Module No. 4

Bayesian and Computational Learning

Lecture -4

Bayesian Belief Network

- Probabilistic models determine the relationship between variables, and then you can calculate the various probabilities of those two values.
- Bayesian Network is also called a Probabilistic Graphical Model (PGM).
- Example:
- *Lung cancer. A patient has been suffering from shortness of breath (called dyspnoea) and visits the doctor, worried that he has lung cancer. The doctor knows that other diseases, such as tuberculosis and bronchitis, are possible causes, as well as lung cancer. She also knows that other relevant information includes whether or not the patient is a smoker (increasing the chances of cancer and bronchitis) and what sort of air pollution he has been exposed to. A positive X-ray would indicate either TB or lung cancer.*

Bayesian Belief Network in artificial intelligence

•Bayesian belief network is key computer technology for dealing with probabilistic events and to solve a problem which has uncertainty.

We can define a Bayesian network as:

•"A Bayesian network is a probabilistic graphical model which represents a set of variables and their conditional dependencies using a directed acyclic graph."

•It is also called a **Bayes network, belief network, decision network**, or **Bayesian model**.

•Bayesian networks are probabilistic, because these networks are built from a **probability distribution**, and also use probability theory for prediction and anomaly detection.

Real world applications are probabilistic in nature, and to represent the relationship between multiple events, we need a Bayesian network. It can also be used in various tasks including **prediction, anomaly detection, diagnostics, automated insight, reasoning, time series prediction**, and **decision making under uncertainty**.

Bayesian Network can be used for building models from data and experts opinions, and it consists of two parts:

- **Directed Acyclic Graph**
- **Table of conditional probabilities.**

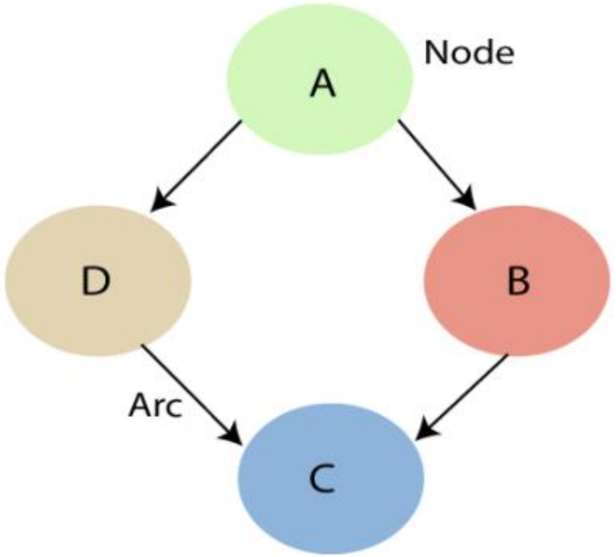
The generalized form of Bayesian network that represents and solve decision problems under uncertain knowledge is known as an **Influence diagram**.

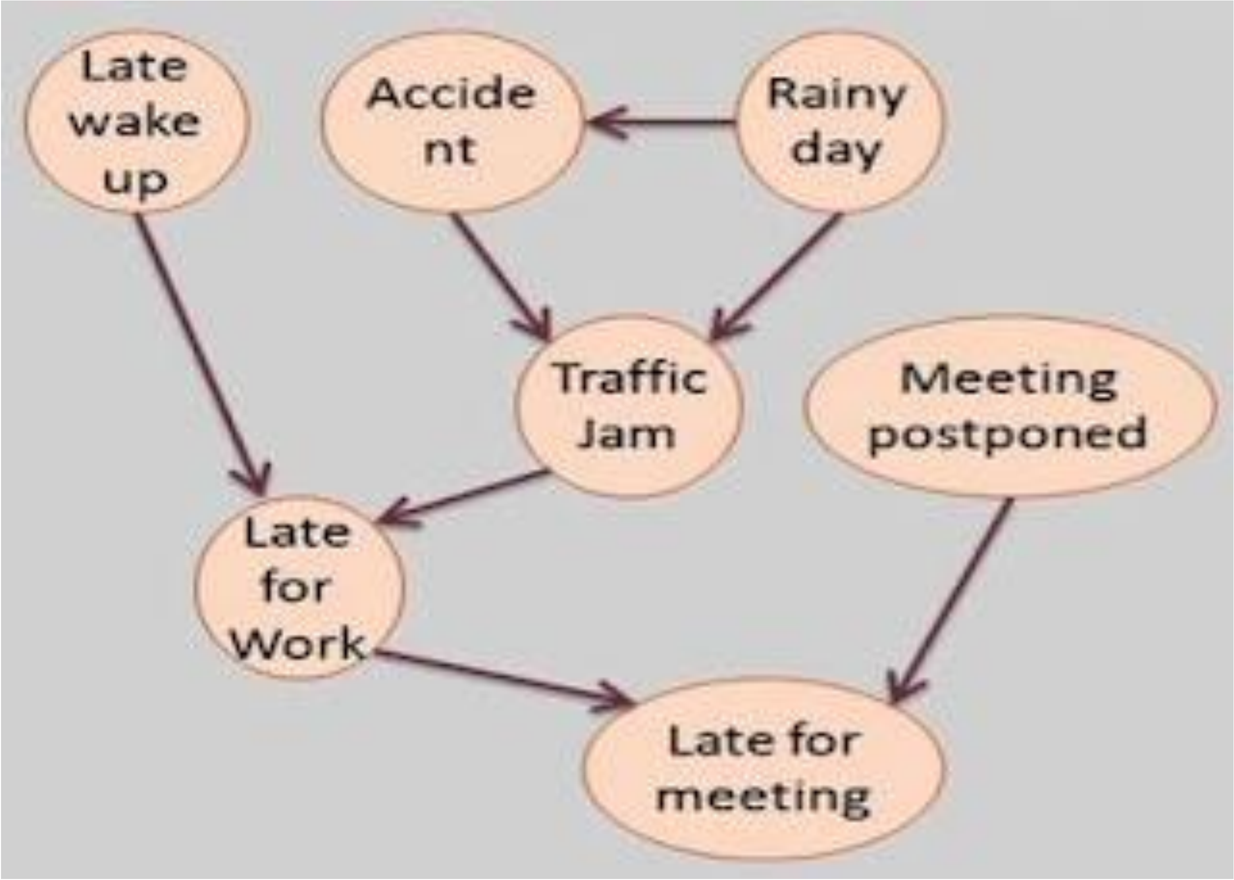
Bayesian Belief Network

- is a graphical representation of different probabilistic relationships among random variables in a particular set.
- It is a classifier with no dependency on attributes i.e it is condition independent.
- Due to its feature of joint probability, the probability in Bayesian Belief Network is derived, based on a condition — P(attribute/parent) i.e probability of an attribute, true over parent attribute.

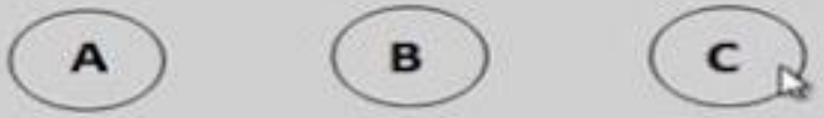
A Bayesian network graph is made up of nodes and Arcs (directed links), where:

- Each node corresponds to the random variables, and a variable can be continuous or discrete.
- Arc or directed arrows represent the causal relationship or conditional probabilities between random variables. These directed links or arrows connect the pair of nodes in the graph. These links represent that one node directly influence the other node, and if there is no directed link that means that nodes are independent with each other
- In the above diagram, A, B, C, and D are random variables represented by the nodes of the network graph.
- If we are considering node B, which is connected with node A by a directed arrow, then node A is called the parent of Node B.
- Node C is independent of node A.

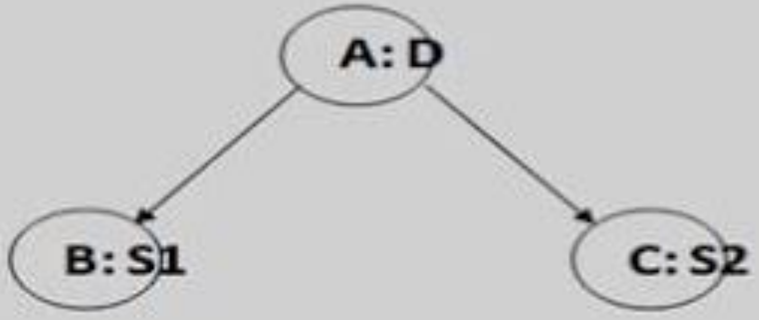



Late wake up
Accide nt
Rainy day
Traffic Jam
Meeting postponed
Late for Work
Late for meeting

- Bayesian Networks are more restrictive, where the edges of the graph are directed, meaning they can only be navigated in one direction.
- This means that cycles are not possible, and the structure can be more generally referred to as a directed acyclic graph (DAG).
- A probabilistic graphical model, such as a Bayesian Network, provides a way of defining a probabilistic model for a complex problem by stating all of the conditional independence assumptions for the known variables, whilst allowing the presence of unknown (latent) variables.



Marginal Independence:
$p(A,B,C) = p(A)\,p(B)\,p(C)$



Conditionally independent effects:
$p(A,B,C) = p(B|A)p(C|A)p(A)$
B and C are conditionally independent
Given A

Independent Causes:
$p(A,B,C) = p(C|A,B)p(A)p(B)$
“Explaining away”

Markov dependence:
$p(A,B,C) = p(C|B)\,p(B|A)p(A)$

The Bayesian network has mainly two components:

- **Causal Component**
- **Actual numbers**

•Each node in the Bayesian network has condition probability distribution **$P(X_i \mid Parent(X_i))$**, which determines the effect of the parent on that node.

•Bayesian network is based on Joint probability distribution and conditional probability.

Joint probability distribution:

If we have variables x1, x2, x3,....., xn, then the probabilities of a different combination of x1, x2, x3.. xn, are known as Joint probability distribution.

$P[x_1, x_2, x_3, \ldots, x_n]$, it can be written as the following way in terms of the joint probability distribution.

$$= P[x_1 \mid x_2, x_3, \ldots, x_n] P[x_2, x_3, \ldots, x_n]$$

$$= P[x_1 \mid x_2, x_3, \ldots, x_n] P[x_2 \mid x_3, \ldots, x_n] \ldots P[x_{n-1} \mid x_n] P[x_n].$$

In general for each variable Xi, we can write the equation as:

$$P(X_i \mid X_{i-1}, \ldots\ldots\ldots, X_1) = P(X_i \mid Parents(X_i))$$

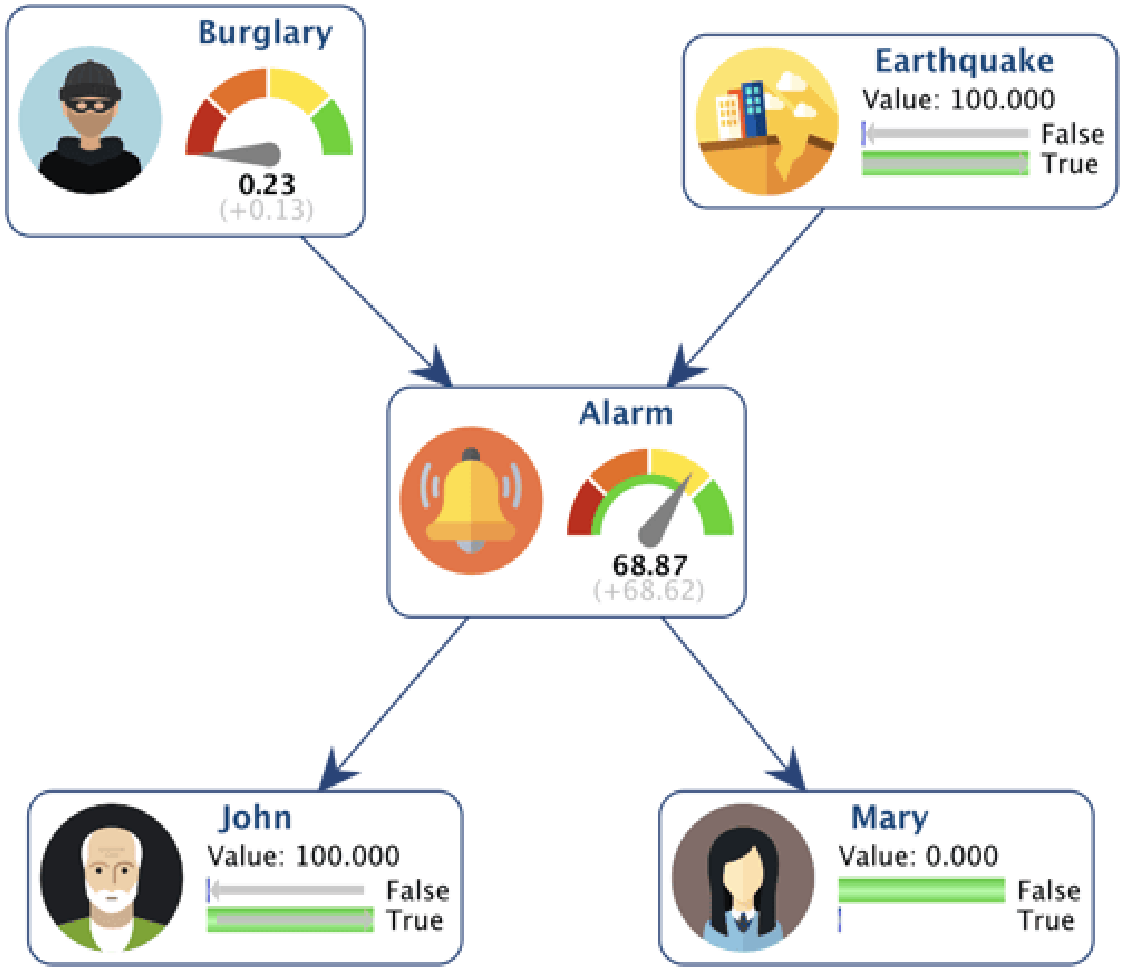
The joint probability distribution

- A generic entry in the joint probability distribution $P(x_1, \ldots, x_n)$ is given by:

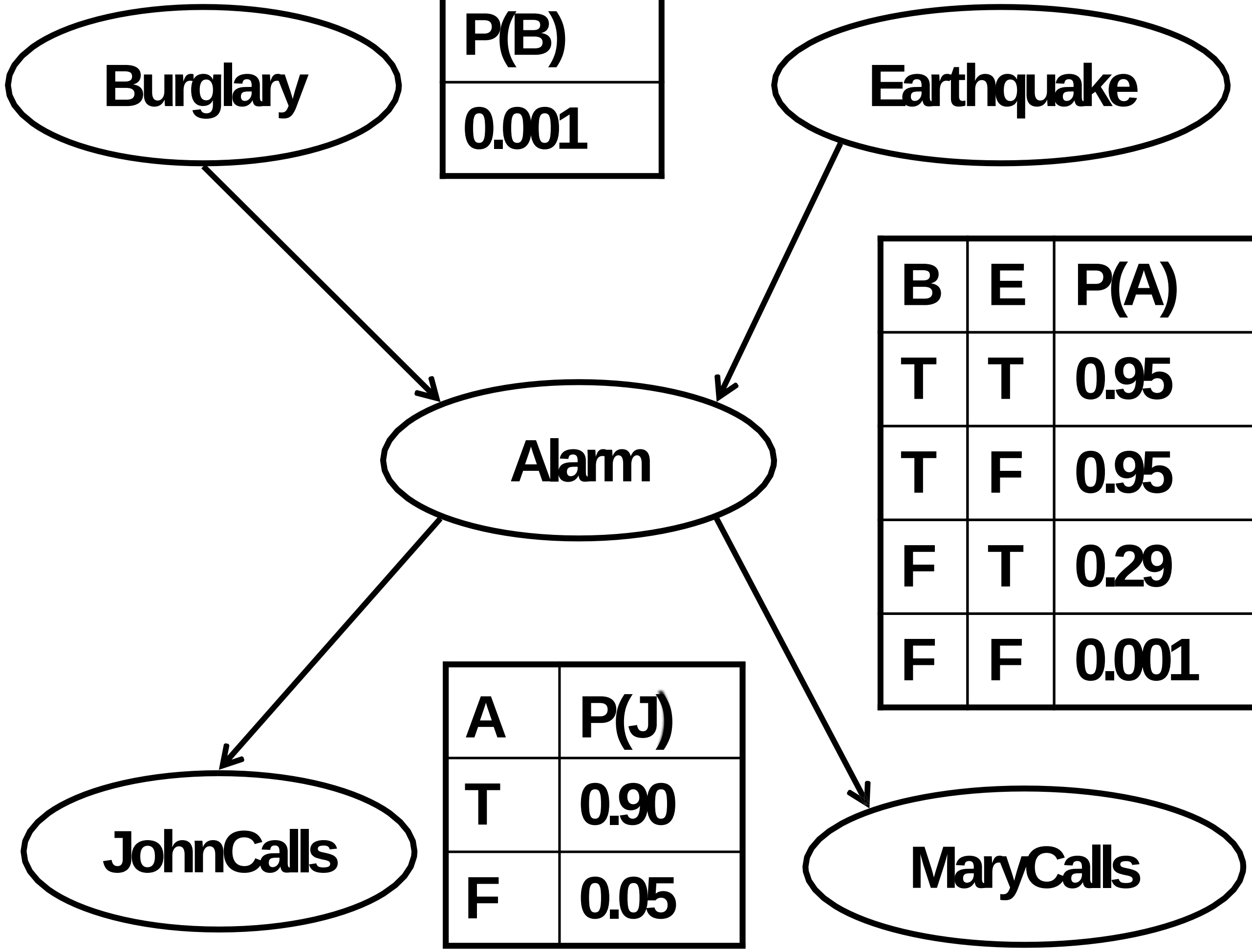
$$P(x_1,\ldots,x_n) = \prod_{i=1}^{n} P(x_i \mid Parents(X_i))$$

Example

- ❑ **Burglar alarm at home**
    - Fairly reliable at detecting a burglary
    - Responds at times to minor earthquakes
- ❑ **Two neighbors, on hearing alarm, calls police**
    - John always calls when he hears the alarm, but sometimes confuses the telephone ringing with the alarm and calls then, too.
    - Mary likes loud music and sometimes misses the alarm altogether



Belief Network Example



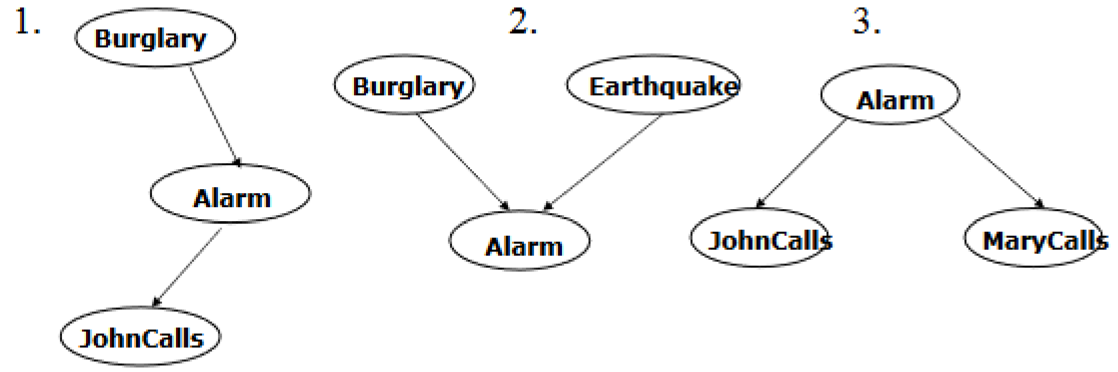
P(B)
0.001

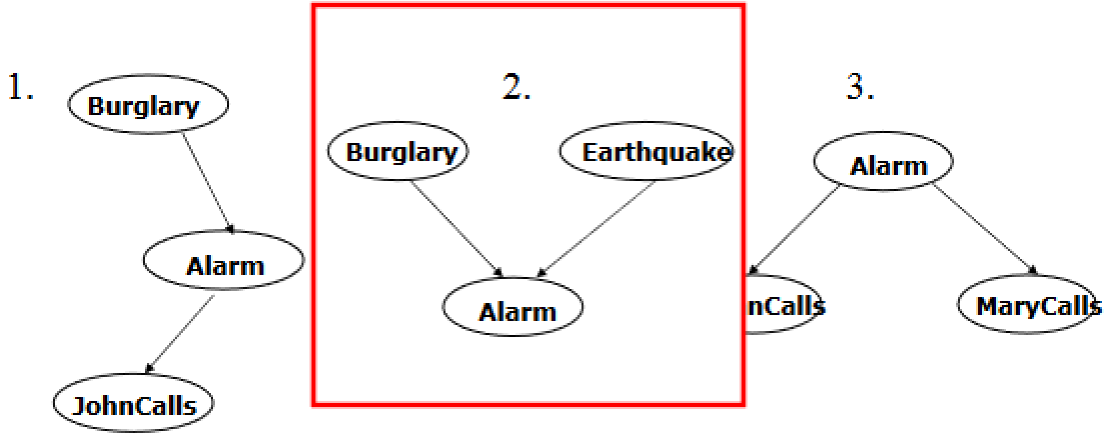
P(E)
0.002

B	E	P(A)
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

A	P(J)
T	0.90
F	0.05

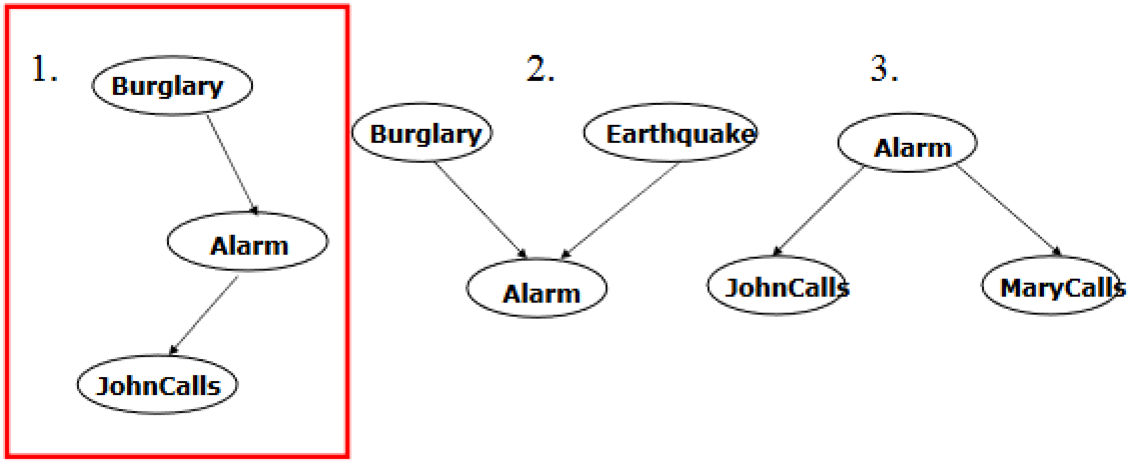
A	P(M)
T	0.70
F	0.01





2. Burglary **is independent** of Earthquake (not knowing Alarm)
   Burglary and Earthquake **become dependent** given Alarm !!

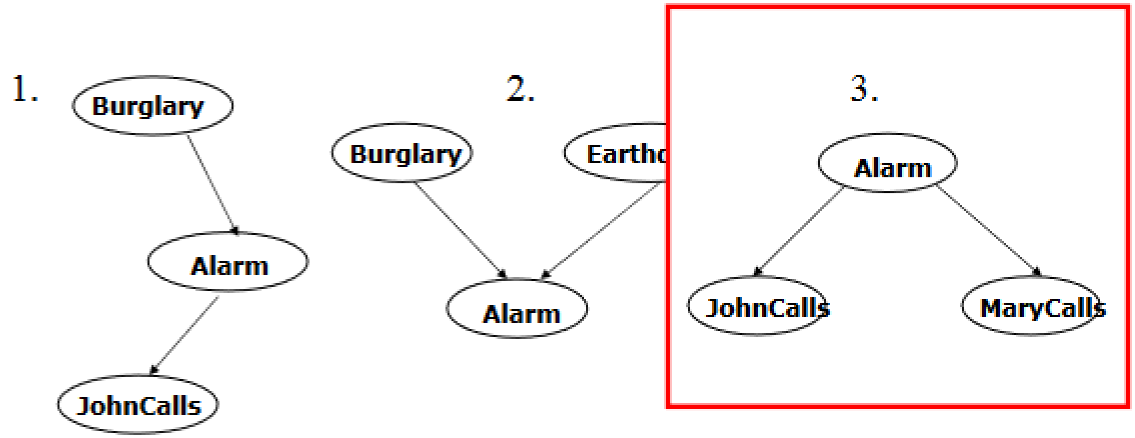
$$P(B,E) = P(B)P(E)$$



1. JohnCalls **is independent** of Burglary given Alarm

$$P(J \mid A,B) = P(J \mid A)$$

$$P(J,B \mid A) = P(J \mid A)P(B \mid A)$$



3. MaryCalls **is independent** of JohnCalls given Alarm

$$P(J \mid A,M) = P(J \mid A)$$

$$P(J,M \mid A) = P(J \mid A)P(M \mid A)$$

The joint probability distribution

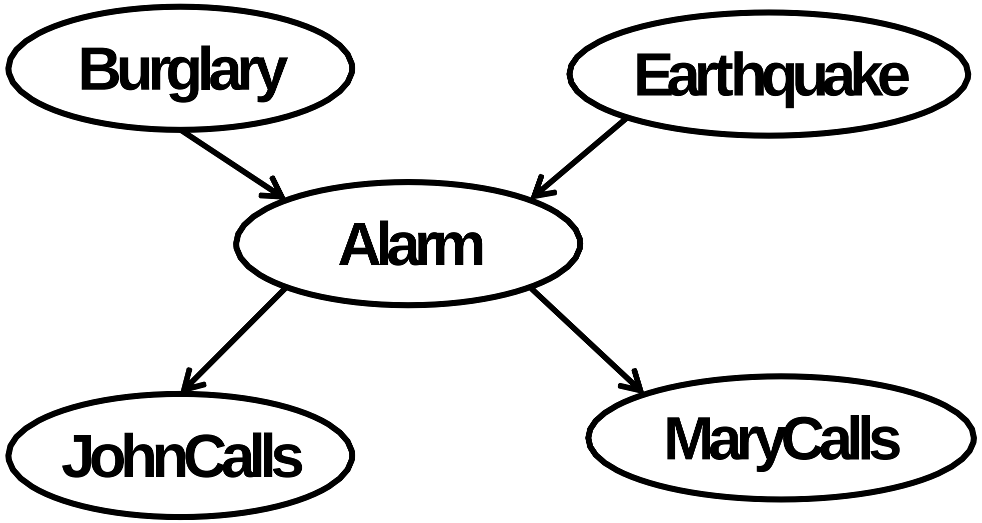
- Probability of the event that the alarm has sounded but neither a burglary nor an earthquake has occurred, and both Mary and John call:

$$P(J \wedge M \wedge A \wedge \neg B \wedge \neg E)$$

$$= P(J \mid A)\ P(M \mid A)\ P(A \mid \neg B \wedge \neg E)\ P(\neg B)\ P(\neg E)$$

$$= 0.9 \text{ X } 0.7 \text{ X } 0.001 \text{ X } 0.999 \text{ X } 0.998$$

$$= 0.00062$$

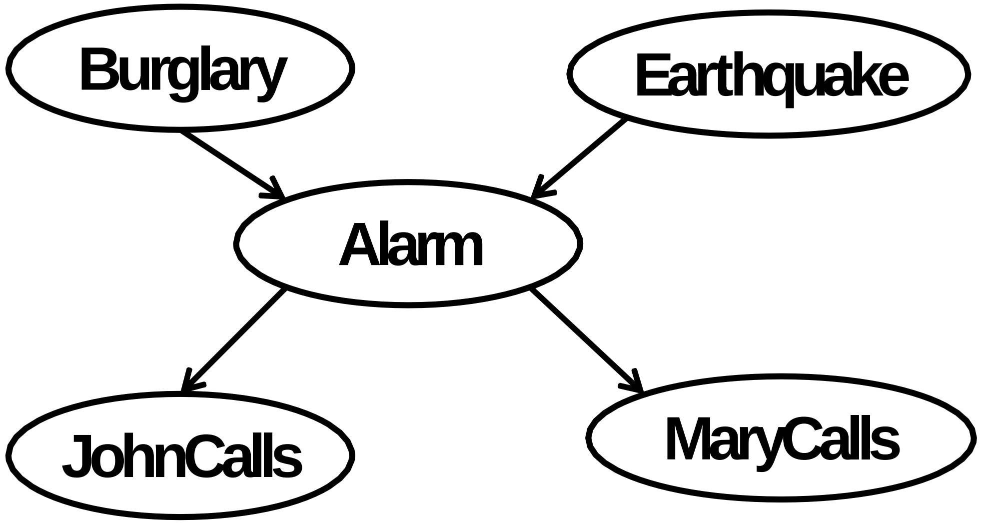


The joint probability distribution

- Computation of the probabilities of several different event combinations of the Burglary-Alarm belief network example:

$P(B) = 0.001$
$P(B') = 1 - P(B) = 0.999$

$P(E) = 0.002$
$P(E') = 1 - P(E) = 0.998$



The joint probability distribution

- Computation of the probabilities of several different event combinations of the Burglary-Alarm belief network example:

$$P(A) = P(AB'E') + P(AB'E) + P(ABE') + P(ABE)$$

$$= P(A \mid B'E').P(B'E') + P(A \mid B'E).P(B'E) + P(A \mid BE').P(BE') + P(A \mid BE).P(BE)$$

$$= 0.001 \times 0.999 \times 0.998$$

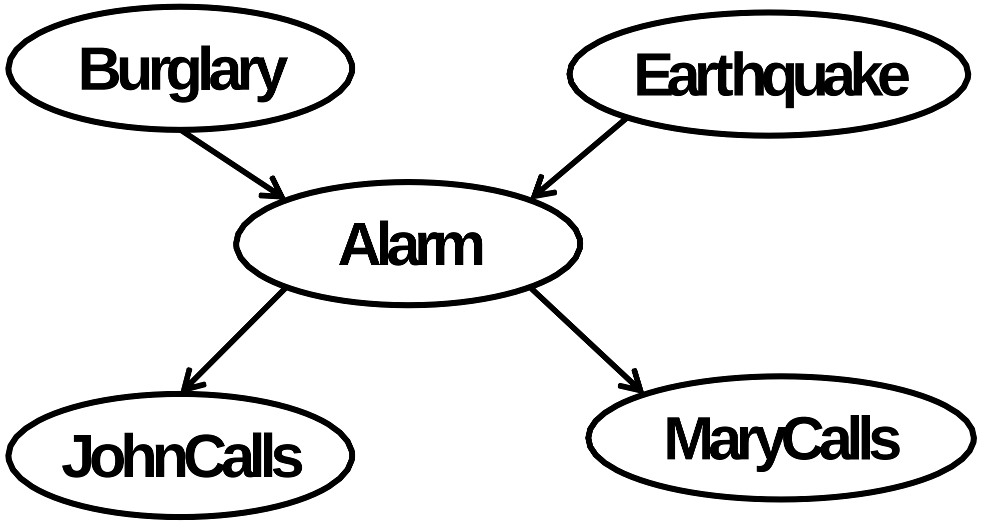
$$+ 0.29 \times 0.999 \times 0.002$$

$$+ 0.95 \times 0.001 \times 0.998$$

$$+ 0.95 \times 0.001 \times 0.002$$

$$= 0.001 + 0.0006 + 0.0009$$

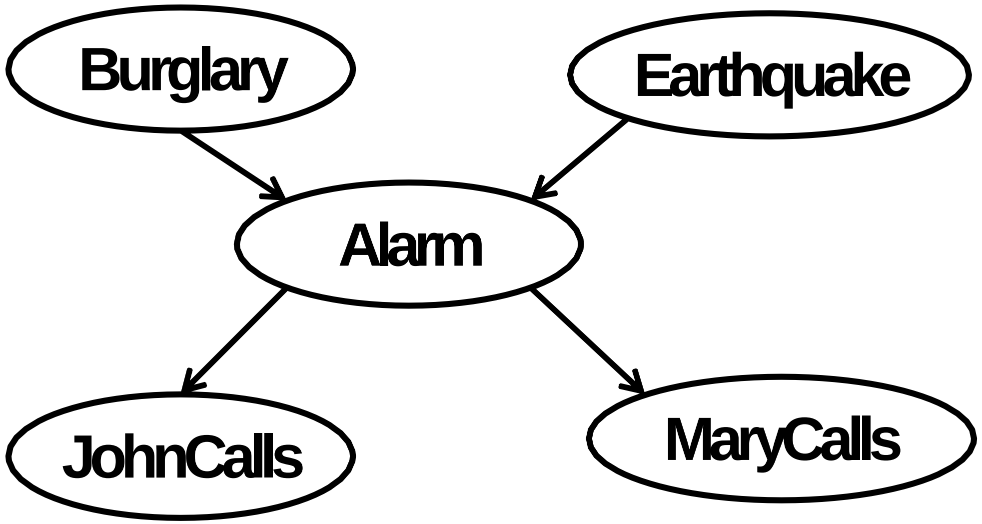
$$= 0.0025$$



The joint probability distribution

- Computation of the probabilities of several different event combinations of the Burglary-Alarm belief network example:

$$P(J) = P(JA) + P(JA')$$
$$= P(J \mid A).P(A) + P(J \mid A').P(A')$$
$$= 0.9 \times 0.0025 + 0.05 \times (1 - 0.0025)$$
$$= 0.052125$$

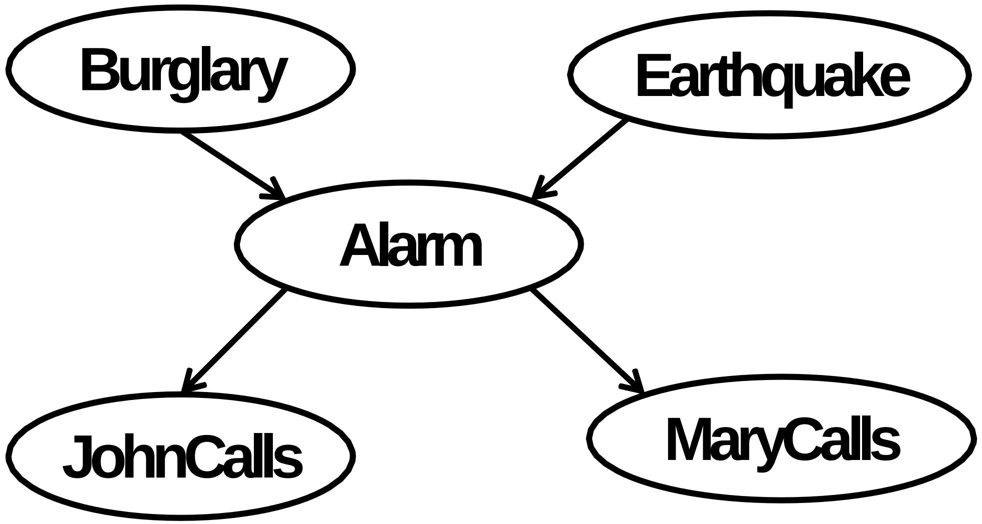


$$P(AB) = P(ABE) + P(ABE')$$
$$= 0.95 \times 0.001 \times 0.002 + 0.95 \times 0.001 \times 0.998$$
$$= 0.00095$$

The joint probability distribution

- Computation of the probabilities of several different event combinations of the Burglary-Alarm belief network example:

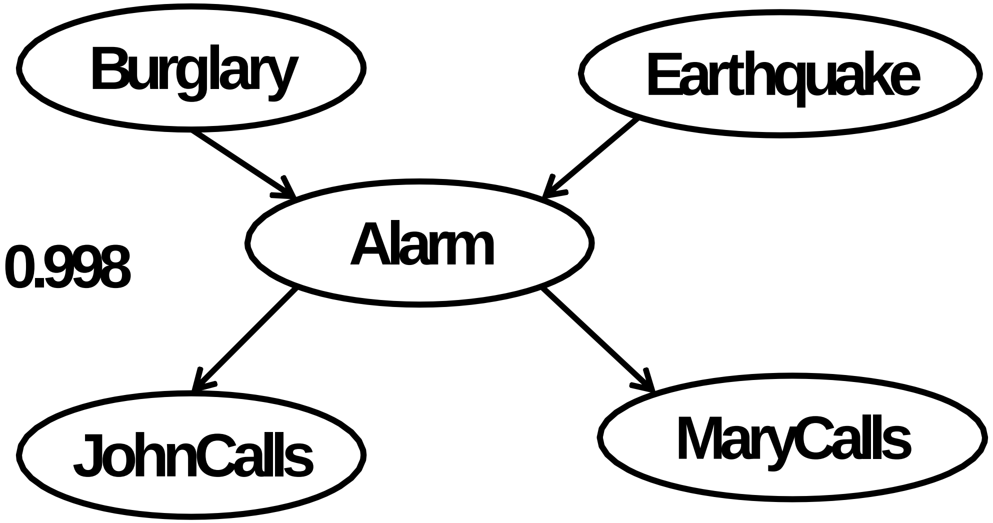
$$
\begin{aligned}
P(A'B) &= P(A'BE) + P(A'BE') \\
&= P(A' \mid BE).P(BE) + P(A' \mid BE').P(BE') \\
&= (1 - 0.95) \times 0.001 \times 0.002 \\
&+ (1 - 0.95) \times 0.001 \times 0.998 \\
&= 0.00005
\end{aligned}
$$



$$
\begin{aligned}
P(AE) &= P(AEB) + P(AEB') \\
&= 0.95 \times 0.001 \times 0.002 + 0.29 \times 0.999 \times 0.002 \\
&= 0.00058
\end{aligned}
$$

The joint probability distribution

- **Computation of the probabilities of several different event combinations of the Burglary-Alarm belief network example:**



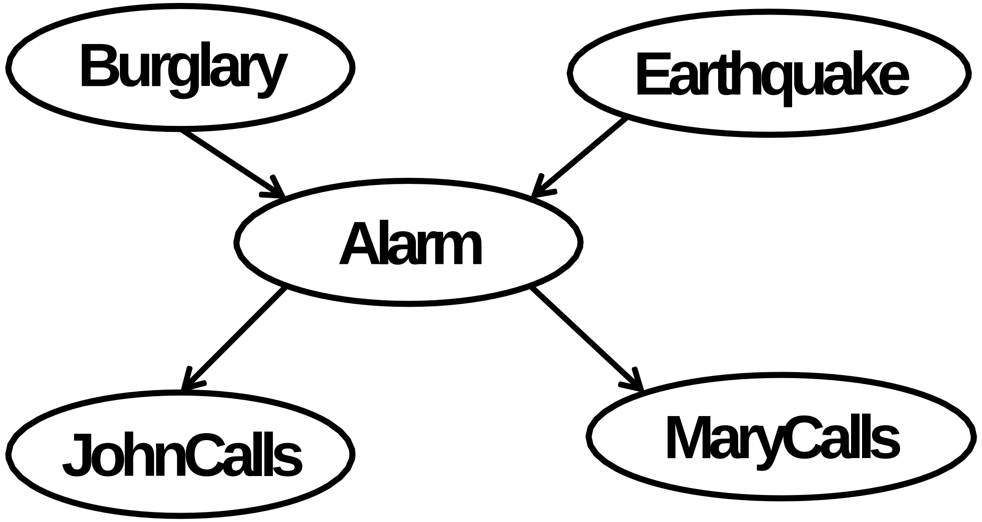
$$
\begin{aligned}
P(AE') &= P(AE'B) + P(AE'B') \\
&= 0.95 \times 0.001 \times 0.998 + 0.001 \times 0.999 \times 0.998 \\
&= 0.001945
\end{aligned}
$$

$$
\begin{aligned}
P(A'E') &= P(A'E'B) + P(A'E'B') \\
&= P(A' \mid BE').P(BE') + P(A' \mid B'E').P(B'E') \\
&= (1 - 0.95) \times 0.001 \times 0.998 + (1 - 0.001) \times 0.999 \times 0.998 \\
&= 0.996
\end{aligned}
$$

The joint probability distribution

- Computation of the probabilities of several different event combinations of the Burglary-Alarm belief network example:

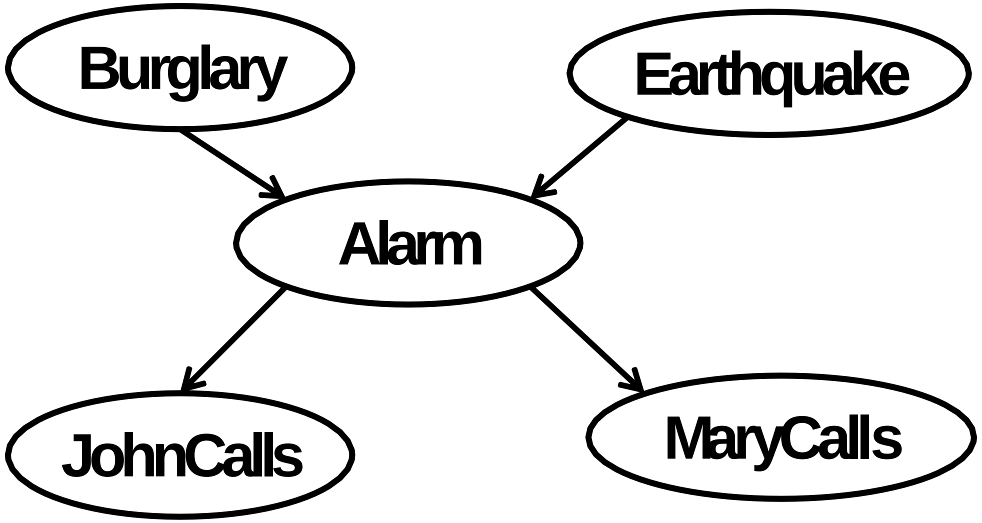
$$
\begin{aligned}
P(JB) &= P(JBA) + P(JBA') \\
&= P(J \mid AB).P(AB) + P(J \mid A'B).P(A'B) \\
&= P(J \mid A).P(AB) + P(J \mid A').P(A'B) \\
&= 0.9 \times 0.00095 + 0.05 \times 0.00005 \\
&= 0.00086
\end{aligned}
$$



The joint probability distribution

- Computation of the probabilities of several different event combinations of the Burglary-Alarm belief network example:

$P(J \mid B) = P(JB) / P(B) = 0.00086 / 0.001 = 0.86$



The joint probability distribution

- Computation of the probabilities of several different event combinations of the Burglary-Alarm belief network example:

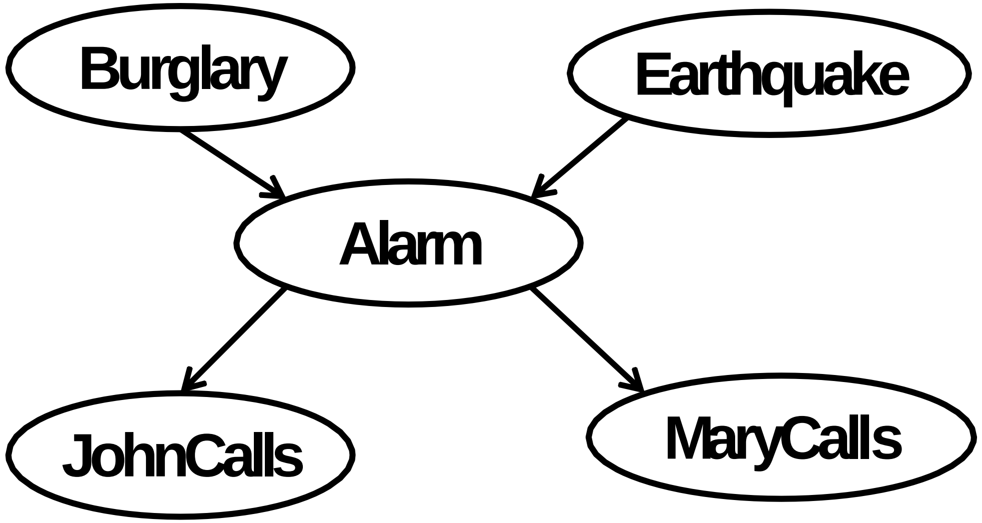
$$P(MB) = P(MBA) + P(MBA')$$

$$= P(M \mid AB).P(AB) + P(M \mid A'B).P(A'B)$$

$$= P(M \mid A).P(AB) + P(M \mid A').P(A'B)$$

$$= 0.7 \times 0.00095 + 0.01 \times 0.00005$$

$$= 0.00067$$



The joint probability distribution

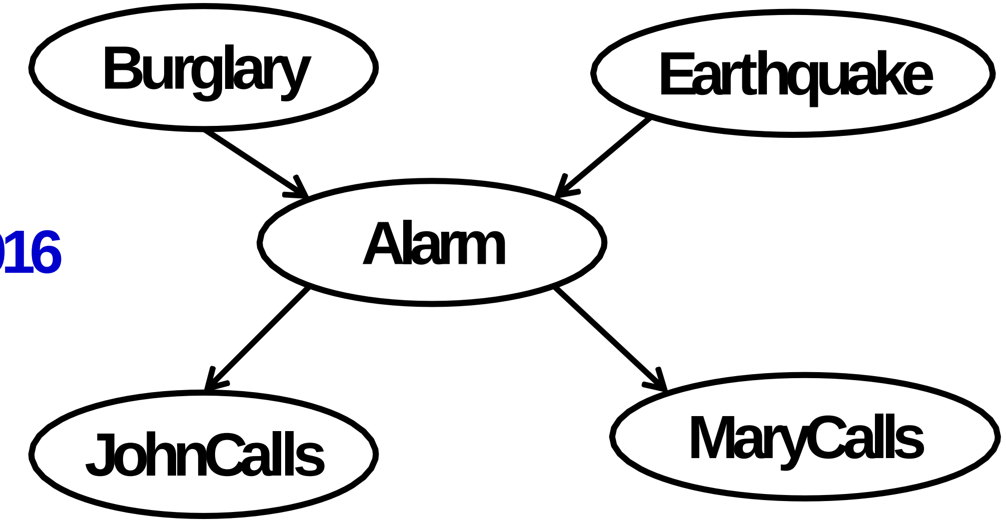
- **Computation of the probabilities of several different event combinations of the Burglary-Alarm belief network example:**

$P(M \mid B) = P(MB) / P(B) = 0.00067 / 0.001 = 0.67$

$P(B \mid J) = P(JB) / P(J) = 0.00086 / 0.052125 = 0.016$

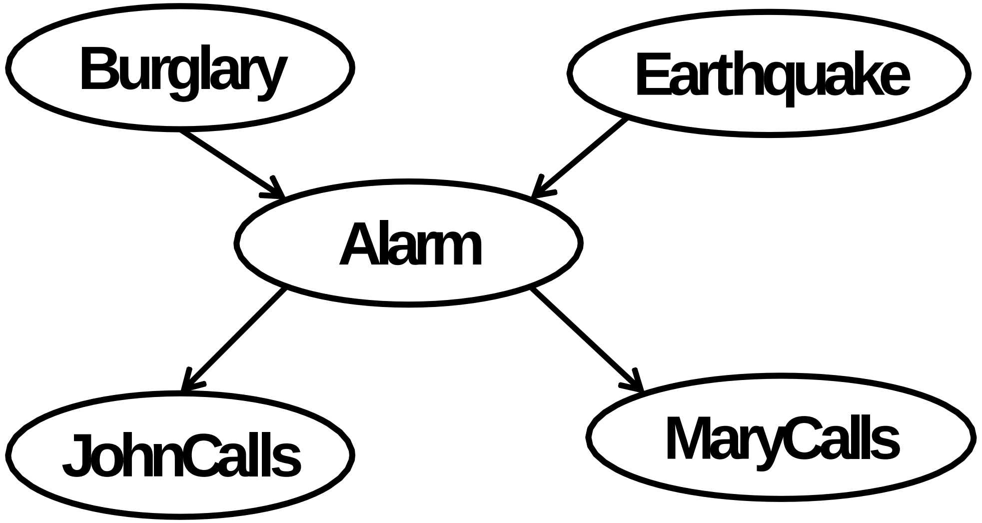
$P(B \mid A) = P(AB) / P(A) = 0.00095 / 0.0025 = 0.38$

$$P(B \mid AE) = P(ABE) / P(AE) = [\, P(A \mid BE).P(BE) \,] / P(AE)$$
$$= [\, 0.95 \times 0.001 \times 0.002 \,] / 0.00058$$
$$= 0.003$$



The joint probability distribution

- Computation of the probabilities of several different event combinations of the Burglary-Alarm belief network example:



$$\begin{aligned} P(AJE') &= P(J \mid AE').P(AE') \\ &= P(J \mid A).P(AE') \\ &= 0.9 \times 0.001945 \\ &= 0.00175 \end{aligned}$$

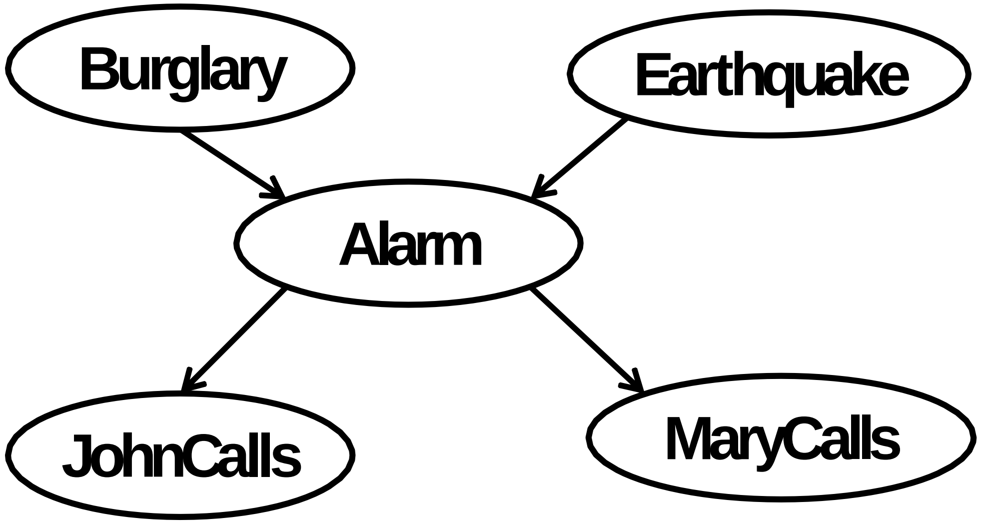
$$\begin{aligned} P(A'JE') &= P(J \mid A'E').P(A'E') \\ &= P(J \mid A').P(A'E') \\ &= 0.05 \times 0.996 \\ &= 0.0498 \end{aligned}$$

$$P(JE') = P(AJE') + P(A'JE') = 0.00175 + 0.0498 = 0.05155$$

The joint probability distribution

- Computation of the probabilities of several different event combinations of the Burglary-Alarm belief network example:

$$P(A \mid JE') = P(AJE') / P(JE') = 0.00175 / 0.05155 = 0.03$$



The joint probability distribution

- Computation of the probabilities of several different event combinations of the Burglary-Alarm belief network example:

$P(BJE') = P(BJE'A) + P(BJE'A')$

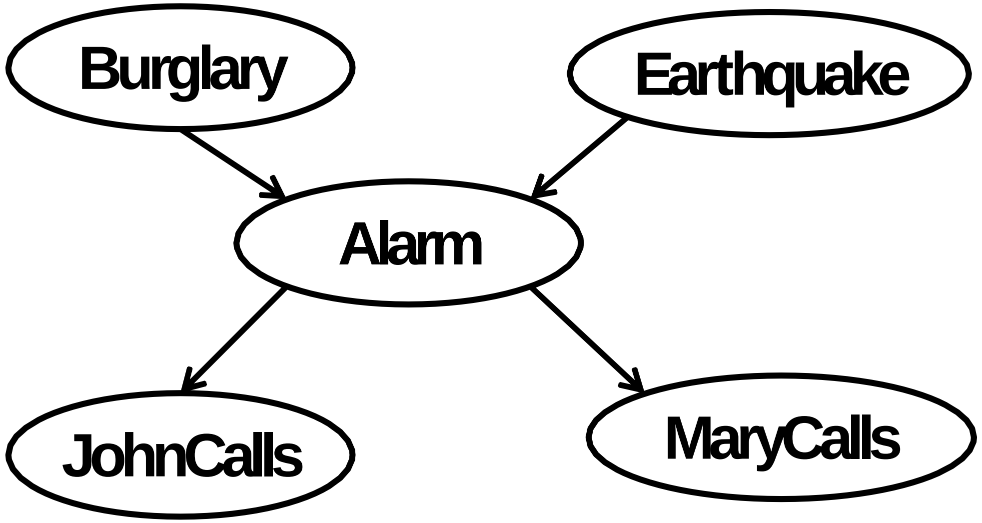
$= P(J \mid ABE').P(ABE') + P(J \mid A'BE').P(A'BE')$

$= P(J \mid A).P(ABE') + P(J \mid A').P(A'BE')$

$= 0.9 \times 0.95 \times 0.001 \times 0.998 + 0.05$

$\times (1 - 0.95) \times 0.001 \times 0.998$

$= 0.000856$



$P(B \mid JE') = P(BJE') / P(JE') = 0.000856 / 0.05155 = 0.017$

What is the probability of the event that the alarm has sounded and no burglary but an earthquake has occurred and both Mary and John call?

- P(M ^ J ^ !B ^ E) = P(J |A) * P(M|A) * P(A |!B ^E) * P(!B) *P(E)
- 0.90 x 0.70 x 0.29 x 0.999 x 0.002 = 0.00036

What is the probability of the event that the alarm has sounded but neither a burglary nor an earthquake has occurred and John call and Mary didn't call?

P(J ^ ~M ^ A ^ ~B ^ ~E) = P(J|A) x P(~M|A) x P(A|~B^~E) x P(~B) x P(~E)
= 0.90 x 0.30 x 0.001 x 0.999 x 0.998 = 0.00027

Conditional independence

$$P(x_1,\ldots,x_n)$$
$$= P(x_n \mid x_{n-1},\ldots,x_1)P(x_{n-1},\ldots,x_1)$$
$$= P(x_n \mid x_{n-1},\ldots,x_1)P(x_{n-1} \mid x_{n-2},\ldots,x_1)$$
$$\ldots P(x_2 \mid x_1)P(x_1)$$

$$= \prod_{i=1}^{n} P(x_i \mid x_{i-1},\ldots,x_1)$$

- The belief network represents conditional independence:

$$P(X_i \mid X_i,\ldots,X_1) = P(X_i \mid Parents(X_i))$$

Inferences using belief networks

- **Diagnostic inferences (from effects to causes)**
  - **Given that JohnCalls, infer that**

    P(Burglary | JohnCalls) = 0.016

- **Causal inferences (from causes to effects)**
  - **Given Burglary, infer that**

    P(JohnCalls | Burglary) = 0.86

    and P(MaryCalls | Burglary) = 0.67

Inferences using belief networks

- **Intercausal inferences (between causes of a common effect)**
  - Given Alarm, we have

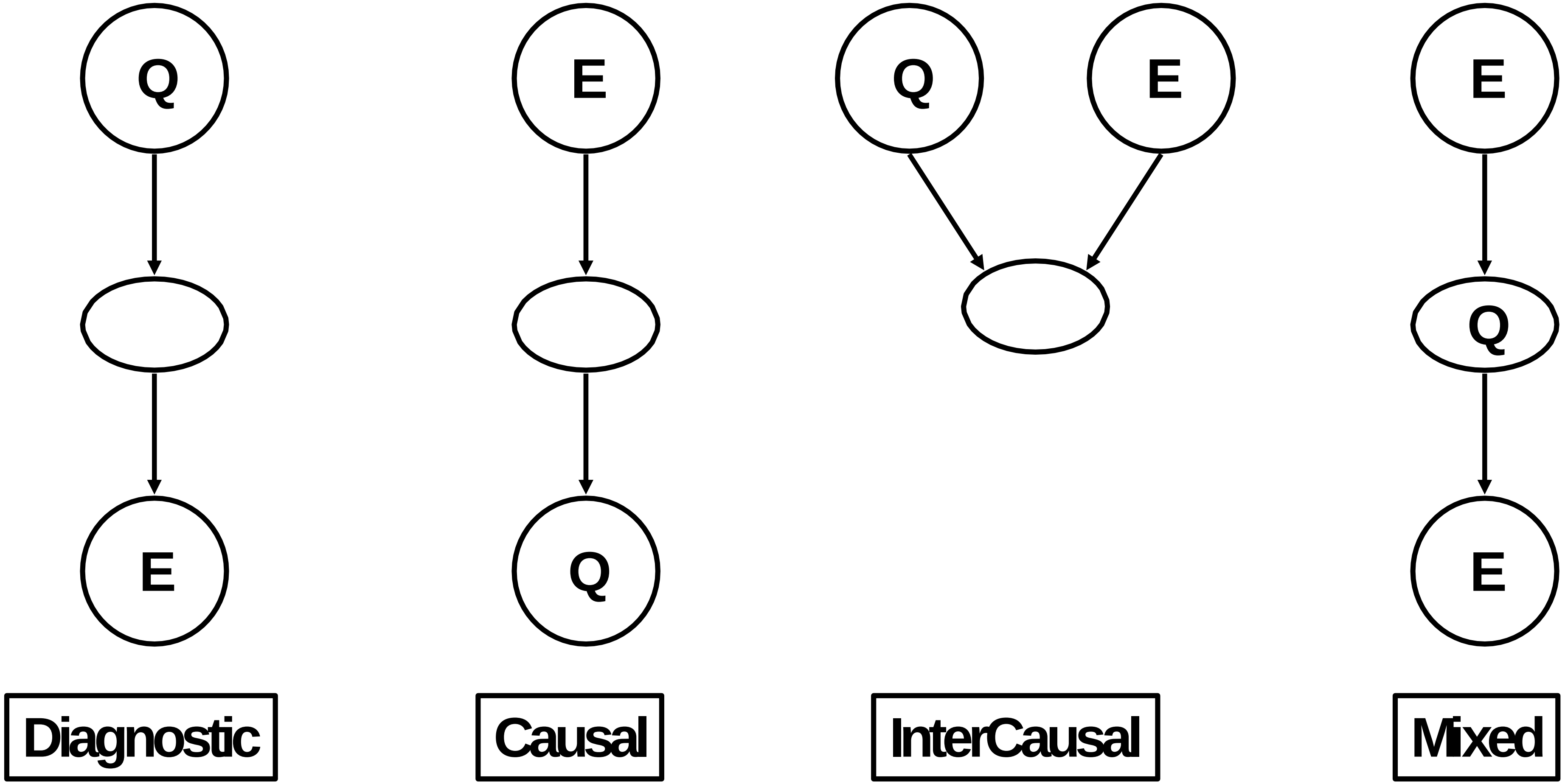
    $$P(Burglary \mid Alarm) = 0.376.$$

  - If we add evidence that Earthquake is true, then $P(Burglary \mid Alarm \wedge Earthquake)$ goes down to 0.003

- **Mixed inferences**
  - Setting the effect JohnCalls to true and the cause Earthquake to false gives

    $$P(Alarm \mid JohnCalls \wedge \neg Earthquake) = 0.003$$

The four patterns



Answering queries

- We consider cases where the belief network is a poly-tree
  - There is at most one undirected path between any two nodes