Artificial Intelligence

Prof. P. Dasgupta

Department of Computer Science & Engineering

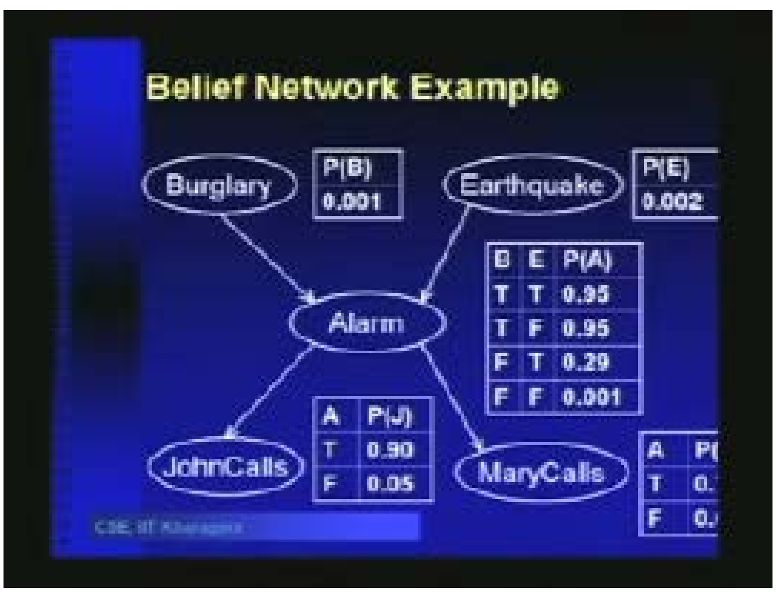
Indian Institute of Technology, Kharagpur

Lecture No - 23

Reasoning with Bayes Networks

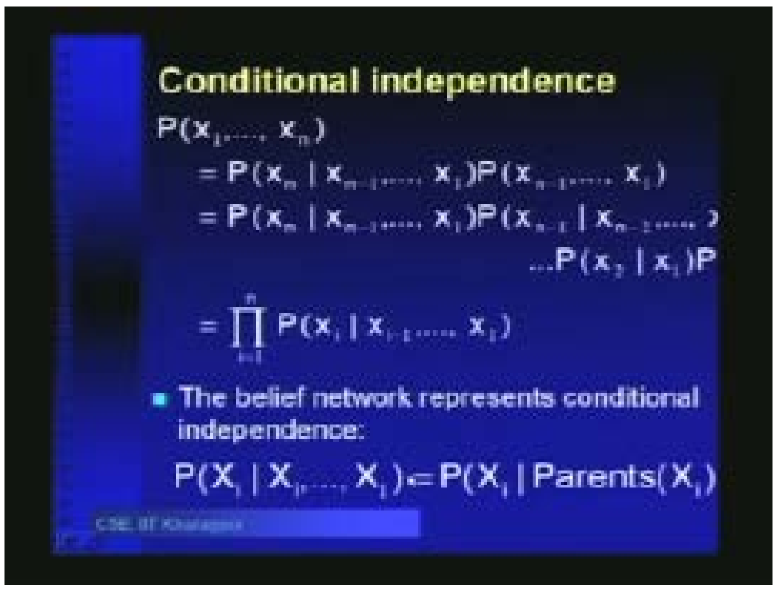
In this lecture, we are going to study reasoning with Bayes networks. In the previous lecture, we had seen how to create a Bayes network and what is the meaning of a Bayes network. Today, we are going to see how an existing Bayes network can be used to reason about different kinds of events. We will continue with our example on the burglar's alarm and try to see how to compute different kinds of probabilities values using the given network.

(Refer Slide Time: 01:26)



Just to recall, very briefly, is that what actually we have in this belief network is that we have conditional independence, in the sense that the probability that a given variable takes a given value, given the values of all the other variables which precede it in some ordering, is given by the probability, given the values of the parents. For Example, if we look here, then, the probability of John calls, given alarm and earthquake, is the same as probability of John calls, given alarm. Parents of John calls, namely burglary, earthquake, etc., do not affect John calls directly, if we are given the value of alarm. If we are given whether the alarm has rang or not.

(Refer Slide Time: 01:52)



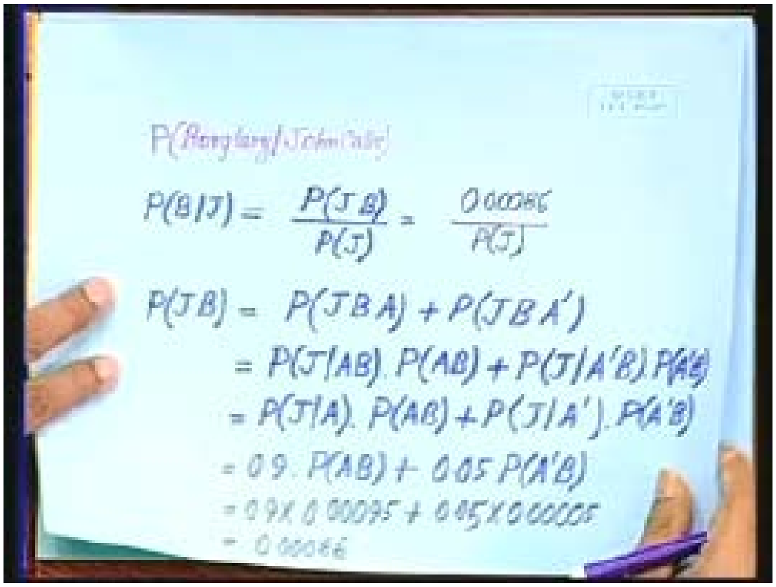
This is what is the representation of belief network or Bayes networks. Attempts to do what we shall now do, is that we will try first with some Examples, see how we can compute different kinds of probabilities with this. Then, we will study the general algorithm for computing those even probability. We will start by looking at what is called diagnostic inference from effects to causes. What we are interested in finding out is probability of burglary, given John calls. Probability- okay, I will use short forms- so, for burglary, I will use B and for John calls, I will use J, and so on. Now, the question is, how do we compute this value from the given belief network? The given belief network has only some conditional probability values.

This is how we are going to do: we will compute PB given J first as probability of J and B, so, I will write that as JB, divided by PJ. Now, let us see why we are doing this. If we go back to the belief network, then, you will see that according to this belief network, John calls is actually an effect, can be an indirect effect of the burglary, because if the burglary takes place, then, the alarm may go off and then John calls may happen. But what we are trying to compute here is probability of burglary, given John calls, so, we are given that John has called and we are interested in finding out what is the probability that actually a burglary has taken place.

This is a diagnostic inferencing, so, in order to find out these kinds of probabilities, see, in these conditional probability tables, we are given probability of John calls, given alarm probability of alarm, given burglary and so on. But here, we require it the other way around: we want probability of burglary given John calls, so, in the first step that we have done here, as you can see, that we have converted it into a form which is suitable for us. So, we have converted this using Bayes rule to PJ and B divided by PJ.

Now, how do we compute PJ and B? To compute PJ and B, we break it up into PJBA plus A prime, where A is alarm. We can always do this based on any variables, but we are doing it on A, because we know that J depends on A from the belief network structure. Now, we have to compute PJBA and PJBA dash, so, this we can compute again as PJ given AB right times PAB and the other 1 is PJ given A dash B times PA dash B.

(Refer Slide Time: 18:21)

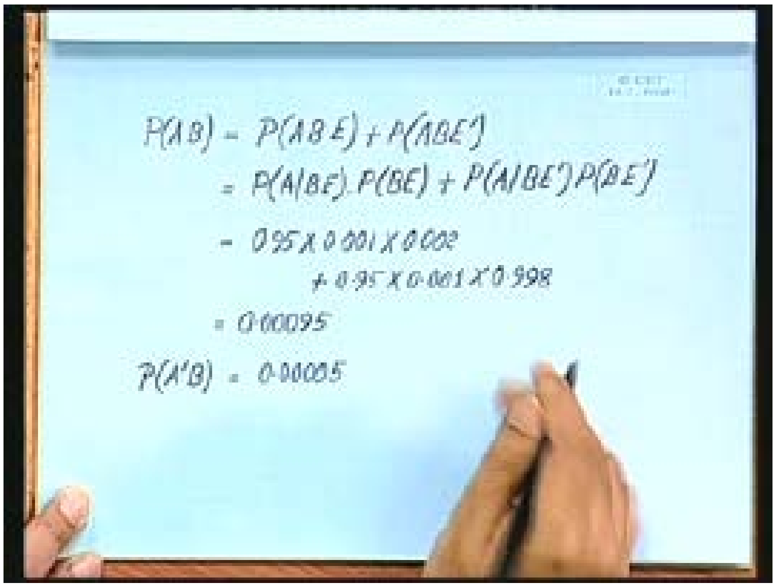


Now, PJ given AB is the same as PJ given A. Why? Because if we look at the belief network structure, then, you will see that PJ given A and burg B is the same as PJ given A, because A separates B from J. We will have this simply as PJ given A, and then, the next term is PAB plus PJ given A dash- for the same reason- B will disappear from here times PA dash B. Now, this I can get from the conditional probability table. PJ given A, if you look at the belief network, is given as 0.9, and PJ given A prime is 0.05. This is 0.9 times PAB plus 0.05 times PA dash B.

See, what we are doing is- essentially, we are going from bottom upwards, because that is the way in which it is easier to compute in the belief network, because conditional probabilities are given from bottom to in terms of the parents. Now, we still have to compute PAB and PA dash B, so, let us continue how to compute PAB. So, if we to compute PAB, we can compute that as PABE plus PABE dash. We are introducing this E, because A has 2 parents: 1 is B and the other is E.

Then again we can rewrite this as- yes. (Students speaking). PAB- (Students speaking). Yes, you can write it that way also, as what we are doing here is, we are writing it in this way, because now, what we can do is, we can just look up this thing as follows: this is going to be probability of A given BE times probability of- and B and E are independent. Similarly, this is probability of A given BE dash times probability of BE dash. Again, what we have here is this value we can get from the belief network.

(Refer Slide Time: 13:14)

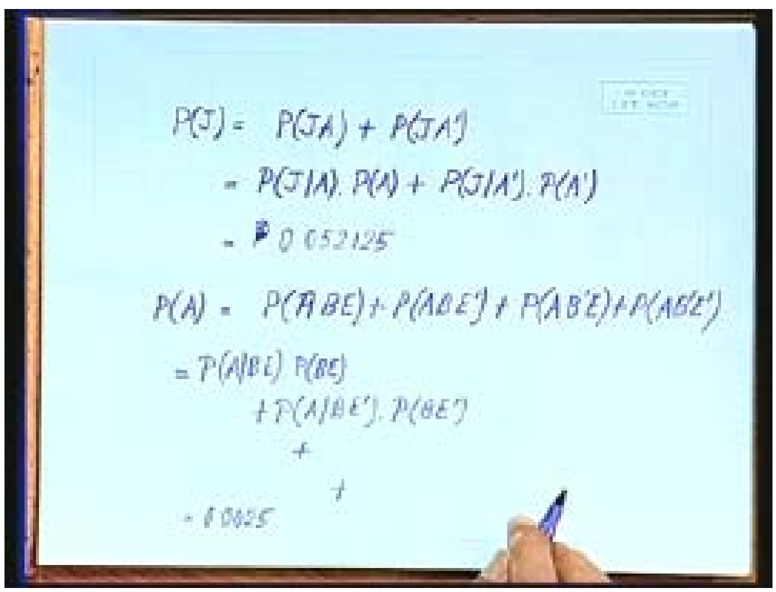


So, if we look at the belief network probability of A given B and E is 0.95 and probability of A given B and E dash is also 0.95. If you use those values and we get 0.95 here times PBE here, B and E are independent, so, that times PB times PE. PB times PE will be 0.001 times 0.002 plus, this is going to be 0.95 times probability of B, which is 0.001 plus probability of E dash, which is 0.998, and so this comes out to be 0.00095, right? Then, similarly, we have to compute PA dash B and if we use a similar calculation, exactly similar to this, then, we find that to be equal to 0.00005.

Now, if we coming back to our original thing, we now have the values of PAB and PA dash B, so, we can write this as 0.9 times PAB was 0.00095 plus 0.05 times 0.00005 and this comes to 0.00086. After doing everything here, we have succeeded in computing the value of PJB, but we still have to compute the value of PJ in order to get this. We now have the numerator, which is 0.00086, and we have to have the denominator PJ here. Now, how do we compute PJ? How can we compute PJ? We want to find out just the probability of John calls, so, PJ- yes, how do we compute this?

Again, we will break this up into PJA plus PJA dash and then, we go again similarly as PJ is broken up into PJ given A, which we can find from the conditional probability table times PA plus PJ given A dash times PA dash and then, again, PJ given A we can find from here. PA and PA dash- how do we compute PA and PA dash? Again, we just break it up into- (Students speaking)- 4 cases, right? So, we break it up into PA- (Students speaking). Yes, say, BE plus PABE dash plus PAB dash E plus PAB dash E dash. And then, we can break up each of these as PA given BE times PBE plus PA given BE dash times PBE dash and so on.

(Refer Slide Time: 18:12)



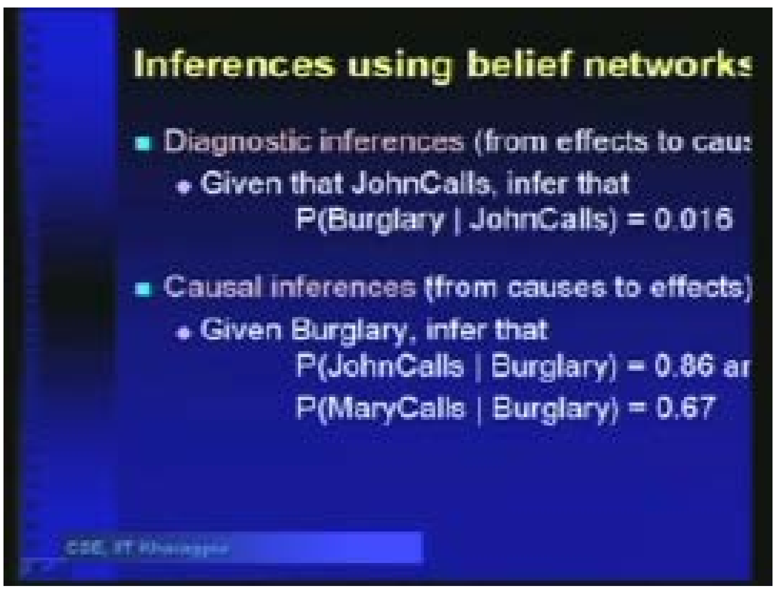
We get 4 terms like this and in each of them, the first term can be obtained from the Bayes network, because we are given the probability of alarm, given BE and BE dash and B dash E and B dash E dash and the second term in the second term B and E are independent variables. So therefore, this will split up into PB and P E and PB dash and P

E dash and we can find out those probabilities again from the belief network. If we do all this, then, finally, we will find out that this PA is going to come out to be 0.0025.

And similarly, when we compute PA dash, that we can obtain bY1 minus this thing and then, when we substitute the values back here, with the values of PJA and PJA dash, then, we get the value of PJ as 0.052125, right? Having such and then, once we have this value, then, we go back to our original thing and here, instead of PJ, we now use 0.052125 and then, this gives us 0.016.

Yes- (Students speaking). PA given B, yes, so, it will become PA, given BE plus E dash. All this analysis put together gives us this P burglary, given John calls is 0.016.

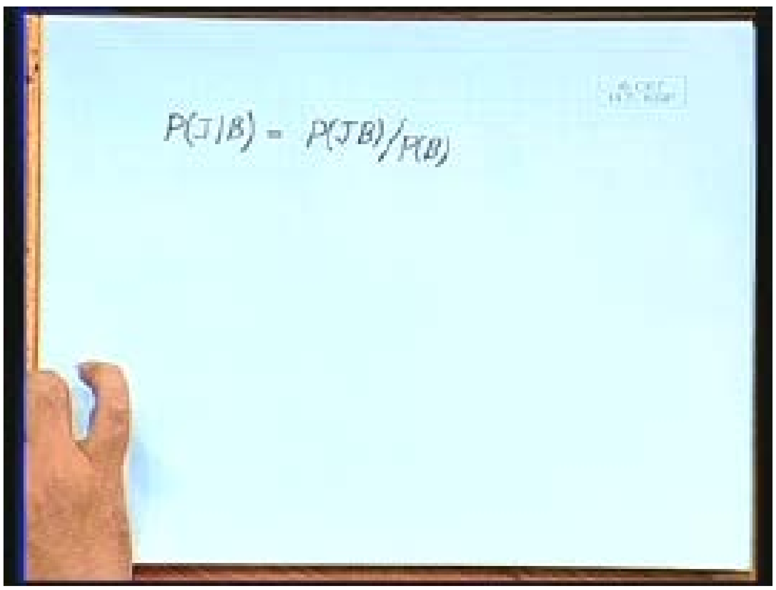
(Refer Slide Time: 19:58)



See, what essentially we are doing is, in the first step, we are converting it into a form where we have the cause rather the effect in terms of the cost. Suppose we have a variable A which influences a variable B. We are trying to recast the computation so that we compute the probability of the effect given the cause, rather than the cause given the effect. For Example, in some cases where we have casual inferences where we go from causes to effects, you will see that the computation is usually much easier.

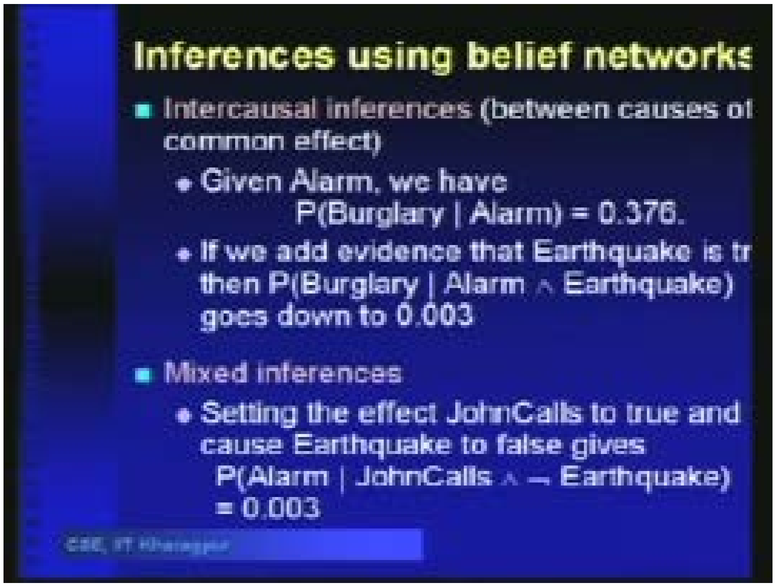
Suppose we want to compute P John calls, given burglary and P Mary calls, given burglary, so, these we have in the right shape, only thing is that John calls does not directly depend on burglary, it depends on whether the alarm has gone off or not. So, if you look at PJ given B, now, see, we previously computed PB given J. Now we are looking at PJ given B, so, this is the causal inference.

(Refer Slide Time: 21:13)



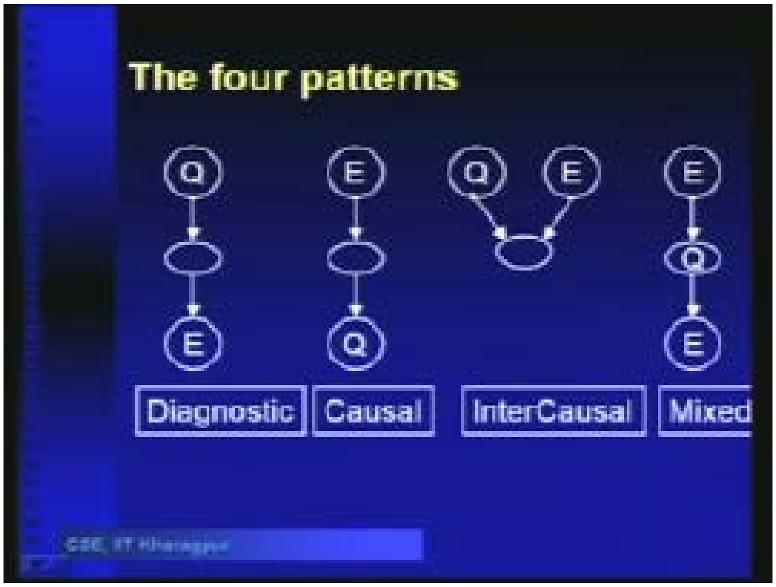
We will rewrite this as PJB divided by PB and then, again, we proceed just as previously. Similarly, for PM given B, we will do the same style. Is it clear? What about inter-causal inferences between causes of a common effect? Suppose we are looking at PB given A. We use the same style PB given A, convert it into PAB divided by A.

(Refer Slide Time: 22:28)



We are always converting it into the joint probability distributions queries and then, using the conditional independence of Bayes network to filter out certain variable and evaluate the rest. With this background, let us- okay, the complete analysis of all these different kinds of events for this particular Example is available also, in the course web page. There is pdf file there, which shows the detailed computations of each of these, so, if you are interested, you can download that and have a look at it. These are the 4 kinds of patterns that we have seen: diagnostic, causal, inter-causal and mixed, where Q is the query and E is the evidence.

(Refer Slide Time: 24:02)



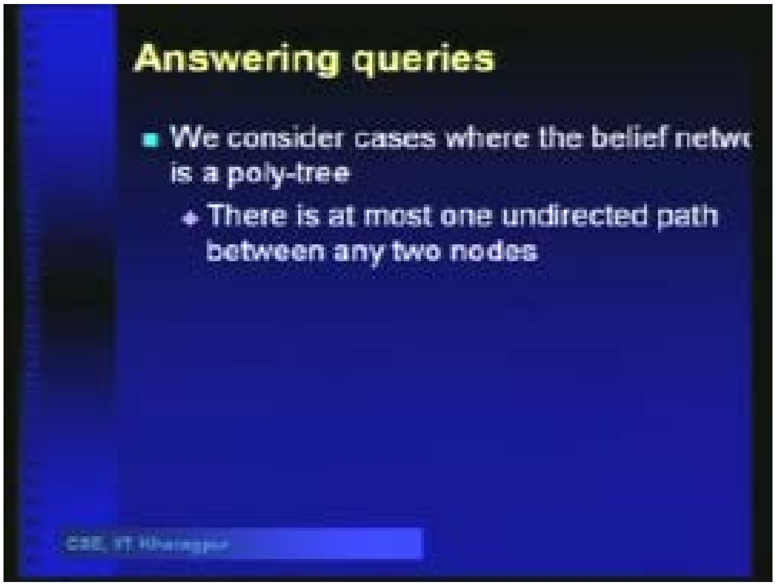
E is the evidence; E is what is given. See, in the first case, what is given to us is a successor in the belief network of the query, like, for example, here the query could be that probability of burglary given John calls. John calls is the evidence and burglary is the query that we want to determine. Then, causal is where the query is further down in the belief network. It is like probability of John calls, given burglary and this is like probability of burglary, given earthquake. (Students speaking).

P of burglary given alarm- (Students speaking). No, it will just simply be causal probability of burglary, given alarm. No, that is going to be diagnostic. Inter-causal is where you are- they are brothers; the query and the evidence are brothers, they are siblings in the network and mixed is where you have evidence at the top and evidence at the bottom and you are looking at the query in the middle. Suppose we want to ask what is the probability of the alarm, given earthquake and John calls. (Students speaking). Hmm. (Students speaking). Oh, yes.

We can- probability of John calls, given Mary calls- no, they are not independent. (Students speaking). They are not independent because if 1 calls, then, that is in increased probability that the other will also call you. See, if John has called, then, he has probably heard the alarm, so therefore, it is also increased probability that Mary also calls. Probability that John calls given Mary calls, I think it will come in mixed; it will come under mixed. (Students speaking). No, it will not come under inter-causal, inter-causal is where they are together, they are affecting something, which is, they have a common successor, but here, they have a common predecessor. (Students speaking). Probability of burglary- (Students speaking).

Yes, burglary and earthquake are independent events, so, probability of burglary, given earthquake is the same of probability of burglary. If you recollect, in the last lecture, we talked about conditional independence and when in the belief network, 2 events are conditionally independent, it also depends on what evidence you have. Like, for example, when we are talking about the car starting example, then, when you knew whether the car starts starts or not, then, the events of the radio and the petrol were no longer independent. Otherwise, they were independent.

(Refer Slide Time: 28:05)

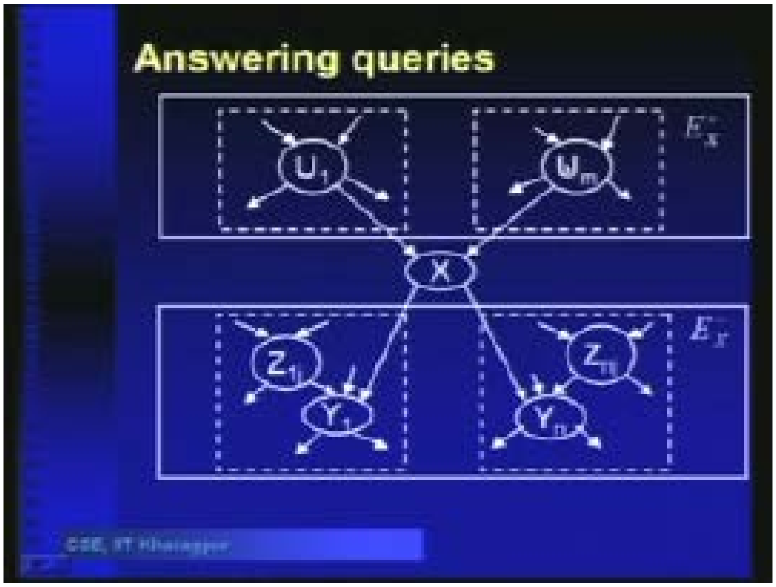


So, mixed here is going to encompass everything else. This is the general thing and this general thing is what we are going to study. These other things are just special cases of them, but the algorithm that we are going to study now is going to encompass all of these, so, it is going to work for all 4 of these cases. Now, we are moving into the algorithm and it is going to be similar in flavor to the analysis that we did here for this example. Now, we are going to consider belief networks, which are poly-trees, which means that there is utmost 1 undirected path between any 2 nodes.

The example that we saw here for the burglar's alarm is also a poly-tree, because if you take any pair of nodes, then, there is utmost 1 undirected path between the 2 nodes. This is the general view that we are going to consider, so, our query variable is X, so, we want to find out the probability of X, given a set of evidence E. What is evidence E? Evidence E is a set of events whose values are given to us. For Example, evidence could be that the alarm could be an evidence, earthquake could be an evidence, Mary calls could be an evidence.

Evidence is the given values for some of the variables in the belief network. It means that we are given the values of some of the variables in the belief network and we have to compute the probability of X, given those values. We are going to represent the set of evidence variables as E will denote the set of evidence variables and if you look at X, then, X will have some parents. These parents are called U1 through UM. Let us say that X has M parents, U1 through UM.

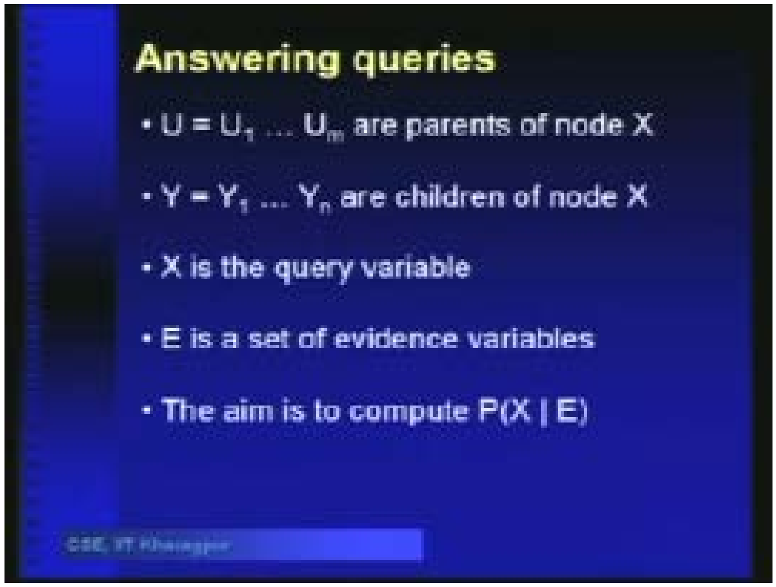
(Refer Slide Time: 29:57)



And X has n successors numbered Y1 through Yn. Each of these successors can have some parents, some other parents. We are going to call the successors of Yi as Y Zi 1, Zi 2, Zi 3, Zi 4, like that. Is this fine? And then, we use EX plus and EX minus to denote the set of evidence that we have preceding X and succeeding X. If you look at all the ancestors of X, then, out of those, the nodes which belong to the evidence, which means whose values are given to us- they are clubbed under the set EX plus.

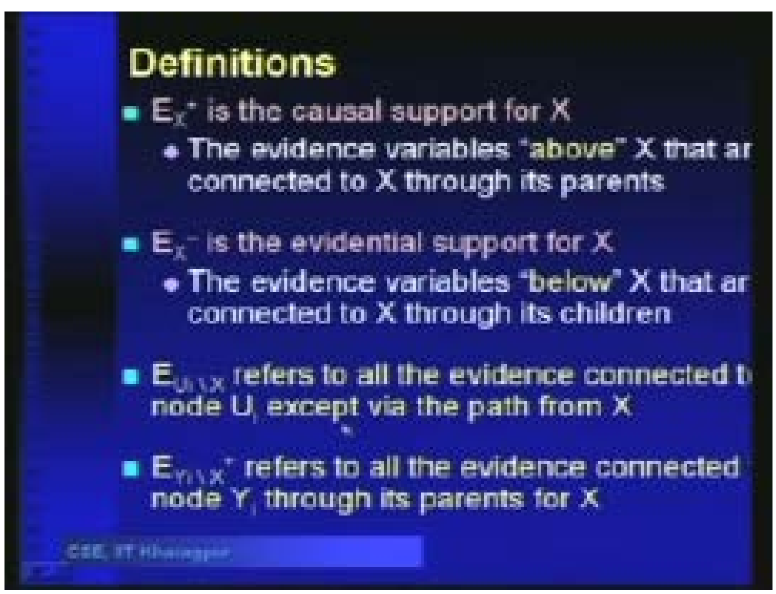
And similarly, EX minus is the set of evidence nodes which are among the successors or descendants of X. May not be immediate children, but descendants of X. Is this clear? Here is what I said just now: U1 through UM are parents of node X. Y1 through Yn are children of X. X is the query variable, E is the set of evidence variables. The aim is to compute PX given E.

(Refer Slide Time: 31:40)



EX plus is called the causal support of X. This is a set of evidence variables above X that are connected to X through its parents.

(Refer Slide Time: 32:46)

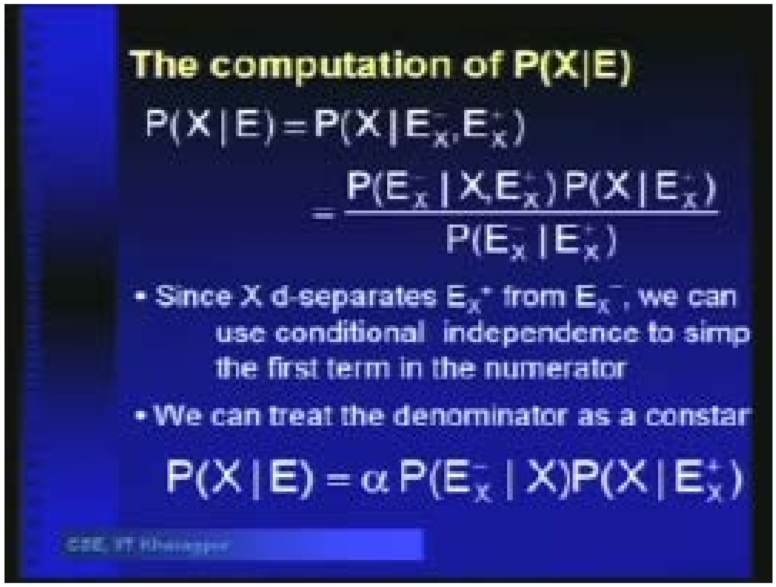


EX minus is the evidential support for X; the evidence variables below X that are connected to X, to its children. In addition, we have 2 other things. Just try to remember this also: what is UI? Parent of X. So, E UI and this is UI except X, these operator in set theoretic terms means except. UI except X refers to all the evidence connected to node UI except via the path from X, right? It means that UI is a parent of X, so, whatever evidence is there in the ancestors of UI comes under UI, except X. Not clear? Let us go back to this picture.

Here is UI right now. UI also has some parents and some ancestors; the set of evidence nodes that are there in the ancestors of UI is called UI except X; not only just the ancestors but also the successors which are not through X. It is all the evidence that is connected to UI, except those that are through X, so, if I disconnect, if I remove this edge that connects UI to X, if I remove this edge and then look at the set of evidence nodes that are connected to UI, then, that is UI except X, right? Similarly, EYi except X plus refers to all evidence connected to node Yi through its parents, except X. It should be except X.

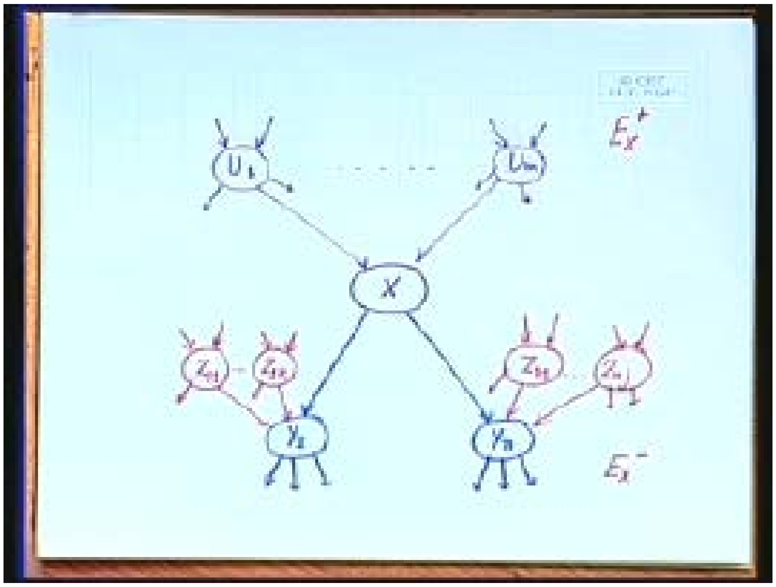
If you look at Yi and there is evidence nodes in the ancestors of Yi except those through X, so, that comes under Eyi except X plus. Is it clear? It will become clearer when we actually go down into the analysis. Now, let us start with the computation. We are infer some quite detailed analysis. Now, observe carefully. Let me do 1 thing: let me take down this picture here, so, let us recap this picture once again.

(Refer Slide Time: 35:02)



We have X here; we have U1 through UM, each of these are parents of X and then, towards the bottom, we have the children of X. So, Y1 through Yn and each of these children in turn, can have their own parents. Z1 through say Z1 k, each of them can have their own parents and children and similarly, our Y1 can also have other children. Similarly, each of these nodes, so, this is a general scenario under which we are analyzing the probability of X, given the evidence. Evidence nodes are in the top above X, which are denoted as EX plus and there are some evidence nodes below X, which are denoted by EX minus.

(Refer Slide Time: 37:28)

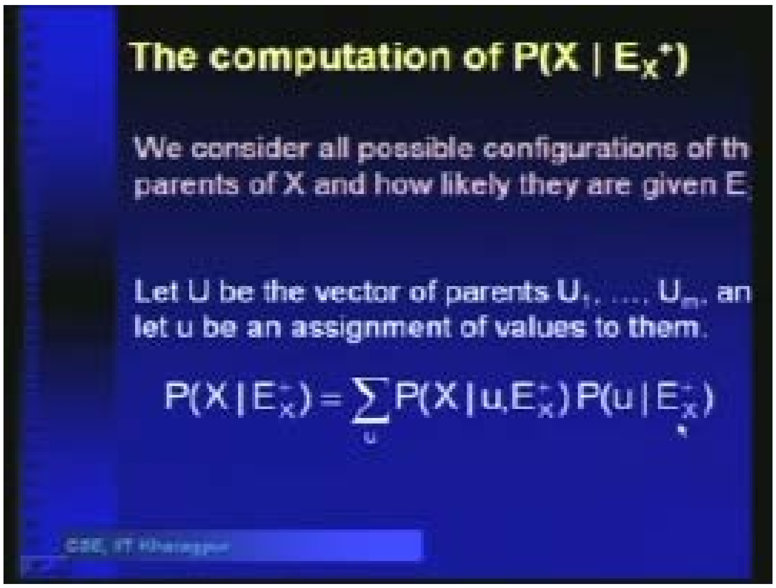


Now, let us go into the analysis. We are interested in computing PX given E, so, we can write E in as the conjunction of the 2 sets EX plus and EX minus, because that is the set of evidence nodes that we have. This is just rewriting E as EX plus and EX minus. Then, we can break it up using Bayes rule, as P EX minus given XEX plus and PX given EX plus and then, the denominator is P E EX minus given EX plus. Now, what we are going to do is, since XD-separates EX plus from EX minus, recall that the kind of network that we are looking at in that- there is only 1 path between every pair of nodes.

So, any path which is from EX minus to EX plus has to necessarily go through X, so therefore, XD-separates EX plus from EX minus and we can use conditional independence to simplify this first term to simply P EX minus given X. And the second term is P EX PX given EX plus, and the denominator is just the probability of the evidence EX minus given the evidence EX plus. So, we treat the denominator as a constant and just we will name it as alpha and we know how to compute this for a given set of evidence nodes. We are going to come back to this later.

Now, we have 2 main terms to compute: P EX minus given X and PX given EX plus. What we have effectively done is, we have reduced both of these terms to the form where we have the effect here and the cause here, because EX minus is all the evidence towards the bottom the set of events, towards the bottom of the network and obviously, they are causally affected by X. So, both of these are like causal terms; we have reduced it into that form. Now, we will compute firstly the first term PX given EX plus. Let U be the vector of parents U1 through UM and let small u be an assignment of values to them.

(Refer Slide Time: 40:53)

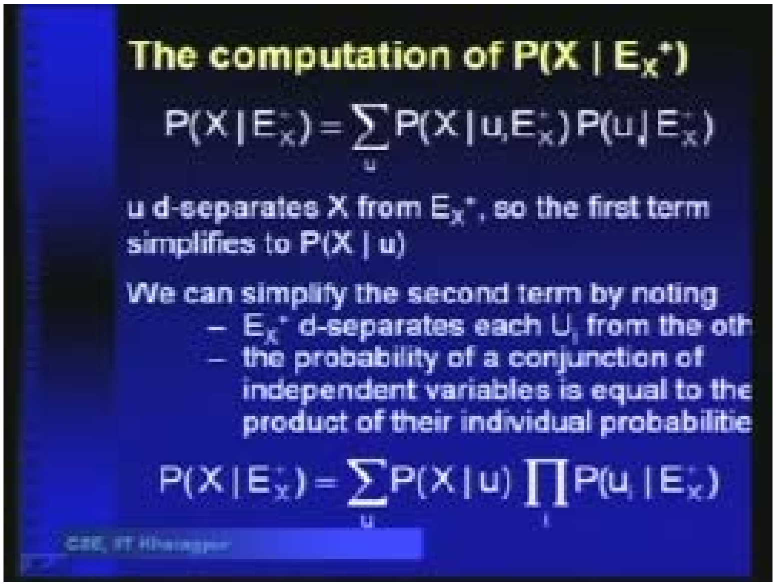


We are going to break up PX given EX plus into the sum of probability of X given U EX plus times probability of U given EX plus. Recall that we had done this previously when we were looking at the probability of John calls given burglary; when we were looking at probability of John calls given burglary- that was exactly what we had done. We had broken it up. Let us go back there. Not this one; yes, when we actually started off here, when we wanted to compute PJB, we actually broke it up into PJBA plus PJAB dash in terms of the parents.

This is exactly what we are doing out here, when we compute this in terms of the parents. What we are doing is, we are inserting all possible combinations of the parent value. Suppose I have parent U1, U2, U3- so, this is getting broken down into U1, U2, U3, U1 bar, U2, U3, U1, U2 bar, U3, and so on. For each of those cases and then, we have the probability of that case given EX plus. If your belief network is such that in a given node, you have too many predecessors or too many parents, then, this term is going to become pretty long.

So, the analysis is going to become quite complex when you have belief networks which has nodes having many parents. That is why it is very important to be able to contain the size of the belief network by proper ordering of the variable. Let us continue with this. This is what we have. Now, u are the parents of XD-separates X from the remaining EX plus, because if you look at the network here, all the evidence that we have on the other side of this is being d-separated from X by the U.

(Refer Slide Time: 45:15)

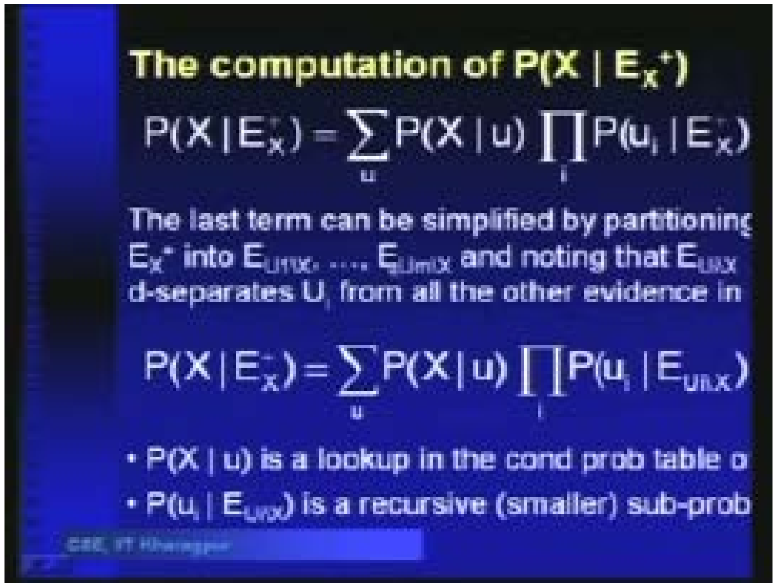


Moreover, what happens is that if you look at all the evidence that is around UI and around Uj, they are also d-separated by X. If you look at U2, there cannot be any common nodes between- there cannot be any other path from U1 to U2 except the 1 which is through X, because of the structure of the network that we have assumed. Therefore, because of this d-separation by X, the evidence around U1 and the evidence around U2 are d-separated and therefore, the probability of each of these UIs given the evidence are all independent. Using that, what we are doing here is, we simplify the first term here- PX U EX plus as PX given U- this we can do, because U is d-separating EX plus from X.

Then, the second term can be simplified, because this U- what is this U? This is a actually a vector, so, it is U1, U2 dash, U3 or U2 dash, U1, U3 dash. It is that vector given EX plus now, if you look at that vector, then, each of those individual items U1, U2, dash U3, etc. are independent, given EX plus why? Because they are d-separated by X. So, I can split them up into the individual UIs given EX plus and the product, because they are independent, so, it is going to split up into the product of their individual probability.

We observe this kind of thing when we looked at probability of alarm given- when we looked at probability of alarm, given- sorry, when we looked at the probability of burglary and earthquake- and we split up into probability of burglary and probability of earthquake. Having split up this thing, then, we look at the last term and it can be simplified by partitioning this evidence into U1 except X through UM except X and noting that UI except XD-separates UI from all the other evidence in EX plus.

(Refer Slide Time: 46:30)



Now, what we are essentially saying here is that, see, there can be some evidence nodes around U1 and there can be some evidence nodes around E2 and some evidence nodes around M, but these are all independent. These are independent of each other because they are being d-separated to it.

(Refer Slide Time: 47:08)

We can actually rewrite this EX plus as each evidence around each EY, so that the evidence around UI is given by EUI except X and that is what we can break up the individual probability terms into.

(Refer Slide Time: 45:15)

If we look at this again, then, this is going to- this second term is going to be written like this. Now, this term PX given U is a- look up in the conditional probability table of X and this is a recursively smaller sub-problem. Recursively now, we will compute P UI given UI except X. The algorithm so far will continue up to this step. Use the conditional probability here and then use recursively- invoke the algorithm to compute this. Now, all through this thing, what we have essentially done is, we have finished a computation of this first term here.

The first term P EX minus given X, sorry, this term we have computed PX given EX plus this term's computation is done. We have an algorithm for doing it. Computing this term is a little more complex, so, we will leave it for the next lecture where we will see how we can compute that and that will complete the algorithm for inferencing using Bayes networks.