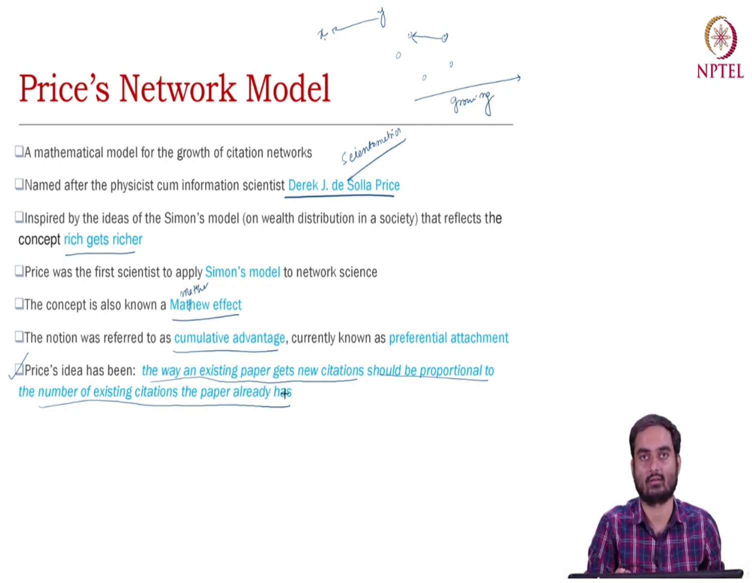
Social Network Analysis
Prof. Tanmoy Chakraborty
Department of Computer Science and Engineering
Indraprastha Institute of Information Technology, Delhi

Chapter - 03
Lecture - 06

So, in the last lecture we completed Barabasi Albert Model and we discussed two approaches continuum approach and master equations and we saw that using both the approaches you know we could you know we can see that the Barabasi Albert Model actually produces networks which follow power law degree distribution. And BA model is possibly the first model which actually shows that and I mean how you can think of a power law kind of a model right power law kind of a network, how you can synthetically generate power law network.

So, but it is not the case that you know Barabasi Albert model or Barabasi Albert were the first to explain this long tail or heavy tail powered law kind of degree distribution ok. You know long before that long before Barabasi Albert around 1916, 1956, 1960 around that time right the hunch of you know creating such a model which follows power law kind of a network which produces power law kind of network you know that was there and I mean if you go back to the history right. So, the first such model was actually proposed by D J. Solla Price.

(Refer Slide Time: 01:51)



Now, here price was a was basically a physicist cum information scientist and he is known as the father of the field called scientometrics. Scientometrics is basically a field you know scientometrics; scientometrics is basically a field which basically deals with measures and analysis of scientific papers, scientific articles can be papers, patents right and D J. Solla Price is known as the father of the field scientometrics, right.

So, he actually you know came up with such a model which produces a power law kind of degree distribution in the synthetic network ok. So, I thought you know this should be discussed because this is people generally overlook this model, but this might be the first model which produces this power law kind of distribution ok.

So, but price actually you know showed this the model for the citation network, let me recap what is citation network. In the citation network nodes are you know scientific papers, scientific articles and a scientific article can cite another scientific article right through a citation link. And remember this is a growing network; this is a growing network in the sense like in the sense that the nodes if joined in a system right those nodes will never be deleted if a paper is published the paper will not be deleted ok.

So, citation network is a directed acyclic growing network ok think why it is as why it is acyclic, why it does not have a cycle right. Directed acyclic growing network and in this particular network I mean when price kind of tried to model the growth of citation network, so he came up with this model ok.

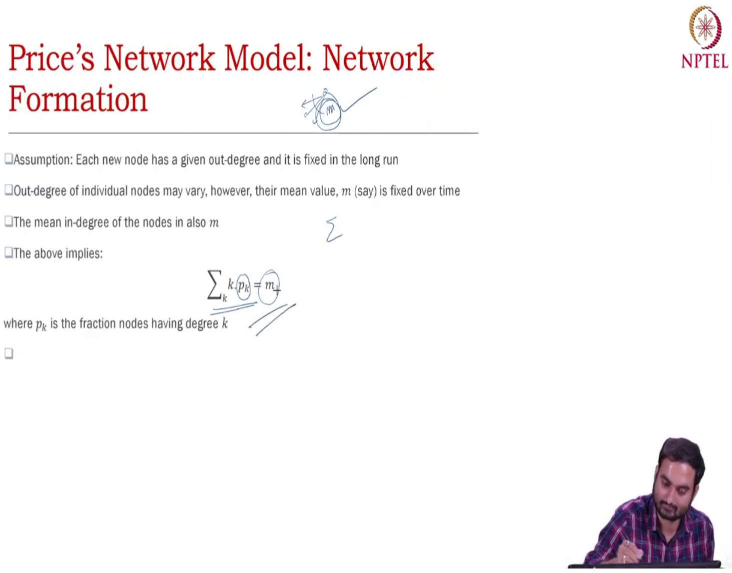
So, according to you know price or even according to there is something called Simon's model or Simon's model. So, according to this Simon's model or price model you know when a new citation comes in or new paper comes in it is highly likely that the new paper will cite an existing paper with high degree. Meaning that if paper x has already been cited by many other papers, it is also highly likely that a new paper will also cite this paper and this is quite natural. If you think about it this is quite natural we generally tend to cite those papers which are already very famous ok.

So, this is same as rich gets richer phenomena that we mentioned in the last time this is also called Matthews effect M a t t h e w Matthews effect ok and this rich get richer the rich gets richer phenomenon this particular term was not there earlier. So, price invented the term called cumulative advantage ok.

Now, what is cumulative advantage? Cumulative advantage is same as preferential attachment basically says that over time you cumulate your preference meaning you cumulate your citations and that would give you advantage is called cumulative advantage ok.

So, the idea is as follows; the idea is that when a the way an existing paper gets new citations should be proportional to the number of existing citations that the paper already has ok. This is very same as Barabasi Albert model but remember Barabasi Albert model does not explicitly talk about directed graph, but here we explicitly talk about directed graphs ok.

(Refer Slide Time: 06:02)



So, in the price model, so the assumption is that when a particular node comes in right when a particular node comes in it again comes with m number of citations m number of references.

So, remember when we mentioned out degree edges in case of citation network out degree edges outward edges are basically references and inward edges are citations ok. So, a note comes with m number of references, m number of outward edges right.

Now, people may argue that how can we fix you know the number m because papers have different references for example, if a paper is a short paper you will get hardly 5 to 10 references if it is a long 10 page paper you will get 20 to 25, 30 citations, if it is a survey paper you will get hundred 100, 200 citations I mean references ok.

So, how can we fix the number m? That is a very valid argument. So, he said that ok forget about you know forget about fixing m, but we can at least assume that the average number of references or the average number of edges that a node you know brings in that is same; that is m. And that is a valid assumption why?

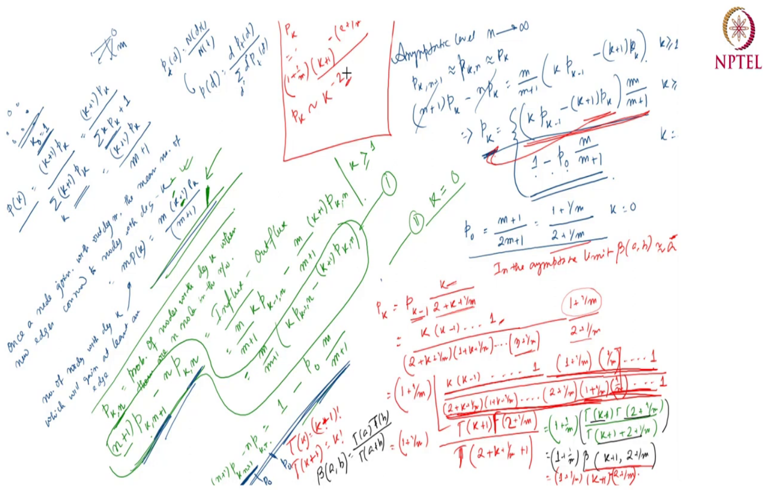
Because we know that the sum of you know degree of all the nodes is two into number of edges. So, if we assume that the total number of edges in the system is fixed right, we can also assume that the degree is also average degree is also kind of same right.

So, let us assume that the average number of; average number of edges that a particular node brings in is m ok, this is average out degree or average outward edges, average references of a particular node ok.

So, then let us you know define some quantities again here. So, what is p k? p k is the fraction of vertices with degree k, right. So, if we sum over sum across all p k. So, k times p k and you sum it you basically sum them. So, that will give you the average degree. So, the average degree is m ok; the average degree m. So, the average in degree is also m average out degree is also m although individual level they are different, but the average is same ok.

So, now let us again use whiteboard and derive the same thing. So, we will again derive that the power law distribution is being followed in case of price model ok.

(Refer Slide Time: 08:58)



So, a new node comes in with m number of edges outward edges this is the paper scientific article and there are already some nodes existing nodes there right and let us assume that; let us assume that on an average again the existing nodes have one in citation in degree one citation ok.

This is again reasonable because price mentioned that you know when a paper is published you can assume that the paper already cites itself. So, there is already one citation when a paper gets published. So, in that way we can assume that the existing nodes right when we start the system there is you have some existing nodes and existing nodes have on an average degree 1.

Now, let one node come in and you decide you know how to attach that node with the other nodes. So, initially we will assume that there are k 0 number of edges per node or k 0 degree per node which is 1 ok.

So, the number of. So, what is the probability of a node degree k? So, this would be in the same way in the last day we mentioned if you remember in the last lecture we mentioned that we mentioned a quantity right which is which was this one which was p t d right N d t by N t and then from here we derived p d which was d p t d this was discussed the master equation d dash sum over all d dash p t d dash.

Here also we will do the same thing ok, but this time we have k plus 1 degree because we have this initial degree. So, instead of writing k we write k plus 1. And what is this? The denominator the numerator is this one and this is k p k plus summation of p k is 1.

So, what is summation of k p k? This is the average degree and you have already seen this is m right already seen this is m ok. So, this is p k now right now here also we do the same thing influx, outflux ok. So, once a new node a node joins right without degree m, the mean number of new edges connected to nodes with degree k is what? m p k last day we were also in the last lecture we will also you know saw this thing the same equation.

So, this is what this is m times k plus 1 p k by m plus 1 and this is also same as this is same as number of nodes with degree k which will gain at least an edge; at least an edge right which is same as this one. So, and this equation we have also seen in the master equation part right.

So, let us again denote another notation which is you know p k n this one also we discussed in the last lecture. What is this? This is the probability of nodes with degree k when there are n nodes in the network ok.

So, we again you know we again calculate influx minus outflux right. So, what is influx? Influx is when this new node joins in right minus when the new node joins in how many nodes are there with degree k and when there were when there were n number of nodes before joining this new node there were n number of nodes right how many nodes were there with degree k? Because those nodes will essentially gain an edge right this one right this is based on this one.

Now, let us look at this equation and based on that let us calculate influx and outflux. So, ok. So, when there are n number of nodes and degree k minus 1 we look at degree nodes with degree k minus 1 those nodes will gain an edge.

So, how many such nodes are there, how many such nodes are there with degree k minus one which will get affected? This would be m plus 1 right. So, when it is k this is k plus 1 when this is k minus 1 this would be k. So, k right p k minus 1 comma n right. So, this is influx and what is outflux? Outflux would be m plus 1 k plus 1 because though the nodes with degree k will now become k plus 1.

So, how many nodes are there with degree k when the number of nodes is n? Right. So, this would be ok when k greater than 1 greater than equals to 1 you keep m into m plus 1 outside we have k p k minus 1 m n minus k plus 1 p k comma n this is equation 1.

What happens when k equals to 0 this is k greater than equals to 1 when k goes to 0 what happens? So, k equals to 0 means what? k equals to 0 means a node with in degree 0; the node with in degree 0. So, when a new node comes in right how many. So, we need to calculate influx. So, how many nodes would be there with in degree 0? Only one because the node which comes in that node has out degree m, but in degree 0.

So, influx would be one and how many nodes will actually move from degree 0 to degree 1? That would be your out flux that we need to calculate right and that would be we can calculate from this equation right. It will be by m plus 1 because k equals to 0. So, we have only m by m plus 1 times p 0 right. So, n plus 1 I am repeating this line again p k n plus 1 minus n p k n. So, this would be this one when k equals to 0 ok.

So, right now if you only look at this equation ok and you rearrange it you will get this one, you will get because in the again in the asymptotic level when n tends to infinity p k, n plus 1 this one p k, n; p k, n those would all be same as p k. So, we replace this by p k ok.

So, essentially we will have what? We will have n plus 1 p k minus n p k to m by m plus 1 k p k minus 1 minus k plus 1 p k ok and this is when k greater than equals to 1. So, if we rearrange it you will get p k equals to k p k minus 1 minus k plus 1 p k times 1 because this term and this term will vanish this is k greater than equals to 1.

What happened with k equals to 0? When k equals to 0 if you calculate it here again if you rearrange it you will get you will get this is this would be 1 minus p 0 right I mean I am just writing these the these equations without making any changes right ok.

Now, essentially when k goes to 0 this is also p 0 and this is also p 0 is not it. So, what would happen? So, if you just consider this part ok you will see that this would be p 0 m plus 1 p 0 would be m plus 1 by 2 m plus 1 this is not recursive ok, this would be 1 by 1 plus m right 2 by 1 plus m when k equals to 0 ok. Because this is p 0 this is. So, p 0 equals to 1 minus p 0 times m by m plus 1. So, you move all p 0's in the left side and then you calculate ok. So, this one we will use when we use when we unfold the recursive this recursive equation alright.

So, let us unfold it ok. So, you have p k which is I mean if you rearrange this equation a little bit you will see that this would be p k minus 1 because you move p k in the left side right you move p k in the left side, so you will y have I mean a particular quantity with pk and the remaining part would be p k minus 1.

So, this would be k by 2 plus k plus 1 by m and then we unfold it ok. So, k times k minus 1 we unfold it till 1 and this would be 2 plus k plus 1 by m, then 1 plus k by k plus 1 by m dot dot dot we have 3 plus 1 by m and then we will have p 0 ok. Now, think about it you have k minus 1 here and therefore, k here you have 0 1 0 here therefore, 1 here ok right and also look at the difference between this and this ok. So, this minus this the denominator and numerator if you take the difference this would be 2 plus 1 by m.

So, if this is 1, so this should also be 3 plus 1 by m then only the difference would be 2 plus 1 by m ok think about it carefully and you will understand it. Now, I replace p 0 here by this equation. So, we will have what? We will essentially have 1 plus 1 by m 2 plus 1 by m ok.

So, let me move let me keep this thing outside. So, if I keep this thing outside ok we will have dot dot 1 here 2 plus k plus 1 by m 1 plus k plus 1 by m dot dot dot 3 plus 1 by m 2 plus 1 by m right 2 plus 1 by m ok. I multiply 1 plus 1 by m in the numerator as well as in the denominator, I can do that ok.

So, what is this one? 1 times, 2 times 3 dot dot dot times k this is factorial of k right. So, factorial of k we know that if you do not know there is something called gamma function ok, gamma k is essentially factorial of k minus 1. So, gamma k gamma k plus 1; so gamma k plus 1 is factorial of k. So, this part is essentially gamma this part is basically gamma function this is gamma factorial of k plus 1 ok.

So, let me again write it 1 plus 1 by m ok in the numerator we have gamma k plus 1 times 1 plus 1 by m ok and in the numerator; in the numerator we have we can we also try to write this in terms of gamma right. So, we have the last term is 1 plus 1 by m we need to add some more terms here right then, so that this will be gamma this part ok. So, I basically multiply 1 plus I multiply this with 1 by m right 1 by m 1 by m minus 1 dot dot dot 1, I also need to do the same thing in the denominator this would be 1 by m dot dot dot 1 ok.

So, if you do that then the this would be basically a factorial 1 from 1 till 2 plus k plus 1 by m and this would be gamma of it should be gamma of 2 plus k plus 1 by n plus 1 because gamma k plus 1 is factorial of k. So, this is gamma 2 plus k plus 1 by m plus 1 and if you look at this part right this is also factorial from 1 to m plus 1 to 1 plus 1 by m right this can also be written as gamma 2 plus 1 by m ok.

So, now the entire equation is as follows, the entire equation is you see a nice equation is emerging right. So, this is 1 plus 1 by m gamma k plus 1 gamma 2 plus 1 by m by gamma k plus 1 plus 2 plus 1 by m ok and if you know there is another function called beta function ok. So, beta a b beta a comma b beta takes two parameters this is basically gamma a times gamma b by gamma a plus b. So, this is essentially a beta function you see here gamma a plus gamma b by gamma a plus b.

So, this is 1 plus 1 by m beta k plus 1 comma 2 plus 1 by m ok there is another nice theorem which says that in the asymptotic limit right in the asymptotic limit this beta a comma b would tend to e to the power minus p right. So, if we think of this in the asymptotic limit this would be 1 plus 1 by m beta k plus 1 comma 2 plus 1 by m would be k plus 1 to the power 2 plus 1 by m ok.

So, let me write it afresh what we have got. So, we have got p k is 1 plus 1 by m times k plus 1 to the power 2 plus 1 by m; 2 plus 1 by m and I am sorry this is minus right I am sorry this is minus, so we will have minus here ok. So, will be the power would be minus. So, this would be minus 1 plus 1 by m k plus 1 minus 2 plus 1 by m ok. It basically says that p k is proportional to k to the power minus 2 which is power law where the coefficient is minus the coefficient is 2 here or whatever minus 2 here ok.

So, price according to price model you will also get power law a network with power law degree distribution, but he showed this thing on the citation network and at that time you also you can also understand that you know there were there was no such computational facilities right in access to large data set therefore, price was unable to you know simulate his analytics you know on the real world network right which later on Barabasi Albert Barabasi Albert model actually did simulation on the real world network ok.

So, I stop here and in the next lecture the next lecture is going to be the last lecture on this particular topic on this particular chapter. I will briefly talk about the remaining few models and then we finish this chapter ok.

Thank you.