Scalable Multi-Scale Visual Analytical Tools for Health Science

Robert Light and Daniel O'Donnell

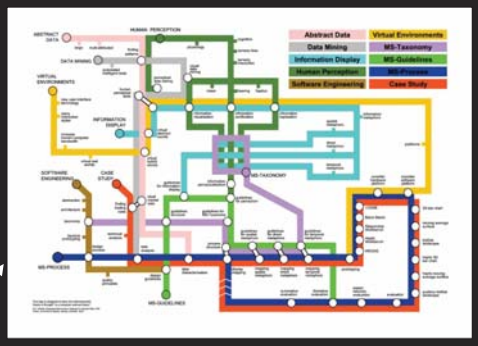
Cyberinfrastructure for Network Science Center
School of Informatics and Computing, Indiana University, USA

Center for Computational Research, SUNY Buffalo
March 29, 2016

Language Communities of Twitter - Eric Fischer - 2012

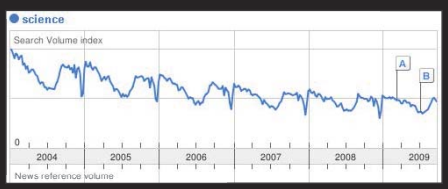
Terra bytes of data

Descriptive & Predictive Models



Find your way

Find collaborators, friends



Identify trends

Descriptive Models

Multiple levels: Micro … Macro

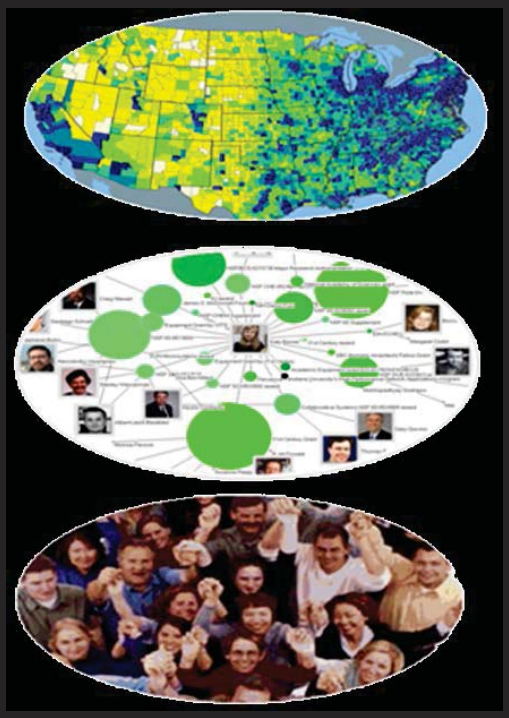
Answering: When? Where? What? With Whom?

Different Levels of Abstraction/Analysis

Macro/Global
Population Level

Meso/Local
Group Level

Micro
Individual Level

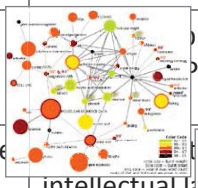
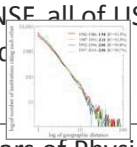


Type of Analysis vs. Level of Analysis

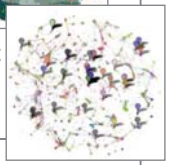
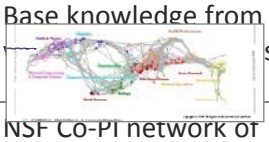
	Micro/Individual (1-100 records)	***Meso/Local (101–100,000 records)***	***Macro/Global (100,000 < records)***
Statistical Analysis/Profiling	Individual person and their expertise profiles	Larger labs, centers, universities, research domains, or states	All of NSF, all of USA, all of science.
Temporal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

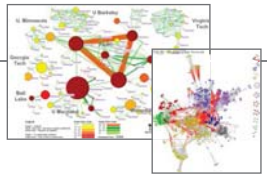
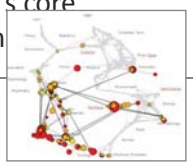
Type of Analysis vs. Level of Analysis

	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

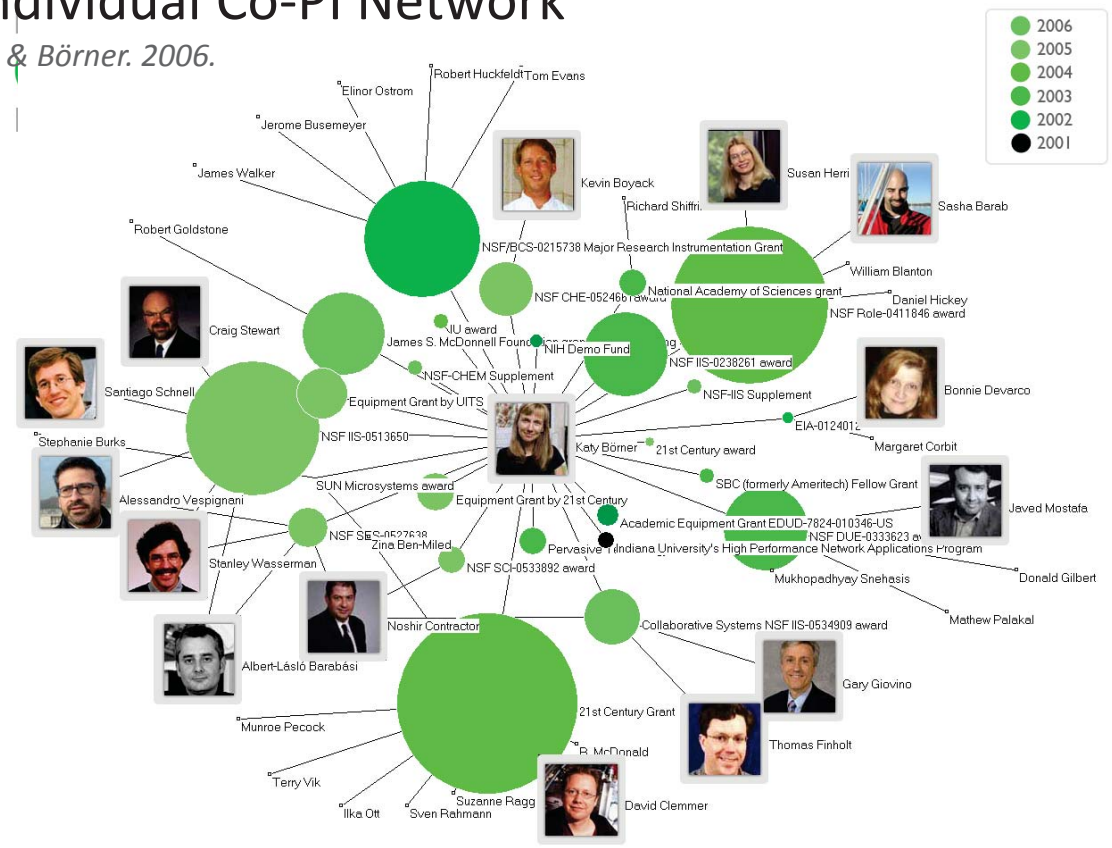


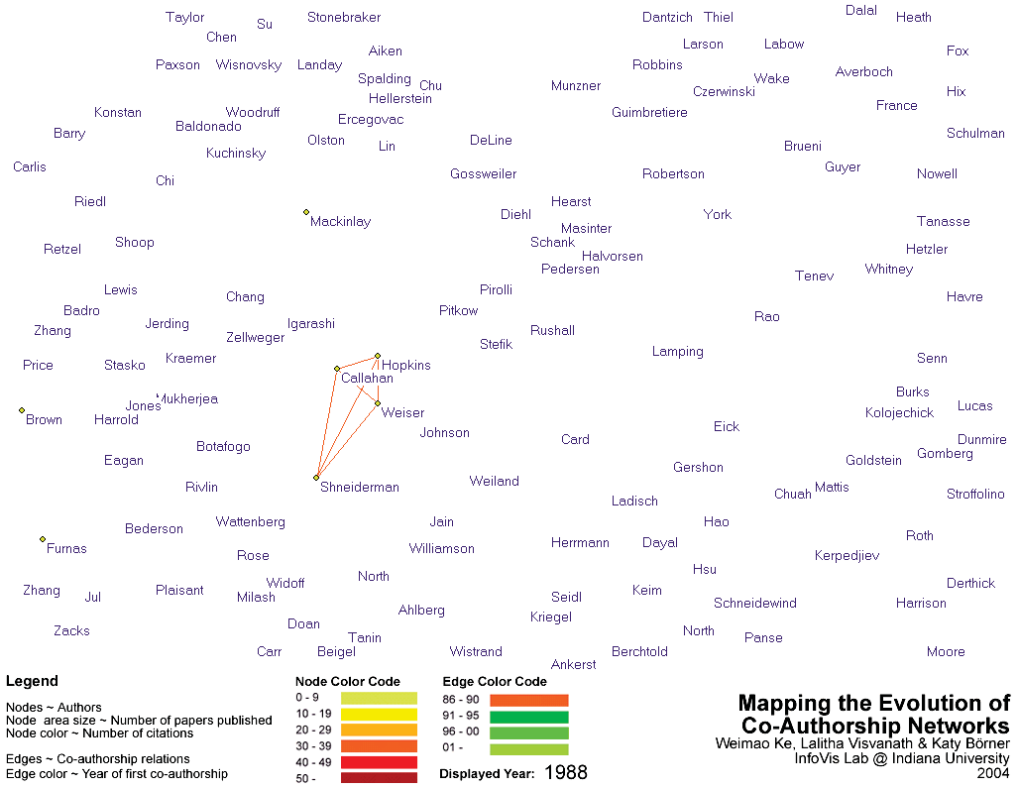
Individual Co-PI Network

Ke & Börner. 2006.



Mapping the Evolution of Co-Authorship Networks

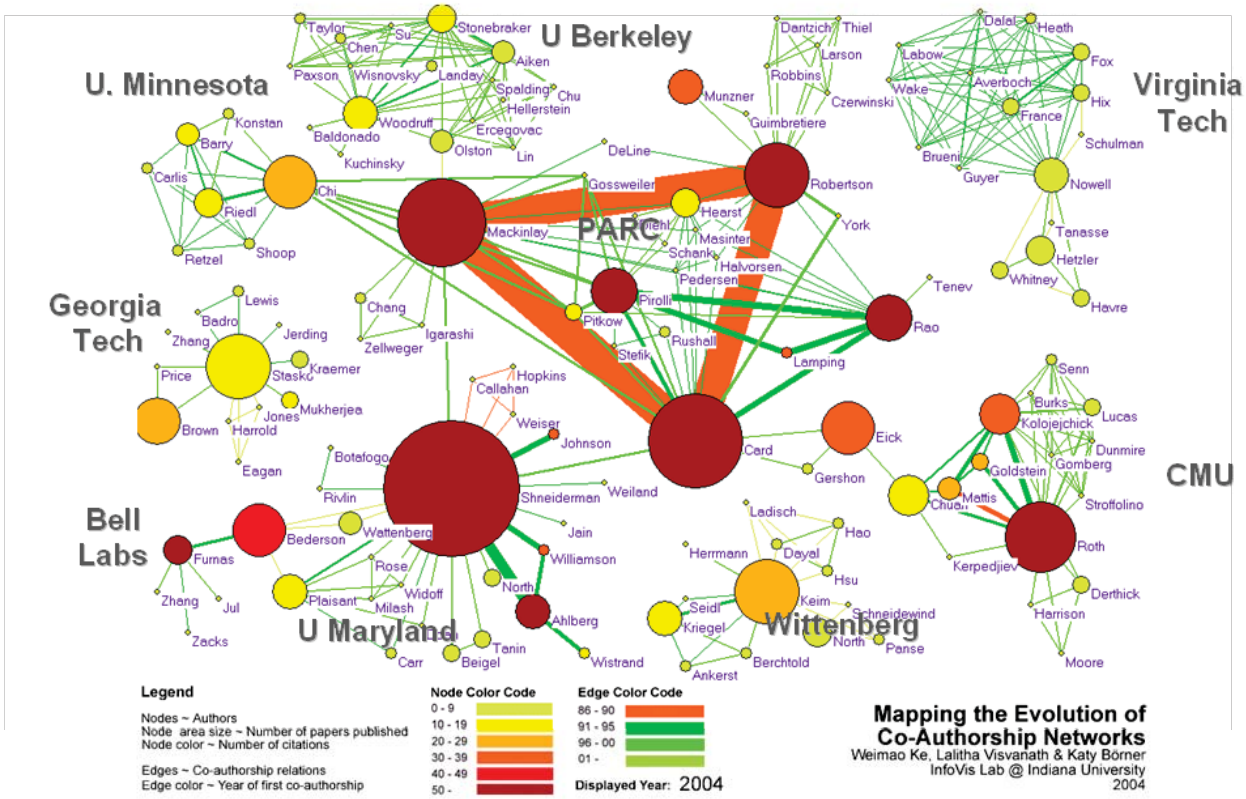
Ke, Visvanath & Börner. 2004. Won 1st prize at the IEEE InfoVis Contest.



Mapping the Evolution of Co-Authorship Networ

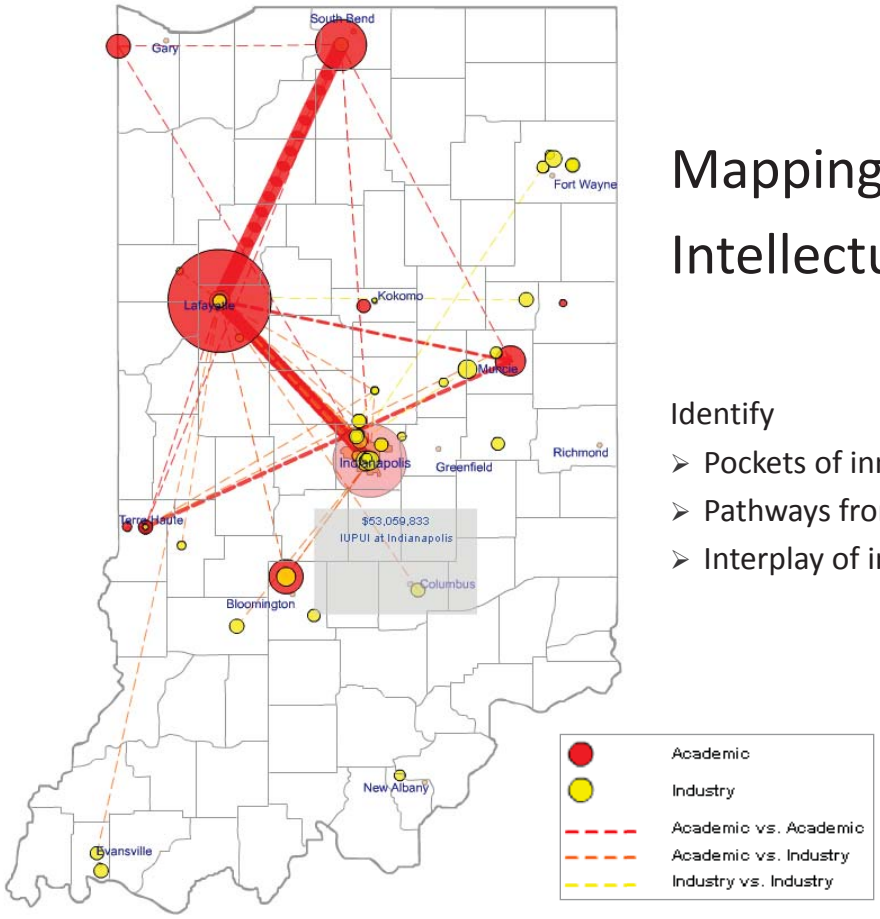
2

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



9

3



Mapping Indiana's Intellectual Space

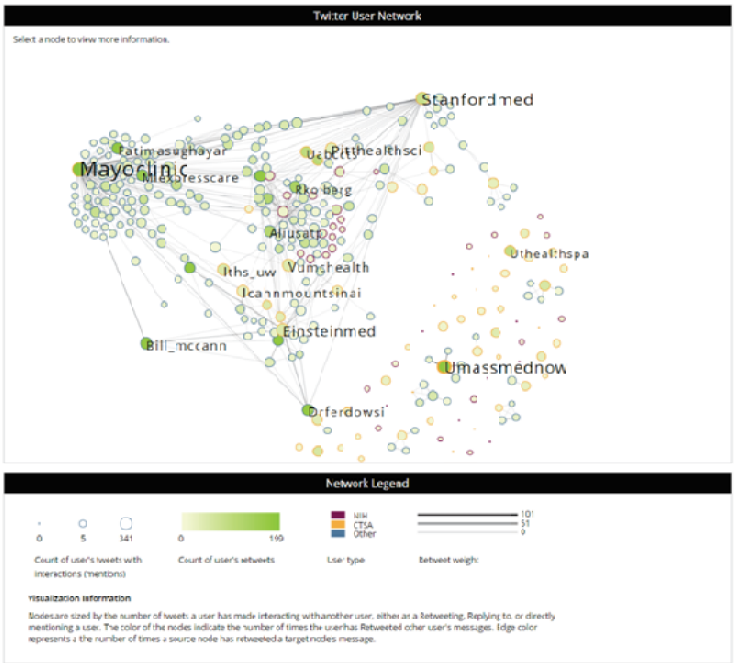
Identify

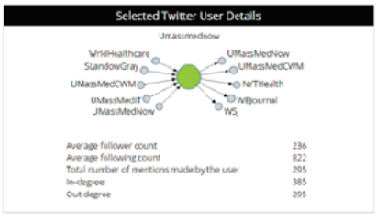
- Pockets of innovation
- Pathways from ideas to products
- Interplay of industry and academia

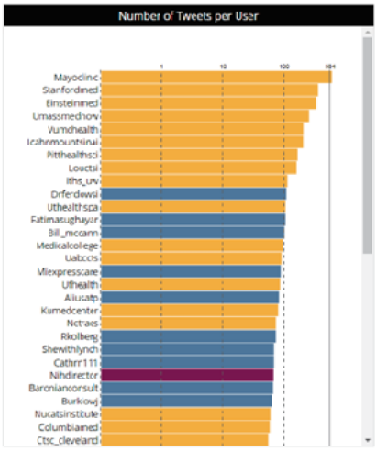
10

Twitter Network of NIH and STSI User Accounts

Tang, Börner, Liu, Light, & Simpson. 2016. SMS-VAT: A Scalable Multi-Scale Visual Analytical Tool

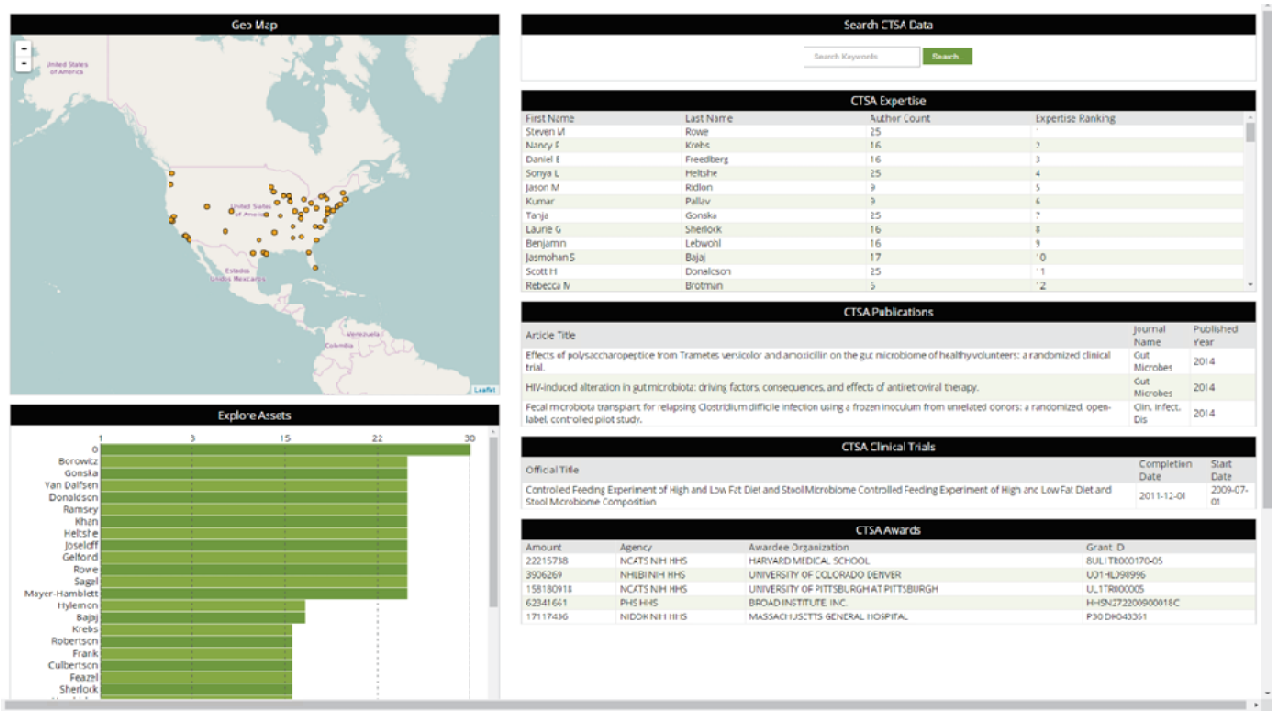






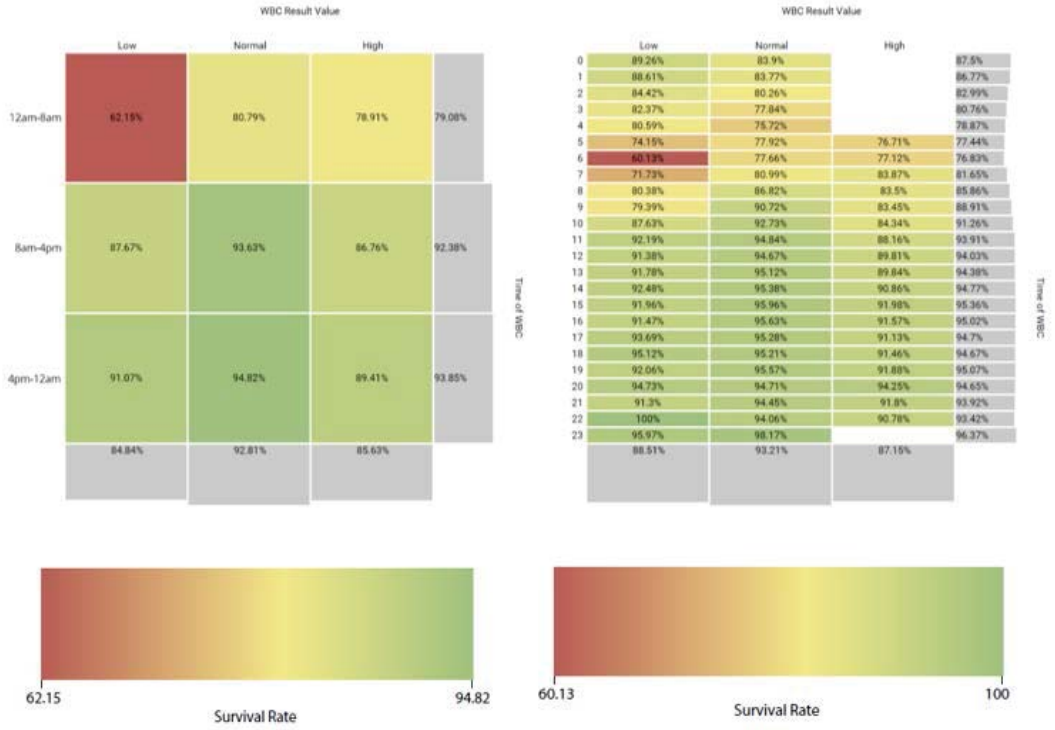
CTSI Expertise Visualization

Tang, Börner, Liu, Light, & Simpson. 2016. SMS-VAT: A Scalable Multi-Scale Visual Analytical Tool



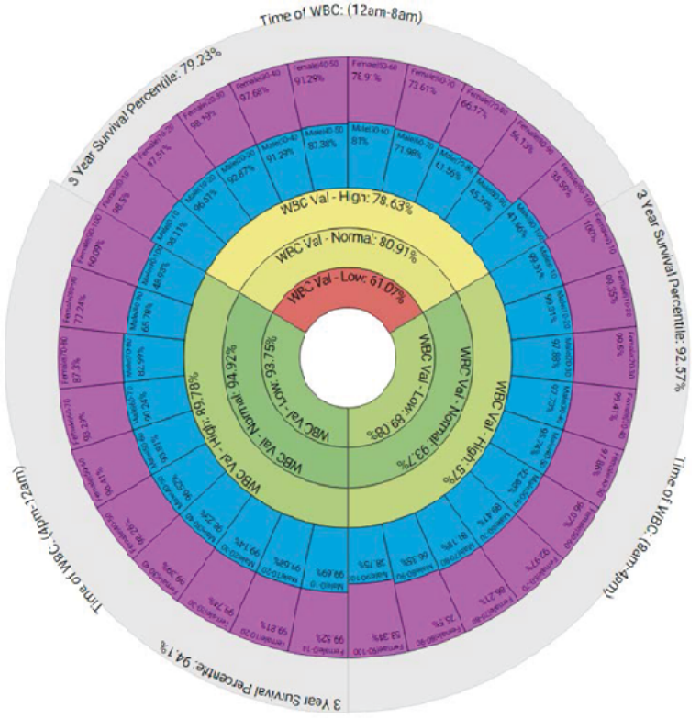
Mapping Big Biomedical Data

Weber and Börner. 2015. Visualizing Healthcare System Dynamics in Biomedical Big Data



Mapping Big Biomedical Data

Weber and Börner. 2015. Visualizing Healthcare System Dynamics in Biomedical Big Data



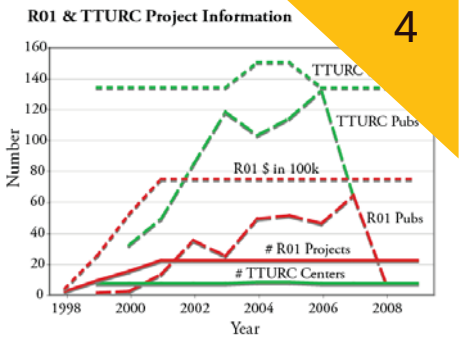
Mapping Transdisciplinary Tobacco Use Research Centers Publications

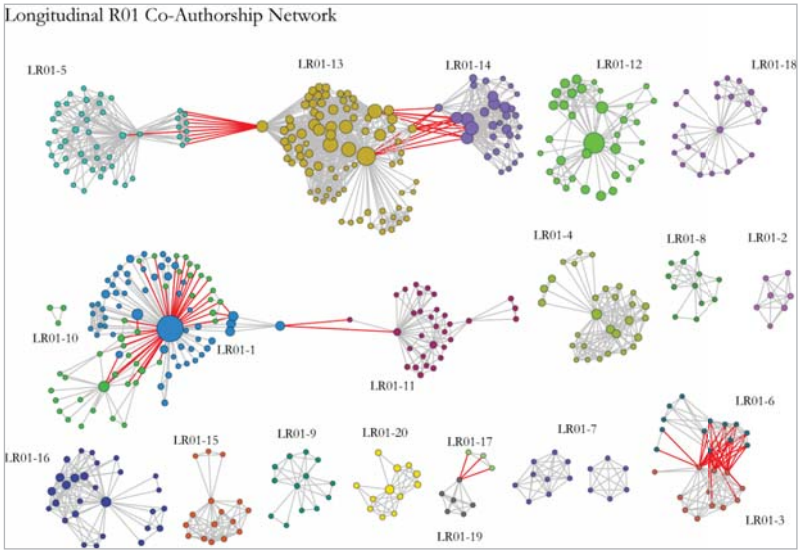
4

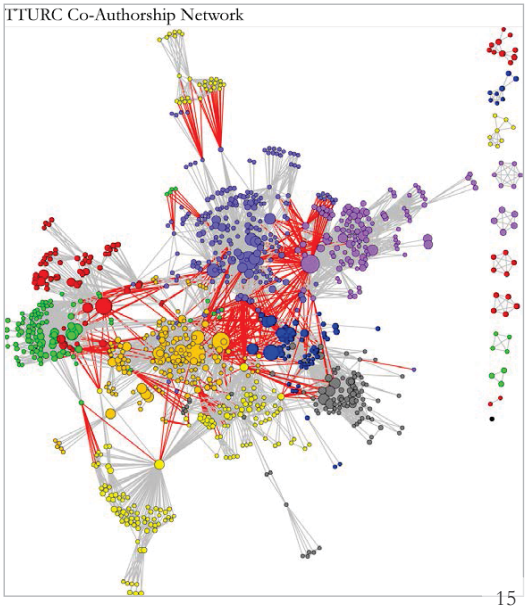
Compare R01 investigator-based funding with TTURC Center awards in terms of number of publications and evolving co-author networks.

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

Supported by NIH/NCI Contract HHSN261200800812





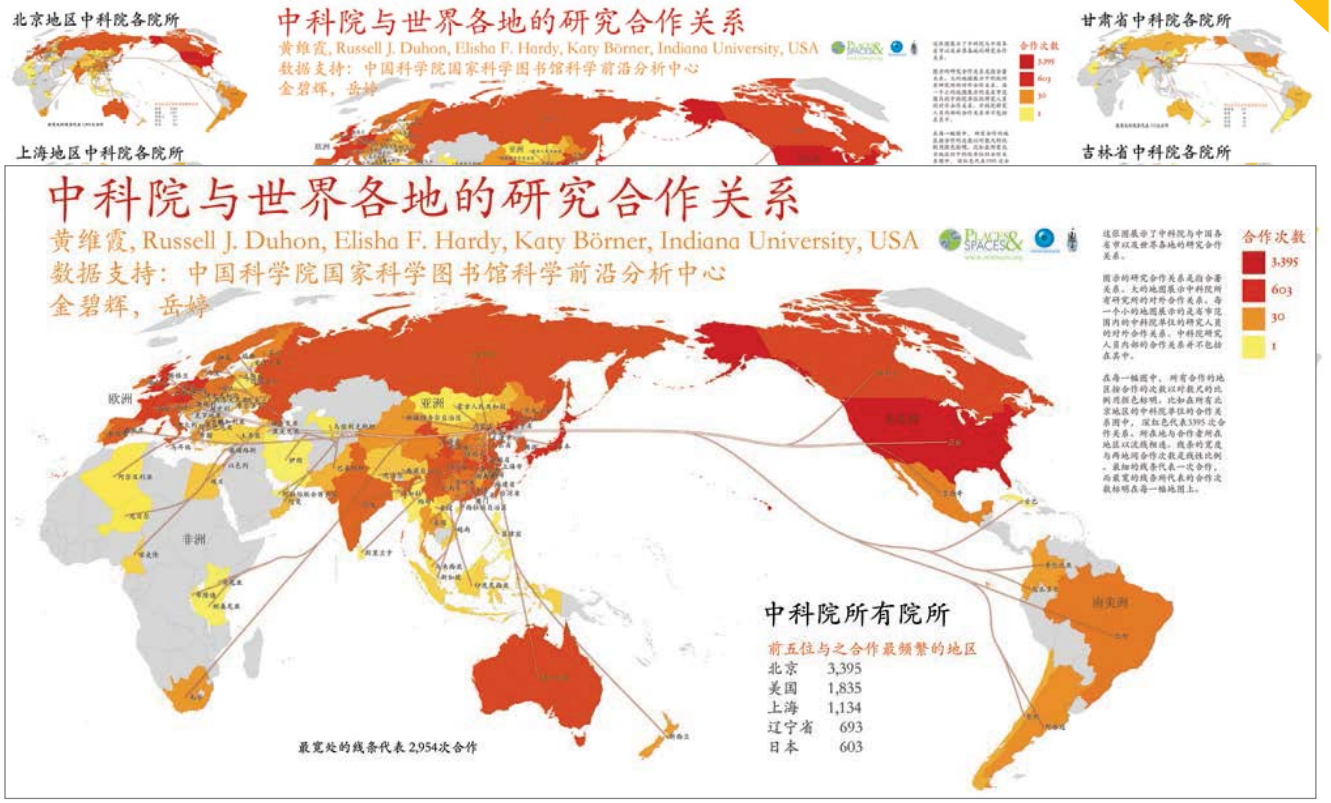


15

Research Collaborations by the Chinese Academy of Sciences

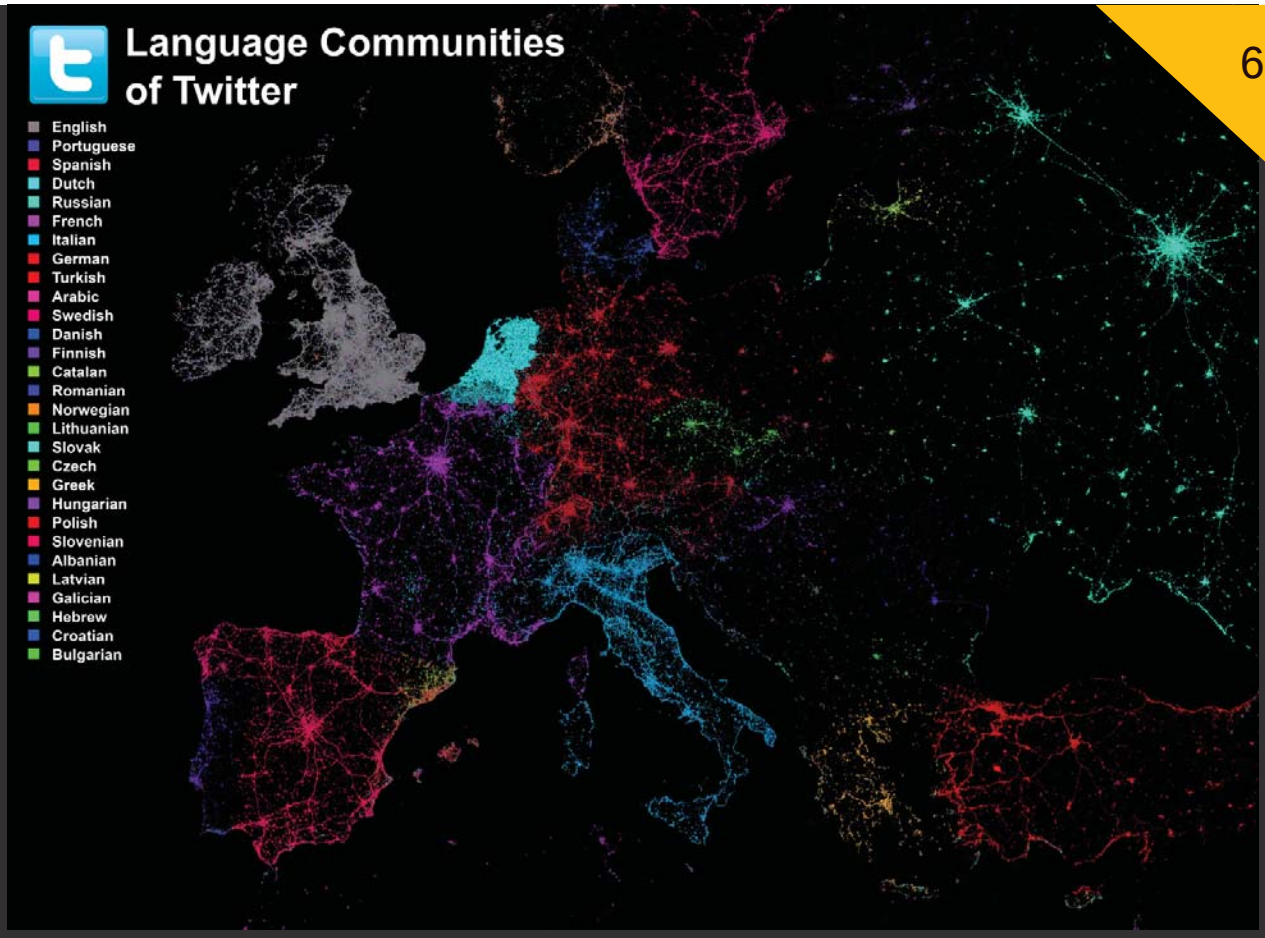
5

Huang, Duhon, Hardy & Börner



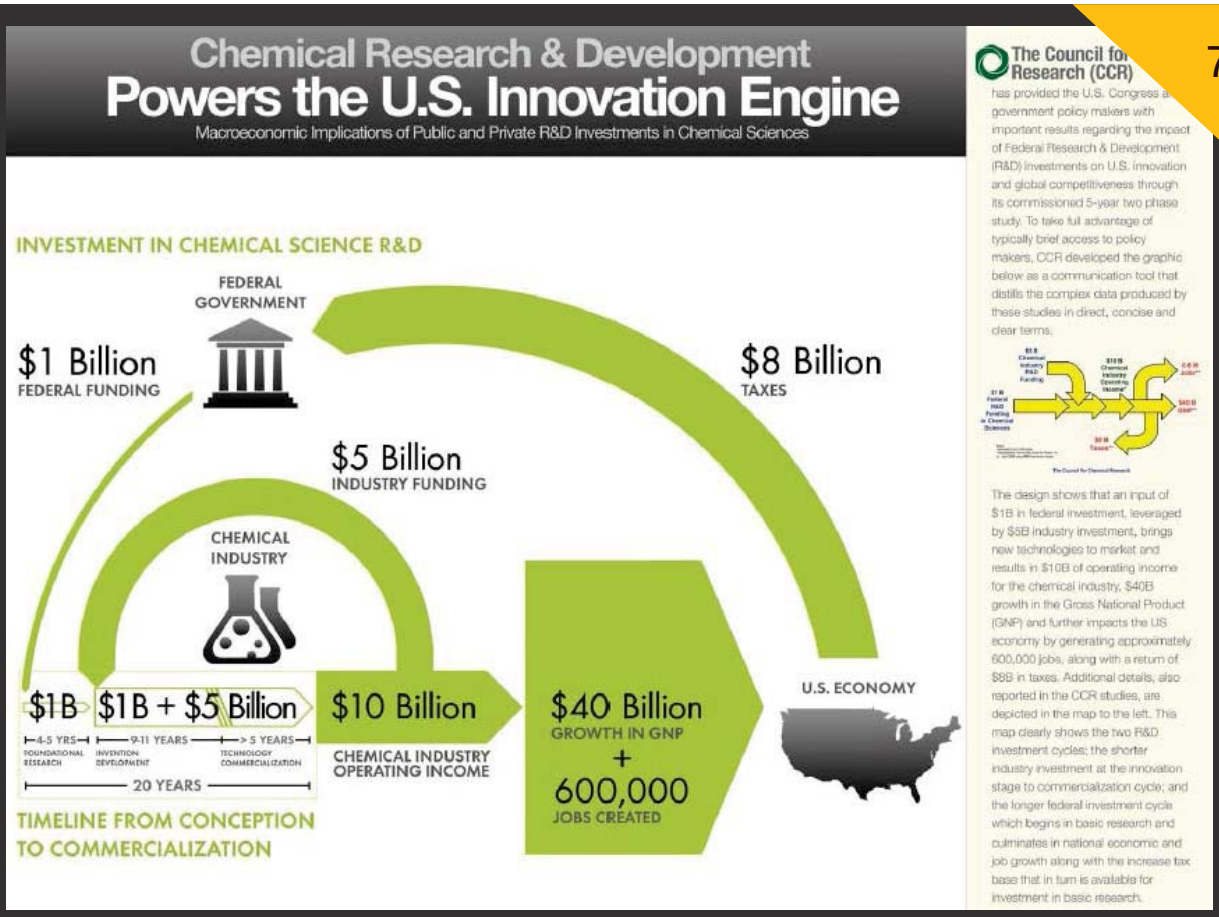
16

6



Language Communities of Twitter - Eric Fischer - 2012

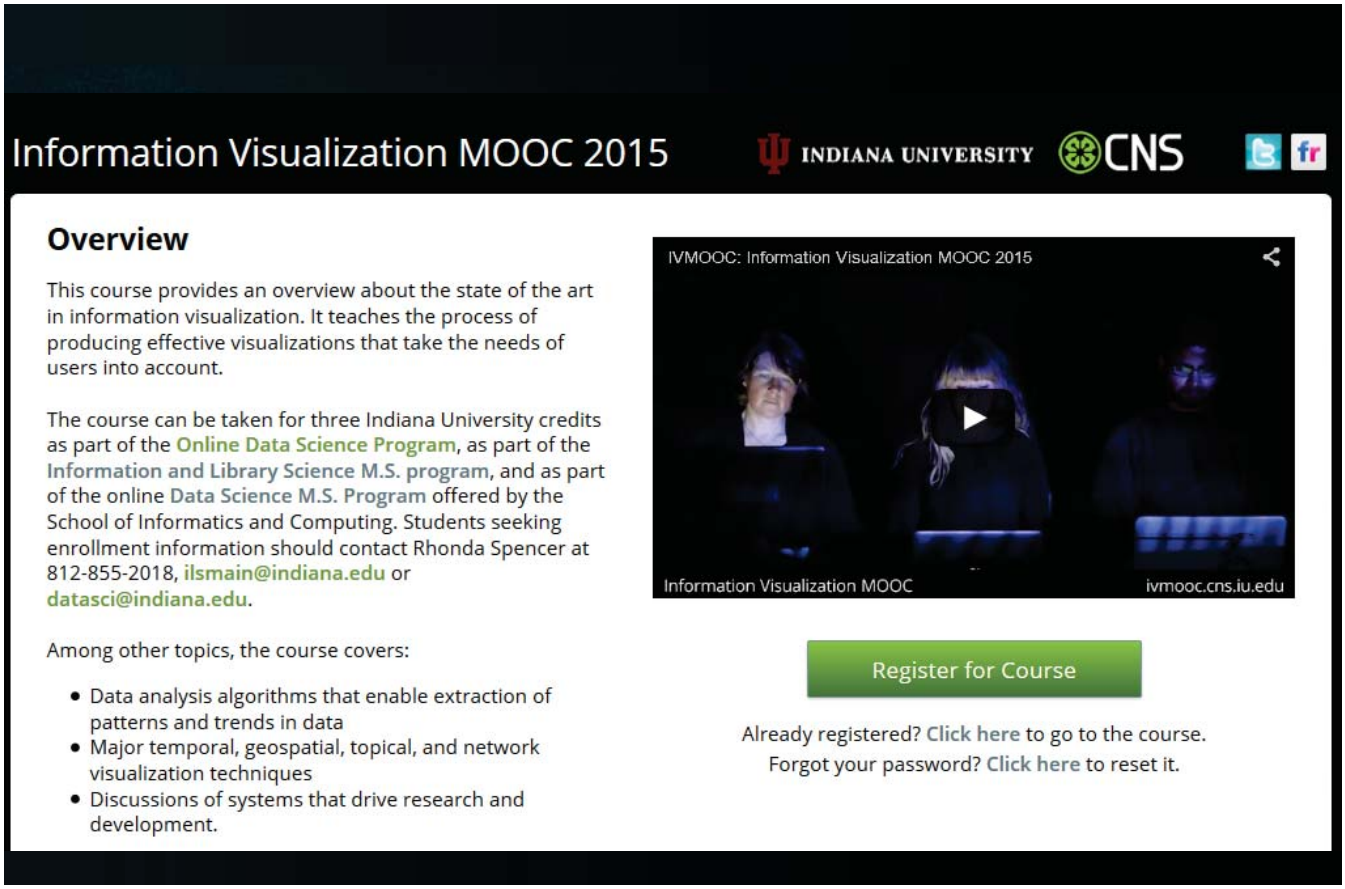
7



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

Empowering Anyone to Visualize STI

Example: The Information Visualization MOOC



Register for free at http://ivmooc.cns.iu.edu. Class restarted in January 13, 2015.

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

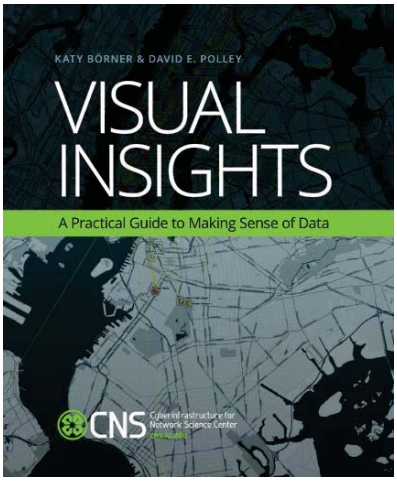
- **Session 5** – "With Whom:" Trees
- **Session 6** – "With Whom:" Networks
- **Session 7** – Dynamic Visualizations and Deployment

Final Exam

Part 2: Students work in teams on client projects.

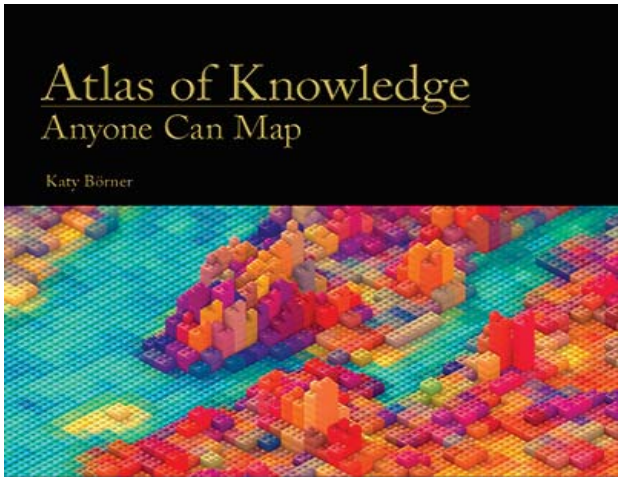
Final grade is based on Class Participation (10%), Midterm (30%), Final Exam (30%), and Client Project(30%).

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.

Visualization Frameworks

How to Classify Different Visualizations?

By

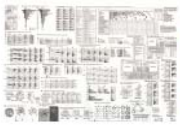
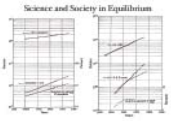
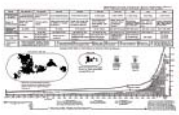
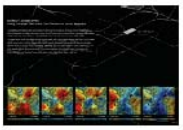
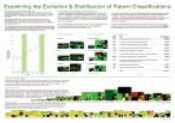
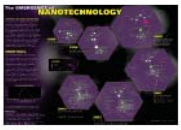
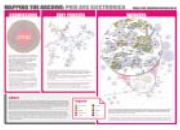
- User insight needs?
- User task types?

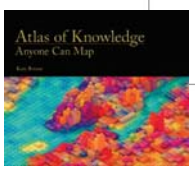
- Data to be visualized?
- Data transformation?

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

- Or ?

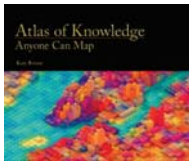
Tasks

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



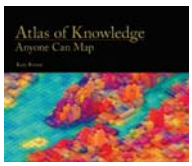
See page 5

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



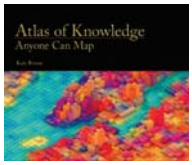
See page 24

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



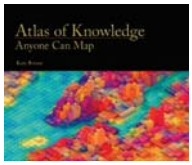
See page 24

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.

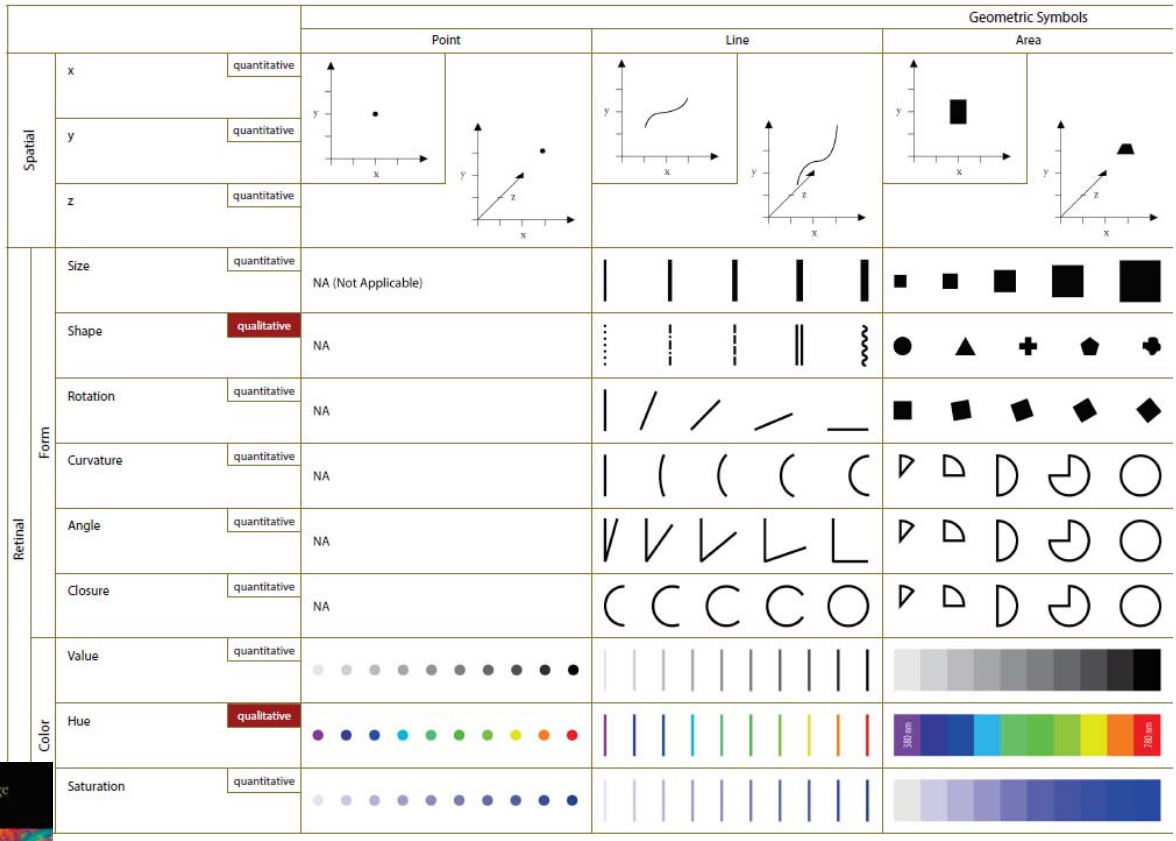
Types

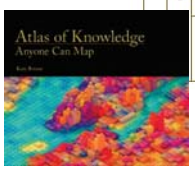
Insight Need Types page 26	Data Scale Types page 28	Visualization Types page 30	Graphic Symbol Types page 32	Graphic Variable Types page 34	Interaction Types page 26
• categorize/cluster • order/rank/sort • distributions (also outliers, gaps) • comparisons • trends (process and time) • geospatial • compositions (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

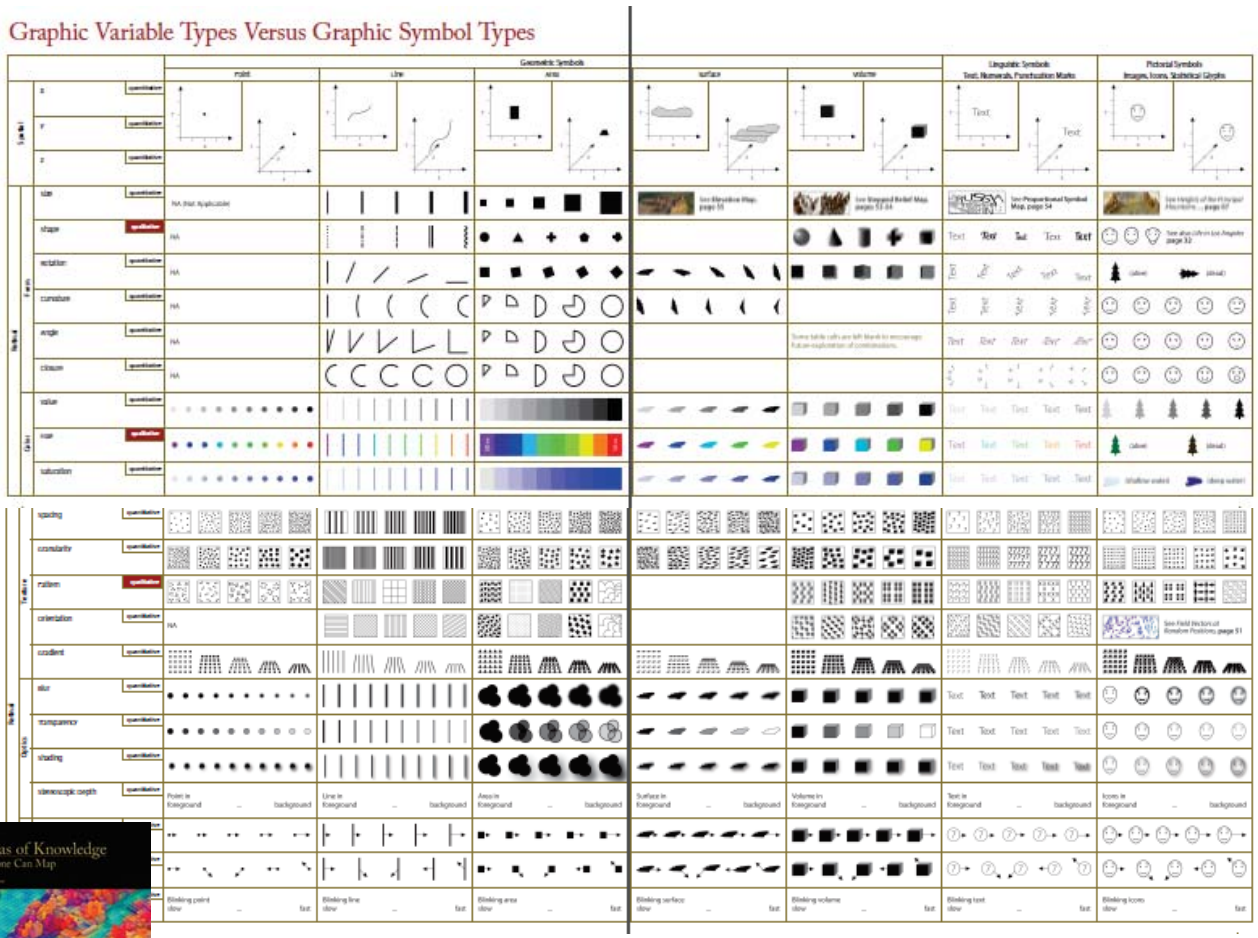
Graphic Variable Types Versus Graphic Symbol Types





See page 36

Graphic Variable Types Versus Graphic Symbol Types



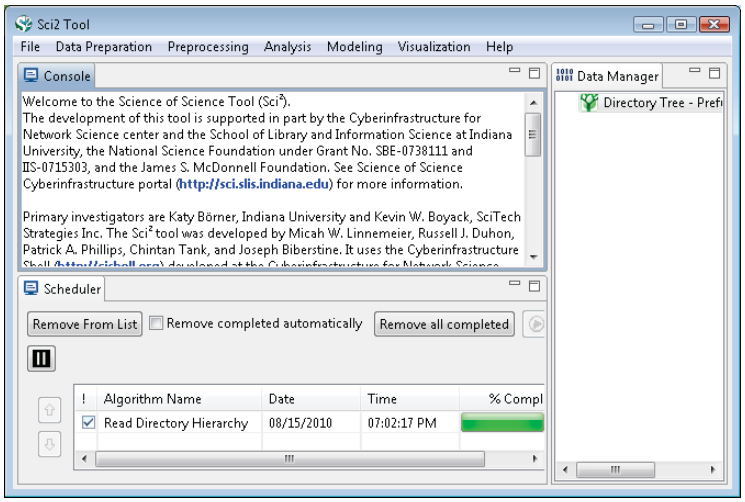
See pages 36-39

Sci2 Tool – OSGi/CIShell-based Macroscope

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

Use

- **Menu** to read data, run algorithms.
- **Console** to see work log, references to seminal works.
- **Data Manager** to select, view, save loaded, simulated, or derived datasets.
- **Scheduler** to see status of algorithm execution.



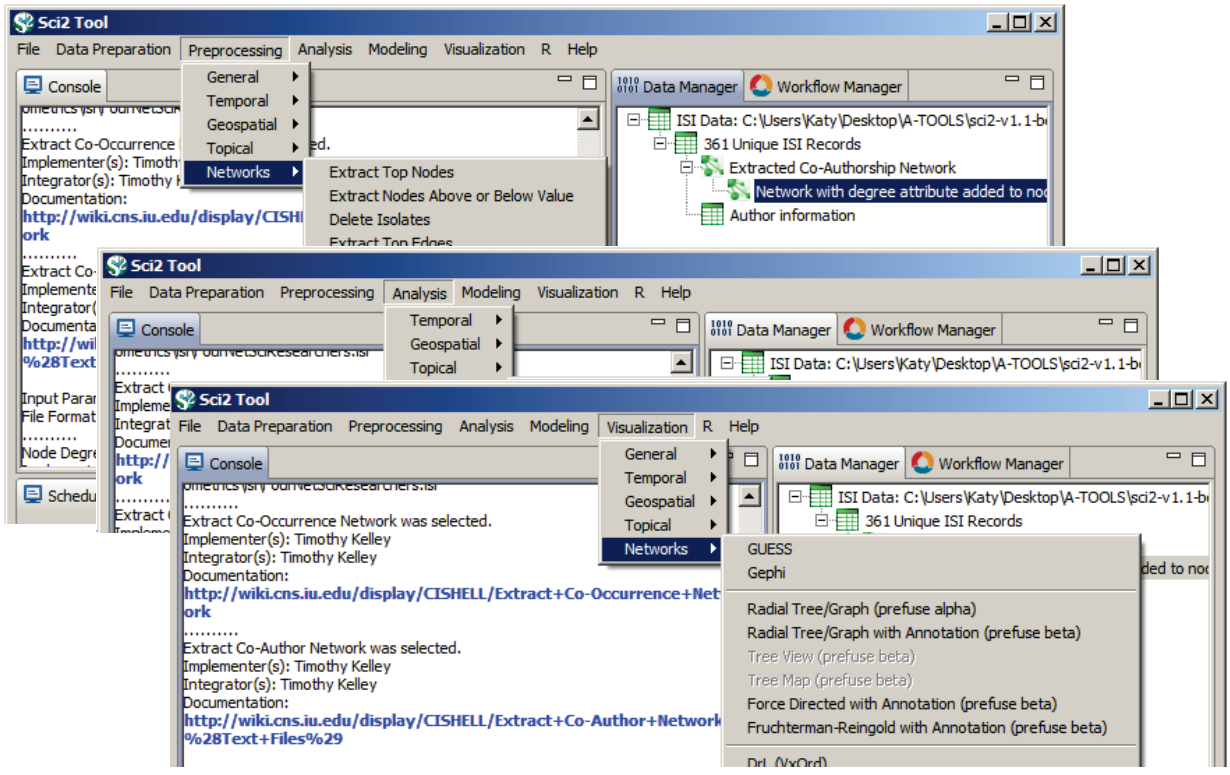
All workflows are recorded into a log file (see /sci2/logs/…), and can be re-run for easy replication. If errors occur, they are saved in a error log to ease bug reporting.

All algorithms are documented online; workflows are given in Sci2 Manual at http://sci2.wiki.cns.iu.edu

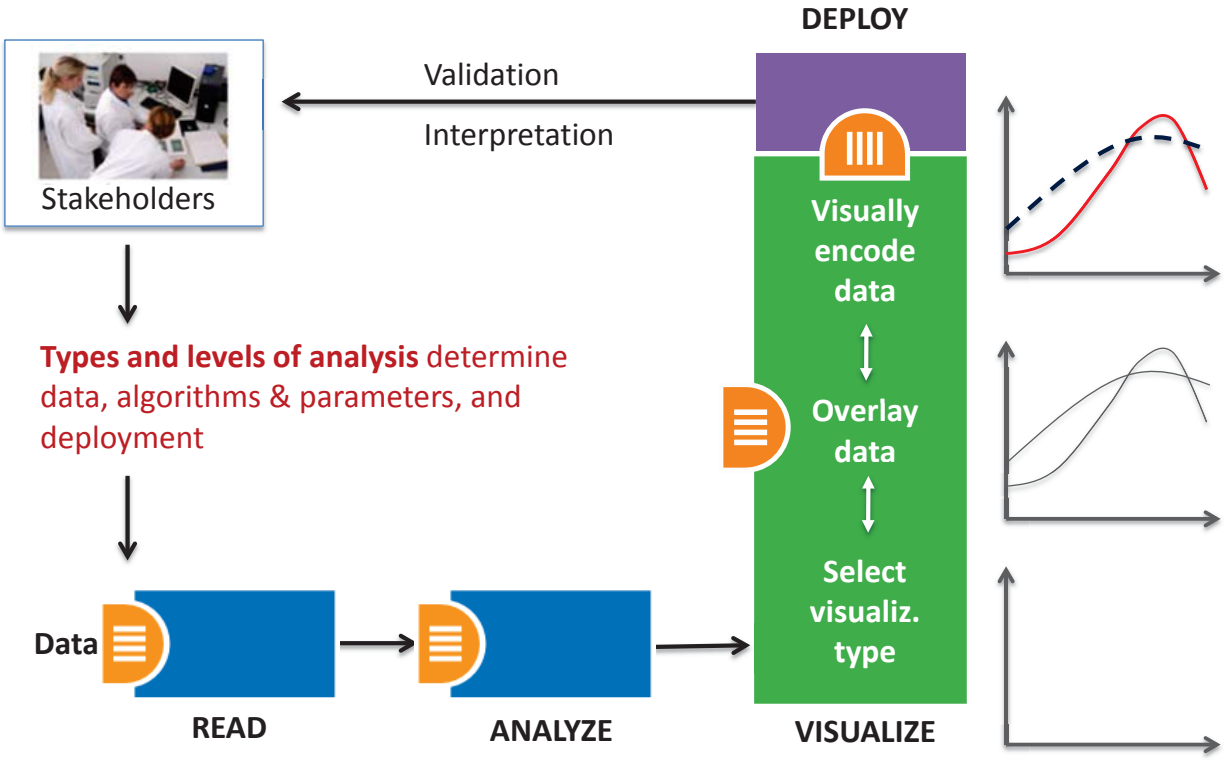
CNS Cyberinfrastructure for Network Science Center

Sci2 Tool Interface Components

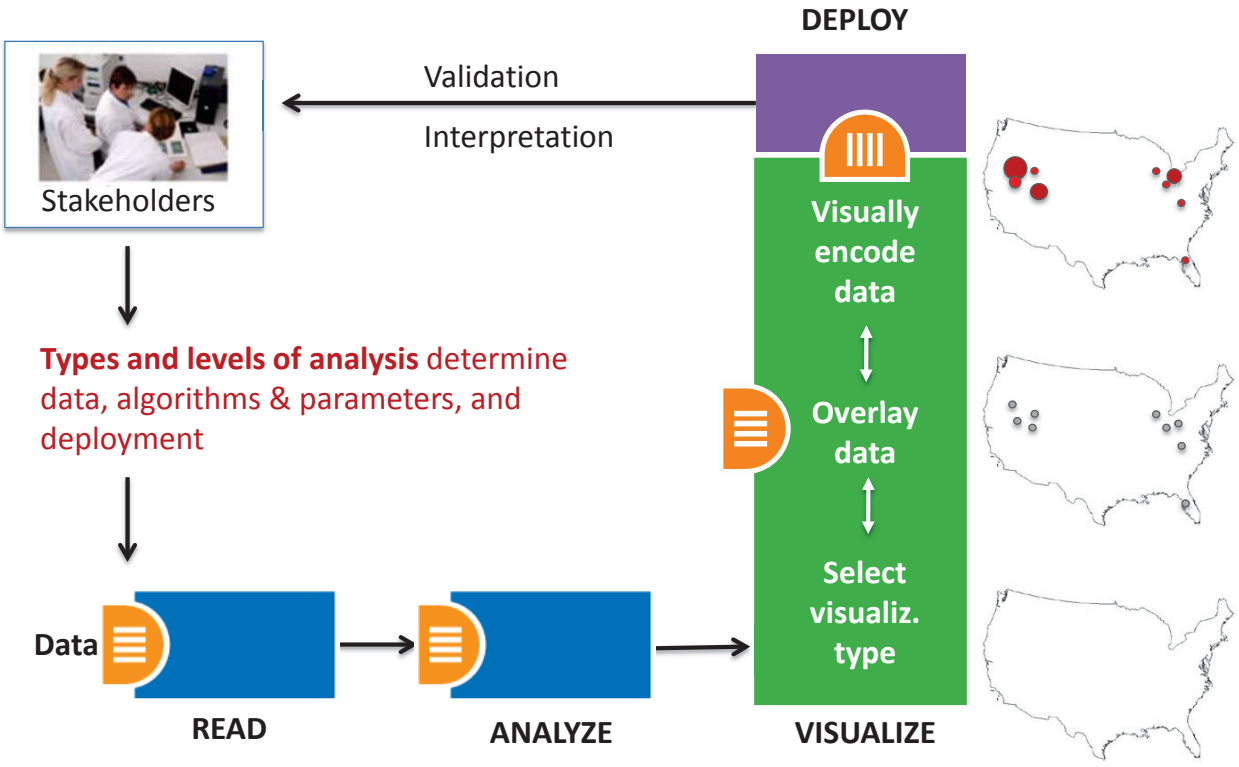
Download for free at http://sci2.cns.iu.edu



Needs-Driven Workflow Design



Needs-Driven Workflow Design



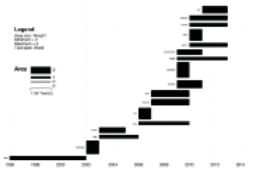
Load **One** File and Run **Many** Analyses and Visualizations

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

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



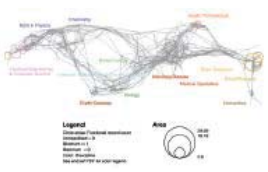
Geospatial Analysis–p. 52

Geospatial Analysis–p. 52

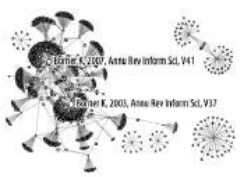
Load **One** File and Run **Many** Analyses and Visualizations

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

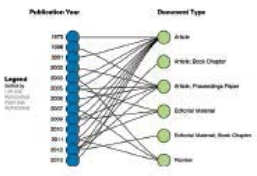
Topical Analysis–p. 56



Paper Citation Network–p. 60



Bi-Modal Network–p. 60



Co-author and many other bi-modal networks.

Existing Interfaces for Health-related Data

9) NIH RePORTER: Visual Interface to Biomedical Funding Data in U.S.

10) CIShell/Sci2 World and Science Visualizations of NIH RePORTER Data

11) NIH RePORTER: NIH Map

12) BBSRC: Visual Interface to Biomedical Funding Data in UK

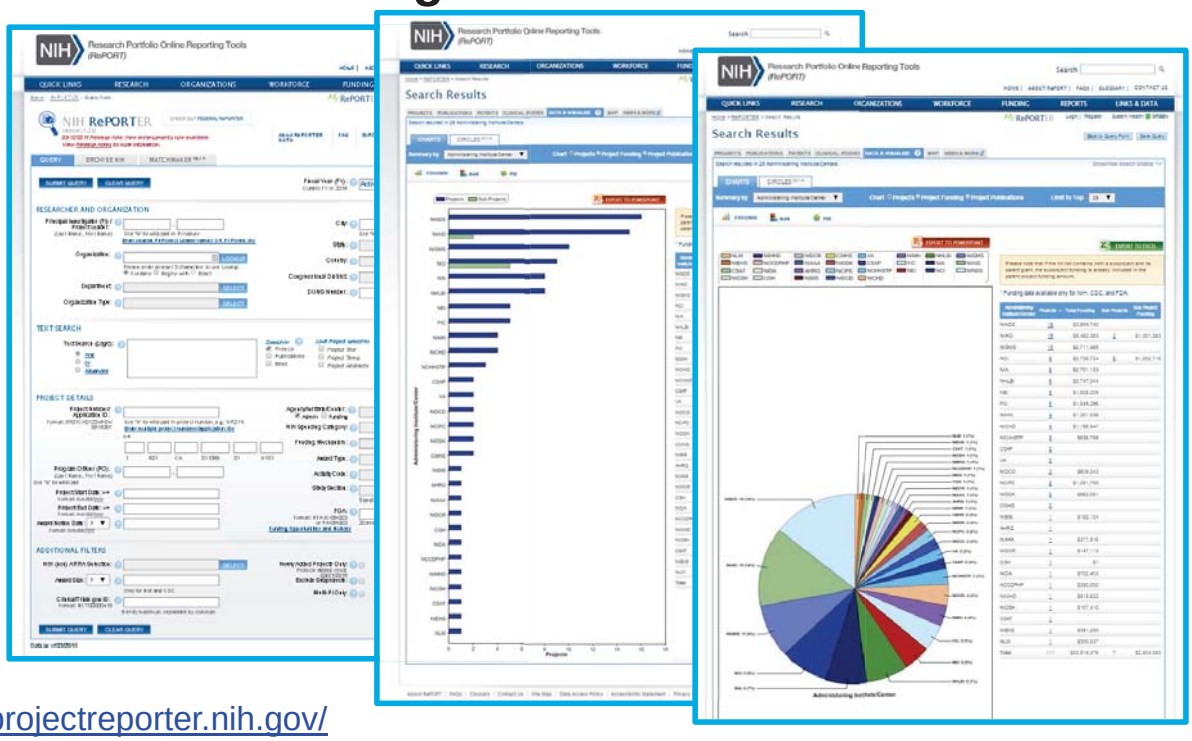
13) IAI Multidimensional Analysis

14) Scraawl: Twitter Analysis

15) Illuminated Diagram: Searchable World and Science Maps

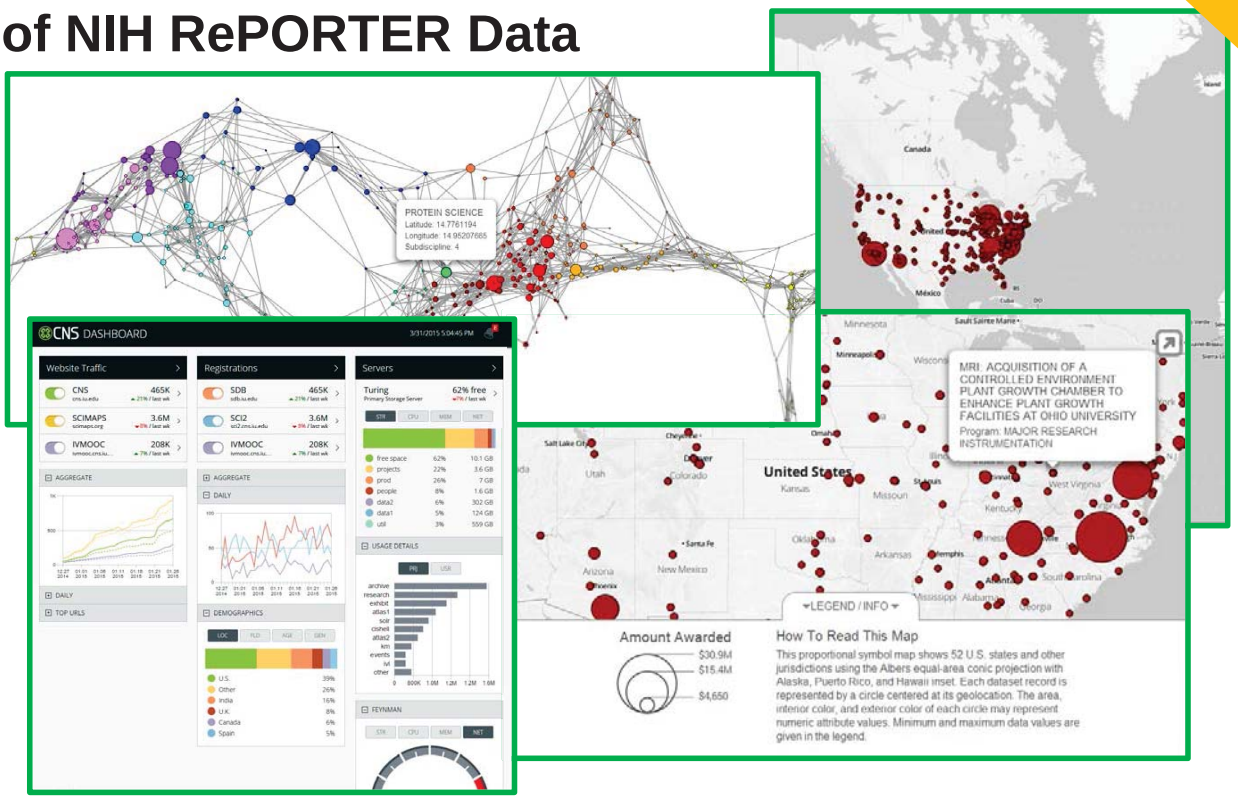
CNS Cyberinfrastructure for Network Science Center

NIH RePORTER: Visual Interface to Biomedical Funding Data in US



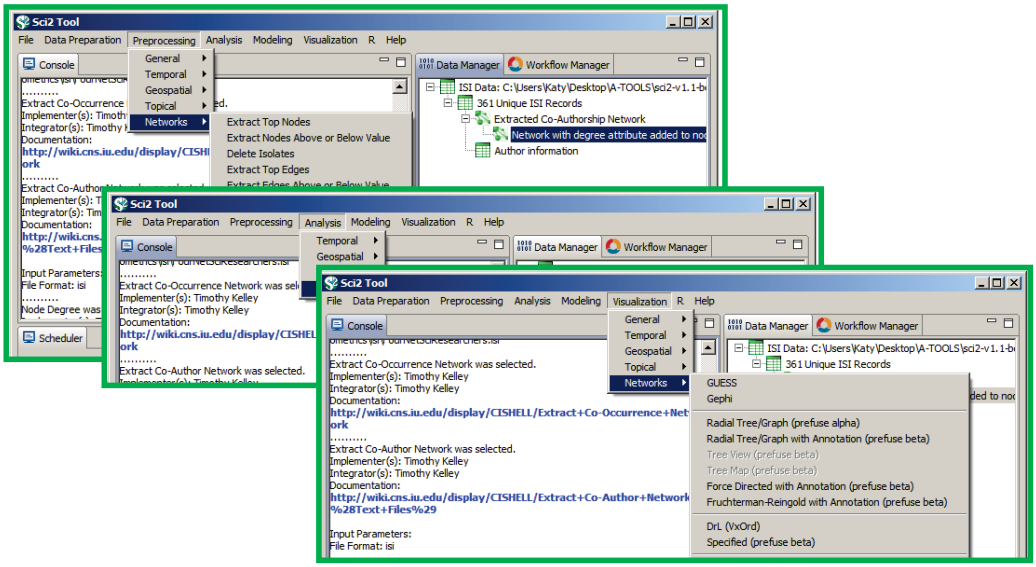
http://projectreporter.nih.gov/

CIShell/Sci2 World and Science Visualizations of NIH RePORTER Data

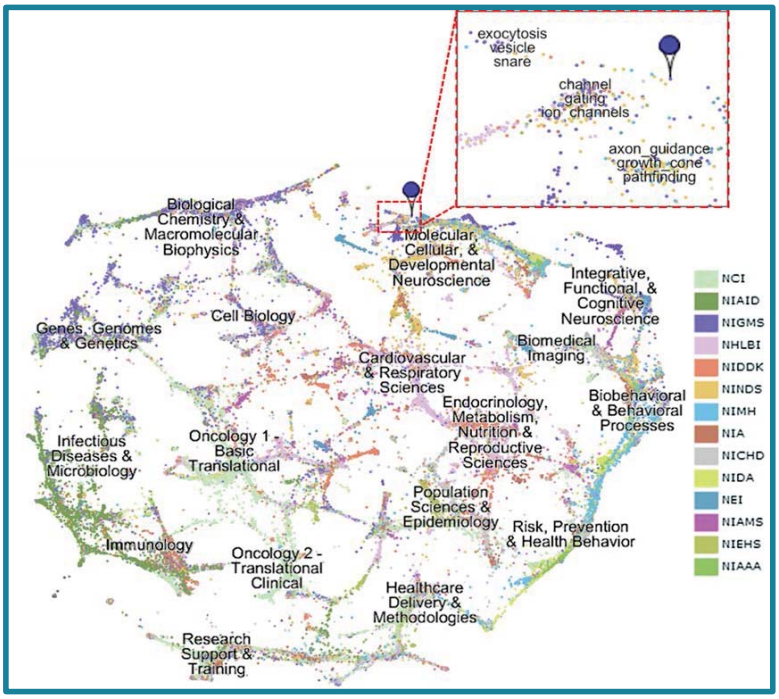


CNS Cyberinfrastructure for Network Science Center

Sci2 Desktop

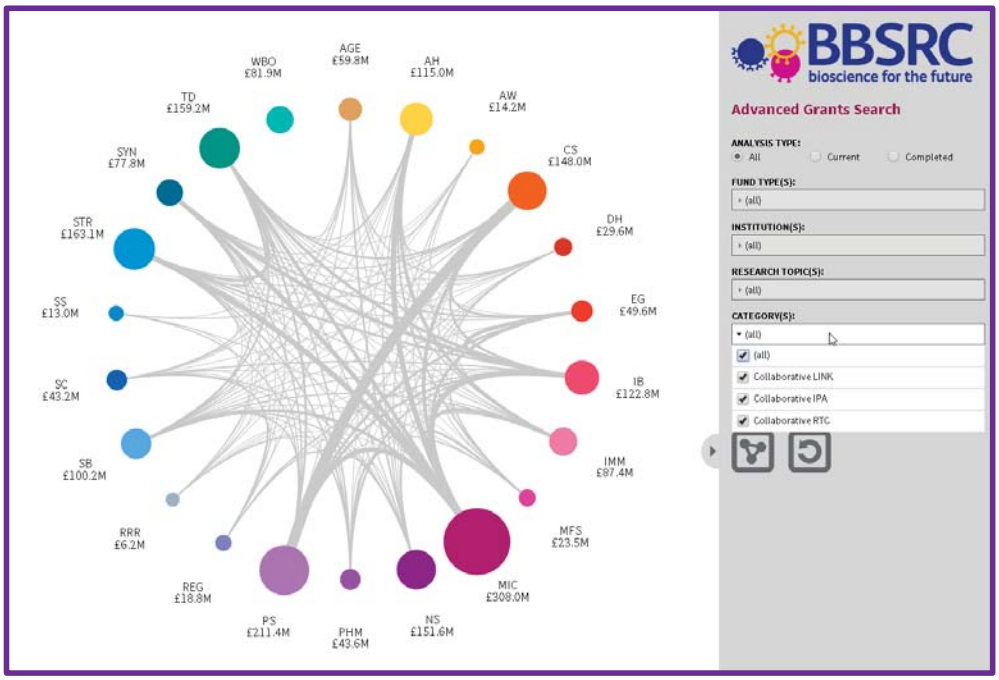


NIH RePORTER: NIH Map



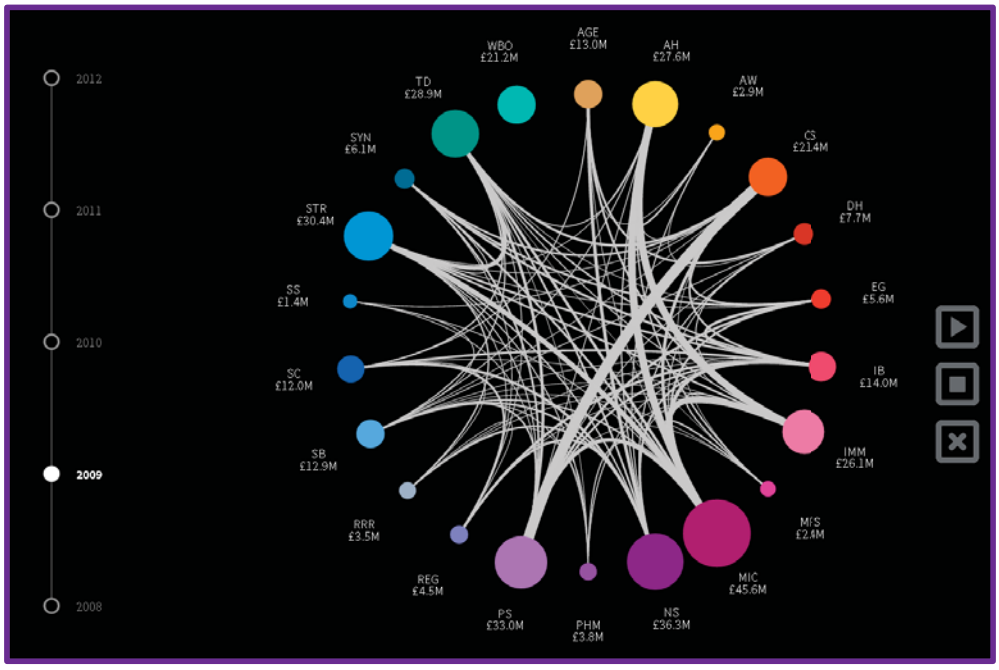
http://nihmaps.org/

BBSRC: Visual Interface to Biomedical Funding Data in UK



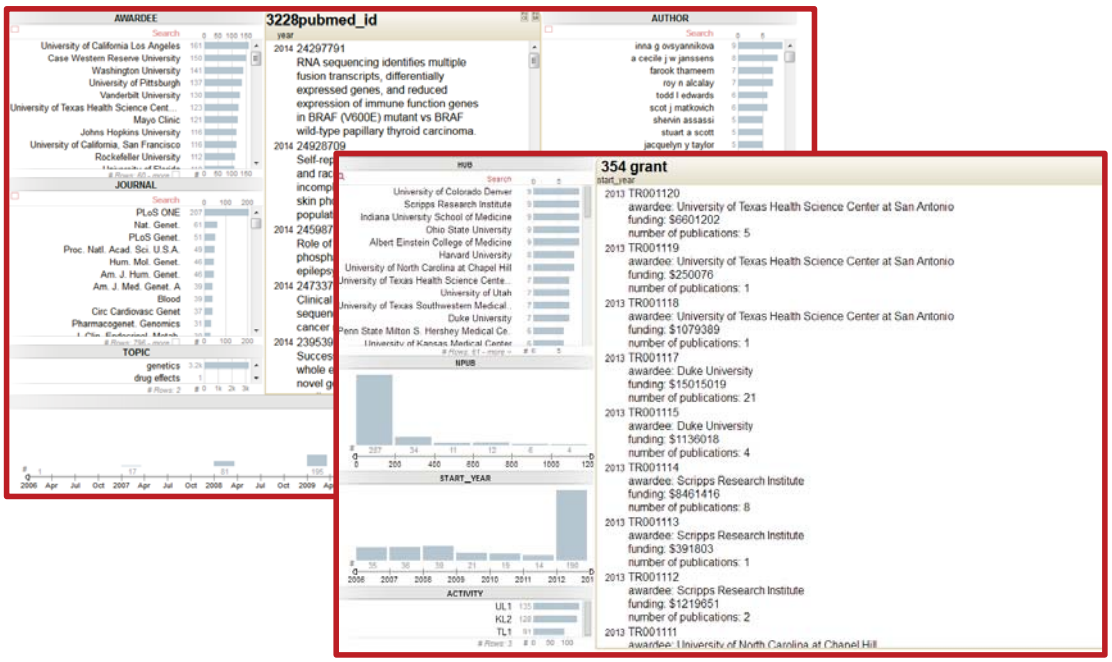
http://www.bbsrc.ac.uk/

BBSRC: Temporal animation

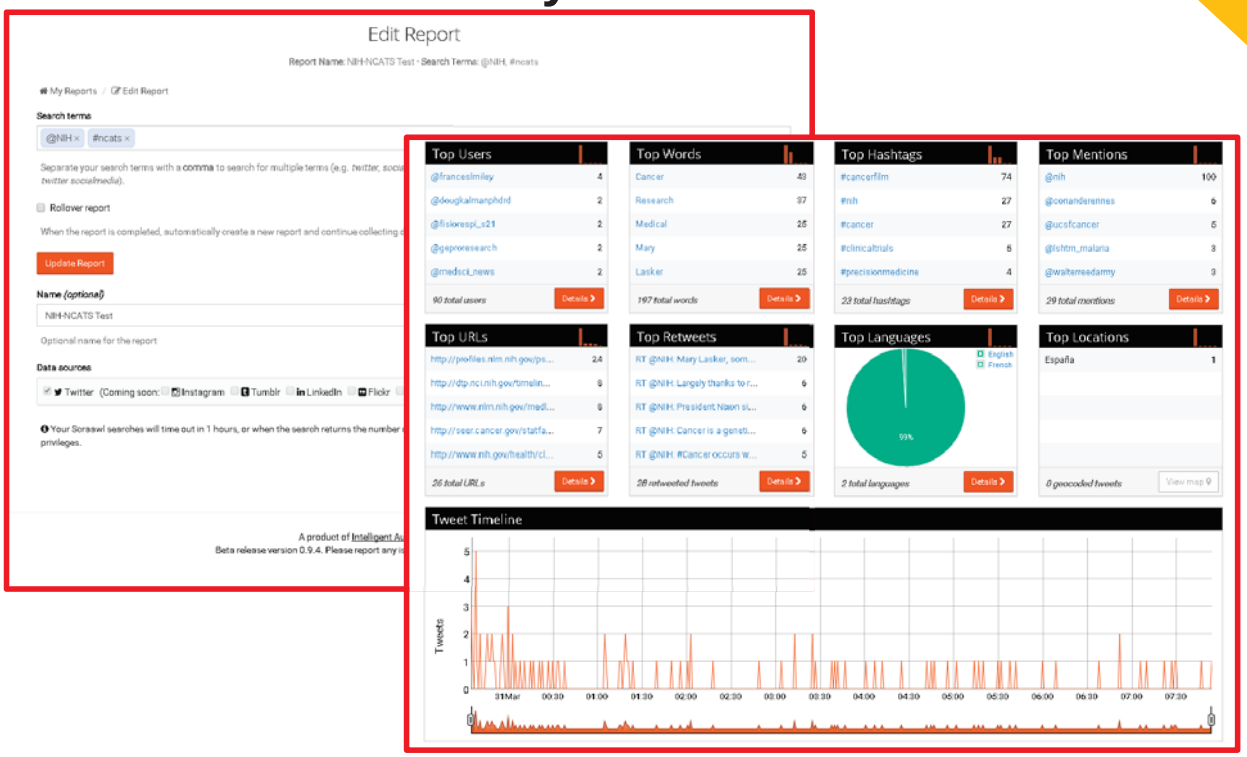


CNS Cyberinfrastructure for Network Science Center

IAI Multidimensional Analysis

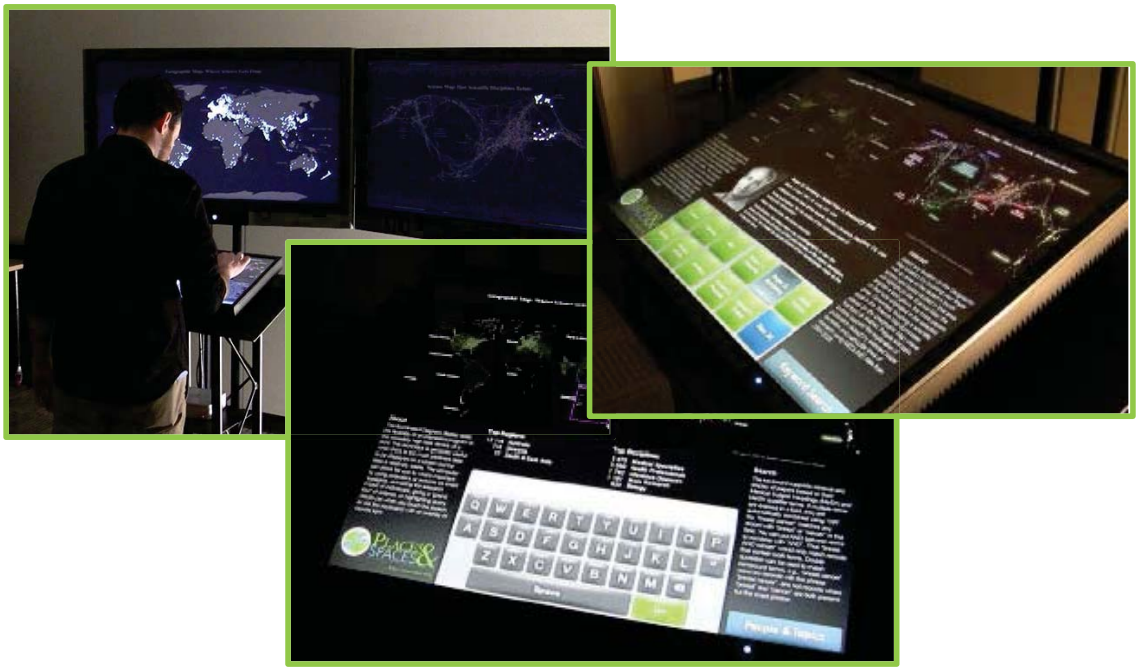


Scraawl: Twitter Analysis



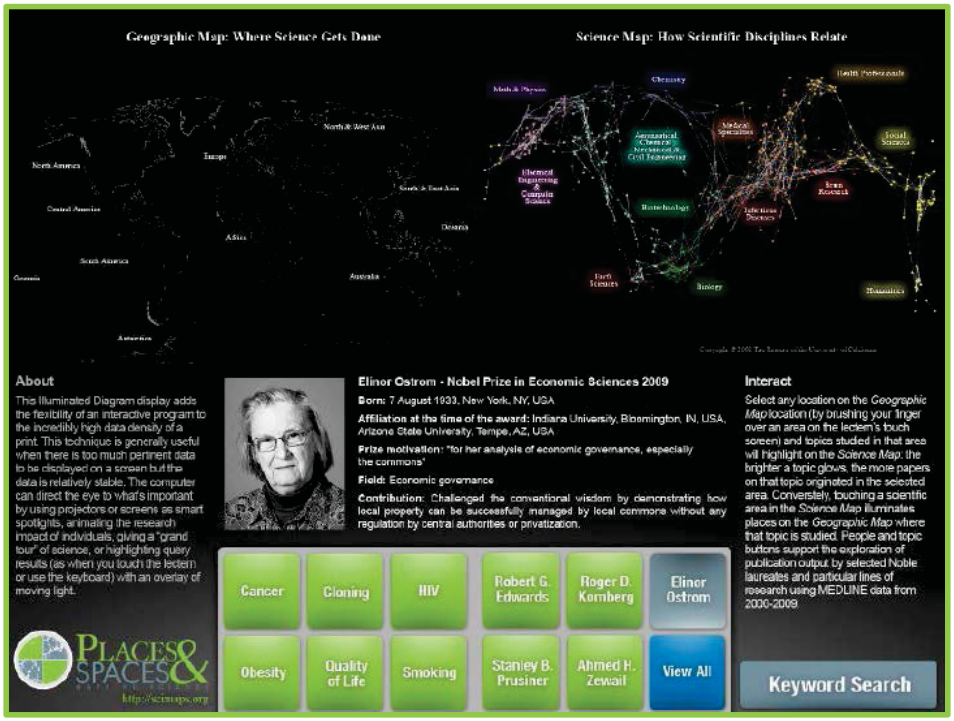
https://www.scraawl.com/

Illuminated Diagram: Searchable World and Science Maps



http://cns.iu.edu/interactive_displays.html

Illuminated Diagram: Search detail



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/

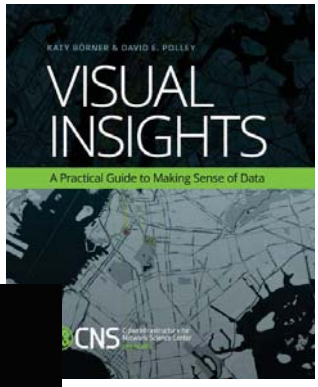
Börner, Katy (2010) **Atlas of Science: Visualizing What We Know**. The MIT Press. http://scimaps.org/atlas

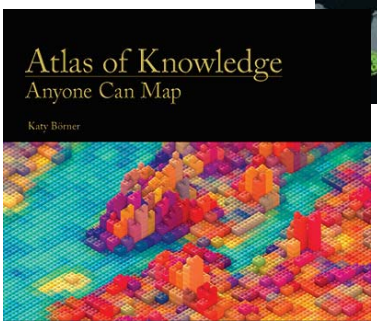
Scharnhorst, Andrea, Börner, Katy, van den Besselaar, Peter (2012) **Models of Science Dynamics**. Springer Verlag.

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 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