Artificial Intelligence

Probabilistic Reasoning

CS 444 – Spring 2019

Dr. Kevin Molloy
Department of Computer Science
James Madison University

JAMES MADISON UNIVERSITY

CS RESEARCH SEMINAR

ISAT/CS 243 12:20-1:10
Brown bag lunch. All are welcome.

- 2/22 - Dr. Jason Forsyth
- 3/1 - Dr. Michael Lam
- 3/15 - Dr. Chris Mayfield
- **3/29 - Dr. Nathan Sprague**
- 4/5 - Dr. Michael Stewart
- 4/12 - Mrs. Eliza Shoemaker & Mr. Randy Shoemaker
- 4/19 - Ms. Becky Wild & Mr. Adam Blalock

Bayesian Networks

A simple, graphical notation for conditional independence assertions and hence for compact specification of full joint distribution

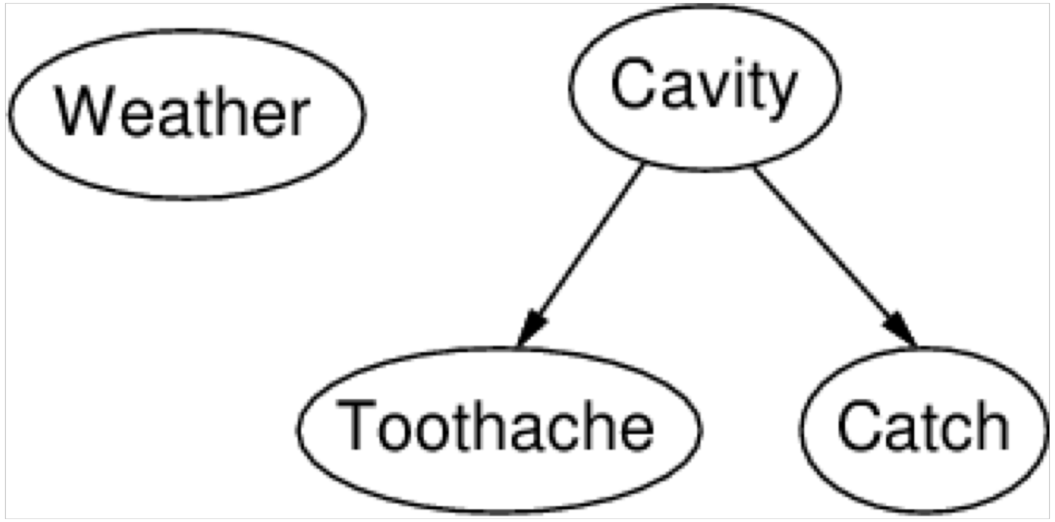
Syntax:

- A set of nodes, one per variable
- A directed, acyclic graph (link is approximately "directly influences")
- a conditional distribution for each node given its parents $P(X_i \mid Parents(X_i))$.

In the simplest case, conditional distribution represented as a conditional probability table (CPT) giving the distribution over X_i for each combination of parents values.

Example of a Bayesian Networks

Topology of network encodes conditional independence assertions:



Weather is independent of the other variables

Toothache and *Catch* are conditionally independent given *Cavity*.

Example of a Bayesian Networks

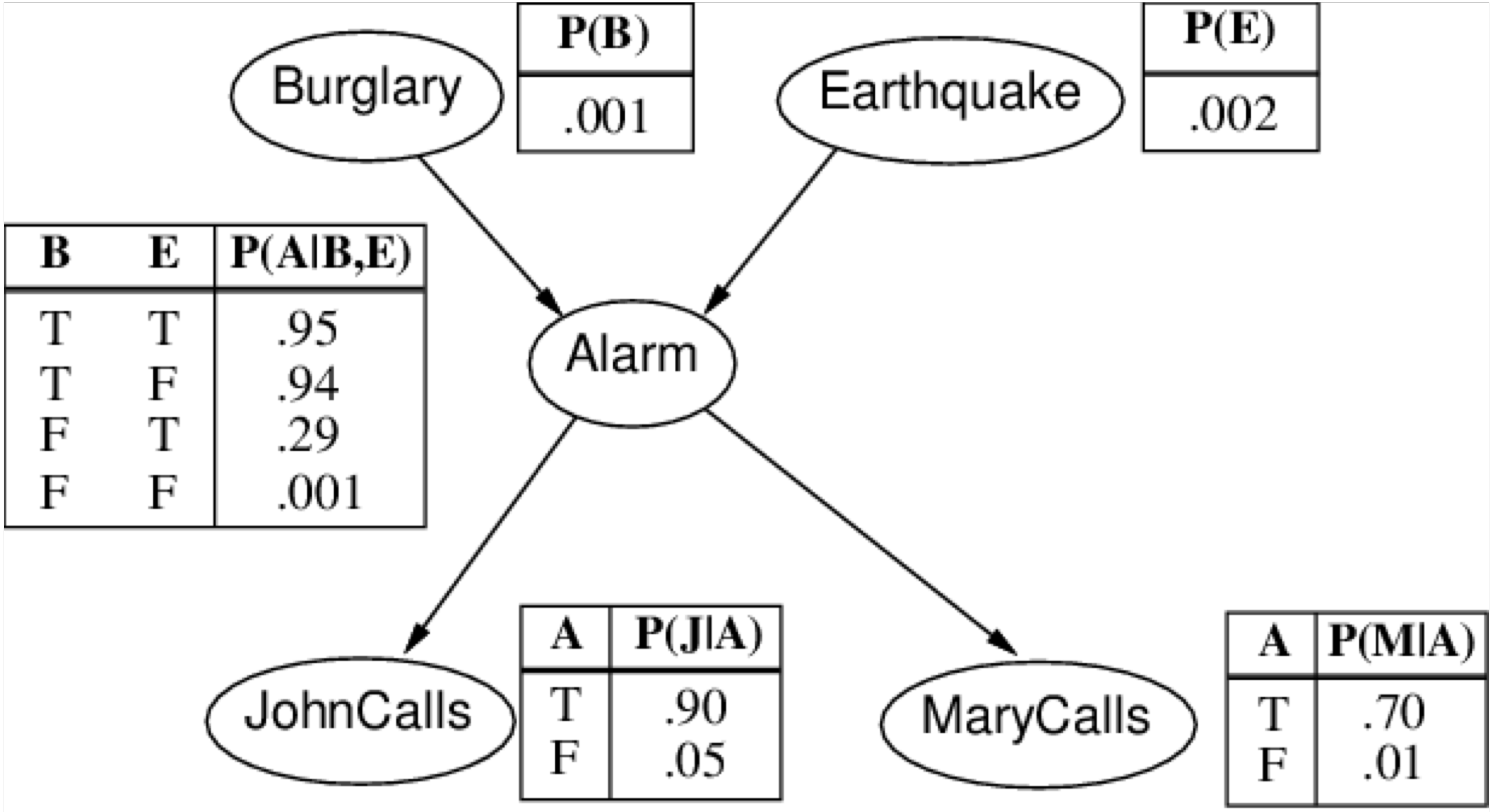
I'm at work, neighbor John calls to say my alarm is ringing, but my neighbor Mary doesn't call. Sometimes it's set off by monitor earthquakes. Is there a burglar?

Variables: Burglar, Earthquake, Alarm, JohnCalls, MaryCalls

Network topology reflects "causal" knowledge:

- *A burglar can set off the alarm*
- *An earthquake can set the alarm off*
- *The alarm can cause Mary to call*
- *The alarm can cause John to Call.*

Example of a Bayesian Networks



P(B)
.001

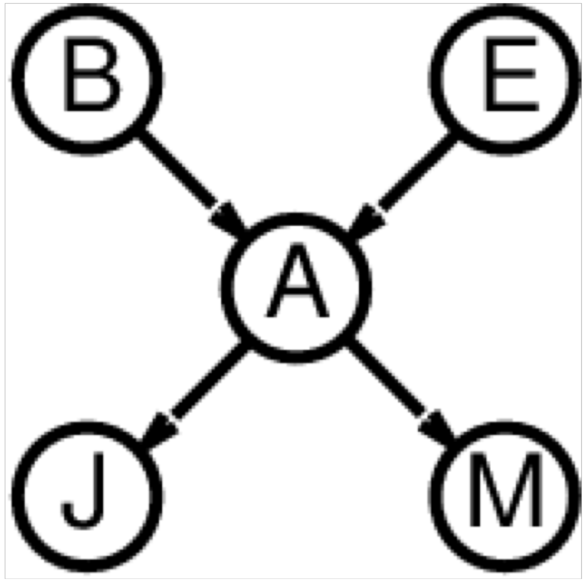
P(E)
.002

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

Compactness



A CPT for Boolean X_i with k Boolean parents.

Has:

2^k rows for the combinations of parent values

Each row requires on number *p for* X_i *= true*

(the number for X_i = false is simply 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 the burglary net, 1 + 1 + 4 + 2 + 2 = 10 numbers (vs. 25 – 1 = 31).

Global Semantics

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

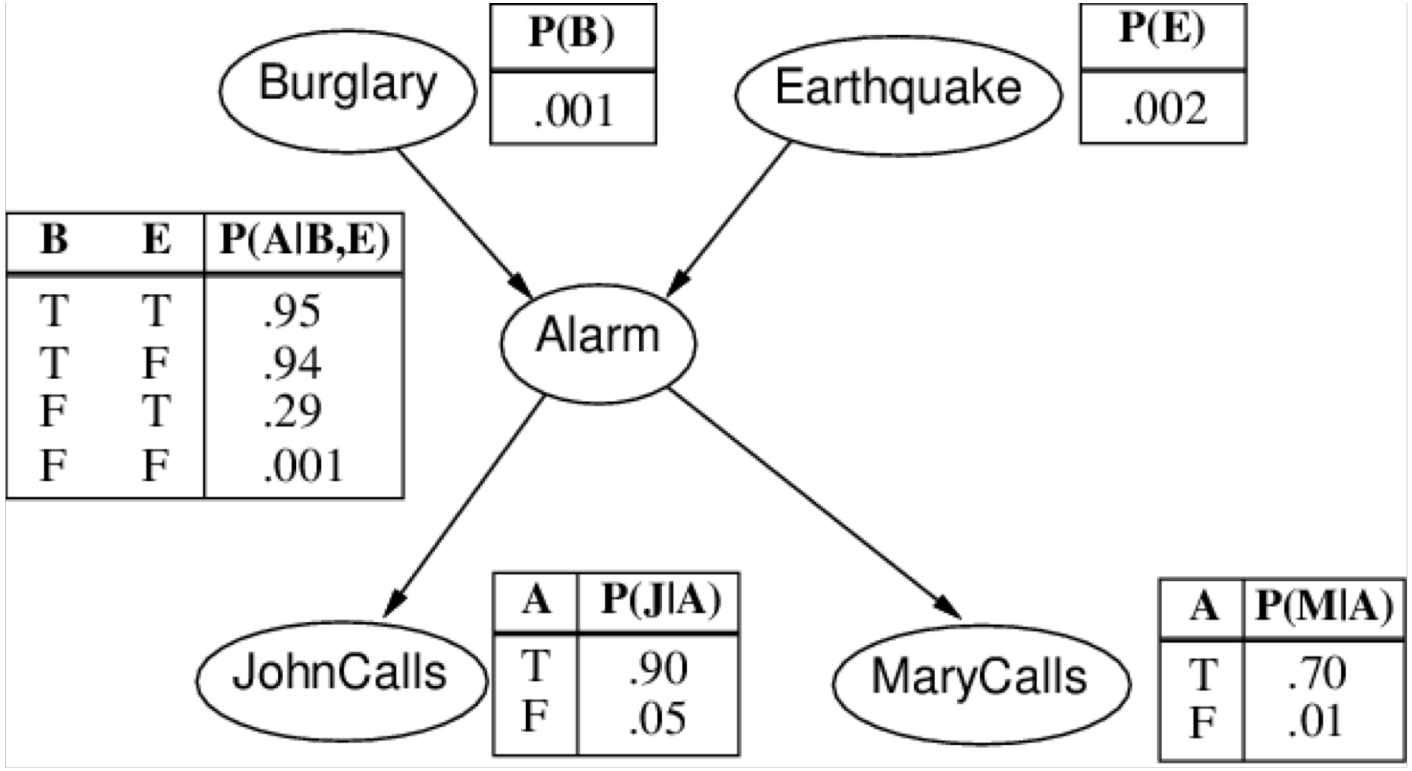
$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | parents\ (X_i))$$

e.g. P(j ∧ m ∧ a ∧ ¬b ∧ ¬e)

= P(j | a)P(M | a) P (a | ¬b, ¬e)P(¬b) P(¬e)

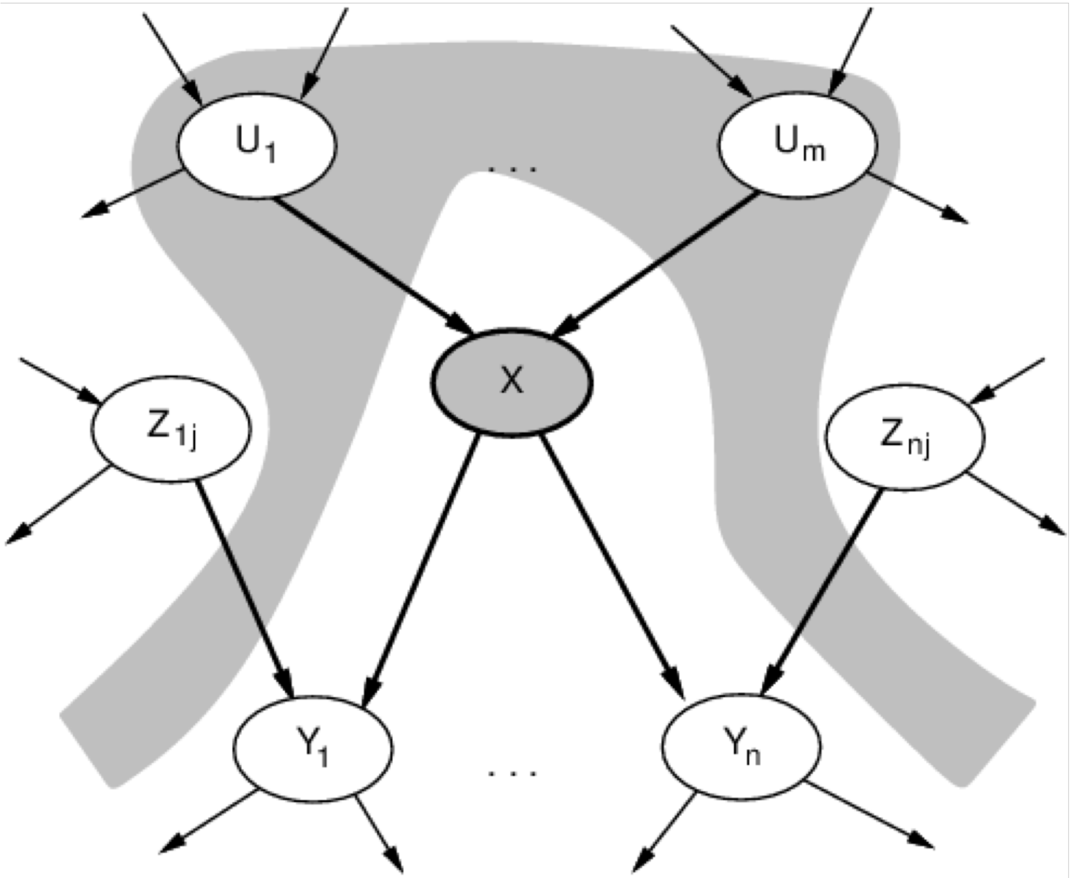
= 0.9 x 0.7 x 0.001 x 0.999 x .998

≈ 0.00063



Local Semantics

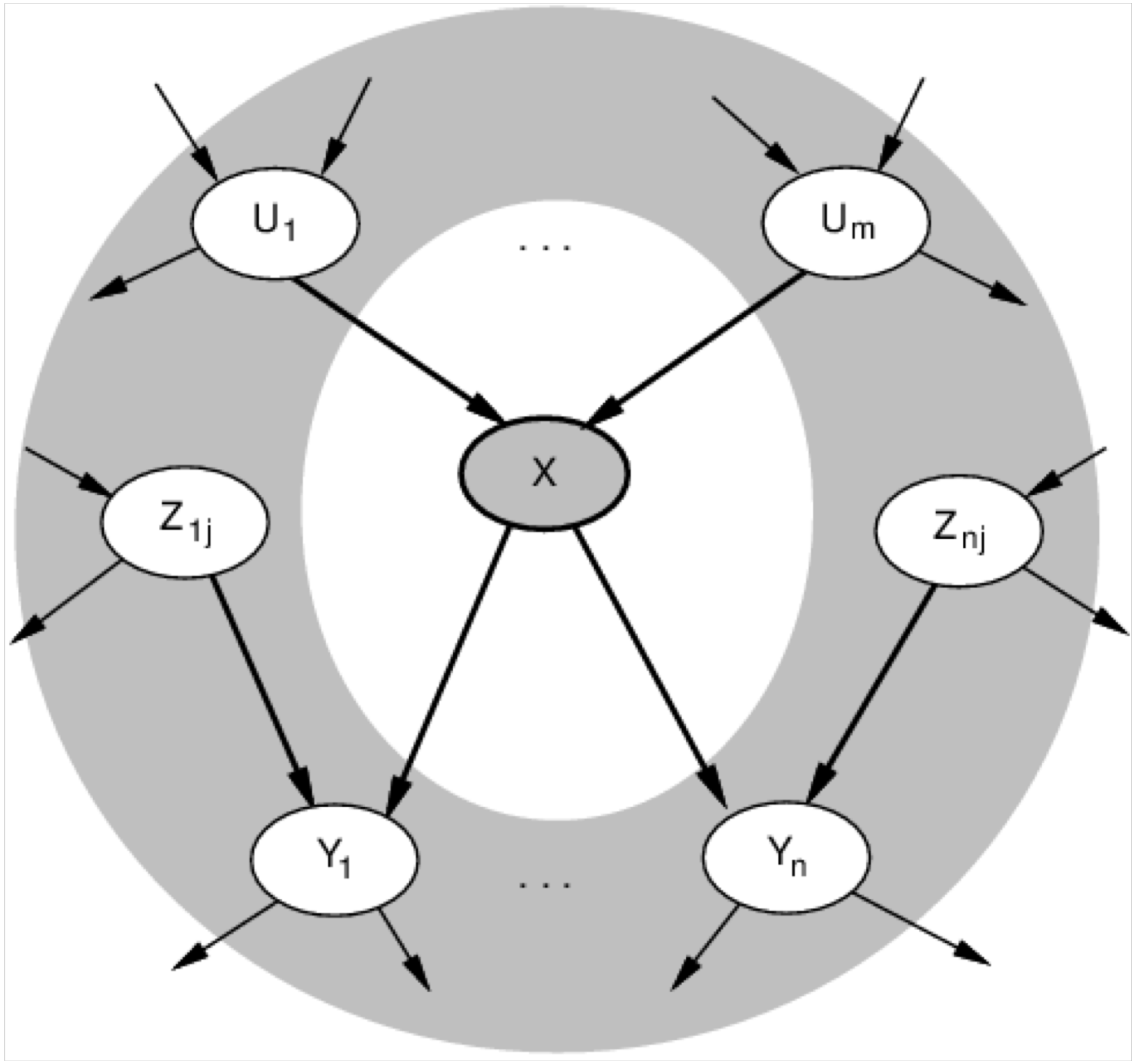
Local semantics: each node is conditionally independent of its nondescenants given its parents



Theorem: Local semantics ⇔ global semantics

Markov Blanket

Each node is conditionally independent of all others given its Markov blanket: parents + children + children's parents



Constructing a Bayesian Network

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

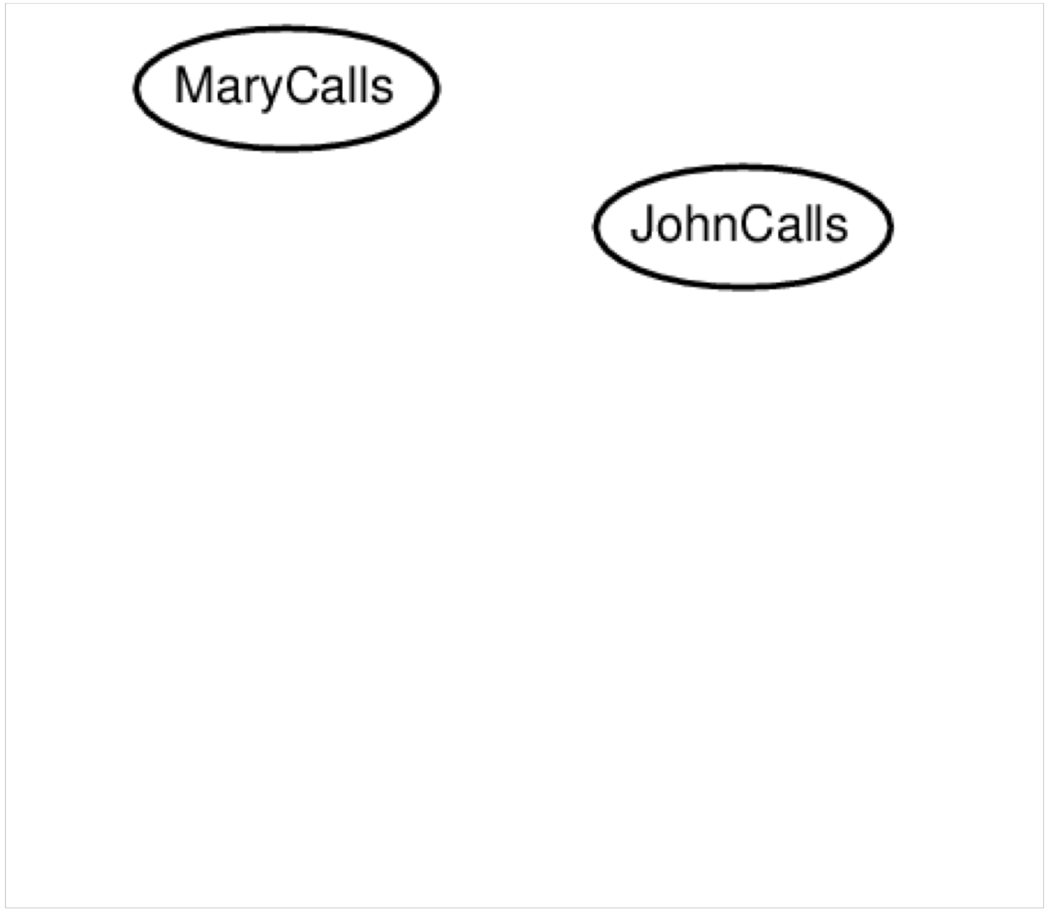
1. Choose an ordering of variables: $X_1, \ldots, X_n$
2. For I = 1 to n
   - Add X_i to the network
   - Select parents from $X_1, \ldots, X_{i-1}$ such that $P(X_i \mid Parents(X_i)) = P(X_i \mid X_1, \ldots, X_{i-1})$

This choice of parents guarantees the global semantics

Example

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

P(J | M) = P(J). No.

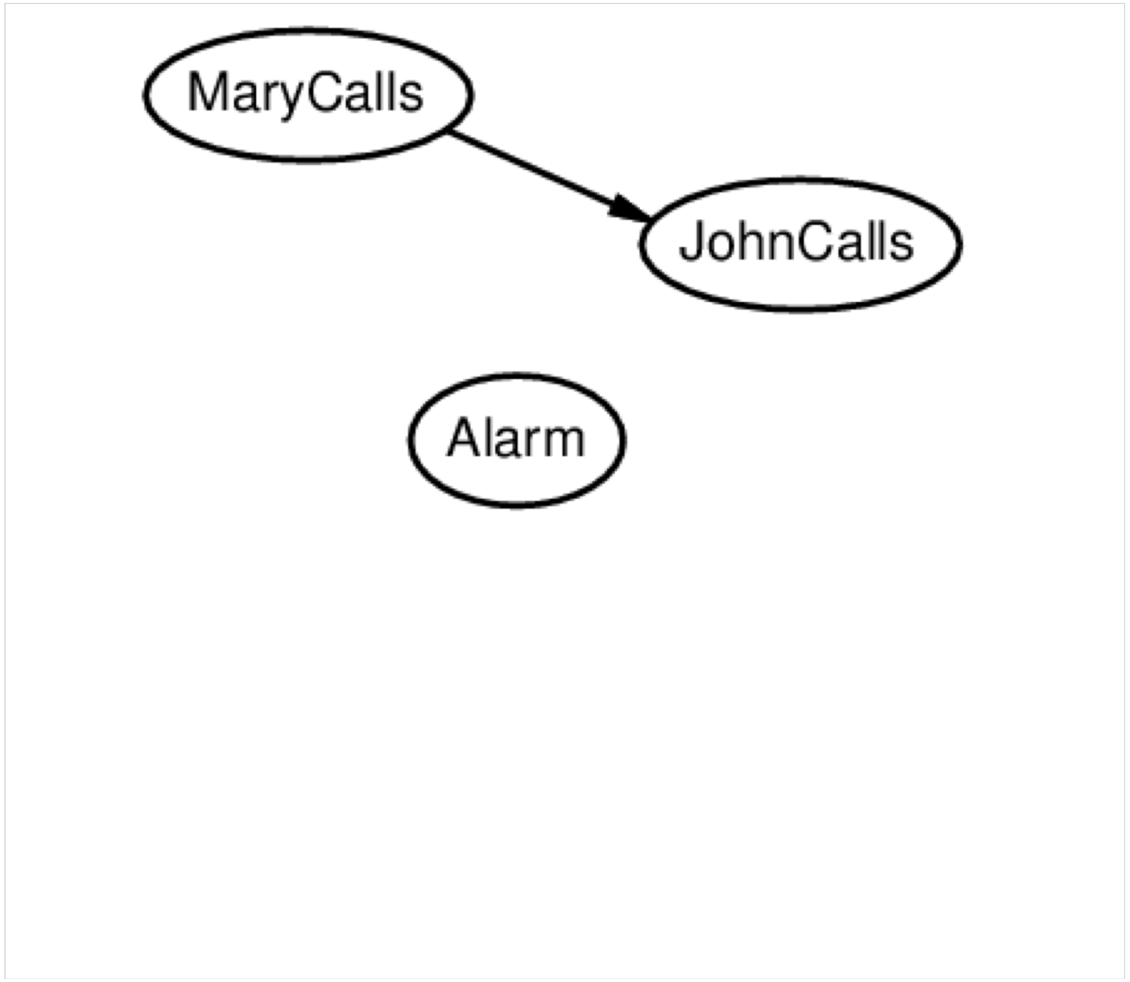


Example

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

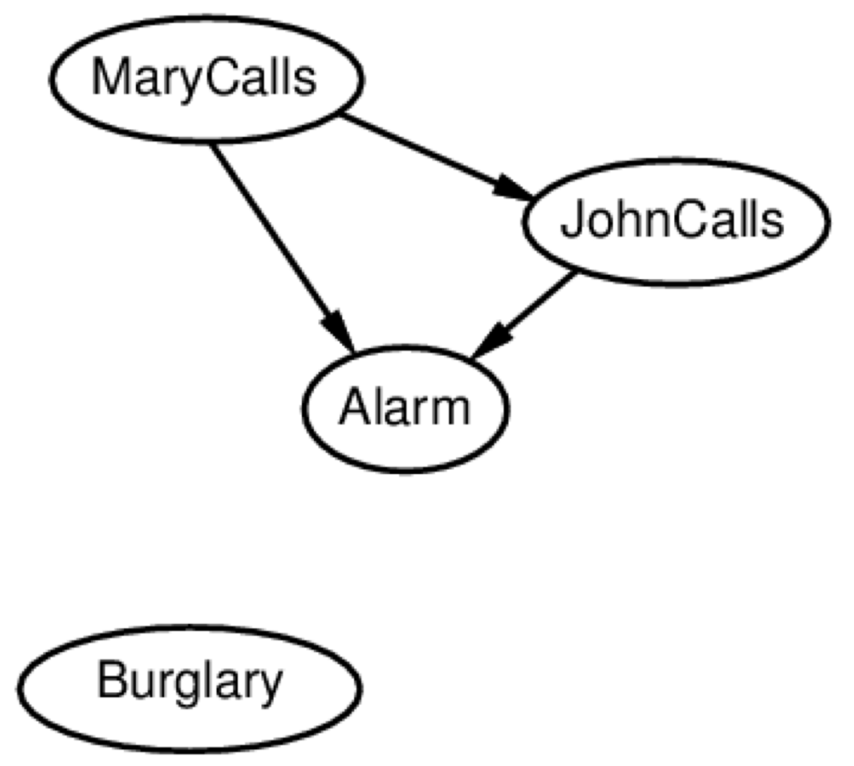
Suppose we choose the ordering M, J, A B, E

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, HJ, M) = P(B)?



Example

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

P(J | M) = P(J). No.

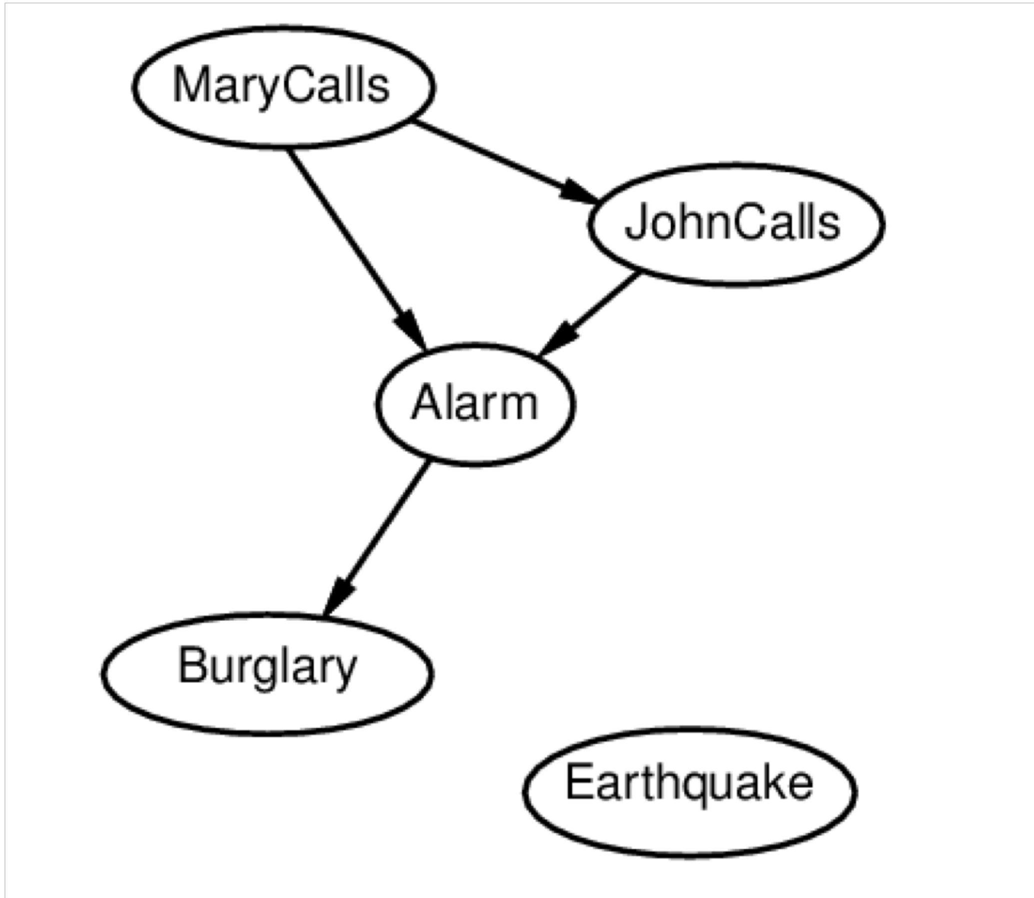
P(A | J, M) = P(A|J) P(A|J,M) = P(A). No.

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

P(B | A, HJ, 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.

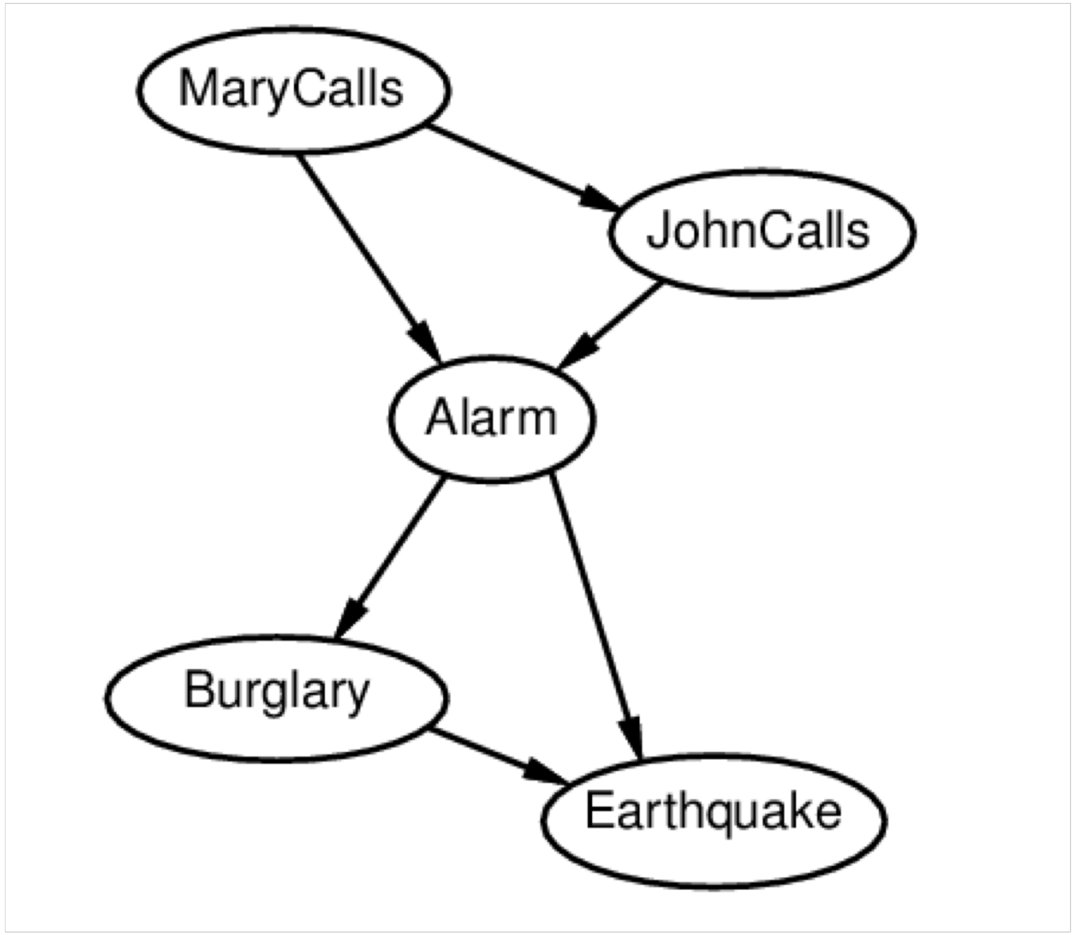
P(A | J, M) = P(A|J) P(A|J,M) = P(A). No.

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

P(B | A, HJ, M) = P(B)? No.

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

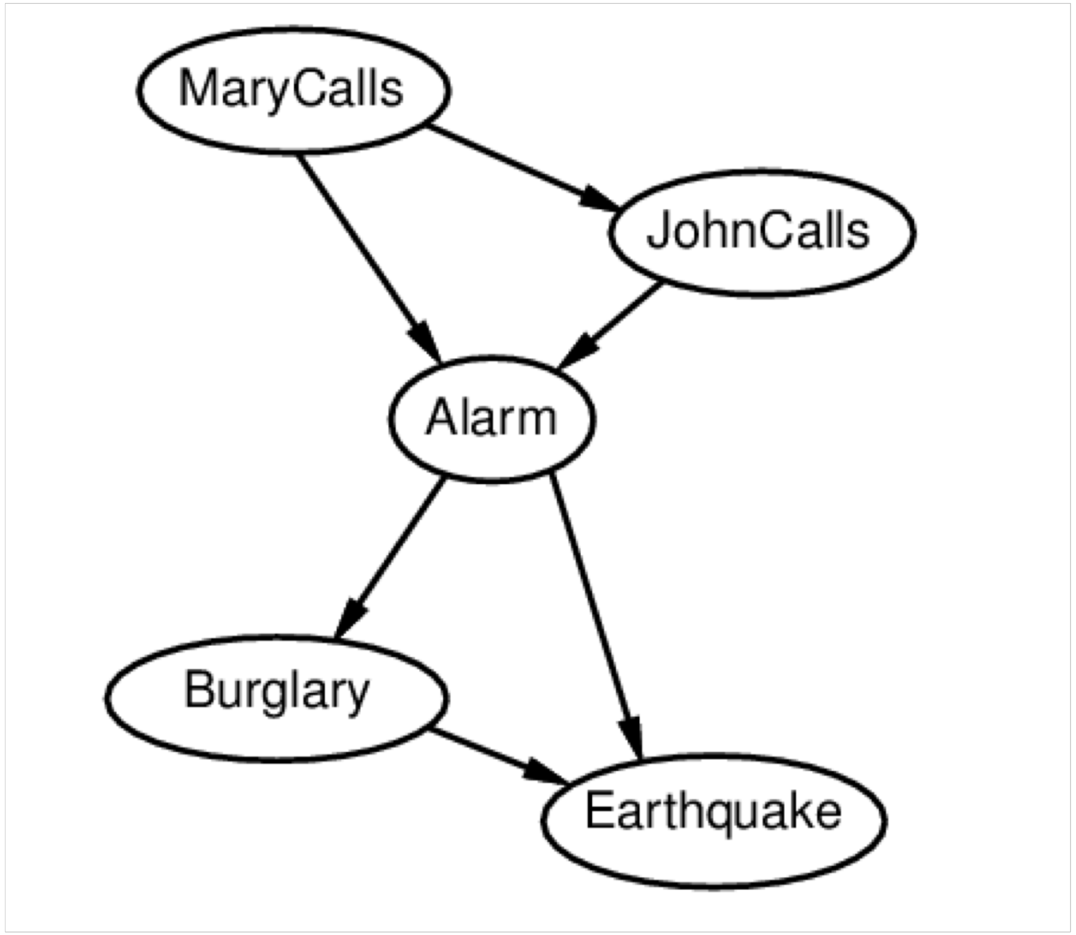
P(E |B, A, J, M) = P(E| A, B)? Yes.



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, resulting in the network being less compact: 1 + 2 + 4 + 2 + 4 = 13 numbers needed

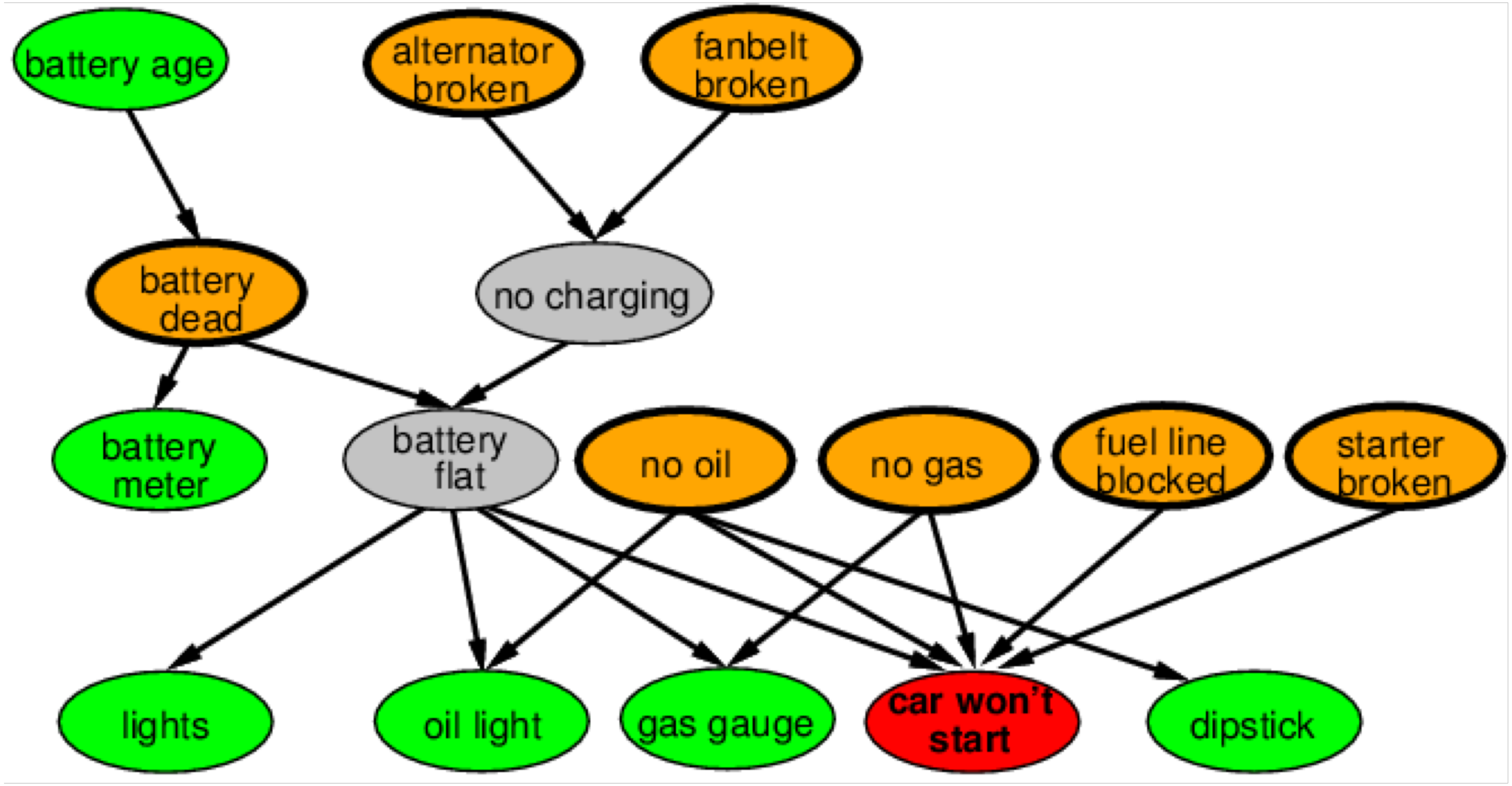


Example: Car Diagnosis

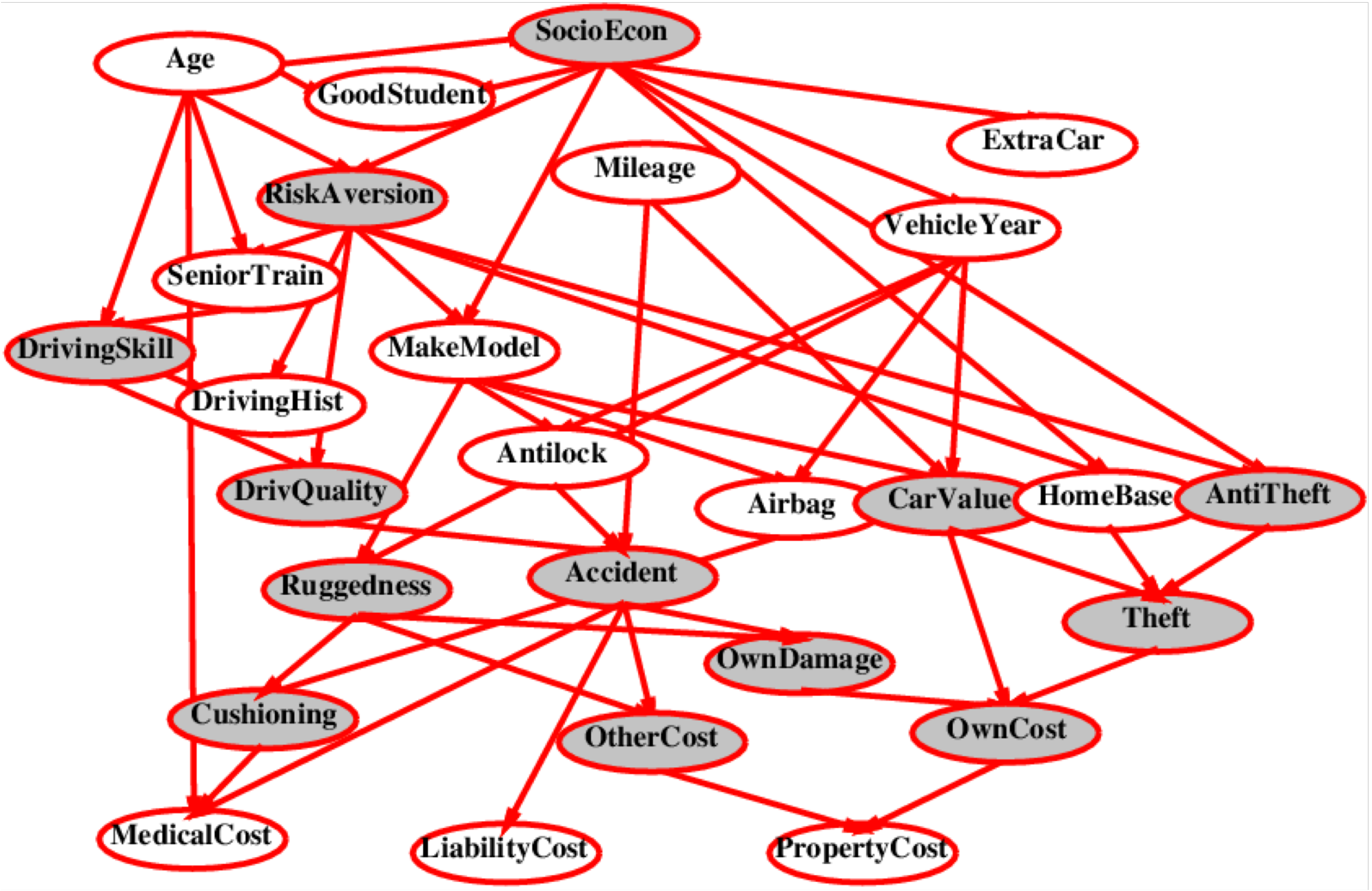
Initial evidence: car won't start

Testable variables (Green), "Broken, so fix it" variables are orange

Hidden variables (gray) ensure sparse structure, reduce parameters



Example: Car Insurance



JMU JAMES MADISON UNIVERSITY.

Compact Conditional Distributions

CPT grows exponentially with number of parents

CPT becomes infinite with continuous-valued parent or child

Some Exercises

We have a bag of 3 biased coins: a, b, c with probabilities of coming up heads of 20%, 60%, and 80% respectively. One coin is drawn radmonly from the bag (with equal likelihood od drawing each of the 3 coins), and then the coin is flipped 3 times to generate the outcomes X_1, X_2, and X_3.

1. Draw the Bayesian network corresponding to this setup and define the necessary CPTs.
2. Calculate which coin was most likely to have been drawn from the bag if the observed flips come out heads twice and tails once.

Some Exercises

Consider this Bayesian network.

1. If no evidence is observed, are Burglary and Earthquake independent? Explain why/why not.
2. If we observe Alarm = true, are Burglary and Earthquake independent? Justify your answer by calculating whether the probabilities involved satisfy the definition of conditional independence.