

Understanding (Big) Data by Using Macroscopes

Katy Börner

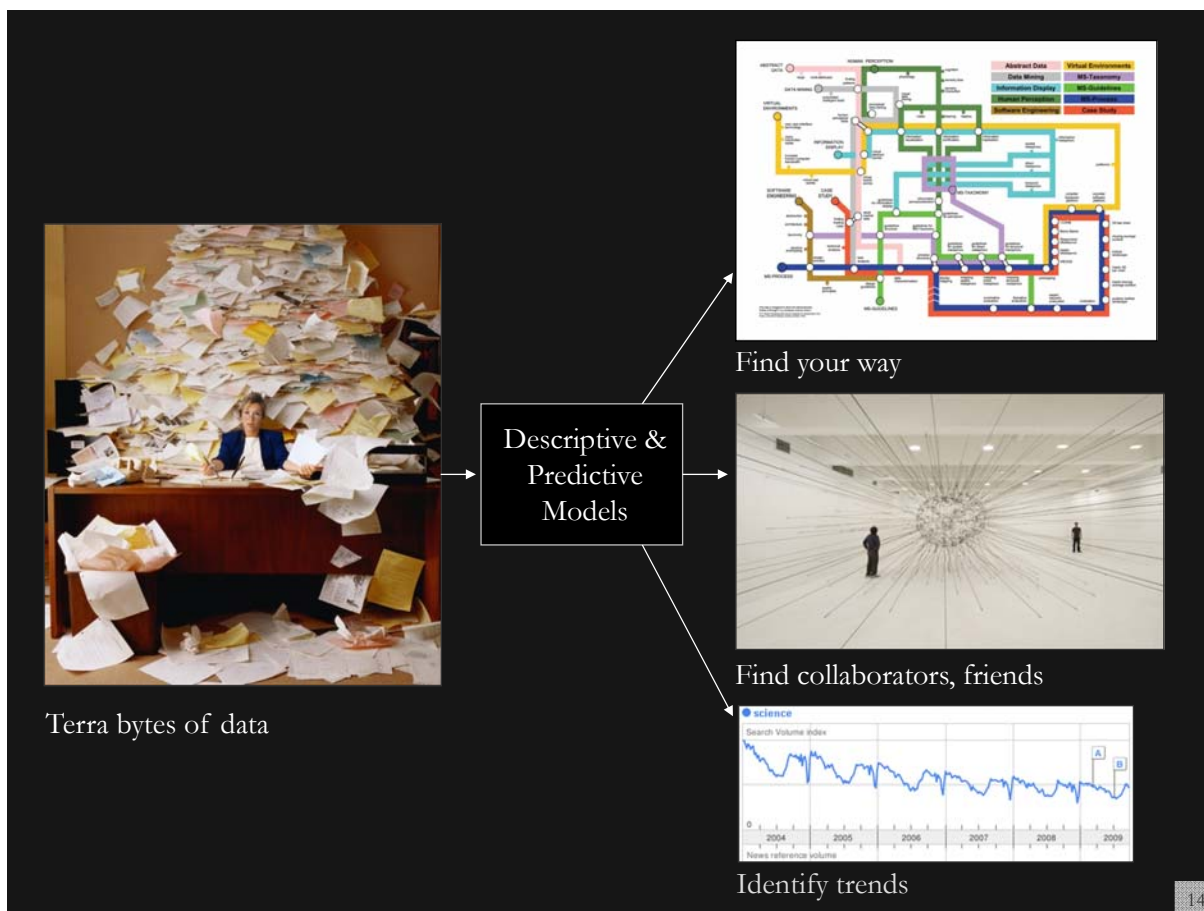
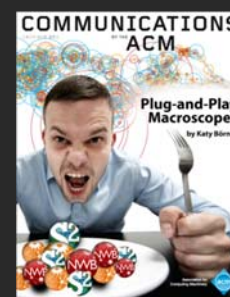
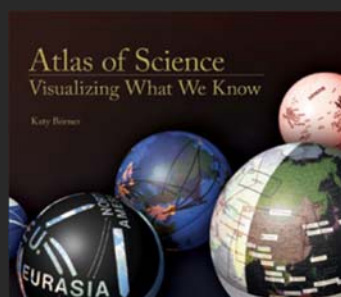
Cyberinfrastructure for Network Science Center, Director
Information Visualization Laboratory, Director
Department of Information and Library Science
School of Informatics and Computing
Indiana University, Bloomington, IN
katy@indiana.edu



With special thanks to the members at the Cyberinfrastructure for Network Science Center; the Sci2, NWB teams, and the IVMOOC students

Digital HPS Meeting, Bloomington, IN

September 7, 2013



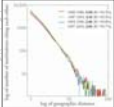
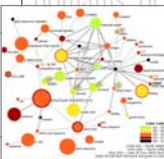



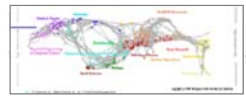
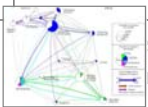
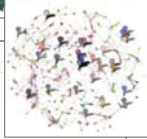

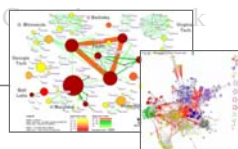
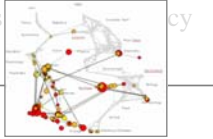
Type of Analysis vs. Level of Analysis

	Micro/Individual (1-100 records)	Meso/Local (101-10,000 records)	Macro/Global (10,000 < records)
Statistical Analysis/Profiling	Individual person and their expertise profiles	Larger labs, centers, universities, research domains, or states	All of NSF, all of USA, all of science.
Temporal Analysis (When)	Funding portfolio of one individual	Mapping topic bursts in 20-years of PNAS	113 Years of Physics Research
Geospatial Analysis (Where)	Career trajectory of one individual	Mapping a states intellectual landscape	PNAS publications
Topical Analysis (What)	Base knowledge from which one grant draws.	Knowledge flows in Chemistry research	VxOrd/Topic maps of NIH funding
Network Analysis (With Whom?)	NSF Co-PI network of one individual	Co-author network	NIH's core competency



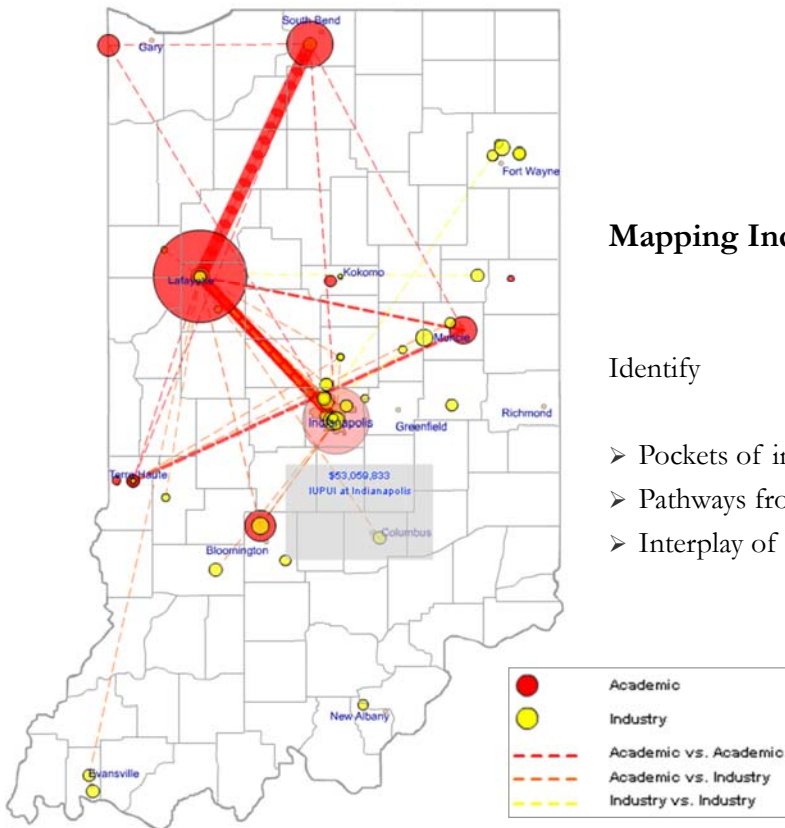
15

Type of Analysis vs. Level of Analysis

	Micro/Individual (1-100 records)	Meso/Local (101-10,000 records)	Macro/Global (10,000 < records)
Statistical Analysis/Profiling	Individual person and their expertise profiles	Larger labs, centers, universities, research domains, or states	All of NSF, all of USA, all of science. 
Temporal Analysis (When)	Funding portfolio of one individual	Mapping topic bursts in 20-years of PNAS 	113 Years of Physics Research 
Geospatial Analysis (Where)	Career trajectory of one individual	Mapping a states intellectual landscape 	PNAS publications 
Topical Analysis (What)	Base knowledge from which one grant draws. 	Knowledge flows in Chemistry research 	VxOrd/Topic maps of NIH funding 
Network Analysis (With Whom?)	NSF Co-PI network of one individual 	Co-author network 	NIH's core competency 



16



Mapping Indiana's Intellectual Space

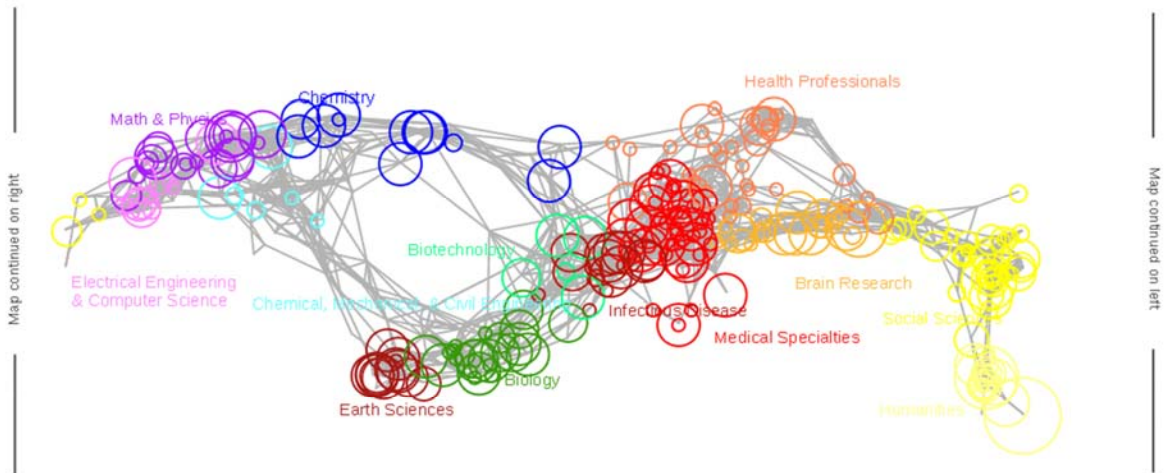
Identify

- Pockets of innovation
- Pathways from ideas to products
- Interplay of industry and academia

17

Mapping the Intersection of Science & Philosophy

Murdock, Jaimie, Robert Light, Colin Allen, and Katy Börner. Joint Conference on Digital Libraries (2013)



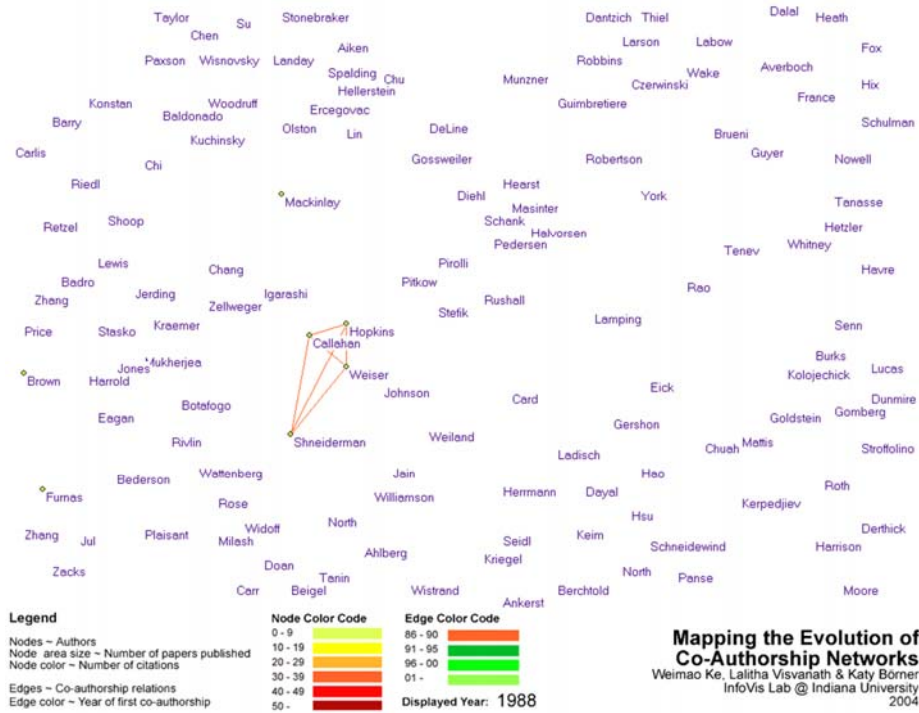
©2008 The Regents of the University of California and SciTech Strategies.
Map updated by SciTech Strategies, OST, and CNS in 2011.

Philosophical content on the UCSD Map of Science. The size of each circle corresponds to the number of SEP editorial areas citing material from the UCSD Map of Science subdiscipline (minimum: 0, maximum: 43). Circle color denotes 13 major disciplines of science.

18

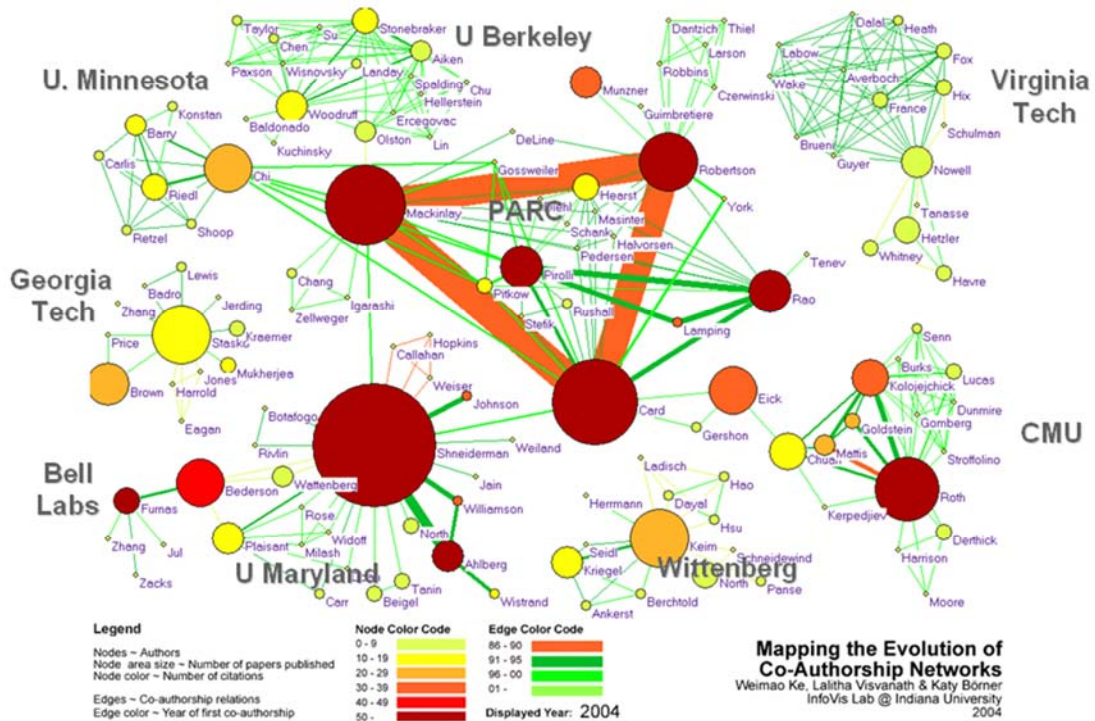
Mapping the Evolution of Co-Authorship Networks

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



Mapping the Evolution of Co-Authorship Networks

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

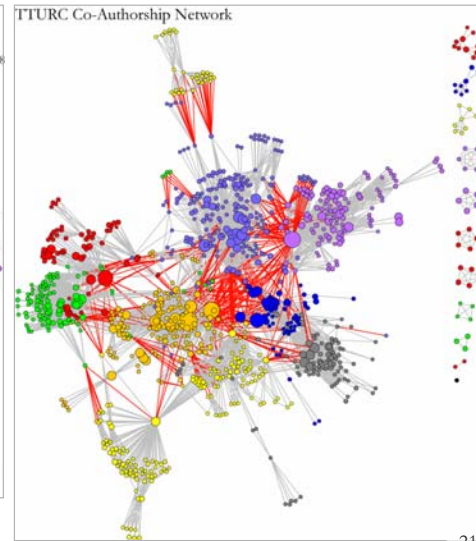
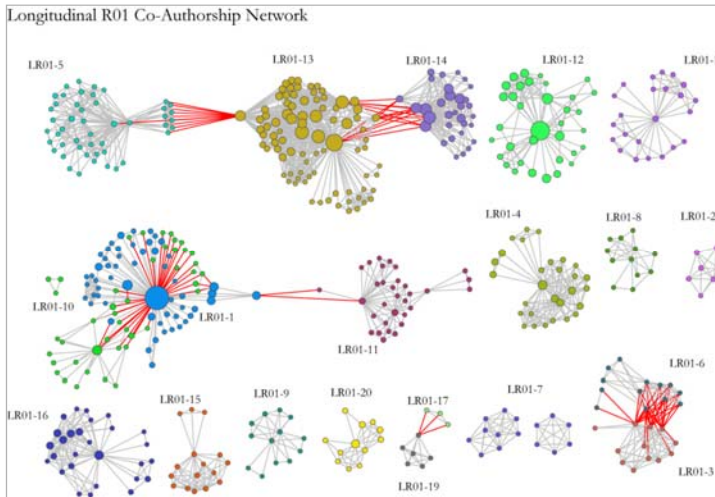
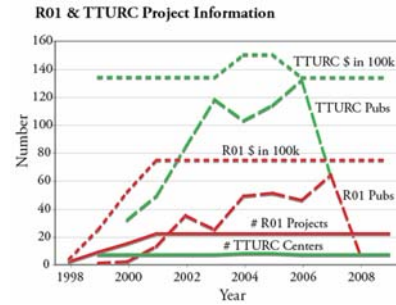


Mapping Transdisciplinary Tobacco Use Research Centers Publications

Compare R01 investigator based funding with TTURC Center awards in terms of number of publications and evolving co-author networks.

Zoss & Börner, *forthcoming*.

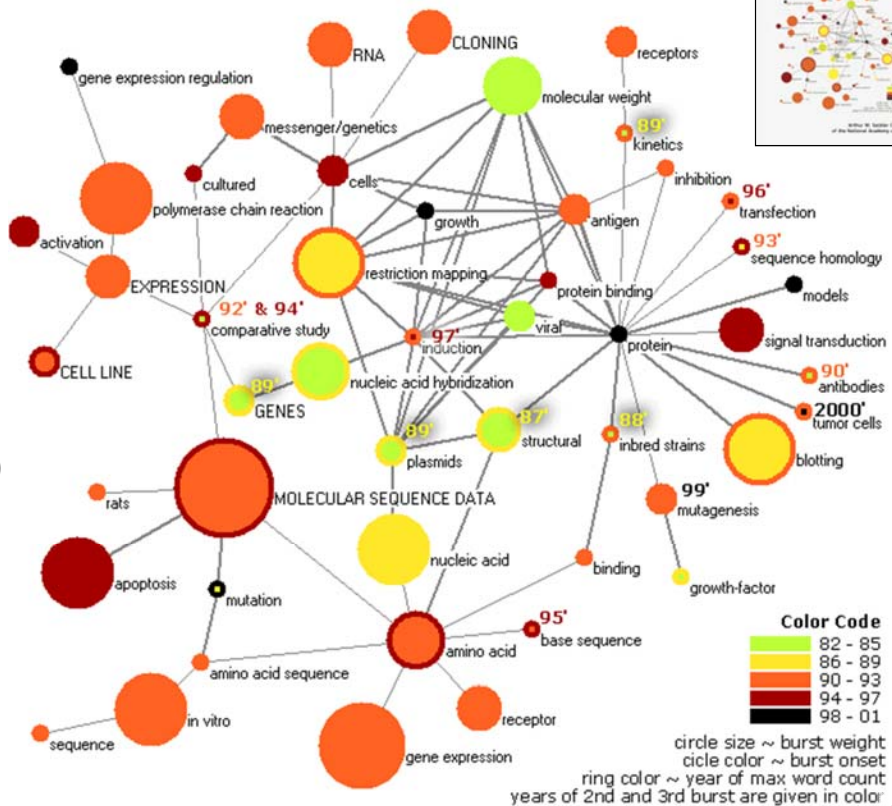
Supported by NIH/NCI Contract HHSN261200800812



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.

Mane & Börner. (2004) PNAS, 101(Suppl. 1): 5287-5290.



Geographic Map: Where Science Gets Done

Science Map: How Scientific Disciplines Relate

About

This Illuminated Diagram display adds the flexibility of an interactive program to the incredibly high data density of a print. This technique is generally useful when there is too much pertinent data to be displayed on a screen but the data is relatively stable. The computer can direct the eye to what's important by using projectors or screens as smart spotlights, animating the research impact of individuals, giving a "grand tour" of science, or highlighting query results (as when you touch the lectern or use the keyboard) with an overlay of moving light.

<http://scimaps.org>

Top Five Continents

- North America - 4,000 records
- South & East Asia - 3,589
- Australia - 2,431
- Africa - 2,208
- South America - 1,562

Top Five Scientific Disciplines

- Math & Physics - 4,000 records
- Health Professionals - 3,589
- Social Sciences - 2,431
- Aeronautical, Chemical, Mechanical & Civil Engineering - 2,208
- Humanities - 1,562

Search

The keyboard supports retrieval and display of papers based on their Medical Subject Headings (MeSH) and MeSH qualifier terms. If multiple terms are entered in a field, they are automatically combined using "OR". So, "breast cancer" matches any record with "breast" or "cancer" in that field. You can put AND between terms to combine with "AND". Thus "breast AND cancer" would only match records that contain both terms. Double quotation can be used to match compound terms, e.g., "breast cancer" retrieves records with the phrase "breast cancer", and not records where "breast" and "cancer" are both present, but the exact phrase.

Q

W

E

R

T

Y

U

I

O

P

A

S

D

F

G

H

J

K

L

"

Z

X

C

V

B

N

M

⌫

Space

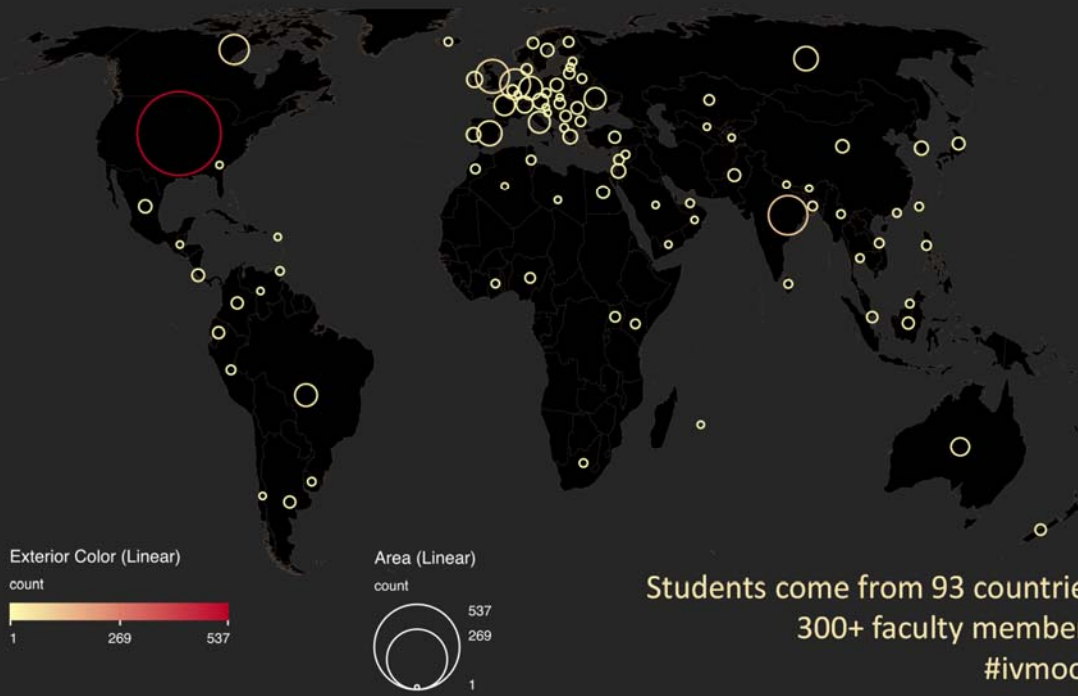
Go

People & Topics

27

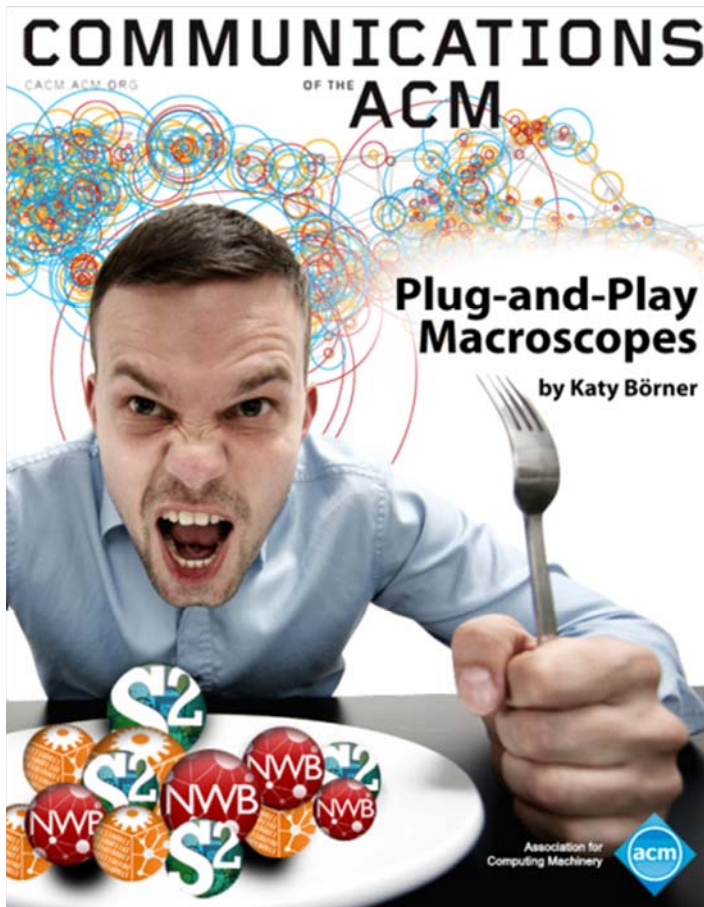
Learning how to use Macroscopic tools by taking the Information Visualization MOOC

The Information Visualization MOOC
ivmooc.cns.iu.edu



Students come from 93 countries
300+ faculty members
#ivmooc

29



Börner, Katy. (2011).
Plug-and-Play Macroscopes.
Communications of the ACM,
54(3), 60-69.

Video and paper are at
<http://www.scivee.tv/node/27704>

Forthcoming Book:

The Historian's Macroscope by
Shawn Graham, Ian Milligan,
& **Scott Weingart**, Imperial
College Press, London,
2014/2015.

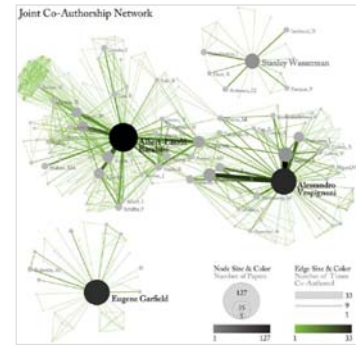
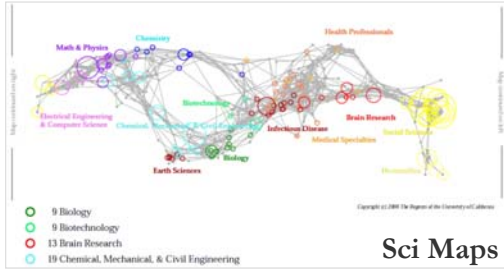
<http://themascope.org>

30



Sci² Tool answers When, Where, What, and With Whom questions

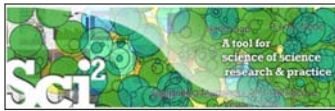
Using temporal, geospatial, topical, and network analysis and visualizations plugins.



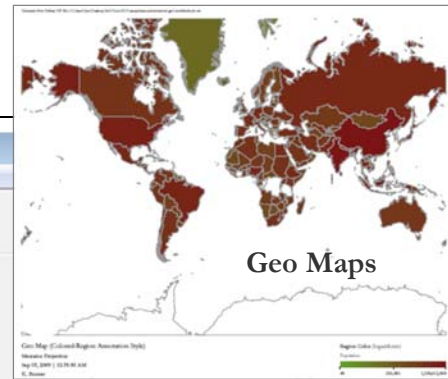
Horizontal Time Graphs



Börner, Katy, Huang, Weixia (Bonnie), Linnemeier, Micah, Dubon, Russell Jackson, Phillips, Patrick, Ma, Nianli, Zoss, Angela, Guo, Hanning & Price, Mark. (2009). *ReTe-Netzwerk-Red: Analyzing and Visualizing Scholarly Networks Using the Scholarly Database and the Network Workbench Tool*. *Proceedings of ISSI 2009: 12th International Conference on Scientometrics and Informetrics, Rio de Janeiro, Brazil, July 14-17*. Vol. 2, pp. 619-630.



Sci² Tool Vis cont.



Sci² Tool

File Preprocessing Modeling Analysis Visualization Scientometrics Help

Console

Welcome to the Science of Science Tool (Sci²). The development of this tool is supported in Network Science center and the School of Li Indiana University, the National Science Foundation and IIS-0715303, and the James S. McDonnell Cyberinfrastructure portal (<http://sci.slis.indiana.edu>).

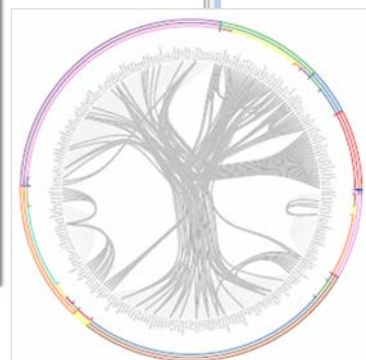
The primary investigators are Katy Börner, In SciTech Strategies Inc. The Sci² tool was developed by J. Duhon, Patrick A. Phillips, Chintan Tank, a Cyberinfrastructure Shell (<http://cishell.org>) for Network Science Center (<http://cns.slis.indiana.edu>). Many algorithm plugins were derived from the Network Workbench Tool (<http://nwb.slis.indiana.edu>).

Please cite as follows:
Sci² Team. (2009). Science of Science Tool. In SciTech Strategies Inc., <http://sci.slis.indiana.edu>.

Scheduler

Remove From List Remove completed

!	Algorithm Name	Date	Time	% Con
<input checked="" type="checkbox"/>	Extract Co-Author Network	09/03/2009	00:15:20 AM	100%
<input checked="" type="checkbox"/>	Load and Clean ISI File	09/03/2009	00:15:05 AM	100%



Visualizing Isis: A Geospatial and Topical Analysis of the *History of Science*

David E. Hubbard (Texas A&M University), Anouk Lang (University of Strathclyde), Kathleen Reed (Vancouver Island University), Anelise Hanson ShROUT (Davidson College) and Lyndsay D. Troyer (Colorado State University)

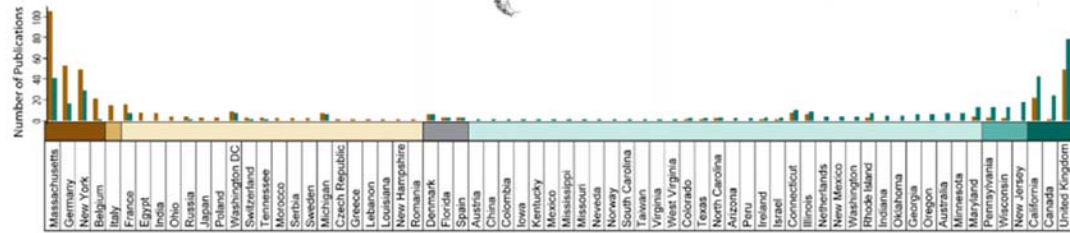
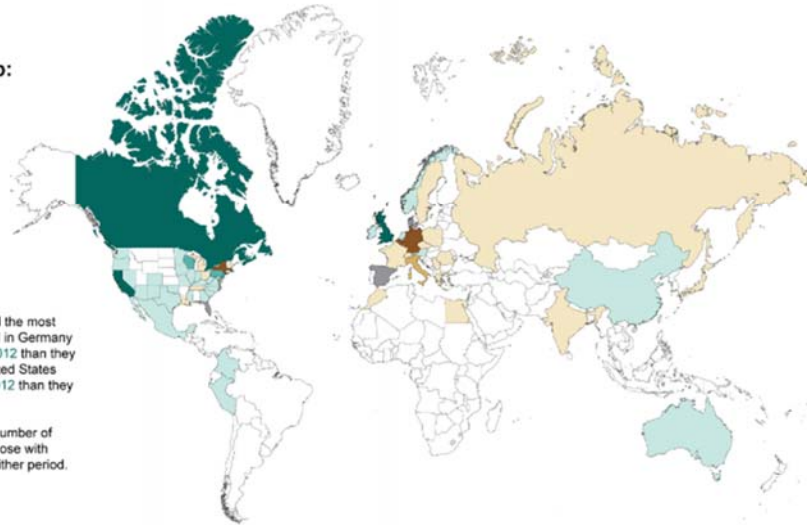
Changes in *Isis* Authorship: 1913-1937 to 1988-2012.

This map and chart show geographic shifts in *Isis* authors' locations across two time periods of 25 years each: 1913-1937 and 1988-2012.

Countries and states tinted **brown** had a greater number of articles published in the 1913-1937 period, while those tinted **green** had a greater number published in the 1988-2012 period.

Germany and the United States experienced the most extreme shifts in authorship. Authors located in Germany published 37 fewer articles between 1988-2012 than they did in 1913-1937. Authors located in the United States published 31 more articles between 1988-2012 than they did in 1913-1937.

Countries and states in **gray** had the same number of published articles in both periods, while those with no color (i.e., white) had no publications in either period.



Visualizing Isis: A Geospatial and Topical Analysis of the *History of Science*

David E. Hubbard (Texas A&M University), Anouk Lang (University of Strathclyde), Kathleen Reed (Vancouver Island University), Anelise Hanson ShROUT (Davidson College) and Lyndsay D. Troyer (Colorado State University)

Thematic trends over time

In *Isis*'s early years, the journal showed considerable interest in non-western cultures. The appearance of *Arab* and *Egyptian* in the 1920s reflect Sarton's interests in that decade. In the 1970s, *Isis*'s interest shifted focus to the West. German emerged as scholars began to pay attention to 19th-century German universities, which supported *Wissenschaft* (the study of science that involves systematic research and teaching), and from which grew many of the American research universities that became the primary locus for American scientific research for many years.

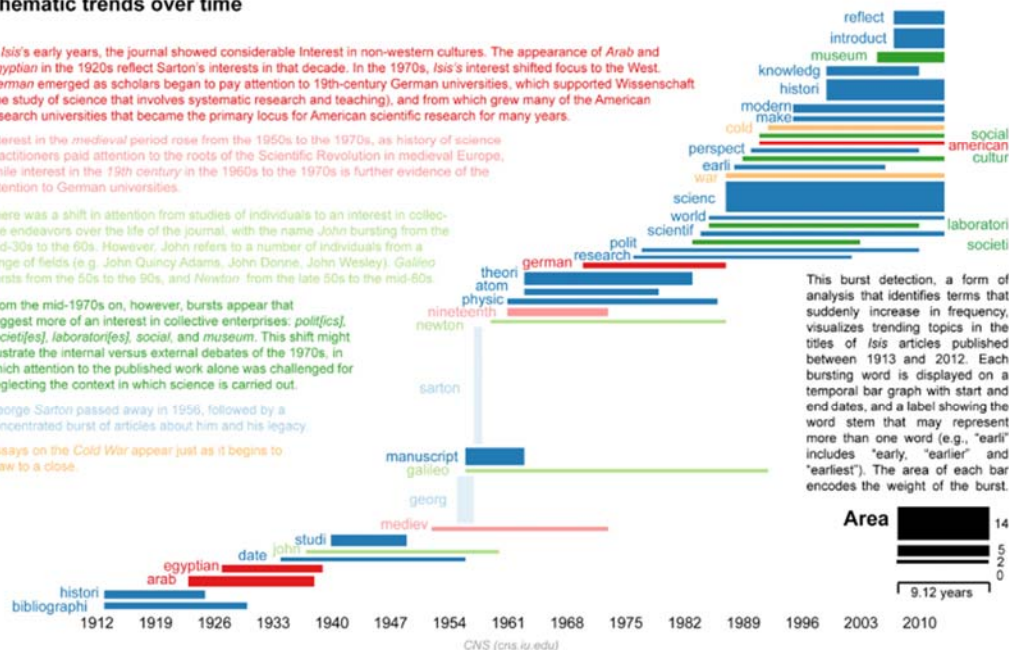
Interest in the medieval period rose from the 1950s to the 1970s, as history of science practitioners paid attention to the roots of the Scientific Revolution in medieval Europe, while interest in the 19th century in the 1960s to the 1970s is further evidence of the attention to German universities.

There was a shift in attention from studies of individuals to an interest in collective endeavors over the life of the journal, with the name *John* bursting from the mid-30s to the 60s. However, *John* refers to a number of individuals from a range of fields (e.g. John Quincy Adams, John Donne, John Wesley). Galileo bursts from the 50s to the 90s, and *Newton* from the late 50s to the mid-60s.

From the mid-1970s on, however, bursts appear that suggest more of an interest in collective enterprises: *politics*, *societ[ies]*, *laborator[ies]*, *social*, and *museum*. This shift might illustrate the internal versus external debates of the 1970s, in which attention to the published work alone was challenged for neglecting the context in which science is carried out.

George Sarton passed away in 1956, followed by a concentrated burst of articles about him and his legacy.

Essays on the *Cold War* appear just as it begins to draw to a close.



This burst detection, a form of analysis that identifies terms that suddenly increase in frequency, visualizes trending topics in the titles of *Isis* articles published between 1913 and 2012. Each bursting word is displayed on a temporal bar graph with start and end dates, and a label showing the word stem that may represent more than one word (e.g., "earli" includes "early," "earlier" and "earliest"). The area of each bar encodes the weight of the burst.

Mapping NEH awards and MEDLINE publications, 1980-2009

NEH Grants:

41,258 grants of 47,197 started between 1980 and 2009, encompassing 3.21 billion of the 3.77 billion dollars awarded.

Geo-coding by zipcode:

36,512 of 41,258 grants encompassing 3.13 billion of a potential 3.21 billion were geocoded to 3,510 distinct locations.

Science-coding by topic:

37,132 of 41,258 grants encompassing 2.09 billion of a potential 3.21 billion were mapped to 42 distinct subdisciplines.



MEDLINE publications:

12.95 million papers were published between 1980 and 2009.

Geo-coding:

Not possible with the data we have.

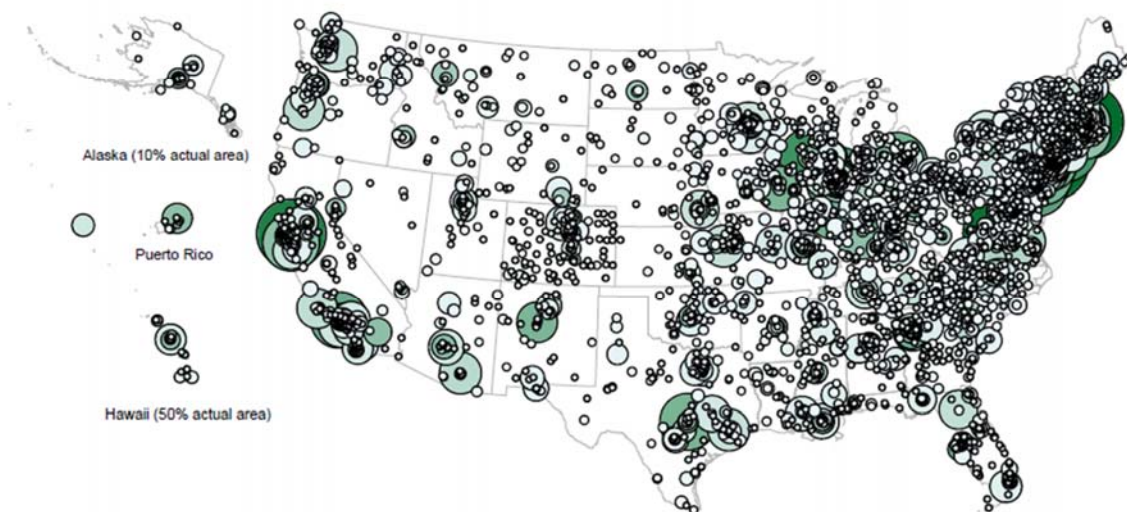
Science-coding by journal:

11.62 million of 12.95 million papers were science located (89.7%). Science located 5,941 out of 14,561 journals (40.8%) to 415 distinct subdisciplines.

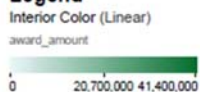


Geospatial Visualization (Proportional Symbol Map)

NEH Grants by ZIP Code (1980-2009)
Apr 09, 2013 | 01:09:37 PM EDT



Legend



Area (Linear)
count



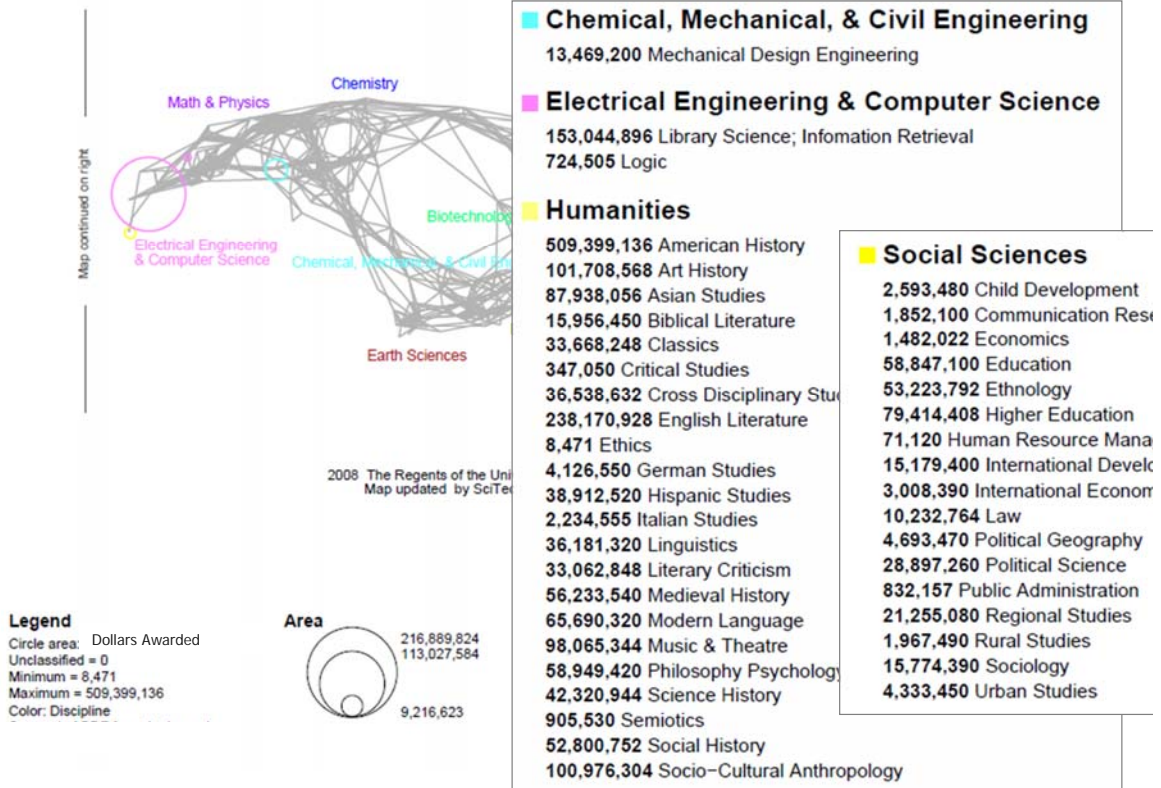
CNS (cns.iu.edu)

How to Read this Map

This *proportional symbol map* shows 52 U.S. states and other jurisdictions using the Albers equal-area conic projection with Alaska, Puerto Rico, and Hawaii inset. Each dataset record is represented by a circle centered at its geolocation. The area, interior color, and exterior color of each circle may represent numeric attribute values. Minimum and maximum data values are given in the legend.

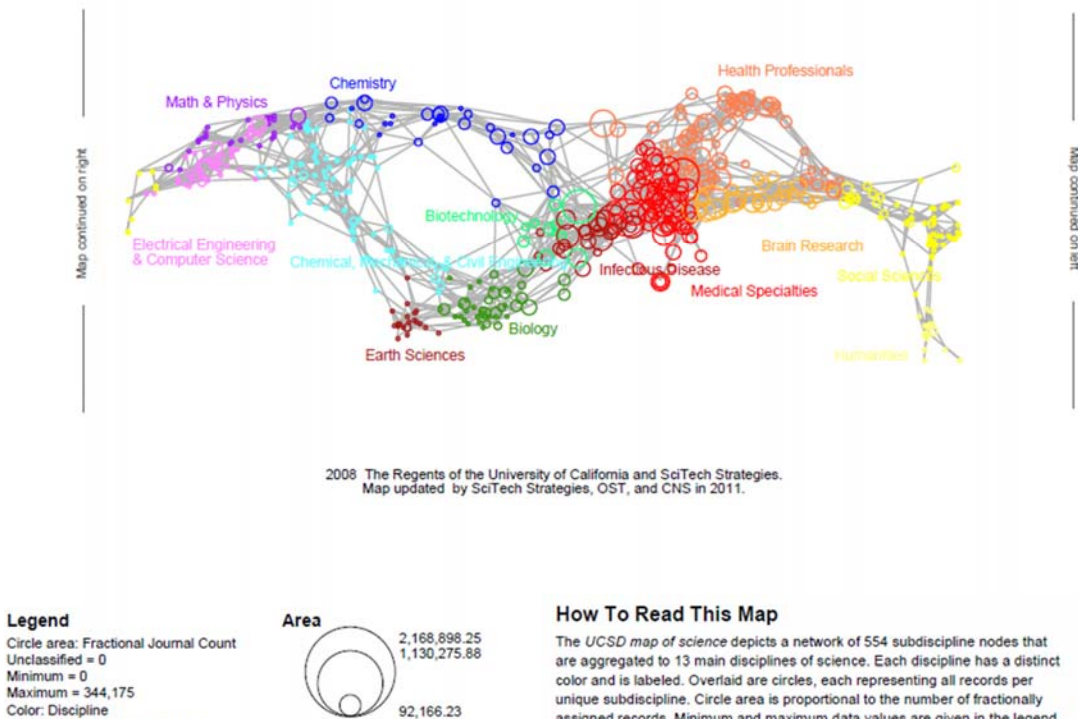
Topical Visualization: UCSD Map of Science

NEH grants, mapped based on categories



Topical Visualization: UCSD Map of Science

Medline papers, mapped based on journal names



References

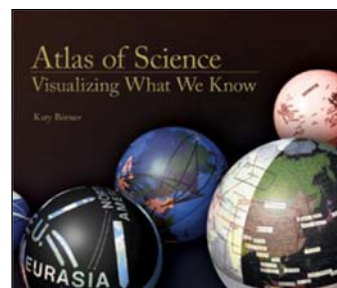
Börner, Katy, Chen, Chaomei, and Boyack, Kevin. (2003). **Visualizing Knowledge Domains**. In Blaise Cronin (Ed.), *ARIST*, Medford, NJ: Information Today, Volume 37, Chapter 5, pp. 179-255.
<http://ivl.slis.indiana.edu/km/pub/2003-borner-arist.pdf>

Shiffrin, Richard M. and Börner, Katy (Eds.) (2004). **Mapping Knowledge Domains**. *Proceedings of the National Academy of Sciences of the United States of America*, 101(Suppl_1).
http://www.pnas.org/content/vol101/suppl_1/

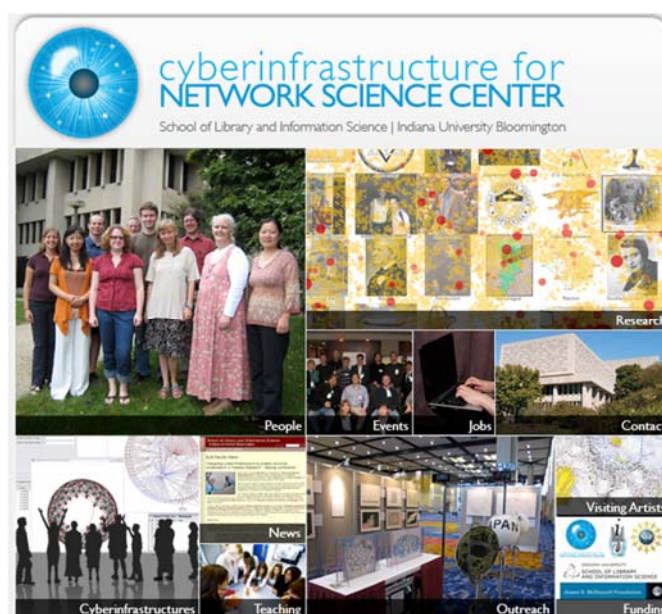
Börner, Katy, Sanyal, Soma and Vespignani, Alessandro (2007). **Network Science**. In Blaise Cronin (Ed.), *ARIST*, Information Today, Inc., Volume 41, Chapter 12, pp. 537-607.
<http://ivl.slis.indiana.edu/km/pub/2007-borner-arist.pdf>

Börner, Katy (2010) **Atlas of Science**. MIT Press.
<http://scimaps.org/atlas>

Scharnhorst, Andrea, Börner, Katy, van den Besselaar, Peter (2012) **Models of Science Dynamics**. Springer Verlag.



39



All papers, maps, tools, talks, press are linked from <http://cns.iu.edu>

CNS Facebook: <http://www.facebook.com/cnscenter>

Mapping Science Exhibit Facebook: <http://www.facebook.com/mappingscience>

40