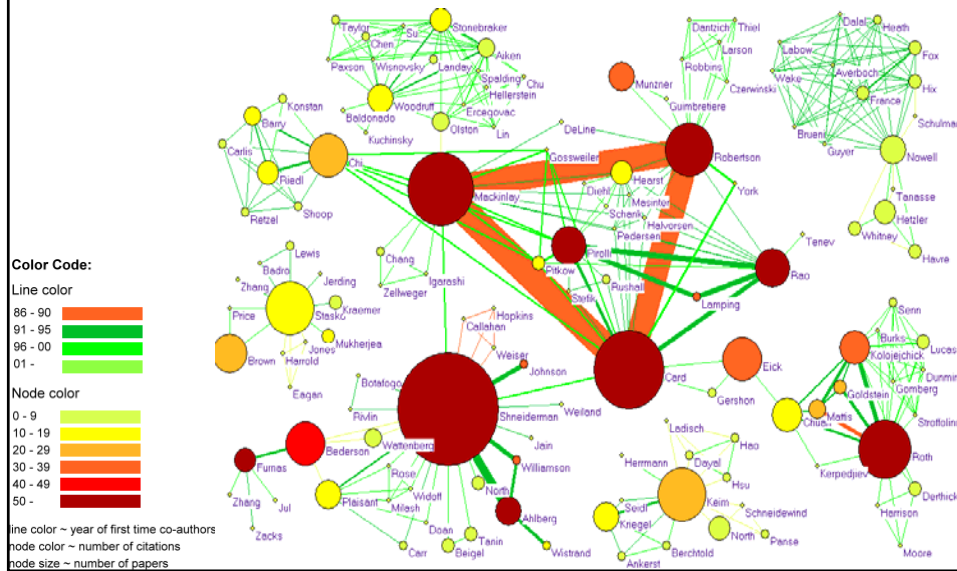


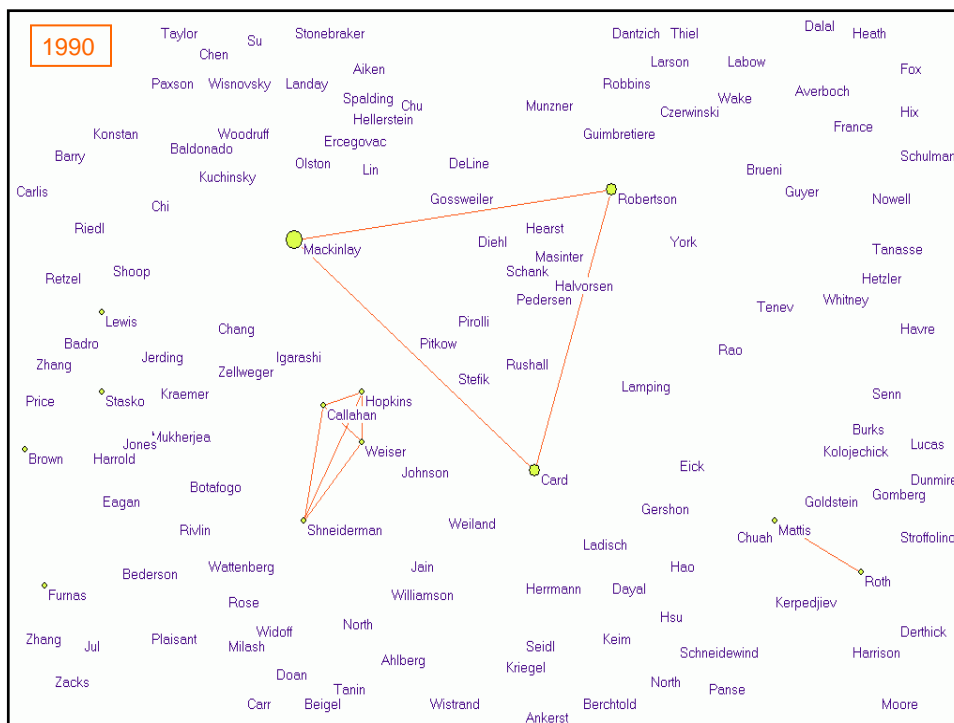
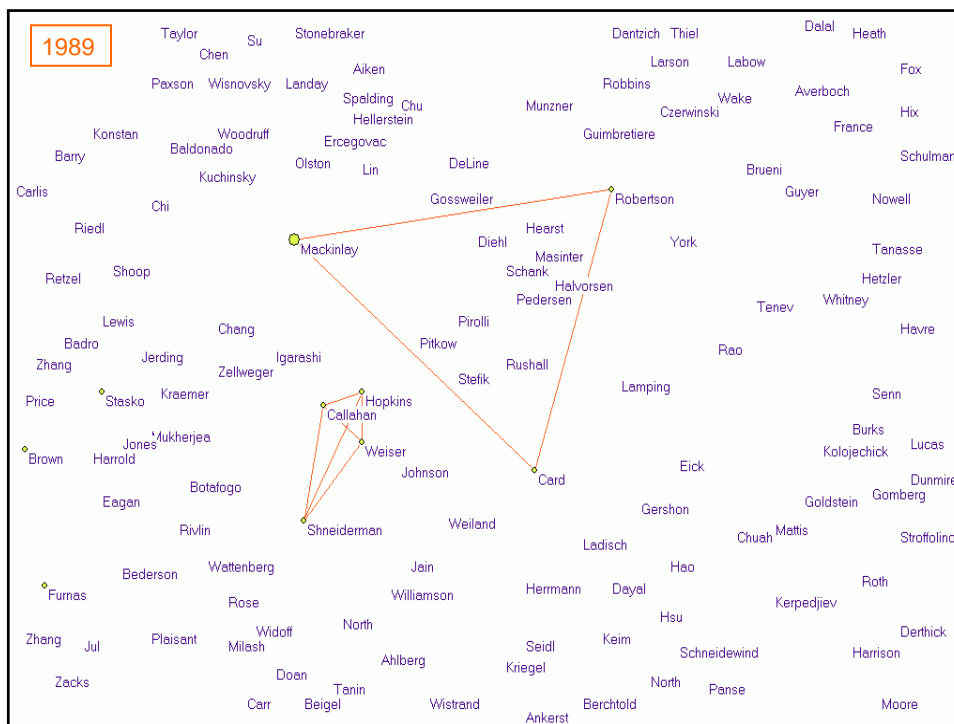
Overview

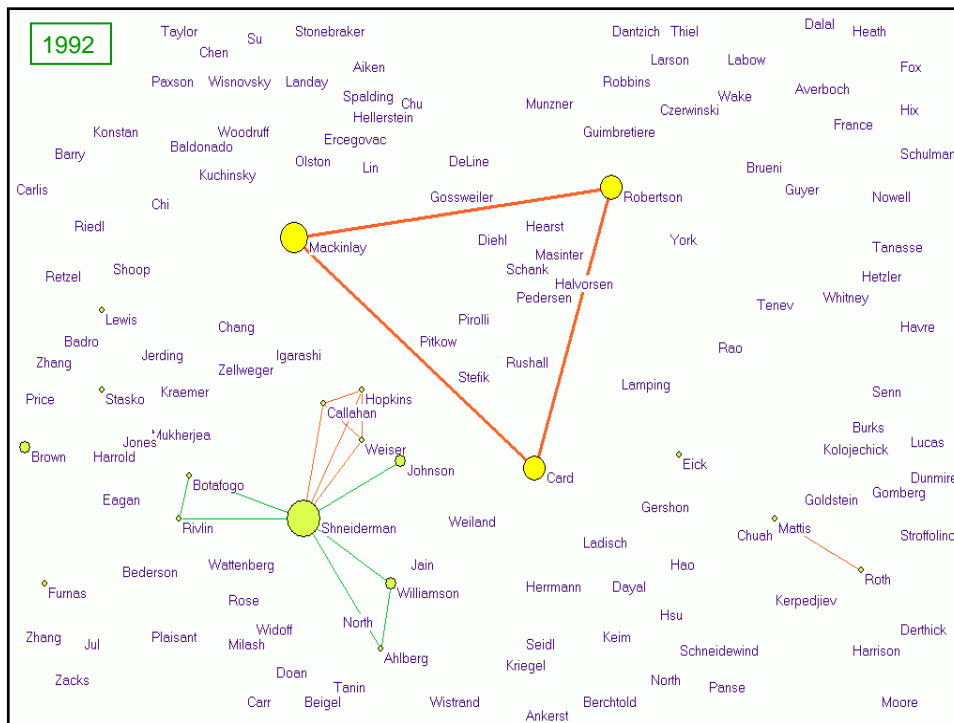
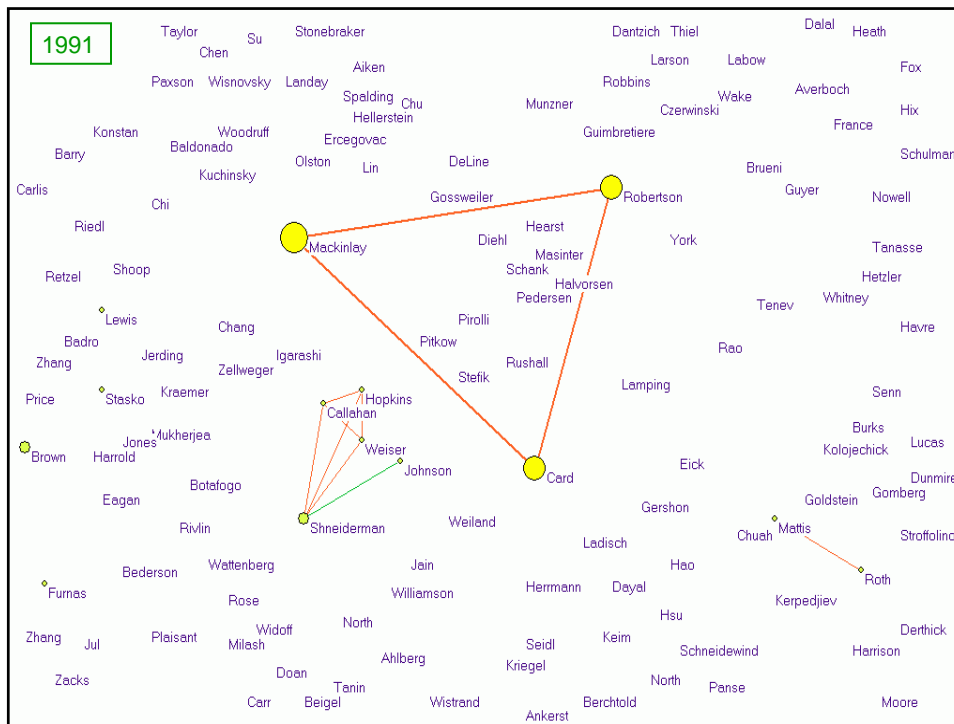
1. Motivation for Mapping Knowledge Domains
2. Mapping the Structure and Evolution of
Scientific Disciplines
All of Sciences
3. Challenges and Opportunities

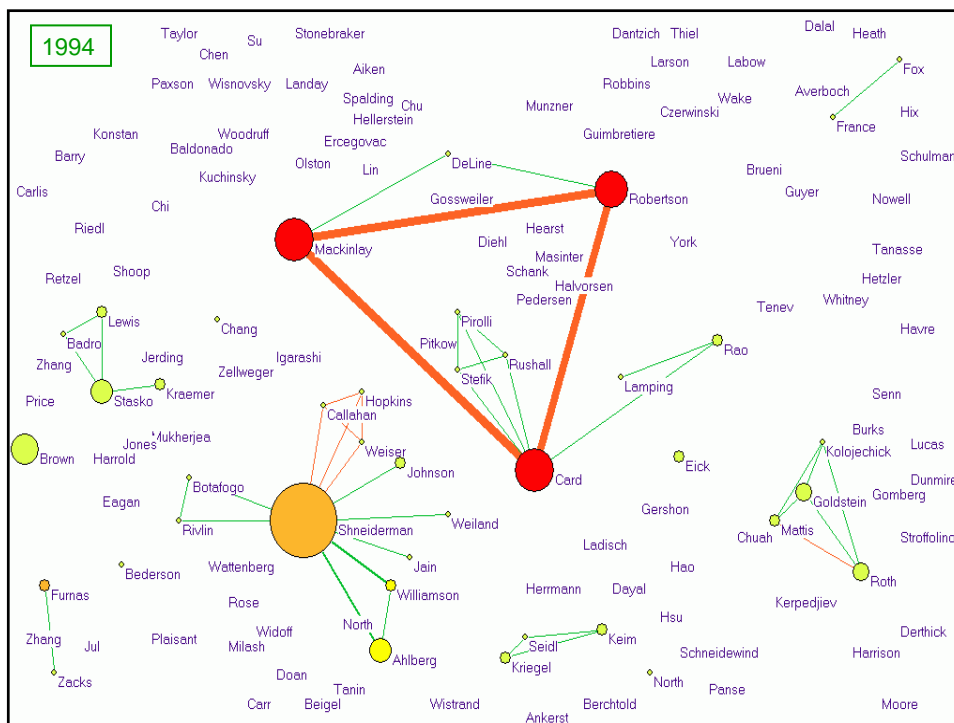
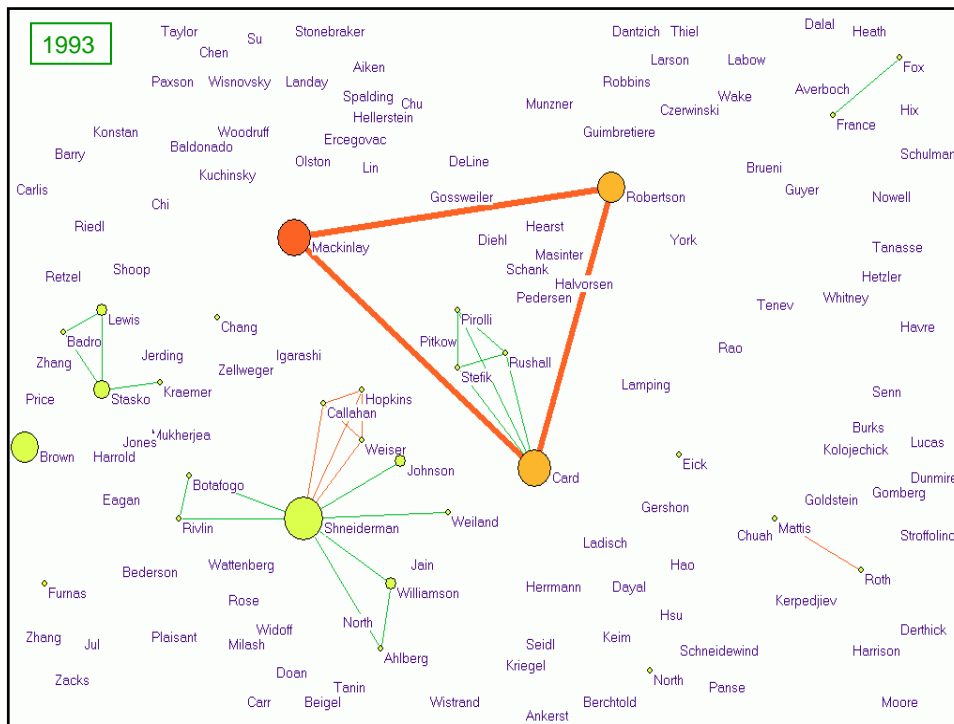
Mapping the Evolution of Co-Authorship Networks

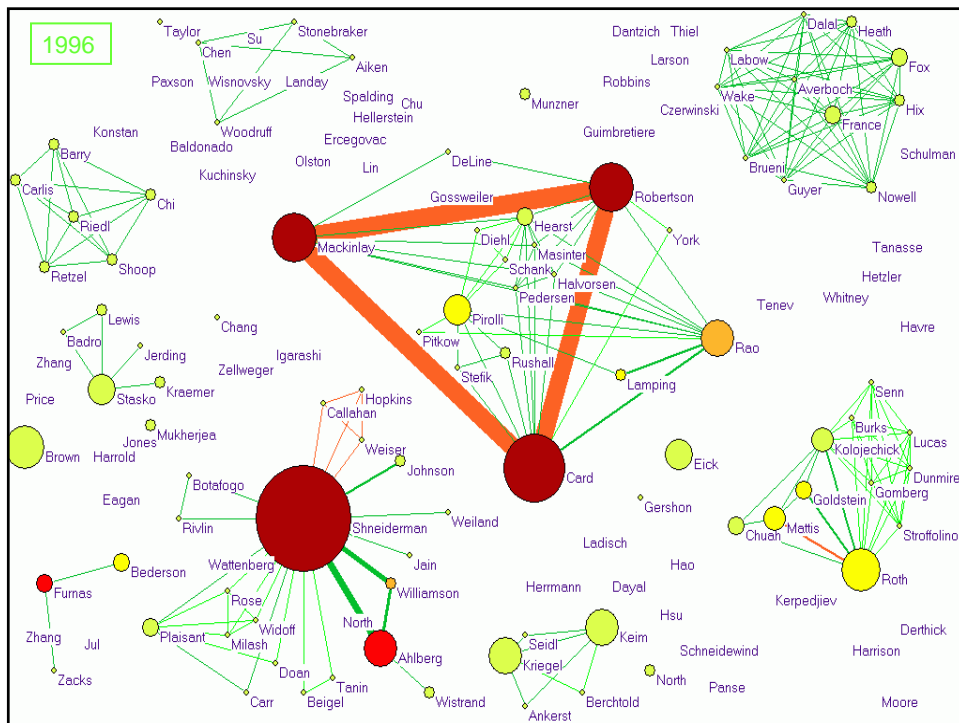
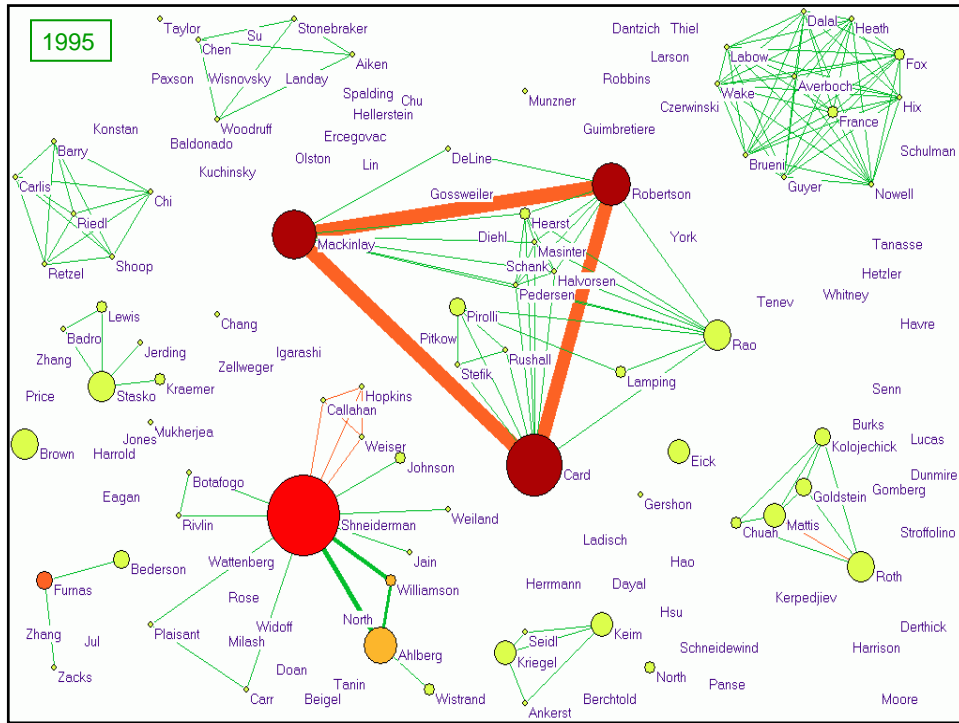
Ke, Viswanath & Börner, (2004) *Won 1st price at the IEEE InfoVis Contest.*

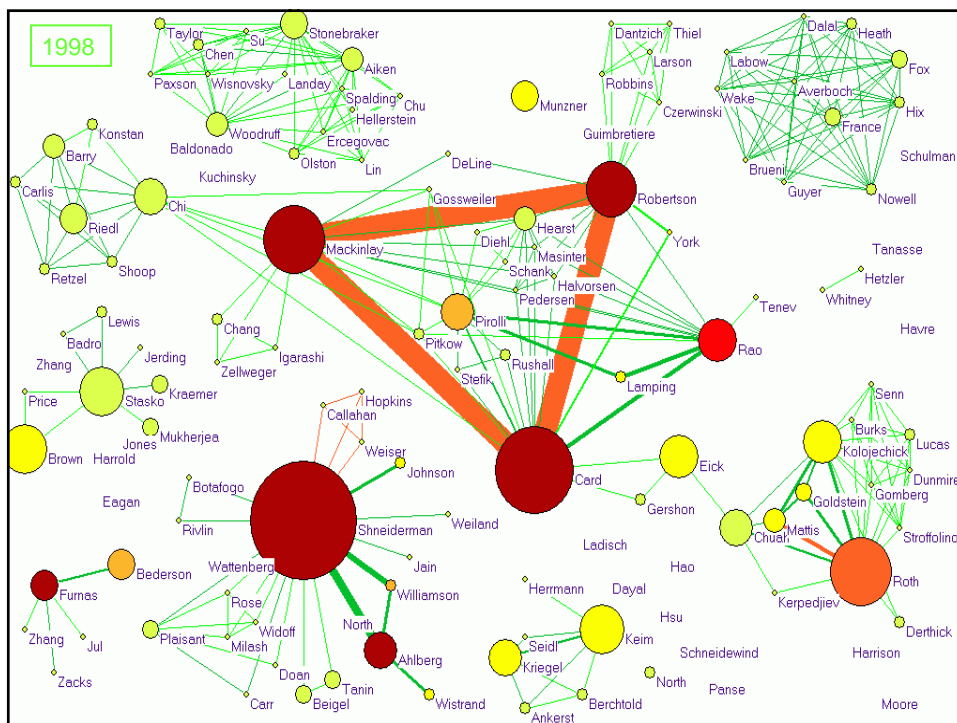
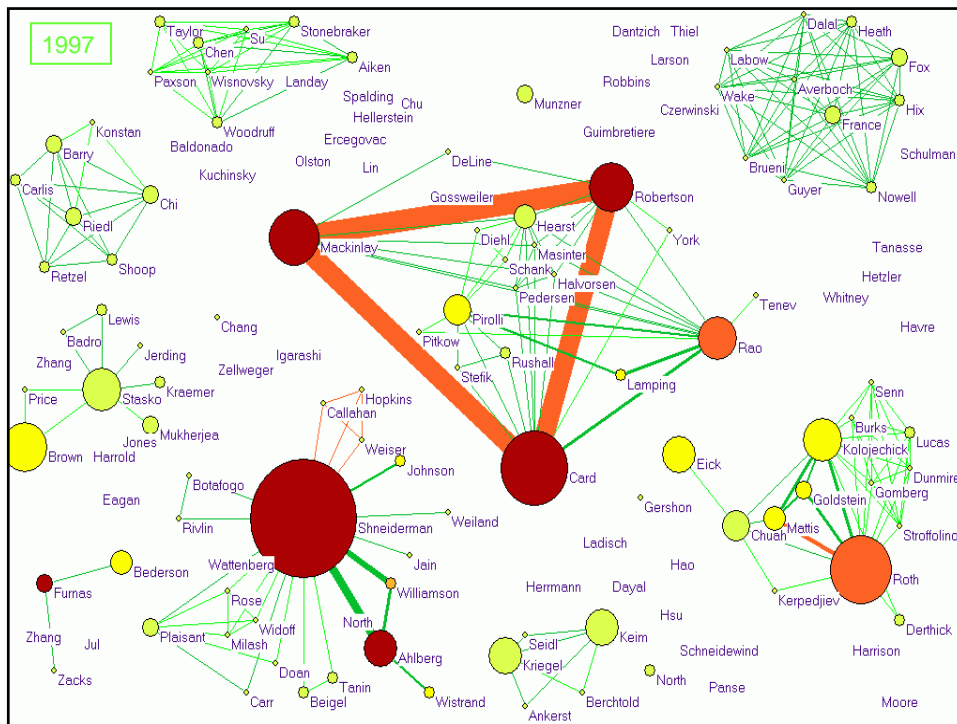


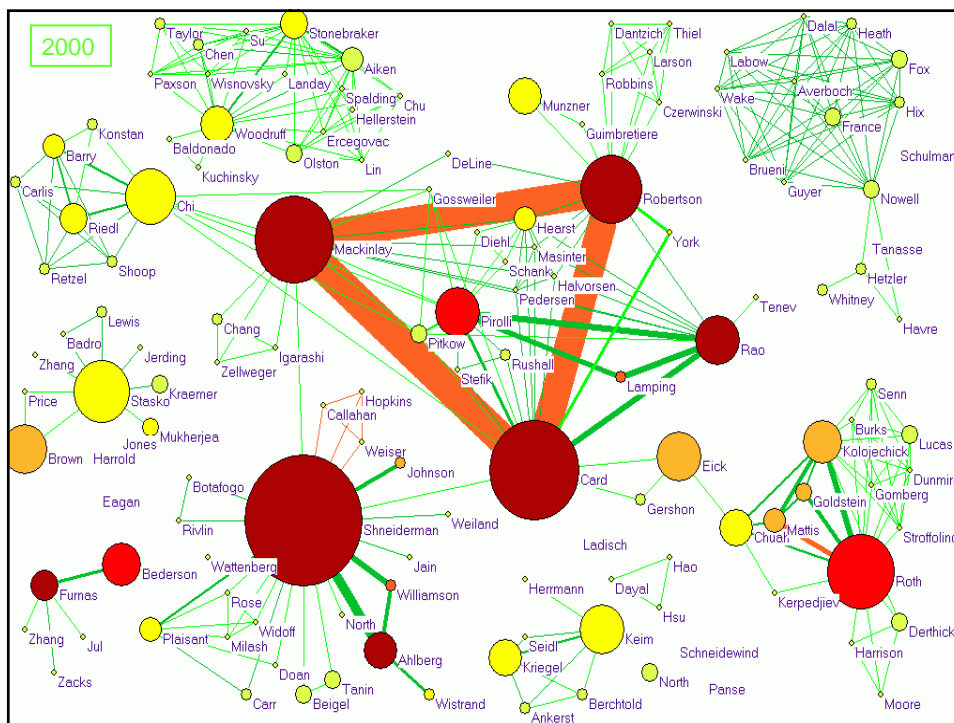
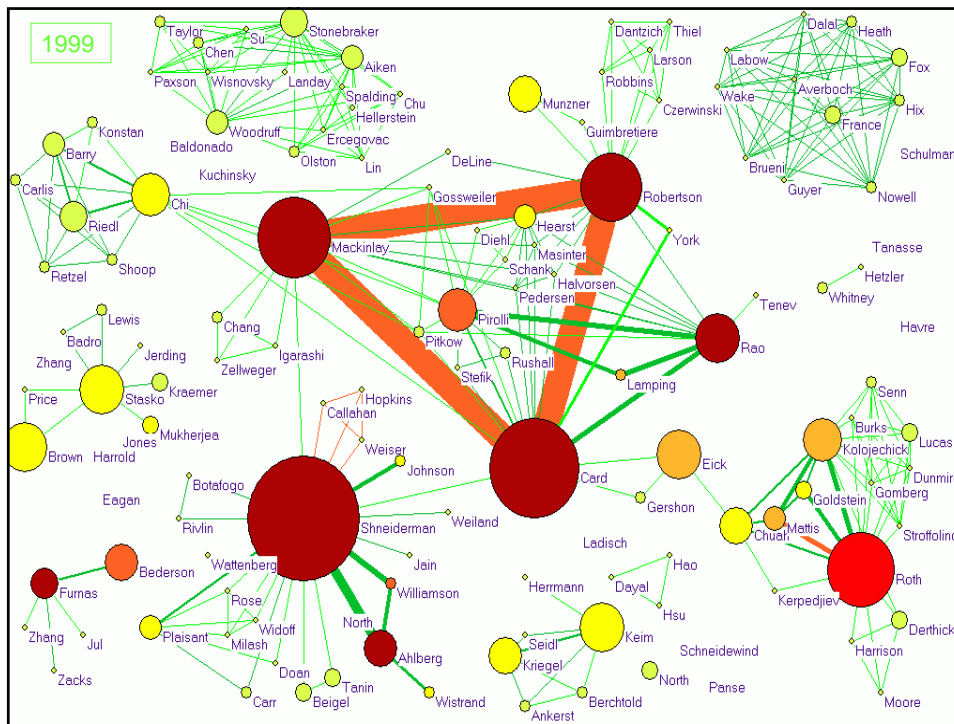


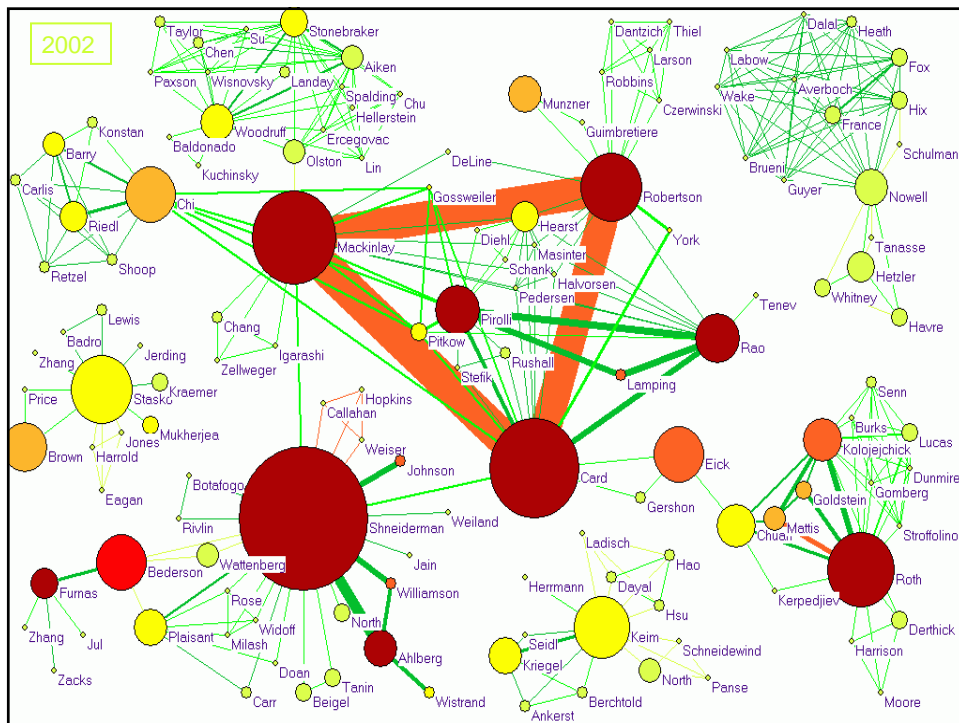
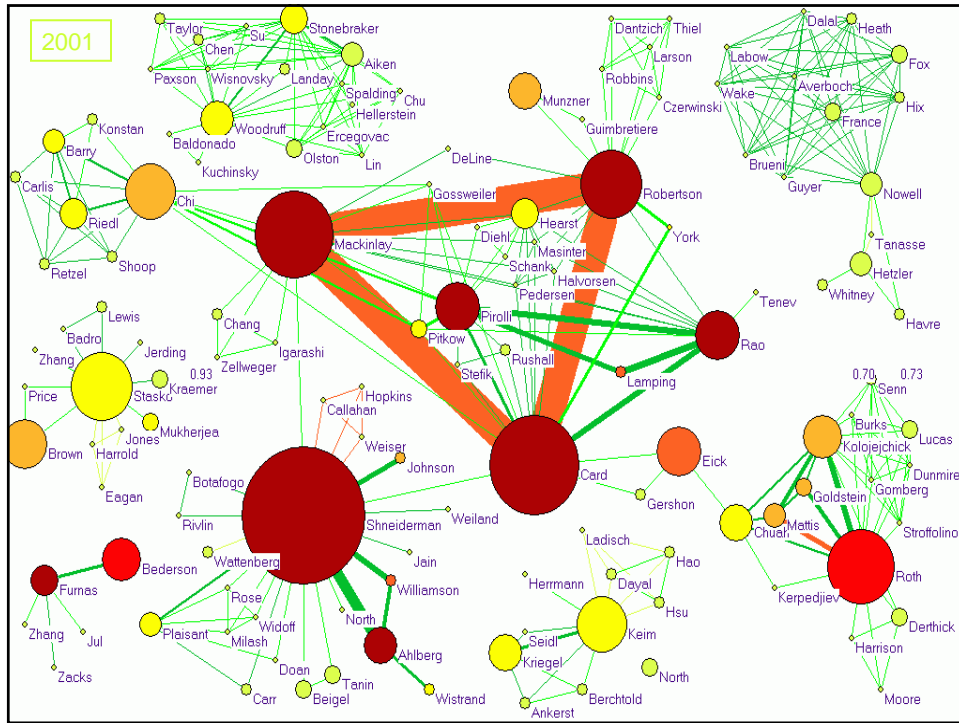


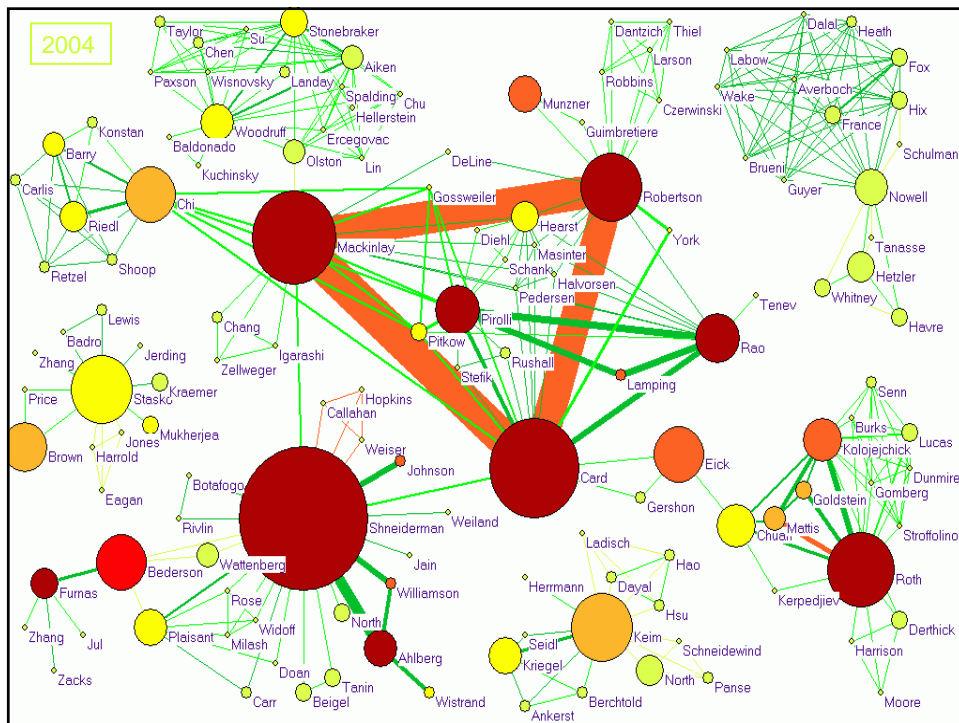
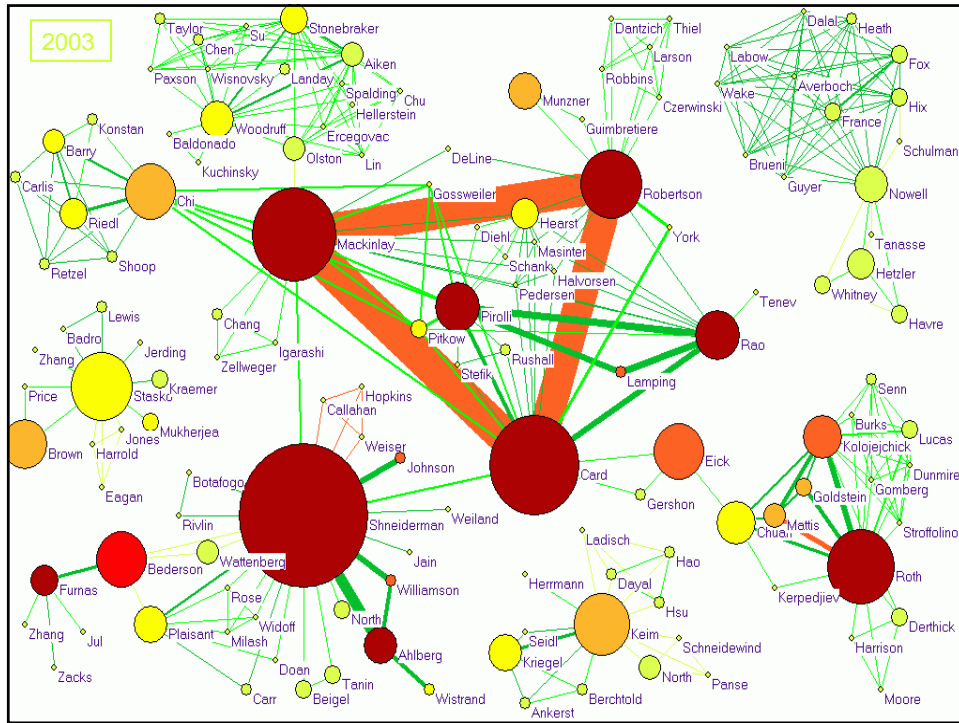


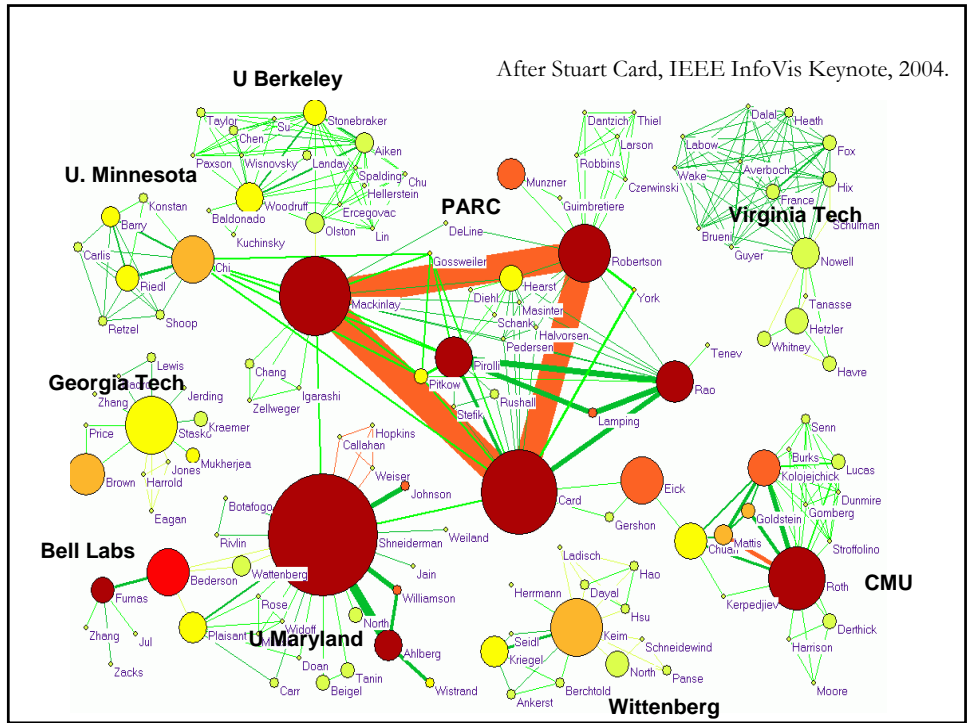


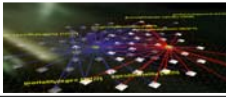













1. Motivation for Mapping Knowledge Domains / Computational Scientometrics

Knowledge domain visualizations help answer questions such as:


- What are the major research areas, experts, institutions, regions, nations, grants, publications, journals in xx research?
- Which areas are most insular?
- What are the main connections for each area?
- What is the relative speed of areas?
- Which areas are the most dynamic/static?
- What new research areas are evolving?
- Impact of xx research on other fields?
- How does funding influence the number and quality of publications?



Answers are needed by funding agencies, companies, and researchers.

Shiffrin & Börner (Eds). (2004) Mapping Knowledge Domains. PNAS, 101(Suppl_1):5266-5273.

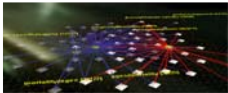
Mapping Knowledge Domains, Katy Börner, Indiana University



User Groups

- **Students** can gain an overview of a particular knowledge domain, identify major research areas, experts, institutions, grants, publications, patents, citations, and journals as well as their interconnections, or see the influence of certain theories.
- **Researchers** can monitor and access research results, relevant funding opportunities, potential collaborators inside and outside the fields of inquiry, the dynamics (speed of growth, diversification) of scientific fields, and complementary capabilities.
- **Grant agencies/R&D managers** could use the maps to select reviewers or expert panels, to augment peer-review, to monitor (long-term) money flow and research developments, evaluate funding strategies for different programs, decisions on project durations, and funding patterns, but also to identify the impact of strategic and applied research funding programs.
- **Industry** can use the maps to access scientific results and knowledge carriers, to detect research frontiers, etc. Information on needed technologies could be incorporated into the maps, facilitating industry pulls for specific directions of research.
- **Data providers** benefit as the maps provide unique visual interfaces to digital libraries.
- Last but not least, the availability of dynamically evolving maps of science (as ubiquitous as daily weather forecast maps) would dramatically improve the communication of scientific results to the **general public**.

Mapping Knowledge Domains, Katy Börner, Indiana University 23

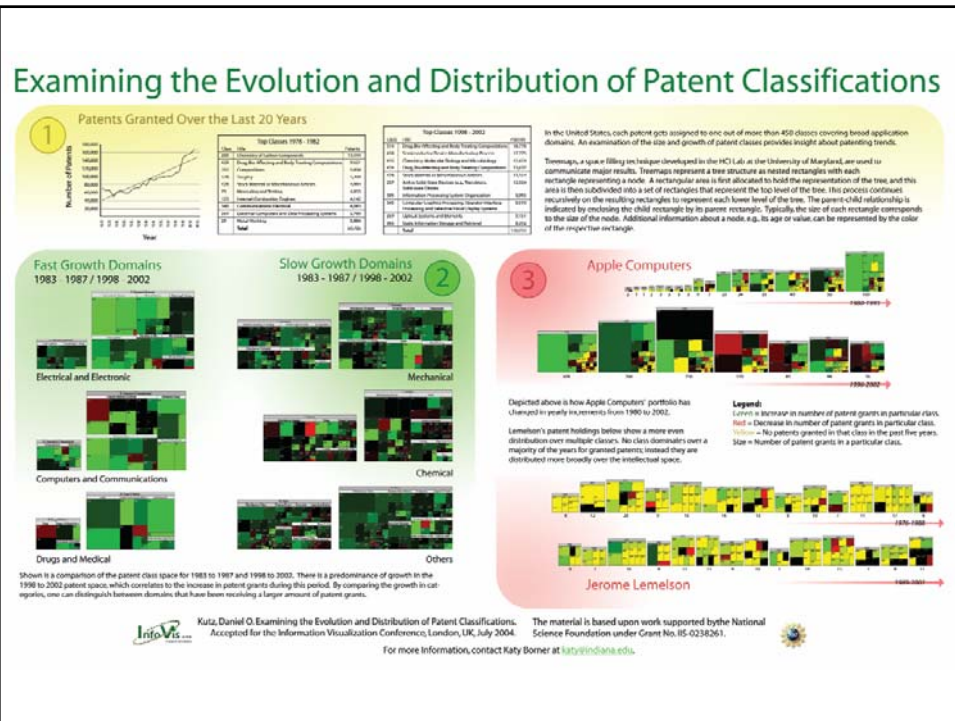
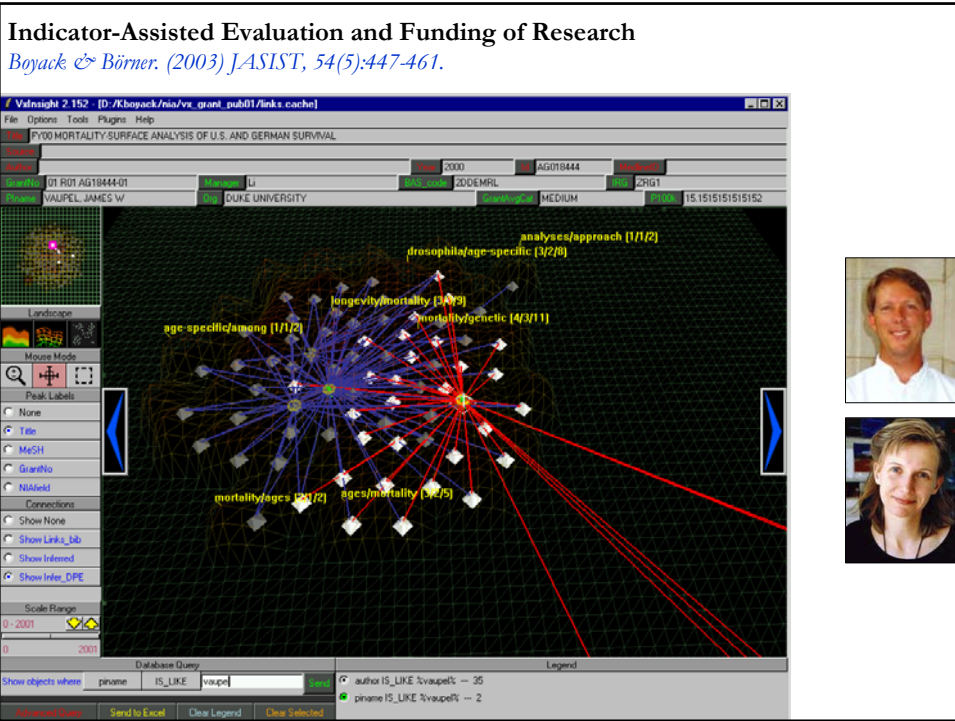


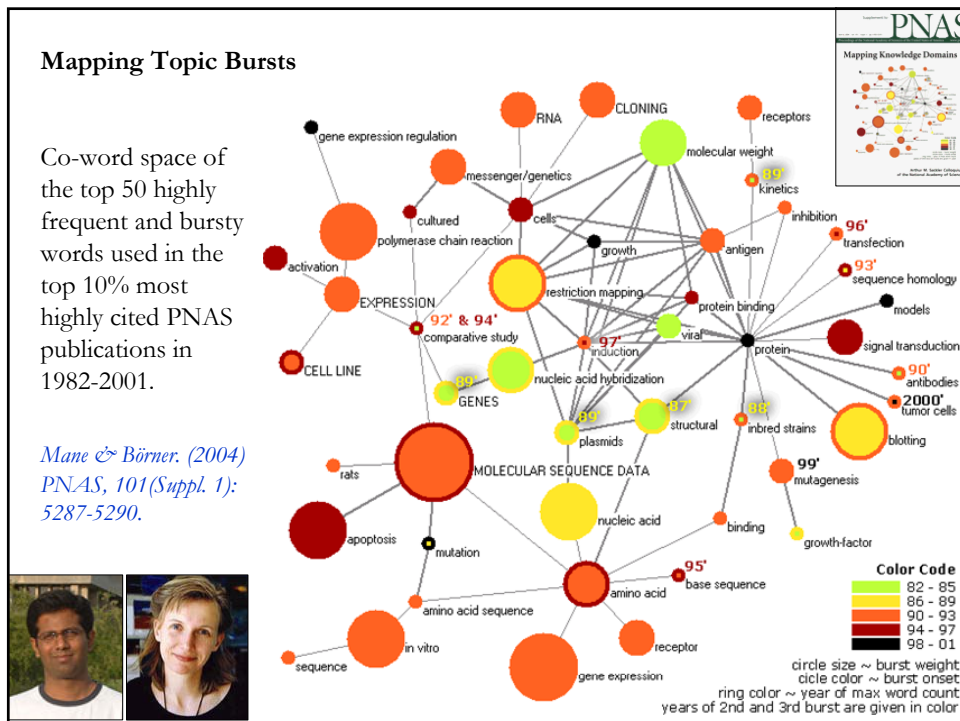
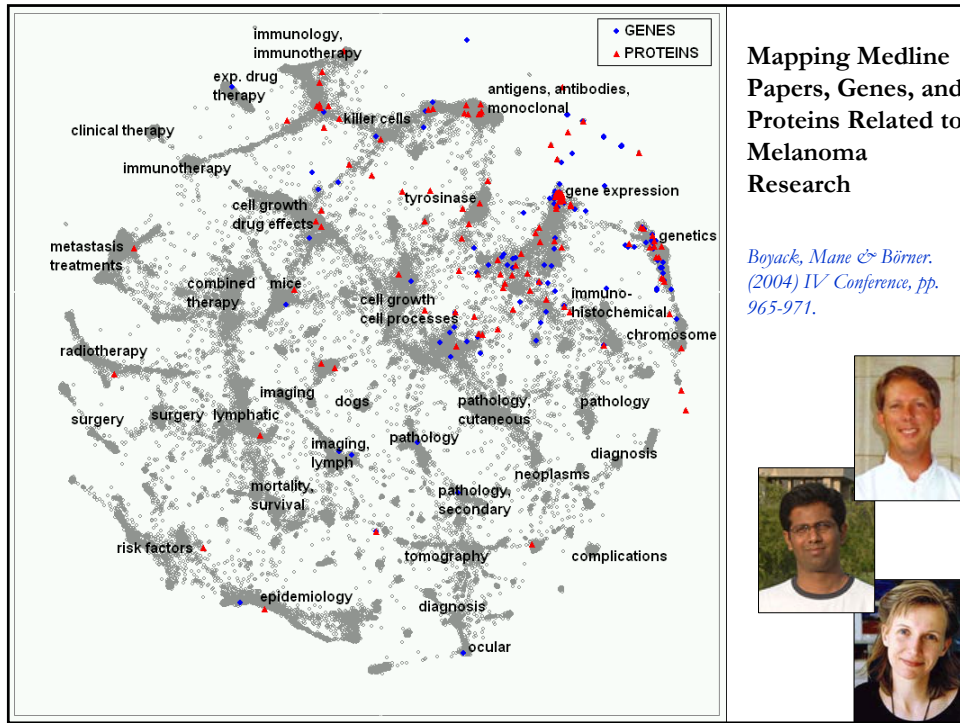
2. Mapping the Structure and Evolution of Knowledge Domains

| 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. | 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, Topics 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 |
| BROADENING By citation By terms | | | | | |

Börner, Chen & Boyack.. (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.

Mapping Knowledge Domains, Katy Börner, Indiana University 24





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

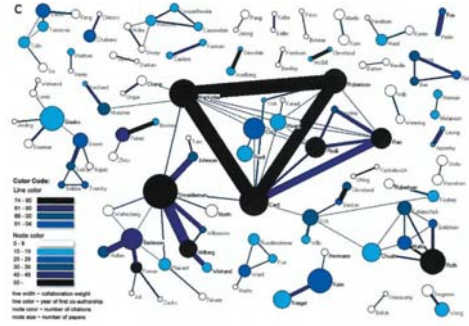
Börner, Dall'Asta, Ke & Vespignani (2005) *Complexity*, 10(4):58-67.

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.



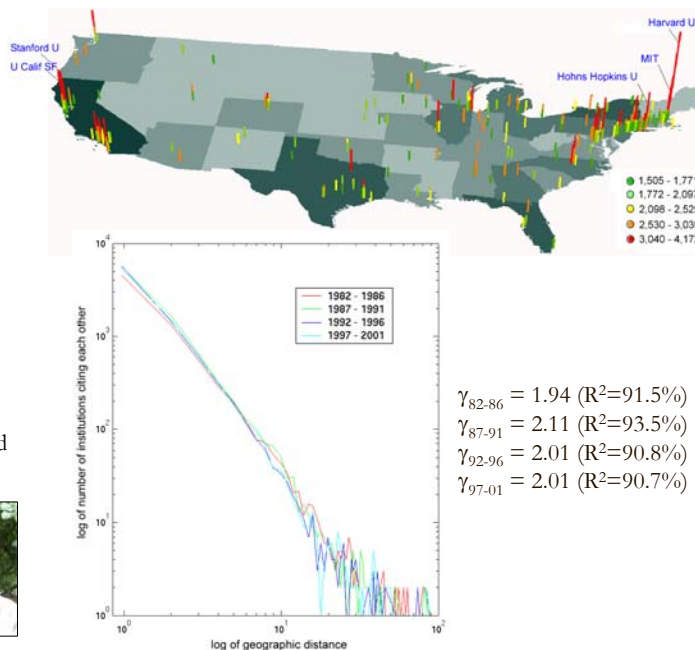
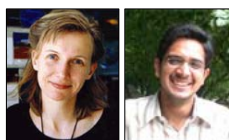
Spatio-Temporal Information Production and Consumption of Major U.S. Research Institutions

Börner & Penumathy. (2005) *Scientometrics Conference*.

Does Internet lead to more global citation patterns, i.e., more citation links between papers produced at geographically distant research institutions?

Analysis of top 500 most highly cited U.S. institutions.

Each institution is assumed to produce and consume information.





Mapping all of Sciences

(in English speaking domain, based on available data)

Subsequent slides are based on

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

Comparing different similarity measures

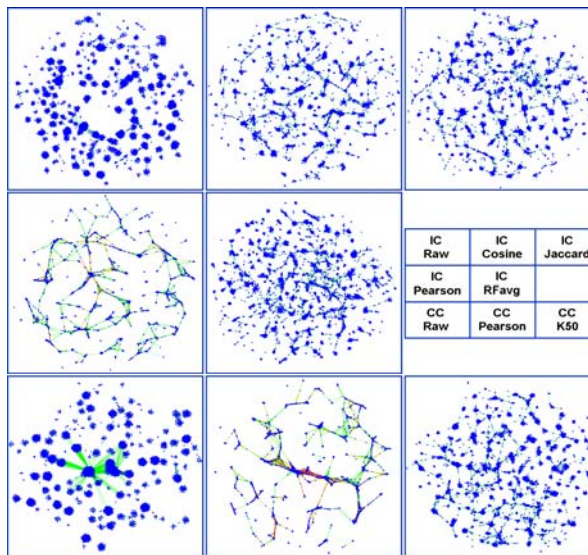
ISI file year 2000, SCI and SSCI:
7,121 journals.

Ten different similarity metrics

- 6 Inter-citation (raw counts, cosine, modified cosine, Jaccard, RF, Pearson)
- 4 Co-citation (raw counts, cosine, modified cosine, Pearson)

Maps were compared based on

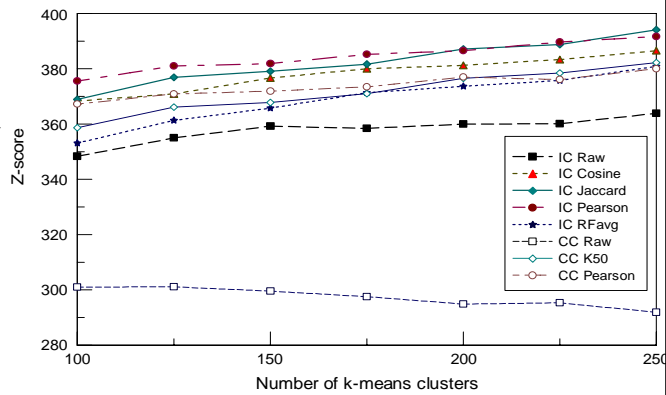
- regional accuracy,
- the scalability of the similarity algorithm, and
- the readability of the layouts.



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

Selecting the similarity measure with the best regional accuracy

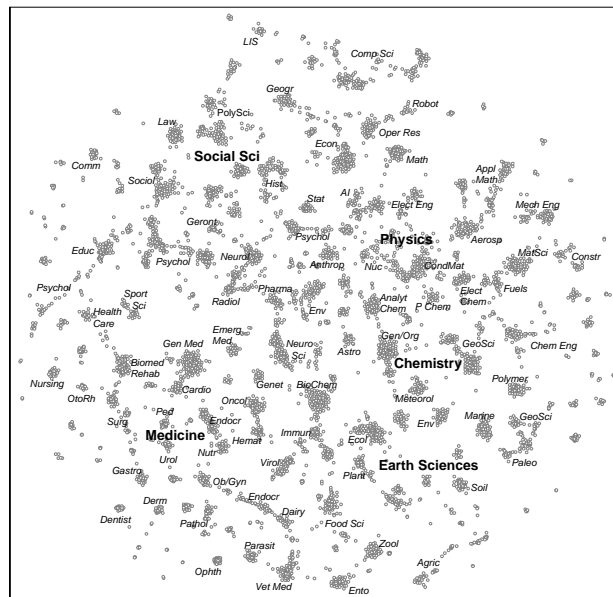
- For each similarity measure, the VxOrd layout was subjected to k-means clustering using different numbers of clusters.
- Resulting cluster/category memberships were compared to actual category memberships using entropy/mutual information method by Gibbons & Roth, 2002.
- Increasing Z-score indicates 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*.

A map of all science & social science

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

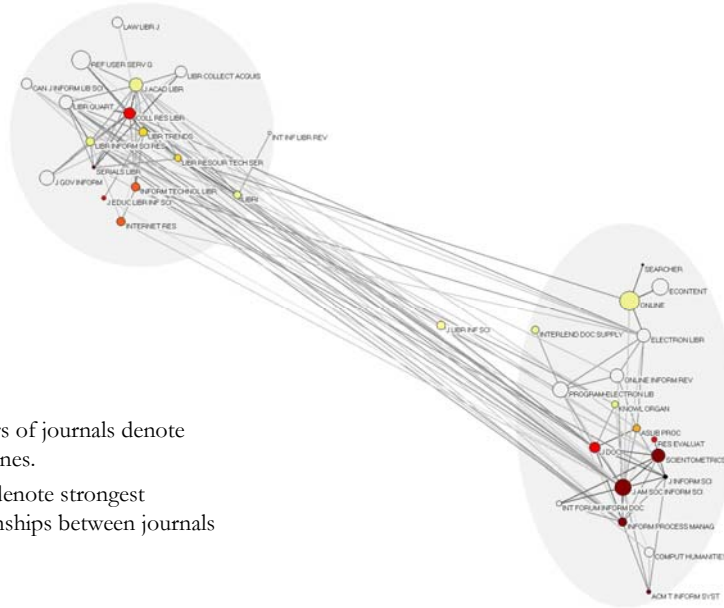


Structural map: Studying disciplinary diffusion

- The 212 nodes represent clusters of journals for different disciplines.
- Nodes are labeled with their dominant ISI category name.
- 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).

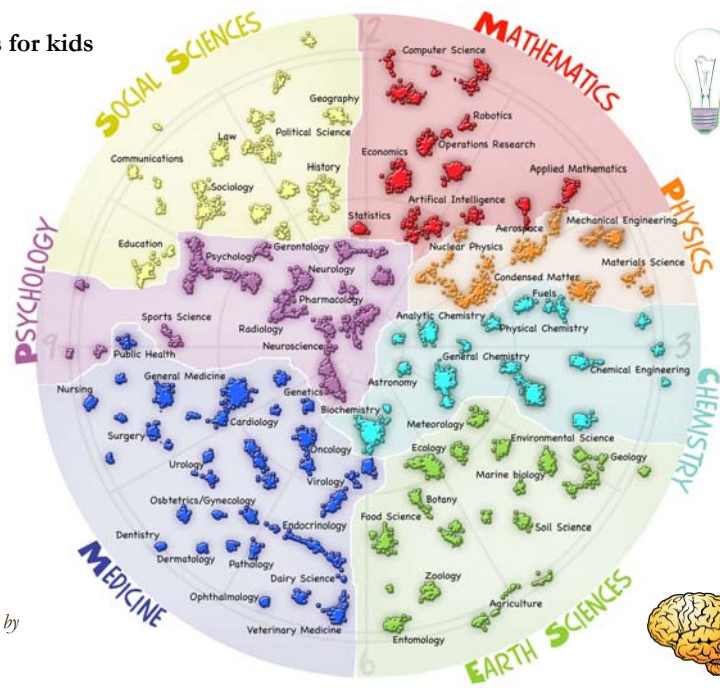


Zoom into structural map



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

Science maps for kids

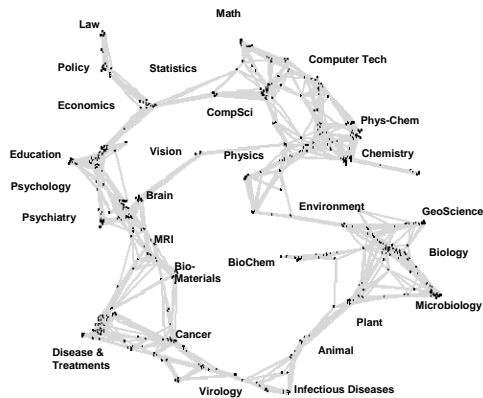


Base map modified by Ian Aliman, IU.

Latest 'Base Map' of sciences

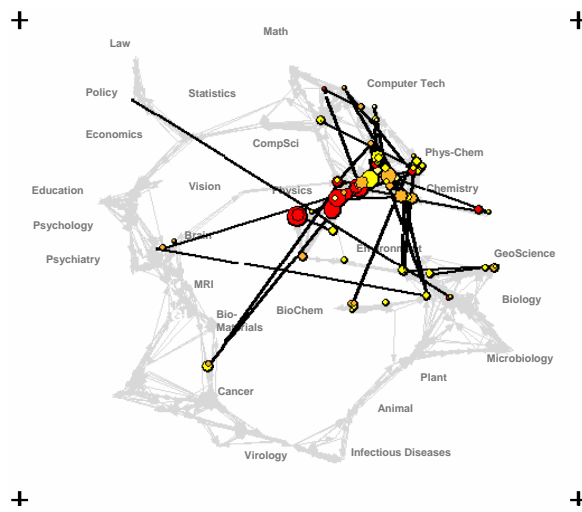
Presented by Kevin Boyack at AAG, 2005.

- Uses combined SCIE/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



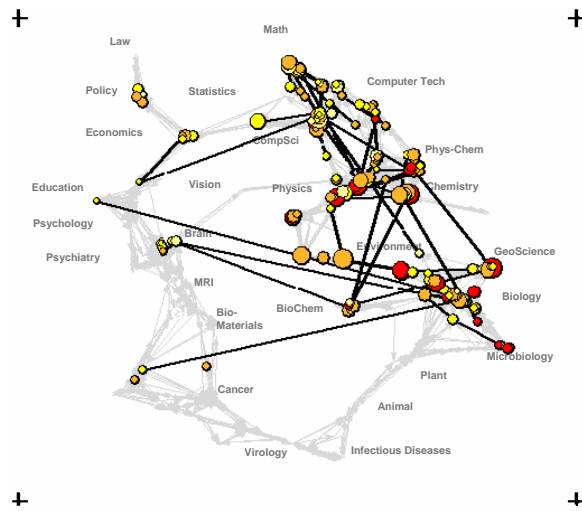
Science Map Applications: Identifying Core Competency

Funding patterns of the US Department of Energy (DOE)



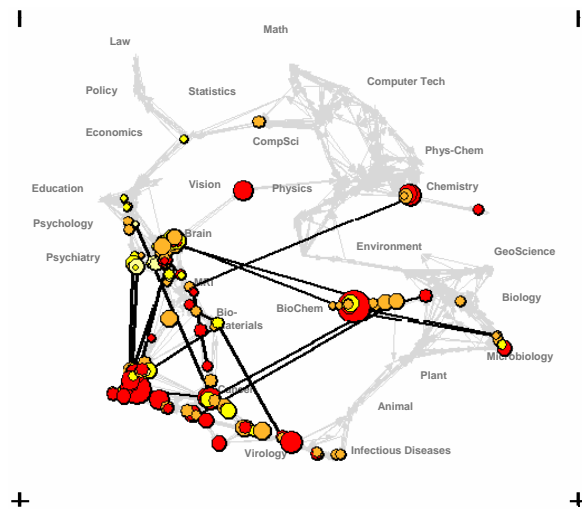
Science Map Applications: Identifying Core Competency


Funding patterns of the National Science Foundation (NSF)



Science Map Applications: Identifying Core Competency

Funding patterns of the National Institutes of Health (NIH)





3. Challenges and Opportunities

Map sciences on a small (regional) and a large scale:

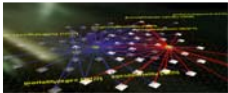
- Develop techniques, tools, and infrastructures that can continuously harvest, integrate, analyze, and visualize the growing stream of scholarly data.
- Educate scholars, practitioners, and the general public about alternative means to access humanity's collective knowledge.

Increase our understanding of the structure and evolution of sciences:

- Model the co-evolution of scholarly networks
Börner, Katy, Marin, Jeegar and Goldstone, Robert. (2004). The Simultaneous Evolution of Author and Paper Networks. Proceedings of the National Academy of Sciences of the United States of America, 101(Suppl_1):5266-5273. Also available as cond-mat/0311459.
- Model the diffusion of knowledge in evolving network ecologies.

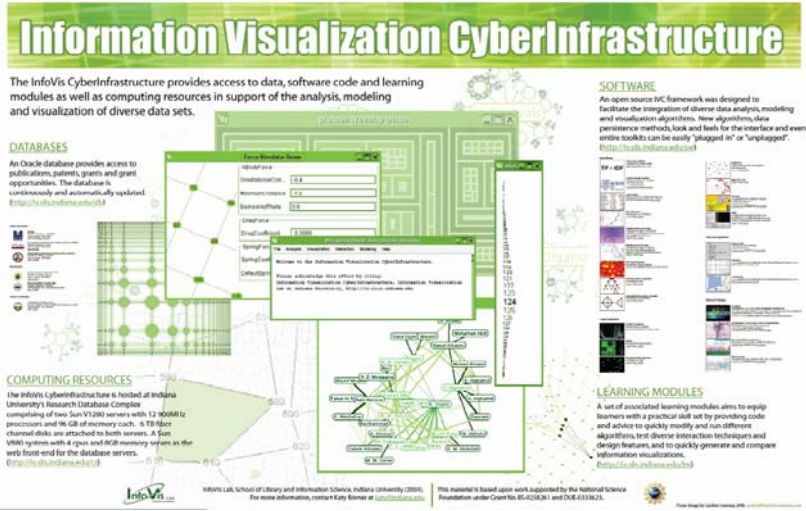
Mapping Knowledge Domains, Katy Börner, Indiana University

43



InfoVis Cyberinfrastructure at IUB

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



Information Visualization CyberInfrastructure

The InfoVis CyberInfrastructure provides access to data, software code and learning modules as well as computing resources in support of the analysis, modeling and visualization of diverse data sets.

DATABASES
 An Oracle database provides access to publications, patents, grants and grant opportunities. The database is continuously and automatically updated.
<https://slis.indiana.edu/ib/>

COMPUTING RESOURCES
 The InfoVis CyberInfrastructure is hosted at Indiana University's Research Database Complex comprising of two Sun V1200 servers with 12 900MB L2 processors and 64 GB of memory each. 6 TB fiber channel disks are attached to both servers. A Sun V880 system with 4 cpus and 8GB memory servers in the web front end for the database servers.
iv@slis.indiana.edu

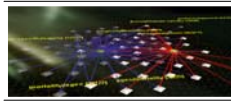
SOFTWARE
 An open source MVC framework was designed to facilitate the integration of diverse data analysis, modeling and visualization algorithms. New algorithms, data persistence methods, look and feels for the interface and even entire toolkits can be easily "plugged in" or "unplugged".
<https://tools.indiana.edu/>

LEARNING MODULES
 A set of associated learning modules aims to equip learners with a practical skill set by providing code and advice to quickly modify and run different algorithms, test diverse interactive techniques and design features, and to quickly generate and compare information visualizations.
<https://tools.indiana.edu/learn/>

Indiana University School of Library and Information Science, Indiana University (2005). For more information, contact Katy Börner at kborner@slis.indiana.edu. This material is based upon work supported by the National Science Foundation under Grants IRI-0208261 and IRI-0318624. Photo: Roger Lauder, Creative Commons.

Mapping Knowledge Domains, Katy Börner, Indiana University

44



3. Challenges and Opportunities

Map sciences on a small (regional) and a large scale:

- Develop techniques, tools, and infrastructures that can continuously harvest, integrate, analyze, and visualize the growing stream of scholarly data.
- Educate scholars, practitioners, and the general public about alternative means to access humanity's collective knowledge.

Increase our understanding of the structure and evolution of sciences:

- Model the co-evolution of scholarly networks
Börner, Katy, Mann, Jeeagar and Goldstone, Robert. (2004). The Simultaneous Evolution of Author and Paper Networks. Proceedings of the National Academy of Sciences of the United States of America, 101(Suppl_1):5266-5273. Also available as cond-mat/0311459.
- Model the diffusion of knowledge in evolving network ecologies.

Mapping Knowledge Domains, Katy Börner, Indiana University

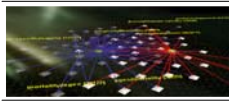
45

This physical & virtual science exhibit compares and contrasts first maps of our entire planet with the first maps of all of sciences.

<http://vw.indiana.edu/places&spaces/>







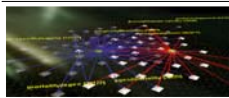
3. Challenges and Opportunities

Map sciences on a small (regional) and a large scale:

- Develop techniques, tools, and infrastructures that can continuously harvest, integrate, analyze, and visualize the growing stream of scholarly data.
- Educate scholars, practitioners, and the general public about alternative means to access humanity's collective knowledge.

Increase our understanding of the structure and evolution of sciences:

- Model the co-evolution of scholarly networks
Börner, Katy, Maru, Jeegar and Goldstone, Robert. (2004). The Simultaneous Evolution of Author and Paper Networks. Proceedings of the National Academy of Sciences of the United States of America, 101(Suppl_1):5266-5273. Also available as cond-mat/0311459.
- Model the diffusion of knowledge in evolving network ecologies.



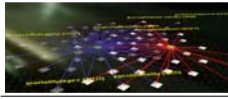
Acknowledgements

I would like to thank the students in the InfoVis Lab at IU and my collaborators for their contributions to this work.



Support comes from the School of Library and Information Science, Indiana University's High Performance Network Applications Program, a Pervasive Technology Lab Fellowship, an Academic Equipment Grant by SUN Microsystems, and an SBC (formerly Ameritech) Fellow Grant. This material is based upon work supported by the National Science Foundation under Grant No. DUE-0333623 and IIS-0238261.





References

- Boyack, Kevin W., Klavans, R. and Börner, Katy. (in press). Mapping the Backbone of Science. *Scientometrics*.
- 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.
- Katy Börner. (in press) Semantic Association Networks: Using Semantic Web Technology to Improve Scholarly Knowledge and Expertise Management. In Vladimir Geroimenko & Chaomei Chen (eds.) *Visualizing the Semantic Web*, Springer Verlag, 2nd Edition, chapter 11.
- 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*, 10(4): pp. 58 - 67. Also available as [cond-mat/0502147](#).
- Ord, Terry J., Martins, Emilia P., Thakur, Sidharth, Mane, Ketan K., and Börner, Katy. (2005) Trends in animal behaviour research (1968-2002): Ethoinformatics and mining library databases. *Animal Behaviour*, 69, 1399-1413. [Supplementary Material](#).
- Mane, Ketan K. and Börner, Katy. (2004). [Mapping Topics and Topic Bursts in PNAS](#). *Proceedings of the National Academy of Sciences of the United States of America*, 101(Suppl. 1):5287-5290. Also available as [cond-mat/0402380](#).
- Börner, Katy, Maru, Jeegar and Goldstone, Robert. (2004). [The Simultaneous Evolution of Author and Paper Networks](#). *Proceedings of the National Academy of Sciences of the United States of America*, 101(Suppl_1):5266-5273. Also available as [cond-mat/0311459](#).