

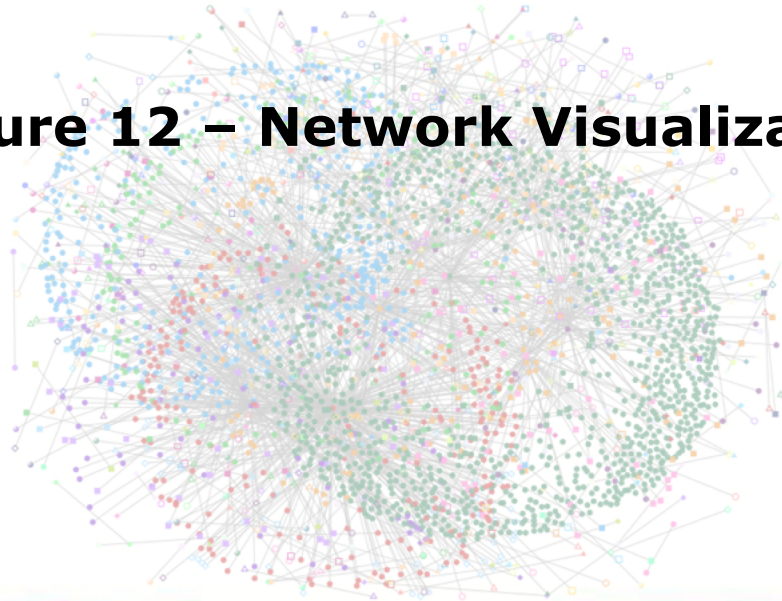
CTI-3A3

Applied Social Network Analysis

Dr. Warih Maharani

wmaharani.staff.telkomuniversity.ac.id

Lecture 12 – Network Visualization



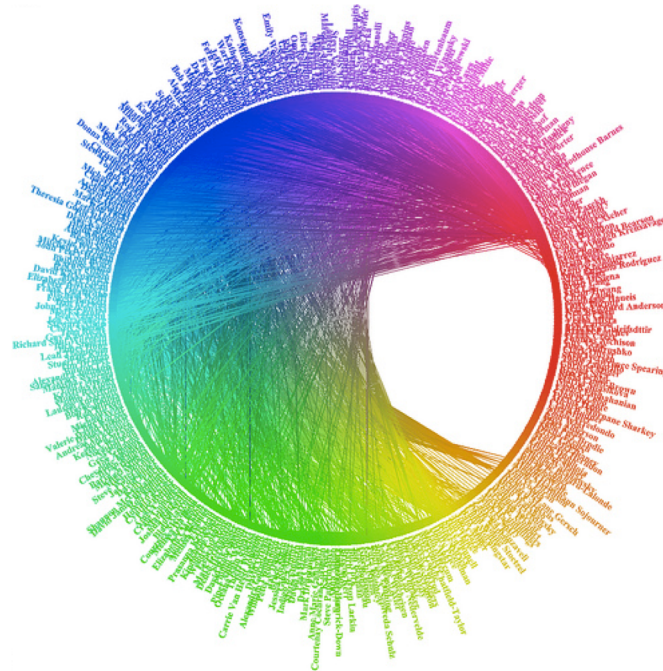
Outline

- ▶ Reviews: Network
- ▶ Network Visualization
- ▶ Gephi

Course Learning Outcomes (CLO)

- ▶ use software to analyze and visualize network formation and evolution;
- ▶ use software to simulate the dynamics of networks based on social network models.

Reviews: Network



What is a Network?

▶ Graph – Network visualized

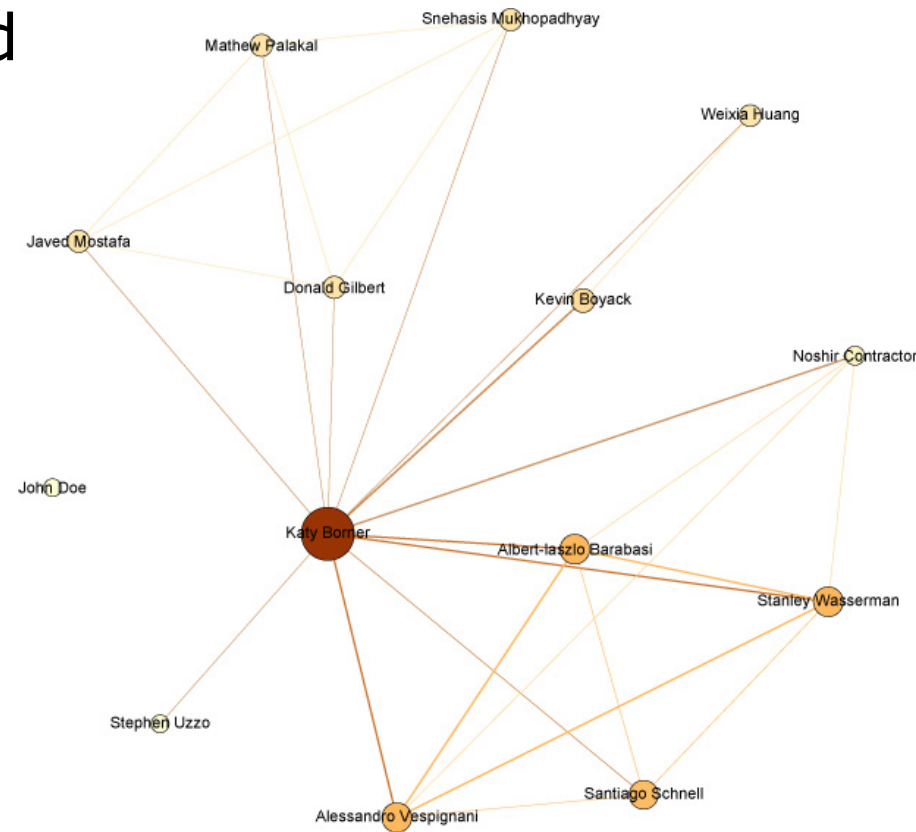
- Nodes
- Edges
- Components

▶ Representations

- Matrices
- Graphs
- Edge and Node Lists

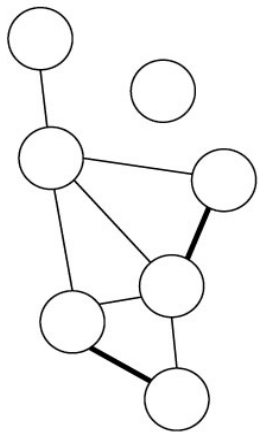
▶ Data Formats

- Tabular
- XML
- Text
- JSON



General types of networks

Undirected Networks



Nodes:



Edges:



Node Degree:

Number of edges connected to nodes

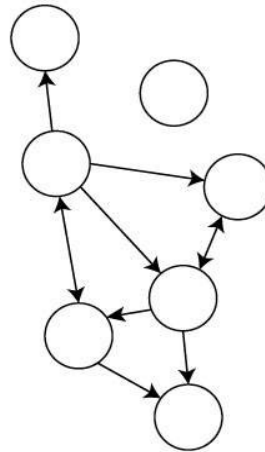
Isolates:

Nodes that are not connected to the rest of the network

Edge Weight:

Demonstrates relative importance of relationships

Directed Networks



Edge Direction:

Directional relationship is represented by arrows

In-Degree:

Number of incoming edges

Out-Degree:

Number of outgoing edges

Other types of networks and graphs:

- Hierarchical networks (tree networks)
- Bipartite Networks
- Multigraphs
- Hypergraphs

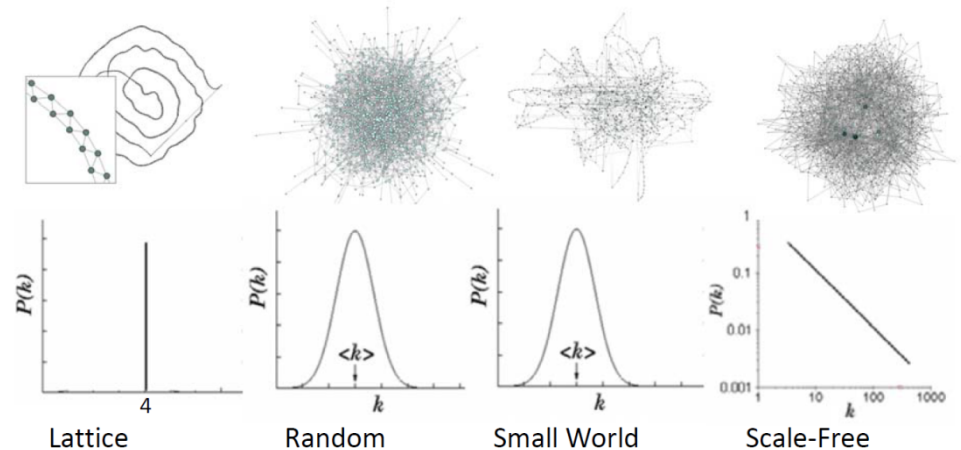
Graph Features

General Topologies

- Random Graphs network
- Watts-Strogatz / Small World network
 - gene networks, food chains, voter networks, power grids
- Barabasi-Albert Scale Free network
 - Internet, Citation Networks, Social Network

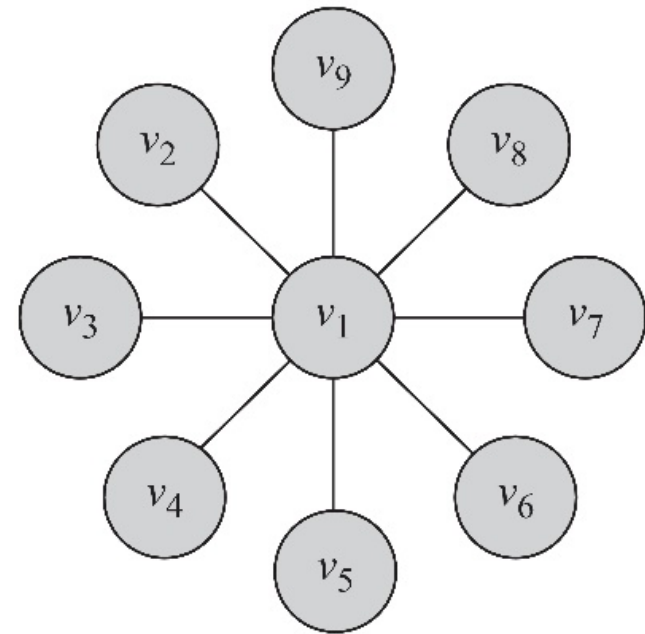
Measurements

- Node and Edge Counts
- Network Components
- Giant Component
- Avg. degree distribution
- Avg. Clustering
- Density
- Avg. Path Length
- Diameter

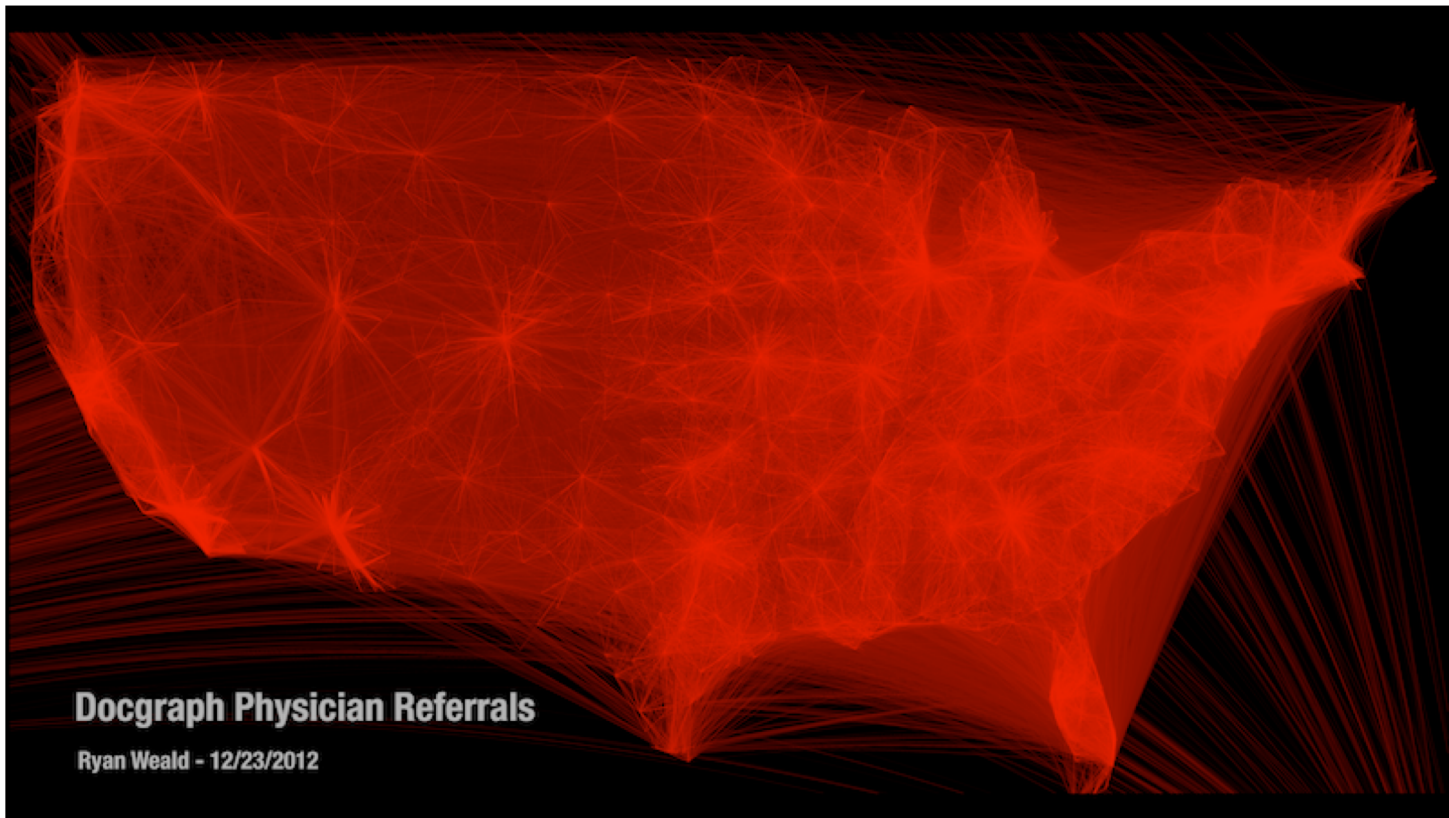


Centralities

- Degree Centrality
- Betweenness centrality
- Closeness centrality
- Eigenvector centrality

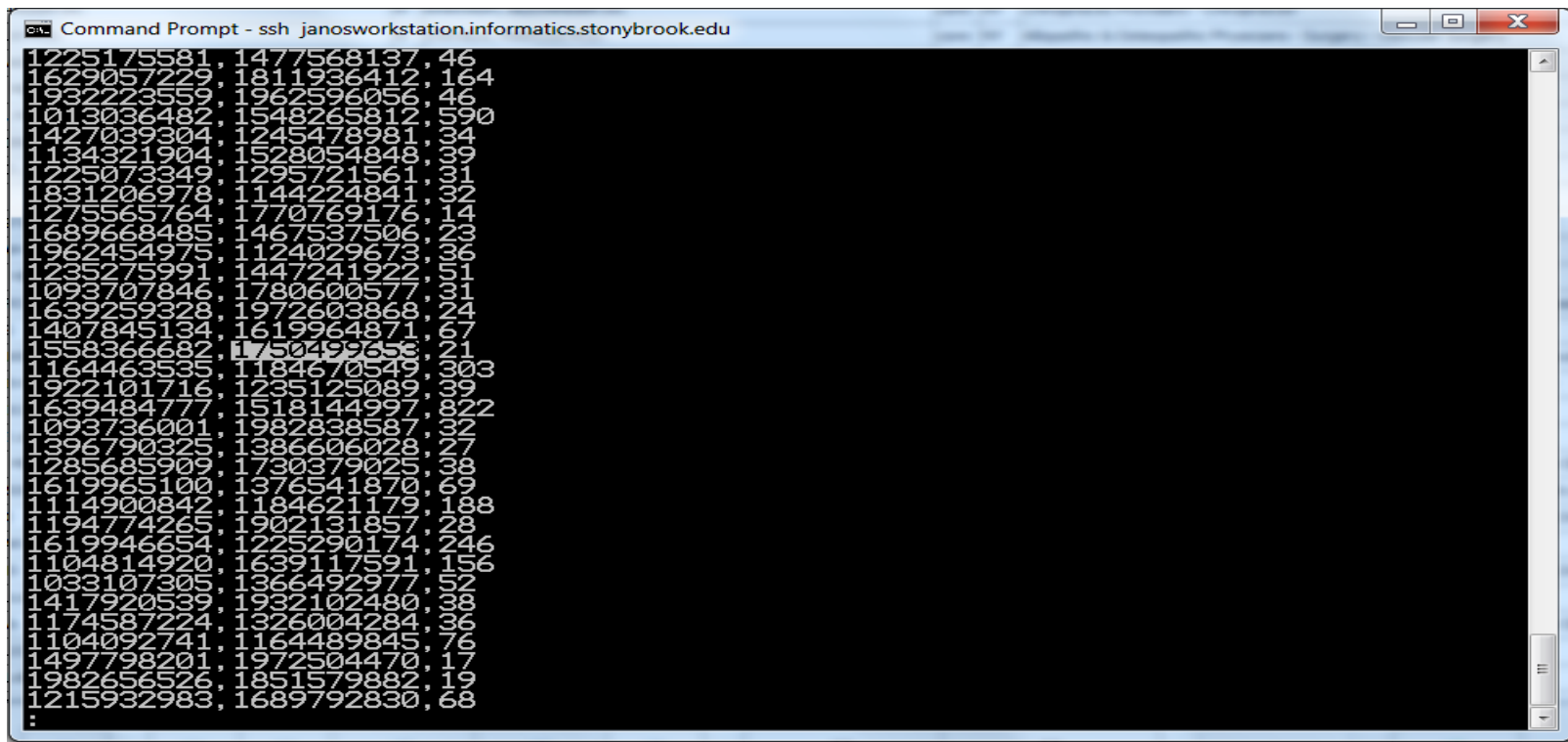


Geographic visualization



<http://isurfsoftware.com/blog/2012/12/13/visualizing-geographic-connections-between-us-doctors/>

DocGraph data



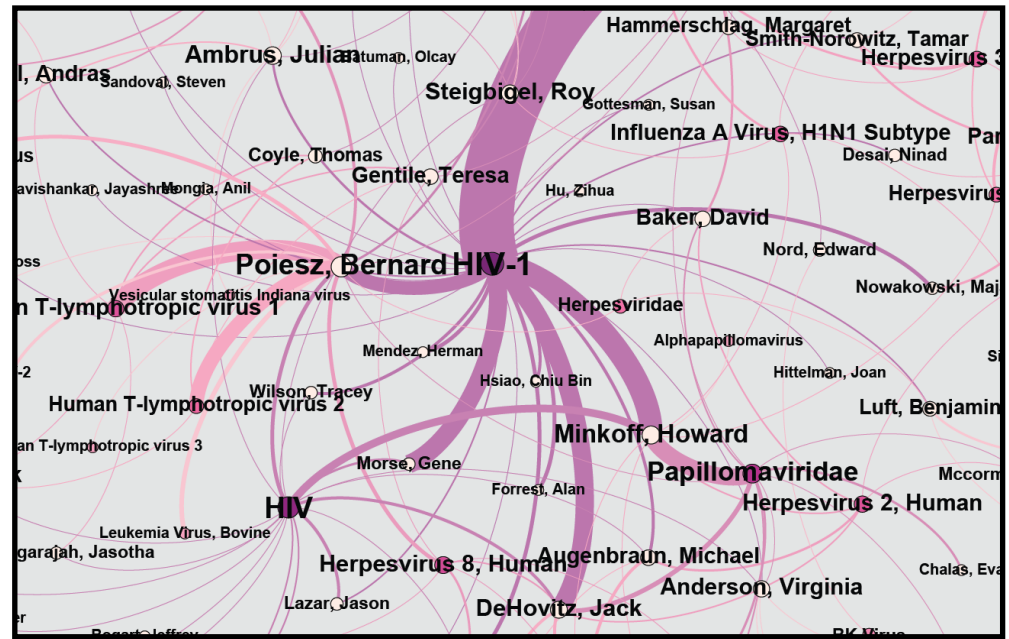
```
cmd Command Prompt - ssh_janosworkstation.informatics.stonybrook.edu
1225175581, 1477568137, 46
1629057229, 1811936412, 164
1932223559, 1962596056, 46
1013036482, 1548265812, 590
1427039304, 1245478981, 304
1134321904, 1528054848, 39
1225073349, 1295721561, 31
1831206978, 1144224841, 32
1275565764, 1770769176, 14
1689668485, 1467537506, 23
1962454975, 1124029673, 36
1235275991, 1447241922, 51
1093707846, 1780600577, 31
1639259328, 1972603868, 24
1407845134, 1619964871, 67
1558366682, 1750499658, 21
1164463535, 1184670549, 303
19221081716, 1235125089, 39
1639484777, 1518144997, 822
1093736001, 1982838587, 32
1396790325, 1386606028, 27
1285685909, 1730379025, 38
1619965100, 1376541870, 69
1114900842, 1184621179, 188
1194774265, 1902131857, 28
1619946654, 1225290174, 246
1104814920, 1639117591, 156
1033107305, 1366492977, 52
1417920539, 1932102480, 38
1174587224, 1326004284, 36
1104092741, 1164489845, 76
1497798201, 1972504470, 17
1982656526, 1851579882, 19
1215932983, 1689792830, 68
:
```

Tabular data

	A	B	D	E	F	G	H
1	NPI	Entity_Type_Code	Employer_Identification_Number_EIN	Provider_Organization_Name_Legal_Business_Name	Provider_Last_Name_Legal_Name	Provider_First_Name	Provider_Midd
2	1679576722	1			WIEBE	DAVID	A
3	1588667638	1			PILCHER	WILLIAM	C
4	1497758544	2	<UNAVAIL>	CUMBERLAND COUNTY HOSPITAL SYSTEM, INC			
5	1306849450	1			SMITSON	HAROLD	LEROY
6	1215930367	1			GRESSOT	LAURENT	
7	1023011178	2	<UNAVAIL>	NAPA VALLEY HOSPICE & ADULT DAY SERVICES			
8	1932102084	1			ADUSUMILLI	RAVI	K
9	1841293990	1			WORTSMAN	SUSAN	
10	1750384806	1			BISBEE	ROBERT	
11	1669475711	1			SUNG	BIN	SHENG
12	1578566626	1			KUIPERS	WARREN	D.
13	1487657532	1			HUEBERT	ALLISON	L
14	1396748448	1			YOUNGMAYKA	CYNTHIA	
15	1205839354	1			DIFILIPPO	EMIL	A
16	1114920261	1			THACKER	RICHARD	RANDALL
17	1023011079	2	<UNAVAIL>	ADVANTAGE HOME HEALTH CARE, INC.			
18	1932102985	1			ROTHSTEIN	MARK	TERRY
19	1841293891	1			GIBBS	ELMER	RICKEY
20	1750384707	1			MARKMAN	ALAN	WILLIAM
21	1669475612	1			TROTCHIE	DEBBIE	C
22	1578566527	1			DYSART	STANLEY	H
23	1487657433	2	<UNAVAIL>	PEKIN MRI, LLC			
24	1396748349	1			GRUNERT	GEORGE	M
25	1205839255	1			GOLDBERG	STEVEN	M
26	1114920162	1			DUBOSE	JON	

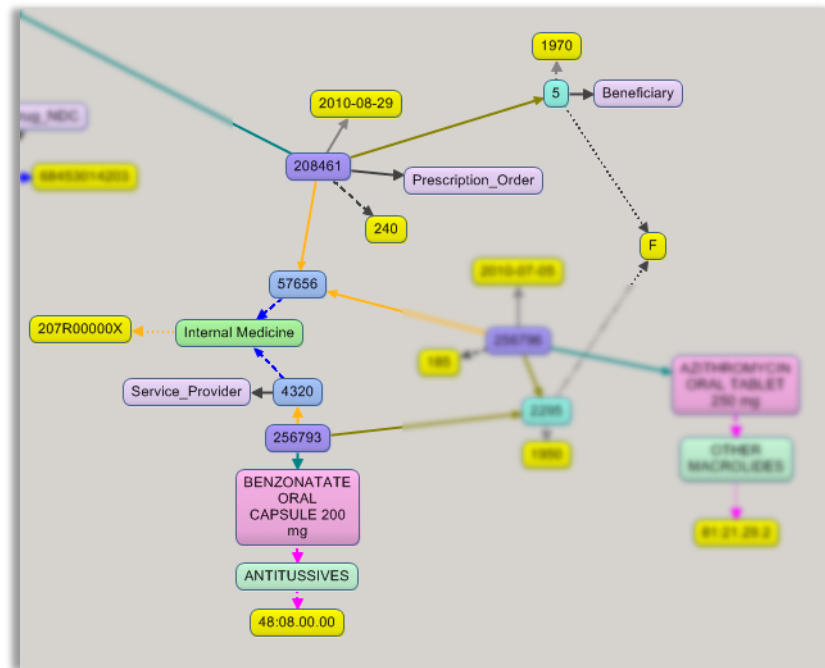
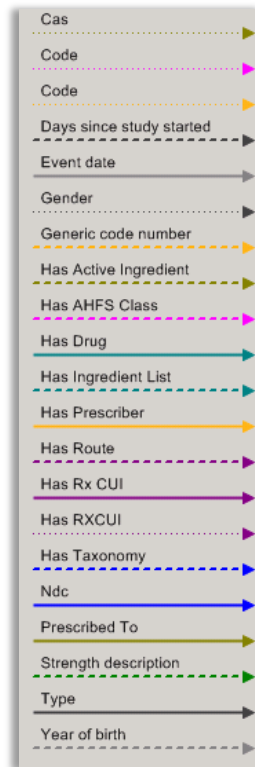
Graph data

- Relation between authors and MeSH terms from PubMed



Graphs in healthcare

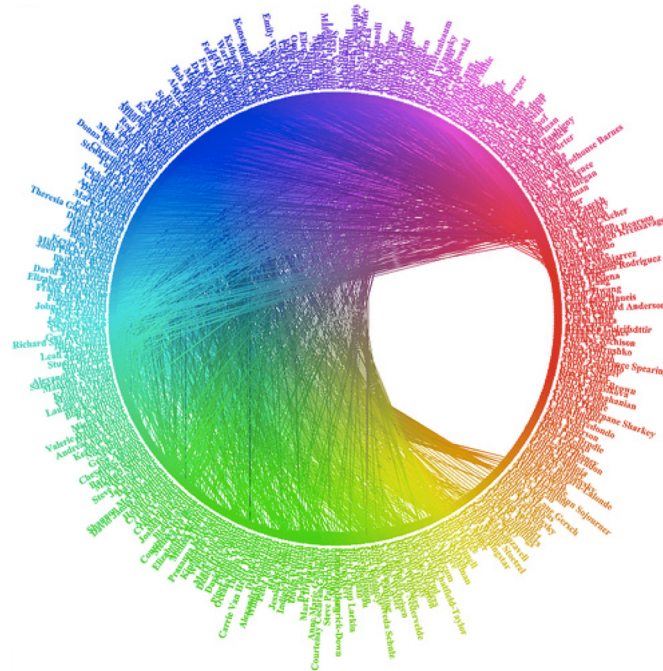
- Prescriber and patient (bipartite)
 - NCPDP data with NPI
- Referral data sets
- Shared patients
 - DocGraph
- Social networks
 - Tweeting about a disease
- Limited by imagination



Generating GraphML

- ▶ XML based file format for graphs
- ▶ Readable by a large number of tools
 - Gephi
 - Mathematica
 - igraph (R)
- ▶ NetworkX a Python library for graphs which can export to GraphML
- ▶ GraphML is not a file format for really large graphs
- ▶ GraphML is not readable by d3.js

Network Visualization



Information Visualization

- ▶ external visual representation of data, exploits perceptual system to reduce human cognitive load
- ▶ find appropriate visual metaphor for data that is not implicitly spatial

Visualizing Network Data

- ▶ Visualize the data associated with a network
 - Understand data, not network themselves
- ▶ Coping with large data volumes
 - Hundreds of nodes
 - Thousands of links
 - Data from time periods
- ▶ Overcome the map clutter problem

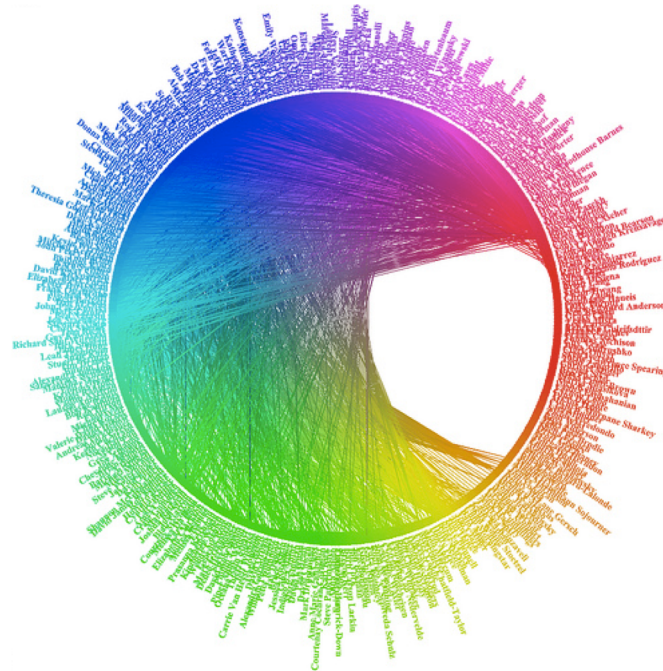
Network Tools

- ▶ **NetworkX**, network analysis package (Python)
- ▶ **igraph**, network analysis tools (Python, C++, R)
- ▶ **graph-tool**, network analysis and visualization software (Python, C++)
- ▶ **GraphLab**, scalable network analysis (Python, C++)

Network Visualization

- › [Cytoscape](#), network visualization software
- › [yEd Graph Editor](#), network visualization software
- › [Graphviz](#), network visualization software
- › [Gephi](#), network visualization software
- › [graph-tool](#), network analysis and visualization software
- › [webweb](#), network visualization tool joining Matlab and d3
- › [MuxViz](#), multilayer analysis and visualization platform

Gephi



Gephi

- ▶ Exploratory Network Analysis
 - Static and Dynamic Networks
- ▶ Visualization
 - Node & Edge Size, Color, & Labels
 - Layout algorithms
- ▶ Filtering and Partitioning
- ▶ Cluster Analysis
- ▶ Loads and exports a variety of network formats
 - Streaming Graphs
- ▶ Extensible Plugin Library
- ▶ Save raster and vector graphics
- ▶ Open source tool
 - [Gephi GitHub](#)
- ▶ Large user community

Gephi Resources

- ▶ Gephi Support - <https://gephi.org/users/support/>
- ▶ Bug Reporting - <https://github.com/gephi/gephi/issues>
- ▶ Gephi Tutorials - <https://gephi.org/users/>
- ▶ Gephi User Community
 - Volunteer Opportunities
<https://gephi.org/users/contribute/>
 - Developers - <https://gephi.org/developers/>
 - User Support - <http://forum-gephi.org/>

Gephi Interface

Gephi tool layout and visualization framework:
user is free to rearrange the environment, move panels,
show/hide windows, etc.

The GUI is set by default for three task families grouped as
Overview, Data Laboratory and Preview.

- ▶ **Overview:** graph analysis and manipulation mode.
- ▶ **Data Laboratory:** data tables.
- ▶ **Preview:** visual tuning before vectorial/raster export.

(<https://github.com/gephi/gephi/wiki/GUI>)

The screenshot displays the Gephi 0.9.1 interface with a central graph view and several surrounding panels. The graph shows a dense network of nodes and edges. The interface includes a menu bar (File, Workspace, Tools, Window, Plugins, Help), a toolbar (Overview, Data Laboratory, Preview), and a workspace area. The Overview Pane is highlighted with an orange border and contains several toolbars and panels:

- Appearance X:** A toolbar with icons for nodes and edges, and a text box describing its functions: "Pane lets user control size and color of nodes and edges based on network partitions, node and edge rankings, and clustering results and scale values to splines."
- Graph X:** A toolbar with icons for dragging and configuring, and a text box: "Tools let you select nodes and edges, and color nodes and edges based on paths."
- Context X:** A panel showing basic network statistics: "Nodes: 410" and "Edges: 2083", and a text box: "Basic network stats and filter stats."
- Filters X / Statistics:** A panel with a "Reset" button and a "Library" of filters (Attributes, Dynamic, Edges, Operator, Topology, Saved queries), and a text box: "Pane contains network statistical analysis algorithms. And network filters for subsets and partitioning based on node and edge variables".
- Appearance X (bottom):** A panel with a "Run" button and a text box: "Pane lets user select a network layout, and adjust the layout algorithm parameters".
- Appearance X (bottom):** A panel with a "Presets..." button and a "Reset" button, and a text box: "These tools let you re-center the network, and rest color, size and label attributes."
- Appearance X (bottom):** A panel with a "Filter" button and a text box: "Tools let you adjust network labels attributes, and take snapshots of the graph viewer."
- Appearance X (bottom):** A panel with a "Filter" button and a text box: "Navigate workspaces".

Select the node or edge lists data, and configure the sheet

Add node edges, imports data (nodes and edges lists) to create new networks in blank workspaces, and exports data tables for a network.

A REGEX filter for nodes and edges table columns in data table.

Nodes	Label	number_of_authored...	times_cited	blondel_community_level_0	blondel_community_level_1
n9		9	1389	community_0	community_0
n4		4	48	community_0	community_0
n4		4	48	community_0	community_0
n2		2	17	community_0	community_0
n4		4	48	community_0	community_0
n4		4	48	community_0	community_0
n3		3	41	community_0	community_0
n5	Chaves, Sandra S.	5	365	community_0	community_0
n5	Gubareva, Larisa	5	307	community_0	community_0
n12	Hall, Henrietta	2	17	community_0	community_0
n13	Wallis, Teresa	4	48	community_0	community_0
n14	Villanueva, Julie	4	48	community_0	community_0
n15	Xu, Xiyao	4	48	community_0	community_0
n16	Bresee, Joseph	10	1712	community_0	community_0
n17	Cox, Nancy	10	1995	community_0	community_0
n28	Tappero, Jordan W.	3	58	community_3	community_3
n30	Nyenswah, Tolbert G.	6	58	community_1	community_1
n31	Montgomery, Joel M.	5	42	community_1	community_1
n32	Neatherlin, John	3	25	community_1	community_1
n35	Singleton, James A.	10	933	community_54	community_45
n36	Flannery, Brendan	2	51	community_2	community_2
n37	Fry, Alicia	3	310	community_2	community_2
n41	Pesik, Nicki	2	52	community_3	community_3
n42	Brown, Clive M.	3	61	community_3	community_3
n43	Aranas, Aaron E	3	54	community_16	community_16
n48	Cohen, Nicole J.				
n51	Hale, Christa				
n55	Holton, Kelly				
n65	Clacke, Kevin P.				

Adds & Merge data column tools

Column editing tools

Create columns fitting Boolean criteria and regex functions. Useful for filtering.

Converts data fields from standard to dynamic (temporal data fields)

The screenshot shows the Gephi 0.8.2 interface in the Preview mode. The main window displays a dense network graph with numerous nodes and edges. On the left, the 'Preview Settings' panel is open, showing various configuration options for nodes, edges, and labels. Several callout boxes are overlaid on the interface, pointing to specific settings:

- Preset layouts, and a layout configuration saving feature:** Points to the 'Presets' section at the top of the settings panel.
- Node border and opacity attributes:** Points to the 'Nodes' section, specifically 'Border Color' and 'opacity'.
- Node labels attribute selection, including label size, color, length, etc.:** Points to the 'Node Labels' section, including 'Font', 'Proportional size', 'Color', and 'Shorten label'.
- Edge size, color, type, and opacity and scaling attributes:** Points to the 'Edges' section, including 'Show Edges', 'Thickness', 'Rescale weight', 'Color', 'Opacity', 'Curved', and 'Radius'.
- Edge label attributes:** Points to the 'Edge Labels' section, including 'Show Labels', 'Font', 'Color', 'Shorten label', 'Max characters', 'Outline size', 'Outline color', and 'Outline opacity'.
- Refresh and exporting the Preview tools, and zoom resets:** Points to the bottom of the interface, including the 'Refresh' button, 'Export: SVG/PDF/PNG' options, and the 'Background Reset zoom' control.

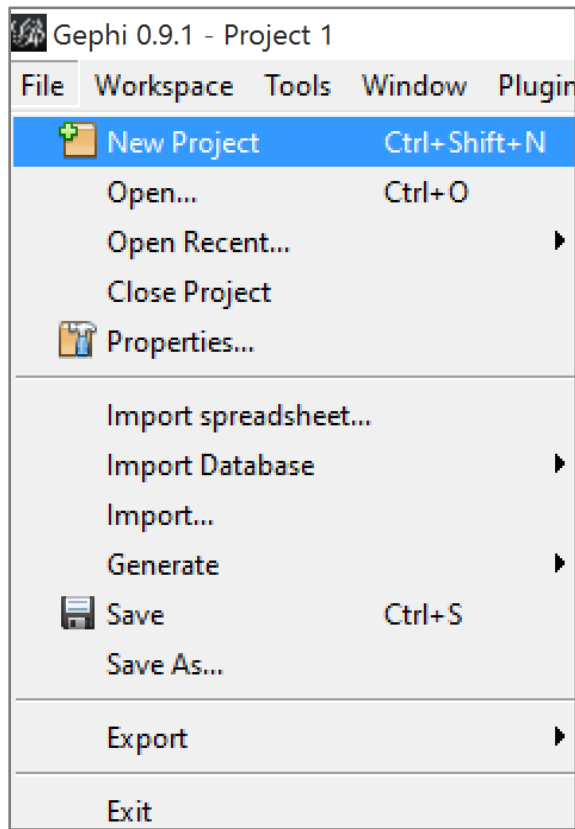
Creating a Visualization in Gephi

How to get data into Gephi

- Loading the data in manually through user interface
 - Overview Window
 - Load in pre-formatted network
 - Data Laboratory
 - Load in Node and Edge Lists CSV files
- Stream in data using Gephi graph API using plugins
- Pass graph data to Gephi using graph API
 - Sci2 network extraction to Gephi

	Edge List/Matrix Structure	XML Structure	Edge Weight	Attributes	Visualization Attributes	Attribute Default Value	Hierarchical Graphs	Dynamics
CSV	■	■						
DL Ucinet		■	■					
DOT Graphviz		■		■				
GDF		■	■	■	■			
GEXF		■	■	■	■	■	■	■
GML		■	■	■	■			
GraphML		■	■	■	■			
NET Pajek	■		■		■			
TLP Tulip								
VNA Netdraw		■	■					
Spreadsheet*							■	

Creating a New Project



Create a New Project

After you've started Gephi, you will need to create a new project.

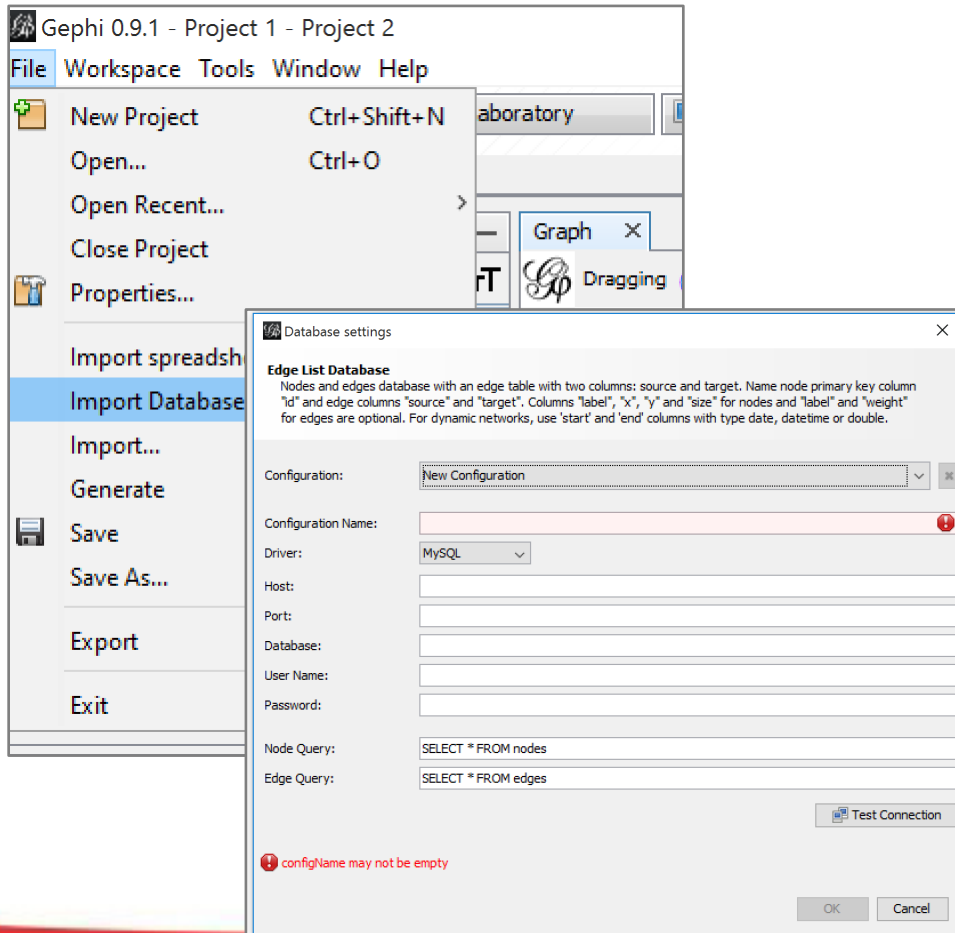
You can select new project from the file menu or select Ctrl+Shift+N on your keyboard.

Loading a Random Graph

- ▶ Creating sample networks
 - Random Graphs
 - Dynamic Networks
 - Multigraphs

The screenshot displays the Gephi 0.9.1 software interface. The 'File' menu is open, with the 'Generate' option highlighted. A sub-menu is visible, showing 'Random Graph...' as the selected option. In the foreground, the 'Random Graph' dialog box is open, showing the 'Number of nodes' field set to 410 and the 'Wiring probability' field set to 0.025. The background shows a network graph visualization in the 'Graph' window, which is currently in a 'Dragging (Configure)' state.

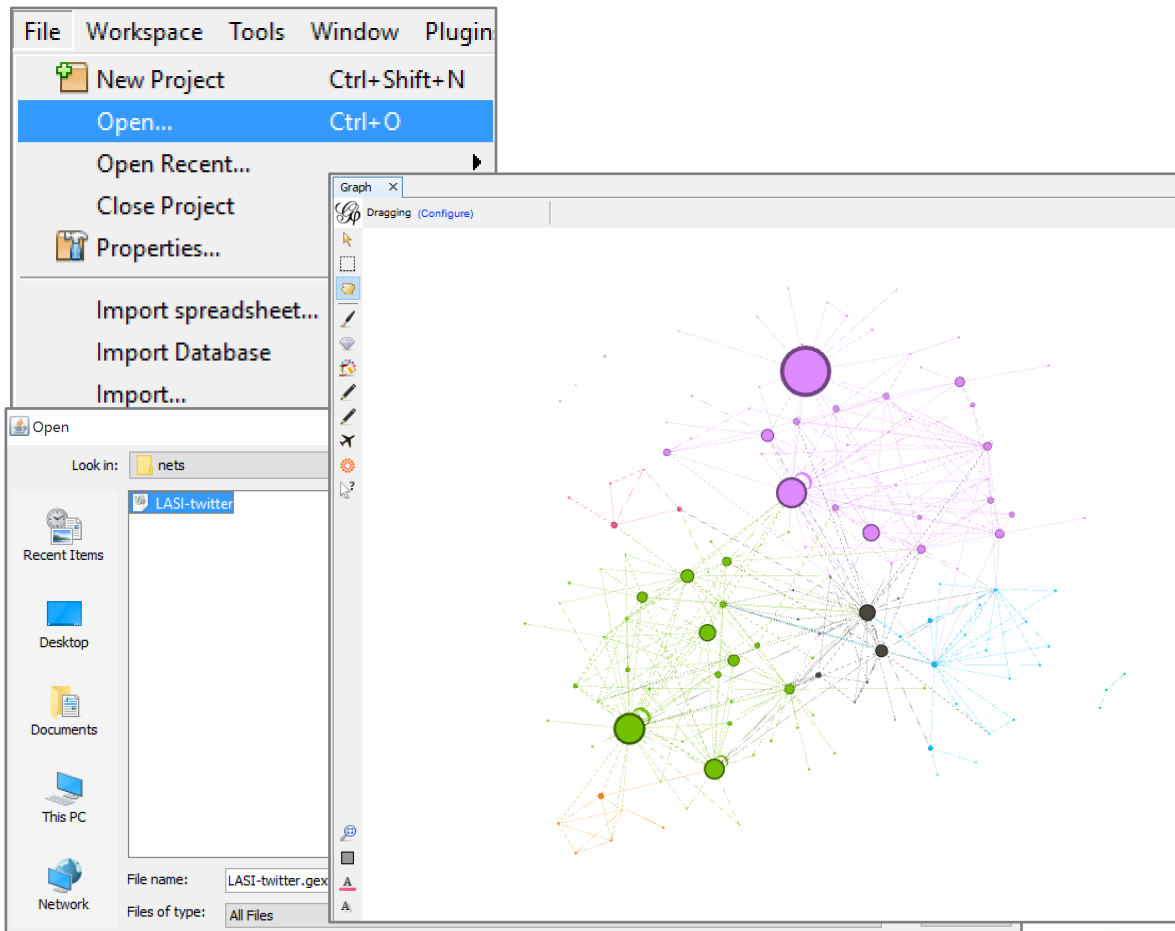
Streaming in an edge list to Gephi



Connecting to a Database

While not the focus of this presentation, documentation for [import a network edge list database](#) is found on the [Gephi GitHub](#).

Loading in Data through GUI



Loading a network data file

To load a formatted network file into Gephi, in the File menu select "Open..." or Ctrl+O on the keyboard.

Loading a Network from Spreadsheets

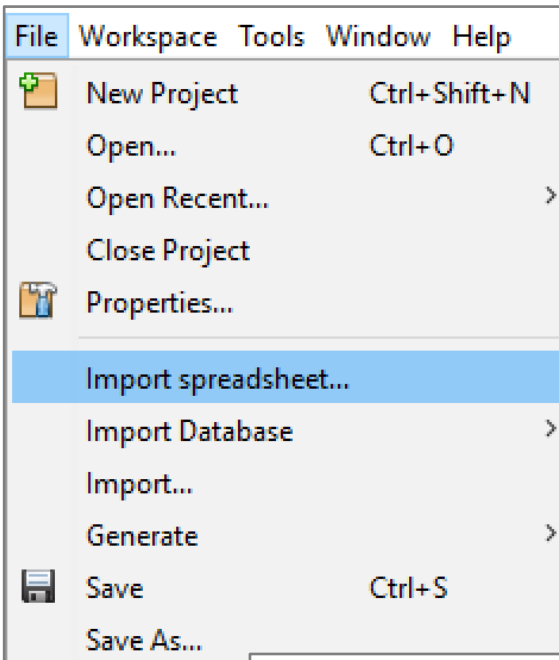
Gephi can import networks from tabular data (as CSVs), if you can provide a node* and edge list using Gephi's *Import Spread Sheet tool*.

- (If using) Node list file should contain an "ID" field that references the values found in the "Source" and "Target" fields of the edge list.
- Edge list must have a "Source" and "Target" field; these are mandatory and can't be deselected
- An Edge List "Weight" field with a numeric data type will be automatically recognized by the import wizard as well
- If a column name already exists in the project space you will be able to use it, but the data type of the column is already set and can't be changed, and
 - Imported data will be parsed to fit the existing column type

Written instructions can be found here:

<https://github.com/gephi/gephi/wiki/Import-CSV-Data>

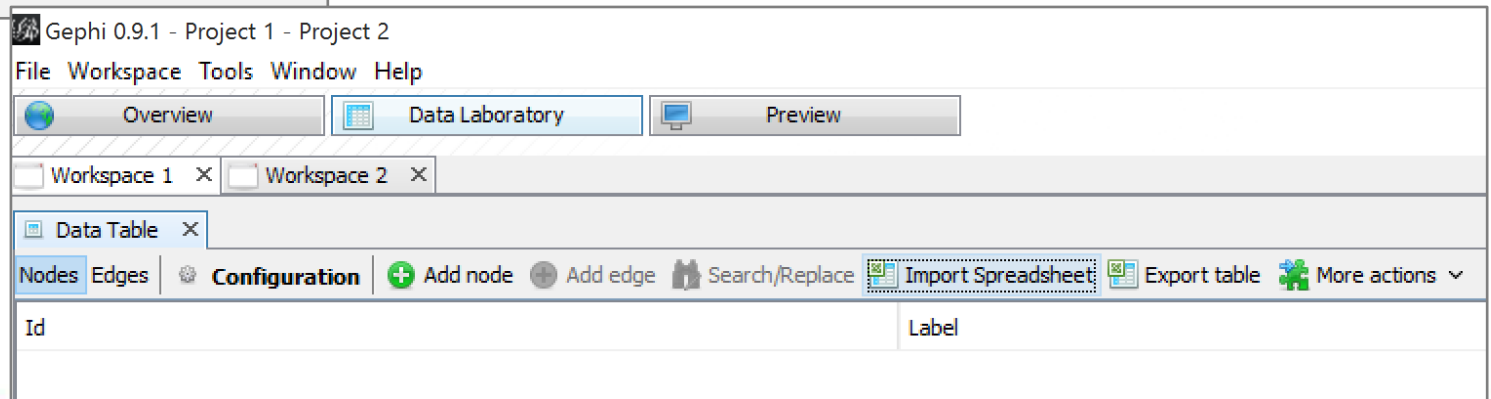
1



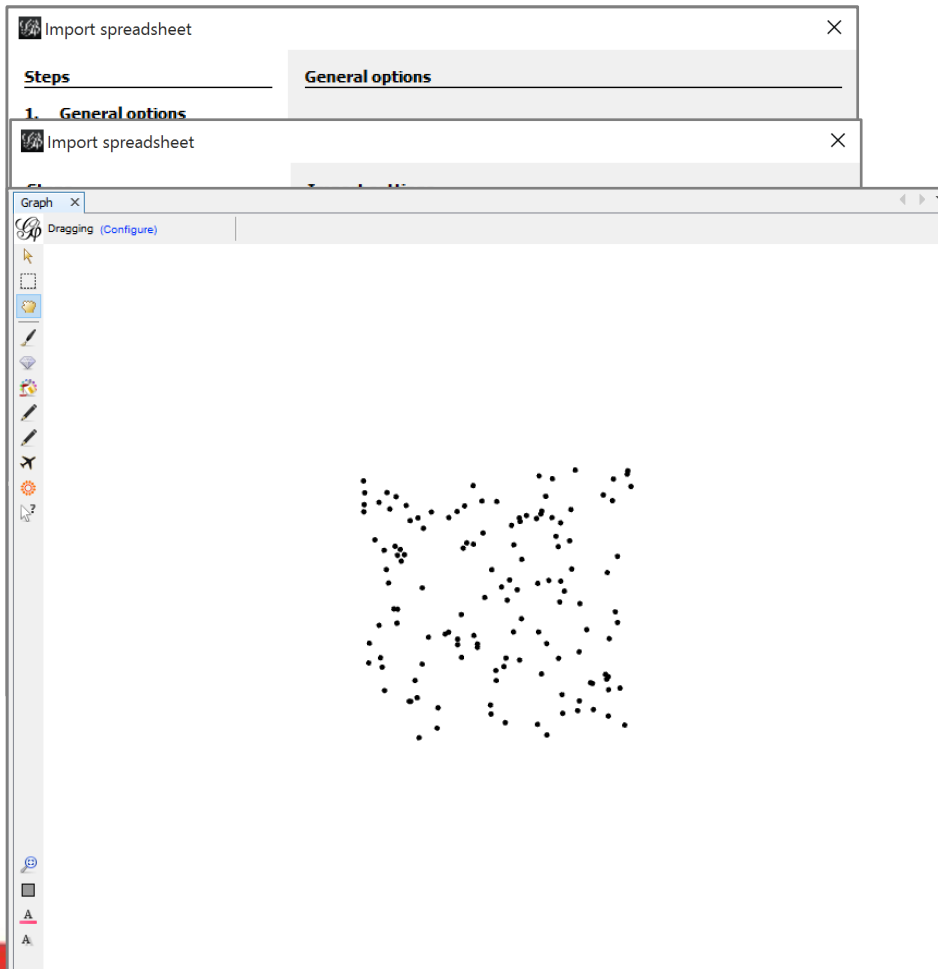
Loading a Network from Spreadsheets

- 1) The spreadsheet import tool may be engaged from the file menu and selecting “Import spreadsheet...” or via
- 2) the Data Laboratory window’s “Data Table” view by selecting “Import spreadsheet...” button.

2



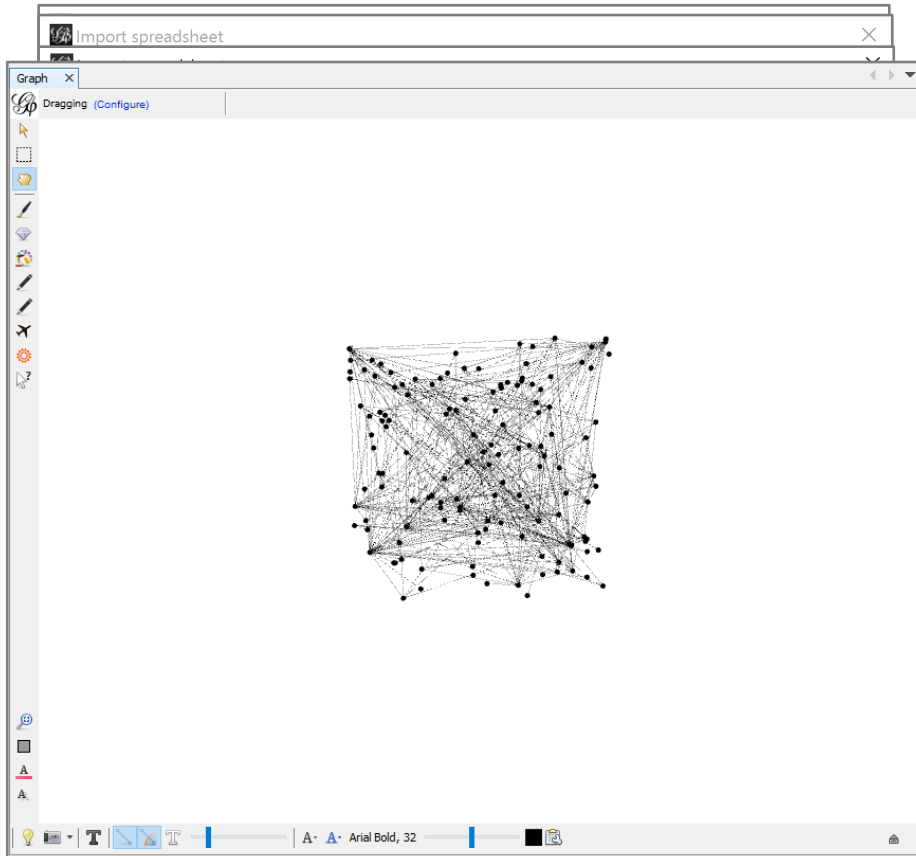
Loading a Network from Spreadsheets



When creating a network, it is easier to start with the node list first.

1. Choose the node list CSV file to import (LASI-Nodes.csv)
2. Make sure the field As Table is set to “Nodes table”, and select Next.
3. The columns will be identified in the table, allowing a user to select the data type, then select Finish.
 - except for preset fields like id and label
 - for column names that have been encountered before
 - A field may be unselected by the check box next to the column name.
4. Looking at the Overview window, shows that the nodes have been loaded.

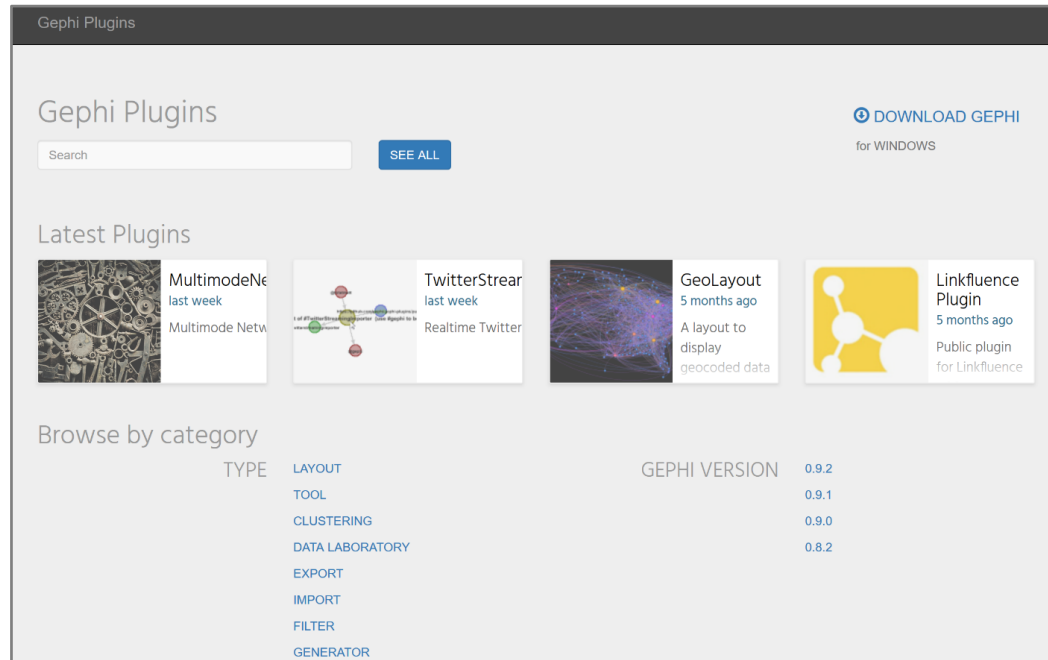
Loading a Network from Spreadsheets



Back in the Data Laboratory, select the “Import spreadsheet...” button, and load in the edge list.

1. Choose the edge list CSV file to import (LASI-Edges.csv)
2. Make sure the field As Table is set to “Edges table”, and select Next.
3. The columns will be identified in the table, allowing a user to select the data type, then select Finish.
 - except for preset fields like source, target, label, and weight.
 - for column names that have been encountered before
 - a field may be unselected by the check box next to the column name.
4. Looking at the Overview window, shows that the edges have been loaded.

Finding and Adding Plugins to Gephi

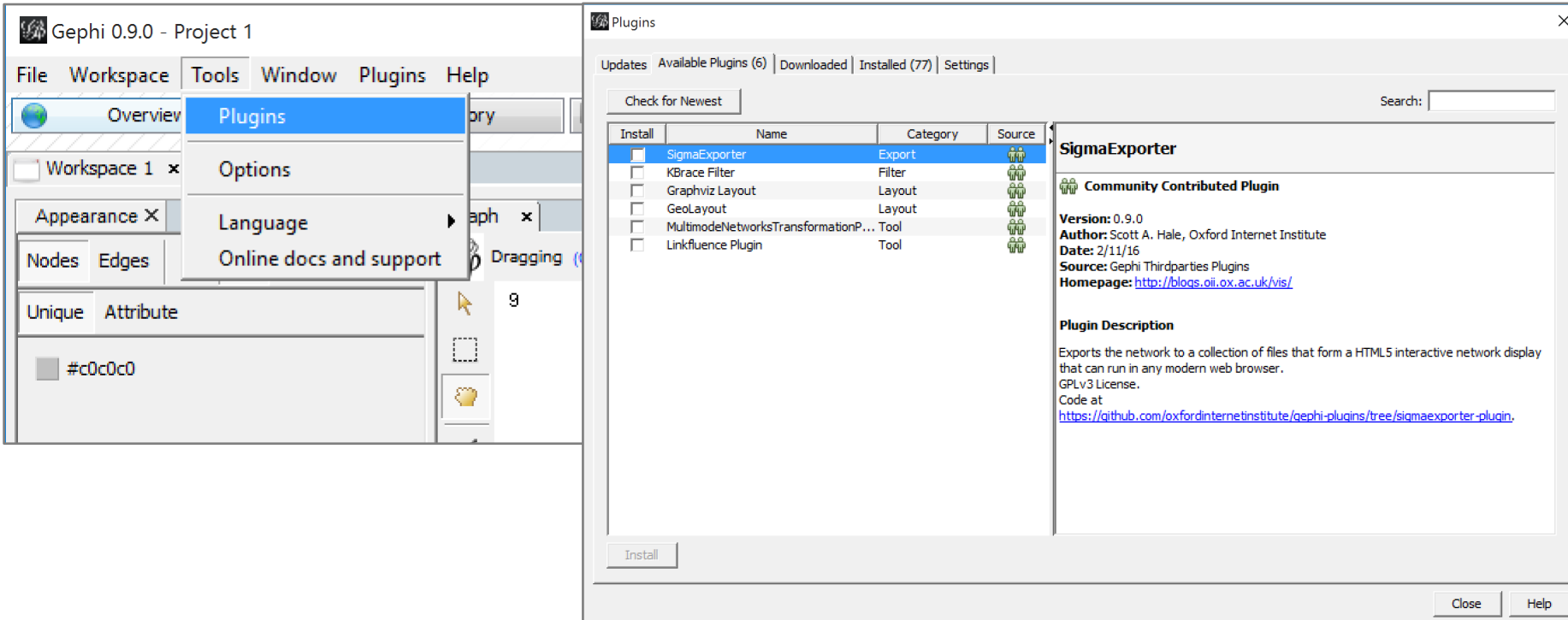


The screenshot shows the Gephi Plugins website interface. At the top, there is a search bar with the text "Search" and a "SEE ALL" button. To the right, there is a link to "DOWNLOAD GEPHI for WINDOWS". Below the search bar, there is a section titled "Latest Plugins" with four plugin cards: "MultimodeNetwork" (last week), "TwitterStream" (last week), "GeoLayout" (5 months ago), and "Linkfluence Plugin" (5 months ago). Below this, there is a "Browse by category" section with a list of categories: LAYOUT, TOOL, CLUSTERING, DATA LABORATORY, EXPORT, IMPORT, FILTER, and GENERATOR. To the right of this list, there is a "GEPHI VERSION" section with a list of versions: 0.9.2, 0.9.1, 0.9.0, and 0.8.2.

Gephi is extensible to user created plugins, that allow users to add new analysis, layout, export features to Gephi. Users can download plug-ins manually (<https://gephi.org/plugins>) or through the user interface in Gephi.

Make sure that plugins are compatible with the version of Gephi that you are running.

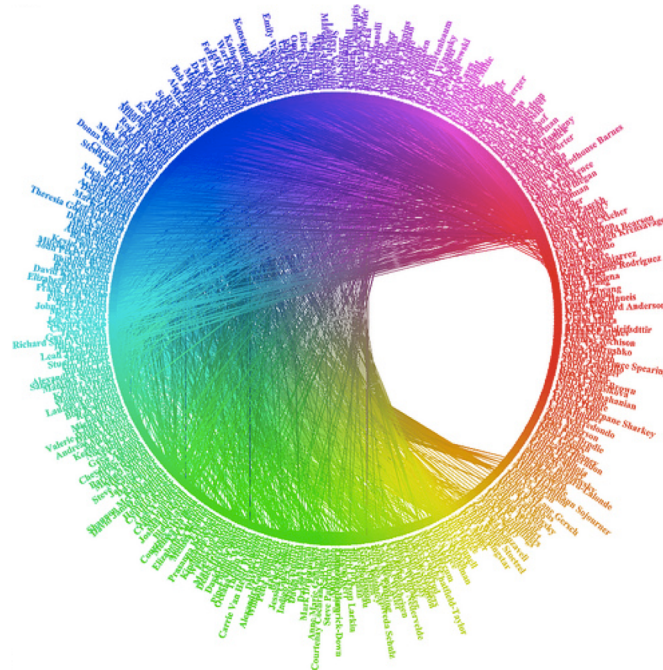
Finding and Adding Plugins to Gephi



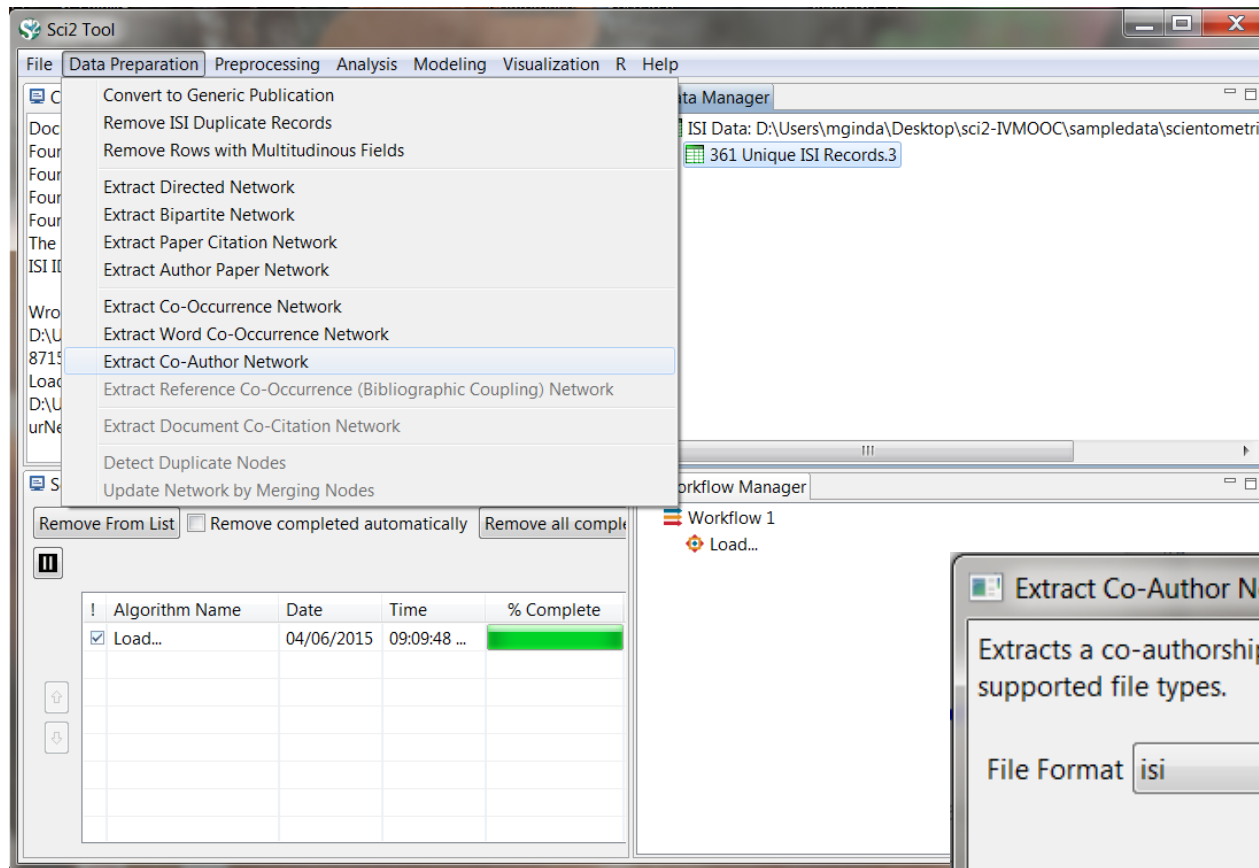
Adding plugins through the user interface is done through the Tool menu, select "Plugins" from the menu. A new window will pop-up, giving you the option to:

- install updates to existing plugins,
- check for plugins that are listed as available for your release, or
- manually add plugins downloaded from Gephi's plugin site; and
- see plugins that are install (which can be activated or deactivated) and see settings.

Co-Authorship Analysis in Gephi



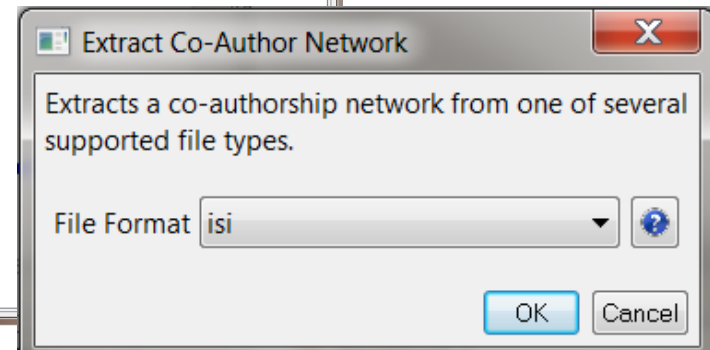
Co-Author Network Analysis



Load *FourNetSciResearchers.isi* located in Sci2 Directory

Select the data 361 Unique ISI Records in the Data manager and the algorithm *Extract Co-Author Network* in the Data Preparation menu.

A pop-up window will appear; select the format ISI.



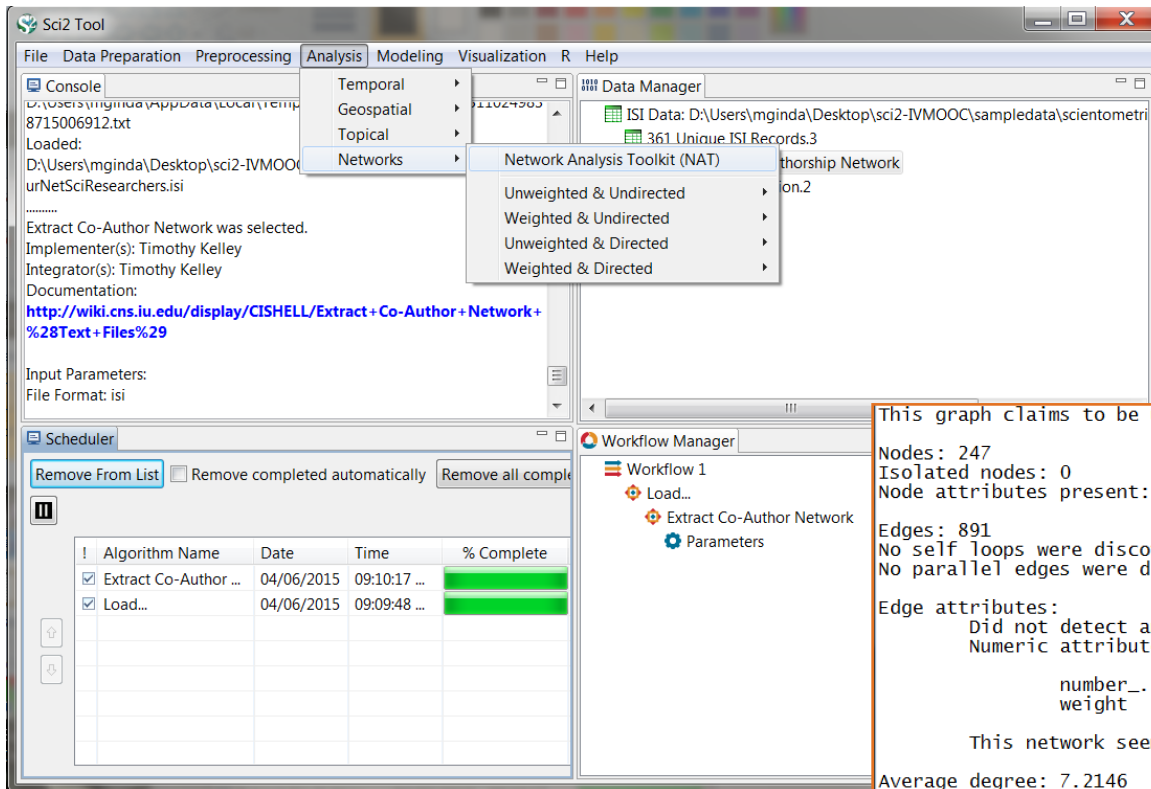
Co-Author Network Analysis

After creating your initial network, it is a good idea to get a brief overview of its statistical properties.

Sci2 has a built-in analysis toolkit to perform these basic statistics.

Select the network output file in the data manager, and then in the menu select *Analysis -> Networks -> Network Analysis Toolkit (NAT)*

The output should read:



```

This graph claims to be undirected.
Nodes: 247
Isolated nodes: 0
Node attributes present: label, number_of_authored_works, times_cited

Edges: 891
No self loops were discovered.
No parallel edges were discovered.

Edge attributes:
Did not detect any nonnumeric attributes.
Numeric attributes:

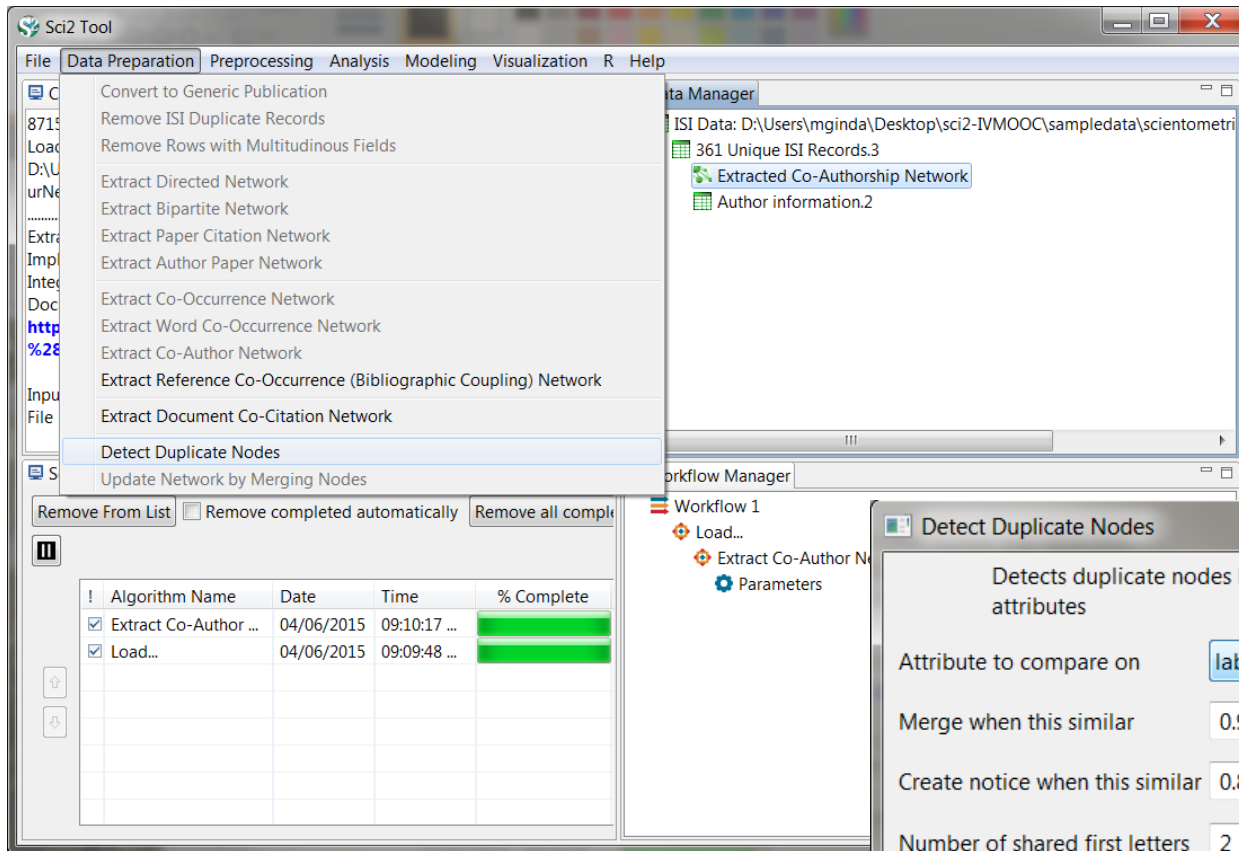
           min    max    mean
number_...  1     33    1.76094
weight      1     33    1.76094

This network seems to be valued.

Average degree: 7.2146
This graph is not weakly connected.
There are 3 weakly connected components. (0 isolates)
The largest connected component consists of 194 nodes.
Did not calculate strong connectedness because this graph was not directed.

Density (disregarding weights): 0.0293
Additional Densities by Numeric Attribute
  
```

Co-Author Network Analysis



The screenshot shows the Sci2 Tool interface. The 'Data Preparation' menu is open, highlighting 'Detect Duplicate Nodes'. The 'Data Manager' window shows the 'Extracted Co-Authorship Network' selected. A 'Detect Duplicate Nodes' dialog box is open in the foreground, showing the following settings:

Algorithm Name	Date	Time	% Complete
Extract Co-Author ...	04/06/2015	09:10:17 ...	100%
Load...	04/06/2015	09:09:48 ...	100%

One of the challenges of a co-author network is determining if your data set has duplicate names (e.g. John P. Smith and J P Smith).

To detect duplicate nodes, we will want to select the network in the data manager, and then select *Data Preparation -> Detect Duplicate Nodes*.

A pop-up window will appear, for this demo we will keep the input parameters.

Co-Author Network Analysis

Let's look at the output file *Text Log: Noteworthy nodes that will NOT be merged*.

The screenshot shows a software interface with three main windows:

- Left Window:** A list of nodes with their IDs and names. The list includes:
 - 0.9393939971923828 similar: "Anderson, I"
 - 0.9333333373069763 similar: "Tosatti, E"
 - 0.9259259104728699 similar: "Albert, I"
 - 0.9259259104728699 similar: "Albert, R"
 - 0.9259259104728699 similar: "Albert, L"
 - 0.9259259104728699 similar: "Jensen, P"
 - 0.9166666865348816 similar: "Jeong, H"
 - 0.9141414165496826 similar: "Anderson, I"
 - 0.8888888359069824 similar: "Lee, S"
 - 0.8842592239379883 similar: "Jeong, Hw"
 - 0.8796296119689941 similar: "Albert, L"
 - 0.8796296119689941 similar: "Albert, R"
 - 0.8796296119689941 similar: "Albert, I"
 - 0.8796296119689941 similar: "Albet, R"
- Top-Right Window (Data Manager):** A context menu is open over a file named "Text Log: Noteworthy nodes that will NOT be merged". The menu options are: Save, View, View With..., Rename, Discard, Merge tables based on label, Text Log: Nodes that will be merged, and Author information.2.
- Bottom-Right Window (Notepad):** A text file named "Nodes that will be merged2769568192464013332.txt" is open. The content is a merge report:


```
Merge reportSimilarly named entities will be merged into the one with the longest name.
Merge 1 =====Farkas, Ij will have the following merged in:
Farkas, I===== Merge 2 =====Brockman, Jb will have the following merged in:
Brockman, J===== Merge 3 =====Jeong, Hw will have the following merged in:
Jeong, H===== Merge 4 =====Lee, Cs will have the following merged in:
Lee, SLee, C===== Merge 5 =====Pfeifer, Ma will have the following merged in:
Pfeifer, M===== Merge 6 =====Hornbaker, Dj will have the following merged in:
Hornbaker, D===== Merge 7 =====Wasserman, Ss will have the following merged in:
Wasserman, S===== Merge 8 =====Anderson, Cj will have the following merged in:
Anderson, CEnd of merge report.
```

Right click the file in the data manager, and select *view* or *view with...* which will allow you to open the file in a text editor like notepad.

We can repeat this process with the file listing nodes that will be merged.

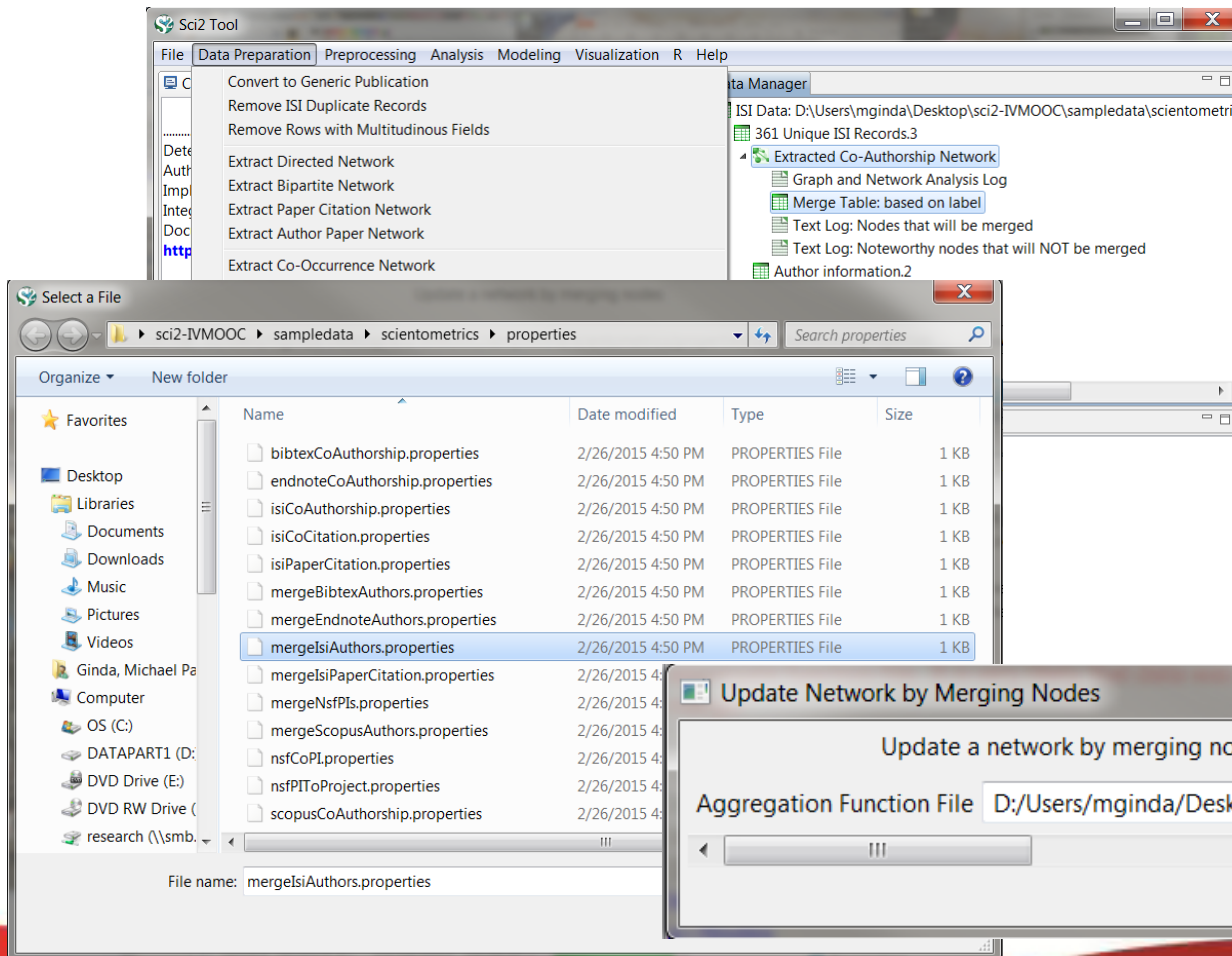
Co-Author Network Analysis

After we've identifying our duplicate nodes, we need to merge these duplicates.

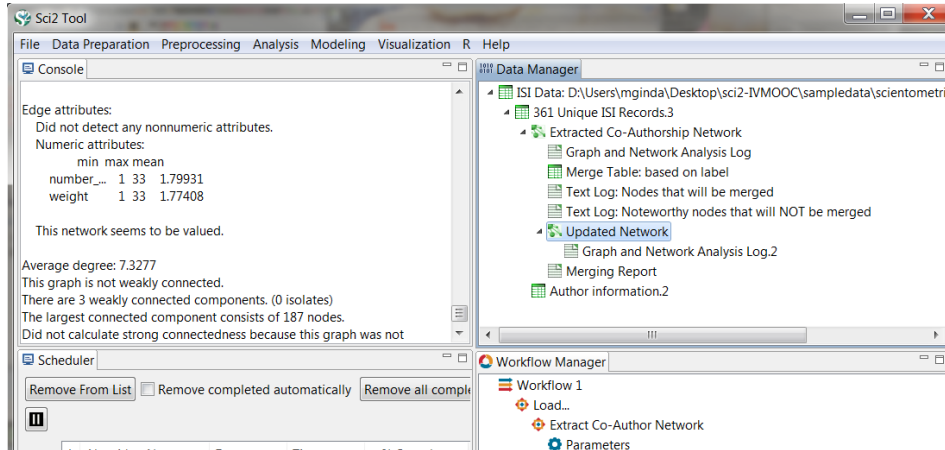
Select the network file and the *Merge Table: based on label* file in the data manager, and then select *Data Preparation -> Update Network by Merging Nodes*

A box will appear that allows us to use an aggregation function file (property files). Select browse, and navigate to the *Sci2 directory sampledata -> scientometrics -> properties* and select *MergelsiAuthors.properties*

Select open, and then OK.



Co-Author Network Analysis



We've now updated our network, so let's re-run the network analysis toolkit algorithm to see how our network has been effected by our work.

What changes do you notice to the network statistics?

This graph claims to be undirected.

Original Network

Nodes: 247
 Isolated nodes: 0
 Node attributes present: label, number_of_authored_works, times_cited

Edges: 891
 No self loops were discovered.
 No parallel edges were discovered.

Edge attributes:
 Did not detect any nonnumeric attributes.
 Numeric attributes:

	min	max	mean
number_...	1	33	1.76094
weight	1	33	1.76094

This network seems to be valued.

Average degree: 7.2146
 This graph is not weakly connected.
 There are 3 weakly connected components. (0 isolates)
 The largest connected component consists of 194 nodes.
 Did not calculate strong connectedness because this graph was not directed.

Density (disregarding weights): 0.0293
 Additional Densities by Numeric Attribute

This graph claims to be undirected.

Revised Network

Nodes: 238
 Isolated nodes: 0
 Node attributes present: label, number_of_authored_works, times_cited

Edges: 872
 No self loops were discovered.
 No parallel edges were discovered.

Edge attributes:
 Did not detect any nonnumeric attributes.
 Numeric attributes:

	min	max	mean
number_...	1	33	1.79931
weight	1	33	1.77408

This network seems to be valued.

Average degree: 7.3277
 This graph is not weakly connected.
 There are 3 weakly connected components. (0 isolates)
 The largest connected component consists of 187 nodes.
 Did not calculate strong connectedness because this graph was not directed.

Density (disregarding weights): 0.0309
 Additional Densities by Numeric Attribute

Co-Author Network Analysis

Let's start to analyze the updated network. To start, let's find the degree for each node.

Select the updated network in the data manager, and then select in the menu *Analysis -> Networks -> Unweighted & Undirected -> Node Degree*

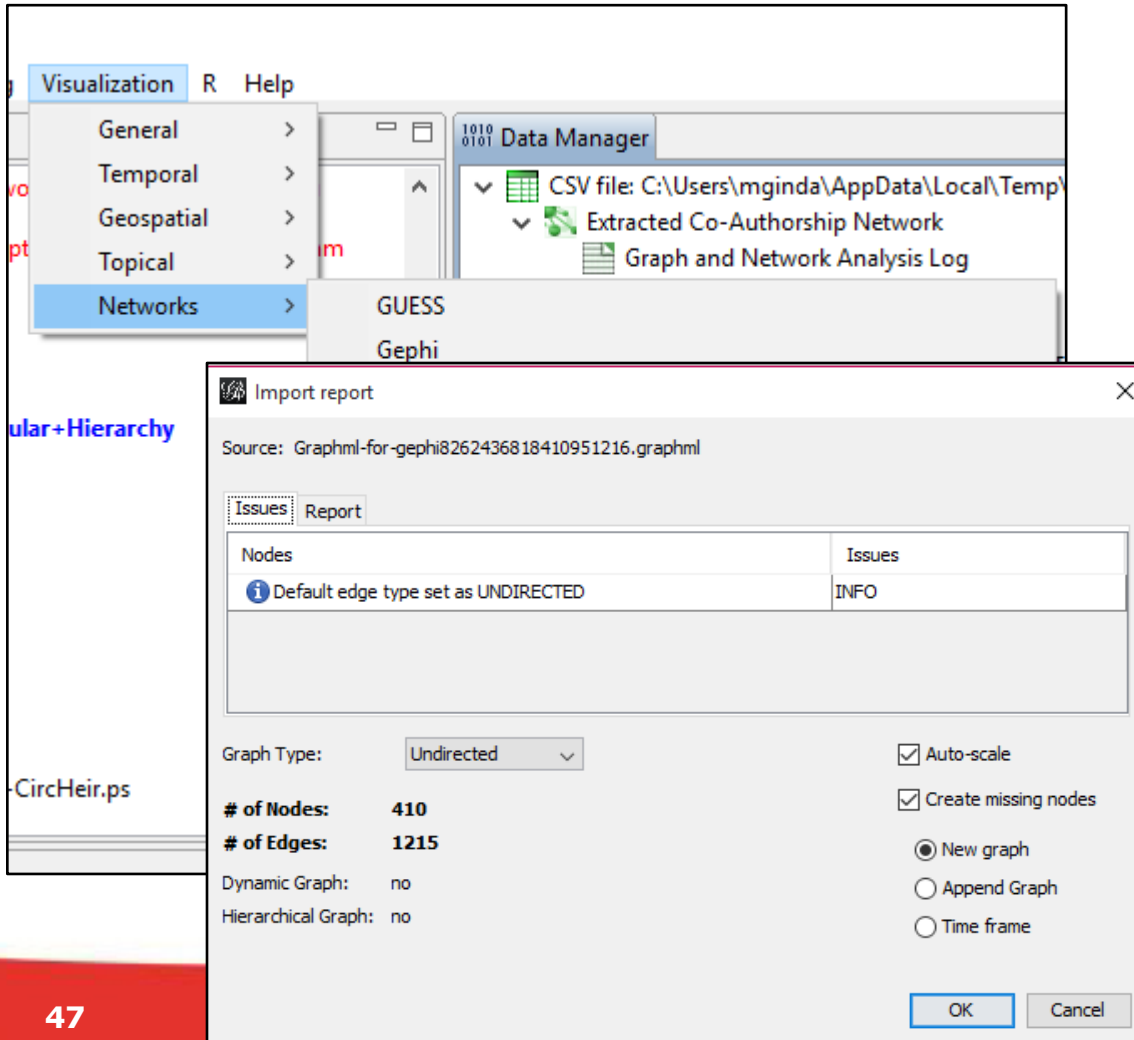
A new network file will be output that has appended a degree to each node in your network file.

To see the distribution of node degrees, use the same menu path above, except you will need to select algorithm *Degree Distribution*. A pop-up window will appear, for now, just hit OK. Two data files will appear, we'll select the first.

To visualize this file, select *Visualization -> General -> GnuPlot*

The screenshot displays the Sci2 Tool interface. The 'Analysis' menu is open, showing the path: *Analysis -> Networks -> Unweighted & Undirected -> Node Degree*. The 'Data Manager' window shows a selected network file. The console window shows the command: `xe -load usegnuplot.txt -persist`. A 'gnuplot graph' window displays a degree distribution plot with red '+' markers. The plot shows a high frequency of nodes with low degrees (around 1-10) and a long tail of nodes with higher degrees (up to 140). The x-axis ranges from 0 to 140, and the y-axis ranges from 0 to 0.2.

Co-Author Network Visualization in Gephi



Next we will visualize the network in Gephi.

Navigate to **Visualization > Networks > Gephi**.

The algorithm is a bridge that passes the network data to Gephi. The program will automatically start. The tool produces an Import Report. It lets you select the network type, gives load errors, etc.

Next, is a brief walk through of Gephi's three main sections, and outline various functions and tools available.

Gephi: Initial Layout of Network

The screenshot shows the Gephi 0.8.2 interface with two configuration panels open. The left panel is for 'ForceAtlas 2' and the right panel is for 'YifanHu's Multilevel'.

ForceAtlas 2 Configuration:

- Threads: 2
- Behavior Alternatives:
 - Dissuade Hubs:
 - LinLog mode:
 - Prevent Overlap:
 - Edge Weight Influence: 1.0
- Tuning:
 - Scaling: 30.0
 - Stronger Gravity:
 - Gravity: 0.08
- Performance:
 - Tolerance (speed): 0.1
 - Approximate Repulsion:
 - Approximation: 1.2

YifanHu's Multilevel Configuration:

- Barnes-Hut's properties:
 - Quadtree Max Level: 10
 - Theta: 1.2
- Multi-level:
 - Minimum level size: 3
 - Minimum coarsening rate: 0.75
- Yifan Hu's properties:
 - Step ratio: 0.97
 - Optimal Distance: 1.0 (with a sub-field for 0.97)

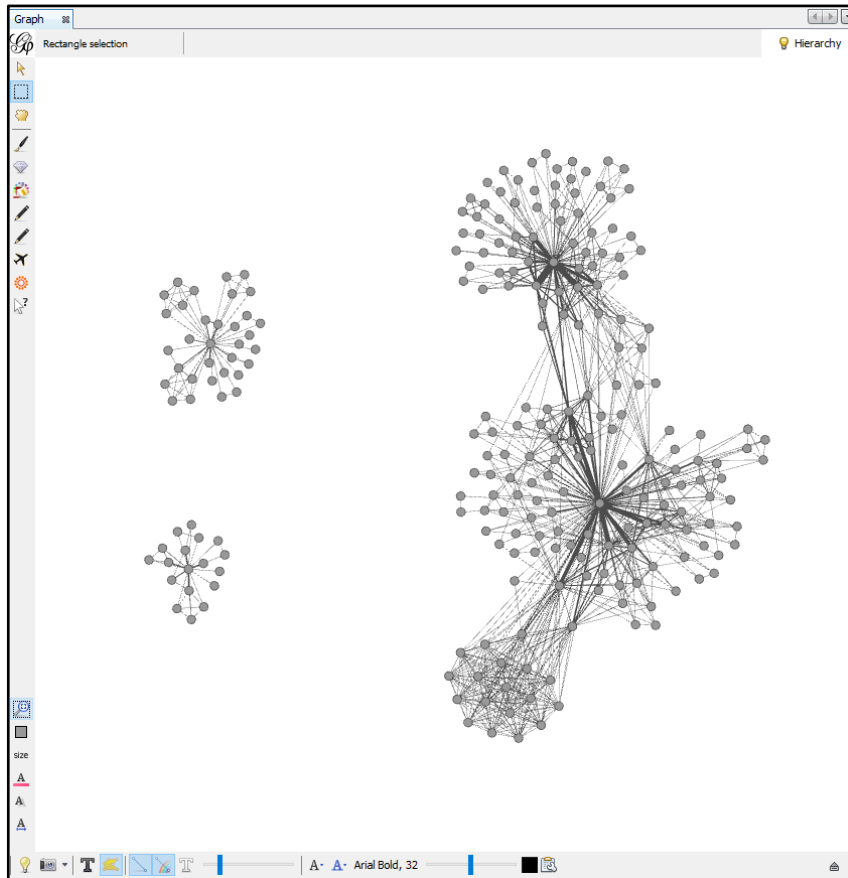
Now we can start an visualization and analysis of the network.

First we will adjust the layout of the network.

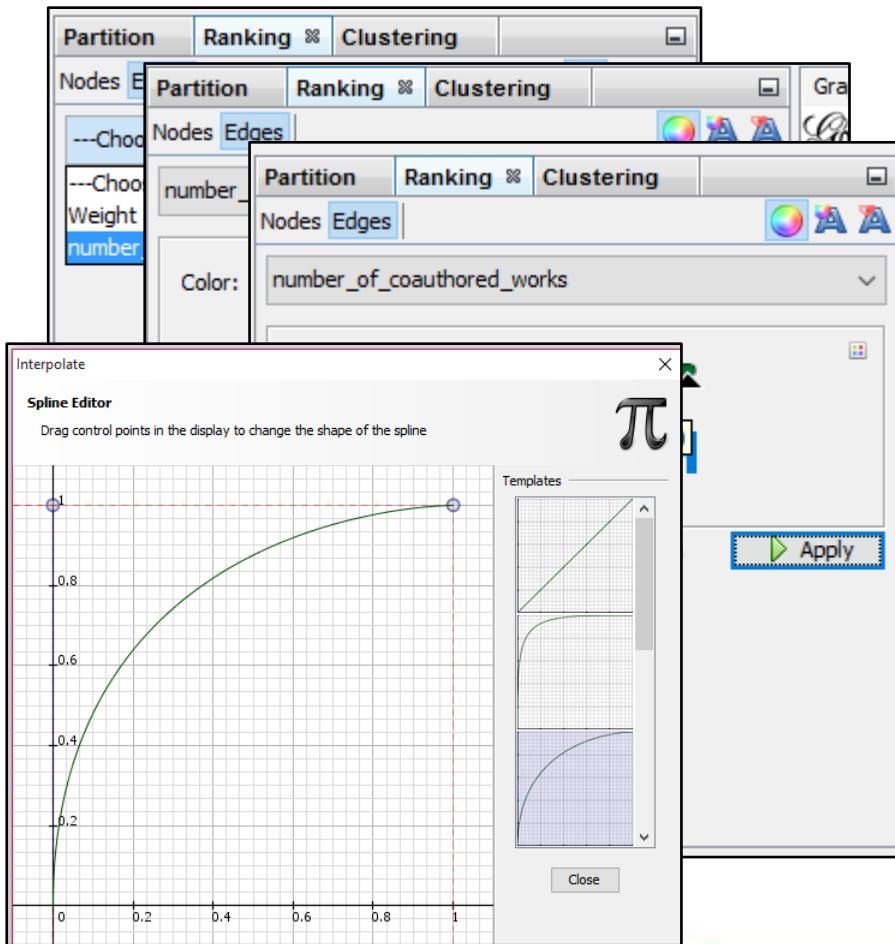
From the layout pane, select the "ForceAtlas2" layout algorithm and enter the following parameters, and then select "Run".

You may also select YifanHu's Multilevel force network layout.

Gephi: Initial Layout of Network



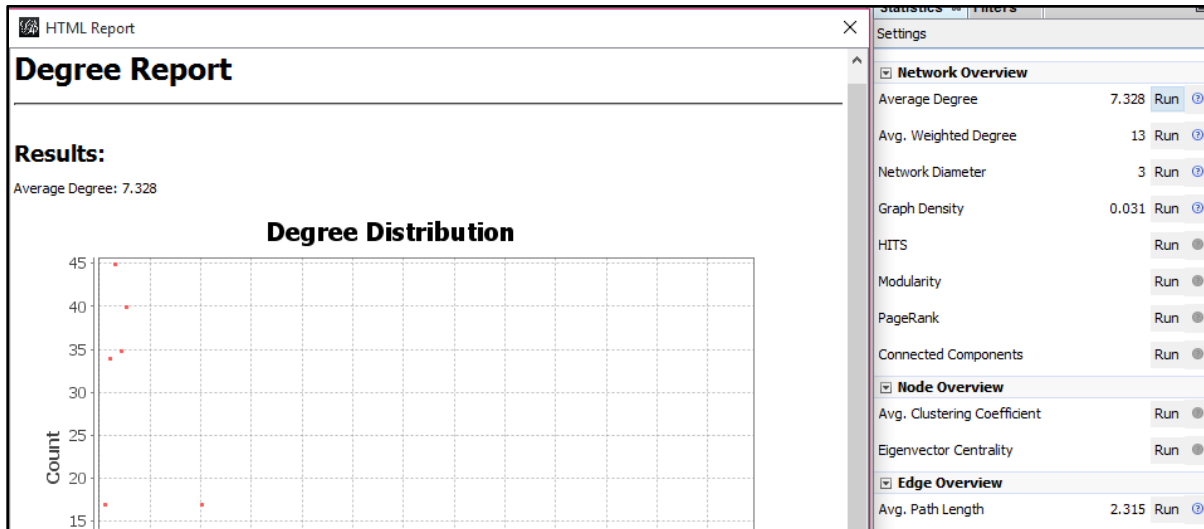
Gephi: Edge Color



Next, we can adjust the edge color by selecting the Edge tab in the Ranking window.

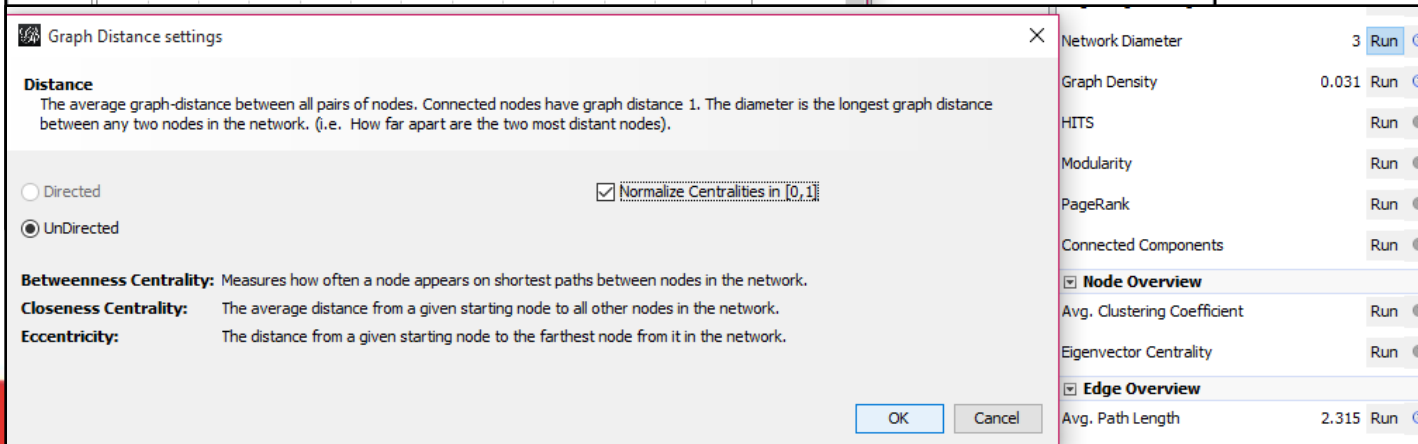
- ▶ In the drop down menu, select "number_of_coauthored_works".
- ▶ Select the small square in the right corner of the Color Range box. This lets us choose new color ranges for variables.
- ▶ You may also set the color range and values to apply the colors to, or adjust color scaling variables by adjusting the spline.

Gephi: Network statistics



Gephi provides a variety of node and edge statistics to help understand the relationships, clustering, paths, centrality, and communities within a network.

Try implementing the Average Degree, and Network Diameter statistics, which we will next visualize.



Graph Distance settings

Distance
The average graph-distance between all pairs of nodes. Connected nodes have graph distance 1. The diameter is the longest graph distance between any two nodes in the network. (i.e. How far apart are the two most distant nodes).

Directed
 UnDirected

Normalize Centralities in [0, 1]

Betweenness Centrality: Measures how often a node appears on shortest paths between nodes in the network.
Closeness Centrality: The average distance from a given starting node to all other nodes in the network.
Eccentricity: The distance from a given starting node to the farthest node from it in the network.

OK Cancel

Network Diameter 3 Run

Graph Density 0.031 Run

HITS Run

Modularity Run

PageRank Run

Connected Components Run

Node Overview

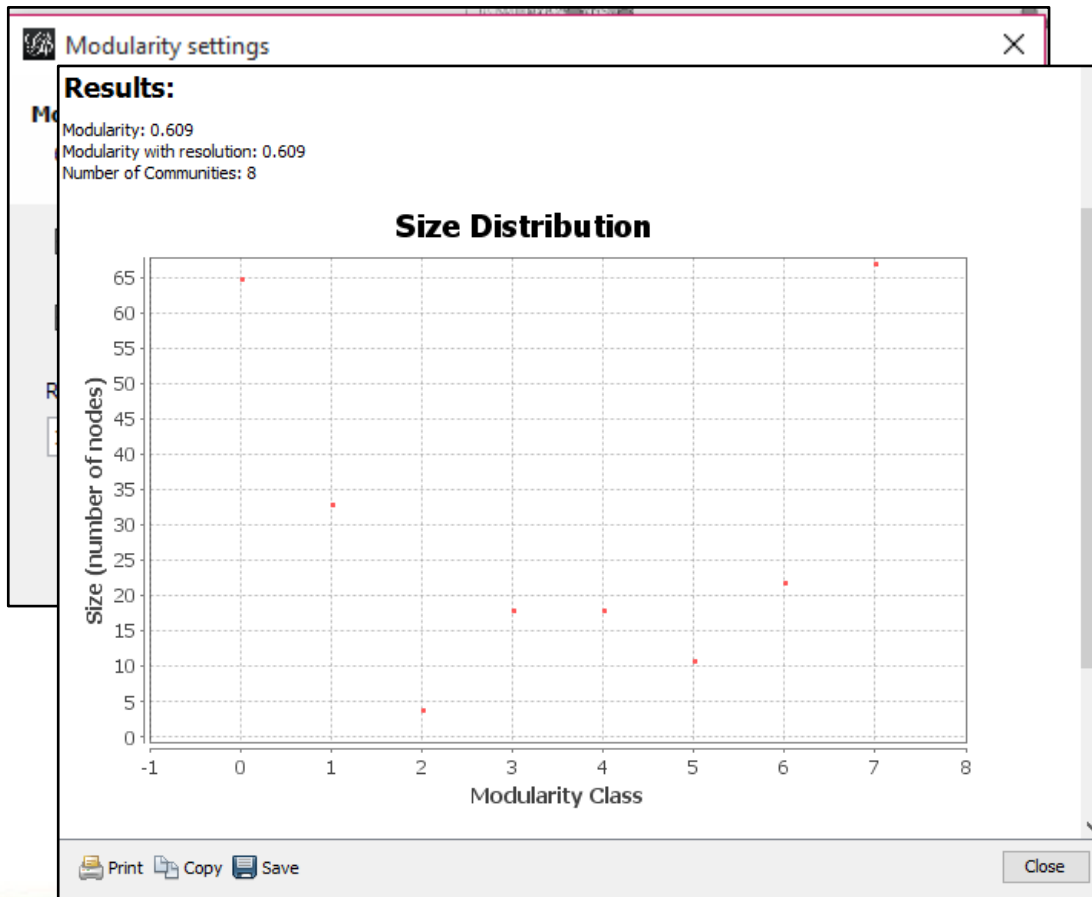
Avg. Clustering Coefficient Run

Eigenvector Centrality Run

Edge Overview

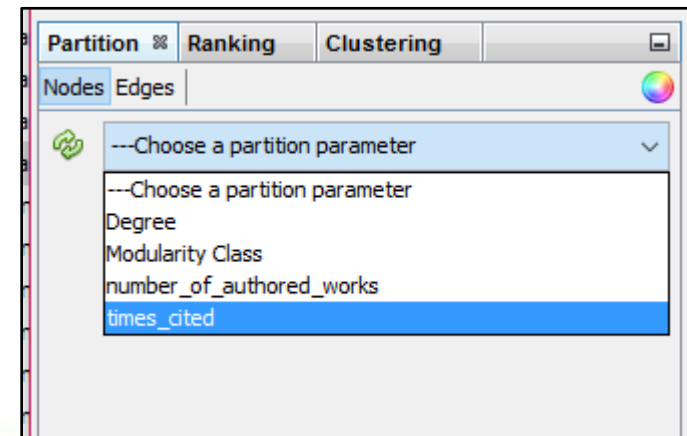
Avg. Path Length 2.315 Run

Gephi: Network statistics/Community Detection

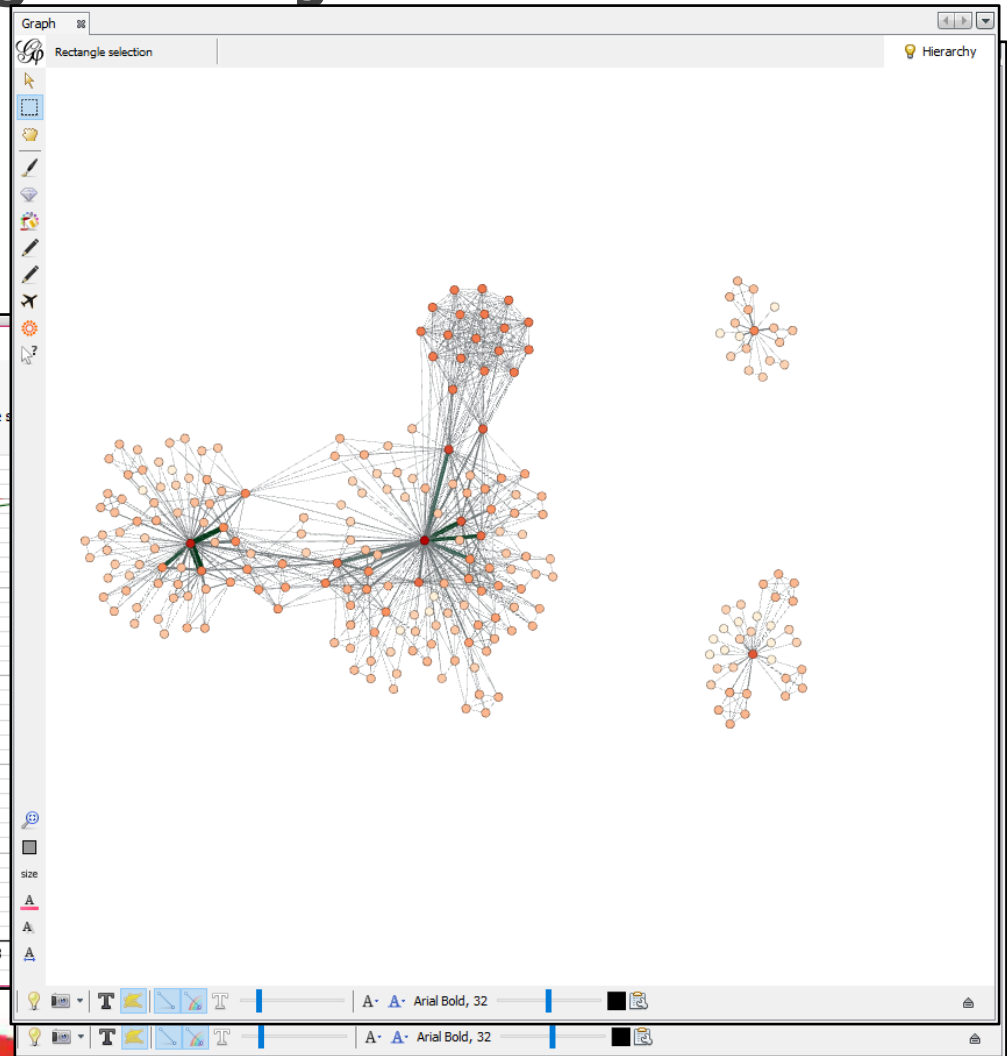
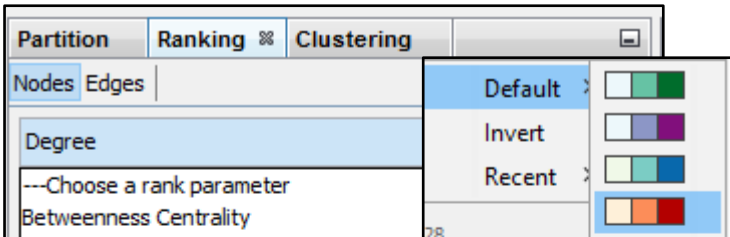


The modularity statistical algorithm calculates how the connectedness of a network, and the Blondel Communities that exist in the network. The communities are added as a partition to the nodes.

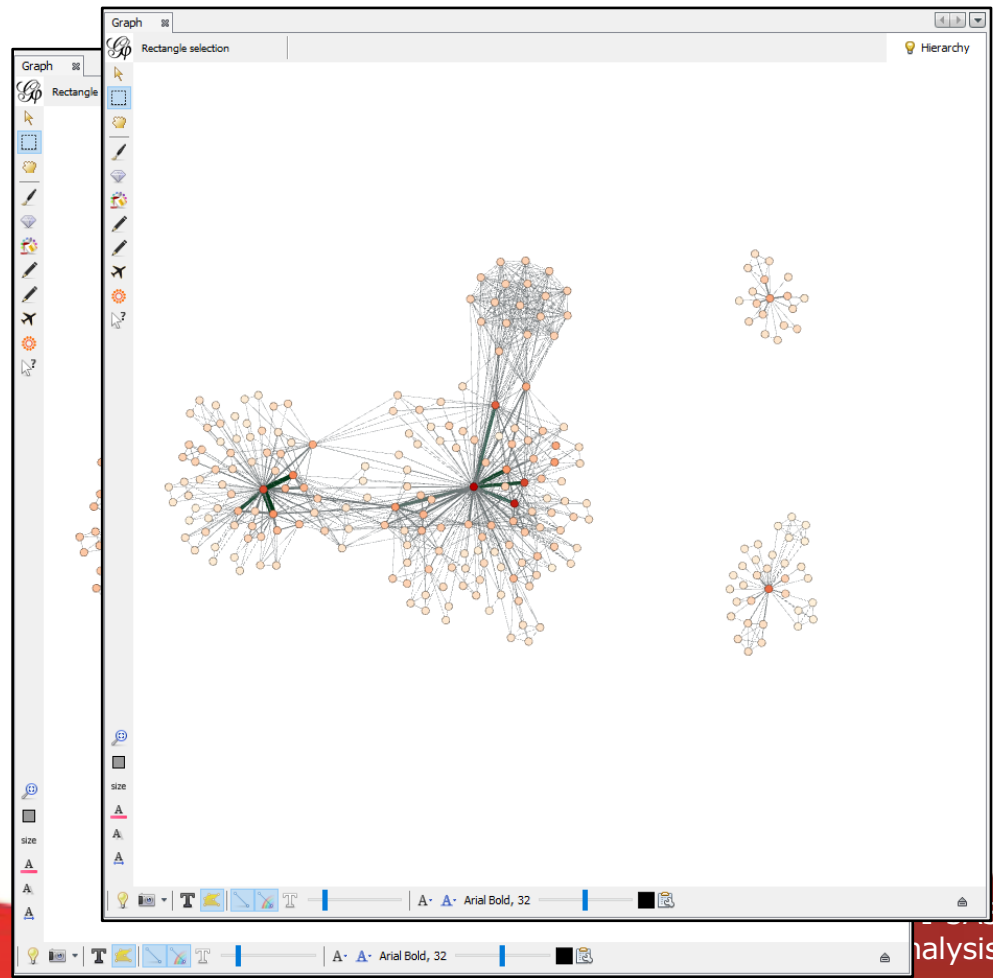
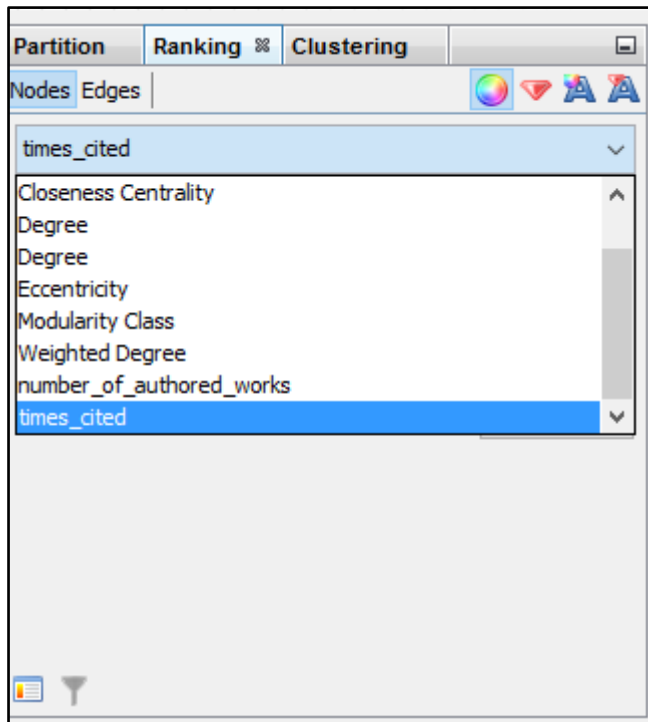
The modularity categories may be applied to the network from the Partitions window.



Gephi: Node Color Ranking & Scaling – Degree & Times Cited



Gephi: Node Color Ranking & Scaling – Betweenness Centrality



Gephi: Node Labels – Data Laboratory

The screenshot shows the Gephi 0.8.2 interface with the Data Laboratory window open. The main data table contains the following data:

Nodes	Id	Label	number_of_authored...	times_cited	blondel_community_level_0	blondel_community_level_1
● Finelli, Lyn	n1	Finelli, Lyn	9		1389 community_0	community_0
● Blanton, Lenee	n3	Blanton, Lenee	4		48 community_0	community_0
● Brammer, Lynnette	n4	Brammer, Lynnette				
● Smith, Sophie	n5	Smith, Sophie				

The 'Add column - Settings' dialog box is open, showing the following configuration:

- Add column to nodes**
- Title:** Label 2
- Type:** String

Below the dialog box, a portion of the data table is visible, showing columns for 'Betweenness C...', 'Modularity Cl...', and 'Label2':

	Betweenness C...	Modularity Cl...	Label2
1	0.004	3	Wasserman, Ss
7	0.007	0	Vicsek, T
8	0.267	7	Vespignani, A
4	0.02	7	Vazquez, A
3	0.02	6	Stanley, He
1	0.006	7	Pastor-satorras, R
9	0.013	0	Oltvai, Zn
2	0.009	7	Munoz, Ma
5	0.018	0	Kahng, B
1	0.017	1	Garfield, E
2	0.007	7	Barrat, A
2	0.434	0	Barabasi, Al
8	0	1	

The bottom toolbar contains various actions: Add column, Merge columns, Delete column, Clear column, Copy data to other column, Fill column with a value, Duplicate column, Create a boolean column from regex match, Create column with list of regex matching groups, Negate boolean values, and Convert column to dynamic.

Gephi: Node Labels – Overview

The image displays the Gephi software interface. The main window shows a network graph with nodes and edges. The 'Labels' tab is selected in the bottom toolbar. The 'Label text settings' dialog box is open, showing the 'Nodes' tab. The 'Label2' attribute is selected for display as a label. The 'Global' settings for nodes and edges are also visible.

Label text settings

Nodes Edges Show properties

Select attributes to display as labels

- Id
- Label
- number_of_authored_works
- times_cited
- Degree
- Weighted Degree
- Eccentricity
- Closeness Centrality
- Betweenness Centrality
- Modularity Class
- Label2

OK Cancel

Configure...

Global Nodes Edges Labels

Node Edge

Font: Arial Bold, 9 Color:

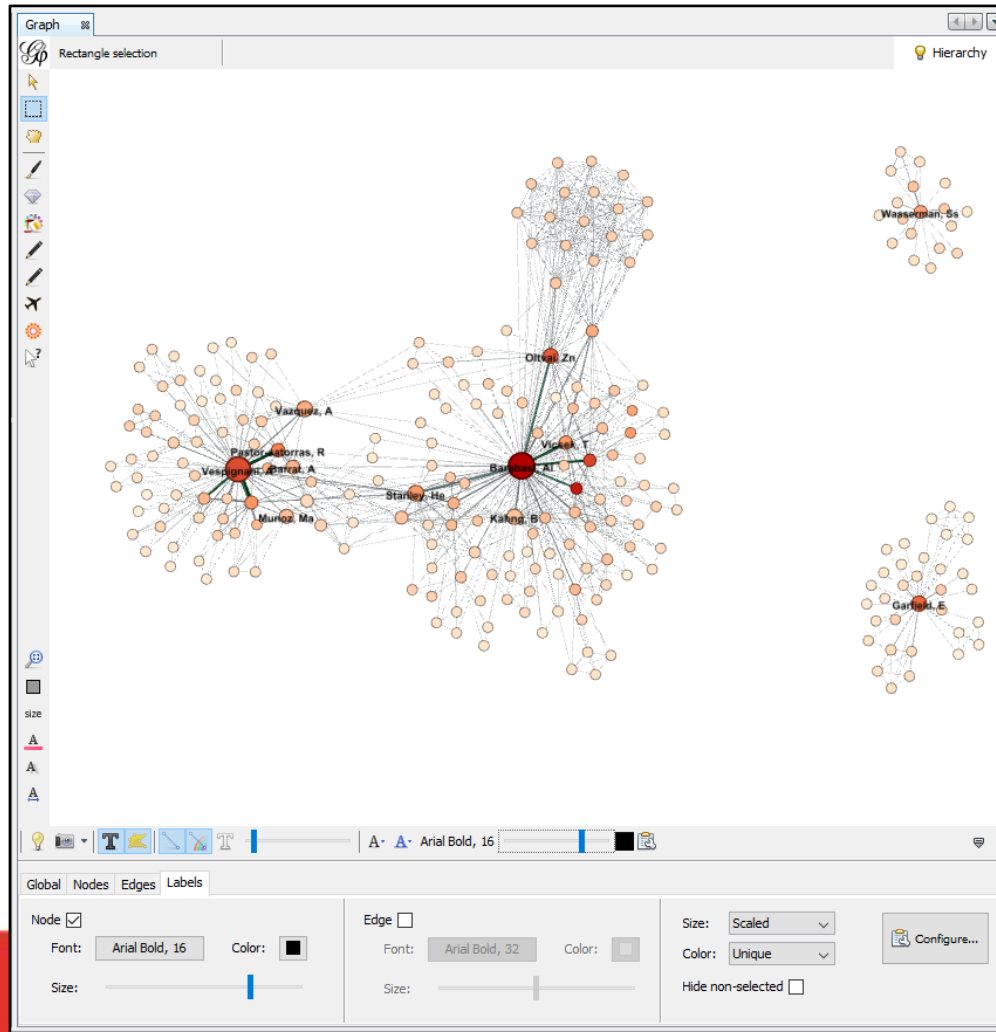
Size:

Font: Arial Bold, 32 Color:

Size:

Size: Scaled Color: Unique Hide non-selected

Gephi: Node Labels – Overview



Gephi: Final visualization – Node Parameters

The screenshot displays the Gephi 0.8.2 interface with the 'Nodes' panel open. The 'Node Labels' section is expanded, showing various parameters for node labels. A 'Preview Settings - Border Color' dialog box is overlaid on the right, showing options for 'Parent' and 'Custom' border colors. A blue arrow points from the 'Custom' radio button to a black color swatch. The background shows a network graph visualization with orange nodes and edges.

Category	Parameter	Value	
Nodes	Border Width	0.025	
	Border Color	custom [0,0,0]	
	opacity	100.0	
Node Labels	Show Labels	<input checked="" type="checkbox"/>	
	Font	Arial 10 Plain	
	Proportional size	<input checked="" type="checkbox"/>	
	Color	custom [0,0,0]	
	Shorten label	<input type="checkbox"/>	
	Max characters	30	
	Outline size	0.0	
	Outline color	custom [255,255,255]	
	Outline opacity	80.0	
	Box	<input type="checkbox"/>	
	Box color	parent	
	Box opacity	100.0	
	Edges	Show Edges	<input checked="" type="checkbox"/>
Thickness		1.0	
Rescale weight		<input type="checkbox"/>	
Color		mixed	
Opacity		100.0	
Curved		<input checked="" type="checkbox"/>	
Radius		0.0	
Edge Arrows		Size	3.0
		Show Labels	<input type="checkbox"/>
Edge Labels		Show Labels	<input type="checkbox"/>
	Font	Arial 10 Plain	
	Color	original	
	Shorten label	<input type="checkbox"/>	
	Max characters	30	
	Outline size	0.0	
	Outline color	custom [255,255,255]	
Outline opacity	80.0		

Gephi: Final visualization – Node Parameters

The screenshot displays the Gephi 0.8.2 interface with the 'Node Labels' and 'Preview Settings - Font' dialog boxes open.

Node Labels Panel:

- Show Labels:
- Font: defaultStyle
- Proportional size:
- Color: custom [0,0,0]
- Shorten label:
- Max characters: 30
- Outline size: 0.0
- Outline color: custom [255,255,255]
- Outline opacity: 80.0
- Box:
- Box color: parent
- Box opacity: 100.0

Preview Settings - Font Dialog:

- Font: Arial
- Font Style: Plain
- Size: 10
- Preview text: The quick brown fox jumps over the lazy dog
- Buttons: OK, Cancel

Bottom Panel:

- Preview ratio: 100%
- Export: SVG/PDF/PNG
- Refresh button

Gephi: Final visualization – Edge Color Parameters

The screenshot displays the Gephi 0.8.2 interface. The 'Preview Settings' panel is open, showing the 'Edges' section. The 'Preview Settings - Color' dialog box is overlaid on top, allowing for the configuration of edge colors. The dialog includes the following options:

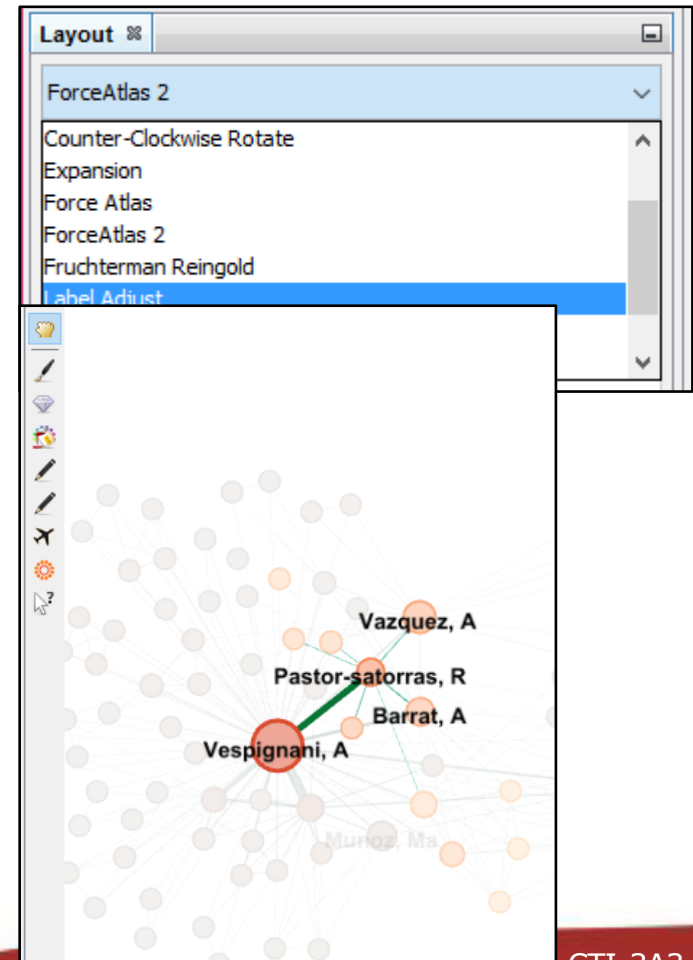
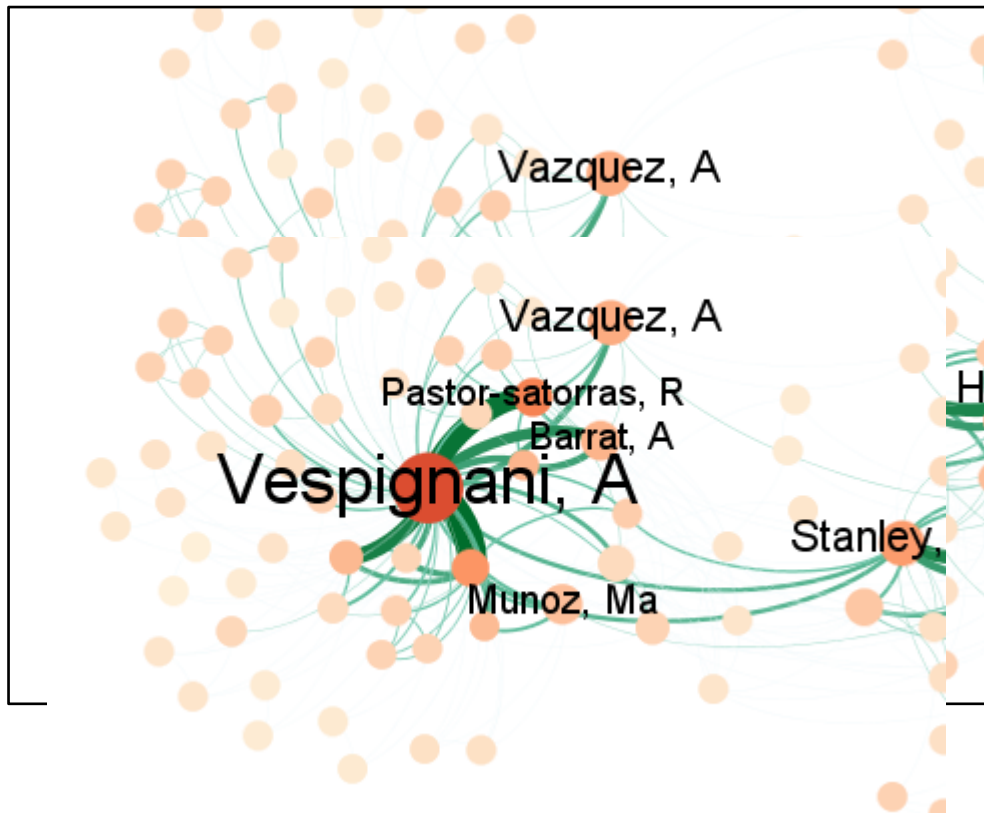
- Edge Color:** Configures the color of the edges. Edges can either a color on their own (original) or uses incident nodes color.
- Original
- Mixed
- Source
- Target
- Custom

The 'Edges' section in the 'Preview Settings' panel shows the following parameters:

- Show Edges:
- Thickness: 20.0
- Rescale weight:
- Color: original
- Opacity: 100.0
- Curved:
- Radius: 0.0
- Edge Arrows: Size: 3.0
- Edge Labels: Show Labels: ; Font: Arial 10 Plain; Color: original; Shorten label: ; Max characters: 30; Outline size: 0.0

The network visualization in the background shows a complex graph with nodes and edges, some of which are highlighted in red. The 'Preview ratio' is set to 100%.

Gephi: Label Adjustments



Gephi: Exporting the visualization and networks

The screenshot displays the Gephi 0.8.2 interface with a network visualization of orange nodes and edges. The 'Preview Settings' panel on the left is open, showing various configuration options for nodes and edges. An 'Export' dialog box is overlaid on the right, showing the 'Results' folder as the save location. The dialog lists two PDF files: 'CDC-Wos-CoAuth-Blondel-CircHeir.pdf' and 'CDC-Wos-CoAuth-Force2-PubsCitations.pdf'. The 'File name' field contains '4NetSci2Researchers' and the 'Files of type' dropdown is set to 'PDF Files (*.pdf)'. The 'Export' button is highlighted in blue.

Preview Settings

- Nodes**
 - Border Width: 0.025
 - Border Color: parent
 - opacity: 100.0
- Node Labels**
 - Show Labels:
 - Font: Arial 10 Plain
 - Proportional size:
 - Color: custom [0,0,0]
 - Shorten label:
 - Max characters: 30
 - Outline size: 0.0
 - Outline color: custom [255,255,255]
 - Outline opacity: 80.0
 - Box:
 - Box color: parent
 - Box opacity: 100.0
- Edges**
 - Show Edges:
 - Thickness: 20.0
 - Rescale weight:
 - Color: original
 - Opacity: 100.0
 - Curved:
 - Radius: 0.0
 - Edge Arrows**
 - Size: 3.0
 - Edge Labels**
 - Show Labels:
 - Font: Arial 10 Plain
 - Color: original
 - Shorten label:
 - Max characters: 30
 - Outline size: 0.0
 - Outline color: custom [255,255,255]
 - Outline opacity: 80.0

Export Dialog

Save in: Results

- CDC-Wos-CoAuth-Blondel-CircHeir.pdf
- CDC-Wos-CoAuth-Force2-PubsCitations.pdf

File name: 4NetSci2Researchers

Files of type: PDF Files (*.pdf)

Export: SVG/PDF/PNG

Gephi: Exporting the visualization and networks

The screenshot displays the Gephi 0.8.2 interface. The main window shows a network visualization with nodes and edges. A 'Preview' window is open, showing a zoomed-in view of the network. The 'File' menu is open, highlighting the 'Export' option. The 'Export' dialog box is also open, showing the 'Save in' location as 'Results' and the 'File name' as 'FourNetSci2.graphml'. The 'Files of type' dropdown is set to 'GraphML Files (*.graphml)'. The 'Graph' options are set to 'Full'.

File Menu:

- New Project (Ctrl+Shift+N)
- Open... (Ctrl+O)
- Open Recent...
- Close Project
- Properties...
- Import Spigot...
- Import Database
- Generate
- Save (Ctrl+S)
- Save As...
- Export**
 - Graph file...
 - SVG/PDF/PNG file...
- Exit

Export Dialog:

- Save in: Results
- File name: FourNetSci2.graphml
- Files of type: GraphML Files (*.graphml)
- Graph: Full The complete graph Visible only Only visible nodes and edges

Let's practice in Gephi

References

- ▶ <https://gephi.org/users/>



THANK YOU