

Remember; test on Wednesday! 50 minutes, closed book.

The ~~direct~~ method for constructing Bayesian networks

Let  $V$  be a finite set of finite propositional variables,  $(\Omega, F, P)$  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 descendants 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 descendants. 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

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

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

- based on this definition

## The 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  $X_1, \dots, X_n$
2. For  $i = 1$  to  $n$   
add  $X_i$  to the network  
select parents from  $X_1, \dots, X_{i-1}$  such that  
$$P(X_i | Parents(X_i)) = P(X_i | X_1, \dots, X_{i-1})$$

This choice of parents guarantees the global semantics:

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

## An example [Russell & Norvig, Ch. 14]

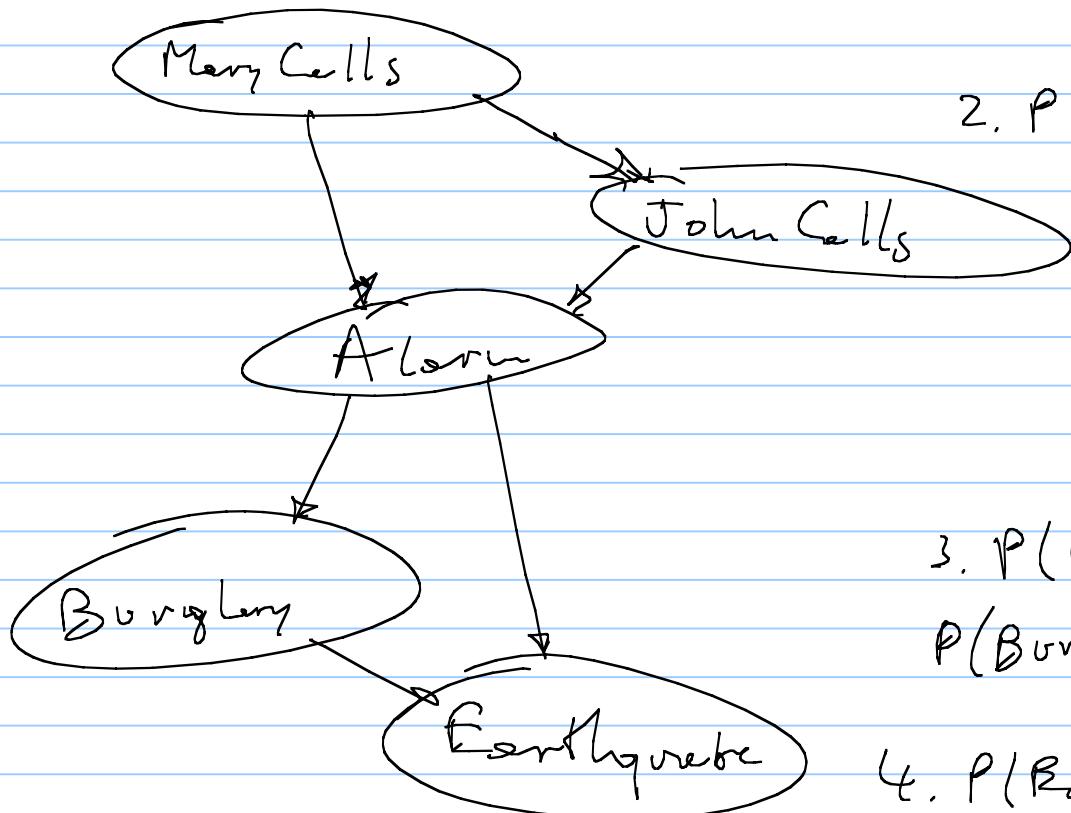
I'm at work, neighbor John calls to say my alarm is ringing, but neighbor Mary doesn't call. Sometimes it's set off by minor earthquakes. Is there a burglar?

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  $M, J, A, B, E$



$$1. P(\text{John Calls}) = P(\text{John Calls} | \text{Many Cells}) ?$$

$$2. P(\text{Alarm}) = P(\text{Alarm} | \text{Many Cells}, \text{John Calls}) ?$$

$$P(\text{Alarm} | \text{Many Cells}) =$$

$$P(\text{Alarm} | \text{", John Calls}) ? \text{No}$$

$$P(\text{Alarm} | \text{John Calls}) =$$

$$P(\text{Alarm} | \text{", Many Cells}) ? \text{No}$$

$$3. P(\text{Burglary} | \text{Alarm}) =$$

$$P(\text{Burglary} | \text{Alarm}, \text{Many Cells},$$

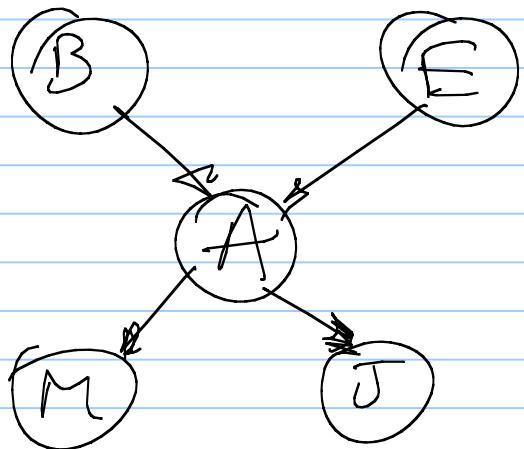
$$\text{John Calls}) ? \text{Yes}$$

$$4. P(\text{Earthquake} | \text{Alarm}) \neq$$

$$P(\text{Earthquake} | M, J, A, B) =$$

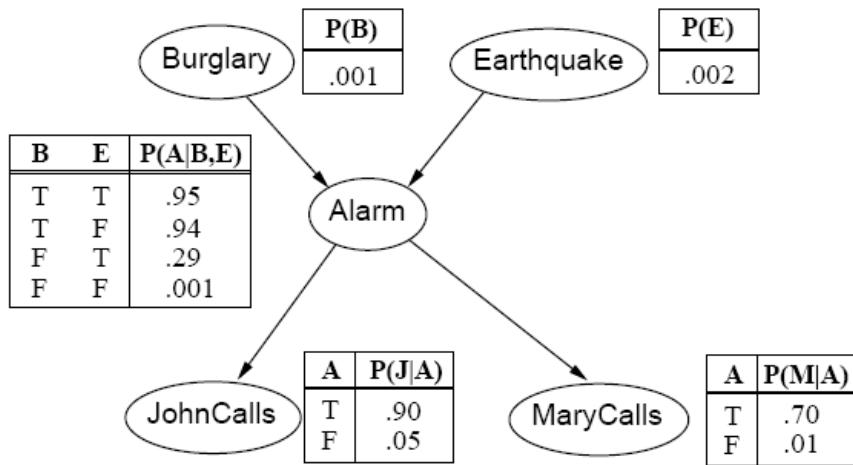
$$= P(\text{Earthquake} | A, B).$$

Choose instead  $\langle B, E, A, M, J \rangle$ .

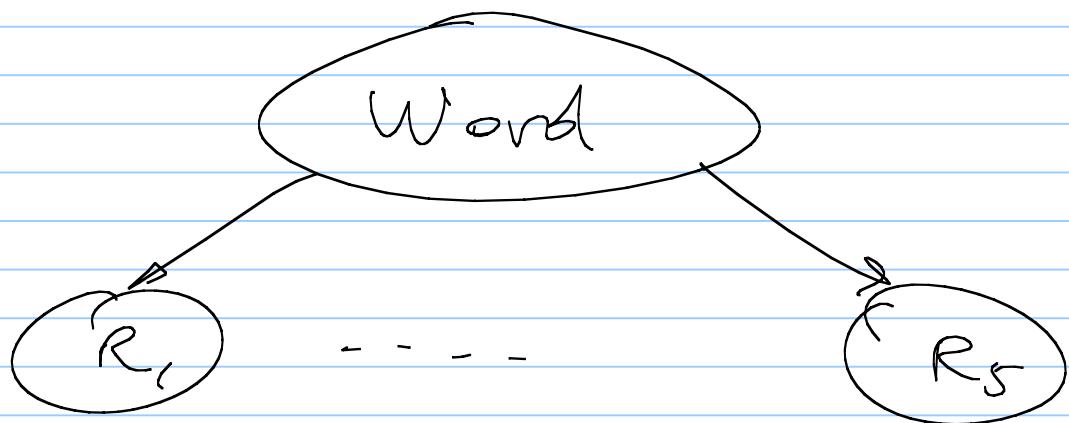


An order in which the edges are directed causally  
always results in a sparser network.

• Empirical observation



Result with CPTs,



Bad, b/c the state space of word is too large