Independence in Belief Networks

CMPUT 366: Intelligent Systems

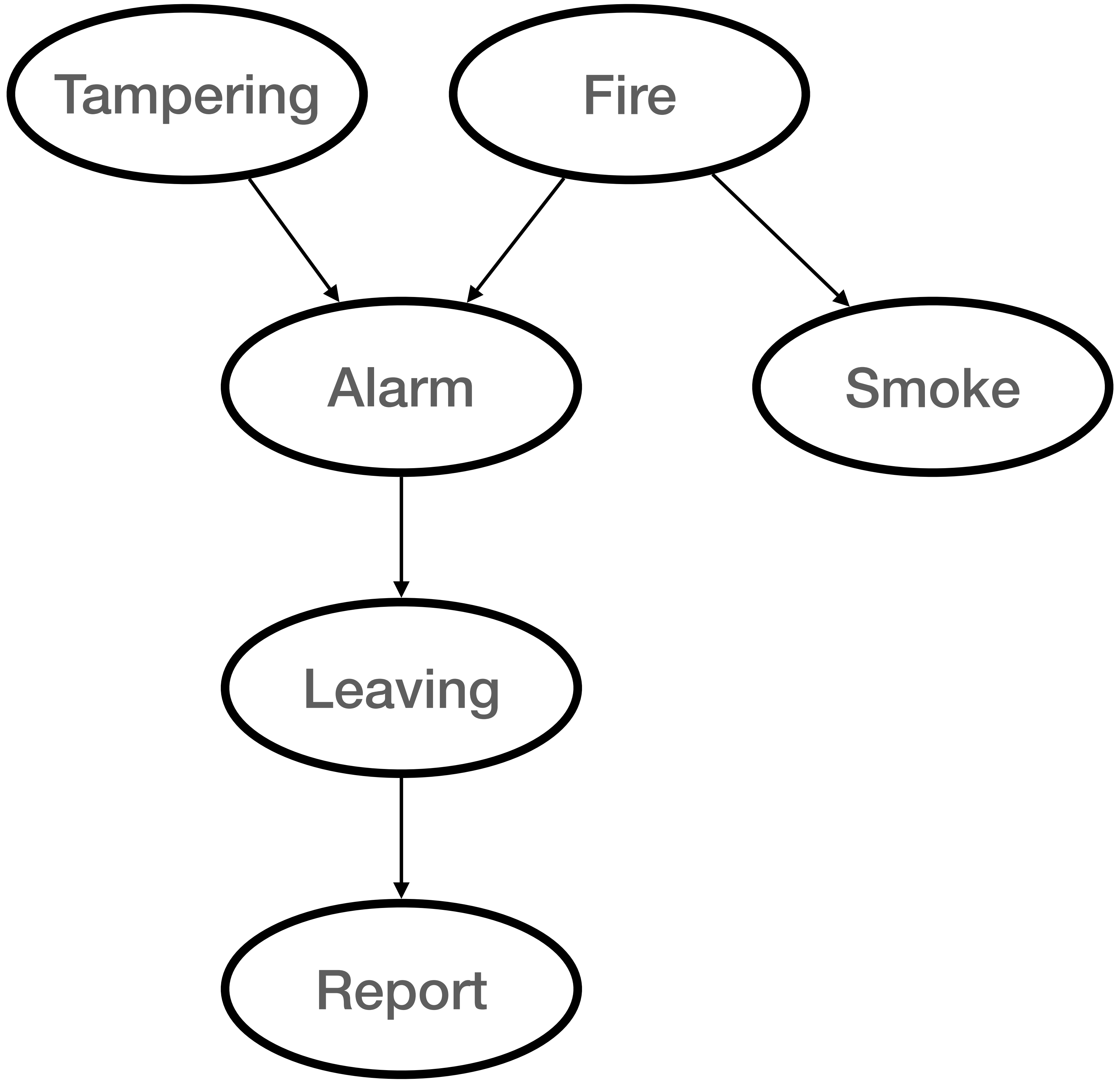
P&M §8.4

Lecture Outline

1. Recap
2. Belief Networks as Factorings
3. Independence in Belief Networks

Recap: Belief Network Semantics

- Graph representation represents a specific **factorization** of the full **joint distribution**
  - Distribution on each node **conditional on its parents**
  - **Marginal distributions** on nodes with no parents
  - **Product** of these distributions is the joint distribution
  - Not every possible factorization is a **correct** factorization
- **Semantics:**
  Every node is **independent** of its **non-descendants**, **conditional only** on its **parents**



Recap: Mechanically Constructing Belief Networks

Given a **joint distribution** we can mechanically construct a **correct** encoding:

1. Order the variables $X_1, X_2, \ldots, X_n$ and associate each one with a **node**
2. For each variable X_i:

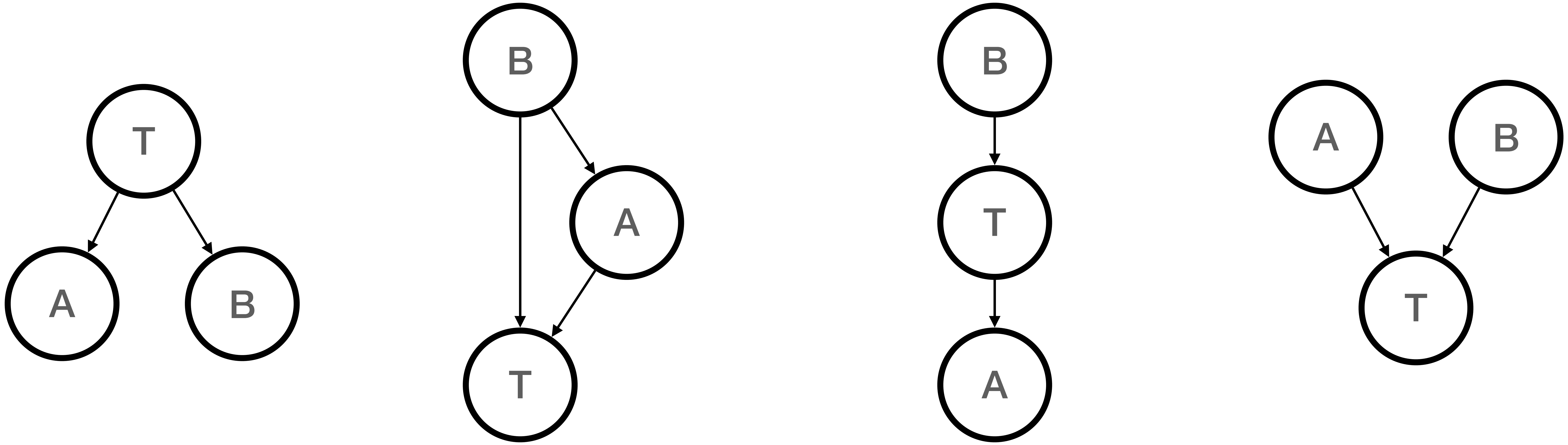
   (i) Choose a **minimal** set of variables $parents(X_i)$ from $X_1, \ldots, X_{i-1}$ such that $P(X_i \mid parents(X_i)) = P(X_i \mid X_1, \ldots, X_{i-1})$

   (ii) i.e., **conditional** on $parents(X_i)$, X_i is **independent** of all the other variables that are **earlier** in the ordering

   (iii) Add an **arc** from each variable in $parents(X_i)$ to X_i

   (iv) Label the node for X_i with the **conditional probability table** $P(X_i \mid parents(X_i))$

Belief Networks as Factorings



- A joint distribution can be factored in **multiple** different ways
  - *Every* variable ordering induces at least one correct factoring (**Why?**)
- A belief network represents a **single** factoring
- Some factorings are correct, some are incorrect

Questions:

1. Does applying the **Chain Rule** to a given variable ordering give a **unique** factoring?
2. Does a given variable ordering correspond to a **unique** Belief Network?

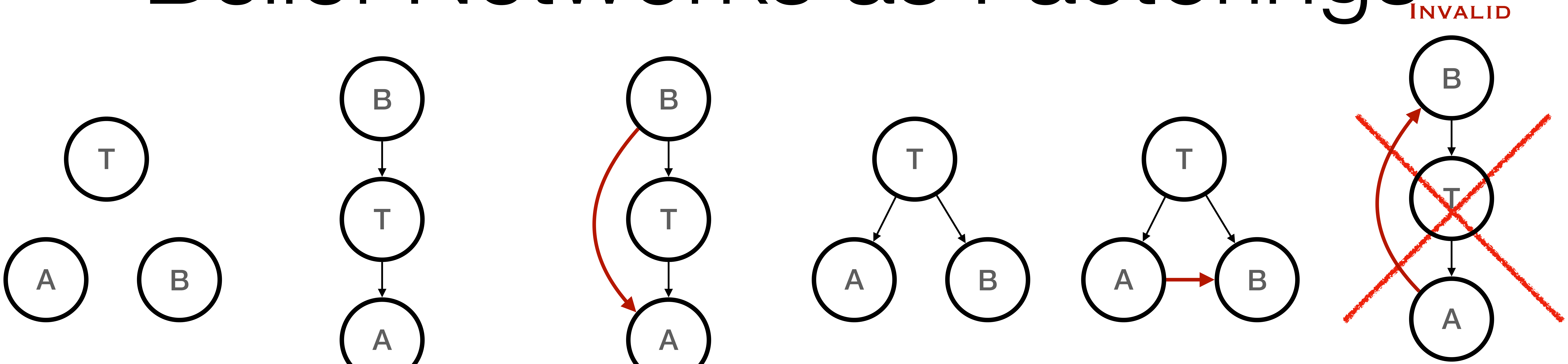
Correct and Incorrect Factorings in the Clock Scenario

Which of the following are **correct** factorings of the joint distribution $P(A, B, T)$ in the Clock Scenario?

1. $P(A)P(B)P(T)$
2. $P(A)P(B \mid A)P(T \mid A, B)$ Chain rule(A,B,T): $P(A)P(B \mid A)P(T \mid A, B)$
3. $P(T)\boxed{P(B \mid T)}P(A \mid T)$ Chain rule(T,B,A): $P(T)\boxed{P(B \mid T, A)}P(A \mid T)$

Which of the above are a **good** factoring for the Clock Scenario? **Why?**

Belief Networks as Factorings



$P(A)P(B)P(T)$

$P(B)P(T \mid B)P(A \mid T)$

$P(B)P(T \mid B)P(A \mid B, T)$

$P(T)P(A \mid T)P(B \mid T)$

$P(T)P(A \mid T)P(B \mid A, T)$

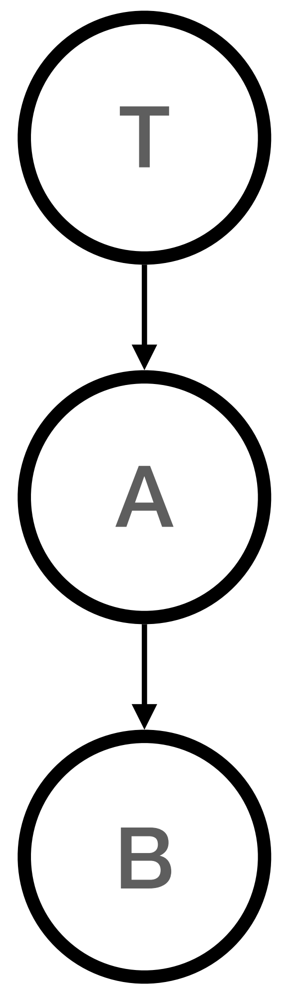
~~$P(B \mid A)P(T \mid B)P(A \mid T)$~~

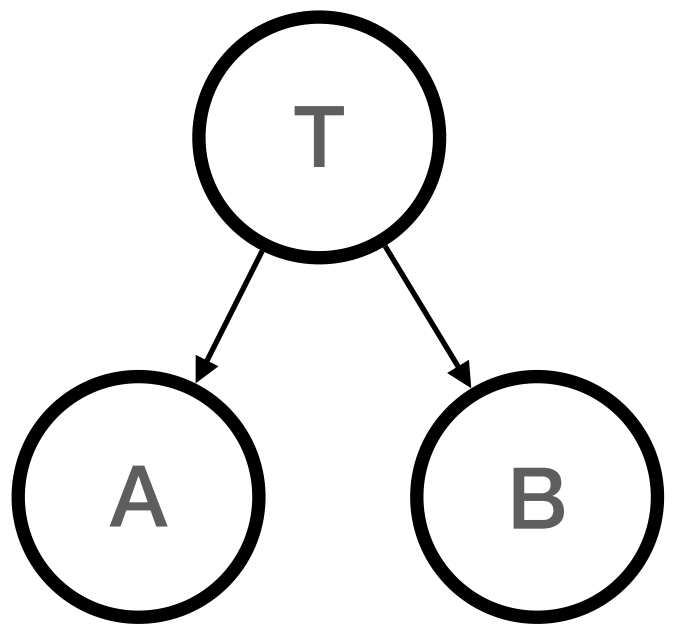
Question: What factoring is represented by each network?

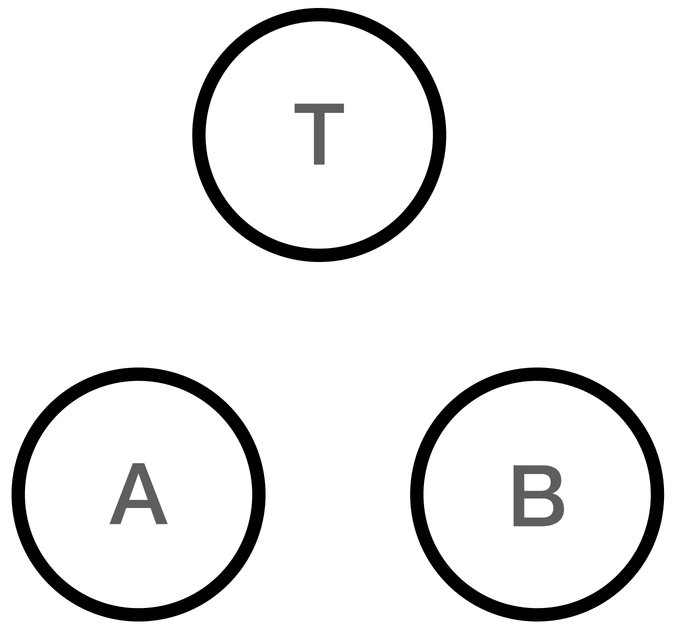
Conditional independence guarantees are represented in belief networks by the absence of edges.

Variations on the Clock Scenario

- A valid belief is only "correct" or "incorrect" with respect to a given joint distribution
- A **single network** may be correct in one scenario and incorrect in another
- **Telephone Clock Scenario:** Alice looks at the clock, then tells Bob the time over a noisy phone connection
- **Desert Islands Clock Scenario:** Alice is on Island A. Bob is on Island B. The clock is on Island C. Alice and Bob cannot see or hear each other, or the clock.







Independence in a Joint Distribution

Question: How can we answer questions about independence using the **joint distribution**?

Examples using $P(A, B, T)$:

1. Is A independent of B?
   - $P(A = a \mid B = b) = P(A = a)$ **for all** $a \in \text{dom}(A), b \in \text{dom}(B)$?
2. Is T independent of A?
   - $P(T = t \mid A = a) = P(T = t)$ **for all** $a \in \text{dom}(A), t \in \text{dom}(T)$?
3. Is A independent of B given T?
   - $P(A = a \mid B = b, T = t) = P(A = a \mid T = t)$ **for all** $a \in \text{dom}(A), b \in \text{dom}(B), t \in \text{dom}(T)$?

$$P(A, B) = \sum_{t \in T} P(A, B, T = t)$$

$$P(A, T) = \sum_{b \in B} P(A, B = b, T)$$

$$P(B, T) = \sum_{a \in A} P(A = a, B, T)$$

$$P(A) = \sum_{b \in B} P(A, B = b)$$

$$P(B) = \sum_{a \in A} P(A = a, B)$$

$$P(T) = \sum_{a \in A} P(A = a, T)$$

$$P(A \mid B, T) = \frac{P(A, B, T)}{P(B, T)}$$

$$P(A \mid B) = \frac{P(A, B)}{P(B)}$$

$$P(A \mid T) = \frac{P(A, T)}{P(T)}$$

$$P(T \mid A) = \frac{P(A, T)}{P(A)}$$

Belief Network Semantics

Definition: A belief network represents a joint distribution that can be factored as

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid parents(X_i))$$

Theorem: (Belief Network Semantics)
Every node is **independent** of its **non-descendants**, **conditional only** on its **parents**

Proof:

1. X_j is a descendant of $X_i \implies i < j$
2. For all $i > j$, $P(X_i \mid parents(X_i), X_j) = P(X_i \mid parents(X_i))$
3. For all $i < j$, if j is not a descendant of i, then $P(X_i \mid parents(X_i), X_j) = P(X_i \mid parents(X_i))$

Belief Network Semantics: Proof (2)

$$Z \doteq \{X_1, \ldots, X_{i-1}\} \backslash parents(X_i) \cup \{X_j\}$$
$$= \left\{X_k \mid 1 \leq k \leq i-1, k \neq i, k \neq j, X_k \notin parents(X_i)\right\}$$

$$P(X_1, \ldots, X_i) = P(X \mid parents(X_i))P(X_1, \ldots, X_{i-1}) = P(X_i \mid parents(X_i)) \prod_{k=1}^{i-1} P(X_k \mid parents(X_k))$$ def. belief network

$$P(X_i, X_j, parents(X_i)) = \sum_Z P(X_1, \ldots, X_i) = \sum_Z P(X_i \mid parents(X_i)) \prod_{k=1}^{i-1} P(X_k \mid parents(X_k))$$ marginalization

$$P(X_i \mid parents(X_i), X_j) = \frac{P(X_i, parents(X_i), X_j)}{P(parents(X_i), X_j)}$$ def. conditional probability

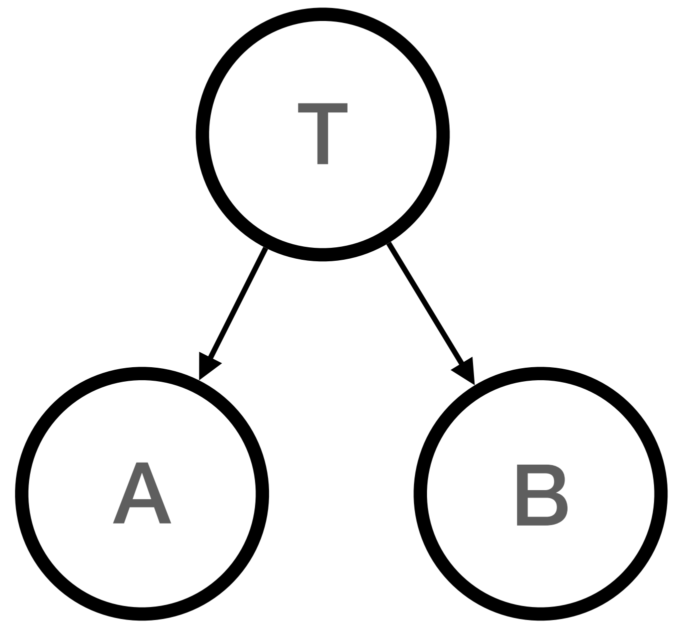
$$= \frac{\sum_Z P(X_i \mid parents(X_i)) \prod_{k=1}^{i-1} P(X_k \mid parents(X_k))}{\sum_Z \prod_{k=1}^{i-1} P(X_k \mid parents(X_k))}$$

$$= P(X_i \mid parents(X_i)) \underbrace{\frac{\sum_Z \prod_{k=1}^{i-1} P(X_k \mid parents(X_k))}{\sum_Z \prod_{k=1}^{i-1} P(X_k \mid parents(X_k))}}_{=1} = P(X_i \mid parents(X_i)) \blacksquare$$

Independence in a Belief Network

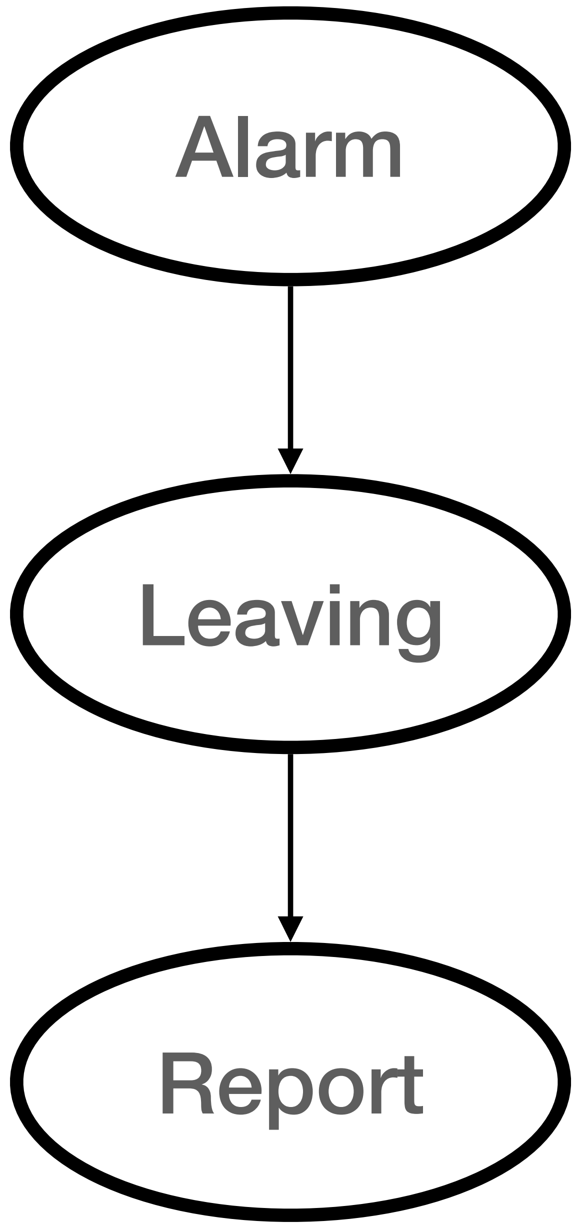
Belief Network Semantics:
Every node is **independent** of its **non-descendants**, **conditional only** on its **parents**

- We can use the semantics of a correct belief network to answer questions about independence
- Examples using the belief network at right:
  1. Is **T** independent of **A**?
  2. Is **A** independent of **B** given **T**?
  3. Is **A** independent of **B**?



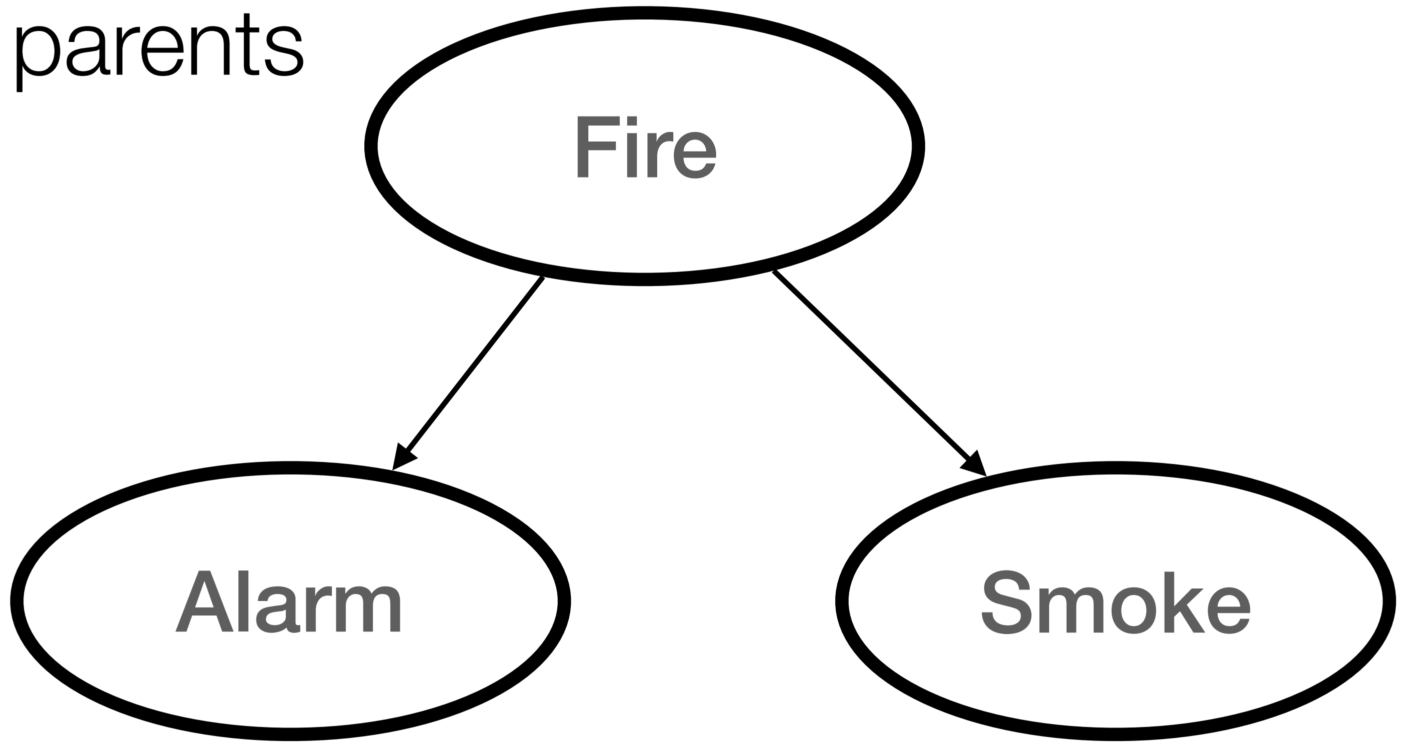
Chain

- **Question:** Is **Report** independent of **Alarm** given **Leaving**?
  - *Intuitively:* The only way learning **Report** tells us about **Alarm** is because it tells us about **Leaving**; but **Leaving** has already been observed
  - *Formally:* **Report** is independent of its non-descendants given only its parents
    - **Leaving** is **Report's** parent
    - **Alarm** is a non-descendant of **Report**
- **Question:** Is **Report** independent of **Alarm**?
  - *Intuitively:* Learning **Report** gives us information about **Leaving**, which gives us information about **Alarm**
  - *Formally:* **Report** is independent of **Alarm** given **Report's** parents; but the question is about **marginal** independence



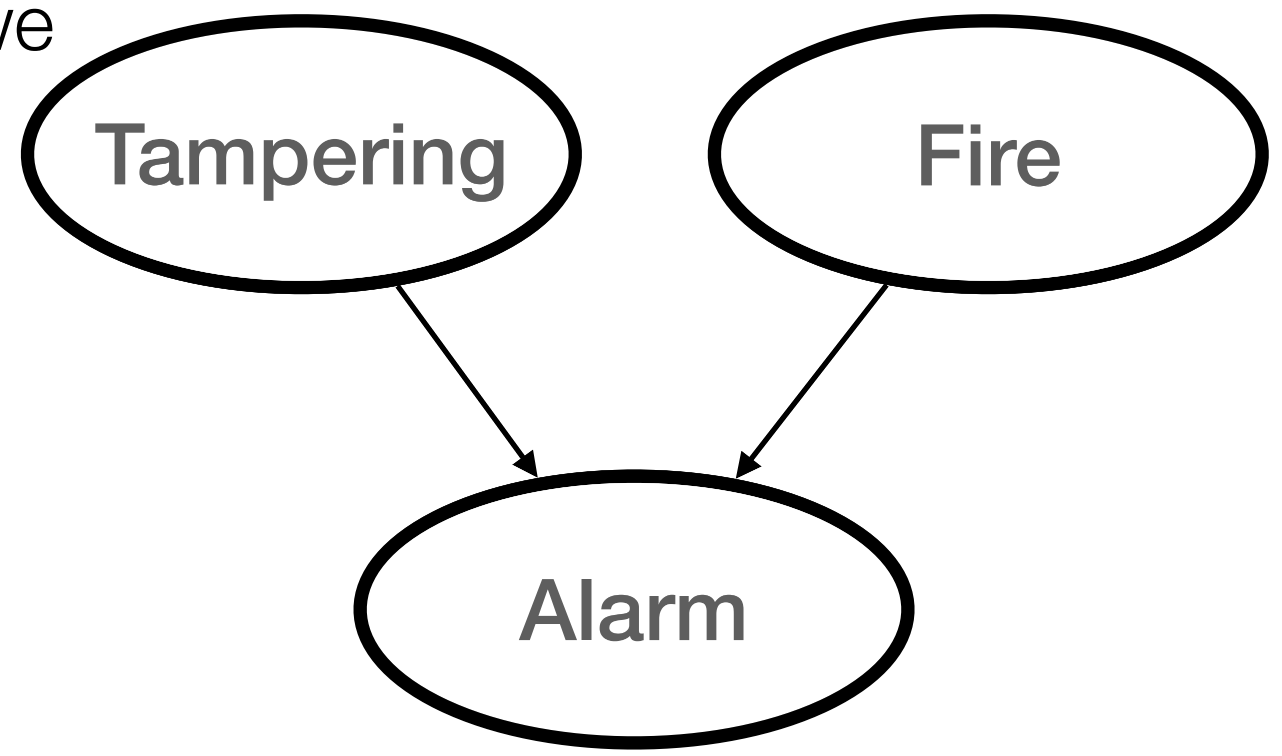
Common Ancestor

- **Question:** Is **Alarm** independent of **Smoke** given **Fire**?
  - *Intuitively:* The only way learning **Smoke** tells us about **Alarm** is because it tells us about **Fire**; but **Fire** has already been observed
  - *Formally:* **Alarm** is independent of its non-descendants given only its parents
    - **Fire** is **Alarm's** parent
    - **Smoke** is a non-descendant of **Alarm**
- **Question:** Is **Alarm** independent of **Smoke**?
  - *Intuitively:* Learning **Smoke** gives us information about **Fire**, which gives us information about **Alarm**
  - *Formally:* **Alarm** is independent of **Smoke** given only **Alarm's** parents; but the question is about **marginal independence**



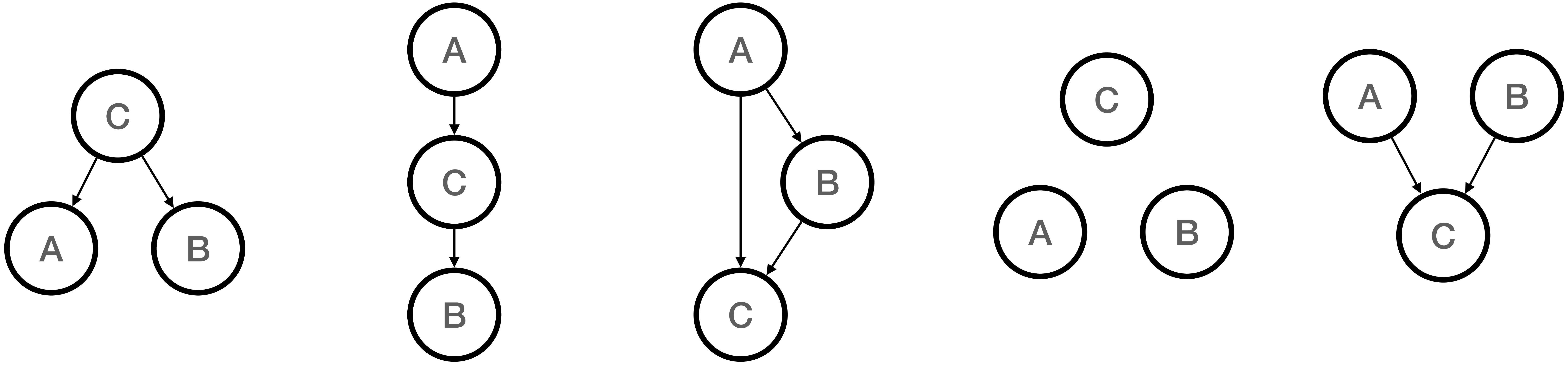
Common Descendant

- **Question:** Is **Tampering** independent of **Fire** given **Alarm**?
  - *Intuitively:* If we know **Alarm** is ringing, then both **Tampering** and **Fire** are more likely. If we then learn that **Fire** is false, that makes it more likely that the **Alarm** is ringing because of **Tampering**.
  - *Formally:* **Tampering** is independent of **Fire** given **only** **Tampering's** parents; but we are conditioning on one of Tampering's **descendants**
    - Conditioning on a **common descendant** can make independent variables dependent through this **explaining away** effect
- **Question:** Is **Tampering** (marginally) independent of **Fire**?
  - *Intuitively:* Learning **Tampering** doesn't tell us anything about whether a **Fire** is happening
  - *Formally:* **Tampering** is independent of **Fire** given **Tampering's** parents
    - **Tampering** has no parents, so we are always conditioning on them
    - **Fire** is a non-descendant of **Tampering**



Correctness of a Belief Network

A belief network is a **correct** representation of a joint distribution when the belief network answers "yes" to an independence question **only if** the **joint distribution** answers "yes" to the same question.



Questions:

1. Is A independent of B in the above belief networks?
2. Is A independent of B given C in the above belief networks?

Summary

- A belief network represents a specific **factoring** of a joint distribution
  - More than one belief network can correctly represent a joint distribution
  - A given belief network may be correct for one underlying joint distribution and incorrect for another
- A **good** belief network is one that encodes as many **true** conditional independence relationships as possible
- It is possible to read the conditional independence guarantees made by a belief network directly from its **graph structure**