Bayes Nets: Construction and Conditional Independence

Outline

I. Construction of Bayes nets

II. Conditionally independent relations

III. Efficient representations of conditional distributions

* Figures are either from the textbook site or by the instructor.

* A few slides are based on lecture notes by Dr. Jin Tian.

I. Construction of the Bayesian Network

$$\boldsymbol{P}(X_i \mid X_{i-1}, \dots, X_1) = \boldsymbol{P}(X_i \mid \mathit{Parents}(X_i)) \quad \text{for } i = 2, \dots, n$$

The Bayesian network is correct only if X_i is conditionally independent of any X_j, $1 \le j \le i-1$, such that $X_j \notin \mathit{Parents}(X_i)$.

I. Construction of the Bayesian Network

$\boldsymbol{P}(X_i \mid X_{i-1}, \dots, X_1) = \boldsymbol{P}(X_i \mid \textit{Parents}(X_i))$ for $i = 2, \dots, n$

The Bayesian network is correct only if X_i is conditionally independent of any X_j, $1 \le j \le i-1$, such that $X_j \notin \textit{Parents}(X_i)$.

Construction algorithm

1. Determine the set of variables that are required to model the domain.

I. Construction of the Bayesian Network

$$\boldsymbol{P}(X_i \mid X_{i-1}, \dots, X_1) = \boldsymbol{P}(X_i \mid \textit{Parents}(X_i)) \quad \text{for } i = 2, \dots, n$$

The Bayesian network is correct only if X_i is conditionally independent of any X_j, $1 \le j \le i-1$, such that $X_j \notin \textit{Parents}(X_i)$.

Construction algorithm

1. Determine the set of variables that are required to model the domain.
2. Order them as $X_1, X_2, \dots, X_n$.

   Any order works, although network compactness depends on how much the order – whether causes precede effects – is respected.

I. Construction of the Bayesian Network

$\boldsymbol{P}(X_i \mid X_{i-1}, \dots, X_1) = \boldsymbol{P}(X_i \mid \mathit{Parents}(X_i))$ for $i = 2, \dots, n$

The Bayesian network is correct only if X_i is conditionally independent of any X_j, $1 \le j \le i-1$, such that $X_j \notin \mathit{Parents}(X_i)$.

Construction algorithm

1. Determine the set of variables that are required to model the domain.
2. Order them as $X_1, X_2, \dots, X_n$.

   Any order works, although network compactness depends on how much the order – whether causes precede effects – is respected.
3. For $i = 1$ to n do

   a) Chose a minimal set of parents for X_i from $X_1, X_2, \dots, X_{i-1}$ such that

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

I. Construction of the Bayesian Network

$\boldsymbol{P}(X_i \mid X_{i-1}, \ldots, X_1) = \boldsymbol{P}(X_i \mid \mathit{Parents}(X_i))$ for $i = 2, \ldots, n$

The Bayesian network is correct only if X_i is conditionally independent of any X_j, $1 \le j \le i-1$, such that $X_j \notin \mathit{Parents}(X_i)$.

Construction algorithm

1. Determine the set of variables that are required to model the domain.

2. Order them as $X_1, X_2, \ldots, X_n$.

   Any order works, although network compactness depends on how much the order – whether causes precede effects – is respected.

3. For $i = 1$ to n do

   a) Chose a minimal set of parents for X_i from $X_1, X_2, \ldots, X_{i-1}$ such that

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

   b) Add a directed edge from every parent to X_i.

I. Construction of the Bayesian Network

$$\boldsymbol{P}(X_i \mid X_{i-1}, \dots, X_1) = \boldsymbol{P}(X_i \mid \mathit{Parents}(X_i)) \quad \text{for } i = 2, \dots, n$$

The Bayesian network is correct only if X_i is conditionally independent of any X_j, $1 \le j \le i-1$, such that $X_j \notin \mathit{Parents}(X_i)$.

Construction algorithm

1. Determine the set of variables that are required to model the domain.
2. Order them as $X_1, X_2, \dots, X_n$.

   Any order works, although network compactness depends on how much the order – whether causes precede effects – is respected.
3. For $i = 1$ to n do

   a) Chose a minimal set of parents for X_i from $X_1, X_2, \dots, X_{i-1}$ such that

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

   b) Add a directed edge from every parent to X_i.

   c) Write down the conditional probability table (CPT), $\boldsymbol{P}(X_i \mid \mathit{Parents}(X_i))$.

Construction (cont'd)

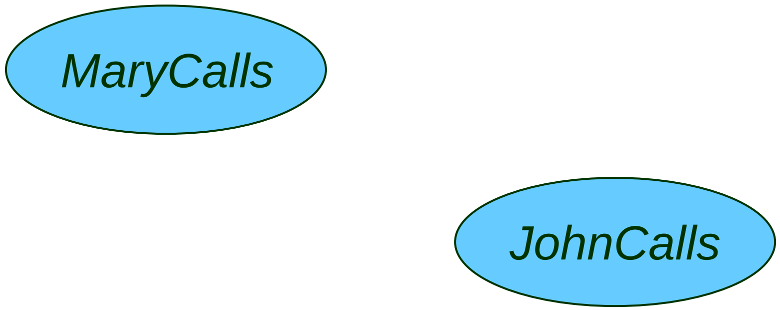
Chosen order: *MaryCalls*, *JohnCalls*, *Alarm*, *Burglary*, *Earthquake*.

Construction (cont'd)

Chosen order: *MaryCalls*, *JohnCalls*, *Alarm*, *Burglary*, *Earthquake*.

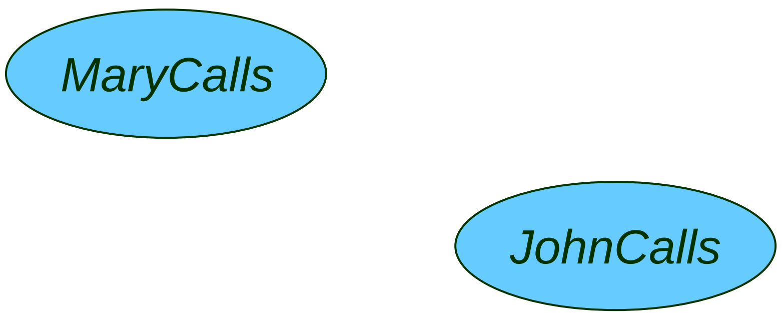
Construction (cont'd)

Chosen order: *MaryCalls*, *JohnCalls*, *Alarm*, *Burglary*, *Earthquake*.



Construction (cont'd)

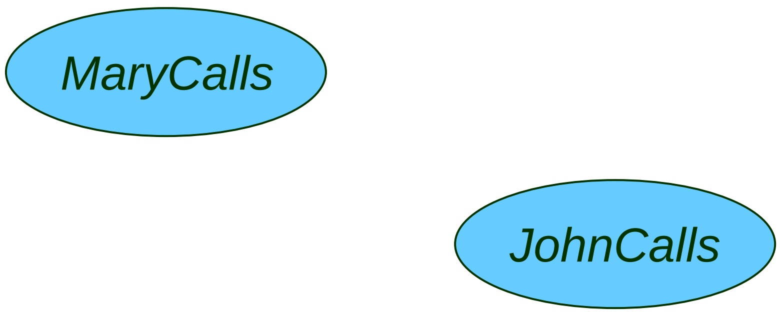
Chosen order: *MaryCalls*, *JohnCalls*, *Alarm*, *Burglary*, *Earthquake*.



$P(j \mid m)$ $\quad$ $P(j)$

Construction (cont’d)

Chosen order: *MaryCalls*, *JohnCalls*, *Alarm*, *Burglary*, *Earthquake*.

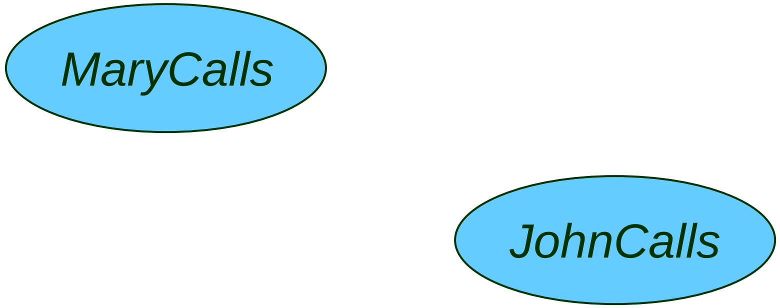


$P(j \mid m)$ $P(j)$

// If May calls, that probably means
// the alarm has gone off, which
// makes John more likely to call.

Construction (cont'd)

Chosen order: *MaryCalls*, *JohnCalls*, *Alarm*, *Burglary*, *Earthquake*.

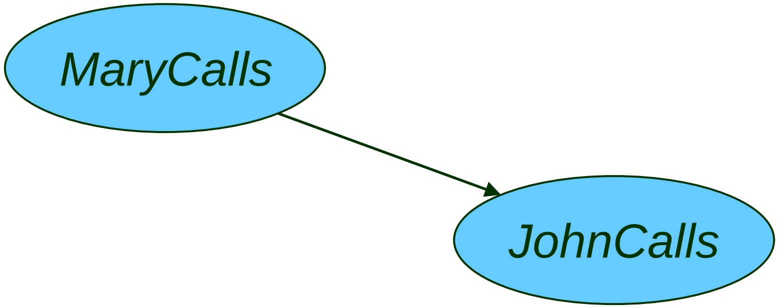


$$P(j \mid m) > P(j)$$

// If May calls, that probably means
// the alarm has gone off, which
// makes John more likely to call.

Construction (cont'd)

Chosen order: *MaryCalls*, *JohnCalls*, *Alarm*, *Burglary*, *Earthquake*.

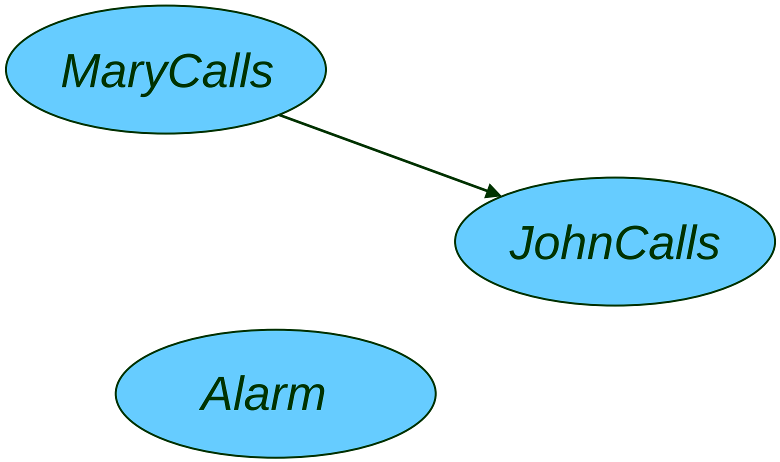


$$P(j \mid m) > P(j)$$

// If May calls, that probably means
// the alarm has gone off, which
// makes John more likely to call.

Construction (cont'd)

Chosen order: *MaryCalls*, *JohnCalls*, *Alarm*, *Burglary*, *Earthquake*.

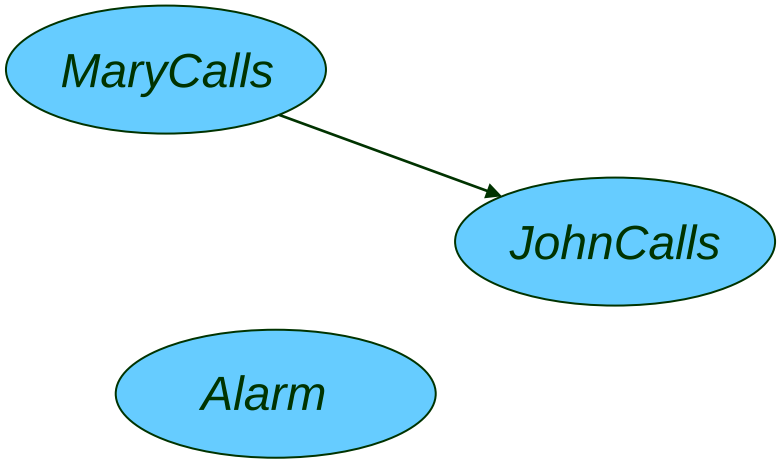


$$P(j \mid m) > P(j)$$

// If May calls, that probably means
// the alarm has gone off, which
// makes John more likely to call.

Construction (cont'd)

Chosen order: *MaryCalls*, *JohnCalls*, *Alarm*, *Burglary*, *Earthquake*.



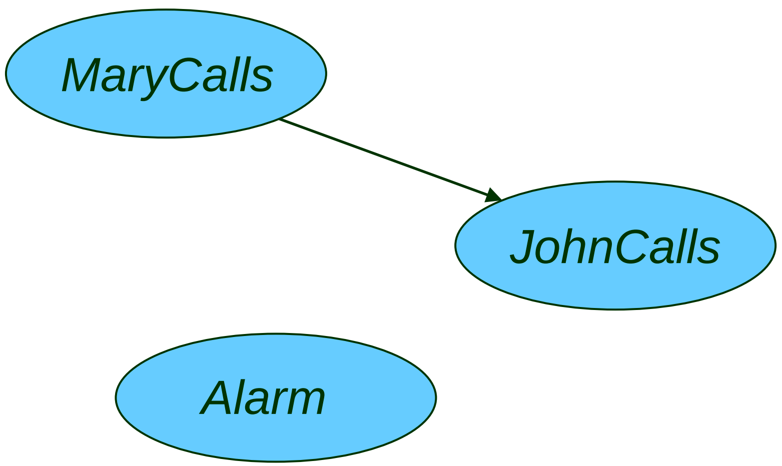
$P(j \mid m) > P(j)$

// If May calls, that probably means
// the alarm has gone off, which
// makes John more likely to call.

$P(a \mid m, j) \quad P(a \mid j), P(a \mid m), P(a)$

Construction (cont'd)

Chosen order: *MaryCalls*, *JohnCalls*, *Alarm*, *Burglary*, *Earthquake*.



$P(j \mid m) > P(j)$

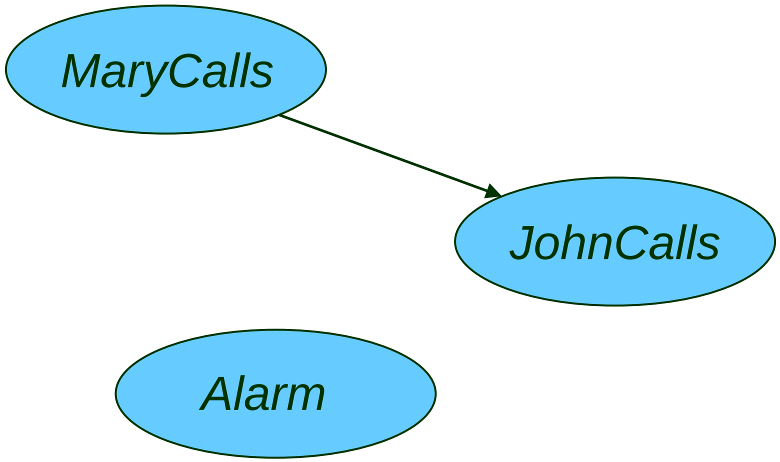
// If May calls, that probably means
// the alarm has gone off, which
// makes John more likely to call.

$P(a \mid m, j) \qquad P(a \mid j), P(a \mid m), P(a)$

// If both Mary and John call, the alarm
// is more likely to go off than if just
// one calls.

Construction (cont'd)

Chosen order: *MaryCalls*, *JohnCalls*, *Alarm*, *Burglary*, *Earthquake*.



$P(j \mid m) > P(j)$

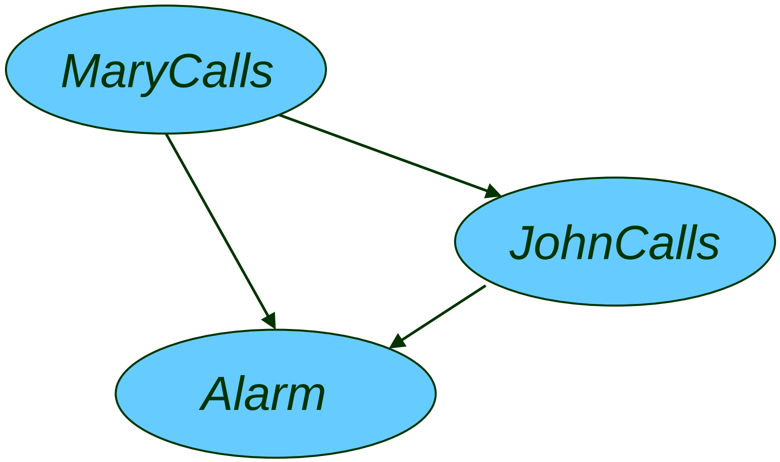
// If May calls, that probably means
// the alarm has gone off, which
// makes John more likely to call.

$P(a \mid m, j) > P(a \mid j), P(a \mid m), P(a)$

// If both Mary and John call, the alarm
// is more likely to go off than if just
// one calls.

Construction (cont'd)

Chosen order: *MaryCalls*, *JohnCalls*, *Alarm*, *Burglary*, *Earthquake*.



$P(j \mid m) > P(j)$

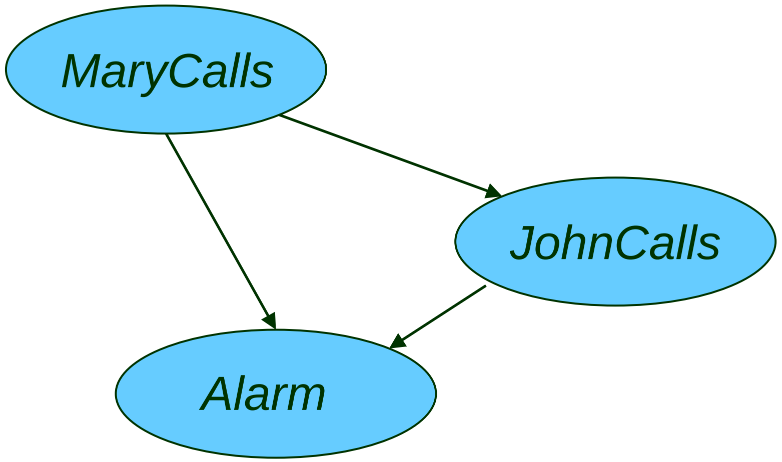
// If May calls, that probably means
// the alarm has gone off, which
// makes John more likely to call.

$P(a \mid m, j) > P(a \mid j), P(a \mid m), P(a)$

// If both Mary and John call, the alarm
// is more likely to go off than if just
// one calls.

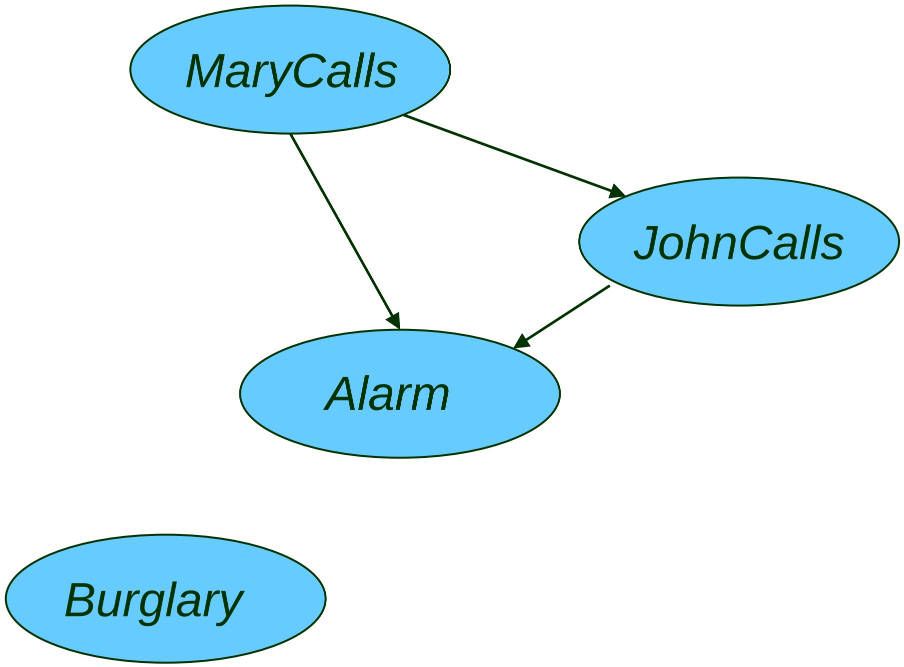
Construction for the Burglary Example

MaryCalls, *JohnCalls*, *Alarm*, *Burglary*, *Earthquake*.



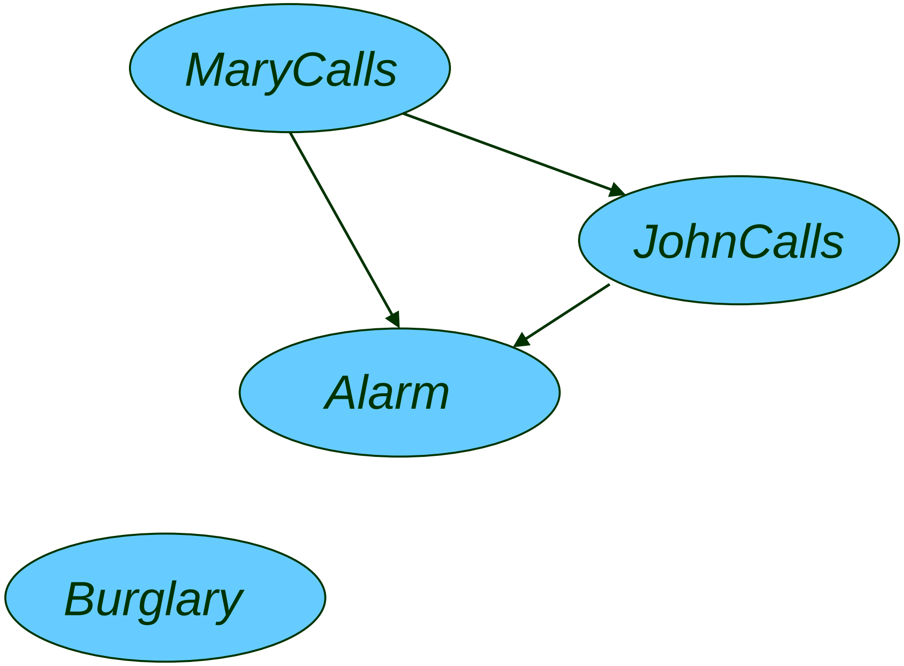
Construction for the Burglary Example

MaryCalls, *JohnCalls*, *Alarm*, *Burglary*, *Earthquake*.



Construction for the Burglary Example

MaryCalls, *JohnCalls*, *Alarm*, *Burglary*, *Earthquake*.

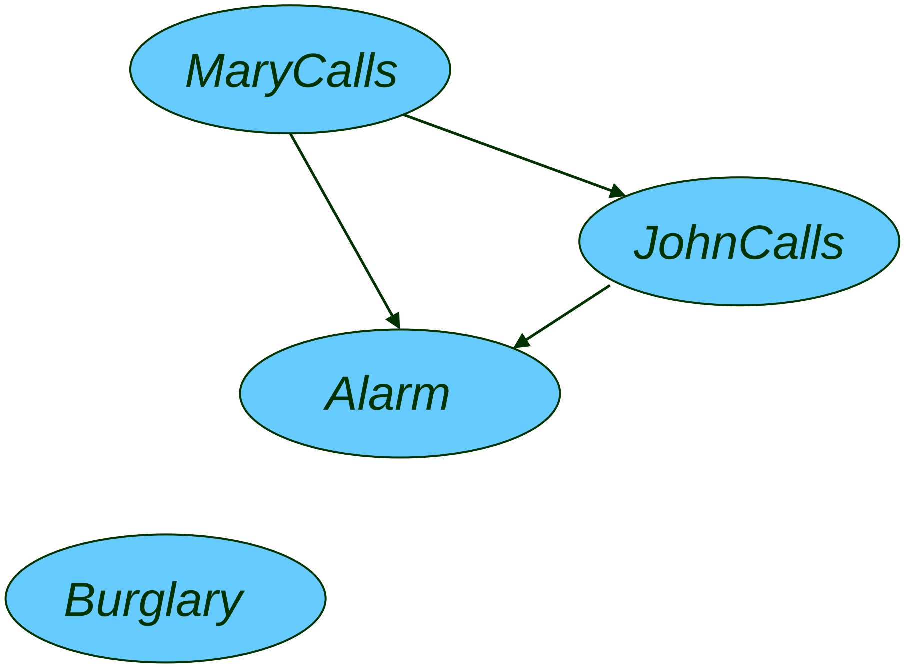


$\boldsymbol{P}(B \mid A, J, M) \qquad \boldsymbol{P}(B \mid A)$

// If the value of A (either a or $\neg a$) is
// known, then the call from John or
// Mary does not add any information
// about burglary.

Construction for the Burglary Example

MaryCalls, *JohnCalls*, *Alarm*, *Burglary*, *Earthquake*.

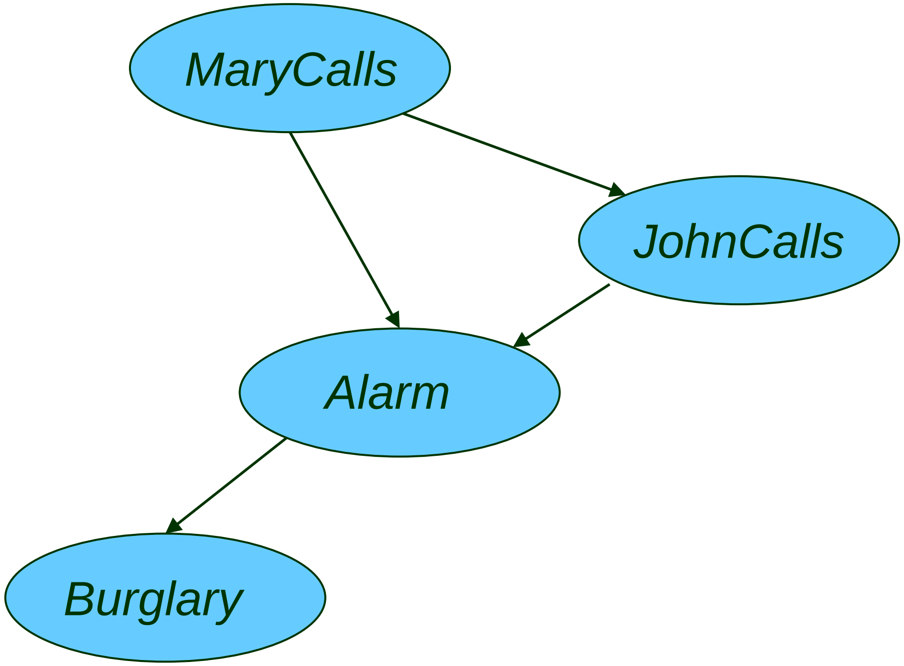


$$\boldsymbol{P}(B \mid A, J, M) = \boldsymbol{P}(B \mid A)$$

// If the value of A (either a or $\neg a$) is
// known, then the call from John or
// Mary does not add any information
// about burglary.

Construction for the Burglary Example

MaryCalls, *JohnCalls*, *Alarm*, *Burglary*, *Earthquake*.

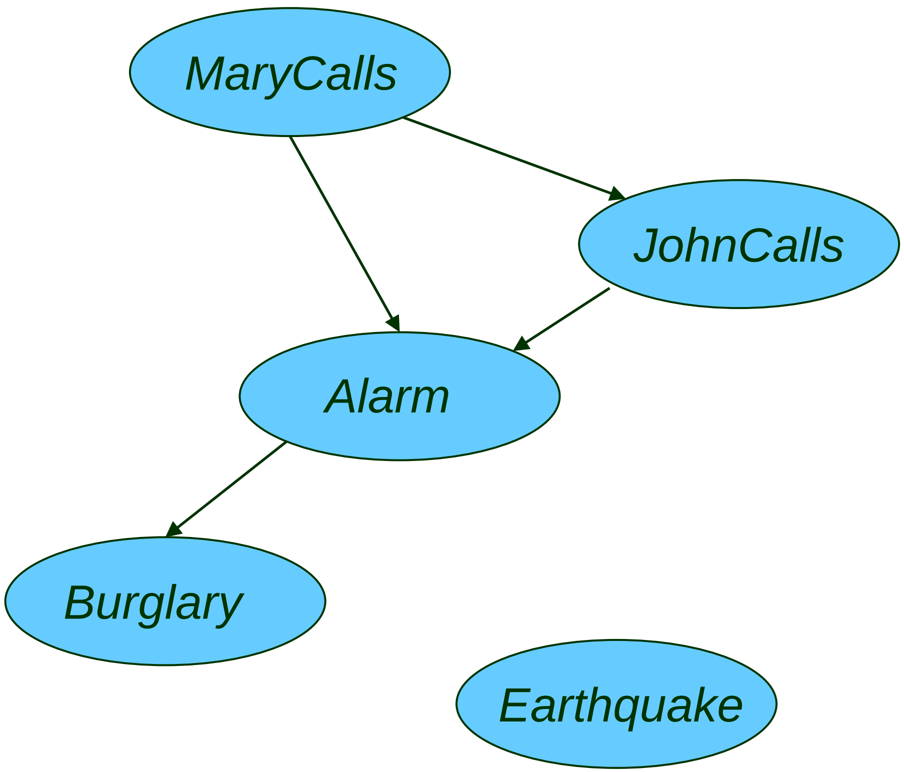


$$\boldsymbol{P}(B \mid A, J, M) = \boldsymbol{P}(B \mid A)$$

// If the value of A (either a or $\neg a$) is
// known, then the call from John or
// Mary does not add any information
// about burglary.

Construction for the Burglary Example

MaryCalls, *JohnCalls*, *Alarm*, *Burglary*, *Earthquake*.

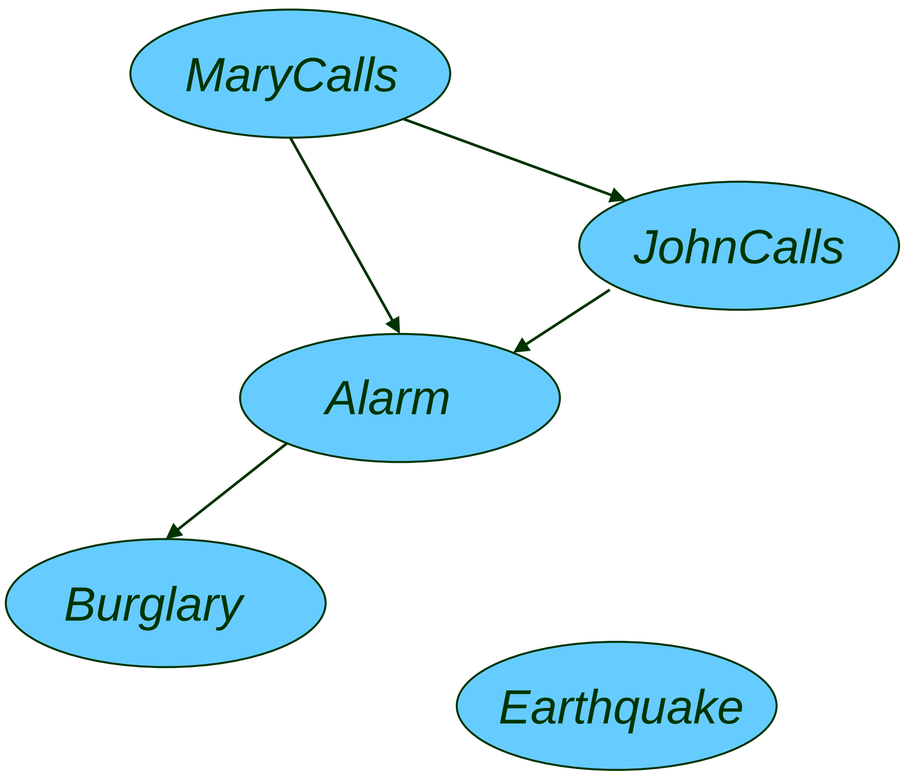


$$\boldsymbol{P}(B \mid A, J, M) = \boldsymbol{P}(B \mid A)$$

// If the value of A (either a or $\neg a$) is
// known, then the call from John or
// Mary does not add any information
// about burglary.

Construction for the Burglary Example

MaryCalls, *JohnCalls*, *Alarm*, *Burglary*, *Earthquake*.



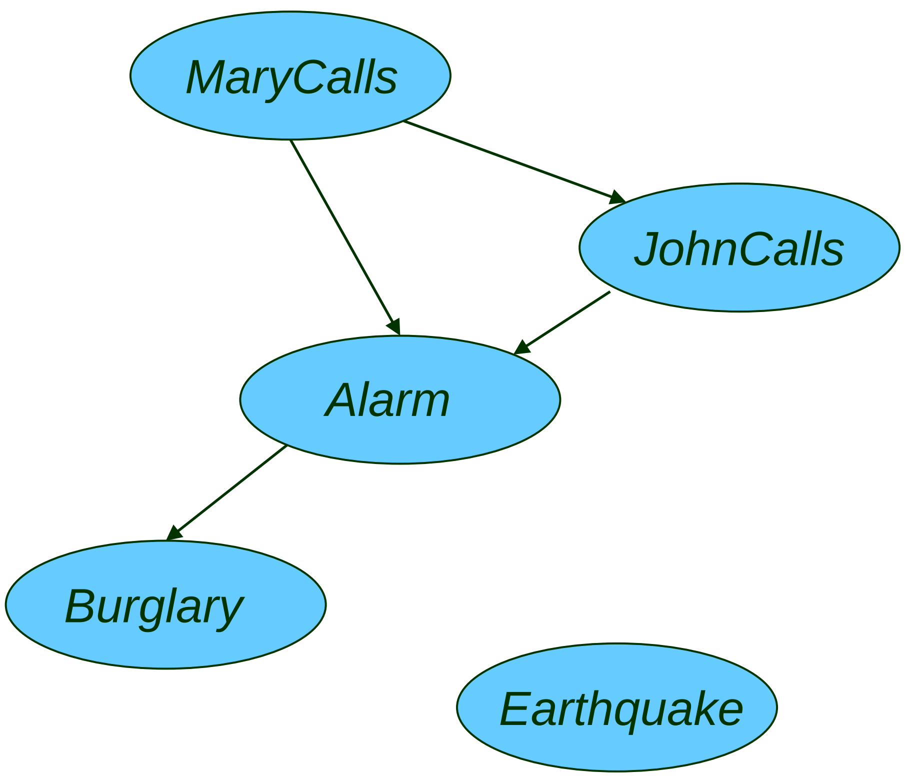
$$\boldsymbol{P}(B \mid A, J, M) = \boldsymbol{P}(B \mid A)$$

// If the value of A (either a or $\neg a$) is
// known, then the call from John or
// Mary does not add any information
// about burglary.

$$P(e \mid a, b) \qquad P(e \mid a), P(e \mid b)$$

Construction for the Burglary Example

MaryCalls, *JohnCalls*, *Alarm*, *Burglary*, *Earthquake*.



$$\boldsymbol{P}(B \mid A, J, M) = \boldsymbol{P}(B \mid A)$$

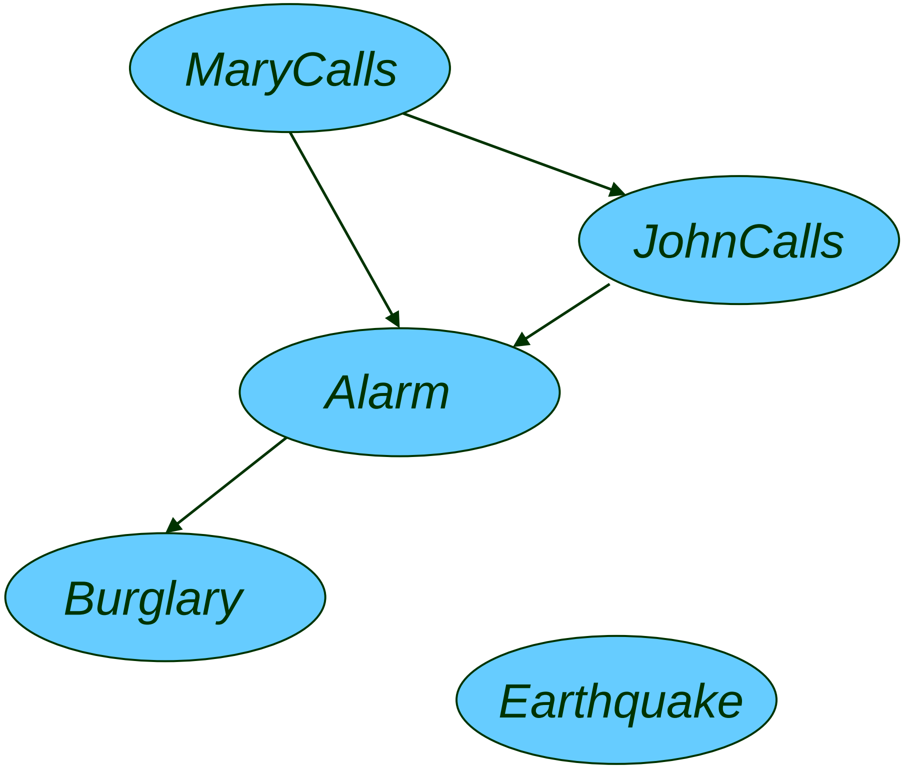
// If the value of A (either a or $\neg a$) is
// known, then the call from John or
// Mary does not add any information
// about burglary.

$$P(e \mid a, b) \qquad P(e \mid a), P(e \mid b)$$

// If the alarm is on, it is more likely that
// there has been earthquake. If there
// has been a burglary, it is slightly more
// likely that it happened after an
// earthquake. In the occurrences of
// both events, the chance of earthquake
// occurrence is even higher.

Construction for the Burglary Example

MaryCalls, *JohnCalls*, *Alarm*, *Burglary*, *Earthquake*.



$$\boldsymbol{P}(B \mid A, J, M) = \boldsymbol{P}(B \mid A)$$

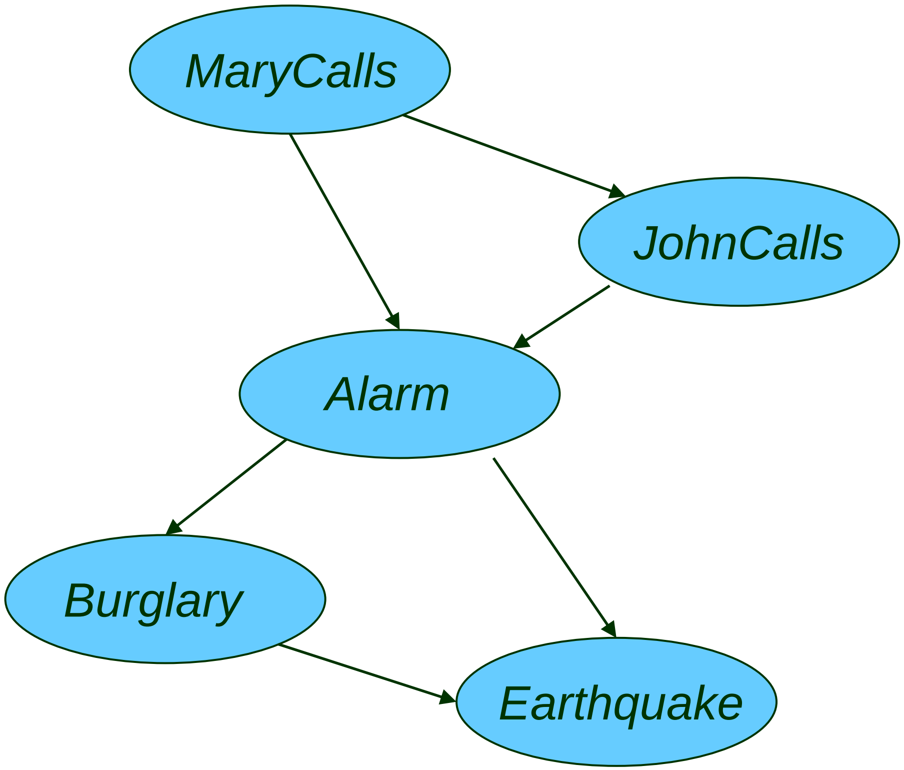
// If the value of A (either a or $\neg a$) is
// known, then the call from John or
// Mary does not add any information
// about burglary.

$$P(e \mid a, b) > P(e \mid a), P(e \mid b)$$

// If the alarm is on, it is more likely that
// there has been earthquake. If there
// has been a burglary, it is slightly more
// likely that it happened after an
// earthquake. In the occurrences of
// both events, the chance of earthquake
// occurrence is even higher.

Construction for the Burglary Example

MaryCalls, *JohnCalls*, *Alarm*, *Burglary*, *Earthquake*.



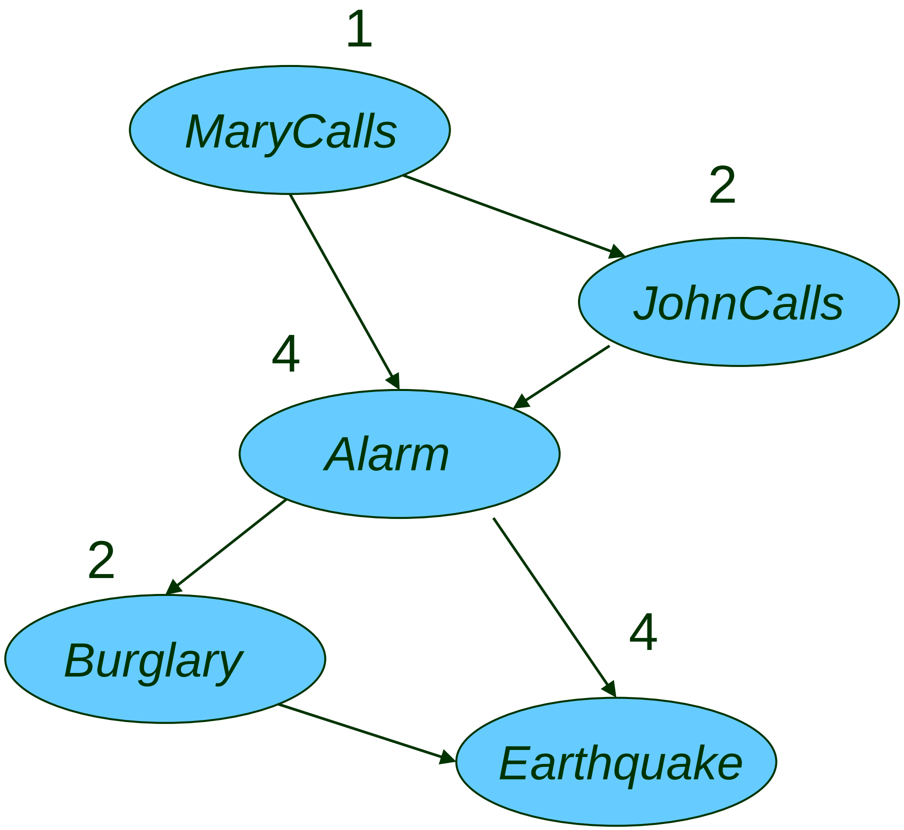
$$\boldsymbol{P}(B \mid A, J, M) = \boldsymbol{P}(B \mid A)$$

// If the value of A (either a or $\neg a$) is
// known, then the call from John or
// Mary does not add any information
// about burglary.

$$P(e \mid a, b) > P(e \mid a), P(e \mid b)$$

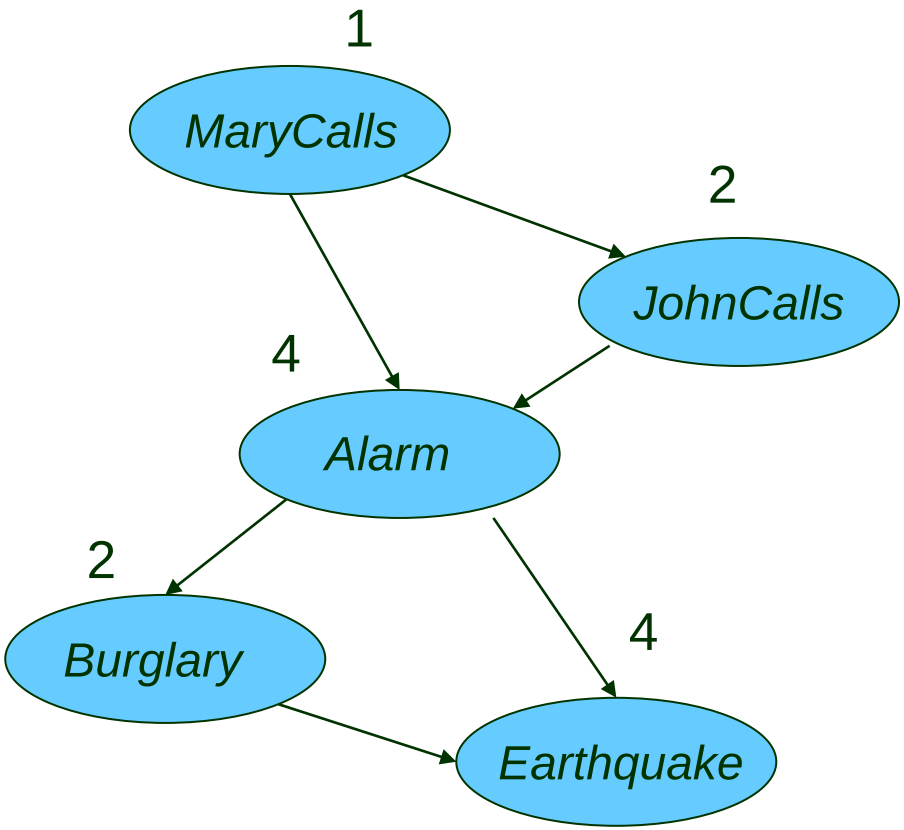
// If the alarm is on, it is more likely that
// there has been earthquake. If there
// has been a burglary, it is slightly more
// likely that it happened after an
// earthquake. In the occurrences of
// both events, the chance of earthquake
// occurrence is even higher.

Node Ordering Matters



$1 + 2 + 4 + 2 + 4 = 13$
conditional probabilities

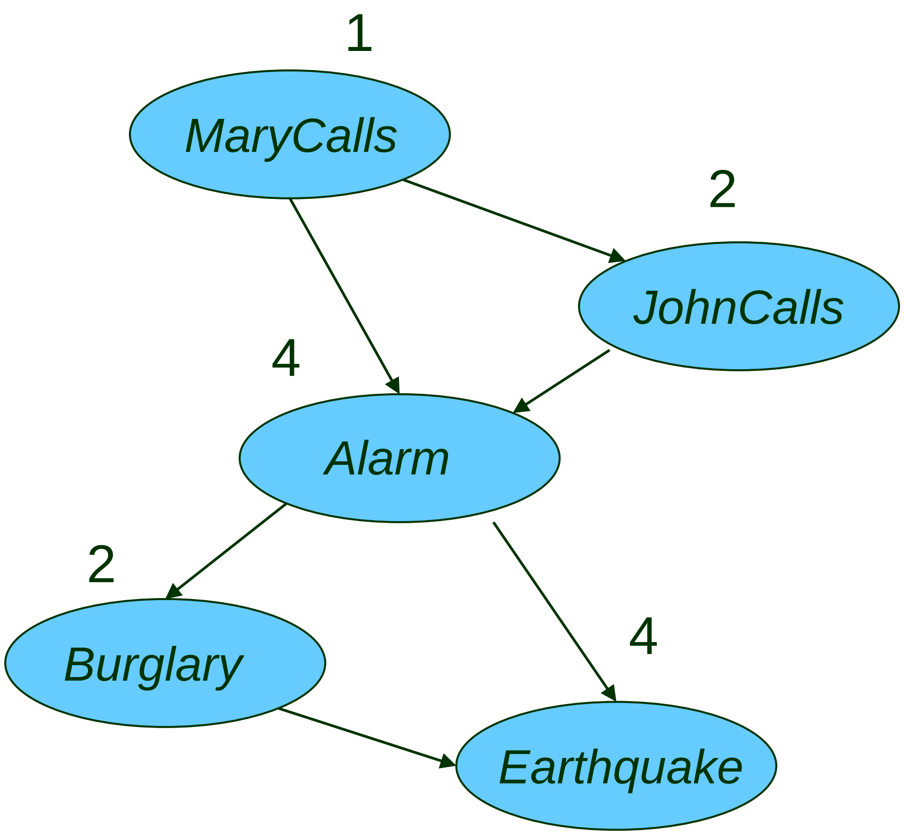
Node Ordering Matters



$1 + 2 + 4 + 2 + 4 = 13$
conditional probabilities

- More conditional probabilities than needed.

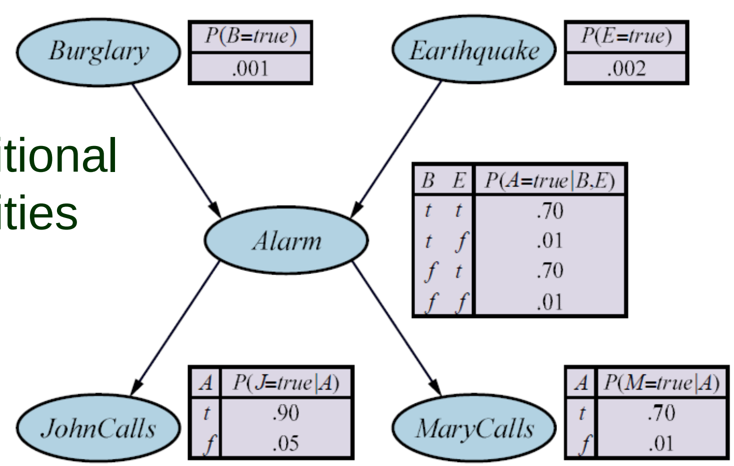
Node Ordering Matters



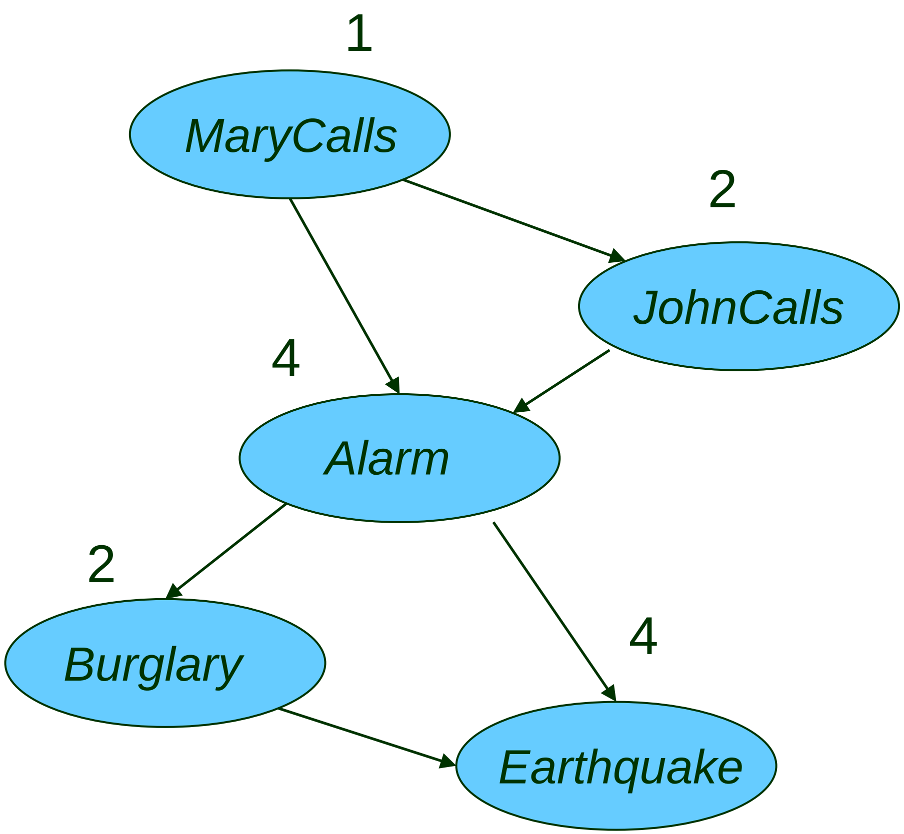
$1 + 2 + 4 + 2 + 4 = 13$
conditional probabilities

♣ More conditional probabilities than needed.

10 conditional probabilities



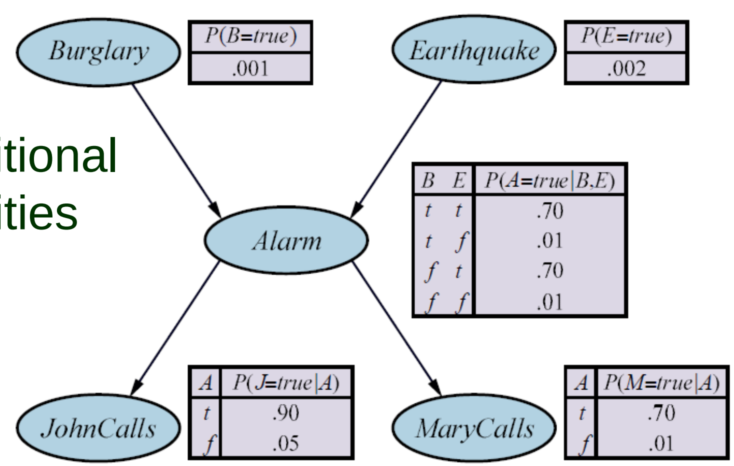
Node Ordering Matters



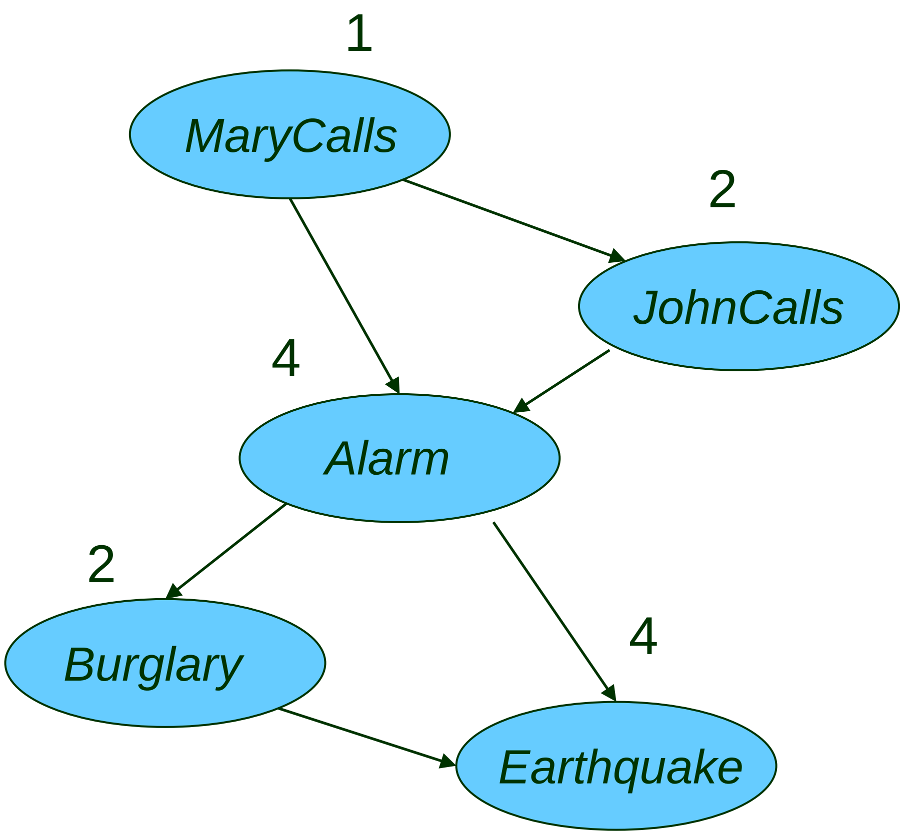
- More conditional probabilities than needed.
- Assessment of unnatural probabilities, e.g., $\boldsymbol{P}(\mathit{Earthquake} \mid \mathit{Burglary}, \mathit{Alarm}\}$.

$1 + 2 + 4 + 2 + 4 = 13$
conditional probabilities

10 conditional probabilities



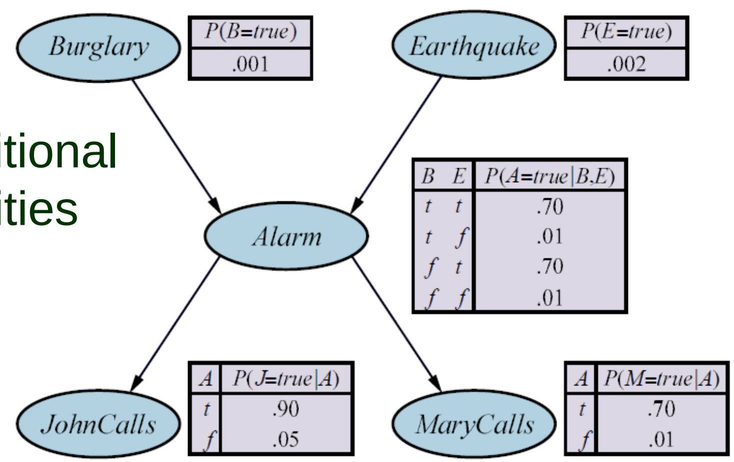
Node Ordering Matters



$1+2+4+2+4=13$
conditional probabilities

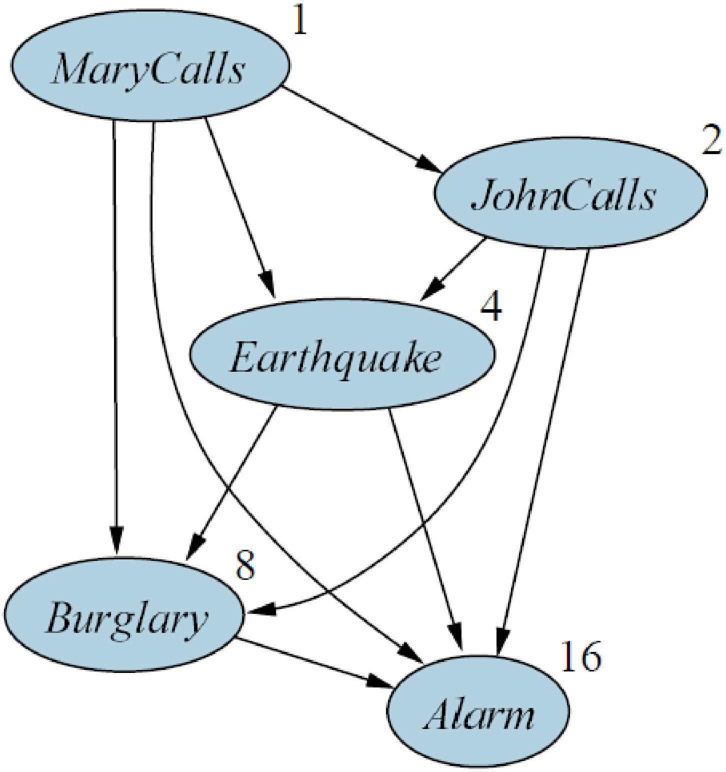
- ♣ More conditional probabilities than needed.
- ♣ Assessment of unnatural probabilities, e.g., $\boldsymbol{P}(\textit{Earthquake} \mid \textit{Burglary, Alarm})$.
- ♦ Sticking to a causal model results in fewer probabilities that are also easier to come up with.

10 conditional probabilities



Bad Node Ordering

MaryCalls, *JohnCalls*, *Earthquake, Burglary*, *Alarm*.



$1 + 2 + 4 + 8 + 16 = 31$ distinct probabilities (exactly the same as the full joint distribution)!

Roles of Casualty

- Deciding conditional independence is hard in noncausal directions.
(Causal models and conditional independence seem hardwired for humans!)

- Assessing conditional probabilities is hard in noncausal directions.

- The interpretation of directed acyclic graphs as carriers of independence assumptions does not necessarily imply causation

- The ubiquity of DAG models in statistical and AI applications stems (often unwittingly) primarily from their causal interpretation.

- In practice, DAG models are rarely used in any variable ordering other than those which respect the direction of time and causation.

Compactness Of Bayes Nets

- The full joint distribution contains 2^n numbers.

Compactness Of Bayes Nets

- ♠ The full joint distribution contains 2^n numbers.

- It is reasonable to assume that each random variable is directly influenced by $\leq k$ others (i.e., every node has $\leq k$ parents in a BN).

Compactness Of Bayes Nets

♠ The full joint distribution contains 2^n numbers.

- It is reasonable to assume that each random variable is directly influenced by $\leq k$ others (i.e., every node has $\leq k$ parents in a BN).

- The conditional probability table (CPT) for each node has size $\leq 2^k$.

Compactness Of Bayes Nets

- ♠ The full joint distribution contains 2^n numbers.

- It is reasonable to assume that each random variable is directly influenced by $\leq k$ others (i.e., every node has $\leq k$ parents in a BN).

- The conditional probability table (CPT) for each node has size $\leq 2^k$.

- With n Boolean variables, the network has $\leq n \cdot 2^k$ numbers.

Compactness Of Bayes Nets

♠ The full joint distribution contains 2^n numbers.

- It is reasonable to assume that each random variable is directly influenced by $\leq k$ others (i.e., every node has $\leq k$ parents in a BN).

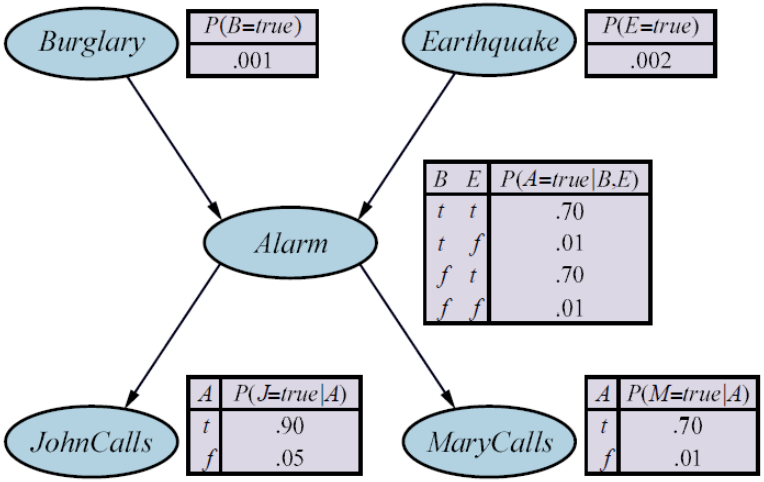
- The conditional probability table (CPT) for each node has size $\leq 2^k$.

- With n Boolean variables, the network has $\leq n \cdot 2^k$ numbers.

♦ To avoid a fully connected network, leave out links that represent slight dependencies.

II. Non-Descendants Property

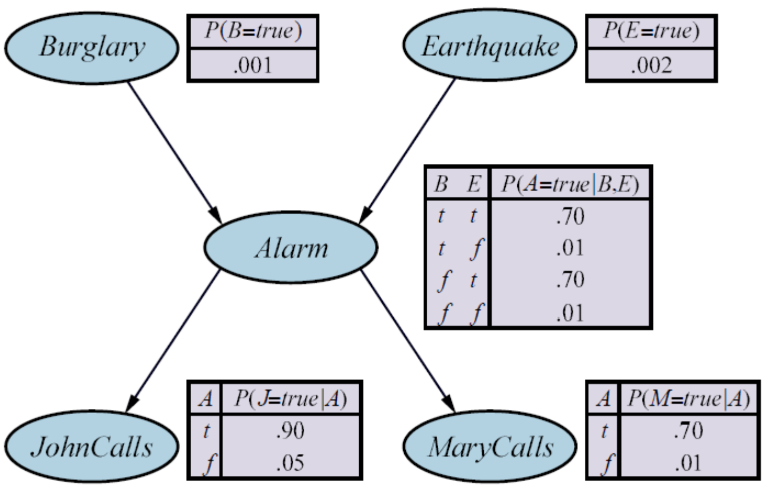
Every variable is conditionally independent of its non-descendants, given the values of its parents.



Given the value of *Alarm*, *JohnCalls* is independent of *Burglary*, *Earthquake*, and *Marycalls*.

II. Non-Descendants Property

Every variable is conditionally independent of its non-descendants, given the values of its parents.

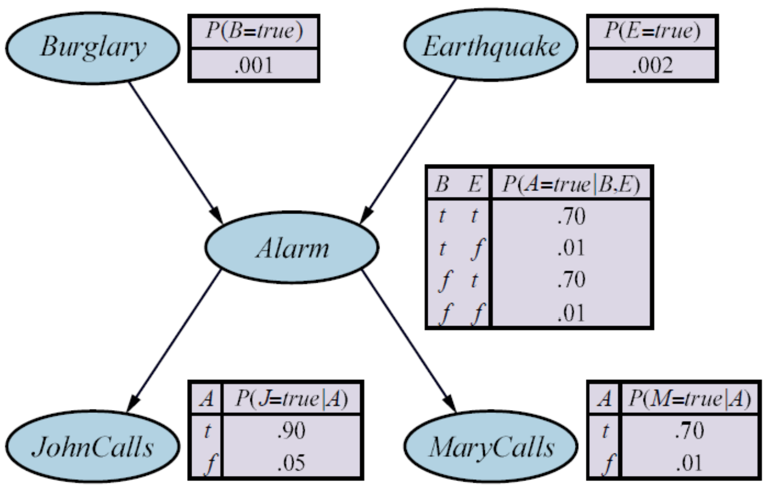


Given the value of *Alarm*, *JohnCalls* is independent of *Burglary*, *Earthquake*, and *Marycalls*.

Non-descendants property

II. Non-Descendants Property

Every variable is conditionally independent of its non-descendants, given the values of its parents.



Given the value of *Alarm*, *JohnCalls* is independent of *Burglary*, *Earthquake*, and *Marycalls*.

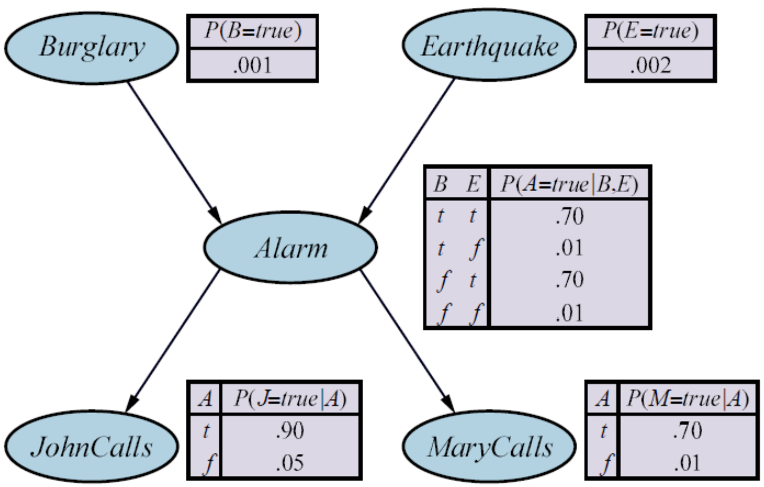
Non-descendants property

// network parameter interpretation

$$P(x_i \mid Parents(X_i)) = \theta_i(x_i \mid parents(X_i))$$

II. Non-Descendants Property

Every variable is conditionally independent of its non-descendants, given the values of its parents.



P(B=true)
.001

P(E=true)
.002

B	E	P(A=true | B,E)
t	t	.70
t	f	.01
f	t	.70
f	f	.01

A	P(J=true | A)
t	.90
f	.05

A	P(M=true | A)
t	.70
f	.01

Given the value of *Alarm*, *JohnCalls* is independent of *Burglary*, *Earthquake*, and *Marycalls*.

// network parameter interpretation

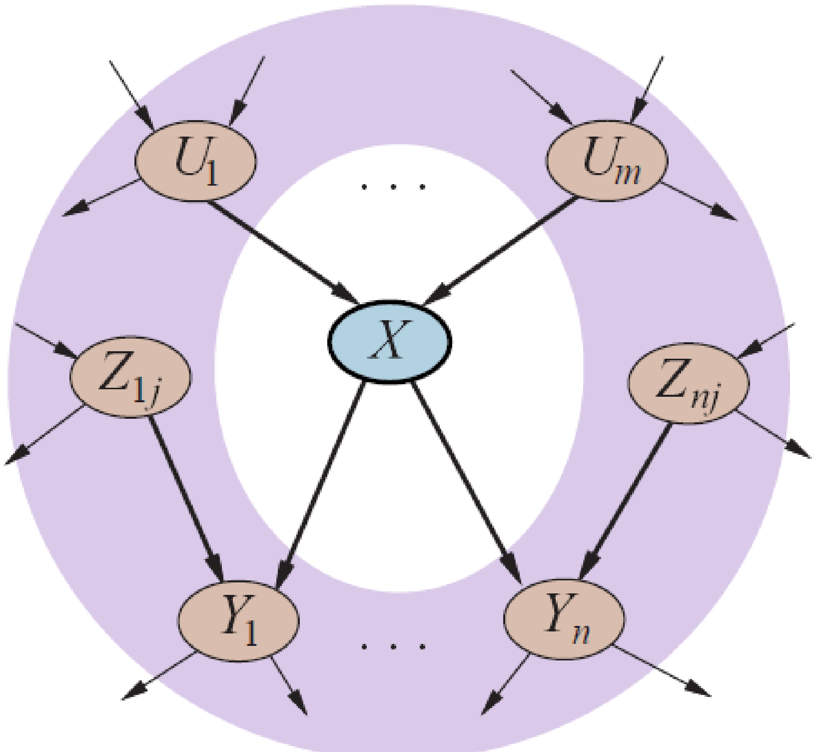
$$\underbrace{\text{Non-descendants property} \quad P(x_i \mid \mathit{Parents}(X_i)) = \theta_i(x_i \mid \mathit{parents}(X_i))}$$

$$\Downarrow$$

The full joint distribution $P(x_1, \dots, x_n) = \prod_{i=1}^{n} P(x_i \mid \mathit{Parents}(X_i))$

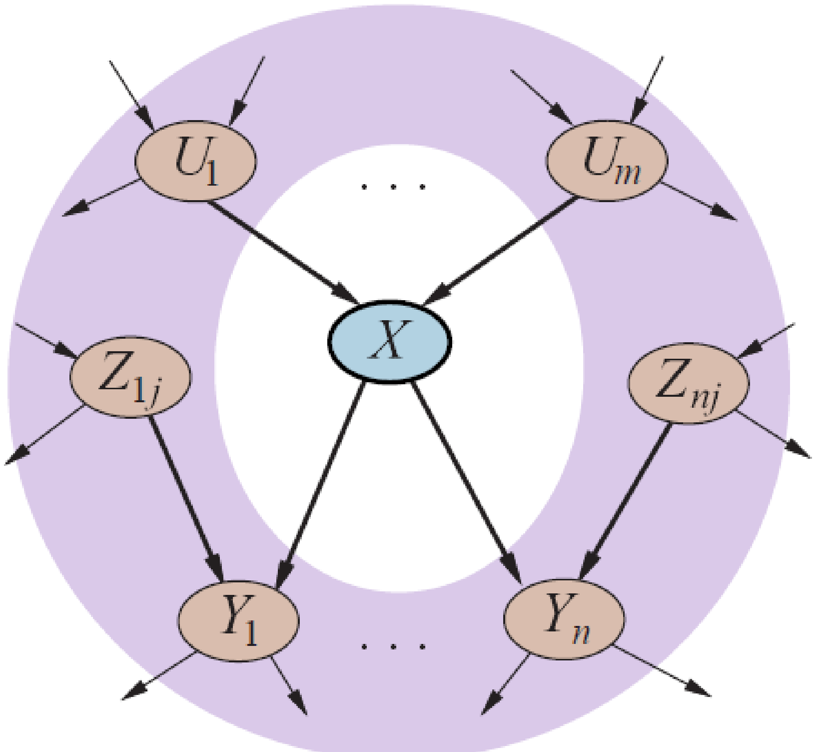
Markov Blanket

The *Markov blanket* of a node consists of its parents, children, and children's parents.



Markov Blanket

The *Markov blanket* of a node consists of its parents, children, and children's parents.



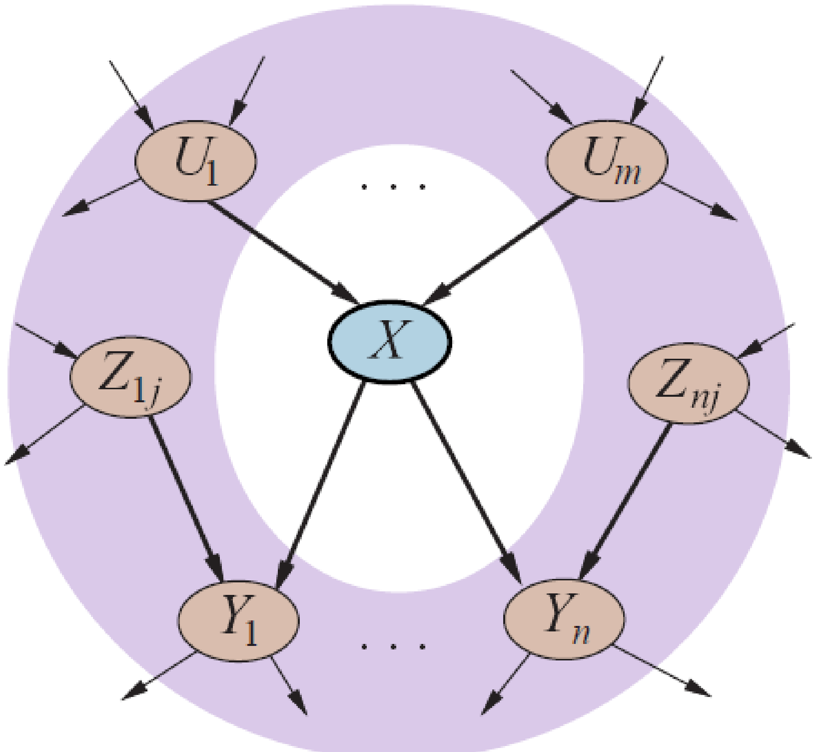
non-descendants property

⇩

A node is conditionally independent of all other nodes given its Markov blanket.

Markov Blanket

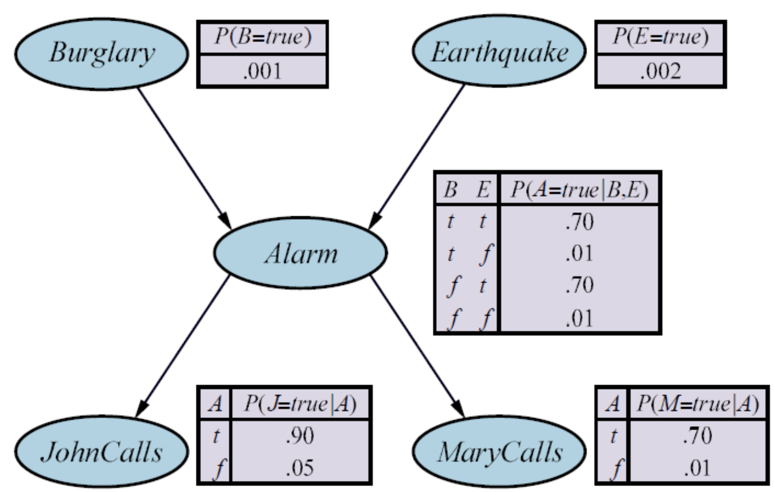
The *Markov blanket* of a node consists of its parents, children, and children's parents.



non-descendants property

⇩

A node is conditionally independent of all other nodes given its Markov blanket.

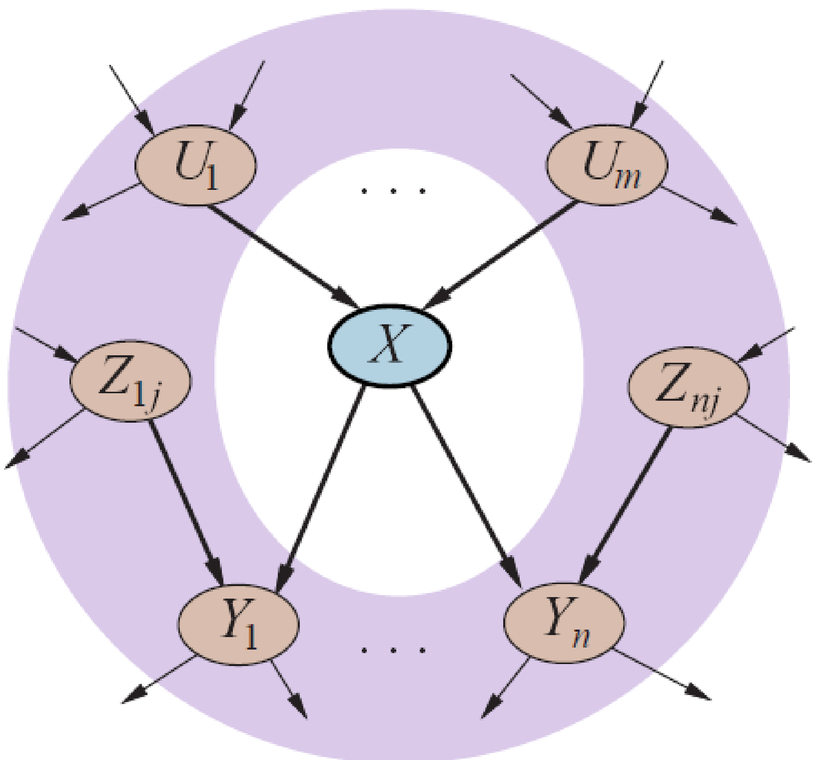


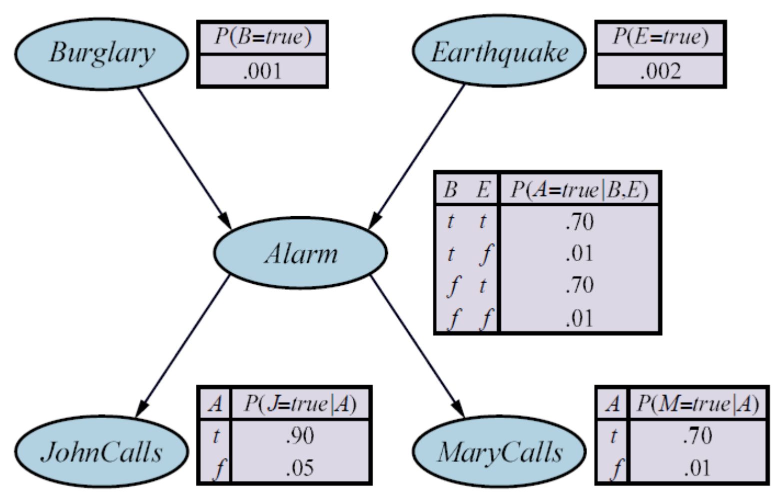
Markov Blanket

The *Markov blanket* of a node consists of its parents, children, and children's parents.

non-descendants property

A node is conditionally independent of all other nodes given its Markov blanket.





Given *Alarm* and *Earthquake*, *Burglary* is independent of *JohnCalls* and *Marycalls*.

D-Separation

Q: Is a set of nodes X conditionally independent of another set Y, given a third set Z?

D-Separation

Q: Is a set of nodes X conditionally independent of another set Y, given a third set Z?

This question can be answered as follows:

1. Start with the *ancestral subgraph* consisting of X, Y, Z, and their ancestors (and edges between them).

D-Separation

Q: Is a set of nodes X conditionally independent of another set Y, given a third set Z?

This question can be answered as follows:

1. Start with the *ancestral subgraph* consisting of X, Y, Z, and their ancestors (and edges between them).
2. Replace all directed edges with undirected edges.

D-Separation

Q: Is a set of nodes X conditionally independent of another set Y, given a third set Z?

This question can be answered as follows:

1. Start with the *ancestral subgraph* consisting of X, Y, Z, and their ancestors (and edges between them).
2. Replace all directed edges with undirected edges.
3. Add an (undirected) edge between every two nodes that share a common child. The resulting graph is the *moral graph* of the ancestral subgraph.

D-Separation

Q: Is a set of nodes X conditionally independent of another set Y, given a third set Z?

This question can be answered as follows:

1. Start with the *ancestral subgraph* consisting of X, Y, Z, and their ancestors (and edges between them).
2. Replace all directed edges with undirected edges.
3. Add an (undirected) edge between every two nodes that share a common child. The resulting graph is the *moral graph* of the ancestral subgraph.
   a. If Z blocks all paths between X and Y in the moral graph, then Z *d-separates* X and Y. In this case, X is conditionally independent of Y, given Z.

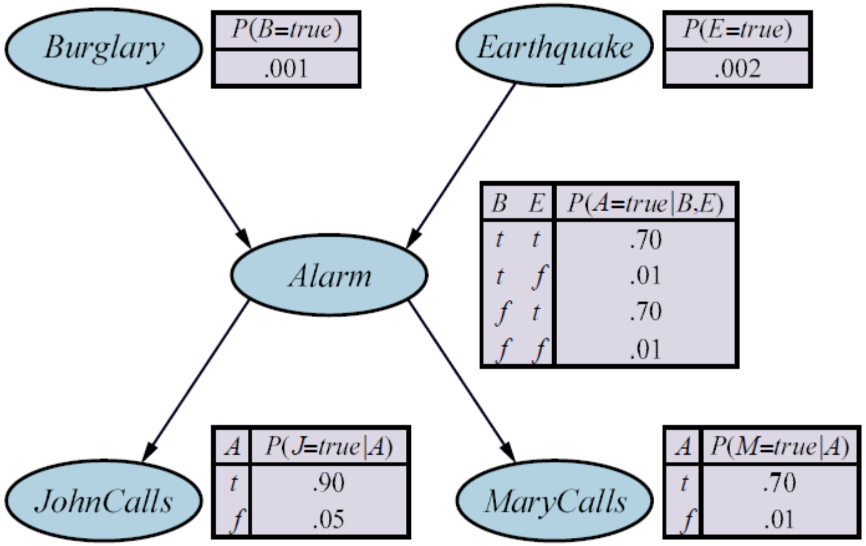
D-Separation

Q: Is a set of nodes X conditionally independent of another set Y, given a third set Z?

This question can be answered as follows:

1. Start with the *ancestral subgraph* consisting of X, Y, Z, and their ancestors (and edges between them).
2. Replace all directed edges with undirected edges.
3. Add an (undirected) edge between every two nodes that share a common child. The resulting graph is the *moral graph* of the ancestral subgraph.
   a. If Z blocks all paths between X and Y in the moral graph, then Z *d-separates* X and Y. In this case, X is conditionally independent of Y, given Z.
   b. Otherwise, X and Y are not necessarily conditionally independent, given Z.

Examples

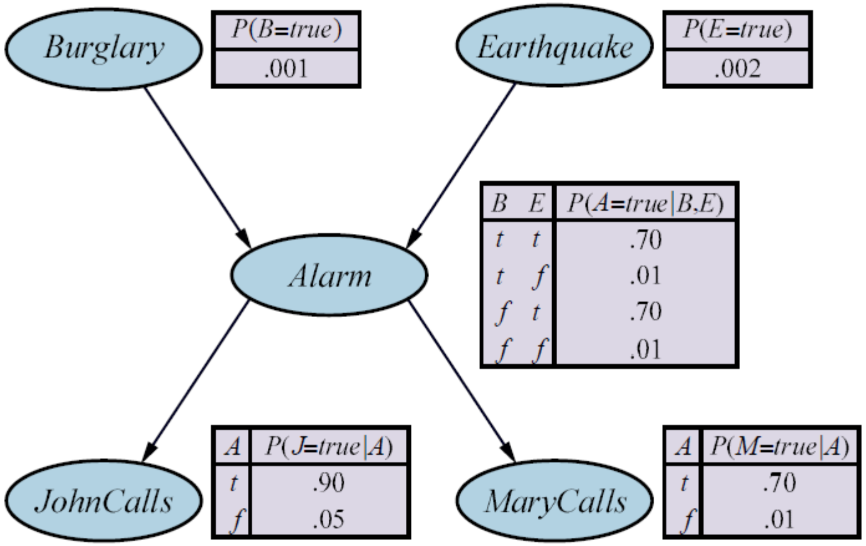


1. $X = \{$ *Burglary* $\}$

   $Y = \{$ *Earthquake* $\}$

   $Z = \{\ \}$

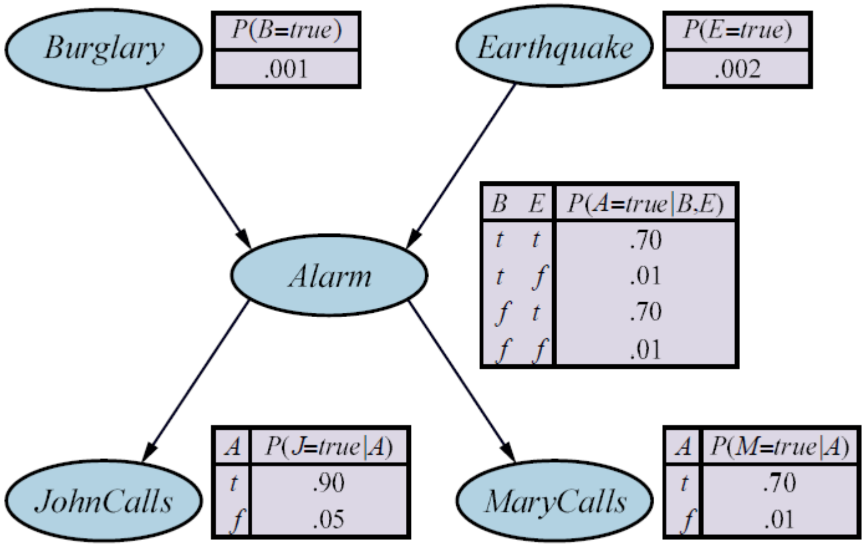
Examples



1. $X = \{$ *Burglary* $\}$
   $Y = \{$ *Earthquake* $\}$
   $Z = \{\,\}$

Moral graph:

Examples



1. $X = \{$ *Burglary* $\}$

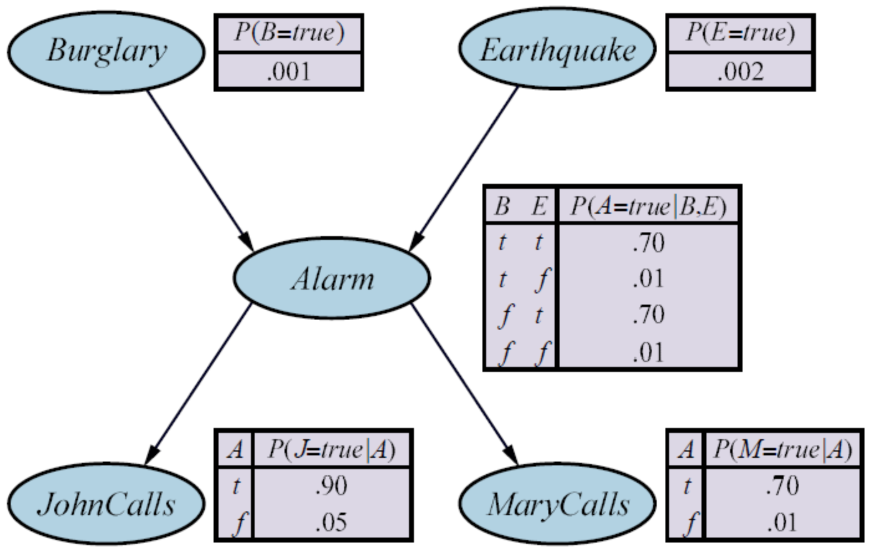
   $Y = \{$ *Earthquake* $\}$

   $Z = \{\ \}$

Moral graph:

X and Y are separated, thus d-separated by Z. They (*Burglary* and *Earthquake*) are independent given the empty set.

Examples



1. X = { *Burglary* }

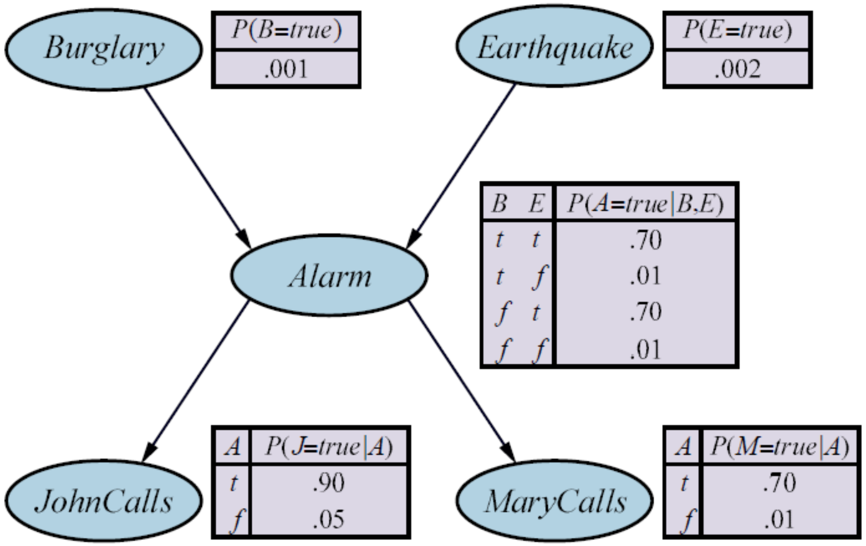
   Y = { *Earthquake* }

Moral graph:

X and Y are separated, thus d-separated by Z. They (*Burglary* and *Earthquake*) are independent given the empty set.

Z = { *Alarm* }

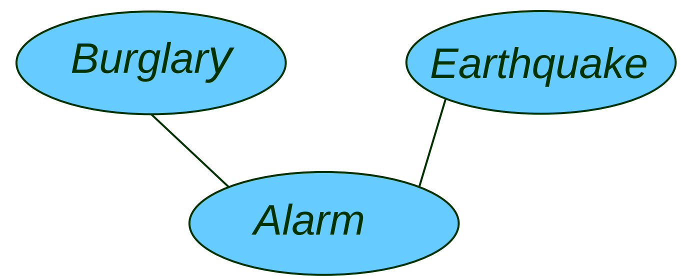
Examples



1. X = { *Burglary* }

   Y = { *Earthquake* }

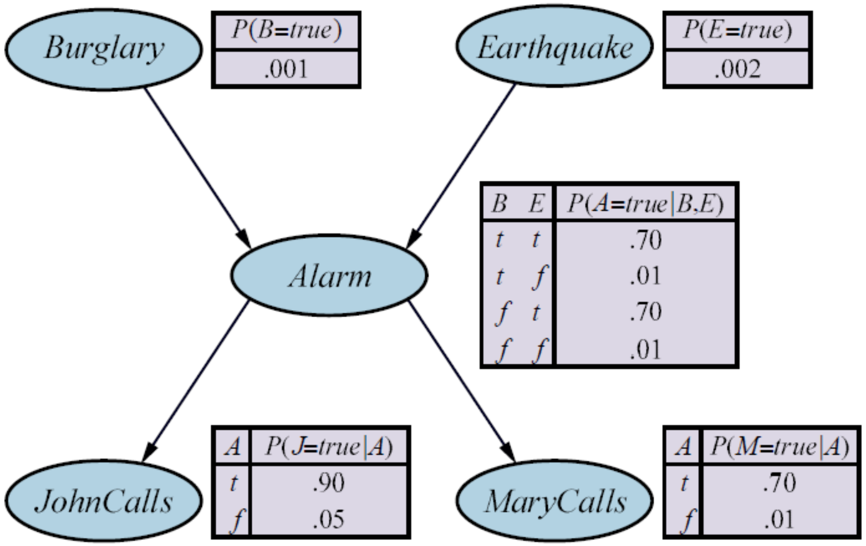
Moral graph:



X and Y are separated, thus d-separated by Z. They (*Burglary* and *Earthquake*) are independent given the empty set.

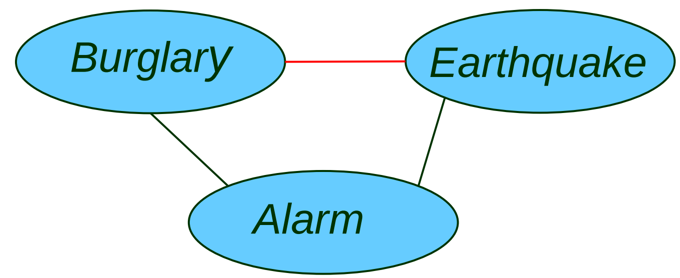
Z = { *Alarm* }

Examples



1. X = { *Burglary* }
   Y = { *Earthquake* }

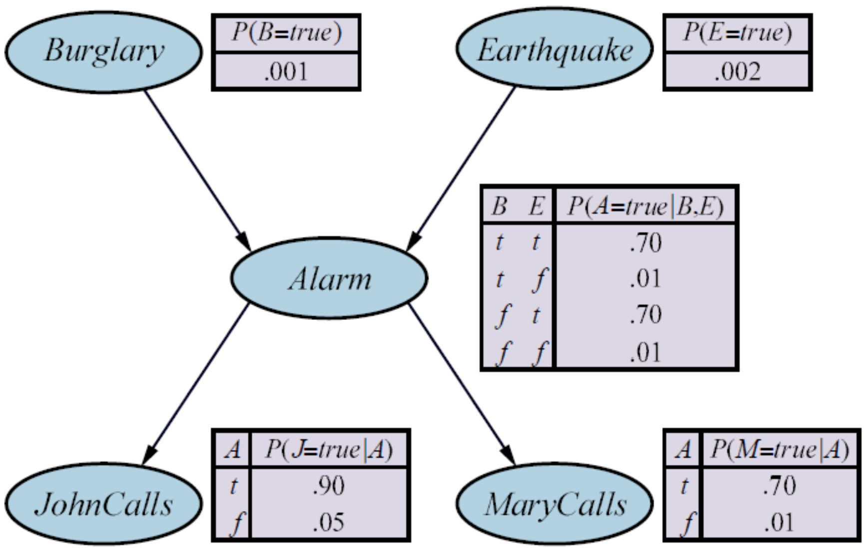
Moral graph:



X and Y are separated, thus d-separated by Z. They (*Burglary* and *Earthquake*) are independent given the empty set.

Z = { *Alarm* }

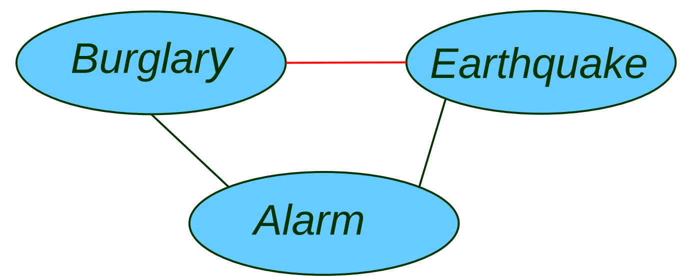
Examples



1. X = { *Burglary* }

   Y = { *Earthquake* }

Moral graph:

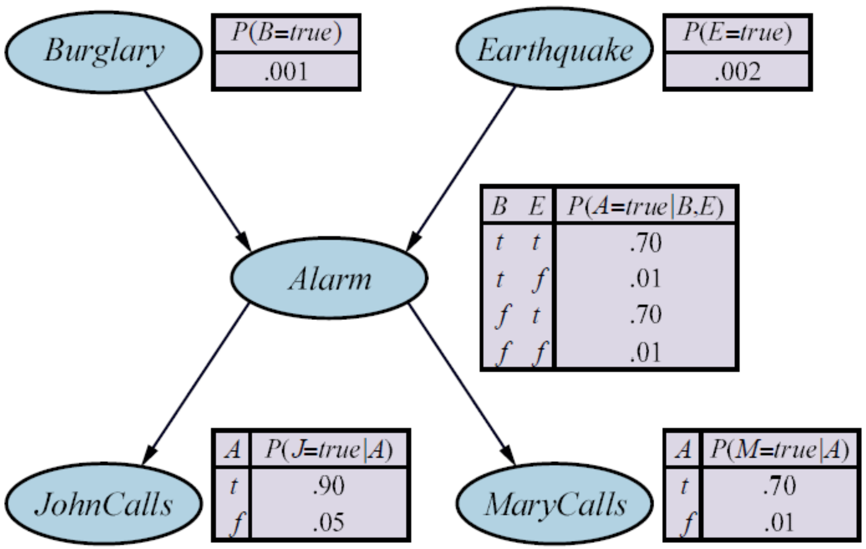


X and Y are separated, thus d-separated by Z. They (*Burglary* and *Earthquake*) are independent given the empty set.

Z = { *Alarm* }

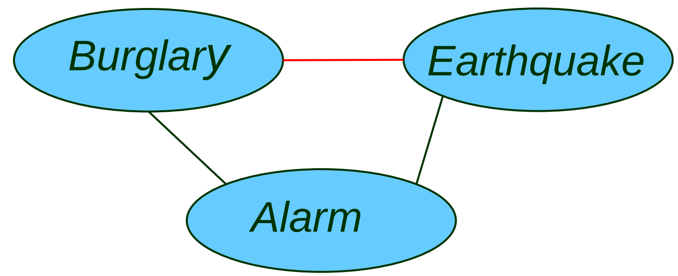
Burglary and *Earthquake* are not necessarily independent given *Alarm*.

Examples



1. X = { *Burglary* }
   Y = { *Earthquake* }

Moral graph:



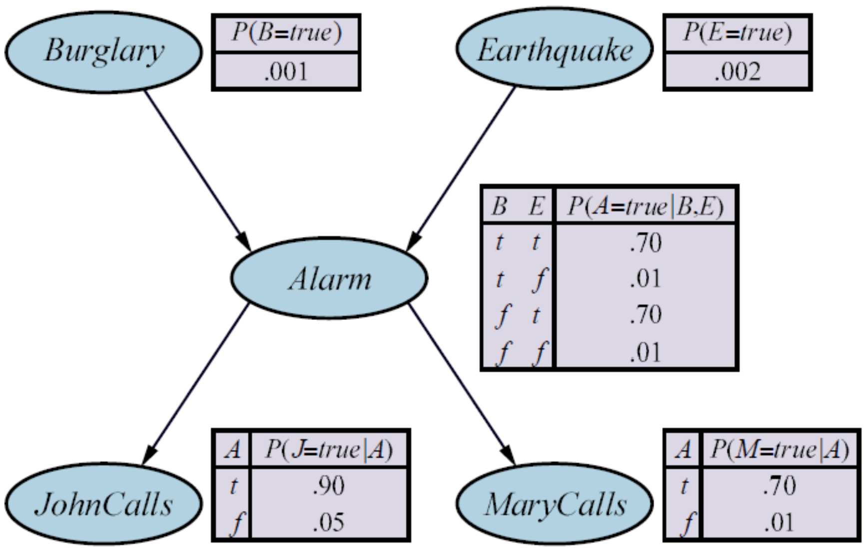
X and Y are separated, thus d-separated by Z. They (*Burglary* and *Earthquake*) are independent given the empty set.

Z = { *Alarm* }

Burglary and *Earthquake* are not necessarily independent given *Alarm*.

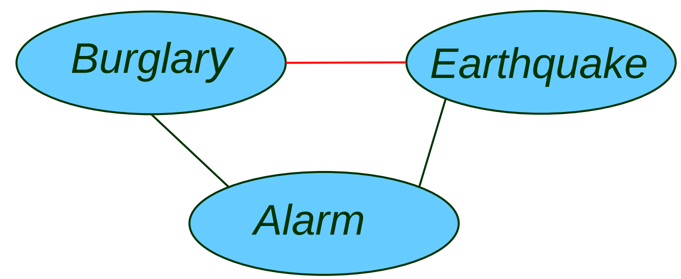
2. X = { *JohnCalls* }
   Y = { *MaryCalls* }
   Z = { *Alarm* }

Examples



1. X = { *Burglary* }
   Y = { *Earthquake* }

Moral graph:

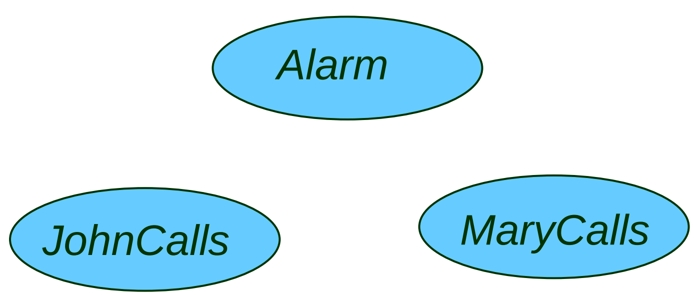


X and Y are separated, thus d-separated by Z. They (*Burglary* and *Earthquake*) are independent given the empty set.

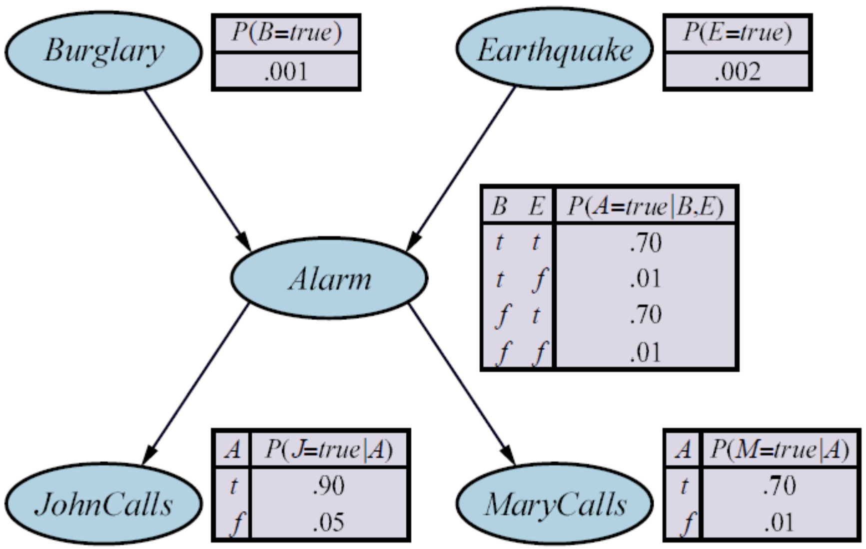
Z = { *Alarm* }

Burglary and *Earthquake* are not necessarily independent given *Alarm*.

2. X = { *JohnCalls* }
   Y = { *MaryCalls* }
   Z = { *Alarm* }



Examples



P(B=true)
.001

P(E=true)
.002

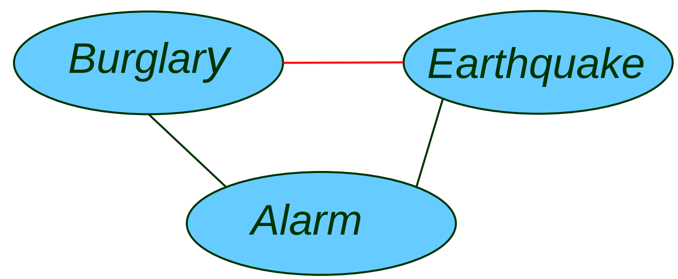
B	E	P(A=true|B,E)
t	t	.70
t	f	.01
f	t	.70
f	f	.01

A	P(J=true|A)
t	.90
f	.05

A	P(M=true|A)
t	.70
f	.01

1. X = { *Burglary* }
   Y = { *Earthquake* }

Moral graph:

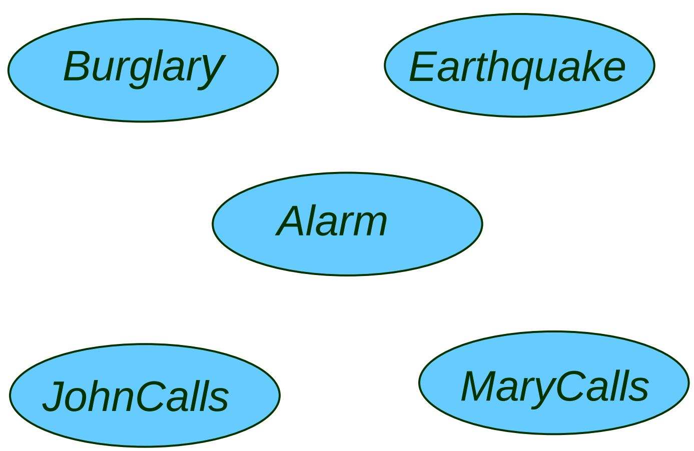


X and Y are separated, thus d-separated by Z. They (*Burglary* and *Earthquake*) are independent given the empty set.

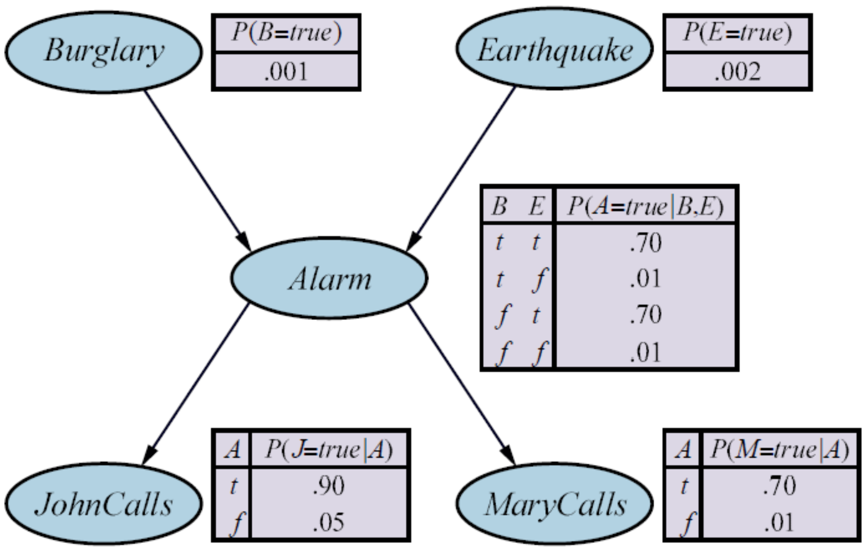
Z = { *Alarm* }

Burglary and *Earthquake* are not necessarily independent given *Alarm*.

2. X = { *JohnCalls* }
   Y = { *MaryCalls* }
   Z = { *Alarm* }

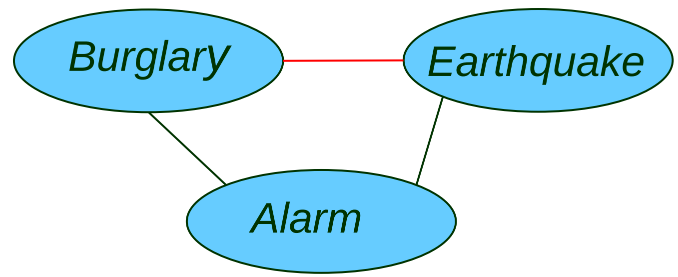


Examples



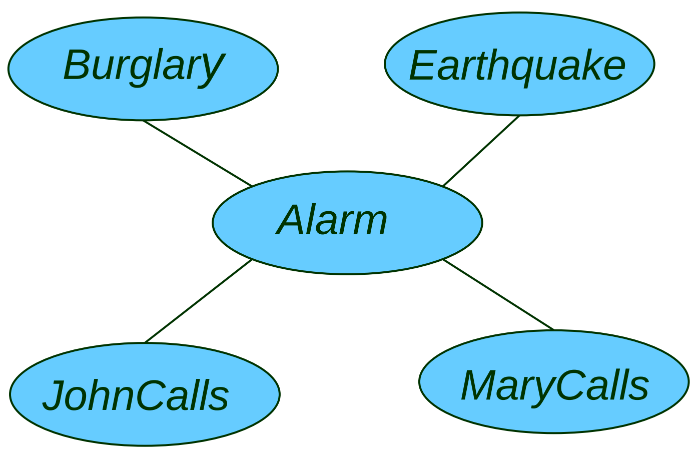
1. X = { *Burglary* }
   Y = { *Earthquake* }

Moral graph:



X and Y are separated, thus d-separated by Z. They (*Burglary* and *Earthquake*) are independent given the empty set.

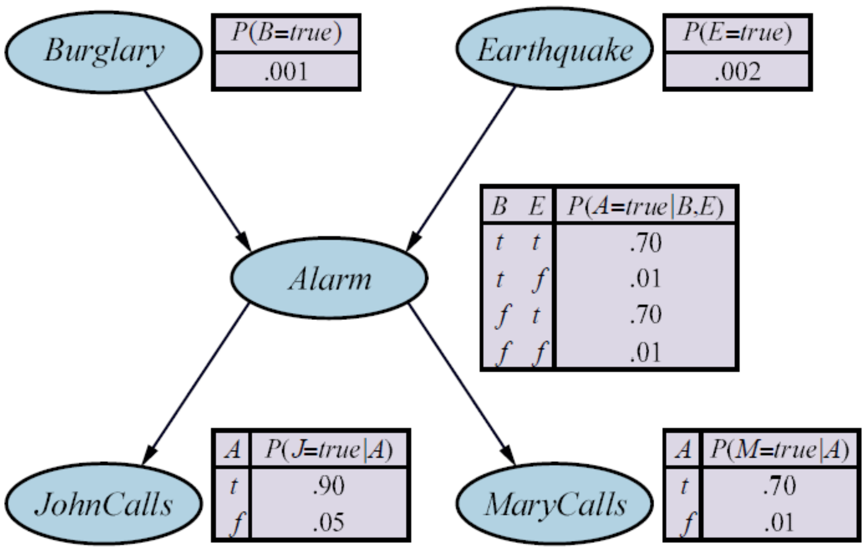
2. X = { *JohnCalls* }
   Y = { *MaryCalls* }
   Z = { *Alarm* }



Z = { *Alarm* }

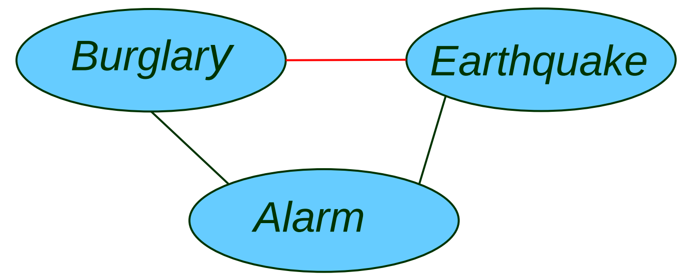
Burglary and *Earthquake* are not necessarily independent given *Alarm*.

Examples



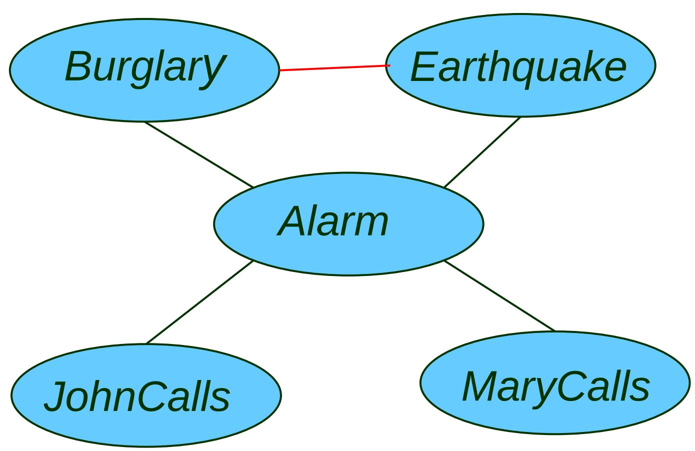
1. X = { *Burglary* }
   Y = { *Earthquake* }

Moral graph:



X and Y are separated, thus d-separated by Z. They (*Burglary* and *Earthquake*) are independent given the empty set.

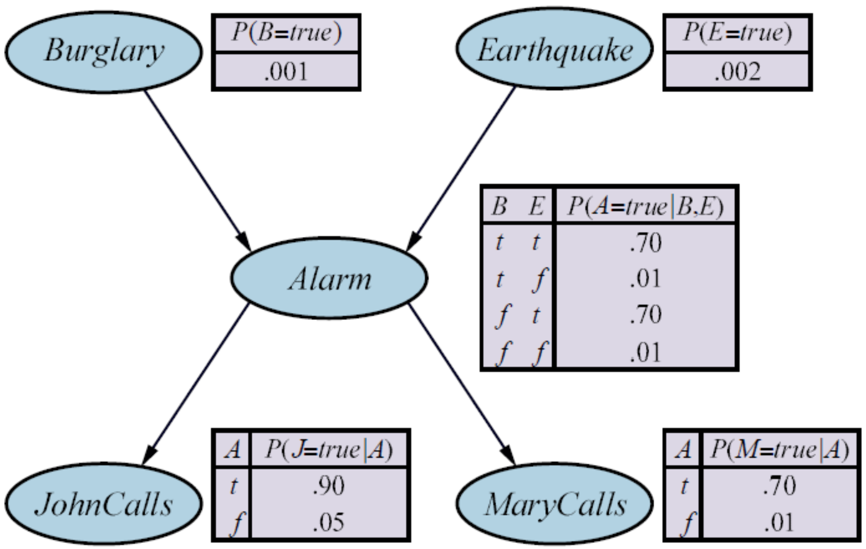
2. X = { *JohnCalls* }
   Y = { *MaryCalls* }
   Z = { *Alarm* }



Z = { *Alarm* }

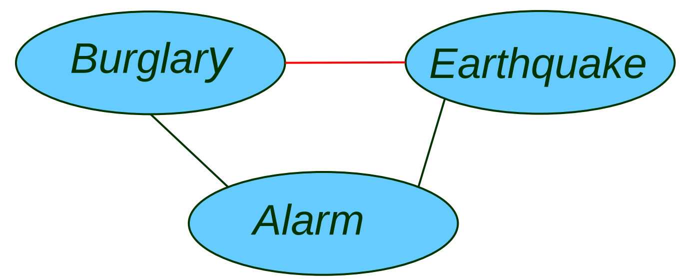
Burglary and *Earthquake* are not necessarily independent given *Alarm*.

Examples



1. X = { *Burglary* }
   Y = { *Earthquake* }

Moral graph:

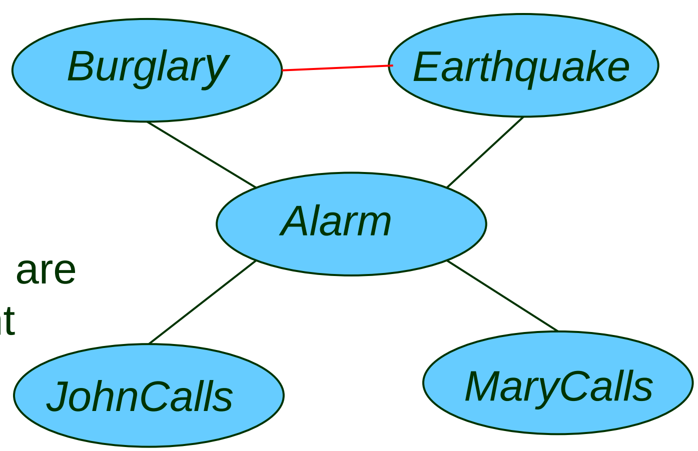


X and Y are separated, thus d-separated by Z. They (*Burglary* and *Earthquake*) are independent given the empty set.

Z = { *Alarm* }

Burglary and *Earthquake* are not necessarily independent given *Alarm*.

2. X = { *JohnCalls* }
   Y = { *MaryCalls* }
   Z = { *Alarm* }



JohnCalls and *MayCalls* are conditionally independent given *Alarm*.

III. Efficient Representations

- The size 2^k of CPT for a node with k parents is the worst-case scenario in which the relationships with the parents are arbitrary.

III. Efficient Representations

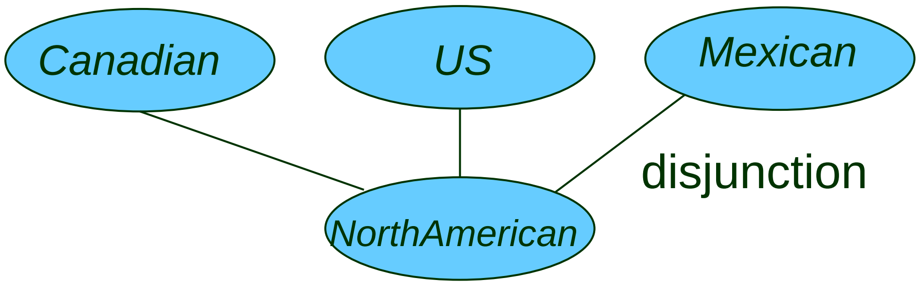
- ♣ The size 2^k of CPT for a node with k parents is the worst-case scenario in which the relationships with the parents are arbitrary.

- ♦ Deterministic nodes. Such a node has its value specified exactly by those of its parents, with no uncertainty. Specify deterministic functions in a BN.

III. Efficient Representations

- The size 2^k of CPT for a node with k parents is the worst-case scenario in which the relationships with the parents are arbitrary.

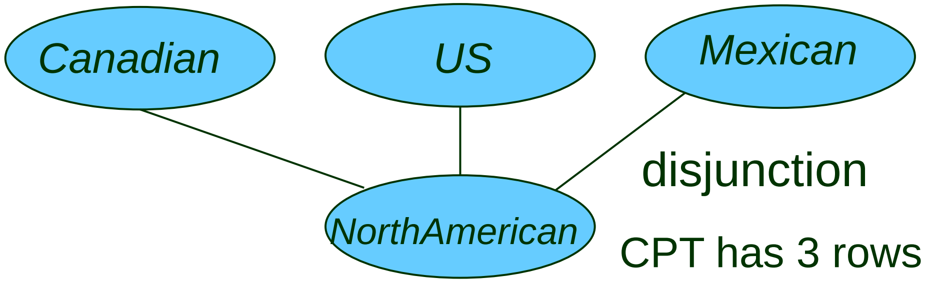
- Deterministic nodes. Such a node has its value specified exactly by those of its parents, with no uncertainty. Specify deterministic functions in a BN.



III. Efficient Representations

- The size 2^k of CPT for a node with k parents is the worst-case scenario in which the relationships with the parents are arbitrary.

- Deterministic nodes. Such a node has its value specified exactly by those of its parents, with no uncertainty. Specify deterministic functions in a BN.



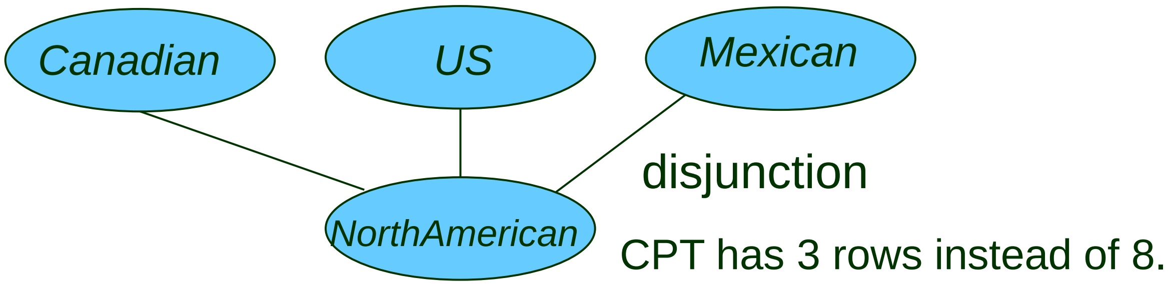
disjunction

CPT has 3 rows instead of 8.

III. Efficient Representations

- The size 2^k of CPT for a node with k parents is the worst-case scenario in which the relationships with the parents are arbitrary.

- Deterministic nodes. Such a node has its value specified exactly by those of its parents, with no uncertainty. Specify deterministic functions in a BN.

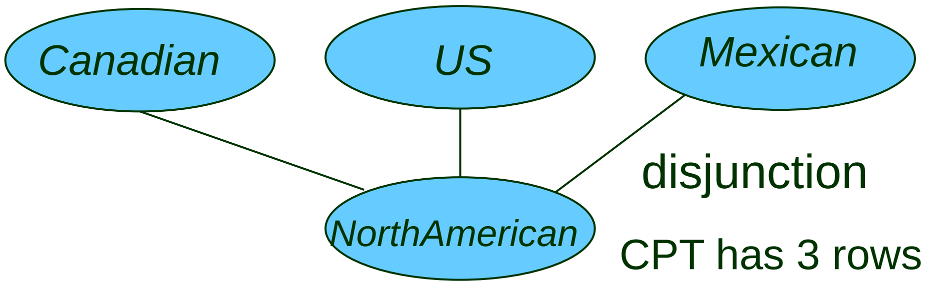


- Context-specific independence. A variable is conditionally independent of some of its parents given certain values of others.

III. Efficient Representations

- ♣ The size 2^k of CPT for a node with k parents is the worst-case scenario in which the relationships with the parents are arbitrary.

- ♦ Deterministic nodes. Such a node has its value specified exactly by those of its parents, with no uncertainty. Specify deterministic functions in a BN.



disjunction

CPT has 3 rows instead of 8.

- ♦ Context-specific independence. A variable is conditionally independent of some of its parents given certain values of others.

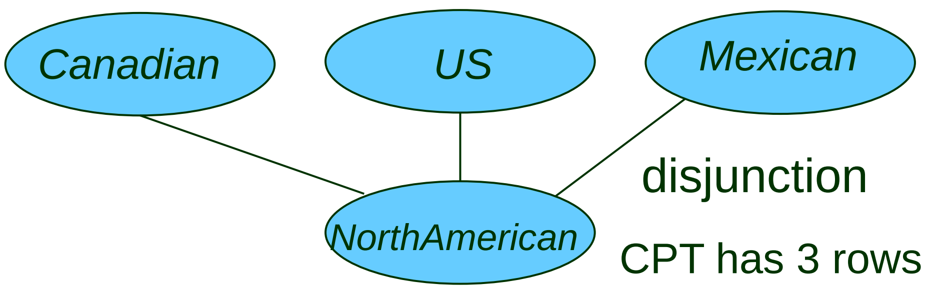
$$\boldsymbol{P}(\textit{Damage} \mid \textit{Ruggedness}, \textit{Accident}) = $$
$$\text{if } (\textit{Accident} = \text{false}) \text{ then } d_1 \text{ else } d_2(\textit{Ruggedness})$$

// if no accident, damage to
// your car does not depend
// on ruggedness.

III. Efficient Representations

- The size 2^k of CPT for a node with k parents is the worst-case scenario in which the relationships with the parents are arbitrary.

- Deterministic nodes. Such a node has its value specified exactly by those of its parents, with no uncertainty. Specify deterministic functions in a BN.



disjunction

CPT has 3 rows instead of 8.

- Context-specific independence. A variable is conditionally independent of some of its parents given certain values of others.

$$\mathbf{P}(Damage \mid Ruggedness, Accident) =$$
$$\text{if } (Accident = \text{false}) \text{ then } d_1 \text{ else } d_2(Ruggedness)$$

some distributions

// if no accident, damage to
// your car does not depend
// on ruggedness.

Noisy-OR Relation

The causal relationship between a node and its parents may be *inhibited*.

- All the possible causes are listed.
- Inhibition of each parent is independent of inhibition of any other parents.

Noisy-OR Relation

The causal relationship between a node and its parents may be *inhibited*.

- All the possible causes are listed.
- Inhibition of each parent is independent of inhibition of any other parents.

$Cold$	Flu	$Malaria$	$P(fever \mid \cdot)$	$P(\neg fever \mid \cdot)$
f	f	f	0.0	1.0
f	f	t	0.9	**0.1**
f	t	f	0.8	**0.2**
f	t	t	0.98	$0.02 = 0.2 \times 0.1$
t	f	f	0.4	**0.6**
t	f	t	0.94	$0.06 = 0.6 \times 0.1$
t	t	f	0.88	$0.12 = 0.6 \times 0.2$
t	t	t	0.988	$0.012 = 0.6 \times 0.2 \times 0.1$

Noisy-OR Relation

The causal relationship between a node and its parents may be *inhibited*.

- All the possible causes are listed.
- Inhibition of each parent is independent of inhibition of any other parents.

$Cold$	Flu	$Malaria$	$P(fever \mid \cdot)$	$P(\neg fever \mid \cdot)$
f	f	f	0.0	1.0
f	f	t	0.9	**0.1**
f	t	f	0.8	**0.2**
f	t	t	0.98	$0.02 = 0.2 \times 0.1$
t	f	f	0.4	**0.6**
t	f	t	0.94	$0.06 = 0.6 \times 0.1$
t	t	f	0.88	$0.12 = 0.6 \times 0.2$
t	t	t	0.988	$0.012 = 0.6 \times 0.2 \times 0.1$

- Whatever inhibits cold from causing a fever is independent of whatever inhibits flu from causing a fever.
- *Fever* is false if and only if all its parents are inhibited.

Noisy-OR Relation

The causal relationship between a node and its parents may be *inhibited*.

- All the possible causes are listed.
- Inhibition of each parent is independent of inhibition of any other parents.

$Cold$	Flu	$Malaria$	$P(fever \mid \cdot)$	$P(\neg fever \mid \cdot)$
f	f	f	0.0	1.0
f	f	t	0.9	**0.1**
f	t	f	0.8	**0.2**
f	t	t	0.98	$0.02 = 0.2 \times 0.1$
t	f	f	0.4	**0.6**
t	f	t	0.94	$0.06 = 0.6 \times 0.1$
t	t	f	0.88	$0.12 = 0.6 \times 0.2$
t	t	t	0.988	$0.012 = 0.6 \times 0.2 \times 0.1$

- Random variables:

  $X_1 \equiv$ *Cold*, $X_2 \equiv$ *Flu*, $X_3 \equiv$ *Malaria*

- Whatever inhibits cold from causing a fever is independent of whatever inhibits flu from causing a fever.
- *Fever* is false if and only if all its parents are inhibited.

Noisy-OR Relation

The causal relationship between a node and its parents may be *inhibited*.

- All the possible causes are listed.
- Inhibition of each parent is independent of inhibition of any other parents.

$Cold$	Flu	$Malaria$	$P(fever \mid \cdot)$	$P(\neg fever \mid \cdot)$
f	f	f	0.0	1.0
f	f	t	0.9	**0.1**
f	t	f	0.8	**0.2**
f	t	t	0.98	$0.02 = 0.2 \times 0.1$
t	f	f	0.4	**0.6**
t	f	t	0.94	$0.06 = 0.6 \times 0.1$
t	t	f	0.88	$0.12 = 0.6 \times 0.2$
t	t	t	0.988	$0.012 = 0.6 \times 0.2 \times 0.1$

- Random variables:

  $X_1 \equiv$ *Cold*, $X_2 \equiv$ *Flu*, $X_3 \equiv$ *Malaria*

- Inhibition probabilities:

- Whatever inhibits cold from causing a fever is independent of whatever inhibits flu from causing a fever.
- *Fever* is false if and only if all its parents are inhibited.

Noisy-OR Relation

The causal relationship between a node and its parents may be *inhibited*.

- All the possible causes are listed.
- Inhibition of each parent is independent of inhibition of any other parents.

$Cold$	Flu	$Malaria$	$P(fever \mid \cdot)$	$P(\neg fever \mid \cdot)$
f	f	f	0.0	1.0
f	f	t	0.9	**0.1**
f	t	f	0.8	**0.2**
f	t	t	0.98	$0.02 = 0.2 \times 0.1$
t	f	f	0.4	**0.6**
t	f	t	0.94	$0.06 = 0.6 \times 0.1$
t	t	f	0.88	$0.12 = 0.6 \times 0.2$
t	t	t	0.988	$0.012 = 0.6 \times 0.2 \times 0.1$

- Random variables:

  $X_1 \equiv \mathit{Cold}$, $X_2 \equiv \mathit{Flu}$, $X_3 \equiv \mathit{Malaria}$

- Inhibition probabilities:

- Whatever inhibits cold from causing a fever is independent of whatever inhibits flu from causing a fever.
- *Fever* is false if and only if all its parents are inhibited.

Noisy-OR Relation

The causal relationship between a node and its parents may be *inhibited*.

- All the possible causes are listed.
- Inhibition of each parent is independent of inhibition of any other parents.

$Cold$	Flu	$Malaria$	$P(fever \mid \cdot)$	$P(\neg fever \mid \cdot)$
f	f	f	0.0	1.0
f	f	t	0.9	**0.1**
f	t	f	0.8	**0.2**
f	t	t	0.98	$0.02 = 0.2 \times 0.1$
t	f	f	0.4	**0.6**
t	f	t	0.94	$0.06 = 0.6 \times 0.1$
t	t	f	0.88	$0.12 = 0.6 \times 0.2$
t	t	t	0.988	$0.012 = 0.6 \times 0.2 \times 0.1$

- Random variables:

  $X_1 \equiv Cold,\ X_2 \equiv Flu,\ X_3 \equiv Malaria$

- Inhibition probabilities:

$q_1 = q_{\text{cold}}$ // false alarm

$= P(\neg fever \mid cold, \neg flu, \neg malaria) = 0.6$

- Whatever inhibits cold from causing a fever is independent of whatever inhibits flu from causing a fever.
- *Fever* is false if and only if all its parents are inhibited.

Noisy-OR Relation

The causal relationship between a node and its parents may be *inhibited*.

- All the possible causes are listed.
- Inhibition of each parent is independent of inhibition of any other parents.

Cold	*Flu*	*Malaria*	$P(fever \mid \cdot)$	$P(\neg fever \mid \cdot)$
f	f	f	0.0	1.0
f	f	t	0.9	**0.1**
f	t	f	0.8	**0.2**
f	t	t	0.98	$0.02 = 0.2 \times 0.1$
t	f	f	0.4	**0.6**
t	f	t	0.94	$0.06 = 0.6 \times 0.1$
t	t	f	0.88	$0.12 = 0.6 \times 0.2$
t	t	t	0.988	$0.012 = 0.6 \times 0.2 \times 0.1$

- Random variables:

  $X_1 \equiv Cold,\ X_2 \equiv Flu,\ X_3 \equiv Malaria$

- Inhibition probabilities:

$q_1 = q_{\text{cold}}$ // false alarm

$= P(\neg fever \mid cold, \neg flu, \neg malaria) = 0.6$

- Whatever inhibits cold from causing a fever is independent of whatever inhibits flu from causing a fever.
- *Fever* is false if and only if all its parents are inhibited.

Noisy-OR Relation

The causal relationship between a node and its parents may be *inhibited*.

- All the possible causes are listed.
- Inhibition of each parent is independent of inhibition of any other parents.

$Cold$	Flu	$Malaria$	$P(fever \mid \cdot)$	$P(\neg fever \mid \cdot)$
f	f	f	0.0	1.0
f	f	t	0.9	**0.1**
f	t	f	0.8	**0.2**
f	t	t	0.98	$0.02 = 0.2 \times 0.1$
t	f	f	0.4	**0.6**
t	f	t	0.94	$0.06 = 0.6 \times 0.1$
t	t	f	0.88	$0.12 = 0.6 \times 0.2$
t	t	t	0.988	$0.012 = 0.6 \times 0.2 \times 0.1$

- Random variables:

  $X_1 \equiv Cold,\ X_2 \equiv Flu,\ X_3 \equiv Malaria$

- Inhibition probabilities:

$q_1 = q_{\text{cold}}$ // false alarm

$= P(\neg fever \mid cold, \neg flu, \neg malaria) = 0.6$

$q_2 = q_{\text{flu}}$

$= P(\neg fever \mid \neg cold, flu, \neg malaria) = 0.2$

- Whatever inhibits cold from causing a fever is independent of whatever inhibits flu from causing a fever.
- *Fever* is false if and only if all its parents are inhibited.

Noisy-OR Relation

The causal relationship between a node and its parents may be *inhibited*.

- All the possible causes are listed.
- Inhibition of each parent is independent of inhibition of any other parents.

$Cold$	Flu	$Malaria$	$P(fever \mid \cdot)$	$P(\neg fever \mid \cdot)$
f	f	f	0.0	1.0
f	f	t	0.9	**0.1**
f	t	f	0.8	**0.2**
f	t	t	0.98	$0.02 = 0.2 \times 0.1$
t	f	f	0.4	**0.6**
t	f	t	0.94	$0.06 = 0.6 \times 0.1$
t	t	f	0.88	$0.12 = 0.6 \times 0.2$
t	t	t	0.988	$0.012 = 0.6 \times 0.2 \times 0.1$

- Random variables:

  $X_1 \equiv Cold,\ X_2 \equiv Flu,\ X_3 \equiv Malaria$

- Inhibition probabilities:

$q_1 = q_{\text{cold}}$ // false alarm

$= P(\neg fever \mid cold, \neg flu, \neg malaria) = 0.6$

$q_2 = q_{\text{flu}}$

$= P(\neg fever \mid \neg cold, flu, \neg malaria) = 0.2$

- Whatever inhibits cold from causing a fever is independent of whatever inhibits flu from causing a fever.
- *Fever* is false if and only if all its parents are inhibited.

Noisy-OR Relation

The causal relationship between a node and its parents may be *inhibited*.

- All the possible causes are listed.
- Inhibition of each parent is independent of inhibition of any other parents.

$Cold$	Flu	$Malaria$	$P(fever \mid \cdot)$	$P(\neg fever \mid \cdot)$
f	f	f	0.0	1.0
f	f	t	0.9	**0.1**
f	t	f	0.8	**0.2**
f	t	t	0.98	$0.02 = 0.2 \times 0.1$
t	f	f	0.4	**0.6**
t	f	t	0.94	$0.06 = 0.6 \times 0.1$
t	t	f	0.88	$0.12 = 0.6 \times 0.2$
t	t	t	0.988	$0.012 = 0.6 \times 0.2 \times 0.1$

- Random variables:

  $X_1 \equiv \mathit{Cold},\ X_2 \equiv \mathit{Flu},\ X_3 \equiv \mathit{Malaria}$

- Inhibition probabilities:

$q_1 = q_{\text{cold}}$ // false alarm

$= P(\neg \mathit{fever} \mid \mathit{cold}, \neg \mathit{flu}, \neg \mathit{malaria}) = 0.6$

$q_2 = q_{\text{flu}}$

$= P(\neg \mathit{fever} \mid \neg \mathit{cold}, \mathit{flu}, \neg \mathit{malaria}) = 0.2$

$q_3 = q_{\text{malaria}}$

$= P(\neg \mathit{fever} \mid \neg \mathit{cold}, \neg \mathit{flu}, \mathit{malaria}) = 0.1$

- Whatever inhibits cold from causing a fever is independent of whatever inhibits flu from causing a fever.
- *Fever* is false if and only if all its parents are inhibited.

Noisy-OR Relation

The causal relationship between a node and its parents may be *inhibited*.

- All the possible causes are listed.
- Inhibition of each parent is independent of inhibition of any other parents.

$Cold$	Flu	$Malaria$	$P(fever \mid \cdot)$	$P(\neg fever \mid \cdot)$
f	f	f	0.0	1.0
f	f	t	0.9	**0.1**
f	t	f	0.8	**0.2**
f	t	t	0.98	$0.02 = 0.2 \times 0.1$
t	f	f	0.4	**0.6**
t	f	t	0.94	$0.06 = 0.6 \times 0.1$
t	t	f	0.88	$0.12 = 0.6 \times 0.2$
t	t	t	0.988	$0.012 = 0.6 \times 0.2 \times 0.1$

- Random variables:

  $X_1 \equiv Cold,\ X_2 \equiv Flu,\ X_3 \equiv Malaria$

- Inhibition probabilities:

$q_1 = q_{\text{cold}}$ // false alarm

$= P(\neg fever \mid cold, \neg flu, \neg malaria) = 0.6$

$q_2 = q_{\text{flu}}$

$= P(\neg fever \mid \neg cold, flu, \neg malaria) = 0.2$

$q_3 = q_{\text{malaria}}$

$= P(\neg fever \mid \neg cold, \neg flu, malaria) = 0.1$

- Whatever inhibits cold from causing a fever is independent of whatever inhibits flu from causing a fever.
- *Fever* is false if and only if all its parents are inhibited.

Noisy-OR Relation

The causal relationship between a node and its parents may be *inhibited*.

- All the possible causes are listed.
- Inhibition of each parent is independent of inhibition of any other parents.

$Cold$	Flu	$Malaria$	$P(fever \mid \cdot)$	$P(\neg fever \mid \cdot)$
f	f	f	0.0	1.0
f	f	t	0.9	$\mathbf{0.1}$
f	t	f	0.8	$\mathbf{0.2}$
f	t	t	0.98	$0.02 = 0.2 \times 0.1$
t	f	f	0.4	$\mathbf{0.6}$
t	f	t	0.94	$0.06 = 0.6 \times 0.1$
t	t	f	0.88	$0.12 = 0.6 \times 0.2$
t	t	t	0.988	$0.012 = 0.6 \times 0.2 \times 0.1$

- Random variables:

  $X_1 \equiv Cold,\ X_2 \equiv Flu,\ X_3 \equiv Malaria$

- Inhibition probabilities:

$q_1 = q_{\text{cold}}$ // false alarm

$= P(\neg fever \mid cold, \neg flu, \neg malaria) = 0.6$

$q_2 = q_{\text{flu}}$

$= P(\neg fever \mid \neg cold, flu, \neg malaria) = 0.2$

$q_3 = q_{\text{malaria}}$

$= P(\neg fever \mid \neg cold, \neg flu, malaria) = 0.1$

- Whatever inhibits cold from causing a fever is independent of whatever inhibits flu from causing a fever.
- *Fever* is false if and only if all its parents are inhibited.

Noisy-OR Relation

The causal relationship between a node and its parents may be *inhibited*.

- All the possible causes are listed.
- Inhibition of each parent is independent of inhibition of any other parents.

Cold	*Flu*	*Malaria*	$P(fever \mid \cdot)$	$P(\neg fever \mid \cdot)$
f	f	f	0.0	1.0
f	f	t	0.9	**0.1**
f	t	f	0.8	**0.2**
f	t	t	0.98	$0.02 = 0.2 \times 0.1$
t	f	f	0.4	**0.6**
t	f	t	0.94	$0.06 = 0.6 \times 0.1$
t	t	f	0.88	$0.12 = 0.6 \times 0.2$
t	t	t	0.988	$0.012 = 0.6 \times 0.2 \times 0.1$

- Random variables:

  $X_1 \equiv \mathit{Cold},\ X_2 \equiv \mathit{Flu},\ X_3 \equiv \mathit{Malaria}$

- Inhibition probabilities:

$q_1 = q_{\text{cold}}$ // false alarm

$= P(\neg \mathit{fever} \mid \mathit{cold}, \neg \mathit{flu}, \neg \mathit{malaria}) = 0.6$

$q_2 = q_{\text{flu}}$

$= P(\neg \mathit{fever} \mid \neg \mathit{cold}, \mathit{flu}, \neg \mathit{malaria}) = 0.2$

$q_3 = q_{\text{malaria}}$

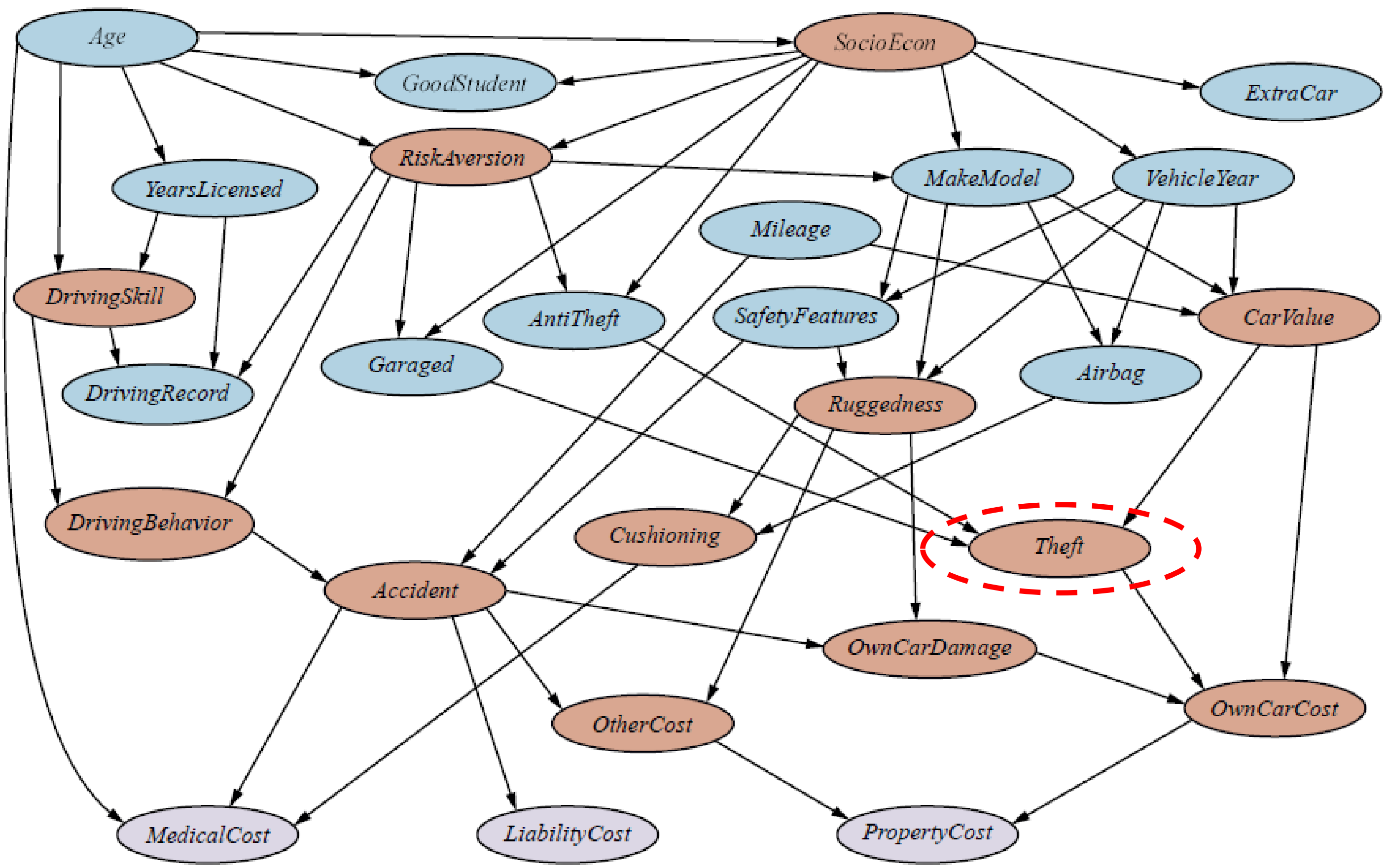
$= P(\neg \mathit{fever} \mid \neg \mathit{cold}, \neg \mathit{flu}, \mathit{malaria}) = 0.1$

- Whatever inhibits cold from causing a fever is independent of whatever inhibits flu from causing a fever.
- *Fever* is false if and only if all its parents are inhibited.

$$P(x_i \mid \mathit{parents}(X_i)) = 1 - \prod_{\{j : X_j = \mathit{true}\}} q_j$$

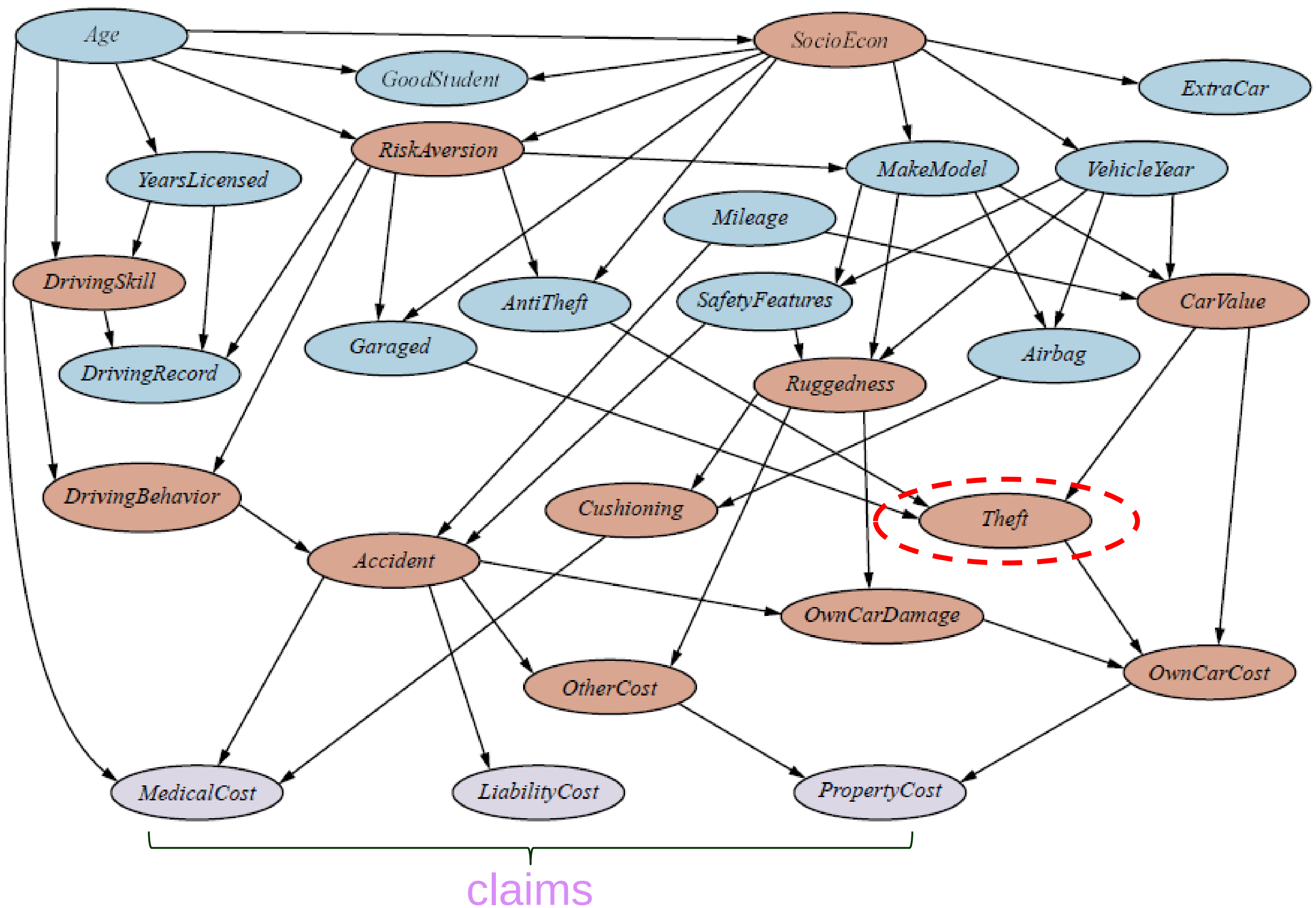
Case Study: Car Insurance

A car insurance company processes a car insurance application to decide on the annual premium based on the anticipated claims it will pay out for the applicant.



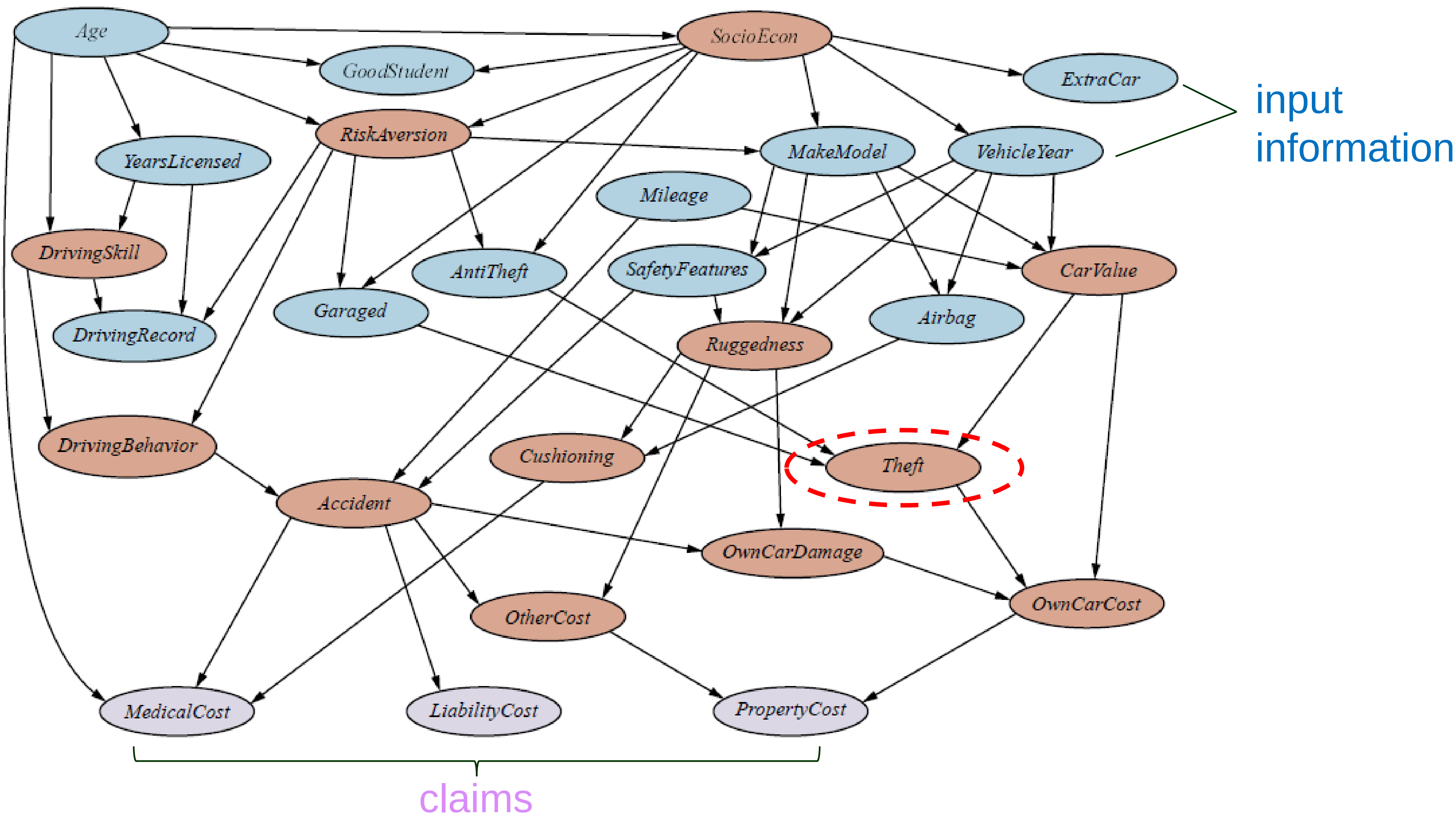
Case Study: Car Insurance

A car insurance company processes a car insurance application to decide on the annual premium based on the anticipated claims it will pay out for the applicant.



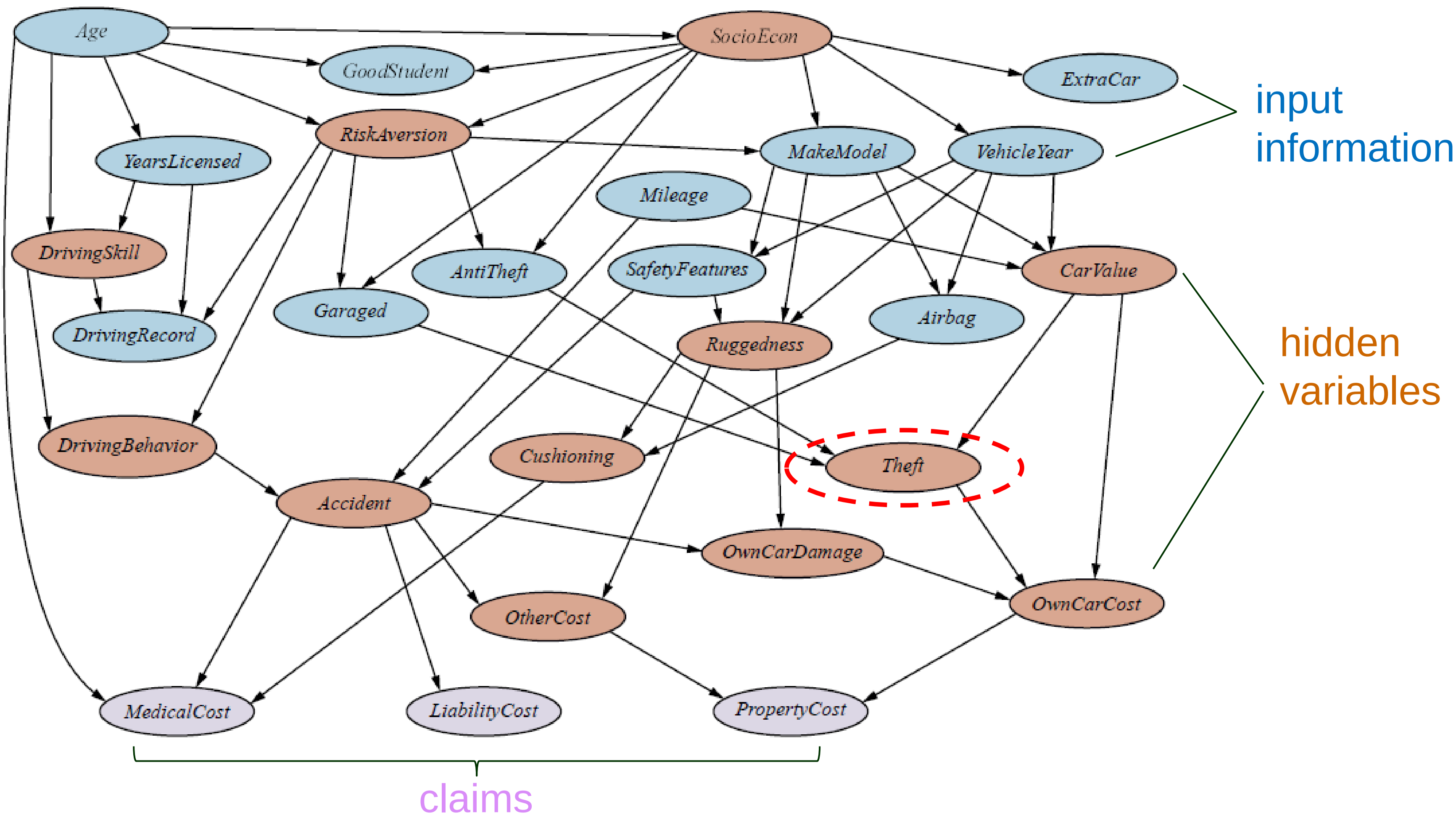
Case Study: Car Insurance

A car insurance company processes a car insurance application to decide on the annual premium based on the anticipated claims it will pay out for the applicant.



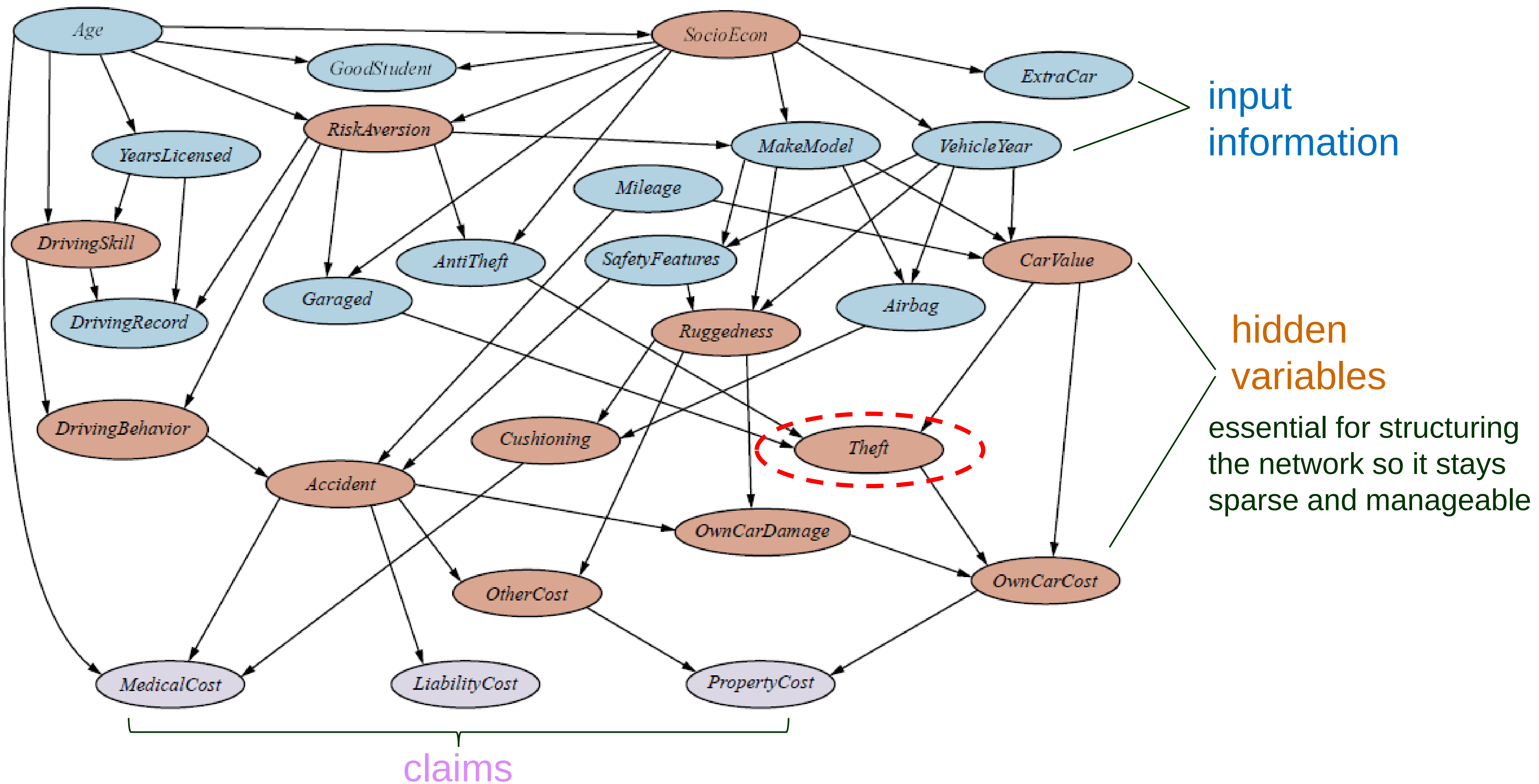
Case Study: Car Insurance

A car insurance company processes a car insurance application to decide on the annual premium based on the anticipated claims it will pay out for the applicant.



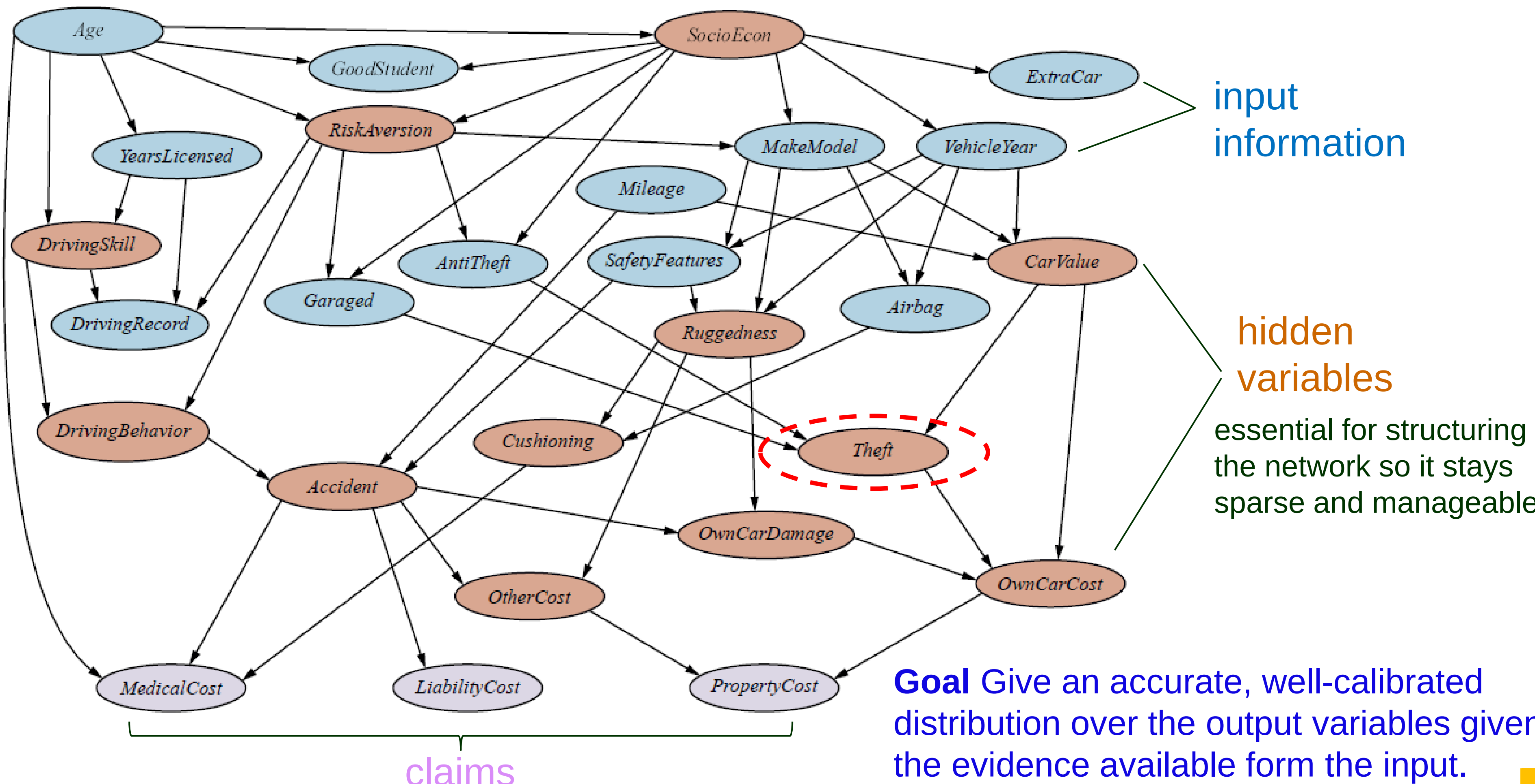
Case Study: Car Insurance

A car insurance company processes a car insurance application to decide on the annual premium based on the anticipated claims it will pay out for the applicant.



Case Study: Car Insurance

A car insurance company processes a car insurance application to decide on the annual premium based on the anticipated claims it will pay out for the applicant.



Goal Give an accurate, well-calibrated distribution over the output variables given the evidence available form the input.