

# 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

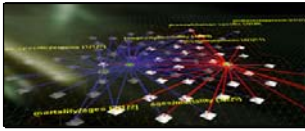
- NWB Team (2009) Network Workbench Tool, User Manual 1.0.0, <http://nwb.slis.indiana.edu/Docs/NWBTool-Manual.pdf>
- Scott Weingart, Hanning Guo, Katy Borner, Kevin W. Boyack, Micah W. Linnemeier, Russell J. Duhon, Patrick A. Phillips, Chintan Tank, and Joseph Biberstine (2010) [Science of Science \(Sci2\) 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](http://sci.slis.indiana.edu/registration/docs/Sci2_Tutorial.pdf)

3

### [#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

4



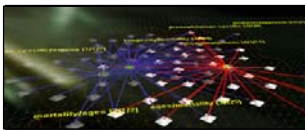
## Map Types

**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 Tutorial #5**

5

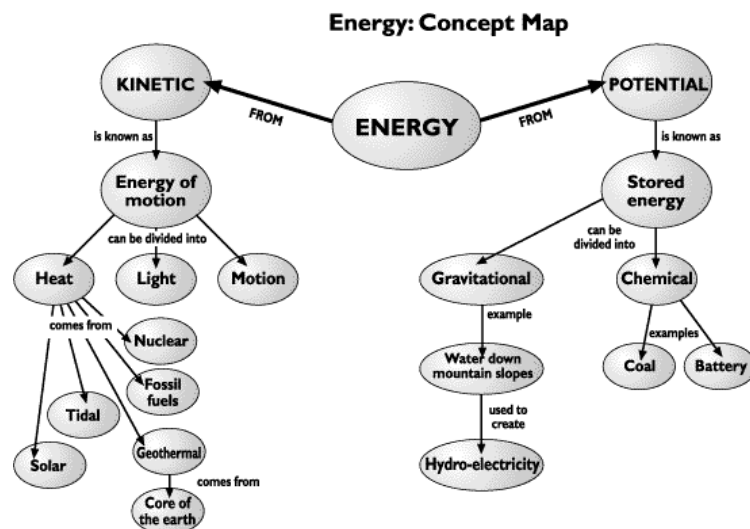


## Concept Maps

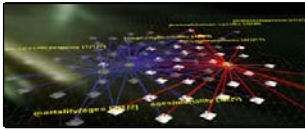
There are at least nine different categories of connecting words between nodes:

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

*(West et al, 1991, p 101).*



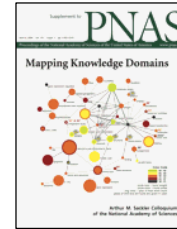
6



## 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.

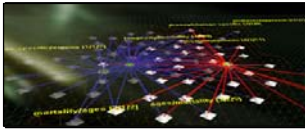
See also <http://scimaps.org> and **Tutorial #1**.

7

## [#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

8



## 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.	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, <b>Topics</b> Pathfinder networks (PFNet) Self-organizing maps (SOM) includes SOM, ET-maps, etc.	INTERACTION Browse Pan Zoom Filter Query Detail on demand  ANALYSIS
BROADENING By citation By terms				CLUSTER ANALYSIS  SCALAR Triangulation Force-directed placement (FDP)	

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.

9



### NIH Datasets

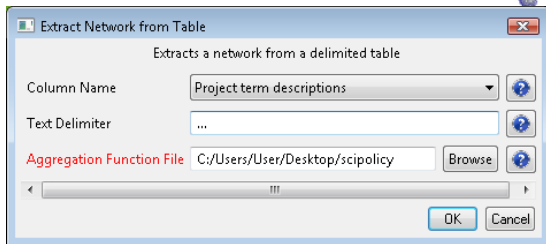
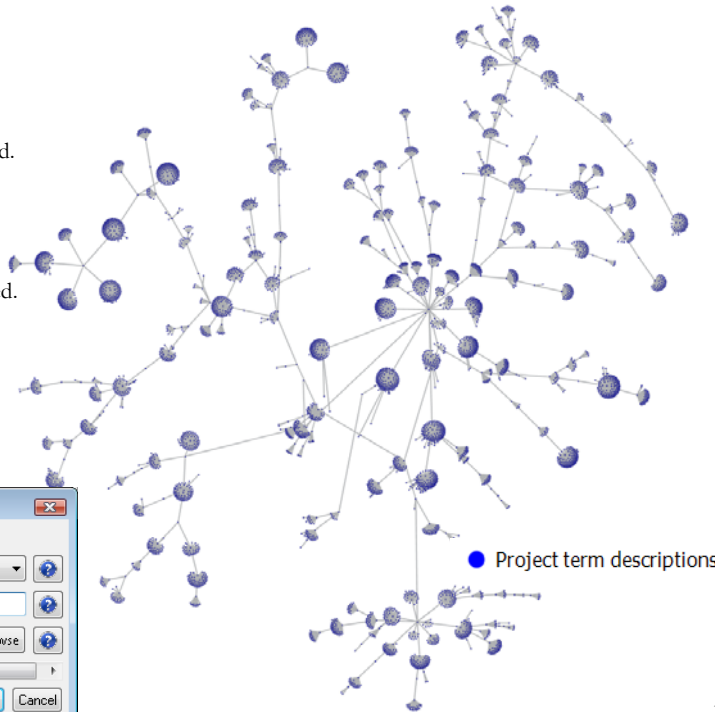
#### Using NIH RePORTER

- 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

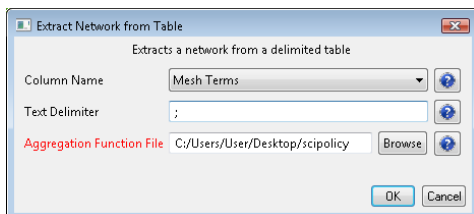
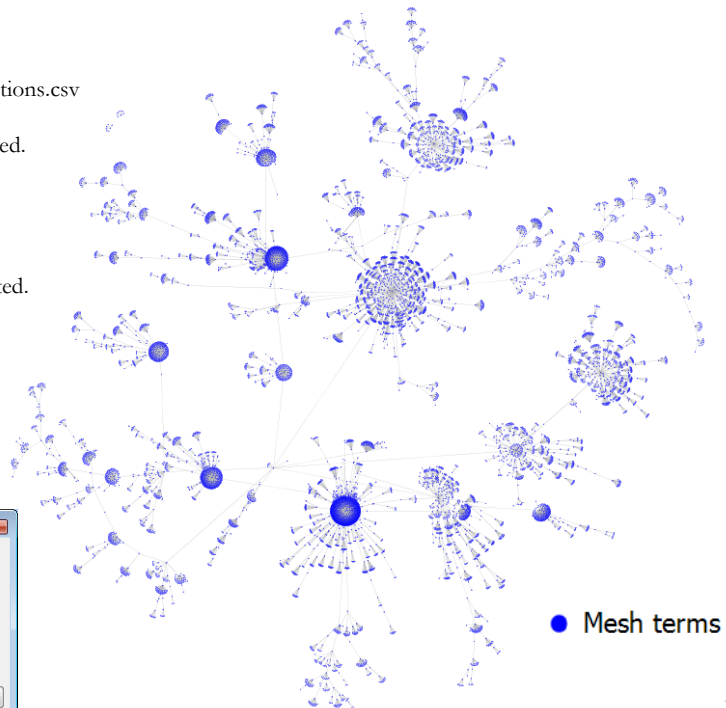


11



## NIH CTSA Publications: Co-Mesh Terms Occurrence Network

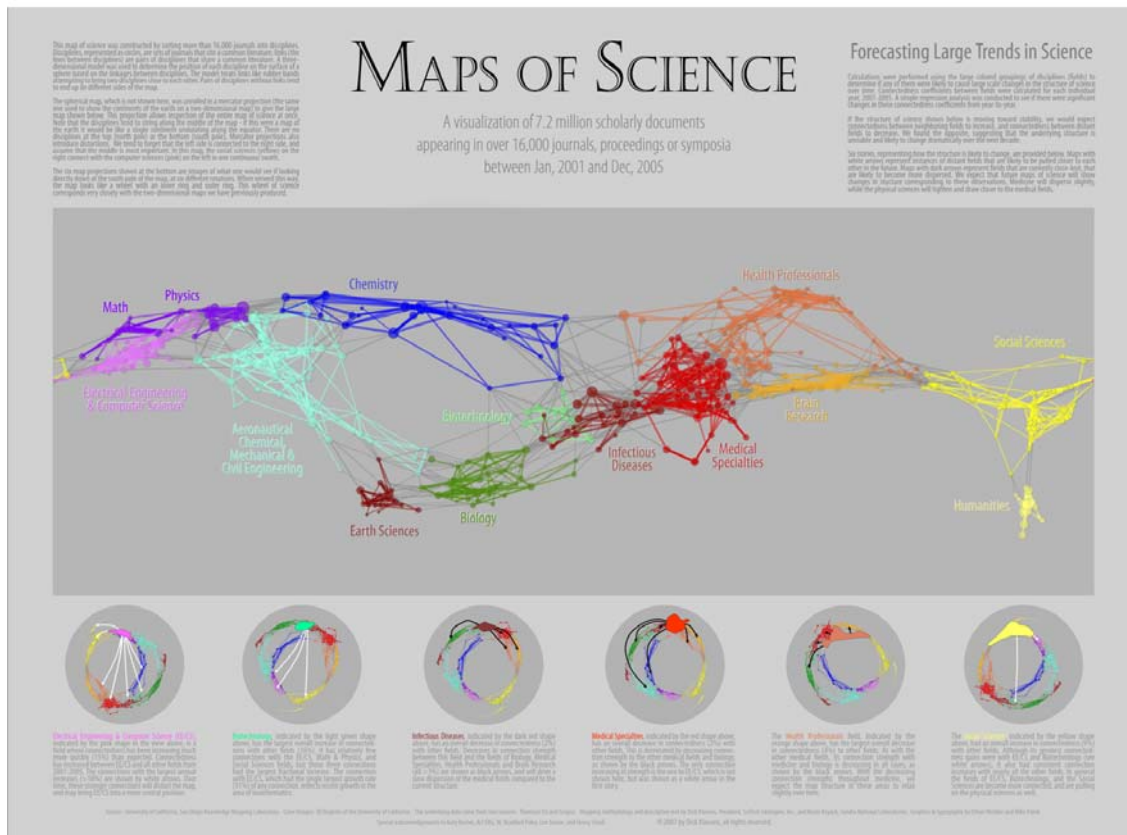
Load... was selected.  
Loaded: ...\\NIH-data\\NIH-CTSA-Publications.csv  
.....  
Extract Co-Occurrence Network was selected.  
Input Parameters:  
Text Delimiter: ;  
Column Name: Mesh Terms  
.....  
Network Analysis Toolkit (NAT) was selected.  
Nodes: 10218  
Edges: 163934



12

# [#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

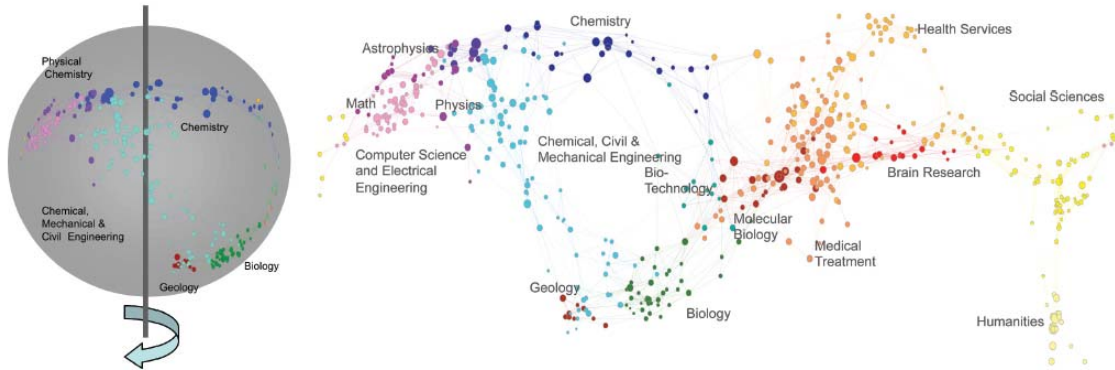


## UCSD Map of Science

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).



15



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

To make the UCSD Science Map and new geomaps available via the Sci<sup>2</sup> menu, simply add

Name	Size	Type	Date Modified
edu.iu.scipolicy.visualization.geomaps_0.0.1.jar	4,864 KB	Executable Jar File	6/24/2010 5:41 PM
edu.iu.scipolicy.visualization.scimaps_0.0.1.jar	1,507 KB	Executable Jar File	6/18/2010 3:17 PM
org.cishell.reference.gui.persistence_1.0.0.jar	61 KB	Executable Jar File	6/24/2010 5:41 PM
org.cishell.utilities_1.0.0.jar	72 KB	Executable Jar File	6/24/2010 5:41 PM

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

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

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: [wjdecke@ucsd.edu](mailto:wjdecke@ucsd.edu)

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

16



- 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:

The screenshot shows the Network Workbench Tool interface with the 'Modeling' menu open. The menu items listed are: Random Graph, Watts-Strogatz Small World, Barabási-Albert Scale-Free, Can, Chord, Hypergrid, PRU, TARL, Discrete Network Dynamics (DND), and Evolving Network (Weighted). To the right, the XML code for this menu is displayed:

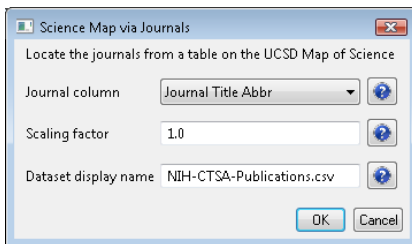
```
<top_menu name="Modeling">
<menu pid="edu.iu.nwb.modeling.erdosrandomgraph"/>
<menu pid="edu.iu.nwb.modeling.smallworld"/>
<menu pid="edu.iu.nwb.modeling.barabasiabert"/>
<menu type="break"/>
<menu pid="edu.iu.iv.modeling.p2p.can.CanAlgorithm"/>
<menu pid="edu.iu.iv.modeling.p2p.chord.ChordAlgorithm"/>
<menu pid="edu.iu.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>
```



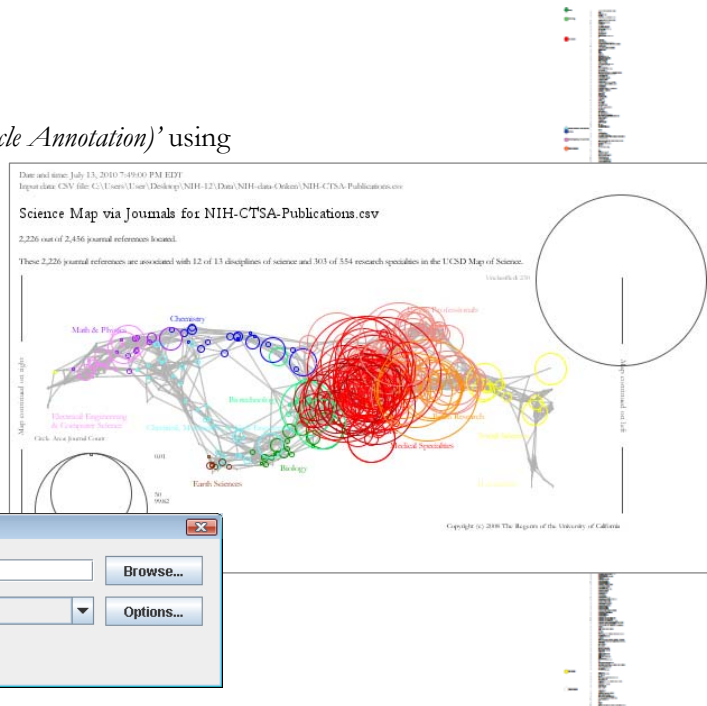
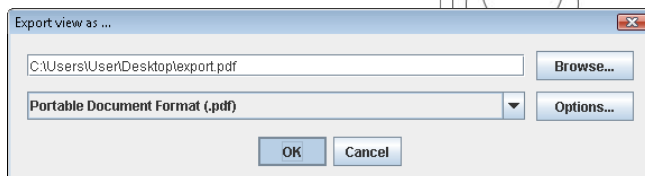
## NIH CTSA Grants: Topic Coverage of Publications

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.

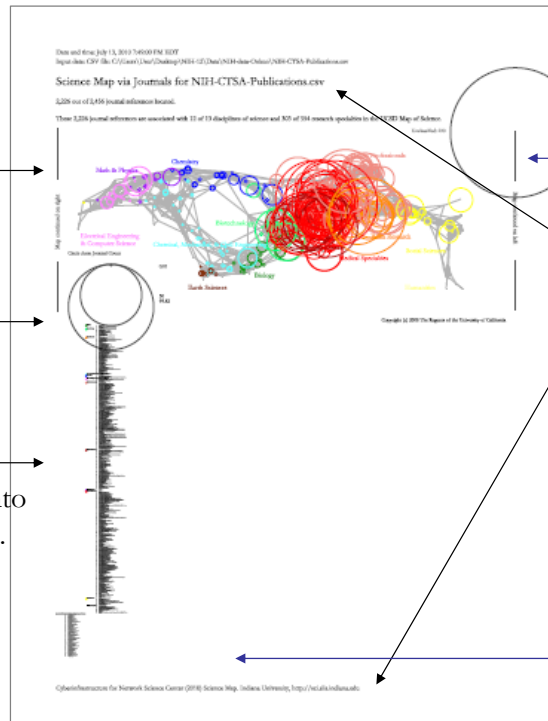


## How to Read the UCSD Map

UCSD Science Map with data overlay.

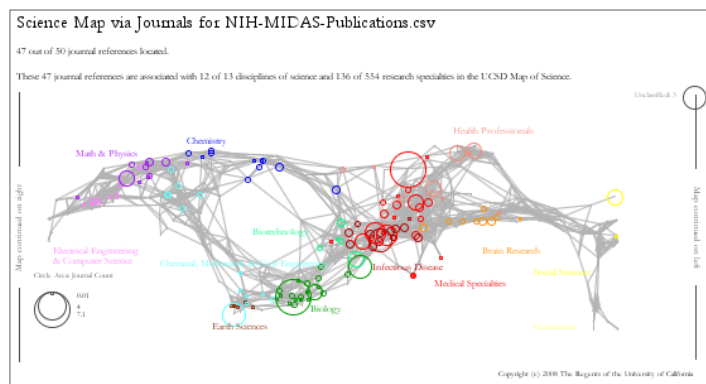
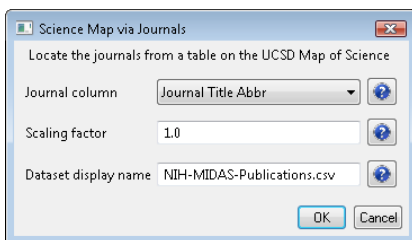
Map legend of circle area size coding

Listing of all data records organized into UCSD science areas.



## NIH MIDAS Funding: Topic Coverage of Publications

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



- Biology**
    - 1 Bio: Evolutionary Biology
    - 1 Journal Of Evolutionary Biology
    - 1 Mathematical Biosciences
    - 3 Plos Medicine
    - 3 Proceedings Of The Royal Society - Biological Sciences (Series B)
  - Biotechnology**
    - 1 Plos Computational Biology
  - Chemical, Mechanical, & Civil Engineering**
    - 1 International Journal Of Health Geographics
  - Health Professionals**
    - 1 Bio: Medical Informatics And Decision Making
    - 1 Journal Of The Royal Society, Interface (Biometric Research) / The Royal S...
    - 1 Maryland Medicine (Man): A Publication Of Medicine, The Maryland State Medic...
  - Infectious Diseases**
    - 3 Vaccine
  - Math & Physics**
    - 1 Mathematical And Computer Modelling
  - Medical Specialties**
    - 6 American Journal Of Epidemiology
    - 2 Emerging Infectious Diseases
    - 1 Epidemiology
    - 1 Epidemiology And Infection
    - 1 International Journal Of Infectious Diseases
    - 1 Lancet Infectious Diseases
    - 1 New England Journal Of Medicine
  - Social Sciences**
    - 1 Biostatistics
  - Multiple Categories**
    - 1 Journal Of Theoretical Biology
    - 2 Physical Review E: Statistical, Nonlinear, And Soft-Matter Physics
    - 5 Proceedings Of The National Academy Of Sciences Of The United States Of Ame...
    - 1 Science
- Journals which could not be located on the map of science:**  
 1 Health Care



## NIH NIGMS Funding: Topic Coverage of Publications

Science locate all 18,448 MIDAS papers

Load... was selected.

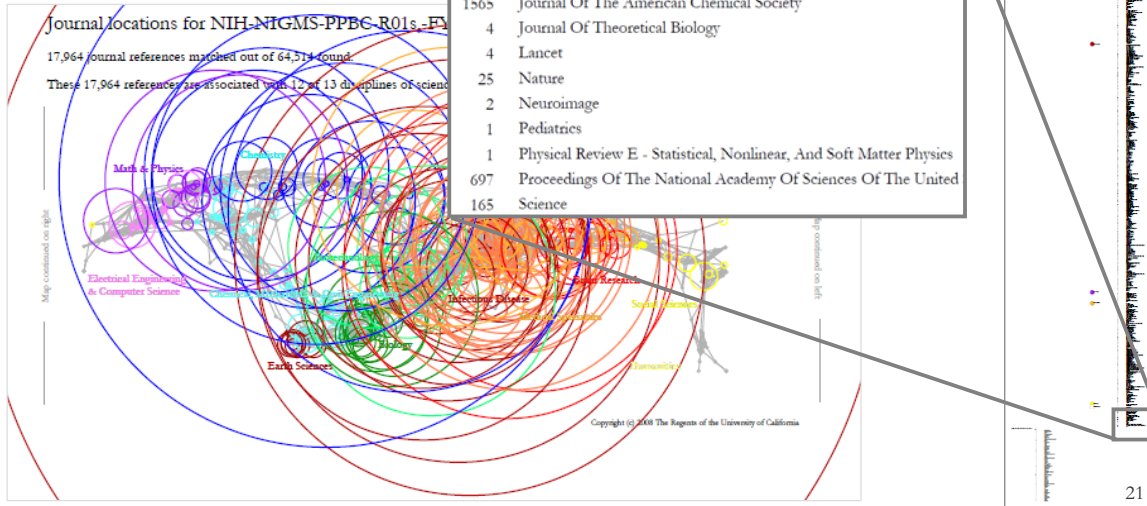
Loaded: C:\A\...NIH-NIGMS-PPBC-R01s,-FY08-Publications.csv

Science Map (Circle Annotation) was selected.

Input Parameters:

Journal column: Journal Title Abbr

Saved: NIH-NIGMS-PPBC-R01s,-FY08-Publications.ps



## 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:

Science Map via 554 Fields (Circle Annotations)

Locate UCSD area tagged records on the UCSD Map of Science

UCSD area: UCSD Map field name

Label: Knowledge areas or terms

Value: Value, e.g., total \$ amount spent

Scaling factor: 1.0

Dataset display name: Sample.csv

OK Cancel

## [#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

23



### Comparing the Portfolios of Different Funding Organizations

**MAPS OF SCIENCE**



Overview	Detail	Disciplinary Maps	Competency Maps	Paradigm Maps	Posters	Education
		Institutional Strategy: NIH				
		Institutional Strategy: NSF				
		Institutional Strategy: DOE				
		Institutional Strategy: NASA				

**BETTER MAPS • BETTER DECISIONS**



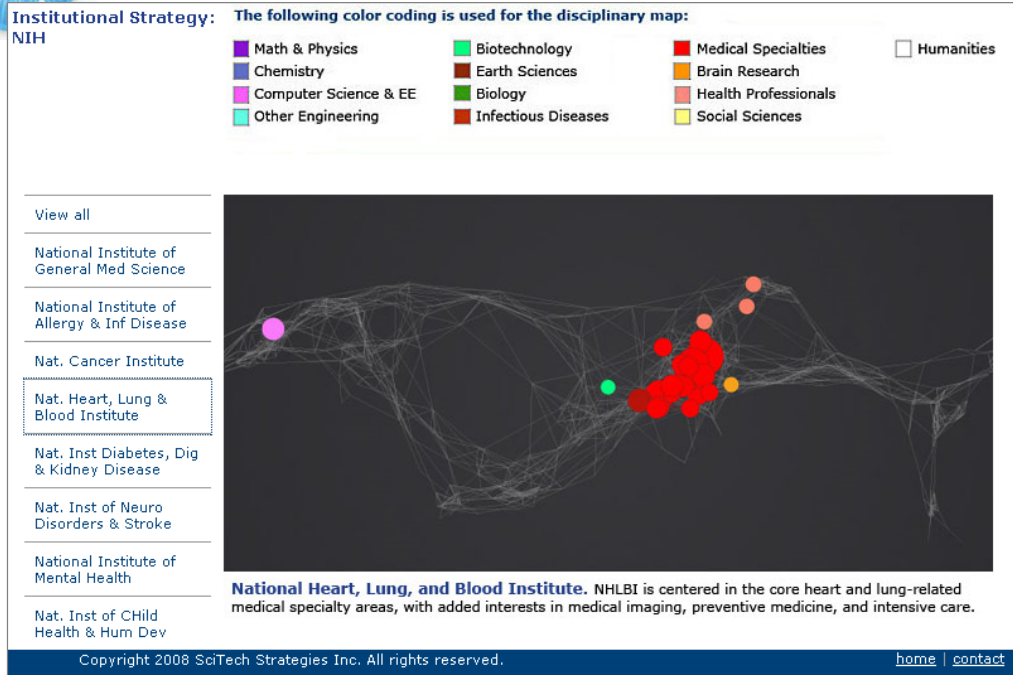
Copyright 2008 SciTech Strategies Inc. All rights reserved. [home](#) | [contact](#)

<http://mapofscience.com>

24

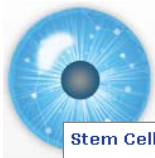


## Comparing the Portfolios of Different NIH

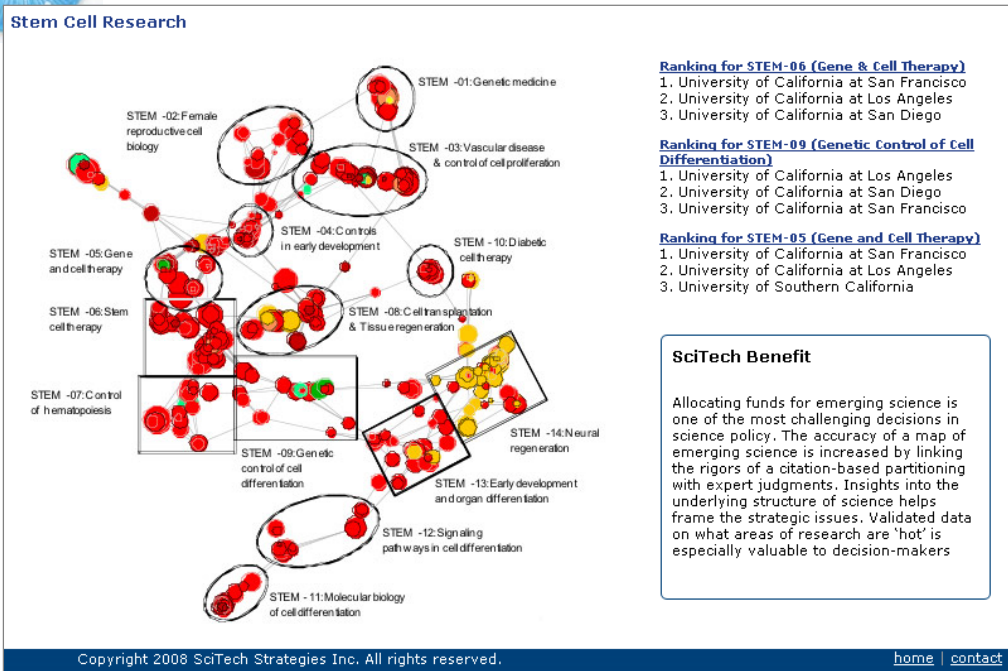


<http://mapofscience.com>

25



## Comparing Stem Cell Research Portfolios



<http://mapofscience.com>

26



## Topic Map of NIH Funding by Ned Talley et al

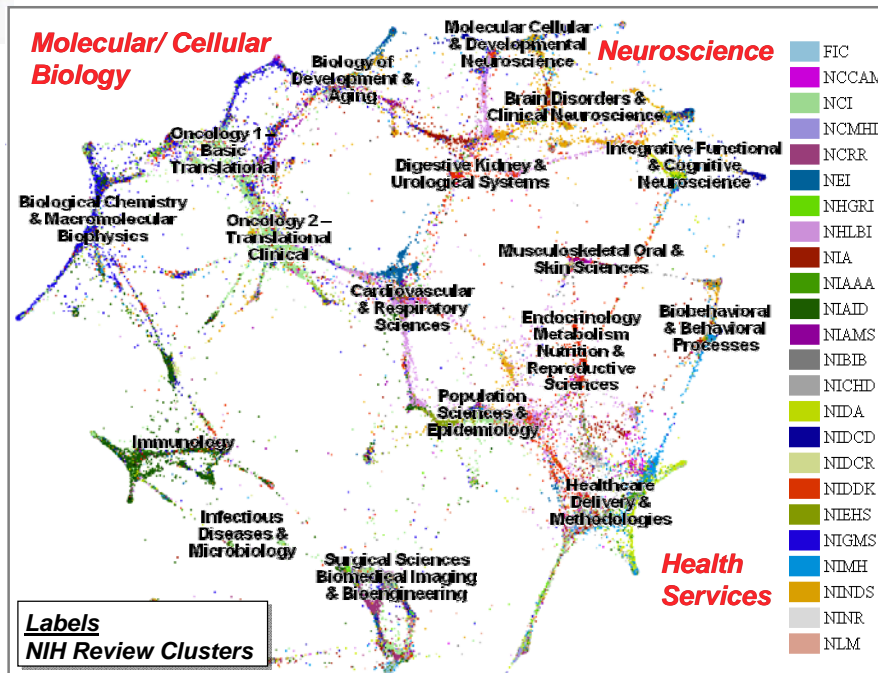
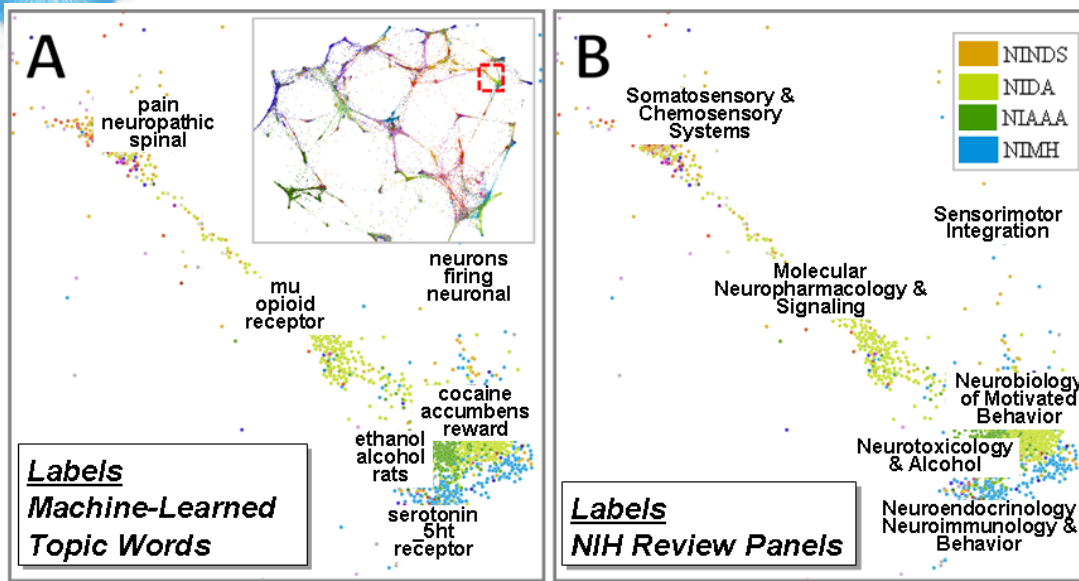


Figure 1. Graphical layout of NIH grants provides an organizational framework with compelling global structure.

27



## Comparing the Portfolios of Different Funding Organizations



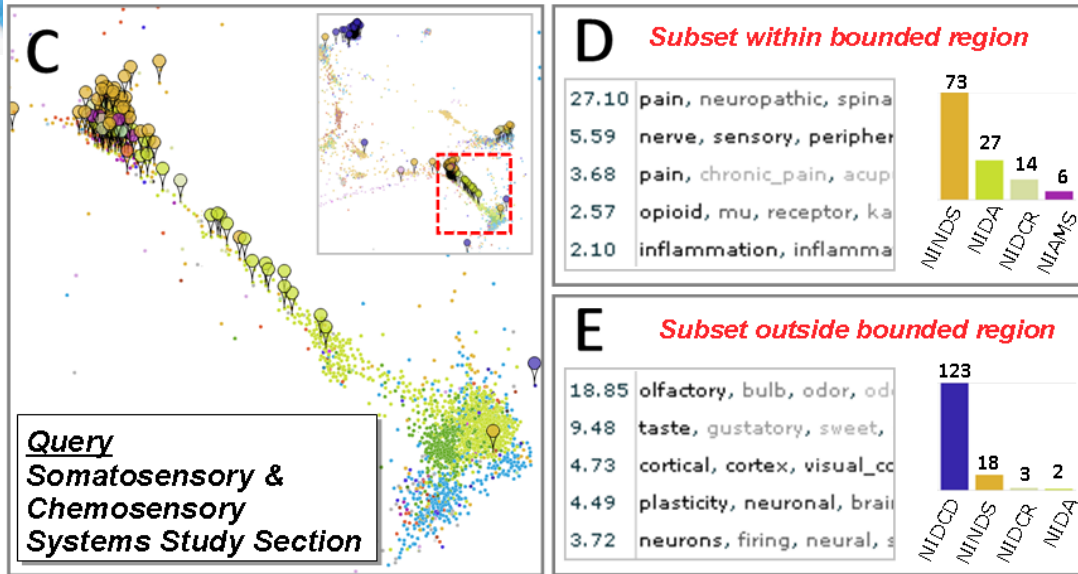
Relationships between machine-learned categories and NIH Institute/Review organization.

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/IFCNIIRG/>).

28



## 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/jfconry/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).

29



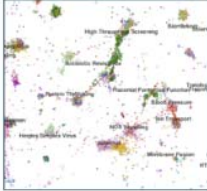
## Topic Map of NIH Funding by Ned Talley et al

# NIH MAPS

Hosted by BIRN

Home About Browse the Maps! Documentation Survey

### Navigating Topic Maps of NIH Grants




Attempting to make sense of the large-scale organization of a complex field of research is both very challenging and very common. All working scientists benefit from being able to understand the landscape of both scientific funding and publication and yet the scale of both the published literature and funding agencies portfolios are too large to be able to understand easily. We want to provide *practical mapping tools* that can help working scientists navigate this bewildering landscape.

Here, we present a highly-interactive mapping system based on a publicly available collections of funded scientific grants from the National Institutes of Health (NIH, source: CRISP). These documents comprise a (somewhat) comprehensive view of federally funded biomedical research in the United States (given that funding is also provided by other, non-NIH sources).

We hope that this tool allows scientists to examine how projects are

### Acknowledgements

This work is funded by the National Institute of Neurological Diseases and Stroke (NINDS) and the National Institute of General Medical Science (NIGMS). Early Support was provided by NSF (IIS-0513650) and unrestricted funds from the Information Sciences Institute.

ChalkLabs  
UCIRVINE  


## [#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

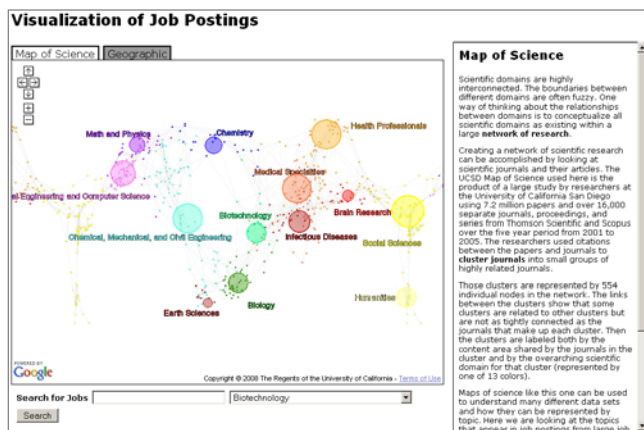
31



### 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 Sc&T Jobs. Angela Zoss, Michael Connover, Katy Börner (2010).*

<http://cns-nd3.slis.indiana.edu/mapjobs/geo/scivis.html?limit=5000>

32



# Reference Mapper

*Dubon & Börner, forthcoming.*

**(a) Overview**

**Date and input directory**

**Basic counts**

**Overlay of all matched journal references from all PDF files on 554 scientific disciplines (nodes) in UCSD Map of Science**

**Circle size denotes # references**

**Listing of all references grouped by 13 science areas**

**(b) Visual Index**

**For each PDF file: Basic counts and thumbnail science map**

**Max 18 per page**

**(c) Details**

**For each PDF file: Overlay of all matched journal references on 554 scientific fields (nodes) in UCSD Map of Science**

**Circle size denotes # references**

**Colors and names of science areas that are cited**

**Alphabetic listing of cited journals and # of times cited**

**(d) Top-10 Most Similar**

**Top-n most similar PDF files identified based on journal name co-occurrences. The similarity of each PDF file to itself is 1**

**Overlay of matched journal references from all above listed PDF files on UCSD Map of Science and grouping by 13 science areas**

33

## [#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

34



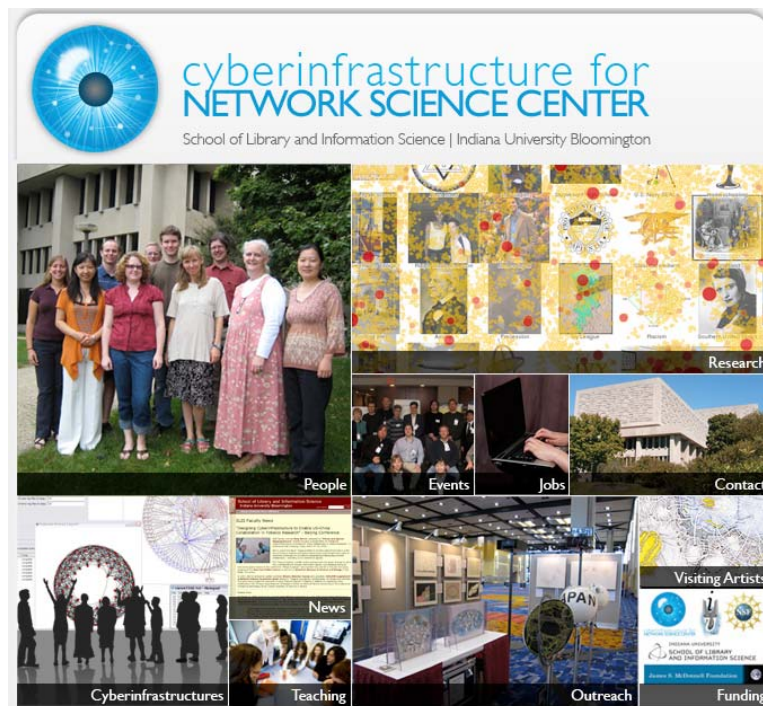
## Exercise

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.

35



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

36