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

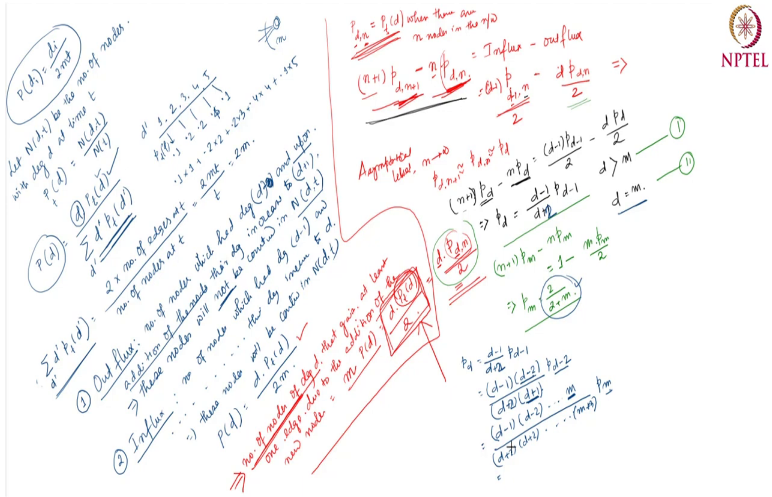
Chapter - 03
Lecture - 05

So, in the last lecture we have started discussing Barabasi – Albert model, BA model and we basically try to understand whether it I mean BA model follows you know power law D distribution. And, we mentioned that there are two ways to show that a model actually follows I mean whether a model follows power law or not. And, this two ways – one is the continuum approach and the second way is the is the master equation.

And in the continuum approach we have showed in the last day that you know that the d distribution for the Barabasi – Albert model basically follows power law. So, we derived this equation if you remember p k is proportional to k to the power minus 3, right. So, today we will you know discuss the master equation right the master approach and we will show that we will show basically the same thing that the power law d distribution is followed by Barabasi – Albert model. But the way we approach this thing it is little bit different from the continuum approach that we discussed in last time ok.

So, let us get started I am not using any slide I am just you know deriving the equations slowly so that you understand each and every equation you know properly.

(Refer Slide Time: 01:51)



So, in the last class we have seen that you know this P d, right P of d i the probability of you know of a node v i having degree d i is basically d i by 2 mt, right. So, this equation we used in the last lecture right, but here we will use little bit different ways right to basically you know denote the same thing. So, let us define a quantity.

Let us define let N d comma t right be the number of nodes with degree d at time t. I hope you remembered the Barabasi – Albert you know the mechanism that it follows at every time stamp t a particular node is coming adjoining the network with a fixed degree m, ok. And, the and each of these m edges is going to be connected with one of the existing edges based on the probability based on the preferential attachment.

And, this preferential attachment is basically proportional to the current degree of a particular node basically indicates that if a node has higher degree it is highly likely that the node will you know attract a new edge right compared to the node having say a lower degree, ok. So, let N d you know N d comma t be the number of nodes having degree d at time t.

Therefore, we define this quantity probability of you know at time t probability of a particular you know node with degree d right probability of a particular degree d is N d comma t by N t. What is N t? N t is the number of nodes at time t and N d t is the number of nodes with degree d at time t ok. So, in the continuum approach we use this equation, right. Can we write the same thing in a different ways? Can we write the same thing probability of you know of a degree d can we write in this ways d times P t d divided by sum of all d right let us write d dash ok p t d dash, think about it.

So, this is the degree and this is the probability of that particular degree at particular time ok. And, and the denominator is sum of all the degrees, but proportionately, ok. So, the denominator is kind of weighted average right say you know you have five degrees, right 1, 2, 3, 4, 5 for example, right. So, your d dash varies from 1 to 5 and you have P t d dash right say for degree 1 this is 0.1, this is 0.2 0.2 this is 0.5, this is a 0.4 and this is 0.1.

So, the sum would be the denominator would be what? 0.1 times 1 plus 0.2 times 2 plus 0.2 times 3 plus 0.4 times 4 plus 0.1 times 5, this is basically you know weighted average, ok. So, the denominator is essentially you know the average degree of a node average degree of a node in the particular network ok. So, let us you know let us just focus on the denominator ok. So, d dash d dash.

So, d dash is all the d dash indicates d dash is basically the index indicating the possible degrees that you have in the particular network p t d dash, ok. This is basically you know average degree ok. So, average degree is what? So, sum of degree is 2 into number of edges. So, 2 into number of edges and since this is average, right. So, sum of degree is essentially average degree into number of nodes right number of nodes.

I am basically moving this thing in the denominator. So, then this is going to be this is going to be the number of nodes right number of nodes at time t; at time t, number of edges at time t right. What is the number of edges at time t? So, at every time m number of edges are coming in. So, number of edges at time t is m times t. So, 2 mt and what is the number of nodes? Number of nodes is basically t because at every timestamp there is only one node coming in. So, this is 2m, ok. So, the denominator is 2m.

Now, let us define two quantities, right – the first one is called out flux, ok. What is out flux? Out flux is the number of number of number of nodes, right which had degree d minus 1, right and upon you know say addition of the node. So, we are interested in the situation when a particular node is added, ok. So, when a particular node is added the degree will change right the degree of other nodes will change.

Now, we are interested in only those nodes whose degree is d ok. So, when a particular node is added to the system what would happen? So, there are nodes which already have degree d, right and due to the addition of this node it may happen that some of the nodes will basically acquire edges out of this m edges at least one edge will be acquired. And, those nodes whose degree whose current degree is d, right due to the I mean due to the addition of one additional edge the degree will basically become d plus 1.

So, those nodes will not be counted when we calculate you know the number of nodes with degree d at time t because those nodes will have degree d plus 1. So, we call those nodes as the out flux, ok. They are actually you know they are actually moving out of the system and the system is a system that we are considering at current time stamp, ok.

So, number of nodes which whose who which are degree d which has degree d sorry it is it is d not d plus 1 which has degree d and upon addition of the node their degree increases to d plus 1, ok. Therefore, these nodes these nodes will not be counted; not be counted in n of d comma t, ok. So, this is out flux. Now, similarly we have influx.

So, influx constitutes those nodes whose degree was earlier degree was d minus 1 and due to the addition of this new node that degree will become d ok. So, number of nodes which had degree d minus 1 and upon addition of the new node right the same phrase here their degree increases to d, right and these nodes will be counted will not be counted here will be counted here in N dt, ok. So, we will try to analyse we will try to understand how many such nodes are there in the influx, how many such nodes are there in the outflux, ok.

Now, so, here we have seen that P d is d times P t d divided by 2m, right probability of particular degree d. So, now, since remember so, when a particular node comes in. So, one of the edges one of the one of these m edges will join will basically connected to a node, it will not happen that the same node will again be chosen for another edge, right there will not be any parallel edges, right.

So, the number of nodes of degree d that gain at least one edge, right due to the addition of the new node, right. So, we are interested in the number of such nodes which will gain at least one edge, right. So, m number of edges are coming in and what is the probability that a node d a node with degree d will gain an edge and how many such nodes are there?

So, there are so, P d is the you know P d is the probability of nodes with degree d, right you can also say that P d is the fraction of nodes with degree d, right. So, out of this m edges how many edges will going to be attached to nodes with degree d? This many ok. So, I replace P d by this one. So, this is going to be d right P t d by 2 m and m will cancel out, right. So, remember this particular term. This is basically saying that when a new node joins in, how many nodes with degree d will gain an edge ok gain a degree or whatever an edge.

Now, let us denote another quantity. So, many notations are being used. So, it is very important that you know you basically list down all the notations that we are following so that whenever those notations are referred you can easily understand, ok. So, let us denote another notation P sub d comma n. What is this? This is same as P t d probability of a node probability of a degree d at time t when you know when there are n nodes in the system in the network.

So, when there are n nodes in the network what is the probability of a node what is the probability of a degree d ok. So, now let us write. So, now, let us try to understand the number of nodes with degree d right at a particular time t. So, when this new node comes in right what would happen? So, then the number of nodes will become n plus 1, right. So, at n plus 1 so, when the number of nodes is n plus 1, how many nodes are there with degree d?

So, p of d n plus 1 is what? p of d n plus 1 is the probability of a degree d when the number of nodes is n plus 1 or the probability or it is basically a fraction. So, fraction of nodes with degree d when the number of nodes is n plus 1, so, what is the total number of nodes with degree d when the number of nodes is n plus 1? n plus 1 times this fraction, right. So, this is kind of an influx or this is kind of the quantity that we will contribute when you measure the number of nodes with degree d, right and we need to subtract right this quantity n d n, what is this?

This is the number of nodes with degree d when the total number of nodes was n. When this new node joins in there were already some nodes with degree d and those nodes will now become d plus 1. So, those nodes will now be you know we will not contribute to the total number of nodes with degree d anymore. So, how many such nodes are there?

So, total number of nodes is n and the fraction of nodes with degree d when the total number of nodes is n is this quantity. So, this plus this is the net quantity of the number of nodes that we should consider now ok. So, in other words, this is basically influx minus out flux, right. Now, this is one way of writing influx plus outflux. Now, we can also write influx in a different manner.

We can also write influx you know based on this equation, ok. Now, what is influx? Influx is essentially and remember we have already renamed this notation, right we have already renamed it by d into p d comma n by 2. So, this was renamed here, ok. So, what is influx according to this one? Influx is when the system had when the system had n minus 1 number of edges, n minus 1 number of nodes and sorry, a system sorry let me go back.

So, when the system had n number of nodes, how many nodes were there with degree d minus 1? So, this is the fraction. So, p d minus 1 comma n ok. p d minus 1 comma n is the fraction of nodes with degree d minus 1 when there were n number of nodes in the system, ok. And what is the total number of such nodes times n? Sorry. So, we have already seen that the total number of nodes with such degree is this one, right.

So, let us write in there in I mean using this equation. So, if this is d, this is d. So, this is d minus 1, this would be d minus 1. So, this quantity indicates that number of nodes if you look at this phrase again number of nodes with degree d minus 1, that gain at least one edge due to the addition of this new node. Now, this will contribute to the quantity that we are interested in. So, this is influx, ok.

And, what is out flux? Out flux is p of right d comma n, right the number of nodes with degree d when there were n number of nodes in the system right, those will basically gain one edge right. So, the number of such nodes is this one ok of course, by 2. Now, if this is not understood so, we basically try to write influx and outflux in two different ways one based on this quantity other based on this quantity ok.

So, in the asymptotic label in the asymptotic you know label meaning when n tends to infinity ok all these notations right this small notations like p of d n plus 1 right, p of d n right all these notations these two things would be same, this should be p of d when n tends to infinity. So, we replace this by p of d. So, now, the modified equation is this one, n plus 1 p d minus n p d is d minus 1, right p d minus 1 by 2 minus d p d by 2 because in the asymptotic level when n tends to infinity we just ignore we will ignore n at every place ok.

So, now if you if you know solve this equation and if you try to get the values of p d you will get you know p d is d minus 1 by d plus 1 right p d minus 1 you see at this n p d and n p d will cancel out. So, n will vanish and you will get this equation ok. Now, this equation is true when d greater than m. What about d when d equals to m? d is a degree and m is the this one the number of edges that are attached to particular node, right.

When d goes to m, what would happen? When d goes to m, right we so, this p d will become p m, right. So, I will write this equation again in a different manner. So, this will be n plus 1 p m p p m yeah minus n p m equals to now think about the influx. So, it is basically saying that what would be the; what would be the number of nodes with degree m when a new node comes in?

The number of nodes with degree m, ok, so, what is the influx? Number of nodes with degree m how many such nodes are there with degree m when a new node comes in? 1, because the node which is coming that node has degree m. So, the influx would be 1 and what would be the out flux? Out flux we do not know. So, we need to compute. So, out flux would be let us look at you know let us look at this quantity again right or whatever this quantity.

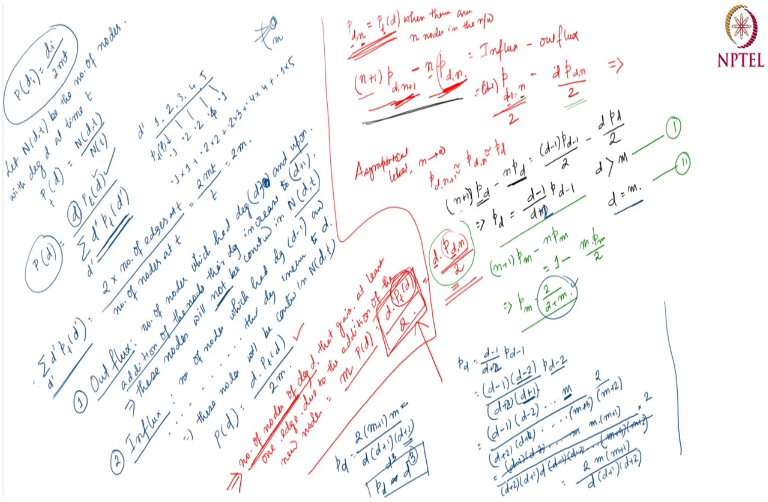
So, this would be m times you know p m by 2, ok. So, here also if we calculate you know if we just calculate the value of p m you will have a p m equals to 2 by 2 plus m. Now, we have two equations when d greater than n, when d equals to m. When d greater than n you have p d this one, when d equals to m you have this one. Now, we need to solve this equation. You see this is a recursive equation, the equation 1, right. p d is dependent on d minus one d minus will d minus one will depend on d minus 2 and so on, right.

So, let us unfold it now, ok and we stop when d goes to m because at d goes to m we have a you know non recursive value. So, p d equals to d minus 1 by d plus 1 p of d minus 1 ok, this is. So, I am replacing d minus 1, right in this equations we will get d minus 1 you know d minus 2 by d plus 1 I think I made a mistake here. So, this should be 2 d minus 1 d d minus 1 d plus 2, ok.

So, this would be 2. So, this is d plus 2 d plus 1 and p of d minus 2, ok and we basically do this same kind of replacement again and again until unless we see this one d minus 1, d minus 2 dot dot dot right we basically do the same replacement until unless we get m, ok. And, in the denominator we have d plus 2, d plus 1 so, what would the last value here? Remember when this is d plus when this is d minus 1, this is d plus 1, right.

So, when so, the difference between this and this is 3, ok now when this is m, this would be m plus 3 ok and this would be p m because here d minus 2 here also d minus 2. This is m, this is m, ok. So, now, what we will do? We will replace p m by this 1. So, we will have let me you know just replace it here.

(Refer Slide Time: 26:33)

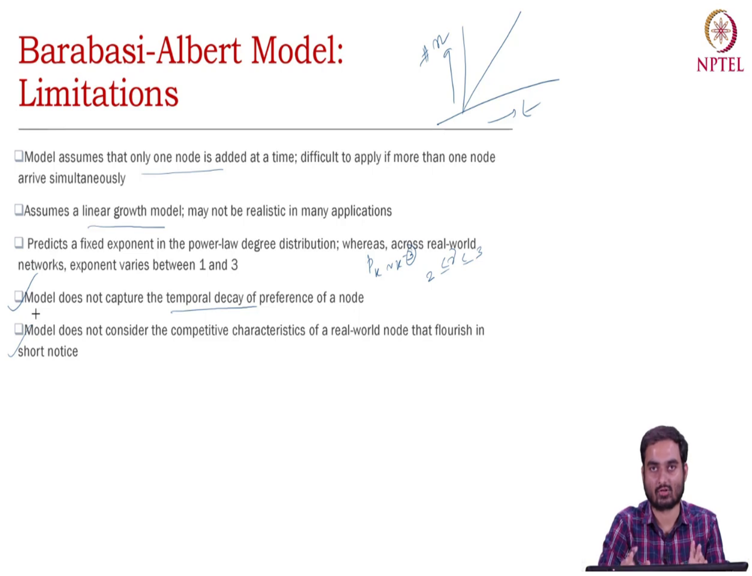


So, we will have 2 by m plus 2 ok. So, now if you look at it carefully so, let me break it further. So, you have d minus 1 d minus 2 dot dot dot m right in the denominator you have d plus 2 d plus 1 d d minus 1 d minus 2 dot dot dot m dot dot dot m plus 3, right and m plus 2 also and we have 2 here. So, this part m into right. So, this part and this part will vanish ok, but we need to multiply this one, right here.

So, this would be; this would be m if you multiply here m and m plus 1, right then this and this will vanish, right. So, we will have 2 m into m plus 1 divided by d d plus 1 d plus 2, ok. So, this is p d. So, p d is 2 into m plus 1 into m and d into d plus 1 and d plus 2 ok. So, m is constant. If you look at the denominator this is d cube, right; d times d plus 1 times d plus 2. So, you will have a; you will have a component which is d to the power 3. So, p d will be proportional to d cube d to the minus 3 ok. This is proportional to 1 by d cube meaning it is proportional to d to the power minus 3.

So, we will see here that the degree distribution here also follows a power law and the coefficient is minus 3 and the same coefficient we obtained in case of I mean when we used continuum approach ok. So, the proof is this approach is little bit tricky, but I suggest if you I mean if you think about it carefully and if you derive it on your own you will understand it, ok. So, this is all about you know Barabasi – Albert model.

(Refer Slide Time: 29:57)



Now, if you look at the limitations right limitations of Barabasi – Albert model. The major limitation is that the major limitation is that you know is that we consider only one node at a time when you know when increasing the network. So, at every time only one node is joining in right, but in real world there are multiple nodes that can join at the same time, ok.

And when multiple node join when multiple nodes join in the network we do not know how to how to you know make them I mean how to come up with a sequence that this will be considered first, this will be considered second and so on. So, this is not clear, ok and this is a linear growth model because every time only one node joins.

Therefore, the growth will be linear with time. So, if we see the t if we see t and number of nodes; number of nodes in the system will see like this, but real world network does not follow this linear growth kind of pattern.

The both the you know models the both the approaches continuum approach and master equation we observed that the coefficient of the power law is minus 3. I mean 3 right we observe that p k is k to the power minus 3. So, this is 3 is gamma, but in the real network real world case gamma actually varies from 2 to 3 even 1 to 4 sometime, right. So, we mean we would not be able to get a variety of such you know scale free graphs using Barabasi – Albert because Barabasi – Albert according to Barabasi – Albert the coefficient is always fixed.

BA model does not capture the temporal decay of preference of a node, what does it mean? So, as the time goes by what happens is that a node actually starts decreasing its importance, ok. Now, think of you know movie actor, actress, right as they become older and older their preference may decrease. But, I am not saying that this is true, but in general their preference will decrease ok. So, this kind of temporal decay of preference is not explicitly present in the Barabasi – Albert.

And, the last problem is that there is no such factor of competitiveness, right. What does it mean? It means that you know when a new node joins; it does not have the power to compete with others. Now, think of a system, think of a you know again scientific research collaboration kind of a system where and when a new scientist joins in the scientist may be brilliant, right and the scientist although he is young, but you know he or she you know has the capability to compete with other researchers.

Similarly, you know movie actor, actress – when a new actor, actress you know joins, right he or she may have you know better performance ability, ok. But, according to Barabasi – Albert model you will act you will have to survive for certain point in time in the system so that you gain certain edge and your degree will increase and then your preferential attachment probability will also increase.

There is no such competitiveness or competition that as soon as a new node joins in he or she has you know that the node has the capability to compete with others to gain you know new edges ok. So, these are the problems of Barabasi – Albert model. Later on many such advance models have been proposed, we will not discuss those models in this particular chapter. I will try to in the next lecture I will try to give you a very brief of you know the other such approaches which basically try to address this limitations ok. With this I stop here.

Thanks.