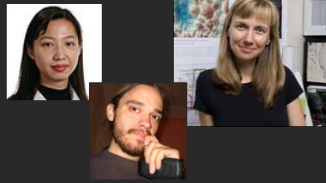
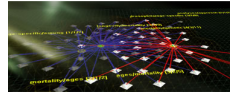


## User Friendly Cyberinfrastructures

Dr. Katy Börner / Bruce Herr & Weixia (Bonnie) Huang  
 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)



Workshop at the Cambridge Colloquium on Complexity and Social Networks  
 Harvard University, Boston, MA.  
 September 21, 2006

## Questions

- Did you attend the Network Science Conference & Workshop 2006?
- What datasets do you use in your research? What is their size? Any NWs?
- What tools/software packages/scripting languages do you use? Any vis?
- Do you program? Develop new algorithms?
- Do you share your algorithms? How?



Katy Börner, Workshop on User-Friendly Cyberinfrastructures

**This Talk has Three Parts:**

1. Why Develop CI in a InfoVis Lab / NetSci Center?
2. Cyberinfrastructure Shell (CIShell) Framework
3. Utilizing CIShell to Serve Different Communities
  - Information Visualization CI
  - Network Workbench



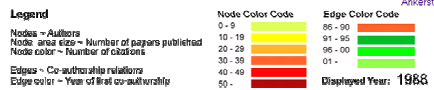
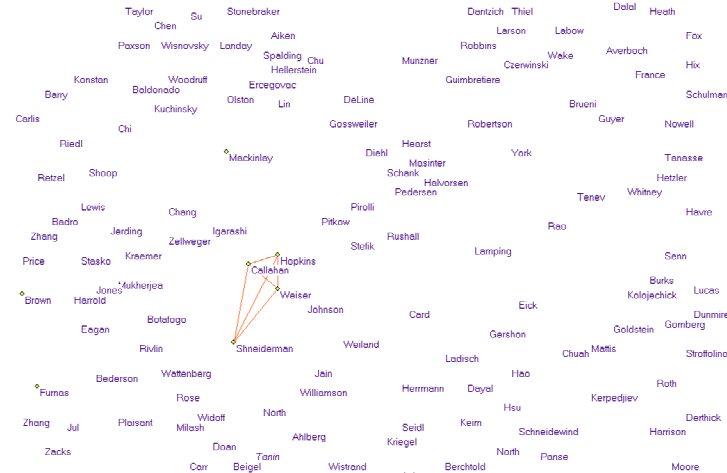
**This Talk has Three Parts:**

- 1. Why Develop CI in a InfoVis Lab / NetSci Center?**
2. Cyberinfrastructure Shell (CIShell) Framework
3. Utilizing CIShell to Serve Different Communities
  - Information Visualization CI
  - Network Workbench



# Mapping the Evolution of Co-Authorship Networks in Information Visualization, 1988 - 2004

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

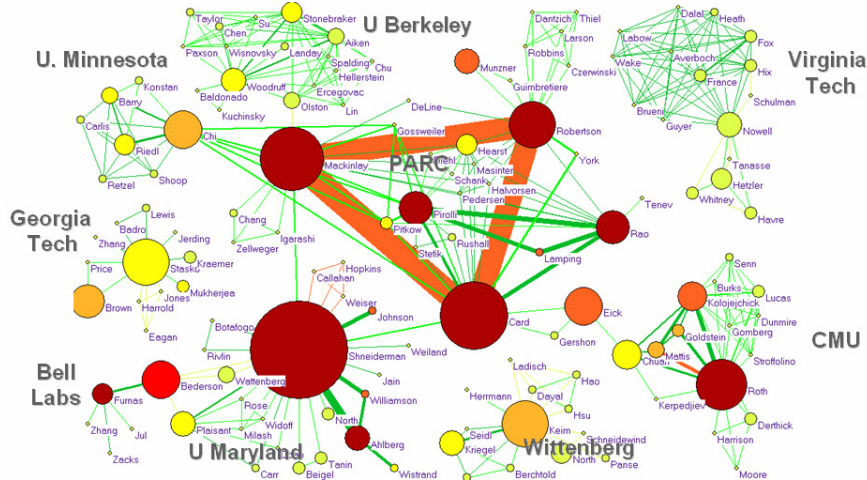


**Mapping the Evolution of Co-Authorship Networks**  
Weimao Ke, Lalitha Viswanath & Katy Börner  
InfoVis Lab @ Indiana University  
2004

Katy Börner, Workshop on User Friendly Cyberinfrastructures

# Mapping the Evolution of Co-Authorship Networks

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

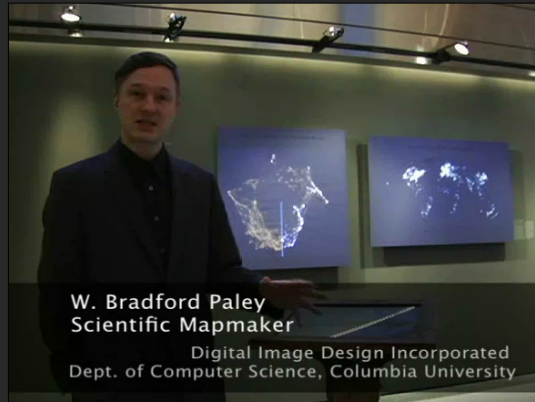


**Mapping the Evolution of Co-Authorship Networks**  
Weimao Ke, Lalitha Viswanath & Katy Börner  
InfoVis Lab @ Indiana University  
2004

## 'Places & Spaces: Mapping Science' Exhibit

<http://scimaps.org/exhibit>

### Illuminate Diagram Display



W. Bradford Paley  
Scientific Mapmaker  
Digital Image Design Incorporated  
Dept. of Computer Science, Columbia University

TOPIC MAP: HOW SCIENTIFIC PARADIGMS RELATE

GEOGRAPHIC MAP: WHERE SCIENCE GETS DONE

You may run your finger over each of these maps to control the lighting on the other: touching a place on the world map will light up topics studied in that place; touching a paradigm on the topic map will light up the places that study that topic.

**Nanotechnology**

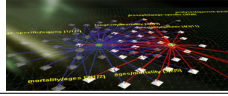
This overlay shows the distribution of nanotechnology within the paradigms of science. The majority of current work in nanotechnology takes place in physics, chemistry, and materials science, at the upper right portion of the map. However, an increasing amount of nanotechnology is being applied in the biological and medical sciences, at the lower right.

<b>All Topics</b>	<b>Nanotechnology</b>	<b>Francis H. C. CRICK</b>	<b>Albert EINSTEIN</b>	<b>Michael E. FISHER</b>	<b>Susan T. FISKE</b>
<i>Sweep through all 776 scientific paradigms</i>	<i>Science on the tiny scale of molecules</i>	<i>Co-discovered DNA's double helix</i>	<i>Revitalized physics with Relativity theories</i>	<i>Models critical phase transitions of matter</i>	<i>Connects perception and stereotypes</i>
<b>Sustainability</b>	<b>Biology &amp; Chemistry</b>	<b>Joshua LEDERBERG</b>	<b>Derek J. de Solla PRICE</b>	<b>Richard N. ZARE</b>	<b>About this display</b>
<i>The science behind our long-term hopes</i>	<i>The interface between these two vital fields</i>	<i>Pioneer in bacterial genetic mechanisms</i>	<i>Known as the "Father of Scientometrics"</i>	<i>Uses laser chemistry in molecular dynamics</i>	<i>People &amp; organizations that helped create it</i>

*We sweep slowly through adjoining related topics, lighting up the places in the world that study each topic. You may select a subset of the topics that deal with these three interesting subjects by touching it.*

*A single person's spreading influence is shown as a series of four snapshots. First, we light only topics and places relating to that person's papers—papers that are still highly cited today. The second lights everything that cites that original work. Note that this first-generation impact extends to far more topics than did the original work. The third snapshot lights science that cites the second; and the fourth lights science that cites the third.*





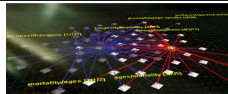
## Why Do we Need Cyberinfrastructures?

### Problem

- There are too many and too complex datasets that need to be correlated and understood to arrive at the best possible decisions.
- There are too many different data formats, different algorithms, different implementations of the same algorithm, different programming languages, different research purposes (modeling, analysis, visualization), different communities and practices.
- The analysis, modeling, and visualization of large datasets requires powerful computing infrastructures.
- Managing 1000+ of different data sets and 100+ of different algorithms requires a means to quickly select the best dataset(s)/algorithm(s).

### Needed is a socio-technical cyberinfrastructure that supports

Easy access to datasets and algorithms, computer resources, their descriptions, and associated learning modules and access to expertise.



## Cyberinfrastructure Shell (CIShell)

<http://cishell.org>



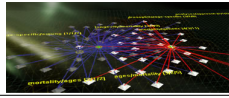
CIShell is an 'empty shell' that supports

- Easy integration of new datasets and algorithms by algorithm developers and
- Easy usage of algorithms by algorithm users.

Its plug-and-play architecture supports the integration and utilization of diverse

- Datasets, e.g., stored in files, databases, streaming 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 – Technical Details



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

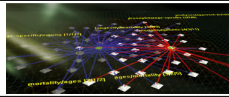
**OSGi** (<http://www.osgi.org>) is

- 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 be connected via RPC/RMI supporting peer-to-peer sharing of data, algorithms, and computing power.

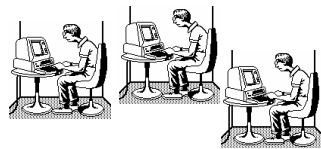
Ideally, CIShell becomes a standard for creating OSGi Services for algorithms. Developed Tools/CI, e.g., IVC & NWB, provide a reference GUI for underlying services.



## Serve Algorithms Developers & Users



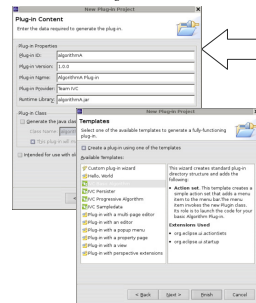
### Developers



### Users



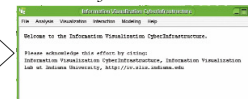
### CIShell Wizards



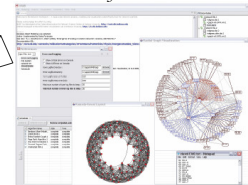
### CIShell

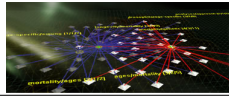


### IVC Interface



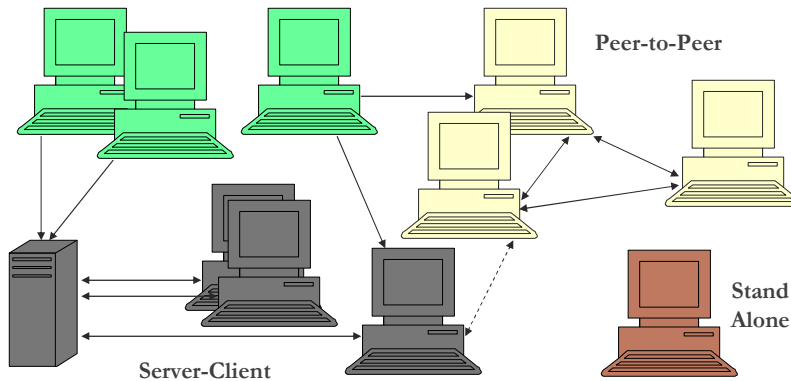
### NWB Interface





## Deployment

### Data-Algorithm Repositories



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

*Katy Börner, Workshop on User Friendly Cyberinfrastructures*

15

### This Talk has Three Parts:

1. Why Develop CI in a InfoVis Lab / NetSci Center?
2. Cyberinfrastructure Shell (CIShell) Framework
3. Utilizing CIShell to Serve Different Communities
  - Information Visualization CI
  - Network Workbench





### 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 data sets.

**DATABASES**  
An Open-Database provides access to publications, posters, grants and grant opportunities. The database is continuously and automatically updated.  
<http://www.infovis.indiana.edu/>

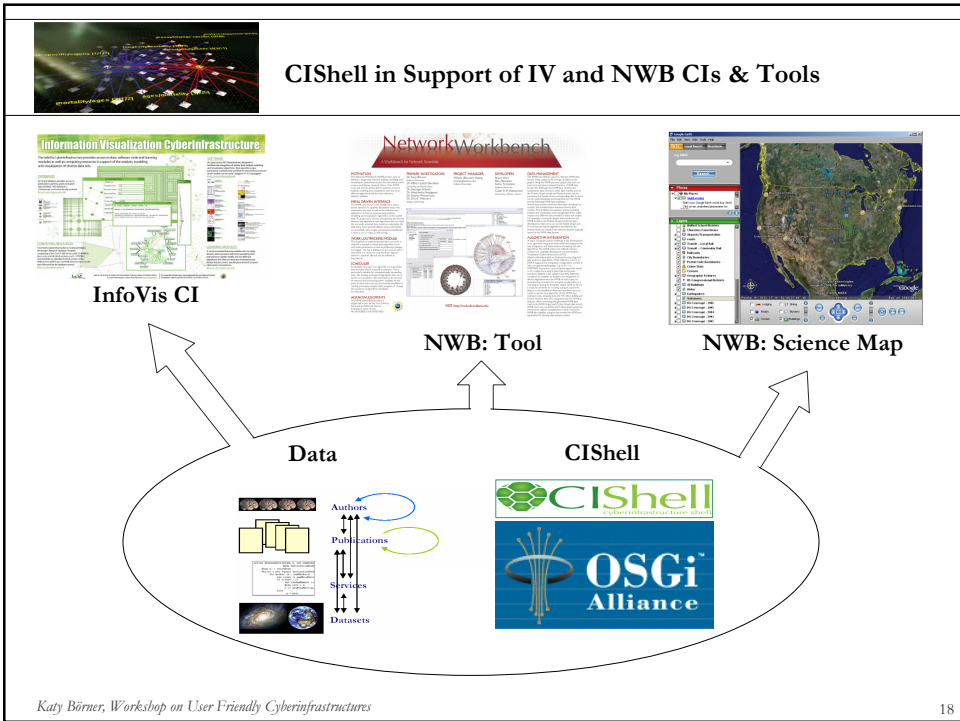
**COMPUTING RESOURCES**  
The InfoVis CyberInfrastructure is located at Indiana University's Research Database Complex, consisting of three Sun T2000 servers with 124 nodes, processors and 48 GB of memory each. A 72-hour cluster file system is available to all users. A team of system administrators is available to assist with user needs. For more information, please visit our web page at <http://www.infovis.indiana.edu/>

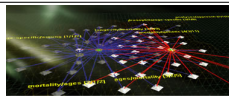
**SOFTWARE**  
An open source MVC framework was designed to facilitate the integration of diverse data analysis, modeling and visualization algorithms. New algorithms, data performance methods, look and feel for the interface and new user tools can be easily integrated in our framework.  
<http://www.infovis.indiana.edu/>

**LEARNING MODULES**  
A set of associated learning modules aims to equip human users in general and to assist in providing public and advice to specific users and use different algorithms and diverse research techniques and design features, used to quickly generate and compare information visualizations.

*CAREER: Visualizing Knowledge Domains. NSF IIS-0238261 award (Katy Börner, \$451,000) Sept. 03-Aug. 08.  
<http://in.slis.indiana.edu/>*

*SEI: Network Workbench: A Large-Scale Network Analysis, Modeling and Visualization Toolkit for Biomedical, Social Science and Physics Research. NSF IIS-0313650 award (Katy Börner, Albert-Laszlo Barabasi, Santiago Schnell, Alessandro Vespignani & Stanley Wasserman, Eric Wernert (Senior Personnel), \$1,120,926) Sept. 05 - Aug. 08.  
<http://nwb.slis.indiana.edu>*





## InfoVis Cyberinfrastructure

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

# 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 data sets.

### DATABASES

An Oracle database provides access to publications, patents, grants and grant opportunities. The database is continuously and automatically updated. (<http://iv.slis.indiana.edu/db>)



### COMPUTING RESOURCES

The InfoVis Cyberinfrastructure is hosted at Indiana University's Research Database Complex comprising of two Sun V1280 servers with 12 900MHz processors and 96 GB of memory each. 6 TB fiber channel disks are attached to both servers. A Sun V880 system with 4 cpus and 8GB memory serves as the web front-end for the database servers. (<http://iv.slis.indiana.edu/cr>)

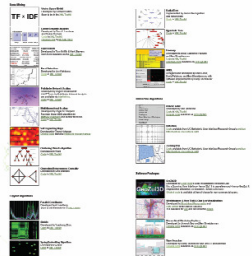


InfoVis Lab, School of Library and Information Science, Indiana University (2001). For more information, contact Katy Börner at [kborner@indiana.edu](mailto:kborner@indiana.edu)

This material is based upon work supported by the National Science Foundation under Grant No. 05-0238261 and DUE-0338623.

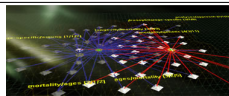
### SOFTWARE

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 feels for the interface and entire toolkits can be easily "plugged in" or "unplugged". (<http://iv.slis.indiana.edu/ivf>)



### 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://iv.slis.indiana.edu/lm>)



## IVC Database (<http://iv.slis.indiana.edu/db>)

### Papers and Patents



**Medline**  
Number of Entries: 11,883,477  
Years covered: 1963-2002  
Size: 135 MB (gunzipped)



**Proceedings of the National Academy of Science (PNAS)**  
Number of Entries: 16,189  
Years covered: 1997-2002  
Size: 583 MB



**United States Patent and Trademark Office (Patents)**  
Number of Entries: 2,582,847  
Years covered: 1976-2003  
Size: 350 MB

### Grant Awards



**National Science Foundation (NSF)**  
Number of Entries: 181,132  
Years covered: 1985-2002  
Size: 400 MB



**National Institute of Health (NIH)**  
Number of Entries: 1,003,521  
Years covered: 1972-1992 and 1994-2002  
Size: 2.3 GB

### Funding Opportunities



**Community of Science (COS)**  
Number of Entries: 38,154 (5,000 new entries per month)  
Years covered: 2001-present  
Size: 60 MB

# 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 data sets.

## DATABASES

An Oracle database provides access to publications, patents, grants and grant opportunities. The database is continuously and automatically updated. (<http://ivis.Indiana.edu/db>)



## COMPUTING RESOURCES

The InfoVis CyberInfrastructure is hosted at Indiana University's Research Database Complex comprising of two Sun V1280 servers with 12 900MHz processors and 96 GB of memory each. 6 TB fiber channel disks are attached to both servers. A Sun V880 system with 4 cpus and 8GB memory serves as the web front-end for the database servers. (<http://ivis.Indiana.edu/cr>)



## 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 feels for the interface and even entire toolkits can be easily "plugged in" or "unplugged". (<http://ivis.Indiana.edu/iv>)

...to equip ...ing code ...ferent algorithms, test diverse interaction techniques and design features, and to quickly generate and compare information visualizations. (<http://ivis.Indiana.edu/ivm>)



InfoVis Lab, School of Library and Information Science, Indiana University (2004). For more information, contact Katy Börner at [katy@indiana.edu](mailto:katy@indiana.edu)

This material is based upon work supported by the National Science Foundation under Grant No. IS-0228261 and DUE-0339623.



Poster Design by Carolea Gosselin, DMA, [carolea@indiana.edu](mailto:carolea@indiana.edu)

# 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 data sets.

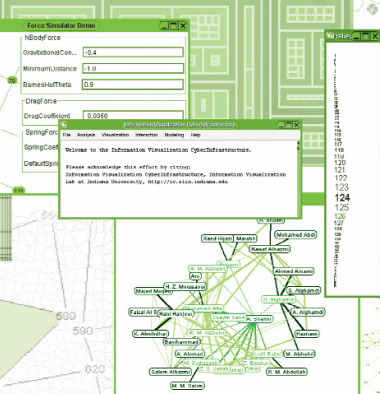
## DATABASES

An Oracle database provides access to publications, patents, grants and grant opportunities. The database is continuously and automatically updated. (<http://ivis.Indiana.edu/db>)



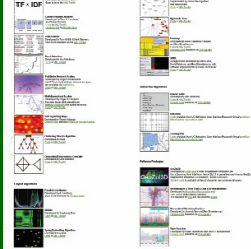
## COMPUTING RESOURCES

The InfoVis CyberInfrastructure is hosted at Indiana University's Research Database Complex comprising of two Sun V1280 servers with 12 900MHz processors and 96 GB of memory each. 6 TB fiber channel disks are attached to both servers. A Sun V880 system with 4 cpus and 8GB memory serves as the web front-end for the database servers. (<http://ivis.Indiana.edu/cr>)



## 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 feels for the interface and even entire toolkits can be easily "plugged in" or "unplugged". (<http://ivis.Indiana.edu/iv>)



## 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://ivis.Indiana.edu/ivm>)



InfoVis Lab, School of Library and Information Science, Indiana University (2004). For more information, contact Katy Börner at [katy@indiana.edu](mailto:katy@indiana.edu)

This material is based upon work supported by the National Science Foundation under Grant No. IS-0228261 and DUE-0339623.



Poster Design by Carolea Gosselin, DMA, [carolea@indiana.edu](mailto:carolea@indiana.edu)

# 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 data sets.

## DATABASES

An Oracle database provides access to publications, patents, grants and grant opportunities. The database is continuously and automatically updated. (<http://ivslis.indiana.edu/db>)



## COMPUTING RESOURCES

The InfoVis CyberInfrastructure is hosted at Indiana University's Research Database Complex comprising of two Sun V1280 servers with 12 900MHz processors and 96 GB of memory each. 6 TB fiber channel disks are attached to both servers. A Sun V880 system with 4 CPUs and 8GB memory serves as the web front-end for the database servers. (<http://ivslis.indiana.edu/cr>)

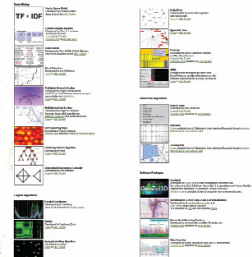


InfoVis Lab, School of Library and Information Science, Indiana University (2004). For more information, contact Katy Börner at [katyb@indiana.edu](mailto:katyb@indiana.edu)

This material is based upon work supported by the National Science Foundation under Grant No. IS-0228261 and CUE-0339623.

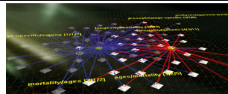
## SOFTWARE

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 feels for the interface and even entire toolkits can be easily "plugged in" or "unplugged". (<http://ivslis.indiana.edu/iv>)



## 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://ivslis.indiana.edu/lm>)



## Time Series Analysis



[Learning Modules](#) > Visualizing Time Series Data

[Description](#) | [Usage Hints](#) | [Learning Task](#) | [Discussion](#) | [References](#) | [Acknowledgments](#)

### Learning Module

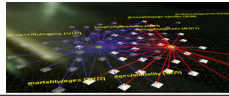
<http://iv.slis.indiana.edu/lm/lm-time-series.html>

#### Description

A time series is a sequence of events/observations which are ordered in one dimension, e.g., time. Frequently, successive observations depend on each other and it makes sense to display them in a (time) sorted fashion, e.g., as a scatter plot. Alternatively, one could be interested to know how many observations of a certain value have been made. Here one would sort the observations by value, count the number of observations for each value and derive a histogram. Time series data can be continuous, i.e., there is an observation at every instant of time see figure below, or discrete, i.e., observations exist for regularly or irregularly spaced intervals.



Time series are recorded, analyzed and used in diverse domains of science. Check out the [Time Series Data Library](#) maintained by Rob Hyndman and Muhammad Akram for numerous data sets from Agriculture, Chemistry, Crime, Demography, Ecology, Finance, Health, Hydrology, Industry, Labour market, Macro-Economics, Meteorology, Micro-Economics, Physics, Production, Sales, Simulated series, Sport, Transport & Tourism or Utilities.



## Visualizing Tree Data

### Learning Module

<http://iv.slis.indiana.edu/lm/lm-trees.html>



[Learning Modules](#) > Visualizing Tree Data

[Description](#) | [Usage Hints](#) | [Learning Task](#) | [Discussion](#) | [References](#) | [Acknowledgments](#)

#### Description

Many data sets come in tree format. There are family trees, organizational charts, classification hierarchies, and directory structures. The figure below shows an inheritance tree by Ernst Haeckel ('Stammbaum' in German). Read also [To Draw a Tree](#) by Pat Hanrahan.



[Click image for larger version](#)

A tree graph is a set of straight line segments (edges) connected at their ends containing no closed loops (cycles). You can also call it a simple, undirected, connected, acyclic graph (or, equivalently, a connected forest). A tree with  $n$  nodes has  $n-1$  graph edges. All trees are bipartite graphs.

Many trees have a root node and are called rooted trees. Trees without a root node are called free trees. Subsequently, we will only consider rooted trees. In rooted trees, all nodes except the root node have only one parent node. Nodes which have no children are called leaf nodes. All other nodes are referred to as intermediate nodes.

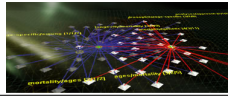
*Katy Börner, Workshop on User Friendly Cyberinfrastructures*

25

### This Talk has Three Parts:

1. Why Develop CI in a InfoVis Lab / NetSci Center?
2. Cyberinfrastructure Shell (CIShell) Framework
3. **Utilizing CIShell to Serve Different Communities**
  - Information Visualization CI
  - **Network Workbench**





## Network Workbench



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



**Software Team:** Team Lead: Weixia (Bonnie) Huang  
Software Developers: Bruce Herr & Ben Markines  
Algorithm Developers: Santo Fortunato & Cesar Hidalgo



**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>



*Katy Börner, Workshop on User Friendly Cyberinfrastructures*

27

# NetworkWorkbench

A Workbench for Network Scientists

### MOTIVATION

The Network Workbench (NWB) project aims to develop a large-scale network analysis, modeling and visualization cyberinfrastructure for biomedical, social science, and physics research. Users of the NWB tools and platform will be able to perform network analysis, modeling and visualization with the most effective algorithms and the best reference datasets available.

### MENU DRIVEN INTERFACE

The NWB tool shown in the middle has a menu-driven interface. It supports file/dataset load, view, conversion, and save as well as the selection and application of diverse preprocessing, analysis, modeling, and visualization algorithms on the loaded data. To guide users' choices among many and diverse datasets and algorithms, only algorithms that can read the currently selected data model are selectable. All data entry forms provide default values, information on acceptable value ranges, instantaneous feedback if a value is out of range, as well as help.

### WORK LOG TRACKING MODULE

The sequence of steps performed by a user such as what file is loaded or saved, what algorithm is run with what parameters, as well as preference changes are logged. The log is displayed in the console and is also saved as a record in a log file. Error logs are saved in a separate file and can be utilized as log reports.

### SCHEDULER

A scheduler lets users run algorithms at a particular date and time and in a specified sequence. This is particularly suitable for computationally demanding jobs. The number and type of algorithms that run in series or in parallel is only restricted by the amount of memory and processing power available. At any point in time, users can see all currently scheduled or running processes, monitor their progress, or change the sequence of algorithms scheduled for execution.

### ACKNOWLEDGMENTS

The NWB cyberinfrastructure is supported in part by the 21st Century Fund and the National Science Foundation under Grants No. IIS-0230261 and IIS-0513650.



### PRIMARY INVESTIGATORS

Dr. Katy Börner  
Indiana University  
Dr. Albert-Laszlo Barabasi  
University of Notre Dame  
Dr. Santiago Schnell  
Dr. Alessandro Vespignani  
Dr. Stanley Wasserman  
Dr. Eric A. Wernert  
Indiana University

### PROJECT MANAGER

Weixia (Bonnie) Huang  
(huang@slis.indiana.edu)  
Indiana University

### DEVELOPERS

Bruce Herr  
Ben Markines  
Santo Fortunato  
Cesar A. H. Ramaciotti  
University of Notre Dame

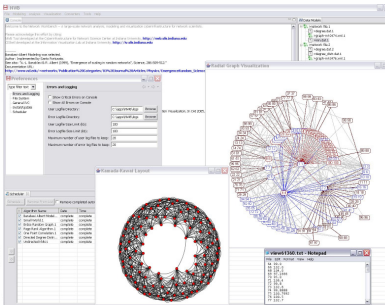
### DATA MANAGEMENT

The NWB tool delivers a generic, efficient NWB data format which supports the storage of million node graphs. Using the NWB persistent plugin, the tool can load, view, and save a network through a NWB data format file. Although the NWB data model is the fundamental data structure, other data models, such as the Prefuse Graph model and Matrix model, and the persisters that handle those corresponding data formats can be easily developed and integrated into the NWB tool by following NWB data templates.

Several data model converters have been developed to conduct the transformation between diverse data models. This facilitates the pipeline of data modeling, analysis, and visualization even though algorithms might require very different data models for input and output. For example, a converter plugin that transforms the NWB model to the Prefuse Graph model has been developed so that users can use the Radial Graph and Force Directed Layout algorithms provided by the Prefuse library to visualize the network dataset originally stored in the NWB data format.

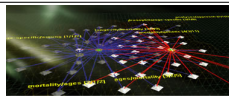
### ALGORITHM INTEGRATION

A major computer science challenge is the development of an algorithm integration framework that supports the easy integration and dissemination of existing and new algorithms. The NWB utilizes the CytoShell software architecture originally developed in the Information Visualization Cyberinfrastructure (IVC) (<http://india.indiana.edu>) to facilitate the easy plug and play of diverse algorithms. While CytoShell is written in Java, it supports the integration of algorithms written in other programming languages, e.g., in C++ or FORTRAN. In practice, a pre-compiled algorithm needs to be wrapped as a plugin that implements basic interfaces defined in the CytoShell Core APIs. Different templates are available to facilitate the integration of diverse algorithms into the NWB. In most cases, no programming is required to integrate an algorithm as a new plugin. A plugin developer simply needs to fill out a sequence of methods for creating a plugin, export the plugin to the installation directory, and then users are ready to use the new algorithm via the NWB tool interface menu. Drawing from the IVC, other JUNG and Prefuse libraries have been integrated into the NWB as plugins. After converting the generalized NWB data model into JUNG Graph and Prefuse Graph data model, NWB users can run JUNG and Prefuse graph layouts to interactively explore visualizations of their networks. NWB also supplies a plugin that invokes the XINGrace application for plotting data analysis results.



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

28



## NWB Tool: Interface Elements

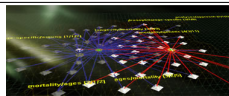


The screenshot shows the NWB Tool interface with several key elements labeled:

- Load Data:** Points to the 'File' menu.
- Select Preferences:** Points to the 'Preferences' dialog box.
- List of Data Models:** Points to the 'Data Models' panel on the right.
- Console:** Points to the 'Console' window at the bottom left.
- Scheduler:** Points to the 'Scheduler' window at the bottom left.
- Visualize Data:** Points to the 'Visual Graph Visualization' window showing a network graph.
- Open Text Files:** Points to the 'Open Text Files' window at the bottom right.

Katy Börner, Workshop on User Friendly Cyberinfrastructures

29



## List of Algorithms (partially implemented)



### Modeling

Random Network Model  
Random

### Preferential Attachment Algorithms

Barabasi-Albert Model  
Dorogovtsev-Mendes-Samukhin  
Fitness  
Vertices/edges deletion  
Copying strategy  
Finite vertex capacity  
TARL

### Rewiring algorithms

Rewiring based on degree distribution  
Watts Strogatz Small World Model

### Peer-to-Peer Models

Structured  
CAN Model  
Chord Model

### Unstructured

PRU Model  
Hypergrid Model

### Measurement

Edge/Node level  
node degree  
BC value of nodes/edges  
Max flow edge  
Hub/Authority value for nodes  
Distribution of node distances (Hop plot)

### Local (directed and weighted versions)

Clustering Coefficient (Watts Strogatz)  
Clustering Coefficient (Newman)  
k-Core Count  
Distributions (Plot and gamma, and  $R^2$ )  
Degree Distributions (in, out, total) (Directed/Total Degree Distribution)  
Degree Correlations (in-out, out-out, out-in, in-in, total-total)  
Clustering Coefficient over k  
Coherence for weighted graphs  
Distribution of weights  
Probability of degree distribution

### Global

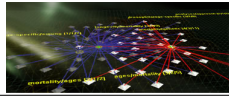
Density  
Square of Adjacency Matrix  
Giant Component  
Strongly Connected Component  
Betweenness Centrality  
Diameter  
Shortest Path = Geodesic Distance  
Average Path Length

What other algorithms do you use?

Motif Identification  
Page Rank  
Closeness centrality  
Reach centrality  
Eigenvector centrality  
Minimum Spanning Tree

Katy Börner, Workshop on User Friendly Cyberinfrastructures

30



## List of Algorithms (partially implemented)



### Basic Processes on Networks

#### Search

k Random-Walk Search  
 Depth First Search  
 p-rand Breadth-First Search  
 P2P  
 CAN Search  
 Chord Search

#### Epidemics Spreading

SIR  
 SIS

#### Graph Matching

Simple Match  
 Similarity Flooding  
 ABSURDIST

### Clustering

#### Based on Attributes

Hierarchical Clustering  
 Single Link  
 Complete Link  
 Average Link  
 Ward's Algorithm

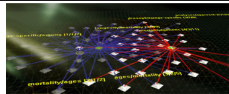
#### Based on Network Structure

Newman Girvan  
 Clauset-Newman-Moore  
 Newman  
 Ceconi-Parisi  
 Simulated annealing of modularity  
 Caldarelli  
 Weak Component Clustering  
 vanDongen (random walk)  
 Cfinder (Clique percolation method)  
 Reichardt, Bornholdt (q-potts model)

### Visualization

Distribution  
 Scatterplot  
 Histogram  
 Geospatial  
 Circle layout  
 Grid-based  
 Dendrogram  
 Treemap  
 Hyperbolic tree  
 Radial Tree  
 Sparse Matrix Visualization  
 Kamada-Kawaii  
 Fruchterman-Rheingold  
 Orthogonal Layout  
 k-core visualization

What other algorithms do you use?



## Download

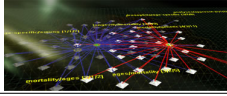
- Download the IVC from <http://sourceforge.net/projects/ivc>
- Click 
- Download the ivc.1.0.4.full-installer.jar (with plugins).
- Install.
- Run.
- Enjoy.

### Note:

If you have a Mac and the most recent JVM installed then you can download the empty IVC shell without plugins and get super old plugins from the update site at <http://iv.slis.indiana.edu/ivc/update>.

Or you wait for the new release that will be available on October 2<sup>nd</sup>, 2006.





## 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. <http://ella.slis.indiana.edu/~katy/paper/06-cishell.pdf>
- Börner, Katy, Sanyal, Soma and Vespignani, Alessandro. Network Science: A Theoretical and Practical Framework. (in press) In Blaise Cronin (Ed.), *Annual Review of Information Science & Technology*, Volume 41, Medford, NJ: Information Today, Inc./American Society for Information Science and Technology. <https://ivl.slis.indiana.edu/publications/borner-2006-netsci.pdf>
- See also publications linked from Network Workbench web site at <http://nwb.slis.indiana.edu/>

### For further information please contact:

Weixia (Bonnie) Huang, NWB Software Team Lead, Indiana University, [huangb@indiana.edu](mailto:huangb@indiana.edu)

Bruce Herr, CISHell Architect, Indiana University, [bherr@indiana.edu](mailto:bherr@indiana.edu)

