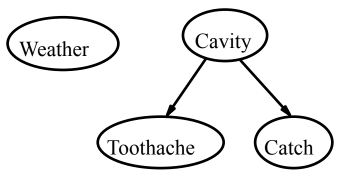
Bayesian networks

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 implies direct influence)
  - A conditional distribution for each node given its parents
- Conditional distributions
  - For each X_i, $P(X_i|Parents(X_i))$
  - In the form of a conditional probability table (CPT)
    - Distribution of X_i for each combination of parent values

Example

Topology of network encodes conditional independence assertions



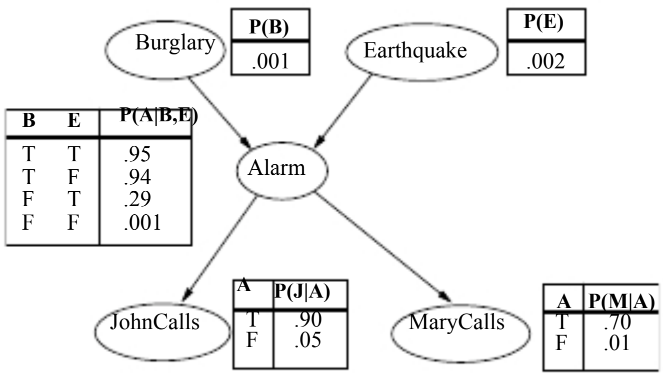
- Weather is independent of the other variables
- Toothache,Catch are conditionally independent given Cavity

Example

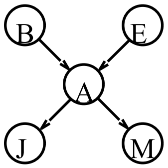
I’m at work, neighbor John calls to say my alarm is ringing, but neighbor Mary doesn’t call. Sometimes it’s set off by minor earthquakes. Is there a burglar?

- Variables: 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

Example (Contd.)



Compactness



A CPT for Boolean X_i with k Boolean parents

- 2^k rows for the combinations of parent values
- Each row requires one number

Each variable has no more than k parents

- The complete network requires $O(n \cdot 2^k)$ numbers
- Grows linearly with n
- Full joint distribution requires $O(2^n)$

Example: Burglary network

- Full joint distribution requires $2^5 - 1 = 31$ numbers
- Bayes net requires 10 numbers

Global semantics

- Full joint distribution
  - Can be written as product of local conditionals
- Example:

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

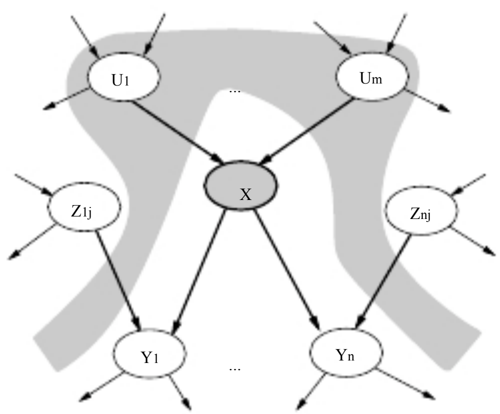
- Example:

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

- Can we compute $P(b|j,\neg m)$?

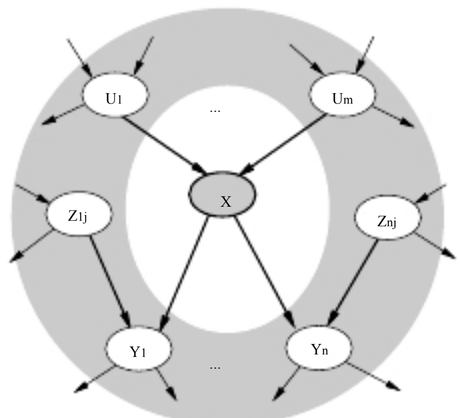
Local semantics

Each node is conditionally independent of its nondescendants given its parents



Markov blanket

Each node is conditionally independent of all others given its Markov blanket, i.e., parents + children + children's parents

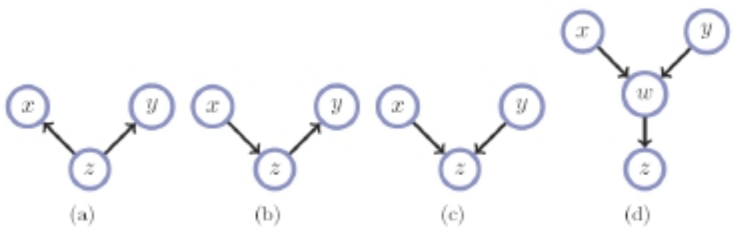


Conditional Independence in BNs

Which BNs support $x_1 \perp x_2 | x_3$

- For (a), x_1,x_2 are dependent, x_3 is a collider
- For (b)-(d), $x_1 \perp x_2 | x_3$

Conditional Independence (Contd.)



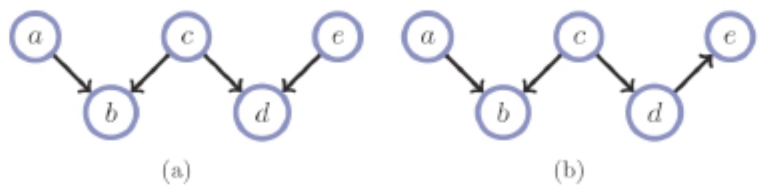
Which BNs support $x \perp y|z$

- For (a)-(b), z is not a collider, so $x \perp y|z$
- For (c), z is a collider, so x and y are conditionally dependent
- For (d), w is a collider, and z is a descendent of w, so x and y are conditionally dependent

d-connection, d-separation

- Definition (d-connection): X,Y,Z be disjoint sets of vertices in a directed graph G. X,Y is d-connected by Z iff $\exists$ an undirected path U between some $x \in X, y \in Y$ such that for every collider C on U
  - Either C or a descendent of C is in Z
  - No non-collider on U is in Z
- Otherwise X and Y are d-separated by Z
- If Z d-separates X and Y, then $X \perp Y | Z$ for all distributions represented by the graph

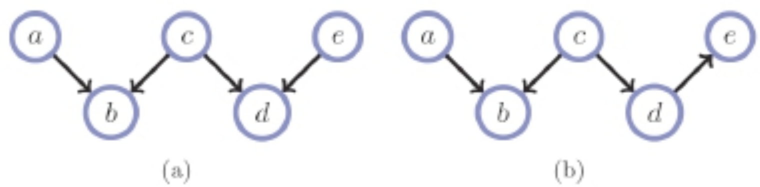
Conditional Independence (Contd.)



Examples

- For (a), a $\perp$ e|b; but a,e are dependent given {b,d}

Conditional Independence (Contd.)



Examples

- For (a), a $\perp$ e|b; but a,e are dependent given {b,d}
- For (b) a and e are dependent given b; c and e are unconditionally dependent

Constructing Bayesian networks

Choose an ordering of variables $X_1,\ldots,X_n$
For i = 1 to n
  Add X_i to the network
  Select parents from $X_1,\ldots,X_{i-1}$ such that

$$P(X_i|Parents(X_i)) = P(X_i|X_1,\ldots,X_{i-1})$$

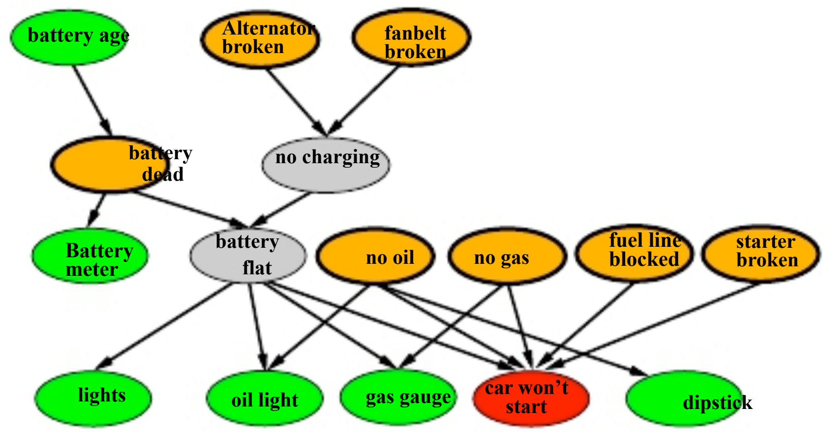
This choice of parents guarantees global semantics

$$P(X_1,\ldots,X_n) = \sum_{i=1}^{n} P(X_i|X_1,\ldots,X_{i-1})$$

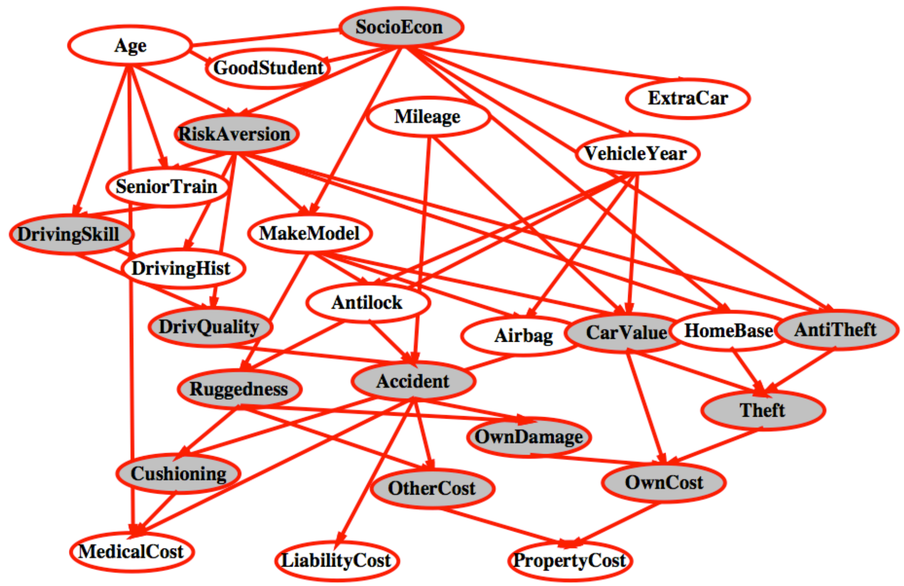
$$= \sum_{i=1}^{n} P(X_i|Parents(X_i))$$

Example: Car diagnosis

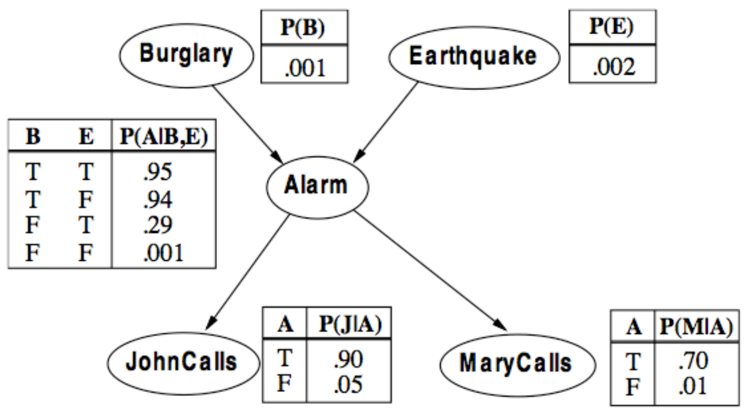
Initial evidence: car won't start
Testable variables (green), "broken, so fix it" variables (orange)
Hidden variables (gray) ensure sparse structure, reduce parameters



Example: Car insurance



Inference



How can we compute P(b|j,¬m)?