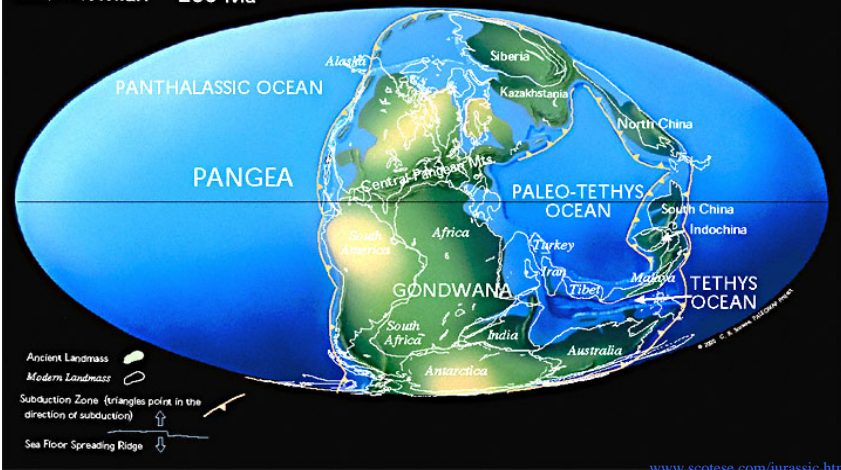
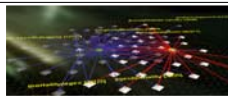


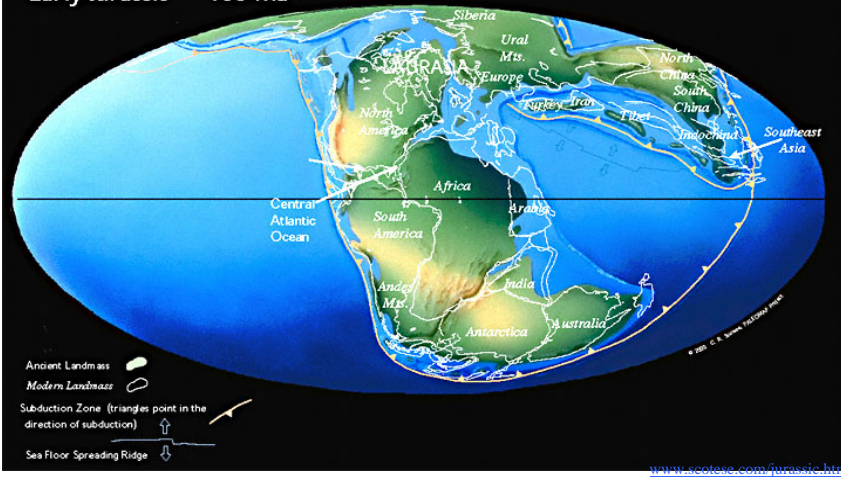
Late Permian 255 Ma



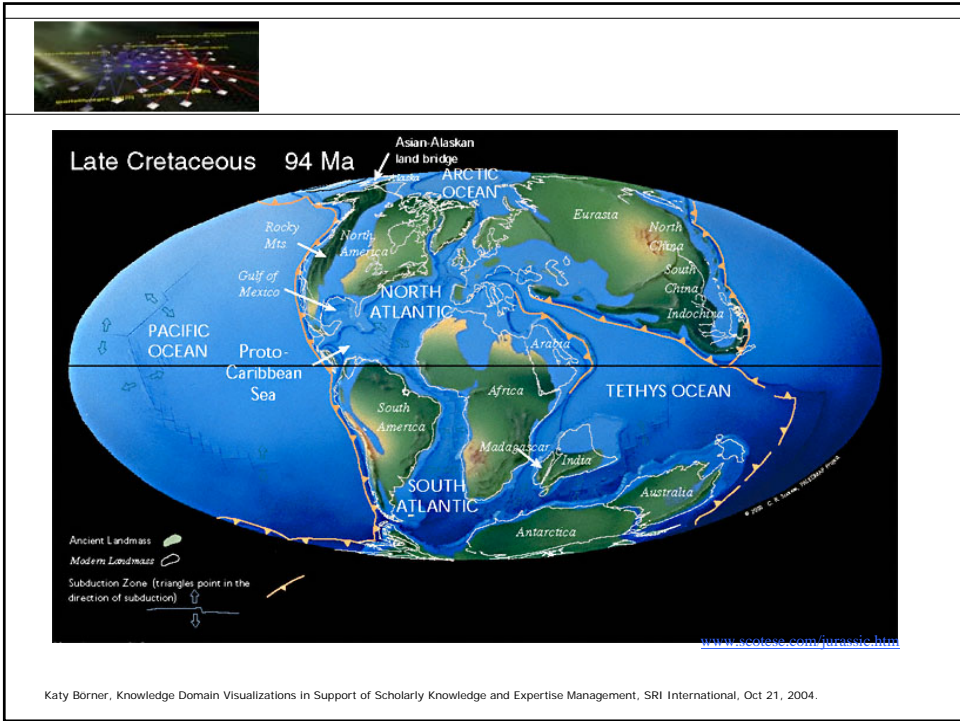
Katy Börner, Knowledge Domain Visualizations in Support of Scholarly Knowledge and Expertise Management, SRI International, Oct 21, 2004.



Early Jurassic 195 Ma



Katy Börner, Knowledge Domain Visualizations in Support of Scholarly Knowledge and Expertise Management, SRI International, Oct 21, 2004.



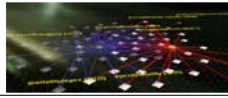
Session 4: Guidelines to Determine Specific Fields of S&E in SRS's Surveys

Knowledge Domain Visualizations in Support of Scholarly Knowledge and Expertise Management

Katy Börner
School of Library and Information Science

INDIANA UNIVERSITY
BLOOMINGTON
katy@indiana.edu

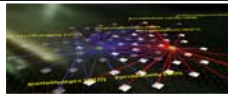
Panel Meeting on SRS's Evaluation of its Science and Engineering Taxonomies,
NSF sponsored Workshop, SRI International, Arlington, VA, Oct 21st, 2004.



Overview

1. Motivation
2. Knowledge Domain Analysis and Visualization
3. Cyberinfrastructure for InfoVis/KDVis Research
4. Conclusions

Katy Börner, Knowledge Domain Visualizations in Support of Scholarly Knowledge and Expertise Management, SRI International, Oct 21, 2004.



1. Motivation

Facing the Information Flood:

- Information available in electronic form doubles every 18 months.
- Human perception stays constant.
- Major means of accessing humanity's knowledge and expertise are search interfaces.

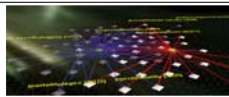
Opportunities & Challenges:

Apply advanced data mining and visualization techniques to study the structure and evolution of S&E development. Present results as visual maps to shift user's mental load from slow reading to faster perceptual processes such as visual pattern recognition.

Facilitated by:

- CPU speed & hard disk sizes have increased by two orders of magnitude.
- Bandwidth: Since the invention of the web browser, international IP bandwidth deployments have more than doubled each year.
- Monitor resolution has increased by a factor of 4 (800x600 -> 1600x1200).

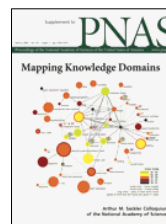
Katy Börner, Knowledge Domain Visualizations in Support of Scholarly Knowledge and Expertise Management, SRI International, Oct 21, 2004.



2. Analysis and Visualization of Knowledge Domains

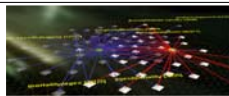
Based on publication, patent, grant, etc. data to help answer questions such as:

- 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.

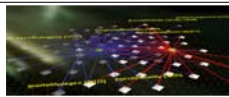
Katy Börner, Knowledge Domain Visualizations in Support of Scholarly Knowledge and Expertise Management, SRI International, Oct 21, 2004.



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**.

Katy Börner, Knowledge Domain Visualizations in Support of Scholarly Knowledge and Expertise Management, SRI International, Oct 21, 2004.



Process of Mapping Knowledge Domains

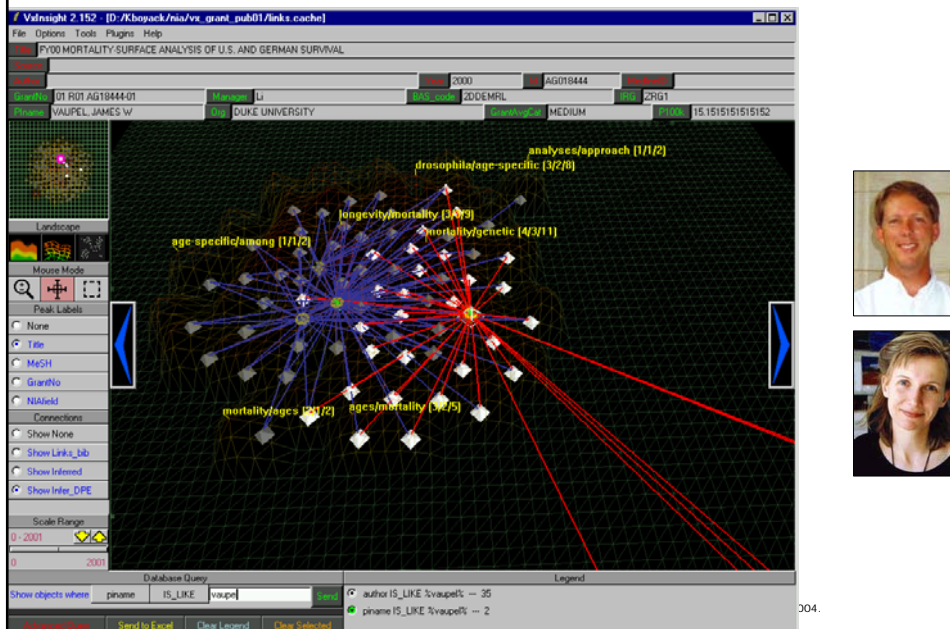
DATA EXTRACTION	UNIT OF ANALYSIS	MEASURES	LAYOUT (often one code does both similarity and ordination steps)		DISPLAY
			SIMILARITY	ORDINATION	
SEARCHES ISI INSPEC Eng Index Mbdline ResearchIndex Patents etc.	COMMON CHOICES Journal Document Author Term	COUNTS/FREQUENCIES Attributes (e.g. terms) Author 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) ind. 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, Katy, Chen, Chaomei, and Boyack, Kevin. (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.

Katy Börner, Knowledge Domain Visualizations in Support of Scholarly Knowledge and Expertise Management, SRI International, Oct 21, 2004.

Indicator-Assisted Evaluation and Funding of Research

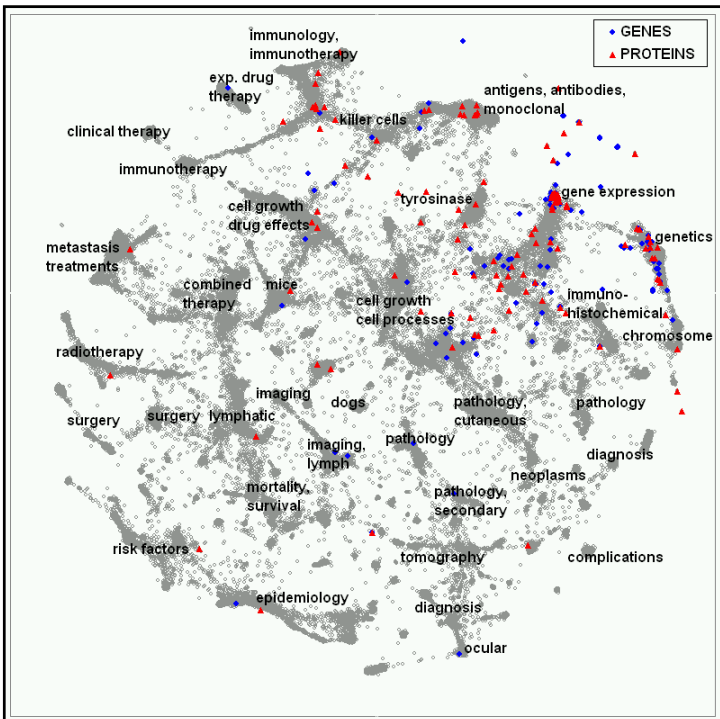
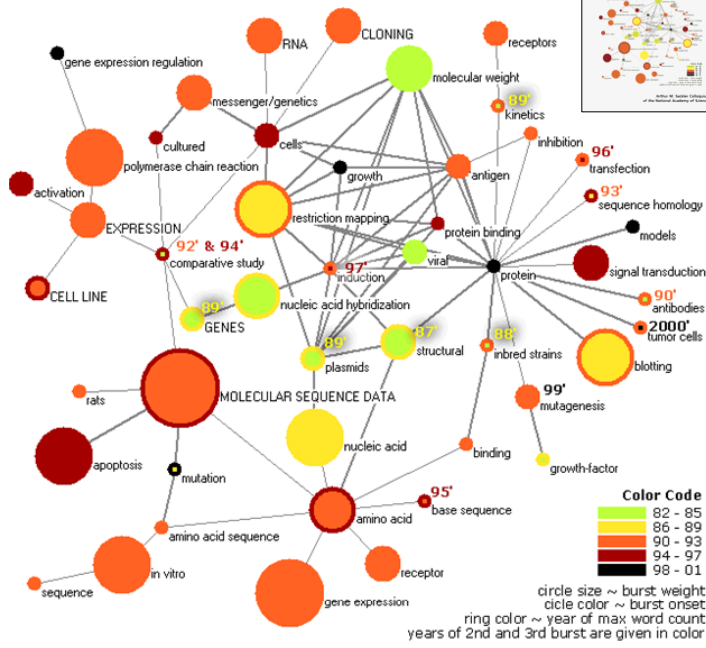
Visualizing the influence of grants on the number and citation counts of research papers (Boyack & Börner, 2003)



Mapping Topic Bursts

(Mane & Börner, 2004)

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.

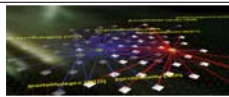


Mapping Medline Papers, Genes, and Proteins Related to Melanoma Research

(Boyack, Mane & Börner, 2004)

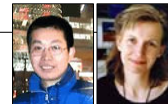


Oct 21, 2004.



Co-PI Map of Current IDM Awardees

(Ke & Börner, 2004)



Legend

Node size: # awarded grants

Node Inner Color: # unique

Co-PIs (0 white; 1

YellowGreen; 2 Green; 3

PineGreen; 4 Orange; 5

Red; 6 Maroon)

Node Border Color: Grant

Source (Career "Yellow";

Peach "Blue"; ITR

"Green"; SGER "Pink";

other "White"; Multi-

Grants "Red")

Edge Width: # times people

Co-PI'd

Edge Color: First year of Co-

PIship (1999 Maroon; 2000

Red; 2001 Orange; 2002

PineGreen; 2003 Green;

2004 YellowGreen)

Career awardees are not

showing except they have

other IDM grant support as

well.



Katy Börner, Knowledge Domain Visualizations in Support of Scholarly Knowledge and Expertise Management, SRI International, Oct 21, 2004.

Mapping the Evolution of Co-Authorship Networks

Won 1st price at the IEEE InfoVis Contest

(Ke, Viswanath & Börner, 2004)



Color Code:

Line color

86 - 90

91 - 95

96 - 00

01 -

Node color

0 - 9

10 - 19

20 - 29

30 - 39

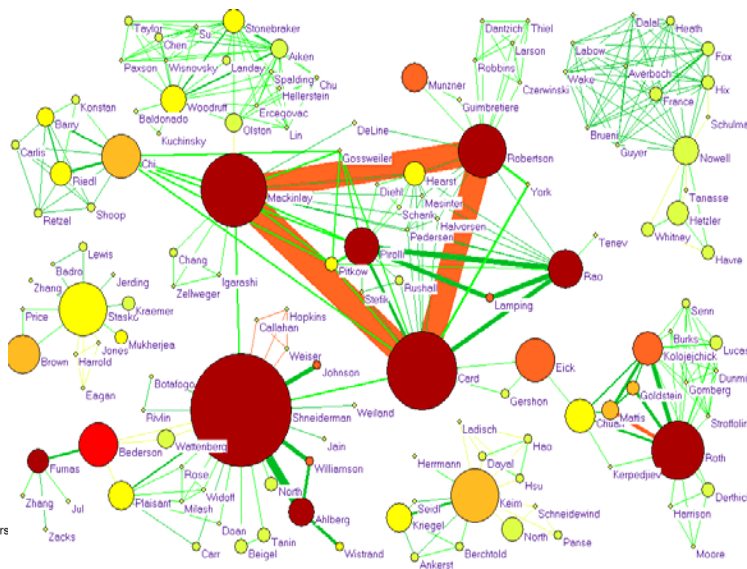
40 - 49

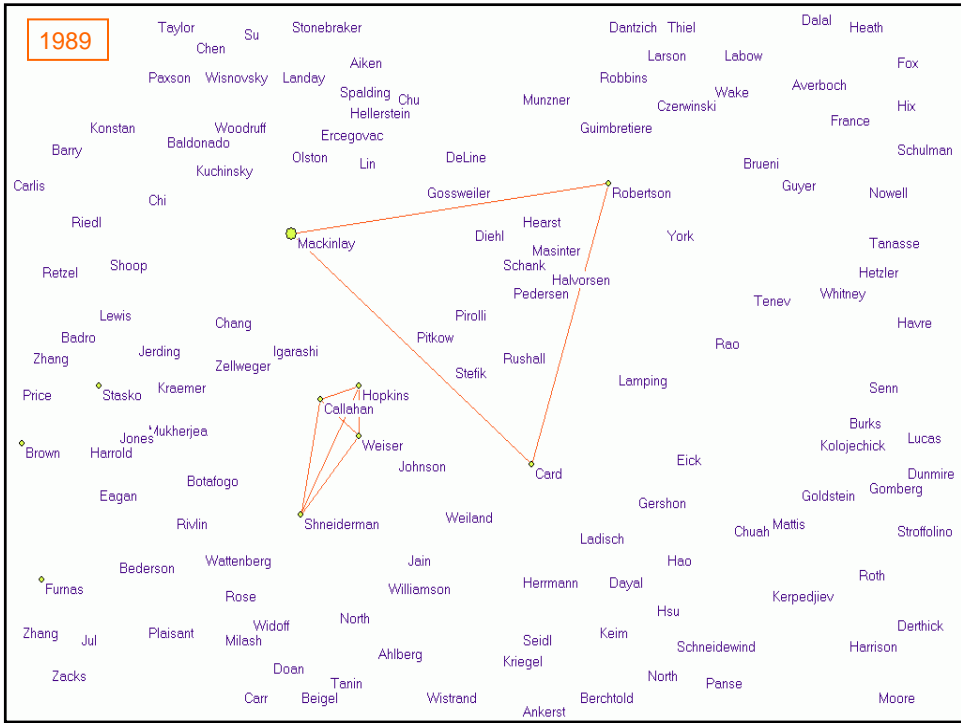
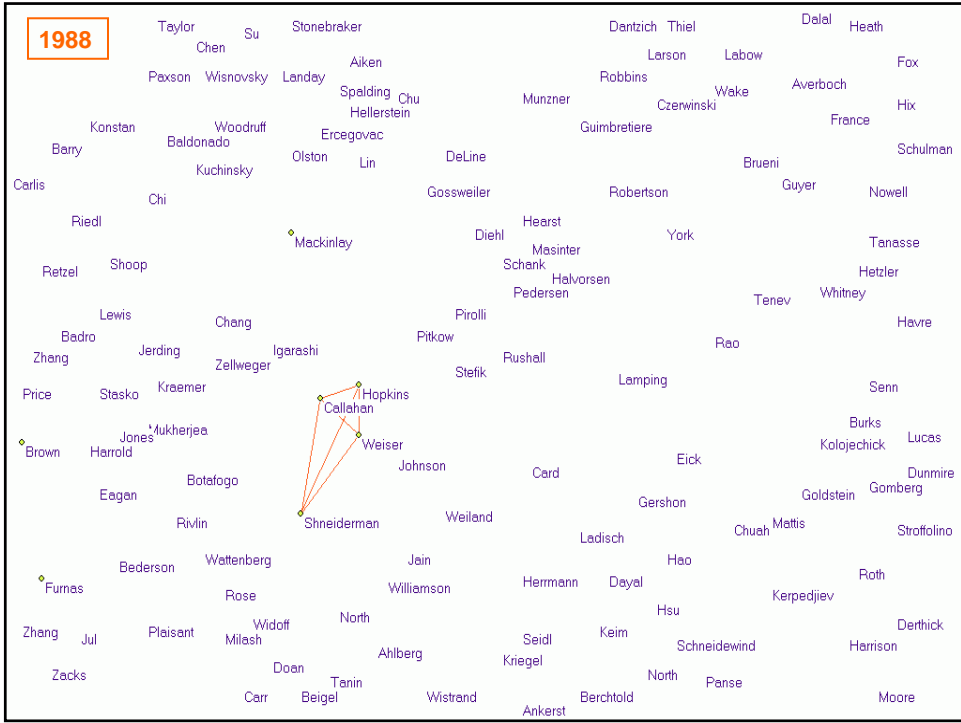
50 -

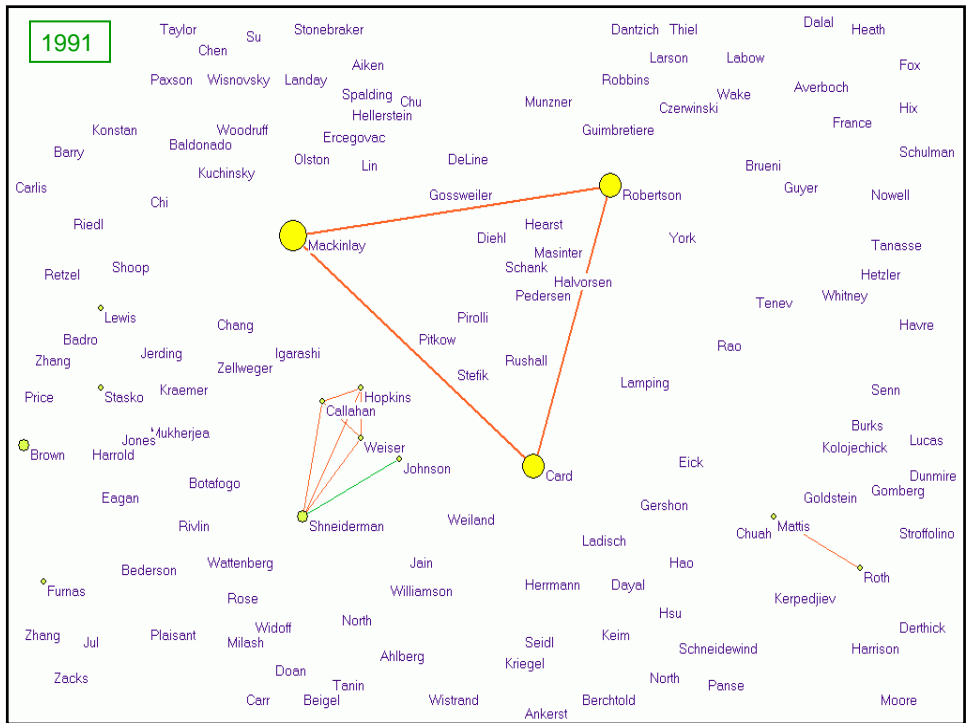
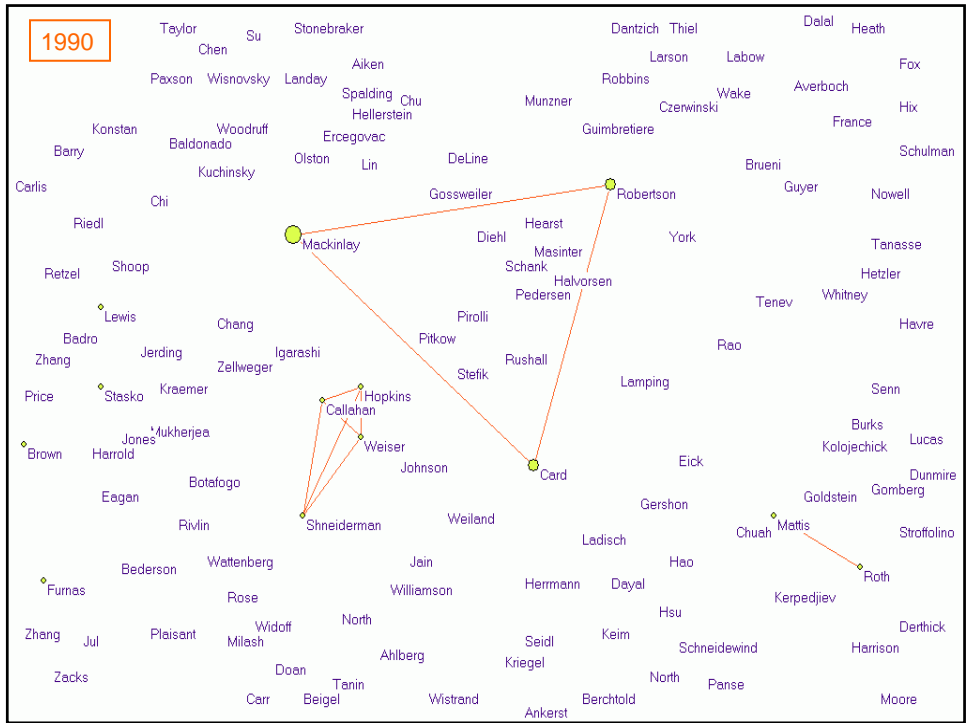
line color ~ year of first time co-authors

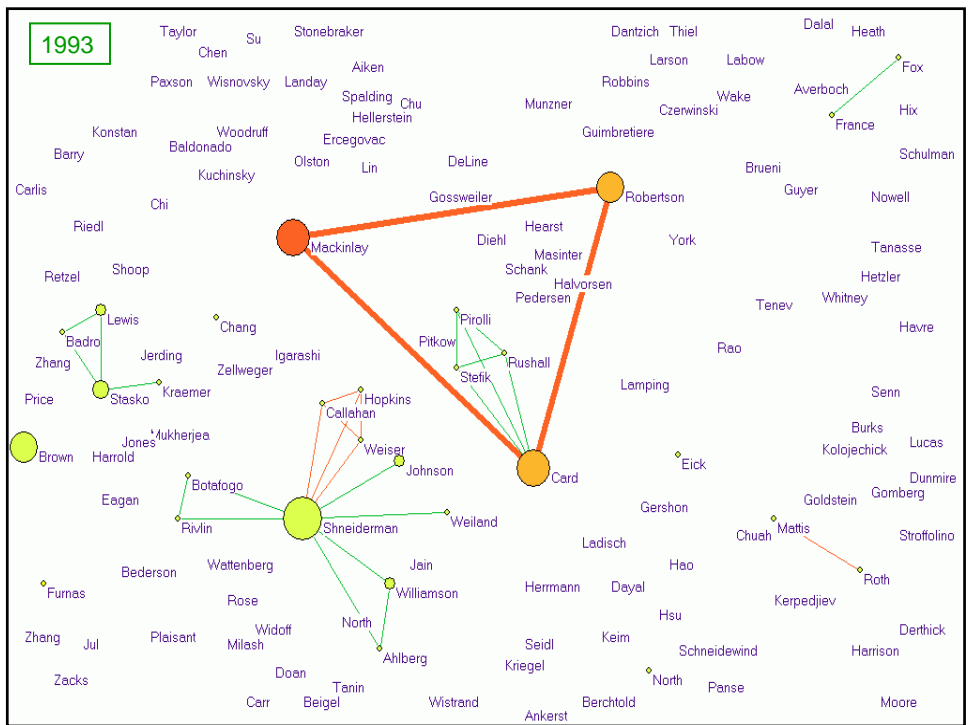
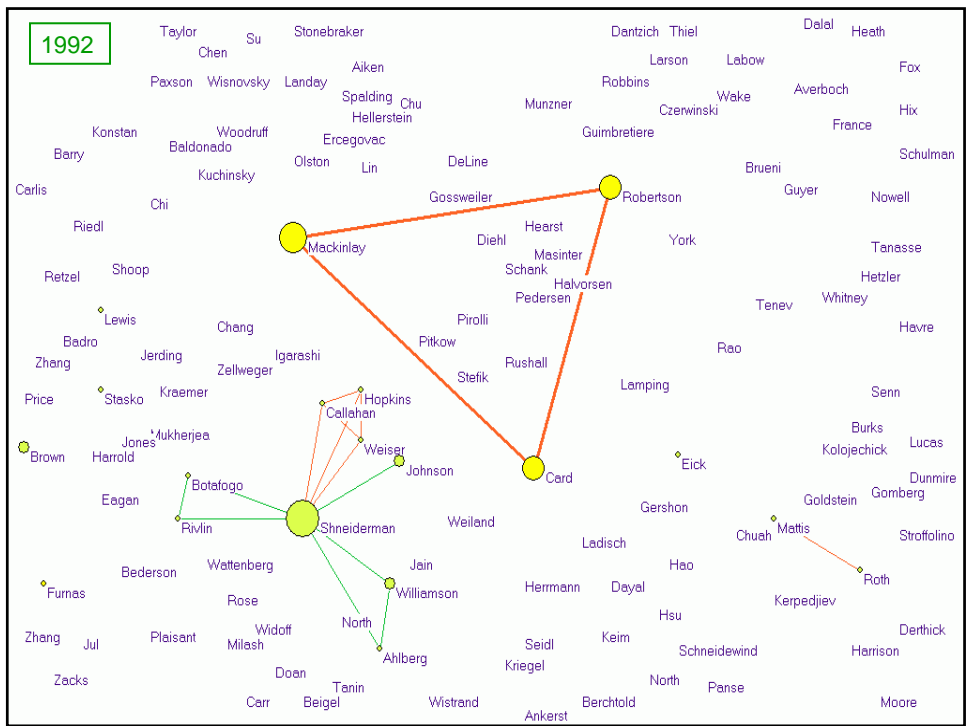
node color ~ number of citations

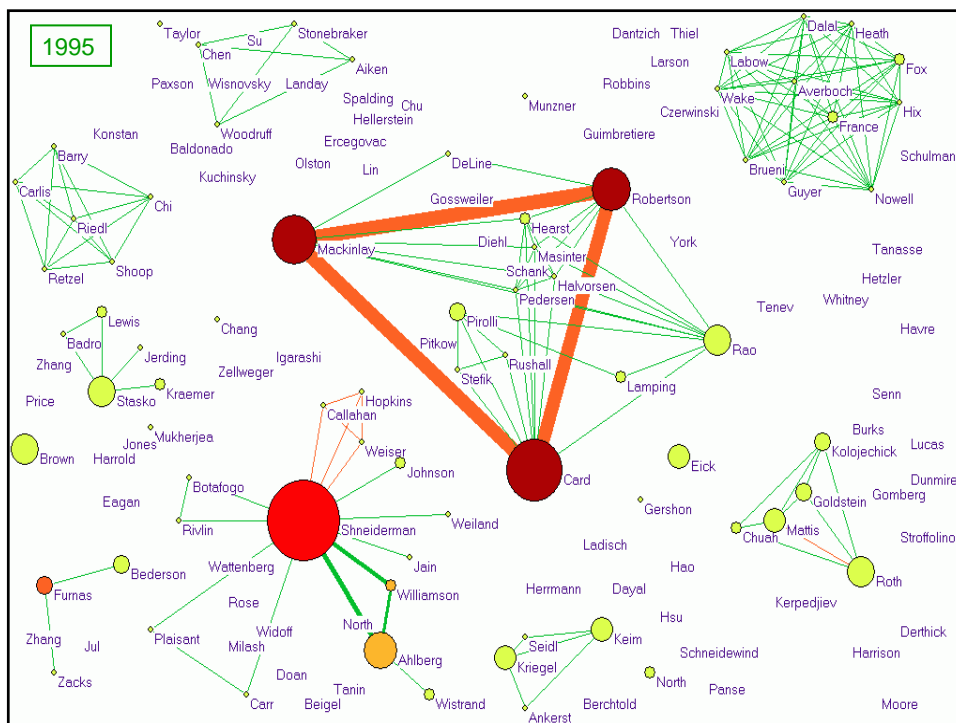
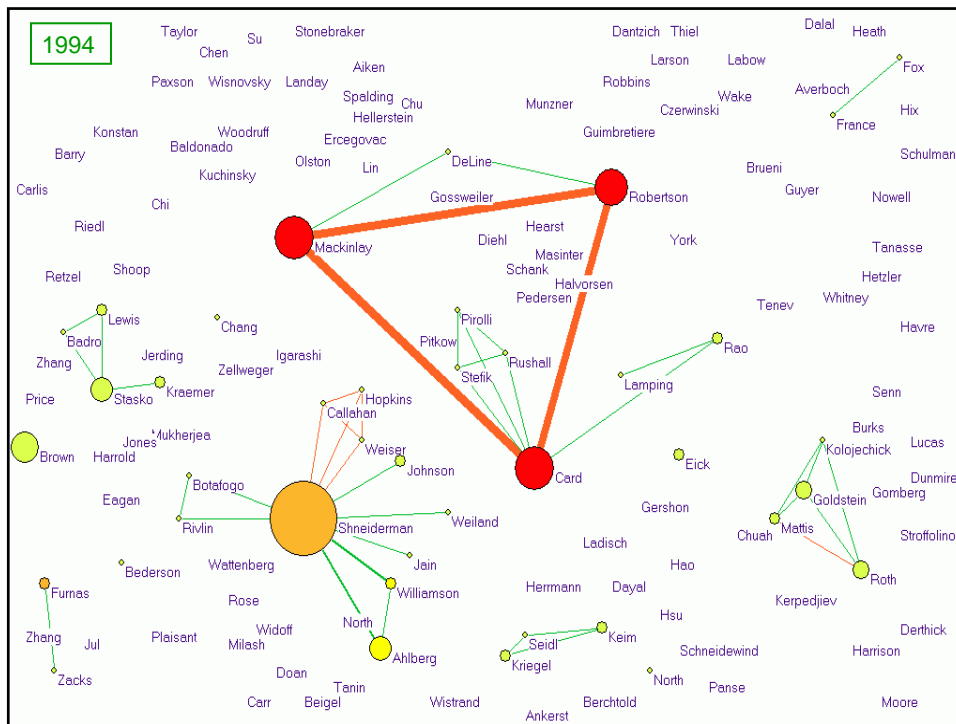
node size ~ number of papers

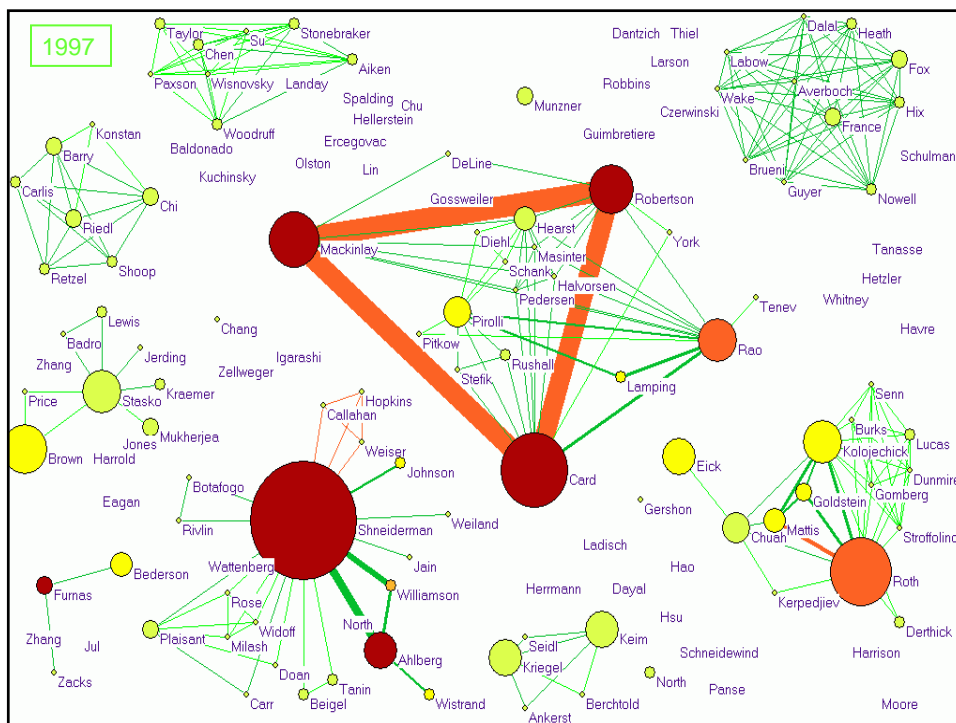
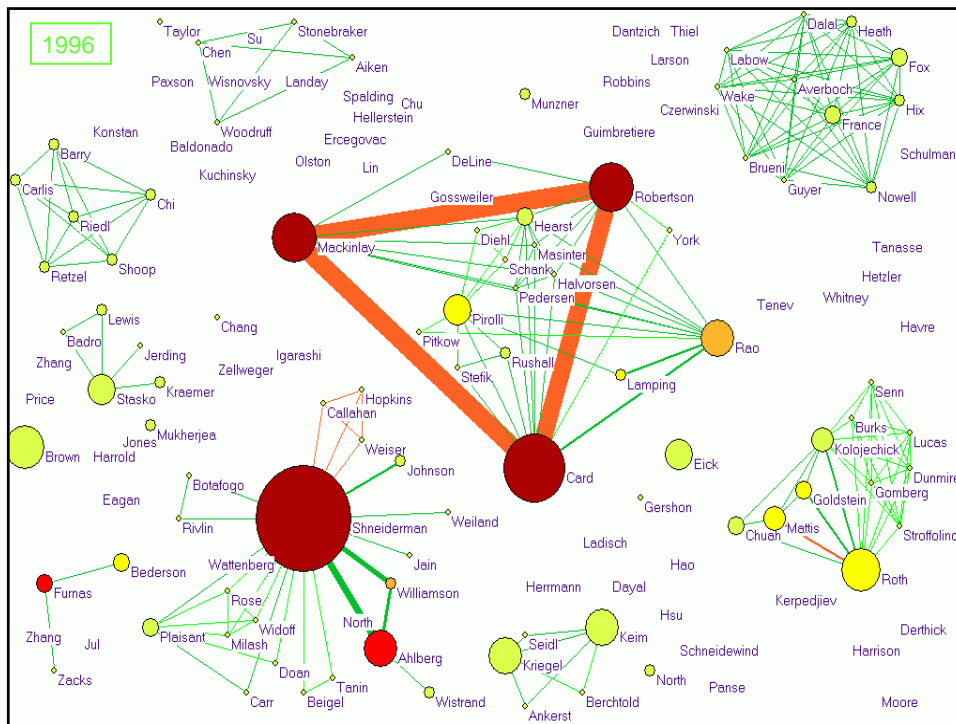


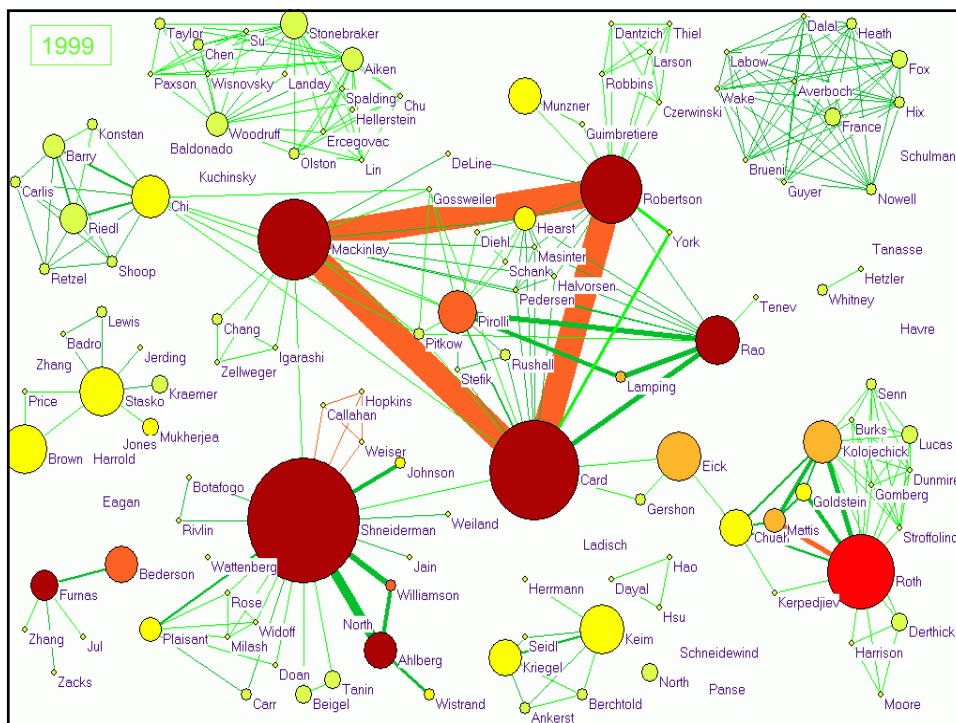
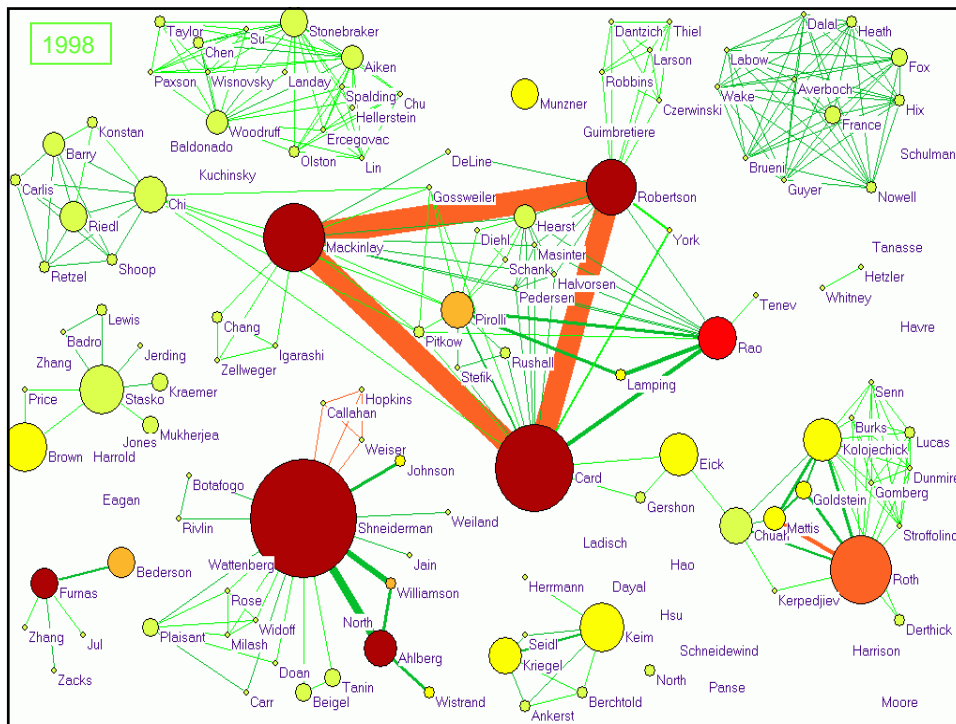


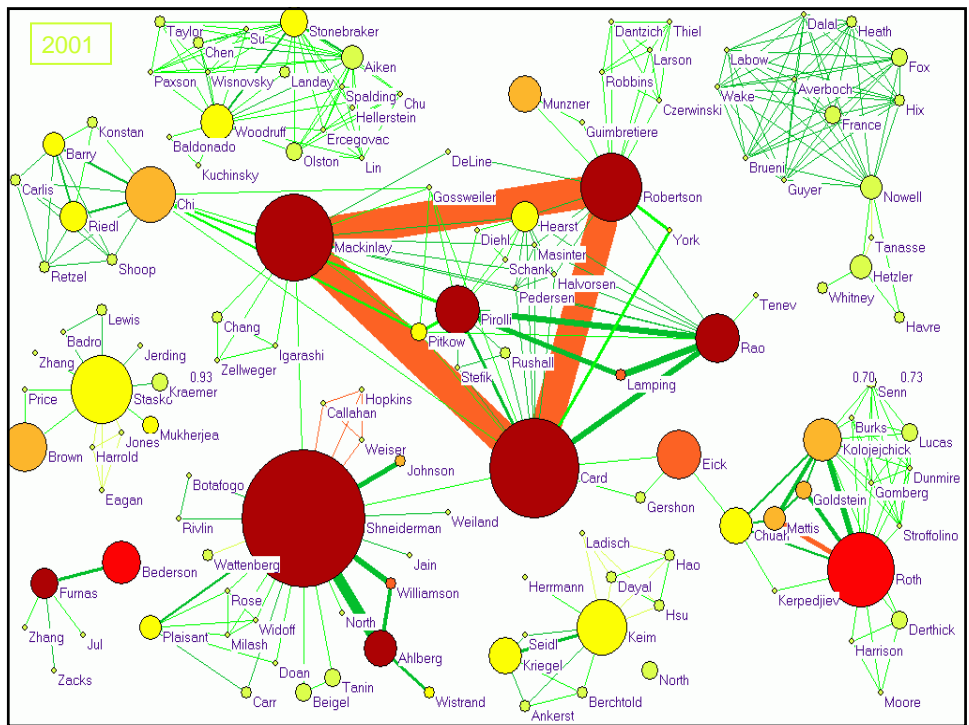
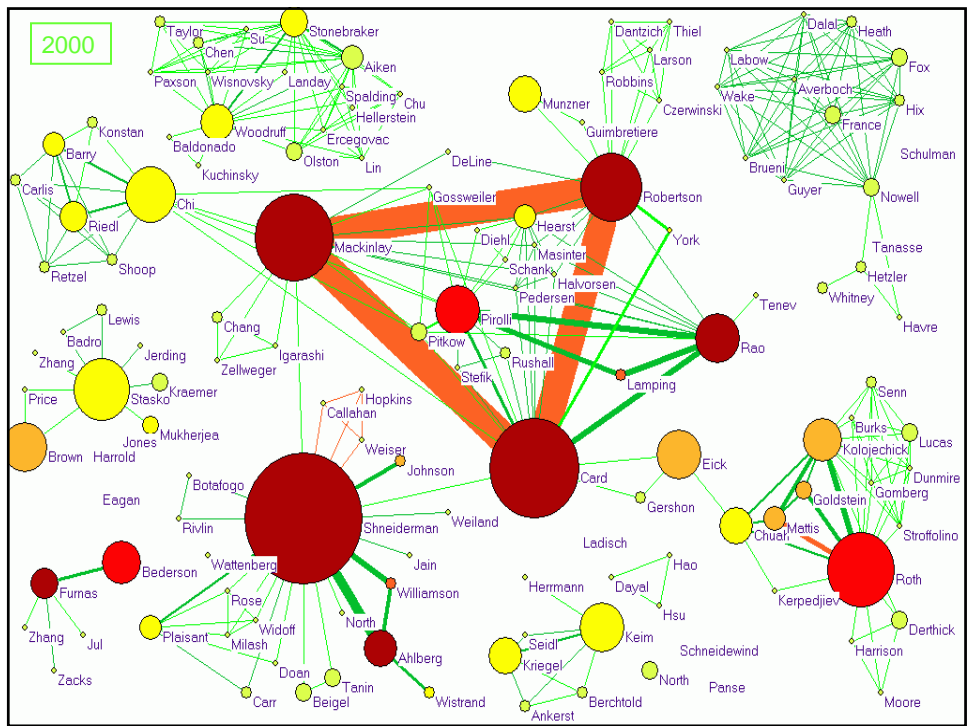


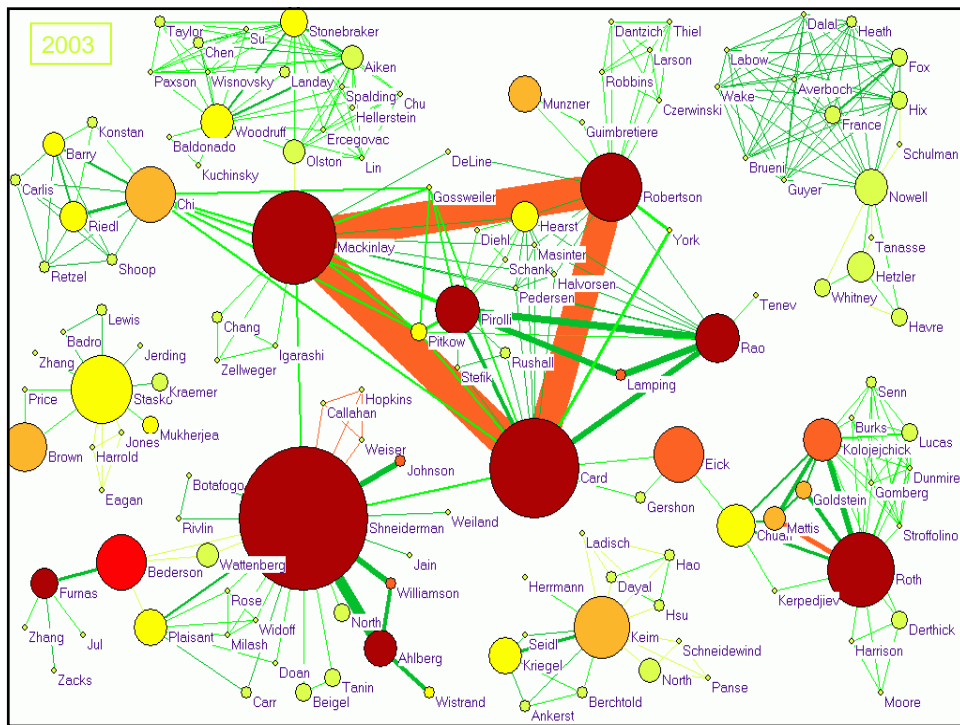
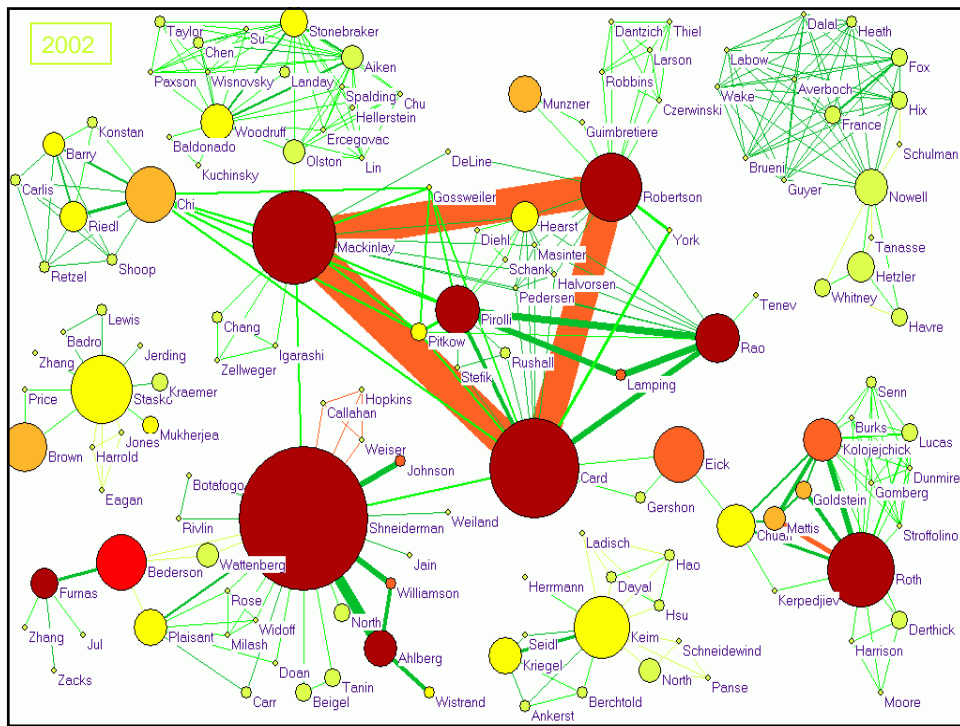


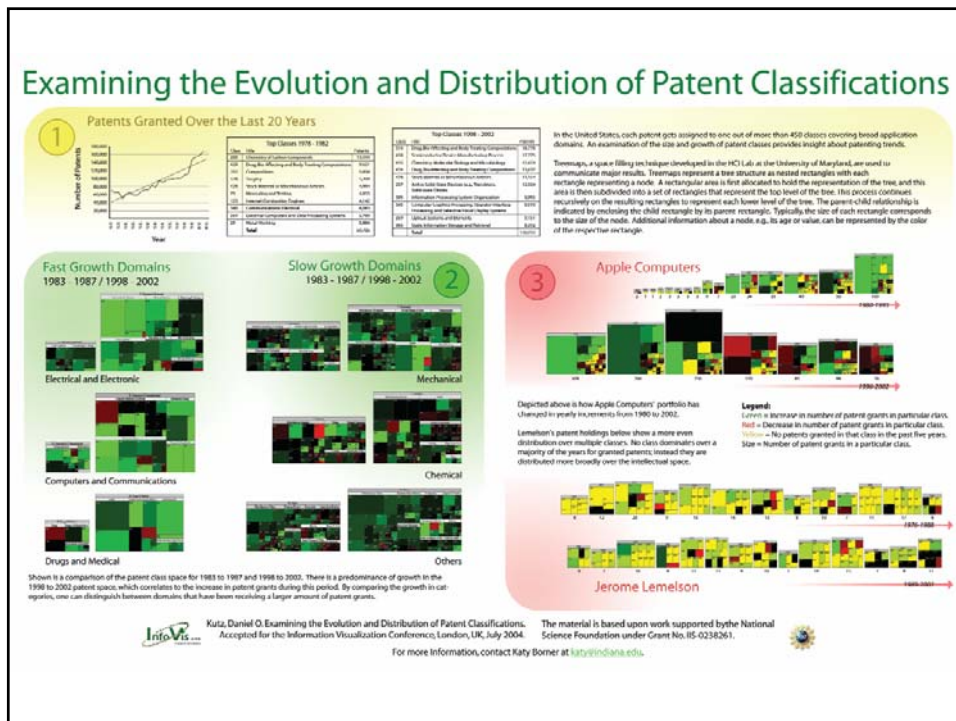
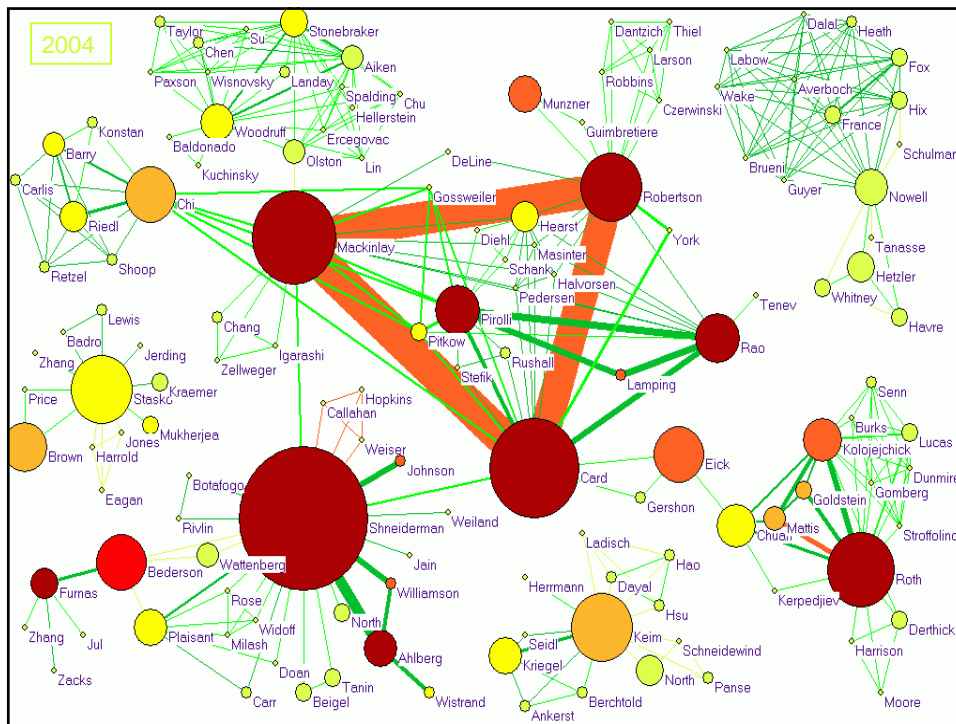


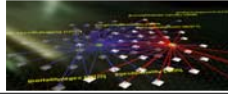












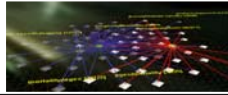
- Diverse attempts have been made to generate maps of science.
- Most have concentrated on specific knowledge domains due to data availability and scalability of algorithms.
- Cartographic metaphors seem to work well as they exploit the map reading skills people acquire in their education.
- Ideally, maps of science would resemble weather forecast maps in that they not only show the structure but also the dynamics of scientific evolution and progress.

It is just today, that we have the data, code and compute power to study science using the scientific methods of science as suggested by Derek J. deSolla Price about 40 years ago.

However, generating a map of science requires a computational effort common in physics or biology but not in the social sciences.

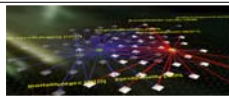
However, maps of science will benefit every field.

Katy Börner, Knowledge Domain Visualizations in Support of Scholarly Knowledge and Expertise Management, SRI International, Oct 21, 2004.



3. Cyberinfrastructure for InfoVis / KDVis Research

Katy Börner, Knowledge Domain Visualizations in Support of Scholarly Knowledge and Expertise Management, SRI International, Oct 21, 2004.



3. Cyberinfrastructure for InfoVis / KDVis Research

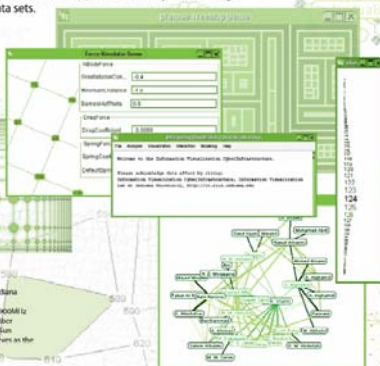
Information Visualization CyberInfrastructure

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.

<http://ivc.slis.indiana.edu/db>



SOFTWARE

An open source IVC 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".

<http://ivc.slis.indiana.edu/db>



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 use different algorithms, test diverse interaction techniques and design features, and to quickly generate and compare information visualizations.

<http://ivc.slis.indiana.edu/db>

COMPUTING RESOURCES

The InfoVis CyberInfrastructure is housed at Indiana University's Research Database Complex comprising of two Sun V120 servers with 12 400MB i2 processors and 16 GB of memory each. 18 78-fiber channel disks are attached to both servers. A Sun V180 system with 4 gpus and 8GB memory is open on the web front-end for the database servers.

<http://ivc.slis.indiana.edu/db>



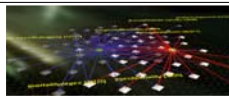
IRWIN LAB, School of Library and Information Science, Indiana University (2005). For more information, contact Katy Börner at kborner@slis.indiana.edu

This material is based upon work supported by the National Science Foundation under Grant Nos. IRI-0226261 and DUE-0318624.



From <http://ivc.slis.indiana.edu/db>

Katy Börner, Knowledge Domain Visualizations in Support of Scholarly Knowledge and Expertise Management, SRI International, Oct 21, 2004.



IVC Database (<http://ivc.slis.indiana.edu/db>)

Papers and Patents



Medline
Number of Entries: 11,693,477
Years covered: 1963-2002
Size: 135 MB (gunzipped)



Proceedings of the National Academy of Science (PNAS)
Number of Entries: 16,169
Years covered: 1987-2002
Size: 583 MB



United States Patent and Trademark Office (Patents)
Number of Entries: 2,582,647
Years covered: 1976-2003
Size: 350 MB

Grant Awards



National Science Foundation (NSF)
Number of Entries: 181,132
Years covered: 1985-2002
Size: 400 MB



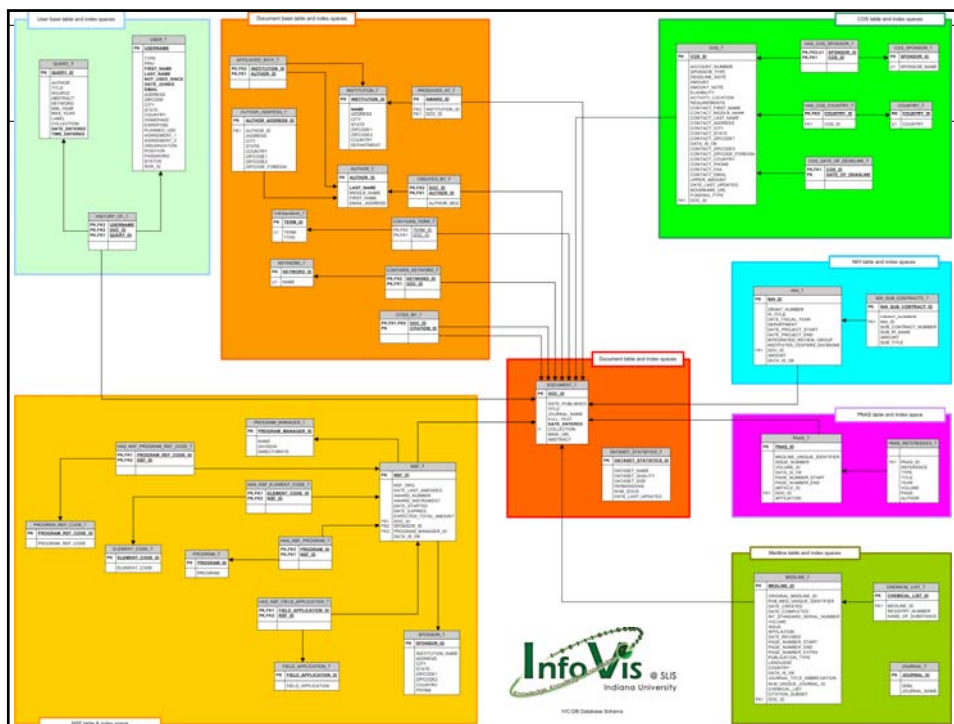
National Institute of Health (NIH)
Number of Entries: 1,003,521
Years covered: 1972-1992 and 1994-2002
Size: 2.3 GB

Funding Opportunities



Community of Science (COS)
Number of Entries: 38,154 (5,000 new entries per month)
Years covered: 2001-present
Size: 60 MB

Katy Börner, Knowledge Domain Visualiz:



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. (<http://ivis.indiana.edu/>)

SOFTWARE

An open source IVC 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". (<http://ivis.indiana.edu/>)

COMPUTING RESOURCES

The InfoVis CyberInfrastructure is hosted at Indiana University's Research Database Complex, comprising of two Sun V230 servers with 12 9000Mhz processors and 96 GB of memory each. 6 TB fiber channel disks are attached to both servers. A Sun V880 system with 4 cpus and 8GB memory serves as the web front-end for the database servers. (<http://ivis.indiana.edu/>)

...to equip coding code algorithms, test diverse interaction techniques and design features, and to quickly generate and compare information visualizations. (<http://ivis.indiana.edu/>)

InfoVis Lab, School of Library and Information Science, Indiana University (2004). For more information, contact Katy Börner at kborner@indiana.edu. This material is based upon work supported by the National Science Foundation under Grants No. IS-0228261 and DUE-0319623.

Katy Börner, Knowledge Domain Visualizations in Support of Scholarly Knowledge and Expertise Management, SRI International, Oct 21, 2004.

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. (<http://ivis.indiana.edu/db/>)



COMPUTING RESOURCES

The InfoVis CyberInfrastructure is hosted at Indiana University's Research Database Complex comprising of two Sun X1280 servers with 12 900MHz processors and 96 GB of memory each. 6 TB fiber channel disks are attached to both servers. A Sun VM80 system with 4 cpus and 8GB memory serves as the web front-end for the database servers. (<http://ivis.indiana.edu/>)

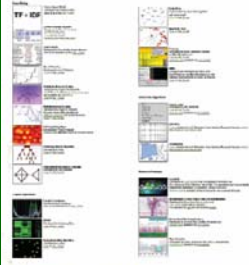


InfoVis Lab, School of Library and Information Science, Indiana University (2004).
For more information, contact Katy Börner at kborner@slis.indiana.edu

This material is based upon work supported by the National Science Foundation under Grant Nos. IIS-0228261 and DUE-0339624.

SOFTWARE

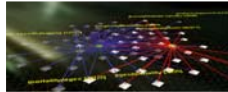
An open source IVC framework was designed to facilitate the integration of diverse data analysis, modeling and visualization algorithms. New algorithms, data persistence methods, look and feels for the interface and entire toolkits can be easily "plugged in" or "unplugged". (<http://ivis.indiana.edu/iv/>)



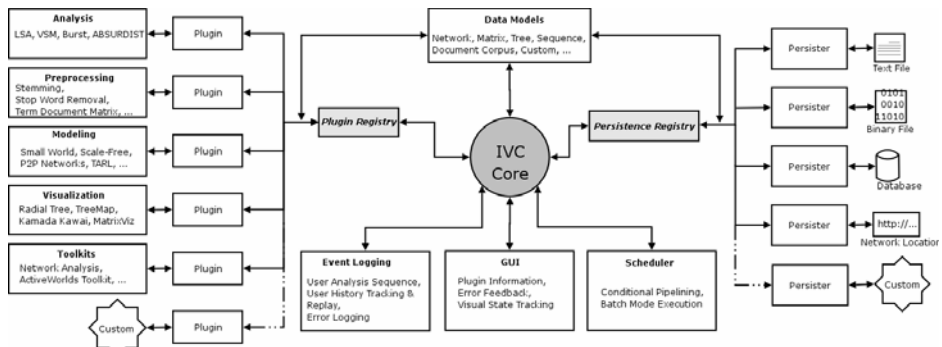
LEARNING MODULES

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 interaction techniques and design features, and to quickly generate and compare information visualizations. (<http://ivis.indiana.edu/iv/>)

Katy Börner, Knowledge Domain Visualizations in Support of Scholarly Knowledge and Expertise Management, SRI International, Oct 21, 2004.



IVC Software Framework (<http://ivis.indiana.edu/iv/>)



Katy Börner, Knowledge Domain Visualizations in Support of Scholarly Knowledge and Expertise Management, SRI International, Oct 21, 2004.



Workshop on Information Visualization Software Infrastructures

Sat. Oct. 9th, 2004, 10am-6pm
Room Hill Country A-B

Right before the [IEEE Symposium on Information Visualization \(InfoVis\) 2004.](#)

Workshop Chairs

- Jean-Daniel Fekete**, INRIA Futurs, France (Author of [The InfoVis Toolkit](#)), Jean-Daniel.Fekete@inria.fr, Home page: <http://www.lri.fr/~fekete/index.en.html>
- Katy Börner**, Indiana University, USA (Co-Author of the [InfoVis Cyberinfrastructure](#)) katy@indiana.edu, Home page: <http://ella.slis.indiana.edu/~katy/>

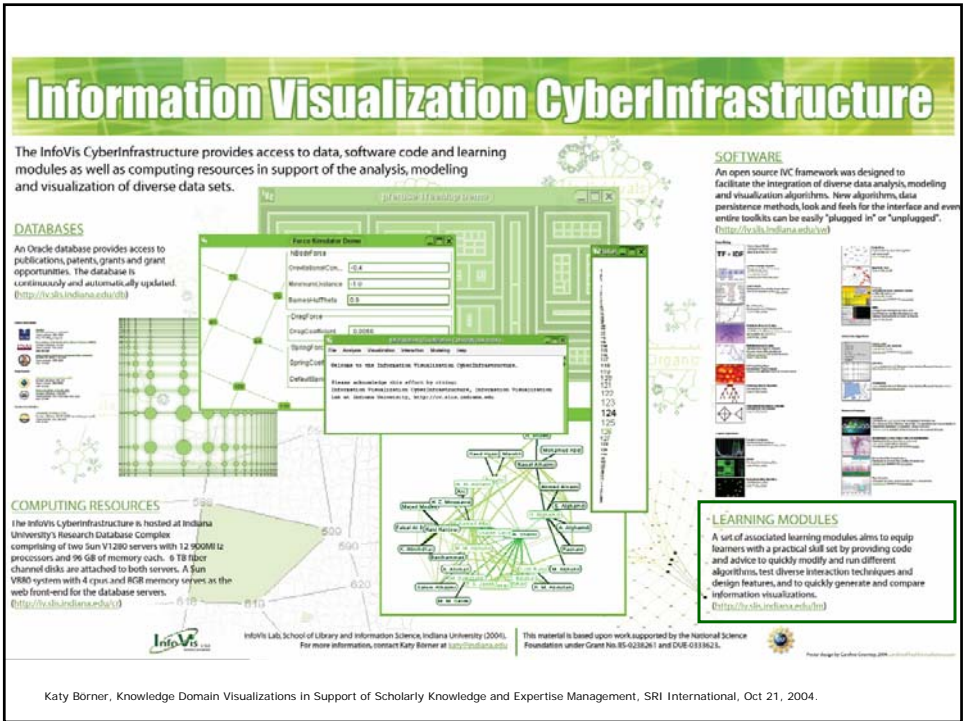
Description

Information visualization systems and toolkits are becoming available for a large range of visualization and interaction techniques and are used in diverse application domains. This workshop is aimed at gathering experts involved in building such infrastructures to share their views, understand the issues involved and trying to find ways to avoid fragmentation and improve collaborations.

To participate in the workshop, you should submit a semi-structured position paper explaining your view of what an infrastructure should provide, describe what you consider as the main challenges for such infrastructures and describe the capabilities of toolkits of systems you have already built, following a form available [here](#) by Sept 30th, 2004.

<http://www.indiana.edu/irvi2004/>

Katy Börner, Knowledge Domain Visualizations in Support of Scholarly Knowledge and Expertise Management, SRI International, Oct 21, 2004.



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. (<http://ivis.slis.indiana.edu/>)

COMPUTING RESOURCES

The InfoVis Cyberinfrastructure is hosted at Indiana University's Research Database Complex consisting of two Sun V2200 servers with 12 9000Mhz processors and 96 GB of memory each. 6 TB fiber channel disks are attached to both servers. A Sun V880 system with 4 cpus and 4GB memory serves as the web front-end for the database servers. (<http://ivis.slis.indiana.edu/>)

SOFTWARE

An open source IVC 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". (<http://ivis.slis.indiana.edu/>)

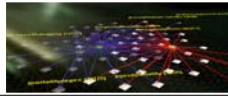
LEARNING MODULES

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 interaction techniques and design features, and to quickly generate and compare information visualizations. (<http://ivis.slis.indiana.edu/>)

InfoVis Lab, School of Library and Information Science, Indiana University (2004). For more information, contact Katy Börner at katy@indiana.edu

This material is based upon work supported by the National Science Foundation under Grants No. IS-0228261 and DUE-0338623.

Katy Börner, Knowledge Domain Visualizations in Support of Scholarly Knowledge and Expertise Management, SRI International, Oct 21, 2004.



IVC Learning Modules (<http://iv.slis.indiana.edu/lm>)



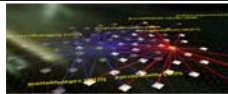
Learning Modules

Most information visualizations are highly interactive. While a number of excellent textbooks exist, the two-dimensional printouts on paper often cannot convey their true visual appearance and interactive performance. Several textbooks come with accompanying web sites that contain snapshots of user interfaces as well as animations and movies. However, none of them facilitates the exploration, application, evaluation, and comparison of algorithms.

This web page will provide access to a number of learning modules. Each learning module comes with an:

- ◆ Introduction
- ◆ Discussion of Existing Algorithms
- ◆ Learning Task
 - ◊ A challenging scenario to use the code in the [XML Toolkit](#) to visualize a data set to support a specific user group.
- ◆ Programming Exercise
 - ◊ An programming exercise plus an explanation of possible solutions.
- ◆ Opportunities & Challenges, and
- ◆ References to research papers, online demos, (commercial) applications)

Katy Börner, Knowledge Domain Visualizations in Support of Scholarly Knowledge and Expertise Management, SRI International, Oct 21, 2004.

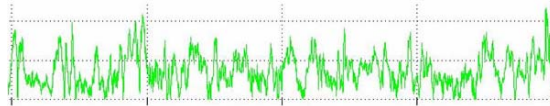


[Learning Modules](#) > Visualizing Time Series Data

[Description](#) | [Usage Hints](#) | [Learning Task](#) | [Discussion](#) | [References](#) | [Acknowledgments](#)

Description

A time series is a sequence of events/observations which are ordered in one dimension, e.g., time. Frequently, successive observations depend on each other and it makes sense to display them in a (time) sorted fashion, e.g., as a scatter plot. Alternatively, one could be interested to know how many observations of a certain value have been made. Here one would sort the observations by value, count the number of observations for each value and derive a histogram. Time series data can be continuous, i.e., there is an observation at every instant of time see figure below, or discrete, i.e., observations exist for regularly or irregularly spaced intervals.



Time series are recorded, analyzed and used in diverse domains of science. Check out the [Time Series Data Library](#) maintained by Rob Hyndman and Muhammad Akram for numerous data sets from Agriculture, Chemistry, Crime, Demography, Ecology, Finance, Health, Hydrology, Industry, Labour market, Macro-Economics, Meteorology, Micro-Economics, Physics, Production, Sales, Simulated series, Sport, Transport & Tourism or Utilities.

Katy Börner, Knowledge Domain Visualizations in Support of Scholarly Knowledge and Expertise Management, SRI International, Oct 21, 2004.

Time Series Analysis & Visualization

<http://iv.slis.indiana.edu/lm/lm-time-series.html>

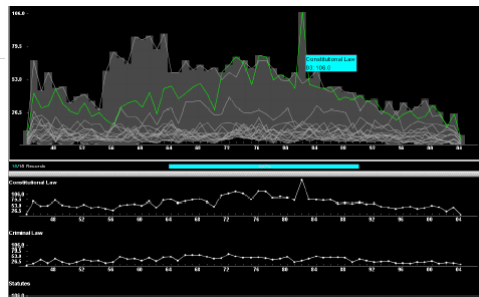
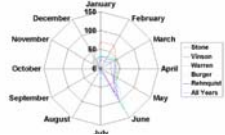
Visualizing the Work of the United States Supreme Court Based on Time Data and Top Level West Topics

by Peter A. Hook & Rongke Gao

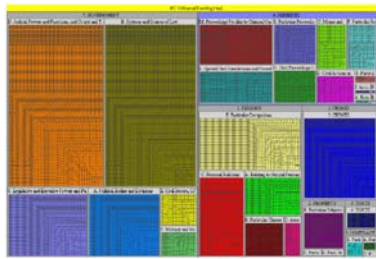


Stone Vinson Warren Burger Rehnquist

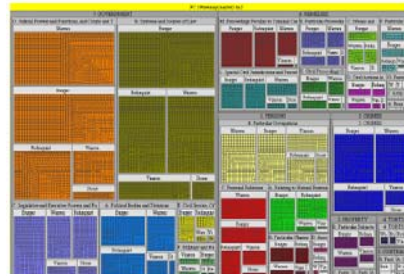
Topics Per Month Per Court



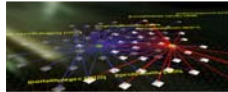
Top fifteen most occurring topics from 1944 to 2004 in Timesearcher



All topics grouped by West Category and Sub-Category grouped over the entire lengths of the data set



All topics by West Category and Sub-Category grouped corresponding to the five chief justices



4. Conclusions

Given access to appropriate databases knowledge domain analysis and visualization techniques can be employed to

- Understand the structure and evolution of scientific disciplines.
- Correlate research input (funding) and research output.
- Track and communicate the development of S&E fields at proper levels of detail.
- Identify major players (individuals or institutions).
- Understand the result of / the need for cross-disciplinary S&E.
- Correlate taxonomies with the data they describe.

Mapping Humanity's Knowledge and Expertise in the Digital Domain

At the **101st Annual Meeting** of the **Association of American Geographers** Denver, CO: April 5-9, 2005.

Session Organizers

Katy Börner, **Indiana University**
André Skupin, **University of New Orleans**

Sponsors

Cartography and GIS specialty groups

Description

This session will bring together leading researchers and practitioners that aim to develop techniques, tools, and infrastructures to map humanity's knowledge and expertise for the improvement of science and education.

Knowledge and expertise is typically extracted from digitally available literature, news, computer mediated communication data as well as from information about the producers and consumers of those data sets. Advanced data analysis techniques in combination with spatial metaphors, geographic principles, and cartographic methods are applied to organize, visualize, and communicate the semantic relationships inherent in the data.

The ultimate goal of this work might be an interactive cartographic map of all of science, with continents representing the major research areas such as, e.g., biology or physics, dots denoting major authors, PIs, papers or news, dynamically evolving research frontiers, blinking 'hot' papers and topics, etc. This map could be used to teach and understand the evolving structure of all of science, to identify major experts, to find and read the most relevant papers and news, to see the effects of resource allocation decisions, to study social networks, etc. Last but not least, it would provide a unique bird's eye view of major experts in specific areas and mankind's knowledge in general.

Some of the leading-edge research on this topic is found where geography intersects with information/library science, computer science, and cognitive science. We invite papers on the broad foundations, computational methods, software systems, and evaluation of such data analyses and visualizations, as they have emerged in this interdisciplinary endeavor.

<http://vm.indiana.edu/aag05>

