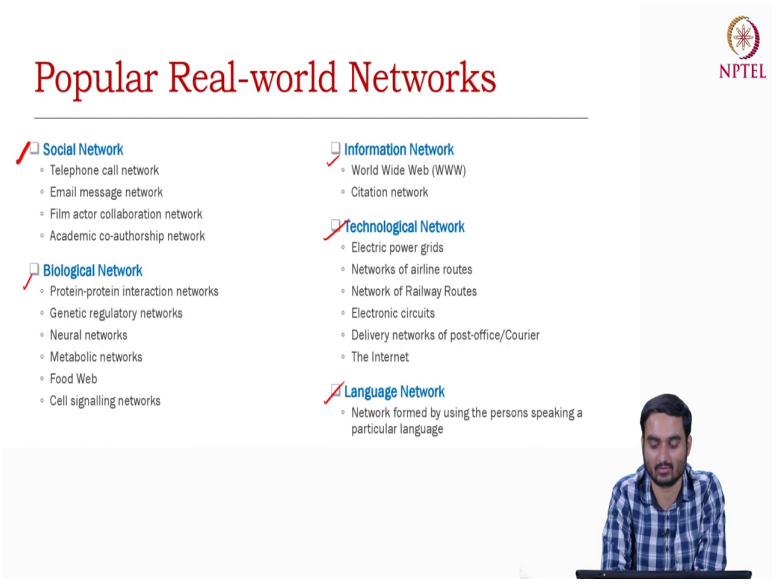
Social Network Analysis
Prof. Tanmoy Chakraborty
Department of Computer Science and Engineering
Indraprastha Institute of Information Technology, Delhi

Lecture - 03
Social Network Analysis Chapter - 01

We will discuss different types of you know real world networks; I mean the popular real world networks that we generally talk about right.

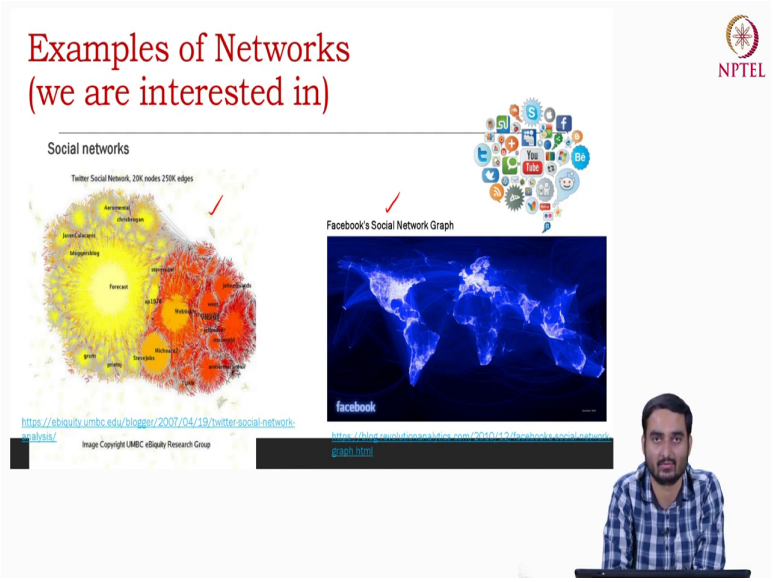
(Refer Slide Time: 00:39)



So, if you look at the literature you can see that broadly there are 5 types of networks that we talk about; social network right, biological network, information network, technological network and language network, ok. We will give examples of each of this one by one, but in general social networks you all know.

Biological networks I mean you can think of say protein-protein interaction networks or say you know neural network I mean interactions between neurons, food network and so on. Information networks include World Wide Web, citation network you know. Technological networks include power grid, airline network, railway network. Language network includes you know say word co-occurrence network and so on and so forth. We will discuss each of these one by one.

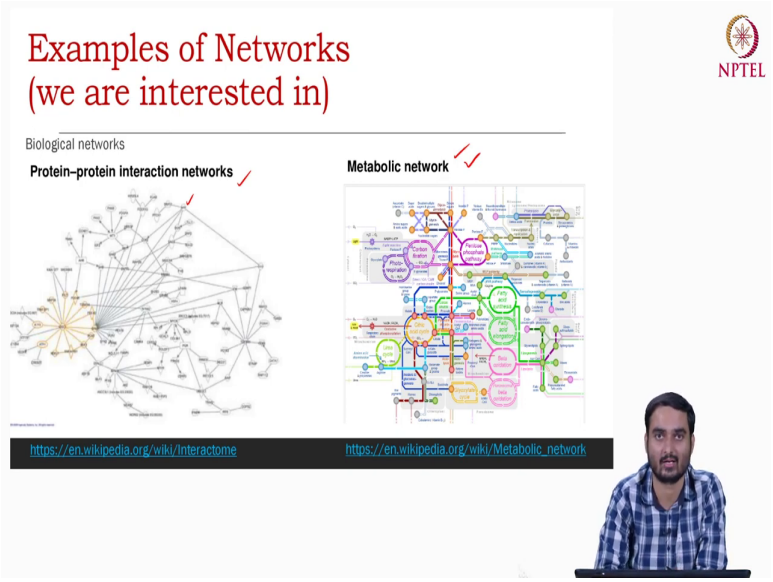
(Refer Slide Time: 01:27)



In fact, you know I basically tell my students that you know you can imagine you can think of a network from almost every data set, from almost every application right. Network is just an abstraction of a complex system as I mentioned earlier right. So, if you are given a simple problem right where you know a common person cannot see any network any notion of network right, you can think of a network out of it right and that is the beauty of this particular course.

So, let us look at social network right. So, I mean we have been discussing about Twitter network, multiple times follower following network nodes are users and links can be follower followings and so on. Similarly, we have Facebook friendship networks where nodes are nodes you know users and links are friendship relationships and so on and so forth and this is very obvious.

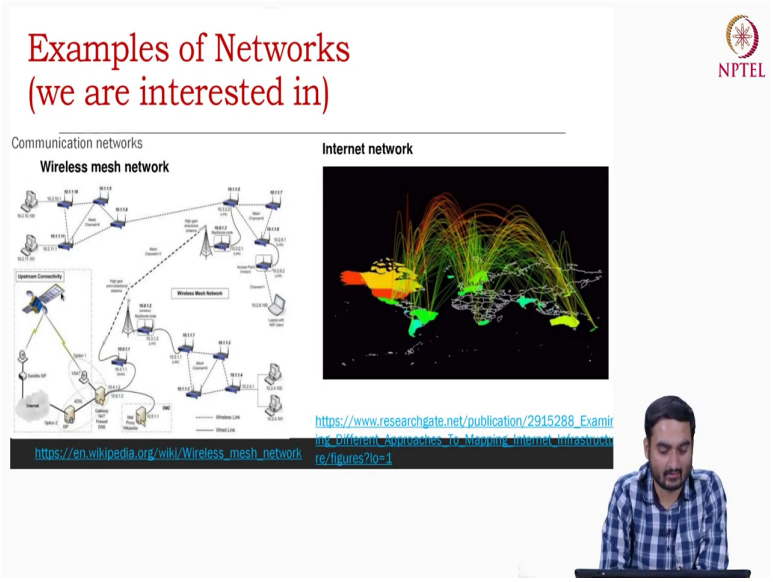
(Refer Slide Time: 02:20)



We have biological network for example, protein-protein interaction network where nodes are you know proteins and two proteins interact during a different metabolic processes in our body and you can connect proteins accordingly right. Similarly, we have you know metabolic network where you know basically it describes the relationship between you know small say you know metabolites right and enzymes proteins which basically interact with them during different you know biochemical reactions for example, right.

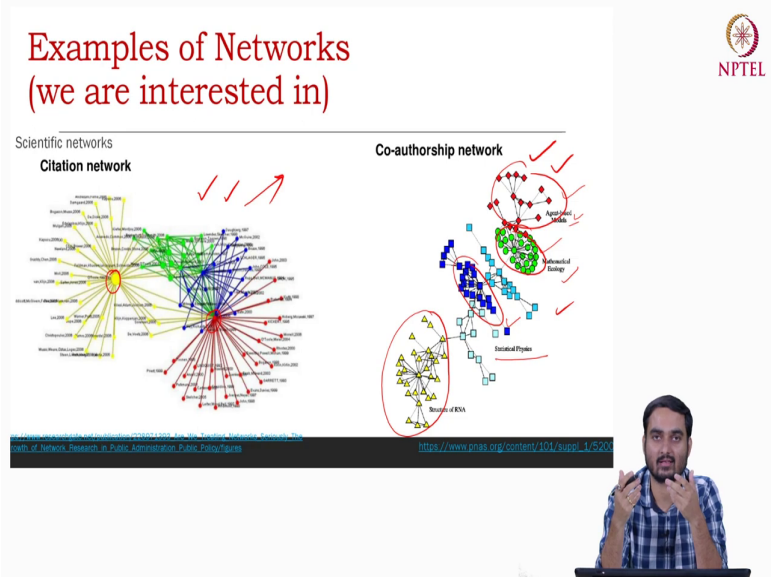
This is a metabolic network and there are tons of papers on protein protein interaction networks mostly in the bioinformatics computational biology domain.

(Refer Slide Time: 03:07)



You can think of other communication network like a wireless mesh kind of network, where say within an organization you can think of different routers computers and communications between routers. In fact, you can also think of communications through satellites that can also be a network, right. Of course, you have a big internet network where you know the whole internet is also considered you know broadly as a network.

(Refer Slide Time: 03:37)



We have scientific networks like a citation network we discussed earlier where papers are nodes and citations are the links, this is a directed network and we have a co-authorship

network where nodes are co-authors nodes are authors and if two authors work together right. You can think of them as co-authors and you can connect them right. So, interestingly if you look at citation network this is a example this is a small example of a citation network. Look at this node right look at this node they have high citations, right.

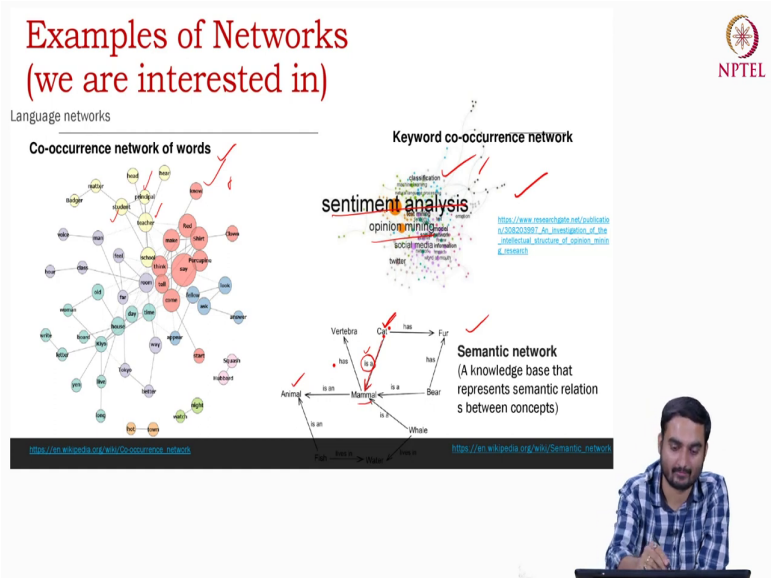
In fact, since this is a directed network, you have in degree you have inward edges and outward edges right. So, think of a paper which has a lot of inward edges. So, inward edges indicate citations, outward edges indicate references, right. So, if a paper has a lot of citations meaning basically the paper is very important therefore, people are citing it, but if a paper say you know has a lot of outgoing edges indicating a lot of references that paper is also important that papers might be a book that paper might be a survey paper or a literature review paper right.

We will discuss how this the notion of inward edge and outward edge basically you know interplay with each other and you can think of interesting metrics right, out of this the notion of directionality of an edge. Co-authorship network if you see the network right you see that you know there is a closed group, there is another closed group here nodes are densely connected right.

Here you see yellow nodes densely connected. Red nodes are also densely connected right and you know every such group has its own identity for example, this red group indicates researchers working on agent based models. So, green group indicates researchers working in mathematical ecology right, there is this blue group working on statistical physics and so on and so forth.

So, you see that a cluster multiple such clusters basically emerge from a network right which might be interesting to study and we will discuss in a separate chapter you know how we how we can detect such clusters or communities from a network ok.

(Refer Slide Time: 06:08)



Language network: one simple example can be a co-occurrence network of words right, where nodes can be words and if two words co-occur together in a sentence for example, or very close by in a sentence say within a boundary within a window of window of 3 words or 4 words you basically connect them in the network.

So, nodes are words and if two nodes co-occur together multiple times you can connect them. For example, you see that you know the word like teacher right, principal, student they occur very frequently they co-occur very frequently and therefore, they are connected right and these kind of network is very important to study you know to automatically detect you know synonymous words or say antonyms right or say holonym homonym and so on and so forth, whole bunch of things in natural language processing right.

Another example of a network is keyword co-occurrence network where nodes are keywords and if two keywords again co-occur together multiple times you can connect them right. In scientific papers you may have seen that you need to specify keywords right in the paper. So, now if you see that you know phrases like keywords like sentiment analysis and opinion mining right. They are very close and you see that these keywords you know appear together very frequently.

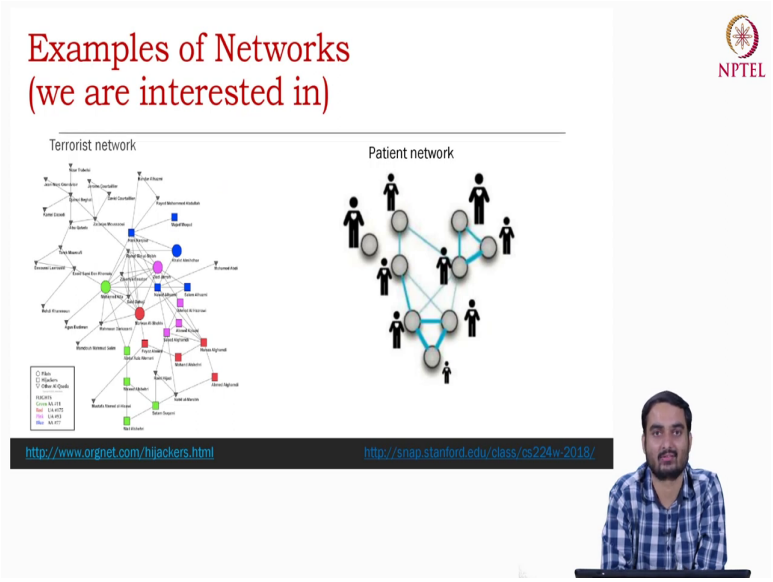
So, meaning that you know these phrases are very important sorry these phrases are actually you know very linked. You see here in this particular network sentiment analysis and opinion

mining they are close by right and so in this network nodes are keywords and two keywords are connected if they co-occur together multiple times in different papers.

There is another network called semantic network, ok. In semantic network now this is basically knowledge graph you may have heard about the term knowledge graph right, a knowledge base or knowledge graph, where nodes are different entities indicating different granularities of knowledge's. For example, you see that cat is a mammal right. So, cat is an entity, this is a node mammal is a node right whale is a node, animal is a node and so on and so forth.

So, cat is a mammal. So, therefore, there is a link from cat to mammal and the relationship is a right. So, this is you can think of this as an attributed network where edges are associated with some sort of attributes right. Is a or has right and so on and so forth leaps in and so on and so forth right, these are different attributes of edges right. So, now this is called semantic network, ok.

(Refer Slide Time: 09:10)



Similarly, we have we can think of many other interesting networks like you know terrorist network. I mentioned last day that you know where in this particular network nodes can be terrorists and if two terrorists went together for a similar mission or if two terrorists were arrested together right you can basically connect them through links right.

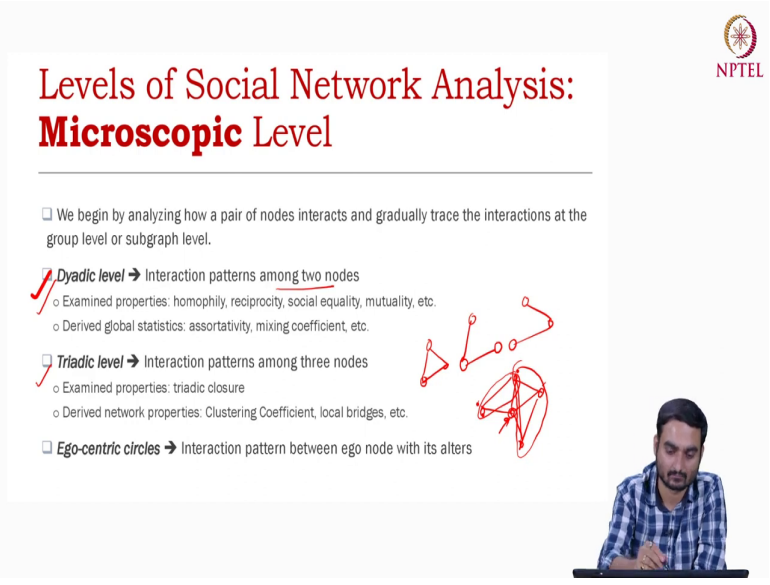
In fact, there are very interesting studies where people basically keep track of how such movement happens. I mean I am not very aware of this, but if you look at studies there are multiple books on you know how you can model such activities using networks, ok. There is another type of network called patient network and this is very important in again in the computational biology domain where you basically want to study how a particular epidemic spreads or a virus spreads.

So, here nodes are patients and if two patients are interacting if two patients come closer together you can connect them right. So, now if say let us assume that you have in a hospital you have this kind of network right and you have certainly seen that a virus has started spreading right.

So, you will immediately understand that through which path this virus has basically spread because you know that these people have been infected right and you also know that these people have been frequently interacting with other people right. So, essentially it means that you may want to protect those people who have already you know been interacting with the already infected patients and you want to protect them either through vaccinations or some other ways.

So, now we will look at a network from different angles right. We will basically inspect the network as a whole we then zoom in right and try to look at a part of the network. We further zoom in and look at you know even more fine grained entities in a network right.

(Refer Slide Time: 11:26)



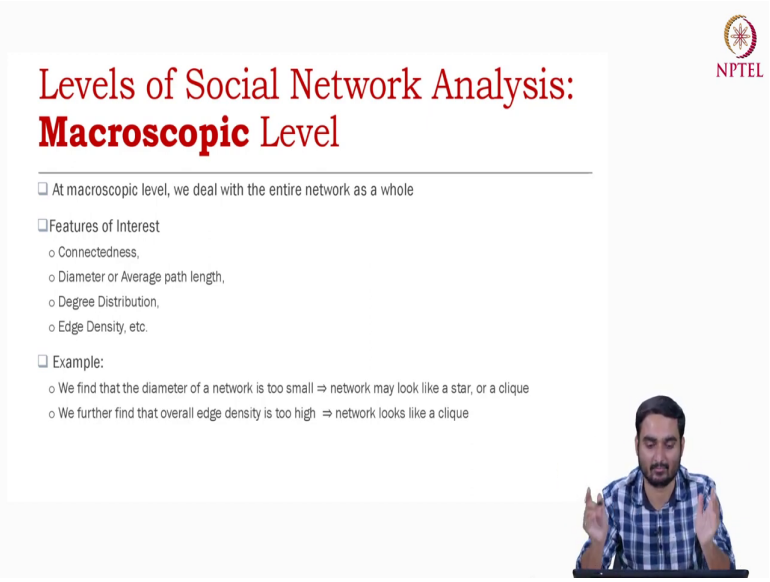
So, we will discuss three levels of granularity of a network, one is called microscopic level, the other is called macroscopic level. So, you have microscopic level you have macroscopic level of analysis of network and in between you have something called mesoscopic level mesoscopic view of a network ok.

So, let us start with microscopic level. So, when you talk about microscopic view of a network we basically you know we basically analyse nodes and edges ok. We look at nodes properties, we look at edges properties, we look at how two nodes interact right, how say three nodes interact and so on. We do not go further ok. So, we look at different properties of nodes for example, degree, centrality we will discuss what is centrality later right and so on.

We look at how two nodes interact right and this is called a dyadic level of interaction. We will look at how three nodes interact this is called triadic level of interaction right. For example, say this is a triadic level of interaction or this is another type of triadic level of interaction right. This is another type of triadic level of interaction and so on. So, we look at dyadic level of interaction, triadic level of interaction and we also look at egocentric circle. We have already discussed what is ego net ego network right. We will see that say let us say this is a ego network, ok.

And let us say this is a structure and this is ego, this is ego and these are the altars, ok. So, you see that here there is a circle ok meaning a closely connected nodes right. This is also a closely connected group ok. So, these are called circles egocentric circles. This is also a kind of a microscopic level view because you are only looking at a particular node and its surrounding neighbours and that is all ok.

(Refer Slide Time: 13:33)

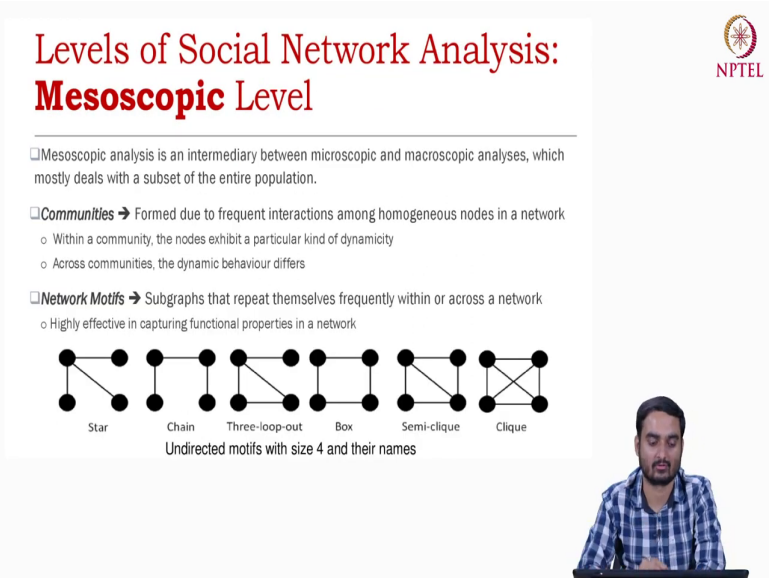


Now, let us look at macroscopic level, the entire network as a whole right. We can think of you know number of nodes, degree distribution ok this the term degree distribution may not be very familiar to you, but we will discuss in the next chapter. There is something called the diameter of a network which is the you know longest shortest path right. You may have heard about something called shortest path, you take a pair of nodes and look at the shortage path you take all pairs of nodes right.

You can see which one is the longest right. So, this is the diameter, that is called the diameter of a network, it is a network property. Similarly you can think of edge density of a network right. How many edges can be formed? How many edges can be possible in a network of node n, n number of nodes n c 2? And how many edges are actually there in the network right?

So, you take a fraction of actual number of edges divided by I mean the fraction of actual number of edges and the possible number of edges and that will give you something called edge density right. So, this is basically looking at the network as a whole micro macroscopic level.

(Refer Slide Time: 14:42)

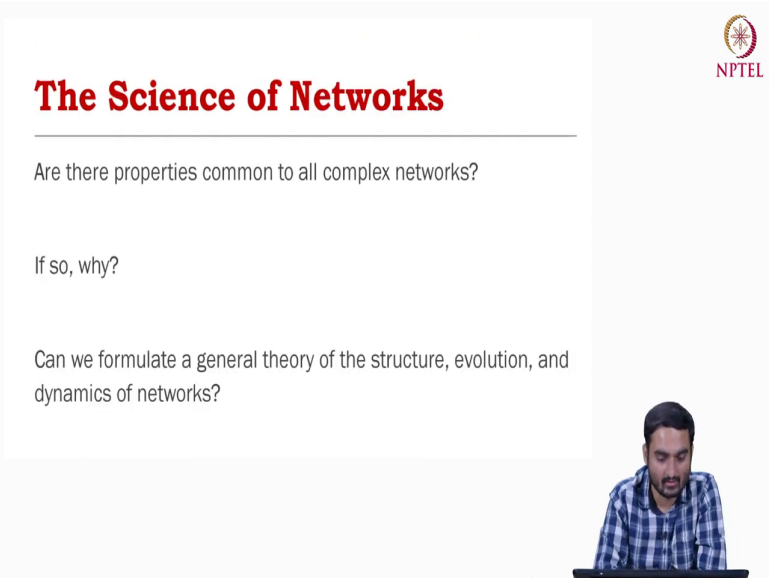


In between microscopic and macroscopic there is something called mesoscopic view of a network ok. And what is this? Mesoscopic view we look at specific regions of a network. For example I already mentioned about something called clusters or communities where multiple nodes interact together frequently and form a dense group right, you can think of this as a mesoscopic structure of a network.

You have some multiple such communities and these communities then form the entire network right. You can also think of something called network motifs and this motif structure is very useful in biological network particularly, where now what is motif? Motif is basically a recurrent you know sub structures which appear in a network right. For example, if you think of this [FL] this is a star network right and this star network appears very very frequently in a network. So, that network gives you a separate indication.

If you see a chain this is a chain right this chain the this is called chain motif right. Now if you see a lot of such chain motifs present in a network the that network can be different from a network having a lot of star motifs for example, right. So, again motif analysis is a different you know different direction of network analysis in general which we are not covering because this is more related to biological network. So, but of course, in social network also we have studied we see plenty of cases where motifs are useful.

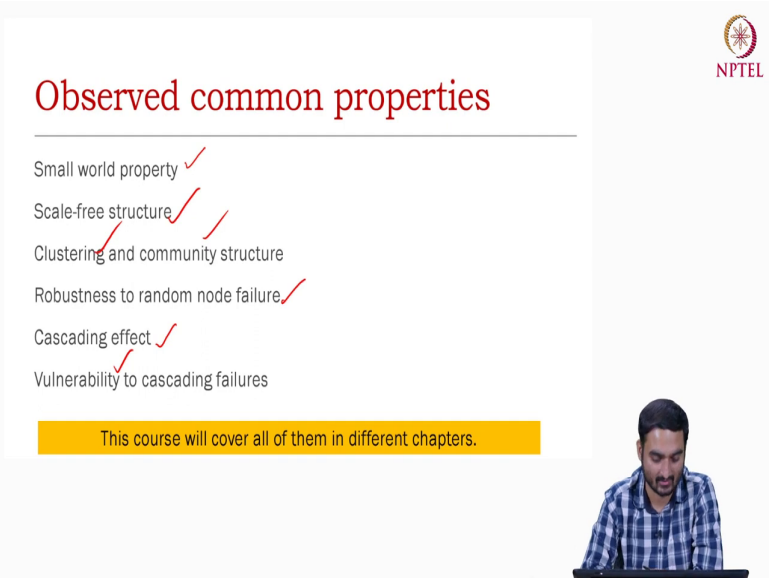
(Refer Slide Time: 16:20)



Now, you know this if you look at the science of network analysis. So, the question that we ask is that are these properties common across networks? Let us say you take a Facebook network and you take a protein protein interaction network right and you see that some of the properties are some of the microscopic mesoscopic macroscopic properties are common across networks, then what do you conclude right? Would you be able to conclude that these two networks are same or have similar properties? What are the different properties right?

So, the question that we ask is that can we formulate a general theory of the structure of the structure evolution and the dynamics of a network ok.

(Refer Slide Time: 17:11)

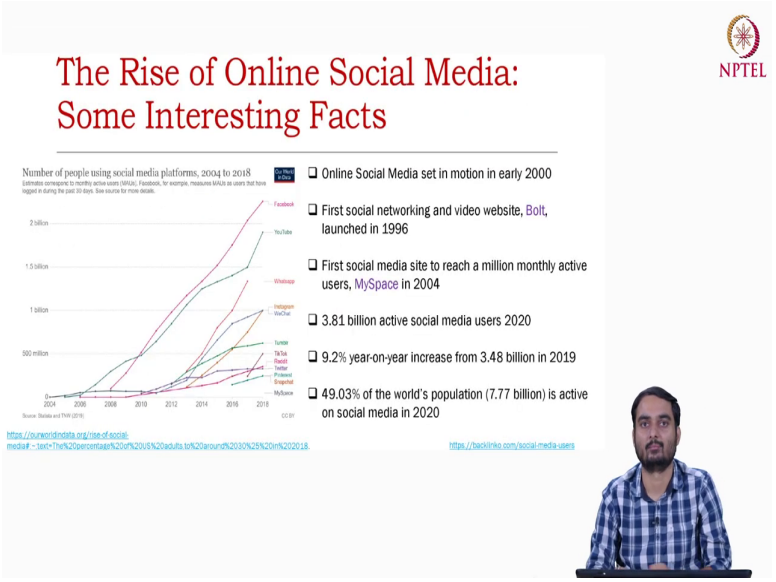


The common the observed property the common observed property that we often see in a network includes something called small world property ok. We will discuss in the next chapter what is small world property. It basically says that you know the world is very small meaning that if you want to move from one node to another node you do not need to traverse a lot ok.

And there is a very interesting property called 6 degree of separation ok. It basically says that you know there are 6 hoops on an average between pair any pairs of any pair of nodes in the network ok. We will also discuss something called scale free property right. We will discuss clustering community structure we will discuss something called the robustness of a network robustness of a network to different attacks different adversarial attacks. We will discuss something called cascade effect, ok.

The vulnerability to different cascading failure. For example, say there is a power grid electric power grid and suddenly you see that one node has got damaged right. And if one node got damaged it may happen that this damage may get propagated through the entire network right entire power grid, but that will you know create devastation right. So, we will need to stop such power failure, such failure of nodes right. So, what would be the strategies through which we can stop the spread of such failure?

(Refer Slide Time: 18:49)

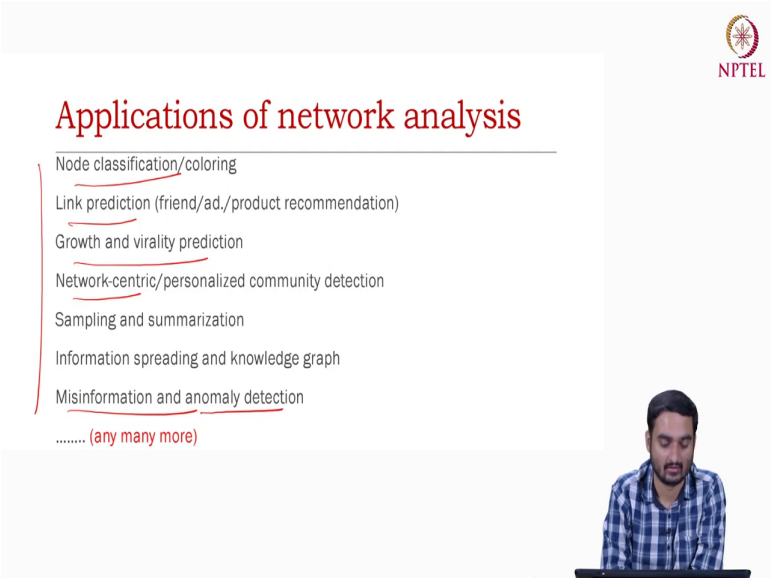


So, this course will cover all these things in different chapters and we will basically learn different mathematical formulations properties which can define this characteristics of a network. Now this is a motivational slide I am pretty sure that you do not need to I do not need to motivate you why you know social media has become a part and parcel of for life.

If you see the user engagement over time right from say the year 2000 to 2021 2022 there is a massive growth exponential growth of usage particularly during this you know lockdown time this pandemic time right and a lot of content are being generated every minute right. So, people sometimes say that you know social network is a proxy of our society right.

You see many cases where you know people take opinion public poll right. You have you know may have been invited to vote for certain polls right certain decision right and you know on the in the online social media since you are observing the patterns you are observing the different opinions. You are consuming different information in different ways right. You are one of the stakeholders right who can participate in this polling right. So, therefore, this is very important ok.

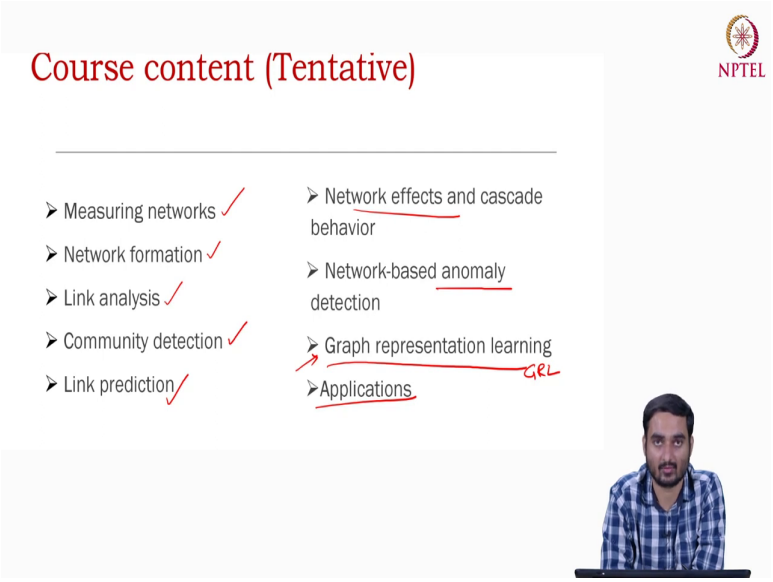
(Refer Slide Time: 20:19)



So, these are the applications that we will also cover in this particular lecture, we will cover something called node classification where we classify nodes depending upon the properties. We predict links right this is also called recommendation link recommendation or and link recommendation has a lot of applications in the product recommendation friendship recommendation ad recommendation and so on. We will discuss something called a growth and virality of messages right of networks in general.

We will discuss node centric network centric properties personalized properties of nodes. We discuss how misinformation spreads over social network, we discuss you know how you can identify nodes which are abnormal for example, say fraud nodes right fraud users and there is something called anomaly detection through which we look at how we can identify such you know fraud nodes or outlier nodes.

(Refer Slide Time: 21:18)



Now, this is the tentative course content that we are going to cover in this lecture series. We will start by you know by measuring a network. We quantify in the next lecture we will try to quantify a network, then we will something we will discuss something called network formation. We will discuss many such models through which you can actually mimic the way a network is formed, a way a network evolves over time right.

We will discuss you know random growth model, we will discuss preferential achievement models and many such models those models are mostly borrowed from physics, but they are highly applicable here in social network as well. We will discuss link analysis where you specifically look at how to characterize a link and edge right and how we will see how different social theories are basically are useful to characterize a link.

We will discuss community detection community or cluster is a very important property of a network and we will see how we can detect communities efficiently. We discussed something called link prediction ok and we have already mentioned the application in recommendation system, but the problem is you know when we need to predict something for future given that we have network you know at the current time stamp this is difficult because you do not know what is going to happen right.

So, link prediction is very very challenging and we will discuss some you know some algorithms which have become very popular because of their you know the way they are they basically work. We discussed network effect and cascade behaviour particularly how this

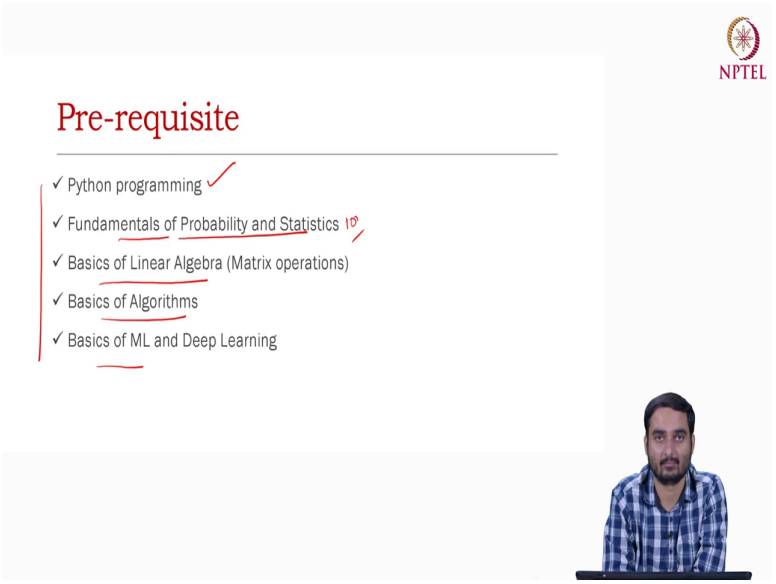
epidemic spreads over network online and offline network. We discussed network based anomaly detection anomaly detection or abnormality detection is also studied in data mining.

But, when it comes to network structure approaches are very different and we will see how we identify anomalous nodes anomalous entities it can be anomalous edges. It can be anomalous sub graphs right from the network. Then we will discuss a very interesting in fact, this is a recent trend in network analysis something called graph representation learning network representation learning GRL ok in short. So, we will see how network is mapped, how a network is mapped to an embedding space right to an Euclidean space for example, right.

And when you map a network to a vector space say on an Euclidean space things would become very easy. For example, now a node is represented by a vector right. So, you can do whole bunch of vector operations matrix multiplication and so on and so forth to solve different applications, but to understand this chapter you need to know a basics of deep learning, basics of machine learning we do not need to go into details of that, but a basics of machine learning and deep learning might be useful.

And then we will conclude this lecture by you know by giving you ideas about some applications for example, fraud detection. We look at fraud detection particularly right, we will we look at something called collusion black market driven activities in online social network. How you can detect such activities, we will also discuss very briefly about recommender systems like particularly friendship recommendation system and so on right.

(Refer Slide Time: 25:05)



So, these are the prerequisites. So, we so it is highly recommended that you learn python programming, you I assume that I am assuming that you have fair bit of ideas about probabilities and statistics I mean probability statistics 101 is good enough. Again I am assuming that you have ideas basic ideas about linear algebra particularly matrix operations you have ideas about basic algorithm design right and it would be great if you also learn basics of machine learning and deep learning ok.

With this I would like all of you to learn together the skill of networking.

Thank you.