Towards Scholarly Marketplaces

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

Talk in Trey Ideker's Group, UCSD

2007.08.06

CIShell | **Building Market Places not Cathedrals**

- ➢ Design & implementation of 'software glue' that can interlink datasets and algorithms written in different languages using different data formats.
- ➢ The smaller the glue or 'CI Shell', the more likely it can be maintained.
- ➢ Dataset and algorithm 'plugins' are provided by application holders/ community.
- ➢ Applications resemble custom 'fillings'.

Cyberinfrastructure Shell (CIShell)

http://cishell.org

CIShell is an 'empty shell' that supports

➢ Easy integration of new datasets and algorithms by <u>algorithm developers</u> and

➢ Easy usage of algorithms by <u>algorithm users.</u>

Its <u>plug-and-play architecture</u> supports the integration and utilization of diverse

➢ Datasets, e.g., stored in files, databases, steaming data.

➢ Algorithms, e.g., data processing, analysis, modeling, visualization.

➢ Interfaces, e.g., remote services, scripting engines, peer-to-peer clients.

➢ Services, e.g., workflow support, scheduler.

Hence, it can be used for custom UI/Toolkit development.

CIShell – Needs of Algorithm Developers & Users

CIShell – Needs of Algorithm Developers & Users

Developers

Users

CIShell Wizards

CIShell

IVC Interface

NWB Interface
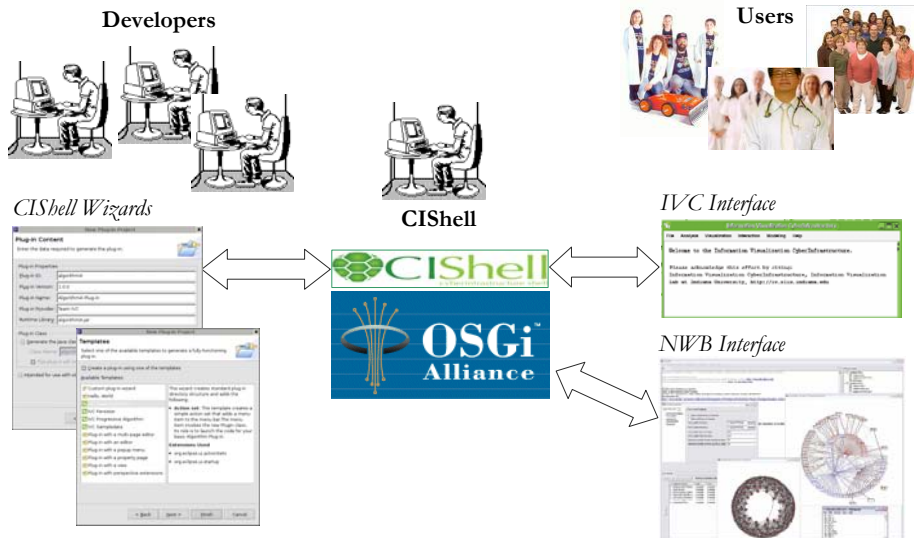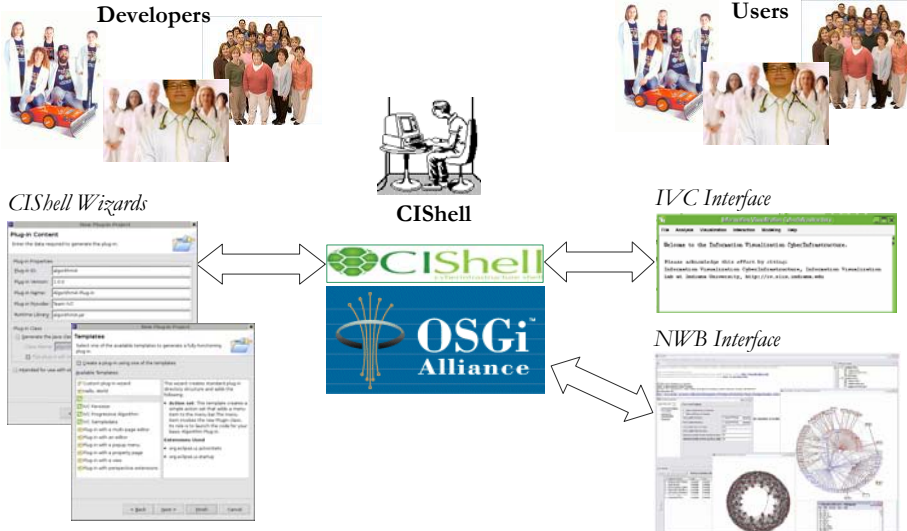
CIShell – Technical Details

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 7 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 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. Developed Tools/CI, e.g., IVC & NWB, provide a reference GUI for underlying services.
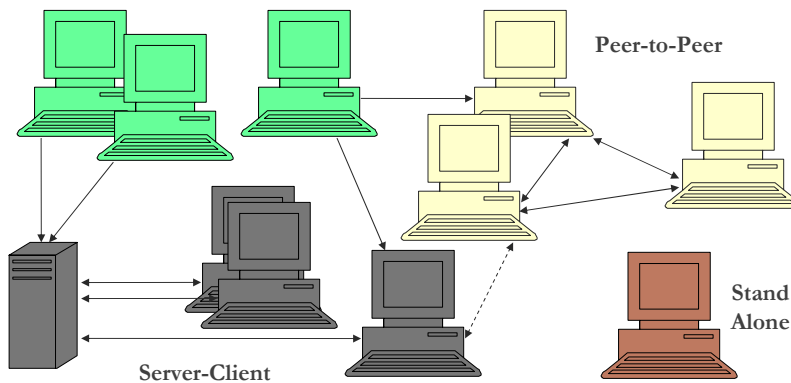
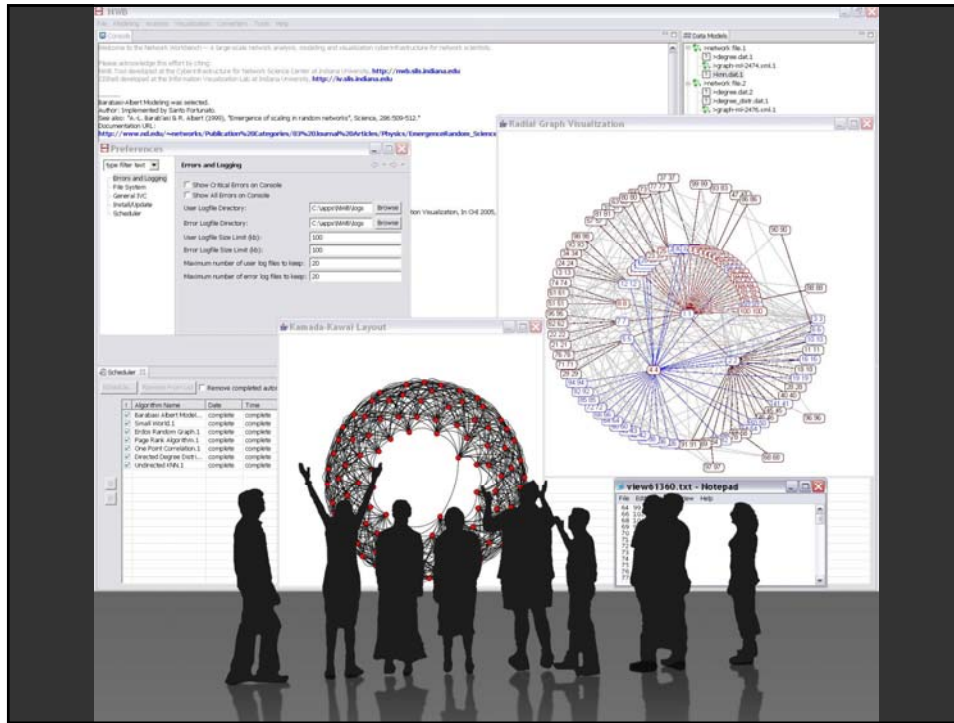CIShell – Technical Details

CIShell layer cake.

CIShell – Deployment

Data-Algorithm Repositories

Peer-to-Peer

Stand Alone

Server-Client

CIShell applications can be deployed as distributed data and algorithm repositories, stand alone applications, peer-to-peer architectures, and server-client architectures.

NetworkWorkbench	Network Workbench (NWB)

Investigators:	Katy Börner, Albert-Laszlo Barabasi, Santiago Schnell, Alessandro Vespignani & Stanley Wasserman, Eric Wernert
Software Team:	Lead: Weixia (Bonnie) Huang
	Developers: Bruce Herr, Ben Markines, Santo Fortunato, Cesar Hidalgo, Ramya Sabbineni, Vivek S. Thakre, & Russell Duhon
Goal:	Develop a large-scale network analysis, modeling and visualization toolkit for biomedical, social science and physics research.
Amount:	$1,120,926 NSF IIS-0513650 award.
Duration:	Sept. 2005 - Aug. 2008
Website:	http://nwb.slis.indiana.edu

NetworkWorkbench   **NWB Advisory Board**

- Ulrik Brandes, University of Konstanz, Germany (Graph Theory)
- Noshier Contractor, Northwestern University (Communication Theory)
- Mark Gerstein, Yale University (Bioinformatics)
- James Hendler, Rensselaer Polytechnic Institute (Semantic Web)
- Jason Leigh, Electronic Visualization Laboratory, University of Illinois at Chicago (Visualization & CI)
- Neo Martinez, Pacific Ecoinformatics and Computational Ecology Lab (Biology)
- Michael Macy, Cornell University (Sociology)
- Stephen North, AT&T (Graph Visualization)
- Tom Snijders, University of Groningen (Social Network Analysis)

NetworkWorkbench   **NWB CI Deliverables**

Glue:
- CIShell   Core programmer team lead by Bonnie Huang
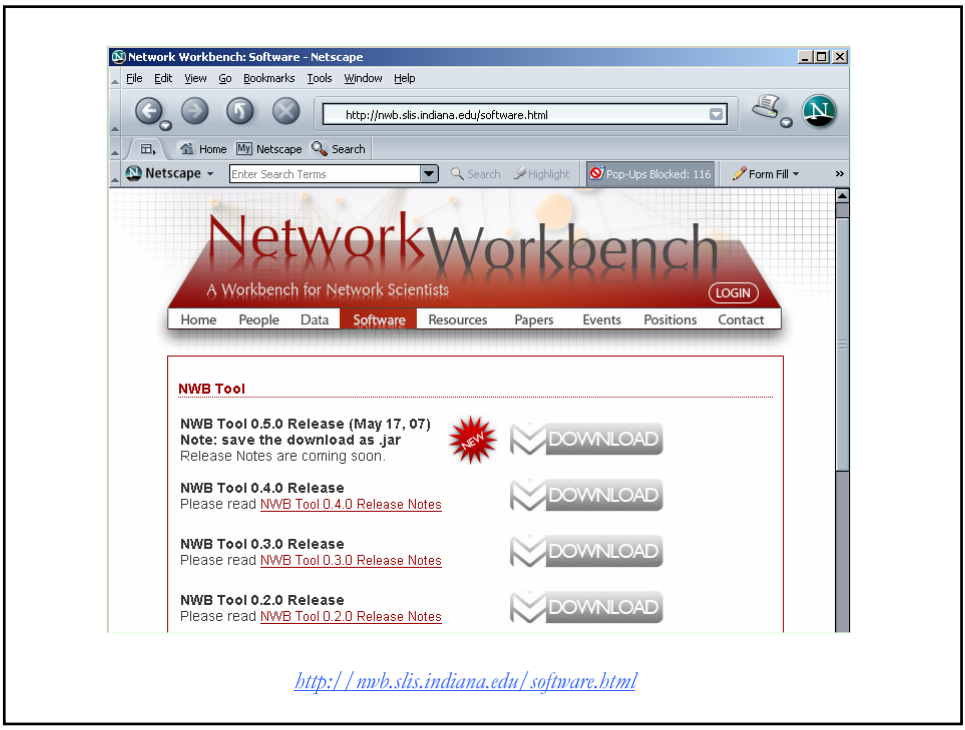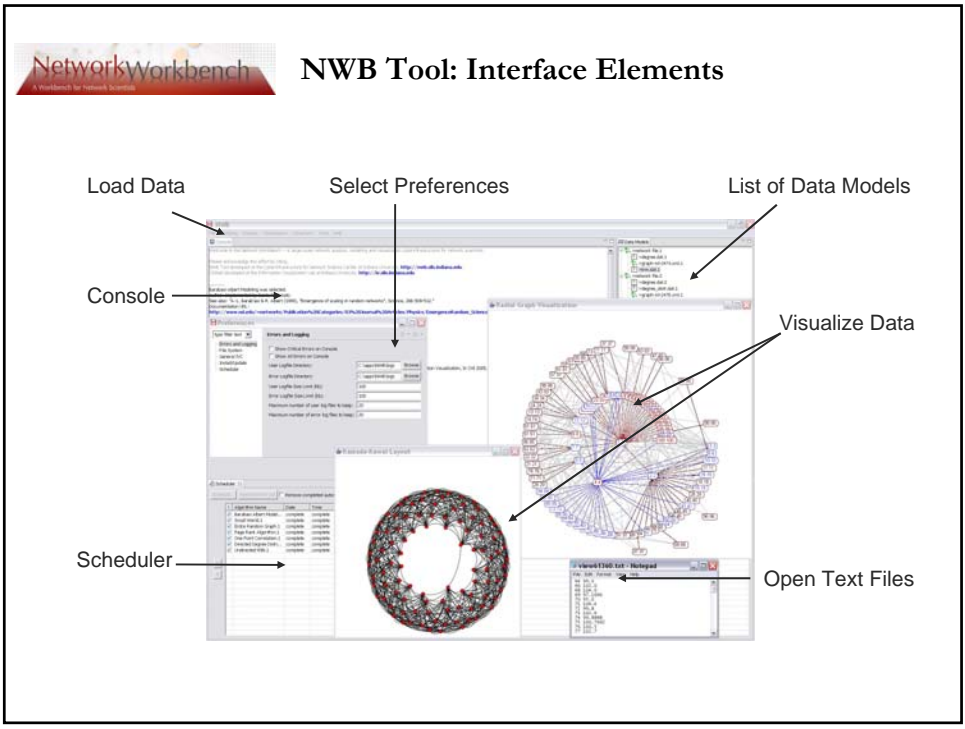
Tools, Services & Portals:
- NWB Tool   Lead by Alex Vespignani with input from other PIs
- SciMaps Service Online   Lead by Katy Borner
- Bio Tool   Lead by Laszlo Barabasi & Santiago Schnell

All three are prototypical instantiations of CIShell serving as reference implementations.

Documentation/Registry/Market Place:
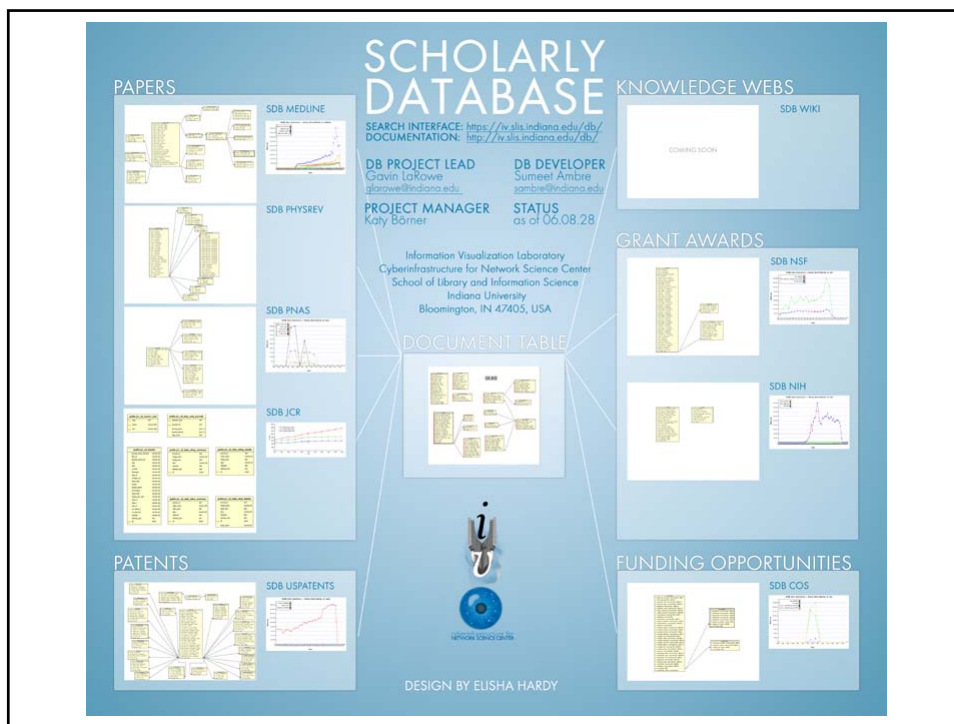- NWB Community Wiki   Lead by Katy Borner

NWB Tool: Interface Elements

Load Data   Select Preferences   List of Data Models

Console

Visualize Data

Scheduler

Open Text Files

http://nwb.slis.indiana.edu/software.html

NWB Tool 0.2.0: List of Algorithms

Category	Algorithm	Language
Preprocessing	Directory Hierarchy Reader	JAVA
Modeling	Erdös-Rényi Random	FORTRAN
	Barabási-Albert Scale-Free	FORTRAN
	Watts-Strogatz Small World	FORTRAN
	Chord	JAVA
	CAN	JAVA
	Hypergrid	JAVA
	PRU	JAVA
Visualization	Tree Map	JAVA
	Tree Viz	JAVA
	Radial Tree / Graph	JAVA
	Kamada-Kawai	JAVA
	Force Directed	JAVA
	Spring	JAVA
	Fruchterman-Reingold	JAVA
	Circular	JAVA
	Parallel Coordinates (demo)	JAVA
Tool	XMGrace	

Analysis Algorithm	Language
Attack Tolerance	JAVA
Error Tolerance	JAVA
Betweenness Centrality	JAVA
Site Betweenness	FORTRAN
Average Shortest Path	FORTRAN
Connected Components	FORTRAN
Diameter	FORTRAN
Page Rank	FORTRAN
Shortest Path Distribution	FORTRAN
Watts-Strogatz Clustering Coefficient	FORTRAN
Watts-Strogatz Clustering Coefficient Versus Degree	FORTRAN
Directed k-Nearest Neighbor	FORTRAN
Undirected k-Nearest Neighbor	FORTRAN
Indegree Distribution	FORTRAN
Outdegree Distribution	FORTRAN
Node Indegree	FORTRAN
Node Outdegree	FORTRAN
One-point Degree Correlations	FORTRAN
Undirected Degree Distribution	FORTRAN
Node Degree	FORTRAN
k Random-Walk Search	JAVA
Random Breadth First Search	JAVA
CAN Search	JAVA

https://nwb.slis.indiana.edu/community

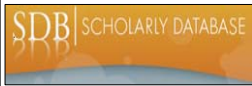**Scholarly Database: Web Interface**

Search across publications, patents, grants.

Download records and/or (evolving) co-author, paper-citation networks.


Register for free access at https://sdb.slis.indiana.edu.

**SDB | SCHOLARLY DATABASE**

Scholarly Database: # Records & Years Covered

Datasets available via the Scholarly Database (* future feature)

Dataset	# Records	Years Covered	Updated	Restricted Access
Medline	13,149,741	1965-2005	Yes	
PhysRev	398,005	1893-2006		Yes
PNAS	16,167	1997-2002		Yes
JCR	59,078	1974, 1979, 1984, 1989 1994-2004		Yes
USPTO	3,179,930	1976-2004	Yes*	
NSF	174,835	1985-2003	Yes*	
NIH	1,043,804	1972-2002	Yes*	
Total	**18,021,560**	**1893-2006**	**4**	**3**

Aim for comprehensive time, geospatial, and topic coverage.
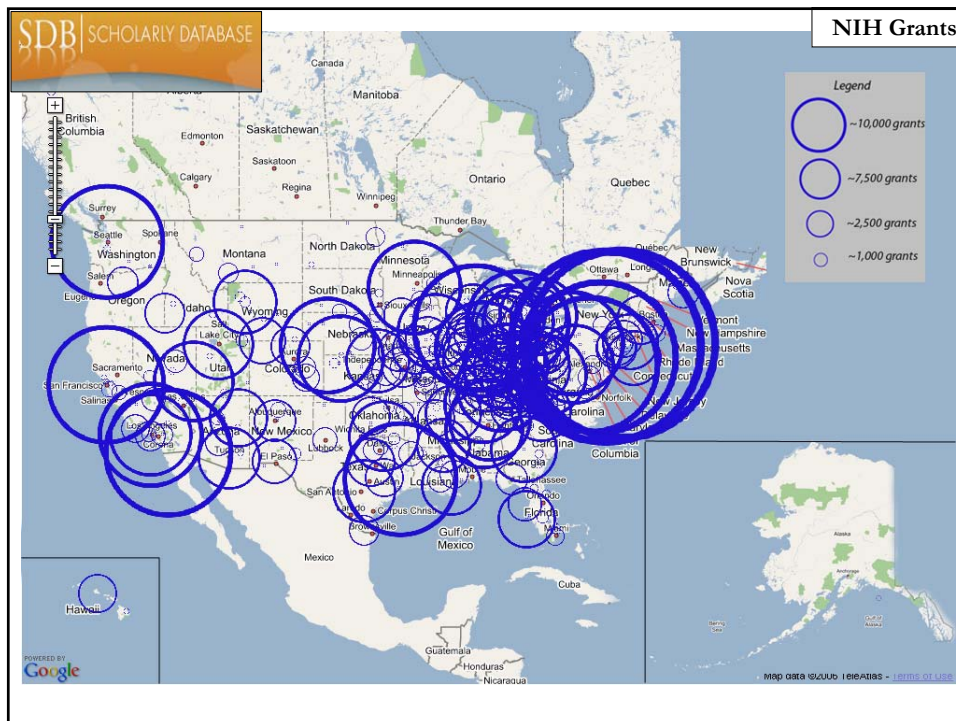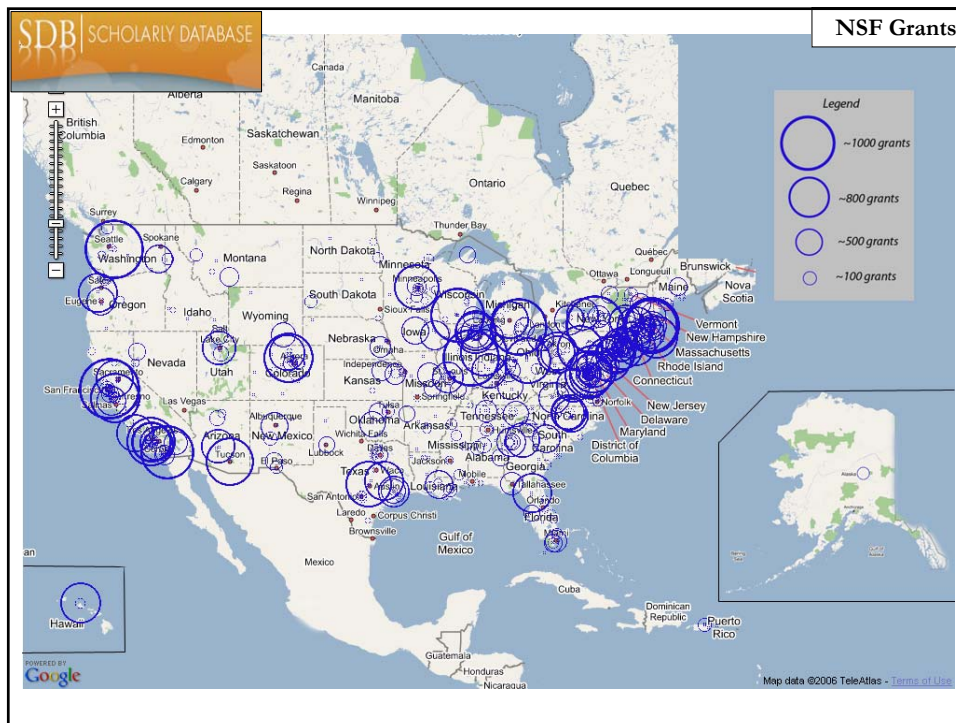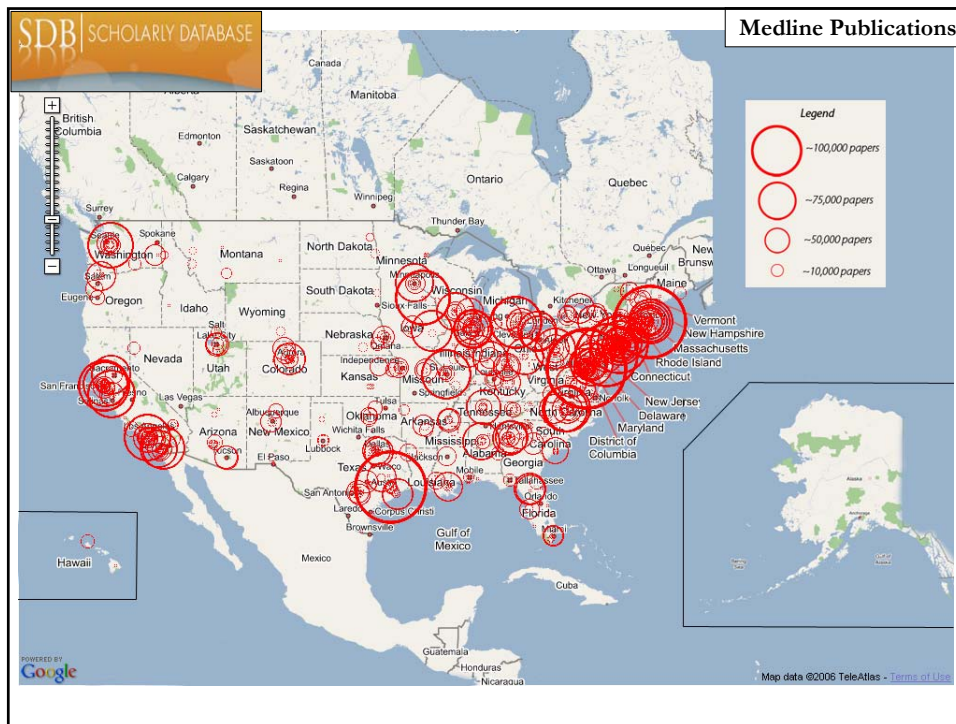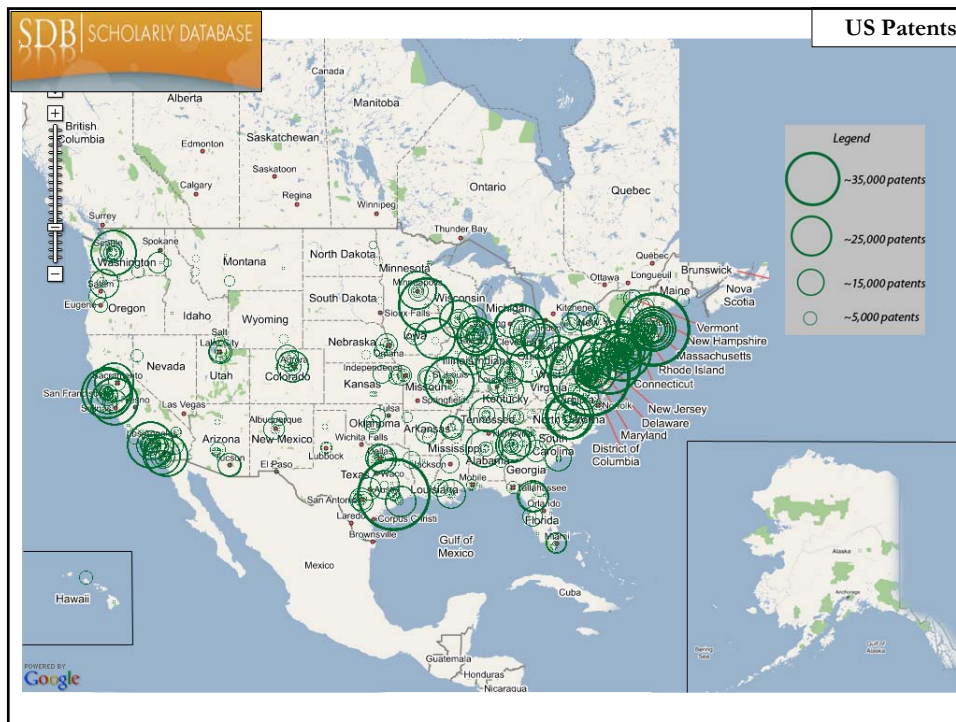
SDB | SCHOLARLY DATABASE

NIH Grants

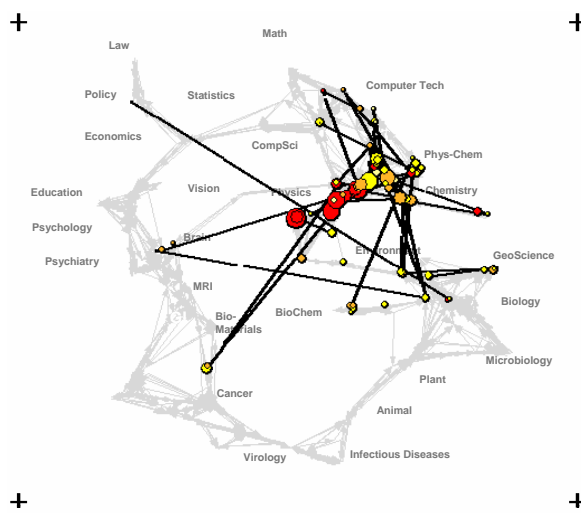**Medline Publications**

NSF Grants

Science map applications: Identifying core competency
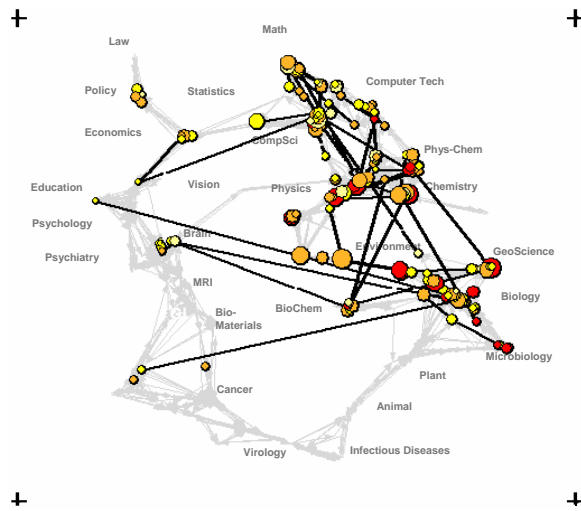*Kevin W. Boyack, Katy Börner & Richard Klavans, 2007*

Funding patterns of the US Department of Energy (DOE)

Science map applications: Identifying core competency
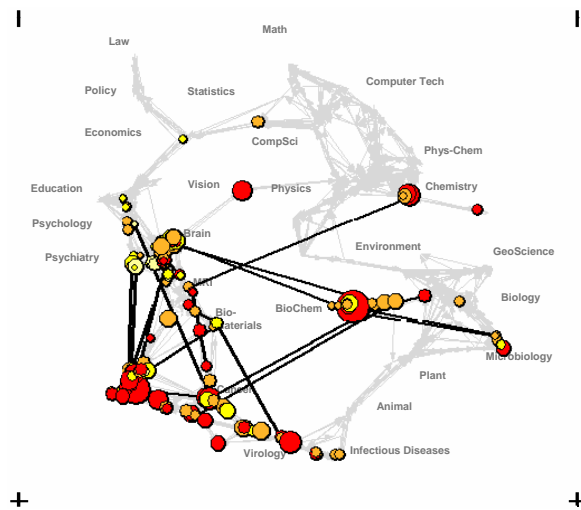*Kevin W. Boyack, Katy Börner & Richard Klavans, 2007*

Funding Patterns of the National Science Foundation (NSF)

Science map applications: Identifying core competency
Kevin W. Boyack, Katy Börner & Richard Klavans, 2007
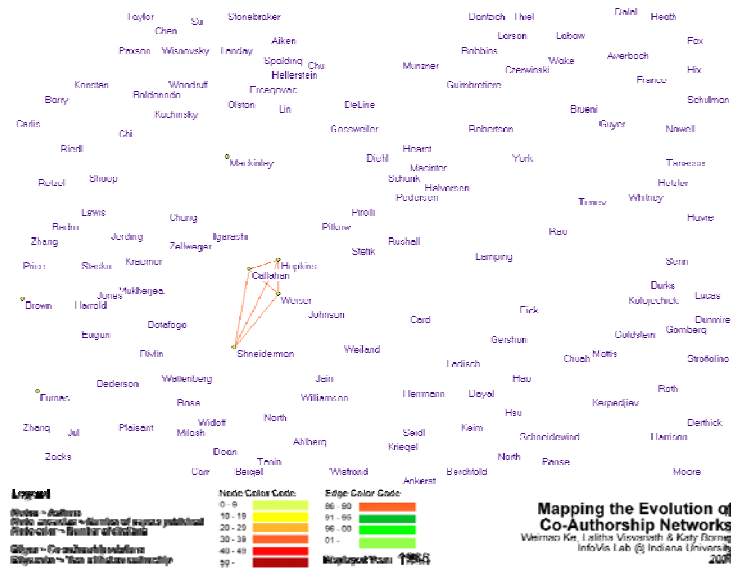
Funding Patterns of the National Institutes of Health (NIH)

13

Mapping, Managing and Utilizing
Our Collective Scholarly Knowledge

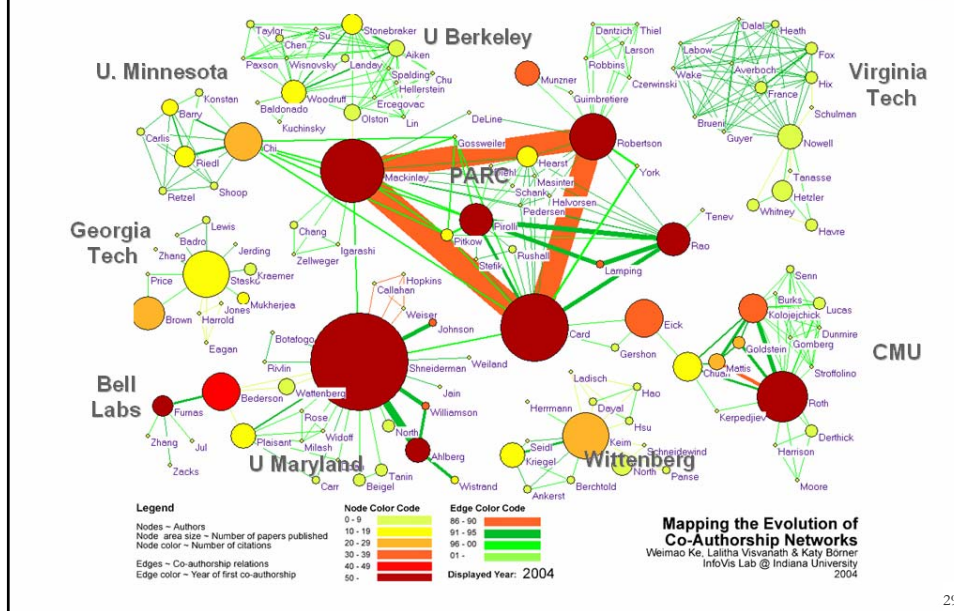**Mapping the Evolution of Co-Authorship Networks**

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

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 representation.
- Visualization of the growth of weighted co-author network.
- Centrality measures to identify author impact.
- Global statistical analysis of paper production and citations in correlation with co-authorship team size over time.
- Local, author-centered entropy measure.
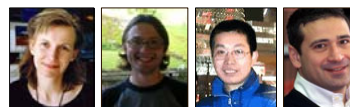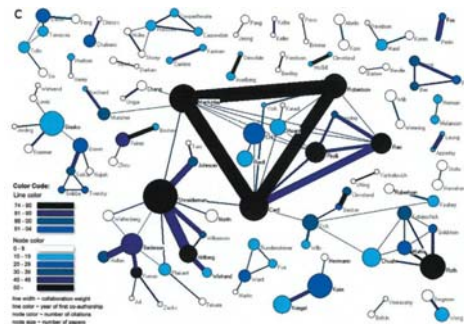
Spatio-Temporal Information Production and Consumption of Major U.S. Research Institutions

Börner, Katy, Penumarthy, 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.
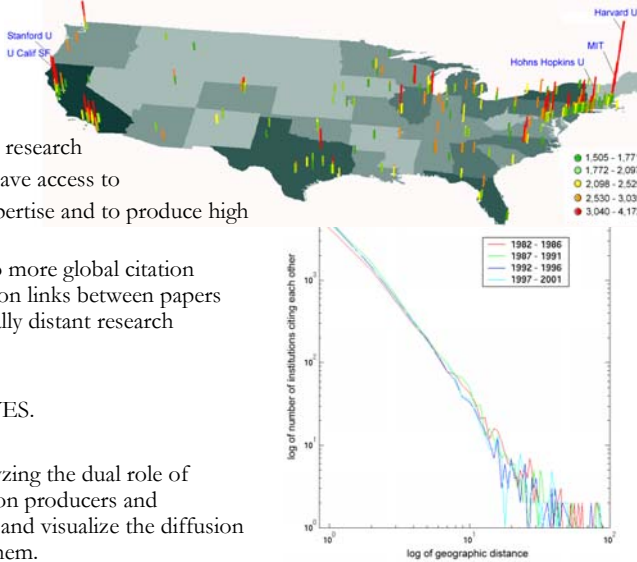
Research questions:

1. Does space still matter in the Internet age?
2. Does one still have to study and work at major research institutions in order to have access to high quality data and expertise and to produce high quality research?
3. Does the Internet lead to more global citation patterns, i.e., more citation links between papers produced at geographically distant research instructions?

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.

Places & Spaces: Mapping Science
a science exhibit that introduces people to maps of sciences, their makers and users.
http://scimaps.org

Exhibit Curators:
Dr. Katy Börner &
Julie Smith, Indiana University

Places & Spaces: Mapping Science exhibit, see also http://scimaps.org.

Our Sponsors

➢ I-IKM: "Visualizing Network Dynamics" Competition at the International Conference on Network Science 2007. NSF IIS-0724282 award (Katy Börner) April 07 - March. 08.

➢ Creative Metaphors to Stimulate New Approaches to Visualizing, Understanding, and Rethinking Large Repositories of Scholarly Data. NSF award (Katy Börner) June 07 - May 09.

➢ Mapping Science Exhibit at the 233rd National Meeting & Exposition of the American Chemical Society in Chicago, IL. NSF award (Katy Börner, March 15, 07- March 14, 08)

➢ Collaborative Research: Social Networking Tools to Enable Collaboration in the Tobacco Surveillance, Epidemiology, and Evaluation Network (TSEEN). Collaborative Systems NSF IIS-0534909 award (Katy Börner, March 15, 06 - Feb 28, 09). Collaborative proposal with Noshir S. Contractor, NCSA, Tom Finholt, University of Michigan, and Gary Giovino, University at Buffalo.

➢ Modeling the Structure and Evolution of Scholarly Knowledge. James S. McDonnell Foundation grant in area Studying Complex Systems (Katy Börner & Robert L. Goldstone) Jan. 06 - Dec. 08.

➢ SEI: NetWorkBench: A Large-Scale Network Analysis, Modeling and Visualization Toolkit for Biomedical, Social Science and Physics Research. NSF IIS-0513650 award (Katy Börner, Albert-Laszlo Barabasi, Santiago Schnell, Alessandro Vespignani & Stanley Wasserman, Eric Wernert (Senior Personnel)) Sept. 05 - Aug. 08.

➢ Center of Excellence for Computational Diagnostics. 21st Century Grant (Susanne Ragg, David Clemmer, Sven Rahmann, and Ilka Ott, Terry Vik, R Clement McDonald, Nunroe Pecock, Zina Ben Miled & Katy Börner) Sept. 04 - Aug. 07.

➢ Quest Atlantis: Advancing a Socially-Responsive Meta-Game for Learning. NSF Role-0411846 award (Sasha Barab & Susan Herring, Daniel Hickey, William Blanton, Katy Börner (Senior Personnel)) Sept. 04 - Aug. 07.

➢ CAREER: Visualizing Knowledge Domains. NSF IIS-0238261 award (Katy Börner) Sept. 03-Aug. 08.
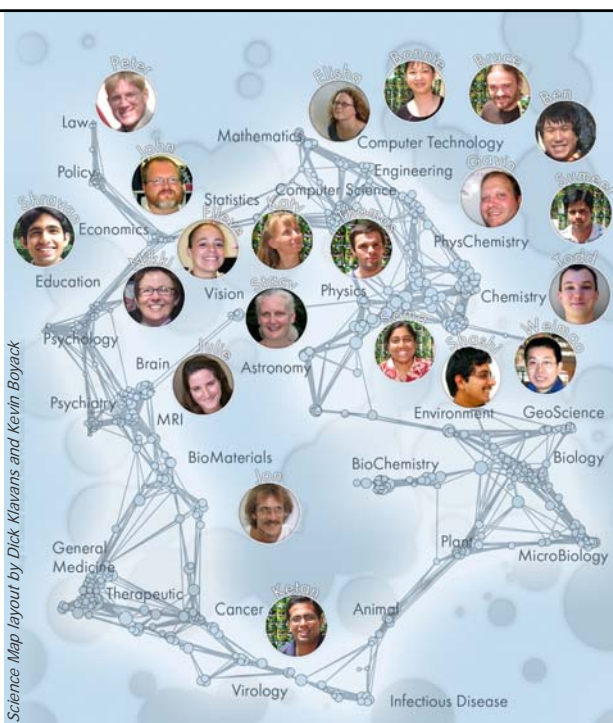
References

➢ Bruce Herr, Weixia Huang, Shashikant Penumarthy, Katy Börner. Designing Highly Flexible and Usable Cyberinfrastructures for Convergence. Submitted to William S. Bainbridge (Ed.) Progress in Convergence. Annals of the New York Academy of Sciences.

➢ Börner, Katy. Mapping All of Science: How to Collect, Organize and Make Sense of Mankind's Scholarly Knowledge and Expertise. Accepted for *Environment and Planning B*, Special Issue on *Mapping Humanity's Knowledge and Expertise in the Digital Domain*.

➢ Börner, Katy, Penumarthy, 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.

➢ Holloway, Todd, Božicevic, Miran and Börner, Katy. Analyzing and Visualizing the Semantic Coverage of Wikipedia and Its Authors. Accepted for *Complexity.* Also available as cs.IR/0512085.

➢ Katy Börner. (2006) Semantic Association Networks: Using Semantic Web Technology to Improve Scholarly Knowledge and Expertise Management. In Vladimir Geroimenko & Chaomei Chen (eds.) *Visualizing the Semantic Web*, Springer Verlag, 2nd Edition, chapter 11, pp. 183-198.

➢ Boyack, Kevin W., Klavans, R. and Börner, Katy. (2005). Mapping the Backbone of Science. *Scientometrics,* 64(3), 351-374.

➢ Hook, Peter A. and Börner, Katy. (2005) Educational Knowledge Domain Visualizations: Tools to Navigate, Understand, and Internalize the Structure of Scholarly Knowledge and Expertise. In Amanda Spink and Charles Cole (eds.) *New Directions in Cognitive Information Retrieval.* Springer-Verlag, Netherlands, chapter 5, pp. 187-208.

➢ Börner, Katy, Dall'Asta, Luca, Ke, Weimao and Vespignani, Alessandro. (April 2005) Studying the Emerging Global Brain: Analyzing and Visualizing the Impact of Co-Authorship Teams. *Complexity*, special issue on *Understanding Complex Systems*, 10(4): pp. 58 - 67. Also available as cond-mat/0502147.

➢ Ord, Terry J., Martins, Emília P., Thakur, Sidharth, Mane, Ketan K., and Börner, Katy. (2005) Trends in animal behaviour research (1968-2002): Ethoinformatics and mining library databases. *Animal Behaviour*, 69, 1399-1413. Supplementary Material.

➢ Mane, Ketan K. and Börner, Katy. (2004). Mapping Topics and Topic Bursts in PNAS. *Proceedings of the National Academy of Sciences of the United States of America*, 101(Suppl. 1):5287-5290. Also available as cond-mat/0402380.

➢ Börner, Katy, Maru, Jeegar and Goldstone, Robert. (2004). The Simultaneous Evolution of Author and Paper Networks. *Proceedings of the National Academy of Sciences of the United States of America*, 101(Suppl_1):5266-5273. Also available as cond-mat/0311459.

Science Map layout by Dick Klavans and Kevin Boyack

The End.