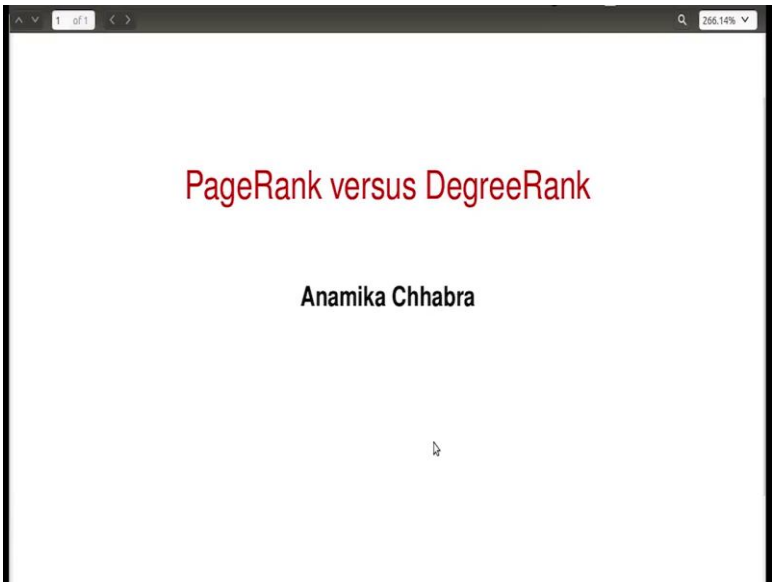
Social Networks Link Analysis
Prof. S. R. S. Iyengar
Prof. Anamika Chhabra
Department of Computer Science
Indian Institute of Technology, Ropar

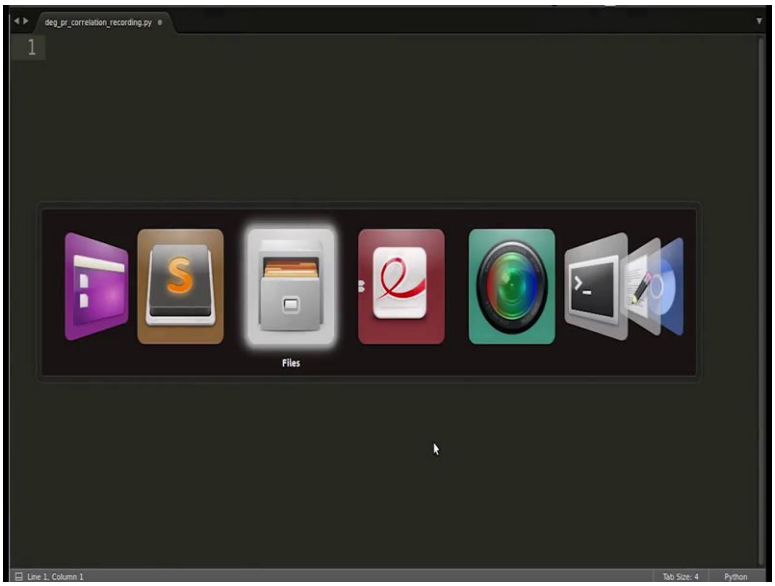
Lecture – 86
Degree Rank versus Page Rank

(Refer Slide Time: 00:05)



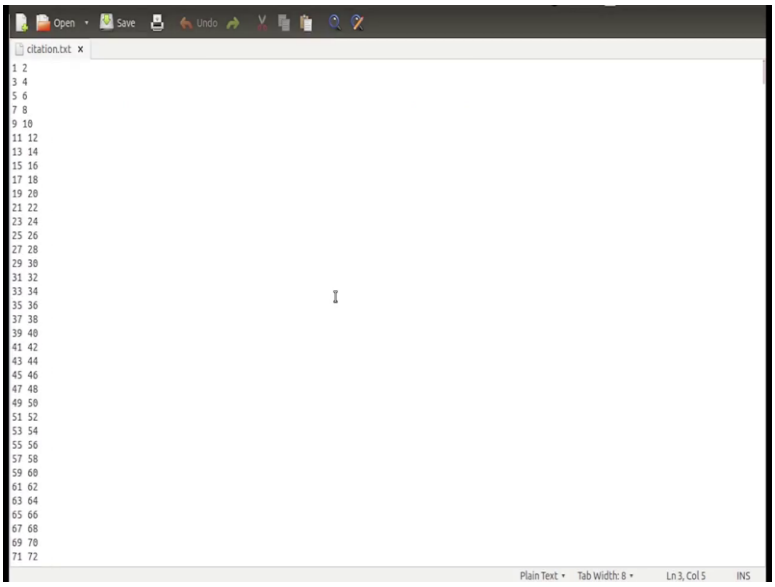
Hey everyone, in this video we going to see the relationship between Degree Rank and Page Rank that is we are going to see whether they correlate with each other or not? We are going to see whether the nodes which have less degree do they have less page rank and vice versa. So, in order to check this, we are going to use citation network which is a directed graph I already have downloaded it from snap repository, so I am going to use that. So, let us start the implementation.

(Refer Slide Time: 00:34)



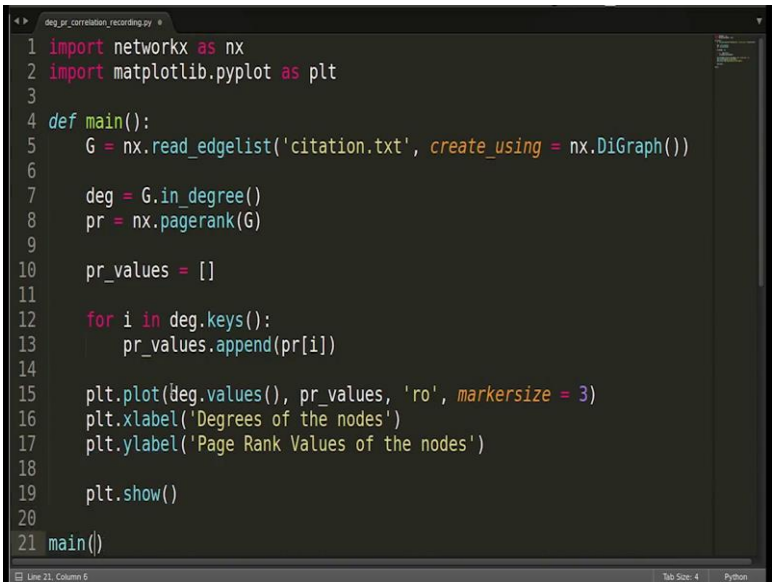
Let me show you the citation data set how it looks like?

(Refer Slide Time: 00:38)



So, this is the citation data set as you can see it is in the form of edge list. Every row contains an edge these are the IDs of the research papers. So, if you take an edge it indicates that the first paper is citing the second paper. So, the paper which there is ID 17 is citing in the paper with the ID 18. So, this is a sort of information that we have.

(Refer Slide Time: 01:04)



Let us implement this now. Since the network is in the form of a edge list we are going to use the function read edge list in order to read it into a graph for object. First, let me import networkx we are going to plot there is a so let me import matplotlib ok.

So, let us start the main function we have to read the network. So, G = nx.read_edgelist, the network name is citation.txt. Since we have to create a directed graph, we will write create using is equal to nx.DiGraph.

Now since this is a directed graph, we are going to keep a track of the in degrees of the nodes that is basically what we will compare with the page rank. So, we have to check whether a node which has more in degree that is a greater number of in links. Does it have more page rank is well or not and vice versa.

So, let us first get all the in degrees we can use a function G.in_degree for that purpose. Now what this function return is a dictionary. So, this is going to be a dictionary where the keys are going to the nodes and the values are going to be the in_degree values of these nodes.

Similarly, in order to compute the pr we are going to use the inbuilt functions from networkx. We will write the page rank = nx.pagerank(G), again page rank function will return a dictionary where the keys are the nodes and the values are the page rank values of these nodes.

Since these two are two different dictionaries and we must plot them. We basically we are going to keep the in_degree values on the x axis and the corresponding page rank values on the y axis. Since there is a two different dictionaries let us keep one of them in the order of the other.

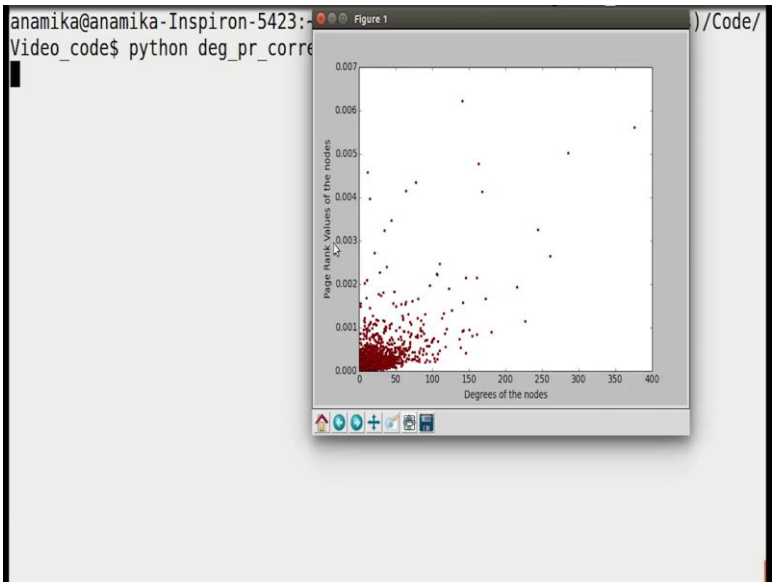
So, I am going to create a list to keep the page rank values in the same order in which degree values are coming. So, I am creating this list page rank values is equal to a list. So, I am going to create this list which will have the page rank values of the nodes in the same order in which they are pairing in this degree dictionary ok.

So, for that I am starting a loop in deg.keys. So, deg.keys will give the keys in the sequence, then we will append the page rank values pr_values.append. What do we have to append? We have to append the page rank value of the corresponding node i ok. So, we will write pr[i] ok, so we got the both the things in the same sequence and then we can plot them.

So, I will write plt.plot on the x axis we have in degrees which are available in this dictionary. So, we will write deg.values corresponding to these degree values what are the page rank values is what is available in pr_values list that we just created. Let us display it in red circles; we can also change the marker size ok.

Let us also give the labels, degrees of the nodes and we can also give the y label on y axis we have page ranks, page rank values ok. After that we have to show we will write plt.show sorry ok. Then we will call this main function they should work. Let us see how it works?

(Refer Slide Time: 05:55)



As you can see in the plot degree and page rank values are not correlated with each other. Because if they if they had been correlated, they would have been linear sort of plot over here which is not there. And, you can see that there are few nodes which have high in degree; however, very less page rank.

For example, if you take this node it has high in degree, but if you look at its page rank which is less. In the context of citation network, it indicates that this is a research paper which is being cited by several other papers. However, those papers are not so important right that is why this paper got less page rank.

Similarly, if you look at the other side you have a few nodes which have very less in degree; however, extremely high page rank. Now in the context of citation network, it indicates that these are the research papers which are being cited by very fewer other papers. However, those papers are important themselves that is those papers are being cited by a lot of other papers, or those papers are being cited by a lot of important papers. So, the conclusion is that in degree and page rank values are not correlated with each other.