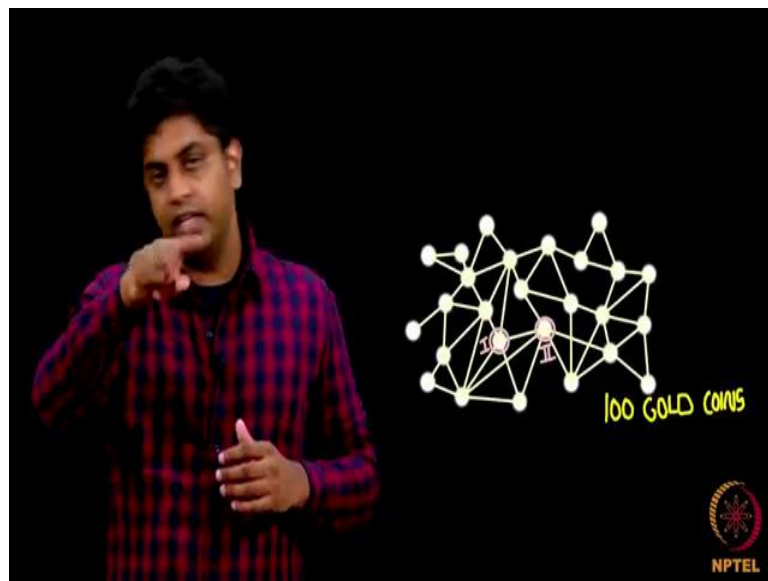


Social Networks
Prof. S. R. S. Iyengar
Department of Computer Science
Indian Institute of Technology, Ropar
Link Analysis

Lecture – 79
Google Page Ranking Using Web Graph

Let us now, get back to the web graph question, 2 questions that we observed. We asked rather, how do you collect the web graph? How are you going to use it to solve your search problem? I am going to collect the web graph by taking random walks I collected. As you are collecting, as you are visiting nodes through random walks, you give points to each node.

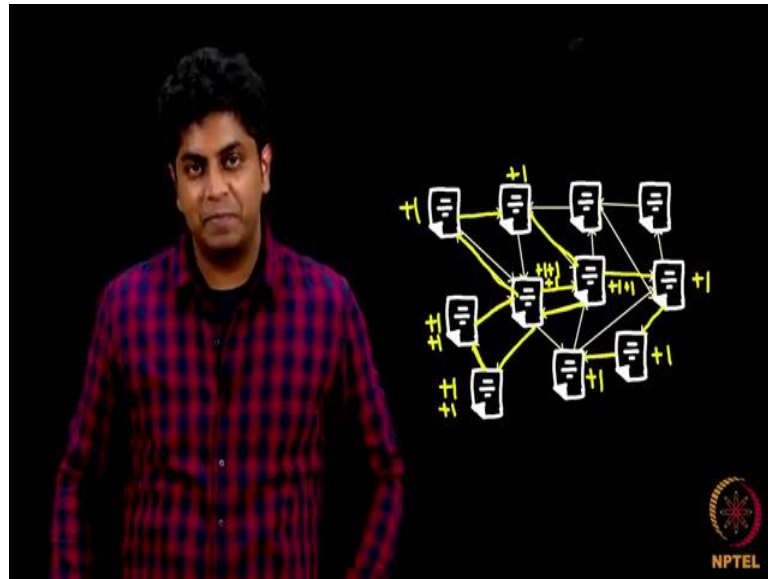
(Refer Slide Time: 00:36)



For example, on this 30-node network which we did in let us say, second game the random coin dropping game. We dropped the coins and we saw which node got the highest gold coins.

On the web graph we are going to do the same thing. We are going to walk on this web graph.

(Refer Slide Time: 00:53)



We keep walking, keep walking, keep walking, we walk like let us say billion times and keep dropping the coins and we look at which node has the highest coins. And when you search for a keyword in the search engine what Google does, is this. It firstly, keeps walking on the web that is called crawling; it keeps crawling on the web keeps dropping the gold coins and sees which node has the highest gold coin.

Every single page has basically, the gold coin accumulation points ok. When you type in let us say a keyword namely ah, I will say Britney Spears; if you type in the keyword Britney Spears we all know she is very popular, very popular pop singer. There is a there are lot of pages which talk about Britney Spears, which page should come first? What Google will do is it will go and retrieve all those pages, which has anything to do with Britney Spears. You will have some 10000 pages maybe, even more and as an example let me say 10000 pages.

Now, Google does not display in fact, it displays all the 10000 pages, but it should display the most relevant page in the beginning. What is the most relevant page? That page, that has accumulated the maximum gold coins; this is how Google works. Google just displays all these 10000 results in descending order of the points that page is accumulate as Google crawls the web and drops a gold coin, on each page that it visits. Very simple, very straight forward, very easy to understand algorithm, very easy to implement technique it works like a charm.

Google harnessed the potential of what is called a web graph, where web pages are related by hyperlinks and Google as an industry thought one should as a before it became Google, in fact in fact, story is very interesting, these 2 people Larry Page and Sergey Brin, they were sitting in a old garage and they were trying to implement this; this particular algorithm of crawling through the web. In fact, their crawling was also very suspicious that people thought that they are trying to hack something and things like that, in fact, they gone into trouble because of this. You can read up the literature for this.

I mean articles on this, in fact, there is a nice documentary made on this as well. I strongly suggest that you watch it. So, these people were doing something out of the box, which was like crawling the entire web and then giving points to each node and using that to give you the right search results. Work like a charm something that is unexpected which is the web graph seemingly useless of what use you think a person like me and you, remember when we had this coffee table discussion in 1990, do not you think? By no means one could think of seemingly useless ah data like the web graph coming in use in a problem such as search.

And that seemingly useless data made this industry a big billion-dollar industry for a simple reason that it avoided hiring thousands of people which is like the first thought that one will get when it comes to solving the search problem. All that they did was take the web graph. How easy is to take? It is easy to take just crawl, take random walk; you will get all the web graph. Not only will you get as you crawl, as you take random walk, drop a gold coin and see which page has how many gold coins. And then when someone asks for a keyword search and show all the pages which has this keyword but sort them according to the gold coins accumulated by this web pages and that is precisely what Google does.