# CS 581 – Advanced Artificial Intelligence

## Topic: Bayesian Networks

**Mustafa Bilgic**

http://www.cs.iit.edu/~mbilgic

https://twitter.com/bilgicm

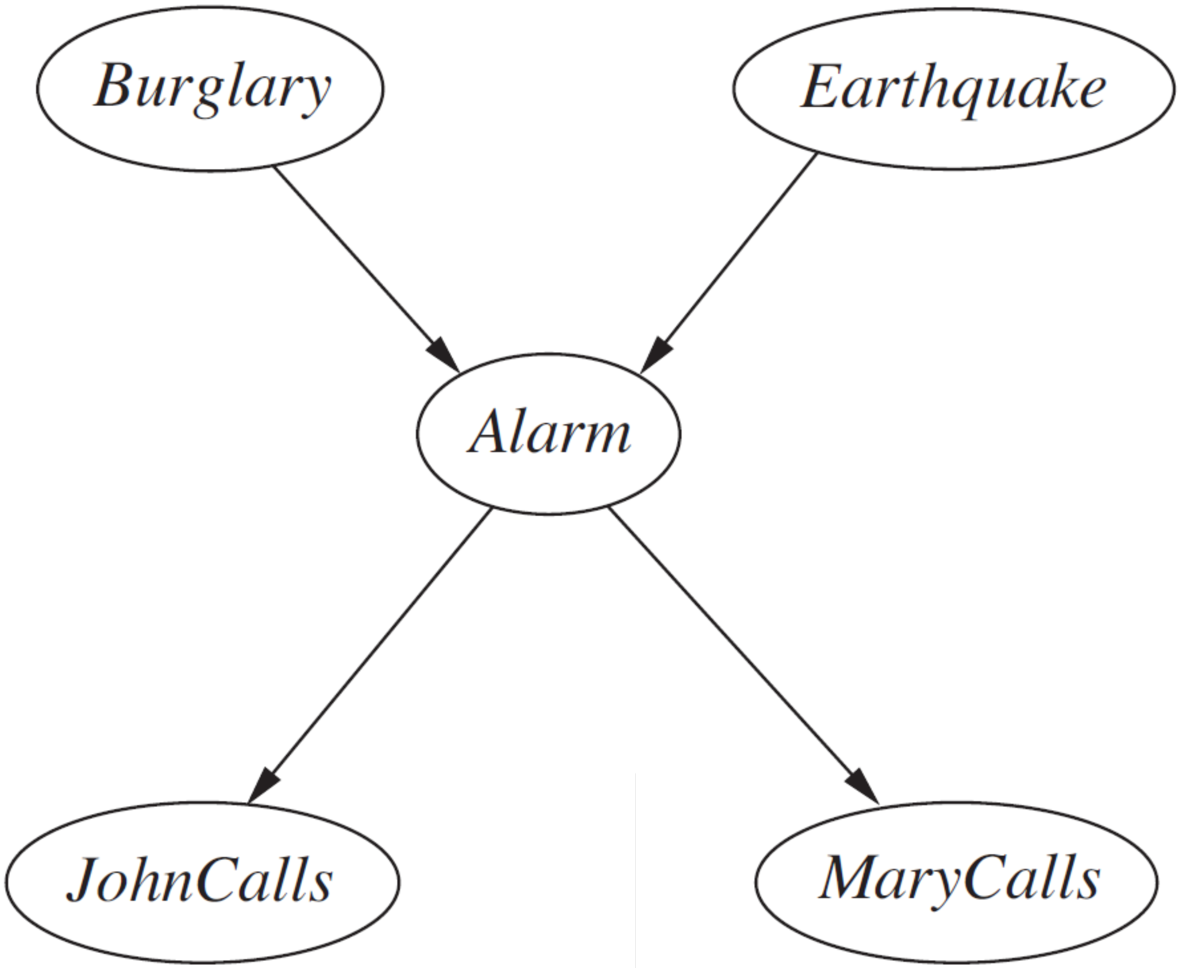
# Motivation

- Efficient, intuitive, and modular representation of probability distributions
  - Represent joint and conditional distributions
- Structured and efficient inference
  - Answer probability and MAP queries
- BN structure represents correlation but can be used to answer causality questions under certain conditions

# An Example

- Five binary variables
  - Earthquake, Burglary, Alarm, MaryCalls, JohnCalls
- Assume the following
  - E and B are uncorrelated
  - E and M are related only through A; similarly, E, J, and A
  - B and M are related only through A; similarly, B, J, and A
  - M and J are directly related through A; undirectly related through E and B; otherwise, M and J are unrelated
- One approach
  - Represent and estimate the full join $P(E, B, A, M, J)$
    - How many parameters?
    - What can you tell about the relationships between the variables?
- Alternative approach
  - Bayesian network (next slide)

# Burglary Example



# Possible Queries

- $P(B \mid J = true)$
- $P(B \mid M = true, J = true)$
- $P(M \mid B = true)$
- $P(M \mid B = false)$
- $P(M, J \mid B = true)$
- $P(M \mid J = true)$
- …

# We'll Cover

- Bayesian networks (in detail)
  - https://en.wikipedia.org/wiki/Bayesian_network
- Hidden Markov Models (in detail)
  - https://en.wikipedia.org/wiki/Hidden_Markov_model
- Dynamic Bayesian networks (brief)
  - https://en.wikipedia.org/wiki/Dynamic_Bayesian_network
- Influence diagrams (in detail)
  - https://en.wikipedia.org/wiki/Influence_diagram
- Causal networks (brief)

# Bayesian Networks

- Random variables = nodes
- Direct relationships = directed edges
- BNs capture independencies
  - More compact than full joint representation
- Graphs provide
  - Graph theory / efficient reasoning
  - Intuition

# Directed Graphs

- A **graph** consists of **nodes** and **edges**
- **Nodes:** $\mathcal{X}=\{X_1, X_2, \ldots, X_n\}$
- **Undirected Edge:** $X_i - X_j$
- **Directed Edge:** $X_i \rightarrow X_j$
- A graph is **directed** if its *all* edges are directed

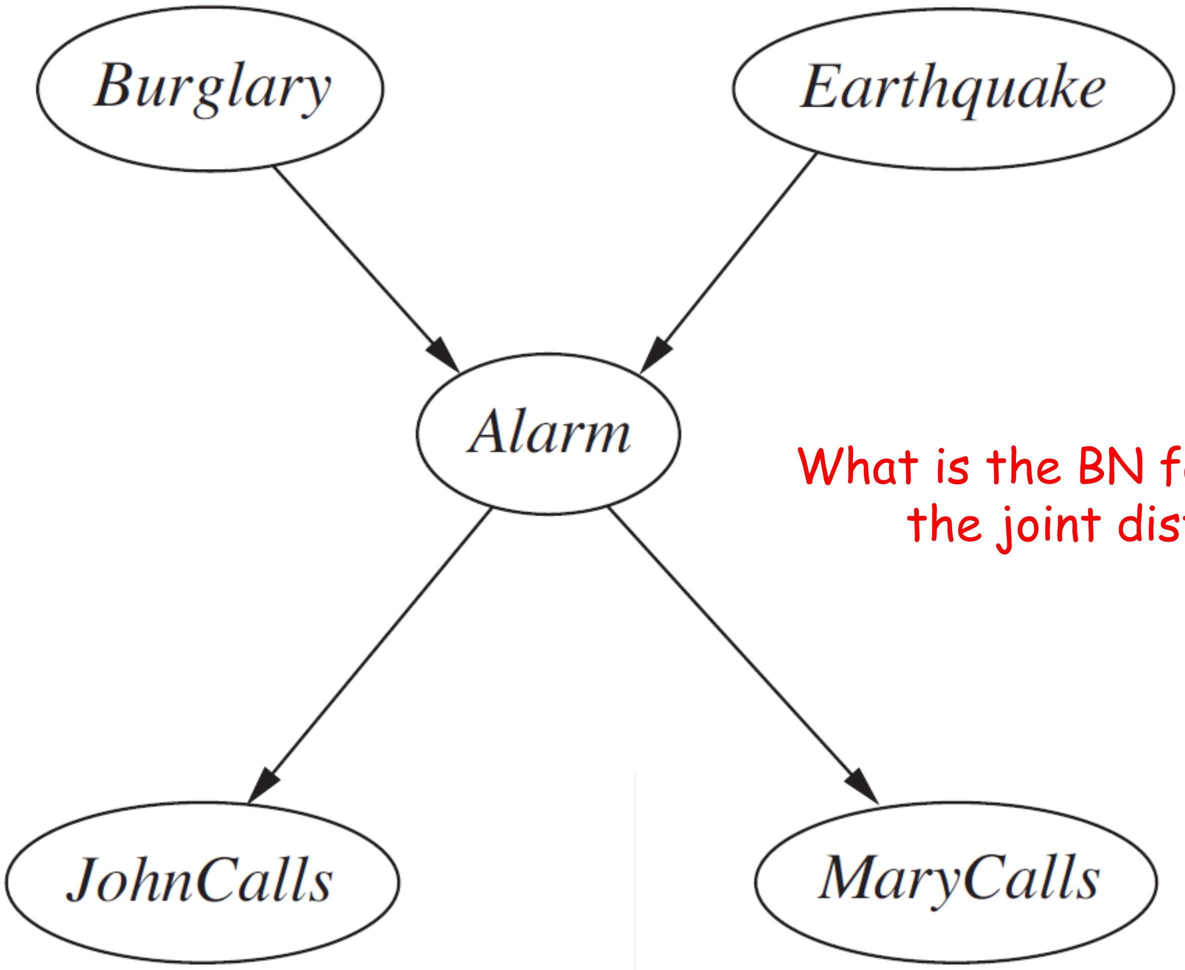
# RELATIONSHIPS

- $X_i \rightarrow X_j$
  - $X_i$ is the **parent**
  - $X_j$ is the **child**
- $X_i$ is an **ancestor** of $X_j$ if there is a directed path from $X_i$ to $X_j$
- $X_i$ is a **descendant** of $X_j$ if there is a directed path from $X_j$ to $X_i$
- **Nondescendants**$(X_i) \equiv \mathcal{X} \setminus$ Descendants$(X_i)$

# Bayesian network Factorization

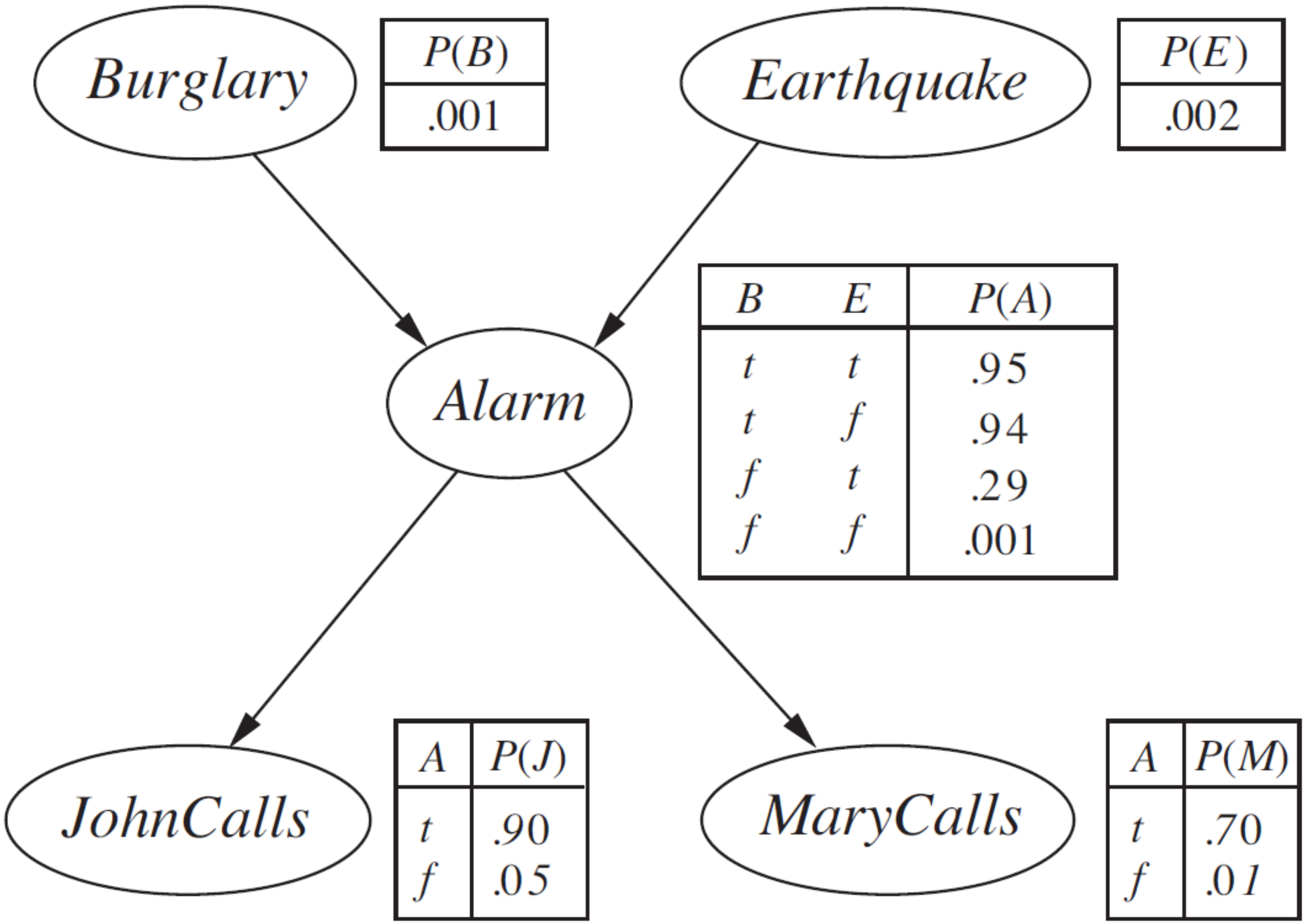
$$P\left(X_1,\ldots,X_n\right)=\prod_i P\left(X_i \mid Pa\left(X_i\right)\right)$$

# BURGLARY EXAMPLE




What is the BN factorization of the joint distribution?

# Burglary Example




| P(B) |
|---|
| .001 |

| P(E) |
|---|
| .002 |

| B | E | P(A) |
|---|---|---|
| t | t | .95 |
| t | f | .94 |
| f | t | .29 |
| f | f | .001 |

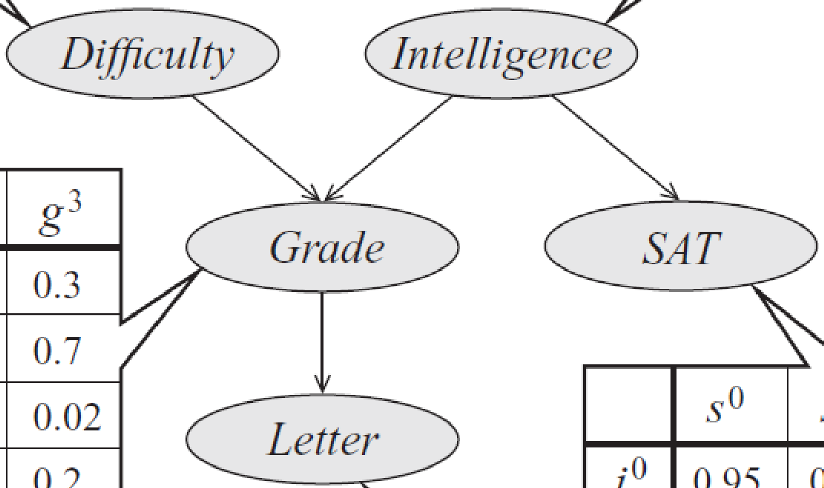
| A | P(J) |
|---|---|
| t | .90 |
| f | .05 |

| A | P(M) |
|---|---|
| t | .70 |
| f | .01 |

# Student Example

| $d^0$ | $d^1$ |
|---|---|
| 0.6 | 0.4 |

| $i^0$ | $i^1$ |
|---|---|
| 0.7 | 0.3 |




| | $g^1$ | $g^2$ | $g^3$ |
|---|---|---|---|
| $i^0, d^0$ | 0.3 | 0.4 | 0.3 |
| $i^0, d^1$ | 0.05 | 0.25 | 0.7 |
| $i^1, d^0$ | 0.9 | 0.08 | 0.02 |
| $i^1, d^1$ | 0.5 | 0.3 | 0.2 |

| | $s^0$ | $s^1$ |
|---|---|---|
| $i^0$ | 0.95 | 0.05 |
| $i^1$ | 0.2 | 0.8 |

| | $l^0$ | $l^1$ |
|---|---|---|
| $g^1$ | 0.1 | 0.9 |
| $g^2$ | 0.4 | 0.6 |
| $g^2$ | 0.99 | 0.01 |

# Independencies

- X is independent of its non-descendants given its parents
  - $X \perp \text{Non-descendants}(X) \mid \text{Parents}(X)$
- D-separation

# EXAMPLES

- X causes Y and Y causes Z; no direct relationship between X and Z
  - $X \rightarrow Y \rightarrow Z$
  - Nothing is marginally independent of each other
  - $Z \perp X \mid Y$
- Y causes both X and Z; no direct relationship between X and Z
  - $X \leftarrow Y \rightarrow Z$
  - Nothing is marginally independent of each other
  - $Z \perp X \mid Y$
- Both X and Z cause Y; no direct relationship between X and Z
  - $X \rightarrow Y \leftarrow Z$
  - X and Z are marginally independent
  - X and Z become dependent when the value of Y is known

# Independencies – D-separation

- Definition: Observed ≡ Its value is known
- Causal trail
  - $X \rightarrow Y \rightarrow Z$; E.g., Burglary → Alarm → MaryCalls
  - X and Z are independent if Y is observed
- Evidential trail
  - $X \leftarrow Y \leftarrow Z$; E.g., MaryCalls ← Alarm ← Burglary
  - X and Z are independent if Y is observed
- Common cause
  - $X \leftarrow Y \rightarrow Z$; E.g., JohnCalls ← Alarm → MaryCalls
  - X and Z are independent if Y is observed
- Common effect
  - $X \rightarrow Y \leftarrow Z$; E.g., Burglary → Alarm ← Earthquake
  - X and Z are marginally independent, but they become dependent if Y or any of Y's descendants are observed

# Reasoning Patterns

- **Causal reasoning**
  - From causes to effects
    - E.g., Burglary to Alarm to MaryCalls
    - E.g., Intelligence to Grade to Letter
- **Evidential reasoning**
  - From effects to the causes
    - E.g., JohnCalls to Alarm to Earthquake
    - E.g, Letter to Grade to Difficulty
- **Explaining away/inter-causal reasoning**
  - Causes of a common effect interact
    - E.g., Earthquake, Burglary, and Alarm (and Alarm's descendants)
    - E.g., Difficulty, Intelligence, and Grade (and Grade's descendants)

# Inference in Bayesian Networks

- There are several methods, some are exact and some are approximate
- We will study only one in this class
- *Variable Elimination*

# Variable Elimination

- Let
  - **V** be the set of all variables, **Q** be the set of query variables, **E** be the set of evidence variables
  - P(**Q**|**E**) be the query

1. Write down the joint dist. using the Bayesian network structure
2. Set the variables in **E** to their respective values
3. Sum over all variables in $\mathbf{V} \setminus (\mathbf{Q} \cup \mathbf{E})$
   a) Pick an order for variables in $\mathbf{V} \setminus (\mathbf{Q} \cup \mathbf{E})$
   b) For each variable $V_i$ in $\mathbf{V} \setminus (\mathbf{Q} \cup \mathbf{E})$, create a new factor by
      - Multiplying all the factors that contains $V_i$, and
      - Summing over possible values of $V_i$
4. Normalize the last remaining factor (this step is unnecessary if **E** is empty)

# Irrelevant

- Let
  - **V** be the set of all variables, **Q** be the set of query variables, **E** be the set of evidence variables
  - P(**Q** | **E**) be the query
- $Y \in \boldsymbol{V} \setminus \{\boldsymbol{Q} \cup \boldsymbol{E}\}$ is irrelevant iff
  - $Y \notin Ancestors\ of\{\boldsymbol{Q} \cup \boldsymbol{E}\}$
    - or
  - $Y \perp \boldsymbol{Q} \mid \boldsymbol{E}$
- Examples