Social Network Analysis
Prof. Tanmoy Chakraborty
Department of Computer Science and Engineering
Indraprastha Institute of Information Technology, Delhi

Chapter - 02
Lecture - 06
Lecture - 01

Hi everyone. Welcome to the course on Social Network Analysis and this is Chapter 2, this chapter is on Network Measures. So, in the last chapter we have actually discussed you know why social network analysis is an important topic or a subject for an academic curriculum and why people really care about you know understanding social networks right. What are the potential applications, we have seen you know different types of networks, social networks.

It can be social networks, it can also be say biological networks, information networks and how people essentially model write a complex systems in terms of a network right. We have also seen you know properties, you know we have briefly talked about different types of properties which we generally use to characterize a social network right. So, in this particular chapter we will measure a network ok.

Now, what do we mean by measuring a network? We will basically look at different components of a network or different entities, different parts of a network right. If you remember correctly in the last you know chapter we have talked about 3 levels of analyzing a network, microscopic level and mesoscopic level.

We will discuss different you know properties for every level of a network right and then we will try to characterize you know different components of a network. It can be an edge node or a sub graph and so on and so forth ok. So, let us get started.

(Refer Slide Time: 02:12)

So, I will start this chapter by you know discussing these two interesting you know social media content right, both of these content went viral, I am pretty sure you guys are aware of you know the virality of this content.

The first one is Gangnam Style right and this went viral in 2012 and this is we all know that this is basically a South Korean pop song and this went viral in 2012. And the second one is an Indian song, Kolaveri Di right and this also went viral within a very you know small amount of time right, within a very small span of time ok.

So, I mean of course, this is I mean what is interesting here, I mean the rhythm was good the music was catchy, the video may also be catchy right. So, what is new about these two songs right? Why is suddenly we you know need to discuss about these two songs in this particular chapter? Ok. So, you know there are many papers, if you look at the papers right there are actually papers which studied why say songs like Gangnam style went viral ok.

So, this is not only about the rhythm or the songs quality or the video right of course, we all agree that these things, these things were actually good in this particular song, but there might be some other factors which actually helped these two songs went viral ok. So, let us talk about Gangnam style right. So, Gangnam style if you remember the history right, you know when this particular video was uploaded right Britney Spear right. So, she tweeted about this song right and that created a huge amount of tweets or retweets right.

According to the Wikipedia after Britney Spears tweet you know there were around 1.3 million tweets containing the term Gangnam style right and that too within a very few days ok. So, here one factor could be the tweet by a very popular celebrity like Britney Spear right. And if you look at you know songs like Kolaveri Di again according to Wikipedia upon release of this particular song you know the hashtag Kolaveri Kolaveri.

So, that topped the you know that talked as a hashtag as a trending hashtag on twitter right, particularly in the Indian context. Again this is the time is around you know I think 2 2011 November right. And you know immediately after that we have seen cases where celebrity started you know retweeting about it, it also I think it was also played in US radio station right.

In November 2011 right it had I think more than 10.5 million YouTube views right; so and the rest is history. So, if you look at you know social, I mean if you look at standard you know the sources authenticated sources you will see the you know the studies on this on these two songs separately.

So, one important lesson that one can learn is that, you know if you really want to make something viral right you convince some celebrities right. Say you can convince say Shah Rukh Khan for example, right or you can convince some political leaders for example, you can convince some actors, some actress, who are celebrities so called celebrities on social media right.

You may also have heard about you know something called viral marketing right, say for example, you have launched your own start-up right and you do not have enough money to promote your start-up right. What you can do? You can you know you can hand over some of your products to some celebrities and let them try your products right. And then you then if they are satisfied with your products and then you request them to write about your products on say social media like Instagram, on twitter right that can be one way.

And you know there are actually cases, there are I know plenty of such start-ups who actually you know adopted this particular techniques right. They did not have enough money to promote their start-ups and they basically you know give these products to some celebrities, some well-known figures on social media and the and those celebrities actually tweeted posted about their start-ups and then automatically the start-ups got traction ok.

So, this is one of the strategies right, but there is no free lunch right. So, meaning that you I mean when you convince somebody right, either you give him or her money right, but you know the amount of money that this celebrities will charge, that might be you know higher than the total budget of your start-up for example, right. So, in some ways you have to convince popular users on social media to promote your product or to promote your brand for example, right.

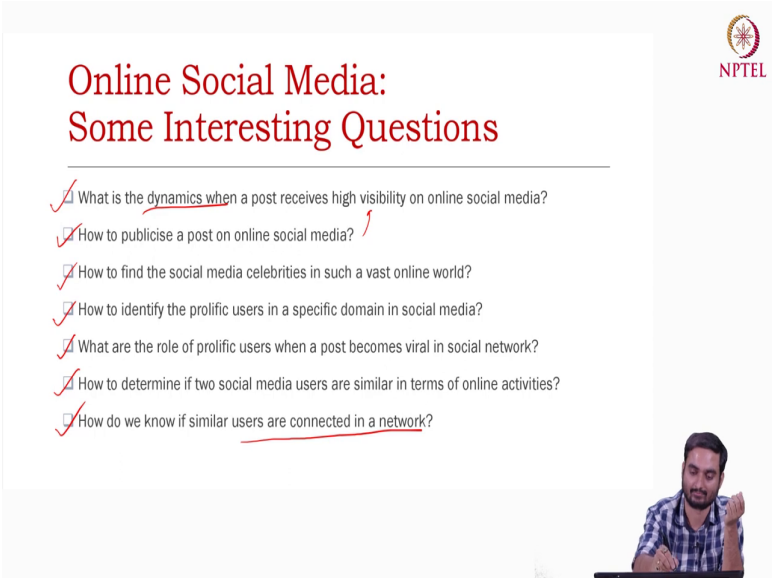
So, immediately the question is how do we measure, who is a popular user right. Now, since you know that say you know people like ShahRukh Khan or you know Sachin Tendulkar or Britney Spear they are actually popular because of their you know because of their work. But sometimes from a naive users point of view, who say for example, who does not have much idea about the activities right.

So, let us say; let us say you give that user a particular network right. So, by the network what I mean to say is that you have set of nodes and set of edges. How do you know how do you measure or how do you identify or how does a naive user identify that these nodes are popular nodes right, these users are popular users.

So, in this particular chapter we are going to discuss about many such measures by which you can identify the popularity of a node right or a popular popularity of an edge for example.

So, I have given you one case study where you know popularity metric right is used, but you can also use it in many other cases and remember the notion of popularity is not sacrosanct meaning that depending upon the applications you can think of different types of popularities right. We will discuss about many such you know notions in the following slides.

(Refer Slide Time: 09:40)



So, what are some interesting questions that we are going to discuss in this chapter? The first question is what is the dynamics when a post receives high visibility on online social media? Right; is this natural? For example, if I post something right versus a popular user or a celebrity post something right, the quality of the posting right the quality of the text the content may be same or may be similar right. But why my post you know does not receive enough popularity whereas, the other post receives a lot of popularity.

What is the dynamics behind it? So, we will discuss about it, how to you know publicize a post on online social media right and this is related to the previous question that we have asked. The third one is how to find social media celebrities, you know the one that I was mentioning right in such a vast online world right. If you say do not know let say you are analyzing the social network of say France for example, right.

And you do not have ideas about the, you know the kind of work that the celebrities do. For example, who is the celebrity in the space of movie, who is a celebrity in the space of sports and so on and so forth, right. You are only giving a network. Now, based on this how do you measure the popularity?

So, fourth one is how to identify the prolific users in a specific domain in social media. So, popularity is a very generic term right. Now, if you focus on a particular domain right, let us say sports or let us say movie right. Now, in this particular you know genre or a you know

domain how do you identify that ok, this user is popular or this user is more popular than the other user right.

The fifth one is what are the role of prolific users when a post becomes viral in social networks. Now, remember it is not the case that every time when a particular post becomes viral there is a celebrity or a popular user behind it, it may not be the case. In fact, there are plenty of such cases right, where we have seen that normal you know simple tweet. For example, you know the I mean you may have seen songs which became very viral for example, this Kacha Badam right.

Now, this song went so viral, but there is no I mean its not the case that you know it started from a from a post of a celebrity, right. So, there might be other factors and we will try to dig deeper into understanding the other factors behind such virality. Now, in this particular chapter, we will briefly talk about some of the factors, but we have a separate chapter altogether where we will discuss you know how particular post becomes viral and what are the dynamics behind it.

The next question is; how to determine if two social media users are similar in terms of online activities? Let us say you have user x user y, right. Now, user x has you know its own network structure, user y has its own network structure, meaning that the neighbours are different. For example, right if the neighbours are same then you can also say that ok they are same.

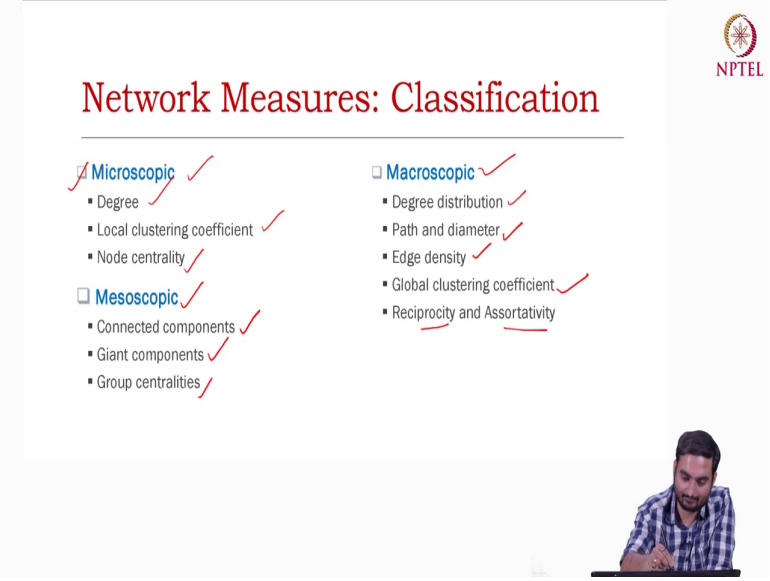
But say let us say they are different, but structurally all the neighbours are different, but structurally how these neighbours are actually distributed, structurally it is same ok. Say, let us say 2 nodes you know are located at two different regions of a network, but if you look at the distribution of fast of neighbours of both these nodes you see that you know they look like a star like structure.

So, can we say that these two nodes are equivalent or same right. Now, why this is important? This is important in in cases like say you want to replace a node by another node, say you want to replace an actor by another actor right. If you know the if you somehow characterize the users based on the neighbour distributions or based on other factors, you will say that ok the these two users are equivalent. So, in certain applications in certain activities you can use user x then user y or you can actually you know sign you can actually use node x instead of node y right.

Because you know you also need to understand that nodes or users may not be available every time. So, say for example, if you want to again let us take the same example of virality right, viral marketing. So, you approach somebody, but that person says that, look I am not available for another two to three weeks. So, you need to basically approach to another person ok.

So, the last one is how do we know if similar users are connected in a network ok. So, as I mentioned earlier, that let us say two users are similar in terms of some activities, but they are far apart from each other. So, how do you note that they are connected? Right again there are different notions of you know connectedness we will discuss this notions one by one. So, primarily we will basically try to answer you know this you know 7 questions, 6 to 7 questions in the remaining part of the chapter ok.

(Refer Slide Time: 15:16)



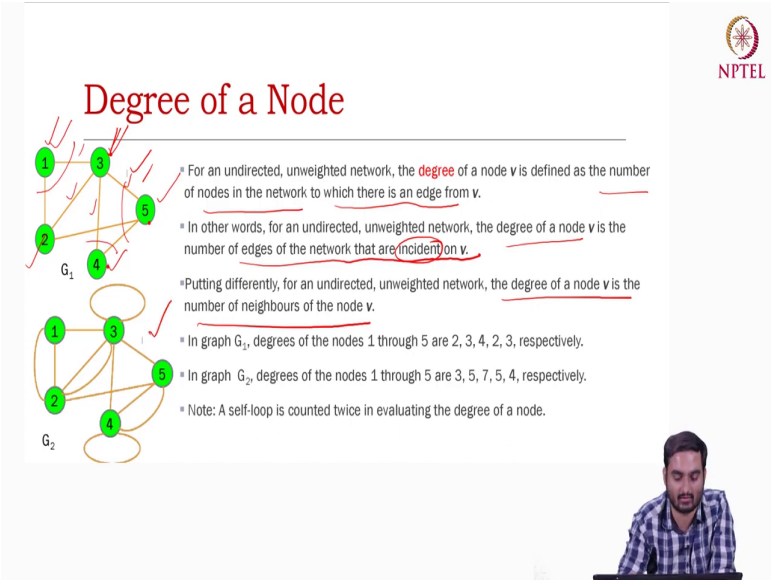
So, now let us look at the metrics that we are going to discuss. So, we will basically categorize metrics into, you know into 3 buckets. We have microscopic level of microscopic level you know in within this microscopic level we have metrics like degree, clustering coefficient and node centrality, we will discuss each of them one by one. We have macroscopic level by where we look at the entire network as a whole you know microscopic level we look at specific entities like nodes and edges.

The fine grained entities in a network macroscopic level we look at the entire network as a whole and we will discuss degree distribution path and diameter, edge density, global

clustering coefficient, something called reciprocity and assortativity. And the third category is mesoscopic you know level of analysis, where we look at. So, you remember last class we discussed that you know mesoscopic level is in between microscopic and macroscopic. So, at mesoscopic level we look at different substructures of a network, it can be a component, it can be a motif right, it can be a group, it can be a community and so on and so forth.

So, particularly here we will discuss about you know something called connected components, giant component and if possible group centralities right. This is the; this is the you know flow of this chapter.

(Refer Slide Time: 16:59)



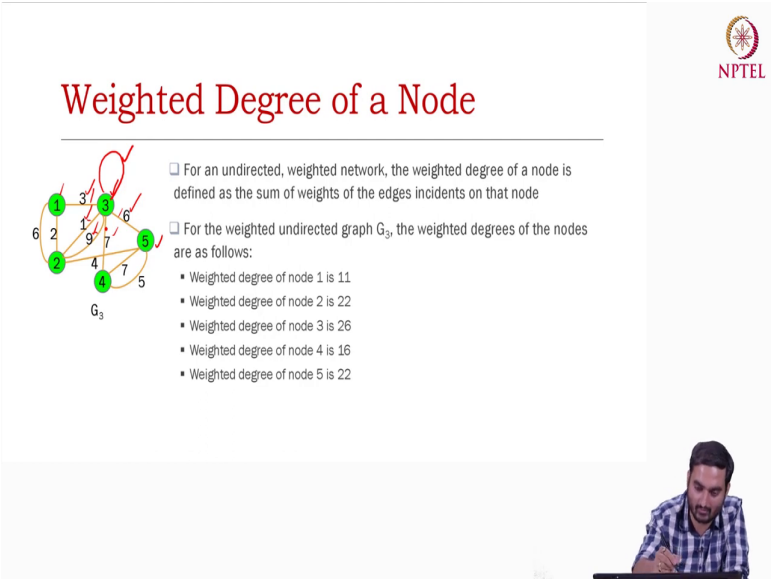
So, let us start with degree. This is the very basic you know very basic quantity of a node right. So, what is degree? So, we know many of you know that essentially you know degree is the number of edges right that are attached to a particular node. Now, this is a node centric property. So, for an undirected unweighted network there is no direction associated with the edge, there is no weight associated with the edge.

For an undirected unweighted network like G 1, right in this particular figure right the degree of a node v is defined by the number of nodes in the network to which there is an edge from v right. For example, right if you look at node 3 the node 3, there are there are 4 nodes, 1, 2, 4 and 5 and you see that there is an edge from 1 to 3, from 2 to 3 from 4 to 3 and then 5 to 3, right. So, degree of node 3 is 4.

Now, this you know this is apparently very simple metric right, but later we will see that you know even a simple degree, metric right I mean the beauty of this metric is in a different aspects, we will discuss about that later. So, this is undirected right and say I mean you can also of course, interpret it in different way for example, you can also say that the degree of a node is the number of edges of the network that are incident right, incident on v.

Now, this terminology is very important ok. So, I am going slow because these terminologies you need to remember throughout the course right, degree incident to a particular node and so on and so forth right. So, you can also say that the degree of a node is the number of neighbours of a particular of the node v right. So, for example, here as I already mentioned node 3, has degree 4, node 1 has degree 2, node 4 has degree 2, node 5 has degree 3 and so on and so forth, right. Similarly for G 2, both G 1 and G 2 are undirected and unweighted.

(Refer Slide Time: 19:22)



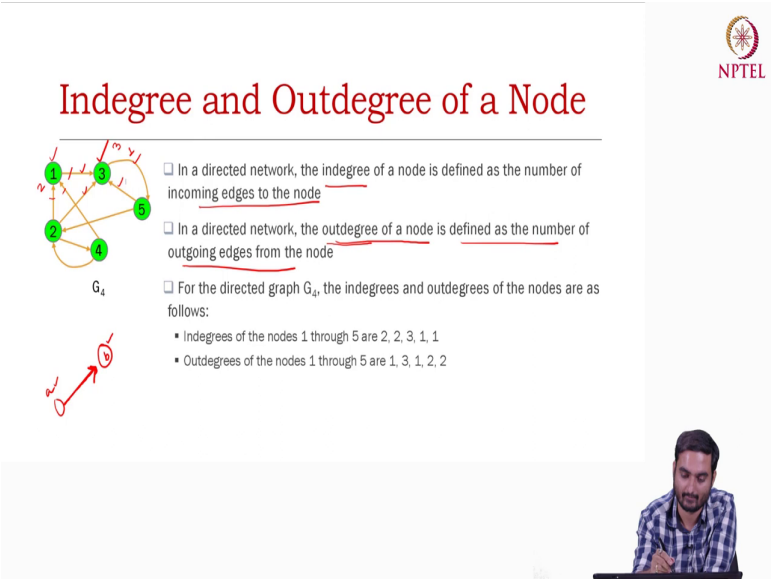
Let us think of weighted network right, where weights are associated with edges ok. So, for an undirected weighted network, the degree of a node is sum of weights of all the edges associated with the particular node right. Say for example, for node 3 you see that there are 4 edges, this one, this one, actually 5 edges, this one, this one, this one. Remember there is there are 2 parallel edges right, between 2 and 3 you see there are 2 edges. So, these are parallel edges ok.

So, the degree of node 3 would be 3 plus 1 plus 9 plus 7 plus 6 ok. Similarly, for degree 5, similarly or node 5 or node 1 ok. So, this is very simple, but remember one thing, if a node

has self-loop for example, like this ok. So, if a node has self-loop that particular edge contributes to two in the degree calculation ok. For example, you know let us go back and let us look at node 5 let us say there is a self loop like this ok.

So, what is the degree of node 4, node 5 the degree of node 5 is 1, 2, 3, 4 and 5 right. So, essentially, we need to look at the you know the edge, number of edges that are incident on a particular node ok.

(Refer Slide Time: 21:02)

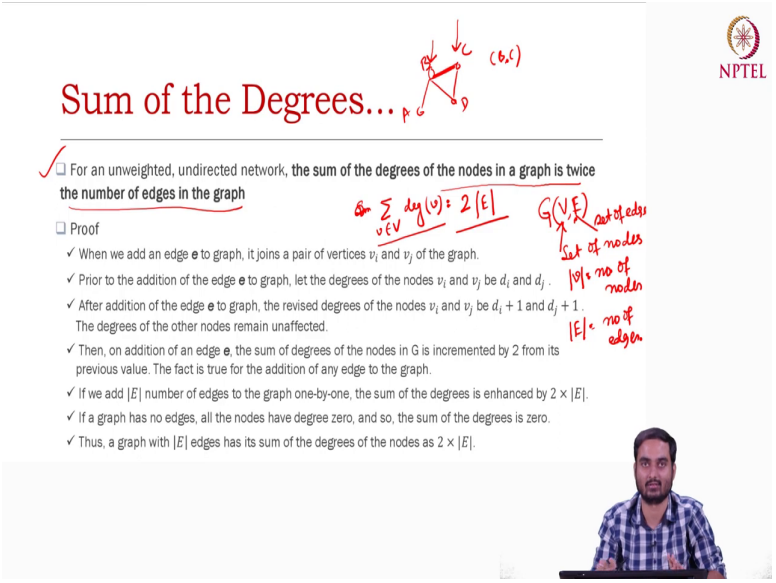


Now, let us think of a directed graph. So far we have discussed about undirected graph, directed graph. Let us think of a directed graph where there is a direction associated with an edge right. So, automatically you can think of two different notions of a degree, one is called indegree and the other is called outdegree. So, what is indegree? Indegree of a node is defined by so the as the name you know implies, indegree of a node is defined by the number of incoming edges to a particular node, right.

Remember here when you talk about an edge, now this edge actually contributes in 2 ways. So, if there is an edge from a to v, say a to b right, let us say there is an edge from a to b it means that this edge indicates an outward link for node a and an inward link for node b ok. So, indegree of a node is defined by the number of incoming edges to a particular node. Similarly the outdegree of a node is defined by the number of outgoing edges from a particular node ok.

So, now we can think so for every node you can think of the indegree and the outdegree right. So, let us look at this example G 4, graph G 4, let us look at node 3. So, node 3 has you know 1, 2 and 3; 3 inward edges. So, the indegree is 3 and there is one outward edge this 1. So, the outdegree is 1 ok. Similarly, node 1, the indegree is 2 these two edges and the outdegree is 1 ok. So, when you talk about degree in case of directed graph, we actually need to talk about both indegree and outdegree right.

(Refer Slide Time: 22:58)



Now, there is a very interesting property. Now, this property is very intuitive, but it turns out to be very very you know important; in later part of this course we will see different theorems where this this simple theory is used right. So, it is said that for an unweighted undirected graph ok, I will use the term network and graph interchangeably, both are same ok. For an undirected unweighted graph the sum of the degrees of the nodes in a graph is twice the number of edges in the network.

What does it say? It is saying that let us first look at; now remember this is unweighted undirected. So, there is no motion of indegree outdegree, there is only a degree right. Now, you take all the nodes and their corresponding degrees ok. You sum them up when you sum them up, the sum is same as twice the number of edges present in the graph ok.

So, sum of degree right, say right this is 2 into number of edges. So, this is mod ok, I represent a graph G by V comma E, where V is a set of this is set of nodes right and this is set of edges. When I write the number of nodes I will use mod v, this is the cardinality of the set

right. So, this is the number of nodes number of nodes and I will use mod E as a number of edges ok. Remember these terminologies ok and the notations as well.

So, you see that sum of degrees of nodes is 2 into number of edges. So, why is it so? Now, it is very simple to understand. Now, think of a network or think of a graph like this ok, ok. Say A, B, C. D each edge, let us look at this edge B comma C ok, this edge contributes two times to the degree, one time to the calculation of degree of B, another time to the calculation of degree of C right. So, every edge contributes 2 times to the degree of node right.

So, therefore, if you take the sum of degrees right, each edge contributes 2 times. So, it should be 2 into number of edges ok, think about it; go back you draw a graph right and you think about it. This is a very simple you know theorem, that is often used in graph theory and social network as well.

So, I stop here. So, in the next part of this chapter, we will discuss you know quantities which actually use degree as a fundamental quantity ok. We will discuss about something called degree distribution and we will see that this degree distribution can be useful to quantify networks in a different manner. Even you know if you know the degree distribution of a particular network, you will blindly say that look this network should look like this.

Let us think of a huge network right and you are not able to analyze the network, you know by just looking at it manually, by manual inspection. What you do? You calculate degree; degree calculation is very straight forward and then you plot degree distribution and then if it follows a particular pattern, we will say that look this network actually looks like this ok.

Therefore, you know I have started off by saying that all the degree the notion of degree is a very simple thing, but it has a lot of applications ok. So, I stop here. Next part we will discuss about degree distribution.

Thank you.