

# Computational Scientometrics

*Studying science by scientific means*



**Dr. Katy Börner**

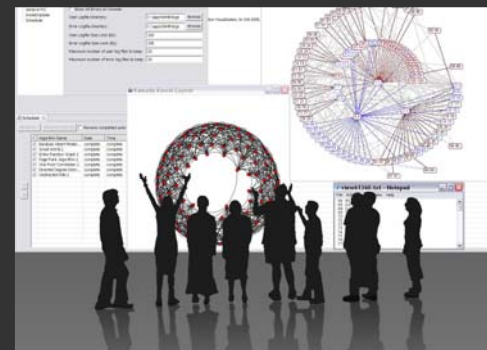
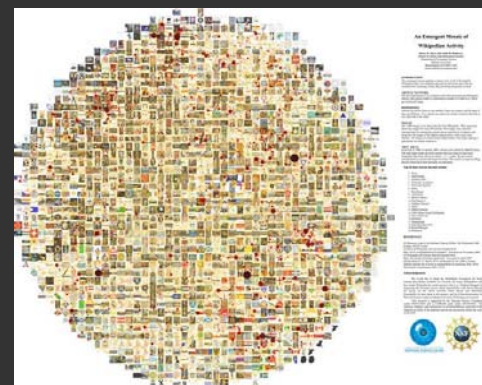
Cyberinfrastructure for Network Science Center, Director  
Information Visualization Laboratory, Director  
School of Library and Information Science  
Indiana University, Bloomington, IN

[katy@indiana.edu](mailto:katy@indiana.edu)

*Designing cyberinfrastructure to enable US-China collaboration in tobacco control and research*

*Beijing, China*

*March 27-29, 2008*



# What is the scientific aim of your research in informatics?

By example of a recent collaboration with Prof. Yanbo Han's group at the Institute of Computing Technology, Chinese Academy of Sciences

---

## ***1. Data exchange and integration***

US team will help Chinese team in querying Medline and other datasets in the SDB to, e.g., acquire all English publications by Chinese authors in a specified biology research area. The query results will be used in the NSFC supported project: *Research on Service Virtualization and Exploratory Service Composition Mechanisms for End-User Programming* (NSFC grant no. 60573117) to identify major experts, understand collaboration dynamics, and avoid duplication of research.

US team will support Chinese team in utilizing the algorithms, tools, and services available in the open source Network Workbench (NWB) tool.

US team will share multiple gene/protein datasets collected in the NWB project with the Chinese team.

## ***2. Data analysis and visualization***

US team will extract meaningful networks, analyze, visualize and compare US and Chinese data. For instance, we are interested to compare the topic coverage of tobacco research in US and in China. We will share the data analysis and visualization technologies and results with our Chinese partner and write joint publications to report at conferences on Digital Government research.

US team will assist Chinese team in the bibliometric analysis of scholarly networks related to their biomedical research.

## ***3. Technical solutions, skills, and experience exchange***

Both teams will exchange technical knowledge, skills, and experience and discuss further collaborative opportunities. US team will support Chinese team in integrating many datasets, algorithms, tools, and services available in the NWB tool to the NSFC funded VINCA (A Visual and Personalized Business-level Composition Language for Chaining Web-based Services) framework and make them available to Chinese scientists.

## ***4. Mapping Science exhibit support***

US team will bring *Mapping Science* Exhibit to Beijing to help bridge language and cultural boundaries and promote scientific, government, and education activities to the general public in China.

# How can technology improve data collection, data management, data analysis, project implementation, research collaboration, network development, and/or capacity building?

The Cyberinfrastructure for Network Science Center built and serves diverse infrastructures:

**Information Visualization CyberInfrastructure**

The InfoVis CyberInfrastructure provides access to data, software code and learning modules as well as computing resources in support of the analysis, modeling and visualization of diverse datasets.

**DATABASES**  
An Oracle database provides access to publications, grants, grants and grant opportunities. The database is continuously and automatically updated.  
<http://slis.indiana.edu>

**SOFTWARE**  
An open source IVC framework was designed to facilitate the integration of diverse data analysis, modeling and visualization algorithms. New algorithms, data persistence methods, look and feel for the interface and even entire toolkits can be easily "plugged in" or "unplugged".  
<http://slis.indiana.edu>

**COMPUTING RESOURCES**  
The InfoVis CyberInfrastructure is hosted at Indiana University's Research Database Complex, comprising of two Sun V1290 servers with 12 900MHz processors and 56 GB of memory each. 6 TB Fibre Channel disks are attached to both servers. A Sun VMIO system with 4 CPUs and 6GB memory system is the web front-end for the database servers.  
<http://slis.indiana.edu>

**LEARNING MODULES**  
A set of associated learning modules aims to equip learners with a practical skill set by providing code and advice to quickly modify and run different algorithms, test diverse interaction techniques and design features, and to quickly generate and compare information visualizations.  
<http://slis.indiana.edu>

**InfoVis Lab**

- People
- Research
- Teaching
- Publications
- Presentations

- Calls & Events
- Datasets
- Software
- Hardware
- Funding

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

**CIShell**  
cyberinfrastructure shell

**SDB** SCHOLARLY DATABASE

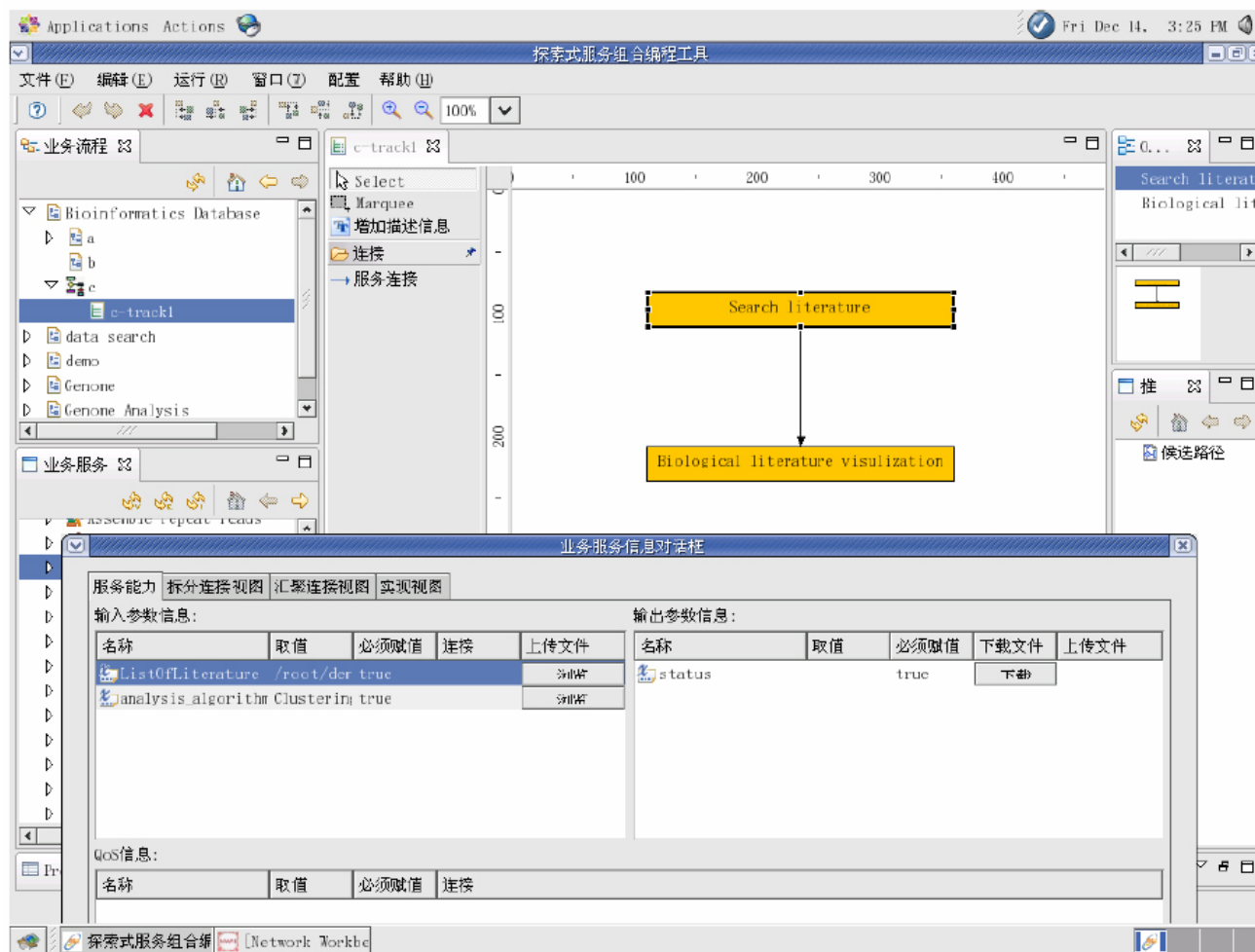
Scholarly Database  
<http://sdb.slis.indiana.edu>

**Network Workbench**  
A Workbench for Network Scientists

SEI: Network Workbench: A Large-Scale Network Analysis, Modeling and Visualization Toolkit for Biomedical, Social Science and Physics Research. NSF IIS-0513650 award (Katy Börner, Albert-László Barabási, Santiago Schnell, Alessandro Vespignani & Stanley Wasserman, Eric Wernert (Senior Personnel), Sept. 05 - Aug. 08. <http://nw/b.slis.indiana.edu>

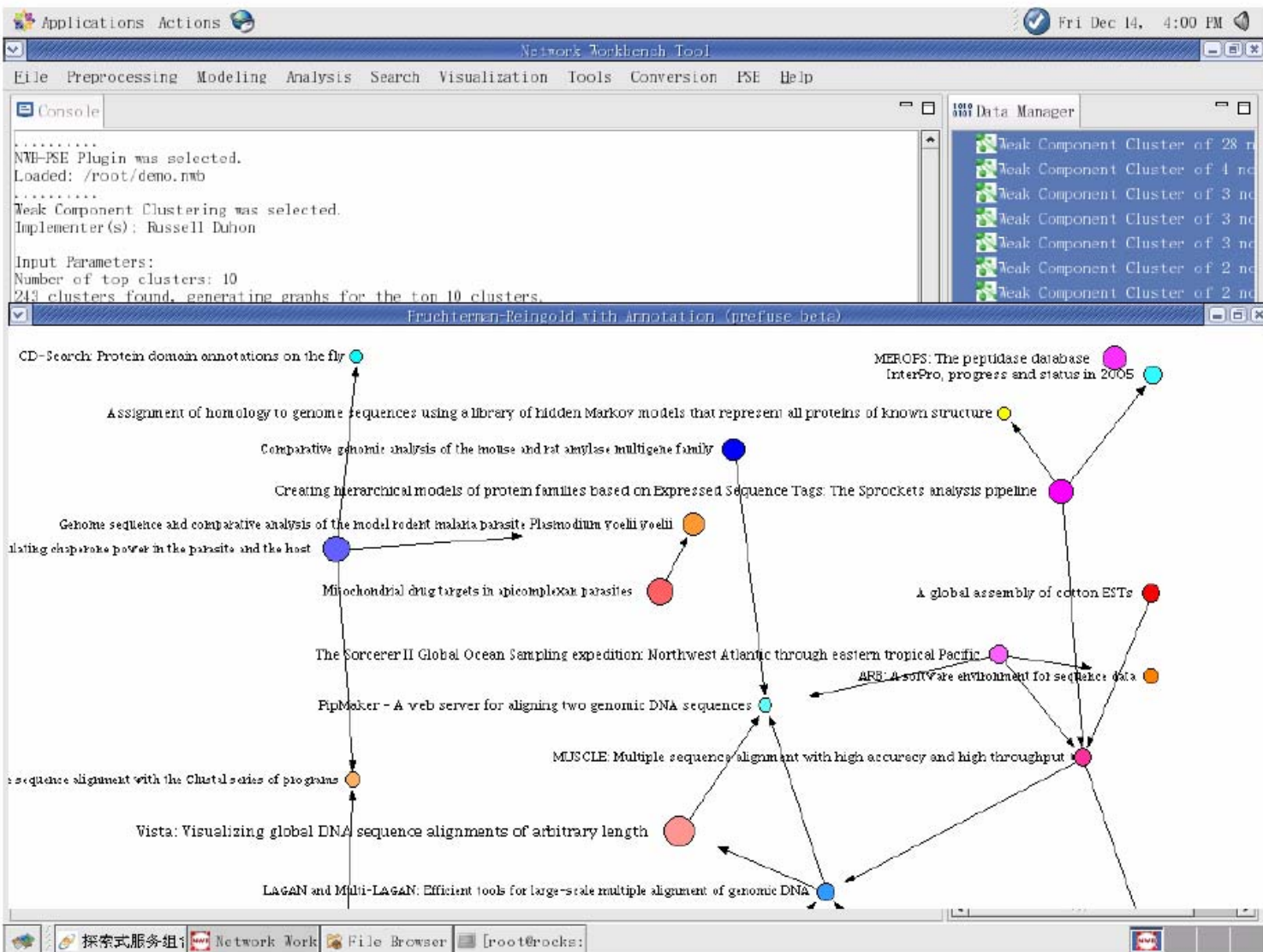
NSF SBC Sun Microsystems HPNAP High Performance Network Applications Program INDIANA 21st Century Research & Technology Fund

Provide case examples demonstrating how technology you developed addressed research and project problems in other research areas or in tobacco control.



**Figure1:** Screenshot of the VINCA4Science PSE tool. Biologists can now use this tool to analyze relevant literature before conducting a biological experiment. They simply use PSE to construct a workflow including searching related literature, extracting co-author or paper-citation networks from search results, and identifying the most relevant and important publications. While PSE acts as an easy to use front end, all algorithms required for the analysis and visualization of scholarly datasets are provided by the NWB tool.





**Figure 2:** The VINCA4Science PSE tool sends the dataset and the configuration parameters to the NWB tool. The PSE communication channel is used to invoke a set of algorithms in the NWB tool. Biologists can now analyze relevant literature and design biological experiments without leaving the VINCA4Science PSE tool.

# What were your learning experiences for future efforts?

It is highly beneficial to combine the best of both worlds/cultures.

We are very interested to continue the collaboration and are looking forward to bring the Mapping Science exhibit <http://scimaps.org/nslc/> to China in May/June 2008.



## CONTACT:

Exhibit Curators:

**Dr. Katy Börner** <[katy@indiana.edu](mailto:katy@indiana.edu)> and Elisha Hardy <[efhardy@indiana.edu](mailto:efhardy@indiana.edu)>  
Indiana University, Bloomington, IN

Exhibit Advisor:

Weixia (Bonnie) Huang <[huangb@indiana.edu](mailto:huangb@indiana.edu)>

## WHERE:

**The National Science Library of the Chinese Academy of Sciences**

BeiCiHuanXiLu, No.33

Beijing, China 100080

Phone: 010-82626684

**WHEN:** May 17th, 2008 - June 30th, 2008.