**INFORMATION VISUALISATION**

**GEPHI Assignment**

**Name : Harsh Anand Awasthi**

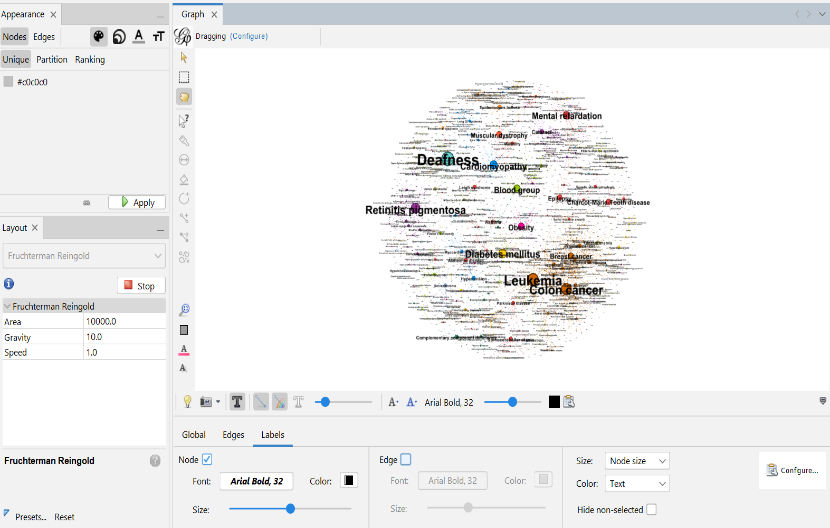
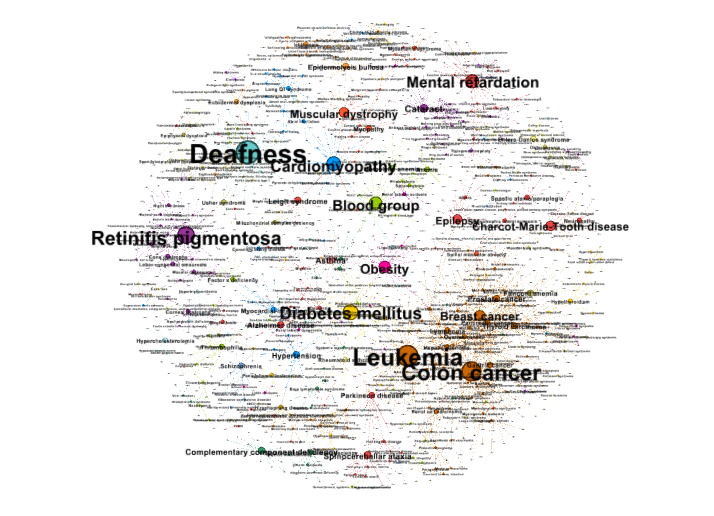
**RA NO. : RA2211027010020 (BATCH 1)**

**Looking at a network of disorders and disease using Gephi**

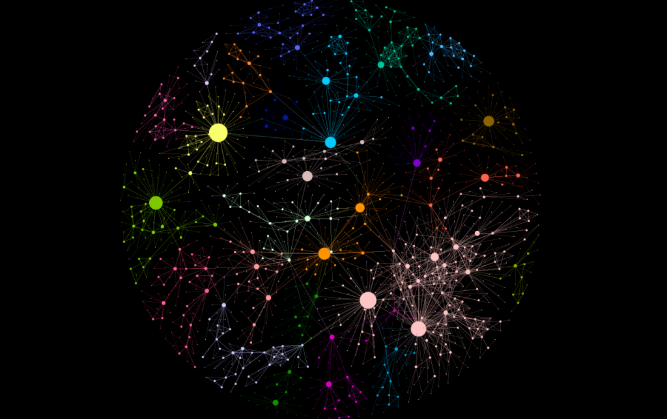
The Diseasome dataset came as a .gexf file, which was a graph file. In the overview tab, I could choose the type of layout I wanted. I chose to do a Fruchterman Reingold layout. This turns the visual in a circle, and places the related network around the same area. From here I went to statics to do a modularity run using weights, which detects the community. To make the colors consistent by community, I went to appearance > partition by modularity and changed the colors accordingly. This made my visualization more colorful and was clear where the community were. In the preview, I could do more modifications to the visuals; like labeling, font and color. This was simple to do compared to the overview. I made another network visual, but with a different layout, Force Atlas 2. This basically makes a linear model, that gives distance with different communities. After using Force Atlas2 layout, I took the same steps in appearance to improve the visuals. When all the visualizations were done, I exported the files and saved it as JPEG images.

**Frutchermen Reginald Layout**

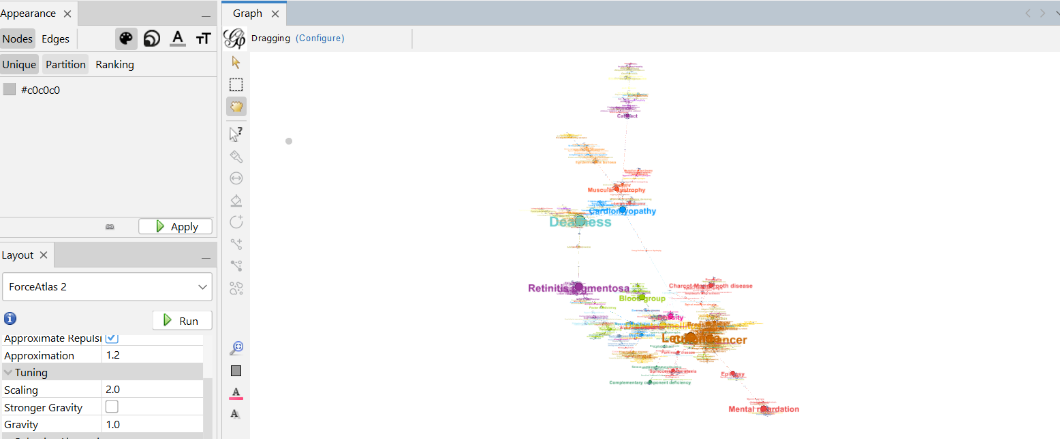
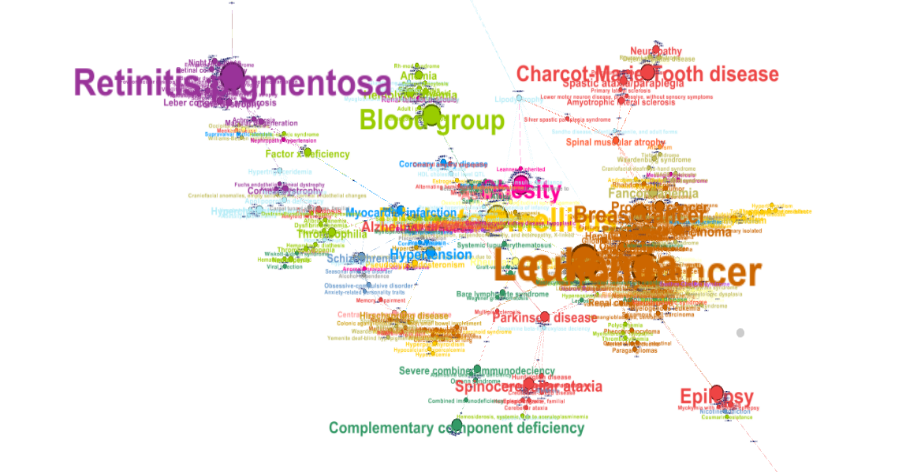
This is the Fruchterman Reingold layout. The different communities are grouped by color. The label size depends on the size of the node. It is clear that cancers are the most dominant disease out of them all. I did not know deafness had a great portion too! There are certain clusters around the layout, like cancer, diabetes, mental health and more. I removed the ‘gene’ labels because they made the visual even more text heavy than it is right now. It looks better with less text. I decided to keep the others because I wanted to represent how dominant some diseases are compared to others.

****

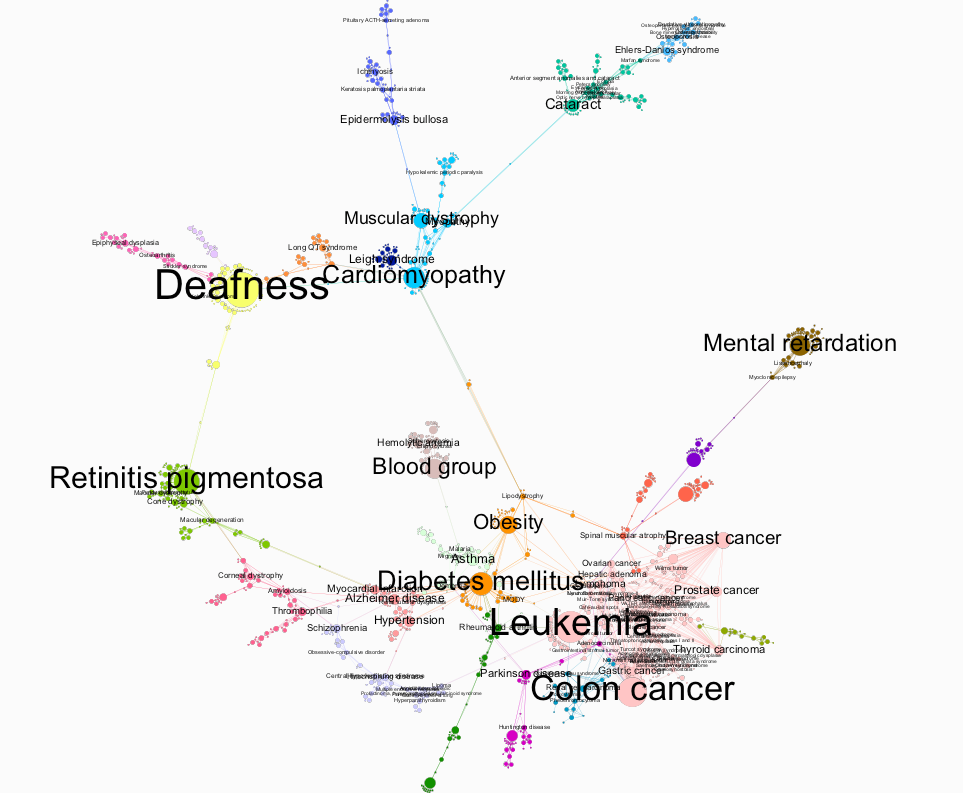
**Frutchermen Reginald with black background**

****

**Force Atlas 2 Layout**

****

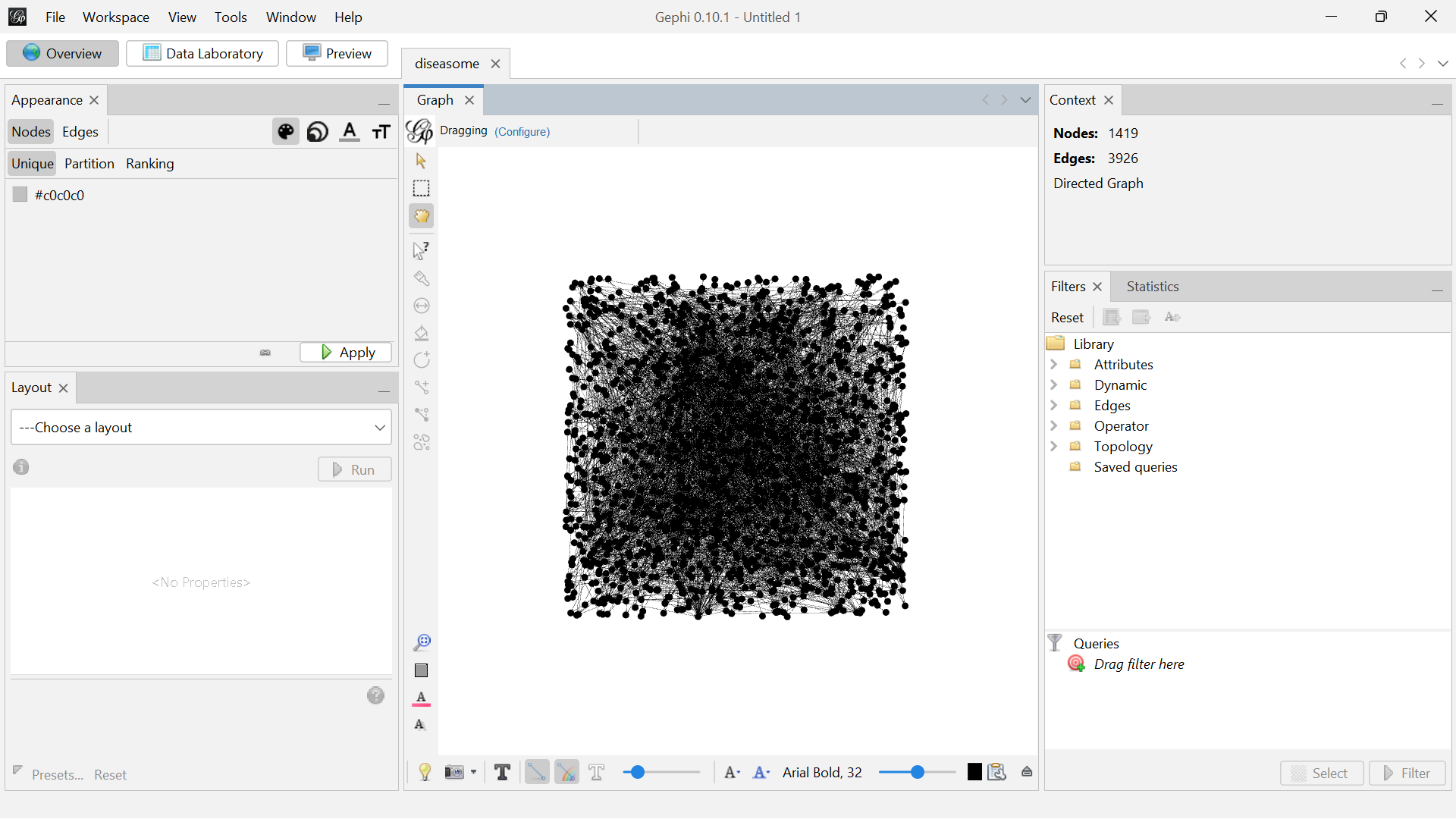
Force Atlas 2 made the visualization more clustered compared to Fruchterman Reingold layout. I had to remove a lot of labels that had in-degree of 1-5 because it made the dominant labels hard to read. This way the visual looked cleaner. Like the other layout, this also clearly delivers the message that cancer is one of the most common disease compared to others that exist today

****

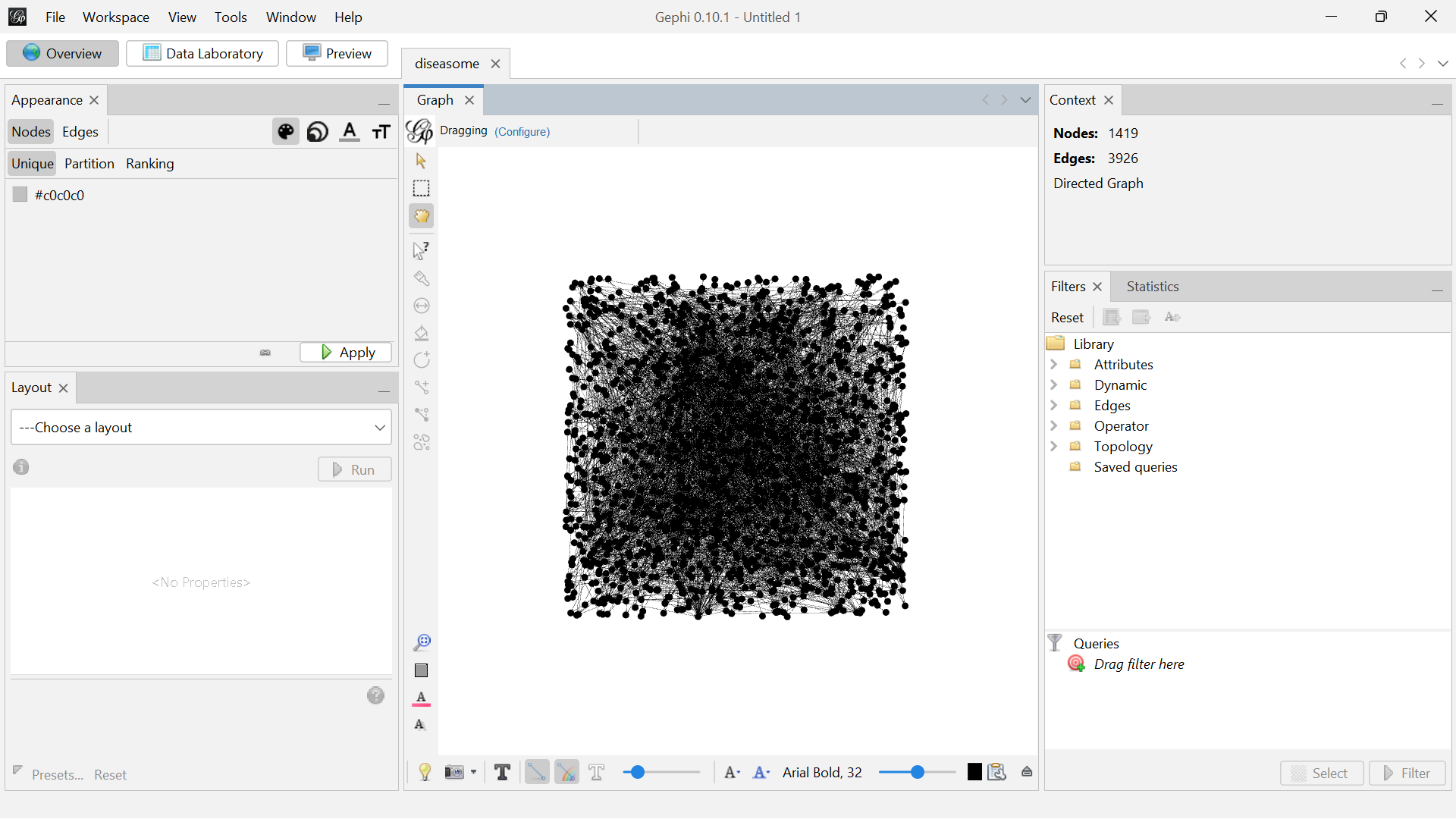
Here are some extra things I tried like different background, text and node colors and also selected major diseases like cancer and deafness for better focus

****

A network of disorders and disease genes linked by known disorder–gene associations, indicating the common genetic origin of many diseases. Genes associated with similar disorders show both higher likelihood of physical interactions between their products and higher expression profiling similarity for their transcripts, supporting the existence of distinct disease-specific functional modules.

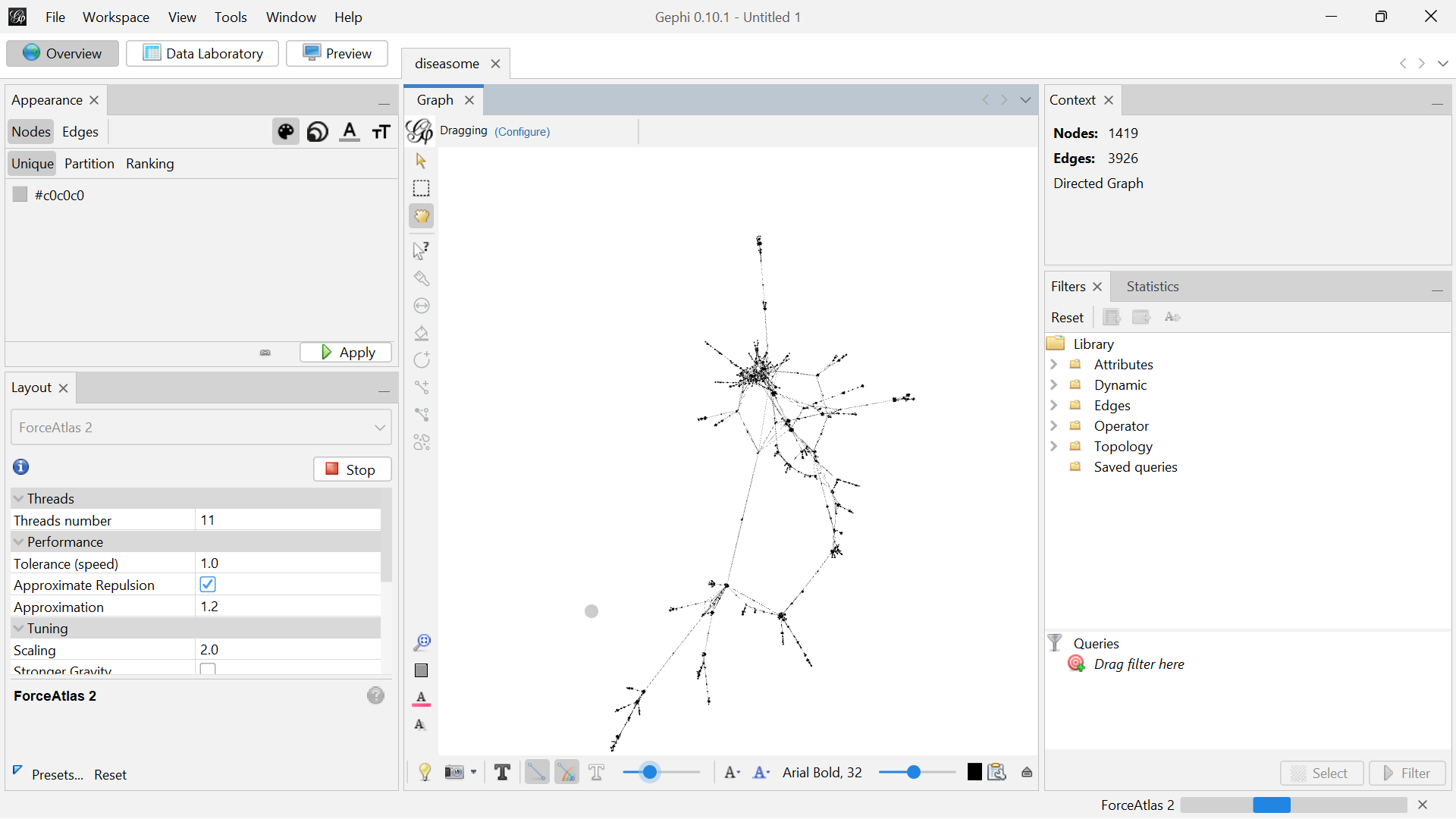


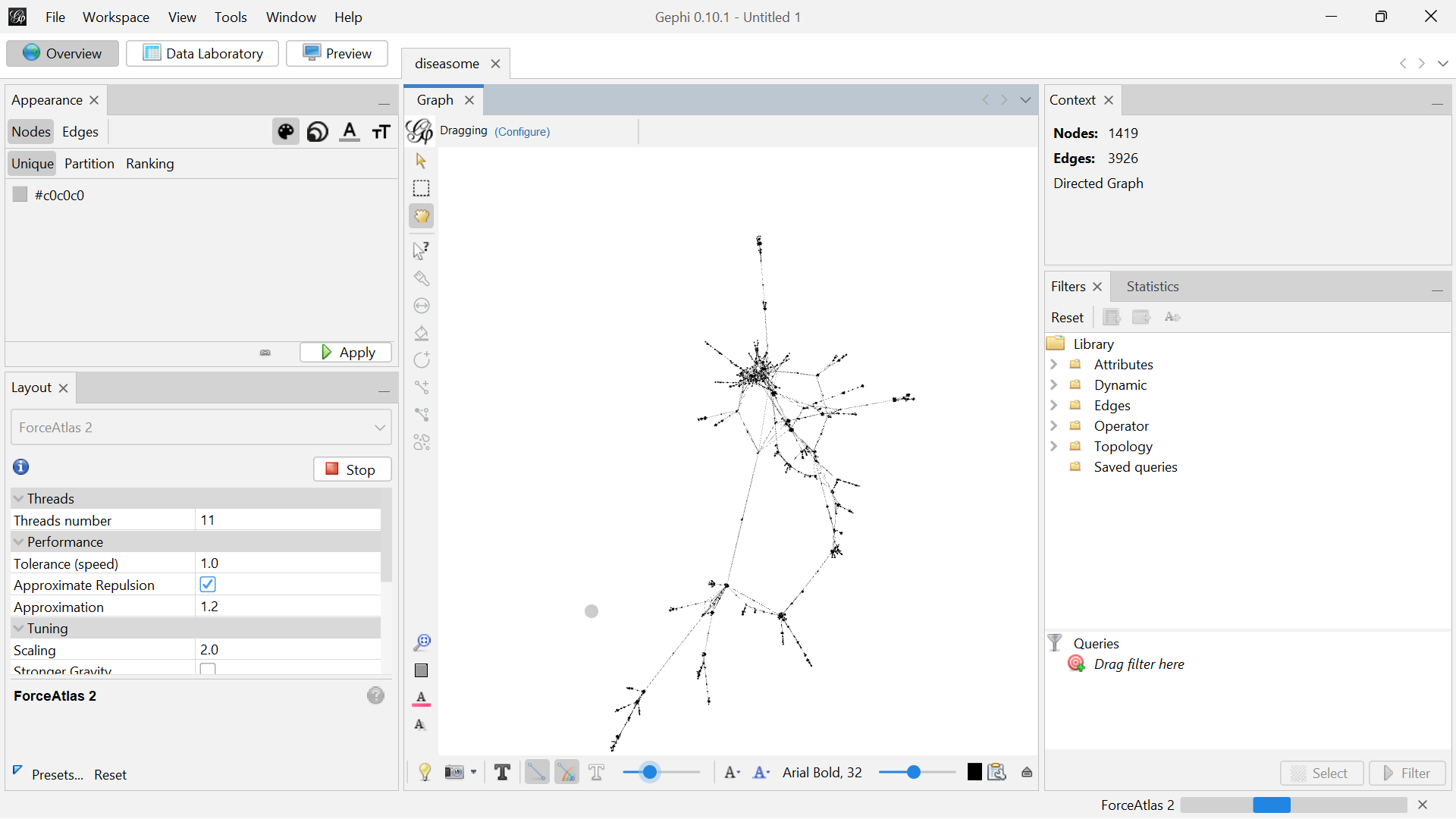
**Initial Visualization:**

****

**Layout Alteration:**

**Force Atlas 2** is a popular layout algorithm used in Gephi, a software for network visualization and analysis. It’s especially suited for visualizing large, complex networks. The algorithm positions nodes based on a force-directed approach, simulating physical forces like attraction and repulsion to reveal the network’s structure in an intuitive way.

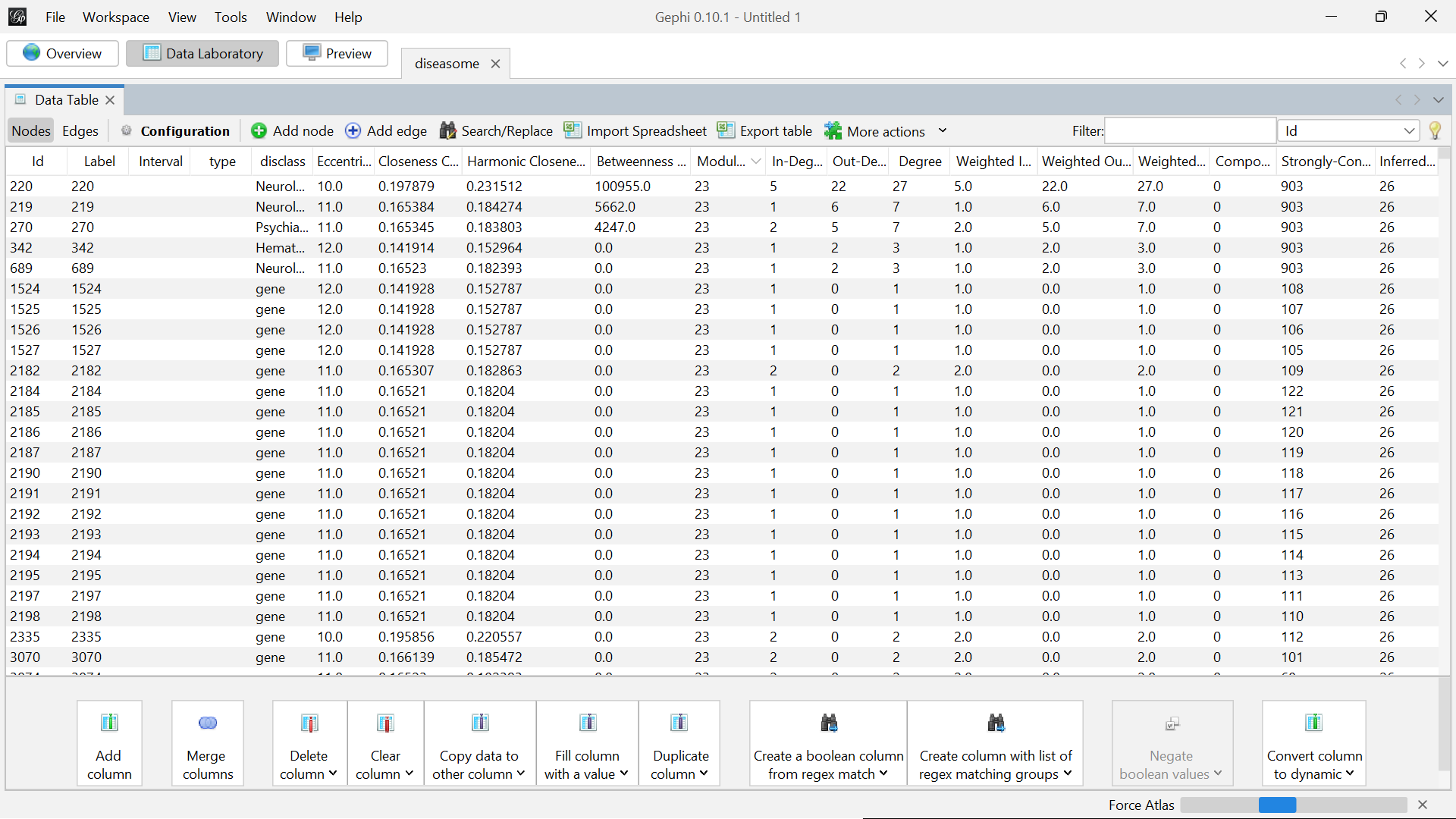
****

****

**Basic Statistics:**

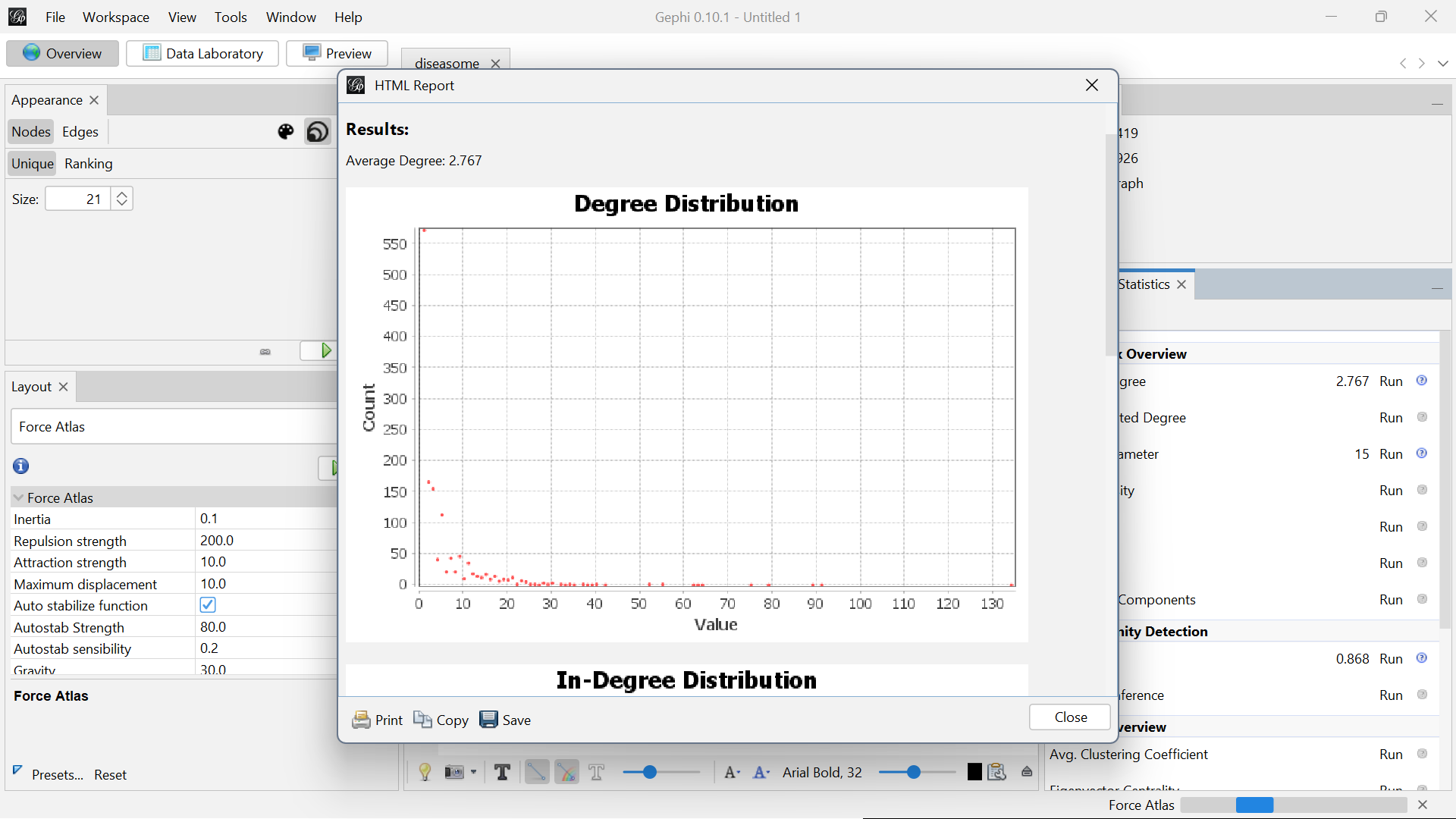
**Data Laboratory:**

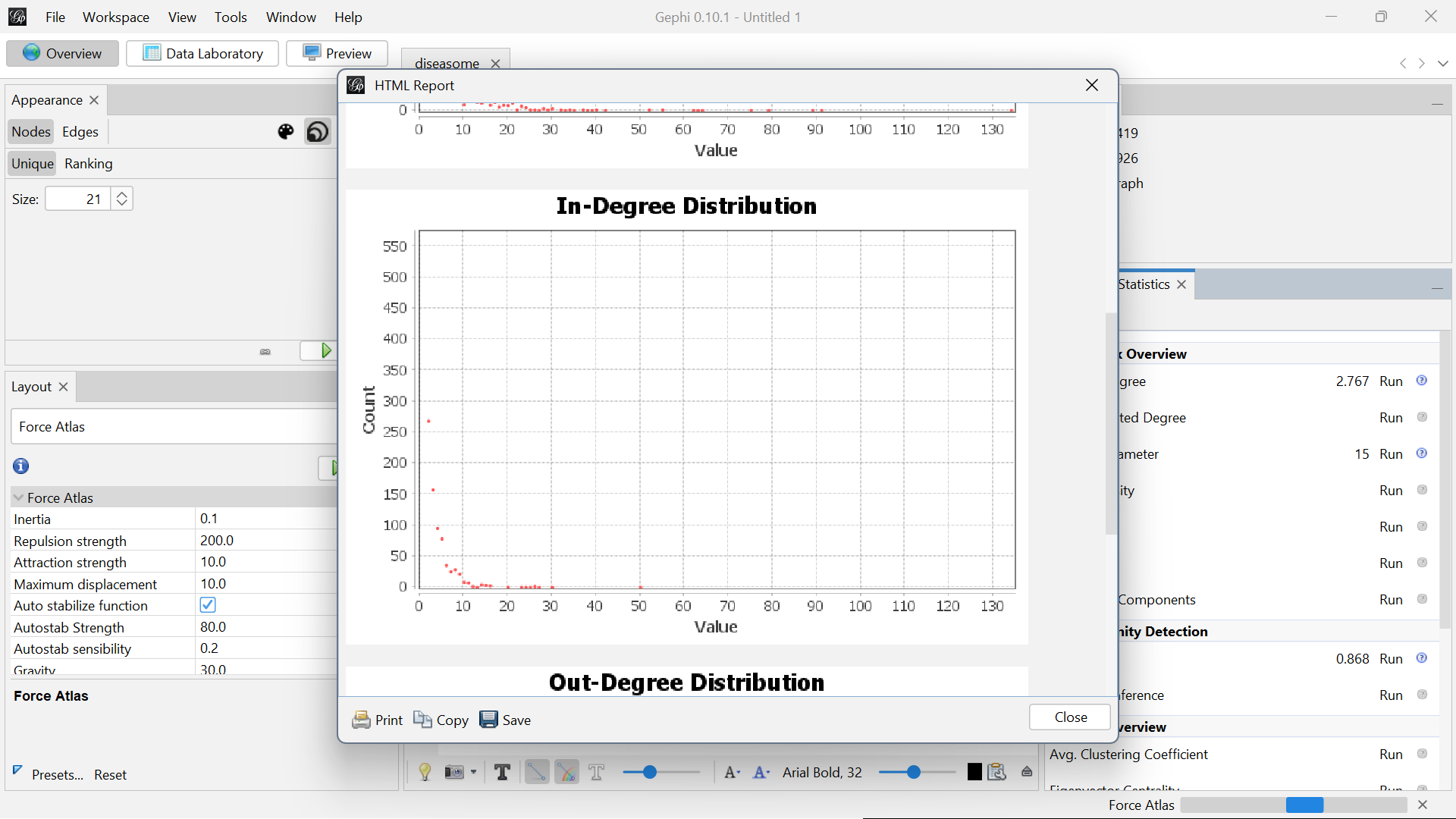
In Gephi, the Data Laboratory is a workspace that allows you to interact with your data in a tabular format, similar to how you would work with spreadsheets. It provides a user-friendly interface to view, edit, and manipulate your graph’s node and edge data directly.

****

**Average Degree:**

Average Degree is a network metric that represents the average number of connections (edges) each node has within the graph. It is a fundamental measure that gives an idea of how connected the nodes are on average.

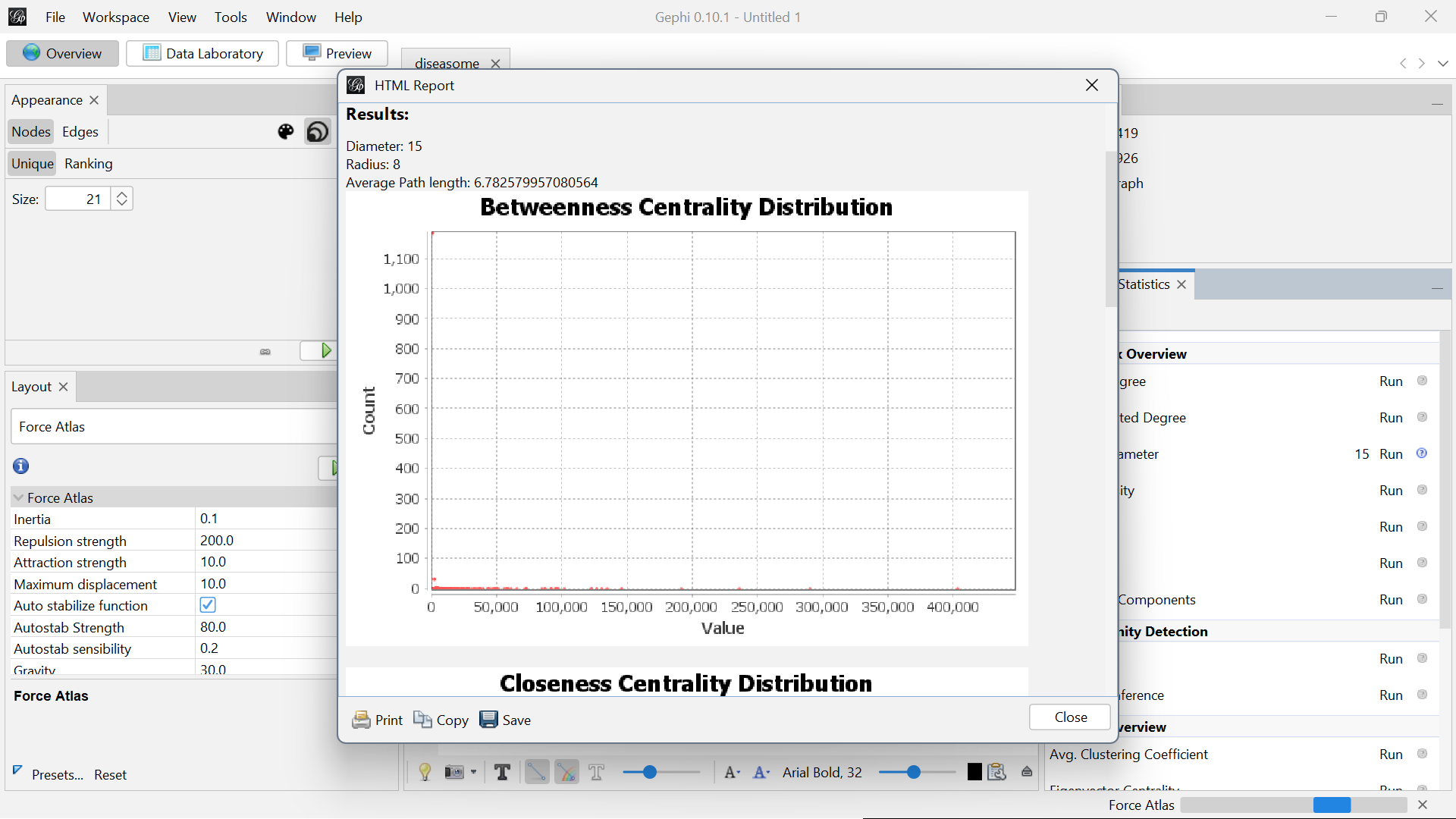
****

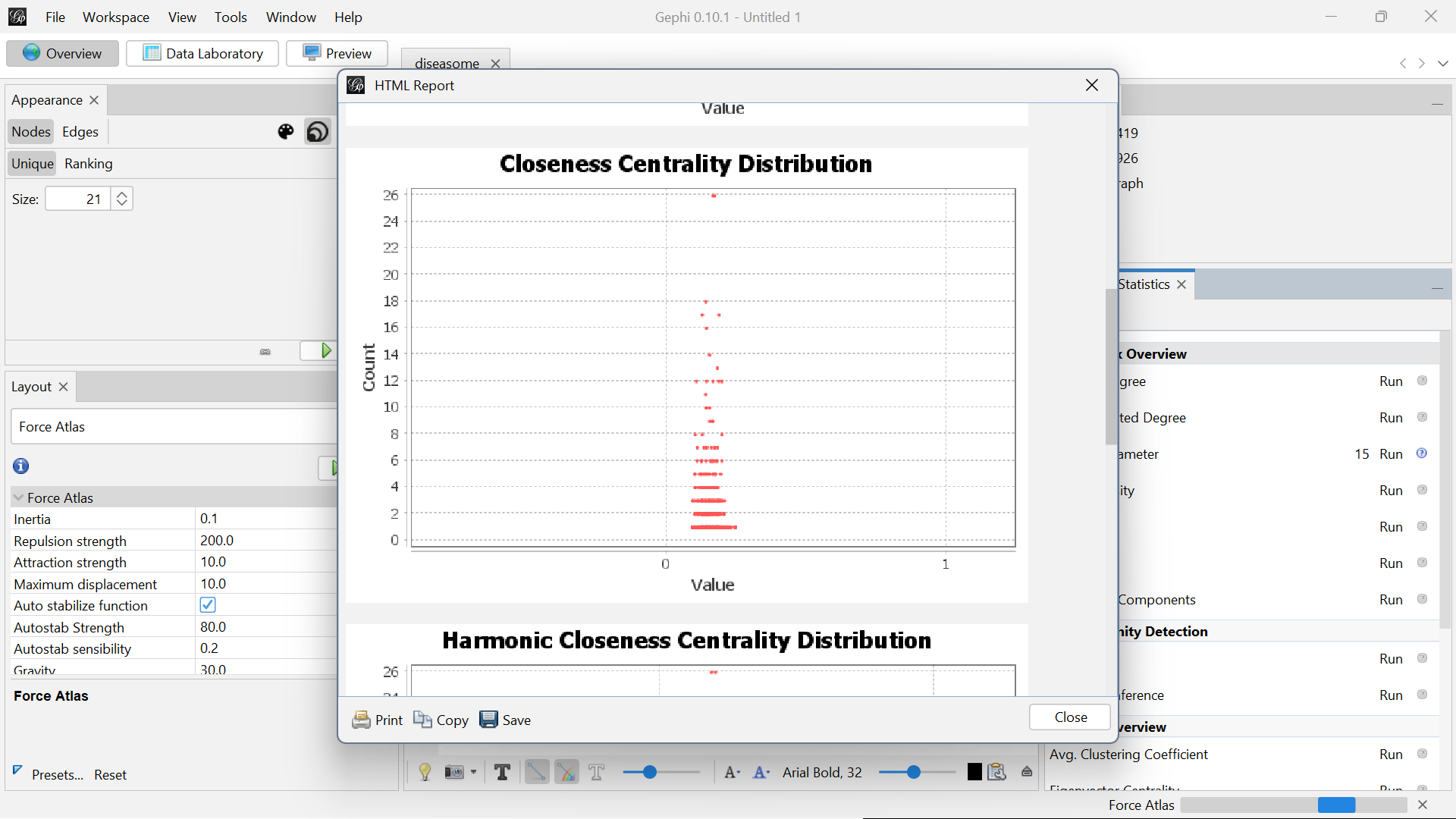
****

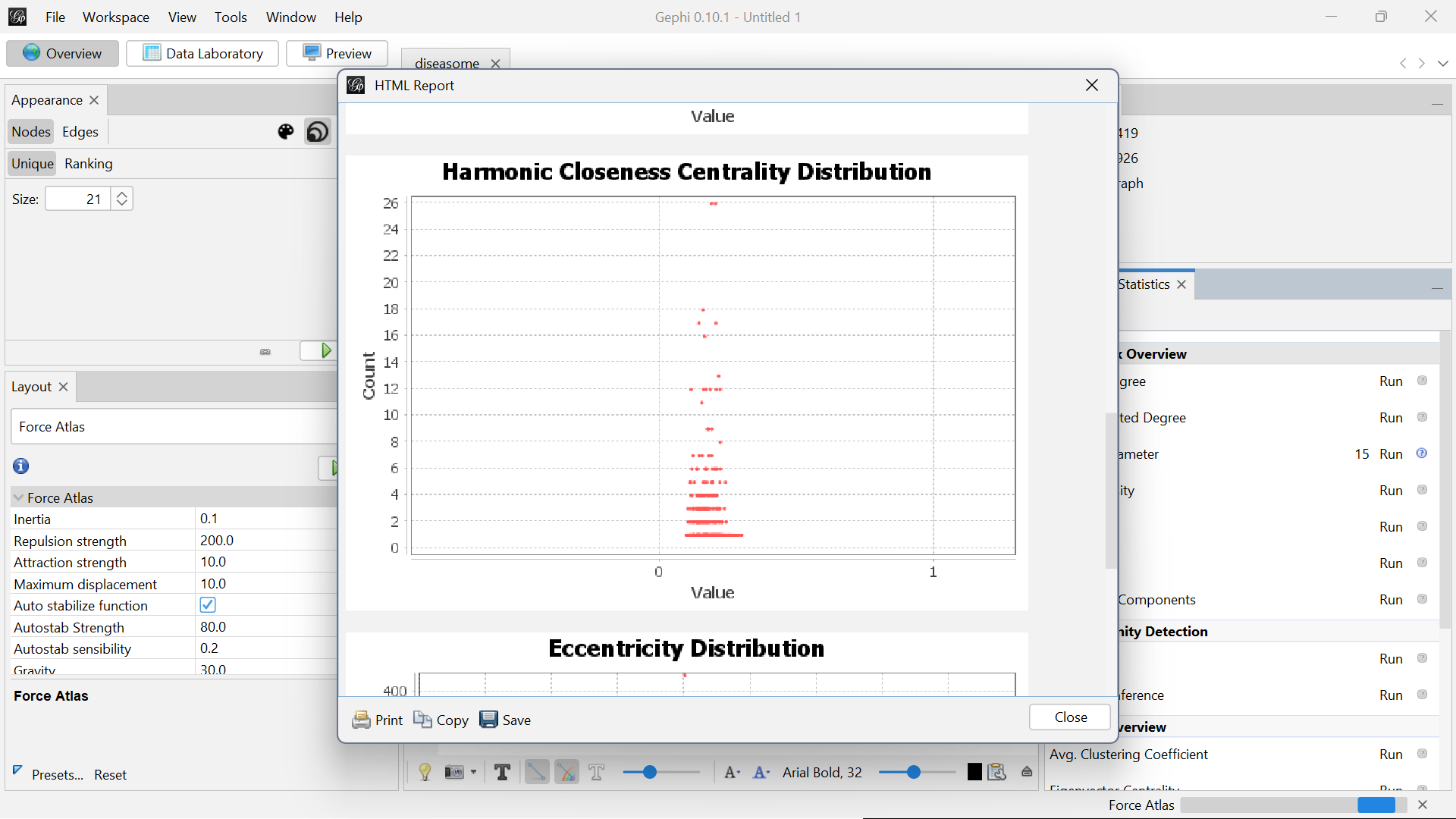
****

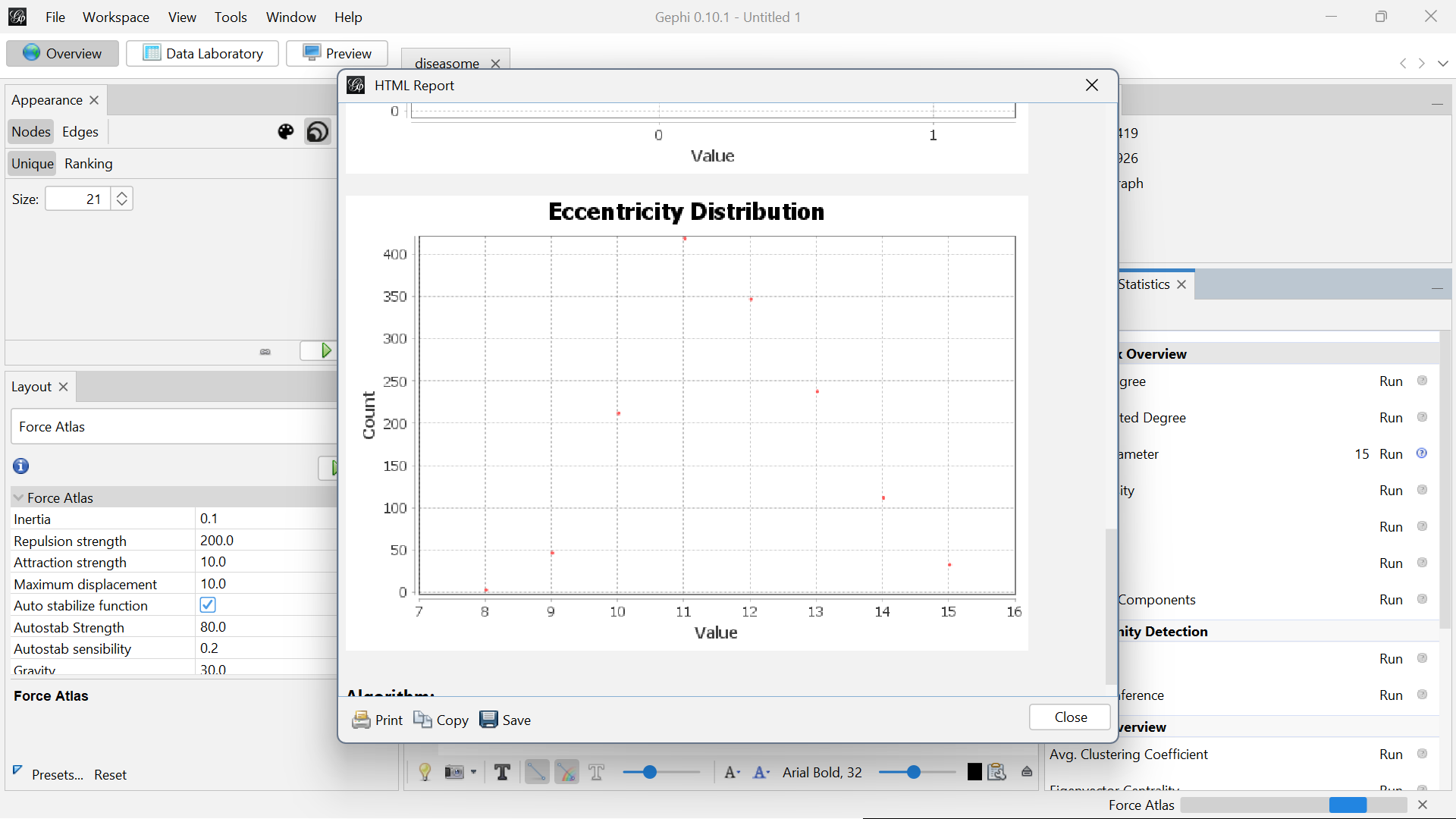
**Network Diameter:**

Network Diameter is a measure of the longest shortest path between any two nodes in a graph. It gives an indication of the "size" of the network in terms of how far apart nodes can be from one another.

****

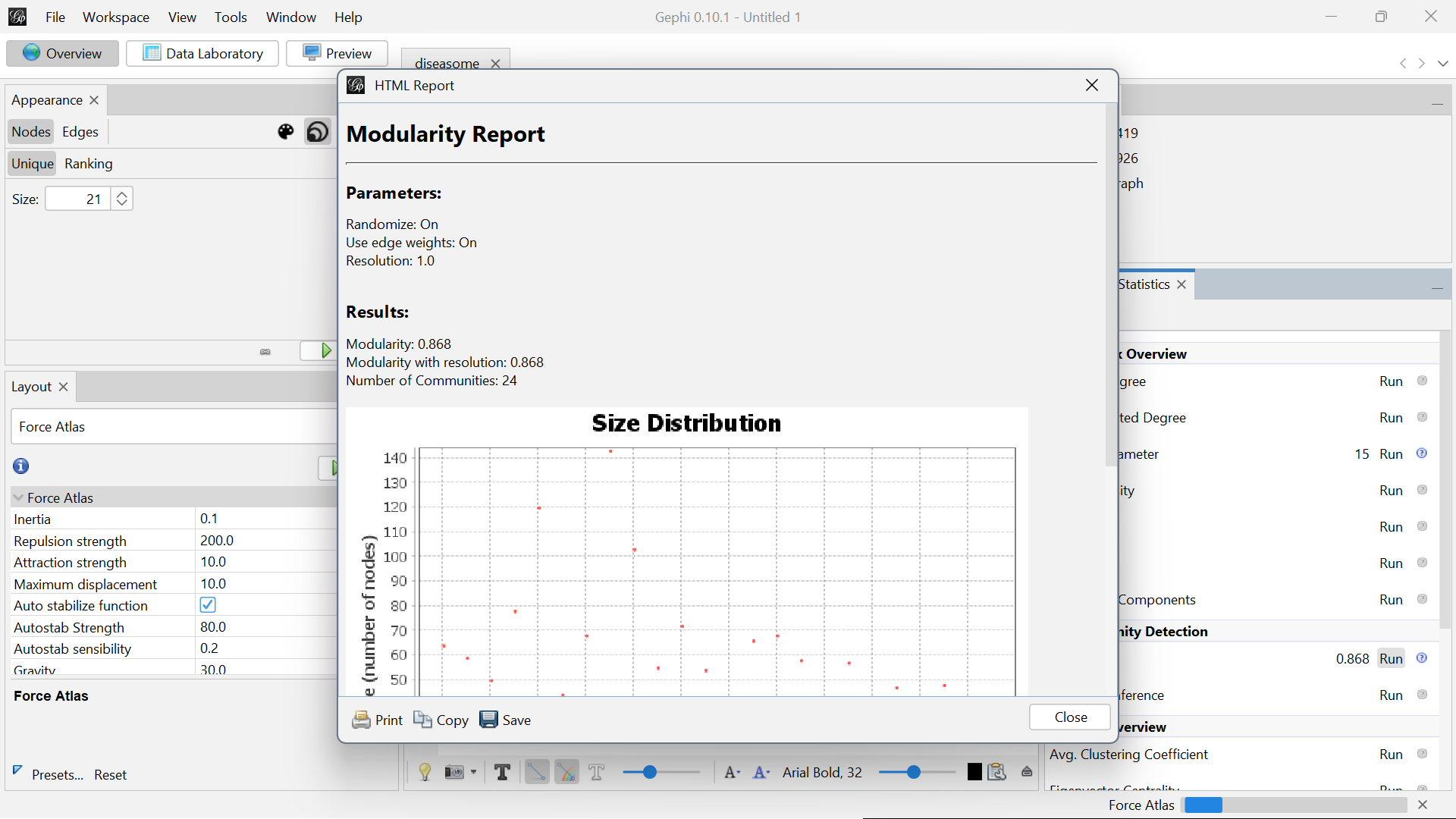
****

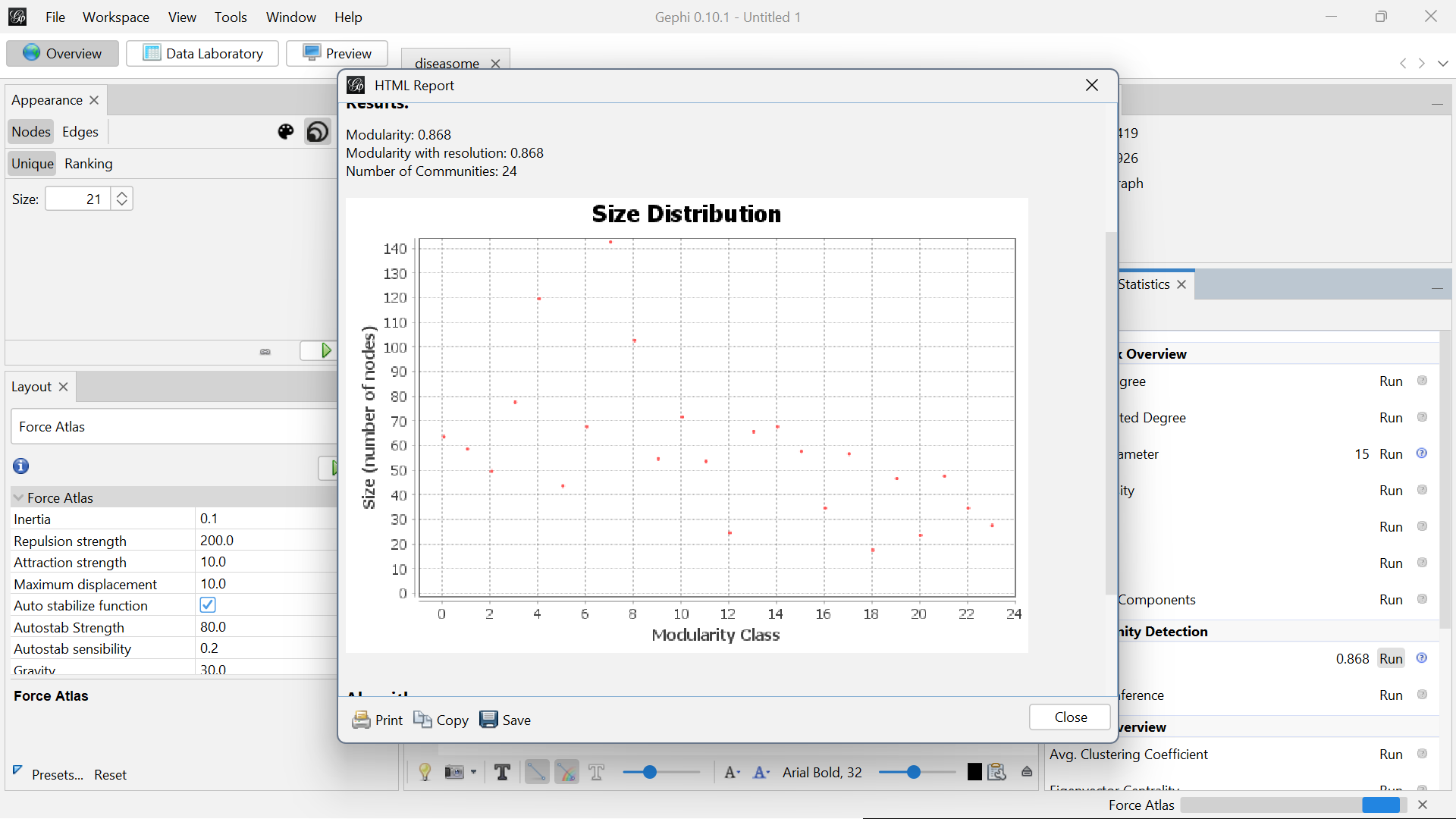
****

****

**Modularity Report:**

Modularity Report provides information about the community structure within a network. Modularity is a measure that evaluates the strength of division of a network into modules or communities. These communities are groups of nodes that are more densely connected internally than with the rest of the network.

****

****

**Appearance Settings:**

**Based on Nodes:**

Appearance settings allow you to customize how nodes (and edges) are visually represented in your graph. These settings help you highlight key patterns and insights by adjusting node size, color, labels, and more based on different attributes.

Appearance Settings Based on Nodes:

1. Node Color:

You can color nodes based on:

* Partition: Assigns colors based on categorical data (e.g., communities, types, categories). Each distinct value is assigned a different color.
* Ranking: Colors nodes based on a numerical or ordinal attribute (e.g., degree, betweenness centrality). You can set a gradient or spectrum to represent different ranges of values.
* Attribute Values: Directly map colors to specific attribute values in your dataset.

1. Node Size:

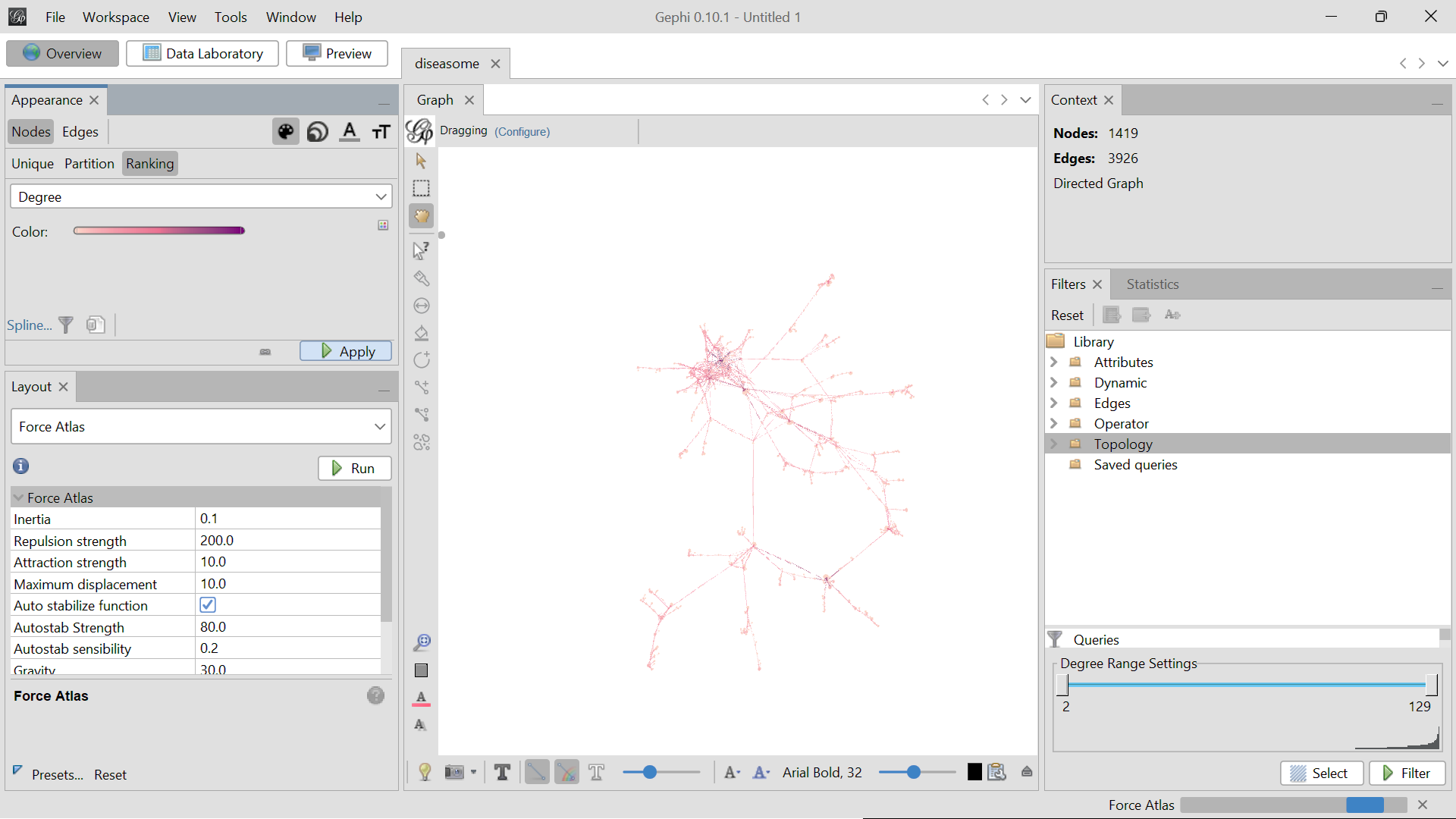
You can scale the size of nodes based on:

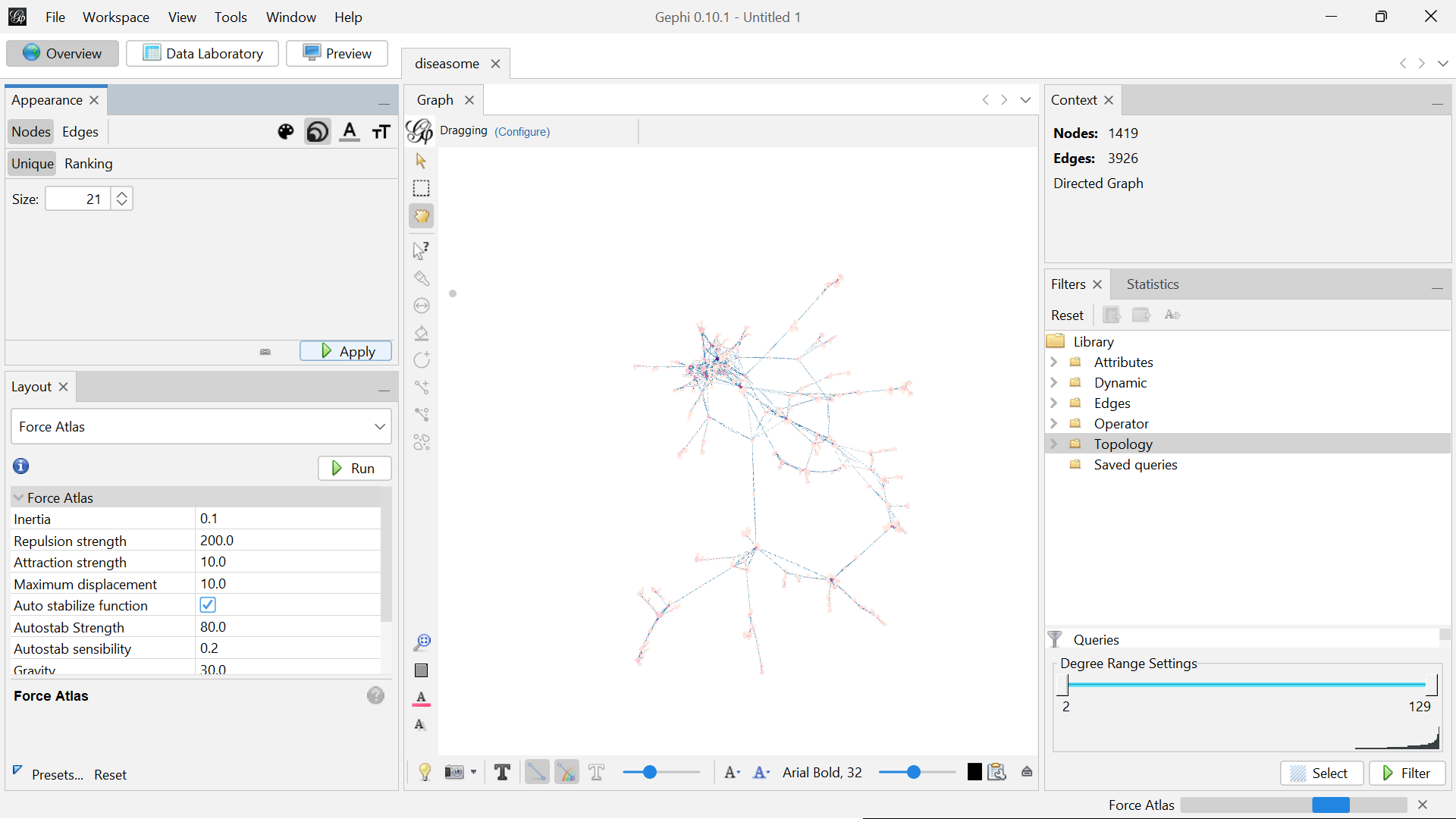
* Ranking: Adjusts node size according to a numeric attribute like degree, centrality, or other metrics. Larger values result in larger node sizes, helping to visually highlight more important or connected nodes.
* Fixed Size: Assign all nodes the same size, useful when you want uniformity in node representation.

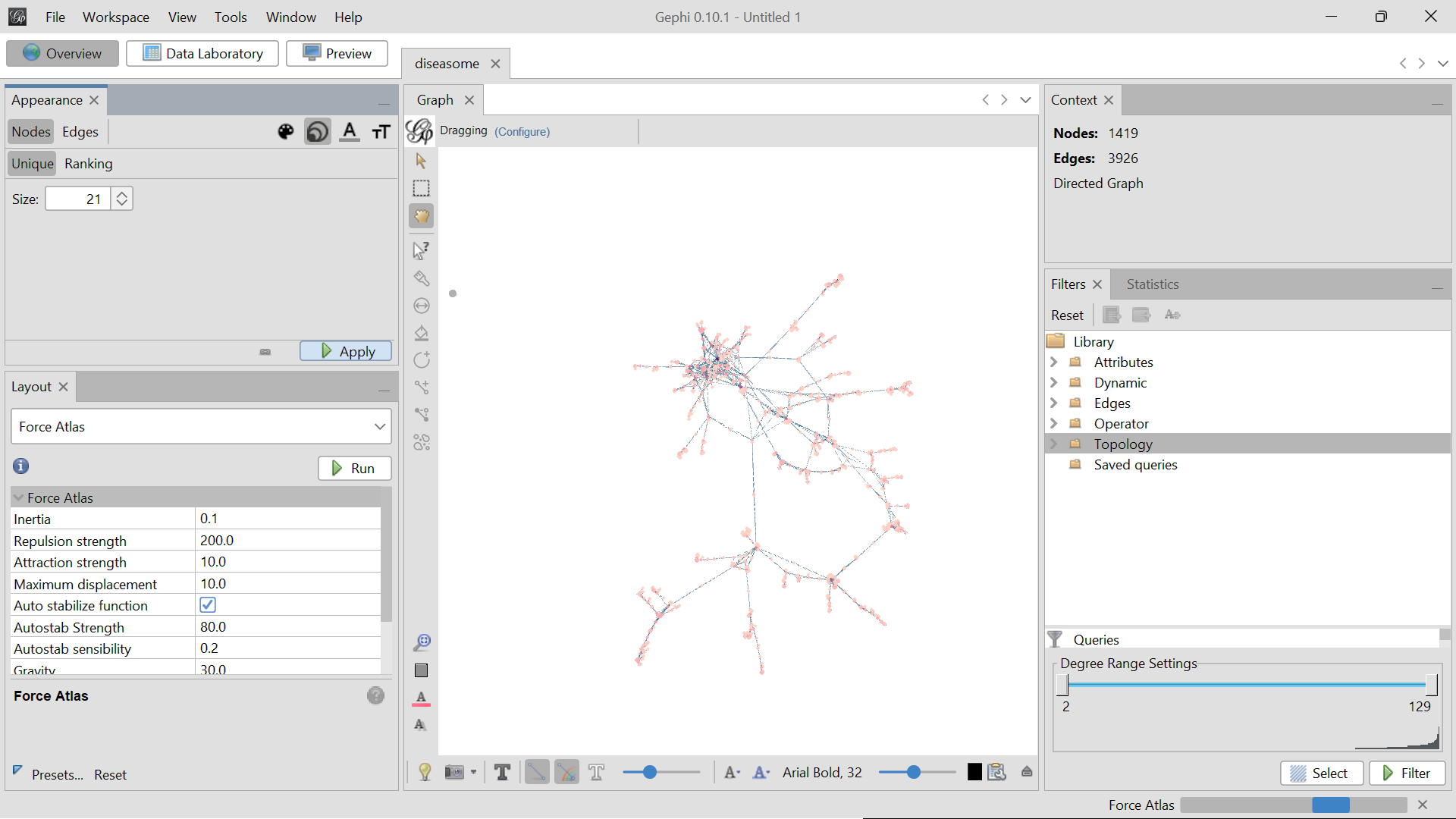
1. Node Label:

* Label Visibility: Toggle node labels on or off.
* Label Size: You can set label sizes to scale based on attributes (e.g., a node with a higher degree could have a larger label).
* Label Content: Choose what information is displayed as a label, such as a node’s name, ID, or any other attribute.
* Label Color: Adjust label colors either uniformly or based on node colors.

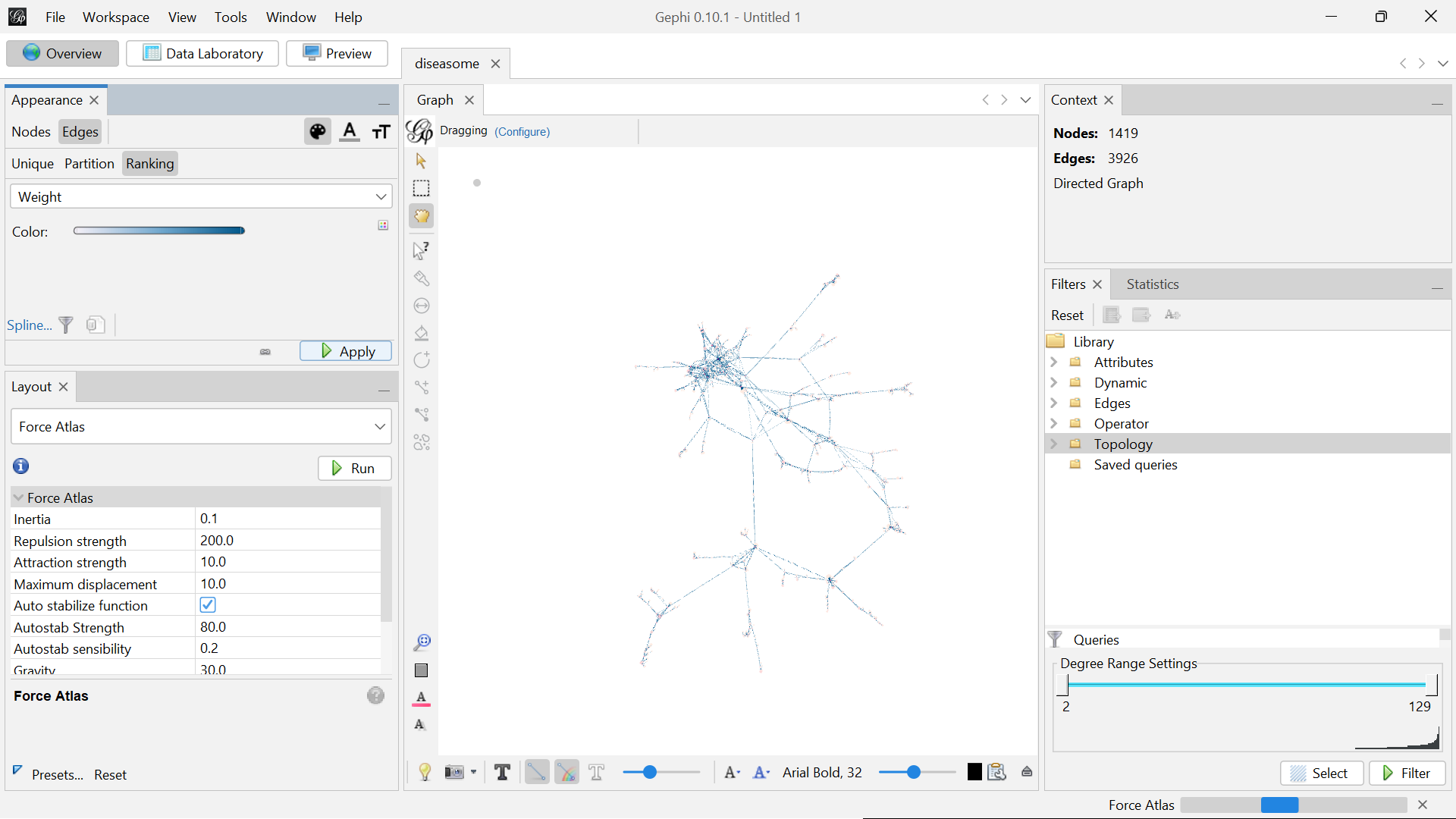
****

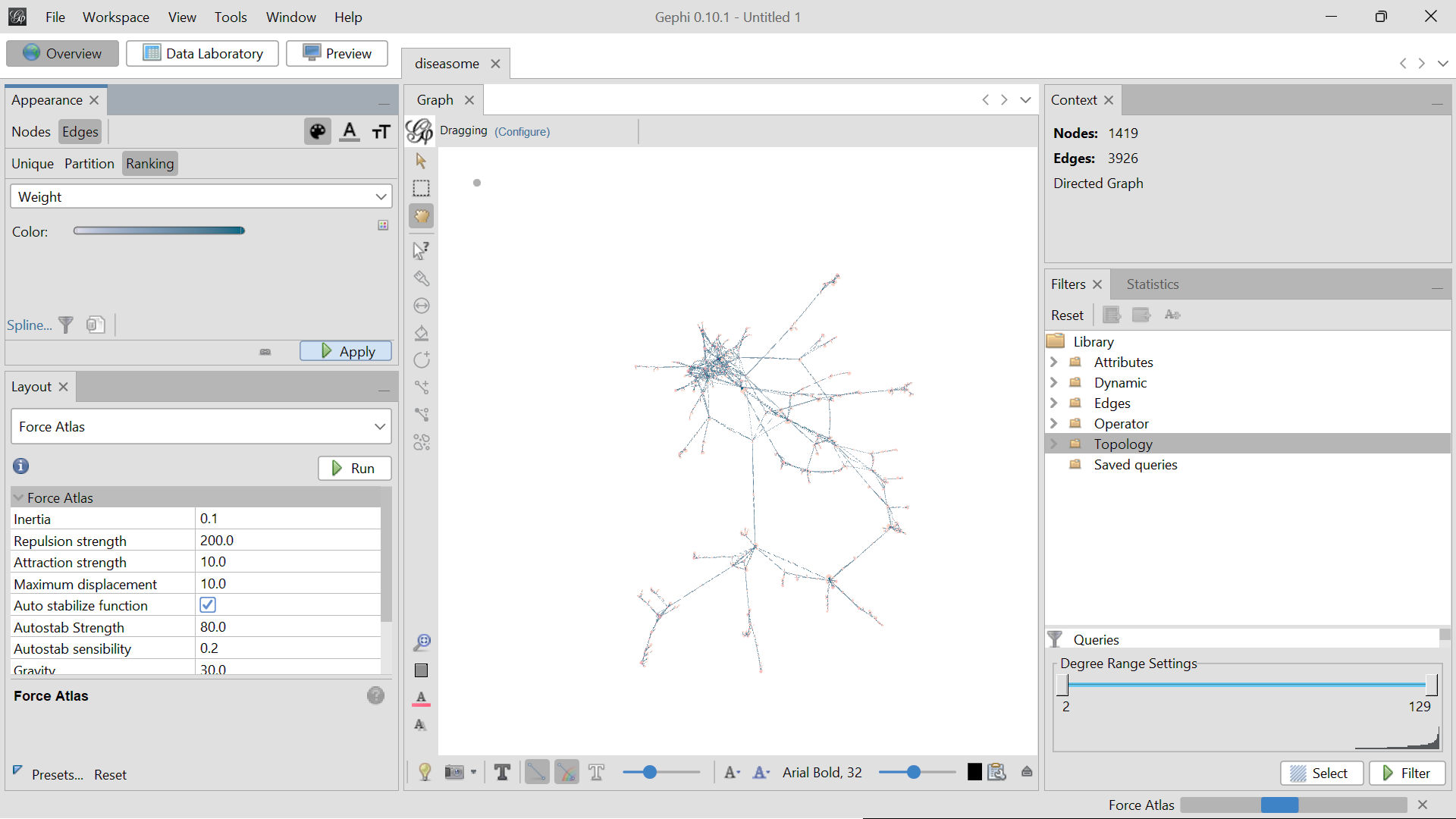
****

****

****

**Based on Edges:**

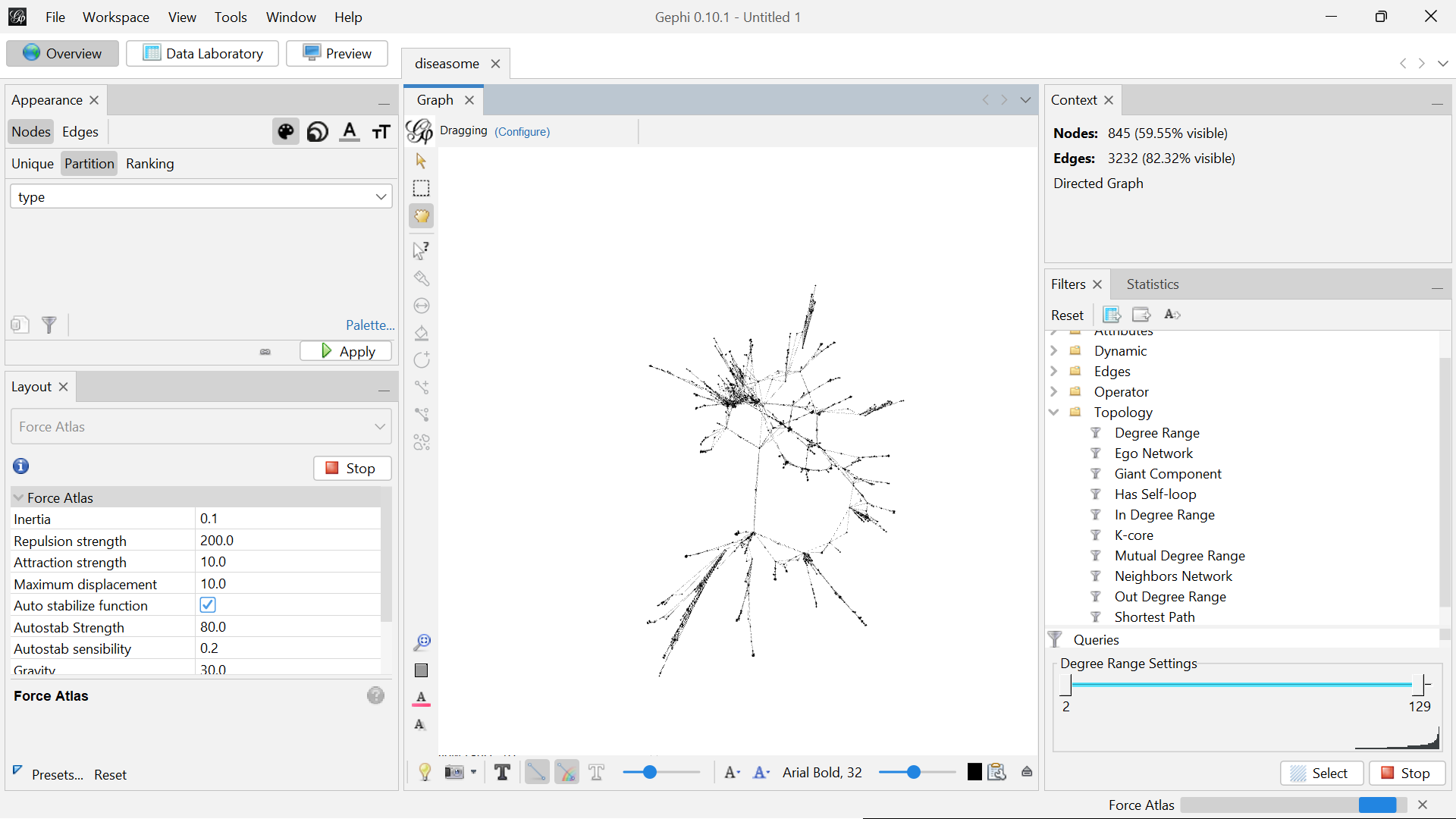
****

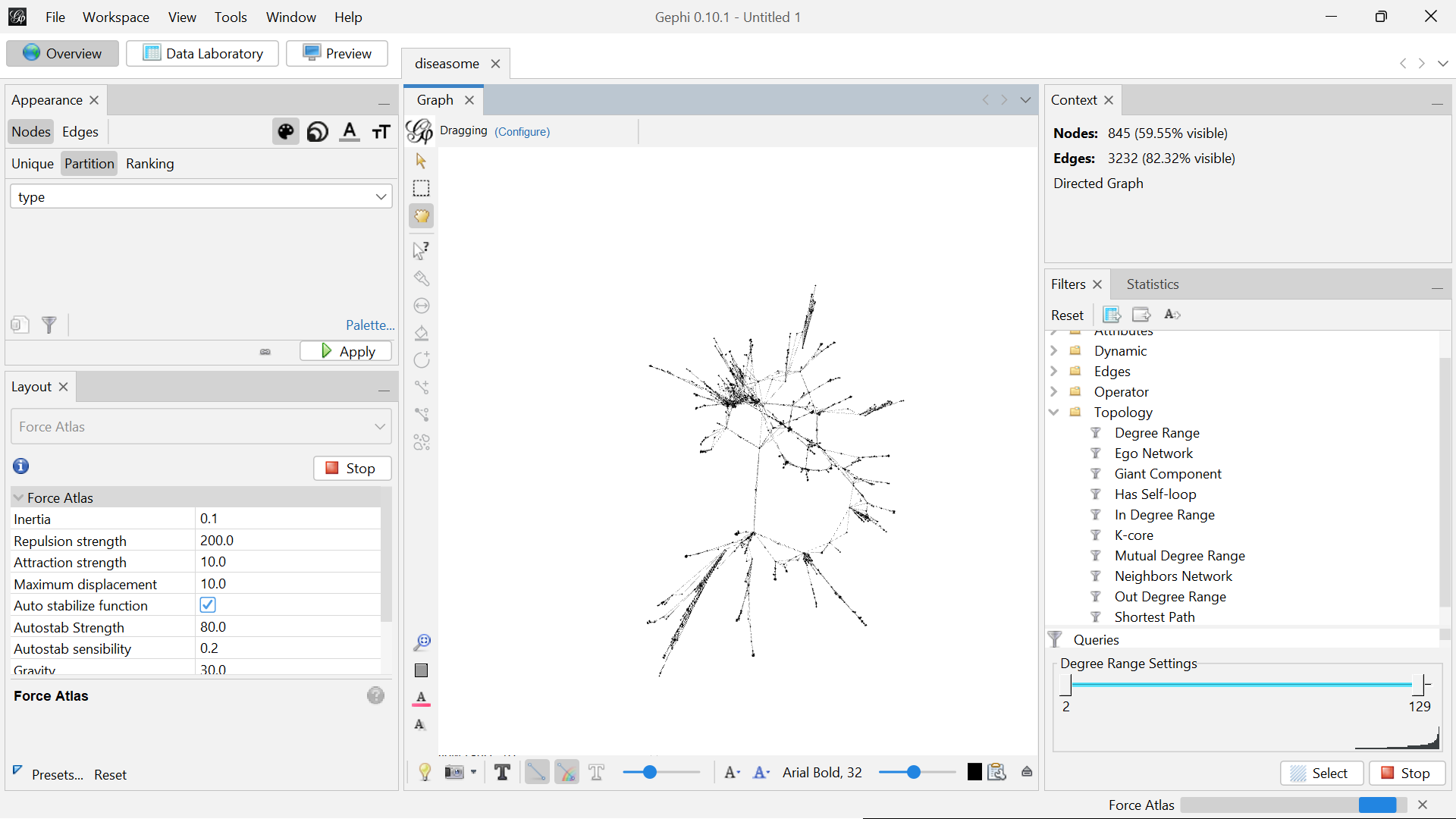
****

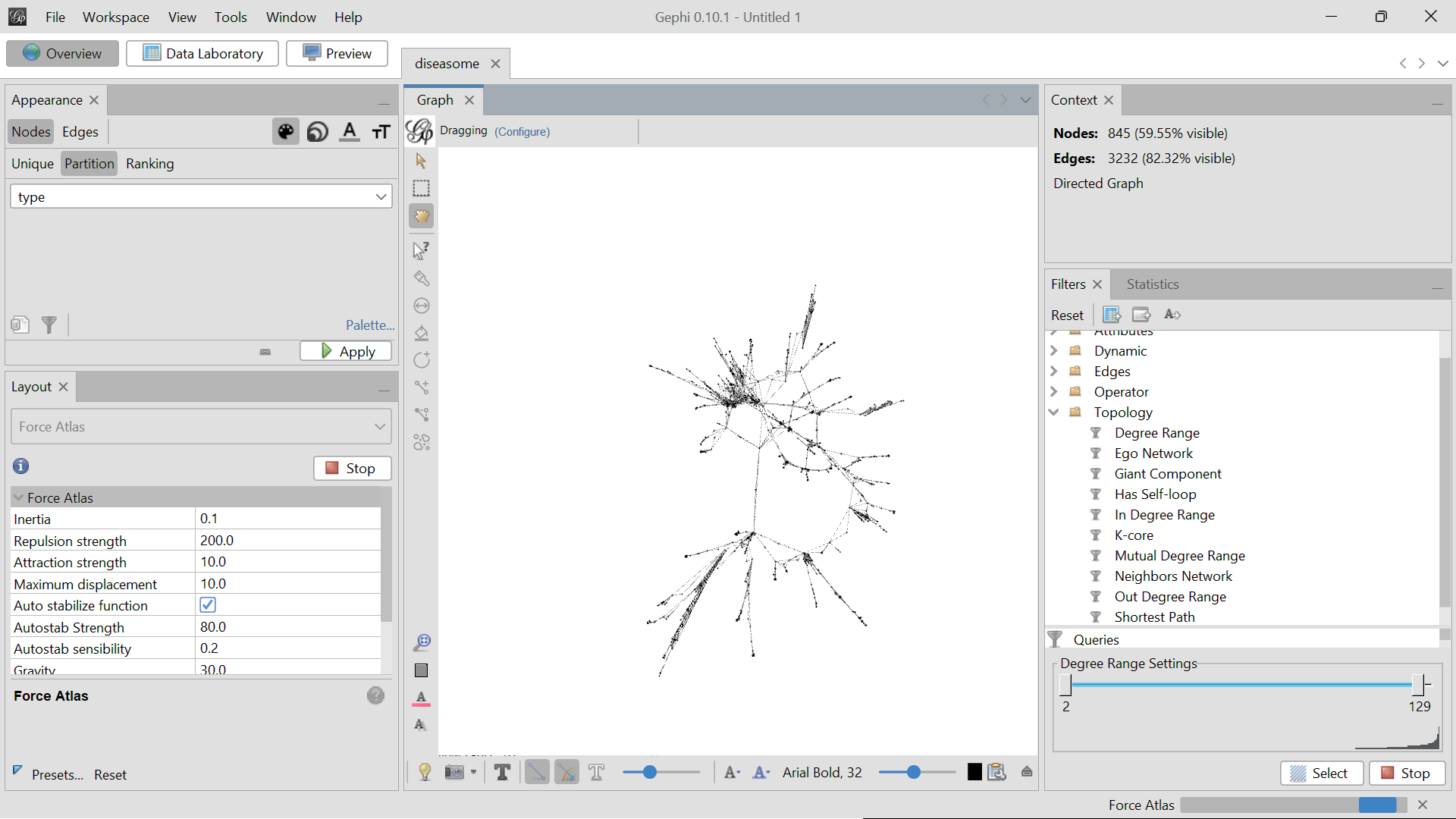
**Use cases of Topologies:**

**Degree Range**

The Degree Range option under the Topology tab in Gephi is a tool used to filter and visualize nodes based on their degree (the number of connections each node has). It allows you to select and view nodes within a specified range of degrees, helping you focus on nodes with particular levels of connectivity.

****

****

****