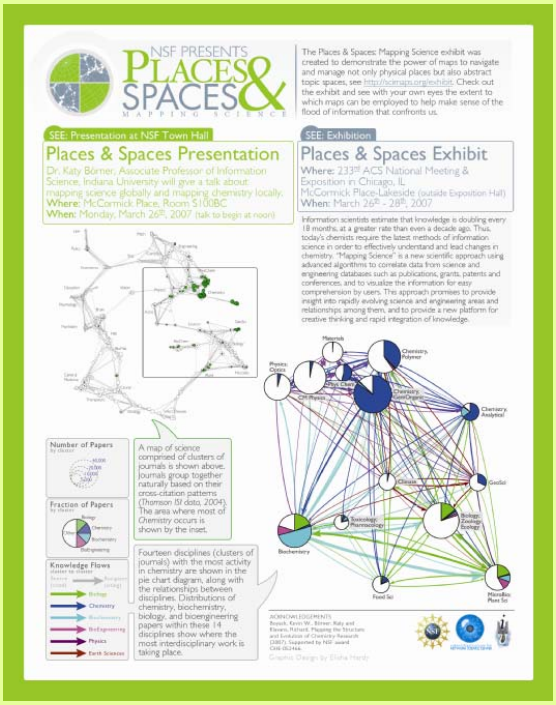
Mapping the Structure and Evolution of Chemistry Research

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

Collaborative work with Kevin W. Boyack & Richard Klavans

CHE-0524661

June 26, 2007 @ ISSI, Madrid, Spain



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

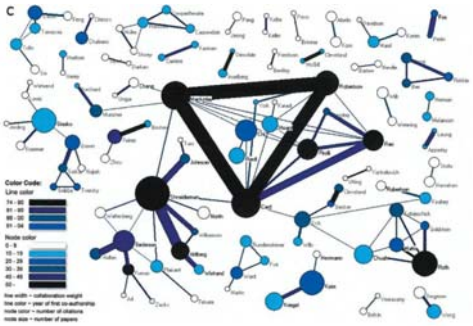
Börner, Dall'Asta, Ke & Vespignani (2005) Complexity, 10(4):58-67.

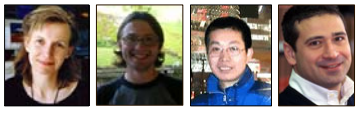
Research question:

- 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.



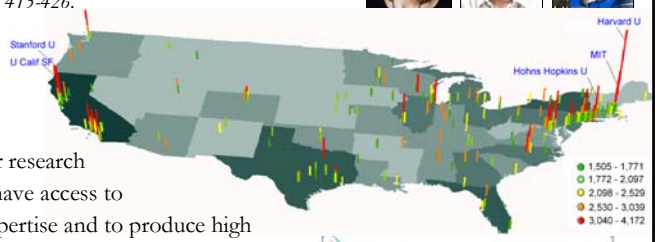


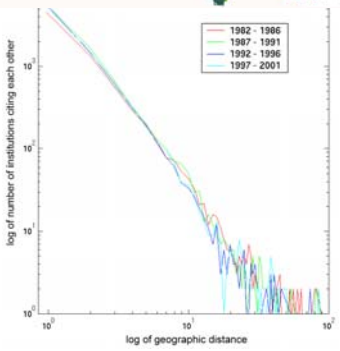
Spatio-Temporal Information Production and Consumption of Major U.S. Research Institutions

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

Research questions:

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



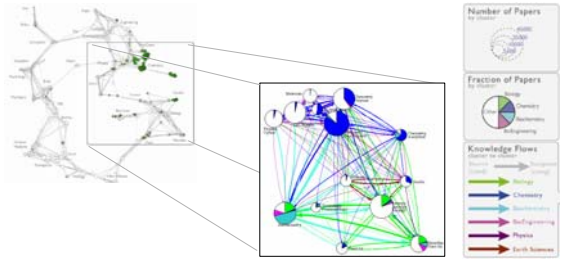


Contributions:

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

Overview

1. Introduction
2. Base map of science
3. Mapping chemistry



References:

Boyack, Kevin W., Börner, Katy and Klavans, Richard. Mapping the Structure and Evolution of Chemistry Research, Accepted for *11th International Conference on Scientometrics and Informetrics*, Madrid, Spain, June 25-27, 2007.

Boyack, Kevin W., Klavans, Richard and Börner, Katy. (2005). Mapping the Backbone of Science. *Scientometrics.* 64(3), 351-374.

Shiffrin, Richard M. and Börner, Katy. (April 2004). Mapping Knowledge Domains. *Proceedings of the National Academy of Sciences of the United States of America*, 101(Suppl_1).

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

Börner, Katy, Sanyal, Soma and Vespignani, Alessandro (2007). **Network Science**. In Blaise Cronin (Ed.), Annual Review of Information Science & Technology, Information Today, Inc./American Society for Information Science and Technology, Medford, NJ, volume 41, chapter 12, pp. 537-607.

Why map science and/or chemistry?

- How does our collective scholarly knowledge grow over time?
- What major areas of science exist and how are they interlinked?
- Which areas are major knowledge producers; which ones are consumers?

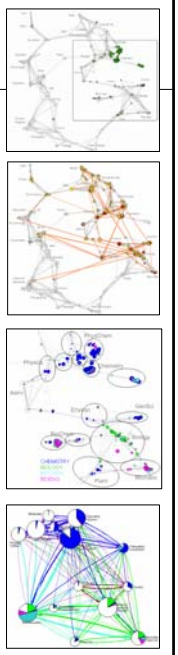
Computational scientometrics – the application of bibliometric/scientometric methods to large-scale scholarly datasets – and the communication of results via maps of science might help us answer these questions.

- Chemistry is a field that is undergoing significant change. Interdisciplinary research has increased over time and the lines between chemistry and the life sciences have seemingly blurred.
- This study maps the structures of *Chemistry*, *Biochemistry*, *Biology*, and *Bioengineering*, and their interactions over 30 years using journal citation patterns.

Börner: "Mapping the Evolving Interface of Mainstream Chemistry and the Fields of Biochemistry, Biology, and Bioengineering.

7

How to map science and/or chemistry?

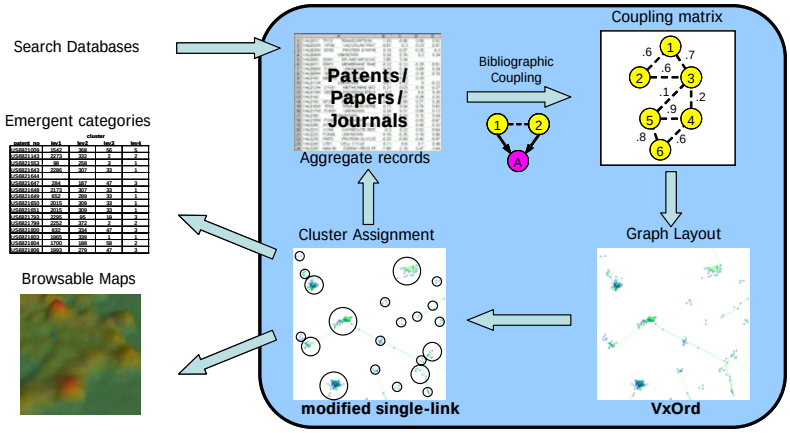


1. **Base Map.** Information from the combined Science (SCIE) and Social Science (SSCI) Citations Indexes from 2002 was used to generate a disciplinary map of 7,227 journals and 671 journal clusters.
2. **Overlays** of DOE, NIH, NSF funding.
3. **Chemistry Disciplines.** Clusters relevant to study the structure and evolution of chemistry were identified using JCR categories and were further clustered into 14 disciplines.
4. **Knowledge Diffusion among Disciplines.** The changing scientific composition of these 14 disciplines and their knowledge exchange via citation linkages was computed. The result is visually communicated using the 2002 map of science introduced in 1.)
5. **Discussion.** Major changes on the dominance, influence, and role of *Chemistry*, *Biology*, *Biochemistry*, and *Bioengineering* over these 30 years are discussed. We conclude with a discussion and suggestions for future work.

Börner: "Mapping the Evolving Interface of Mainstream Chemistry and the Fields of Biochemistry, Biology, and Bioengineering.

8

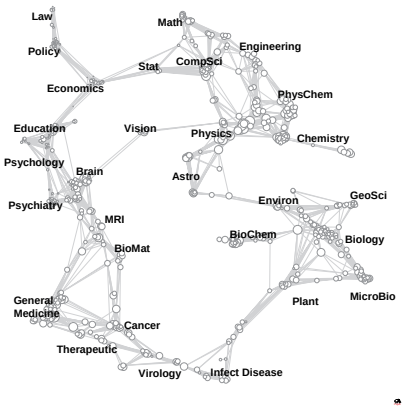
General mapping process (multi-step)



Börner: "Mapping the Evolving Interface of Mainstream Chemistry and the Fields of Biochemistry, Biology, and Bioengineering.

Generation of 2002 paper-level base map

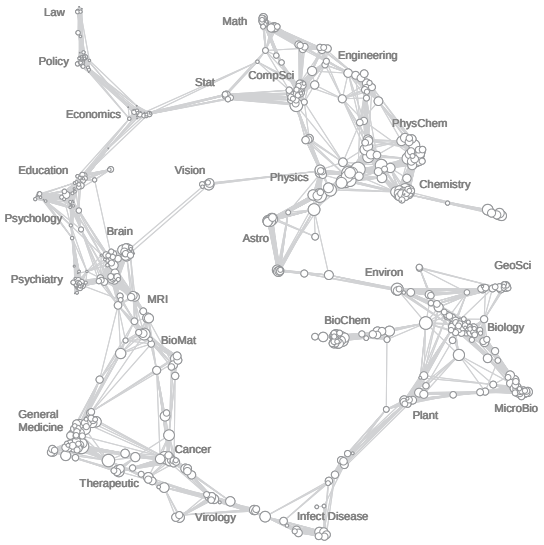
- **Combined SCIE/SSCI from 2002**
  - 1.07M papers, 24.5M references, 7,300 journals
  - Bibliographic coupling of papers, aggregated to journals, counts are normalized using cosine index.
- First step
  - remove 25 multidisciplinary journals (MD)
  - run graph layout and clustering on remaining journals
  - resulting in 646 clusters
- Second step
  - add MD journals back as single-journal clusters, total is now 671 clusters
  - re-aggregate coupling counts at the cluster level
  - layout positions of 671 clusters using VxOrd
  - by association, this gives x,y positions for each journal



Börner: "Mapping the Evolving Interface of Mainstream Chemistry and the Fields of Biochemistry, Biology, and Bioengineering.

2002 disciplinary map

Node – cluster of journals
Edge – similarity of journal clusters based on bibliographic coupling.

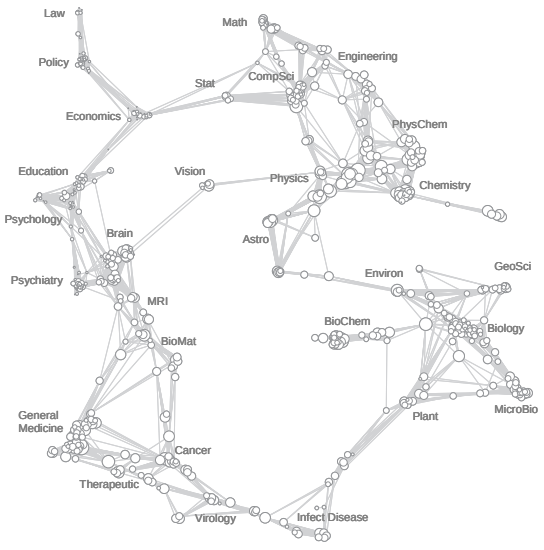


Börner: "Mapping the Evolving Interface of Mainstream Chemistry and the Fields of Biochemistry, Biology, and Bioengineering.

11

2002 disciplinary map

Node – cluster of journals
Edge – similarity of journal clusters based on bibliographic coupling.



Börner: "Mapping the Evolving Interface of Mainstream Chemistry and the Fields of Biochemistry, Biology, and Bioengineering.

14

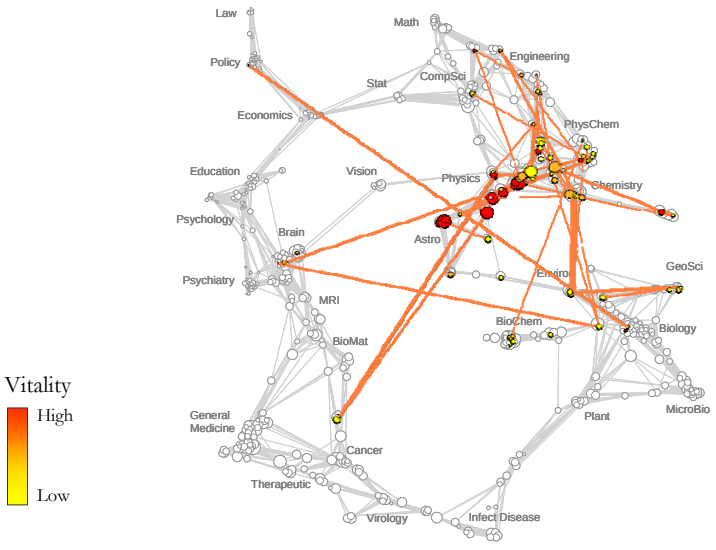
Funding overlays

- DOE – simple lookup of author institutions (LANL, LLNL, SNL, ANL, BNL, …)
- NIH and NSF: Used RaDiUS data from 1999
  - Matched funded PI and institution from 1999 funding data to first author and institution from the 2002 ISI publication data.
  - Not comprehensive, but representative.
  - Undercounts actual funding profiles (e.g., if PI wasn't first author, if publications appeared earlier or later than 2002, …).
  - Some false hits – could use text analysis to narrow further.

Börner: "Mapping the Evolving Interface of Mainstream Chemistry and the Fields of Biochemistry, Biology, and Bioengineering.

15

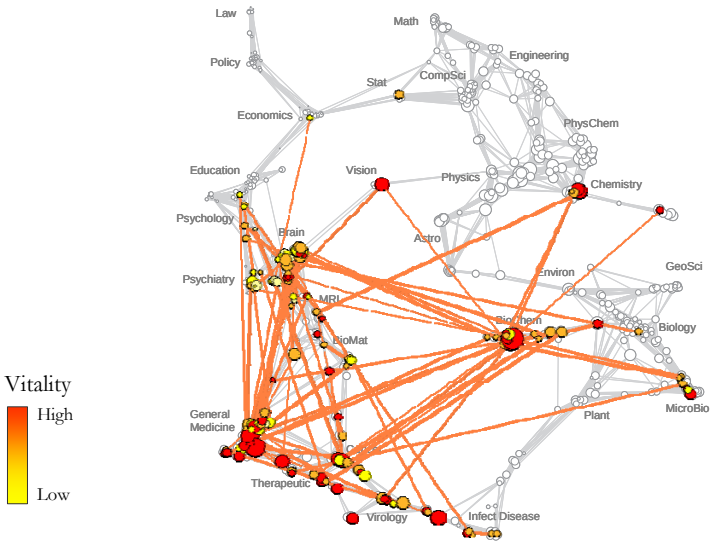
Funding overlay – DOE laboratories



Börner: "Mapping the Evolving Interface of Mainstream Chemistry and the Fields of Biochemistry, Biology, and Bioengineering.

16

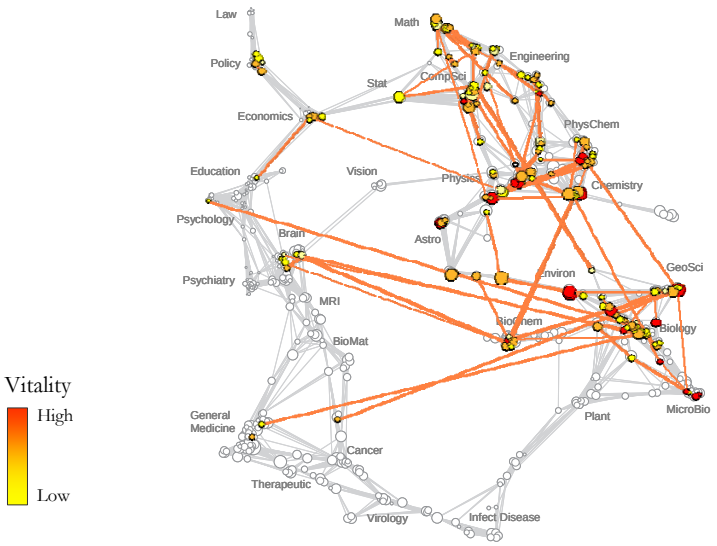
Funding overlay – NIH funded research (extramural)



Börner: "Mapping the Evolving Interface of Mainstream Chemistry and the Fields of Biochemistry, Biology, and Bioengineering.

17

Funding overlay – NSF funded research



Börner: "Mapping the Evolving Interface of Mainstream Chemistry and the Fields of Biochemistry, Biology, and Bioengineering.

18

How to map chemistry

- Data
  - 1993-2004 JCR (citing:cited journal counts).
  - Need comparable data for back years.
  - Obtained journal citing:cited pairs for 1974, 79, 84, 89.
  - Summed pairs to citing:cited journal counts.
  - Thus, full set of citing:cited journal counts at 5 year intervals 1974-2004.

- Used 2002 base map as starting point.
- Added unique journals for other years to existing (2002) structure.
  - Order of adding years: 2004, 1999, 1994, 1989, 1984, 1979, 1974
- Identified four fields and 14 disciplines within map.
- Fractional counting of publications by field and discipline.
- Graphical display of results.

Adding unique journals

- Example, for 2004, if [newjnl] was not in the 2002 list.
  - Intercitation counts were read from table for all newjnl::existingjnl pairs.
  - Counts were aggregated to a newjnl::existingcluster matrix.
  - Only counts to citedyr >= pubyear-9 were included (JCR standard).
  - Cosine index values were generated for the newjnl::existingcluster matrix.
  - Newjnl was assigned to the existingcluster with the highest cosine value.
- For the next year, 1999, if [newjnl] was not in the combined 2002,2004 list
  - same process.
- All journals for all years were assigned to one of the 671 clusters.

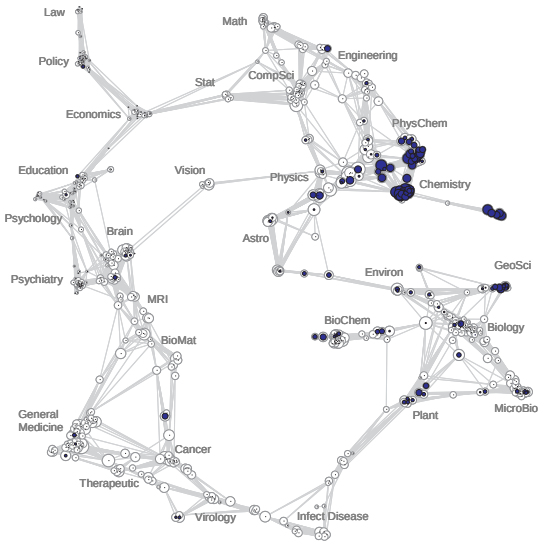
Four fields of interest / JCR categories

Field	JCR Categories	
Chemistry	DW – Chemistry, Applied DX – Chemistry, Medicinal DY – Chemistry, Multidisciplinary EA – Chemistry, Analytical EC – Chemistry, Inorganic & Nuclear EE – Chemistry, Organic	EI – Chemistry, Physical HQ – Electrochemistry II – Engineering, Chemical GC – Geochemistry & Geophysics UH – Physics, Atomic, Molecular & Chemical
Biology	CU – Biology CX – Biology, Miscellaneous DR – Cell Biology HY – Developmental Biology	HT – Evolutionary Biology PI – Marine & Freshwater Biology QU – Microbiology WF – Reproductive Biology
Biochemistry	CO – Biochemical Research Methods CQ – Biochemistry & Molecular Biology	individual journals: Science, Nature, PNAS
Bioengineering	DA – Biophysics IG – Engineering, Biomedical	DB – Biotechnology & Applied Microbiology QE – Materials Science, Biomaterials

Fractional counting of papers

- For each journal:
  - Find JCR category assignments.
    - For example, the journal *Bioelectrochemistry* is in 4 JCR categories
      - CQ (Biochem & Mol Bio) – Biochemistry
      - CU (Biology) – Biology
      - DA (Biophysics) – Bioengineering
      - HQ (Electrochemistry) – Chemistry
  - Split paper counts among fields.
    - If there are 64 papers in *Bioelectrochemistry*, each of the four fields will get 16 papers.
  - Most journals are assigned to only 1 or 2 JCR categories.

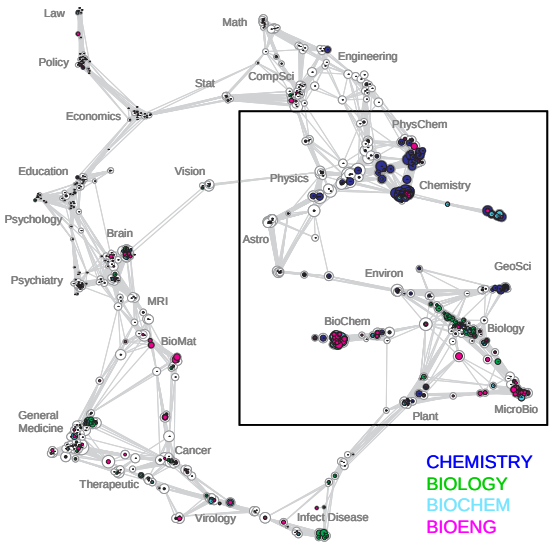
Chemistry on the 2002 base map



Börner: "Mapping the Evolving Interface of Mainstream Chemistry and the Fields of Biochemistry, Biology, and Bioengineering.

23

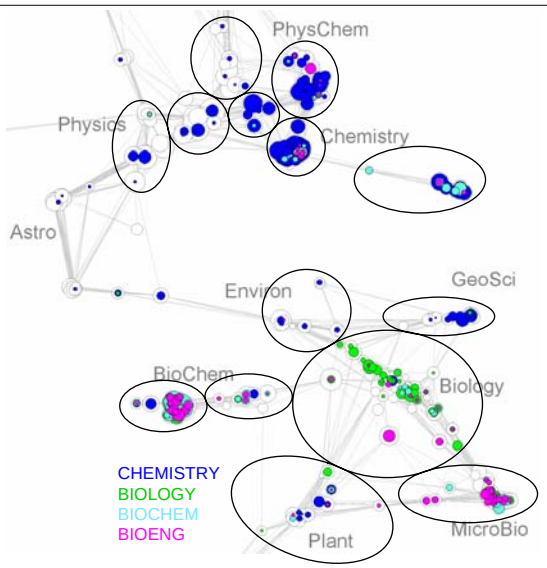
Four fields on the 2002 base map



Börner: "Mapping the Evolving Interface of Mainstream Chemistry and the Fields of Biochemistry, Biology, and Bioengineering.

24

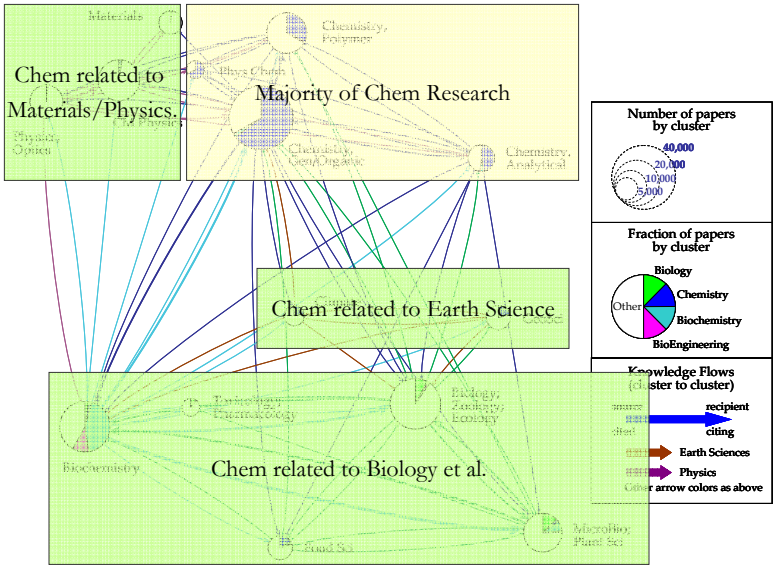
Definition of 14 disciplines



Börner: "Mapping the Evolving Interface of Mainstream Chemistry and the Fields of Biochemistry, Biology, and Bioengineering.

25

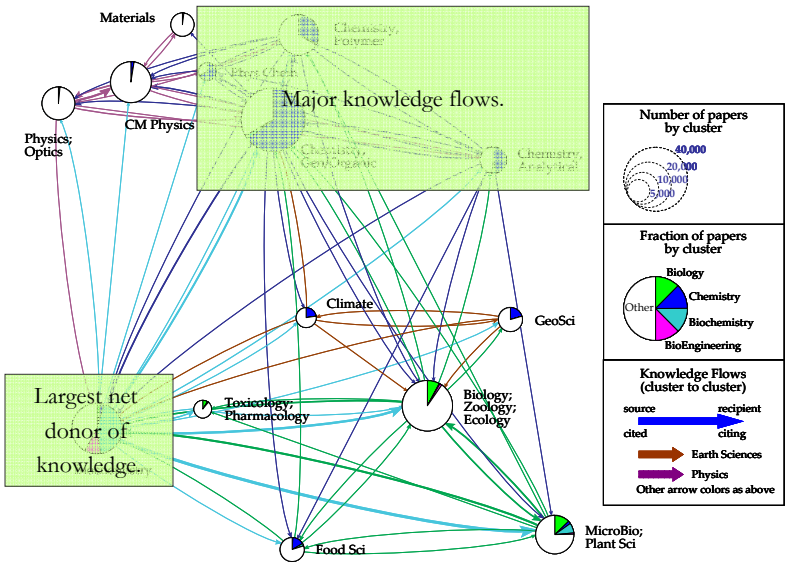
Symbolic map of 14 disciplines; 1974



Börner: "Mapping the Evolving Interface of Mainstream Chemistry and the Fields of Biochemistry, Biology, and Bioengineering.

26

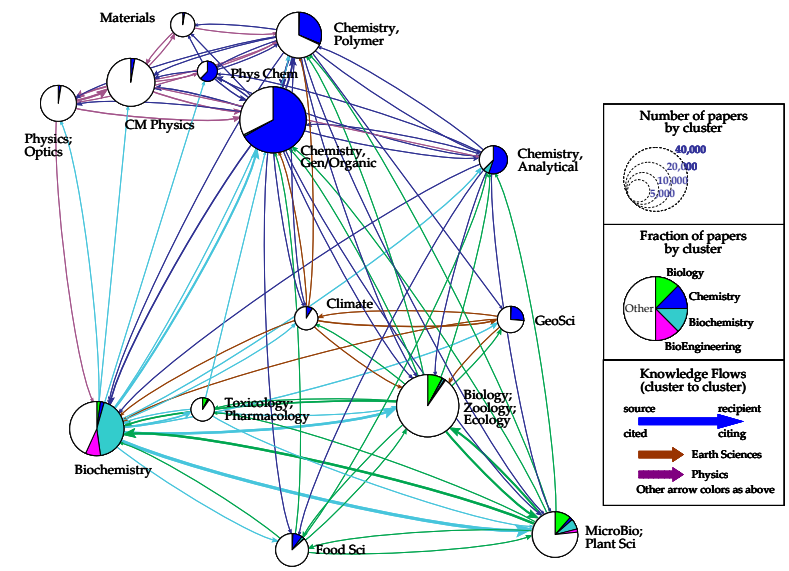
Symbolic map of 14 disciplines; 1974



Börner: "Mapping the Evolving Interface of Mainstream Chemistry and the Fields of Biochemistry, Biology, and Bioengineering.

27

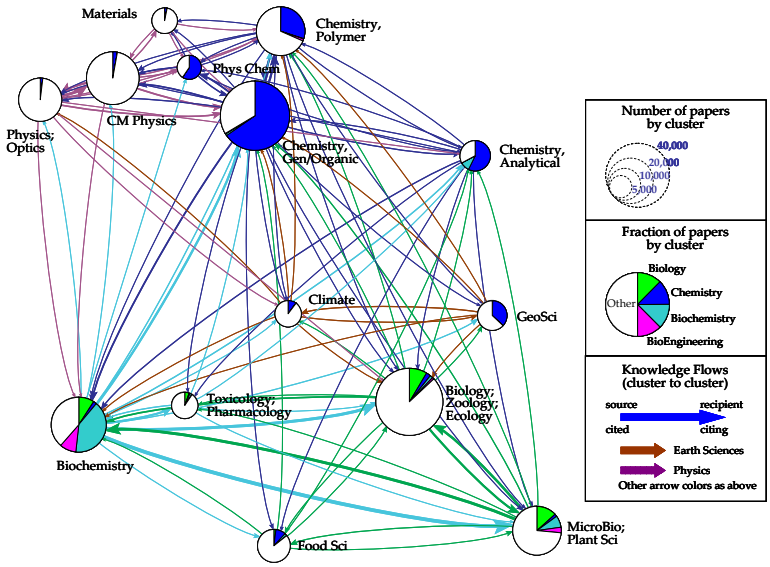
Symbolic map of 14 disciplines; 1979



Börner: "Mapping the Evolving Interface of Mainstream Chemistry and the Fields of Biochemistry, Biology, and Bioengineering.

28

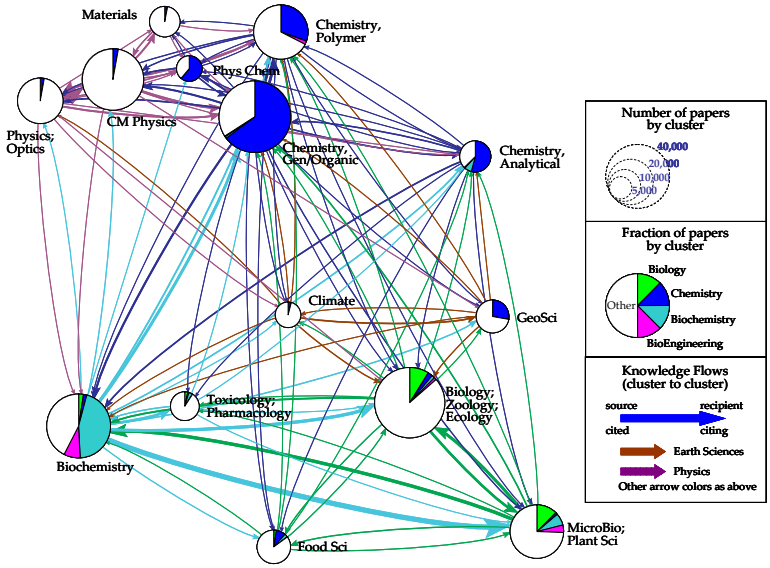
Symbolic map of 14 disciplines; 1984



Börner: "Mapping the Evolving Interface of Mainstream Chemistry and the Fields of Biochemistry, Biology, and Bioengineering.

29

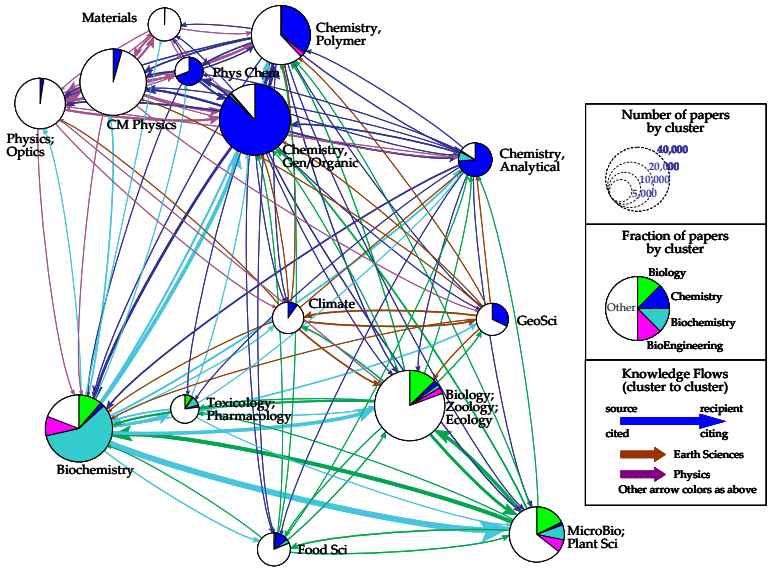
Symbolic map of 14 disciplines; 1989



Börner: "Mapping the Evolving Interface of Mainstream Chemistry and the Fields of Biochemistry, Biology, and Bioengineering.

30

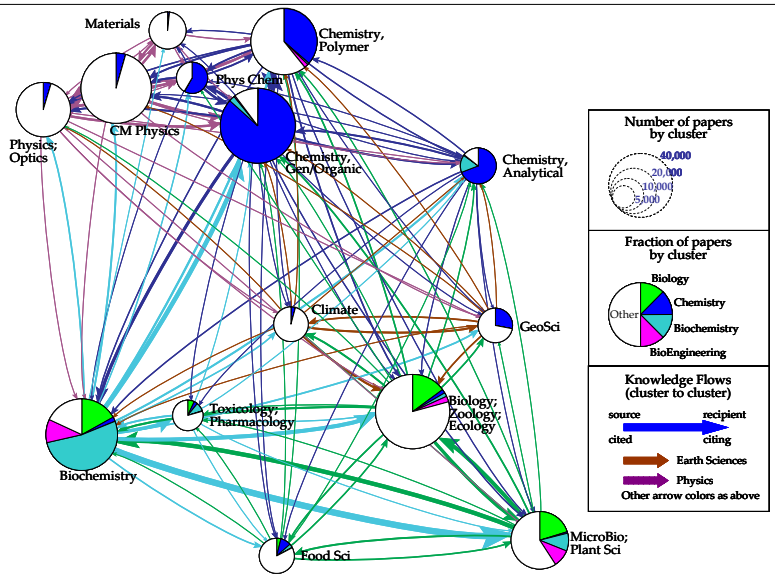
Symbolic map of 14 disciplines; 1994



Börner: "Mapping the Evolving Interface of Mainstream Chemistry and the Fields of Biochemistry, Biology, and Bioengineering.

31

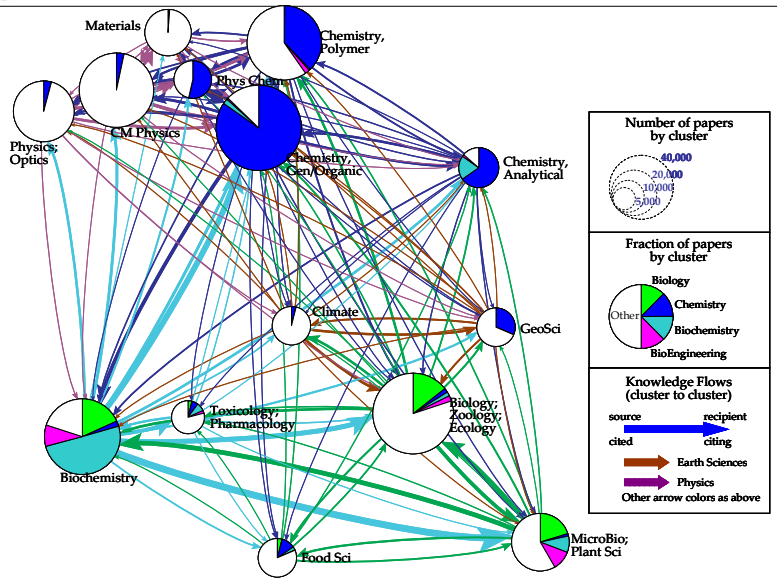
Symbolic map of 14 disciplines; 1999



Börner: "Mapping the Evolving Interface of Mainstream Chemistry and the Fields of Biochemistry, Biology, and Bioengineering.

32

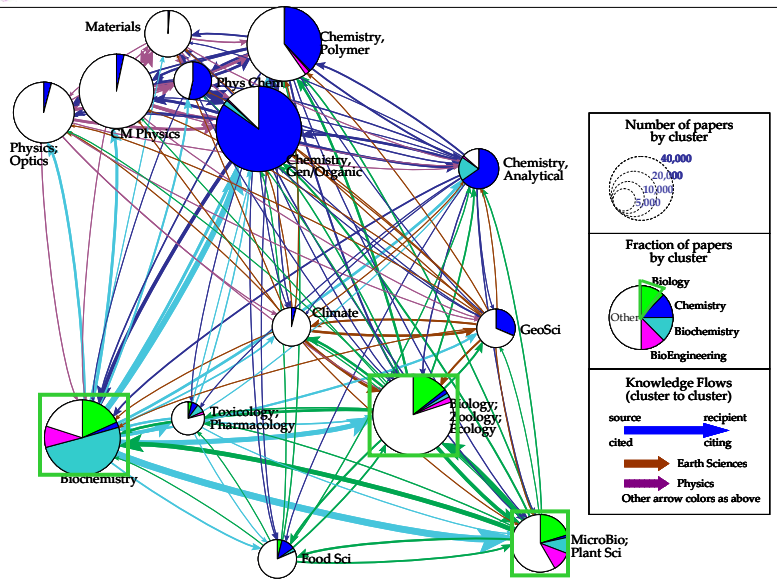
Symbolic map of 14 disciplines; 2004



Börner: "Mapping the Evolving Interface of Mainstream Chemistry and the Fields of Biochemistry, Biology, and Bioengineering.

33

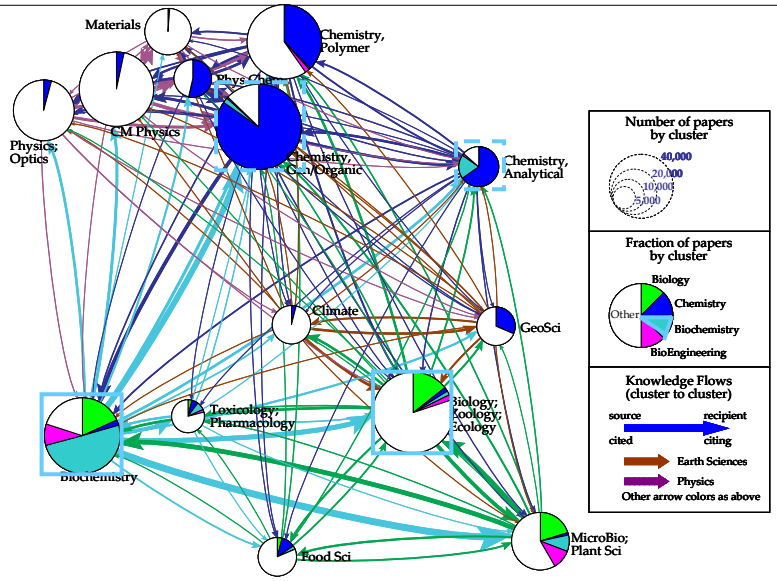
Symbolic map of 14 disciplines; 2004



Börner: "Mapping the Evolving Interface of Mainstream Chemistry and the Fields of Biochemistry, Biology, and Bioengineering.

34

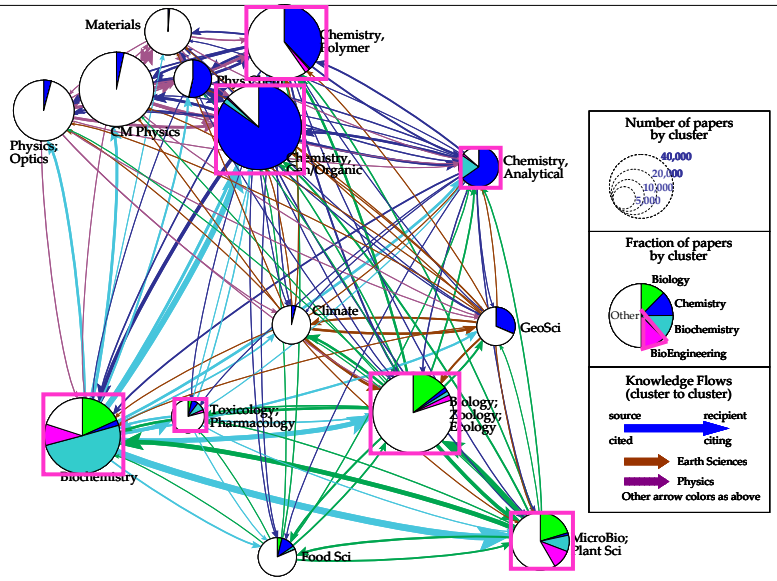
Symbolic map of 14 disciplines; 2004



Börner: "Mapping the Evolving Interface of Mainstream Chemistry and the Fields of Biochemistry, Biology, and Bioengineering.

35

Symbolic map of 14 disciplines; 2004



Börner: "Mapping the Evolving Interface of Mainstream Chemistry and the Fields of Biochemistry, Biology, and Bioengineering.

36

Summary

- Maps show the growth, distribution, and knowledge flows between *Chemistry*, *Biology*, *Biochemistry*, and *Bioengineering*.
- Over the past 30 years, *Biochemistry* and *Bioengineering* are moving steadily into *Chemistry* territory, and are having a large influence on the general knowledge base.
- *Chemistry's* impact on the knowledge base is growing, but at a slower rate.

- Journal-level data provides no information about the topics at the interface between fields, thus limiting the strategic decisions that can be made based on the mapping exercise.
- Folding in patent and or commercial data would provide a basis to study the impact of research on innovation and product development. It might very well be the case that some areas of science change their impact from a generator of cited scholarly knowledge to a generator of commercially valuable and hence patented and/or disclosed knowledge.
- Paper-level data would support the
  - identification of topics on the interfaces between fields, knowledge flows at topical levels, and detailed trends at these micro-levels.
  - analysis of the trajectories and impact of single researchers, teams, institutions, or nations.

Börner: "Mapping the Evolving Interface of Mainstream Chemistry and the Fields of Biochemistry, Biology, and Bioengineering.

37

The End.