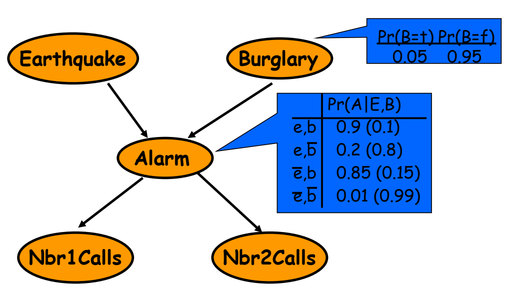
Bayesian Networks

CSE 473

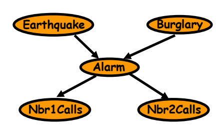
Bayes Nets

- In general, joint distribution *P* over set of variables ($X_1 \times \ldots \times X_n$) requires exponential space for representation & inference
- BNs provide a graphical representation of *conditional independence* relations in *P*
  - usually quite compact
  - requires assessment of fewer parameters, those being quite natural (e.g., causal)
  - efficient (usually) inference: query answering and belief update

An Example Bayes Net



Earthquake Example (cont'd)

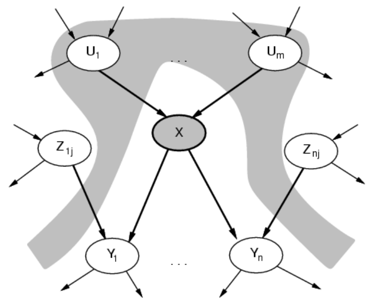


- **If we know *Alarm*, no other evidence influences our degree of belief in *Nbr1Calls***
  - $P(N1|N2,A,E,B) = P(N1|A)$
  - also: $P(N2|N1,A,E,B) = P(N2|A)$ and $P(E|B) = P(E)$
- **By the chain rule we have**
  - $P(N1,N2,A,E,B) = P(N1|N2,A,E,B) \cdot P(N2|A,E,B) \cdot P(A|E,B) \cdot P(E|B) \cdot P(B)$
  - $= P(N1|A) \cdot P(N2|A) \cdot P(A|B,E) \cdot P(E) \cdot P(B)$
- **Full joint requires only 10 parameters (cf. 32)**

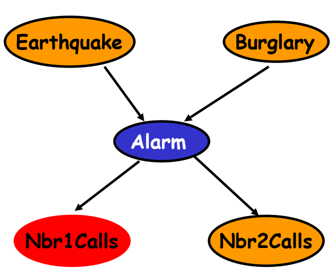
BNs: Qualitative Structure

- Graphical structure of BN reflects conditional independence among variables
- Each variable *X* is a node in the DAG
- Edges denote *direct probabilistic influence*
    - usually interpreted *causally*
    - parents of *X* are denoted *Par(X)*
- ***X* is conditionally independent of all nondescendents given its parents**
    - Graphical test exists for more general independence
    - "Markov Blanket"

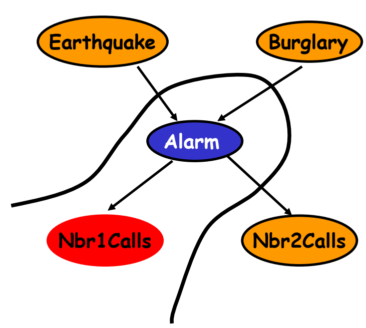
Given Parents, X is Independent of Non-Descendants



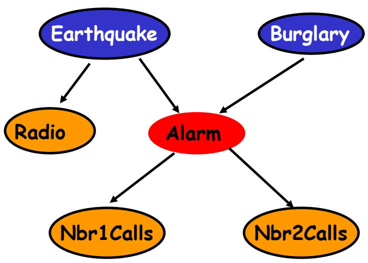
For Example



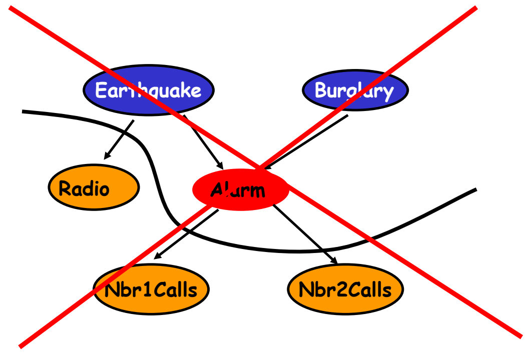
For Example



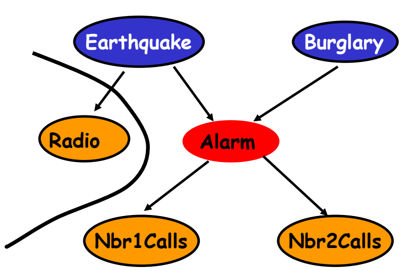
For Example



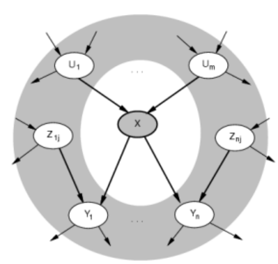
For Example



For Example

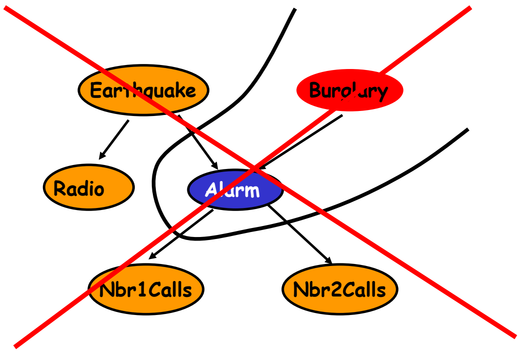


Given Markov Blanket, X is Independent of All Other Nodes

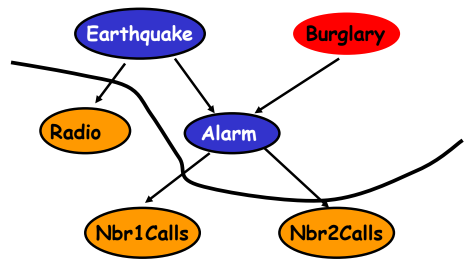


MB(X) = Par(X) ∪ Childs(X) ∪ Par(Childs(X))

For Example



For Example



Conditional Probability Tables

- For complete spec. of joint dist., *quantify* BN
- For each variable X, specify **CPT**: $P(X \mid Par(X))$
  number of params *locally* exponential in $|Par(X)|$
- If $X_1, X_2, \ldots X_n$ is any topological sort of the network, then we are assured:

$$P(X_n, X_{n-1}, \ldots X_1) = P(X_n \mid X_{n-1}, \ldots X_1) \cdot P(X_{n-1} \mid X_{n-2}, \ldots X_1) \ldots P(X_2 \mid X_1) \cdot P(X_1)$$

$$= P(X_n \mid Par(X_n)) \cdot P(X_{n-1} \mid Par(X_{n-1})) \ldots P(X_1)$$

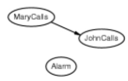
Bayes Net Construction Example

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

$P(J \mid M) = P(J)$?

Example

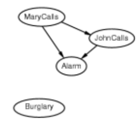
- Suppose we choose the ordering *M, J, A, B, E*
-



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

Example

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

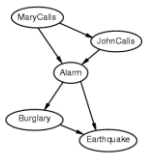
P(J | M) = P(J)?
No
P(A | J, M) = P(A | J)? P(A | J, M) = P(A)? **No**
P(B | A, J, M) = P(B | A)? **Yes**
P(B | A, J, M) = P(B)? **No**
P(E | B, A ,J, M) = P(E | A)?
P(E | B, A, J, M) = P(E | A, B)?

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, J, 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 contd.

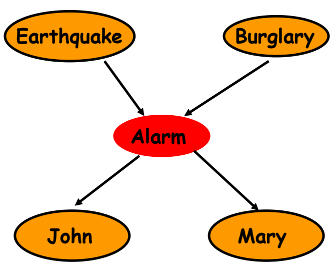


- Deciding conditional independence is hard in noncausal directions
- (Causal models and conditional independence seem hardwired for humans!)
- Network is less compact: 1 + 2 + 4 + 2 + 4 = 13 numbers needed

Inference in BNs

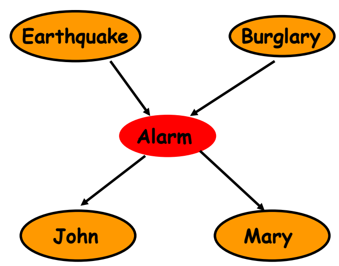
- The graphical independence representation
  yields efficient inference schemes
- We generally want to compute
  Pr(X), or
  *Pr(X|**E**)* where ***E*** is (conjunctive) evidence
- Computations organized by network topology
- One simple algorithm:
  variable elimination (VE)

P(B | J=true, M=true)



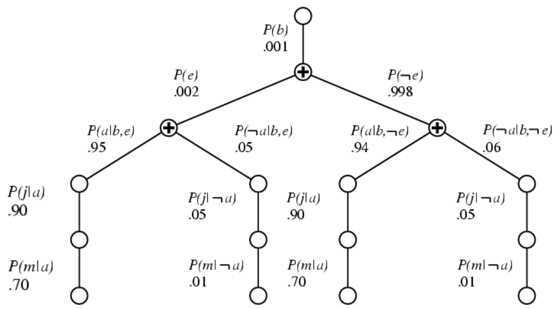
$$P(b|j,m) = \alpha \sum_{e,a} P(b,j,m,e,a)$$

P(B | J=true, M=true)



$$P(b|j,m) = \alpha P(b) \sum_{e} P(e) \sum_{a} P(a|b,e)P(j|a)P(m,a)$$

Structure of Computation



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

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

Example of VE: P(N1)

P(N1)

$= \Sigma_{N2,A,B,E}\ P(N1,N2,A,B,E)$

$= \Sigma_{N2,A,B,E}\ P(N1|A)P(N2|A)\ P(B)P(A|B,E)P(E)$

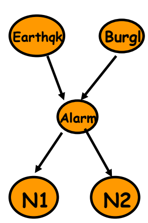
$= \Sigma_{A}P(N1|A)\ \Sigma_{N2}P(N2|A)\ \Sigma_{B}P(B)\ \Sigma_{E}P(A|B,E)P(E)$

$= \Sigma_{A}P(N1|A)\ \Sigma_{N2}P(N2|A)\ \Sigma_{B}P(B)\ f1(A,B)$

$= \Sigma_{A}P(N1|A)\ \Sigma_{N2}P(N2|A)\ f2(A)$

$= \Sigma_{A}P(N1|A)\ f3(A)$

$= f4(N1)$



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