

Updated June 8, 2010



Sci² Tool

A tool for science of science research & practice

Sample NIH Workflow

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



SPIRES

Scientific Publication Information Retrieval and Evaluation System



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

Network Visualizations Using SPIRES Data

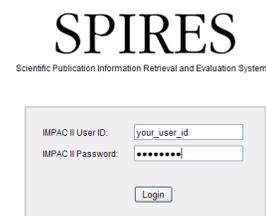
The NIH Scientific Publication Information Retrieval and Evaluation System (SPIRES) provides NIH users with access to bibliographic records of articles in PubMed that have acknowledged support from NIH-funded projects. Each downloadable SPIRES record includes author names, MeSH terms by which each article 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 *Science of Science (Sci²) Tool User Manual* 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 <http://spires.era.nih.gov/spires/login.cfm>. Users can login using their IMPAC II credentials.

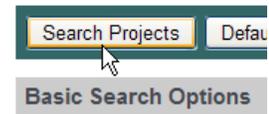


2. Identifying a set of NIH-funded projects for analysis

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



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



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

A screenshot of a form titled "Add selected projects to a cart". The form contains the following elements: a text input field for "New cart name" with the value "My Project Cart"; a text input field for "Cart comment"; a radio button selected for "Share new cart with other users from same institute?" with "Yes" selected; a dropdown menu for "Existing cart name" with the value "--Select an existing project cart--"; and two buttons at the bottom: "Add now" and "Cancel".

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

From the top navigation tabs, select 'Search Publications.'



Select 'Custom Download' output.

A screenshot of the "Basic Search Options" form. It includes an "Output Option:" dropdown menu with "Standard List" selected and a "Saved Query:" dropdown menu with "Custom Download" selected.

Use the project cart created in step 2.

A screenshot of the "Shopping Carts" section. It shows a label "Saved Project Carts:" followed by the text "Include projects in any of these carts:" and a dropdown menu with "My Project Cart" selected.

After entering criteria to define the set of publications to be retrieved (e.g., year of publication), select 'Search Publications.'



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.

And download these data in an Excel file.

Save the Excel file in csv format:

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.

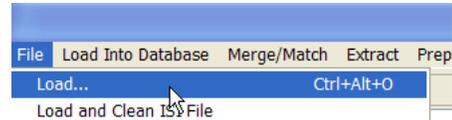
A	B	C	D	E	F	G	
1	Rec	PubMed ID	Publication Title	Authors	Mesh Terms	Project Numbers	Publish Year
1	1	9512521	Cell cycle-dependent transcriptional and proteolytic regulation of Pts2 in Caulobacter.	Kelly, A J; Sackett, M J; Din, N; Guardokus, E; Brun, Y Y	Bacterial Proteins/genetics; Bacterial Proteins/physiology; Base Sequence; Binding Sites/genetics; Caulobacter crescentus/chemistry; Caulobacter crescentus/genetics; Caulobacter crescentus/physiology; Cell Cycle/genetics; Cell Cycle/physiology; Cell Division/genetics; Cell Division/physiology; Cytoskeletal Proteins; DNA-Binding Proteins; GTP-Binding Proteins/genetics; GTP-Binding Proteins/physiology; Gene Expression Regulation, Bacterial; Hydrolysis; Molecular Sequence Data; Mutation/genetics; Mutation/physiology; Promoter Regions (Genetics)/genetics; Protein Processing, Post-Translational; Research Support, Non-U.S. Gov't; Research Support, U.S. Gov't, P.H.S.; Sequence Homology, Nucleic Acid; Time Factors; Transcription Factors; Transcription, Genetic/genetics; Transcription, Genetic/physiology	R01GM051966,T32GM007757,F32GM007757	1998
2							

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

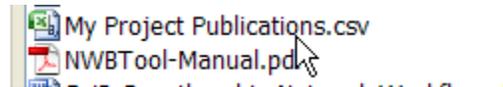
After extracting the downloaded Sci² application, click the icon to run the program.



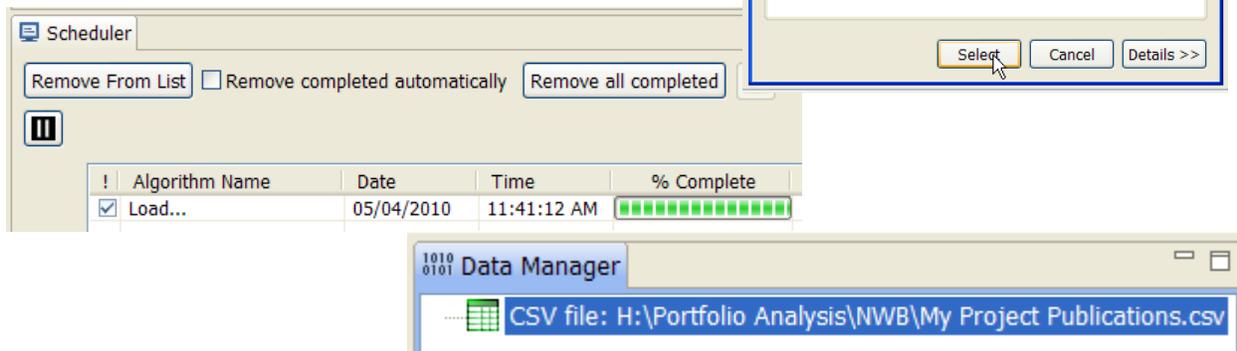
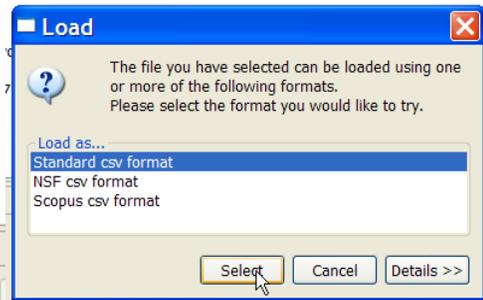
Select 'Load...' from the File menu



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

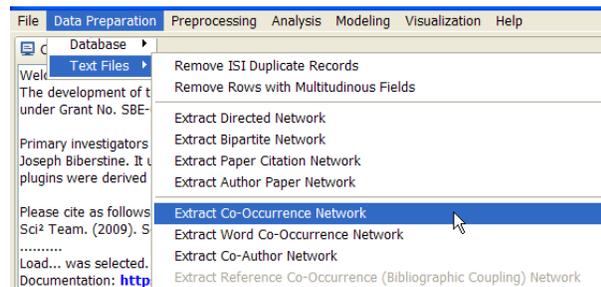


Select 'Standard csv format.' The Scheduler will indicate when the load has completed and the Data Manager will display the new data set.



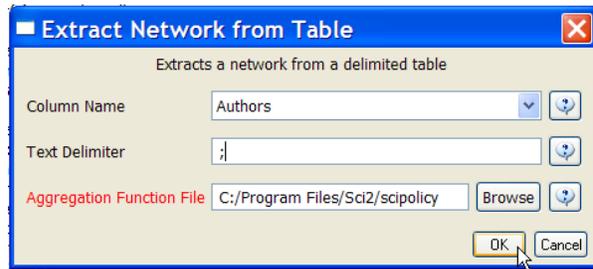
6. Extracting a co-occurrence network

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

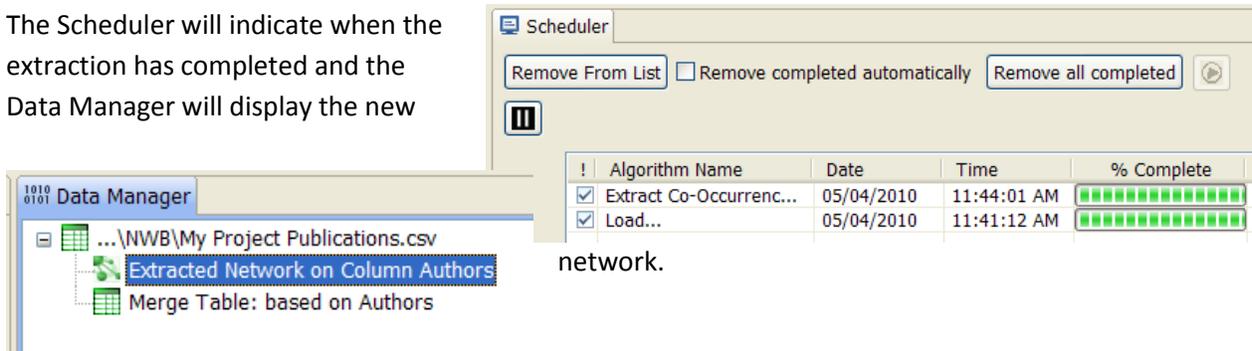


To create a co-author network diagram, select the column 'Authors' and the text delimiter ";" (semicolon), the delimiter that appears between author names in the SPIRES download.

(Ignore the Aggregation Function File parameter.)

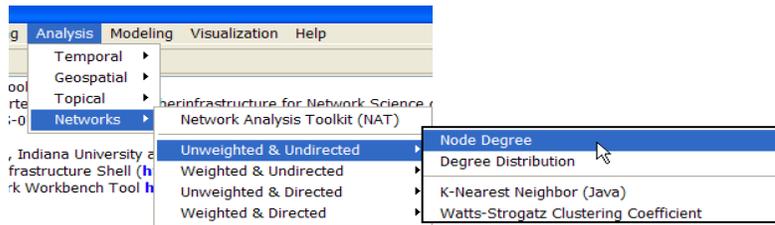


The Scheduler will indicate when the extraction has completed and the Data Manager will display the new

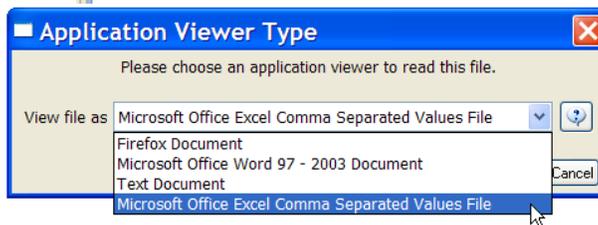
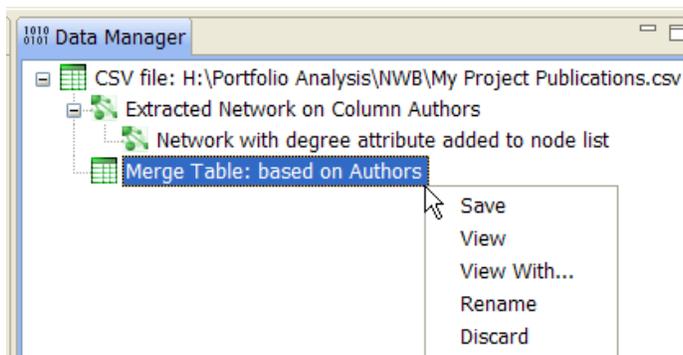


network.

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

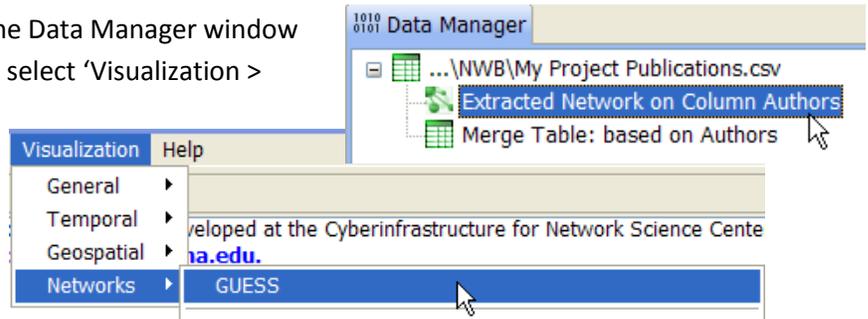


Open the listing of authors in the Merge Table by right clicking in the Data Manager, selecting 'View With...', and selecting Excel from the list of applications available. In Excel, the list of uniquely-identified authors can be sorted and examined to identify names that may pose problems for the analysis,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

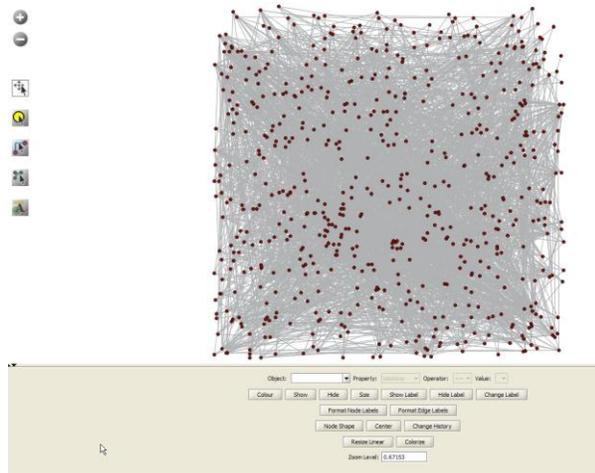


7. Displaying the graph

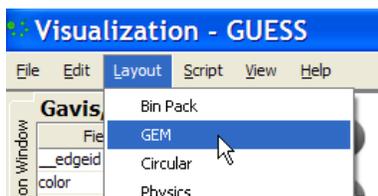
Highlight the Extracted Network in the Data Manager window and from the Sci² top navigation bar, select 'Visualization > Networks > GUESS' (the Graph Exploration System).



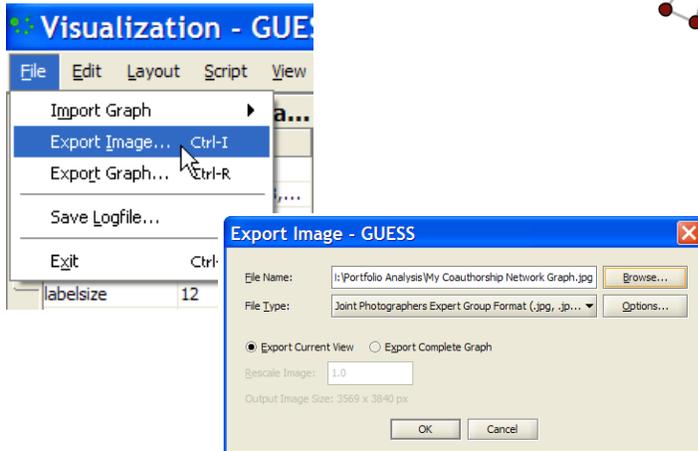
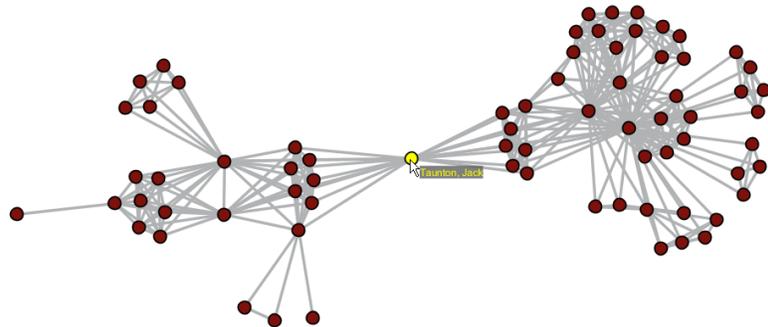
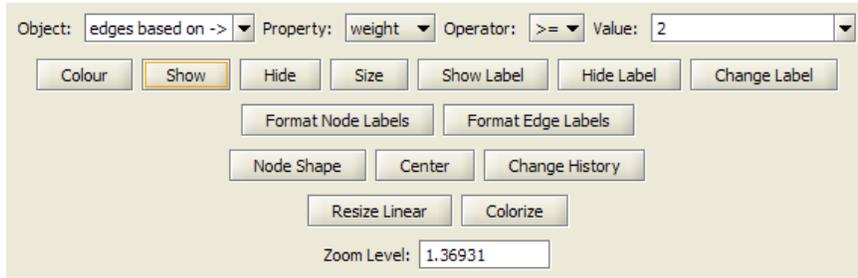
The initial network in GUESS might look something like this:



Select 'GEM' from the GUESS layout menu and use Bin Pack to "pack" the nodes more closely in space.



Zoom-in using the  button in the network display to see portions in more detail and use the GUESS 'Graph Modifier' panel to 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.

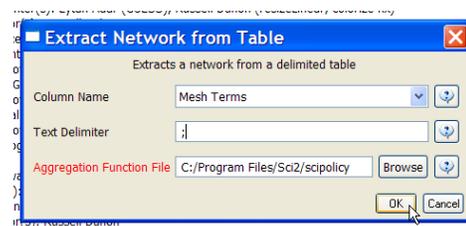


Other Type of Networks

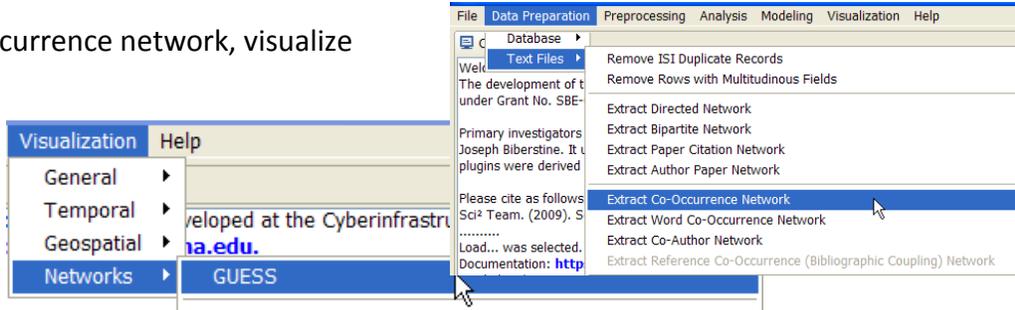
Similar steps, combined with other options, can be used to perform a variety of different analyses, such as co-occurrences of MeSH terms (i.e., topic analysis) and co-acknowledgment networks.

Topic Analysis

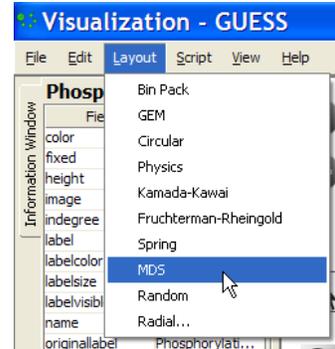
Select the column of MeSH terms and their “;” delimiter.



Extract a co-occurrence network, visualize using GUESS



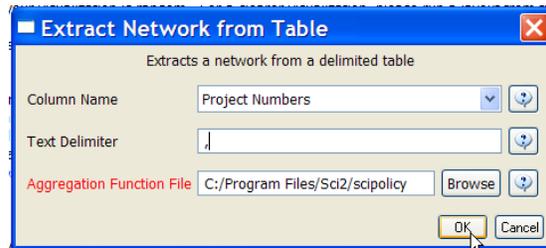
and select the Multi-Dimensional Scaling layout. In GUESS, MDS does a multi-dimensional scaling on the graph where node-node distances are defined by the connecting edge weight (in this case, co-occurrence frequencies).



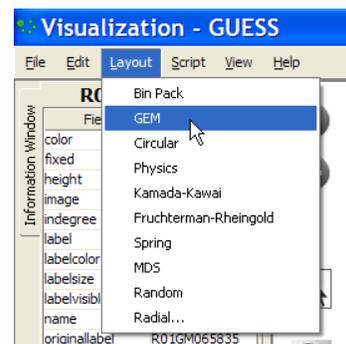
Co-Acknowledgement Networks

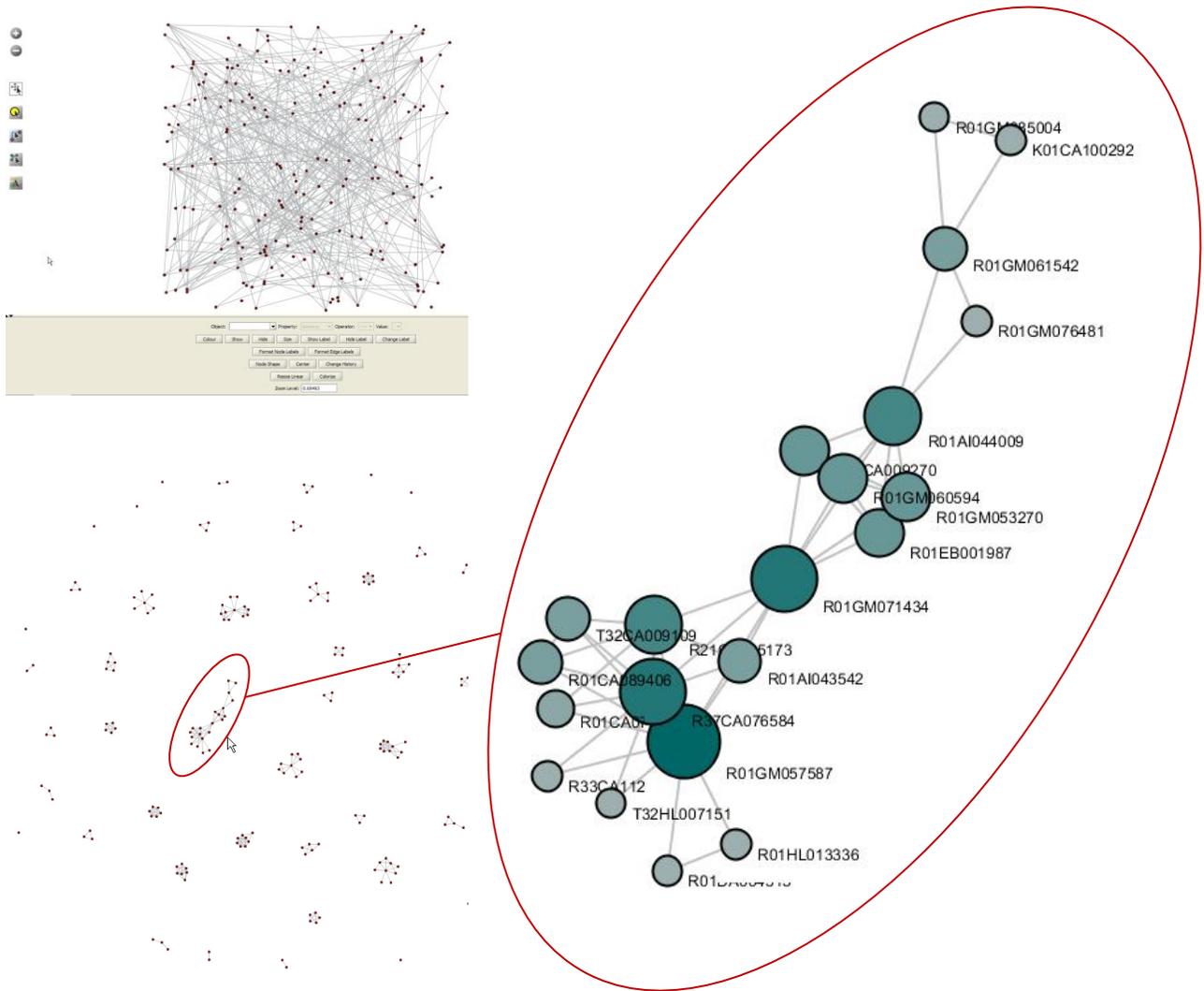
A co-acknowledgement network will reveal sets of research projects which tend to be acknowledged together in research publications.

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



Use the GEM layout in GUESS and zoom in to 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) [Science of Science \(Sci2\) Tool User Manual](#). Cyberinfrastructure for Network Science Center, School of Library and Information Science, Indiana University, Bloomington.

Katy Borner (2009) [The Science of Science Tool and Its Utility for Research](#). Cyberinfrastructure for Network Science Center, School of Library and Information Science, Indiana University, Bloomington.