**Privacy and Security in Online Social Networks**
Department of Computer Science and Engineering
Indian Institute of Technology, Madras

**Lecture – 28**
Gephi Network Visualization

Gephi is an open source network analysis and visualization software. Gephi is widely used in a number of research projects in academia, journalism, digital humanities etcetera. Gephi can also input data of social networks like Facebook and Twitter and generate graphs and clusters.

(Refer Slide Time: 00:57)
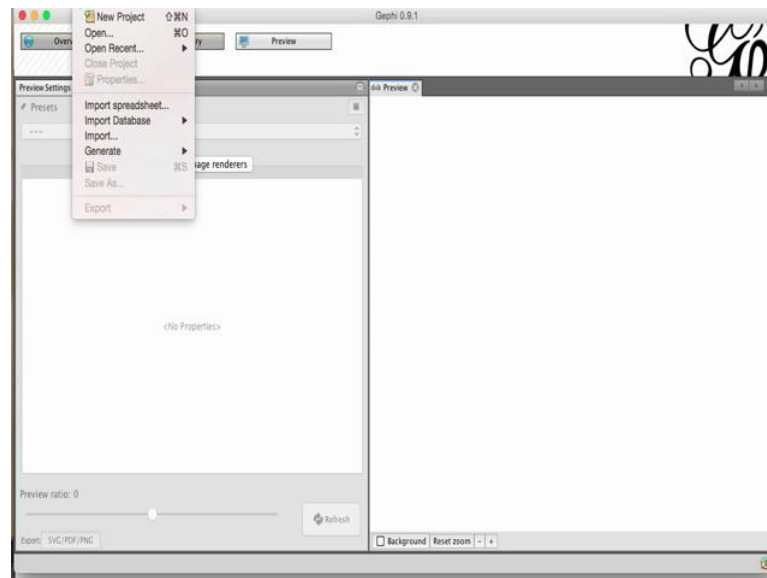
In this tutorial, we will learn in detail how to use gephi. Let us begin by understanding how to load data in gephi. Gephi can read many file formats including gml, graphml, pajek net file and uci net, dl files. You can open any of those by simply going to file and then clicking on open.
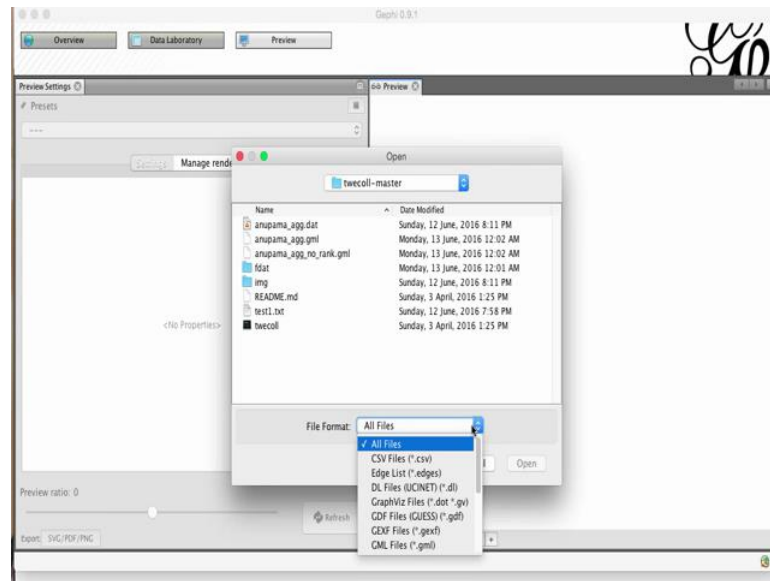
You can also look at the complete list of supported file formats by clicking on the drop down menu. Gephi can also read excel and csv files.
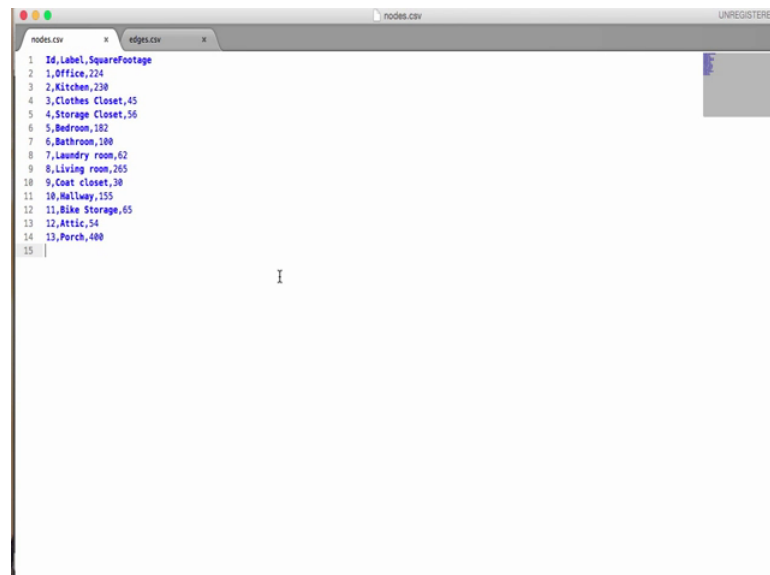
 To import data from csv, you will usually need to prepare two files, one containing nodes and their attributes and the other containing an edge list and edge attributes. The csv file containing nodes needs to include a column named id containing unique node ids. The edge list csv should have columns titled source and target containing node ids of the start and end node for each edge. In addition, you can also include a column called type indicating the type of each edge that is directed or undirected.
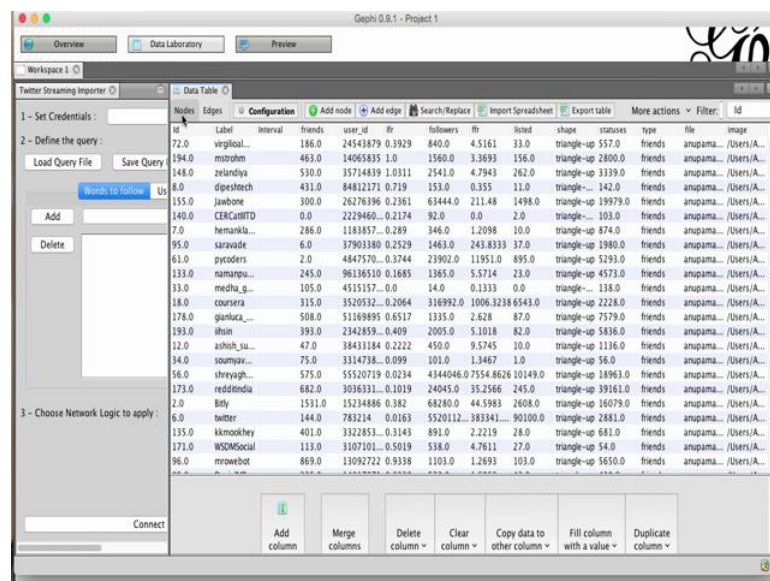
(Refer Slide Time: 02:22)
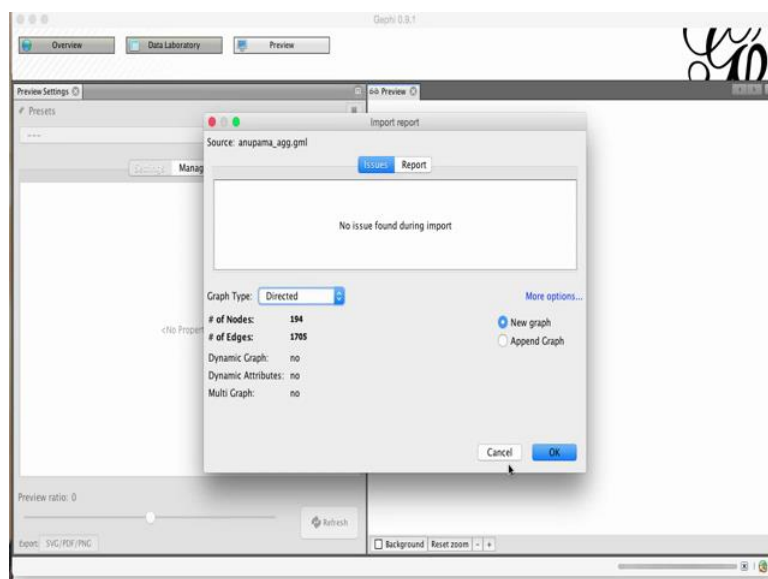
We already learnt about graphml format in the previous tutorial on social network analysis. Let us open an already existing gml file for this demo. If the data is in correct format, gephi will give you a summary of the data, which is the number of nodes, the number of edges and other information about the edge. Click on ok to successfully import the file.
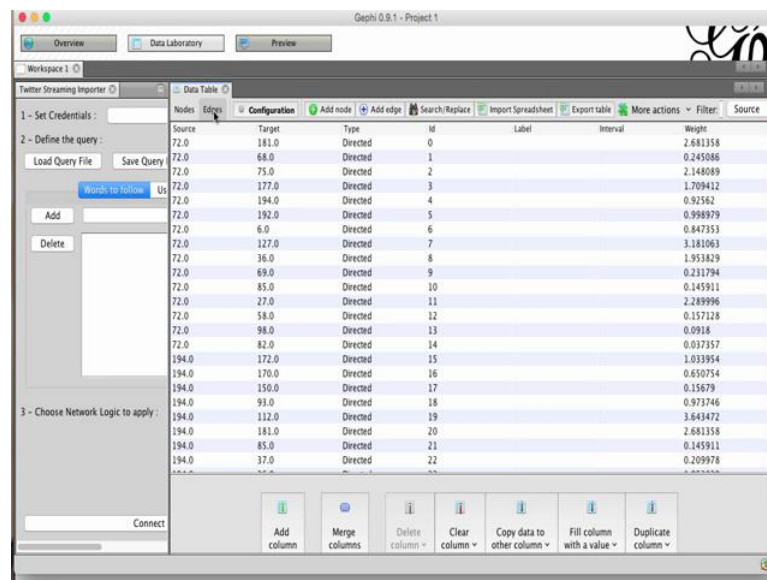
(Refer Slide Time: 02:36)
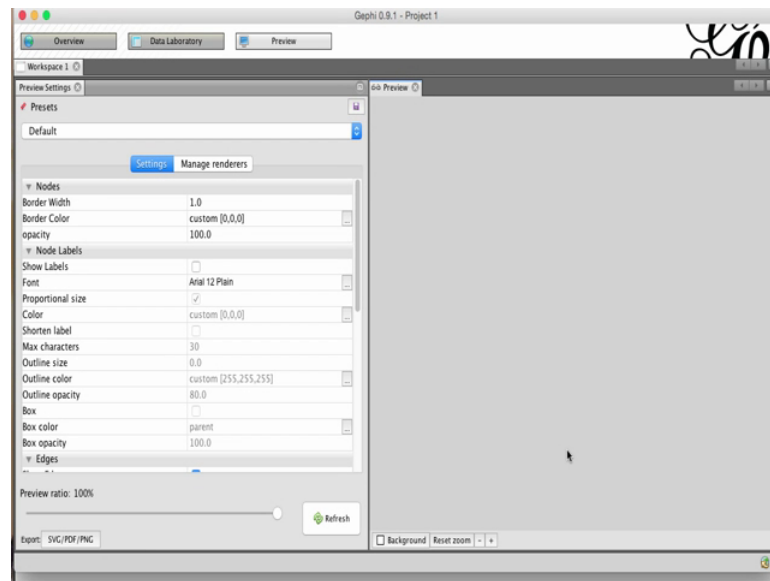
(Refer Slide Time: 02:44)

Now, we can view the data by going to the data laboratory tab in gephi. On right, you can see that there is a data table, which has two tabs - nodes and edges. The nodes tab contains the data table for nodes and node attributes.
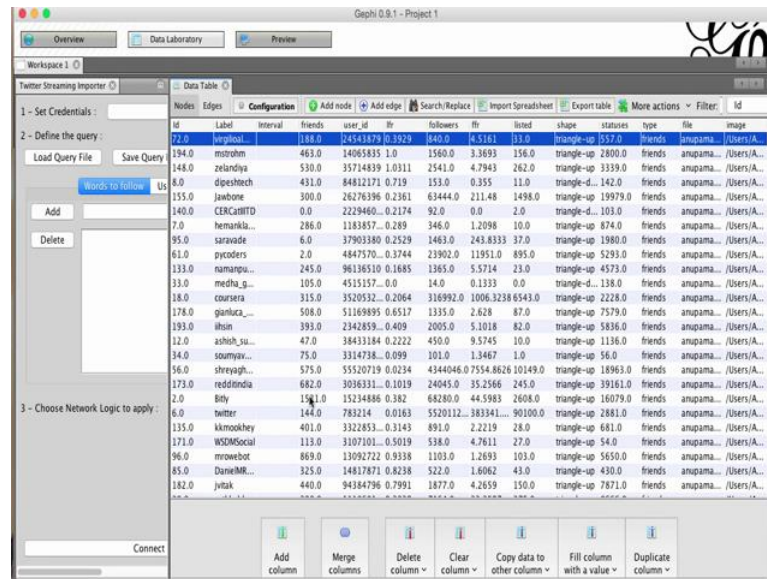
(Refer Slide Time: 02:59)

And the edges tab contains the data table for edges and edge attributes. You will also notice an import spreadsheet button which is used to get nodes and edges data directly from excel or csv files. The panel at the bottom can help you manipulate data columns and modify the data. You can also change individual values directly by clicking on them and typing their new values.

(Refer Slide Time: 03:28)

For instance, let us go to the nodes table, and change the number of friends of the first row from 186 to 188.

(Refer Slide Time: 03:45)
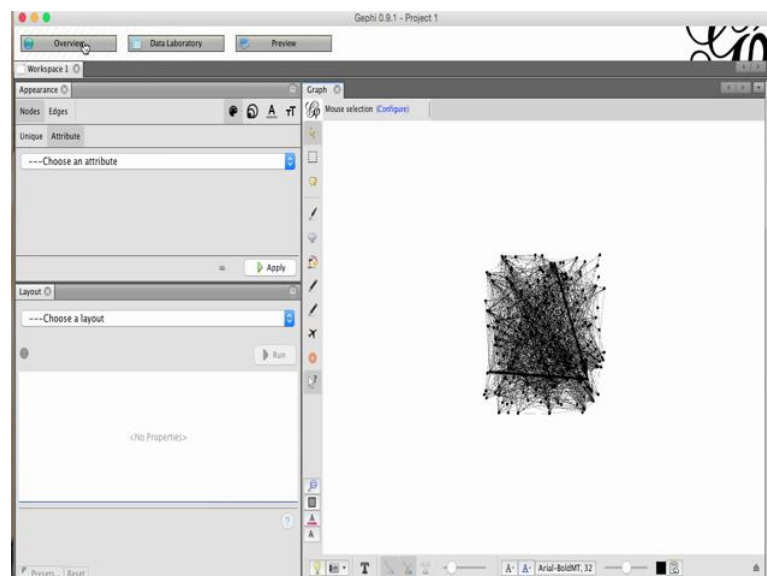
Now, let us look at the over view tab to calculate the network measures and set up network visualization. Once you click the over view tab, you will notice the visualization

of your data in the blank space in the middle. Over view tab has many options like appearance, layout and a several panels to change the color, size and attributes of the visualization. Let us look at each one of them one by one. The appearance tab on the top left lets you change the node edge color and size based on categorical and continuous attributes. You can select various visual properties by going through the options.

The layout tab can help you select and customize one of the available layout algorithms. Various options on the left of the visualization help you with interactive selection of nodes and edges. And also lets you change the size and color of the nodes and edges manually. The edge pencil and the node pencil can also help you add nodes and edges manually. The options towards the bottom let you resize the color and size of the characteristics of nodes and edges.

(Refer Slide Time: 05:04)
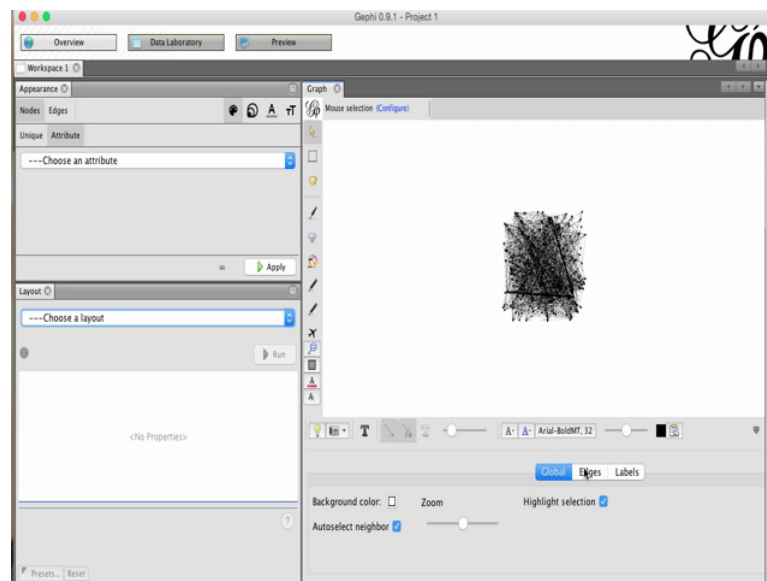
The panel at the bottom can be expanded by clicking on the small up arrow on bottom right. This panel can help you change the color, size and other characteristics applying to all nodes edges and labels.

(Refer Slide Time: 05:33)
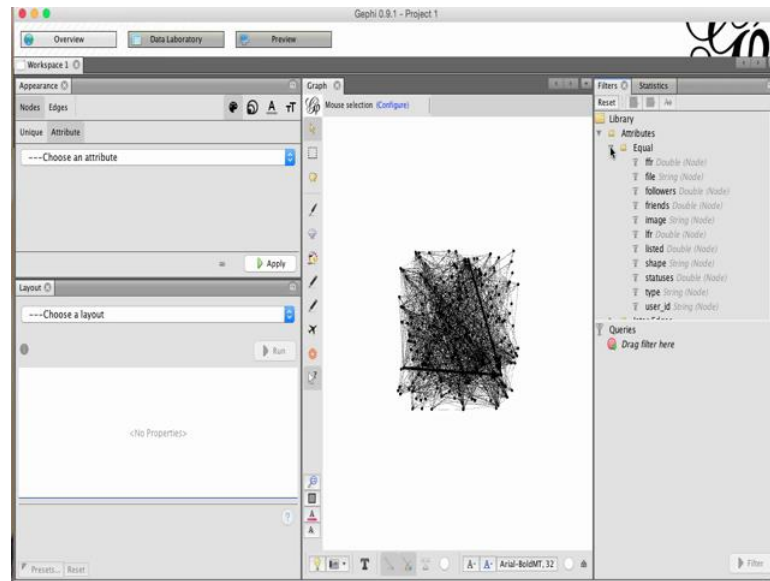
Now, let us look at the filters and statistics tab on the right. You can apply filters to select specific nodes and edges from your network. The filters are applied to dragging and dropping the specific filter in the query space. There are several filters based on attributes like, equal can help you select elements with particular attribute values.

(Refer Slide Time: 06:02)
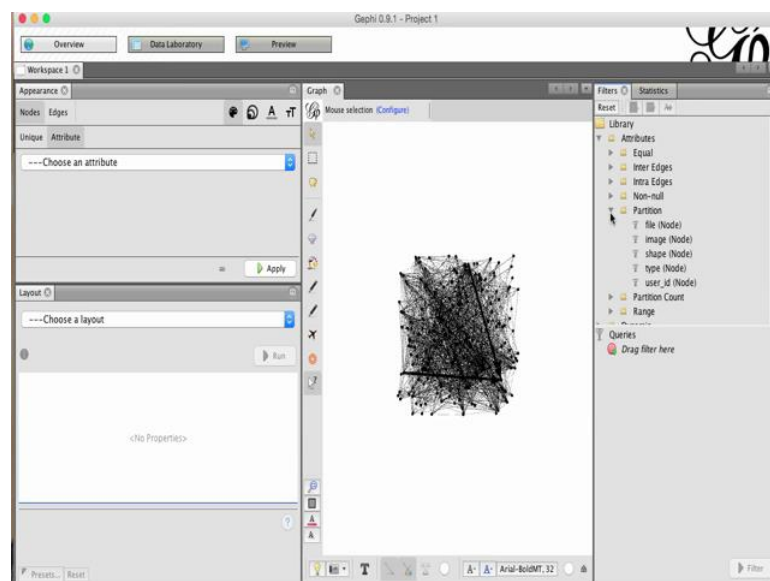
The partition filter can help you select the different levels of categorical attributes.

(Refer Slide Time: 06:14)
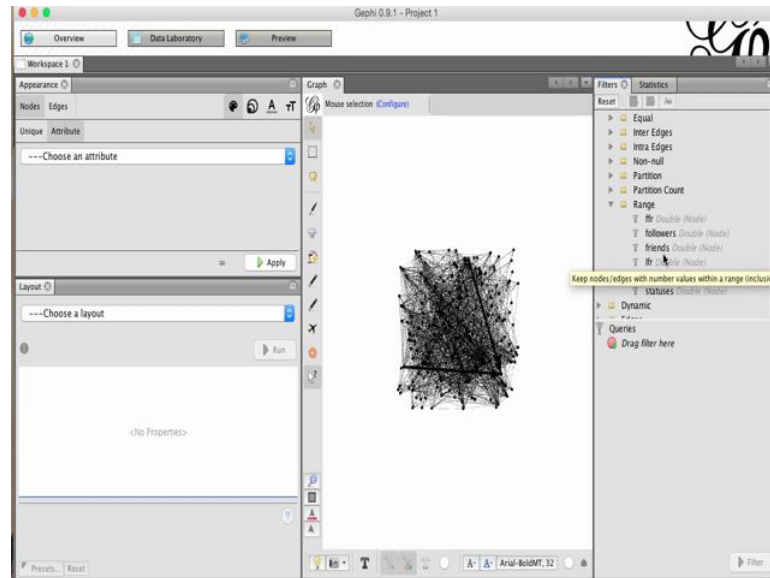
With the range filter, you can select nodes and edges with attribute values in a particular range. For instance, followers between 100 to 1000. Filters based on edges allow you to select edges with different properties like edge weight, edge type and self loops.

(Refer Slide Time: 06:32)
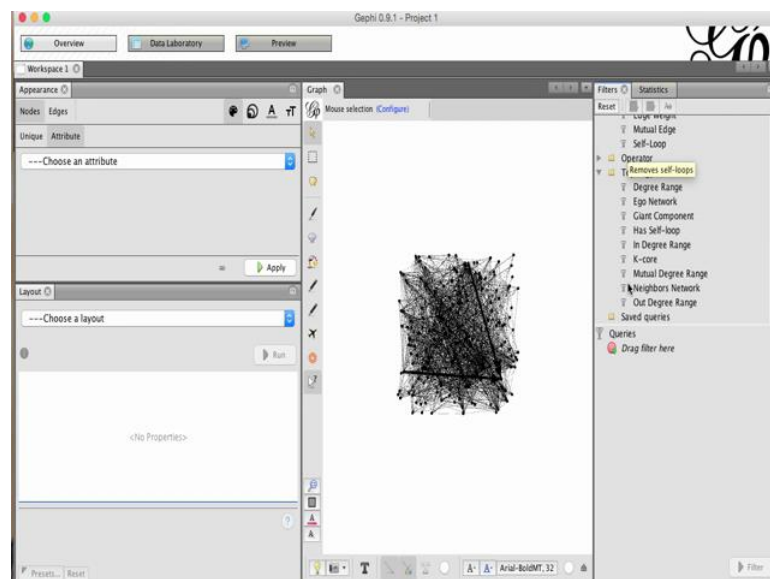
The topology filter allows for selection based on the network structure like components, k cores, degree, ranges etcetera.
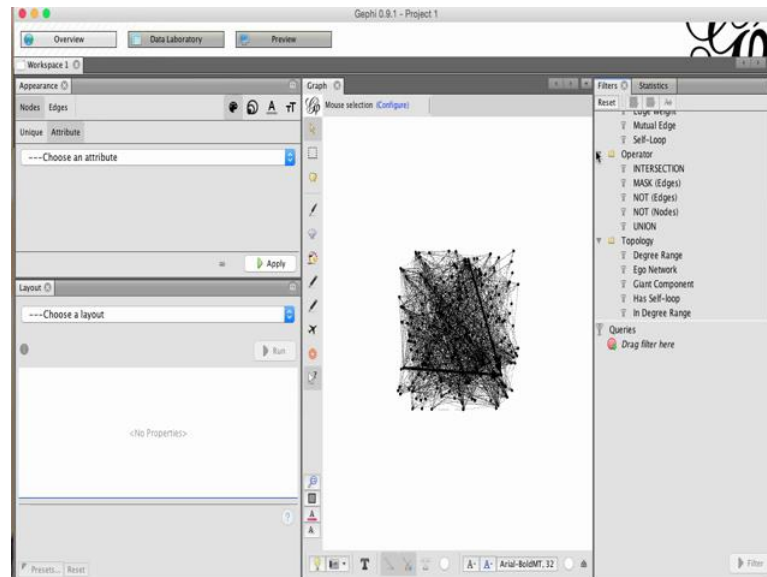
The operator allows you to combine other filters in different ways. For instance, find your intersection, union, compliment etcetera.
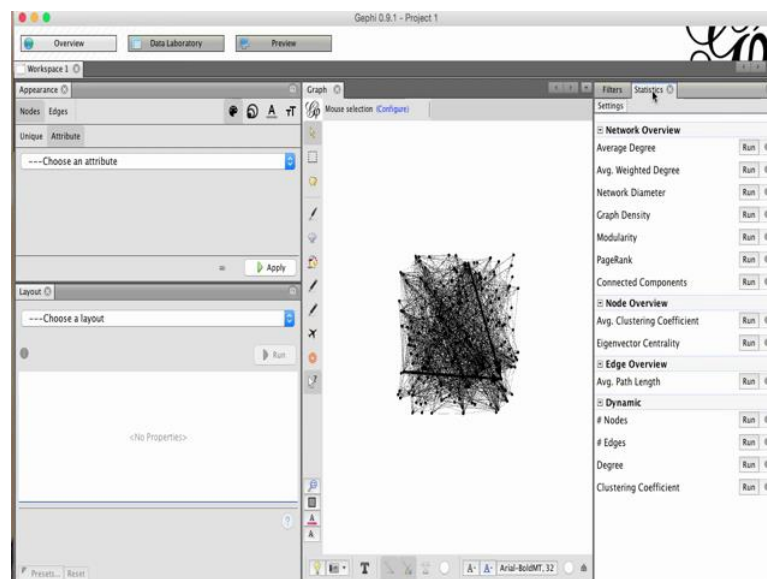
Now let us look at the statistics tab, which is used to calculate the network node edge statistics. Once you click on this statistics tab, you will be able to see the different measures whose value can be calculated by clicking on the individual run button. For our data, let us try calculating the value of these measures.

(Refer Slide Time: 07:24)
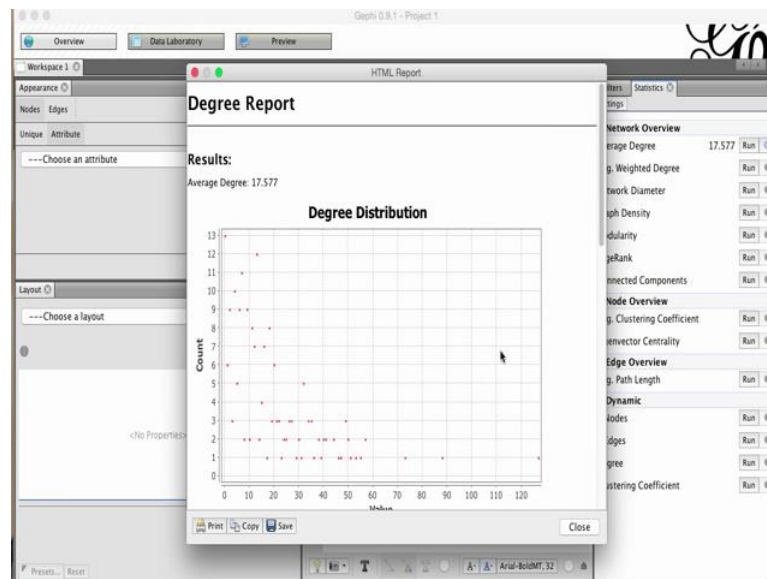
Let us begin by computing the average degree. Once you click on run, there will be a degree report generated, which will tell you the average degree. In our case, it is 17.577 and also give you a degree distribution graph where the x-axis is the value of the degree distribution, and the y value is the number of nodes with that particular degree distribution. It will also give specific reports of the in-degree and out-degree values and corresponding distributions. Next, let us run the average weighted degree. In our data, the weight of the edges is one, therefore, the graph is a straight line.
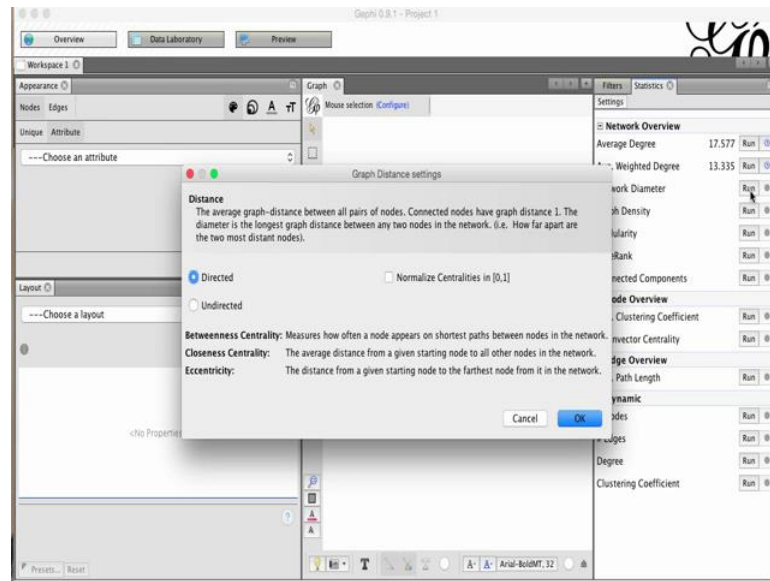
Next, let us look at the network diameter. Network diameter is the average distance between all pair of nodes. We will select the directed option since we have a directed network, and click on OK.
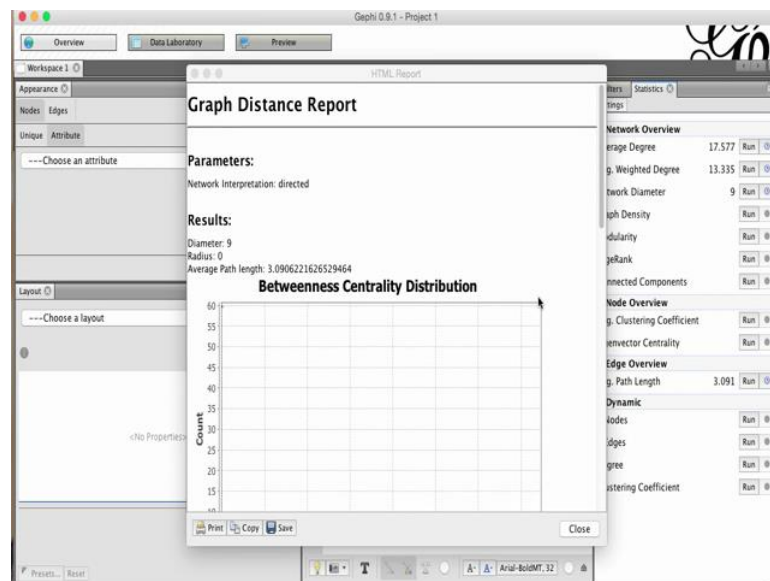
This <mark>will</mark> generate graph distance report <mark>which</mark> says that the average network diameter is 9.

(Refer Slide Time: 08:41)
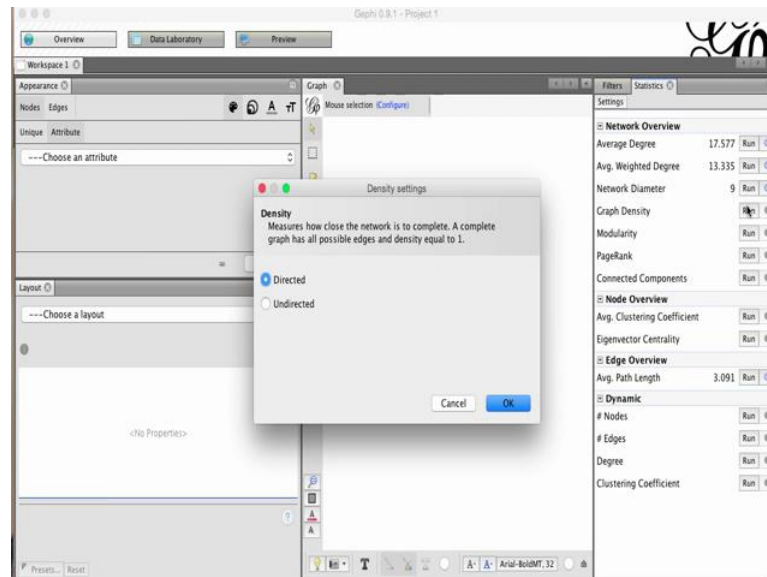
Now let us <mark>compute</mark> the graph density. The graph density measures how close the network is to complete. A complete network means that all the nodes in the graph are connected to each other. Click on OK.

(Refer Slide Time: 09:02)
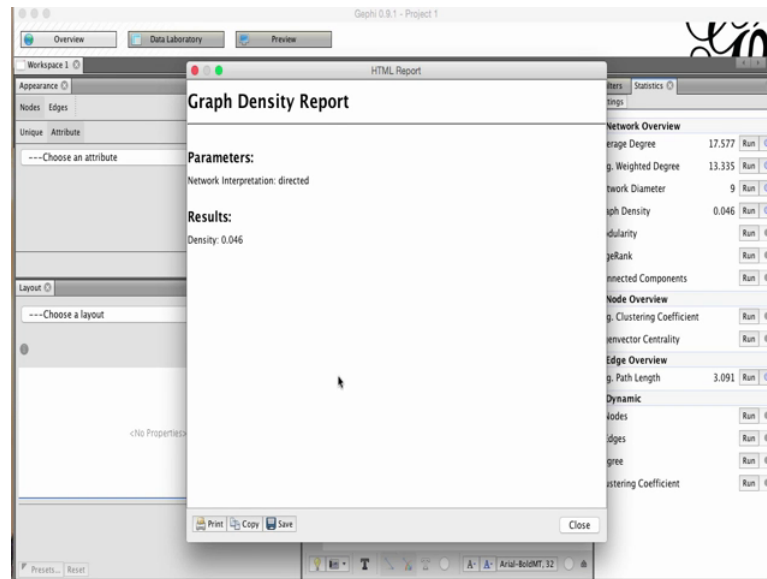
And the graph density in our case is 0.046.

(Refer Slide Time: 09:07)
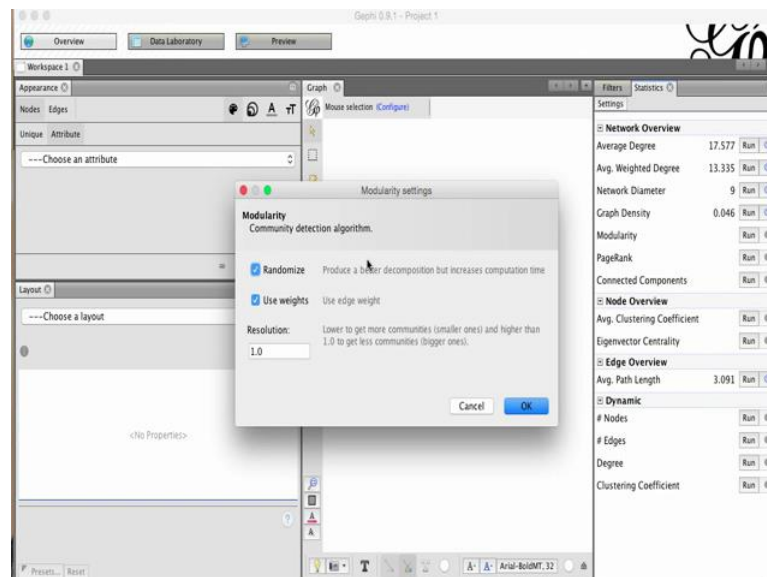
Now, let us compute modularity. Modularity is the measure for communities in a network which we have already covered in the previous tutorial. For our data, we will

unselect the use weight option, because we do not have a weighted graph and click on ok.

(Refer Slide Time: 09:24)
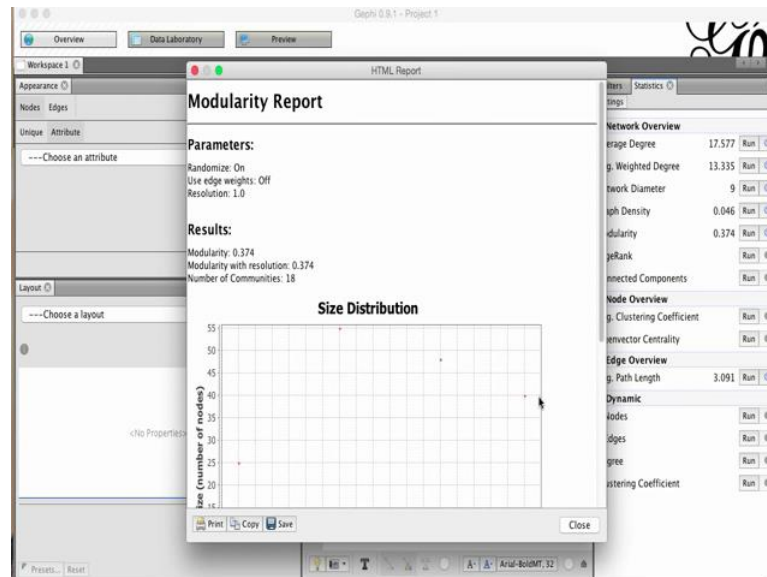
In our case, we can see that the numbers of communities are 18 and the modularity score is 0.374.

(Refer Slide Time: 09:41)
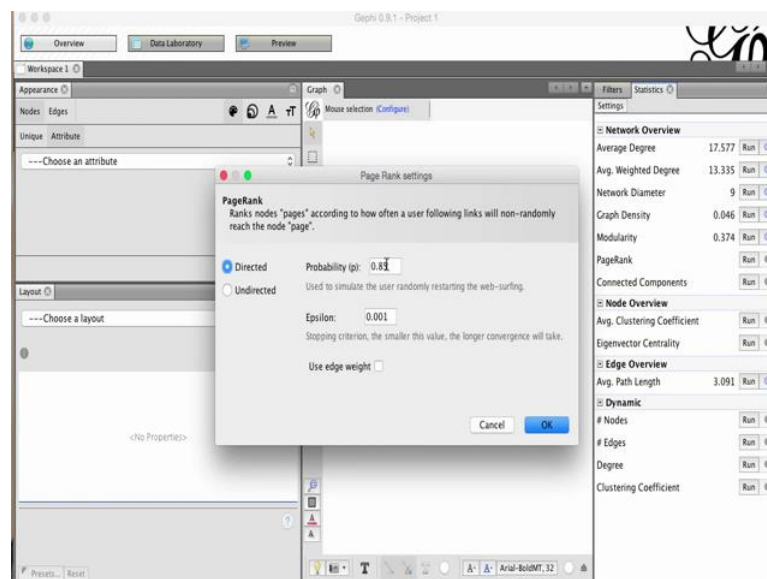
Now let us compute the page rank. You can change the p and epsilon values as per the page rank algorithm or keep it as default and click on OK.

(Refer Slide Time: 09:50)
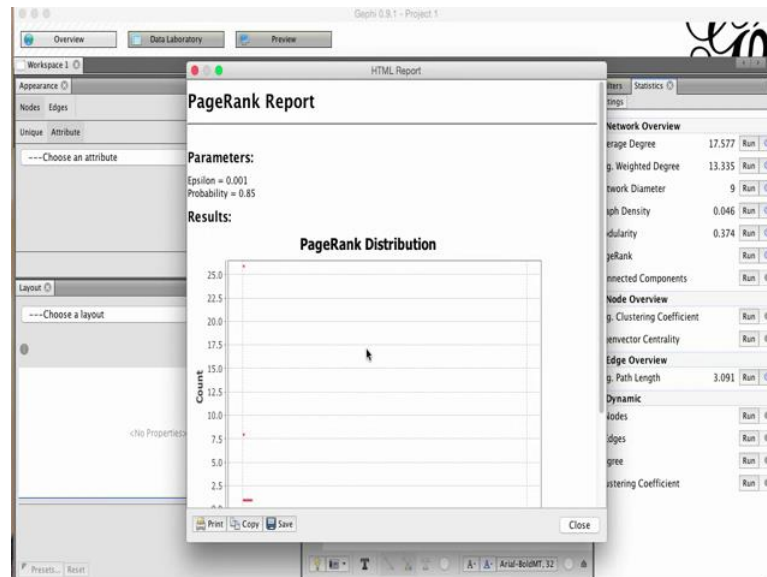
This will compute the values of the page rank algorithm for your graphs.

(Refer Slide Time: 10:04)
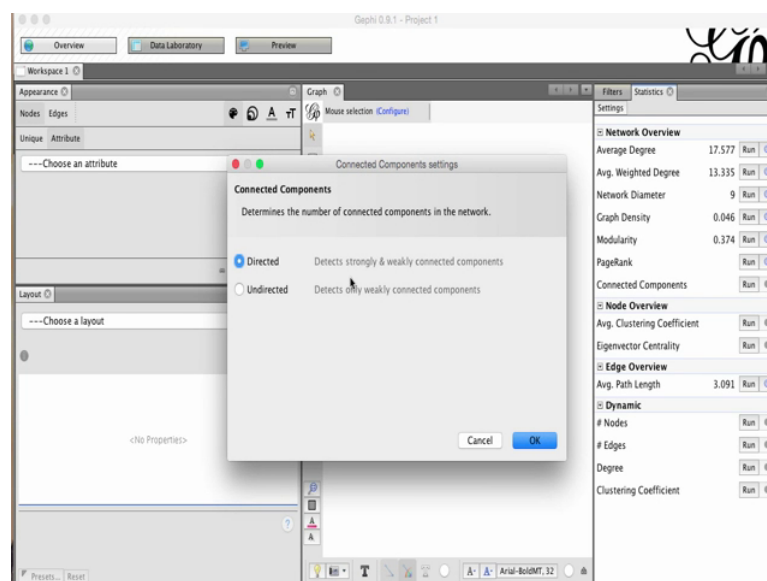
Let us see how many connected components are there in our data. Click on OK.

(Refer Slide Time: 10:08)
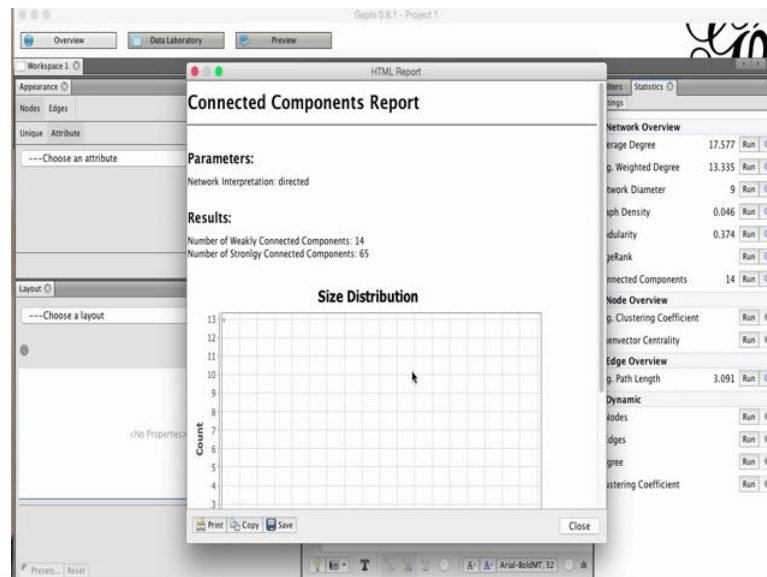
We can see that we have 65 strongly connected components; and 14 weakly connected components in a data.

(Refer Slide Time: 10:27)
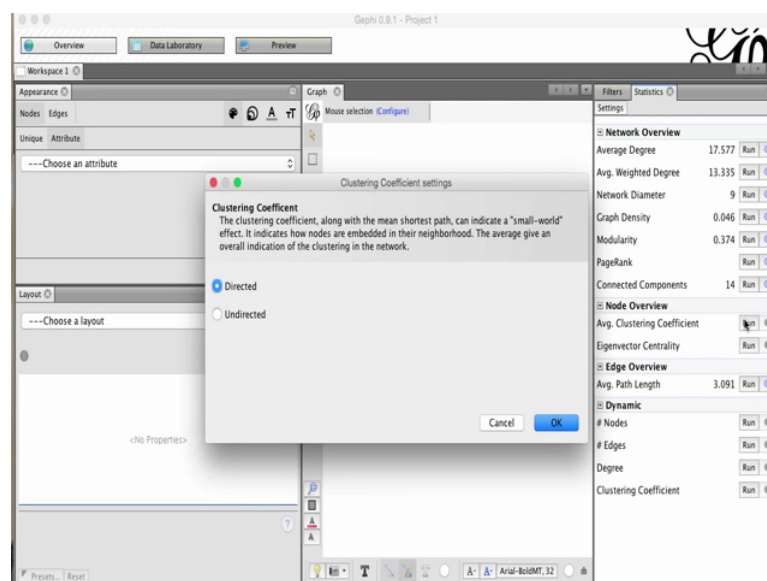
Now let us also look at the node specific attributes which is the average clustering coefficient. Let us click on run.

(Refer Slide Time: 10:33)
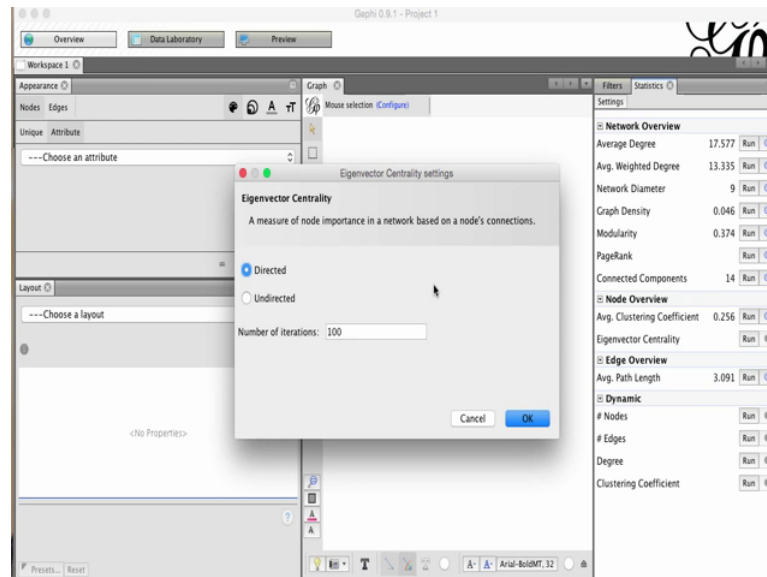
And let us also compute the eigen vector centrality of the nodes.

(Refer Slide Time: 10:38)
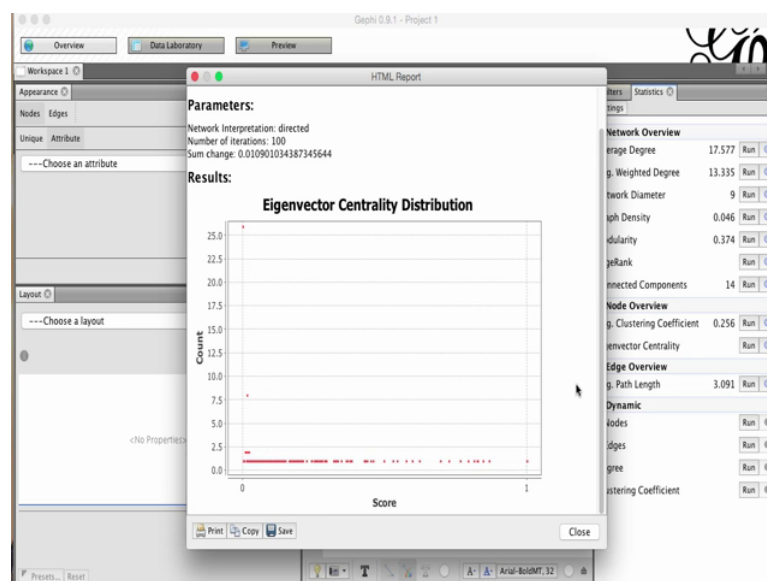
Once these measures are calculated, many of these will be available in the data laboratory and can also be used for the visualization. For example, after computing the eigen vector centrality, it will allow you to resize nodes based on that particular attribute.

(Refer Slide Time: 11:14)
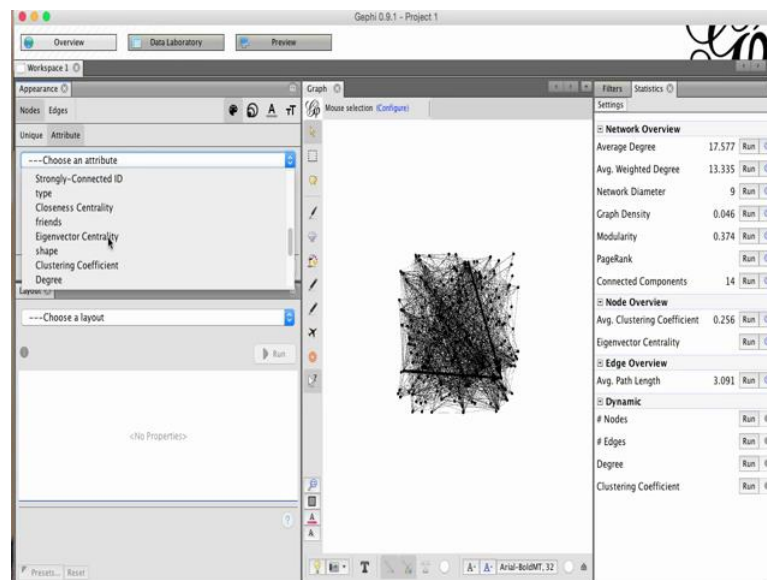
We can see it there in the left panel. Let us go to nodes, attributes, and you will be able to see all the measures which we have computed just now, like the strongly connected components, eigenvector centrality, clustering co-efficient etcetera. All of these options were not present earlier.

(Refer Slide Time: 11:50)
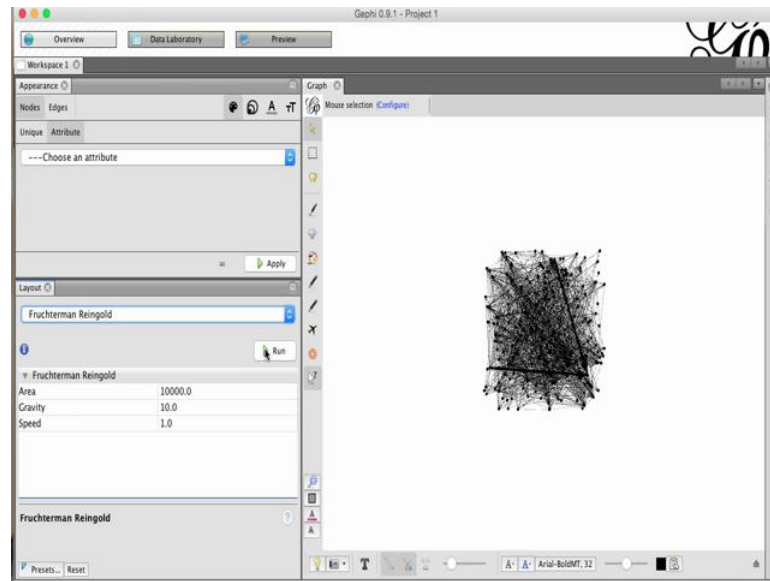
Now, let us collapse the filters and statistic tab and play around with the visualization of the data a bit. Currently, because the data is very dense, the nodes and edges are all mixed up. Therefore, let us choose a layout so that we can see the data more efficiently, click on choose a layout drop down menu and select Fruchterman Reingold algorithm. And then click on run.

(Refer Slide Time: 12:02)
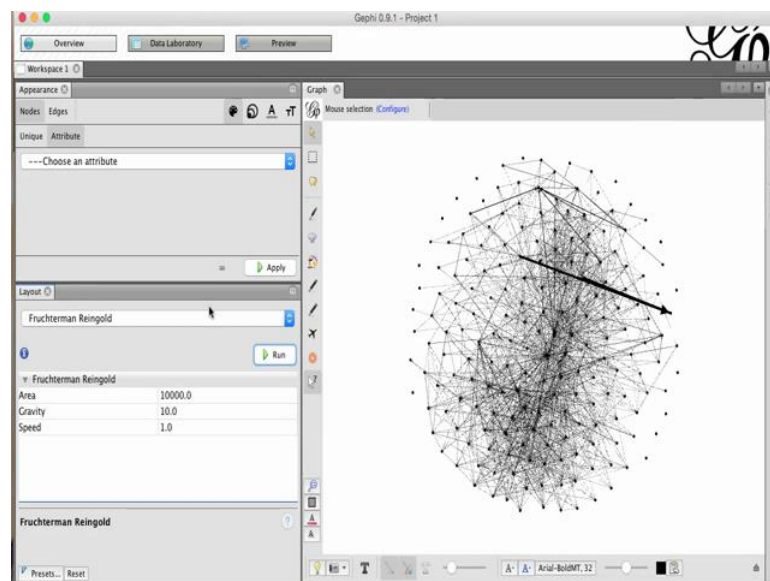
Let the algorithm run for a while. And after it starts to look stabilized, click on stop.

(Refer Slide Time: 12:31)
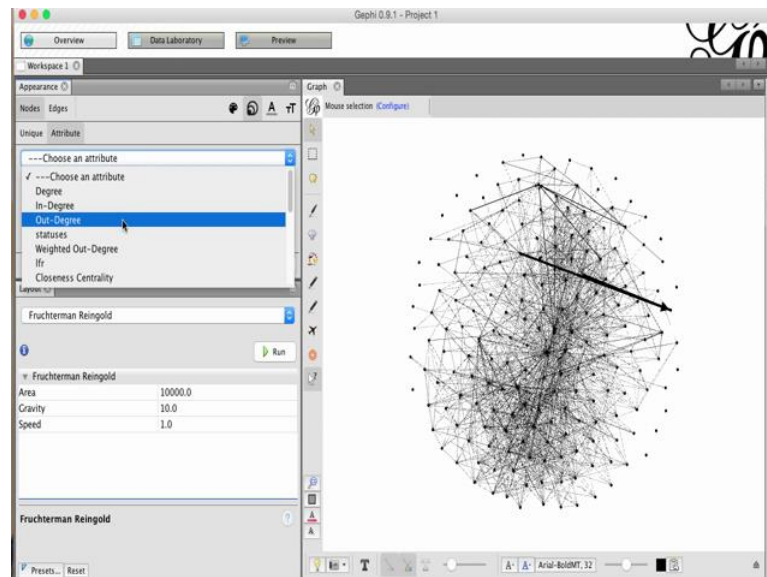
Now let us adjust the size of the nodes and edges based on the various attributes, which we just calculated. Let us first change the size of the nodes. In the appearance tab, on the top left, click on nodes, then select the size from the right side options, click on attribute and let us choose an attribute. Let us say we want to size the nodes according to how many followers that specific node has. Click on followers and then apply.

(Refer Slide Time: 12:50)
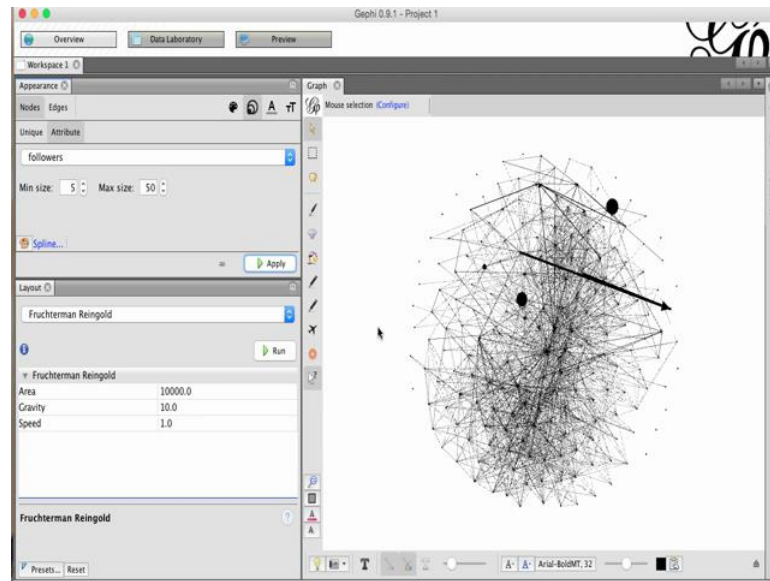
You will notice that some nodes now are bigger. These are probably the nodes which have large number of followers and which has celebrity accounts in the twitter network.

(Refer Slide Time: 13:19)
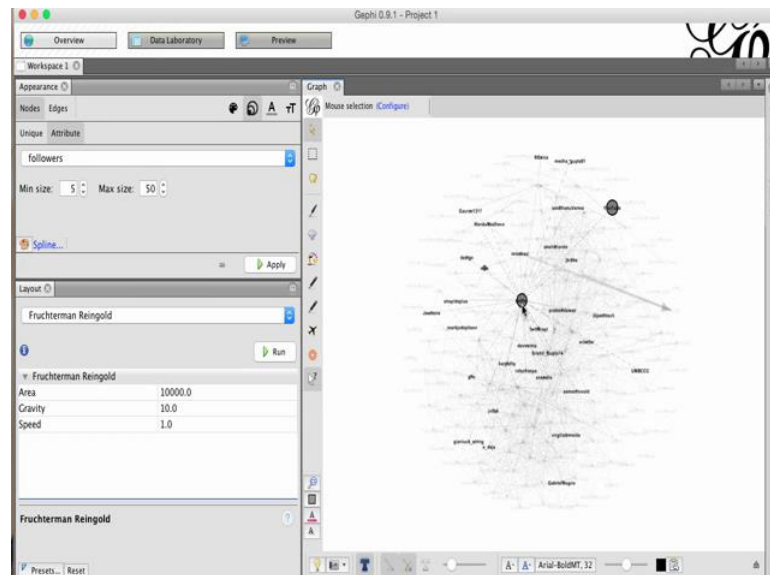
We can see which users these are by enabling the node labels. To enable the node labels click on T in the bottom panel which is show node labels. Now, you will be able to see

which are these users which have a large number of followers. This particular node is the official account of twitter; and this one is YouTube.

(Refer Slide Time: 13:56)
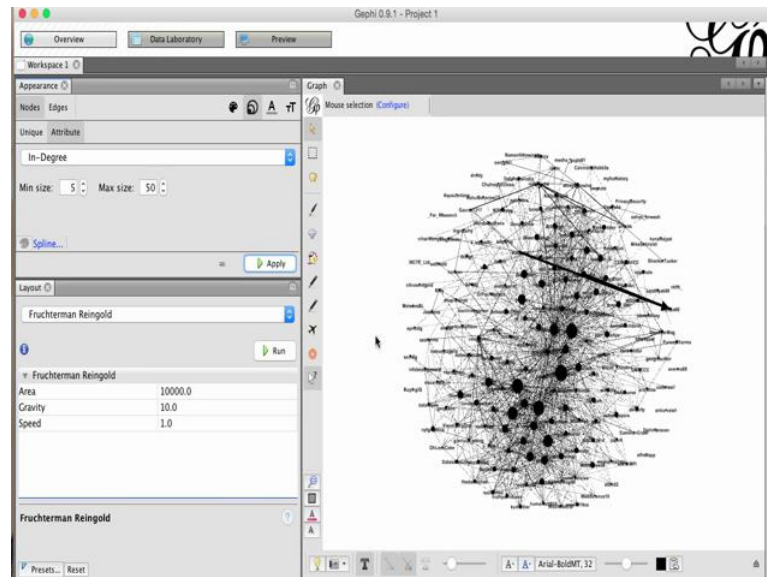
Let us try changing the parameter based on which we want to size the nodes. Click on the drop down menu again, and let us select in-degree this time, click on apply. Since, it is a two hop network of my own twitter data, therefore the in-degree would signify that these are the accounts which are most followed in my two hop twitter network. We can adjust the minimum and the maximum size of the nodes by changing these values.

(Refer Slide Time: 14:28)
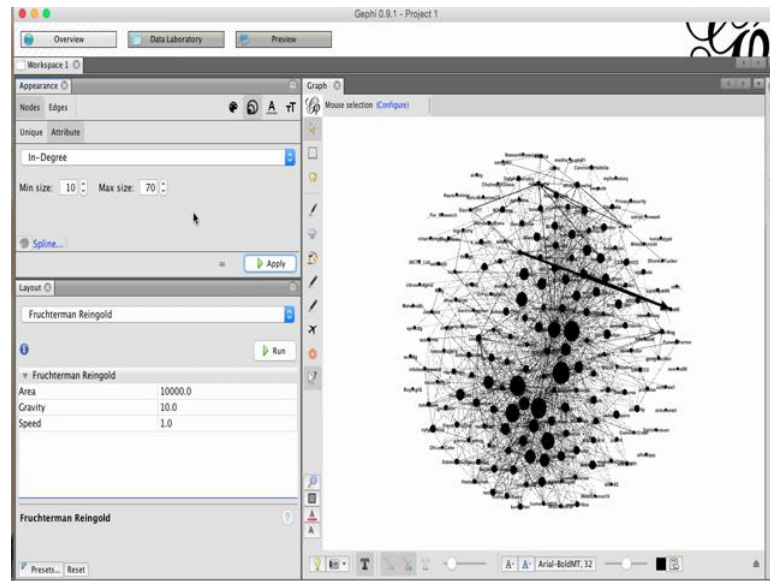
For instance, let us change the minimum size from 5 to 10 to make the even smallest node bigger and change the maximum size to let us say 70. Click on apply. And you will notice that the size of all the nodes is now bigger.

(Refer Slide Time: 15:06)
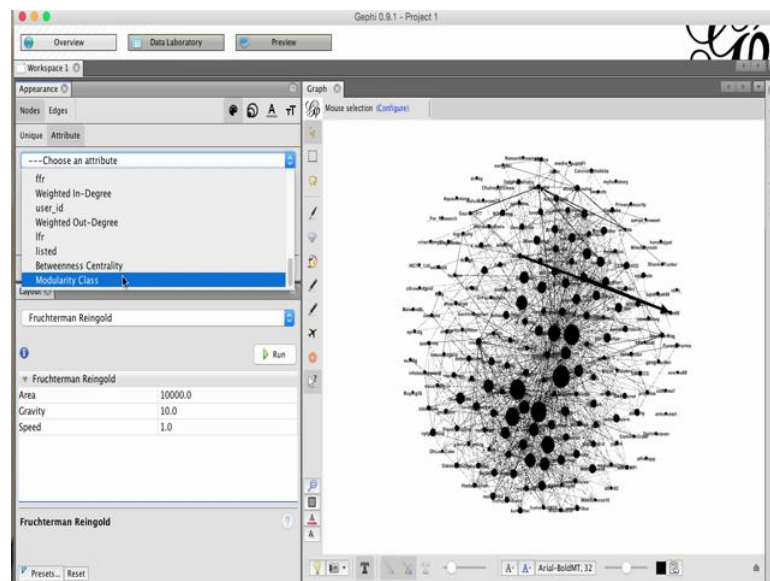
Now to make the graph more visible, let us try changing the color of the nodes. Now again, select nodes, and after that from the right, instead of size, select the color option, then go to attributes., and click on the drop down menu. Let us color the nodes based on the community, which they belong to. To do that, we will need to select the option modularity class.

(Refer Slide Time: 15:12)
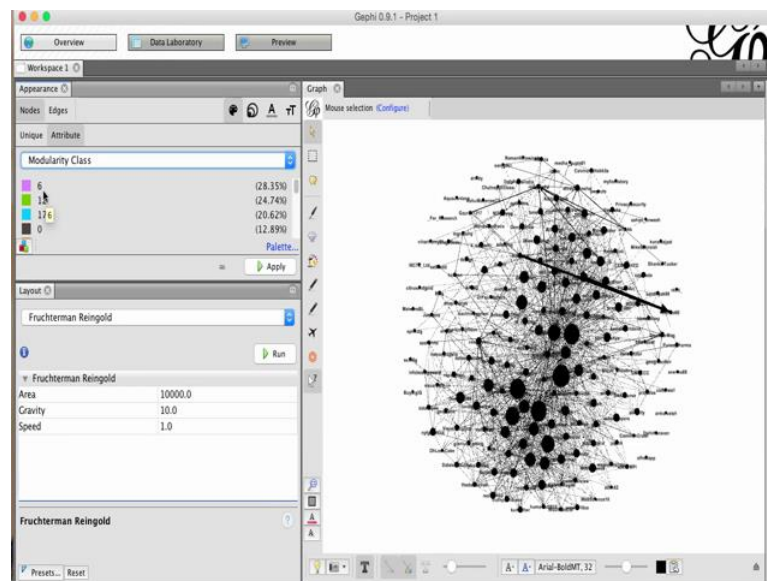
As soon as you click the modularity class option, it will tell you how many nodes exist in each community, click on apply.

(Refer Slide Time: 15:26)
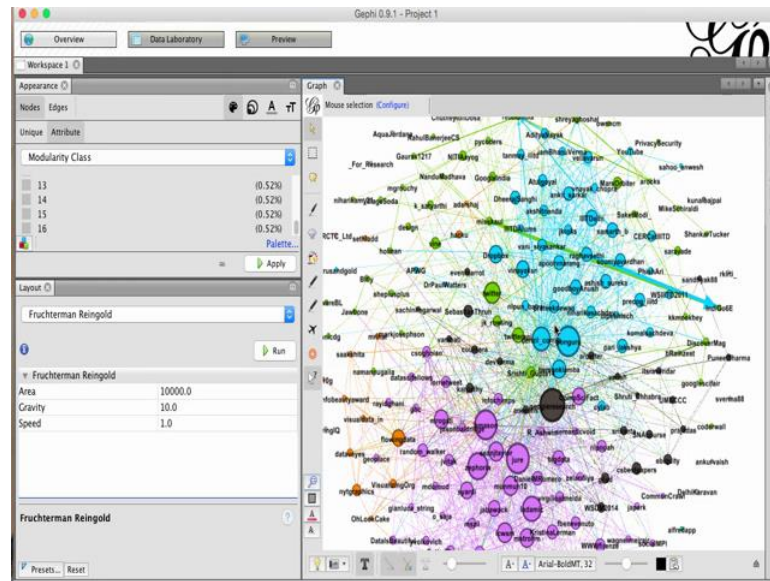
Now, the nodes are better visible you can zoom in by centering the graph and then scrolling down. Let us look at few more options. In the nodes, we can also change the label color by click on the A option. Let us color the labels based on the attributes available in the drop down menu. Let us choose out-degree and click on apply.

(Refer Slide Time: 15:52)
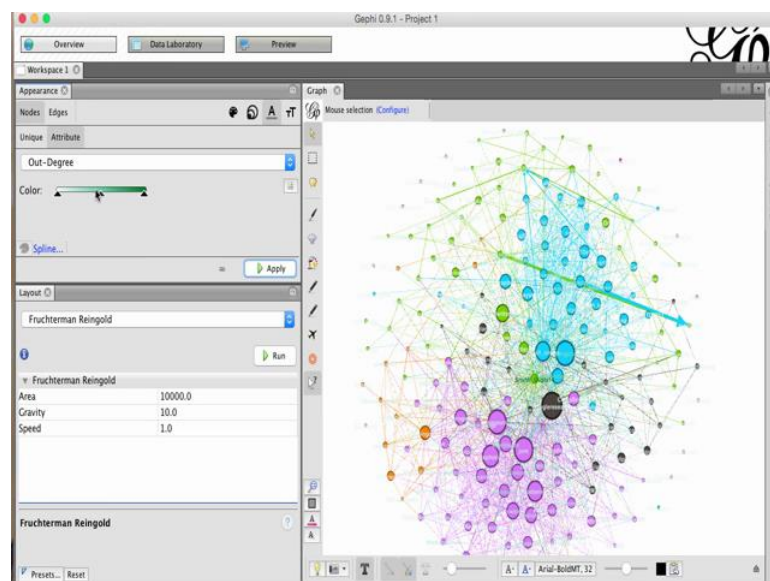
If you notice that the labels are not visible, you can change the color pallet. We can also change the size of the labels by clicking on the T option. Click on T, and either we can change the size of all the node labels.
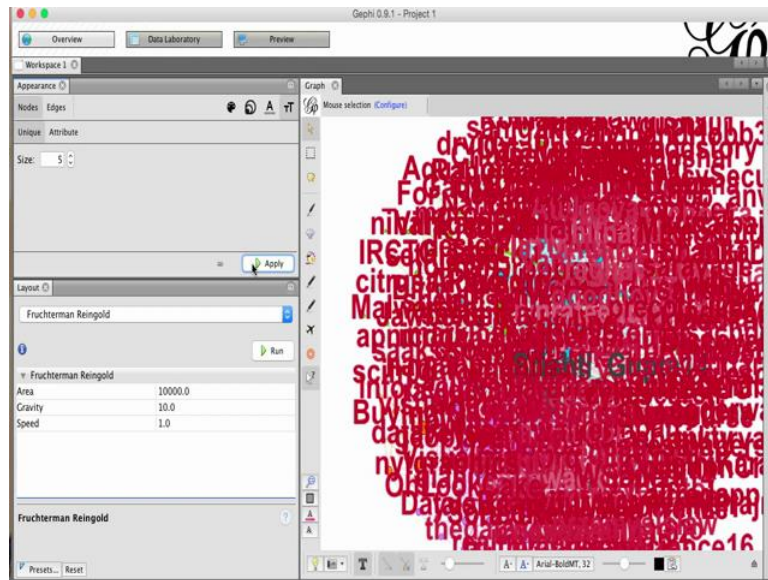
Let us change from 1 to 5 and see what happens, the text becomes too big.

Let us change it back to 1, we can also resize the node labels by clicking on the attribute tab, and have a continuous function. Let us choose out-degree and click - apply. Now the nodes which have higher out-degree have a larger label which means that users who are following other users a lot will have a larger text. In this case, this particular account srishti underscore gupta 14 is the account which follows maximum number of users in my twitter data.

(Refer Slide Time: 17:08)
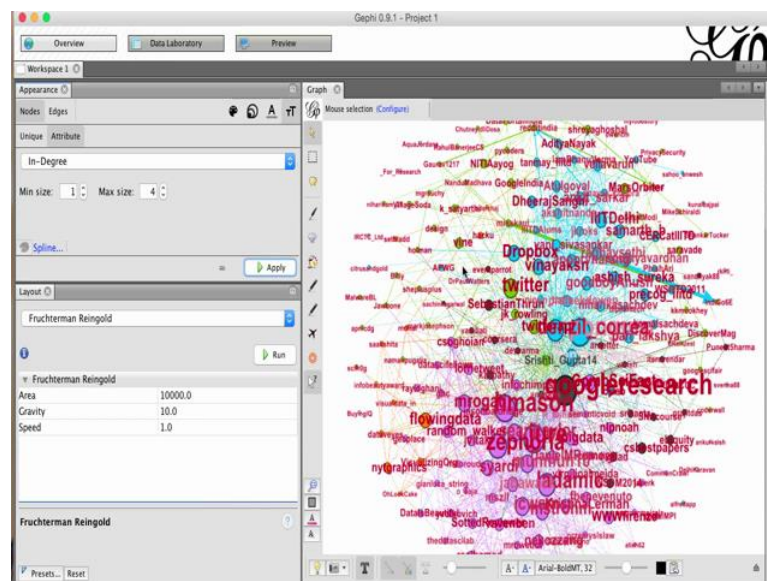
Let us see what happens if we choose in-degree. Select in-degree from the drop down menu and click on apply. Now, the size of the labels will be in line with the size of the nodes, because that is how we have sized our nodes.
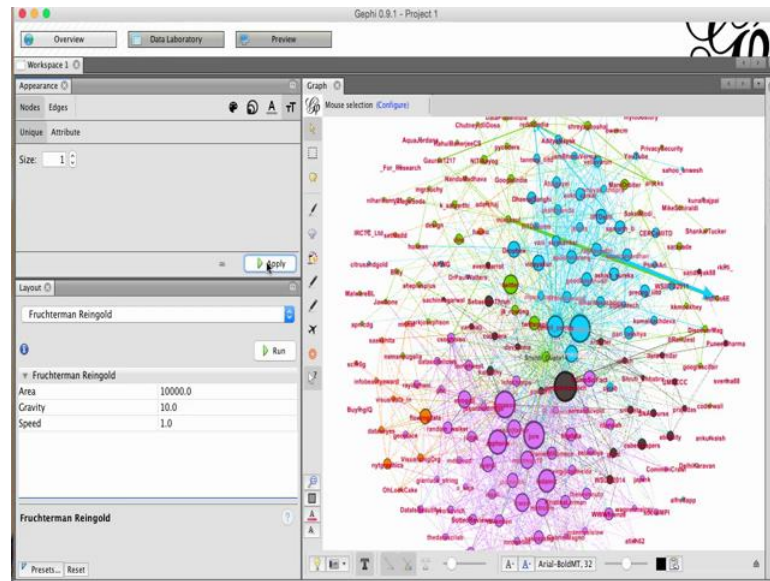
Let us change the node labels back to the unique attribute by selecting one so that all the nodes are of the same size. Now we have all the node labels back to the same size which is equal to 1.
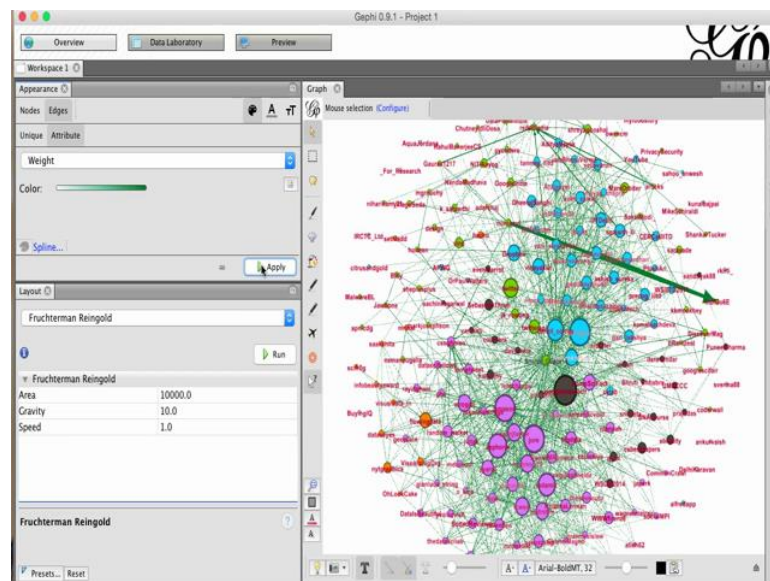
Let us explore more options to adjust the visualization of these graphs. We can also change the appearance of the edges; we can change the color, the label color and the label size of the edges. In our case, we do not have any labels for the edges. Therefore, we can play around with the color of the edges. Click on color, then attributes and then choose the available attributes. In our case, it is weight, click on apply.

(Refer Slide Time: 18:27)
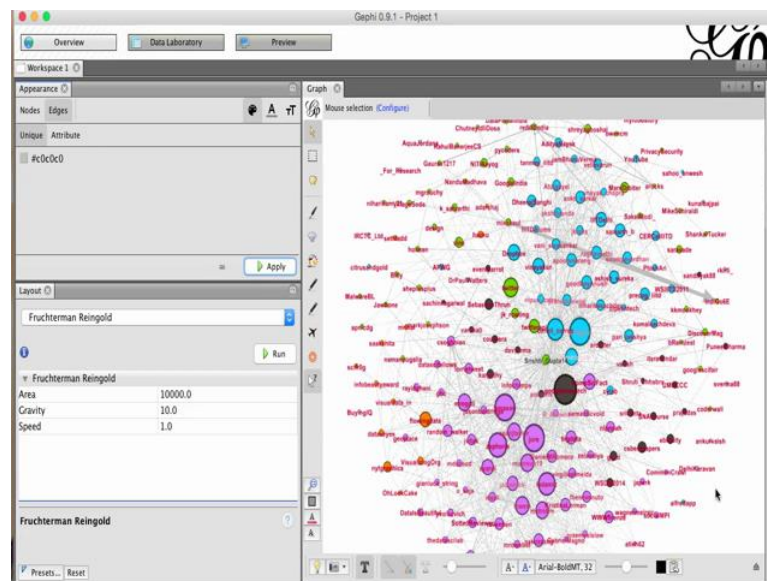
If we want to give the same color to all the edges, click back on unique and select the edge color and then click on apply. This will make all the edges appear grey.
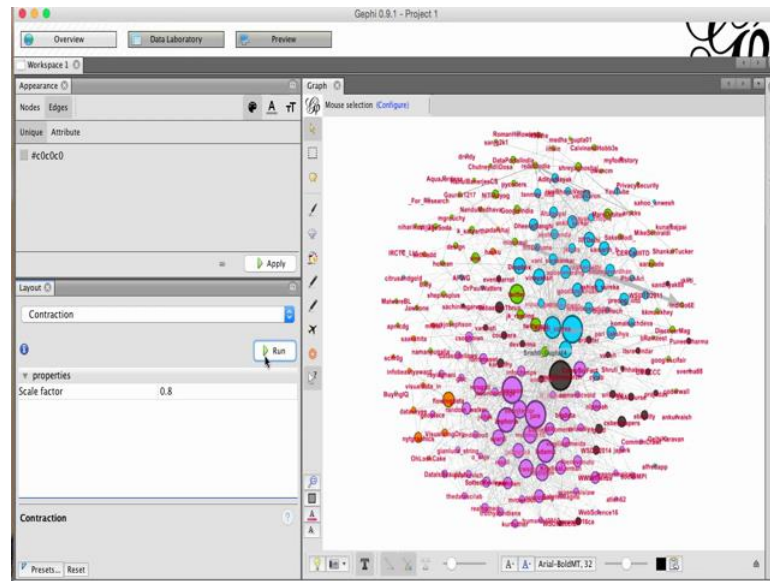
You can also try running the other available layouts in the layout tab. For instance, let us try running contraction; contraction will just zoom out the graphs in the current layout.
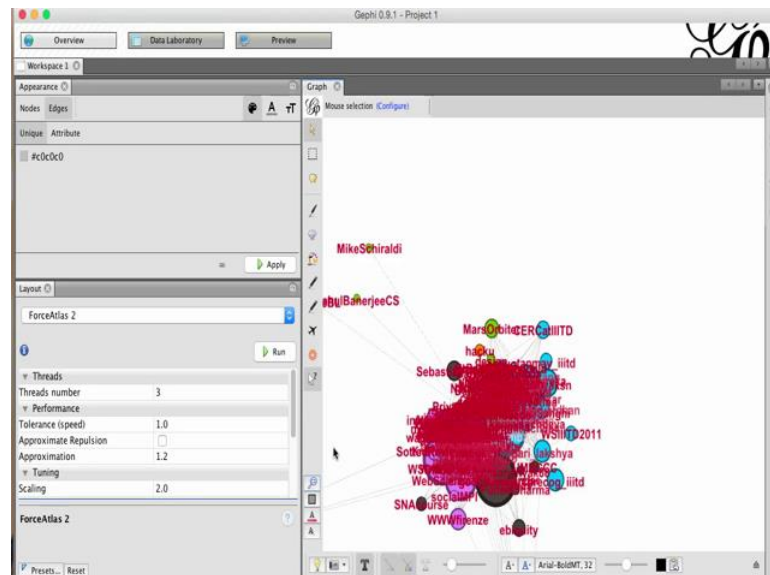
Let us try ForceAtlas. After the graph stabilizes, click on stop to analyze the graph further. Remember that if your graph goes off the screen, then find out this small

magnifying glass, which is used to center the graph in your blank space. And then you scroll down to zoom in the graph in the particular area which you want to see.

(Refer Slide Time: 19:48)
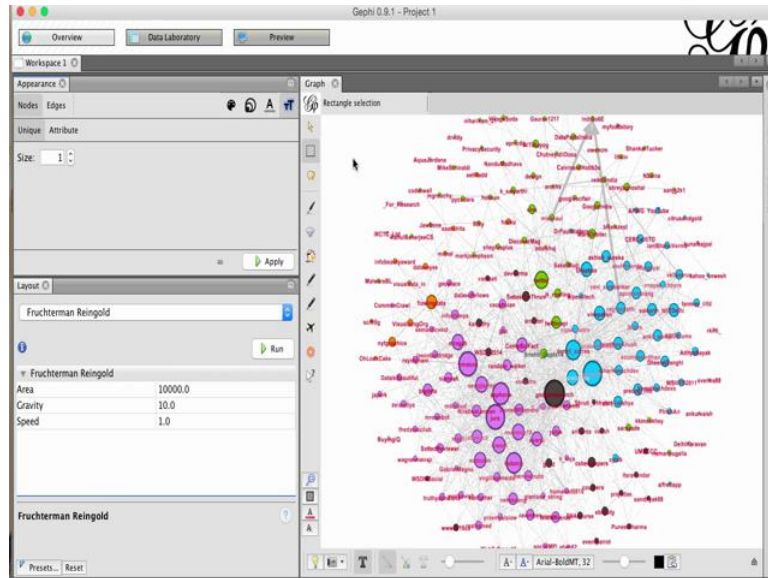
Let us again select the Fruchterman Reingold algorithm and click on run. When the graph seems stabilized, click on stop. Scroll up to zoom out of the magnified graph.

(Refer Slide Time: 20:22)
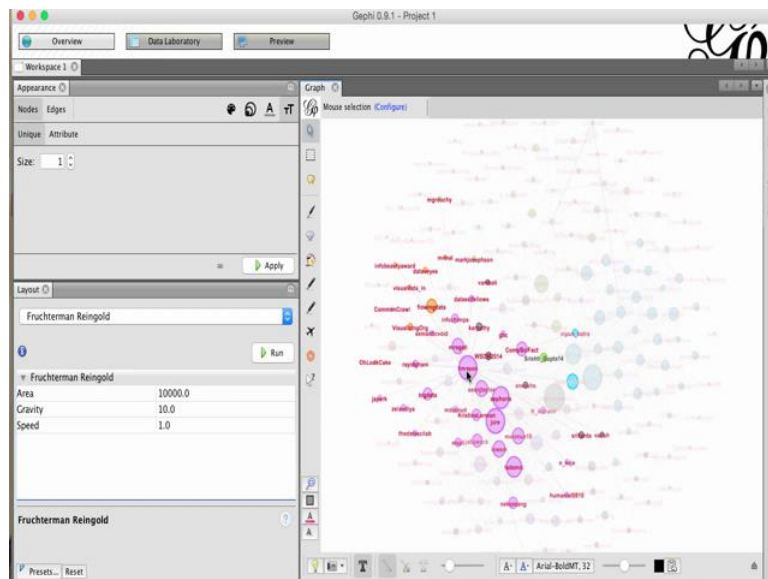
Now, let us look at more visualization options provided by Gephi. Let us look at the options on the left of the layout. The first option is the direct selection tool. It can help you select specific nodes and their neighbors. Click on the direct selection nodes and hover over a particular node. For instance, when we hover over this node it selects the particular node and it is neighbors. Now it's neighbors are automatically getting selected because of an option which has been opted n.

(Refer Slide Time: 20:53)
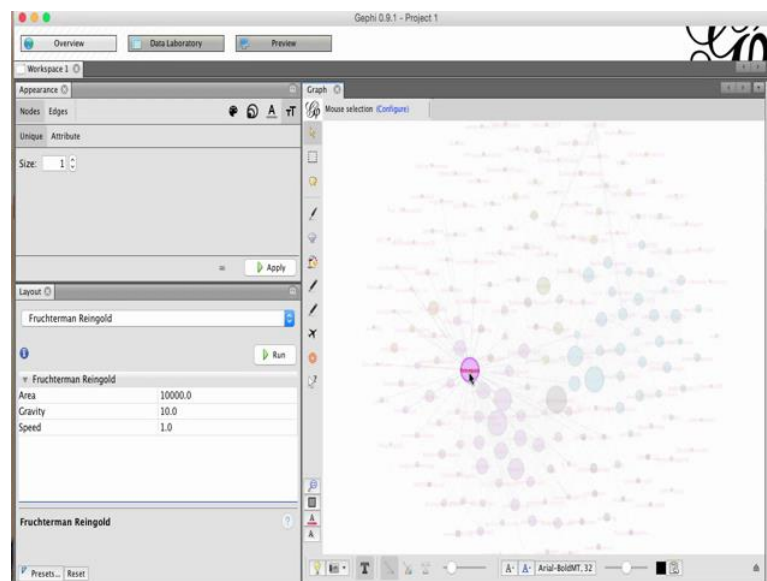
Let us look at that option. Notice the up arrow on the bottom panel, expand it, go to global, and you will see that the auto select neighbor option is ticked. Uncheck that and press the down arrow again. Now again hover over the previous node, this time the neighbors of this particular node will not be selected. If we want to turn back that option then there comes the up arrow again, and again select the auto select neighbor option. Now you will see the same happening.
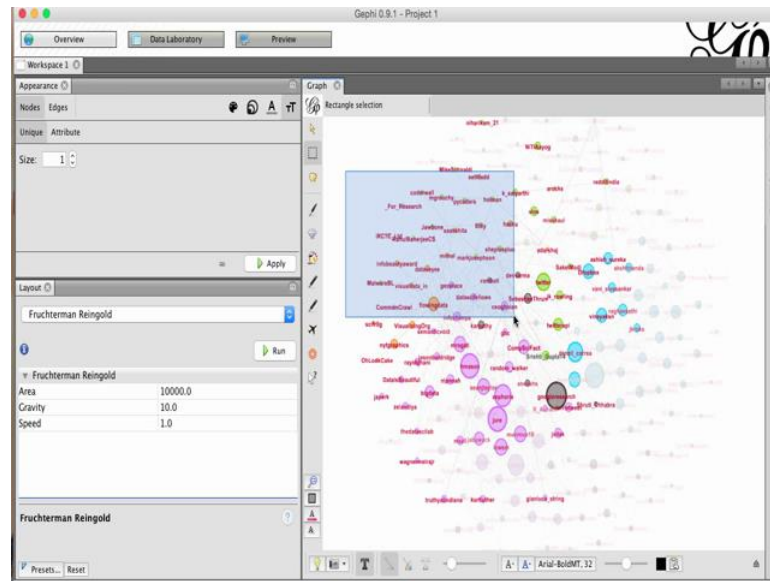
Let us look at the next option. The next option is the rectangular selection. Click on the rectangular selection. This will enable you to select a part of the node. Drag and select a rectangular area to select all the nodes and their neighbors in that particular area.
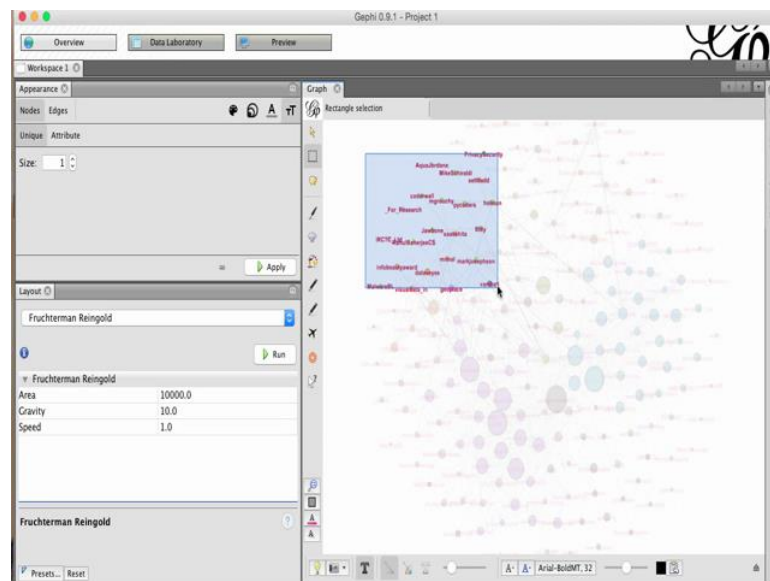
If you want to select only the specific nodes in the selective rectangular area, then we can again click on the arrow on the bottom panel and uncheck the auto select neighbor. Click the down arrow. And this time, when you select a particular area, only the nodes within that area will be selected.

(Refer Slide Time: 22:06)
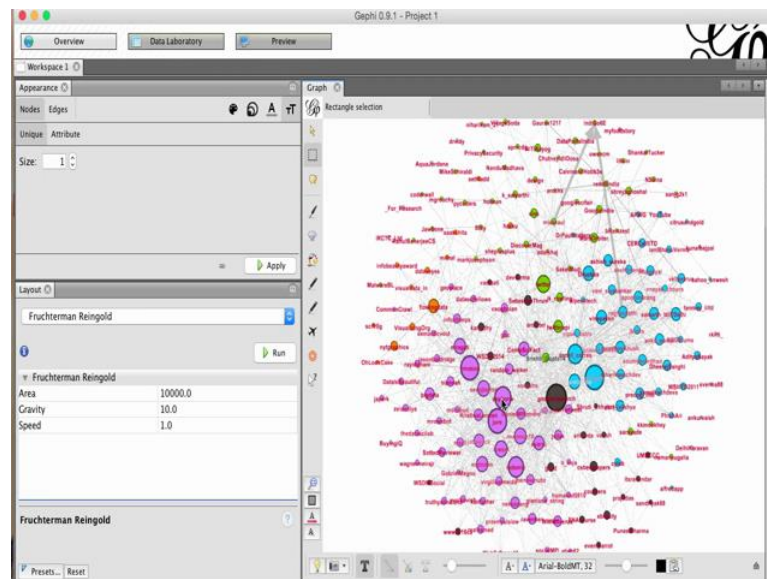
In this graph, we can see that two of these particular nodes are very close to each other.

(Refer Slide Time: 22:34)
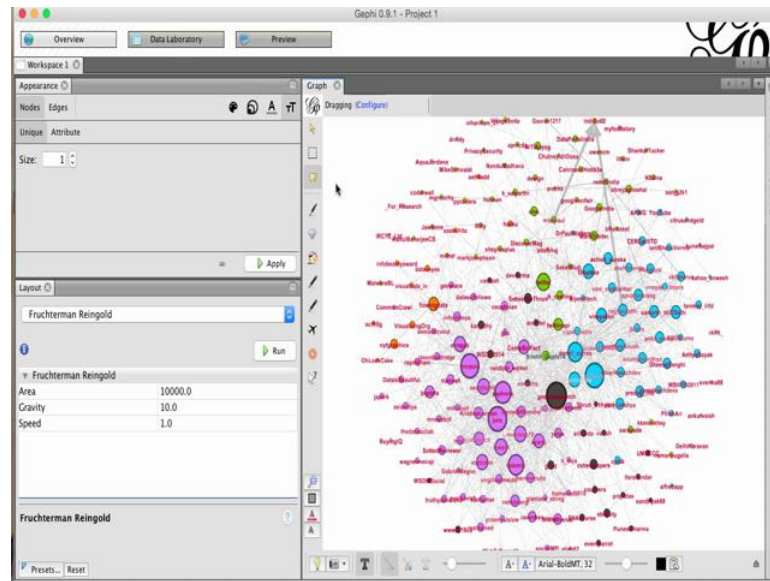
Let us say we want to make a small adjustment and make these nodes a little far away. To do that, we can select the drag option by clicking on it. Then, going to the node which is being overlapped with the other and dragging it just a little further off. The drag tool is used for manual adjustment of the nodes.

(Refer Slide Time: 23:07)
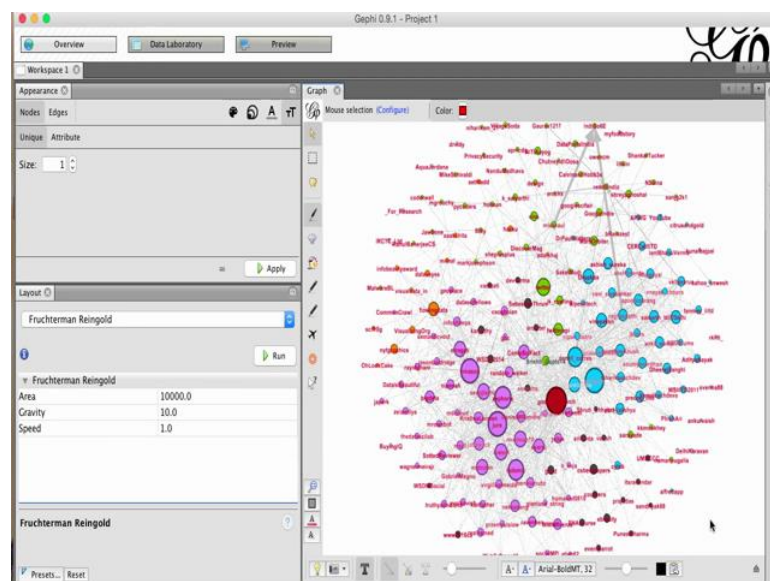
Let us look at the next option. The next option is the painter tool, which is used to color nodes by selecting the specific nodes. Select painter nodes and select the color, which you want to color your new node. Then go to your particular node and click on it. This will give it the color, which we have selected.
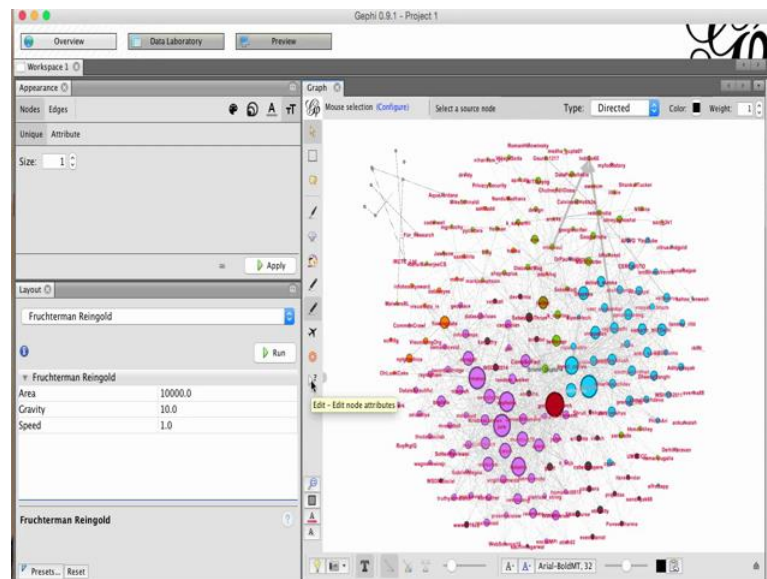
(Refer Slide Time: 24:25)

Now let us look at the option to add additional nodes using the node pencil. Select the node pencil and click somewhere in the empty space. This will create a node. You can add as many nodes as the number of new clicks, which you make. You can also add edges between the newly created nodes by clicking on the edge pencil and creating specific edges. You need to first select the source node and then the target node. Let us create some more edges.

(Refer Slide Time: 24:56)
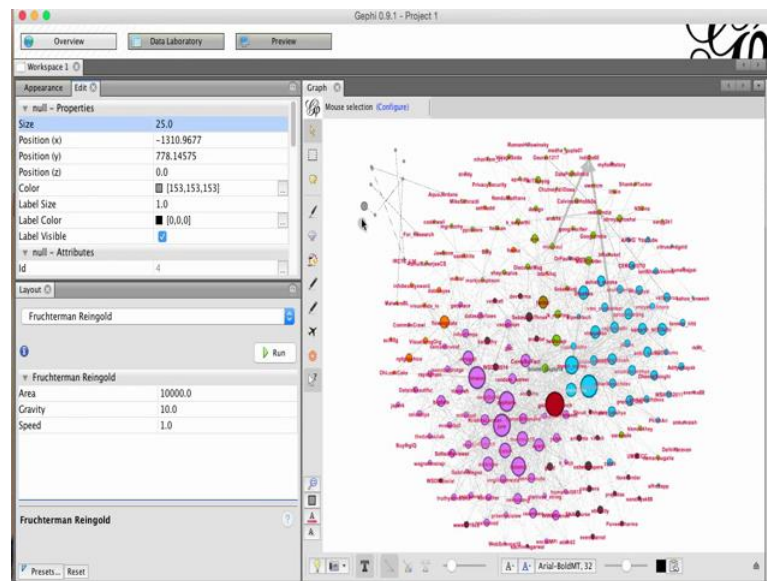
We can also edit the node attributes by selecting the edit tool. Click on the edit tool and then click on a specific node. This will elaborate the properties of the nodes in the left panel. We can change the node properties, as you like. Let us change the size of the node from 10 to 25; you can see that the size of the node has increased. Similarly, you can change the other parameters.

(Refer Slide Time: 25:15)
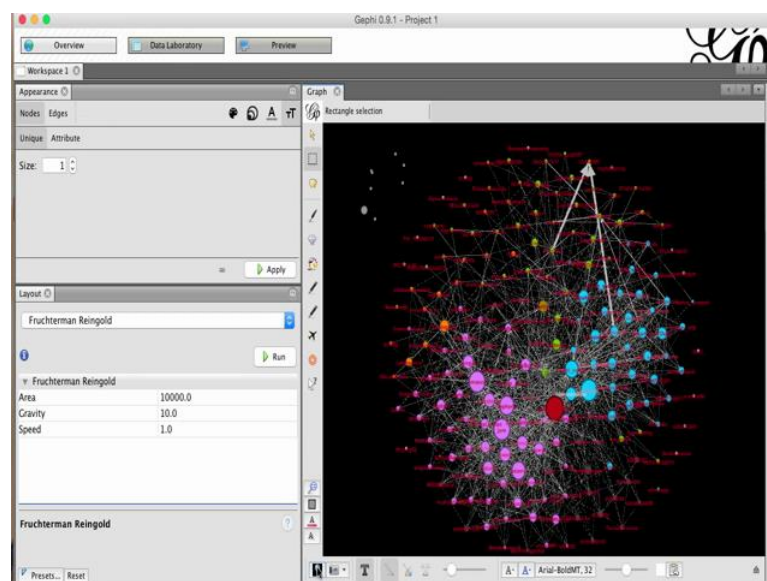
You can also change the background color of the Gephi visualization by clicking on the bulb icon on the bottom. Clicking on the bulb icon will toggle between black and white backgrounds.

(Refer Slide Time: 25:36)
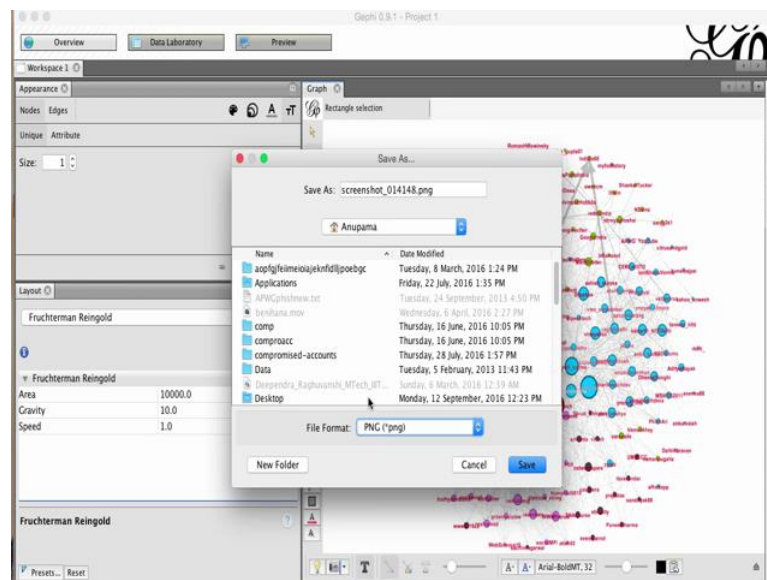
Gephi also lets you take a screen shot of the generated visualization by clicking on the camera button. This will generate a screen shot in png format, and you can save it anywhere you want.

(Refer Slide Time: 26:14)
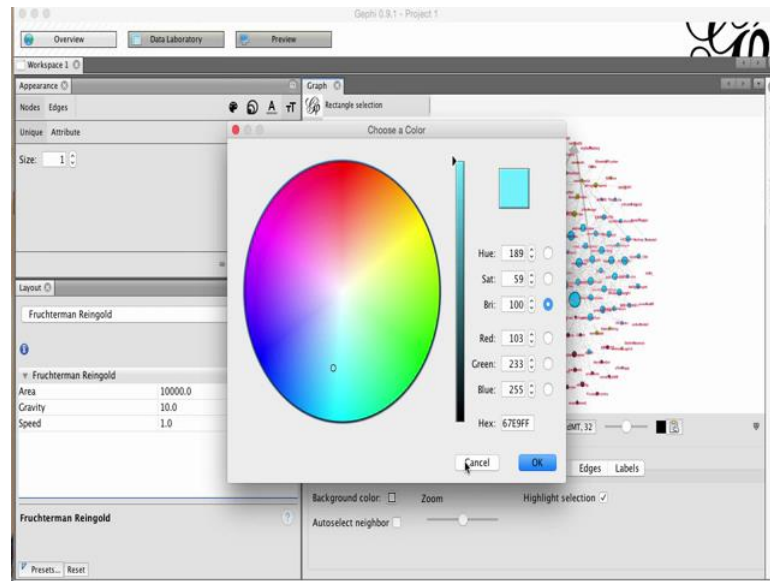

Now, let us look at the options in the bottom panel in more detail. Expand the bottom panel using the up arrow at the bottom, and click on global. From where you can customize the background color not just black and white, but also a customized color. Click on the background color and you can choose any color, which you want. You can change it back to the default color by again clicking on the background color option.

(Refer Slide Time: 26:38)
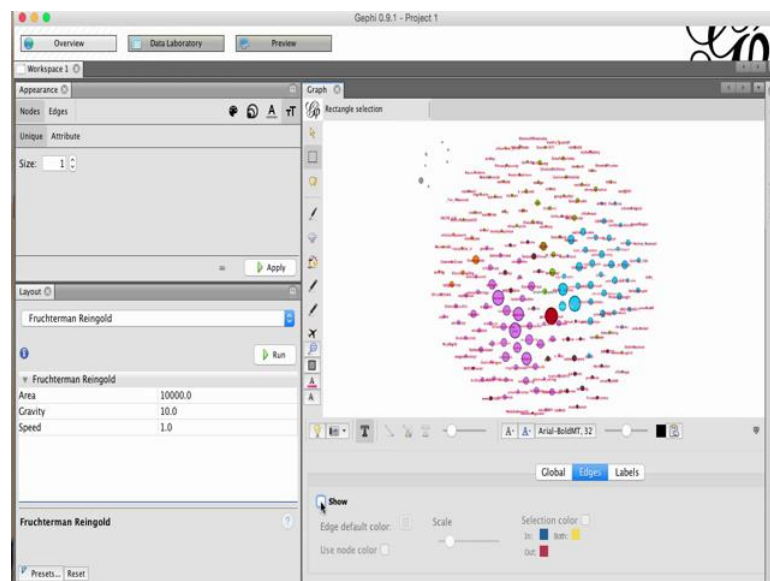
Let us click at the edges tab. The show option is clicked. If you uncheck the option, all the edges will be invisible.

(Refer Slide Time: 26:44)
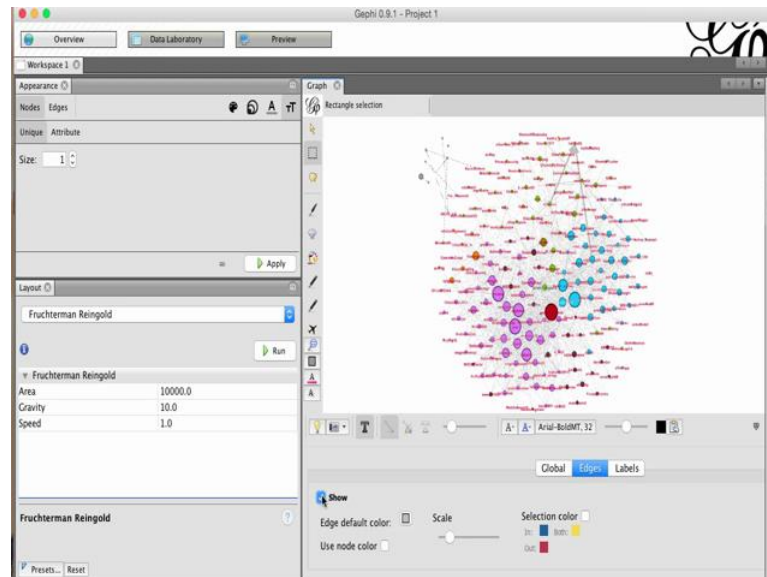
Click it back to make the edges visible. From this panel, you can also decide whether you want to keep the color of the edges same as the nodes or not. If you choose the use node color option then all the edges will become of the color of these specific nodes, which they are attached to.
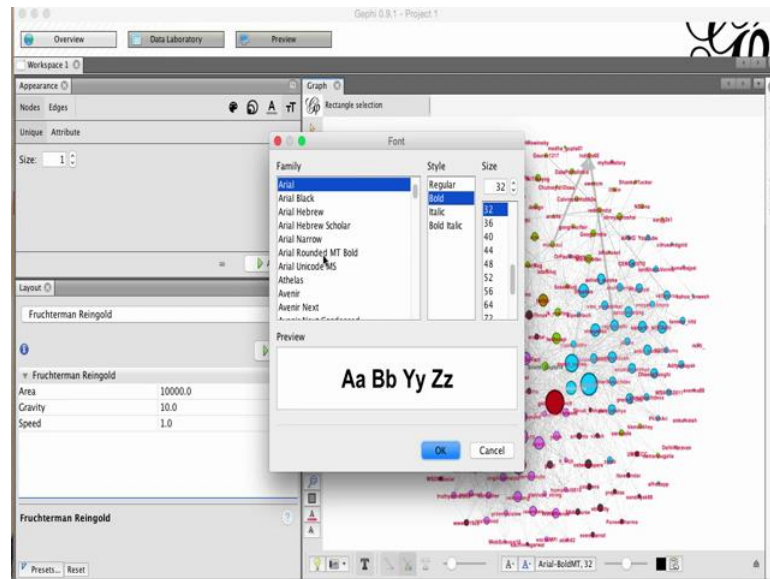
You can also change the font of the node labels by clicking on the font value at the bottom. Select the new font, which you want, and click on ok, this will change the font for all the node labels.
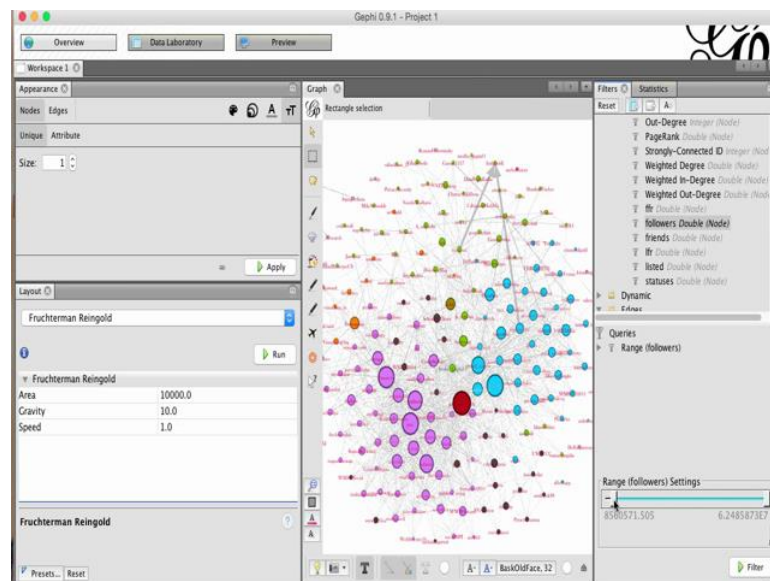
Now, let us look at the filters tab in Gephi on the right. Click on the filters tab. And let us select a filter to choose specific nodes and edges. Let us click on the attributes filters and click on range. Click on followers and define the range.
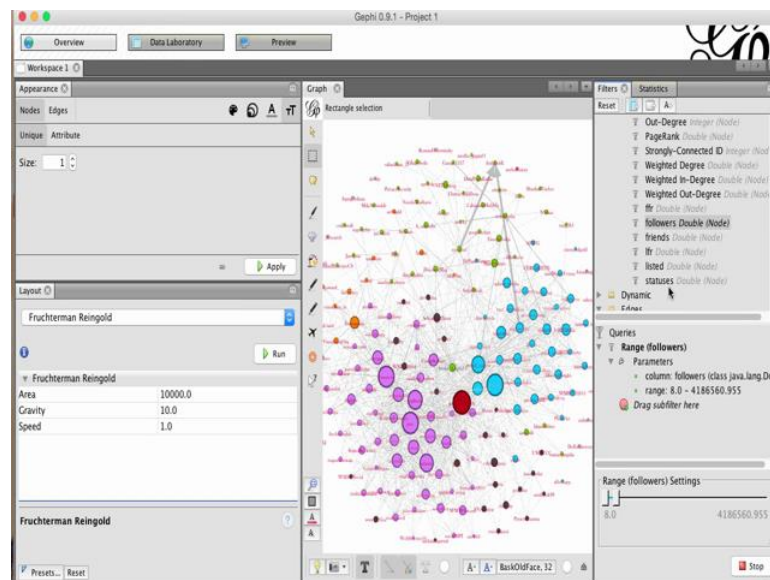
(Refer Slide Time: 28:35)

Let us say we want to keep the number of nodes, let us say we want to keep the nodes which have the followers between 8 to 4 million. Use this sliders accordingly and click on filter. This operation will remove the nodes which do not fall in to the particular range which we have defined.
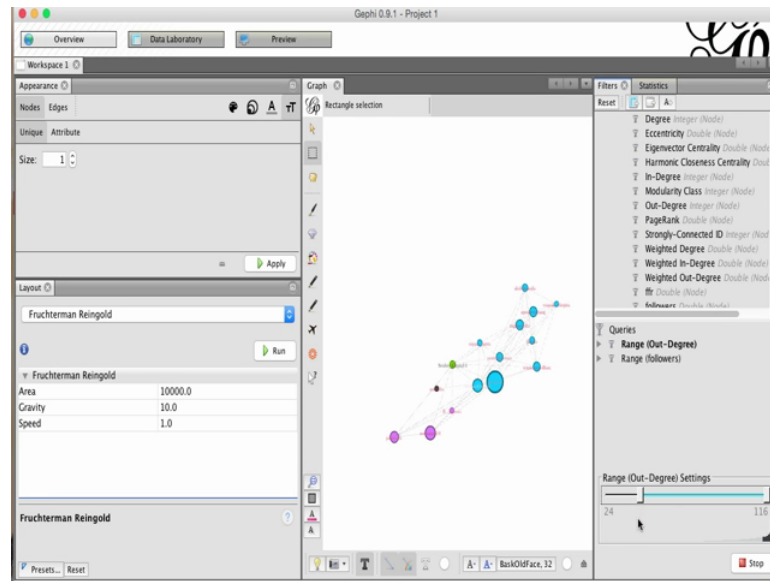
We can also have additional filters. For instance, let us choose a filter based on the out-degree. Select it in the queries and drag the slider to change the values according to what you want. We can see that there are only few nodes, which have out-degree between 24 to 116. If click on stop then we will be able to see the entire network. Let us change the value of the range of the out-degree settings.
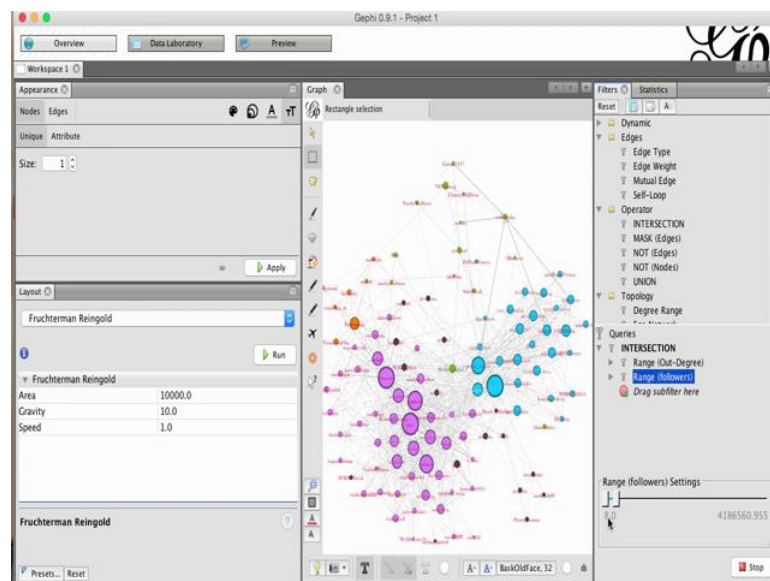
Remember that we also learned about the operator filter. We can also use multiple operations to combine the filters in the queries tab. Let us try using the intersection. Select the already existing filters, which is range based on out-degree and range based on followers as the sub filters in the intersection filter. This will give an overlap of the two filters which is the nodes which have an out-degree between 4 to 116, and also the nodes which have the number of followers between 8 to 4.1 millions. If you click on stop, you will be able to see the entire network.

(Refer Slide Time: 31:16)
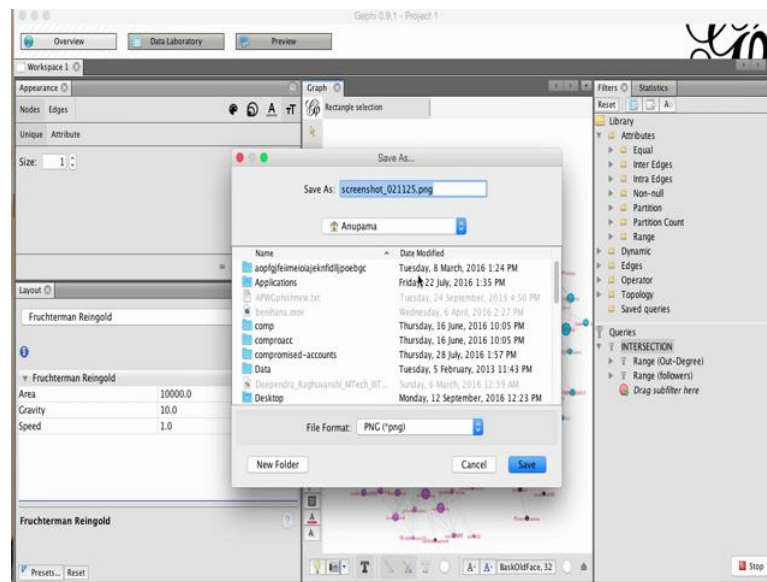
You can use various other filters in similar fashion. You can also save the filtered graph by first applying the filter, and then clicking on the screenshot tool to save this specific graph.
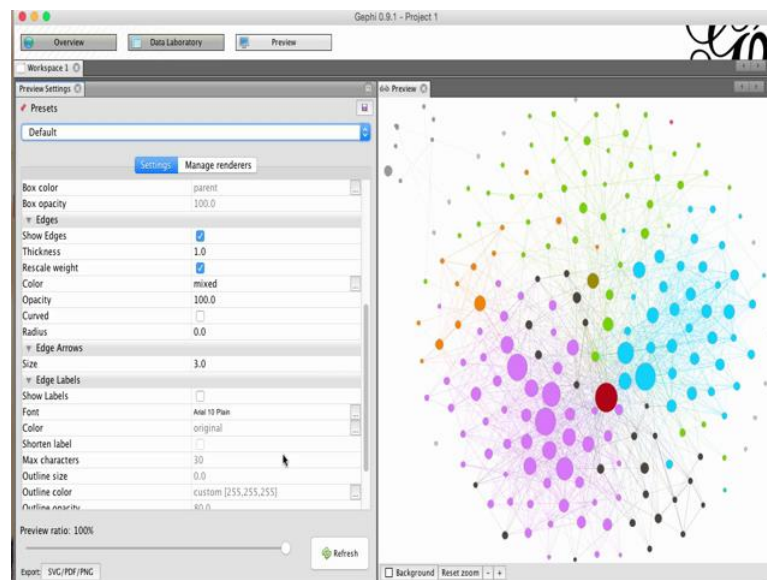
(Refer Slide Time: 32:36)
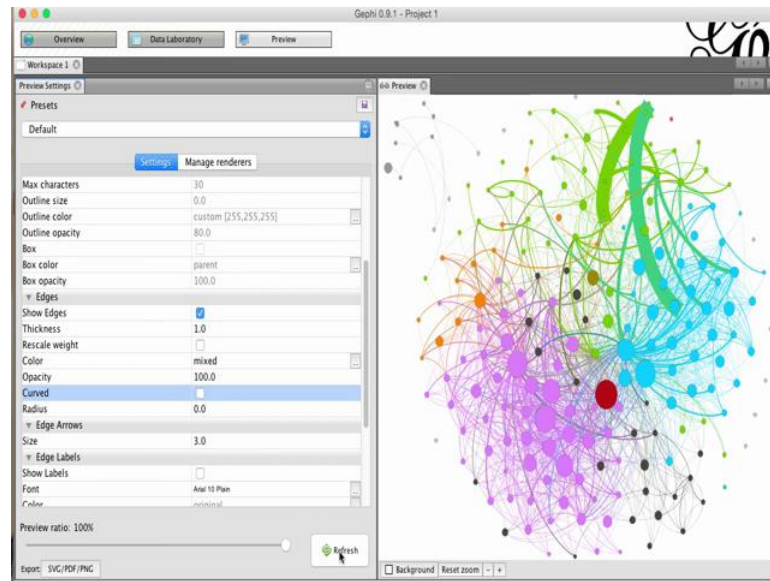
Now, let us collapse the filters tab and look at the preview tab of Gephi. The preview tab will be on the top right. The preview tab provides advanced option to adjust the visualization of the generated graph. Once you click on refresh, you will be able to see the graph visualization on your right.

(Refer Slide Time: 32:43)
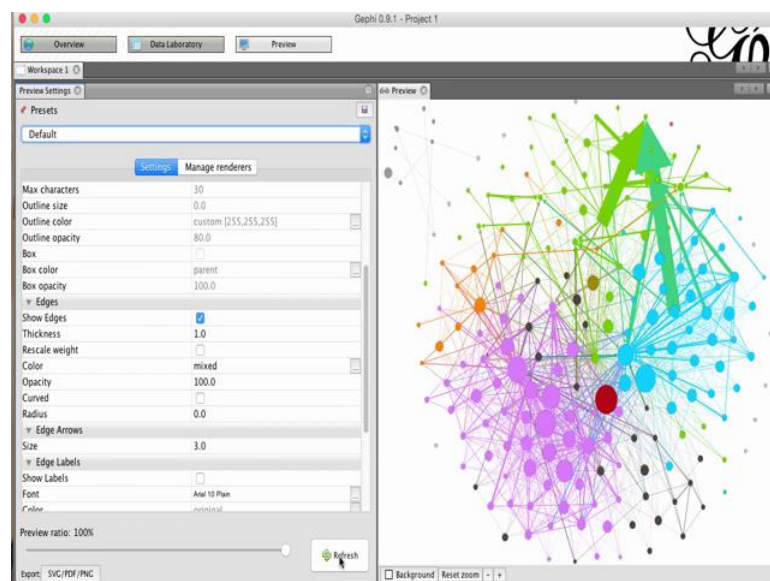
In this graph visualization, the edges are curved, because that is what is present in the default options; scroll down in the presets default in edges and unselect the curved option again click on refresh. Now the edges will be straight and not curved.
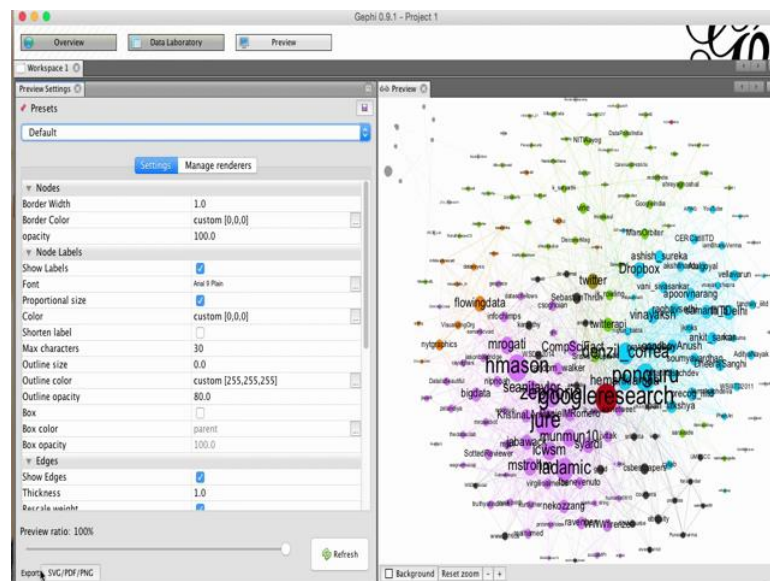
(Refer Slide Time: 33:10)

If you want all the edges to be of the same thickness, then click on rescale weight and click on refresh. This will cause all the edges to be of the same thickness.
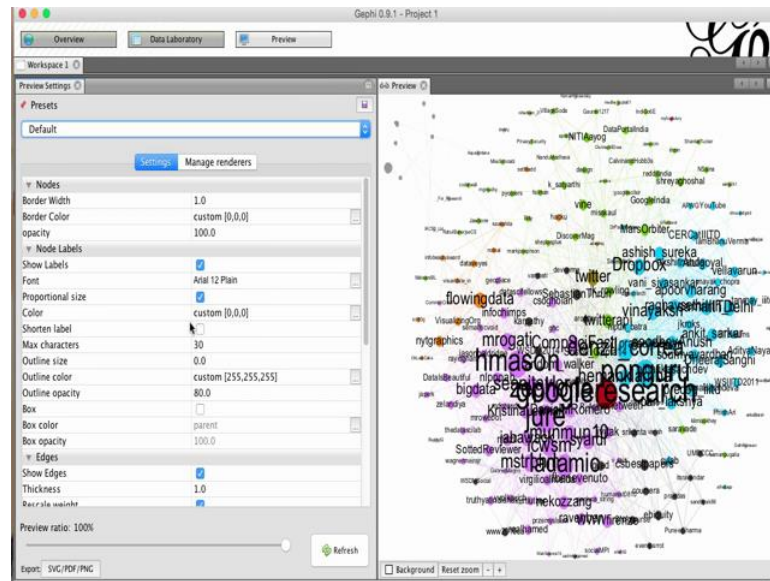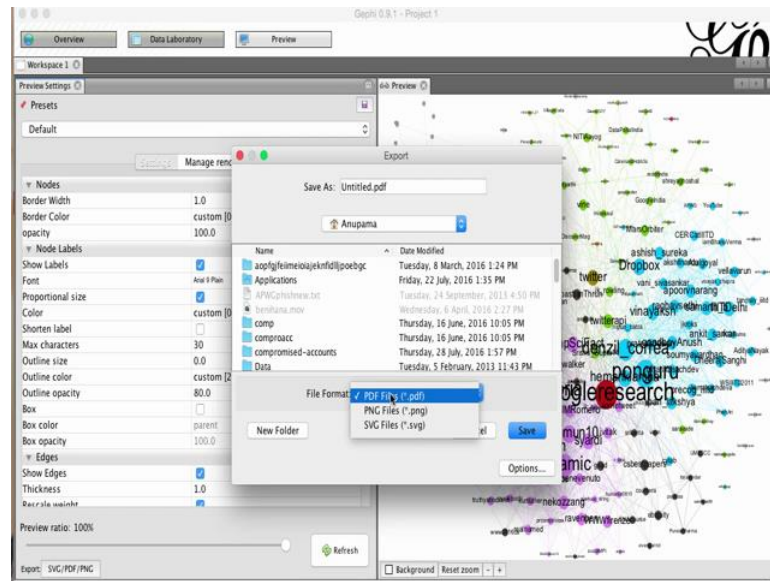
Right now by default, we are not even able to see the node labels. To be able to see the node labels, click on the show labels option under the node labels, and again press on refresh, now you will be able to see the node labels and you can adjust your font size.
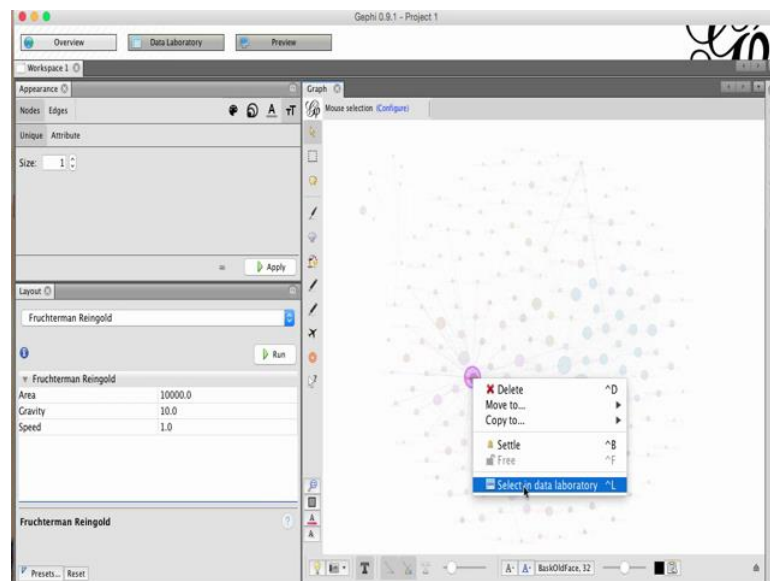
Let us change the font size to 9, and press refresh.

(Refer Slide Time: 34:08)


The preview tab has it's own export command at the bottom left. If you click on the export command, it will give you several options to store the generated file in either pdf format, png or svg.
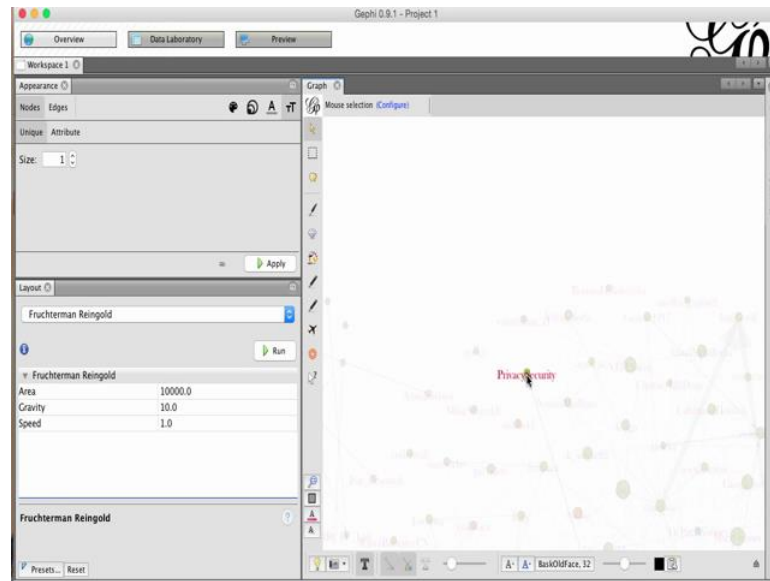
(Refer Slide Time: 34:55)

Now let us go back to the overview tab to view the visualization. In case, you want to know in detail about a particular node in the data laboratory, then you can first go to the select tool, hover over particular node, right click, and then click on the select in data laboratory option.
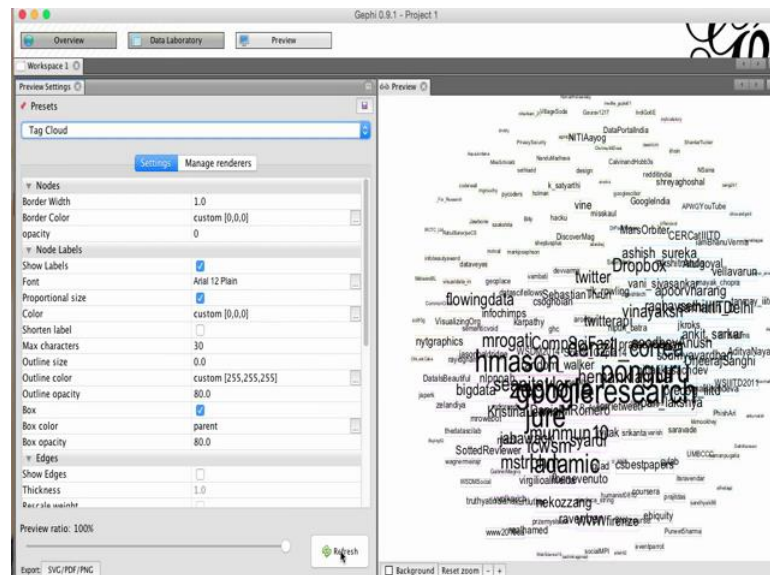
(Refer Slide Time: 35:25)
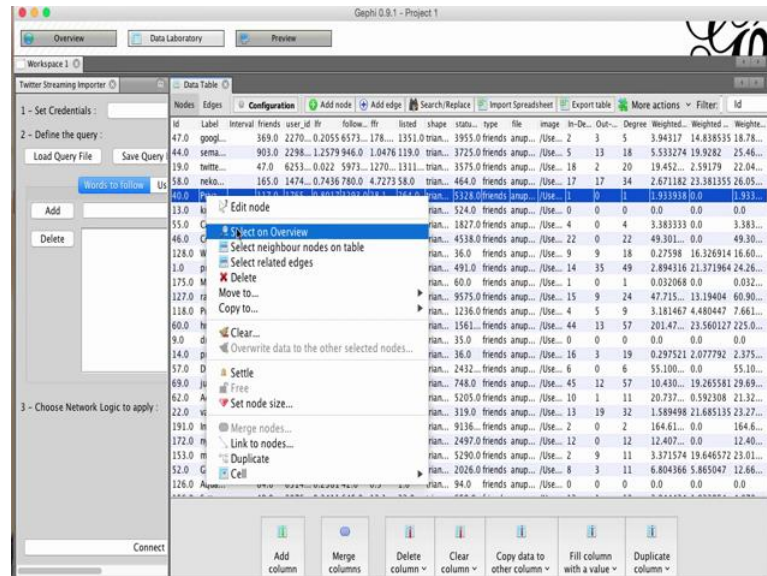
If you go to the data laboratory now, that particular row will be highlighted. And you can look at the particular node in more detail and the nodes tab. Let say, you want to select a particular node in the visualization from the data laboratory, then go to the particular row, right click and then select select on overview.
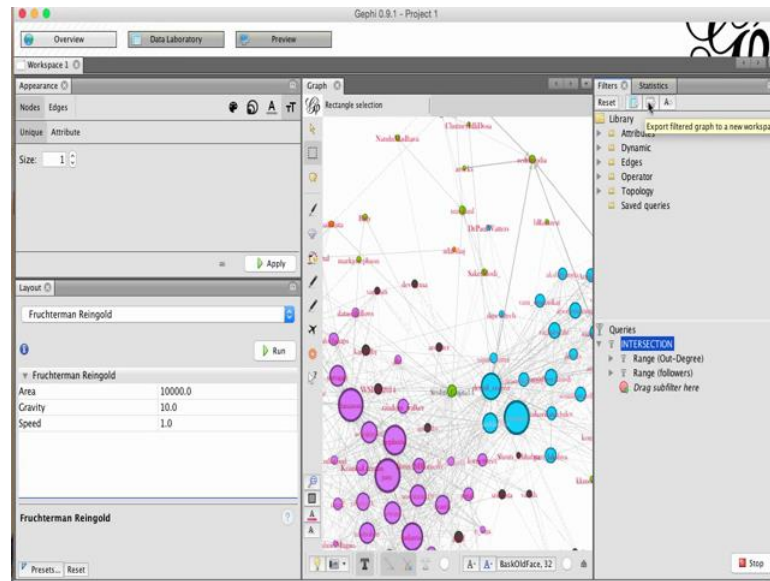
Now when you go to the overview, that particular node will be highlighted. This is an easy way to refer to nodes back and forth from the data laboratory to the visualization.

Let us go back to the preview tab. Let us look at more options in the drop down menu; click on the tag cloud and press refresh. This will generate a tag cloud of the node labels.

We can also select a subset of the graph in a new work space while maintaining the current work space. To do that, let us click on the filters tab again, and apply a filter. Let us choose the previously existing filter which we just tried and click on filter.
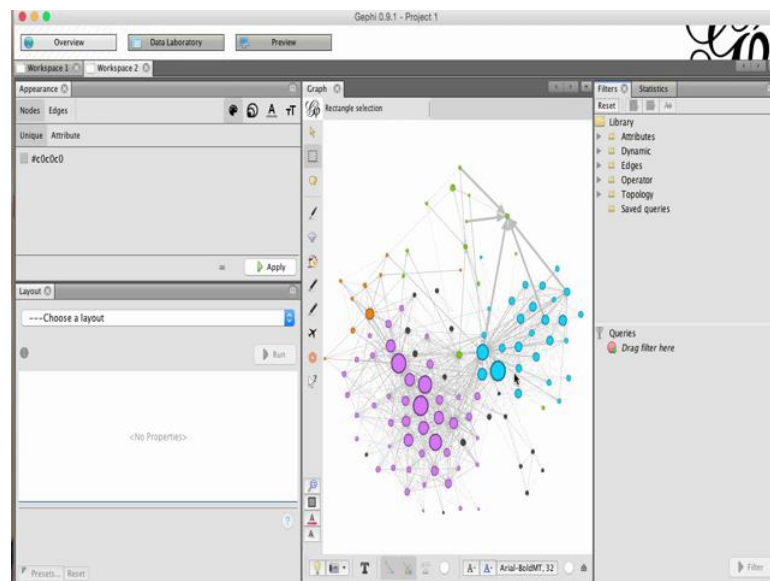
Look at the second option at the top of the filters tab which says export filter graph to a new work space. Click on the option and it will generate a new work space which is work space two. Go to the work space two. And you will be able to see only the subset of the graph based on the filter which we choose earlier. Now, this does not have any existing filters applied to it, you can consider it as a fresh graph to work over and apply further filters, change in visualization etcetera over it.
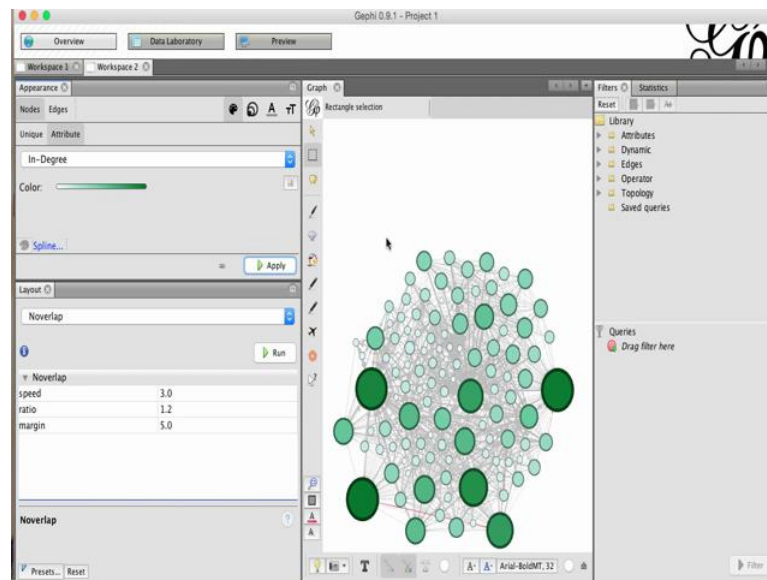
(Refer Slide Time: 37:44)
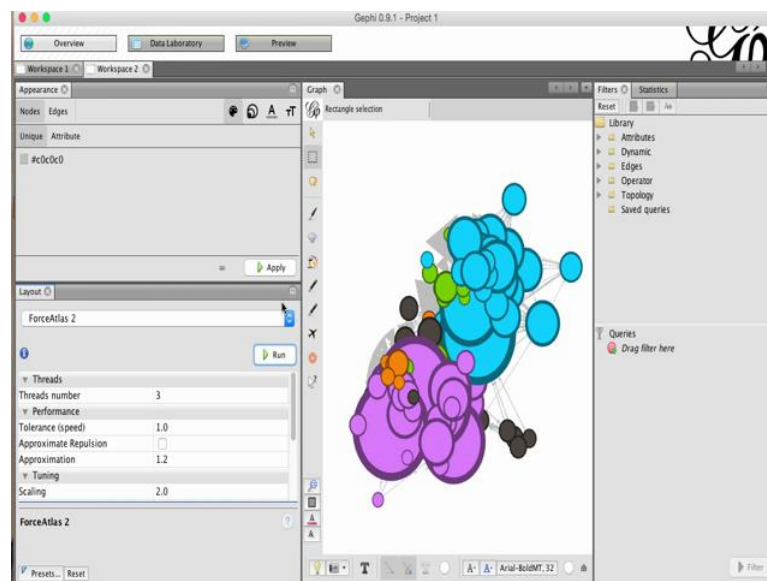
As an example let us try changing the layout of this sub graph. Click on forceatlas 2 and apply the algorithm. Click on stop after the graph stabilizes. You can zoom in to see the resulting graph.

(Refer Slide Time: 37:53)
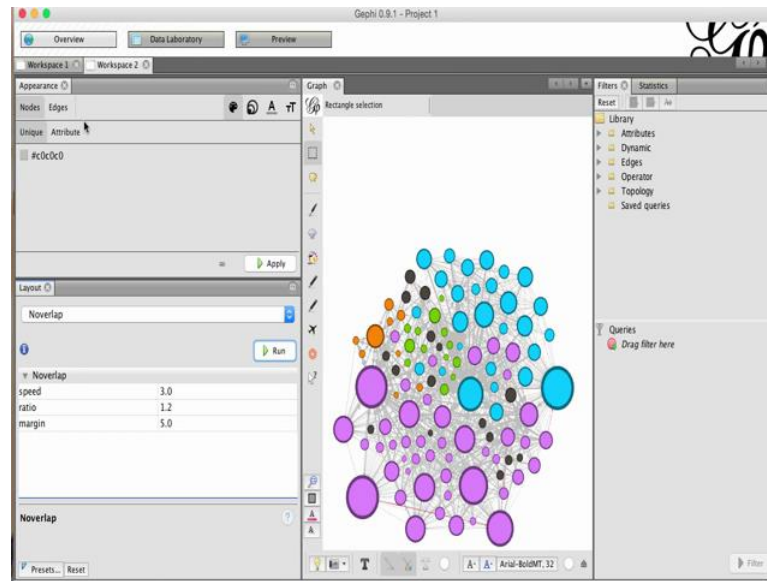
You can also click on no overlap, so that the nodes no longer overlap each other.
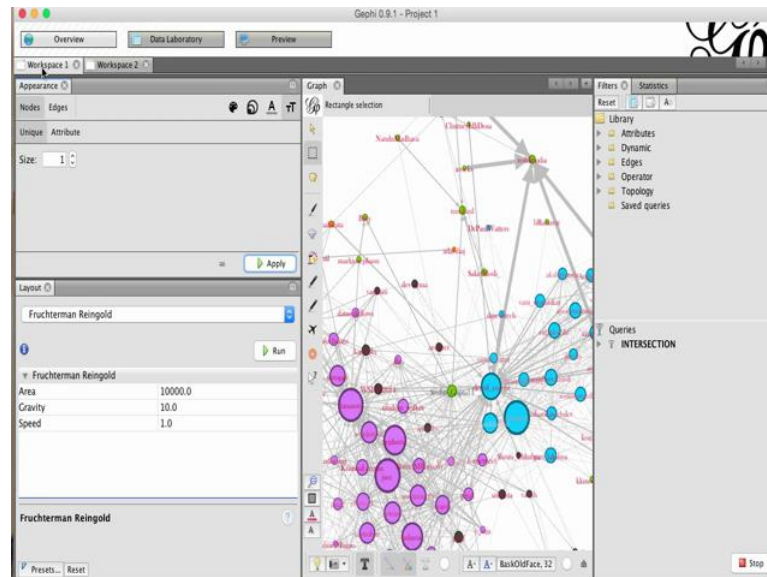
(Refer Slide Time: 38:37)

You can also recolor the nodes based on any new attributes, which you want. For instance, now for this sub graph, let us choose an attribute which is in-degree and apply.

This way you can create new work spaces by filtering portions of the original graph and have more detailed analysis on sub section of a graph.

(Refer Slide Time: 38:47)

To go back to the original graph, you can just click on work space one from the top left corner.

This completes our tutorial on Gephi. And you can generate more graphs based on any data, which has nodes and edges.