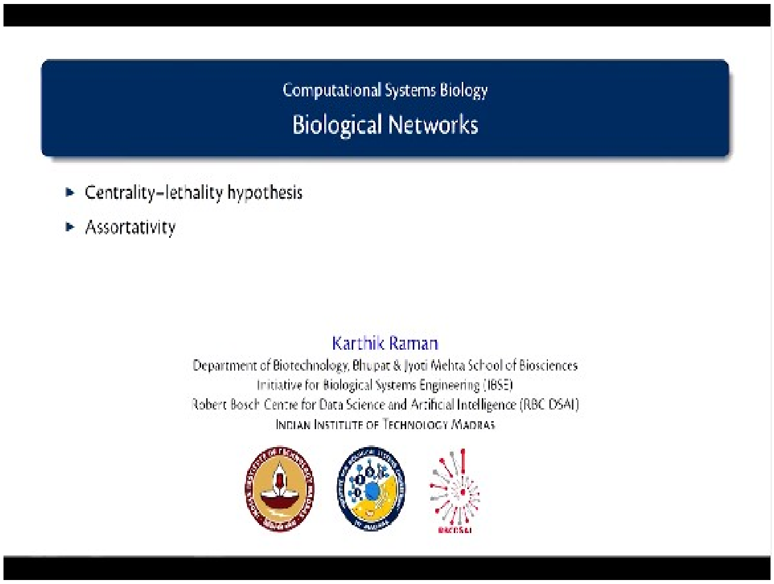
Computational Systems Biology

Karthik Raman
Department of Biotechnology
Indian Institute of Technology - Madras

Lecture – 21
Biological Networks

(Refer Slide Time: 00:11)



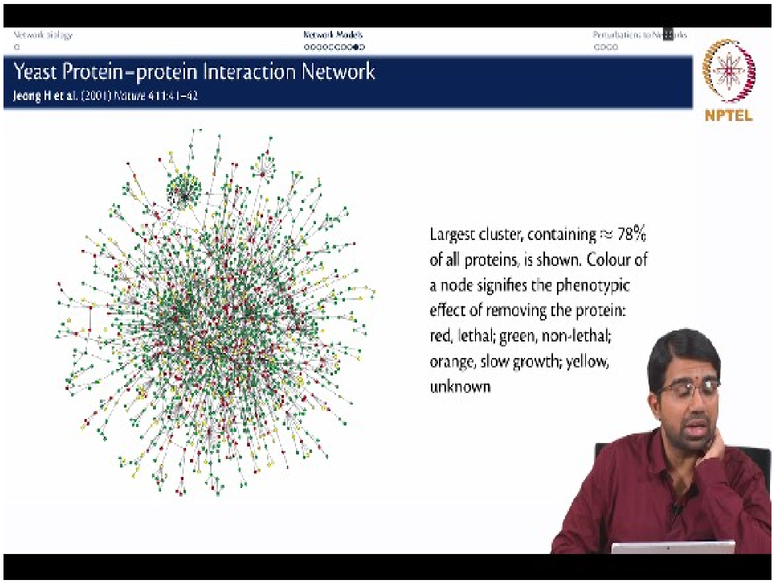
Computational Systems Biology
Biological Networks

- Centrality–lethality hypothesis
- Assortativity

Karthik Raman
Department of Biotechnology, Bhupat & Jyoti Mehta School of Biosciences
Initiative for Biological Systems Engineering (IBSE)
Robert Bosch Centre for Data Science and Artificial Intelligence (RBC DSAI)
INDIAN INSTITUTE OF TECHNOLOGY MADRAS

In today's lecture, let us look at biological networks and how network theory is applied to biology and a very important aspect of this is something known as the centrality-lethality hypothesis which we will discuss in detail and we will also look at the concept of assortativity in networks and whether biological networks are assortative or disassortative.

(Refer Slide Time: 00:33)



Yeast Protein–protein Interaction Network

Jeong H et al. (2001) Nature 411:41–42

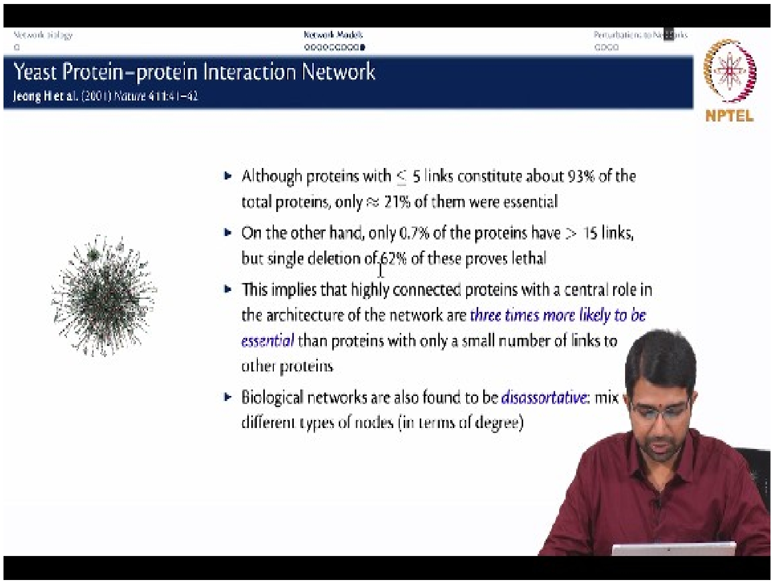
Largest cluster, containing ≈ 78% of all proteins, is shown. Colour of a node signifies the phenotypic effect of removing the protein: red, lethal; green, non-lethal; orange, slow growth; yellow, unknown

So, this is a very popular picture that you tend to see in most biology textbooks or systems biology textbooks and so on. This is a classic experiment. This picture actually shows you the results of a classic set of biological experiments which test the essentiality of every node, every gene in yeast, right. So, this is the yeast protein-protein interaction network experimentally derived and every node has a particular colour.

So, if a node is coloured red, it means the removal of that protein from yeast causes the yeast to die, so it is a lethal deletion. If a node is coloured green, it is non-lethal. If a node is coloured orange, it is a slow grower and if a node is coloured yellow, we will not have experiments that tell you what happens when you delete the node. So, from this, do you already observe something? Is there anything that is sort of apparent from the picture? Hubs are green or hubs are red?

"Professor - student conversation" So, many hubs are actually red, right and some hubs could be green and so on but you find that majority of the hubs are actually red and if you see all the peripheral nodes, you will find that they are all mostly green. The nodes which have very few connections.

(Refer Slide Time: 02:00)



Yeast Protein–protein Interaction Network

Jeong H et al. (2001) *Nature* 411:41–42

- Although proteins with $\leq$ 5 links constitute about 93% of the total proteins, only $\approx$ 21% of them were essential
- On the other hand, only 0.7% of the proteins have $>$ 15 links, but single deletion of 62% of these proves lethal
- This implies that highly connected proteins with a central role in the architecture of the network are *three times more likely to be essential* than proteins with only a small number of links to other proteins
- Biological networks are also found to be *disassortative*: mix different types of nodes (in terms of degree)

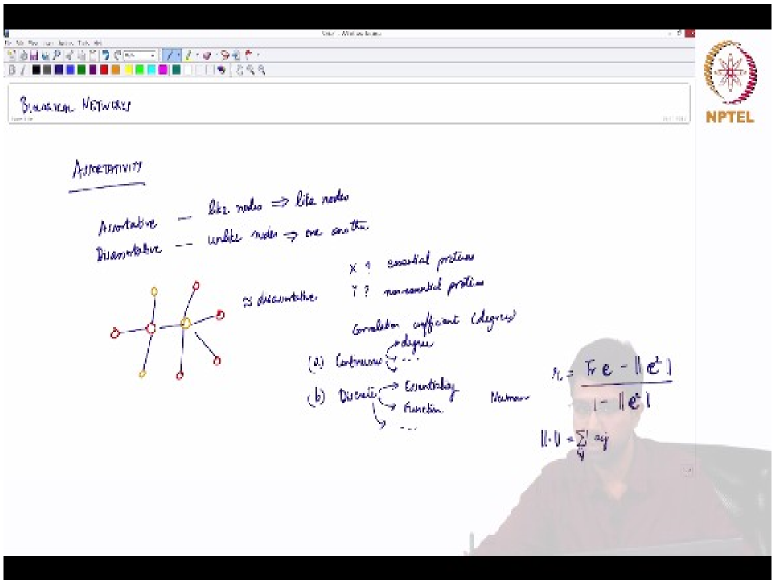
So, what you observe, so in fact if you see, this network actually contains 78% of all proteins in yeast. So, this is the giant strong component I was talking to you about earlier, right. So practically all the nodes, 80% of the nodes fall into a single connected component and although proteins have less than 5 links constitute about 93% of the proteins, only 20% of them are actually essential. On the other hand, there are very few proteins which have a large number of links, more than 15 or so on like just 0.7% but nearly 62% of these are lethal, so the red nodes.

So, these are all red and these are all green. So, what this tells you is that centrality is correlated with lethality and this is actually called the lethality-centrality hypothesis, very important hypothesis. This is in biological networks because the notion of lethality would not arise elsewhere. So, in biological networks, you try to prioritize proteins for lethality or try to rank proteins based on their centrality and you want, these are the proteins that you want to test for lethality.

So, if you want to look at a pathogenic organism's network and you want to try to knock it out, right. You want to kill a disease-causing organism, you may want to find the proteins that are most highly connected in the organism and try to knock them out. There are other things you have to worry about like that should not have a homologue in human and so on.

But this is essentially a strategy wherein you find the most central protein in the pathogenic organism and try to knock it out. Yeah, so this has been, so we have actually studied across organisms. So, there is a paper I probably have it couple of slides down the line and there is also this very interesting concept called assortativity.

(Refer Slide Time: 03:56)



BIOLOGICAL NETWORKS

ASSORTATIVITY

Assortative — like nodes ⇒ like nodes

Disassortative — unlike nodes ⇒ one another.

≈ disassortative

X ? essential proteins

Y ? nonessential proteins

Correlation coefficient (degrees)

(a) Continuous → degree, ...

(b) Discrete → Essentiality, → Function, → ...

Newman

$$r = \frac{\mathrm{Tr}\, e - \|e^2\|}{1 - \|e^2\|}$$

$$\|\cdot\| = \sum_{ij} a_{ij}$$

So, let’s just quickly look at the concept. So, let’s look at what is assortativity? So, networks can be assortative or disassortative. Assortative is like nodes connect to like nodes. Here, unlike nodes connect to one another. So, what is assortativity? When like nodes connect to like nodes, we say that the graph is assortative. When like nodes, I mean nodes of different types connect to one another, you can call the graph disassortative, right.

As an extreme example, a bipartite graph is sort of perfectly disassortative, right. But if you take any general graph, something like this, so in this graph, you have nodes of 2 types. You have red nodes and orange nodes, right and you see that predominantly nodes of different types connect to one another, okay. You do have one red node connecting to another red, then also this is not a bipartite graph but this is something that you would call more or less disassortative.

How do you study or how do you quantify disassortativity or assortativity? So, there are 2 types of measures that you can have. A, if your measure is continuous and B, if your measure is discrete. So, when it is discrete, you may have something like essentiality, function, or any other kind of property. Continuous, you could have degree or other centrality measures and so on. So, for a discrete case, Newman has proposed a formula which goes something like

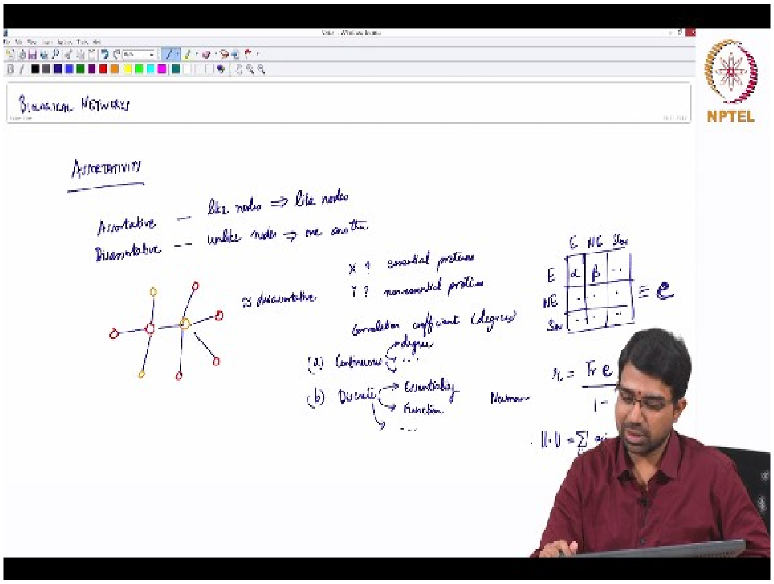
$$r = \frac{Tr\, e - ||e^2||}{1 - ||e^2|}$$

where

$$||.|| = \sum_{ij} a_{ij}$$

So, $||.||$ means sums of all the elements in the matrix. And what is this matrix in the first place?

(Refer Slide Time: 07:34)



BIOLOGICAL NETWORKS

ASSORTATIVITY

Assortative — like nodes ⇒ like nodes

Dissortative -- unlike nodes ⇒ one another.

≈ dissortative

X ? essential proteins

Y ? non-essential proteins

Correlation coefficient (degrees)

(a) Continuous — degree ...

(b) Discrete → Essentiality → Function ...

Newman

$$\begin{array}{c|ccc} & E & NE & Slow \\ \hline E & \alpha & \beta & \cdot \\ NE & \cdot & \cdot & \cdot \\ Slow & \cdot & \cdot & \cdot \end{array} \equiv e$$

$r = \frac{Tr\, e}{1-}$

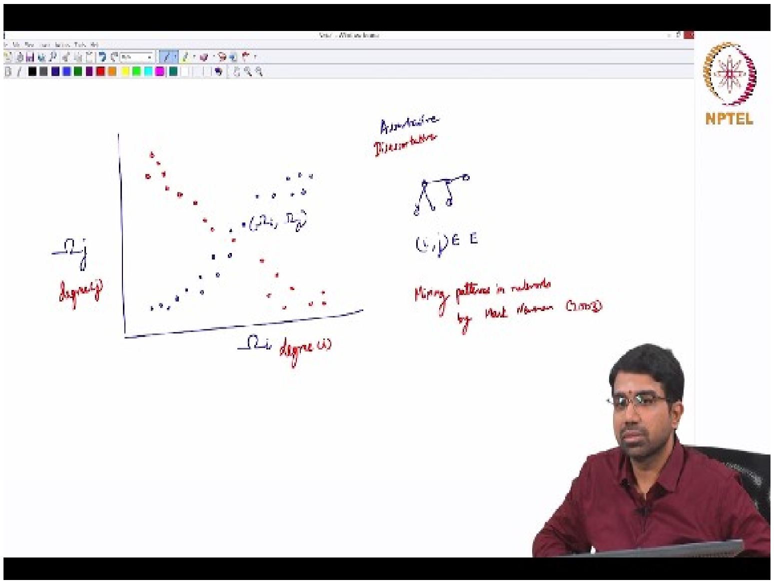
$||\cdot|| = \sum a_{i}$

It is a matrix structured like this. So, let’s say this is essential, non-essential and unknown, 3 classes, okay. This is essential, non-essential, unknown. We will have some value here, some value here, so on. This is your e matrix, right and this is the trace of the matrix which is essentially. **“Professor - student conversation”**. So, by just counting in the network, I can count how many edges go from E to E, how many edges go from E to NE, E to U and so on. You know if you go back to the previous graph that we had, the protein network, how many are green to green, green to red, red to yellow and so on, right. We had 4 classes there, slow as well, so maybe that is a more interesting class to have here. So essential, non-essential, slow. Fraction. What is the trace of a matrix? Sum of all the diagonal elements.

"Professor - student conversation" This happens to be a better way to quantify it. So, this number actually goes from -1 to +1. So, -1 would be perfectly disassortative, +1 would be perfectly assortative and you know values in between are neither assortative nor disassortative. So, this is the discrete case. What happens in the continuous case?

Would you be able to apply this? Well, yes. You make a couple of bins, make it discrete and apply this, right. So, you say anything > 20 links is high degree, anything < 5 links is low degree, other things are intermediate degree, so high, low, intermediate, you make 3 classes, make this matrix and you can do, do some calculations like this.

(Refer Slide Time: 10:04)



The alternative would be actually try and plot a graph, right. So, let's take an arbitrary graph here. For this graph, for every node, you can compute the property of interest or for every edge, right. So, for every (i, j) ∊ E (edges in the graph), can you plot the properties? Let's say the property we want is some Ω. So, you put Ω_i here and Ω_j here, right. So, you essentially plot some statistics. It could degree, it could be anything else, it could be betweenness centrality, nodes of high betweenness centrality connect to nodes of high betweenness centrality.

And you can essentially make a plot of this sort. So, if you end up with the plot that looks like this, it's going to be assortative. If you end up with a plot that looks like this, it is going to be disassortative. That makes sense? You can compute some correlation coefficient you know, a

Spearman correlation coefficient or a Pearson correlation coefficient to see what kind of correlation is there between.

So, so that network is assortative on degree but disassortative on betweenness, right. If one statistic, one Ω gives you assortative behaviour and the other Ω gives you disassortative behaviour, fair enough, right. So, you will find that biological networks are typically disassortative when it comes to essentiality, disassortative when it comes to degree. Actually heavily disassortative.

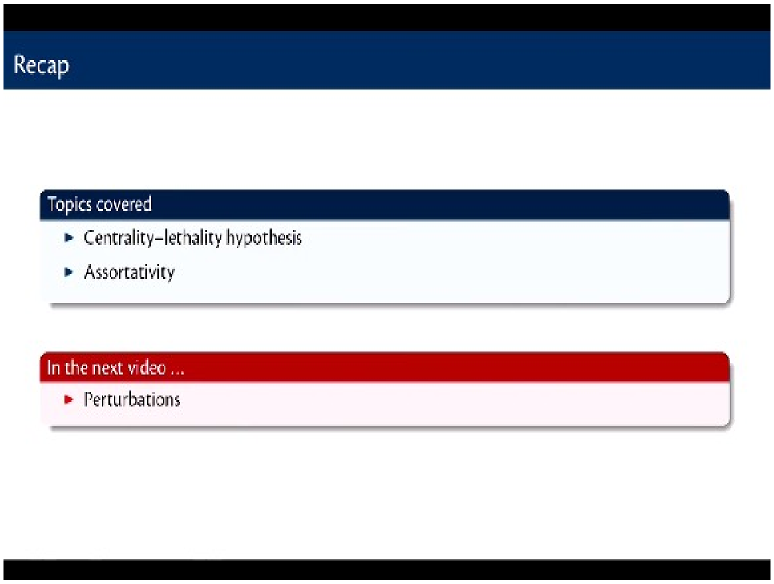
You may not find much assortativity in biological networks. There is a very good paper called mixing patterns in networks by Mark Newman. This is where he describes these, the formula for assortativity that I showed you on the previous slide and so on. For example, especially disassortativity in degree, let's talk about robustness. High degree nodes don't just connect to one another.

Right, then you remove one high degree node, other high degree node also loses counterpart and so on. But it is very difficult to make such sweeping statements about these networks because in practice, these are all global properties, not local properties, right. So, you will have counter examples everywhere. So, for example, you will find a number of essential proteins that have low degree and a number of high degree nodes that are non-essential and so on, right just to give an example for the, a counter example for the centrality-lethality.

But in general, you see that these properties are by and large followed. So, I will also post a reference to our paper where we studied this across 20 different organisms, right and we find that both, betweenness centrality and degree, seem to give you a good amount of correlation with lethality. Higher the betweenness centrality or higher the degree, higher the chance of the protein being essential. Shreya, you had a question?

"Professor - student conversation" Yes, yes, yes. The degree of i, so, let's say for example, let's say, this is degree(i) and degree(j). One example for omega would be degree but you can have betweenness, you can have any other interesting property.

(Refer Slide Time: 14:29)



Recap

Topics covered

- Centrality–lethality hypothesis
- Assortativity

In the next video ...

- Perturbations

In this video, I hope, you got an interesting insight into how biological networks are studied using the tools from graph theory and we looked at the centrality-lethality hypothesis and also the concept of assortativity. In the next video, we will look at a very important aspect of systems biology as I always mention which is what happens to a network when you start perturbing it and how do you study perturbations?