

Belief Networks

CMPUT 261: Introduction to Artificial Intelligence

P&M §8.3

Assignment #1

- Assignment #1 is due **TODAY**
- Submissions will be accepted until **11:59pm TONIGHT**

Recap: Independence

Definition:

Random variables X and Y are **marginally independent** iff

$$P(X = x \mid Y = y) = P(X = x)$$

for all values of $x \in \text{dom}(X)$ and $y \in \text{dom}(Y)$.

Definition:

Random variables X and Y are **conditionally independent given Z** iff

$$P(X = x \mid Y = y, Z = z) = P(X = x \mid Z = z)$$

for all values of $x \in \text{dom}(X)$, $y \in \text{dom}(Y)$, and $z \in \text{dom}(Z)$.

Recap: Chain Rule

Definition: Chain rule (of probabilities)

$$\begin{aligned} P(\alpha_1, \dots, \alpha_n) &= P(\alpha_1) \times P(\alpha_2 \mid \alpha_1) \times \dots \times P(\alpha_n \mid \alpha_1, \dots, \alpha_{n-1}) \\ &= \prod_{i=1}^n P(\alpha_i \mid \alpha_1, \dots, \alpha_{i-1}) \end{aligned}$$

Recap: Chain Rule

Definition: Chain rule (of probabilities)

$$\begin{aligned} P(X_1 = x_1, \dots, X_n = x_n) &= P(X_1 = x_1) \times P(X_2 = x_2 \mid X_1 = x_1) \times \dots \times P(X_n = x_n \mid X_1 = x_1, \dots, X_{n-1} = x_{n-1}) \\ &= \prod_{i=1}^n P(X_i = x_i \mid X_1 = x_1, \dots, X_{i-1} = x_{i-1}) \end{aligned}$$

$P(W, X, Y, Z)$

$P(W, X, Y)$

$P(W, X)$

$$P(W, X, Y, Z) = \underbrace{P(W)P(X \mid W)}_{P(W, X)} P(Y \mid W, X) P(Z \mid W, X, Y)$$

Recap:

Exploiting Independence

- Explicitly specifying an entire **unstructured joint distribution** is tedious and unnatural
- We can exploit **conditional independence**:
 - Conditional distributions are often more **natural** to write
 - Joint probabilities can be extracted from conditionally independent distributions by **multiplication**

Lecture Outline

1. Recap & Logistics
2. Belief Networks as Factorings
3. Querying Joint Probabilities
4. Querying Independence

After this lecture, you should be able to:

- Define a belief network
- Construct a belief network that corresponds to a given factoring
- Recover a factoring that is consistent with a given belief network
- Compute joint probabilities using a belief network
- Identify independence relationships encoded by a given belief network

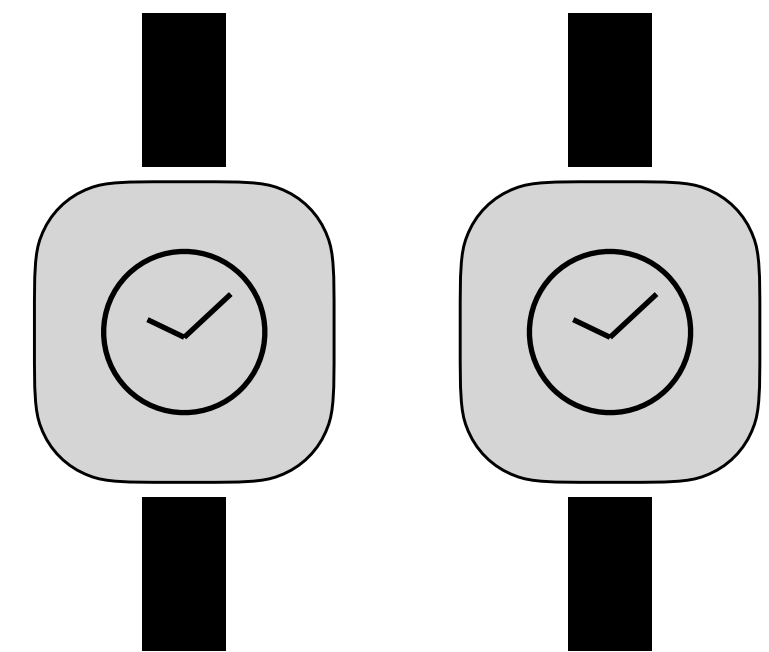
Factoring Joint Distributions

- We can **always** represent a joint distribution as a product of factors, even when there is **no** marginal or conditional independence (**why?**)

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

- Question:** How much space can we save with this factored representation?
- When we do have independence, we can **simplify** some of these factors:

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



Two watches

Random variables:

A - Time on Alice's clock

B - Time on Bob's clock

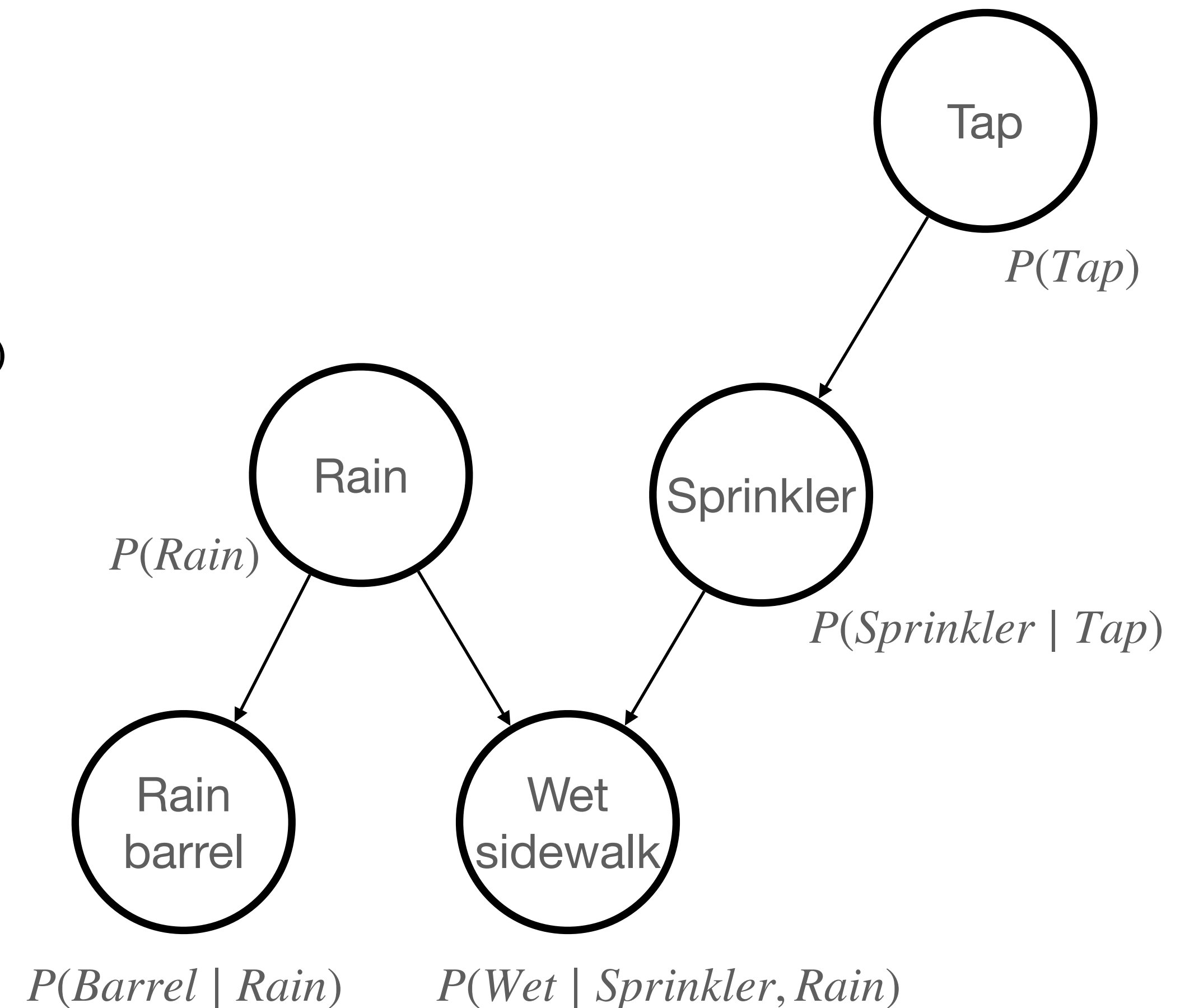
T - Actual time

Belief Networks, informally

We can represent a particular factoring of a joint distribution as a directed acyclic graph:

$$P(\text{Tap}, \text{Rain}, \text{Sprinkler}, \text{Wet}, \text{Barrel}) = \\ P(\text{Tap})P(\text{Rain})P(\text{Sprinkler} \mid \text{Tap})P(\text{Wet} \mid \text{Sprinkler}, \text{Rain})P(\text{Barrel} \mid \text{Rain})$$

- **Nodes** are **random variables**
- Every variable has *exactly one* **factor** in the factoring
- The node's **parents** are the variables that its factor **conditions on**
 - (We'll sometimes say that the factor "depends on" its parents, but that is very imprecise)
- **More** independence means **fewer** arcs (**why?**)



Belief net: Rain barrel has parent Rain, Wet Sidewalk has parents Rain and Sprinkler, Sprinkler has parent Tap

Belief Networks

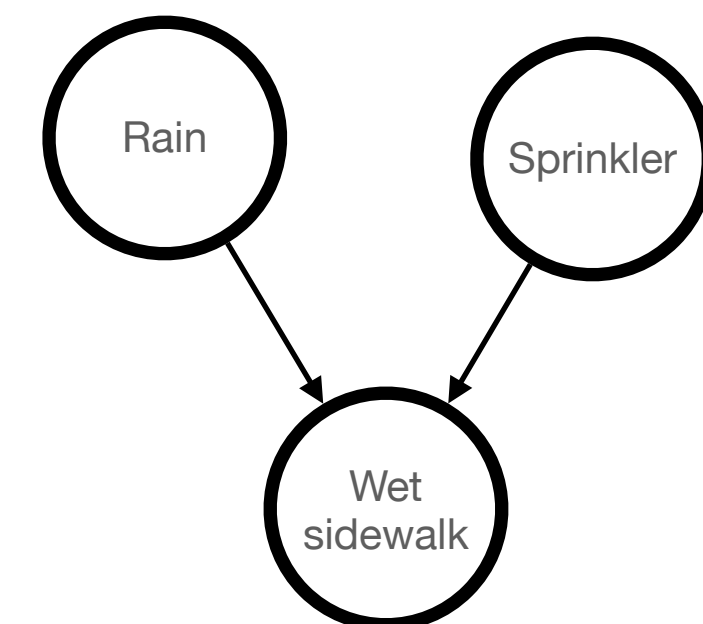
Definition:

A **belief network** (or **Bayesian network**) consists of:

1. A directed acyclic graph, with each node labelled by a **random variable**
2. A **domain** for each random variable
3. A **conditional probability table** for each variable given its **parents**

A table with one row for each **combination** of values of **itself** and **its parents**, and the corresponding conditional probability

Wet	Sprinkler	Rain	P(W S,R)
0	0	0	1.0
1	0	0	0.0
0	1	0	0.5
1	1	0	0.5
0	0	1	0.1
1	0	1	0.9
0	1	1	0.0
1	1	1	1.0



Belief net: Wet sidewalk has parent Rain and Sprinkler

Why is the Graph Encoding Useful?

Encoding the distribution as a graph is useful for a number of reasons:

- Separates the **independence** structure (nodes, arcs) from the **quantitative** probabilities (conditional probability tables)
 - You can often reason about independence without reasoning about actual probability values
- Graph can be specified by reasoning **locally** about independence (i.e., what values fully determine a variable's distribution)
- **Complicated global** independence relationships can then be inferred based on graph structure
- Algorithms that exploit independence can be defined based on the graph structure alone

Watch Scenario

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

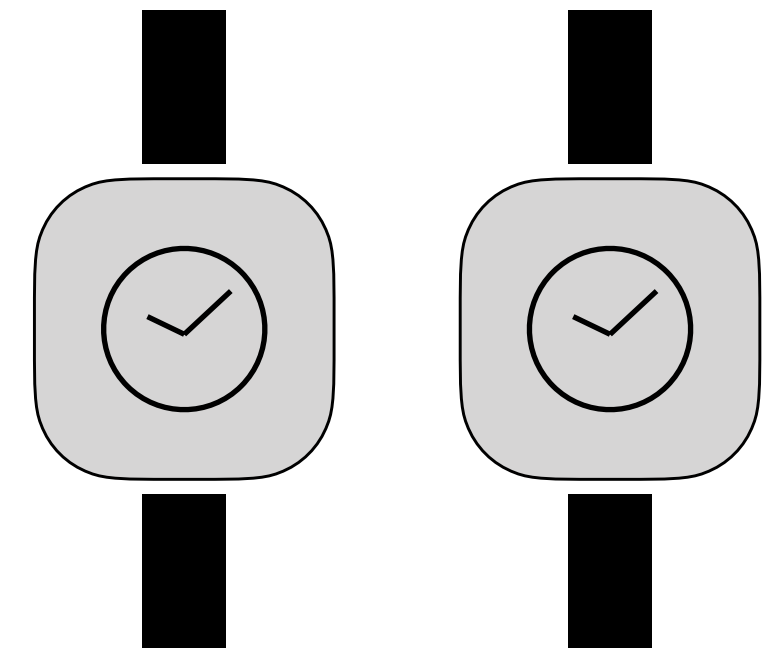
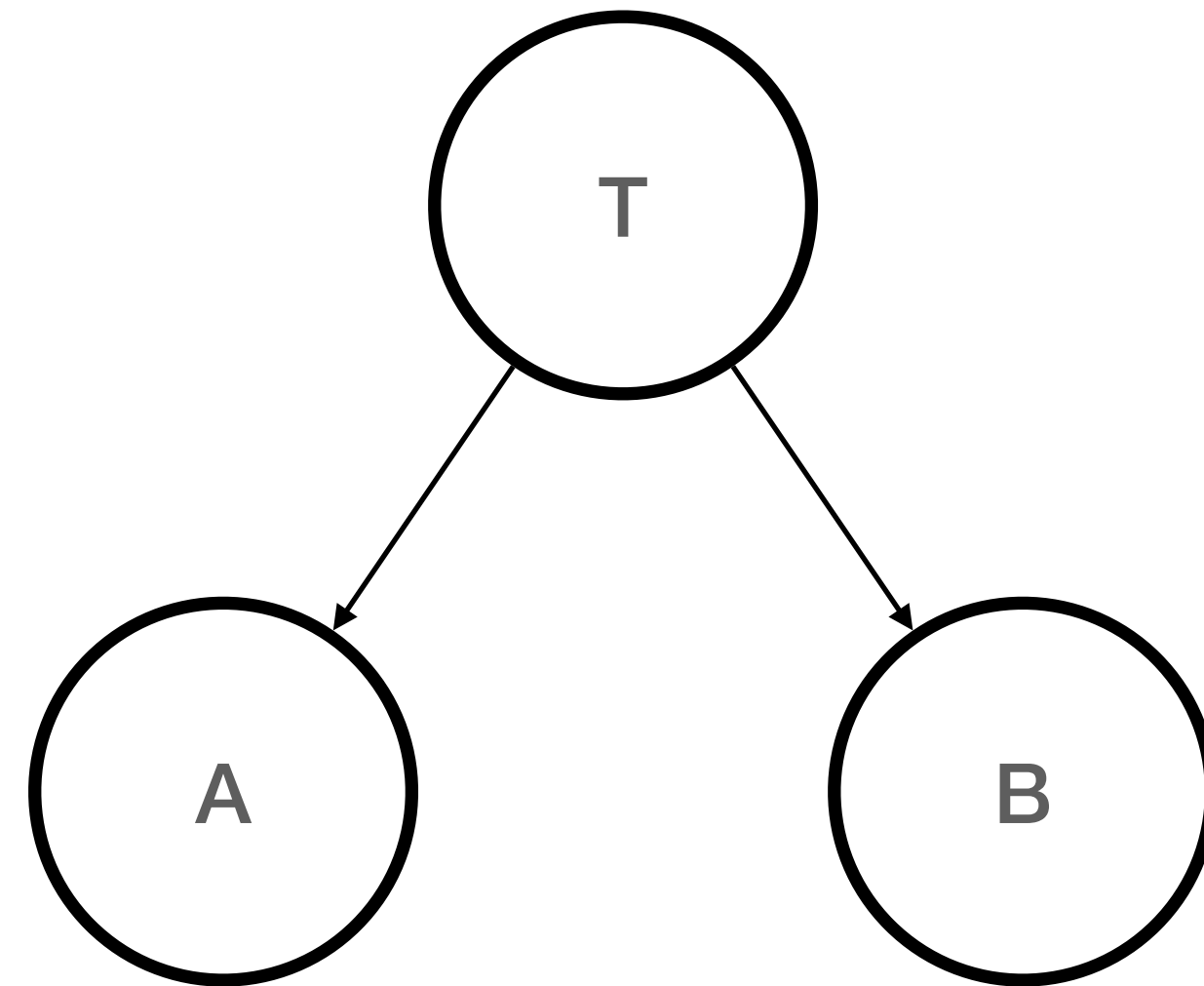


Figure: Two watches



Belief net: A has parent T, B has parent T

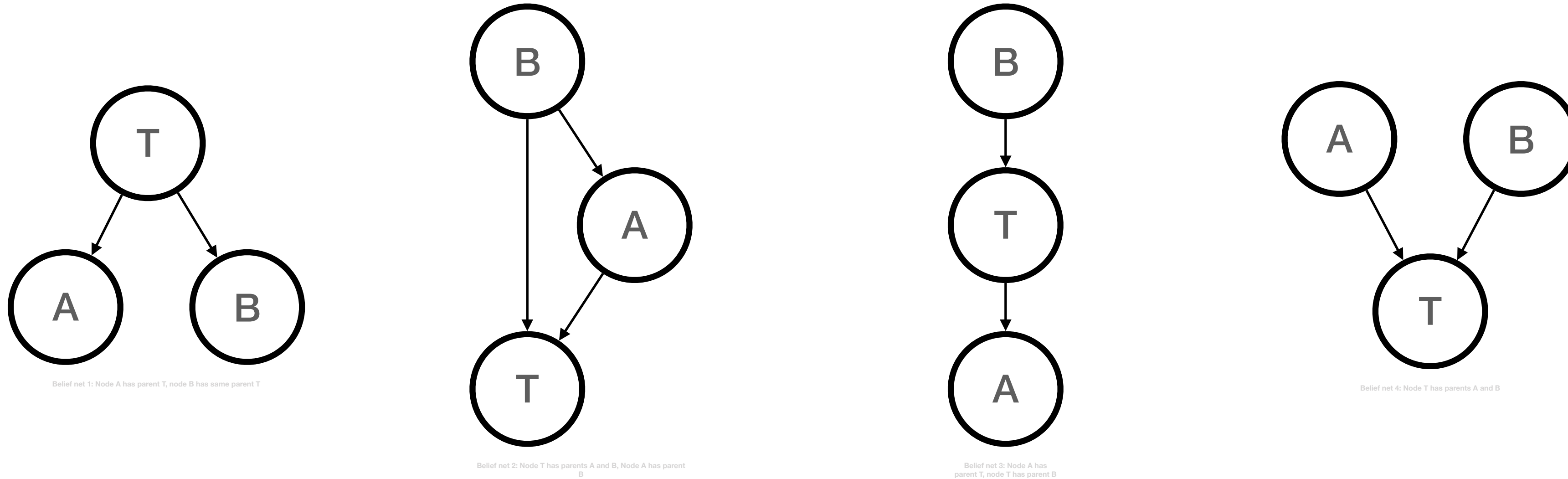
Random variables:

A - Time on Alice's clock

B - Time on Bob's clock

T - Actual time

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
- *For a given joint distribution,* 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

Definition:

A **factoring** of a joint distribution is **correct** when every probability computed by the factoring gives the correct joint probability.

<i>A</i>	<i>B</i>	<i>P(A, B)</i>
0	0	0.45
0	1	0.05
1	0	0.05
1	1	0.45

- In this joint distribution, the factoring $P(A, B) = P(A)P(B)$ is **not correct**
- $P(A = 0) = P(B = 0) = 0.5$
- But $P(A = 0)P(B = 0) = 0.25 \neq P(A = 0, B = 0) = 0.45$

Correct and Incorrect Factorings in the Clock Scenario

Definition:

A **factoring** of a joint distribution is **correct** when every probability computed by the factoring gives the correct joint probability.

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

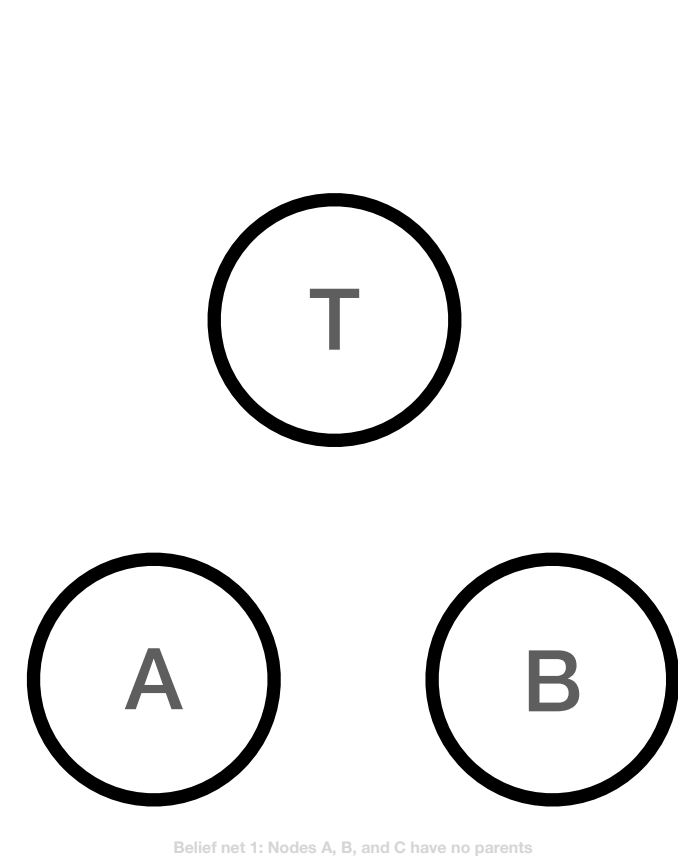
Chain rule(A,B,T): $P(A)P(B \mid A)P(T \mid A, B)$

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

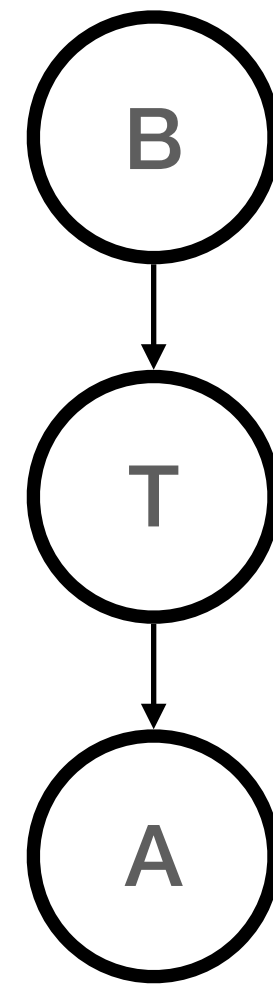
Chain rule(T,B,A): $P(T)P(B \mid T)P(A \mid B, T)$

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

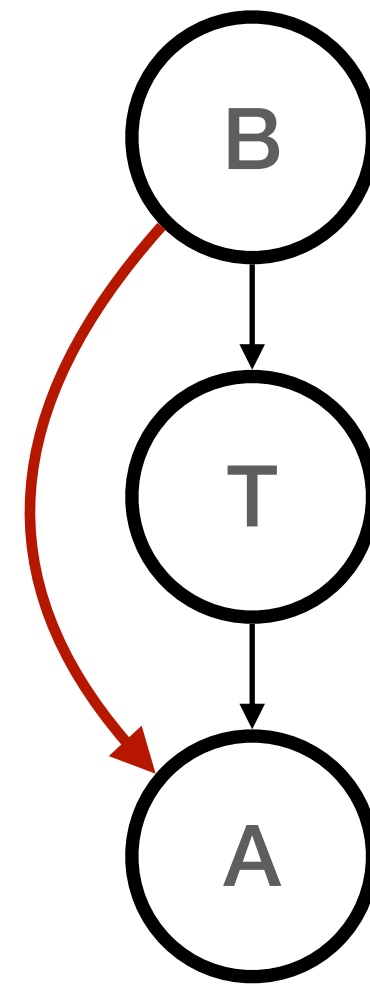
Belief Networks as Factorings



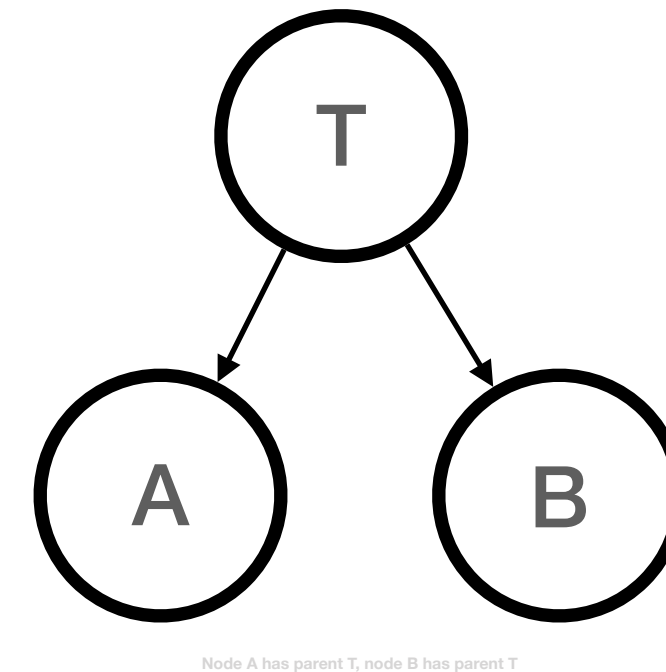
$$P(A)P(B)P(T)$$



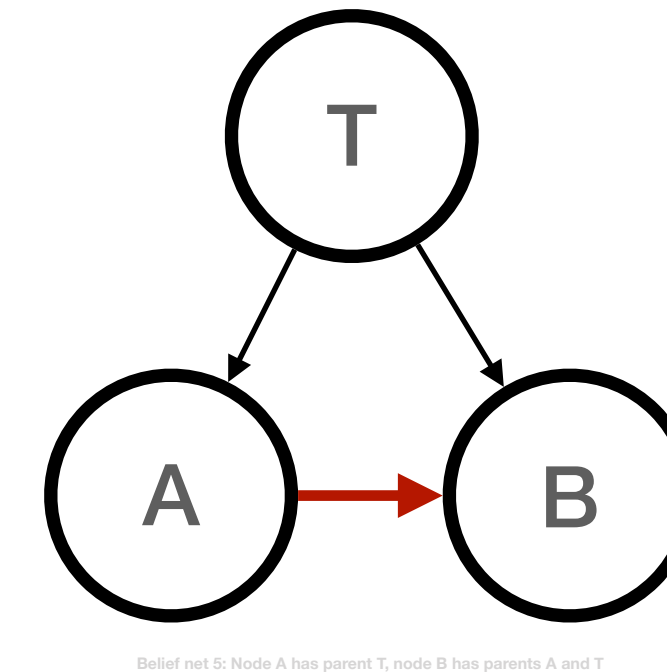
$$P(B)P(T | B)P(A | T)$$



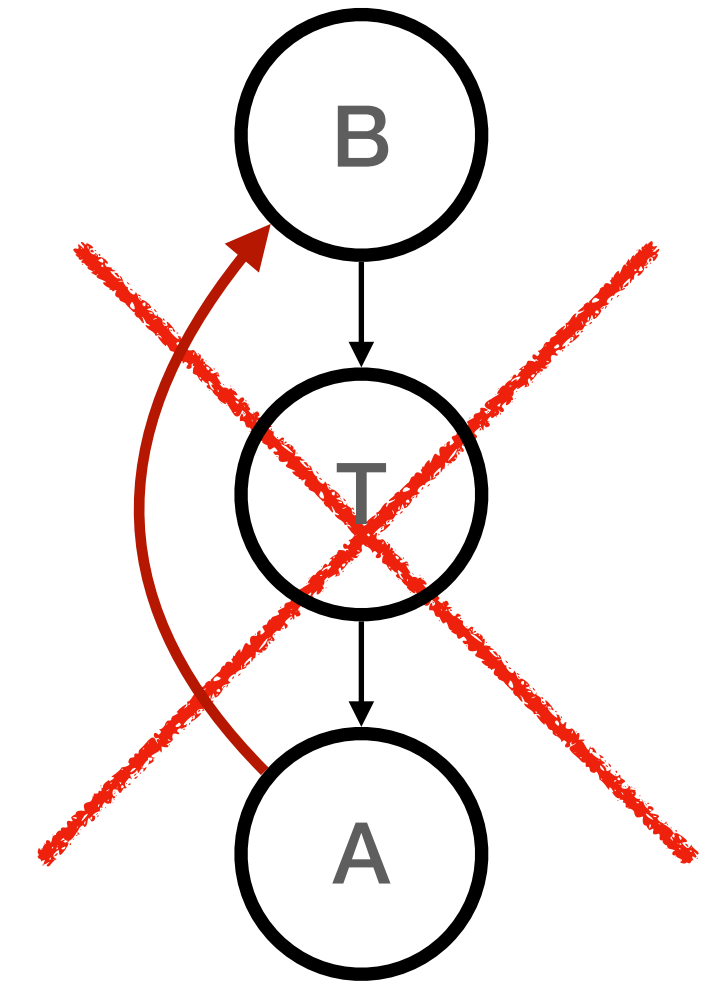
$$P(B)P(T | B)P(A | \textcolor{red}{B}, T)$$



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



$$P(T)P(A | T)P(B | \textcolor{red}{A}, T)$$



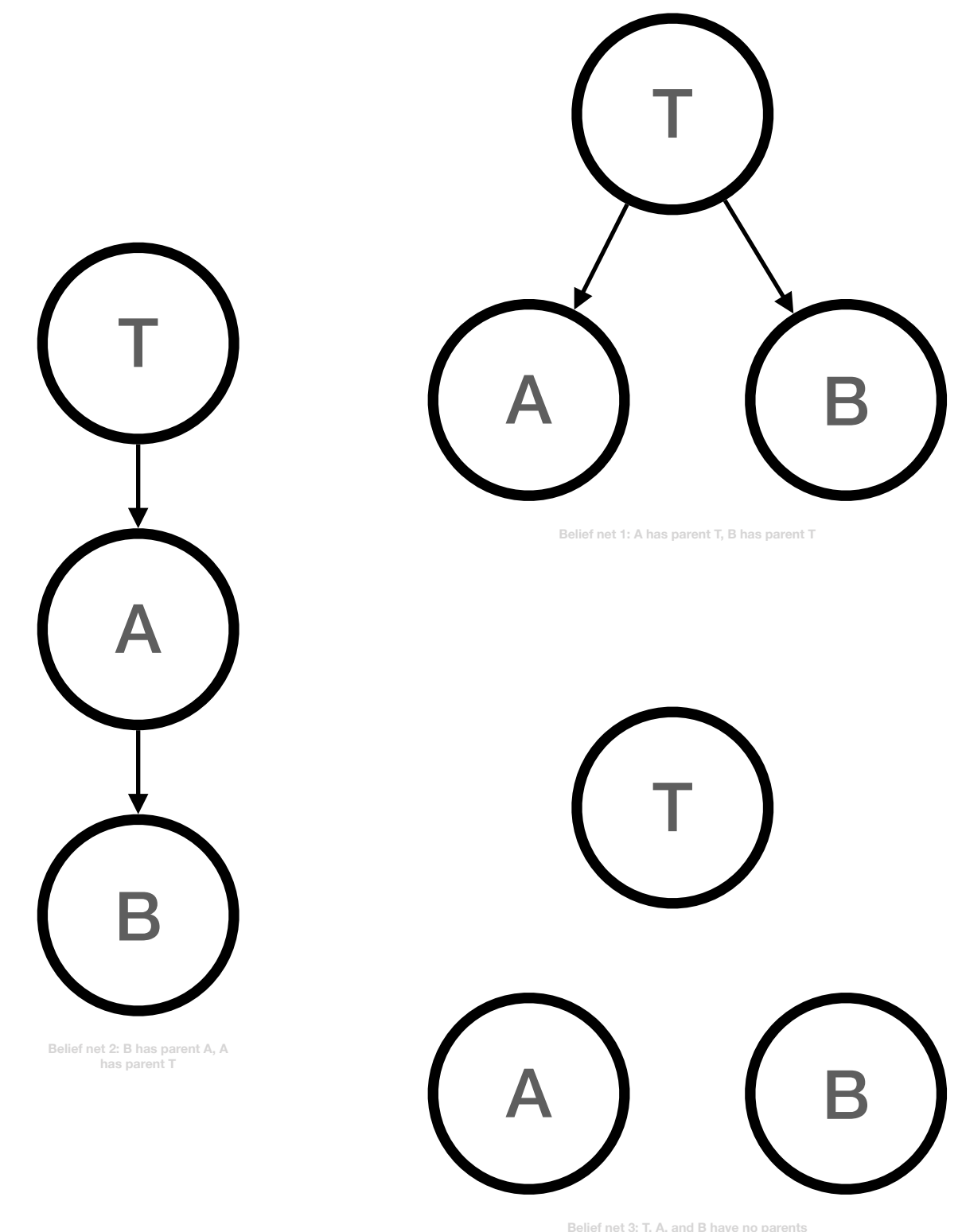
~~$$P(B | \textcolor{red}{A})P(T | B)P(A | T)$$~~

Question: What **factoring** is represented by each network?

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

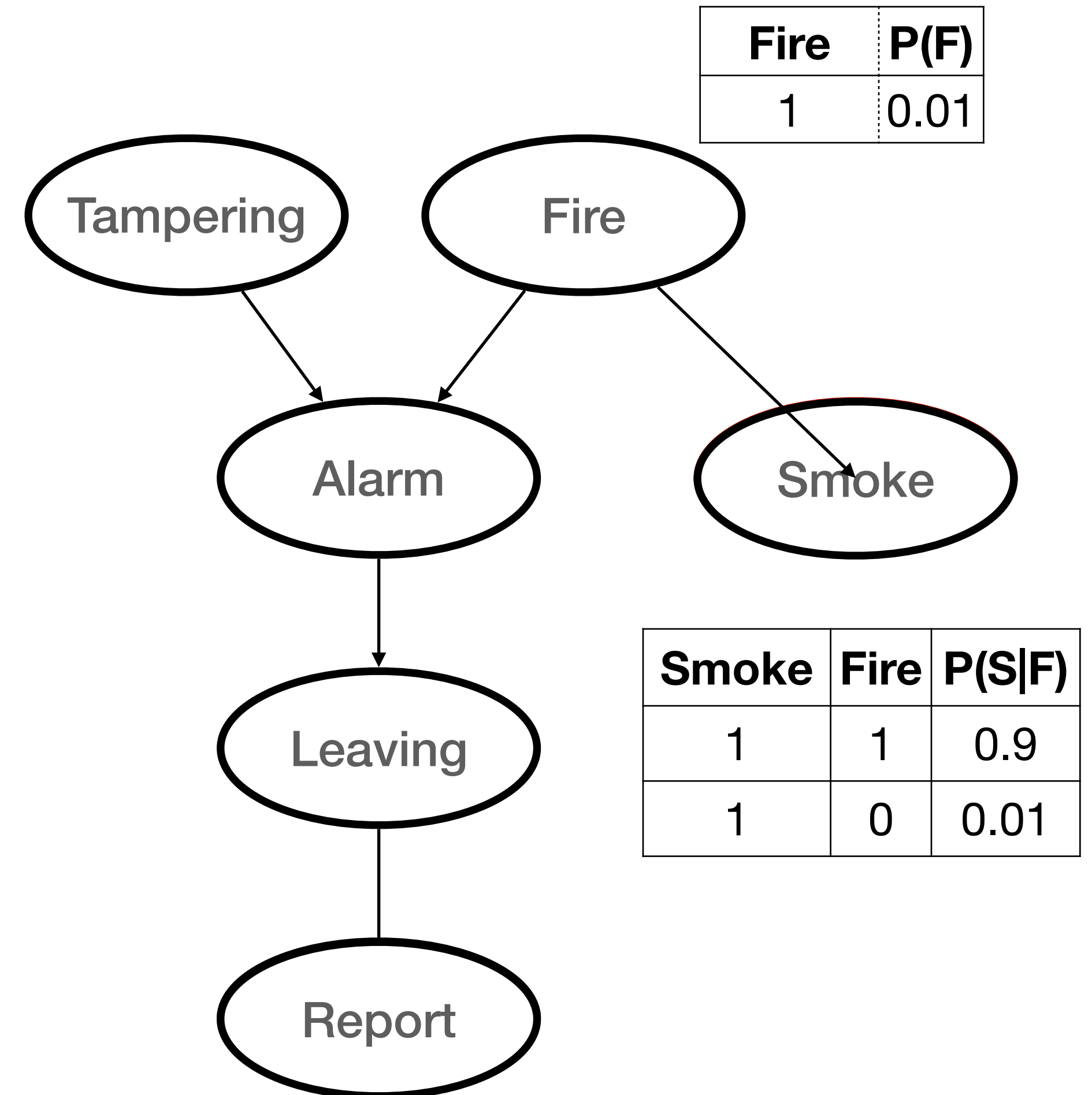
Variations on the Watch Scenario

- A valid belief network 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
- **Shared Watch Scenario:** Bob sets his watch to the time displayed by Alice's watch
- **Dice Watch Scenario:** Alice rolls a sixty-sided die and sets her watch's minutes to the number (minus 1) that comes up. Bob does the same thing.



Queries

- The most common task for a belief network is to query **posterior probabilities** given some **observations**
- **Easy case:**
 - Observations are the **parents** of query target
- More **common** cases:
 - Observations are the **children** of query target
 - Observations have **no straightforward relationship** to the target



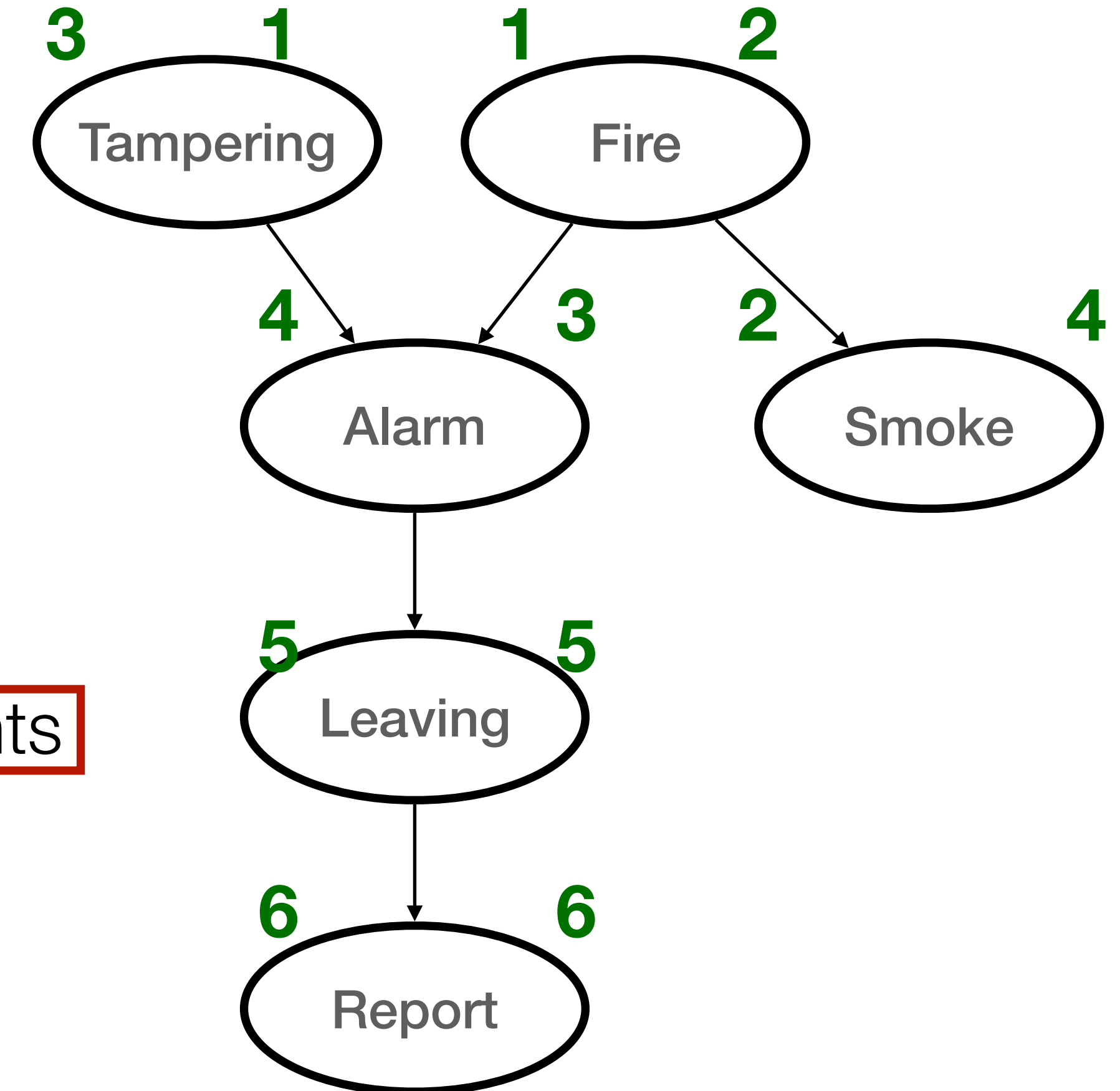
Querying Joint Probabilities: Variable Ordering

To compute joint probability distribution, we need a variable **ordering** that is **consistent** with the graph

for i **from** 1 **to** n :

select an unlabelled variable with no unlabelled parents

label it as i



Question:

Is this **guaranteed** to
exist **at every step**?
Why?

Belief net: Report has parent Leaving, Leaving has Parent Alarm, Alarm has parents Tampering and Fire, Smoke has parent Fire

Querying Joint Probabilities

- Multiply distributions to get joint distribution
- **Example:** Given variable ordering
Tampering, Fire, Alarm, Smoke, Leaving

Questions:

- Why $P(\textit{Fire})$ instead of $P(\textit{Fire} \mid \textit{Tampering})$?
- Why $P(\textit{Smoke} \mid \textit{Fire})$ instead of $P(\textit{Smoke} \mid \textit{Tampering}, \textit{Fire}, \textit{Alarm})$?

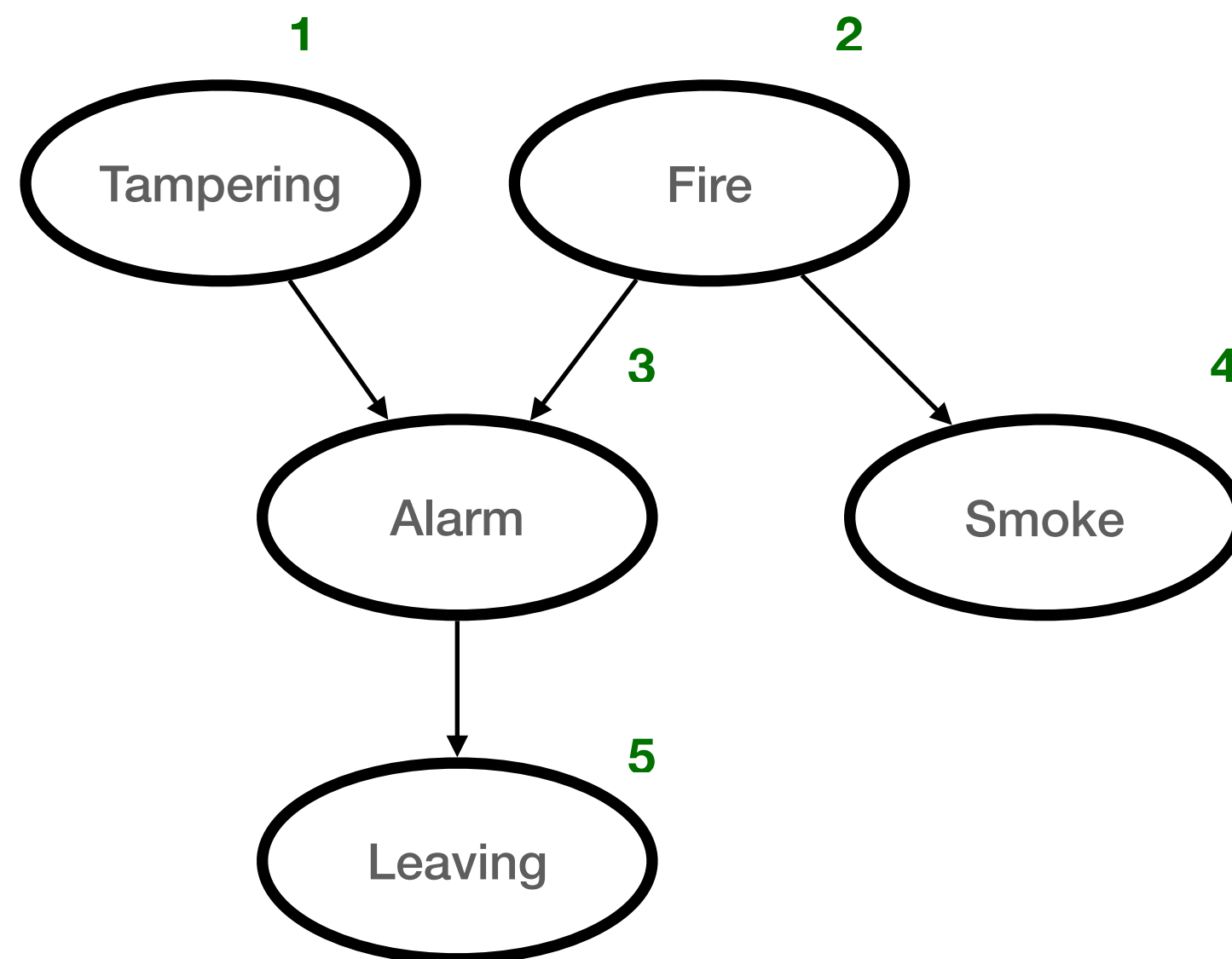
$$P(\textit{Tampering}) = P(\textit{Tampering})$$

$$P(\textit{Tampering}, \textit{Fire}) = P(\textit{Fire})P(\textit{Tampering})$$

$$P(\textit{Tampering}, \textit{Fire}, \textit{Alarm}) = \\ P(\textit{Alarm} \mid \textit{Tampering}, \textit{Fire})P(\textit{Fire})P(\textit{Tampering})$$

$$P(\textit{Tampering}, \textit{Fire}, \textit{Alarm}, \textit{Smoke}) = \\ P(\textit{Smoke} \mid \textit{Fire})P(\textit{Alarm} \mid \textit{Tampering}, \textit{Fire})P(\textit{Fire})P(\textit{Tampering})$$

$$P(\textit{Tampering}, \textit{Fire}, \textit{Alarm}, \textit{Smoke}, \textit{Leaving}) = \\ P(\textit{Leaving} \mid \textit{Alarm})P(\textit{Smoke} \mid \textit{Fire})P(\textit{Alarm} \mid \textit{Tampering}, \textit{Fire})P(\textit{Fire})P(\textit{Tampering})$$



Independence in a Joint Distribution

Question: How can we answer questions about independence using the **full 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 $a \in \text{dom}(A), b \in \text{dom}(B), t \in \text{dom}(T)$?

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

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

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

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

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

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

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

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

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

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

Independence in a Belief Network

Definition:

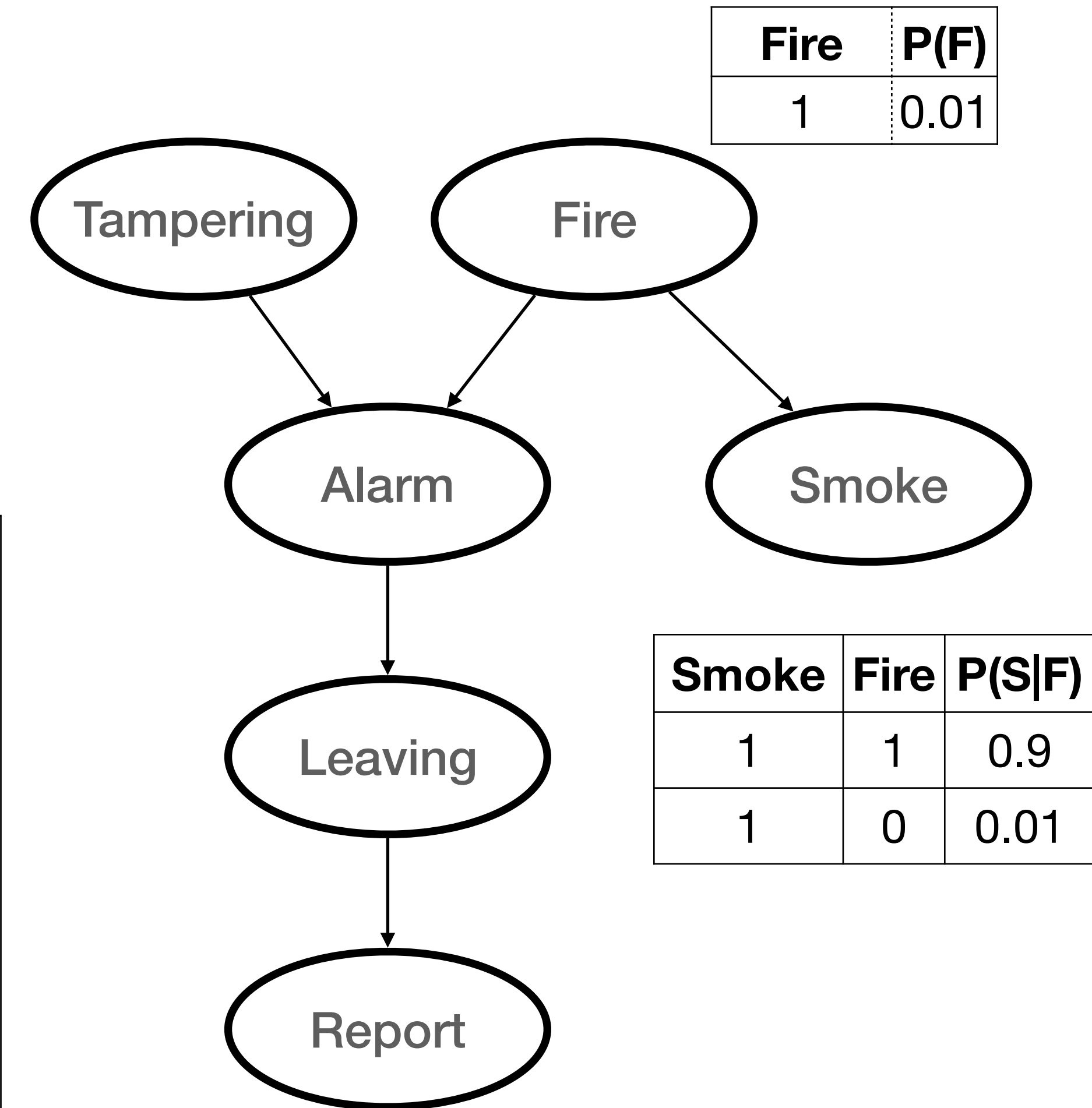
A belief network represents a joint distribution that can be factored as

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

Theorem:

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

- Node u is a **parent** of v if a directed edge $u \rightarrow v$ exists
- Node v is a **descendant** of u if there exists a **directed path** from u to v
- Node v is a **non-descendant** of u if there **does not exist** a directed path from u to v



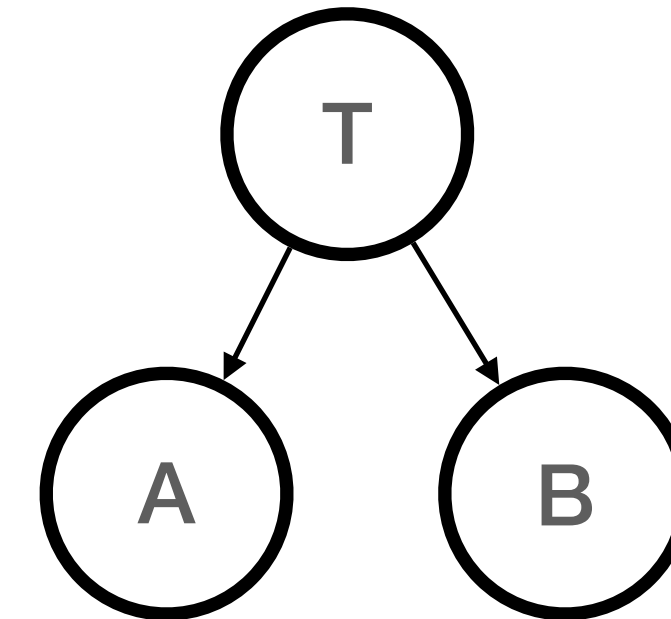
Belief net: Report has parent Leaving, Leaving has Parent Alarm, Alarm has parents Tampering and Fire, Smoke has parent Fire

Querying Independence in a Belief Network

Belief Network Independence:

Every node is **independent** of its **non-descendants**, **conditional only** on its **parents**

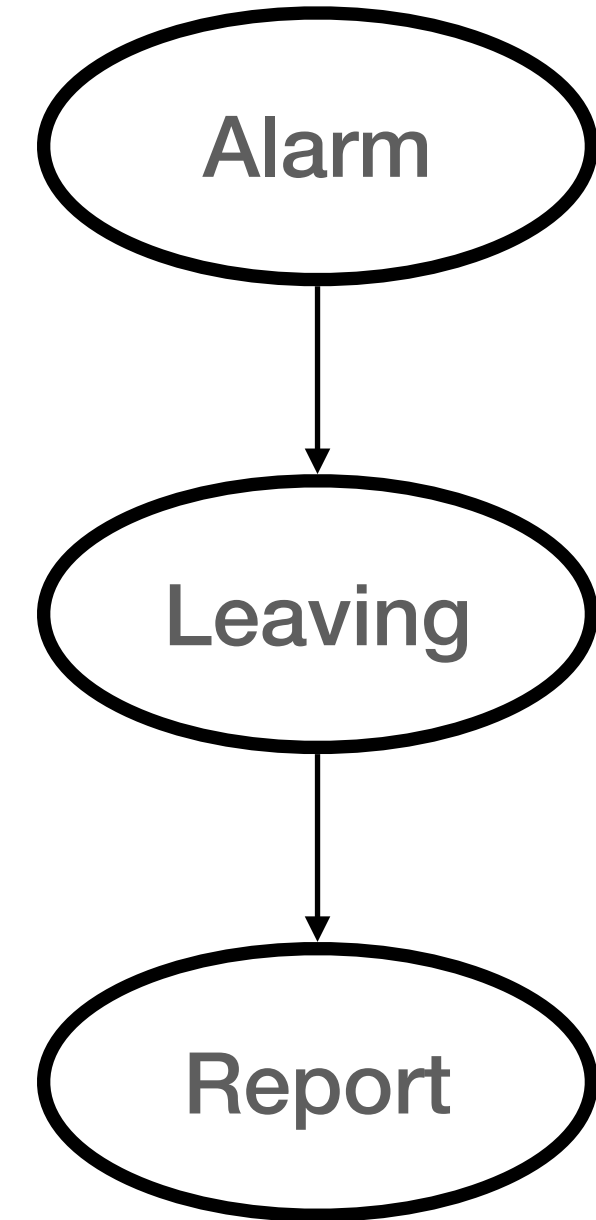
- We can use a correct belief network to efficiently answer questions about independence guarantees without knowing any numbers
- Examples using the belief network at right:
 1. Is **T** independent of **A**?
 2. Is **A** independent of **B** given **T**?
 3. Is **A** independent of **B**?



Belief net: A has parent T, B has parent T

Chain

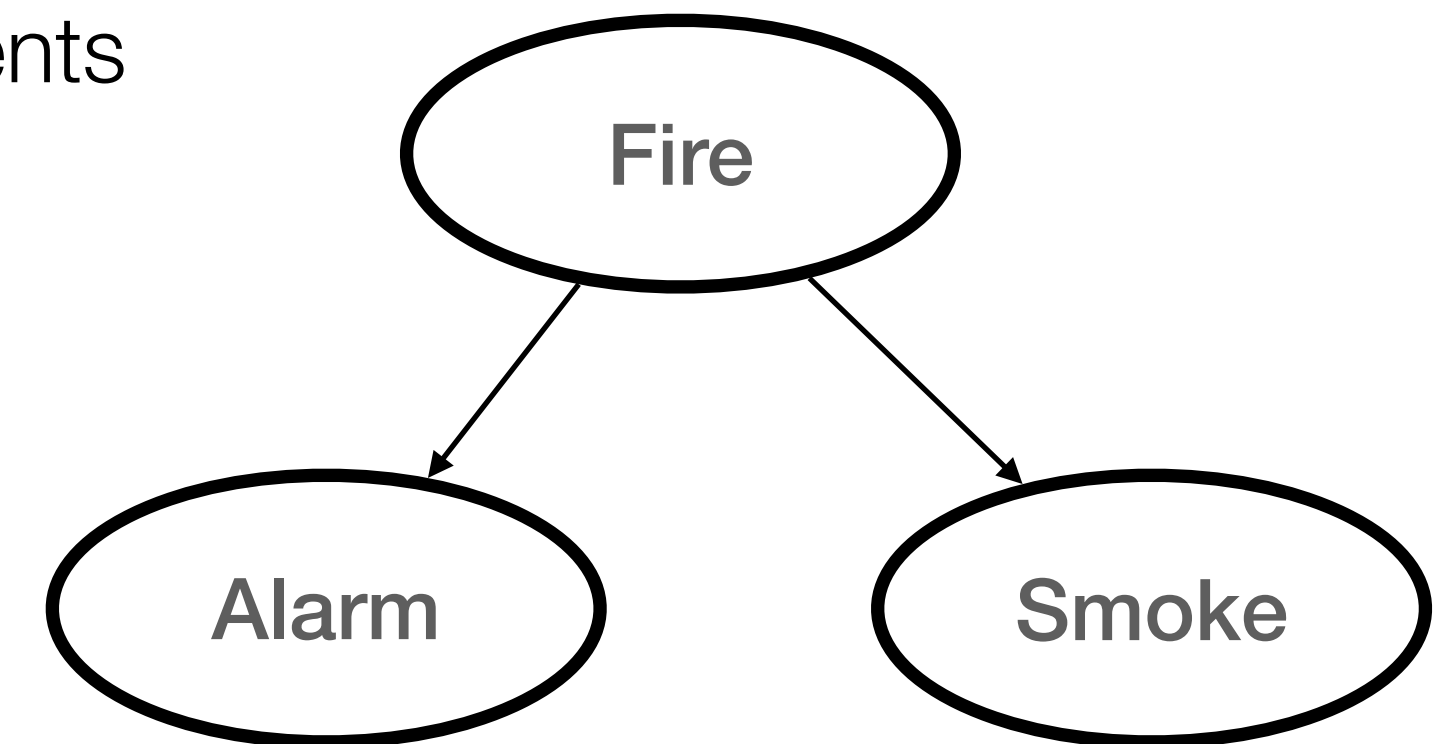
- **Question:** Is **Report** independent of **Alarm** given **Leaving**?
 - *Intuitively:* The only way learning **Report** tells us about **Alarm** is because it tells us about **Leaving**; but **Leaving** has already been observed
 - *Formally:* **Report** is independent of its non-descendants given 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



Belief net: Report has parent Leaving, Leaving has Parent Alarm

Common Ancestor

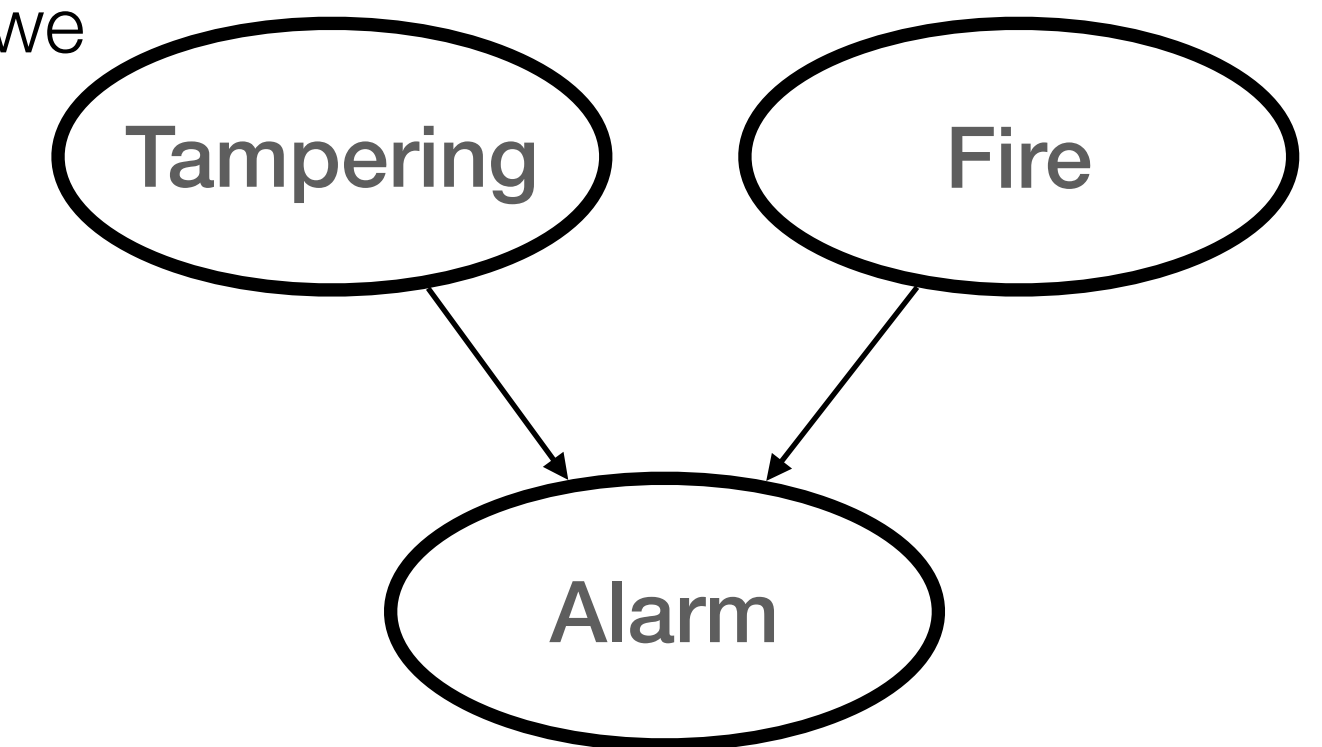
- **Question:** Is **Alarm** independent of **Smoke** given **Fire**?
 - *Intuitively:* The only way learning **Smoke** tells us about **Alarm** is because it tells us about **Fire**; but **Fire** has already been observed
 - *Formally:* **Alarm** is independent of its non-descendants given only its 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**



Belief net: Alarm has parent Fire, Smoke has parent Fire

Common Descendant ("collider")

- **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**

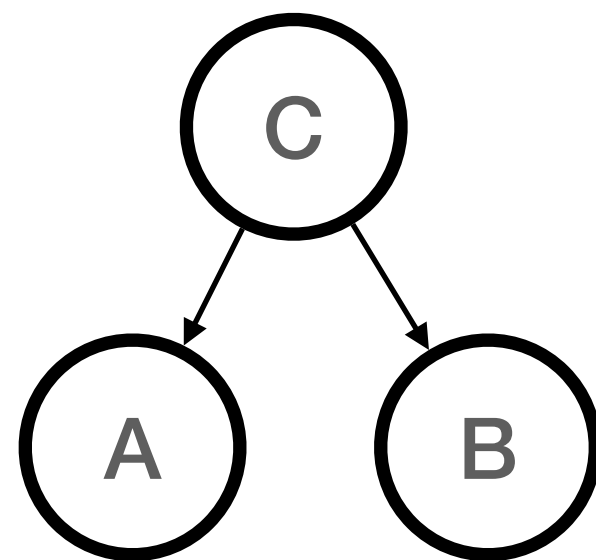


Belief net: Alarm has parents Tampering and Fire

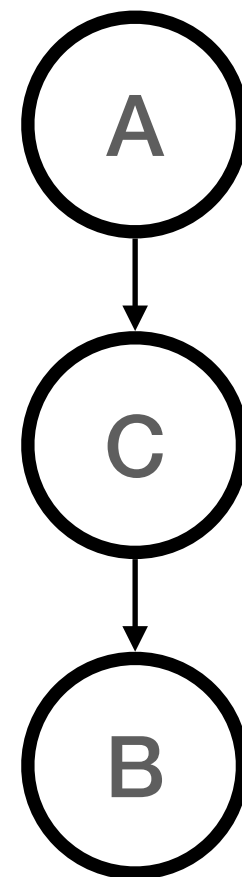
Correctness of a Belief Network

A belief network is a **correct** representation of a joint distribution when the factoring that it represents is a correct factoring of the joint distribution.

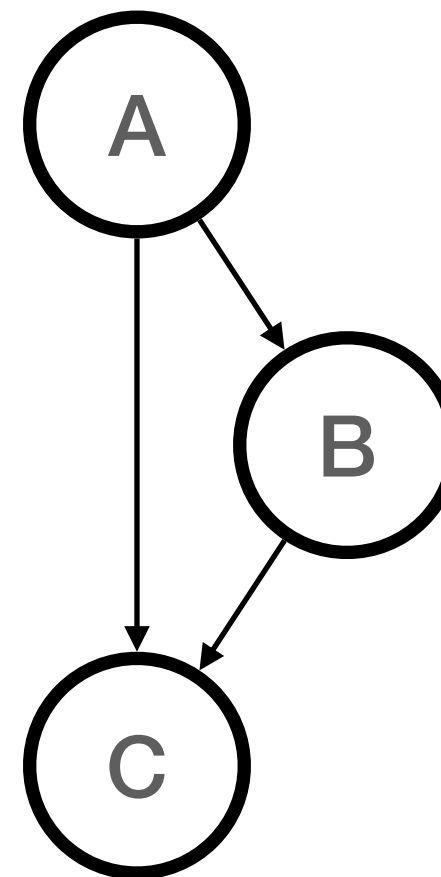
Equivalently: when the belief network answers "yes" to an independence question **only if** the **joint distribution** answers "yes" to the same question.



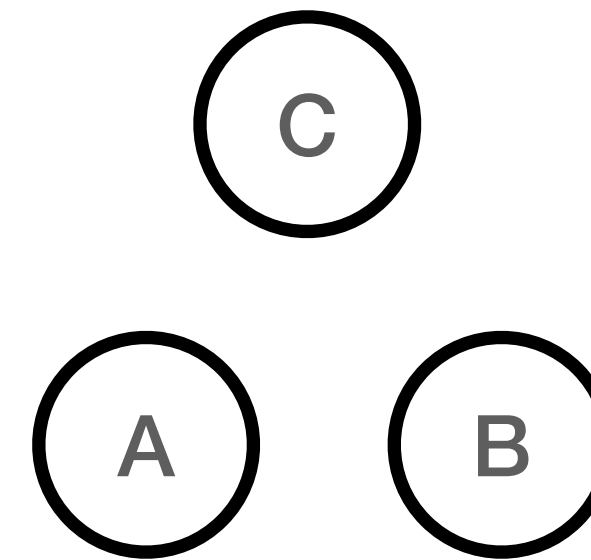
Belief net 1: A has parent C, B has parent C



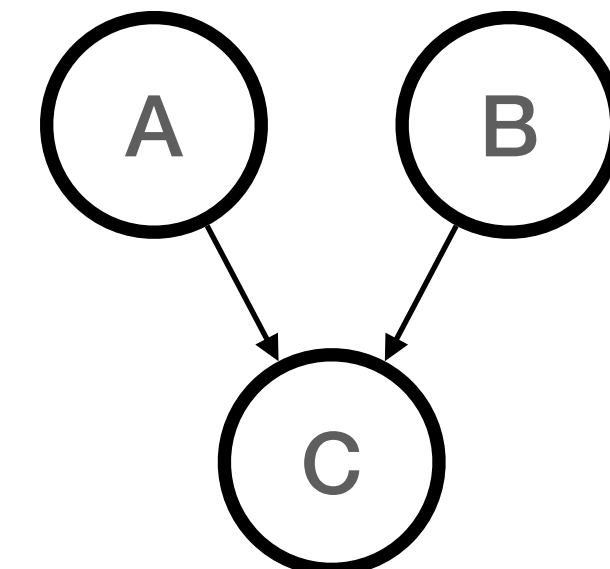
Belief net 2: B has parent C, C has parent A



Belief net 3: C has parents A and B, B has parent A



Belief net 4: A, B, and C have no parents



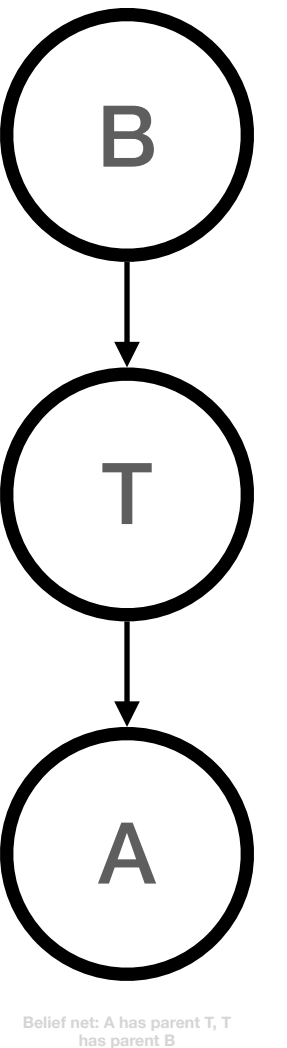
Belief net 5: C has parents A and B

Questions:

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

Causal Network?

- The arcs in belief networks **do not**, in general, represent **causal** relationships!
 - $T \rightarrow A$ is a **causal** relationship if T **causes** the value of A
 - E.g., B doesn't cause T , but this is nevertheless a correct encoding of the joint distribution
- However, reasoning about causal relationships is often a good way to **construct** a **natural** encoding as a belief network
 - We can often reason about causal independence even when we don't know the full joint distribution



Summary

- A belief network represents a specific **factoring** of a joint distribution
 - **Graph structure** encodes conditional independence relationships
 - 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**
- Arcs in a belief network **often** represent **causal** relationships
 - But they don't have to!