Scalable Multi-Scale Visual Analytical Tools for NCATS

Katy Börner

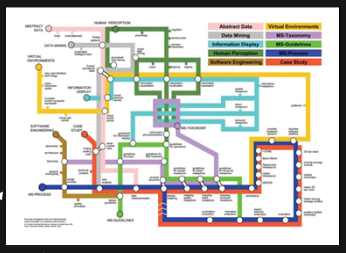
Victor H. Yngve Professor of Information Science
Director, Cyberinfrastructure for Network Science Center
School of Informatics and Computing, Indiana University, USA

Washington DC
April 2, 2015

Language Communities of Twitter - Eric Fischer - 2012

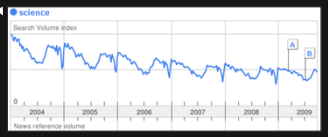
Terra bytes of data

Descriptive & Predictive Models



Find your way

Find collaborators, friends



Identify trends

Scalable Multi-Scale Visual Analytical Tools for NCATS—Questionnaire Results

Number of responses:	5 NIH, 4 NCATS, possible overlap
Gender:	5 female, 4 male
Median age:	38 (average Twitter user is 37.3 years old)
Academic background:	Microbiology, bioinformatics, informatics, computer science, epidemiology, pharma, physics, physician
Native Language:	9 English

3

Scalable Multi-Scale Visual Analytical Tools for NCATS—Questionnaire Results

NIH: In your daily work, what datasets do you use?

Medline Publications, NIH Awards, USTO Patents, Clinical Trials, Program Management, bioactivity data, LinkedIn, ResearchGate

NCATS: What systems/databases do you use?

Medline Publications, eRA (IMPACT, QVR, xTrain), Clinical Trials, Excel, APR, LinkedIn, Research Gate

4

Scalable Multi-Scale Visual Analytical Tools for NCATS—Questionnaire Results

NIH or NCATS	WHEN	WHAT AREAS	WITH WHOM	WHERE
NIH	1	1		
NIH	1	1	1	
NIH		1	1	1
NIH	1	1	1	1
NIH	1	1	1	1
NCATS	1	1	1	1
NCATS	1	1	1	1
NCATS	1	1	1	1
NCATS	1	1	1	
TOTALS	**8**	**9**	**8**	**6**

In your daily work, what general questions do you need to answer?

- ☐ Temporal questions, e.g., WHEN something happens or temporal distributions
- ☐ Geospatial questions, e.g., WHERE something happens or geospatial distributions
- ☐ Topical questions, e.g., WHAT areas of science are emerging, active, or affected
- ☐ Network questions, e.g., WITH WHOM people collaborate or how experts migrate

5

Scalable Multi-Scale Visual Analytical Tools for NCATS—Questionnaire Results

	Bursts of			Diffusion of		
NCATS/NI	Tweets/Topics	$	Experts	Translation	Ideas	Experts
NIH	1	1				
NIH						
NIH	1	1	1			
NIH	1	1	1			
NIH	1	1	1			
NCATS						
NCATS	1	1	1	1	1	1
NCATS	1	1	1	1	1	1
NCATS				1	1	1
TOTALS	**6**	**6**	**5**	**3**	**3**	**3**

Tools now exist that detect bursts of activity, i.e., sudden increases in the usage frequency of words, such as names, institutions, keywords, citation references. How might you use such a tool to improve your daily decision making?

- ☐ Identify surges of interest in publication or social media (e.g., Twitter data) on certain topics (e.g., specific diseases)
- ☐ Identify sudden changes in the number of dollars spent on a topic (across agencies)
- ☐ Identify steep increases in the number of experts that work on a topic

Another set of tools can be used to depict branching and diffusion processes using network visualizations. How might you use such a tool to improve your daily decision making?

- ☐ Analyze and depict the translation of research results into bedside advances of health
- ☐ Visualize the diffusion of ideas between different areas of science
- ☐ Study the circulation of experts between institutions, geolocations, or scientific areas

6

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

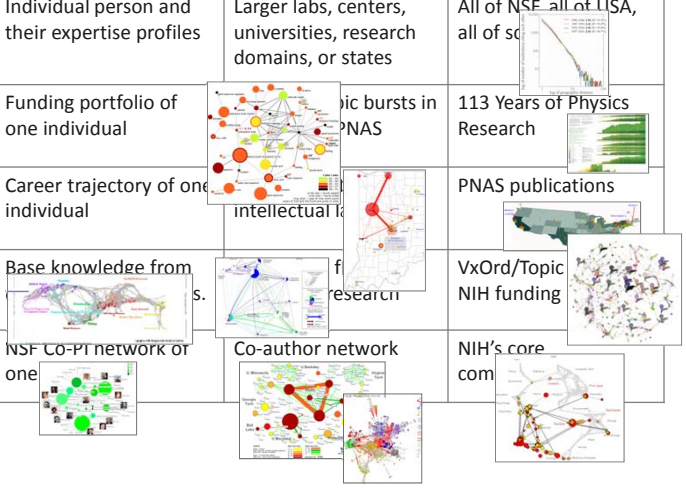
Type of Analysis vs. Level of Analysis

	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

9

Type of Analysis vs. Level of Analysis

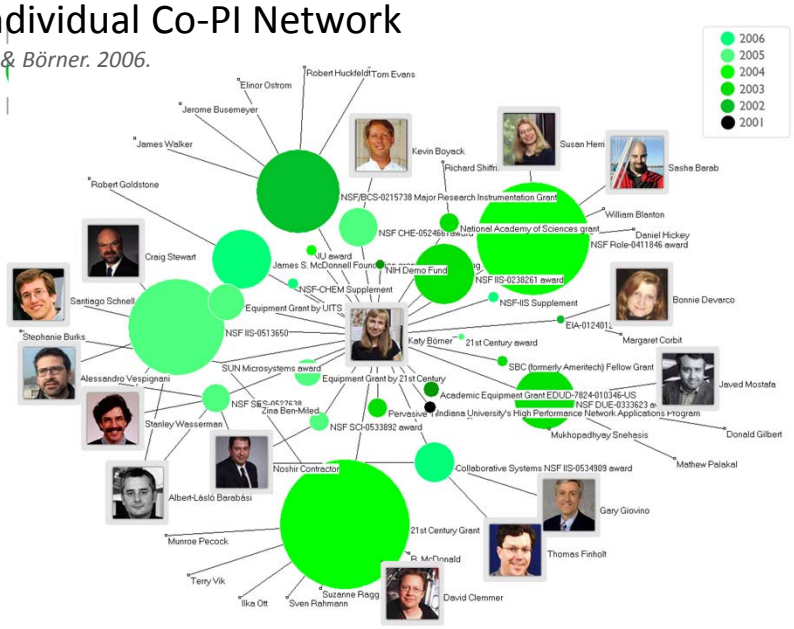
	Micro/Individual (1-100 records)	***Meso/Local (101–100,000 records)***	***Macro/Global (100,000 < records)***
Statistical Analysis/Profiling	Individual person and their expertise profiles	Larger labs, centers, universities, research domains, or states	All of NSF, all of USA, all of sc
Temporal Analysis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



10

Individual Co-PI Network

Ke & Börner. 2006.

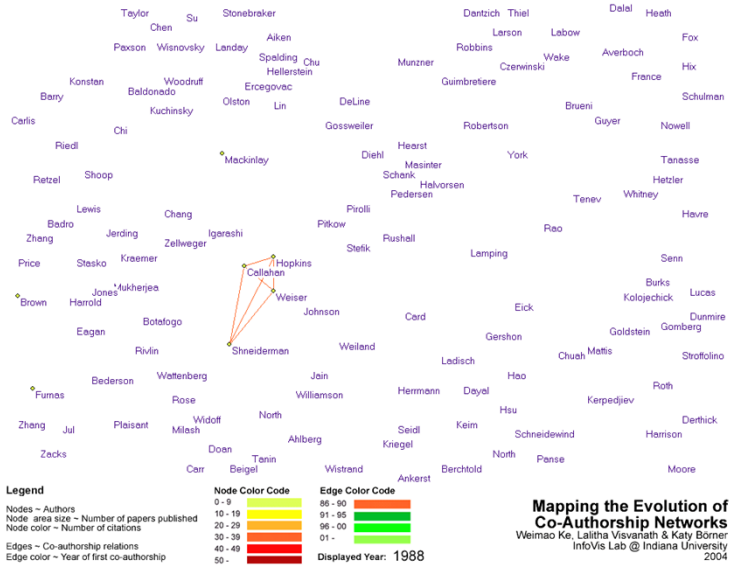


1

11

Mapping the Evolution of Co-Authorship Networks

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

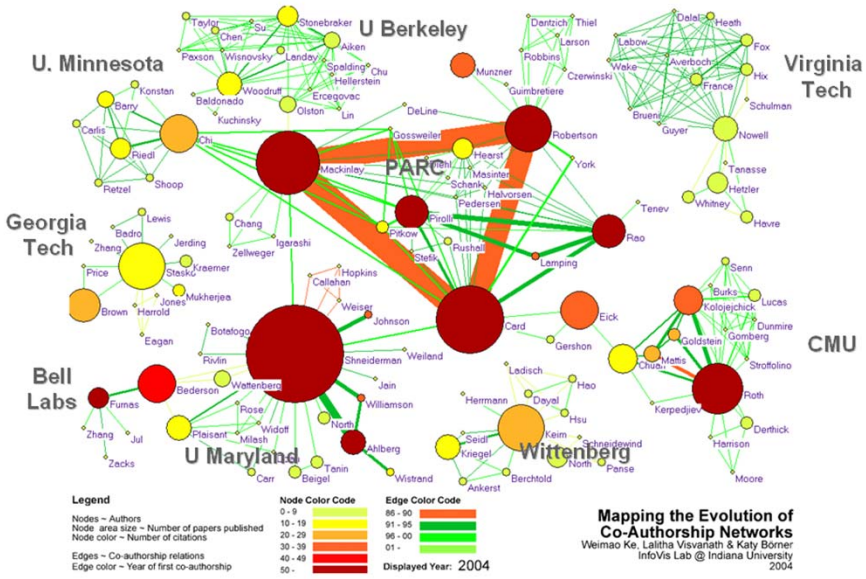


12

Mapping the Evolution of Co-Authorship Networ

2

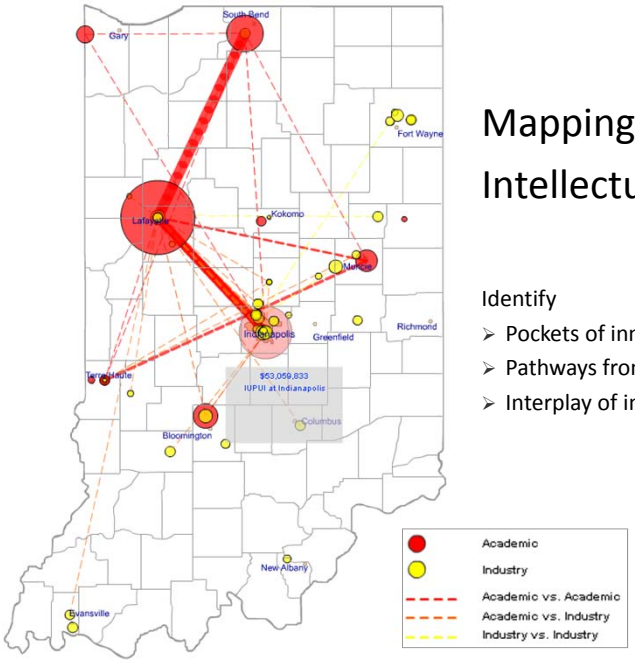
Ke, Visvanath & Börner. 2004. Won 1st prize at the IEEE InfoVis Contest.



13

Mapping Indiana's Intellectual Space

3



Identify

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

14

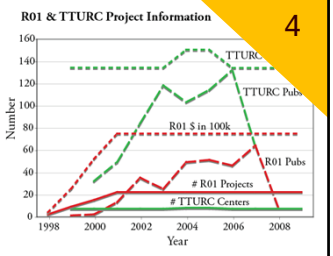
Mapping Transdisciplinary Tobacco Use Research Centers Publications

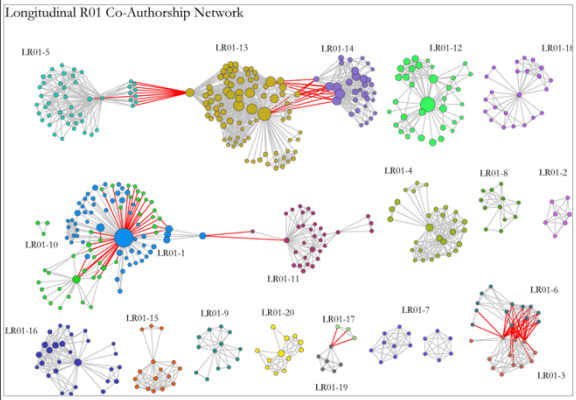
4

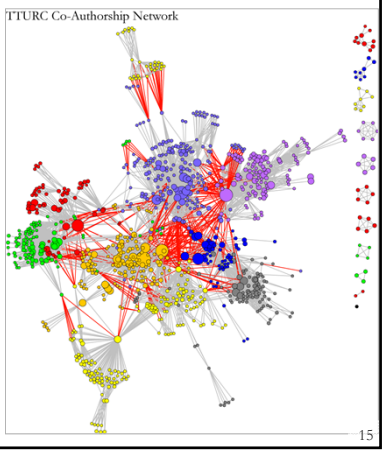
Compare R01 investigator-based funding with TTURC Center awards in terms of number of publications and evolving co-author networks.

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

Supported by NIH/NCI Contract HHSN261200800812





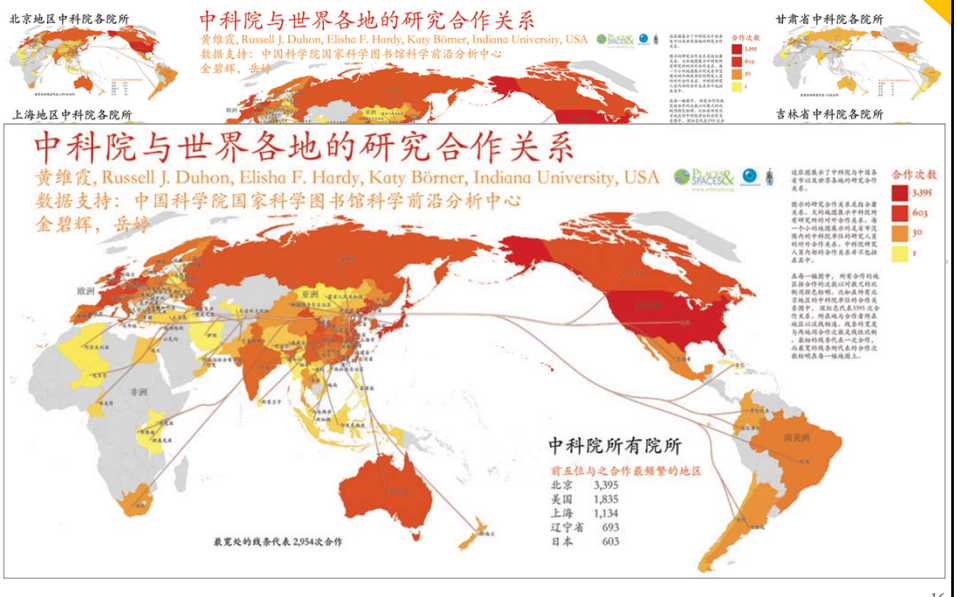


15

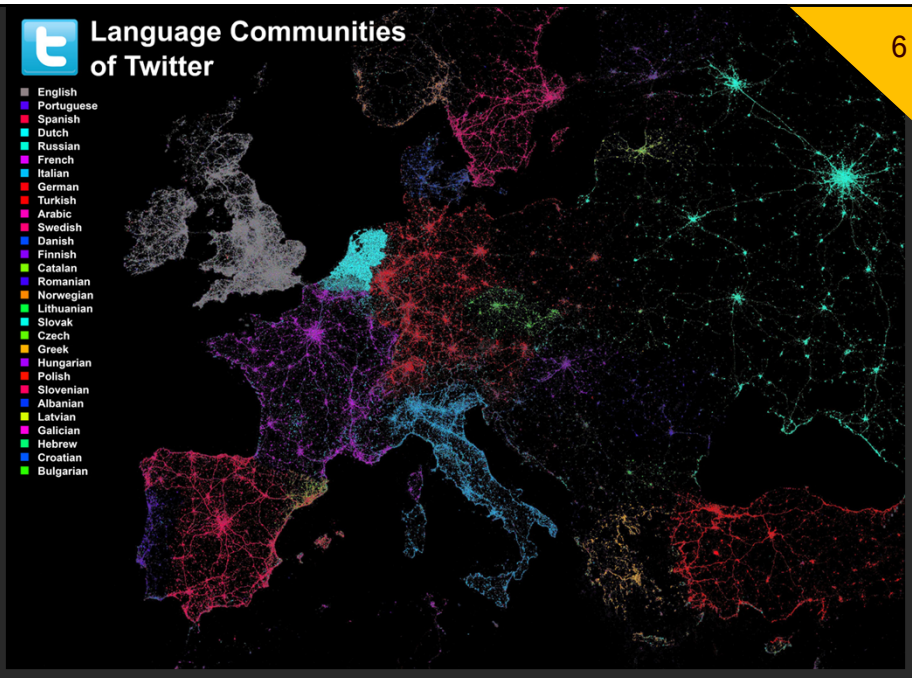
Research Collaborations by the Chinese Academy of Sciences

5

Huang, Duhon, Hardy & Börner

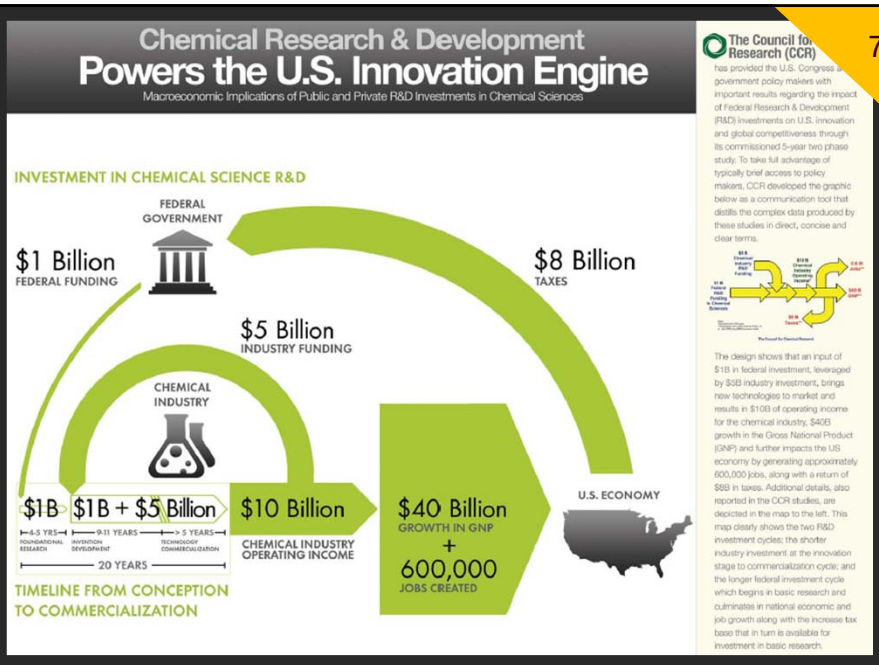


16



6

Language Communities of Twitter - Eric Fischer - 2012

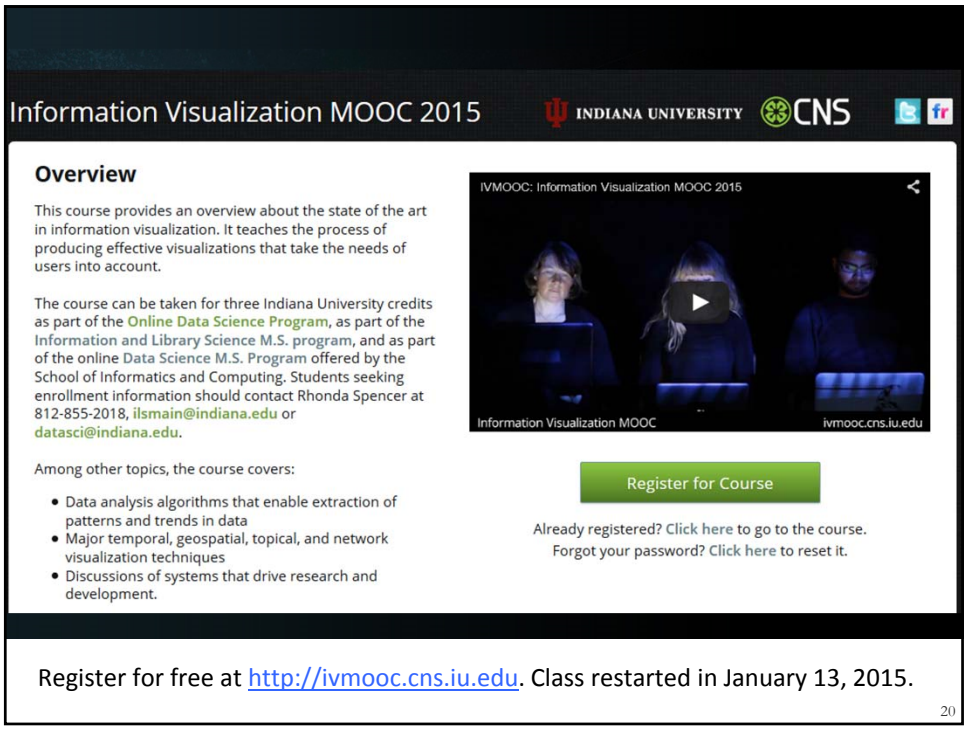


7

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.

20

Course Schedule

8

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

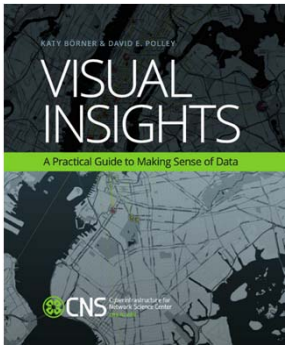
Final Exam

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%).

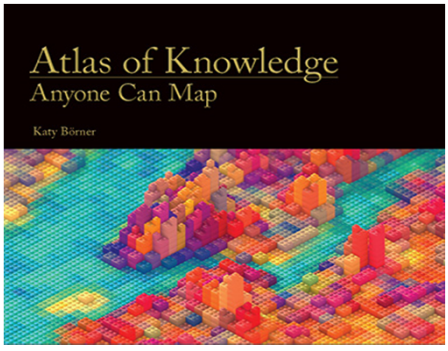
21

Books Used in the IVMOOC



Teaches timely knowledge:

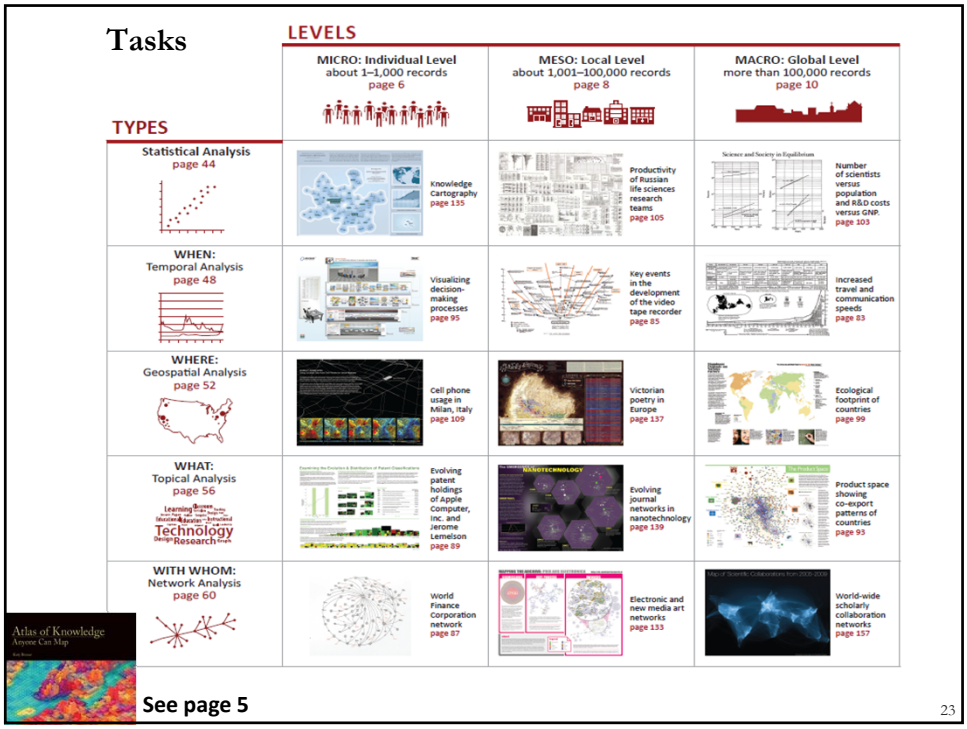
Advanced algorithms, tools, and hands-on workflows.

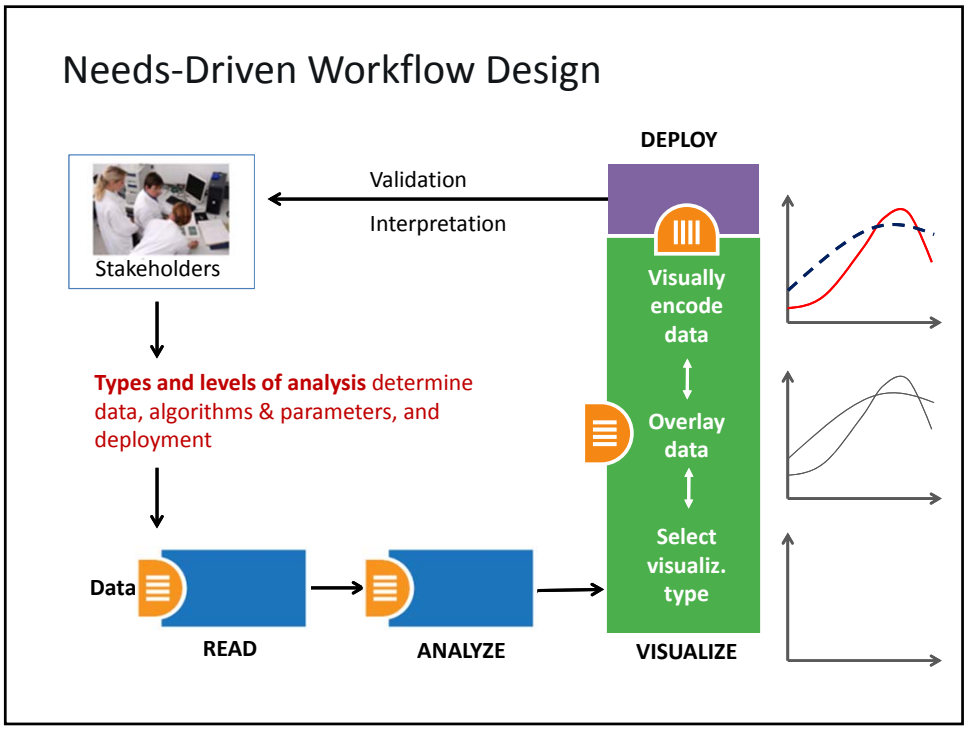


Teaches timeless knowledge:

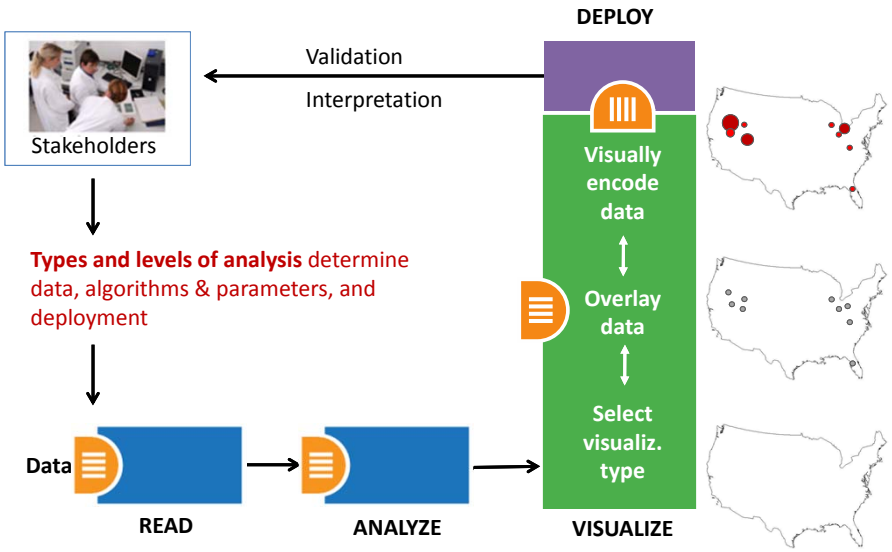
Visualization framework—exemplified using generic visualization examples and pioneering visualizations.

22

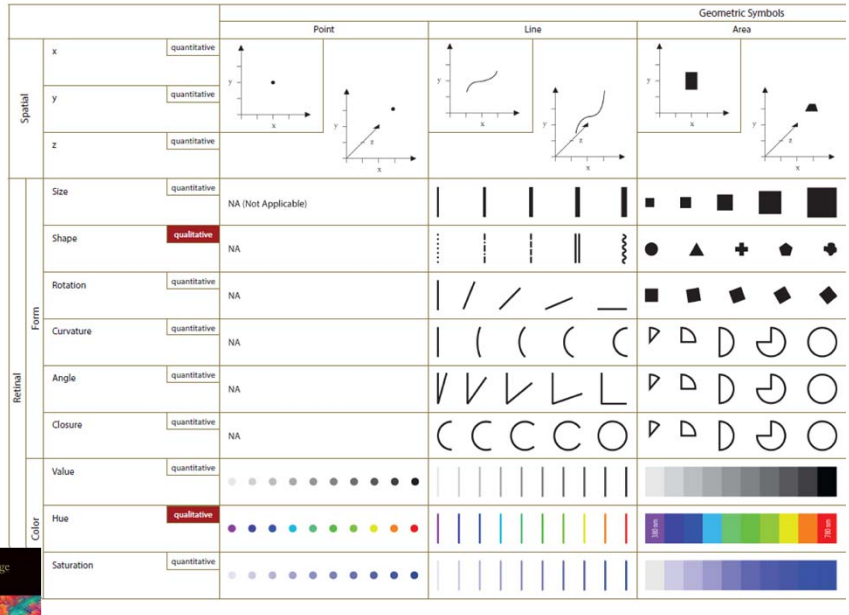

Tasks
LEVELS
MICRO: Individual Level
about 1–1,000 records
page 6
MESO: Local Level
about 1,001–100,000 records
page 8
MACRO: Global Level
more than 100,000 records
page 10
TYPES
Statistical Analysis
page 44
Knowledge Cartography
page 135
Productivity of Russian life sciences research teams
page 105
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
Atlas of Knowledge
Anyone Can Map
See page 5
23

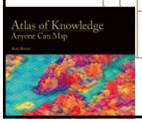

Needs-Driven Workflow Design
DEPLOY
Validation
Interpretation
Stakeholders
Types and levels of analysis determine data, algorithms & parameters, and deployment
Visually encode data
Overlay data
Select visualiz. type
Data
READ
ANALYZE
VISUALIZE

Needs-Driven Workflow Design



Graphic Variable Types Versus Graphic Symbol Types

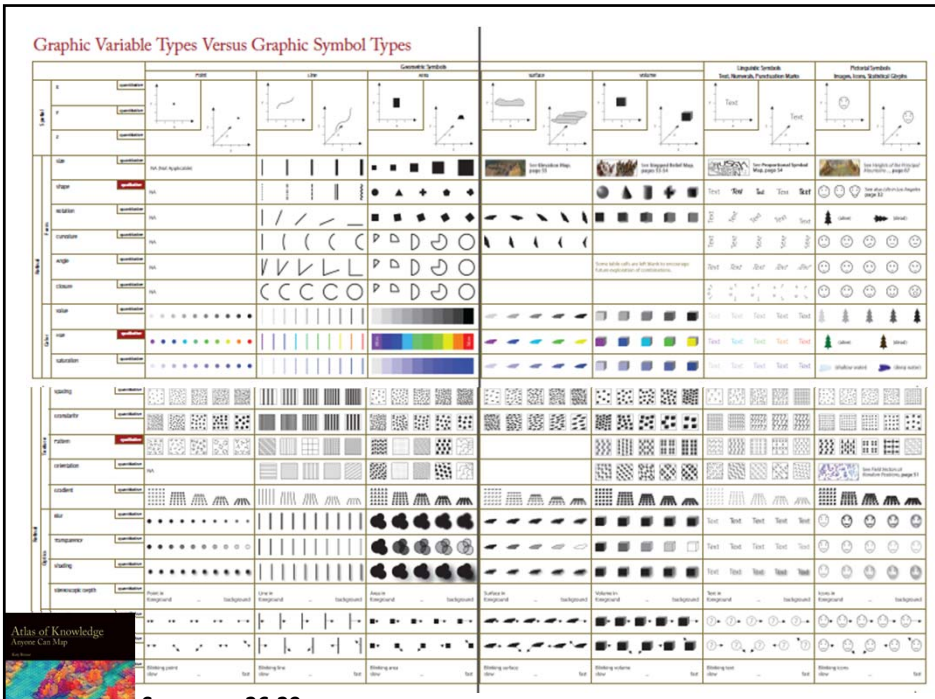




See page 36

26

Graphic Variable Types Versus Graphic Symbol Types



See pages 36-39

27

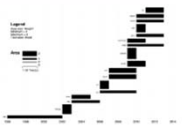
Load **One** File and Run **Many** Analyses and Visualizations

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



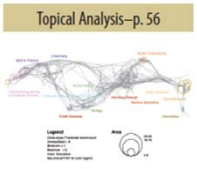
Geospatial Analysis–p. 52

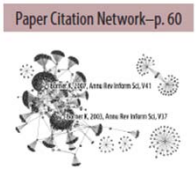
Geospatial Analysis–p. 52

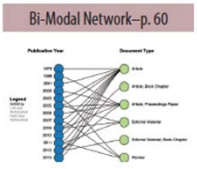
28

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







Co-author and many other bi-modal networks.

29

READINGS

Papers

- Stipelman, Brooke A., Hall, Kara L., Zoss, Angela, Okamoto, Janet, Stokols, Dan, and Börner, Katy (2014) Mapping the Impact of Transdisciplinary Research: A Visual Comparison of Investigator Initiated and Team Based Tobacco Use Research Publications. *The Journal of Translational Medicine and Epidemiology*.
- Bollen, Johan, David Crandall, Damion Junk, Ying Ding, and Katy Börner. 2014. From funding agencies to scientific agency: Collective allocation of science funding as an alternative to peer review. *EMBO Reports* 15 (1): 1-121.
- 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.

Books

- Börner, Katy. 2015. *Atlas of Knowledge: Anyone Can Map*. The MIT Press.
- Börner, Katy, and David E. Polley. 2014. *Visual Insights: A Practical Guide to Making Sense of Data*. Cambridge, MA: The MIT Press.
- Scharnhorst, Andrea, Katy Börner, and Peter van den Besselaar, eds. 2012. *Models of Science Dynamics*. Springer Verlag.
- Börner, Katy, Mike Conlon, Jon Corson-Rikert, and Ying Ding, eds. 2012. *VIVO: A Semantic Approach to Scholarly Networking and Discovery*. Morgan & Claypool Publishers LLC.
- Börner, Katy. 2010. *Atlas of Science: Visualizing What We Know*. The MIT Press.

30

Existing Interfaces for VATS-like 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

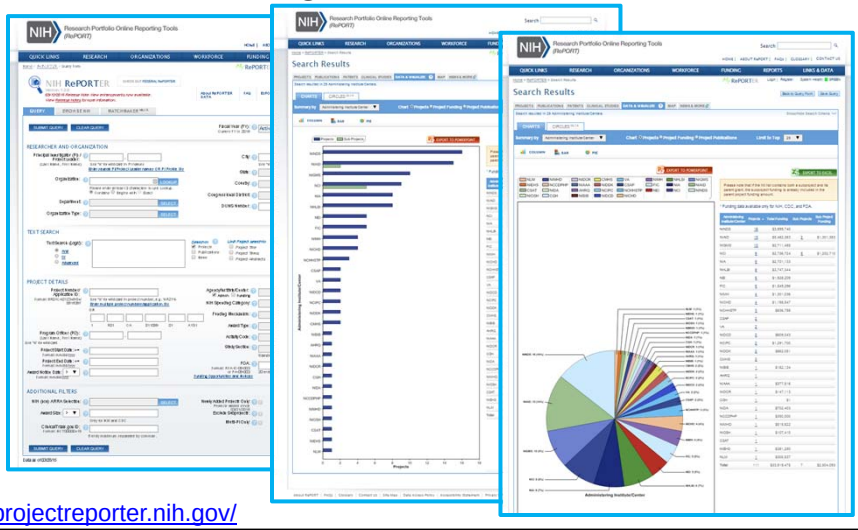
14) Scraawl: Twitter Analysis

15) Illuminated Diagram: Searchable World and Science Maps

CNS Cyberinfrastructure for Network Science Center
VATS Visualizations: Existing Interfaces

9

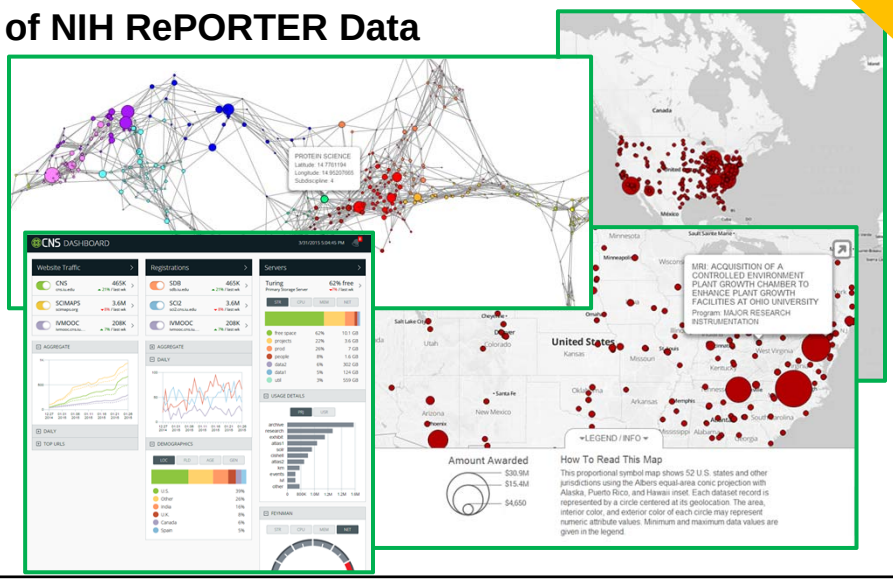
NIH RePORTER: Visual Interface to Biomedical Funding Data in US



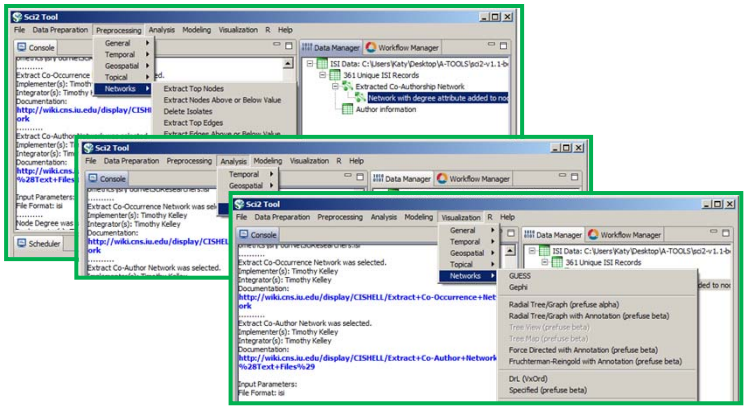
http://projectreporter.nih.gov/

10

CIShell/Sci2 World and Science Visualizations of NIH RePORTER Data

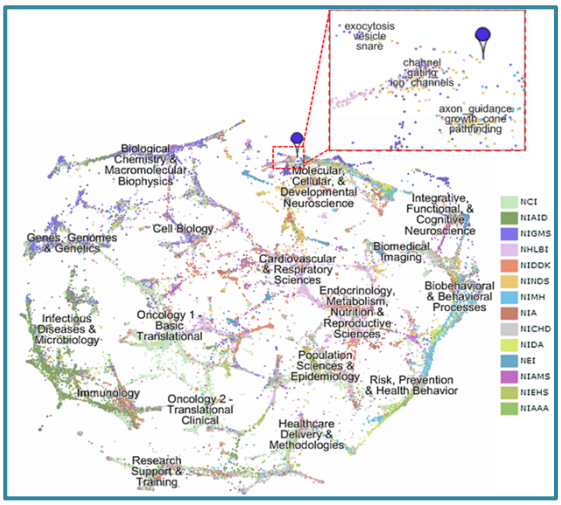


Sci2 Desktop



11

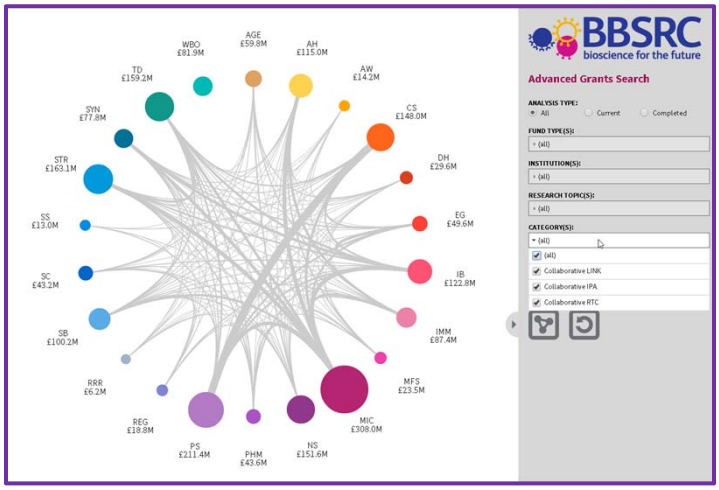
NIH RePORTER: NIH Map



http://nihmaps.org/

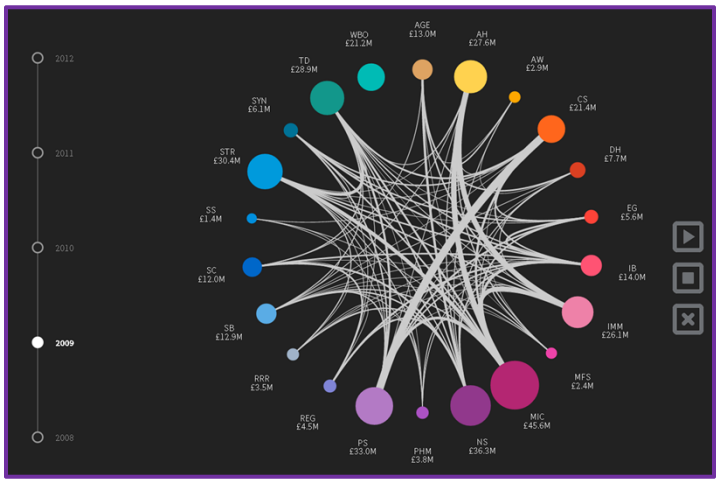
12

BBSRC: Visual Interface to Biomedical Funding Data in UK



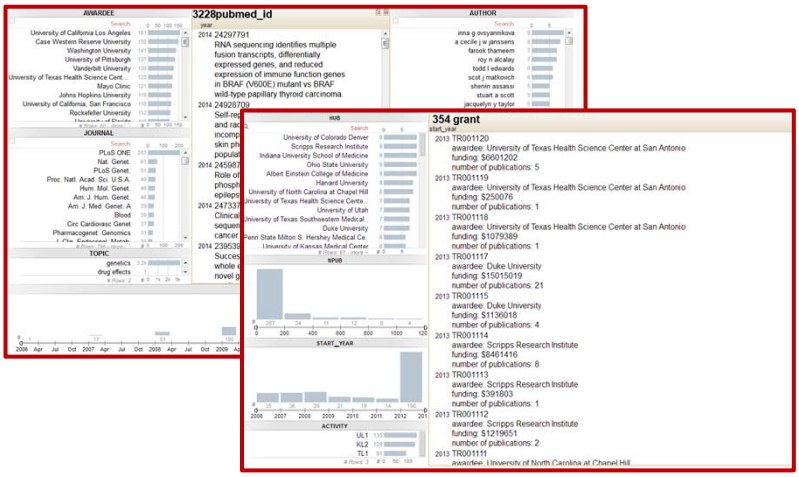
http://www.bbsrc.ac.uk/

BBSRC: Temporal animation

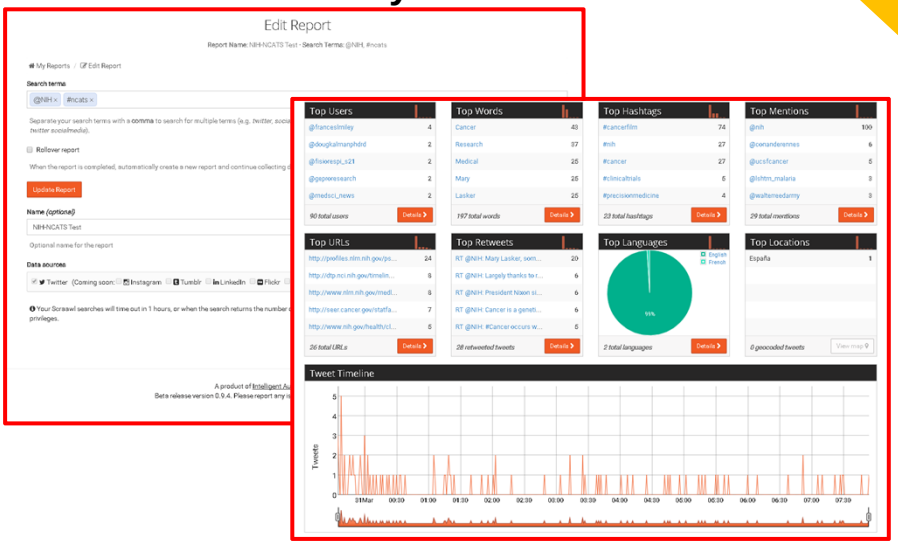


13

IAI Multidimensional Analysis



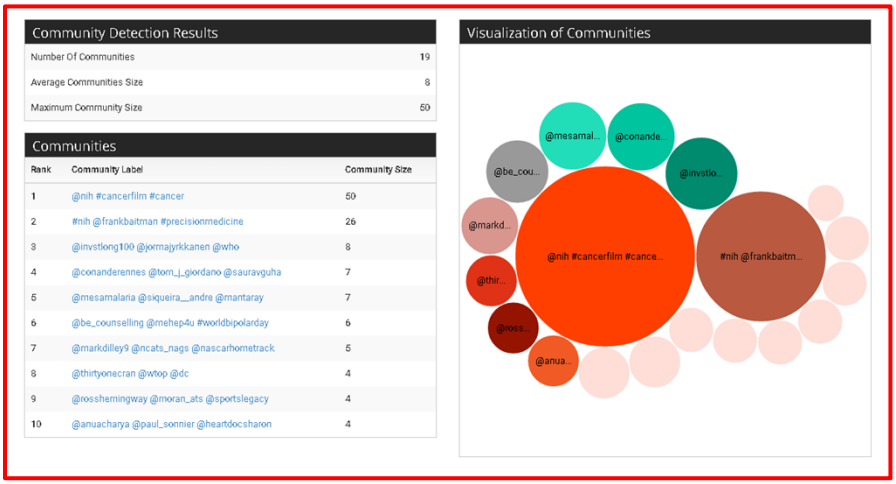
Scraawl: Twitter Analysis



https://www.scraawl.com/

CNS Cyberinfrastructure for Network Science Center
VATS Visualizations: Existing Interfaces

Scraawl: Advanced analysis



Community Detection Results

Number Of Communities	19
Average Communities Size	8
Maximum Community Size	50

Communities

Rank	Community Label	Community Size
1	@nih #cancerfilm #cancer	50
2	#nih @frankbaitman #precisionmedicine	26
3	@invstlong100 @jormajyrkkanen @who	8
4	@conanderennes @tom_j_giordano @sauravguha	7
5	@mesamalaria @siqueira__andre @mantaray	7
6	@be_counselling @mehep4u #worldbipolarday	6
7	@markdilley9 @ncats_nags @nascarhometrack	5
8	@thirtyonecran @wtop @dc	4
9	@rossherningway @moran_ats @sportslegacy	4
10	@anuacharya @paul_sonnier @heartdocsharon	4

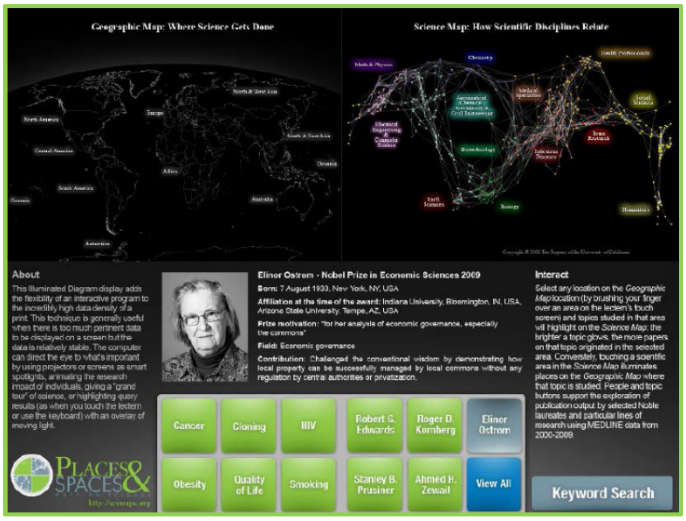
15

Illuminated Diagram: Searchable World and Science Maps

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

CNS Cyberinfrastructure for Network Science Center
VATS Visualizations: Existing Interfaces

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/

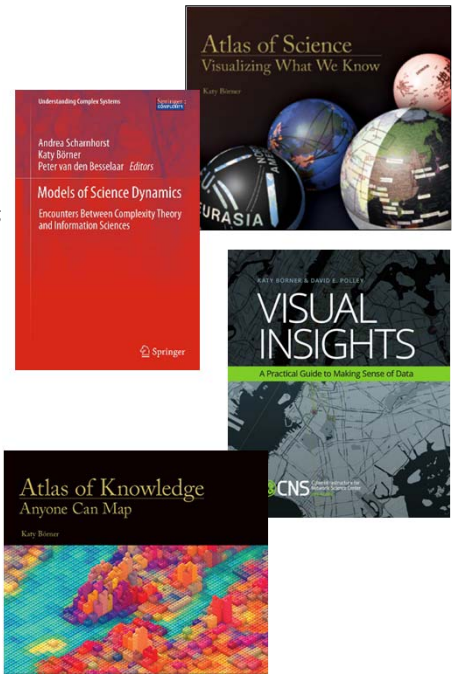
Börner, Katy (2010) **Atlas of Science: Visualizing What We Know**. The MIT Press. http://scimaps.org/atlas

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