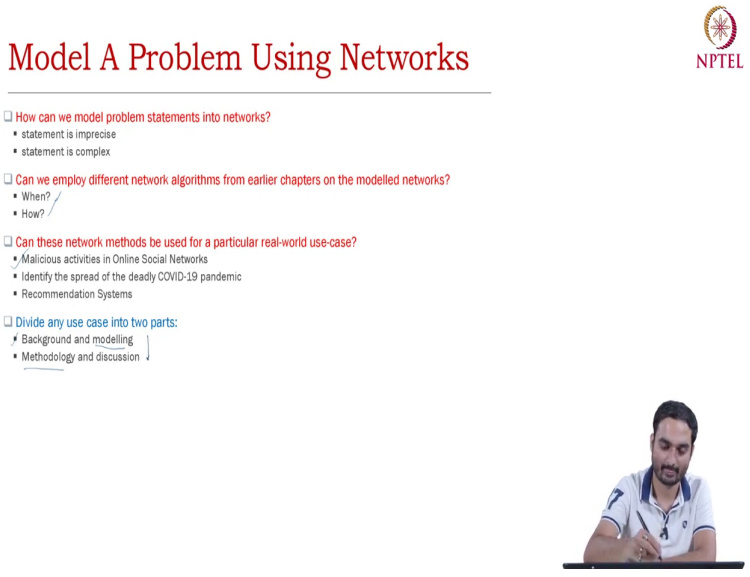
Social Network Analysis
Prof. Tanmoy Chakraborty
Department of Computer Science and Engineering
Indraprastha Institute of Information Technology, Delhi

Chapter - 10
Lecture - 01

Hi everyone, welcome back. So, this is going to be the last chapter of this course. So, we are almost done and today we will discuss some applications and case studies based on the I mean the previous chapters that we have learned so far link analysis, node analysis, community detection right, cascade and so on.

And we have seen we will try to see we will try to understand for different applications how one can you know use different methods of social network analysis right to solve the problems ok. So, let us start by a you know a very basic introduction of how we generally solve a real world problem right.

(Refer Slide Time: 01:10)



So, whenever we are given up real world problem say for example, the problem is to identify fraud users right a fake accounts on social network. So, you are given a lot of information say you know the user history, you know the underlying connections, you know the messages you know who has retweeted which tweet, you know who has message to which person and so

on. So, from all sorts of information what would be our you know important information that we will use for modeling purposes.

This is something that we want to understand. Now depending upon a particular application say the application is fraud detection what would be the approaches or what would be the components that we are going to use to model this problem ok. So, whenever we are given a problem we always think that well this problem is very tough and how do we use it there is no network as such how do we use it.

For example say if we collect data from YouTube right and you have YouTube videos, comments, write replies and so on and the task is to identify say fraudulent comments right fake comments. So, you may wonder where is the network here right there is no network as such, there is no explicit network as such.

So, then we start wondering about you know about how to abstract the data that we have and how to come up with some sort of relations between entities. Here the entities can be say users, the comments or videos and so on right. For example, we want to connect two users based on the set of videos that these two users have viewed together or the set of videos on which these two users have commented together right.

So, there are multiple ways to create a network. So, then the question is what would be the appropriate network construction mechanism should we consider video and based on the common watching property we will connect to users or we will connect to users based on the comments right say similar types of comments. So, we do not know which you know which method would turned out to be useful.

So, in this chapter I will try to give you a glimpse of you know how people generally you know address a particular problem right based on the heterogeneous data that we have right. So, first question is how can we model problem statements into networks. So, the statements is imprecise it is not very concrete statement is complex right can we employ different network algorithms from earlier chapters on the modeling on modeling the network.

For example, when we employ a network algorithms and how we employ network algorithms right can we. So, can these network methods be used for a particular real world use case for example, malicious activities in online social network, detecting malicious activities right,

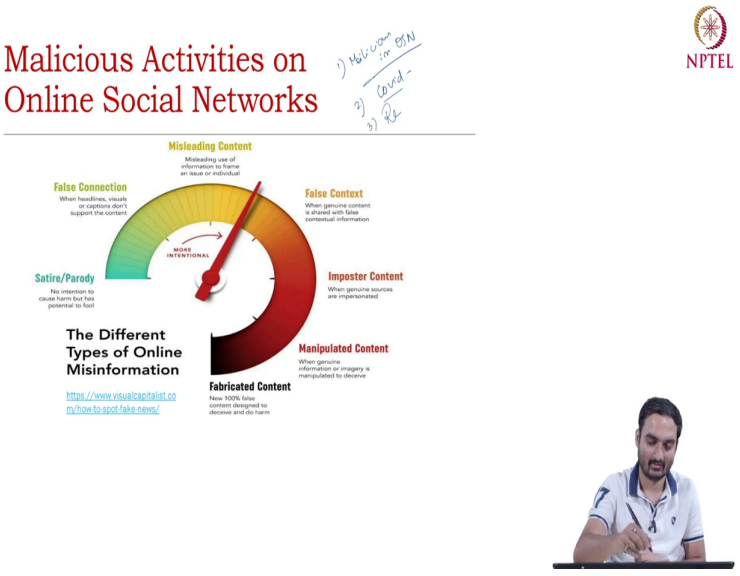
identifying the spread of this deadly COVID-19 pandemic right, designing recommendation system and so on.

So, it is not always the case that if you are given any problem statement you directly start you know coming up with the network and then try to solve it. You may want to use a classical machine learning techniques to solve the problem, but you need to understand that ok for this problem I really need a network construction because things are dependent and if we do not see the entire picture as a whole we would not be able to understand the dynamics right.

So, network plays an important role here to understand the relationship between entities and you know give you the overall picture of the I mean of the entire scenario. So, what we generally do, we divide any use case into two parts we generally understand the background we look at the literature, we also look at you know the previous methods have already been proposed on that particular problem statement.

And then we also look at the modelings and you know the models that people have proposed so far. And then we try to come up with a mythology and then try to evaluate the methodology and discuss the pros and cons of the method ok. So, this is generally the way we attack a problem we approach a particular problem ok.

(Refer Slide Time: 06:19)



So, let us start with the first use case. So, in this chapter we will essentially discuss three four use cases the first one would be you know understanding malicious activities malicious

activities in online social network. The second one would be understanding Covid-19 spread right we will consider we will look into this SIR SIS kind of models and see whether those models are really helpful.

Ah in Covid -19 spread modeling and the third one is a recommendation system. I will not go into the details of all these applications I will try to again give you a overview of this because you know this I mean each of these topics itself is a course altogether. So, therefore, I will try to give you a brief of each of these applications so, that you will understand how to address a problem right.

So, let us start with identifying malicious activities and this is very close to me because I mean my lab has been working on these problems since last 3-4 years and we have done a lot of work on identifying you know fraud users, fake accounts, fake news, misinformation, hate speech and so on. All this I mean all these malicious activities come under you know the overall umbrella of cybercrime and cyber safety ok.

So, social network is kind of a gift to us because we use social network freely with a lot of freedom without thinking much about the repercussion of our content and so on. Whereas, this freedom of speech or this freedom has often been misused by malicious users to you know write fraud comments fake comments fraud activities and so on and so forth.

So, if you look at the statistics if you look at the activities the I think 10 to 15 percent of activities are actually malicious activities right harmful activities ok. So, therefore, it is an utmost interest to the social network administrators Twitter, Facebook to stop the spread of such malicious activities right.

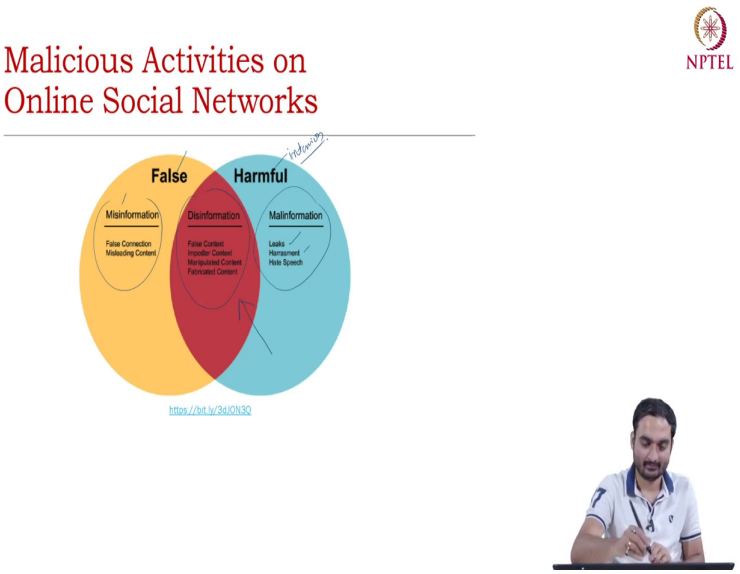
So, I will try to give you an overview of two such malicious activities right; one is in terms of you know identifying something called sock puppets right. We will see how sock puppets are being created by you know by administrator not administrators by controllers they are called puppet masters.

So, how these sock puppets are being created and how they you know manipulate users opinion on social media and particularly how social network analysis techniques can be used to detect such sock puppet accounts right. And then we discuss another malicious activities which basically happens through something called black market services right now what are black market services.

So, say for example, you want to promote your account you want to promote some event, but you do not have enough reach right. So, you approach this black market service there are many such black market services available online we you approach them and you basically you know tell them to boost your account to boost your post right and what they do.

They would systematically boost your event boost your account in an inorganic manner which would be apparently difficult to detect right. But if you understand the network analysis techniques well you would be able to know that ok how things are controlled by black market agencies and how you can identify such collusive activities which basically are controlled by black market activities a black market services.

(Refer Slide Time: 10:31)



So, we will talk about how to detect black market activities right in a systematic manner using social network analysis. So, if you look at the you know the overall spectrum of you know misinformation fraud, fake. So, there are essentially two dimensions of it; one is whether a particular information is false right and the other is what was the intention behind spreading a misinformation right.

So, false information whether it is in a false and the other dimension is whether it is intentionally created to harm somebody right. So, based on these two dimensions right falseness of the information and intention right, we try to categorize fake news. So, fake news is a very broad I would say the broad term right. In fact, people say that fake news is an oxymoron because a news by default is always true.

So, fake news the word fake and news should not appear together side by side anyways, but I am not going through the details of the definition of what is fake news on or not. But, over the years people tried to come up with a strong definition of fake news and they said that let us not call it as fake news let us call it as either misinformation or disinformation or malinformation.

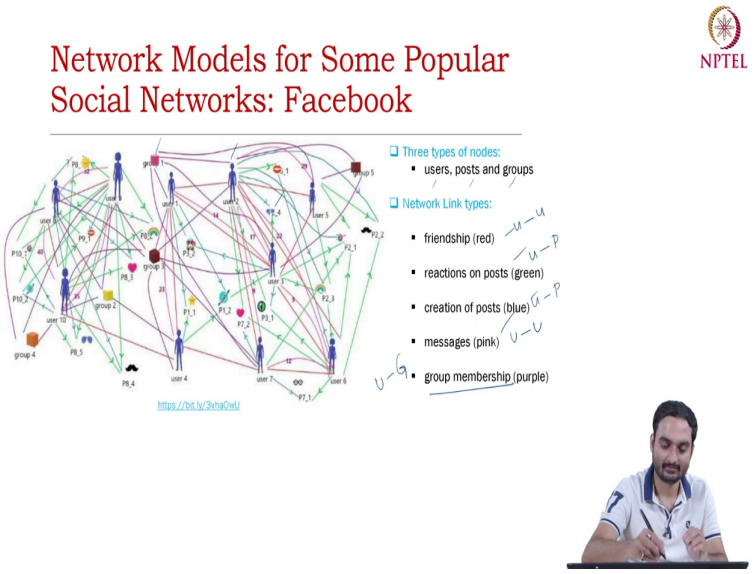
So, this misinformation is basically those information or those news which is false of course, but false connections with some misleading content ok. But, when people write misinformation there may not be a bad intention behind it right. Say for example, you report that you know there were 10 people who were killed yesterday at the in some riot right.

So, you know instead of right. So, let us assume that the actual number is 5, but you mistakenly wrote as 10 right it does not matter because if you look at the repercussion does not have that repercussion right. So, those kind of you know news articles are called misinformation those are false, but the intention is not that bad whereas, in the other spectrum you have something called malinformation which is a kind of a leak.

For example if you look at pornographic videos right where people replace someones face with somebody else and basically try to malign someone right. So, harassment all these video leaks these are false, but these are intentional right to harm somebody or to harm a group of people say a minority people right or a minority group or a society as a whole right. So, those are called malinformation.

In between misinformation and malinformation we have another type of fake news which is called disinformation. So, disinformation is something which is false as well as spread intentionally right and whenever we refer to fake news, we generally refer to disinformation because it basically considers misinformation and malinformation together ok.

(Refer Slide Time: 14:15)

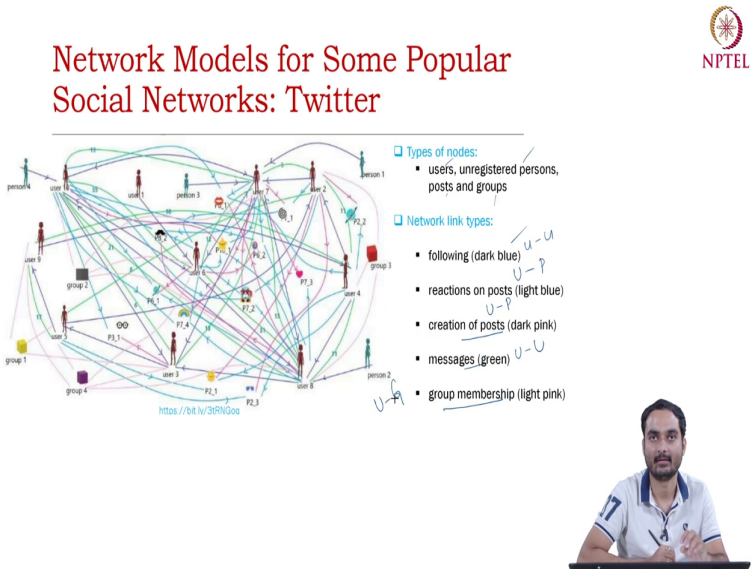


I mean I can speak a lot about misinformation fake news etcetera. But let us not go into that details. If you are really interested you can look at some of the papers published in last 3-4 years there are there are a lot of papers on misinformation fake news. So, in this context if you are given a social network say a Facebook network or a Twitter network route right. How do we use the information to create a network? We have discussed this thing in the previous lecture, but let us look at it again say Facebook right.

So, in case of Facebook you have three types of nodes users, posts and groups right users, groups and say post right. And you have different types of links you have friendship link, you have reactions to a post link, you can also; you can also consider creation of a post link, you can also have messages link if somebody sends a message to someone else you can think of a link.

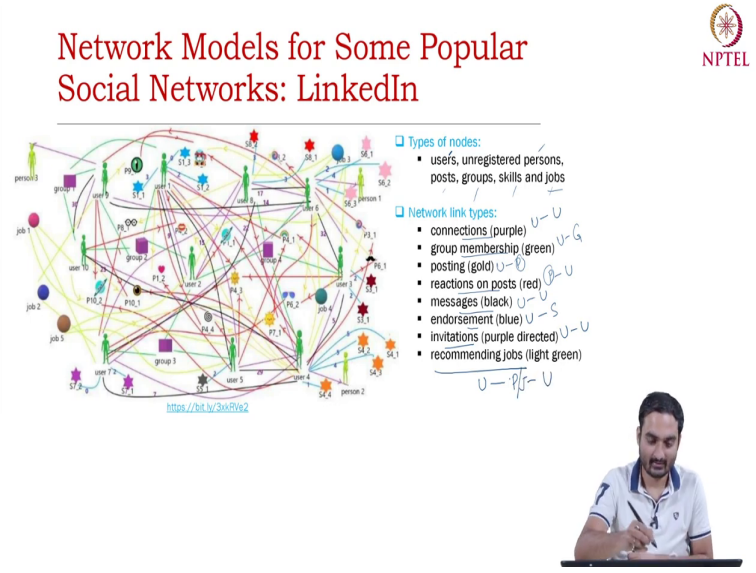
So, this is user to user this is user to post, this is user to post, this is user to user and this is group membership user to group right.

(Refer Slide Time: 15:54)



Similarly, in case of Twitter you can think of users right unregistered persons post, groups as different types of nodes heterogeneous nodes and you have following right; u to u reactions on post U to P creation of post U to P messages U to U and group membership U to G ok.

(Refer Slide Time: 16:30)

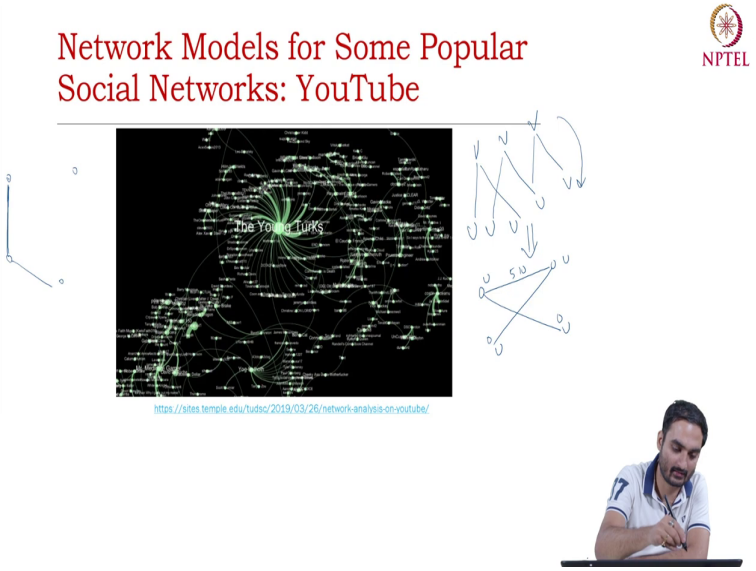


Similarly, if you look at professional network like LinkedIn you can think of nodes like users right. Again unregistered persons post, you know groups, skills and jobs and then you can have links like connections between user to user group membership between user to group

then posting between user to post reaction on post between post to user messages between user to user.

Now, this post can be jobs can be any other normal post. It can be endorsement between a user and a skill right, it can be invitations between user and user, it can be recommending a job between user to post the post is a job and then to user right.

(Refer Slide Time: 17:39)



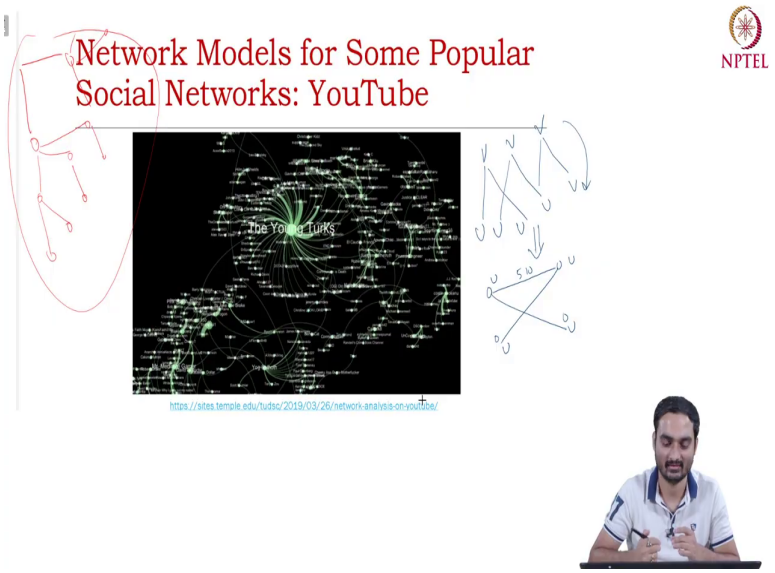
If you think of a network or online media like YouTube right as I was mentioning there is no explicit network right Twitter, Facebook, LinkedIn there are explicit networks. But Twitter you there is no explicit user interaction network. So, what you can do? You can think of you know different videos right.

And you can create bipartite graphs, user cross videos and from this bipartite graph you can think of a unipartite graph a projection to the user partition where nodes can be users. And if two users have watched at least one video together then you can connect this to two users. And if they have watched multiple videos then the number of videos would be the weights of the edges say 5 10 and so on.

So, in this way in fact, this is one way another approach can be say you have users and users have written different comments right you look at the comments. And look at the similarity between comments and then based on that you connect two users right and the weight of this link would be the similarity of the comments that they have posted right. So, this can be

another way to create a network similarly you can also think of a video centric network where say you have a video right.

(Refer Slide Time: 19:30)



And you have comments right level 1 comments and then somebody has replied to this comment level 2 comments right. So, you can think of a tree like structure under a particular video. And then you can connect multiple videos to another level where this connection happens based on the topical similarity of videos for example, both the videos are on Indian election. So, you can connect videos together and you can think of another heterogeneous tree like this ok.

So, what I basically wanted to say is that there are different types of graphs that one can construct from a data set. Now, which graph is useful for which purposes this would be dependent on the application that we have it also depends on you know your previous expertise right; how quickly you understand which network would be useful all depends on you know how much work you have done in this particular field in the past right.

So, in the next lecture we will discuss you know one application as I mentioned earlier we will try to identify shock puppets right dummy accounts which are controlled by different puppet masters. And these are fake accounts, fraud accounts they are used for malicious purposes. And we will see how people have used network analysis a bit of natural language processing techniques for detecting shock puppets ok.

Thank you.