Network Workbench Tool
For Large Scale Network Analysis, Modeling, and Visualization

Two-Hour Workshop

Katy Börner and the NWB Team @ IUB
Victor H. Yngve Associate Professor of Information Science
Director of the Cyberinfrastructure for Network Science Center
School of Library and Information Science, Indiana University
10th Street & Jordan Avenue, Wells Library 021
Bloomington, IN, 47405, USA
E-mail: katy@indiana.edu

Project Details

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

Software Team: Lead: Micah Linnemeier
Members: Patrick Phillips, Russell Duhon, Tim Kelley & Ann McCranie
Previous Developers: Weixia (Bonnie) Huang, Bruce Herr, Heng Zhang,
Duygu Balcan, Bryan Hook, Ben Markines, Santo Fortunato, Felix
Terkhorn, 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
Website: http://nwb.slis.indiana.edu
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/
- Tom Snijders, University of Groningen  http://stat.gamma.rug.nl/snijders/
- Noshir Contractor, Northwestern University  http://www.spcomm.uiuc.edu/nosh/

Outline

1. Exemplary Network Science Research by NWB PIs
   - Computational Proteomics
   - Computational Economics
   - Computational Social Science
   - Computational Scientometrics
   - Computational Epidemics

2. NWB Tool Challenges and Opportunities

3. NWB Tool Overview

4. NWB Tool for Scientometrics Research

5. Discussion of Future Work
**Computational Proteomics**

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


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

---

**Computational Economics**

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

Computational Social Science

Studying large scale social networks such as Wikipedia

Vizzards 2007 Entry


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


2. NWB Challenges and Opportunities

- 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
  - Algorithm developers/users are not computer scientists

- Different tools (Pajek, UCINet, Guess, Cytoscape, R, …)
- Different communities, practices, cultures
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.
- Supports many file formats.
- Easy integration of new algorithms thanks to CShell/OSGi.

Cyberinfrastructure Shell (CShell)
- An open source, software framework for the integration and utilization of datasets, algorithms, tools, and computing resources.
- Extends OSGi industry standard.
CIShell – Builds on OSGi Industry Standard

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

OSGi ([http://www.osgi.org](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 8 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.

NWB 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.
- Supports many file formats.
- Easy integration of new algorithms thanks to CIShell/OSGi.

**Cyberinfrastructure Shell (CIShell)**

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

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

https://nwb.slis.indiana.edu/community/
Major features in v1.0.0 beta 2 Release

- Installs and runs on Windows, Linux and Mac OS X.
- Provides over 80 modelling, analysis and visualization algorithms. Half of them are written in Fortran, others in Java.
- Supports large scale network modelling and analysis for certain workflows (over 100,000 nodes)
- Supports several 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 graph exploration/visualization tool.
NWB Tool Overview

1. Download, install, and run.
2. Load, view, convert, save data.
3. Read and visualize a directory hierarchy.
4. Load a network, compute its basic properties, and explore it in GUESS.
Download, install, and run

Goto [http://nwb.slis.indiana.edu](http://nwb.slis.indiana.edu)

**NWB Tool 1.0.0 beta 2**
(development release)
*November 19th, 2008*

Select your operating system from the pull down menu.

Save as *.jar file.

Install and run.

Session log files are stored in `*yourwbdirectory*/logs` directory.

---

**NWB Tool Interface Components**

Welcome to the Network Workbench.

The Network Workbench is supported in part by the NSF DMS-0350950 award. The Network Workbench was developed by Dr. Albert Lluch Barchin, Dr. Santiago Schnell, Dr. Alessandro Vespignani, Dr. Stanley Wasserman, and Dr. Eric A. Lorenz.

The NWB tool was developed by Weiting Huang, Russell Balon, Mehli Lerman, Timothée Kelsey, Edgar Bier, Mariana Bona, Dwayne Herr, Santa Futumbi, Ben Martin, Peter Forlino, Hong Zhang, Megha Ranjan, Gihan Kodagoda, Rayma Sabbar, Vinay Thakur, Sane Sarem, Ann McNee, Alessandro Vespignani, and Kevin Brown. It uses the Cyberinfrastructure Tool (http://castellanos.developed at the Cyberinfrastructure for Network Science Center (http://cnsi.isis.indiana.edu) at Indiana University.

Please cite as follows:
### File, Preprocessing, Modeling, and Visualization Menus

<table>
<thead>
<tr>
<th>File</th>
<th>Preprocessing</th>
<th>Modeling</th>
<th>Visualization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Load...</td>
<td>Extract Top Nodes</td>
<td>Random Graph</td>
<td>GCESS</td>
</tr>
<tr>
<td>Load and Clean Sf File</td>
<td>Extract Nodes Above or Below Value</td>
<td>Watts-Strogatz Small World</td>
<td>GnuPlot</td>
</tr>
<tr>
<td>Head Directory Hierarchy</td>
<td>Delete High Degree Nodes</td>
<td>Barabasi-Albert Scale-Free</td>
<td>Del. Tree</td>
</tr>
<tr>
<td>Datasets</td>
<td>Delete Random Nodes</td>
<td></td>
<td>Tree Map</td>
</tr>
<tr>
<td></td>
<td>Delete Isolates</td>
<td></td>
<td>Tree View</td>
</tr>
<tr>
<td></td>
<td>Extract Top Edges</td>
<td></td>
<td>Prefuse beta</td>
</tr>
<tr>
<td></td>
<td>Extract Edges Above or Below Value</td>
<td></td>
<td>雷达树图 (Prefuse alpha)</td>
</tr>
<tr>
<td></td>
<td>Remove Self Loops</td>
<td></td>
<td>雷达树图与注释 (Prefuse beta)</td>
</tr>
<tr>
<td></td>
<td>Trim by Degree</td>
<td></td>
<td>简单树图 (Prefuse beta)</td>
</tr>
<tr>
<td></td>
<td>Pathfinder Network Scaling</td>
<td></td>
<td>Force Directed</td>
</tr>
<tr>
<td></td>
<td>empo (nodes)</td>
<td></td>
<td>与注释 (Prefuse beta)</td>
</tr>
<tr>
<td></td>
<td>Brick Wall Sampling (in nodes)</td>
<td></td>
<td>Kamada-Kawai (JUNG)</td>
</tr>
<tr>
<td></td>
<td>Node Sampling</td>
<td></td>
<td>Force-Connection (JUNG)</td>
</tr>
<tr>
<td></td>
<td>Edge Sampling</td>
<td></td>
<td>Fruchterman-Reingold (JUNG)</td>
</tr>
<tr>
<td></td>
<td>Symmetrize</td>
<td></td>
<td>Fruchterman-Reingold与注释 (Prefuse beta)</td>
</tr>
<tr>
<td></td>
<td>Dichotomize</td>
<td></td>
<td>斯普林 (JUNG)</td>
</tr>
<tr>
<td></td>
<td>Multiparticle Joining</td>
<td></td>
<td>小世界 (Prefuse alpha)</td>
</tr>
<tr>
<td></td>
<td>Normalize Text</td>
<td></td>
<td>并行坐标 (演示)</td>
</tr>
<tr>
<td></td>
<td>Slice Table by Time</td>
<td></td>
<td>LaNet</td>
</tr>
</tbody>
</table>

### Analysis Menu and Submenus

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Unweighted &amp; Undirected</th>
<th>Unweighted &amp; Directed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Network Analysis Toolkit (NAT)</td>
<td>Node Degree, Degree Distribution</td>
<td>Node Indegree</td>
</tr>
<tr>
<td>Unweighted &amp; Undirected</td>
<td>K–Nearest Neighbor</td>
<td>Node Outdegree</td>
</tr>
<tr>
<td>Weighted &amp; Undirected</td>
<td>Watts–Strogatz Clustering Coefficient</td>
<td>Indegree Distribution</td>
</tr>
<tr>
<td>Unweighted &amp; Directed</td>
<td>Watts–Strogatz Clustering Coefficient Over k</td>
<td>Outdegree Distribution</td>
</tr>
<tr>
<td></td>
<td>Diameter</td>
<td>k–Nearest Neighbor</td>
</tr>
<tr>
<td></td>
<td>Average Shortest Path</td>
<td>Single Node In-Out Degree Correlations</td>
</tr>
<tr>
<td></td>
<td>Shortest Path Distribution</td>
<td>Page Rank</td>
</tr>
<tr>
<td></td>
<td>Node Betweenness Centrality</td>
<td>Dyad Reciprocity</td>
</tr>
<tr>
<td></td>
<td>Connected Components</td>
<td>Arc Reciprocity</td>
</tr>
<tr>
<td></td>
<td>Weak Component Clustering</td>
<td>Adjacency Transitivity</td>
</tr>
<tr>
<td></td>
<td>Extract K–Core</td>
<td>Weak Component Clustering</td>
</tr>
<tr>
<td></td>
<td>Annotate K–Coreness</td>
<td>Extract and Annotate Attractors</td>
</tr>
<tr>
<td></td>
<td>Weighted &amp; Undirected</td>
<td>Extract K–Core</td>
</tr>
<tr>
<td></td>
<td>Average Clustering Coefficient</td>
<td>Annotate K–Coreness</td>
</tr>
<tr>
<td></td>
<td>Average Nearest Neighbor Degree</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Average Strength</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Degree and Strength</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Endpoint Degree</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Strength Distribution</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Weight Distribution</td>
<td></td>
</tr>
</tbody>
</table>

[Network Workbench](http://nwb.slis.indiana.edu)
Integrated Tools

Gnuplot
portable command-line driven interactive data and function plotting utility
http://www.gnuplot.info/

GUESS
exploratory data analysis and visualization tool for graphs and networks.
https://nwb.slis.indiana.edu/community/?n=VisualizeData.GUESS.

Supported Data Formats

In November 2008, the NWB tool supports loading the following input file formats:

- GraphML (*.xml or *.graphml)
- XGMML (*.xml)
- Pajek .NET (*.net) & Pajek .Matrix (*.mat)
- NWB (*.nwb)
- TreeML (*.xml)
- Edge list (*.edge)
- CSV (*.csv)
- ISI (*.isi)
- Scopus (*.scopus)
- NSF (*.nsf)
- Bibtex (*.bib)
- Endnote (*.enw)

and the following network file output formats:

- GraphML (*.xml or *.graphml)
- Pajek .MAT (*.mat)
- Pajek .NET (*.net)
- NWB (*.nwb)
- XGMML (*.xml)
- CSV (*.csv)

These formats are documented at
NWB Ecology of Data Formats and Converters

Not shown are 15 sample datasets, 45 data preprocessing, analysis, modeling and visualization algorithms, 9 services.

13 Supported data formats

6 Output formats for diverse visualization algorithms

8 Intermediate data formats

35 Supported by data converters.

DATA FORMATS
- External File
- Validated File Format
- In-Memory Java Object

DATA CONVERTERS
- File Validator
- File Format to In-Memory Java Object converter
- File Format to File Format
- In-Memory Java Object to In-Memory Java Object

VISIT
http://nwb.sis.indiana.edu
https://nwb.sis.indiana.edu/community
http://www.cisshell.org

DOWNLOAD: NWB Tool
http://nwb.sis.indiana.edu/software.html
The 
*`yournwbdirectory`/sampledata* directory provides sample datasets from the biology, network, scientometrics, and social science research domains:

/biology
/network
/scientometrics
/scientometrics/bibtex
/scientometrics/csv
/scientometrics/endnote
/scientometrics/isi
  ➢ FourNetSciResearchers.isi
/scientometrics/nsf
  ➢ Cornell.nsf
  ➢ Indiana.nsf
  ➢ Michigan.nsf
/socialscience
  ➢ florentine.nwb

The blue ones are used in this tutorial.
Property Files and Python Scripts

The "*yournwdirectory*/" directory also contains

```
/sampledata/scientometrics/properties  // Used to extract networks and merge data
- bibtexCoAuthorship.properties
- endnoteCoAuthorship.properties
- isiCoAuthorship.properties
- isiCoCitation.properties
- isiPaperCitation.properties
- mergeBibtexAuthors.properties
- mergeEndnoteAuthors.properties
- mergeIsiAuthors.properties
- mergeNsPIs.properties
- mergeScopusAuthors.properties
- nsfCoPI.properties
- scopusCoAuthorship.properties
```

```
/sampledata/scripts/GUESS  // Used to do color/size/shape code networks
- co-author-nw.py
- co-PI-nw.py
- paper-citation-nw.py
- reference-co-occurrence-nw.py
```

Network Workbench (http://nwb.slis.indiana.edu)
Load, View and Save (Convert) Data

Use 'File > Load File' to load florentine.nwb in sample datasets in '*/samplendata/socialscience'.

The loaded file will appear in the Data Manager window.

Right click loaded file to save, view, rename, or discard.

Data Converter Graph

There is no central data format.

Instead, data formats used in different communities and required by the different algorithms are supported.
NWB Tool Overview

1. Download, install, and run.
2. Load, view, convert, save data.
3. Read and visualize a directory hierarchy.
4. Load a network, compute its basic properties, and explore it in GUESS.

Reading and Visualizing a Directory Hierarchy

Use ‘File > Read Directory Hierarchy’ with parameters

Visualize resulting ‘Directory Tree - Prefuse (Beta) Graph’ using
- ‘Visualization > Tree View (prefuse beta)’
- ‘Visualization > Tree Map (prefuse beta)’
- ‘Visualization > Balloon Graph (prefuse alpha)’
- ‘Visualization > Radial Tree/Graph (prefuse alpha)’
Different views of the /nwb directory hierarchy.

Note the size of the /plugin directory.

---

**NWB Tool Overview**

1. Download, install, and run.
2. Load, view, convert, save data.
3. Read and visualize a directory hierarchy.
4. Load a network, compute its basic properties, and explore it in GUESS.
Select florentine.nwb in Data Manager.

- Run ‘Analysis > Network Analysis Toolkit (NAT)’ to get basic properties.
  
  ```
  This graph claims to be undirected.
  Nodes: 16
  Isolated nodes: 1
  Node attributes present: label, wealth, totalities, prarities
  mdeg: 27
  No self loops were discovered.
  No parallel edges were discovered.
  Edge attributes: 
  Numeric attributes:
    Example value
    marriage... 1
    business... 1
  Did not detect any numeric attributes.
  This network does not seem to be a valued network.
  Average degree: 1.375
  This graph is not weakly connected.
  There are 3 weakly connected components (1 isolates).
  The largest connected component consists of 15 nodes.
  Did not calculate strong connectedness because this graph was not directed.
  
  Density (disregarding weights): 0.225
  ```


- Select network and run ‘Visualization > GUESS’ to open GUESS with file loaded.

- Apply ‘Layout -> GEM’.

Network Workbench (http://nwb.slis.indiana.edu).
Pan: "grab" the background by holding left-click and moving your mouse.

Zoom: Using scroll wheel, press the "+" and "-" buttons in the upper-left hand corner, or right-click and move the mouse left or right. Center graph by selecting 'View -> Center'.

Select to select/move single nodes. Hold down 'Shift' to select multiple.

Right click to modify Color, etc.

Graph Modifier: Select "all nodes" in the Object drop-down menu and click 'Show Label' button.

Select "nodes based on ">="", then select "wealth" from the Property drop-down menu, "">="" from the Operator drop-down menu, and finally type "50" into the Value box. Then a color/size/shape code.
Interpreter:

Uses Jython a combination of Java and Python.

Try
colorize(wealth, white, red)

resizeLinear(sitebetweenness, 5, 25)

---

**Network Workbench**

---

**NWB Tool for Scientometrics Research**


2. Loading ISI files of major network science researchers, extracting, analyzing and visualizing paper-citation networks and co-author networks.

3. Loading text files and detecting Bursts.
NWB Tool for Scientometrics Research


2. Loading ISI files of major network science researchers, extracting, analyzing and visualizing paper-citation networks and co-author networks.

3. Loading text files and detecting Bursts.

Download NSF awards data from http://www.nsf.gov/awardsearch/
Analyzing NSF Data

Active NSF Awards on 11/07/2008:

- Indiana University 257
  (there is also Indiana University at South Bend Indiana University Foundation, Indiana University Northwest, Indiana University-Purdue University at Fort Wayne, Indiana University-Purdue University at Indianapolis, Indiana University-Purdue University School of Medicine)
- Cornell University 501
  (there is also Cornell University – State, Joan and Sanford I. Weill Medical College of Cornell University)
- University of Michigan Ann Arbor 619
  (there is also University of Michigan Central Office, University of Michigan Dearborn, University of Michigan Flint, University of Michigan Medical School)

Save files as csv but rename into .nsf.
Or simply use the files saved in
`*yournwbdirectory*/sampledata/scientometrics/nsf`.

Extracting Co-PI Networks

Load NSF data, selecting the loaded dataset in the Data Manager window, run
`Scientometrics > Extract Co-Occurrence Network` using parameters:

Two derived files will appear in the Data Manager window: the co-PI network and a merge table. In the network, nodes represent investigators and edges denote their co-PI relationships. The merge table can be used to further clean PI names.

Running the `Analysis > Network Analysis Toolkit (NAT)` reveals that the number of nodes and edges but also of isolate nodes that can be removed running
`Preprocessing > Delete Isolates`.

Select `Visualization > GUESS` to visualize. Run `co-PI-nw.py` script.
Select network after removing isolates and run ‘Analysis > Unweighted and Undirected > Weak Component Clustering’ with parameter

Indiana’s largest component has 19 nodes, Cornell’s has 67 nodes, Michigan’s has 55 nodes.

Visualize Cornell network in GUESS using same .py script and save via ‘File > Export Image’ as jpg.
Largest component of Cornell U co-PI network

Node size/color ~ total award money
Top-50 total award money nodes are labeled.

Top-10 investigators by total award money

for i in range(0, 10):
    print str(nodesbytotalawardmoney[i].label) + ": " + str(nodesbytotalawardmoney[i].totalawardmoney)

<table>
<thead>
<tr>
<th>Indiana University</th>
<th>Cornell University</th>
<th>Michigan University</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frank Lester: 6,402,330</td>
<td>Sandip Tiwari: 72,094,578</td>
<td>Kensall Wise: 32,164,404</td>
</tr>
<tr>
<td>Michael Lynch: 6,361,796</td>
<td>Donald Bilderback: 47,360,053</td>
<td>Georg Raithel: 23,832,421</td>
</tr>
<tr>
<td>Craig Stewart: 6,216,352</td>
<td>Ernest Fontes: 29,380,053</td>
<td>Roseanne Sension: 23,812,921</td>
</tr>
<tr>
<td>William Snow: 5,434,796</td>
<td>Hasan Padamsee: 18,292,000</td>
<td>Theodore Norris: 23,350,921</td>
</tr>
<tr>
<td>Douglas V. Houweling: 5,068,122</td>
<td>Melissa Hines: 13,099,545</td>
<td>Paul Berman: 23,350,921</td>
</tr>
<tr>
<td>Miriam Zolan: 5,000,627</td>
<td>Timothy Fahey: 7,223,112</td>
<td>Robert Schoeni: 21,991,140</td>
</tr>
<tr>
<td>Carla Caceres: 5,000,627</td>
<td>Jon Kleinberg: 7,165,507</td>
<td>Wei-Jun Jean Yeung: 21,991,140</td>
</tr>
</tbody>
</table>
NWB Tool for Scientometrics Research


2. Loading ISI files of major network science researchers, extracting, analyzing and visualizing paper-citation networks and co-author networks.

3. Loading text files and detecting Bursts.

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
Eugene Garfield
1,525 papers

# papers/citations for last 20 years

Data Acquisition from Web of Science (cont.)

Can download 500 records max.
Exclude Current Contents articles

Network Workbench (http://nwb.slis.indiana.edu)
Data Acquisition from Web of Science (cont.)

1. **GARFIELD E**
   - **Title**: Citation Analysis as a Tool in Journal Evaluation - Journals Can Be Ranked by Frequency and Impact of Citations for Science Policy Studies
   - **Source**: Science 170 (1970): 473-479
   - Times Cited: 125

2. **GARFIELD E**
   - **Title**: Citation Indexes for Science - A New Dimension in Documentation Through Association of Ideas
   - **Source**: Science 212 (1981): 218-219
   - Times Cited: 96

3. **GARFIELD E**
   - **Title**: Does an impact factor lie to you?
   - **Source**: British Medical Journal 313 (1996): 411-413
   - Times Cited: 1096

4. **GARFIELD E**
   - **Title**: Citation Indexing for Studying Science
   - Times Cited: 124

5. **GARFIELD E**
   - **Title**: Science Citation Index - New Dimension in Indexing - Unique Approach Underlies Versatile Bibliographic Systems for Communicating & Evaluating Information
   - **Source**: Science 144 (1964): 548-554
   - Times Cited: 222

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

---

Stanley Wasserman

35 papers

- **Title**: Wasserman D.
  - **Title**: Network Processes Within an Interdisciplinary Field - An Empirical-Test
  - **Source**: Administrative Science Quarterly 34 (3): 454-478
  - Times Cited: 152

- **Title**: Wasserman S, Pattison D.
  - **Title**: Long models and logistic regression for social networks: An introduction to Markov graphs and P
  - **Source**: Psychometrika 61 (3): 401-425
  - Times Cited: 121

- **Title**: Fincher M, Meyer MM, Wasserman S.
  - **Title**: Statistical Analysis of Multiple Socioeconomic Relations
  - **Source**: Journal of the American Statistical Association 80 (389): 51-67
  - Times Cited: 42

- **Title**: Wasserman S.
  - **Title**: Analyzing Social Networks As Stochastic Processes
  - **Source**: Journal of the American Statistical Association 75 (370): 290-301
  - Times Cited: 28

- **Title**: Iacobucci D, Wasserman D.
  - **Title**: A General Framework for the Statistical Analysis of Sequential Dyadic Interaction Data
  - **Source**: Psychological Bulletin 103 (3): 376-388
  - Times Cited: 40

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

<table>
<thead>
<tr>
<th>Name</th>
<th>Age</th>
<th>Total # Cites</th>
<th>Total # Papers</th>
<th>H-Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eugene Garfield</td>
<td>82</td>
<td>1,525</td>
<td>672</td>
<td>31</td>
</tr>
<tr>
<td>Stanley Wasserman</td>
<td>122</td>
<td>35</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>Alessandro Vespignani</td>
<td>42</td>
<td>451</td>
<td>101</td>
<td>33</td>
</tr>
<tr>
<td>Albert-László Barabási</td>
<td>40</td>
<td>2,218</td>
<td>126</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>41</td>
<td>16,920</td>
<td>159</td>
<td>52</td>
</tr>
</tbody>
</table>

*(Dec 2007)* *(Nov 2008)*

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

### Extract Co-Author Network


To extract the co-author network, select the ‘361 Unique ISI Records’ table and run ‘Scientometrics > Extract Co-Author Network’ using isi file format:

The result is an undirected network of co-authors in the Data Manager. It has 248 nodes and 891 edges.

To view the complete network, select the network and run ‘Visualization > GUESS > GEM’. Run `Script > Run Script…`. And select Script folder > GUESS > co-author-nw.py`.

Network Workbench ([http://nwb.slis.indiana.edu](http://nwb.slis.indiana.edu))
Load ‘*yourwbdirectory*/sampledata/scientometrics/isi/FourNetSciResearchers.isi’ using ‘File > Load and Clean ISI File’.

To extract the paper-citation network, select the ‘361 Unique ISI Records’ table and run ‘Scientometrics > Extract Directed Network’ using the parameters:

The result is a directed network of paper citations in the Data Manager. It has 5,335 nodes and 9,595 edges.

To view the complete network, select the network and run ‘Visualization > GUES’.
Run ‘Script > Run Script …’ and select ‘*yourwbdirectory*/script/GUES/paper-citation-nw.py’.

Network Workbench (http://nwb.slis.indiana.edu).
NWB Tool for Scientometrics Research


2. Loading ISI files of major network science researchers, extracting, analyzing and visualizing paper-citation networks and co-author networks.

3. Loading text files and detecting Bursts.

Network Workbench (http://nwb.slis.indiana.edu)
Open and Preprocess SDB zip file

Load medline_medline_master.csv to NWB.

Run 'Preprocessing > Normalize Text' with a space as New Separator.
Run ‘Analysis > Textual > Burst Detection’ with parameters: and space as a separator.

Sort result by burst weight

<table>
<thead>
<tr>
<th>Word</th>
<th>Length</th>
<th>Weight</th>
<th>Strength</th>
<th>Start</th>
<th>End</th>
</tr>
</thead>
<tbody>
<tr>
<td>care</td>
<td>1</td>
<td>Infinity</td>
<td>Infinity</td>
<td>1988</td>
<td>1988</td>
</tr>
<tr>
<td>water</td>
<td>1</td>
<td>29.8883</td>
<td>29.8883</td>
<td>2002</td>
<td>2002</td>
</tr>
<tr>
<td>countri</td>
<td>10</td>
<td>27.03612</td>
<td>27.03612</td>
<td>1990</td>
<td>1999</td>
</tr>
<tr>
<td>protect</td>
<td>2</td>
<td>23.32114</td>
<td>23.32114</td>
<td>2002</td>
<td>2002</td>
</tr>
<tr>
<td>farm</td>
<td>1</td>
<td>23.23146</td>
<td>30.42535</td>
<td>2008</td>
<td>2008</td>
</tr>
<tr>
<td>villag</td>
<td>2</td>
<td>22.33649</td>
<td>30.42535</td>
<td>2008</td>
<td>2008</td>
</tr>
<tr>
<td>blood</td>
<td>5</td>
<td>22.12166</td>
<td>22.12166</td>
<td>1996</td>
<td>2000</td>
</tr>
</tbody>
</table>

5. Discussion of Future Work

- Improving GUESS usability.
- Creating wizard for integrating compiled algorithms.
- Algorithms can communicate with other algorithms at runtime using streams.
- Develop components to connect and query SDB.
- Customize Menu – Users can re-organize the algorithms for their needs.
EpiC will Build on and Extend NWB

Resources

Publications
- http://nwb.slis.indiana.edu/pub.html

Community Wiki, Tutorials, FAQ
- https://nwb.slis.indiana.edu/community
- http://nwb.slis.indiana.edu/doc.html

Software
- http://cishell.org
- http://nwb.slis.indiana.edu/download.html

Developer Resources
- http://cns-trac.slis.indiana.edu/trac/nwb

This work is partially supported by NSF IIS-0513650.
I would like to thank Micah Linnemeier, Patrick Phillips, Elisha Hardy, Mark Price, Hanning Guo, and Angela Zoss for their contributions to this workshop tutorial.

http://cns.slis.indiana.edu