Mapping Science ~ History and Future

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

Several slides were taken from a talk by Kevin W. Boyack  
for the UCGIS Summer Meeting, June, 2009.

Visualization for Collective, Connective & Distributed Intelligence  
Dynamic Knowledge Networks ~ Synthetic Minds  
Stanford University, CA: August 12, 2009
Early Maps of the World  VERSUS  Early Maps of Science

3D
Physically-based
Accuracy is measurable
Trade-offs have more to do with granularity
2-D projections are very accurate at local levels
Centuries of experience
Geo-maps can be a template for other data

n-D
Abstract space
Accuracy is difficult
Trade-offs indirectly affect accuracy
2-D projections neglect a great deal of data
Decades of experience
Science maps can be a template for other data

Kevin W. Boyack, UCGIS Summer Meeting, June, 2009
Towards a World Map
Portolan chart of the central and western Mediterranean and part of the Atlantic - Bartolome Olives - 1559
In 1696, the first accurate map (shown below left) of the Earth was drawn by César-François Cassini de Thury based on 40 points (given in red) of accurate latitude and longitude. The north-south position (latitude) of any point on Earth could be determined via star paths. To measure the east-west position (longitude), exact time measurement was essential: one minute of uncertainty implied a 10-mile margin of error in location. Inspired by Galileo’s work, the mapmakers used the planet Jupiter as a “clock in the sky.” They carefully recorded the motions of Jupiter’s moons (see Cassini’s 1668 table of the eclipses of Jupiter’s moons below).
In 1744, Cassini’s team started to map France in a rigorous fashion using triangulation. In the late 1700s, the world’s first national land survey of France was completed. In 1870, Captain George Everest embarked to map India by triangulation. For generations, a vast network of repeating sightline triangles was meticulously measured and recorded (see map below). What resembles a pattern of eyelashes on the northern border represents the sightlines to stations built above treetops. While analyzing the triangles in the calculating offices of Calcutta, the mapmakers discovered the highest peak in the world: Mount Everest.
Towards a Map of all Sciences
2002 ‘Base Map’ of Science


- Uses combined SCI/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)
Towards a Consensus Map of Science

Milestones of Mapping Science

1934

2007

Zoom into one map and legend
John D. Bernal was a world renowned physicist, a historian of science, and a sociologist of science. He is considered to have produced one of the first ‘maps’ of science.
Ellingham’s “Relations Between the Branches of Natural Science and Technology” with an overlay of “Abstracts or Groups of Abstracts Covering A Very Wide Field” (1948)

Garfield, 1964

Historiograph of DNA Development

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

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

Garfield, Sher, & Torpie (1964). "The Use of Citation Data in Writing the History of Science." Air Force Office of Scientific Research under contract F49(638)-1256.
Using co-citation to create domain maps

Henry Small.
Head of research at ISI, now Thomson Reuters Scientific.


SPIRE, Themescape, 1995

Pacific Northwest Labs introduces a mapping tool based on text

2.5-D representation of intensity of "themes" using topography

Later spinoff of same technology used in patent analysis products

Sandia National Labs introduces an interactive browsing tool for exploring “maps”

Primarily for exploring citation-based maps, but ultimately used in science studies and genomics

Zoom, pan, query, etc. capabilities

“Knowledge mining with VxInsight: Discovery through interaction.”
Author Co-citation Map, White, 1998

120 highly-cited authors in Information Science

Layout using multi-dimensional scaling

Old, 2001

Utilizing spatial information systems for non-spatial data analysis

Authors in Information Science

Topography added

3D representation also

http://php.indiana.edu/~jold/SLIS/L710/L710.htm
Chen, 2001

Software: CiteSpace I

Four-step procedure for visualizing intellectual structures using co-cited documents

Newman, 2001

Physicist bringing new tools to the problem

2-generation co-authorship graph of Mark himself (center node)

Critical Path and Trajectories of Individuals
Kreutz & Druffel

Longitudinal Social Network Shapeshift
Moody, McFarland & Rosenblat-DaMalla

Intramedia Map of Data by Cuginno et al.
Talia

Eventmap View of 2006 Usenet Returns
Brin & Fisher

Cross-Map of Authors' Research
Moreno

Backbone of Science
Bresin, Khil & Binnewies

Profile Visualization API
Hern & Cards

3D Large Graph Exploration System
Asher

Author-Name Disambiguation
Tennert et al.

PubNet
Dougher et al.

CNI Researcher Link Viewer
Schlter et al.

QuteFly
Buwalash
All of Science: Boyack et al., 2003

Map of over 7,000 journals

Over 200 clusters of journals based on inter-citation statistics

Graph layout techniques to position clusters

Inaccurate locations: it appears there is a center to science, but it is an artifact of the graph layout

All of Science: Boyack et al., 2003

Map of over 7,000 journals

Over 200 clusters of journals based on inter-citation statistics

Graph layout techniques to position clusters

Inaccurate locations: it appears there is a center to science, but it is an artifact of the graph layout

Modified from Boyack et al. by Ian Aliman, Indiana University.
Map of over 7,000 journals

Over 200 clusters of journals based on inter-citation statistics

Graph layout techniques to position clusters

Inaccurate locations: it appears there is a center to science, but it is an artifact of the graph layout
Cambrian explosion ~ seemingly rapid appearance of most major groups of complex animals around 530 million years ago.
All of Science: Klavans, 2007

Map of over 16,000 journals and proceedings from ISI and Scopus databases (2001-2005 fileyears)

Two levels of clustering

Spherical layout of 554 clusters

Mercator projection of spherical layout

View from “south pole” is nearly circular

New work is built on existing work. Each of the examples below cites a series of works that developed in a progressive fashion, as one born from the other:

- Garfield’s original historiography of DNA research (1962); his long-term development of HistCite (first published in 2004); and his exhibit map (2006), which incorporates a re-rendering of the 1962 historiography and the application of HistCite.
- Tobler’s early works on the visualization of flow, his Flow Mapper tool (1987), and the tool’s application in geospatial and network journal data (2005).
- Batty et al.’s work on the geography of science (2003 and 2006).
- Boyack and Klavan’s work toward a base map of science followed by the creation of a series of maps (2005–2007).

Over time, former tools are subsumed by new tools, software APIs, and libraries. Examples include the Information Visualization Cyberinfrastructure (2003), Fekete’s The InfoVis Toolkit (2004), and the Network Workbench (2006). Mashups also emerge, such as Herr et al.’s Interactive Google Map of 2006 Society for Neuroscience Abstracts.
Authors are mortal. Papers are immortal.

Monsters = "the unknown" or voids.

Impact of funding on science.

Good and bad years.

Science as accumulation of knowledge.

"Scholarly brick laying".

Standing on the shoulders of giants.

Densely knit communities.

The importance of weak links.

Areas of science are tube shaped.

This drawing attempts to shows the “structure” of science.

Many are interested to understand the “dynamics” of science.
Introduction

E. O. Wilson writes in Consilience: The Unity of Knowledge (1998): "Features that distinguish science from pseudoscience are repeatability, economy, measurement, heuristics, and consilience." Please see Berners’s recent presentation at the A Deeper Look at the Visualization of Scientific Discovery NSF Workshop for a general introduction of the needs and the resources provided here.

Needs Analysis

As part of the "Towards a Macrostructure for Science Policy Decision Making" NSF SBE-0758111 award, interviews with science policy makers are conducted to identify what science of science’ research results and tools might be most desirable and effective. So far, 80 formal, one-hour interviews have been conducted with science policy makers at university campus level, program officer level, and division director level for governmental, state, and private foundations. Data compilation will start in October 2008 and resulting report can be ordered by sending a request to Mark Price (markprice@indiana.edu).

Conceptualization of Science

A science of science requires a theoretically grounded and practically useful conceptualization of the structure and evolution of science. A special journal issue entitled "Science of Science: Conceptualizations and Models of Science" edited by Katz Börner, Indiana University & Andrea Schambor.. Royal Netherlands Academy of Arts and Sciences invites contributions on this topic. It will be published in the Journal of Informetrics 3(1) in January 2009.

Scholarly Database

The Scholarly Database (SDB) at Indiana University aims to serve researchers and practitioners interested in the analysis, modeling, and visualization of large-scale scholarly datasets. The database currently provides access to over 20 million papers, patents and grants. Resulting datasets can be downloaded in bulk. Register for free access at https://sdb.slis.indiana.edu/.

Cyberinfrastructures

The SciNetnetics filling of the Network Workbench (NWB) Tool provides a unique distributed, shared resources environment for large-scale network analysis, modeling, and visualization. Thomson Scientific/ISI, Scopus and Google Scholar data, EndNote and BibTex files, or NSF awards can be read and diverse networks can be extracted and studied. Download User Manual with focus on Cyberinfrastructures.

http://sci.slis.indiana.edu
This is the only mockup in this slide show.

Everything else is available today.
Papers, maps, cyberinfrastructures, talks, press are linked from http://cns.slis.indiana.edu