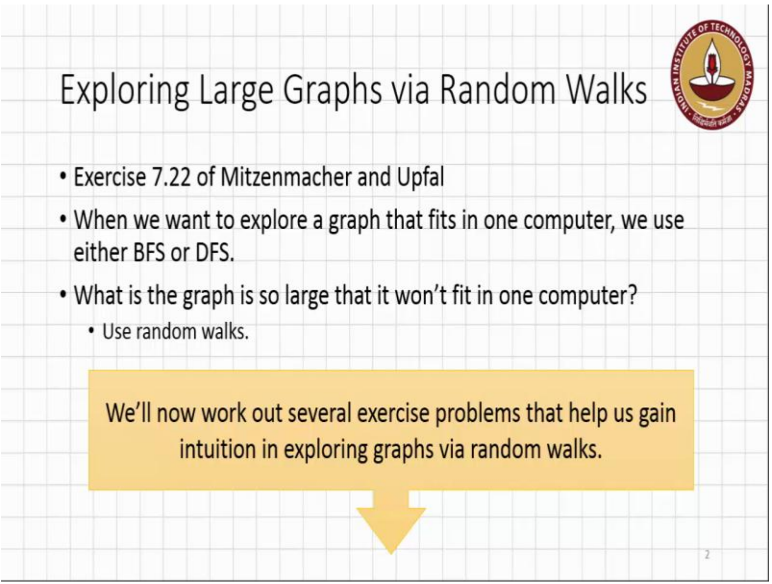
Algorithms for Big Data
Prof. John Ebenezer Augustine
Department of Computer Science and Engineering
Indian Institute of Technology, Madras

Week - 5
Lecture – 32
Exploring Graphs via Random Walks

(Refer Slide Time: 00:22)

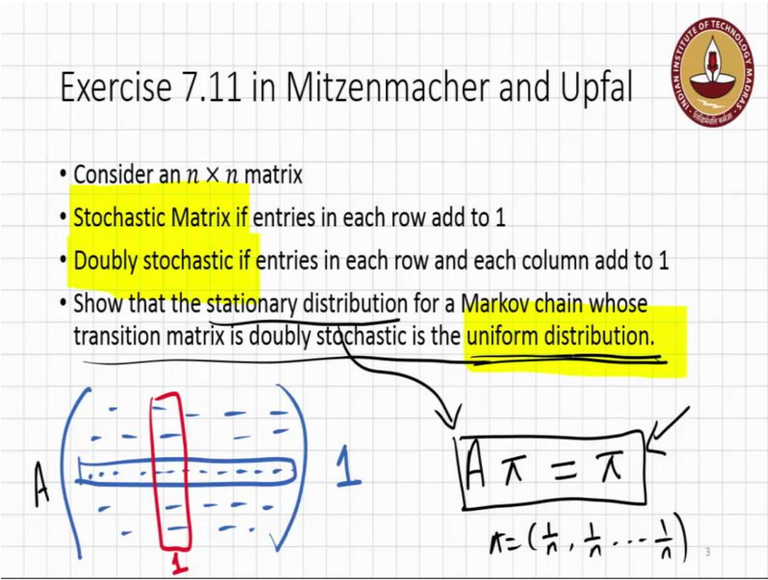


Let us now explore some problems based on Random Walks. Random walks (Refer Time: 00:20) being extremely useful because they are a very lightweight way to explore graphs. You know, when we have a graph that fits into one computer, the entire graph is into the memory of one computer. We can DFS or BFS - depth for search, breadth for search. But when a graph is really large and it has to, it does not fit into one computer, but actually requires an entire data centre to hold that graph, you know we encounter such graphs for example, the entire web is stored in the data centres of Google, for example.

So, in situations like that where there is a very large graph and you need to explore that graph, DFS or BFS is not a viable option. We will need a more lightweight option and one of the techniques used to explore such very large graphs is random walks. And in general, random walks are useful to explore large objects. So, we want to gain a good understanding of how random walks work.

So, what we are going to do today is look at several problem based on chapter 7 of the Mitzenmacher and Upfal are the textbooks that we have been following so far and we will explore these problems to gain a good intuition for how random walks work in exploring large graphs.

(Refer Slide Time: 01:56)



Let us start with the, with the few warm up questions. So, here is the question. Consider an n cross n matrix. Now, we call the matrix that stochastic matrix if entries in each row add up to 1. So, for example, if you have a matrix, an n cross n matrix and it takes any arbitrary row and if the entries in that row add up to 1 and if this is true for every row, then we call such a matrix is stochastic matrix. Now, in addition to the rows adding up to a 1, if each column were also to add up to 1 in that case and if this must be in particular true for all columns in that case, then the graph is called a doubly stochastic, again the matrix is called Doubly stochastic matrix.

And you may recall from our lessons on Markov chain and random walks, that matrices are used to represent transitions in a Markov chain, right. So, let us consider a Markov chain where such a graph, a doubly stochastic graph is the transition matrix. The claim is that such a Markov chain is going to have a uniform distribution and so this is a question that I would like you to think about why is that the case.

So, to kind of get you started on trying to establish this let us recall what we mean by the uniform distribution. What do we mean by that? Well, if this is the transition matrix, so let us say this is the transition matrix A then, what we have is for a uniform, for stationary distribution say pi, A times pi should equal to pi itself. So, this is the requirement for the stationary distribution.

So, what we need to show is that the uniform distribution is in fact, the stationary distribution. So, what we will need to show is that if you set pi equal to 1 over n, 1 over n and so on up to 1 over n and in then this condition should hold. So, that is a thing that you need to verify to show that when the transition matrix is a doubly stochastic matrix, then the Markov chain will have uniform distribution. So, take a few moments to work that out.

So, that will be your first sort of warm up exercise and of course, let me remind you of course, that transition matrix form a key component of our understanding of Markov chains and Random Walks. So, it is important to get a good intuition on stationary distribution and transition matrices.