

Envisioning Biomedical Science

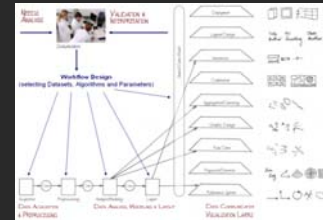
Dr. Katy Börner

Cyberinfrastructure for Network Science Center, Director
Information Visualization Laboratory, Director
School of Library and Information Science
Indiana University, Bloomington, IN
katy@indiana.edu

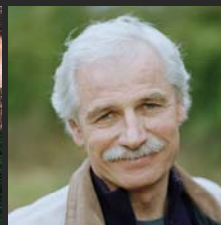
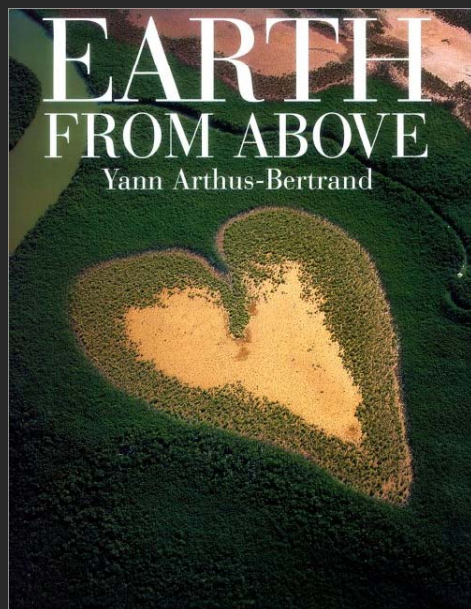
With special thanks to Kevin W. Boyack, Micah Linnemeier,
Russell J. Duhon, Patrick Phillips, Joseph Biberstine, Chintan Tank
Nianli Ma, Angela M. Zoss, Hanning Guo, Mark A. Price,
Scott Weingart

*Seminar, The Beckman Institute for Advanced Science and Technology
University of Illinois, IL*

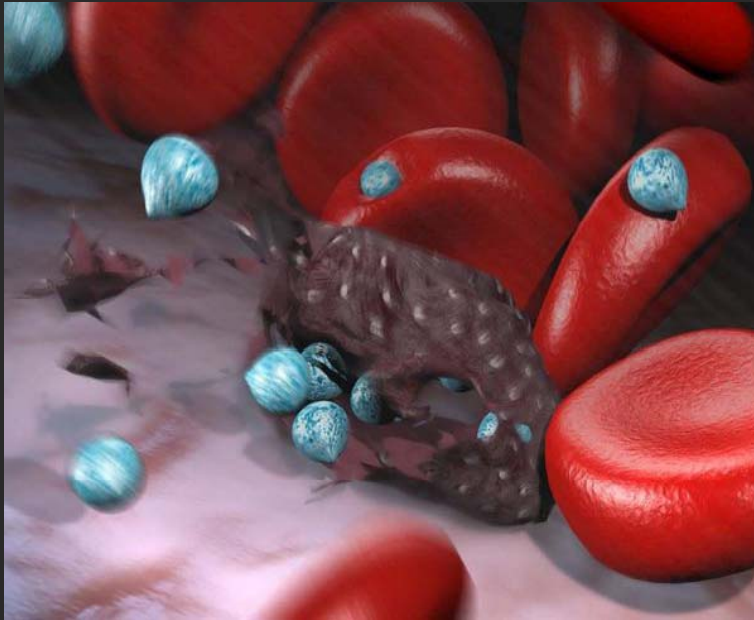
December 3rd, 2009



Biomedical Science from Above



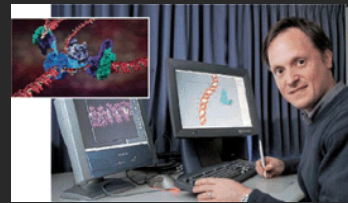
<http://www.home-2009.com>



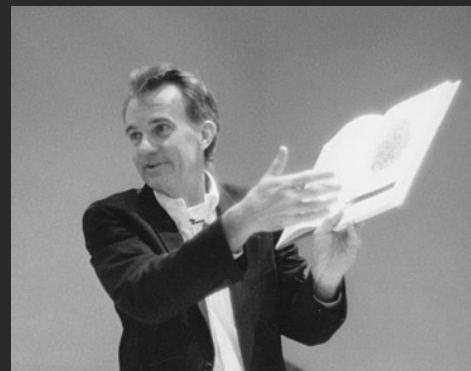
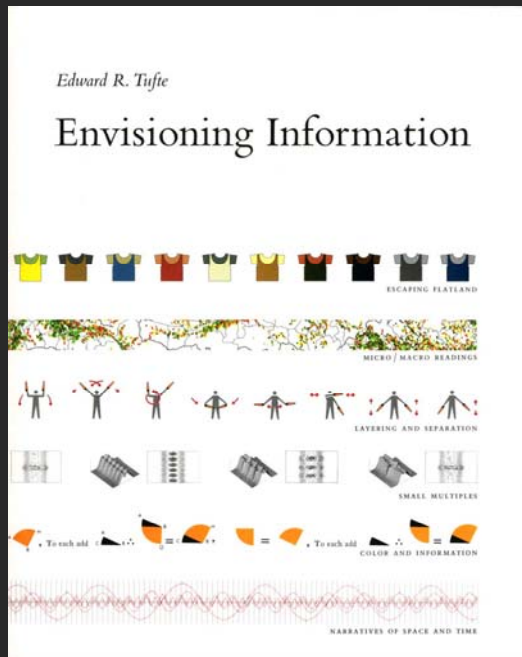
<http://www.malarialifecycle.com>

The Whole Brain Catalog:

<http://wholebraincatalog.org>



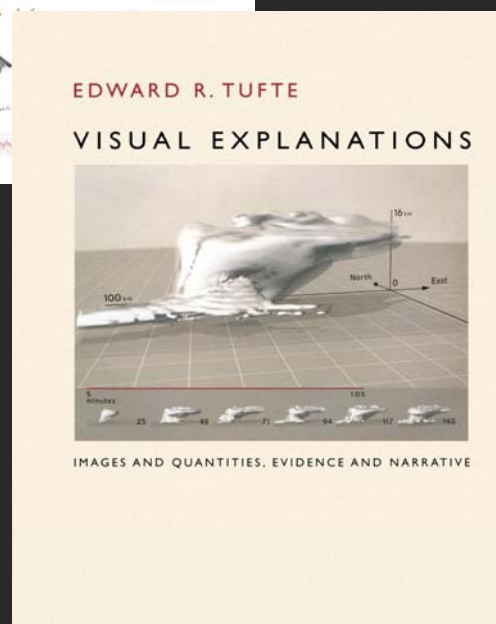
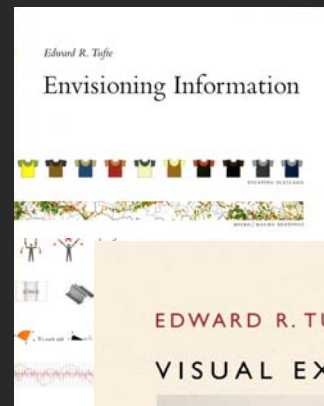
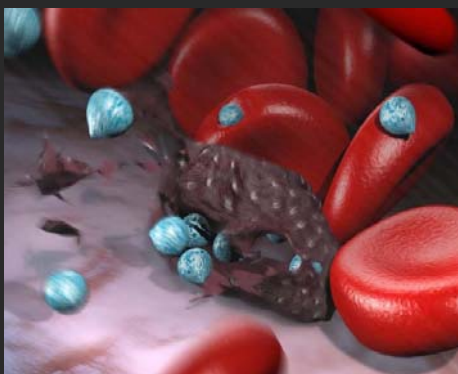
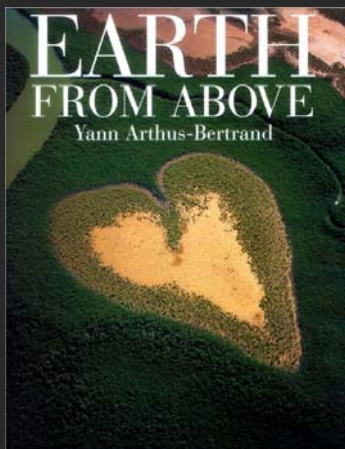
Drew Berry

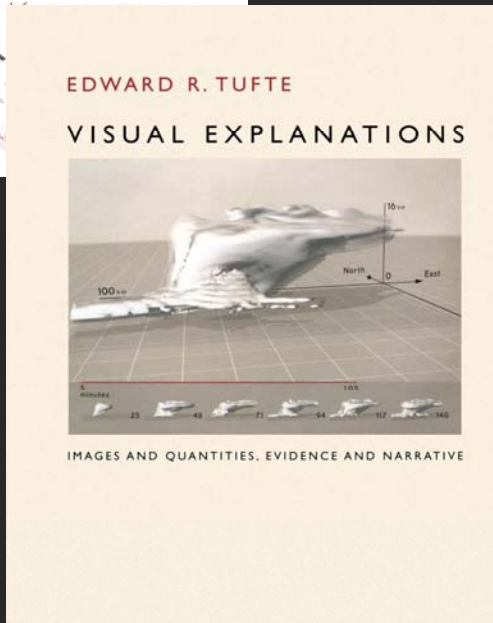
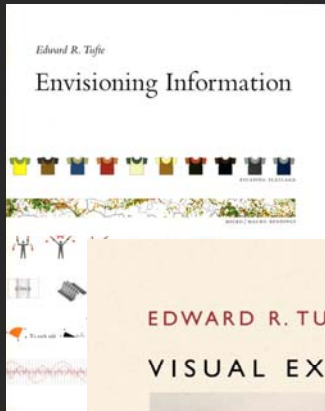


All three care deeply about

1. Data,
 2. Existing expertise and insight needs, and
 3. Are able to acquire the resources it takes to
- Spent months/years to deeply understand the problem/possible solutions.
 - Render data optimally for the human perceptual-cognitive system – given our current understanding of human perception/cognition and technology.

The result are insightful yet perceptually stunning, intellectually stimulating, and emotion provoking imagery.





Today's massive amounts of streaming data cannot be rendered by hand.

How to use computers to envision biomedical science?

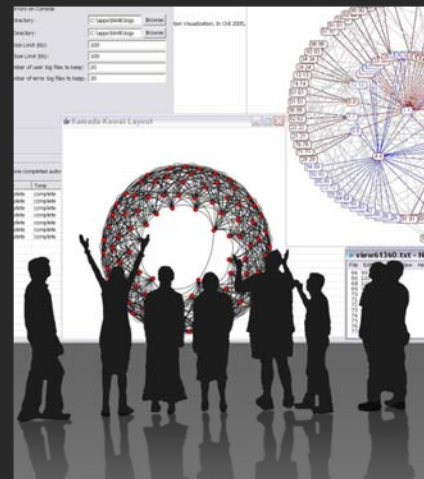
How to combine data mining and visualization algorithms to explore and communicate biomedical science?



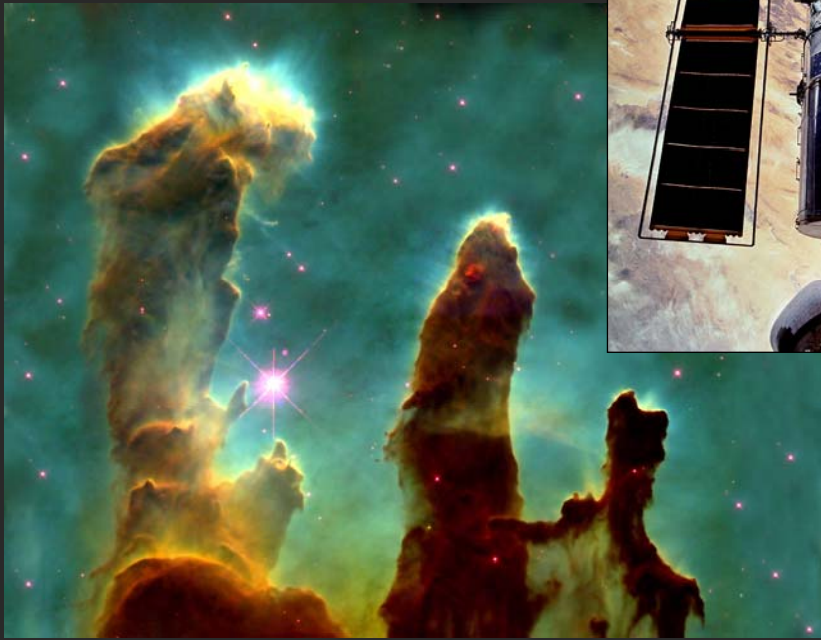
Microscopes



Telescopes



Macrosopes



Structure of the Remaining Talk

1.) Type of Analysis vs. Level of Analysis

Exemplified in Biomedicine

2.) Needs-Driven Workflow Design

using a modular data acquisition/analysis/modeling/
visualization pipeline as well as modular visualization layers.

Implementation in different plug-and-play tools/CIs



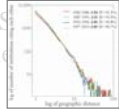
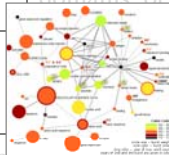


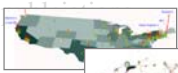
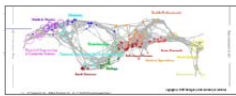



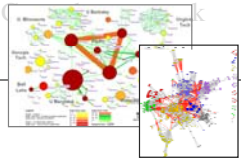

1.) Type of Analysis vs. Level of Analysis

| | <i>Micro/Individual (1-100 records)</i> | <i>Meso/Local (101-10,000 records)</i> | <i>Macro/Global (10,000 < records)</i> |
|---------------------------------------|--|---|---|
| Statistical Analysis/Profiling | Individual person and their expertise profiles | Larger labs, centers, universities, research domains, or states | All of NSF, all of USA, all of science. |
| Temporal Analysis (When) | Funding portfolio of one individual | Mapping topic bursts in 20-years of PNAS | 113 Years of Physics Research |
| Geospatial Analysis (Where) | Career trajectory of one individual | Mapping a states intellectual landscape | PNAS Publications |
| Topical Analysis (What) | Base knowledge from which one grant draws. | Knowledge flows in Chemistry research | VxOrd/Topic maps of NIH funding |
| Network Analysis (With Whom?) | NSF Co-PI network of one individual | Co-author network | NSF's core competency |

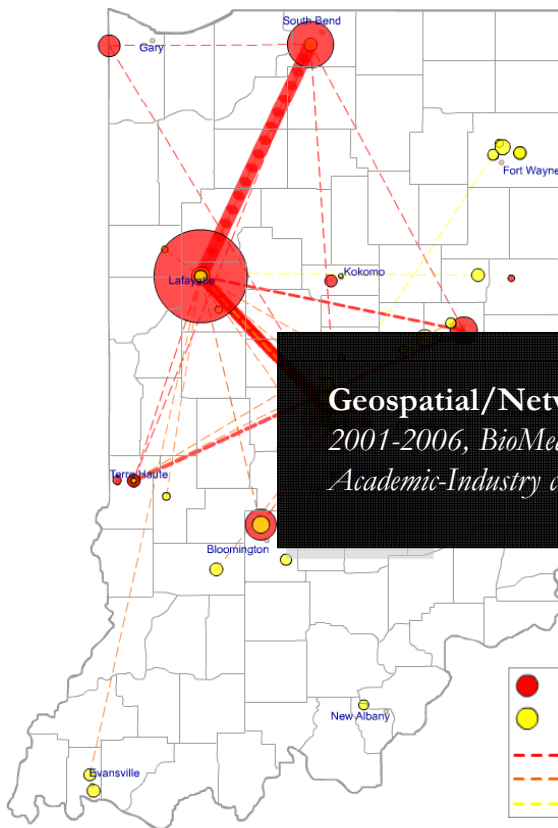
11



Type of Analysis vs. Level of Analysis

| | <i>Micro/Individual (1-100 records)</i> | <i>Meso/Local (101-10,000 records)</i> | <i>Macro/Global (10,000 < records)</i> |
|---------------------------------------|--|--|---|
| Statistical Analysis/Profiling | Individual person and their expertise profiles | Larger labs, centers, universities, research domains, or states | All of NSF, all of USA, all of science.  |
| Temporal Analysis (When) | Funding portfolio of one individual | Mapping topic bursts in 20-years of PNAS  | 113 Years of Physics Research  |
| Geospatial Analysis (Where) | Career trajectory of one individual | Mapping a states intellectual landscape  | PNAS Publications  |
| Topical Analysis (What) | Base knowledge from which one grant draws.  | Knowledge flows in Chemistry research  | VxOrd/Topic maps of NIH funding  |
| Network Analysis (With Whom?) | NSF Co-PI network of one individual  | Co-author network  | NSF's core competency  |

12

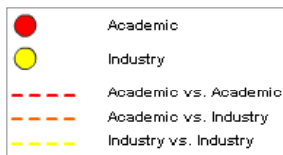


Mapping Indiana's Intellectual Space

Geospatial/Network Analysis

2001-2006, BioMed, IN Scope

Academic-Industry collaborations and knowledge diffusion

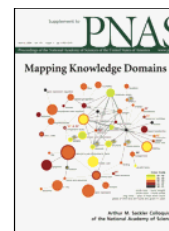
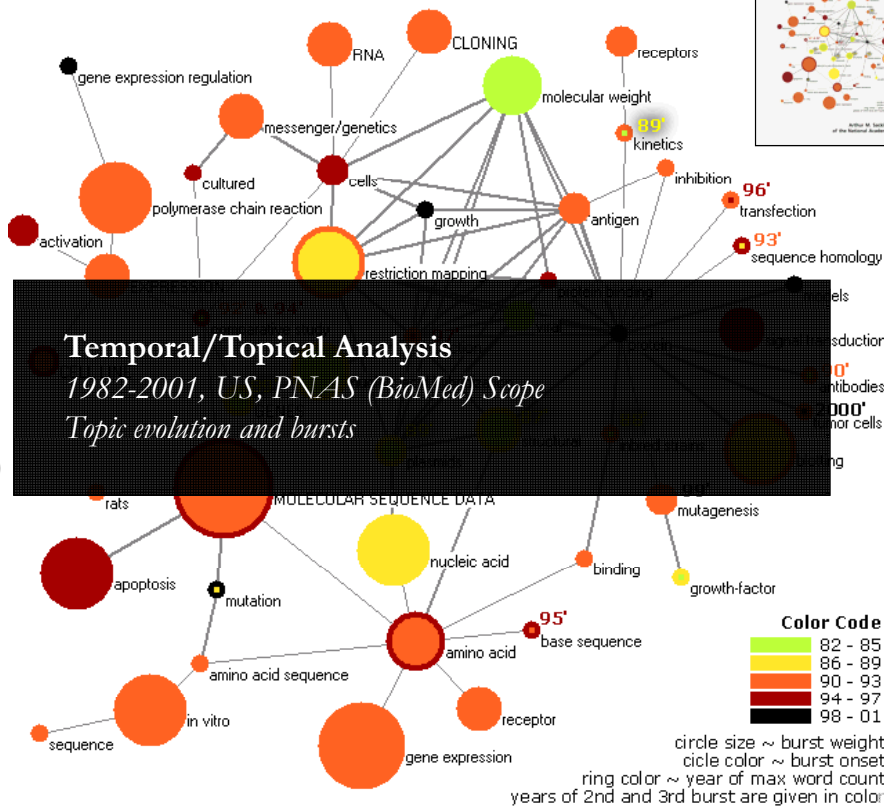


13

Mapping Topic Bursts

Co-word space of the top 50 highly frequent and bursty words used in the top 10% most highly cited PNAS publications in 1982-2001.

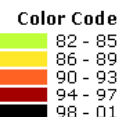
Mane & Börner. (2004) PNAS, 101(Suppl. 1): 5287-5290.



Temporal/Topical Analysis

1982-2001, US, PNAS (BioMed) Scope

Topic evolution and bursts



circle size ~ burst weight
 circle color ~ burst onset
 ring color ~ year of max word count
 years of 2nd and 3rd burst are given in color

14

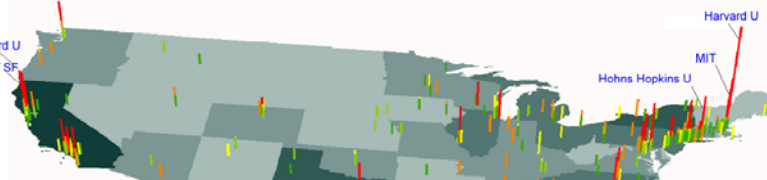
Spatio-Temporal Information Production and Consumption of Major U.S. Research Institutions

Börner, Katy, Penumarthu, Shashikant, Meiss, Mark and Ke, Weimao. (2006)
Mapping the Diffusion of Scholarly Knowledge Among Major U.S. Research Institutions. Scientometrics. 68(3), pp. 415-426

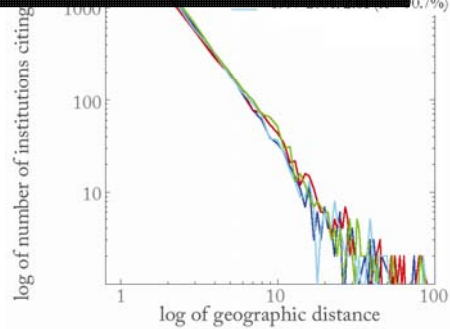


Research questions:

1. Does space still matter in the Internet age?
2. Does one still have to study and work at institutions in order to produce high quality data quality research?
3. Does the Internet change patterns, i.e., more produced at geographically distant research institutions?



Temporal/Geospatial Analysis
 1982-2001, US, PNAS (BioMed) Scope
 Citation impact and knowledge diffusion



Contributions:

- Answer to Qs 1 + 2 is YES.
- Answer to Qs 3 is NO.
- Novel approach to analyzing the dual role of institutions as information producers and consumers and to study and visualize the diffusion of information among them.

Research Collaborations by the Chinese Academy of Sciences

By Weixia (Bonnie) Huang, Russell J. Dubon, Elisha F. Hardy, Katy Börner, Indiana University, USA

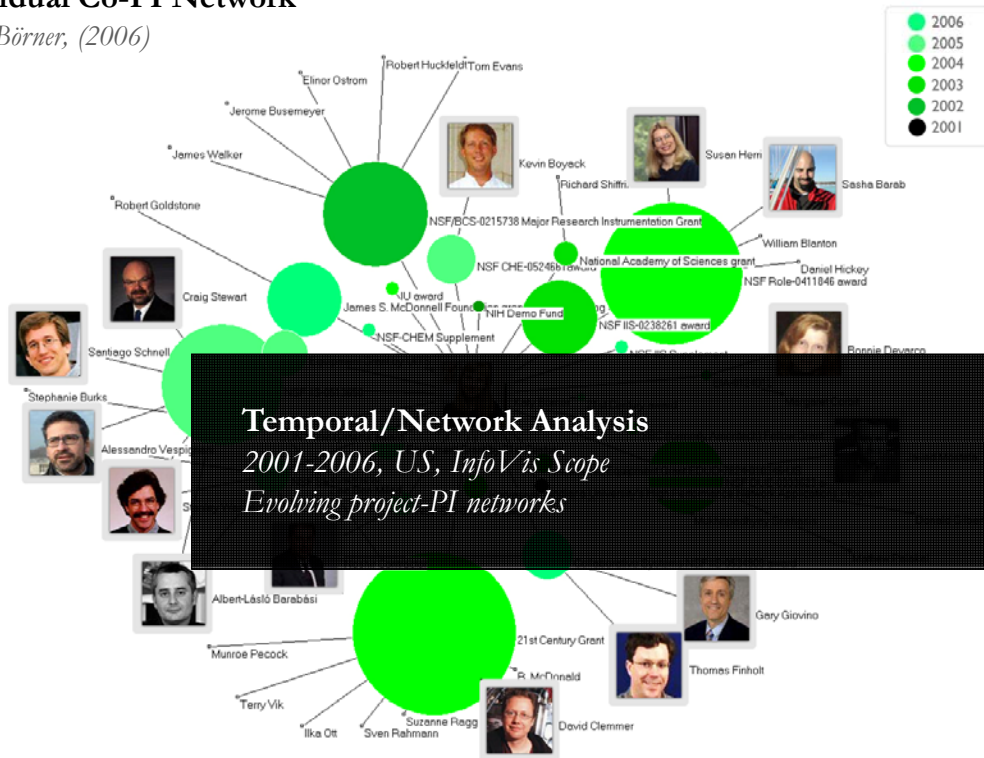


Geospatial Analysis
 World, Chinese Academy of Science
 Collaboration and knowledge diffusion via co-author networks

This map highlights the research collaborations of the Chinese Academy of Sciences with locations in China and countries around the world. The large geographic map shows the research collaborations of all CAS institutes. Each smaller geographic map shows the research collaborations by the CAS researchers in one province-level administrative division. Collaborations between CAS researchers are not included in the data. On each map, locations are colored on a logarithmic scale by the number of collaborations from red to yellow. The darkest red is 3,395 collaborations by all of CAS with researchers in Beijing. Also, flow lines are drawn from the location of focus to all locations collaborated with. The width of the flow line is linearly proportional to the number of collaborations with the locations it goes to, with the smallest flow lines representing one collaboration and the largest representing differing amounts on each geographic map.

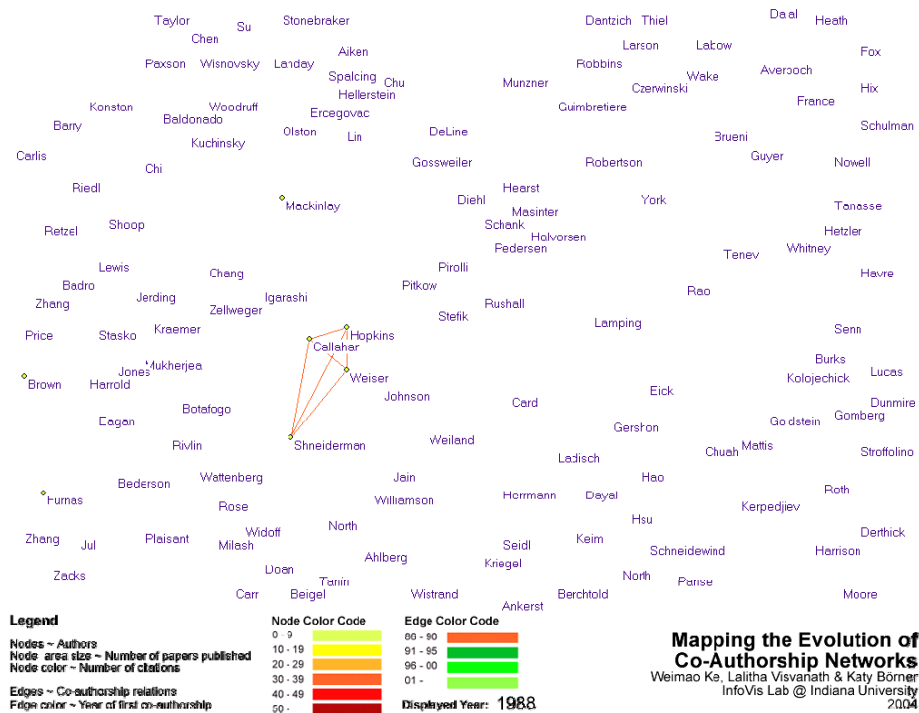
Individual Co-PI Network

Ke & Börner, (2006)



Mapping the Evolution of Co-Authorship Networks

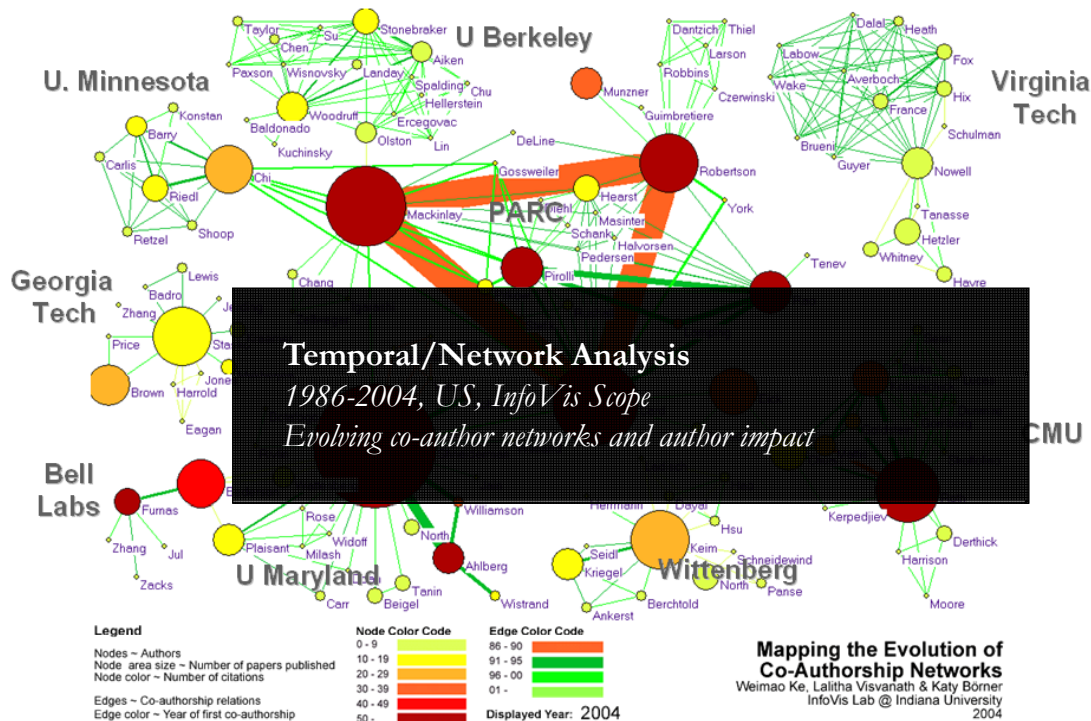
Ke, Visvanath & Börner, (2004) Won 1st price at the IEEE InfoVis Contest.



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

Mapping the Evolution of Co-Authorship Networks

Ke, Visvanath & Börner, (2004) Won 1st price at the IEEE InfoVis Contest



19

Studying the Emerging Global Brain: Analyzing and Visualizing the Impact of Co-Authorship Teams

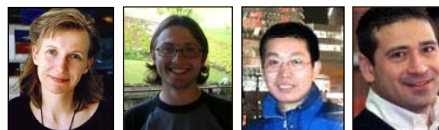
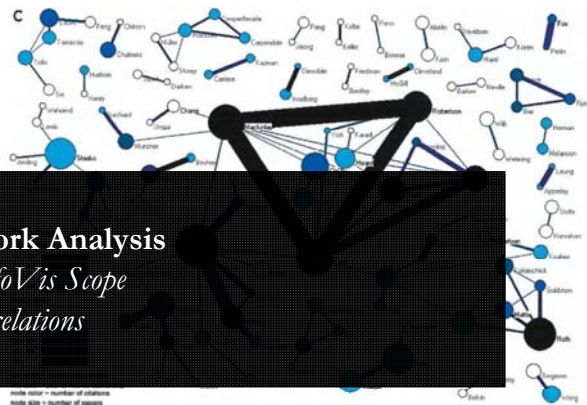
Börner, Dall'Asta, Ke & Vespignani (2005) Complexity, 10(4):58-67.

Research question:

- Is science driven by prolific single experts or by high-impact co-authorship teams?

Contributions:

- New approach to allocate citational credit.
- Novel weighted graph
- Visualization of the co-author network
- Centrality measure impact.
- Global statistical analysis of paper production and citations in correlation with co-authorship team size over time.
- Local, author-centered entropy measure.

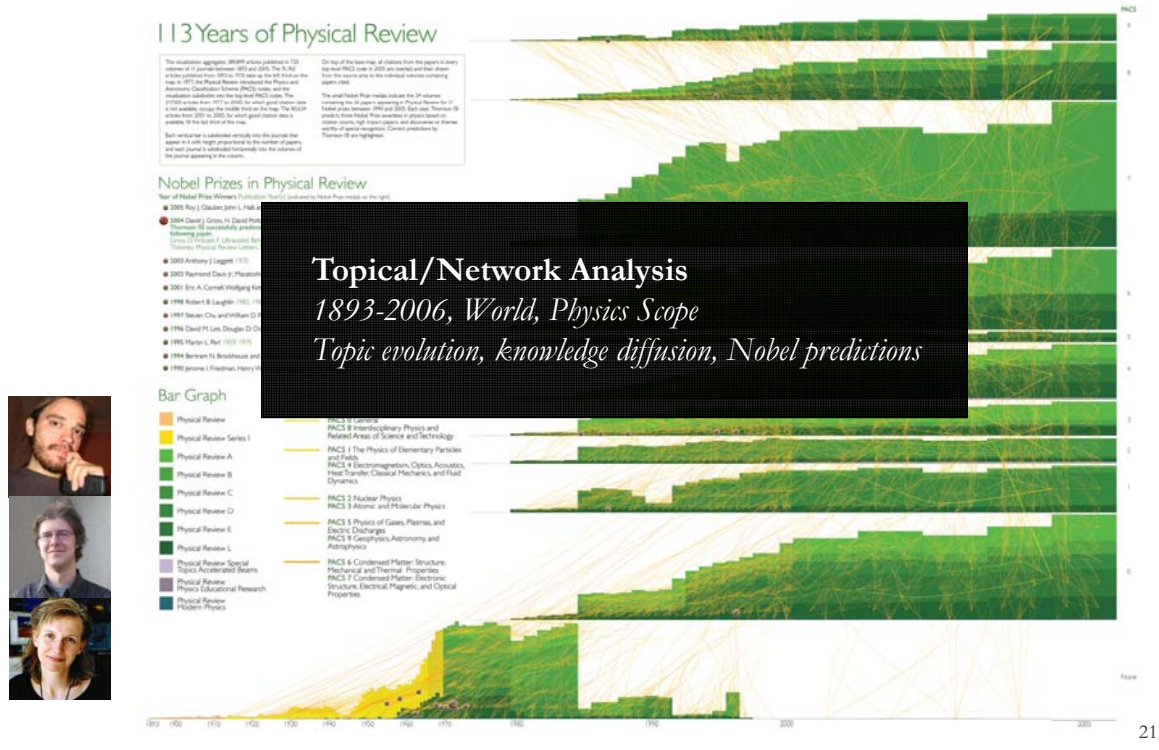


20

113 Years of Physical Review

http://scimaps.org/dev/map_detail.php?map_id=171

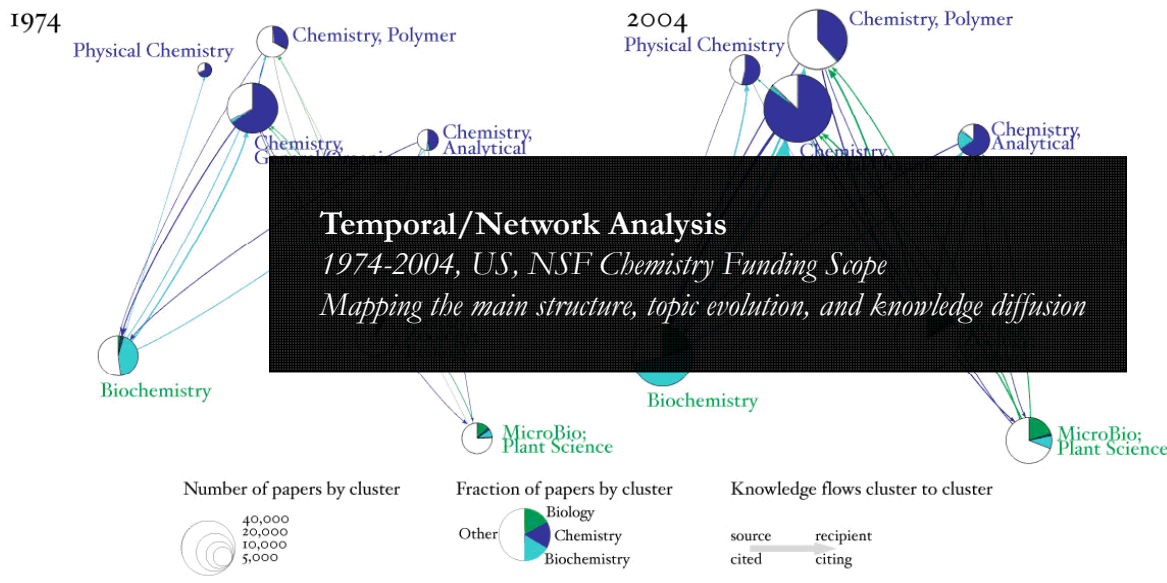
Bruce W. Herr II and Russell Dubon (Data Mining & Visualization), Elisha F. Hardy (Graphic Design), Shashikant Penumarthy (Data Preparation) and Katy Börner (Concept)



Topical Composition and Knowledge Flow Patterns in Chemistry Research for 1974 and 2004

Kevin W. Boyack, Katy Börner, & Richard Klavans (2007)

Chemistry - Biology Interface

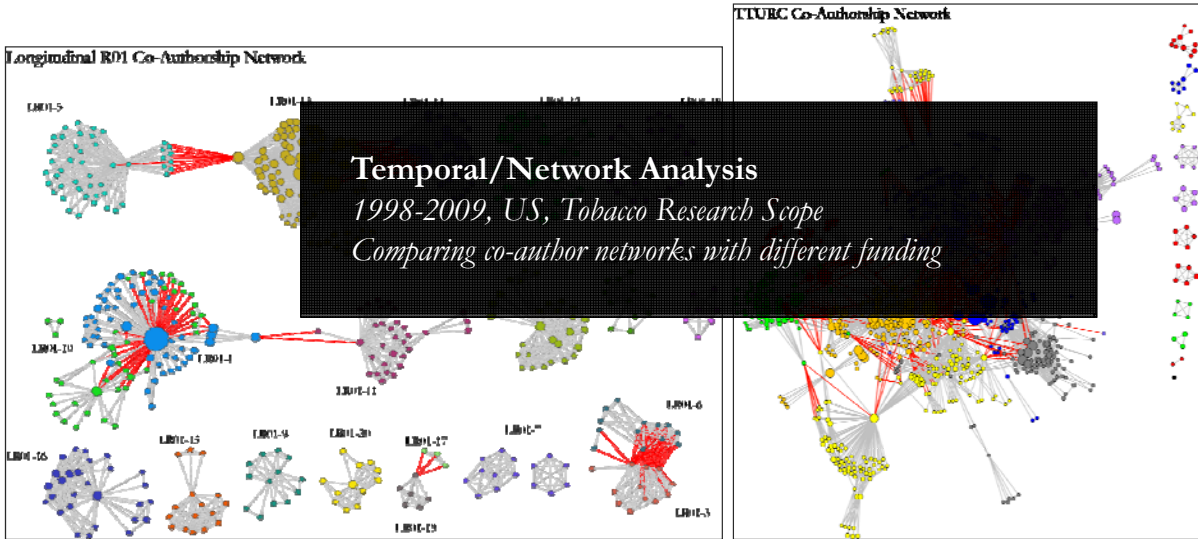
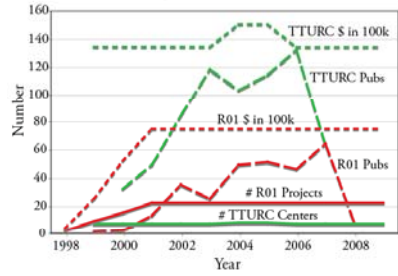


Mapping Transdisciplinary Tobacco Use Research Centers Publications

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

Zoss & Börner, *forthcoming*.

R01 & TTURC Project Information



Reference Mapper

Dubon & Börner, *forthcoming*.

(a) Overview

(b) Visual Index

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

For each PDF file:
Basic counts and thumbnail science map

Max 18 per page

Topical/Network Analysis
2009, US, NSF Funding
Grouping interdisciplinary funding proposals for review

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

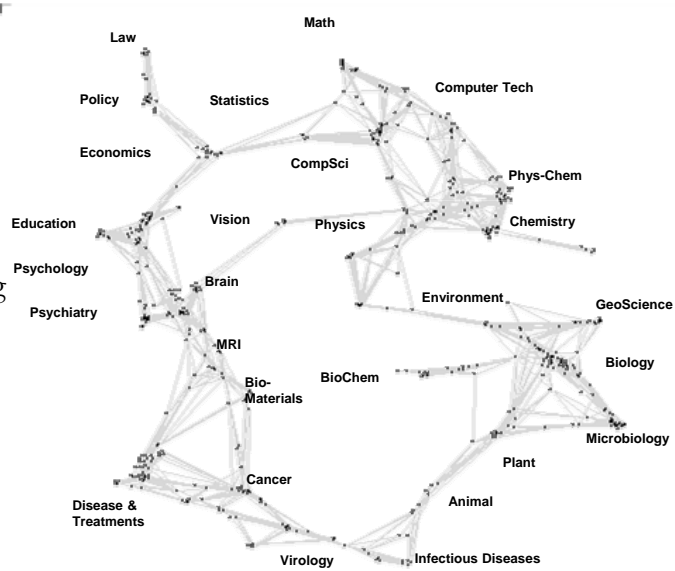
Top-n most similar PDF files identified based on journal name co-occurrences
The similarity of each PDF file to itself is 1

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

Latest 'Base Map' of Science

Kevin W. Boyack, Katy Börner, & Richard Klavans (2007). *Mapping the Structure and Evolution of Chemistry Research*. 11th International Conference on Scientometrics and Informetrics. pp. 112-123.

- Uses combined SCI/SSCI from 2002
 - 1.07M papers, 24.5M references, 7,300 journals
 - Bibliographic coupling of papers, aggregated to journals
- Initial ordination and clustering of journals gave 671 clusters
- Coupling counts were reaggregated at the journal cluster level to calculate the
 - (x,y) positions for each journal cluster
 - by association, (x,y) positions for each journal

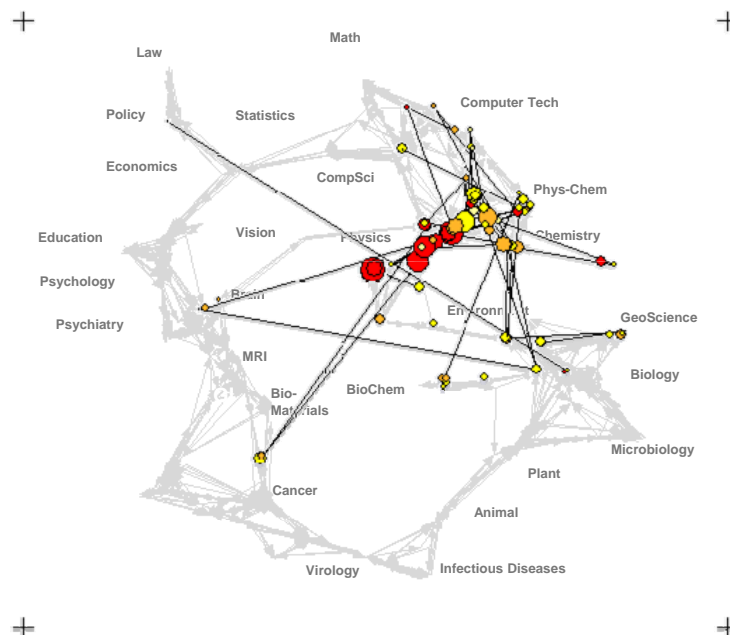


25

Science map applications: Identifying core competency

Kevin W. Boyack, Katy Börner, & Richard Klavans (2007).

Funding patterns of the US Department of Energy (DOE)

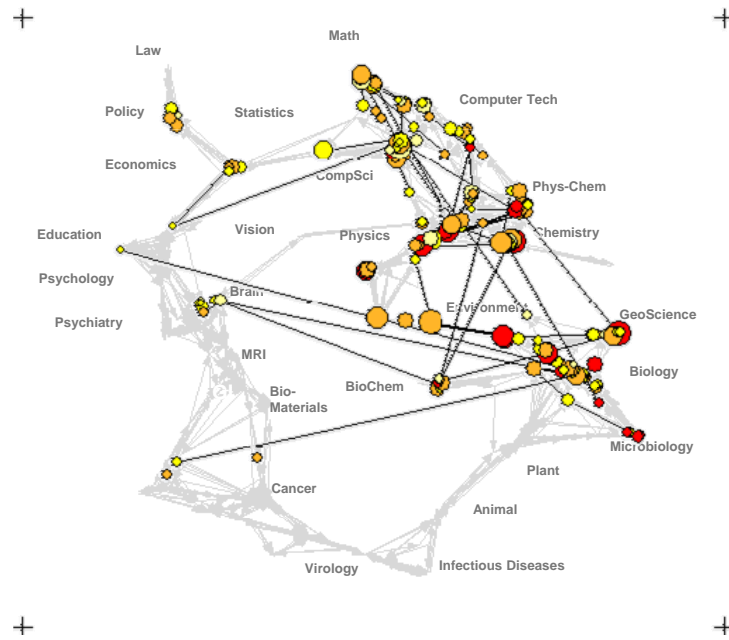


26

Science map applications: Identifying core competency

Kevin W. Boyack, Katy Börner, & Richard Klavans (2007).

Funding Patterns of the National Science Foundation (NSF)

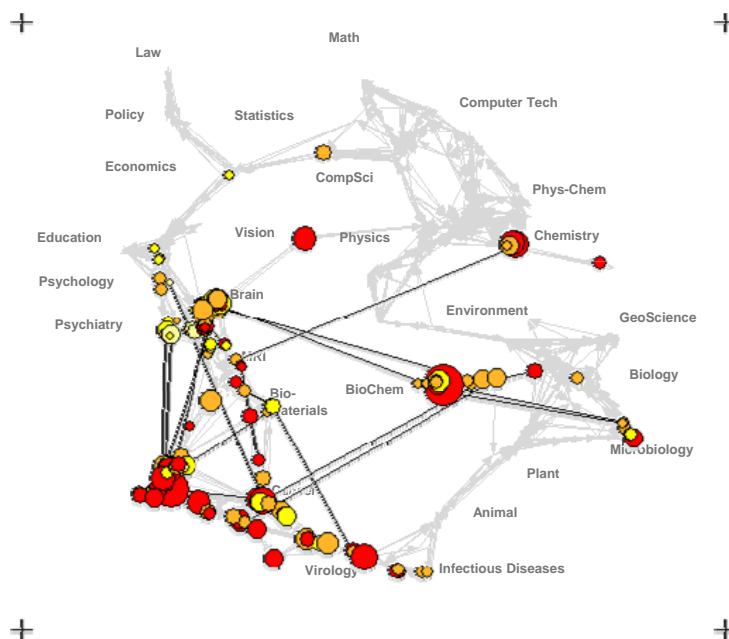


27

Science map applications: Identifying core competency

Kevin W. Boyack, Katy Börner, & Richard Klavans (2007).

Funding Patterns of the National Institutes of Health (NIH)

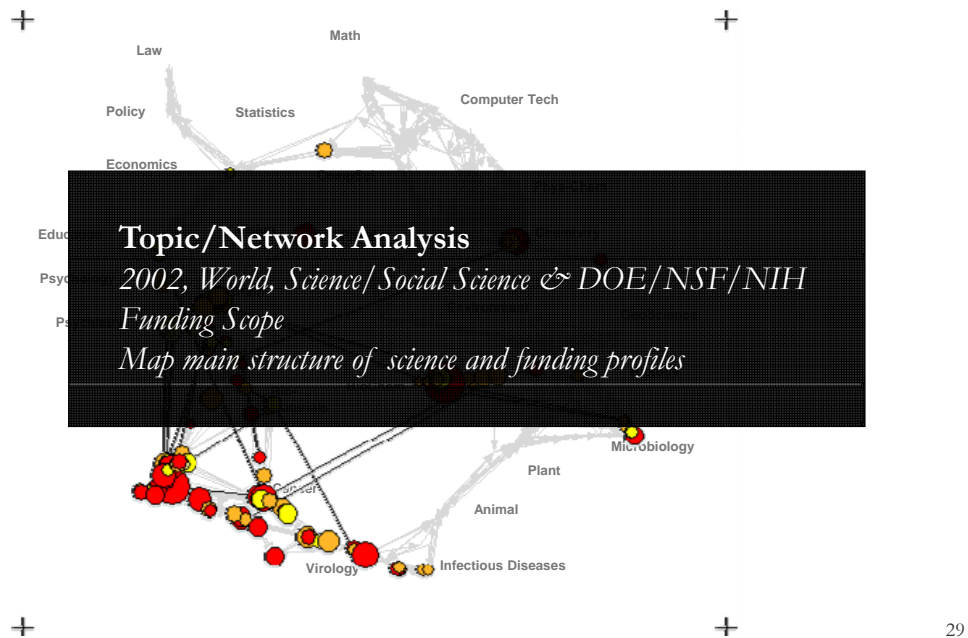


28

Science map applications: Identifying core competency

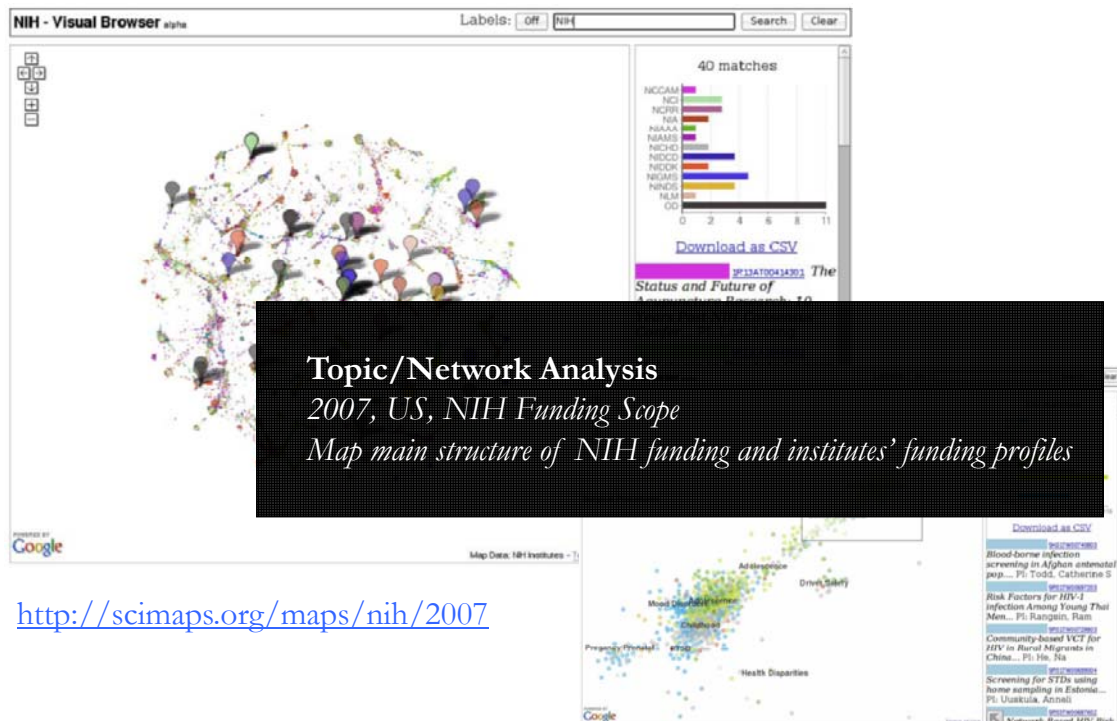
Kevin W. Boyack, Katy Börner, & Richard Klavans (2007).

Funding Patterns of the National Institutes of Health (NIH)



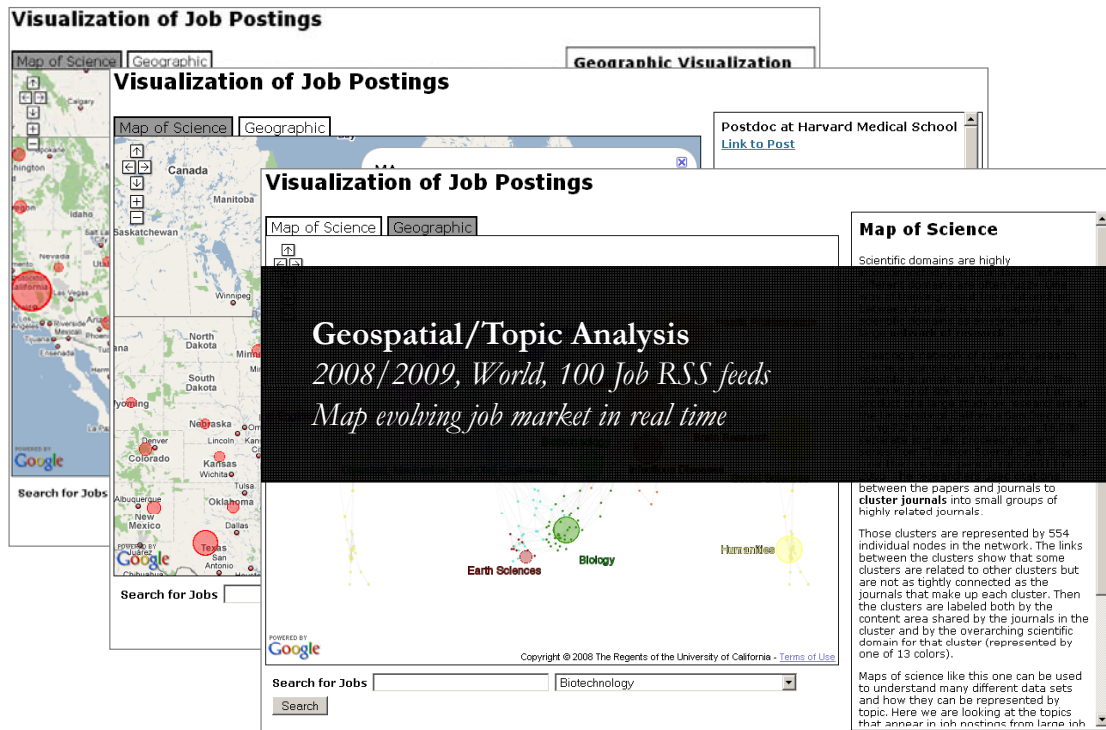
Interactive Science Map of NIH Funding

Herr II, Bruce W., Talley, Edmund M, Burns, Gully APC, Newman, David & La Rowe, Gavin. (2009).



Interactive World and Science Map of S&T Jobs

Angela Zoss, Michael Connover, Katy Börner (in preparation).



31

Structure of the Remaining Talk

1.) Type of Analysis vs. Level of Analysis

Exemplified in Biomedicine

2.) Needs-Driven Workflow Design

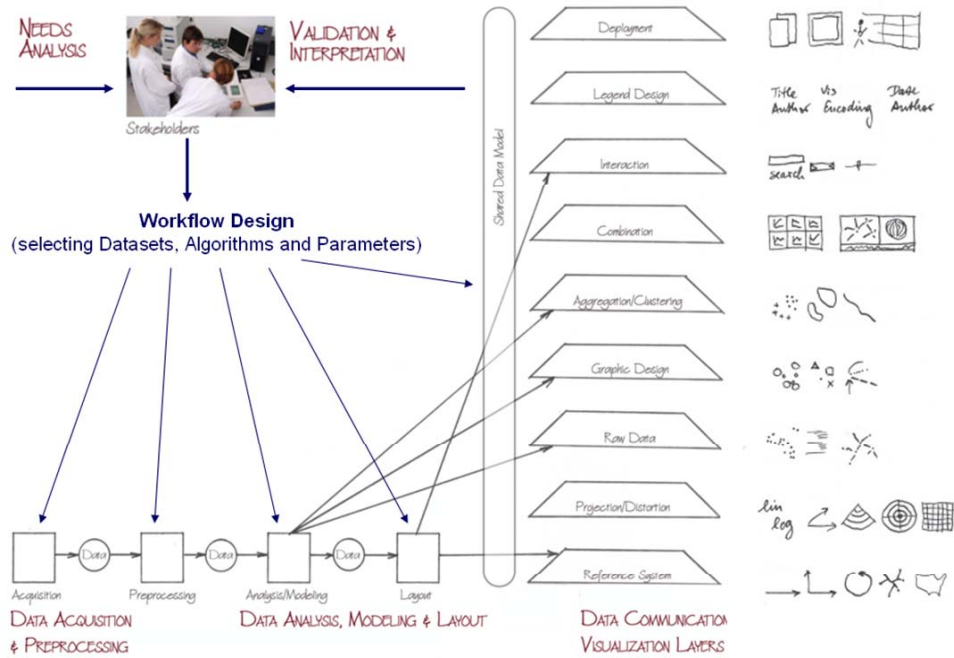
using a modular data acquisition/analysis/modeling/
visualization pipeline as well as modular visualization layers.

Implementation in different plug-and-play tools/CIs



2.) Needs-Driven Workflow Design

using a modular data acquisition/analysis/modeling/ visualization pipeline as well as modular visualization layers.



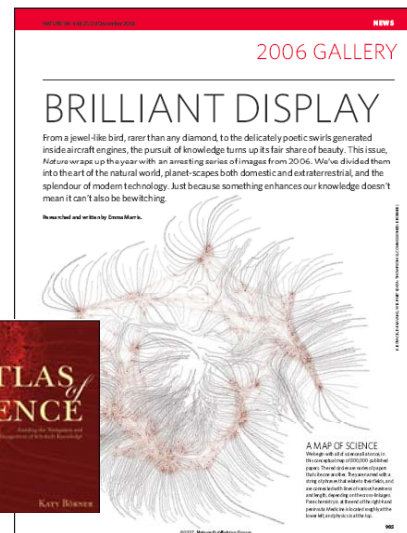
Börner, Katy (2010) *Atlas of Science*. MIT Press. 33

Computational Scientometrics: Studying Science by Scientific Means

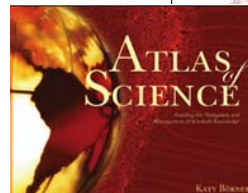
Börner, Katy, Chen, Chaomei, and Boyack, Kevin. (2003). **Visualizing Knowledge Domains**. In Blaise Cronin (Ed.), *ARIST*, Medford, NJ: Information Today, Inc./American Society for Information Science and Technology, Volume 37, Chapter 5, pp. 179-255.
<http://ivl.slis.indiana.edu/km/pub/2003-borner-arist.pdf>



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, Sanyal, Soma and Vespignani, Alessandro (2007). **Network Science**. In Blaise Cronin (Ed.), *ARIST*, Information Today, Inc./American Society for Information Science and Technology, Medford, NJ, Volume 41, Chapter 12, pp. 537-607.
<http://ivl.slis.indiana.edu/km/pub/2007-borner-arist.pdf>

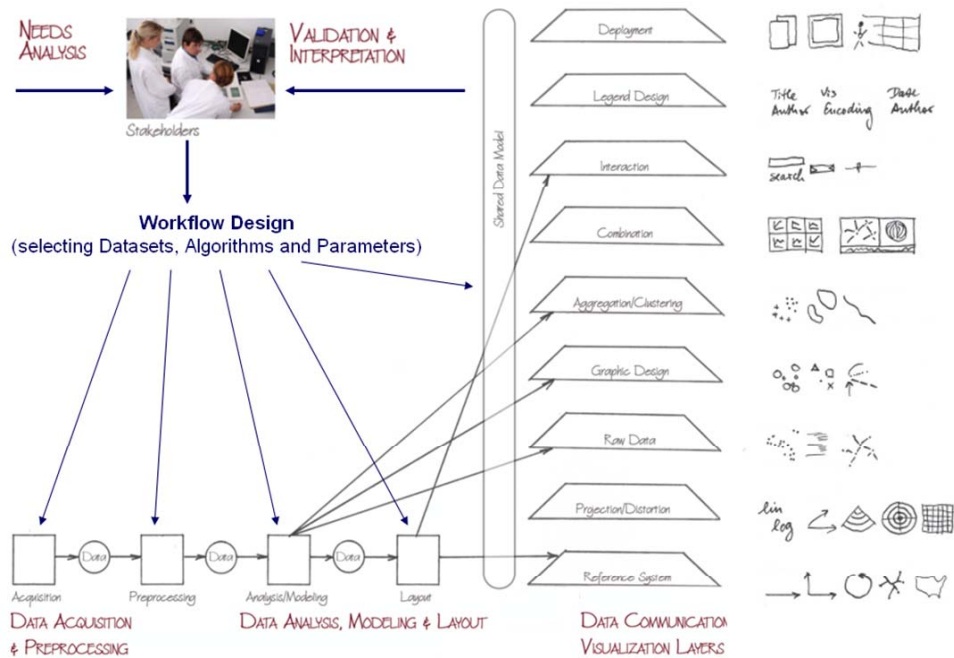


Börner, Katy (2010) *Atlas of Science*. MIT Press.
<http://scimaps.org/atlas>



2.) Needs-Driven Workflow Design

using a modular data acquisition/analysis/modeling/ visualization pipeline as well as modular visualization layers.



Börner, Katy (2010) *Atlas of Science*. MIT Press. 35



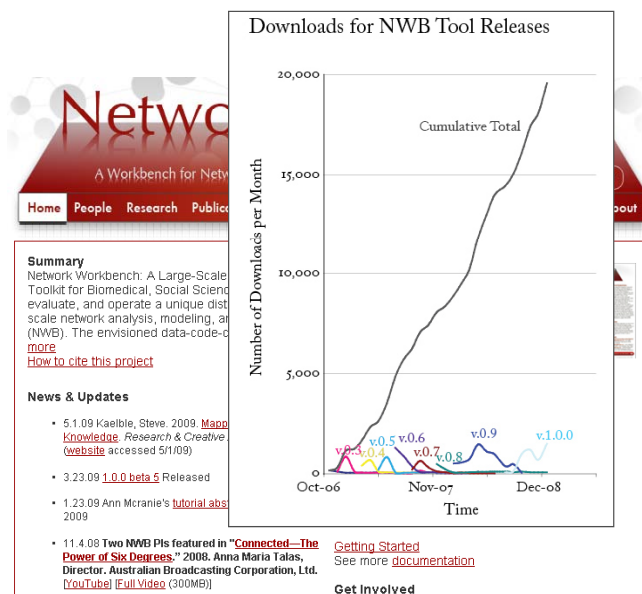
Network Workbench Tool

<http://nwb.slis.indiana.edu>

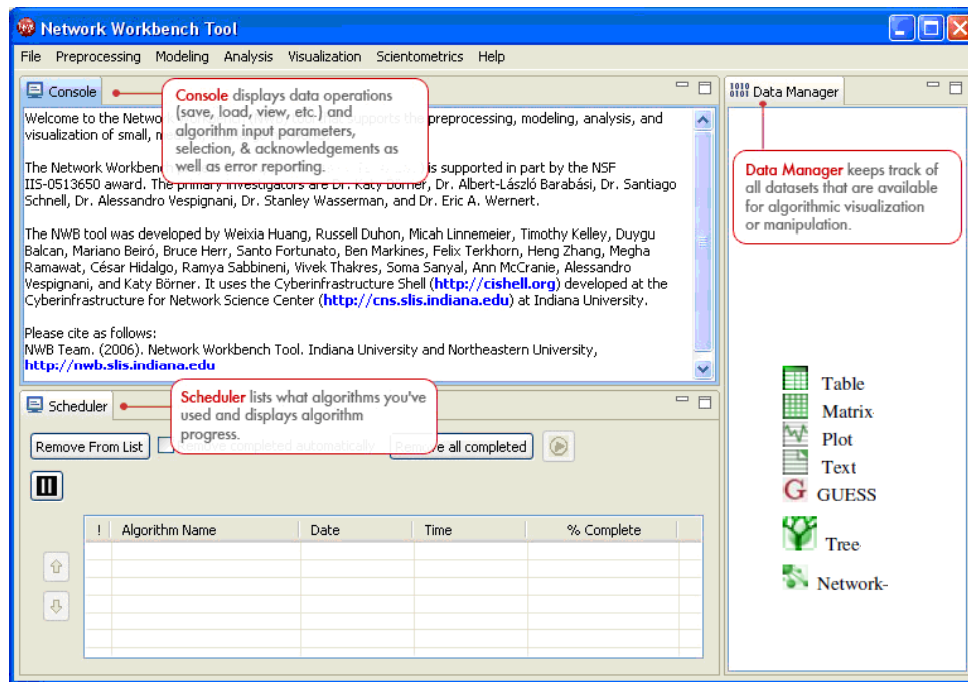
The Network Workbench (NWB) tool supports researchers, educators, and practitioners interested in the study of biomedical, social and behavioral science, physics, and other networks.

In Aug. 2009, the tool provides more 160 plugins that support the preprocessing, analysis, modeling, and visualization of networks.

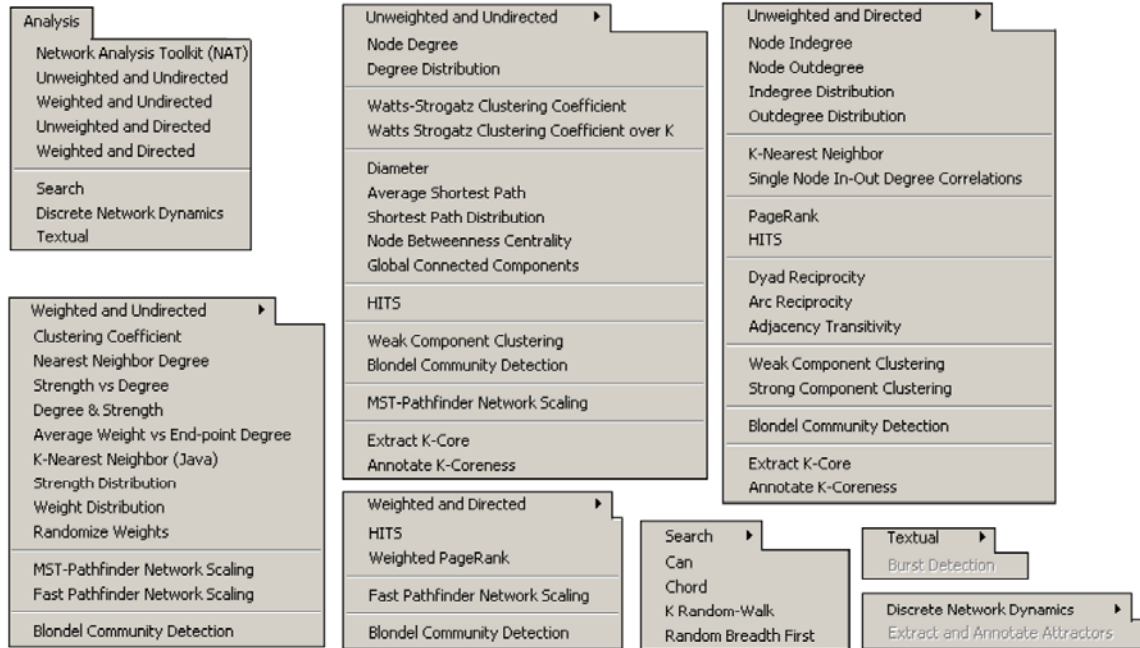
It has been downloaded more than 35,000 times since December 2006.



Herr II, Bruce W., Huang, Weixia (Bonnie), Penumarthy, Shashikant & Börner, Katy. (2007). *Designing Highly Flexible and Usable Cyberinfrastructures for Convergence*. In Bainbridge, William S. & Roco, Mibail C. (Eds.), *Progress in Convergence - Technologies for Human Wellbeing* (Vol. 1093, pp. 161-179), *Annals of the New York Academy of Sciences*, Boston, MA.



| File | Preprocessing | Modeling | Visualization |
|------------------------------------|------------------------------------|---------------------------------|---|
| Load... | Extract Top Nodes | Random Graph | GUESS |
| Load and Clean ISI File | Extract Nodes Above or Below Value | Waltz-Stringatz Small World | Group Plot |
| Read Directory Hierarchy Datasets | Remove Node Attributes | Darabási-Albert Scale-Free | DrL (VxOrd) |
| Save... | Delete High Degree Nodes | Can | Specified (prefuse beta) |
| View... | Delete Random Nodes | Chorc | Circular (JUNG) |
| View with... | Delete Isolates | Hypergrid | Radial Tree/Graph (prefuse alpha) |
| Merge Node and Edge Files | Extract Top Edges | PRU | Radial Tree/Graph with Annotation (prefuse beta) |
| Split Graph to Node and Edge Files | Extract Edges Above or Below Value | TARL | Tree Map (prefuse beta) |
| Tests | Remove Edge Attributes | Discrete Network Dynamics (DND) | Tree view (prefuse beta) |
| Preferences | Remove Self Loops | Evolving Network (Weighted) | Balloon Graph (prefuse alpha) |
| Exit | Trim by Degree | | Force Directed with Annotation (prefuse beta) |
| | Snowball Sampling (n nodes) | | Kamada-Kawai (JUNG) |
| | Node Sampling | | Fruchterman-Reingold (JUNG) |
| | Edge Sampling | | Fruchterman-Reingold with Annotation (prefuse beta) |
| | Symmetrize | | Spring (JUNG) |
| | Dichotomize | | Small World (prefuse alpha) |
| | Multipartite Joining | | Parallel Coordinates (demo) |
| | Normalize Text | | LaNet |
| | Slice Table by Time | | Circular Hierarchy |



The screenshot displays the 'Analysis' menu and its various submenus. The main 'Analysis' menu includes: Network Analysis Toolkit (NAT), Unweighted and Undirected, Weighted and Undirected, Unweighted and Directed, Weighted and Directed, Search, Discrete Network Dynamics, and Textual. Submenus include:

- Unweighted and Undirected:** Node Degree, Degree Distribution, Watts-Strogatz Clustering Coefficient, Watts Strogatz Clustering Coefficient over K, Diameter, Average Shortest Path, Shortest Path Distribution, Node Betweenness Centrality, Global Connected Components, HITS, Weak Component Clustering, Blondel Community Detection, MST-Pathfinder Network Scaling, Extract K-Core, Annotate K-Coreness.
- Unweighted and Directed:** Node Indegree, Node Outdegree, Indegree Distribution, Outdegree Distribution, K-Nearest Neighbor, Single Node In-Out Degree Correlations, PageRank, HITS, Dyad Reciprocity, Arc Reciprocity, Adjacency Transitivity, Weak Component Clustering, Strong Component Clustering, Blondel Community Detection, Extract K-Core, Annotate K-Coreness.
- Weighted and Undirected:** Clustering Coefficient, Nearest Neighbor Degree, Strength vs Degree, Degree & Strength, Average Weight vs End-point Degree, K-Nearest Neighbor (Java), Strength Distribution, Weight Distribution, Randomize Weights, MST-Pathfinder Network Scaling, Fast Pathfinder Network Scaling, Blondel Community Detection.
- Weighted and Directed:** HITS, Weighted PageRank, Fast Pathfinder Network Scaling, Blondel Community Detection.
- Search:** Can, Chord, K Random-Walk, Random Breadth First.
- Textual:** Burst Detection, Discrete Network Dynamics, Extract and Annotate Attractors.

Investigators: Katy Börner, Albert-Laszlo Barabasi, Santiago Schnell, Alessandro Vespignani & Stanley Wasserman, Eric Wernert



Software Team: Lead: Micah Linnemeier
 Members: Patrick Phillips, Russell Duhon, Tim Kelley & Ann McCranie
 Previous Developers: Weixia (Bonnie) Huang, Bruce Herr, Heng Zhang, Duygu Balcan, Bryan Hook, Ben Markines, Santo Fortunato, Felix Terkhorn, Ramya Sabbineni, Vivek S. Thakre & Cesar Hidalgo



Goal: Develop a large-scale network analysis, modeling and visualization toolkit for physics, biomedical, and social science research.

Amount: \$1,120,926, NSF IIS-0513650 award

Duration: Sept. 2005 - Aug. 2009

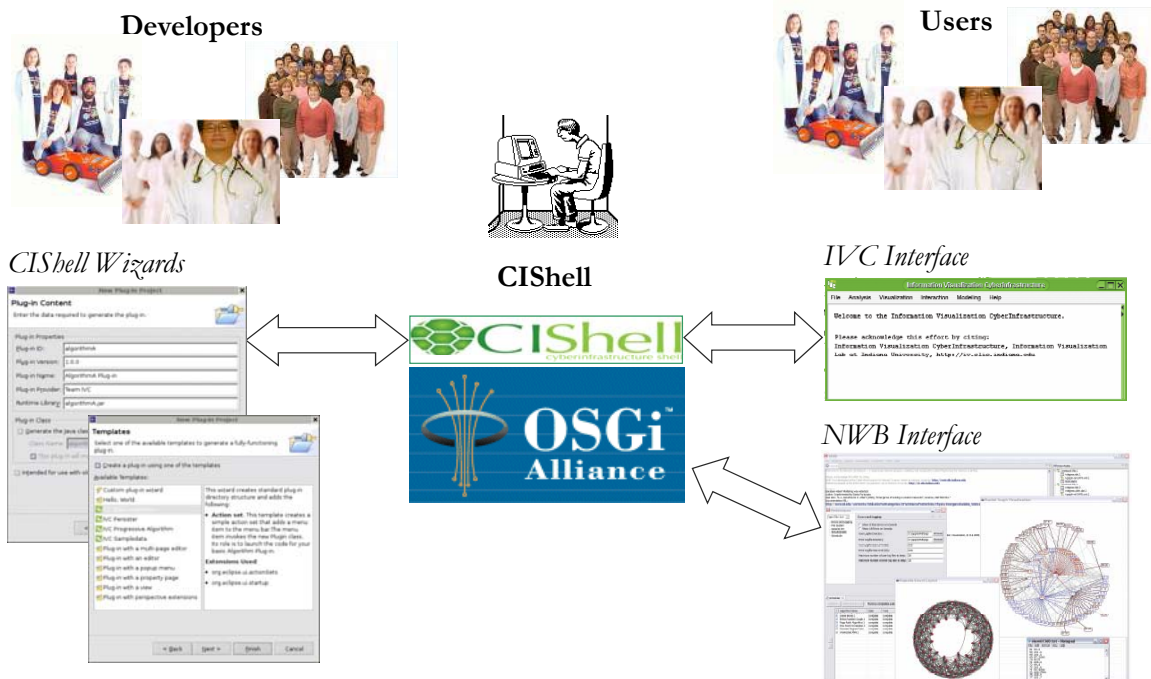
Website: <http://nwb.slis.indiana.edu>

NWB Advisory Board:

- James Hendler (Semantic Web) <http://www.cs.umd.edu/~hendler/>
- Jason Leigh (CI) <http://www.evl.uic.edu/spiff/>
- Neo Martinez (Biology) <http://online.sfsu.edu/~webhead/>
- Michael Macy, Cornell University (Sociology) <http://www.soc.cornell.edu/faculty/macy.shtml>
- Ulrik Brandes (Graph Theory) <http://www.inf.uni-konstanz.de/~brandes/>
- Mark Gerstein, Yale University (Bioinformatics) <http://bioinfo.mbb.yale.edu/>
- Stephen North (AT&T) <http://public.research.att.com/viewPage.cfm?PageID=81>
- Tom Snijders, University of Groningen <http://stat.gamma.rug.nl/snijders/>
- Noshir Contractor, Northwestern University <http://www.spcomm.uiuc.edu/nosh/>



CIShell – Serving Non-CS Algorithm Developers & Users





CIShell – Builds on OSGi Industry Standard

CIShell is built upon the Open Services Gateway Initiative (OSGi) Framework.

OSGi (<http://www.osgi.org>) is

- A standardized, component oriented, computing environment for networked services.
- Successfully used in the industry from high-end servers to embedded mobile devices since 8 years.
- Alliance members include IBM (Eclipse), Sun, Intel, Oracle, Motorola, NEC and many others.
- Widely adopted in open source realm, especially since Eclipse 3.0 that uses OSGi R4 for its plugin model.

Advantages of Using OSGi

- Any CIShell algorithm is a service that can be used in any OSGi-framework based system.
- Using OSGi, running CIShells/tools can be connected via RPC/RMI supporting peer-to-peer sharing of data, algorithms, and computing power.

Ideally, CIShell becomes a standard for creating OSGi Services for algorithms.



NWB Tool: Supported Data Formats

Personal Bibliographies

- Bibtext (.bib)
- Endnote Export Format (.enw)

Data Providers

- Web of Science by Thomson Scientific/Reuters (.isi)
- Scopus by Elsevier (.scopus)
- Google Scholar (access via *Publish or Perish* save as CSV, Bibtext, EndNote)
- Awards Search by National Science Foundation (.nsf)

Scholarly Database (all text files are saved as .csv)

- Medline publications by National Library of Medicine
- NIH funding awards by the National Institutes of Health (NIH)
- NSF funding awards by the National Science Foundation (NSF)
- U.S. patents by the United States Patent and Trademark Office (USPTO)
- Medline papers – NIH Funding

Network Formats

- NWB (.nwb)
- Pajek (.net)
- GraphML (.xml or .graphml)
- XGMML (.xml)

Burst Analysis Format

- Burst (.burst)

Other Formats

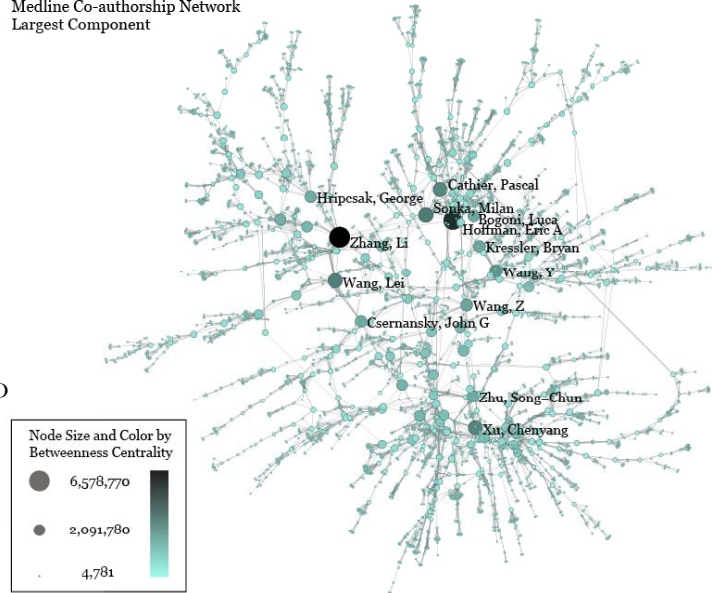
- CSV (.csv)
- Edgelist (.edge)
- Pajek (.mat)
- TreeML (.xml)

- NWB tool can be used for data conversion. Supported output formats comprise:
 - CSV (.csv)
 - NWB (.nwb)
 - Pajek (.net)
 - Pajek (.mat)
 - GraphML (.xml or .graphml)
 - XGMML (.xml)

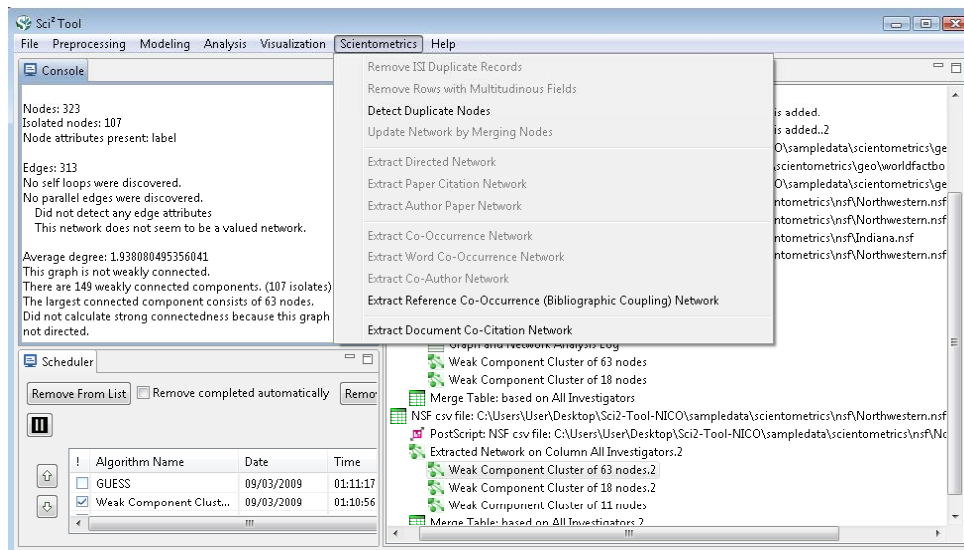
- GUESS
 - Supports export of images into common image file formats.

- Horizontal Bar Graphs
- saves out raster and ps files.

Medline Co-authorship Network
Largest Component



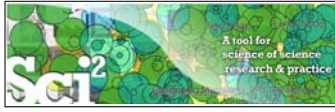
Sci² Tool for Science of Science Research and Practice



Acknowledgments

This work is supported in part by the Cyberinfrastructure for Network Science center and the School of Library and Information Science at Indiana University, the National Science Foundation under Grant No. SBE-0738111 and IIS-0513650, and the James S. McDonnell Foundation.





Sci² Tool: Algorithms

See <https://nwb.slis.indiana.edu/community>

Preprocessing

Extract Top N% Records
Extract Top N Records
Normalize Text
Slice Table by Line

Extract Top Nodes
Extract Nodes Above or Below Value
Delete Isolates

Extract top Edges
Extract Edges Above or Below Value
Remove Self Loops
Trim by Degree
MST-Pathfinder Network Scaling
Fast Pathfinder Network Scaling

Snowball Sampling (in nodes)
Node Sampling
Edge Sampling

Symmetrize
Dichotomize
Multipartite Joining

Geocoder

Extract ZIP Code

Modeling

Random Graph
Watts-Strogatz
Small World
Barabási-Albert Scale-Free
TARL

Analysis

Network Analysis Toolkit (NAT)
Unweighted & Undirected

Node Degree
Degree Distribution

K-Nearest Neighbor (Java)
Watts-Strogatz Clustering Coefficient
Watts Strogatz Clustering Coefficient over K

Diameter
Average Shortest Path
Shortest Path Distribution
Node Betweenness Centrality

Weak Component Clustering
Global Connected Components

Extract K-Core
Annotate K-Coreeness

HTTS

Weighted & Undirected

Clustering Coefficient
Nearest Neighbor Degree
Strength vs Degree
Degree & Strength
Average Weight vs End-point Degree
Strength Distribution
Weight Distribution
Randomize Weights

Blondel Community Detection

HTTS

Unweighted & Directed

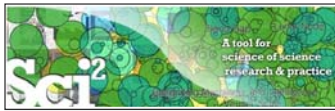
Node Indegree
Node Outdegree
Indegree Distribution
Outdegree Distribution

K-Nearest Neighbor
Single Node in-Out Degree Correlations

Dyad Reciprocity
Arc Reciprocity
Adjacency Transitivity

Weak Component Clustering
Strong Component Clustering

47



Sci² Tool: Algorithms cont.

See <https://nwb.slis.indiana.edu/community>

Extract K-Core
Annotate K-Coreeness

HTTS
PageRank
Weighted & Directed
HTTS
Weighted PageRank

Textual

Burst Detection

Visualization

GnuPlot
GUESS
Image Viewer

Radial Tree/Graph (prefuse alpha)
Radial Tree/Graph with Annotation
(prefuse beta)
Tree View (prefuse beta)
Tree Map (prefuse beta)
Force Directed with Annotation
(prefuse beta)
Fruchterman-Reingold with Annotation
(prefuse beta)

DrL (VxOrd)
Specified (prefuse beta)

Horizontal Line Graph

Circular Hierarchy

Geo Map (Circle Annotation Style)

Geo Map (Colored-Region Annotation Style)

*Science Map (Circle Annotation)

Scientometrics

Remove ISI Duplicate Records
Remove Rows with Multitudinous Fields
Detect Duplicate Nodes
Update Network by Merging Nodes

Extract Directed Network

Extract Paper Citation Network
Extract Author Paper Network

Extract Co-Occurrence Network

Extract Word Co-Occurrence Network
Extract Co-Author Network
Extract Reference Co-Occurrence
(Bibliographic Coupling) Network

Extract Document Co-Citation Network

* Requires permission from UCSD
All four+ save into Postscript files.

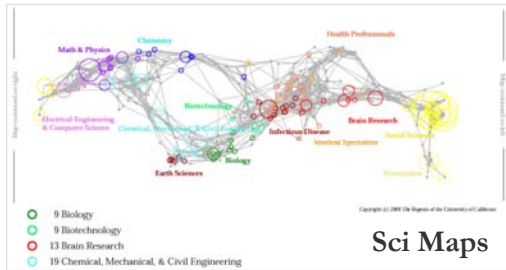
[General Network extraction](#)

48

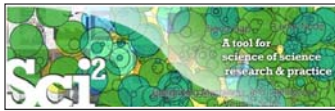


Sci² Tool

Plugins that render into Postscript files:



Börner, Katy, Huang, Weixia (Bonnie), Linnemeier, Micah, Dubon, Russell Jackson, Phillips, Patrick, Ma, Nianli, Zoss, Angela, Guo, Hanning & Price, Mark. (2009). *Rete-Netzwerk-Red: Analyzing and Visualizing Scholarly Networks Using the Scholarly Database and the Network Workbench Tool*. *Proceedings of ISSI 2009: 12th International Conference on Scientometrics and Informetrics, Rio de Janeiro, Brazil, July 14-17. Vol. 2, pp. 619-630.*



Temporal: Horizontal Bargraphs

Horizontal Line Graph

Takes tabular data and generates PostScript for a horizontal line graph.

Label: Full Project Number (including subproject ID)

Start Date: Project start date

End Date: Project end date

Size By: FY Total Costs

Date Format: Month-Day-Year Date Format (U.S., e.g. 10/15/2010)

Page Width: 8.5

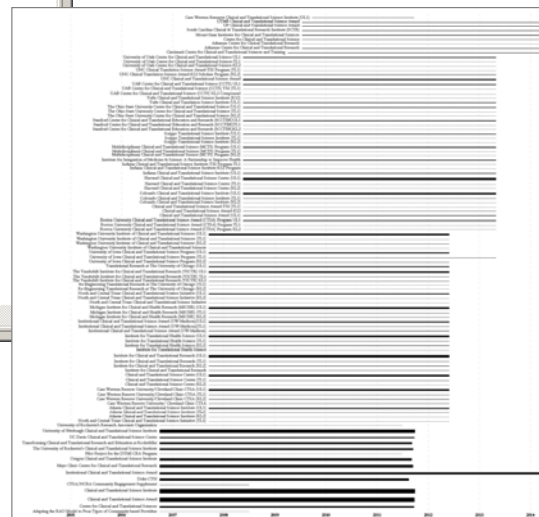
Page Height: 11.0


Scale Output?

Area size equals numerical value, e.g., award amount.

Text, e.g., title []

Start date End date





Science of Science Cyberinfrastructure
— P O R T A L —

Provided by the [Cyberinfrastructure for Network Science Center](#) at Indiana University.

Introduction
E. O. Wilson writes in *Consilience: The Unity of Knowledge* (1998): "Features that distinguish science from pseudoscience are repeatability, economy, mensuration, heuristics, and consilience." Please see Börner's [recent presentation](#) at the *A Deeper Look at the Visualization of Scientific Discovery* NSF Workshop for a general introduction of the needs and the resources provided here.

Needs Analysis
As part of the "TLS: Towards a Macroscopic for Science Policy Decision Making" NSF SBE-0738111 award, interviews with science policy makers are conducted to identify what science of science research results and tools might be most desirable and effective. So far, 30 formal, one-hour interviews have been conducted with science policy makers at university campus level, program officer level, and division director level for governmental, state, and private foundations. Data compilation will start in October 2008 and resulting report can be ordered by sending a request to Mark Price (maaprice@indiana.edu).

Conceptualization of Science
A 'science of science' requires a theoretically grounded and practically useful conceptualization of the structure and evolution of science. A special journal issue entitled "Science of Science: Conceptualizations and Models of Science" edited by [Katy Börner](#), Indiana University & [Andrea Scharnhorst](#), Royal Netherlands Academy of Arts and Sciences invites contributions on this topic. It will be published in the *Journal of Informetrics* 3(1) in January 2009.

Scholarly Database
The [Scholarly Database \(SDB\)](#) at Indiana University aims to serve researchers and practitioners interested in the analysis, modeling, and visualization of large-scale scholarly datasets. The database currently provides access to over 20 million papers, patents and grants. Resulting datasets can be downloaded in bulk. Register for free access at <https://sdb.slis.indiana.edu/>.

Cyberinfrastructures
The Scientometrics filling of the [Network Workbench \(NWB\) Tool](#) provides a unique distributed, shared resources environment for large-scale network analysis, modeling, and visualization. Thomson Scientific/ISI, Scopus and Google Scholar data, EndNote and Bibtext files, or NSF awards can be read and diverse networks can be extracted and studied. Download [User Manual with focus on Scientometrics](#).

<http://sci.slis.indiana.edu>

51



OSGi/CIShell Adoption

CIShell/OSGi is at the core of different CIs and a total of 169 unique plugins are used in the

- **Information Visualization** (<http://iv.slis.indiana.edu>),
- **Network Science (NWB Tool)** (<http://nwb.slis.indiana.edu>),
- **Scientometrics and Science Policy (Sc² Tool)** (<http://sci.slis.indiana.edu>), and
- **Epidemics** (<http://epic.slis.indiana.edu>) research communities.

Most interestingly, a number of other projects recently adopted OSGi and one adopted CIShell:

Cytoscape (<http://www.cytoscape.org>) lead by Trey Ideker, UCSD is an open source bioinformatics software platform for visualizing molecular interaction networks and integrating these interactions with gene expression profiles and other state data (Shannon et al., 2002).

Taverna Workbench (<http://taverna.sourceforge.net>) lead by Carol Goble, University of Manchester, UK is a free software tool for designing and executing workflows (Hull et al., 2006). Taverna allows users to integrate many different software tools, including over 30,000 web services.

MAEviz (<https://wiki.ncsa.uiuc.edu/display/MAE/Home>) managed by Shawn Hampton, NCSA is an open-source, extensible software platform which supports seismic risk assessment based on the Mid-America Earthquake (MAE) Center research.

TEXTrend (<http://www.textrend.org>) lead by George Kampis, Eötvös University, Hungary develops a framework for the easy and flexible integration, configuration, and extension of plugin-based components in support of natural language processing (NLP), classification/mining, and graph algorithms for the analysis of business and governmental text corpuses with an inherently temporal component.

As the functionality of OSGi-based software frameworks improves and the number and diversity of dataset and algorithm plugins increases, the capabilities of custom tools will expand.

Mapping Science Exhibit – 10 Iterations in 10 years

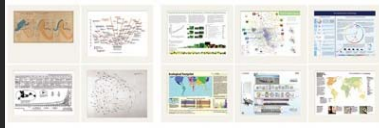
<http://scimaps.org>



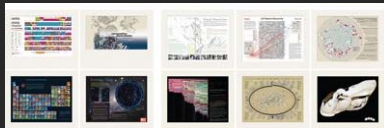
The Power of Maps (2005)



Science Maps for Economic Decision Makers (2008)



The Power of Reference Systems (2006)



Science Maps for Science Policy Makers (2009)

Science Maps for Scholars (2010)

Science Maps as Visual Interfaces to Digital Libraries (2011)

Science Maps for Kids (2012)

Science Forecasts (2013)

The Power of Forecasts (2007)



How to Lie with Science Maps (2014)



Exhibit has been shown in 72 venues on four continents. Currently at
 - NSF, 10th Floor, 4201 Wilson Boulevard, Arlington, VA
 - Wallenberg Hall, Stanford University, CA
 - Center of Advanced European Studies and Research, Bonn, Germany
 - Science Train, Germany.



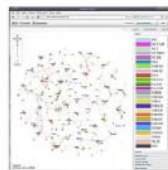
53

A Topic Map of NIH Grants 2007

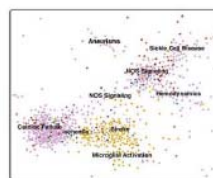
ChalkLabs UCI UCIRVINE

Bruce W. Herr II (ChalkLabs & IU), Gully Burns (IS), David Newman (UCI), Edmund Talley (NIH)

The National Institutes of Health (NIH) is organized as a multitude of institutes and Centers whose missions are primarily focused on distinct diseases. However, disease etiologies and therapies blur scientific boundaries, and thus there is tremendous overlap in the kinds of research funded by each institute. This creates a daunting landscape for decisions on research directions, funding allocations, and policy formulations. Shown here is devised an interactive topic map for navigating this landscape online at www.scimaps.org. Institute abbreviations can be found at www.nih.gov/od.



Topic modeling, a statistical technique that automatically learns semantic categories, was applied to assess projects in terms used by researchers to describe their work, without the biases of keywords or subject headings. Grant similarities were derived from their topic mixtures, and grants were then clustered on a two-dimensional map using a force-directed simulated annealing algorithm. This analysis creates an interactive environment for assessing grant relevance to research categories and to NIH institutes in which grants are localized.



Cardiac Diseases Research

An area of the map focused on cardiovascular function and dysfunction. Cardiac failure is primarily funded by NIDDK and is typically clustered next to Stroke (NINDS) since these are the two major vascular emergencies associated with ischemia, which results from unobstructed blood supply. Also localized in this area are grants focused on Microglial Activation.

Neural Circuits Research

An area of the map focused on neural circuits, which show the diversity of topics and NIH institutes that fund research in this area, such as Cardiovascular Regulation, primarily funded by NIDDK. Visual Processing, primarily funded by NEI and Splicing, primarily funded by NINDS. For color coding, see legend on the upper left inset.

Cerebral Pathways

An area of the map focused on cerebral pathways, which show the diversity of topics and NIH institutes that fund research in this area, such as Cardiovascular Regulation, primarily funded by NIDDK. Visual Processing, primarily funded by NEI and Splicing, primarily funded by NINDS. For color coding, see legend on the upper left inset.

National Cancer Institute (NCI)

- TOP 10 TOPICS
1. Oncology Clinical Trials
 2. Cancer Treatment
 3. Cancer Therapy
 4. Carcinogenesis
 5. Risk Factor Analysis
 6. Cancer Chemotherapy
 7. Metastasis
 8. Leukemia
 9. Prostatectomy
 10. Cancer Chemoprevention

National Institute of General Medical Sciences (NIGMS)

- TOP 10 TOPICS
1. Bioactive Organic Synthesis
 2. X-ray Crystallography
 3. Protein-DB
 4. Computational Models
 5. Yeast Biology
 6. Microbiology
 7. Enzymes/Mechanisms
 8. Protein Complexes
 9. Invertebrate/Zebrafish Genetics
 10. Cell Division

National Heart, Lung, and Blood Institute (NHLBI)

- TOP 10 TOPICS
1. Cardiac Failure
 2. Pulmonary Injury
 3. Genetic Linkage Analysis
 4. Cardiovascular Disease
 5. Atherosclerosis
 6. Hemostasis
 7. Blood Pressure
 8. Arteriosclerosis, Atherosclerosis
 9. Gene Association
 10. Lipoproteins

National Institute of Mental Health (NIMH)

- TOP 10 TOPICS
1. Mood Disorders
 2. Schizophrenia
 3. Behavioral Intervention Studies
 4. Mental Health
 5. Depression
 6. Cognitive Behavior Therapy
 7. PTSD Prevention
 8. Genetic Linkage Analysis
 9. Adolescence
 10. Childhood

Herr II, Bruce W., Gully Burns, David Newman, Edmund Talley. 2007.

A Topic Map of NIH Grants 2007. Bloomington, IN.

54

Happiness Depends on Various Factors

Social scientists are starting to include relative happiness with hard data on economic status, health, and other factors as they assess quality of life. They rely on surveys of "subjective well-being"—how good people feel about their lives. A world map of one "happiness index" shows many, but not all, wealthy northern countries faring well. Residents of sub-Saharan Africa and the former Soviet Union, meanwhile, report particularly low levels of contentment.

Any attempt to measure happiness will fall short—each life is a series of joys, struggles, and sorrows, and satisfaction can depend as much on outlook as on circumstances. Averages obscure the happy moments in struggling nations, as well as people who suffer from poor health, poverty, or discrimination in countries that rank high. Still, happiness indices can help researchers move beyond simple economics as they track progress—or backsliding—over time.

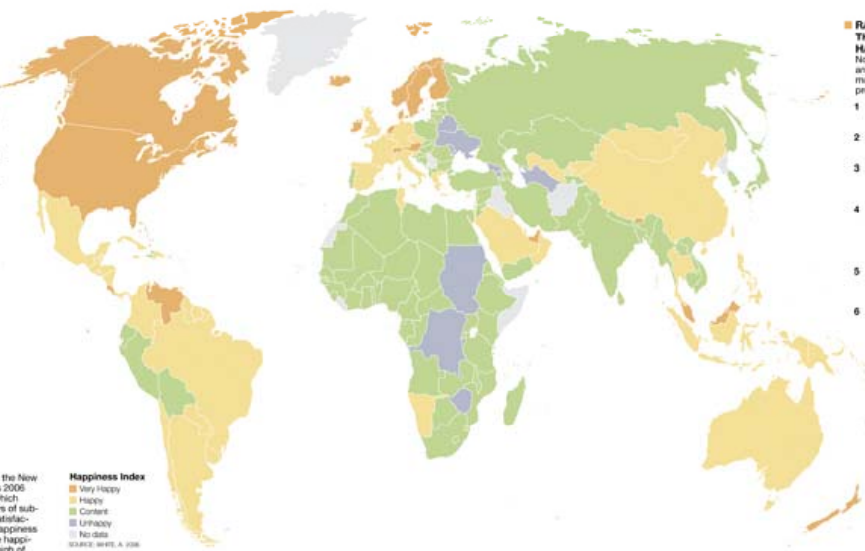
MEASURING THE INTANGIBLE

The map is derived from the New Economics Foundation's 2006 "Happy Planet Index," which drew on over 100 surveys of subjective well-being. Its "satisfaction with life scale"—a happiness index—ranks the relative happiness of nations, from a high of 273 (Denmark and Switzerland) to a low of 100 (Burundi).

Happiness Index

Very happy
Happy
Content
Unhappy
Sad

SOURCE: NEW ECON. FOUN.



RANKING THE WORLD'S HAPPIEST PLACES

Northern Europe, North America, and several wealthy countries make the list, but so do many less prosperous island nations.

- 1 DENMARK SWITZERLAND
- 2 AUSTRIA ICELAND
- 3 BAHAMAS FINLAND SWEDEN
- 4 BHUTAN BRUNEI CANADA IRELAND LUXEMBOURG
- 5 COSTA RICA MALTA NETHERLANDS
- 6 ANTIQUA AND BARBUDA MALAYSIA NEW ZEALAND NORWAY SEYCHELLES ST. KITTS AND NEVIS UNITED ARAB EMIRATES UNITED STATES VANUATU VENEZUELA

DEFINING WELL-BEING

By comparing the happiness index to data from the UN, the CIA, and other sources, a U.K. psychologist determined that good health and health care, enough money for fundamental needs, and access to basic education are the most important factors for subjective well-being. European countries top all three measures.



HEALTH

Japan boasts the world's longest life expectancy—one measure of overall health. Swaziland, at the other end of the scale, is plagued by poverty, disease, and violence. Disparities in access to health care divide many countries into haves and have-nots.



WEALTH

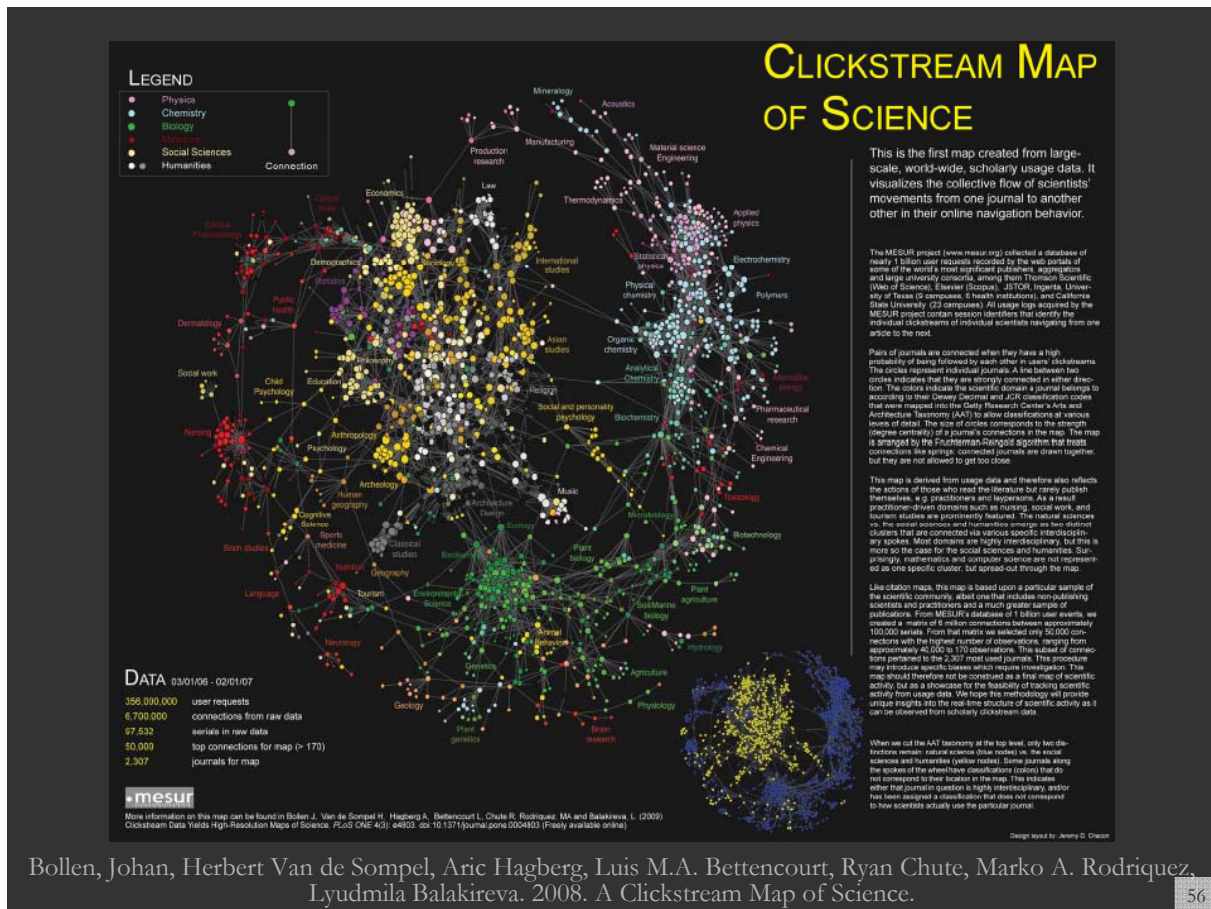
Money still can't buy love, or happiness, and wealthier people aren't always more content. Still, tiny Luxembourg, which takes top rank in per capita Gross Domestic Product (GDP), also rates a 253 on the happiness index. Real poverty means real misery, a fate shared by billions.



EDUCATION

Residents of Australia can expect to spend more time in school—an average of almost 21 years—than citizens of any other country. But only a basic education is needed to see a significant jump in overall happiness. Around the world, hundreds of millions lack even that.

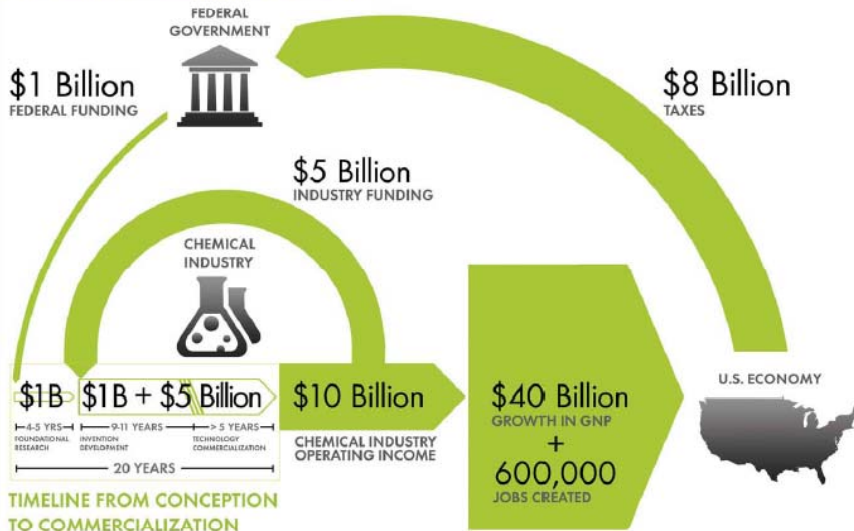
PHOTOS: (TOP) THE GLOBE PHOTOS; (MIDDLE) GETTY IMAGES; (BOTTOM) SHUTTERSTOCK



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

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

INVESTMENT IN CHEMICAL SCIENCE R&D



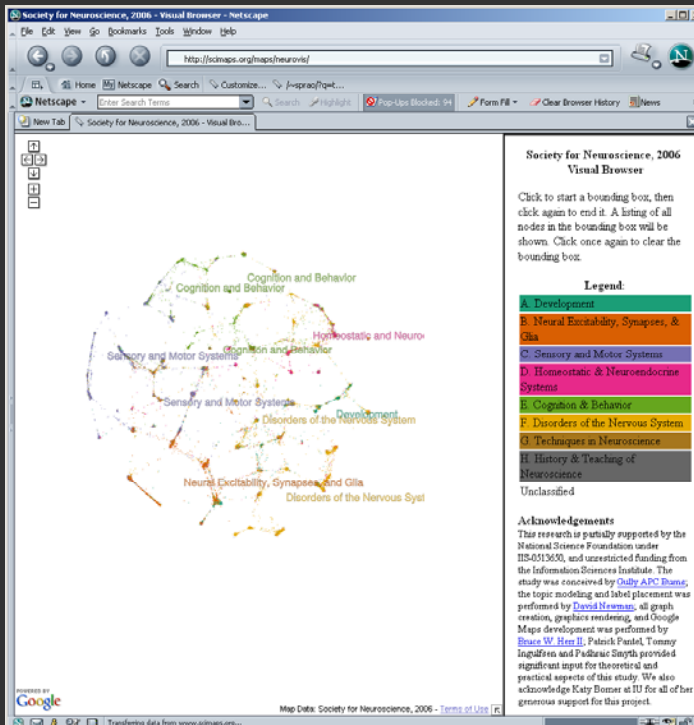
The Council for Chemical Research (CCR)

has provided the U.S. Congress and government policy makers with important results regarding the impact of Federal Research & Development (R&D) investments on U.S. innovation and global competitiveness through its commissioned 5-year two phase study. To take full advantage of typically brief access to policy makers, CCR developed the graphic below as a communication tool that distills the complex data produced by these studies in direct, concise and clear terms.

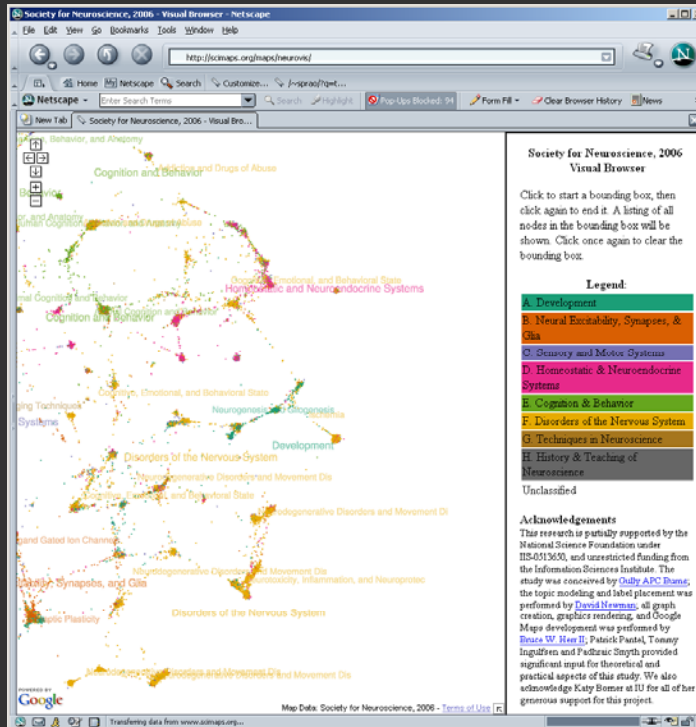


The design shows that an input of \$1B in federal investment, leveraged by \$5B industry investment, brings new technologies to market and results in \$10B of operating income for the chemical industry, \$40B growth in the Gross National Product (GNP) and further impacts the US economy by generating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

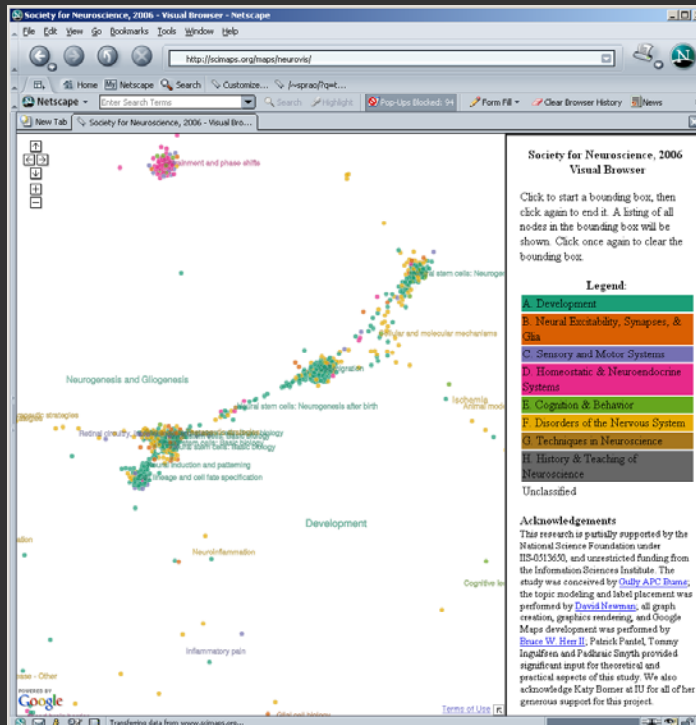
Council for Chemical Research. 2009. Chemical R&D Powers the U.S. Innovation Engine. Washington, DC. Courtesy of the Council for Chemical Research.



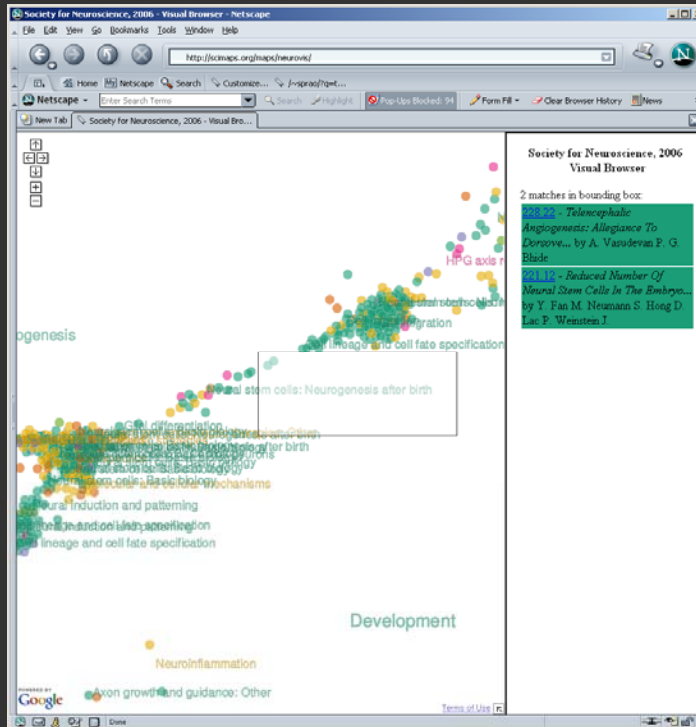
Bruce W. Herr II, Gully Burns (USC), David Newman (UCI), Society for Neuroscience, 2006
Visual Browser, 2007, <http://scimaps.org/maps/neurovis/>



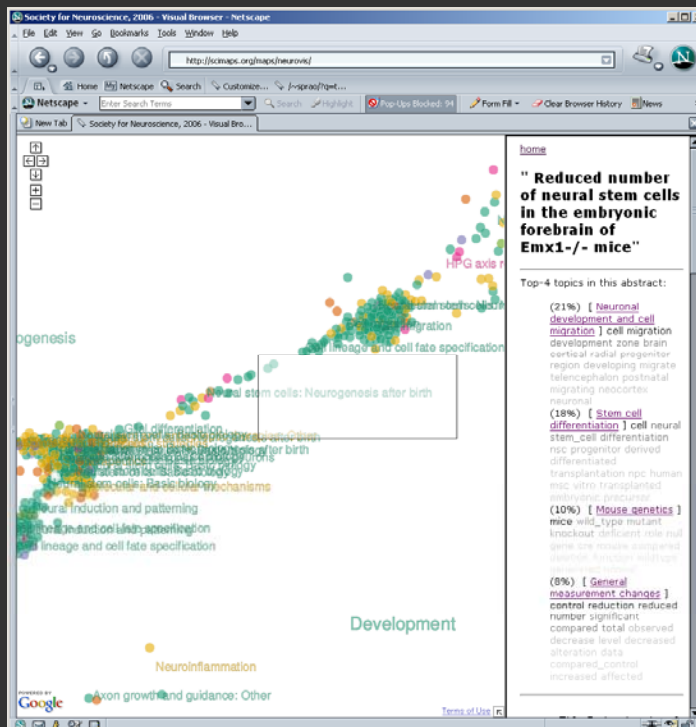
Bruce W. Herr II, Gully Burns (USC), David Newman (UCI), Society for Neuroscience, 2006 Visual Browser, 2007, <http://scimaps.org/maps/neurovis/>



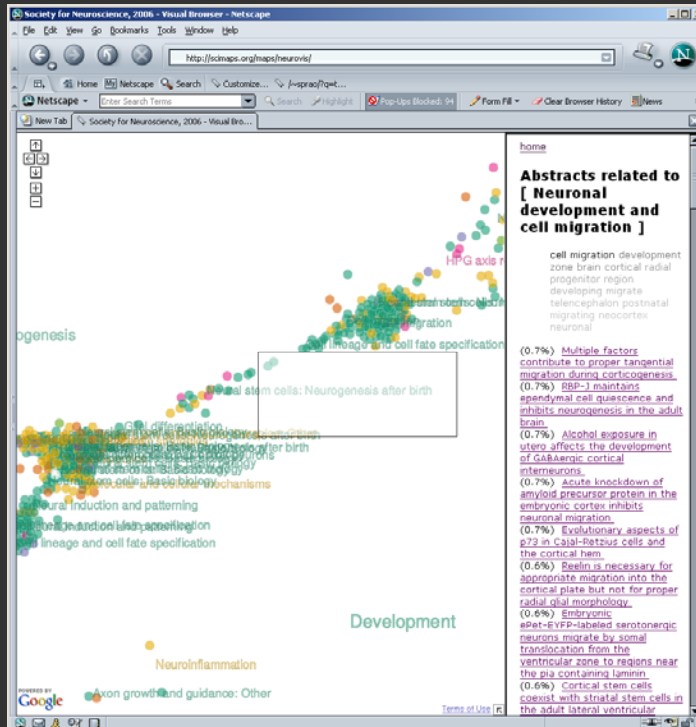
Bruce W. Herr II, Gully Burns (USC), David Newman (UCI), Society for Neuroscience, 2006 Visual Browser, 2007, <http://scimaps.org/maps/neurovis/>



Bruce W. Herr II, Gully Burns (USC), David Newman (UCI), Society for Neuroscience, 2006
 Visual Browser, 2007, <http://scimaps.org/maps/neurovis/>



Bruce W. Herr II, Gully Burns (USC), David Newman (UCI), Society for Neuroscience, 2006
 Visual Browser, 2007, <http://scimaps.org/maps/neurovis/>



Bruce W. Herr II, Gully Burns (USC), David Newman (UCI), Society for Neuroscience, 2006
 Visual Browser, 2007, <http://scimaps.org/maps/neurovis/>

Illuminated Diagram Display

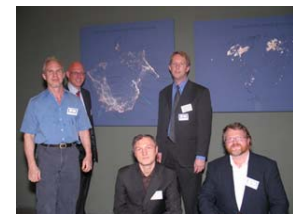
W. Bradford Paley, Kevin W. Boyack, Richard Kalvans, and Katy Börner (2007)
Mapping, Illuminating, and Interacting with Science. SIGGRAPH 2007.

Questions:

- Who is doing research on what topic and where?
- What is the 'footprint' of interdisciplinary research fields?
- What impact have scientists?

Contributions:

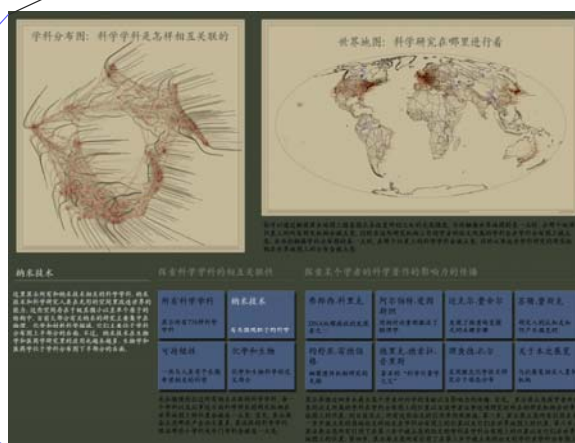
- Interactive, high resolution interface to access and make sense of data about scholarly activity.



Large-scale, high resolution prints illuminated via projector or screen.



Interactive touch panel.





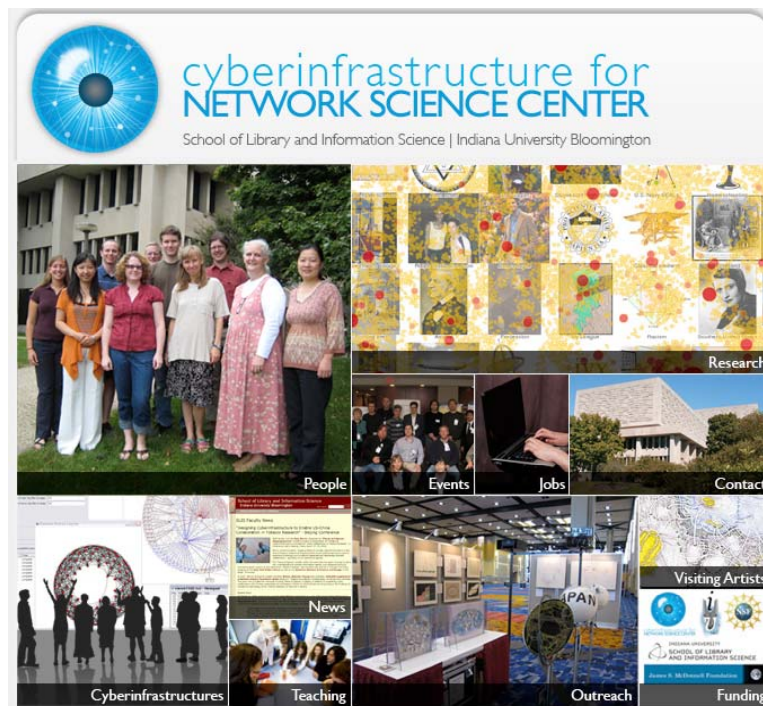
Debut of 5th Iteration of Mapping Science Exhibit at MEDIA X was on May 18, 2009 at Wallenberg Hall, Stanford University, <http://mediax.stanford.edu>, <http://scaleindependentthought.typepad.com/photos/scimaps>



Science Maps in "Expedition Zukunft" science train visiting 62 cities in 7 months
 12 coaches, 300 m long
 Opening was on April 23rd, 2009 by German Chancellor Merkel
<http://www.expedition-zukunft.de>

This is the only mockup in this slide show.

Everything else is available today.



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