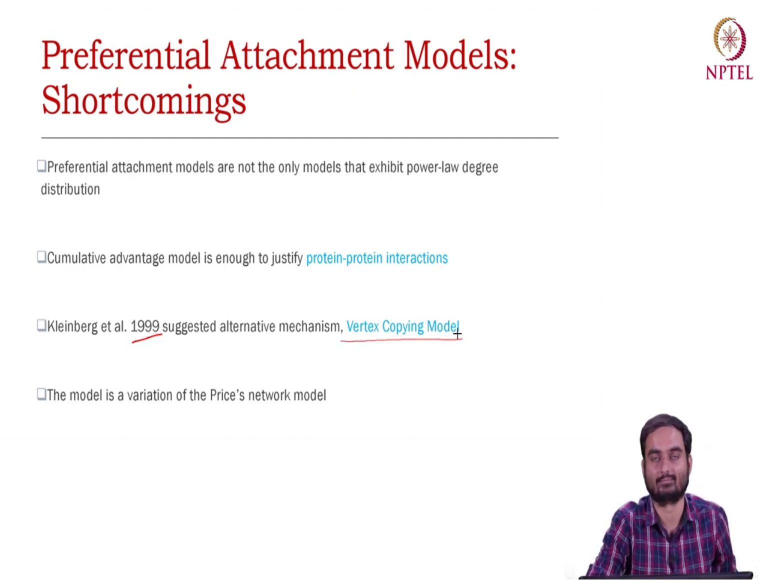
Social Network Analysis
Prof. Tanmoy Chakraborty
Department of Computer Science and Engineering
Indraprastha Institute of Information Technology, Delhi

Chapter - 03
Lecture - 07

So, in the last few lectures we have been discussing about models, and we have discussed many models; random graph model right, Watts-Strogatz model, ring lattice model, then Barabasi-Albert model, then price model right. And Barabasi-Albert models, price model, these models are sophisticated in the sense like they are able to produce networks with power-law d distribution right.

But, if you think of the mechanisms, they basically follow preferential attachment policy right, or cumulative advantage policy.

(Refer Slide Time: 00:58)

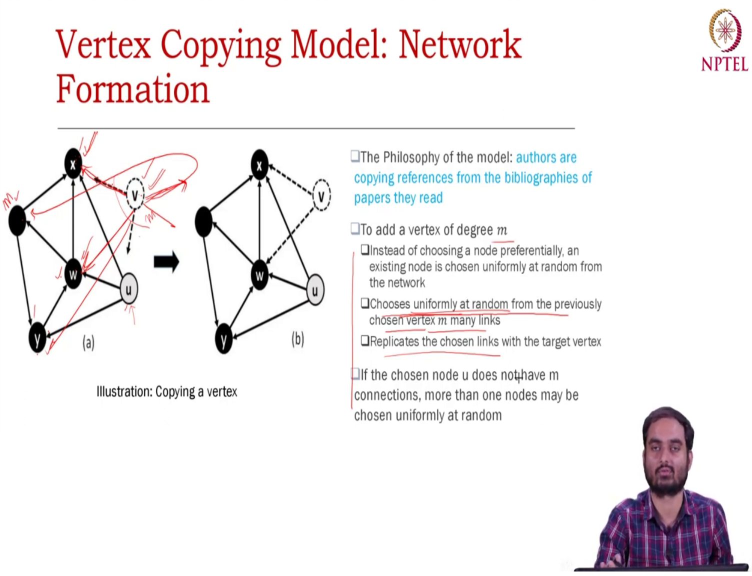


So, cumulative advantage policy basically says that rich node will get richer over time. But, if you think of and this you know rich get richer rich gets richer phenomena this is applicable to social networks right. This is applicable to citation networks, but this may not be applicable to biological networks like protein-protein interaction networks; because there is no such compulsion that a protein will interact to another protein right with high degree ok.

So, therefore, although you explain you know power-law degree distribution you may not be able to realistically explain a biological network, a growth of a biological network which also follows power-law right, using Barabasi-Albert model. Therefore, people started thinking of you know different other mechanisms by which you can also generate a power-law a scale free network, but not following the preferential attachment policy ok.

So, in this particular direction Jon Kleinberg right, he is a; he is a computer scientist in Cornell. So, Jon Kleinberg in 1999, who is also the who is also the inventor of heats algorithm, we discussed in the last chapter ok. So, he proposed this model called Vertex Copying Model ok.

(Refer Slide Time: 02:24)



Now, what is Vertex Copying Model? Vertex Copying Model is you know very unique idea. The idea is that when a new node joins, so, this new node v joins right, and let us assume that again v has n number of out degree, outward edges right. So, when v joins, v will not get attached with other nodes based on the preferentially attachment ok. Instead, what we will do, we will you know we will choose a node an existing node uniformly at random right, and starts copying that nodes edge distribution ok.

Let us say v has randomly chosen u, as the node which v wants to copy ok. So, say u is connected to x, u is connected to w, u is connected to y right. So, v what then what we will do? We will we will also select x w and y ok. But, v has only 2 nodes. So, among these 3 nodes x, w, and y; we will choose 2 nodes. So, let us say we will choose x and w, and v will be connected to x and v will be connected to w.

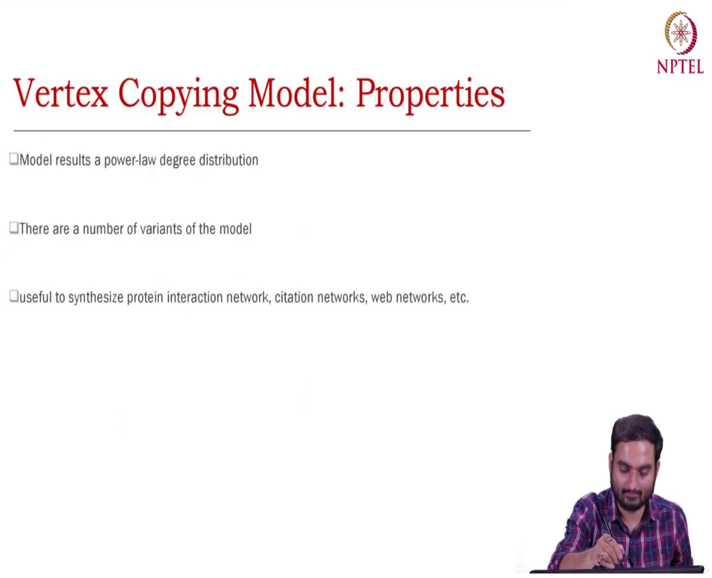
So, what happens, what has happened? It v basically has copied u's edge the edge distribution that u follows ok. Now, you may argue that ok, what happens if v has say degree out degree 4. So, v would first choose u and v will be connected to x w and y, and there is one remaining edge; what we will do, we will choose another node from the remaining node set. Again uniformly at random right, and starts copying one of its edge.

For example, let us say; let us say you know v has chosen w ok. And w is connected to w is connected to x, y and this node. So, this is m let us say. Now, v is already connected to the x and y. So, v will not be connected to x and y now, then v will be connected to m. So, this node, this edge will be connected to m ok.

So, what is the idea here? So, you basically choose uniformly, when a new node joins with degree m, you choose you basically choose uniformly at random from the previous from the previously chosen vertex m right. Replicate those links and if you are exhausted with the links, if some edges are still there you keep on repeating the same thing again and again ok.

So, this is called Vertex Copying Model, I am not going to the theoretical details of it, but if you look at the theory, it also proves that this model would eventually you know generate a network which follows power-law; but this is a non preferential attachment based model ok.

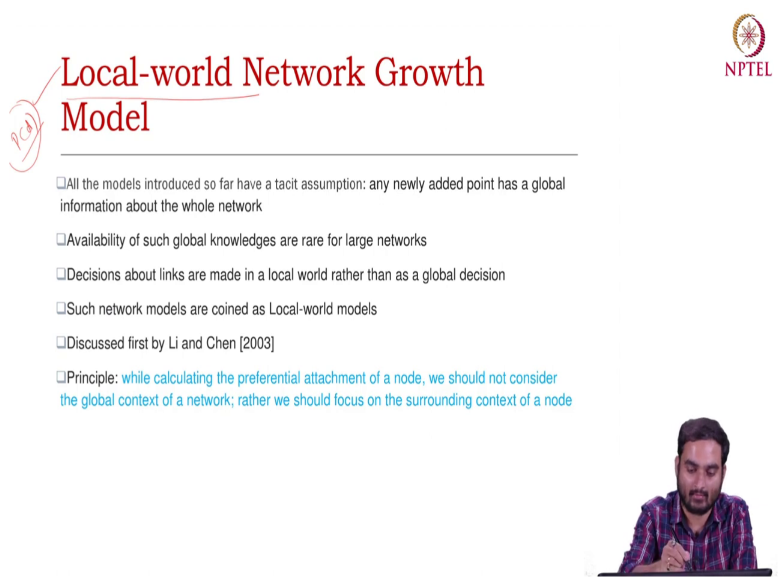
(Refer Slide Time: 05:40)



(Refer Slide Time: 05:43)

So, now let us look at some other models right. So, we in the preferential attachment models we discussed the other limitations. For example, there is there is no notion of competition, there is no notion of you know decay of prestige over time and so on and so forth.

(Refer Slide Time: 06:01)



So, later on you know several models have been proposed, one of such models is this Local-world network growth model. In the in this particular model, the intuition is that you know that when you compute preferential attachment, when you compute p d right, which is d divided by the total sum right. The particular quantity p d, it should not be based on the inter network, it should not be a global property. So, in Barabasi-Albert model, price model, we assume that we calculated p d based on the inter network ok.

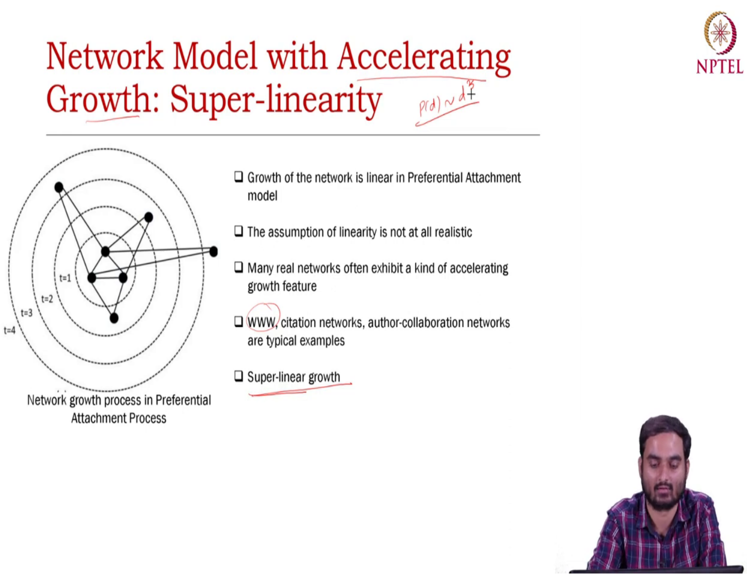
But, you know many times the inter networks are not available before and the second you know reason is that say for example, in case of citation network right. In citation network when a paper you know in the field of say AI right, joins the system right. It is less likely that the node will cite a paper from say database or the paper from a theoretical algorithm kind of paper.

So, what the paper would do? The paper would essentially look at the other papers which are published in the AI domain right, and based on the preferential attachment probability with respect to that particular domain the new node will choose another node.

What I mean to say is that when I write a paper in the area of AI right, I will only look at those highly cited papers in the area of AI and then I cite. It is not that I will look at you know I will look at the citations with respect to all the areas right.

So, the idea was simple here, the idea was that this p d should be calculated based on the local property, not based on the global property ok.

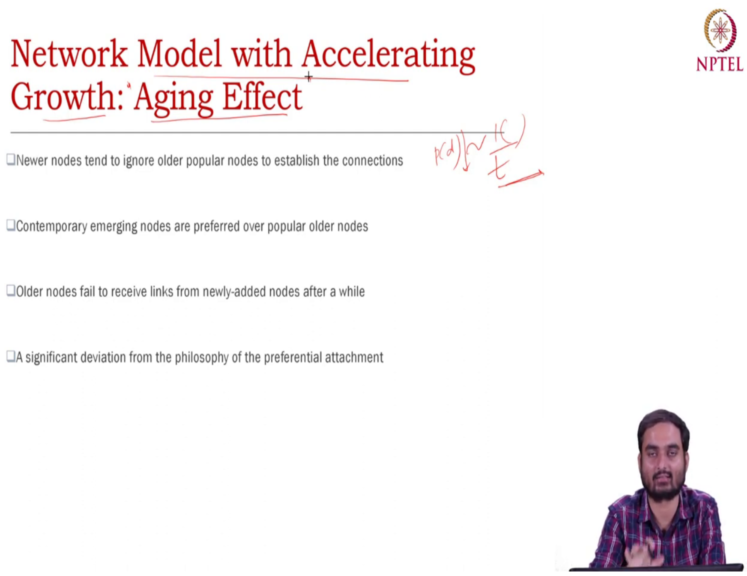
(Refer Slide Time: 07:48)



Another model is called the network model with accelerating growth ok. It basically says that so far we assume that the p d, the probability of a node v degree d is proportional to d which is linear right; but it may not be linear, it may not it may be super linear right. It may be you know p d is proportional to d to the power you know say d square or d cube right.

So, in fact this World Wide Web citation networks right, if you look at the actual growth they actually follow super linear growth model. So, linear growth model may not be enough ok. So, people then try to come up with super linear. In fact, sublinear growth model also, hybrid model also right.

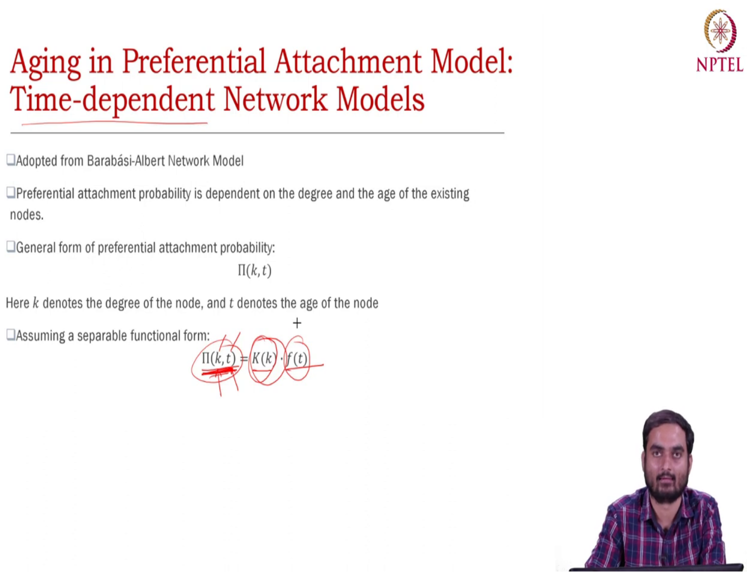
(Refer Slide Time: 08:41)



There is also another interesting idea of aging effect ok, recency effect. What is aging effect? The idea is that you know as the node becomes older and older, it starts reducing its importance. And I mentioned in case of say actor-actor network right, in general actor-actress when you know they become older, it is less I mean the their popularity will also reduce in general ok.

So, this aging effect can be quantified separately into the preferential attachment model right. So, this when you calculate p d there should be some you know some quantity which should be inversely proportional to t right, as t grows as t increases p d should decrease. So, this is called aging effect and one needs to consider this aging effect explicitly into the growth model.

(Refer Slide Time: 09:46)

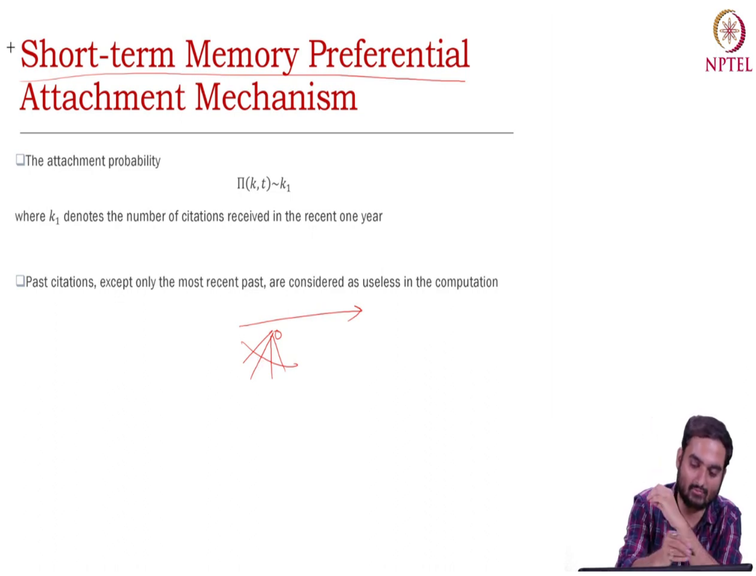


In fact, there are also models which in fact, in the aging model or what people suggested is that let us make it time dependent; because the factor that I mentioned this is time dependent ok.

So, then that is total attachment probability right, of a node with degree k should also be dependent on t. So, this function is essentially a function of both k and t. And this function with a which is a biparametric function, this can be split into two other functions which is a function of k times the function of t.

Now, you define f and t you know based on your applications or based on your assumptions ok. This is aging model.

(Refer Slide Time: 10:32)



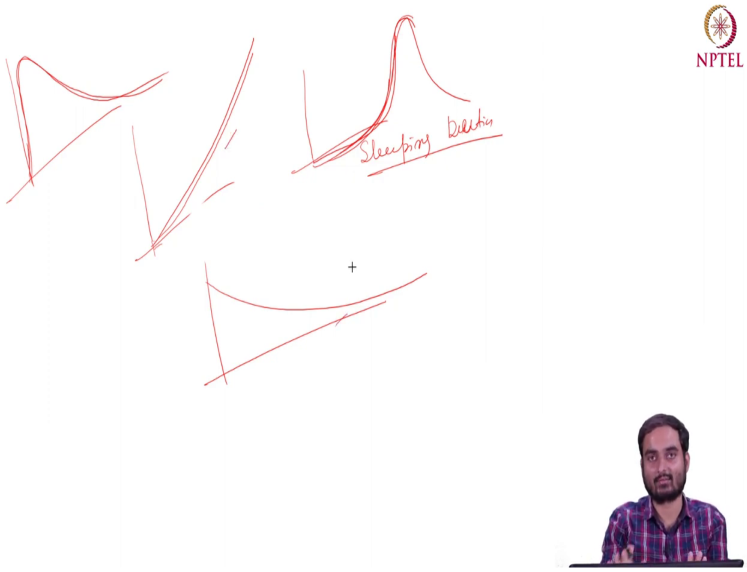
So, there is another idea called short term memory preferential attachment. So, here you know the idea of memory actually comes into the picture. So, the idea of memory is that. So, it basically says that when you calculate preferential attachment probability, you may want to forget about older citations or older ages right.

What it basically says is that, as the time you know moves ok, say a node initially had got a lot of edges right. Those edges should not be considered at the current time period, which we should ignore you know the prestige when the prestige which this node had received long time back.

Now, this factor is important, because this factor now enables the notion of competition right. So, once you start forgetting about the old prestige you basically allow the new nodes right, to kind of compete with the old nodes. Because old nodes now do not have that much prestige right, because some of the prestige has now gone.

So, the notion of competition will come and the new nodes will also have advantages to acquire edges to increase degree over time ok. So, this was also proposed. So, there are many other you know network growth models, which I have not mentioned here you know people looked at not only the power-law distribution right, but also other factors. For example, in one of our research in the lab, we showed that you know in case of citation network.

(Refer Slide Time: 12:30)



If you look at the if you look at the fine grained citation distribution of nodes right. If you think of difference different types of papers, and their citation growth you essentially think you can think of you know 4 - 5 patterns for example, there are papers which starts receiving citations immediately after publications you can think of this kind of growth and then decrease over time.

There are papers whose citation increases over time papers like transformer ok. There are papers which do not receive citations at the initial stage, but later it gives it gets you know popularity. So, this kind of papers are called sleeping beauties ok, sleeping beauties.

So, sleeping beauties are those scientific papers which do not receive enough citation in the initial stage of their career right; their timeline first 5 6 7 years, but after certain time people start realizing the importance of those papers and the importance grows over time ok.

So, this is basically increase and then decrease. This is continuously increasing you can have also patterns like which decays over time right, which in this pattern you see that after certain point in time you the citation starts increasing then decreasing and so on and so forth ok. Monotonically increasing, monotonically decreasing.

So, there are essentially 6 5 6 patterns ok. And the citation network growth model and in general if you combine everything, you will of course, get a this kind of distribution power-law distribution; but when you magnify this one right you will get 5-6 types of distributions right.

Interestingly none of the growth models including price model, Barbasi-Albert model would be able to mimic the 6 different patterns. So, we showed that ok, let us try to come up with a model which includes the you know aging factor and other factors as well to mimic you know this 6 different patterns. But, there are many such actually you know research happened in the past which you know basically suggested us how to generate synthetic networks.

So, we stop here with the with the takeaway that these kind of models would be very handy, would be very useful when you know you want to test your own algorithm on small networks, and these models will essentially you know give you flexibility to generate networks of different sizes, and you test your algorithms on these networks and you freeze one model one algorithm, which seems to be you know effective compared to other models and then you move that algorithm to the production level ok. Therefore, this chapter is very important with this I stop here.

Thank you.