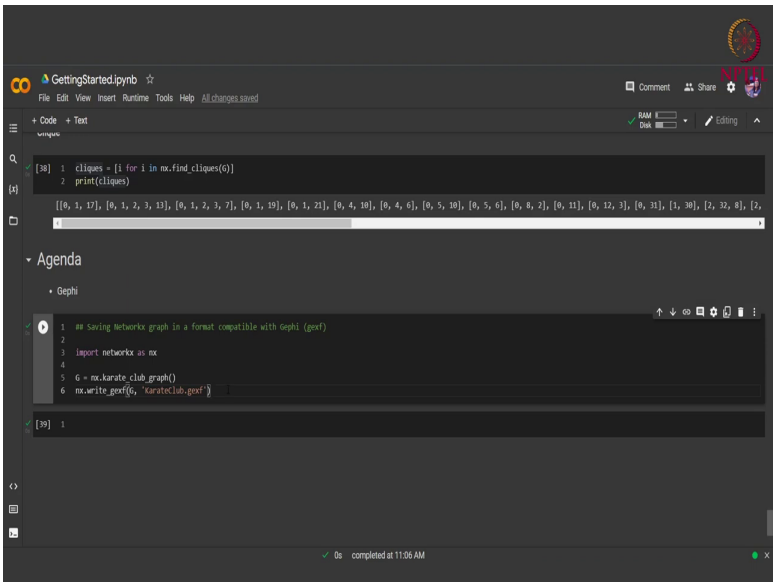
**Social Network Analysis**
**Prof. Shivani Kumar**
**Department of Computer Science and Engineering**
**Indraprastha Institute of Information Technology, Delhi**

**Lecture - 26**
**Tutorial - 4**

Hello everyone. Welcome back to the tutorial of Social Network Analysis. So, we have already seen different attributes of network x and how we can use it to analyze the graph and maybe we for some basic visualization of the network; however, network x is provides us with very limited options for visualization. And for a network to be understand understood fully and to be analyzed fully, we might need to visualize it in a more better and interactive manner.

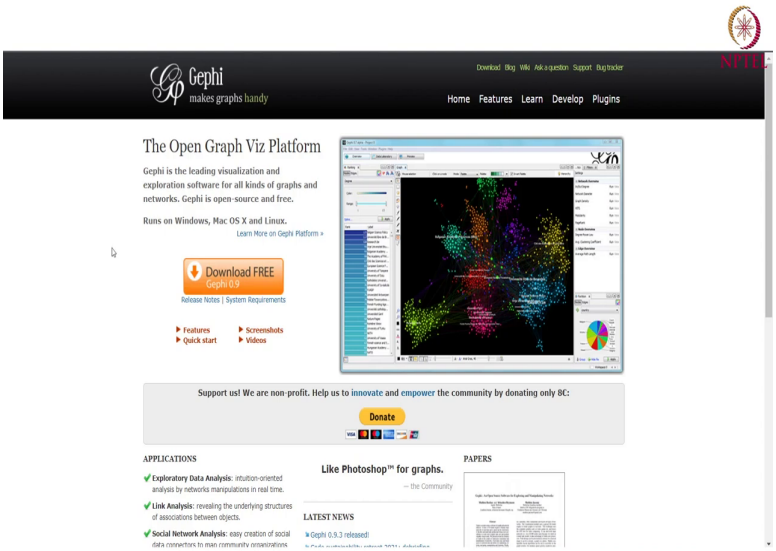
In order to do that, there are various softwares that are present to be downloaded or online, both in the free and the open source domain. One such software that we will discuss in the tutorial today is called Gephi.

(Refer Slide Time: 01:15)




So, Gephi is basically an open source software, which can be downloaded for Windows, Mac or even Linux. And it can be used to visualize as well as do some basic analysis of the network. So, you can download Gephi.

(Refer Slide Time: 01:32)



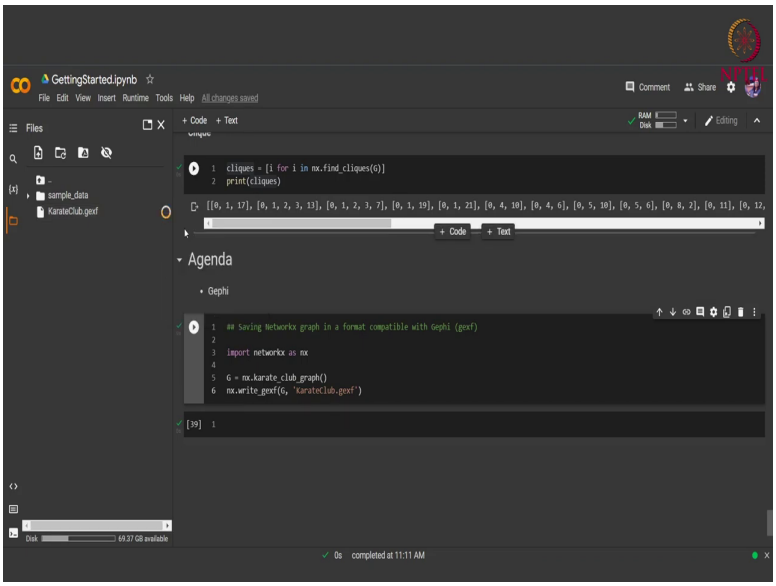

From this Gephi dot org website and it is as I already mentioned it is available for Windows, Mac and Linux. And after you have downloaded it just install it in your systems and then we are good to go. Now, since this software is already present in my system. I will just tell you what to do after that is it is installed.

So, we already saw network x right. So, now, whatever graph that we create in network x. We can save it in a format that is suitable for the Gephi software. So, the format it can be graph m l also it can be g e x f. So, network x it provides us with a bridge in order to save a graph in a suitable manner.

So, that we can open this graph in the Gephi software and visualize or analyze it further. So, we again we as done in the previous tutorial. We will take this karate club graph the sample graph and we will just write this graph in the dot g e x f format file. So, in order to do that, we

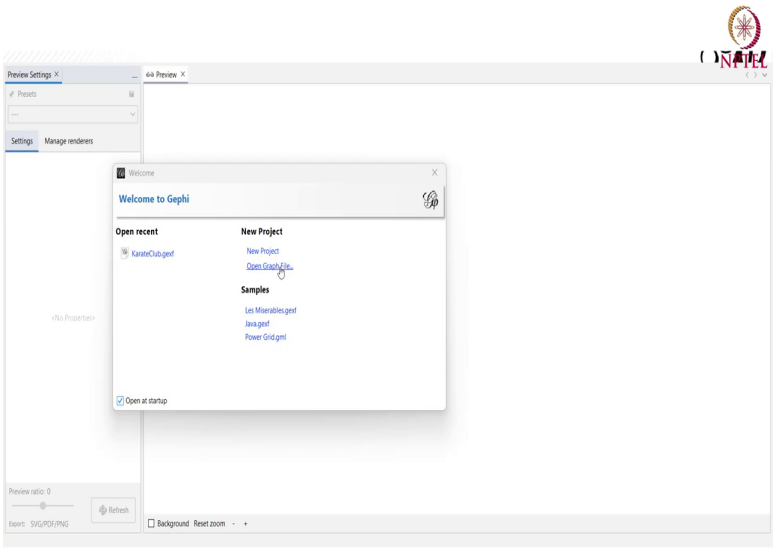
simply call the right dot g e x f function of the network x library and we run this step. So, now our karate club graph has been written in this in this particular file. Now, how to access this file in Google collaborator?.

(Refer Slide Time: 03:13)



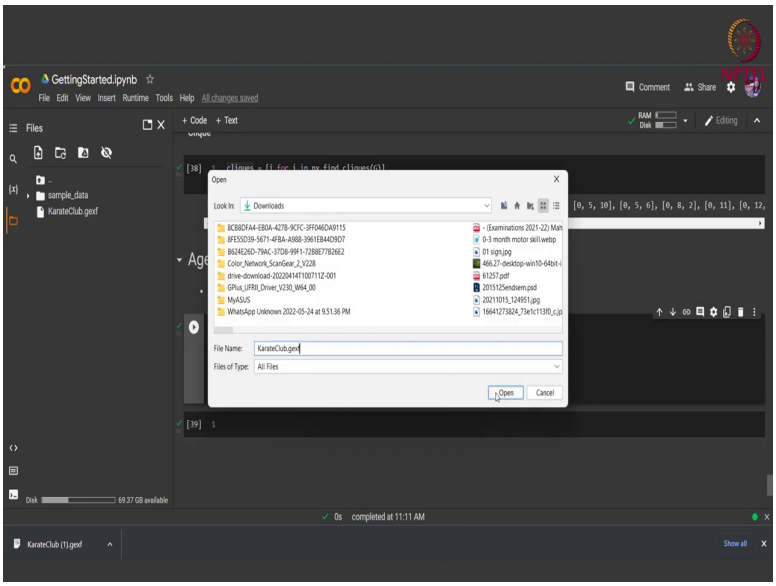

We just open, this files tab here and we see that the graph that we have just written is present here as a file. We go to the options of this and we download this file for us. So, here we have the file. Now, the next step would be to open this file in the Gephi software. So, we open the Gephi software and when we will open it.

(Refer Slide Time: 03:45)

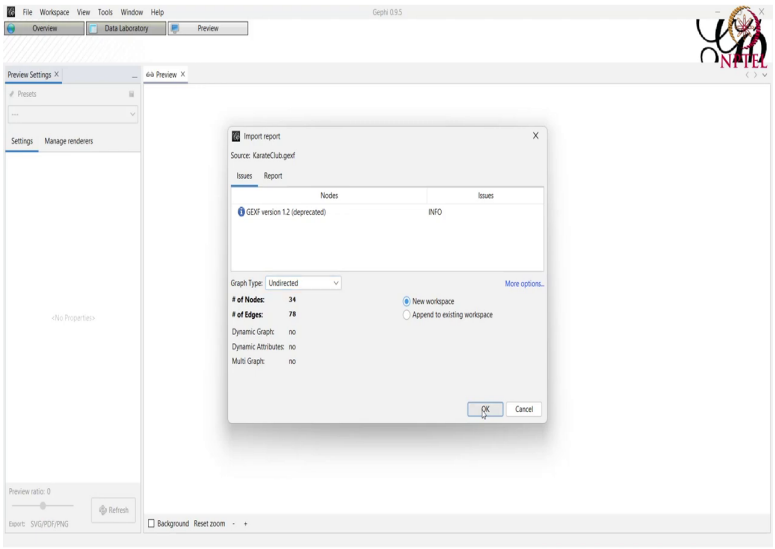



You will see a screen like this. Here we have some different options like, we can either create a new project from scratch where we will be creating a graph and the edges and the nodes in the Gephi software itself. We can open a graph file that is something that we will be doing. And also it provides that with some already present different samples graph, just to play around and get to know the software well.

(Refer Slide Time: 04:19)



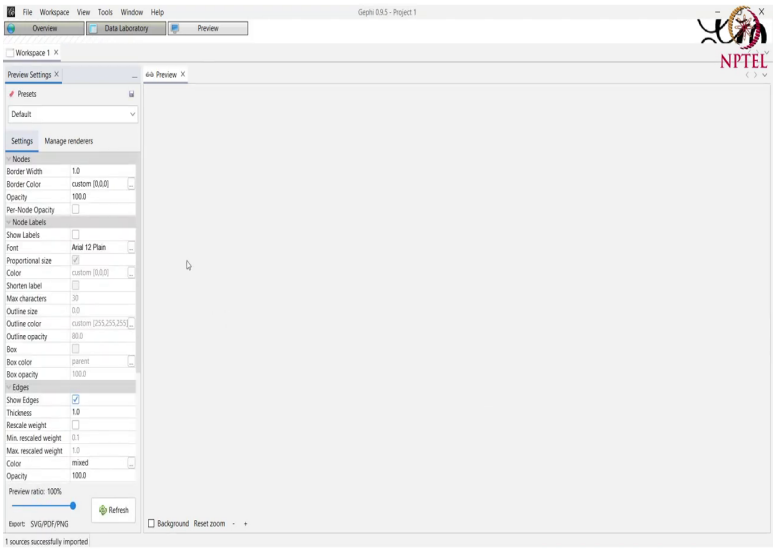

(Refer Slide Time: 04:30)



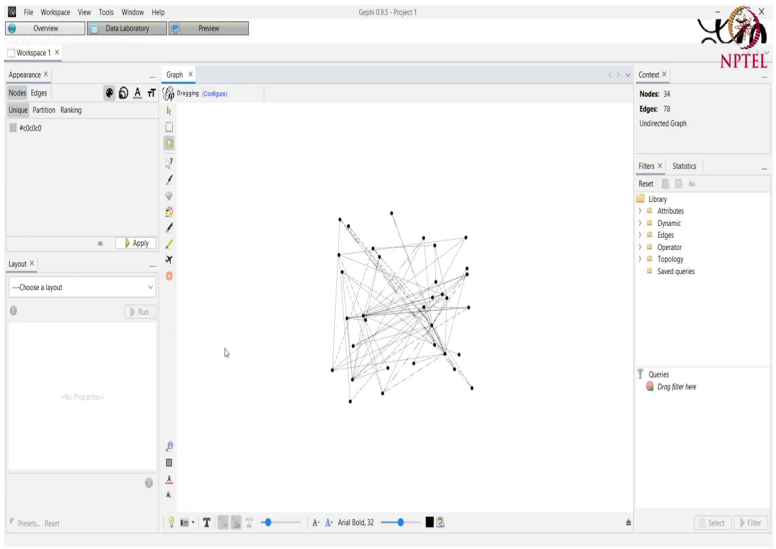

We will open the graph file that we have downloaded. That is the karate club dot g e x f file. We will open this file. And we see that we are shown with some information of the network. That is there are total of 34 nodes, that are present in the network and the number of edges are 78 and the graph type is undirected, that we want.

So, we simply look at these components we see that they are good to go and we press on ok. So, that a graph based on these nodes and edges can be initialized.

(Refer Slide Time: 04:58)



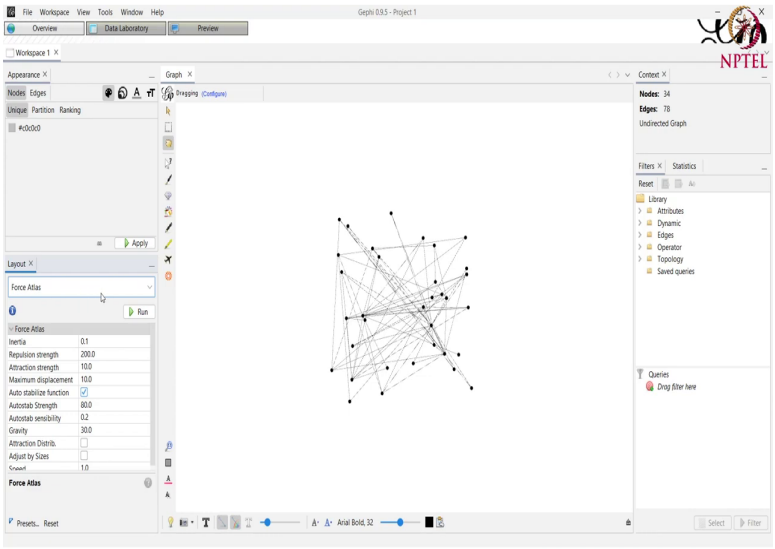

(Refer Slide Time: 05:02)




s

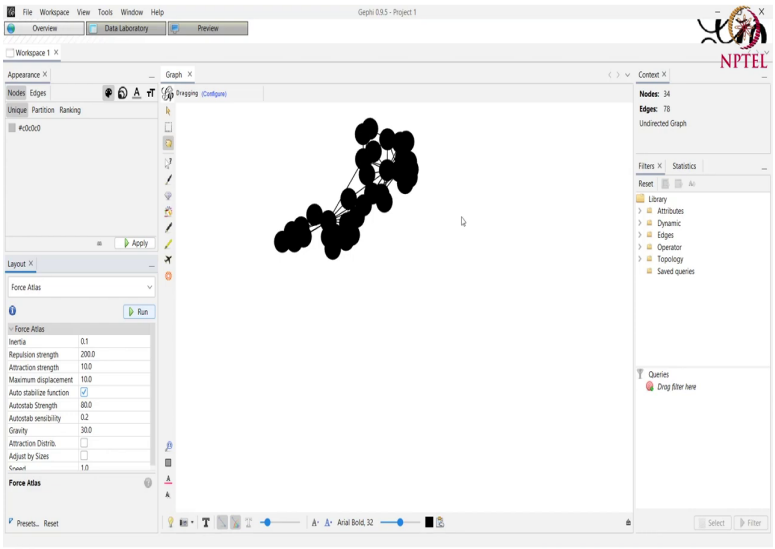
After pressing on ok, I am going in this overview tab. We can see that here a graph is coming which basically has 34 nodes and 78 edges as described in the in this right hand side top box and, but; however, right now this graph although we can understand some structure of it, but still the overall structure might not be as clear as we might want. So, we might want to change the layout of the graph a bit. We can do that in Gephi by going into this tab here at the bottom left corner.

(Refer Slide Time: 05:44)



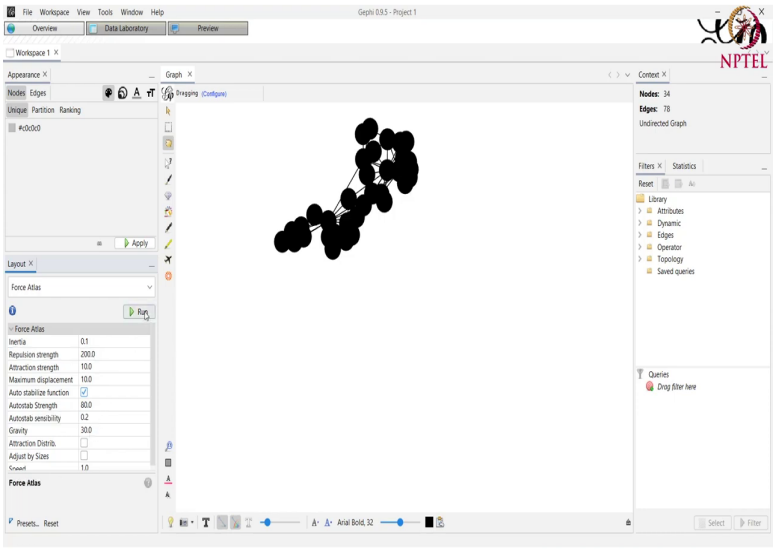

And selecting a layout out of all the different options of layout that we have. So, we select this Force Atlas layout. So, in this layout basically each of the node and the edges they are governed by some physics property that we define here. That is the repulsion strength between 2 nodes, should be around 200 that is the strength that the 2 nodes repel each other or are visualized away from each other on the on our screen right. So, and similarly we have the attraction strength and other different at different properties of the nodes. Now, based on this layout we run this layout on our graph.

(Refer Slide Time: 06:32)

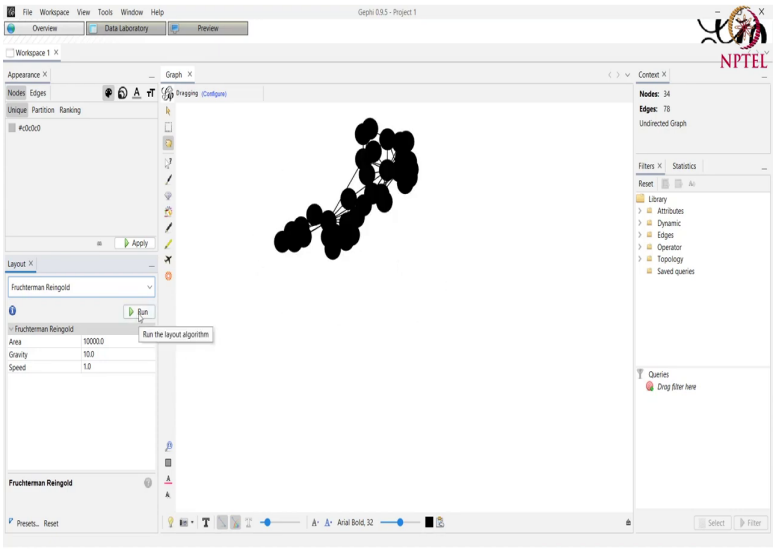



And we see that now our graph looks something like this, but it somehow became even difficult to understand correct. So, we might want to increase this repulsion strength, because we can see that these nodes they are very much close together and we want it to be spread apart. So, that we can we are able to visualize the structure of the graph.

(Refer Slide Time: 07:03)

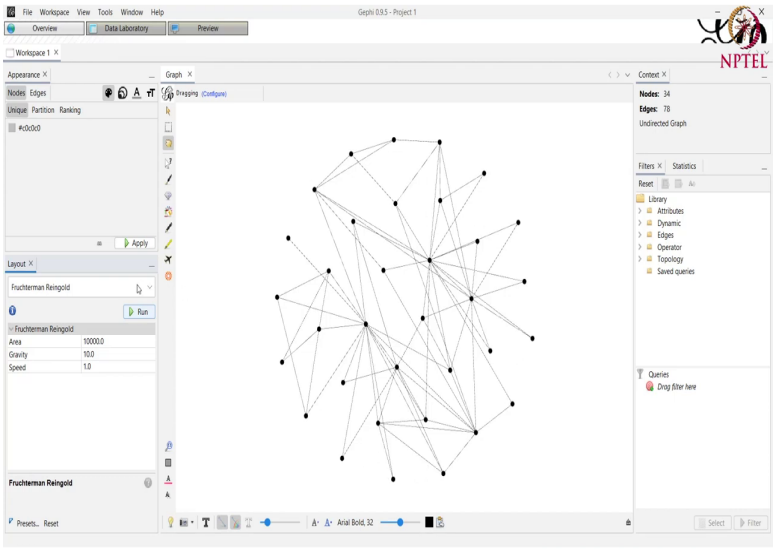



(Refer Slide Time: 07:27)

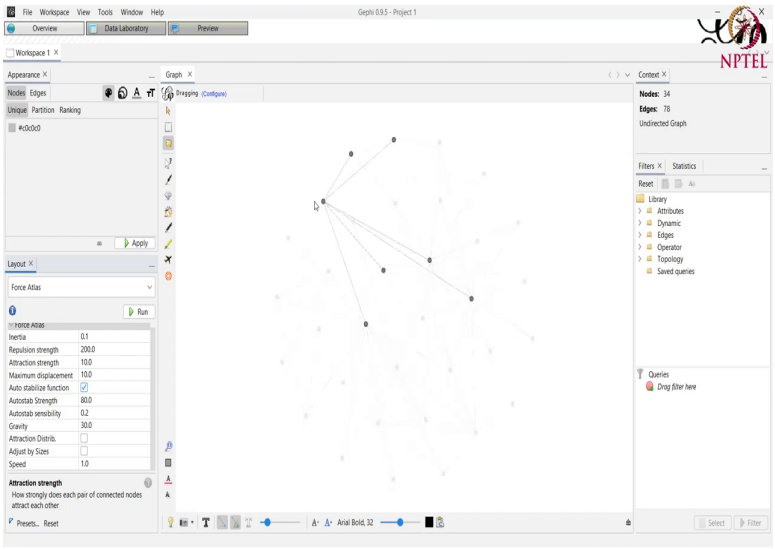



So, we just we will just increase this repulsion strength by maybe a factor of you know 10. So, we just here we increase it. Or maybe we can change the layout altogether. We will do the, we can use this layout that is Fruchterman Reingold, which will basically change this layout into a circular fashion. So, we run this, we reset our graph. And we can see here that the layout that we have now is basically in a structure in a circular manner where each node is like it is a some distance apart.

(Refer Slide Time: 07:48)

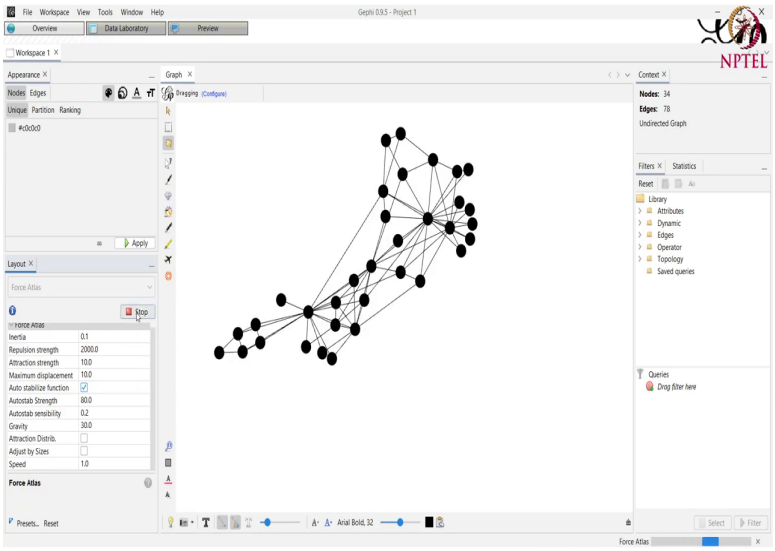



(Refer Slide Time: 08:01)




Right, now, another thing that we can do here is that we can rank these nodes somehow. So, before going into the ranking function of it, as you can see when I hover my mouse over a particular node. The one half neighbor of that node are highlighted right. And I can also click and drag this node to wherever I want, right like this.

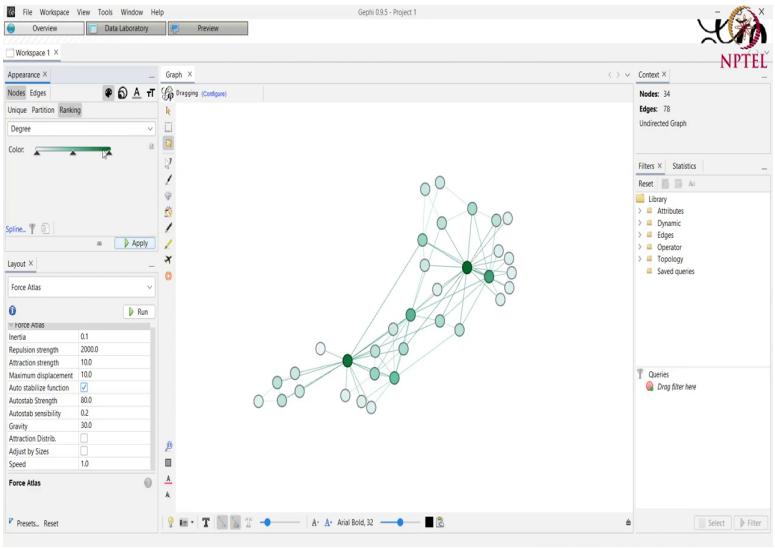
(Refer Slide Time: 08:39)




And I again this run this force atlas maybe and we can see that now the graph looks something like this. We will change this value of repulsion strength and then run it again. And we can see that now the graph like the all the nodes they are spread apart a bit more and now the graph looks something like this right.

Now, another thing here that we can do in Gephi is called Ranking, that is. We can change the color and the size of the nodes and edges, based on the different properties of the nodes and edges. So, we go in this Ranking tab here at the top left corner of our screen.

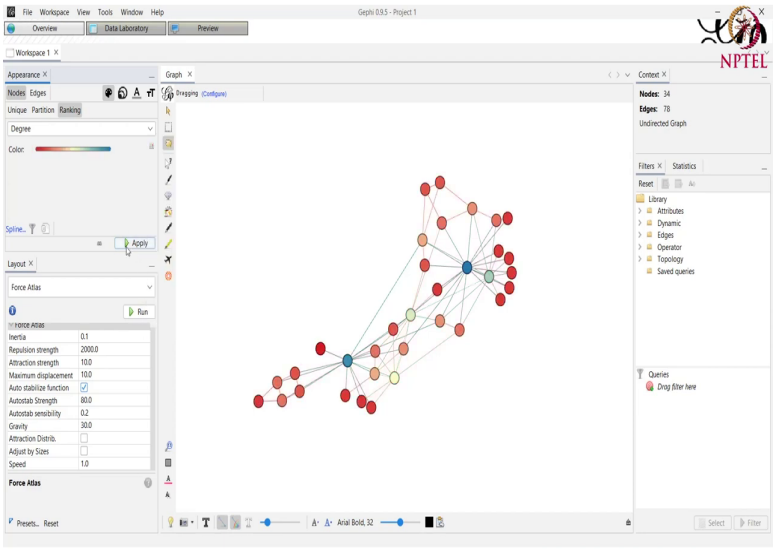
(Refer Slide Time: 09:29)




And, we select an attribute that we want to consider for a Ranking. So, here we select the degree attribute, right. And then we are shown some the you know the scale of colors or the palette of colors that we want.

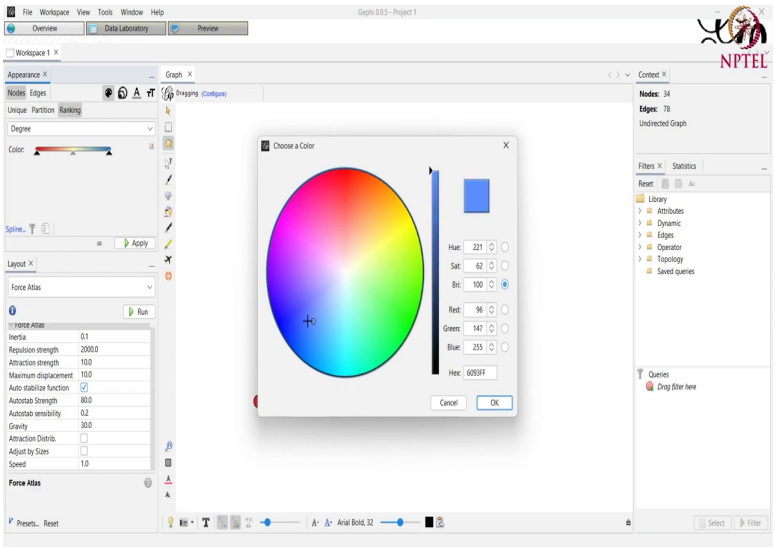
So, a lower degree node would be colored closer to White whereas, a higher degree node would be colored dark Green. So, we apply this ranking to our graph and we see that the higher degree nodes they are colored on the darker side of the Green whereas, the lower degree nodes they are colored on the Whiter side right. We can also change this color palette by like either of the two ways.

(Refer Slide Time: 10:18)

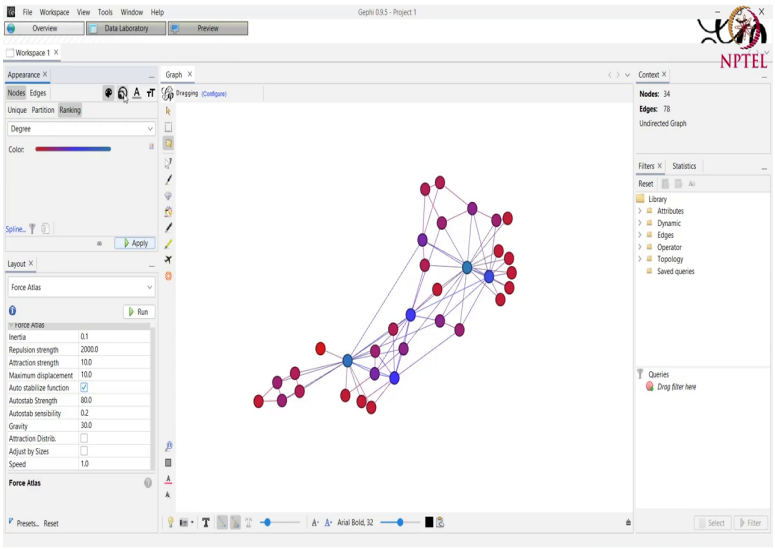



We can either select a default color palette by going here. For example, let us select this and apply we have this higher degree node with value Blue the middle ones with White and the lower ones with Red or we can select a particular this triangle that is present here. We can select this triangle and choose from any of these colors. So, for example, we might want this to be Blue, we select Blue.

(Refer Slide Time: 10:37)

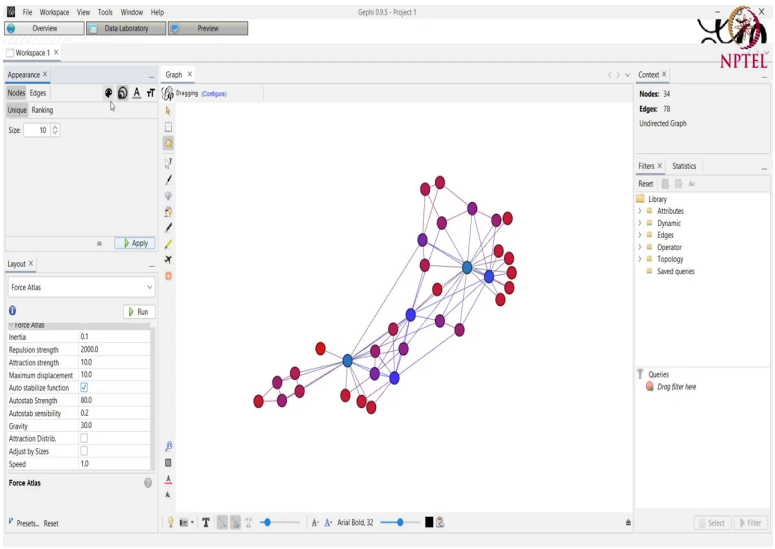



(Refer Slide Time: 10:46)

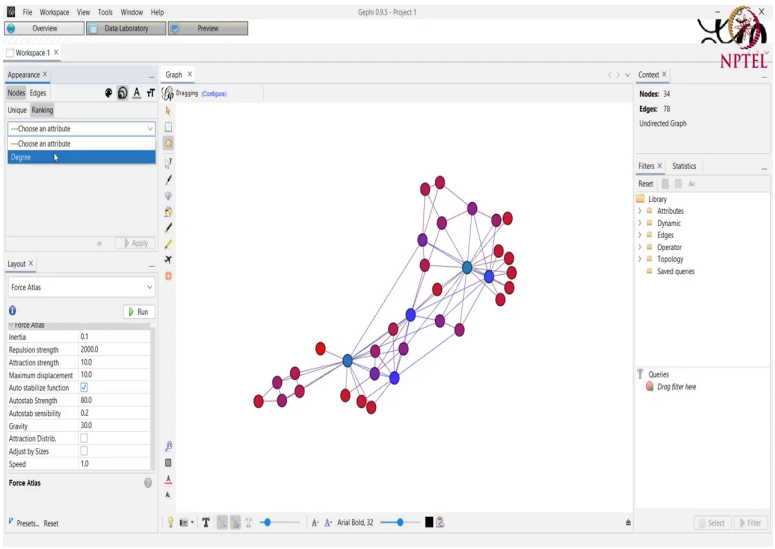



And we apply and we can see that the corresponding degree values they have changed to the said color. Then apart from the just the color of the nodes, we can also modify the sizes of the node just like if you remember we did in the last tutorials with the network x draw functions here also we can modify the size of the node, based on the on the value of different attributes.

(Refer Slide Time: 11:12)

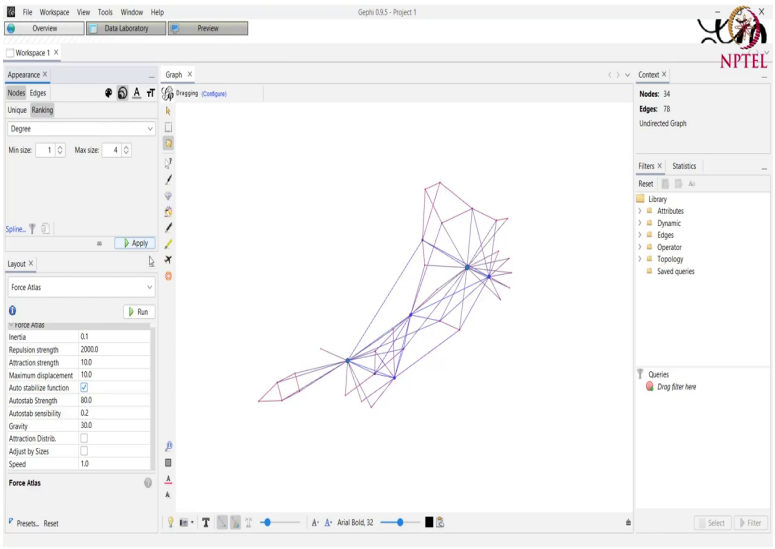



(Refer Slide Time: 11:18)

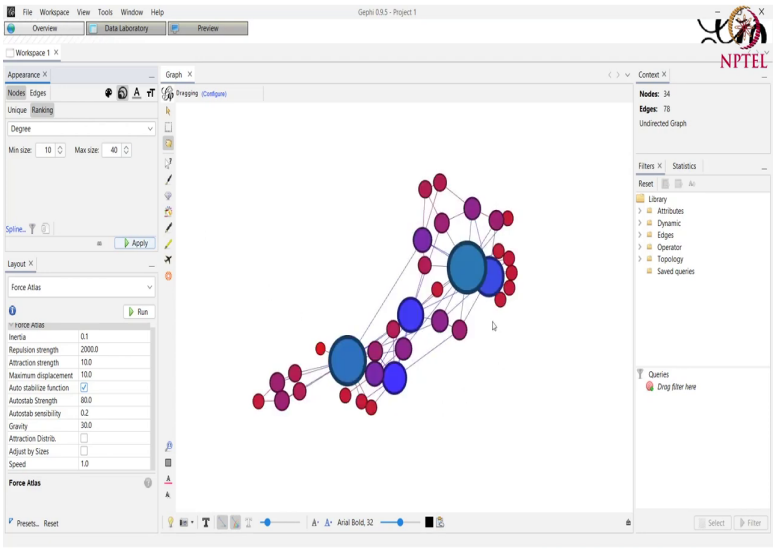



So, we were here in the Ranking tab of the color of the color property of the node. We will just go to one right hand side and select the size attribute of the node and again go to the Ranking module and then we will select an attribute by which we want to change the size of the node that is degree here.

(Refer Slide Time: 11:43)



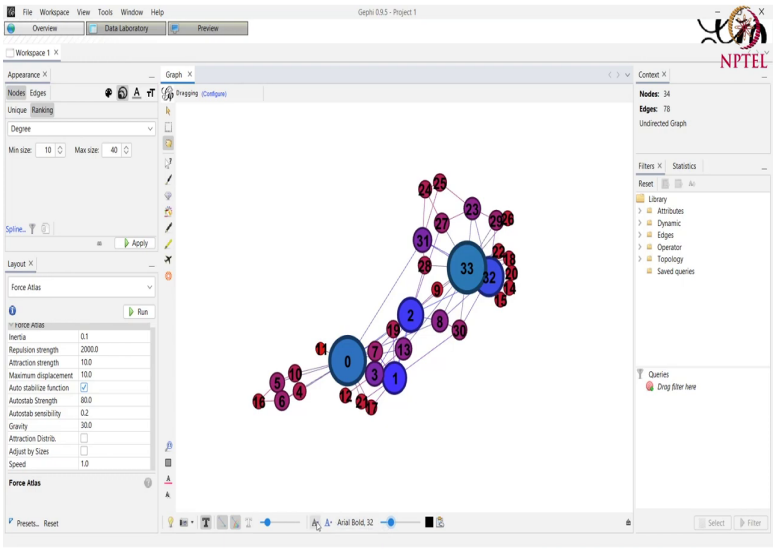

(Refer Slide Time: 11:54)



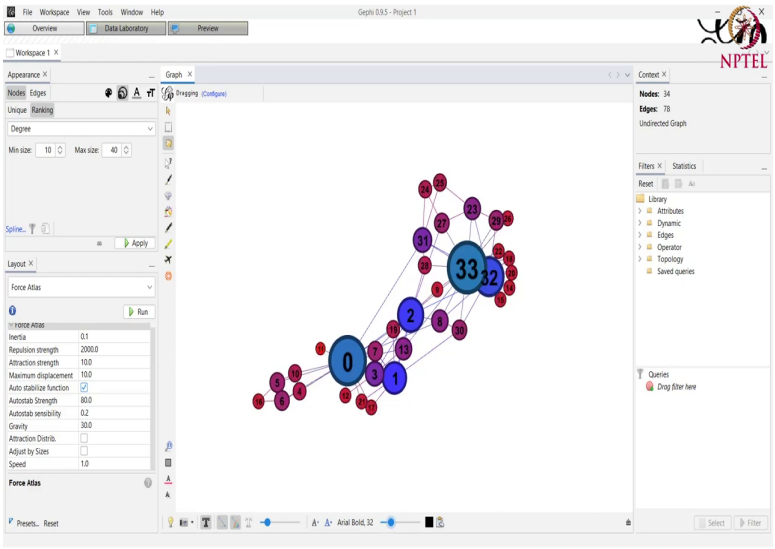

And we select the minimum and maximum size. So, right now the minimum maximum size are small. So, we can see that by applying it and we see that the size is quite small. So, we just increase the size by maybe a value of 10. So, now, the minimum size is 10 and the maximum is 40. We apply this and we see that the higher the degree of a node the larger that node appears in our graph.

Now, again these nodes they can also like we can do one more thing that is we can show the labels or the ids of the nodes on them right. So, in order to do that we just click here. That is show node labels and these ids of the node that they are shown here right. We can change the size of these labels or an interesting thing that we can do, is that we can change the size of each node, based on the size of the node. That is the label size would be bigger for a node that has that is bigger in size.

(Refer Slide Time: 12:25)

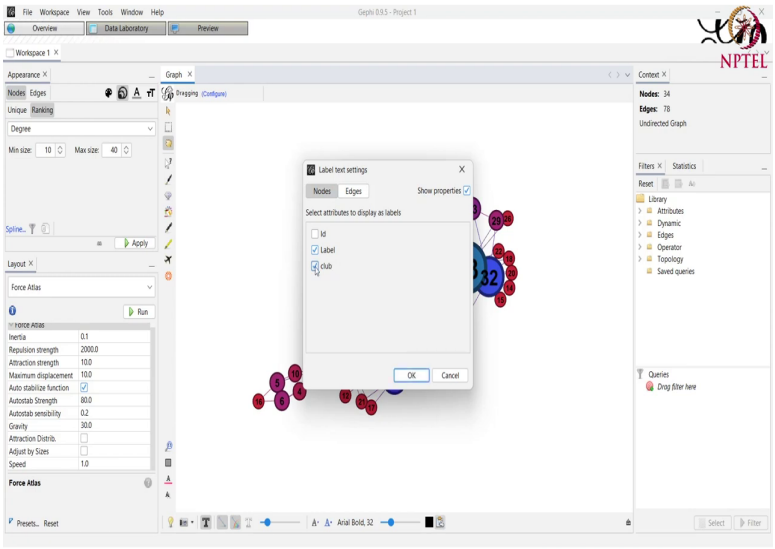



(Refer Slide Time: 12:40)



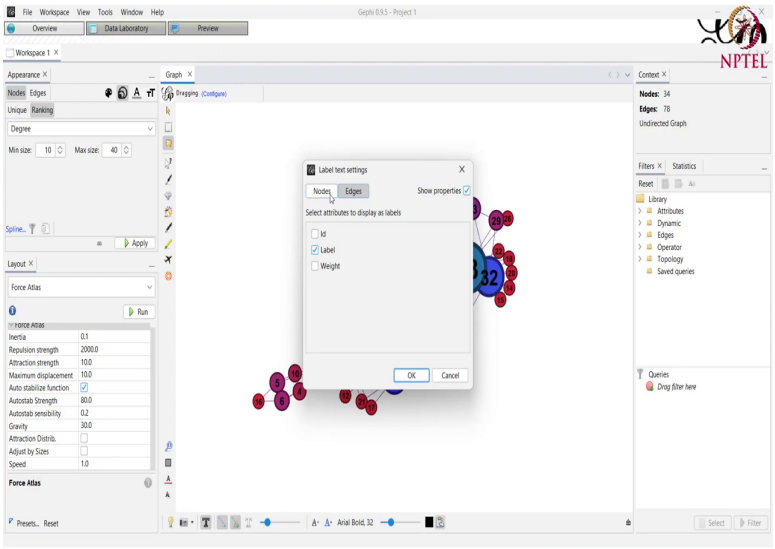

So, we change this and we see that like 0, 33, 33, 32 they are coming to be big with big node labels. Whereas, the smaller nodes smaller degree node that is 11, 16, 26 they are coming to be smaller with smaller node labels. We can also select what kind of labels we want to show here by going to this option here.

(Refer Slide Time: 13:19)

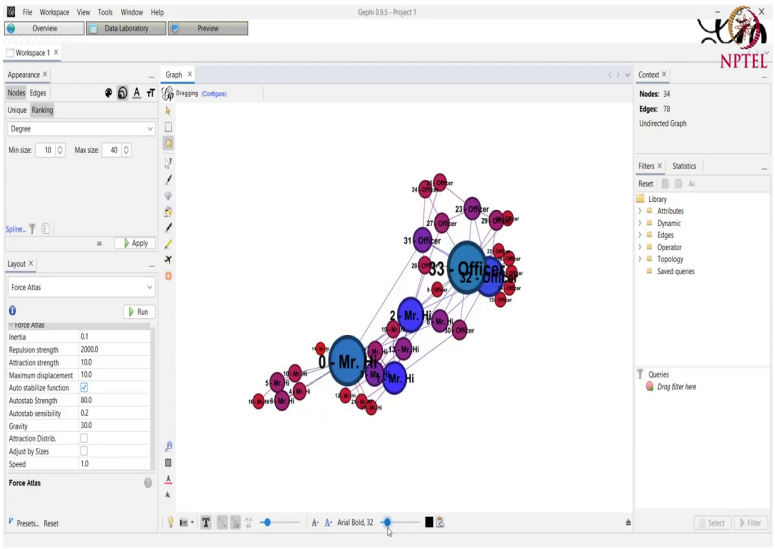



So, each node it basically has three labels right now. That is an i d of the node, the label of the node that is already shown and the club to which it belong. So, here in the karate club graph, basically each node belongs to one of the two club that is present in the network. So, probably we also want to see the club to which the node belongs, we select that or we want to see the club here.

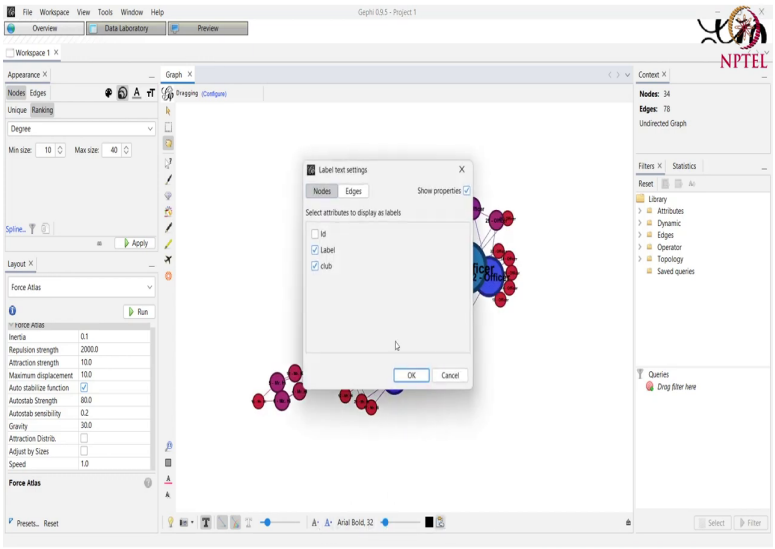
(Refer Slide Time: 13:49)



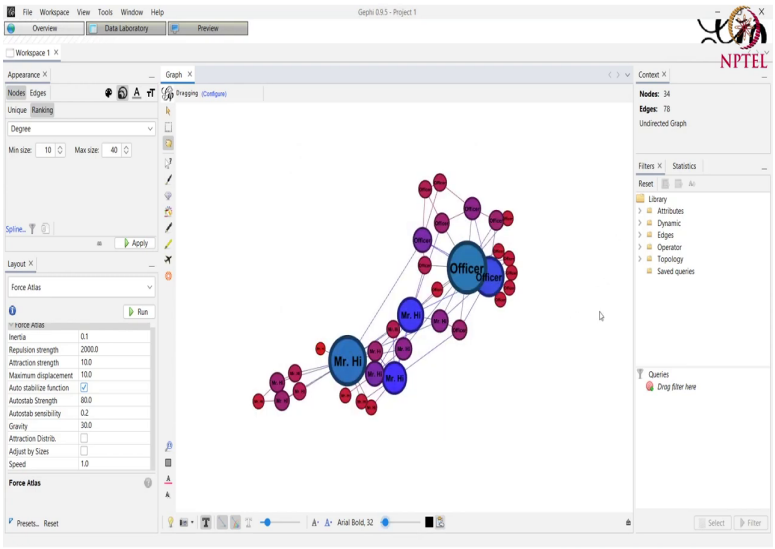
(Refer Slide Time: 14:00)




(Refer Slide Time: 14:09)




(Refer Slide Time: 14:11)

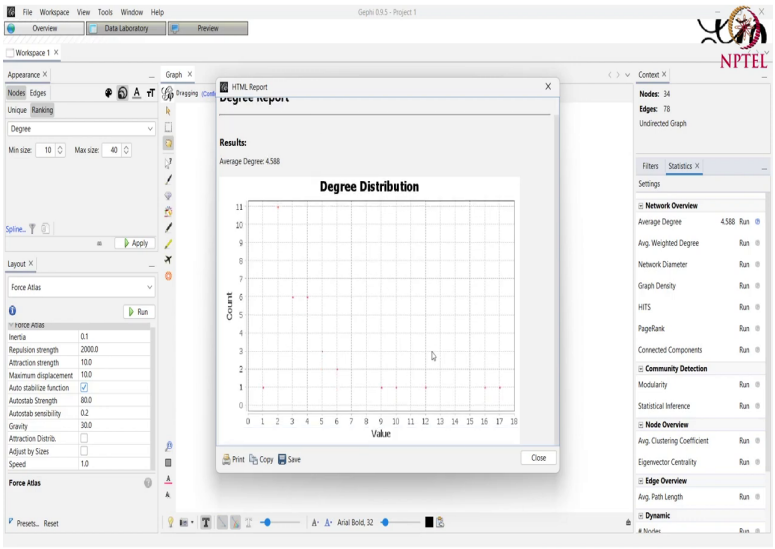



We can also change the different attribute that you want to be shown for the edges also. But right now let us just stick to the nodes. So, we select the value of club also and we click on ok. And for each club for each node we are also shown the club. So, we might want the label to not be there in just the club we press ok.

And then we can see that these two higher degree nodes, they belong to two different clubs. This one for officer and we can see that most of its neighbor also belongs to the club officer. This one to mister hi and we can see that most of the neighbors of this node also belongs to mister hi with just one belonging to officer right.

Then the other thing other very interesting thing that can be done in Gephi software is to analyze the graph in different fashion. So, in order to do that we just go to the statistics tab at the right hand side of our screen.

(Refer Slide Time: 14:58)



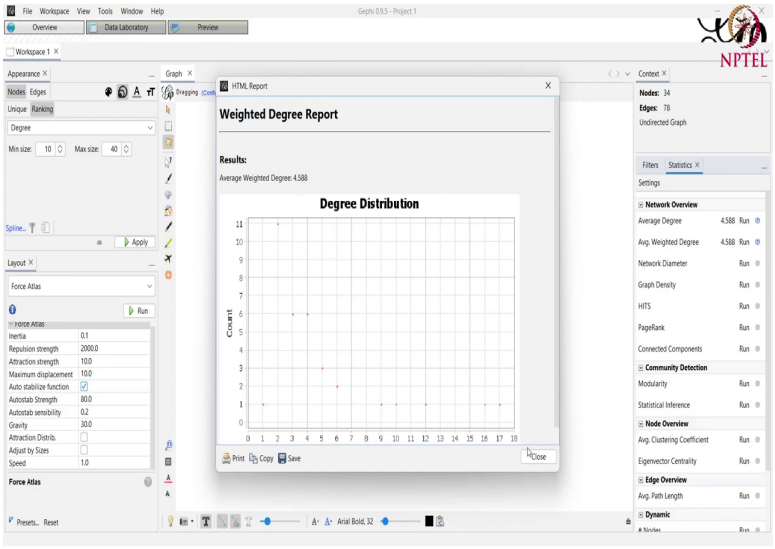

We see here the tab statistics we go here and we can see that Gephi provides us with a lot of possible statistics that can be counted for the graph.

So, first thing is just to calculate the average degree of the of all the nodes that are present in the graph into calculate this we need to run this algorithm here. And we are shown with a degree distribution. So, the average degree is coming out to be 4.588 and this is the degree distribution that we have. That is the degree 0, comes sorry the degree 1 comes for one node only.

Then the degree 2, comes for 11 nodes. Then 3 and 4 both comes from 6 number of nodes. Then 5 comes from 3 and so on. We get this degree distribution as well as the average degree of the of our network.

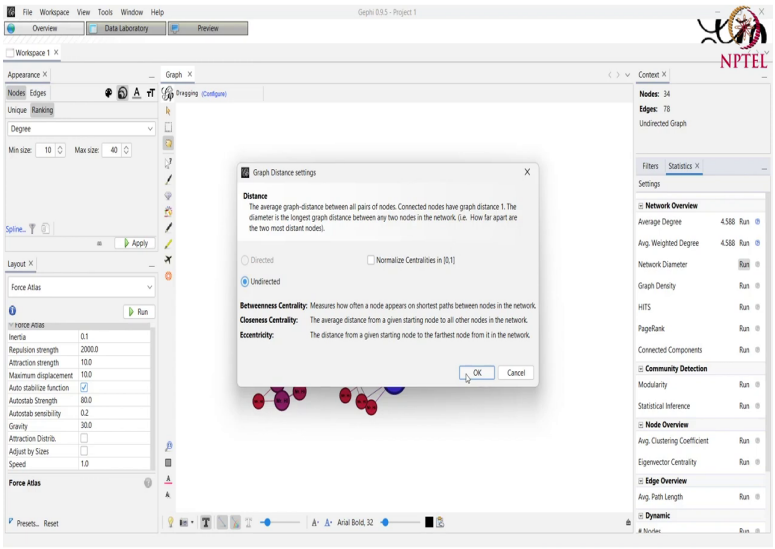
We can save this report, we can copy this report or we can print this report, we close this report right now and here we can see that this algorithm is run. We can also calculate the average weighted degree, but since this is an undirected unweighted graph. The weight for each edge and each node is one. So, the weighted degree is same to a as the average degree.

(Refer Slide Time: 16:16)




Then we can also calculate the network diameter. Now, basically the diameter as already discussed it is the longest shortest path. So, to in order to calculate the diameter we need to visit the distance between the path between all the nodes that are present. So, while we are doing that while we are visiting the path between all possible pair of nodes. So, Gephi while doing that also calculates different centrality measure that is the Betweenness Centrality, Closeness Centrality and Eccentricity.

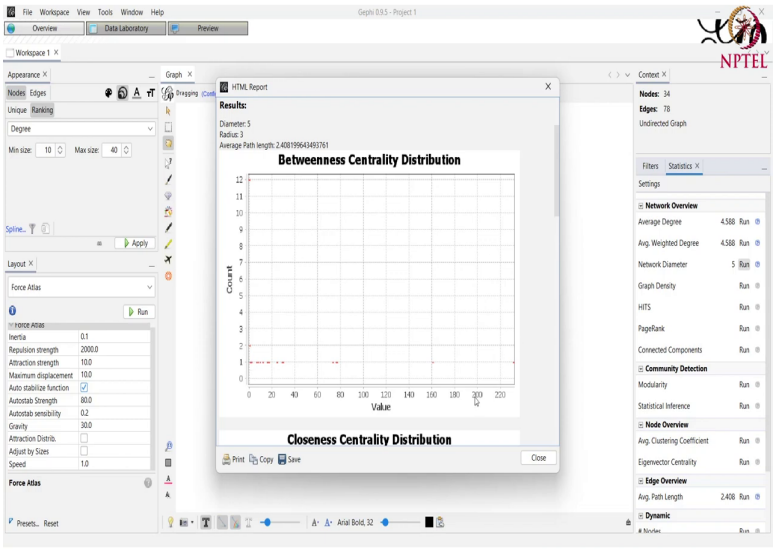
(Refer Slide Time: 16:54)




Since it is already visiting the path between each node. So, there are just some of the like some of the options that we can provide that is either we want to normalize centrality between 0 and 1 and 1 and the graph is directed or undirected. So, for a directed graph, we can select the option of undirected here where we consider the graph as undirected for the calculation of the diameter.

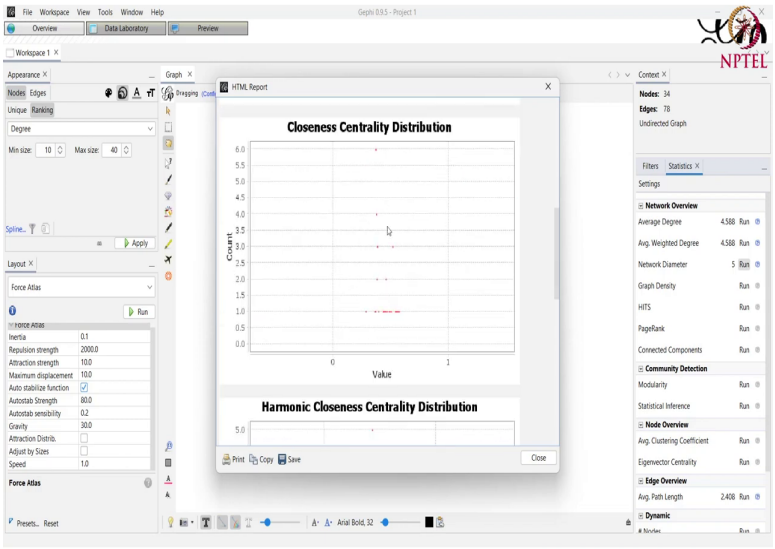
But since we have the undirected graph we just have this option available for us. We press ok.

(Refer Slide Time: 17:37)

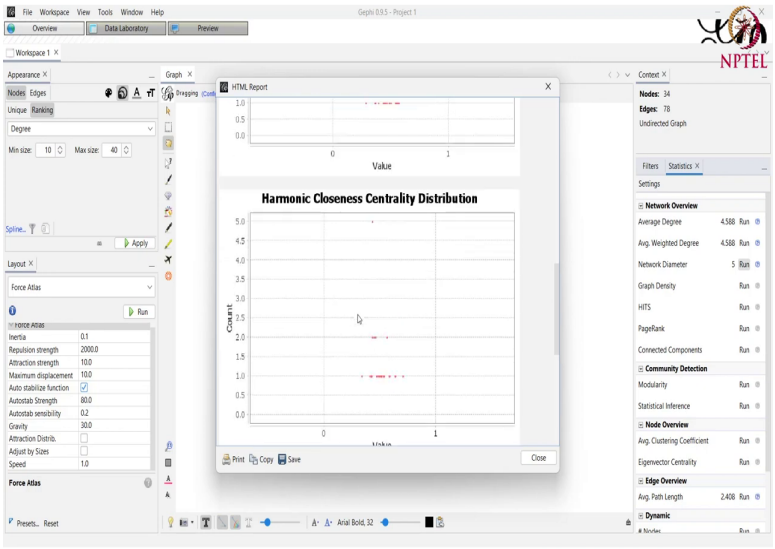



And we see that the diameter of the of our network is 5 that is the longest shortest distance is 5. And the betweenness centrality distribution is coming out to be something like this that is the value from 0 to maybe like 20 or 30 is coming out of or 1 and so on right.

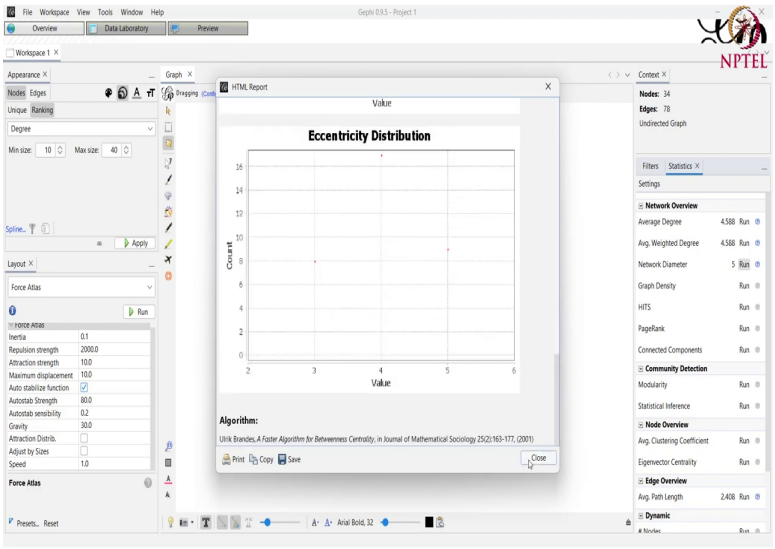
(Refer Slide Time: 18:03)




(Refer Slide Time: 18:05)



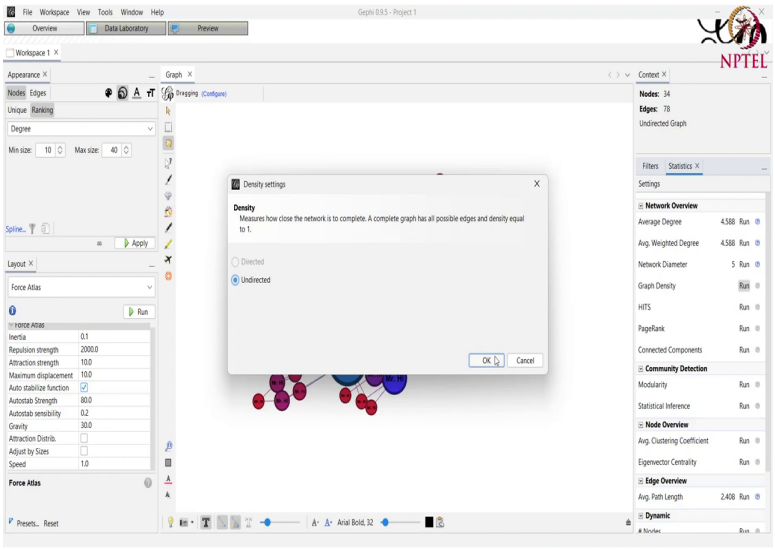

(Refer Slide Time: 18:07)



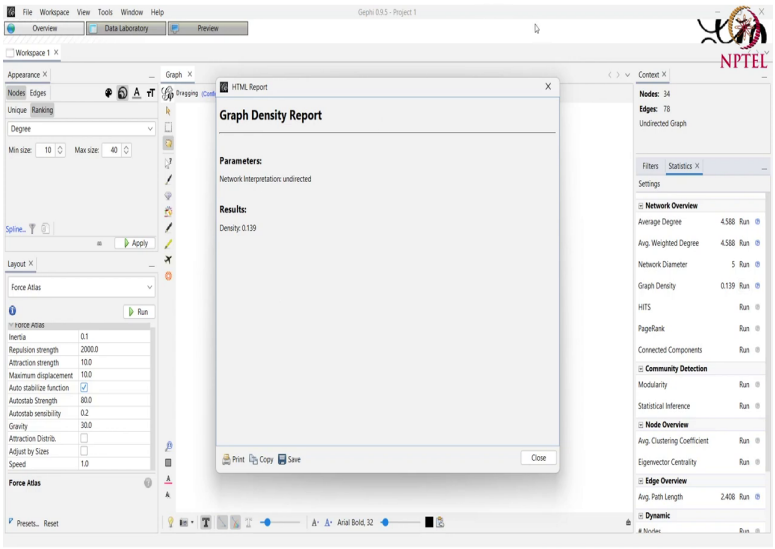

Then we have this closeness centrality distribution. And the harmonic closeness centrality distribution. And the eccentricity distribution. Sorry, we close it and apart from diameter and the weighted degree and the average weighted degree we can also calculate the density and the other importance measure like hits or page rank.

So, density is again calculated using this density algorithm by running this algorithm since it is an undirected graph we can select this undirected option here. And we see that the density that is the number of actual edges that are present in the network divided by the possible number of edges that can be present in the network.

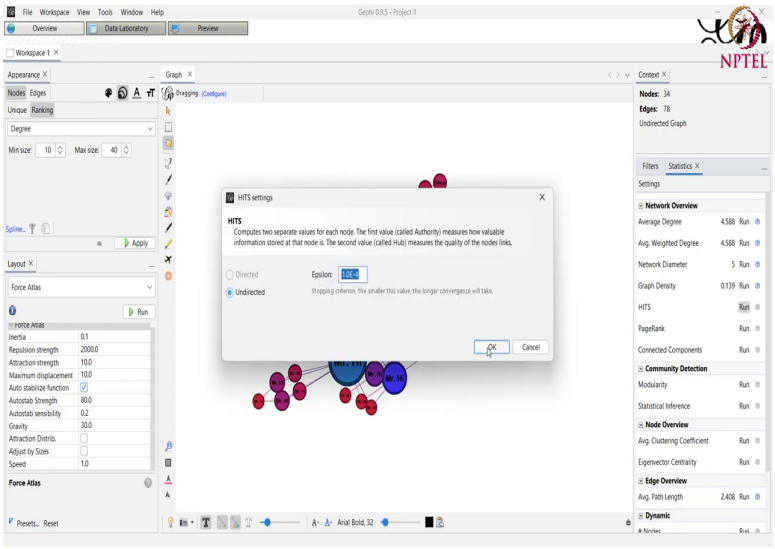
(Refer Slide Time: 18:31)




(Refer Slide Time: 18:36)

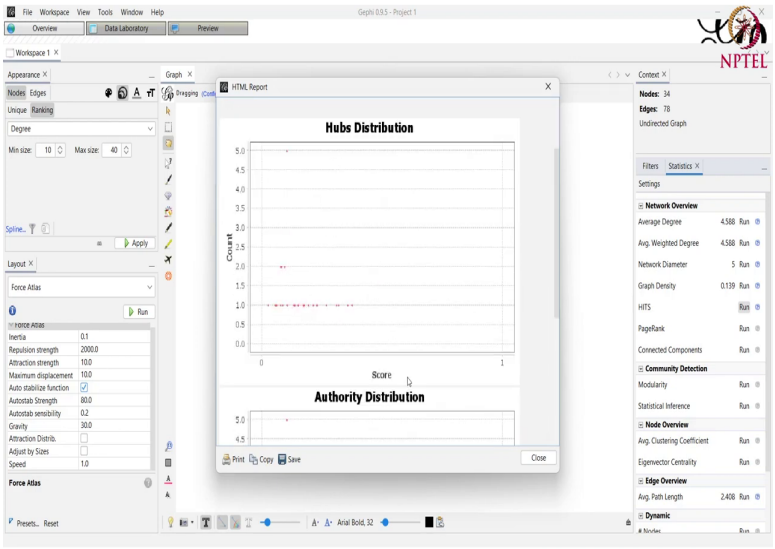



(Refer Slide Time: 19:03)

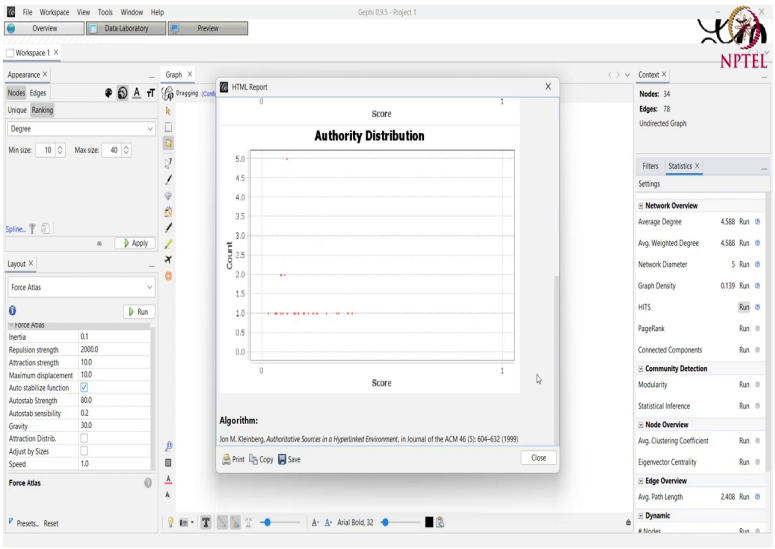



So, that value the density is coming out to be 0.139. We can also calculate the hits value that is the hub and the authority scores of the of each of the nodes of our value of our graph. And we see that this is basically the distribution of the hubs and the authority value.

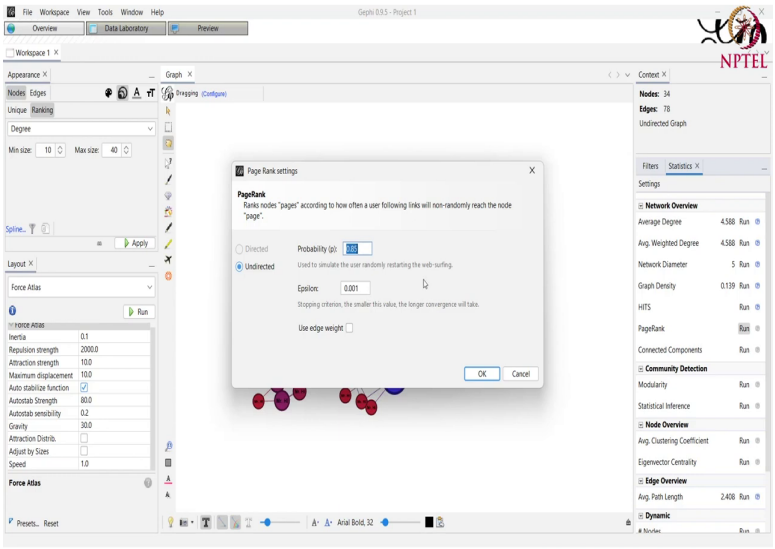
(Refer Slide Time: 19:06)




(Refer Slide Time: 19:11)

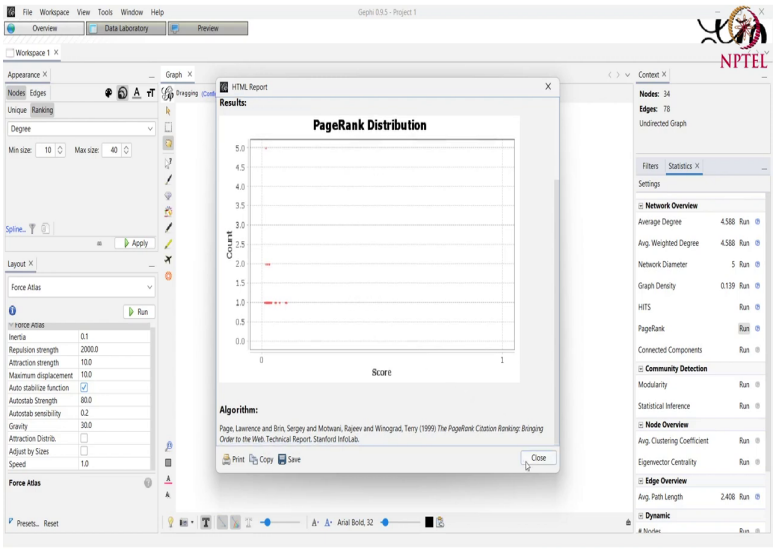



(Refer Slide Time: 19:28)



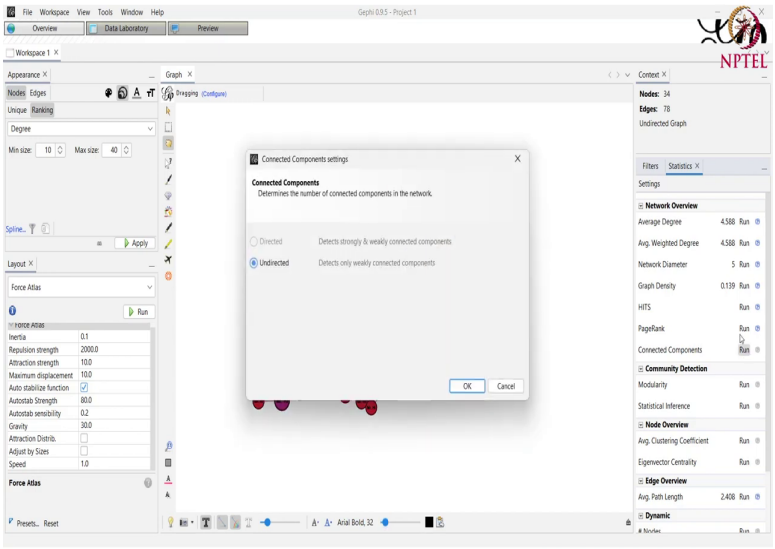

So, this basically tells us the importance of a particular node. The how authoritative that node is or how much of a hub that node is based on this algorithm. We can also calculate the page rank for each node. By simply using this by providing the different attributes of the paging algorithm that is the probability of random work and the epsilon value we pass.

(Refer Slide Time: 19:37)

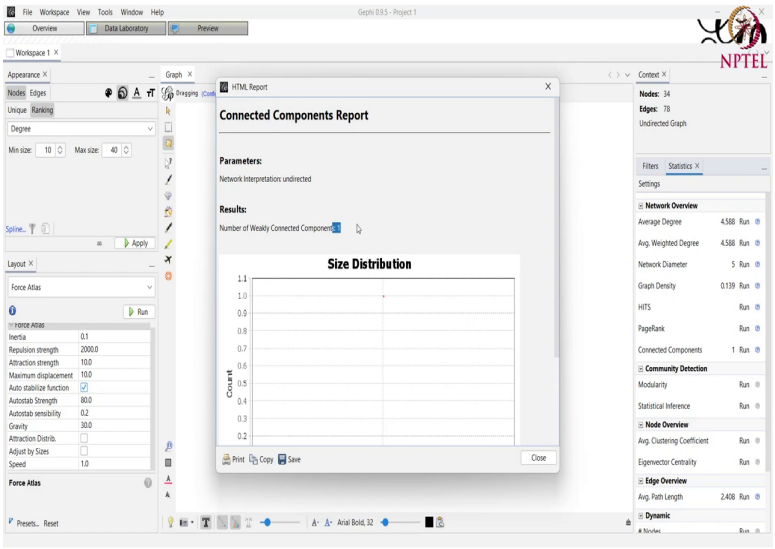



We select ok and we see that this is basically the distribution of the page rank value, that we are getting for all the nodes for this network. Then we can also see how many number of connected components and they are there are present in our graph by simply running this algorithm.

(Refer Slide Time: 19:55)

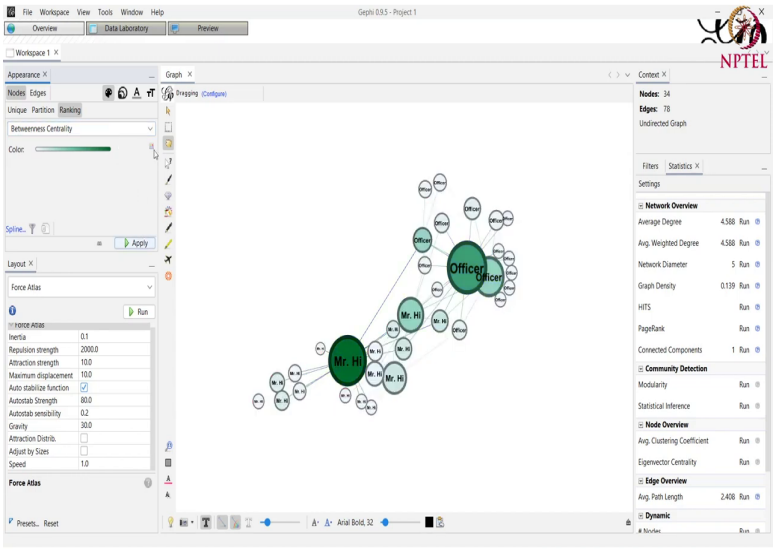



(Refer Slide Time: 19:57)




Selecting the undirected option and we see since all the nodes are reachable by all the other nodes that is there is just a single connected component present here. Now, we calculated these metrics, now we can also rank our nodes based on these metrics instead of the degree that we used. So, again we go to this color tab here, go to this ranking tab and now earlier we were just getting the option of degree, but now if we just click on this drop down box, we are getting the option a full we are getting a lot of options, because we already calculated a lot of metric by running these different algorithms that were present.

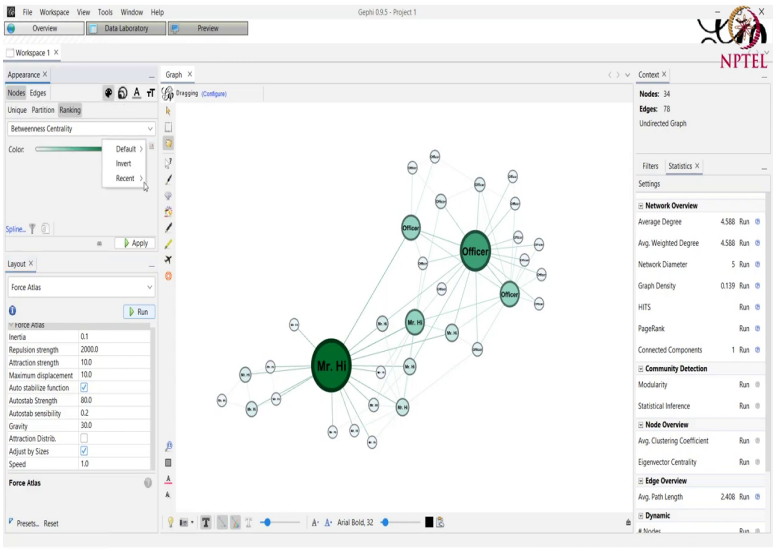
(Refer Slide Time: 20:33)



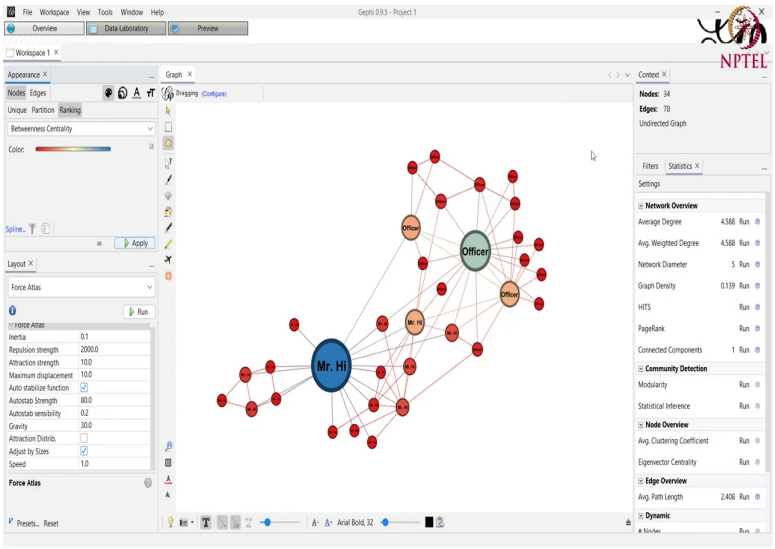

So, now we can we can basically modify the visualization of our graph based on these attributes. For example, we select betweenness centrality here we apply it and the color changes based on the value of betweenness centrality and the size also we might want to change for the betweenness centrality here. Again the let us select a larger scale for our you know graph sizes. And we can see that the betweenness centrality is higher for mister hi and lower for officer. All the degree was higher for officer right.

And, now what we can do. So, right now we can see that these two nodes are probably overlapping there are some other overlapping nodes as well. So, we can adjust now when we have modify the sizes we can go to this layout tab. And we can adjust the layout by the sizes of the node. So, we just enable this option we run it again. And we see that the graph is more spread out and it is more easily visible. Now, let us just change this color palette a bit.

(Refer Slide Time: 21:39)



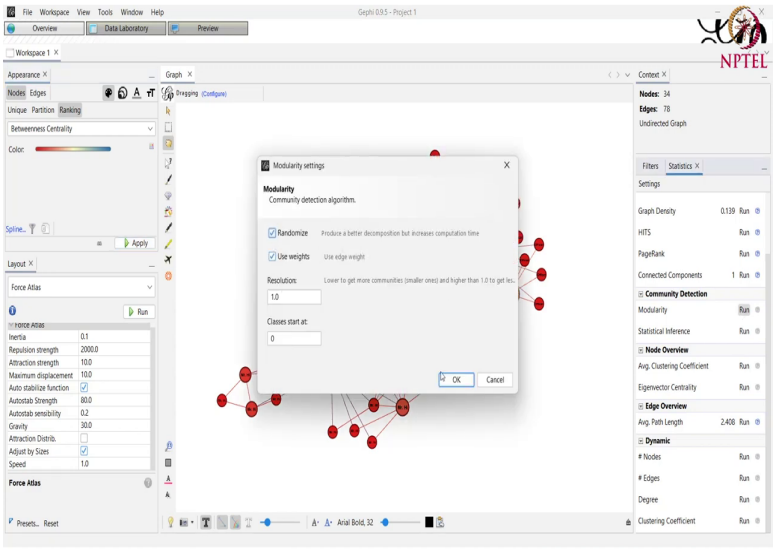

(Refer Slide Time: 21:55)



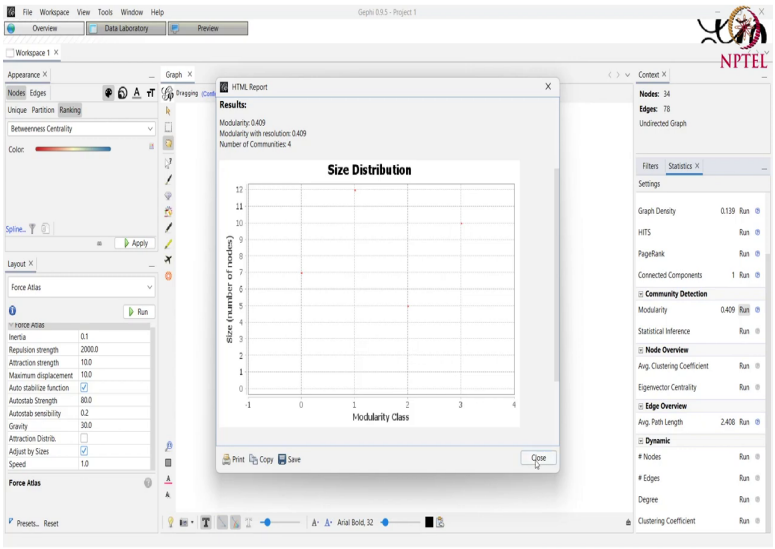

So, that it is more easier to see, yeah. So, now, apart from this ranking and these different metrics that can be created we can do one thing we can perform community detection here. So, what is community detection. So, basically any community detection mechanism or algorithm, it tries to find different groups in the network which are highly connected in like the nodes inside a group is highly connected, when compared to the nodes between two groups right.

So, we can perform community detection here by calculating the modularity values. So, modularity is again a metric. We will not be going into detail of the modularity and community detection here, but modularity is basically a value which tells us how interconnected a community is and how less connected a community is to the other communities. So, we run this modularity algorithm here and we just select some of the option that are present to us.

(Refer Slide Time: 23:14)

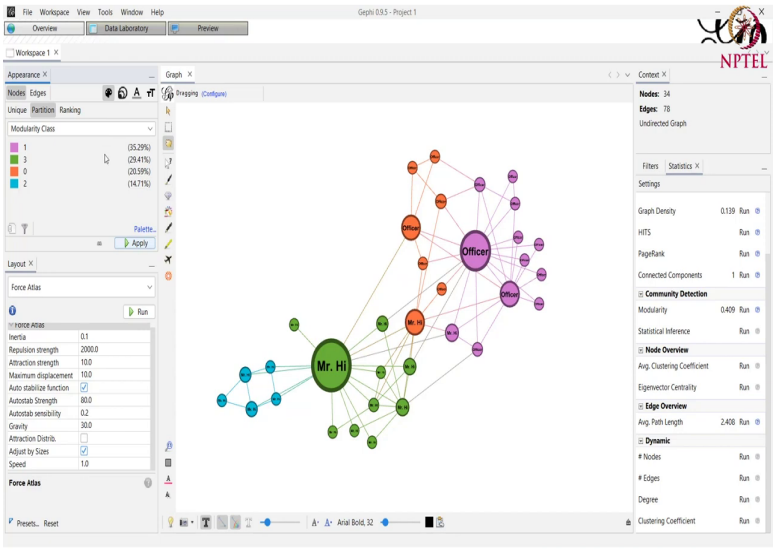



(Refer Slide Time: 23:18)




We select ok and we are basically we get basically all the modularity values associated with the different nodes in our network. We close it and now we can use this modularity value to perform a partition that is to perform the community detection mechanism. To do that we go to this left hand side tab that we have and go into this partition tab. Now, we select like we have different attributes here, according to which we can have different partitions.

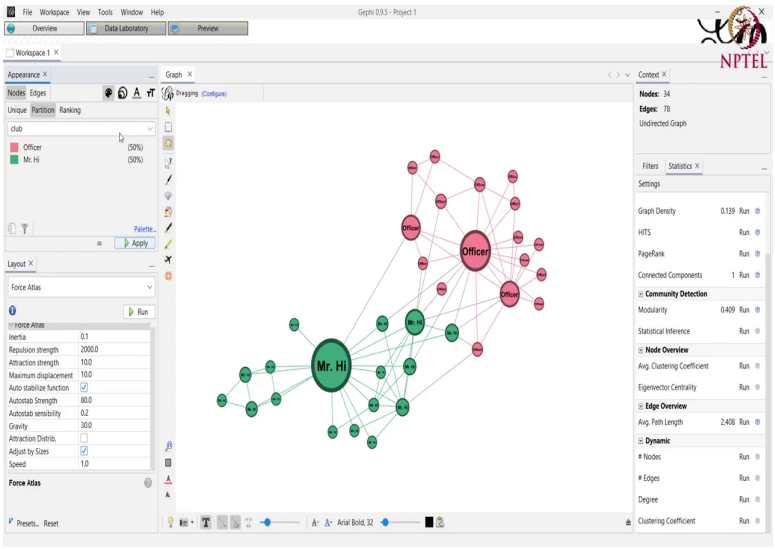
(Refer Slide Time: 24:00)




But we select this modularity class value. And we see that this algorithm it identifies four different communities right. And each community is assigned with some color. This is like the default color that is present here. We run this we just apply this partition value. And we see that these are the four communities that are identified by the modularity value. If you consider the modularity value that is this is one community this blue nodes these green nodes are one community then we have one with pink and one with orange. So, these are the four communities.

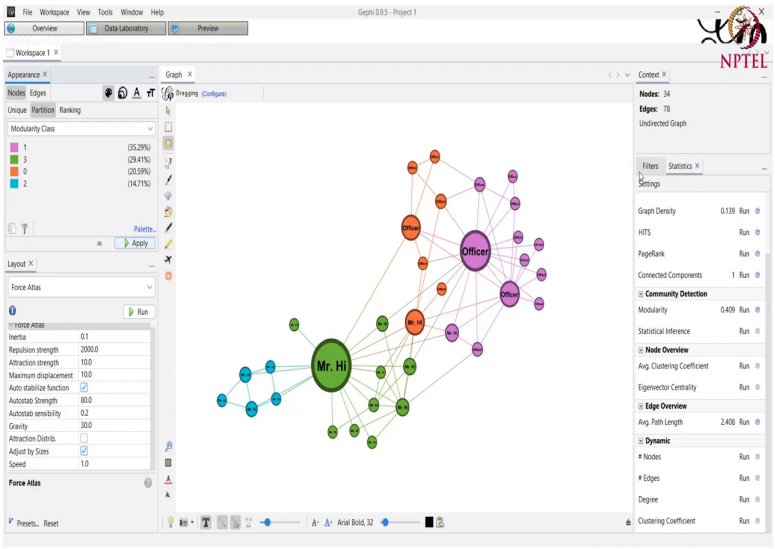
Now, a community can be like it can be of different things. So, for example, here in this network the most basic community that we can think of are the two communities of the two different professors, that is Mister Hi and Officer. So, in order to see these two ground truth communities.

(Refer Slide Time: 25:07)

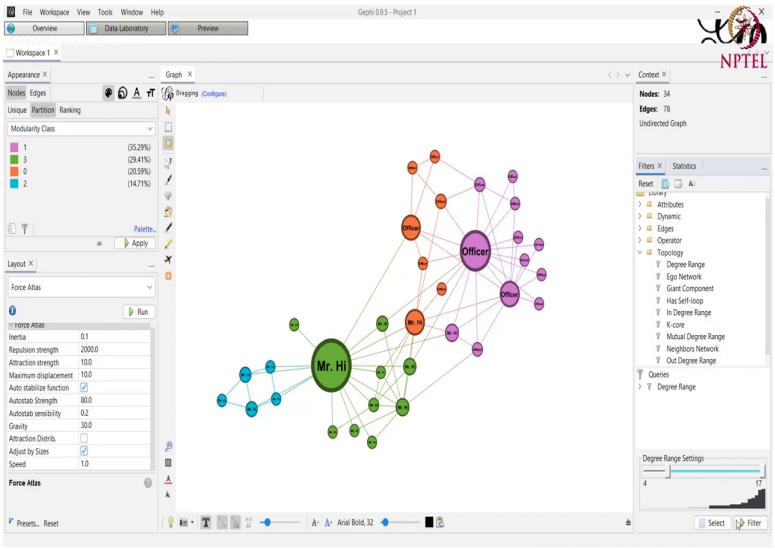



We can select like in the partition module only we can select the club as the attribute. So, the club is the already present node attribute, that we have in the ground truth we select this and we apply these coloring here and we see that now we have just two communities these are basically the ground truth communities, where these nodes belong the green nodes belong to the Mister Hi club and these nodes belong to the Officer club. So, whereas, this is the ground truth community detection.

(Refer Slide Time: 25:41)



(Refer Slide Time: 26:06)



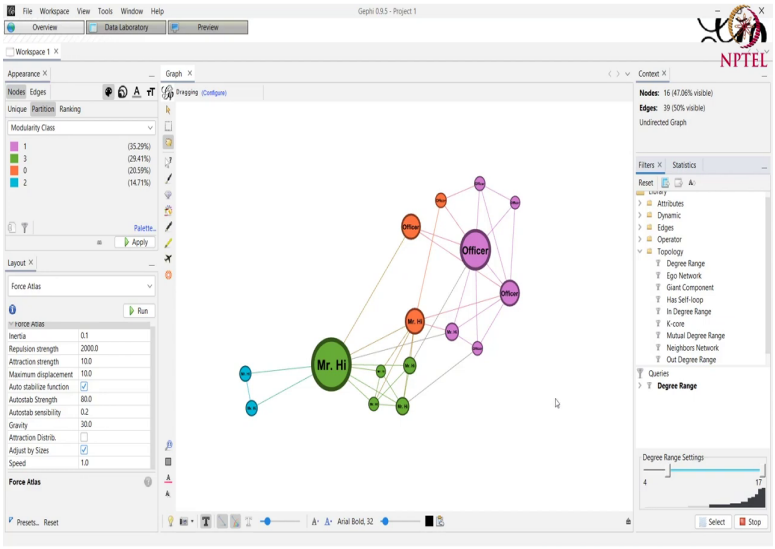

That is when we have the two communities, but if we select the modularity value we are getting four communities with this kind of a partition right. Now, apart from partitioning the nodes in this manner we can also perform some filtering.

Now, filtering can be done using this tab here. That we have on the right hand side. There are different like there are very many options to filter the graphs and the networks and the edges from we just show one example here where we go to the topology and use this degree range for as one of our filters.

So, we just click on this degree range and drag it to the query box here. Sorry the filter box here. And we select the range of the degrees that we want to be shown on our screen. So, suppose we want that all the node that have degree less than 4 to not be shown as our as a part

of our network. So, we select the lower value the lower bound of the degree as 4 and we click on this filter function.

(Refer Slide Time: 27:02)

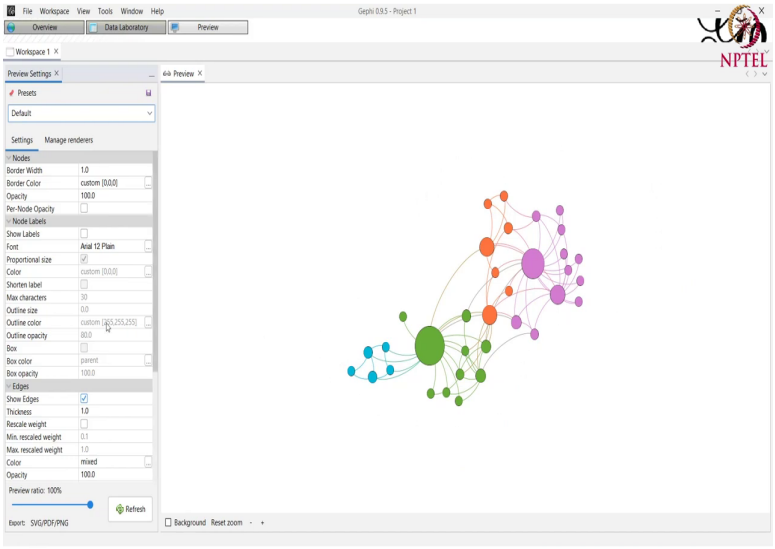



And we see and we see that all the nodes that have values less than 4 are now vanished from our screen. And we can only see the nodes that have a degree of greater than 4 in the original network. Not here this mister hi may have or this node may have a degree 2, but in the original network it had a degree 4 or greater than 4 right.

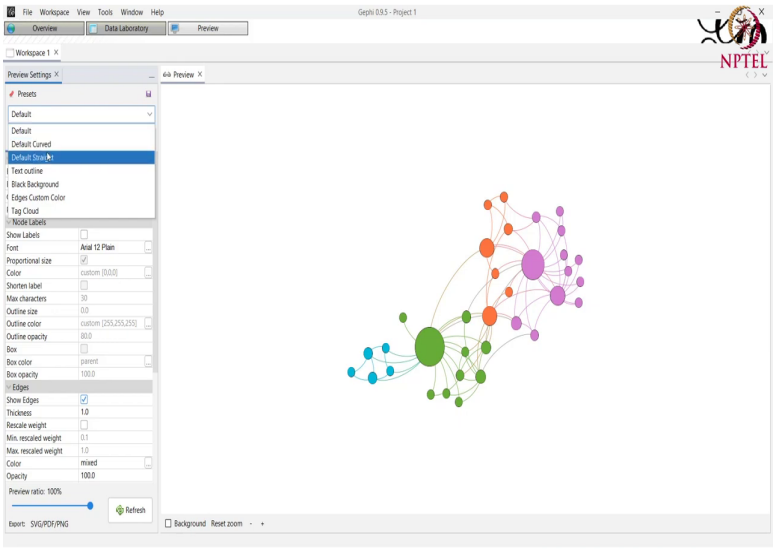
Now, apart from this partitioning, filtering and layout and ranking mechanism. The last thing that we might want to do with in Gephi is to now to after all the processing of the all the analyzing that we have done. We might want to save this graph in a way that we that the visualized network can be used in maybe our report or any other system. So, in order to do that we go to this preview tab on the top.

And we select the different attributes of how we want our network to look like when we save it. So, these default options we just refresh our system and we see that this is the kind of network that is coming right now. We might want it to be the lines to be straight. So, we select the default straight option here then again we refresh it.

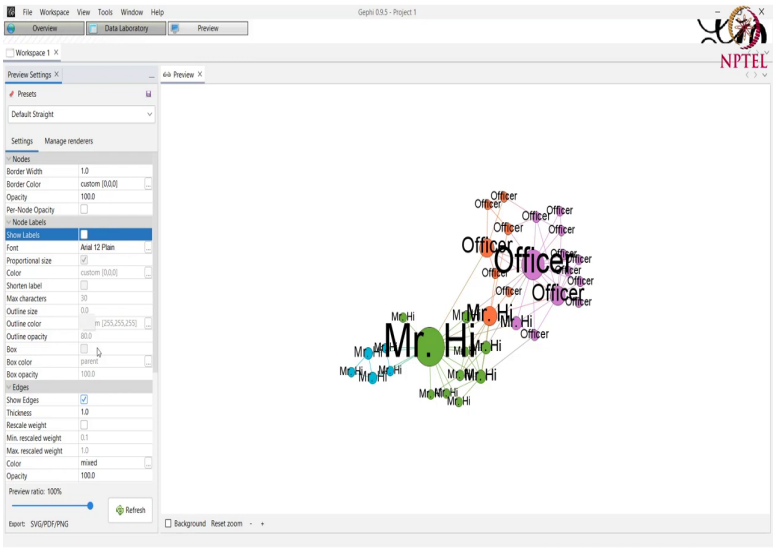
(Refer Slide Time: 28:15)




(Refer Slide Time: 28:35)

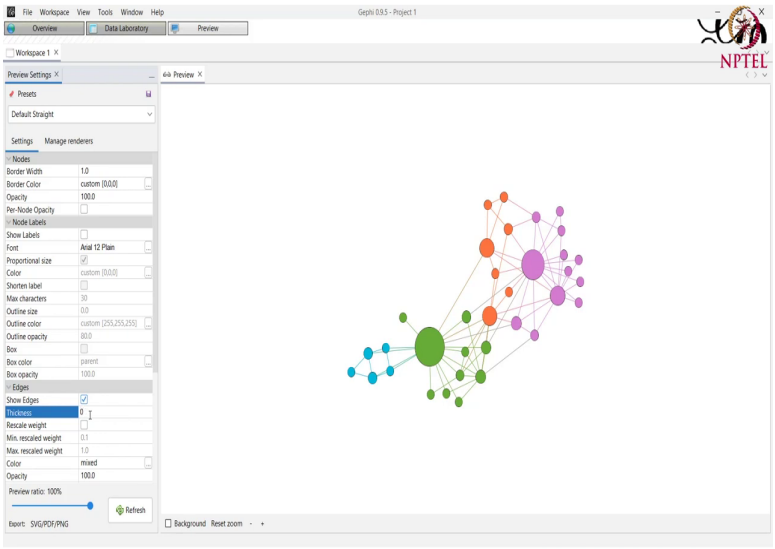



(Refer Slide Time: 28:43)

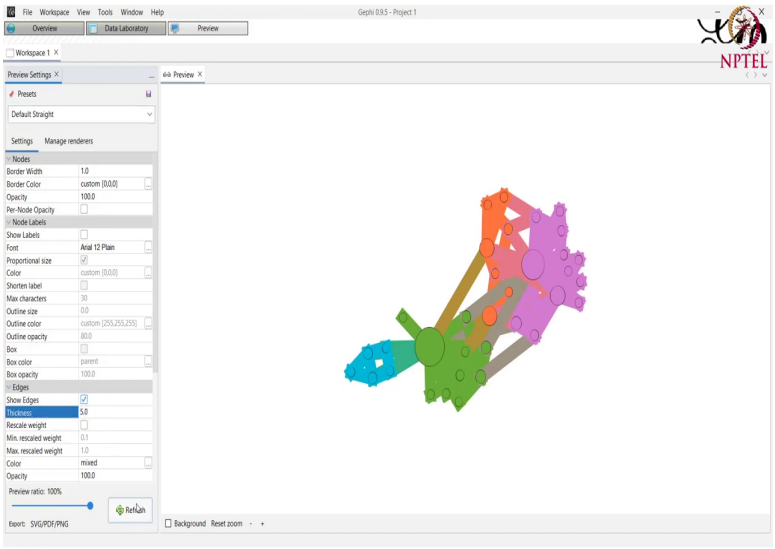



And we can see that the edges are straight and we also have the labels here. We might not want to have these labels we can select we can like disable this option of show labels in this node label option here. Then again refresh it and we do not have the labels anymore. We there might be a case where we want to increase the thickness of the edges.

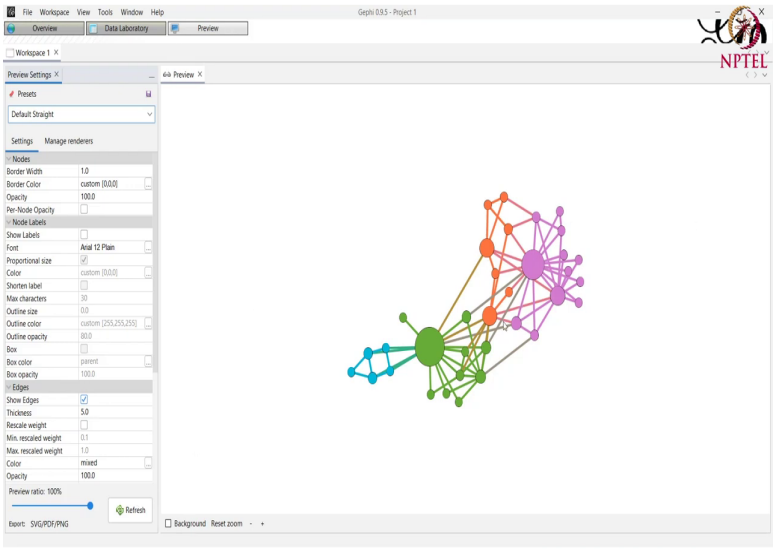
(Refer Slide Time: 29:01)



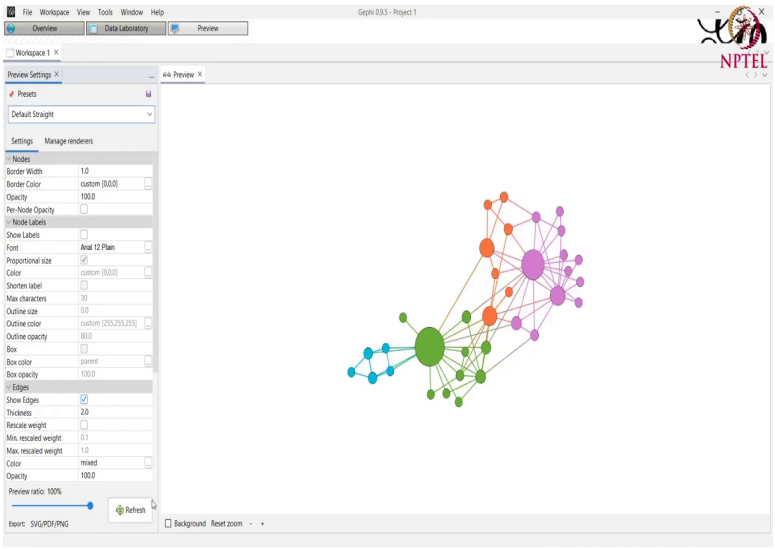

(Refer Slide Time: 29:09)


File Workspace View Tools Window Help
Gephi 0.9.5 - Project 1
Overview
Data Laboratory
Preview
Workspace 1
Preview Settings
Preview
Presets
Default Straight
Settings
Manage renderers
Nodes
Border Width
Border Color
Opacity
Per-Node Opacity
Node Labels
Show Labels
Font
Arial 12 Plain
Proportional size
Color
Shorten label
Max characters
Outline size
Outline color
Outline opacity
Box
Box color
Box opacity
Edges
Show Edges
Thickness
Rescale weight
Min. rescaled weight
Max. rescaled weight
Color
Opacity
Preview ratio: 100%
Refresh
Export: SVG/PDF/PNG
Background
Reset zoom
NPTEL

(Refer Slide Time: 29:10)



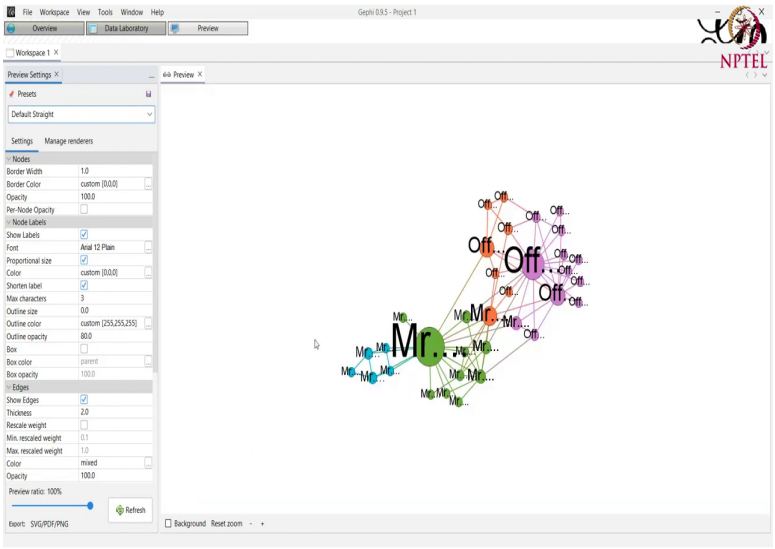

(Refer Slide Time: 19:17)



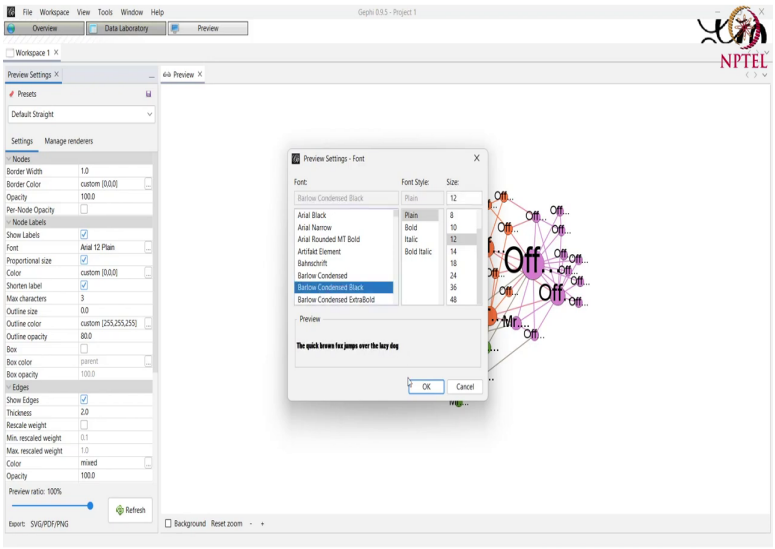

So, for example, 30 here and we refresh it. And it is quite big. We might want to do it 5 here and refresh it and we have a big like you know thicker edges that we have right. Let us just make it even thinner maybe 2 is fine yeah

Then we can show the labels we can have it in proportional size. We can also shorten the labels that instead of a whole officer or mister hi coming we can shorten the label and select the character as probably 3. And we see that only three a characters for each labels are shown.

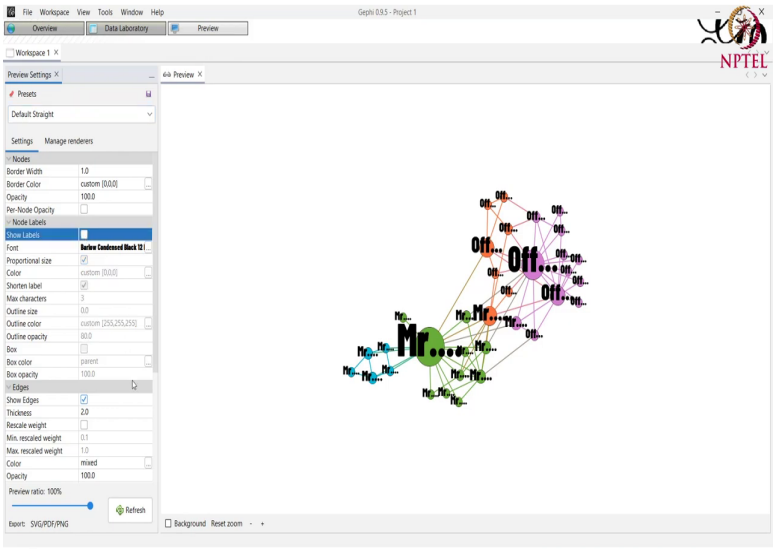
(Refer Slide Time: 29:43)




(Refer Slide Time: 29:48)

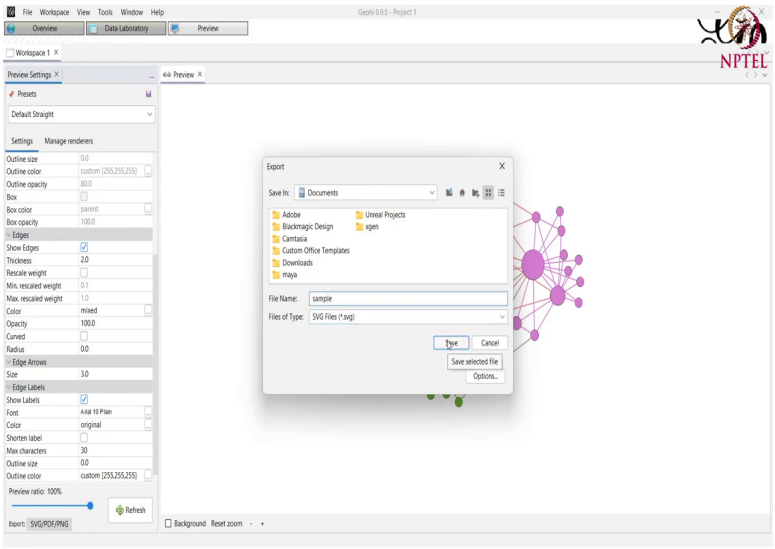



(Refer Slide Time: 29:55)




We can select the different font and the different size of the font that we want. So we select a font, we do this and we can have it like this. So, let us just disable the show label option here we have a graph like this then we have other functions for example, the edges arrow for a directed graph. It will have an arrow for we if the edges also have some kind of attributes associated to it. We can also have those labels shown on the network and so on. Finally after we are satisfied with this preview of the network, we can save it in either s v g p d f or p n g format.

(Refer Slide Time: 30:30)




By using this export functionality here. We select this and we simply just like we name our graph some way we just name its sample here. And we select the types for example, let us select an s v g type and we just press on save. And then if we just go to the location where we have saved it. We see that we have this sample at s v g here we click on it and we can see that we have this graph with us which can be open in this like this an any h m l format h m l supported software right.

So, today we I gave you a very brief overview of the Gephi software. So, this was just to get you started and get you motivated towards using the software. Of course, there are a number of more options a number of more different tweaking and the things that you can try in the software for the visualization for the analysis as well as for the preview aspect of the things.

And I would encourage you to have a look of all these different attributes on your own and explore the software.

Thank you.