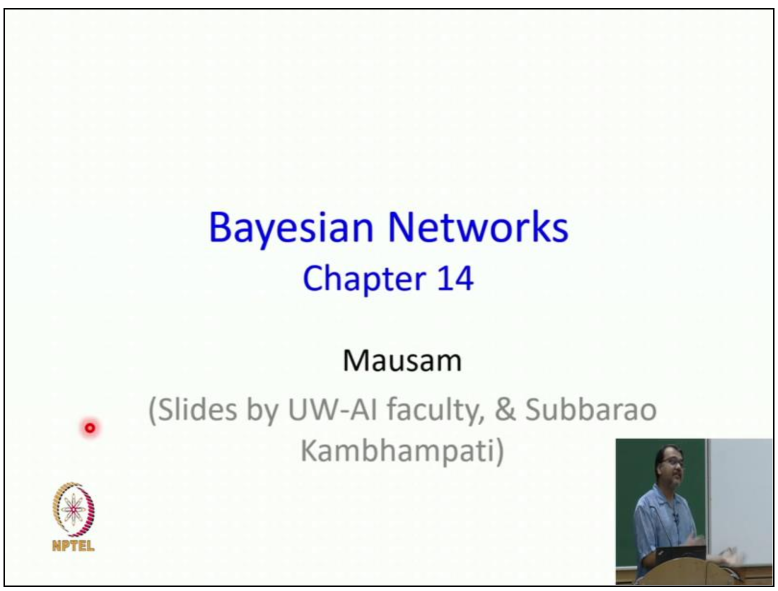
**Artificial Intelligence**
**Prof. Mausam**
**Department of Computer Science and Engineering**
**Indian Institute of Technology - Delhi**

**Lecture-53**
**Bayesian Networks, Syntax, Part - 1**

We are going to start sort of a focus on Bayesian networks. Now, in terms of the big picture where we have been trying to do some various knowledge representation activities and the general idea of knowledge representation has been how do I explain the problem or the world dynamics or the way the world works in general to the machine, right. And we have talked about language of logic.

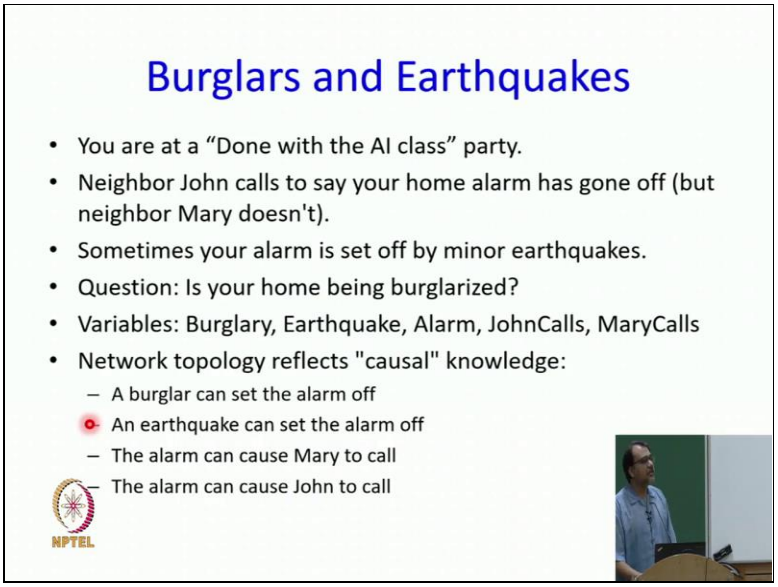
And then, we at some point said, language of logic is good, but not sufficient which is what let us from the traditional AI which was very logic driven to very Modern AI which is highly probability, what was that change we have eluded to it. But today we are going to discuss that first.

**(Refer Slide Time: 01:02)**




And then we will say like we have the language of logic in the world of deterministic environments. And we will have the language of Bayesian networks in the world of probabilistic environment.

**(Refer Slide Time: 01:17)**




So this is going to be our running example. And this is a very famous Burglars and earthquakes example, In this example we say that you are done with AI class. They will come at time that will happen. I know at this point, it looks really long race, but they will come a time when you will be done with the AI class will be like yes, I am done. We can have a party, right. Not that I am stopping you from having those right now, I know that because you have assignments so and so forth. This is very hard for you, so I understand.

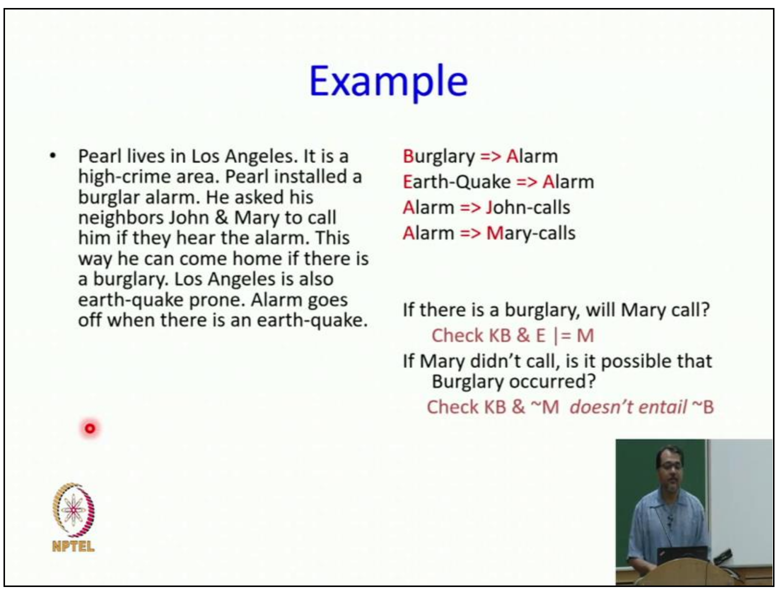
Now, suppose, you have Paul, he was doing his class and he finished his AI class, he went to the party and suddenly your neighbor his neighbor John called him to say alarm has gone off at the home. But his neighbor Mary does not do that? You know that sometimes alarm can be set off by minor earthquake is a small earthquake and the alarm start sleeping, there is no burglary. The question you want to answer is this is your home being burglarized?

Should you leave the very important done with the AI class party? And go back and check you know what all has been stolen or should I just wanted? You know, I remember every distinct example where my uncle has gone off on a trip with family and suddenly he gets a call from the policeman saying your house has been burglarized. We are standing in front of your house 24 by 7 to make sure the burglars do not take anything more. Please come back soon.

It is like if they are standing there 24 by 7 and the burglars have taken what they have to take let us finish my trip and then come back home. That was the other way of looking at it. So, you know you can say ha, my things have already been stolen, let me at least enjoy the party. That is the other way of saying it. But for now the question is not what we should do the question is is the house being burglarized, is the house being burglarized and noticed that some point, you have to realise that we cannot answer this question deterministically.

Are you sure that the house has been burglarized. Are you sure that the house has not been burglarized; you are not sure. You cannot make a deterministic statement about it. And you have to realise there are many things that I will place here, a burglar can set off the alarm, the earthquake can also set off the alarm. The alarm can cause Mary to call, an alarm can cause John to call. Now, suppose you want to model that logic let us think about why we cannot do it. Although I think it should be obvious too many of you let us make sure that we think about it.

**(Refer Slide Time: 04:15)**

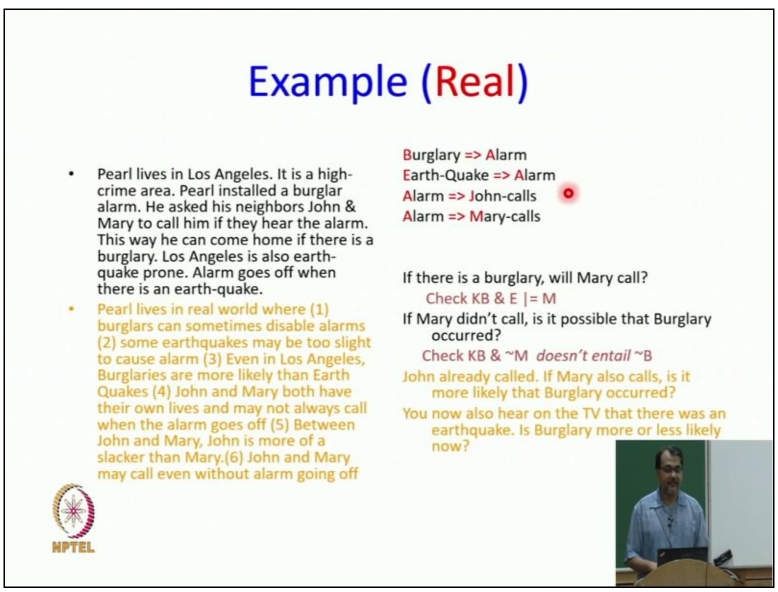



Suppose Pearl lives in Los Angeles, it is high crime area. And he has installed this Burglar alarm and he has asked his neighbors John and Mary to call him whenever they hear the alarm. This is where he can come home if there is any burglary. Los Angeles also on the West Coast, it is earthquake zones. Sometimes minor alarm, minor earthquakes caused. If you are using the world of logic, to the present day you would say burglary implies alarm, you would say the earthquake implies alarm, you would say alarm implies John, going to call, alarm implies Mary is going to come.

This is sort of what logic will allow you to do? It will also allow you to ask the question, if there is burglary will Mary call? You can ask the question that I have this KB of 4 implications and burglary has happened is a mistake here. B has happened will Mary call and the answer may be yes. And you can alternatively ask the question if Mary did not call, is it possible that burglary occurred? And then, you can also answer the question, No, it is not possible. By the way, it does not even represent the possibility that John calls and Mary does not call.

To a logical world this, this logical way of looking at things this world is just impossible. Right, you see that? Because earthquake happened, then alarm will go off, alarm happened then John will call and Mary will call. Maybe you can say that if alarm has not John can still call so that is possibility exists, may be the possibility exists. That is fine. But really what is going on in the real world right is something like this.

**(Refer Slide Time: 06:10)**

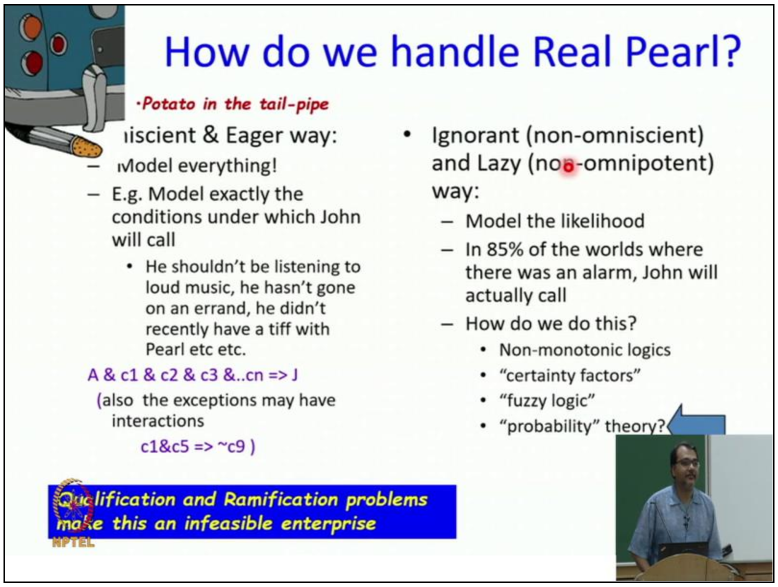



He lives in the real world not in the fictitious world, where Burglars have the ability to disable alarms? You know, Earthquakes may too slide to cause alarm sometimes, an earthquake with no alarm, even in Los Angeles, Burglaries are more common than earthquakes. But notice that we have no way of explaining this information to the logical system. John and Mary actually have lives. They are not just sitting there waiting for them to listen to the alarm and give beauty or Pearl a call. That is not going to happen. If they have life, right?

Sometimes, they may be doing things am not here that the alarm has gone off, right? Of course, between John and Mary, Mary is more sincere John is less. Sometimes, John gets drunk and does not call. Recently, John and UPA also had a Tift, it is possible that you heard, ha, I am not going to call, let his house be burglarized. I think this can happen. And now the question is John is called. If Mary also calls, does it make it more likely that burglary has happened?

Ask that question, this is the kind of question that logic has absolutely no way of thinking about because it has no way of asking the question likelihood whereas Probabilistic models do. Suppose how do we handle the Uriah pearl? Now, there is also the possibility, you may hear on TV today there was an earthquake. Now you ask the question is burglary is more or less right now. So, how do we handle the real Pearl? How do we handle the real burglary.

**(Refer Slide Time: 08:04)**




For example, there are two possible ways: one possible way is the, you enumerate all possible scenarios, right. The uses C1 is when John is not listening to music C2 is when John has not gone on an event C3 is when he has not had a Tift with Uriah, et cetera et cetera. And then you say, that if burglary has happened and C1 happened and C2 has happened and C3 has happened and so on then he is going to call. You can already see what is the problem with this?

What is the problem with this? Too many possibilities and we cannot enumerate all of the them, right. This is called the qualification problem. One such example of the qualification problem is my car does not start. What is going on? And you see that your car does not start, you have checked the engine, you have checked the battery, you have checked the various wires that are going in and out and you finally have no idea and finally realised that the potato was stuck in the tailpipe.

Now, any car diagnosis system is not going to model the fact that there is a potato in the tail pipe. This beautiful example is called the potato in the tailpipe problem also called the qualification and ramification problem. The qualification problem is that one particular event may have many causes it is impossible to anybody dollar. Ramification problem is that one particular event has many effects and it is not possible to enumerate all the problem.

So, this qualification and ramification problems make an invisible. So what is the alternative? Alternative is you say, ha, John may be out, they may be of a theft, he may have be listening to loud music to run, let us all put it into one scenario, so he does not call. In practice, how often would that have done that would let say happen 15% the time. So, we can simply say that it is 85% of the world is when there is alarm; John is going to make a call. And in 15% the world is irrespective of the reason why he did not make call, would not make it.

People just do not worry about enumerating the reasons. Question: First is interesting observation.  Having been criticizing the model that requires you to put in more information that is more detail oriented, whereas I am saying that the better model is the one which has no detail relatively less detail and has put it into the probability. Should not a model which is more detail oriented and more fine grained be a better model. And what is our answer to that?

Let us not give about efficiency, all models are wrong. The question is which one is more useful right? Now, can you enumerate of such possibilities? You may or may not be able to, can you even observe all the possibilities. You may or may not be able to. If you do not know whether John has gone of an errant, and if you do not know whether John has been hearing to loud music or not, It is very hard for you to make inferences about Whether John has called or not, or if he is going to call.

There may be alternative reasons, why John can call? He may just call to trouble you, you know that you are having a party he is stuck at house living alone. He is just calling you to trouble you so that you come back and then he say oh, I thought the alarm went off, anything can happen. You do not know whether that is what is going on in his mind or not. You cannot observe all such things.

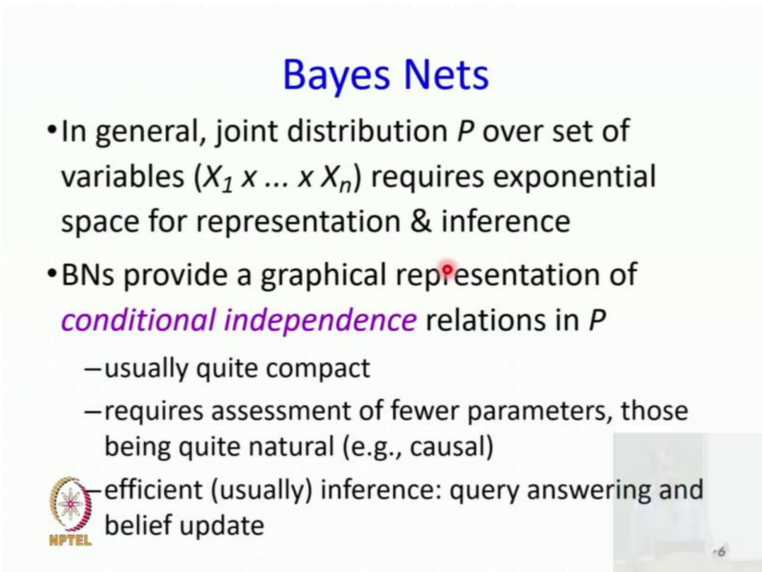
And if you cannot observe all such things, the best you can do in this scenario is to make an likelihood judgement to make an observational based on the observation get a sense of what do I believe in this point of time, right. Again, the short answer to your question would be the; if you can put in the details, if you can observe the details, you can always put it in the probabilistic model that will give you more power.

But of course, if you start putting in all the details in the problem, then you can also have to start asking questions ok, what is the probability that the call if he is drunk, but he has not had a tiff; what about if he is not drunk, but has had a tiff and he is listening to loud music? So, enumerate all such possibility that will make it harder for you to model it effective. So, there may be challenges from any side, always the question you have to ask is, given the current situation, how much I model how much I do not model. What am I not modelling? How do I capture the existence of that other possibility?

Probabilistic model gives a good query. Now, with this background, I think some many of you would agree that probabilistic models are important. And what is the equivalent language from logic into probability world and one of the languages is the language of Bayesian networks. So, last class we talked about in a simple inferences, independence and conditional independence based theorem but now we are going to use all those ideas into one graph, a graph structure which represents the problem.

In this graph structure and many other similar graph structures collectively are called probabilistic graphical models. You would hear the term Pigeons mentioned by a lot of the AI researchers even today, but definitely five years ago. So, probabilistic graphical models were the age in 2000, early 2010, late 90s. There, there was a huge amount of work that was being done in probabilistic graphical models, many different kind of models have come out, Bayesian network Dynamic Bayesian network, Malka network, conditional random fields etcetera.

**(Refer Slide Time: 14:30)**

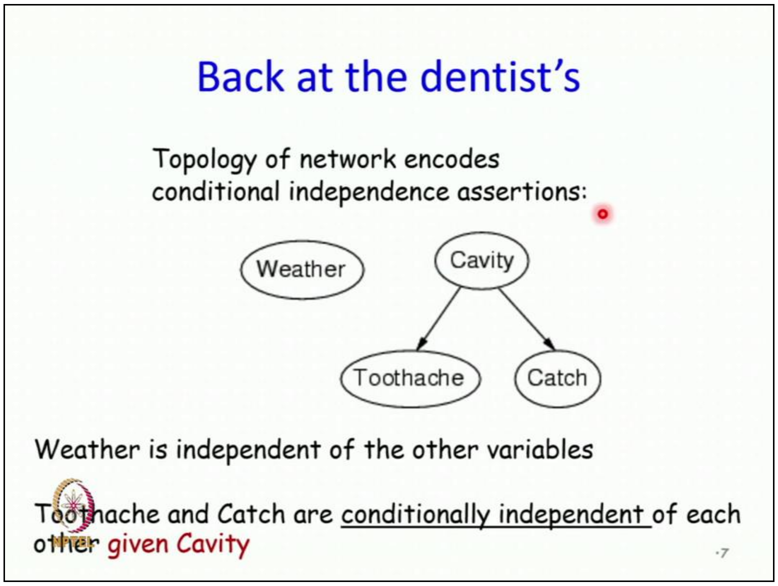



So, the next thing you may remember is that we said that for any n random variables, If I give you the joint distribution, I can compute every query but we said that the joint distribution is not possible to enumerate. So, in general, what we need is a joint distribution and in general, the joint distribution P, we will require exponential space. However, by using Bayesian networks, we are going to make it compact that is our goal.

We will give you a factorization of joint probability distribution that allows us to store it with very limited number of parameters. So, Bayesian networks provide a graphical representation. And what is the fundamental basis on which these graph would be arranged? That would be conditional Independence and all of this will be clear in the next 10 minutes for you. It would be usually compact; they would require assessment of fast fewer parameters than a full exponential one.

And inference may or may not be efficient one. Sometimes, it will be efficient, sometimes, it will not be.

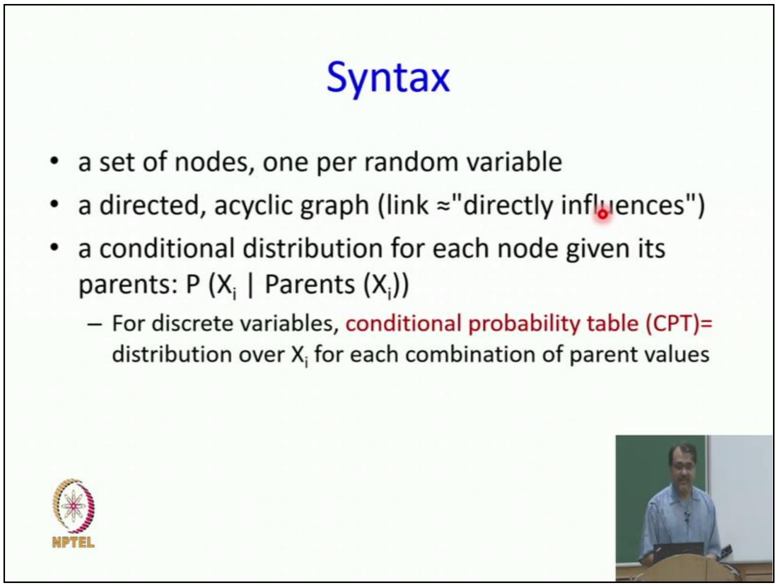
**(Refer Slide Time: 15:42)**




So, we discussed last time that if you are the dentist if you have to somehow manage to place all these variables into some kind of graph structure. It would be what? It would be whether it is completely independent of cavity, tooth ache and catch, so there will be no edge between weather and them. On the other hand cavity would be the one cause for multiple effects. The effects would be tooth ache and catch. They would be edge from cavity to tooth ache, there would be edge from cavity to catch.

But there will be no edge from toothache to catch because toothache does not directly inference catch or visa vis. After two classes, you will be able to read of this graph and say toothache and catch are conditional independent of each other, given cavity. In fact, this graph structures says this in so many words. So, we also have to learn how to read such graphs ok. But first let us talk about Syntax.

**(Refer Slide Time: 16:31)**



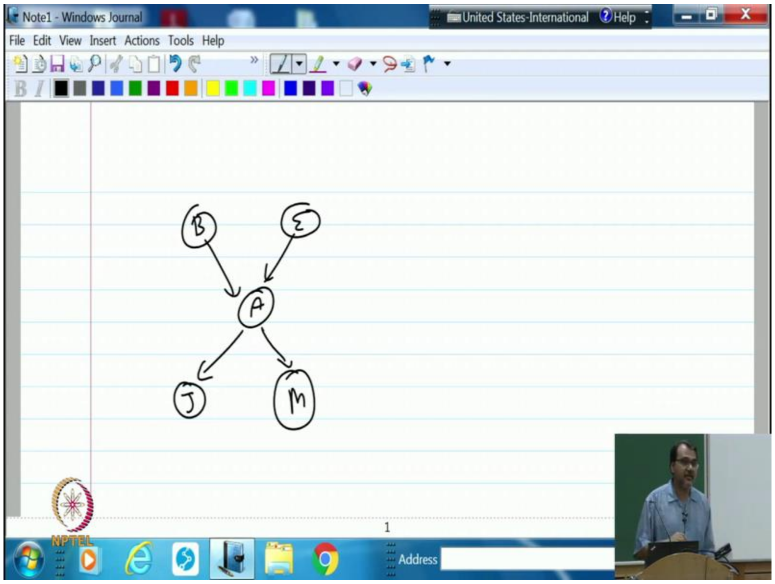

So, the syntax is a set of nodes one per random variable right? Here, there were 4 nodes one per random variable. It will be a directed acyclic graph notice that this is directed and it is a cyclic and the; what is the notion of a link is not a somewhat fuzzy, but for now just imagine the link means it directly influences. So, X1, X2 there is a link that means X1 somehow directly influences X2. So weather does not influence anything any of these 3 cavity influences tooth ache. Moreover, we have to somehow talk about the probabilities. We have not been talking about probabilities.

So, additionally with each node, I am going to associate a conditional probability table, Ok. And this condition probability table would be distribution for each node given its parents so, the probability of Xi given all the parents of Xi. So, if you are discrete variables are people have studied continuous variables. Also we will in this class not talk about continuous variables, but their ways to say what if it does a child is continuous, the parent is discrete.

The parents are continuous, child is continuous. What if the parent is continuous, child is discrete people have thought about functions for that, functional forms and the book talks about that. You should read them. But for now, we will just worry about discrete variable. So we will say that for discrete variable, the conditional probability distribution table over XI for each combination of parents values, ok. This is what we have discussed. How many parameters we need?

So, before we go there, we have annoying this burglars and earthquake world. We have burglary, earthquake alarm, John call and Mary call. If I had to ask you to tell me what should be the graph structure? Just intuitively, though you do not understand, what does the graph mean really? What would you say? Who would be the parent of what? What directly influences what? Earthquake influences alarm.

**(Refer Slide Time: 18:50)**




Earthquake influences alarm what else? Burglary also inferences alarm. In fact, alarm can go off because of two specific reasons like what else? Alarm influence John call. John does not care for burglary directly. He only listens to the alarm and the alarm goes off. Who make call on the other hand for Mary, same thing. Only alarm influences Mary. If I simply asked you to make a graph structure, you would have made this graph structure, right? And that exactly is the graph structure.

For this Bayesian network the most compact graph structure, I should not say that is the graph structure because if you do it in some different order you can come up with the different graph, Ok. That is a little complicated to think about. It is very complicated to think about how to do this in a different order. But machines have a problem with this. In practice what happens is the human gives the graph structure.

The humans I will say the graph structure and you will have the probability that human cannot say that should the probability 0.9, 0.91, 0.92. So, humans are not very good at this. Humans are qualitative people. You cannot say that you know, hundred rupees is ok for me hundred one is really expensive. You cannot say that. That you cannot say that this is my cut off. Have you seen this EMI option to say, what is the maximum, how many to pay you want to pay maximum for you 10000 rupees and 10001 rupees is very similar.

So it is very hard for us to draw line. It is very hard first to give threshold what we can sort of give ballpark. We can say this is expensive. This is to keep et cetera. So, humans will be really good at giving the best graph structure many times. But machines are better with computing Probabilities and this is our typical division of labour happens with all the methods for doing what is called structure learning which basically means learn the graph structure once.