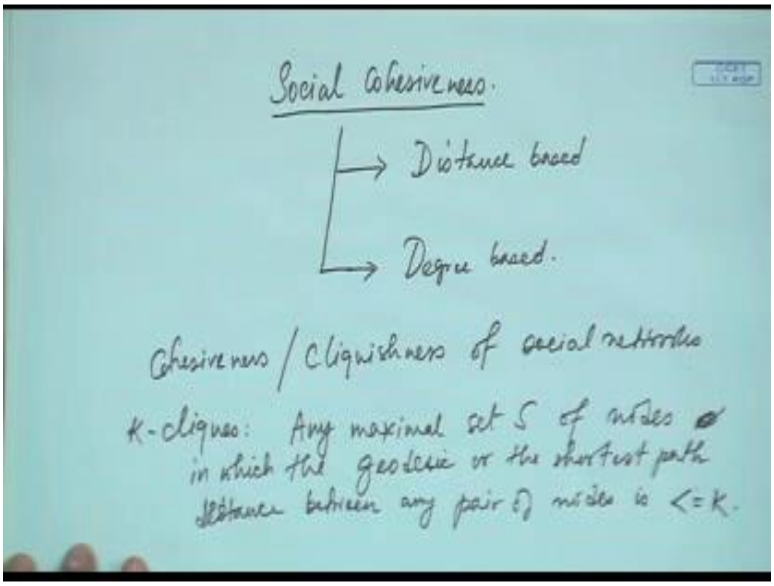
Complex Network: Theory and Application
Prof. Animesh Mukherjee
Department of Computer Science and Engineering
Indian Institute of Technology, Kanpur

Lecture – 09
Social Network Principles – II

We will continue with the Social Network Principles. Last day we have looked into the idea of assortativity or homophily as well as we have looked into the idea of signed networks, signed graphs and structural holes. Today, we will talk about Social Cohesiveness.

(Refer Slide Time: 00:38)



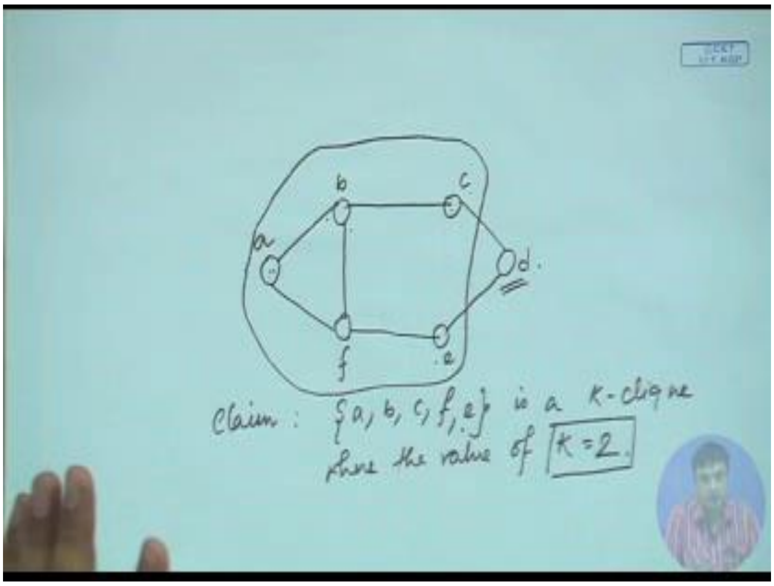
Now, social cohesiveness can be defined broadly in two different ways; one class is distance based, whereas the other class is degree based. So, we will start-off with distance based measures and then we will slowly go in to degree based measures. So, one of the first types of distance based measures is actually, cohesiveness or cliquishness of social networks can be measured as I said through these two different ways either distance based or degree based methods. So, the first method we will talk about is first definition that we will come across is k-cliques.

A k-clique is defined as follows. So, clique is actually a complete graph as you know. So, given a clique it is a complete graph. So, every pair of node has an edge between them. Now, we will relax this definition in k-cliques. So, we imagine that in social network it is very difficult to get complete graph, instances of complete graph. What we will get is most likely say, for instances, 1 or 2 edges are missing from that graph. So, in a complete graph distance between every pair of nodes is 1. Now, we will probably not be fortunate enough to always get examples of such networks, such sub-networks and social networks. What we will get is we will get probably networks where the distance between every pair of nodes is, say 2 or 3 not exactly 1 fine.

But these are also cliquish in nature than for cases where say the value of k is 10 or 12. So, the point is that even when we do not get evidences of complete graphs, we can very well get evidences of little relaxed versions of complete graphs and the first relaxation that we will talk about is k-cliques.

So, the idea is a k-cliques is defined as follows, any maximal set s of nodes in which the geodesic or the shortest path distance between any pair of nodes is less than or equal to k. So, that is the definition of k-cliques. In this particular type of graph, what you have is that take any pair of nodes the distance between them is a shortest path distance between them is bounded by k that is from where the name k-clique comes in. So, we will take an example and see for instance, let us consider this particular example.

(Refer Slide Time: 05:10)



So, in this example I make a following claim. The claim is that a, b, c, f, e is a k-clique where the value of k is equal to 2. So, actually if you look at these nodes a, b, c, f and e this is the set of nodes that we were talking about. Now, this set of nodes is a k-clique and I claim that the value of k in this case is equal to 2.

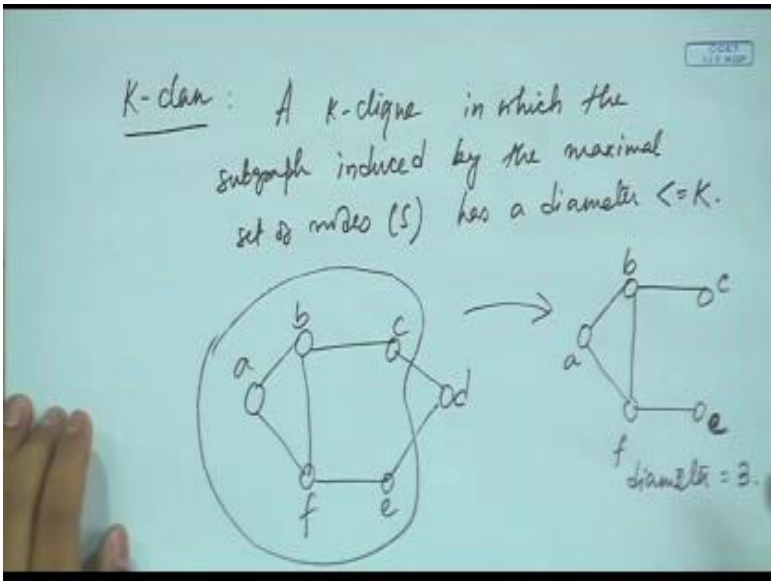
Now, you immediately find something strange. So, what happens is, let us examine the shortest path between every pair of nodes. So, between a and b it is 1, between a and f it is 1, between a and c it is 2, between a and e it is 2. Now, between b and f it is 1, between f and e it is 1, between b and c it is 1.

Now, what happens between c and e, if you look between c and e there is a path of length 2 actually and since your definition does not restrict you so that you cannot count this path, your definition does not restrict you in any way so that you cannot count this path. So, k becomes equal to 2 basically even while d is not a part of the set as per definition, there actually exist a path between c and e of length 2 and that is why this particular sub structure becomes 2-clique even while d is not a part of this maximal set.

So, let us understand this carefully once again, even while d is a not a part of this maximal set since this definition does not constrain us on not including this path. So, definition does not ever tell us that you cannot include this path; there is no such clause in the definition. So, that is why we can definitely include this path and we can and if we include this path, we see that there is a path of distance 2 between c and e and that is why even while d is not a part of the k-clique the distance between c and e becomes 2 and that is why this is 2-clique.

Obviously, this is a problem in the definition; there is a short coming in the definition. Now, the point is like how to overcome this short coming and that brings us to the next cohesive structure. So, basically the point is that the k-cliques might not be as cohesive as they look because, for instance take this example d is not in the sets, but then it is actually building up path between r and c and e and you can actually assume that path of distance 2 between c and e and that actually makes this k-clique, that actually makes this this subset become 2-clique even while d is not part of this set. So, then in order circumvent this problem the next definition that we introduce is that of a k-clan.

(Refer Slide Time: 09:28)

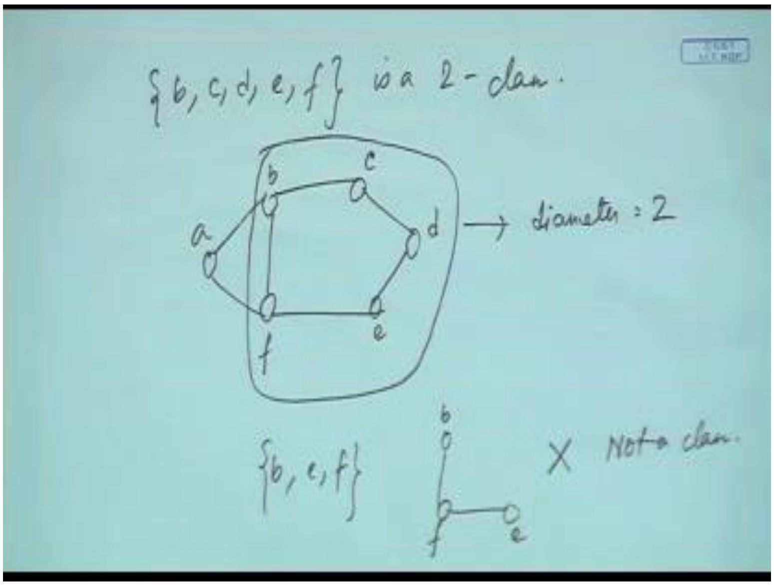


Now, a k-clan is nothing, but a k-clique in which the sub-graph induced by the maximal set of nodes, say s has a diameter less than equal to k. So, see the difference in the definition. So, k-clique was just a maximal subset of nodes where we are constraining that every pair of nodes in that subset should have a distance less than or equal k and that was being satisfied by the previous example. Now, we have restricted the definition in a way such that what we see here is that you have the same subset, but then is the sub-graph induced by the nodes in the subset, the sub-graph induced by the nodes in that maximal subset of k-clique that should have a diameter less than or equal to k.

Now, if you take the example once again the previous example, you see now no longer this particular group of nodes if you induce a sub-graph with this particular of nodes let us induce the sub-graph, the induce sub-graph will be and given this sub-graph, what is the diameter? The diameter of this sub-graph is the distance between c and e which is diameter is equal to 3.

So, basically this is a 3-clan, this is no longer 2-clan because here you do not have the liberty to include the path that goes through d actually which in the earlier case in the k-clique case, you have this liberty to include that path even while d was not part of the maximal set. Here, since you have the restricted definition of the induce sub-graph, now you can no longer include that path and in this particular example, now your diameter is basically equal to 3 and therefore, it is a 3-clan and no longer a 2-clan. So, examples of some 2-clan from this same graph would be say, for example, b, c, d, e, f is a 2-clan.

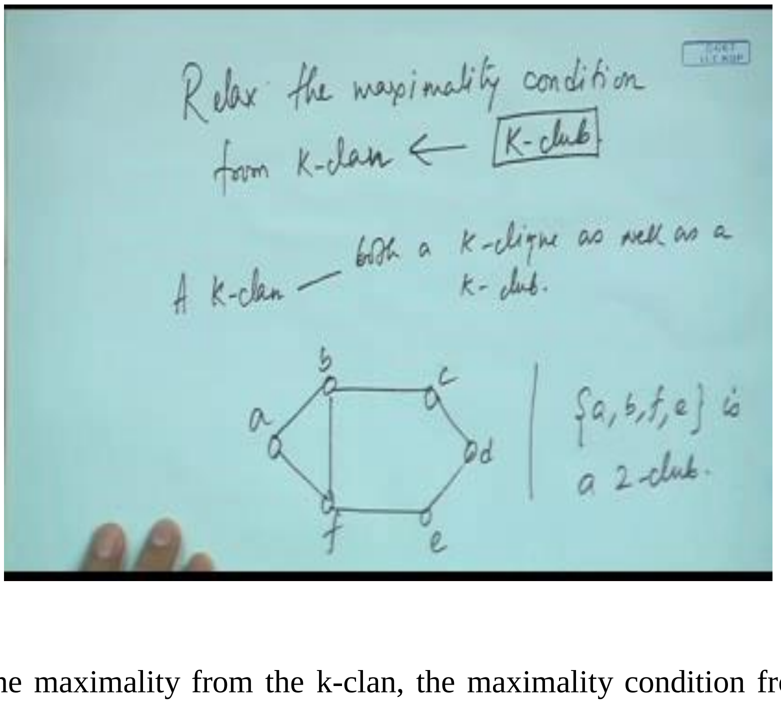
(Refer Slide Time: 12:39)



Let me draw the graph once again for clarity of understanding, now if you take the sub-graph b, c, d, e, f that is this here the diameter is equal to 2 and hence it is a 2-clan; however, let us say take the example of the sub-graph b, e, f which is basically this particular sub-graph. This is not a 2-clan why because remember a 2-clan has to be in the first place a 2-clique and for being a 2-clique the set has to be the maximal set so that you do not have the possibility of including anything else and still keeping this set maximal. So, here if you take the b, e, f then you have possibilities of including other nodes and still keeping the clan to be a 2-clan. So, this is not a maximal set b, e, f. In fact, the set b, c, d, e, f is the maximal set that is why b, c, d, e, f is a 2-clan, whereas b, e, f is not a 2-clan.

Now, however b, e, f if you look carefully is also a cohesive structure. So, in order to admit this particular cohesive, this particular cohesive notion of cohesiveness, we will relax the maximality condition from k-clan.

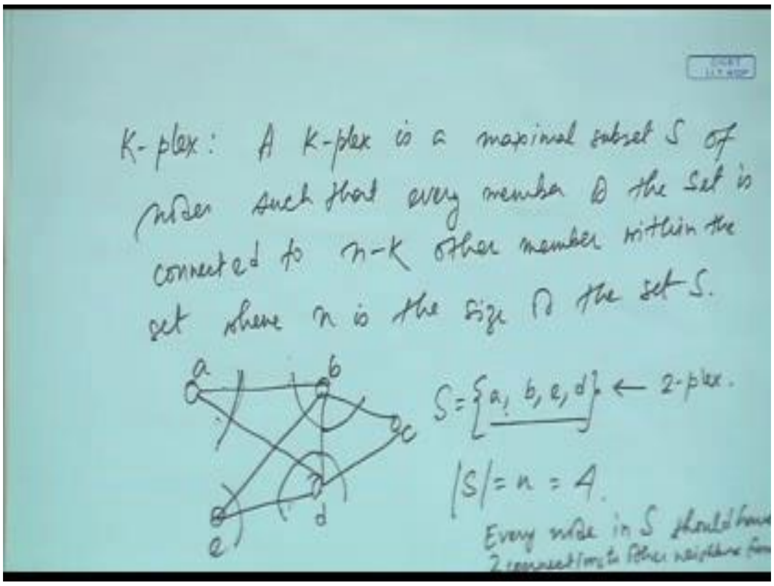
(Refer Slide Time: 14:50)



If you relax the maximality from the k-clan, the maximality condition from the k-clan what we arrive at is a k-club, this is another form of cohesive substructure. So, basically a k-clan is both a k-clique as well as a k-club. However, the vice versa are not true right if you have a k-clan then it is automatically a k-clique by definition as well as a k-club by definition. However, the reverse in each case is not true, for instance if you take the same example once again, the same example in this case a, b, f, e is a 2-club. However, this is not a 2-clan.

Now, these were sum of the distance based notion of social cohesiveness. Now, we will discuss at least two notions of social cohesiveness based on degree the first one among these is what we called a k-plex.

(Refer Slide Time: 16:47)



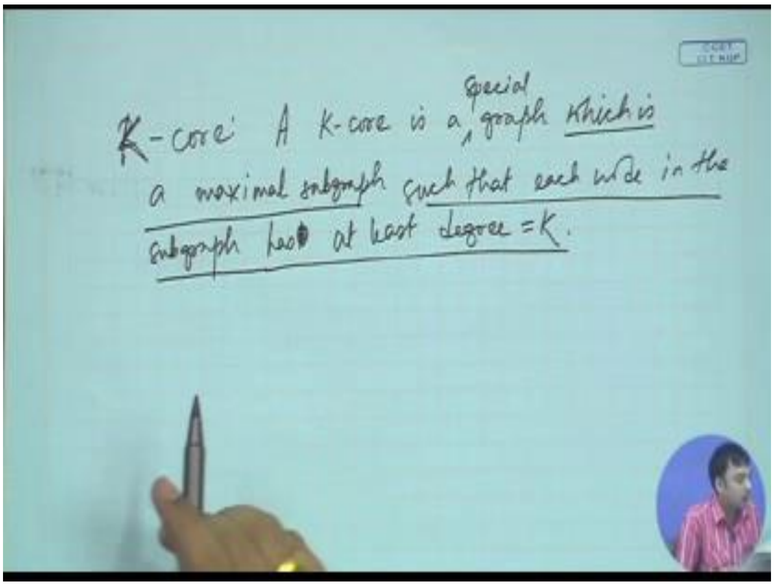
So, a k-plex is again a maximal subset s of nodes such that every member of the set is connected to n minus k, other members within the set where n is the size of the set s. Basically, this is a degree based notion, this is one of the first degree based notion of social cohesiveness.

Let us take an example and see, in this particular example, if we take the subset a, b, e and d then this actually qualifies as a 2-plex, why let us see for instance, take the node a, now node a, what is the degree of that node? So, that is denoted by, there are edges going out of this particular, there are two edges going out of this particular node. Now, let us take the size, what is the size? The size of this particular set is s is, mode s is equal to n is equal to 4 so that means, if you have a 2-plex then every node in s should have two connections to other neighbors from s.

Now, let us investigative this is happening take this graph. So, a has degree to and both of its connections. It has two connections every node has to have at least two connections. So, it has two connections and both this connections are to two other members from the same set s. Let us say b, b has three connection, all three have been in this set of a, b, e, d, e has four connections and however, at least out of this four there are at least which are from this set.

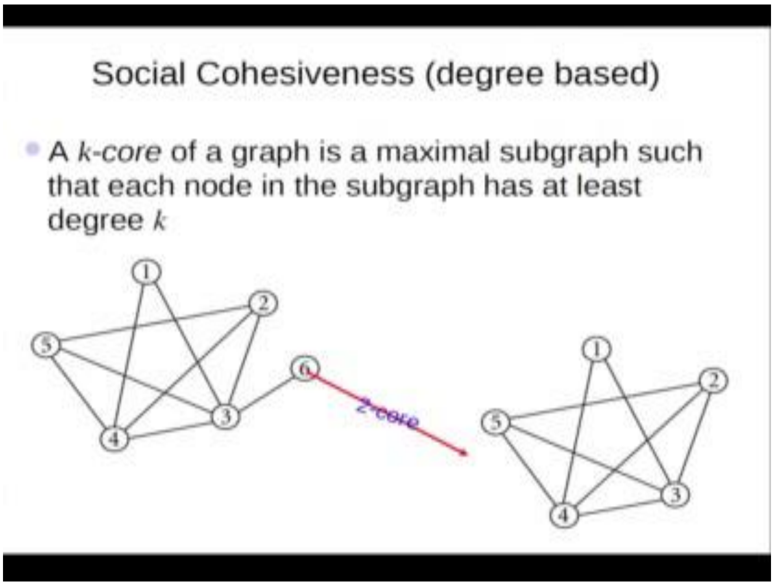
So, it is more than 2. So, it has to have at least two connections. So, it has more than two connections actually two nodes within the set. Similarly e has two connections to all other members within the set d again has 1,2,3,4 connection going out. However, at least three among them are two members within the set s. So, basically this particular set a, b, e, d actually correctly qualifies as a 2-plex and the last definition that we will talk about is 2-core or k-core basically.

(Refer Slide Time: 20:51)



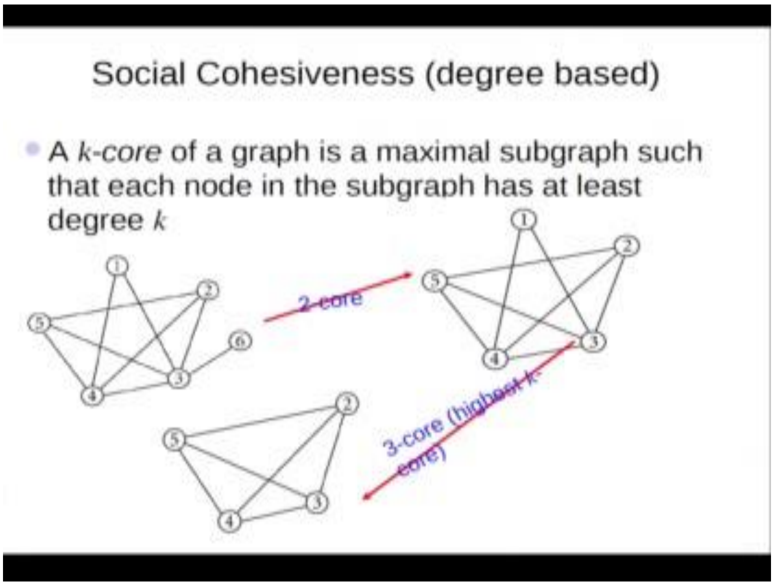
So, a k-core is a graph which is a maximal sub-graph such that each node in the sub-graph has at least degree equal to k. So, a k-core is a special graph which is a maximal sub-graph such that each node in the sub-graph has at least degree equal to k. So, if you look at the example on the slides have given an example on the slides.

(Refer Slide Time: 22:08)



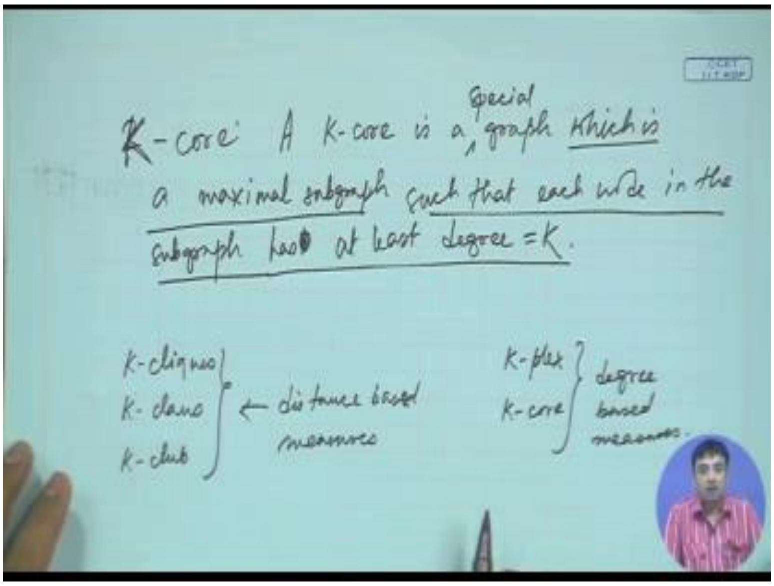
Let us concentrate on the on that example. So, in the slides you have this 6 node network to start with. Now, a valid 2-core of this network would be a maximal sub-graph, where all nodes should have at least degree 2, you see this particular graph every node has at least degree 2. So, node number 2 has degree 3 node number has 1 to the 3, 4 degree 4 node number 1 has degree 2 node number 4 has 1, 2, 3, 4 degree node number 5 has degree 3. So, everybody has node degree greater than or equal to 2. So, that is why this particular sub-graph actually qualifies as a 2-core for the original graph.

(Refer Slide Time: 22:53)



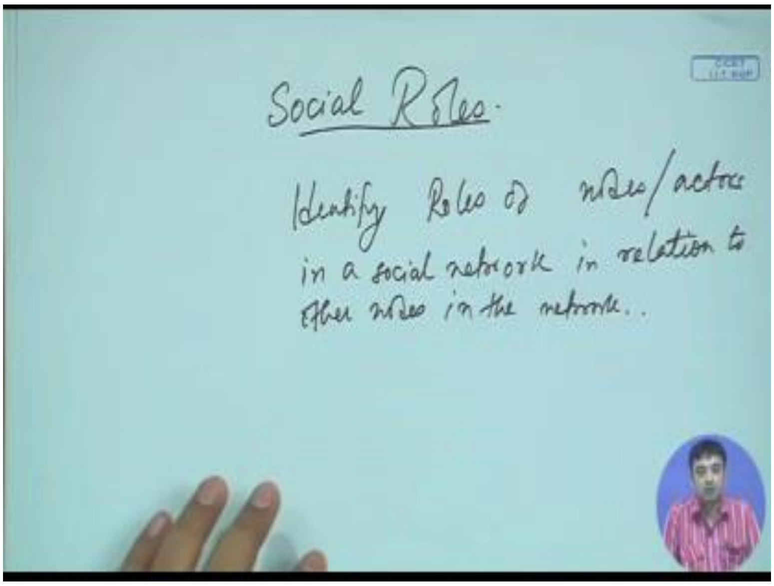
Similarly, from the 2-core you can actually also build up 3-core where you ensure that all the nodes in this particular sub-graph has at least degree 3. So, if you take this particular sub-graph of 2, 3, 4 and 5 you see all nodes have at least degree 3 and so therefore, this is a 3-core for the original graph and this actually the highest core that you can build from the source graph. So, this is another notion of degree based cohesiveness of graphs. So, recapitulate what are the different measures that we have learnt.

(Refer Slide Time: 23:30)



We have learnt k-cliques, we have learnt k-clans, we have learnt k-clubs all these three are distance based measures, whereas we have learnt k-plex and k-core which are degree based measures. Now, following this we will talk about one of another very, very important concept in social networks or especially in friendship network or relationship networks is what we called social roles.

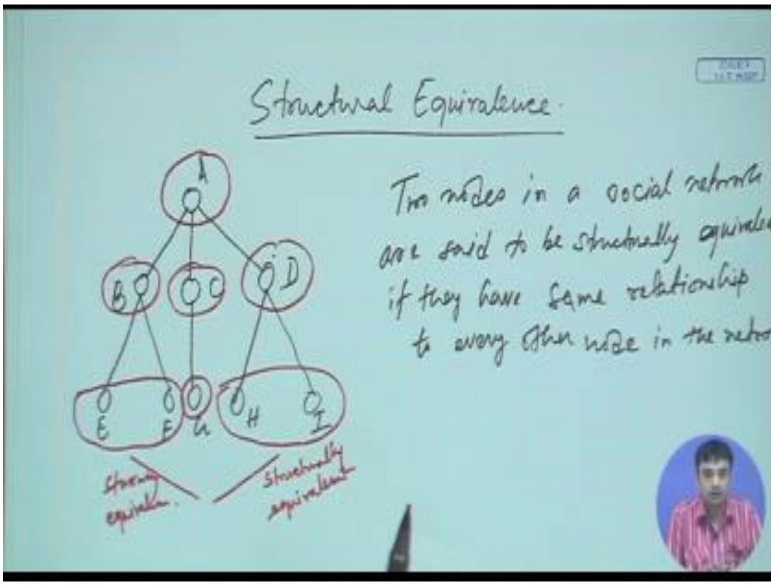
(Refer Slide Time: 24:25)



So, basically, the task is to identify roles of nodes or actors in a social network in relation to other nodes in the network, for instance, let us take and a simple example, the teacher student example in a. So, let us consider social network where I am an entity as a teacher where as you are an entity as a student.

So, this is a teacher-student relationship, and the teacher-student relationship can be thought of as my counter-part role as a teacher and your counterpart role as a student and both of us are perhaps connected to this organization of Indian Institution of Technology, where you are actually consuming knowledge, whereas I am giving knowledge. So, this is the relationship of the two role players in social networks and the question is like how you we can try to quantify these notions of relationships or what we call equivalences in social networks. So, one of the first examples that we deal with is the idea of structural equivalence.

(Refer Slide Time: 26:17)



And we will take up the example of a very interesting graph called the Wasserman-Faust graph to illustrate these particular phenomena, let us first draw the Wasserman-Faust graph. So, given the this Wasserman-Faust graph, in the context of this graph in general any graph structural equivalence is basically defined as follows; two nodes in a social network are said to be structural equivalence, if they have same relationship to every other node in the network . So, basically the idea is that if you take a pair of nodes and if you investigate the basically the neighborhood of this node.

So, the question is like whether the neighborhood of this this par this pair of nodes actually perfectly overlaps or not that is the question we ask. So, if these two nodes have a neighborhood which perfectly overlaps. Then these two nodes are called as perfectly substitutable nodes they have the same set of relationships to all other nodes in the networks. So, if I am connected to say, I am structural equivalent to you then the nodes that i will be connected to you will be also connected to them and only to them and nobody else this is the idea of structure so that means, that somebody in this network can perfectly substitute myself with you that is the idea of perfect substitutability or structural equivalence in a social network.

And once you have such a definition, immediately this graph this Wasserman-Faust graph that I have drawn here can be classified into structural equivalence classes. So, let us see some of the nodes which are structurally equivalent. So, if you look investigate this particular graph, you will see that the nodes h and i they have the exactly same set of neighbor which is only d and the nodes e and f are structurally equivalent.

So, these two pairs of nodes are structurally equivalent structurally equivalent and these two pair this pair is again structurally equivalent because e and f has the same neighborhood that is only b and h and i has the same neighborhood that is only d. So, h and i is a structurally equivalent set e and f is a structurally equivalent set all other nodes.

This particular node is all other nodes are structural equivalent classes of themselves, only this pair here e and f are perfectly substitutable because they exactly share the same set of neighbors and h and i are also perfectly substitutable because they exactly share the same set of neighbors, which is d here. All the other nodes individual define their own structural equivalent classes, they are not structurally equivalent to any other node. So, this is the idea of structural equivalence.

In the next lecture, we will see how to quantify the notion of structural equivalence.