Conditional Probability Table

- A **node without parents** has an unconditional probability

P(Burglary)
.001

- The **sum of probabilities** must be 1
  - If there is a single input variable, it is possible to remove the redundancy

Alarm (A)	P(JohnCalls| .)	P(-JohnCalls | .)
True	0.90	0.10
False	0.05	0.95

Alarm (A)	P(JohnCalls| .)
True	0.90
False	0.05

- **A node with k parents** has 2^k possible rows in the table

Burglary	Earthquake	P(Alarm | .)
T	T	.95
T	F	.94
...	...	...

SCIENCE ACADEMY

- ***Semantics of Bayesian Networks***
- Constructing a Bayesian Network

Bayesian Networks: Semantics

- There are **two equivalent semantic interpretations** of a Bayesian Network

1. **Joint Distribution View**
   - The network encodes the *joint probability distribution* over all variables
   - Computed as the product of local conditional probabilities:

$$P(X_1, \dots, X_n) = \prod_{i=1}^{n} P(X_i \mid \text{Parents}(X_i))$$

   - Useful for constructing models and understanding overall behavior
2. **Conditional Independence View**
   - The structure encodes *conditional independency* between variables
   - Useful for inference and reasoning
   - A variable is conditionally independent of its non-descendants given its parents

Chain Rule for a Joint Distribution

- A **joint distribution** can always be expressed using the chain rule for any:
  - Subset of its RVs
  - Ordering of the RVs

1. You **express one variable** conditionally to the remaining ones

$$\Pr(x_1, \ldots, x_{n-1}, x_n) = \Pr(x_n | x_{n-1}, \ldots, x_1) \Pr(x_{n-1}, \ldots, x_1)$$

2. Apply the same formula **recursively**, until you get an unconditional probability

$$\begin{aligned}
&\Pr(x_1, x_2, \ldots, x_{n-2}, x_{n-1}, x_n) \\
&= \Pr(x_n | x_{n-1}, \ldots, x_1) \Pr(x_{n-1}, \ldots, x_1) \\
&= \Pr(x_n | x_{n-1}, \ldots, x_1) \Pr(x_{n-1} | x_{n-2}, \ldots, x_1) \Pr(x_{n-2}, \ldots, x_1) \\
&\ldots \\
&= \Pr(x_n | x_{n-1}, \ldots, x_1) \Pr(x_{n-1} | x_{n-2}, \ldots, x_1) \Pr(x_{n-2} | x_{n-3}, \ldots, x_1) \ldots \Pr(x_2 | x_1) \Pr(x_1) \\
&= \prod_{i=1}^{n} \Pr(x_i | x_{i-1}, \ldots, x_1)
\end{aligned}$$

Statement Probability from Bayesian Network

- The **full joint distribution** represents the probability of an assignment to each variable $X_i = x_i$:

$$\Pr(x_1, ..., x_n) \triangleq \Pr(X_1 = x_1 \wedge ... \wedge X_n = x_n)$$

- To **evaluate a Bayesian network**
  - Sort the nodes in topological order
    - There are several orderings consistent with the directed graph structure
  - Use the chain rule with the topological ordering:

$$\Pr(X_1, ..., X_n) = \prod_{i=1}^{n} \Pr(X_i | X_{i-1}, ..., X_1)$$

  - Since the probability of each node is conditionally independent of all its predecessors given its parents

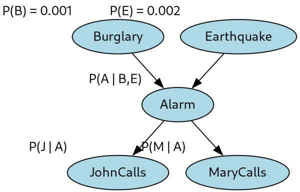
$$\Pr(X_i | X_{i-1}, ..., X_1) = \Pr(X_i | Parents(X_i))$$

  - Express the joint probability in terms of the Conditional Probability Tables (CPTs):

$$\Pr(X_1, ..., X_n) = \prod_{i=1}^{n} \Pr(X_i | Parents(X_i))$$

Statement Probability From Bayes Nets: Example

- Given Pearl LA example, you want to compute the probability that:
  - The alarm has sounded: $Alarm$
  - Neither a burglary nor an earthquake has occurred: $\neg Burglary \wedge \neg Earthquake$
  - Both John and Mary call: $JohnCalls, MaryCalls$



- **Solution**
  - Compute the probability as a product of conditional probabilities from the Bayesian Network

$$
\begin{aligned}
&\Pr(JohnCalls, MaryCalls, Alarm, \neg Burglary, \neg Earthquake) \\
&= \Pr(JohnCalls|Alarm)\cdot \\
&\qquad \Pr(MaryCalls|Alarm)\cdot \\
&\qquad \Pr(Alarm|\neg Burglary \wedge \neg Earthquake)\cdot \\
&\qquad \Pr(\neg Burglary) \cdot \Pr(\neg Earthquake)
\end{aligned}
$$

SCIENCE ACADEMY

- Semantics of Bayesian Networks
- ***Constructing a Bayesian Network***

Constructing a Bayesian Network

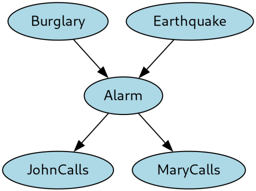
1. **Gather domain knowledge**
   - Identify key variables and their potential interactions
   - List all relevant random variables necessary to describe the system
2. **Order the nodes** according to cause-effects dependencies
   - So that the Bayesian network is minimal
3. For each node, **pick the minimum set of parents** $Parents(X_i)$
   - Add edges to represent the dependencies
   - Avoid redundant connections
4. **Estimate the conditional probability** $\Pr(X_i | Parents(X_i))$ for each node
   - Gather data or expert opinion
   - Use statistical techniques if necessary
5. **Validate the model**
   - Have domain experts review it
   - Ensure that the network is a Directed Acyclic Graph (DAG)
   - Test the network by predicting known outcomes and comparing with actual data

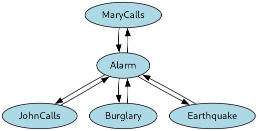
Bayesian Networks: Properties

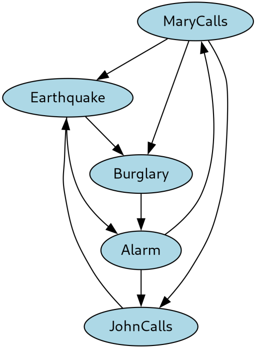
- Bayesian networks are a representation with several interesting properties
  - **Complete**
    - Encode all information in a joint probability
  - **Consistent** (non-redundant)
    - In a Bayesian network, there are no redundant probability values
    - One (e.g., a domain expert) can't create a Bayesian network violating probability axioms
  - **Compact** (locally structured, sparse)
    - Each subcomponent interacts directly with a limited number of other components
    - Typically yields linear (not exponential) growth in complexity
    - Sometimes we ignore real-world dependency to keep the graph simple
- In **fully connected systems**
  - Each variable is influenced by all others
  - The Bayesian network has the same complexity as the joint probability

Ordering of Nodes

- The complexity of the Bayesian network depends on the choice in ordering the nodes



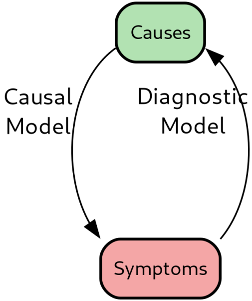




- The graph is "minimal" in terms of connectivity when all edges are causal

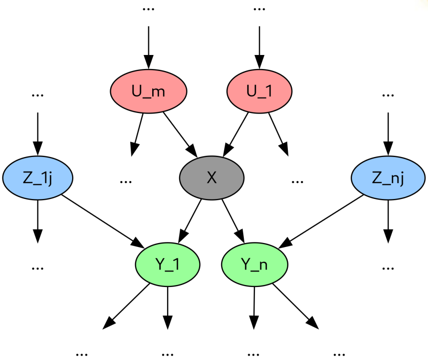
Causal vs Diagnostic Models

- A **causal model** goes from causes to symptoms
  - E.g., *Burglary* $\rightarrow$ *Alarm*
  - Simpler (i.e., fewer and more robust dependencies)
  - "Easier" to estimate
- A **diagnostic model** goes from symptoms to causes
  - E.g., *MaryCalls* $\rightarrow$ *Alarm* or *Alarm* $\rightarrow$ *Burglary*
  - Tenuous / unstable
  - Difficult to estimate
  - This is what we care about: use Bayes' rule to invert the probability



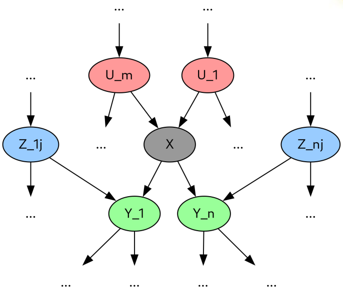
Markov Blanket of a Node

- The **Markov blanket** of a node X consists of:
  1. The **parents** of X
     - The nodes that influence X
  2. The **children** of X
     - The nodes that are directly influenced by X
  3. The **spouses** of X
     - The nodes that are parents of the children nodes
     - I.e., "co-parent"



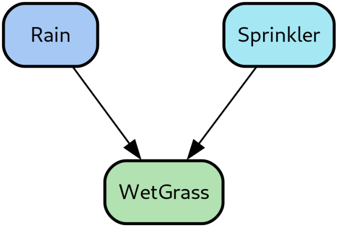
Conditional Independence on Markov Blanket

- In a Bayesian network, each variable is conditionally independent of :
  - **Its predecessors** given its parents (by construction)
  - **All other nodes** in the network given its Markov blanket, i.e., its parents, its children, and its spouses
- The Markov blanket of a node X_i:
  - Contains all the nodes necessary to predict the state of the node X_i, making the network irrelevant
  - Enables efficient and localized inference



How Can a Node Be Influenced by Its Children?

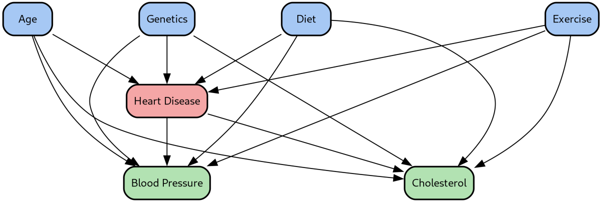
- A descendant can influence its ancestor indirectly through "explaining away" (diagnostic model)
  - Information flows both ways (casual and diagnostic)
  - Evidence about the descendant can change what we believe about the ancestor through dependent paths
- E.g.,
  - Consider the Bayesian network for the Garden World
  - You know the grass is wet
  - This evidence increases the probability of either causes *Rain* or *Sprinkler*
  - If you find out that the *Sprinkler* was on, this "explains away" the *WetGrass*, and the probability of *Rain* goes down
  - The evidence from a descendant *WetGrass* can update your belief about an ancestor (*Rain*)



SCIENCE ACADEMY

Markov Blanket: Medical Example

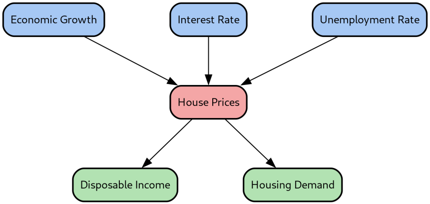
- Consider risk factors and outcomes for heart disease



- **Target node**
- **Parent nodes** (direct influence of H, risk factors)
- **Children nodes** (directly influenced by H, outcomes)
- Note that A, G, D, E also influence BP and C so they are **spouse nodes** of H
- Knowing the state of A, G, D, E, BP, C (Markov Blanket) allows to compute H, without any other information

Markov Blanket: Economic Example

- Consider factors affecting house prices in a particular region
- **Target node**
  - House prices
- **Parent nodes**
  - Economic growth
  - Interest rate
  - Unemployment rate

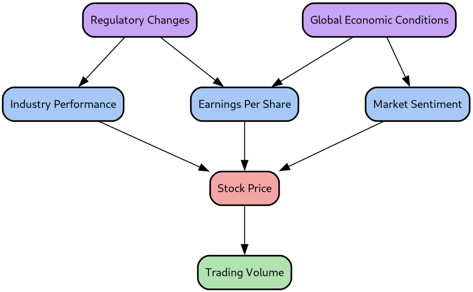


- **Children nodes**
  - Disposable income
    - The house price affects how much money people have left after housing costs
  - Demand for houses
    - Higher prices can reduce demand

SCIENCE ACADEMY

Markov Blanket: Finance Example

- Consider factors affecting an individual company's stock price
- **Target node**
  - *SP*: Stock Price
- **Parent nodes**
  - *IP*: Industry performance
  - *EPS*: Earnings per share
  - *MS*: Market sentiment



- **Children nodes**
  - *TV*: Trading volume
    - Changes in stock price influence how much stock is being traded
- **Grandparents nodes**
  - *RC*: Regulatory changes in the technology sector
    - Influences *IP* and *EPS*, but not directly *TV*
  - *GE*: Global economic conditions
    - Influences *MS* and *EPS*, but not directly *TV*

SCIENCE ACADEMY