Methods of Artificial Intelligence: Lecture

9. Session: Vagueness and Uncertainty II (FINAL PART)

Kai-Uwe Kühnberger, Nohayr Muhammad

Winter Term 2022/2023

January 20th, 2023

Overview

- Bayesian Networks
  - Bayes' Rule
  - Example
  - Constructing Bayesian Networks

Bayesian Networks

Bayes' Rule & Example

Bayes' Rule

Product rule $P(a \wedge b) = P(a|b)P(b) = P(b|a)P(a)$

$\Rightarrow$ Bayes' rule $P(a|b) = \dfrac{P(b|a)P(a)}{P(b)}$

- Example: *Sore throat* $\rightarrow$ *flu*
  - Known facts
    - *P*(*sore throat*) = 0.3
    - *P*(*flu*) = 0.4
    - *P*(*sore throat* | *flu*) = 0.6

$$P(flu \mid sorethroat) = \frac{P(sorethroat \mid flu)P(flu)}{P(sorethroat)}$$

  - With *P* = 0.8 we can infer that someone has the flu if she reports a sore throat

Towards Bayesian Networks

Idea: A single cause directly influences several effects and all effects are conditionally independent given the cause.

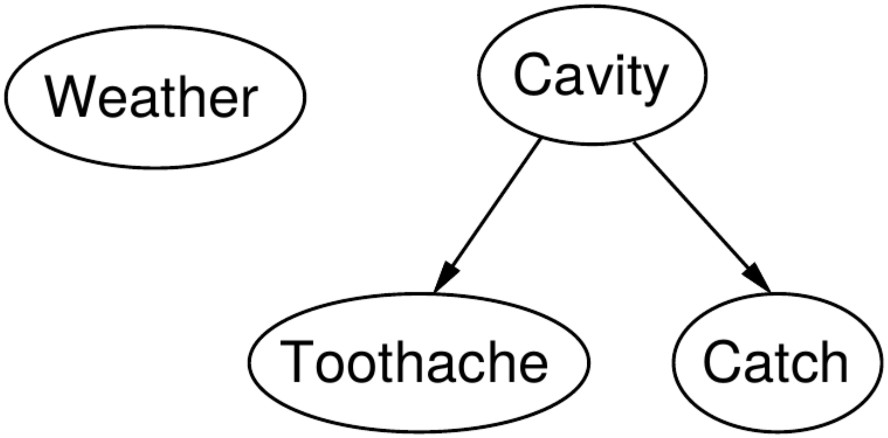
This is an example of a naive Bayes model:

$$\mathbf{P}(Cause, Effect_1, \ldots, Effect_n) = \mathbf{P}(Cause)\Pi_i \mathbf{P}(Effect_i | Cause)$$

Total number of parameters is **linear** in n

Towards Bayesian Networks

Topology of network encodes conditional independence assertions:

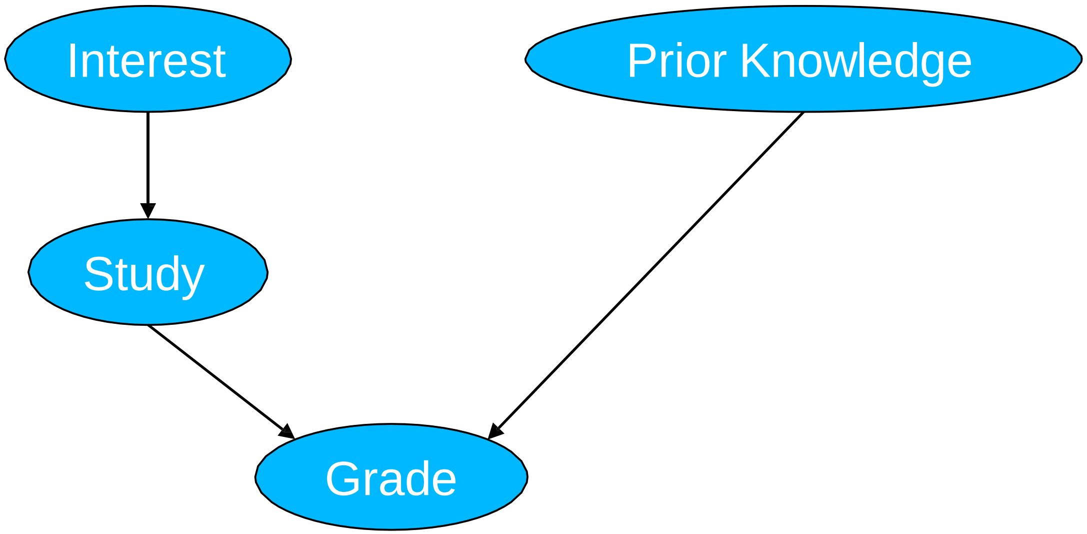


$Weather$ is independent of the other variables

$Toothache$ and $Catch$ are conditionally independent given $Cavity$

Bayesian Networks

- Bayesian network implicitly specifies a joint probability distribution $P(X_1, X_2, \ldots, X_n)$ over random variables
- Explicitly, we only have to define n conditional probability distributions $P(X_i \mid \text{Parents}(X_i))$ for $i=1,\ldots,n$
- $\text{Parents}(X_i)$ denotes a subset of $\{X_1, X_2, \ldots, X_n\}$ called the set of parents of X_i



$\text{Parents}(I) = \{\}$

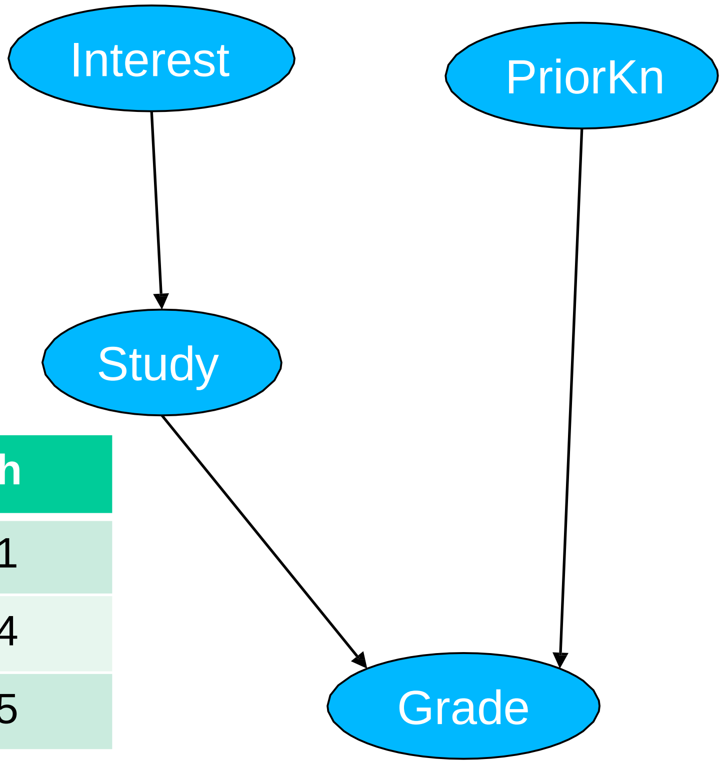
$\text{Parents}(PK) = \{\}$

$\text{Parents}(S) = \{I\}$

$\text{Parents}(G) = \{S, PK\}$

Bayesian Networks Example

Interest	P(Interest)
low	0.1
normal	0.7
high	0.2



PK	P(PK)
no	0.8
yes	0.2

Study	I=low	I=n	I=h
little	0.6	0.3	0.1
normal	0.3	0.4	0.4
much	0.1	0.3	0.5

Grade	(S=little, PK=no)	(S=little, PK=yes)	…
1.0	0.01	0.02	…
1.3	0.02	0.04	…
…	…	…	…

Bayesian Networks Definition

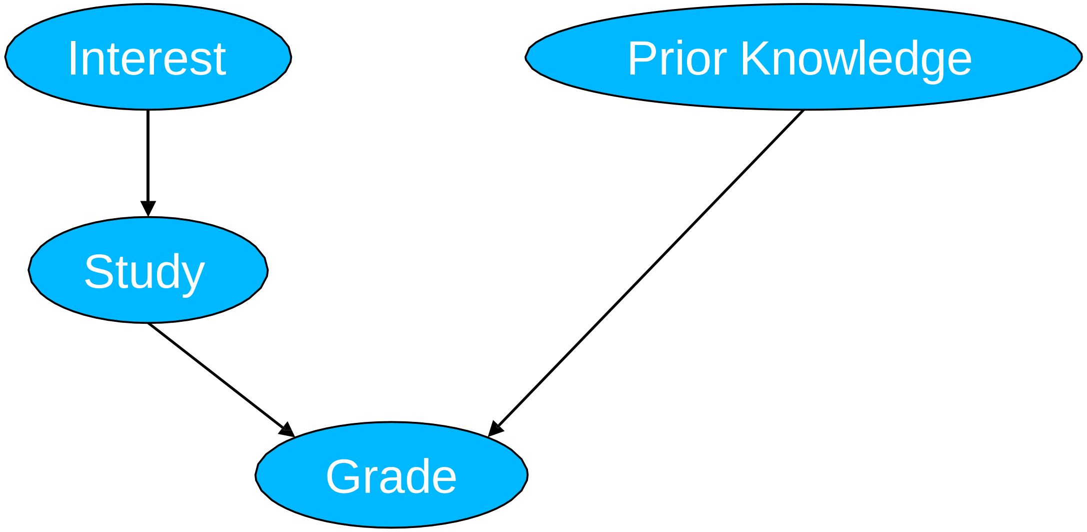
- Formally, a Bayesian network is a pair ($P(X_1, \ldots, X_n)$, G), where
  - $P(X_1, \ldots, X_n)$ is a joint distribution over random variables $X_1, \ldots, X_n$
  - G is a directed, acyclic graph (DAG) over $X_1, \ldots, X_n$ (BN Structure)
  - P factorizes according to G, that is,

$$P(X_1, \ldots, X_n) = P(X_1 \mid Parents(X_1)) * \ldots * P(X_n \mid Parents(X_n)),$$

where $Parents(X_i)$ is the set of parents of X_i in G

- Important: a Bayesian network is not just a DAG, but also consists of a probabilistic model

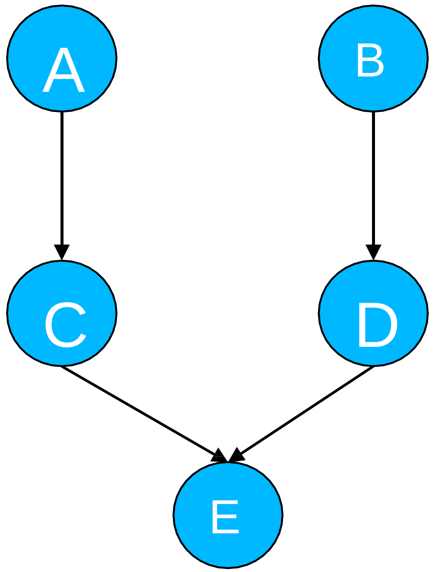
Bayesian Networks Factorization



$$P(I, PK, S, G) = P(I) * P(PK) * P(S \mid I) * P(G \mid S, PK)$$

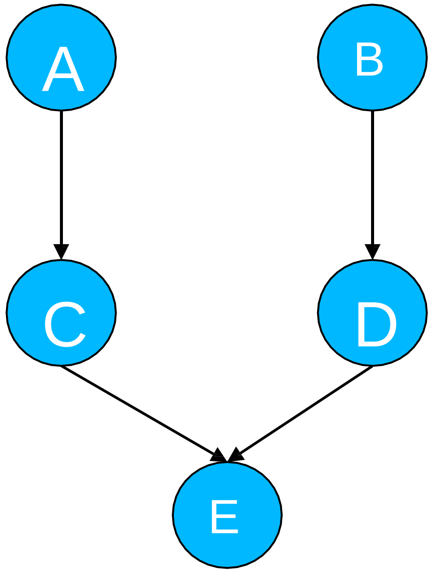
Exercise

- Assume we are given a Bayesian network (P, G) and the BN structure G looks as follows:



- Write down the factorization of P.

Solution



$$P(A,B,C,D,E) = P(A) * P(B) * P(C \mid A) * P(D \mid B) * P(E \mid C, D)$$

Bayesian Networks

Constructing a Bayesian Network

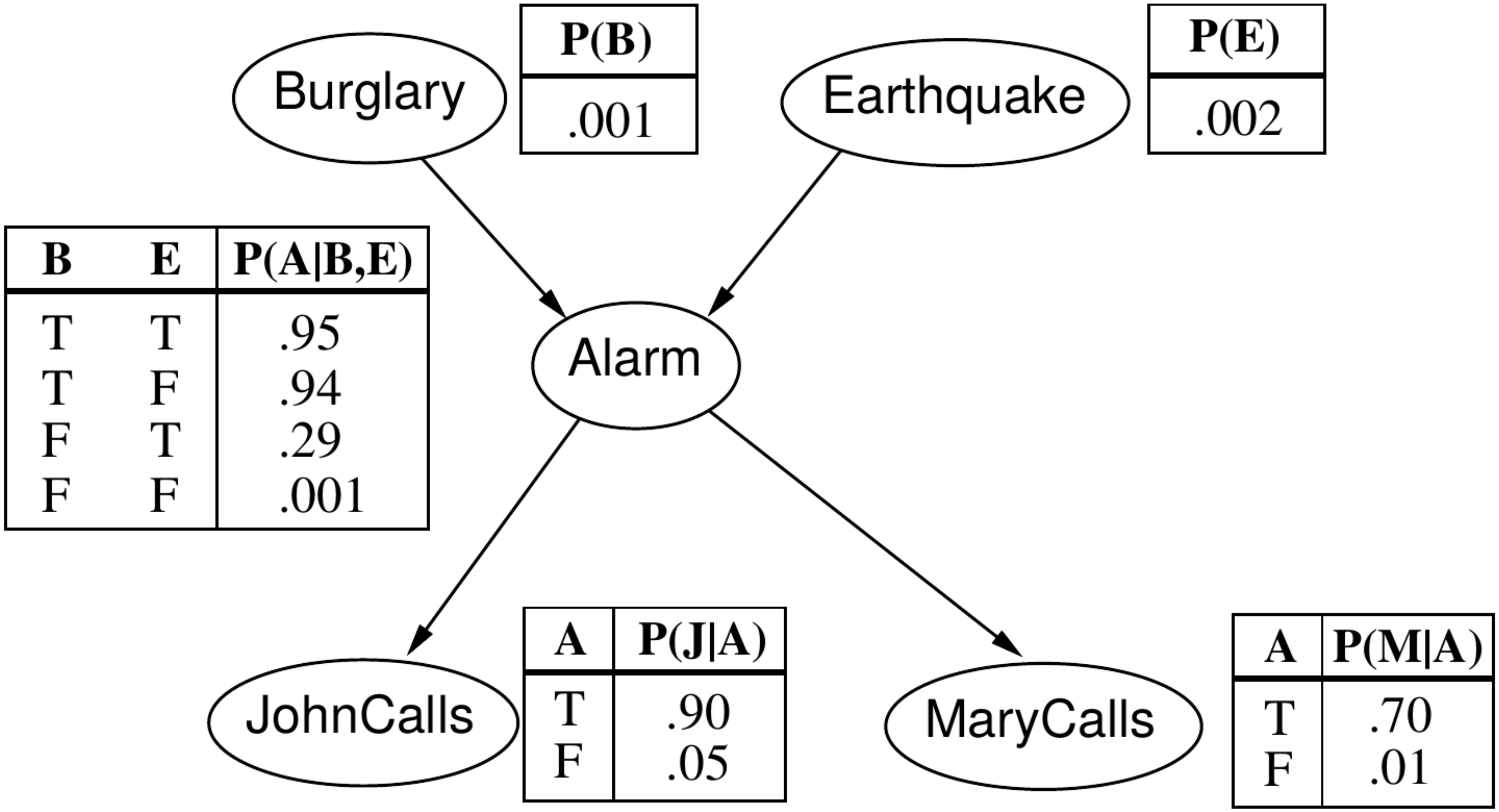
Example

Scenario:

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: $Burglar$, $Earthquake$, $Alarm$, $JohnCalls$, $MaryCalls$
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



P(B)
.001

P(E)
.002

B	E	P(A|B,E)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

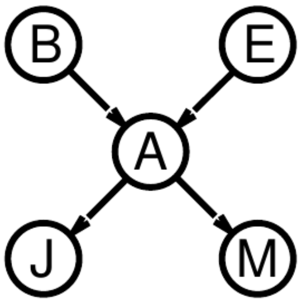
A	P(J|A)
T	.90
F	.05

A	P(M|A)
T	.70
F	.01

Properties

- CPT: Conditional Probability Table

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 $X_i = true$ (the number for $X_i = false$ 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 numbers)

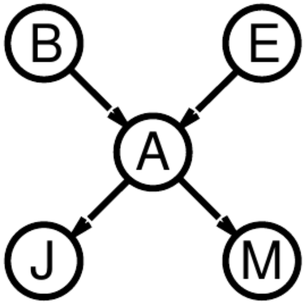
Global Semantics

"Global" semantics defines the full joint distribution as the product of the local conditional distributions:

$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | parents(X_i))$$

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

$$\begin{aligned} &= P(j|a)P(m|a)P(a|\neg b, \neg e)P(\neg b)P(\neg e) \\ &= 0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998 \\ &\approx 0.00063 \end{aligned}$$



Constructing Bayesian Networks

Need a method such that a series of locally testable assertions of conditional independence guarantees the required global semantics

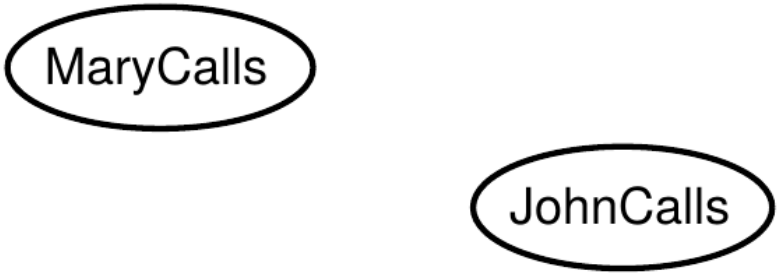
1. Choose an ordering of variables $X_1, \ldots, X_n$
2. For $i = 1$ to n
   add X_i to the network
   select parents from $X_1, \ldots, X_{i-1}$ such that
   $\mathbf{P}(X_i | Parents(X_i)) = \mathbf{P}(X_i | X_1, \ldots, X_{i-1})$

This choice of parents guarantees the global semantics:

$$\begin{aligned} \mathbf{P}(X_1, \ldots, X_n) &= \prod_{i=1}^{n} \mathbf{P}(X_i | X_1, \ldots, X_{i-1}) \quad \text{(chain rule)} \\ &= \prod_{i=1}^{n} \mathbf{P}(X_i | Parents(X_i)) \quad \text{(by construction)} \end{aligned}$$

Example

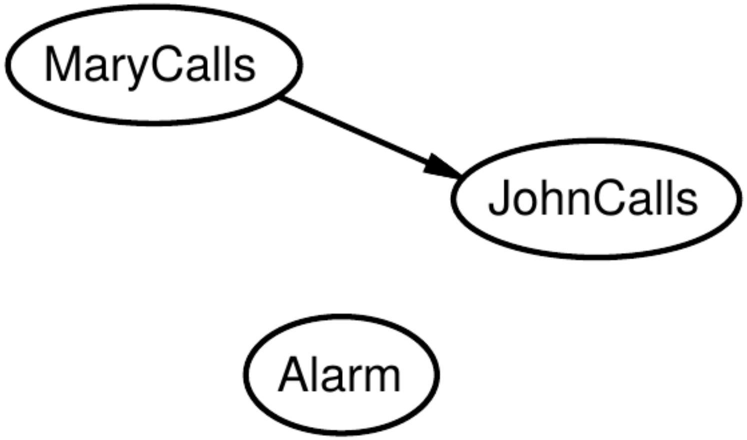
Suppose we choose the ordering M, J, A, B, E



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

Example

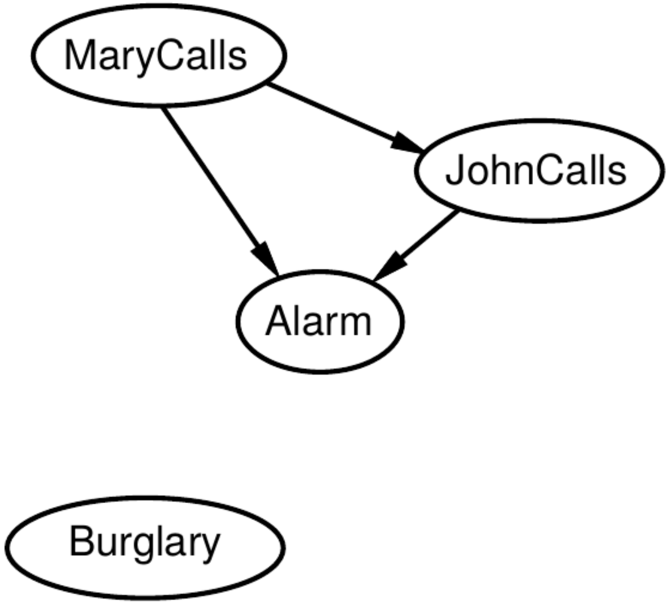
Suppose we choose the ordering M, J, A, B, E



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

Example

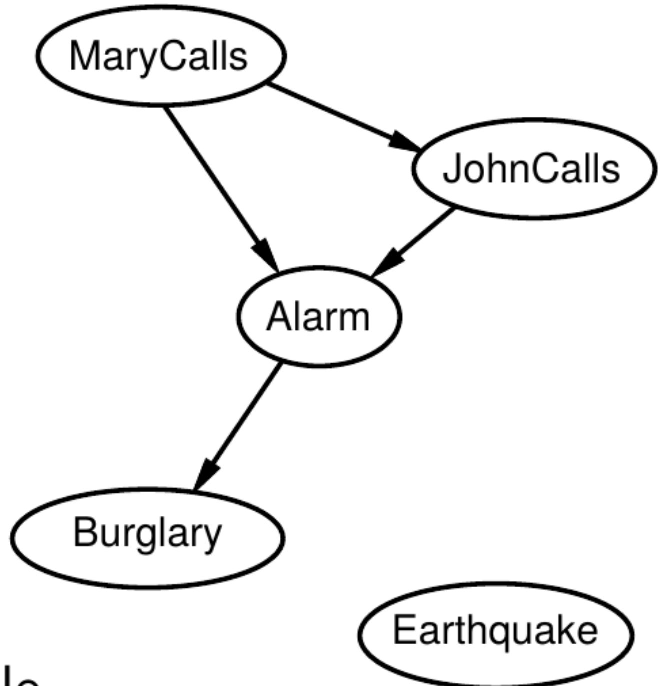
Suppose we choose the ordering M, J, A, B, E



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

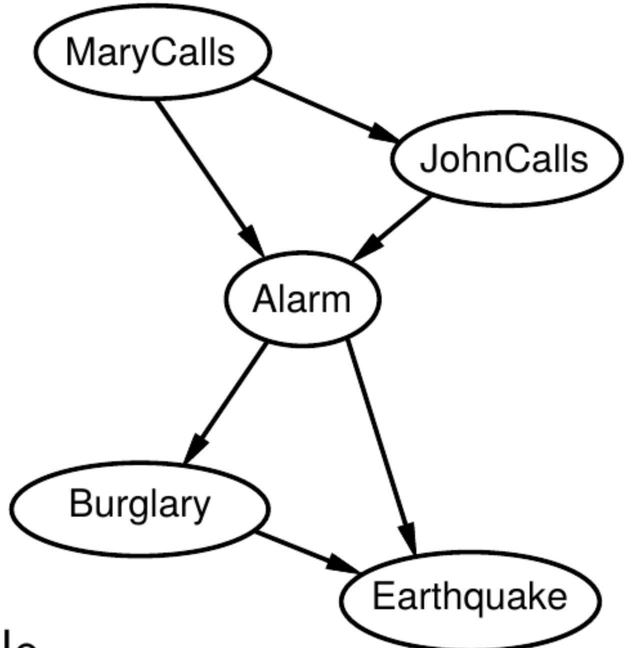
$P(B|A, J, M) = P(B)$? No

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

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

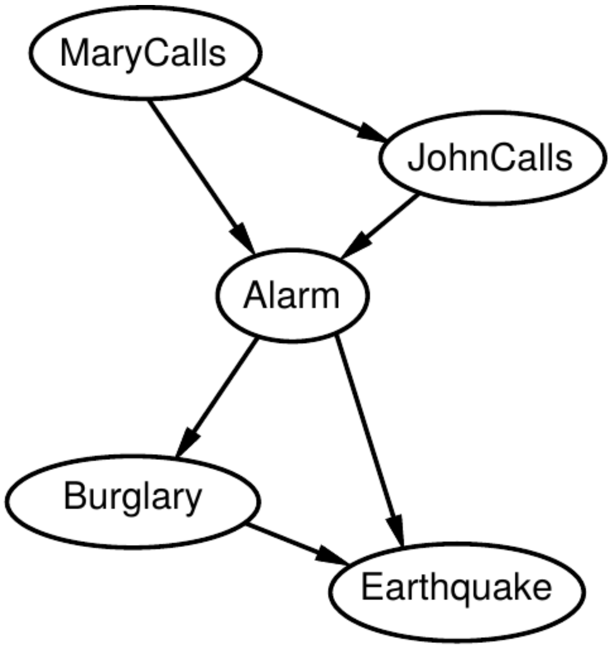
Example

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



Deciding conditional independence is hard in noncausal directions

(Causal models and conditional independence seem hardwired for humans!)

Assessing conditional probabilities is hard in noncausal directions

Network is less compact: $1+2+4+2+4=13$ numbers needed

Further Readings

Most topics can be found in:

- S. Russell, P. Norvig.
  Artificial Intelligence - A modern approach.
  Pearson Education. 2010. 3rd Edition.
    Sections 13.1: Introduction to acting under uncertainty
    Sections 13.2–3: Basic probability and probabilistic logic notions
    Sections 14.7.3: Fuzzy sets and fuzzy logics

- T.J. Ross.
  Fuzzy logic with engineering applications.
  John Wiley & Sons. 2009.

- D. Koller, N. Friedman.
  Probabilistic graphicals models: principles and techniques.
  MIT Press. 2009.