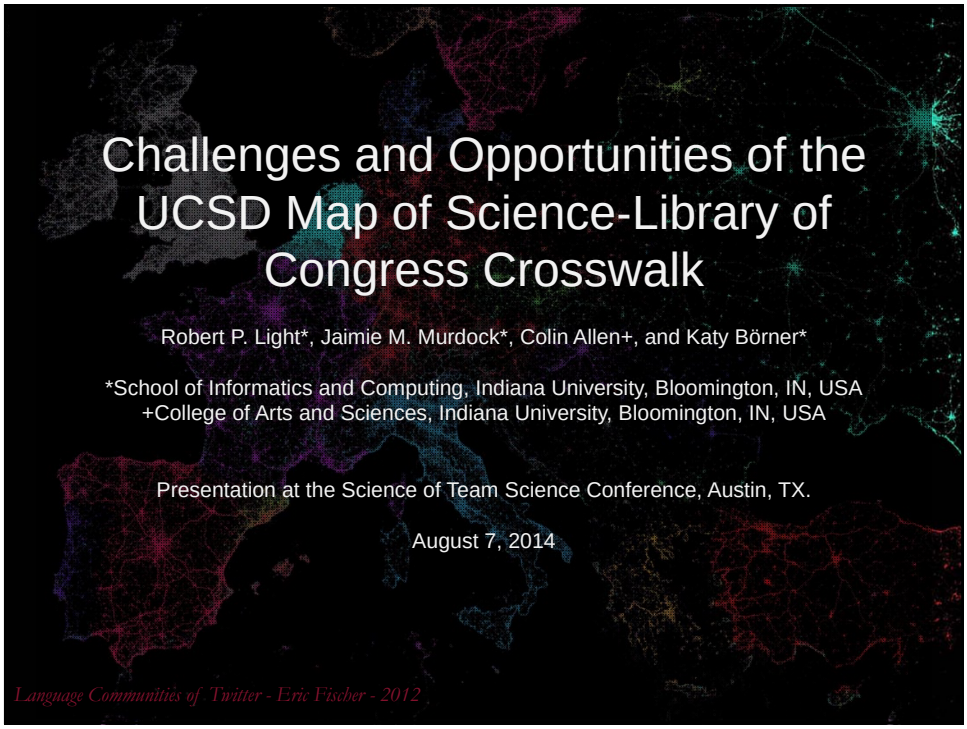


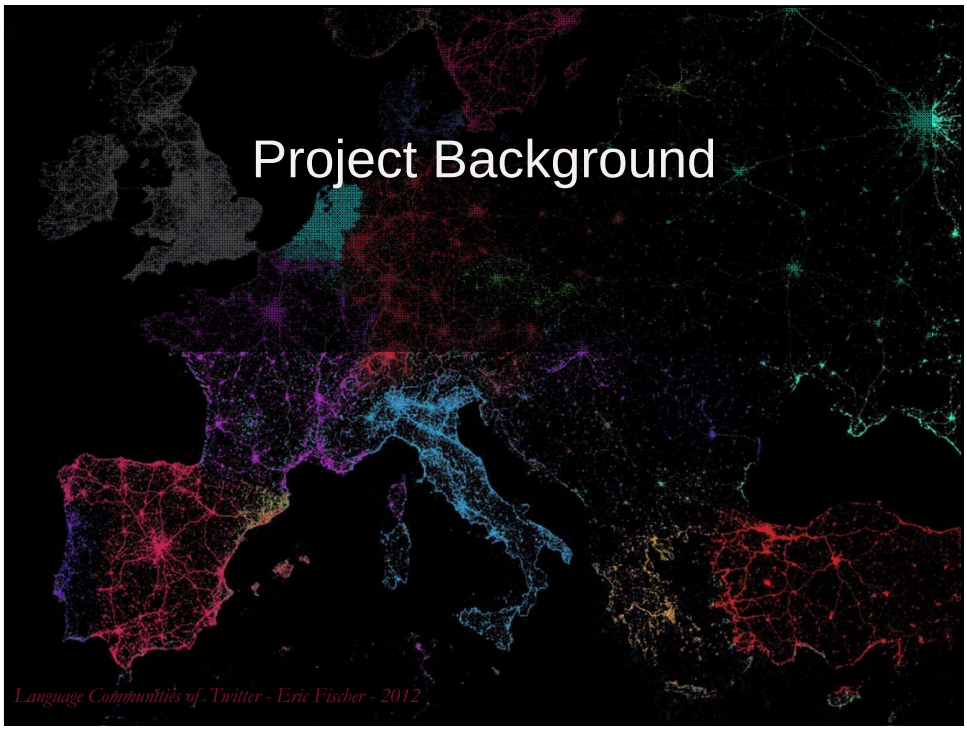
Challenges and Opportunities of the UCSD Map of Science-Library of Congress Crosswalk

Robert P. Light*, Jaimie M. Murdock*, Colin Allen+, and Katy Börner*

*School of Informatics and Computing, Indiana University, Bloomington, IN, USA
+College of Arts and Sciences, Indiana University, Bloomington, IN, USA

Presentation at the Science of Team Science Conference, Austin, TX.

August 7, 2014



Project Background

Digging by Debating (DbyD):

From Big Data Text Repositories to Argument Analysis

- **The InPhO Group**, Indiana University: Colin Allen with Robert Rose, Jaimie Murdock, Jun Otsuka, Doori Lee, Indraneel Dey
- **The Cyberinfrastructure Network Science Center**, Indiana University: Katy Börner with Robert Light
- **The International Centre for Public Pedagogy (ICPuP)**, University of East London (UEL): Andrew Ravenscroft with Simon McAlister
- **ARG:dundee**, University of Dundee: Chris Reed with John Lawrence
- **Centre for Digital Philosophy**, University of Western Ontario: David Bourget

InPhO · UEL University of East London · ARG:dundee Argumentation Research Group · CNS Cyberinfrastructure for Network Science Center cns.iu.edu · centre for digital philosophy

1

From Big Data to Argument Analysis

Linking massive datasets to specific arguments, where 'text is data'

Project Goals

- Uncover and represent the key argumentative structures of digitized documents from a large philosophy/science corpus;
- Allow users to find and interpret detailed arguments in the broad semantic landscape of books and articles, and to support innovative interdisciplinary research and better-informed critical debates

Data Sources

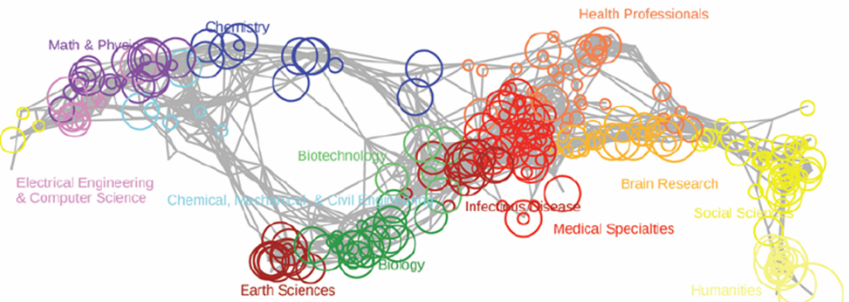
Stanford Encyclopedia of Philosophy, HathiTrust Collection, PhilPapers

4 Levels of Analysis: Macro (Sci/Phil Maps) to Micro (detailed arguments)

1. *Visualizing* points of contact between philosophy and the sciences
2. *Topic modeling* to identify the volumes/pages 'rich' in a chosen topic;
3. Identify and map *key arguments*; apply a novel analysis framework for propositions and arguments;
4. *Sentence modeling* to get back to HathiTrust materials

2

Mapping Philosophy in the Sciences
(analysis level 1)

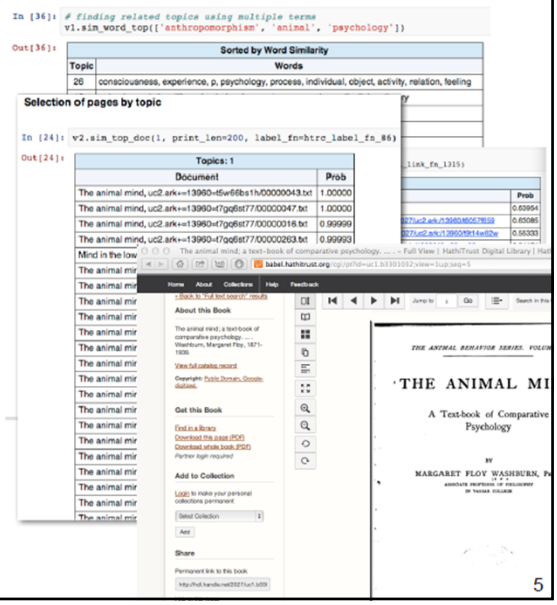


- **UCSD Map of Science:** generated using more than 12M papers and their references from Elsevier's Scopus and Thomson Reuters' Web of Science (25,000 journals), see http://sci.cns.iu.edu/ucsdmap
- Shows **554 subdiscipline nodes** aggregated into **13 color-coded disciplines**.
- Overlaid are citations made by the **Stanford Encyclopedia of Philosophy to** visualize the impact of the sciences on philosophy.
- **Node sizes** scale from 0 (no circle) to 43. Highest numbers are in **Humanities, Earth Sciences, and Math & Physics.**

3

Topic Modeling in HT Books
(analysis level 2)

- LDA Topic Modeling: Bayesian method generates set of "**topics**" – **probability distributions** over terms in the corpus
  - **Every topic contains every term** – different probabilities in the different topics
  - The **number of topics** is a user-selected parameter
- Finds the set of topics best able to reproduce term distributions in the **documents**
- Documents may be whole **volumes**, **chapters**, **articles**, single **pages**, even individual **sentences** – modelers' choice

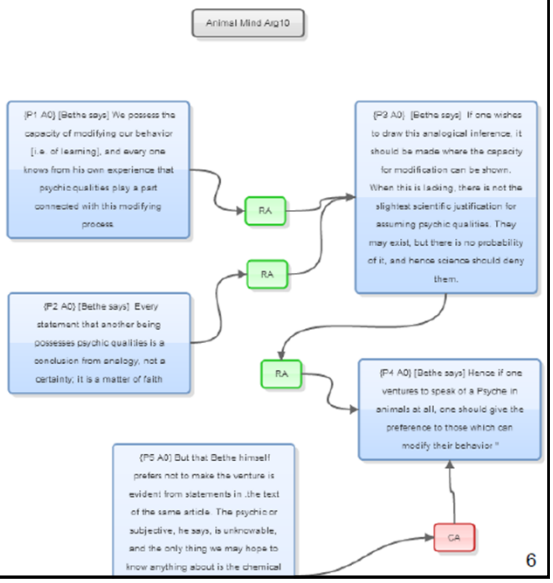


5

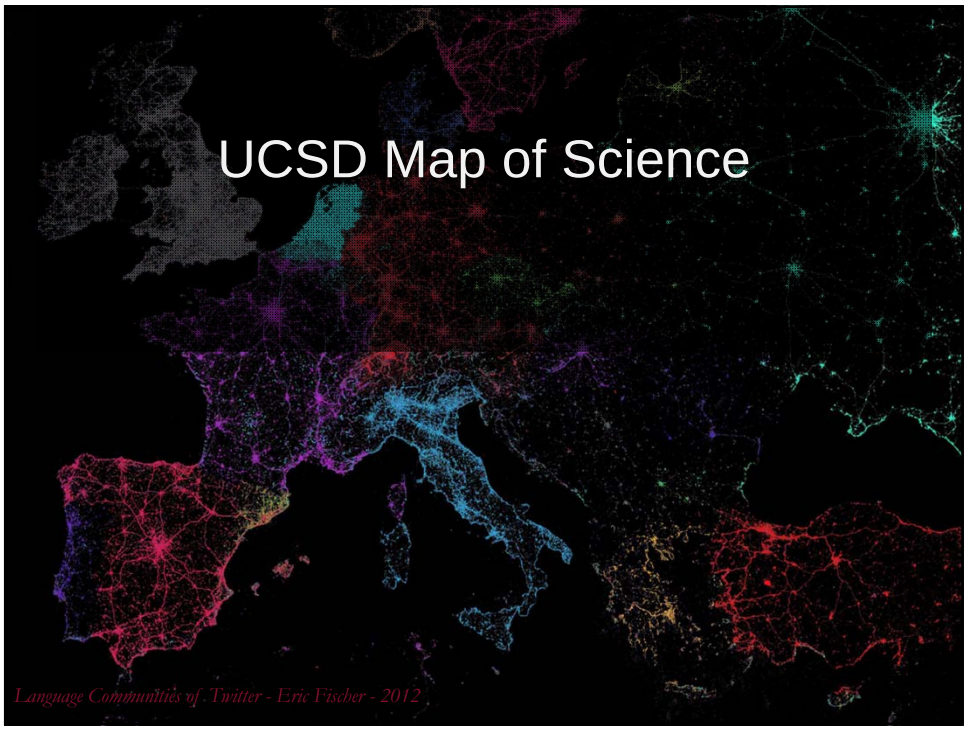
Argument Selection, Mapping and Analysis (analysis level 3)

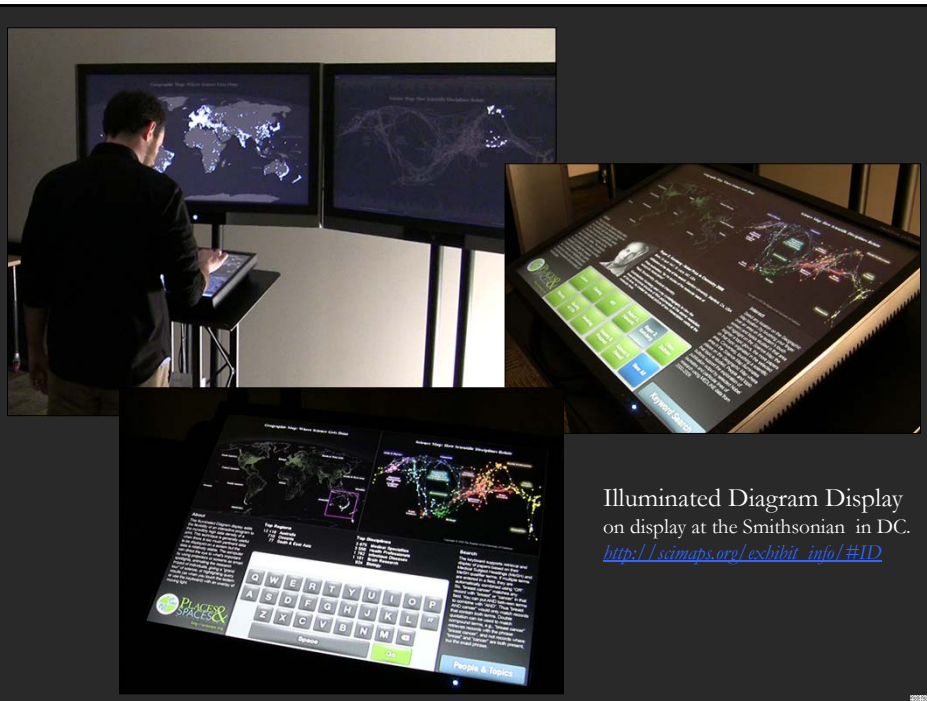
1. Identifying arguments from rated pages (currently human/ manual, but developing algs for automation)
2. Mapping of key arguments with OVA mapping tool:

- Provides a formal framework, or 'lens', for investigation and comparative analysis, e.g. role, structure, status
- More indirectly: meaning, importance and context



6



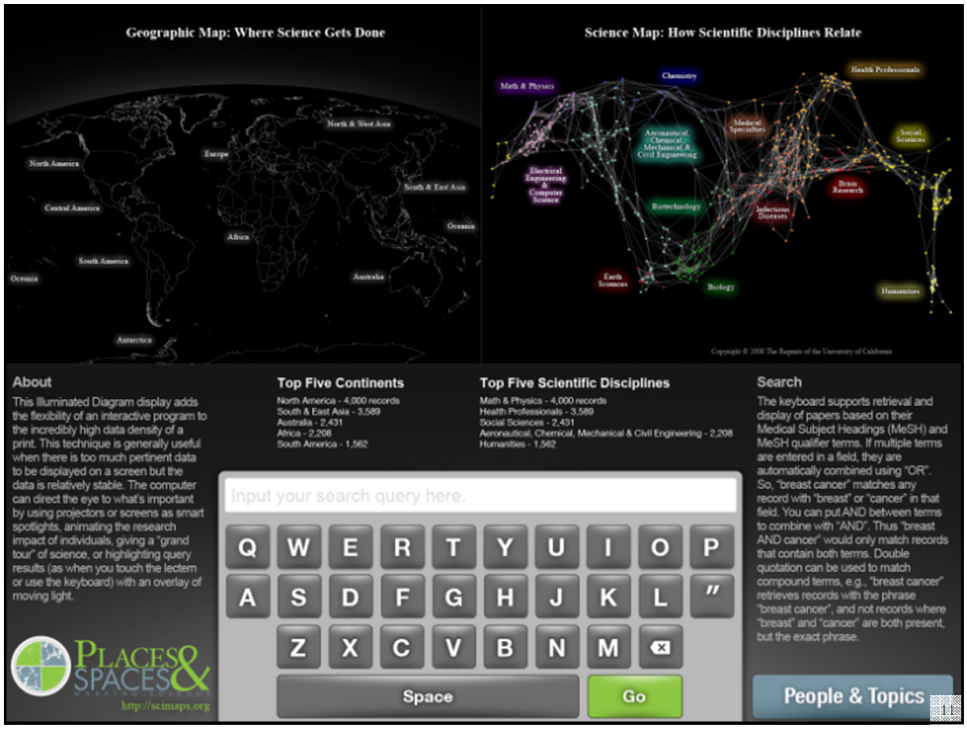


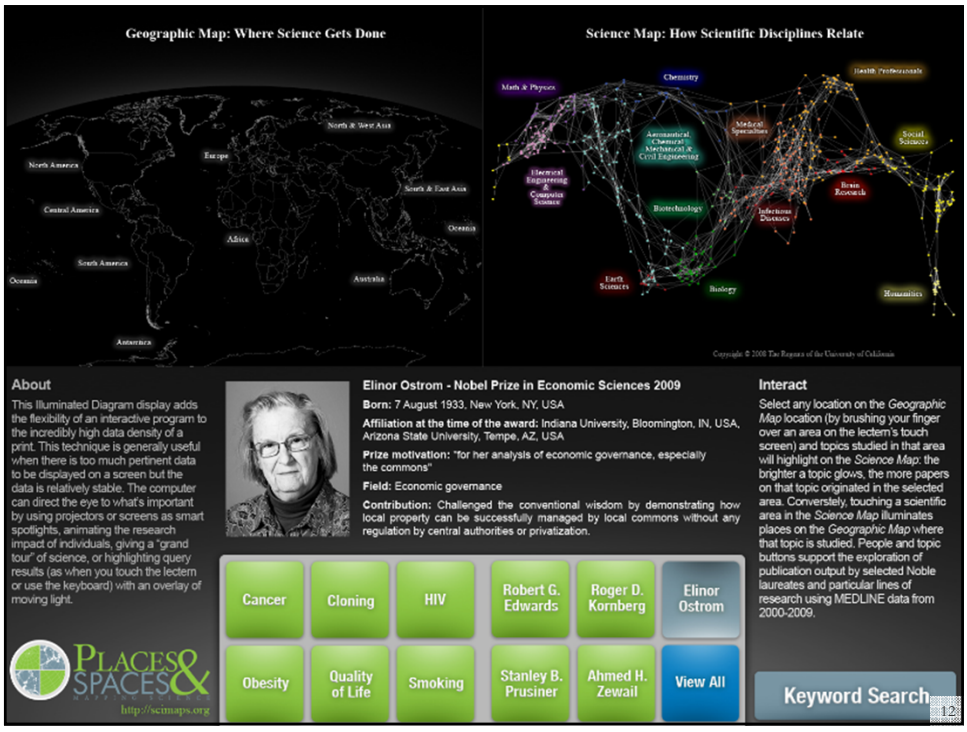
Illuminated Diagram Display
on display at the Smithsonian in DC.
http://scimaps.org/exhibit_info/#ID

9

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

10


Geographic Map: Where Science Gets Done
North & West Asia
North America
Europe
South & East Asia
Central America
Africa
Oceania
South America
Oceania
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
11


Geographic Map: Where Science Gets Done
North & West Asia
North America
Europe
South & East Asia
Central America
Africa
Oceania
South America
Oceania
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
12

Full Update (10 years of Scopus and WoS)

by Börner, Klavans, Patek, Zoss, Biberstine, Light, Larivière, Boyack

Deployment: The UCSD map of science data is available at http://sci.cns.iu.edu/ucsdmap/

Data
The 2010 UCSD map of science and classification system covering 10 years (2001-2010) of Web of Science data and 8 years (2001-2008) of Scopus data with subdiscipline assignments by SciTech Strategies.

1. Data as MS AccessDB and as MS Excel file (identical info as MS AccessDB) as well as data dictionary and database schema.
2. Network .net file to visually render science map. Also provided as .net file with discipline nodes and names.

Usage Conditions
This map is shared under the Creative Commons, Attribution-NonCommercial-ShareAlike 3.0 Unported (CC BY-NC-SA 3.0) license (http://creativecommons.org/licenses/by-nc-sa/3.0/). That is, you are free to share, e.g., to copy, distribute and transmit the work, and to remix, i.e., to adapt the work under the following conditions:

- Attribution — You must attribute the work in the following manner (but not in any way that suggests that they endorse you or your use of the work): Cite the above paper and use the following acknowledgment text: "*The authors wish to acknowledge The Regents of the University of California, SciTech Strategies, Observatoire des Sciences et des Technologies, and the Cyberinfrastructure for Network Science Center for making the 2010 UCSD Map of Science and Classification System available for this work.*"
- Noncommercial — You may not use this work for commercial purposes.

Please cite as: Börner, Katy, Richard Klavans, Michael Patek, Angela Zoss, Joseph R. Biberstine, Robert Light, Vincent Larivière, and Kevin W. Boyack (2012) Design and Update of a Classification System: The UCSD Map of Science. *PLoS ONE* 7(7): e39464. doi:10.1371/journal.pone.0039464

Full Update (10 years of Scopus and WoS)

by Börner, Klavans, Patek, Zoss, Biberstine, Light, Larivière, Boyack

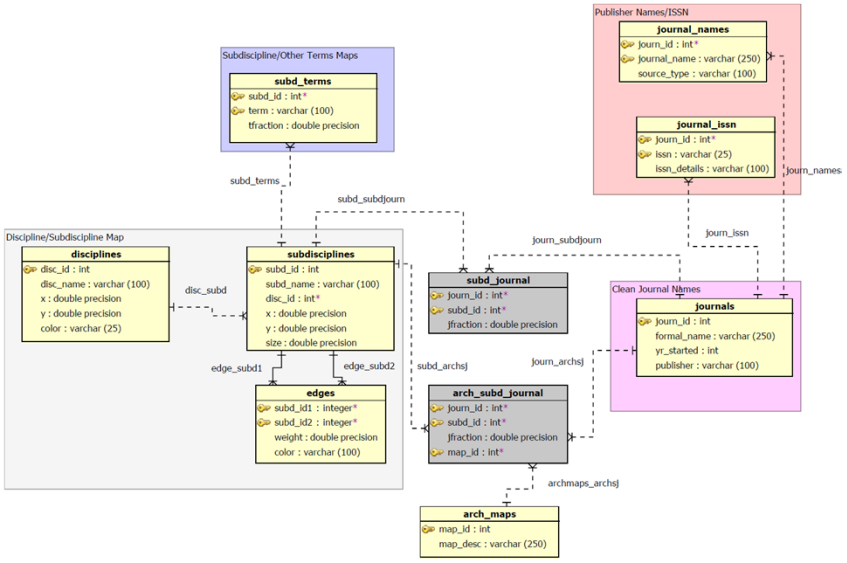
Data: The 2010 UCSD map of science and classification system covers ten years (2001-2010) of data from Thomson Reuters' Web of Science and eight years (2001-2008) of Elsevier's Scopus, specifically the fractional assignment of about 25,000 journal names to 554 subdisciplines grouped into 13 disciplines of science.

The counts for major record types are given here:

1. 13 disciplines with labels and color codes
2. 554 subdisciplines with x, y positions and size
3. 15,849 journals captured by 5-year map
4. 25,258 journals captured by 10-year map
5. 13,520 journal names used by Thomson Reuters
6. 22,253 journal names used by Scopus
7. 21,630 Scopus journal ID numbers
8. 19,988 ISSN numbers
9. 66,759 terms

See Data Dictionary in Supplement 2 in
http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0039464

Full Update (10 years of Scopus and WoS)
by Börner, Klavans, Patek, Zoss, Biberstine, Light, Larivière, Boyack

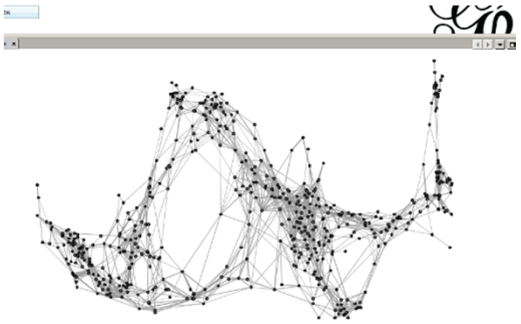


UCSD map table schema http://sci.cns.iu.edu/ucsdmap/data/UCSDmapDBSchema.pdf

Full Update (10 years of Scopus and WoS)
by Börner, Klavans, Patek, Zoss, Biberstine, Light, Larivière, Boyack

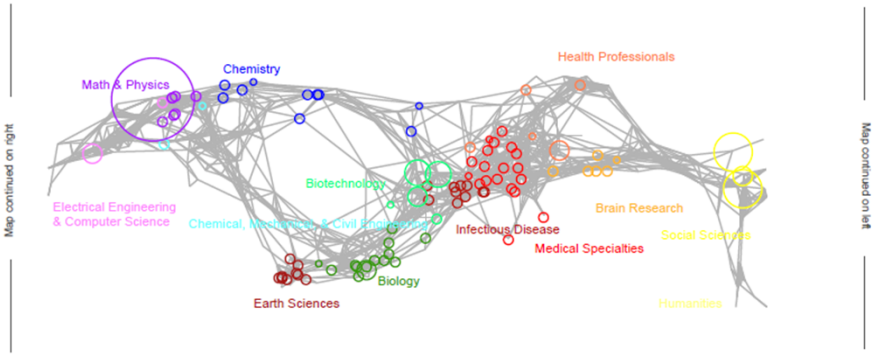
Note: There are no standards on how to render .net files!

Some define the zero point on the top left (e.g., GUESS), while others define the bottom left point as 0,0 (e.g., Gephi, Pajek). This only becomes important when rendering a dataset that has a predefined left and right, top and bottom such as the UCSD map of science.
Simply multiply all node's y-position with -1 to solve this issue.



Topical Visualization

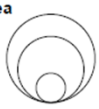
Generated from 361 Unique ISI Records
14 out of 109 records were mapped to 94 subdisciplines and 12 disciplines.
May 27, 2014 | 06:54 AM CEST



Legend

Circle area: Fractional record count
Unclassified = 95
Minimum = 0
Maximum = 25
Color: Discipline
See end of PDF for color legend.

Area



How To Read This Map

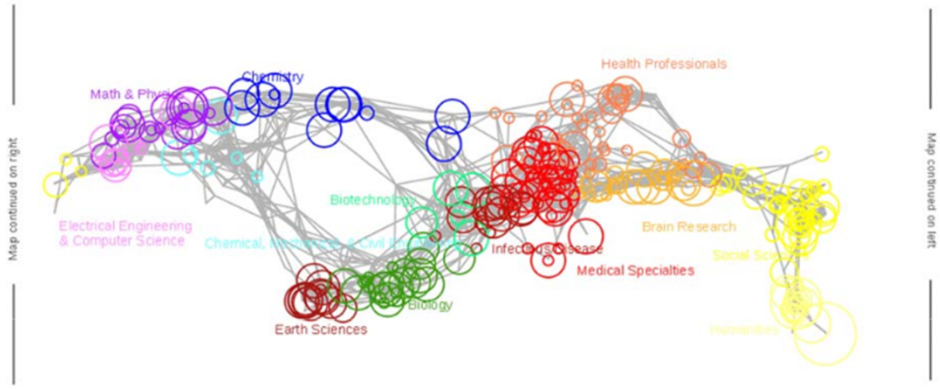
The *UCSD map of science* depicts a network of 554 subdiscipline nodes that are aggregated to 13 main disciplines of science. Each discipline has a distinct color and is labeled. Overlaid are circles, each representing all records per unique subdiscipline. Circle area is proportional to the number of fractionally assigned records. Minimum and maximum data values are given in the legend.

CNS (cns.iu.edu)

12

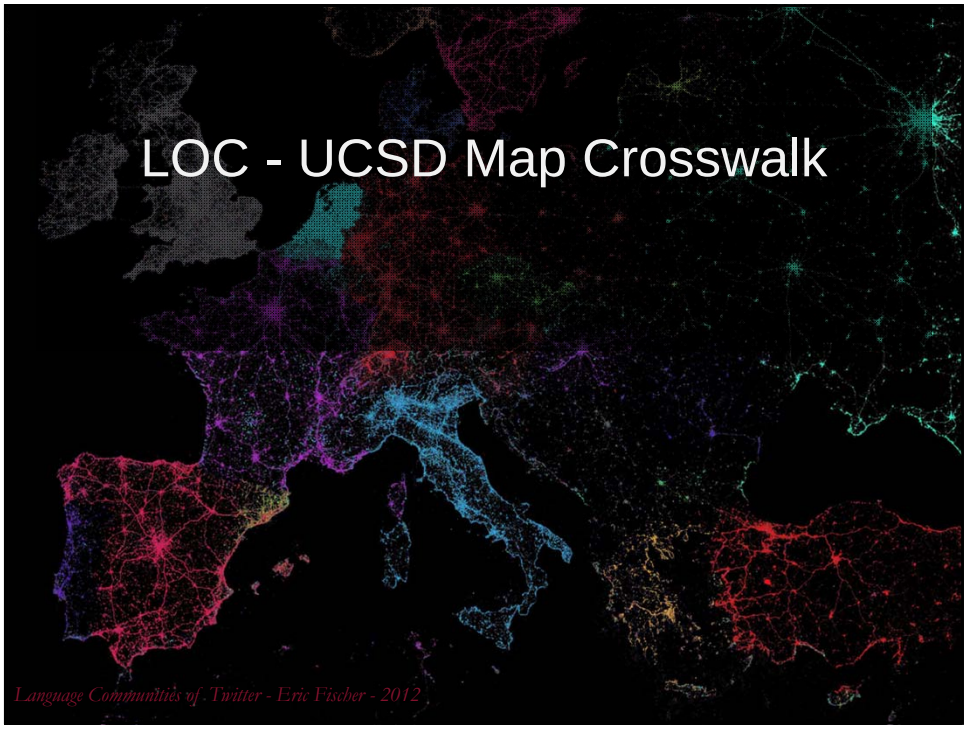
Mapping the Intersection of Science & Philosophy

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



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.

13



Goal

Empirically measure and visualize the cross-pollination of sciences and philosophy through paper citation data.

Using books from the **HathiTrust scanned books collection**, we overlay books onto the UCSD Map of Science to highlight areas of science which overlap with philosophical discussion.

Do philosophers pay more attention to biology or physics? Geology or anthropology? Scientometric and text mining methods can suggest hypotheses in the early stages of an investigation.

20

The Challenge

Numerous classification systems have been devised to manage different types of information. Examples are

- Dewey Decimal System
- Library of Congress Classification Numbers (LCCN)
- International Patent Classification

Each is crafted by different groups to cover different entities with different goals.

This creates substantial challenges when scientists from varied domains attempt to form teams and combine their resources and data.

21

One Solution

1. 2,010 journals with **LCCN** data, provided by IU's Wells Library, were mapped by title to the **UCSD Map of Science**, with the number of titles that share a LCCN and a UCSD subdiscipline giving weight to each linkage.
2. The 776 volumes from the 1,315 volume corpus of Hathi Trust materials that included LCCNs were mapped to UCSD Map of Science as follows:
   - Those volumes that shared an LCCN with some number of mapped journals were directly mapped based on the location of those journals.
   - In cases where a direct match was not possible, the algorithm iterated up the hierarchy, taking the most narrow subsection that included the unmatched numbers as well as some number of mapped journals.
   - Once a match was found, the location of the volume on the map was determined to be the center of all the LoC subsections that contributed to that mapping (one in the case of an exact match).

Example: One of the 1,315 books with the title *Evolution* has a call number of B818 (Evolution, Holism), but as no journals in the USCD map of science share that call number, it is instead mapped as B808-849 (Special topics and schools of philosophy). The location of each subsection is the average of the locations of the subdisciplines to which it maps. Carrying the previous example further, four journals map in the B808-849 range. Three of these map to Philosophy/Psychology, while the other maps to Contemporary Philosophy. Since we do not weight by the number of journals mapping to a specific subdiscipline, the location for B808-849 and any journals that map to that location is the midpoint between the points for Philosophy/Psychology and Contemporary Philosophy.

22

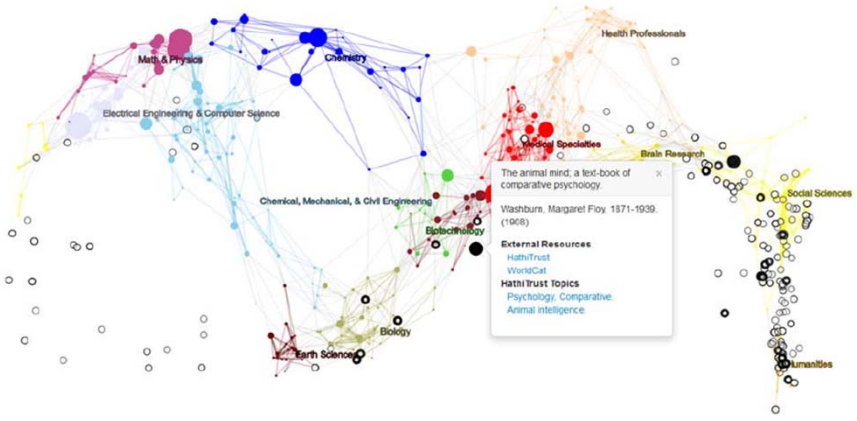


Figure 1: Draft mapping of Animal Psychology volumes onto UCSD Map of Science

Overlay of volumes in the InPhO's HTRC 1,315 corpus on *anthropomorphism*. There are currently 776 volumes on the map due to incomplete HTRC metadata. Books belonging to selected subsets are indicated with darker circles as shown in the key, bottom left. Hovering over the black dot displays the title of the volume, and clicking the dot will link to the volume in the HTRC. The scroll wheel can be used to zoom in and out. https://inpho.cogs.indiana.edu/scimap/scits

23

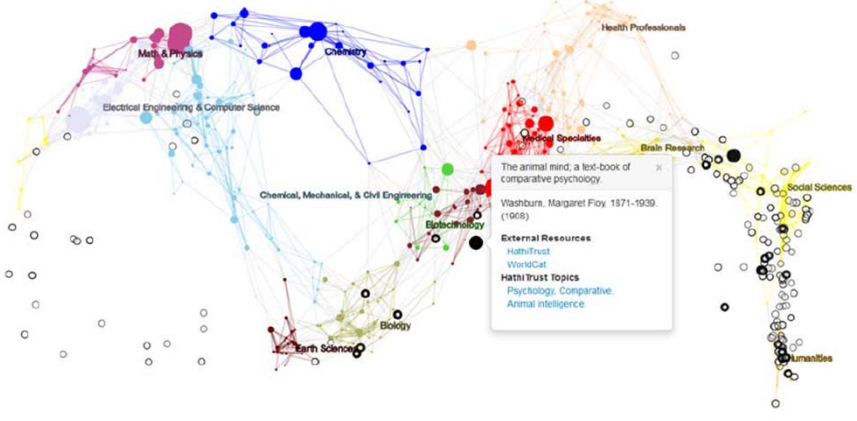
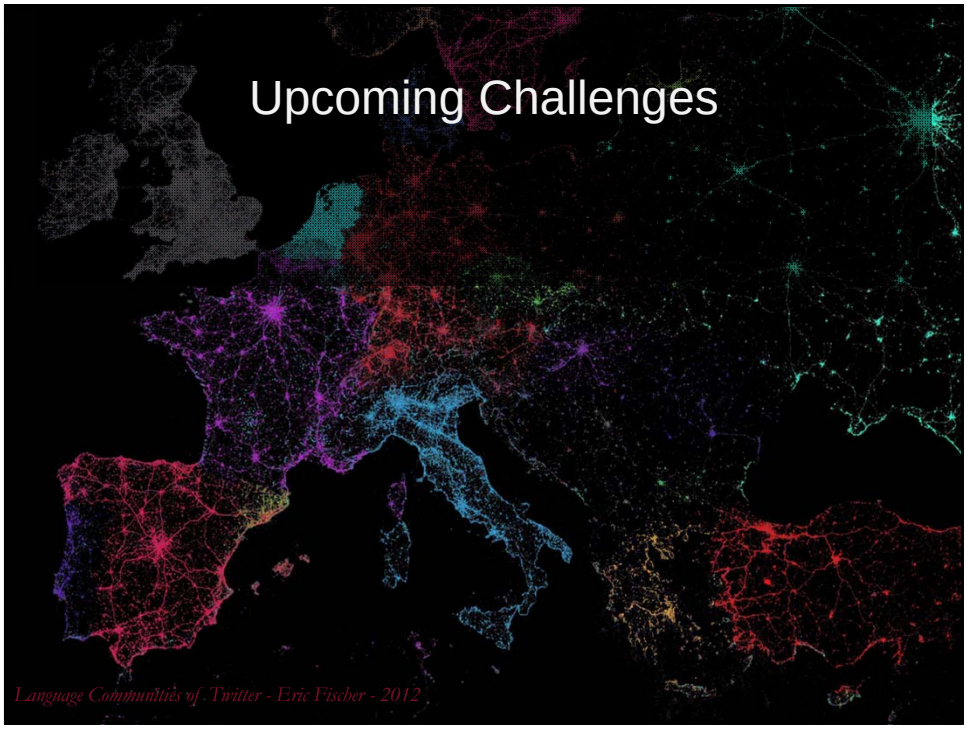


Figure 1: Draft mapping of Animal Psychology volumes onto UCSD Map of Science

4 Levels of Analysis: Macro (Sci/Phil Maps) to Micro (detailed arguments)

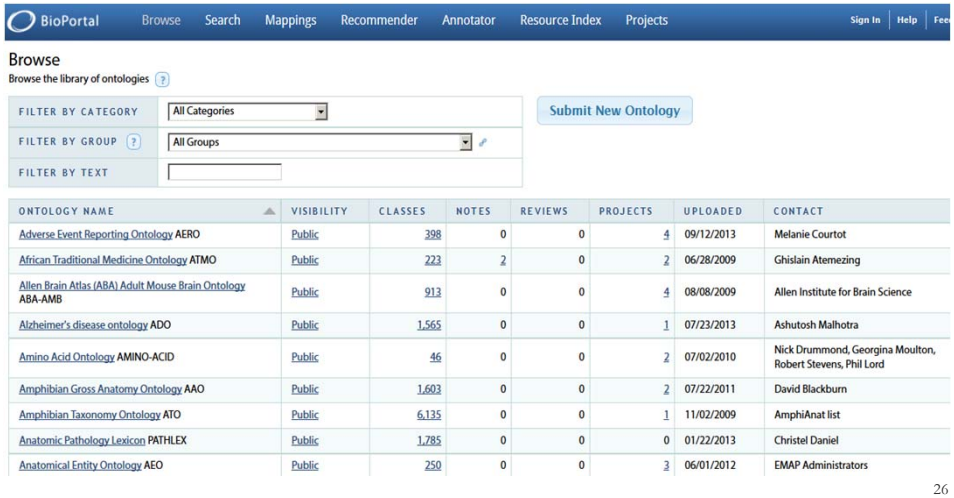
1. *Visualizing* points of contact between philosophy and the sciences
2. *Topic modeling* to identify the volumes/pages 'rich' in a chosen topic;
3. Identify and map *key arguments*; apply a novel analysis framework for propositions and arguments;
4. *Sentence modeling* to get back to HathiTrust materials

24



Need Many More Crosswalks!

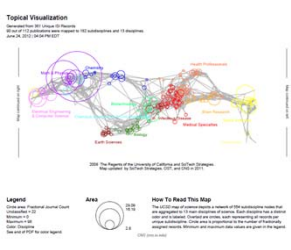
Can learn/benefit from other efforts, e.g., 378 aligned ontologies
http://bioportal.bioontology.org/ontologies



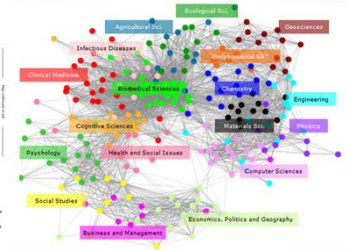
ONTOLOGY NAME	VISIBILITY	CLASSES	NOTES	REVIEWS	PROJECTS	UPLOADED	CONTACT
Adverse Event Reporting Ontology AERO	Public	398	0	0	4	09/12/2013	Melanie Courtot
African Traditional Medicine Ontology ATMO	Public	223	2	0	2	06/28/2009	Ghislain Atemezing
Allen Brain Atlas (ABA) Adult Mouse Brain Ontology ABA-AMB	Public	913	0	0	4	08/08/2009	Allen Institute for Brain Science
Alzheimer's disease ontology ADO	Public	1,565	0	0	1	07/23/2013	Ashutosh Malhotra
Amino Acid Ontology AMINO-ACID	Public	46	0	0	2	07/02/2010	Nick Drummond, Georgina Moulton, Robert Stevens, Phil Lord
Amphibian Gross Anatomy Ontology AAO	Public	1,603	0	0	2	07/22/2011	David Blackburn
Amphibian Taxonomy Ontology ATO	Public	6,135	0	0	1	11/02/2009	AmphiAnat list
Anatomic Pathology Lexicon PATHLEX	Public	1,785	0	0	0	01/22/2013	Christel Daniel
Anatomical Entity Ontology AEO	Public	250	0	0	3	06/01/2012	EMAP Administrators

26

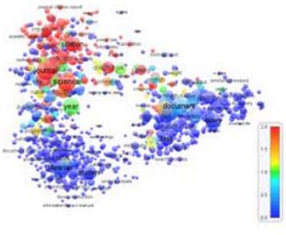
Need Advise on What Science Basemap to Use



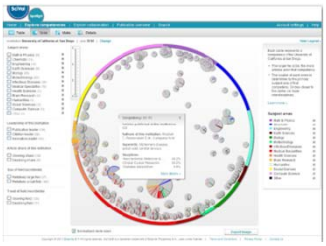
UCSD Map



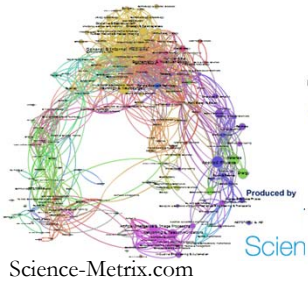
Loet et al science maps ISI categories



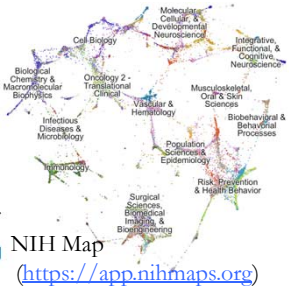
http://vosviewer.com



Elsevier's SciVal Map



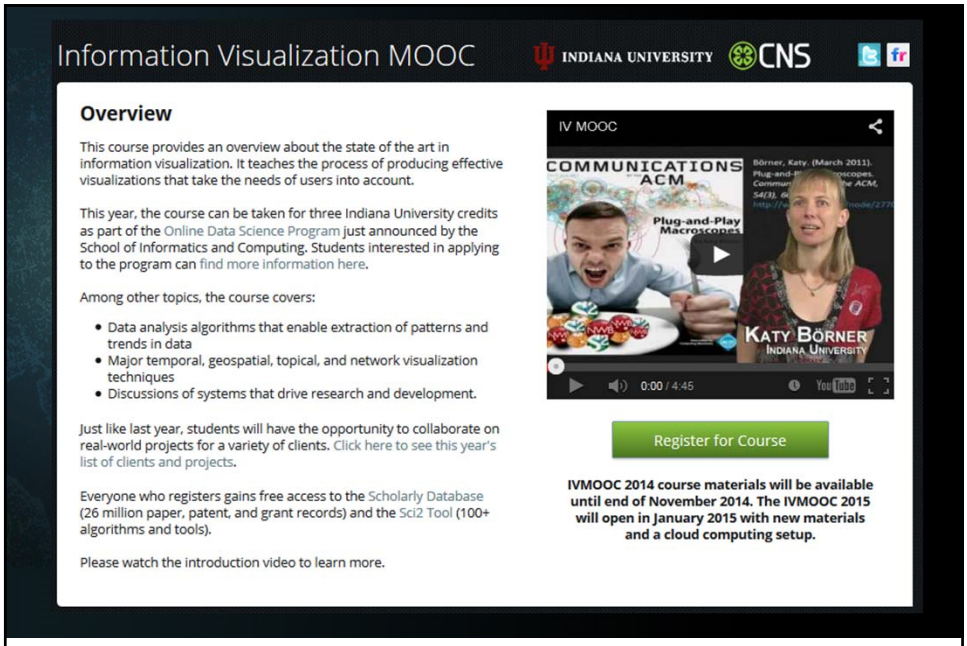
Science-Metrix.com



NIH Map (https://app.nihmaps.org)

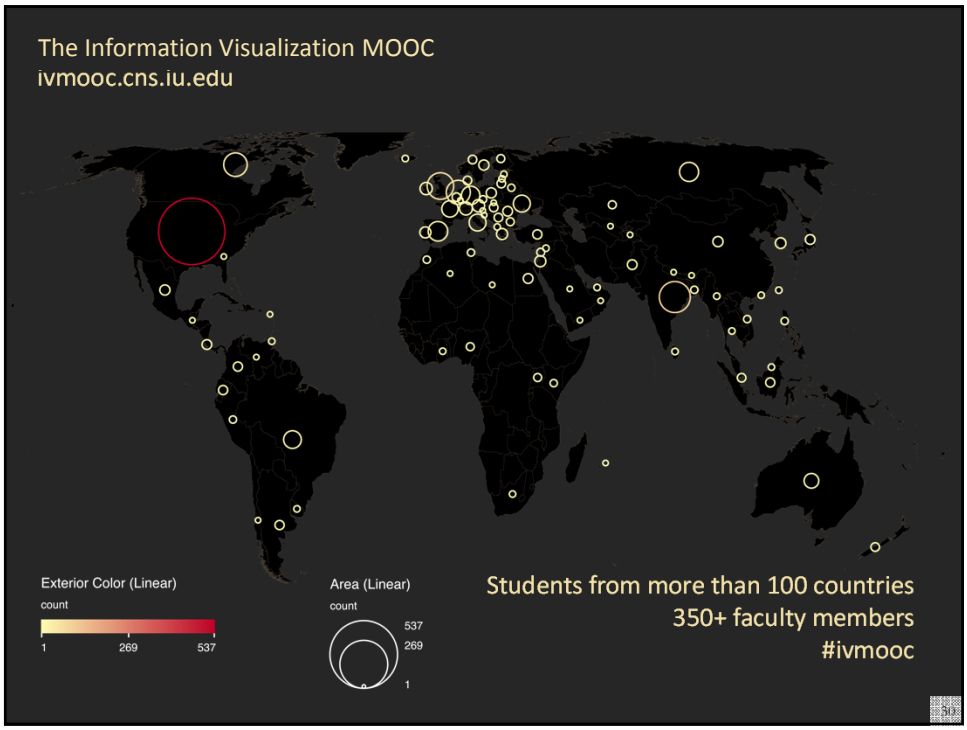
27

Upcoming Opportunities



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

29



30

CNS Cyberinfrastructure for Network Science Center

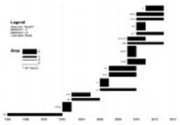
Load **One** File and Run **Many** Analyses and Visualizations

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

Statistical Analysis–p. 44

Location	Count	# Citations
Netherlands	13	292
United States	9	318
Germany	11	36
United Kingdom	1	2

Temporal Burst Analysis–p. 48



Geospatial Analysis–p. 52

Geospatial Analysis–p. 52

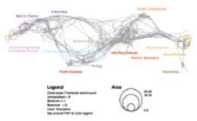
31

CNS Cyberinfrastructure for Network Science Center

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

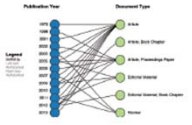
Times Cited	Publication Year	City of Publisher	Country	Journal Title (Full)	Title	Subject Category	Authors
12	2011	NEW YORK	USA	COMMUNICATIONS OF THE ACM	Plug-and-Play Macroscopes	Computer Science	Borner, K
18	2010	MALDEN	USA	CTS-CLINICAL AND TRANSLATIONAL SCIENCE	Advancing the Science of Team Science	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.

32

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.

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**. MIT Press.

33

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

34