Science of Science Research and Tools
Tutorial #06 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
10a-noon, July 14, 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 Tutorials in 12 Days at NIH—Overview

[#06] Topical Analysis & Mapping

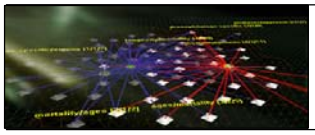
- General Overview
- Designing Effective Topic Maps
- Sci2-Term Co-Occurrence Analysis and Networks
- Sci2-Science Maps With Circle Annotation
- Sci2-Animations
- Outlook
- Exercise: Identify Promising Topical Analyses of NIH Data

Recommended Reading

- NWB Team (2009) Network Workbench Tool, User Manual 1.0.0, http://nwb.slis.indiana.edu/Docs/NWBTool-Manual.pdf
- Scott Weingart, Hanning Guo, Katy Borner, 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. http://sci.slis.indiana.edu/registration/docs/Sci2_Tutorial.pdf

[#06] Topical Analysis & Mapping

- General Overview
- Designing Effective Topic Maps
- Sci2-Term Co-Occurrence Analysis and Networks
- Sci2-Science Maps With Circle Annotation
- Sci2-Animations
- Outlook
- Exercise: Identify Promising Topical Analyses of NIH Data

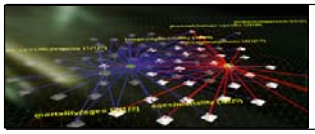


Map Types

Concept maps of conceptual spaces, are a representational tool used in the fields of education and psychology. Concept maps have also been referred to as mind maps, pattern notes, brain patterns, spider maps, networks, semantic maps, semantic networks, and semantic webs

Domain maps of abstract semantic spaces aim to serve today's explorers navigating the world of science. These maps are generated through a scientific analysis of large-scale scholarly datasets in an effort to connect and make sense of the bits and pieces of knowledge they contain.

Cartographic maps of physical places have guided mankind's explorations for centuries. They enabled the discovery of new worlds while also marking territories inhabited by unknown monsters. Without maps, we would be lost. **See <u>Tutorial #5</u>**

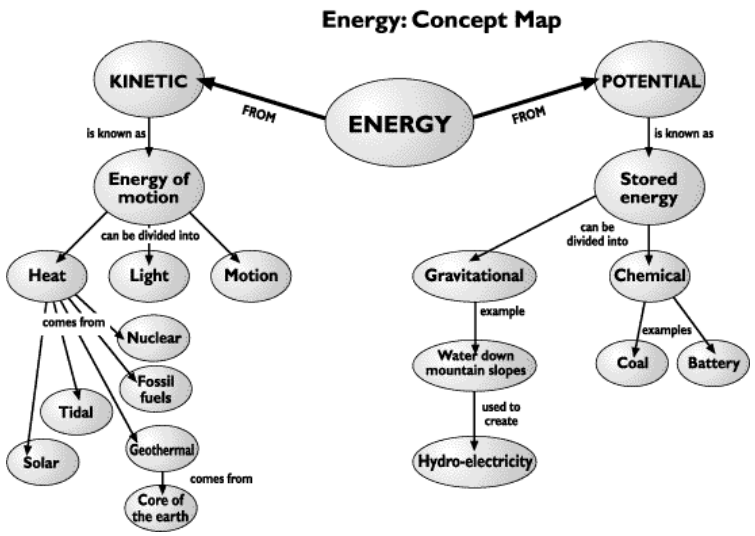


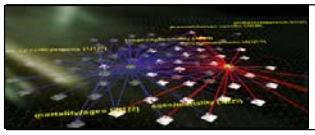
Concept Maps

There are at least nine different categories of connecting words between nodes:

- Subsuming
- Similarity
- Quantity
- Enabling
- Causal
- Timing
- Dissimilarity
- equivalence, and
- categorizing.

(West et al, 1991, p 101).

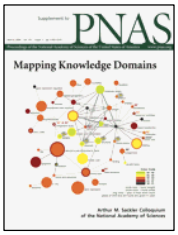




Domain Maps

Domain maps help answer questions such as:

- What are the major research areas, experts, institutions, regions, nations, grants, publications, journals in xx research?
- Which areas are most insular?
- What are the main connections for each area?
- What is the relative speed of areas?
- Which areas are the most dynamic/static?
- What new research areas are evolving?
- Impact of xx research on other fields?
- How does funding influence the number and quality of publications?

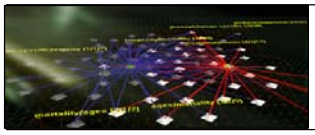


Answers are needed by funding agencies, companies, and researchers.

See also http://scimaps.org and **Tutorial #1.**

[#06] Topical Analysis & Mapping

- General Overview
- Designing Effective Topic Maps
- Sci2-Term Co-Occurrence Analysis and Networks
- Sci2-Science Maps With Circle Annotation
- Sci2-Animations
- Outlook
- Exercise: Identify Promising Topical Analyses of NIH Data



Process of Analyzing and Mapping Knowledge Domains

DATA EXTRACTION	UNIT OF ANALYSIS	MEASURES	LAYOUT (often one code does both similarity and ordination steps)		DISPLAY
			SIMILARITY	ORDINATION	
SEARCHES ISI INSPEC Eng Index Medline ResearchIndex Patents etc. BROADENING By citation By terms	COMMON CHOICES Journal Document Author Term	COUNTS/FREQUENCIES Attributes (e.g. terms) Author citations Co-citations By year THRESHOLDS By counts	SCALAR (unit by unit matrix) Direct citation Co-citation Combined linkage Co-word / co-term Co-classification VECTOR (unit by attribute matrix) Vector space model (words/terms) Latent Semantic Analysis (words/terms) incl. Singular Value Decomp (SVD) CORRELATION (if desired) Pearson's R on any of above	DIMENSIONALITY REDUCTION Eigenvector/ Eigenvalue solutions Factor Analysis (FA) and Principal Components Analysis (PCA) Multi-dimensional scaling (MDS) LSA, **Topics** Pathfinder networks (PFNet) Self-organizing maps (SOM) includes SOM, ET-maps, etc. CLUSTER ANALYSIS SCALAR Triangulation Force-directed placement (FDP)	INTERACTION Browse Pan Zoom Filter Query Detail on demand ANALYSIS

Börner, Katy, Chen, Chaomei, and Boyack, Kevin. (2003) Visualizing Knowledge Domains. In Blaise Cronin (Ed.), Annual Review of Information Science & Technology, Volume 37, Medford, NJ: Information Today, Inc./American Society for Information Science and Technology, chapter 5, pp. 179-255.

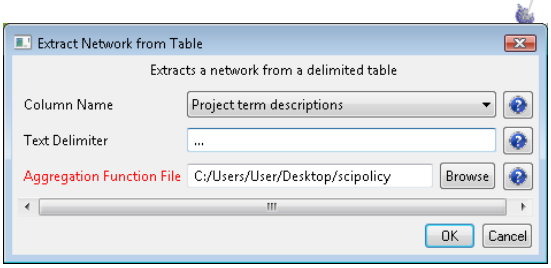
NIH Datasets

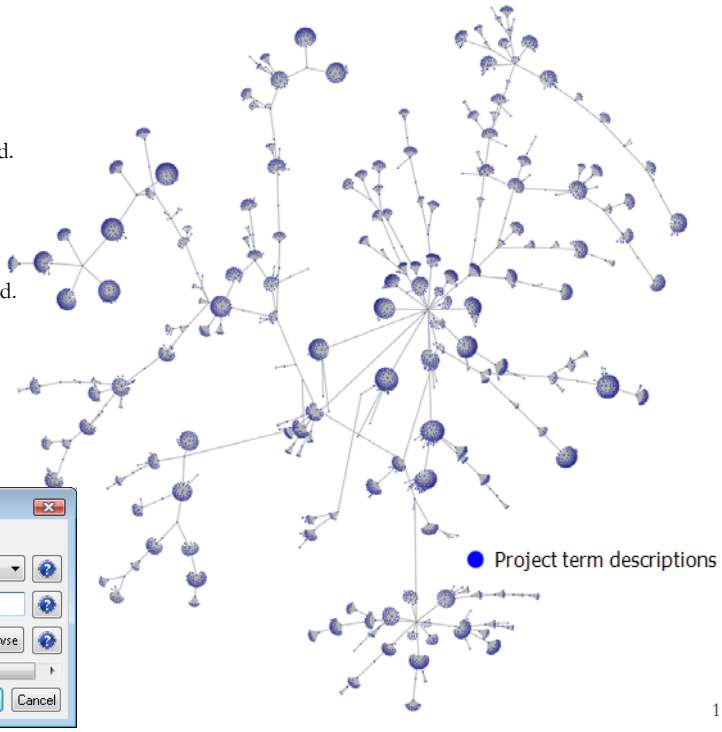
Using NIH RePORTER

- NIH CTSA Funding (534 records, $1,210,288,444 total 'FY Total Cost', Sept. 2006-June 2011) and linked Publications (2,456 records)
- NIH NIGMS PPBC R01s Funding (935 records, $280,825,758 total 'FY Total Cost', Jan. 1976-Aug. 2013) and linked Publications (18,448 records)
- NIH MIDAS Funding (54 records, 35,477,829 total 'FY Total Cost', May 2004-Jan. 2011) and linked Publications (50 records, http://www.nigms.nih.gov/Initiatives/MIDAS/Publications.htm lists 69 on 10/19/2009)

NIH CTSA Grants:
Co-Project Term Descriptions Occurrence Network

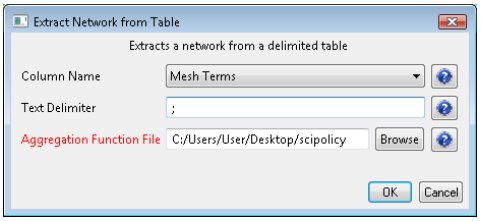
Load... was selected.
Loaded: …\NIH-CTSA-Grants.csv
.....
Extract Co-Occurrence Network was selected.
Input Parameters:
Text Delimiter: ...
Column Name: Project term descriptions
.....
Network Analysis Toolkit (NAT) was selected.
Nodes: 5722
Isolated nodes: 2
Edges: 353218

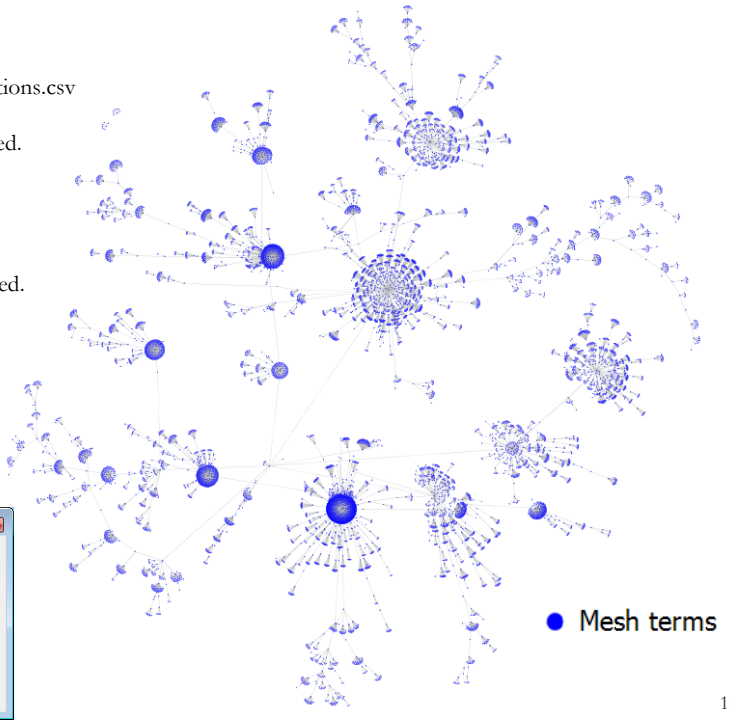




NIH CTSA Publications:
Co-Mesh Terms Occurrence Network

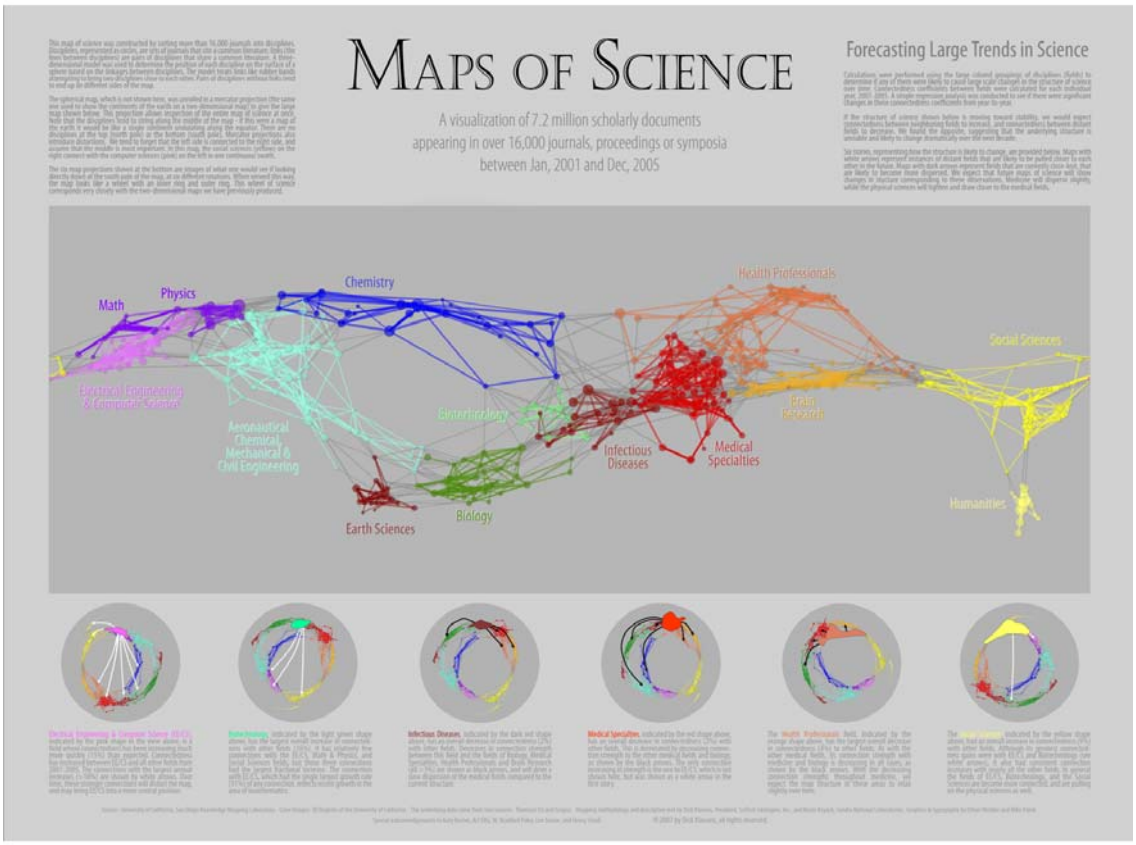
Load... was selected.
Loaded: …\NIH-data\NIH-CTSA-Publications.csv
.....
Extract Co-Occurrence Network was selected.
Input Parameters:
Text Delimiter: ;
Column Name: Mesh Terms
.....
Network Analysis Toolkit (NAT) was selected.
Nodes: 10218
Edges: 163934





[#06] Topical Analysis & Mapping

- General Overview
- Designing Effective Topic Maps
- Sci2-Term Co-Occurrence Analysis and Networks
- Sci2-Science Maps With Circle Annotation
- Sci2-Animations
- Outlook
- Exercise: Identify Promising Topical Analyses of NIH Data



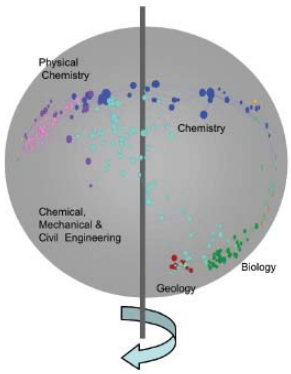
Maps of Science: Forecasting Large Trends in Science - Richard Klavans, Kevin Boyack - 2007

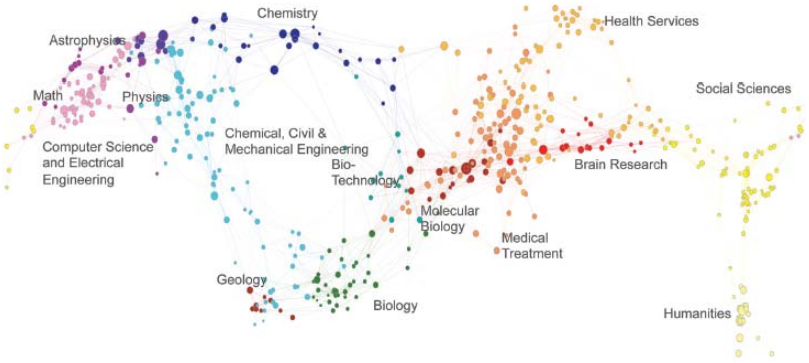
UCSD Map of Science

The UCSD Map of Science was generated using 7.2 million papers published in over 16,000 separate journals, proceedings, and series from Thomson Scientific and Scopus over the five year period from 2001 to 2005.

A combination of bibliographic coupling and keyword vectors was used to group papers and journals into 554 journal clusters.

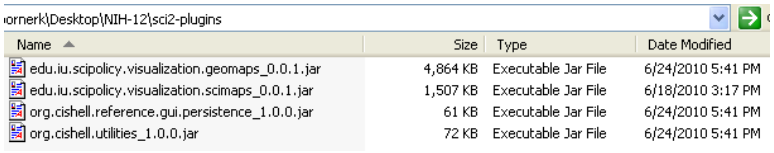
Each cluster is labeled both by the content area shared by the journals in the cluster and by the overarching scientific domain for that cluster (represented by one of 13 colors).





CIShell – Add new Plugins, e.g., UCSD Science Map

To make the UCSD Science Map and new geomaps available via the Sci² menu, simply add



to the '*yourdirectory/plugin*' directory and restart the tool.

The files were made available in /sci2-plugins directory on the computers in the tutorial room.

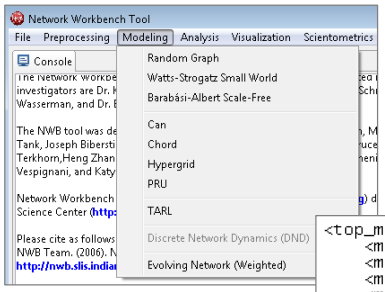
The rights to the UCSD map are owned by the Regents of UCSD. Usage does not require a separate, signed agreement or an additional request to our office if consistent with the permission. As a courtesy, please send information on how the map is being used to

William J. Decker, Ph.D., Associate Director, Technology Transfer Office
University of California, San Diego, 9500 Gilman Drive Dept. 0910, La Jolla, CA 92093
phone:858-822-5128, fax: 858-534-7345, e-mail: wjdecker@ucsd.edu

To delete algorithms that you do not use, simply delete the corresponding *.jar files in the plugin directory.

CIShell – Customize Menu

- The file *'yourtooldirectory/configuration/default_menu.xml'* encodes the structure of the menu system.
- In NWB Tool, the Modeling menu (left) is encoded by the following piece of xml code:

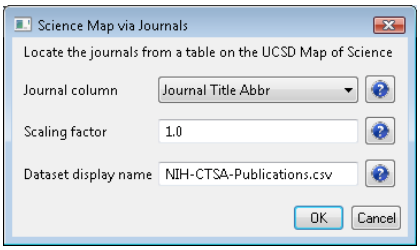


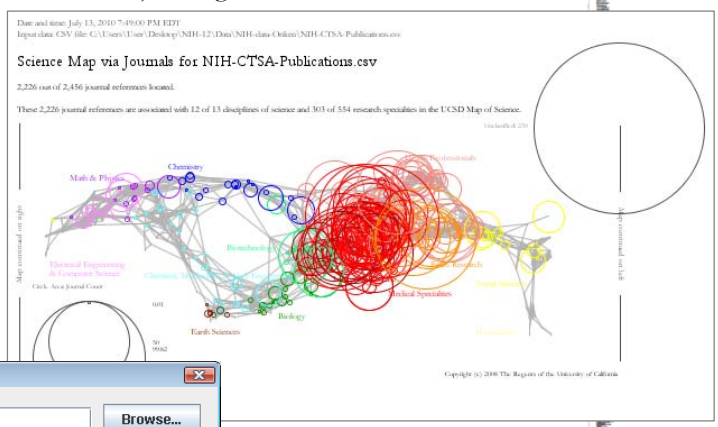

```
<top_menu name="Modeling">
    <menu pid="edu.iu.nwb.modeling.erdosrandomgraph"/>
    <menu pid="edu.iu.nwb.modeling.smallworld"/>
    <menu pid="edu.iu.nwb.modeling.barabasialbert"/>
    <menu type="break"/>
    <menu pid="edu.iu.iv.modeling.p2p.can.CanAlgorithm"/>
    <menu pid="edu.iu.iv.modeling.p2p.chord.ChordAlgorithm"/>
    <menu pid="edu.id.iv.modeling.p2p.hypergrid.Hypergrid"/>
    <menu pid="edu.iu.iv.modeling.p2p.pru.PruAlgorithm"/>
    <menu type="break"/>
    <menu pid="edu.iu.iv.modeling.tarl.TarlAlgorithm"/>
    <menu type="break"/>
    <menu pid="edu.iu.nwb.modeling.discretenetworkdynamics.DNDAlgorithm"/>
    <menu type="break"/>
    <menu pid="edu.iu.nwb.modeling.weighted.evolvingnetwork"/>
</top_menu>
```


NIH CTSA Grants: Topic Coverage of Publications

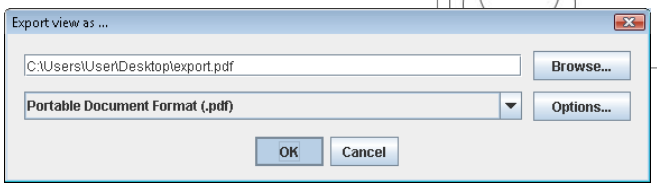
Load NIH-CTSA-Publications.csv

Run *'Visualization > Science Map (Circle Annotation)'* using





To see the science map overlay on right.



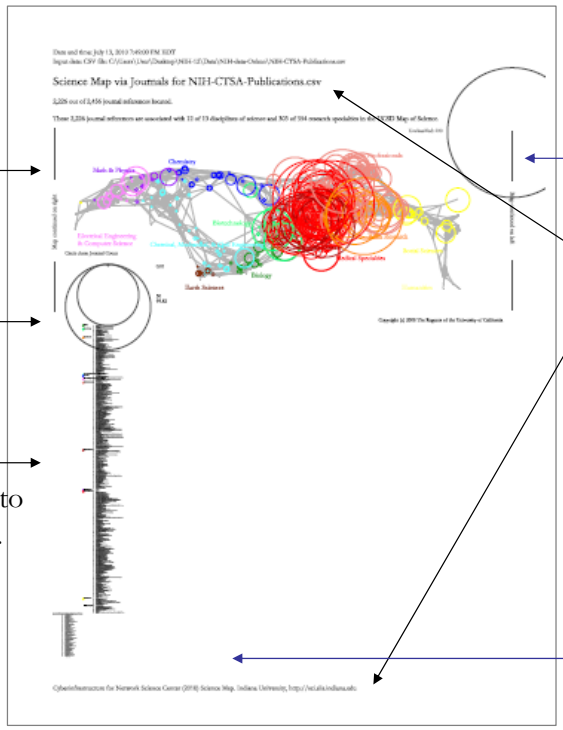
The 'Scaling factor' can be used to increase or decrease the size of the maximum circle.

How to Read the UCSD Map

UCSD Science Map with data overlay.

Map legend of circle area size coding

Listing of all data records organized into UCSD science areas.



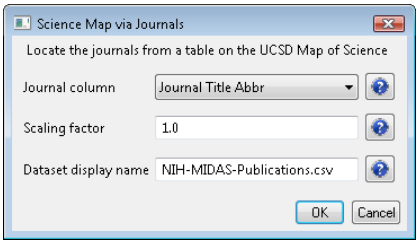
Circle of non-located, e.g., 'Unclassified' records.

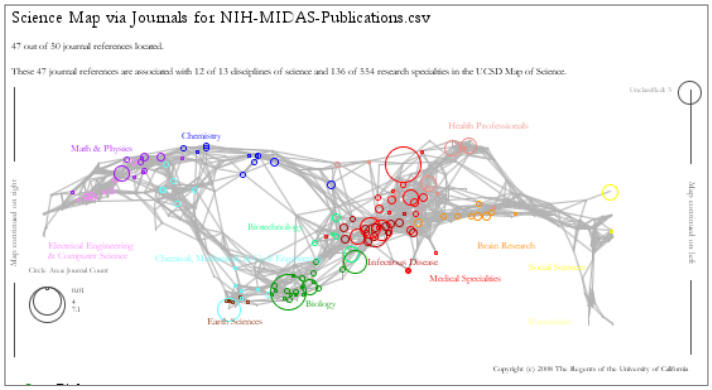
Header and footer with information when this map was created, by whom and using what data set.

Listing and circle of non-located, e.g., 'Unclassified' records.

NIH MIDAS Funding: Topic Coverage of Publications

Load NIH-MIDAS-Publications.csv via '*Visualization* > *Science Map (Circle Annotation)*' and



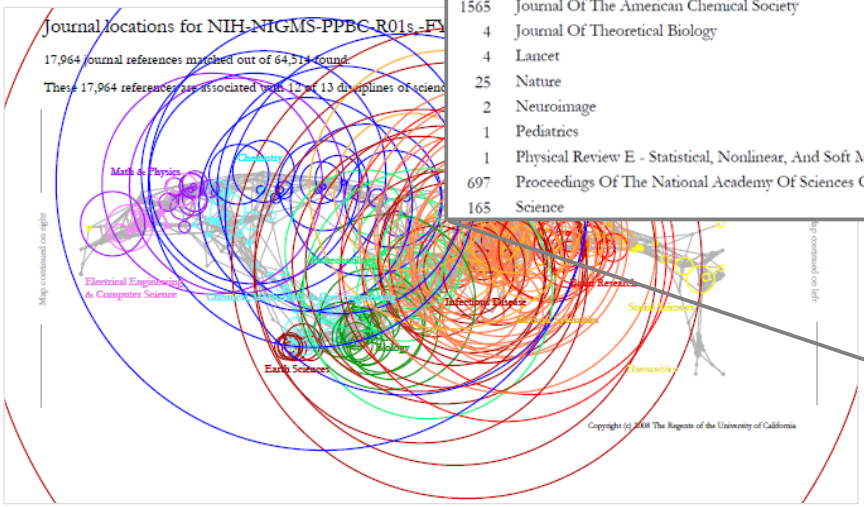


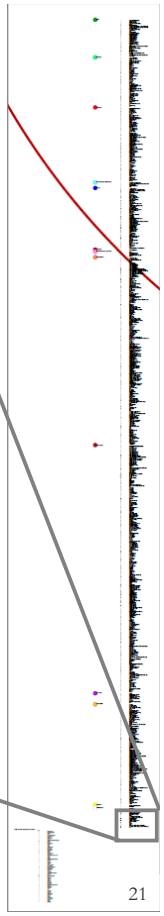
NIH NIGMS Funding: Topic Coverage of Publications

Science locate all 18,448 MIDAS papers

Load... was selected.
Loaded: C:\...\NIH-NIGMS-PPBC-R01s,-FY08-Publications.csv
.....
Science Map (Circle Annotation) was selected.
Input Parameters:
Journal column: Journal Title Abbr
Saved: NIH-NIGMS-PPBC-R01s,-FY08-Publications.ps

1	Jama : The Journal Of The American Medical Association
2205	Journal Of Biological Chemistry
4	Journal Of Chromatography A
1565	Journal Of The American Chemical Society
4	Journal Of Theoretical Biology
4	Lancet
25	Nature
2	Neuroimage
1	Pediatrics
1	Physical Review E - Statistical, Nonlinear, And Soft Matter Physics
697	Proceedings Of The National Academy Of Sciences Of The United
165	Science



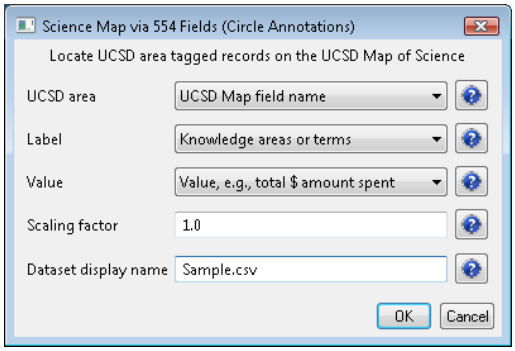


Aligning Topic Terms to the UCSD Map of Science

It is possible to align the 554 fields of science in the UCSD map and "Knowledge Areas or Terms" used internally at an agency.

Knowledge areas or terms	Value, e.g., total $ amount spent	UCSD Map field name
Brain	47405	14
Cancer	677865	15

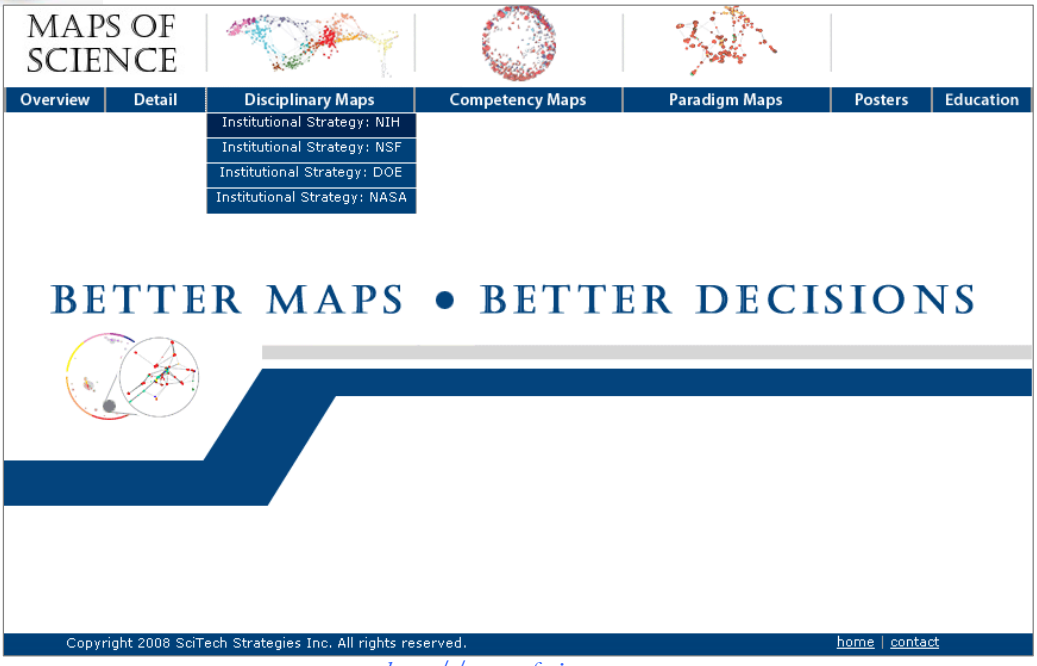
Load data as csv file then run *Visualization* > *Topical* > *Science Map via 554 Fields* using parameters:



[#06] Topical Analysis & Mapping

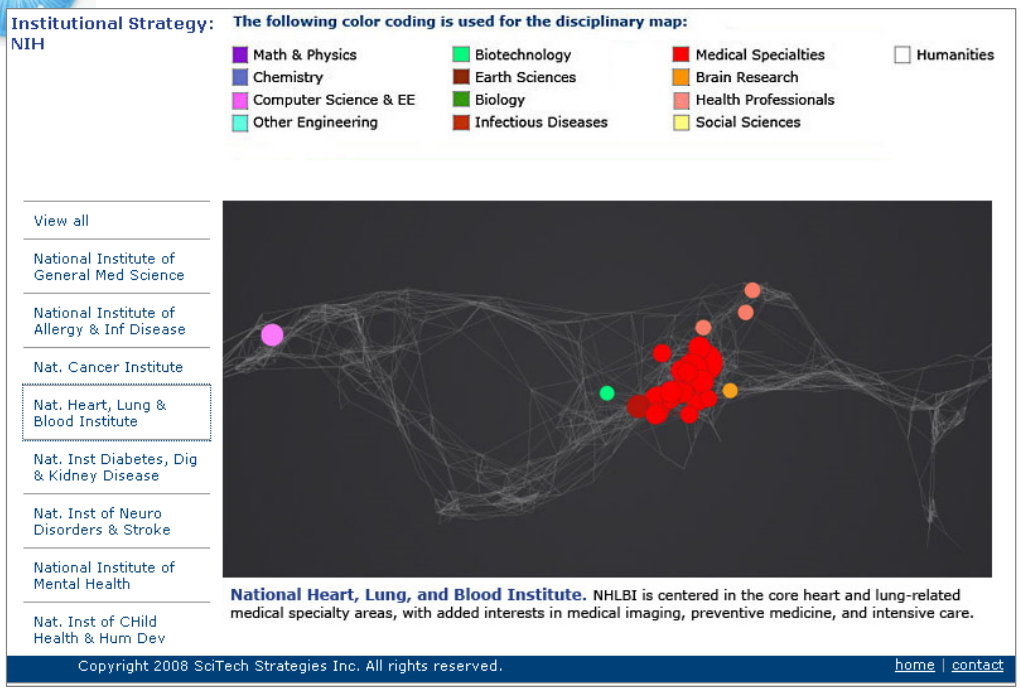
- General Overview
- Designing Effective Topic Maps
- Sci2-Term Co-Occurrence Analysis and Networks
- Sci2-Science Maps With Circle Annotation
- Sci2-Animations – Interactive Maps of Science
- Outlook
- Exercise: Identify Promising Topical Analyses of NIH Data

Comparing the Portfolios of Different Funding Organizations



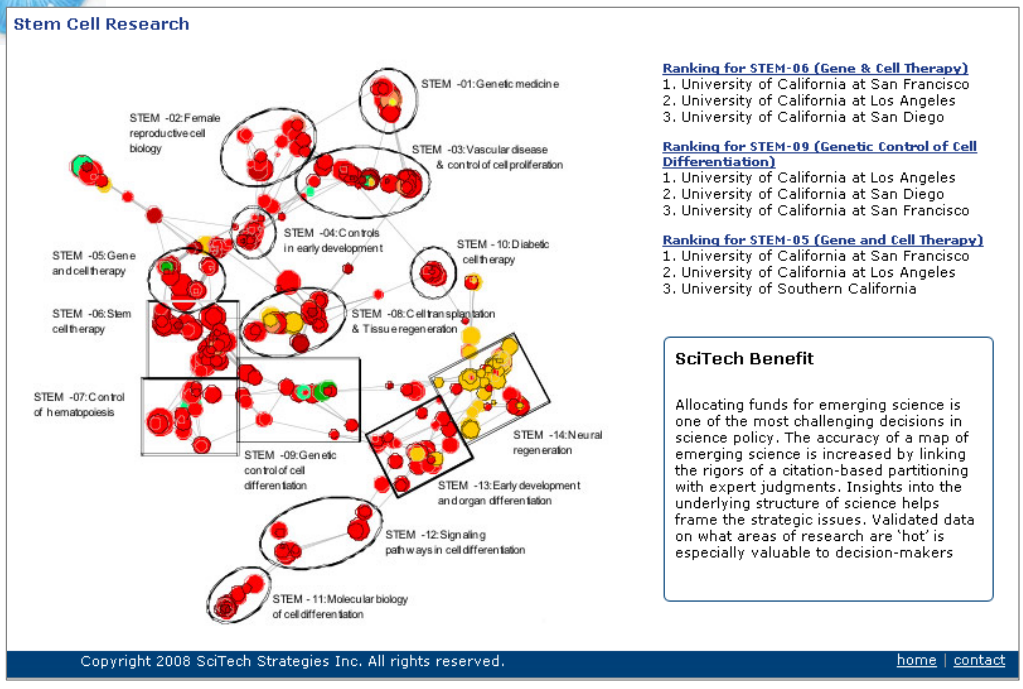
http://mapofscience.com

Comparing the Portfolios of Different NIH



http://mapofscience.com

Comparing Stem Cell Research Portfolios



http://mapofscience.com

Topic Map of NIH Funding by Ned Talley et al

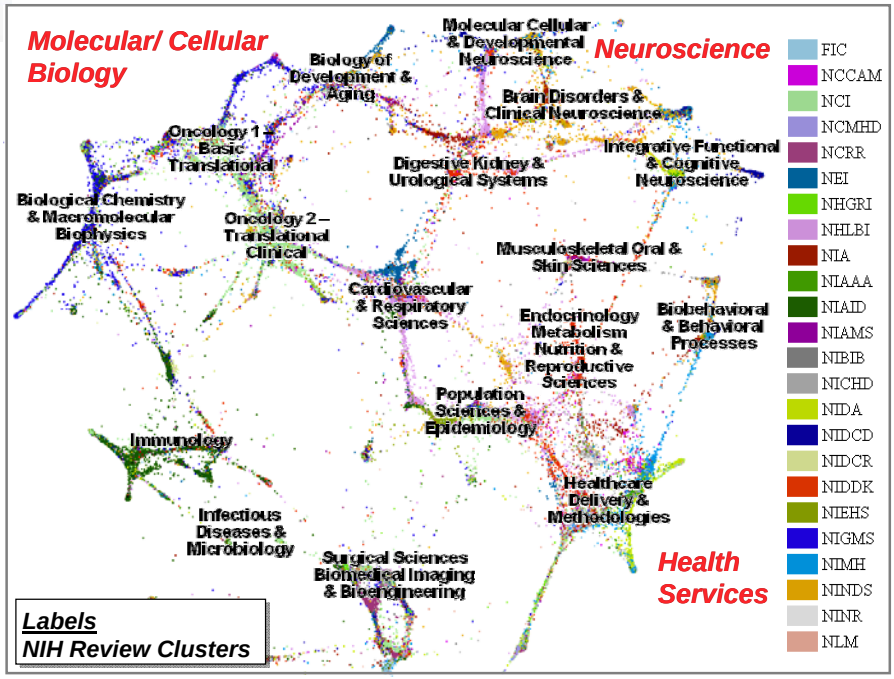
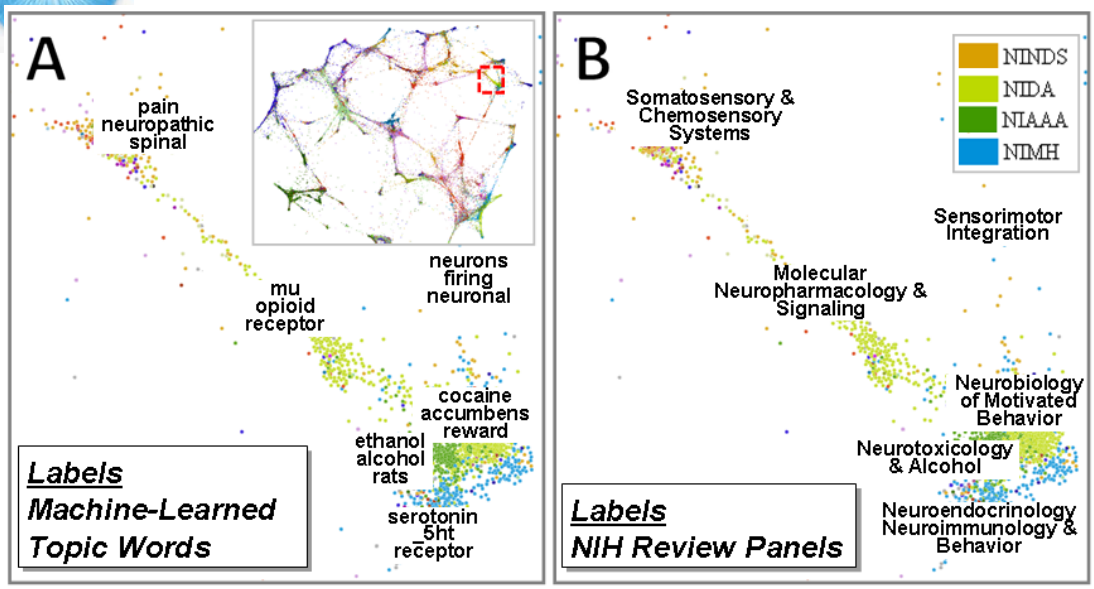


Figure 1. Graphical layout of NIH grants provides an organizational framework with compelling global structure.

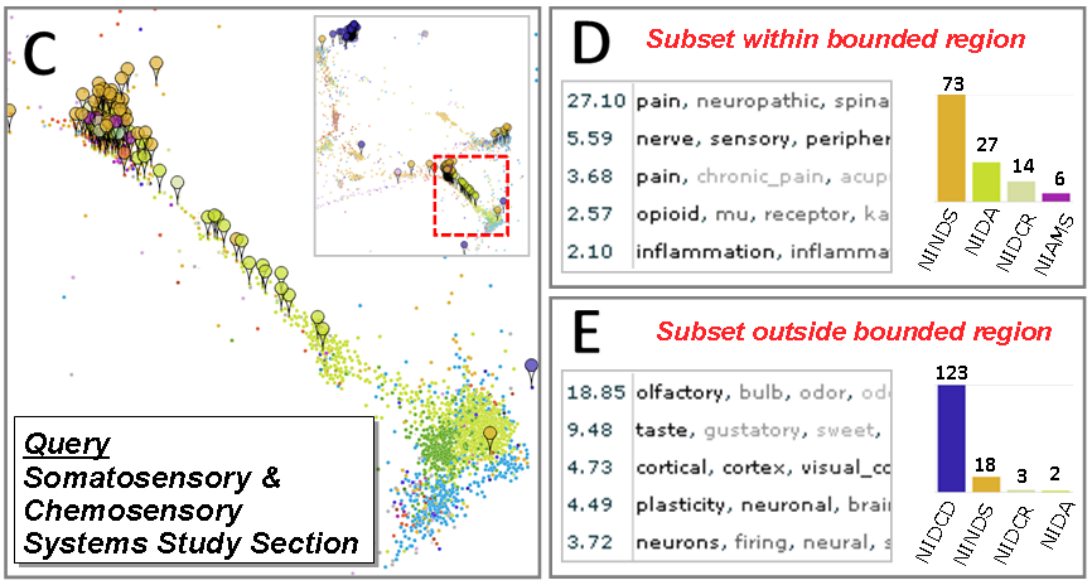
Comparing the Portfolios of Different Funding Organizations



Relationships between machine-learned categories and NIH Institute/Review organization.

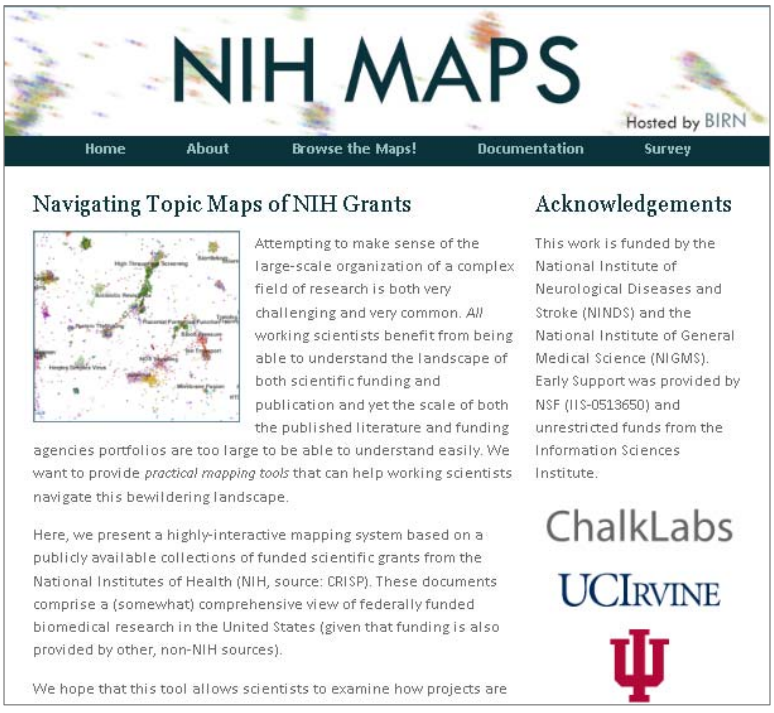
A and B. Color coding from a selected region of the graphical layout (indicated by the red-dashed box in the inset) shows the Institute organization of different research categories. Labels in panel A are derived from the title words of the highest represented topic in the underlying document clusters. Panel B shows the same view but with labels from the highest represented NIH standing review panels (five of the six panels are from the "Integrative, Functional, and Cognitive Neuroscience" Integrated Review Group, http://cms.csr.nih.gov/PeerReviewMeetings/CSRIRGDescriptionNew/IFCNIRG/).

Comparing the Portfolios of Different Funding Organizations



Relationships between machine-learned categories and NIH Institute/Review organization.
C. Query of the database for the review panel represented in B, the Somatosensory & Chemosensory Systems Study Section, http://cms.csr.nih.gov/peerreviewmeetings/csrirgdescriptionnew/ifcnirg/scs.htm). Retrieved awards are represented as inverted teardrop-shaped markers. Panel D indicates the topic and Institute representation of awards shown in panel C. Panel E shows the topic and Institute representation of all other awards reviewed by the study section (i.e., outside the red-dashed box). The top five topics for each subset are represented by the title words from each topic (shown in descending order, with font grayscale proportional to word probability).

Topic Map of NIH Funding by Ned Talley et al



http://www.nihmaps.org

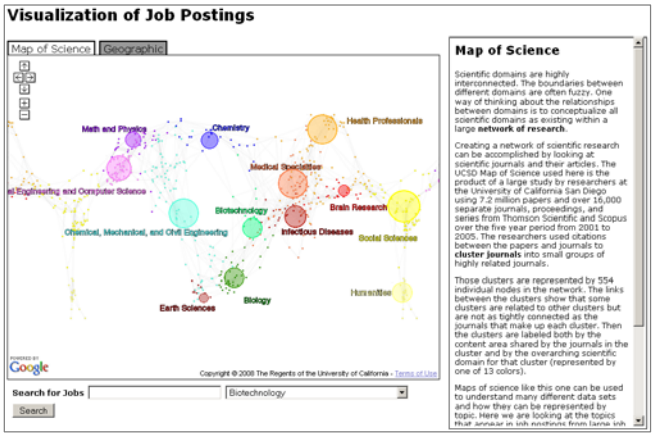
[#06] Topical Analysis & Mapping

- General Overview
- Designing Effective Topic Maps
- Sci2-Term Co-Occurrence Analysis and Networks
- Sci2-Science Maps With Circle Annotation
- Sci2-Animations
- Outlook
- Exercise: Identify Promising Topical Analyses of NIH Data

Outlook

Planned extensions of Sci2 Tool:

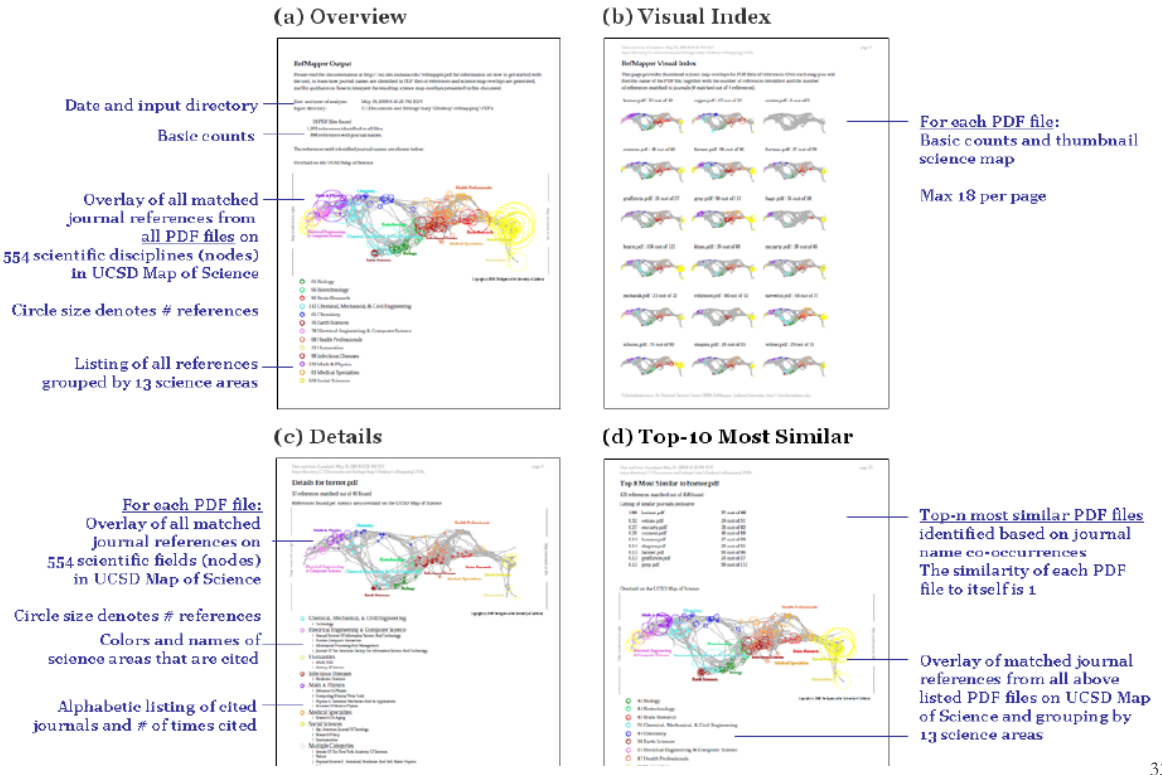
- (Flowmap) network overlays for science maps.
- Easy means to render maps online—upload multiple datasets to compare.



Interactive World and Science Map of S&T Jobs. Angela Zoss, Michael Connover, Katy Börner (2010).
http://cns-nd3.slis.indiana.edu/mapjobs/geo/scivis.html?limit=5000

Reference Mapper

Duhon & Börner, forthcoming.



[#06] Topical Analysis & Mapping

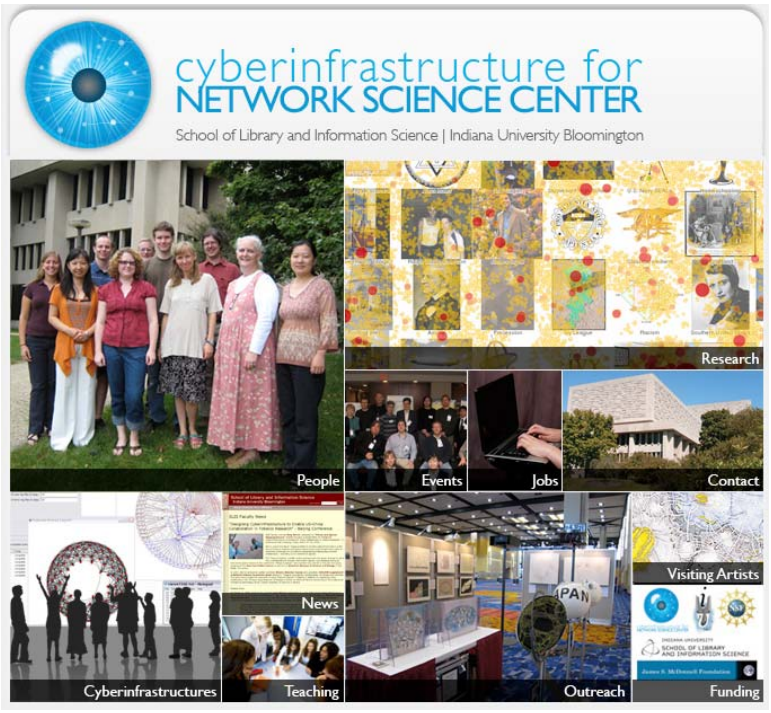
- General Overview
- Designing Effective Topic Maps
- Sci2-Term Co-Occurrence Analysis and Networks
- Sci2-Science Maps With Circle Annotation
- Sci2-Animations
- Outlook
- Exercise: Identify Promising Topical Analyses of NIH Data

Exercise

Please identify a promising topical analysis of NIH data.

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