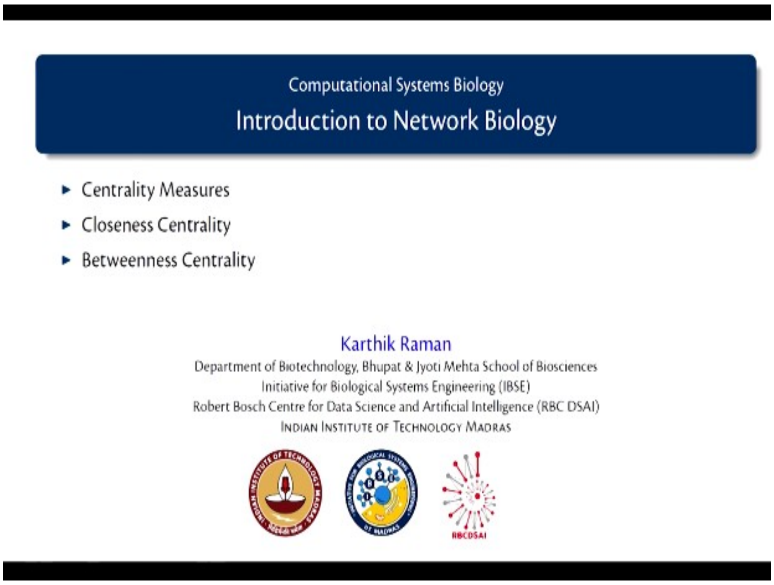
Computational Systems Biology
Karthik Raman
Department of Biotechnology
Indian Institute of Technology – Madras

Lecture - 16
Introduction to Network Biology

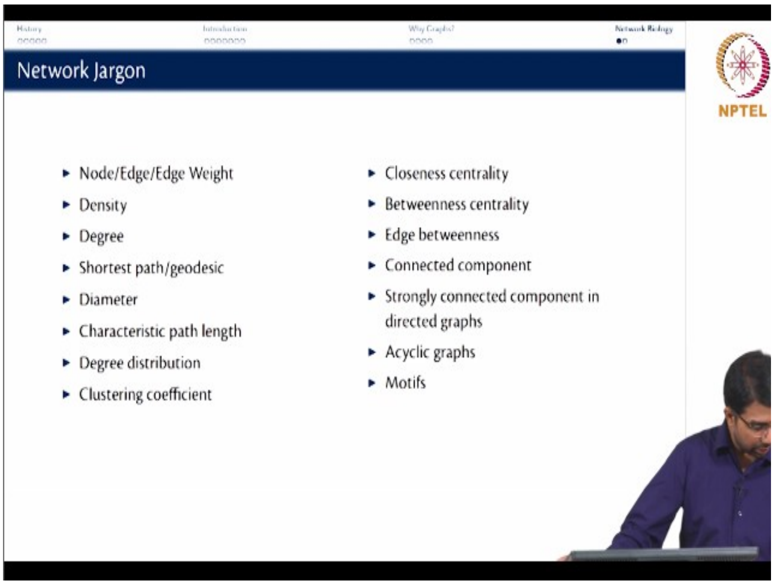
In today's video, we will look at centrality measures and particularly we will look at closeness and betweenness centrality.

(Refer Slide Time: 00:15)



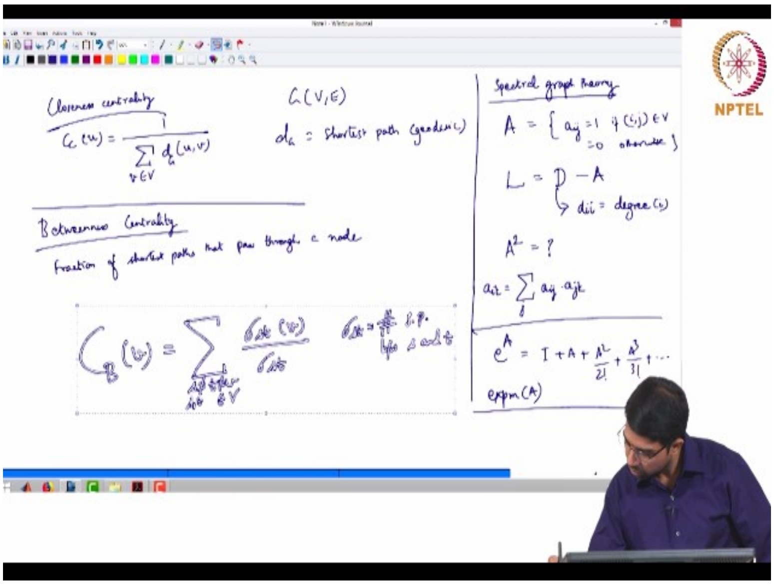
And then there are a bunch of interesting centrality measures, right.

(Refer Slide Time: 00:25)



The first of them is something known as closeness centrality.

(Refer Slide Time: 00:40)



So, this essentially tries to compute how far a node is from every other node in the network. So, that would be called farness, you invert it, you get closeness. So, you essentially compute, for a given vertex, you compute the distance between all other vertices, this is farness. You take a reciprocal of that, that is closeness centrality. There are different such measures.

So, d_G is basically distance on the graph G. So, we always talk about a graph V, E, where V is basically the set of vertices or maybe you know I should call it d_G, right, this would be a slightly better notation. So far, we have only looked at degree **"Professor - student conversation"** always shortest, d is geodesic. Okay now that you asked that, let's just take a small detour.

We will look at an very interesting concept. So, there is a whole branch of graph theory known as spectral graph theory, which looks studies basically a bunch of matrices. So, you all know one very interesting matrix that is connected to graphs, which is the adjacency matrix, this is basically, yes, this is an adjacency matrix, it could be weighted whatever and there is a Laplacian matrix, which is basically D-A and D is basically is a diagonal matrix where dii is degree of the ith node, it is 0 otherwise.

Many interesting properties of these matrix, so the eigen values of this matrix will give you clustering in the graph, a bunch of interesting properties like that, I will let you to study a little more on that, but let's just come back to the adjacency matrix and can you tell me what is A^2 ? Maybe I should reserve this for the lab and you actually try it and figure out what it looks like.

Exactly, so A to the 1 essentially tells you if there is a path of length 1 between a node i and j. A to the 2 basically will tell you if there is a path of length 2 between i and j and A to the 3 will tell you if there is a path of length 3 between i and j, you have to just do the matrix multiplication and convince yourself of this. This is actually a pretty cool trick to use paths. **"Professor - student conversation"** It is A^2 because you will multiply Σ aij.bjk right.

Okay let's look at what is A^2, right, so it is Σaij.ajk, right. So, this basically tells you if there is a way to go from i to j and then j to k, right. So 2 ways of getting there, right, essentially that, so just look at it closely, you will be able to see that, so the ik'th element is going to be given by Σaij.ajk. For, in fact you know, across all j's and k's, right so, I should put a, sorry not k.

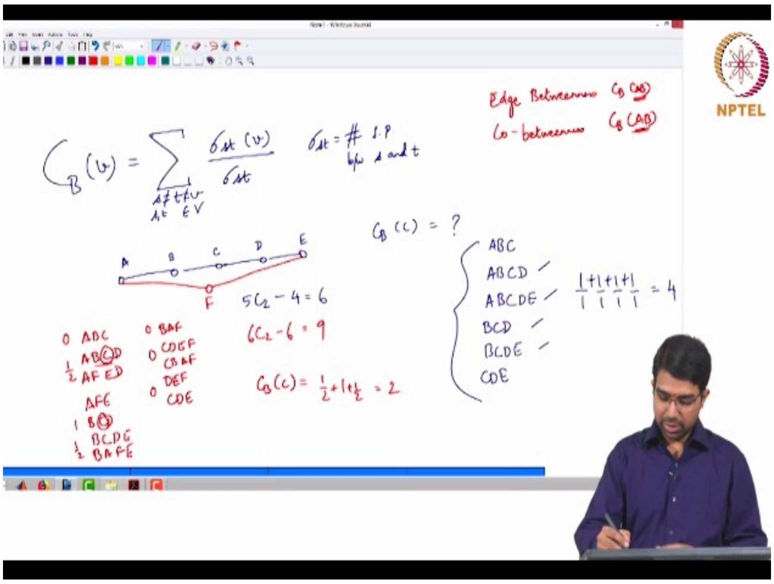
So, you will have 3 for loops right when you do a matrix multiplication. **"Professor - student conversation"** that would be like your diameter or something like that, it wouldn't be the most efficient way to compute the diameter, but you could do something like, you can actually do e^A even, right, that will tell you if there is a path at all in some sense, right, e^A is actually what is e^A?

So, it is basically not A no? It is I+A, it is like $1+X+X^2/2!$, right this is e^A, although this is really not how you would compute any of these things, it is just fun to know about it and in fact as we will see in the MATLAB session, this is how you get it, expm(A). So far we have only been looking at in a sense, individual nodes, right. There is nothing that tells you a little more about the network by virtue of its position about a node, by virtue of its position in the network and so on.

So, the next concept is quite interesting, it is called betweenness centrality. What does betweenness mean? It tries to measure how often a node is between 2 other nodes and how often is this given node sitting on a path between 2 other nodes, right. So, can you think of how this might be calculated? It is essentially fraction of shortest paths that pause through a given node starting from any node across all possible nodes. So, we essentially define $C_B(v)$ as, for all s and t, so σ_{st} number of shortest paths between s and t, and $\sigma_{st}(v)$ is the number of shortest paths between s and t that pass through v, v can't be s or t. So, you ignore edges. You don't consider edges to be paths, the first step.

No, it is only, so you should, v also belongs to V but it is summed only over all s and t. So, this is betweenness centrality.

(Refer Slide Time: 11:21)



Let's just look at a small example. So here what is the betweenness centrality of C? So, the explanation I gave you is somewhat approximate right. So, if you look at this, what is this? Is this a fraction of paths going through a node? It is sum of fractions of paths right. So, the first thing you need to do is to enumerate the shortest paths. So, let's just try and enumerate the shortest paths.

ABC is a shortest path. ABCD is a shortest path, ABCDE is a shortest path. AB is an edge. So, as we said s not equal to v, right. So, you don't consider whenever, so BCD, BCDE, CDE Is that right? So, you have 5 nodes, right, how many shortest paths will you have $5C_2$, because all nodes are kind of in the same connected, they are all connected right. There is no node that is hanging separately, minus 4 edges, which is 6 shortest paths accounted here.

So, we have all the shortest path ready here. So now let's do. So, first let's look at CB of C. So, what are all the paths that C passes through, I mean pass through C, this, so it is going to be 4 by, it is the terminus know, so we said s not equals to t not equals v. It is actually 4/4 equals 1. So, for all s and t so between A and D, A and E, B and D, B and E, sorry it is actually 1+1+ 1+1 no. Let's make it a little harder.

Or somewhat more interesting, you now have a new node F. Now what are the shortest paths? ABC is a shortest path, ABCD is no longer a shortest path, it becomes AFD. **“Professor - student conversation”** no right, so just hold on, let’s finish this example and get back to it right. So, you basically compute, so s and t is first A and D, then A and E and so on, right.

For all possible s and t, right you cycle over all s and ts. You can finally again normalize it, right. So, if you finally want to rank it and so on. We will try to take a look at it when we do the network biology lab maybe tomorrow or so. So, you have ABC, AFD right what happens to this becomes AFDE right, is it right? yeah. Then you have BCD is still a shortest path. BCDE is still a shortest path.

CDE is still a shortest path, FDE is also a shortest path and then you have FAB, FDC, FABC is not a shortest path. So, actually this doesn’t turn out to be a very good example I think. Let’s just work with this now. So now what are the shortest paths? ABC is the shortest path, ABCD is the shortest path as is AFED, right. So, you will see that the betweenness of B and C will now drop.

Right because only half of the shortest path between A and D pass though C. AFED what else so ABC, ABCD, ABCDE is no longer a shortest path it is just AFE and BCD is still a shortest path, BCDE is a shortest path as is BAFE. BAF is a shortest path, CDEF is the shortest path, CBAF is the shortest path. You have to actually worry about it in all the other directions as well, but as long as you do one direction correctly, there is no problem, right.

So, how many paths should you have you should have $6C_2$ -5 right which is basically 10. We have listed 11, so some of them there are alternate paths, so that is okay, **“Professor - student conversation”** all pairs of nodes, number of shortest paths, it will tell you what are the number of pairs of nodes between which you have to find the path. **“**

So AC is a node. AB is not a node so AC, AD is done, AE is done, now BA is not to be worried about, BF is done, BD is done, BE is done, so then CA, CA we do not need to worry about, we will only look at in one direction, because ABC is already there. CF, so we have CBAF and CDEF that is done. Then CDE is missed out. We have CDE and then we have DEF, then you have still one missing path I guess.

But essentially if you now start looking at $C_B(C)$ right, so it will be 1/2 +1+ 1/2. I think there is one more path that we missed, that's right we have 2 edges not 1, fair enough. So, now this is going to be 2, so if you see this number by itself is not very helpful right, you need to look at it in the context of other numbers that you have. If you had a number between 0 and 1 it is nice.

Right we can say something is 1, you know all shortest paths pass through it, something is 0., no shortest path passes through it. But now you have different numbers so you need to normalize them. So, we will look at this when we do the lab session where it will become a little more clearer okay. There is a very identical concept called edge betweenness, but I think we will just hold on to it.

Or let us just finish up with edge betweenness. So, how would you extend this concept to edges? This essentially computes the betweenness of a given node in a network right. You can extend the identical concept to edges which is edge betweenness. How many times does an edge occur between other edges?

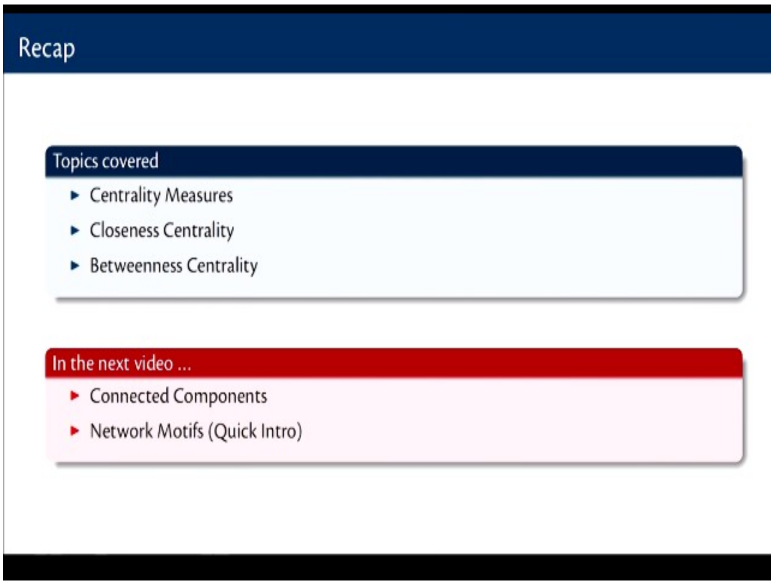
Right you look at all the same paths, but instead of looking for you are now looking for C right, instead now you start looking for say BC or something. So, BC occurs once here. Only once in fact and so on. So, you basically look at how many times a single edge occurs between other edges in a given network and you also have a very similar concept called co-betweenness.

So, co-betweenness is basically very similar to edge betweenness. Here, if you say it means that AB should have an edge between them, here AB need not have an edge between them right. How many times are paths passing though A and B in the network? So, you take a list of shortest paths like this and you just count how many of them pass through A and C, something like that.

So, AC doesn't have an edge as far as this network is concerned right. So, this tells you some sort of relative importance of 2 nodes. So, you can choose to define it in whichever way right. You can either ignore edges or to keep edges, but usually we ignore edges we only look at paths, but it really doesn't make a difference, right. It will sort of add a little bit of redundant information to every measurement right.

So, we usually choose to leave out the edges, fair enough. So, in the afternoon we will continue with this maybe we will finish though this slide finally. It is almost done. We only have like a small bunch of concepts, I wouldn't bother much about motifs. We will do motifs much more seriously in a later class. So, we will just look at you know connected components, and we will be done.

(Refer Slide Time: 25:06)



So, I hope you had a good introduction to various centrality measures in today's lecture. I encourage you to go and read a lot more about centrality measures. There are many centrality measures that have been developed to study different kinds of properties in different types of networks. In the next video, we will look at some more network concepts such as connected components and I will give you a quick introduction to network motifs which we will study in greater detail a little later on.