

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

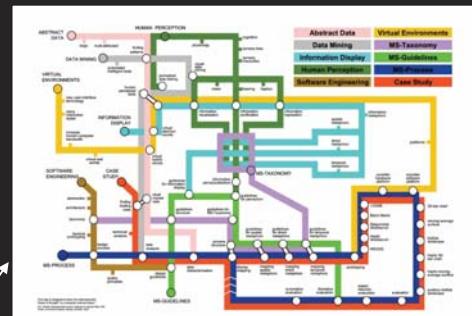
Center for Computational Research, SUNY Buffalo
March 29, 2016

Language Communities of Twitter - Eric Fischer - 2012



Terra bytes of data

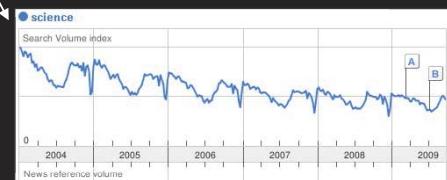
Descriptive &
Predictive
Models



Find your way



Find collaborators, friends



Identify trends

Descriptive Models

Multiple levels: Micro ... Macro

Answering: When? Where? What? With Whom?

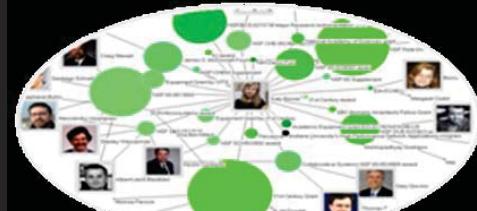
3

Different Levels of Abstraction/Analysis

Macro/Global
Population Level



Meso/Local
Group Level



Micro
Individual Level



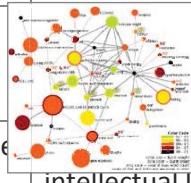
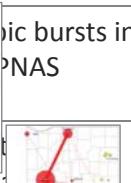
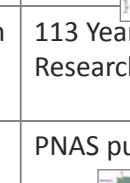
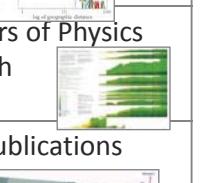
4

Type of Analysis vs. Level of Analysis

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

5

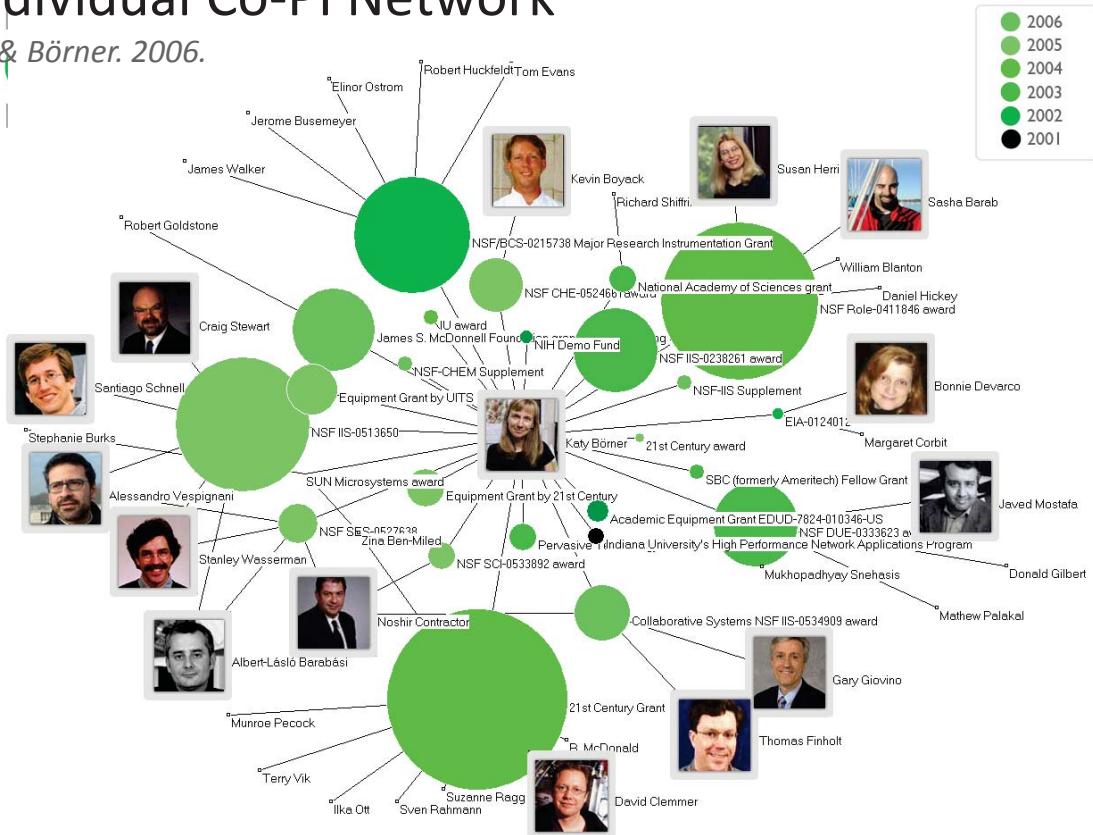
Type of Analysis vs. Level of Analysis

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

6

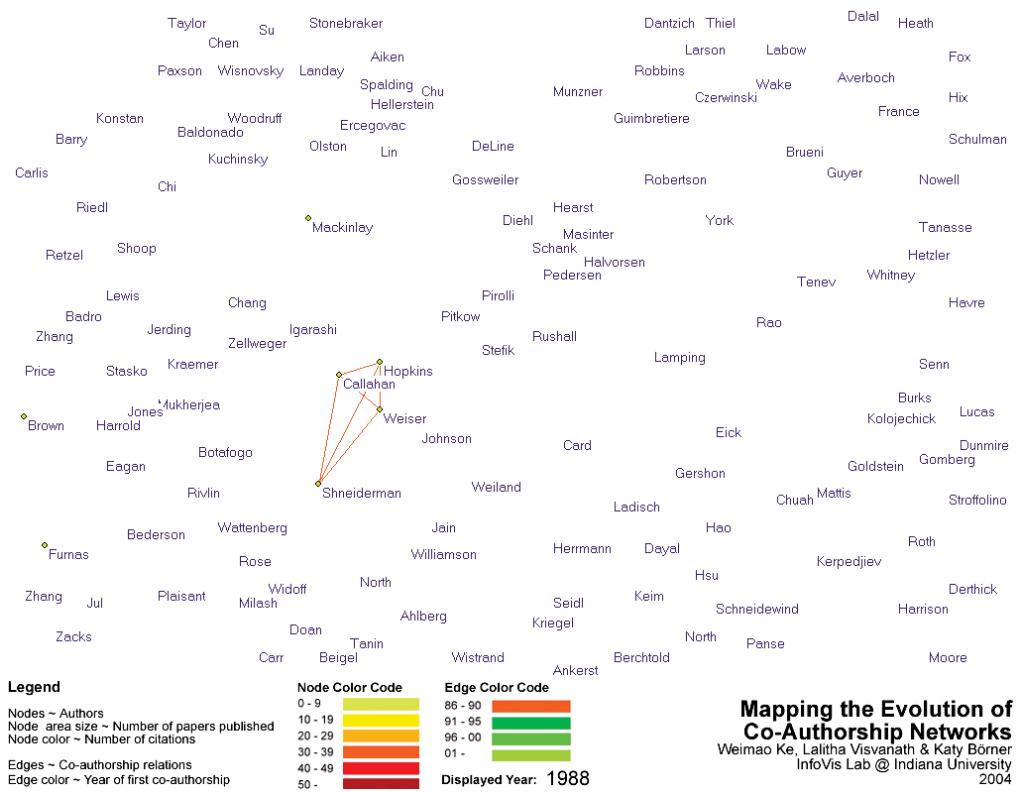
Individual Co-PI Network

Ke & Börner. 2006.



Mapping the Evolution of Co-Authorship Networks

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

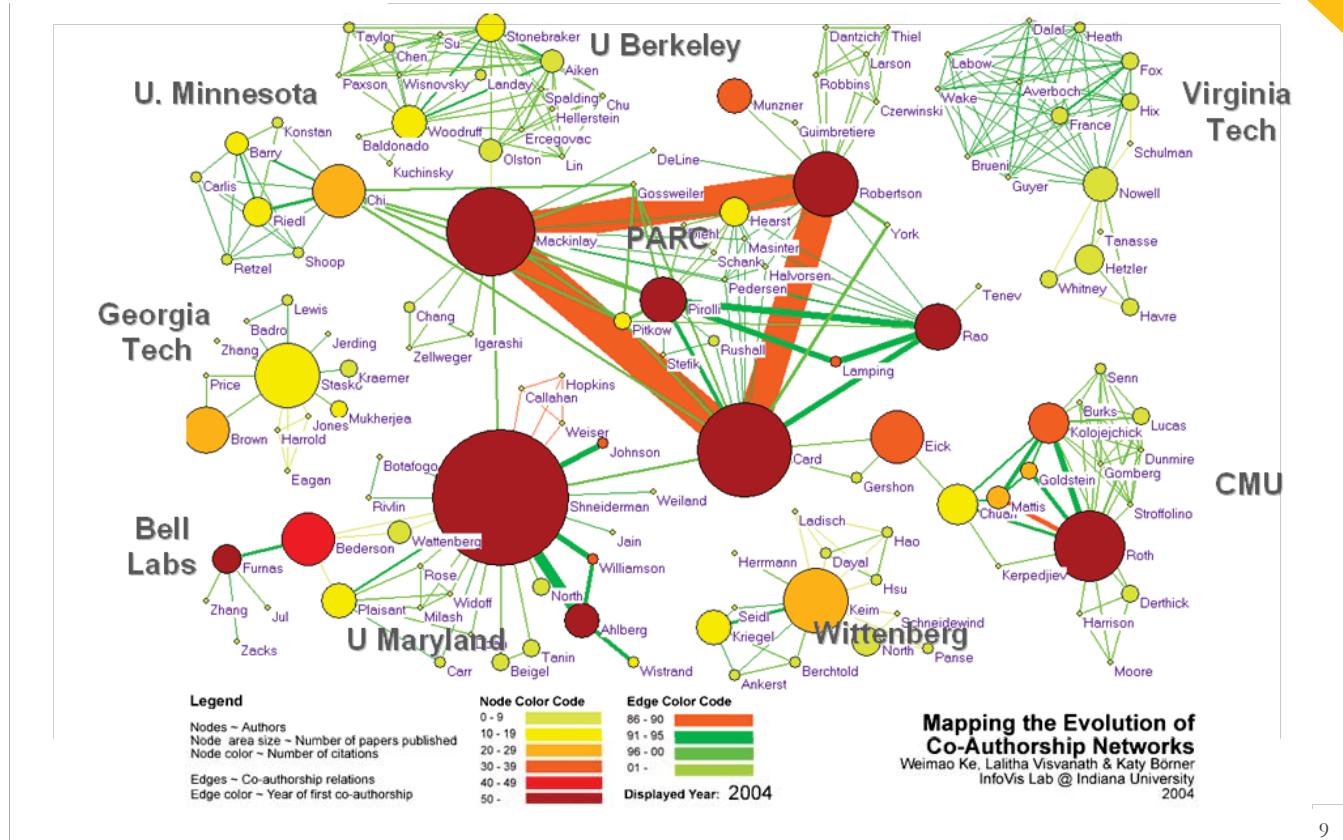


Mapping the Evolution of Co-Authorship Networks
Weimao Ke, Lalitha Visvanath & Katy Börner
InfoVis Lab @ Indiana University
2004

Mapping the Evolution of Co-Authorship Networks

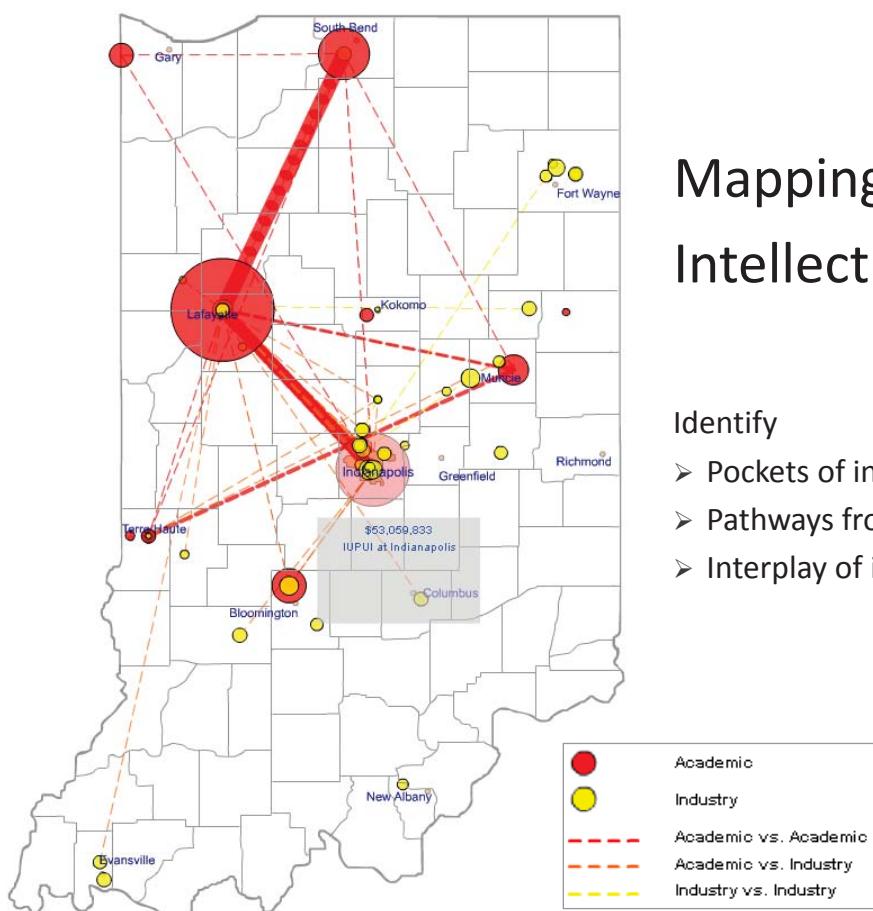
2

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



9

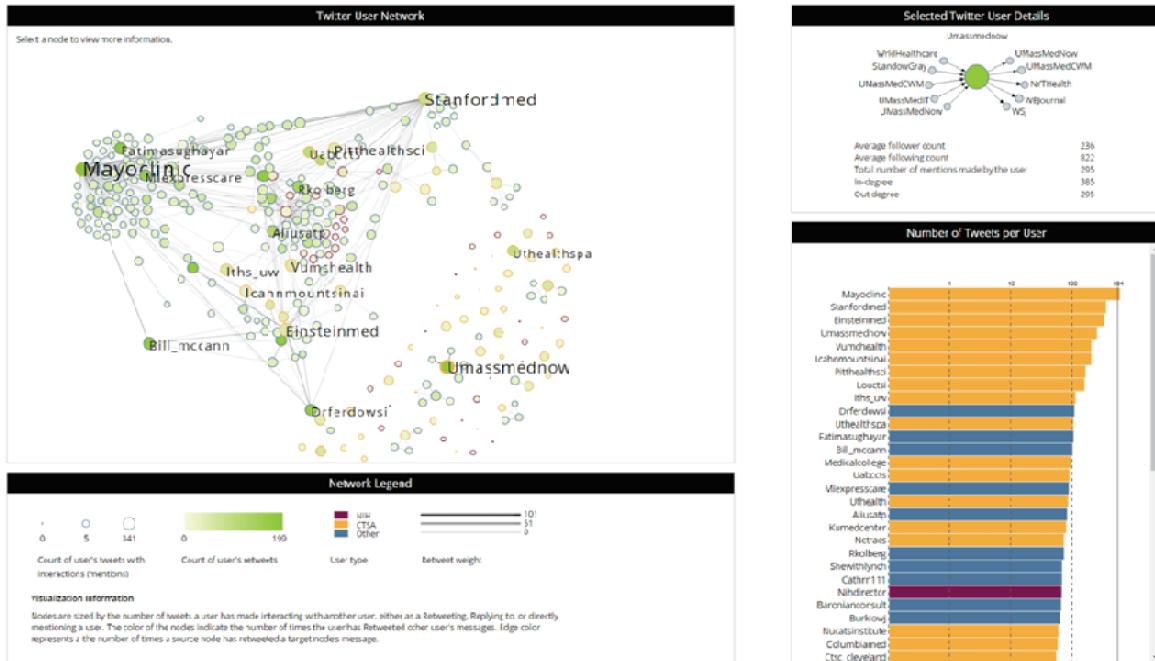
3



10

Twitter Network of NIH and STSI User Accounts

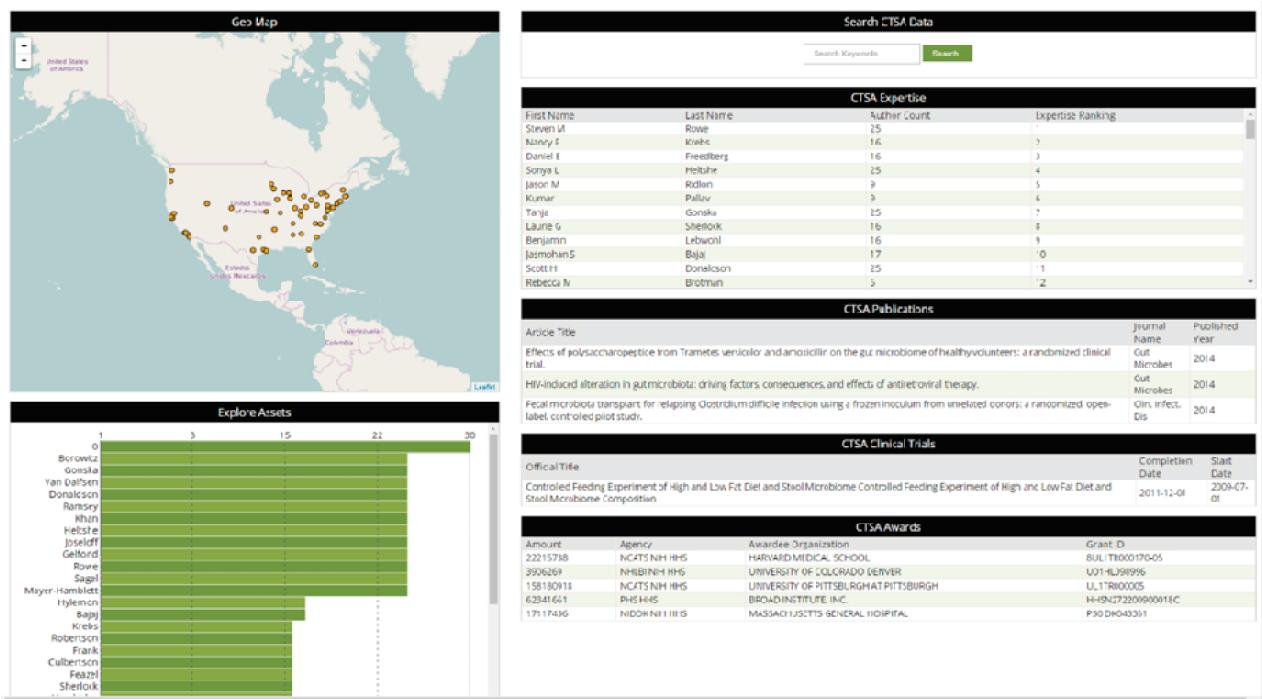
Tang, Börner, Liu, Light, & Simpson. 2016. SMS-VAT: A Scalable Multi-Scale Visual Analytical Tool



11

CTSI Expertise Visualization

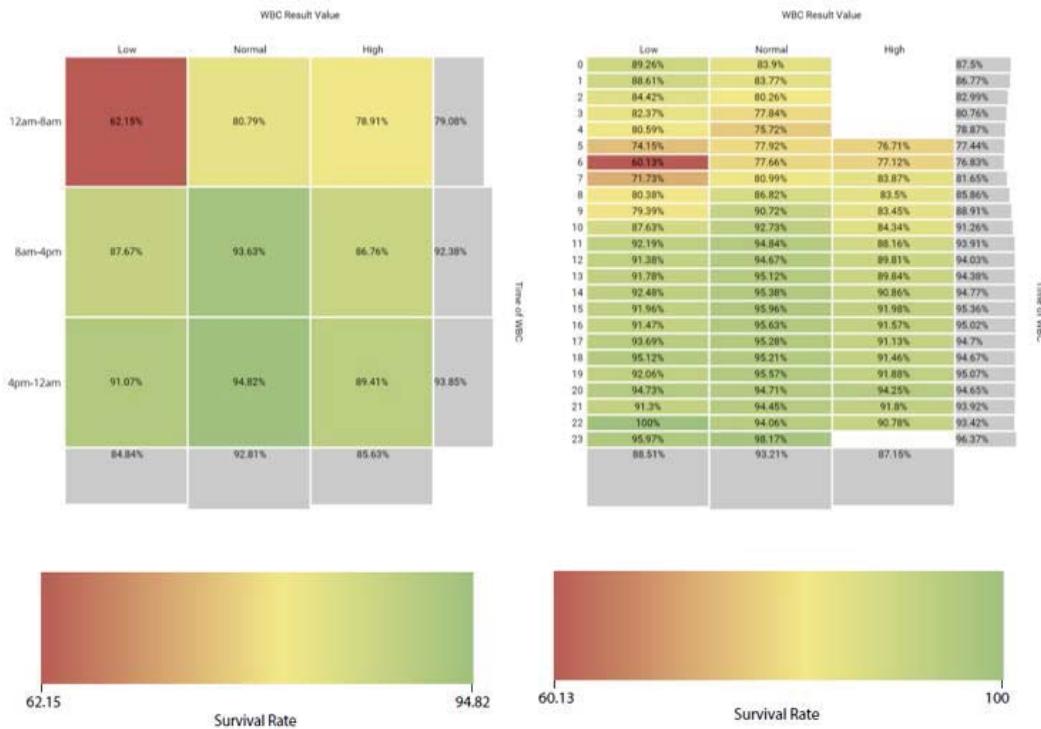
Tang, Börner, Liu, Light, & Simpson. 2016. SMS-VAT: A Scalable Multi-Scale Visual Analytical Tool



12

Mapping Big Biomedical Data

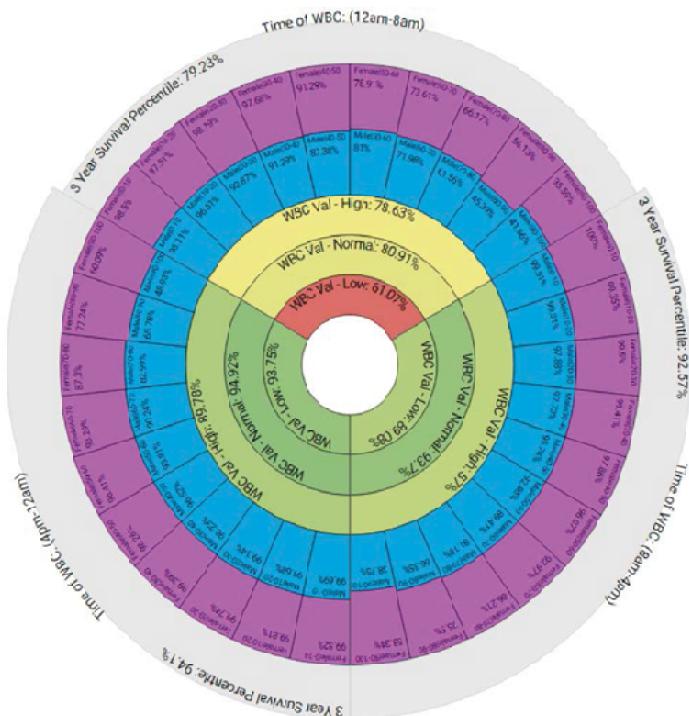
Weber and Börner. 2015. Visualizing Healthcare System Dynamics in Biomedical Big Data



13

Mapping Big Biomedical Data

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



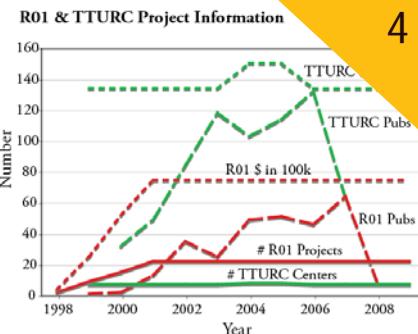
14

Mapping Transdisciplinary Tobacco Use Research Centers Publications

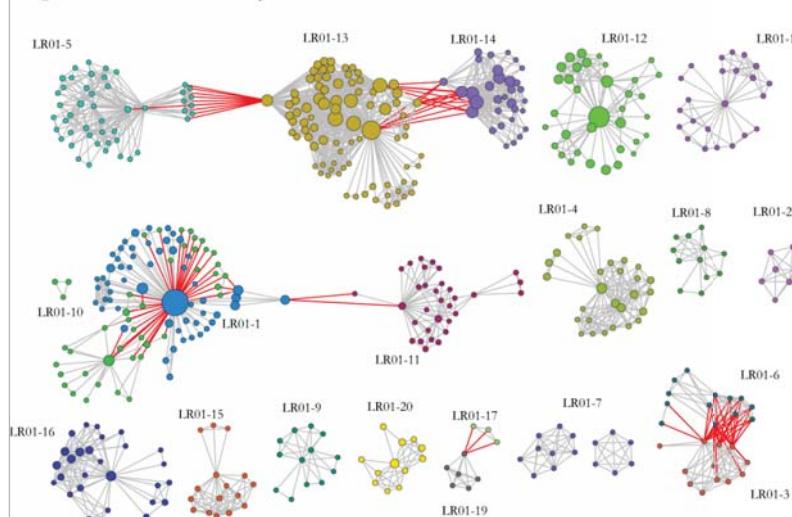
Compare R01 investigator-based funding with TTURC Center awards in terms of number of publications and evolving co-author networks.

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

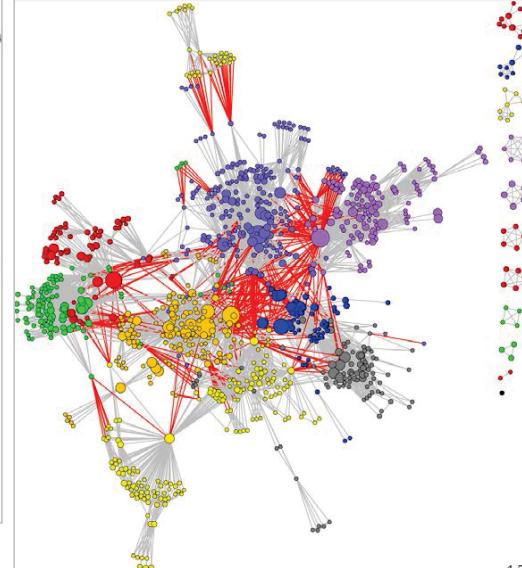
Supported by NIH/NCI Contract HHSN261200800812



Longitudinal R01 Co-Authorship Network



TTURC Co-Authorship Network

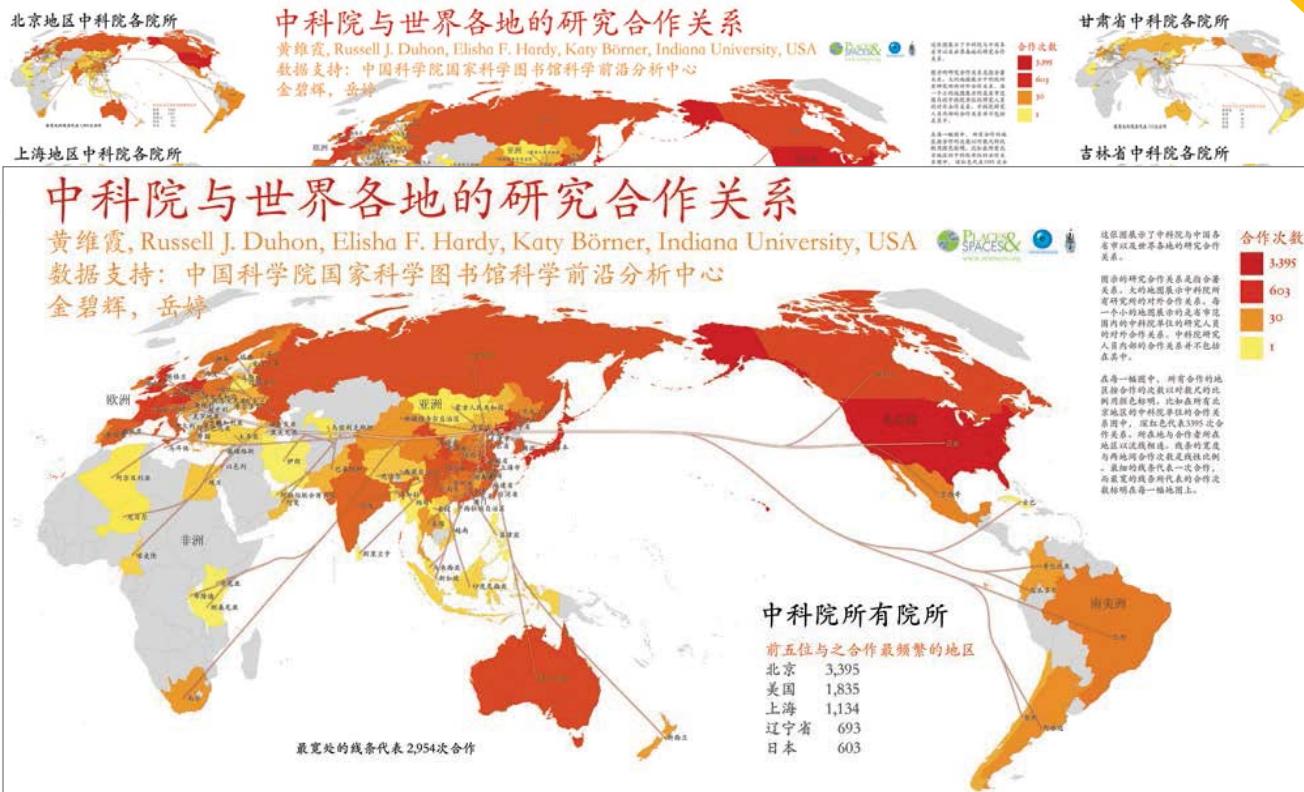


4

Research Collaborations by the Chinese Academy of Sciences

Huang, Duhon, Hardy & Börner

5



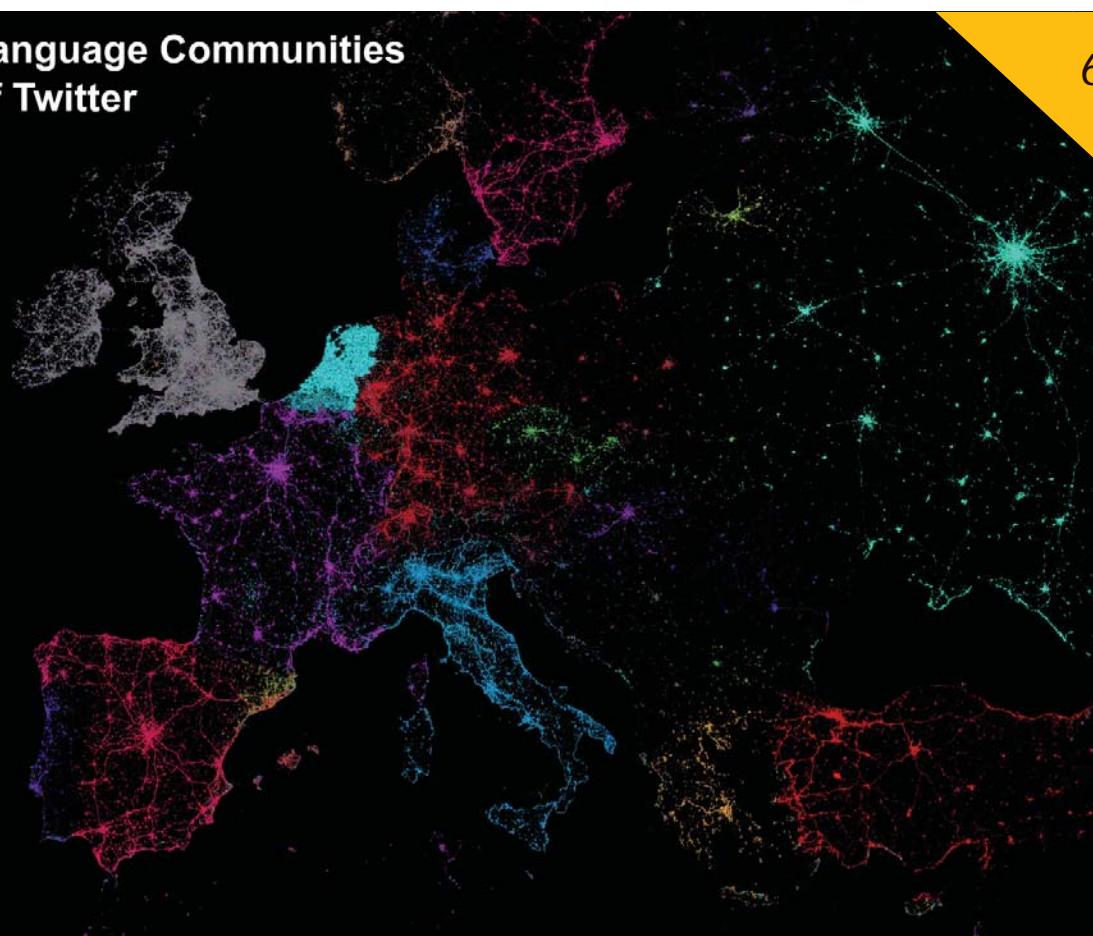
16



Language Communities of Twitter

6

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



Language Communities of Twitter - Eric Fischer - 2012

17

Chemical Research & Development Powers the U.S. Innovation Engine

Macroeconomic Implications of Public and Private R&D Investments in Chemical Sciences

The Council for Chemical Research (CCR)

7

INVESTMENT IN CHEMICAL SCIENCE R&D



TIMELINE FROM CONCEPTION
TO COMMERCIALIZATION

The design shows that an input of \$1B in federal investment, leveraged by \$5B industry investment, brings new technologies to market and results in \$10B of operating income for the chemical industry, \$40B growth in the Gross National Product (GNP) and further impacts the US economy by generating approximately 600,000 jobs, along with a return of \$8B in taxes. Additional details, also reported in the CCR studies, are depicted in the map to the left. This map clearly shows the two R&D investment cycles: the shorter industry investment at the innovation stage to commercialization cycle; and the longer federal investment cycle which begins in basic research and culminates in national economic and job growth along with the increase tax base that in turn is available for investment in basic research.

18

Empowering Anyone to Visualize STI

Example: The Information Visualization MOOC

19

Information Visualization MOOC 2015



Overview

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

The course can be taken for three Indiana University credits as part of the [Online Data Science Program](#), as part of the Information and Library Science M.S. program, and as part of the online Data Science M.S. Program offered by the School of Informatics and Computing. Students seeking enrollment information should contact Rhonda Spencer at 812-855-2018, ilsmain@indiana.edu or datasci@indiana.edu.

Among other topics, the course covers:

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



Register for Course

Already registered? [Click here](#) to go to the course.

Forgot your password? [Click here](#) to reset it.

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

Course Schedule

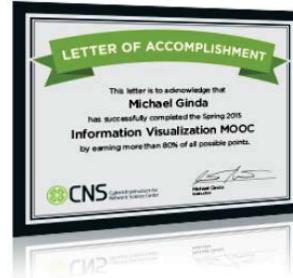
Part 1: Theory and Hands-On

- **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

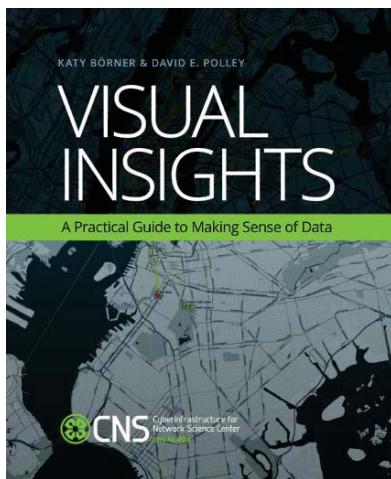


Part 2: Students work in teams on client projects.

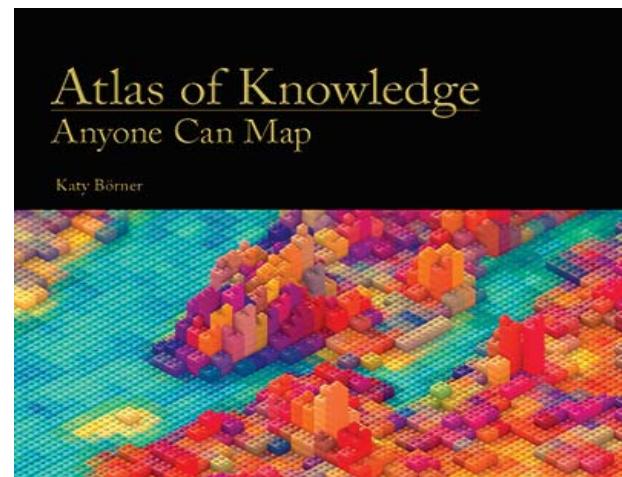
Final grade is based on Class Participation (10%), Midterm (30%), Final Exam (30%), and Client Project(30%).

21

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.

22

Visualization Frameworks



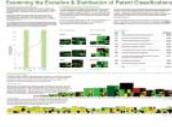
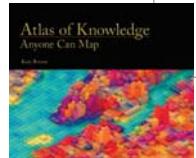
How to Classify Different Visualizations?

By

- User insight needs?
- User task types?
- Data to be visualized?
- Data transformation?
- Visualization technique?
- Visual mapping transformation?
- Interaction techniques?
- Or ?

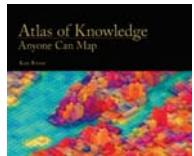


Tasks

LEVELS	
MICRO: Individual Level about 1–1,000 records page 6	MESO: Local Level about 1,001–100,000 records page 8
	
Statistical Analysis page 44 	 Knowledge Cartography page 135
WHEN: Temporal Analysis page 48 	 Visualizing decision-making processes page 95
WHERE: Geospatial Analysis page 52 	 Cell phone usage in Milan, Italy page 109
WHAT: Topical Analysis page 56 	 Evolving patent holdings of Apple Computer, Inc., and Jerome Lemelson page 89
WITH WHOM: Network Analysis page 60 	 World Finance Corporation network page 87
Atlas of Knowledge Anyone Can Map 	
See page 5	

25

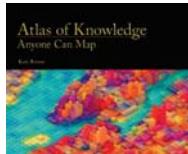
Insight Need Types page 26	Data Scale Types page 28	Visualization Types page 30	Graphic Symbol Types page 32	Graphic Variable Types page 34	Interaction Types page 26
<ul style="list-style-type: none"> categorize/cluster order/rank/sort distributions (also outliers, gaps) comparisons trends (process and time) geospatial compositions (also of text) correlations/relationships 	<ul style="list-style-type: none"> nominal ordinal interval ratio 	<ul style="list-style-type: none"> table chart graph map network layout 	<ul style="list-style-type: none"> geometric symbols point line area surface volume linguistic symbols text numerals punctuation marks pictorial symbols images icons statistical glyphs 	<ul style="list-style-type: none"> spatial position retinal form color optics motion 	<ul style="list-style-type: none"> overview zoom search and locate filter details-on-demand history extract link and brush projection distortion



See page 24

26

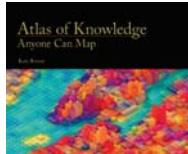
Basic Task Types		Bertin, 1967	Wehrend & Lewis, 1996	Few, 2004	Yau, 2011	Rendgen & Wiedemann, 2012	Frankel, 2012	Tool: Many Eyes	Tool: Chart Chooser	Börner, 2014
selection	categorize				category					categorize/cluster
order	rank	ranking						table		order/rank/sort
	distribution	distribution							distribution	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

27

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
<ul style="list-style-type: none"> categorize/cluster order/rank/sort distributions (also outliers, gaps) comparisons trends (process and time) geospatial compositions (also of text) correlations/relationships 	<ul style="list-style-type: none"> nominal ordinal interval ratio 	<ul style="list-style-type: none"> table chart graph map network layout 	<ul style="list-style-type: none"> geometric symbols point line area surface volume linguistic symbols text numerals punctuation marks pictorial symbols images icons statistical glyphs 	<ul style="list-style-type: none"> spatial position retinal form color optics motion 	<ul style="list-style-type: none"> overview zoom search and locate filter details-on-demand history extract link and brush projection distortion



See page 24

28

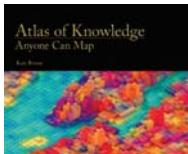
Visualization Types (Reference Systems)

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.

29

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
<ul style="list-style-type: none">• categorize/cluster• order/rank/sort• distributions (also outliers, gaps)• comparisons• trends (process and time)• geospatial• compositions (also of text)• correlations/relationships	<ul style="list-style-type: none">• nominal• ordinal• interval• ratio	<ul style="list-style-type: none">• table• chart• graph• map• network layout	<ul style="list-style-type: none">• geometric symbols• point• line• area• surface• volume• linguistic symbols• text• numerals• punctuation marks• pictorial symbols• images• icons• statistical glyphs	<ul style="list-style-type: none">• spatial• position• retinal• form• color• optics• motion	<ul style="list-style-type: none">• overview• zoom• search and locate• filter• details-on-demand• history• extract• link and brush• projection• distortion



See page 25

30

Graphic Variable Types Versus Graphic Symbol Types

		Geometric Symbols		
		Point	Line	Area
Spatial	x	quantitative		
	y	quantitative		
	z	quantitative		
Retinal	Size	quantitative	NA (Not Applicable)	
	Shape	qualitative	NA	
	Rotation	quantitative	NA	
	Curvature	quantitative	NA	
	Angle	quantitative	NA	
	Closure	quantitative	NA	
	Value	quantitative		
	Hue	qualitative		
Color	Saturation	quantitative		

Atlas of Knowledge
Anyone Can Map

See page 36

31

Graphic Variable Types Versus Graphic Symbol Types

		Geometric Symbols			Qualitative Symbols			Pictorial Symbols		
		Point	Line	Area	Surface	Volume	Text	Text	Text	Image, Icons, Statistical Graphs
Spatial	x	quantitative								
	y	quantitative								
	z	quantitative								
Retinal	Size	quantitative	NA (Not Applicable)							
	Shape	qualitative	NA							
	Orientation	quantitative	NA							
	Curvature	quantitative	NA							
	Angle	quantitative	NA							
	Closure	quantitative	NA							
	Value	quantitative								
	Hue	qualitative								
Color	Saturation	quantitative								

Atlas of Knowledge
Anyone Can Map

See pages 36-39

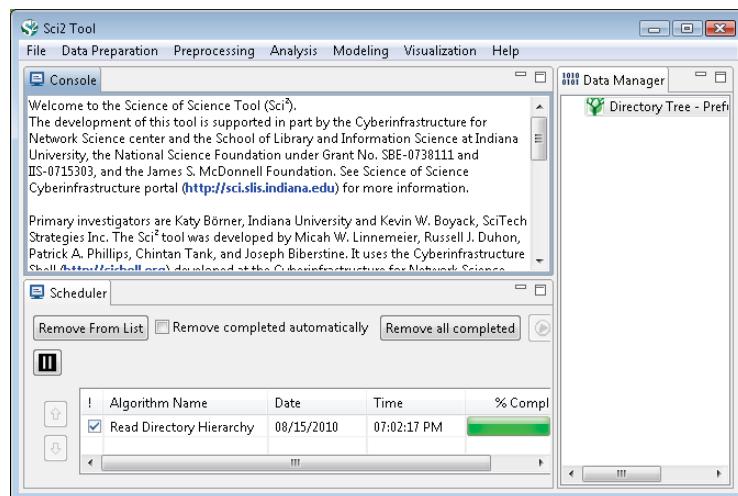
32

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.



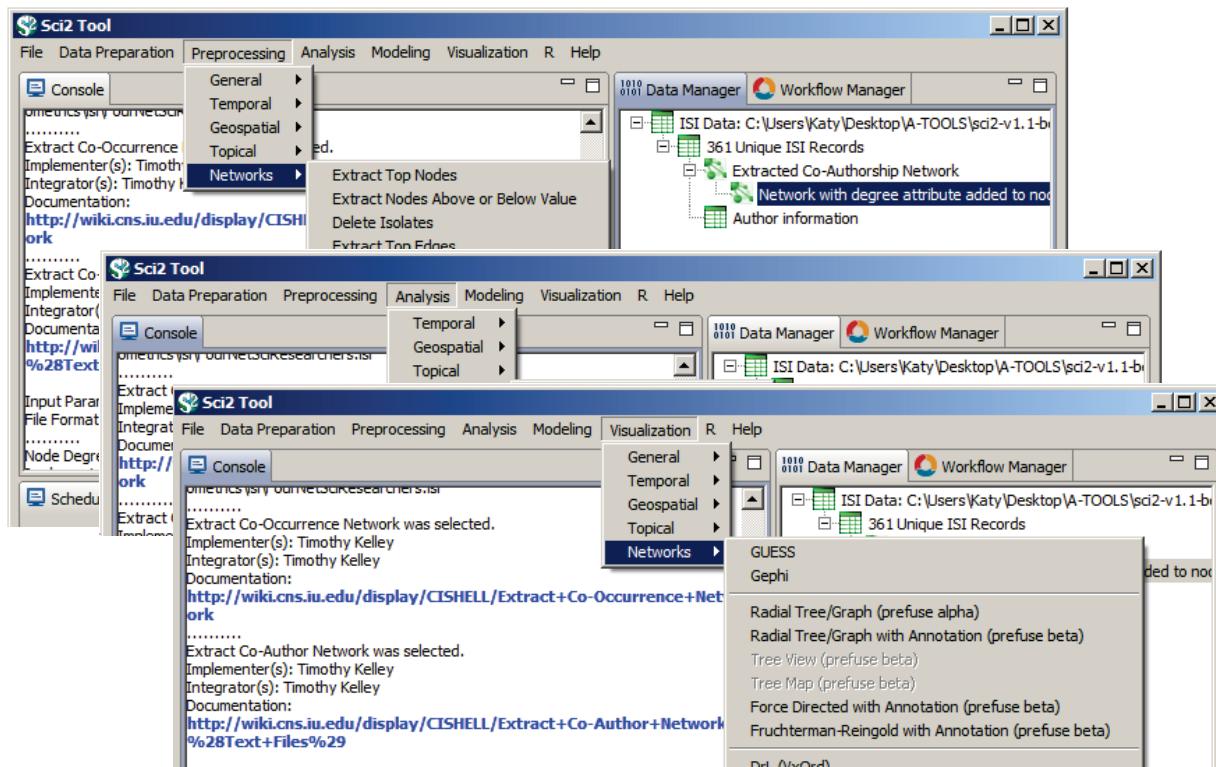
All workflows are recorded into a log file (see `/sci2/logs/...`), and can be re-run for easy replication. If errors occur, they are saved in a error log to ease bug reporting.

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

33

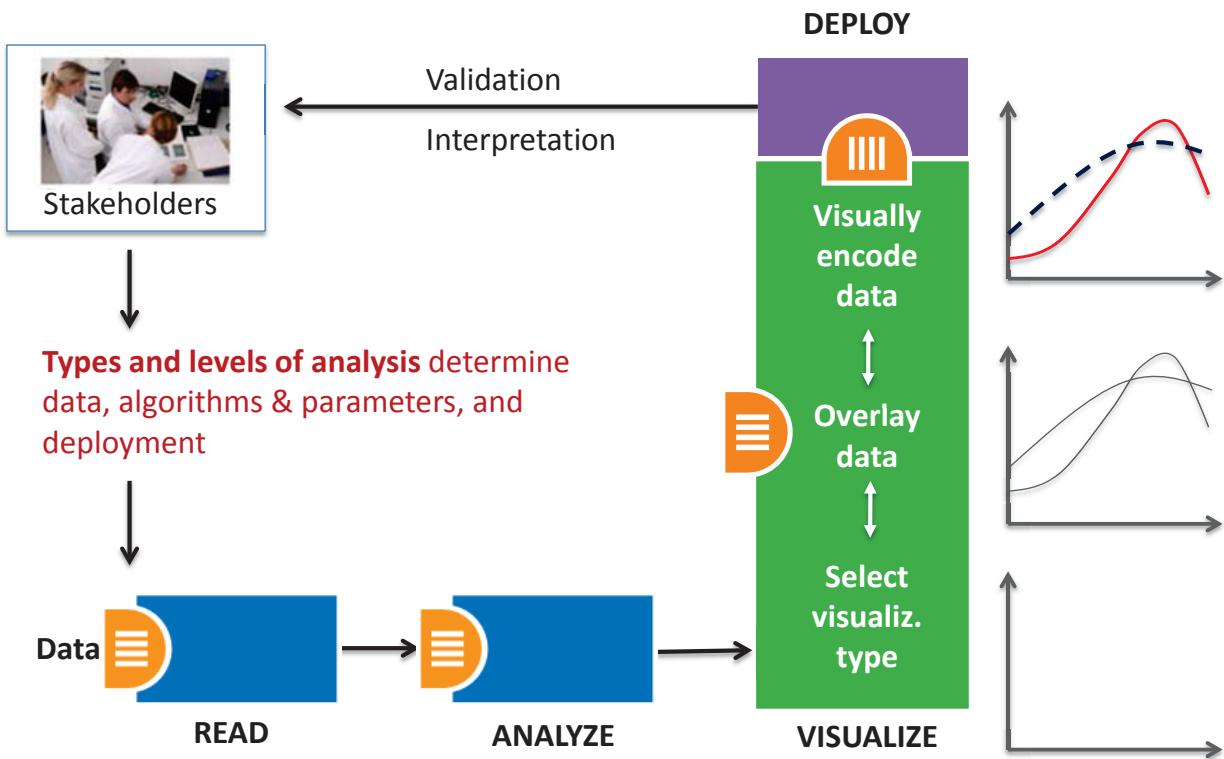
Sci2 Tool Interface Components

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



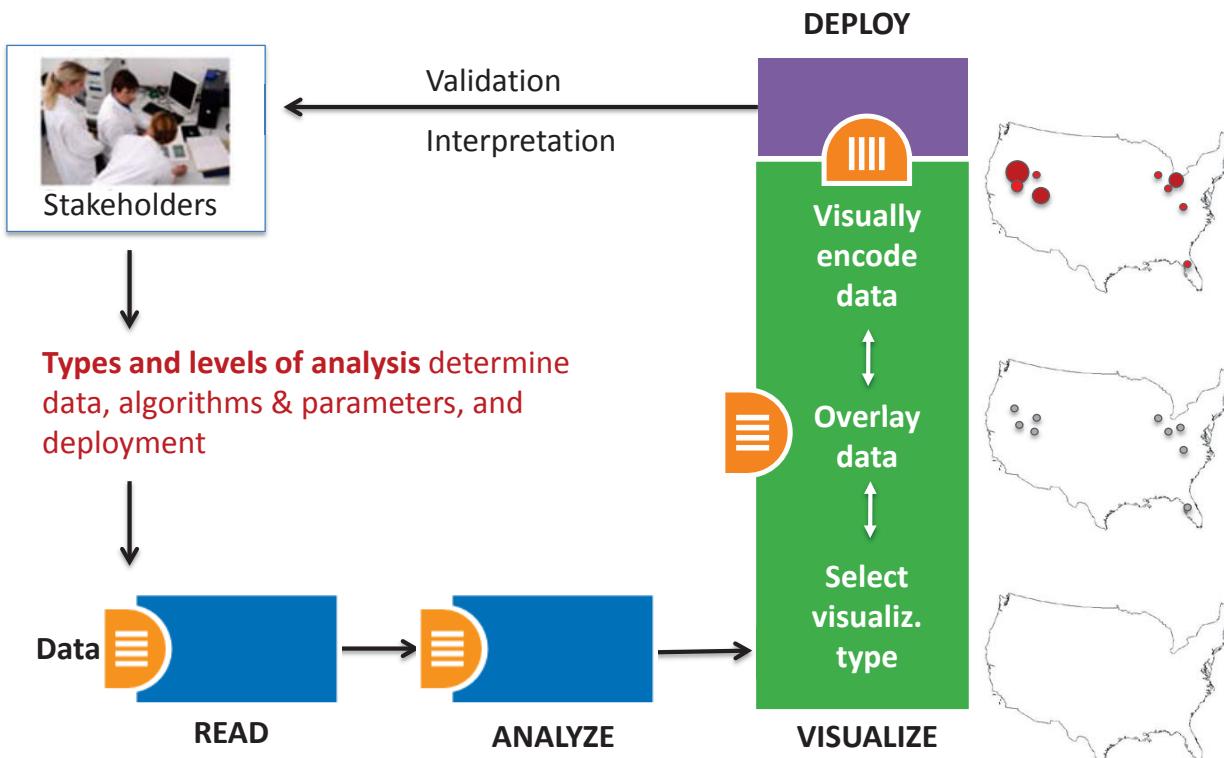
34

Needs-Driven Workflow Design



35

Needs-Driven Workflow Design



36

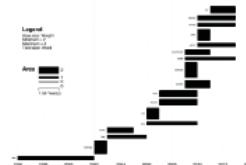
Load One File and Run Many Analyses and Visualizations

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 Macrosopes	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

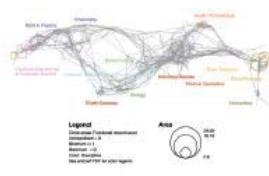


37

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 Macrosopes	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

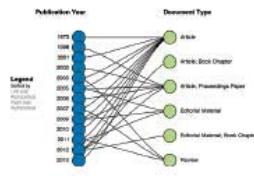
Topical Analysis—p. 56



Paper Citation Network—p. 60



Bi-Modal Network—p. 60



Co-author and many other bi-modal networks.

38

Existing Interfaces for Health-related Data

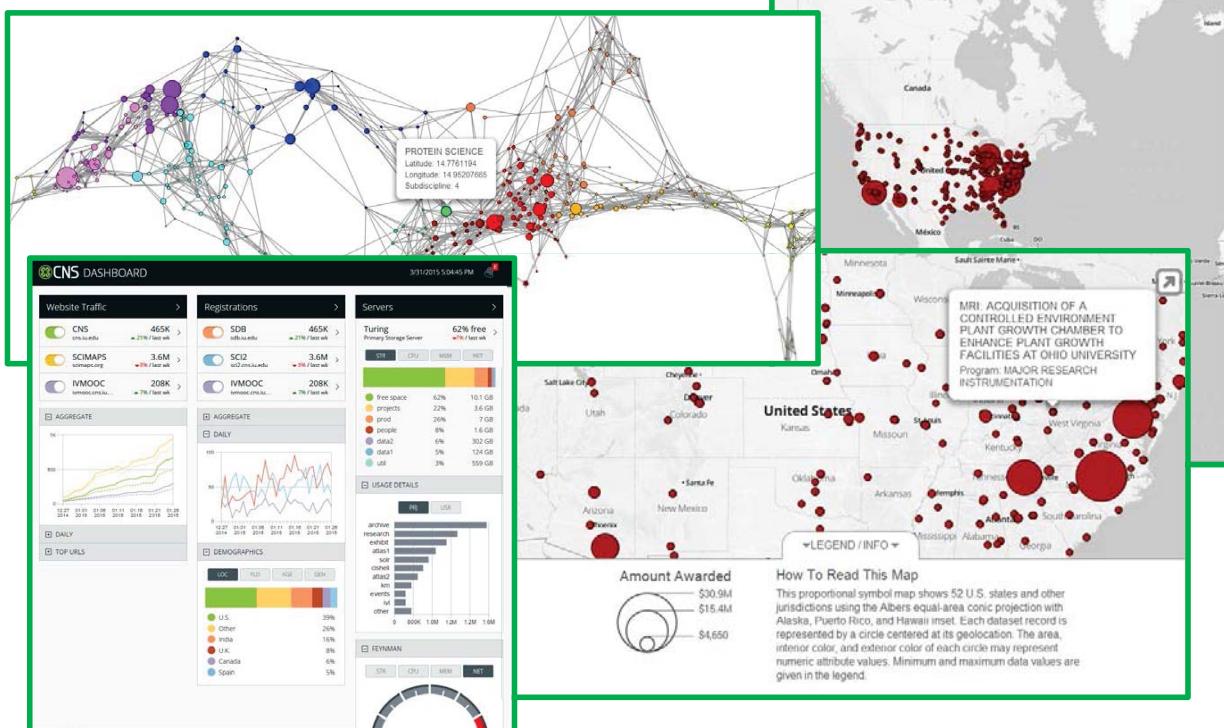
- 9) NIH RePORTER: Visual Interface to Biomedical Funding Data in U.S.
- 10) CIShell/Sci2 World and Science Visualizations of NIH RePORTER Data
- 11) NIH RePORTER: NIH Map
- 12) BBSRC: Visual Interface to Biomedical Funding Data in UK
- 13) IAI Multidimensional Analysis
- 14) Scraawl: Twitter Analysis
- 15) Illuminated Diagram: Searchable World and Science Maps

NIH RePORTER: Visual Interface to Biomedical Funding Data in US

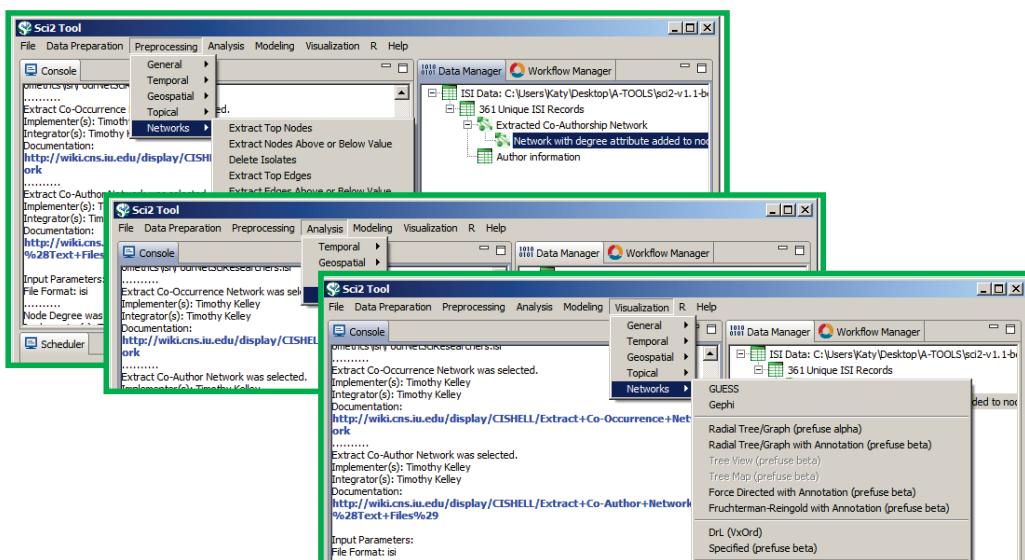
9



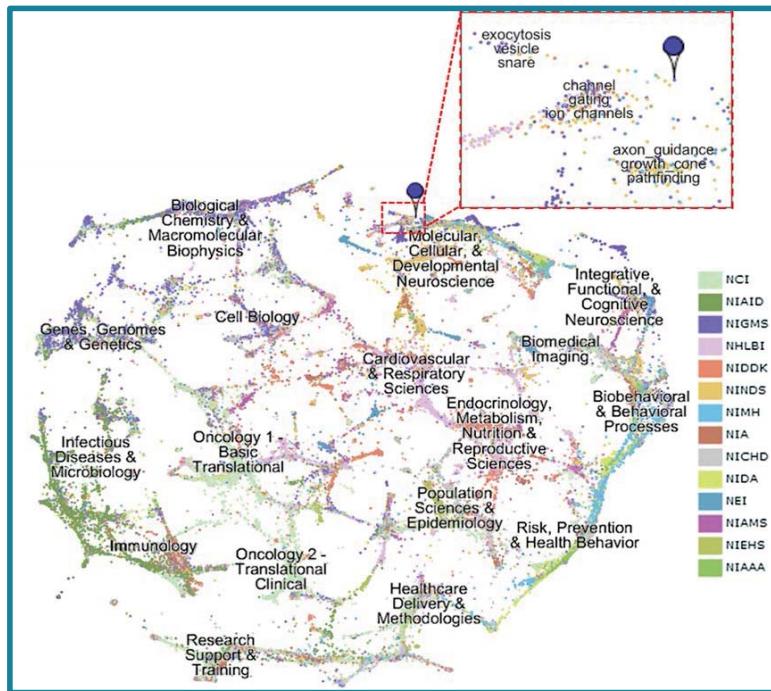
CISShell/Sci2 World and Science Visualizations of NIH RePORTER Data



Sci2 Desktop

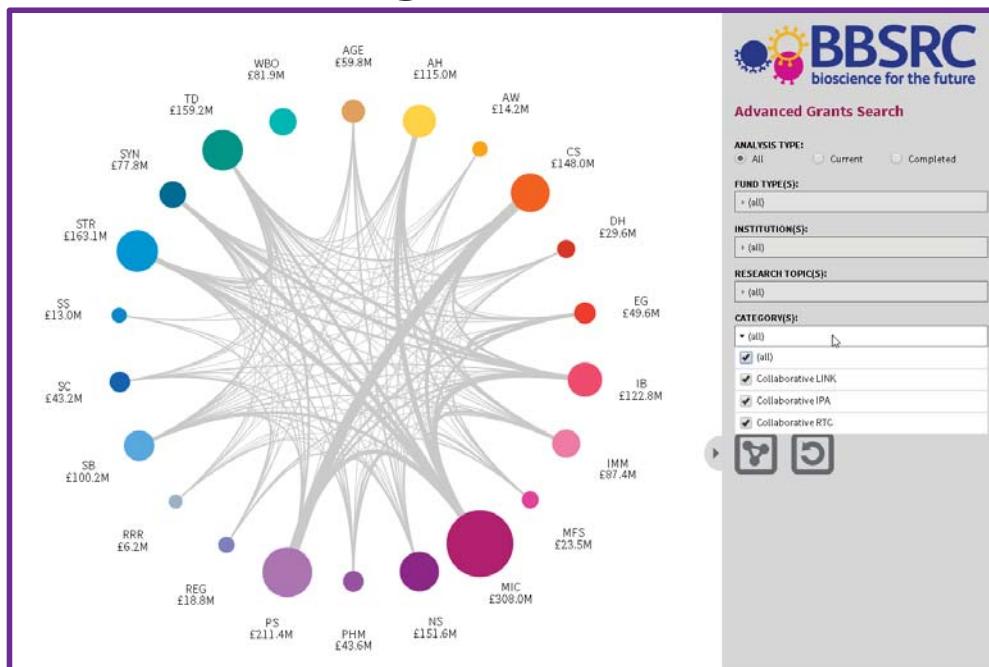


NIH RePORTER: NIH Map



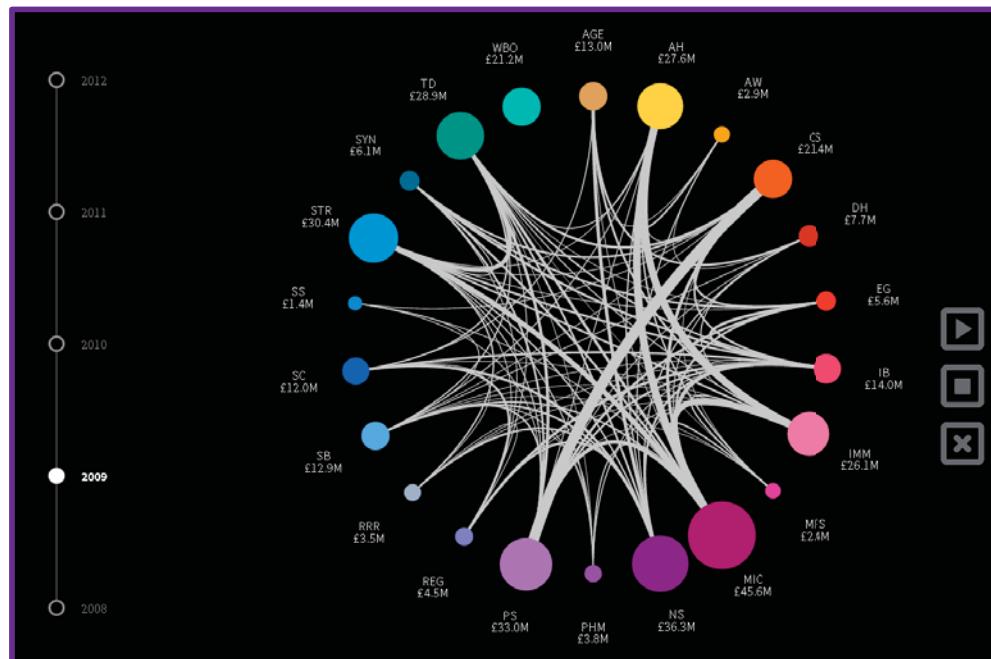
<http://nihmaps.org/>

BBSRC: Visual Interface to Biomedical Funding Data in UK



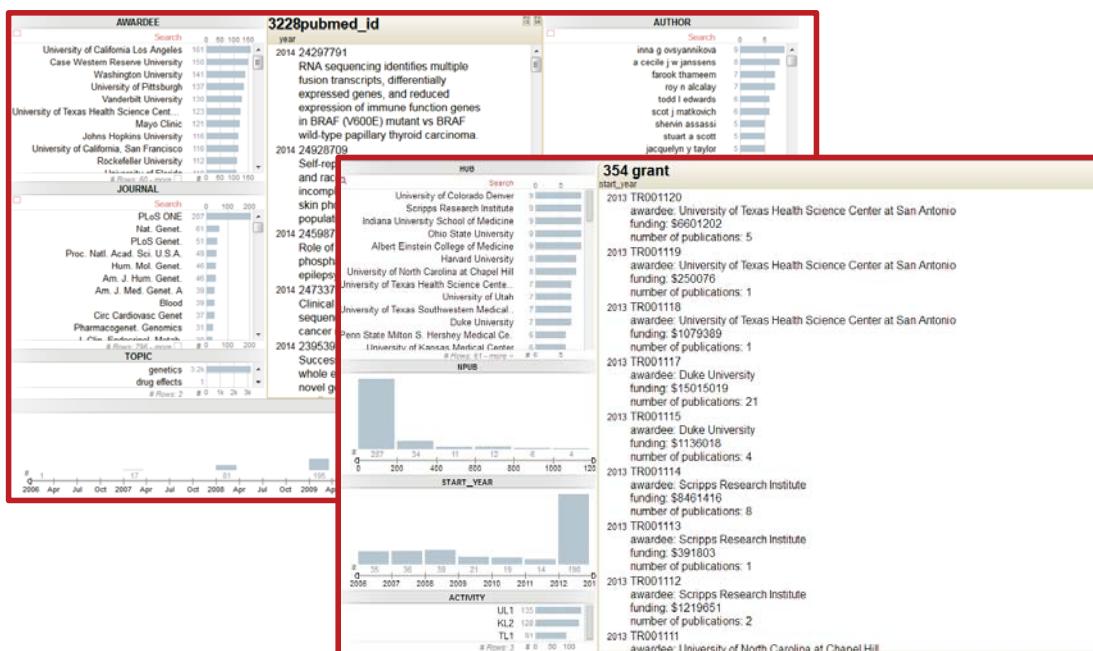
<http://www.bbsrc.ac.uk/>

BBSRC: Temporal animation



IAI Multidimensional Analysis

13



Scraawl: Twitter Analysis

[Edit Report](#)

Report Name: NIH-NCATS Test · Search Term: @NIH, #ncats

[My Reports](#) / [Edit Report](#)

Search terms
@NIH x #ncats x

Separate your search terms with a comma to search for multiple terms (e.g. twitter, social media).

Rollover report
When the report is completed, automatically create a new report and continue collecting tweets.

[Update Report](#)

Name (optional)
NIH-NCATS Test

Optional name for the report

Data sources
 Twitter (Coming soon!) Instagram Tumblr LinkedIn Flickr

Your Scraawl searches will time out in 1 hours, or when the search returns the number of tweets specified by the search privileges.

90 total users | 197 total words | 23 total hashtags | 29 total mentions | 26 total URLs | 26 retweeted tweets | 2 total languages

Top Users

User	Count
@francesimiley	4
@dougalmanphd	2
@lisionesp_s21	2
@qipresearch	2
@medsci_news	2

Top Words

Word	Count
Cancer	43
Research	37
Medical	25
Mary	25
Lasker	25

Top Hashtags

Hashtag	Count
#cancerfilm	74
@nih	27
#cancer	27
#clinicaltrials	6
#precisionmedicine	4

Top Mentions

Mention	Count
@nih	100
@conandremes	6
@usccancer	5
@istem_malaria	3
@waltneedarmy	3

Top URLs

URL	Count
http://profiles.nih.gov/p.../	24
http://dp.nci.nih.gov/melin.../	9
http://www.nih.m.nih.gov/med.../	9
http://www.cancer.gov/staff/...	7
http://www.nih.gov/health/c...	5

Top Retweets

Retweet	Count
RT @NIH_MaryLasker: som...	20
RT @NIH: Largely thanks to r...	6
RT @NIH_President_Naom S...	6
RT @NIH: Cancer is a gene...	6
RT @NIH: #Cancer occurs w...	5

Top Languages

99% English | 1% French

Top Locations

Location	Count
España	1

0 geocoded tweets | View map

Tweet Timeline

A product of Intelligent Agents. Beta release version 0.9.4. Please report any issues.

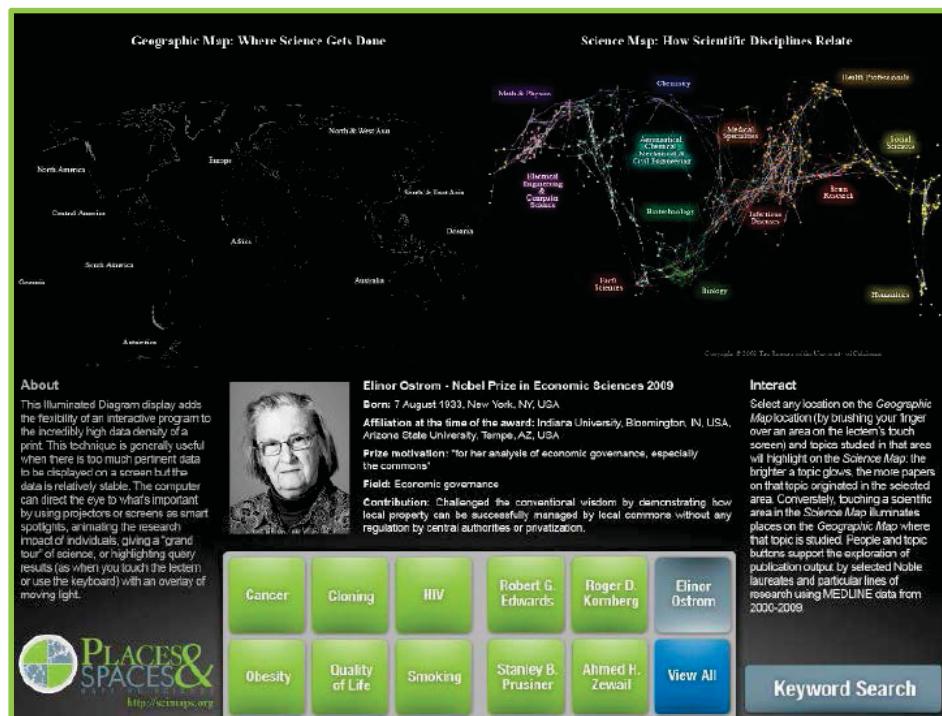
<https://www.scraawl.com/>

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/

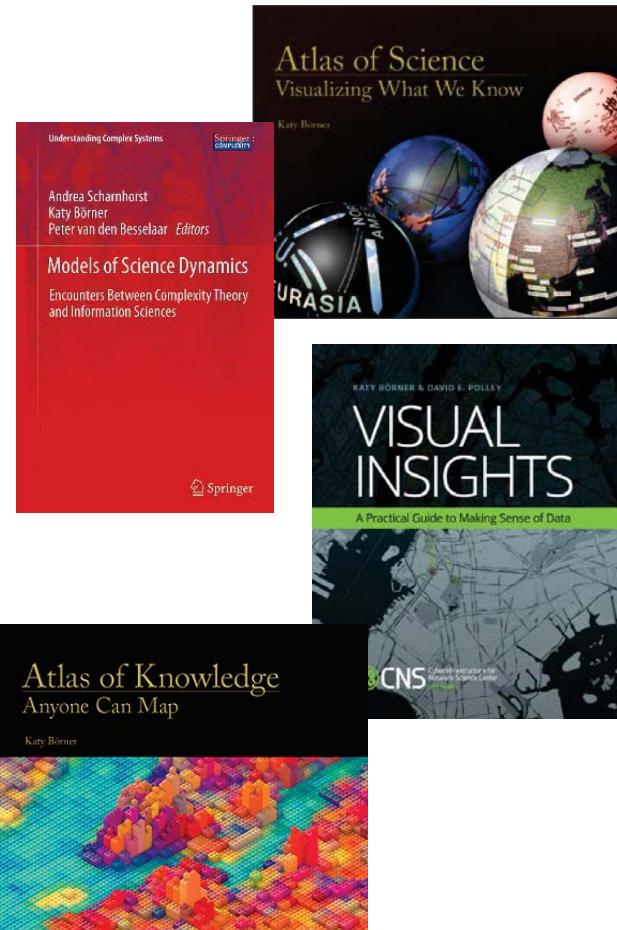
Börner, Katy (2010) **Atlas of Science: Visualizing What We Know**. The MIT Press. <http://scimaps.org/atlas>

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>



The screenshot shows the CNS website homepage. At the top, there's a navigation bar with links for About Us, Research, Development, Teaching, Outreach, Videos, News & Events, and Connect With Us. A search bar and social media icons (Facebook, Twitter, LinkedIn) are also at the top. Below the navigation is a large image of four people working together around a table with laptops. To the right of this image is a green sidebar with the text: "We work closely with clients to provide custom-made data, visualization, and software solutions". The main content area is divided into several sections: "Research" (with a link to "Open Data and Open Code for Big Science of Science Studies"), "Development" (with a link to "Behind the scenes of the design and development of AcademyScope"), "Videos" (with a link to "Watch Katy Börner's full presentation from TEDxBloomington"), "Teaching" (with a link to "Successful IVMOOC will be offered again in January of 2014"), "Latest News" (with a link to "Put your money where your citations are: a proposal for a new funding system"), "Outreach" (with a link to "See some of the most fascinating data visualizations in the world"), "Upcoming Events" (listing events like "Katy Börner attends PIUG 2013 Northeast Conference" on Oct 1), and "Our Products" (with a link to "We work closely with clients to provide custom-made data, visualization, and software solutions").

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>