NetworkWorkbench

A Workbench for Network Scientists

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
Duration:	Sept. 2005 - Aug. 2009
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

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?

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.

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

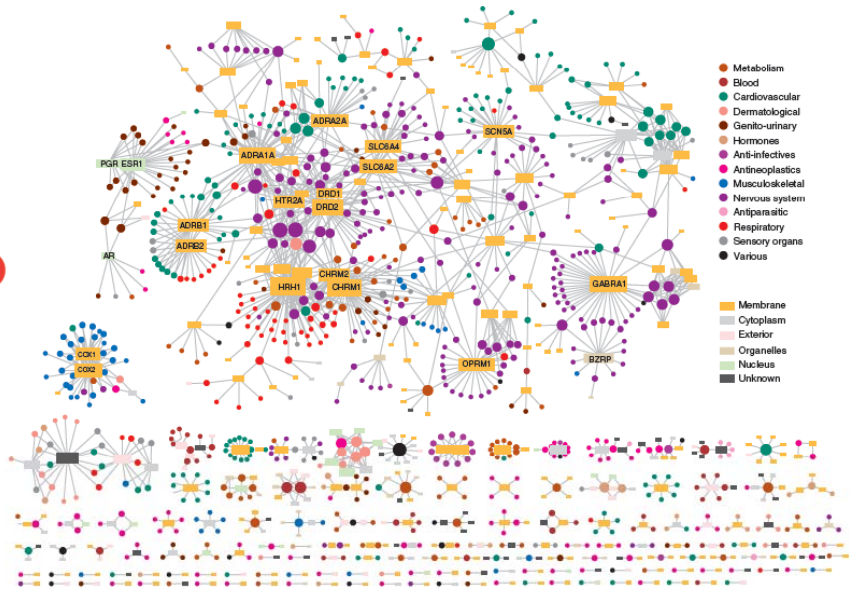


Figure 2 Drug–target network (DT network). The DT network is generated by using the known associations between FDA-approved drugs and their target proteins. Circles and rectangles correspond to drugs and target proteins, respectively. A link is placed between a drug node and a target node if the protein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

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).

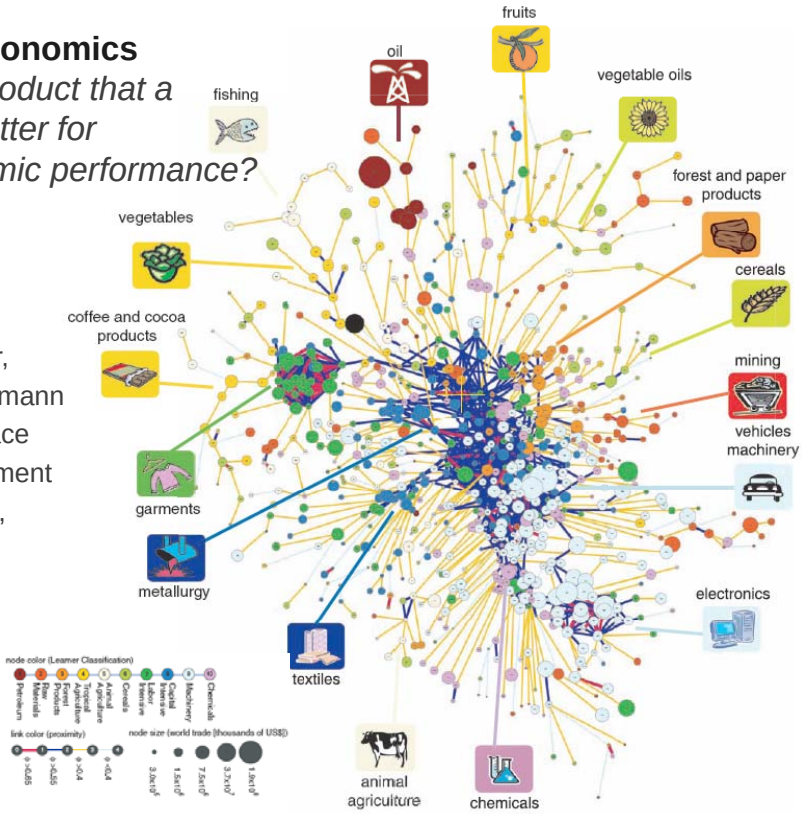


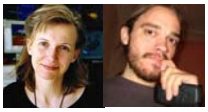
Fig. 1. The product space. **(A)** Hierarchically clustered proximity (φ) matrix representing the 775 SITC-4 product classes exported in the 1998–2000 period. **(B)** Network representation of the product space. Links are color coded with their proximity value. The sizes of the nodes are proportional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

Computational Social Science

Studying large scale social networks such as Wikipedia

Vizzards 2007 Entry

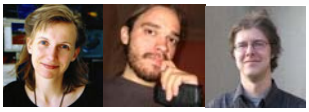
Second Sight: An Emergent Mosaic of Wikipedian Activity, The NewScientist, May 19, 2007

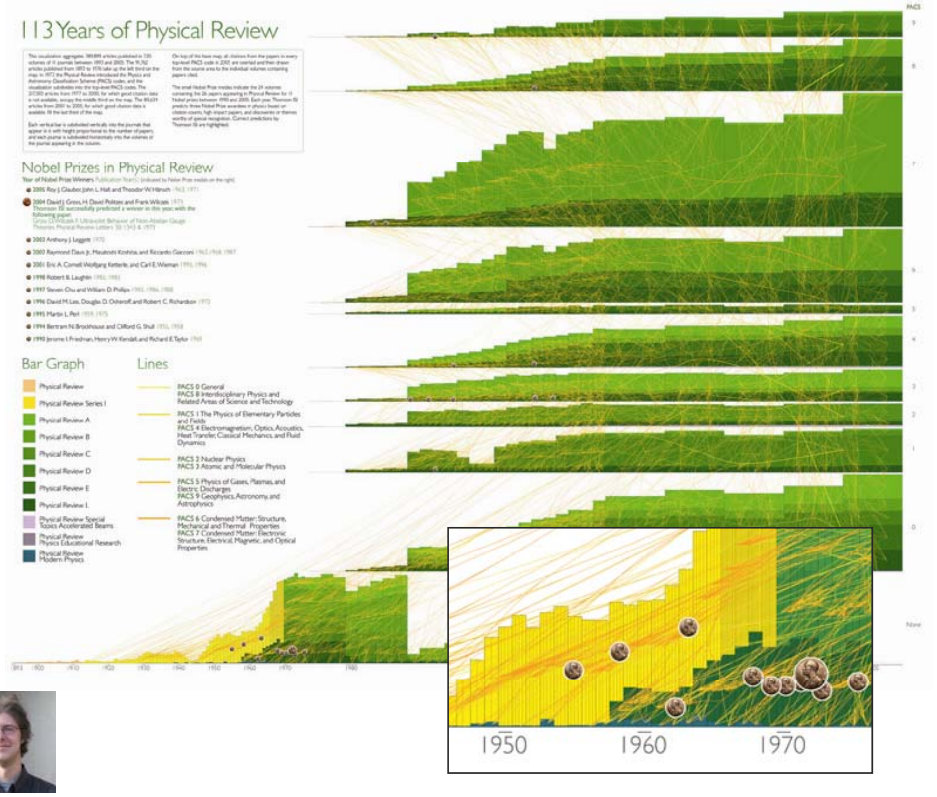


Computational Scientometrics

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)





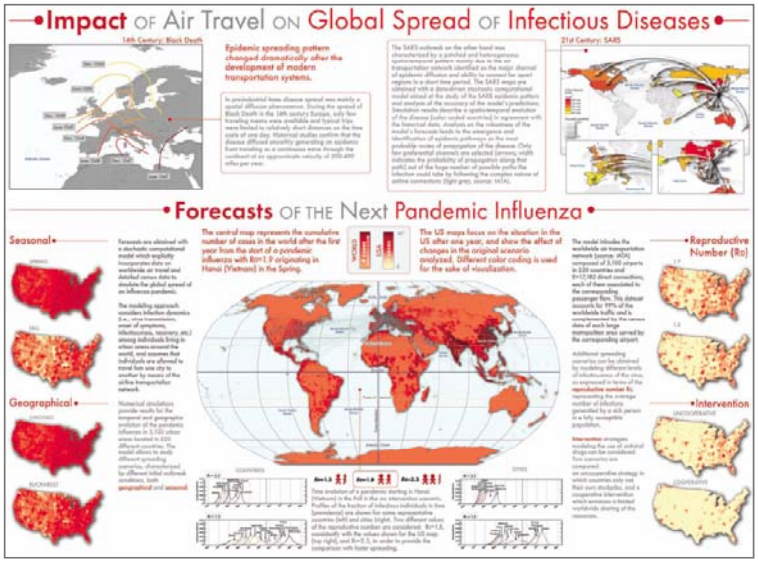
Computational Epidemics

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

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

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

NWB Deliverables

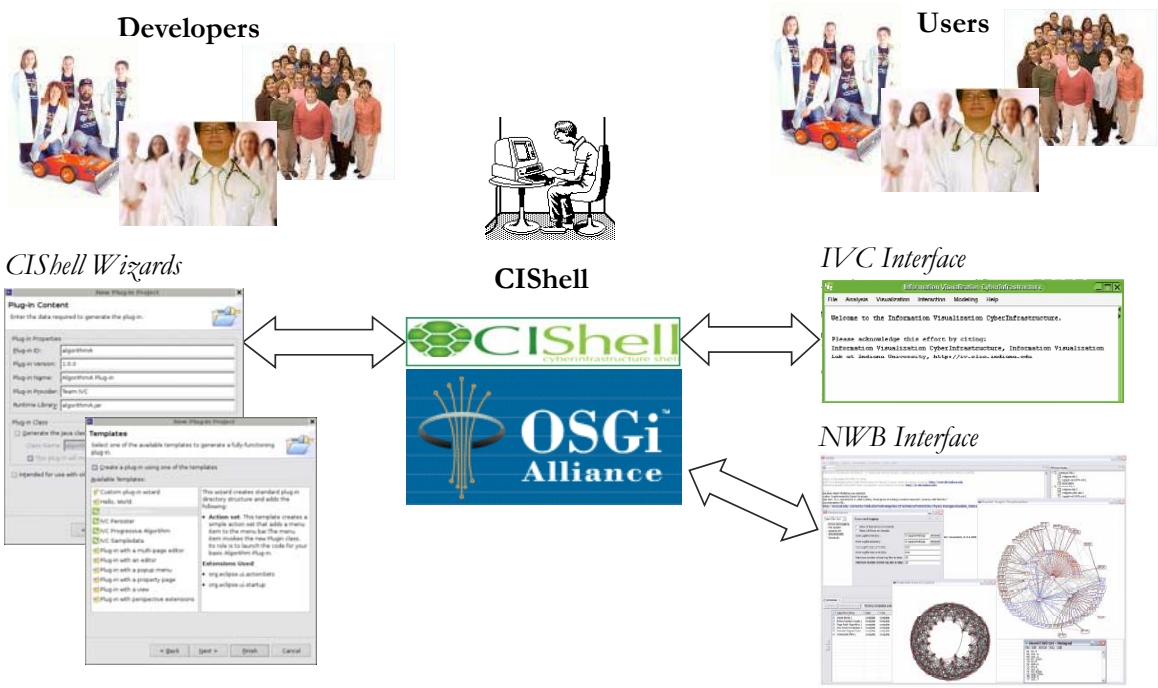
Network Workbench (NWB) Tool

- 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.

CIShell – Serving Non-CS Algorithm Developers & Users



CIShell – Builds on OSGi Industry Standard

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 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

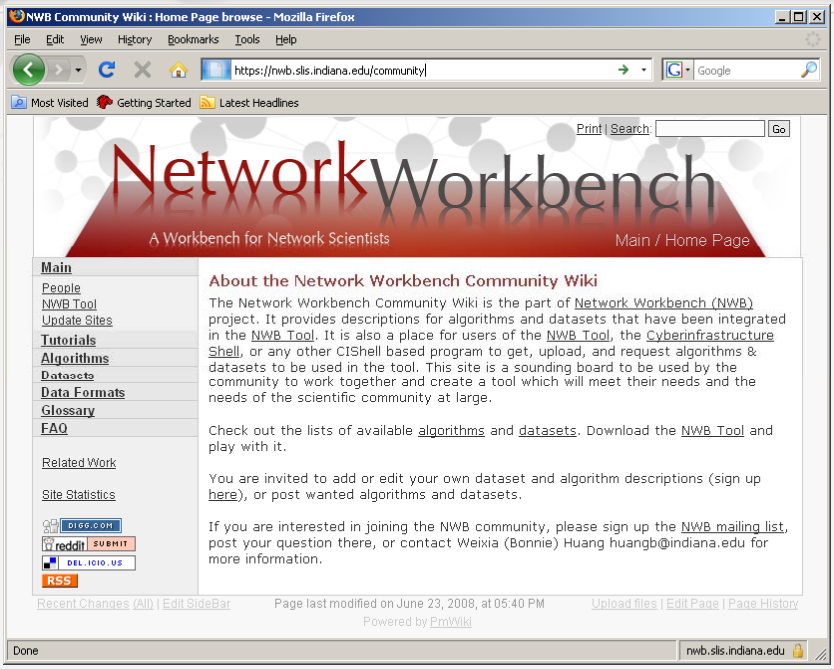
- A network analysis, modeling, and visualization toolkit for physics, biomedical, and social science research.
- Install and run on multiple Operating Systems.
- Supports many file formats.
- Easy integration of new algorithms thanks to CIShell/OSGi.

Cyberinfrastructure Shell (CIShell)

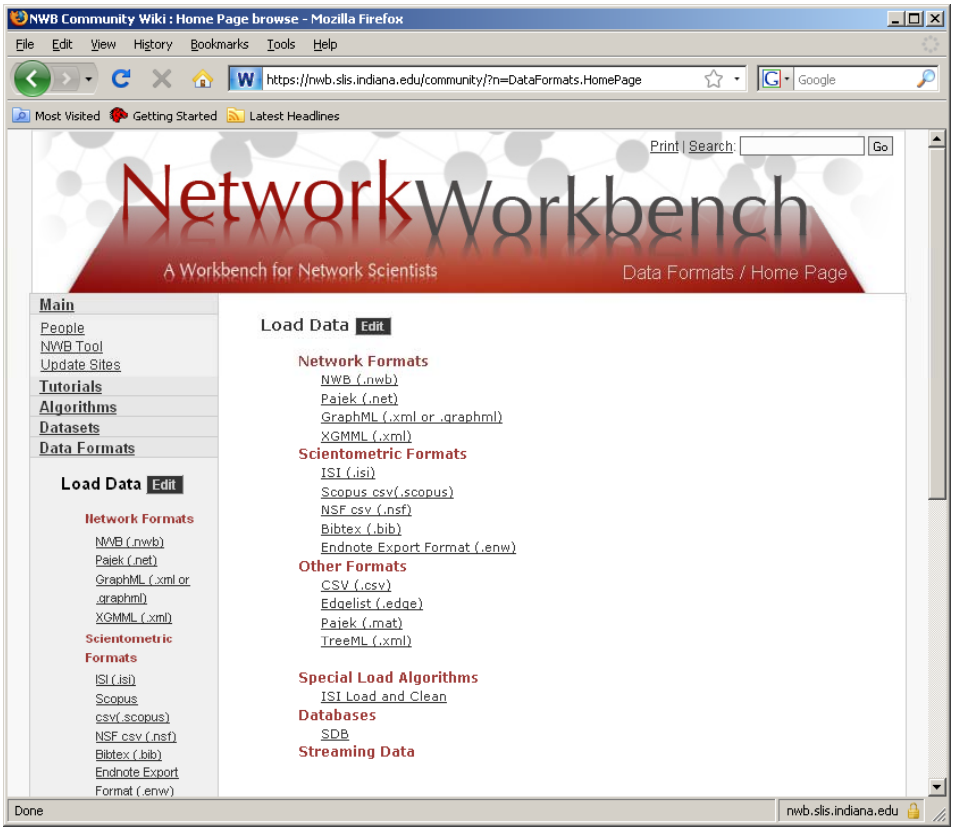
- An open source, software framework for the integration and utilization of datasets, algorithms, tools, and computing resources.
- Extends OSGi industry standard.

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.



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



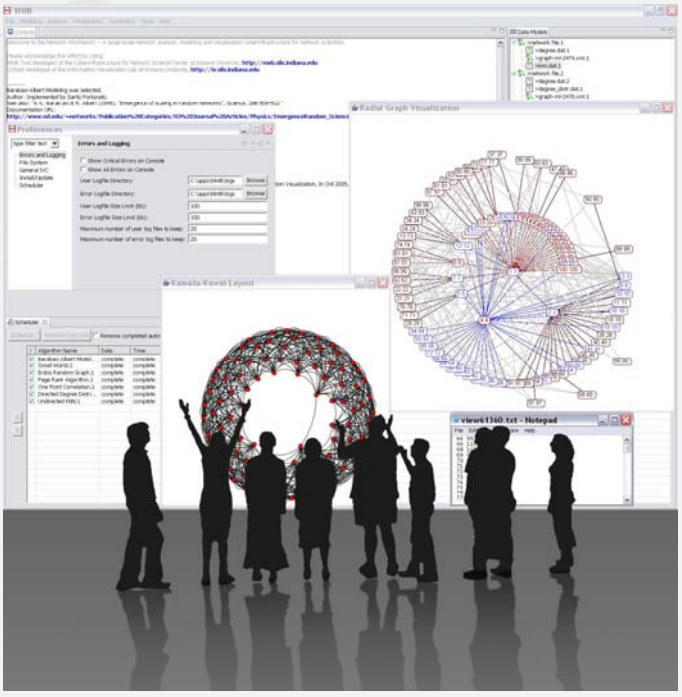
 https://nwb.slis.indiana.edu/community/

NWB Tool Release (Nov 2008)

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.

Demo



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.

NetworkWorkbench
A Workbench for Network Scientists

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

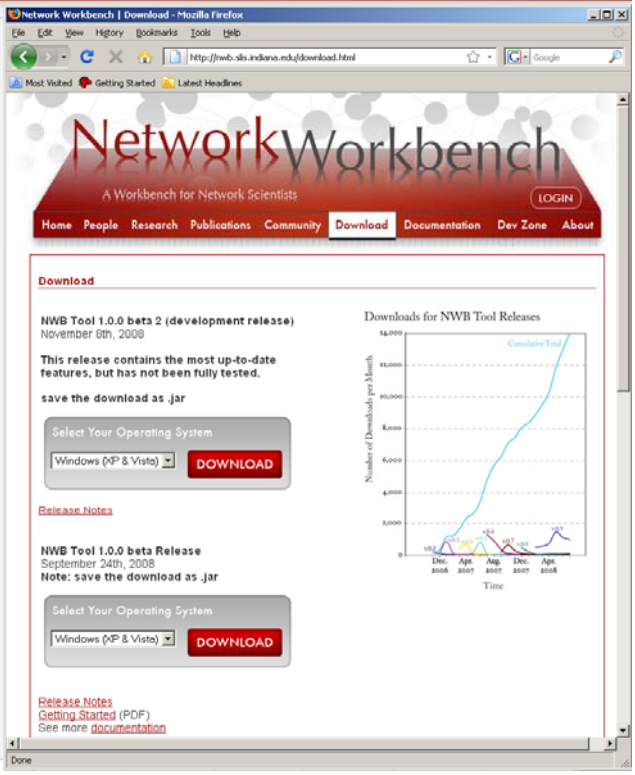
NWB Tool 1.0.0 beta 2
(development release)
November 19th, 2008

Select your operating system from the pull down menu.

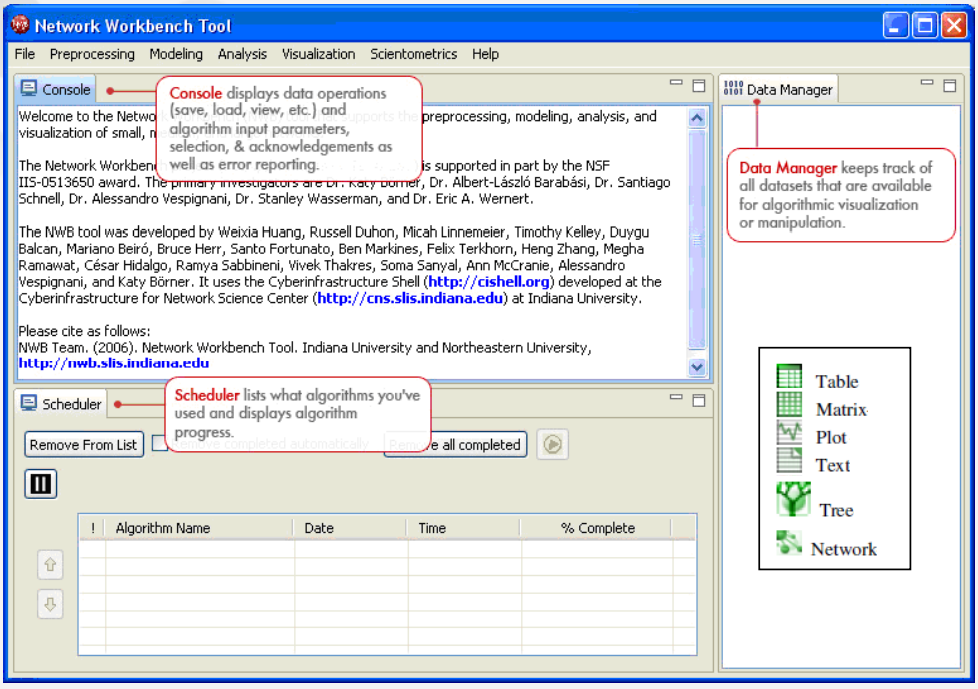
Save as *.jar file.

Install and run.

Session log files are stored in '**yournwbdirectory*/logs'* directory.

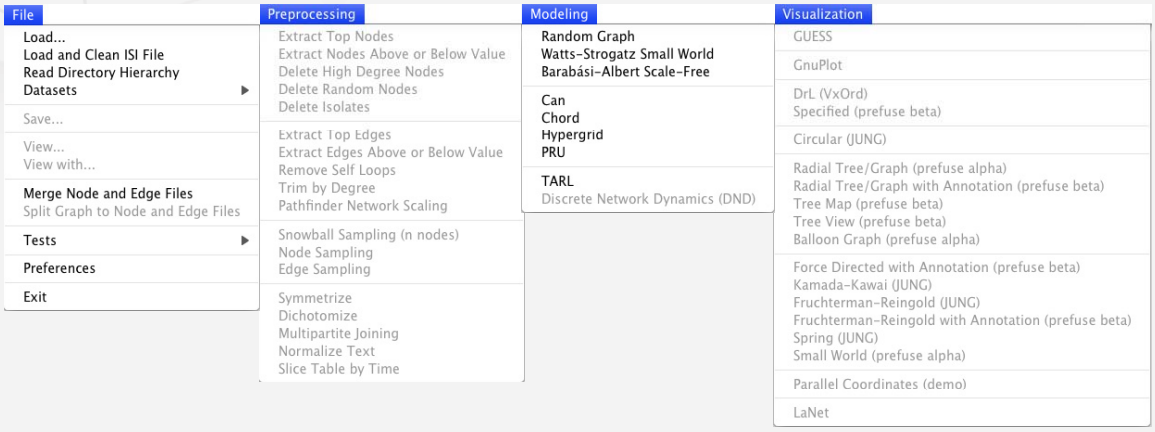


NWB Tool Interface Components



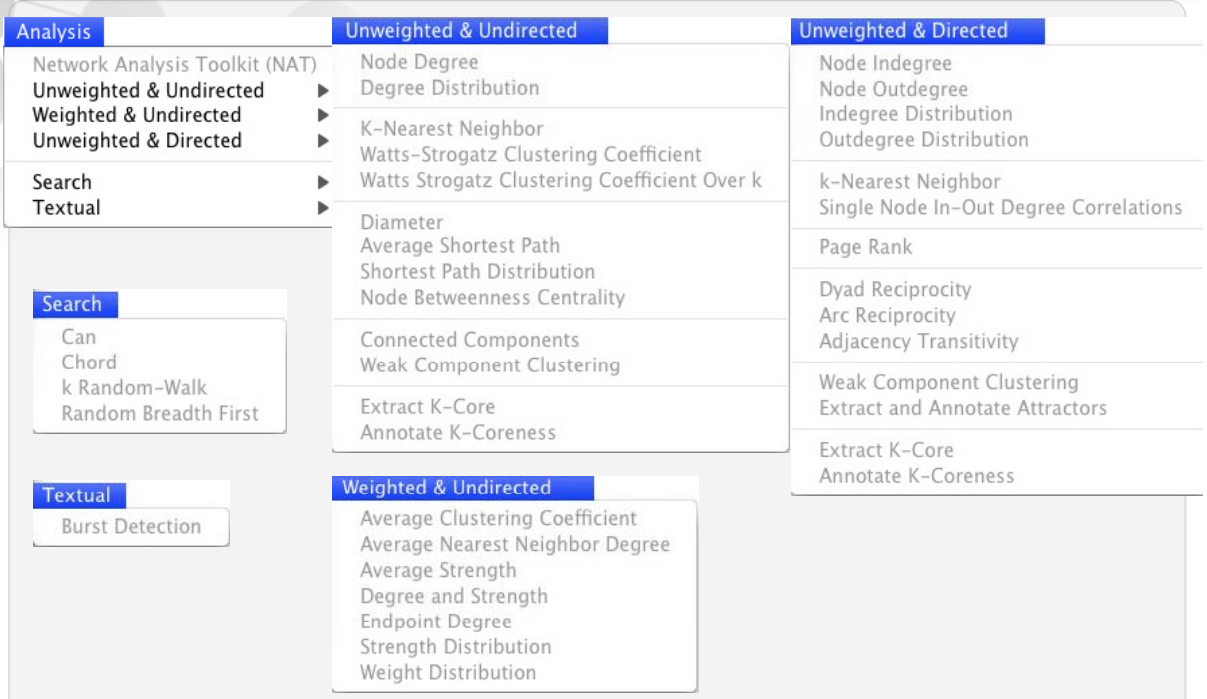
NetworkWorkbench
A Workbench for Network Scientists

File, Preprocessing, Modeling, and Visualization Menus

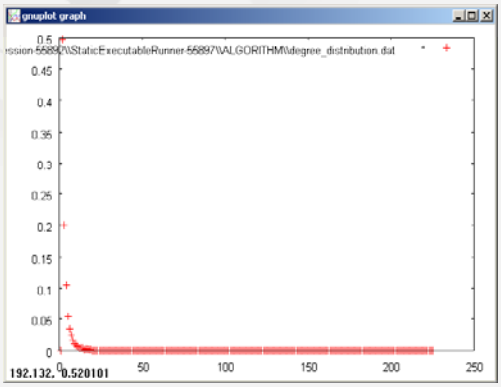


NetworkWorkbench
A Workbench for Network Scientists

Analysis Menu and Submenus

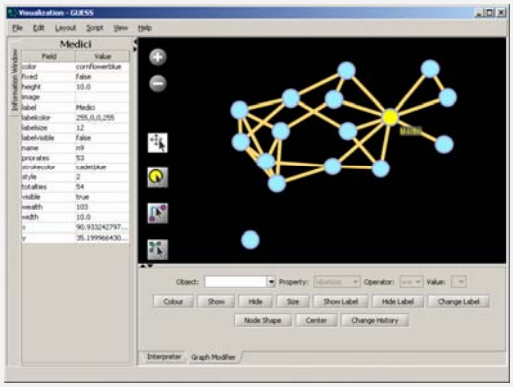


Integrated Tools



Gnuplot

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



GUESS

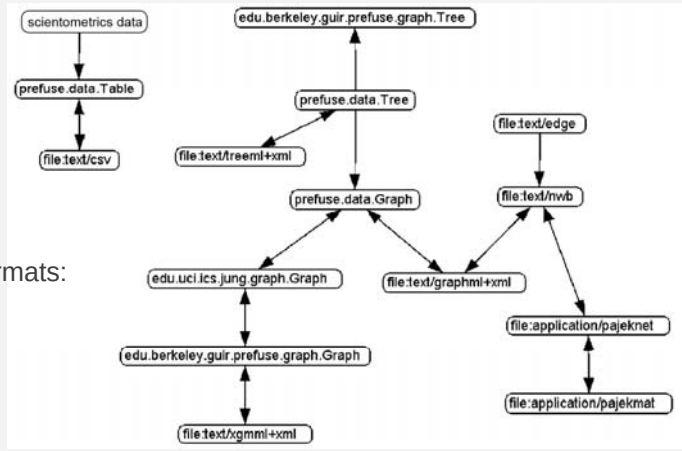
exploratory data analysis and visualization tool for graphs and networks.

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:

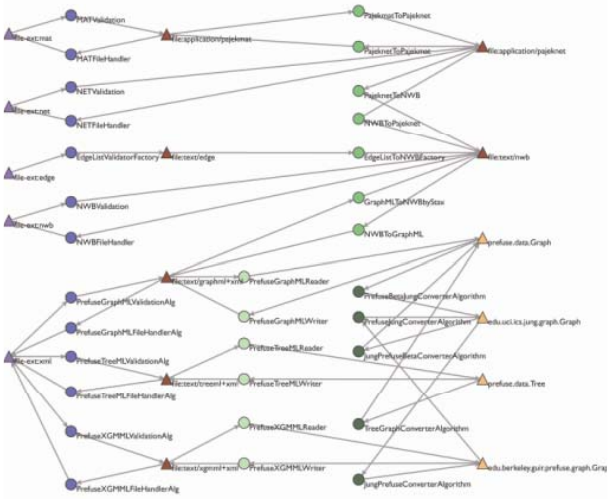
- GraphML (*.xml or *.graphml)
- Pajek .MAT (*.mat)
- Pajek .NET (*.net)
- NWB (*.nwb)
- XGMML (*.xml)
- CSV (*.csv)

These formats are documented at https://nwb.slis.indiana.edu/community/?n=DataFormats.HomePage.

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

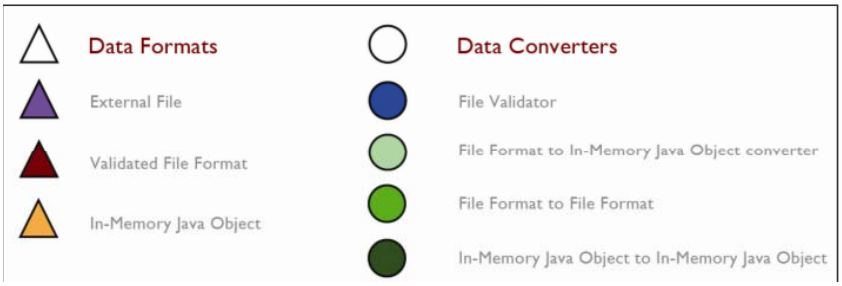


6
Output formats for diverse visualization algorithms

8
Intermediate data formats

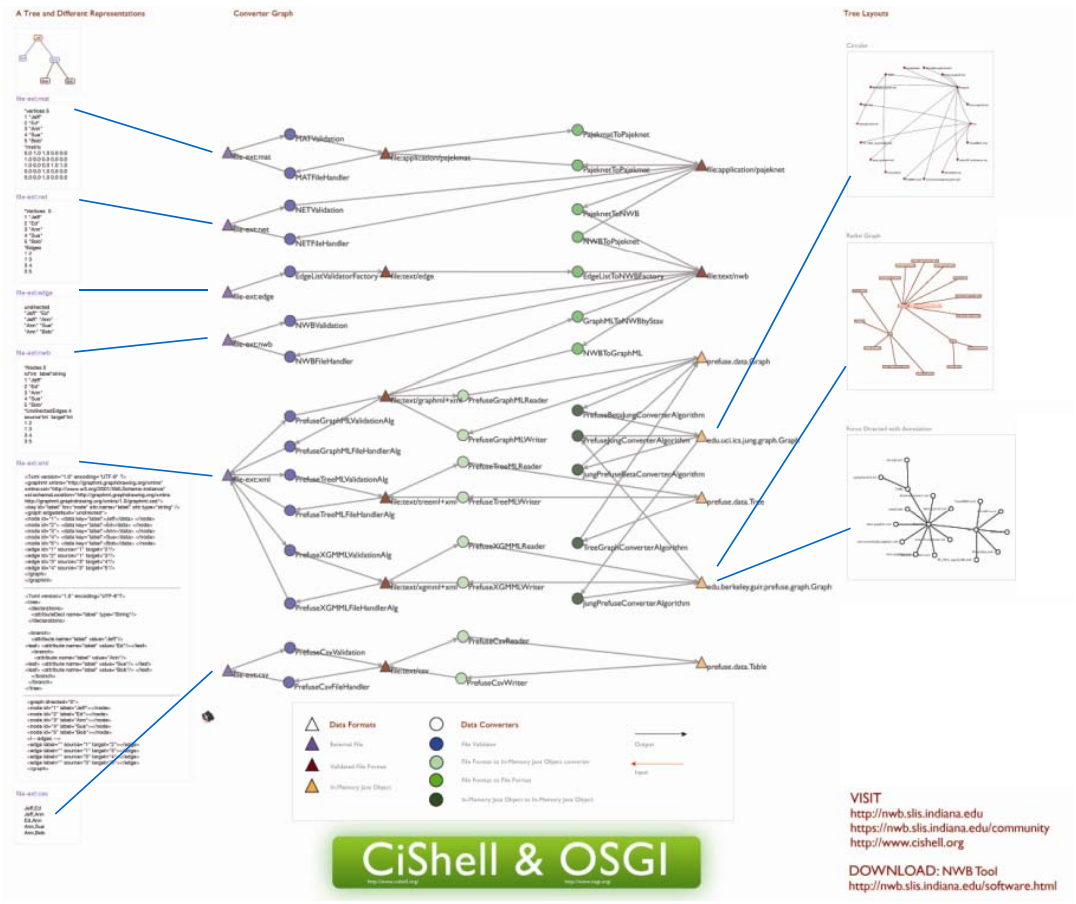
Supported by

35
data converters.



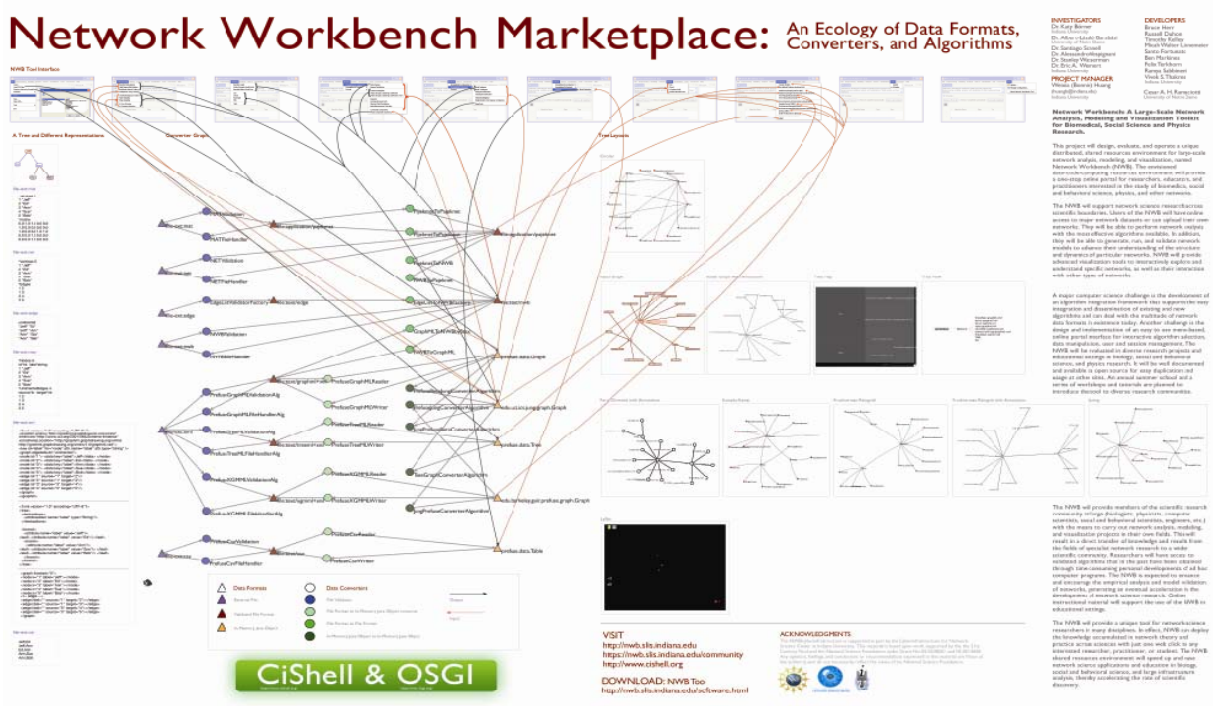
VISIT
http://nwb.slis.indiana.edu
https://nwb.slis.indiana.edu/community
http://www.cishell.org

DOWNLOAD: NWB Tool
http://nwb.slis.indiana.edu/software.html



VISIT
http://nwb.slis.indiana.edu
https://nwb.slis.indiana.edu/community
http://www.cishell.org

DOWNLOAD: NWB Tool
http://nwb.slis.indiana.edu/software.html



NetworkWorkbench
A Workbench for Network Scientists

Sample Datasets

The '**yournwbdirectory*/sampledata'* directory provides sample datasets from the biology, network, scientometics, 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

/scientometrics/scopus
/socialscience
- florentine.nwb

The blue ones are used in this tutorial.

Property Files and Python Scripts

The '**yournwbdirectory*/*" 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
- mergeNsfPIs.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

NWB Tool Overview

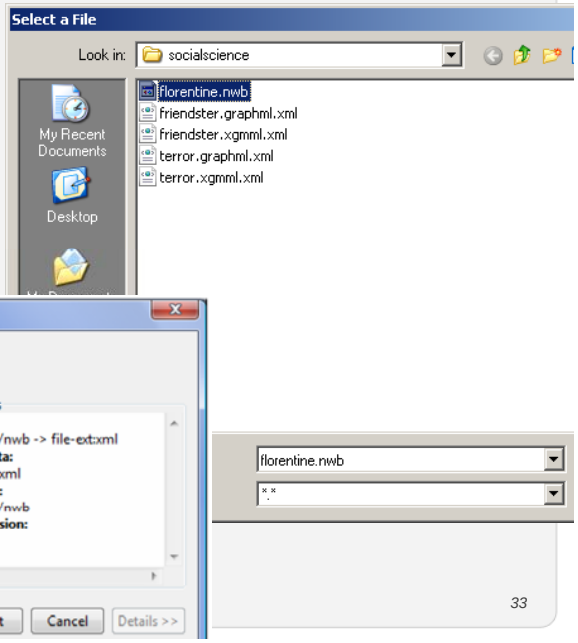
1. Download, install, and run.
2. **Load, view, convert, save data.**
3. Read and visualize a directory hierarchy.
4. Load a network, compute its basic properties, and explore it in GUESS.

Load, View and Save (Convert) Data

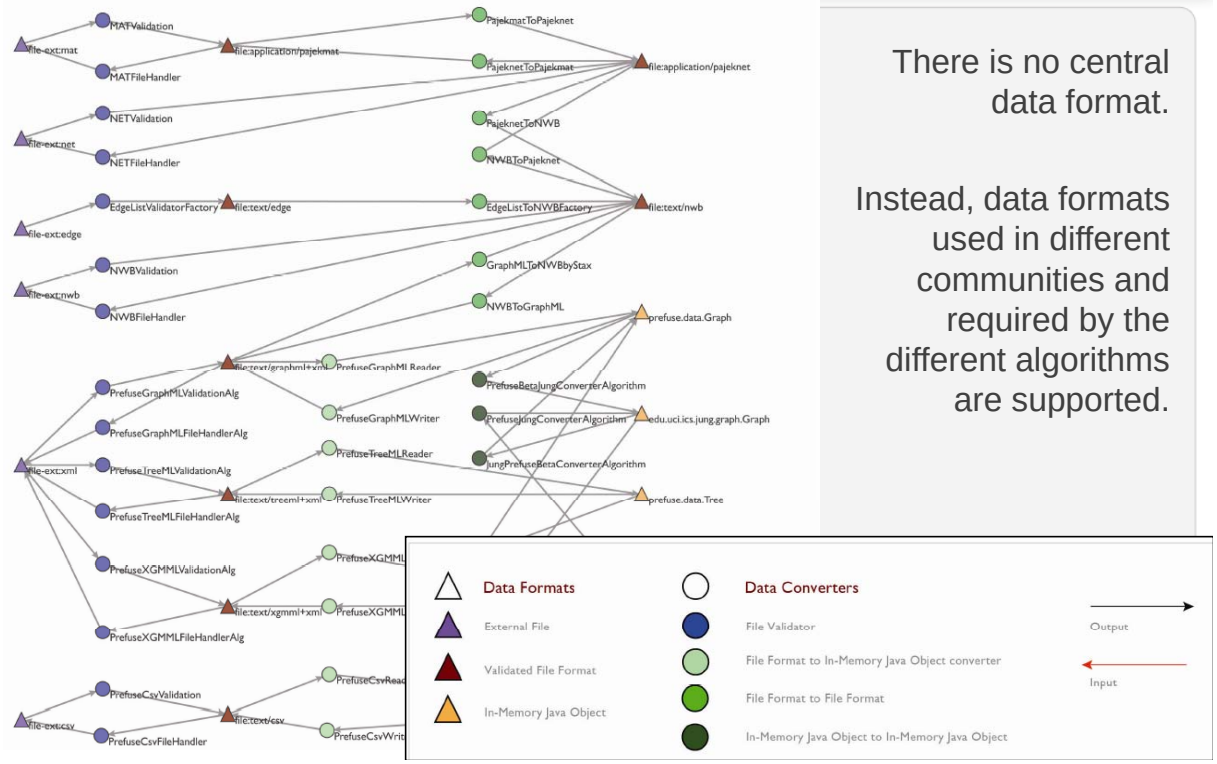
Use *'File > Load File'* to load *florentine.nwb* in sample datasets in '**yournwbdirectory*/sampledata/socialscience*'.

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.

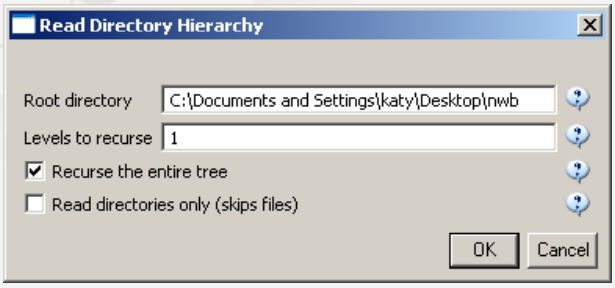
Instead, data formats used in different communities and required by the different algorithms are supported.

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.

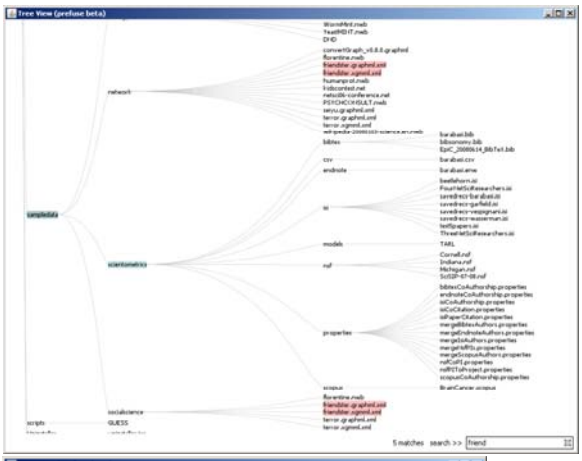
Reading and Visualizing a Directory Hierarchy

Use *'File > Read Directory Hierarchy'* with parameters



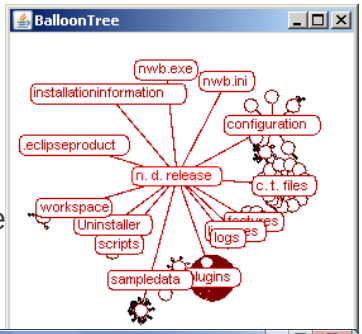
Visualize resulting *'Directory Tree - Prefuse (Beta) Graph'* using

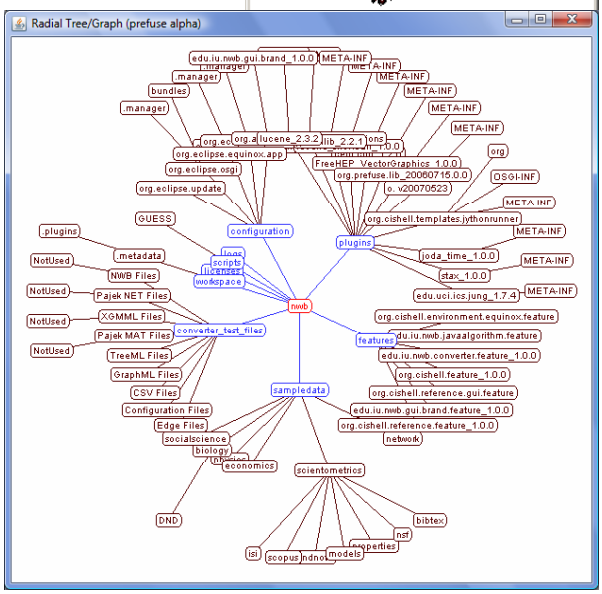
- *'Visualization > Tree View (prefuse beta)'*
- *'Visualization > Tree Map (prefuse beta)'*
- *'Visualization > Balloon Graph (prefuse alpha)'*
- *'Visualization > Radial Tree/Graph (prefuse alpha)'*



Different views of the /nwb directory hierarchy.

Note the size of the /plugin directory.





NetworkWorkbench
A Workbench for Network Scientists

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.**

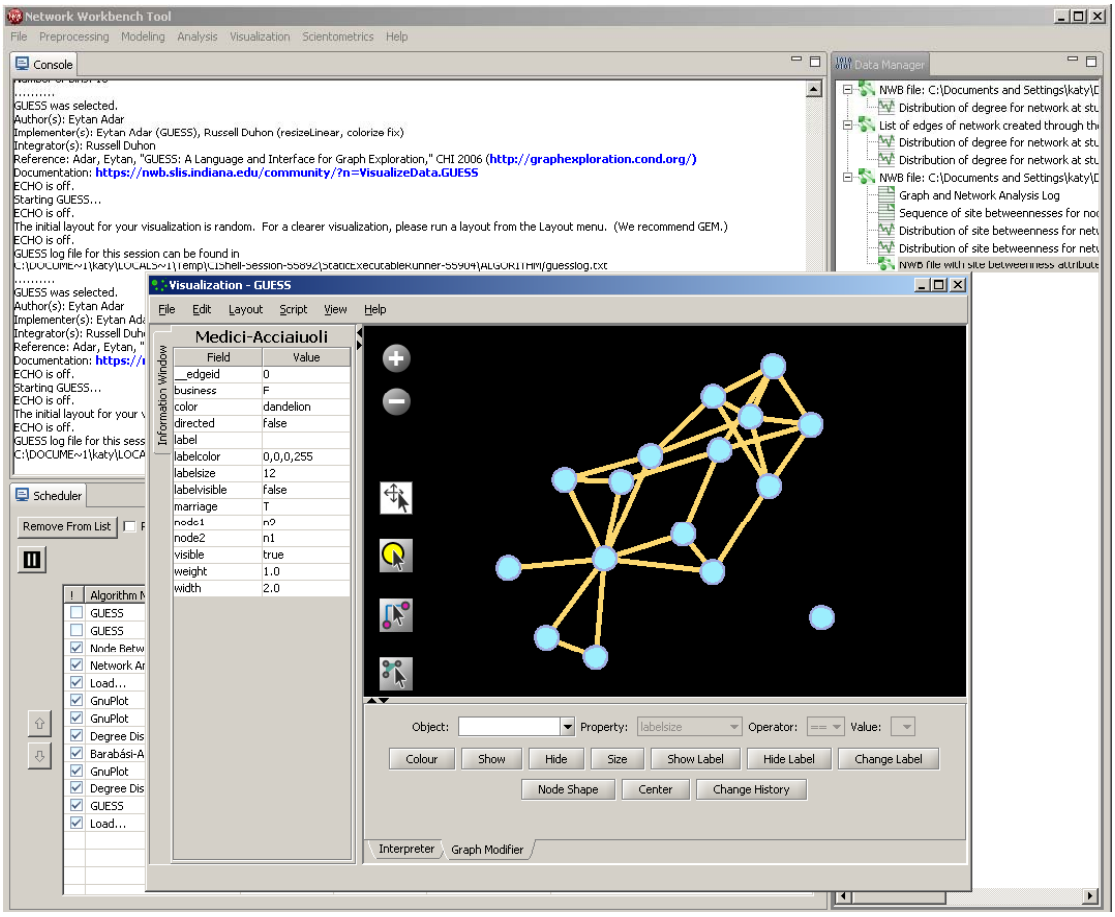
Compute Basic NW Properties & View in GUESS

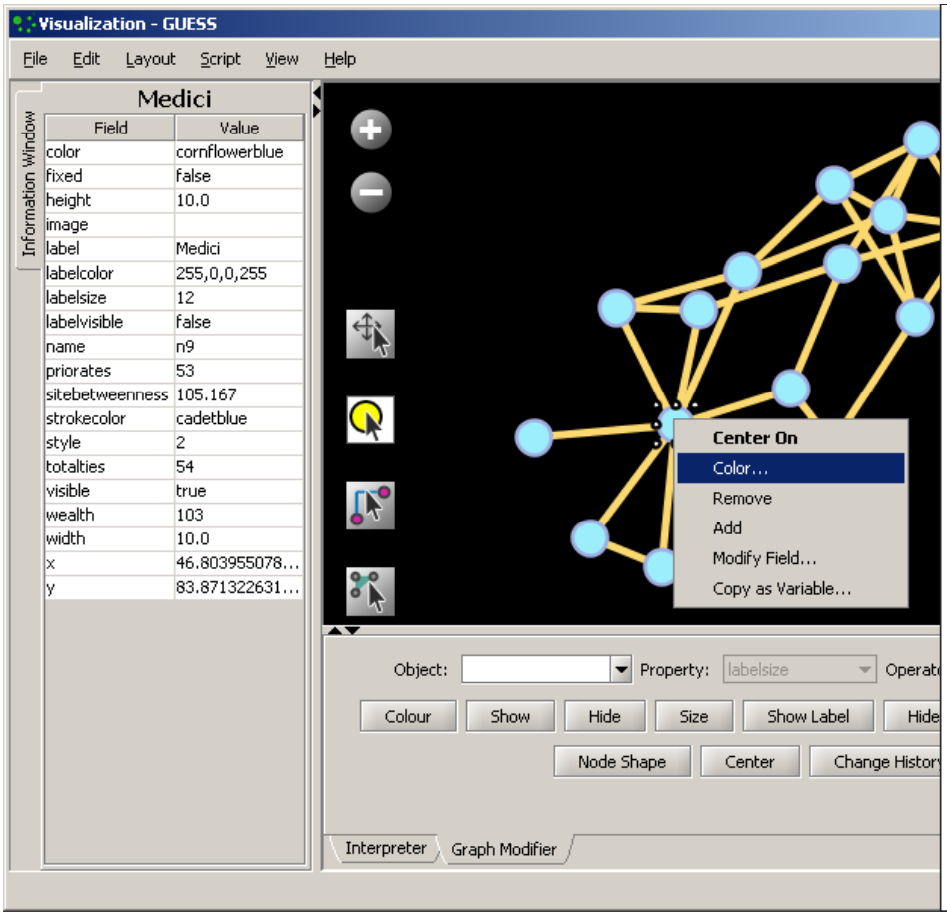
Select *florentine.nwb* in Data Manager.

- 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, totalties, priorates

Edges: 27
No self loops were discovered.
No parallel edges were discovered.
Edge attributes:
  Nonnumeric attributes:
                        Example value
        marriag...      T
        busines...      F

  Did not detect any numeric attributes
  This network does not seem to be a valued network.

Average degree: 3.375
This graph is not weakly connected.
There are 2 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
```

- Optional: Run 'Analysis -> Unweighted & Undirected -> Node Betweenness Centrality' with default parameters.
- Select network and run *'Visualization > GUESS'* to open GUESS with file loaded.
- Apply 'Layout -> GEM'.





Pan:

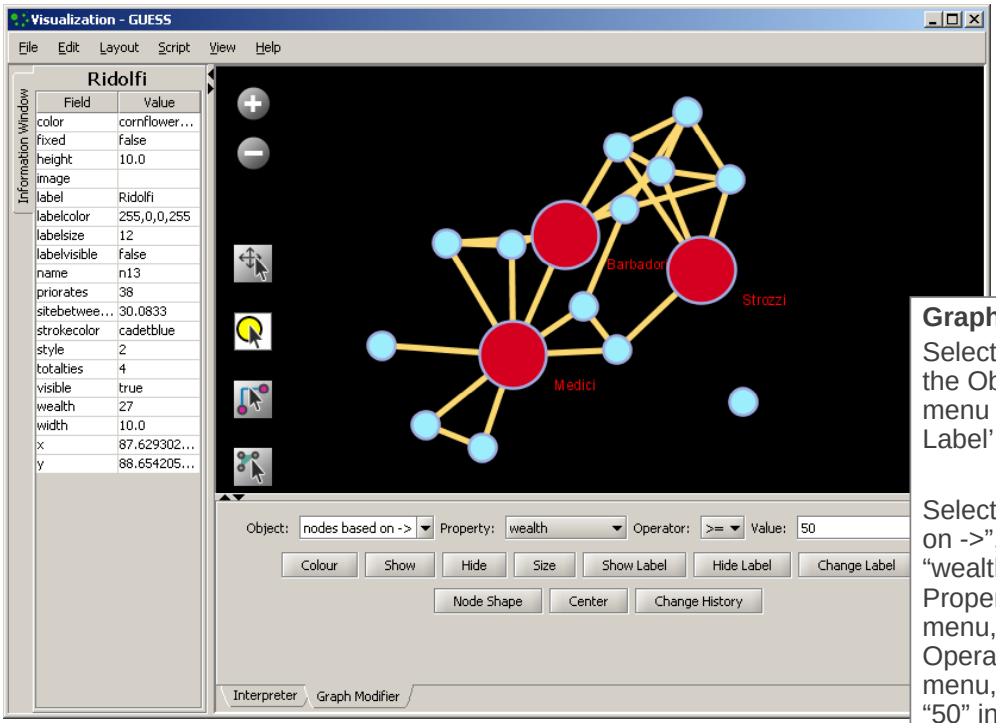
"grab" the background by holding left-click and moving your mouse.

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.

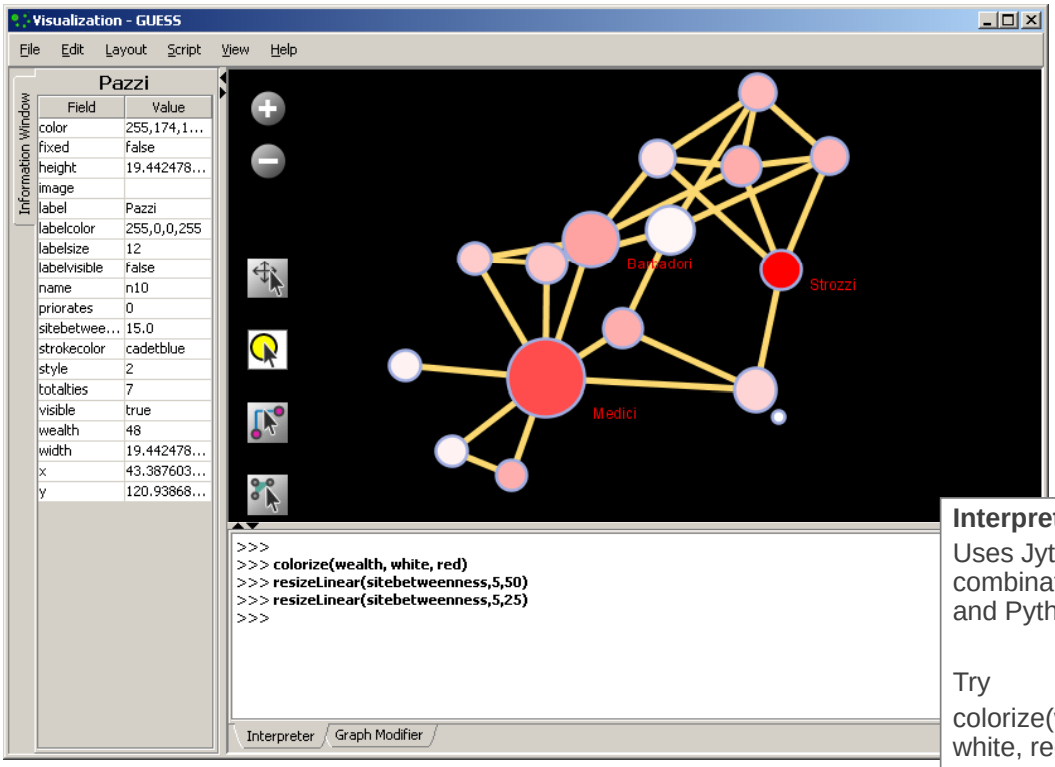
Right click to modify Color, etc.



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

NetworkWorkbench
A Workbench for Network Scientists

NWB Tool for Scientometrics Research

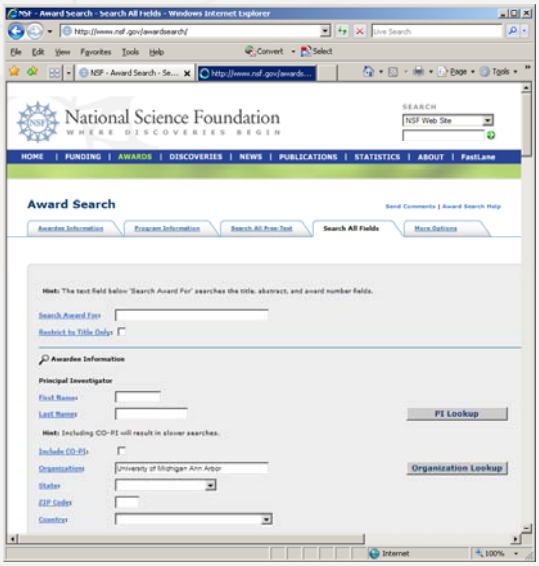
1. **Loading NSF datasets with currently active NSF funding for Indiana U, Cornell U, and Michigan U, extracting, and comparing Co-PI networks.**
2. **Loading ISI files of major network science researchers, extracting, analyzing and visualizing paper-citation networks and co-author networks.**
3. **Loading text files and detecting Bursts.**

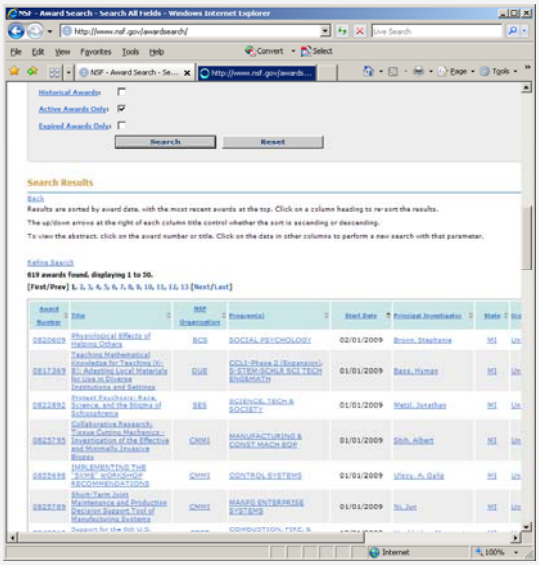
NWB Tool for Scientometrics Research

1. **Loading NSF datasets with currently active NSF funding for Indiana U, Cornell U, and Michigan U, extracting, and comparing Co-PI networks.**
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.**

Analyzing NSF Data

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)

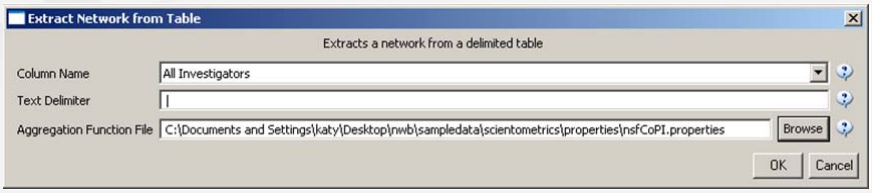
Save files as csv but rename into .nsf.

Or simply use the files saved in
'**yournwbdirectory*/sampledata/scientometrics/nsf/*'.

Analyzing NSF Data

Extracting Co-PI Networks

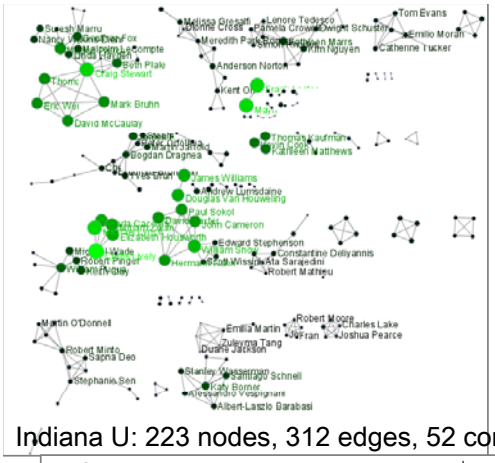
Load NSF data, selecting the loaded dataset in the Data Manager window, run '*Scientometrics > Extract Co-Occurrence Network'* using parameters:



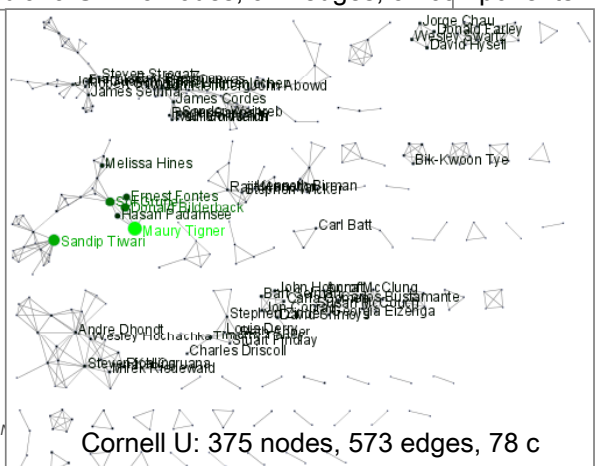
Two derived files will appear in the Data Manager window: the co-PI network and a merge table. In the network, nodes represent investigators and edges denote their co-PI relationships. The merge table can be used to further clean PI names.

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.



Indiana U: 223 nodes, 312 edges, 52 components



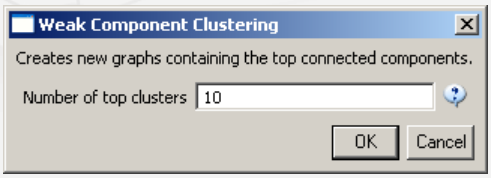
Cornell U: 375 nodes, 573 edges, 78 c

U of Michigan: 497 nodes, 672 edges, 117 c

NetworkWorkbench

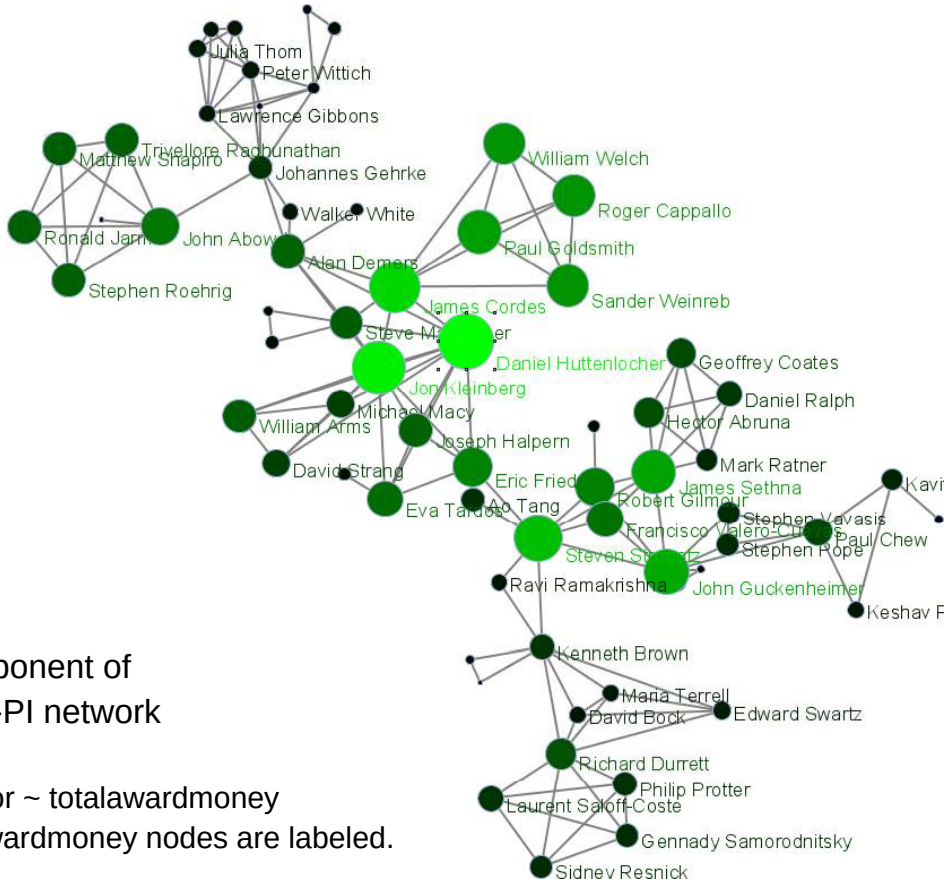
A Workbench for Network Scientists

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 ~ totalawardmoney
Top-50 totalawardmoney nodes are labeled.

NetworkWorkbench
A Workbench for Network Scientists

Top-10 investigators by total award money

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

Indiana University		Cornell University		Michigan University	
Curtis Lively:	7,436,828	Maury Tigner:	107,216,976	Khalil Najafi:	32,541,158
Frank Lester:	6,402,330	Sandip Tiwari:	72,094,578	Kensall Wise:	32,164,404
Maynard Thompson:	6,402,330	Sol Gruner:	48,469,991	Jacquelynne Eccles:	25,890,711
Michael Lynch:	6,361,796	Donald Bilderback:	47,360,053	Georg Raithel:	23,832,421
Craig Stewart:	6,216,352	Ernest Fontes:	29,380,053	Roseanne Sension:	23,812,921
William Snow:	5,434,796	Hasan Padamsee:	18,292,000	Theodore Norris:	23,35,0921
Douglas V. Houweling:	5,068,122	Melissa Hines:	13,099,545	Paul Berman:	23,350,921
James Williams:	5,068,122	Daniel Huttenlocher:	7,614,326	Roberto Merlin:	23,350,921
Miriam Zolan:	5,000,627	Timothy Fahey:	7,223,112	Robert Schoeni:	21,991,140
Carla Caceres:	5,000,627	Jon Kleinberg:	7,165,507	Wei-Jun Jean Yeung:	21,991,140

NWB Tool for Scientometrics Research

1. Loading NSF datasets with currently active NSF funding for Indiana U, Cornell U, and Michigan U, extracting, and comparing Co-PI networks.
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.

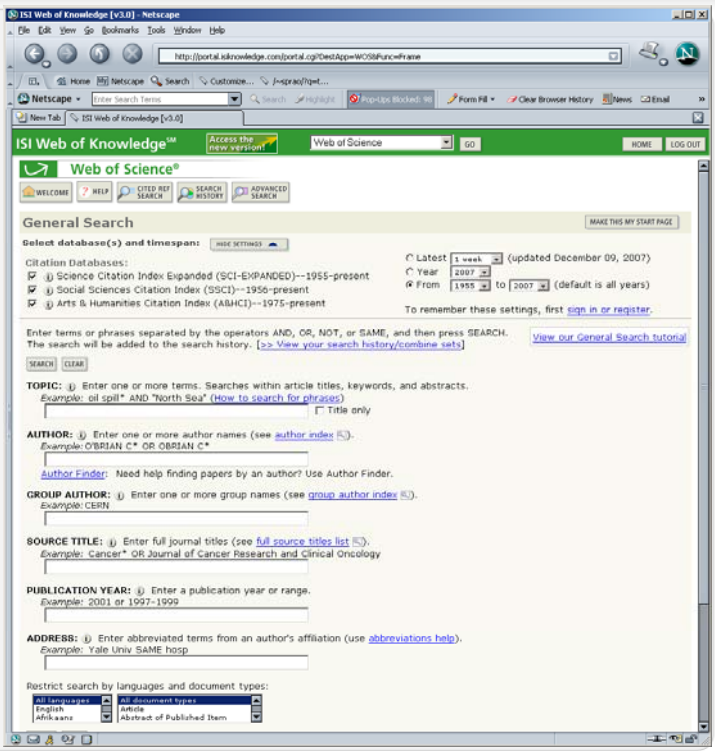
Data Acquisition from Web of Science

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

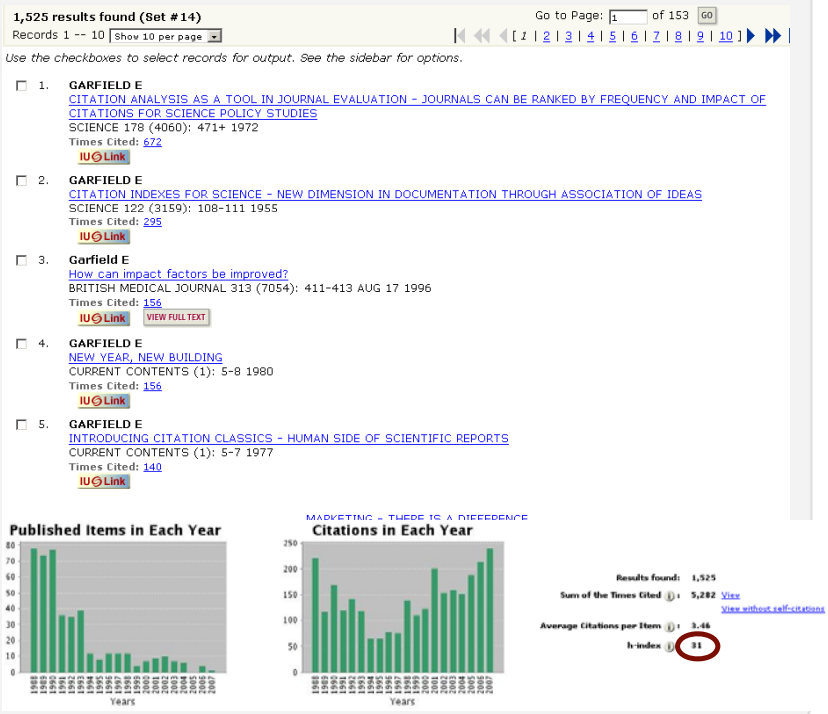


Data Acquisition from Web of Science (cont.)

Eugene Garfield

1,525 papers

papers/citations for last 20 years

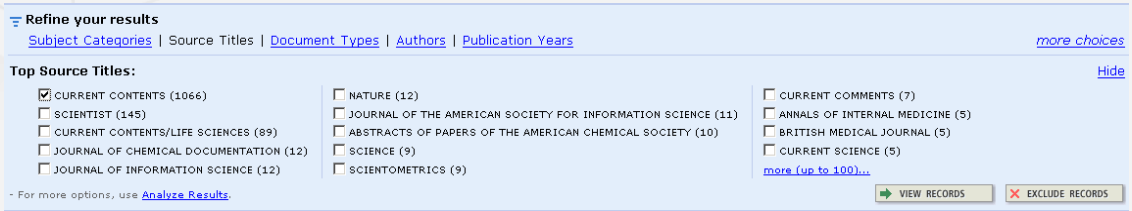


Network Workbench (http://nwb.slis.indiana.edu).

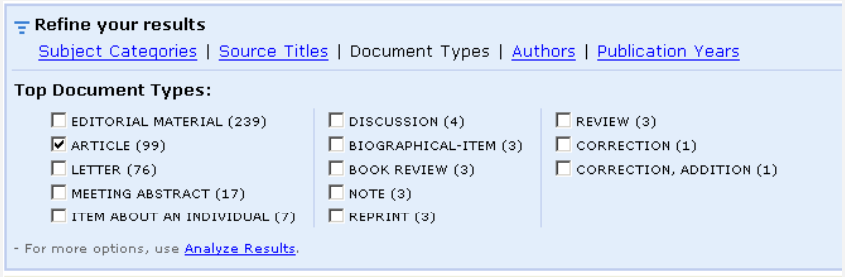
Data Acquisition from Web of Science (cont.)

Can download 500 records max.

Exclude Current Contents articles



Include only articles. Download 99 articles.



Data Acquisition from Web of Science (cont.)

99 results found
Records 1 -- 10 Show 10 per page
Go to Page: 1 of 10 GO
[1 | 2 ...]

Use the checkboxes to select records for output. See the sidebar for options.

1. **GARFIELD E**
CITATION ANALYSIS AS A TOOL IN JOURNAL EVALUATION - JOURNALS CAN BE RANKED BY FREQUENCY AND IMPACT OF CITATIONS FOR SCIENCE POLICY STUDIES
SCIENCE 178 (4060): 471+ 1972
Times Cited: 672
IUGLink

2. **GARFIELD E**
CITATION INDEXES FOR SCIENCE - NEW DIMENSION IN DOCUMENTATION THROUGH ASSOCIATION OF IDEAS
SCIENCE 122 (3159): 108-111 1955
Times Cited: 295
IUGLink

3. **Garfield E**
How can impact factors be improved?
BRITISH MEDICAL JOURNAL 313 (7054): 411-413 AUG 17 1996
Times Cited: 156
IUGLink VIEW FULL TEXT

4. **GARFIELD E**
CITATION INDEXING FOR STUDYING SCIENCE
NATURE 227 (5259): 669& 1970
Times Cited: 134
IUGLink

5. **GARFIELD E**
SCIENCE CITATION INDEX-NEW DIMENSION IN INDEXING - UNIQUE APPROACH UNDERLIES VERSATILE BIBLIOGRAPHIC SYSTEMS FOR COMMUNICATING + EVALUATING INFORMATION
SCIENCE 144 (361): 649& 1964
Times Cited: 109
IUGLink

Data Acquisition from Web of Science (cont.)

Stanley Wasserman

35 papers

35 results found (Set #7) (Why 35?)
Records 1 -- 10 Show 10 per page
Go to Page: 1 of 4 GO
[1 | 2 | 3 | 4]

Use the checkboxes to select records for output. See the sidebar for options.

1. GALASKIEWICZ J, WASSERMAN S
MIMETIC PROCESSES WITHIN AN INTERORGANIZATIONAL FIELD - AN EMPIRICAL-TEST
ADMINISTRATIVE SCIENCE QUARTERLY 34 (3): 454-479 SEP 1989
Times Cited: 122
IUGLink

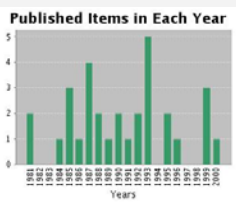
2. Wasserman S, Pattison P
Logit models and logistic regressions for social networks .1. An introduction to Markov graphs and p
PSYCHOMETRIKA 61 (3): 401-425 SEP 1996
Times Cited: 99
IUGLink

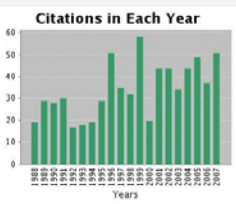
3. FIENBERG SE, MEYER MM, WASSERMAN SS
STATISTICAL-ANALYSIS OF MULTIPLE SOCIOMETRIC RELATIONS
JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION 80 (389): 51-67 1985
Times Cited: 47
IUGLink

4. WASSERMAN S
ANALYZING SOCIAL NETWORKS AS STOCHASTIC-PROCESSES
JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION 75 (370): 280-294 1980
Times Cited: 38
IUGLink

5. IACOBUCCI D, WASSERMAN S
A GENERAL FRAMEWORK FOR THE STATISTICAL-ANALYSIS OF SEQUENTIAL DYADIC INTERACTION DATA
PSYCHOLOGICAL BULLETIN 103 (3): 379-390 MAY 1988
Times Cited: 34
IUGLink

papers/citations for last 20 years



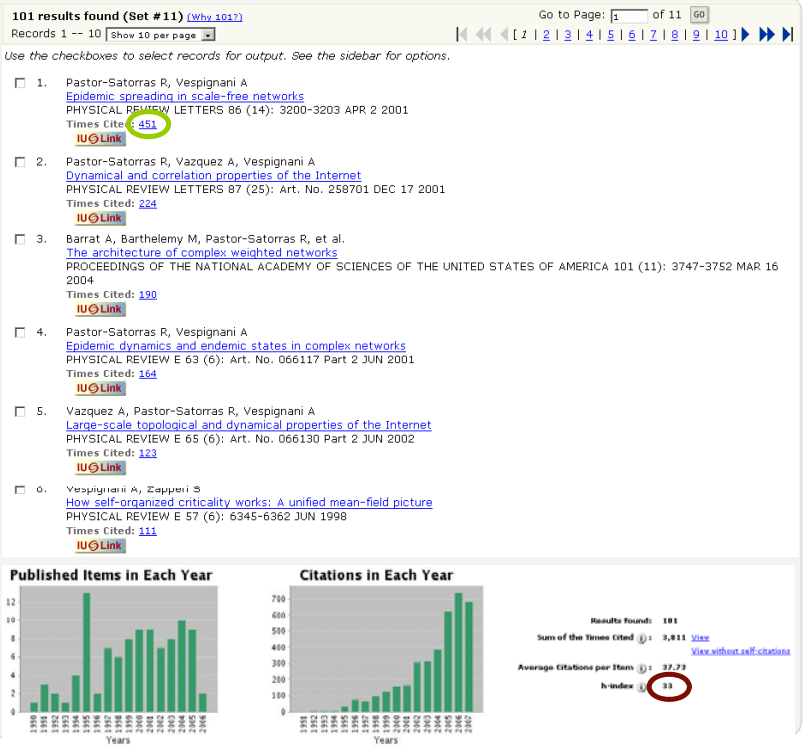


Data Acquisition from Web of Science (cont.)

Alessandro Vespignani

101 papers

papers/citations for last 20 years



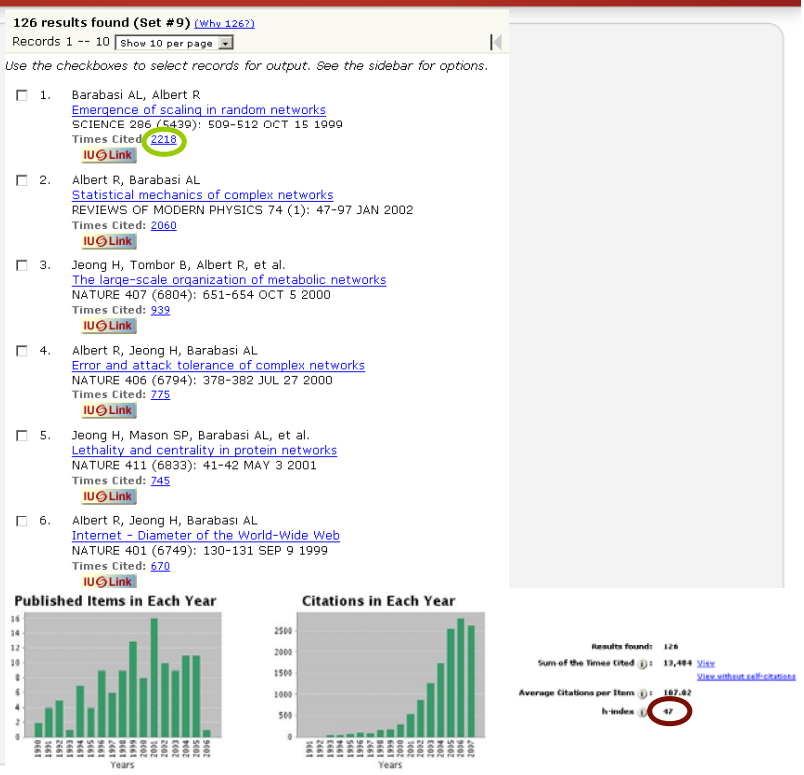
Network Workbench (http://nwb.slis.indiana.edu).

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).

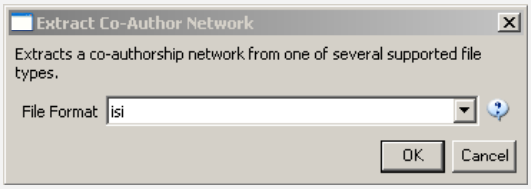
Comparison of Counts

	Age	Total # Cites	Total # Papers	H-Index	
Eugene Garfield	82	1,525	672	31	
Stanley Wasserman		122	35	17	
Alessandro Vespignani	42	451	101	33	
Albert-László Barabási	40	2,218	126	47	*(Dec 2007)*
	41	16,920	159	52	*(Nov 2008)*

Extract Co-Author Network

Load*yournwbdirectory*/sampledata/scientometrics/isi/FourNetSciResearchers.isi' using *'File > Load and Clean ISI File'.*

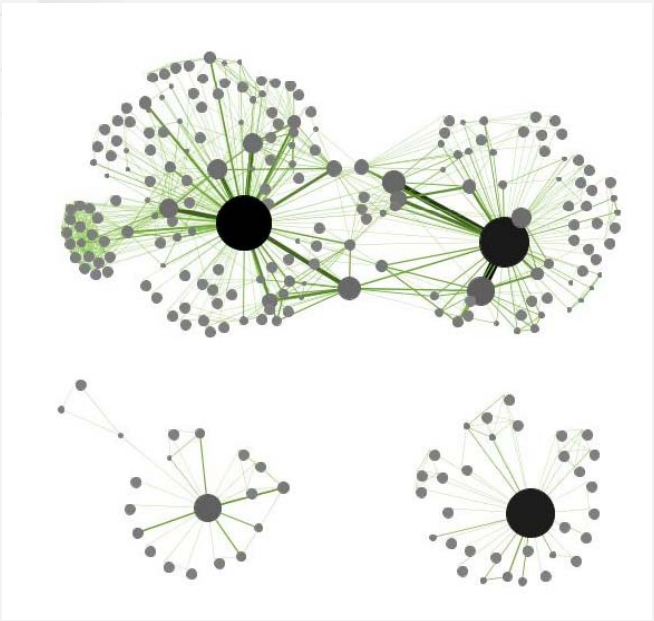
To extract the co-author network, select the *'361 Unique ISI Records'* table and run *'Scientometrics > Extract Co-Author Network'* using isi file format:



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

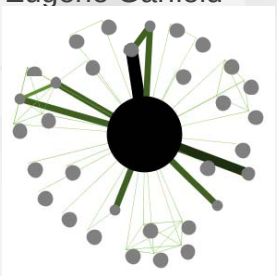
To view the complete network, select the network and run *'Visualization > GUESS > GEM'*. Run *Script > Run Script… . And select Script folder > GUESS > co-author-nw.py* .

Network of Wasserman, Vespignani and Barabási

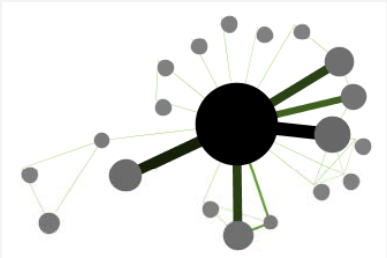


Comparison of Ego-Centric Co-Author Networks

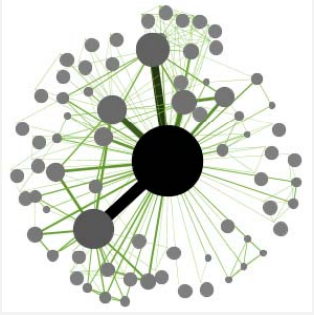
Eugene Garfield



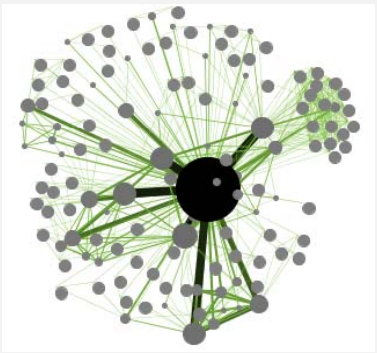
Stanley Wasserman



Alessandro Vespignani



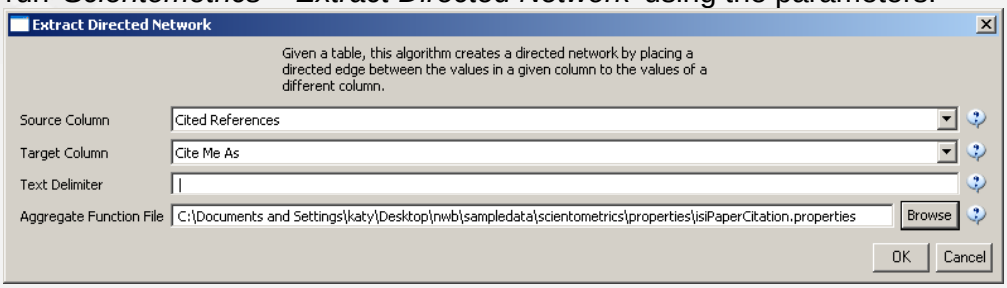
Albert-László Barabási



Paper-Citation Network Layout

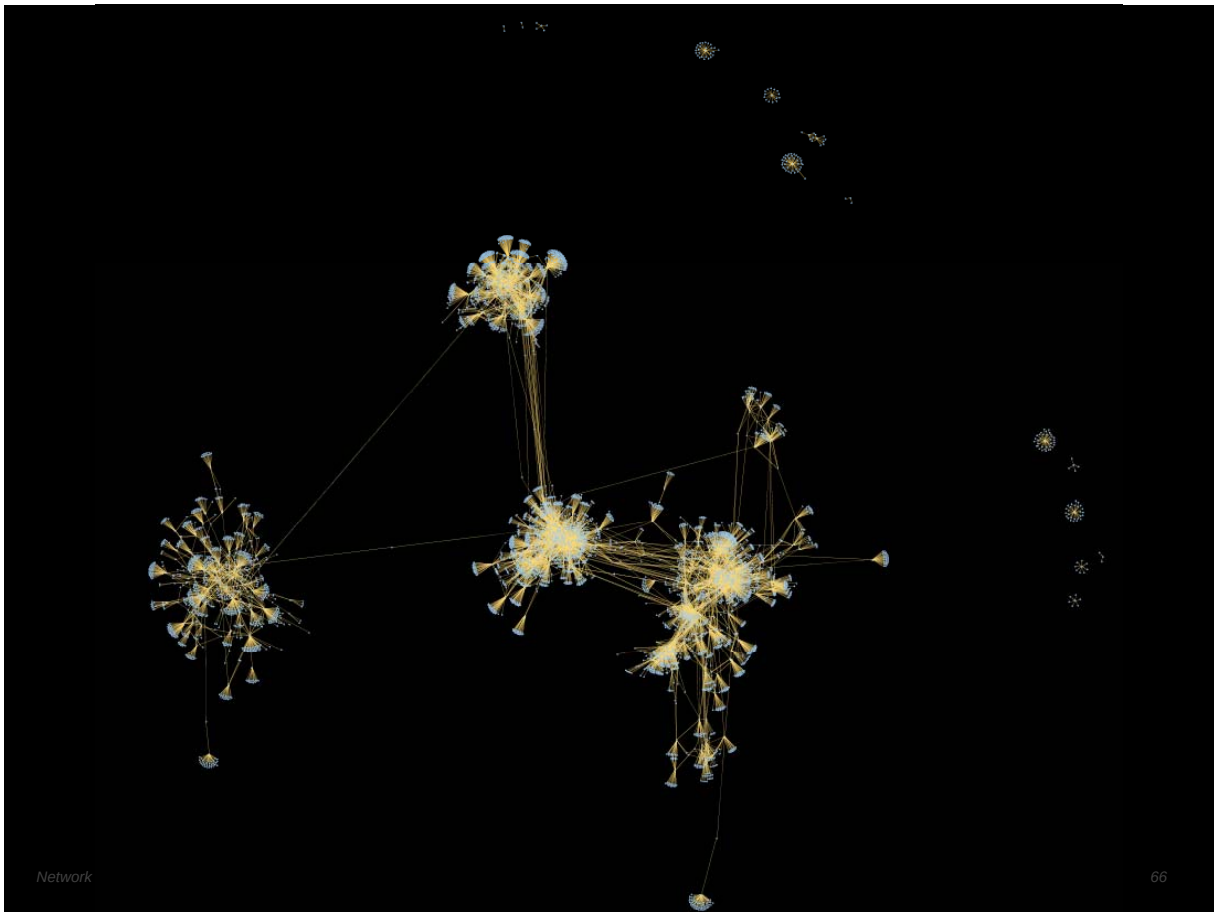
Load *'*yournwbdirectory*/sampledata/scientometrics/isi/FourNetSciResearchers.isi'* using *'File > Load and Clean ISI File'.*

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



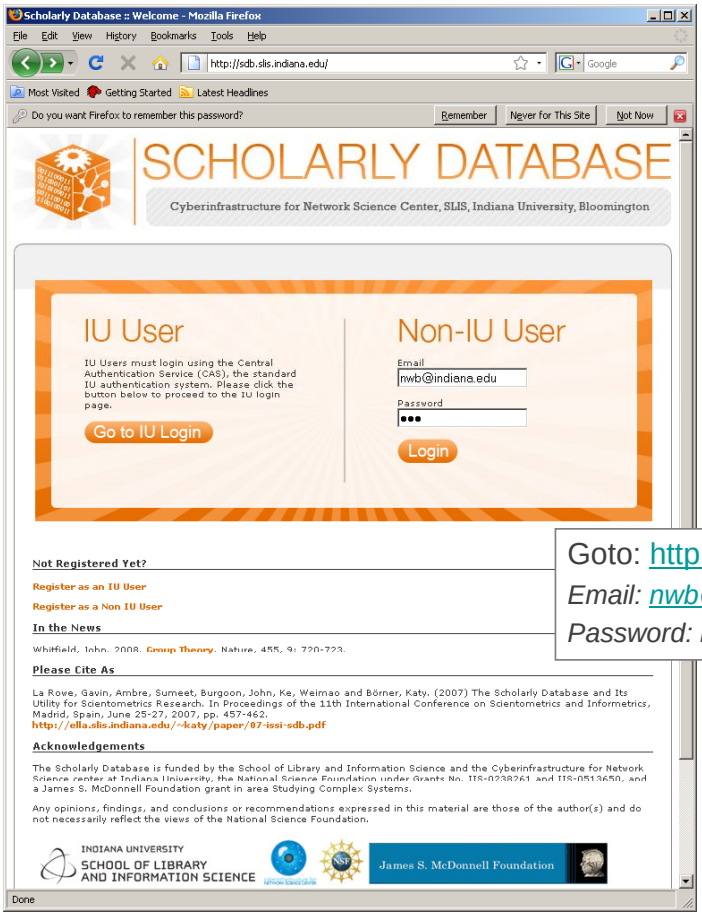
The result is a directed network of paper citations in the Data Manager. It has 5,335 nodes and 9,595 edges.

To view the complete network, select the network and run *'Visualization > GUESS'*. Run '*Script > Run Script …' and select 'yournwbdirectory*/script/GUESS/paper-citation-nw.py'.*

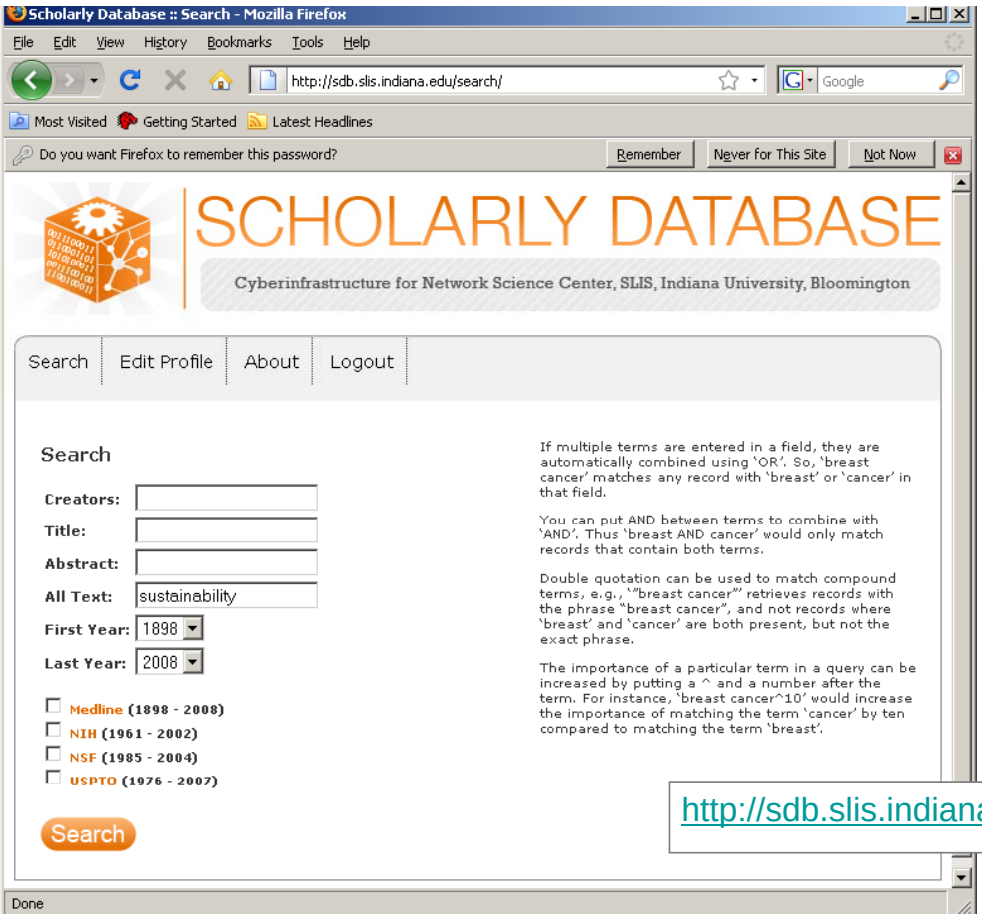


NWB Tool for Scientometrics Research

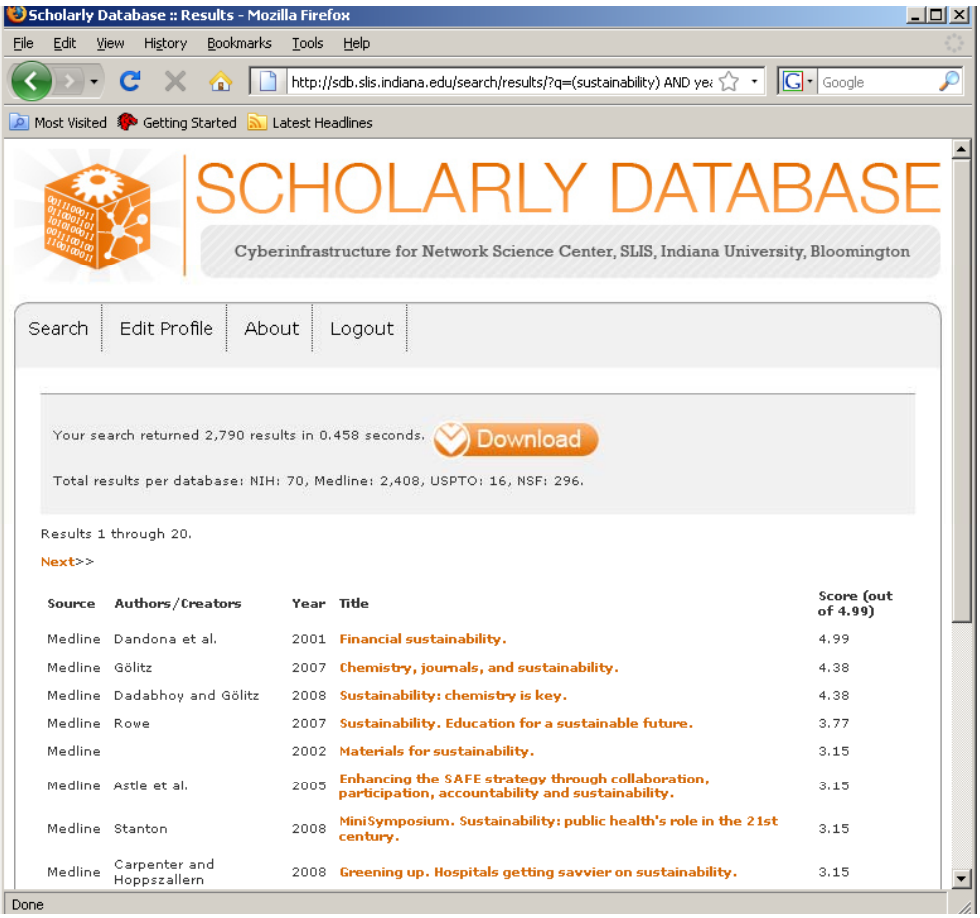
1. Loading NSF datasets with currently active NSF funding for Indiana U, Cornell U, and Michigan U, extracting, and comparing Co-PI networks.
2. Loading ISI files of major network science researchers, extracting, analyzing and visualizing paper-citation networks and co-author networks.
3. **Loading text files and detecting Bursts.**

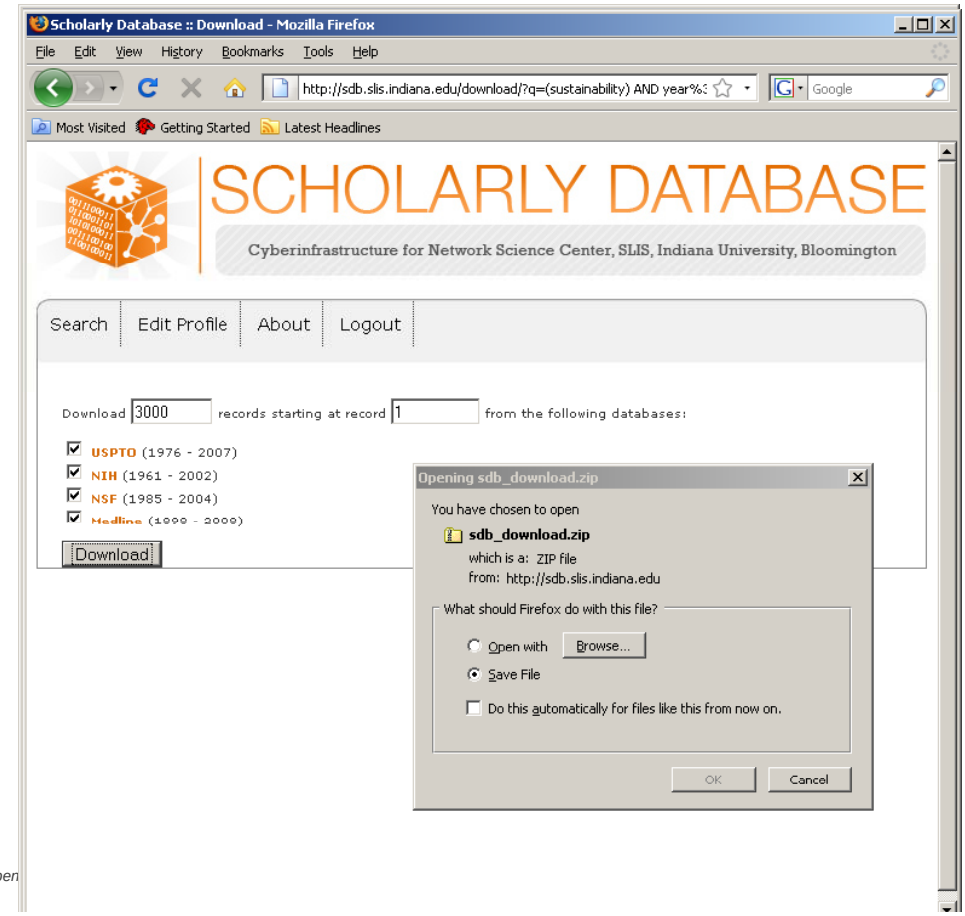


Goto: http://sdb.slis.indiana.edu
Email: nwb@indiana.edu
Password: nwb



http://sdb.slis.indiana.edu



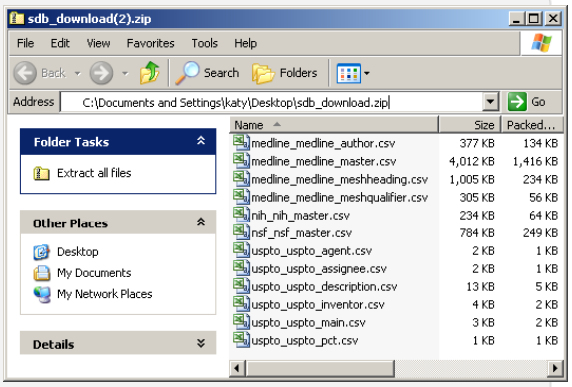


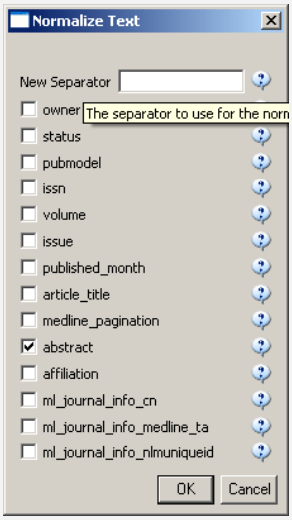
NetworkWorkbench
A Workbench for Network Scientists

Open and Preprocess SDB zip file

Load medline_medline_master.csv to NWB.

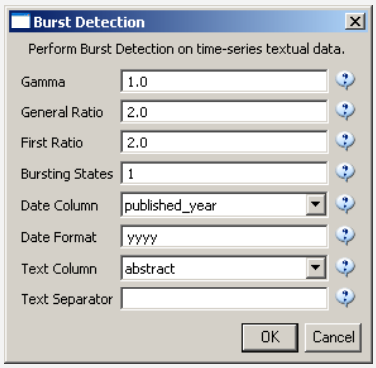
Run *'Preprocessing > Normalize Text'* with a space as New Separator.





Burst Analysis

Run ‘Analysis > Textual > Burst Detection’ with parameters:
and space as a separator.



Sort result by burst weight

||

V

Word	Length	Weight	Strength	Start	End
care	1	Infinity	Infinity	1988	1988
water	1	29.8883	29.8883	2002	2002
countri	10	27.03612	27.03612	1990	1999
protect	1	26.88557	26.88557	2002	2002
farm	1	23.32114	23.32114	2005	2005
villag	2	23.273	40.65081	2008	
crop	2	22.33649	30.42535	2008	
educ	2	22.14556	26.98588	1995	1996
blood	5	22.12166	22.12166	1996	2000

5. Discussion of Future Work

- Improving GUESS usability.
- Creating wizard for integrating compiled algorithms.
- Algorithms can communicate with other algorithms at run-time 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
- GUESS Manual http://guess.wikispot.org/manual

Software

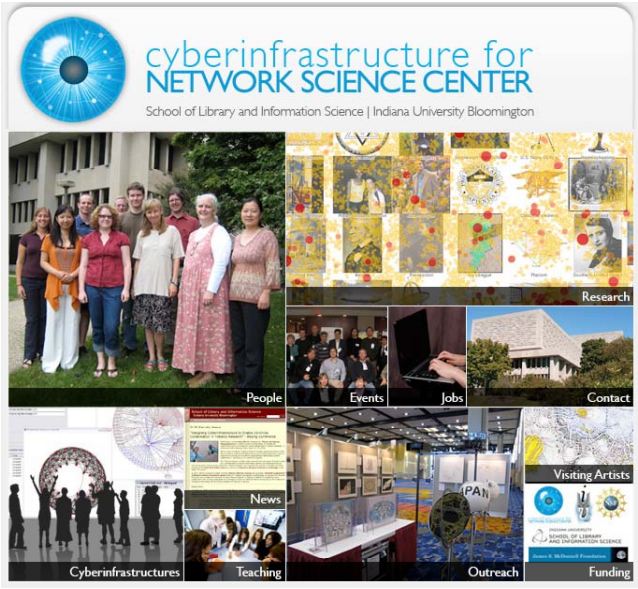
- http://cishell.org
- http://nwb.slis.indiana.edu/download.html

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