W-209 - Assignment 5 – Gephi

In this assignment you will use Gephi (https://gephi.org/) to generate a network visualization for data from one of the datasets in the link below.  Import the data and use Gephi to examine relationships among the nodes, edges, and communities in the data. Utilize several encoding techniques (e.g., color, size, directionality, etc.) to highlight patterns in the network, looking for several different patterns. During this exploration process:

* Experiment with different layout algorithms and examine how these affect the network visualization
* Explore centrality measures, community structures, network density, and paths
* Use filtering and labeling techniques as appropriate.

Select **three** distinct views that you’ve created that showcase compelling patterns in the data.  At least one of these views should utilize an additional plugin (Tools > Plugins) in Gephi. Turn in a PDF document to the ISVC submission page that shows a screen capture of each of these three views (with a labeled heading for each).

Please select a dataset from: <https://github.com/gephi/gephi/wiki/Datasets>

Do **not** use the “Jazz musicians network” dataset as this is difficult to work with for the assignment.

# Overview

For my visualization I chose the diseasome dataset from <http://gephi.org/datasets/diseasome.gexf.zip>. The diseasome is a network of disorders and disease genes linked by known disorder-gene associations, indicating the common disease origin of many diseases.

The code and data sources for assignment 4 is available in GitHub at

## Gephi Pre-requisites

The below additional plugins are loaded and used in this assignment

1. Yifan Hu
2. Force Atlas 2

## Data Sources

Shown below are examples of the original data sources for the node and edges table

**Example**: Original Nodes table

|  |  |  |  |
| --- | --- | --- | --- |
| **Id** | **Label** | **type** | **disclass** |
| 55 | Deafness | disease | Ear,Nose,Throat |
| 888 | Enlarged vestibular aqueduct | disease | Ear,Nose,Throat |
| 889 | Pendred syndrome | disease | Ear,Nose,Throat |
| 1329 | MTP | gene | gene |
| 1338 | CNGA3 | gene | gene |

**Example**: Original Edges table

|  |  |  |  |
| --- | --- | --- | --- |
| **Source** | **Target** | **Type** | **Id** |
| 30 | 1420 | Directed | 36 |
| 30 | 1421 | Directed | 37 |
| 30 | 3553 | Directed | 38 |
| 30 | 3539 | Directed | 39 |

# Visualization clusters of related diseases and genes

For the first visualization, I use the original dataset and extract clusters of related disorders and genes.

## Steps

1. Create a new workspace
2. Import the nodes and edges datasets CSV file