Science of Science Research and Tools
Tutorial #12 of 12

Dr. Katy Börner
Cyberinfrastructure for Network Science Center, Director
Information Visualization Laboratory, Director
School of Library and Information Science
Indiana University, Bloomington, IN
http://info.slis.indiana.edu/~katy

With special thanks to Kevin W. Boyack, Micah Linnemeier,
Russell J. Duhon, Patrick Phillips, Joseph Biberstine, Chintan Tank
Nianli Ma, Hanning Guo, Mark A. Price, Angela M. Zoss, and
Scott Weingart

Invited by Robin M. Wagner, Ph.D., M.S.
Chief Reporting Branch, Division of Information Services
Office of Research Information Systems, Office of Extramural Research
Office of the Director, National Institutes of Health

Suite 4090, 6705 Rockledge Drive, Bethesda, MD 20892
noon-2p, July 28, 2010

---

12 Tutorials in 12 Days at NIH—Overview

1.	Science of Science Research	**1st Week**
2.	Information Visualization	
3.	CIShell Powered Tools: Network Workbench and Science of Science Tool	

4.	Temporal Analysis—Burst Detection	**2nd Week**
5.	Geospatial Analysis and Mapping	
6.	Topical Analysis & Mapping	

7.	Tree Analysis and Visualization	**3rd Week**
8.	Network Analysis	
9.	Large Network Analysis	

10.	Using the Scholarly Database at IU	**4th Week**
11.	VIVO National Researcher Networking	
12.	**Future Developments**	

[#12] Future Developments

➢ Validation Studies

➢ Needed Data/Documentation

➢ Needed and New Tool Functionality

➢ Needed Documentation/Tutorials

➢ Promising Research Questions

➢ Exercise: Identify Promising Collaborations

Recommended Reading

Börner, Katy (2010) Atlas of Science. MIT Press. *http://scimaps.org/atlas*

Börner, Katy, Bettencourt, Luis M. A., Gerstein, Mark & Uzzo, Stephen Miles (Eds.), Knowledge Management and Visualization Tools in Support of Discovery. (2009). NSF CDI Initiative Workshop Report, National Science Foundation, Indiana University. *http://vw.slis.indiana.edu/cdi2008/whitepaper.html*

3

[#12] Future Developments

➢ **Validation Studies**

➢ **Needed Data/Documentation**

➢ **Needed and New Tool Functionality**

➢ **Needed Documentation/Tutorials**

➢ **Promising Research Questions**

➢ **Exercise: Identify Promising Collaborations**

4

OPINION

Let's make science metrics more scientific

To capture the essence of good science, stakeholders must combine forces to create an open, sound and consistent system for measuring all the activities that make up academic productivity, says **Julia Lane**.

SUMMARY
- Existing metrics have known flaws
- A reliable, open, joined-up data infrastructure is needed
- Data should be collected on the full range of scientists' work
- Social scientists and economists should be involved

5

Validating Science Maps

Boyack, Kevin W., Klavans, Richard & Börner, Katy. (2005). Mapping the Backbone of Science. Scientometrics. Vol. 64(3), 351-374.

Eight alternative measures of journal similarity were applied to a data set of 7,121 journals covering over 1 million documents in the combined Science Citation and Social Science Citation Indexes. For each journal similarity measure we generated two-dimensional spatial layouts using the force-directed graph layout tool, VxOrd. Next, mutual information values were calculated for each graph at different clustering levels to give a measure of structural accuracy for each map. The best co-citation and inter-citation maps according to local and structural accuracy were selected and are presented and characterized. These two maps are compared to establish robustness. The inter-citation map is then used to examine linkages between disciplines.
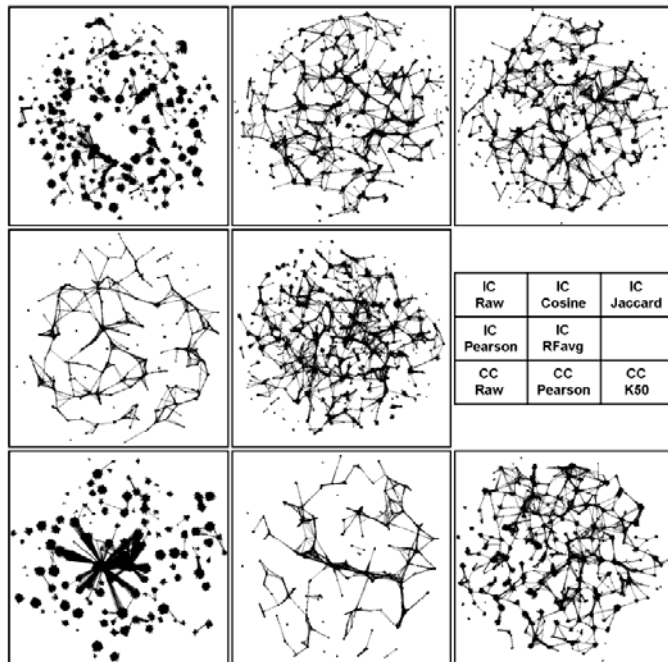

IC Raw	IC Cosine	IC Jaccard
IC Pearson	IC RFavg	
CC Raw	CC Pearson	CC K50

6

Table 1. Summary of validation results for maps based on eight similarity measures.

Measure	Local accuracy @ 95% coverage[1]	Scalability[1]	Z-score for 200 clusters	Clustering (qualitative)
IC-Raw	60.1%	High	360.0	Too few, loose
IC-Cosine	80.2%	High	381.3	Good balance
IC-Jaccard	79.5%	High	387.1	Good balance
IC-Pearson	71.7%	Low	386.5	Too tight
IC-RFavg	80.2%	High	373.3	Good balance
CC-Raw	25.6%	High	294.9	Too few, loose
CC-Pearson	65.3%	Low	377.0	Too tight
CC-K50	71.4%	High	376.6	Good balance

Figure 4. Map of science generated using the IC-Jaccard similarity measure. The map is comprised of 7,121 journals from year 2000. Large font size labels identify major areas of science. Small labels denote the disciplinary topics of nearby large clusters of journals
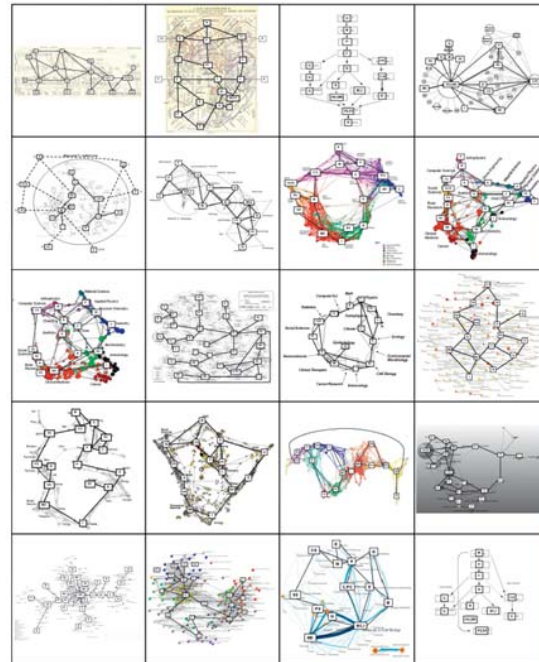
Validating Science Maps

Klavans, R., & Boyack, K. W. (2009). Toward a consensus map of science. Journal of the American Society for Information Science and Technology, 60(3), 455-476.

A consensus map of science is generated from an analysis of twenty existing maps of science. These twenty maps occur in three basic forms: hierarchical, centric, and non-centric (or circular). The consensus map, generated from consensus edges that occur in at least half of the input maps, emerges in a circular form. The ordering of areas is as follows: mathematics is (arbitrarily) placed at the top of the circle, and is followed clockwise by physics, physical chemistry, engineering, chemistry, earth sciences, biology, biochemistry, infectious diseases, medicine, health services, brain research, psychology, humanities, social sciences, and computer science.

The circular map of science is found to have a high level of correspondence with the twenty existing maps, and has a variety of advantages over hierarchical and centric forms.

Table 1: Characteristics of twenty comprehensive maps of science. Abbreviations SC, SS, AH, and PR refer to Thomson Scientific's Science, Social Science, Arts & Humanities, and Proceedings Citation databases, respectively.

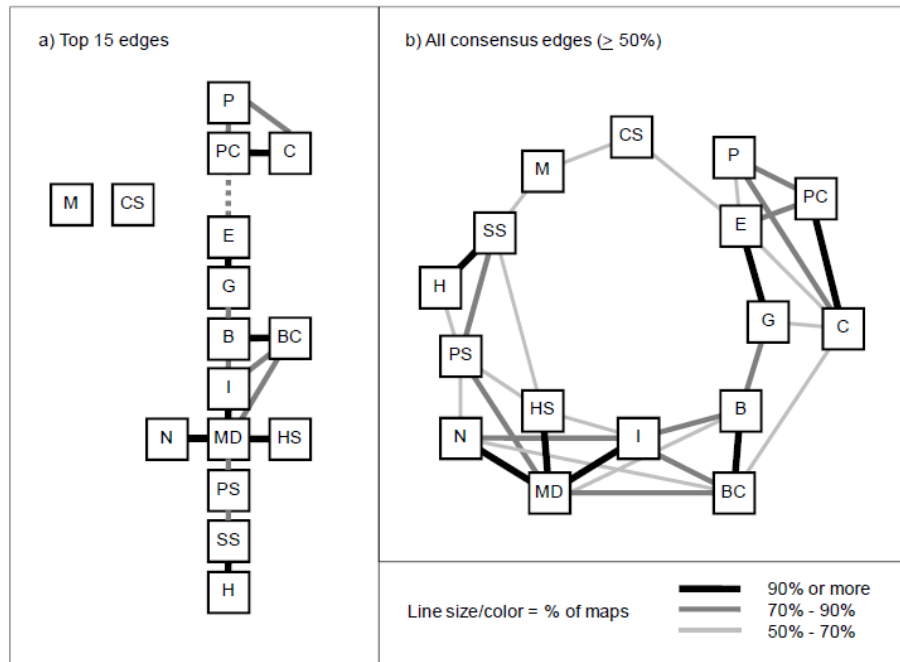| Researcher(s) & Reference	Map Name	Method	Elements	# Clust	Database & Year	Form
(Bernal, 1939)	Bernal	Expert		14, 110		Hierarchical
(Ellingham, 1948)	Ellingham	Expert		13, 51, 130		Hierarchical & Non-centric
(Balaban & Klein, 2006)	Balaban-I	Expert	16 fields	16		Hierarchical & Centric
(Griffith, Small, Stonehill, & Dey, 1974)	Small74	Reference papers	1,150 pap	41	SC, 1972 Q1	Centric
(Small & Garfield, 1985)	Small85	Reference papers	~11,000 pap	51	SC+SS, 1983	Hierarchical & Centric
(Small, 1999)	Small99	Reference papers	36,720 pap	35	SC+SS, 1995	Hierarchical
(Klavans & Boyack, 2008) [a]	KB-Para	Reference papers	800k pap	776	SC+SS, 2003	Non-centric
(Klavans & Boyack, 2007)	KB06-TS	Reference papers	1.9M pap	283	SC+SS, 2004	Non-centric
(Klavans & Boyack, 2007)	KB06-SC	Reference papers	2.1M pap	554	Scopus, 2004	Non-centric
(Bassecoulard & Zitt, 1999)	B-Z	Journals	~2,000 jnl	29	SC/JCR, 1993	Hierarchical & Centric
Klavans, unpublished, 2002	K02	Journals	5,647 jnl	69	SC+SS+AH, 2000	Non-centric
(Boyack, Klavans, & Börner, 2005)	Backbone	Journals	7,121 jnl	205	SC+SS, 2000	Non-centric
(Boyack et al., 2009)	BBK02-S	Journals	7,227 jnl	671	SC+SS, 2002	Non-centric
(Boyack, 2009)	B03-ST	Journals	8,667 jnl	852	SC+SS+PR, 2003	Non-centric
(Klavans, Boyack, & Patek, 2008) [b]	UCSD	Journals	16,235 jnl	554	SC/SS/AH + Scopus, 2001-05	Non-centric
(Rosvall & Bergstrom, 2008) [c]	Rosvall	Journals	6,116 jnl	87	SC+SS, 2004	Non-centric
(Moya-Anegón et al., 2004)	Scimago-I	Journal categories	25 categ	25	SC+SS+AH, 2000 Spanish papers	Non-centric
(Moya-Anegón et al., 2007) [d]	Scimago-II	Journal categories	219 categ	219	SC+SS+AH, 2002	Centric
(Leydesdorff & Rafols, 2008) [e]	L-R	Journal categories	6,164 jnl; 172 categ	172	SC, 2006	Mixed
(Balaban & Klein, 2006)	Balaban-II	Course prerequisites		11	Texas A&M undergraduate	Centric

Figure 5: Two-dimensional consensus maps of science from twelve of the twenty input maps, excluding the eight input maps from Klavans, Boyack, and Börner.

Validating Science Maps

Accuracy of Models for Mapping the Medical Sciences
Kevin W. Boyack, Richard Klavans, SciTech Strategies Inc.
Katy Börner, Russell J. Duhon, Nianli Ma, Indiana
University, Bob Schijvenaars, Aaron Sorensen, Collexis
Holdings Inc., André Skupin, San Diego State University

This project aims to provide a highly accurate interactive map of medical research that can be easily used by both technical and non-technical users. Phase I of this project compares and determines the relative accuracies of maps of medical research based on commonly used text-based and citation-based similarity measures at a scale of over two million documents.

All work is documented in real time at http://sci.slis.indiana.edu/sts and at a level of detail that supports the exact replication of work.

Evaluate The Efficacy of Self Organizing Maps For The Representation and Organization of (MEDLINE) data

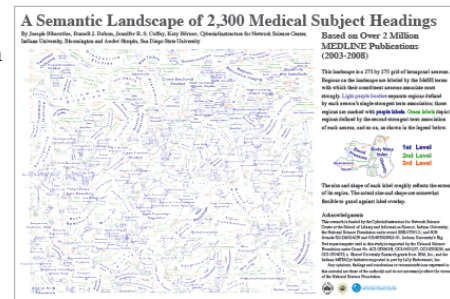*Biberstine, Joseph, Skupin, Andre, Duhon, Russell & Börner, Katy. (2010)*

Dear all,

We are interested to evaluate the efficacy of Self Organizing Mapping Algorithms toward the representation and organization of MEDLINE data. Because we know of your interest in visualization and network science, and particularly because of your expertise in medicine, medical informatics, and/or bioinformatics we would like to invite you to participate in a study sponsored by the Cyberinfrastructure for Network Science Center and Dr. Katy Börner.

This study will examine the use of Self-Organizing Maps to represent vast MEDLINE dataset in a manner that is easy to use and understand. If you would like to participate, we would like to arrange a **½ hour meeting** with you tomorrow**, 7/28 at 2pm (right after the 12th tutorial).**

Your participation will consist of a couple of tasks using the Medline map, and your written impressions of the map and its usefulness. The data we collect during these sessions will form the basis of a future publication, and the feedback we receive on the use of the self-organizing mapping algorithm will help refine the output to become more intuitive and user friendly to professionals who need access to large sets of data in a direct, representational fashion.

Thank you for considering our request.

K. Borner

Based on over 2 million MEDLINE publications (2003-2008) and their 2,300 Medical Subject Headings

275 by 275 grid of hexagonal neurons Regions are labeled by the MeSH terms with which their constituent neurons associate most strongly

13

[#12] Future Developments

➢ **Validation Studies**

➢ **Needed Data/Documentation**

➢ **Needed and New Tool Functionality**

➢ **Needed Documentation/Tutorials**

➢ **Promising Research Questions**

➢ **Exercise: Identify Promising Collaborations**

14

Data (Documentation) Needs, see Tutorials 10 and 11

Informed science and technology policy (and Science of Science Studies) depend on comprehensive and useful data that has high
➢ Accuracy
➢ Integrity (structured & managed)
➢ Consistency
➢ Validity (rules, standards are followed)
➢ Reliability

However, publications, patents, grants are kept in data silos with few interlinkages, incompatible formats, unknown quality and coverage.

Needed Data/Documentation

The quality and topic coverage of data, who provides/updates it, and different linkages has to be known.

Data Providers
Name | Institution | Contact info/email | Geolocation (ZIP if in US, city+country otherwise)

Datasets
Dataset Name | Original Source | URL | # Records | Link to raw data sample | Ontology/structure/data dictionary | topic coverage, e.g., medicine, CS | Type, e.g., patents, funding, genes | derivative datasets, e.g., calculated unique names, geolocations | Available since when?

Services
What tools/services use what datasets?
Service Name | URL | Type of functionality | Available since when?

People—Data Linkages

Show who contributes how many datasets but also what datasets are served by multiple parties. Need listing of

People Name | Dataset Name

Data—Data Linkages

Carol Goble commented that a "semantic exposure turns datasets 'inside out'" and semantically exposed datasets "'snap together' like the metal pieces that make up Terminator".

To show how the different datasets combine, we need a listing of

Dataset Name 1 | Dataset Name 2 | Mapped classes/attributes/linkages, etc. | #matches | # records in dataset 1 | # records in dataset 2

Data—Services Linkages

The number of services that use a dataset is a major indictor of its quality, reliability, and utility. Need listing of

Dataset Name | Service Name(s)

[#12] Future Developments

➢ **Validation Studies**

➢ **Needed Data/Documentation**

➢ **Needed and New Tool Functionality**

➢ **Needed Documentation/Tutorials**

➢ **Promising Research Questions**

➢ **Exercise: Identify Promising Collaborations**

Needed Tool Functionality

➢ TexTrend—linguistic analysis and evolving networks analysis, see http://textrend.org

➢ Epic Tool—network models and diffusion models for epidemics of diseases and other tangible (people, specimen equipment) and intangible (ideas, innovation) objects.

➢ CIShell branding of CIShell powered tools. NWB, Sci2, EpiC, TexTrend, …

➢ New plugins that address the needs identified in the last 12 days.

➢ Read SAS files. Connect to DB.

➢ Universal undo. Stop processes via scheduler.

➢ Dedicated server for Sci2 Tool + new plugins for federal usage (IU's computing infrastructure is fully HIPPA compliant).

➢ Visualization Layers, see **Tutorial 2**.

New Sci2 Tool Functionality –
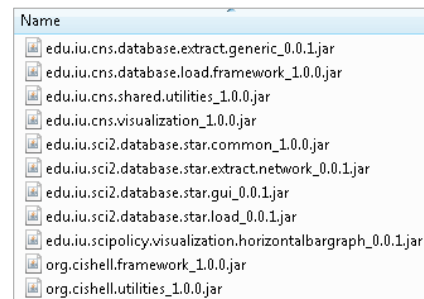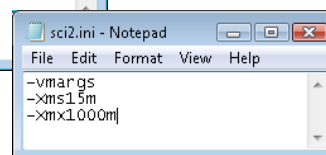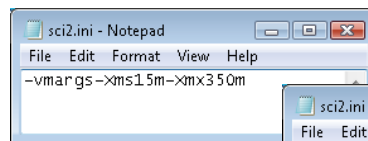## Colored Horizontal Bargraph and STAR Database
(Still need to be tested and documented)

Setup

➢ Copy and replace all plugins from /NewPlugins/plugins directory on memory stick into your Sci2 Tool /plugins directory.

➢ If you have at least 2GB of memory installed, increase memory usage by replacing text in /sci2/sci2.ini file

by

-vmargs

-Xms15m

-Xmx1000m

Name
edu.iu.cns.database.extract.generic_0.0.1.jar
edu.iu.cns.database.load.framework_1.0.0.jar
edu.iu.cns.shared.utilities_1.0.0.jar
edu.iu.cns.visualization_1.0.0.jar
edu.iu.sci2.database.star.common_1.0.0.jar
edu.iu.sci2.database.star.extract.network_0.0.1.jar
edu.iu.sci2.database.star.gui_0.0.1.jar
edu.iu.sci2.database.star.load_0.0.1.jar
edu.iu.scipolicy.visualization.horizontalbargraph_0.0.1.jar
org.cishell.framework_1.0.0.jar
org.cishell.utilities_1.0.0.jar

sci2.ini - Notepad
File   Edit   Format   View   Help
-vmargs-Xms15m-Xmx350m

sci2.ini - Notepad
File   Edit   Format   View   Help
-vmargs
-Xms15m
-Xmx1000m

➢ Run the sci2.exe file to start Sci2 Tool.

New Sci2 Tool Functionality
(Still needs to be tested and documented)

Colored Horizontal Bargraph

➢ Load NIH-CTSA-Grants.csv as a csv file.

➢ Run

with parameter values:

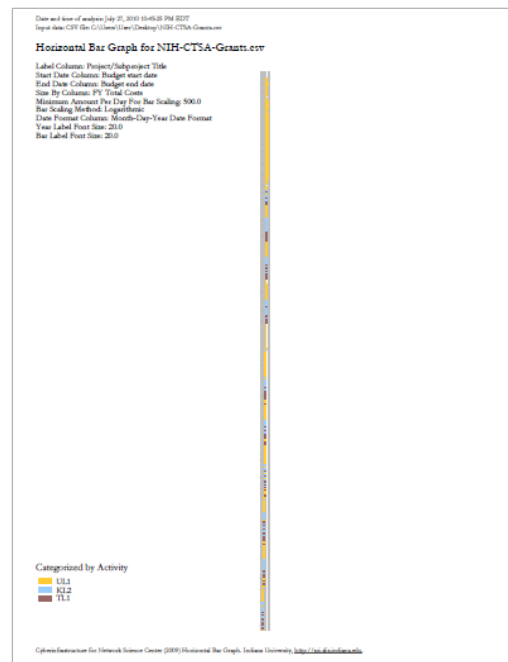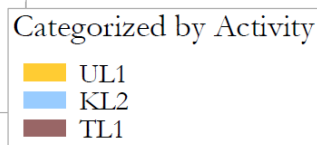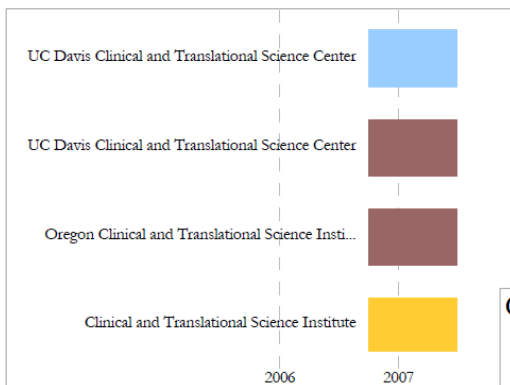➢ Save resulting ps file

➢ View it in PSViewer.

New Sci2 Tool Functionality
(Still needs to be tested and documented)

Colored Horizontal Bargraph cont.

Letter size visualization with legend, see right.
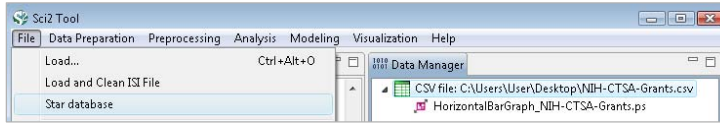
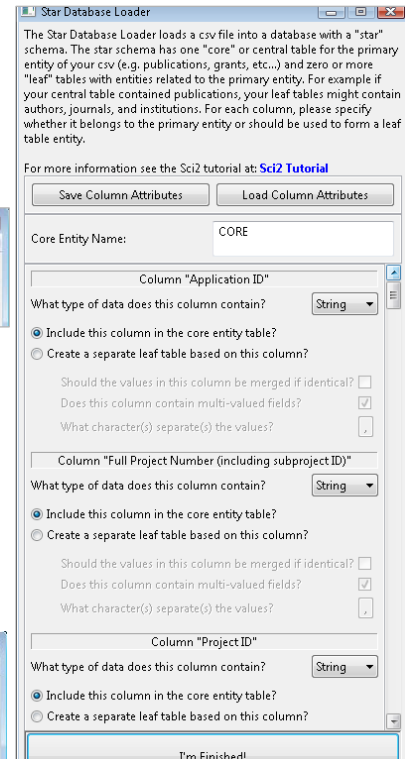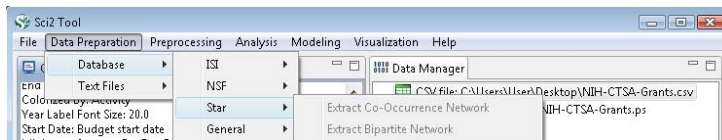Zoom into first awards, below.

New Sci2 Tool Functionality
(Still needs to be tested and documented)

STAR Database
➢ Select NIH-CTSA-Grants.csv in Data Manager.
➢ Run 'File > Star Database'

➢ Use GUI (see right) to setup database.
➢ For tabular CSV data with a large quantity of columns, save chosen attributes for use/modification later using 'Save Column Attributes' button. Reuse with 'Load Column Attributes'.
➢ Select 'I'm Finished' to load data into database.
➢ The resulting Star Database appears as a child in the Data Manager and can be used via

Star Database Loader

The Star Database Loader loads a csv file into a database with a "star" schema. The star schema has one "core" or central table for the primary entity of your csv (e.g. publications, grants, etc...) and zero or more "leaf" tables with entities related to the primary entity. For example if your central table contained publications, your leaf tables might contain authors, journals, and institutions. For each column, please specify whether it belongs to the primary entity or should be used to form a leaf table entity.

For more information see the Sci2 tutorial at: Sci2 Tutorial

Save Column Attributes | Load Column Attributes

Core Entity Name: CORE

Column "Application ID"
What type of data does this column contain? String
◉ Include this column in the core entity table?
○ Create a separate leaf table based on this column?
  Should the values in this column be merged if identical?
  Does this column contain multi-valued fields?
  What character(s) separate(s) the values?

Column "Full Project Number (including subproject ID)"
What type of data does this column contain? String
◉ Include this column in the core entity table?
○ Create a separate leaf table based on this column?
  Should the values in this column be merged if identical?
  Does this column contain multi-valued fields?
  What character(s) separate(s) the values?

Column "Project ID"
What type of data does this column contain? String
◉ Include this column in the core entity table?
○ Create a separate leaf table based on this column?

I'm Finished!

23

[#12] Future Developments

➢ **Validation Studies**
➢ **Needed Data/Documentation**
➢ **Needed and New Tool Functionality**
➢ **Needed Documentation/Tutorials**
➢ **Promising Research Questions**
➢ **Exercise: Identify Promising Collaborations**

24

Sci² Tool
A tool for science of science research & practice

Email Address

Password

Login

Forgot your password?

To recover your account password, please visit our **password recovery page**.

Not registered yet?

Register now

Tutorials

Scott Weingart, Hanning Guo, Katy Börner, Kevin W. Boyack, Micah W. Linnemeier, Russell J. Duhon, Patrick A. Phillips, Chintan Tank, and Joseph Biberstine (2010) **Science of Science (Sci2) Tool User Manual.** Cyberinfrastructure for Network Science Center, School of Library and Information Science, Indiana University, Bloomington.

Katy Börner (2010) Science of Science Research and Tools (12 Tutorials). Reporting Branch, Office of Extramural Research/Office of the Director, National Institutes of Health, Bethesda, MD.

- Tutorial #01: **Science of Science Research**
- Tutorial #02: **Network Science / Information Visualization**
- Tutorial #03: **CIShell Powered Tools: Network Workbench and Science of Science Tool**
- Tutorial #04: **Temporal Analysis—Burst Detection**
- Tutorial #05: **Geospatial Analysis and Mapping**
- Tutorial #06: **Topical Analysis & Mapping**
- Tutorial #07: **Tree Analysis and Visualization**
- Tutorial #08: **Network Analysis and Visualization**
- Tutorial #09: **Large Network Analysis and Visualization**
- Tutorial #10: Using the Scholarly Database at IU
- Tutorial #11: VIVO National Researcher Networking
- Tutorial #12: Future Developments

http://sci.slis.indiana.edu/sci2

[#12] Future Developments

- ➢ **Validation Studies**
- ➢ **Needed Data/Documentation**
- ➢ **Needed and New Tool Functionality**
- ➢ **Needed Documentation/Tutorials**
- ➢ **Promising Research Questions**
- ➢ **Exercise: Identify Promising Collaborations**

Katy's Original Suggestions

1. Genealogy of NIH Funding

Horizontal Bar Graph (HBG) Visualization of RFAs and PAs for all NIH for last 10 years. Size code bars by # awards or $ amount or publication output, and # RCDC concepts as an indicator of interdisciplinarity. Suggested data format:

Institute Name or 'General' | RFAs and PAs title | RFAs and PAs # | start date* | end date *| # awards | total $ awarded to date | # linked publications | RCDC concepts of all awards made separated by ;

2. Genealogy of Extramural Staff

HBG of Program Officers for all NIH data. Size code bars by # awards made, total $ amount for those, publication output, and # RCDC concepts as an indicator of interdisciplinarity. Suggested data format:

Institute Name | Program Officer Name | start date* | end date* | # awards | total $ awarded to date | # publications |RCDC concepts of awards made separated by ;

3. Genealogy of Investigators

HBG of NIH Investigators for all NIH data

Size code bars by # awards received, total $ amount for those, publication output and # RCDC concepts as in indicator of interdisciplinarity. Suggested data format:

Institute Name | Investigator Name | start date | end date | # awards | total $ awarded to date | # publications | RCDC concepts of awards separated by ;

*Use start and end dates of the awards under the RFAs and PAs/Program Officer.

Katy's Original Suggestions cont.

4. Program Officer – PI Network

Show network of who is (co-)funding whom? Program Officers (POs) will be color coded by institute, size coded by number of PIs funded. I am curious to see the trajectories of investigators between different institutes and the 'scholarly' networks that exist.

Suggested data format:

Program Officer name | Program Officer Institute | list of funded PIs separated by ;

5. Program Officer – PI Topic Space

Does the expertise of the POs (according to what they fund – RCDC concepts) match the topics that investigators submit (according to their proposals – RCDC)? How does this network compare with 4.)?

6. Solicitations -> Proposals -> Publications Topic Resemblance

What solicitations attract proposals with what RCD concepts, result in what RCDC concept awards, that are later linked to what MeSH term publications?

Note that RCDC categories/concepts are not assigned to RFAs and PAs.

Brainstorming Results – Existing and Planned Analyses

(See Online Survey at Sharepoint site)

Geographic - Existing Analyses
➢ Analyze the co-funding for individuals by different institutions in multiple NIH grants over time.
➢ Do all NIH grants produce publications? Perform an in-depth analysis of who isn't publishing.
➢ What is the main scientific strength by geography and connection to local economy/trends.

Temporal
➢ How/when/why do PIs leave the NIH system?
➢ Is it more efficient to fund centers or individuals?
➢ How has NIH funding to different individual institutions and institutions types (e.g. sector or Carnegie Classification) changed over time?
➢ Do NIH funding patterns vary over time by the demographic characteristics of PIs and why? (related to Diversity of Workforce)
➢ Develop total career track training, appointments, NIH and other support and publications in the post-academic career
➢ What new discoveries were made using ARRA funds, over time?

Brainstorming Results – Existing and Planned Analyses cont.

(See Online Survey at Sharepoint site)

Topical
➢ How much funding is awarded for each of the current top 5 NIH priorities (Director's Initiatives) per IC?
➢ Emerging scientific concepts for RCDC categories in one FY.
➢ Are genetic vs behavioral studies concentrated on different disease/conditions?
➢ What is the top research funded across ICs and how do they relate?
➢ How many awards were made between years XXX and YYY that included the words X,Y,Z in the title and abstract?
➢ Analyze the major areas of research funded during the tenure of the past 4 or 5 NIH Directors based upon their initiatives and how those funding levels have changed over time.

➢ How will Dr. Collins convince Congress to fund more grants by scientific areas?
➢ Compare NIH research portfolio to health profile of the nation. Are there obvious imbalances?
➢ Compare NIH funding with biomedical journal publications by scientific area
➢ Science - what are we missing and why?
➢ Is there any relationship between review panel composition and whether or not a grant is funded?

Brainstorming Results – Existing and Planned Analyses cont.
(See Online Survey at Sharepoint site)

Networks
➤ Is there clustering among different demographic groups in areas of science?
➤ How many researchers are there per grant?
➤ Unfunded research areas - fostering collaboration.
➤ What areas of biomedical science do institutions specialize in?
➤ Social networking of boiler-plate text in proposals. Who wrote it first and who copied it?

➤ Diffusion of topics (RCDC terms) across a social network map (co-author, co-PI, etc. - based on term occurrence)
➤ Text analytics - are papers cited for the same concepts they were funded to investigate?
➤ How does the # of personnel supported by ARRA grants differ in two reporting sources:
      1) the NIH progress report, All Personnel Report and
      2) ARRA recipient reporting database for FY09 and FY10?
➤ What is the profile of NIH funded researchers compared to NSF and other agencies e.g. in terms of institutional affiliations including training

---

Exercises Conducted After Each Tutorial
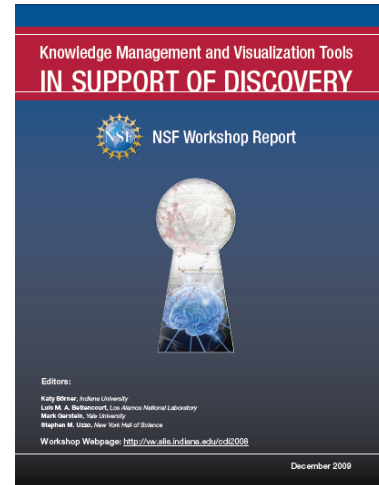*(See digital version of all feedback on Sharepoint site)*

➤ **Temporal Analysis**
➤ **Geospatial Analysis**
➤ **Topical Analysis**
➤ **Tree Analysis**
➤ **Network Analysis**
➤ **Large Network Analysis**
➤ **Scholarly Database**
➤ **VIVO National Researcher Networking**
➤ **Future Developments**

Knowledge Management and Visualization Tools in Support of Discovery NSF Workshop Report

This report summarizes the results of two National Science Foundation (NSF) workshops on "Knowledge Management and Visualization Tools in Support of Discovery" that were inspired by NSF's Cyber-Enabled Discovery and Innovation (CDI) initiative. The report **presents major challenges and opportunities** for the design of more effective tools and CIs in support of scholarly discovery, including a timeline of anticipated science and technology development that will impact tool development; **provide recommendations** on how current lines of work in academia, government, and industry and promising avenues of research and development can be utilized to **design more effective knowledge management, visualization tools and cyberinfrastructures** that advance discovery and innovation in 21st century science.

Börner, Katy, Bettencourt, Luis M. A., Gerstein, Mark & Uzzo, Stephen Miles (Eds.), Knowledge Management and Visualization Tools in Support of Discovery. (2009). NSF CDI Initiative Workshop Report, National Science Foundation, Indiana University.
http://vw.slis.indiana.edu/cdi2008/whitepaper.html

NSF Report – Principal Findings

➢ **Science is interdisciplinary and global.** Researchers and practitioners need easy access to expertise, publications, software, and other resources across scientific and national boundaries.

➢ **Science is data driven.** Access to large amounts of high quality and high-coverage data is mandatory. The "long tail" of data producers/users is larger than the existing major databases and their users.

➢ **Science is computational.** The design of modular, standardized and easy to use cyberinfrastructures is key for addressing major challenges, such as global warming, or a deeper understanding of how science and technology evolves. Ideally, the "million minds" can share, combine, and improve expertise, data, and tools. It is advantageous for scientists to adapt industry standards, defacto or not, than to have to create their own tools.

➢ **Science uses many platforms.** Some sciences thrive on Web services and portals, others prefer desktop tools, while some require virtual reality environments, or mobile (handheld) devices.

➢ **Science is collaborative.** A deeper understanding of how teams "form, storm, norm and perform" will improve our ability to compose (interdisciplinary/international) teams that collaborate effectively. There were also a number of findings specific to the workshop topic "Knowledge Management and Visualization Tools in Support of Discovery":

➢ **Formulas and visual imagery** help communicate results across scientific boundaries with different cultures and languages.

➢ **Advanced data analyses combined with visualizations** are used to identify patterns, trends, clusters, gaps, outliers and anomalies in massive amounts of complex data. Network science approaches seemed particularly useful in the selected biomedical/ecological and SoS domains.

➢ **Scientific domains have different affordances.** For example, intuition and approaches developed in the analysis of scholarly data, which is much more readily available and easier to understand than biomedical/ecological data, could be used to study biomedical/ecological data (which requires a large amount of highly specialized background knowledge).

NSF Report – SoS Principal Recommendations

➤ A decentralized, free **"Scholarly Database"** is needed to keep track, interlink, understand and improve the quality and coverage of Science and Technology (S&T) relevant data.

➤ Science would benefit from a **"Science Marketplace"** that supports the sharing of expertise and resources and is fueled by the currency of science: scholarly reputation. This marketplace might also be used by educators and the learning community to help bring science to the general public and out of the "ivory tower".

➤ A **"Science Observatory"** should be established that analyzes different datasets in real-time to assess the current state of S&T and to provide an outlook for their evolution under several (actionable) scenarios.

➤ **"Validate Science Maps"** to understand and utilize their value for communicating science studies and models across scientific boundaries, but also to study and communicate the longitudinal (1980-today) impact of funding on the science system.

➤ Design an easy to use, yet versatile, **"Science Telescope"** to communicate the structure and evolution of science to researchers, educators, industry, policy makers, and the general public at large. The effect of this (and other science portals) on education and science perception needs to be studied in carefully controlled experiments.

➤ **"Science of (Team) Science"** studies are necessary to increase our understanding and support the formation of effective research and development teams.

➤ **"Success Criteria"** need to be developed that support a scientific calculation of S&T benefits for society for society.

➤ A **"Science Life"** (an analog to Second Life) should be created to put the scientist's face on their science. Portals to this parallel world would be installed in universities, libraries and science museums. The portals would be "fathered and mothered" by domain, as well as learning experts. Their effect on education and science perception should be rigorously evaluated in carefully controlled experiments and improved from a learning science standpoint.

[#12] Future Developments

➤ **Validation Studies**

➤ **Needed Data/Documentation**

➤ **Needed and New Tool Functionality**

➤ **Needed Documentation/Tutorials**

➤ **Promising Research Questions**

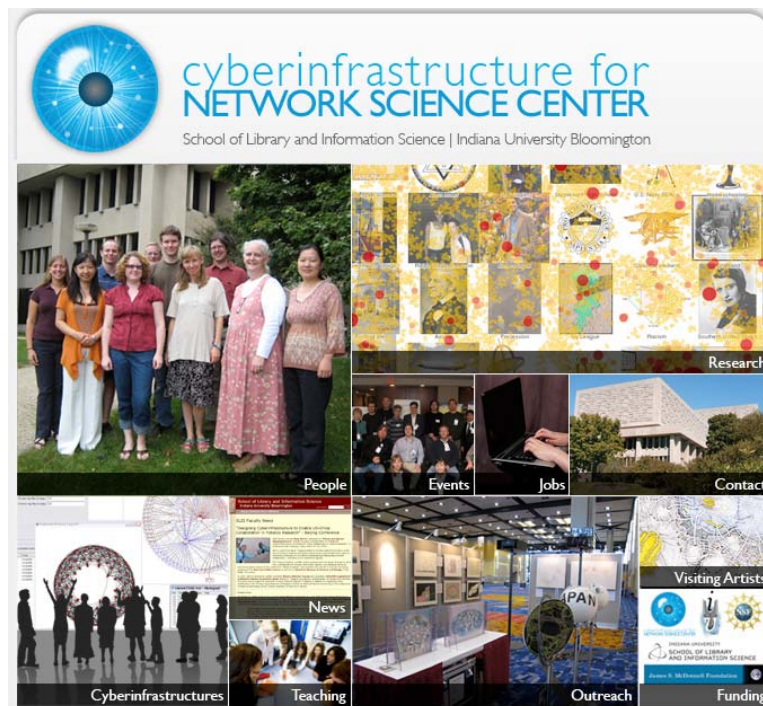➤ **Exercise: Identify Promising Collaborations**

Exercise

Please identify promising collaborations.

Document it by listing
- Project title
- User, i.e., who would be most interested in the result?
- Insight need addressed, i.e., what would you/user like to understand?
- Data used, be as specific as possible.
- Analysis algorithms used.
- Visualization generated. Please make a sketch with legend.

All papers, maps, cyberinfrastructures, talks, press are linked
from http://cns.slis.indiana.edu