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

Chapter - 04
Lecture - 20
Lecture - 01

Welcome to the course on Social Network Analysis once again and today, we will discuss Chapter 4 which is on network analysis particularly link analysis. So, far we have looked at how we can measure a network from different perspective in chapter 2; in chapter 3 we have seen how we can generate synthetic networks which actually look like a real world network.

And, there we have observed that when a network is formed a network when an edge is formed that is not random right it is not I mean the random graph model that we discussed in the last chapter the real world network actually does not follow a random graph model right random network model. So, edges are not formed in random. There are some you know some ways some systematic ways through which edges are linked between nodes.

And, we have seen in case of say Barabasi–Albert model price model this kind of models that edges are basically you know formed based on something called preferential attachment, right. When you basically want to join a particular network you may want to choose a node which is already of high degree for example, or high already has hyper stage ok.

So, now the important question is how do we quantify an edge, ok. Can we come up with different ways to quantify edges again from different perspective and can we see how these edges can be useful for other applications, right? Of course, in the in chapter 6 we will talk about link prediction which is definitely an application of edges right, but link prediction is basically you know it is more of a recommendation system.

But, if you think of network and you know edges right when you quantify edges what are the other things that you can actually infer, ok. Can we say that ok look these nodes are of high importance because of these edges or when we measure importance of nodes right, we should look at these edges not this set of edges and so on and so forth.

PageRank is something that we discussed and if you remember PageRank formulation, PageRank systematically looks you know takes into account edges, the directions of pages

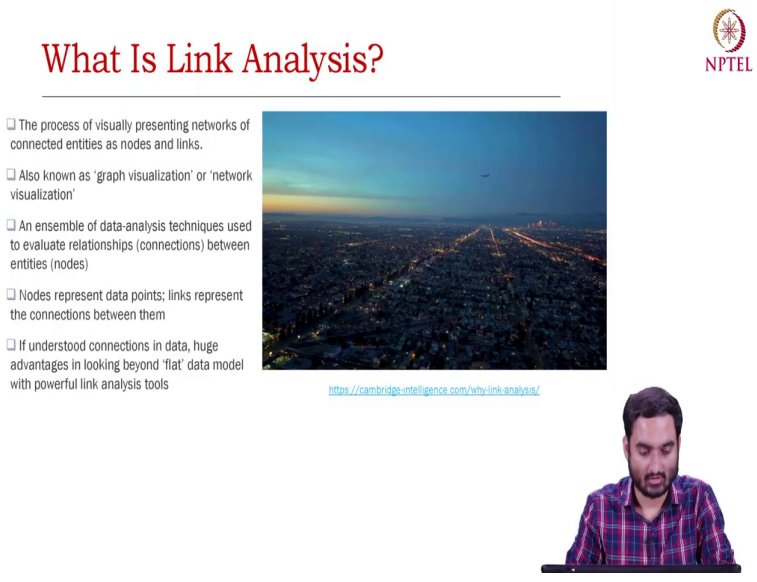
right and then basically prestige moves from one node to another node through edges and so on and so forth, right.

So, in this chapter in this link analysis chapter we will basically try to learn various ways to characterize an a particular link or edge in a network. We also you know connect the network science you know network science domain with the social science theories in general and then we apply these concepts in you know various other applications, ok.

We also try to you know quantitatively measure the structural properties of a network you know via different edge connections. We will learn how you know networks how links and you know edges help in measuring importance of nodes. In a network, we will discuss PageRank we will discuss again you know in a different manner, right. We will discuss DeepRank, SIEM Rank and different other ranking mechanisms that you can think of based on the importance of edge.

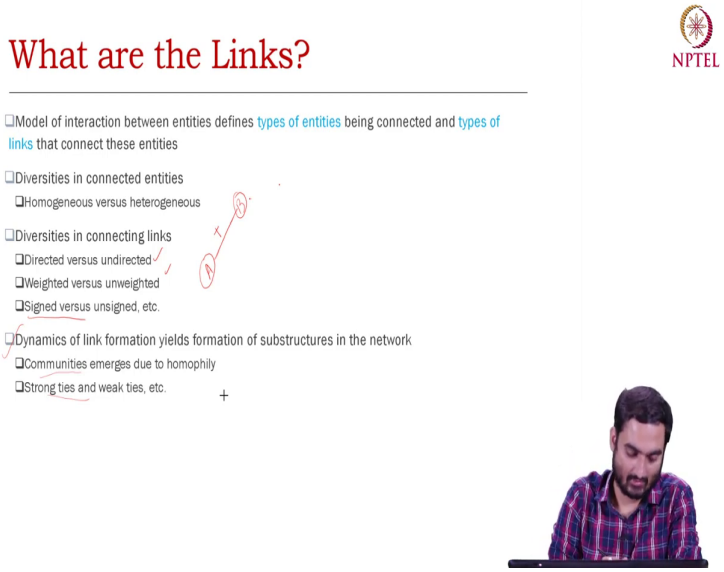
And, we will also look at you know we will deal with heterogeneous network where we will see you know a edges are of; edges are of different types, nodes are also of different types and then how can we use heterogeneous networks for say ranking purposes. We will discuss something called meta-path, right which looks like a path, but in the context of heterogeneous network, right. We will discuss all these concepts in this particular chapter.

(Refer Slide Time: 04:17)



So, it is basically I mean if you think of network analysis. It is basically you know a study to visualize a network in terms of links, right. So, and you know no of course, when we look at nodes right vertices we and we try to quantify the importance of nodes in a network we always take into account the links because links are ultimately entities which basically measure you know right and importance of a node or it basically help you understand that how certain prestige flows from one node to another node and so on, right.

(Refer Slide Time: 04:59)



So, what are the links? So, we have already discussed in the chapter 2 that links can be of different types it can be you know an undirected edge or link can be a directed edge right when you I mean if you think of an undirected edge is basically a two way directed edge. And, it can also be weighted, right? This weight of an edge basically quantifies how important that edge is.

Now, importance can be can be quantified in different ways. For example, in case of flow information flow importance can be the amount of inform information that you know have been have moved between two nodes through this particular link, right. In case of citation for example, say citation network between authors the weight of the edge weight of an edge indicates the number of citations that an author actually you know gave to another author right through scientific papers.

Interestingly, edges can also be signed ok. Sign here means it can be positive edge it can also be a negative edge or it can be a neutral edge, right. So, what do we mean by a signed edge or

a positive edge, right? So, when I say that ok look there is a; there is a; there is a link between A and B and the sign is positive it basically means that A and B have you know friendship relationships right for example, in case of friendship network, right.

In case of say in case of say you know selection networks where nodes are nodes are individuals who actually vote and I mean nodes are of two types – one types of nodes are individuals who vote for certain candidates and other nodes are essentially you know the candidates themselves, right. So, then in that case the link indicates whether an individual you know vote whether an individual votes for a particular candidate or against a particular candidate and so on, right.

You can also think of links and signs in case of say e-commerce network right when you review some e-commerce products, right. You can write positive reviews, you can like the product, you can dislike the product. So, like corresponds to positive edge, dislike corresponds to negative edge in say e I mean platforms like YouTube also you can like a video, dislike a video right a platforms like epinion slash dot there you actually you know specify your friends or your enemies explicitly, ok. So, we will discuss whole bunch of things on sign network.

In general, when we talk about a network it is in general unsigned, unsigned undirected, but depending on the situation it can be signed, directed and weighted, right. And, we have seen in the last chapter that the dynamics of link formation is not random there is always a preference based on which links are connected to different nodes and because of the connections because of this preferential attachment, right you can think of you know emerging communities or emerging clusters.

For example, when we see a network actually exhibits homophily kind of property where nodes are connected with each other based on some sort of common properties right common characteristics, right. And, in that kind of setting we will see that this the communities right clusters will emerge we will talk about community analysis in the next chapter in chapter 5 and we will discuss in this chapter something called strong ties and weak ties.

So, strong ties or strong links are those links which actually connect nodes with high similarities, right whereas weak ties are those edges which actually connect nodes which are nodes which do not interact frequently. For example, say you know I mean if you think of a again user interaction network so, the close friends right the colleagues or a school friends

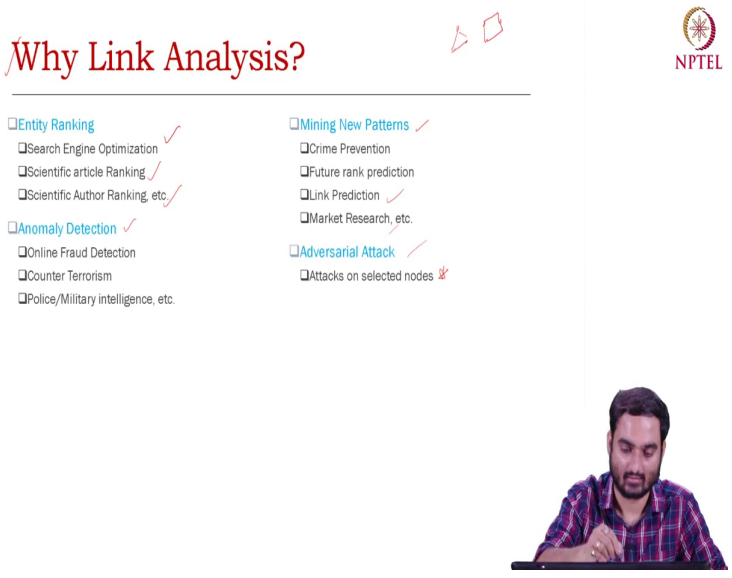
they interact with each other quite frequently. So, they are basically connected through strong ties.

Whereas, you know of a student from one school is connected to student in another school through weak ties or as or an individual in one company is connected to individual in another company through weak ties because they do not interact with each other quite frequently. Now, we will discuss here in this particular chapter that weak ties play a very important role in many aspects.

For example, you know we will see in the next chapter that weak ties are important to detect community structure right because when we break this weak ties the clusters of a graph actually emerge. In case of information flow also weak ties play an important role because essentially weak ties connect a different components of a graph different communities of a graph.

And, if you want to spread one information from one community to another community, you choose an edge which act like an a weak ties bridge right. So, we discussed all these concepts in the last few lectures – bridge articulation point and so on and so forth, ok.

(Refer Slide Time: 10:51)



So, now, look if you think of applications of link analysis right apart from link prediction that we will discuss later, link analysis can be useful for searching purpose, right. We will discuss PageRank algorithm once again, DeepRank algorithm where links play an important role and

ultimately when we use PageRank, we essentially use it for what? We basically use it for ranking nodes right and ranking nodes is basically useful when we search something.

For example, you type a query Harry Potter, right. Now, the this particular term is present in many you know web pages. So, how do you rank web pages? How do we rank relevant pages so that users will so that the user will you know find the relevant web pages at the beginning and the non relevant web pages at the end, right?

Now, ranking plays an important role. So, you can think of page rank here and say you already you already measure the importance of nodes importance of web pages based on page rank and some query comes in, you just fetch all the nodes which are relevant to the particular query and then you rank them based on the page rank.

You already have the you have you have already got the rank, so, then you use that ranking to arrange web pages on the user feed or search feed, right. This is useful for scientific article ranking because if you also again think of page rank on the citation network, right and page rank is a very simple measure, but you can think of more sophisticated ways again based on links and you can also rank scientific articles right either in general or with respect to a certain field, ok.

You can say that ok let us look at page rank and let us use page rank to rank all the scientific papers in a citation network and then you and then given a particular field say AI, right you retrieve all the papers related to AI and you then rank those pages based on based on the PageRank, ok.

Similarly, you can rank scientific authors right researchers; a researchers are again connected through citations; researchers can also be connected through collaboration for example, and then you can invent your own ways try to measure the importance of authors, again based on links and then you rank them.

Anomaly detection we will discuss later that this link analysis plays a very important role very crucial role in anomaly detection. So, when we talk about anomaly detection, fraud detection say you know counter terrorism, terrorist detection. So, some I mean uses of this in say police or military services, right. We will see that certain links I mean in general in general most of the links are normal, right in the sense like they actually connect nodes of similar types, right.

Of course, some of the links can be weak links some of the links can be strong links, but there are links which basically connect nodes of a completely different properties for example, right and those links are suspicious links. So, link prediction link analysis is also useful in case of say transaction network where edges indicate different transactions right say between seller and buyer and it may happen that some of the transactions are fraud, right.

So, how can we come up with metrics to quantify again edges or links based on the extent of you know fraudulent activities, right with respect to links again. We will discuss some of the quantities in the in one the chapters. It is useful for mining new patterns because ultimately say if you think of motives right 2 3 3 node motives or say 4 node motives, right. Ultimately it is a pattern mining approach, right.

You try to mine you try to analyse the network and try to come up with structures and sub structures which do not frequently appear for example, right. It can also be a part of anomaly detection, but in general finding out new patterns from a network would be very useful for link prediction, for again you know some sort of market research, you know crime prevention and so on.

Links are very important to attack a particular network. Now, what do you mean by attacking a network? So, think of think of a network, right and you and say assume that you know you want to you have some algorithm to detect the communities of a network, right. Now, your task is to you know add or delete edges right or nodes in such a way that the community detection algorithms start performing badly.

So, you are attacking a network, you are disrupting some structure of a network – in this case the structure is community structure you try to disrupt the structure of the community structure of a network with the with the modification of network. Now, people may ask that it is very easy because you keep on adding edges right between two community for example, and automatically the inter community edges right that will become a sparse and sparse and sorry, that will become dense and dense and as the inter community edges you know the inter community portion become dense and dense it would be difficult for a community algorithm to detect communities, ok.

But, it is not that easy because you are not allowed to add links or add edges or delete edges right as for your choice because adding an edge actually incurs some cost. Think of a social network right say Facebook friendship network. Adding an edge means what? Adding an

edge means are you are making a friend friendship between two users, right. So, you have to motivate you have to convince these two users to become friend and that is costly, right.

So, and therefore, you can think of it as an optimization problem where you have a budget right and every addition or deletion of an edge incurs some cost, right. So, within the given budget how can you manipulate the network by adding or deleting edges so that with minimal change in the network structure your purpose will be fulfilled. Your purpose in this case is to disrupt the community structure, right.

That is more from I mean you can think of it more from a attackers point of view, right say you are an attacker and you want to attack the community detection algorithm or the network structure itself so that the algorithm you know starts performing badly. From a defenders point of view also this is very important because let us say you know you already you already have an algorithm for detecting clusters right and you had to release the algorithm publicly.

So, all the attackers know that this algorithm is used for community detection right. So, attackers will start attacking your algorithm your network, right. Now, you are a defender, right. So, you also want that you also want to add or delete edges so that attackers will not be able to figure out which edges and nodes have been added or deleted, right.

At the same time they also get fooled, right because what they would do? They would again run the algorithm and let us assume that the attacker is very smart, the attacker knows the inter network structure, right. So, attacker will run the clustering method community detection method right, but you know that these edges have been added ok. Therefore, the algorithm will definitely perform badly on the attacker side, right. So, attacker will not be able to understand the actual community structure ok.

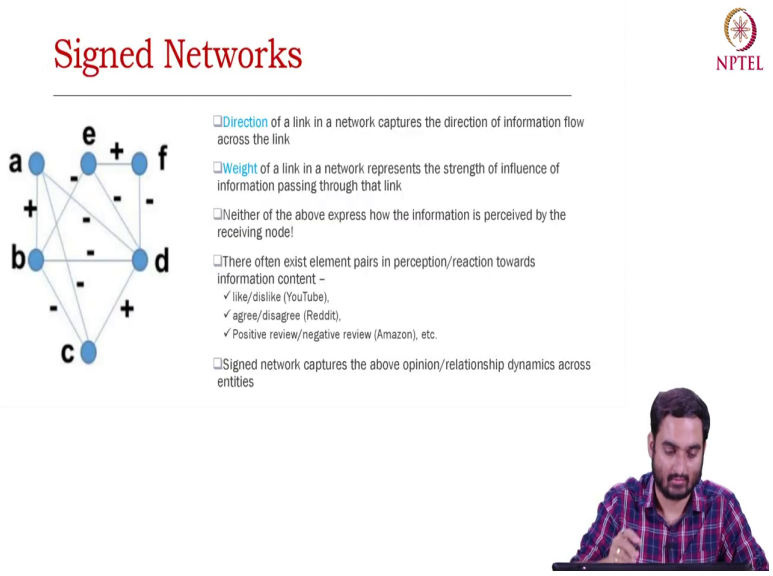
Now, here is the same question people may ask that since you have a complete control on the network structure you can basically keep on adding edges and automatically the attacker will be will not be able to detect the actual cluster because right because that is trivial. But, here again the catch is you cannot add edges you know randomly or right I mean you cannot add multiple edges because when you start adding edges the attackers are smart.

They also understand that these edges should not have formed here, should not have been there right. But these edges are there, therefore, there is some problem. They will definitely understand it. So, here the idea is how you can you know systematically add or delete edges

in into the network so that the attacker will not be able to identify the modification at the same time they also get fooled, ok.

This is a very interesting topic the these days which you know which a lot of you know research communities in so, in network science have been working since last 2 – 3 years.

(Refer Slide Time: 20:19)



So, in the in the next chapter in the in the next lecture, we will start discussing on signed network. We will discuss different properties, we will discuss something called balance theory and status theory, and we will try to relate because this part is more of a social science kind of part we will try to relate you know the social science part of it with the network science part of it, ok.

Thank you.