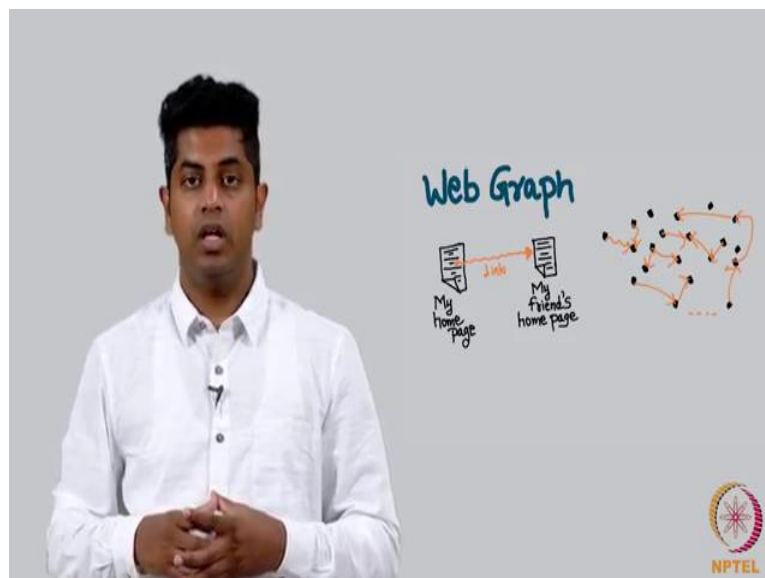


Social Networks
Prof. S. R. S. Iyengar
Department of Computer Science
Indian Institute of Technology, Ropar

Lecture – 17
Handling Real-world Network Datasets
Web Graph

So, let us look at this seemingly useless example, I say useless because it is appearingly useless, but it is extremely useful, I have been highlighting this fact from almost the first lecture. The example is this. It is called the network of the World Wide Web also called the web graph.

(Refer Slide Time: 00:24)



So, what I do is the following, I have my home page and I link the homepage of my friend on my homepage. I say for more details about these scores, please refer to my friend's website here and that here is a hyperlink to the other web page. We have discussed this before if you remember. So, now, if I look at the entire World Wide Web it is a bunch of pages; simply a bunch of pages, is it not. So, what I do is I go to a page, call it a node in my graph, in my graph every node corresponds to a page.

So, take 2 nodes means take 2 web pages, put an edge between them if one points to the other. So, my home page is a node, my friend's home page is another node, I put an edge from my page to my friend's home page saying A; my homepage is adjacent to B; my

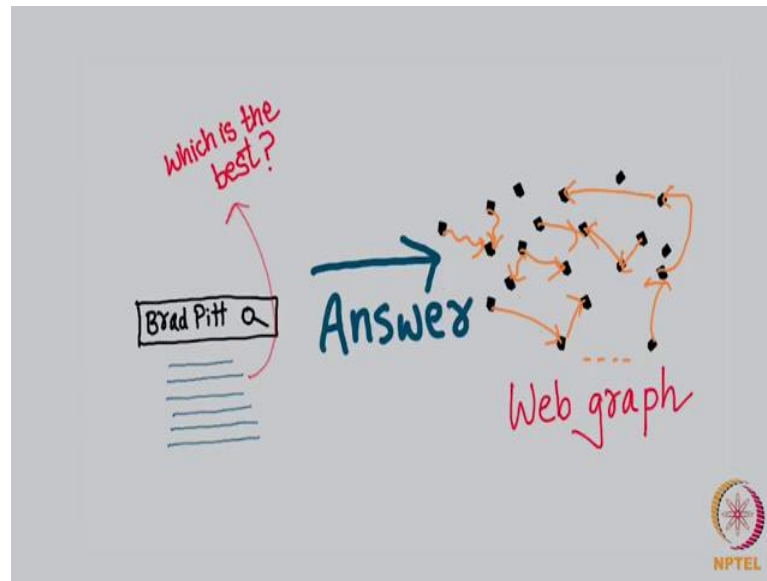
friend's home page and here is the edge. Assume I do this to all possible pages in the world. How does the graph look like? Very weirdly collecting such a graph looks like complete waste of time of what use would such a graph be it does not sound like it is going to be of any use, but as I told you people this resulted in a multibillion dollar industry which today we call google.

(Refer Slide Time: 01:42)



So, google basically harness the fact, but deep within this connections is the clue to the answer; answer to this question as to how can one get the best search result. If someone types a celebrity's name; let us say a Brad Pitt, it has a whole lot of their whole lot of pages on the internet which has the word Brad Pitt which is the best possible hit amongst these; the answer is in this web graph and google unlisted it.

(Refer Slide Time: 02:22)



So, now here is an example of a network dataset called the web graph which is appearing of no use, but a very small process of mining, it gives you an answer to a huge question, but people did not realize is a that important in the 90s, but in the late 90s to Larry Page and (Refer Time: 02:50) came out with this beautiful algorithm with says this is where one can search for an answer for the question, how does one rank once search keywords once you type in a keyword whatever you get how do you rank the output of your search.