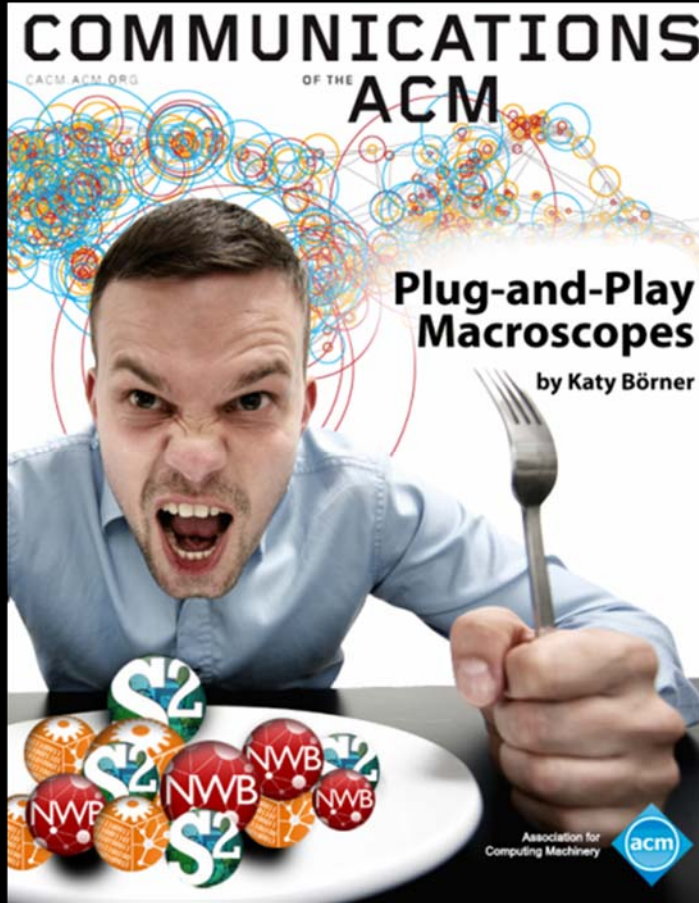


Börner, Katy. (March 2011). Plug-and-Play Macroscopes.
Communications of the ACM, 54(3), 60-69.
 Video and paper are at <http://www.scivee.in/node/27704>

Taming Complexity
 TTI VANGUARD
 Conference in D.C.
 2011.10.04



CNS.IU.EDU
 CISbell.org
 SciMaps.org



Take terra bytes of data

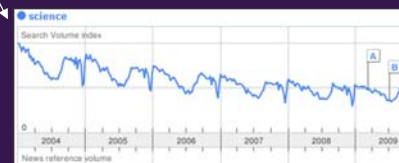
Plug-and-Play
 Macroscopes



Find your way



Find collaborators, friends



Identify trends

Macroscopes

Decision making in science, industry, and politics, as well as in daily life, requires that we make sense of data sets representing the structure and dynamics of complex systems. Macroscopes provide a “vision of the whole,” helping us “synthesize” the related elements and enabling us to detect patterns, trends, and outliers while granting access to myriad details. Rather than make things larger or smaller, **macroscopes let us observe what is at once too great, slow, or complex for the human eye and mind to notice and comprehend.**



Microscopes



Telescopes



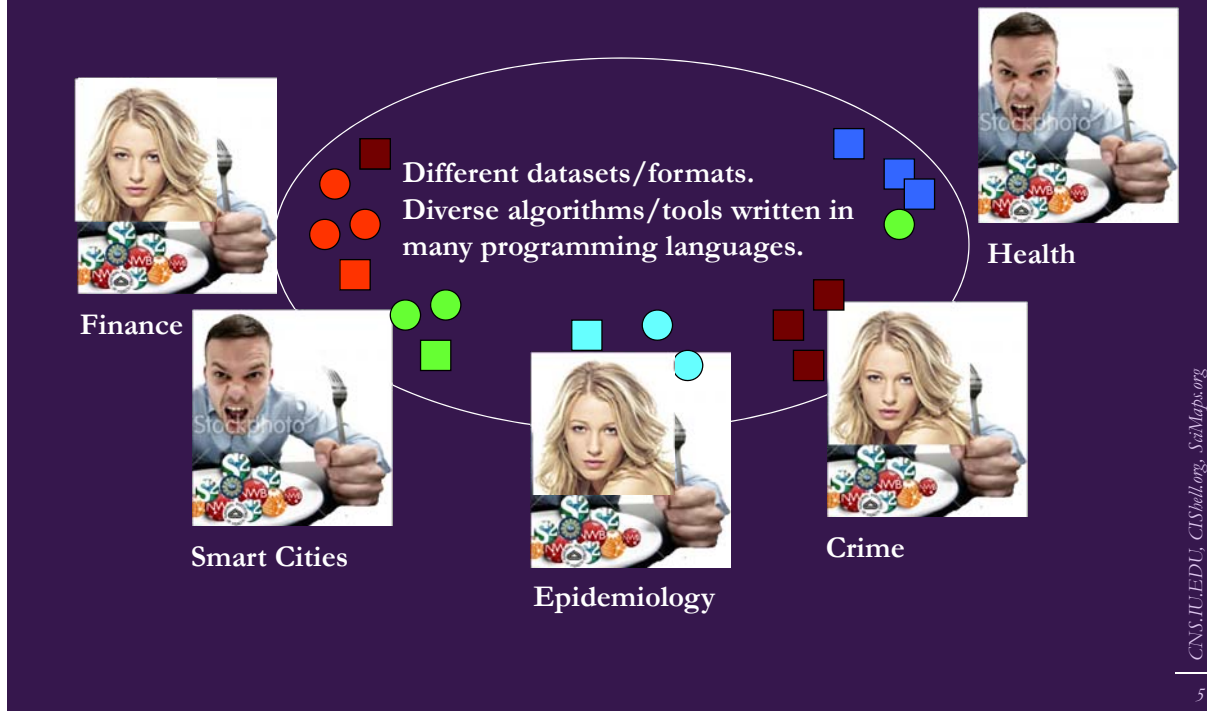
Macroscopes

Plug-and-Play Macroscopes

While microscopes and telescopes are **physical instruments**, macroscopes resemble **continuously changing bundles of software plug-ins**. Macroscopes make it easy to select and combine algorithm and tool plug-ins but also interface plug-ins, workflow support, logging, scheduling, and other plug-ins needed for scientifically rigorous yet effective work.

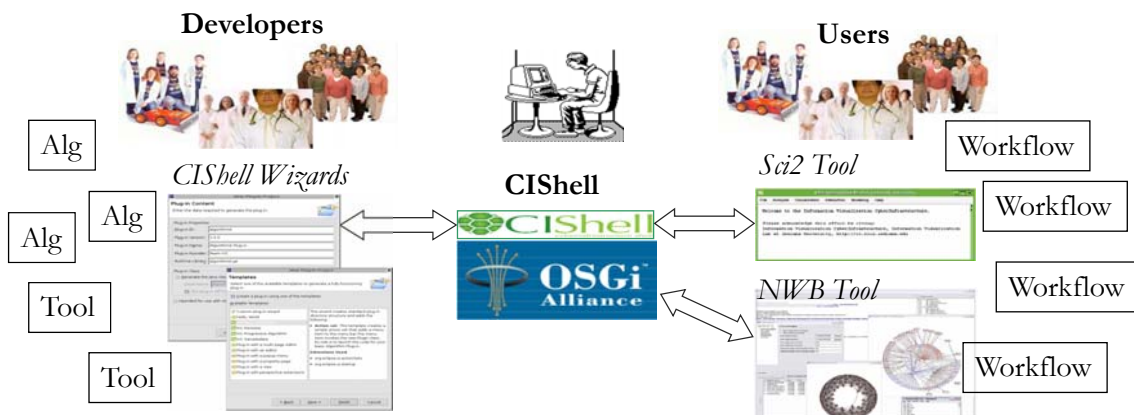
They make it easy to share plug-ins via email, flash drives, or online. To use new plugins, simply copy the files into the plug-in directory, and they appear in the tool menu ready for use. No restart of the tool is necessary. **Sharing algorithm components, tools, or novel interfaces becomes as easy as sharing images on Flickr or videos on YouTube. Assembling custom tools is as quick as compiling your custom music collection.**

Plug-and-Play Macroscopes



OSGi & Cyberinfrastructure Shell (CISHell)

- CISHell (<http://cishell.org>) is an open source software specification for the integration and utilization of datasets, algorithms, and tools.
- It extends the Open Services Gateway Initiative (OSGi) (<http://osgi.org>), a standardized, component oriented, computing environment for networked services widely used in industry since more than 10 years.
- Specifically, CISHell provides “sockets” into which existing and new datasets, algorithms, and tools can be plugged using a wizard-driven process.



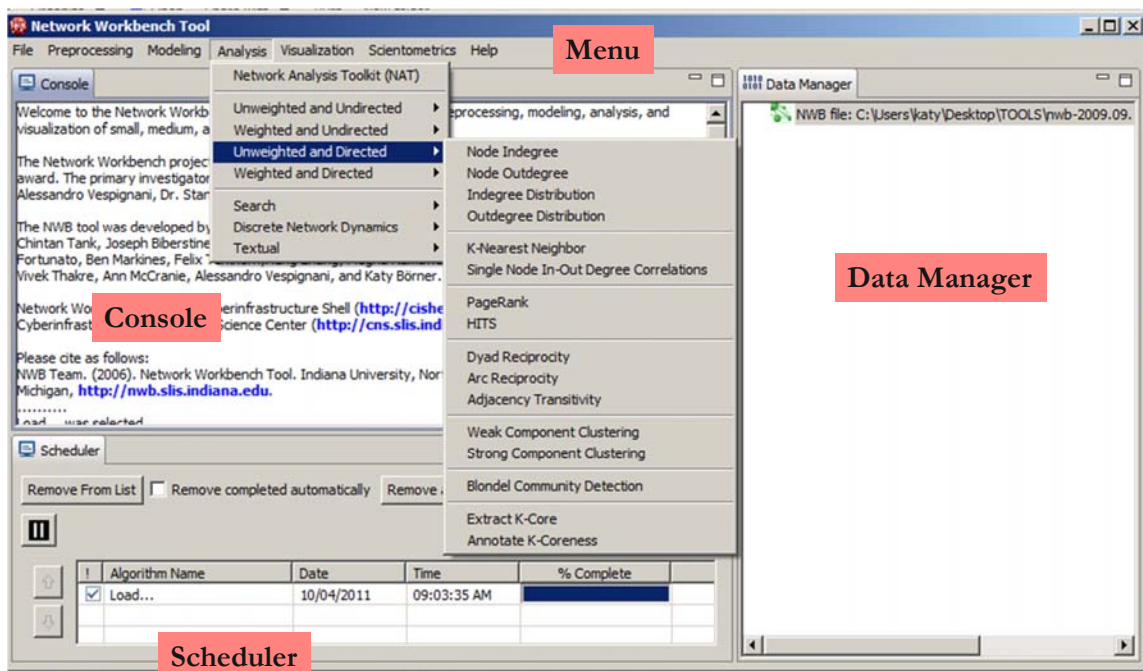
The Network Workbench (NWB) tool supports researchers, educators, and practitioners interested in the study of biomedical, social and behavioral science, physics, and other networks.

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

It has been downloaded more than 110,000 times.



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, Mihail C. (Eds.), *Progress in Convergence - Technologies for Human Wellbeing (Vol. 1093, pp. 161-179)*, Annals of the New York Academy of Sciences, Boston, MA.



Computational Proteomics

What relationships exist between protein targets of all drugs and all disease-gene products in the human protein–protein interaction network?

Yildirim, Muhammed A., Kwan-II Goh, Michael E. Cusick, Albert-László Barabási, and Marc Vidal. (2007). Drug-target Network. Nature Biotechnology 25 no. 10: 1119-1126.



Figure 2 Drug-target network (DT network). The DT network is generated by using the known associations between FDA-approved drugs and their target proteins. Circles and rectangles correspond to drugs and target proteins, respectively. A link is placed between a drug node and a target node if the protein is a known target of that drug. The area of the drug (protein) node is proportional to the number of targets that the drug (the number of drugs targeting the protein). Color codes are given in the legend. Drug nodes (circles) are colored according to their Anatomical Therapeutic Chemical Classification, and the target proteins (rectangular boxes) are colored according to their cellular component obtained from the Gene Ontology database.

Computational Economics

Does the type of product that a country exports matter for subsequent economic performance?

C. A. Hidalgo, B. Klinger, A.-L. Barabási, R. Hausmann (2007) The Product Space Conditions the Development of Nations. Science 317, 482 (2007).

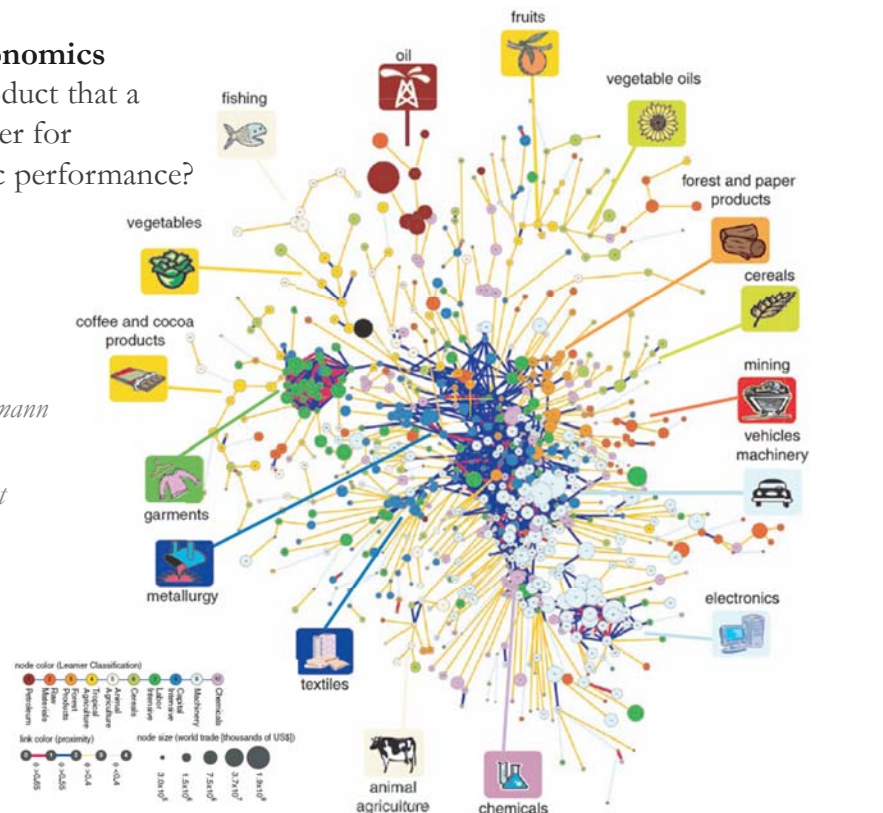


Fig. 1. The product space. (A) Hierarchically clustered proximity matrix representing the 775 SITC-4 product classes exported in the 1998–2000 period. (B) Network representation of the product space. Links are color coded with their proximity value. The sizes of the nodes are proportional to world trade, and their colors are chosen according to the classification introduced by Leamer.

Computational Social Science

Studying large scale social networks such as Wikipedia

Second Sight: An Emergent Mosaic of Wikipedian Activity, The NewScientist, May 19, 2007

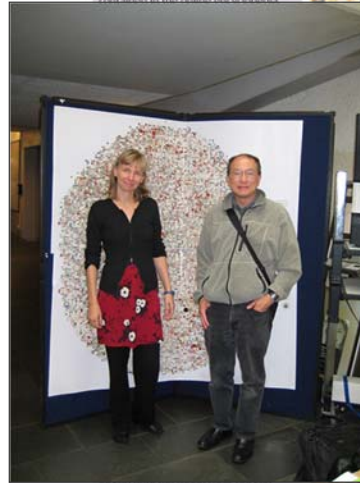


Second sight

Image: Bruce W. Hest and Todd M. Holloway

Power struggle

How do you keep track of the bobbling mass of information that is Wikipedia? This chaotic-looking mosaic is one attempt to show which topics are contained in the online encyclopedia.



...linked with the most-viewed pages at the time of writing include entries on Sheffield Wednesday football club, Mikhail Gorbachev and pigs). The mosaic has been commended in a competition for images that visualise network dynamics, coinciding with this week's International Workshop and Conference on Network Science in Bloomington.

www.newscientist.com



19 May 2007 | NewScientist | 11

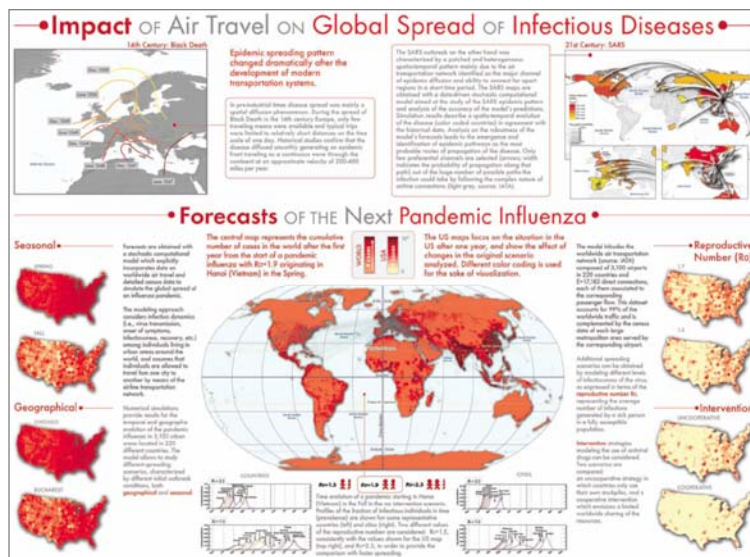
Computational Epidemics

Forecasting (and preventing the effects of) the next pandemic.

Epidemic Modeling in Complex realities, V. Colizza, A. Barrat, M. Barthelemy, A. Vespignani, Comptes Rendus Biologie, 330, 364-374 (2007).

Reaction-diffusion processes and metapopulation models in heterogeneous networks, V. Colizza, R. Pastor-Satorras, A. Vespignani, Nature Physics 3, 276-282 (2007).

Modeling the Worldwide Spread of Pandemic Influenza: Baseline Case and Containment Interventions, V. Colizza, A. Barrat, M. Barthelemy, A.-J. Valleron, A. Vespignani, PLoS-Medicine 4, e13, 95-110 (2007).





Science of Science Tool

<http://sci2.cns.iu.edu>

Sci² Tool v0.5.1 Alpha (May 4th, 2011)

Can be freely downloaded for all major operating systems from

<http://sci2.cns.iu.edu>

Select your operating system from the pull down menu and download.

Unpack into a /sci2 directory.

Run /sci2/sci2.exe

Sci² Manual is at

<http://sci2.wiki.cns.iu.edu>

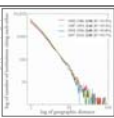


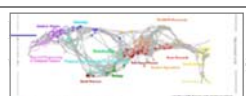



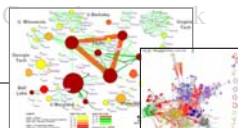



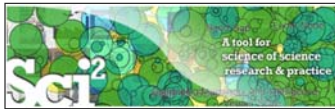
Cite as

Sci² Team. (2009). Science of Science (Sci²) Tool. Indiana University and SciTech Strategies, <http://sci2.cns.iu.edu>



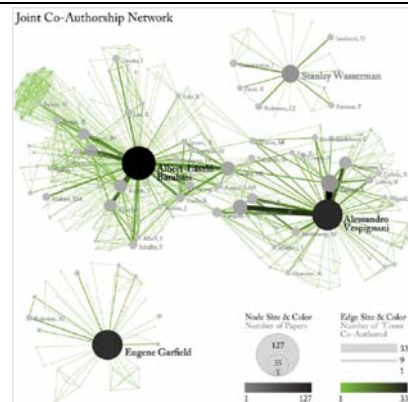
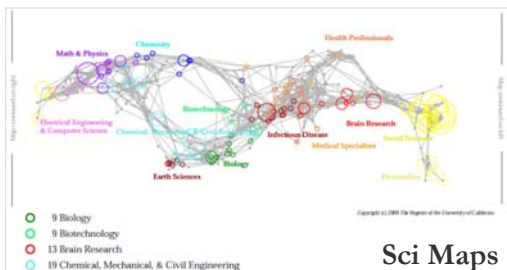
Type of Analysis vs. Level of Analysis

	<i>Micro/Individual</i> (1-100 records)	<i>Meso/Local</i> (101-10,000 records)	<i>Macro/Global</i> (10,000 < records)
Statistical Analysis/Profiling	Individual person and their expertise profiles	Larger labs, centers, universities, research domains, or states	All of NSI, SA, all of sci² 
Temporal Analysis (When)	Funding portfolio of one individual	Topic bursts of PNAS	113 Years of PNAS Research 
Geospatial Analysis (Where)	Career trajectory of one individual	Mapping intellectual	PNAS 
Topical Analysis (What)			VxOrd/Topic r NIH funding 
Network Analysis (With Whom?)	NSI network of one 		NIH's 



Open Code for Replicable S&T Assessment

OSGi/CShell powered tool, see <http://cishell.org>
<http://sci2.cns.iu.edu> | <http://sci2.wiki.cns.iu.edu>



GUESS Network Vis

Horizontal Time Graphs



Börner, Katy, Huang, Weixia (Bonnie), Linnemeier, Micah, Dubon, Russell Jackson, Phillips, Patrick, Ma, Nianli, Zoss, Angela, Guo, Hanning & Price, Mark. (2009). *Reti-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.*



Sci² Tool

Sci² Tool

File Preprocessing Modeling Analysis Visualization Scientometrics Help

Console

Welcome to the Science of Science Tool (Sci²). The development of this tool is supported in Network Science center and the School of Li Indiana University, the National Science Foundation (NSF) Grant IIS-0715303, and the James S. McDonnell Cyberinfrastructure portal (<http://sci.slis.indiana.edu>).

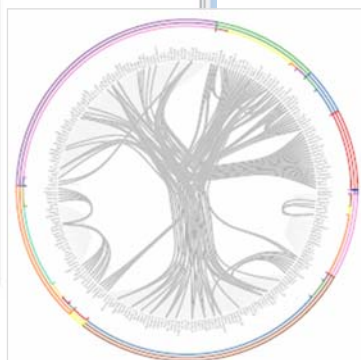
The primary investigators are Katy Börner, In SciTech Strategies Inc. The Sci² tool was developed by J. Duhon, Patrick A. Phillips, Chintan Tank, a Cyberinfrastructure Shell (<http://cishell.org>) for Network Science Center (<http://cns.slis.indiana.edu>). Many algorithm plugins were derived from the Network Workbench Tool (<http://nwb.slis.indiana.edu>).

Please cite as follows:
 Sci² Team. (2009). Science of Science Tool. In SciTech Strategies Inc., <http://sci.slis.indiana.edu>.

Scheduler

Remove From List Remove completed

!	Algorithm Name	Date	Time	% Con
<input checked="" type="checkbox"/>	Extract Co-Author Netw...	09/03/2009	00:15:20 AM	100%
<input checked="" type="checkbox"/>	Load and Clean ISI File	09/03/2009	00:15:05 AM	100%





Network Extraction: Examples

Author co-occurrence network

	A	B
1	Publication	Authors
2	Paper1	A1
3	Paper2	A1;A2;A6
4	Paper3	A1;A3
5	Paper4	A1;A4;A5
6	Paper5	A5;A6
7	Paper6	A1;A2

Extract Network from Table

Extracts a network from a delimited table

Column Name: Authors

Text Delimiter: ;

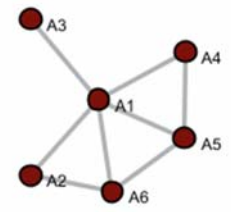
Extract Bipartite network

Extract a bipartite network from two columns in the table. If the column values may list multiple entries, enter the special text which delimits them.

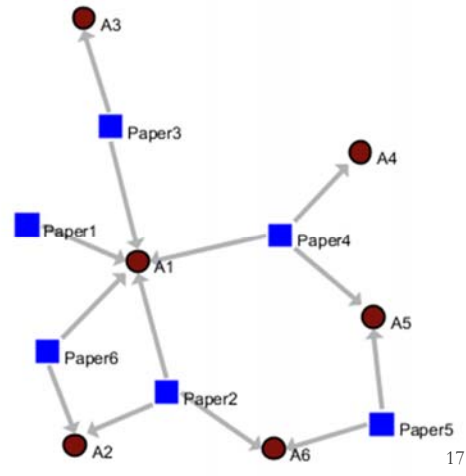
First column: Publication

Second column: Authors

Text Delimiter: ;



Paper-author 2-mode network



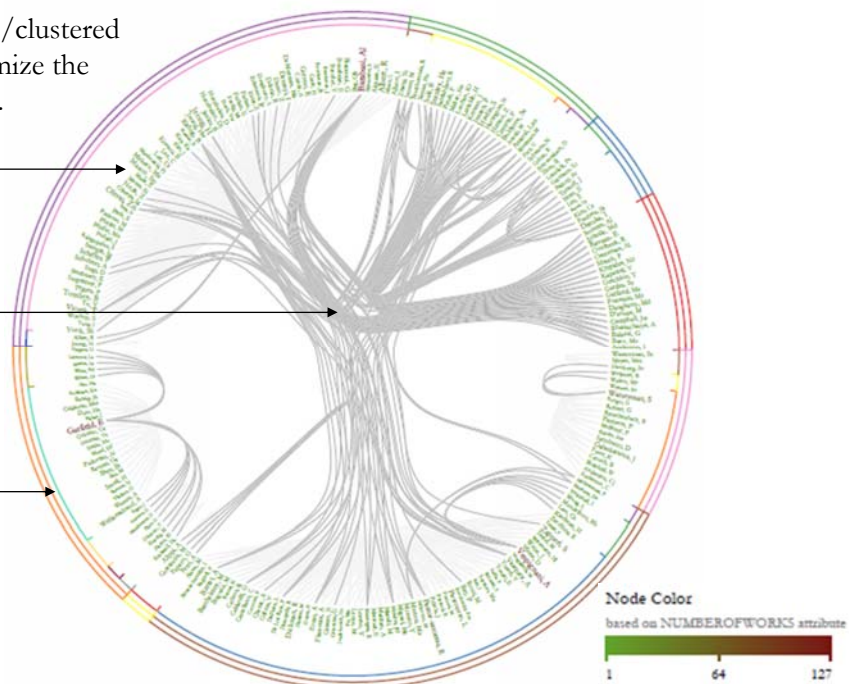
Network Visualization: Circular Hierarchy Visualization

Nodes that are interlinked/clustered are spatially close to minimize the number of edge crossings.

Node labels, e.g., author names.

Network structure using edge bundling.

Color coded cluster hierarchy according to Blondel community detection algorithm.



Note:
Header/footer info, legend, and more meaningful color coding are under development.

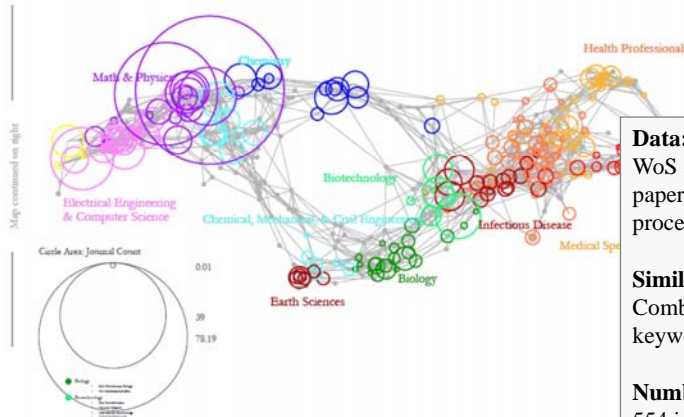


Topic Mapping: UCSD Science Map

Science Map via Journals for FourNetSciResearchers.isi

314 journal references matched out of 361 found.

These 314 references are associated with 13 of 13 disciplines of science and 255 of 554 research specialties in the UCSD Map of Science.



Data:

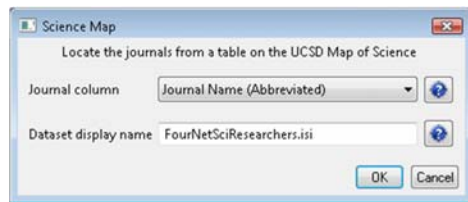
WoS and Scopus for 2001–2005, 7.2 million papers, more than 16,000 separate journals, proceedings, and series

Similarity Metric:

Combination of bibliographic coupling and keyword vectors

Number of Disciplines:

554 journal clusters further aggregated into 13 main scientific disciplines that are labeled and color coded in a metaphorical way, e.g., Medicine is blood red and Earth Sciences are brown as soil.



Geospatial Maps with Congressional Districts

	A
1	Zip code
2	90095
3	4672
4	232980568
5	10032
6	10039242
7	46091500
8	191112434
9	27705
10	981959472
11	10065
12	10065



Identify Congressional District, Latitude, Longitude

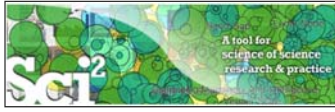
	A	B	C	D
1	Zip code	Congressional District	Latitude	Longitude
2	90095	CA-30	34.0735035	-118.6645815
3	4672	ME-02	45.818717	-69.0290345
4	232980568	VA-03	37.270472	-77.0699835



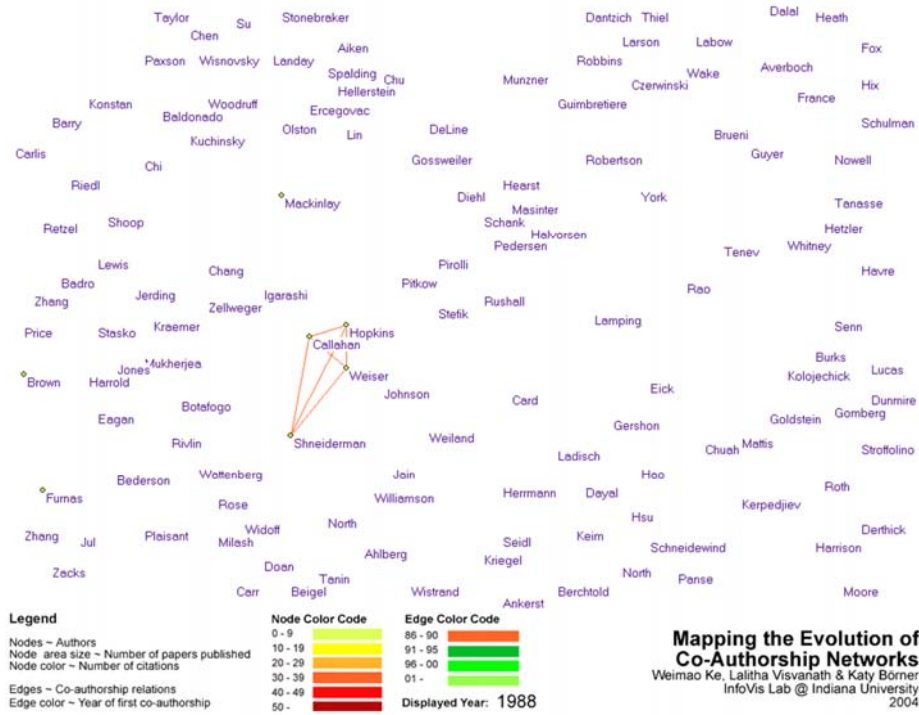
Aggregate/Count identical Congressional Districts

	A	B	C	D
1	Congressional District	Latitude	Longitude	Count
2	CA-30	34.0735035	-118.6645815	4
3	ME-02	45.818717	-69.0290345	2
4	VA-03	37.270472	-77.0699835	1
5	NY-15	40.8341475	-73.9342095	4





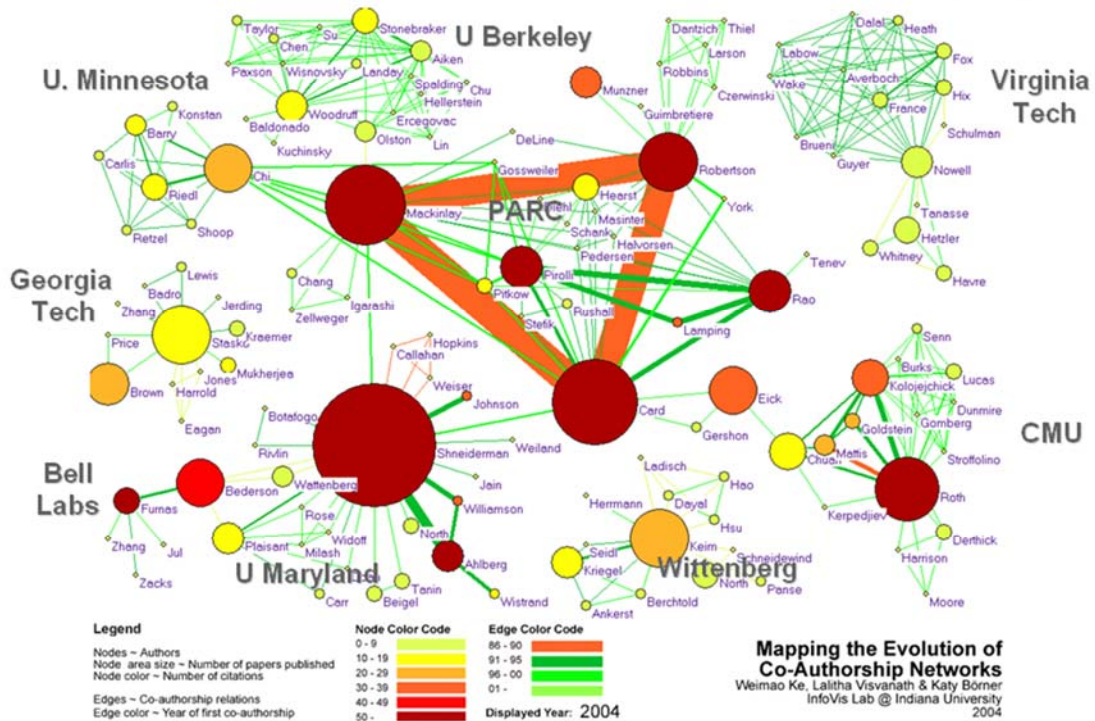
Evolving Collaboration Networks



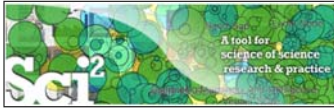
21

Mapping the Evolution of Co-Authorship Networks

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

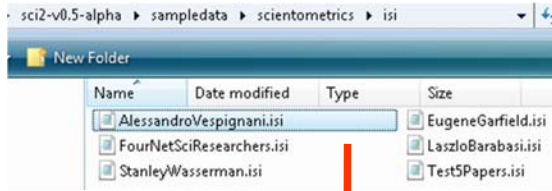


22



Evolving Collaboration Networks

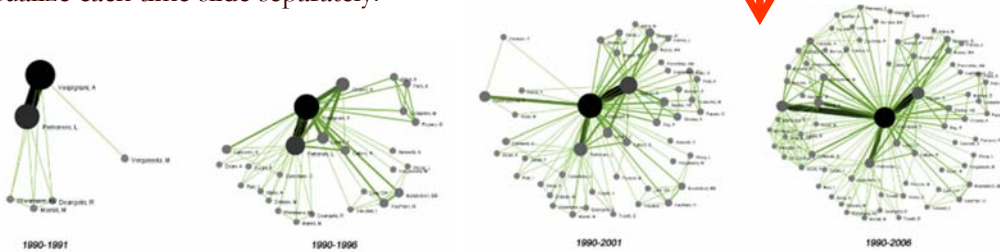
Load isi formatted file



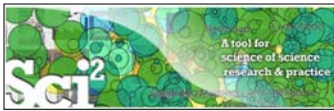
As csv, file looks like:

	A	B	C	D	E	F	G
1	Abstract	Authors	Authors (Full Names)	Beginning	Book Serie	Book Serie	Cited Pate
2	The systematic study of	Colizza, V Barrat, A Barthelemy, M Vespignani, A		2015			
3	Uncovering the hidden r	Colizza, V Flammini, A Serrano, MA Vespignani, A		110			
4	Computer viruses can s	Vespignani, A		135			
5	Mapping the Internet ge	Dall'Asta, L Alvarez-Hamelin, I Barrat, A Vazquez, A Vespignani, A		140			LECTURE NOTES IN

Visualize each time slide separately:



23



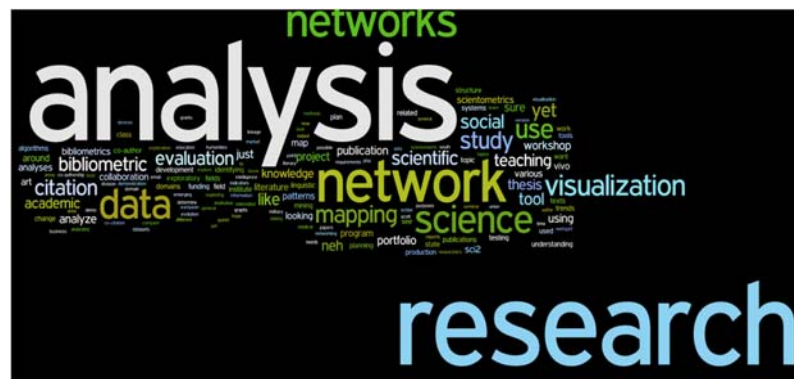
Sci2 Tool Adoption



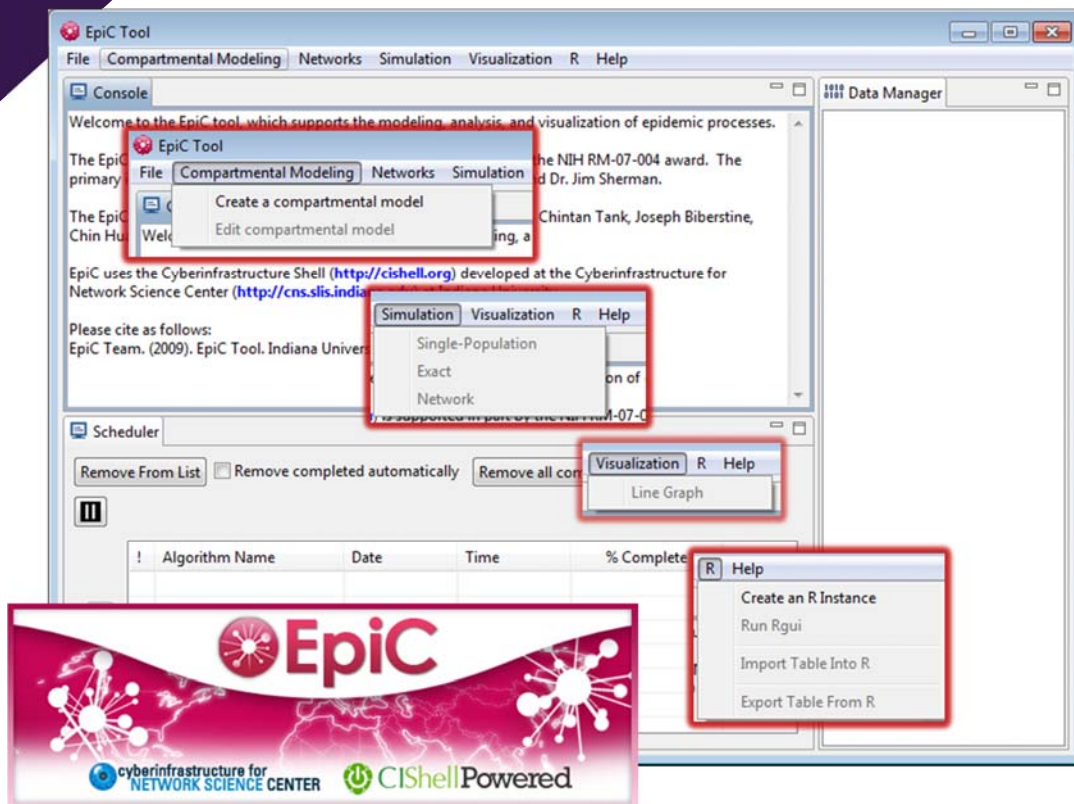
The Sci2 Tool is used by the National Science Foundation, the National Institutes of Health, and the US Department of Agriculture.

“As a new user, I am beginning with very little knowledge of the analyses and modeling techniques that Sci2 enables. I've been able to use my own dataset and follow through some of the workflows to the point of generating the first network and time horizon visuals. That was so exciting I stayed up far past bedtime to get to the visuals.”

*Dr. Suzanne A. Pierce, Center for International Energy and Environmental Policy
Jackson School of Geosciences, The University of Texas at Austin*



24



25

TEXTrend
National Office for Research and Technology, Jedlik Ányos Programme

Home Consortium UseCases Publications Downloads svn Archives Tutorials Support Media

Consortium

Project leader
George Kampis, PhD, DSc

University Press Ltd.
Team leader: George Kampis, PhD, DSc
György Fábri, PhD, CSc
László Gulvás, PhD
Sándor Soós, PhD
Zalán Szokolci, BSc
Zoltán Szász, BSc

HCCI Research Institute of Economics and Enterprises
Team leader: István János Tóth, PhD
Ágnes Czibik, MSc
Ágnes Makó, MSc
Tamás Uhrin, MSc
Zoltán Várhalmi, MSc

Gila Computer Consulting Ltd.
Team leader: Attila Bencsik, MSc
Rita Ádám, MSc
Hennett Bagó, BSc
István Gráf, MSc

TEXTREND: Development of a business and governmental decision support toolbox using trend- and text-analysis tools

The two interconnected objectives of the **TEXTrend project** are (1) the creation of an integrated TEXTrend toolkit and service basis, and (2) the elaboration of **demonstrative applications** in varied fields of business and governmental decisions, exemplified by use cases.

Created at <http://tagcrowd.com>

Created by www.wordle.net

TEXTrend adds WEKA, UIMA, Wordij, CFinder, and more.
See the latest versions of TEXTrend Toolkit modules at <http://textrend.org>

26



About the Cyberinfrastructure Shell

The Cyberinfrastructure Shell (CIShell) is an open source, community-driven platform for the integration and utilization of datasets, algorithms, tools, and computing resources. Algorithm integration support is built in for Java and most other programming languages. Being Java based, it will run on almost all platforms. The software and specification is released under an Apache 2.0 License.

CIShell is the basis of [Network Workbench](#), [TexTrend](#), [SciF](#) and the upcoming [EpiC](#) tool.

CIShell supports remote execution of algorithms. A standard web service definition is in development that will allow pools of algorithms to transparently be used in a peer-to-peer, client-server, or web front-end fashion.

CIShell Features

A framework for easy integration of new and existing algorithms written in any programming language

Using CIShell, an algorithm writer can fully concentrate on creating their own algorithm in whatever language they are comfortable with. Simple tools are provided to then take their algorithm and

Learn More...

- [CIShell Papers](#)
- [CIShell Powered Tools](#)
- [Algorithms](#)
- [Plugins \(coming soon\)](#)
- [Misc. Tool Documentation](#)
- [CIShell Web Services \(coming soon\)](#)
- [Screenshots](#)

Getting Started...

- [Documentation & Developer Resources](#)
- [Download](#)

Getting Involved...

- [Contact Us](#)

CIShell Developer Guide is at <http://cishell.wiki.cns.iu.edu>

Additional Sci2 Plugins are at <http://sci2.wiki.cns.iu.edu/3.2+Additional+Plugins>

27



A number of other projects recently adopted OSGi and/or CIShell:

- USA**
 - *Cytoscape* (<http://cytoscape.org>) Led by Trey Ideker at the University of California, San Diego 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).
 - *MAEviz* (<https://wiki.ncsa.uiuc.edu/display/MAE/Home>) Managed by Jong Lee at NCSA is an open-source, extensible software platform which supports seismic risk assessment based on the Mid-America Earthquake (MAE) Center research.
 - Europe**
 - *Taverna Workbench* (<http://taverna.org.uk>) Developed by the myGrid team (<http://mygrid.org.uk>) led by Carol Goble at the University of Manchester, U.K. 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.
 - *TEXTrend* (<http://textrend.org>) Led by George Kampis at Eötvös Loránd University, Budapest, Hungary supports natural language processing (NLP), classification/mining, and graph algorithms for the analysis of business and governmental text corpuses with an inherently temporal component.
 - *DynaNets* (<http://www.dynanets.org>) Coordinated by Peter M.A. Sloot at the University of Amsterdam, The Netherlands develops algorithms to study evolving networks.
 - *SISOB* (<http://sisob.lcc.uma.es>) An Observatory for Science in Society Based in Social Models.
- 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.

28

Plug-and-Play Macroscopes



EpiC



Converters



Sci2



NWB



TexTrend

Common algorithm/tool pool
Easy way to share new algorithms
Workflow design logs
Custom tools

- IS
- CS
- Bio
- SNA
- Phys



Maps created using Sci2 are travelling in the “Expedition Zukunft” science train visiting 62 cities in 7 months, 12 coaches, 300 m long. <http://www.expedition-zukunft.de>

And they are part of the international Mapping Science exhibit: <http://scimaps.org>

Impact

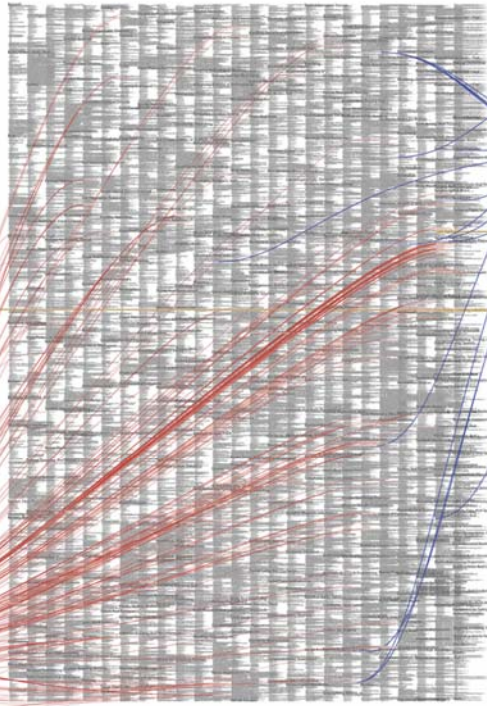
The United States Patent and Trademark Office allows scientists and industry a great service by granting patents to protect inventions. Inventions are categorized in a taxonomy that groups patents by industry, by use, government function, effect or product, and structure. At the time of this writing there are 160,323 categories in a hierarchy that can split as deep as 13 levels. We display the first three levels (13,329 categories) at right in what might be considered a textual map of inventions.

Patent applications are required to be unique and non-obvious, partially by revealing any previous patents that might be similar in nature or provide a foundation for the current invention. In this way we can trace the impact of a single patent, seeing how many patents and categories it affects.

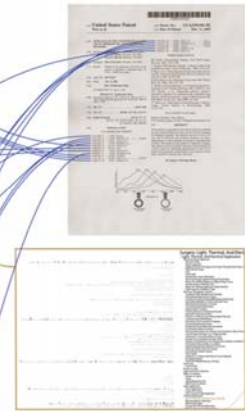
The patent on Gore-tex—a lightweight, durable synthetic fiber—is an example of one that had significant impact. The box below enlarges the section of the hierarchy where it is filed, and the red lines (arranged to start along a time line from 1961 to 2000) point to the 130 categories that contain 162 patents, from waterproof clothing to surgical cosmetic implants, that mention Gore-tex as prior art.



US Patent Hierarchy



Prior Art



New patents often build on older ideas from many categories. Here, blue lines originate in various different categories that contain the patents cited as prior art for a patent on 'gold nanoshells.' Gold nanoshells are a new invention: the spheres (with a diameter ten million times smaller than a human hair) that can be used to make tumors more visible in infrared scans, and have even helped cause complete remission of tumors in tests with laboratory mice. The blue lines show that widely separated categories provided background for this invention.

Keeping categories understandable is an important part of maintaining any taxonomy, including the patent hierarchy. Categories are easier to understand, search, and maintain if they contain elements (patents in this case) that fit well within the definition of the category. The box above shows a tiny bar chart, part of a 'Taxonomy Validator' that helps people decide whether categories are good ones.

Categories can be modified or combined, and sometimes need to be split when they become too large a container problem shared by many classification systems in this information-rich century. And how can we do this more exactly where to split a category in two, for example: if there are hundreds or thousands of elements in it?

The Taxonomy Validator measures a 'distance to prototype' how far each element is from an idealized 'prototypic' element for each bucket. This can be based on statistics, a computational comparison of words, or even human judgment. A simple bar chart can then show how good a category is. A good category has lots of small bars, a generally jagged category is one that might need activity or reorganization, while one that has only one or two tall bars may just mean that one or two elements don't belong. Even simple visualizations like this can ease knowledge work by showing the eye much more than it can fit into memory as words, focusing people on just the right issues, and providing a really readable background to support more informed judgements.

Katy Börner, Elisha Hardy, Bruce W. Herr II, Todd Holloway, and W. Bradford Paley (2006) Taxonomy Visualization of Patent Data.

113 Years of Physical Review

The visualization aggregates 3800+ articles published in 133 volumes of Physical Review between 1893 and 2000. The 133 volumes contain over 100,000 pages of text. The visualization is based on the Physical Review Online (PRL) archive and the complete metadata for the Physical Review Online (PRL) archive.

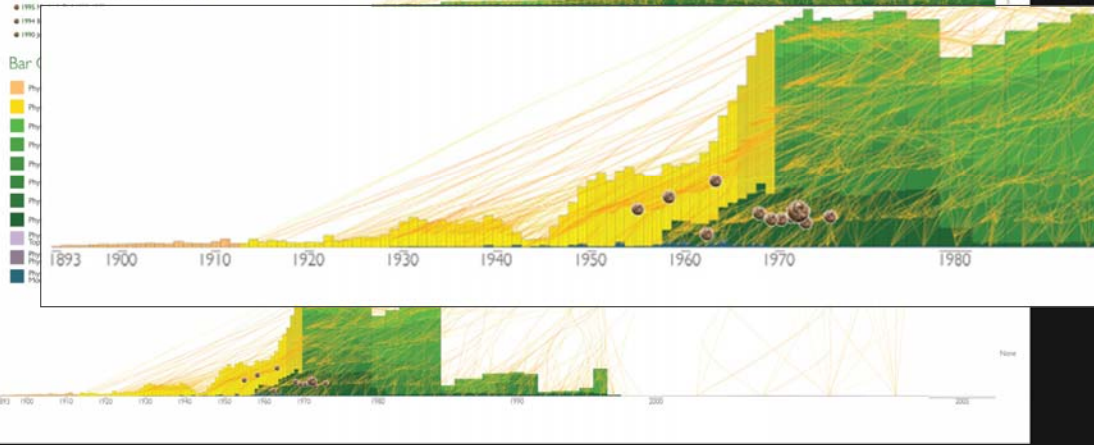
On top of the base map, all citations from the papers in every Physical Review volume in 2000 are mapped and then shown as lines connecting the volumes.

The final Physical Review Online includes the 13 volumes spanning the 113 years appearing in Physical Review for 113 Nobel prizes between 1893 and 2000. Each year, Thomson ISI provides three Nobel Prize awardees in physics based on their research, high school grades, and statements of former teachers and other individuals. Current problems for Thomson ISI are highlighted.

Each article has a calculated verticality into the present that gives it a 'high' proportional to the number of papers and each journal's published horizontally into the volume of the journal appearing in the volume.

Nobel Prizes in Physical Review

- 2000 Roy J. Glauber, John L. Hall, and Theodor W. Hänsch (1981)
- 2000 Charles K. Geil, David Heston, and Herta and Paul Amirian (1973)
- 2000 Anthony J. Leggett (1970)
- 2000 Raymond Davis Jr., Masatoshi Koshiba, and Riccardo Giacconi (1962, 1968, 1987)
- 1999 Eric A. Cornell, Wolfgang Ketterle, and Carl E. Wieman (1996, 1998)
- 1998 Robert B. Laughlin (1982, 1983)
- 1997 Steven Chu and William D. Phillips (1982, 1986, 1998)
- 1996 David N. Lee, Douglas D. Osheroff, and Robert C. Richardson (1972)
- 1995
- 1994
- 1993



Bruce W. Herr, Russell Duhon, Elisha F. Hardy, Shashikant Penumarthy, and Katy Börner (2007) 113 Years of Physical Review.

Examining the Evolution & Distribution of Patent Classifications

Managing Growing Patent Portfolios

Organizations, businesses, and individuals use patents to protect their intellectual property and business models. As market competition increases, patenting innovation and intellectual property rights becomes ever more important.

Managing the staggering number of patents demands new tools and methodologies. Grouping patents by their classifications offers an ideal resolution for better understanding how intellectual borders are established and change over time.

The charts below show the annual number of patents granted from January 1, 1976 to December 31, 2002 in the United States Patent and Trademark Office (USPTO) patent archive, slow and fast growing patent classes, the top-10 fast growing patent subclasses, and two evolving patent portfolios.

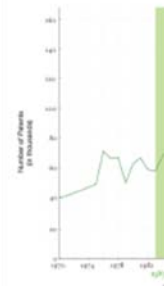
The Structure and Evolution of the Patent Space

The United States Patent and Trademark Office assigns each patent to one of more than 450 classes covering broad application domains. For example, class 514 encompasses all patents dealing with "Drug, Bio-Affecting and Body Testing Compositions." Classes are further broken down by subclasses that have hierarchical associations. As one example, class 455 features subclass 99 entitled "with vehicle."

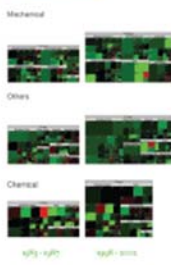
The top-10 fast growing patent classes for 1998-2002 are listed together with the number of patents granted. Most come from the "Computer and Communications" and the "Drugs and Medical" areas.

The evolving hierarchical structure of patent classes and their sizes is represented using treemaps, a space-filling visualization technique developed by Ben Shneiderman at the University of Maryland. A treemap presents a hierarchy as a collection of nested rectangles—demonstrating a parent-child relationship between nodes by nesting the child within the parent rectangle. The size and color of each rectangle represent certain attributes of the nodes.

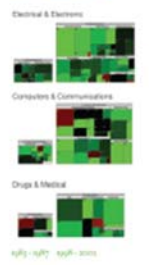
Here, each rectangle represents a class and the area size denotes the total number of patents in that class. The rectangle's color corresponds to percentage increase (green) or decrease (red) in the number of patent granted in that class from the previous interval.



Slow Growing Classes



Fast Growing Classes



Top-10 Subclasses

Class	Title	# of Patents
514	Drug, Bio-Affecting and Body Testing Compositions	18,778
438	Semiconductor Device Manufacturing Process	17,775
431	Chemistry: Molecular Biology and Microbiology	17,474
474	Drug, Bio-Affecting and Body Testing Compositions	13,677
428	Block Material or Miscellaneous Articles	13,314
257	Active Solid-State Device (e.g., Transistors, Solid-State Diodes)	12,974
395	Information Processing System Organization	9,955
345	Computer Graphics Processing, Operator Interface Processing, and Selective Visual Display Systems	9,810
359	Optical Systems and Elements	9,151
365	Static Information Storage and Retrieval	8,392
Total		130,910

Patent Portfolio Analysis

A longitudinal analysis of portfolios reveals different patenting strategies. For each year (given in gray above each treemap), a treemap of all new patents granted to the assignee is shown. The number of patents is given below each treemap. The same size and color coding as above was used. In addition, yellow indicates that no patent has been granted in that class in the last 5 years.

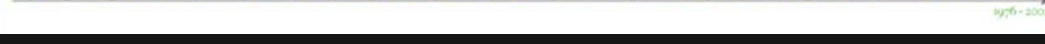
Apple Computer, Inc.

Apple Computer, Inc.'s portfolio starts in 1980 and increases considerably in size over time. In most years, more than half of Apple Computer's patent filings were placed into four classes, namely "Information Processing System Organization," "365 Computer Graphics Processing, Operator Interface Processing, and Selective Visual Display Systems," "362 Image Analysis," and "707 Data Processing: Database and File Management or Data Structures." These four classes are an integral part of Apple Computer, Inc.'s patent portfolio, receiving patents every year.



Jerome Lemelson

The patent portfolio of Jerome Lemelson shows a very different activity pattern. Starting in 1976, he publishes between 6-20 patents each year. However, the predominance of yellow shows that there is little continuity from previous years in regards to the classes into which patents are filed; no class dominates. Instead, more and more new intellectual space is claimed.



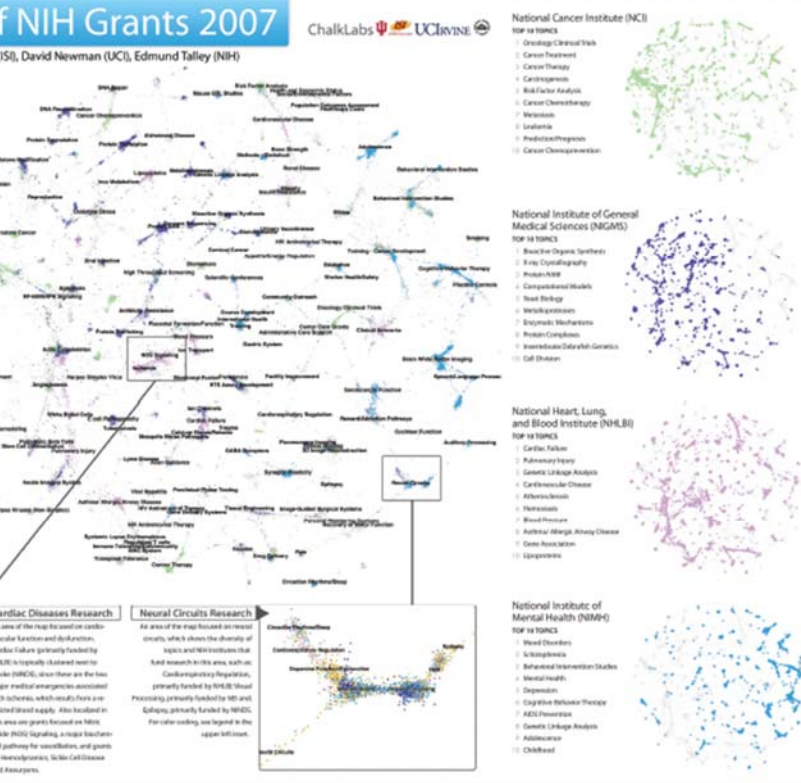
Daniel O. Kutz, Katy Börner & Elisha Hary (2004) Examining the Evolution and Distribution of Patent Classifications.

A Topic Map of NIH Grants 2007

Bruce W. Herr II (ChalkLabs & IU), Gully Burns (ISI), 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 float 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 ncicb.nci.nih.gov/institutes/nci-topics. Institute abbreviations can be found at www.nih.gov/ncicb.

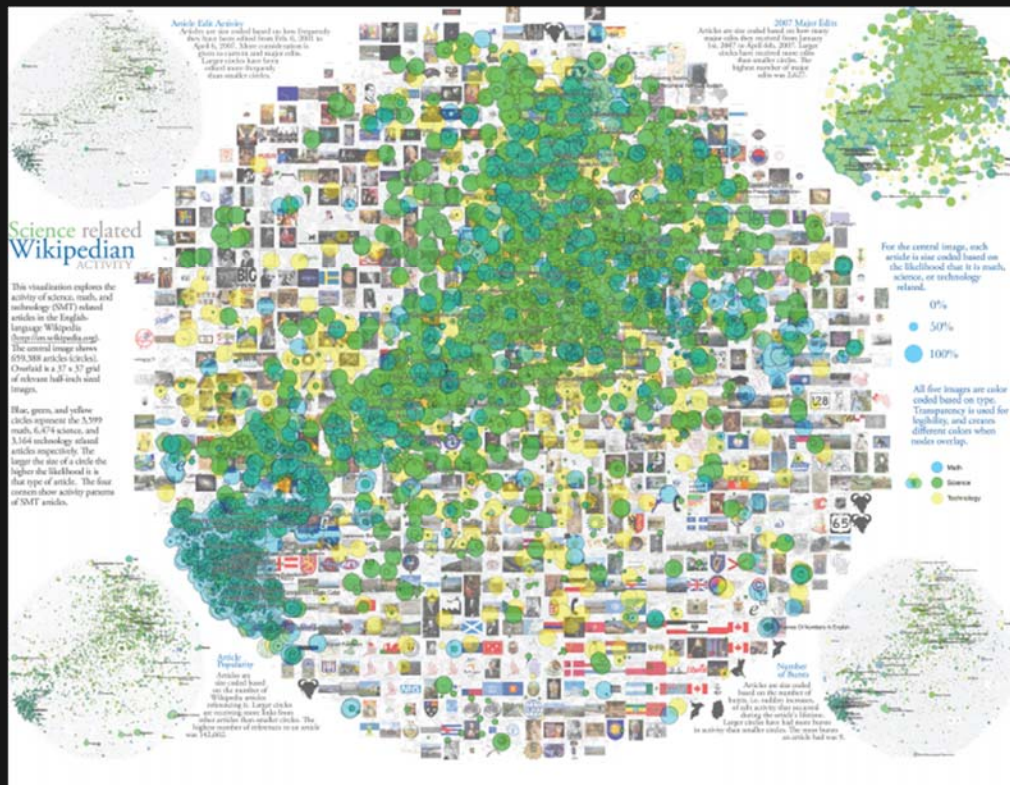
Topic modeling, a statistical technique that automatically learns semantic categories, was applied to assess projects in terms used by researchers to describe their work, without the biases of keywords or subject headings. Grant similarities were derived from their topic mixtures, and grants were then 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.



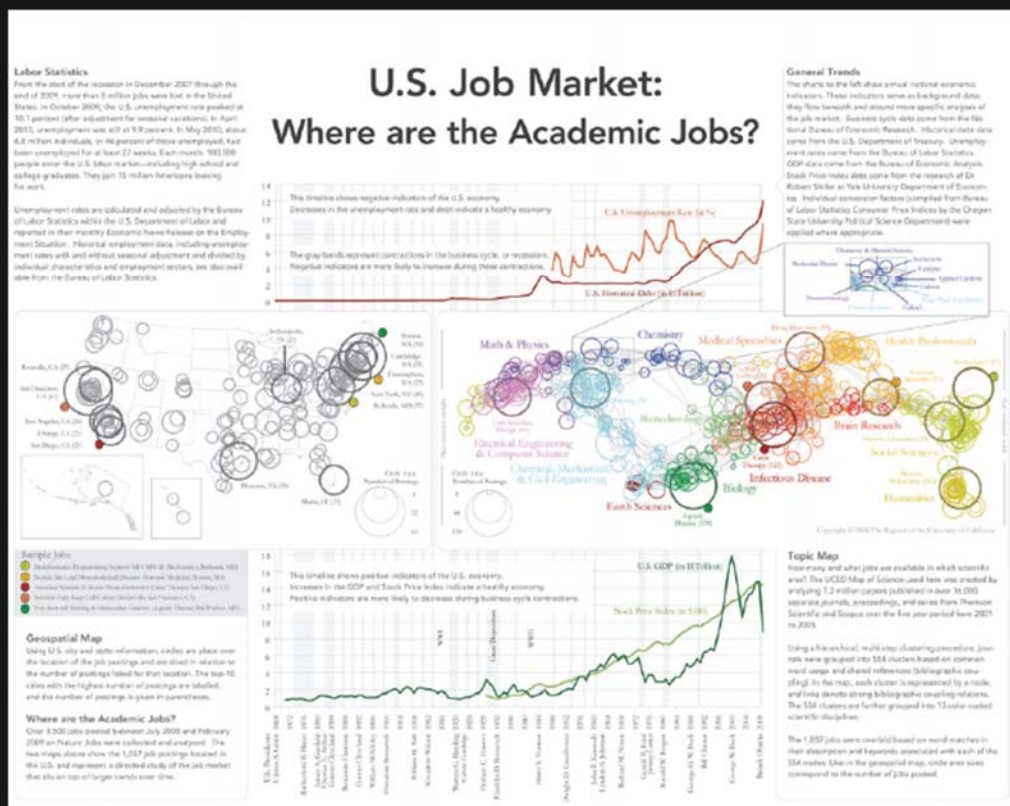
Cardiac Diseases Research
An area of the map focused on cardiovascular function and dysfunction. Cardiac failure primarily funded by NIGMS is notably clustered near the center (NIGMS), along with the two major medical emergencies associated with ischemia, which result from a restricted blood supply. Also located in this area are grants focused on NIGMS (DNA/RNA signaling, a major biochemical pathway for vasodilation, and gene-environmental interactions in the cardiovascular and Alzheimer's).

Neural Circuits Research
An area of the map focused on neural circuits, which shows the diversity of brain and neurobiological research in this area, such as: Cerebellar/brainstem regulation, primarily funded by NHLBI; Visual Processing, primarily funded by NIGMS and NIH; and Sensory, primarily funded by NIGMS. The color coding, not legend in this upper left-hand.

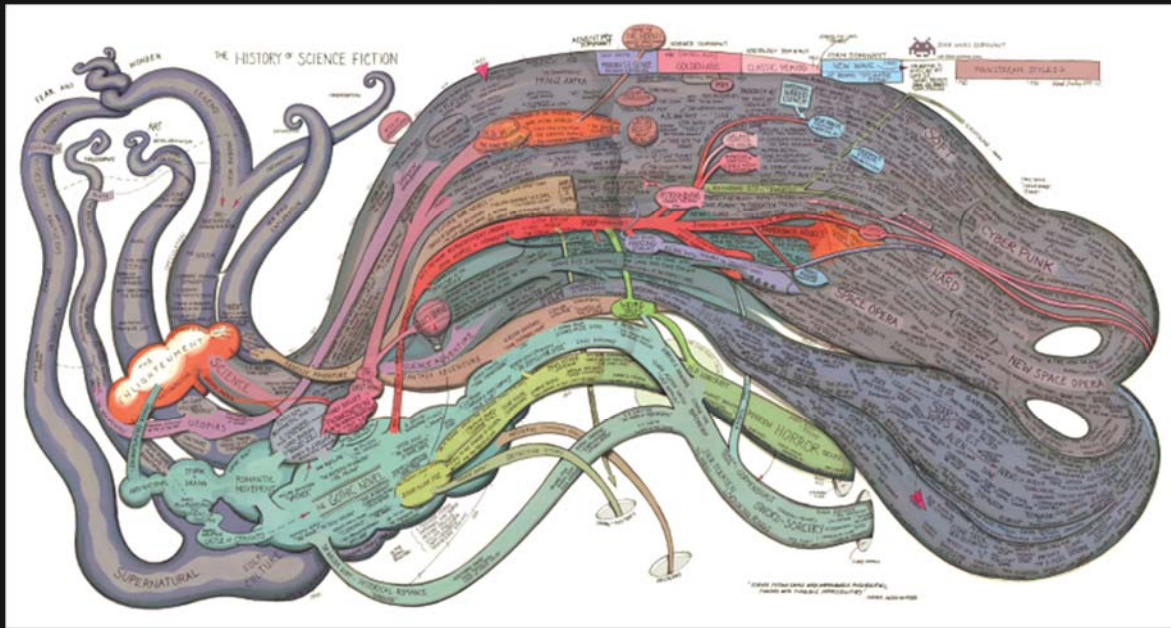
Bruce W. Herr II, Gully Burns, David Newman, Edmund Talley (2007) A Topic Map of NIH Grants.



Bruce W. Herr II, Todd M. Holloway, Elisha F. Hardy, Kevin W. Boyack, and Katy Börner (2007) Science Related Wikipedian Activity.



Angela M. Zoss and Katy Börner (2010) U.S. Job Market: Where are the Academic Jobs?



Ward Shelley. 2011. History of Science Fiction.



Debut of 5th Iteration of the Mapping Science Exhibit at MEDIA X took place at Wallenberg Hall, Stanford University, <http://mediax.stanford.edu>, <http://scaleindependentthought.typepad.com/photos/scimaps>

This is the only mockup in this slide show.
Everything else is available today.



References

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

Shiffrin, Richard M. and Börner, Katy (Eds.) (2004). **Mapping Knowledge Domains**. *Proceedings of the National Academy of Sciences of the United States of America*, 101(Suppl_1).
http://www.pnas.org/content/vol101/suppl_1/

Börner, Katy, Sanyal, Soma and Vespignani, Alessandro (2007). **Network Science**. In Blaise Cronin (Ed.), *ARIST*, Information Today, Inc., 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>

Scharnhorst, Andrea, Börner, Katy, van den Besselaar, Peter (2011) **Models of Science Dynamics**. Springer Verlag.

