Bayesian Networks Inference
Chapter 14

Mausam

(Slides by UW-AI faculty, Stuart Russell & David Page)

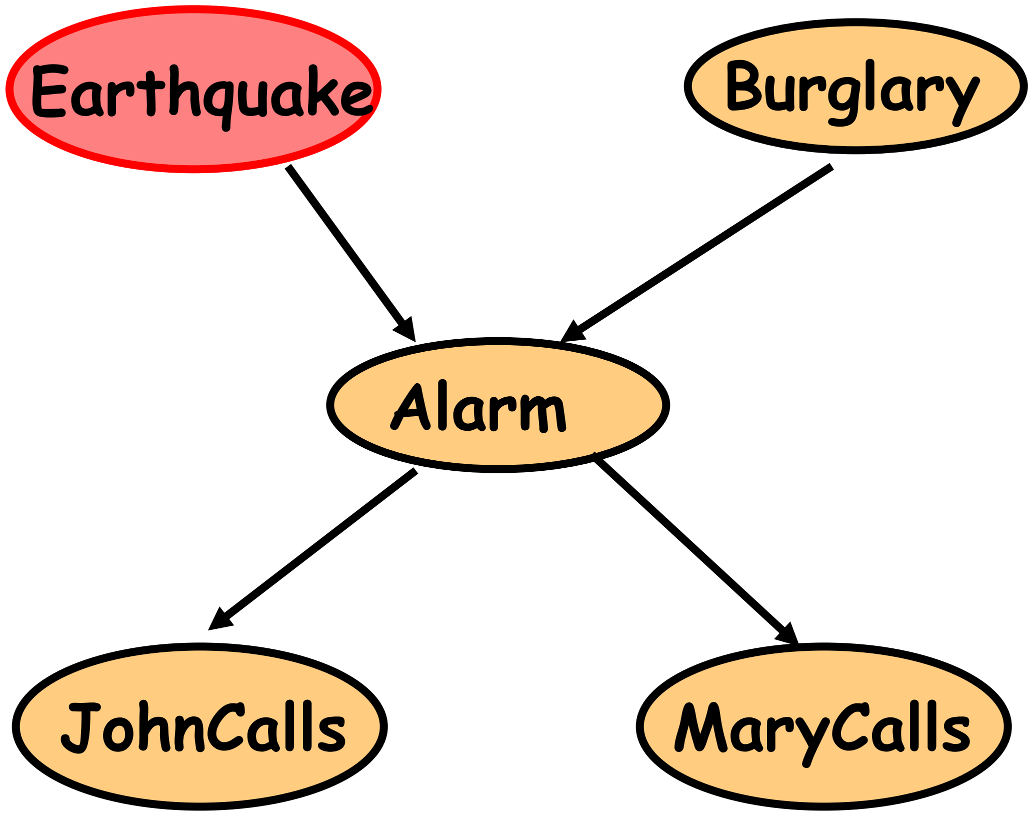
Bayesian Nets: Executive Summary

- Representing full joint distribution is untractable
- Factor using a directed graphical model
- $$P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid Par(X_i))$$
- *X* is conditionally independent of all non-descendents given its parents
- X is cond. independent of all nodes given Markov Blanket
  – MB: parents, children, parents of children
- D-separation: general conditional independence
- Exact Inference: Variable Elimination

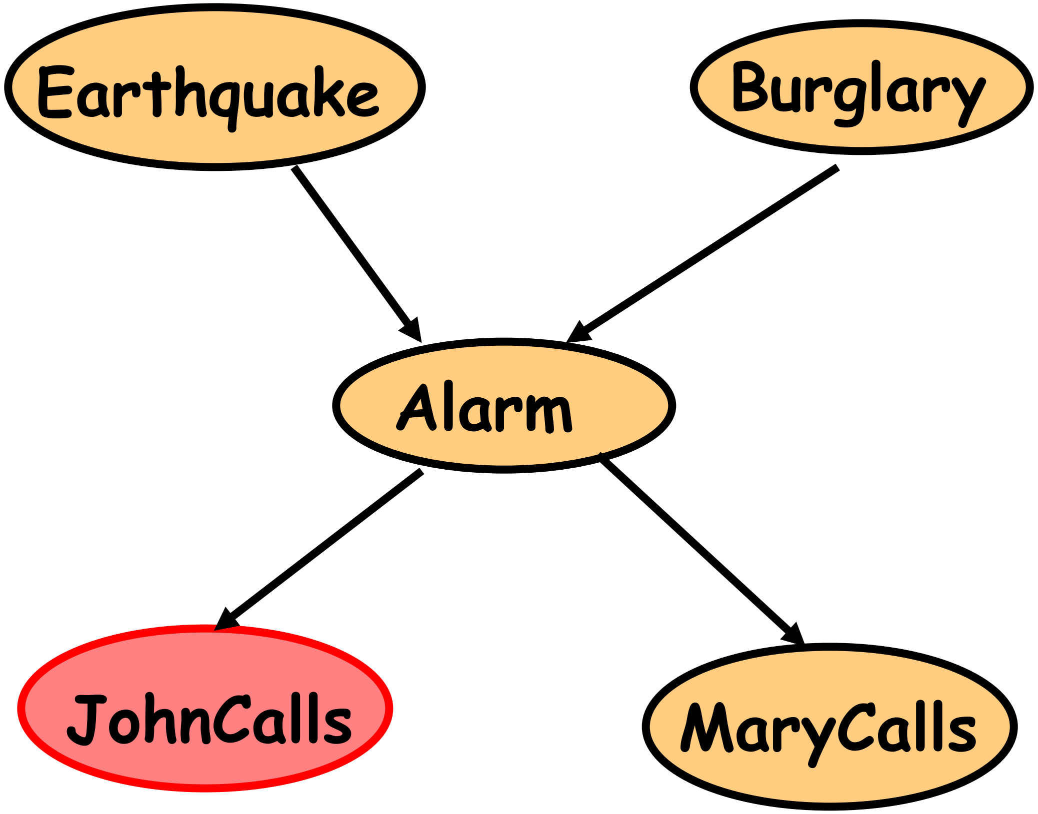
Inference in BNs

- The graphical independence representation
  - yields efficient inference schemes
- We generally want to compute
  - Marginal probability: *Pr(Z),*
  - *Pr(Z|**E**)* where ***E*** is (conjunctive) evidence
    - Z: query variable(s),
    - E: evidence variable(s)
    - everything else: hidden variable
- Computations organized by network topology

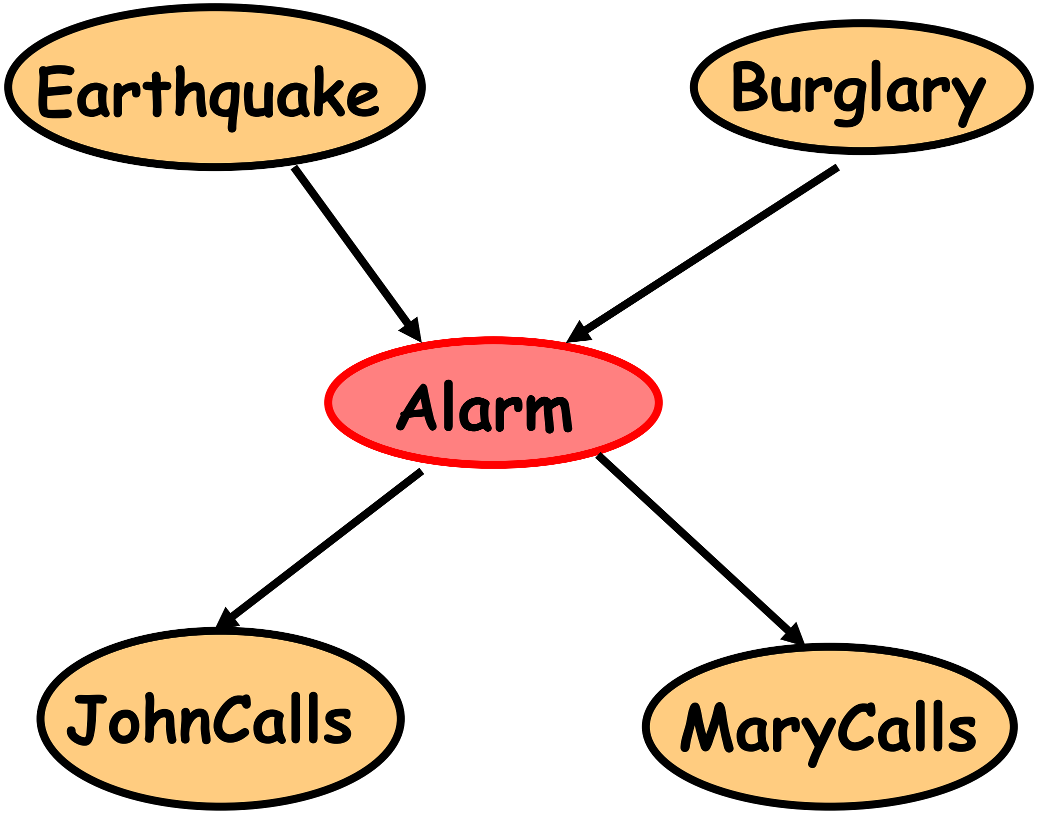
Causal Reasoning: P(j|e)



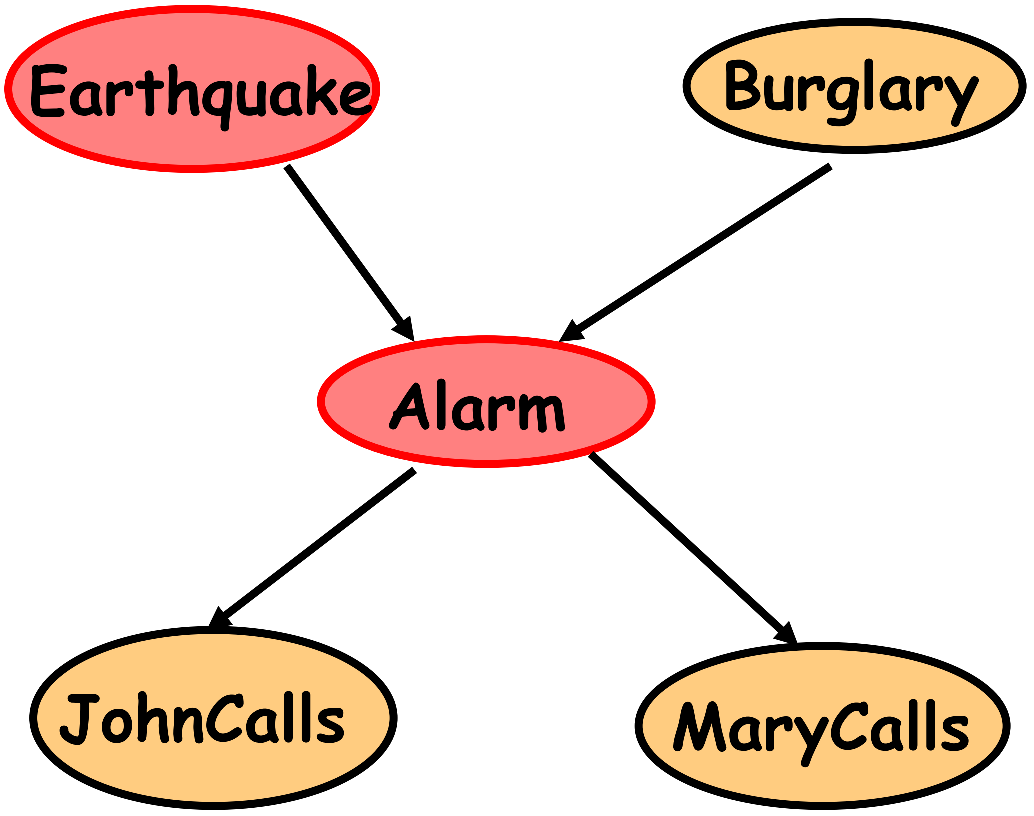
Evidential Reasoning: P(b|j)



Intercausal Reasoning: P(e|a) vs P(b|a)



Intercausal Reasoning: P(b|a,e)



Variable Elimination

- A *factor* is a function from some set of variables into a specific value: e.g., *f(E,A,N1)*
  - CPTs are factors, e.g., *P(A|E,B)* function of *A,E,B*
- VE works by *eliminating* all variables in turn until there is a factor with only query variable
- To eliminate a variable:
  - *join* all factors containing that variable (like DB)
  - *sum out* the influence of the variable on new factor
  - exploits product form of joint distribution

Inference Step 1: Marginal Probability

Rewrite in terms of joint distribution

- Fix the query variables
- Sum over unknown variables

Examples

$$P(j) = \Sigma_{M,A,B,E}\, \mathbf{P(j,M,A,B,E)}$$

$$P(j,\sim m) = \Sigma_{A,B,E}\, \mathbf{P(j,\sim m,A,B,E)}$$

Inference Step 1: Conditional Probability

Rewrite in terms of joint distribution

- Fix the query variables
- Fix the evidence variables
- Sum over unknown variables
- Add a normalization constant

Example:

$$P(b|j,m) = P(b,j,m)/P(j,m) = \alpha P(b,j,m)$$

$$= \alpha \Sigma_{A,E} P(b,j,m,A,E)$$

Compute α s.t.

$$P(b|j,m) + P(\sim b|j,m) = 1$$

Inference Step 2

- Rewrite joint probability using Bayes Net factors

- Examples
  - $\boldsymbol{\Sigma_{M,A,B,E}\, P(j,M,A,B,E)} = \Sigma_{\mathbf{M,A,B,E}} P(B)P(E)P(A|B,E)P(j|A)P(M|A)$
  - $\Sigma_{A,E}\, P(b,j,m,A,E) = \boldsymbol{\Sigma_{A,E}} P(b)P(E)P(A|b,E)P(j|A)P(m|A)$

Inference Step 3

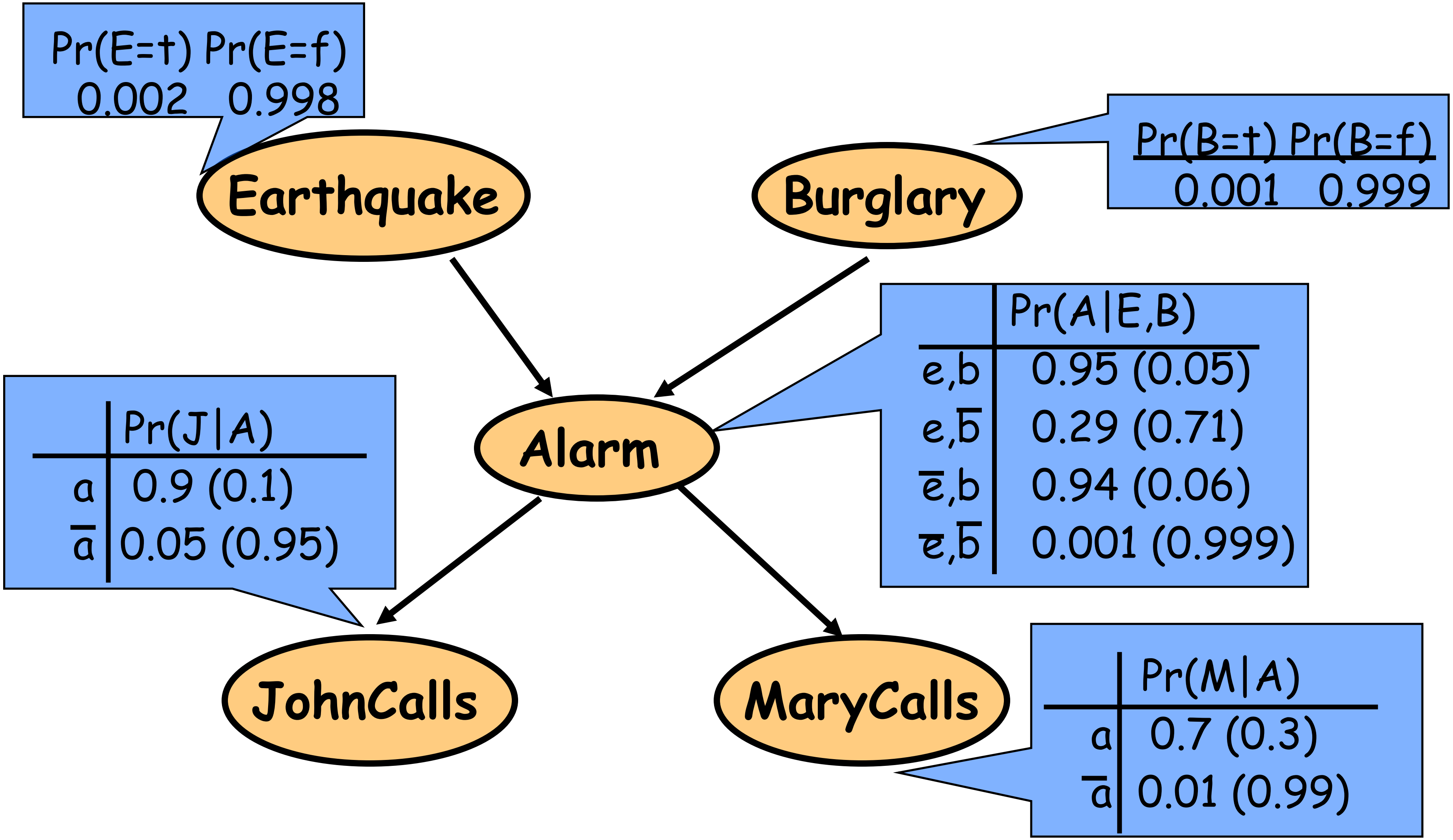
- Choose variable order
- Take summations inside

- Examples
  - $\Sigma_{M,A,B,E} P(B)P(E)P(A|B,E)P(j|A)P(M|A)$
  - Order: M, A, E, B
  - $\Sigma_B P(B)\ \Sigma_E P(E)\ \Sigma_A P(A|B,E)\ P(j|A)\ \Sigma_M P(M|A)$

Inference Step 3 contd

- Example
  - $\Sigma_{A,E} P(b)P(E)P(A|b,E)P(j|A)P(m|A)$
  - Choose order A, E
  - $P(b)\Sigma_E P(E)\ \Sigma_A P(A|b,E)P(j|A)P(m|A)$
  - Choose order E, A
  - $P(b)\Sigma_A P(j|A)P(m|A)\Sigma_E P(E)P(A|b,E)$

Burglars and Earthquakes



Pr(E=t)	Pr(E=f)
0.002	0.998

Pr(B=t)	Pr(B=f)
0.001	0.999

	Pr(A|E,B)
e,b	0.95 (0.05)
e,$\bar{b}$	0.29 (0.71)
$\bar{e}$,b	0.94 (0.06)
$\bar{e}$,$\bar{b}$	0.001 (0.999)

	Pr(J|A)
a	0.9 (0.1)
$\bar{a}$	0.05 (0.95)

	Pr(M|A)
a	0.7 (0.3)
$\bar{a}$	0.01 (0.99)

Detailed Computation

- $\Sigma_B P(B)\ \Sigma_E P(E)\ \Sigma_A P(A|B,E)\ P(j|A)\ \boxed{\cancel{\Sigma_M P(M|A)}}$

  $f_1(A)$

	Pr(M|A)
a	0.7 (0.3)
$\bar{a}$	0.01 (0.99)

Detailed Computation

- $\Sigma_B P(B)\ \Sigma_E P(E)\ \boxed{\Sigma_A P(A|B,E)\ P(j|A)}$

$f_2(B,E)$

	Pr(J|A)
a	0.9 (0.1)
$\bar{a}$	0.05 (0.95)

	Pr(A|E,B)
e,b	0.95 (0.05)
e,$\bar{b}$	0.29 (0.71)
$\bar{e}$,b	0.94 (0.06)
$\bar{e}$,$\bar{b}$	0.001 (0.999)

	f2(B,E)
e,b	0.95*0.9+ 0.05*0.1
e,$\bar{b}$	0.29*0.9+ 0.71*0.1
$\bar{e}$,b	0.94*0.9+ 0.06*0.1
$\bar{e}$,$\bar{b}$	0.001*0.9 + 0.999*0.1

Detailed Computation

- $\Sigma_B P(B)\ \Sigma_E P(E)\ \boxed{\Sigma_A P(A|B,E)\ P(j|A)}$

$f_2(B,E)$

	Pr(J|A)
a	0.9 (0.1)
$\bar{a}$	0.05 (0.95)

	Pr(A|E,B)
e,b	0.95 (0.05)
e,$\bar{b}$	0.29 (0.71)
$\bar{e}$,b	0.94 (0.06)
$\bar{e}$,$\bar{b}$	0.001 (0.999)

	f2(B,E)
e,b	0.86
e,$\bar{b}$	0.332
$\bar{e}$,b	0.852
$\bar{e}$,$\bar{b}$	0.1008

Detailed Computation

- $\Sigma_B P(B) \boxed{\Sigma_E P(E) f_2(B,E)}$

$\quad\quad\quad$ ***$f_3(B)$***

Pr(E=t)	Pr(E=f)
0.002	0.998

	f2(B,E)
e,b	0.86
e,$\bar{b}$	0.332
$\bar{e}$,b	0.852
$\bar{e}$,$\bar{b}$	0.1008

	f3(B)
b	0.002*0.86+0.998*0.852
$\bar{b}$	0.002*0.332+0.998*0.1008

Detailed Computation

- $\Sigma_B P(B) \boxed{\Sigma_E P(E) f_2(B,E)}$

$\underbrace{}_{}$ ***$f_3(B)$***

Pr(E=t)	Pr(E=f)
0.002	0.998

	f2(B,E)
e,b	0.86
e,$\bar{b}$	0.332
$\bar{e}$,b	0.852
$\bar{e}$,$\bar{b}$	0.1008

	f3(B)
b	0.8517644
$\bar{b}$	0.1006653312

Detailed Computation

- $\Sigma_B P(B)\, f_3(B)$

	f3(B)
b	0.8517644
$\bar{b}$	0.1006653312

Pr(B=t)	Pr(B=f)
0.001	0.999

*0.8517644*0.001*
*+ 0.1006653312*0.999*
= 0.10…

Notes on VE

- Each operation is a simple multiplication of factors and summing out a variable
- Complexity determined by size of largest factor
  - in our example, 3 vars (not 5)
  - linear in number of vars,
  - exponential in largest factor elimination ordering greatly impacts factor size
  - optimal elimination orderings: NP-hard
  - heuristics, special structure (e.g., polytrees)
- Practically, inference is much more tractable using structure of this sort

Variable Elimination Heuristic

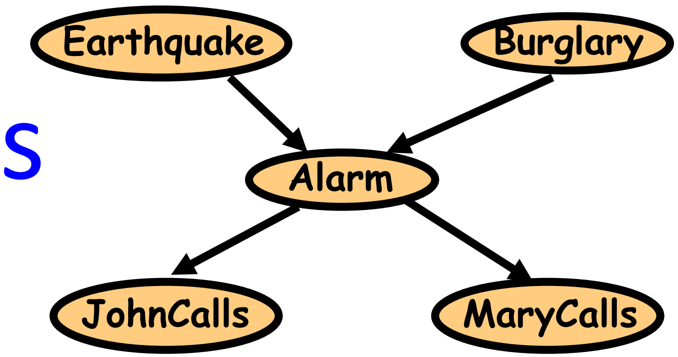
- Min-factor
  - Select var that results in smallest factor

- Other better heuristics exist…

Caution

- Multiplying too many probabilities in practice
- Underflow
  - Take logs and add

Irrelevant variables



$P(J)$

$= \Sigma_{M,A,B,E}\ P(J,M,A,B,E)$

$= \Sigma_{M,A,B,E}\ P(J|A)P(B)P(A|B,E)P(E)P(M|A)$

$= \Sigma_A P(J|A)\ \Sigma_B P(B)\ \Sigma_E P(A|B,E)P(E) \boxed{\Sigma_M P(M|A)}$

$= \Sigma_A P(J|A)\ \Sigma_B P(B)\ \Sigma_E P(A|B,E)P(E)$

$= \Sigma_A P(J|A)\ \Sigma_B P(B)\ f1(A,B)$

$= \Sigma_A P(J|A)\ f2(A)$

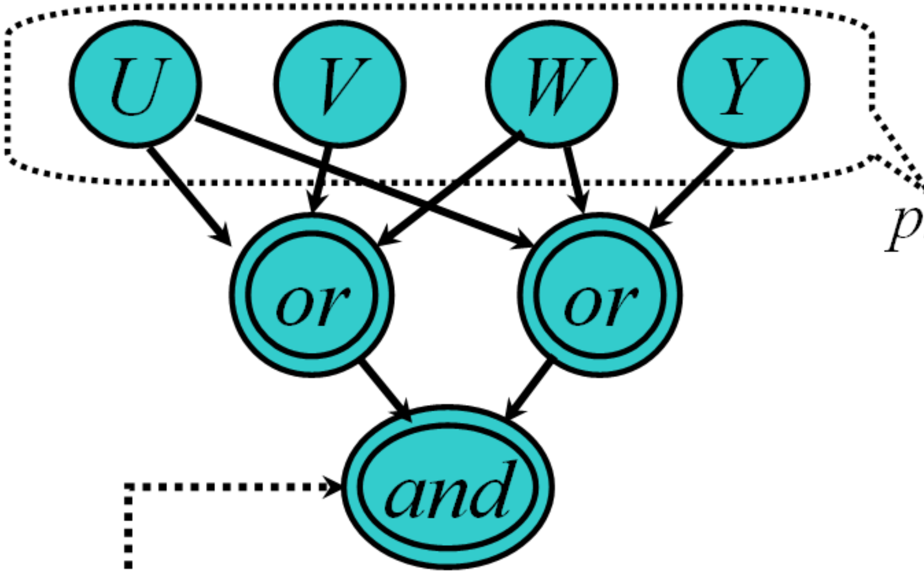
$= f3(J)$ M is irrelevant to the computation

Thm: Y is irrelevant unless Y ϵ Ancestors(Z U E)

Reducing 3-SAT to Bayes Nets

- **Theorem:** Inference in a multi-connected Bayesian network is NP-hard.

Boolean 3CNF formula $\phi = (u \vee \overline{v} \vee w) \wedge (\overline{u} \vee \overline{w} \vee y)$



Probability $(\,\cdot\,) = 1/2^n \cdot \#$ *satisfying assignments of* ϕ

Complexity of Exact Inference

- Exact inference is NP hard
  - 3-SAT to Bayes Net Inference
  - It can count no. of assignments for 3-SAT: #P complete
- Inference in tree-structured Bayesian network
  - Polynomial time
  - compare with inference in CSPs
- Approximate Inference
  - Sampling based techniques