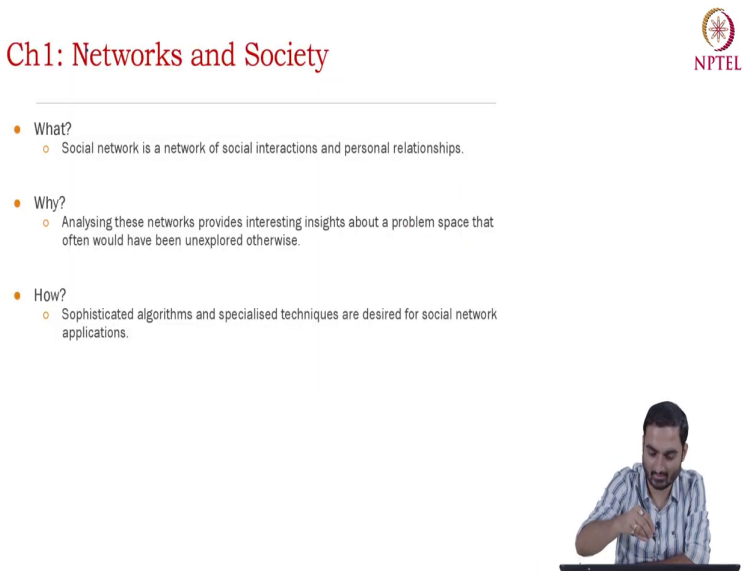
Social Network Analysis
Prof. Tanmoy Chakraborty
Department of Computer Science and Engineering
Indraprastha Institute of Information Technology, Delhi

Lecture - 74
Social Network Analysis Conclusion

Hi everyone. Welcome to the last part of this course. And first of all you know huge congratulations to all of you to you know finish all the 10 chapters. Hopefully you have you know thoroughly enjoyed the chapters and you know I also hope that I was able to you know convince you motivate you why this is an important course. Not only in terms of you know academic purposes, but also for your research purposes right.

And whatever we have discussed was just you know small part of a huge huge area that is evolving you know every day. So, let us together conclude this you know course.

(Refer Slide Time: 01:09)



So, what I will do today. I will briefly talk about you know the gist of all the chapters that we have discussed right here. So, we will try to answer 3 questions, what? Why? And how? Right for individual chapters. So, let us start with the first chapter. So, first chapter we discussed networks and society.

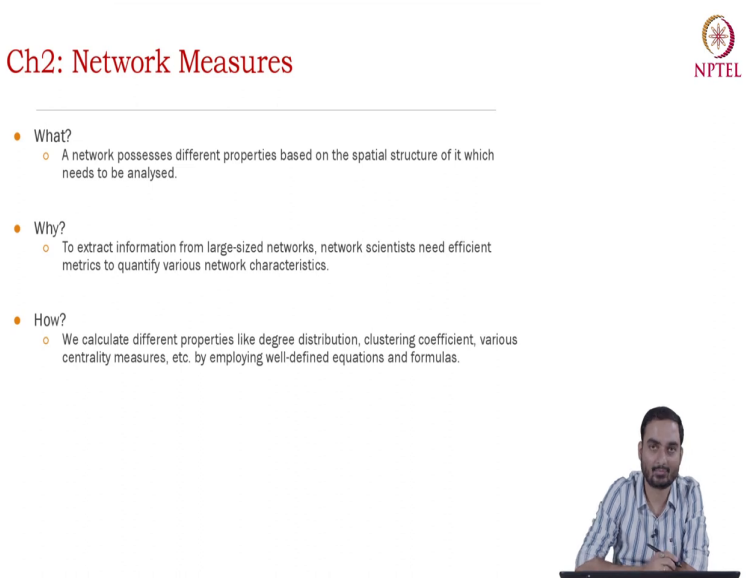
So, this was basically an introductory chapter. Where we basically wanted to understand how you know social network can be visualized, can be conceptualized using a you know any data

set that is given to us right. And why this was important? Because analyzing these networks provides you know interesting insights about the problem space that would that often would have been unexplored otherwise.

So, the beauty of our network structure is that. It gives us you know strong understanding about the interrelations and this interrelations if you do not know the interconnections interrelations between entities objects right as a whole right. So, the problem would be that we would not be able to understand the big picture right. Micro level statistics is ok. You know if you look at the macro level statistics you would not be able to understand if you do not have the introduction you know interaction behavior.

So, therefore, we decided that we will discuss set of sophisticated algorithms to understand the social network structure.

(Refer Slide Time: 02:43)

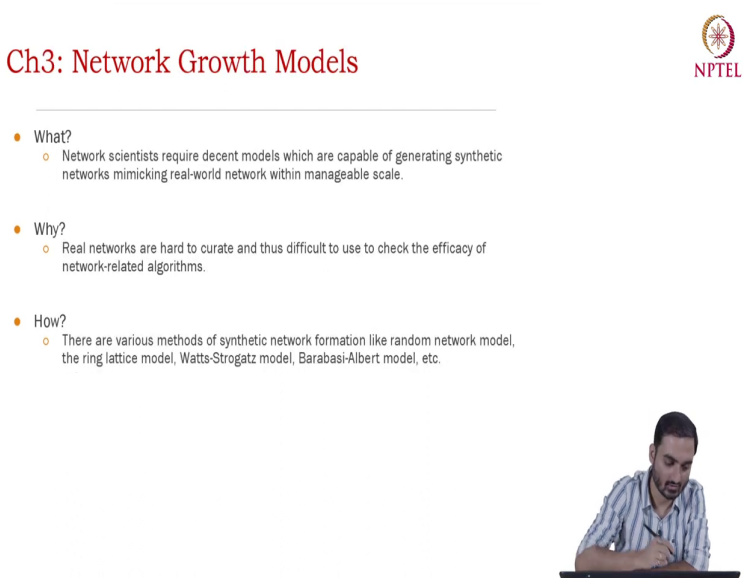


In chapter 2, we discussed different measures right. So, a network possesses different properties based on the spatial structure of it which needs to be analyzed. And why this is important? This is important, because to because we wanted to extract information from large scale networks.

Now, we wanted to understand micro level mezzo level and macro level statistics. And we discussed various types of properties node centric, link centric right. Sub graph centric, even graph centric properties right. So, hopefully you remember a metrics like you know page rank

right. A clustering coefficient, a degree centrality, between a centrality, eigenvector centrality right, a car centrality and so on and so forth.

(Refer Slide Time: 03:37)



In chapter 3, we then started discussing on you know how network evolves over time. The growth pattern of a network. Now, why this was important? This was important because if you see how I mean sometimes what we do to analyze certain things we scrape data right. Now, scraping Twitter for example, or scraping Facebook is extremely challenging right.

So, if we can have a some sort of synthetic a process right. Some sort of you know synthetic process to generate network and then we apply all our algorithms on the synthetic networks and then if those algorithms perform well. You know then we can basically map I mean we can use those algorithms for real world data set. So, in order to create a real world like network structure we tried to understand different types of network generation models.

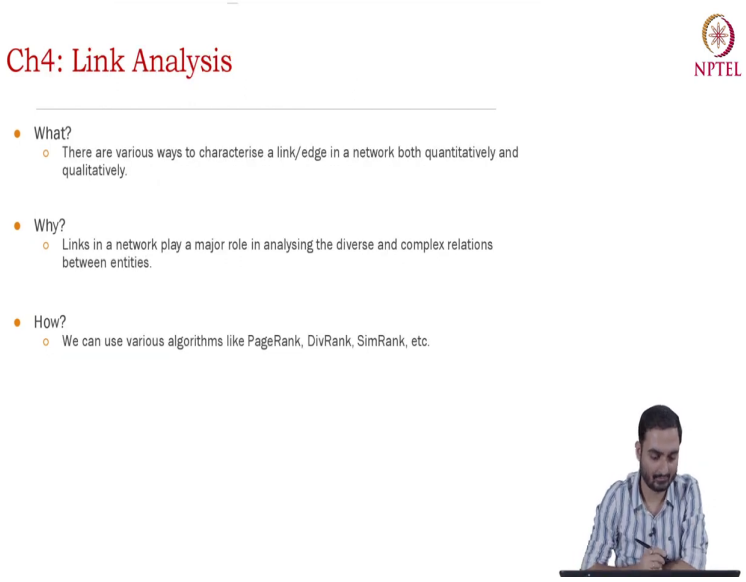
You know as simple as models like random graph models then we have looked at sophisticated models like Barabbas Albert models right; which were which was basically inspired by say price model and even you know other preference level models discussed earlier.

At the same time we also discussed like you know what is topic models and so on and so forth. And we try to realize that you know if we generate networks using this process right.

Would it be possible to say preserve properties like clustering coefficient. Properties like you know power law degree distribution.

Properties like you know 6 degree of separation and so on and so forth right. And we have seen that few models preserve few properties other models preserve few some other properties, but there is no such universal model which is able to preserve all sorts of properties ok.

(Refer Slide Time: 05:39)



Then we you know we moved into the you know details of links. And we discussed link analysis right. So, link is an important you know ingredient of a network links or relations right. So, and depending upon how links are formed. We see properties like communities, we see properties like say you know gateways or gatekeepers or say you know bridge between communities and so on and so forth.

We you know we got motivation from social science. We discussed strong ties, weak ties right, sign network, status theory, balance theory and so on and so forth. And then we try to understand how to measured an importance of a link. We looked at PageRank; we looked at DivRank and other types of measures like SimRank and so on and so forth.

So, these measures are really important. You know in cases like say epidemic spreading when you want to cut a network right. You a what would be the minimal way possible to disconnect

network right by removing a edges right. So, these kind of metrics are useful in those applications.

(Refer Slide Time: 06:59)

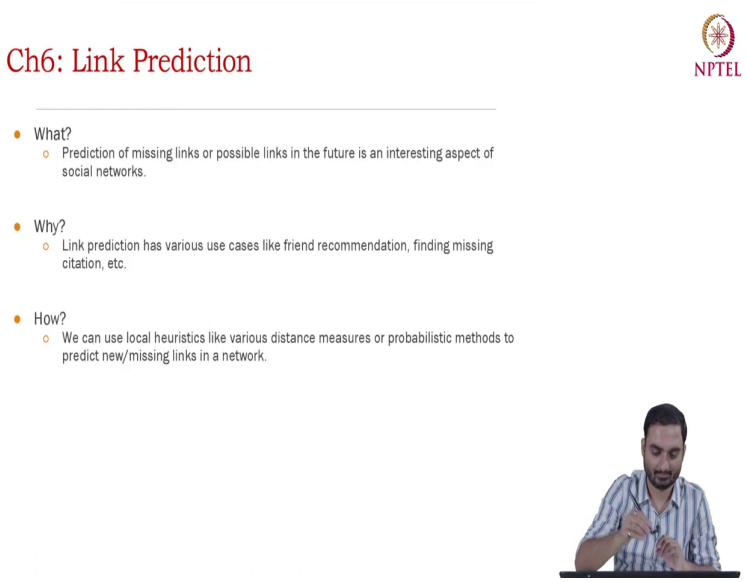
Then we looked at community structures or clusters in a graph right. Communities are roughly defined by sets of nodes, which are strongly connected internally and sparsely connected externally. We try to understand how to quantify a community. Community is an ill defined problem.

So, people have tried to come up with multiple metrics to quantify communities. We have seen metrics like modularity like permanence, chart right and we also looked at different types of community structures disjoint communities, overlapping communities, hierarchical communities, local communities and so on and so forth. We also looked at how to you know measure the accuracy of a community detection algorithms using metrics like a nmi, purity and so on and so forth right.

A community structure is extremely important in applications like recommendation system right, applications like viral marketing, applications like say churn prediction and so on and so forth right. So, we looked at metrics we also looked at how algorithms are you know designed to optimize these metrics to detect the community structures ok. Disjoint community is still easy to detect, but when it comes to overlapping community it is difficult.

And we also looked at generative models like you know b clamp types of models, which basically try to which try to generate networks by constructing communities underline and then we and as a byproduct we get the community structure ok.

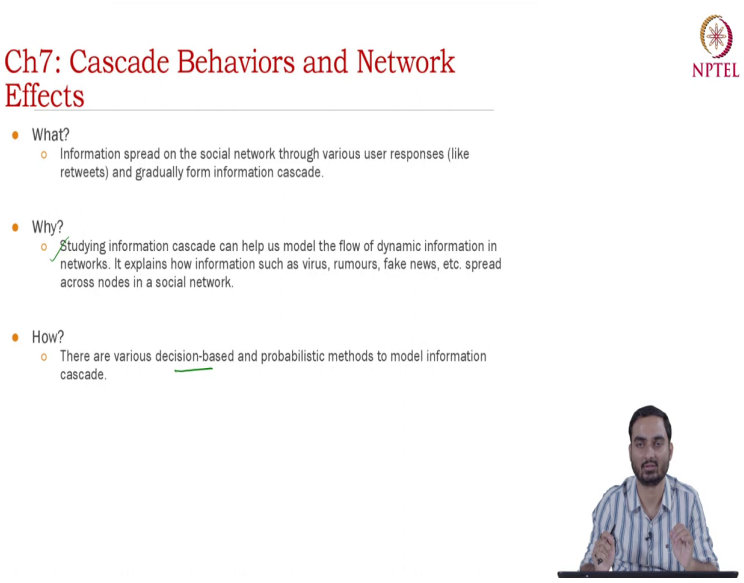
(Refer Slide Time: 08:48)



So, then we looked at link prediction and this is extremely important in the context of recommendation system. Missing link future link we try to understand the difference between these 2 applications. And then we you know we looked at different ways to predict links.

We looked at we started with simple statistical measures. Like Jaccard coefficient, atomic model distance to predict whether 2 nodes is 2 nodes are going to be connected. We also discussed even sophisticated methods right. I mean methods like you know methods, which use random walk based approaches right; random walk with restart particularly to predict links. We also looked at probabilistic methods like to predict links ok.

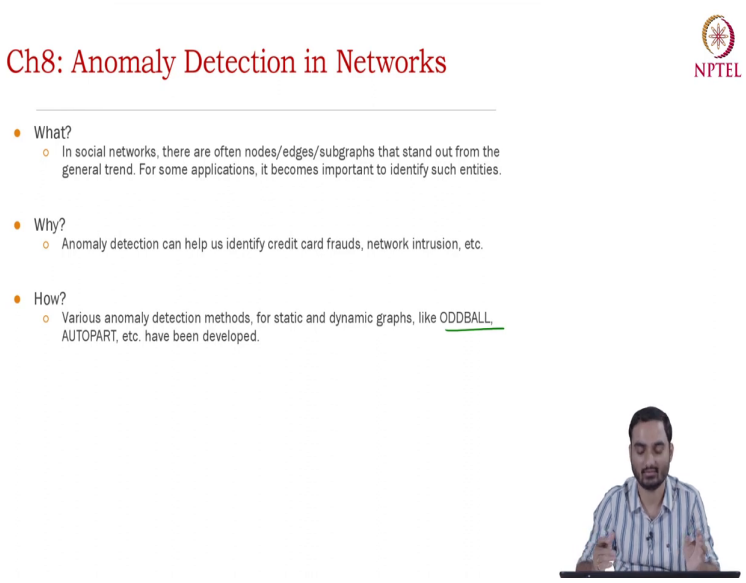
(Refer Slide Time: 09:55)



So, in chapter 7 we discussed you know how cascade evolves over time right how cascade grows. Particularly in the context of a epidemic spread right. We basically studied information cascade that can help us model the flow of dynamic information in networks. It explains how information such as virus, rumor, fake news etcetera spreads spread over the network.

And how can we come up with preventive mechanisms to stop the spread of such information such information on misinformation. It can also be you know epidemic. And we discussed 2 types of methods decision based methods and probabilistic methods to understand the spread of such information on social networks. Now, this is useful not only to understand the dynamics of such spread, but also do come up with preventive mechanisms to stop the spread of you know such information ok.

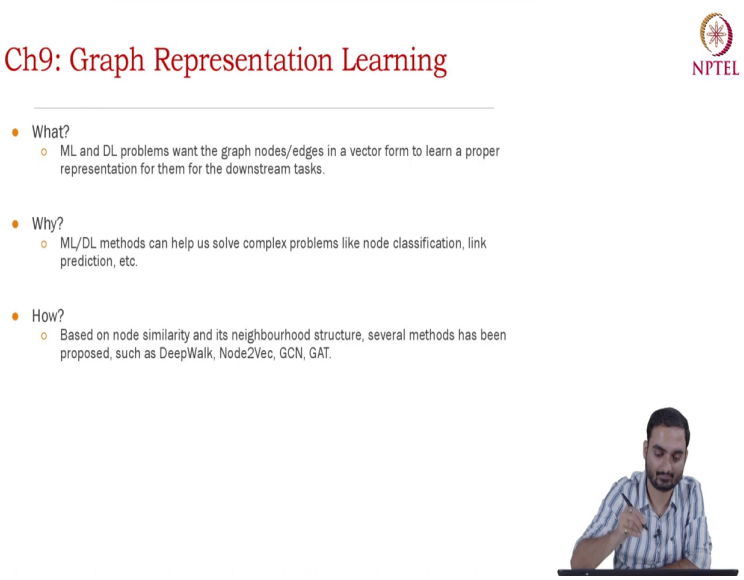
(Refer Slide Time: 11:08)



Chapter 8, in chapter 8 we discussed anomalous behavior right. We again anomalous structure of a network is an ill defined problem. So, it depends on the context the particular application that we are interested in. And we looked at anomalous nodes, anomalous edges, anomalous subgraphs, anomalous structure in static network, in dynamic network, in temporal network, anomalous structure with respect to the overall graph, anomalous structure with respect to a particular community and so on and so forth right.

And this is important for many applications like fraud detection, network intrusion detection and so on and so forth. So, we discussed method like odd ball, auto part. We also looked at generative methods like right and so on and so forth.

(Refer Slide Time: 12:07)



In chapter 9, we elaborately discussed graph representation learning. We started off by understanding neural networks basic neural networks. And then we looked at you know machine learning and deep learning based methods for graph reproduction learning. We looked at methods like DeepWalk, Node2Vec, GCN, graph neural network, Graph Convolutional Network, Graph Attention Network and how these methods are useful for coming up with representations of nodes and edges right.

And these representations are then fed to the downstream task. The task can be say node classification or link prediction right.

(Refer Slide Time: 12:57)

So, and the last chapter was basically application. And we looked at how all the other chapters that we have discussed. How those things can be applied to 3 applications. So, first one was more about frauds right sockpuppets and collusion collusive activity detection in online social network.

We also looked at recommendation system particularly pin sage right. How graph sage kind of methods can be used for recommendation system in platforms like pinterest. And then we also looked at different models for COVID-19 spread modeling right.

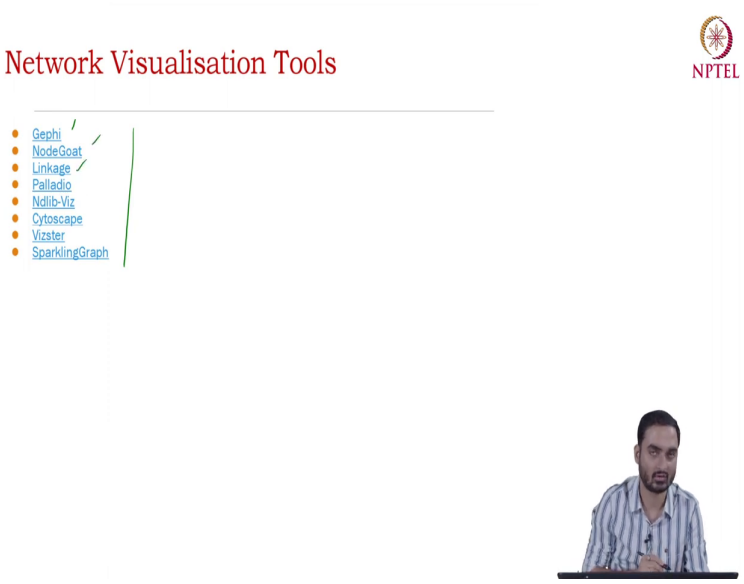
(Refer Slide Time: 13:40)

(Refer Slide Time: 13:43)

So, some additional resources if you are really interested in you know taking this thing forward and do some interesting research projects. These are the links that you should possibly look at. Here you will get lots of data sets, open projects, existing projects baselines codes and so on. So, the first one is of course, this SNAP, Stanford Network Analysis Project by Jure Leskovec. You will get whole bunch of data sets and you know codes that have been designed for different applications of a graphs.

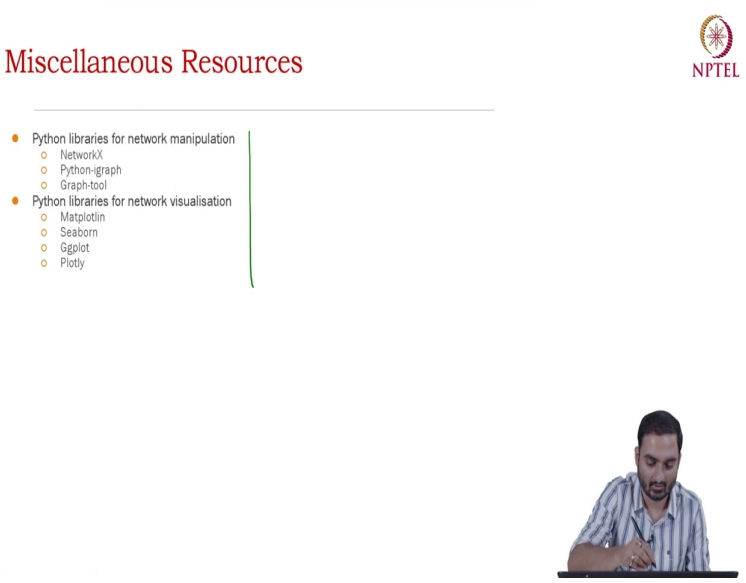
UCI data set is also very useful the for different purposes. You have other network repositories. You have this connect project where you will get a lot of a data set. This is another link this was nicely mentioned by Professor Aaron Clauset. You will also get different types of data sets and of course, the other git links including our links I mean our git link we also provide a lot of data sets for different applications, right.

(Refer Slide Time: 15:01)



These are some of the visualization tools that you can use to visualize your network. Like Gephi right, Linkage, NodeGoat and so on and so forth.

(Refer Slide Time: 15:16)



And of course, since you have already done a lot of you know your assignments. So, you are already aware of this Python packages right. These packages are really useful for different applications. So, this brings us to the end of this chapter. Hope you have enjoyed this thing thoroughly if you are interested to you know contact me.

(Refer Slide Time: 15:47)

Please feel free to you know send your email to this email id, chak dot tanmoy dot iit at the rate gmail dot com or you can even you know look at our website lcs 2 right. You will get you know other opportunities other you know other ways to connect our lab and connect with

me. If you I mean if you have any other suggestions to improve this course further, please feel free to you know get back to me.

Thank you very much and I hope this course you know has motivated you to take this thing as your next set of I mean for your next set of projects. Even this would also help you know start a new carrier or network analysis ok. With this I would like to thank all of you and once again congratulate you all to you know to I mean to be a part of this nice course. Hope to see you know sometimes you know physically.

Thank you very much.