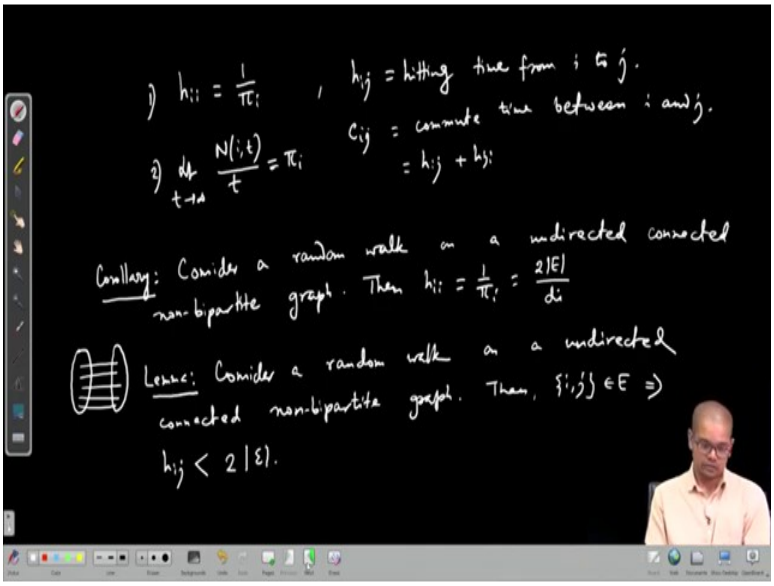
Selected Topics in Algorithm
Prof. Palash Dey
Department of Computer Science and Engineering
Indian Institute of Technology – Kharagpur

Lecture – 26
Monte Carlo Method, Hitting Time, Cover Time

Welcome, so, in the last class we have stated some parameters of Markov chain like cover time, mixing time and heating time and so on. And stated some fact now we will use them to derive some randomized algorithms.

(Refer Slide Time: 00:56)



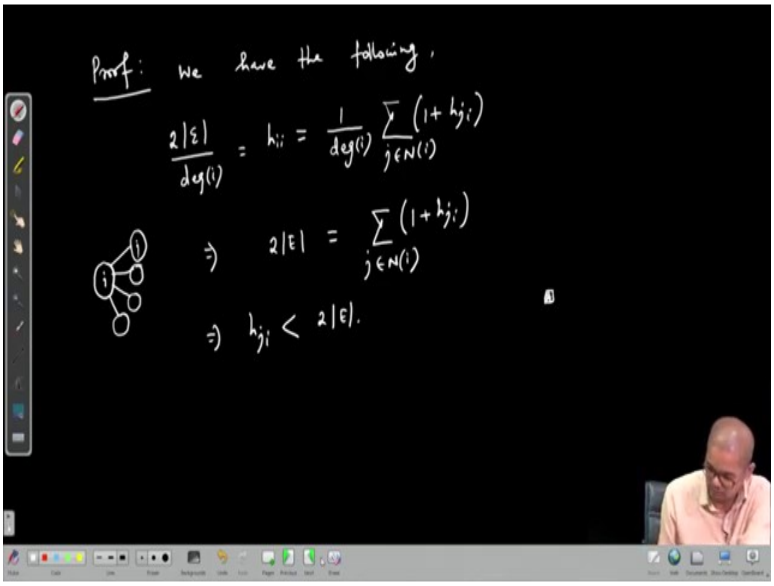
So, let us briefly recall we stated that h_{ii} it is the expected time of starting from i and come back to i. This is π_i is the and what are the other concepts? There was h_{ij}. This is heating time from i to j then there was commute time from i to j is nothing but $h_{ij}+h_{ji}$ and there was cover time. So, this is the one fact that we stated. The second fact is $\frac{\lim_{t\to\infty} N(i,t)}{t} = \pi_i$.

It is the expected it is the number of times that state i is visited in the first t steps this is equal to π_i. So now, using these facts, we can prove this important corollary and just follows immediately. Consider a random walk on undirected no connected non bipartite graph. If it is non bipartite then it is aperiodic. And let you check that if it is a graph is Markov chain is aperiodic if and only the underlying graph is non bipartite.

So, graph is called bipartite if the vertices can be partitioned into two sets, so that all edges are cross edges and these two sets are independent sets they in they induce independent sets. Then $h_{ii} = \frac{1}{\pi_i}$ and stationary distribution π is π_i is proportional degree. So, this is $2\frac{|E|}{d_i}$. Now, using this we can prove this Lemma an interesting result.

That again consider a random walk on undirected connected non bipartite graph. Then, if there is an edge between i and j, i and j belongs to E it is set. Then h_{ij} is less than twice number of edges.

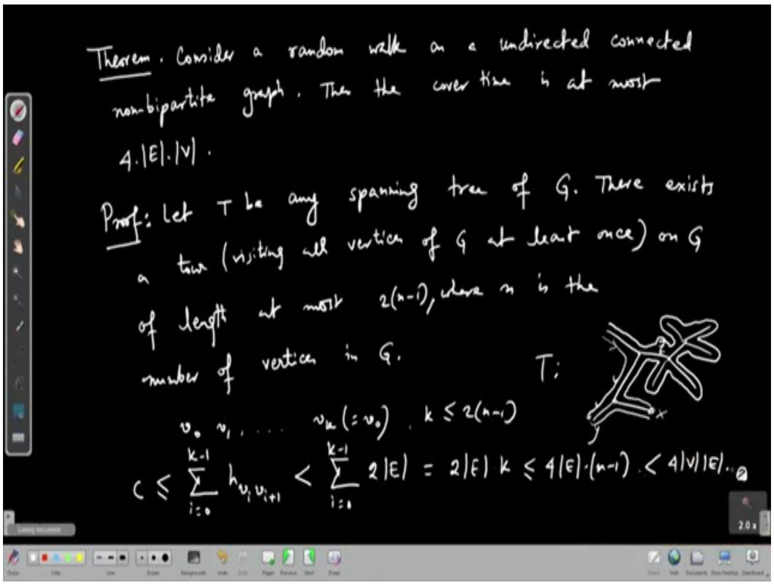
(Refer Slide Time: 05:37)



So, let us prove it, proof, so, we have the following, following chain of inequalities. So, what twice E by degree of i this is h_{ii}. And now look at i and what happens in the next step? It goes to there is no self loop, so, it goes to one of it is neighbour. So, h_{ii} is it goes to one of it is neighbour with probability degree of i and this is $\sum_{j \in N(i)} (1+h_{ji})$ neighbour of i. So, this is $2|E| = \sum_{j \in N(i)} (1+h_{ji})$.

So, this shows that h_{ji} is less than strictly less than twice number of edges which proves the result. It is very easy now using this we will design a very simple connectivity checking algorithm. So, what is the number of steps two? Visit one step. What is the cover time of a random work on a graph? Some bound.

(Refer Slide Time: 07:38)



So, let me do it in the next page, theorem, again consider random walk on an undirected connected non bipartite graph. Then, the cover time is at most four times the number of edges times number of vertices. Proof, so, for that let us take any spanning tree of G. A spanning tree is a sub graph of G, a spanning sub graph which is connected. A connected spanning sub graph which is a tree.

So that is called spanning tree a spanning sub graph which is a tree. So, let T be any spanning tree of G. So, if so then there exists a two visiting all vertices at least once, vertices of G at least once. The T be an is spanning tree and then there is a two on G of length at most twice in – 1. Basically, suppose this is a spanning tree and here is the tree. So, what we can do, if not anything, we can just simply follow this tree.

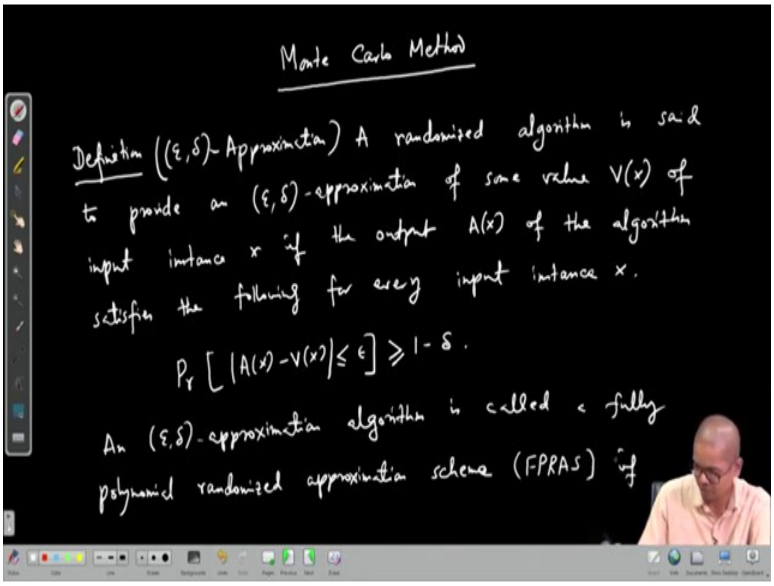
And from this node say if I want to go to this node, so, this node is visited. And so, the next unvisited node is this, x and if x is if there is no edge between y to x then at least what I can do is that from y, I go 2x via following this tree. And from this the next unvisited node is z and if there is some edge. Then we can make shortcut otherwise, I simply follow this tree and go like this.

This way all vertices are covered and if there are some edges then we can. we can shortcut it but even with this the length of this tool, the number of edges is twice the number of edges in the spanning tree which is n – 1, where n is the number of vertices in G. Now, what is the expected time? To reach to so, suppose, this tool is say $v_0, v_1, \ldots, v_k = v_0$. And $k \leq 2(n-1)$.

Now, see the cover Time is less than equal to $\sum_{i=0}^{k-1} h_{v_i, v_{i+1}}$. Now because there is an edge between v_i and v_{i+1}, this is strictly less than twice number of edges. So, this is $\sum_{i=0}^{k-1} 2|E| = 2|E|k \leq 4|E|(n-1)$.

This is less than $4|V||E|$ which concludes the proof. So, a random walk is expected to see all the vertices in $4|V||E|$ many number of steps.

(Refer Slide Time: 14:48)

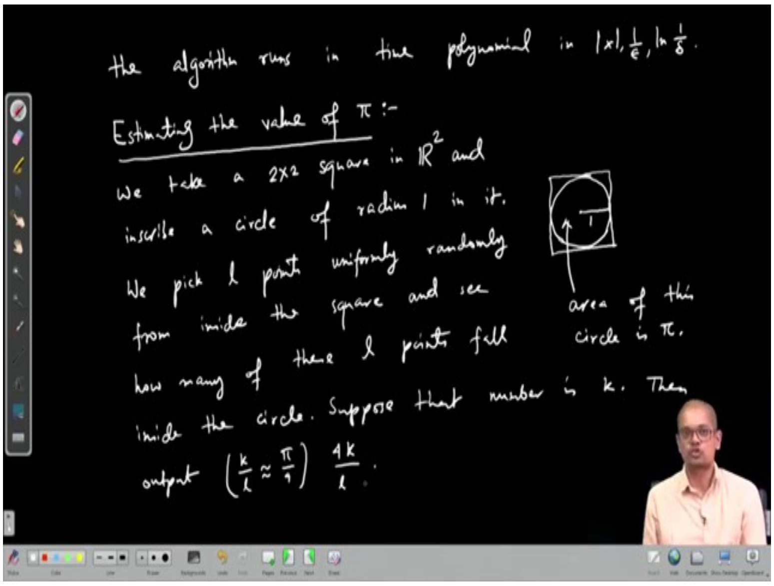


So, our next topic is Monte Carlo method. These are technique to design or to approximately count or estimate some real number to approximately count some number of objects or estimate some real number which is a Monte Carlo technique. So, using Monte Carlo technique what we will do is that we will design approximate counter. So, for that let us formally define, what are the object? (ϵ, δ) approximation.

A randomized algorithm is said to provide an (ϵ, δ) approximation of some value $V(X)$, where X is an input of input instance X. If the output $A(X)$ of the algorithm, satisfies the following for every input instance X. What is it? Probability that line $|A(X) - V(X)| \leq \epsilon$. That means $V(X)$ is ϵ close to $A(X)$ is ϵ close to $V(X)$. This should be the case with probability at least $1-\delta$.

And (ϵ, δ) approximation algorithm is called fully polynomial randomized approximation scheme, in short FPRAS.

(Refer Slide Time: 19:45)



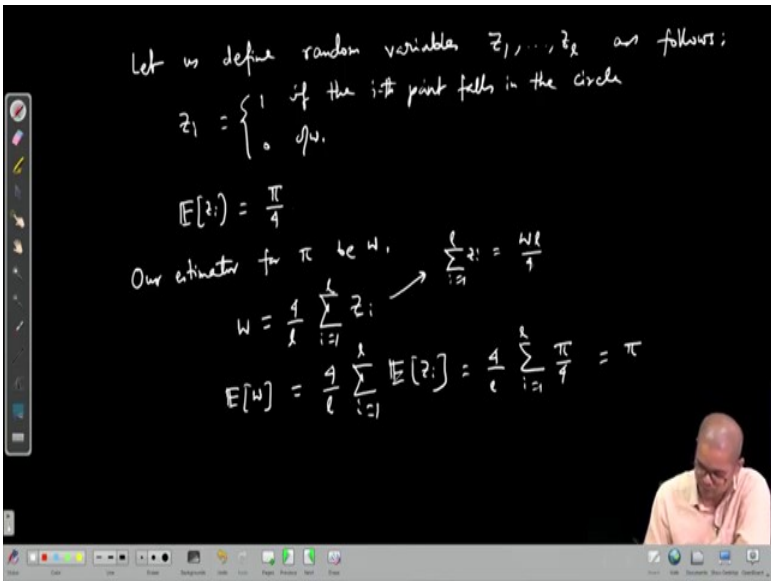
If the algorithm runs in time polynomial in size of the input $\frac{1}{\epsilon}$ and $\log\left(\frac{1}{\delta}\right)$. Now, what we will do is that using Monte Carlo method, we will design an FPRAS for some of the problem. To warm up so, our first problem is the easy problem is estimating the value of π. So, for that what we do is that we take a circle of radius 1 and enclose it in a square of side 2.

We take a 2 cross 2 square in $\mathbb{R}^2$ and inscribe a circle of radius 1 in it. So, the area of this circle is π. So, estimating π is equivalent to estimating the area of this circle. Now, next, what we do is that we draw or we pick l points uniformly randomly from inside the square and see how many of them fall within the circle. And see how many of these l points fall inside the circle.

Suppose that number is k then output. So, what is the probability that a point uniformly drawn from the square falls into the circle. The area of the circle is for π, whereas the area of the square is 4. So, this $\frac{\pi}{4}$ and $\frac{\pi}{4}$ should be close to $\frac{k}{l}$. That is the hope and if $\frac{k}{l}=\frac{\pi}{4}$ then $\pi=\frac{4k}{l}$ this is the probability this should be close to $\frac{\pi}{4}$ so, output.

So, π is like $\frac{4k}{l}$. Now, the question is, what is the error probability? And how big the l should be?

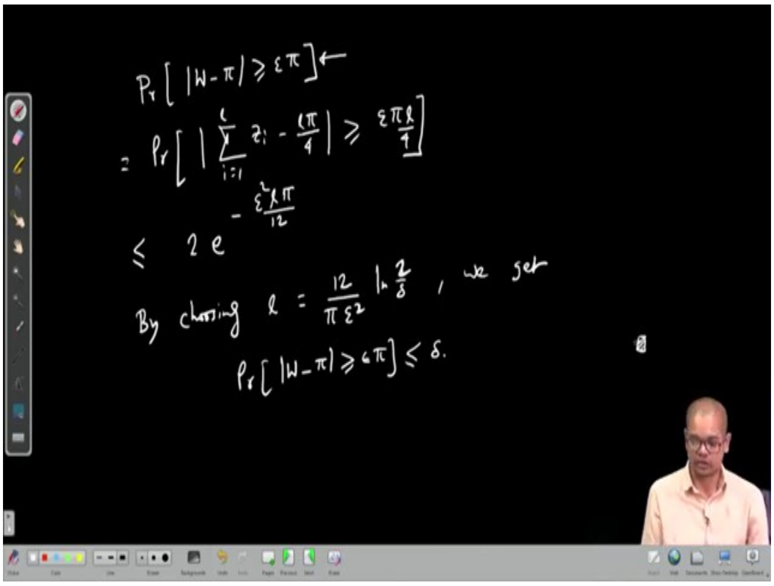
(Refer Slide Time: 26:08)



So that the error margin is tolerable. So, towards that let us define random variable $Z_1, \ldots, Z_l$ many random variables as follows $Z_i = 1$ if the i-th point falls in the circle and 0 otherwise. So, Z_i is the indicator random variable for the event that i-th point falls in the circle. So, expectation of Z_i is the probability that the i-th point falls in the circle and that is $\frac{\pi}{4}$. So, our estimator for π be W.

So, what is W? $W = \frac{4}{l} \sum_{i=1}^{l} Z_i$ this is the number of points that put that fall in the circle and this is multiplied by $\frac{4}{l}$. So, $E[W] = \frac{4}{l} \sum_{i=1}^{l} E[Z_i]$. Expectation of Z_i is $\frac{\pi}{4}$ this is π. So, our output is an unbiased estimate of π but how close our output is?

(Refer Slide Time: 28:58)



What is the probability that it deviates we make an error of epsilon? Probability that $|W-\pi|$ is greater than equal to $\epsilon\pi$. We make a multiplicative ϵ multiplicative error of ϵ. This is probability that $\sum_{i=1}^{l} Z_i$. Now, $\frac{\sum_{i=1}^{l} Z_i}{W}$ is $\frac{4\pi}{l}$. So, if we multiply this inside with $\frac{l}{4}$.

Then we will have $\frac{Wl}{4}$ this is what you get. And now, this is from chart of bound because there is no sum of 0 1 random variable. This is $2e^{-\epsilon^2 l \frac{\pi}{12}}$. $\epsilon^2 \frac{\mu}{3}$ and here is a 4 there so, 4 times 3 is 12. So, by choosing suppose we want to make this error probability less than delta. So, choose $l = \frac{12}{\pi} \epsilon^2 \ln\left(\frac{1}{\delta}\right)$.

Then we get probability that W deviates from π by more than $\epsilon\pi$ is less than equal to δ. This is what we want in FPRAS for epsilon delta approximation. So, we will stop here today.