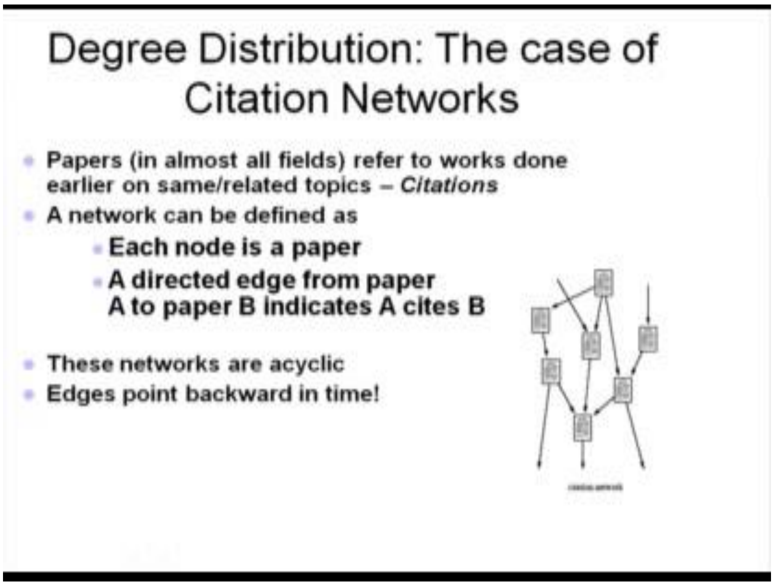
Complex Network: Theory and Application
Prof. Animesh Mukherjee
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur

Lecture - 02
Network Analysis – I

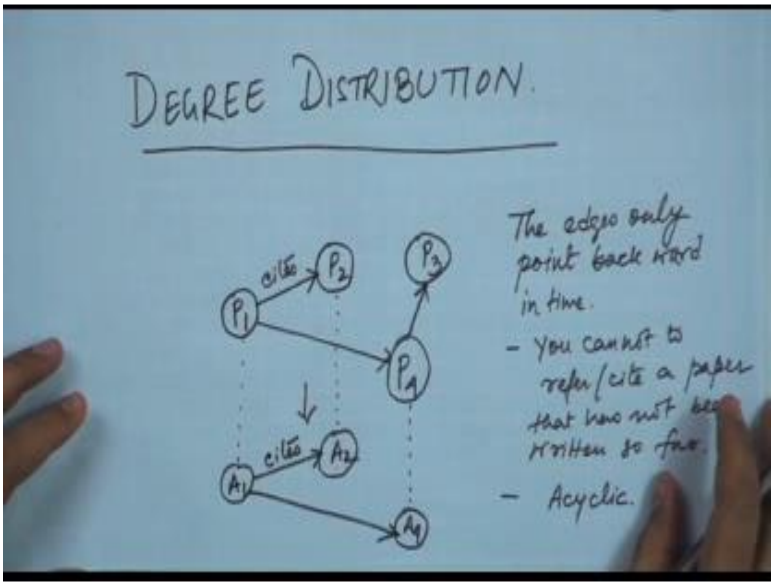
Welcome back to this session on Complex Networks. And, today we will look mostly into network analysis metrics. So, we will start off with Degree Distribution.

(Refer Slide Time: 00:33)



So, this is the first metric that we will look into. This is called Degree Distribution.

(Refer Slide Time: 00:38)



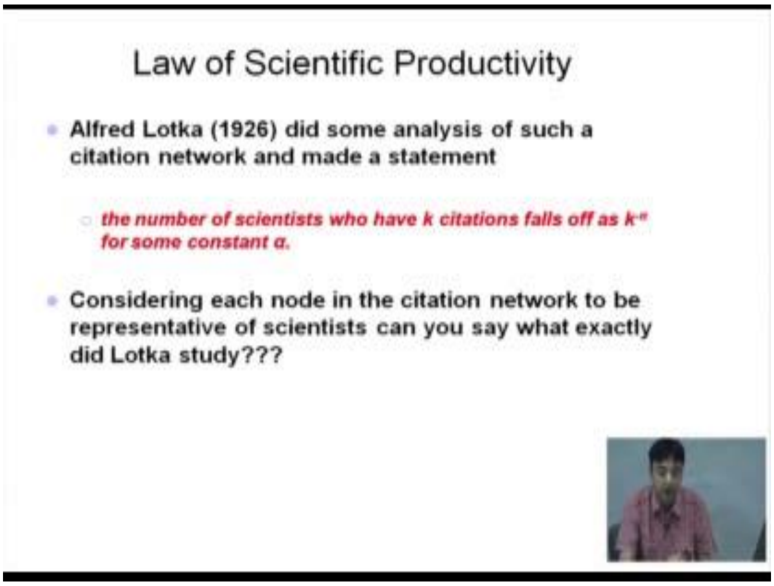
Now, we will try to understand the metric using the example of citation networks that I was talking about in the last lecture. So as I already told you, and as you also see in the slides, the citation network is composed of nodes which are papers say P 1, P 2, P 3, P 4. And, if a paper says P 1 refers or cites a paper P 2, then there is a directed link from P 1 to P 2, indicating that P 1 cites P 2. Similarly, P 1 might also cite P 4 and P 4 may cite P 3, and in this way the network gets constructed.

So, the point to note is that in this particular network the edges only point backward in time. This is easily understood because you cannot refer. This is since; you cannot refer or cite a paper that has not been written so far. So, edges only point backward in time. And, this network, consequently this network is acyclic.

So now given this network, we can measure the degree distribution as follows. Or, first let us understand what it means by degree distribution or when we say degree distribution, what it exactly means? Now, from this paper-paper citation network that I have already drawn in this picture, one can also have a author-author citation network, where an author say A 1 who is writing the paper P 1 actually can cite some author A 2, who has written the paper P 2. So, so say author A 1 has written a paper P 1 and author A 2 has written a paper P 2. So, then A 1 cites A 2 in the author-author citation network.

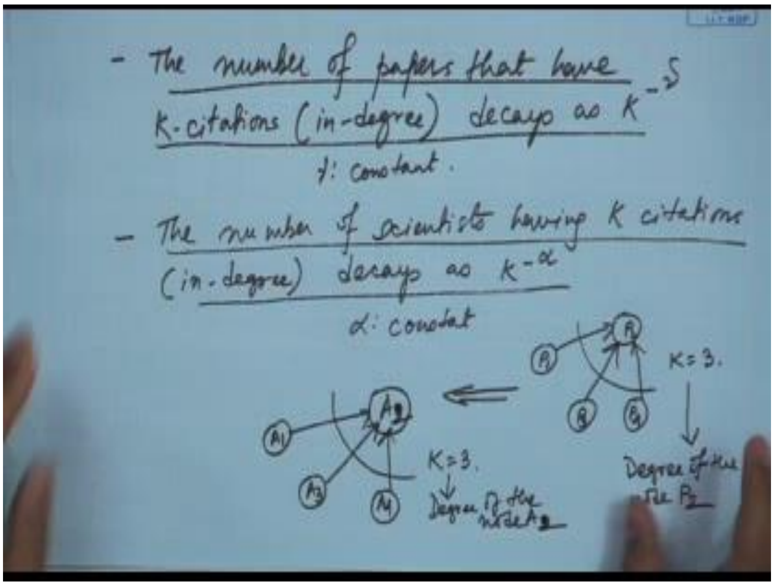
And, similarly say there is some author A 4, who also A 1 cites because P 4 has been written by A 4.

(Refer Slide Time: 04:18)



Now when this construction was made, there was a famous scientist called Alfred Lotka. In 1926, he observed very interesting phenomena. And, what that phenomena was is that you can write in both in terms of papers as well as in terms of authors.

(Refer Slide Time: 04:39)



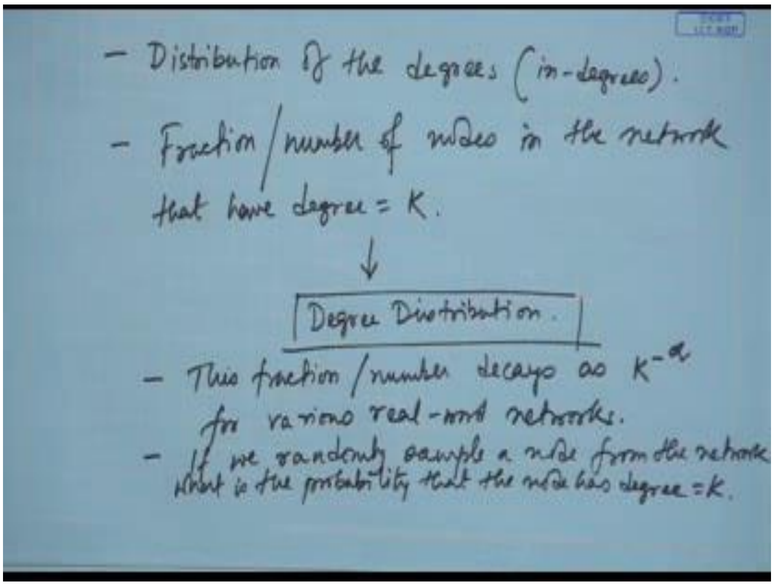
So, the number of papers that have k-citations that is in-degree, so since the citation network is a directed graph, it can have both in-degree and out-degree. So, K citations or in-degree, the number of papers that have K citations decays as K power minus some constant gamma, where gamma is a constant. You can also translate this in the context of the author- author citation network.

So, basically the number of scientists having k-citations, again in-degree decays as K to the power some other constant alpha, where alpha is as I already said is a constant. So, basically if you look at the paper-papers; just to recap, if you look at the paper papers citation network and say there is some paper P 2 who is being cited by three other papers P 1, P 3 and P 4, then in this case K is equal to 3. K is equal to 3 in this particular example.

Similarly, for the author-author correlate, you can have if A 1 is an author of P 1, A 3 is an author of P 3 and A 4 is an author of P 4, then, and all of them are referring to P 2 which has been authored by A 2. Then, the degree here is equal to three. Now, whenever I am talking off, the number of papers that have k-citations decays as K to the power minus gamma or the number of scientists having k-citations decays as K to the power minus alpha.

What am I trying to express in the context of the two networks, the paper-paper citation network and the author-author citation network. It is basically what we are trying to measure? So, as you see here K is 3; here again K is 3. So, this is basically nothing but the degree of the node P 2. Here, this case is nothing but the degree of the node A 3, sorry, A 2; so degree of the node A 2. So, this is basically when we are talking off this number, we are basically talking off the degree. And, what exactly is this expression translating to? The expression is translating to what we are interested in and is the distribution of the degrees or in specific distribution of in-degrees, OK.

(Refer Slide Time: 08:24)

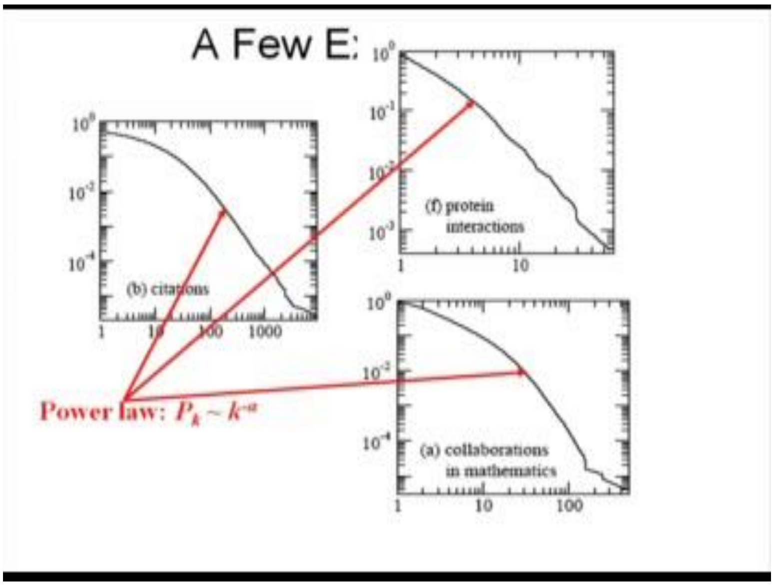


So, what we are exactly measuring is the fraction of nodes, fraction or number of nodes in the network. This could be either the paper-paper network or the author-author network. Fraction or number of nodes in the network that have degree equal to K, this is what is referred to as the degree distribution of a particular network. So, basically (Refer Time: 09:27) the fraction or the number of nodes in the network that have a degree K is defined as the degree distribution. And this, usually this fraction, this fraction or number decays; is seen to decays as K to the power some constant minus alpha for various real world networks.

You can also reinterpret this definition in the following probabilistic term. So, in the probabilistic terms you can say that is if you randomly sample a node from the network what is the probability that the node has degree equal to K.

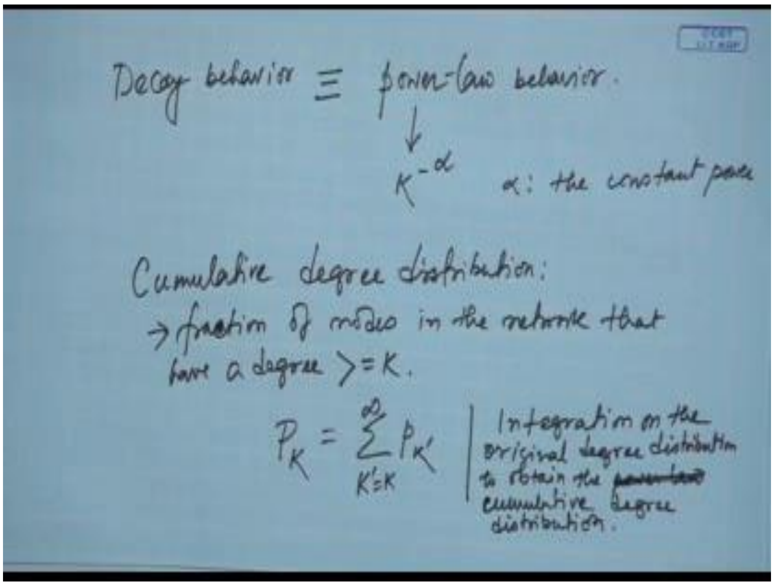
So, this is another way to interpret the definition of degree distribution; is another way of reinterpreting degree distribution. So, you would, you toss a coin and you randomly pick a node from the network and the probability that this node has a degree K is expressed by the, this probability is expressed by the degree distribution.

(Refer Slide Time: 11:18)



So in general if you look at the slides, in general this degree distribution for various real world networks looks like the following three figures that I have shown you in the slides. So, in all these three figures what we have plotted is this degree versus the fraction of nodes having that particular degree. And, this actually follows the decaying behavior.

(Refer Slide Time: 11:52)



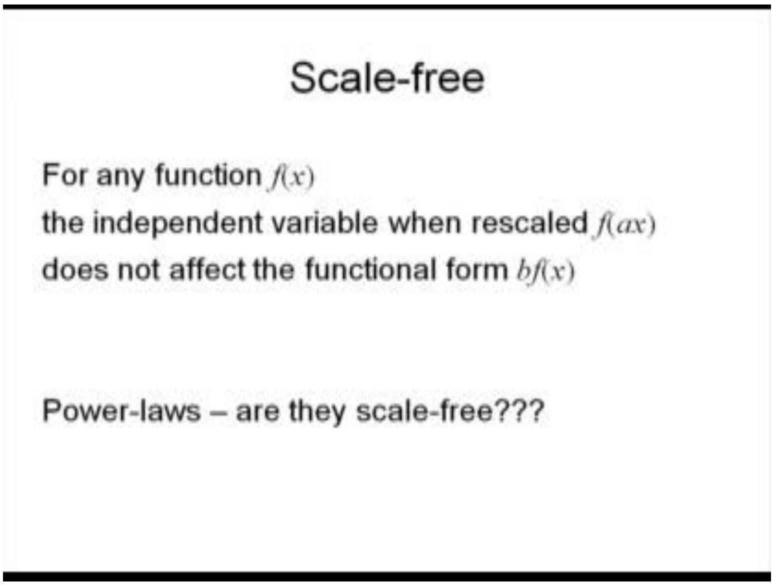
This decaying behavior is sometimes called the; this decay behavior or this decay pattern behavior is sometimes termed as the power-law behavior because this is because the decay happens in some K to the power minus alpha, where alpha is the power, the constant power. So, and you see all the different networks that I have given examples of in the last lecture. Many of them actually follow this kind of a power-law degree distribution. So, their degree distribution has a power-law behavior. So, sometimes it is more like.

So, what happens since these are real world data? In many cases while gathering the data there are some noise incorporated in these data. Sometimes in order to smooth, in order to smooth out the degree distribution functions, rather than looking into the density function, you look into the cumulative density function. Cumulative degree distribution, which is nothing but this can be reinterpreted as fraction of nodes in the network that have a degree greater than or equal to K. Rather than exactly equal to K, you have this loose constraint; fraction of nodes having degree greater than or equal to K.

So, and the expression then for; then, you can translate the original degree distribution into a cumulative degree distribution by the expression, simply small p K, some k to infinity. So, this is you are summing up. So, you are integrating; integration on the original degree distribution to obtain the power-law distribution, sorry, to obtain the cumulative degree distribution.

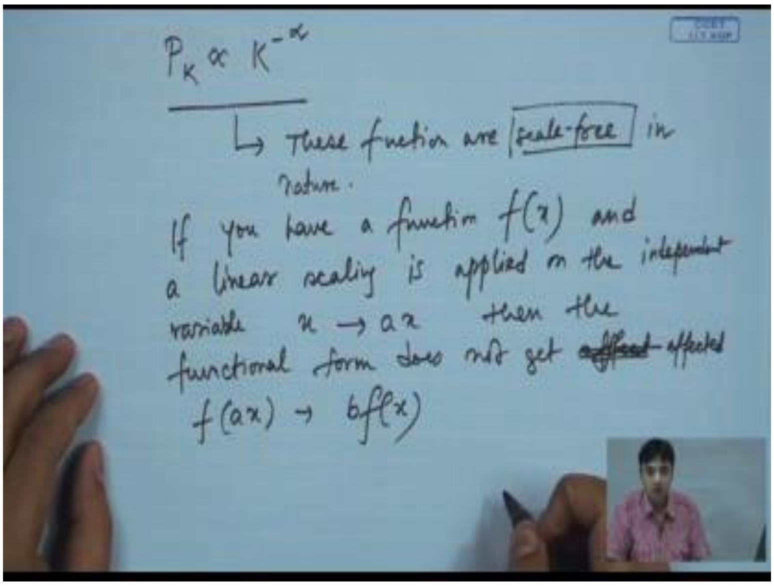
So, basically again reiterating you can express the degree distribution as a raw probability distribution, where it is interpreted as the fraction of nodes having degree equal to some value K. You can also express it in a most smoothened way where you measure the cumulative degree distribution, which is nothing but it is a sum on the raw values, and you measure that. Rather than measuring the exact probability of a node having degree K, you measure the fraction of nodes having in the network that have a degree greater than or equal to K. So, this is one of the first interesting examples that we have seen.

(Refer Slide Time: 15:54)



Now, why is this kind of a structure interesting?

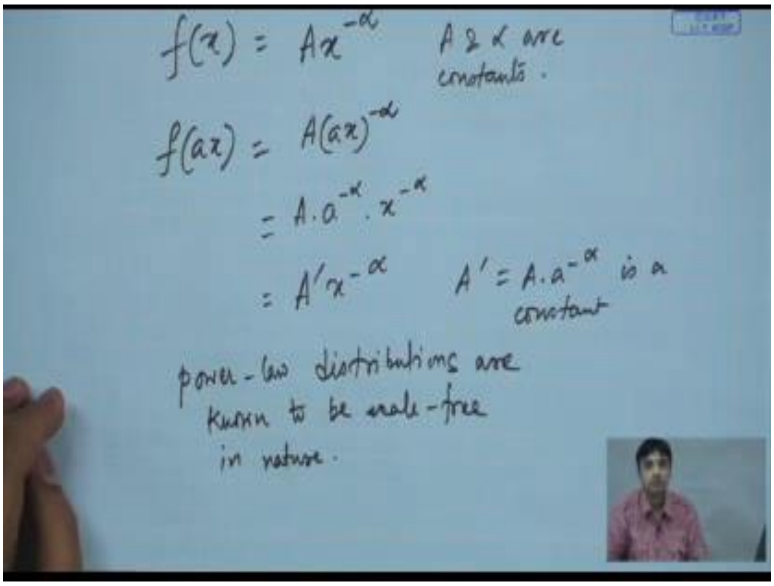
(Refer Slide Time: 16:01)



Now, so this behavior that you see that P K is kind of proportional to K to the power minus alpha is an interesting behavior in itself. So, this power-law behavior is in itself an interesting behavior. And, the interesting part comes out of the fact that these functions are scale-free in nature. So, this is an important term; ‘scale-free’ in nature. Now, what do you mean by a function being scale-free?

So, basically if you have a function say f x and a linear scaling is applied on the independent variable say x is scaled to a x, then the functional form does not get affected. In other words, f a x translates to some other function b f of x. So, the as you see, the constant inside the independent variable comes out as a prefactor of the function and the functional form still remains the same. So, let us see if a power-law function is a scale-free function or not.

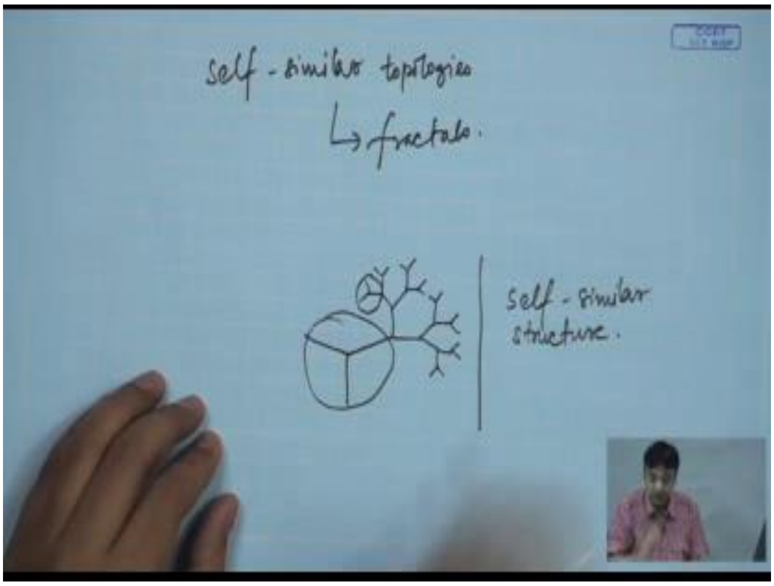
(Refer Slide Time: 18:13)



So, let us say that my f x is some A x to the power minus alpha, where A and alpha are constants. OK. Now, let us say we have f of a x; this should be nothing but A a x to the power minus alpha. So, this can be re written as A into a to the power minus alpha into x to the power minus alpha. Now, this can be written as some a prime x to the power minus alpha, where A prime is equal to A into a to the power minus alpha is a constant. So, you see the new function, the new scaled function also looks exactly similar in shape to the original function f x.

So, that is why power-law distributions are known to be scale-free in nature; fine. So, this is one of the interesting properties of power-law of topologies, which have a power-law distribution. Now, how can we interpret back this power-law distribution in terms of the network topology? So, I will try to give you a very small example. And then, from there try to illustrate the basic idea.

(Refer Slide Time: 20:17)



So, these kind of scale-free structures are very prevalent in the context of self-similar topologies. And, one of the best examples of self-similar topologies is fractals. In case, you are unaware of what fractals are, these are various interesting geometric shapes which have dimensions that are fractional. And, you can know more about them in the, from the Wikipedia entry for fractals. You can go back and it would be interesting to you to have a look into this and make a nice summary out of this. This could be an assignment for you.

So, to go back and look for what are fractals? If you already aware, it is fine. If you are not, it is a good assignment to go back and look in to the definition of fractals; how they look like, how their existence is shown in various real life systems.

But, to give, to motivate you with an example self-similar structures would look like something like this. So as you see, this looks like the branches of a tree. So, basically what is happening is that every individual part is similar to every other individual part, only smaller in size. That is what is happening. So, every individual part here is basically a smaller version of a larger, of one of its larger counterpart. So, such structures are called self-similar structures.

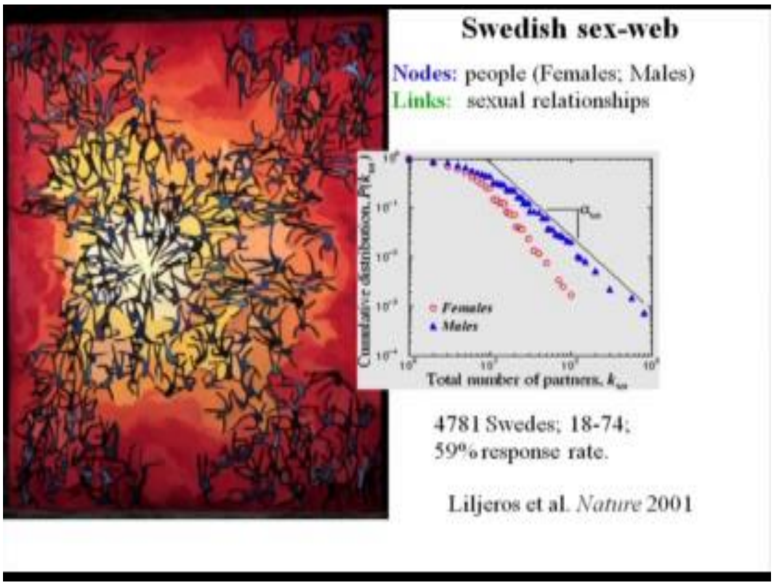
And, what we interestingly note is that for these kinds of complex networks that we encounter like the collaboration network, the organization network that we have seen last day, the terrorist network, the airport network, and all this networks actually show this power-law degree distribution. And, this actually internally indicates that the topology is very self-similar.

So, basically what I mean to say by that is that say suppose, you have this entire complex network. Now, you remove some of the nodes from the periphery of this network. And, you re-estimate the degree distribution. You will again let a power-law. Once again, if you remove the nodes in the periphery, you keep on doing this recursively you will every time land up getting a power-law. So, that is why we call this kind of structures are self-similar structures.

So, there is a core, internal core, then there is a moderate core, then there is a less moderate core and then there is a periphery. In this way; so most of these networks are arranged in such a layered fashion. And, that is why like every layer you move, you again get back a power-law degree distribution or self-similar structures. So, the first layer is removed. The rest of the network again behaves similarly as the original whole. Then again you remove the next layer; the rest of the network behaves as similar to the original whole and so on and so forth. So, that is why these are called self-similar structures.

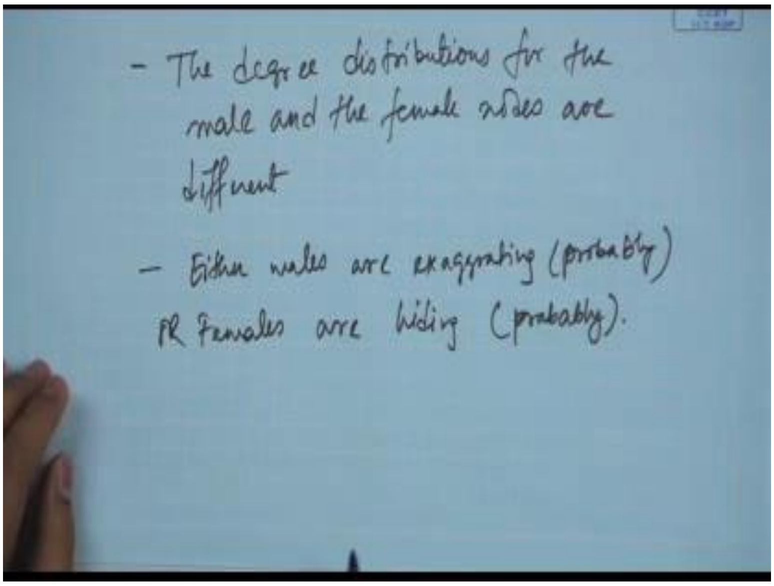
So, they repeat actually. So, even if nodes get remove from the network, this topological property, this topological property of self-similarity actually is retained in the network because it actually sits into the prime organizational principles. That is, you have a core, you have a looser core, you have an even looser core and then you have a final periphery. So, in this way, in this layered fashion these kinds of networks are arranged. And, that is why you observed such a self-similar behavior in this kind of networks.

(Refer Slide Time: 24:47)



So, now to motivate you further as I was talking about this Swedish sex-web last day. I already talked about this example of Swedish sex-web. So, where each node is an individual and an edge actually can be interpreted as a sexual relationship between a pair of individuals. Now, we will assume that this network is a heterogeneous network. And, what happens is that if as I show you in this slide if you draw the degree distribution of the male nodes, so those that are only males in the network, those that are only males in the network actually are represented by the blue triangles in this figure in the slides. Those that are only female nodes, their degree distribution are indicated by the red circles that you see in the figure.

(Refer Slide Time: 25:45)



So, what one can immediately say. So, one interesting observation is that the degree distributions for the male and the female nodes are different. So, the degree distribution of the male and the female nodes seems to be very different actually. But then, how can these be possible? Now, one immediately has to question like how is this possible. See, this has to be a very balanced network. So, every sexual relationship by a male has to be balanced by individual female. If that is not the case, then there can be a discrepancy in the degree distribution.

So, that means there is some problem. As soon as, so unless you look, unless you do this plot, it is very difficult to understand whether there is some problem going on in the construction of the network. So, there is some problem that has gone in to the construction of the network that because the degree distribution of the males and the females should match. Ideally should match in order to balance this sexual relationship. If, since it does not match as we see the in the figure; it does not match, so we can definitely say then there is some problem in the construction. So, what is the construction problem? The construction problem is either males are probably exaggerating or females are hiding; both probably. We do not know which one of this is true, but one of this is definitely happening.

So these, actually unless you have plotted this degree distribution, there is no way to understand that such a thing is happening. So, this actually tells you that the data that you have collected and the network that you have constructed is not a true representation of the real system of the real complex system. So, there is some problem going on. And, probably there is a need for the re-estimation of the survey. So, that was a very interesting observation that these authors, these set of authors actually reported it in one of the very prestigious journals called “Nature” in 2001.

Thank you very much.