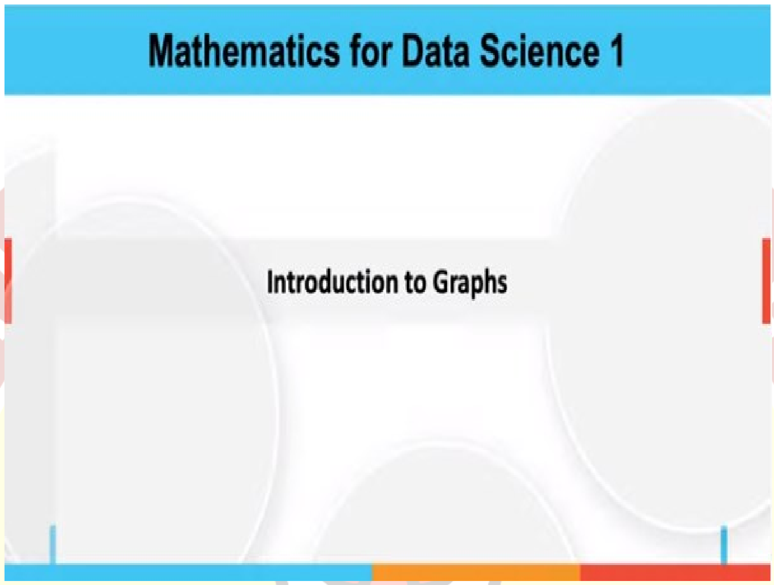
IIT Madras

ONLINE DEGREE

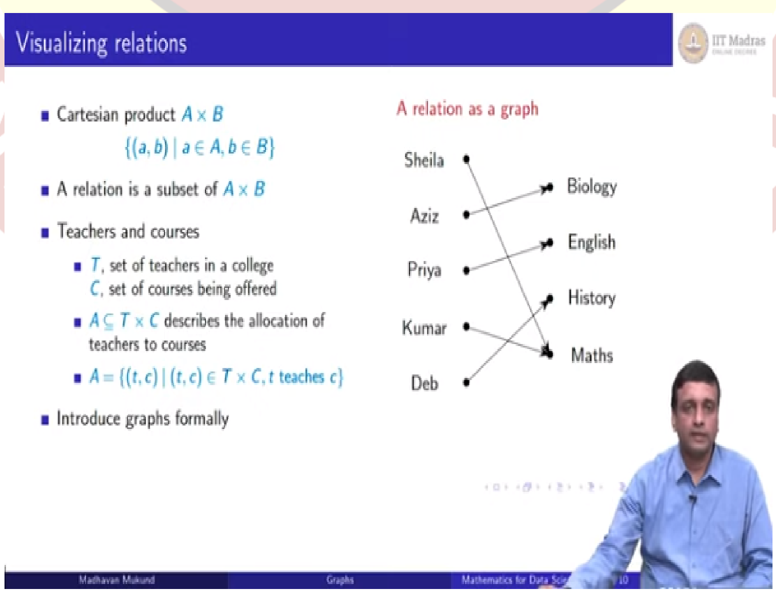
Mathematics for Data Science 1
Professor Madhavan Mukund
Indian Institute of Technology, Madras
Lecture 59
Introduction to Graphs

(Refer Slide Time: 0:09)



So, for the next few weeks we will be looking at graphs, so this will be the concluding section of this maths course, so for the next three weeks we will looking at graph.

(Refer Slide Time: 0:24)



So, we saw graphs in the first week when we were talking about relations, we said that we can take a relation, so what is the relation? We take two set they could be the same set, we take the Cartesian product all pairs. So, we take $A \times B$ and then we take some subset of that Cartesian product, so some pairs, out of the total set of pairs and we say that these pairs are related.

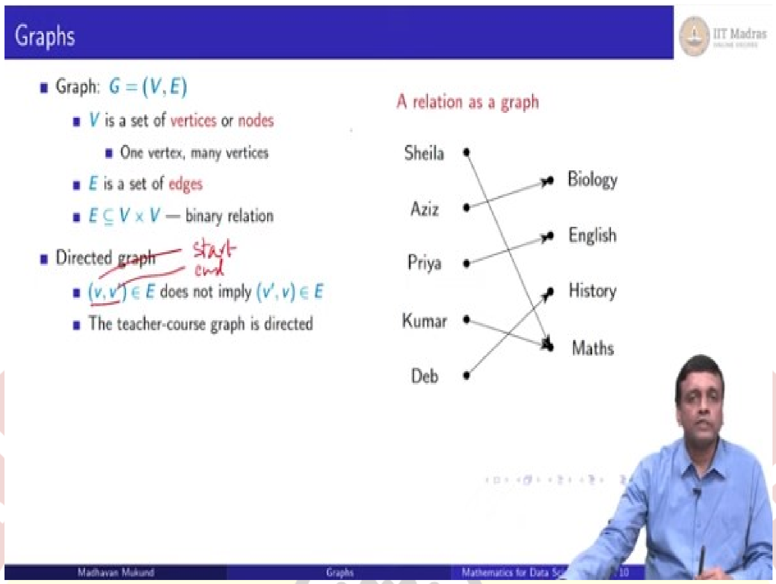
And then we said that we could visualize this, so for example, supposing our set A is a set of teachers, so let us call it T and the set C is set of courses that are being offered in the current semester, then we could have a relation which captures, which teachers are teaching which course.

So, we have $T \times C$ as a set of all possible pairs where the first element is the teacher and the second element is a course and then this relation A, which is a kind of course allocation relation describes how teachers are assigned to courses in the current semester. So, what we said was we could draw this relation as a pictorial form, so we could create these nodes or dots, representing each element in our set.

So since, there are two different types of sets, the set of features instead of courses we write them in two columns like this, so we have 5 teachers and we have 4 courses and whenever a teacher is teaching a course, we connect that teacher to the corresponding course node through an arrow, so in this case, since there are 5 teachers and 4 courses, clearly there must be at least one course which has been taught by 2 different teachers in this case you can see that maths is being taught by Sheila and Kumar.

So, what we are going to do in the next three weeks is to look at this picture that we have drawn earlier just as visualization of a relation, we are going to look at these pictures, these pictures are called graphs and we are going to formally analyze what we can do with these graphs.

(Refer Slide Time: 2:19)



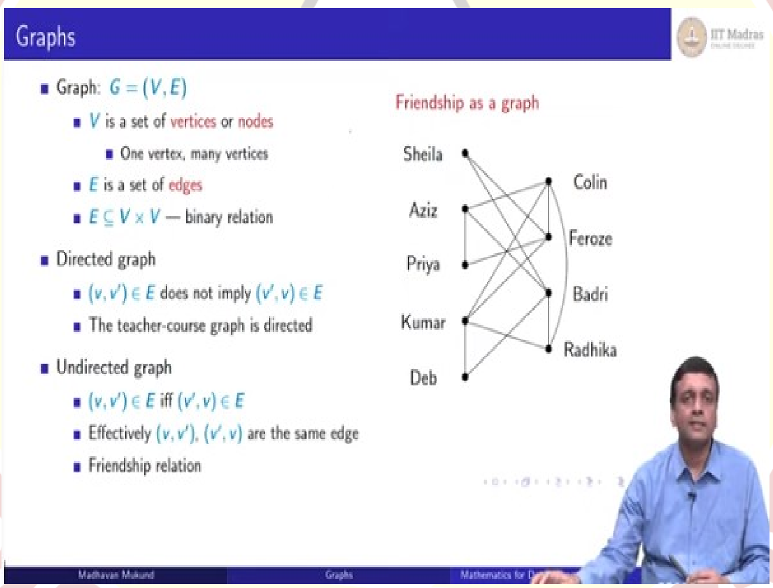
So, to begin with a graph consists of a set of nodes or vertices and edges between them. So typically, a graph therefore, has two components a set of vertices and a set of edges, so vertices is the plural of vertex. So, we have one vertex, many vertices, so we use interchangeably, either the node or vertex, as a name of for these elements and then what edges do is that they connect these vertices. So, notice in this graph that we had drawn before we had earlier two set, we had the set of teachers and we had the set of courses, and we were taking a relation which is a subset of $T \times C$.

But once we put it into the graph we lose or we do not really care about the distinction between T and C, T and C together form the set V of vertices of the corresponding graph, so now there are nine vertices, there is no real separation between the five that came from the teacher set and the four that came from the course set and then the edges were those that the original relations represented namely, those teachers which are teaching the courses.

But in general, E is just a binary relation on the vertices so, it connect some vertices to some other vertices. So, this graph, for example, has a direction, right a teacher teaches a course, we do not have a corresponding edge from a course back to a teacher. So, if v, v′ is an edge it does not necessarily mean that v′, v is an edge in this particular relation it does not even make sense for v′v to be an edge but there could be other relations where it does make sense as we will see.

So, this kind of a graph is called a directed graph, so we have an order, we have a starting vertex for each edge and we have an ending vertex for each edge and you are suppose to go from the start to the end, you cannot go backwards, so think of it like a one way road. So, there is a one-way road from the start vertex to the end vertex, so every edge is of this form, right, it is a pair, so there is a start and there is end, so this is how you should think of an edge. So pictorially we just draw it as a line with an arrow but mathematically, it is an element of $E \times E$ so it is a pair, so the teacher course graph is directed.

(Refer Slide Time: 4:40)

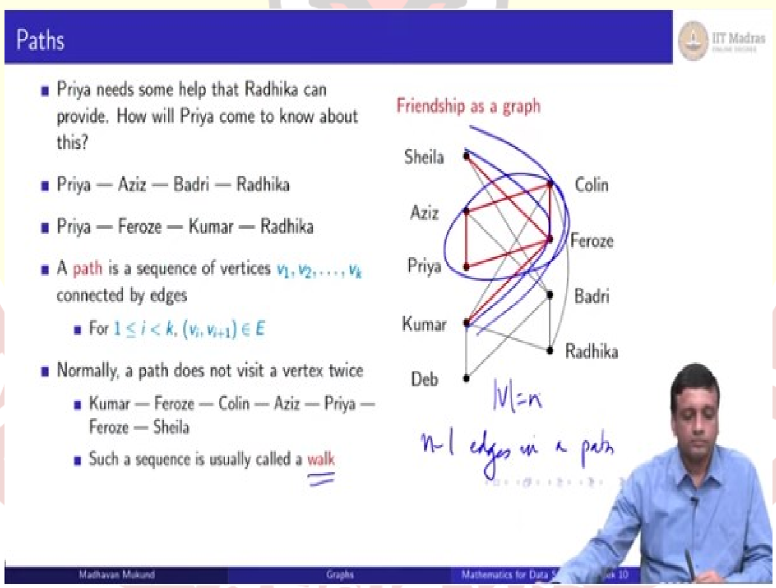


On the other hand, supposing we are just looking at a bunch of people and we are trying to capture, which of them are friends of each other. So, this now becomes an undirected graphs so presumably if Sheila is a friend of Badri, then Badri is also a friend of Sheila, because friendship is not a one directional thing, you cannot be my friend if I am not your friend.

So, in this case, we do not have an arrow, we just have pairs, which represent in some sense a symmetric relation, if you remember we talked about symmetric relations. If a,b is in the relation, then b,a is also in the relation. So, this is what we have here, we have a symmetric graph, which says that if v v′is an edge then v′v is also an edge.

So, here we have seen that there are two types of graphs that we can have, both of them are defined in terms of vertices and edges. In the first case, if the edge relation is not symmetric then we specifically record the direction and we call it a directed graph. If the edge relation is symmetric and this happens very often, so it is useful to think of this as a separate case, of course, we could always represent this by having edges in both directions, nothing to stop us from creating an edge saying that Sheila is a friend of Badri and Badri is a friend of Sheila and having an edge going from Sheila to Badri and one going back, but is much more convenient to just draw a single edge with no arrows indicating that this is symmetric. So, since this is an important special case, this is usually treated separately in graphs and it is called an undirected graph.

(Refer Slide Time: 5:58)



So, what can we do with this graph, other than just visualizing the pictorial relationship between the people? Supposing, we have a situation where Priya needs some help and it turns out that actually Radhika is in a position to provide this help, but as you can see from the graph, there is no direct connection between Priya and Radhika, so Priya and Radhika are not friends, so Priya may not even be aware of the fact that Radhika can be a source of help.

So, what do we do in real life? In real life when we have a problem we reach out to our friends and we say I have a problem do you know somebody who could help me. So, in this case Priya could reach out to her friends who are in this particular graph Aziz and Feroze and then one of them, presumably can reach out to their friends or both of them and so on and eventually somebody will hit upon Radhika. So, one possible scenario is that Priya told Aziz about her problem, Aziz told Badri about this problem that Priya has and Badri says 'Oh, I know that Radhika can solve this problem, so let me put Priya in touch with Radhika'.

So, what we have constructed through the friend relation is a path. It is a sequence, connecting Priya to Radhika even though there is no direct relationship between Priya and Radhika. On the other hand, if Priya had asked Feroze, then Feroze might have propagated this question to Kumar and independently Kumar is also a friend of Radhika, so Kumar would have found out the same thing and told Feroze for why does not Priya contact Radhika. So, this is something that you can do once you have the graphical representation of the relationship, you can look for these long-distance connections, which are called paths.

So formally in a graph, a path is a sequence of vertices, you have a starting vertex, say v1 and an ending vertex say vk and what you want to do is go from v1 to vk by following a sequence of connected edges. So, v1, v2 should be an edge. So, in this case v2 will be a friend of v1, but then v2, v3 is also an edge, so v3 is a friend of v2 and so on. So, you have k minus 1 edge connecting these k vertices, so you can go from 1 to k following these edges and this is called a path.

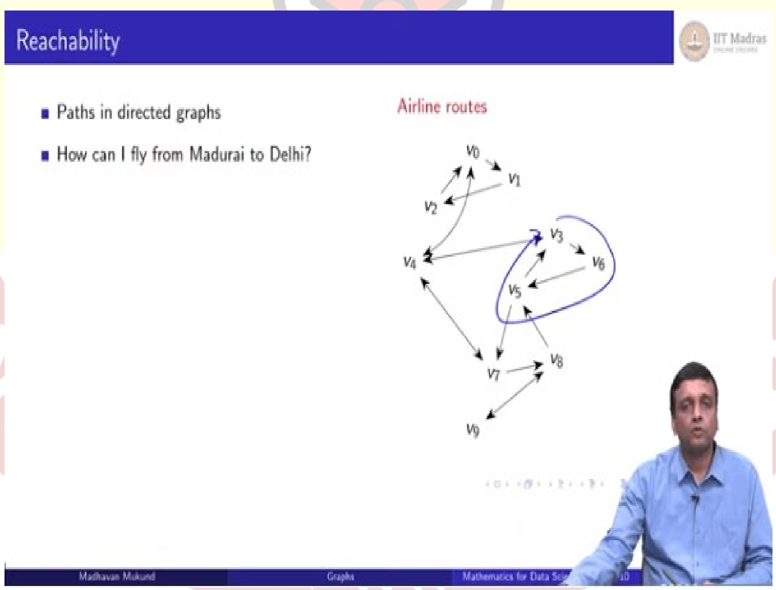
So, one thing is that there is no description in the previous definition as to whether vi in the path can be the same as a later vj in the path. In other words, you go to someplace in the graph and then you come back to that place and then proceed. Now, of course, you can imagine that this is never necessary.

So here is an example. I want to connect Kumar to Sheila but instead of going directly as I would here, right, so this would be a direct connection, so let me draw that in a different color maybe so instead of drawing a direct connection from Kumar to Sheila, I actually took this roundabout route of going around and going. So technically, in graph theory, this is not a path.

So, path should not repeat a vertex, if you have a sequence which starts at the vertex and ends with another vertex and possibly repeats vertices along the way, the graph theoretic for that is a walk, so walk is a more general type of path, in a path we usually assume that there are no repeated vertices and if there are no vertices and there are only some n vertices in the graph, every graph has a finite number of vertices and usually we call it n as the number, then clearly a path cannot have more than n vertices because if I have n plus 1 vertices in my path some vertices shall repeat.

So, this means also in terms of edges the longest accurate path that is path in the strict sense can have at most n minus 1 edges. So, we have at most n minus 1 edges in a path, where the size of v is n.

(Refer Slide Time: 9:44)



So, that was an example that we did in the friend graph, but of course, you can also do paths in directed graphs. So, let us not look at that previous directed graph which is a bit boring because there is no way to go from a teacher to a course and then go anywhere, you just get stuck. So, a more common, directed graph is something that represent say transportation network.

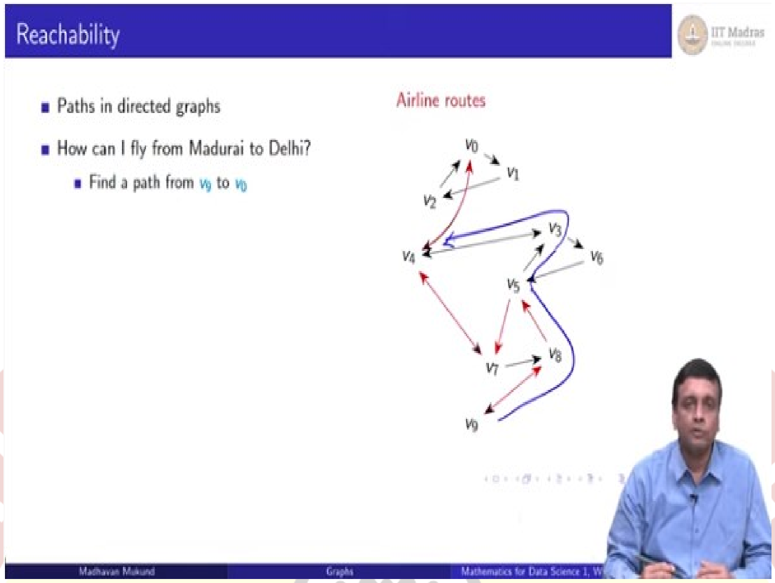
If you have ever looked at an airline map or railway map, you might find graphical representations of the routes that the airline or the railway services. So, for instance, here this might represent an abstract picture of 10 cities which are served by some airline. So, if we assume that this is super impose roughly on a map of India then v0 somewhere in the north, so let us assume maybe that v_0 represents Delhi and v9 the tenth vertex in this graph maybe represent some city in the south, let us say it is Madurai.

Now, there are arrows in this graph, indicating that not all flights go between both cities in both directions, typically between large cities like say between Delhi and Mumbai or between Bangalore and Chennai if you have a flight going in one direction by an airline you also have a flight in the reverse direction, so you can assume that if you can go in one direction, you can also go in another direction. But very often in smaller sectors airlines might operate these kinds of triangular routes, right.

So there is a route, which serves three cities, but it does not go back and forth between each pair of cities, it starts in one city goes to the second one, goes to the third one and returns to the first. So, for instance, in this case if you want to go from v3 to v5 you have to sit through a halt in v6.

So, now given this graph, the same question that we asked before about how Priya would discover that Radhika can help her. In this case we might ask a more direct and natural question saying that I am in Madurai in v9 and is it possible on this airline to travel from v9 to Delhi which is v_0?

(Refer Slide Time: 11:43)

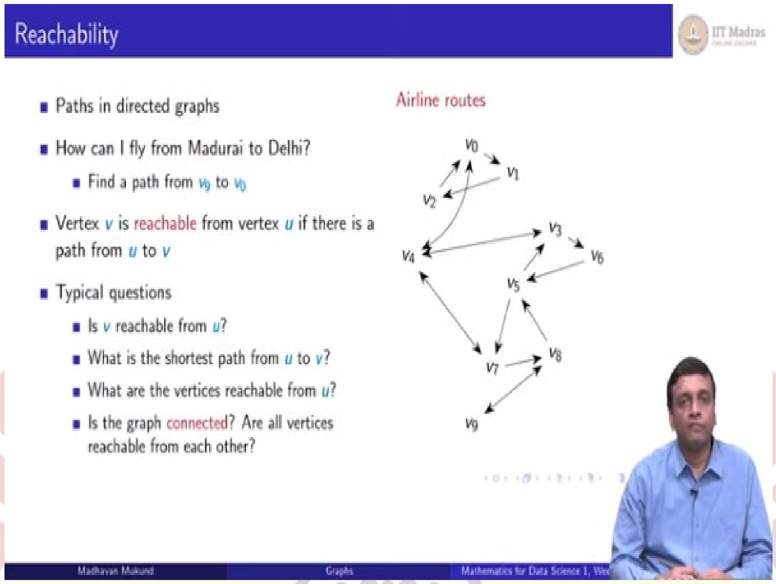


So, I need to find a path and of course, if you look at this picture here is a possible path, I can go from v_9 to v_8 and then v_5 and then to v_7 and then back to v_4 and then up to v_0. So, notice that some of these edges we drew with two arrows like this and this. So, this indicates that it is a directed graph, but we explicitly have an edge in both directions. We have an edge from v_4 to v_0 and we have an edge v_0 to v_4 where as these where there is only one arrow, one arrowhead is a unidirectional edge, that is I can go from v_3 to v_6 but I cannot go back from v_6 to v_3 in this graph.

Now, this is not the only way to go, obviously, so we could for instance, instead of doing this, we could, for instance, at the v_5 we could have followed this path up to v_5, up to there we have no choice because from v_9 we can only go to v_8 and from v_8 we can only go back to v_9 which is not very useful or we can go on to v5 but at v5 there is an option to go to v_3, right and v_3 is connected both ways to v_4, so I can come back to v4 so this is another way to reach v_4.

So, there are two ways to get to the v_4 either via v_3 or via v_7 both of them have roughly the same number of, same number of cities in between so there is not much advantage, but there are multiple paths and we can discover them.

(Refer Slide Time: 13:02)



So, in graph theory what we say is that a vertex is reachable from another vertex, if we can find a path, so we say that v is reachable from u, if there is a path from u to v. So, some of the questions that we might be interested in a graph, of course, the first question is, is a vertex v reachable from u?

This is the kind of question that we asked just now about Madurai and Delhi or about Priya and Radhika, so if Radhika can help, is there a way that Priya can find out about it through her friends networks, so this is a reachability question for a specific pair of vertices.

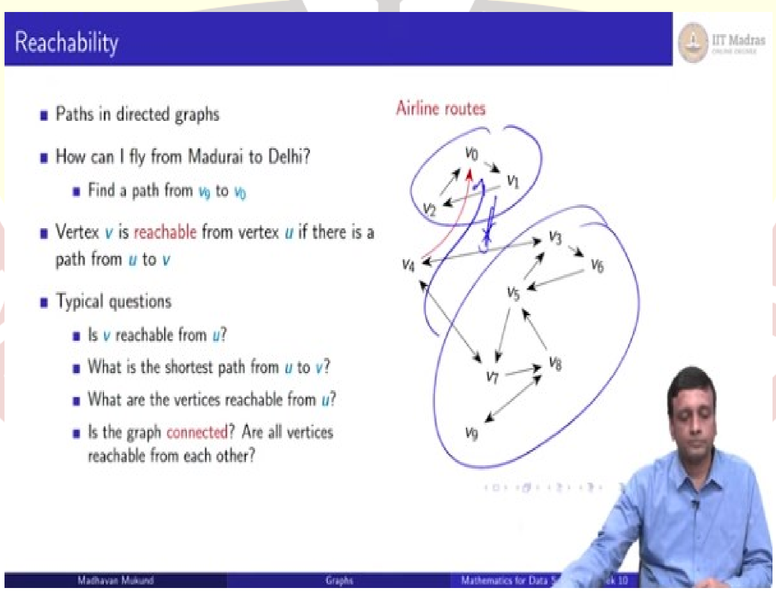
Now, given that this is possible, you might still want to find out the best possible way to do it in terms of say the shortest number of flights. So, when you log into something like MakeMyTrip or any of these travel websites, it will offer you a number of flights direct flight or one hop flight, a two-hop flight and you might prefer to go by a flight which has fewer stops, so that you do not have to waste your time waiting while the plane is on the ground.

So, we might ask for the shortest path, now shortest path for us right now is just in terms of number of edges or number of intermediate vertices, but later on we will see that we could also associate some kind of distance or time with each leg and then we could ask for the shortest path not in terms of the number of hops but in terms of some quantity that we are measuring as we travel, say the time or the distance.

Now, we could also ask a more general question which is that if I started u were all can I reach? So, in particular, if I know this, if I know where all I can reach, then I can answer the first question because if v is one of those vertices that I can reach, then v is reachable from u, so this is a more general questions then asking whether a specific vertex is reachable is asking where all can I go from a given starting vertex.

And then we can ask more general questions about the graph as a whole, so is the graph connected, can I go from everywhere to everywhere? Now, in an undirected graph this basically means that if I start at one vertex, I can reach every other vertex, in a directed graph is a little more complicated, I may be able to go from one vertex to every other vertex but I may not be able to come back. So, I need to go from every vertex to every other vertex.

(Refer Slide Time: 15:15)



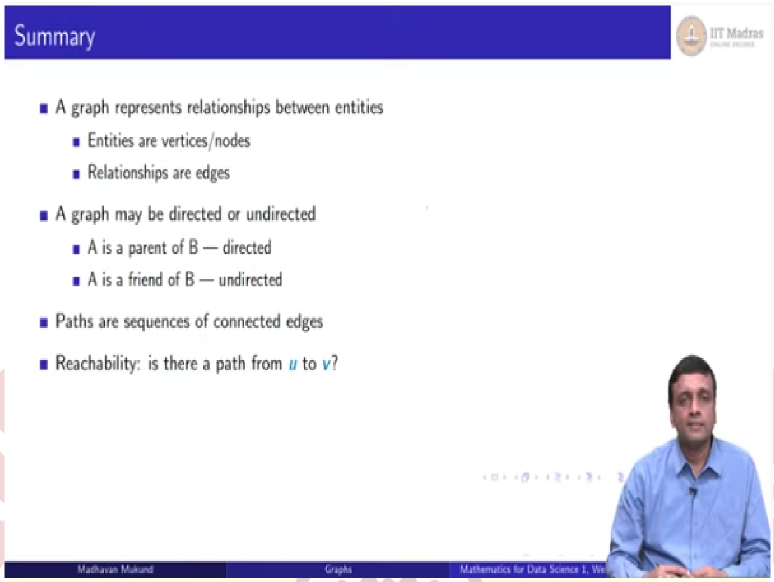
So, let at look at this graph for instance. Supposing, we take this v4 to v3 flight and make it to one directional flight. Now, is it still possible to go? In the earlier graph actually you can check that you can go from everywhere to everywhere, because the crucial edges so we have this section where you can go from everywhere to everywhere, you have this section which you can go from everywhere to everywhere, you have this section where you can go from everywhere to everywhere and all these sections are connected by these bi directional things, so from each of these components you can go to every other component via v4 and then within that component, you can go around in a circle of three and reach any place you want.

So, this original graph is definitely connected, now what happens if you break this bi directional connectivity to v_4 by saying by saying you can only go from v_4 to v_3? You cannot go back. Now, earlier we could take this as an escape route from v_3 to go from this triangle to any other triangle but now we can still go down from v_5 to v_7 and then proceed.

So, even if we make v_4 to v_3 a one directional edge, there is no problem, this graph is connected in the sense that from every city we can reach every other city. However, if we take another edge from v_4, say v_4 to v_0 and make that a single direction then we have an issue, because go we can go from everywhere here, we can go up through that edge, if we are at the top, we cannot come back down, right there is no way to leave that v_0, v_1, v_2 triangle, because there are no edges leaving them. So, now the graph has become not connected in one specific way which is the vertices v_0, v_1, v_2 cannot reach the other vertices.

Everything else can reach, so notice that is a symmetric, the other vertices can reach. So, we are not solving these questions, we are just posing these as typical questions that you might want to ask once you have a graph presented to you as a representation or some information so that information, we started with the motivation that it came from a relation, but it could also be some natural information like this like friendships, or it could be like airline routes.

(Refer Slide Time: 17:19)



So, to summarize a graph represents a relationship between entities, so in the graph these entities are represented as nodes or vertices, and the relationship that we are trying to capture is represented by edges between these and these edges might be directed or undirected. So as a directed graph, we saw one example, which was the airline route. Another example involving people can be family relationships. For example, if you have a group of people and you want to connect people who are parent child. So, if A is a parent of B, you want to say that A is related to B.

Now, clearly this is asymmetric if A is parent of B, then there is no way that B can be a parent of A. So, this could be a graph that you might have seen pictorially in the form of family trees, so people sometimes represent relationships within the family, in terms of a graph where they draw edges between parents and children and then they have a way of connecting people who are married together and all that, but a family tree is a kind of a graph and that would be typically directed because parents' child is asymmetric relation.

On the other hand, as we saw if you have a friend relation it becomes an undirected graph. So, we have these two fundamental types of graphs and the problem that we have looked at right now which is of interest is within the graph to identify paths and through paths, talk about reachability and connectedness. So, we will explore these problems more systematically in the lectures to come.