**Social Networks**
Prof. S. R. S. Iyengar
Department of Computer Science
Indian Institute of Technology, Ropar

Lecture - 22
Handling Real World Network Datasets
**Datasets: Analysing using Gephi**

Hi everyone. In the previous video we saw how we can use network x package on Python to analysis network data sets. In this video we are going to learn some basic features of software called Gephi which is also used for the analysis and visualization of networks. Gephis in open source tool which is written in Java; although Networkz provides a number of functions to analysis networks. However, when it comes to visualization Gephi provides much more flexibility. So, let us get started and look at the interface of Gephi software, I all ready have a Gephi installed in my system which is a one two, but in case you want to know how we can do that I quickly show you that.

(Refer Slide Time: 00:52)
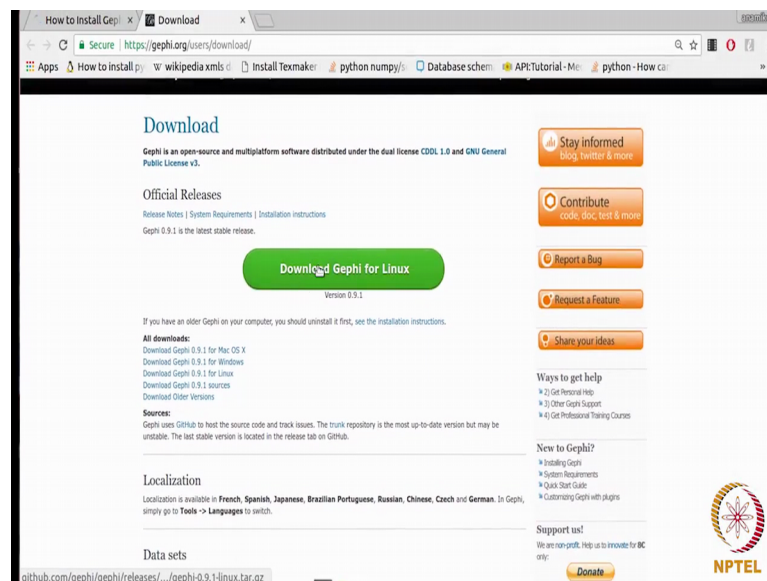
So, it is pretty simple I will just download Gephi for one two. So, we can follow the first link. So, we have to download a file I will click this link and this while that we have to download the version is 0.9.1.
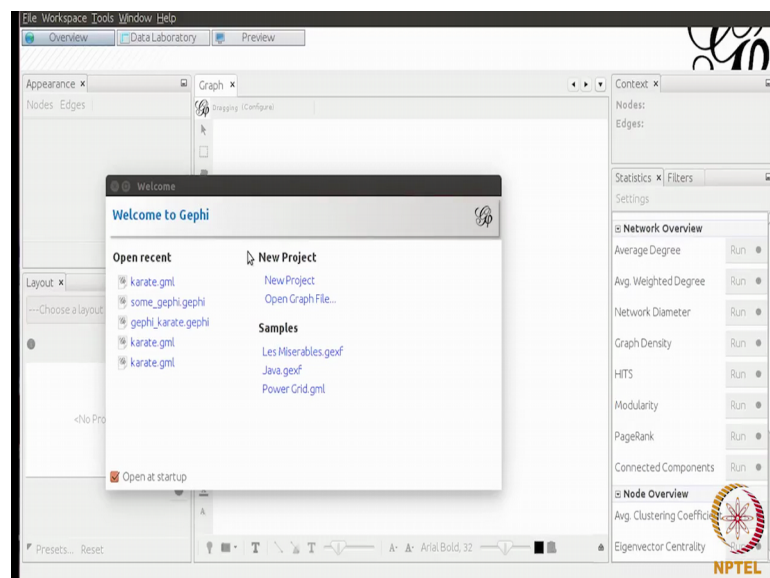
(Refer Slide Time: 01:14)
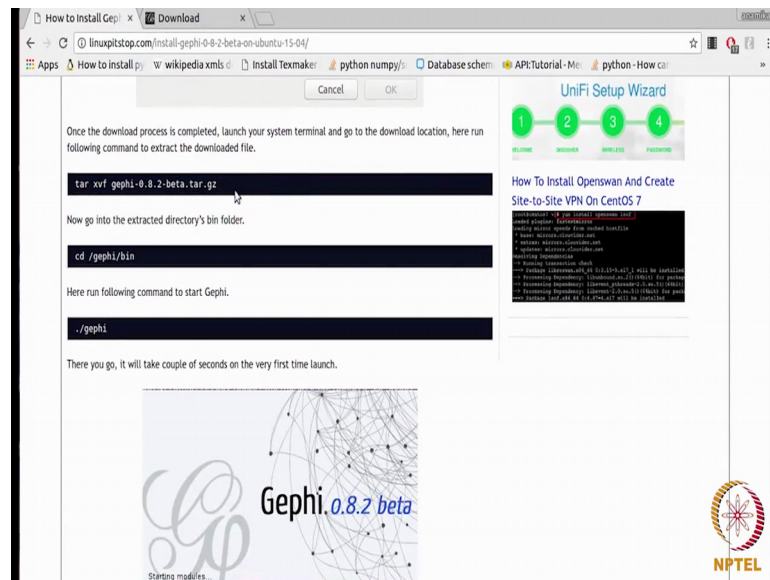
So, we are going to download this file after you download it you where ever you have saved it you will open the terminal there in that folder, and you will run this command to extract the Gephi folder out of it and after that you will get Gephi folder extracted out of the file, and you will go inside bin and after that you will run dots slash Gephi that is how you will run this software.
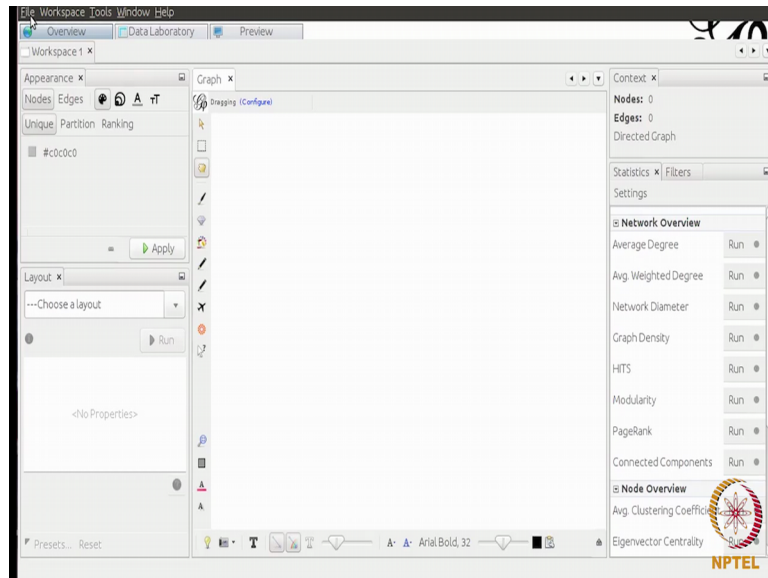
So, I already have a downloaded in my systems. So, I am going to show you how it can started; I am going to open the terminal here.

(Refer Slide Time: 01:49)

Since, I mean Gephi two to bin and then I will write dot slash Gephi. So, that is how we start the software this is the interface of Gephi we are going to start a new project. So, I will click here.

(Refer Slide Time: 02:05)

Now, this are the various options; now in order to analyze a network in Gephi we have to first open that the file of that network here. The most commonly used format by Gephi community is dot g e x f. However, there are lots of other formats which are also supported in Gephi.

(Refer Slide Time: 02:29)
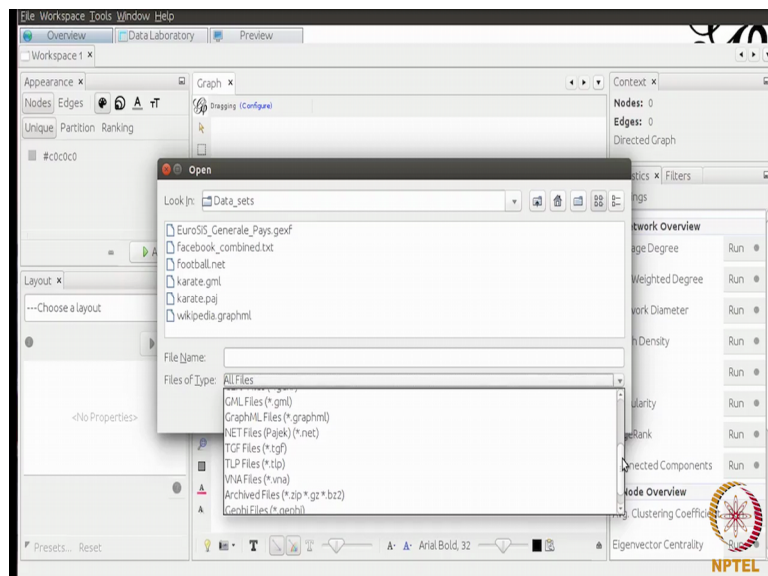
For example let me show you when you click open here this are the formats that are supported by Gephi. As you can see there is a CSV, there is GXF there is GML graph M L dot net Pajik files. So, number of formats are fortunately supported by Gephi; in this video we are going to take a small network in we are already downloaded this karate dot GML in the previous video. So, we are going to continue with that. So, I am selecting this file and I am opening it. So, this is what I get.

(Refer Slide Time: 03:01)
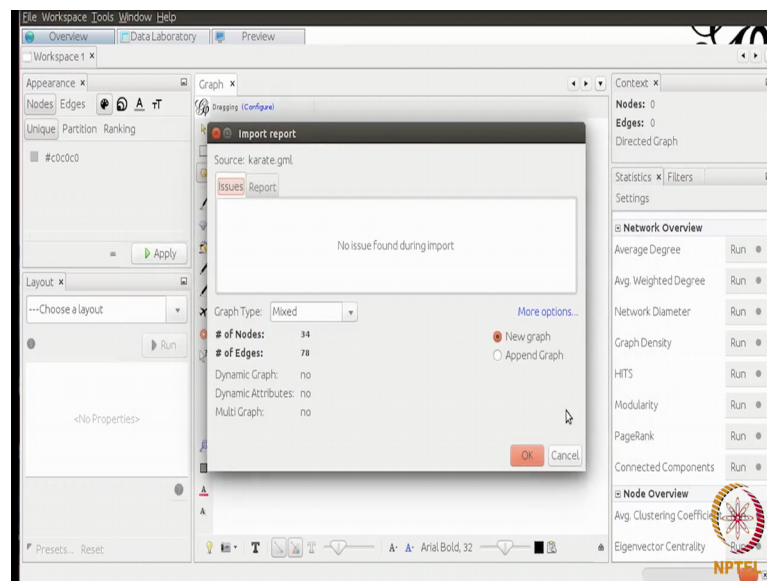
So, this is graph type which is mixed here since it is not mentioned anywhere what kind of graph it is whether it is there it is directed or undirected. Since we know that karate is undirected graph I am going to select undirected here then it showing me the number of nodes and edges and it is not a dynamic graph, and it is not a multi graph I am I want a new graph. So, I am going to click.

(Refer Slide Time: 03:30)
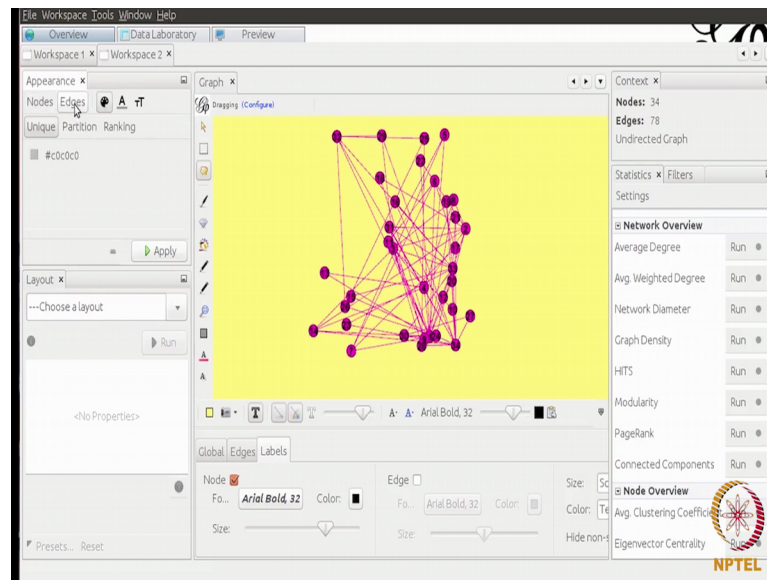
So, as I do that this is the graph that I get let me click here and see more options here. So, on this interface there are few options in the reference side there are few option about bottom there are few options on the right hand side, my aim is to introduce you to this software that we can use it for your analysis.

So, let me quickly start this, on the bottom you see some of the options for example you can change the back ground color. So, you can always change that in case you want then there is a option of zooming, let me zoom in this network. So, that I can visualize it better and then you can go to this edges tab. Here you have this pointer which you can use to in case of thickness of the edges. So, I am going to keep it like this, and then you can go to labels I am telling you the most commonly used features out here. So, when you go to labels and if you click this node, you get the label of the nodes here and this is how you can change the size of the labels as well.

Now this is not looking so good because the size of the node is very small, we can always change that I will show you how we can do that as well and here let me show you can change you can also display the labels of the edges, in this graph there is no label sign to the edges. So, it is not going to make any difference. So, these are the main features of available of here. So, this are some short cuts that you can use for changing the thickness and this is what changing this font size of the label of the nodes. So, that is about it.

Now, let us go to the left hand side panel. So, here we have this appearance panel and under that we have nodes tab and edges tab. So, once this nodes tab is clicked we have some four options over here, the first option is for the color and second option is for the size. Since you want to change the size of the nodes so I am going to show you this first. So, right now the size is 10 let me make it 30, let us say and then I apply and this, what I see. So, labels have gone because the color of the nodes is black. So, we can change that I am going to click this color pallet.

And this is the color which is there by default we can change it let me give some color here and when I apply you see the color of the nodes as well as the edges as change by this is by default it changes the color of the edges as well, we can change that further I will show you the and this for the color of the labels let us keep it black itself as of now and this for the size of the labels.

Now, let us click the edges tab and see what options are here. So, we have the color of the edges since we have to change it, let us change it to some bluish color and apply and the color of the edges as change.

(Refer Slide Time: 06:55)
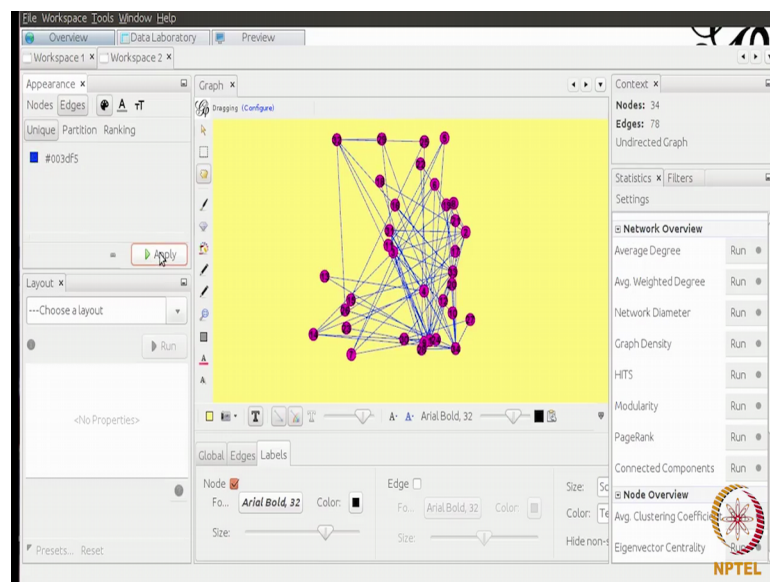
Now, this is for changing of the labels of edges since there are no labels; so would not make any difference, and this is for changing the size of the labels alright. Now in the bottom left corner out here you see an option for changing the label layouts. So, layouts

basically tell us the way the node should be should be position and how the graph should look like.

So, there are number of layouts here and we can change it, I am you can just pair with them I am going to show you one of the layout here which may be used full in certain scenarios. Some times what happens that these nodes are overlapping? So, one node is hidden behind the other so that way you cannot see them properly.
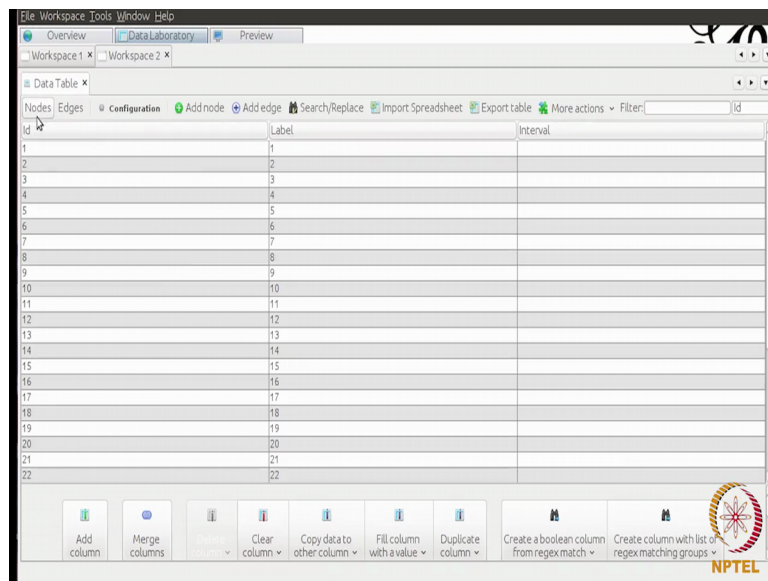
(Refer Slide Time: 07:56)
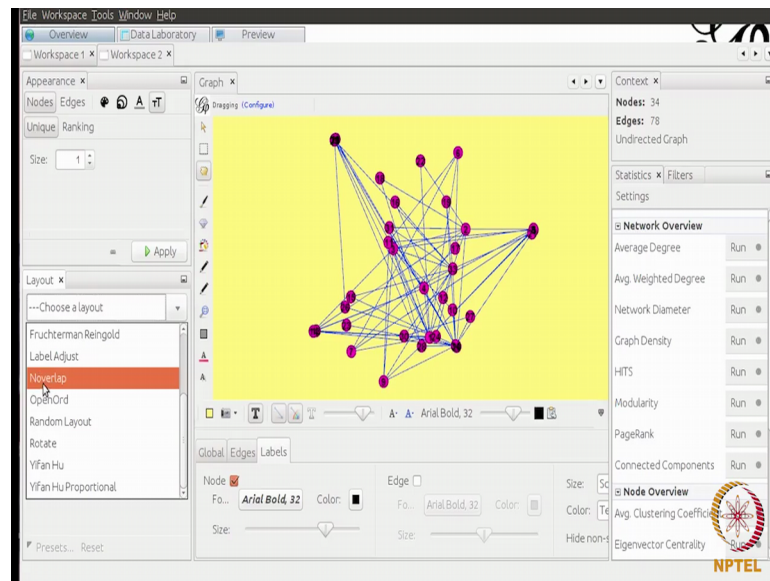
So, I am going to show you one format that can help such a scenario, before that I have to show you one of the nice features of Gephi and that is you can manually position the nodes. Now this is something that you cannot do n network x, if you have to change the positioning you have to right code for that you cannot simply do this.

So, this is one of nice features of Gephi. Now as I was telling you sometimes nodes overlap each other I am expressively over lapping some of the nodes so that I can show you functioning of one of the layouts. So, I am just expressively putting some nodes on top of the other nodes. So, let us go here and see there is one layout that is called no over laps. So, I am taking it and I am joining it and you see the nodes have scattered, the once which are over lapping they are no more over lapping. So, that is what is the effect of this layout. There are a number of other layouts as well so you can just read the documentation of these layouts and see what they do. Let us go to the right panel now on the top you see it just face a number of nodes number of edges and it is an undirected
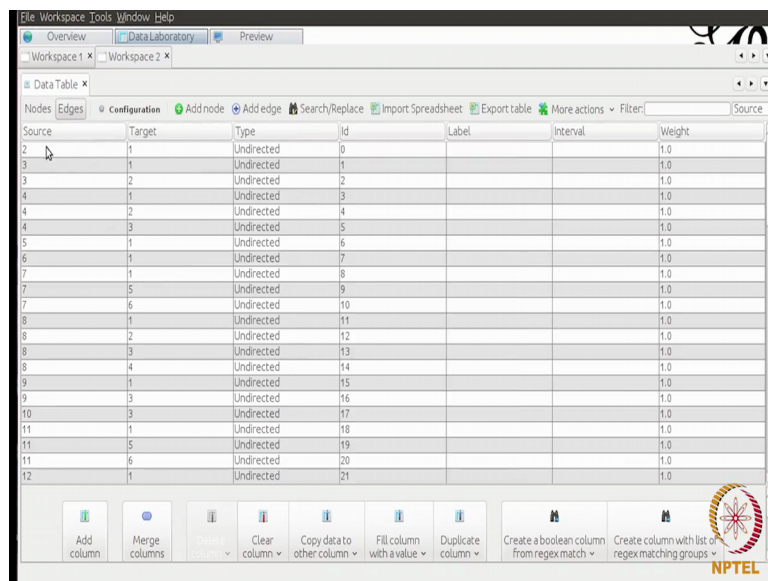
graph. As you go down you have a number of functions that we can apply on the given network.

Now, before we get into these functions I want to show you two things over here. So, we are currently working in this over view tab, there are two more tab on the on the top left corner. So, one of them is data laboratory.

(Refer Slide Time: 09:39)

(Refer Slide Time: 09:47)

Let us see what is there. So, here you see the id of the nodes and the label of the nodes and when you click on the edges you see all the edges in terms of source and target. So,

this is the basic information about the network in the form of a spread sheet. So, there are some other features also available over here for example, you can add new nodes here you can add edges here and you can also export this table.

So, one more feature that should be noted here that when you apply some sort of operation on the network the values that will be calculated out of that operation will also be added to this spread sheet. So, once you apply lot of operations and then you want to save the details, you can just export the table and you can make use of that table in some other tool as well. So, that is nice feature.

(Refer Slide Time: 10:42)
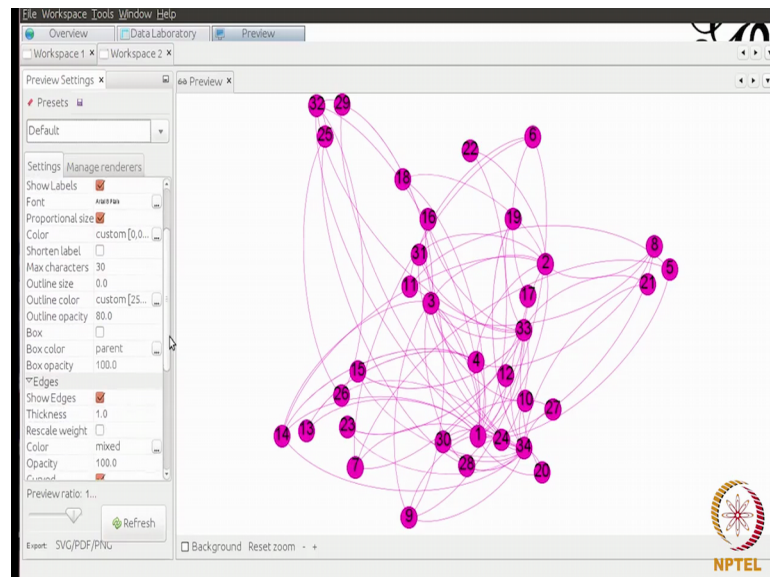
So, let us get back to the over view and let me show you the third tab is well the preview tab, when you click this and you press refresh your network appears over here with a very nice rendering, which you can download and use it the way you want. So, you can export it this formats that is SVG, PDF, PNG you can also change various properties of this network for example, if you want to show the labels you can click here and you can change the forms after clicking it you have to press refresh, now you see the labels here you can also change the font.

For example; let me reduce the forms to 8 because it is not looking so nice here let me press refresh and label as reduced. So, you can change it the way you want there are lot of other features as well for example, you want to show the labels or show the edges or

not of course, you want to show the edges you can just play around this features and you can also change the back ground color let me just show you this features.

(Refer Slide Time: 11:48)
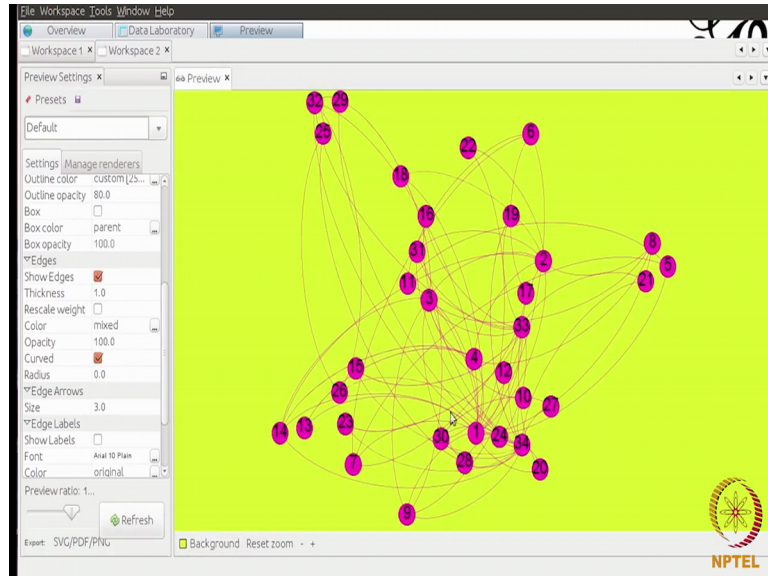
So, this is how you can play around here and then you can download it and you can use it. Let us go back to the over view tab and we were checking the functions that Gephi provides here. So, we have this function average degree let us click on run.

(Refer Slide Time: 12:09)
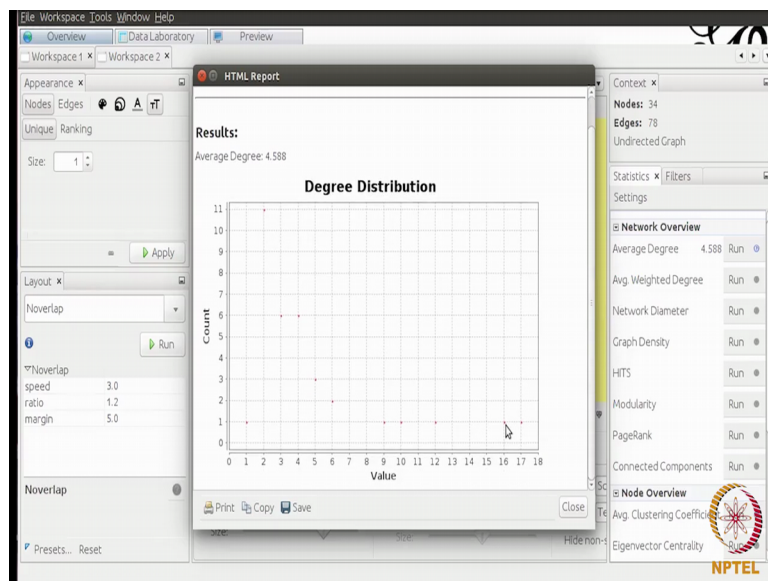
So, just like a button and the average degree and number of other things will be calculated. In this network average degree is open filed and it also given the degree

distribution automatically as you can see this is sort of power loss plot as we (Refer Time: 12:26) previous video as well. Now you can save this plot you can copy it you can print it. So, it is up to you want to use this information. Once you calculate the degree there are a couple of more features that you can see on the left hand side, now let me show you that. So, here nodes tab is clicked and you see one more option here that is ranking.

So, sometimes what we want; we want to change the appearance of certain nodes based on certain properties for example, we want that nodes which have high degree should appear little different be it in terms of color or be it in terms of the size of that node. So, such features are easily available in this software. So, we have this node clicked here and let me click this color here, and when I clicked the ranking here it is asking me how we want to ram the nodes.

(Refer Slide Time: 13:31)
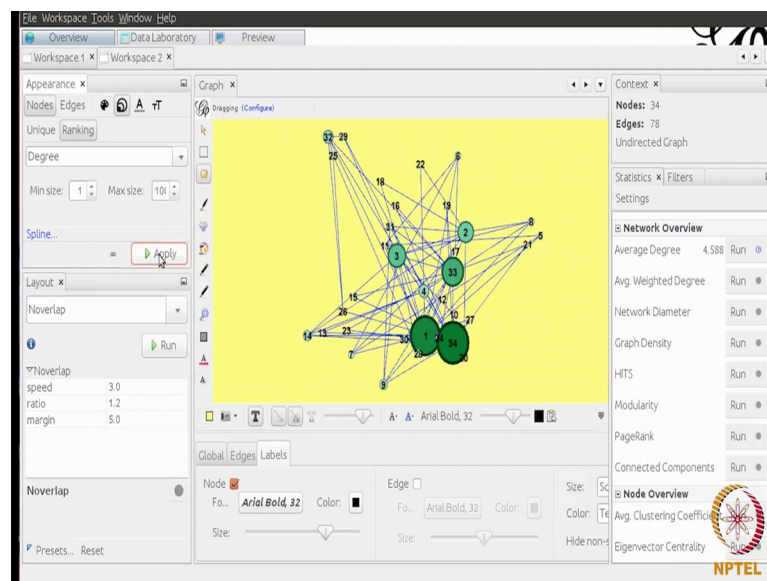
Now, I selected degree here, after this when you press apply you see the color of nodes as change the nodes which have high degree or darker in color and nodes which have less degree are lighter in color. So, you can quickly see by looking at the network you can see which are the nodes that are having high degree. So, this is one visualization feature, similarly if you want that the nodes which have high degree should appear bigger in size, you can also do that. So, I am clicking this button here and I will. So, as of now the size is unique that is 30 right.

So, we want to change it based on degree. So, I will select degree here and the minimum size is one and maximum size is 100, I will go with a head with that press apply. So, this is what you see the size is of the nodes as change based on the degree. As you can see there are two nodes two three nodes which have very high degree compared to the other nodes, so according the size of the nodes as change. So, this is one another nice visualization features so that is about degree, let us check out more features over here we have average weighted degree.

(Refer Slide Time: 14:47)
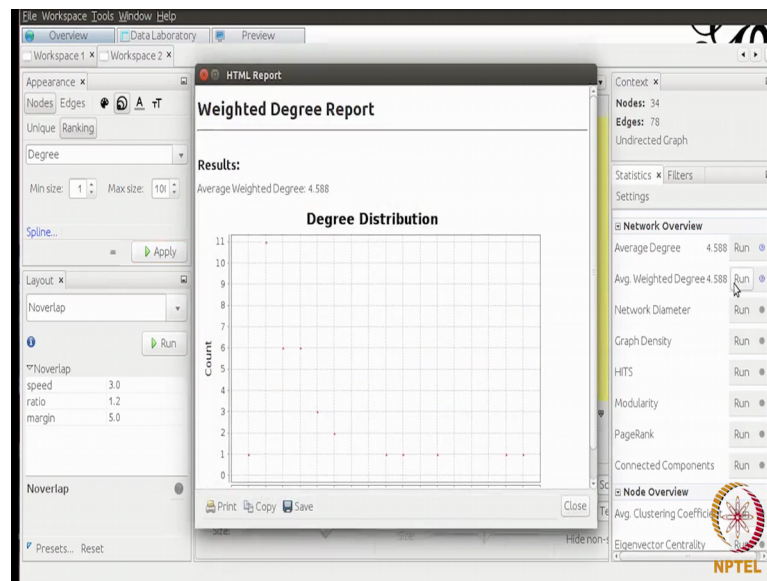
Since this graph is unweighted. So, this is now going to make difference. So, we getting same average degree here, this is because the weight is one for all the edges. We also have a number of other functions here for the example diameter we have density, where modularity, page rank, connected components clustering coefficients. So, when learn the other concepts in the throughout the course, you might want to come back here and then analyze your network based on that parameters that we have learnt. So, that will be nice because now you see what all features this software provides.

So, as when you learn about that feature that you can come back here and import here graph file and then see the features so that will be a good start for you. Let us go back here and as I want to show you one more thing in detail laboratory, since we calculated degree of the nodes that degrees as come over here. So, another column as be added over here where degrees of the nodes have been displayed. So, you can export this table and

you can use it the way you want if you go the preview tab again if you press the refresh button you get that network over here.

(Refer Slide Time: 16:04)
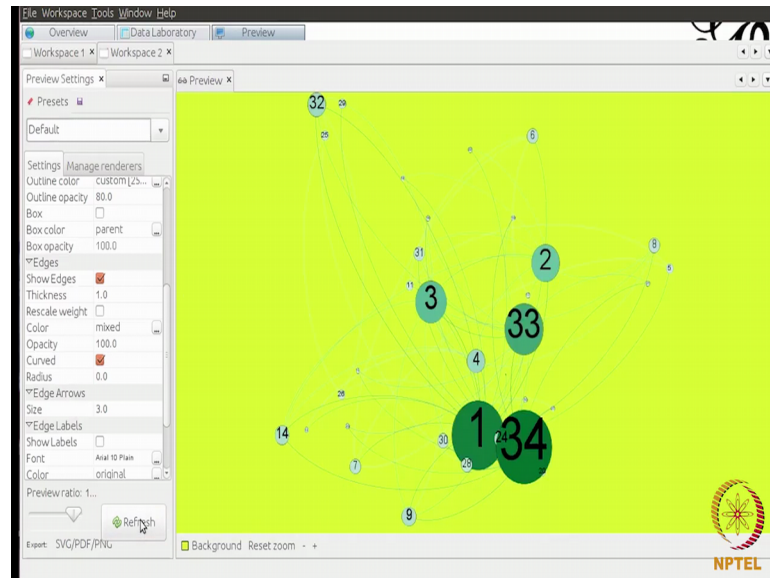
So, this is the final rendering of the network and from the left hand side panel you can change the various properties of this network. So, I think that was pretty much basic introduction to Gephi software and may be a good starting point for you.

Thank you.