Visualizing Science and Technology Data

Katy Börner

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

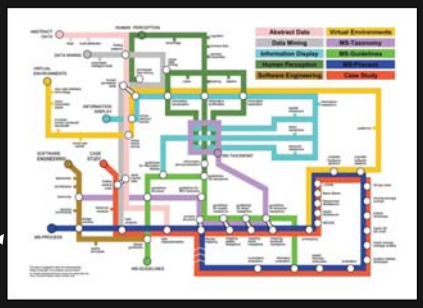
CyberBridges Workshop
CEB Waterview Conference Center, Arlington, VA

August 31, 2015

Language Communities of Twitter - Eric Fischer - 2012

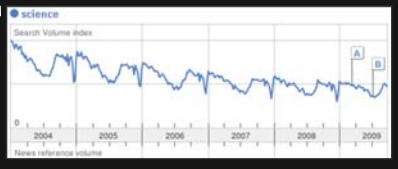
Terra bytes of data

Descriptive & Predictive Models



Find your way

Find collaborators, friends



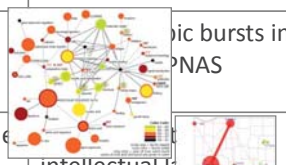
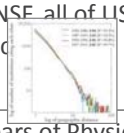
Identify trends

Type of Analysis vs. Level of Analysis

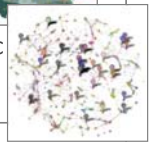
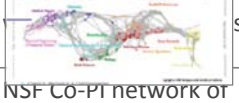
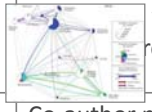
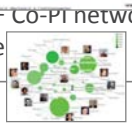
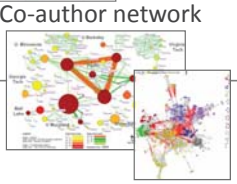
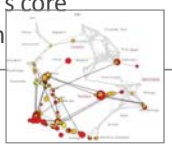
	Micro/Individual (1-100 records)	***Meso/Local (101–100,000 records)***	***Macro/Global (100,000 < records)***
Statistical Analysis/Profiling	Individual person and their expertise profiles	Larger labs, centers, universities, research domains, or states	All of NSF, all of USA, all of science.
Temporal 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

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 sc
Temporal Analysis (When?)	Funding portfolio of one individual	ic bursts in PNAS	113 Years of Physics Research
Geospatial Analysis (Where?)	Career trajectory of one individual	intellectual la	PNAS publications
Topical Analysis (What?)	Base knowledge from s.	f research	VxOrd/Topic NIH funding
Network Analysis (With Whom?)	NSF Co-PI network of one	Co-author network	NIH's core com

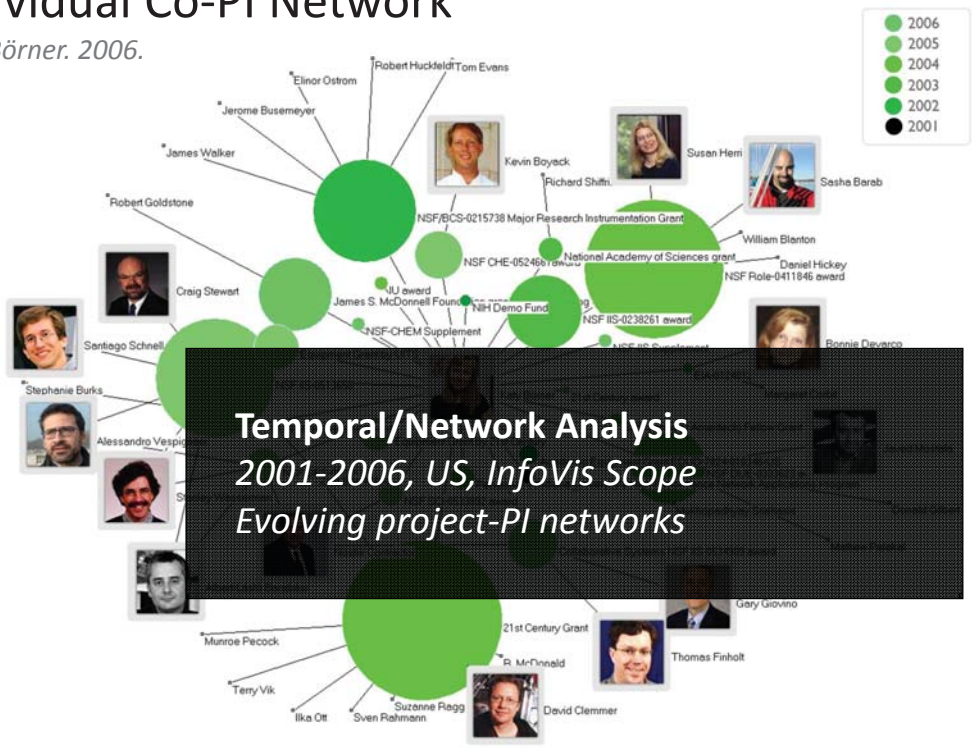





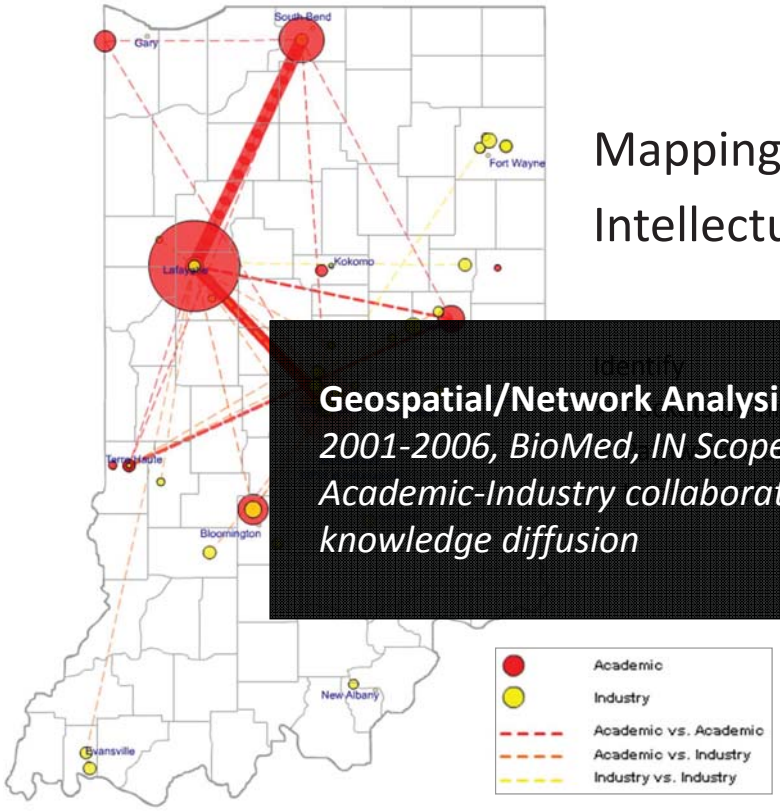
Individual Co-PI Network

Ke & Börner. 2006.



Temporal/Network Analysis
2001-2006, US, InfoVis Scope
Evolving project-PI networks

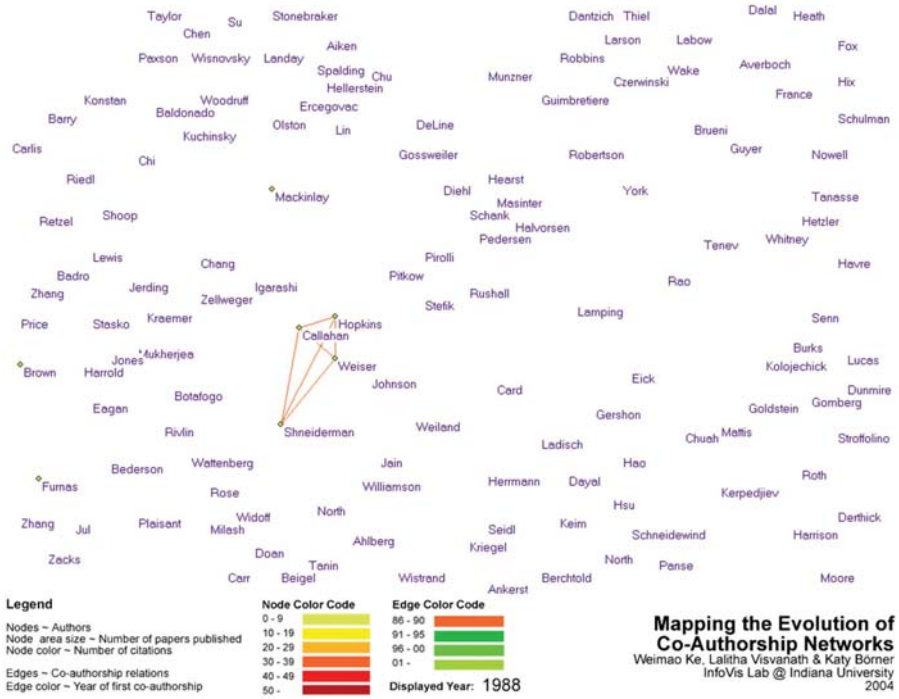
Mapping Indiana's Intellectual Space



Geospatial/Network Analysis
2001-2006, BioMed, IN Scope
Academic-Industry collaborations and knowledge diffusion

Mapping the Evolution of Co-Authorship Networks

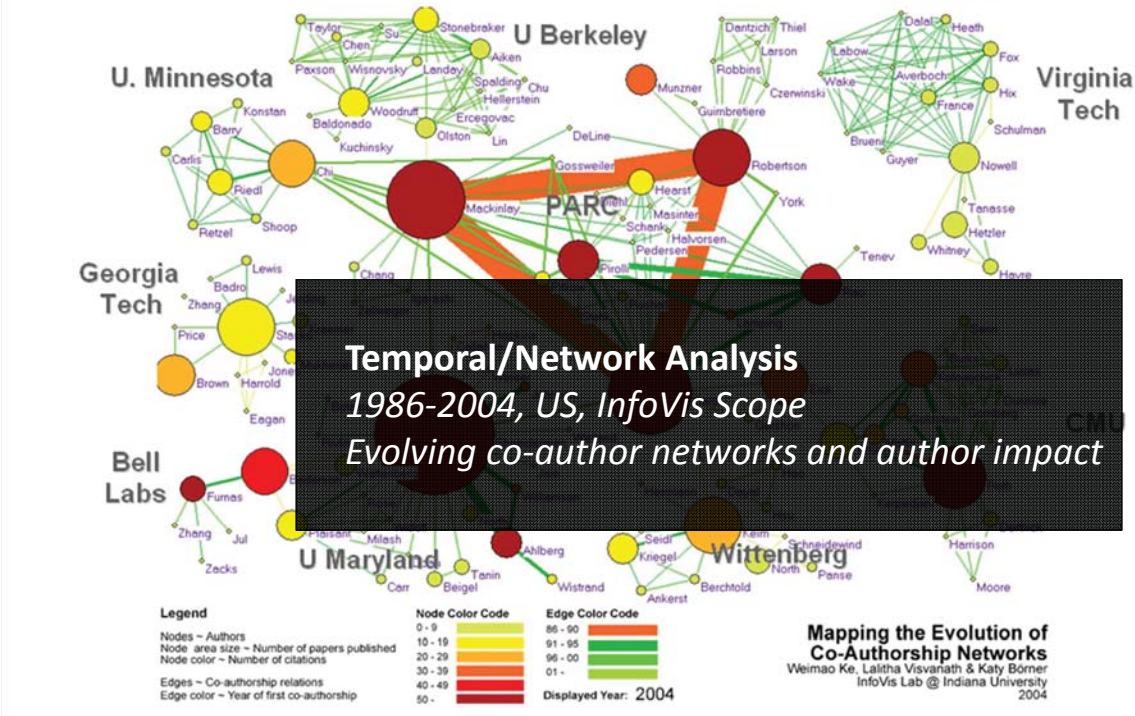
Ke, Visvanath & Börner. 2004. Won 1st prize at the IEEE InfoVis Contest.



7

Mapping the Evolution of Co-Authorship Networks

Ke, Visvanath & Börner. 2004. Won 1st prize at the IEEE InfoVis Contest.



8

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

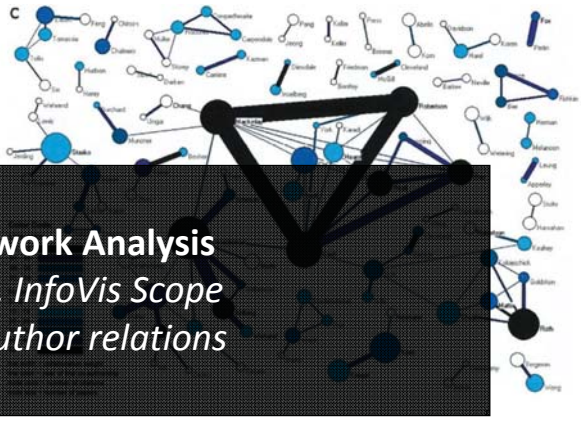
Börner, Dall'Asta, Ke & Vespignani. 2005. Complexity 10 (4):58-67.

Research question:

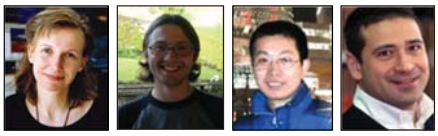
- Is science driven by prolific single experts or by high-impact co-authorship teams?

Contributions:

- New approach to allocate citational credit.
- Novel weighted graph representation.
- Visualization of the growth of weighted co-author network.
- Centrality measures to study author impact.
- Global statistical analysis of paper production and citations in correlation with co-authorship team size over time.
- Local, author-centered entropy measure.



Temporal/Network Analysis
1986-2004, US, InfoVis Scope
Impact of co-author relations

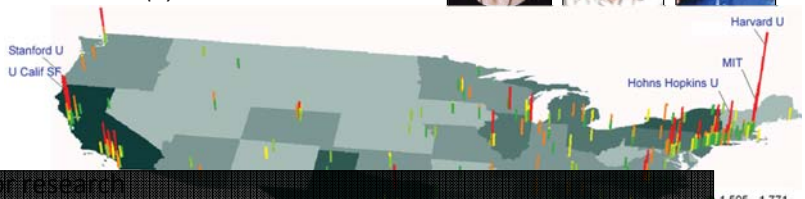


Spatio-Temporal Information Production and Consumption of Major U.S. Research Institutions

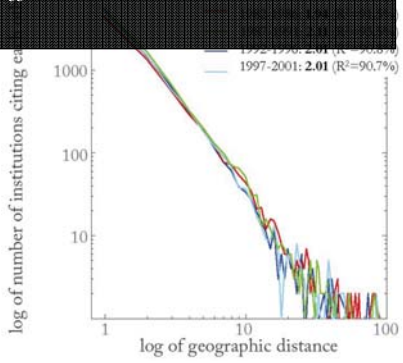
Börner, Penumarthy, Meiss, & Ke.
2006. "Mapping the Diffusion of Scholarly Knowledge Among Major U.S. Research Institutions." Scientometrics 68 (3): 415-426.

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 to collaborate with high-quality researchers?
3. Does the Internet lead to more global citation patterns—i.e., more citation links between papers produced at geographically distant research instructions?



Temporal/Geospatial Analysis
1982-2001, US, PNAS (BioMed) Scope
Citation impact and knowledge diffusion



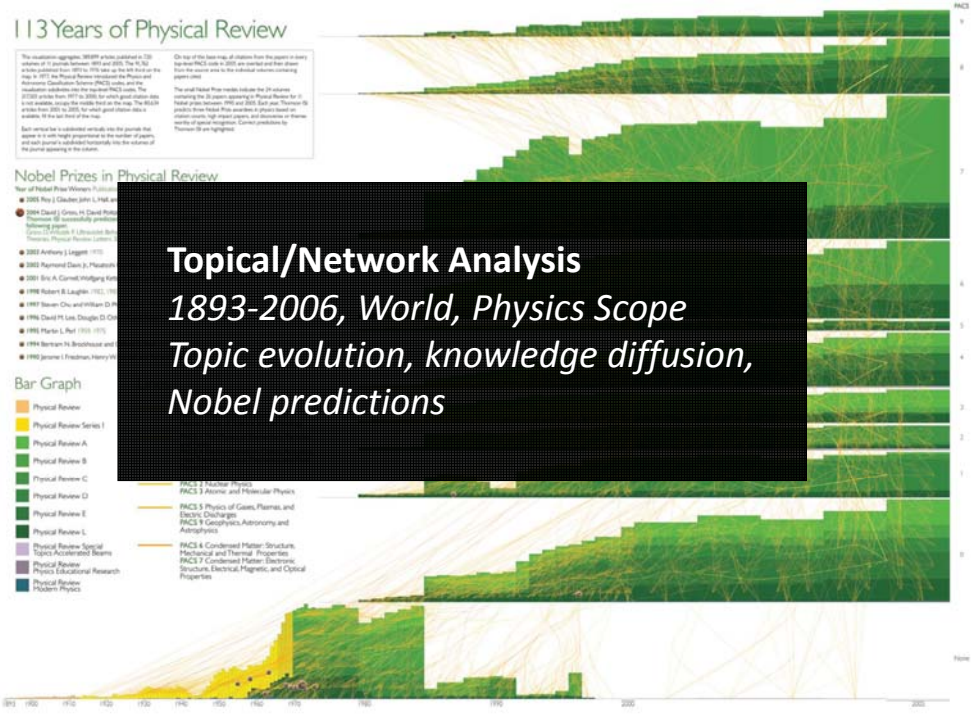
Contributions:

- 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.

113 Years of *Physical Review*

http://scimaps.org/dev/map_detail.php?map_id=171

Herr II, Duhon, Hardy, Penumarthy & Börner.



Topical/Network Analysis

1893-2006, World, Physics Scope

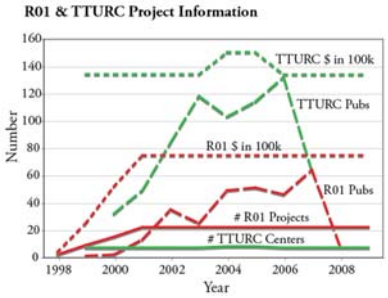
Topic evolution, knowledge diffusion, Nobel predictions

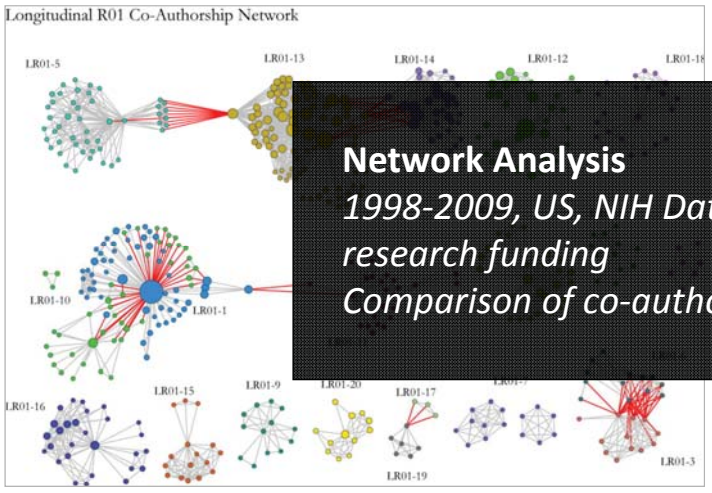
Mapping Transdisciplinary Tobacco Use Research Centers Publications

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





Network Analysis

1998-2009, US, NIH Data on Tobacco research funding

Comparison of co-author networks

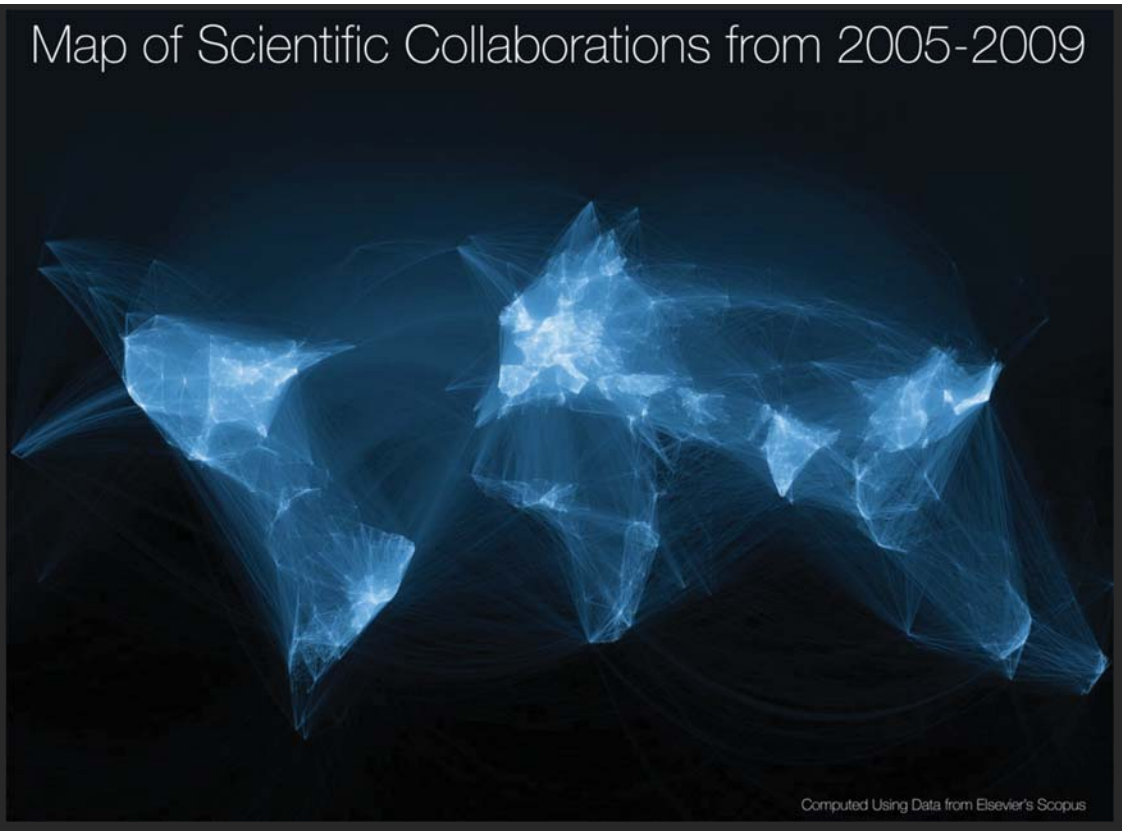
Research Collaborations by the Chinese Academy of Sciences

Huang, Duhon, Hardy & Börner

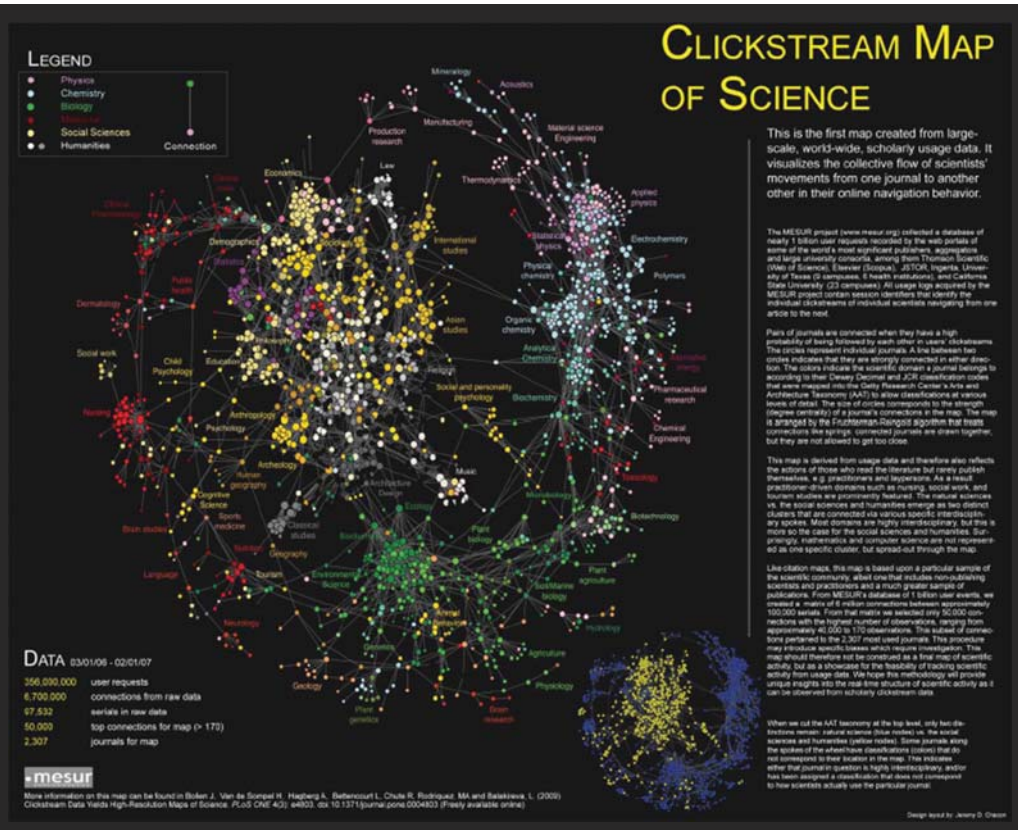
Mapping Topic Bursts

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

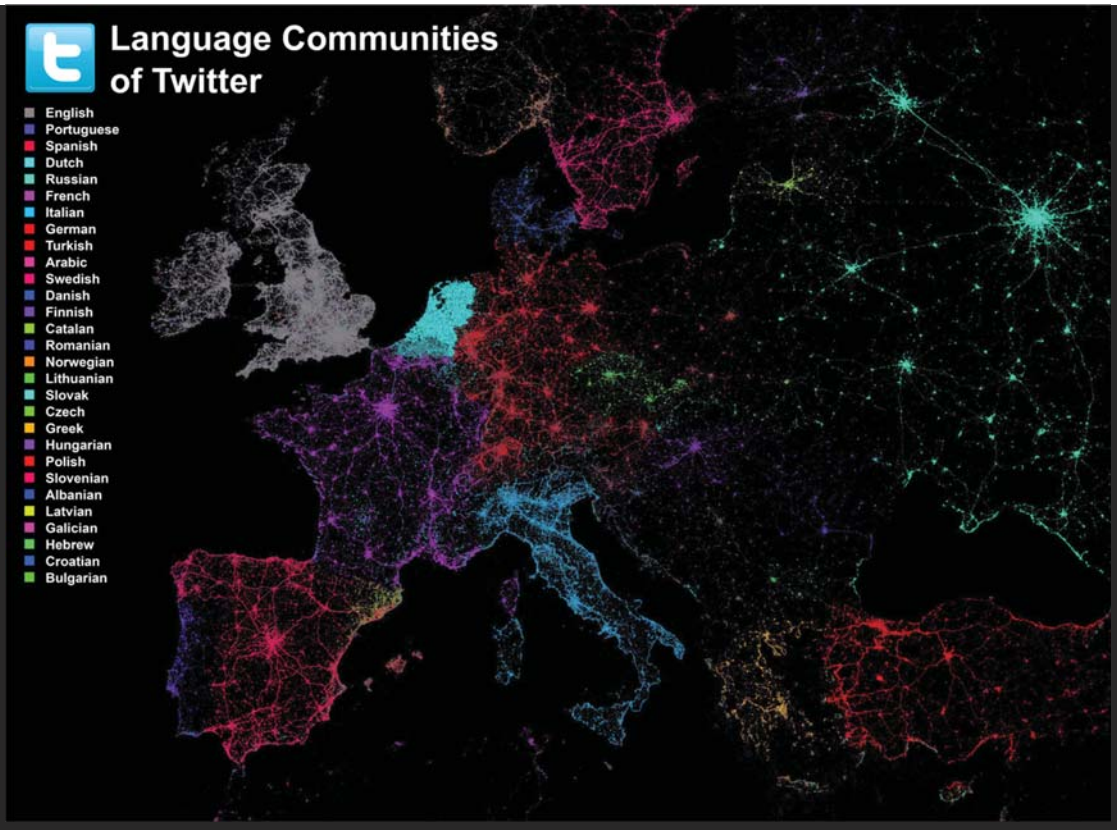
Mane & Börner. 2004. PNAS 101(Suppl. 1): 5287-5290.



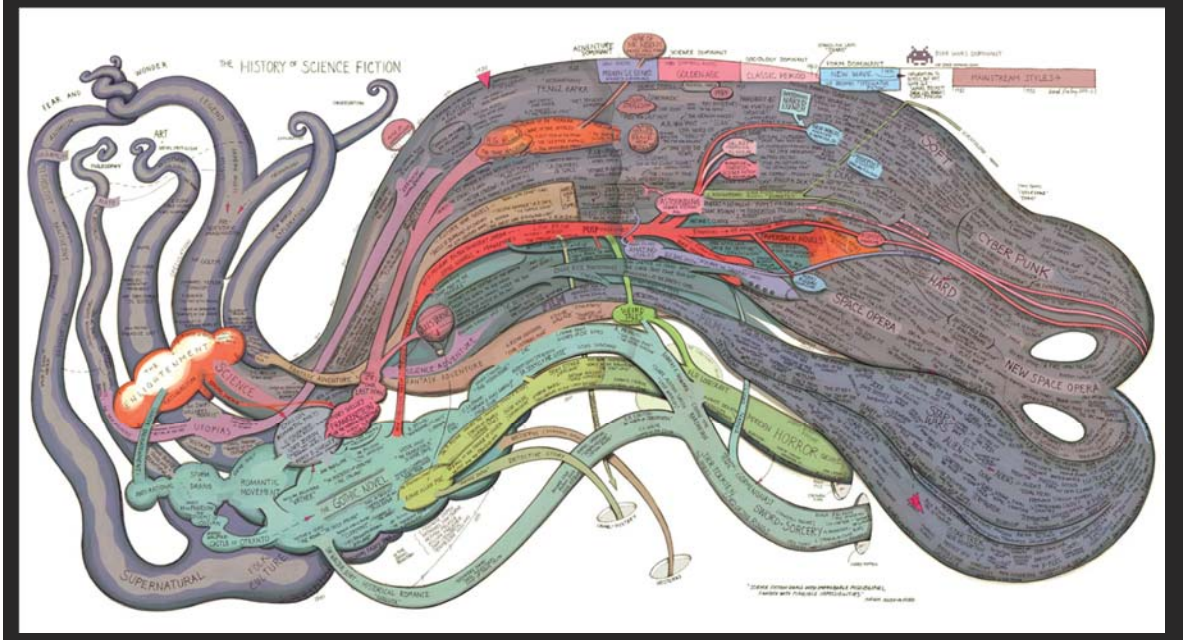
Olivier H. Beauchesne, 2011. Map of Scientific Collaborations from 2005-2009.



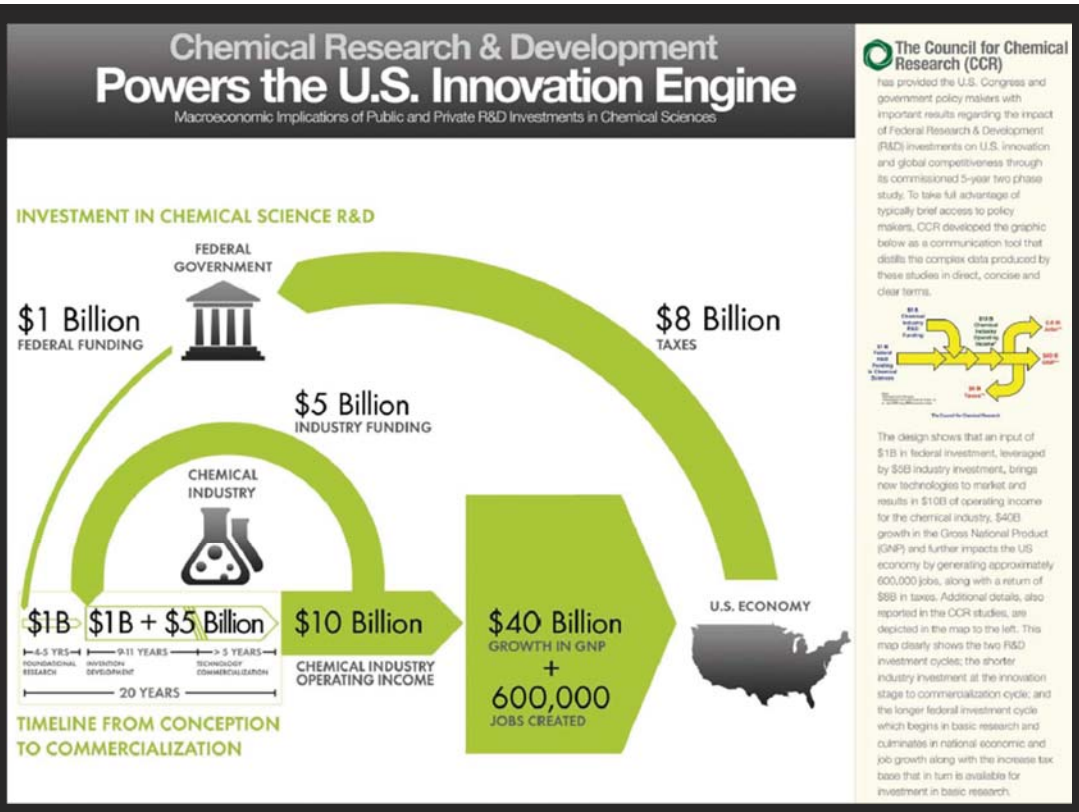
Bollen, Johan, Herbert Van de Sompel, Aric Hagberg, Luis M.A. Bettencourt, Ryan Chute, Marko A. Rodriquez, Lyudmila Balakireva. 2008. A Clickstream Map of Science.



Language Communities of Twitter - Eric Fischer - 2012

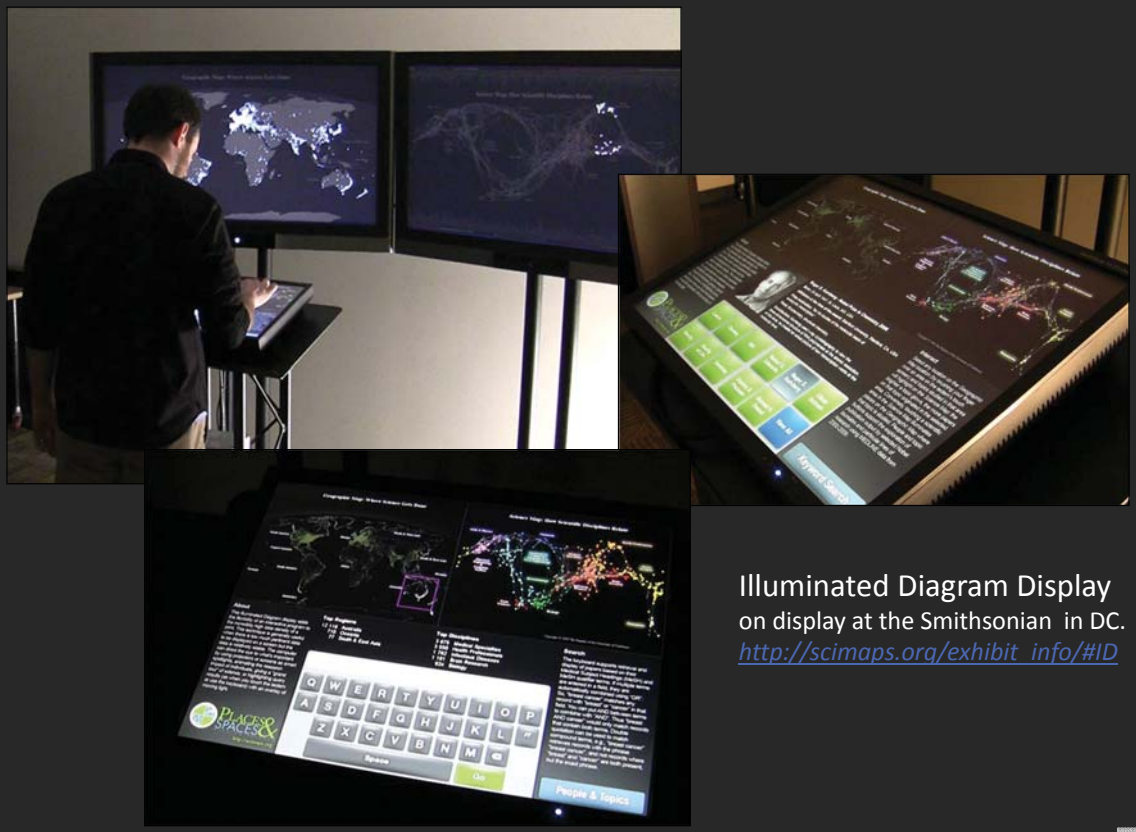


Ward Shelley . 2011. History of Science Fiction.

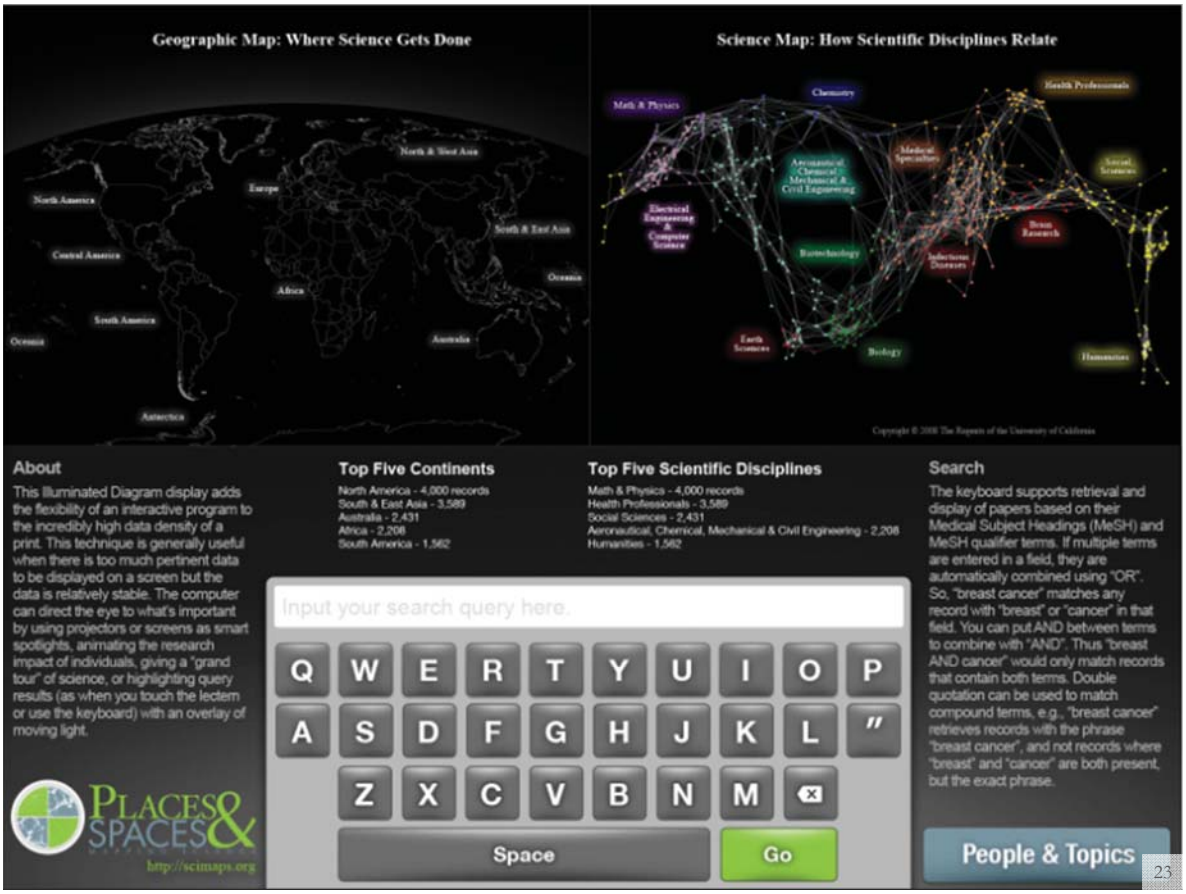


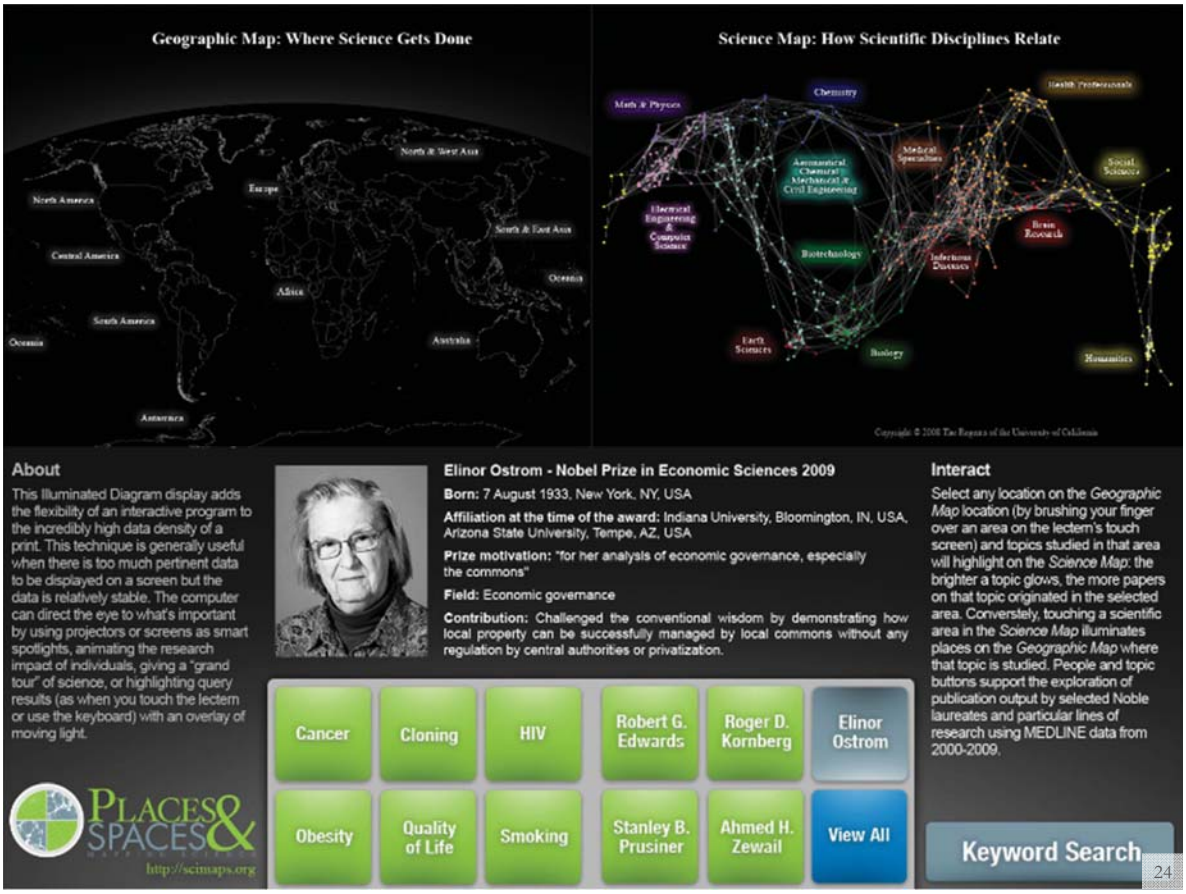
Council for Chemical Research. 2009. Chemical R&D Powers the U.S. Innovation Engine. Washington, DC. Courtesy of the Council for Chemical Research.

Places & Spaces at Northwestern University
May 14 - September 23, 2015

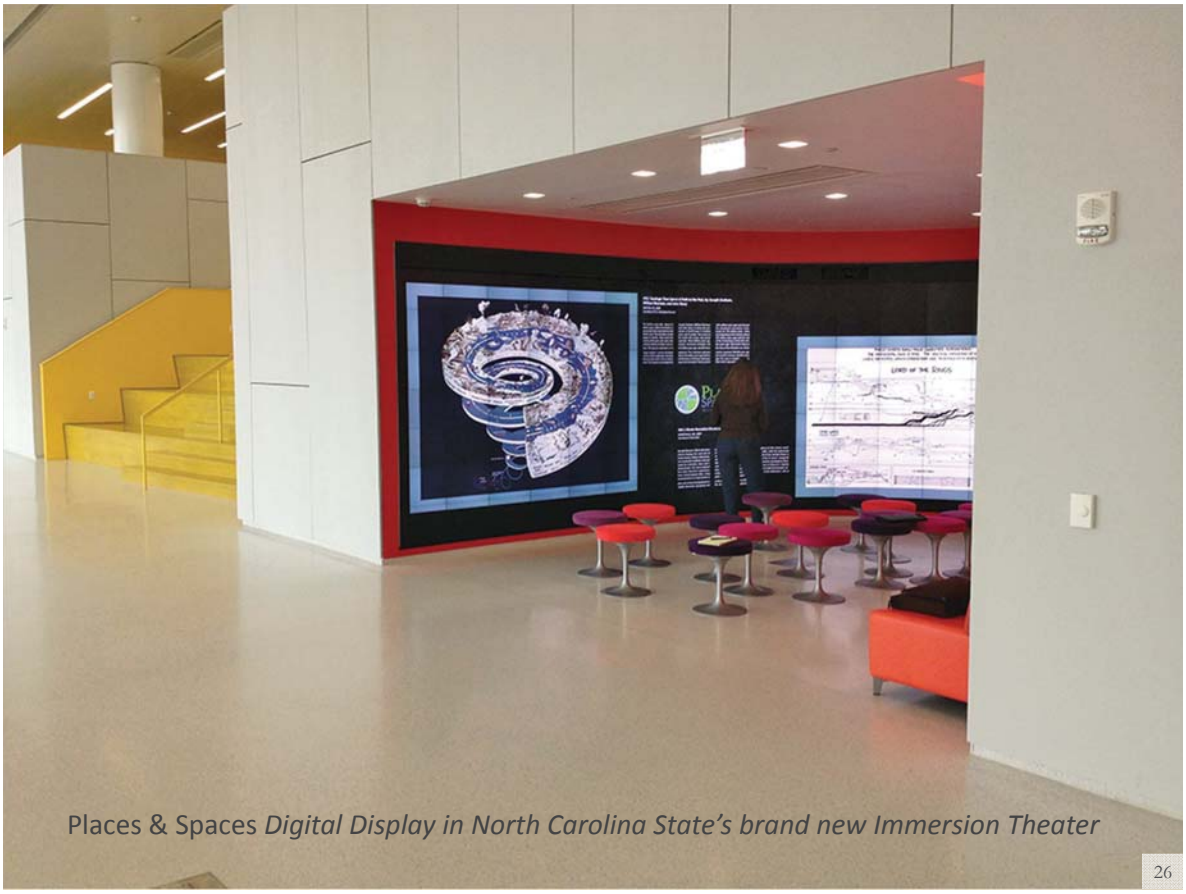


Illuminated Diagram Display
on display at the Smithsonian in DC.
http://scimaps.org/exhibit_info/#ID

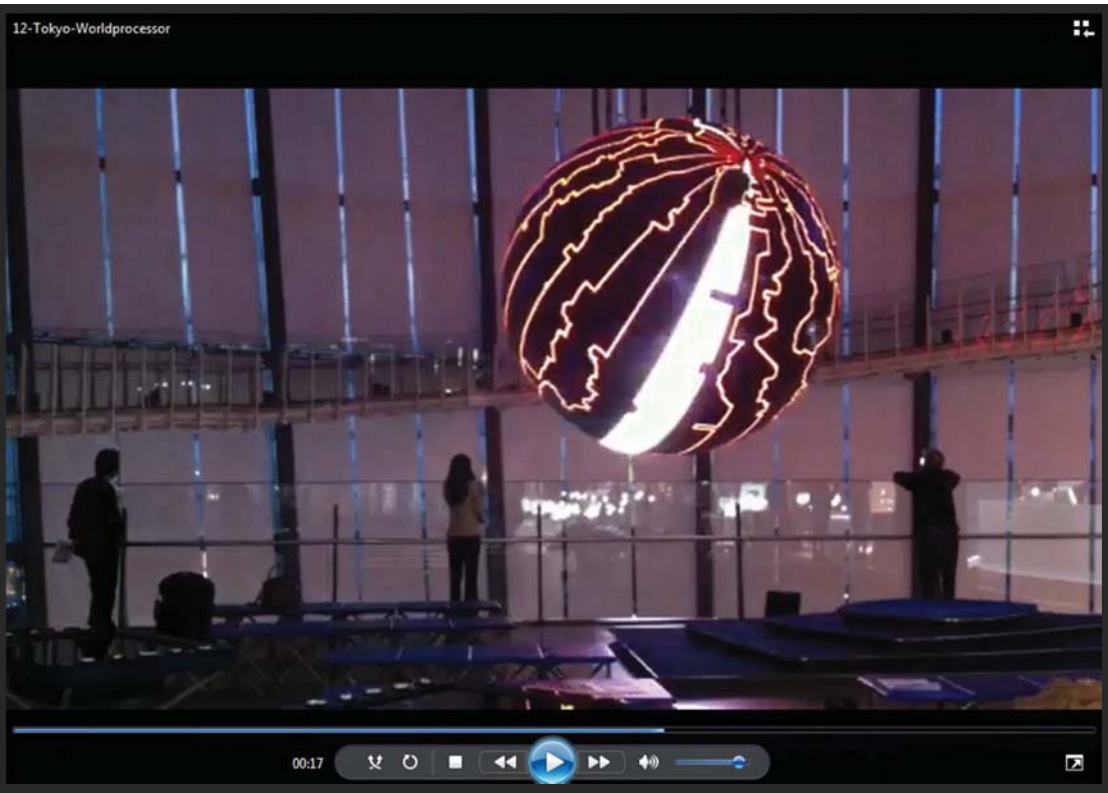

Geographic Map: Where Science Gets Done
North America
Europe
North & West Asia
South & East Asia
Central America
Africa
Oceania
South America
Australia
Antarctica
Science Map: How Scientific Disciplines Relate
Math & Physics
Chemistry
Health Professionals
Medical Specialties
Social Sciences
Aeronautical, Chemical, Mechanical & Civil Engineering
Electrical Engineering & Computer Science
Brain Research
Biotechnology
Infectious Diseases
Earth Sciences
Biology
Humanities
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.
PLACES & SPACES
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
Input your search query here.
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
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.
People & Topics


Geographic Map: Where Science Gets Done
North America
Europe
North & West Asia
South & East Asia
Central America
Africa
Oceania
South America
Australia
Antarctica
Science Map: How Scientific Disciplines Relate
Math & Physics
Chemistry
Health Professionals
Medical Specialties
Social Sciences
Aeronautical, Chemical, Mechanical & Civil Engineering
Electrical Engineering & Computer Science
Brain Research
Biotechnology
Infectious Diseases
Earth Sciences
Biology
Humanities
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.
PLACES & SPACES
http://scimaps.org
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.
Cancer
Cloning
HIV
Robert G. Edwards
Roger D. Kornberg
Elinor Ostrom
Obesity
Quality of Life
Smoking
Stanley B. Prusiner
Ahmed H. Zewail
View All
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.
Keyword Search

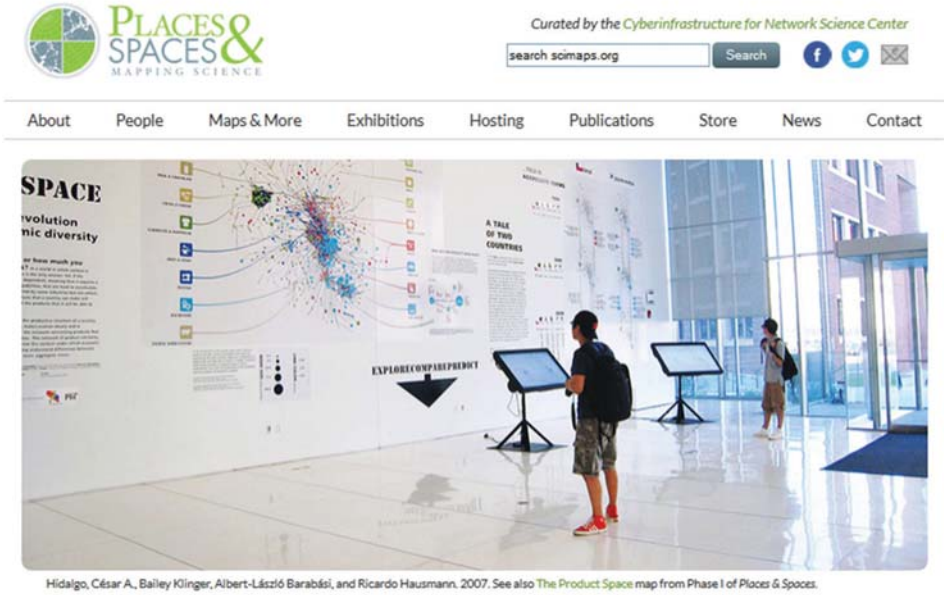
Science Maps in “Expedition Zukunft” science train visiting 62 cities in 7 months 12 coaches, 300 m long Opening was on April 23rd, 2009 by German Chancellor Merkel
http://www.expedition-zukunft.de



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



Ingo Gunther's Worldprocessor globe design on display at the Giant Geo Cosmos OLED Display at the Museum of Emerging Science and Innovation in Tokyo, Japan



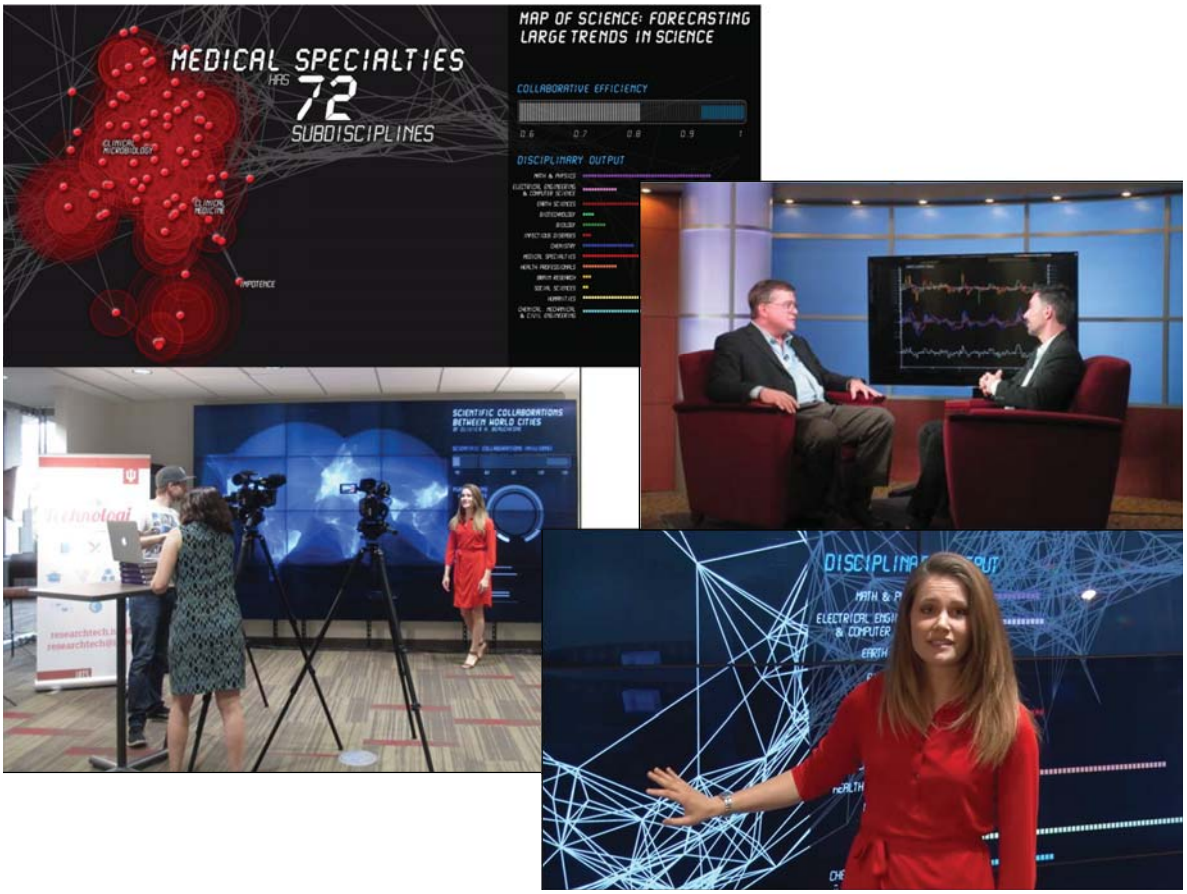
Hidalgo, César A., Bailey Klinger, Albert-László Barabási, and Ricardo Hausmann. 2007. See also The Product Space map from Phase I of *Places & Spaces*.

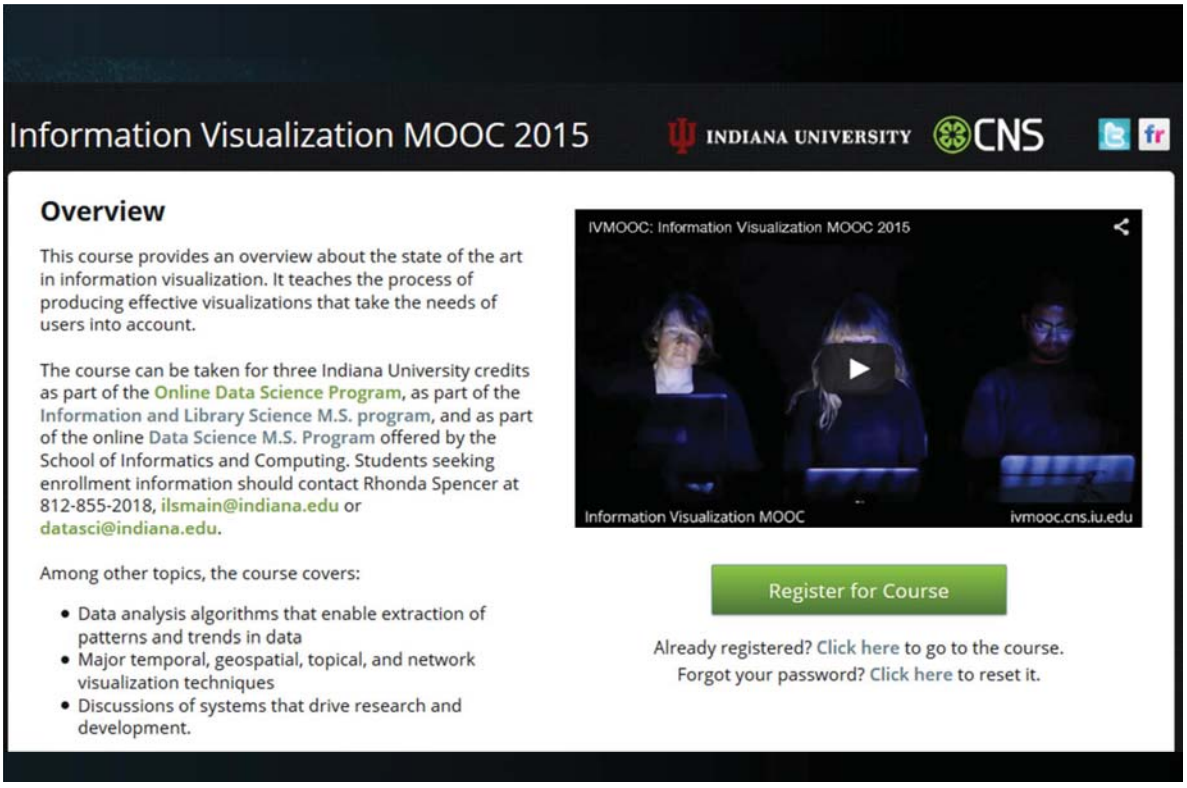
Call for Macroscope Tools for the *Places & Spaces: Mapping Science* Exhibit (2015)

http://scimaps.org/call

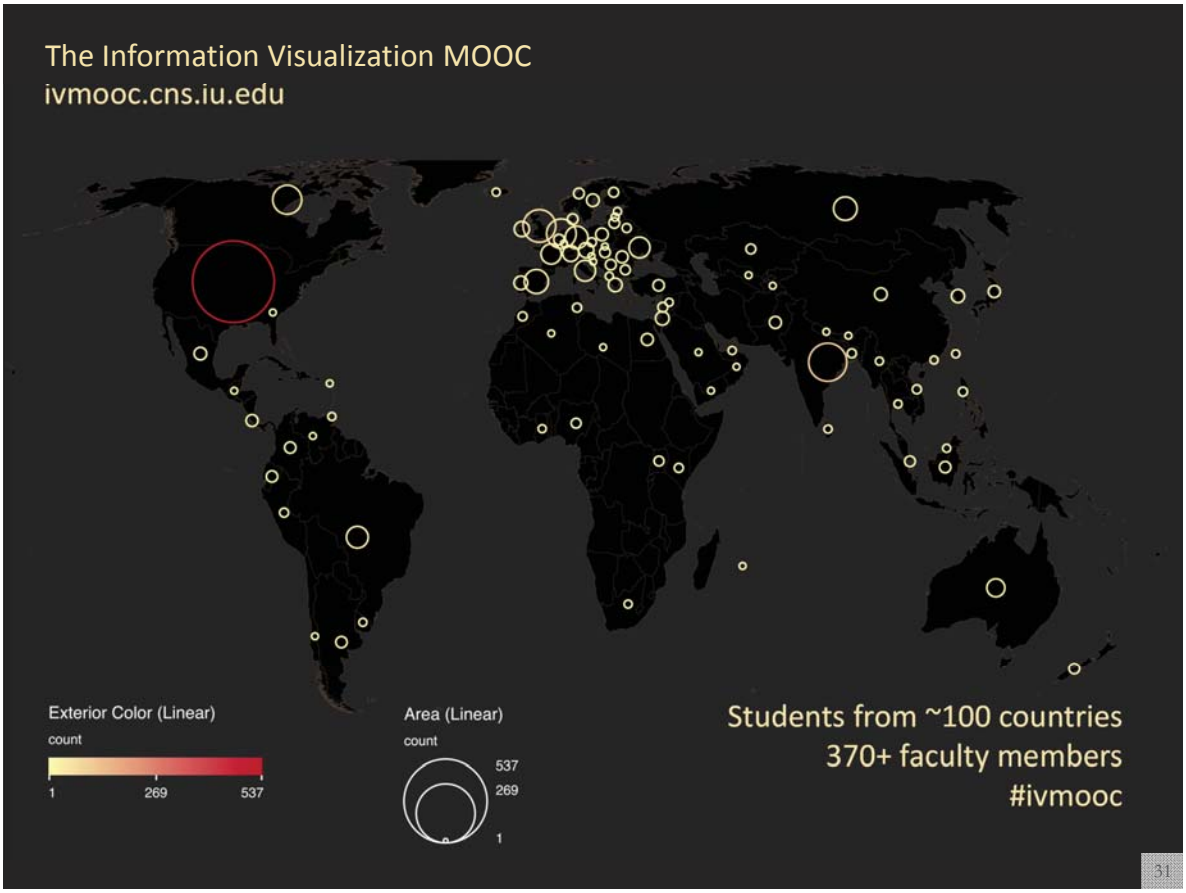
Themes for the upcoming iterations/years are:

- 11th Iteration (2015): Macroscopes for Interacting With Science
- 12th Iteration (2016): Macroscopes for Making Sense of Science
- 13th Iteration (2017): Macroscopes for Forecasting Science
- 14th Iteration (2018): Macroscopes for Economic Decision Makers
- 15th Iteration (2019): Macroscopes for Science Policy Makers
- 16th Iteration (2020): Macroscopes for Scholars





Register for free at http://ivmooc.cns.iu.edu. Class restarts January, 2016.



Course Schedule

Part 1: Theory and Hands-On

- **Session 1** – Workflow Design and Visualization Framework
- **Session 2** – "When:" Temporal Data
- **Session 3** – "Where:" Geospatial Data
- **Session 4** – "What:" Topical Data

Mid-Term

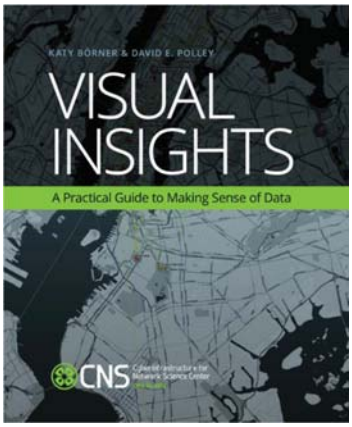
- **Session 5** – "With Whom:" Trees
- **Session 6** – "With Whom:" Networks
- **Session 7** – Dynamic Visualizations and Deployment

Final Exam

Part 2: Students work in teams on client projects.

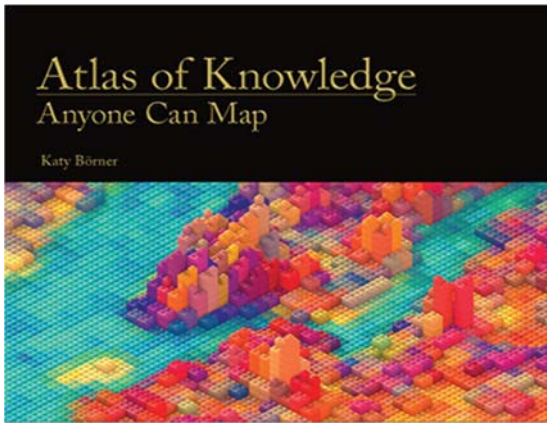
Final grade is based on Class Participation (10%), Midterm (30%), Final Exam (30%), and Client Project(30%).

Books Used in the IVMOOC



Teaches timely knowledge:

Advanced algorithms, tools, and hands-on workflows.



Teaches timeless knowledge:

Visualization framework—exemplified using generic visualization examples and pioneering visualizations.

Course Schedule

Part 1: Theory and Hands-On

- **Session 1** – Workflow Design and Visualization Framework
- **Session 2** – "When:" Temporal Data
- **Session 3** – "Where:" Geospatial Data
- **Session 4** – "What:" Topical Data

Mid-Term

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

Final Exam

Part 2: Students work in teams on client projects.

Final grade is based on Class Participation (10%), Midterm (30%), Final Exam (30%), and Client Project(30%).

Course Schedule

Part 1: Theory and Hands-On

- **Session 1** – **Workflow Design** and Visualization Framework
- **Session 2** – "When:" Temporal Data
- **Session 3** – "Where:" Geospatial Data
- **Session 4** – "What:" Topical Data

Mid-Term

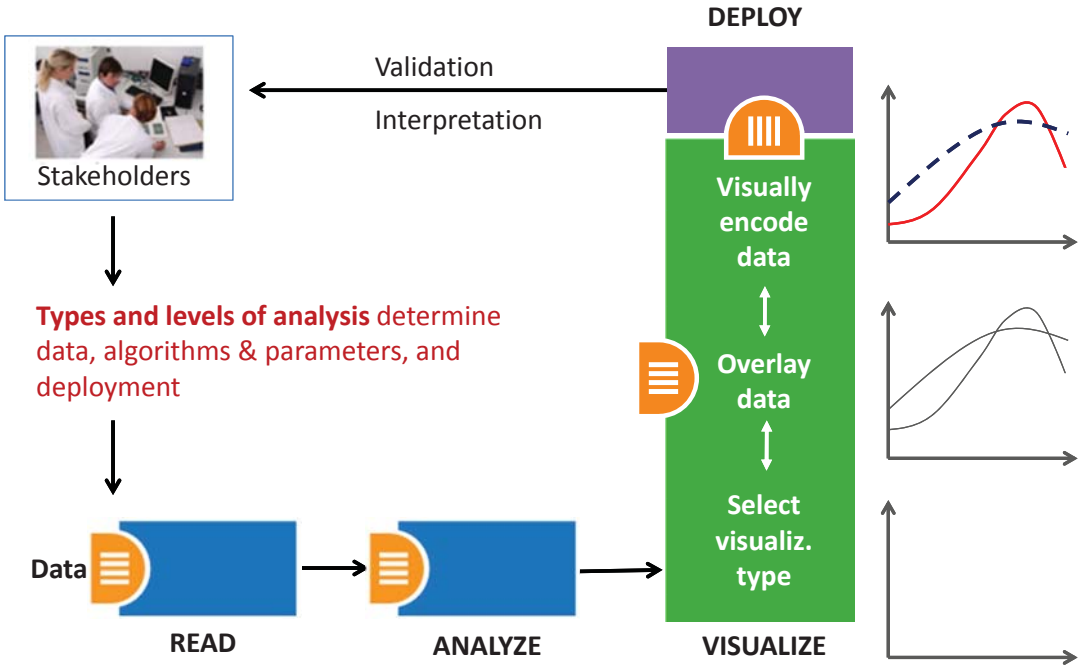
- **Session 5** – "With Whom:" Trees
- **Session 6** – "With Whom:" Networks
- **Session 7** – Dynamic Visualizations and Deployment

Final Exam

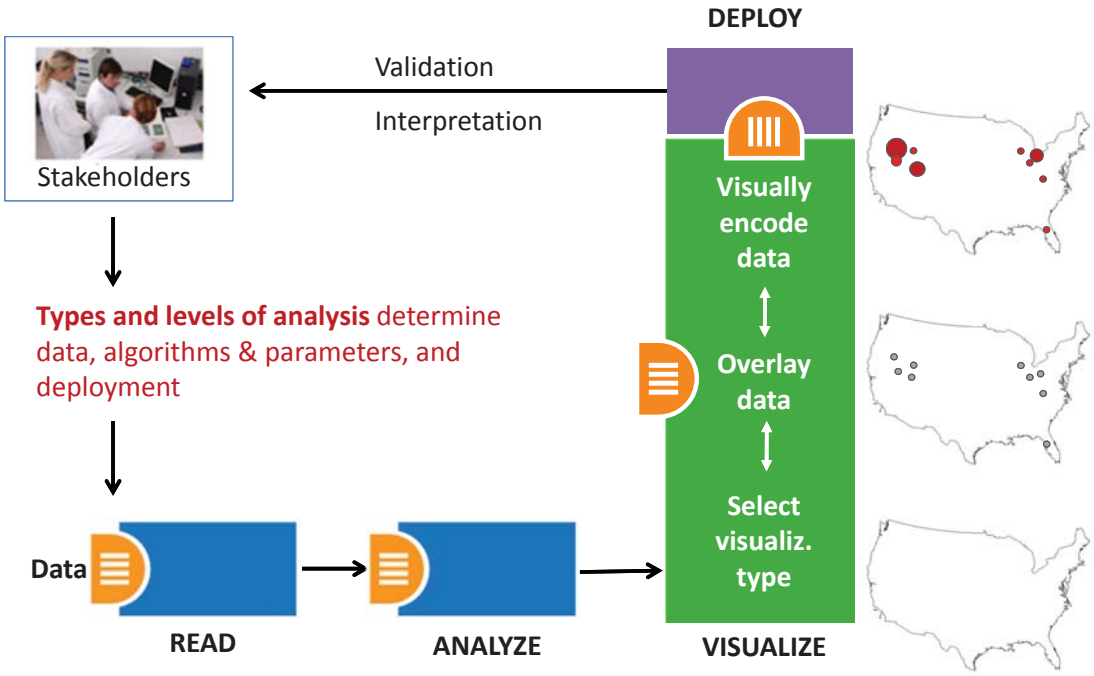
Part 2: Students work in teams on client projects.

Final grade is based on Class Participation (10%), Midterm (30%), Final Exam (30%), and Client Project(30%).

Needs-Driven Workflow Design

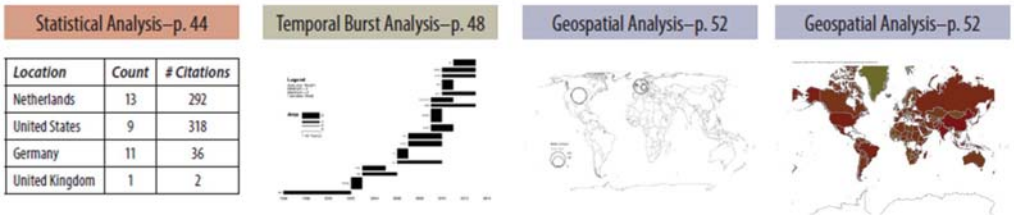


Needs-Driven Workflow Design



Load **One** File and Run **Many** Analyses and Visualizations

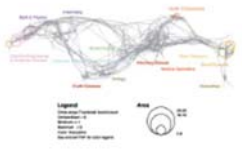
Times Cited	Publication Year	City of Publisher	Country	Journal Title (Full)	Title	Subject Category	Authors
12	2011	NEW YORK	USA	COMMUNICATIONS OF THE ACM	Plug-and-Play Macroscopes	Computer Science	Borner, K
18	2010	MALDEN	USA	CTS-CLINICAL AND TRANSLATIONAL SCIENCE	Advancing the Science of Team Science	Research & Experimental Medicine	Falk-Krzesinski, HJ|Borner, K|Contractor, N|Fiore, SM|Hall, KL|Keyton, J|Spring, B|Stokols, D|Trochim, W|Uzzi, B
13	2010	WASHINGTON	USA	SCIENCE TRANSLATIONAL MEDICINE	A Multi-Level Systems Perspective for the Science of Team Science	Cell Biology |Research & Experimental Medicine	Borner, K|Contractor, N|Falk-Krzesinski, HJ|Fiore, SM|Hall, KL|Keyton, J|Spring, B|Stokols, D|Trochim, W|Uzzi, B



Load **One** File and Run **Many** Analyses and Visualizations

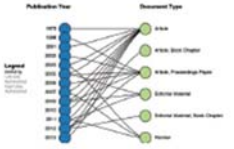
Times Cited	Publication Year	City of Publisher	Country	Journal Title (Full)	Title	Subject Category	Authors
12	2011	NEW YORK	USA	COMMUNICATIONS OF THE ACM	Plug-and-Play Macroscopes	Computer Science	Borner, K
18	2010	MALDEN	USA	CTS-CLINICAL AND TRANSLATIONAL SCIENCE	Advancing the Science of Team Science	Research & Experimental Medicine	Falk-Krzesinski, HJ|Borner, K|Contractor, N|Fiore, SM|Hall, KL|Keyton, J|Spring, B|Stokols, D|Trochim, W|Uzzi, B
13	2010	WASHINGTON	USA	SCIENCE TRANSLATIONAL MEDICINE	A Multi-Level Systems Perspective for the Science of Team Science	Cell Biology |Research & Experimental Medicine	Borner, K|Contractor, N|Falk-Krzesinski, HJ|Fiore, SM|Hall, KL|Keyton, J|Spring, B|Stokols, D|Trochim, W|Uzzi, B

Topical Analysis–p. 56



Paper Citation Network–p. 60

Bi-Modal Network–p. 60



Co-author and many other bi-modal networks.

Course Schedule

Part 1: Theory and Hands-On

- **Session 1** – Workflow Design and **Visualization Framework**
- **Session 2** – "When:" Temporal Data
- **Session 3** – "Where:" Geospatial Data
- **Session 4** – "What:" Topical Data

Mid-Term

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

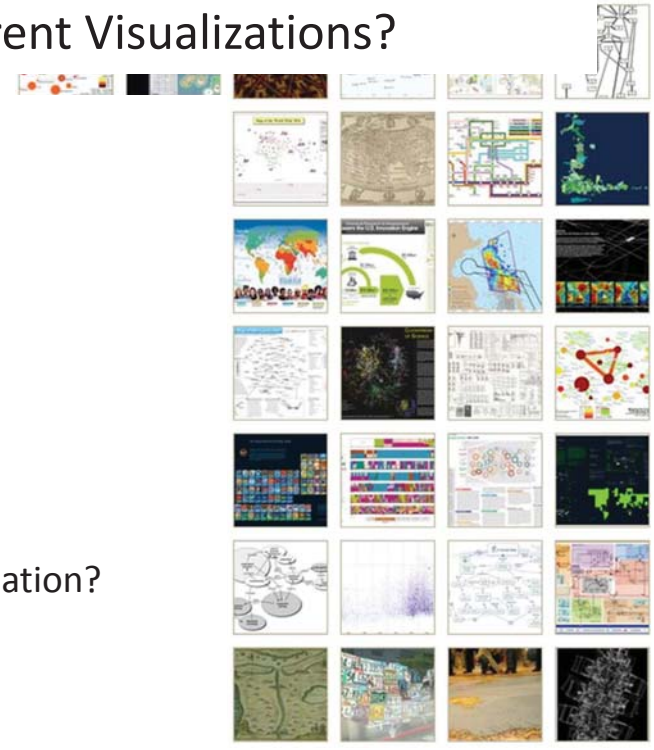
Final Exam

Part 2: Students work in teams on client projects.

Final grade is based on Class Participation (10%), Midterm (30%), Final Exam (30%), and Client Project(30%).

Visualization Frameworks

How to Classify Different Visualizations?



By

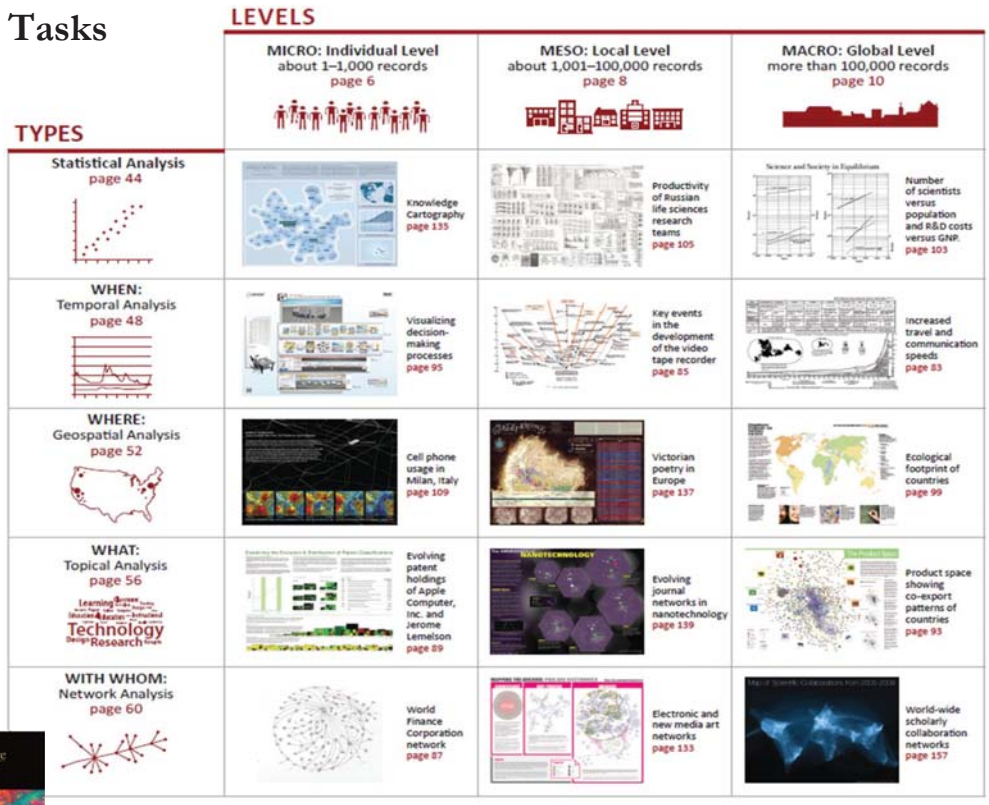
- User insight needs?
- User task types?

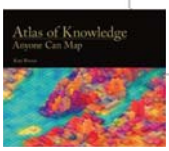
- Data to be visualized?
- Data transformation?

- Visualization technique?
- Visual mapping transformation?
- Interaction techniques?

- Or ?

Tasks



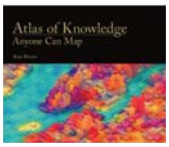


See page 5

Insight Need Types page 26	Data Scale Types page 28	Visualization Types page 30	Graphic Symbol Types page 32	Graphic Variable Types page 34	Interaction Types page 26
• categorize/cluster • order/rank/sort • distributions (also outliers, gaps) • comparisons • trends (process and time) • geospatial • compositions (also of text) • correlations/relationships	• nominal • ordinal • interval • ratio	• table • chart • graph • map • network layout	• geometric symbols point line area surface volume • linguistic symbols text numerals punctuation marks • pictorial symbols images icons statistical glyphs	• spatial position • retinal form color optics motion	• overview • zoom • search and locate • filter • details-on-demand • history • extract • link and brush • projection • distortion

See page 24

Basic Task Types								
Bertin, 1967	Wehrend & Lewis, 1996	Few, 2004	Yau, 2011	Rendgen & Wiedemann, 2012	Frankel, 2012	Tool: Many Eyes	Tool: Chart Chooser	Börner, 2014
selection	categorize			category				categorize/ cluster
order	rank	ranking					table	order/rank/ sort
	distribution	distribution					distribution	distributions (also outliers, gaps)
	compare	nominal comparison & deviation	differences		compare and contrast	compare data values	comparison	comparisons
		time series	patterns over time	time	process and time	track rises and falls over time	trend	trends (process and time)
		geospatial	spatial relations	location		generate maps		geospatial
quantity		part-to-whole	proportions		form and structure	see parts of whole, analyze text	composition	compositions (also of text)
association	correlate	correlation	relationships	hierarchy		relations between data points	relationship	correlations/ relationships



See page 26

Insight Need Types page 26	Data Scale Types page 28	Visualization Types page 30	Graphic Symbol Types page 32	Graphic Variable Types page 34	Interaction Types page 26
• categorize/cluster • order/rank/sort • distributions (also outliers, gaps) • comparisons • trends (process and time) • geospatial • compositions (also of text) • correlations/relationships	• nominal • ordinal • interval • ratio	• table • chart • graph • map • network layout	• geometric symbols point line area surface volume • linguistic symbols text numerals punctuation marks • pictorial symbols images icons statistical glyphs	• spatial position • retinal form color optics motion	• overview • zoom • search and locate • filter • details-on-demand • history • extract • link and brush • projection • distortion

See page 24

Visualization Types (Reference Systems)

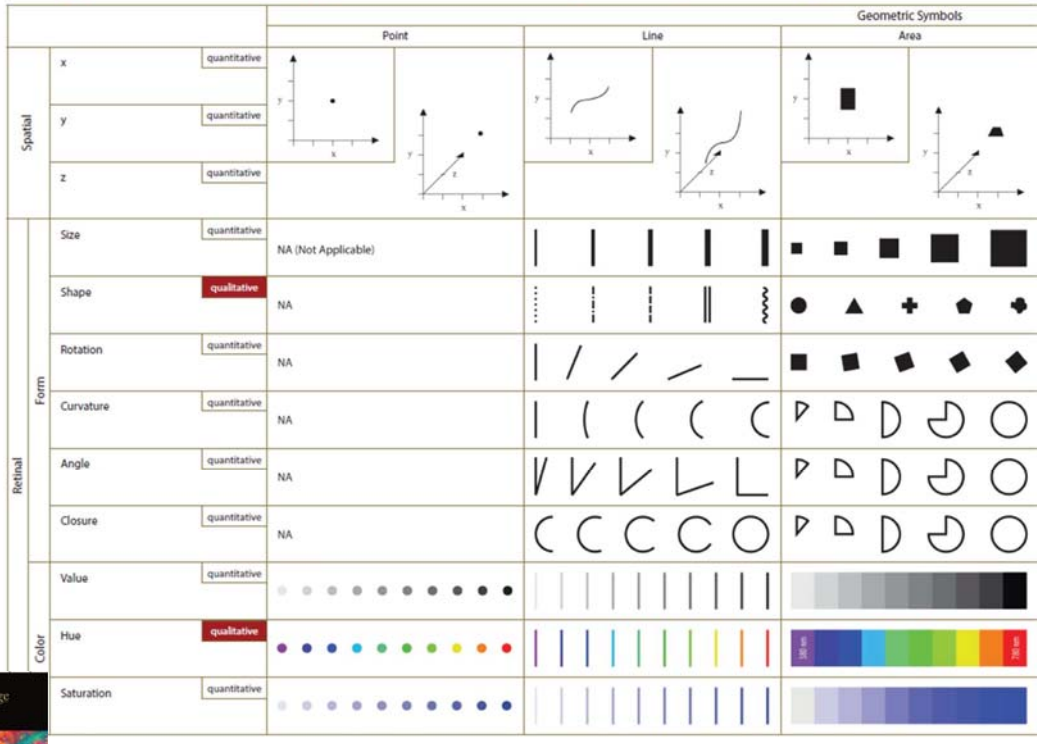
1. **Charts:** No reference system—e.g., Wordle.com, pie charts
2. **Tables:** Categorical axes that can be selected, reordered; cells can be color coded and might contain proportional symbols. Special kind of graph.
3. **Graphs:** Quantitative or qualitative (categorical) axes. Timelines, bar graphs, scatter plots.
4. **Geospatial maps:** Use latitude and longitude reference system. World or city maps.
5. **Network layouts:** Node position might depends on node attributes or node similarity. **Trees:** hierarchies, taxonomies, genealogies. **Networks:** social networks, migration flows.

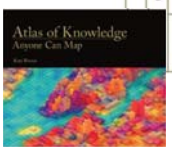
Types

Insight Need Types page 26	Data Scale Types page 28	Visualization Types page 30	Graphic Symbol Types page 32	Graphic Variable Types page 34	Interaction Types page 26
• categorize/cluster • order/rank/sort • distributions (also outliers, gaps) • comparisons • trends (process and time) • geospatial • compositions (also of text) • correlations/relationships	• nominal • ordinal • interval • ratio	• table • chart • graph • map • network layout	• geometric symbols point line area surface volume • linguistic symbols text numerals punctuation marks • pictorial symbols images icons statistical glyphs	• spatial position • retinal form color optics motion	• overview • zoom • search and locate • filter • details-on-demand • history • extract • link and brush • projection • distortion

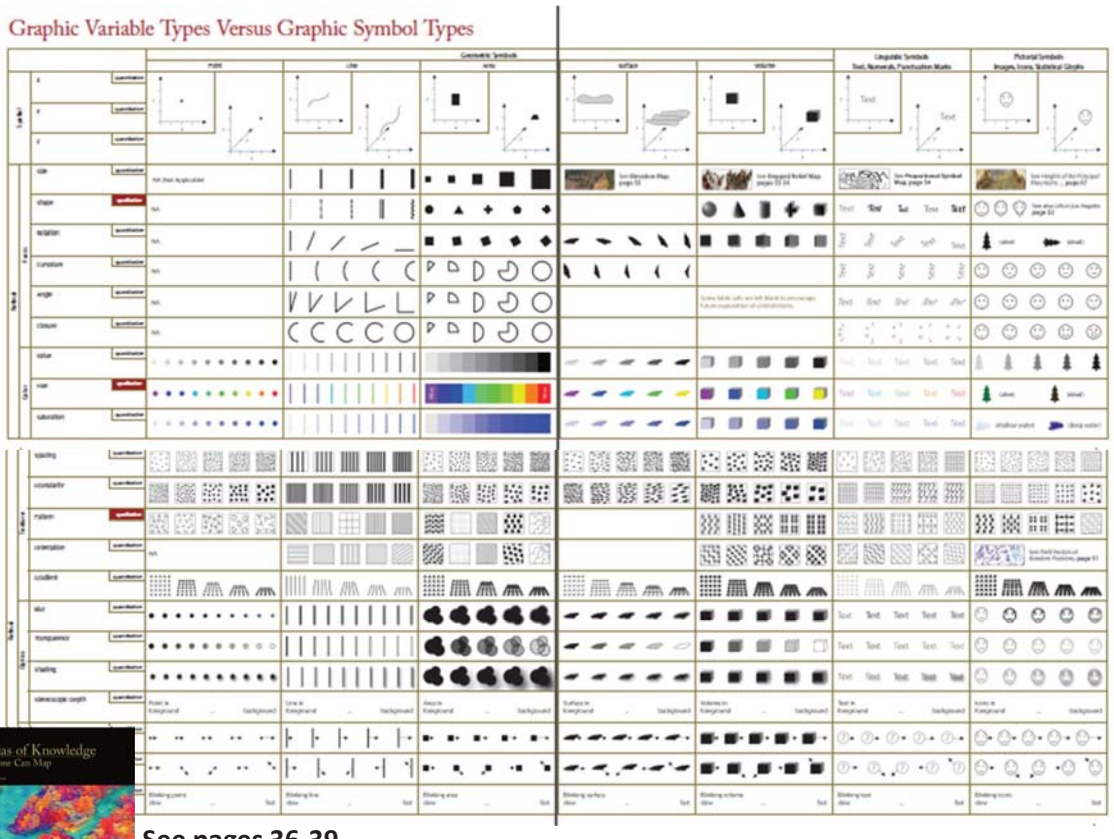
See page 25

Graphic Variable Types Versus Graphic Symbol Types





See page 36



See pages 36-39

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/

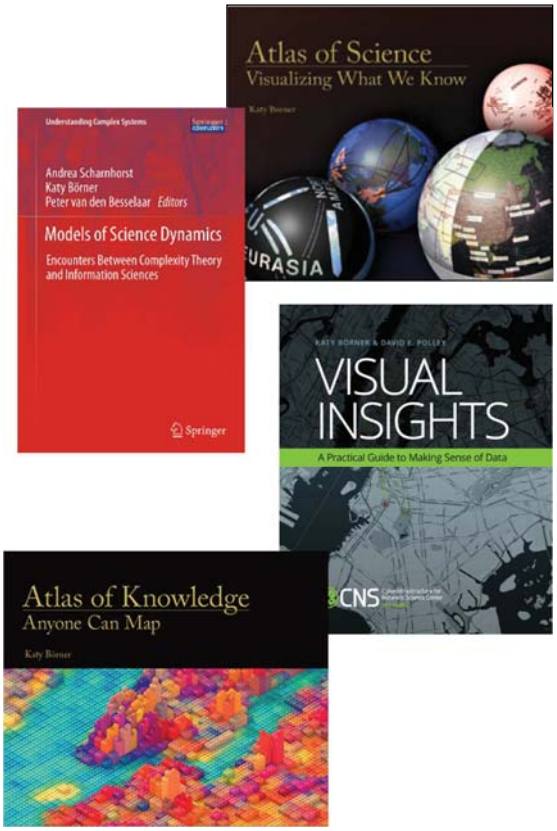
Börner, Katy (2010) **Atlas of Science: Visualizing What We Know**. The MIT Press. http://scimaps.org/atlas

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



All papers, maps, tools, talks, press are linked from http://cns.iu.edu
These slides are at http://cns.iu.edu/docs/presentations

CNS Facebook: http://www.facebook.com/cnscenter
Mapping Science Exhibit Facebook: http://www.facebook.com/mappingscience