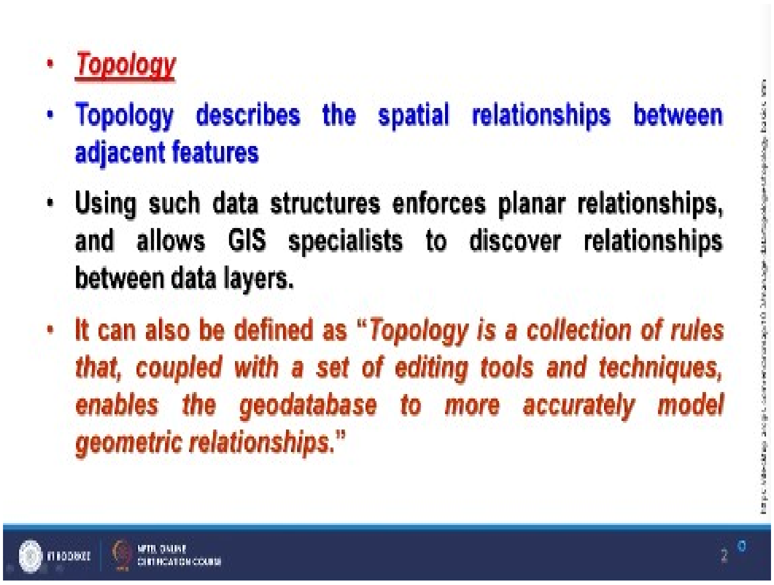
Geographic Information Systems
Prof. A.K. Saraf
Department of Earth Sciences
Indian Institute of Technology – Roorkee

Lecture – 4
Concept of Topology

Hello everyone! and welcome to our new discussion of this Geographic Information Systems course and in this, we are going to discuss the concept of topology. Topology basically a branch of mathematics and which has been implemented into the GIS and it is very essential part of vector data. Otherwise without topology, we will have lot of difficulties. So, initial development of GIS did not have this concept of topology and there were lot of difficulties was there.

But later on, the topology was brought into GIS very successfully. Earlier also in the softwares like in Arc-Info or others PC Arc-info, we used to first digitize the vector data especially line or polygon data and then we used to develop the topology and there used to be lot of errors and their corrections. Nowadays the softwares you know, simultaneously build the topology and therefore errors and everything is resolved at the time of digitization.

(Refer Slide Time: 01:46)

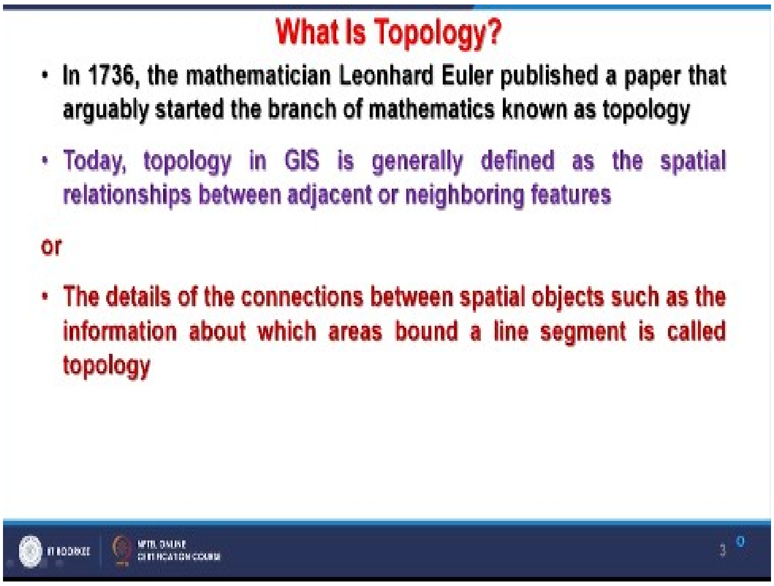


So, basically topology describes the spatial relationships between adjacent features. Adjacent features may be another polygon or different polygons around a polygon or maybe line features or even point features. But mainly the topology in GIS is developed for the polygons because there it is the real challenge. For line or points; it is not difficult at all to develop a topology. So, for topology specifically, it is for polygon data.

Now this is as you know a spatial relationship between adjacent features. So that means this is how it is stores; what is inside and what is outside and also neighborhood information. So, it stores in that way. Now, as you know that everything is stored in GIS as a planar form and therefore planar relationship. So, using this topological concept and vector data especially for like polygon data; this data structures basically enforces planar relationships and allows GIS specialists to discover relationship between data layers.

So not only in one layer; within one layer, you can have knowledge of relationship but also different layers, that information can also be retrieved into GIS or can be analyzed in GIS. So, topology is basically a collection of rules that occupied with a set of editing tools and techniques enables the geodatabase to more accurately model geometrical relationships so geometric here means basically the relationship and since we are talking geographic so, it should be in a geographic model also.

(Refer Slide Time: 03:53)



Now what is basically topology? As I mentioned in the beginning that it is a branch of mathematics and a Leonhard Euler who published a paper in 1736 and after that this branch of mathematics got establishment and it is now known as topology. And topology today in GIS is generally defined as the spatial relationship between adjacent or neighboring features.

Also, in some literature you may find little different definitions. So, I am giving 2-3 examples of such definitions.

The details of connections between a spatial object such as information about which areas bound a line segment is called topology. So you know, as I mentioned that mainly concern of topology is with the polygon data. So in that sense, it is very much required to build a topology or also it can be defined in simpler terms that topology stores the relationships of one spatial object with respect to another.

So, this basically stores the relationships. Now let me give you one example about this topology. In our day to day life, we eat noodles or spaghetti and if suppose there is a plate full of spaghetti or noodles. If you would like to pick one single noodle, you can do it. Though they are in the group but there is no connection between the single noodle with all other noodles.

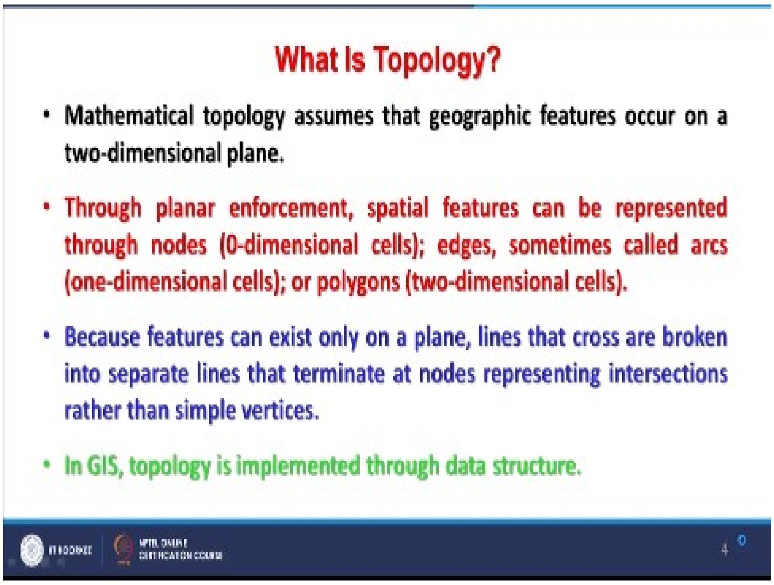
So, every object or every noodle on that plate is lying sort of in isolation; no relationship, no connections between them. And that concept of model is also called a spaghetti model where these neighborhood relationship or adjacent features informations cannot be stored at all. Every object is lying in isolation. So, that is called a spaghetti model. Spaghetti in some countries and noodles are also called Spaghetti especially in Italy.

Now, there are softwares which do not understand or where concept of topology has not been implemented and therefore, there are lots of difficulties if you go for digitizing a polygon. Let me give you an example like Coreldraw; very popular graphics software. Where people if those who do not want to use GIS and do not know about GIS, they are still using Coreldraw.

And what happens in Coreldraw that two adjacent polygons; if you want to draw two polygons and there is a common boundary that means that common boundary will be drawn or digitized twice. And no human can draw exactly the same boundary twice in the same manner because intermediate nodes might be very different and therefore if you zoom this common boundary part, you would notice that there might be some gaps or overlapping.

So, you can say there is a dispute between features if I am having any spaghetti model without having topology but if I construct a topology between say these two polygons. For the common boundary, it would be digitized or stored in the system only once. Both polygons can share very easily without creating any kind of dispute between them and this is what the topology does. It creates the relationship between two adjacent objects; features. Here our features are polygons.

(Refer Slide Time: 08:03)

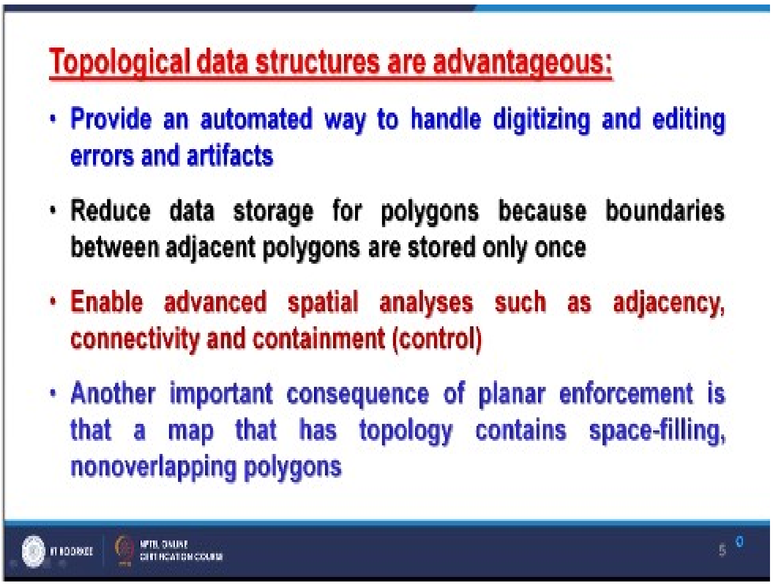


So mathematically, basically topology assumes that geographic features occur on a two dimensional plane and through planar enforcement because we think in terms of a flat or plane or a map. So instead of a 3D surface, we represent that 3D surface in 2D that is why through planar enforcement; special features can be represented through a node that is point. So, that is why zero dimensional cells. Edges (arc), line are one dimensional cells and polygon which are two dimensional cells or vector entities.

And because features can exist only on a plane therefore the line that cross are broken into separate lines and that terminate at nodes representing intersection rather than a simple vertices. So, if there are two lines and they are crossing each other then wherever they are crossing, there is a new node which has to be created. Otherwise these two lines will not know adjacent features or neighborhood information at all.

So, this is what the topology basically does that it creates the information about the neighborhood or adjacent features. So, in GIS topology is implemented through the data structure and this is as I have been mentioning mainly for vector polygon data.

(Refer Slide Time: 09:35)



What are the advantages if we are having topology buildup already there for our polygon features that it is provides automated way to handle digitizing and editing errors and artifacts. As I said in the beginning that in the earlier versions of these softwares, the topology used to be built later on. And when we used to go for building topology, we used to have lot of errors and then artifacts and lot of time used to be spent to correct them.

But whereas now, topology is built while digitizing process is going on simultaneously. So, any error is there, it will be corrected immediately because as you know, this is one of the sort of rule of GIS is that error propagates in GIS. So, whenever you find an error in your data, correct it immediately. Do not leave for others to correct because error propagates in GIS. A minor error today in my data can become a very big error in my results.

And therefore it has to be corrected there itself whenever first time it has been detected. So, as soon as one realizes there is error, immediately should be corrected. Another advantage of a topology buildup is; it loses the data storage because adjacent 2 polygons boundaries will be stored only once that means that the common boundary between 2 polygons is stored once if topology has been built.

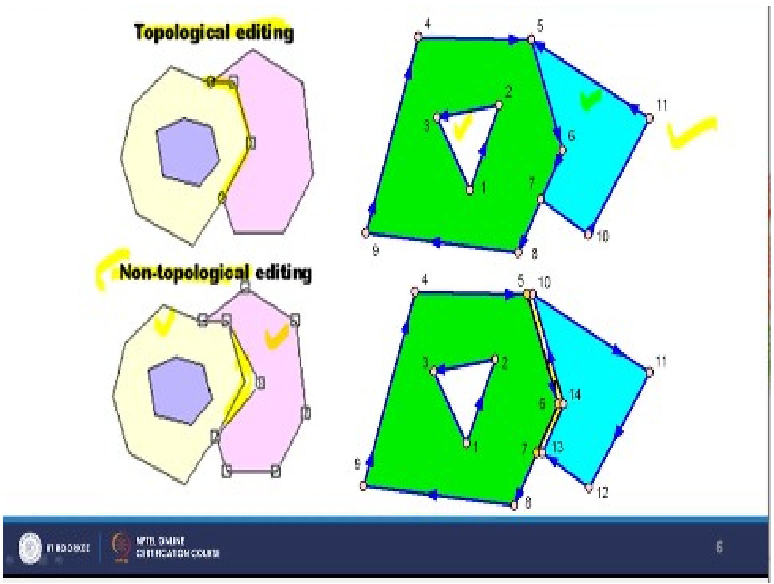
But if it is in a non-topological model like spaghetti model where the common boundary would be stored twice and therefore it would require more storage space. Do not think that we are only going to handle only 2 polygons or 10 polygons. In a real projects; a map can have thousands or thousands of polygons and think that how much unnecessary space it will require and the errors which we will have all the time. So, that will become your inherent errors.

Now, this topology structures or topology models enables advanced spatial analysis. So, once the topology has been built then lot of analysis like adjacency, connectivity, containment, control; all these can be done very easily on GIS platform. Containment word; nowadays we are hearing a lot during this COVID pandemic. So, even for that containment zones when the civil administration in any city is deciding, they too have now started using GIS systems to define what is the containment zone?

Of course, they have to use other information also. And another important consequence of developing topology of planar enforcement is that a map that has topology contains space filling non-overlapping polygons. Recall the discussion when I said that a common boundary between 2 polygons if it is digitized twice it may have sometime overlapping and sometimes gaps. These are called artifacts.

We never intended that there will be some extra polygons and that will create hell of problems by doing such analysis like adjacency, connectivity or containment etcetera. So, when we build the topology, all these errors are removed and there will not be any non-overlapping space filling polygons. So if my map should have 50 polygons, it would have 50 if I have done properly the digitization part. Let me give you an example.

(Refer Slide Time: 13:46)

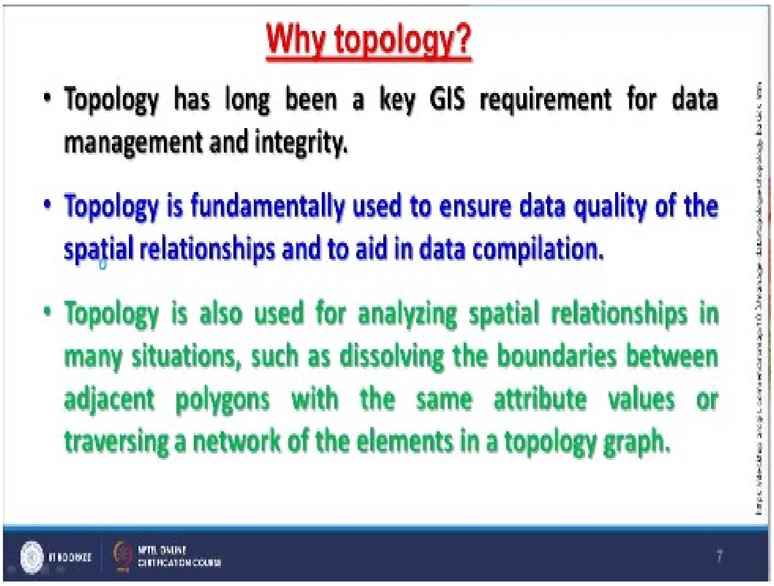


You know the top left one is showing 3 polygons and a common boundary between these two polygons is also shown. So, this is topological editing or things are in topological domain or model. If we go for non-topological modeling then these are the gaps which have been created because the common boundary between yellow polygon and the pink polygon has been digitized twice and while digitizing twice, there is a gap which has been created.

And this gap will become in our layer as one new polygon which was never intended and this new polygon will also have an area and also parameter which is not as per the real data set. So, therefore this is ultimately an error. So, same example is also shown here but little differently that you are again having 3 polygons; white one, green one and this cyan color. In this non-topological one; you are having a big gap between these 2 polygons.

So, the line though it has been digitized perfectly alright but there is a gap. And therefore lot of analysis cannot be performed on non-topological model data when they are in a spaghetti model.

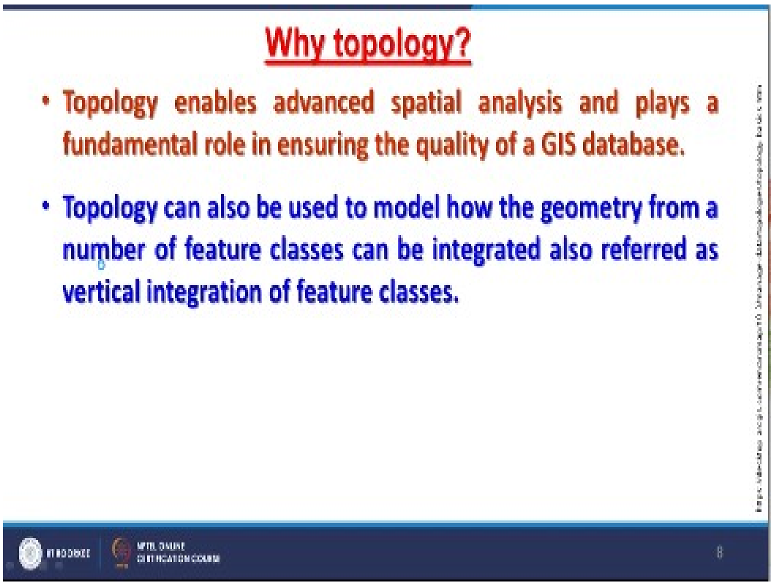
(Refer Slide Time: 15:14)



So, topology has long been a key GIS requirement for data management and integrity. Initially, it was not there. Later on, it started coming but we had to build later on but now it is built basically automatically when we do the digitization work. And topology is used basically to ensure data quality because errors are minimized and to maintain the spatial relationship and it also aid in data compilation.

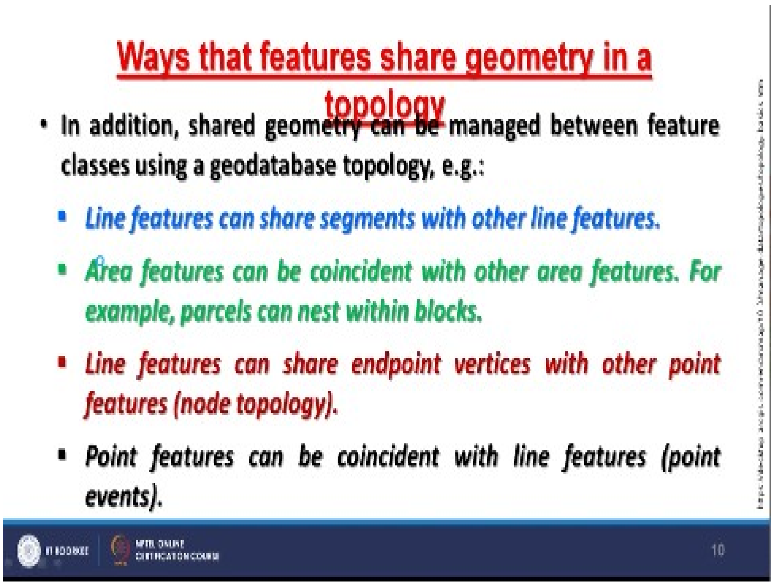
Topology is also used for analyzing spatial relationships in many situations such as dissolving boundaries between adjacent polygons with the same attribute values or traversing a network of elements in a topology graph. So that way, we can also do it.

(Refer Slide Time: 16:16)



Now, we have discussed this point little bit that topology enables advanced spatial analysis and plays a fundamental role in ensuring the quality of GIS database. And topology can also be used to model how the geometry from a number of features classes can be integrated; also referred as vertical integration of feature classes that means topology you can have for different layers also, not within one layer but different layers.

(Refer Slide Time: 16:45)

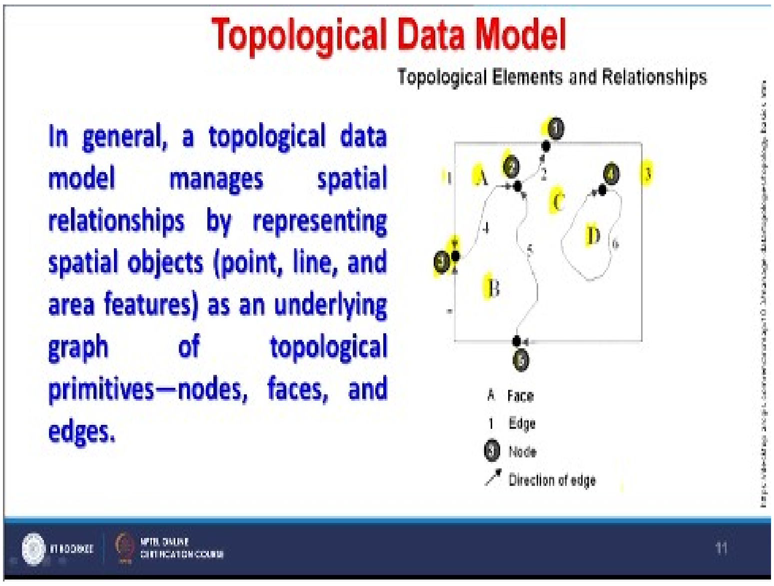


So how this is done? Now this part we will see that the ways the features share the geometry in a topological model is that; features share geometry within a topology among adjacent features; 2 polygons. Example, I have already given. Area features can share boundaries. Line features can share end points; that are the edge node topology but our concern for topology is mainly for what I should say for polygons.

In addition, shared geometry can be merged between features classes into geodatabase. For example, line features can share segments with other line features of different layers. Also, features can be coincident with other area features or other polygon. And then line features can share end point vertices with other point features. So, not only one type of feature but line can be shared with point and so on so forth.

So, all these permutation and combinations are possible once you are having topology. And point features can coincident with the line features also.

(Refer Slide Time: 18:00)

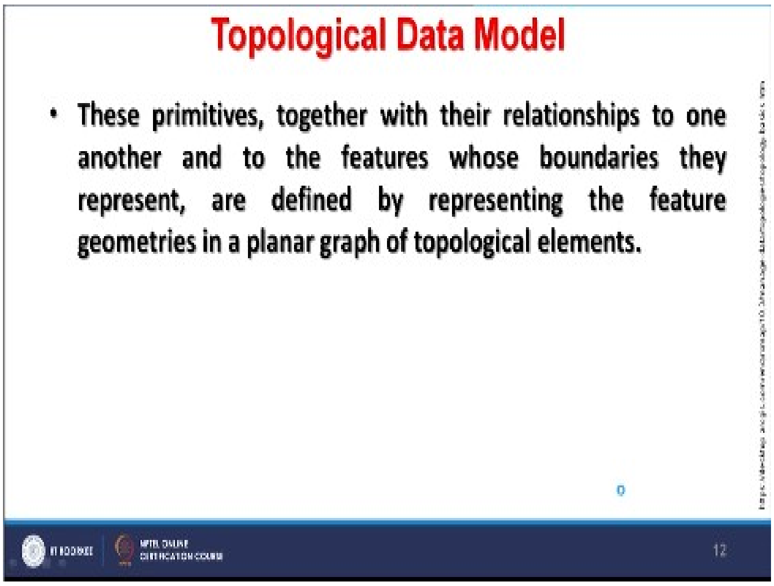


Now, let us see a few examples of the way the data is represented and stored in the system. So, in general a topological data model manages spatial relationship by representing spatial object that is point, line and area features as an underlying graph of topological primitive-node, faces and edges. So what are these terms? So, we will see that these like A, B, C; all we are have nothing but the polygon or faces. Then here 1, 2, 3, 4; these are nodes.

And you would notice that even for outside; there is a line like here and this is coming here. So, begin node and end node is the same here. So outside also; there is a polygon which is marked here or line which is marked here. So, you are having edges like (1, 3), (4, 5), (2, 6). You are having node and of course since it is a line data. So, you will have the direction of edges as well. That direction of edge or line will decide who is on the left or who is on the right.

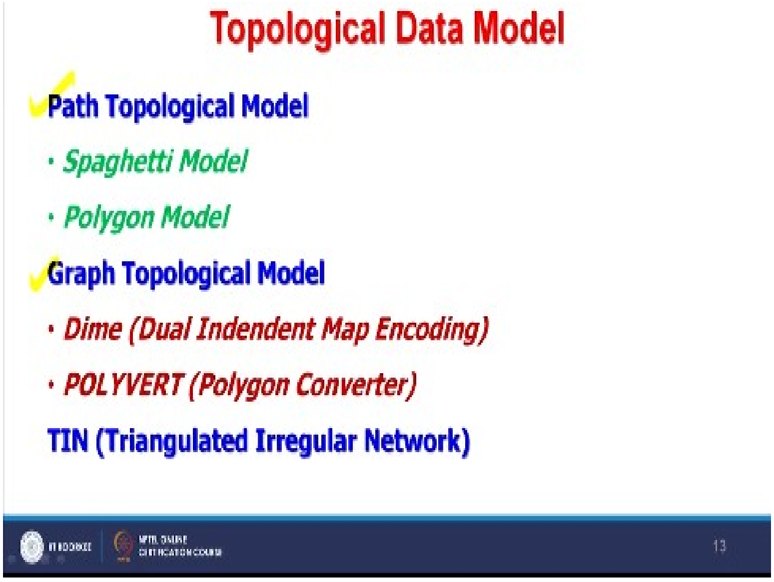
So, when we say left bank or right bank; we decide based on the flow direction. So, if it is flowing from A to B and if we look towards the B then whatever on right hand side is the right bank and whatever on the left hand is the left bank because the direction of edge or line will give us the information who is on right and who is on the left side.

(Refer Slide Time: 19:54)



Computer stores all this information so that more advanced analysis can be performed using GIS. And these primitives like nodes and other things together with their relationship to one another and to the feature whose boundary they represent, are defined by representing feature geometries in a planar graph of topological elements. Now, different types of models are there. Two major models; I have already mentioned one is spaghetti which is non-topological model. Another one is topological model.

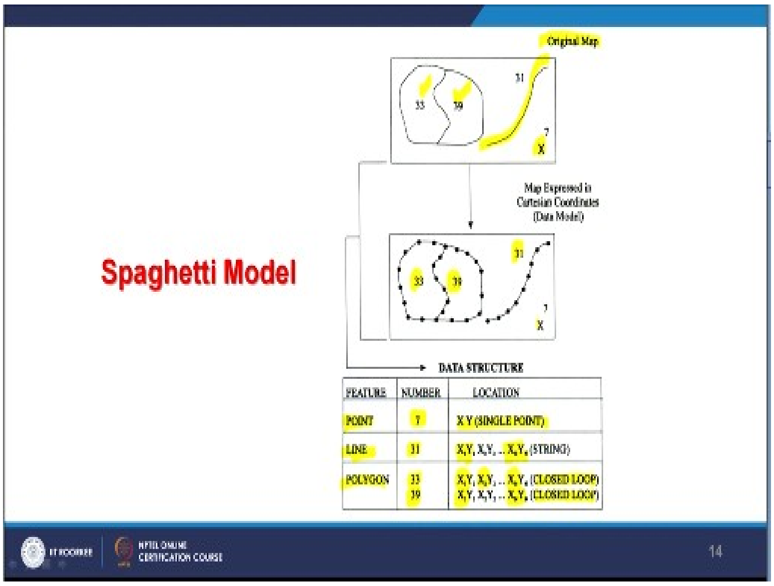
(Refer Slide Time: 20:38)



Now, within topological model there are different types like path topological model and this Spaghetti model will be considered under this path topological model. Polygon model also. So, path topological model and another type is the graph topological model. There is a Dime-Dual Independent Map Encoding and then POLYVERT; there is another graph topological model which is a polygon converter and also one different type of topological model which is TIN- Triangulated Irregular Network.

TIN is a different type of data representation model also like vector-raster so there is TIN. While discussing in the previous lecture about different types of data which are handled in GIS, I have touched upon this TIN. So, TIN is also a type of topological model. None of these are useful except for TIN; other these path and graph are not useful in GIS because we are having a better one which is polygon topological model so that we will discuss.

(Refer Slide Time: 21:41)

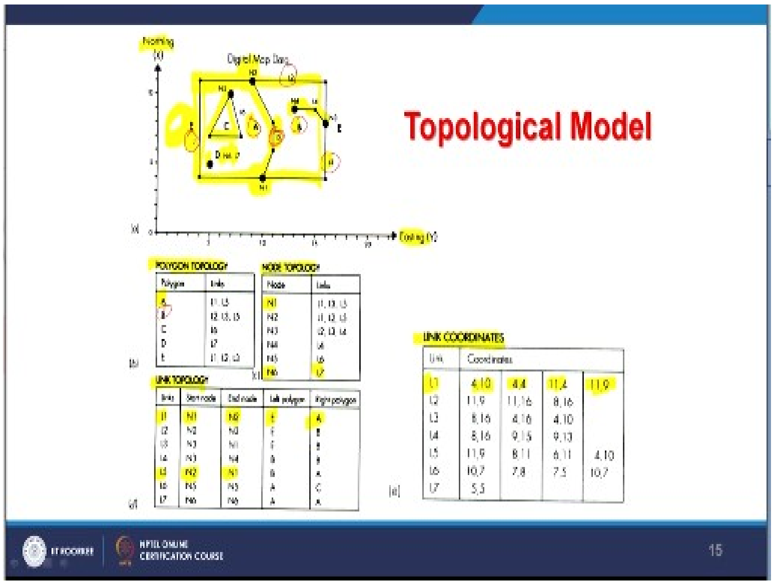


Now, before that we take the example and then we will compare the topological model with the Spaghetti. So, Spaghetti model is a basically non-topological model or we put in a category of path topological model because say just line information. So here if this is the map which is marked as original map. Three different vector entities are present on this map. You are having point data. You are having the line data and you are having the polygons also.

Now if I store this one in the computer; this is how it will be storing that you are having a node for point; you are having many nodes for line and again you are having many nodes for the polygon except that begin node and end node are the same. Now, how it will be stored further in the form of tables that you will have point information. Some code is given. ID is given like here 7 and a set of coordinate that is X and Y.

Then in case of line; here number is shown 31 that ID. And then you are having X1, Y1 to Xn, Yn; a string of points which are connected through a line and then polygon. Now this is shown as 33 and 39; now in order to store 33 polygons, all these points or strings of coordinates are stored. Similarly for 39; X1, Y1 to Xn, Yn are stored. And please note closed loops or closed polyline. Now in this model; the problem between the common boundaries that is not stored, the outside polygon area is also not stored. So, the relationship between adjacent objects features is not stored in Spaghetti model.

(Refer Slide Time: 23:54)



When we compare this Spaghetti model with topological model having almost the same example like here also, we are having the point data. Here we are having the line data. Here we are having polygon data like C is one polygon; A is also one polygon and this B is also one polygon. And outside this is also one polygon which is marked here as E. So, compared to Spaghetti there is one table; now here we are having 4 tables to store the information.

So, though more space will be required but the data; your polygon data, vector or line data, everything will remain intact and quality data that means it would have minimum errors. At least errors due to non-topology will not be there because topology has been constructed and those errors which are due to non-topological models will not be there completely in the system. So, let us see in further details. First, we will see the node topological table.

So each node and then links are stored like if I take the example of node 1. This is node 1 and node 1 is made from 3 links; L1, L3, L5. So, node 1 is made from 3 links L1, L3 and L5. Similarly if I take example of node 6 that is only having link L7. So, node 6 is here and then you are having link L7. This is L7. So, this is node 6 and L7.

Now, the next table will store the information about the link topology. I will come later on polygon topology. Now, this link topology will store information about begin node and end node because this is link between different for lines or polygons. So, it has to store this information. So, if I take again example of L1 link which we have discussed while discussing node topology table. Now, L1 starts with node N1 and ends with node N2.

So, this is the L1; it starts with N1 and it ends with N2. N1 is here; N2 is there so that means the link is this much. Similarly, if I take another one like link 5; so link 5 is this one and which is having again node N2 and N1 so both this one. Further in this, it is stored like in case of link 1; left polygon. So what is on the left side in case of L1 link which is this outside link? So, what is on the left side is E polygon and what is on the right side is A polygon.

Now, we come to the link coordinates. So, each node coordinate are also stored. So, we are having one northern easting example that is geographic coordinates in UTM projection. We can have you know latitude-longitude. So for L1; this link L1 is having N1 node which has got the coordinate then this node on the corner, another intermediate or you know that node and then end node 11. 9 and last is the polygon topology.

Now, polygon topology like if I say polygon A is here and polygon A is made from link L1 and L5. So, this is L1 and this is L5. Similarly, polygon B here is made from link 2, link 3 and link 5. Now this link 5 is the common boundary between 2 polygons A and B but it has been stored only once. And this is what typically happens in case of polygon topological model.

It is not possible with the Spaghetti path topological model. Every detail is stored in this model and that is why we can perform analysis about adjacent features not within that particular layer but across the layers also. So, that is why topology must be prepared, must be generated for all vector entities. And this brings to the end of this discussion. Thank you very much.