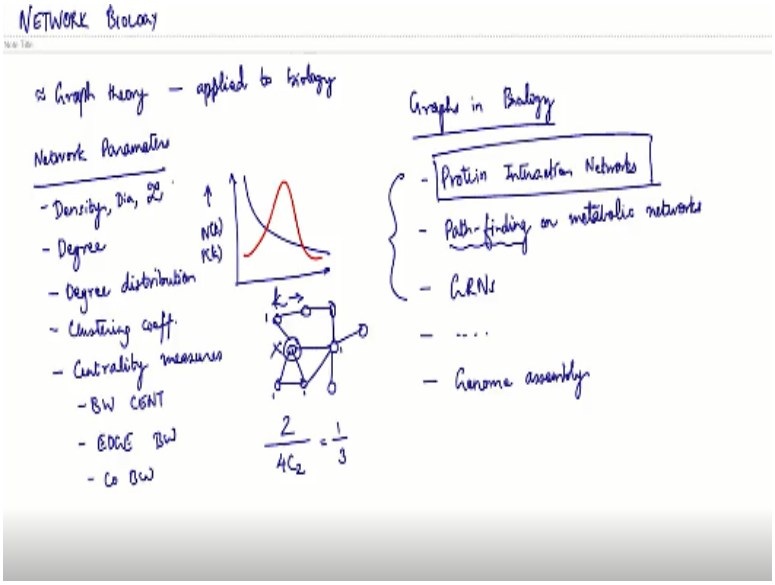
Computational Systems Biology
Karthik Raman
Department of Biotechnology
Indian Institute of Technology-Madras

Lecture – 28
Network Biology: Recap

So, in todays video it is a little long we will try to recap all the concepts we have covered so far in network biology so that you know all the concepts are fixated before we move on to the next topics. So, we look at the basic network parameters, centrality measures, network models, communities, motives and so on. So, welcome today we will take a look at network biology a recap of what we have covered over the last few classes in the context of network biology.

So, what are all the concepts that we have studied so for.

(Refer Slide Time: 00:47)



So, what is network biology it is nothing but graph theory applied to biology. We started with the theory of biology looking Eulers classic contribution of showing that there is no Eulerian path in the bridges of Konigsberg. But beyond that we started saying why we need to study graphs in biology and it turns out that many biological problems map nicely back onto graphs. Can you give me some examples of problems that may map on to graphs in biology?

"Professor - student conversation starts" So, you can have protein interaction networks so

maybe I will start showcasing the problem itself so maybe you want to study do path finding in metabolic networks what else gene regulatory networks **"Professor - student conversation ends.** so there are different types of networks as you say GRNs and so the different types of networks or a problem that we did not go into detail.

But just quite important in the concept in the context of graph theory is genome assembly. How do you assemble genomes this again involves graph theory? So, there are many problems that heavily involve graph theory. We are of course more interested in some of these problems and even there we actually did not look at path finding and so on. But this is a good example of protein interaction network what can we find from protein interaction network.

You can study various network properties you can look at what is the average separation between 2 nodes and a network. What is the maximum separation between 2 nodes and a network? Is the node very cluster? or does it have a particular kind of degree distribution? and so on so we then went on studied several important network parameters. So, and network properties so we looked at things like density.

Which is essentially the fraction of edges that are actually in the graph then we looked at degree most importantly degree distribution, the concept of clustering coefficient and there it is centrality measures. We discussed closeness and between us and so on but there are many other centrality measures including page rank eigen vector centrality and many more. They are different centrality measures that are commonly used.

And many of them have a biological implication as well and I have shared some material corresponding to those. So, maybe we can quickly visit some of these things so what is degree is the number of neighbors that any node has. And for a direct graph some of these evidence slightly change. May want to talk about indegree and out degree and things like that. Very important concept of course is that of degree distribution.

Which is a plot of how degree changes for different nodes of the number of nodes of degree K probability of finding an node is degree K versus K. So, if the graph looks like this then it is a

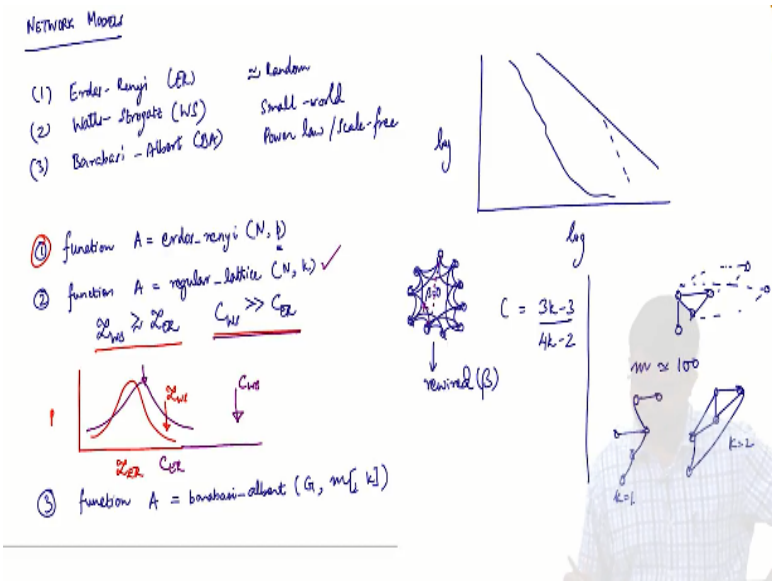
power law so if it looks like this it is something you can have different kinds of degree distributions. What is clustering coefficient it tries to measure how clustered the network is and the simplest cluster is this.

So, it essentially counts the number of triangles in the neighborhood of a particular node. So, if this is a node and it is part of a bigger graph. You will find there does node X has 1 2 3 4 neighbors so there are total of possible 4 choose 2 edges between those nodes but how many of them exist you know 1 here you have 2 here is that yeah. So, this is the clustering coefficient of this node.

And just basically 1/6 or 2/6 1/3 and the centrality measures particularly important is between a centrality. We also looked at the edge betweenness and Co betweenness these are other useful concepts. So, these are all the basic parameters which we can use to characterize in it but some of these are node properties some of these are edge properties some of these are network properties you can also study network averages.

Network average clustering coefficient diameter is already a network property. So, maybe I will say density diameter characteristic path length. Characteristic path length is average separation between a pair of nodes in the network. Okay so once we finished looking at these we then moved over to important network models.

(Refer Slide Time: 07:43)



And we studied 3 classic models Erdos Renyi, Watts Strogatz this is what we call Rando this is what we call small world and this is what we call power law scale free. So, do all networks have to fall into one of these categories not necessarily you might find some partial character in some region in degrees you might find a power law behavior not of that region you might find a small world behavior and you may not find any perceptible pattern in some other region.

So, there is lot of some sort of a misplaced obsession about trying to show that your network is a power law network. People just want to show that every biological network is a power law and typically you might end up with a degree situation looks somewhat like this in a log log plot this need not be a power law right this would be a perfect power law. But you might have you know deviations from it but just not unusual.

Or you might find a nice power law in this domain and then just changes after that right so you have to be careful about claiming that your network is a power law network and what is more important is the fact that are there hubs in your network right is there good clustering in your network right. And we also studied how to generate these network and we tried to quote them I think we should try doing that again.

Because it is sometimes can be a little tricky so how do you generate Erdos Renyi random network. If we were to write to MATLAB function it will look something like function A=

always you want to basically create the adjacency matrix corresponding to any given network of N, P. And then what will you do you will just write to construct an adjacent matrix such that a number an entry in the adjacency matrix will be non-zero.

If it is if you generated a random number that is <=P. Right so you generate a random generate several random numbers right and whenever it is <P for every edge in fact. Whenever it is <P you put a 1 bar otherwise you let it remain a 0 **"Professor - student conversation starts"** you can do something like that page right. So, you may have to scale P down or something like that right.

So, you have to make only half the edges with the upper triangular and then just make a copy of that right **"Professor - student conversation ends** so that is how you would generate and how would the degree distribution look like pass on because you are counting the number of successes of you know something an event that has a probability P. What is the event that has probability P the probability of establishing a link right?

And we are now counting the number of links or the number of successes which is equivalent to binomial or a Poisson. So, now this is for in the second case you start with the regular lattice and N, K you now have to adjust the adjacency matrix. How do you assemble this adjacency matrix? what you know is you have a regular lattice where every node is connected to K nearest neighbors on either side.

And then we can then show the cluster in coefficient is basically we did see an example previously right a cluster in coefficient for a network with 10, 3 or something like that. But this gives you only a regular lattice which now has to be rewired with some probability beta or P whatever right so beta=0 is this regular lattice beta=1 is a completely randomized network. Somewhere in between.

As I do you will find what are known as the small world properties which are basically what are the smaller properties. **"Professor - student conversation starts"** so LWS right yeah and how would you ascertain these so you have to create multiple networks right you cannot you create

multiple ER networks using say this function and you will get a plot **“Professor - student conversation ends.**

This is ER and this is some probability of observing frequency whatever you might find that LWS lies somewhere here. But if this were actually the clustering coefficient plot CWS might actually lie somewhere here. Right because it is much much >CER the mean value of CER of course somewhere here. **“Professor - student conversation starts”** beta you rewire it how do you rewire it pick a random node which is not connected to and then rewire.

So, several link and connect it across the table essentially so you say let us say we sever this link. Let us say we remove this link and then we connect it here it is not a neigh our already and that is where we try and connect it. We can connect with both nodes you have 1 node which is kept constant you detach the other edge plug it somewhere else across the table. So, only pick a node where it is not already connected to **“Professor - student conversation ends.”**

So, now think of how will you write the rewiring code right is quite important to start thinking in terms of how do you code this a password eight especially for some of you may have a biology and are not that familiar with programming so you need to start thinking about how you code each one of these things. Right so the first thing is what should my output look like your rewired matrix.

Essentially your rewired graph should actually be represented as a rewired matrix not adjacency matrix. So, you have to know how to generate an adjacency matrix such that you have rewired the original graph or its corresponding adjacency matrix. So, what you need to do is you first construct this regular graph regular lattice and now pick a node pick an edge from that with a probability beta you now decide if I want to require it or not.

So, you pick the edge and re wire it if it satisfies the probability beta just like what we were doing in the first case we then rewire it and while rewiring just make sure that you are attached to some node that you are not already connected to **“Professor - student conversation starts”** <beta right <beta so in that way you can generate a rewired network. So, the next thing that we

will look at is function A= what does my input to be here initial graph very good.

So, I would start with an A0 or let us call G an initial graph G right then number of incoming nodes and also very good number of it can be optional number of edges that you add in every round every alteration. Every incoming node how many edges does it make with the existing graph. **"Professor - student conversation ends."** So, if we started you could start at the graph like this event and let us say you have 2 incoming nodes which make 2 edges each.

So, you now have to connect this you end up going to going into this and then this you may end up going to going through this as another it could be here and there it could be here something of this sort. So, after every iteration you will be adding NK edges I mean K edges right at the end you will have NK new edges that have been added right. What happens think about it when if I say N is 100=1 what would be the clustering coefficient of the graph.

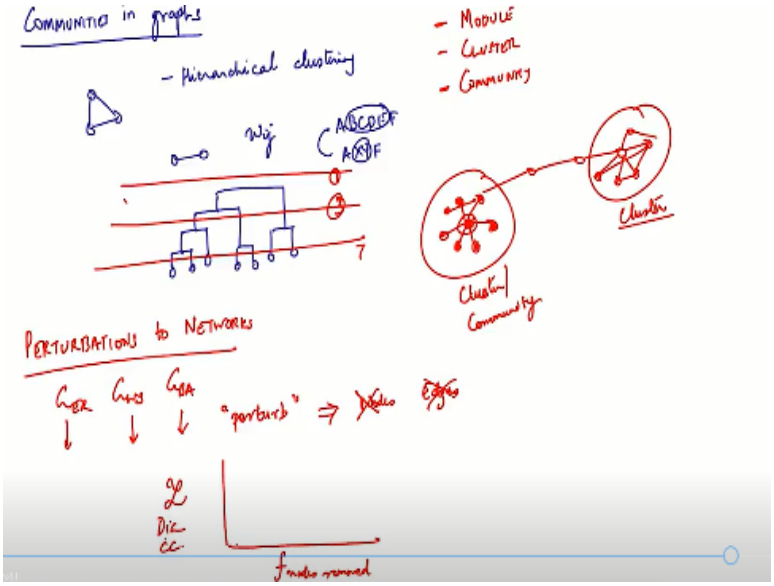
And not something that you would normally be able to answer easily because K=1 it is a special case and I think you may be able to easily see what happens. If every incoming node and let us say we start of with it this is your G initial graph every incoming node makes 1 edge with the existing nodes so what would be the clustering coefficient. You would find there that there is no way triangles can be created because when a triangle to be created.

You need one node to make at least another one so 2 edges and then they should also connect to 2 nodes which are already neighbors right. ok so if I am going to just add 1 node like this I am going to make some change but on the other hand if I were going to 2 nodes I make a few triangles not everything results is a triangle. But you may have you will have some more clustering.

But Barabasi Albert graph do show clustering additional 1 node is a sort of maybe simplifying assumption. So Barabasi Albert graph do show clustering as you can see for yourself so you are getting a better picture now of how you go about building the different **"Professor - student conversation starts"** from the equation there is no cluster**.** So, this is K=2 I have this it is not a cluster **"Professor - student conversation ends."**

So, one thing could be as you know that is the next thing we will get to.

(Refer Slide Time: 21:28)



You will have communities in graphs so that is the difference between communities and clusters sometimes use the term interchangeably. But usually when you say cluster this is what we call it as a cluster at least in the sense of clustering coefficient where the terms are often used loosely and if you are going to talk about communities or cluster. So, usually it is called clustering in graphs right.

Is there any clustering in a given graph? and how do you go about identifying communities in graphs? One way we looked at it by using hierarchical clustering you first connect the nodes closest to each other based on some metric. So, you can have some metric wij which tells you how close to a node. Maybe the one example we looked at was trying to compute the number of node independent paths between a pair of nodes.

So, if you have ABCDEF and AXYF these are 2 node independent paths but there are no paths no nodes it has shared between these 2 paths. So, that could be wij and you could use that to start grouping vertices together and enter it as dendrograms of some sort and so on. And depending upon where you cut this dendrogram if you cut it here you will get 7 clusters if you cut it here you get a single cluster if you cut it here you will get 3 clusters use interchangeably.

So, so one thing so the moment you understand the concept of cluster coefficient you can then start looking at communities and clusters and similar. But the thing is you may find that this is a this is actually truly a cluster, this is a cluster or a community the words are often used interchangeably particularly in the biological community. But see this this is this may be a cluster by virtue of its isolated function and so on.

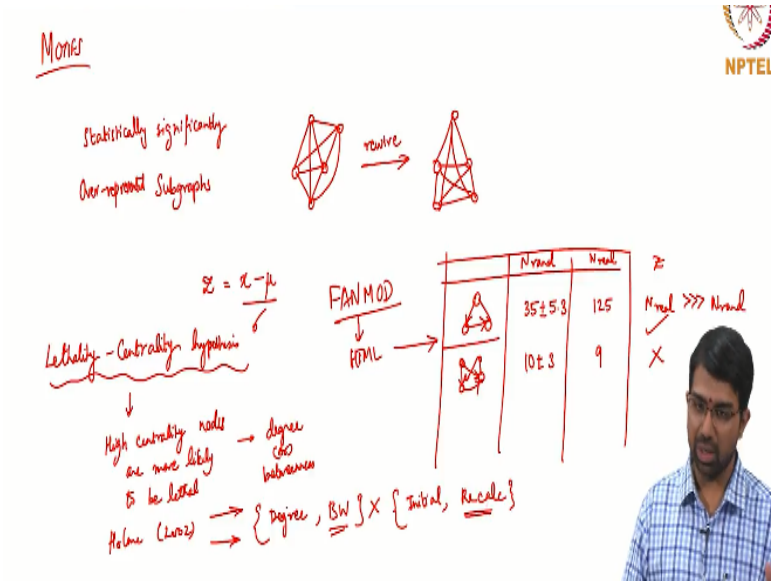
But if you see that are not enough intra community edges. I just think there are there are a few but it is not this tight where you use all of these so this is this is the hub this is not the hub right. So, the hub is only the center node so you call it a hub and spoke formation kind of thing and so but in general all these terms are used loosely. Module, cluster, community so I think this is more preferred in graph theory literature and so on.

This is somewhat more preferred in biological literature this is arbitrary what other concepts we discuss. Modularity so modularity community perturbations. Right how do you quantify perturbations to any network right I think you should try to study all of these things how do you study perturbations. You take you start with any graph right it could be it could be a ER graph it could be a WS graph it could be a DA graph and perturb which means what remove nodes and edges.

And a typical study to do is fraction of nodes removed on the x axis y axis is some parameter that you are studying. IT could be L, it could be dia it could be clustering coefficient whatever and you may find something like this s rather something like this and so on this is and so on. This is something you should study let us try to do this in a lab session you first have to write codes to generate these basic networks.

And then knock out something and then study then finally we will look at motives.

(Refer Slide Time: 27:13)



Motives are nothing but subgraphs rather over represented subgraphs statistically significantly over represented subgraphs. So, as an example are there any motives in this graph this is why you need to understand the difference between count and statistical significance. So, just because something is abundant has high frequency does not mean it is statistically significant. Right what do you mean by statistically significance.

"Professor - student conversation starts" when compared to a null model **"Professor - student conversation ends"** null model is revised version of this network and this network cannot be even rewired. All the edges already exist so how many edges are there in this network. Every node has 4 outgoing edges. So, how will you rewire it we will rewire it to something like this.

This is same thing so all realizations of this network will be the same so none of your motives are statistically significant. This is of course an extreme example to drive significant issues of cause and extreme example to drive home the concept but in practice you will see that you have to do a statistical significance. So, what means of statistical significance what other measures of statistical significance can you do.

A simple way to do is to look at the Z score. What is Z score? You have a population you take its mean and you do not take a particular observation you do x-mu/sigma this is you know really

high you will find that maybe it is going to be statistically significant or you can do a parametric test or a non-parametric test typically an unparametric test to try and identify how whether your observation your motive is statistically significant.

So, in practice how do you do this it is very similar to what we were doing right here you need to plot towards the frequency distribution of your motives of interest in the randomized graphs. It will start with a network you and randomize it multiple times and then your compute how many times am I observing this feed forward loop or bi fan or any of these particular motives. And then you go and study you go in and see how many times it occurs in your real network.

Maybe you call gene regulatory network and then you can say that it seems that feedforward loops are over represented and statistically significant over represented in the equalized gene regulatory network compiled to random rewired versions. So, there is a tool called FAN MOD which helps you find motives it will basically give you a long table of this sort e sword here it will say what is the motive and here it will tell you the random distribution.

And maybe Z score well and so on so would you say this is statistically significant steady + or – 5.3 or whatever so it is a tight distribution on 35 and the real guy is way of 125 which is much greater than. So, you will say that and real so this is a statistically significant motive let us say another motive When you find that the N rand is you will say that this is so this FAN MOD tool basically generates some HTML files which will give you a table such as this.

You can draw any random any network at FAN MOD it works only on windows though. Okay so I think this more or less brings us to the end of network biology. There are many other interesting concepts you can look at dynamics in network you can look at percolation and there are different kinds of perturbations. In fact, maybe one thing we should discuss is how do you really damage a network.

What is the most effective perturbation you can do to a network. One concept I did not review was Lethality centrality hypothesis. What does this state? High centrality nodes or more likely to be lethal and this could come from either degree or betweenness. So, there is a strategy to know

knock out these networks so people evaluated strategy a classic paper by Petter Holme in 2002 I have posted this paper.

So, what they did was they tried 2 strategies degree, betweenness what this means is they tried degree and used only the Initial degree. So, initially you will have some degrees right some node will have a degree of 100 some node will have a degree of 95 90 80 20 whatever. So, you remove those in that or you could recalculate degree after every removal when you remove the node at degree 90 it will take away few other edges from the other nodes.

So, the degree distribution will actually change after you remove the first node right so that is re calculation. Similarly, you can compute betweenness centrality initially rank them based on betweenness centrality and start removing the nodes in order of betweenness centrality. The other option is you remove the first node the highest between the centrality recompute between this in the entire network.

And now start removing the node with the highest betweenness centrality recursively. It turns out that what I have underlined was the strongest perturbation you remove nodes on betweenness centrality and recalculate very iteration ensure that the network disintegrated the fastest. How do you measure network disintegration again measure **"Professor - student conversation starts"** characteristic path length?

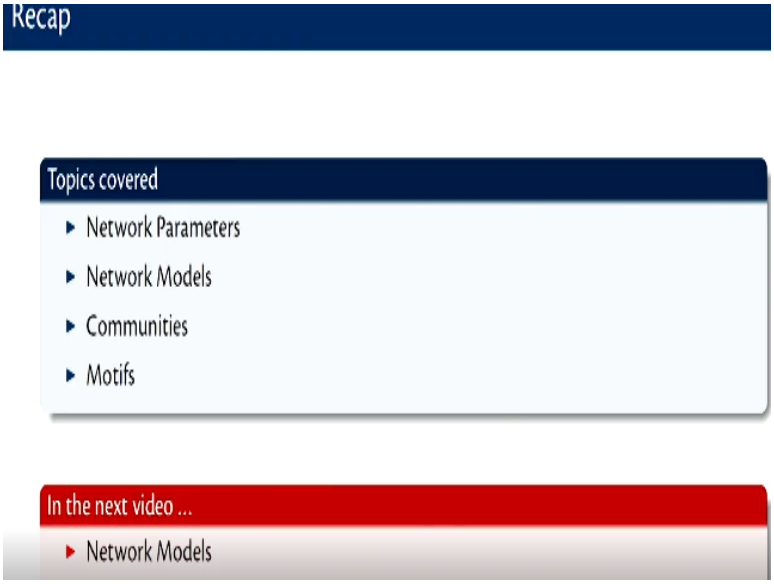
How does characteristic path length change how does diameter change yes you can also use the pair wise disconnectvity that we were talking about in a previous class. You start removing the top most degree and then the next highest next highest and next highest not recalculate **"Professor - student conversation ends"** not recalculate in the degrees. You initially computed a degree vector you use that as your list priority list for removing nodes.

The other option is to recalculate after every perturbation **"Professor - student conversation starts"** we have to recalculate yeah recalculate degree why the difference is not coming that is another strategy so you can either use rank nodes based on degree. Well I can add a 100 other things to this I could say page rank close to centrality I can say eigen vector centrality so on and

so forth.

All those are potential strategies and you do not want to evaluate how good these are as strategies **"Professor - student conversation ends"**.

(Refer Slide Time: 36:52)



So you hope you had a good revision of all the network related concepts in todays video right from the network parameters to the different types of network model communities and so on. In the next video you will do a lab wherein we will visit some of these network models.