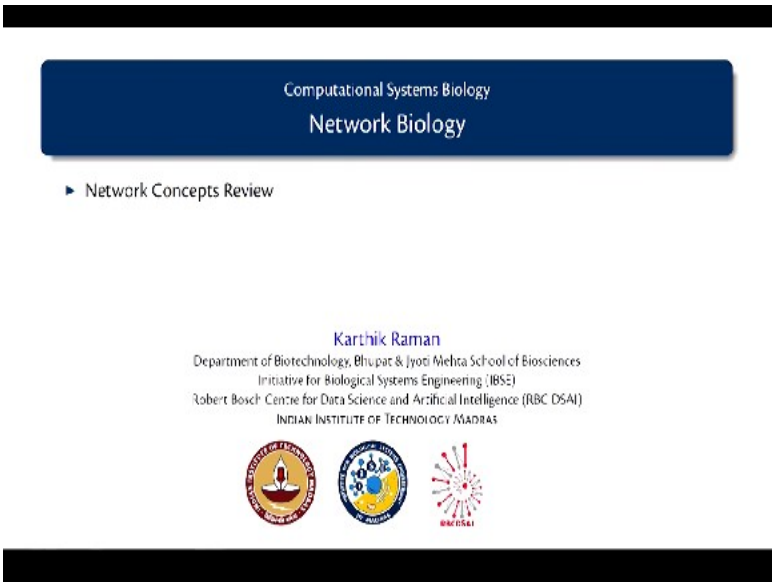
Computational Systems Biology
Karthik Raman
Department of Biotechnology
Indian Institute of Technology - Madras

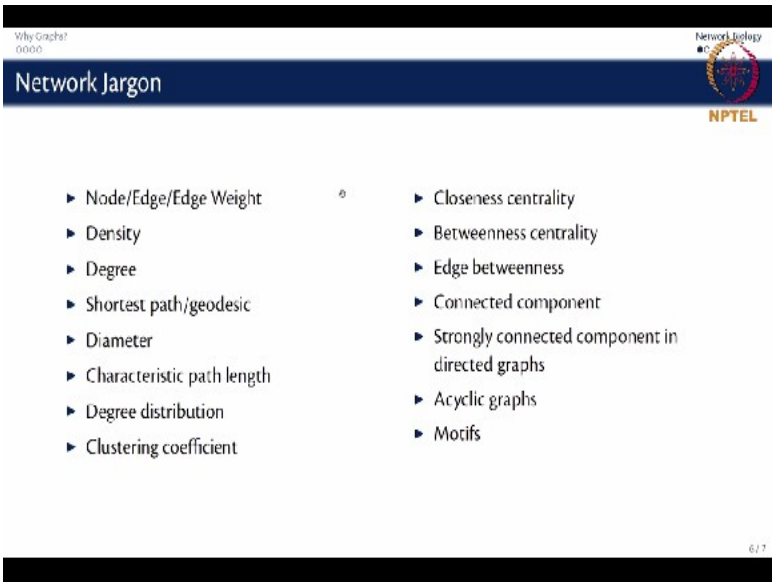
Lecture – 18
Network Biology

(Refer Slide Time: 00:22)



In this video, let us briefly review a lot of the network concepts we have discussed so far so that we are in shape to look at more advanced topics as we move on, okay.

(Refer Slide Time: 00:30)



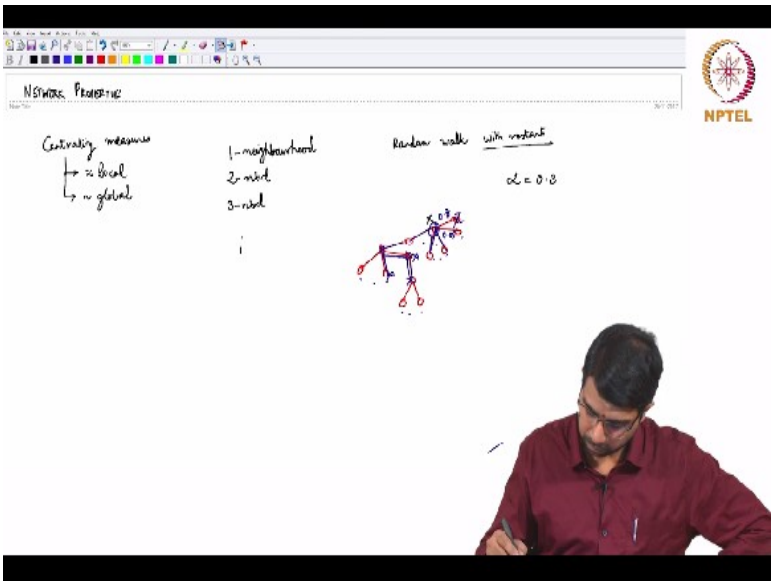
So, let's have a quick recap of what we did in the previous class in terms of the different parameters. Quickly let's look at which parameters correspond to nodes and which parameters correspond to edges and which parameters correspond to the entire network. So, density, would be a network property. Degree, is a node property. You can think of average degree as being a network property.

Shortest path, is basically for any pair of nodes and so on. Diameter, is a network property. Characteristic path length, is a network property. Degree distribution, is a network property. Clustering coefficient, is a node property but you can also average it across the entire network or you can also average it across nodes which have a particular degree, right. So, you typically look at a plot of Ck versus k.

So, C of k is the clustering coefficient of all nodes with degree k versus the degree k itself. It is like Nk versus k is your degree distribution. Ck versus k is something that you may want to check for different kinds of networks. We will look at it a little later today. Closeness centrality, is a node property, as is betweenness centrality. Edge betweenness, as the name sounds is an edge property.

Then of course, you have connected components and so on. So, we have several different properties to study node in a network as well as several global characteristics of networks themselves. There are many more interesting centrality measures. We will try to now do them in the lab session rather than just look at all the definitions and so on. For example, there is page rank and may be briefly just take a look at another interesting property.

(Refer Slide Time: 02:26)



So, we did look at several centrality measures, right and these centrality measures could be local or global in some sense, right. So, is degree a local property? Yes. What about betweenness centrality? It is somewhat more than local because it looks at all the nodes in the network in some sense, right. So, the betweenness centrality of one node is very dependent on the betweenness centrality of other nodes in the network whereas degree, yes to some extent, of course, right.

If a node has degree x, it has to connect to x other nodes but in some sense betweenness centrality is more of a global measure of what's happening, right. So, the other useful things to think about are, what is the, what is the 1-neighbourhood of a node? Or what is the 2-neighbourhood, 3-neighbourhood and so on? What is this? 1-neighbourhood is the list of all neighbours of a node.

2, is neighbours plus neighbours of neighbours. 3, is neighbours plus neighbours of neighbours plus neighbours of neighbours of neighbours, right. So, typically in biological networks you will see that in 4 hops, you might be able to traverse like practically all of the network or something like that, at least in many protein interaction networks and so on. So, these are other properties, right.

What nodes are present in your neighbourhood? Another very useful property is, let's look at a

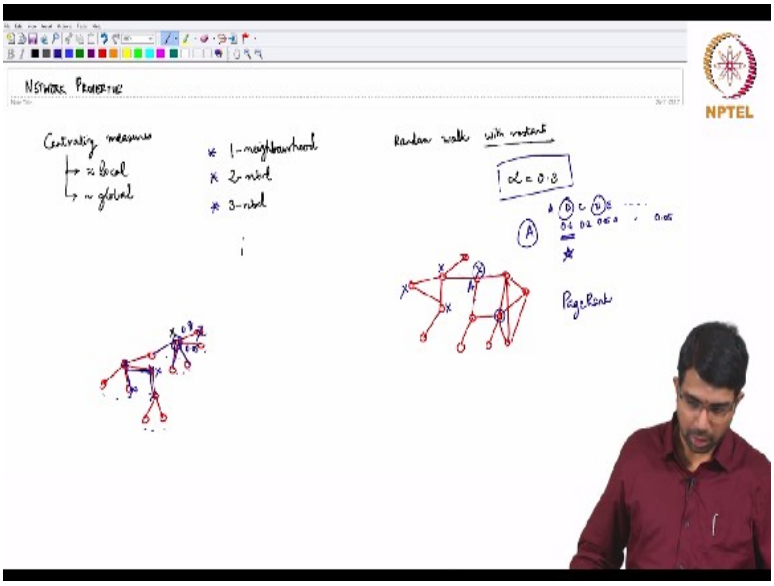
concept and I will come back to the property in a movement. What is a random walk? Let's start with a graph, right, and you could basically start somewhere on this graph, may be let's say you start here, right and what happens when you do a random walk? You pick a random edge to walk through.

So, let's say you take this edge, you walkthrough this edge, you now reach here. After this, you can either go back or go here, then reach here and may be come back here. We go here, come back here, go here, go here, go here, here, here, here and so on, right. You can basically just meander along the network, walk to different points, different nodes on the network, right. So which part of the network will you, so how does the profile of nodes that you visit change depending upon where you are, right?

You will see that if you start at this node, you are very likely to visit these nodes. Whereas if you start at this node, you are very more likely to visit these nodes or even these nodes, right. So, depending upon where you start on the network, you will end up exploring a different bunch of nodes. Does that make sense? And let us say suppose you had even probabilities, let us say you know, you had a 0.8 of probability of taking this edge versus a 0.05 probability of taking this edge and so on, right.

So, you will find that if you start here, you are very likely going to be here in the next iteration. So, this is a plain random walk that is this notion of random walk with restart, so you now define an alpha value, let us say alpha equals 0.3, which means with a 0.3 probability, you will restart your random walk. Let's say you are exploring the same graph that we looked at and you started at this point, right. You will explore, so you may first visit this, then you may come back here and so on or let's just take a fresh graph for a moment.

(Refer Slide Time: 07:04)



So, let's take another graph and let's say we start exploring this graph from this point onwards. So now with arbitrary probability, you might end up here, right. Then you might end up here. Then you might end up here but by this time because your alpha is 0.3, there are very good chances that you might, so at every point you restart with a probability of 0.3, right which means on an average, you will be restarting every 3 steps. You take 3 steps, you would have restarted.

"Professor - student conversation" Restart from the initial point. Restart from the initial point. So, the idea is I want to get a feel for the neighbourhood around me, right. So, we do that, this is one way of looking at the neighbourhood around me, right but this just tells you a long list of what are all the neighbours that I have. It doesn't tell you which neighbour is more important or which neighbour I am more likely to be at and, so on.

Even if you had an unweighted graph, you will see that in this graph you will find that you are very likely to be here. Because you will start here, you might go here, you might reach here. You might go here, you might reach here. You might go here, you might reach here. There are so many ways to reach this node. So, what you want to actually measure is let's say you have a node A and you have its neighbours or not even the neighbours, the list of all nodes in the graph, right.

So, with what probability do you arrive at each of the other nodes or how many times do you

arrive at each of those nodes? So, after any number of steps, let's say 100 steps. Hopefully, it would have averaged out by then, right. So, you essentially compute the stationary probability of visiting the neighbours. So, maybe for this node, you might find something like 0.6, 0.2, 0.05, 0 whatever. It should all add up to 1, right. Something else let's say with another 0.05 somewhere, right.

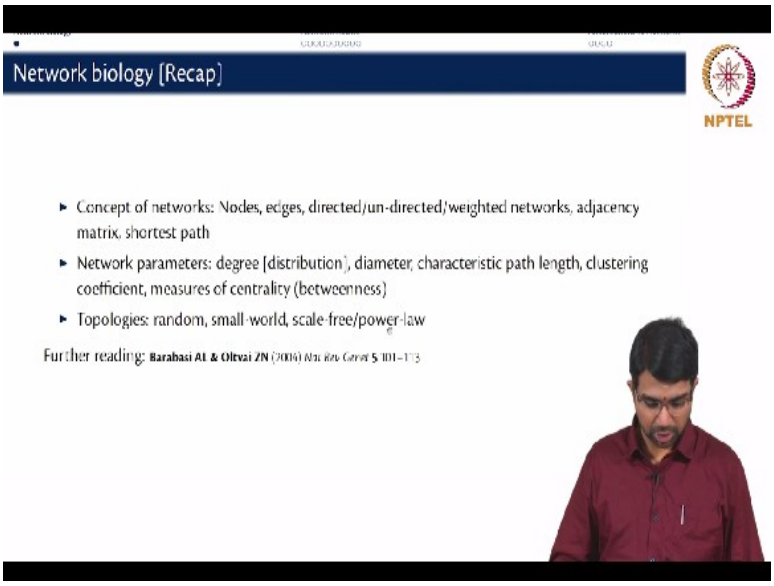
So, this means that B is in my neighbourhood of, in the neighbourhood of A and is also particularly important. So, this gives you some sort of quantification of the neighbourhood. In a sense, another way to quantify your neighbourhood because 1-neighbourhood, 2-neighbourhood are basically you know some unweighted neighbourhood maps. Here this tells you that, a particular node is more important than another node in the neighbourhood, yes. So, you just have to run simulations, right.

You basically start a random walker on, right. You start a random walker on whichever node you want to measure, let's say A, let's assume this is A and you want to now find, do a random walk with restart. So, you start a random walker on A, at A and you allow it to explore the graph, right. And you compute the number of times it visits B, number of times it visits C and so on in one iteration.

You do it for 1000 iterations and average over it. It would have hopefully converged. So, it basically you know uniform probability. So, if this graph is unweighted and you have let's say this is your node that you are starting off and all these 3 edges will have equal probability. So, once you will take this, the other time will take this, so if you do it 3 times, you would have likely visited each of these once or rather if you do it 30 times, you would have likely visited each of these around 9 to 11 times, right, something like that.

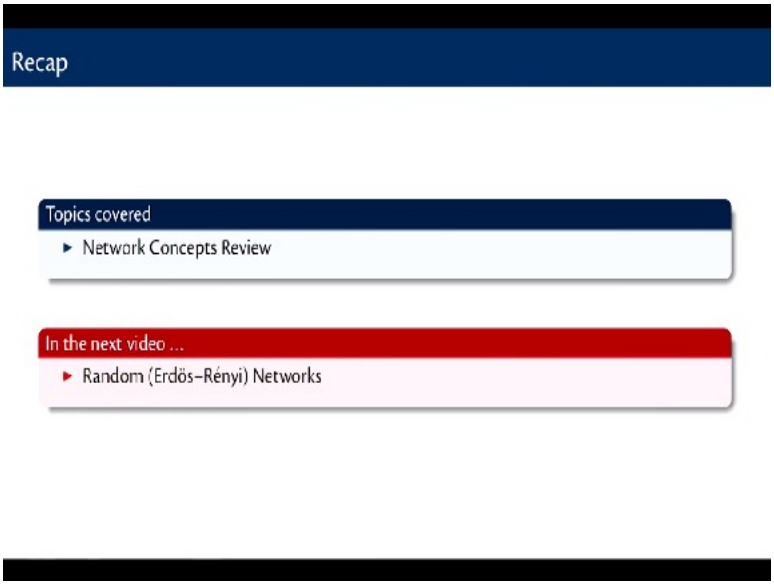
That is what a random walk would give you on one of these graphs. In fact, Google's page rank is intimately connected to this kind of a metric. It again measures the stationary probability of visiting a webpage or something like that. It is a centrality measure that essentially relies on this concept of some sort of a weighted importance to different nodes in a graph, okay.

(Refer Slide Time: 12:22)



So, now given that these are all the, we are now familiar with all the network parameters, the next thing to look at is the network topologies, basically random, small-world and power-law networks.

(Refer Slide Time: 12:33)



So, in today's video, we covered, we reviewed several network concepts and in the next video, we will start looking at the concept of network models and start with a very interesting model of networks which was proposed way back in the 1960s by Erdos and Renyi.