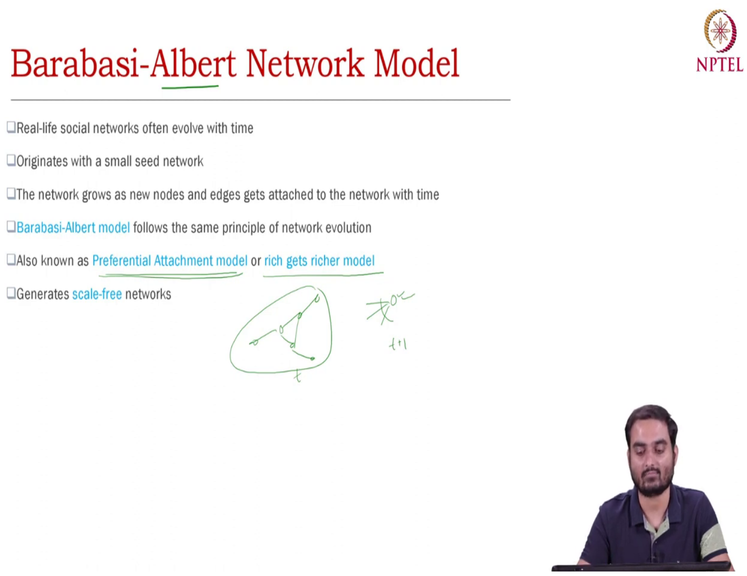
Social Network Analysis
Prof. Tanmoy Chakraborty
Department of Computer Science and Engineering
Indraprastha Institute of Information Technology, Delhi

Chapter - 03
Lecture - 04

So, we have discussed Random Graph Model, we have discussed Regular Ring Lattice Model, we have also discussed watts Strogatz Model, right. And the property that we have not seen so far is the power law degree distribution, right. So, today, we will discuss the new model which is perhaps the one of the one of the similar models, seminal models in the area of Network Growth.

(Refer Slide Time: 00:52)



And this is this was proposed by Albert Barabasi and recall I mean and you know, so (Refer Time: 00:57) Albert. And this is basically called Barabasi-Albert model or BA model, ok.

So, BA model is the first model which explains the power law degree distribution of a network, ok. And this is more realistic. Reason is that if you think of the other model that we have discussed so far, ok, all those models assume that the number of nodes is fixed, ok and number of nodes are given beforehand, you are basically designing mechanisms to rewire or to connect nodes in different ways, right.

Random, in case of random graph model we connect nodes randomly, in case of watts Strogatz model we basically rewire the regular lattice, right in such a way that after certain rewiring we will get a you know a small world network and so on. Barabasi-Albert model is a model which basically says that let the node come one by one, let the node join the system one by one with certain edges. And let the node decide you know which node, which other nodes it wants to get connected, ok.

So, this is also called preferential attachment model. Why it is called preferential attachment model? Because when a node comes in, right let us say, let us say you already have this kind of network at time t, ok. At time t plus 1, a new node comes in with certain edges, right. So, how does this node choose an existing nodes? How does this node choose other existing nodes for the connections? Right.

So, the node will choose based on the degree based on the current degree of the other nodes. So, it means that the other nodes which are already there in the system, the other nodes will acquire a new edge, the probability of acquiring a new edge is proportional to the degree to its current degree, ok.

So, let us say there are two nodes, node 1 has degree 10 and node 2 has degree 1. So, the likelihood that node 1 will receive an edge is higher than node 2 receiving an edge. This is proportional to the degree of nodes at current point in time.

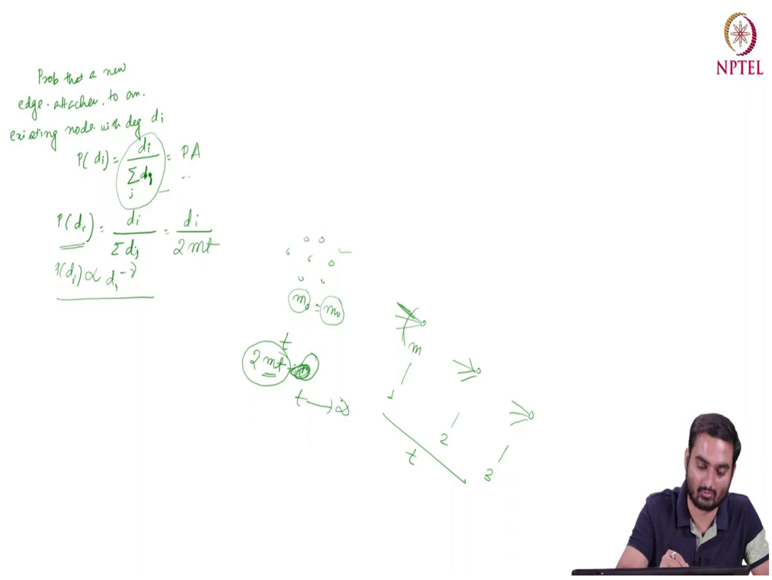
So, this is, based on the preference with respect to the degree of nodes in a network, right. So, therefore, this is called preferential attachment model. This is also called rich gets richer model. The reason is simple. You are basically allowing yourself to be connected to a node which is already rich, right rich in terms of degree, right. The node will the node has already higher degree, and you know you basically give preference to those nodes which have already got higher degree. This is also called rich gets richer model, ok.

We will see that this rich gets richer model will allow you to generate a network which follows power law, scale free network. And this is also called scale free network by the way, this, Barabasi-Albert model is also called scale free model, scale free network, right.

We will also see that this will allow you to generate hubs, right or outliers. Now, real world networks you know real world networks always exhibit this kind of outliers or hub nodes, right with a lot of degree with very high degree, ok. We will see that using this models you will also get nodes with higher degrees, ok.

So, let us get started. Let me let me not use my slides let me use the whiteboard and derive things, I mean gradually, slowly. So, this is little bit mathematically heavy. So, you just need to follow whatever I am explaining here, right.

(Refer Slide Time: 05:25)



So, the probability that a new node, the probability that a new edge, right attaches to an existing node with degree d i is P of d i which is as I mentioned this is based on the current degree this is d i and we normalize because this is probability we normalize it by we normalize it by j, right, so by all the degrees, ok. And this is called preferential attachment.

So, let us try to understand the system first. So, what is happening? So, initially let us assume that you have m number, you have m 0 number of nodes in the system because you have to start from some nodes, right. So, let us assume that you have m 0 number of nodes, and you also have roughly you know you have roughly m 0 number of edges, ok; roughly m 0.

Number of nodes and number of edges is more or less same, it does not matter, right. You can assume that there is no edge. But if you assume there is no edge, then it would be difficult to calculate this probability because nodes will have degree 0. So, let us assume that on an average a node is connected to only one other node in the system. So, roughly m 0 number of nodes; roughly m 0 number of edges. So, you start with this setting, ok.

Now, a new node comes in with degree m. So, this is a drawbacks, but let us assume that each node is coming with the same degree, the degree is same which is m, ok. Now, each of this m edges will choose an existing node based on this preferential attachment, right.

So, this edge will choose a node based on the preferential attachment. Remember one thing, this is a probability. Meaning that it is not always the case that always you will choose a node with high degree, right. With high probability you choose a node with high degree, but also there are there is non-negligible probability based, I mean by which you basically choose a node with low degree. Otherwise low degree will never grow, right. So, this is a probability, ok. It is a probabilistic model. So, this is the setting.

So, we basically need to we need to determine when these nodes are coming in. And remember at every time only one node is coming in, only one node joins. So, at time 1, one node joins; at time 2, another node with m edges joins; at time 3, another node with m edges joins and so on and so forth, right.

So, we will see that over the time t, t many nodes will join, right; m into t m into t number of edges will join, right. We will see how you know joining of this nodes will change the degree, average degree. We will see how the joining of this nodes will change the degree distribution, ok.

So, this P of d i is d i by sum of degree, right. Now, sum of degree is what? Sum of degree is two into number of edges. So, how many edges are there? So, we have, so at time t, mt number of edges will be join, mt number of edges have already joined, right.

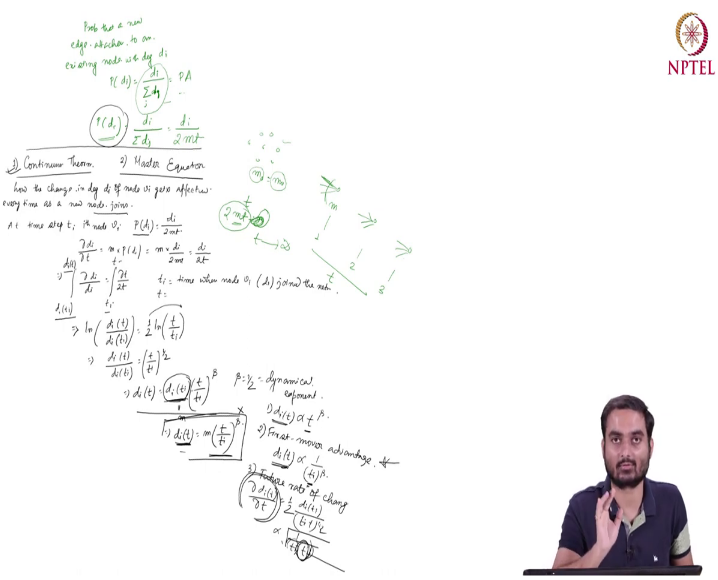
And if you just ignore the last node because at time t you are calculating the number of edges, right you are you will not consider the node; let us assume that you are not considering the node which is joining at time t.

So, how many number of edges are there? You; I mean you can roughly say that, ok this is 2 mt you know minus m, it does not matter because this m is, and as the t as t tends to infinity you will say that this this does not matter. So, you will have only 2 mt. So, 2 into number of edges. So, this would be 2 mt, right. So, this preferential attachment is now d i by 2 mt, right.

So, now we will see, now this is degree, this is the you know a probability of a particular degree d i. We will see that how this degree changes over time and how we can see that this P of d is actually you know proportional to d i, P of d i is proportional to d i to the power some minus gamma, right. Because this is power law. We have to prove this, ok. We will prove this thing, ok.

In order to prove this, we will use a theorem. So, this degree dynamics is generally measured in two ways.

(Refer Slide Time: 11:00)



So, there are two ways to measure it, one is called continuum theorem or continuum approach and the other is called master equation, ok. So, using these two approaches, we will prove separately that the d-distribution follows power law, ok. So, let us start with continuum theorem, continuum approach, ok. So, we basically show that how the change in degree d i; how the change in degree d i of node v i gets affected, every time as a new node joins. So, this is something that we are trying to understand, ok.

So, the new edge, right or the new link, right that gets connected to a node, right is proportional to this one. So, at time step t, right we know that for the ith node v i the probability of getting connection is P d i which is d i by 2 mt, right.

So, what is the rate of change of degree? So, the rate of change of degree is delta d i d t, rate of change of degree d i, think about it rate of change of degree d i when m edges are coming in. So, this is basically m times P d i because m number of edges are coming in, right. So, and each edge will join a node with degree d i is P d i.

So, what is the change of degree d i? Is m times P of di, right. This is basically m times d i 2 mt which is d i by 2 t, ok. So, what we do? We basically take the integration. So, d or del d i d i del t 2 t, right, we will take the integral because ultimately our task is to understand d i t, right. So, we take the integral from d i t i to d i t, right. I will I am explaining what is this what are these notations t i t and t, right.

So, t i is the time when node d node v i, node v i, right when node v i whose degree is d i joined the network, ok and t is something which is a current time. So, these two things are clear.

What is d i t i? d i t i. What is d i t i? d i t i is the degree of is a degree d i at time t i, right. And what is d i t? d i t is a degree d i at time t, right. So, if we do this integral, you will get log of d i t i, d i t i half log of t by t i, ok. And if you know do some modifications, you move half here and you know remove log in both the sides. So, you will get d i t by d i t i equals to t by t i to the power sorry half, ok which is basically d i t, right equals to d i t i times t by t i to the power beta.

And what is beta? Beta is half in our case, right. And beta is called dynamical exponent. So, this is the equation, right. Degree at degree of v i at time t is, degree at time t i degree of v i at time t i times this one, ok.

So, let us try to understand what is this. Degree of a node v i at time t i. So, what is time t i? Time t i is the time when node v i joined and when node v i joined, v i joined with m edges, the degree was m. So, this is m, ok. So, we have d i t is m times t by t i to the power beta, right. This is very important equation. This is a very very important equation. Try to understand.

So, there are 3 observations that we can draw. Number 1, is that d i t is proportional to t to the power beta, right. This is also power law which basically says that the growth, right that the degree growth of node v i is proportional to the time, right.

Number 2, is called first mover advantage. What is first mover advantage? So, you see that d i t is inversely proportional to 1 by t i to the power beta. What is t i? t i is the joining of node v i. So, it means that as t i, if t i is low d i t would be high. Meaning that if the node joined earlier t i would be low, and degree d i would be higher, if the node joined earlier it is likely that the node will gain a lot of degree.

So, the node which actually joined earlier which join first will have advantage in terms of degree in terms of degree growth. So, therefore, this is called first mover advantage, ok. This first mover advantage is very important in the context of research, right.

I mean there are papers actually you know which are not that scientifically concrete, right, scientifically fundamental, but those papers actually started a particular field or a of or a particular research area. Simple idea, but since those papers started a research area they actually got a lot of citations, ok.

The third one is called is future you know future rate of change, right. We basically try to understand that the rate of change of this degree, so this is degree at time t. And how this rate will change? So, we need to take the you know derivative.

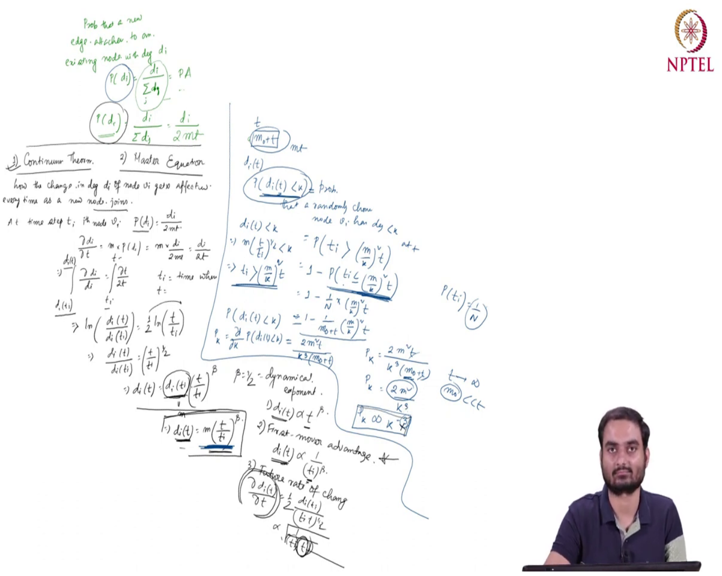
So, we will do we take the derivative with respect to time. So, del d t del t is if we take the derivative here, assuming beta equals to half; so, we will have half, right and m is d i t i, right. So, half d i t i, right by t i t to the power half, ok which is basically proportional to 1 by t i times t to the power half or root over, square root of this one.

So, this is the rate of change. What does it mean? So, it means that since the rate of acquiring new nodes is inversely proportional a rate of, acquiring new nodes meaning rate of change of degree is inversely proportional to the current time stamp the rate of acquiring of new edges by a node slows down with the time. So, as time as time increases this will slow down, ok.

So, the rate with which a node you know is supposed to gain edges is much much higher, at the initial level and as the time increases it will slow down, ok. It is inversely proportional to t, it is also inversely proportional to t i, ok.

So, these are 3 observations that one can draw. But the important thing to note is this equation, ok. We are not yet done. We have to prove that the degree distribution follows power law, right. So, let me erase this part, ok. Let me erase this part and let me use this region of my notepad, right, alright. So, now, what we will derive? We will we will basically; so, let me just you know recap some of the things.

(Refer Slide Time: 22:24)



So, at every time stamp t at every time stamp t, there are at a particular time stamp t how many nodes are there? There are m 0 is the number of nodes at the beginning, plus t because at every time you have one edge, one node. So, m 0 plus t number of nodes are there. And how many edges? mt number of edges are there, right. Of course, mt plus m 0, but you can ignore m 0, ok.

Also we need to remember that d i t is what? This is the degree of node v i at time t, ok. So, let us denote a probability P which is this one. So, it basically says that the probability that a random randomly chosen node v i has degree less than k at t, this is important, right.

So, if you just think of this quantity this is d i t less than k, right and we already know that d i t is this one. So, this is basically I replace d i t by m times t by t i to the power half less than k. If you do the math, you will get t i greater than write m by k square times t, ok. So, I can replace this by this. So, this is same as, this probability is same as P t i greater than m by k square t, right. This is same as 1 minus P of t i less than equals to m by k square times t, ok.

Now, what is this one? This is basically saying that the what is the probability that the time is less than equals to this one. Now, time t i is same as number of nodes joining in, right. So, at every time, every time we have, now if you think of this P t i. What is P t i? Probability of a particular time t i. And this is uniform; always 1 by N. If N is the total number of nodes that we can expect at the end, right because at every time we have one node joining in. So, this is uniform distribution, right.

And this is cumulative, right. This is basically cumulative. So, at every time this many, so you have you want to measure the total till this part. So, this would be 1 minus 1 by N times m by k square into t, right. What is N? N is the total number of nodes. And how many total number of nodes are there? We already know the total number of nodes is m 0 plus t, right. I replace N by this one, ok.

So, this probability P of d i t less than k is this one, right. And this is cumulative. So, how do we get the probability density? We take the derivative. This is cumulative. If you want to get the non-cumulative, we need to take the derivative. So, then P of k, right is essentially d k of P of d i t less than k, right. Meaning we take the derivative of this with respect to k. If we do that we will get 2 m square t by k cube, m 0 plus t, ok. So, we what we get? We get P k which is something that we wanted to measure P d i, right. This is 2 m square by t by k cube m 0 plus t.

If t tends to infinity, what would happen? And if t tends to infinity, m 0 can be negligible, right. This will vanish. This t and this t will vanish. So, we will have P k 2 m square by k 3, k 2 the power 3. What is m? m is the degree of a node which is, I mean degree of a newly joined node this is fixed. So, m is fixed. So, this part is constant. So, we will have P k proportional to k to the power minus 3, right.

So, the degree the probability of a degree degree k is k to the power minus 3 which is basically power law, where the coefficient also ranges between 2 to 3, right. So, so this is the master equation. This is the sorry continuum approach. Using continuum approach you can prove that the d-distribution follows power law.

So, in the next lecture, we will continue with this and we will also show the same thing, the same conclusion by using master equation. Now, why I am explaining these two approaches separately? Because say you want to propose your own model, using this continuum approach and master equation, you can also show that your model also follows the power law property, ok. So, in the next lecture, we will discuss master equation.

Thanks.