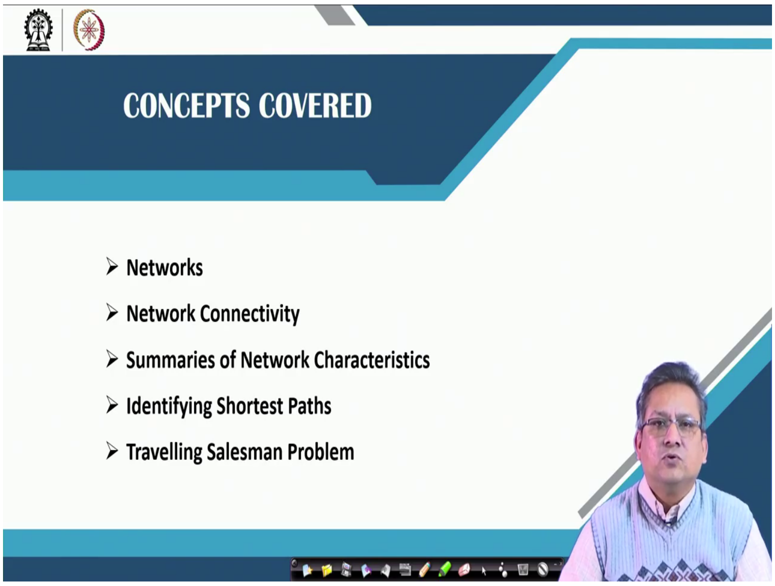
Geo Spatial Analysis in Urban Planning
Prof. Saikat Kumar Paul
Department of Architecture and Regional Planning
Indian Institute of Technology, Kharagpur

Lecture – 14
Network Analysis

Welcome dear students we are in the lecture 14, we have already covered the basics of GIS and GIS operations; how data gets stored, how we store the attributes? So, now, we are into the applications we have covered the basics of GIS. So, today we shall look into the network analysis in GIS.

(Refer Slide Time: 00:41)

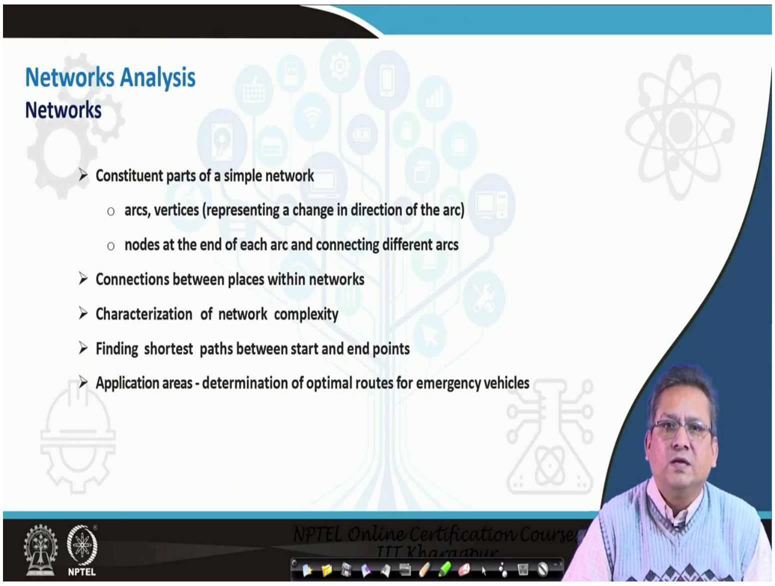


So, the topics that we are going to cover today are we are going to introduce what is a network, though we had already seen or I mean how a network can be coded in GIS, how the data can be stored? We would look into network connectivity concepts, we would see the different types of summaries of network characteristics, it is complexities, connectivitys. So, how to summarize a network, then to identify the shortest path, if we have a network to identify the shortest path.

Now, you come across the application of this shortest path in your Google I mean route finder, say whenever you have or you are travelling in a city and you give your origin and destination it would find the shortest path, either the shortest time distance path or you can find out the shortest cause distance path in a network.

So, one such algorithm which I mean works on finding out the shortest path is Dijkstra's algorithm, wherein you can find the shortest path in a network. Another shortest path problem I mean wherein a person would have to visit n number of nodes, in a given number, in a given network can be solved using a traveling salesman's approach. So, these are the concepts that we are going to touch upon today.

(Refer Slide Time: 02:17)



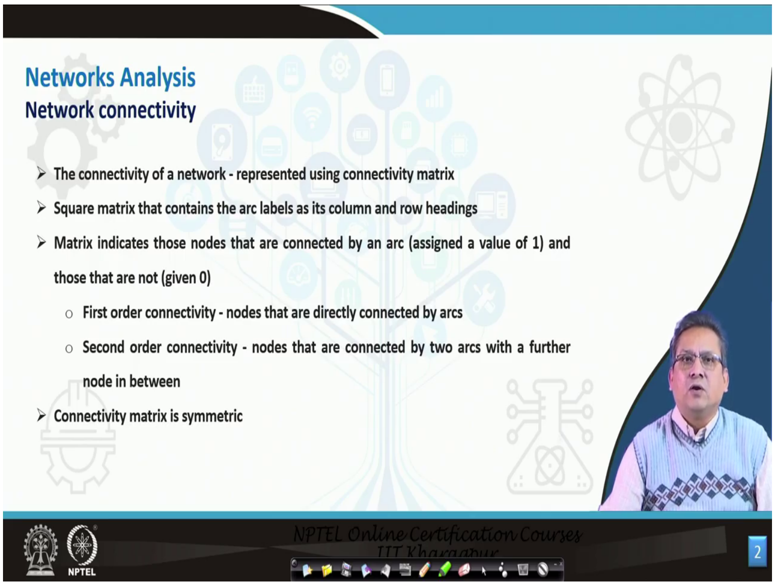
So, let us see what is a network. So, a network is a aggregation of links and nodes and it is very easy to code the data for links put the data for nodes and you can create attributes for the links as well as the nodes. So, in arc GIS we generally for links we call it as arcs and each arc would have it is vertices. So, the order in which you are at is it digitizing would show the direction of the arc. So, otherwise we can also give some indication for directions in a network and we will revisit this while we are discussing further.

Nodes, now the nodes are the endpoint of arcs and they are the points where in the different arcs are connected to each other. So, we can denote a real city or the network of a city I mean that is connected through roads, rail, or other form of transportation network. And, we can establish this network in GIS using arcs and nodes, we can represent such networks in GIS using arcs and nodes.

Now, we can also do a characterization of the network we can do an accessibility analysis, if some of the areas are unserved by network, we can identify that and we can do a characterization of the complexity of the network in a given urban area.

Now, we can also find out the shortest paths between the start and the end point. These algorithms of finding shortest path or visiting n number of nodes, in the most optimal way can be used, in the determination of the shortest corridor, in case of emergency evacuation to route and schedule vehicles. So, these are some of the application areas wherein we can use networks.

(Refer Slide Time: 04:41)



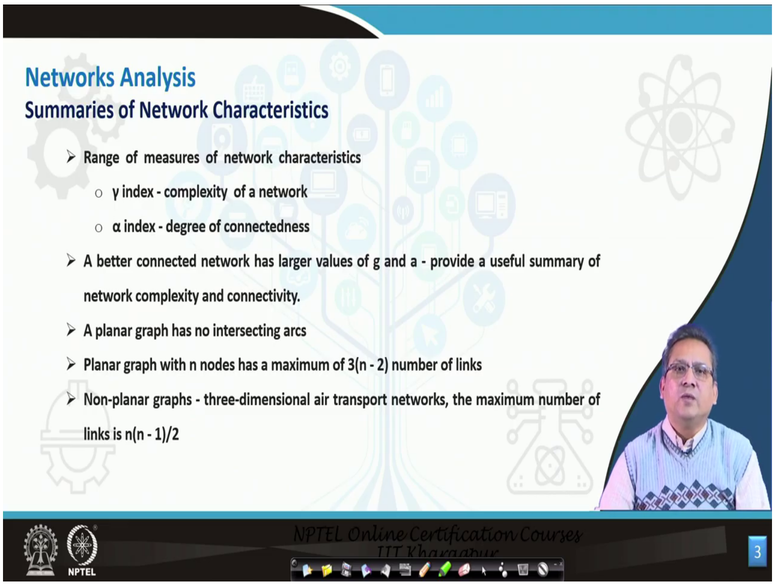
Now, let us see how the networks are connected and how do we represent it? We represent the network connectedness of the network using a connectivity matrix. So, this would be a diagonally symmetric matrix, because I mean the node numbers would be connected to each other. So, I mean this you would have the different nodes as vertical columns and as rows and you would use numbers, probably an impedance value or the distance to travel or the cost to find out the connectivity in the network.

Now, this is basically a square matrix and it would contain the arc level as the column and as row headings. So, these matrix indicates that, if they are connected by arc those nodes the I mean the intersection of these 2 columns and rows would have a value of 1 and 0 otherwise, in case they are not connected. Now, this connectivity between the different nodes can be of different order.

So, let us discuss what is the first order connectivity? So, these are a nodes which are directly connected to each other by arcs. So, such nodes or such connectivity is known as first order connectivity. Now, we have another term which is known as second order connectivity where in the nodes are connected by 2 arcs with a further node in between them. So, you would have I mean 2 nodes connected to each other, through some other node why are some other node.

So, we call a second order of connectivity. Now, as we had already said that this connectivity matrix, wherein we had said that it is a square matrix.

(Refer Slide Time: 06:53)



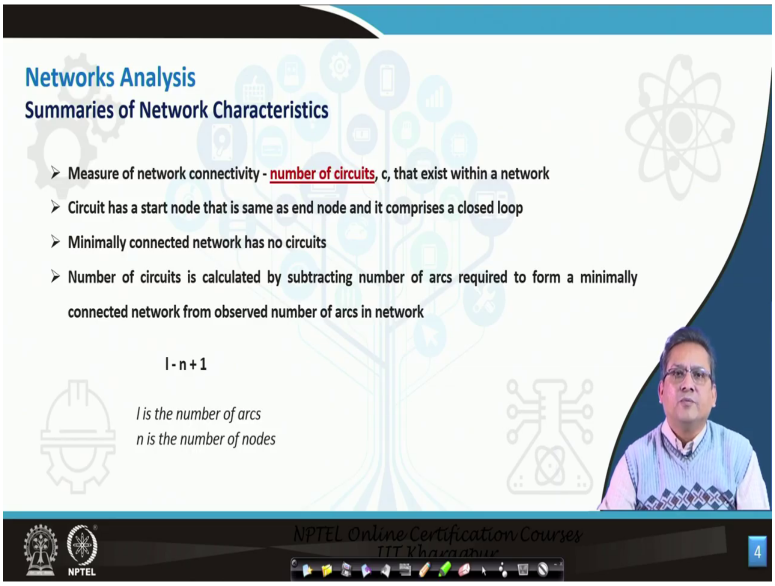
So, it has it is a symmetric matrix and it is a square matrix. Now, the summary of the network characteristics can be given and we have a range of measures to characterize networks. The first among them is known as the gamma index, which gives the complexity of a network and we have alpha index, which gives us the degree of connectedness, we had talked about the first order connectivity, we had talked about the second order connectivity. So, the alpha index it gives us the degree of connectedness.

So, a better connected network would be 1 where these values of gamma and alpha provide summary of the network connectivity's and where these values are larger. So, there is a minor I mean edit in this particular slide, for g please read gamma and for a please read it as alpha, it is some typographical issue because of some fonts in the computer. So, while you are going through this lecture please read g as gamma and a as alpha.

Now, these networks are generally represented through graph theory and there we have a graph which is known as planar graph, which represents networks wherein there are no intersecting arcs or edges or links. As you may call it I mean in different GIS terminologies or books you may come across different terms, but all of them mean the same.

Now, we have these planar graphs, which has n number of node and the maximum possible number of links that it can have is 3 into n minus 2 number of links. Now, there could be non-planar graphs also which are 3 dimensional. For example, we can have a transportation network varying the maximum link is given by n into n minus 1 divided by 2.

(Refer Slide Time: 09:11)

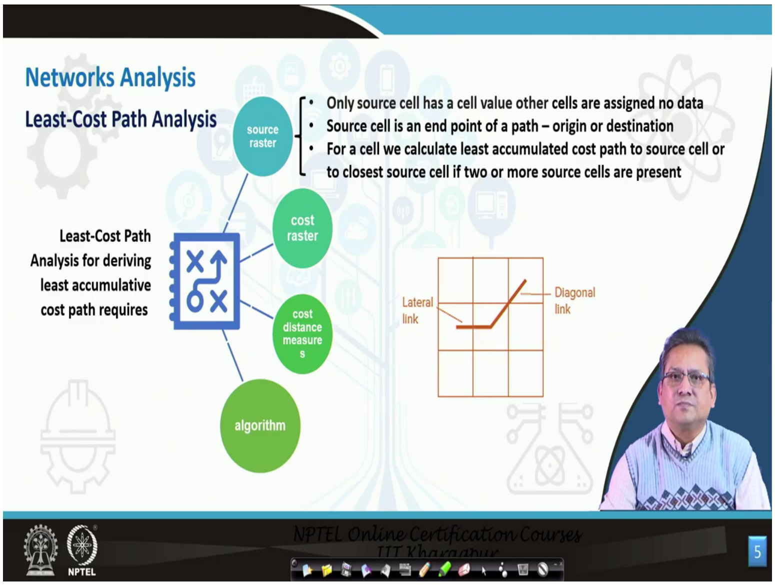


Now, the measure of connectivity is given by a measure a measure c, which is known as number of circuits that exist within a given network. Now, every circuit has a start node and that is the same as the n node. And, therefore, it is a closed loop it comprises a closed loop.

So, if we have a circuit which is minimally connected I mean network which is minimally connected, in that case we would not have any circuits ok.

So, the number of circuits can be calculated by subtracting the number of arcs, that is required for minimally connecting a network from the observed number of arcs in a network. So, you can calculate the number of circuits in a network. So, this is how it is calculated, where l is the number of arcs and n is the number of nodes. So, I mean you can calculate the number of circuits by I mean differencing l minus n plus 1.

(Refer Slide Time: 10:25)

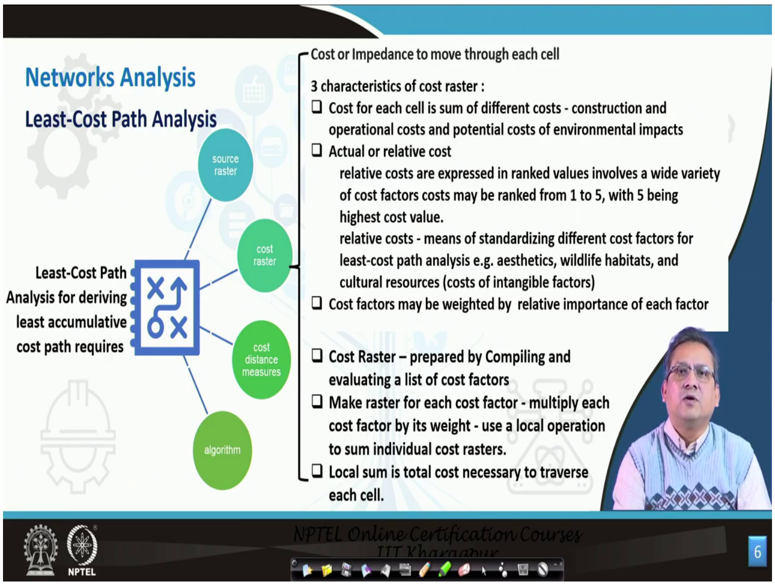


Now, you can see that least cost path analysis is used for I mean the least accumulative cost path is found out. I mean we can use this least cost path analysis to explore the least cost path in a given network. So, in this we require a source raster, we would have a cost raster, we would have a cost distance measure and we would have an algorithm, which would process all these different inputs into an output of the shortest cost path.

So, in this particular graph, in this particular image, you can see that this is a raster wherein the centroids are basically the nodes and these centroids are connected through links. So, this link which goes horizontally or vertically is known as lateral link. And, the link which joins the centroids of diagonally placed elements is known as the diagonal link. Now, in case of source raster you are the only the source raster would and the destination raster would have cell values. And, the other cell values will not have any other data.

So, you can identify in a raster giving the location of those grid points, where your source is and where your destination is where you use is your start point and where is your end point. Now, the source cell is an end point it could be an origin or destination as we have talked about. So, we can calculate the accumulated cost path we shall see an example how we can calculate the accumulated cost path. I mean this cost path is calculated to source cell or to the closest source cell. And, we can find out the distances if 2 or more source cell are presents present, we can find out the alternate your path from the source to the I mean destination.

(Refer Slide Time: 12:57)



Now, we can also find out the cost and the cost implications of traversing between 2 nodes is registered as a cost raster. So, in this analysis we are talking about a raster data model, but we can also do a similar analysis using a vector data model. So, in this case for the cost we have a cost raster. So, the cost raster is either there are a cost to travel across different nodes or it could also be an impedance or a penalty cost of or impedance to move through each cell.

So, it would if you want to move through the different routes in a network it might offer different kind of resistance. So, we shall see, what are the different types of impedances or resistance offered by networks? So, there are three characteristics to a cost raster. So, the cost for each cell is a summation of the different types of cost. For example, construction cost or operation cost, or potential cost of a kind of an environmental impact that might happen, because of pollution or for some other reason. So, you can find out the potential cost such as environmental cost.

So, like you are traveling from a location to a destination. So, the construction of the road would entail some money there would be some operational expenditure because you have to maintain the road. So, that will also I mean you would incur some cost and there would be potential cost in terms of pollution or other impacts so, that needs to that can be accounted. So, we can sum up each of this cost and find out what is the traversal cost from each node to the other node.

So, we can either work the actual cost or we can also work out the relative cost, which are basically ranked values and involves a blanket of cost factors. It may be ranging different I mean factors which could be I mean used for working out the relative cost. So, we can rank the relative cost for relative cost we can rank the values between say 1 to 5, 5 being the highest cost value or it could be in a different scale all together. So, we can rank this values, we can find out the relative cost.

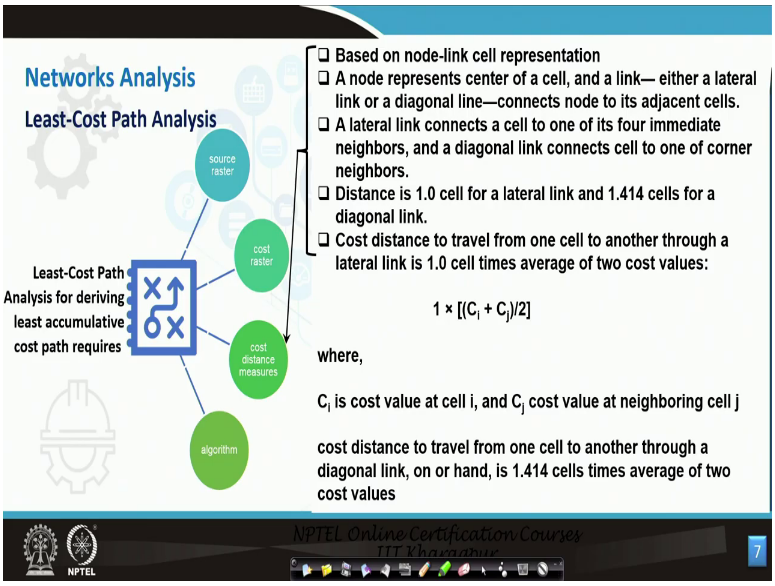
So, what we can do is we can standardize the cost. If we know the I mean aggregated cost the sum of cost over different links like we had talked about construction cost, operation cost, maintenance cost or environmental cost. So, we can aggregate all those costs over all the links and we can standardize that, we can take the summation of that and we can divide the cost of traversing between 2 nodes divided by the total cost of traversing across each of those nodes.

So, that will give us the relative cost. So, it is generally used when you are trying to I mean code intangible values, values that you cannot measure. Like say, suppose aesthetics or I mean some kind of intangible factors, like your cultural resources or say suppose you have a wildlife habitat. So, I mean we tend to work the relative cost. Since, it is very difficult to quantify these kind of intangible factors.

Now, the cost factors can be weighted by relative importance of each factors. So, we can also have a mechanism that we do apportion some kind of weights based on the relative importance of the different factors that we have been talking about. So, probably I mean safeguarding the wildlife habitat could be the prime objective. So, there you may apportion a higher weightage compared to probably aesthetics or some other cost. So, you can have different kind of weights for the different factors. And, the cost raster is compiled by evaluating and summing up all these different cost factors.

So, we make raster's for each of this cost factor and the way we generate the composite cost raster is we multiply each of this cost factor by it is weight. And then what we do is we sum up all the individual raster's to give you the final cost raster. So, I mean we can work out the local sum which is the cost I mean summed up cost to traverse each of the cells from either laterally or diagonally as we had seen.

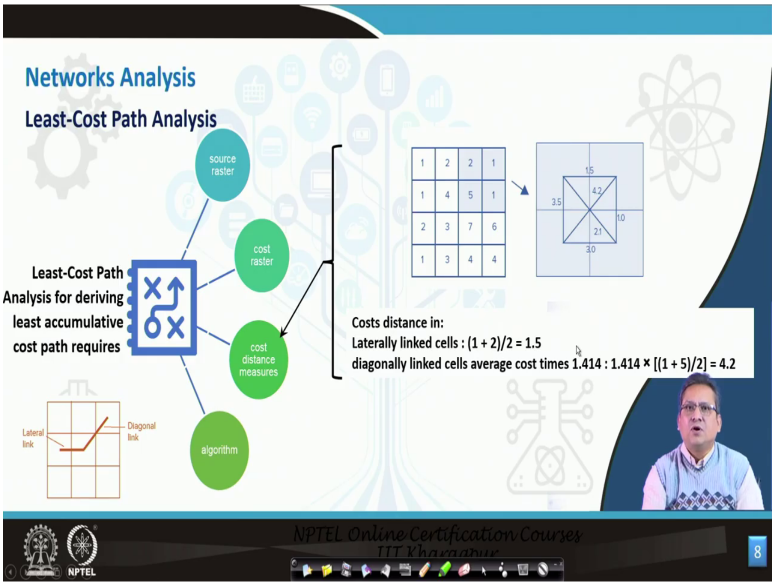
(Refer Slide Time: 18:31)



So, it is based on the node link cell representation. A node represents a cell, and a link and it as we have seen is either lateral or it is diagonal and it connects node to it is adjacent cell. The lateral link connects cells to it is immediate neighbors and the diagonal link connects the corner elements, or the corner neighbors from a given point. Now, this distance if we are traversing in the lateral direction would have to travel 1 cell distance and if we are traveling diagonally we have to travel 1.414 or which is equivalent to root 2.

I mean the centroid between 2 cells. So, that can be worked out I mean you can calculate that and you can work out the centroid between 2 cells. So, in this case in this particular equation you can see that we have product of I mean average of the C 1 and C 2, where C 1 is the cost value at cell i and C j sorry C i and C i and C j divided by 2 where C j is the cost value at the neighboring cell. So, we can work out either the lateral I mean distance to the link or we can find out the diagonal distance.

(Refer Slide Time: 20:17)

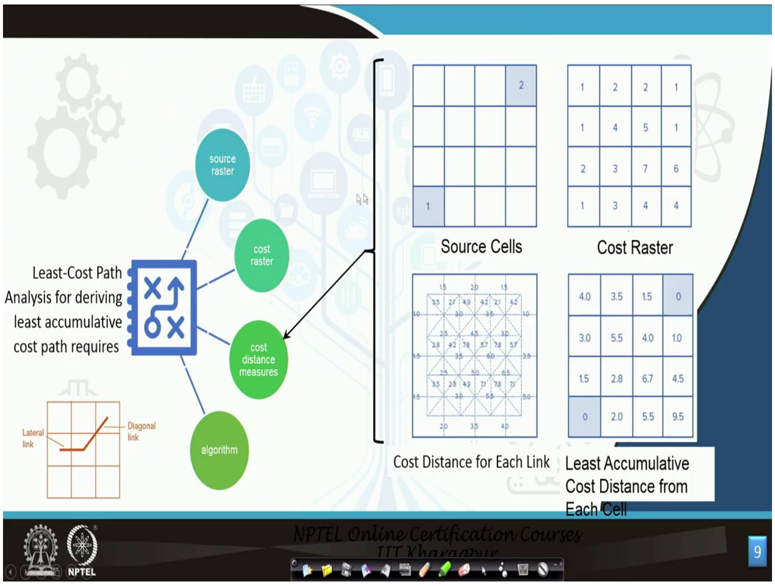


So, we you can see this particular cost raster. So, we have assumed a cost raster. So, in this case this matrix has 4 by 4 elements. So, you can see this top right 4 elements are zoomed in and I mean it has shown by this particular enlarged representation. So, the distance between the laterally linked cells are summation of the values of the particular link distance.

So, suppose this value is 2 and this value is 1. So, the lateral your cost of traversing these 2 nodes, the centroid of this pixel to the centroid of this pixel would be 2 plus 1 that is 3 divided by 2 which is 1.5, but when we are traversing the diagonal link say suppose we are traversing this diagonal link between 5 and 1. So, what we do is we sum up 1 plus 5, which works out to 6 divided by 2 is 3 into 1.414, which comes to a value of 1 4.2.

So, for this diagonal link you can see this value is 4.2. So, similarly this diagonal link has a cost value of 2.1 and similarly the lateral links all have different values. So, you know now how to calculate the cost distance of the lateral or diagonal link.

(Refer Slide Time: 21:59)

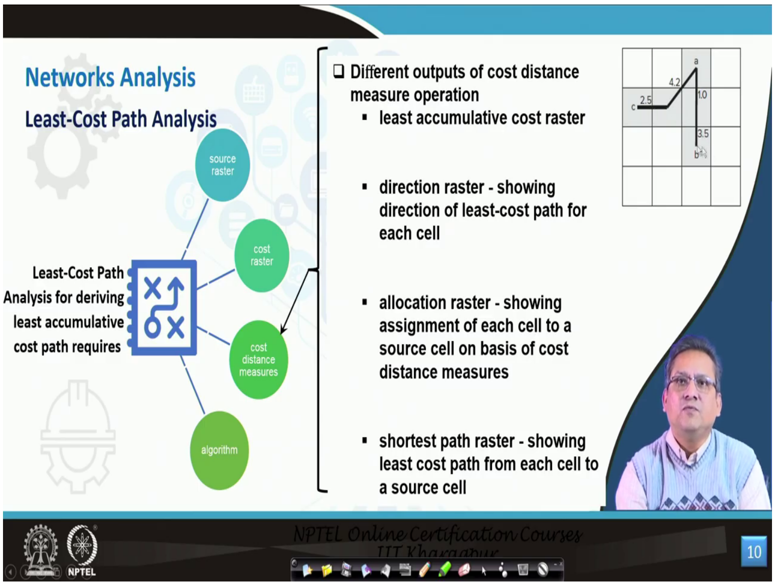


So, we can do same thing in this particular matrix. So, first we had said we had the source raster. So, the source raster we said the source and the destination would have values and the other elements would have null values. So, you can see that in this particular your matrix or the raster representation.

So, you have the cost raster that is the cost of traversing across your lateral links or through your diagonal links. So, you can work out the court cost raster and then as we had seen we had calculated the cost for this particular sector. I mean these 4 pixels; we have done this for all the pixels now. For all the elements in this given raster we have calculated the lateral costs traversing the lateral in a lateral fashion and in a diagonal fashion.

So, from there you can work out if this is your origin and destination, you can work out the distance to traverse to the next cell in a cumulative fashion. So, you can find out the route which would I mean have the minimal cost in this given network ok. So, you can either traverse like this. So, if you traverse from here to this you have a value of 6.7. I mean then the value of 4 is added to this. So, it would have a value of 10.7, but if you traverse along this line probably that will give you the least cost path.

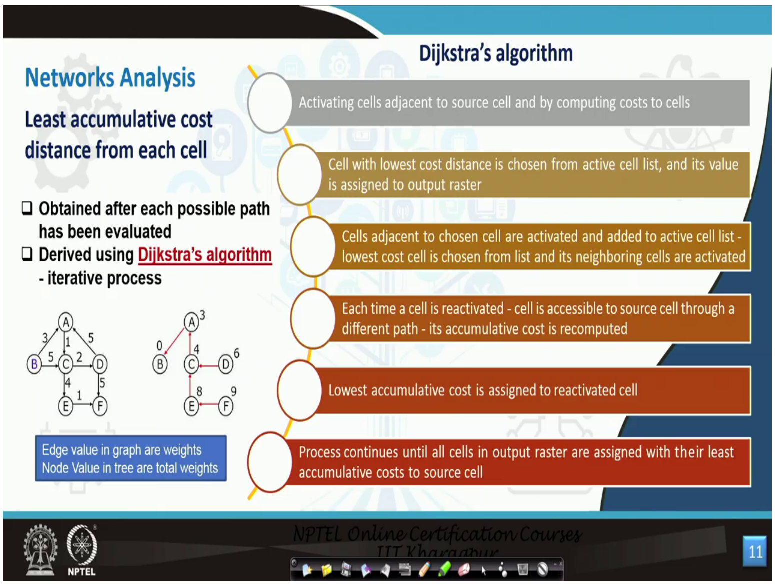
(Refer Slide Time: 23:45)



Now, we have different types of outputs of cost distance measure operations. So, the first output could be a least accumulative cost raster and the next output is the direction raster, it is showing which shows the direction of the least cost path for each of the cell. And, the third output could be a allocation raster showing the assignment of cells to the source cell on basis of the least cost distance measure.

Now, the fourth type of I mean output, that we can have from cause distance measure is the shortest path raster, which shows the least cost path from each cell to a source cell. So, from our earlier example, we mean we have this, I mean route the minimal route if we are traveling from c to a or c to b, we can traverse this route and for the minimal distance, cost distance.

(Refer Slide Time: 24:53)



Now, the least cost distance is obtained after each path is evaluated. Now, we can use a Dijkstra's algorithm and algorithm given by Dijkstra and it is an iterative algorithm, which I mean we have a method and we basically iterate it many times to come to the optimum. So, the steps involved include activating the cells adjacent to the source and we compute the cost to these cells. The second step in this method is that the cell having the lowest or the least cost distance is chosen from the neighboring nodes, I mean if you have active cell list the least cost distance is chosen and it is distance is value is assigned to the output raster.

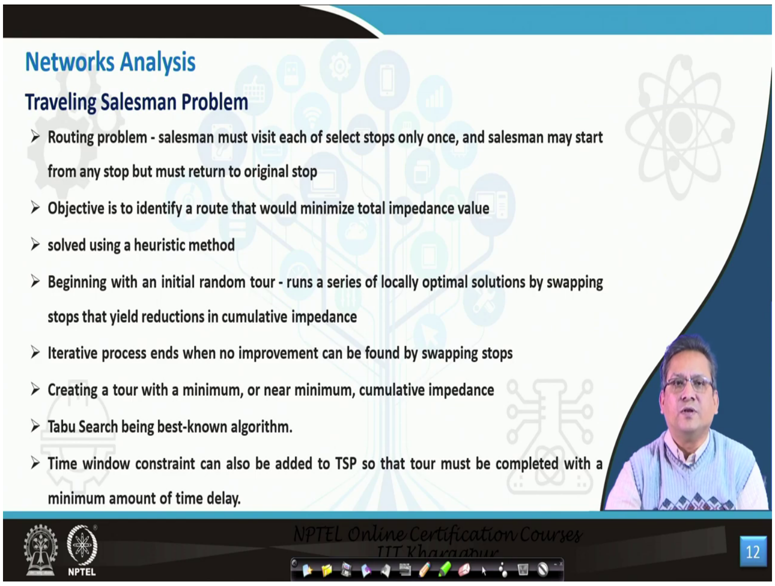
Now, the in the next step what we do is we try to find out the cells adjacent to the chosen cell they are activated and then we added to the active cell list, and the lowest cost cell would be chosen and it is neighboring cells are then activated. So, whenever we are reactivating a cell, a cell is accessible to source and what we is we I mean try to identify alternate routes, and I mean try to pick out the least cost path, and we accumulate the cost, and we re compute the weightages of the alternate possible routes and the least cost path is retained.

So, then what we do is after we have worked out the least cost path it is again assigned to the reactivated cell. So, this process is an iterative one and it would continue till your all the cells in your output raster would be assigned with the latest accumulative cost to the source cell.

Now, let us see this particular network we have this first diagram on your left wherein you have the age values or the link you can say the link impedances. So, which are weights? So, traversing from B to A the cost is 3 and from traversing from your D to A we have the cost as 5. So, your A 2 B is linked and traversing from A to B would have a cost of 3. So, what we do is we based on the directionalities, we try to identify the cost as well as the direction and we find out the I mean aggregated cost or the accumulative cost.

So, we do this for the entire network and we can see that which is the shortest possible route between say link D and link F.

(Refer Slide Time: 28:27)



Now, talking about another method of transportation problem when we are to visit say multiple nodes in a given network, we have an algorithm which is known or traveling salesman's algorithm. So, it is a route routing problem wherein I mean the salesman would start from 1 node and he would traverse through all the nodes that are selected in a given network, but he has to return to the original node.

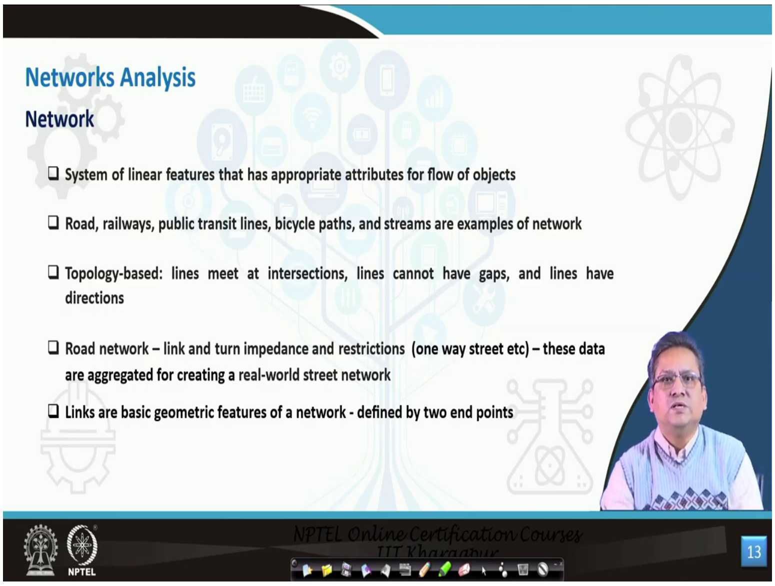
So, our objective of implementing this traveling salesman's problem is that we want to identify the route, varying the there would be a minimum total impedance value or minimum total cost value. Now, it is a heuristic method this traveling salesman's problems are solution is a heuristic method and we begin the initial search using a random tool and runs this process runs a series of locally optimal solution. And, we swap the stops and try to find out if there is a reduction in the cumulative impedance.

Now, this iterative process would end, when there is no improvement there is no change in the value of your cumulative impedance, by swapping the stops I mean in a in our earlier step we have been swapping the stops. So, whenever you see that between 2 successive iterations there is no improvement in terms of the impedance value or the cost value, we can stop the iteration and we can create a tool with a minimum or a near minimum cumulative impedance, I mean always it may not give you the best solution.

So, one of the methods which gives you the best possible results is a Tabu search algorithm for I mean finding out connectivity between n number of nodes. The time window constraint can also be added that, you want to complete this node within some amount of time or within a time delay certain amount of time delay, minimum amount of time delay, you can also apportion time window constraint.

So, then what it will do; it will do is you it will try to pick out the links, which varying the travel time are the least. So, that it would be able to complete this entire tour or travel within that stipulated trying frame.

(Refer Slide Time: 31:17)

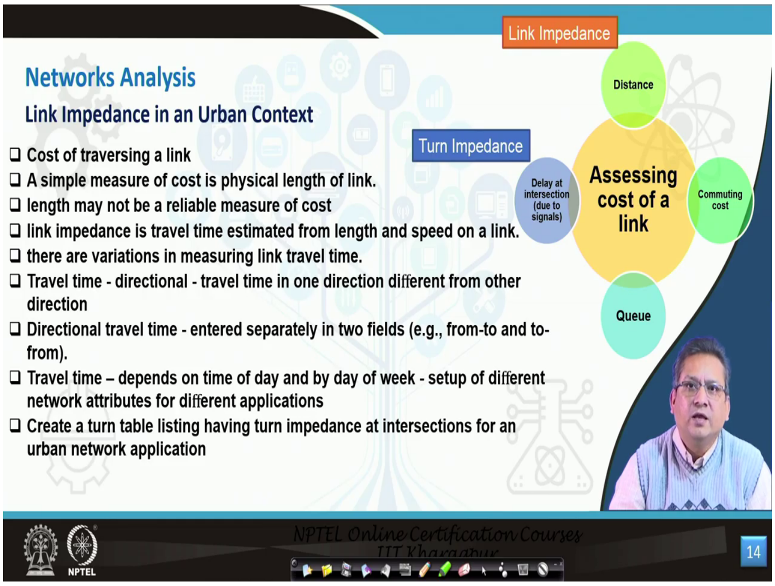


Now, we have a system of linear features with appropriate attributes for flow of objects and these linear features or networks could be a bicycle track, or a network stream of a river or drainage course it could be a railway line, or some public transit corridor, or a road. And, we can create a topology wherein the lines would meet at the intersections and we the lines should not have any gaps. Otherwise, the errors would show up in your analysis, because connectivity is a very important thing in this particular network.

So, whenever you are trying to digitize your networks for doing a network analysis you have to see that, there are no overshoots or undershoots while you are digitizing. So, you can also apportion direction, some of the link could be a one way street. So, you can apportion the turn impedance as well because you know, if you want to travel a turn take a right turn or a left turn, you would have to I mean negotiate a lot of vehicles through I mean in that during that process of turning or if you are taken the u turn.

So, that will delay your travel and it will impede your movement. So, I mean we can have an impedance not only for the links, but we can also have an impedance for the nodes as well. So, these data would be aggregated and we create a real world street networks. So, these geometries again are defined by two points two end points.

(Refer Slide Time: 33:13)



So, we can I mean find out the link impedance in a urban in an urban context and we can find out the cost of traversing a link. So, it could be a simple measure like finding out the length, I mean accounting for this cost in terms of the physical length, which is a reliable measure, I mean you can measure that through GPS or your software would generate those points, if you are digitizing lines on a geo referenced image say a satellite high resolution image or a topo sheet.

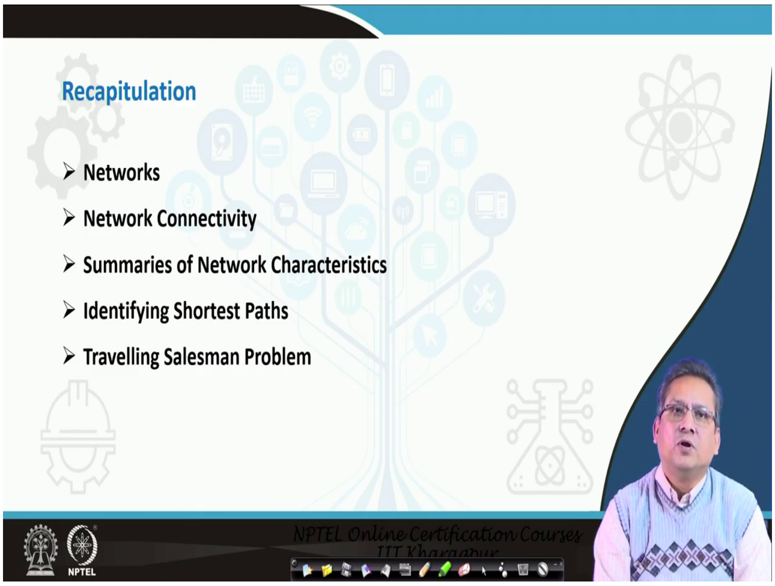
So, you would be able to find out the exact distances between these nodes. So, this length could be a measure of reliable measure of cost. I mean we can have the speed on a link and it could be used to calculate the link impedance from the length as well as the speed on the link. So, there could be different types of link travel time. So, I mean we can find out the directional travel time, I mean your if you travel 2 nodes in different direction whether the travel time is same or different it could be different in most of the cases.

So, we can enter this data I mean the directional travel time separately. So, we have when we are having the topology of the networks, we have those columns like from node and to node. So, we can create two columns like from and to and to and from. So, we can identify the nodes the origin node and the destination node and we can have a value, directional travel time value and from the other end that is from the 2 node to the from node, we can have another travel time value.

So, this travel time would depend on the of the day or the week or the season or in different locations. So, I mean we can include this travel time and I mean in a different context wherein we may have I mean a one way corridor or a two way street. So, we can attribute this travel time. So, when we are doing some network analysis like we had talked about the traveling salesman's, I mean your algorithm or we were talking about Dijkstra's algorithm.

So, there we not only include your link impedance, but we also include the term impedance in doing the calculations. So, you can see that we can assess the cost of a link using a distance function, which would give you the link impedance, we can find out the commuting cost, we can find out the queue length, we can find out the delay at intersection probably due to signals, which will give us the turn impedance. So, all these can be accounted and aggregated into a cost function for accessing a particular link.

(Refer Slide Time: 36:39)



So, a recapitulation of what we had covered in this particular lecture we had talked about networks. We had talked about network connectivity, we had seen some indicators for summarizing network characteristics, we have identified the shortest paths using a Dijkstra's algorithm. And, then we had seen a traveling salesman's problem and how it can be implemented and it can be applied in an urban context.

So, most of the software's like qgis, arcinfo and other software's have these algorithms built in, and you can devise your own I mean application based on either the shortest path or using a traveling salesman's approach.

So, thank you.