

A complex network visualization with nodes and edges in various colors (red, blue, green, purple) on a black background. The nodes are scattered across the frame, with some forming dense clusters and others being isolated. The edges connect these nodes, creating a web-like structure.

Data Visualization Literacy

Katy Börner

Victor H. Yngve Professor of Information Science
Director, Cyberinfrastructure for Network Science Center
School of Informatics and Computing, Indiana University, USA

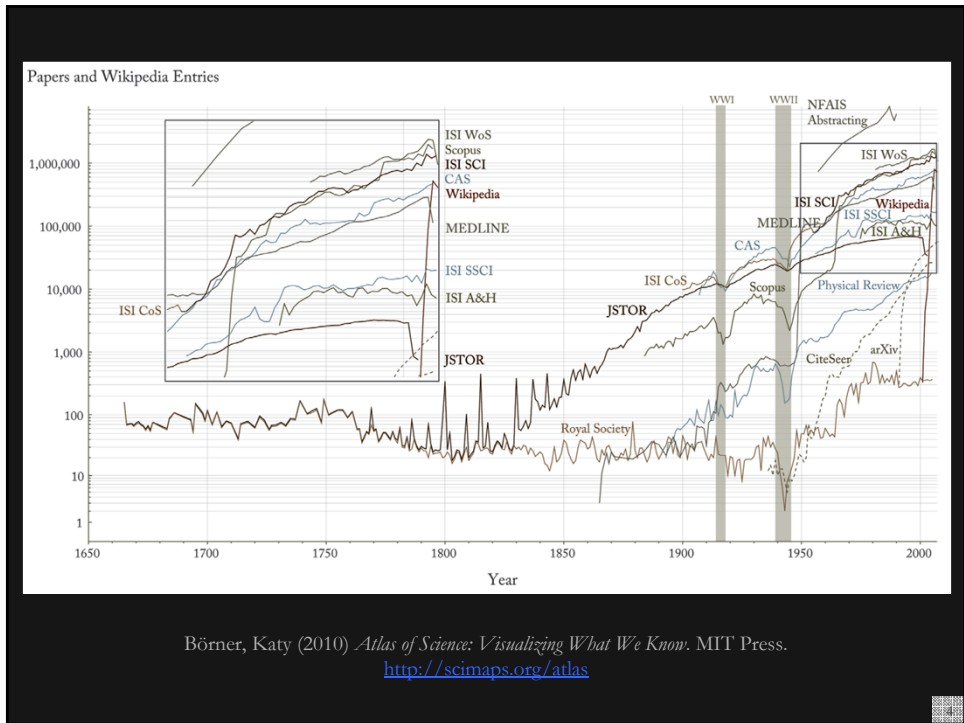
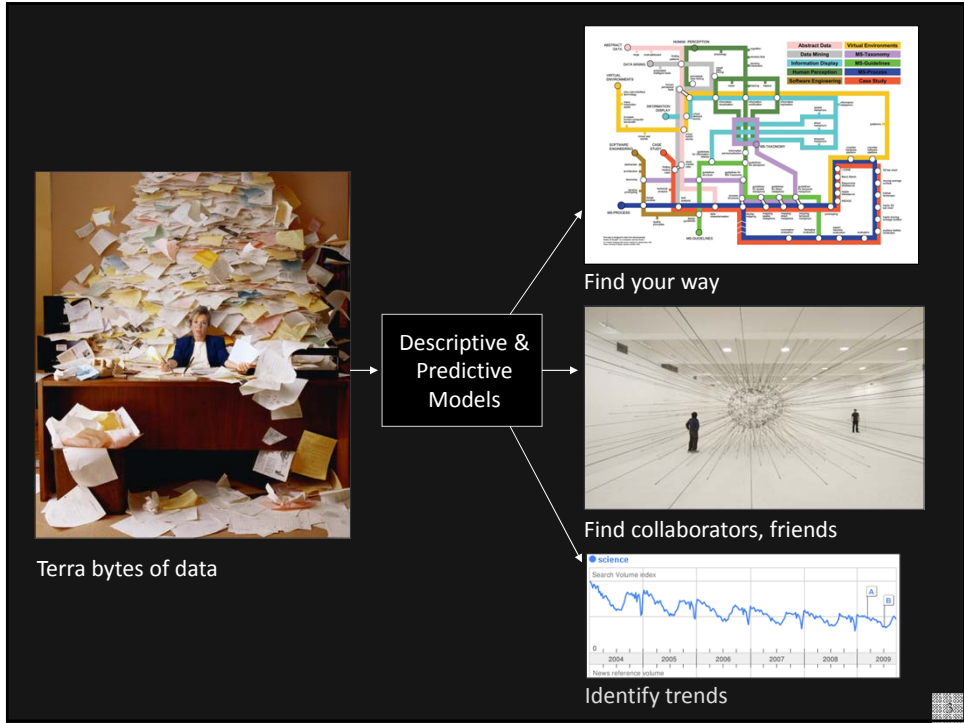
Science & Creativity Annual Conference (KOFAC)
Seoul, Korea

December 5, 2014

Language Communities of Twitter - Eric Fischer - 2012

***Navigating, Managing, and Utilizing
Data, Information, Knowledge, and Expertise
in the Information Age***





*Using Visualizations to Convert BIG Data into Insights
&
to Communicate Insights Gained to Others*



Different Levels of Abstraction/Analysis

Macro/Global
Population Level



Meso/Local
Group Level



Micro
Individual Level

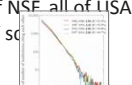





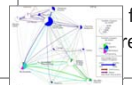
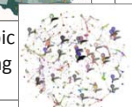
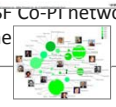
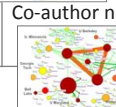
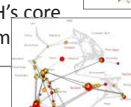


Type of Analysis vs. Level of Analysis

	Micro/Individual (1-100 records)	Meso/Local (101-100,000 records)	Macro/Global (100,000 < 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

9

Type of Analysis vs. Level of Analysis

	Micro/Individual (1-100 records)	Meso/Local (101-100,000 records)	Macro/Global (100,000 < 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 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 

10

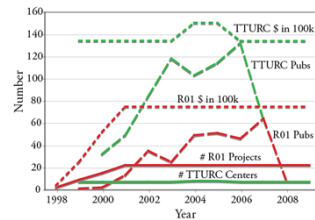
Mapping Transdisciplinary Tobacco Use Research Centers Publications

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

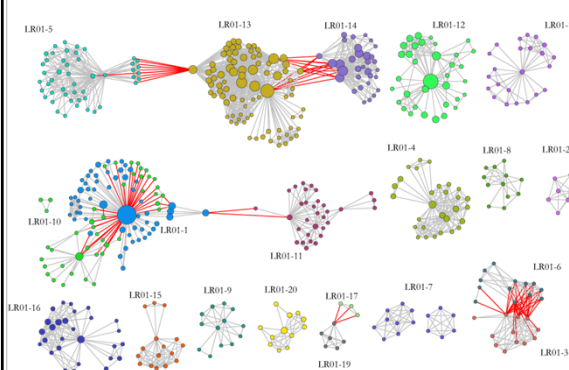
Stipelman, Hall, Zoss, Okamoto, Stokols & Börner, 2014

Supported by NIH/NCI Contract HHSN261200800812

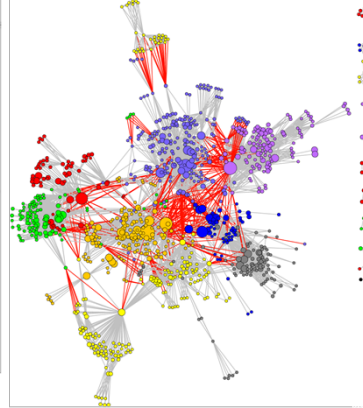
R01 & TTURC Project Information



Longitudinal R01 Co-Authorship Network



TTURC Co-Authorship Network



11

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

Börner, Penumathy, Meiss & Ke (2006) Mapping the Diffusion of Scholarly Knowledge Among Major U.S. Research Institutions. Scientometrics. 68(3), pp. 415-426.

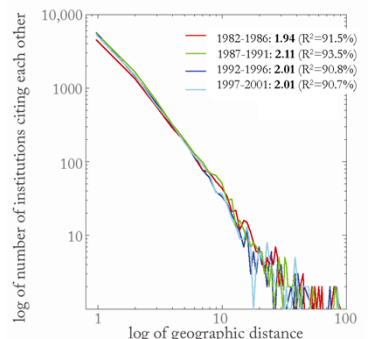
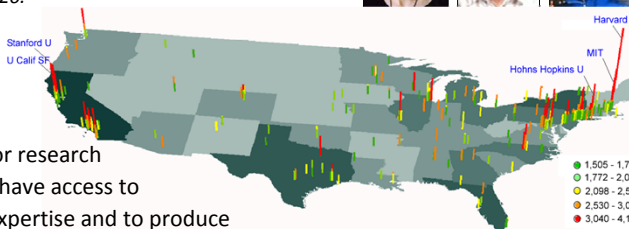


Research questions:

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 institutions?

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.



12

The Global 'Scientific Food Web'

Mazloumian, Amin, Dirk Helbing, Sergi Lozano, Robert Light, and Katy Börner. 2013. "Global Multi-Level Analysis of the 'Scientific Food Web'". *Scientific Reports* 3, 1167.

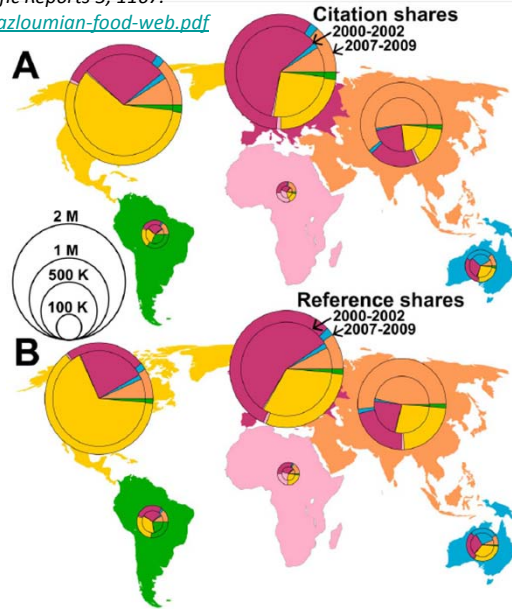
<http://cns.iu.edu/docs/publications/2013-mazloumian-food-web.pdf>

Contributions:

Comprehensive global analysis of scholarly knowledge production and diffusion on the level of continents, countries, and cities.

Quantifying knowledge flows between 2000 and 2009, we identify global sources and sinks of knowledge production. Our knowledge flow index reveals, where ideas are born and consumed, thereby defining a global 'scientific food web'.

While Asia is quickly catching up in terms of publications and citation rates, we find that its dependence on knowledge consumption has further increased.



13

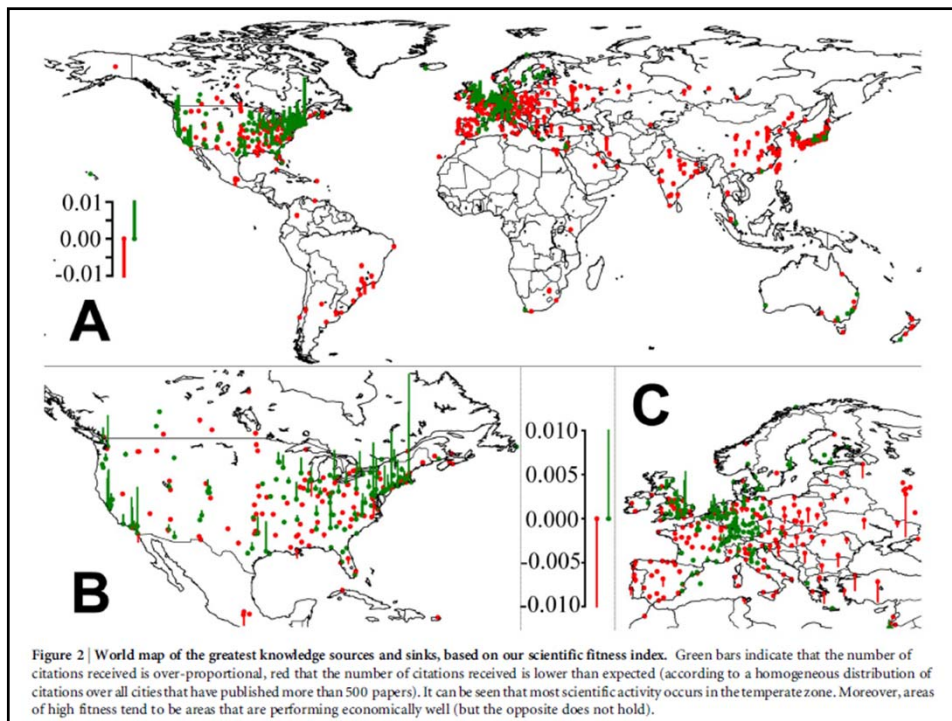


Figure 2 | World map of the greatest knowledge sources and sinks, based on our scientific fitness index. Green bars indicate that the number of citations received is over-proportional, red that the number of citations received is lower than expected (according to a homogeneous distribution of citations over all cities that have published more than 500 papers). It can be seen that most scientific activity occurs in the temperate zone. Moreover, areas of high fitness tend to be areas that are performing economically well (but the opposite does not hold).

Type of Analysis vs. Level of Analysis

	<i>Micro/Individual (1-100 records)</i>	<i>Meso/Local (101-100,000 records)</i>	<i>Macro/Global (100,000 < records)</i>
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 sci
Temporal Analysis (When?)	Funding portfolio of one individual	Topic bursts in PNAS	113 Years of Physics Research
Geospatial Analysis (Where?)	Career trajectory of one individual	Intellectual trajectory	PNAS publications
Topical Analysis (What?)	Base knowledge from research	Research	VxOrd/Topic NIH funding
Network Analysis (With Whom?)	NSF Co-PI network of one	Co-author network	NIH's core com

15

Data Visualization Literacy Study

Can Science Museum Visitors Read Data Visualizations?

More than 1000 youth and adult visitors across six U.S. science museums.

Visitors saw 5 out of 20 visualizations and were asked to answer

- Does this type of data presentation look at all familiar?
- Where might you have seen images like this?
- How do you think you read this type of data presentation?
- What would you call this type of data presentation?

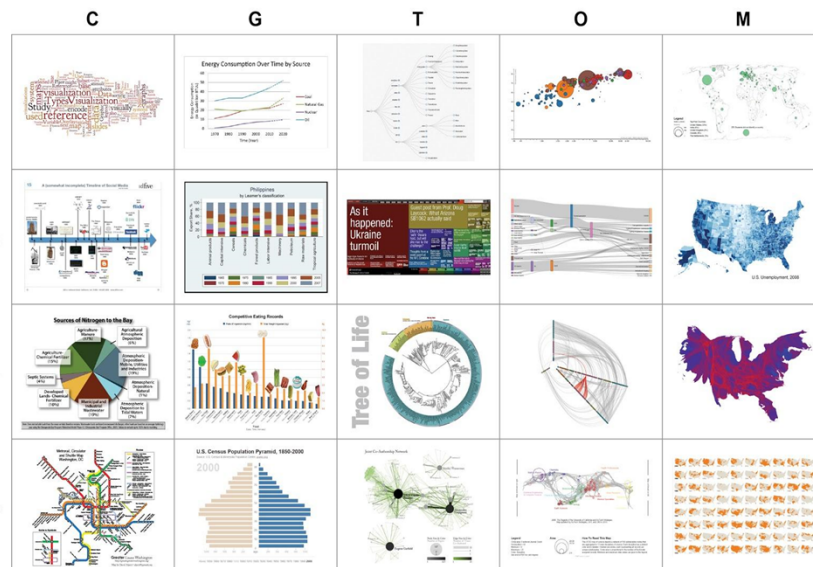
Results show that:

- a very high proportion of the population, both adult and youth, cannot interpret data visualizations beyond very basic reference systems;
- construction of complex visualizations led to more accurate interpretation than deconstruction; and
- individuals are willing to spend time attempting to make meaning in representations depending on their personal interest in the topic.

Joint work with Adam V. Maltese, Russell N. Balliet, Joe Heimlich and COSI, SMM, NYSCI, WonderLab, and the Marian Koshland Science Museum.

17

Can Science Museum Visitors Read These 20 Visualizations?



18

Names Science Museum Visitors Give These Visualizations

	<i>Visualization Type</i>	<i>Technical Label</i>	<i>Equivalent Phrase</i>	<i>Related Phrases</i>	<i>Unrelated</i>	<i>Not Applicable</i>	<i>Doesn't Know</i>
C1	Chart	Word cloud Tag cloud	Word chart Wordle	Info graphic Visual thesaurus Word cluster Word art Word Chart	Messaging Advertising Randomness Mixed up graph	Mess Test Makes me think of words	Not Sure No Idea
O1	Graph	Scatter graph with proportional symbol coding	Scatterplot Bubble Graph Rainbow Scatter Graph Bubble Chart Weighted scatter plot	Population Graph Graph to show Crumb Graph Dot Chart Graph	Paint Dots Countries of the World Open Map Circley Graph Chart Line graph	Things that have color Messy	No Idea Not Sure
M2	Map	Choropleth map	Map of U.S. Density distribution Density Map Map with Concentration	Map Population Chart Population Map Map Percentage Map Saturation concentratio Hue Map	Graph Chart	Boring A Chicken	Not Sure No Idea

Börner, Katy, Russell Balliet, Adam V. Maltese, and Stephen Miles Uzzo. 2015. "Meaning Making Through Data Representation Construction and Deconstruction". Accepted for American Educational Research Association (AERA) Annual Meeting.

19

Places Where Science Museum Visitors Encounter Visualizations

<i>Location Code</i>	<i>Example Response(s)</i>	<i>% of Instances by Relative Age</i>	
		Youth	Adult
Advertisements	Seen it in ads	1.0	2.8
Art	Artwork	0.8	1.1
Books	History Book	8.8	9.7
Data and Research	Research or data that is broken dc	0.8	6.3
Entertainment	American Idol	4.3	3.5
Internet	Websites	6.6	8.2
Magazines & Brochures	National Geographic	1.5	5.8
Maps	Anywhere a map would be	7.1	4.4
Medical	Doctor's office, Psychology tests	1.3	1.7
Museums & Zoos	MOMA in NYC	1.8	0.8
News	Newspapers	2.8	10.7
Politics	Voting	0.8	1.7
Posters & Presentations	Business Presentations	0.8	4.2
Public Spaces	I think I've seen this at the mall	5.1	3.8
School	Classes- History	51.8	19.6
Technology Devices	iPod	2.0	1.0
Work & Everyday Life	Weather, At my job in marketing	2.0	12.4
Other	Everywhere	0.5	1.5
Don't Know	Not sure	0.5	0.7
Total Coded Elements		396	710

Börner, Katy, Russell Balliet, Adam V. Maltese, and Stephen Miles Uzzo. 2015. "Meaning Making Through Data Representation Construction and Deconstruction". Accepted for American Educational Research Association (AERA) Annual Meeting.

20

Visualization Types Sorted from Easy to Hard

Visualization Type	1 (Easiest)	2	3	4	5 (Hardest)
Chart	34.8	4.3	26.1	17.4	17.4
Graph	25.5	24.5	17.9	18.9	13.2
Map	21.7	31.9	21.7	17.4	7.2
Network Layout	4.5	6.0	19.4	25.4	44.8

Study Replication in Other Countries:

We are interested to replicate the study in different environments and countries. Therefore, all study materials have been made available online at <http://cns.iu.edu/2015-VisLit.html>.

Please feel free to contact the authors for more information. We are interested to collaborate on research projects related to data visualization literacy.

Börner, Katy, Russell Balliet, and Adam V. Maltese. (Submitted). "Data Visualization Literacy of Youth and Adult Science Museum Visitors". Information Visualization.

21

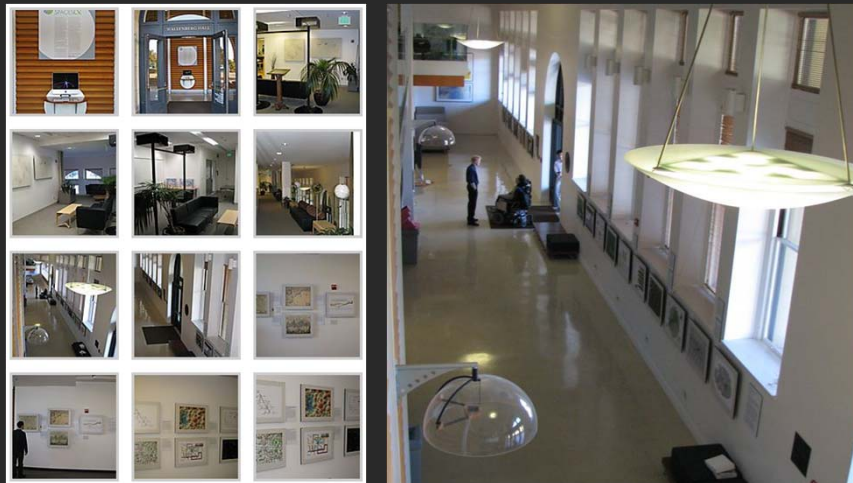
***Bringing Visualizations of BIG Data
to General Audiences***

Places & Spaces: Mapping Science Exhibit

<http://scimaps.org>



Maps are available for sale and the exhibit can be hosted by anyone.

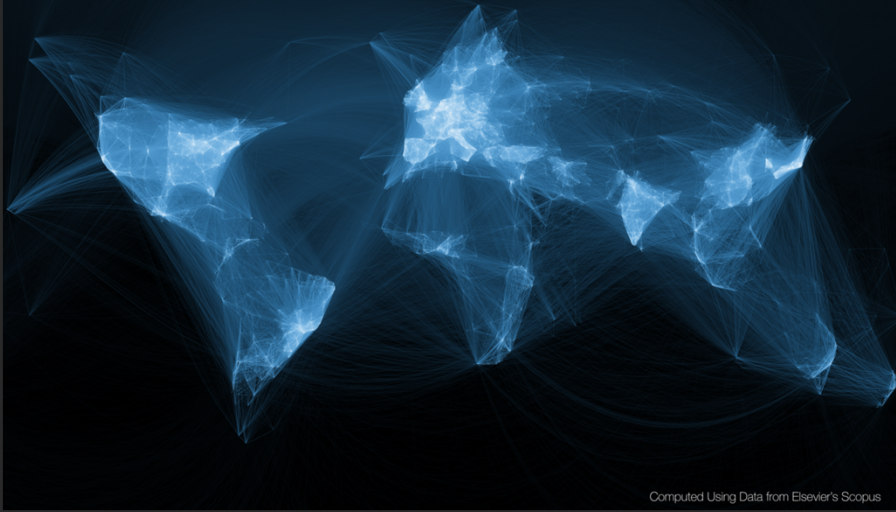


Mapping Science Exhibit on display at MEDIA X, Stanford University

<http://mediax.stanford.edu>, <http://scaleindependentthought.typepad.com/photos/scimaps>



Map of Scientific Collaborations from 2005-2009



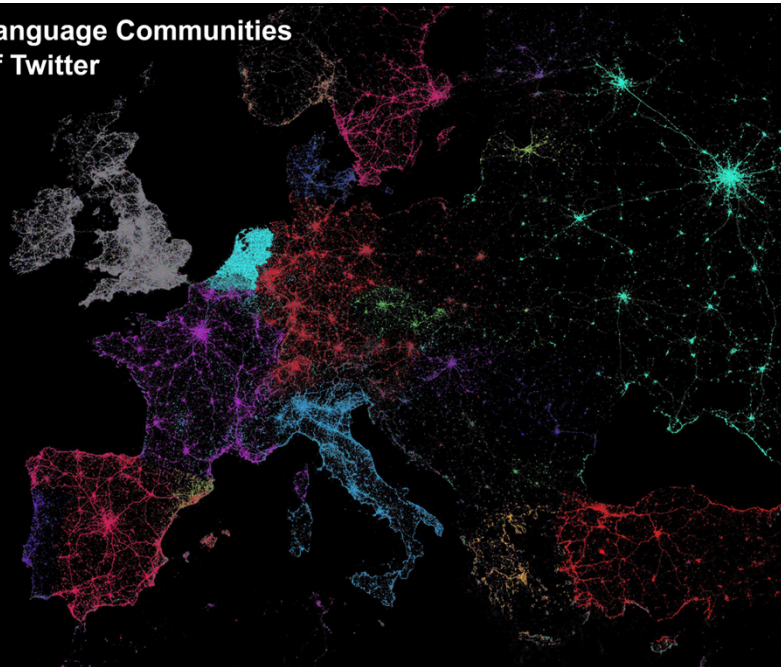
Computed Using Data from Elsevier's Scopus

Olivier H. Beauchesne, 2011. Map of Scientific Collaborations from 2005-2009.



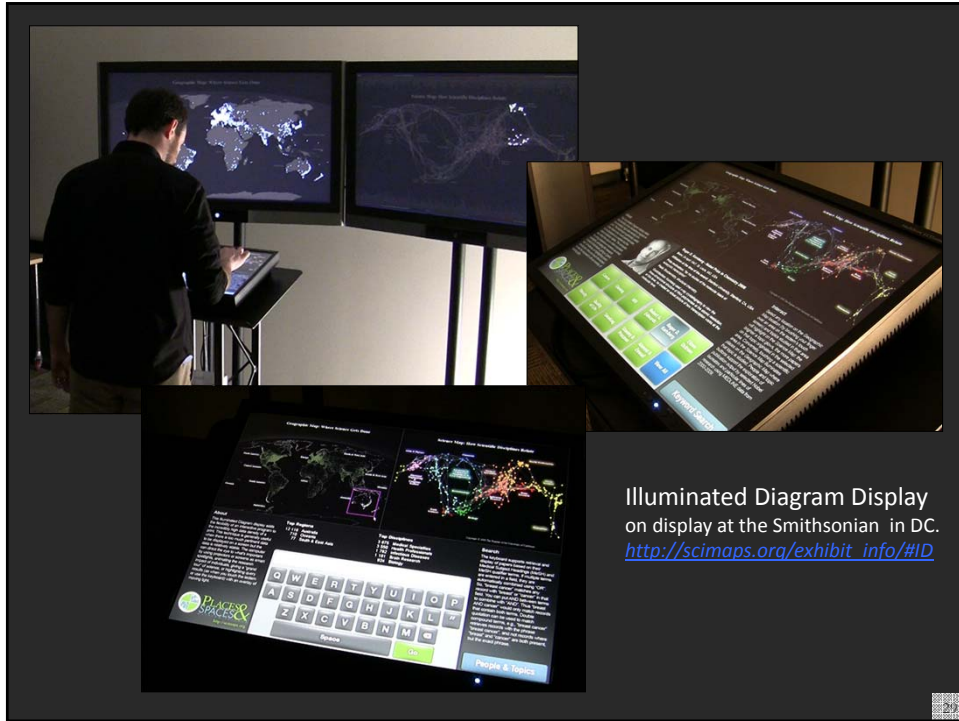
Language Communities of Twitter

- English
- Portuguese
- Spanish
- Dutch
- Russian
- French
- Italian
- German
- Turkish
- Arabic
- Swedish
- Danish
- Finnish
- Catalan
- Romanian
- Norwegian
- Lithuanian
- Slovak
- Czech
- Greek
- Hungarian
- Polish
- Slovenian
- Albanian
- Latvian
- Galician
- Hebrew
- Croatian
- Bulgarian



Language Communities of Twitter - Eric Fischer - 2012





Geographic Map: Where Science Gets Done

Science Map: How Scientific Disciplines Relate

Copyright © 2008 The Regents of the University of California

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 - 2,421
- 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

Input your search query here.

People & Topics

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.

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.

Elinor Ostrom - Nobel Prize in Economic Sciences 2009

Born: 7 August 1933, New York, NY, USA

Affiliation at the time of the award: Indiana University, Bloomington, IN, USA, Arizona State University, Tempe, AZ, USA

Prize motivation: "for her analysis of economic governance, especially the commons"

Field: Economic governance

Contribution: Challenged the conventional wisdom by demonstrating how local property can be successfully managed by local commons without any regulation by central authorities or privatization.

Interact

Select any location on the Geographic Map location (by brushing your finger over an area on the lectern's touch screen) and topics studied in that area will highlight on the Science Map: the brighter a topic glows, the more papers on that topic originated in the selected area. Conversely, touching a scientific area in the Science Map illuminates places on the Geographic Map where that topic is studied. People and topic buttons support the exploration of publication output by selected Noble laureates and particular lines of research using MEDLINE data from 2000-2009.

Cancer	Cloning	HIV	Robert G. Edwards	Roger D. Kornberg	Elinor Ostrom
Obesity	Quality of Life	Smoking	Stanley B. Prusiner	Ahmed H. Zewail	View All

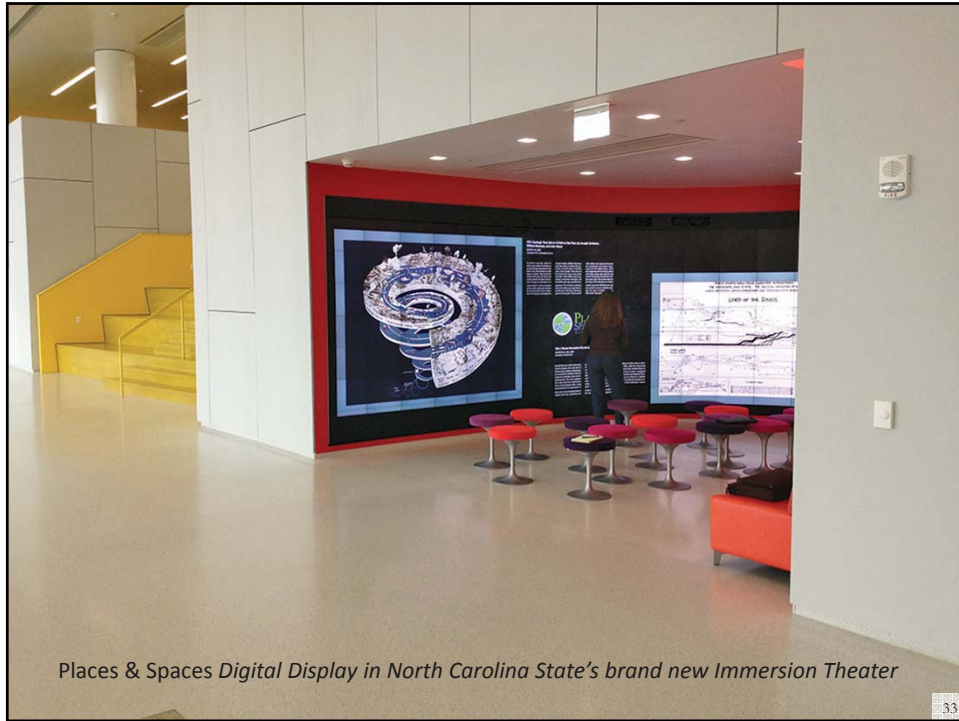
<http://scimaps.org>

Keyword Search

Science Maps in "Expedition Zukunft" science train visiting 62 cities in 7 months 12 coaches, 300 m long Opening was on April 23rd, 2009 by German Chancellor Merkel

<http://www.expedition-zukunft.de>

16



Places & Spaces Digital Display in North Carolina State's brand new Immersion Theater

33

***Empowering ANYONE
to Convert BIG Data into Insights***

34

Overview

This course provides an overview about the state of the art in information visualization. It teaches the process of producing effective visualizations that take the needs of users into account.

This year, the course can be taken for three Indiana University credits as part of the Online Data Science Program just announced by the School of Informatics and Computing. Students interested in applying to the program can find more information here.

Among other topics, the course covers:

- Data analysis algorithms that enable extraction of patterns and trends in data
- Major temporal, geospatial, topical, and network visualization techniques
- Discussions of systems that drive research and development.

Just like last year, students will have the opportunity to collaborate on real-world projects for a variety of clients. Click here to see this year's list of clients and projects.

Everyone who registers gains free access to the Scholarly Database (26 million paper, patent, and grant records) and the Sci2 Tool (100+ algorithms and tools).

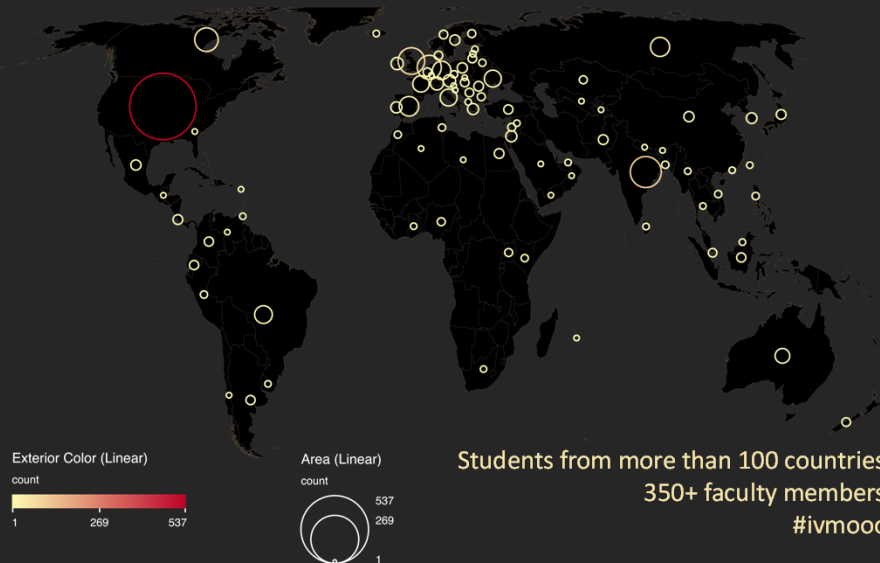
Please watch the introduction video to learn more.



IVMOOC 2014 course materials will be available until end of November 2014. The IVMOOC 2015 will open in January 2015 with new materials and a cloud computing setup.

Register for free at <http://ivmoo.cns.iu.edu>. Class will restart in January 2015.

The Information Visualization MOOC
ivmoo.cns.iu.edu



Course Schedule

- **Session 1** – Workflow design and visualization framework
- **Session 2** – “When:” Temporal Data
- **Session 3** – “Where:” Geospatial Data
- **Session 4** – “What:” Topical Data

Mid-Term

Students work in teams with clients.

- **Session 5** – “With Whom:” Trees
- **Session 6** – “With Whom:” Networks
- **Session 7** – Dynamic Visualizations and Deployment

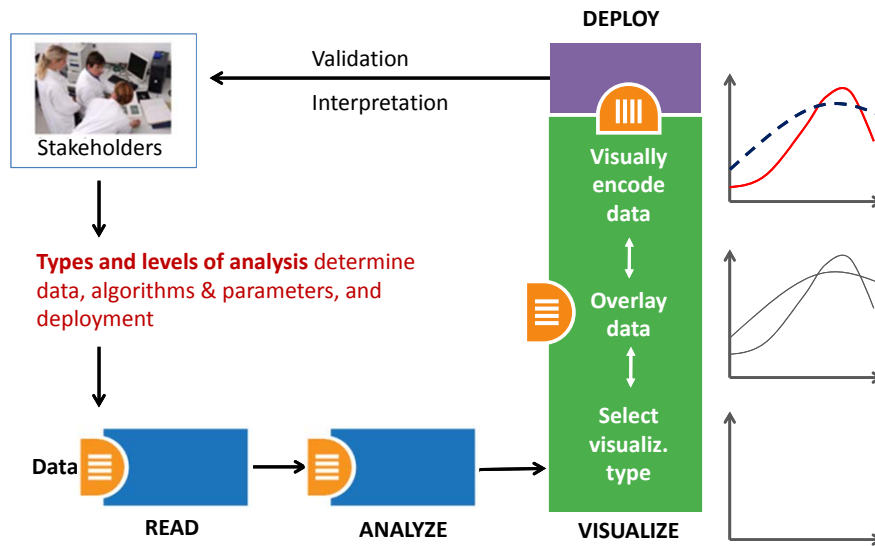
Final Exam

Final grade is based on Midterm (**30%**), Final (**40%**), Client Project (**30%**).

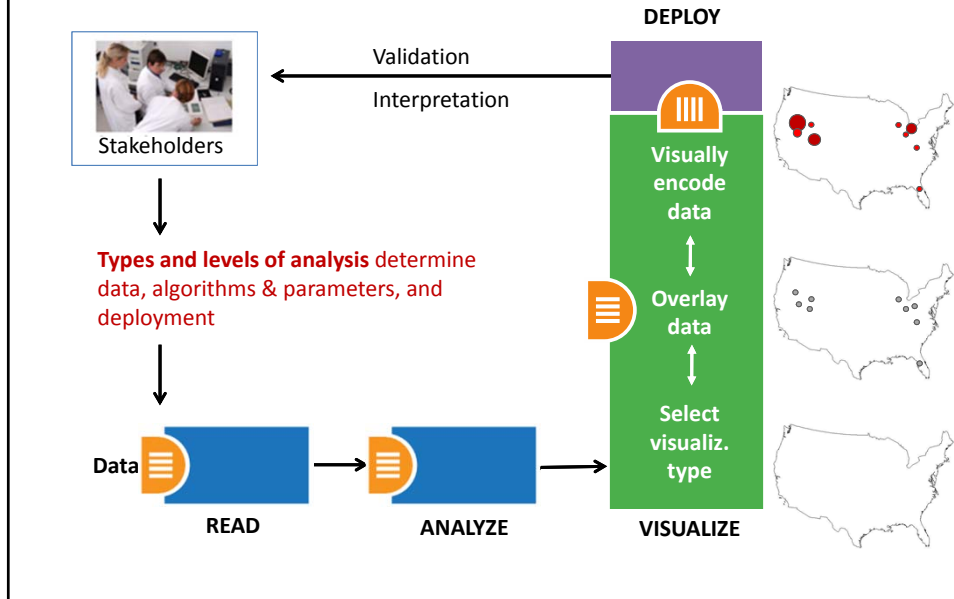


37

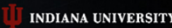



Needs-Driven Workflow Design



Needs-Driven Workflow Design



Clients

Information Visualization MOOC    

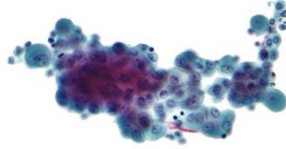
List of Clients

ISIS **Project Title:** Isis: 100 Years
Client Name: Jay Malone
Project goal/scientific or practical value: A visual representation Isis' contributors and locales over the past 100 years. Isis is the journal of the History of Science Society. This representation will provide a dynamic picture of how scholarship in the history of science has shifted over the past century.
Information on dataset(s) to be used: Citation information, author locale, and issue number for Isis publications.
Relevant publications, websites, etc: <http://www.press.uchicago.edu/ucp/journals/journal/isis.html>
Conditions under which students can publish results and/or add project results to their resume: Client would like to approve results.

oycib **Project Title:** e-Xploration
Client Name: Luly
Project goal/scientific or practical value: e-Xploration is an agent-based model for the ethnographic observation and the registry, analysis, and interpretation of social practices in virtual communities for intervention in the development of collaboration and cooperation. This project will analyze the interactions between subjects and objects in a platform collaborative community called OYCIB, a project based on e-Xploration (e-crick.net).
Information on dataset(s) to be used: I can provide a data base in .graphml format for the students. The file .graphml contains the interactions between subjects and objects in a platform collaborative community called OYCIB. In the level of practice, it is not necessary that students know agent-based models for using the database. But, in another level, for example: the collaborate level for the OYCIB development, it is necessary to have basic knowledge in AMS or MAS and another competences like PHP and MySQL.
Relevant publications, websites, etc: <http://www.e-crick.net/logs>
Conditions under which students can publish results and/or add project results to their resume: If any person or institution use my dataset or another info about eXploration (e-crick.net, oycib.net), I need to approve the results and appear as co-author.

<http://ivmooc.cns.iu.edu/clients.html>

Diogo Carmo



Mesothelioma

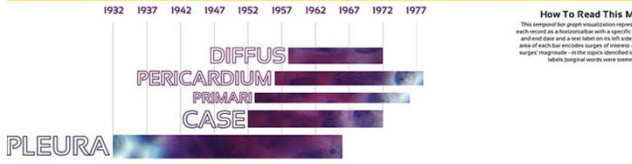
Main title topics in Medline papers

Mesothelioma is, more precisely, malignant mesothelioma is a kind of cancer that develops from transformed cells originating on the mesothelium, the protective lining that covers many of the internal organs of the body. It is usually caused by exposure to asbestos.

The most common anatomical site for the development of mesothelioma is the pleura (the outer lining of the lungs and internal chest walls), but it can also arise in the peritoneum (the lining of the abdominal cavity), and the pericardium (the sac that surrounds the heart, or the outer layer of the heart).

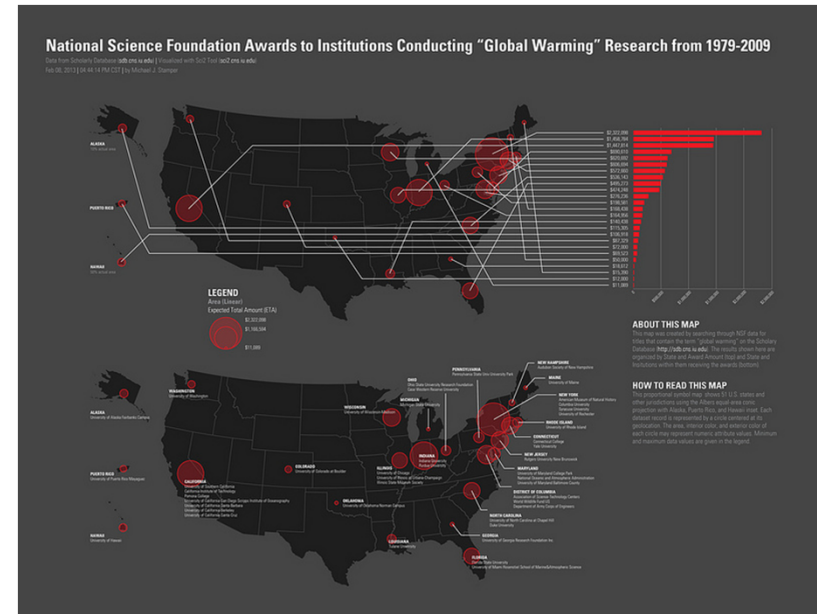
Most people who develop mesothelioma have worked in jobs where they inhaled asbestos, or were exposed to asbestos dust and fibers in other ways. It has also been suggested that working in the family member who worked with asbestos increases their risk for developing mesothelioma. Unlike lung cancer, there seems to be no association between mesothelioma and tobacco smoking, but smoking greatly increases the risk of getting asbestos-related cancers. Some people who were exposed to asbestos have collected damages in asbestos-related lawsuits, including mesothelioma. Compensation via asbestos funds or class action lawsuits is an important issue in law practices regarding mesothelioma.

MALIGNANT PLEURAL CYSTIC BENIGN DIAGNOSIS



How To Read This Map
This map shows the growth mesothelioma research each record as a horizontal bar with a specific start and end date and a vertical bar to its left. The area of each bar includes target of interest - and target magnitude - on the right identified as labels (original words were summed)

Author: Diogo Carmo (<http://diogo.carmo@epi.usp.br>) | Visualization software: Sci2 Team (2009). Science of Science (Sci2) Team, Indiana University and SciTech Strategies, <http://c2.com.us.edu/> | Dataset: Medline Papers, as available in Scholarly Database: <http://pubs.nlm.nih.gov/>; Text and images: Wikipedia: Mesothelioma article, available at <http://en.wikipedia.org/wiki/Mesothelioma> | Data: Sanjour, by Bernd Montag © 2001. All Rights Reserved. This text is hereby released under the Creative Commons Attribution-NonCommercial-ShareAlike license (<http://creativecommons.org/licenses/by-nc-sa/2.0/>)



mistamper_ivmoooc

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.sjis.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/

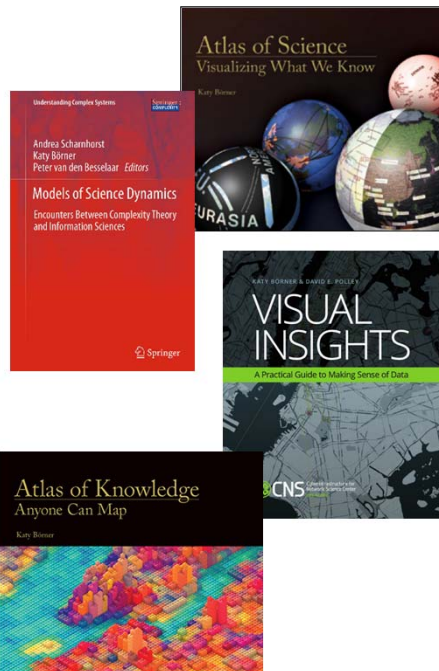
Börner, Katy (2010) **Atlas of Science: Visualizing What We Know**. The MIT Press. <http://scimaps.org/atlas>

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

Katy Börner, Michael Conlon, Jon Corson-Rikert, Cornell, Ying Ding (2012) **VIVO: A Semantic Approach to Scholarly Networking and Discovery**. Morgan & Claypool.

Katy Börner and David E Polley (2014) **Visual Insights: A Practical Guide to Making Sense of Data**. The MIT Press.

Börner, Katy (2015) **Atlas of Knowledge: Anyone Can Map**. The MIT Press. <http://scimaps.org/atlas2>



43

All papers, maps, tools, talks, press are linked from <http://cns.iu.edu>
 These slides will soon be at <http://cns.iu.edu/docs/presentations>
 CNS Facebook: <http://www.facebook.com/cnscenter>
 Mapping Science Exhibit Facebook: <http://www.facebook.com/mappingscience>

44