The stratum (layer) method for constructing Bayesian networks

Let V be a finite set of finite propositional variables, (Ω, F, P) be their joint probability distribution, and $G = (V,E)$ be a dag.
For each $v \in V$, let $c(v)$ be the set of all parents of v and $d(v)$ be the set of all descendents of v. Furthermore, for $v \in V$, let $a(v)$ be $V \setminus \{d(v) \cup \{v\}\}$, i.e., the set of propositional variables in V excluding v and v's descendents. Suppose for every subset $W \subseteq a(v)$, W and v are conditionally independent given c(v); that is, if $P(c(v)) > 0$, then

$$P(v \mid c(v)) = 0 \text{ or } P(W \mid c(v)) = 0 \text{ or } P(v \mid W \cup c(v)) = P(v \mid c(v)).$$

Then, $C = (V, E, P)$ is called a *Bayesian network* [Neapolitan, 1990].

The stratum method is based on the local Markov condition, which is the core of Neapolitan's definition of Bayesian networks.

The structure method [Russell & Norvig, Ch. 14]

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

Here, select a minimal set of parents

This choice of parents guarantees the global semantics:

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

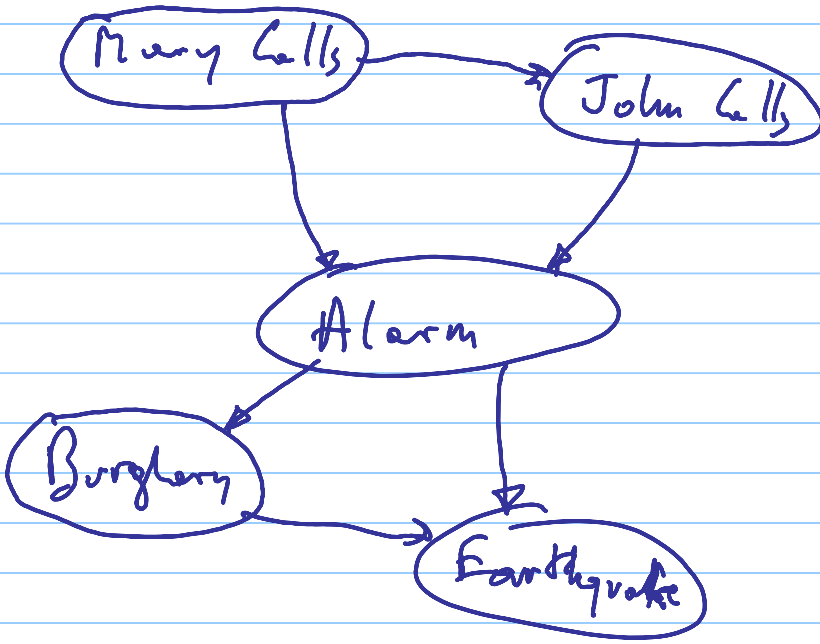
An example [Russell & Norvig, Ch. 14]

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

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



This is not a causal network

1. $P(\text{John Calls}) \stackrel{?}{=} P(J \mid \text{Mary Calls})$ No
2. $P(\text{Alarm}) \stackrel{?}{=} P(\text{Alarm} \mid M, J)$ No

$P(A \mid M) \stackrel{?}{=} P(A \mid M, J)$ No

$P(A \mid J) \stackrel{?}{=} P(A \mid M, J)$ No

3. $P(B) \stackrel{?}{=} P(B \mid M, J, A)$ No

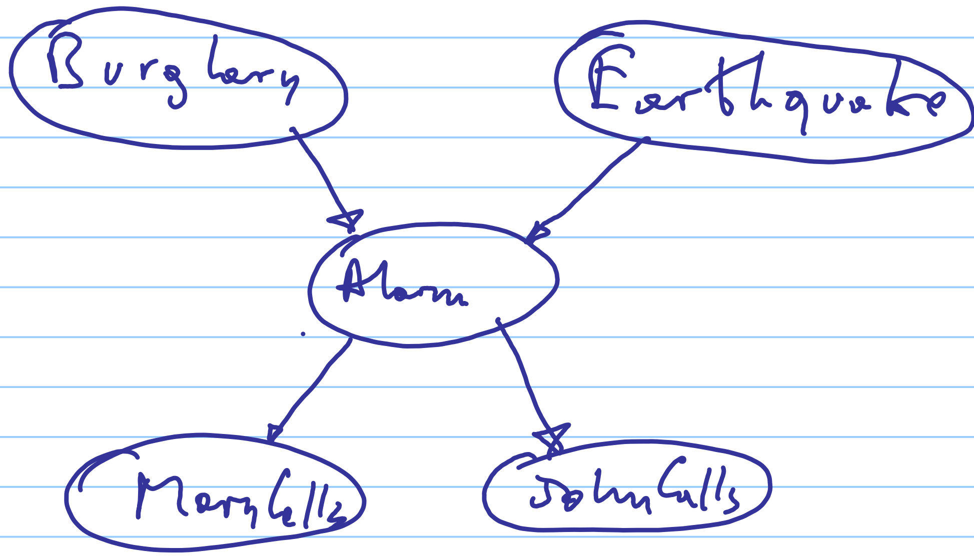
$P(B \mid A) \stackrel{?}{=} P(B \mid A, M, J)$ Yes

4. $P(E) \stackrel{?}{=} P(E \mid A, J, A, B)$ No

$P(E \mid A) \stackrel{?}{=} P(E \mid A, B, M, J)$ No

$P(E \mid A, B) \stackrel{?}{=} P(E \mid A, B, M, J)$ Yes

Suppose we instead choose the ordering $\langle \underline{B}, \underline{E}, \underline{A}, \underline{M}, \underline{J} \rangle$



1. $P(E) \overset{?}{=} P(E|B)$ Yes
2. $P(A) \overset{?}{=} P(A|B,E)$ No

   $P(A|B) \overset{?}{=} P(A|B,E)$ No

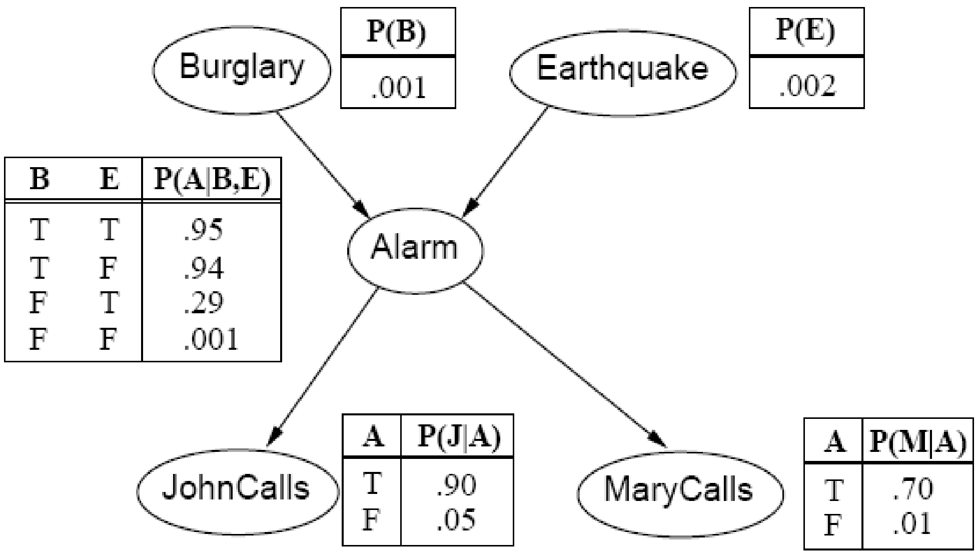
   $P(A|E) \overset{?}{=} P(A|B,E)$ No
3. $P(M) \overset{?}{=} P(M|B,A,E)$ No

   $P(M|A) \overset{?}{=} P(M|A)$ Yes
4. $P(J) \overset{?}{=} P(J|A,M,B,E)$ No

   $P(J|A) \overset{?}{=} P(J|A)$ Yes

This is a causal network

It is an empirical observation that BNs whose DAGs are causal are smaller.



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

A possible choice of CPTs, to complete the BN.