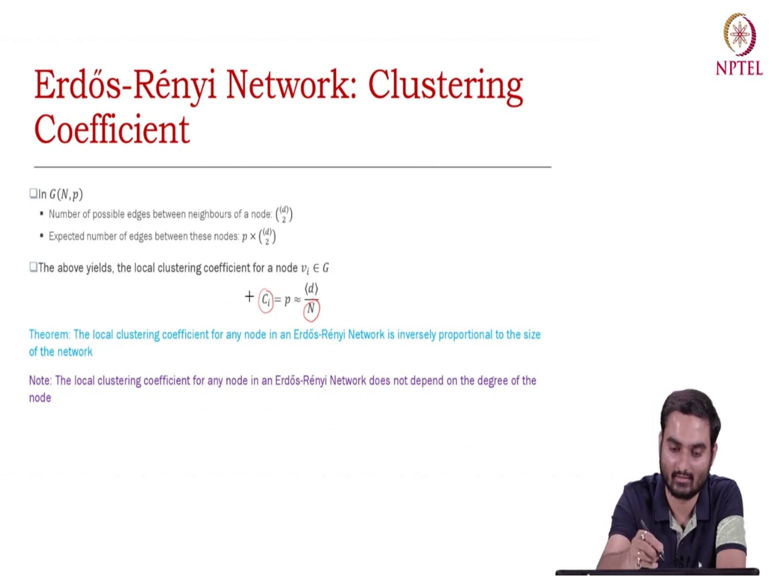
Social Network Analysis
Prof. Tanmoy Chakraborty
Department of Computer Science and Engineering
Indraprastha Institute of Information Technology, Delhi

Chapter - 03
Lecture - 03

So, we have been continuing you know our discussion on network growth model and last lecture we have discussed you know why network growth model is important right why synthetic network generation is important. And we also have discussed a simple model called random graph model ok ER model Erdos Renyi model. And in the ER model we have seen that that the 3 properties that we want to satisfy, one is the high clustering coefficient high local clustering coefficient the second one is scale free property and the third one is small world property.

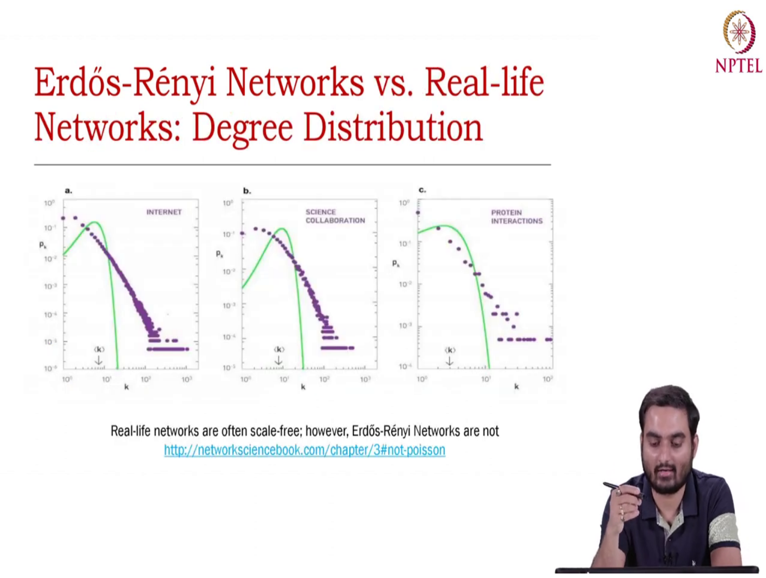
These 3 properties need to be satisfied, but in case of ER model we observed that except small world property the other two properties are not satisfied right.

(Refer Slide Time: 01:19)



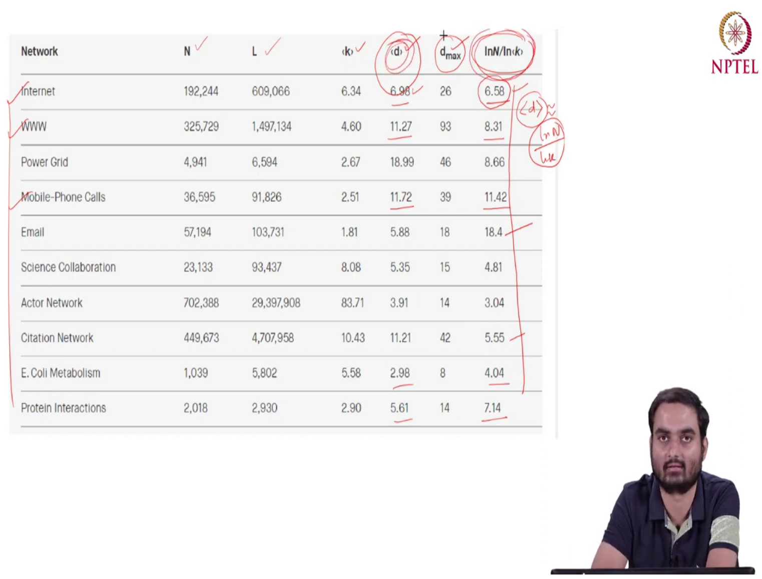
High local clustering coefficient is not satisfied we have seen ester we have seen the last lecture that with the increase of the number of nodes right the clustering coefficient will decrease right. But that should not happen because in the real world network we observed that you know large networks with millions of node the clustering coefficient is still quite high 0.4, 0.5 around right.

(Refer Slide Time: 01:43)



And obviously, the degree distribution follows Poisson distribution or you know binomial distribution in case of ER model which is not the same as the real world model which actually follows power law degree distribution. So, these two properties are not satisfied by the ER model ok.

(Refer Slide Time: 02:12)



Network	N	L	$\langle k \rangle$	$\langle d \rangle$	d_{max}	$\ln N / \ln \langle k \rangle$
Internet	192,244	609,066	6.34	6.98	26	6.58
WWW	325,729	1,497,134	4.60	11.27	93	8.31
Power Grid	4,941	6,594	2.67	18.99	46	8.66
Mobile-Phone Calls	36,595	91,826	2.51	11.72	39	11.42
Email	57,194	103,731	1.81	5.88	18	18.4
Science Collaboration	23,133	93,437	8.08	5.35	15	4.81
Actor Network	702,388	29,397,908	83.71	3.91	14	3.04
Citation Network	449,673	4,707,958	10.43	11.21	42	5.55
E. Coli Metabolism	1,039	5,802	5.58	2.98	8	4.04
Protein Interactions	2,018	2,930	2.90	5.61	14	7.14

(Refer Slide Time: 02:22)

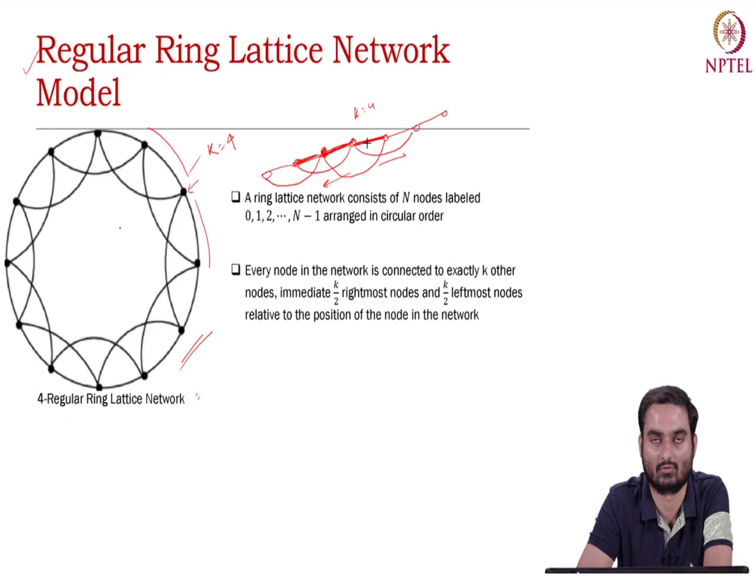
And therefore, you know if you look at the statistics right. Let us look at the statistics and let us try to understand that you know the last lecture we mentioned that the you know the; you know the path length right the average path length which actually you know indicates the small world behaviour, it proportional to the logarithmic of the number of nodes right. It is basically log N by; log N by log d or log k which is the average degree right.

So, if you fix the denominator it basically scales with log N. So, let us see whether you know this approximation of the average path length actually is same as you know the average path length that we generally see in case of real world networks right. So, if you look at the statistics here right there are several real world networks here and you see the number of nodes, number of edges right average degree, average path length, maximum path this is the diameter of the network and this is log N by log k.

So, this is. So, we basically want to see whether the average path length whether it is approximating this one whether this quantity actually approximates the average path length ok. So, this is the; so, this the last row indicates the approximation that we can actually measure. You see that you know the numbers are quite same, 6.98 6.58 in case of internet network, in case of world wide web 11.27 8.31, in case of say mobile phone 11.72, 11.42.

For some networks yes these are not exactly same for example, this one or this one right, but if you look at in general right they are more or less same ok. So, this is the average path length that you can observe right in a real world network and this is the approximation that you can actually observe right. So, we see that this is more or less I mean this quantity is more or less same as the this fraction ok.

(Refer Slide Time: 04:32)



But nevertheless, I mean the you know the final observation is that this ER model random growth model is not enough because of the you know non scale free property and the clustering coefficient which is basically quite low. So, in order to address this problem these two problems, so later on different other models had been proposed and one such model was the regular ring lattice network model ok.

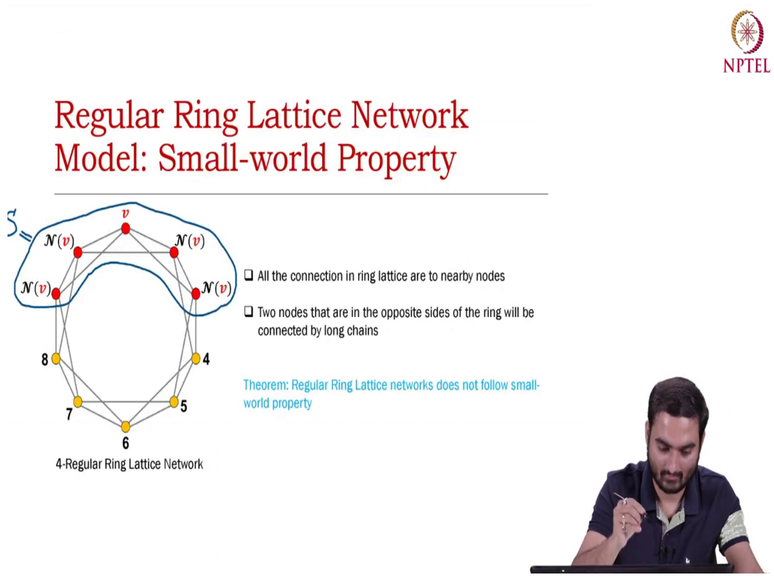
The ring lattice network model is a very simple you know network model that we can think of I mean generally no one will believe that you know this model actually mimics the way real world network grows over time. But at least this is a you know this is a model by which you can show that the clustering coefficient is would be high if you follow you know the ring lattice network growth model ok.

So, let us try to understand what is this ring lattice model ok and this is regular graph meaning that nodes have same degree ok. So, first let us fix the degree of anode in this network. So, let us see let us fix the degree as k ok. So, each node has degree k ok in this particular example k equals to 4. So, this is actually a 4 regular ring lattice network and how this lattice structure is basically constructed?

So, for every node, so every node is connected to k by 2 nodes in the left side of it and k by 2 nodes in the right side of it ok meaning that if you linearized the structure ok it basically would look like this ok. So, say for this particular node if k equals to 4, it means that it is connected to two nodes in the right side and two nodes in the left side. So, this is one such connection this is another such connection ok and this is one such connection this is another such connection in this way.

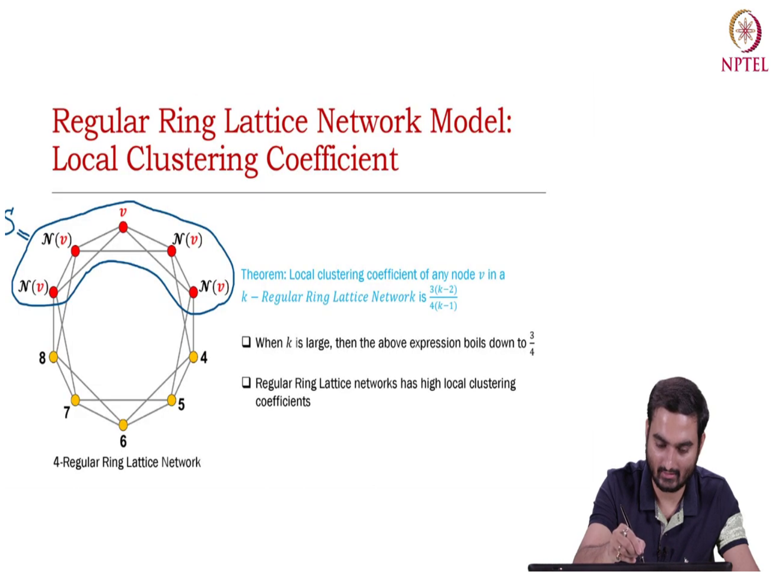
Similarly, for this node you know it is connected to this node already. So, it is will also be connected to this node, this and this ok and so on and so forth. So, this example is for k equals to 4 ok. So, if this is the structure what would be the local clustering coefficient of a node ok? Let us try to understand it first.

(Refer Slide Time: 07:09)



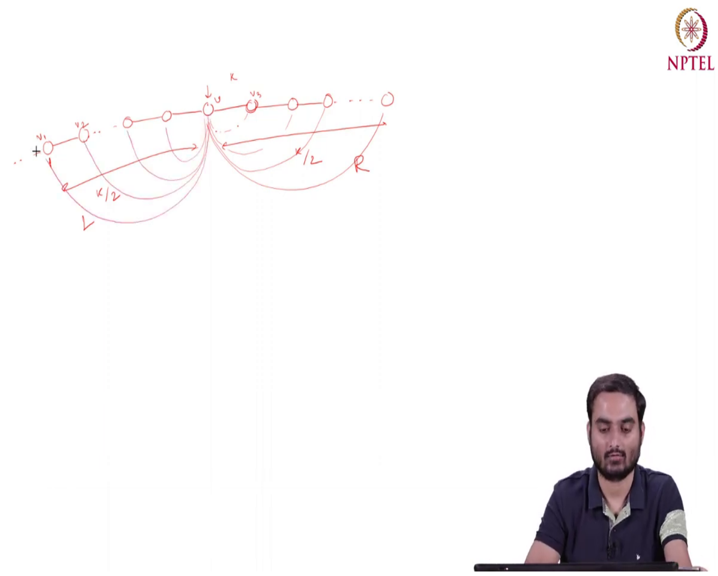
Let us let us think about it ok.

(Refer Slide Time: 07:14)



Let me draw a graph first, let me draw you know a graph.

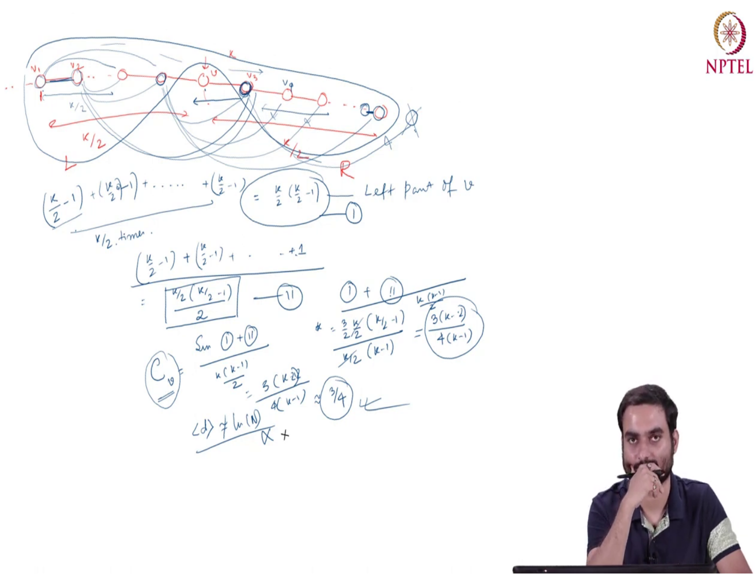
(Refer Slide Time: 07:21)



So, let me you know linearize the lattice structure ok. So, you will have like this ok and let us say ok and let us only focus on a particular node called v alright. So, degree is k therefore, v is connected to k by 2 nodes in this side and k by 2 nodes in this side ok. So, let us say the last node ok the k by 2 th node in the left side. So, this is left this is right k by 2 th node in the left side is v 1, this is v 2. And similarly in the first node in the right side is v 3 and so, on and so, forth ok and how these connections are actually happening?

So, basically it is connected to this one, it is connect like this ok like this right. And we are interested in looking at the number of edges right number of edges among the neighbours of v that is the local clustering coefficient among the neighbours of v.

(Refer Slide Time: 08:53)



So, we are interested in understanding or measuring the number of nodes among these nodes. This is clustering coefficient we take the induced sub graph and then we measure the number of edges not number of nodes number of edges induced I mean number of edges in the graph induced by the neighbours ok. So, now let us think of the number of edges ok.

Remember that v 1 is also connected to this edge sorry not this one. So, it is already connected to v 2 it will be connected to this one this one this one and so on. So, v 1 is also connected to you know k by 2 nodes in the right side ok and all those edges which are actually connected to v 1 the right side of v 1 will be counted while calculating the local clustering coefficient of node v; pause and ponder and think why it is so, ok.

So, how many edges we need to consider for node v 1? How many edges which are associated to v 1 will be considered in this calculation ok? So, we need to consider; we need to consider k by 2 minus 1 edges why k by 2 minus 1? Because v 1 is connected to k by 2 number of nodes in the right side of it right why minus 1? Because v 1 is also connected to v, v 1 is also connected to v, but that edge will not be considered while calculating the clustering coefficient therefore, k by 2 minus 1.

Similarly, for v 2 ok v 2 is v 2 is connected to k by 2 number of; k by 2 number of nodes in the right side. So, this node this node this node right, but we will not consider the connection which is which actually form I mean which actually connect v 2 and v. So, again k by 2 minus 1, k by 2 minus 1 right and so on and so forth. So, we will, so for each of the left side nodes of v, we will add k by 2 minus 1 number of edges. So, dot dot dot dot dot k by 2 minus 1 and how many times? k by 2 times ok right. So, what is the total number of edges? So, this is k by 2 into k by 2 minus 1.

Now, this is this calculation is done with respect to the left side of v right. Remember one thing we have already considered edges like this right for example, the immediate left side of v right for which we have also considered k by 2 minus 1. Therefore, it means that we have already considered these edges right why I am saying so, because when I when I look at the right side of v and look at how many edges we should consider, we should not consider these edges because this has already been considered ok.

So, this number of edges are basically considered with respect to the left part of v what about the right part? So, let us look at v 3. v 3 is the immediate right neighbour of v ok. So, if you think carefully v 3 is also connected to k by 2 nodes in the left side of it and all this k by 2 nodes have already been considered before while we while calculating this part. So, only v 3 is right part will be considered ok.

So, how many edges we will consider for v 3? So, v 3 is; v 3 is also; v 3 is also connected to k by 2 nodes in the right side right. But remember v 3 is last node in the right side v 3 is last node in the right side would be somewhere here which is not part of vs neighbour. Because v 3 has v 3 is located in the first position of the right side of v right and v 3 has k by 2 number of edges in the right side.

So, this edge will also be a will also be connected to v three, but this edge will not be considered while calculating the local clustering coefficient of v because this node is not a part of the neighbours of v right. So, how many edges will be considered in that case? k by 2 minus 1 for v 3 what about v 4? Same calculation k by 2 minus 2 right k by 2 minus 2 and think of the last node right or second last node not the last node second last node ok. In case of second last node only one edge will be considered this one.

So, dot dot dot 1 ok. So, and this is a series and we know that how to calculate it basically this is k by 2 into k by 2 minus 1 by 2 this is the sum. So, the total number of edges among the neighbours of v is this number 1 plus number 2 number 1 plus number 2. So, if we take the sum this is going to be; this is going to be you know 3 by 2 k by 2, k by 2 minus 1 by k by 2 k minus 1 you do the calculation you will get this number ok.

This and this will cancel out so, we will basically have you know 3 k minus 2 by 4 k minus 1 ok. So, if you take the sum of 1 plus 2 and if you also need to divide, sorry I just forgot to mention you also need to divide it by the total number of edge the that the total number of possible edges between total number of possible edges among the k neighbours right that is k into k minus 1 by 2.

So, if we take the sum of 1 plus 2 and you divide it by k into k minus 1 by 2 that is the clustering coefficient of v right and you will get this number 3 k minus 2 by 4 into k minus 1 ok. So, which is in the asymptotic label this is 3 by 4 proportional to 3 by 4. So, what does it indicate? It basically indicates that you know the clustering coefficient of a node local clustering coefficient of a node in a lattice regular lattice network is always 3 by 4.

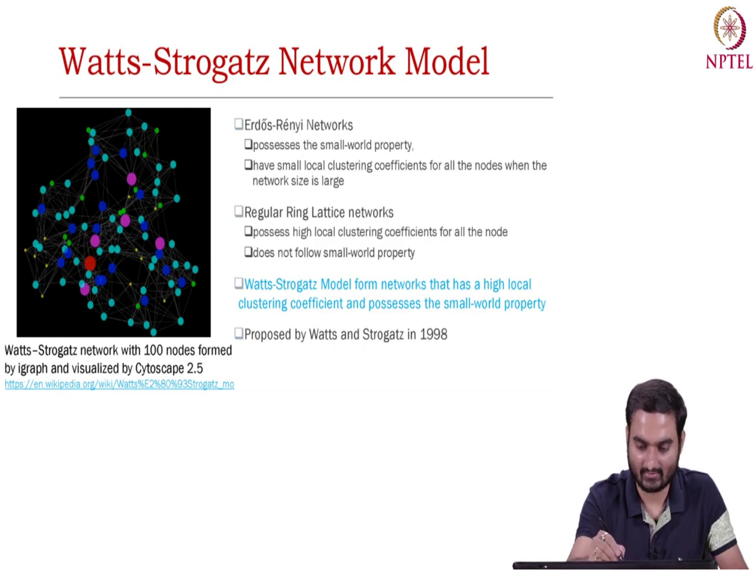
As the network grows number of node let the number of node increase does not matter the clustering coefficient will always remain same 3 by 4 right which is quite large. So, the clustering coefficient property that a synthetic model needs to satisfy ok. So, it is being satisfied by the regular lattice model, but the problem is there are multiple problem the first problem is the small world property will not be preserved right because the average path length would be quite high.

It is a regular graph each node is connected to only the neighbours. So, if you really want to jump from one node to another node, it would not be 6 degree for example, right. So, average path length would not be logarithm of l. So, that is the property which is not followed by the regular lattice ok. And; obviously, the other property is the a power law d distribution which will not be preserved at all because this is ultimately a regular graph.

So, nodes will have same degree therefore, the degree would be uniform right degree will be same. But nevertheless at least we understood that you know in the regular lattice kind of formulation of a graph the clustering coefficient can be preserved whereas, the ER model random graph model the small world property right will be preserved. So, let us try to come up with a model which in one spectrum behaves like a small world like a regular lattice and another spectrum behaves like a random graph ok.

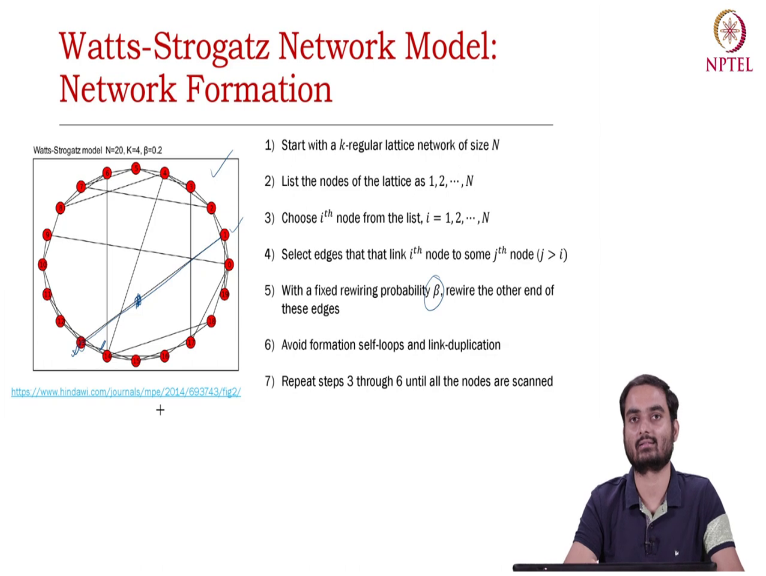
If you think of a spectrum right one in one side of it one extreme side of it you will have a regular lattice and you keep on changing and then you will get a random network right. So, the other extreme will preserve the random network property this extreme will preserve the regular lattice property and in between these two we will have you know we may find; we may find some sort of you know small world property ok or say scale free property ok. So, let us try to understand this.

(Refer Slide Time: 19:16)



So, Watts and Strogatz right this is called Watts Strogatz model. Watts and Strogatz model actually you know combines the idea of regular lattice and random graph models ok. So, what is your suggestion?

(Refer Slide Time: 19:34)



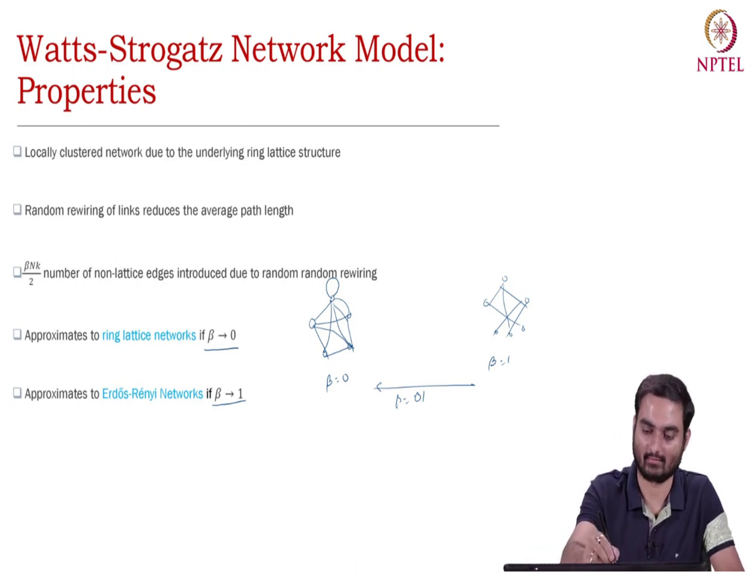
So, their suggestion is as follows. This they are saying that you know let us start with the with the regular lattice ok. Let us start with the regular lattice structure and then what you do you basically choose a node ok and one of its edges ok the node is connected to k other nodes in the graph and let us choose one of the one of the edges. And let us try to rewire it ok meaning that let us try to connect that edge to another node and how do you connect it?

With certain probability; with certain probability beta right you basically rewire it with another node with which the given node has not been connected ok. Meaning that let us say this node 13 ok. You take you take one of its edges for example, you take this edge right and let us and the ok forget about this let us think of this edge ok this edge 13 to 15 ok.

So, you choose this edge and then you rewire it how do you rewire it? You open the connection right you choose I mean one of the one of the remaining nodes present in the graph right again I uniformly at random and you then you connect it. For example, you choose node 1 and you then you connect 13 with 1 ok with certain probability beta ok and this beta is same as the probability p that we mentioned in case of ER model ok.

So, as you know keep on rewiring things the regular lattice structure will break and as you keep doing you actually tend to move towards more random network kind of structure because you are breaking the regular structure regular graph structure and you are making this thing random you are making the connections random ok.

(Refer Slide Time: 21:44)

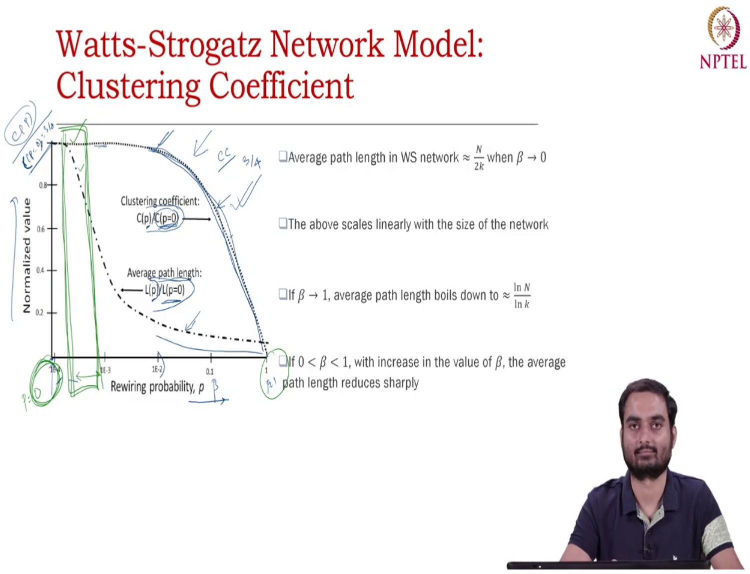


So, if these kind of things happen, then what would be the clustering coefficient? What would be the change of you know small world property and so on and so forth let us try to understand it ok. So, as you as beta tends to a 0 meaning the rewiring probability 0, it would basically follow a regular lattice structure because there is no rewiring case and if it if beta tends to 1 then it becomes more and more random ok. So, there are two spectrum right.

So, there are two spectrums of it see in one spectrum you will see that there is a regular lattice structure like this ok when beta equals to 0. When beta equals to 1 you will have right you will you may have a random network like this same number of edges. But remember one thing when we re-wire we will avoid two things we will avoid the case of self-loop and we will avoid the case of parallel edge right.

It would not happen that you connect this with itself it would not happen, it will also not happen that you connect you basically connect it. So, ok it will not happen that this kind of self-loop structure will form or it will not happen that in parallel edges will be formed ok. So, let us try to understand what happens in between these two. Beta equals to 0 say beta equals to 0.01 what will happen? How the graph looks like ok we will analyse this thing now.

(Refer Slide Time: 23:51)



So, this is the empirical results ok. Let us try to understand this curve this is very important curve. So, the x axis is the rewiring probability beta or p whatever right and the y axis corresponds to two things you see here there are two quantities that we are plotting, one quantity is the clustering coefficient this one and the other quantity is the average path length ok.

So, what we are try to understand is that, what is the clustering coefficient let us let us look at this curve first ok. What is the clustering coefficient of a node at this point when p equals to 0 or beta equals to 0? And, how this clustering coefficient will vary? And, what would be the clustering coefficient when p equals to 1 beta equals to 1?

How the clustering coefficient will change right? And what exact quantity we are plotting here? We are plotting a ratio which is the ratio of the clustering coefficient of the network at a certain value of p and the clustering coefficient when p equals to 0 because when p equals to 0 we know that what is the clustering coefficient. When p equals to 0 meaning that there is no rewiring right p equals to 0 means you have a lattice structure and we know the clustering coefficient is 3 by 4 ok.

So, at this position CCP goes to 0 is 3 by 4 right. And as you change p as you increase p CP value will change and we measure this ratio and we will see whether this ratio increases or decreases. Now, think about it. As you increase p you are adding more randomness ok you are adding more randomness. So, what would happen? The clustering coefficient will decrease because you are now randomly you are doing random rewirings.

So, the connections among your neighbours will decrease. So, the clustering coefficient will decrease. So, the numerator will decrease denominator will remain same the numerator will decrease. So, therefore, the ratio will also decrease. As you see here the ratio is decreasing with the increase in p ok. If p is 1 meaning the network is completely random, we know that the clustering coefficient would be very very less right.

Because in the earlier in model we have seen that the clustering coefficient is inversely proportional to 1 by N and of course, I mean although here N is fixed, but when you make this thing completely random there is no clustering coefficient at all ok. So, this is the change of the clustering coefficient ok. Now, let us look at the change of the average path length and if you do this.

So, and remember this is a simulation results meaning that you have certain you start with the regular lattice and you keep on changing and you are getting you keep on getting different networks and you are basically measuring these things. So, this is empirical results simulation results ok and we are trying to explain the simulation result here.

So, average path length behaves like this, what does it mean? Let us look at the average path length at p equals to 0. So, let us assume that you have certain average path length right. So, as you increase p what you are doing? You are basically letting a node connect with a node which is far from the given node. Because currently in the regular lattice structure a node is connected to only you know the load node is connected to your I mean its immediate right side neighbour or the second of neighbour or so, on immediate right side immediate left side neighbour.

So, as you increase p, you are actually letting the node connect with the other nodes which are far from the; far from the given node right so; obviously, the average path length will also decrease think why ok. So, the average path length will also decrease therefore, this ratio will also decrease. So, you see a decreasing curve like this ok. Both the curves are decreasing, but the way these two curves are decreasing these two quantities are decreasing they are different right.

So, clustering coefficient is decreasing slowly whereas, the average path length is decreasing suddenly. So, in this region you will see that say in this region of p right. If we focus on this part ok or say if you just focus on not this part let us focus on slightly left part of the of it let us focus on this part for example, ok. What you will see? In this particular region you will see that the clustering coefficient is quite high and the average path length is also quite less ok the average path length is quite less.

It basically says that, so this part when p equals to 0 this indicates a regular lattice when p equals to 1 indicates the random graph right, but this part is something which follows two properties of real network high average clustering coefficient and low average path length. So, possibly this part follows the small world property ok. So, this was the proposition of Watts Strogatz. So, you can actually generate a spectrum of graphs using Watts Strogatz method by varying you know by varying the size of by varying the values of p ok.

So, that is ok, but what about the you know this power law distribution? Scale free property is not preserved till now. So, we have understood how to generate a network with high clustering coefficient. We have understood how to network how to generate a network with low average path length right. But we do not know how to make this thing you know I mean how to make a network which would eventually follow the power law degree distribution.

So, in the next lecture we will discuss one such model which is called a Barabasi Albert model which is the first model that allows a network follow the power law degree distribution ok. So, we stop here today.

Thanks.