Sample NIH Workflow

A tool for science of science research & practice

Network Visualizations Using SPIRES Data and the Sci² Tool

Version: 0.0.2 beta

NIH Office of Extramural Research

and

Dr. Katy Börner School of Library and Information Science







Sci² Team. (2009). Science of Science (Sci²) Tool. Indiana University and SciTech Strategies, <u>http://sci.slis.indiana.edu</u>.

Network Visualizations Using SPIRES Data

The NIH Scientific Publication Information Retrieval and Evaluation System (SPIRES) provides NIH users with access to bibliographic records of articles in PubMed that have acknowledged support from NIH-funded projects. Each downloadable SPIRES record includes author names, MeSH terms by which each article is indexed, and a list of the project numbers acknowledged in the article. These data can be analyzed using the Science of Science (Sci²) tool to describe networks among co-authors of publications resulting from a set of NIH-funded projects, co-occurrences of the MeSH terms in those articles (a form of topic analysis), and clusters of research projects based on "co-acknowledgement," the acknowledgement of two or more NIH-funded projects in a single article.

The Sci² tool can be downloaded after registering at

http://sci.slis.indiana.edu/registration/user/. A <u>Science of Science (Sci2) Tool User Manual</u> also is available. The major steps in performing such analyses include:

- 1. Logging into SPIRES
- 2. Identifying a set of NIH-funded projects for analysis (e.g., limiting projects by topic, program class code, RFA/PA, institute, investigator, institution)
- 3. Retrieving the publications that have acknowledged support from those projects
- 4. Downloading these data from SPIRES as a .csv file
- 5. Running the Sci² tool and loading the .csv files into it for analysis
- 6. Extracting a co-occurrence network (for author names, MeSH terms, or acknowledged projects)
- 7. Analyzing and displaying the graph

1. Logging into SPIRES

The URL for SPIRES is <u>http://spires.era.nih.gov/spires/login.cfm</u>. Users can login using their IMPAC II credentials.

2. Identifying a set of NIH-funded projects for analysis

From the home page, select 'Search Projects' to identify a set of projects whose publications are to be analyzed.





Login

After entering criteria to define the set of projects, select 'Search Projects.'

Search Projects Defau 2 **Basic Search Options**

Add to Cart

When the resulting hit list appears, select the projects of interest and add them to a cart.



Basic Search Options

3. Retrieving the publications that have acknowledged support from those projects

From the top navigation tabs, select 'Search Publications.'		Home	Scientif	PIRES	mation Retrieval and Eva	luation System	Utilities	Logout
Select 'Custom Download'	output.	Basic Sea Output Op Saved Qu	otion:	Standard List Standard List Custom Down	_			
Use the project cart	Shopping Carts							
created in step 2.	Saved Project Carts: Include pr	ojects in any o	of these	carts:	My Project Cart			¥
After entering criteria to de (e.g., year of publication), s			o be	retrieved	Search	Publicatio	ons C)efault Cri

4. Download these data from SPIRES as a .csv file

Select the data elements to download from the SPIRES database. Select Authors, MeSH Terms, and Project Numbers for co-authorship, topic analysis, and co-acknowledgement networks, respectively.

Download Data for:	1-183 💌		Saved List Name:	
Pick Items Last Revision Date Mesh Date Journal Title Abbr Journal Issue Journal Volume ISSN Page Number Publish Date Publish Date Publish Year Impact Factor Publish Status		Add Item(s) > Add All >> < Remove Item(s) << Remove All CTRL Click or SHIFT Click to select multiple items	Download It Publication Title Authors Mesh Terms Project Numbers Publish Year	Up Down
And download these data in ar	h Excel file.	Download Excel File	Reset Retrieve List	Save List Save/Manage Query
Save the Excel file in csv		y Project Publications.csv SV (Comma delimited) (*.csv)		 ▼
format:				Save Cancel
				<u>S</u> ave Cancel

Note that you may have to download large SPIRES results sets in several separate, smaller files and combine them in Excel after all are downloaded.

A	B	C	D	E	F	G
1 Rec	PubMed ID	Publication Title	Authors	Mesh Terms	Project Numbers	Publish Year
2	1. 550025	Cell cycle: dependent transcriptional and protechtic regulation of Fis2 in Caulobacter.	Kelly, A.Y. Sackett, M.J.; Din, N; Quandokus, E; Brun, Y V	Bacterial Proteins/genetics", Bacterial Proteins/pivology*, Back Sequence; Binding Sites/genetics", Caulobacter crescentus/themistry: Caulobacter crescentus/genetics", Caulobacter crescentus/genetics", Caulobacter Createristics: Cell Cycle/physiology*, Cell Division/physiology: Cycleskeletal Proteins*, DNA-Binding Proteins*, GTP- Binding Proteins/genetics; DT-Binding Proteins/physiology: Gene Expression Regulation, Bacterial: hydrofysis; Molacular Sequence Data; Mutation/physiology: Promoter Regions (Genetics)/genetics; Protein Srucessing, Oost*Transistional; Research Support, US. Gov't, P.AS: Sequence Homology, Nucleic Acid; Time Factory; Transcription, Genetic/genetics; Prosenter Support, US.		1998

5. Starting Sci² and uploading the .csv file into the Sci² tool for analysis

After extracting the downloaded Sci² application, click the icon to run the program.





And browse to find the csv file created in step 4.



The file you have selected can be loaded using one

or more of the following formats. Please select the format you would like to try.

Select 'Standard csv format.' The Scheduler will indicate when the load has completed and the Data Manager will display the new data set.

							Sco	ous csv format
Remov		rom List Remove com	oleted a	utomatio	cally Remove a	all completed		Select Cancel Details >>
	1	Algorithm Name	Date		Time	% Complete	e	
	\checkmark	Load	05/04/	2010	11:41:12 AM			
				1010 D	ata Manage)	r		- 8
					CSV file:	H:\Portfolio	Ana	ysis\NWB\My Project Publications.csv

6. Extracting a co-occurrence network

From the Sci² top navigation bar, select 'Data Preparation > Text Files > Extract Co-Occurrence Network.'

File Data Preparation	Preprocessing Analysis Modeling Visualization Help					
📃 🖉 🛛 Database 🕨						
Welc Text Files > The development of t	Remove ISI Duplicate Records Remove Rows with Multitudinous Fields					
under Grant No. SBE- Primary investigators Joseph Biberstine. It u plugins were derived	Extract Directed Network Extract Bipartite Network Extract Paper Citation Network Extract Author Paper Network					
Please cite as follows Sci² Team. (2009). S	Extract Co-Occurrence Network					
Load was selected.	Extract Word Co-Occurrence Network					
Documentation: http	Extract Reference Co-Occurrence (Bibliographic Coupling) Network					

To create a co-author network diagram, select the column 'Authors' and the text delimiter ";" (semicolon), the delimiter that appears between author names in the SPIRES download.

(Ignore the Aggregation Function File parameter.)

Extract Network from Table								
Extracts a network from a delimited table								
Column Name	Authors	v 😲						
Text Delimiter	;	•						
Aggregation Function File	C:/Program Files/Sci2/scipolicy	Browse 🔇						
		OK Cancel						

The Scheduler will indicate when the extraction has completed and the Data Manager will display the new

Image: Image:

Merge Table: based on Authors

1999 Data Manager

heduler will indicate when the	🖳 Scheo	lule	er				
tion has completed and the	Remove	e Fr	rom List Remove com	pleted automatio	cally Remove a	all completed	
Aanager will display the new							
		1	Algorithm Name	Date	Time	% Complete	
ita Manager		✓	Extract Co-Occurrenc	05/04/2010	11:44:01 AM		
\NWB\My Project Publications.csv	n	_	Load work.	05/04/2010	11:41:12 AM		
S Extracted Network on Column Authors Network.							

Calculate the number of co-authors or degree for each node using 'Analysis > Networks > Unweighted and Undirected > Node Degree.'

g Analysis Modeling Visualization Help	-
Temporal > Geospatial > rte Topical > Network Analysis Toolkit (NAT)	
, Indiana University e frastructure Shell (h rk Workbench Tool h unweighted & Undirected Unweighted & Directed Weighted & Directed	Node Degree Degree Distribution K-Nearest Neighbor (Java) Watts-Strogatz Clustering Coefficient

Open the listing of authors in the Merge Table by right clicking in the Data Manager, selecting 'View With...,' and selecting Excel from the list of applications available. In Excel, the list of uniquely-identified authors can be sorted and examined to identify names that may pose problems for the analysis, such as very common names or the appearance of several name variants for the same individual.



7. Displaying the graph



Select 'GEM' from the GUESS layout menu and other options GUESS to produce a clearer picture. Use Bin Pack to "pack" the nodes more closely in space.





Zoom-in to see portions in more detail, add node labels, size- or color-code nodes by their properties (e.g., the number of each individual's co-authors, also called node degree), and export the graph for use in presentations or reports.



Other Type of Networks

Similar steps, combined with other options, can be used to perform a variety of different analyses, such as co-occurrences of MeSH terms (i.e., topic analysis) and co-acknowledgment networks.

Topic Analysis

Select the column of MeSH terms and their ";" delimiter.

Extract Netwo	rk from Table	
Extract	s a network from a delimited table	
Column Name	Mesh Terms	v 🔍
Text Delimiter	;	
Aggregation Eurotion File	C:/Program Files/Sci2/scipolicy	Browse
Aggregation Function File	C./Frogram Files/Sci2/scipolicy	BIOWSE V
		OK Canc

File Data Preparation Preprocessing Analysis Modeling Visualization Help Extract a co-occurrence network, visualize Database 1 E c Text Files 🔸 Remove ISI Duplicate Records Weld using GUESS Remove Rows with Multitudinous Fields The development of t Inder Grant No. SBE-Extract Directed Network Extract Bipartite Network Primary investigators Visualization Help Joseph Biberstine. It i Extract Paper Citation Network plugins were derived Extract Author Paper Network General ۲ Please cite as follow Extract Co-Occurrence Netv Temporal Sci² Team. (2009). S veloped at the Cyberinfrastru Extract Word Co-Occurrence Network Geospatial ۲ Extract Co-Author Network 1a.edu. Load... was selected. Extract Reference Co-Occurrence (Bibliographic Coupling) Network ocumentation: http Networks GUESS Þ

and select the Multi-Dimensional Scaling layout. In GUESS, MDS does a mutli-dimensional scaling on the graph where node-node distances are defined by the connecting edge weight (in this case, co-occurrence frequencies).



Co-Acknowledgement Networks

A co-acknowledgement network will reveal sets of research projects which tend to be acknowledged together in research publications.

Select the column Project Numbers, which are delimited in the csv file by commas.

Extract Network from Table								
Extracts a network from a delimited table								
Column Name	Project Numbers	v 😲						
Text Delimiter		•						
Aggregation Function File	C:/Program Files/Sci2/scipolicy	Browse 😲						
		OK Cancel						

Use the GEM layout in GUESS and zoom in to examine find projects that are more frequently acknowledged together in publications. Color code by funding institution, award types, etc.





References

Scott Weingart, Hanning Guo, Katy Borner, Kevin W. Boyack, Micah W. Linnemeier, Russell J. Duhon, Patrick A. Phillips, Chintan Tank, and Joseph Biberstine (2010) <u>Science of Science (Sci2) Tool User Manual.</u> Cyberinfrastructure for Network Science Center, School of Library and Information Science, Indiana University, Bloomington.

Katy Borner (2009) <u>The Science of Science Tool and Its Utility for Research</u>. Cyberinfrastructure for Network Science Center, School of Library and Information Science, Indiana University, Bloomington.