Lecture 25: Science Mapping I

Hello Learners, welcome again to the course. So, in the last week we have discussed descriptive analytics. We have conducted the descriptive analysis on two of the columns of the dataset. So, our dataset was the test dataset where we have done the descriptive analysis on the number of publications and the number of citations. Then we have seen about the how we will calculate the central tendency and the measure of dispersion in R and we have also seen that how we can analyze the Bradford law and the Lotka law in R. So, in this week what we will be doing is we will be discussing the structure and dynamics of the knowledge.

Basically we will be talking about science mapping, what exactly science mapping is and what are the different components. Science mapping is an emerging area comprising scientific citation analysis, visual analytics, data mining and knowledge discovery. Presently a wide variety of science mapping tools are available and few of them we will be discussing in this course. So, in this week we will be discussing what science mapping is, what exactly the network is and how we can analyze those networks.

We will also create some of the networks in R. So, in science mapping we represent the structure of knowledge using some visuals also known as graphs or networks. It shows the relationship between different subjects, sources, topics, institutions and countries and the relationship between them. It has various applications like it helps in answering various questions like who are the influential institutions or countries on a research topic or who are the authors have expertise on the certain topic. Also it helps in understanding the emergence of the discipline and interdisciplinarity of the trending research and many more things which we will be discussing throughout this week.

The main aim of science mapping is to explore the structure and dynamics of scientific knowledge. Various methods have been proposed to study the different dynamics of scientific knowledge like some of the popular techniques of science mapping are co-citation analysis, co-word analysis, keyword co-authorship analysis etc. So, before going ahead let us first see the data that we have extracted from Scopus and we will try to understand what we have already discussed and where we can apply the science mapping techniques. So, in the previous week we have already taken up this quantitative analysis of publication and citation where we have counted the publication and citation. So, we have done this column and we have done this column of the data set then we can also use these three columns to analyze based on our previous week like we can count the length of the papers. So, whenever we download the data from different bibliographic data sources these details are given but sometimes if any of the details are missing like if page count is missing. So, we can easily count so we can easily just do what we do. we will just 1 0 2 2 minus 9 9 3. So, these are simple arithmetic operations and then we can count the length of the paper and based on this analysis we can find out what paper is the most

lengthy paper in that particular source or that particular domain. Then we have already discussed that this is the default column whenever we download the data so this is not to be considered. Then we can just analyze the different language documents so if we are doing the analysis on a particular domain. We can analyze what different kinds of language documents are published on that particular domain.

Then we can do analysis on document type so we can easily count the number of document types like what is the frequency of articles, what is the frequency of other document types like note editorial and all. Then here also we can do the analysis easily of what all documents are in open access and under open access what are the documents are in green open access diamond open access or gold open access. So, what data we have is still to be analyzed. So, the first is that authors have not analyzed this kind of collaboration between these authors. Then also we have this title which is a text data so this particular analysis I will be discussing when I will be talking about the text mining.

Then we have this affiliation data where we can see the collaboration of the different institutions. Then we have this abstract data so here again we can do the analysis of text mining so I will be discussing this when I will be talking about the text data in the upcoming weeks and these are the index keywords. So, bibliographic data sources give the two kinds of keywords so one is author keywords and another is index keywords. So, author keywords are the keywords which are supplied by the author during the submission of the article. So, whenever an article is published, there are keywords which are given by the authors and then the index keywords are the keywords which are identified by the system and they index that publication under those keywords.

So, these are the two data sets which we have to analyze. So, a science mapping is a kind of network analysis where we study the network of authors, institutions or maybe the keywords. So, networks are everywhere we live in a kind of network system. We have a network of family and friends. If we define formally a network it is a structure made up of connected things.

So, what are these connected things? These connected things are known as nodes. So, for understanding purposes, imagine this is a situation where A is a friend of B. So, a friend of B and B is a friend of C and C is a friend of A. So, this is a network of three friends which are A, B and C. If we see these A, B and C these are known as nodes or vertices.

So, these are the nodes and the path between A and B, B and C and C and A is known as link or edge. So, this is what a network is. So, take another example where we are aiming to study the industry collaboration of an institution. So, A is the first university. So, for example this is the first university who has collaborated with industry X and B is a

university.

So, B is another university but B has not collaborated with any industry. So, if we see this network diagram we can easily say that A and X have an edge between them and hence they are connected with each other. So, we can say that A university is collaborating with industry X but B does not have any collaboration with industry X. There is no edge with either of the industry X or any of the industry. So, we can say that B does not have collaborated with any of the industry.

So, we have various applications of network analysis. Some of them are listed here. So, anthropologists use network analysis to map relationships within communities including kinship ties, social hierarchy and patterns of cooperation. By analysing these networks researchers can understand social structure, cultural norms and information flow within the community. Biologists use network analysis to study interaction between proteins within the cell.

Here nodes represent proteins and edge represents physical interaction or functional association. The analysis helps in identifying protein complexes, pathways and regulatory mechanisms underlying cellular processes. Social media network analysis is another application of network analysis which is most popular in recent time. So, social network analysis is used to study communication patterns on social media platforms. Nodes represent user or account and edge represents the interaction such as follow mentions or retweets.

By analysing these networks researchers can understand information dissemination, influence dynamics and community structures on social media. So, network analysis is also used to study the transportation system including road networks, airline routes or subway systems. So, here nodes represent locations and edges represent the transportation link. This kind of analysis optimizes the transportation infrastructure, assess accessibility and to study urban dynamics. And in the end this is what our focus of this week is, citation and collaboration network in the academic literature.

It shows the relationship between individuals or organizations engaged in collaborative activities such as co-authorship network, collaboration in scientific research. Nodes represent authors, research institutions or research topics and edge represents collaboration or joint publication. By doing the analysis we can identify key contributors, research trends and interdisciplinary collaborations. In addition to the example listed here, it is important to note that network analysis has many more applications across various other fields where researchers and practitioners of diverse domains use network analysis to gain valuable insights into the structure, behaviour and interaction within the

network leading to informed decision making and impactful interventions. So, why do the network analysis? So, to understand the structure of a data set to identify nodes, edges and patterns of connection between them.

Also to identify the most influential elite player or actor in a particular network. To identify the relationship between entities in the network. Also to identify the hubs or detect the community in the network. So, to take the informed decision about the changes across diverse fields to understand the emerging fields and the knowledge discovery. So, in the end, to get the predictive insights.

So, to define the concept of a network we will be using the concept of graph theory. So, what graph is? It is just a structure of a network. So, to represent the structure it needs two things. One is a vertex which we call a node and another thing is an edge, the connection between the two vertices. So, vertices are just nodes, actors in the network.

You can say if you are doing the analysis of collaboration of authors. So, vertices will be the institution if we are doing the analysis of collaboration between institutions and edge is a line between those nodes. So, we are doing the analysis of collaboration between institutions. So, what edge will be? So, edge will be the collaboration link between those two institutions or the two authors. So, there are two kinds of graphs.

One is directed graph and another is undirected graph. So, in the directed graph it shows the path but in the undirected graph we have the edge but no direction. So, in simpler words we can say that directed graph is a kind of one way path but the undirected graph is a kind of which shows the mutual relationship. So, let us understand both by this illustration. So, for directed graphs take the example of Twitter and Instagram follow networks.

So, you are aware that on Twitter and Instagram if this is a Mohit and if it follows Mukesh so that means there is a directed edge between Mohit and Mukesh but if you know on Twitter and Instagram it is okay that Mohit follows Mukesh. So, it is not sure whether Mukesh will follow back to Mohit. So, this is an example of a directed graph. Now we take the example of an undirected graph. So, to understand the undirected graph just imagine the network of Facebook.

So, here this is Mohit and again here this is Vijay. So, if Mohit and Vijay are friends on Facebook. So, it is sure that Vijay and Mohit are also friends on Facebook because it is mutual like if Mohit sends the request to Vijay so Vijay will accept and after accepting the request both will have a friend. So, let us take one more example of a collaborative author. So, if you say for example this is the author Priyanka and this is another author

Arvind.

They have collaborated on a paper say for example this is a paper A. So, we can say that Preyanka is the co-author of this paper A and Arvind is also co-author of paper A or we can say in that that Arvind is a co-author of paper A and then Preyanka is also co-author of paper A. So, this is a kind of example of undirected graph and we will be doing most of our analysis on undirected graph. So, an adjacency matrix is used for representation of a graph. It shows the relationship between two nodes.

Adjacency matrix is a kind of square matrix where the number of rows and the number of columns are equal. So, if we are saying that there are n crosses and n matrices are there so both the number of rows that is n and the number of columns that is n both are equal and that n is the number of nodes of that graph. So, let us say x1, x2, x3 and x4 are the nodes. So, let us say this is x1, this is x2, this is x3 and say for example this is x4. So, there is a like these are connected, these are also connected, these are connected, this is also connected.

So, for this particular kind of a network how we will draw the adjacency matrix. So, the matrix will be like this. This is x1, x2, x3, x4, here also x1, x2, x3, x4. So, in this matrix wherever there is an edge we will write the 1 and wherever there is no edge we will write the 0. So, to x1 and x1 we will write the 0 then to x1 and x2 there is an edge so we will write 1 to x1 and x3 there is an edge we will write 1 to x1 and x4 there is no edge so we will write 0.

Similarly to x2 to x1 there is an edge that will be 0, x2 to x3 will be 1 then x2 to x4 0 then in x3, x3 to x1 1, x3 to x2 1, x3 to x0, x3 to x4 1. For x4 there is no edge between x4 and x1 so it will be 0, for x4 to x2 there is no edge it will be 0, x4 to x3 there is edge so we will write 1 and x4 to x4 it will be 0. So, this we call it as this is our adjacency matrix. So, let us say that we have this kind of a data set where we have authors and the paper contributed by them. So, these are the author details and then we have this title of the papers.

So, here we have a total number of authors, so one is this author then this is second author then this is third author and then we have this is fourth author. So, the total number of authors is 4. Now we can represent this by a network graph of 4 authors like Mohit Garg. So, we can say that this is Mohit and then we will say this is Mukesh, we will say this is Preyanka and say this is Vijay. Now we can say that Mohit Garg and Mukesh Vaira have contributed one paper so they are connected.

So, there is an edge between Mohit Garg and Mukesh Vaira. Then we can say that

Mukesh and Priyanka have contributed one paper. So, we can say that this is the path between Mukesh, Vaira and Priyanka. Then we have like Priyanka and Mohit have also contributed one paper. So, this is also an edge between them and in the last we have one paper which is contributed by Mukesh and Vijay.

So, this is our network graph. So, what our edge pair is, our edges are Mohit, Mukesh. Then we have Mohit and Preyanka. Then we have Mukesh and Preyanka and then finally we have Mukesh, Vijay. So, in an undirected graph if you say Mohit and Mukesh or Mukesh and Mohit the both are same. So, if I am saying that Mohit and Mukesh there is an edge so if and then I will say Mukesh, Mukesh.

So, for this undirected graph it is So, we are showing that there is a mutual relationship. So, both contributed to the paper. So, from this we can create the adjacency matrix. So, how will we create the adjacency matrix? So, we will create something like this. So, here we will write Mohit, then Mukesh, then Preyanka and then Vijay.

Similarly here we will write Mohit, Mukesh, Preyanka and Vijay. Now this will be zero. Now if there is an edge between any of the two nodes we will write one. So, for Mohit and Mukesh there is an edge we will write one. For Mohit and Priyanka we have an edge. We will write one.

For Mohit and Vijay there is no edge so we will write zero. For Mukesh and Mohit we have this edge so we will write one. For Mukesh and Mukesh it will be zero. For Mukesh and Preyanka it will be one.

And for Mukesh and Vijay it will be one. For Priyanka, Preyanka to Mohit we have one edge. Preyanka to Mukesh we have one edge. From Preyanka to Preyanka it will be zero. And from Preyanka to Vijay there is no path.

For Vijay, Vijay to Mohit it will be zero. For Vijay to Mukesh there is a path. it will be one. For Vijay to Preyanka zero. And for Vijay to Vijay it will be zero. So, our adjacency matrix is created.

So, if we see that we have a data set of this kind of thing. So, first we have counted the total number of authors that is four. Then we have created the four nodes. First thing is done. After that we have seen the edges between them.

So, what are the edges? So, Mohit and Mukesh. So, this is one edge. So, we have created one edge. Then Mohit and Priyanka have created this edge. Then Mukesh and Priyanka have created this edge.

And then Mukesh and Vijay this is one edge. So, what are these edges? So, these edges represent that Mohit and Mukesh have contributed paper one. Mukesh and Preyanka have contributed paper three. Mohit and Preyanka have contributed paper two. Mukesh and Vijay contributed paper four.

Understand? So, this is the whole network. So, now we will create all this analysis in R. So, how will we do that? Let us see that. So, to do that in R first we need a package. So, that package is igraph. So, if you have not installed the package igraph you can just simply run this.

So, the package was installed. Now we will call the igraph package or package called. So, what is the first thing we will do? We will first define the nodes. So, here first we have identified the nodes. So, our nodes are four and which are those? Mohit, Preyanka, Mukesh and Vijay.

So, the same thing we will do in R also. So, these are our nodes. So, Mohit is one node. Mukesh is one node. Preyanka is one node and Vijay is another node.

And we know how to create that character vector. So, we can create this. So, this is our node. So, if I see the value of nodes. So, there are four nodes. Now after identifying the nodes the next thing we have done? We have identified the edge.

So, if you see here these were the edges. So, we have identified the edges. Mohit, Mukesh, Preyanka, Mukesh, Vijay. So, we will be declaring the edges. So, how will we declare the edge? So, we will be creating the matrix and this is an edge. So, it is showing that Mohit and Mukesh have an edge.

Then Mohit and Preyanka have an edge. Then there is an edge between Mukesh and Preyanka and finally Mukesh and Vijay. And we will do by ncol2 and we will put the byrow true. ncol is basically a column. We have already studied what ncol is and what nrow is.

And then we will see this matrix. So, this is our kind of edge matrix where Mohit and Mukesh have one edge. Then another edge between Mohit and Preyanka. Then another edge between Mukesh and Preyanka and the last is the edge between Mukesh and Vijay.

So, now we will be creating the undirected graph. So, we have the nodes. We have the edge. Now we have to create the graph. So, to create a graph we have this function called graph from edge list and if I run this so for this argument is edges. So, our edges are

already defined and we will say that directed false means it is an undirected graph.

It is not a directed graph. So, we will run this and we will assign this to an object let us say GU. What GU is? GU is an undirected graph of these edges. So, if I run this and now if I see the output of this. So, it is showing that Mohit and Mukesh have an edge.

Mohit and Preyanka have an edge. Then Mukesh and Preyanka there is an edge and finally Mukesh and Vijay there is an edge. So, technically our graph is created but we need to visualize what exactly this graph looks like. So, what will we do? We will plot this graph for better understanding. So, for that we need to call this plot function and here our object name is GU and I am using the vertex color sky blue.

Size of the vertex means node size I am putting to 30. So, this exactly represents the size of the labels then the label distance then this edge color and then the width of the edge. And this is I am putting exactly the title of that graph and then the vertex label color. So, the node label color is in the black color. So, if I run this, our plot is this. So, this is what our plot what we were discussing in that there is a node Mohit which is connected to Preyanka and Mukesh then there is a node which is connected to Mohit and Mukesh then there is a node Mukesh which is connected to Preyanka Mohit and Vijay and then there is a node Vijay which is connected to Mukesh only.

And what are these edges? These edges are showing that Mohit and Preyanka have collaborated for one paper then Preyanka and Mukesh have collaborated for one paper then Mukesh and Mohit have collaborated for one paper and in the last Mukesh and Vijay have collaborated for one paper. So, our graph is created. So, this is how we can create the graph in R. So, here I will be visualizing only the small graphs for complete analysis on those graphs we will be discussing in the next week and then in this plot we can change all those colors. We can simply put here either green then if you see it automatically identified that this is green or we can say for example we can see the number of colors.

So, we can take maybe violet red for edge color and then label color we will have let us say if I run this my graph is created. Also we can use the color codes here that we have discussed in the previous week. So, we have different color codes like we can have the hexadecimal code and we can write here like we can write here hexadecimal code if I run this. So, if you see these are the labels Mohit, Preyanka, Mukesh and Vijay.

So, this is how you can plot the graph in R. Now this is what we have discussed about the undirected graph with no weights but the graph can have some weights also. So, let us discuss the graph with some weights. So, this is a kind of a situation where the number of

authors is 4. So, the total number of authors is 4 only but this time we have a total of 8 publications.

And so our nodes are Mohit then Mukesh, Preyanka and then Vijay. If you see the edges we have an edge between Mohit, Mukesh. So, this is done then this is also that only that edge then we have Priyanka and Mohit. So, Mohit and Preyanka then we have Mukesh and Preyanka, Mukesh and Preyanka and Mukesh and Preyanka. We have 3 papers but the edge will be Mukesh and Preyanka.

Then this is done, this is done, this is done. And we have this Mukesh and Vijay and Mukesh and Vijay. So, we have this Mukesh and Vijay 2 times, 2 publications. Mukesh and Priyanka have 3. Mohit and Priyanka we have 1 and we have Mukesh and Mohit 2.

So, in total we have 8 publications, 2 plus 1, 3, 3 plus 3, 6 and 8. So, total publication we have total publication we have 8. Now if we see the edges list, these are the 4 edges. So, this is what we have the edge. So, if we represent this kind of a data set into this kind of a graph, some information is missed here.

What is that information? These 3, 2, 1 and 2. So, how can this problem be solved? This problem can be solved if we can solve this kind of relationship here and if we can give some weight to it. So, for Mohit and Preyanka there is only 1 publication. For Mohit and Mukesh there are 2 publications. For Mukesh and Preyanka there are 3 publications and Mukesh and Vijay 2. So, what can we do? We will make this edge only this much thickness but we will just increase the thickness of this edge and then this is the most thickness edge will be there because it has the most number of publications and then okay.

So, now if we represent this data into this graph one can easily say that for this particular data set the pair of Mukesh and Preyanka has collaborated the highest number of publication and what those number of publications are 3 and who has collaborated the minimum number of publications those are Mohit and Preyanka with only 1 publications. So, for this how will we create the adjacency matrix? It is very simple: the edge will be the same as what we have discussed here. So, the edges will be there but now we will give the weight and the weight is we will give Mohit and Mukesh to 2 and Mukesh and Preyanka to 3 and then Mukesh and Vijay to 2. So, now we will be creating the weighted adjacency matrix for this example.

So, Mohit, Mukesh, Preyanka and Vijay and here Mohit, Mukesh, Preyanka and Vijay. Now from Mohit to Mohit it will be 0 Mohit to Mukesh we have 2 earlier it was we had 1 only but now because the weight is 2 so here it will 2 will come then Mohit to Preyanka it

will be 1 then Mohit to Vijay 0 then Mukesh to Mohit 2 then Mukesh to it will be 0 Mukesh to Preyanka 3 then Mukesh to Vijay 2 then Preyanka to Mohit is 1 then Preyanka to Mukesh it will be 3 Preyanka to Preyanka it will be 0 and Preyanka to Vijay 0 Vijay to Mohit 0 Vijay to Mukesh 2 and Vijay to Preyanka 0 and Vijay to Vijay 0. So, this is a weighted adjacency matrix for this example. So, let us now see the weighted graph in R. So, our package we will call this library so our nodes we have already created so these are our notes.

Now we will define the weights to edges. So, how will we create? So, for Mohit and Mukesh there are 2 publications so we will assign the 2 weights. For Mohit and Preyanka there is only 1 publication we will assign the 1 weight. So, for Mukesh and Preyanka there are 3 publications we will assign the weight 3 and for Mukesh and Vijay there are 2 publications we will assign the weight 2. And now we will have 3 columns and by row it will be 2. If we see the output of these edges it is created so for Mohit and Mukesh the weight is 2 for Mohit and Preyanka the weight is 1 for Mukesh and Preyanka the weight is 3 and for Mukesh and Vijay the weight is 2.

So, our edges matrix is created and these are the weights. Now if I have to create a graph for this we will use the function graph from the edge list and these are our edges. So, what exactly are these? I am saying that you take these edges. If I run this so it is saying that these are the edges and again I am saying that it is not a directed graph.

So, if I run this my graph is created. If I see this. So, here the relationship between nodes has been created. Now we will plot this. So, how will we plot? We will use this function plot. So, here this is our graph then our this is the color of the nodes this is the size then I am putting the nodes whatever the nodes I have created. So, if I run this, So, what exactly we did here is so if you see in the earlier one without the weight we have simply assigned the weight to all the edges but here so what exactly this it is calling if I see the out of this so it is assigning the weight accordingly between these thing.

So, this if you see the thickness between Mukesh and Priyanka the highest so this is exactly the 3 is. So, if we zoom it so if you see this thin line between Mohith and Priyanka it is showing that they have collaborated less for a publication but if you see here the thickness between Mukesh and Priyanka so we can easily say that these are the most collaborative author for this particular data set and then we have Mukesh and Mohith and Mukesh and Vijay. So, similarly the way we have changed the color in the previous graph we can change the color here. So, here if we want to see this directed graph how we will do that so instead of this false we will write true and now if I run this.

So, this is a directed graph. So, this is how we can create the directed graph in R. So, this

is about some of the key concepts of the graph so we have discussed the directed graph and undirected graph and we have also seen how we can assign the weight to the edges of the graph. So, now we will discuss some of the key measures in network analysis. So, one is that in centrality measures some of the common centrality measures are degree centrality, betweenness centrality, closeness centrality, eigenvector centrality measure. So, degree betweenness and closeness centrality were proposed by Freeman in the 1970s and related to unweighted networks. So, let us first discuss the degree centrality measure and then I will discuss the betweenness centrality and the closeness centrality measures.

So, degree centrality counts the number of edges between the nodes. So, it helps in knowing the number of connections a node has and it shows the popularity of the node. So, say for example, this is our graph we have already discussed. So, this is Mohit then this is Preyanka then this is Vichai. So, here Mohit is connected to two nodes. What are those two nodes so one is this one and another is this one so for Mohit there are only two connections. Now if we see the connection of Preyanka so for Preyanka there are two connections one is with Mukesh and another is with Mohit.

So, for this also two now for Mukesh if we see the connection this is one two three so for Mukesh it has three connections one with Mohit one with Preyanka and another with Vijay and Vijay it has only one connection with Mukesh only. So, if we see here Mukesh has the highest degree of centrality because it has the highest number of connections in the network. So, we can say that Mukesh is the influential node in this particular network. Also we can normalize this degree of centrality measure so how we normalize so we just simply divide the degree of centrality by n minus 1 n is the number of nodes so if we count the normalized degree of centrality for Mohit, Mohit normalize it will be 2 upon 4 minus 1 equal to 2 by 3 so it will be 0.

67. Now if we count for Preyanka the normalized will be 2 upon 3 0.67 if we see Mukesh for 3 upon 4 minus 1 3 upon 3 1 and then Vijay 1 upon 3 it will be 0.33. So, we see here again Mukesh has the highest normalized degree of centrality so now we will see this computation in R. So, how will we do that in R? So, this is our nodes then we can define our edges so these are edges and then we can create a graph so we are creating an undirected graph. Now if I see the degree of this so for Mohit 2, Mukesh 3, Preyanka 2, Vijay 1 so if I see here on our analysis so Mohit 2, Preyanka 2, Mukesh 3 and Vijay 1.

So, now if we calculate the normalized degree centrality so we will do like this under degree function we will call that normalize is true if I do like this so for this Mukesh this is a one normalized degree centrality but for Mohit and Preyanka 0.66 and for Vijay 0.33 for the same we got it here. So, this is about the degree centrality measures so now we will discuss the betweenness centrality measure.

So, betweenness centrality measure is based on the shortest path it shows the nodes that form the bridges between other nodes. So, to calculate the betweenness centrality measure the formula is centrality measure of a node is say for example we are calculating for node X in a graph it will be so what this is this is total number of shortest path that passes from A to B and then what this is this is the total number of all those shortest path that passes from node X. So, let us understand this within our example. So, our example is this: we have this node Mohit, we have another node Priyanka, we have another node Mukesh and we have another node Vijay and this is the edge. Now if we have to calculate the betweenness centrality measure of Mohit so what we have to do is to calculate the path between Preyanka and Mukesh. Now the shortest path between Preyanka and Mukesh is this. There is another path from Preyanka to Mukesh that is this path but this is a long path compared to this because here Preyanka has to move to only one step but here two steps.

Similarly for Preyanka to Vijay this is the shortest path one step this one the total two steps and if we go by this way so already Preyanka has to move two steps here so for this it will be three steps. So, these are the two shortest way for Preyanka so Preyanka to Mukesh and then Preyanka to Vijay is via Mukesh and another path between Mukesh and Vijay this is the only shortest way Mukesh can also come via Priyanka and Mohit and then Vijay it will be longer way so this is the only shortest path but none of these shortest path passes through Mohit so betweenness centrality measure of Mohit is zero. Now if we calculate in a similar way betweenness centrality of Preyanka, we need to see the path between Mohit and Vijay. So, we see the shortest path for Mohit to Mukesh so this is the shortest path then to Mohit to Vijay this is the shortest path Mohit go this way this will be lengthy way so here also there is no path where Preyanka is coming in between so betweenness centrality measure of Preyanka is also zero. Now if we calculate the betweenness centrality of Vijay so if we calculate the betweenness centrality of Vijay we have to calculate the shortest path between Mohit and Mukesh then Mohit and Preyanka and then Preyanka and Mukesh.

So, from this graph we can easily see that the shortest path to Preyanka and Mukesh is this one to Mohit and Preyanka is this one and to Mohit and Mukesh is this one and nowhere Vijay is coming in between so the betweenness centrality measure of Vijay is also zero. Now if we calculate the betweenness centrality measure of Mukesh so we need to calculate the shortest path between Mohit and Preyanka and Mohit Vijay and then Preyanka to Vijay. So, for Mohit to Preyanka there is one path but Mukesh is not coming but Mohit to Vijay this is the only shortest path and this is the only path is there and Mukesh is coming in between and between Preyanka and Vijay also Mukesh is coming so the betweenness centrality measure of Mukesh is 2. So, we can say that the betweenness

centrality measure of Mukesh is the highest and the other three have the zero betweenness centrality measure Mukesh has the most connection in the network.

So, let us see this computation in R how we exactly calculate the betweenness measure in R. So, to calculate the betweenness measure in R we have this function betweenness or if I run this so you can see that for Mukesh the betweenness is 2 for others Mohit, Preyanka and Vijay it is 0. It is very simple. Now we will discuss the third centrality measure is closeness. So, similar to betweenness measures centrality we have another measure which is a closeness centrality measure.

So, it exactly finds the nodes that are closest to all other nodes in the network. The closeness score helps in identification of the expert influence in the network. So, say for example again we will go with our graph. So, to calculate the closeness centrality measure we need to first calculate the shortest distance between one node to all other nodes then the closeness centrality measures will be the inverse of the sum of all those distances. So, we will count the closeness centrality measure for each of our nodes.

So, for Mohit to Preyanka if you see for Mohit to Preyanka it is one step Mohit to Preyanka the shortest distance is one step. There is another way like if we go like this: Mohit to Mukesh and Preyanka there are two steps but the shortest one step is directly connected. Then for Mohit to Mukesh we have one step and for Mohit to Vijay we have two steps. So, because after moving the first step will reach Mukesh and then the second step is Vijay.

So, it is two steps. So, closeness centrality measures of Mohit will be one upon one plus one plus two that is one by four. So, it will be equal to 0.25. Now we will calculate the closeness centrality of Preyanka. So, for Preyanka to Mohit one step for Preyanka to Mukesh one step for Preyanka to Vijay.

So, if we count closeness centrality for Preyanka it will be one upon one plus one plus two it will be one by four equal to 0.25. So, now if we count the closeness centrality measure for Mukesh. So, for Mukesh we have Mukesh to Preyanka one step Mukesh to Mohit one step Mukesh to Vijay one step. And if we count the cc Mukesh it will be one upon one plus one plus one equal to one by three which is equal to 0.

33. Now if we count the closeness centrality measure for Vijay. So, for Vijay to Preyanka it is two for Vijay to Mohit it is two for Vijay to Mukesh it is one. So, cc for Vijay will be one upon one plus two plus two it will be one upon five it will be 0.2. So, here if you see that Mukesh again has the highest closeness centrality measure compared to the other three nodes.

So, if you see Mukesh has 0.33 of centrality measure then Mohit and Preyanka have 0.25 and then Vijay has the least centrality measure of 0.2. So, what it means is that Mukesh is well connected in the network, it is connected to other nodes and then Vijay has the least centrality measure. So, we can say that Vijay is the least connected in the network; it has a closeness centrality measure of 0.20 and the other two nodes have the same closeness centrality measure so they are equally connected.

So, there is another measure known as the eigenvector centrality measure. So, this eigenvector centrality measure was given by a sociologist named Philips Buncis in 1987. So, what exactly does it calculate? So, it does not only consider the connection between the nodes but also see with which these nodes are connected. So, let us see an example. So, say for example this is the case here Preyanka then this is Mukesh Mohit this is to Vijay this is Jukkur work.

Now here if we see that both Mukesh and Preyanka have the degree centrality of 3 because Mukesh has 3 connections with other nodes. So, one is this one one is this one one is this one so it has 3 of degree centrality measure then same with Preyanka 1, 2 and 3. But if you see here Preyanka has one influential connection with Jukkur work. So, from here we can conclude that Preyanka is more influential compared to Mukesh. So, this is what an eigenvector centrality measure is: it captures global influence of the nodes that focus not only on quantity but also on quality.

So, one of the famous applications of eigenvector centrality measures is the Page Rank algorithm. So, let us see what Page Rank algorithm is. So, have you ever imagined how Google works? How about when you just simply enter some keywords in the search bar and Google retrieves the results and then after retrieving the result it ranks those all results. So, how exactly does it work? So, it basically computes this Page Rank algorithm based on which it orders the number of pages in a result. This Page Rank algorithm was given by these two people Larry Page and Sergey Brin.

So, they are the founders of Google also. So, what they did was they computed the ranking of the web pages using this algorithm. And they presented this computation in the 7th worldwide conference held in 1998. So, in Page Rank algorithm the rank of a web page is not computed just based on the link but the normalization of all the links are done by measuring the importance and number of outbound links to the neighboring pages. So, let us understand the Page Rank algorithm with an example. Now consider this is an example of Mohit then we have here Priyanka then here we have Mohit.

So, you consider that Mohit has cited Priyanka just you consider that Mohit has cited

Priyanka in their paper. Then Priyanka has cited Mukesh and Mukesh has also cited Priyanka in their paper. And then Mukesh has cited Mohit in the one paper. So, this is one kind of a citation network where graphs are made based on how they have been cited.

So, this is how they have cited. Now to calculate the Page Rank of a node what we have to do is a formula. So, this is the formula of Page Rank of a given node. So, what exactly is it? It is a Page Rank of the node from where the input link is coming and these are the leaving nodes. Let us calculate the Page Rank of all the three nodes here. What will we do? We will make the table like this: Mohit then Priyanka then Mukesh.

So, at the first iteration we will assume that the rank of all the nodes are equal. So, what will we do? We will make sure that there are three nodes. So, we will give the 1 by 3 here, 1 by 3 here, 1 by 3 here. Now if we have to calculate the Pagerank of Mohit, what will it be? It will be from where the input link is coming from Mukesh. So, Page Rank of Mukesh divided by the leaving node to Mukesh.

So, what exactly is the Page Rank of Mukesh? So, this was the Page Rank of Mukesh, the old rank. So, 1 by 3 and the number of leaving nodes to Mukesh is 2. So, 1 is this lm and 1 is this lm. So, it will be 2. So, it will be 1 by 6. Now if we have to calculate the Pr of Priyanka, so it will be Pr of Mohit by l of Mohit plus Pr of Mukesh by l of Mukesh.

Why? Because Priyanka is getting input from Mohit also and then input from Mukesh also. So, Priyanka is cited by Mohit also and Mukesh also. So, now we will calculate Pr of Mohit is 1 by 3 and the leaving nodes are 1. Then Pr of Mukesh is 1 by 3 divided by 2.

So, it will be 1 by 3 plus 1 by 6 equal to 2 by 2 plus 1 by 6. So, it is equal to 3 by 6 1 by 2. So, the rank Pr of Priyanka is 1 by 2, and the rank of Mohit is 1 by 6. What will be the rank of Mukesh? So, for Mukesh, it will be Pr of Priyanka then l of Priyanka.

So, what exactly Pr of Priyanka? 1 by 3 divided by then leaving nodes to Priyanka only 1. So, it will be 1 by 3. So, this equals 0.16, this equals 0.50 and this equals 0.33. So, from here we can say that Priyanka has the high PageRank value compared to Mukesh and Mohit.

Mukesh has the second highest PageRank value. Then Mohit has the least PageRank value. And our highest PageRank value is for Priyanka. So, this is just an example of how we compute the PageRank of the nodes. But here I have concluded the PageRank based on just one iteration. But we have to take many iterations to calculate the PageRank with accuracy.

So, when we will be analyzing a huge amount of data, it will be quite difficult for us to do the analysis in a manual way. So, we have to use some software. So, in R we can do it easily. So, we will see how we can easily calculate this PageRank of this graph. So, this is our graph. So, what I am saying is that, so these are the edges.

So, there is an edge between Priyanka and Mukesh, then Mohit and Priyanka, then Mukesh and Priyanka and then Mukesh and Mohit. So, I am considering this as an edge. So, we will create the graph and this time we are creating a directed graph.

So, this is how we will do it. Our graph is created. So, we will see the graph. So, it is said that there is an edge between Priyanka and Mukesh. So, if I see it here, there is an edge between Priyanka and Mukesh. Then I am saying that Mohit to Priyanka, Mohit to Priyanka also, that is okay.

Then we have another edge, Mukesh to Priyanka. So, we have Mukesh to Priyanka, there is an edge. Then we have the edge Mukesh to Mukesh.

So, we have the edge from Mukesh to Mukesh. So, we have another edge from Mukesh. So, this is our graph. So, let's visualize it. So, I am just visualizing it. So, this is my graph. So, if you see here, the double arrow between Priyanka and Mukesh is showing that Mukesh has sighted Priyanka and Priyanka has sighted Mukesh. What if you see between Mohit and Priyanka and Mukesh and Mohit, there is a single arrow link. So, what exactly it is showing is that Mohit has sighted Priyanka and then Mukesh has sighted Mohit.

So, to calculate the page rank, what we will do, we have this simple function page rank GD. So, GD is our graph and if we do compute this, our page rank is calculated.

So, page rank for Priyanka is 0.39, for Mukesh 0.38 and for Mohit is 0.214. So, this is what we have also got. Our value was not these but this was the case that Priyanka has the highest page rank value then followed by Mukesh and then in the last Mohit because a lot of iteration is already done. But we have done only one iteration. So, this is an example of a page rank algorithm. So, in this lecture, I have discussed the concept of science mapping and how we can do the science mapping using network analysis.

We have discussed the different kinds of networks. We have discussed the directed network then undirected network. Then we discussed the centrality measure to map the strength of a network. We have discussed the degree of centrality measures. Then we have discussed the betweenness centrality measure, closeness centrality measure and in the end we have discussed the eigenvector centrality measure and we have taken the example of the page rank algorithm. So, this is all about the first lecture of this week. In

the next lecture, I will be discussing some of the techniques of science mapping that includes bibliographic coupling, co-citation analysis, co-authorship analysis and then covert analysis. So, see you in the next lecture. Thank you. Thank you.