Science of Science Research and Tools
Tutorial #08 of 12

Dr. Katy Börner
Cyberinfrastructure for Network Science Center, Director
Information Visualization Laboratory, Director
School of Library and Information Science
Indiana University, Bloomington, IN
http://info.slis.indiana.edu/~katy

With special thanks to Kevin W. Boyack, Micah Linnemeier, Russell J. Duhon, Patrick Phillips, Joseph Biberstine, Chintan Tank Nianli Ma, Hanning Guo, Mark A. Price, Angela M. Zoss, and Scott Weingart

Invited by Robin M. Wagner, Ph.D., M.S.
Chief Reporting Branch, Division of Information Services
Office of Research Information Systems, Office of Extramural Research
Office of the Director, National Institutes of Health

Suite 4090, 6705 Rockledge Drive, Bethesda, MD 20892
10a-noon, July 20, 2010

12 Tutorials in 12 Days at NIH—Overview

1st Week
1. Science of Science Research
2. Information Visualization
3. CIShell Powered Tools: Network Workbench and Science of Science Tool

2nd Week
4. Temporal Analysis—Burst Detection
5. Geospatial Analysis and Mapping
6. Topical Analysis & Mapping

3rd Week
7. Tree Analysis and Visualization
8. Network Analysis
9. Large Network Analysis

4th Week
10. Using the Scholarly Database at IU
11. VIVO National Researcher Networking
12. Future Developments

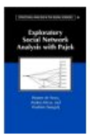
12 Tutorials in 12 Days at NIH—Overview

[#08] Network Analysis and Visualization

- General Overview
- Designing Effective Network Visualizations
- Notions and Notations
- Sci2-Reading and Extracting Networks
- Sci2-Analysing Networks
- Sci2-Visualizing Networks
- Outlook
- Exercise: Identify Promising Network Analyses of NIH Data

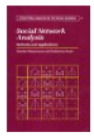
Recommended Reading

- NWB Team (2009) Network Workbench Tool, User Manual 1.0.0, http://nwb.slis.indiana.edu/Docs/NWBTool-Manual.pdf

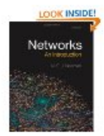


Exploratory Social Network Analysis with Pajek by de Nooy, Wouter
(9)
$35.19

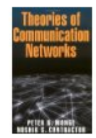
Models and Methods in Social Network Analysis by Peter J. Carrington
(1)
$18.14



Social Network Analysis: Methods and Applications by Katherine Faust
(9)
$31.20



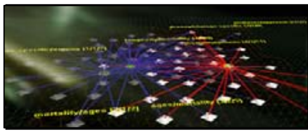
Networks: An Introduction by Mark Newman
$68.90



Theories of Communication Networks by Peter R. Monge
(7)
$19.25

[#08] Network Analysis and Visualization

- General Overview
- Designing Effective Network Visualizations
- Notions and Notations
- Sci2-Reading and Extracting Networks
- Sci2-Analysing Networks
- Sci2-Visualizing Networks
- Outlook
- Exercise: Identify Promising Network Analyses of NIH Data

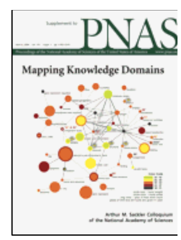


Sample Networks

- Communication networks
  - Internet, telephone network, wireless network.
- Network applications
  - The World Wide Web, Email interactions
- Transportation network/ Road maps
- Relationships between objects in a data base
  - Function/module dependency graphs
  - Knowledge bases

Network Properties

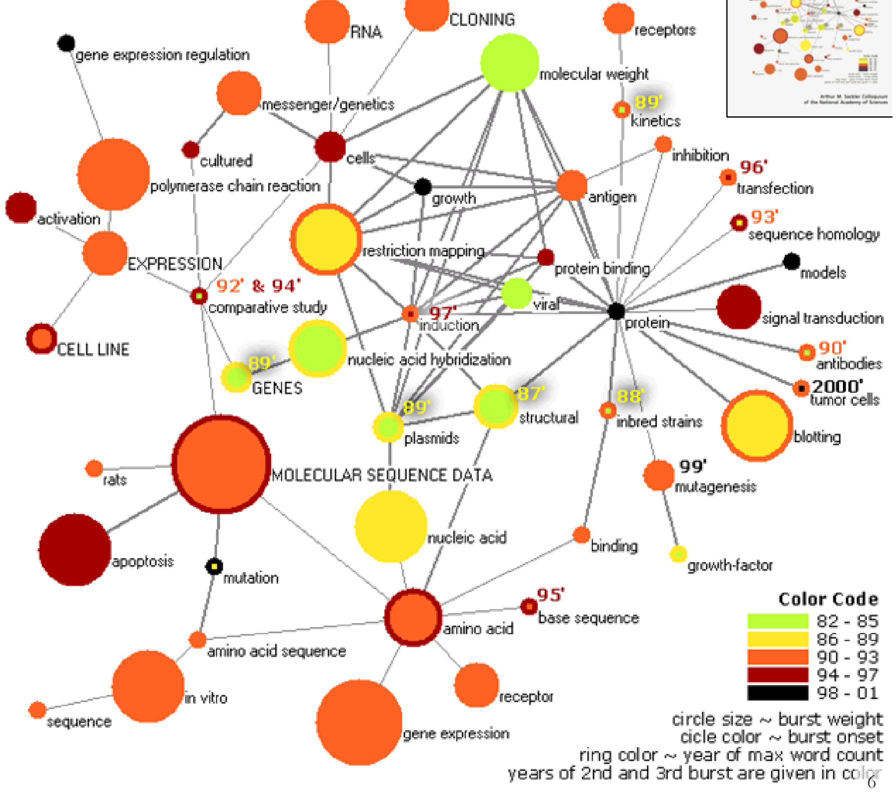
- Directed vs. undirected
- Weighted vs. unweighted
- Additional node and edge attributes
- One vs. multiple node & edge types
- Network type (random, small world, scale free, hierarchical networks)



Reducing the number of edges via pathfinder network scaling.

Co-word space of the top 50 highly frequent and bursty words used in the top 10% most highly cited PNAS publications in 1982-2001.

(Mane & Börner, 2004)



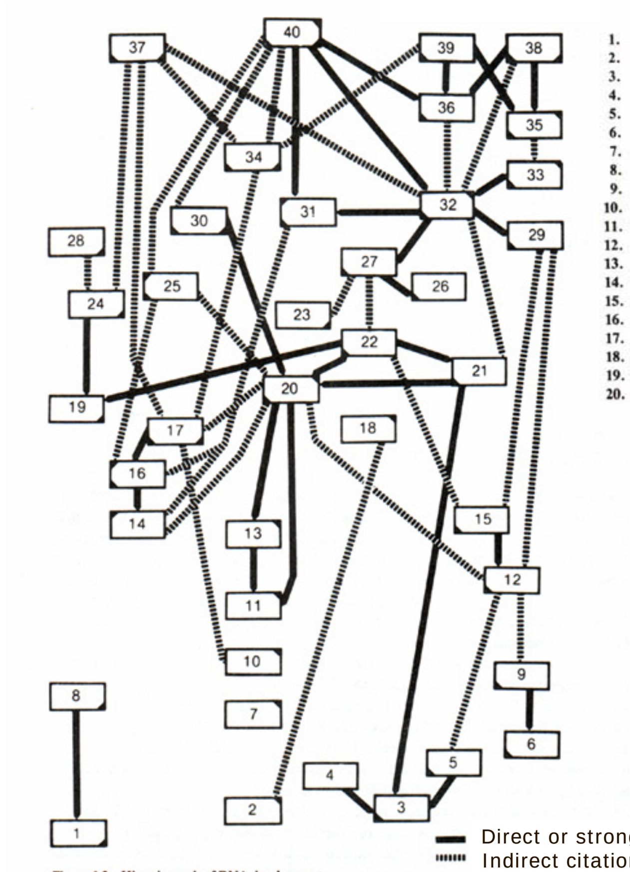


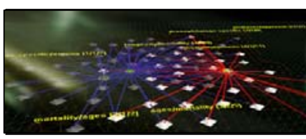
Figure 6.3 Historiograph of DNA development.

Direct or strongly implied citation
Indirect citation

KEY

1. Braconnot 1820
2. Mendel 1865
3. Miescher 1871
4. Flemming 1879
5. Kossel 1886
6. Fischer and Piloty 1891
7. DeVries 1900
8. Fischer 1907
9. Levene and Jacobs 1909
10. Muller 1926
11. Griffith 1928
12. Levene with Mori and London 1929
13. Alloway 1932
14. Stanley 1935
15. Levene and Tipson 1935
16. Bawden and Pirie 1936-1937
17. Caspersson and Schultz 1938-1939
18. Beadle and Tatum 1941
19. Martin and Synge 1943-1944
20. Avery, MacLeod, and McCarty 1944
21. Chargaff 1947
22. Chargaff 1950
23. Pauling and Corey 1950-1951
24. Sanger 1951-1953
25. Hershey and Chase 1952
26. Wilkins 1953
27. Watson and Crick 1953
28. DuVigneaud 1953
29. Todd 1955
30. Palade 1954-1956
31. Fraenkel-Conrat 1955-1957
32. Ochoa 1955-1956
33. Kornberg 1956-1957
34. Hoagland 1957-1958
35. Jacob and Monod 1960-1961
36. Hurwitz 1960
37. Dintzis 1961
38. Novelli 1961-1962
39. Allfrey and Mirsky 1962
40. Nirenberg and Matthaei 1961-1962

Historiograph of
DNA Development
(Garfield, Sher, & Torpie, 1964)

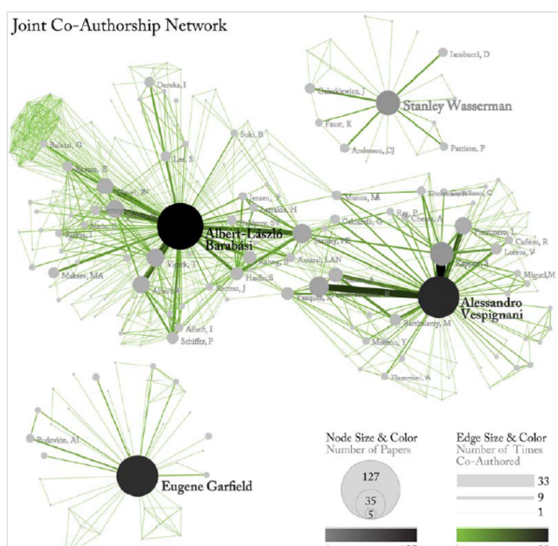


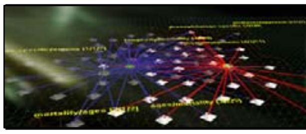
Force Directed Layout – How does it work?

The algorithm simulates a system of forces defined on an input graph and outputs a locally minimum energy configuration. Nodes resemble mass points repelling each other and the edges simulate springs with attracting forces. The algorithm tries to minimize the energy of this physical system of mass particles.

Required are
- A force model
- Technique for finding locally minimum energy configurations.

P. Eades,"A heuristic for graph drawing" Congressus Numerantium, 42,149-160,1984.



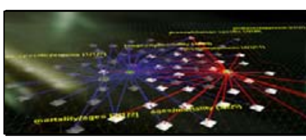


Force Directed Layout cont.

Force Models

Force Model	Formula	Example of usage
Spring Force	***F = k(1-a)*** *k- stiffness of spring* *a- natural length of spring*	Assigning different k and a to different edges to separate nodes by different distances.
Gravity Force	**F = g/r²** *g- associated with mass of node, usually equals 1.*	Apply gravity force between node pairs to prevent node overlapping.
Electrical and Magnetic Force	**F = eE** **F = qB** *E- electric field strength* *B- magnetic field strength*	Changes nodes distribution along a direction.

A simple algorithm to find the equilibrium configuration is to trace the move of each node according to Newton's 2nd law. This takes time O n^3, which makes it unsuitable for large data sets. Rob Forbes (1987) proposed two methods that were able to accelerate convergence of a FDP problem 3-4 times. One stabilizes the derivative of the repulsion force and the other uses information on node movement and instability characteristics to make a predictive extrapolation.



Force Directed Layout cont.

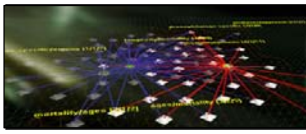
Most existing algorithms extend Eades' algorithm (1984) by providing methods for the intelligent initial placement of nodes, clustering the data to perform an initial coarse layout followed by successively more detailed placement, and grid-based systems for dividing up the dataset.

GEM (Graph EMbedder) attempts to recognize and forestall non-productive rotation and oscillation in the motion of nodes in the graph as it cools, see

Frick, A., A. Ludwig and H. Mehldau (1994). A fast adaptive layout algorithm for undirected graphs. Graph Drawing, Springer-Verlag: 388-403.

Walshaw's (2000) multilevel algorithm provides a "divide and conquer" method for laying out very large graphs by using clustering, see

Walshaw, C. (2000). A multilevel algorithm for force-directed graph drawing. 8th International Symposium Graph Drawing, Springer-Verlag: 171-182.



Force Directed Layout cont.

VxOrd (Davidson, Wylie et al. 2001) uses a density grid in place of pair-wise repulsive forces to speed up execution and achieves computation times order O(N) rather than O(N2). It also employs barrier jumping to avoid trapping of clusters in local minima.

Davidson, G. S., B. N. Wylie and K. W. Boyack (2001). "Cluster stability and the use of noise in interpretation of clustering." Proc. IEEE Information Visualization 2001: 23-30.

An extremely fast layout algorithm for visualizing large-scale networks in three-dimensional space was proposed by (Han and Ju 2003).

*Han, K. and B.-H. Ju (2003). "A fast layout algorithm for protein interaction networks." Bioinformatics **19**(15): 1882-1888.*

Today, the algorithm developed by Kamada and Kawai (Kamada and Kawai 1989) and Fruchterman and Reingold (Fruchterman and Reingold 1991) are most commonly used, partially because they are available in Pajek.

*Fruchterman, T. M. J. and E. M. Reingold (1991). "Graph Drawing by Force-Directed Placement." Software-Practice & Experience **21**(11): 1129-1164.*

*Kamada, T. and S. Kawai (1989). "An algorithm for drawing general undirected graphs." Information Processing Letters **31**(1): 7-15.*

[#08] Network Analysis and Visualization

- General Overview
- Designing Effective Network Visualizations
- Notions and Notations
- Sci2-Reading and Extracting Networks
- Sci2-Analysing Networks
- Sci2-Visualizing Networks
- Outlook
- Exercise: Identify Promising Network Analyses of NIH Data

Notions and Notations

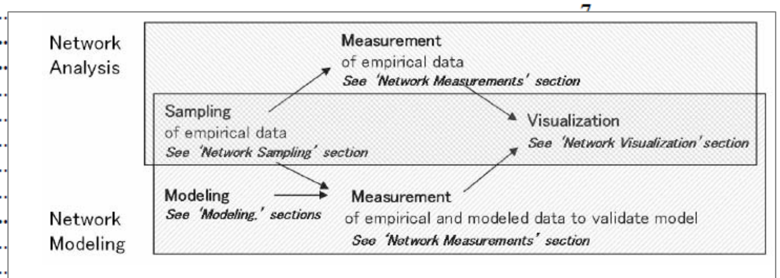


Figure 1. General network science research process.

Börner, Katy, Sanyal, Soma and Vespignani, Alessandro (2007). Network Science. In Blaise Cronin (Ed.), ARIST, Information Today, Inc./American Society for Information Science and Technology, Medford, NJ, Volume 41, Chapter 12, pp. 537-607. http://ivl.slis.indiana.edu/km/pub/2007-borner-arist.pdf

Notions and Notations

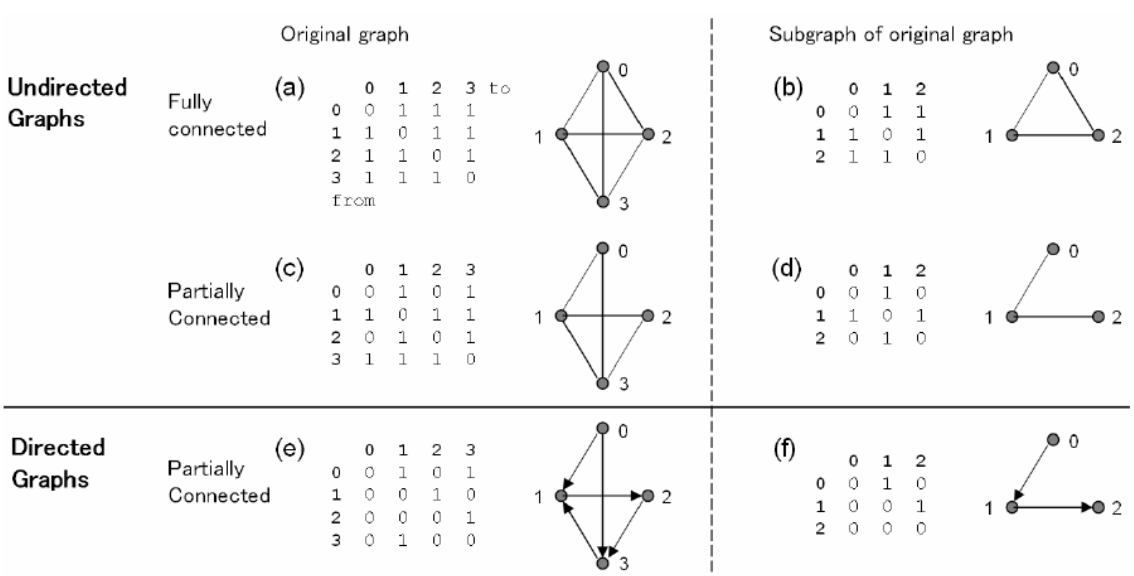


Figure 2: Adjacency matrix and graph presentations of different undirected and directed graphs.

Börner, Katy, Sanyal, Soma and Vespignani, Alessandro (2007). Network Science. In Blaise Cronin (Ed.), ARIST, Information Today, Inc./American Society for Information Science and Technology, Medford, NJ, Volume 41, Chapter 12, pp. 537-607. http://ivl.slis.indiana.edu/km/pub/2007-borner-arist.pdf

Notions and Notations

2.2.1 Node Degree

In undirected graphs, the degree k of a node is termed the number of edges connected to it. In directed graphs, the degree of a node is defined by the sum of its in-degree and its out-degree, $k_i = k_{in,i} + k_{out,i}$, where the *in-degree* $k_{in,i}$ of the node i is defined as the number of edges pointing to i; its *out-degree* $k_{out,i}$ is defined as the number of edges departing from i. In terms of the adjacency matrix, we can write

$$k_{in,i} = \sum_j A_{ji} \ , \ \ k_{out,i} = \sum_j A_{ij} \ . \qquad (1)$$

For an undirected graph, with a symmetric adjacency matrix, $k_{in,i} = k_{out,i} \equiv k_i$ holds. For example, node 1 in Figure 2a has a degree of three. Node 1 in Figure 2e has an in-degree of two and an out-degree of one.

2.2.2 Nearest Neighbors

The nearest neighbors of a node i are the nodes to which it is connected directly by an edge, so the number of nearest neighbors of the node is equal to the node degree. For example, node 1 in Figure 2a has nodes 0, 2, and 3 as nearest neighbors.

2.2.3 Path

A *path* P_{i_0,i_n} that connects the nodes i_0 and i_n in a graph $G = (V, E)$ is defined as an ordered collection of n+1 nodes $V_P = \{i_0, i_1, ..., i_n\}$ and n edges $E_P = \{(i_0, i_1), (i_1, i_2), ..., (i_{n-1}, i_n)\}$, such that $i_\alpha \in V$ and $(i_{\alpha-1}, i_\alpha) \in E$, for all α. The *length* of the path P_{i_0,i_n} is n. For example, the path in Figure 2f that interconnects nodes 0, 1, and 2 has a length of two.

Börner, Katy, Sanyal, Soma and Vespignani, Alessandro (2007). Network Science. In Blaise Cronin (Ed.), ARIST, Information Today, Inc./American Society for Information Science and Technology, Medford, NJ, Volume 41, Chapter 12, pp. 537-607. http://ivl.slis.indiana.edu/km/pub/2007-borner-arist.pdf

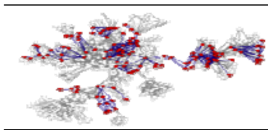
Notions and Notations

Betweenness centrality is a measure that aims to describe a node's position in a network in terms of the flow it is able to control. As an example, consider two highly connected subgraphs that share one node but no other nodes or edges. Here, the shared node controls the flow of information, for example, rumors in a social network. Any path from any node in one subgraph to any node in the other subgraph leads through the shared node. The shared node has a rather high betweenness centrality. Mathematically, the betweenness centrality is defined as the number of shortest paths between pairs of nodes that pass through a given node (Freeman, 1977). More precisely, let $L_{h,j}$ be the total number of shortest paths from h to j and $L_{h,i,j}$ be the number of those shortest paths that pass through the node i. The betweenness b of node i is then defined as $b_i = \sum L_{h,i,j} / L_{h,j}$, where the sum runs over all h,j pairs with $j \neq h$. An efficient algorithm to compute betweenness centrality was reported by Brandes (2001). The betweenness centrality is often used in transportation networks to provide an estimate of the traffic handled by different nodes, assuming that the frequency of use can be approximated by the number of shortest paths passing through a given node. It is important to stress that while the betweenness centrality is a local attribute of any given node, it is calculated by looking at all paths among all nodes in the network and therefore it is a measure of the node centrality with respect to the global topology of the network.

Börner, Katy, Sanyal, Soma and Vespignani, Alessandro (2007). Network Science. In Blaise Cronin (Ed.), ARIST, Information Today, Inc./American Society for Information Science and Technology, Medford, NJ, Volume 41, Chapter 12, pp. 537-607. http://ivl.slis.indiana.edu/km/pub/2007-borner-arist.pdf

[#08] Network Analysis and Visualization

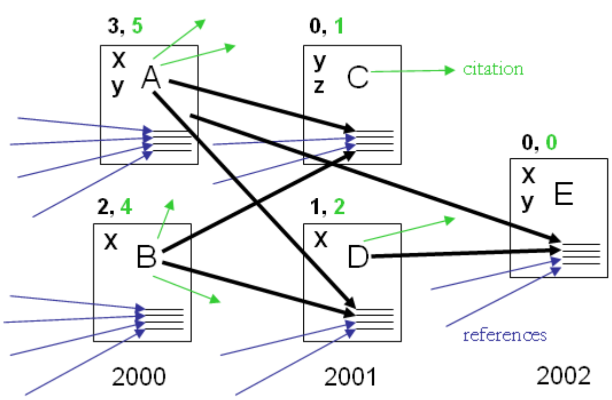
- General Overview
- Designing Effective Network Visualizations
- Notions and Notations
- Sci2-Reading and Extracting Networks
- Sci2-Analysing Networks
- Sci2-Visualizing Networks
- Outlook
- Exercise: Identify Promising Network Analyses of NIH Data



Network Extraction - Examples

Sample paper network (left) and four different network types derived from it (right).
From ISI files, about 30 different networks can be extracted.

Papers A-E written by authors x, y, z over 3 years.
Each paper happens to have 4 references.

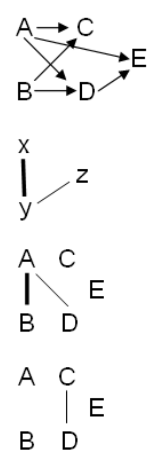


Paper-Paper Citation Network
Papers are connected via direct citation links. Arrows represent information flow from older papers to younger papers.

Author-Author (Co-Author) Network
x and y co-author papers A and E together
y and z co-author papers A and E

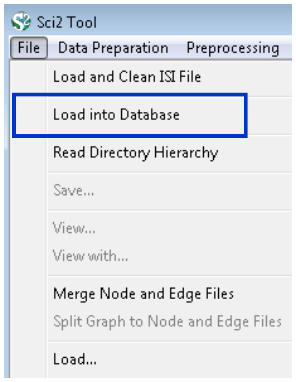
Document Co-Citation (DCA) Network
A and B are co-cited by C and D
A and D are co-cited by E

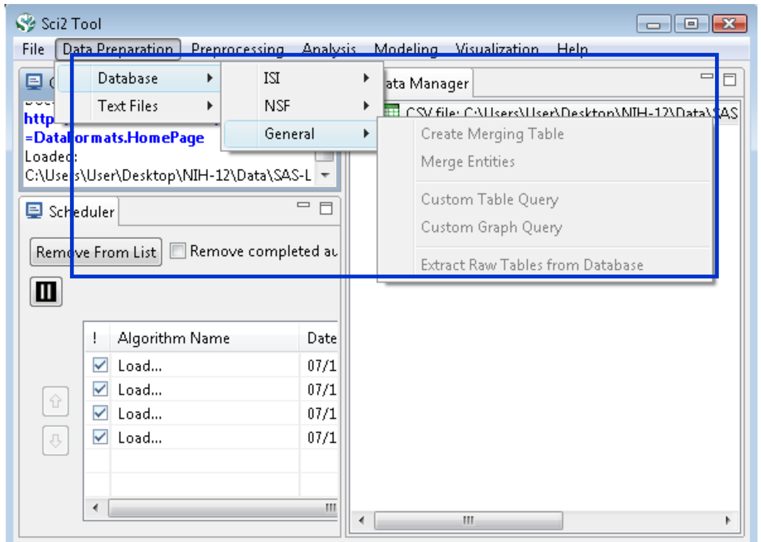
Reference Co-Occurrence (Bibliographic Coupling) Network
C and D are bibliographically coupled as they both cite/reference A and B.



Local citation counts (within this dataset) are given in **black** and global citation counts (ISI times cited) are given in green above each paper.

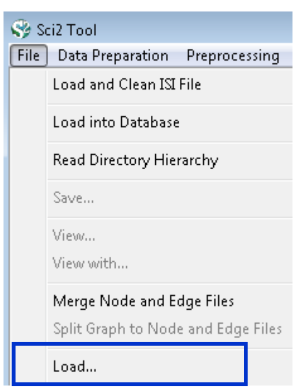
Extract Networks with Sci2 Tool – Database

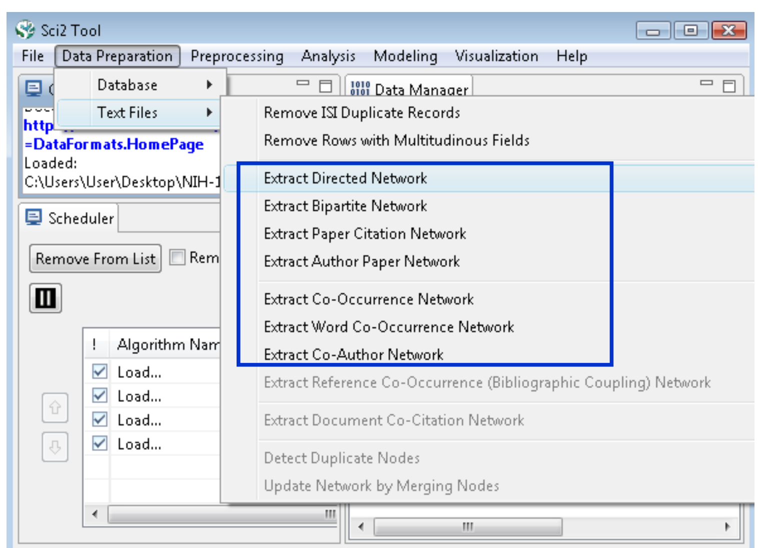




See Science of Science (Sci2) Tool User Manual, Version Alpha 3, Section 3.1 for a listing and brief explanations of all plugins. [*http://sci.slis.indiana.edu/registration/docs/Sci2_Tutorial.pdf*](http://sci.slis.indiana.edu/registration/docs/Sci2_Tutorial.pdf)
See also **Tutorial #3**

Extract Networks with Sci2 Tool – Text Files





See Science of Science (Sci2) Tool User Manual, Version Alpha 3, Section 3.1 for a listing and brief explanations of all plugins. [*http://sci.slis.indiana.edu/registration/docs/Sci2_Tutorial.pdf*](http://sci.slis.indiana.edu/registration/docs/Sci2_Tutorial.pdf)
See also **Tutorial #3**

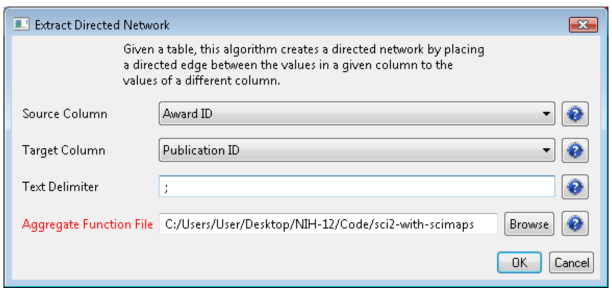
Fake NIH Dataset of Awards and Resulting Publications

Ten existing awards and a fake set of resulting publications.

Award ID	Publication ID
C06CA058690	9485464;9096302
C06CA059267	20527532;8858722;20427856;20185186;20019401;10587228
C06RR011192	16913728;16362150;19490921
C06RR012176	9714740;19490921
C06RR012488	15345738;11994348;12586855;12865481
C06RR012511	19896513;19487298;19214230
C06RR012512	18991629;17125941;18636192;16621538;18595716;17504144;17350279;17134906;19155177
C06RR012537	18207467;17318410;17961182;19490921
C06RR013551	16136041
C06RR014469	17621683

Load resulting using 'File > Load > Fake-NIH-Awards+Publications.csv' as csv file format.

Extract author bipartite grant to publications network using 'Data Preparation > Text Files > Extract Directed Network' using parameters:



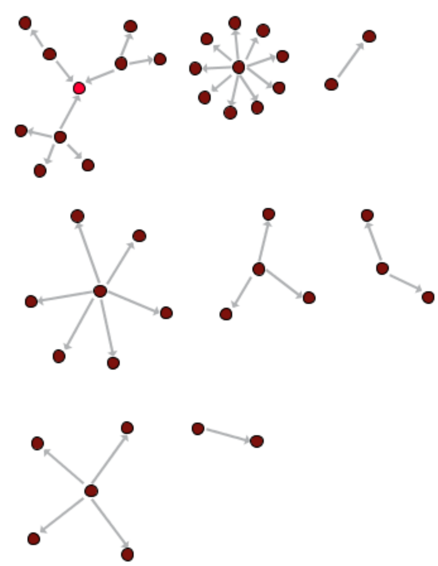
Fake NIH Dataset cont.

Network Analysis Toolkit (NAT)

This graph claims to be directed.
Nodes: 43
Isolated nodes: 0
Edges: 35
No self loops were discovered.
No parallel edges were discovered.
Did not detect any edge attributes
This network does not seem to be a valued network.

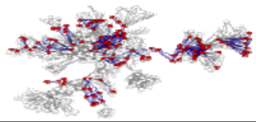
Average total degree: 1.6279
Average in degree: 0.814
Average out degree: 0.814
This graph is not weakly connected.
There are **8 weakly connected components**. (0 isolates)
The largest connected component consists of 10 nodes.

Density (disregarding weights): 0.0194

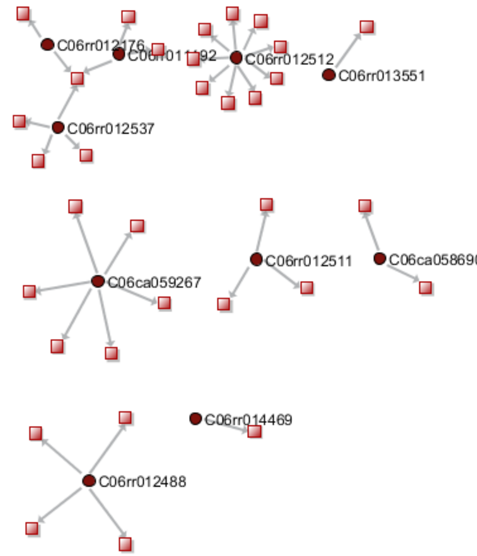


GUESS

GEM Layout, Bin pack



Fake NIH Dataset cont.

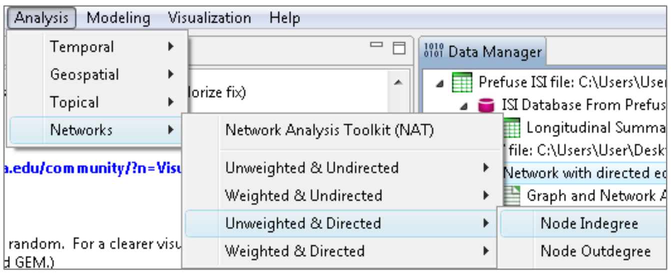


In Sci2

Node Indegree was selected.

……….

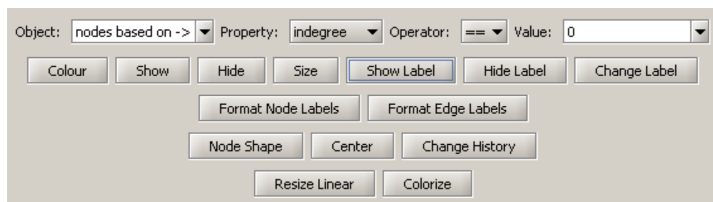
Node Outdegree was selected.

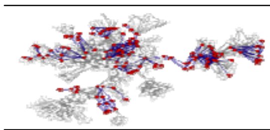


GUESS

GEM Layout, Bin pack

Color using Graph Modifier





Fake NIH Dataset cont.

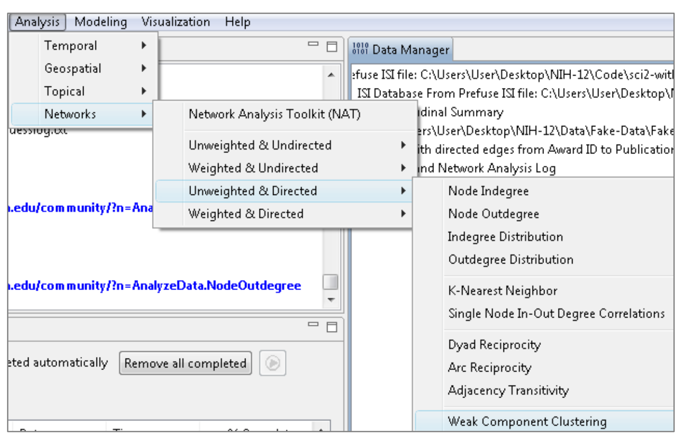
In Sci2

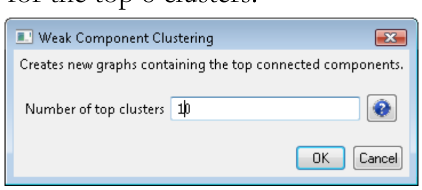
Weak Component Clustering.

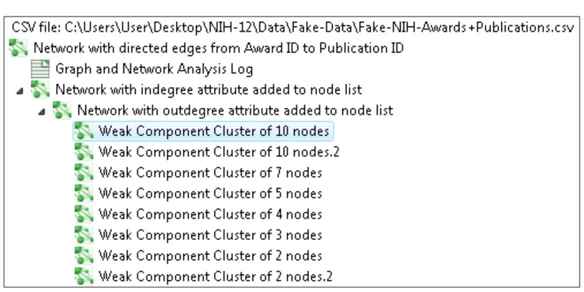
Input Parameters:

Number of top clusters: 10

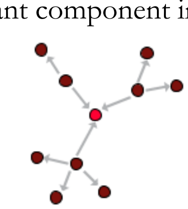
8 clusters found, generating graphs for the top 8 clusters.





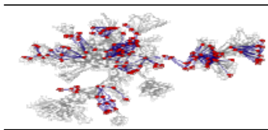


Visualize giant component in GUESS



[#08] Network Analysis and Visualization

- General Overview
- Designing Effective Network Visualizations
- Notions and Notations
- Sci2-Reading and Extracting Networks
- Sci2-Analyzing Networks
- Sci2-Visualizing Networks
- Outlook
- Exercise: Identify Promising Network Analyses of NIH Data



Couple Network Analysis and Visualization to Generate Readable Layouts of Large Graphs

Discover Landmark Nodes based on

- Connectivity (degree or BC values)
- Frequency of access

(Source: Mukherjea & Hara, 1997; Hearst p. 38 formulas)

Identify Major (and Weak) Links

Identify the Backbone

Show Clusters

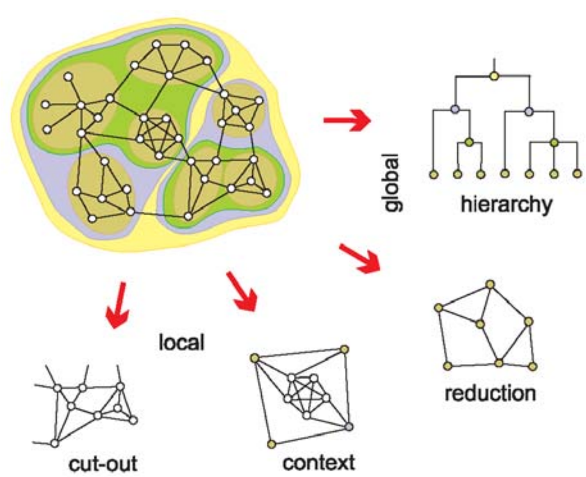


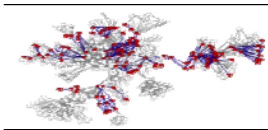
Figure 2: *Approaches to deal with large networks*

Pajek Tutorial

See also Ketan Mane's Qualifying Paper
http://ella.slis.indiana.edu/~kmane/phdprogress/quals/kmane_quals.pdf
http://ella.slis.indiana.edu/~katy/teaching/ketan-quals-slides.ppt

[#08] Network Analysis and Visualization

- General Overview
- Designing Effective Network Visualizations
- Notions and Notations
- Sci2-Reading and Extracting Networks
- Sci2-Analysing Networks
- Sci2-Visualizing Networks
- Outlook
- Exercise: Identify Promising Network Analyses of NIH Data



Network Visualization

General Visualization Objectives

- Representing structural information & content information
- Efficient space utilization
- Easy comprehension
- Aesthetics
- Support of interactive exploration

Challenges in Visualizing Large Networks

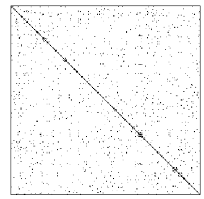
- Positioning nodes without overlap
- De-cluttering links
- Labeling
- Navigation/interaction

General Network Representations

Matrices

1	0	0	6	0
0	10.5	0	0	0
0	0	.015	0	0
0	250.5	0	-280	33.32
0	0	0	0	12

Structure Plots

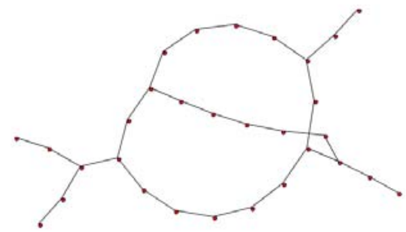


Equivalenced representation of US power network

Lists of nodes & links

```
*Vertices 3
1 "Doc1" 0.0 0.0 0.0 ic Green bc Brown
2 "Doc2" 0.0 0.0 0.0 ic Green bc Brown
3 "Doc3" 0.0 0.0 0.0 ic Green bc Brown
*Arcs
1 2 3 c Green
2 3 5 c Black
*Edges
1 3 4 c Green
```

Network layouts of nodes and links



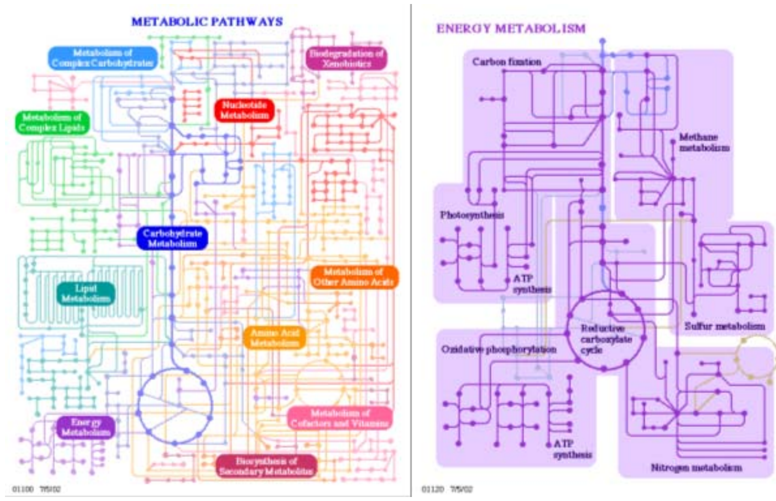
Aesthetic Criteria for Network Visualization

- Symmetric.
- Evenly distributed nodes.
- Uniform edge lengths.
- Minimized edge crossings.
- Orthogonal drawings.
- Minimize area / bends / slopes / angles

Optimization criteria may be relaxed to speed up layout process.

(Source: Fruchterman & R. alg p. 76, see Table & discussion Hearst, p 88)

Aesthetic Network Visualization



http://www.genome.ad.jp/kegg/pathway/map/map01100.html

Small Networks

- Up to 100 nodes
- All nodes and edges and most of their attributes can be shown.

General mappings for

nodes

- # -> (area) size
- Intensity (secondary value) -> color
- Type -> shape 

edges

- # -> thickness
- Intensity, age, etc. -> color
- Type -> style

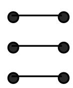
Medium Size Networks

- Up to 10,000 nodes
- Most nodes can be shown but not all their labels.
- Frequently, the number of edges and attributes need to be reduced.

Major design strategies:

Show only important nodes, edges, labels, attributes

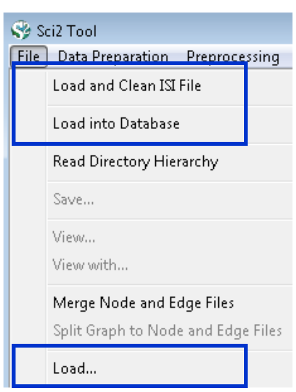
Order nodes spatially

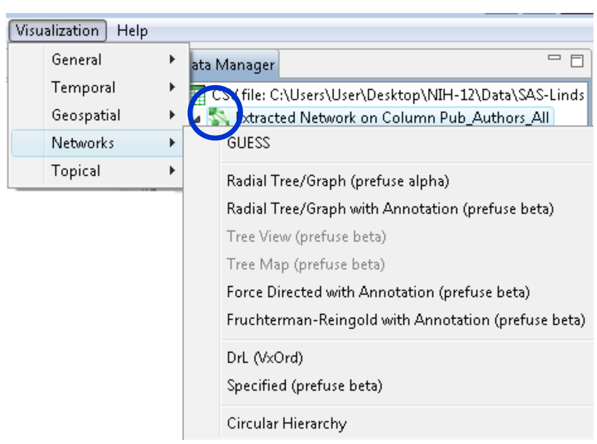


Reduce number of displayed nodes

3 •—•

Visualize Networks with Sci2 Tool





See Science of Science (Sci2) Tool User Manual, Version Alpha 3, Section 3.1 for a listing and brief explanations of all plugins. [*http://sci.slis.indiana.edu/registration/docs/Sci2_Tutorial.pdf*](http://sci.slis.indiana.edu/registration/docs/Sci2_Tutorial.pdf)

See also **Tutorial #3**

NIH Datasets Used

Using NSF Awards Search

- NSF Medical AND Health Awards (283 awards, $152,015,288 total, Sept 2003-July 2014)

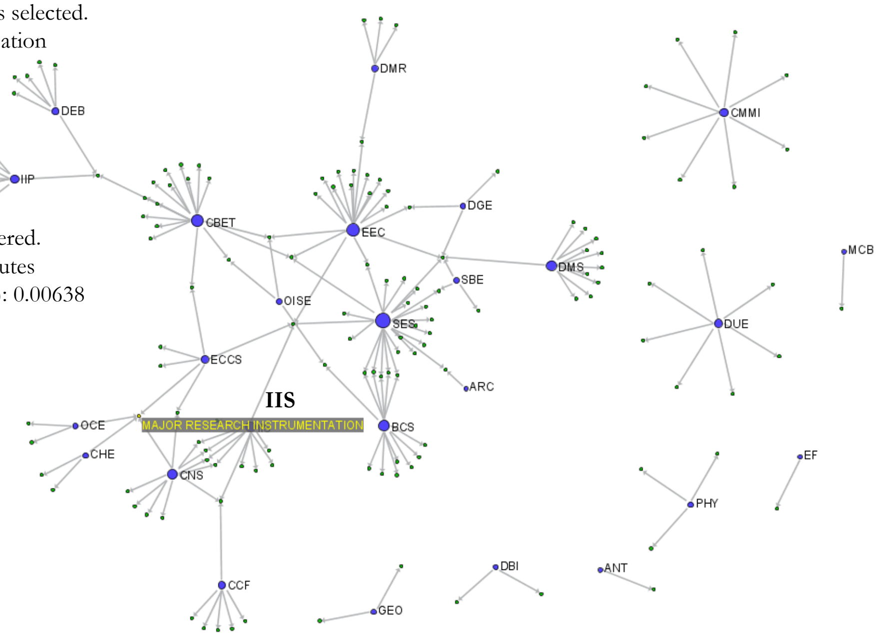
Using NIH RePORTER

- NIH CTSA Funding (534 records, $1,210,288,444 total 'FY Total Cost', Sept. 2006-June 2011) and linked Publications (2,456 records)

NSF Medical+Health Funding: Bimodal Network of NSF Organization to Program(s)

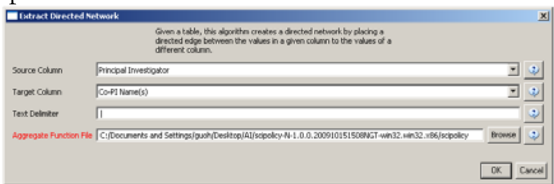
Extract Directed Network was selected.
Source Column: NSF Organization
Text Delimiter: |
Target Column: Program(s)

Nodes: 167
Isolated nodes: 0
Edges: 177
No parallel edges were discovered.
Did not detect any edge attributes
Density (disregarding weights): 0.00638



NSF Medical+Health Funding:
Extract Principal Investigator: Co-PI Networks

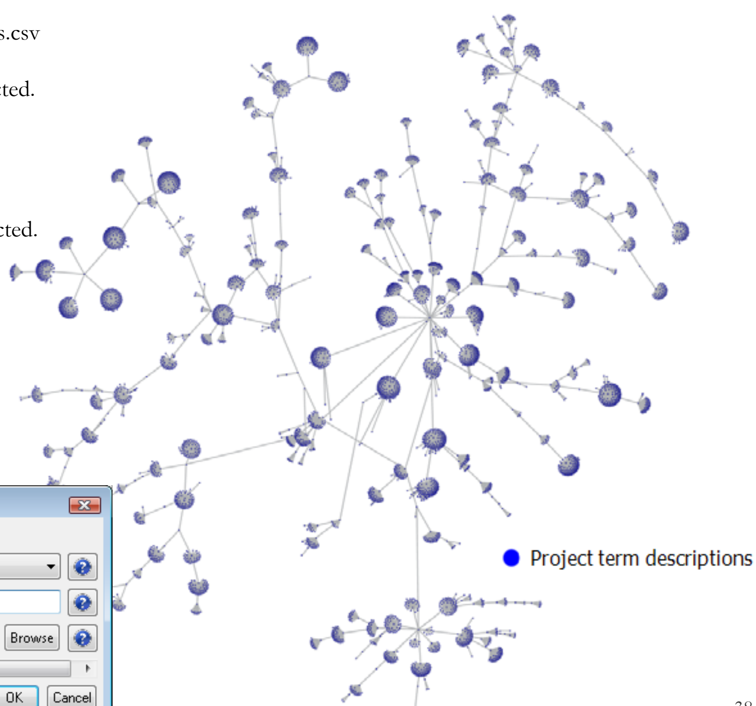
- Load into NWB, open file to count records, compute total award amount.
- Run '*Scientometrics > Extract Directed Network*' using parameters:

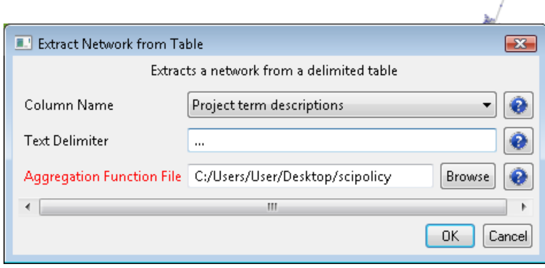


- Select *"Extracted Network .."* and run *'Analysis > Network Analysis Toolkit (NAT)'*
- Remove unconnected nodes via *'Preprocessing > Delete Isolates'.*
- Run '*Analysis > Unweighted & Directed Network > Node Indegree / Node Outdegree*'.
- *'Visualization > GUESS'* , layout with GEM, Bin Pack
- Use Graph Modifier to color/size network.

NIH CTSA Grants:
Co-Project Term Descriptions Occurrence Network

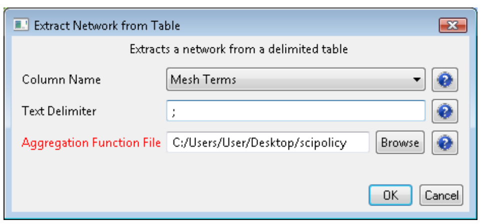
Load... was selected.
Loaded: …\NIH-data\NIH-CTSA-Grants.csv
....
Extract Co-Occurrence Network was selected.
Input Parameters:
Text Delimiter: ...
Column Name: Project term descriptions
....
Network Analysis Toolkit (NAT) was selected.
Nodes: 5723
Isolated nodes: 3
Edges: 353218

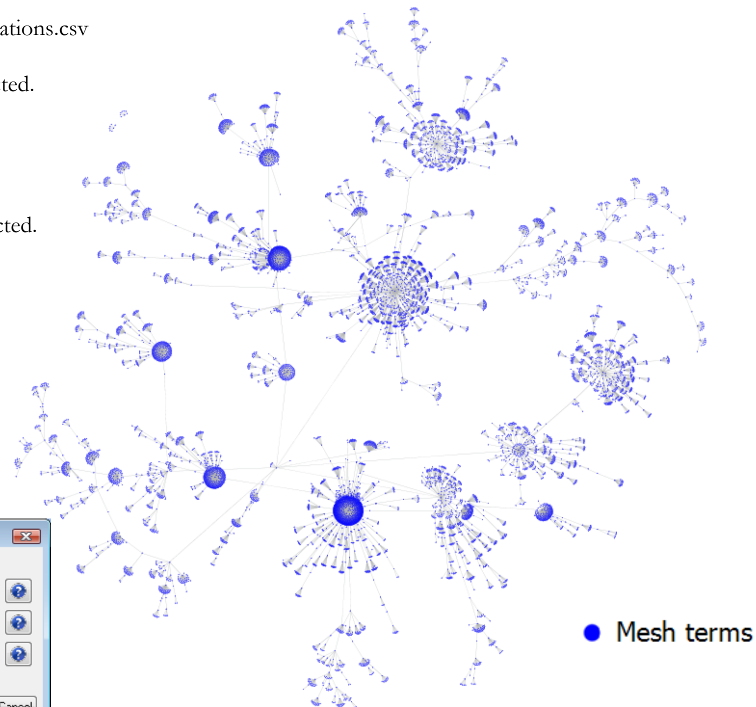




NIH CTSA Publications:
Co-Mesh Terms Occurrence Network

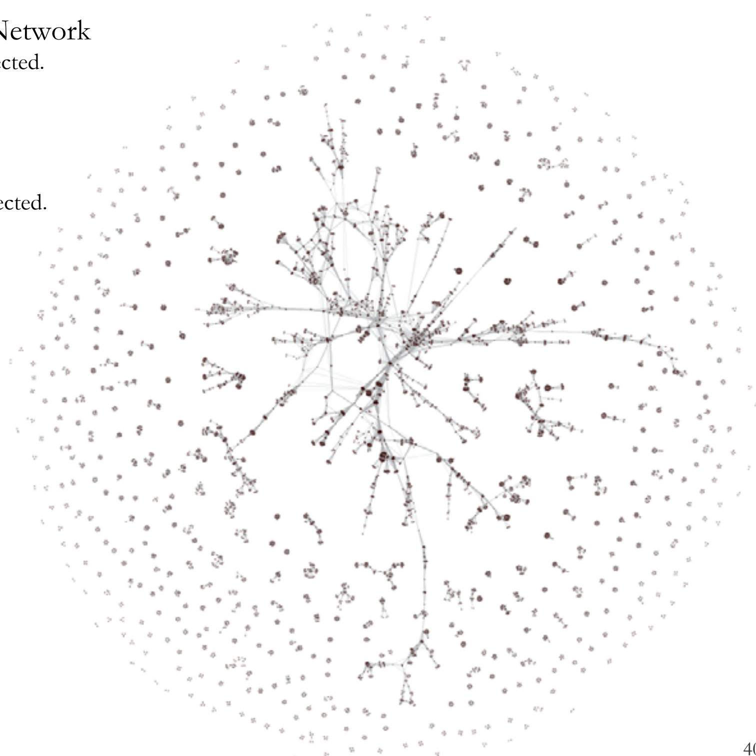
Load... was selected.
Loaded: …\NIH-data\NIH-CTSA-Publications.csv
....
Extract Co-Occurrence Network was selected.
Input Parameters:
Text Delimiter: ;
Column Name: Mesh Terms
....
Network Analysis Toolkit (NAT) was selected.
Nodes: 10218
Edges: 163934

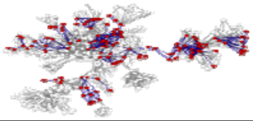




NIH CTSA Grants:
Publication Co-Author Network

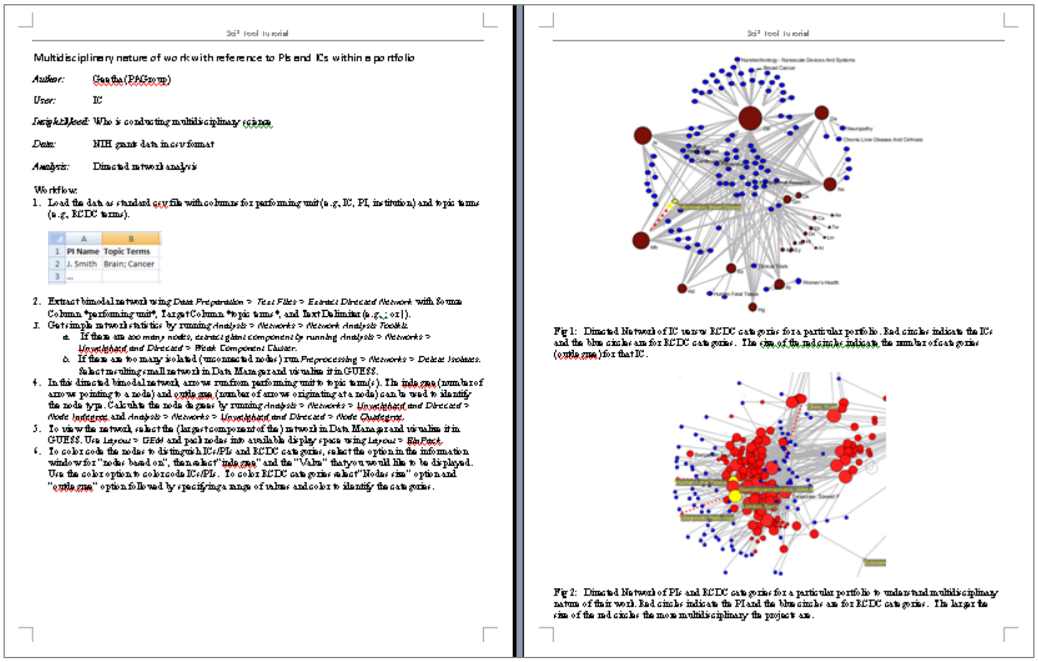
Extract Author Co-occurrence Network
Extract Co-Occurrence Network was selected.
Input Parameters:
Text Delimiter: ;
Column Name: Authors
....
Network Analysis Toolkit (NAT) was selected.
Nodes: 8680
Isolated nodes: 27
Edges: 50160





Visualize multidisciplinary nature of work with reference to PIs and ICs within a portfolio by Geetha Senthil (PAGroup)

Please see Sci2-Tutorial-Geetha-Senthil.pdf



Sci² Tool Tutorial

Multidisciplinary nature of work with reference to PIs and ICs within a portfolio

Author: Geetha (PAGroup)

User: IC

Insight Need: Who is conducting multidisciplinary science

Data: NIH grants data in csv format

Analysis: Directed network analysis

Workflow:

1. Load the data as standard csv file with columns for performing unit (e.g., IC, PI, institution) and topic terms (e.g., RCDC terms).

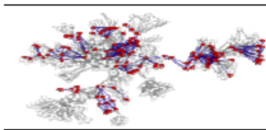
	A	B
1	PI Name	Topic Terms
2	J. Smith	Brain; Cancer
3	...	

2. Extract bimodal network using *Data Preparation > Text Files > Extract Directed Network* with Source Column "performing unit", Target Column "topic terms", and Text Delimiter (e.g., ; or |).
3. Get simple network statistics by running *Analysis > Networks > Network Analysis Toolkit.*
   a. If there are too many nodes, extract giant component by running *Analysis > Networks > Unweighted and Directed > Weak Component Cluster.*
   b. If there are too many isolated (unconnected nodes) run *Preprocessing > Networks > Delete Isolates.* Select resulting small network in Data Manager and visualize it in GUESS.
4. In this directed bimodal network, arrows run from performing unit to topic term(s). The indegree (number of arrows pointing to a node) and outdegree (number of arrows originating at a node) can be used to identify the node type. Calculate the node degrees by running *Analysis > Networks > Unweighted and Directed > Node Indegree* and *Analysis > Networks > Unweighted and Directed > Node Outdegree.*
5. To view the network, select the (largest component of the) network in Data Manager and visualize it in GUESS. Use *Layout > GEM* and pack nodes into available display space using *Layout > Bin Pack.*
6. To color code the nodes to distinguish ICs/PIs and RCDC categories, select the option in the information window for "nodes based on", then select "indegree" and the "Value" that you would like to be displayed. Use the color option to color code ICs/PIs. To color RCDC categories select "Nodes size" option and "outdegree" option followed by specifying a range of values and color to identify the categories.

Sci² Tool Tutorial

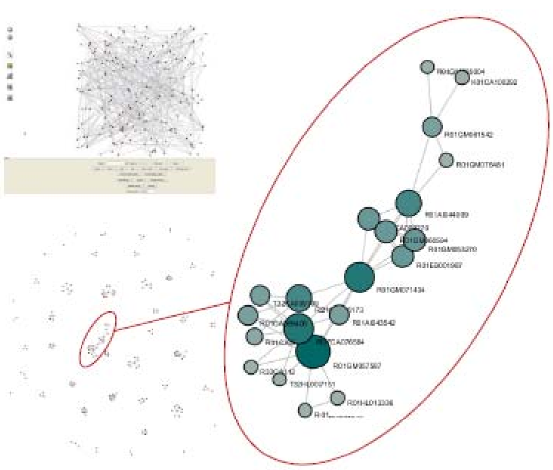
Fig 1: Directed Network of IC versus RCDC categories for a particular portfolio. Red circles indicate the ICs and the blue circles are for RCDC categories. The size of the red circles indicate the number of categories (outdegree) for that IC.

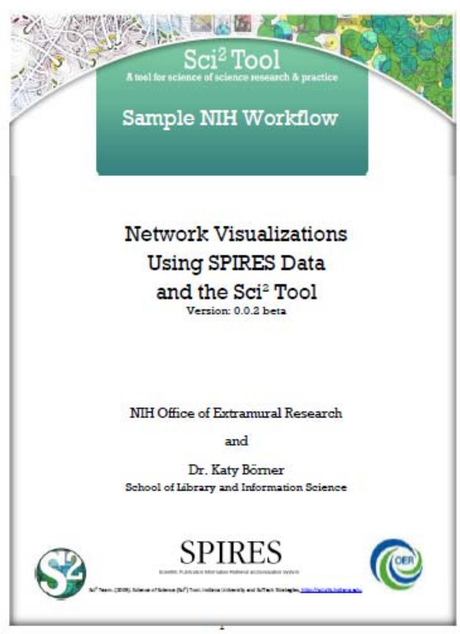
Fig 2: Directed Network of PIs and RCDC categories for a particular portfolio to understand multidisciplinary nature of their work. Red circles indicate the PI and the blue circles are for RCDC categories. The larger the size of the red circles the more multidisciplinary the projects are.



Network Visualizations Using SPIRES Data and the Sci² Tool by NIH Office of Extramural Research and Katy Börner

Please see Sci2 Tutorial, Network Visualizations Using SPIRES Data, 2010-06-01.pdf and My Project Publications.csv





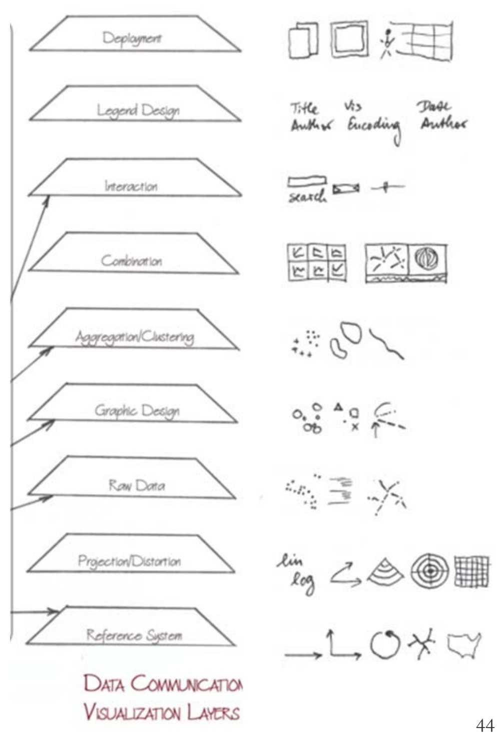
[#08] Network Analysis and Visualization

- General Overview
- Designing Effective Network Visualizations
- Notions and Notations
- Sci2-Reading and Extracting Networks
- Sci2-Analysing Networks
- Sci2-Visualizing Networks
- Outlook
- Exercise: Identify Promising Network Analyses of NIH Data

Outlook – Visualization Layers

See **Tutorial #02**

- **Deployment** of results is enabled through paper printouts, online animations, or interactive, three-dimensional, audiovisual environments.
- The **Legend Design** delivers guidance on the purpose, generation, and visual encoding of the data. Mapmakers should proudly sign their visualizations, adding credibility as well as contact information.
- In many cases, it is desirable to **Interact** with the data, that is, to zoom, pan, filter, search, and request details on demand. Selecting a data entity in one view might highlight this entity in other views.
- Sometimes it is beneficial to show multiple simultaneous views of the data, here referred to as **Combination**.
- Frequently, **Aggregation/Clustering** techniques are applied to identify data entities with common attribute values or dense connectivity patterns.
- **Graphic Design** refers to the visual encoding of data attributes using qualities such as size, color, and shape coding of nodes, linkages, or surface areas.
- Placing the **Raw Data** in a reference system reveals spatial patterns.
- **Projections/Distortions** of the reference system help emphasize certain areas or provide focus and context.
- **Reference Systems** organize the space.



Outlook - OSGi/CIShell Adoption

See **Tutorial #03**

A number of other projects recently adopted OSGi, among them are:

Cytoscape (http://www.cytoscape.org) lead by Trey Ideker, UCSD is an open source bioinformatics software platform for visualizing molecular interaction networks and integrating these interactions with gene expression profiles and other state data (Shannon et al., 2002).

TEXTrend (http://www.textrend.org) lead by George Kampis, Eötvös University, Hungary develops a framework for the easy and flexible integration, configuration, and extension of plugin-based components in support of natural language processing (NLP), classification/mining, and graph algorithms for the analysis of business and governmental text corpuses with an inherently temporal component.

As the functionality of OSGi-based software frameworks improves and the number and diversity of dataset and algorithm plugins increases, the capabilities of custom tools will expand.

Run ***Cytoscape*** out of Sci2 Tool by adding org.textrend.visualization.cytoscape_0.0.3.jar to the /plugin directory.

Soon, general 'star database' will be available. NIH database is planned.

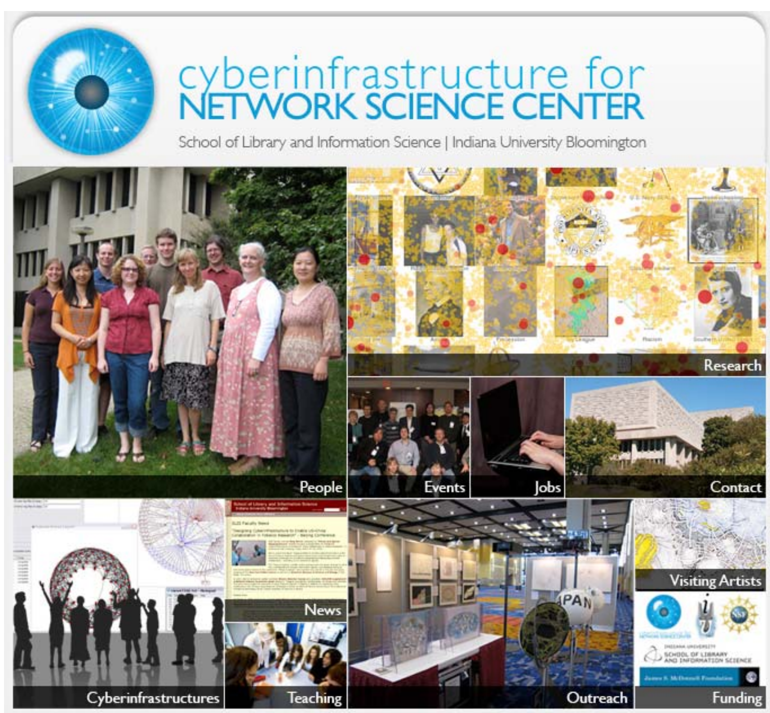
[#08] Network Analysis and Visualization

- General Overview
- Designing Effective Network Visualizations
- Notions and Notations
- Sci2-Reading and Extracting Networks
- Sci2-Analysing Networks
- Sci2-Visualizing Networks
- Outlook
- Exercise: Identify Promising Network Analyses of NIH Data

Exercise

Please identify a promising network analysis of NIH data.

Document it by listing
- Project title
- User, i.e., who would be most interested in the result?
- Insight need addressed, i.e., what would you/user like to understand?
- Data used, be as specific as possible.
- Analysis algorithms used.
- Visualization generated. Please make a sketch with legend.



All papers, maps, cyberinfrastructures, talks, press are linked from http://cns.slis.indiana.edu