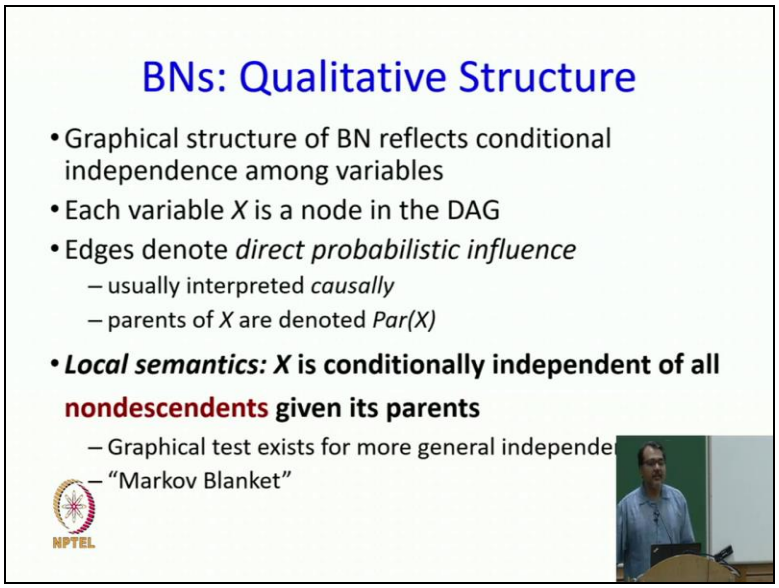
Artificial Intelligence
Prof. Mausam
Department of Computer Science and Engineering
Indian Institute of Technology - Delhi

Lecture-55
Bayesian Networks: Conditional Independences and D-Separation- Part-3

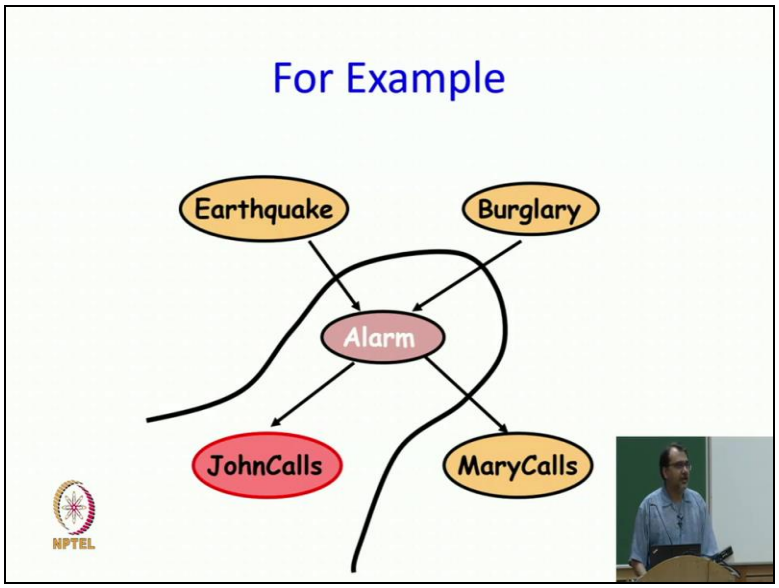
So now, what is we are going to do is we are going to use the basic network think about what are the condition Independence exemptions it is making. And how can I read the graph structure to figure out what condition Independence are being said by this?

(Refer Slide Time: 00:35)



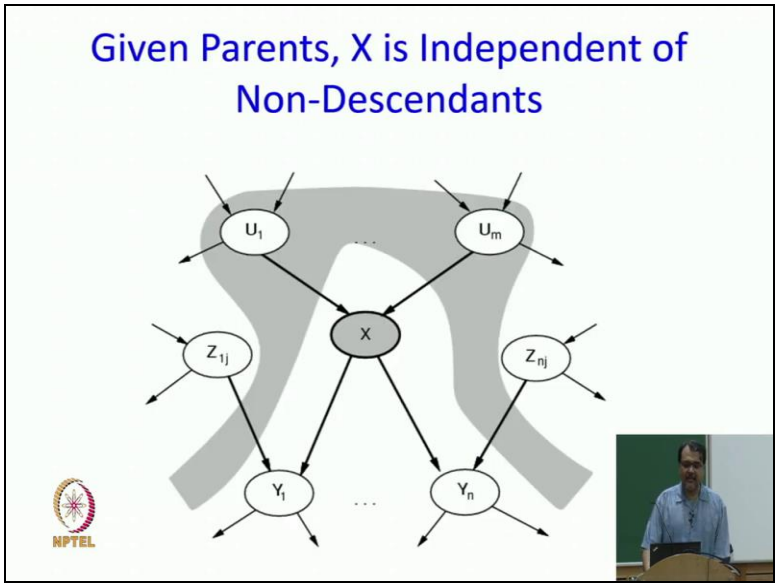
And graph structure of Bayesian network reflects some condition Independence is that the graph structure is saying ahm and there are some local semantics for it.

(Refer Slide Time: 00:45)



And I will start by asking questions ok. The first question is, is John called conditionally independent of earthquake given alarm? We have discussed this but let us do this again. Is John cal conditionally independent of earthquake given alarm? Yes. Is John call condition different of burglary given alarm? Yes. Is John cal condition independent of Mary call given alarm? So in fact, this is true, right. So, once I know alarm, John call is independent of everything else in this graph. Is there a General principle here?

(Refer Slide Time: 01:32)



And the general principle is this. Given my parents, I am independent of all non descendants, ok so in this case my parents alarm I am John calls and my non descendants are earthquake, burglary and Mary call since I have no descendants, everything else is not decided. This and this

grey thing shows that once these are all the nodes that are having cut off. Once I know my parents nothing else matters to me but my Children will matter to me.

That is also true in many families. If you want to understand about me, let us put it this way. if you knew my parents, you will be able to understand about me quite well. After that my grandfather or other people do not matter exactly true because sometimes they say that in some traits come from grandfather and grandparents that how this is not for nature this is for nurture. So, this is how my parents treated me, depends on their parents.

But once you know my parents do not need to go to their parents to understand me. They treated me the way they treated me based on who they are. And once I know who they are, then you can figure out who I am. So, in other ways, but my children still give information. Because I am treating my children and by knowing my children you will be able to understand more about me. But no other people matter.

That is what this particular theorem is saying that if a random variable parents are given, they are known to you, then the node is condition independent of all the non descendants, right. All parents must be given. Good point. In this example, right, right, right. So in this example, we are talking about X. So, we are talking about X and the parents are U1 and Um here. And X is the node here, if the parents U1 and Um are given; now it is conditionally independent of all the non descendants.

What are the non descendants? Parents of U1 and Um and their parents are Z1j conditionally independent of the Zi1, right but I am not conditionally independent of Yi which am I descendant. If I know about Yi, I also get to know about Zi that is correct. But additionally I know about Zi but do not know Yi that does not change my X. So, think about it this way. See, what is Zi? Zi is sort of my wife, my spouse because we have a child together, if you think about it in humanistic words.

So if you have not seen the children and if you only meet the Yi, this thing is saying you cannot tell about the husband. No, it is not be humanly appropriate but what is the intuition see the

intuition is that and Zi and X may both be the reasons for why Yi turned out in a certain way. If you see Yi you have to figure out ok, is it because of Zi or is it because of X.

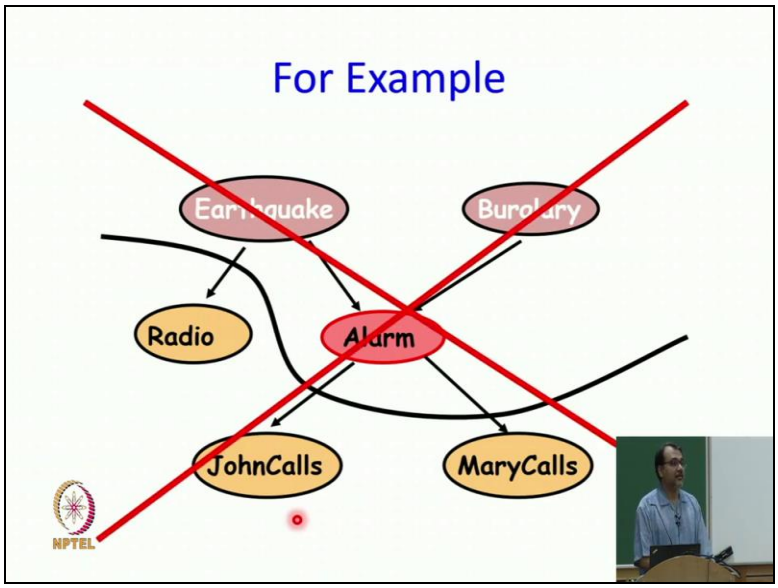
By knowing Zi and Y, I can explain Yi away by Zi. Like for example, this is exactly that case where, earthquake and burglary are the two parents of alarm. So, by knowing earthquake and alarm do I know something about burglary? But otherwise by knowing earthquake do I know something about burglary? Exactly it is the same setup, right.

Earthquake and burglary are independent of each other. But by knowing alarm they became, that they become dependent on each other. By knowing the child, husband wife become dependent on each other, because there is in the child together. Not the child may have turned out one way because of one parent or the other parent and child has turned out one way because of one parent then reduces my probability that it is because of the other parent. That is the; that is the analogy in other words.

But forget the human world, I think this specific example will make sense to you. Earthquake and Burglary are independent of each other but given alarm, earthquake and burglary are dependent on each other. All non descendants are independent given the parents. That is why No, given Yi means given that child so whether the statement does not say anything about what happens if the child is given. It is a very specific statement.

It is saying that if I give you the parents and only the parents right, then I am conditionally independent of my non descendants. Good. It is to some more examples and we will come back. Will X dependent on Zi, then it will. We will, we will come to that. Is there we will we will do this in more details. Give me a few slides and exactly going to come, ok. How about this?

(Refer Slide Time: 07:23)



Is this alarm independent of everything given earthquake and burglary. Alarm is independent of everything given earthquake and burglary. Ahm, Why? We just said that the alarm is called; a node is independent of Non descendants given the parents. Are they given the parents? Yes, alarm parents are earthquake and burglary. Are they asking only about the non descendant? We are also asking about descendants. So first of all the theorems says no.

But intuitively all said yes, I think about this. So the question you always have to ask this, by knowing this thing do I additional get information about the variable? So the question is I already know it is burglary, I know their values. By additionally knowing the value of John call does that change my probability distribution about alarm. That question you have to answer. So, let us think about it. Suppose, let us begin, we keep any value.

Say, burglary happened. Or let us take a easier example. Let us say earthquake happened? Now the probability of alarm going off is 0.2 earlier right? Now that John called. Does that increase my probability that alarm might have gone off? Earlier it was weekly, it was weekly likely that the alarm would have gone off. 0.29. But not John has also called. If john has called, obviously I would have increased my probability of alarm might have gone off.

Let us say both John and Mary called. Then the probability will become very close to one another. Therefore by additional knowing John, even though I knew earthquake and burglary, I

get more information about alarm. But, alarm which basically means that they are not conditionally independent, I mean, you can do it in two ways. You can do this intuitive thinking this that is actually incredibly important.

It is extremely important in the sense that you have to ask the question define if I also given you this variable, does that changed my evidence, does that change the information that I have got the variable at hand? But you can also apply the rule that given the parents my nodes condition dependent of all non descendants. Non descendants John calls, is non descendants. Given earthquake and burglary what is alarm condition independent of?

Just additionally in knowing about radio gives me no more information, why? Because radio can only influence alarm for radio says that I heard on radio that alarm has happened. So, what basically that says if I hear on radio, it can only influence earthquake, but we already know earthquake. Radio would have affected alarm, if I did not know the earthquake. I would really strongly encourage you to do these kinds of questions at home, because the more you are able to ask the question and this is exactly the words you use.

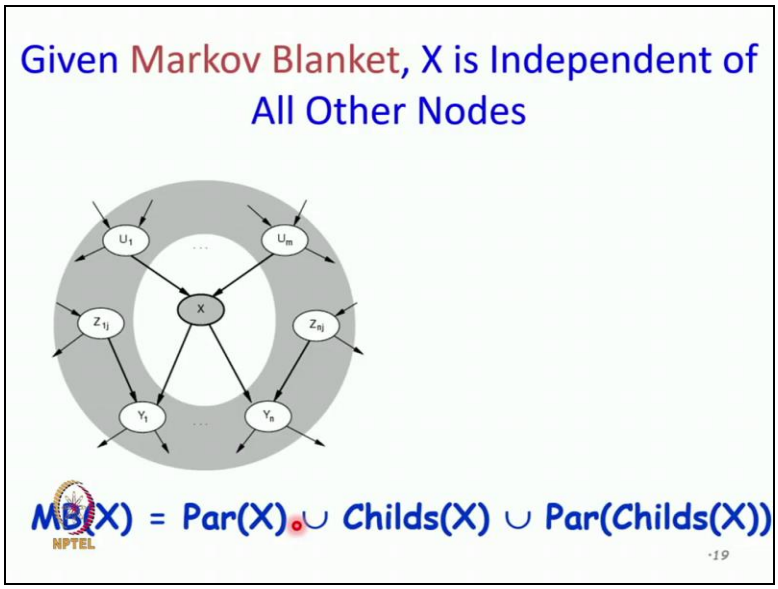
I already know earthquake and burglary, by additionally knowing about this variable radio, do I get more information about alarm? This is exactly the way you should formulate the question and then think about. The more you will think about it the more efficient clear. Yes, Happened or not happened. Are radio and John call similarly affecting alarm? The question is how are they affecting alarm? In one case, John calls is child of an alarm. In one case radio is not the child of Alarm.

They are not similarly affecting alarm. The question is if you knew about video that you knew the probability of earthquake, we already knew the earthquake. No, we do not know about alarm. So again, very important, what is given to us? We are given earthquake and burglary; those two things are given to me which basically means I do not have to estimate them. They are given I know their values. I know whether it is true or false I am interested in alarm.

I am interested in weather my information about alarm increases or not. So, I do not know alarm I am estimating alarm. And now I am asking the question that if I additionally tell you about radio, does that increase your evidence about alarm decrease or remains the same. If I additionally tell you about John calls, does that will increase the evidence about alarm, reduces or remains the same if it is increasing or reduces they are not independent.

If it remains the same they are independent. Very good this is an important stuff. So the second parameter is given Markov blanket I am independent of all the nodes in the world. Like I am independent of everything given some set of nodes which are very close to me that is my family. So, what is my family? My parents, my children and believe it or not, yes, that is true.

(Refer Slide Time: 13:32)



The other parents of my children, now, of course because this is a Bayesian network, you cannot use the word wife and husband. We can say spouses but then that reflects a different world. Not the one that we are currently living in. It is the world of 10, some 1000 years later. So, what is my Markov Blanket? Markov Blanket is my parents because they get all the evidence, I am getting from my ancestors.

Ages is my children because they give all the evidence, they gift to my descendants. And also the other parents of my children because they are also the one that are affecting my children and so that makes a collective whole. This is called the Markov Blanket. So let us ask some questions.

Given an alarm, is burglary condition independent of everything else? Given alarm is burglary conditionally independent of everything else. What is the Markov blanket of burglary? Alarm and earthquake so, therefore the answer is no.

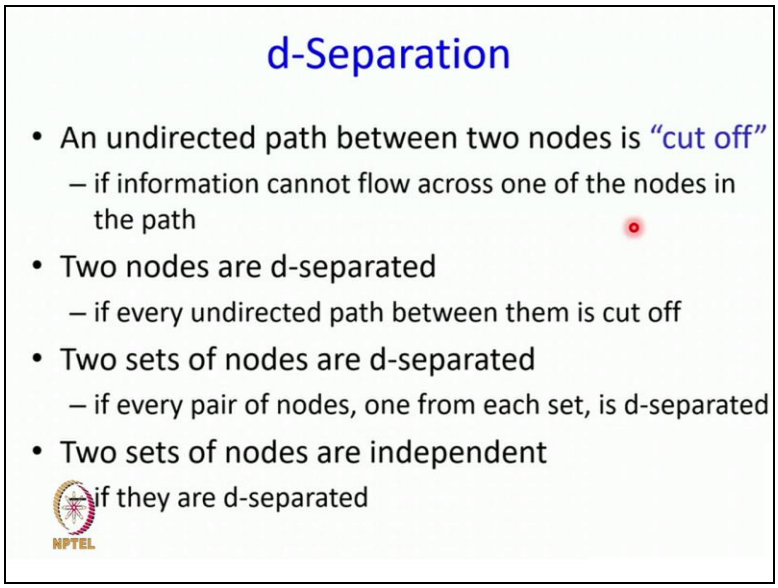
Can you tell me intuitively what is going on? Why is earthquake important because alarm could also have caused by earthquake. So, by knowing earthquake I increase my or decrease my probability of earthquake. Earthquake has happened burglary probability decreases. If earthquake has not happened burglary property increases. This however is correct. So, by knowing earthquake and alarm, burglary is conditionally independent of everything.

So, now I think it is a good time to stop. We have done 2 theorems. One theorem is what happens given parents. One theorem is what happens given Markov blanket. Of course, you might be asking the question given burglary is John calls independent of radio? You can ask that question. And neither of the terms will apply because neither burglary, the parents of John calls nor the burglary is in Markov blanket of John calls, right.

How do we generally analyze the given Bayesian Network? And that happens using a principle of D-separation. A Bayesian network structure represents several conditional Independence assumptions. The structure itself in quote some independence exemptions and one such Independence information that we did was that given my parents, me as a random variable given a random variables parents a variable is independent of all non descendants.

Non Descendants matter but not descendants are independent given the parents and second thing we discussed is given the Markov Blanket, X is independent of all other nodes and what is the definition of Markov blanket the definition of Markov blanket is parents, children and parents of children, other parents of the children so that is your closest of family. Given your closest family and nothing else matters to you. You are independent of everything. But of course these rules are applicable in specific cases.

(Refer Slide Time: 16:47)



In one case, if all your parents were given in the other case all the Markov Blanket nodes were given. You cannot make it reverse that if one part of the Markov blanket node is not given then the Independence does not hold you. You cannot say that. It is only in One Direction given the Markov blanket you are independent of everything. So, you might ask, in general, how do we test independence of a given, given set of pairs of nodes, given some other nodes.

This is a general question and answer to the general question is the concept of D-Separation. So you will have just physically follow the book follow the rules and if you are able to understand the rules, then you would be able to do it. This is very simple. It says first forget the directionality, forget the arrows. Think of a graph with undirected path. First step, an undirected path between two nodes is cutoff. Cut off means information cannot flow ok from one node to the other.

An undirected path between two nodes is cutoff information cannot flow across another to know that information cannot flow. Then, they will be, if information cannot flow it will be independent. Why because of the notional Independence. Notional independence is that two nodes are dependent adding any new information about the first does not add any new information about the second then they cannot influence each other.

Then, obviously they will be independent. So now, the next step is 2 nodes are D-separated. If every undirected path between them is off. So there are two nodes there are many many parts that connects the 2 nodes and if all of them are cut off that means information cannot flow from one node to the other, then, they will be called D-separated. They will be D-separated in this. You can extend it by saying if two sets of nodes are separated if this is a set of nodes.

If this is a set of node is separated, if every pair of node is separated. A B C nodes D E F there should not be a valid information pass from A to D, A to E, A to F, B to D, B to E, B to F and so on. So, all pairs of them should be in 32. And 2 sets of nodes are independent if they are D-separated right because information cannot flow. So, that is the general definition that you have to look at what are the ways in which a node can influence another node?

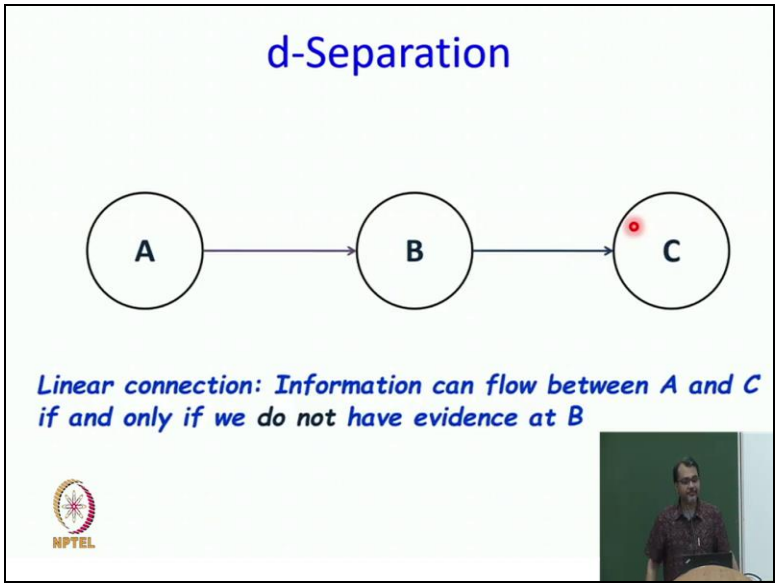
And if you are not able to find any such way then there will be independent to each other and they are D-separated. However, there are 3 rules on when is a path considered cutoff and when is the path not considered cutoff and that actually requires directionality ok. So, there are three rules to check if an undirected path between the two nodes is cutoff or not. If a path in two nodes is cut off or not and one of them is called the linear connection.

Suppose I want to ask you the question between burglary, alarm and John calls, now look at the structure that they create. They create a linear structure, burglary is the parent of alarm, alarm is the parent of John. Now, the question you ask is in this linear structure when can information flow from burglary to John call. How can it flow? It can flow only through alarm. If I give you an alarm, if I give you alarm are they independent of each other? We can answer it in 2 different ways.

We can say let us say the alarm is given to me, by adding information about burglary does it change my distribution about whether John is going to call or not. Now John just listens to the alarm and makes the call whether John knows the cause of the alarm or not literally does not matter. By knowing Burglary, I will not get a change in the distribution of John calls. Another way of saying this is given my parents; I am conditionally independent of all non descendants.

So which node should I apply this? John calls let us apply this to John calls. What are the parents of John calls alarm, given my parents, that means given alarm. I am conditionally independent of all the non descendants. What are the descendants of John call? So it is conditional independent of all the John calls given the alarm. However, you see it. That is fine.

(Refer Slide Time: 21:52)



Basically you can recognize that if this particular linear structure A goes to B, B goes to C information can flow between A and C only if you do not know B, because for information, information to flow from A, it needs to go through B, right. It influences B because B changes, so, that influences C therefore C changes. If you already know B, the path is blocked off it is cutoff and so information from A cannot go to B. Any questions on the 1st case, ok? Everybody ok?

Let us look at second case. The second case is what is called the diverging connection? And for that let us look at the specific example. Let us look at the triple, John calls, alarm and Mary Calls. This is what is called a diverging connection because they share the common parents. So, from one parent we have 2 diverging edges to two other nodes. ok Now we again asked the same question. How does information flow from John to Mary? How does it flow?

Let us say that John call and I know about the fact that John has called does that change my distribution about Mary calls? Without knowing anything is John has called that means likely

that the alarm went off. And if the alarm went off probability increases than Mary calling probability, so, therefore we does not have to be deterministic, but as long as you have a change in the probability it can decrease also it is all fine.

On the other hand suppose I give you alarm suppose that tell you the value of alarm, now does additional knowing the John has called changes the probability distribution of Mary calls? Because, both of them independently listen to the alarm and based on that make the decision. They are both dependent on the alarm. If I know alarm they become independent. This is what is called the diverging connection.

In this diverging connection, I have a structured B is the parents of A and C and again, we ask the question when can information flow from A to C, the information can only flow through B. I as a child tell me something about my parents and since it now know something more about the parents that say something more about another. And if I give you B then this passes cutoff right, any questions, on diverging connection ok?

Now becomes the more interesting connection and that is called the converging connection. And for converging connection I will look at earthquake, alarm and burglary as my example. So, in this case alarm, is the common child of both earthquake and burglary. Now let us ask the question when can earthquake influence burglary? It is a very interesting question; some of few might have figured it out for others, think about it. I will ask you 2 questions.

Question Number one: Is earthquake independent of burglary? Question number 2: Is earthquake conditionally independent of burglary given alarm. You can answer in intuitively I do not mind. Or you can use another theorem that we have studied or you can think about the fact that this is earthquake and burglary we are talking about. You can do whatever you want. But let us talk about the first question, so is earthquake independent of burglary?

Earthquake is independent of burglary let us just think in simple terms, If I tell you that is earthquake that does not tell me that your house is been burgled as long as we do not model inside that there is earthquake and therefore is friendly people live in earth have gone crazy and

suddenly some people can make advantage of that, then, we have not use that particular model. Some model of earthquake in dependent and burglary is independent now, interesting question.

Is Earthquake conditionally independent of burglary given alarm, there is somebody not serious everybody says no that is pretty impressive. Let us think about it. Alarm is given to you; let us say there was a alarm. You can do it in both ways as long as influence is one way it is fine. Let us say there was an alarm. The question you have to ask is by knowing there is earthquake does that change my probability distribution about burglary and vice versa. Any one question is fine.

They are equivalent questions. Suppose I tell you there was earthquake, what does that tell you about burglary? The probability has decreased, there was alarm, something must have happened oh, earthquake has happened. Therefore burglary is already happened. This is the chain of reasoning that is going on in your mind. Again these are probabilistic statement with the point is by knowing that one of the parent is the cause, of myself, the Other parents become less likely to this.

In other words, this is a phenomenon called explaining away. It is a technical term is basically says that it has many cores, many ways of explaining something, I use one to explain away the phenomenon and the others are less likely of explanation. So, in this particular case, look what is happening in the converging connection. Earthquake and burglary are not independent given alarm, right and this we need to generalize and so I will make it one more step complicated is earthquake independent of burglary given John calls.

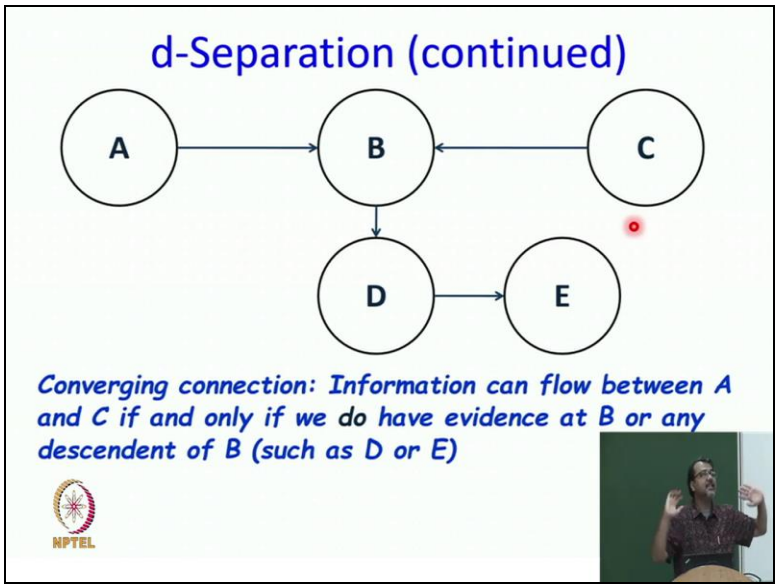
Again let us think intuitively, John calls. By knowing earthquake does that changed my probability distribution of burglary? It has, it is not independent, why because John had called it must have been an alarm at least that increases my probability of alarm, because the probability of alarm has increases, I am interested in explaining that out. Even Earthquake has happened I have now explained it now, my burglary probability goes up.

So, therefore, by knowing John has called, now the earthquake and burglary have become independent. This is one interesting phenomenon, where both the random variables, you know,

are independent of each other on priority. Earthquake was living in its own life looking at the earth's you know the Rock movement and causing tremors. Burglars are living in their own life. And were just looking at doors which are locked and then who have elderly in there happily making their burglary.

And suddenly by knowing alarm, they have become dependent on each other. By knowing alarm because they are two independent causes of the same effect, because such an effect may have occurred they have now become dependent. So, this is called the converging connection. And the rule for converging connection is slightly tricky.

(Refer Slide Time: 30:07)



So we will look at ABC, A and B are both parents of B and A and C are both parents of B. But B has its own children also that is D E and C so on. The converging connection says contrary to diverging connection and linear connection says that information can flow from A to C if we do have evidence. Either in B or in any child A, because the children of B eventually change the probability distribution of B and now that I have more information of B information can flow from A to C.

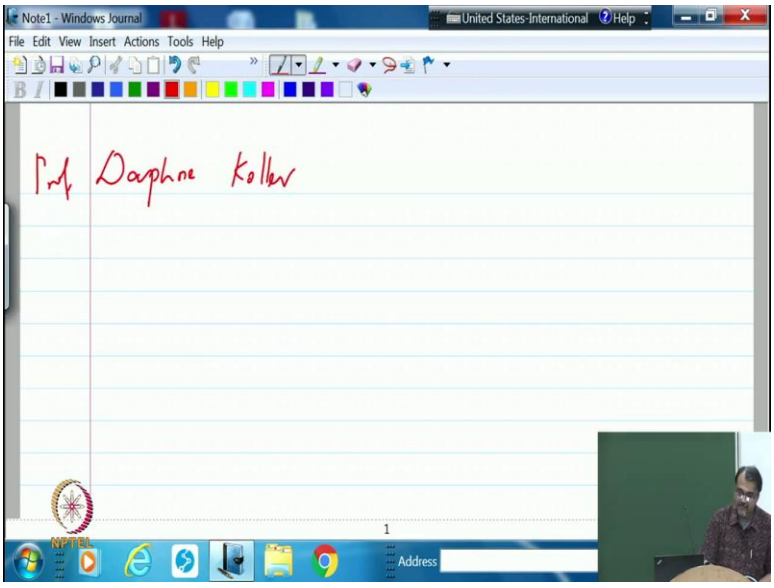
So, A and C are not cut off they were originally cut off and they become not cut off if I know B in any child of B. And again if you get confused go back to these set of rules. This is very very, crisp. I will ask you if a set of node X1 versus set of nodes X2 are independent of each other

given evidence E, so then you look at each pair of nodes look at all parts between the pair of nodes and see if information can flow.

If you find anyone pathway information can flow they are not independent of each other. And in every part that you check whether information can flow or not, either you will use the linear connection at every step you will be going down this path at every step, either it will be linear connection or it will be a converging connection or will be a diverging connection keep following those rules.

And if you find that those rules still make this path, information flow then it is not stipulated, information cannot flow you check the next path. You find no such path when information cannot flow then they are independent, ok. Try to do some practice, it is important if you think that is became too confusing to you, I will tell you the one person who explains it extremely beautifully is Daphne Koller. Have you talked about Daphne Koller earlier?

(Refer Slide Time: 32:21)



Daphne Koller is a very famous professor, who was a professor in Stanford. She is amazingly productive, prolific researcher. She has produced a large number of extremely successful PhD Students who are faculty members in their own right? She also is a co-founder of Coursera, along with Andrew Ng. Checkout her video on how to compute the influence between probabilistic graphical models.

But also if you ever get to the point where you take a full class on probabilistic graphical models, where are you teach, learn, not only Bayesian network, other kinds of graphical models like Markov networks, dynamic vision networks, conditional random fields. There are other kinds of networks, Cycle graphs, networks which are both directed and undirected at the same time, there networks which are, which handle actions which influence diagrams et cetera.

All of those come inside this big area called probabilistic graphical models, it is a probabilistic model but you are using a class structure to arrange the joint distribution and if you ever get there you will have to read a book written by Daphne Koller, bible of probabilistic graphic model. So now, that we have discussed what are the condition Independence assumptions expressed by graph structure, we will do the next important thing about vision networks. Ok.

So, now, we are done with semantics, we have done with syntax earlier now we will look at the semantics, what does the network say, what does the probabilistic graphical model say and that is called the semantics and now we will talk about Inference. That is what you will do in the next class and we can stop here.