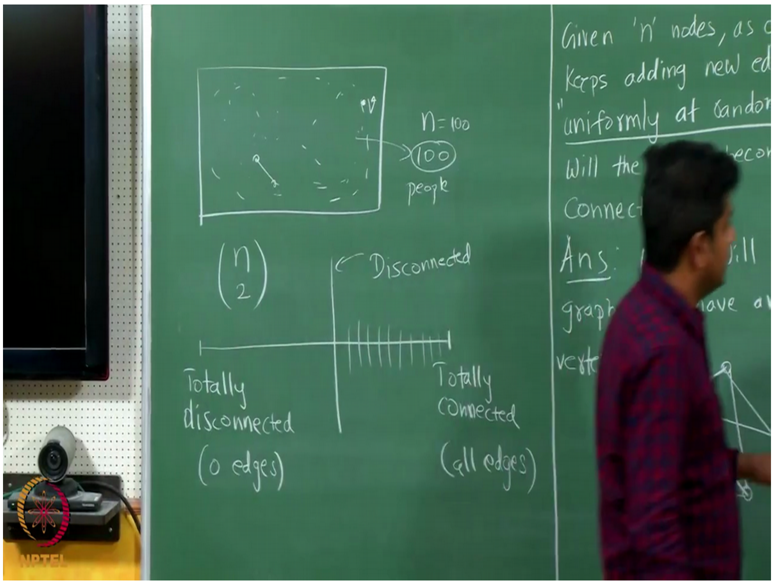
Social Networks
Prof. S. R. S. Iyengar
Department of Computer Science
Indian Institute of Technology, Ropar

Lecture – 24
Handling Real-world Network Datasets
Advanced Material: Emergence of Connectedness

Let us see something nice and deep with some mathematical analysis, let us continue with our question of when exactly do we see a connected graph as we keep putting edges.

(Refer Slide Time: 00:21)



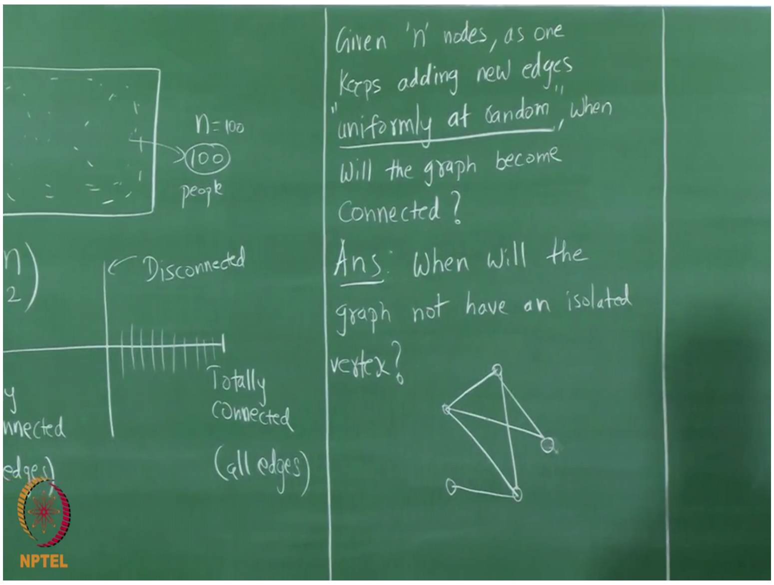
So, the question is this, the same question that we been discussing so far, assume there are 100 people or let us say n people n is equal to 100 here, a particular example there are n people here, if you put all the edges what is all the edges? N choose 2 or let us say 100 choose 2, if you put all the edges; obviously the graph becomes connected.

Now, when you put all the edges and remove just 1 edge it remains connected. If you keep plugging out edges 1 at a time from this complete graph of n choose to edges, there is a time when it becomes disconnected. I repeat take 100 nodes put all possible edges between any 2 nodes, which happens to be 100 choose to and from this you start removing 1 edge at a time remove and then throw it, keep doing this you will finally, get a disconnected graph somewhere at some point. The question is the investigation of when exactly you will get disconnected graph as you keep removing edges.

So, complete graph no completely disconnected graph; complete graph means all edges completely disconnected graph means absolutely no edges, as you keep removing edges you come closer and closer to this, there is a point when the graph becomes disconnected right. So, I am going to use the word totally disconnected which stands for no edges at all and then I am going to say totally connected which means it is a complete graph. 0 edges no edge at all, all edges enchoose to edges.

As you keep removing edges you will observe that there is a point when the graph becomes disconnected, what is that point should you remove half the edges or only a few edges when removed will make the graph disconnected or should you remove a lot of edges? This question is same as take no edges keep putting 1 edge uniformly at random, when exactly will the graph become connected. So, we will write this down for clarity sake, given n nodes as 1 keeps adding new edges.

(Refer Slide Time: 03:26).



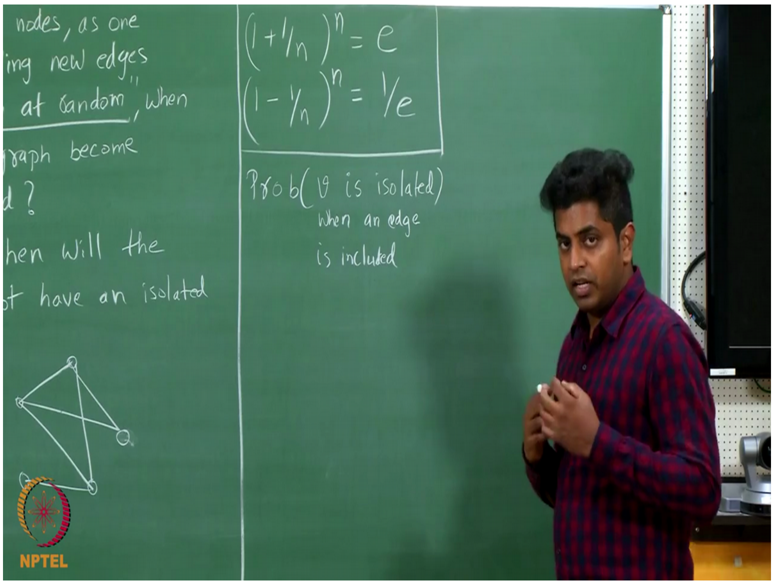
I use the word uniformly at random by this I mean, I do not have any preference for the edges I keep putting edges uniformly at random randomly basically. As I keep putting adding edges uniformly at random, when will the graph become connected? Now a question like this does not have an answer, we always say mostly if you put so many edges becomes connected; whatever mean what do you mean this? If you say I need to lose weight, I need to lose 5 kilo grams of weight, you will say then you start running half an hour a day for the next 1 month, this is just guess work and you will say an average if you run half an hour day in let us say 1 month time, you probably lose 2 to 3 kilograms of weight I am not a dietitian I am not a exercise expert, I am just telling you an example right.

Similarly, I am asking this question as you keep putting edges what do you guess? Will be the number of edges it takes for you to make the graph become connected. Let us answer this question with a small observation. I will answer this question by asking this question when will the graph not have an isolated vertex. To answer this question I am asking another question, I will ask you this question as you keep putting edges when is the time you do not have even a single point without edges, what do you mean by that?

So, let me take 1 2 3 4 5 vertices here, as you know how many possible edges can put; little bit of observation tells you 10 edges totally. Let me put 1 edge when I put 1 edge this is isolated, this is isolated, this is isolated, this was isolated does not continue to be isolated because you have put an edge. Now I will put another edge let us say this there are 2 isolated vertices now, still 2 isolated vertices only 1 isolated vertices vertex only 1, none there are no isolated vertices right now correct? After putting 1 2 3 4 5 6 edges isolated vertices came to an end, but then I should be putting this edges uniformly at random, I should not try to put edges just in order to make the graph appear connected without any isolated edges.

Please note you cannot have isolated vertex sorry; you cannot you can continue to have an isolated vertex by simply putting all possible edges within the remaining vertices, not touching our vertex that is all possible. But the question is uniformly at random when you start putting edges, when does the graph become connected? Let us answer this question.

(Refer Slide Time: 07:46)



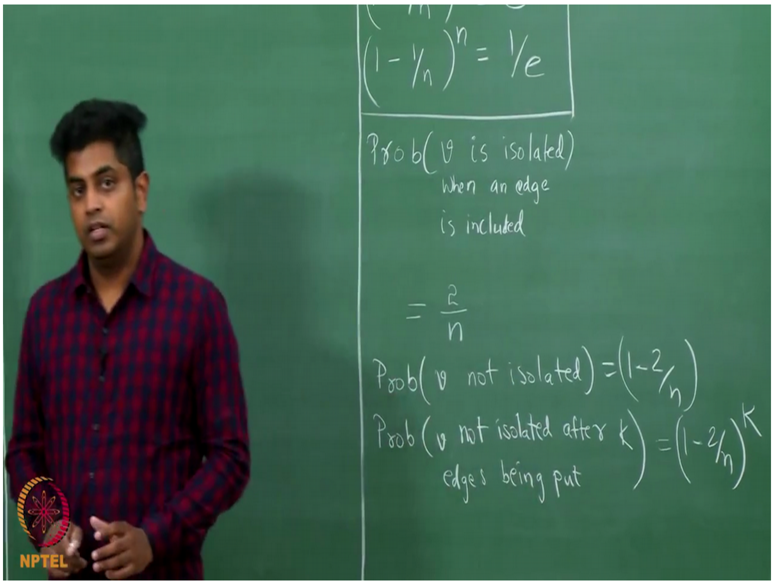
So, we have all seen this result which says 1 plus 1 over n whole to the n, is as n become bigger and bigger, this quantity is between 2 and 3 this is called e we have all seen this, it is some basic calculus limits basically. As n increases 1 plus 1 by n whole to the n becomes e for example, if n is 10, 1 plus 1 by 10 which is 1.1 to the power of 10, 1.1 to the power of 10 is around 2.7 which is e value. 1 plus 1 by 20 whole to the power 20 is roughly equal to a number between 2 and 3.

So, it is observed that this is the limit of this is a constant. Little bit of observation says we have all observed this are 11 12th standard mathematics, this is 1 over e do not worry much if you we recollect this things or I am trying to say is 1 plus 1 by n to the whole n as n becomes bigger and bigger becomes a quantity between 2 and 3 which is e, 1 minus 1 by n whole to the n let us say n is 10 then this becomes 1 minus 1 by 10 which is 0.9, 0.9 to the power of 10 is 1 over 2.7. As n is very very bigger it becomes very very accurate now we are going to use only this 2 things ok.

What is the probability let us get back to the question, what is the probability of you having a vertex v becoming isolated? When an edge is included there are n nodes you put 1 edge and v a particular vertex v is isolated, what is the probability what you I mean by this? Look at this, here are 100 nodes this is my vertex v, and we put 1 edge what is a probability that v continues to be isolated when you put 1 edge. In this big a graph when you put an edge where there no edges are put an edge, it is very unlikely that this edge will fall on v, it will be some alpha and beta some 2 vertices it may not be b v, what is the probability that is v?

There are 100 people v is one of them, you choose 2 people to put an edge, what is a probability that it is v? It is 2 by 100 simple observation, it may be pause and thing for a minute I am going to I am assuming that you know that is 2 by 100. So, when I take when I put an edge and that edge not being on v on a graph with n vertices is 2 divided by n.

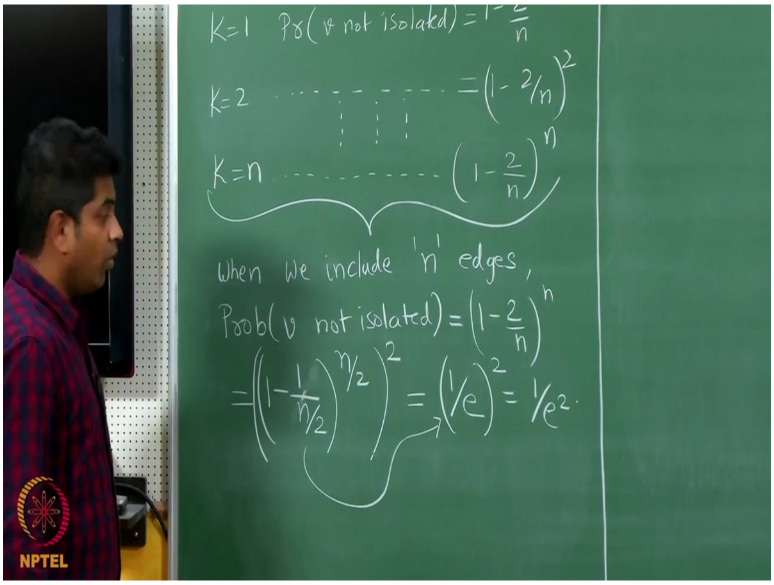
(Refer Slide Time: 11:01).



As I keep doing this you know the probability of me the probability of v having rain today if it let us say 1 by 4, the probability of not raining today is 3 by 4 correct. So, let me say probability of v not being isolated is 1 minus 2 by n, when you put 1 edge. You see probability of it raining today if its 1 by 4, probability of it not raining today is 3 by 4. The probably the probability of not raining today and tomorrow is 3 by 4 into 3 by 4, the probability of not seeing any kind of rain happening today tomorrow day after is simply product of 3 by 4, 3 by 4, 3 by 4 again straight forward concept of probability right.

So, probability of v not being isolated after K edges being put is, 1 minus 2 by n, times 1 minus 2 by n, times 1 minus 2 by n so on K times right so on K times. So, we saw that the probability is 1 minus 2 by n whole to the K, the power probability of v not being isolated after K edges being put.

(Refer Slide Time: 13:06)

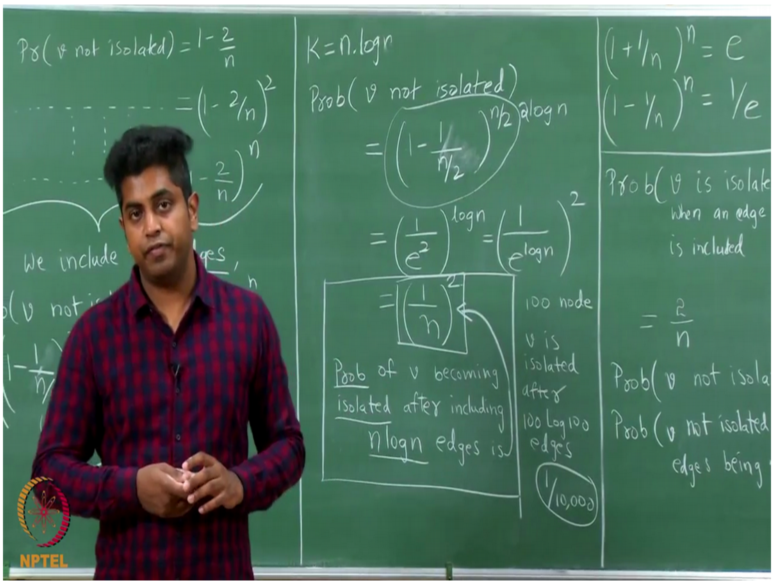


Now, let me ask this question when K is equal to 1, the probability of v not being isolated is 1 minus 2 by n whole to the K, K is 1 correct. If K equals 2 this will be 1 minus 2 by n whole square. If K is 3 it will be q, if let us say K is equal to n then this is going to be 1 minus 2 by n whole to the n, observe this let me observe this what am I saying. When v include n edges which means K equals n, the probability of v not being isolated is 1 minus 2 by n whole to the n, which is 1 minus 1 divided by n by 2, 2 goes down and becomes n by 2, whole to the power of n I will do a small find tuning here.

I divide this by 2 and multiply it by 2, usual trick we do in mathematics. If something does not look very nice you make it appear nice by multiply something on both sides multiplying and dividing by the same number adding something on LHS and RHS we know those tricks. So, I am writing it. So, that this appears like sometime that we know already, in place of n we have n by 2 1 minus some 1 over something to the power of the same thing is 1 by. So, this is 1 by e which means is equal to 1 over e this e is 1 by e, 1 over e square, what I mean by this? E is a number between of 2 and 3 ok.

So, let me say 2 e is 2, if e is 2 then the probability of e is not being isolated is 1 by 4; is a quarter probability that pause and think what I am saying there is a quarter probability 1 over let us say 2 square or 2.7 square actually whatever a constant, there is roughly quarter probability that when you put n edges on n vertices you put n edges, there is some decent probability 1 by 4, but your v is not isolated it is a very small probability, but then I will do a small trick, I will take K equals n times log n ok.

(Refer Slide Time: 16:21)



Do not ask me why I am taking n equal to log n, just for fun you can take whatever you want alright there is a reason I am taking this way, but for the time being assume I am taking slightly more than n which is n log n, if K equal to log n the probability of v not being isolated happens to be.

Let me rewrite this 1 minus 1 over n by 2 this thing, instead of n I put n log n here, n log n here, n log n here and here this becomes same thing I repeat 1 divided by n by 2 whole to the n by 2 times log n, we know this part is we saw its e correct. So, there is a 2 here, 2 here this becomes 1 over e square times log n, which is 1 over e to the log n let us say whole square I just take a log here and put 2 that side, which is equal to 1 over we know e to the log n is n 1 over n the whole square. Let me take a minute and stare at this and observe what just happens. The probability of a vertex v becoming isolated after including n log n edges is this. I repeat the probability for vertex v becoming isolated after including n log n edges is 1 over n square; which means if you take a graph with 100 vertices and if you only put 100 log 100 edges in that graph, then you will find a vertex v isolated with 1 over 100 square probability. So, by that I mean 100 node network v is isolated after putting 100 log 100 edges, the probability for this to happen is 1 over 10,000 is what my observation says.

My observation says when you take 100 nodes a vertex v is isolated with very small probability if you put 100 log 100 edges, which means when you take a graph g with n nodes you take a graph g with n nodes by just putting n log n edges, with a very high probability you will not have any isolated vertices; not having isolated vertices does not guarantee you of a connected graph, but then not having an even isolated vertex, but the graph is disconnected is very unlikely.

I told you before correct you cannot have 2 components, you cannot even have 1 disconnected vertex because it is very small probability. So, this tells you that as you keep putting more and more edges very soon the graph becomes connected; when you put n log n edges the graph with a very high probability does not have any isolated vertices, does not have any isolated components there is no isolated vertex having an isolated isolated component is very unlike correct. With a very high probability the graph is connected when you put just n log n edges.

So, we saw if 1 puts only n log n edges in a graph, the graph becomes connected. You see what is n log n given a graph with let us say 1 million nodes, 1 million times log 1 million so many edges suffices to make the graph become connected. Point to note log n means number of digits in n. So, you are not multiplying n by a big factor here when I say n log n; by n log n I mean something almost as good as n. So, if you have n number of vertices by putting just a few edges as many as the vertices itself, n times log n you achieve connectedness.

Now, we saw the theoretical framework for it, do not worry much about all the proof that I gave you it is not very important I just gave it for completeness sake. All that you need to know is the following let me go slowly, all that you need to know is this. Given n nodes you keep putting edges you keep putting n log n edges, then the graph becomes connected; how true is this statement. Whatever I wrote on the broad may not be true the mathematical statements I have derived is in no way a validation for what is going to happen if I am going to program and see it maybe or may not be, I do not know I am at least not convinced. So, let us do one thing, let us take our laptops and try to write piece of code and then check if it is really true or not.