Belief Networks

CMPUT 366: Intelligent Systems

P&M §8.3

Lecture Outline

1. Recap & Logistics
2. Belief Networks
3. Queries
4. Constructing Belief Networks

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 $x \in dom(X)$ and $y \in dom(Y)$.

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

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

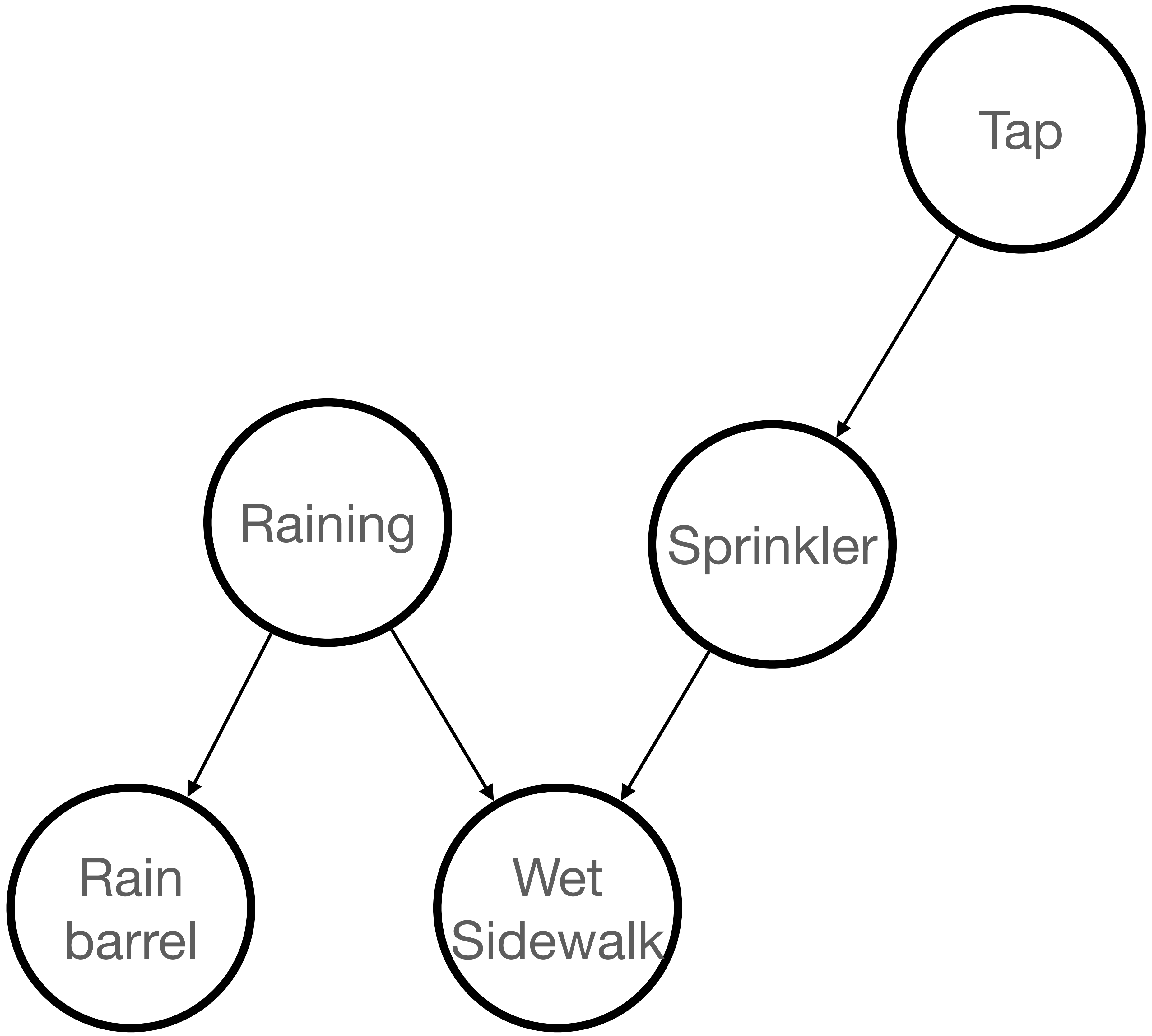
for all values of $x \in dom(X)$, $y \in dom(Y)$, and $z \in dom(Z)$.

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**

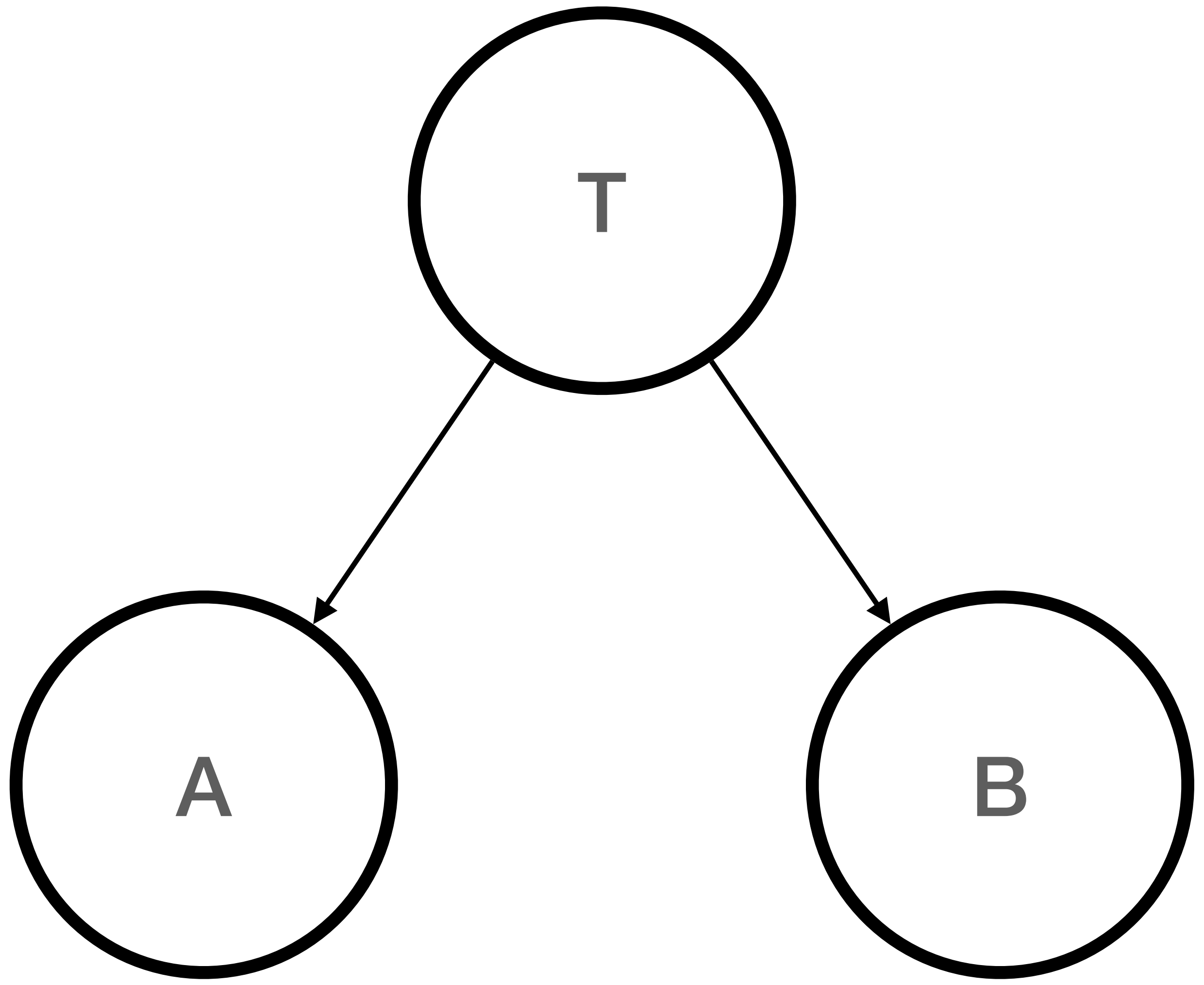
Belief Networks, informally

- We can represent the pattern of **dependence** in a distribution as a **directed acyclic graph**
- **Nodes** are random **variables**
- Arc to each node from the variables on which it **depends**



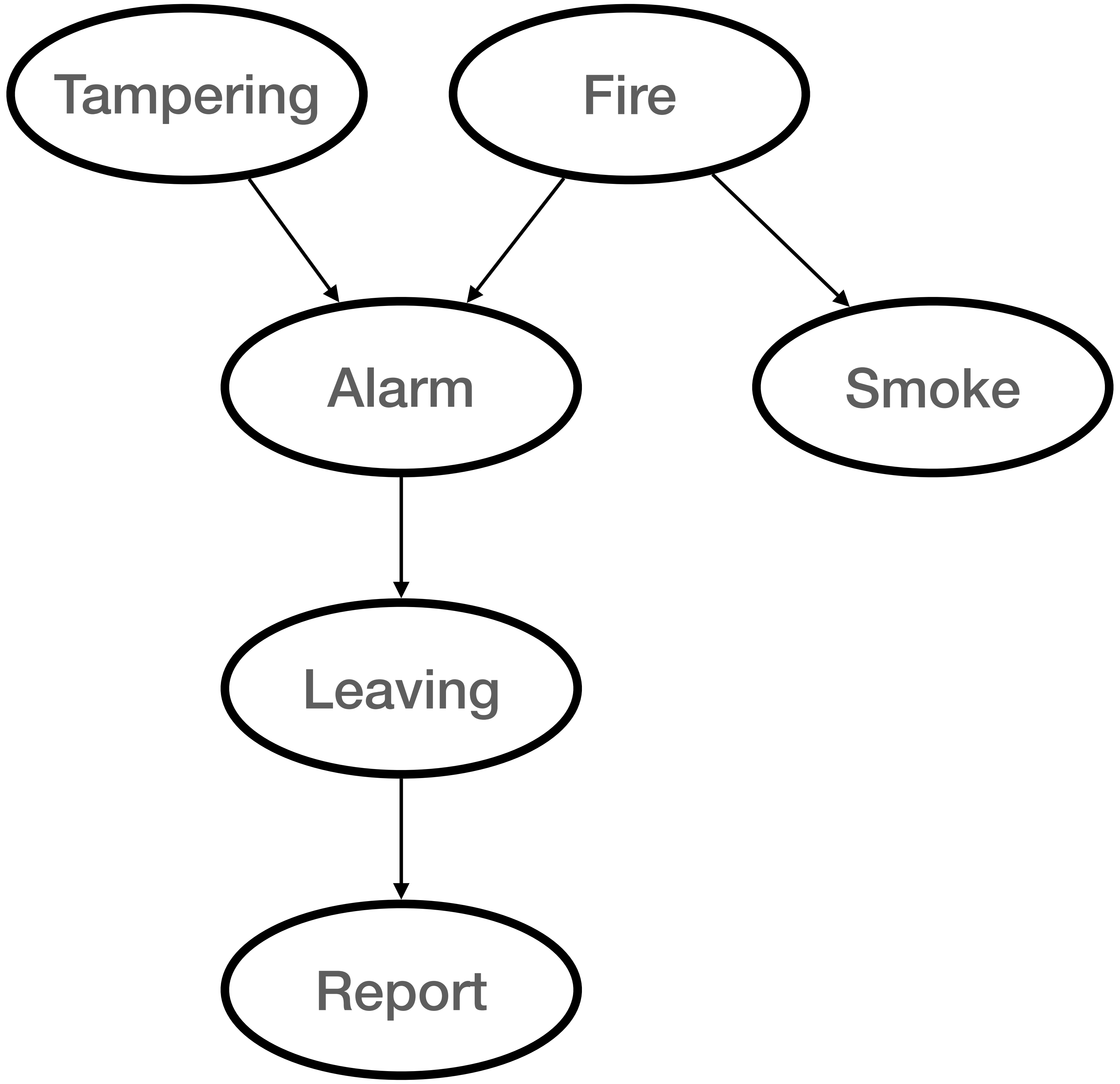
Clock Scenario

- **Alice's** observation depends on the **actual time**
- So does **Bob's**
- Neither depends on **each other's** observation



Fire Alarm Scenario

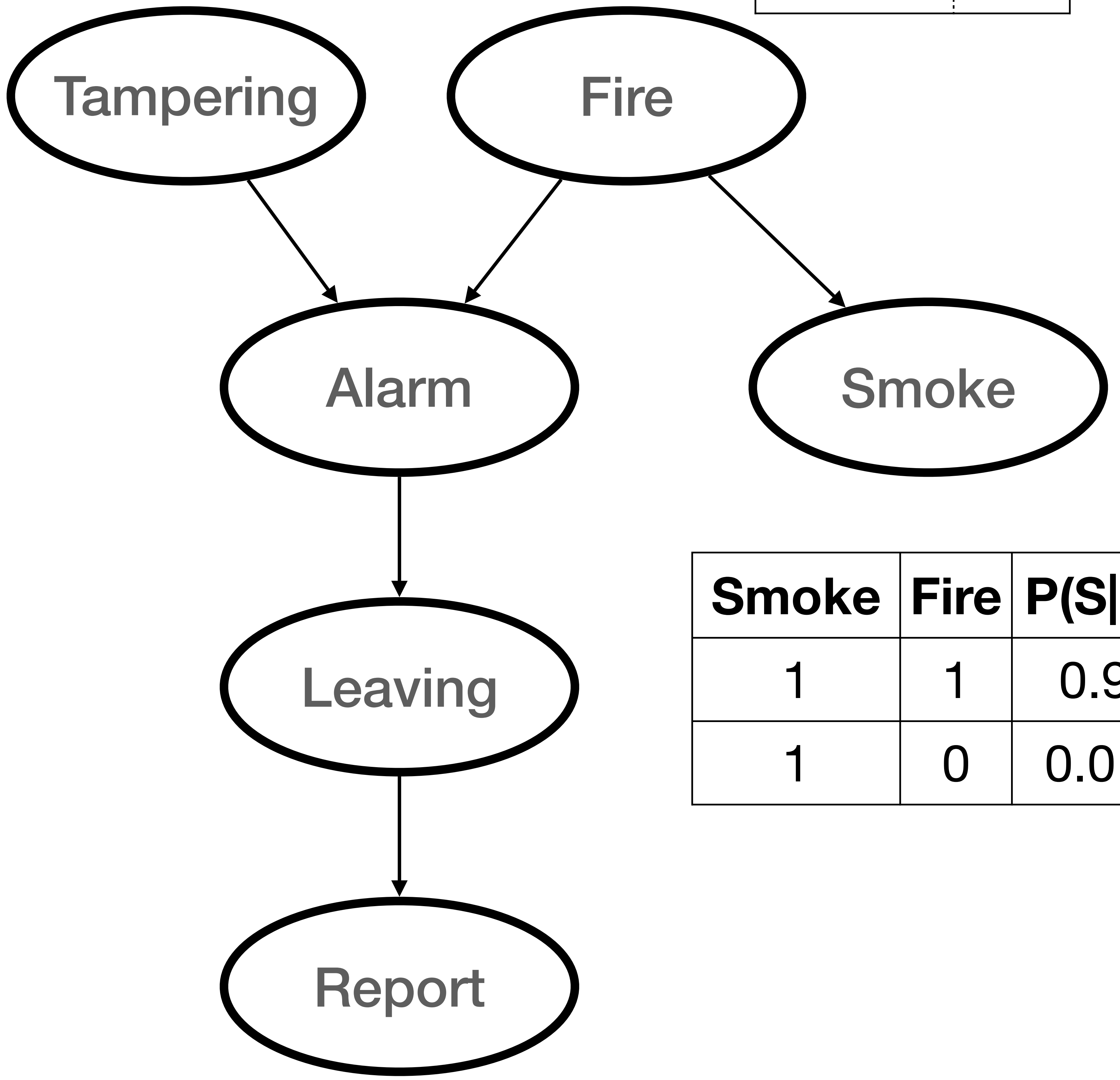
- Agent wants to deduce whether there is a **fire** in the building next door
- The fire **alarm** detects heat from fires
  - But it can also be set off by **tampering**
- A fire causes visible **smoke**
- People usually **leave** the building as a group when the fire alarm goes off
- When lots of people leave the building, our friend will **tell** us (**report** to us)



Conditional Probabilities

- 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
- **Theorem:**
  Every node is **independent** of its **non-descendants**, **conditional** 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

Fire	P(F)
1	0.01



Smoke	Fire	P(S|F)
1	1	0.9
1	0	0.01

Belief Networks

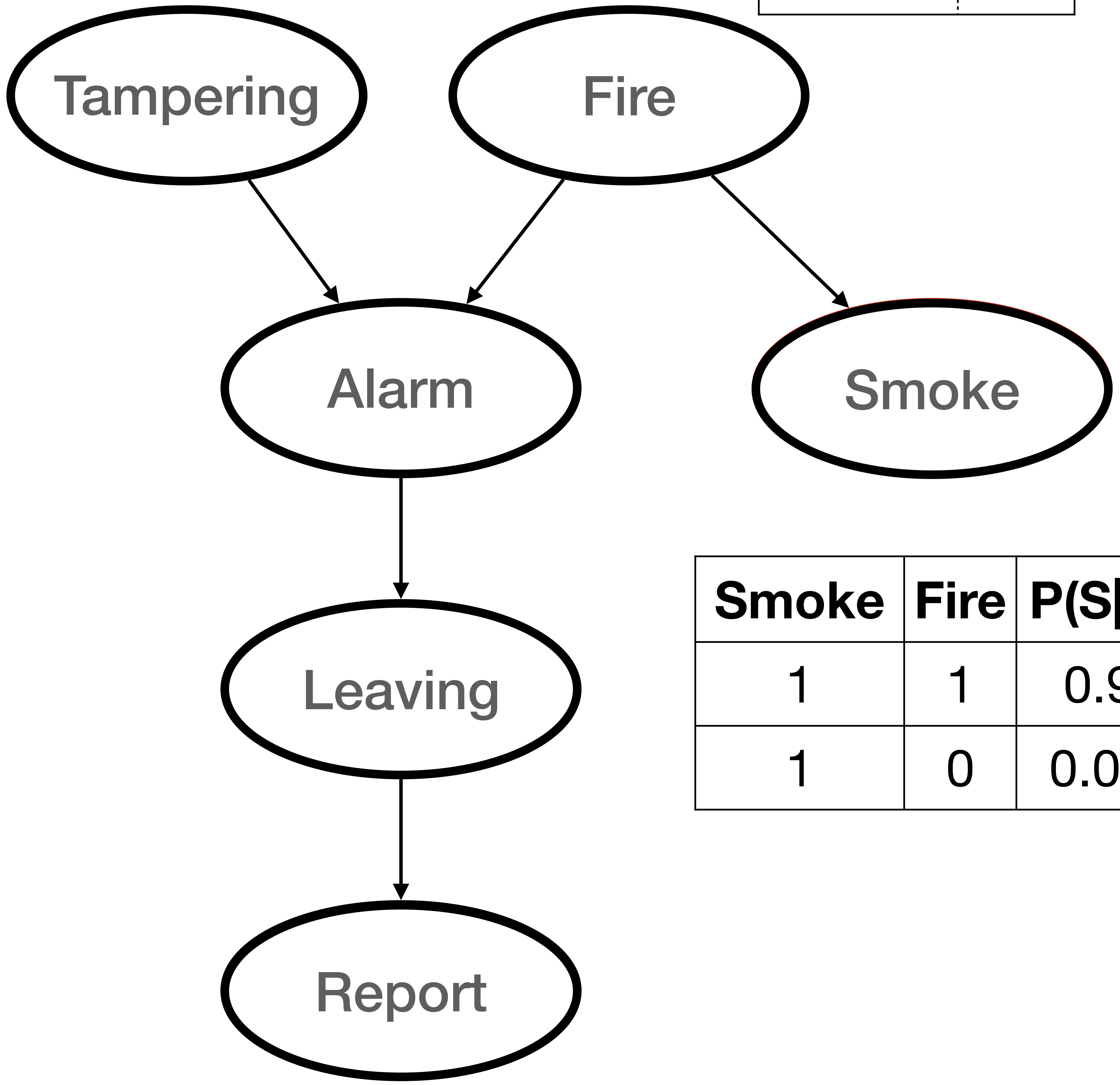
Definition:
A **belief network** (or **Bayesian network**) consists of:

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**

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

Fire	P(F)
1	0.01



Smoke	Fire	P(S|F)
1	1	0.9
1	0	0.01

Extracting Joint Probabilities: Variable Ordering

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

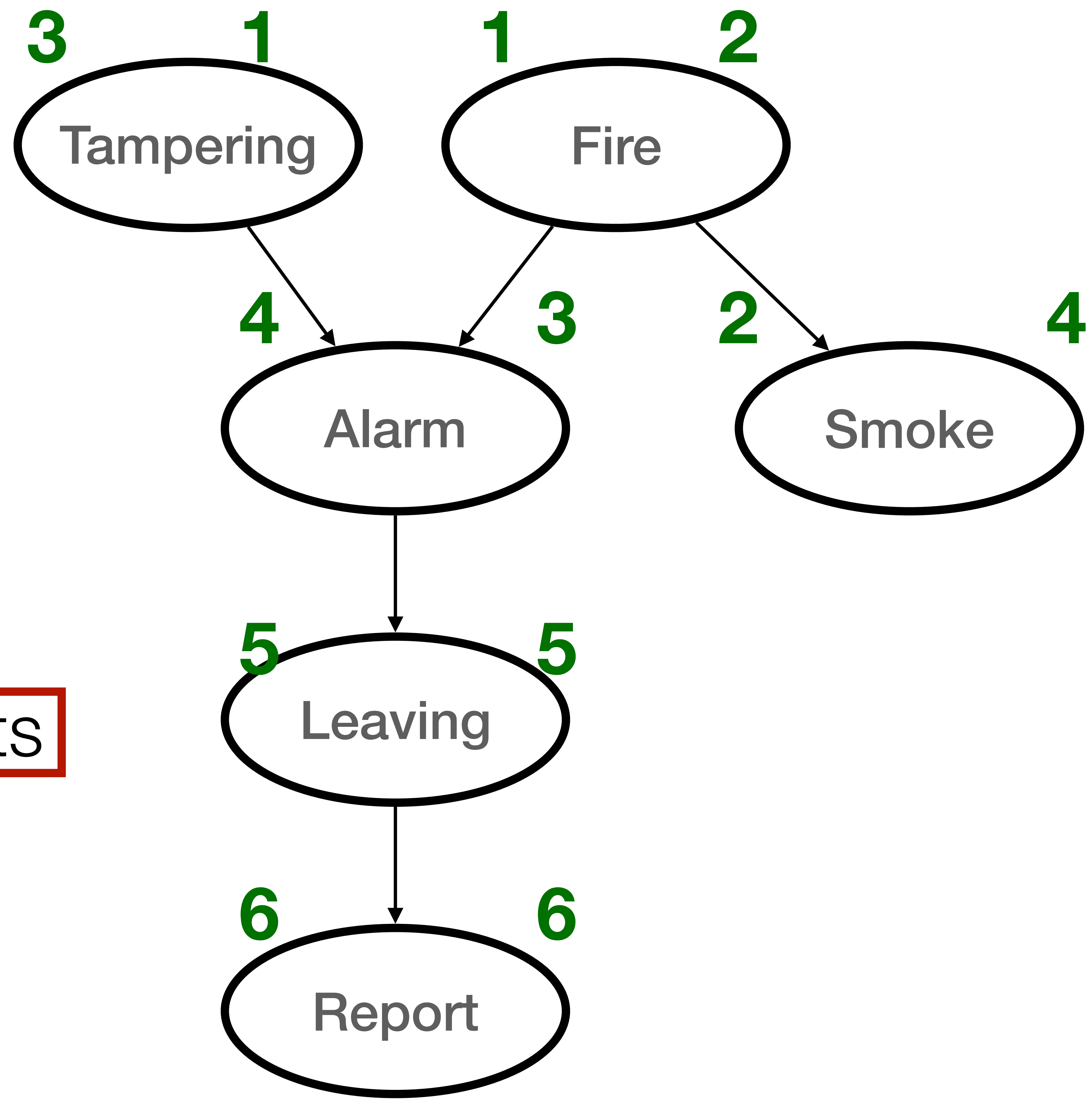
for i **from** 1 **to** n:

  select an unlabelled variable with no unlabelled parents

  label it as i

Question:

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

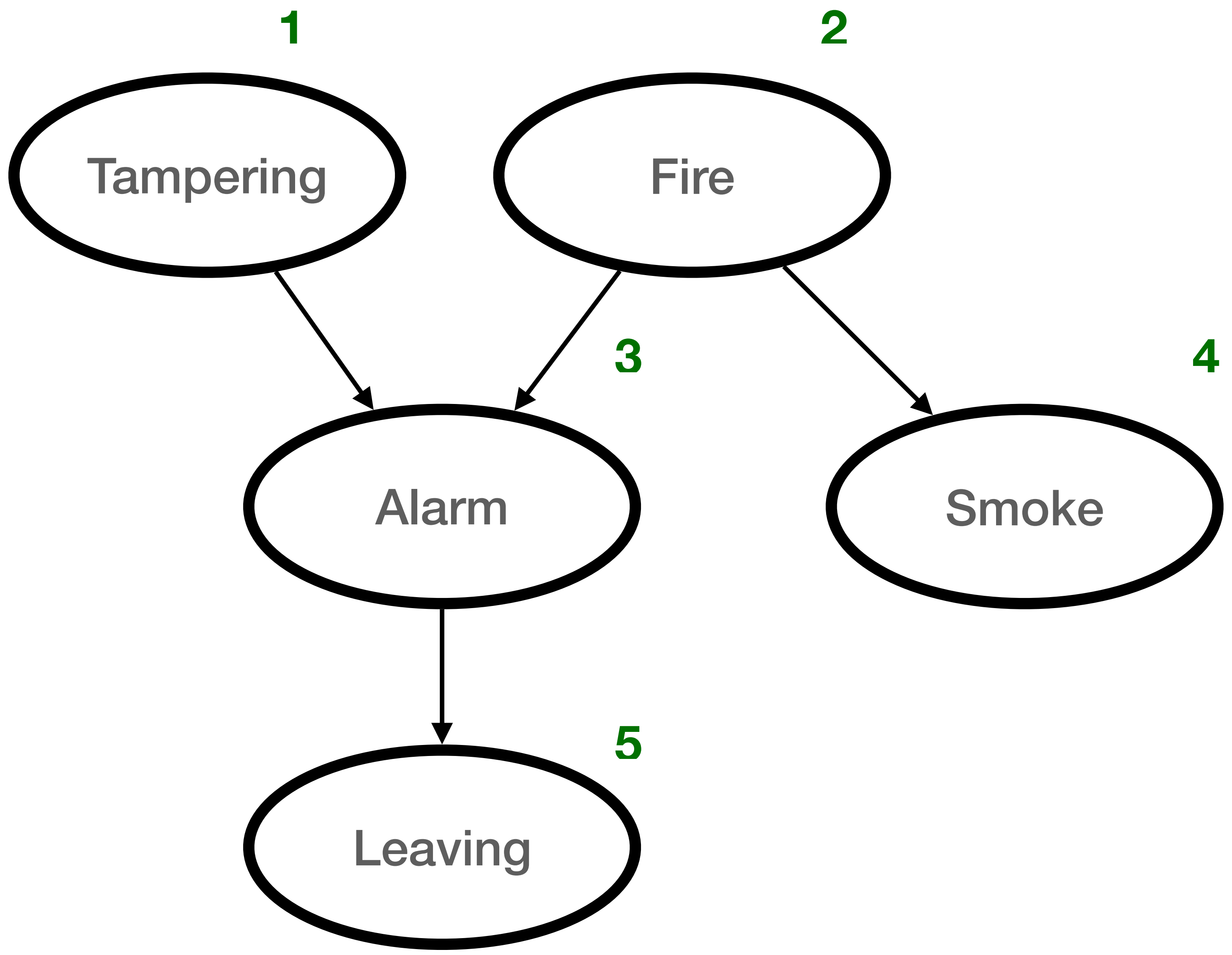


Extracting Joint Probabilities

- Multiply joint distributions in **variable order**
- **Example:** Given variable ordering
  Tampering, Fire, Alarm, Smoke, Leaving

Questions:

1. Why $P(Fire)$ instead of $P(Fire \mid \cancel{Tampering})$?
2. Why $P(Smoke \mid Fire)$ instead of $P(Smoke \mid \cancel{Tampering}, Fire, \cancel{Alarm})$?



$$P(Tampering) = P(Tampering)$$

$$P(Tampering, Fire) = P(Fire)P(Tampering)$$

$$P(Tampering, Fire, Alarm) =$$
$$P(Alarm \mid Tampering, Fire)P(Fire)P(Tampering)$$

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

$$P(Tampering, Fire, Alarm, Smoke, Leaving) =$$
$$P(Leaving \mid Alarm)Pr(Smoke \mid Fire)P(Alarm \mid Tampering, Fire)P(Fire)P(Tampering)$$

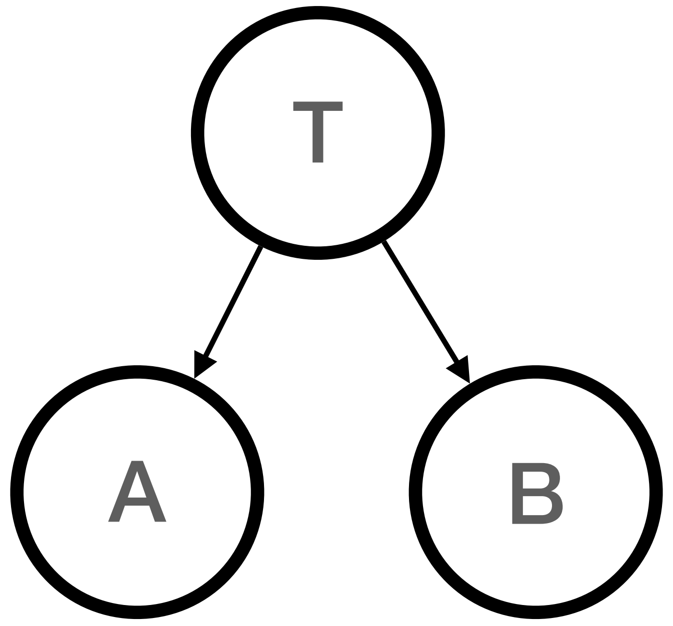
Questions:

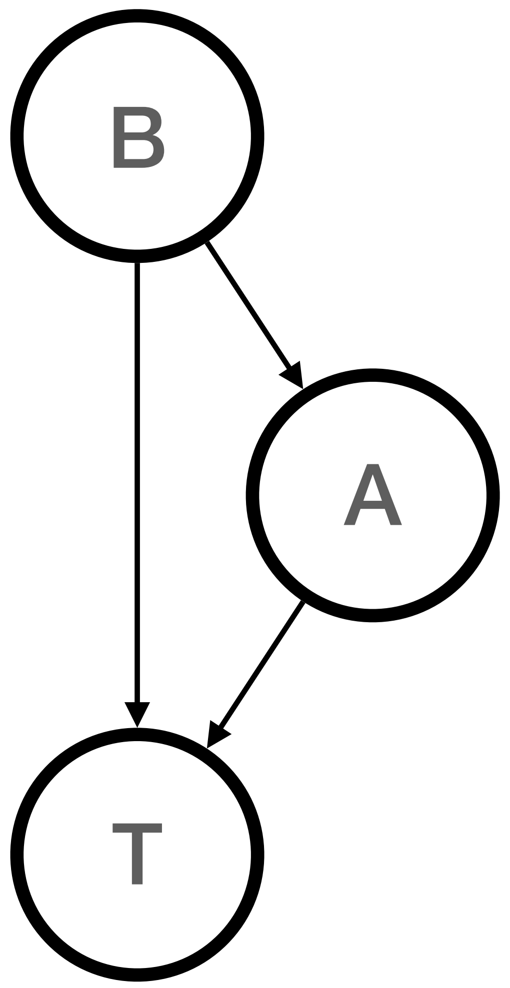
1. Which of the graphs at the right is a **correct** encoding of the Clock scenario? **Why**?

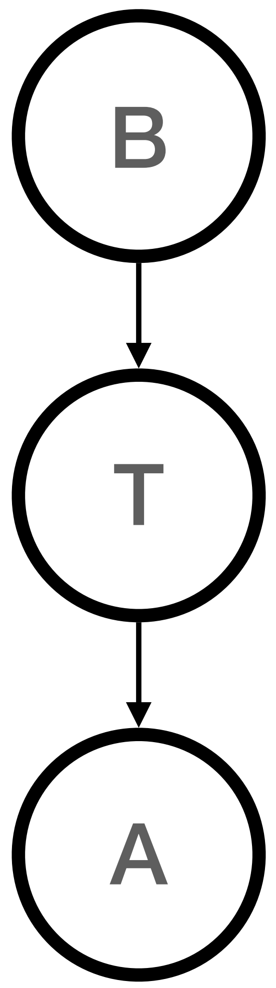
2. Which of the graphs at the right is a **good** encoding? **Why**?

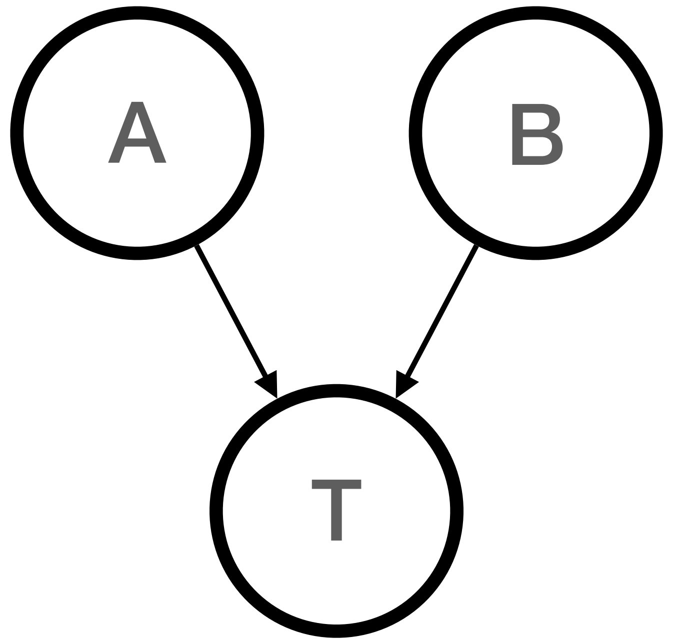
Constructing Belief Networks

- A belief network is **correct** if it encodes true conditional independence relationships: All nodes are independent of their non-descendants given their parents
- A joint distribution can, in general, have **many** correct encodings as belief networks
- Some encodings are **better** than others:
    - They represent **natural** relationships
    - They are more **compact** (they require fewer probabilities)









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:

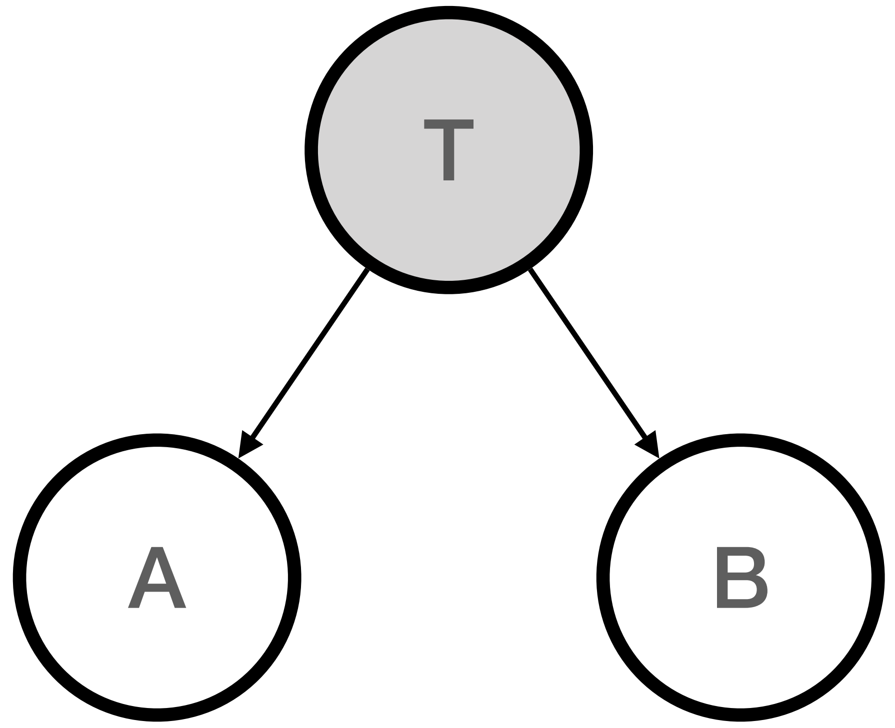
   (i) Choose a **minimal** set of variables $parents(X_i)$ from $X_1, \ldots, X_{i-1}$ such that $P(X_i \mid parents(X_i)) = P(X_i \mid X_1, \ldots, X_{i-1})$

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

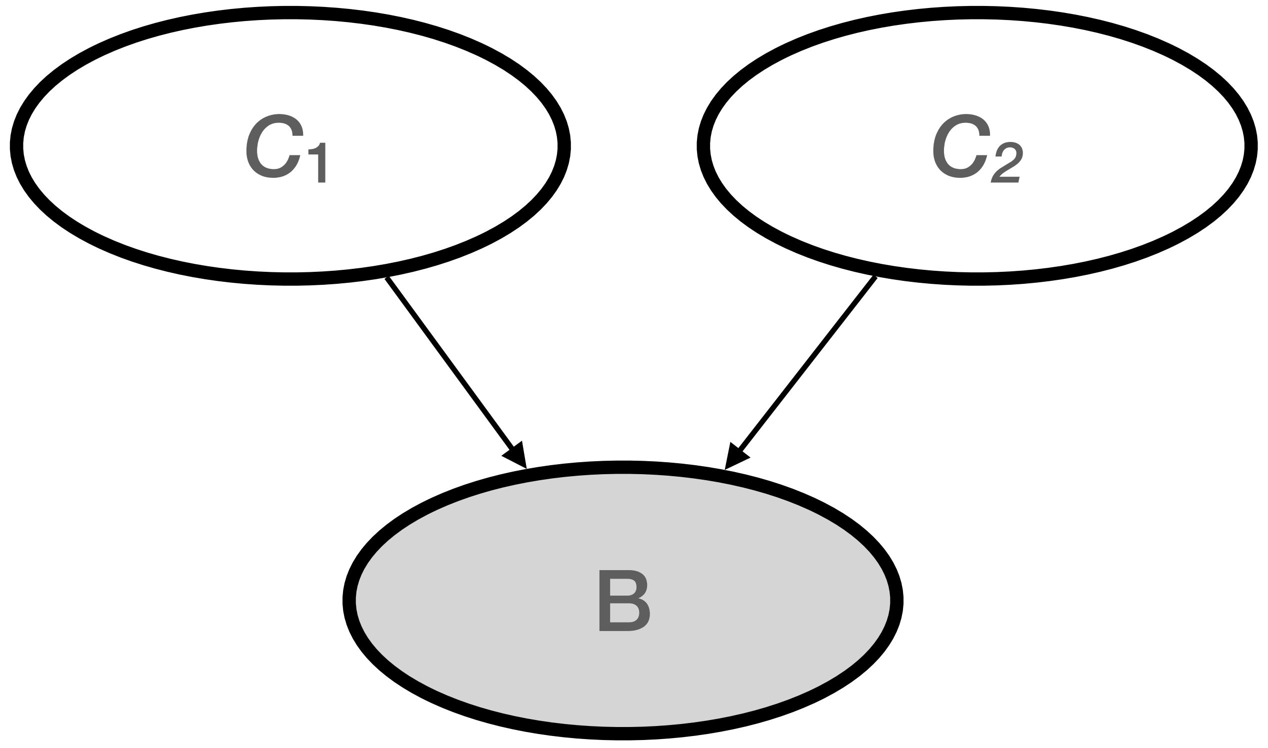
   (iii) Add an **arc** from each variable in $parents(X_i)$ to X_i

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

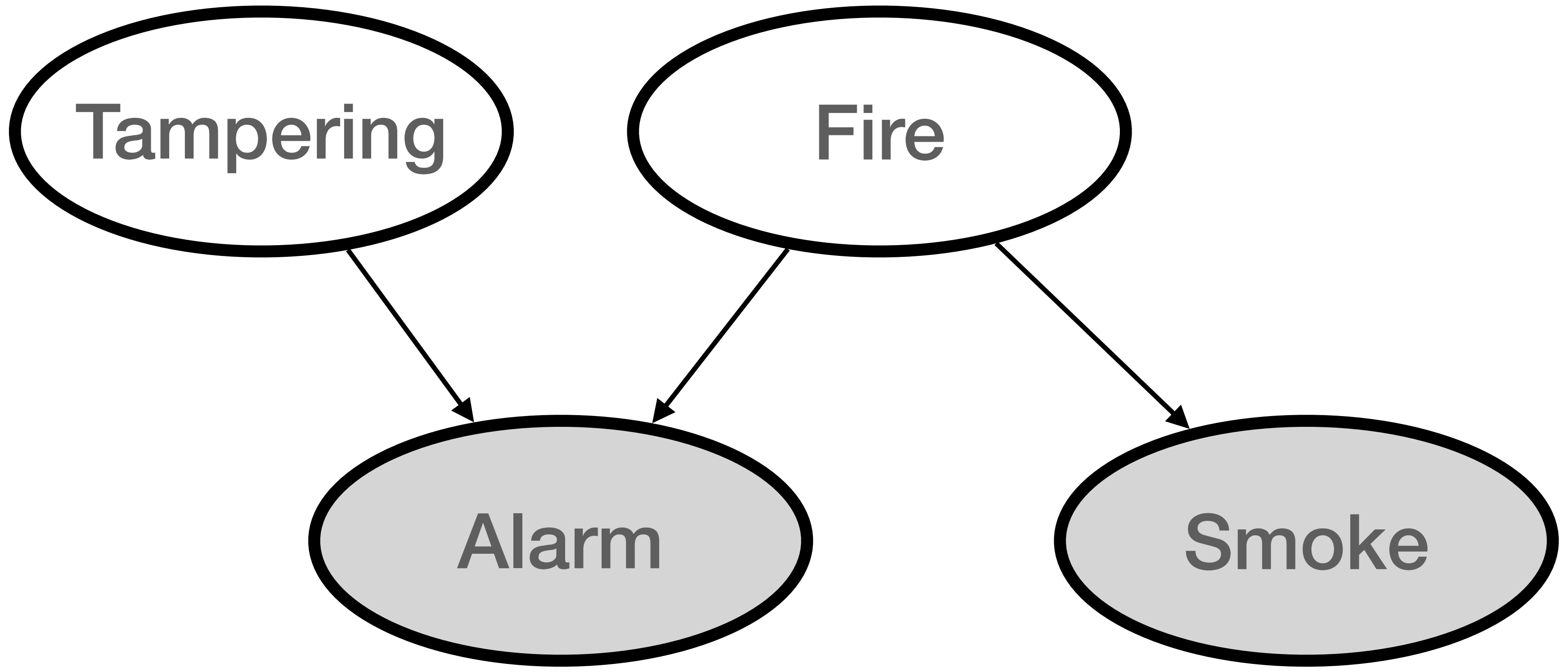
Observing Children



- Observing a **parent** renders conditionally dependent nodes conditionally independent



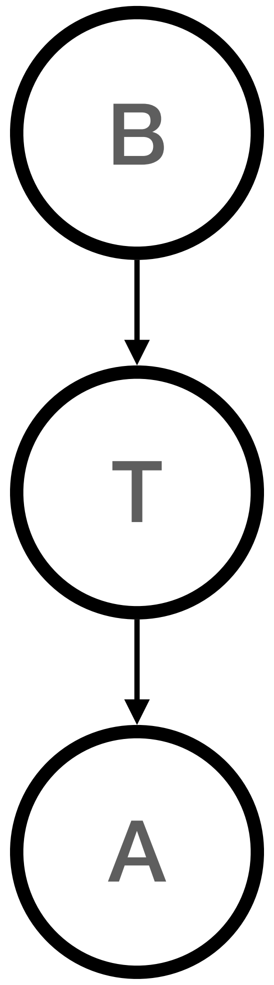
- Observing children can render conditionally independent nodes conditionally dependent
  - Extreme example: The Coins scenario: **B** = $C_1 \wedge C_2$
  - Observing both **B** and C_1 uniquely determines C_2



- Similar effect called **explaining away:**
  - We start with **prior** probabilities of Tampering and Fire
  - **Question:** If we observe that **Alarm** is ringing, how are these **posterior** probabilities **different**?
  - **Question:** If we then observe **Smoke**, how do these **posterior** probabilities **change**?

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

- Belief networks represent a **factoring** of a joint distribution
  - **Graph structure** encodes conditional independence relationships
  - Can query **posterior probabilities** of subsets of variables given **observations**
- Each joint distribution has **multiple correct representations** as a belief network
  - Some are more **compact** than others
  - Some are more **natural** than others
- Arcs in a belief network **often** represent **causal** relationships
  - But they don't have to!