Mapping Knowledge Domains

Katy Börner
School of Library and Information Science
INDIANA UNIVERSITY
BLOOMINGTON
katy@indiana.edu

School and Workshop on “Structure and Function of Complex Networks”,
Abdus Salam International Centre for Theoretical Physics (ICTP),
Trieste, Italy, May 24th, 2005.

Overview

1. Motivation for Mapping Knowledge Domains / Computational Scientometrics

2. Mapping the Structure and Evolution of
   Scientific Disciplines
   All of Sciences

3. Challenges and Opportunities
Mapping the Evolution of Co-Authorship Networks


Color Code:
Line color
99 - 99
98 - 98
97 - 97
96 - 96
Node color
0 - 9
10 - 19
20 - 29
30 - 39
40 - 49
50 -
- line color = year of first time co-author
- node color = number of citations
- node size = number of papers

1988

K. Börner
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.

### 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.

---

### 2. Mapping the Structure and Evolution of Knowledge Domains

<table>
<thead>
<tr>
<th>DATA EXTRACTION</th>
<th>UNIT OF ANALYSIS</th>
<th>MEASURES</th>
<th>LAYOUT/stylesheet (also can do both similarity and ordinal maps)</th>
<th>DISPLAY</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEARCHES</td>
<td>COMMON</td>
<td>COUNTS/FREQUENCIES</td>
<td>SCALAR (perl by unit matrix)</td>
<td>SCALAR</td>
</tr>
<tr>
<td>USE</td>
<td>CHOICES</td>
<td>Document, Author, Term, Co-citations</td>
<td>Dimension Reduction, Interaction, Intersection, Police's fit on any of above</td>
<td>SCALAR</td>
</tr>
<tr>
<td>MEDICINE</td>
<td></td>
<td>By year, By term</td>
<td>VECOR (perl by unit matrix)</td>
<td></td>
</tr>
<tr>
<td>RESEARCH</td>
<td></td>
<td>Thresholds, By counts, By year, By term</td>
<td>VECOR (perl by unit matrix)</td>
<td>VECOR</td>
</tr>
<tr>
<td>PROVIDERS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---


K. Börner
Indicator-Assisted Evaluation and Funding of Research

Mapping Medline Papers, Genes, and Proteins Related to Melanoma Research
Mapping Topic Bursts

Co-word space of the top 50 highly frequent and bursty words used in the top 10% most highly cited PNAS publications in 1982-2001.


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


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


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

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

Each institution is assumed to produce and consume information.

\[
\begin{align*}
\gamma_{82-86} &= 1.94 \quad (R^2=91.5\%) \\
\gamma_{87-91} &= 2.11 \quad (R^2=93.5\%) \\
\gamma_{92-96} &= 2.01 \quad (R^2=90.8\%) \\
\gamma_{97-01} &= 2.01 \quad (R^2=90.7\%)
\end{align*}
\]

Mapping all of Sciences

(in English speaking domain, based on available data)

Subsequent slides are based on

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.


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.

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

Zoom into structural map
Science maps for kids

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
- Model the diffusion of knowledge in evolving network ecologies.
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
  

- Model the diffusion of knowledge in evolving network ecologies.
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
- 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.