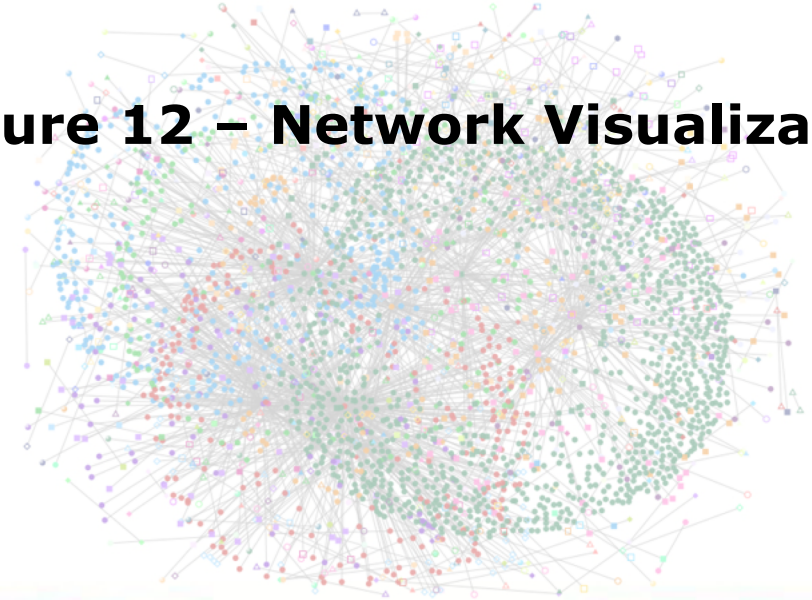
Fakultas Informatika
School of Computing
Telkom University

CTI-3A3
Applied Social Network Analysis

Dr. Warih Maharani

wmaharani.staff.telkomuniversity.ac.id



Lecture 12 – Network Visualization

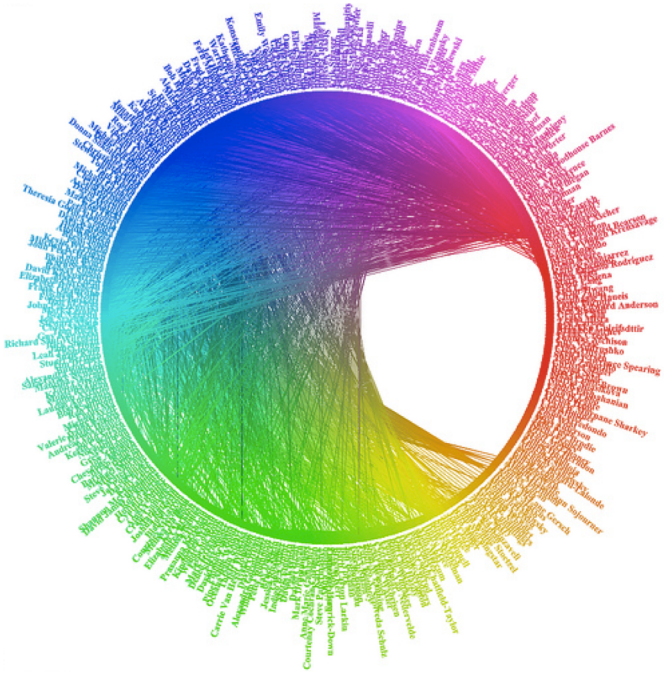
Outline

- Reviews: Network
- Network Visualization
- Gephi

Course Learning Outcomes (CLO)

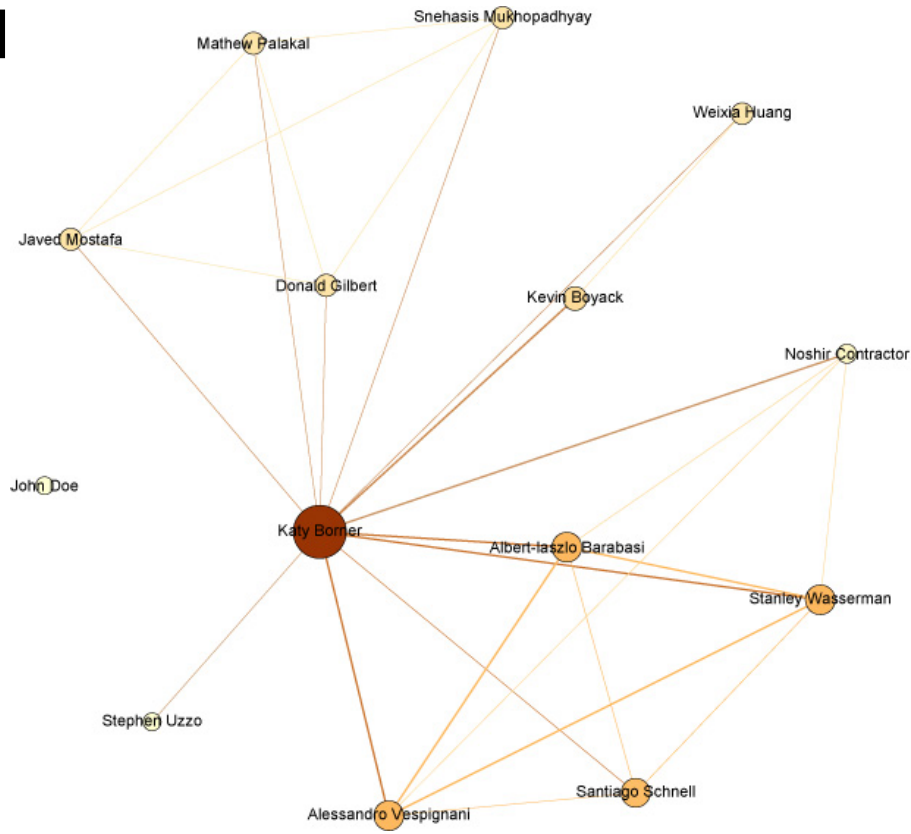
- use software to analyze and visualize network formation and evolution;
- use software to simulate the dynamics of networks based on social network models.

Reviews: Network



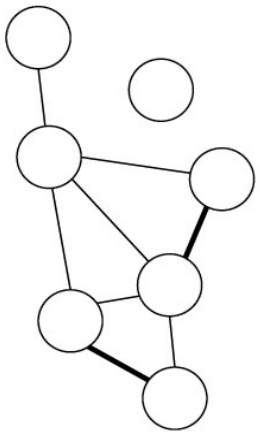
What is a Network?

- 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:

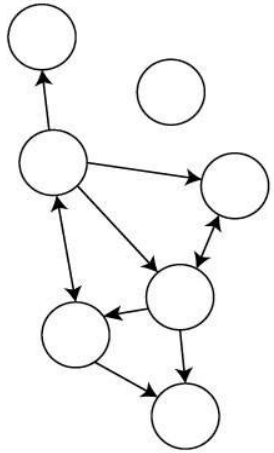
Edges:

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

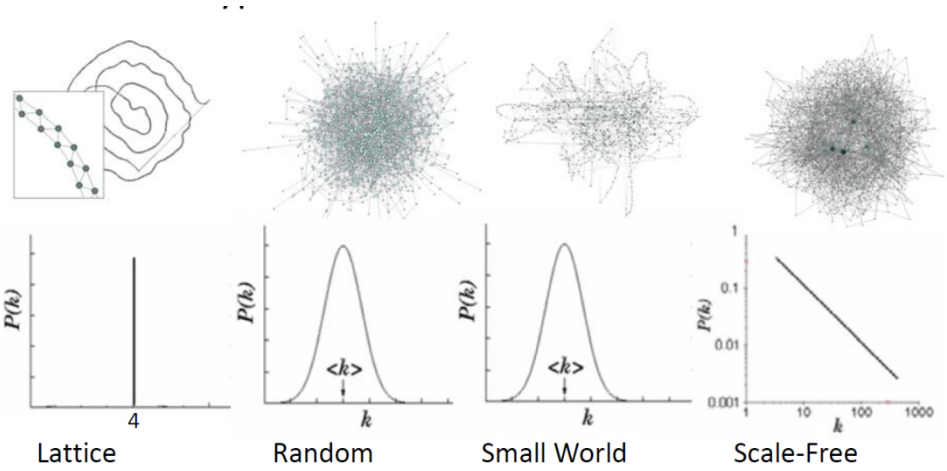
Out-Degree:
Number of outgoing edges

Other types of networks and graphs:

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

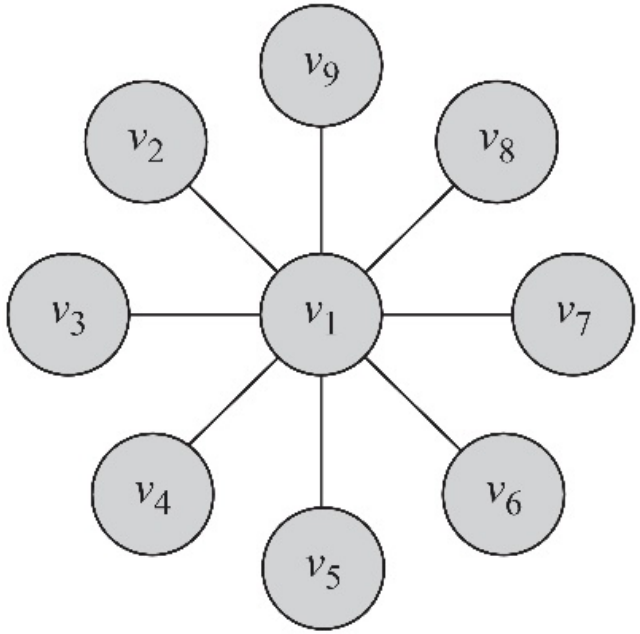
Graph Features

- **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

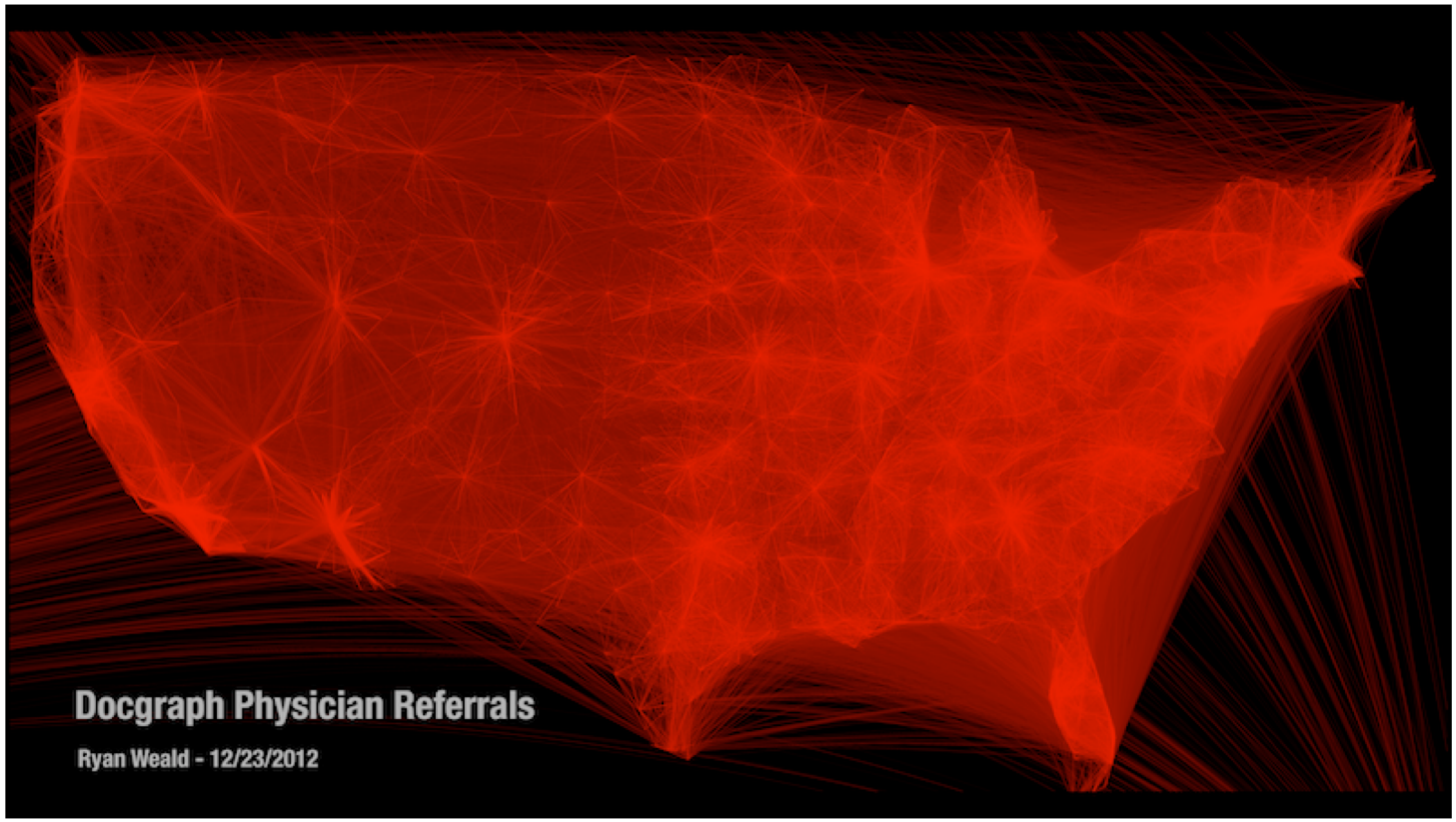


Centralities

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

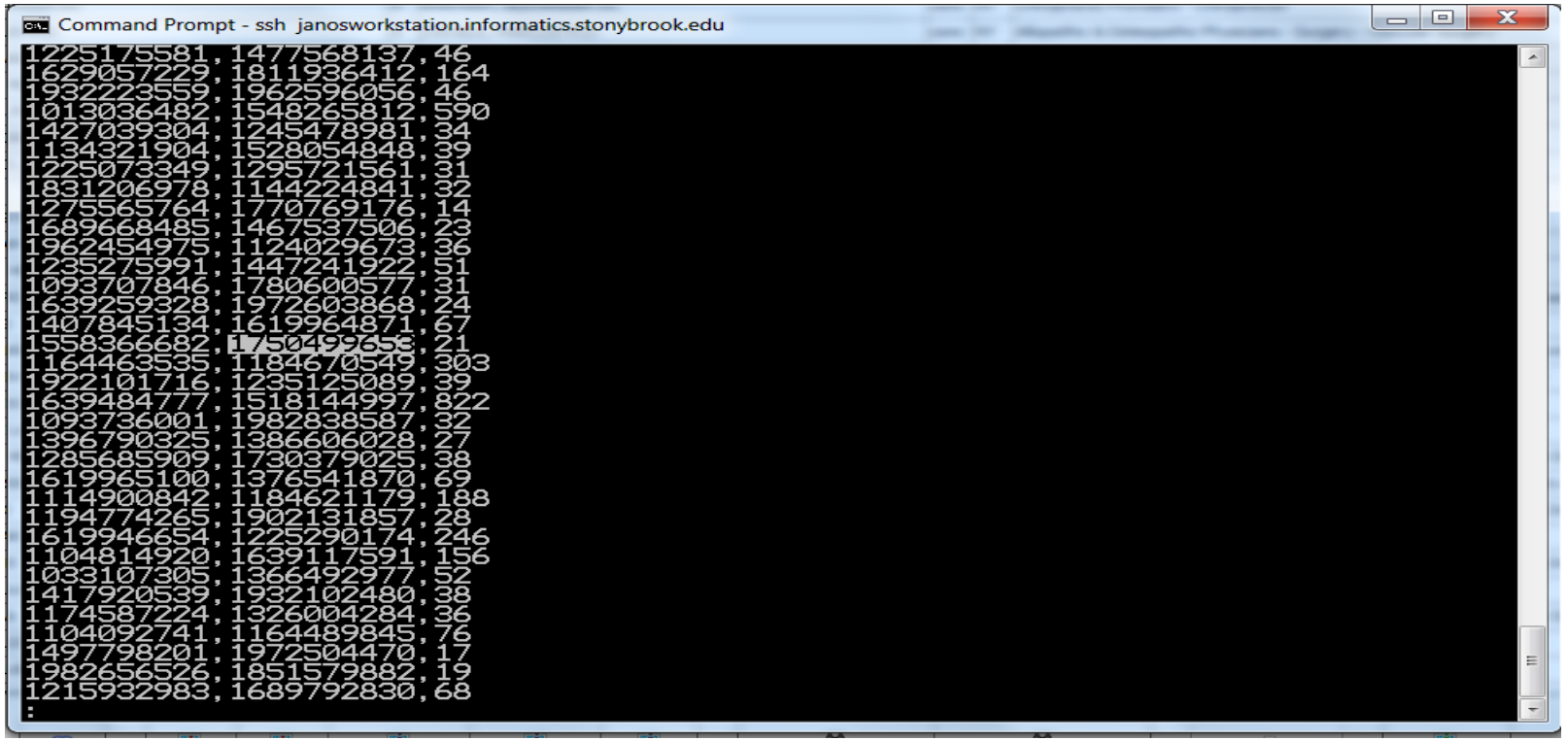


Geographic visualization

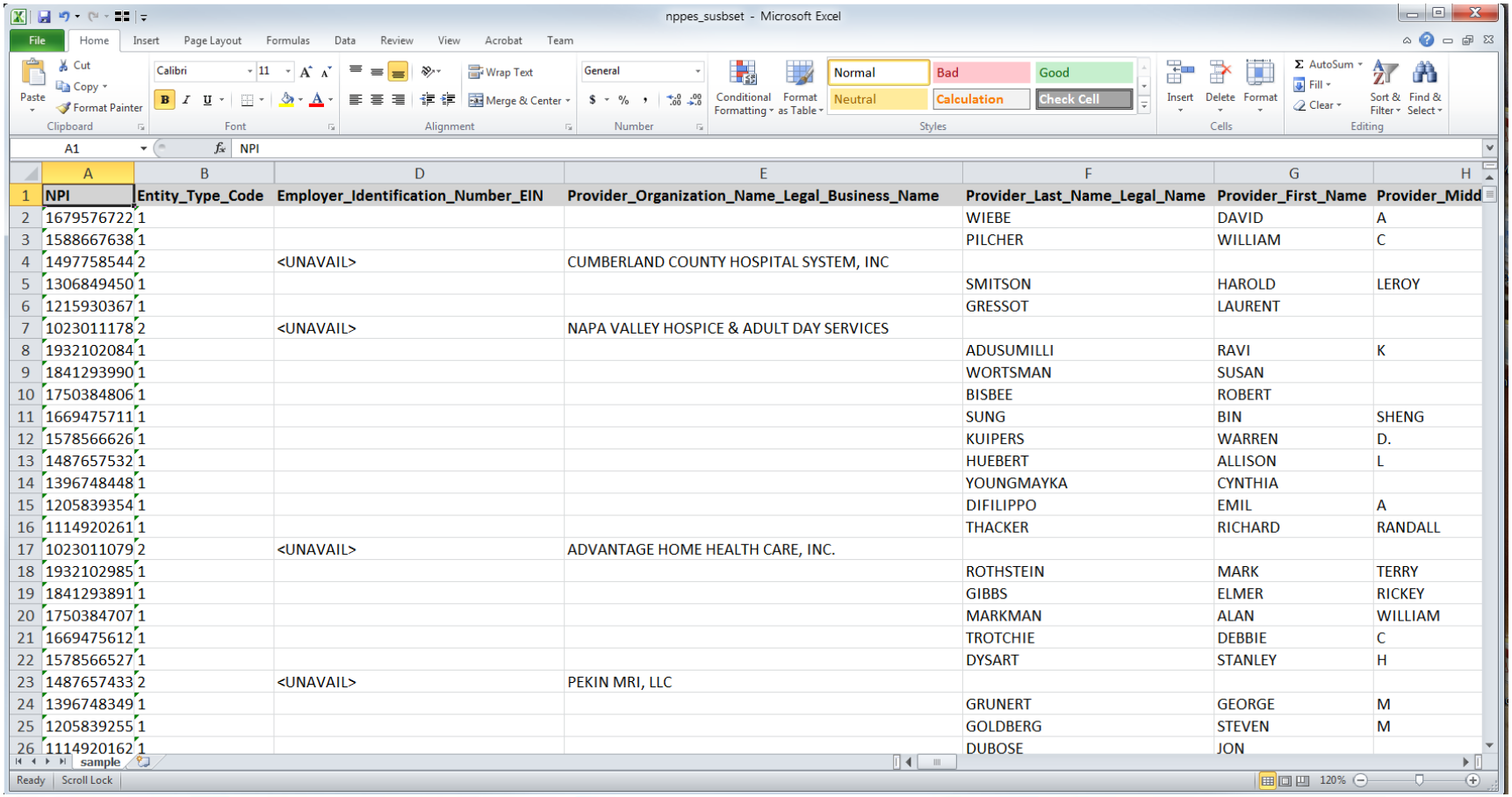


http://isurfsoftware.com/blog/2012/12/13/visualizing-geographic-connections-between-us-doctors/

DocGraph data



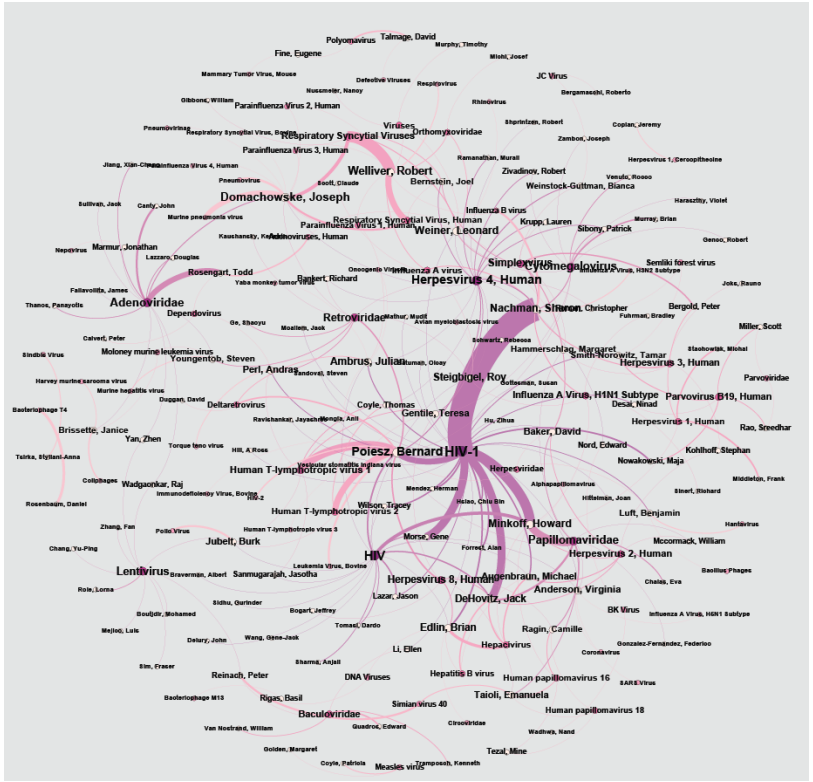
Tabular data

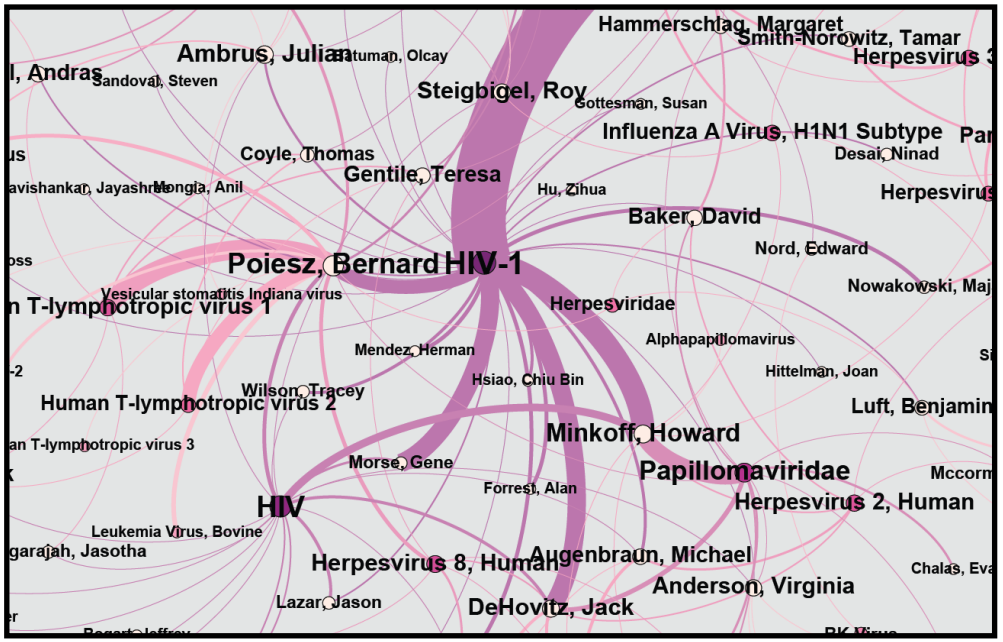


	A	B	D	E	F	G	H
1	NPI	Entity_Type_Code	Employer_Identification_Number_EIN	Provider_Organization_Name_Legal_Business_Name	Provider_Last_Name_Legal_Name	Provider_First_Name	Provider_Midd
2	1679576722	1			WIEBE	DAVID	A
3	1588667638	1			PILCHER	WILLIAM	C
4	1497758544	2	<UNAVAIL>	CUMBERLAND COUNTY HOSPITAL 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

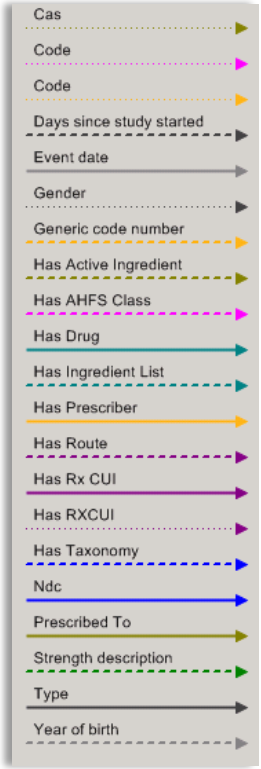
- Relation between authors and MeSH terms from PubMed

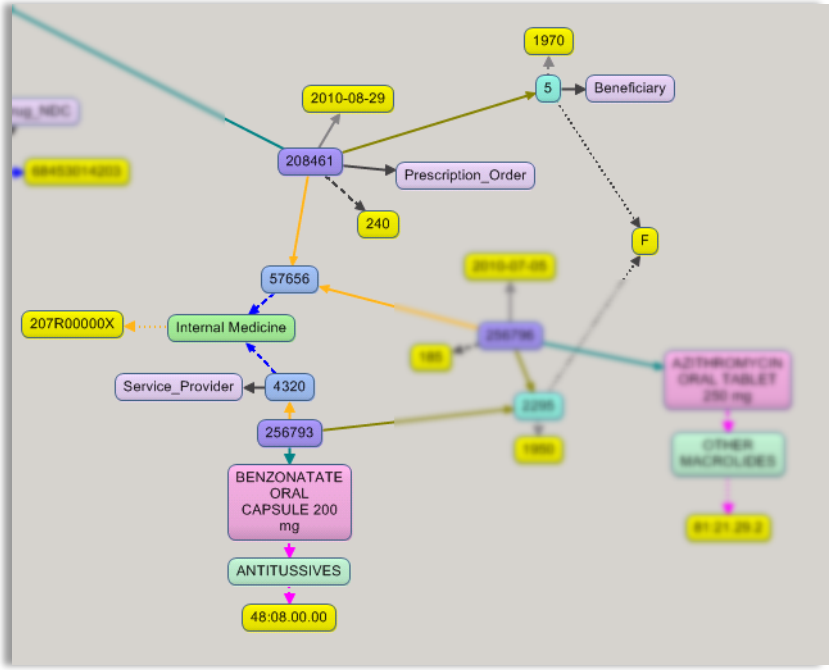




Graphs in healthcare

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

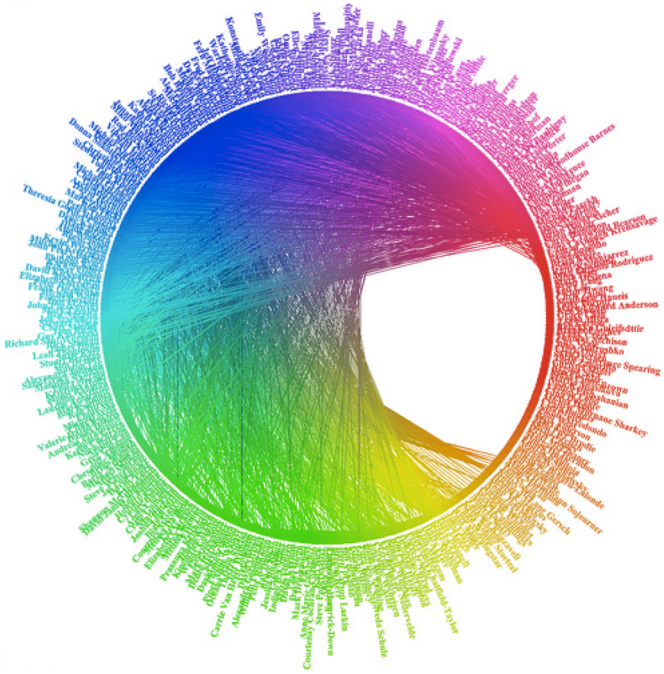




Generating GraphML

- 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

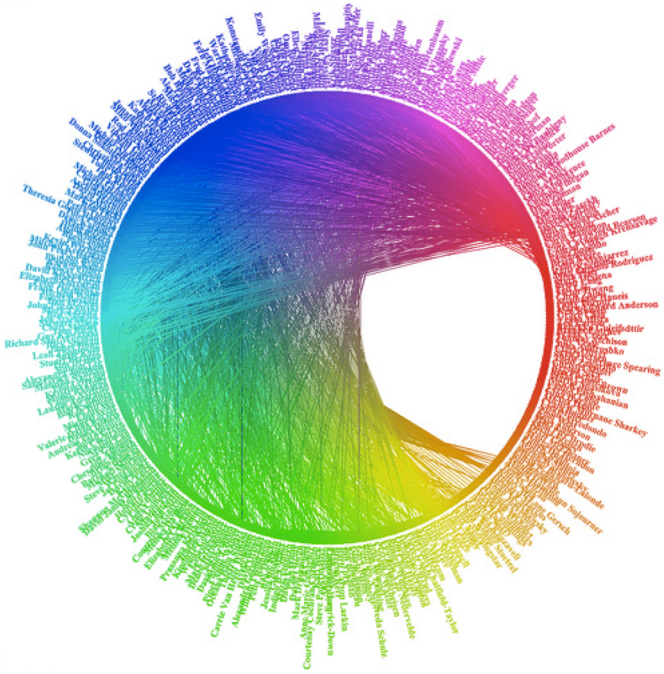
Network Tools

- **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

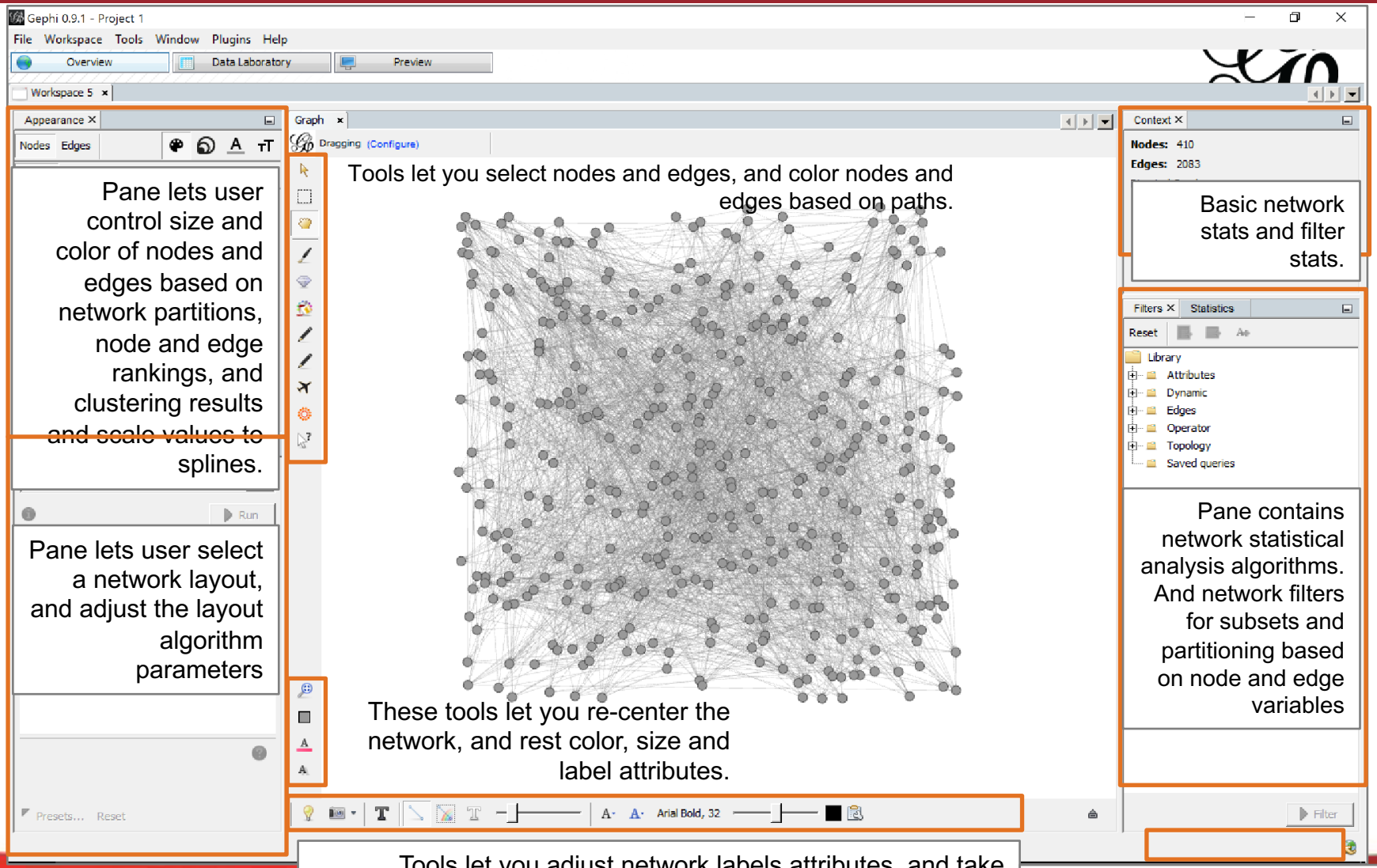
Gephi tool layout and visualization framework:
user is free to rearrange the environment, move panels, show/hide windows, etc.

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

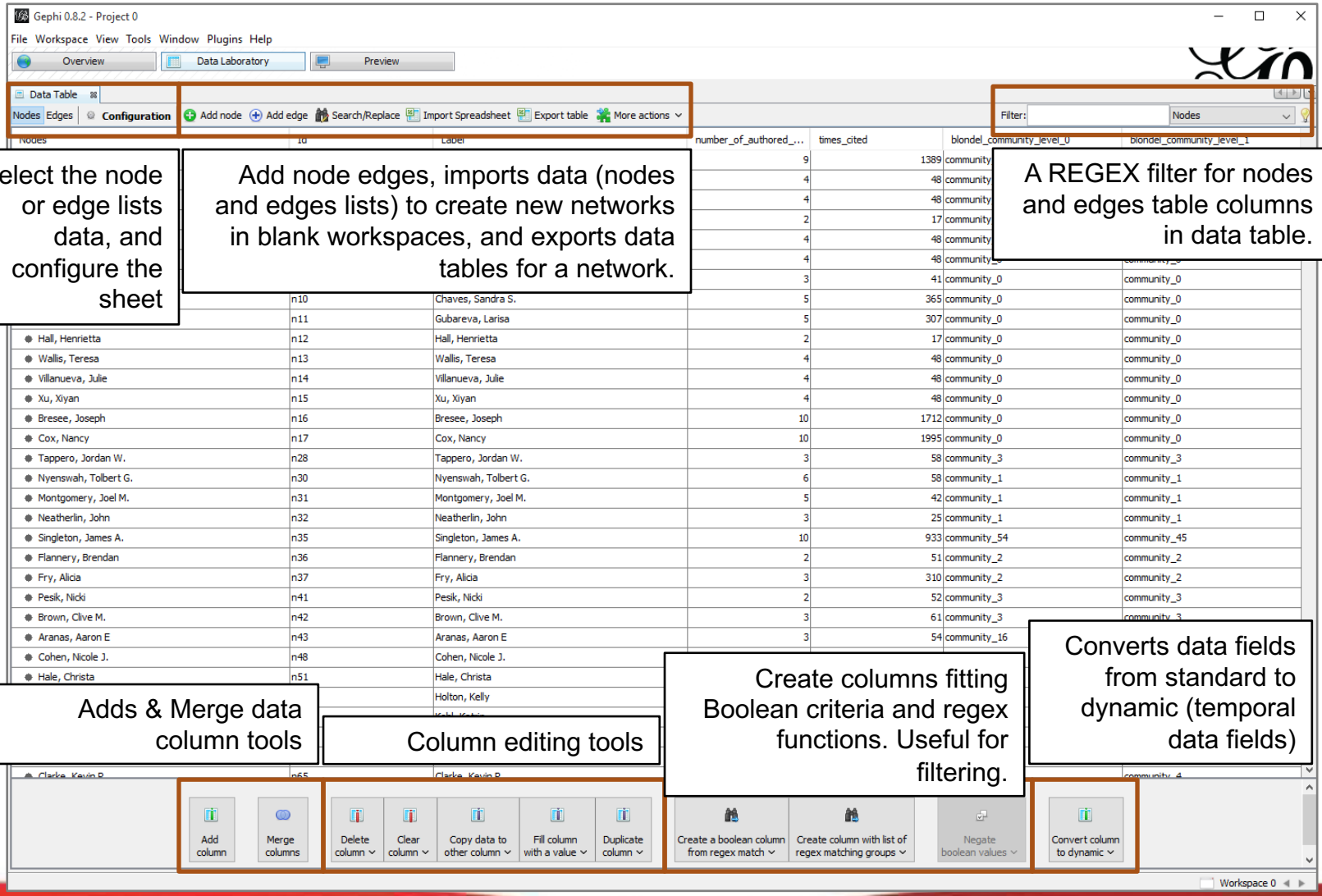
- **Overview**: graph analysis and manipulation mode.
- **Data Laboratory**: data tables.
- **Preview**: visual tuning before vectorial/raster export.

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

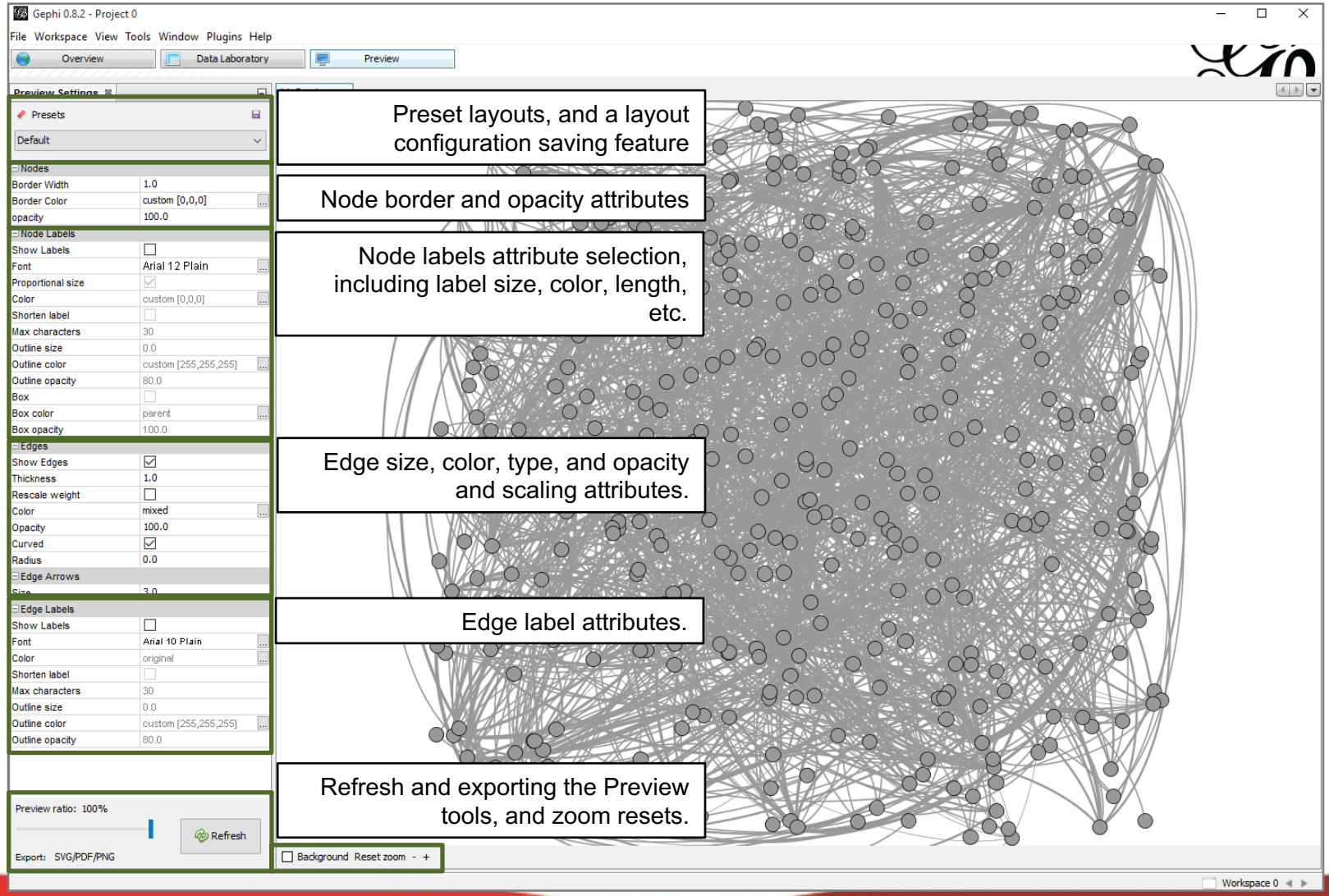
Gephi Interface: Overview Pane



Gephi Interface: Data Laboratory



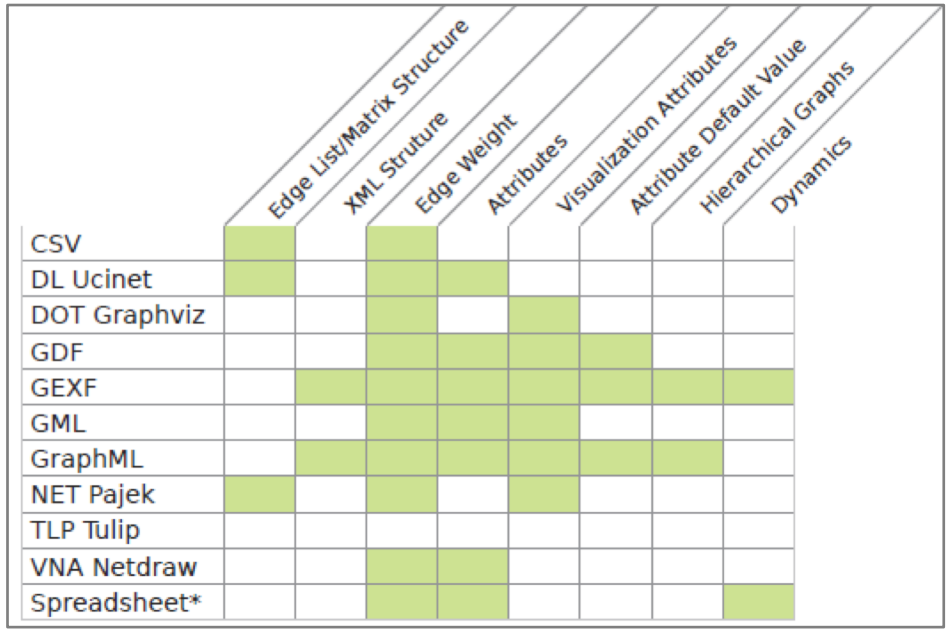
Gephi Interface: Preview



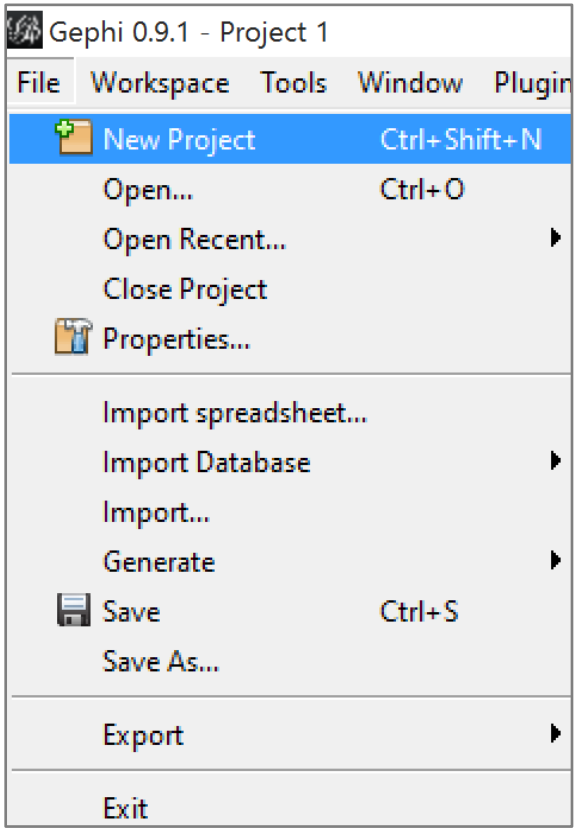
Creating a Visualization in Gephi

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



Creating a New Project

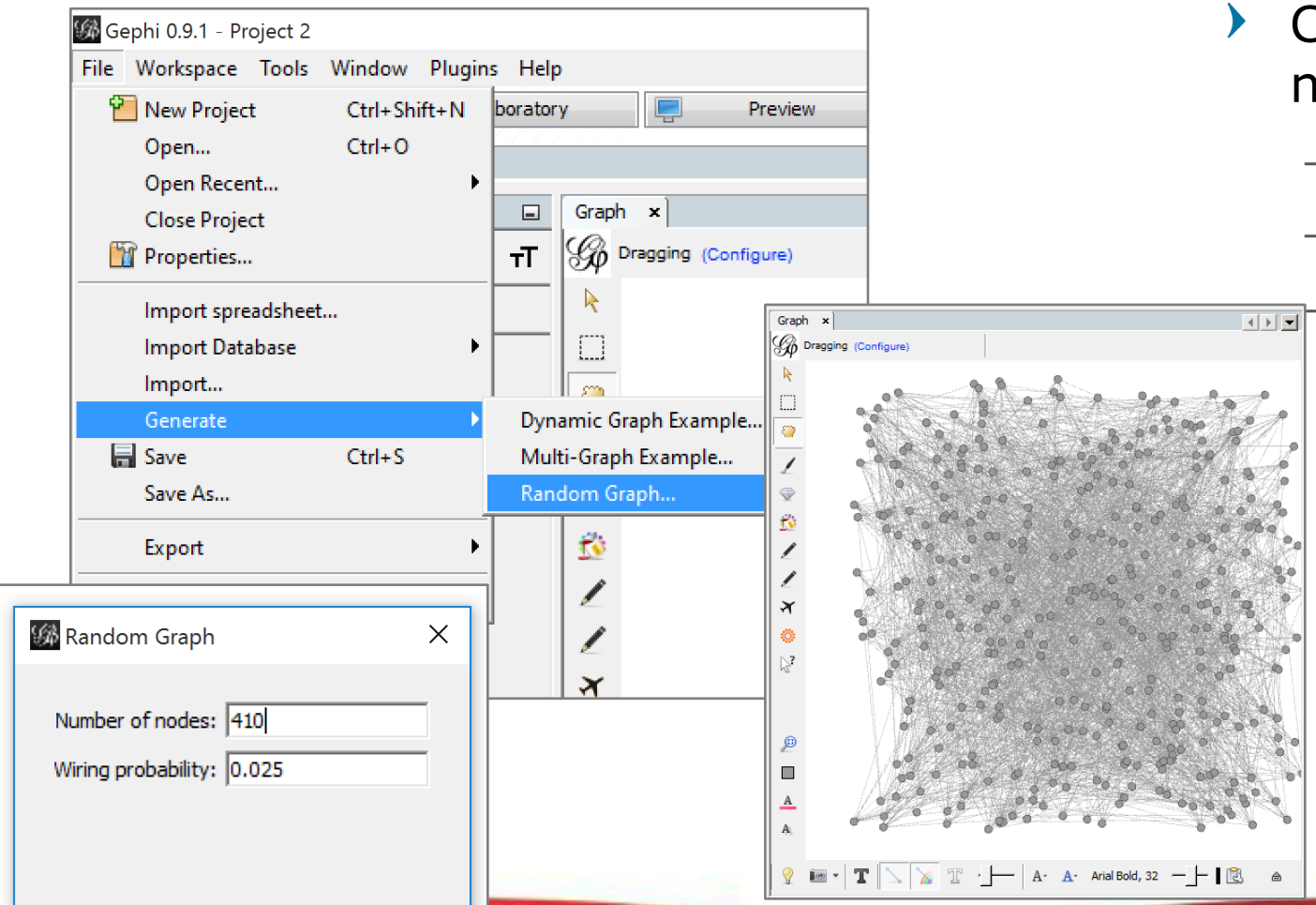


Create a New Project

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

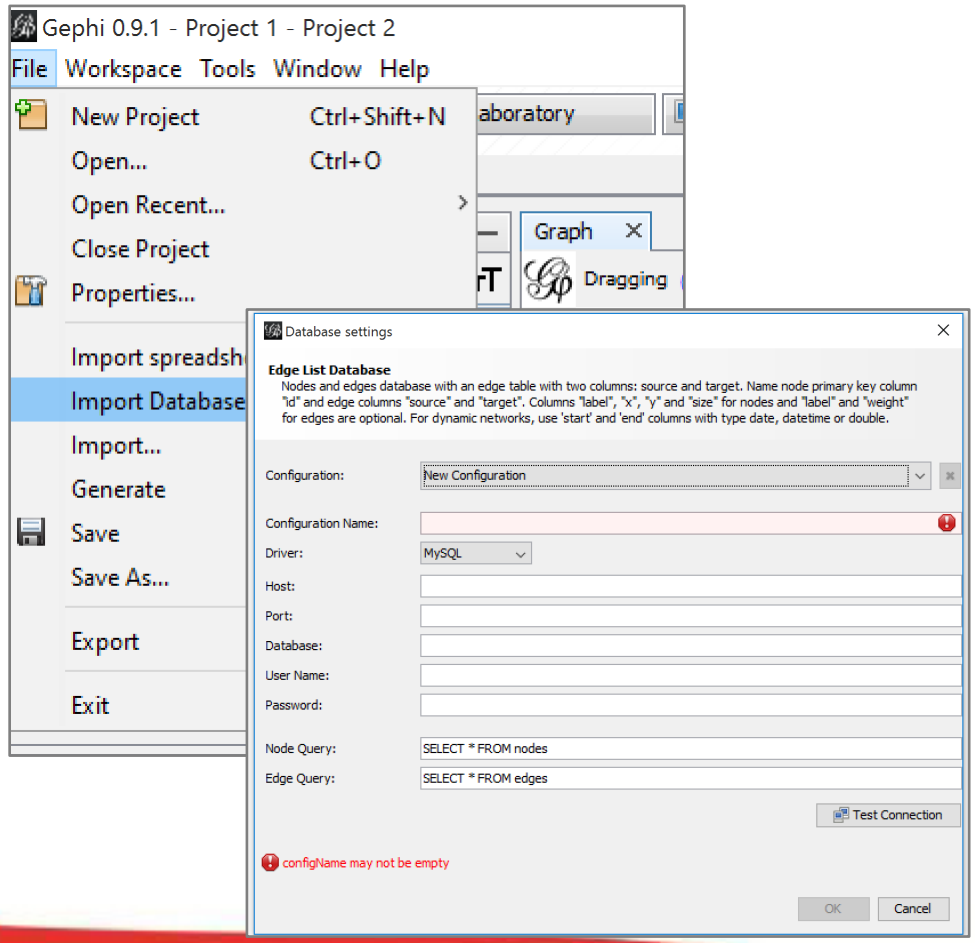
You can select new project from the file menu or select Ctrl+Shft+N on your keyboard.

Loading a Random Graph



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

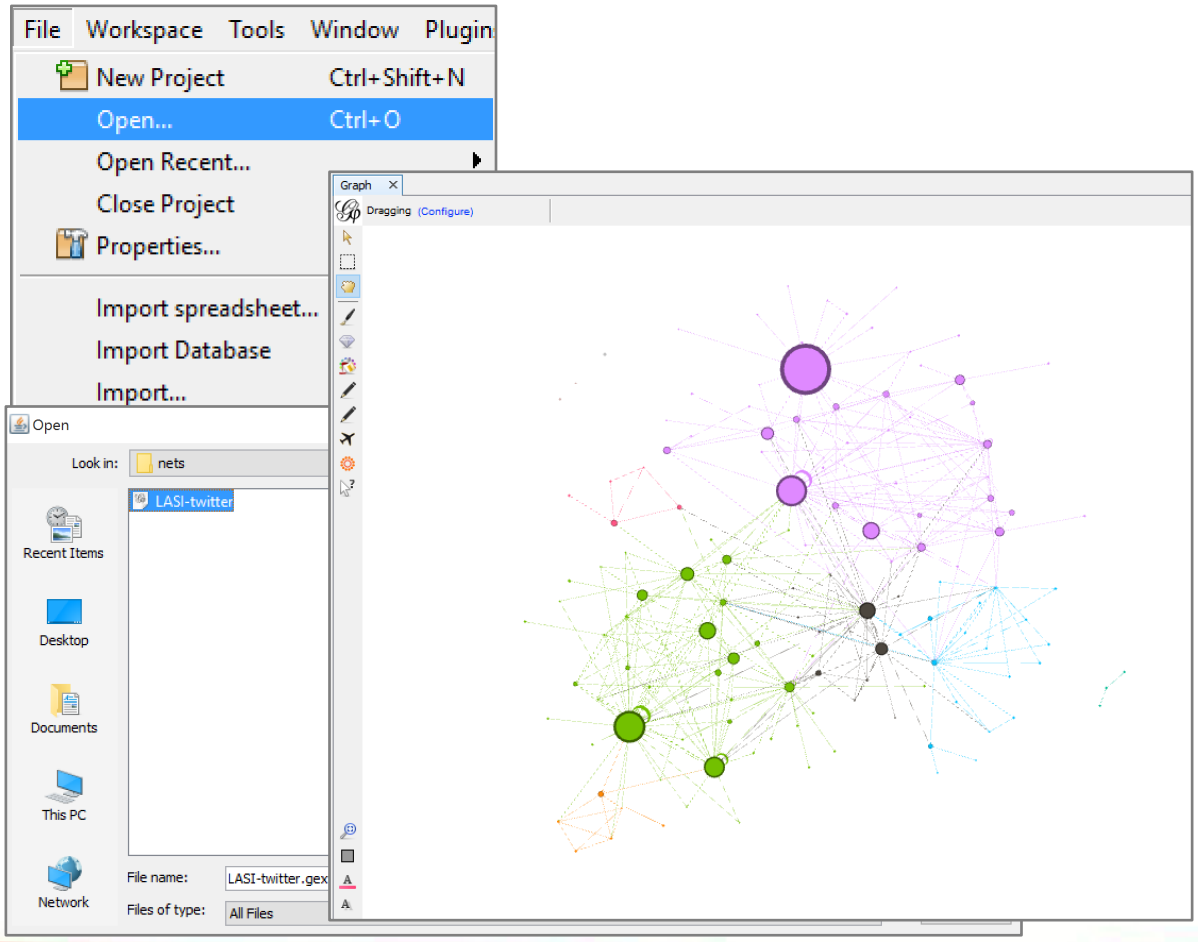
Streaming in an edge list to Gephi



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

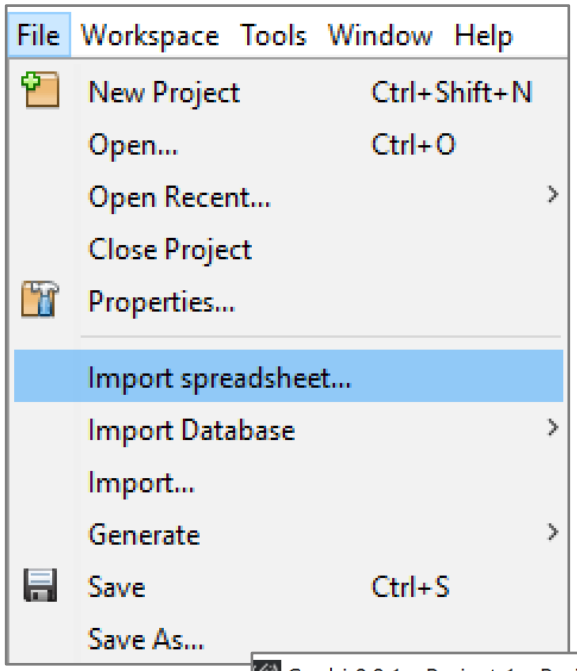
Gephi can import networks from tabular data (as CSVs), if you can provide a node* and edge list using Gephi's *Import Spread Sheet tool*.

- (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

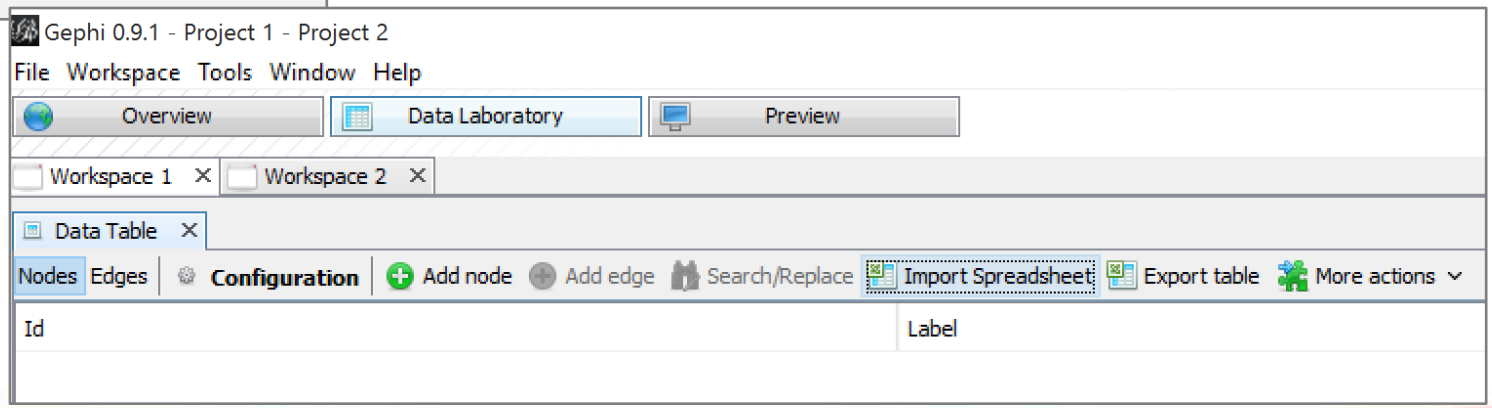
Loading a Network from Spreadsheets

1

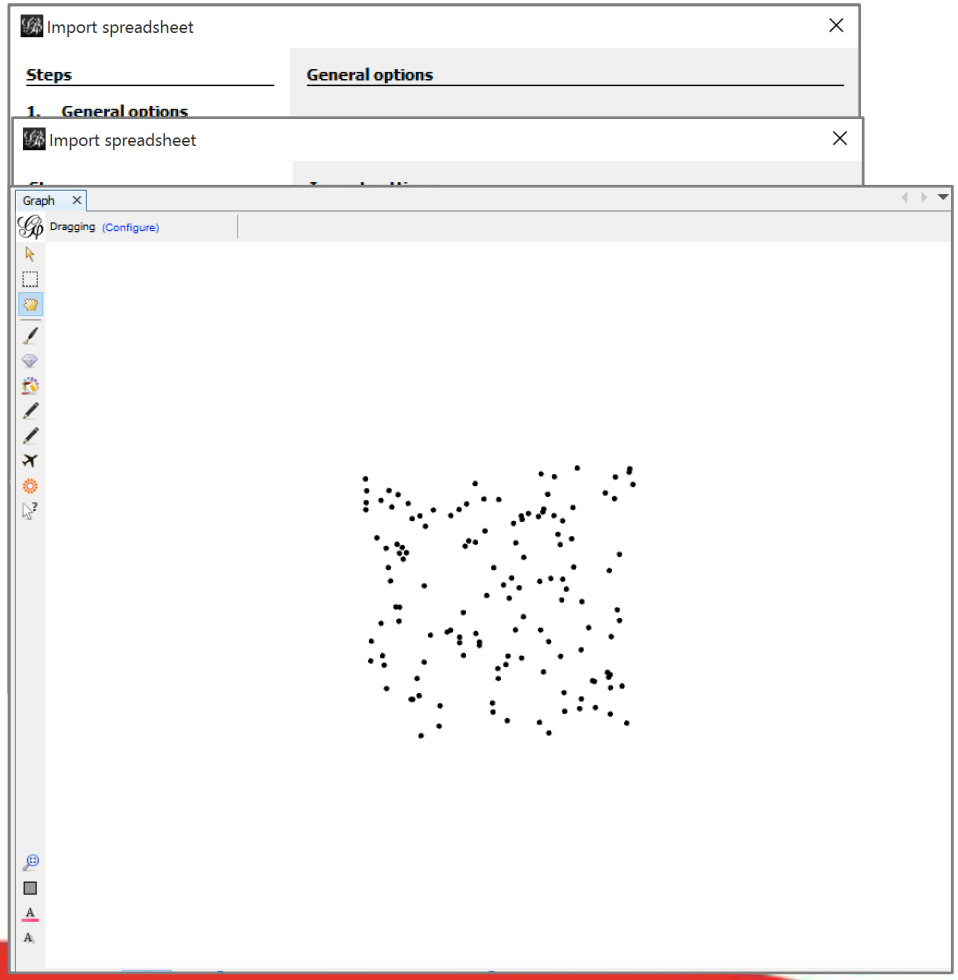


1) The spreadsheet import tool may be engaged from the file menu and selecting “Import spreadsheet…”
or via
2) the Data Laboratory window’s “Data Table” view by selecting “Import spreadsheet…” button.

2



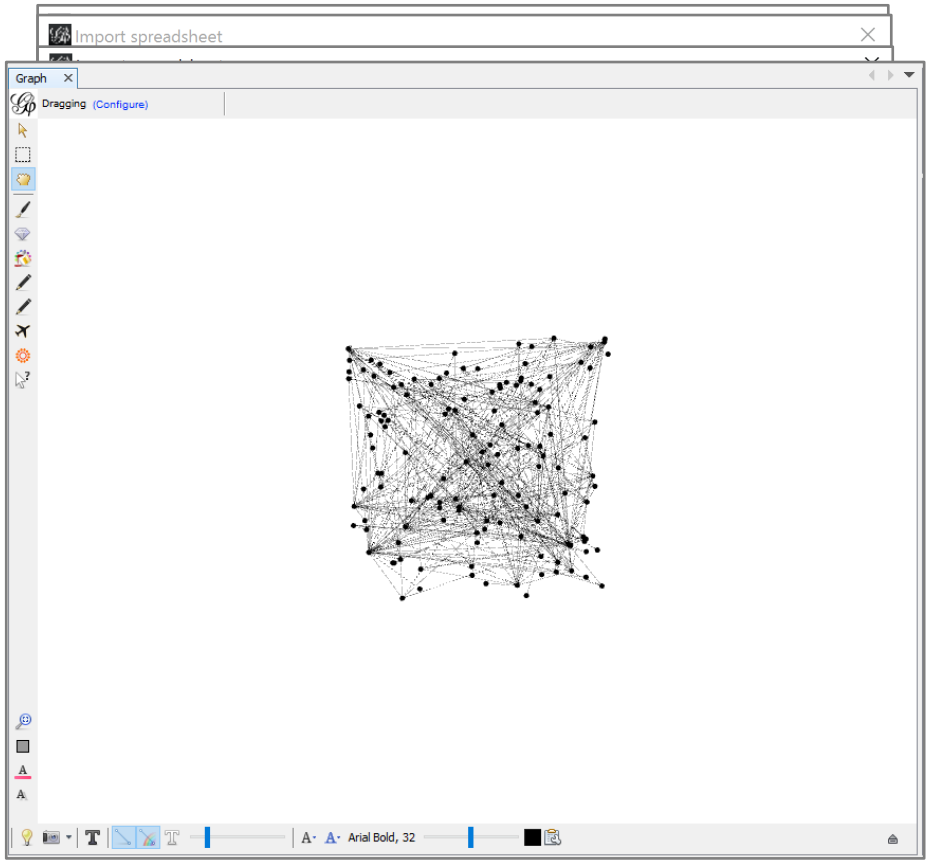
Loading a Network from Spreadsheets



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.

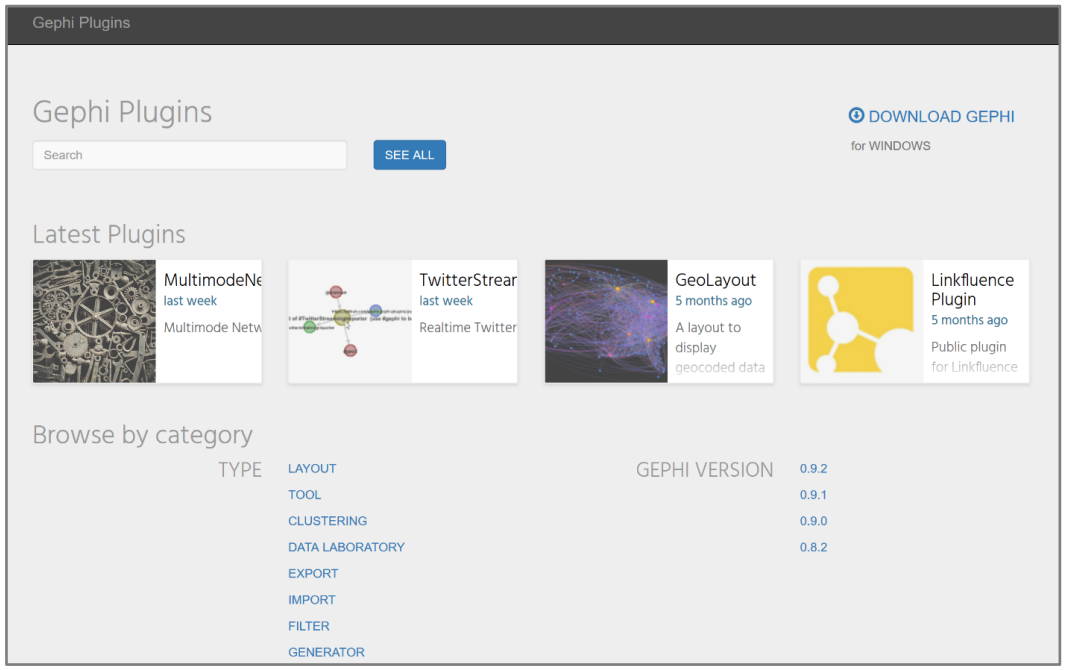
Loading a Network from Spreadsheets



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.

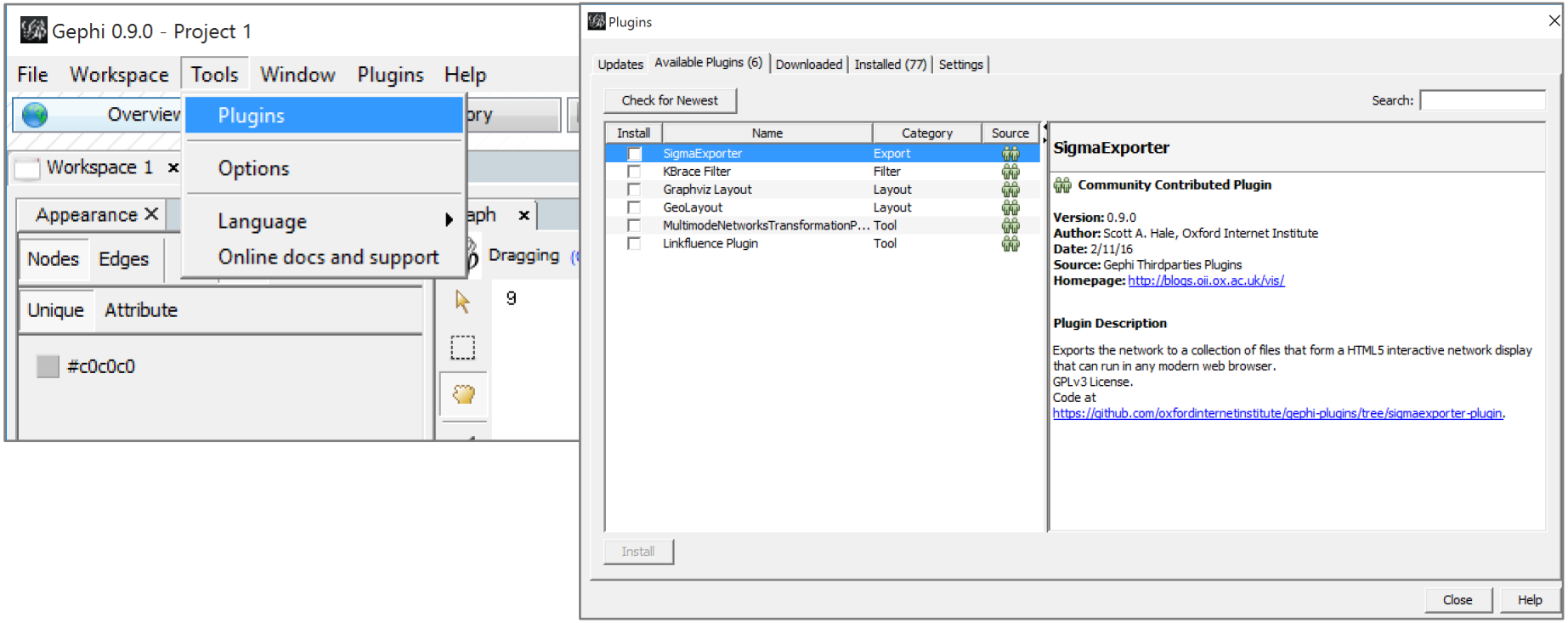
Finding and Adding Plugins to Gephi



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.

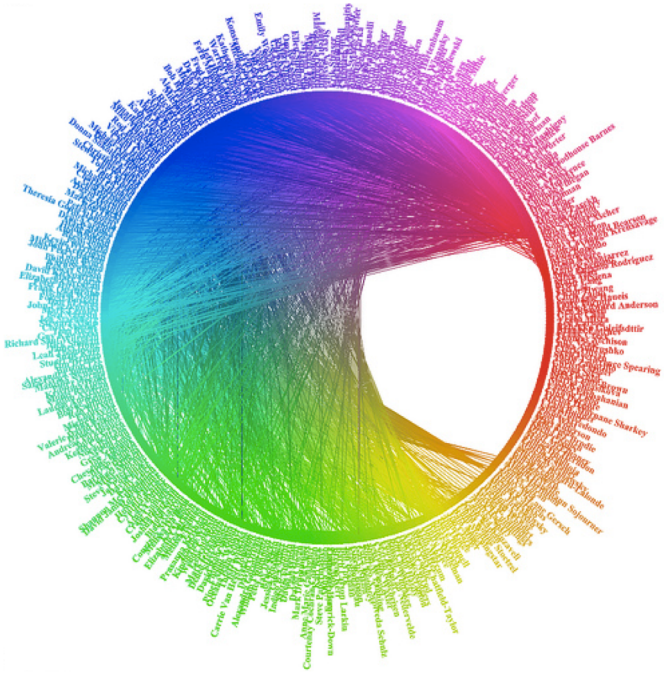
Finding and Adding Plugins to Gephi



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:

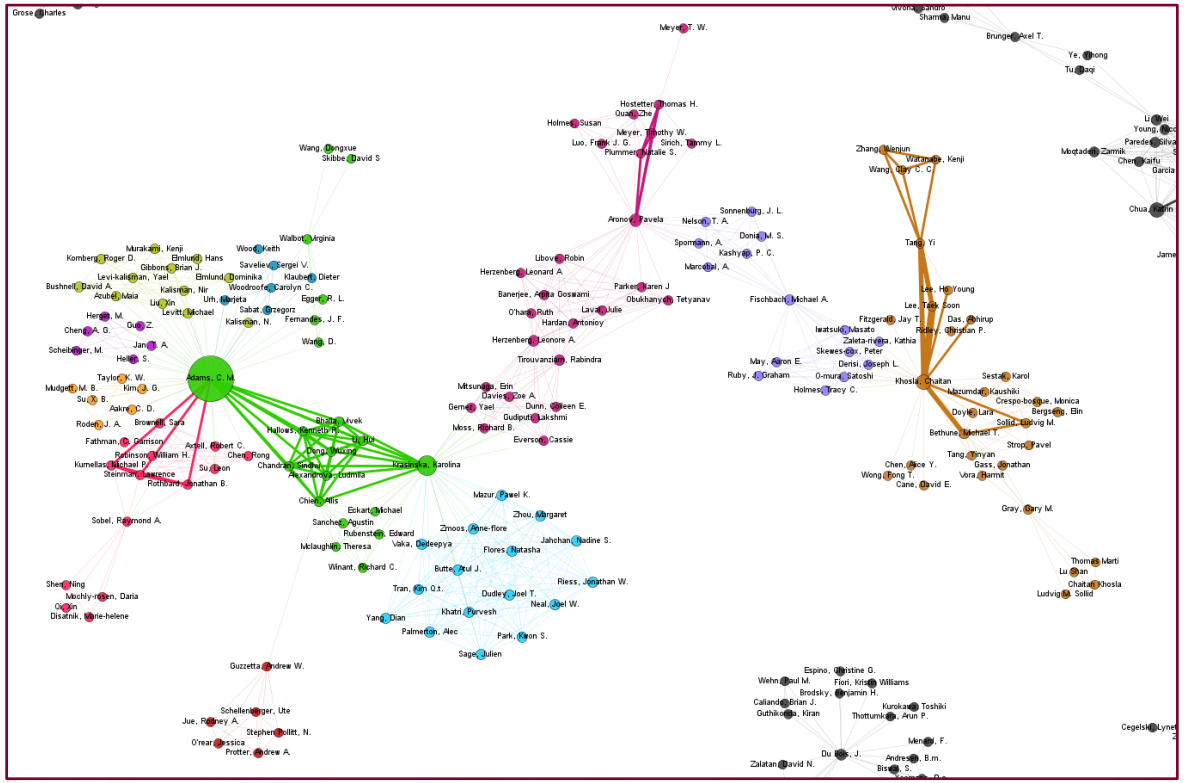
- install updates to existing plugins,
- check for plugins that are listed as available for your release, or
- manually add plugins downloaded from Gephi's plugin site; and
- see plugins that are install (which can be activated or deactivated) and see settings.

Co-Authorship Analysis in Gephi



Co-Author Network Analysis

What is the purpose of looking at co-author networks? What can they tell us?

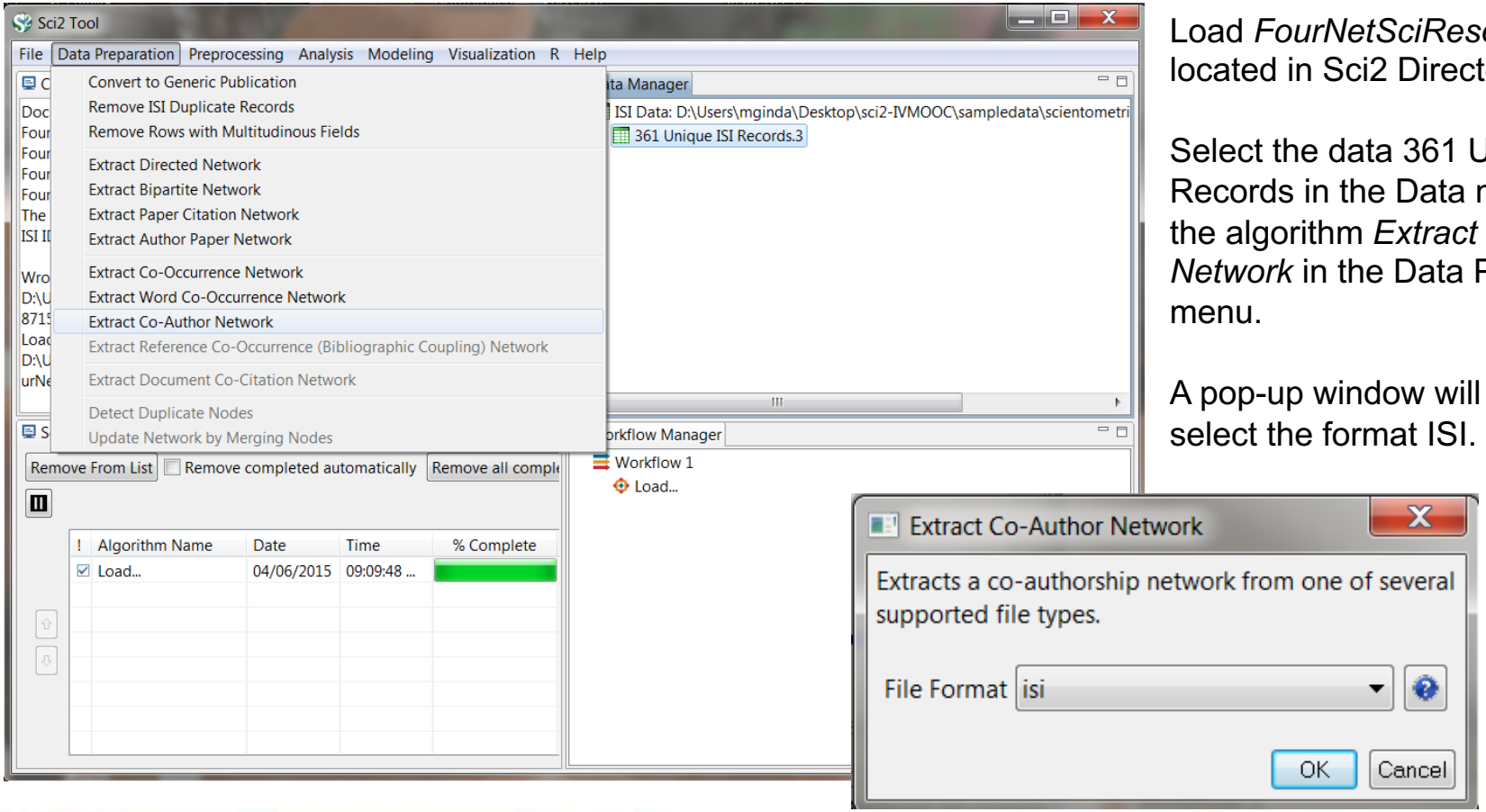


This is a visualization of a citation dataset from a researcher that administers a Core research facility at Stanford University.

The objective of the researcher who provided this data set was to understand

1. Which researchers using her lab were publishing articles?
2. Which researchers collaborate frequently in the facility?
3. Who has the most citation impact?

Co-Author Network Analysis

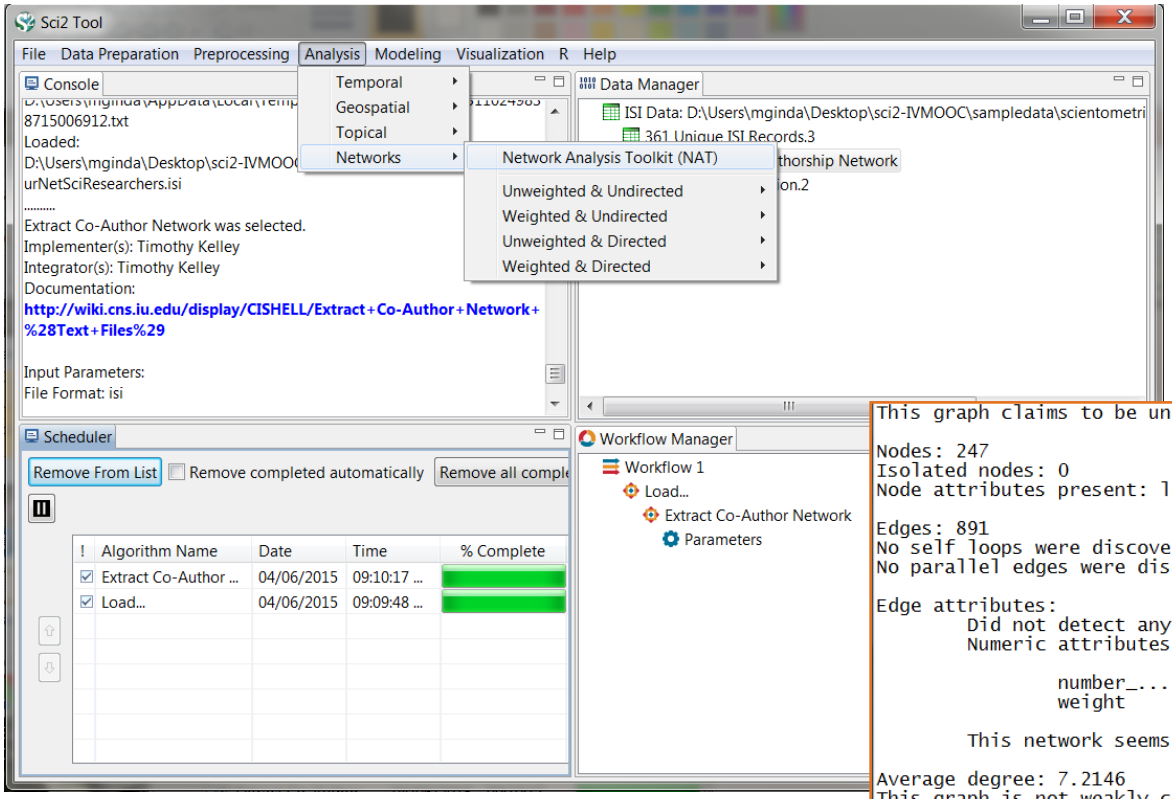


Load *FourNetSciResearchers.isi* located in Sci2 Directory

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.

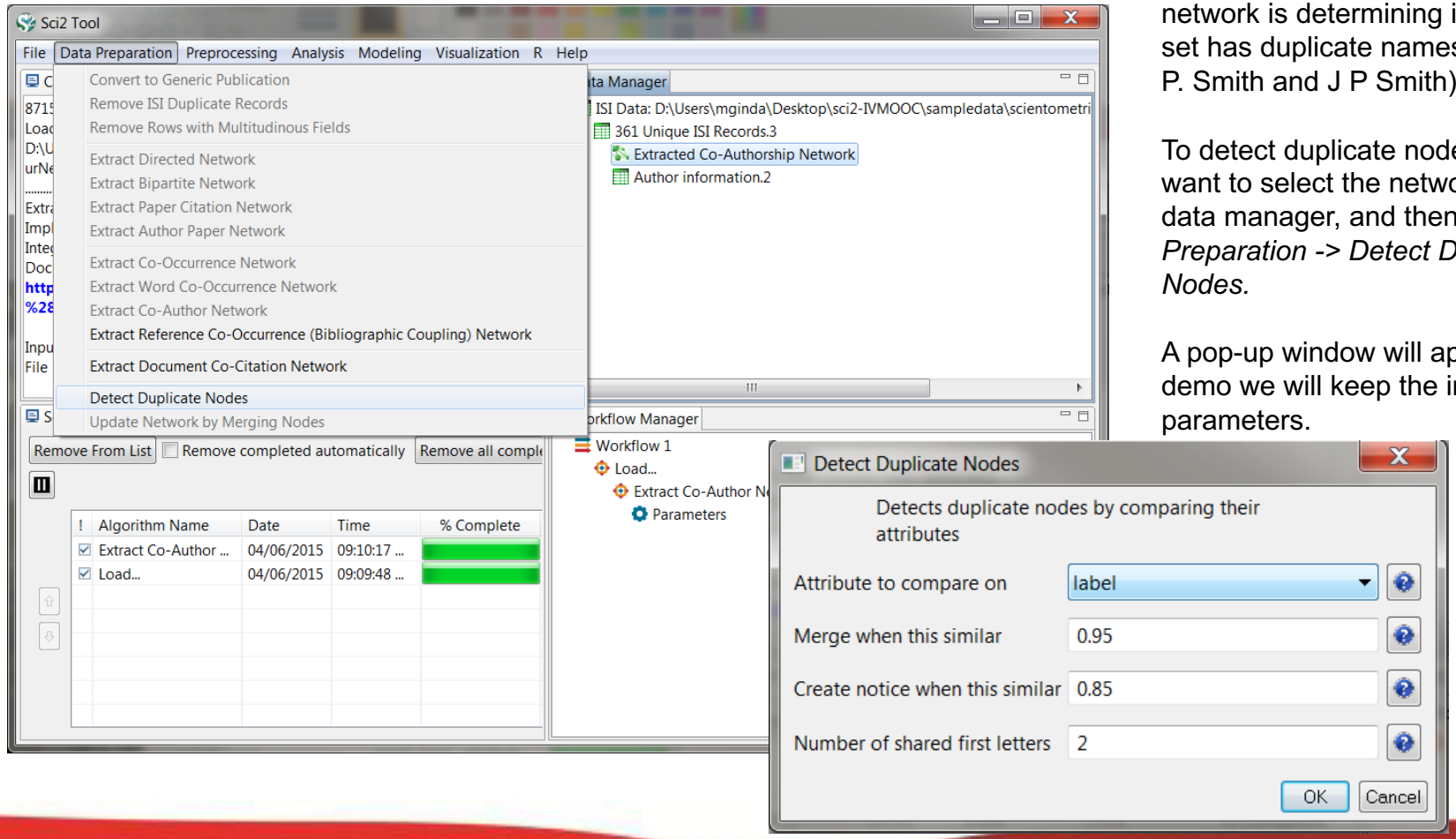
Sci2 has a built-in analysis toolkit to perform these basic statistics.

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

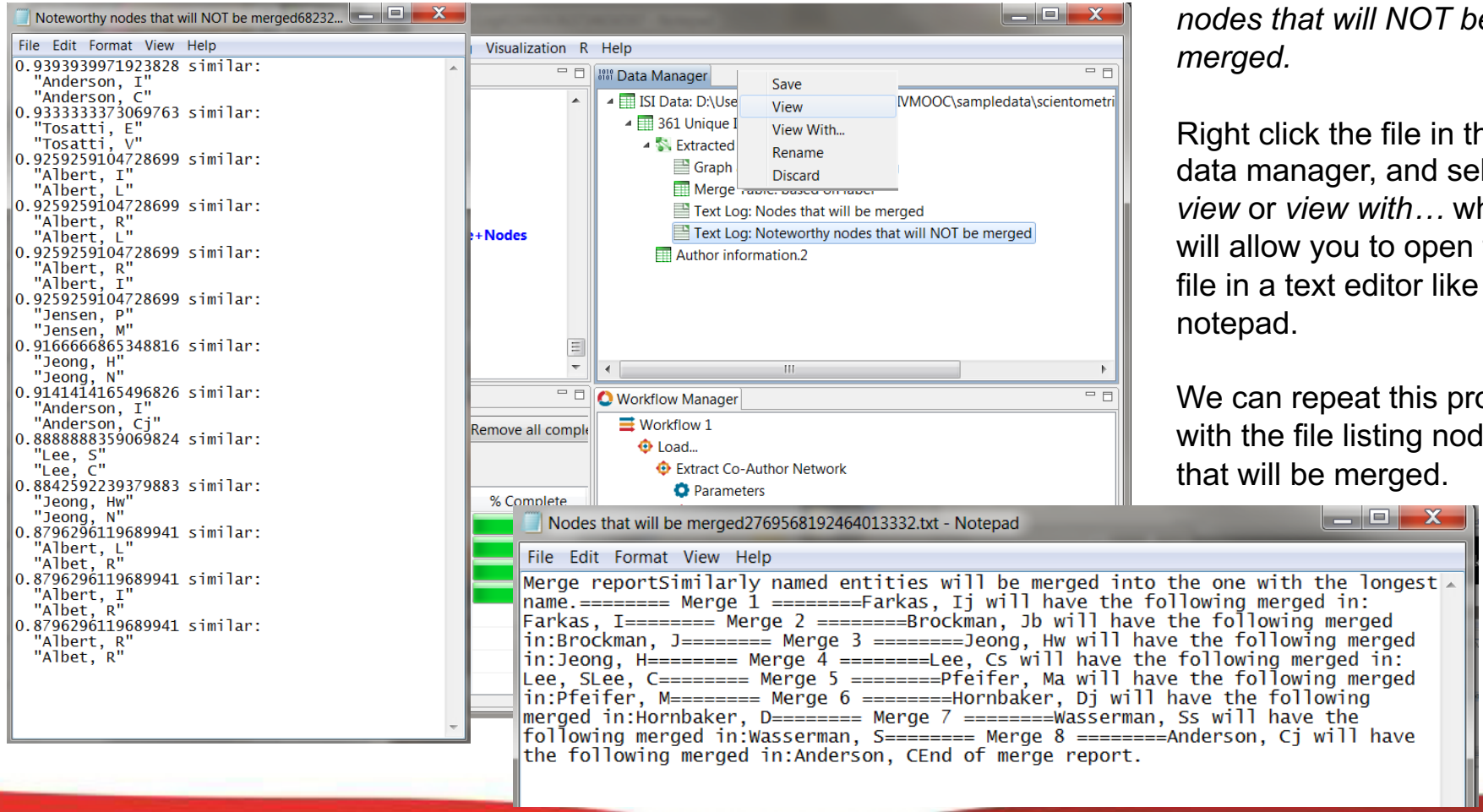


One of the challenges of a co-author network is determining if your data set has duplicate names (e.g. John P. Smith and J P Smith).

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

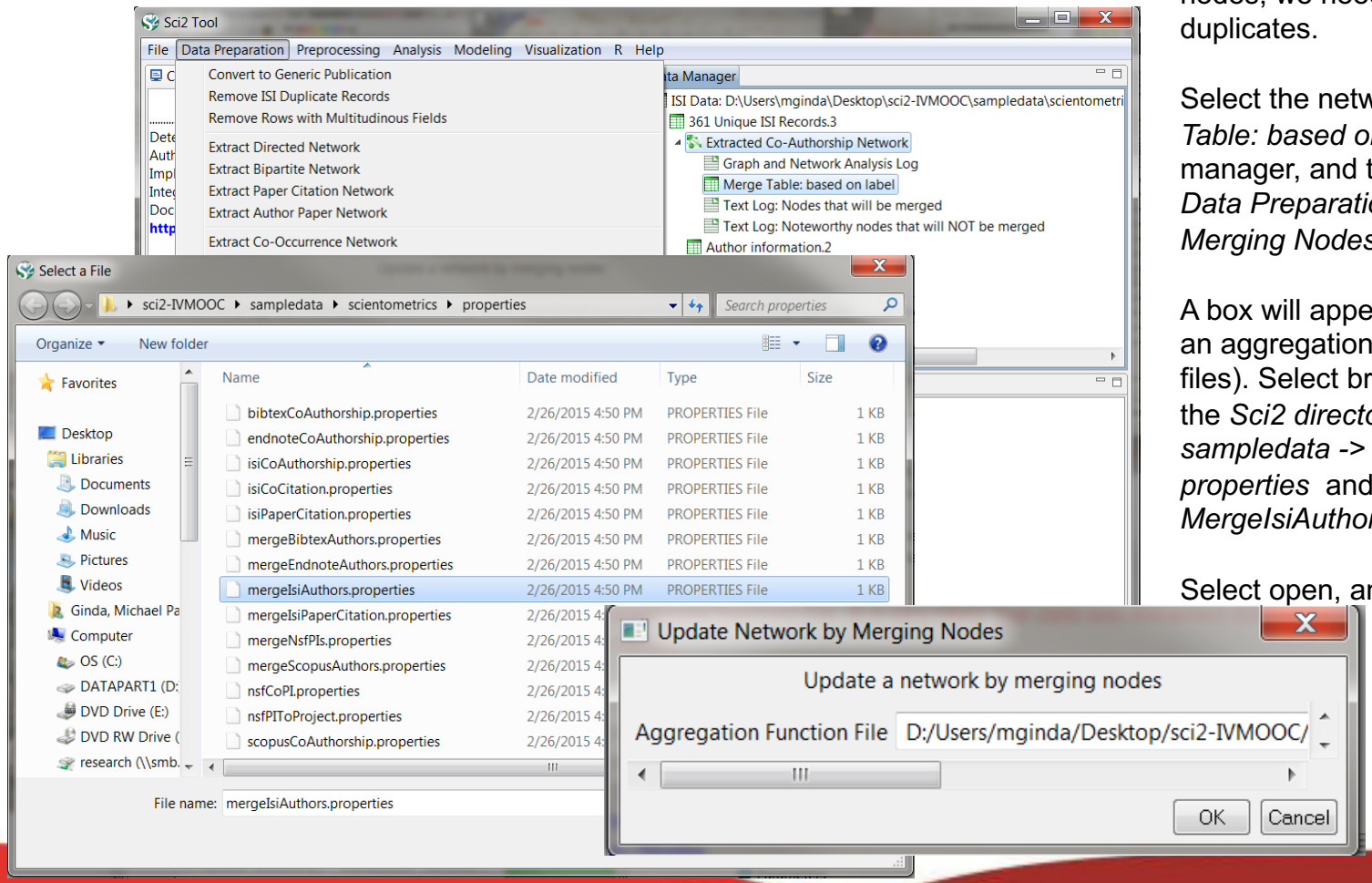


Let's look at the output file *Text Log: Noteworthy nodes that will NOT be merged.*

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



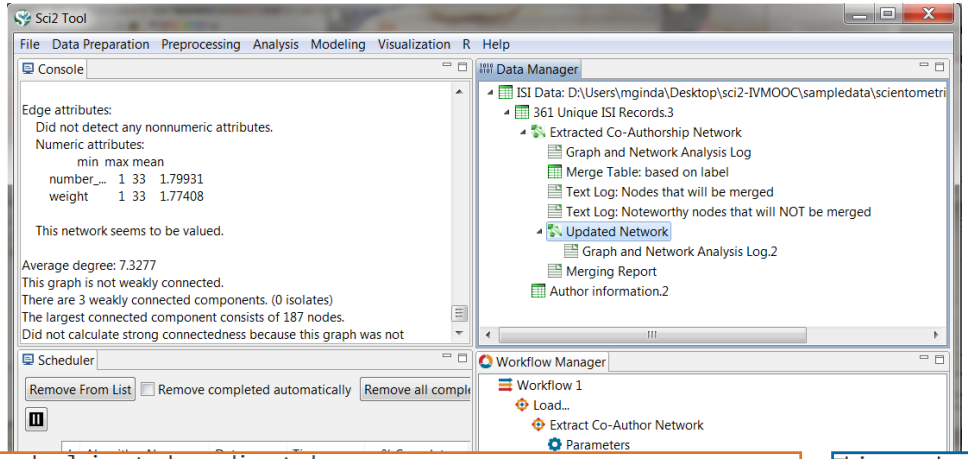
After we've identifying our duplicate nodes, we need to merge these duplicates.

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 *MergeIsiAuthors.properties*

Select open, and then OK.

Co-Author Network Analysis



We've now updated out network, so lets 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?

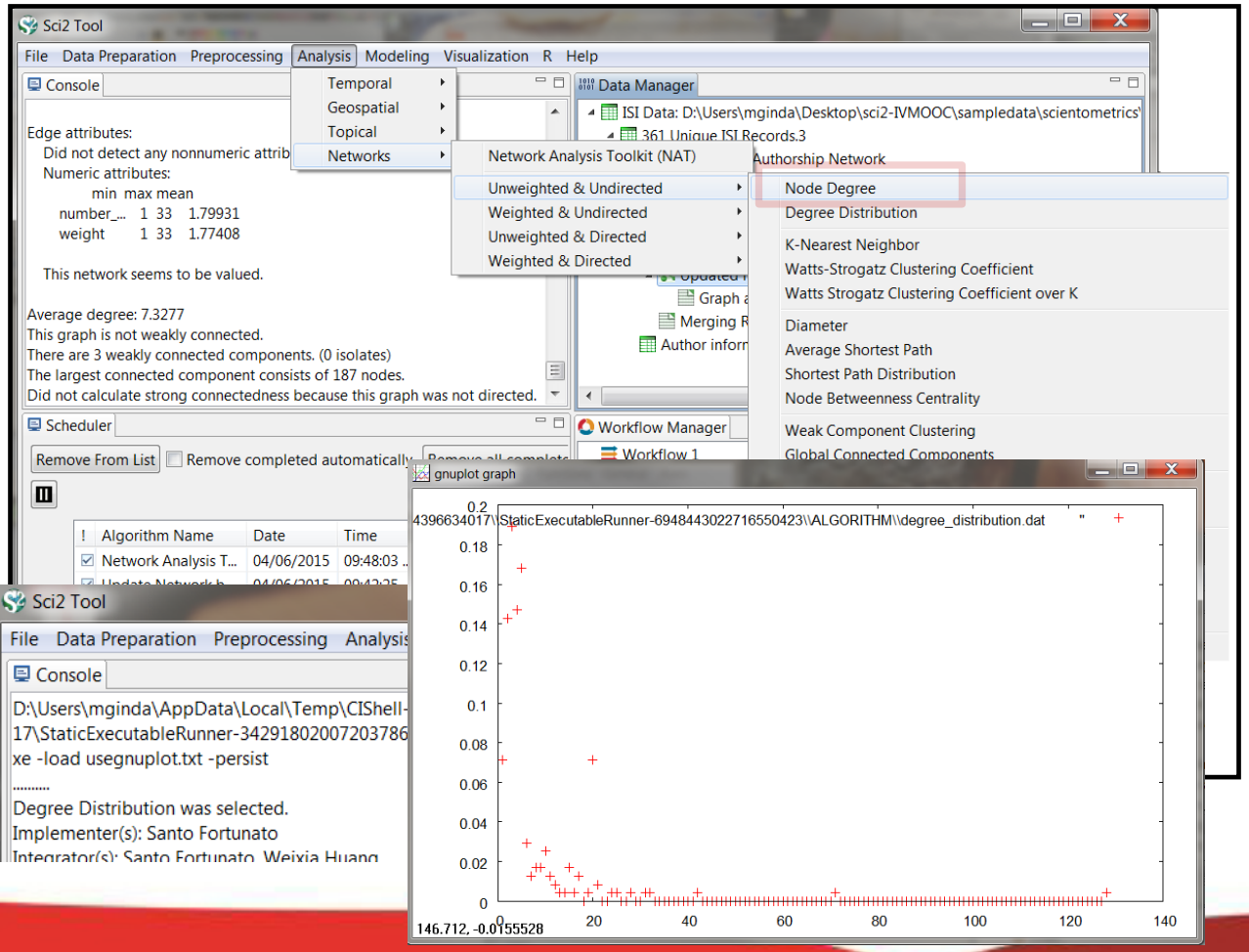
Original Network

```
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
```

Revised Network

```
This graph claims to be undirected.

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, lets find the degree for each node.

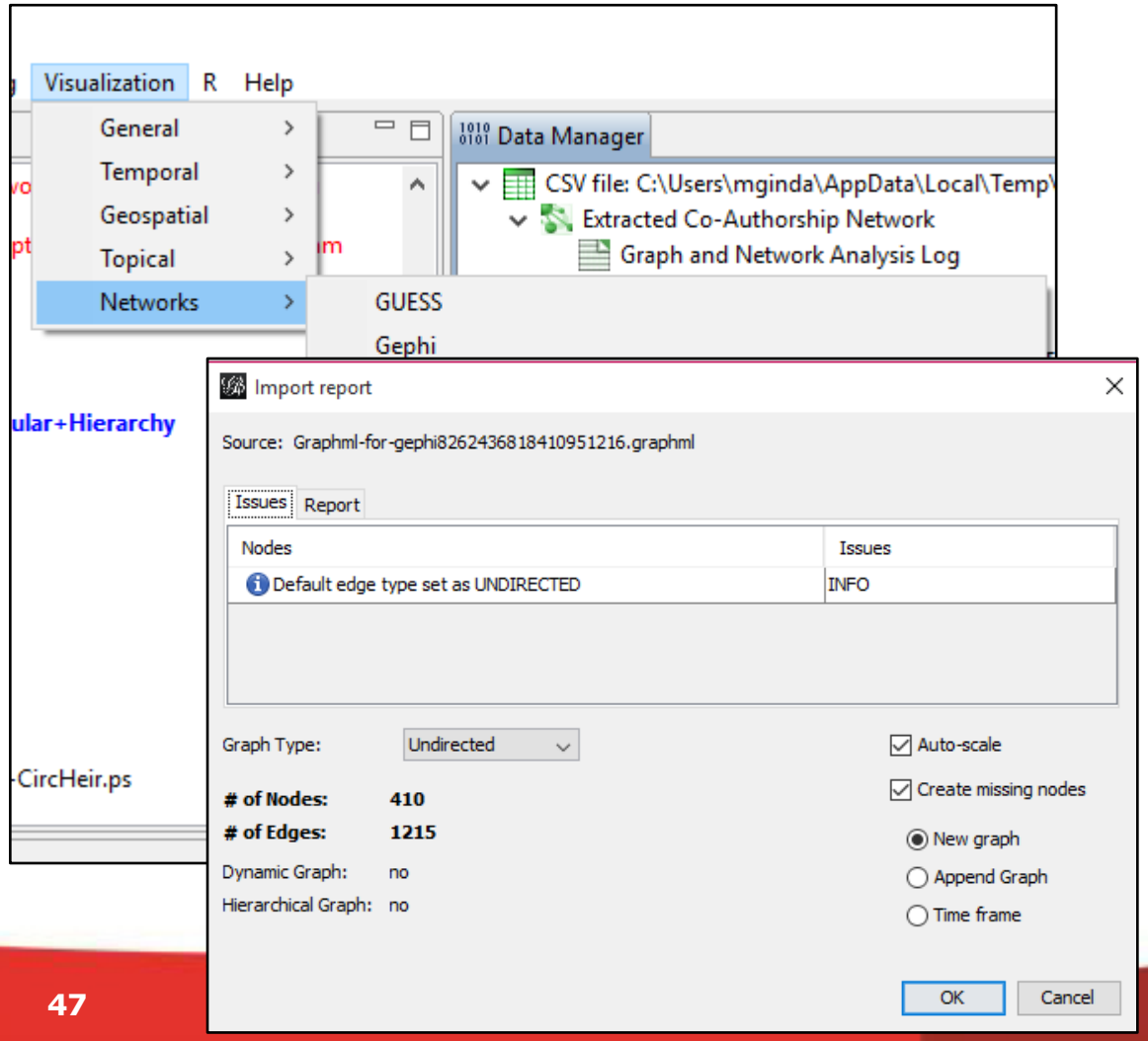
Select the updated network in the data manager, and then select in the menu *Analysis -> Networks-> Unweighted & Undirected -> Node Degree*

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*

Co-Author Network Visualization in Gephi



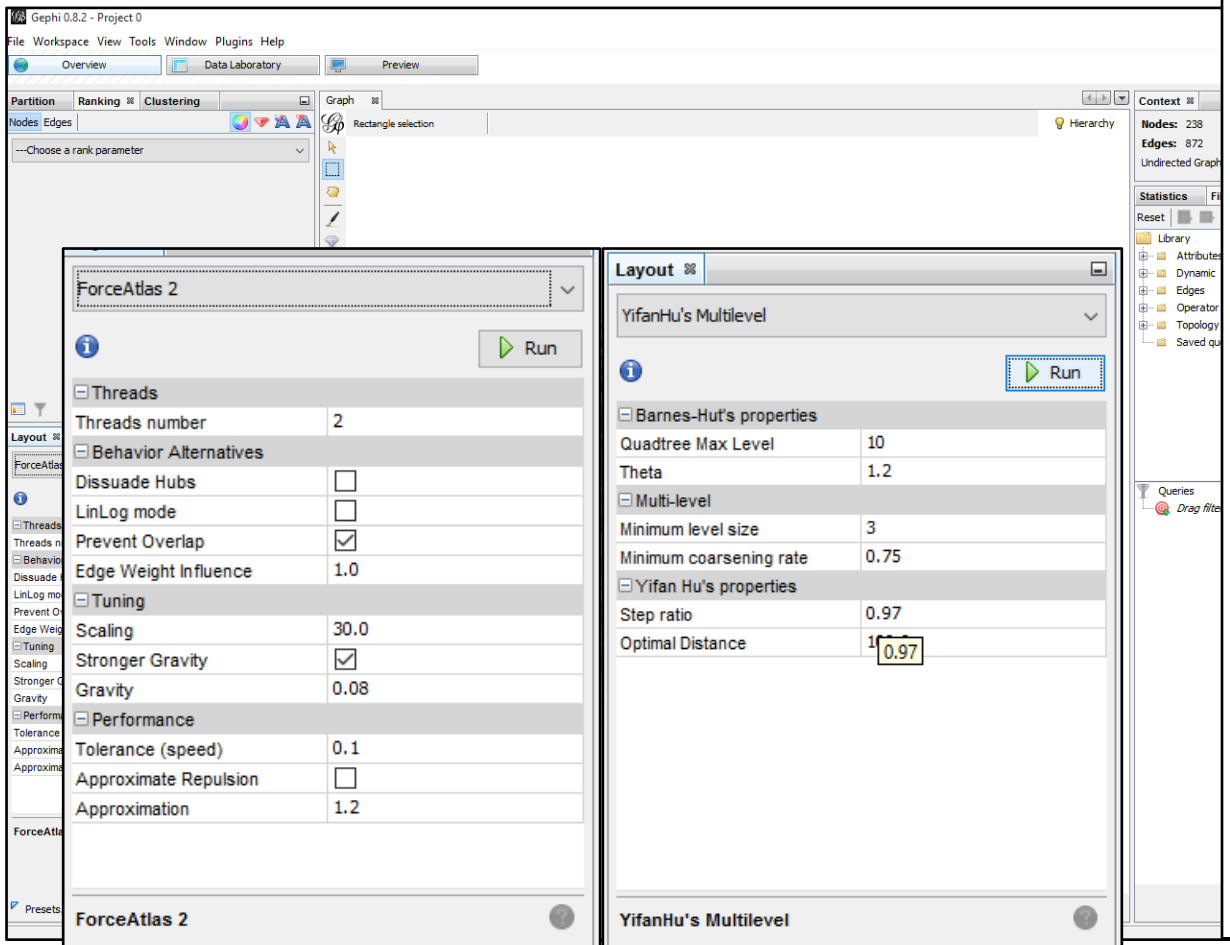
Next we will visualize the network in Gephi.

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



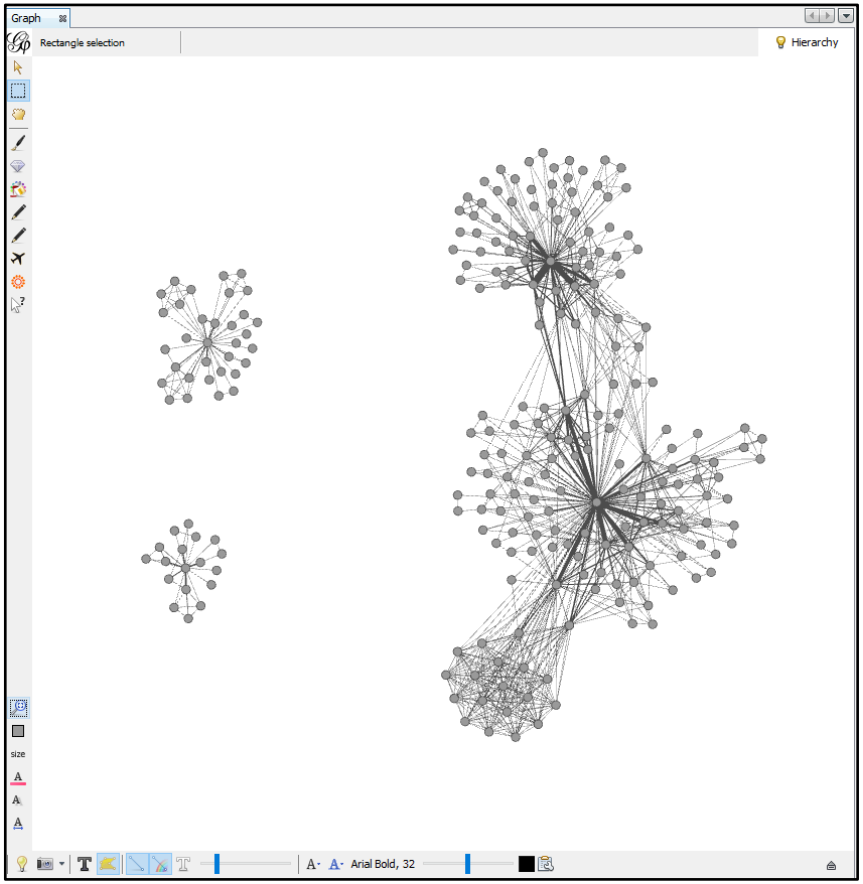
Now we can start an visualization and analysis of the network.

First we will adjust the layout of the network.

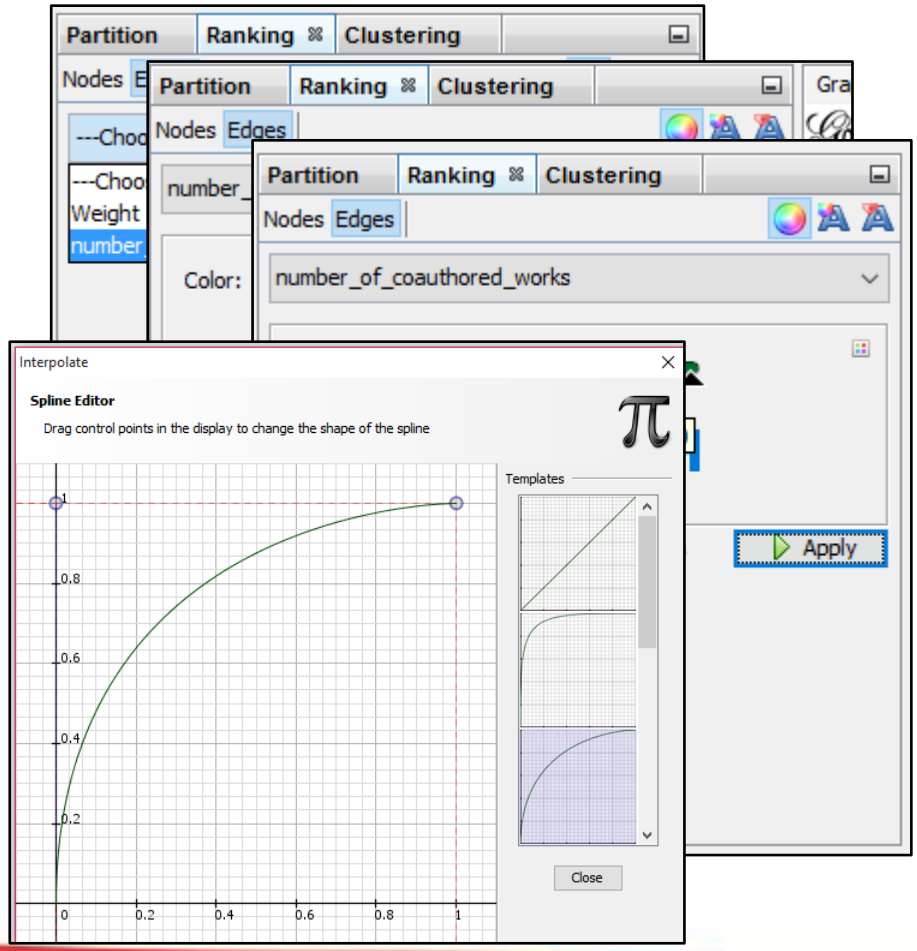
From the layout pane, select the "ForceAtlas2" layout algorithm and enter the following parameters, and then select "Run".

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

Gephi: Initial Layout of Network



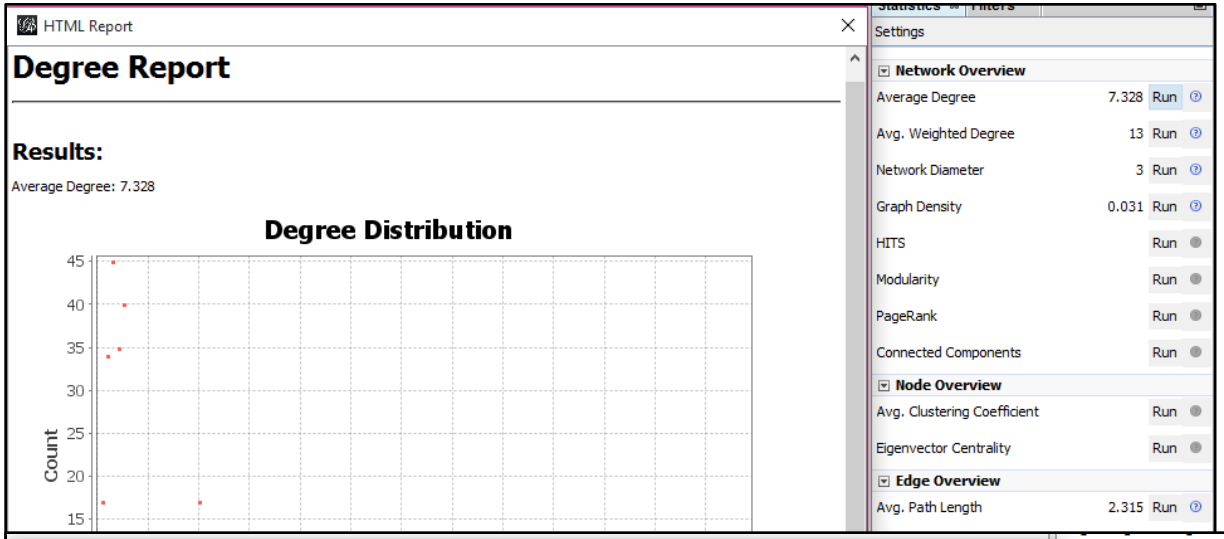
Gephi: Edge Color

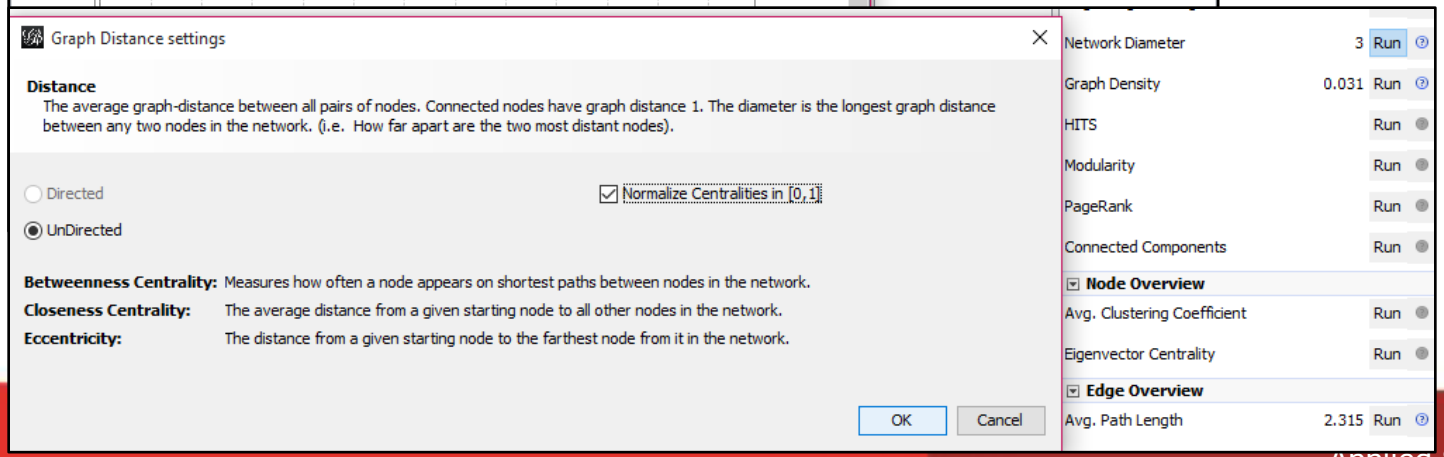


Next, we can adjust the edge color by selecting the Edge tab in the Ranking window.

- 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

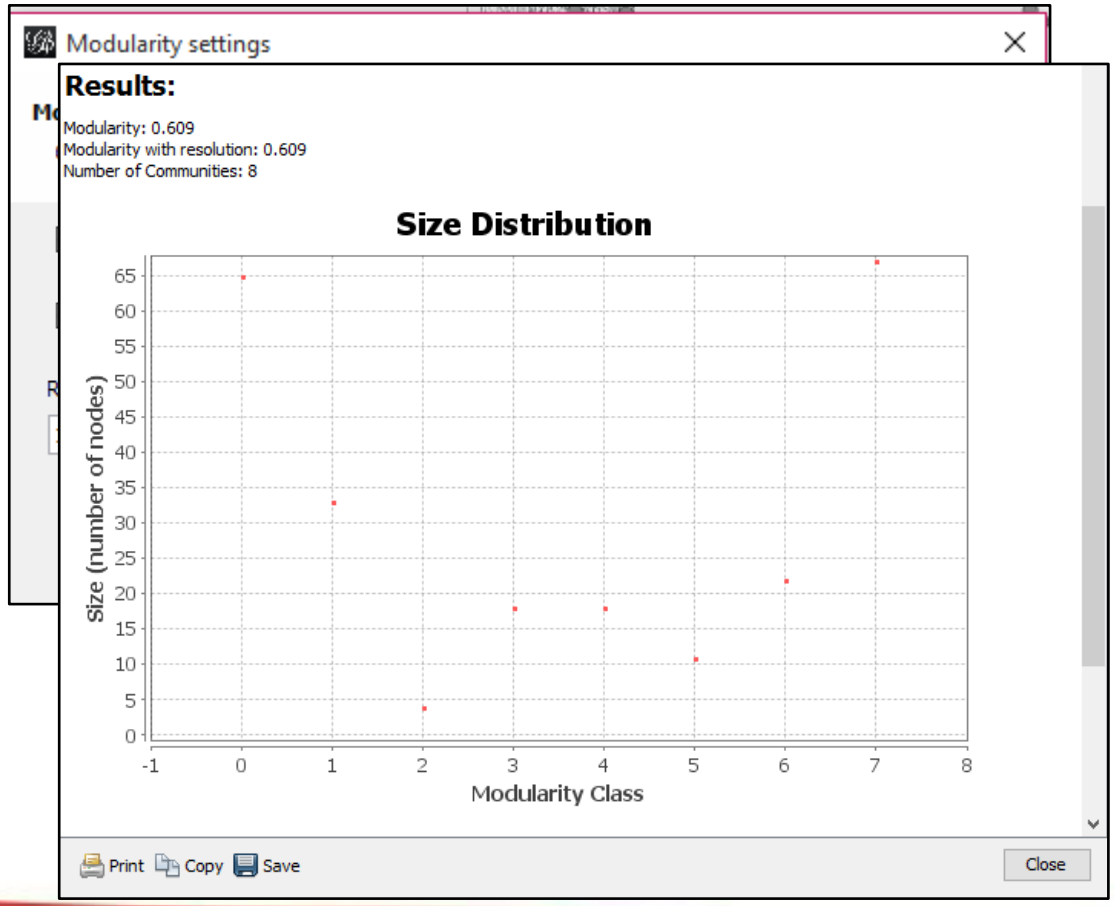




Gephi provides a variety of node and edge statistics to help understand the relationships, clustering, paths, centrality, and communities within a network.

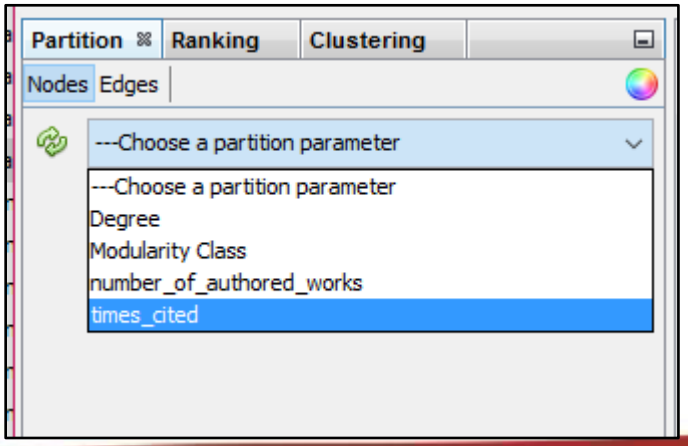
Try implementing the Average Degree, and Network Diameter statistics, which we will next visualize.

Gephi: Network statistics/Community Detection

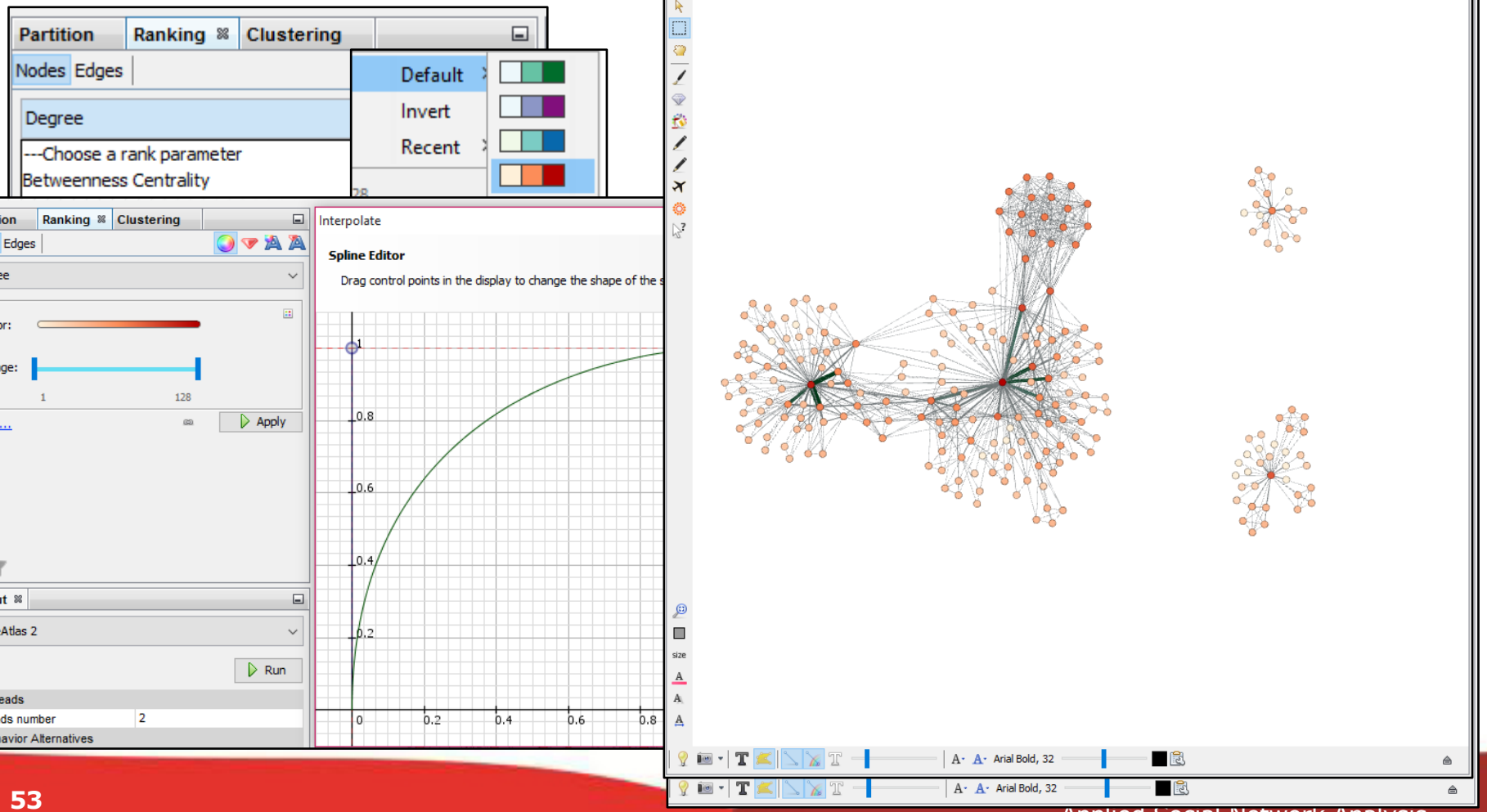


The modularity statistical algorithm calculates how the connectedness of a network, and the Blondel Communities that exist in the network. The communities are added as a partition to the nodes.

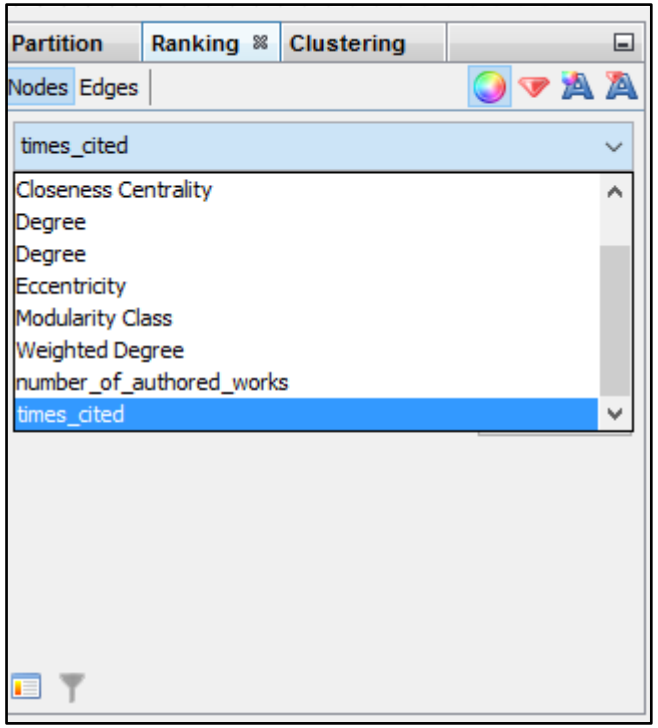
The modularity categories may be applied to the network from the Partitions window.

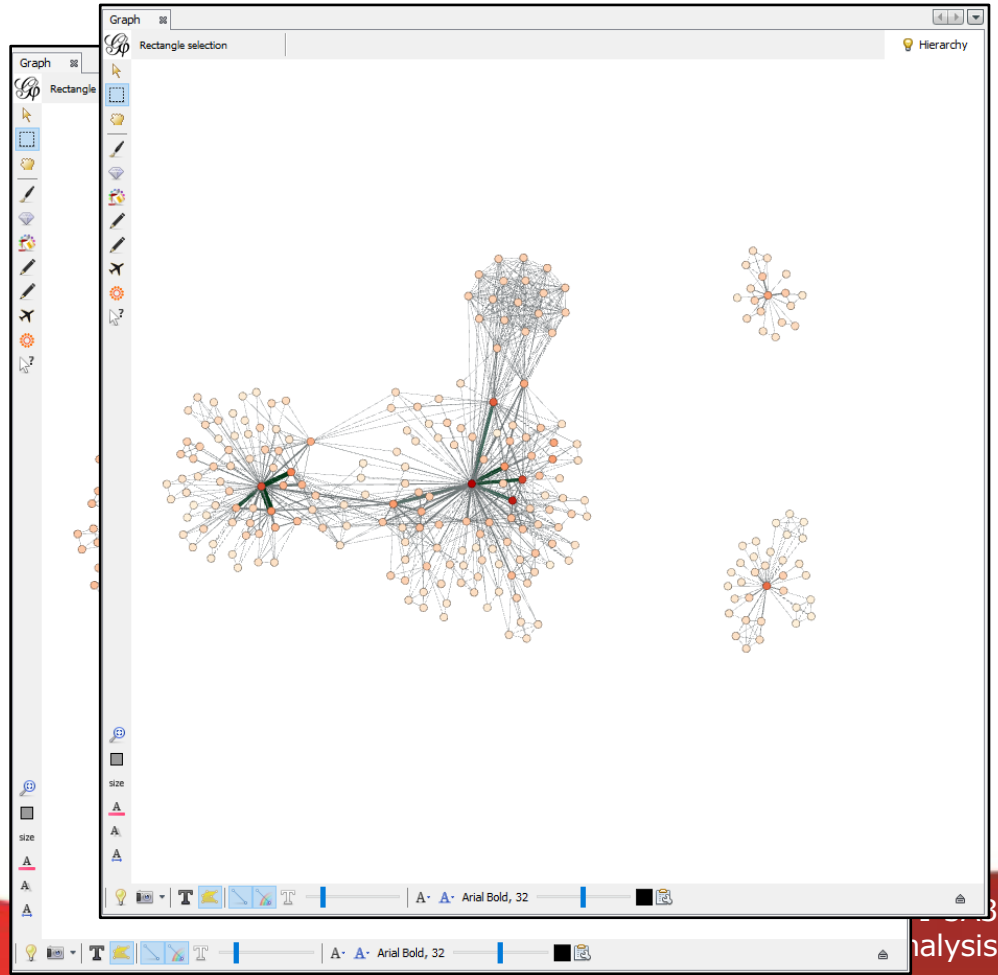


Gephi: Node Color Ranking & Scaling – Degree & Times Cited

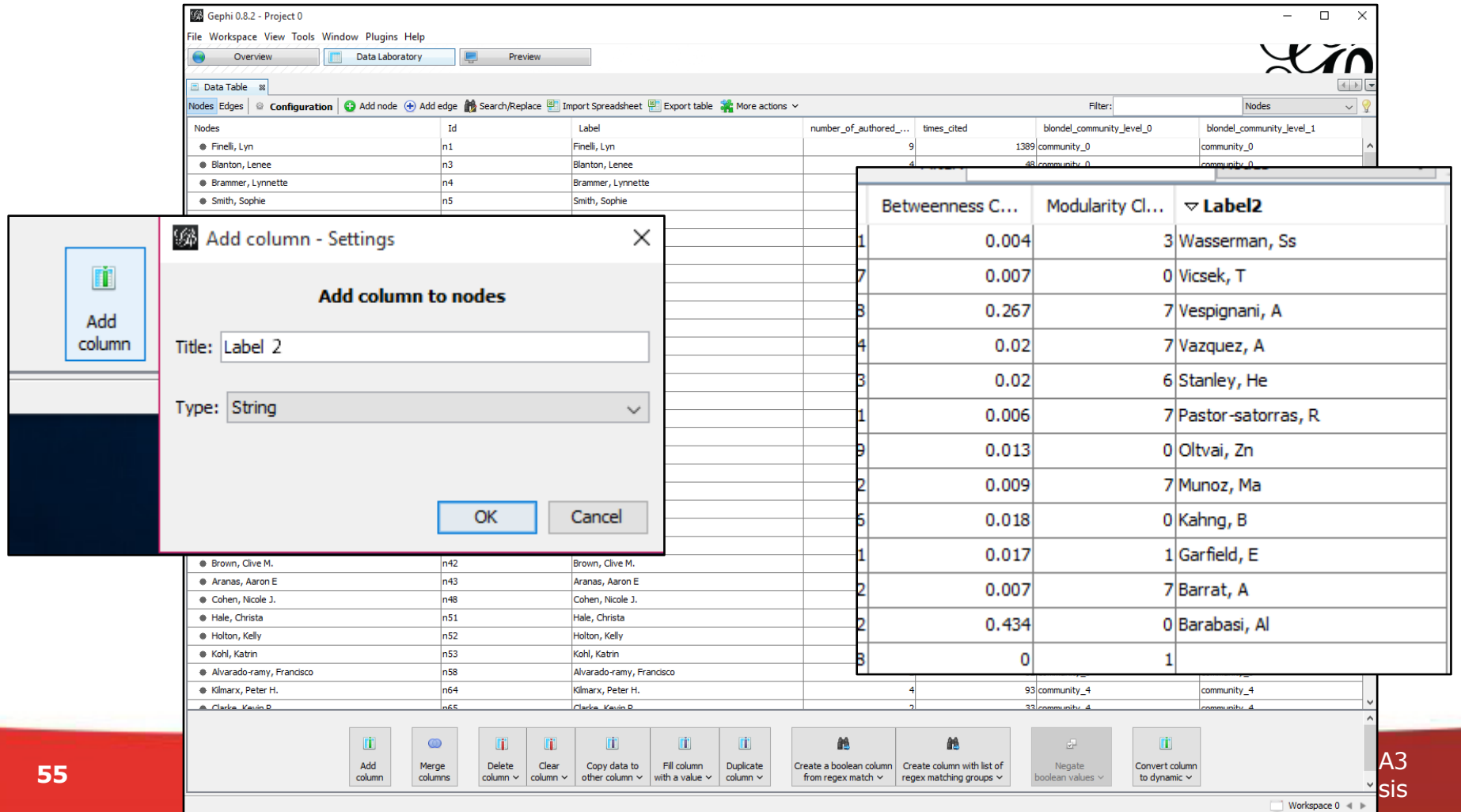


Gephi: Node Color Ranking & Scaling – Betweenness Centrality



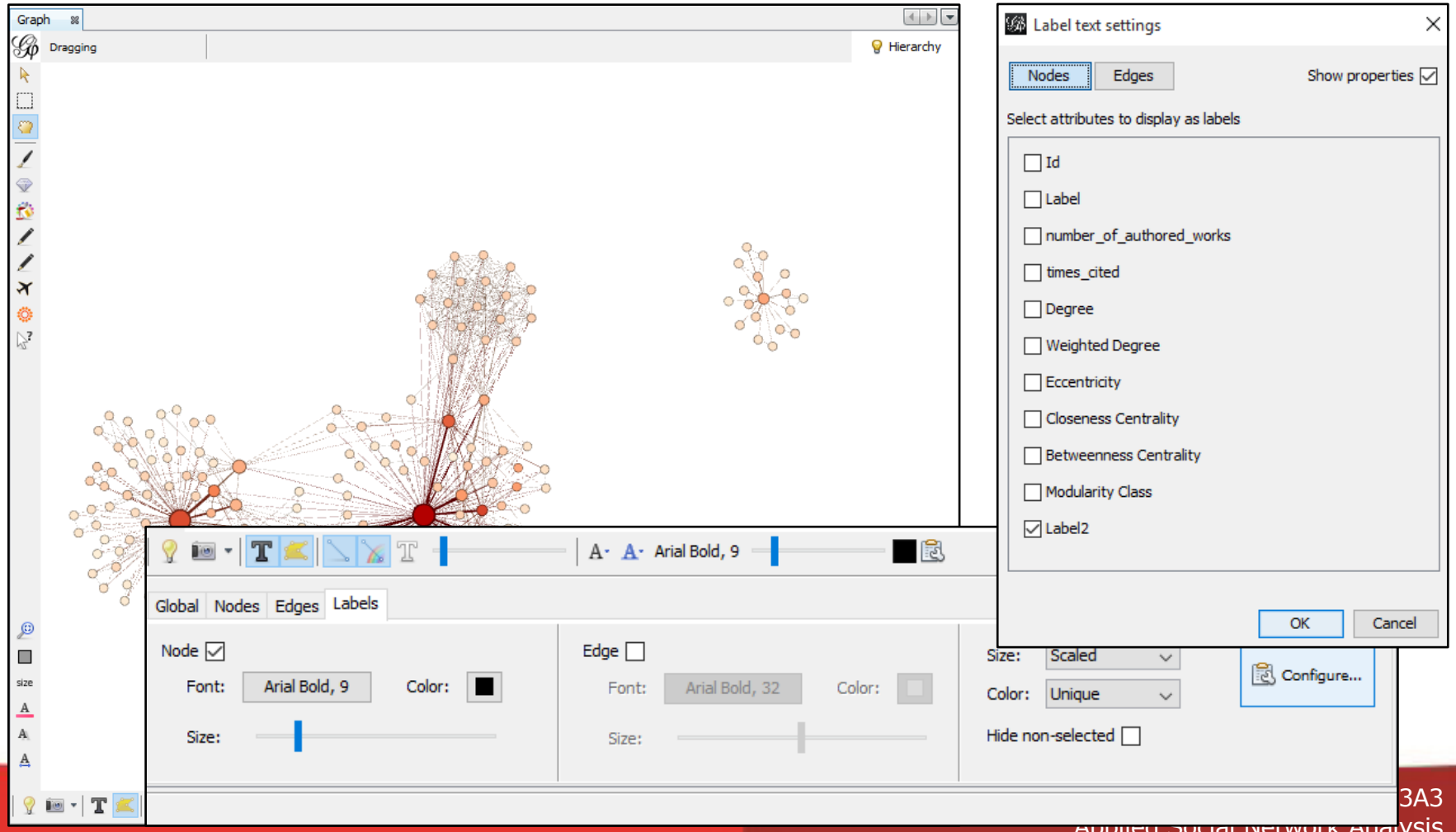


Gephi: Node Labels – Data Laboratory

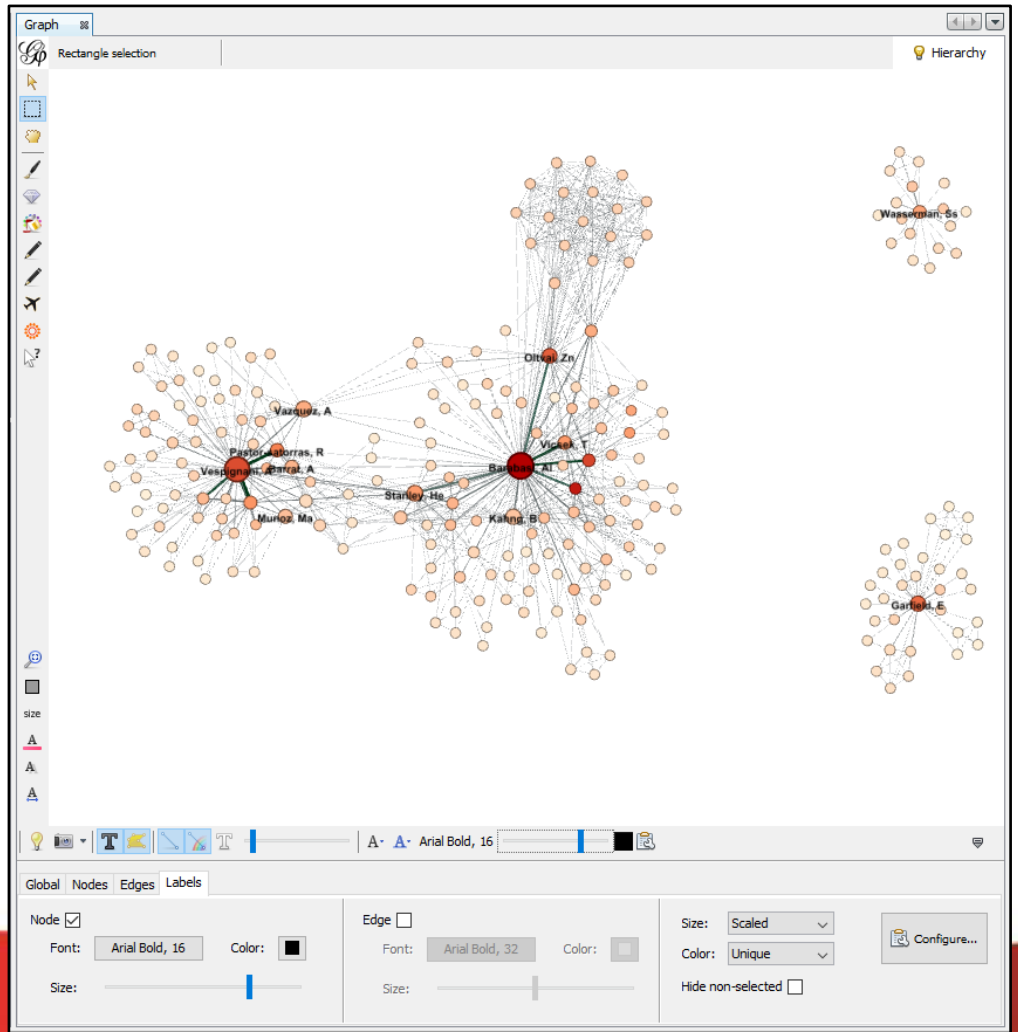


A3
sis

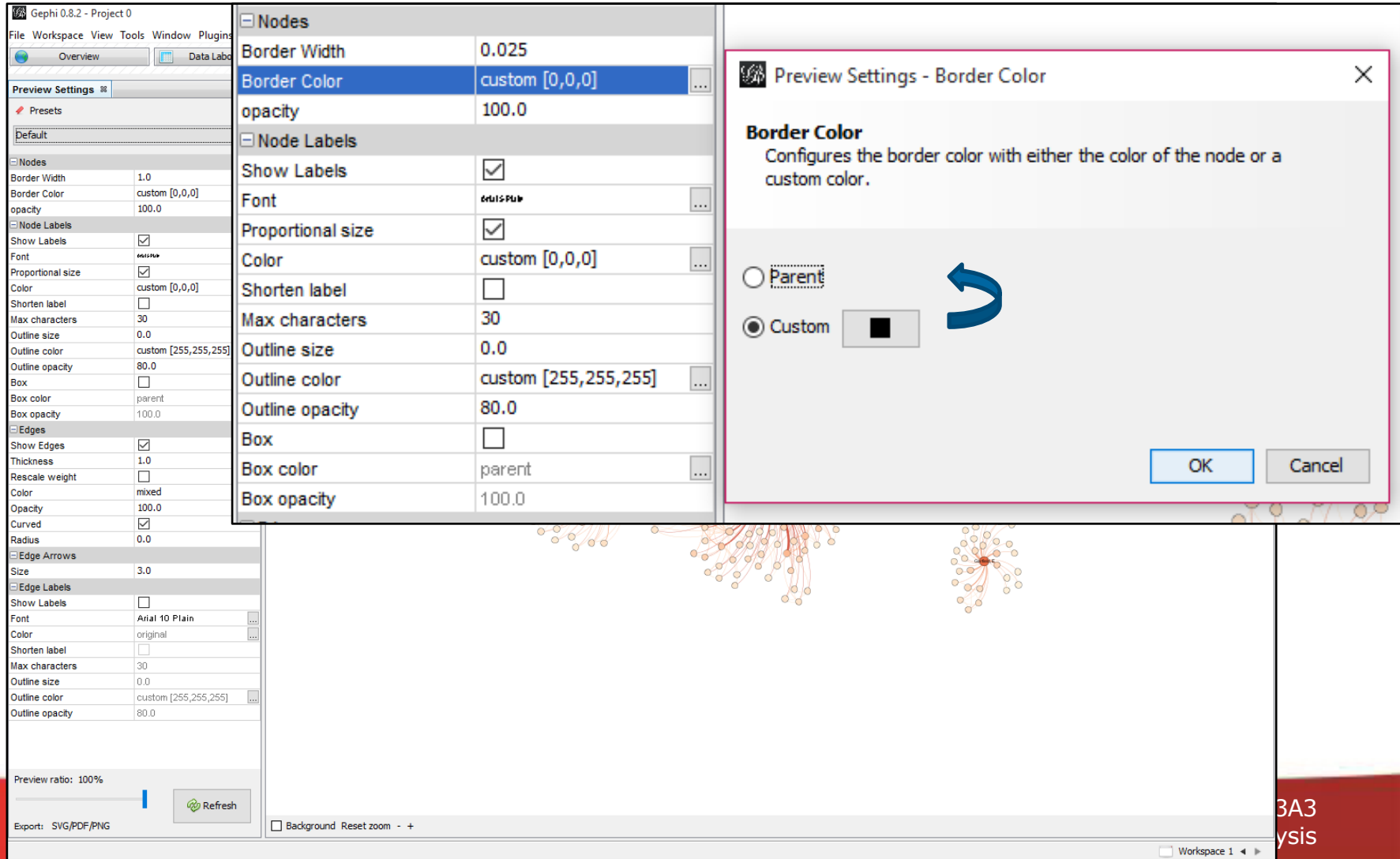
Gephi: Node Labels – Overview



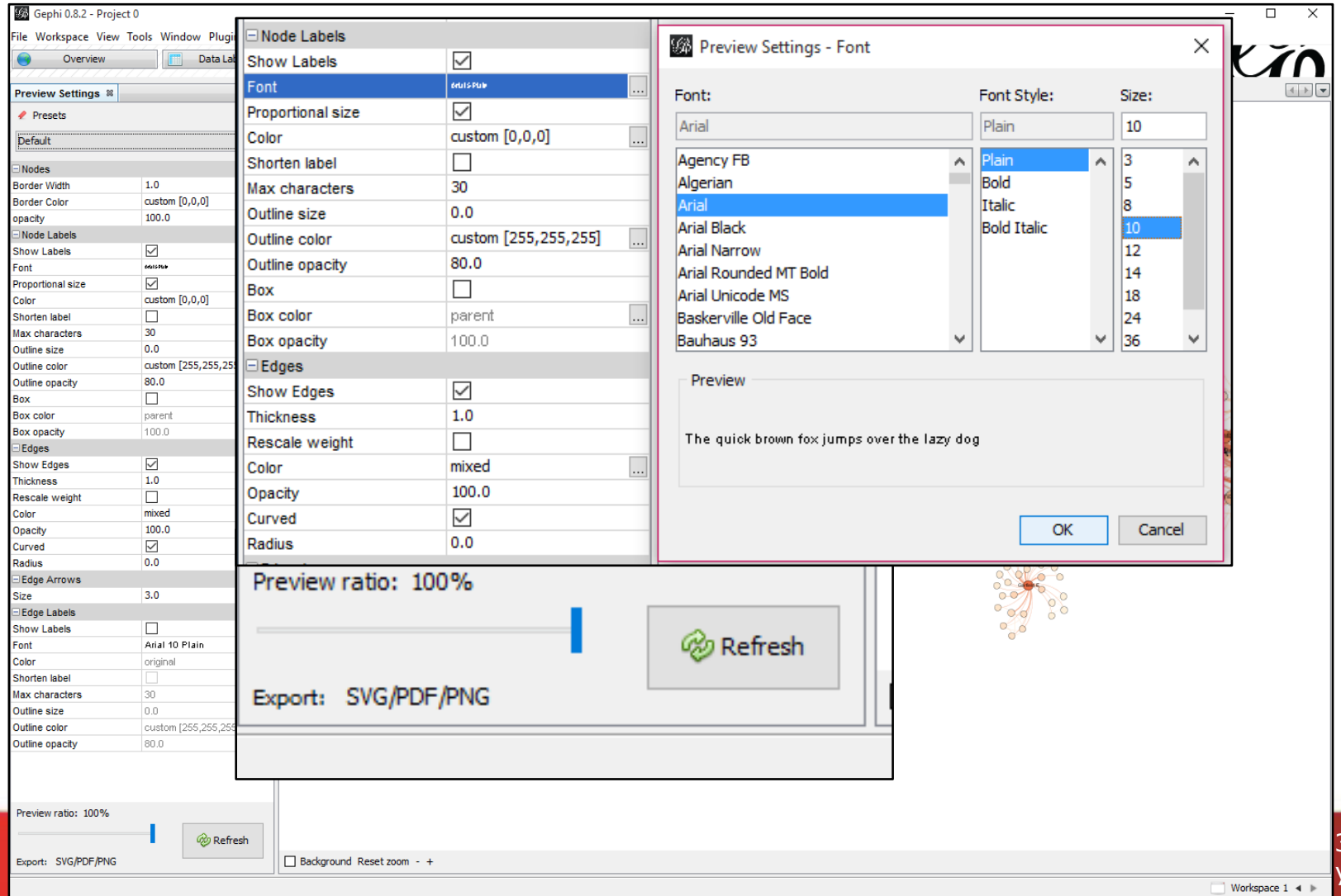
Gephi: Node Labels – Overview



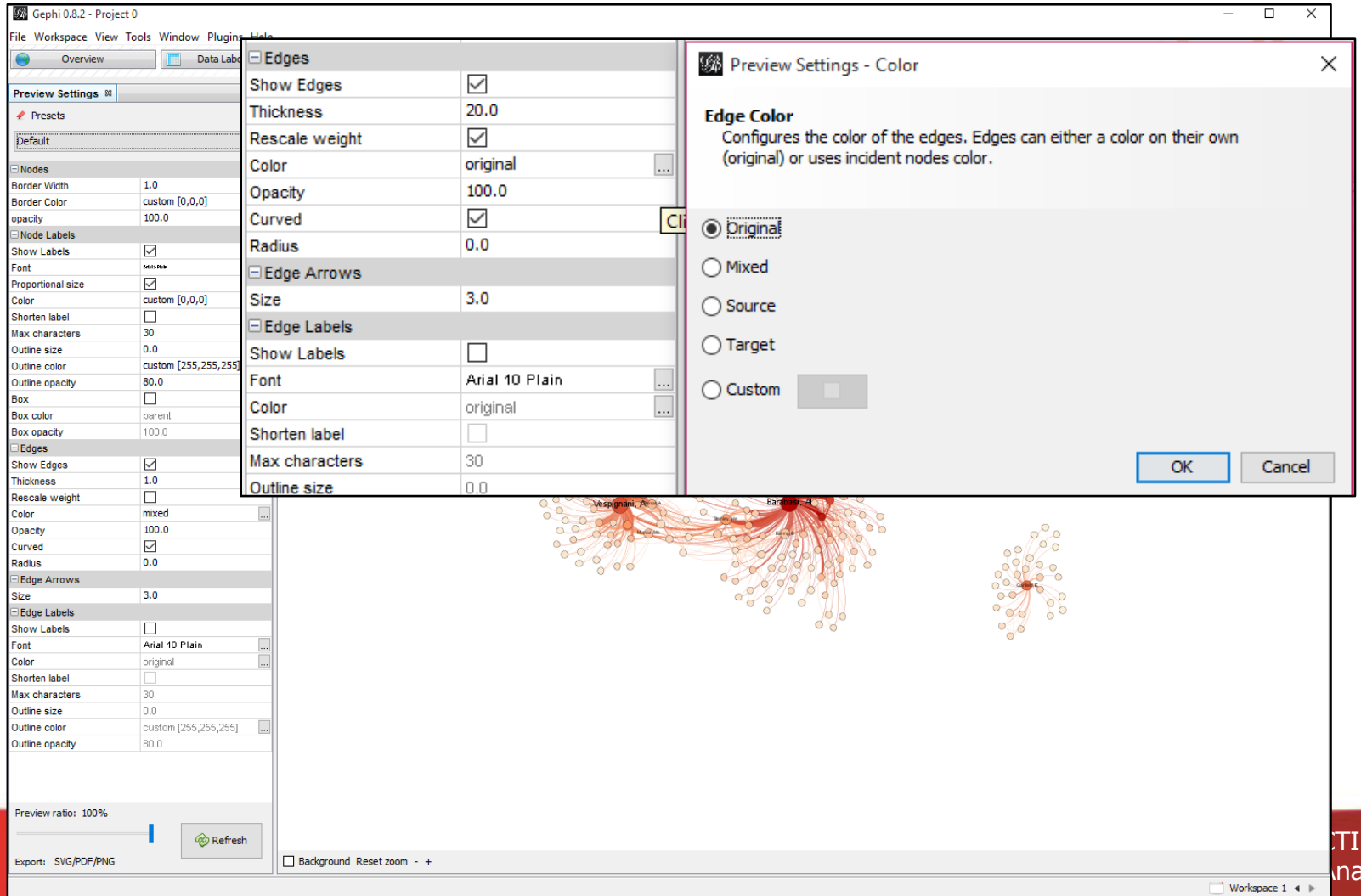
Gephi: Final visualization – Node Parameters



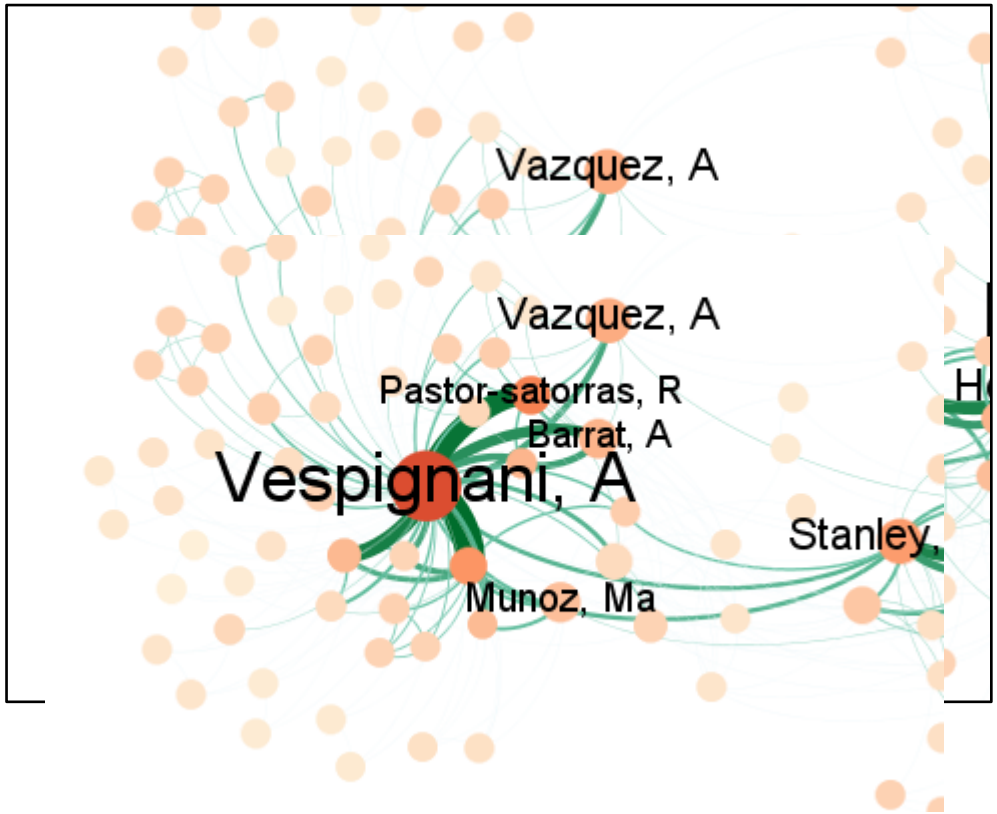
Gephi: Final visualization – Node Parameters

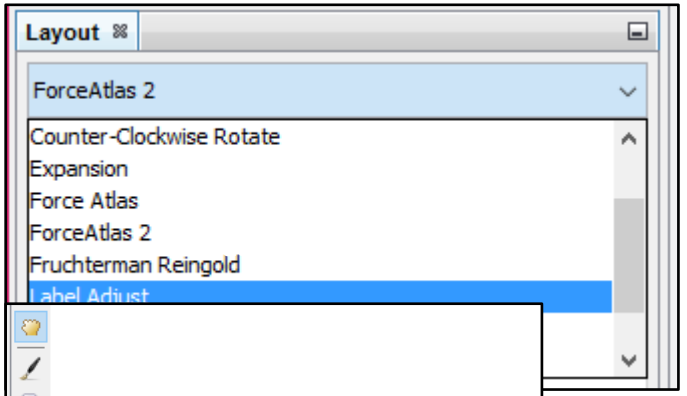


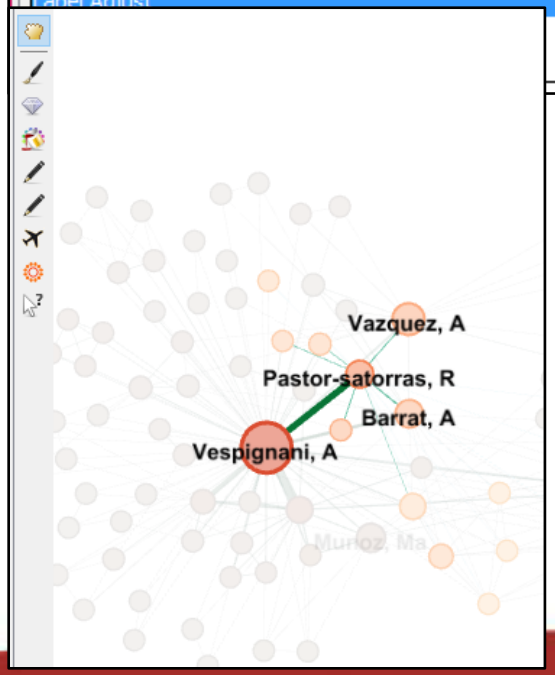
Gephi: Final visualization – Edge Color Parameters



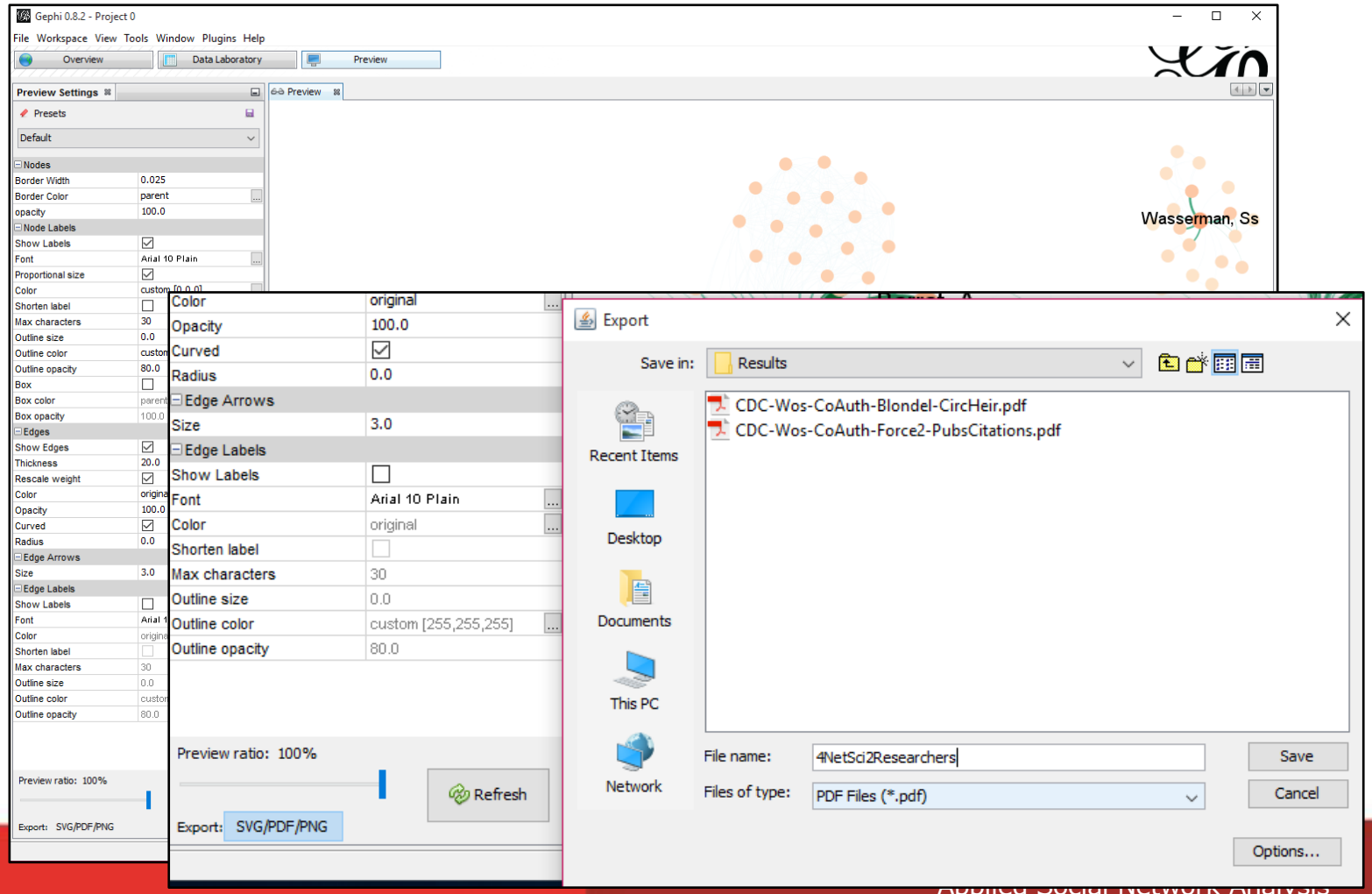
Gephi: Label Adjustments



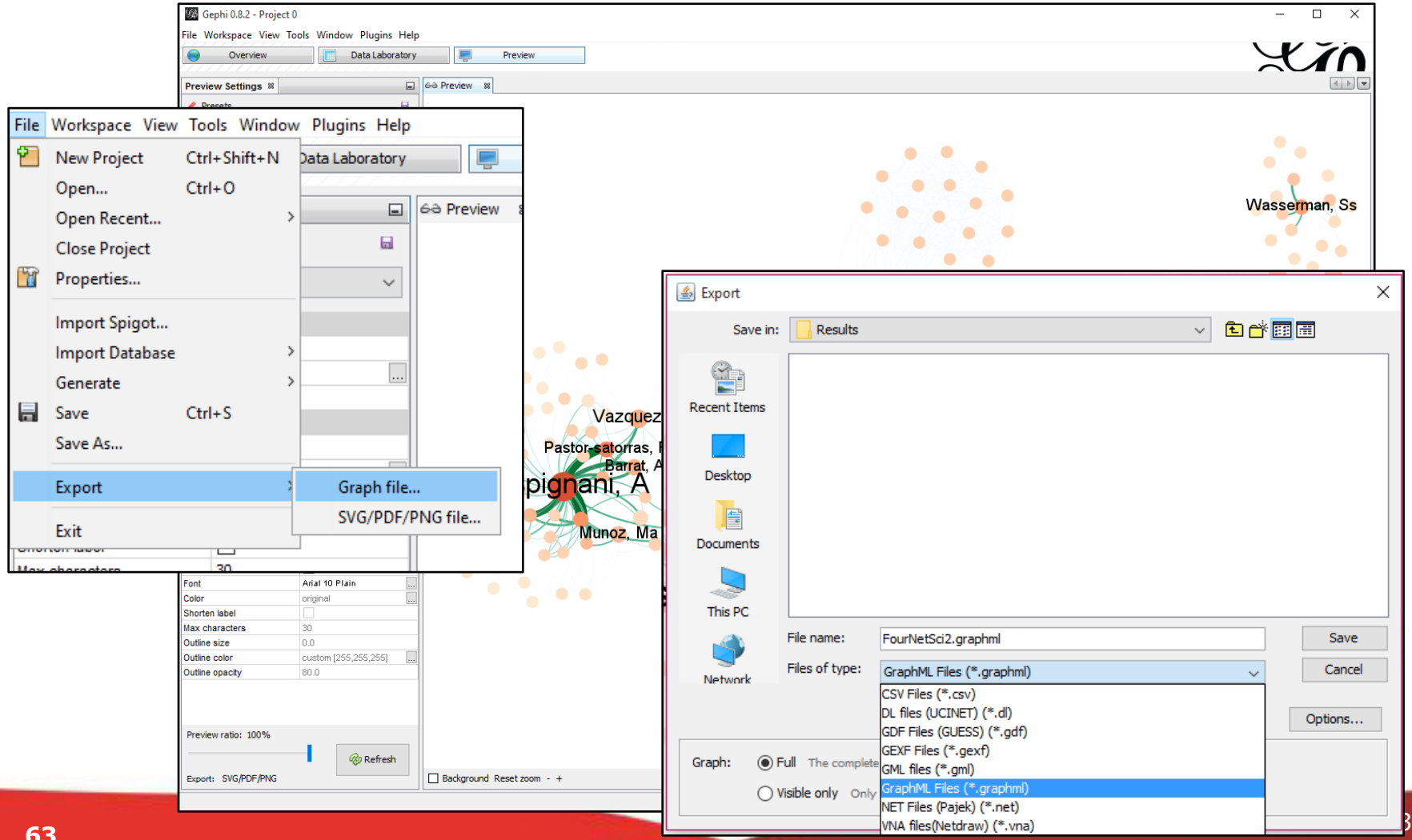




Gephi: Exporting the visualization and networks



Gephi: Exporting the visualization and networks



Let’s practice in Gephi

References

- https://gephi.org/users/

THANK YOU