NetworkWorkbench

A Workbench for Network Scientists

Network Analysis, Modeling, and Visualization

NWB Team

http://nwb.slis.indiana.edu

Indiana University, Bloomington, IN

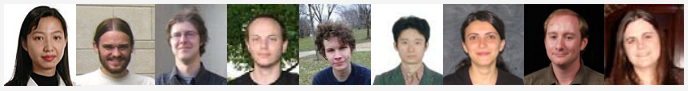
NetworkWorkbench

A Workbench for Network Scientists

Project Details

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

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



Goal:	Develop a large-scale network analysis, modeling and visualization toolkit for physics, biomedical, and social science research.
Amount:	$1,120,926, NSF IIS-0513650 award
Duration:	Sept. 2005 - Aug. 2008
Website:	http://nwb.slis.indiana.edu

Project Details (cont.)

NWB Advisory Board:

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

Outline

- ❑ NWB Research Results
- ❑ NWB Tool Overview and Demo
- ❑ ~~NWB Tool in Bioinformatics Research~~
- ❑ NWB Tool for Scientometrics Research
- ❑ Discussion of CIShell and Future Work

NWB Research Results

- Computational Social Science
- Computational Scientometrics
- Computational Economics
- Computational Proteomics
- Computational Epidemics

Computational Social Science
Studying large scale social networks such as Wikipedia

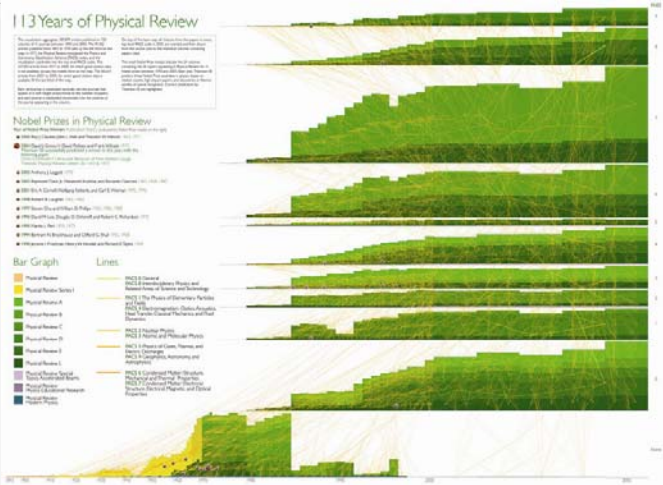
Vizzards 2007 Entry

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

NetworkWorkbench
A Workbench for Network Scientists

113 Years of Physical Review

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



Network Workbench (http://nwb.slis.indiana.edu).

7

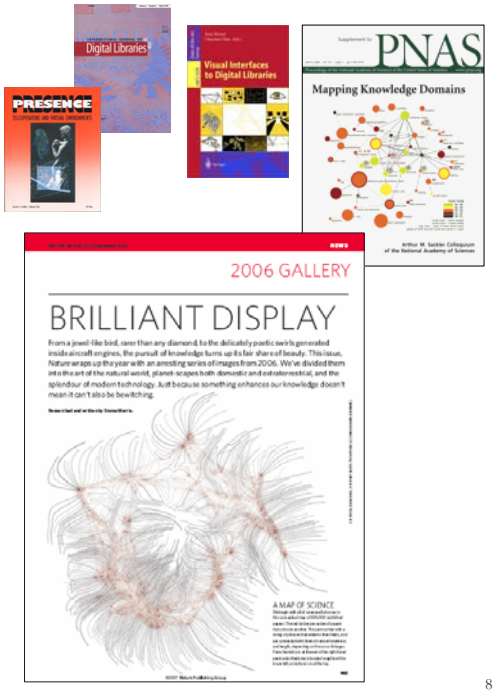
Computational Scientometrics

Studying science by scientific means

Börner, Katy, Chen, Chaomei, and Boyack, Kevin. (2003). **Visualizing Knowledge Domains.** In Blaise Cronin (Ed.), Annual Review of Information Science & Technology, Volume 37, Medford, NJ: Information Today, Inc./American Society for Information Science and Technology, chapter 5, pp. 179-255.

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).

Places & Spaces: Mapping Science exhibit, Currently on display at the American Museum for Science and Energy, Oak Ridge, TN, see also http://scimaps.org.



8

NetworkWorkbench
A Workbench for Network Scientists

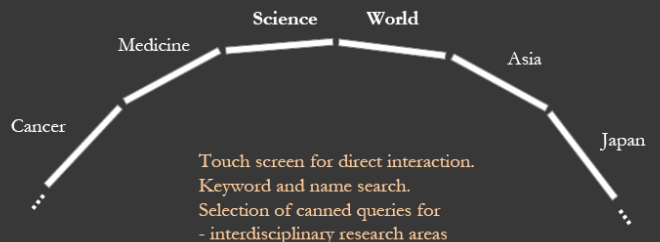
Illuminated Diagram Display

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

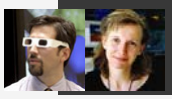
NetworkWorkbench
A Workbench for Network Scientists

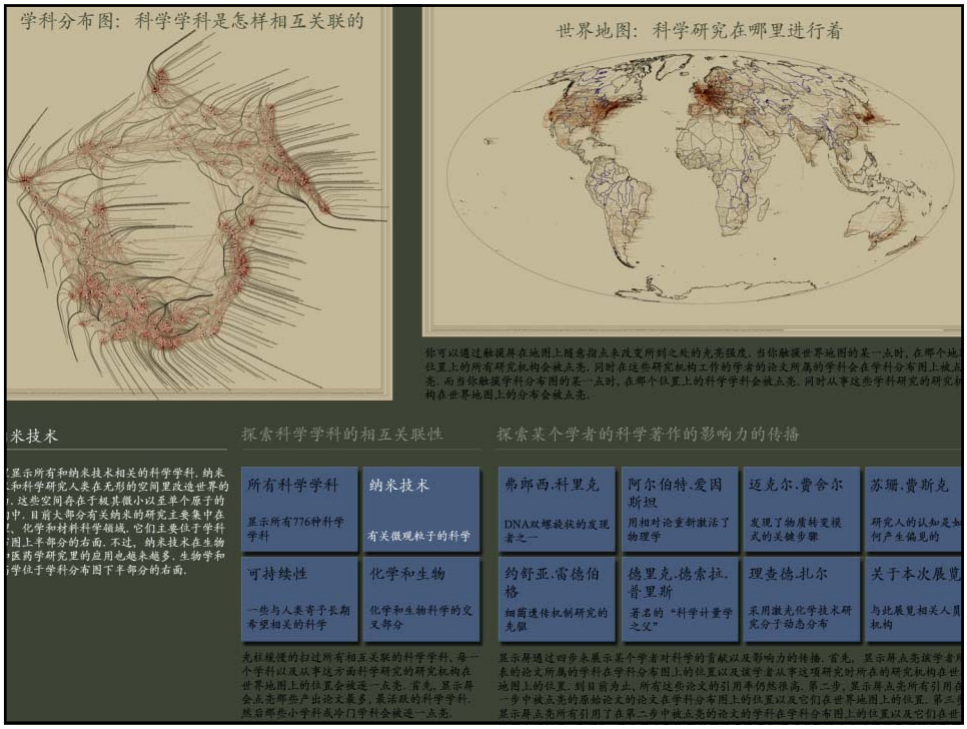
Re-implementation of Illuminated Diagram Software (in progress)
by Advanced Visualization Lab, Indiana University

Drives unlimited number of ID screens.



Touch screen for direct interaction.
Keyword and name search.
Selection of canned queries for
- interdisciplinary research areas
- famous people
- activity patterns, e.g., bursts, trends, etc.





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).

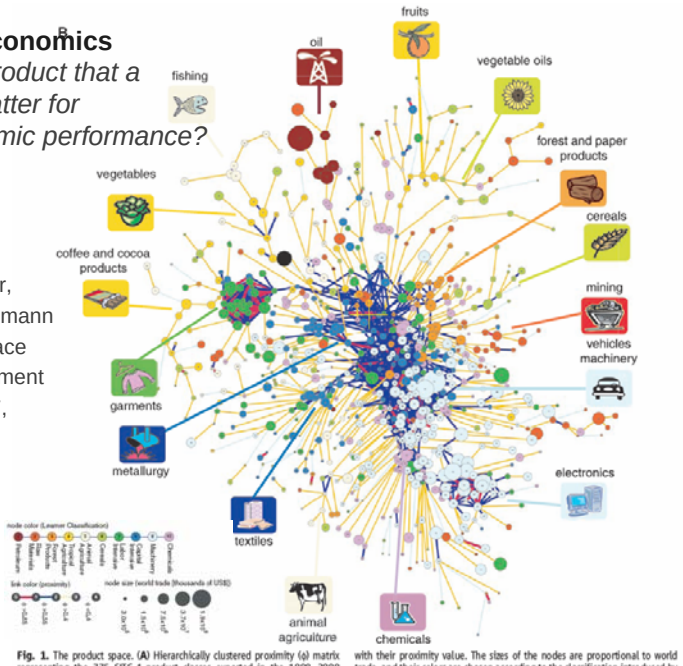


Fig. 1. The product space. (A) Hierarchically clustered proximity (ϕ) matrix representing the 775 SITC-4 product classes exported in the 1998–2000 period. (B) Network representation of the product space. Links are color coded with their proximity value. The sizes of the nodes are proportional to world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

Computational Proteomics

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

Yildriim, 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.

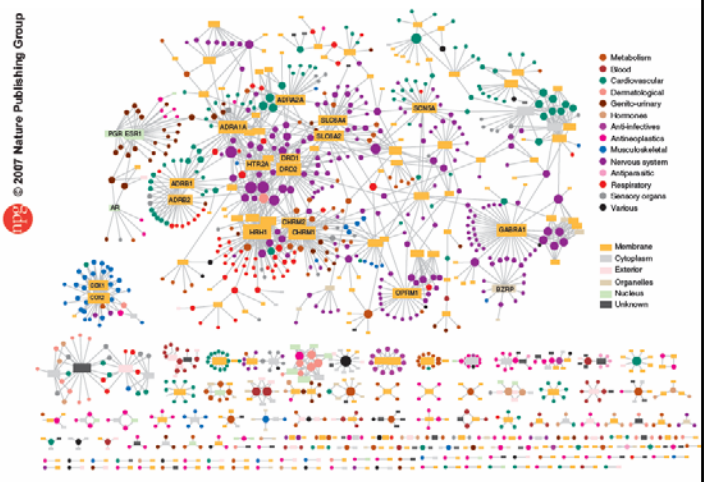
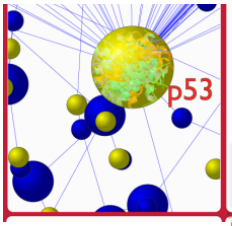


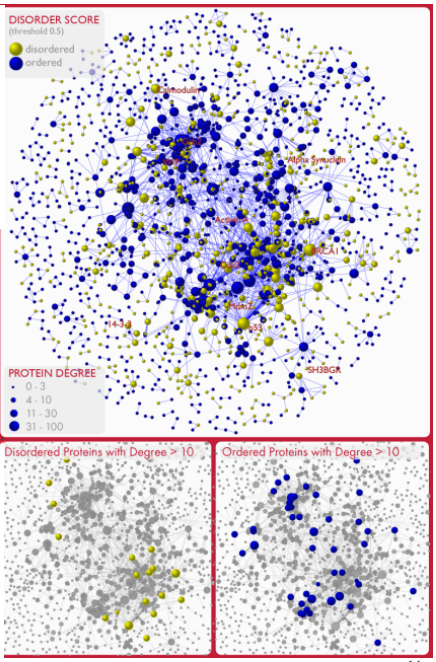
Figure 2 Drug–target network (DT network). The DT network is generated by using the known associations between FDA-approved drugs and their target proteins. Circles and rectangles correspond to drugs and target proteins, respectively. A link is placed between a drug node and a target node if the protein is a known target of that drug. The area of the drug (protein) node is proportional to the number of targets that the drug has (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.

Network Workbench (http://nwb.slis.indian

Computational Proteomics

S. Schnell, S. Fortunato, and S. Roy (2007). Is the intrinsic disorder of proteins the cause of the scale-free architecture of protein-protein interaction networks? *Proteomics* 7, 961-964.





Network Workbench (http://nwb.slis.indiana.edu

14

14

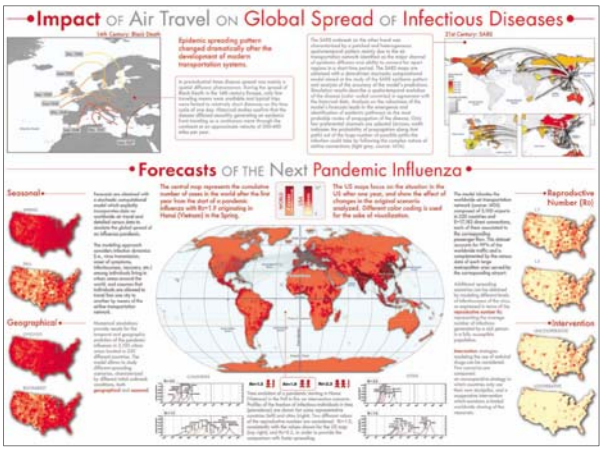
Computational Epidemics

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).

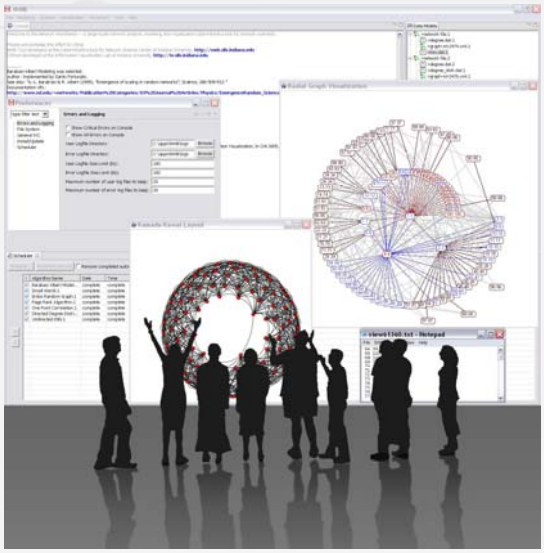
Modeling the Worldwide Spread of Pandemic Influenza:
Baseline Case and Containment Interventions, V. Colizza, A. Barrat, M. Barthelemy, A.-J. Valleron, A.Vespignani, PloS-Medicine 4, e13, 95-110 (2007).



Network Workbench (http://nwb.slis.indiana.edu).

NetworkWorkbench
A Workbench for Network Scientists

The NWB Tool



Network Workbench (http://nwb.slis.indiana.edu). 16

Challenges in Network Science Research

- Data
  - Different data formats
  - Different data models
- Algorithms
  - Different research purposes (preprocessing, modeling, analysis, visualization, clustering)
  - Different implementations of the same algorithm
  - Different programming languages
- Match between Data and Algorithms
- Different communities and practices
- Different tools (Pajek, UCINet, Guess, Cytoscape, R, NWB tool)

Major Deliverables

Network Workbench (NWB) Tool

- A network analysis, modeling, and visualization toolkit for physics, biomedical, and social science research.
- Install and run on multiple Operating Systems.
- Uses Cyberinfrastructure Shell Framework underneath.

Cyberinfrastructure Shell (CIShell)

- An open source, software framework for the integration and utilization of datasets, algorithms, tools, and computing resources.

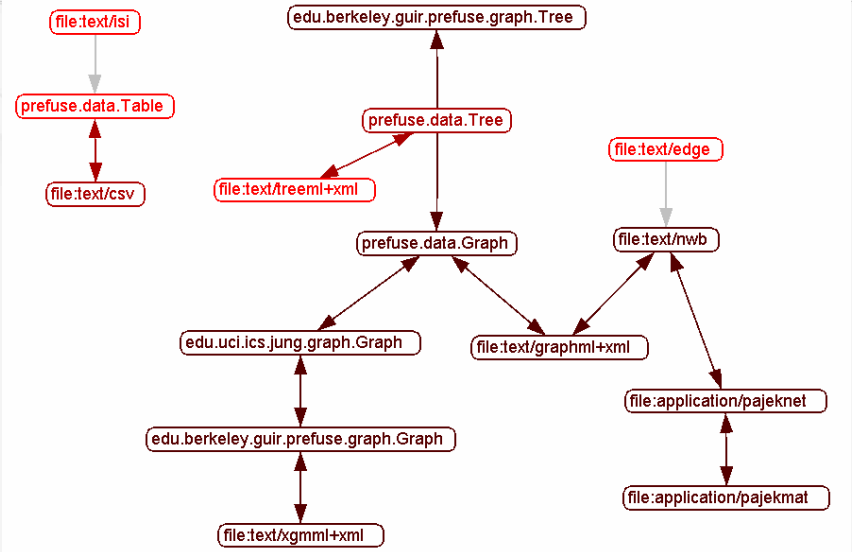
NWB Community Wiki

- A place for users of the NWB Tool, the Cyberinfrastructure Shell (CIShell), or any other CIShell-based program to request, obtain, contribute, and share algorithms and datasets.
- All algorithms and datasets that are available via the NWB Tool have been well documented in the Community Wiki.

Supported File Formats in NWB Tool

- Can load, view, process and save the following file formats:
  - GraphML (.xml or .graphml)
  - XGMML (.xml)
  - Pajek .net (.net)
  - Pajek .mat(.mat)
  - NWB (.nwb)
  - TreeML (.xml)
  - Edge list (.edge)
  - CSV (.csv)
  - isi (.isi)
- Can load two CSV files (node list and edge list) and construct a network.
- Can load an isi file, extract co-authorship network and update graph by merging nodes if needed.

Converter Graph in NWB tool v0.8.0



NWB Tool Major Deliverables

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

Major features in v0.8.0 Release

- Installs and runs on Windows, Linux x86 and Mac OsX.
- Provides over 60 modeling, analysis and visualization algorithms. Half of them are written in Fortran, others in Java.
- Supports large scale network modeling and analysis (over 100,000 nodes)
- Supports various visualization layouts with node/edge annotation.
- Provides several sample datasets with various formats.
- Supports multiple ways to introduce a network to the NWB tool.
- Supports automatic Data Conversion.
- Provides a Scheduler to monitor and control the progress of running algorithms.
- Integrates a 2D plotting tool – Gnuplot (requires pre-installation on Linux and Mac).
- Integrates GUESS (runs on Linux and Mac. *Windows forthcoming)

NWB Tool – Algorithms (Implemented)

Category	Algorithm	Language
Preprocessing	Random Node Deletion	JAVA
	High Degree Node Deletion	JAVA
	Pathfinder Network Scaling	JAVA
	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
	TARL	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

Analysis Algorithm	Language
Node Betweenness Centrality	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
Chord Search	JAVA
Weak Component Clustering	JAVA
Tool: GnuPlot	

NetworkWorkbench
A Workbench for Network Scientists

Summary

NWB tool and CIShell provide

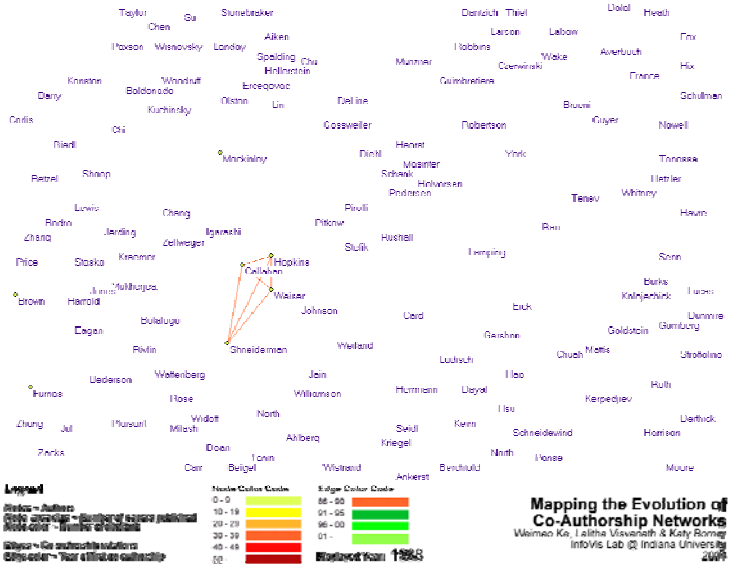
- A testbed for diverse algorithm implementations
- A mechanism to quickly integrate an algorithm and disseminate it through the NWB tool and community wiki.
- A bridge between what application users need and what algorithm developers can provide.

NetworkWorkbench
A Workbench for Network Scientists

NWB Tool for Scientometrics Research

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

Ke, Viswanath & Börner (2004)



25

Data Acquisition from Web of Science

Download all papers by

- Eugene Garfield
- Stanley Wasserman
- Alessandro Vespignani
- Albert-László Barabási

from

- Science Citation Index Expanded (SCI-EXPANDED)--1955-present
- Social Sciences Citation Index (SSCI)--1956-present
- Arts & Humanities Citation Index (A&HCI)--1975-present

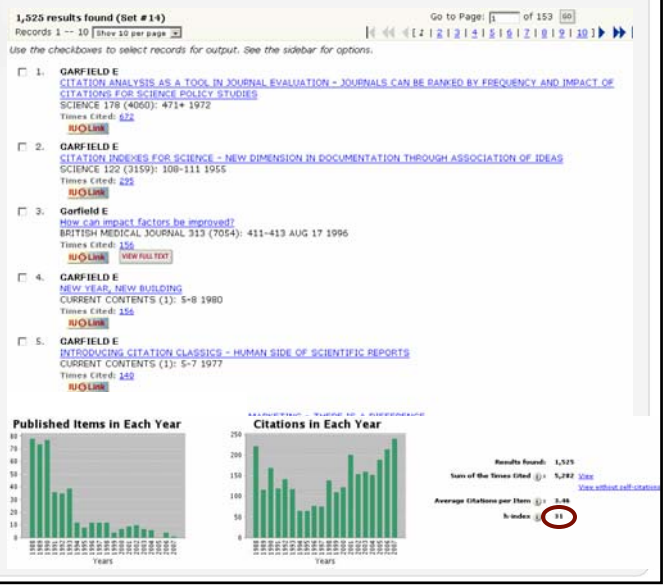
Network Workbench (http://nwb.slis.indiana.edu).

Data Acquisition from Web of Science (cont.)

Eugene Garfield

1525 papers

papers/citations for last 20 years

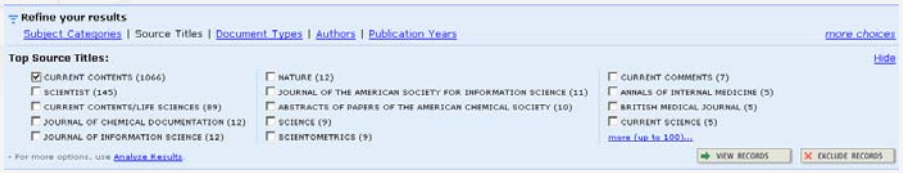


Network Workbench (http://nwb.slis.indiana.edu).

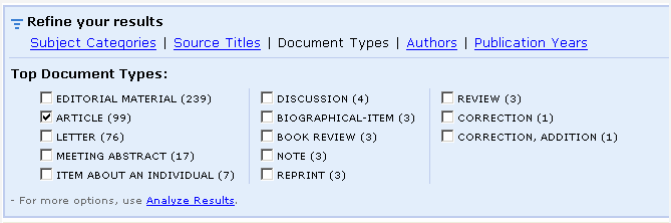
Data Acquisition from Web of Science (cont.)

Can download 500 records max.

Exclude Current Contents articles

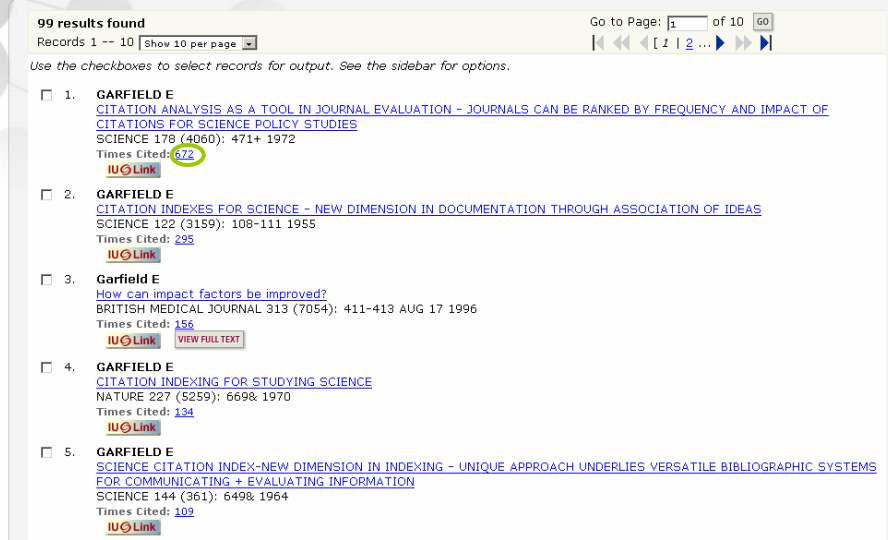


Include only articles. Download 99 articles.



Network Workbench (http://nwb.slis.indiana.edu). 28

Data Acquisition from Web of Science (cont.)



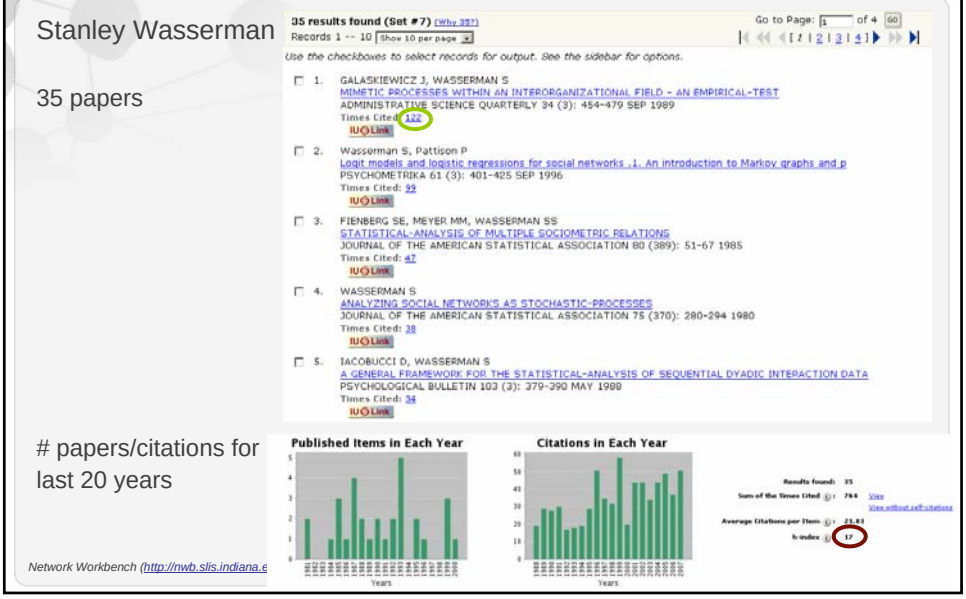
Network Workbench (http://nwb.slis.indiana.edu). 29

Data Acquisition from Web of Science (cont.)

Stanley Wasserman

35 papers

papers/citations for last 20 years



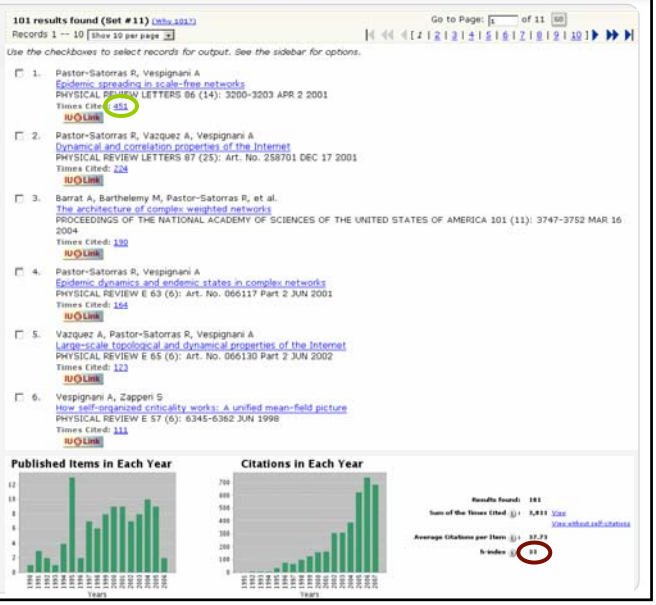
Network Workbench (http://nwb.slis.indiana.e

Data Acquisition from Web of Science (cont.)

Alessandro Vespignani

101 papers

papers/citations for last 20 years



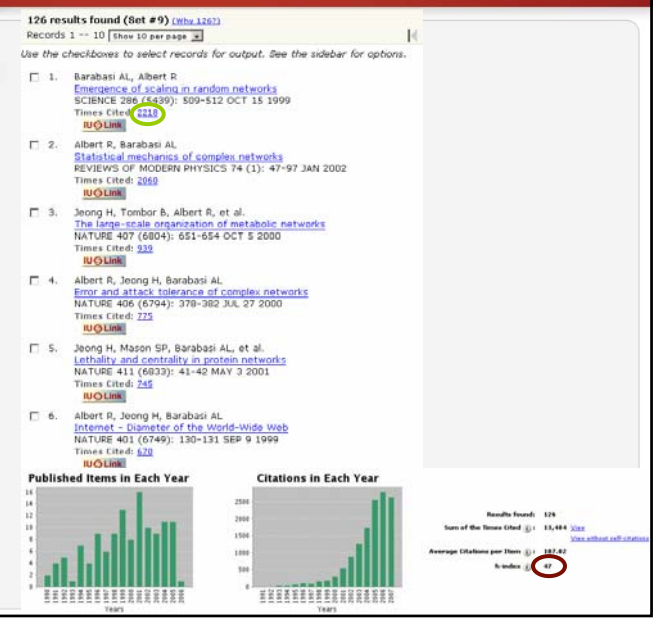
Network Workbench (http://nwb.slis.indiana.edu).

Data Acquisition from Web of Science (cont.)

Albert-László Barabási

126 papers

papers/citations for last 20 years



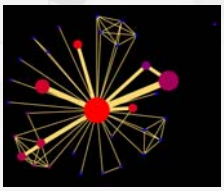
Network Workbench (http://nwb.slis.indiana.edu).

Comparison of Counts

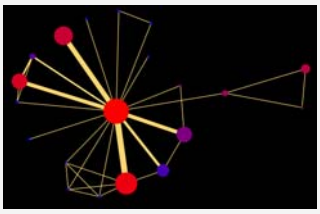
	Age	Highest Cited Paper	H-Index
Eugene Garfield	82	672	31
Stanley Wasserman		122	17
Alessandro Vespignani	42	451	33
Albert-László Barabási	40	2218	47

Comparison of Networks

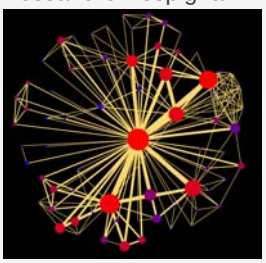
Eugene Garfield



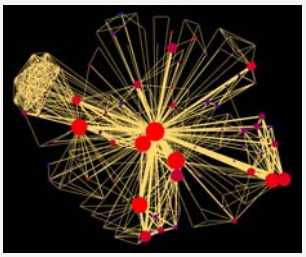
Stanley Wasserman



Alessandro Vespignani

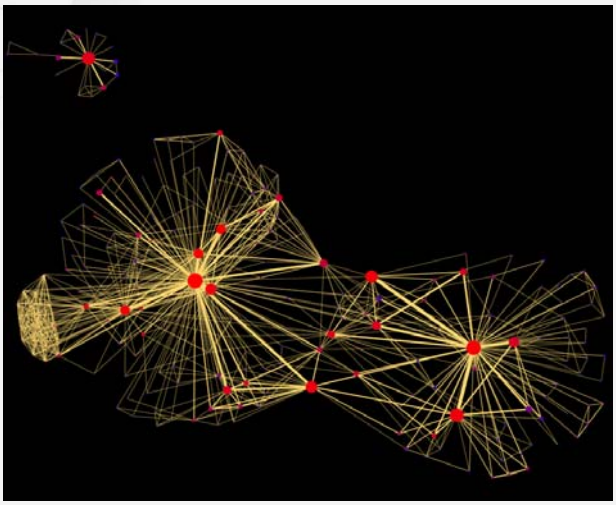


Albert-László Barabási



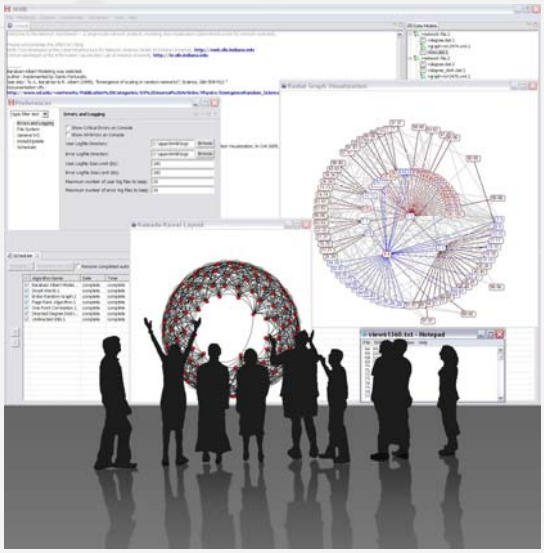
NetworkWorkbench
A Workbench for Network Scientists

Network of Wasserman, Vespignani and Barabási



NetworkWorkbench
A Workbench for Network Scientists

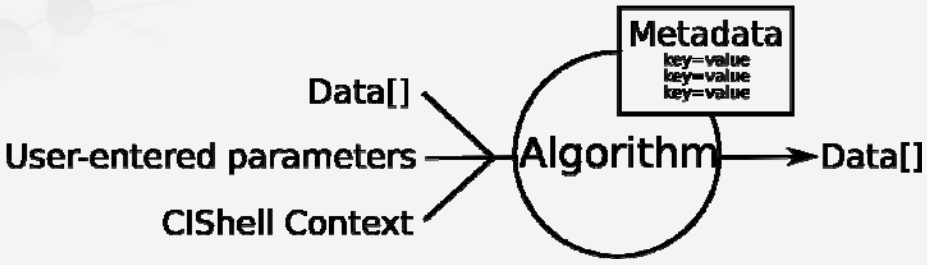
Demo



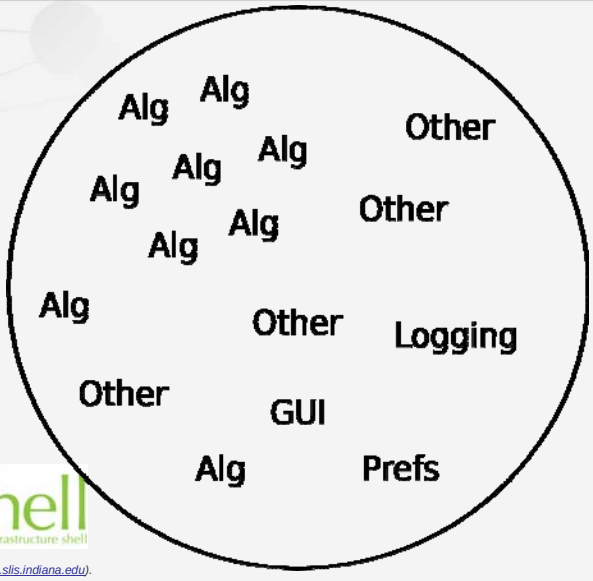
CIShell Framework

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.

Algorithm Definition

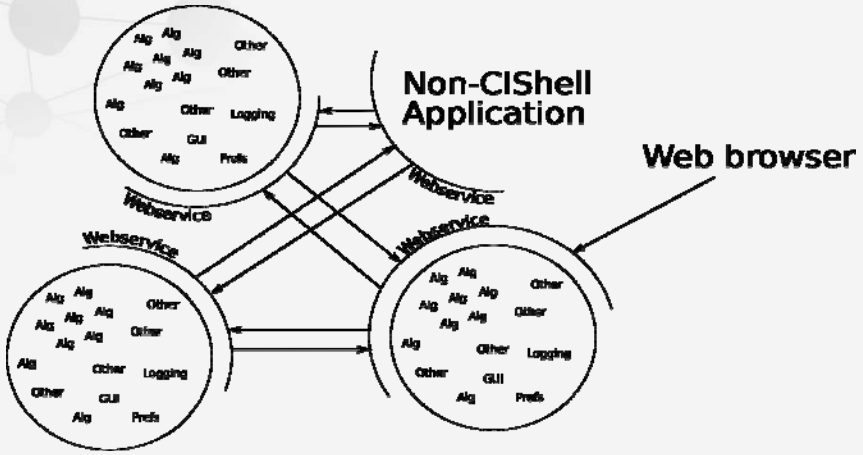


Pooling Algorithms



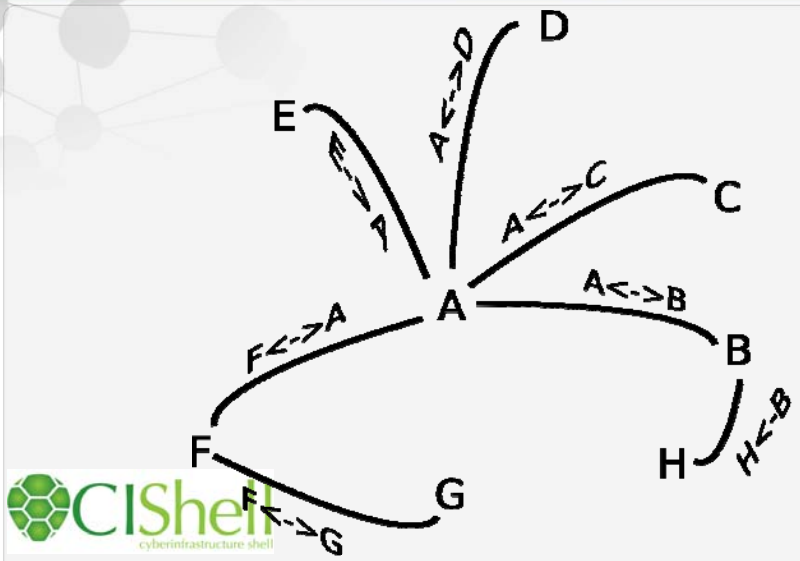
Network Workbench (http://nwb.slis.indiana.edu). 39

Inter-Pool Interaction



Network Workbench (http://nwb.slis.indiana.edu). 40

Data Conversion



Adding New Plugins

- Using update sites
- Using OSGi Console Magick™!
- Dropping plugins into the plugins directory
- Using the NWB Community Wiki

Creating your own plugins

- Wizard-driven templates ease development
- Documentation Forthcoming
  - CIShell Specification
  - CIShell Developer's Guide
- Some preliminary documentation is available at http://cishell.org
- A future workshop will address this
- We are available for consulting

Upcoming Events

- New release (v0.8.0) of the NWB tool and a complete user manual with tutorials (v1.0) will be ready for Sunbelt 2008.
- An end-user workshop is scheduled in the middle of February at IUB (Alex for physics and internet research, Ann & Stan for social network research)
- CIShell specification and CIShell/NWB algorithm developer guide will be available in late February.
- Workshop for algorithm developers will be planned accordingly.

Future Work

Add features to serve communities including Physics, Biology, Social Science, and Scientometrics.

- Integrate classic datasets
- Support the most popular data formats for biology and social science research.
- Develop the converters to bridge those formats to the current formats supported by NWB tool.
- Design and deliver better visualization algorithms and modularity
- Develop components to connect and query SDB
- R bridge
- Customize Menu – Users can re-organize the algorithms for their needs
- Continue integrating best algorithm implementations

References

- Hidalgo, César A. and C. Rodriguez-Sickert. Persistence, Topology and Sociodemographics of a Mobile Phone Network. 2007. (Submitted to Physica A)
- Hidalgo, C.A., B. Klinger, A. L. Barabási, and R. Hausmann. The Product Space and its Consequences for Economic Growth. *Science*. Vol. 317 (2007, July 27): 482-487.
- Börner, Katy. Making Sense of Mankind's Scholarly Knowledge and Expertise: Collecting, Interlinking, and Organizing What We Know and Different Approaches to Mapping (Network) Science. *Environment and Planning B: Planning and Design*. Vol. 34(5), 808-825, Pion.
- Yildriim, Muhammed A., Kwan-Il Goh, Michael E. Cusick, Albert-László Barabási, and Marc Vidal. (2007). Drug-target Network. *Nature Biotechnology* 25 no. 10: 1119-1126.
- Vespignani, Alessandro, Soma Sanyal, and Katy Börner. (2007). Network Science. In *Annual Review of Information Science & Technology,* vol. 41, ed. Blaise Cronin, 537-607. Medford, NJ: Information Today, Inc./American Society for Information Science and Technology.
- Herr II, Bruce W., Weixia (Bonnie) Huang, Shashikant Penumarthy, and Katy Börner. (2007). Designing Highly Flexible and Usable Cyberinfrastructures for Convergence. In *Progress in Convergence – Technologies for Human Wellbeing*, vol. 1093, eds. William S. Bainbridge and Mihail C. Roco, 161-179. Boston: Annals of the New York Academy of Sciences.

References (Cont.)

- Colizza, V., A. Barrat, M. Barthelemy, and A. Vespignani. (2007). Epidemic modeling in complex realities. *Comptes Rendus Biologie* 330: 364-374. Elsevier.
- Colizza, Vittoria, Romualdo Pastor-Satorras, and Alessandro Vespignani. (2007). Reaction-diffusion processes and metapopulation models in heterogeneous networks. *Nature Physics* 3: 276-282. Nature Publishing Group.
- Vermeirssen, Vanessa, M. Inmaculada Barrasa, César A. Hidalgo, Jenny Aurelle B. Babon, Reynaldo Sequerra, Lynn Doucette-Stamm, Albert-László Barabási, and Albertha J. M. Walhout. (2007). Transcription factor modularity in a gene-centered C. elegans core neuronal protein-DNA interaction network. *Network Genome Research*. Cold Spring Harbor Laboratory Press.
- Börner, Katy, Elisha F. Hardy, Bruce W. Herr II, Todd Holloway, and W. Bradford Paley. (2007). Taxonomy Visualization in Support of the Semi-Automatic Validation and Optimization of Organizational Schemas. *Journal of Informetrics* 1 *(3)*: 214-225. Elsevier.
- More papers at http://nwb.slis.indiana.edu/papers.html

Comments & Questions

Websites

- http://nwb.slis.indiana.edu
- https://nwb.slis.indiana.edu/community
- http://cishell.org
- http://cns-trac.slis.indiana.edu/trac/nwb/

NSF IIS-0513650 award

Thank You