

Vorlesung

Grundlagen der

Künstlichen Intelligenz

Reinhard Lafrenz / Prof. A. Knoll

Robotics and Embedded Systems
Department of Informatics – I6
Technische Universität München

www6.in.tum.de

lafrenz@in.tum.de

089-289-18136

Room 03.07.055



Wintersemester 2012/13

14.1.2013



Chapter 14

Probabilistic Reasoning

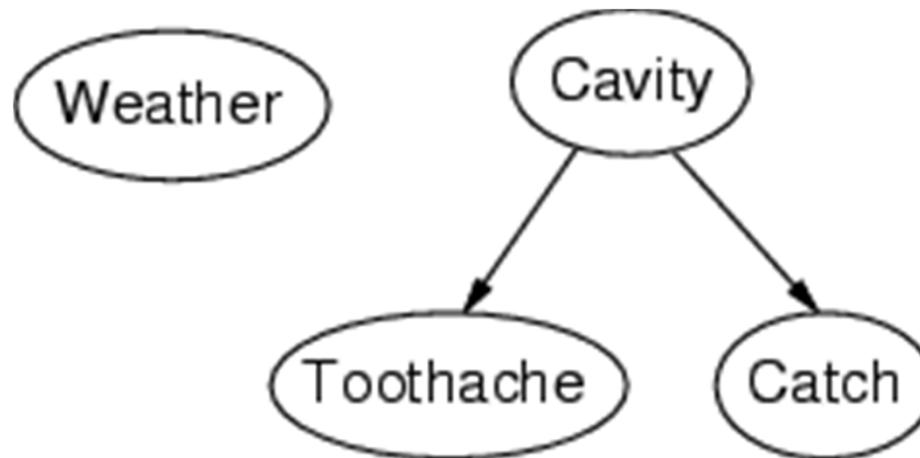
Bayesian networks

- A simple, graphical notation for conditional independence assertions and hence for compact specification of full joint distributions
- Syntax:
 - a set of nodes, one per variable
 - a directed, acyclic graph (link \approx "directly influences")
 - a conditional distribution for each node given its parents:
$$\mathbf{P}(X_i | \text{Parents}(X_i))$$
- In the simplest case, conditional distribution represented as a **conditional probability table** (CPT) giving the distribution over X_i for each combination of parent values
This quantifies the effect of the parents on the node.



Example

- Topology of network encodes conditional independence assertions:



- *Weather* is independent of the other variables
- *Toothache* and *Catch* are conditionally independent given *Cavity*

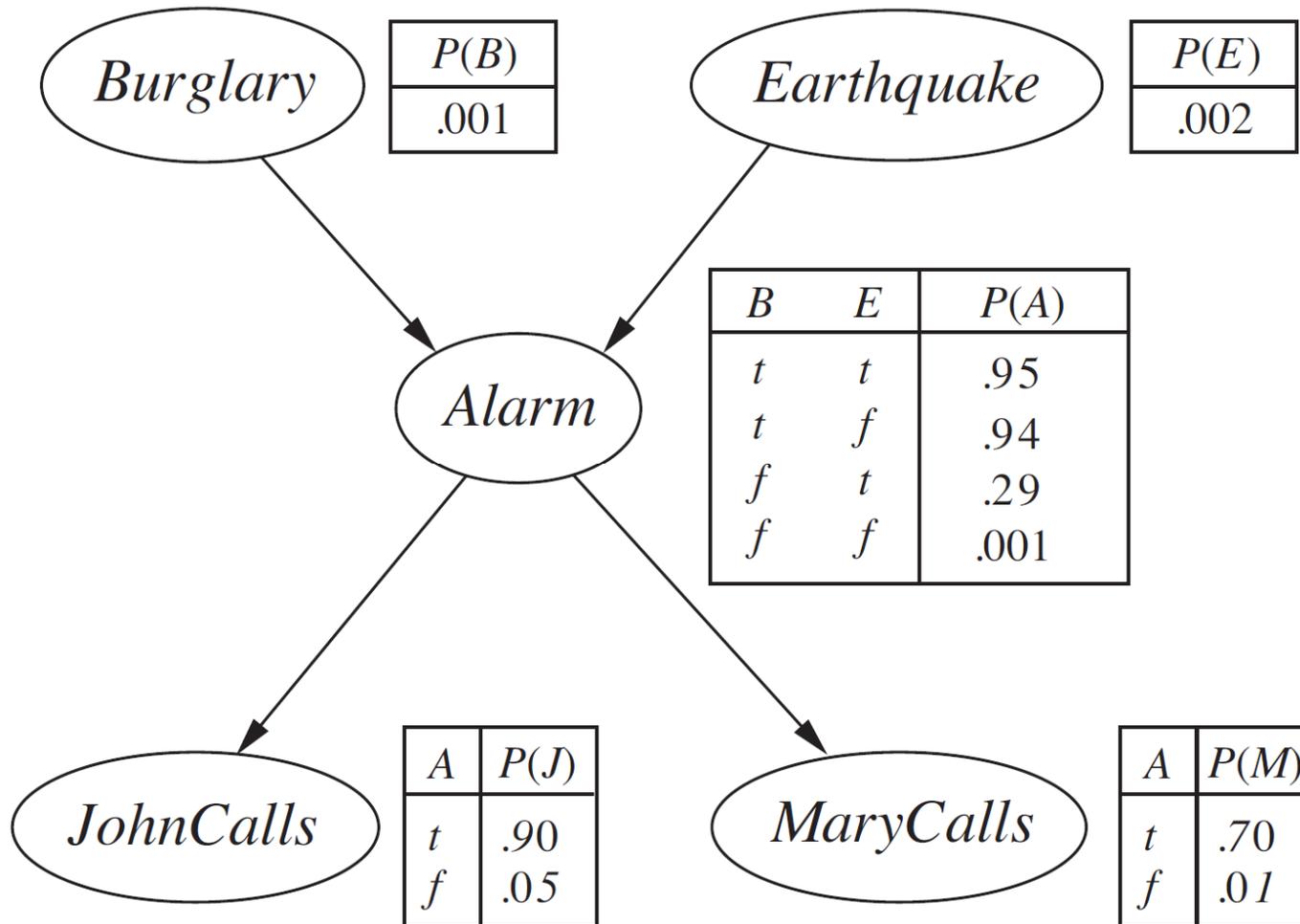


Example

- I'm at work, neighbor John calls to say my alarm is ringing, but neighbor Mary doesn't call. Sometimes it's set off by minor earthquakes. Is there a burglar?
- Variables: *Burglary*, *Earthquake*, *Alarm*, *JohnCalls*, *MaryCalls*
- Network topology reflects "causal" knowledge:
 - A burglar can set the alarm off
 - An earthquake can set the alarm off
 - The alarm can cause Mary to call
 - The alarm can cause John to call



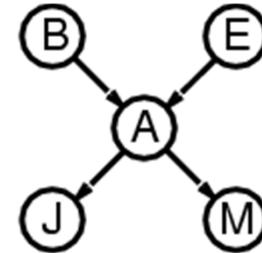
Example contd.



Compactness

- A CPT for Boolean X_i with k Boolean parents has 2^k rows for the combinations of parent values

- Each row requires one number p for $X_i = \text{true}$ (the number for $X_i = \text{false}$ is just $1-p$)



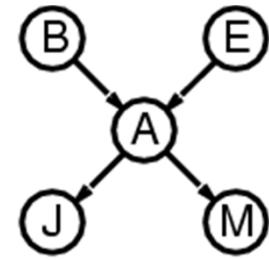
- If each variable has no more than k parents, the complete network requires $O(n \cdot 2^k)$ numbers
- I.e., grows linearly with n , vs. $O(2^n)$ for the full joint distribution
- For burglary net, $1 + 1 + 4 + 2 + 2 = 10$ numbers (vs. $2^5 - 1 = 31$)



Semantics

The full joint distribution is defined as the product of the local conditional distributions:

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



e.g., $\mathbf{P}(j \wedge m \wedge a \wedge \neg b \wedge \neg e)$

$$= \mathbf{P}(j | a) \mathbf{P}(m | a) \mathbf{P}(a | \neg b, \neg e) \mathbf{P}(\neg b) \mathbf{P}(\neg e)$$

$$= 0.90 * 0.7 * 0.001 * 0.999 * 0.998$$

$$\approx 0.00063$$



Constructing Bayesian networks

1. Choose an ordering of variables X_1, \dots, X_n
2. For $i = 1$ to n
 - add X_i to the network
 - select parents from X_1, \dots, X_{i-1} such that
$$\mathbf{P}(X_i | \text{Parents}(X_i)) = \mathbf{P}(X_i | X_1, \dots, X_{i-1})$$
with $\text{Parents}(X_i) \subseteq \{X_1, \dots, X_{i-1}\}$

This choice of parents guarantees:

$$\mathbf{P}(X_1, \dots, X_n) = \prod_{i=1}^n \mathbf{P}(X_i | X_1, \dots, X_{i-1})$$

(chain rule)

$$= \prod_{i=1}^n \mathbf{P}(X_i | \text{Parents}(X_i))$$

(by construction)



Example

Suppose we choose the ordering M, J, A, B, E

MaryCalls

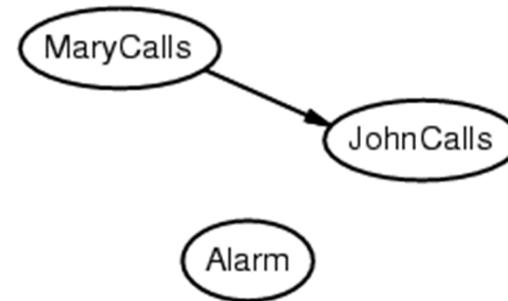
JohnCalls

$$P(J | M) = P(J)?$$



Example

Suppose we choose the ordering M, J, A, B, E



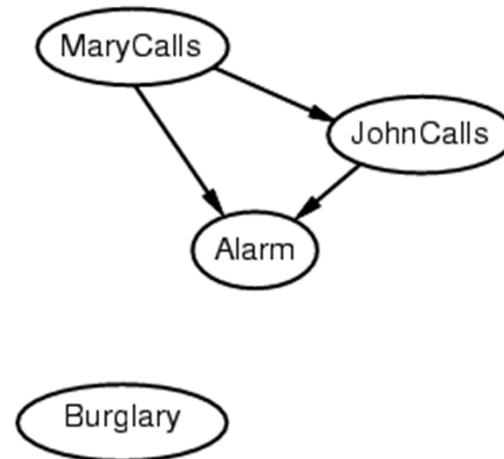
$P(J | M) = P(J)$? **No**

$P(A | J, M) = P(A | J)$? $P(A | J, M) = P(A)$?



Example

Suppose we choose the ordering M, J, A, B, E



$P(J | M) = P(J)$? **No**

$P(A | J, M) = P(A | J)$? $P(A | J, M) = P(A)$? **No**

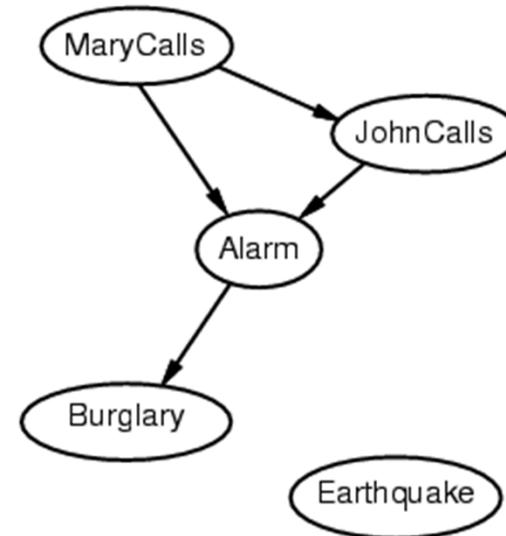
$P(B | A, J, M) = P(B | A)$?

$P(B | A, J, M) = P(B)$?



Example

Suppose we choose the ordering M, J, A, B, E



$P(J | M) = P(J)$? **No**

$P(A | J, M) = P(A | J)$? $P(A | J, M) = P(A)$? **No**

$P(B | A, J, M) = P(B | A)$? **Yes**

$P(B | A, J, M) = P(B)$? **No**

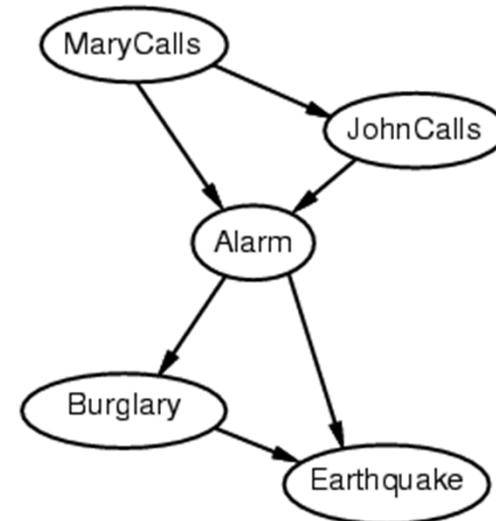
$P(E | B, A, J, M) = P(E | A)$?

$P(E | B, A, J, M) = P(E | A, B)$?



Example

Suppose we choose the ordering M, J, A, B, E



$P(J | M) = P(J)$? **No**

$P(A | J, M) = P(A | J)$? $P(A | J, M) = P(A)$? **No**

$P(B | A, J, M) = P(B | A)$? **Yes**

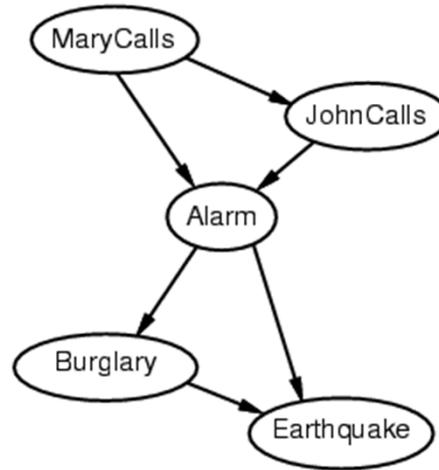
$P(B | A, J, M) = P(B)$? **No**

$P(E | B, A, J, M) = P(E | A)$? **No**

$P(E | B, A, J, M) = P(E | A, B)$? **Yes**



Example contd.

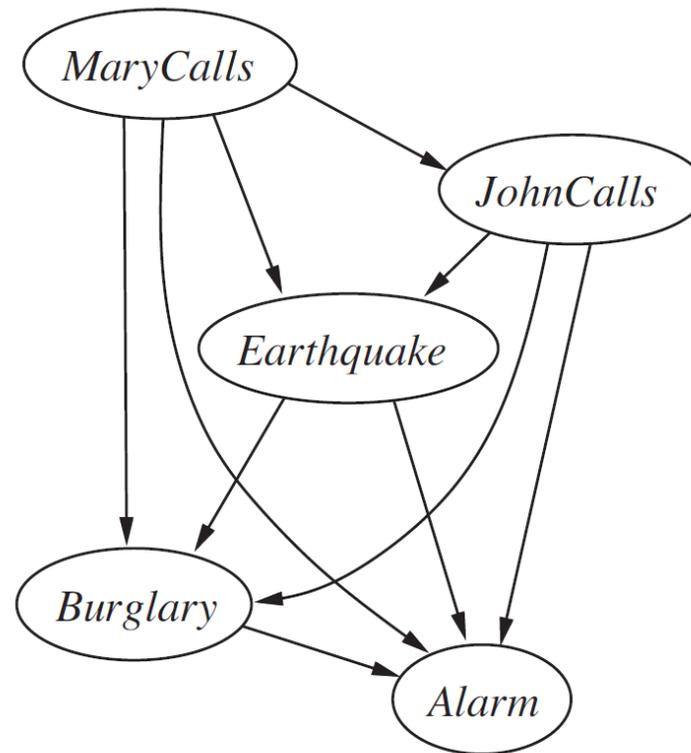


- Deciding conditional independence is hard in noncausal directions
- (Causal models and conditional independence seem hardwired for humans!)
- Network is less compact: $1 + 2 + 4 + 2 + 4 = 13$ numbers needed



Example contd.

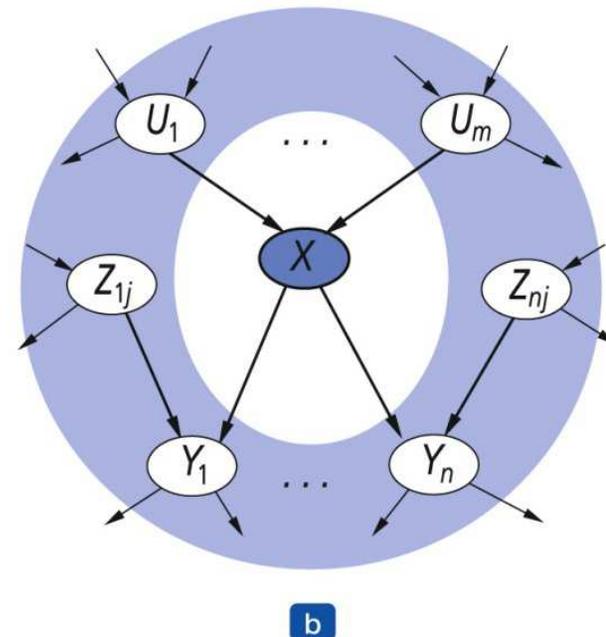
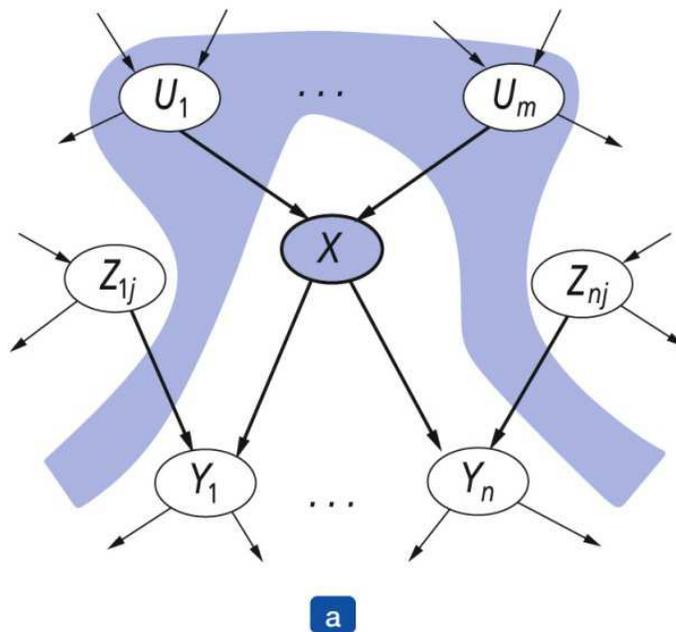
Even worse structure if order is M, J, E, B, A



Structure and conditional independence

- a) Local semantics: each node is conditionally independent of its nondescendants given its parents
- b) Topological semantics: each node is conditionally independent of all others given its

Markov blanket: parents + children + children's parents



Compact conditional distributions

- CPT grows exponentially with number of parents
- CPT becomes infinite with continuous-valued parent or child

Solution: **canonical** distributions that are defined compactly

Deterministic nodes are the simplest case:

$$X = f(\text{Parents}(X)) \text{ for some function } f$$

E.g., Boolean functions

$$\text{NorthAmerican} \Leftrightarrow \text{Canadian} \vee \text{US} \vee \text{Mexican}$$

E.g., numerical relationships among continuous variables

$$\frac{\partial \text{Level}}{\partial t} = \text{inflow} + \text{precipitation} - \text{outflow} - \text{evaporation}$$



Compact conditional distributions contd.

- Noisy-OR distributions model multiple noninteracting causes
- 1) Parents U_1, \dots, U_k include all causes (can add leak node)
- 2) Independent failure probability q_i for each cause alone

$$P(X_i | Parents(X_i)) = 1 - \prod_{j: X_j = true} q_i$$

<i>Cold</i>	<i>Flu</i>	<i>Malaria</i>	$P(Fever)$	$P(\neg Fever)$
F	F	F	0.0	1.0
F	F	T	0.9	0.1
F	T	F	0.8	0.2
F	T	T	0.98	$0.02 = 0.2 \times 0.1$
T	F	F	0.4	0.6
T	F	T	0.94	$0.06 = 0.6 \times 0.1$
T	T	F	0.88	$0.12 = 0.6 \times 0.2$
T	T	T	0.988	$0.012 = 0.6 \times 0.2 \times 0.1$

Number of parameters linear in number of parents



Summary

- Bayes nets provide a natural representation for (causally induced) conditional independence
- Topology + CPTs = compact representation of joint distribution
- Generally easy for (non)experts to construct
- Canonical distributions (e.g., noisy-OR) = compact representation of CPTs

