Introduction to Artificial Intelligence

COMP307

Building a Bayesian Network

Yi Mei

yi.mei@ecs.vuw.ac.nz

Outline

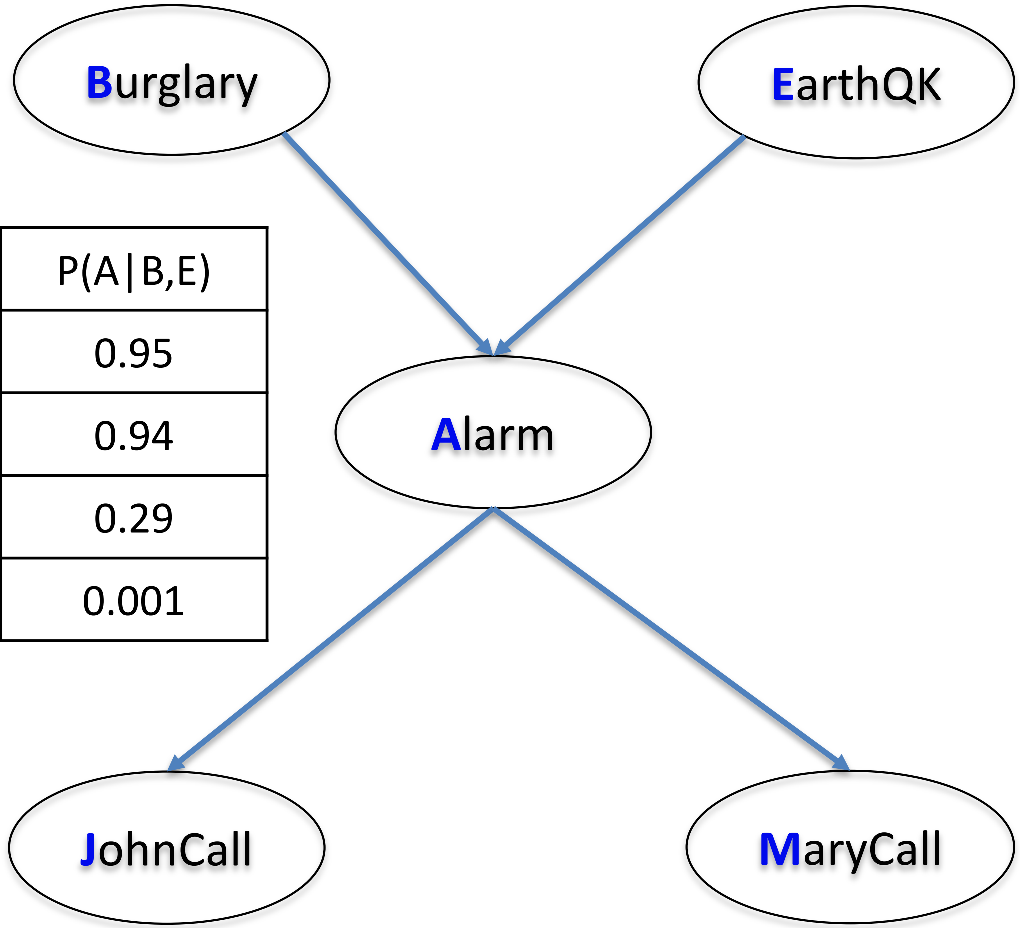
- Number of Free Parameters
- Building a BN
- Nodes Ordering and Compactness
- Summary

Alarm Network Revisited

- Do we need to store P(B = F), P(A = F | B = T, E = T)?
- How many probabilities need to be stored?

B	P(B)
T	0.001

E	P(E)
T	0.002



B	E	A	P(A|B,E)
T	T	T	0.95
T	F	T	0.94
F	T	T	0.29
F	F	T	0.001

A	J	P(J|A)
T	T	0.9
F	T	0.05

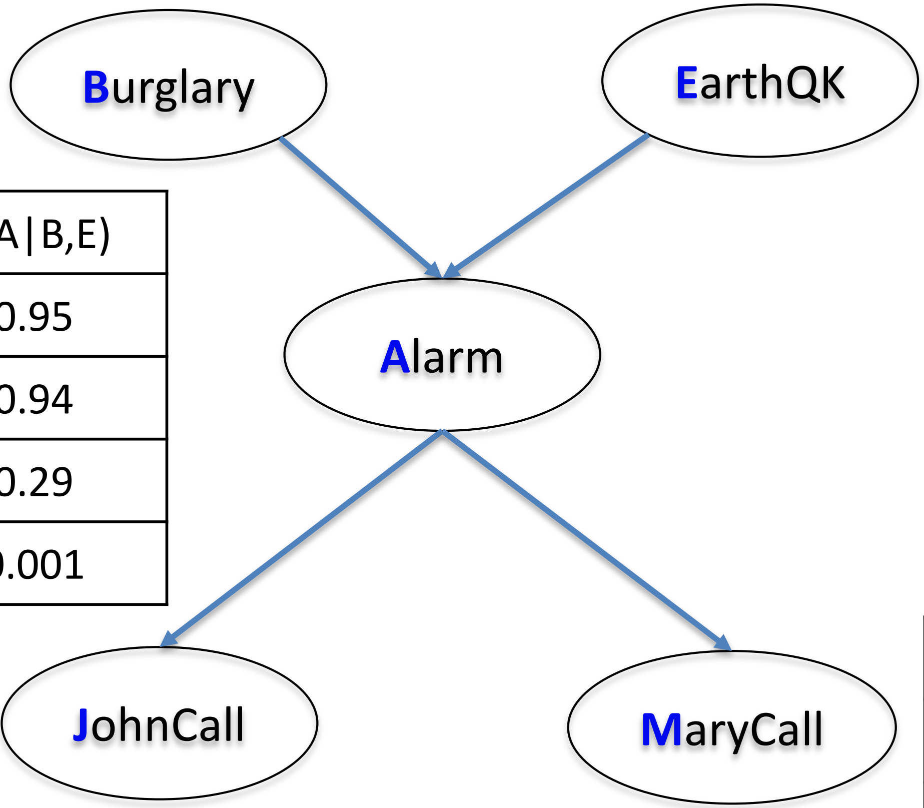
A	M	P(M|A)
T	T	0.7
F	T	0.01

Alarm Network Revisited

- CPT size: ignore the last possible value (can be derived)
- Number of free parameters in a model is the number of variables/probabilities that cannot be derived, but have to be estimated
  - Number of free parameters in the alarm network: 1+1+4+2+2=10

B	P(B)
T	0.001

E	P(E)
T	0.002



B	E	A	P(A|B,E)
T	T	T	0.95
T	F	T	0.94
F	T	T	0.29
F	F	T	0.001

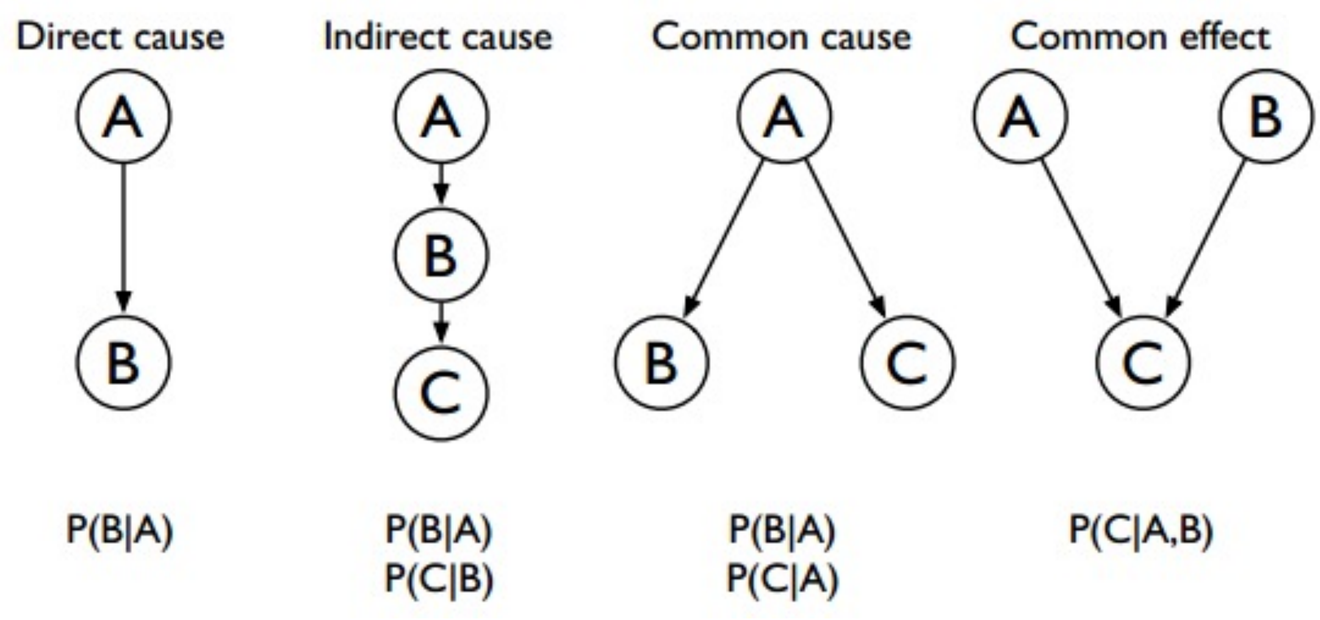
A	J	P(J|A)
T	T	0.9
F	T	0.05

A	M	P(M|A)
T	T	0.7
F	T	0.01

Number of Free Parameters

- Calculate the CPT size (number of free parameters) for the following
  - Assume: $|A| = 2, |B| = 2, |C| = 2$, they are all Boolean (binary) variables

- Example: direct cause
  - $|A| - 1 + |A| \times (|B| - 1) = 2 - 1 + 2 \times 1 = 3$
- Other cases?



Number of Free Parameters

- In general, for a Bayesian network with factorization

$$P(X_1, \dots, X_n) = P(X_1|parents(X_1)) * \cdots * P(X_n|parents(X_n))$$

- The number of free parameters of X_i is

Number of probs estimated for each condition

$$(|X_i| - 1) * \prod_{Y \in parents(X_i)} |Y|$$

Number of conditions

- A Bayesian network with smaller number of free parameters is desired because it
  - Requires less memory
  - More efficient to do reasoning (less variables involved for calculating posterior probabilities)

- When building a Bayesian network, we should minimise the number of parents of each variable $parents(X_i)$

Building a BN from Domain Knowledge

- 1. Identify a set of **random variables** that describe the problem, using domain knowledge.
- 2. Build the **directed acyclic graph**, i.e., the **directed links** between the random variables based on domain knowledge about the causal relationships between the variables.
- 3. Build the **conditional probability table** for each variable, by estimating the necessary probabilities using domain knowledge or historical data.

Steps 1 and 3 are directly from domain knowledge/data, here we discuss how to build the DAG for step 2.

Building the DAG of a BN

- Pearl's Network Construction Algorithm (A way):
  1. Choose an **order** for the variables
  2. While there are variables left
     a) add the next variable X_i to the network
     b) add arcs to the X_i node from a minimal set of nodes (parents) already in the network, such that the conditional independency property is satisfied: $P(X_i \mid X'_1, \dots, X'_m) = P\left(X_i \,\middle|\, Parents(X_i)\right)$, where $X'_1, \dots, X'_m$ are all the variables preceding X_i
     c) Define the conditional probability table for X_i

Steps 2b) requires to know the conditional independence between variables from domain knowledge

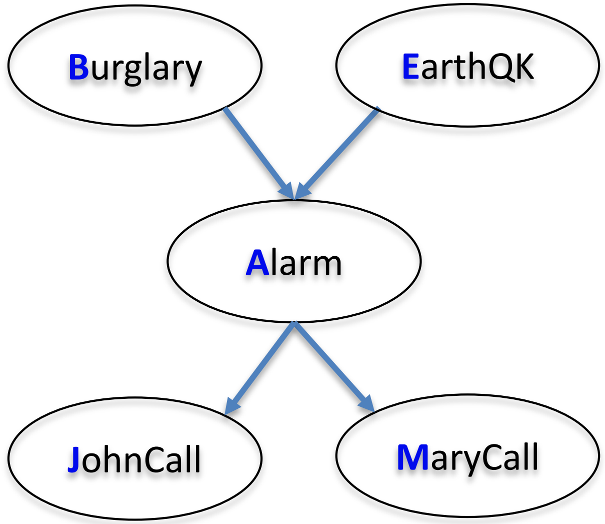
Compactness and Node Ordering

- Compactness:
  - The more compact the model is, the smaller the CPT size
  - Less computer memory, more computationally efficient
  - Over dense networks fail to represent independencies explicitly
  - Over dense networks fail to represent the causal dependencies in the domain

- The compactness depends on getting the node ordering "right." The optimal order is to add the root causes first, then the variable(s) they influence directly, and continue until leaves are reached.

Building BN

- Given the node order, how to add the links?
- Suppose we choose the order as B, E, A, J, M

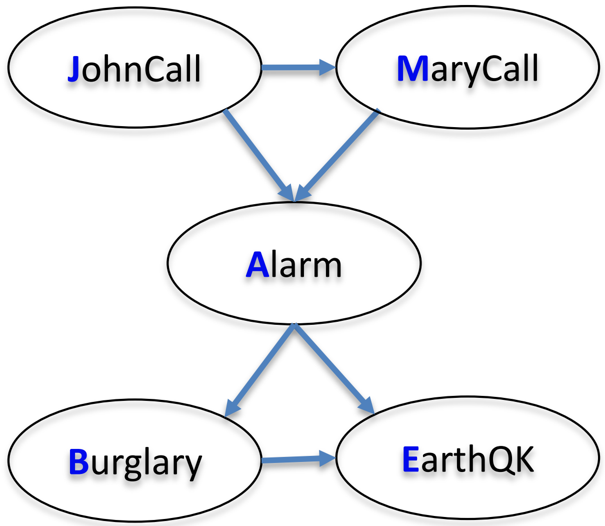


- **Step 1**: Add node B
- **Step 2**: Add node E
  - $P(E \mid B) = P(E)$? Yes, no link
- **Step 3**: Add node A
  - $P(A \mid B, E) = P(A)$? No
  - $P(A \mid B, E) = P(A \mid B)$? No
  - $P(A \mid B, E) = P(A \mid E)$? No, B -> A and E -> A
- **Step 4**: Add node J
  - $P(J \mid B, E, A) = P(J \mid A)$? Yes, A -> J, no link from B or E to J
- **Step 5**: Add node E
  - $P(M \mid B, E, A, J) = P(A)$? Yes, A -> M, no link from B, E or J to M

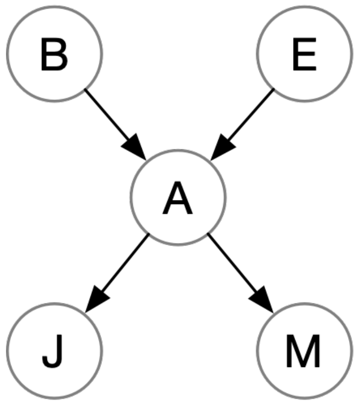
Building BN

- Given the node order, how to add the links?
- Suppose we choose the order as J, M, A, B, E

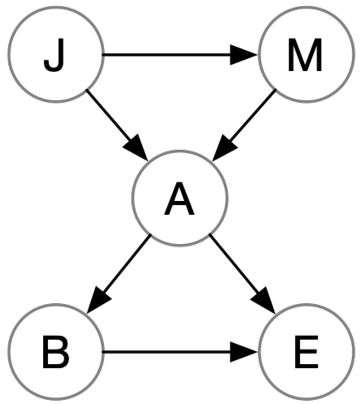
- **Step 1**: Add node J
- **Step 2**: Add node M
  - P(M I J) = P(M)? No, J -> M
- **Step 3**: Add node A
  - P(A I M, J) = P(A)? No
  - P(A I M, J) = P(A I J)? No
  - P(A I M, J) = P(A I M)? No, M -> A and J -> A
- **Step 4**: Add node B
  - P(B I M, J, A) = P(B)? No
  - P(B I M, J, A) = P(B I A)? Yes, A -> B, no link from M or J to B
- **Step 5**: Add node E
  - P(E I M, J, A, B) = P(E)? No
  - P(E I M, J, A, B) = P(E I A)? No
  - P(E I M, J, A, B) = P(E I B)? No
  - P(E I M, J, A, B) = P(E I A, B)? Yes, A -> E, B -> E, no other link



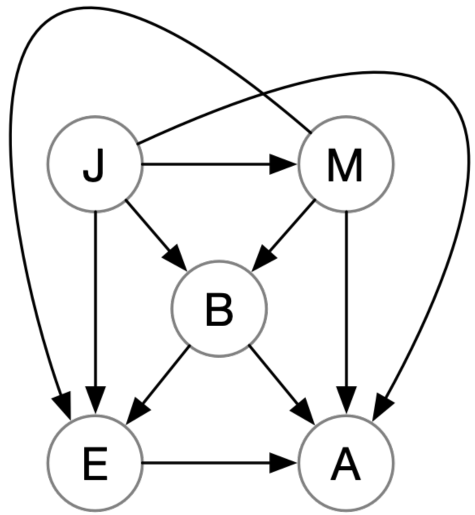
Ordering and Compactness



B -> E -> A -> J -> M



J -> M -> A -> B -> E



J -> M -> B -> E -> A

Are they essentially the same?

How many free parameters in each BN?

Summary

- Building Bayesian network
  - Minimise the conditional dependency table size
- Order of nodes make difference
- Usually put cause first, and then effects
- Make fewer parents (links)