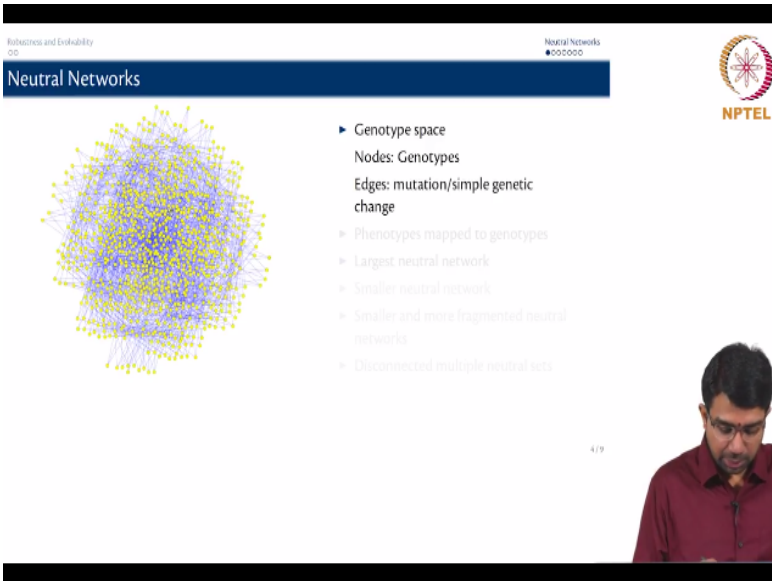
Computational Systems Biology
Karthik Raman
Department of Biotechnology
Indian Institute of Technology – Madras

Lecture - 98
Robustness and Evolvability

So in this last lecture on robustness and evolvability we will closely study the concept of neutral networks which are very important for our understanding of robustness and evolvability and how they can coexist in biological systems and also discuss genotype-phenotype mapping which is central to how we build these genotype networks or phenotype networks and the corresponding neutral networks.

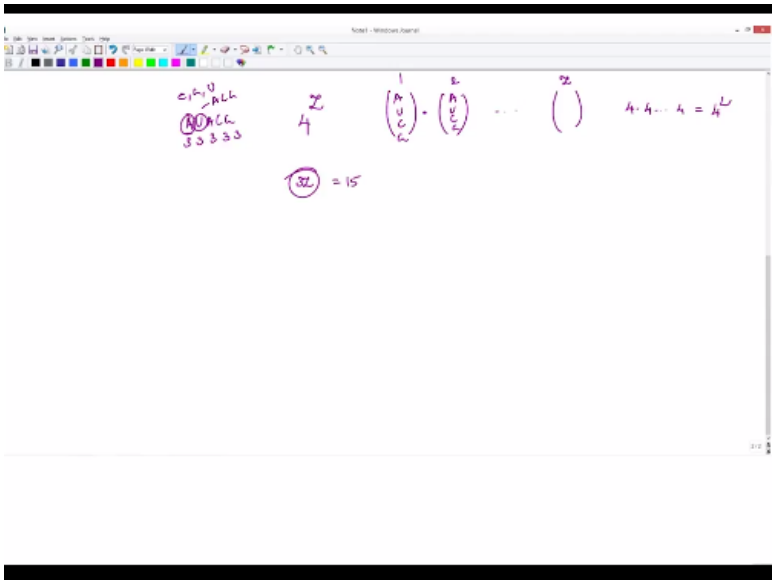
Let us try to understand the concept of neutral networks. This is very central to our whole understanding of robustness and evolvability in biological systems.

(Refer Slide Time: 00:48)



So you now start imaging a genotype space right where you have, what I show hear on the left is a genotype space. So every circle, every yellow coloured node here is the genotype and 2 nodes are connected by an edge if they can be reached by a simple mutation. So they are genotype neighbours. So if you want to have a quick example.

(Refer Slide Time: 01:17)



Let us consider an RNA sequence of the sort right, it is a very small RNA sequence. If you did point mutations, you can get marry more RNA sequences. So how many RNA sequences can you get from this. You can have 3 variations here. You can have 3 variations here. You can have 3 variations here. You can have 3 variations here. You can have 3 variations here. So what is the total number of possibilities, 3 + 3 + 3 + 3.

You have only 15, you cannot combine the two, because that would be more than one mutation. So before we get there what is the total number of let us say you have an RNA sequence of length L. What is the total possible number of such RNA sequences, how many different RNA sequences of length L can you have, 4 to the L, right? Because you can have A, U, C or G in position 1 * A, U, C or G in position 2, so on till position L.

So this 4 * 4 * 4 = 4 to the L. So you can imagine the space is massive, if you are talking about 30, an RNA sequence of length 30 you are about 4 to the 30 or like 2 to the 60 which was basically going to be somewhere around 10 to the 18, very massive number, but were as the number of neighbours are only 3L because you have, so you have A here, you can have C, G or U here.

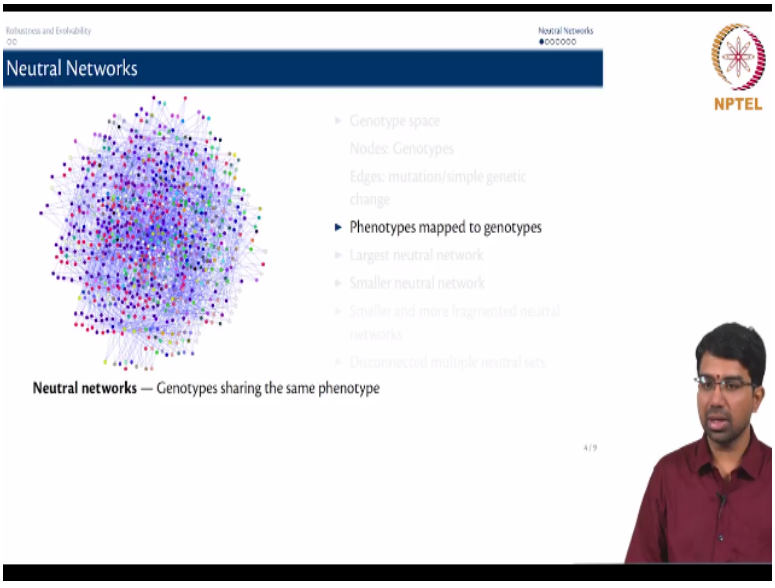
You have U here, you can have A, C or G here and so on. So each of these would be a separate sequence, you have a total of 3L sequences, 3L neighbours, in this case it is going to be 15 neighbours. So if there were the genotype space for size 5, how many nodes would we have 4 to the 5 or 2 to the 10 which is 1024. You have about a thousand nodes here right and how many edges will you have? Every node will have 15 edges to each of those neighbors.

So you will have a regular graph wherein there are 1024 nodes each with 15 outgoing edges. So the total number of edges will actually be 1024 * 15/2 because you will have them both ways, that is going to be a reasonably dense graph, might look somewhat like this. This is nothing but the genotype space, does it make sense? you understand what a genotype is, like you know DNA or RNA or any underlying aspect you could even think of it as a network topology, we will come to that in a moment.

But genotype is the underlying genetic makeup of an organism whatever that is subject to change. Phenotype is a sort of higher level observation. So it is a trait for example, can an organism grow in a particular environment or not, or you know hair colour, eye colour, these are common phenotypic traits or it could be say RNA structure, protein structure so on and so forth.

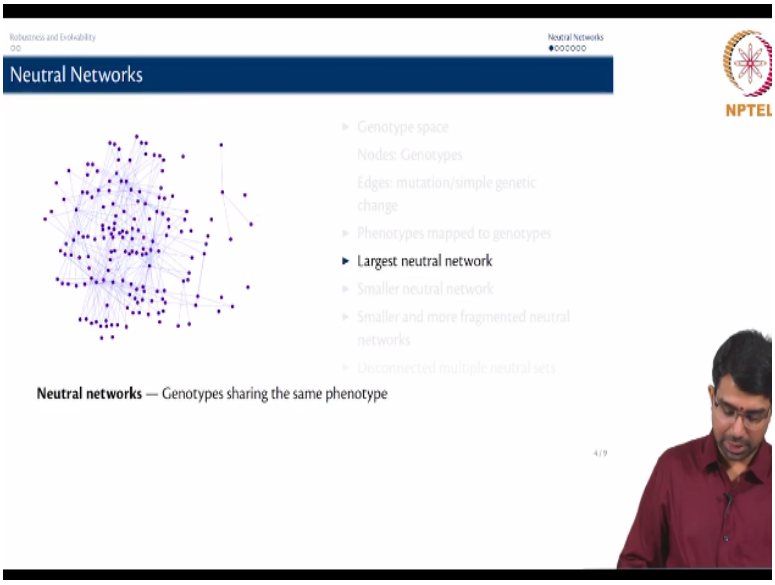
So all these could be potential phenotypes. So now let us take this genotype network and map the phenotypes on to this, which means we essentially colour up the network.

(Refer Slide Time: 04:47)



So every colour now represents a particular phenotype. If you see a red colour all the red nodes together form is like you know the same phenotype, all the violet nodes are one phenotype, all the grey nodes are one phenotype and so on and so forth. Now what is the neutral network? These are networks of genotypes that share the same phenotype, meaning, if I pulled out one colour out of this graph, that would be a neutral network.

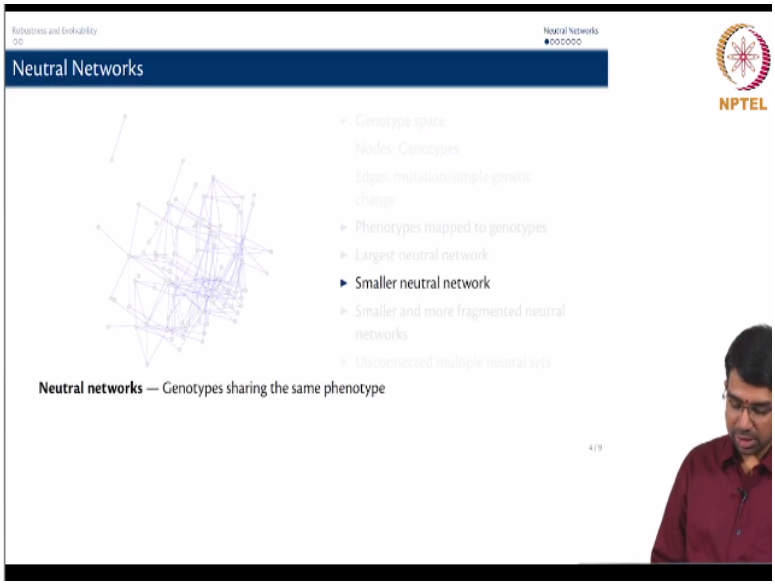
(Refer Slide Time: 05:17)



For example, this is the largest neutral network on this graph. It is technically a set, not a network because you have certain nodes that are disconnected and so on, but you can see that this is the network. What is this network mean for us? What is the node again? It is a genotype. Do all, and what is unique about all these nodes? They share the same phenotype, which means that if you are here, if you start on this node, you can mutate here, then mutate here, then mutate here, here, here, to any of the neighbours without changing your phenotype.

What does that mean? It means it is robust, it can change genotype, but phenotype really does not change. So that phenotype is actually robust.

(Refer Slide Time: 06:15)

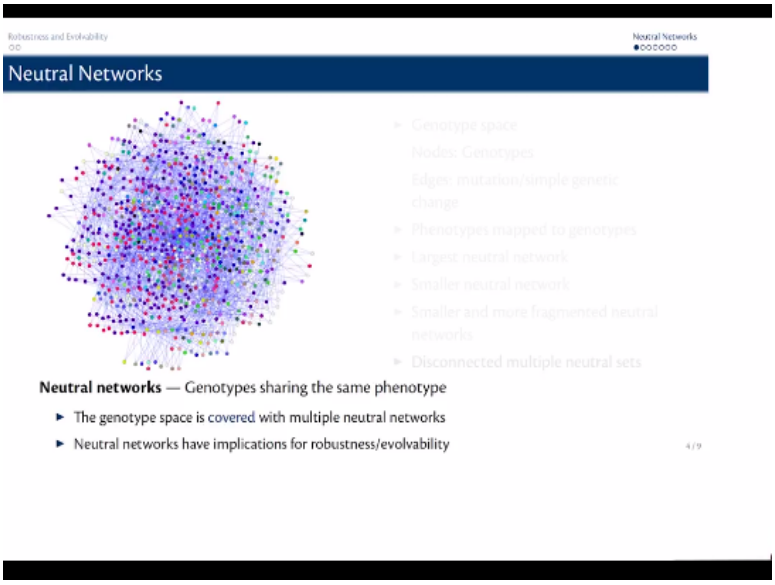


This is a smaller neutral network, incrementally smaller fragmented neutral network.

(Refer Slide Time: 06:22)

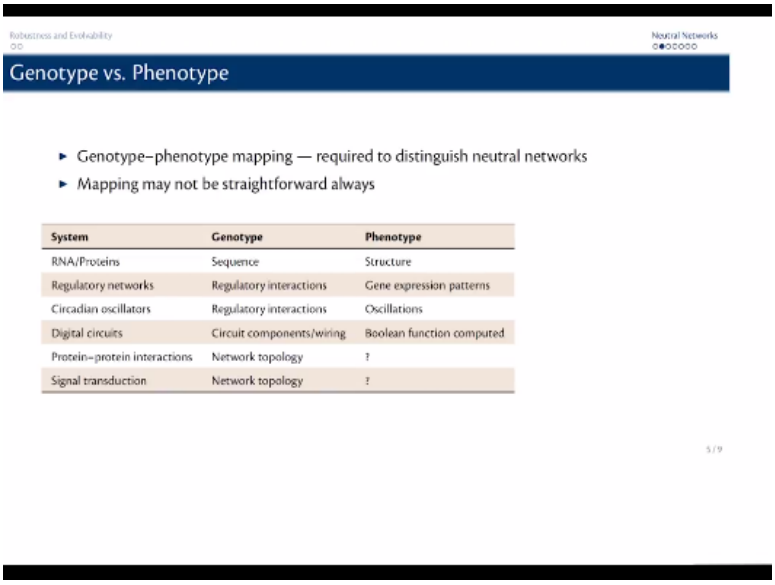
And it is just multiple neutral sets, disconnected multiple neutral sets. So what implication does this have.

(Refer Slide Time: 06:28)



So you can see that neutral networks you know just completely lays the entire genotype space. So the genotype space is basically overlaid with several neutral networks and these neutral networks have implications for both robustness and evolvability. We will see how that is.

(Refer Slide Time: 06:47)



So let us first understand this genotype-phenotype mapping because that central because if you have to think about what a colour means and so on, you need to have a mental picture of what is a genotype and what is the phenotype. So a very straightforward example of genotype and phenotype is sequence and function. Function is usually harder, so maybe sequence and structure as you see here.

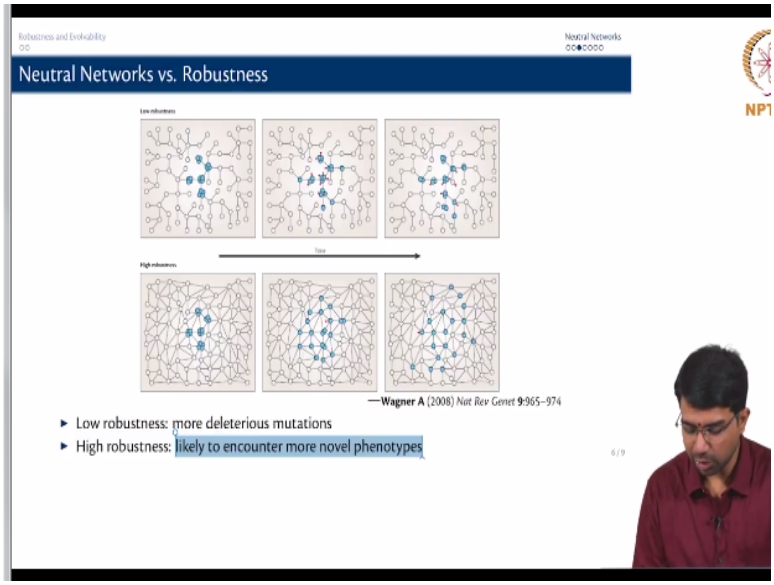
So RNA or proteins you can think of sequence being the genotype and structure being the phenotype. For regulatory network, you can think of the regulatory interactions of the regulatory topology, being the genotype and gene expression pattern being the function. For a circadian oscillator I am giving these examples because these are from interesting published studies in the past.

So for circadian oscillators the regulatory interactions can again be the function and be the genotype and whether or not an oscillation is observed can be the phenotype and we also studied digital circuits in the past where in the circuit or components can be the circuit components of the wiring can be the genotype and what is the Boolean function that is computed by the circuit that can become the phenotype.

Then we can try to design a robust circuit and so on. Similarly, here you can try to design a robust oscillator and things like that, but it is a little harder if you are looking at say protein-protein interactions or signal transduction and so on, it is difficult to really map what is you know sensible genotype or you know the genotype is maybe not that difficult but how do you make the phenotype.

And to add to this one can really imagine metabolic networks where the presence or absence of reactions can or genes can be considered as the genotype and phenotype can be computed using flux balance analysis. You can just do an FBA simulation to see if the organism grows or not and use that as a phenotype. This mapping may not be a straightforward always, it could be a little tricky.

(Refer Slide Time: 08:41)



So let us look at how neutral networks play a role for robustness. Do you see any similarity between the 2 graphs, the nodes are actually the same, the number of nodes are actually the same? Whereas the network on the top has far fewer edges and the network on the bottom. You can clearly say the average degree of the network on top is much lower than the average degree of the network in the bottom.

So now let us imagine all these coloured dots that you see there, let us just imagine that they are organisms evolving on these neutral networks. So what is evolution mean here, what happens when a mutation happens? It will move along an edge; can it move outside the neutral network?

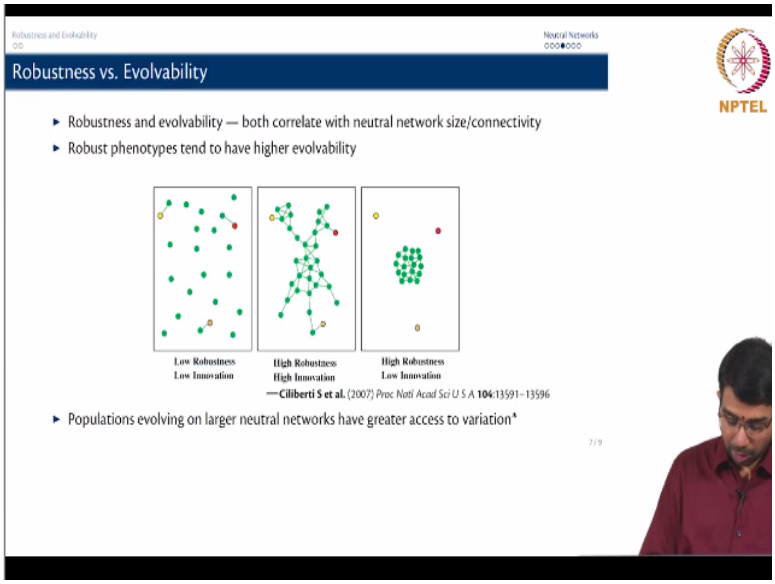
Yeah, because this does not account for the entire space, this is just a neutral network, this is where the genetic space then, it catalogues all possible mutations that can occur, this is only a subset of the genotype space you know essentially a network layer of the same colour that we saw in a couple of slides ago. Wherein you start with these with some node here and you can take any of the edges.

There are some potential edges that are not on this as well, but what is the issue with those edges, taking those edges will change your phenotype, perfect, very good. So now if you let them evolve for a while, what happens? you see that here there are many more edges that are taken, which take you out of the neutral network and for all you know that could be a lethal phenotype.

So let us assume it is a lethal phenotype and those organisms die, which is indicated by the small red arrows that you see and if you keep proceeding you will see that all these in the neutral network on top, the organisms are still stuck around wherever they started more or less, whereas here they have been able to percolate to distant parts of the genotype space, like in this neutral network you are very likely to have more deleterious mutations.

Whereas if you have high robustness you are more likely to be able to explore the space and as a consequence you might encounter novel phenotypes. We will see what that is in the next slide.

(Refer Slide Time: 11:06)



So you are talking about the last point that I meant. So robustness and evolvability both somewhat correlate with neutral network size and connectivity. Size alone is not sufficient; connectivity alone is not sufficient as you will see here. So this is a network which is highly connected and you know but somewhat very tightly connected. This is a network that is completely disconnected.

This network is intermediate, it is you know, it is well connected has what is the important parameter, network parameter that differs between these 3, average degree maybe, what else another very simple parameter, clustering, yes, more obvious, clustering does vary I agree, most obvious diameter. Diameter here is essentially 1 or something. Here it is very large could be like you know 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or something like that at least.

Whereas here the diameter is very small it could be just be 1, 2, 3, 4 or something like that and of course there are disconnected sets and so on. So what does this mean? So here if I start off on this genotype, I can never reach the red phenotype, here if I start on this genotype, I can again not reach the red phenotype at all. Whereas here if I start on this genotype, I can have several neutral mutations that can finally take me towards red.

I can reach the red phenotype here. So this is what one calls a novel phenotype, potentially a novel phenotype. You could argue this could be a lethal phenotype as well, but the thing is you have access to several more phenotypes in the neighborhood. If I were to look the entire space like in the previous figure or even here. So if I am these nodes, on this large neutral network.

I can probably move a lot and suddenly I find myself close to so many other phenotypes or if I go to this region on space which is still accessible through neutral mutations, I can again access many more phenotypes or if I go to this region of space I can still access some more genotypes with different phenotypes in the neighbourhood, does that make sense? So all this has implications for robustness and evolvability.

I would say robustness is nothing but average degree, that is not necessary robustness right because robustness is the ability to maintain the same phenotype. Evolvability would be that, evolvability it is the ability to access new phenotypes in the neighborhood. So robust robust phenotypes tend to have higher evolvability, meaning, if you are robust like this phenotype you could have access to more genotypes with different phenotypes in your neighbourhood.

What do you mean by quickly, if I had other you know nearby genotypes with different colours that is still going to be reasonably good, that still going to be evolvable, here this is not evolvable because this is just, this does not seem to be connected to any other phenotype

at all. **"Professor - student conversation starts"** If I had phenotypes like if you had like multiple red phenotypes all around it then would it still be robust?

Yes, assuming that you know this is still the network, you still have enough phenotypes, enough genotypes that share the same phenotype, right, the size of this neutral network is still large. So you can still accommodate several mutations whereas here you seem to be able to accommodate even more mutations. Right if you count the number of edges here it is clearly higher.

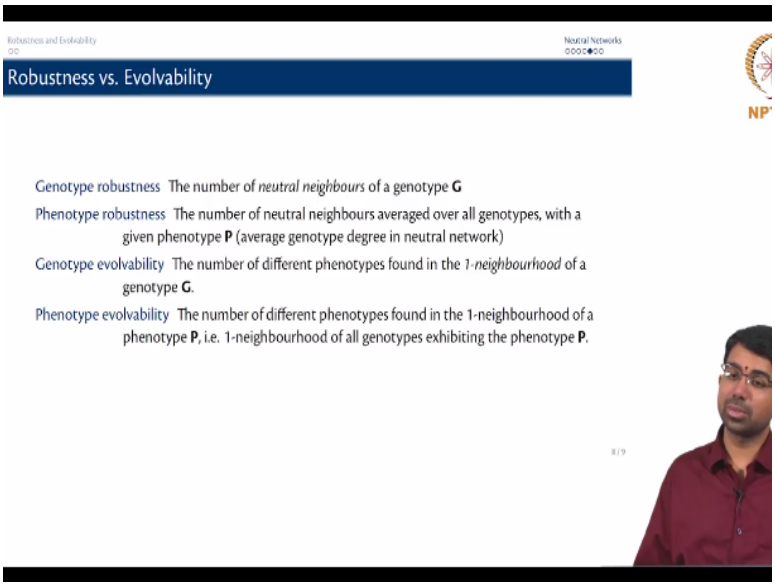
So if it is more likely to stay within the same genotype as robust but it should also be able to (()) (14:40) yeah. **"Professor - student conversation ends"** But you can imagine that if you have a network that is very large then very likely in most genotype spaces you are going to have access to new phonotypes in the neighborhood, because as we saw the way at least we define the genotype spaces and so on, they are pretty regular graphs right.

So every node is going to have n number of neighbours and if some K of them are the same phenotype that is the average degree, that is going to be a robustness, but the rest of them are going to have different phenotypes and for a larger network the neighborhood size will be even larger. So we look at those definitions in a moment. And populations evolving on large neutral networks have greater access to variation.

This is a large neutral network, so you have good access to variation. I can just move all around this place and suddenly come close to a red phenotype or some other phenotype and so and on. So the fact that this neutral network is structured thus becomes very important, right. What is the structure? So just cannot just look at the, if you look at the number of nodes it is the same in all cases.

If you look at the number of edges it might be somewhat similar here, but you also need to look at the diameter and the connectivity and so on and so forth.

(Refer Slide Time: 15:51)



So now let us look at some definitions, some interesting definitions. What is genotype robustness? So it is the number of neutral neighbours of a genotype G, does it make sense. So number of neighbours that share the same phenotype. Meaning these are the edges if I am at a particular genotype, these are the edges that I can take without disturbing my phenotype. So what is phenotype robustness, is the number of neutral neighbours averaged over all genotypes with a given phenotype.

So this is nothing but the average degree of the entire neutral network, does it make sense. What is a genotype again? It is a node, very good. What is the phenotype? It is potentially the entire neutral network. So if you want now compute robustness of the phenotype, you essentially look at the entire neutral network right. So what is the average degree across that neutral network.

What is genotype evolvability? It is obviously the number of different phenotypes found in the immediate neighbourhood of a genotype or essentially the number of non-neutral neighbours, you are right. So genotype robustness and genotype evolvability are negatively correlated, but we never look at genotype robustness right. You always look at phenotypic robustness.

How does a particular behaviour remain robust, right, so it turns out that phenotype evolvability and phenotype robustness are positively correlated? What is phenotype evolvability? It is the number of different phenotypes found in the immediate neighbourhood

of a given phenotype that is the one neighborhood of all genotypes expressing the same phenotype or exhibiting the same phenotype, one immediate neighbourhood, it is a network.

So the neighbourhood of the entire network. What is my entire zone of access, right? So let us go back to the. So in this case the one neighborhood of the neutral network contains how many phenotypes, in the center, it contains 3 phenotypes, right, here one neighbourhood contains 0, here neighbourhood does contain 3, but the network itself is completely disconnected.

So potentially this is what is innovation, why is this innovation it is interesting to understand. So let us say you start here, you are a normal RNA sequence or your metabolic network that is just growing on some substrate and so on. You keep harbouring certain mutations. So what kind of mutations are these? These are neutral mutations. If you stay on the network, they are all neutral mutations.

If you fall of the network they could be non-neutral mutations and so on, but what is a neutral mutation? The phenotype remains the same, so it could be some phenotype. Obviously you are often looking at a particular phenotype, there are other phenotypes that might change. So because phenotype is not the whole thing. So you may look at ability to say grow in one carbon source that is your phenotype, that ability might remain.

But as you keep neutrally moving there you might find that other abilities improve or worsen as the case maybe and this is very interestingly found even in protein structures and so on, wherein you find that there are like let us say you have a protein that carries out a particular function. Let us say it is alcohol dehydrogenase. So what is important for this function? some active site residues, right.

But therefore as long as you keep the active site constant and you change a few other things in other parts of the protein, the protein can retain it is ADH function, alcohol dehydrogenase function, but at the same time mutate. It can harbour mutations in other parts of the structure and maybe it now develops an active site that can do a succinate dehydrogenase. So it can bind succinate substrate or something like that.

So because this was constant, the organism survived in the first place. So it slowly harboured neutral mutations which enabled to then get close to. So this is let us say alcohol dehydrogenase, it is slowly moving, moving, moving, mutating. Alcohol dehydrogenase ability has remained unchanged all around, but suddenly it is close to picking up a succinate dehydrogenase ability.

Then it has some function ability. So how would you generally define a phenotype? usually function right and even when you say because even if you say structure you are always finally concerned about function being the phenotype. Structure is usually a proxy for function, it might change, it would change, right. So in that case you know nothing would be a neutral mutation right.

So every small, but potentially you know some amino acids changing may not significantly affect the structure as well. So you may have to consider it that way, right, because if you then say I am going to permit no change in the structure then no mutation can ever be neutral. So if you are talking about DNA sequence to protein structure, yes, obviously because of the degeneracy there going to be many silent mutations and so on.

But even they are not essentially silent because they can change the expression which can finally change the function, phenotype and so on, right, because what sequence is there, supposed GGG, how does it get translated? GGG anticodon has to bind right, which is carried by the GGG t-RNA. If the GGC tRNA is less in conservation compared to GGG tRNA the expression of the protein itself is going to change.

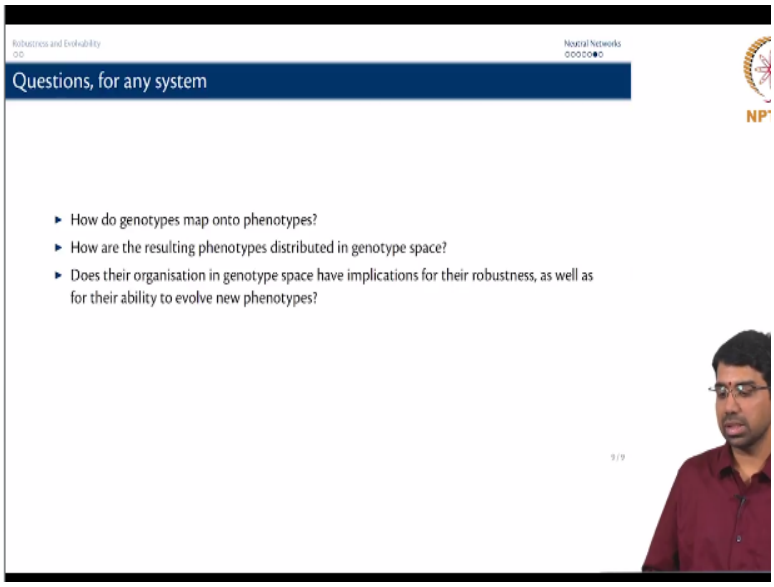
So neutral mutations may not always be neutral, so how do you actually define neutral becomes you know a little bit of a contentious point, but essentially the idea is this that if you have a large neutral network you can accommodate, harbour more mutations without damaging the phenotype and it also gives you the ability to get closer to newer phenotypes because you started off here you can go around all along and reach some place like this.

Whereas here you cannot go anywhere and here you can just keep circling around here. You can still never reach one of the novel phenotypes. So because of these definitions which are actually very reasonable, this is a very interesting paper published by Andreas Wagner in

2008, titled Robustness versus Evolvability: A Paradox Resolved. This gave a very good resolution of the seemingly paradoxical relationship between robustness and evolvability.

Because robustness is the ability to resist change and evolvability is the ability to get new changes and innovations and so on and you find that phenotype robustness and phenotype evolvability are very possibly correlated whereas genotype evolvability and genotype robustness are quite negatively correlated.

(Refer Slide Time: 22:30)

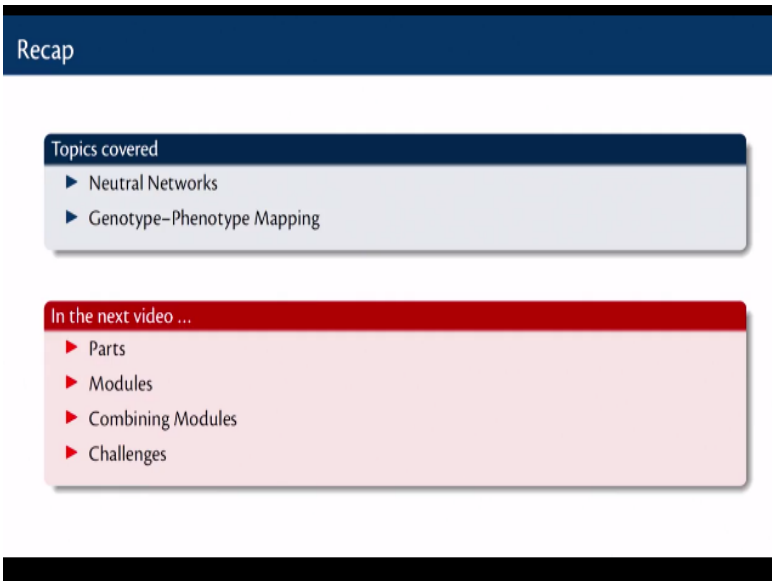


So for any system you have to answer the following questions. What is a genotype-phenotype map? What are your genotypes? What are your phenotypes? So classic examples are I take gene networks as my genotype and phenotype is the ability to show an oscillation, is it oscillate or not. So everything that oscillates forms one phenotype and how are the resulting phenotypes distributed in the genotype space.

Does this organisation in genotype space have implications for the robustness as well as evolvability. So these are the kind of questions one ask of the system. There are many interesting papers that have been published in this space and I will share some of them with you, can have a look.

So this is essentially a brief and quick introduction to evolutionary systems biology. It is a very important field inside of systems biology wherein we answer a lot of questions through in silico evolution and so on and very important for say synthetic biology even, how do you design robust genotypes or robust phenotypes and so on and so forth.

(Refer Slide Time: 23:27)



So in today's video we overviewed the concepts of a neutral network and how we perform a genotype-phenotype mapping that enables us to construct and analyse these topologies and the degree and the various aspects of these neutral networks and so a lot of work in this area comes from the lab where I did my postdoctoral research which is the lab of professor Andreas Wagner at the University of Zurich.

So I suggest that you can go and look at some of those very interesting papers that are coming out of that lab and following this we will switch gears and move to the final topics in this course. We will look at synthetic biology wherein we will study the different parts, modules and how these modules have been combined in biological systems. So synthetic biology is essentially a very engineering approach to biology, where one wants to design different kinds of circuits and modules and so on and we will also look at what are all the challenges in synthetic biology.