Bibliometrics & Research Assessment

A SYMPOSIUM FOR LIBRARIANS & INFORMATION PROFESSIONALS

presented by NIH Library nihlibrary.nih.gov

SLA

(Network) Data Visualization Literacy

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

Bibliometrics & Research Assessment: A Symposium for Librarians & Information Professionals

NIH Library in Building 10, Bethesda, MD
Live stream is at http://videocast.nih.gov

October 31, 2016

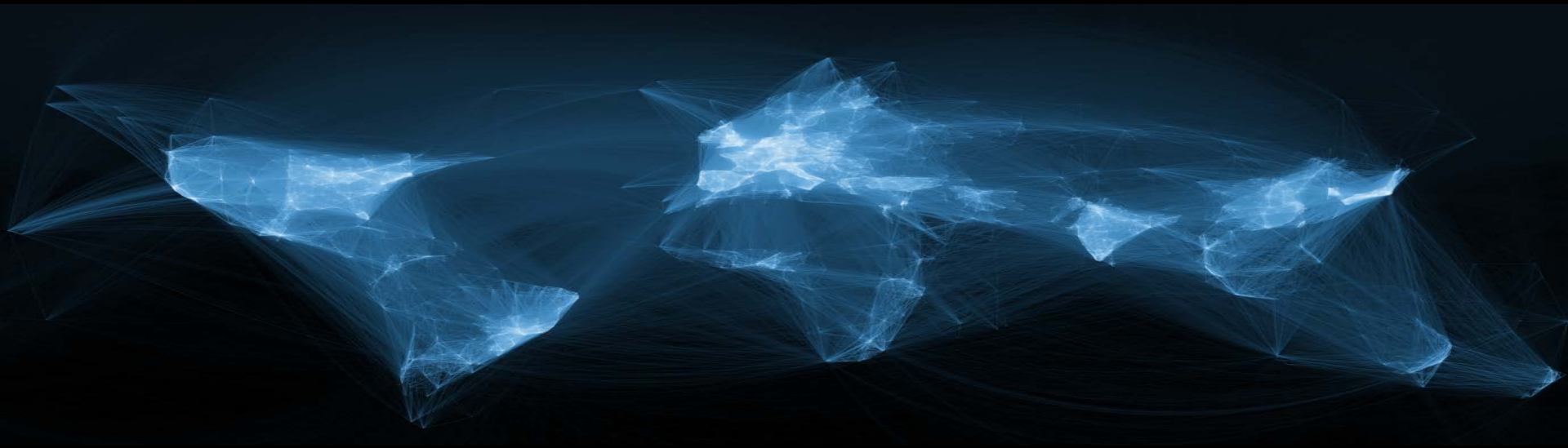
Bibliometrics & Research Assessment

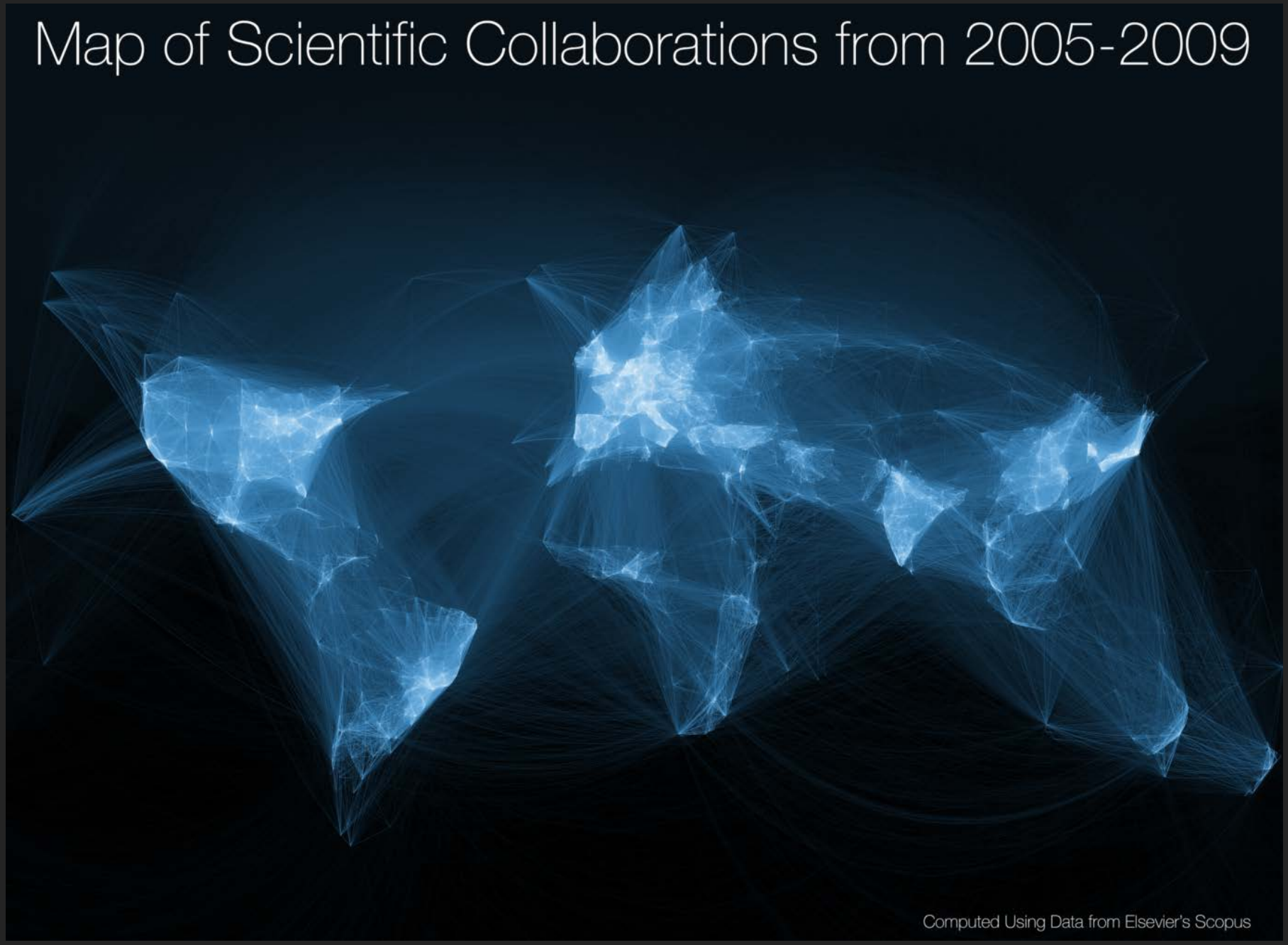
A SYMPOSIUM FOR LIBRARIANS & INFORMATION PROFESSIONALS

presented by NIH Library nihlibrary.nih.gov

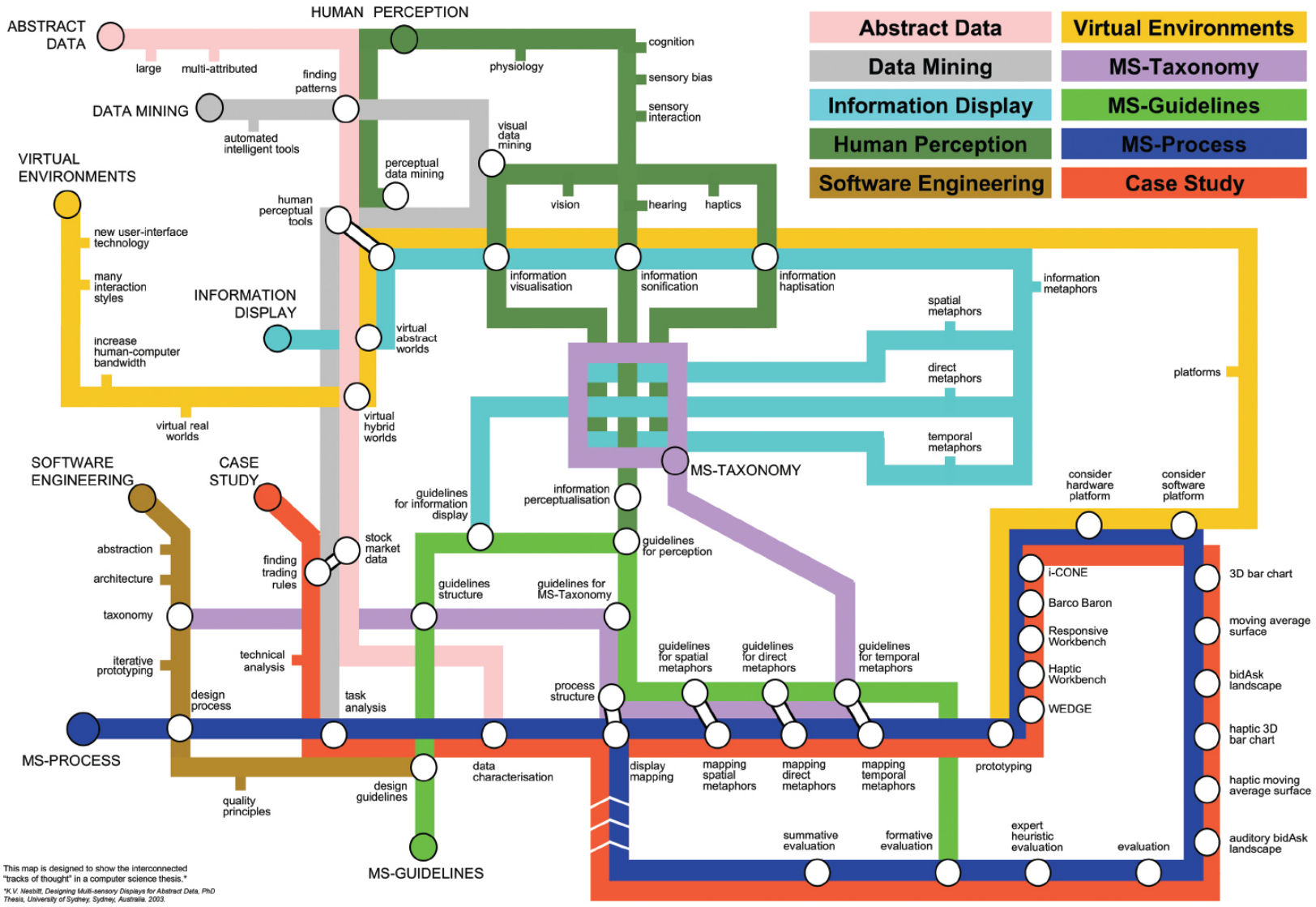
SLA Connecting People and Information Maryland Chapter

Visualizing Networks





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



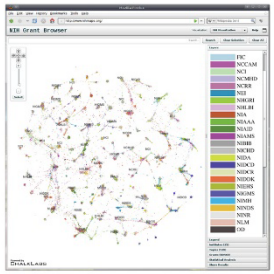
Keith V. Nesbitt. 2004. Ph.D. Thesis Map.

A Topic Map of NIH Grants 2007

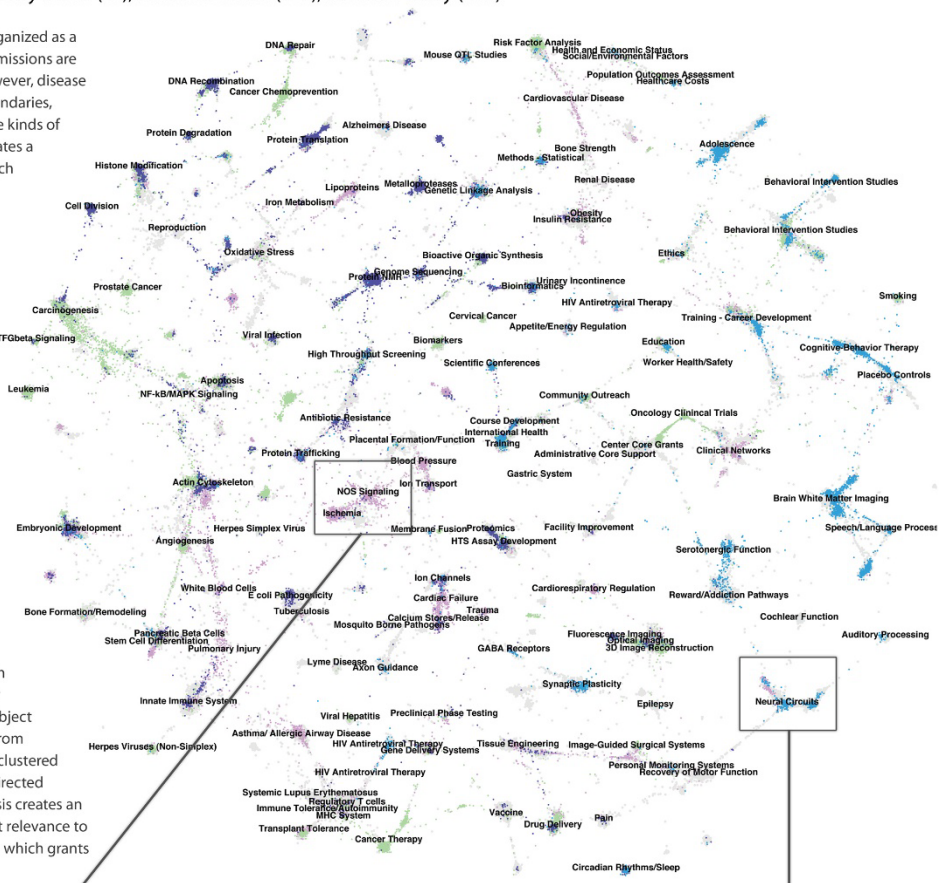
ChalkLabs UCIrvine

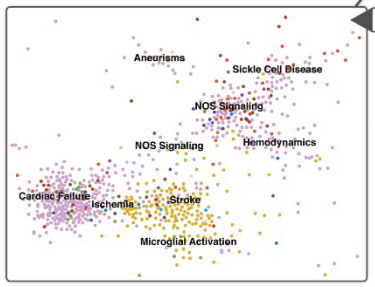
Bruce W. Herr II (Chalklabs & IU), Gully Burns (ISI), David Newman (UCI), Edmund Talley (NIH)

The National Institutes of Health (NIH) is organized as a multitude of Institutes and Centers whose missions are primarily focused on distinct diseases. However, disease etiologies and therapies flout scientific boundaries, and thus there is tremendous overlap in the kinds of research funded by each Institute. This creates a daunting landscape for decisions on research directions, funding allocations, and policy formulations. Shown here is devised an interactive topic map for navigating this landscape, online at www.nihmaps.org. Institute abbreviations can be found at www.nih.gov/icd.



Topic modeling, a statistical technique that automatically learns semantic categories, was applied to assess projects in terms used by researchers to describe their work, without the biases of keywords or subject headings. Grant similarities were derived from their topic mixtures, and grants were then clustered on a two-dimensional map using a force-directed simulated annealing algorithm. This analysis creates an interactive environment for assessing grant relevance to research categories and to NIH Institutes in which grants are localized.



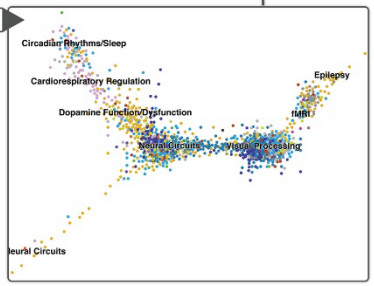


Cardiac Diseases Research

An area of the map focused on cardiovascular function and dysfunction. Cardiac Failure (primarily funded by NHLBI) is topically clustered next to Stroke (NINDS), since these are the two major medical emergencies associated with ischemia, which results from a restricted blood supply. Also localized in this area are grants focused on Nitric Oxide (NOS) Signaling, a major biochemical pathway for vasodilation, and grants on Hemodynamics, Sickle Cell Disease, and Aneurysms.

Neural Circuits Research

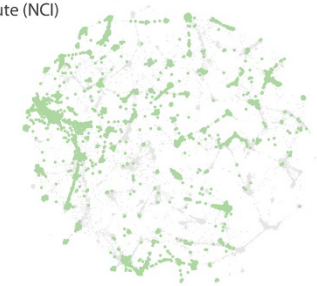
An area of the map focused on neural circuits, which shows the diversity of topics and NIH Institutes that fund research in this area, such as: Cardiorespiratory Regulation, primarily funded by NHLBI; Visual Processing, primarily funded by NEI; and Epilepsy, primarily funded by NINDS. For color coding, see legend in the upper-left inset.



National Cancer Institute (NCI)

TOP 10 TOPICS

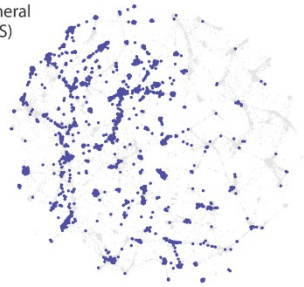
1. Oncology Clinical Trials
2. Cancer Treatment
3. Cancer Therapy
4. Carcinogenesis
5. Risk Factor Analysis
6. Cancer Chemotherapy
7. Metastasis
8. Leukemia
9. Prediction/Prognosis
10. Cancer Chemoprevention



National Institute of General Medical Sciences (NIGMS)

TOP 10 TOPICS

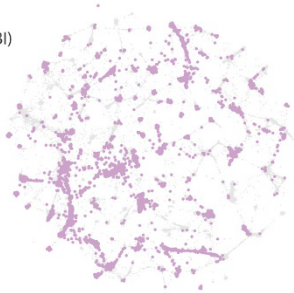
1. Bioactive Organic Synthesis
2. X-ray Crystallography
3. Protein NMR
4. Computational Models
5. Yeast Biology
6. Metalloproteases
7. Enzymatic Mechanisms
8. Protein Complexes
9. Invertebrate/Zebrafish Genetics
10. Cell Division



National Heart, Lung, and Blood Institute (NHLBI)

TOP 10 TOPICS

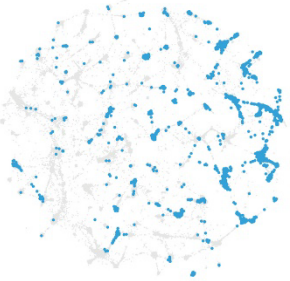
1. Cardiac Failure
2. Pulmonary Injury
3. Genetic Linkage Analysis
4. Cardiovascular Disease
5. Atherosclerosis
6. Hemostasis
7. Blood Pressure
8. Asthma/ Allergic Airway Disease
9. Gene Association
10. Lipoproteins



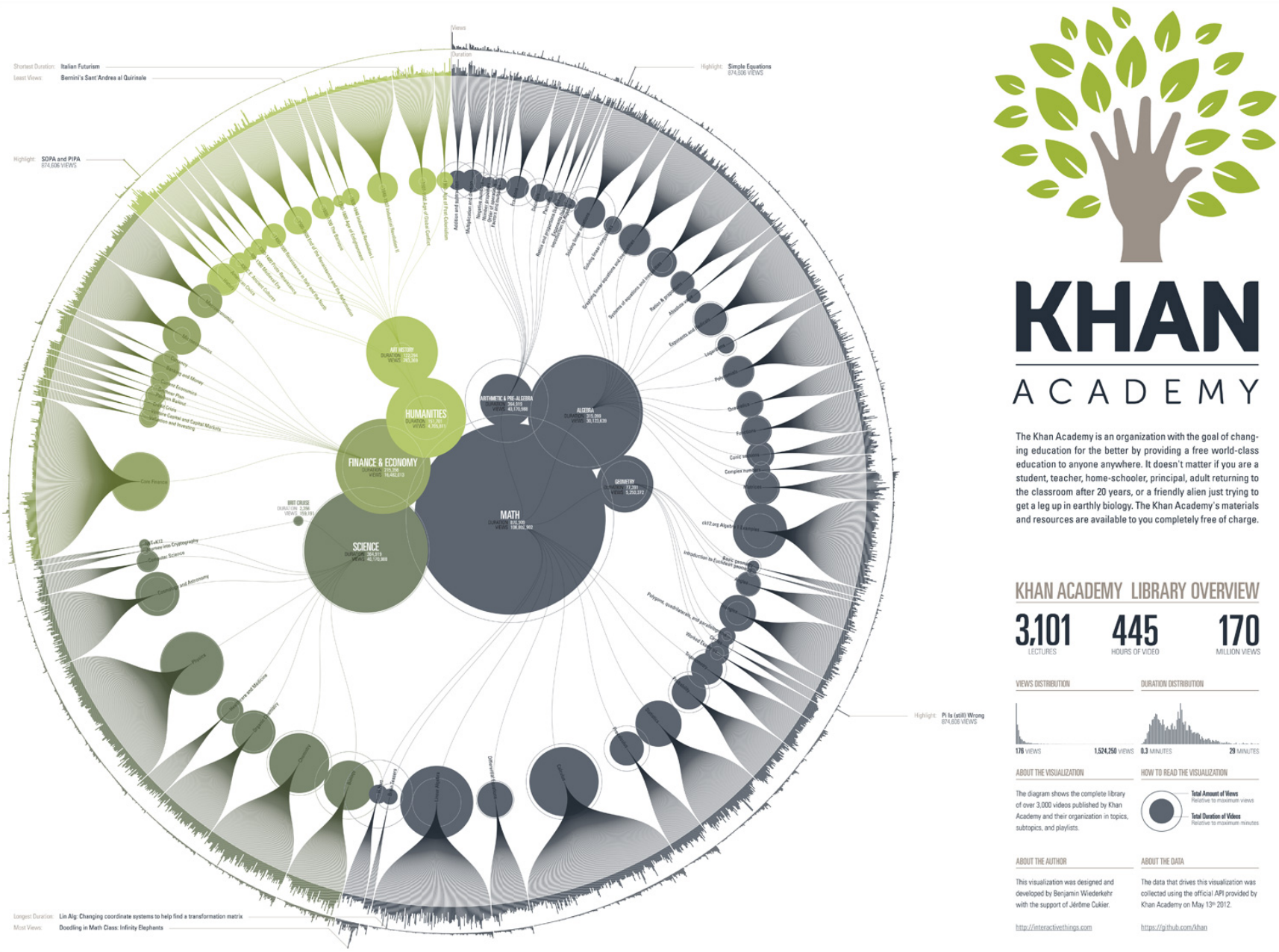
National Institute of Mental Health (NIMH)

TOP 10 TOPICS

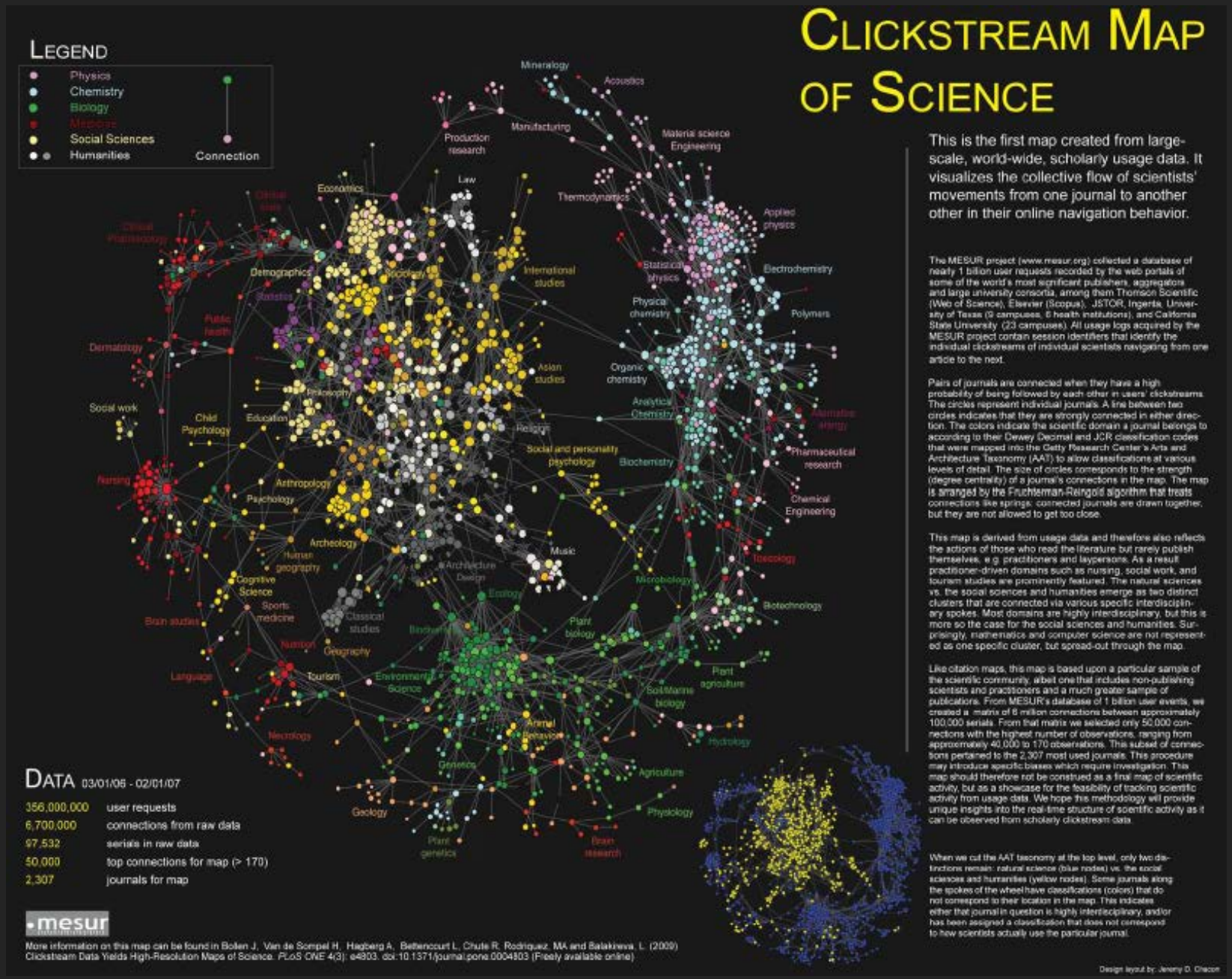
1. Mood Disorders
2. Schizophrenia
3. Behavioral Intervention Studies
4. Mental Health
5. Depression
6. Cognitive-Behavior Therapy
7. AIDS Prevention
8. Genetic Linkage Analysis
9. Adolescence
10. Childhood



Bruce W. Herr II, Gully A.P.C. Burns, David Newman, Edmund Talley. 2009. *A Topic Map of NIH Grants 2007.*



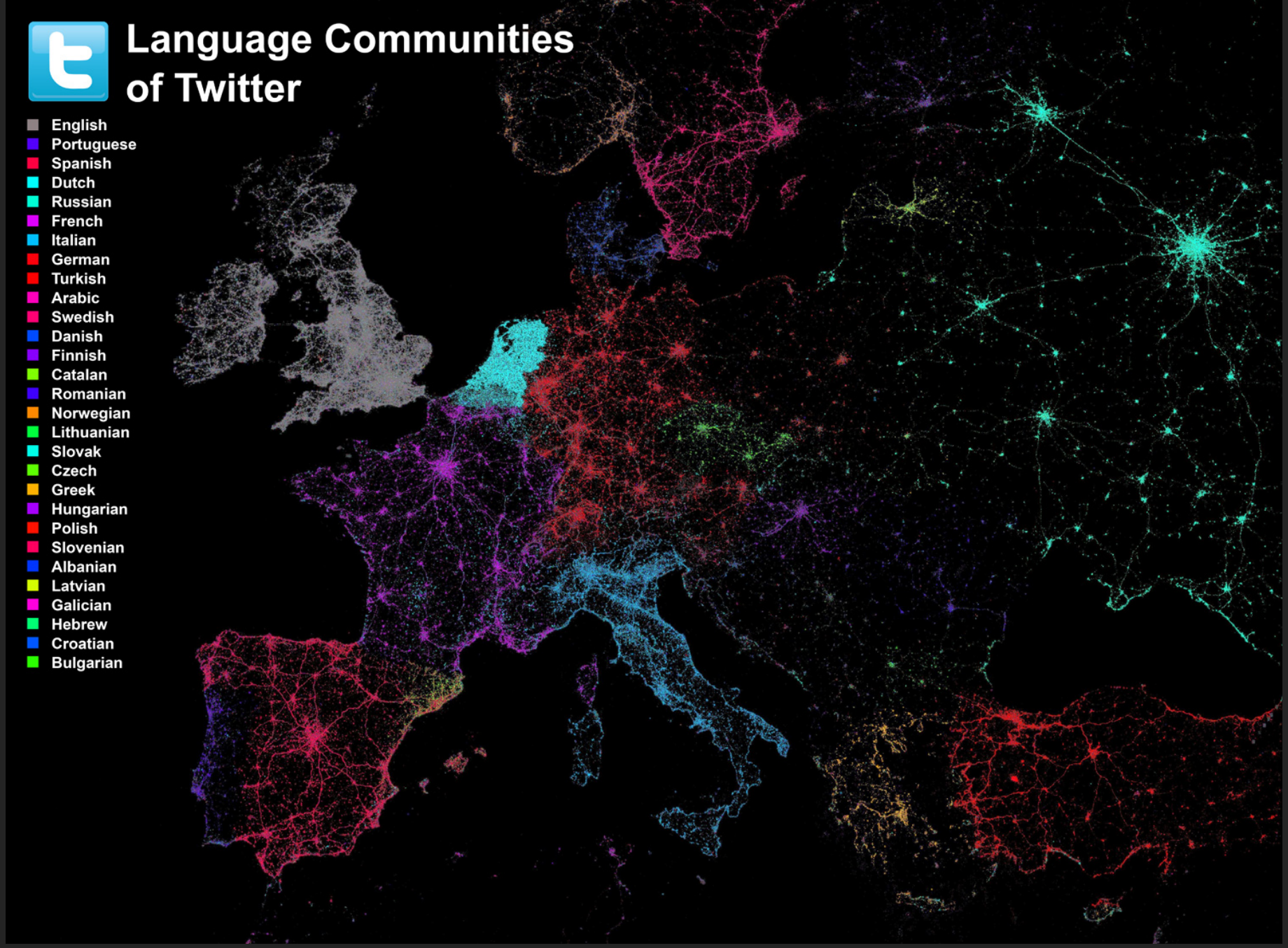
Benjamin Wiederkehr and Jérôme Cukier. 2012. Khan Academy Library Overview.



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.

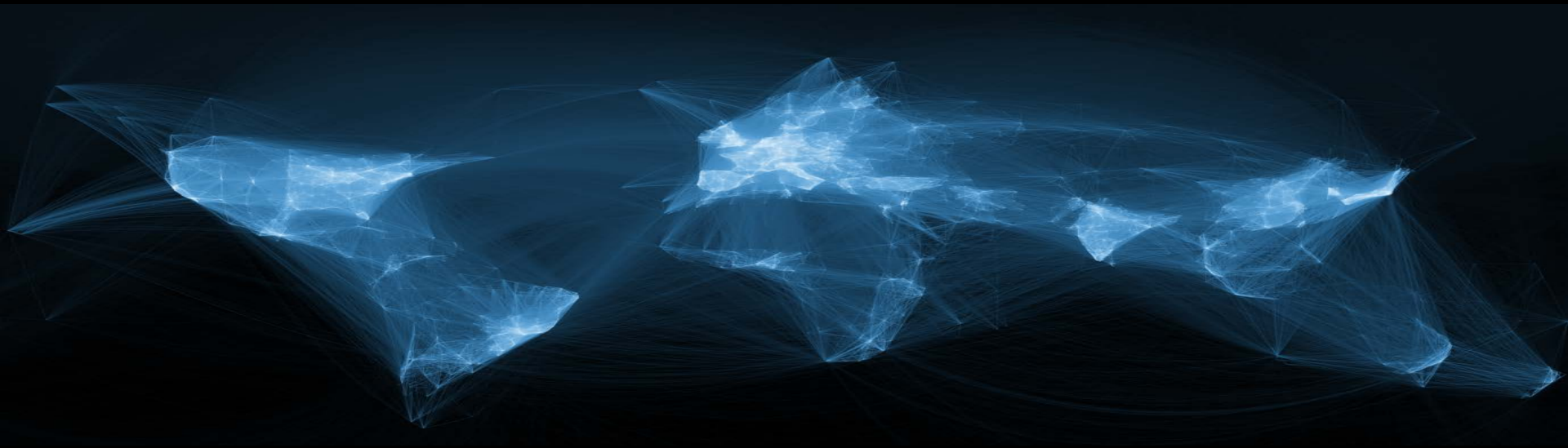
Martin Vargic. 2014. Map of the Internet.


FACEBOOK
PINTEREST
MYSPACE
TWITTER
TUMBLR
BLOGSPOT
TAGGED
BING
YAHOO!
ASK
REDDIT
DIGG
WIKIA
WIKIPEDIA
INFORMATION SEA
RENREN
WEIBO
QQ
NYTIMES
CNN
BBC
SEA OF ART
DEVIANTART
4CHAN
9GAG
GOOGLEPLUS
GMAIL
Google Search
ABOUT
MAIL.RU
SOHU
360
163
YELP
BRITANNICA
DATA OCEAN
GOOGLE
YOUTUBE
GOOGLE DRIVE
YANDEX
BUZZFEED
CHROME
INTERNET EXPLORER
NORTH DATASTREAM
THE CAPITOL
MICROS
Land of Indexes
GULF OF GOOGLE
VIMEO
IGN
GAMESPOT
LEAGUE OF LEGENDS
RAPID SHARE
DROPBOX
SOUTH DATASTREAM
MOZILLA
CLOUD
OPERA
BLUEHOST
SAFARI
DEVELOPER SEA
CONDUIT
APPSTO
APP SEA
DESPICABLE SEA
MEGA
MEGA UPLOAD
PREZI
SLIDE SHARE
FLICKR
SEA OF ARCHIVES
BITTORRENT
WAREZ SEA
ISOHUNT
PHOTOBUCKET
GO DADDY
HOST GATOR
WORDPRESS.COM
GOOGLE PLAY
ANDROID
SOFTONIC
PLENTY OF FISH
ADULT FRIEND FINDER
MYFREECAMS
LIVE JASMIN
YOUPORN
REDTUBE
HARDSEX TUBE
PIRATE SEA



Eric Fischer. 2012. Language Communities of Twitter.

Increasing Data Visualization Literacy



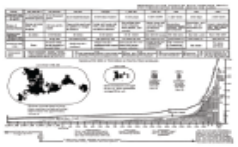
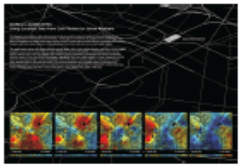
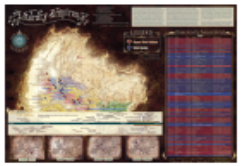
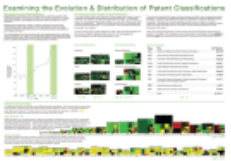
Data Visualization Literacy defined:

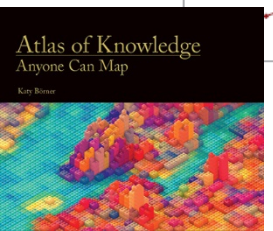
Data visualization literacy (ability to read, make, and explain data visualizations) requires

- literacy (ability to read and write text, e.g., in titles, axis labels, legend),
- visual literacy (ability to find, interpret, evaluate, use, and create images and visual media),
- and data literacy (ability to read, create, and communicate data).

The course can be taken for free or for Indiana University credits as part of the Online Data Science Program or the ILS M.S. program. Register for free at http://ivmooc.cns.iu.edu. Next class starts January 10, 2017.

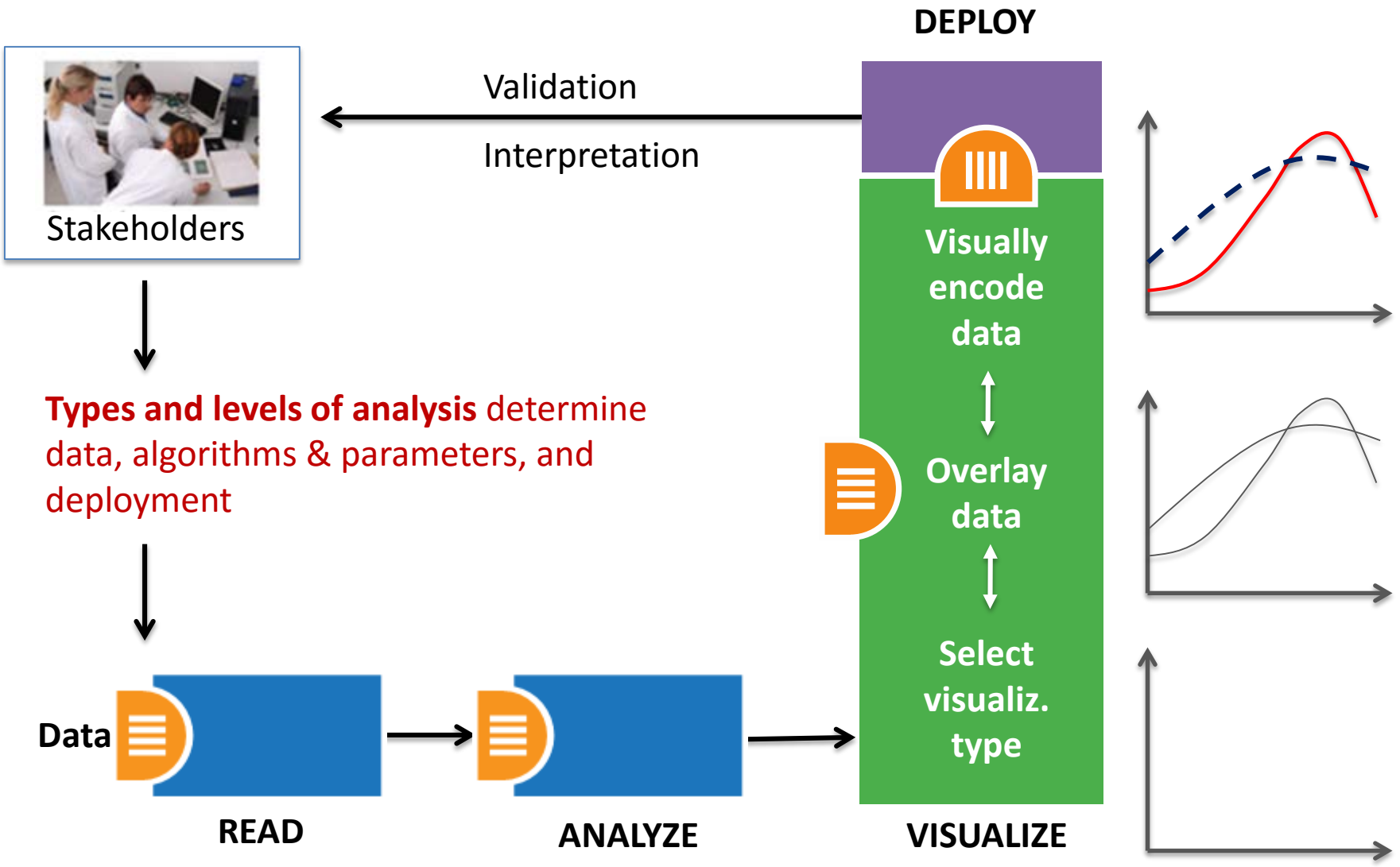
Tasks

TYPES	LEVELS		
	MICRO: Individual Level about 1–1,000 records page 6	**MESO: Local Level** about 1,001–100,000 records page 8	**MACRO: Global Level** more than 100,000 records page 10
Statistical Analysis page 44	 Knowledge Cartography page 135	 Productivity of Russian life sciences research teams page 105	 Number of scientists versus population and R&D costs versus GNP. page 103
WHEN: Temporal Analysis page 48	 Visualizing decision-making processes page 95	 Key events in the development of the video tape recorder page 85	 Increased travel and communication speeds page 83
WHERE: Geospatial Analysis page 52	 Cell phone usage in Milan, Italy page 109	 Victorian poetry in Europe page 137	 Ecological footprint of countries page 99
WHAT: Topical Analysis page 56	 Evolving patent holdings of Apple Computer, Inc. and Jerome Lemelson page 89	 Evolving journal networks in nanotechnology page 139	 Product space showing co-export patterns of countries page 93
WITH WHOM: Network Analysis page 60	 World Finance Corporation network page 87	 Electronic and new media art networks page 133	 World-wide scholarly collaboration networks page 157

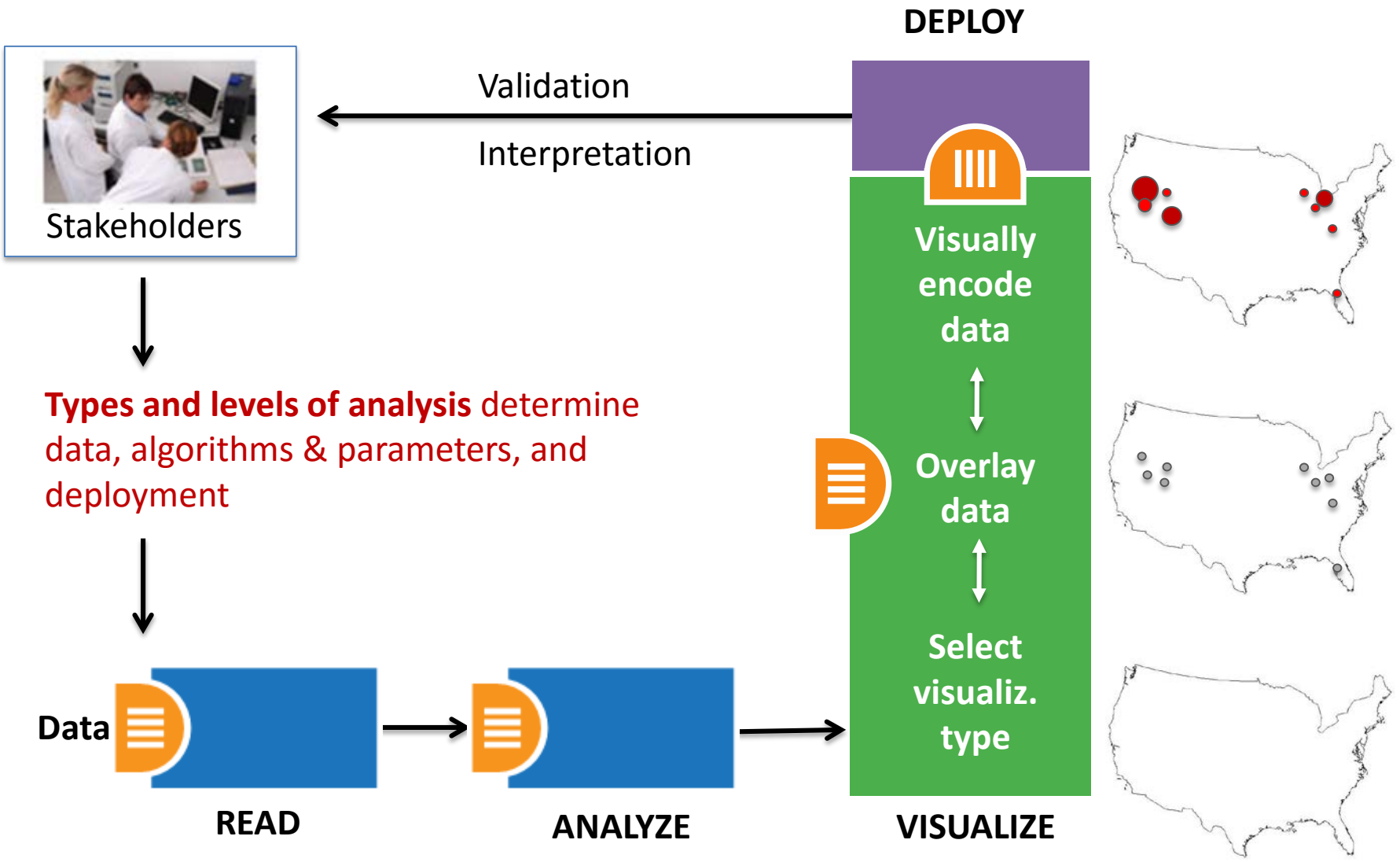


See page 5

Needs-Driven Workflow Design



Needs-Driven Workflow Design



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

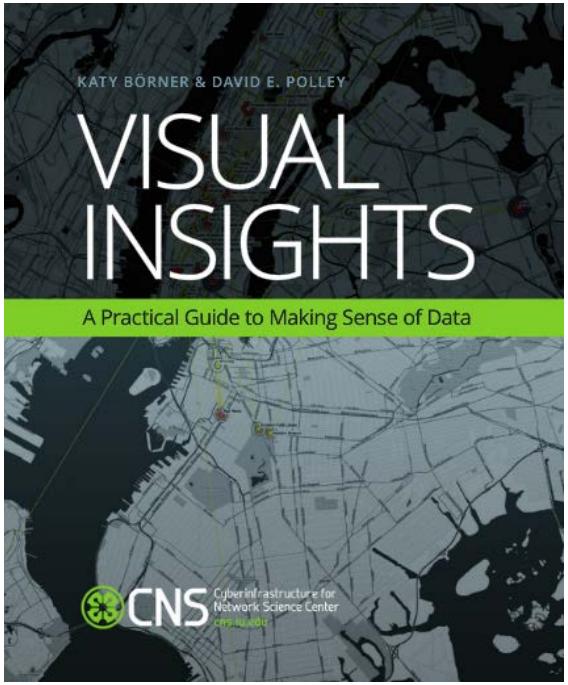
- **Session 5** – “With Whom:” Trees
- **Session 6** – “With Whom:” Networks
- **Session 7** – Dynamic Visualizations and Deployment

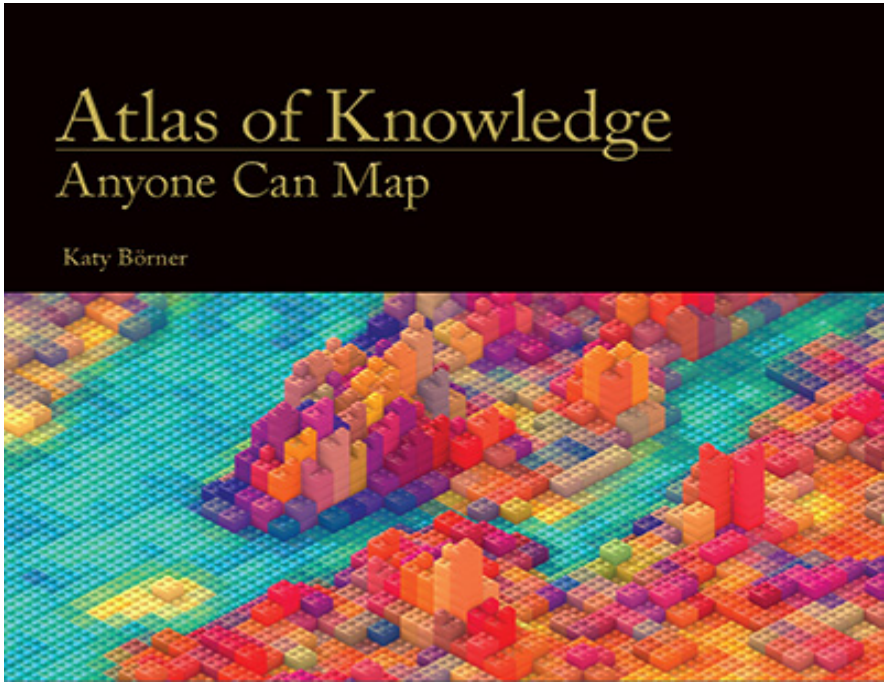
Final Exam

Part 2: Students work in teams on client projects.

Final grade is based on Homework (10%), Midterm (20%), Final Exam (30%), Client Project (30%), and Class Participation (10%).

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.

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

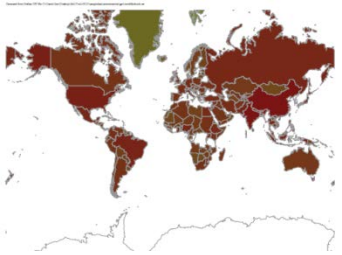
Statistical Analysis–p. 44

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

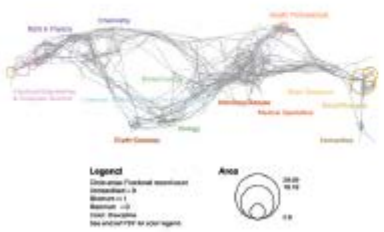
Geospatial Analysis–p. 52



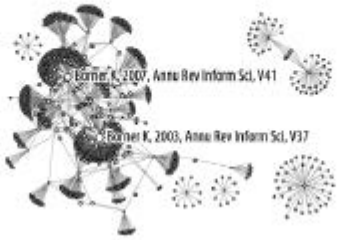
Load **One** File and Run **Many** Analyses and Visualizations

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

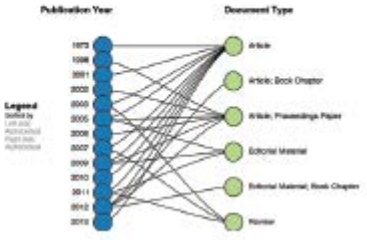
Topical Analysis–p. 56



Paper Citation Network–p. 60

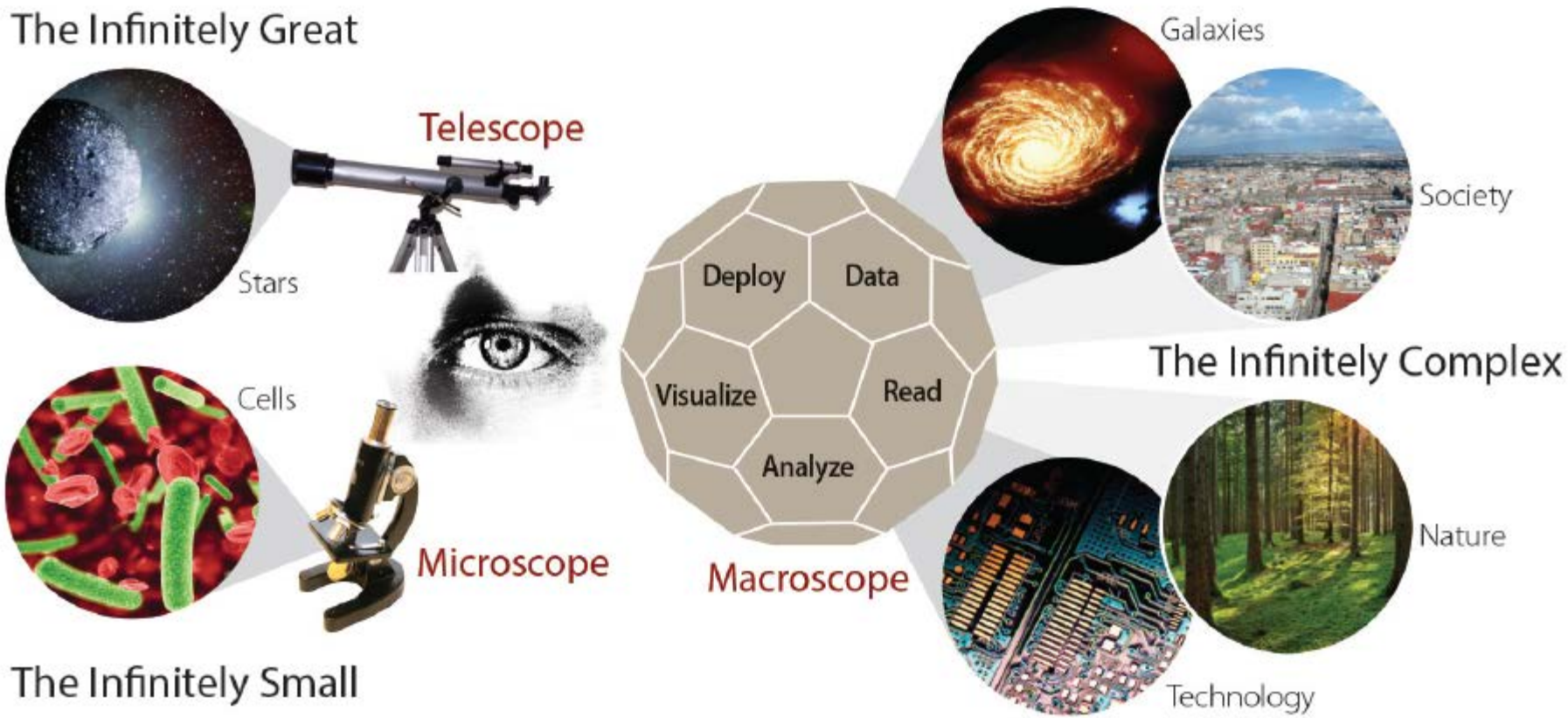


Bi-Modal Network–p. 60



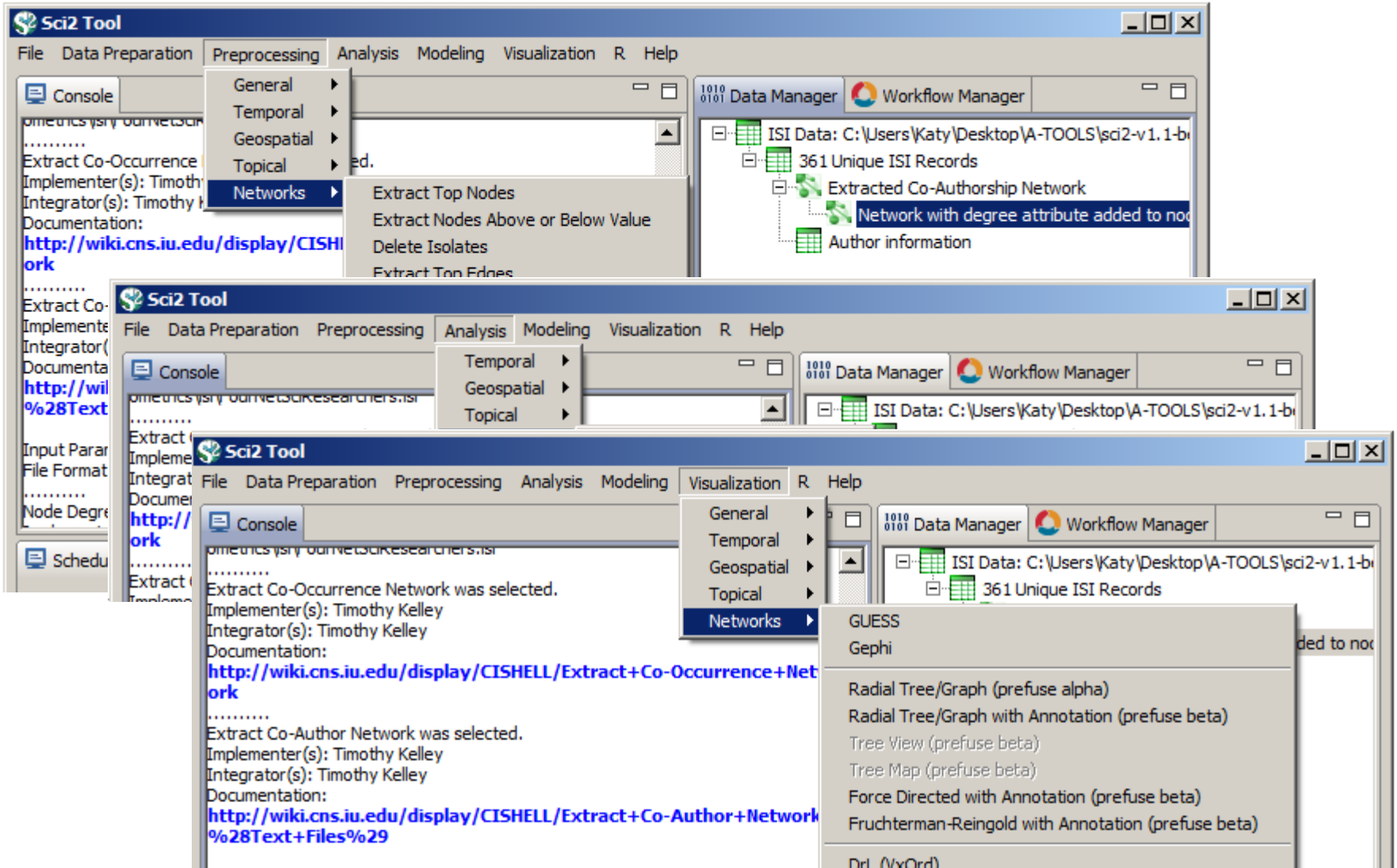
Co-author and many other bi-modal networks.

Microscopes, Telescopes, Macroscopes Plug-and-Play Macroscopes



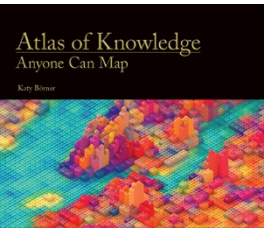
Sci2 Tool Interface Components

Download tool for free at http://sci2.cns.iu.edu



Visualization Framework

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



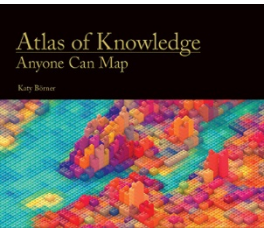
See page 24

Visualization Framework

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

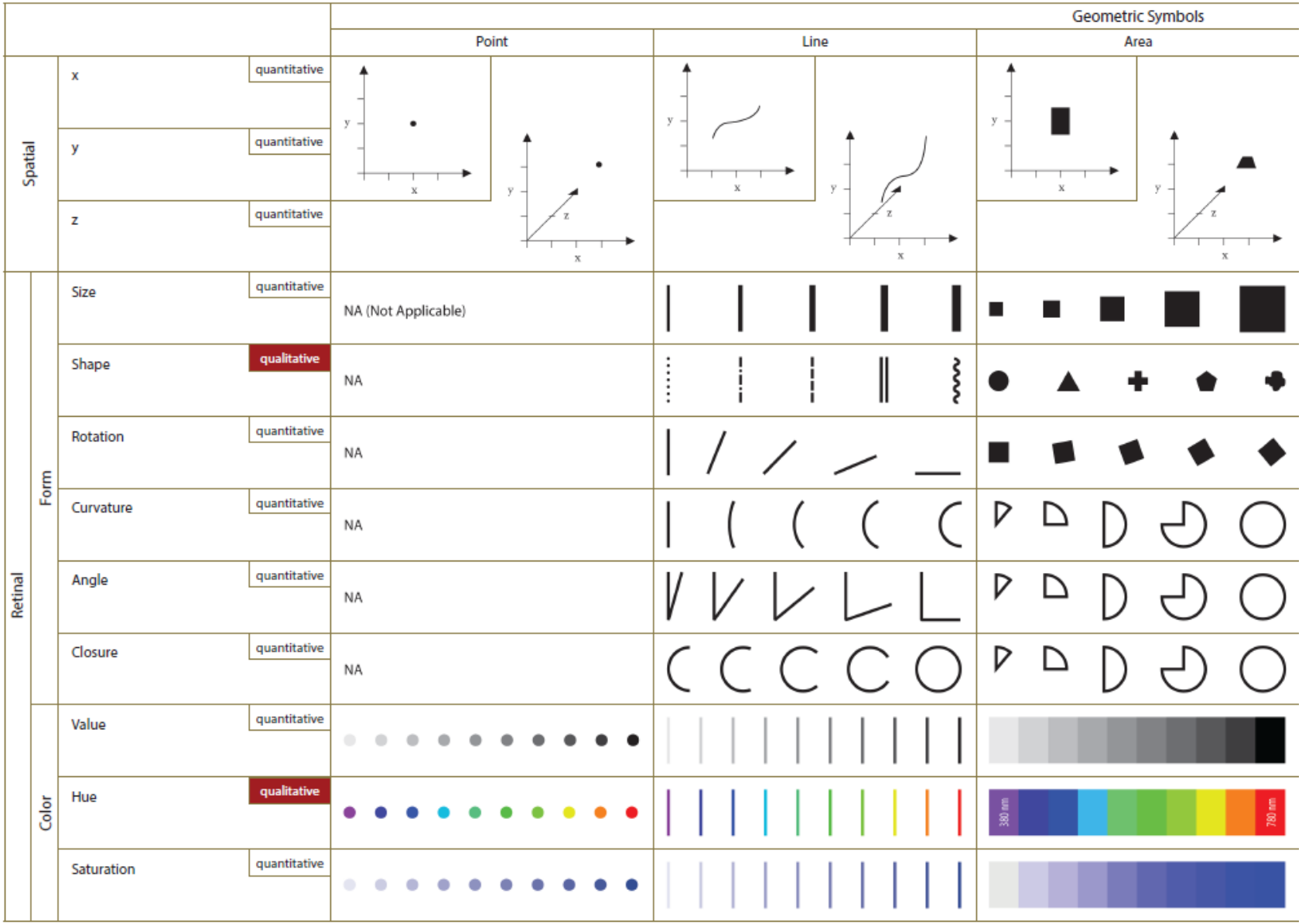
Visualization Framework

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

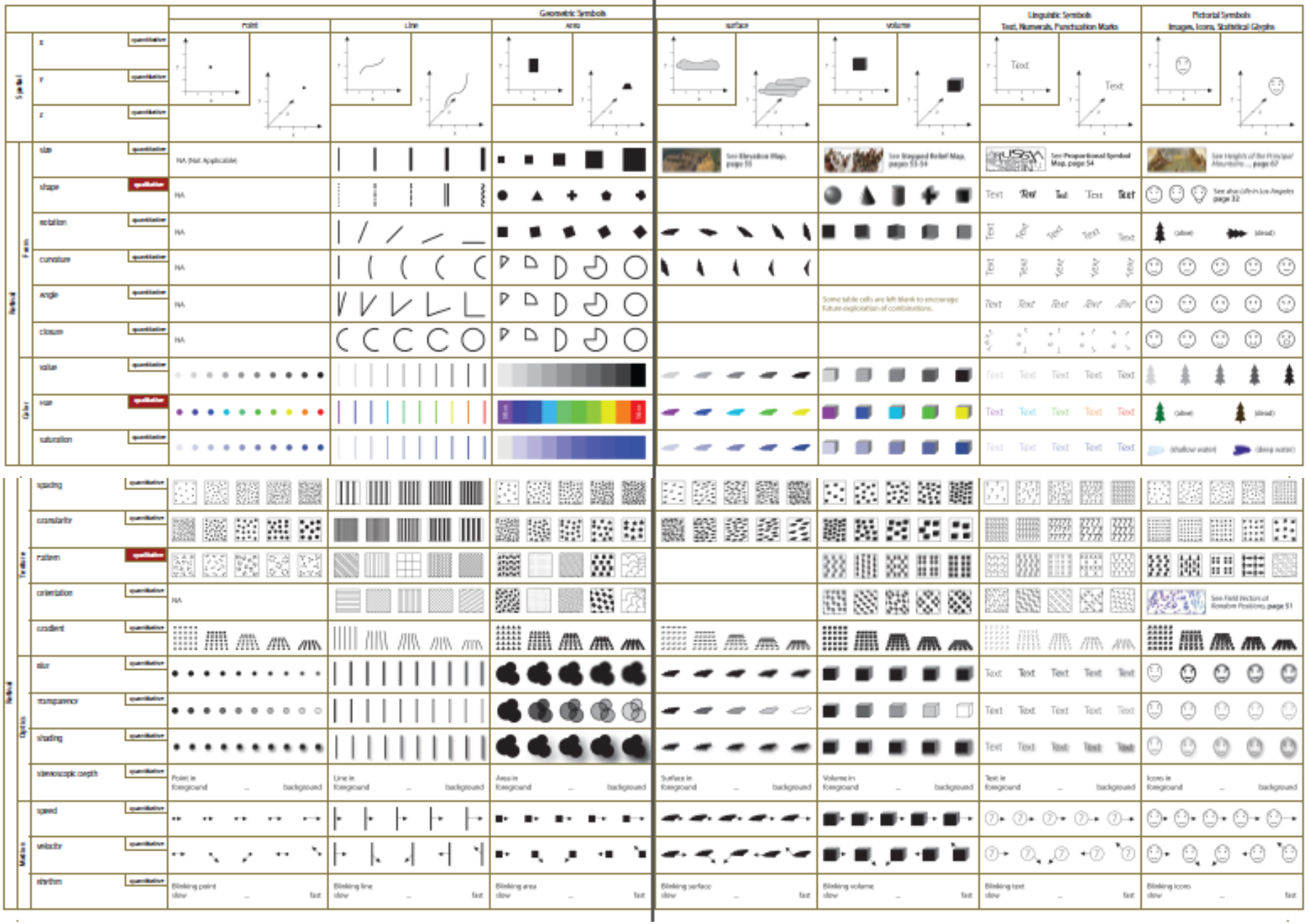


See page 24

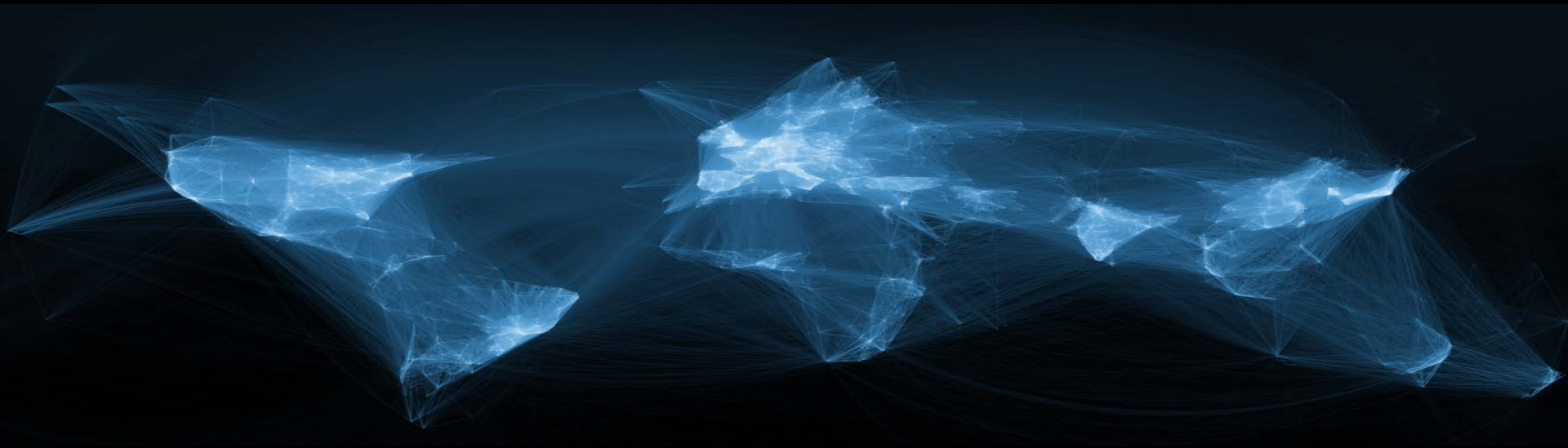
Graphic Variable Types Versus Graphic Symbol Types



Graphic Variable Types Versus Graphic Symbol Types



Descriptive Models



The Global 'Scientific Food Web'

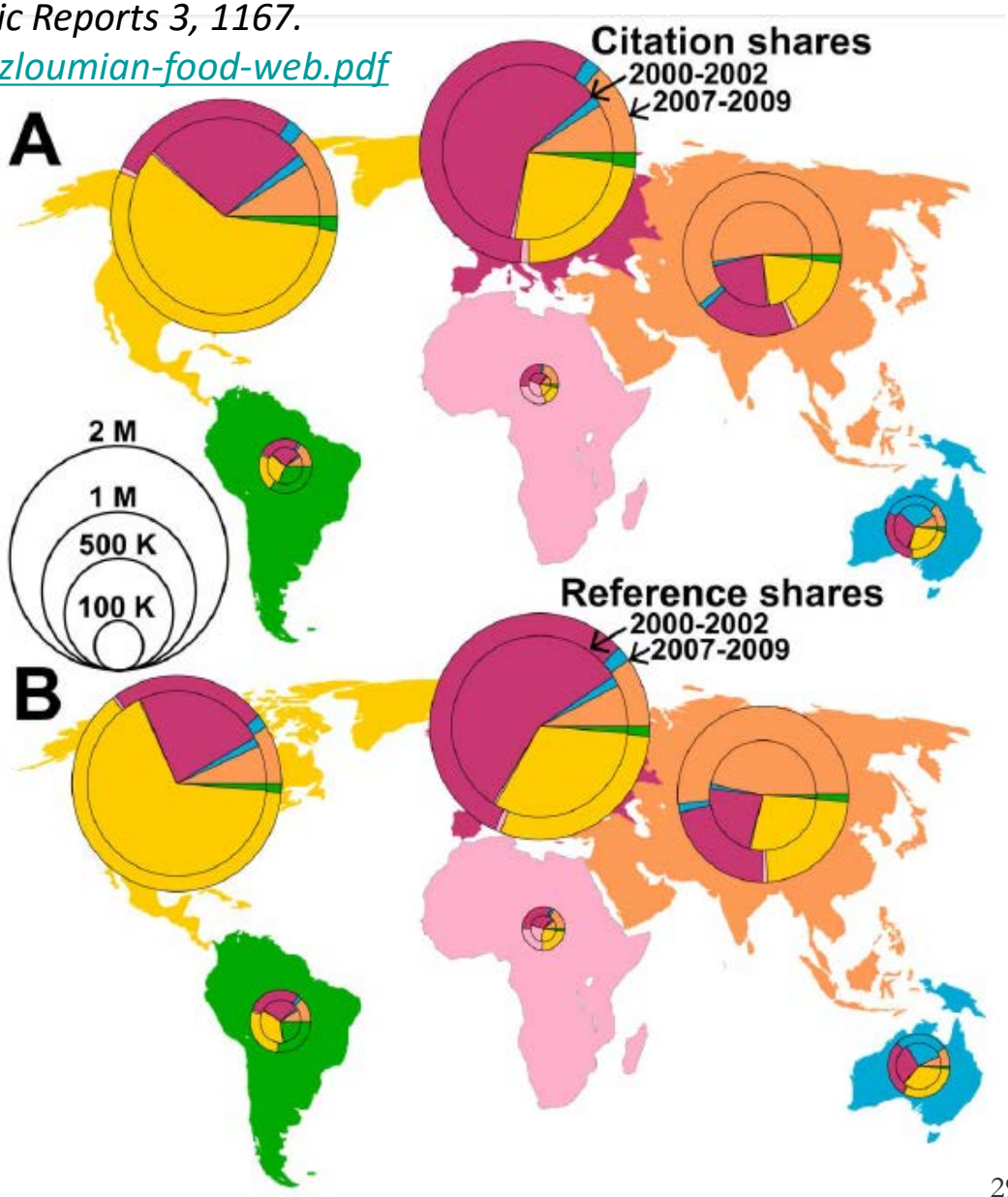
Mazloumian, Amin, Dirk Helbing, Sergi Lozano, Robert Light, and Katy Börner. 2013. "Global Multi-Level Analysis of the 'Scientific Food Web'". Scientific Reports 3, 1167.
http://cns.iu.edu/docs/publications/2013-mazloumian-food-web.pdf

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.



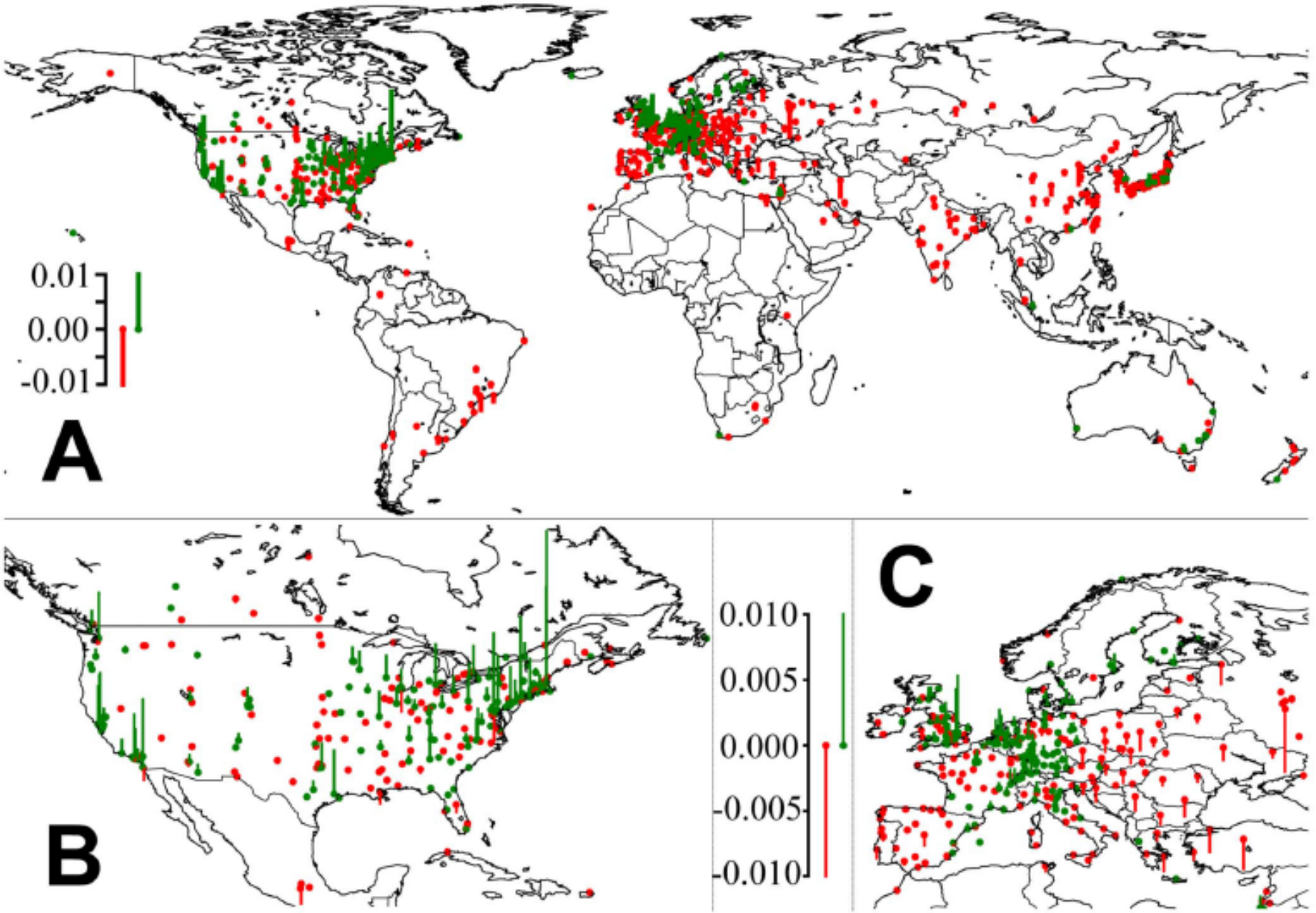


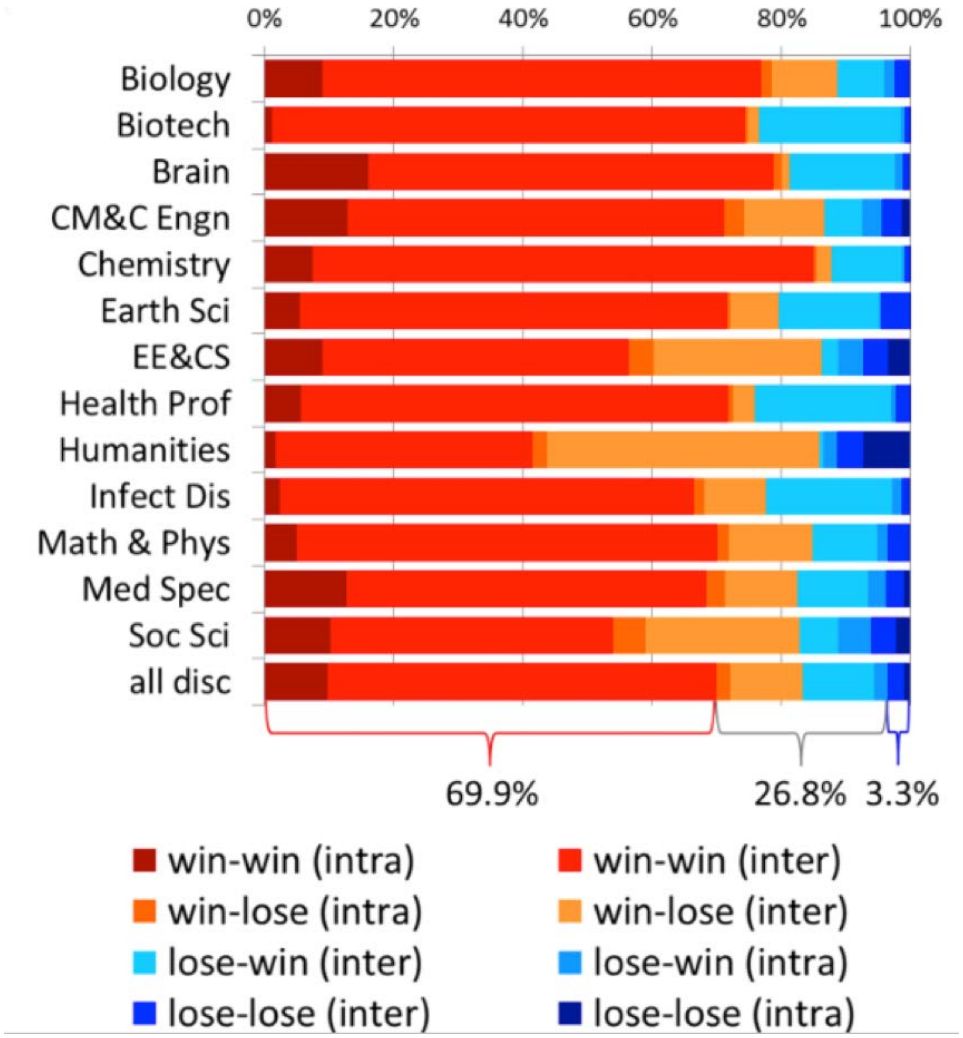
Figure 2 | World map of the greatest knowledge sources and sinks, based on our scientific fitness index. Green bars indicate that the number of citations received is over-proportional, red that the number of citations received is lower than expected (according to a homogeneous distribution of citations 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).

Long-Distance Interdisciplinarity Leads to Higher Scientific Impact

Larivière, Vincent, Stefanie Haustein, and Katy Börner. 2015. PLOS ONE DOI: 10.1371.

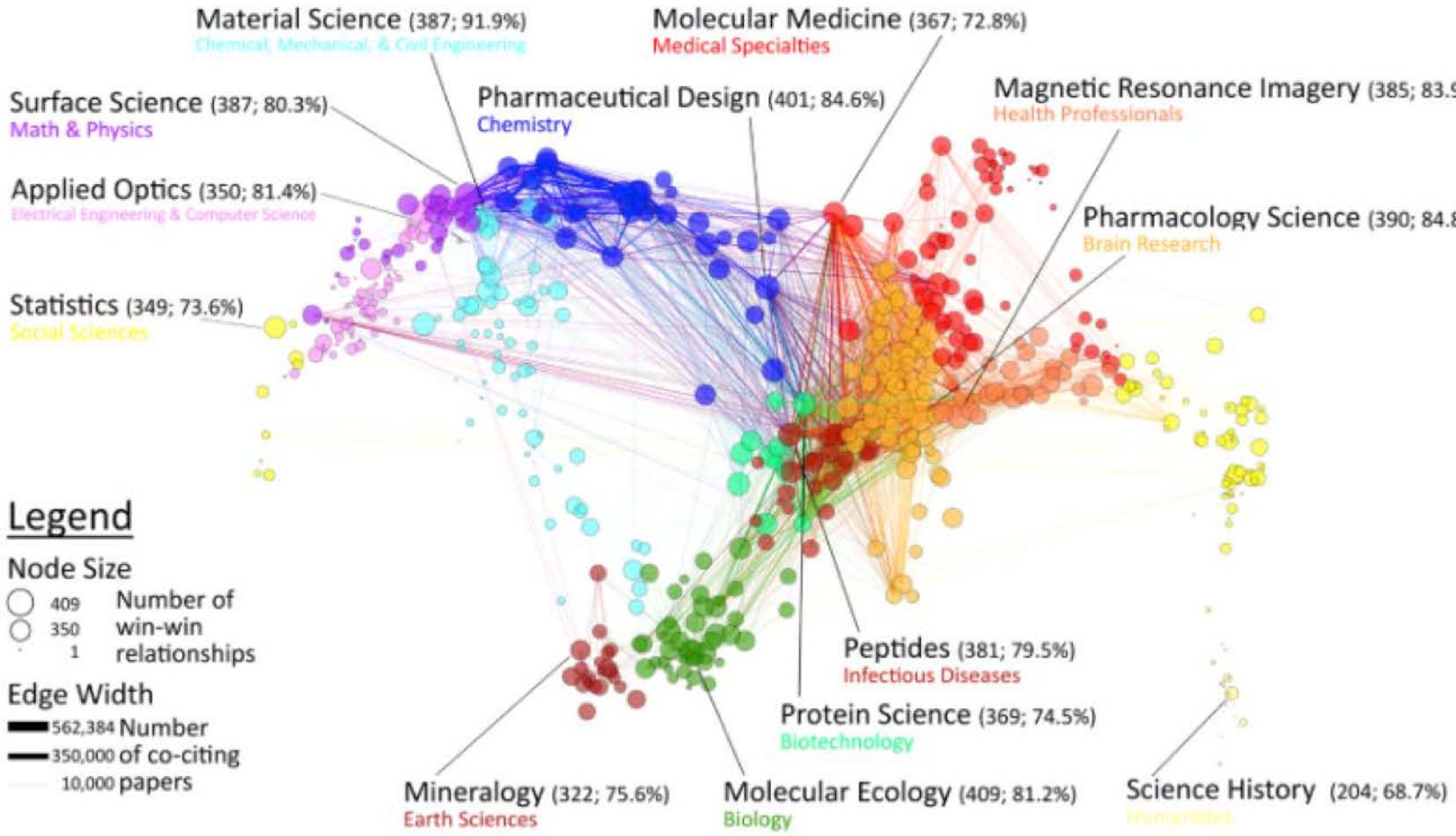
Data: 9.2 million interdisciplinary research papers published between 2000 and 2012.

Results: majority (69.9%) of co-cited interdisciplinary pairs are “win-win” relationships, i.e., papers that cite them have higher citation impact and there are as few as 3.3% “lose-lose” relationships. UCSD map of science is used to compute “distance.”



A1 Number of papers citing win-win relationships (≥10,000 citing papers)

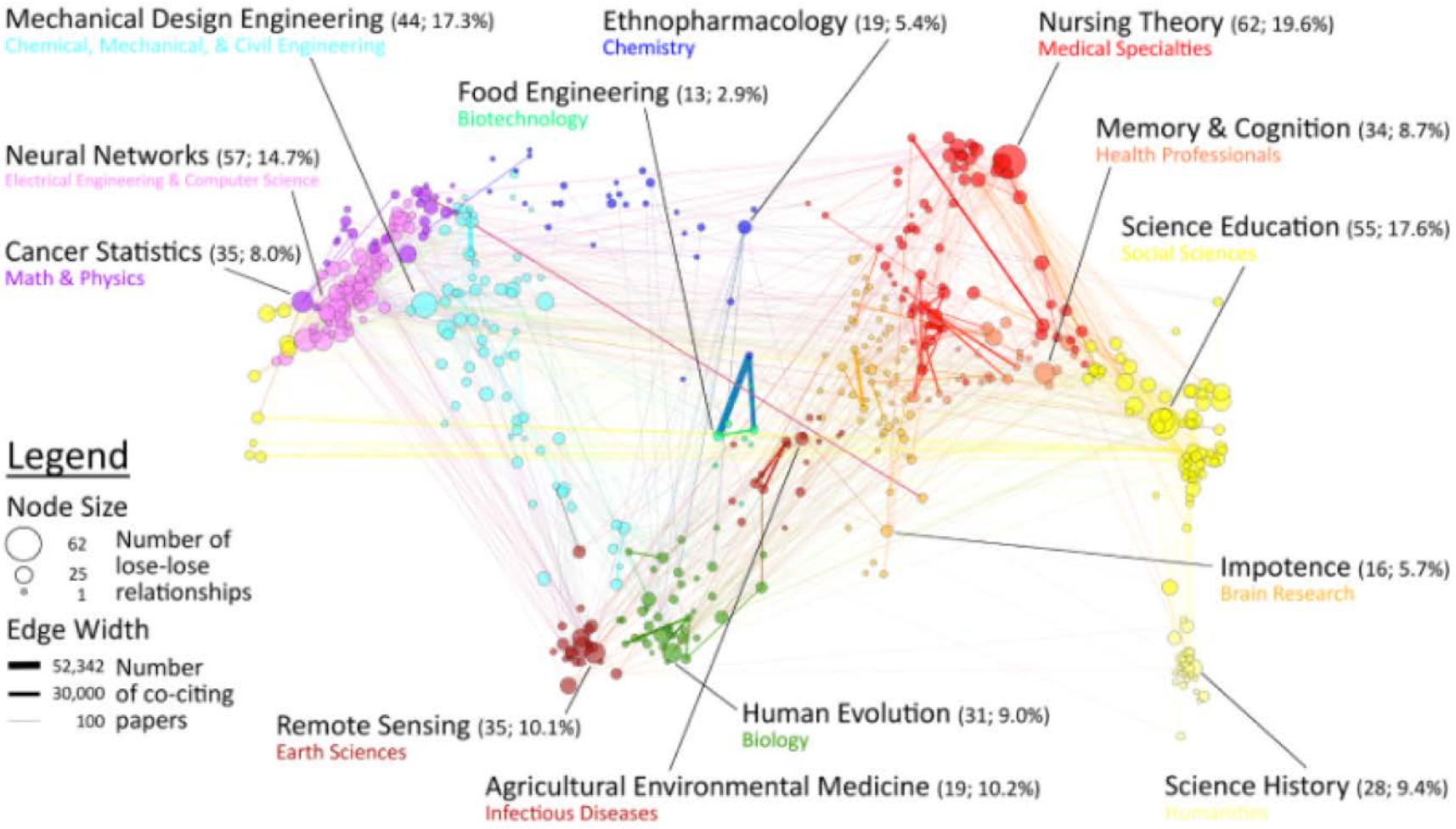
citing papers



2,940 (5.19%) of 56,614 win-win edges

node color: discipline | edge color: mix of adjacent nodes | labels: subdiscipline with highest number of win-win relationships per discipline (number and percentage of win-win relationships)

B1 Number of papers citing lose-lose relationships (≥100 citing papers)



1,204 (44.4%) of 2,712 lose-lose edges

node color: discipline | edge color: mix of adjacent nodes | labels: subdiscipline with highest number of lose-lose relationships per discipline (number and percentage of lose-lose relationships)

Web of Science as a Research Dataset

Date:

November 14-15, 2016

Meeting Place:

Social Science Research Commons (SSRC),
Woodburn Hall, Room 200
1100 East Seventh Street
Bloomington, IN 47405

Web Indiana University Campus Map »

Workshop Goals

This practical workshop brings together data scientists and data stewards from research centers that are using the Web of Science™ at scale. We will explore WoS from the perspective of a research dataset and work together on practical ways to better support our research in the future. While the main focus will be on the Web of Science, the results should be extensible to all similar metadata aggregations. This unique focus—bringing data stewards and data scientists from these centers together to work on shared needs in tandem with the Web of Science team—will enable us to redefine and fully repurpose WoS to fit our research goals. We intend to launch an ongoing community in which we will learn techniques and develop tools to improve the data that underlies our research.

Advance Preparations

- Data stewards will provide a short profile of how WoS as a dataset is being implemented in the context of their research center/university and the technical, content, and other challenges they are facing.
- Researcher data scientists will prepare a short profile of current research projects leveraging the WoS dataset, focusing on key challenges such as linking, disambiguating, mining, etc. that, if solved, would offer greater research opportunities.

Organizers:

Katy Börner

Victor H. Yngve Distinguished Professor of Information Science, Department of Information and Library Science, School of Informatics and Computing, Indiana University, Bloomington; Director, Cyberinfrastructure for Network Science Center & Curator of Mapping Science exhibit, Bloomington, IN
katy@indiana.edu

Eamon Duede

Executive Director, Knowledge Lab. Administrator, Metaknowledge Research Network, University of Chicago
eduede@uchicago.edu

James Pringle

Head of Industry Development & Innovation at Thomson Reuters IP & Science

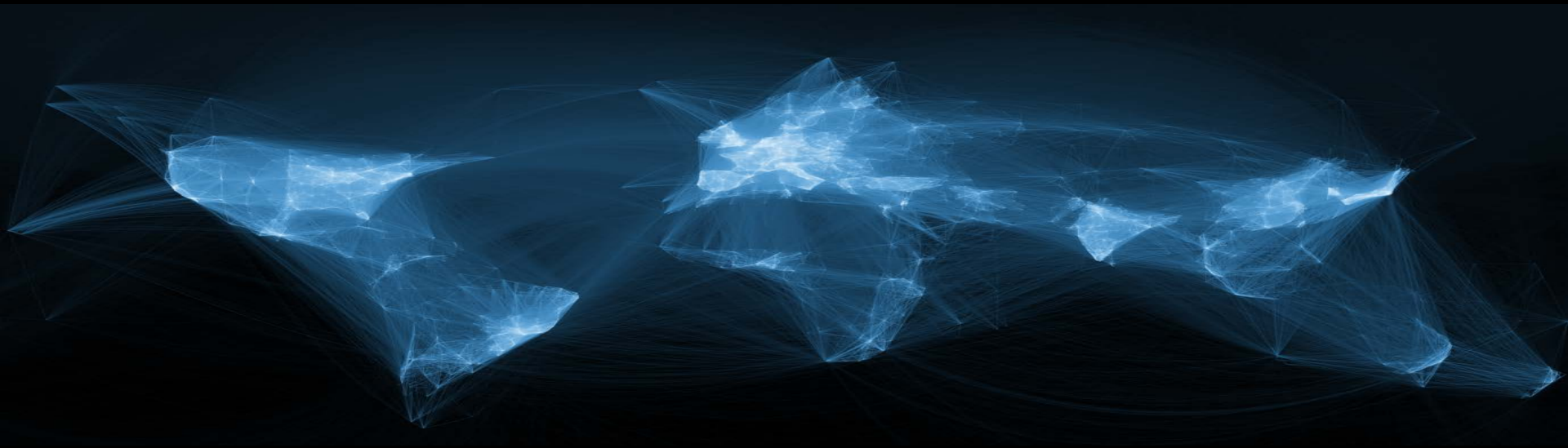
Bibliometrics & Research Assessment

A SYMPOSIUM FOR LIBRARIANS & INFORMATION PROFESSIONALS

presented by NIH Library nihlibrary.nih.gov

SLA Connecting People and Information Maryland Chapter

Communicating STI to Different Audiences



Places & Spaces: Mapping Science Exhibit, online at http://scimaps.org

PLACES & SPACES
MAPPING SCIENCE

Curated by the Cyberinfrastructure for Network Science Center

search scimaps.org Search

About People Maps & More Exhibitions Hosting Publications Store News Contact

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 (2017) http://scimaps.org/call

Background and Goals

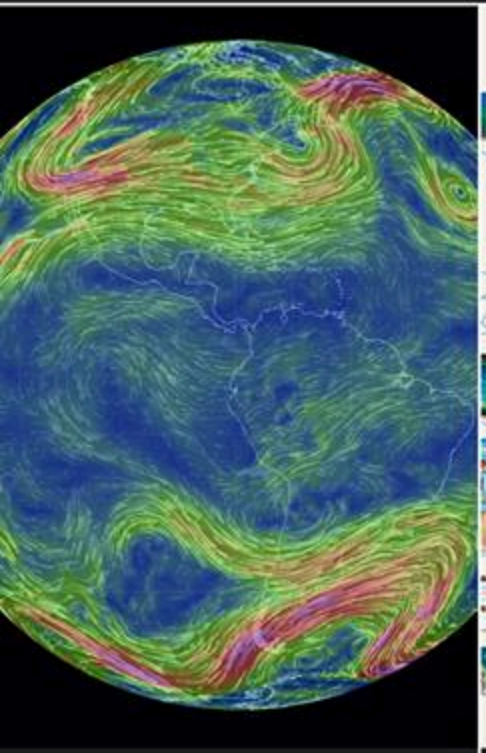
The *Places & Spaces: Mapping Science* exhibit was created to in
communicate human activity and scientific progress on a glol
that enable the close inspection of large-scale maps in public
conferences; (2) novel, interactive macroscope tools that let

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

MACROSCOPES FOR INTERACTING WITH SCIENCE

PLACES & SPACES
MAPPING SCIENCE
scimaps.org



Earth

AcademyScope

Mapping Global Society

Charting Culture

http://scimaps.org/iteration/11

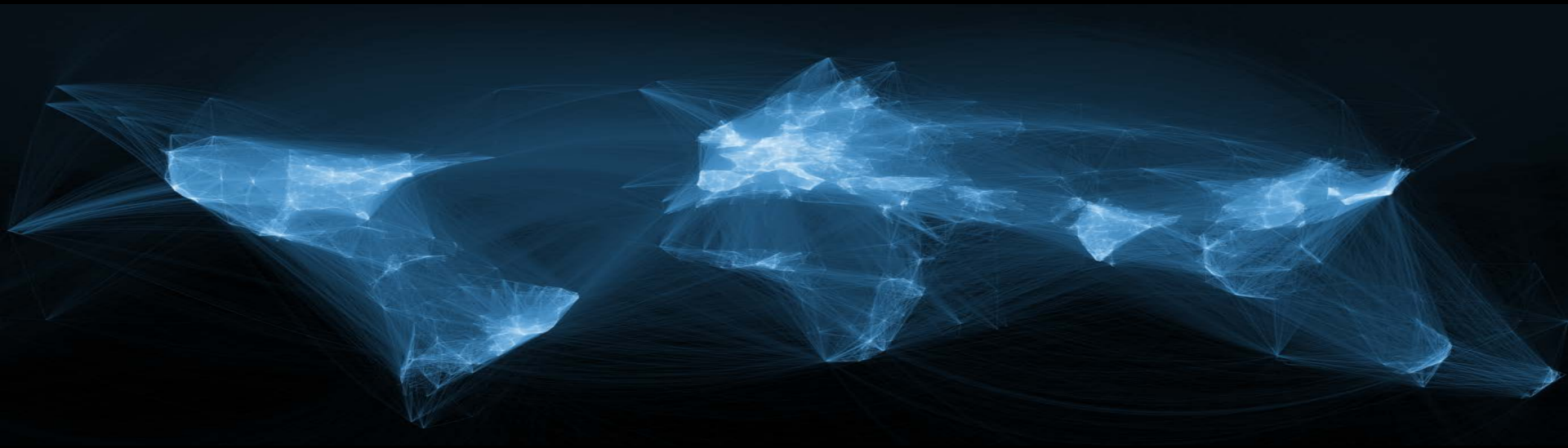
Places & Spaces Exhibit at the David J. Sencer CDC Museum, Atlanta, GA
January 25-June 17, 2016

Seeing for Action - Using Maps and Graphs to Protect the Public's Health.

Maps of Health Exhibit
David J. Sencer CDC Museum
Atlanta, GA

Jan 25-Jun 17, 2016

Predictive Models



Modelling Advantage

Models are widely used in the construction of scientific theories as they help

- Make assumptions explicit
- Describe the structure and dynamics of systems
- Communicate and explain systems
- Suggest possible interventions
- Identify new questions

Modelling Challenges

Comprise among others:

- Model utility and usability
- Model credibility and validation
- Model extendibility and reproducibility
- Model sharing and retrieval

Modelling Opportunities: Data-Driven Decision Making

Now available:

- high-quality, high coverage, interlinked data
- cost-effective storage and computation
- validated, scalable algorithms
- visualization and animations capabilities

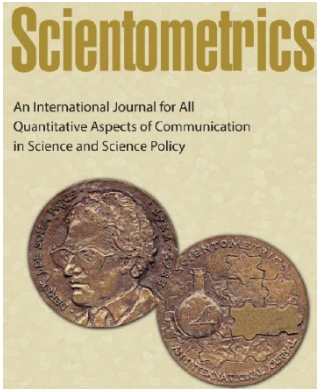
Government, academic, and industry leaders discussed challenges and opportunities associated with using big data, visual analytics, and computational models in STI decision-making.

Conference slides, recordings, and report are available via http://modsti.cns.iu.edu/report

Special Issue of *Scientometrics*:

Simulating the Processes of Science, Technology, and Innovation

Bruce Edmonds, Andrea Scharnhorst, Katy Börner & Staša Milojević (Editors)



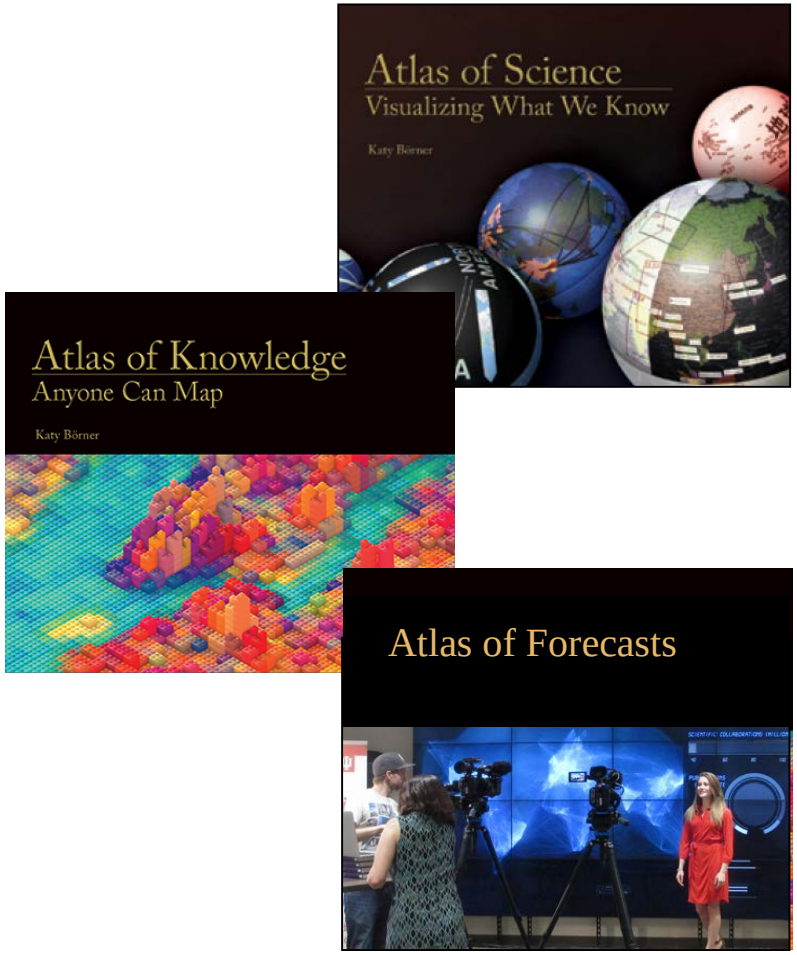
- ***Rogier De Langhe:*** Towards the Discovery of Scientific Revolutions in Scientometric Data
- ***Sabine Brunswicker, Sorin Matei, Michael Zentner, Lynn Zentner and Gerhard Klimeck:*** Creating Impact in the Digital Space: Digital Practice Dependency in Scientific Developer Communities
- ***Johan Bollen et al.:*** An Efficient System to Fund Science: From Proposal Review to Peer-to-Peer Distributions
- ***Petra Ahrweiler:*** Agent-based Simulation for Science, Technology and Innovation Policy
- ***David Chavalarias:*** What's Wrong With Science? Modeling Collective Discovery Processes With the Nobel Game
- ***Jeff Alstott, Giorgio Triulzi, Bowen Yan and Jianxi Luo:*** Mapping Technology Space by Normalizing Patent Technology Networks

Atlas Trilogy

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

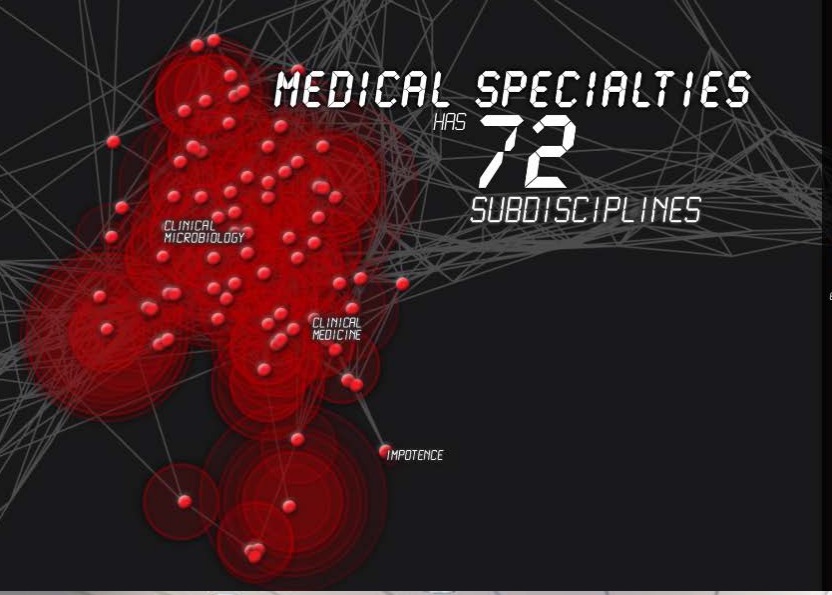
Börner, Katy (2015) **Atlas of Knowledge: Anyone Can Map**. The MIT Press.
http://scimaps.org/atlas2

Börner, Katy (2018) **Atlas of Forecasts: Predicting and Broadcasting Science, Technology, and Innovation**. The MIT Press.



Upcoming Sackler Colloquium on "**Modelling and Visualizing Science and Technology Developments**" will take place in October 2017 at the Beckman Center, Irvine, CA.

Science Forecast S1:E1, 2015



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