Introduction to Artificial Intelligence

COMP307

Uncertainty and Probability 4: Building a Bayesian Network

Yi Mei

yi.mei@ecs.vuw.ac.nz

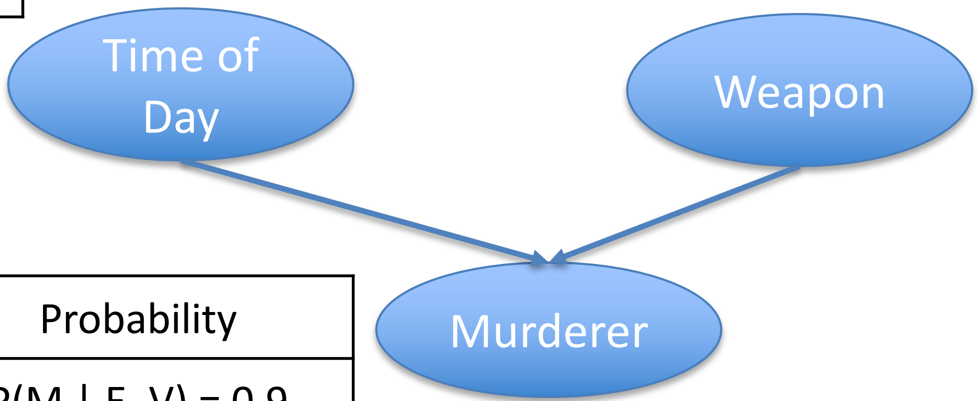
Outline

- A Bayesian Network Example
- Conditional Probability Table Size
- Building a BN
- Nodes Ordering and Compactness
- Summary

Bayesian Network for Lazy Detective

Time	Probability
E	P(T = E) = 0.05
Midnight	P(T = M) = 0.95

Weapon	Probability
V	P(W = V) = 0.8
S	P(W = S) = 0.2



T	W	Murderer	Probability
E	V	Maid	P(M | E, V) = 0.9
E	S	Maid	P(M | E, S) = 0.55
M	V	Maid	P(M | M, V) = 0.35
M	S	Maid	P(M | M, S) = 0.05
E	V	Butler	P(B | E, V) = 0.1
E	S	Butler	P(B | E, S) = 0.45
M	V	Butler	P(B | M, V) = 0.65
M	S	Butler	P(B | M, S) = 0.95

Use normalization rule to save space

Bayesian Network for Lazy Detective

- Simplified CPT: ignore the last possible value (can be derived)
- Number of free parameters in a model is the number of variables that cannot be derived, but have to be estimated

Number of free parameters in the following BN: 1+1+4=6

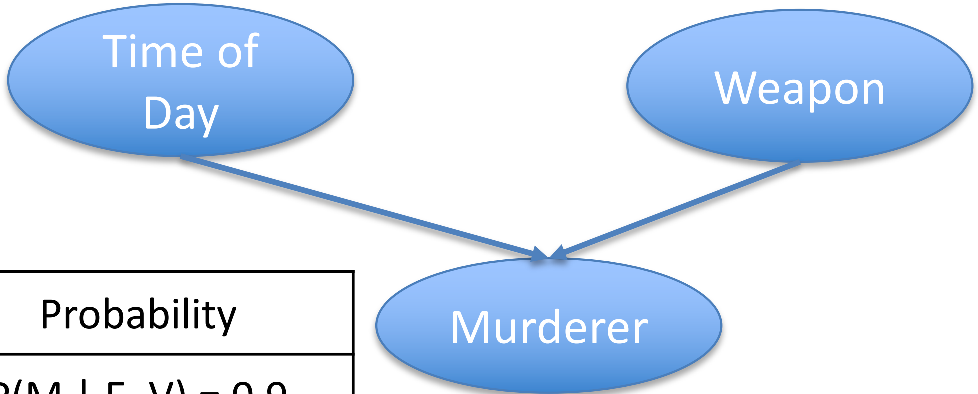
* 简化 CPT: 忽略最后一个可能的值(可以派生)
* 模型中自由参数的数量是不能派生但必须估计的变量的数量
 -下面的贝叶斯网络图中的Free Parameter数量是1+1+4=6

Weapon	Probability
V	P(W = V) = 0.8

$P(W = S) = 1 – P(W = V)$

Time	Probability
E	P(T = E) = 0.05

$P(T = M) = 1 – P(T = E)$



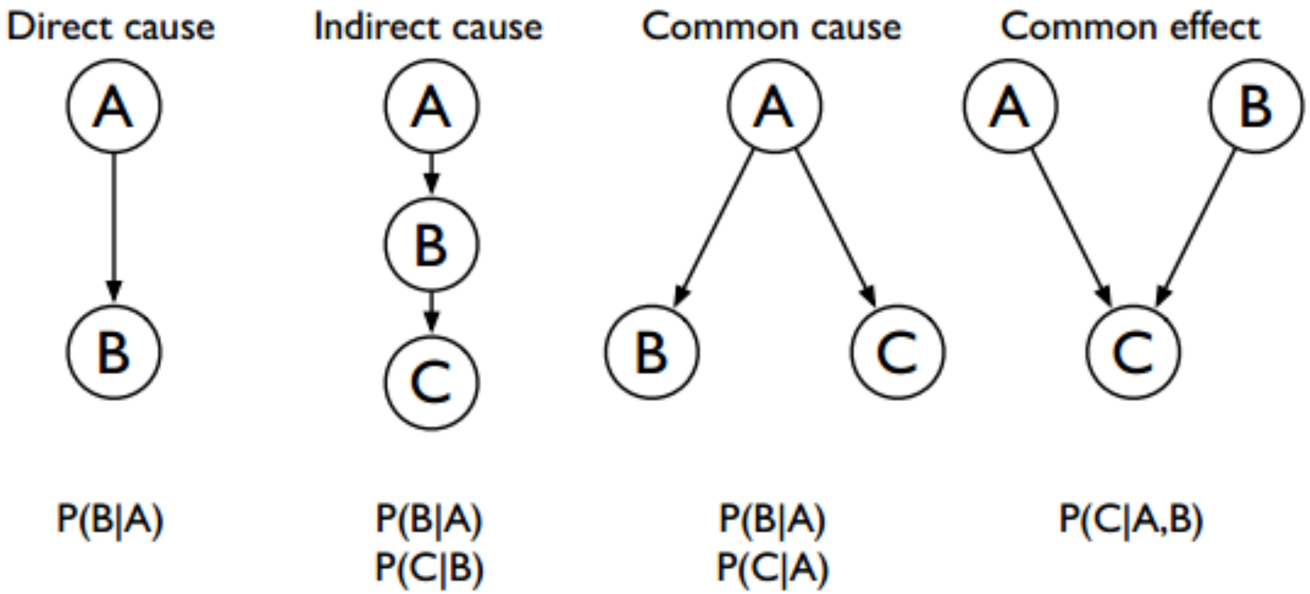
T	W	Murderer	Probability
E	V	Maid	P(M | E, V) = 0.9
E	S	Maid	P(M | E, S) = 0.55
M	V	Maid	P(M | M, V) = 0.35
M	S	Maid	P(M | M, S) = 0.05

$P(B \mid E, V) = 1 – P(M \mid E, V)$
$P(B \mid E, S) = 1 – P(M \mid E, S)$
$P(B \mid M, V) = 1 – P(M \mid M, V)$
$P(B \mid M, S) = 1 – P(M \mid M, S)$

Number of Free Parameters

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

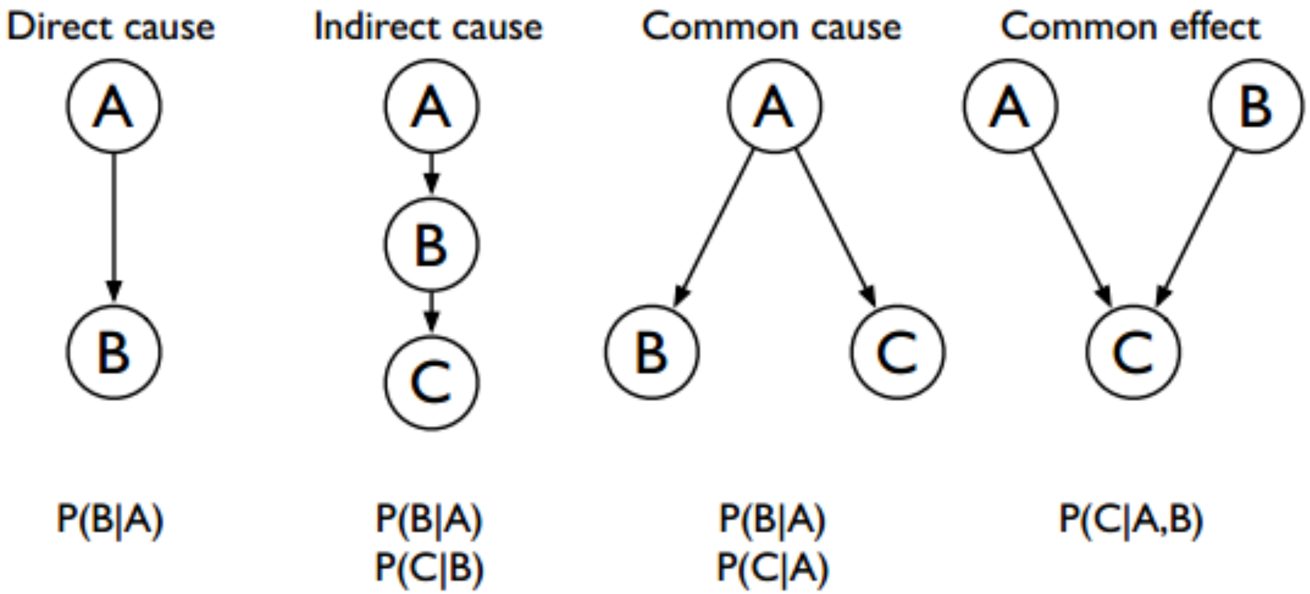
- Example: direct cause
  - $|A| - 1 + |A|\times(|B| - 1) = 2 - 1 + 2\times1 = 3$
- Other cases?



Number of Free Parameters

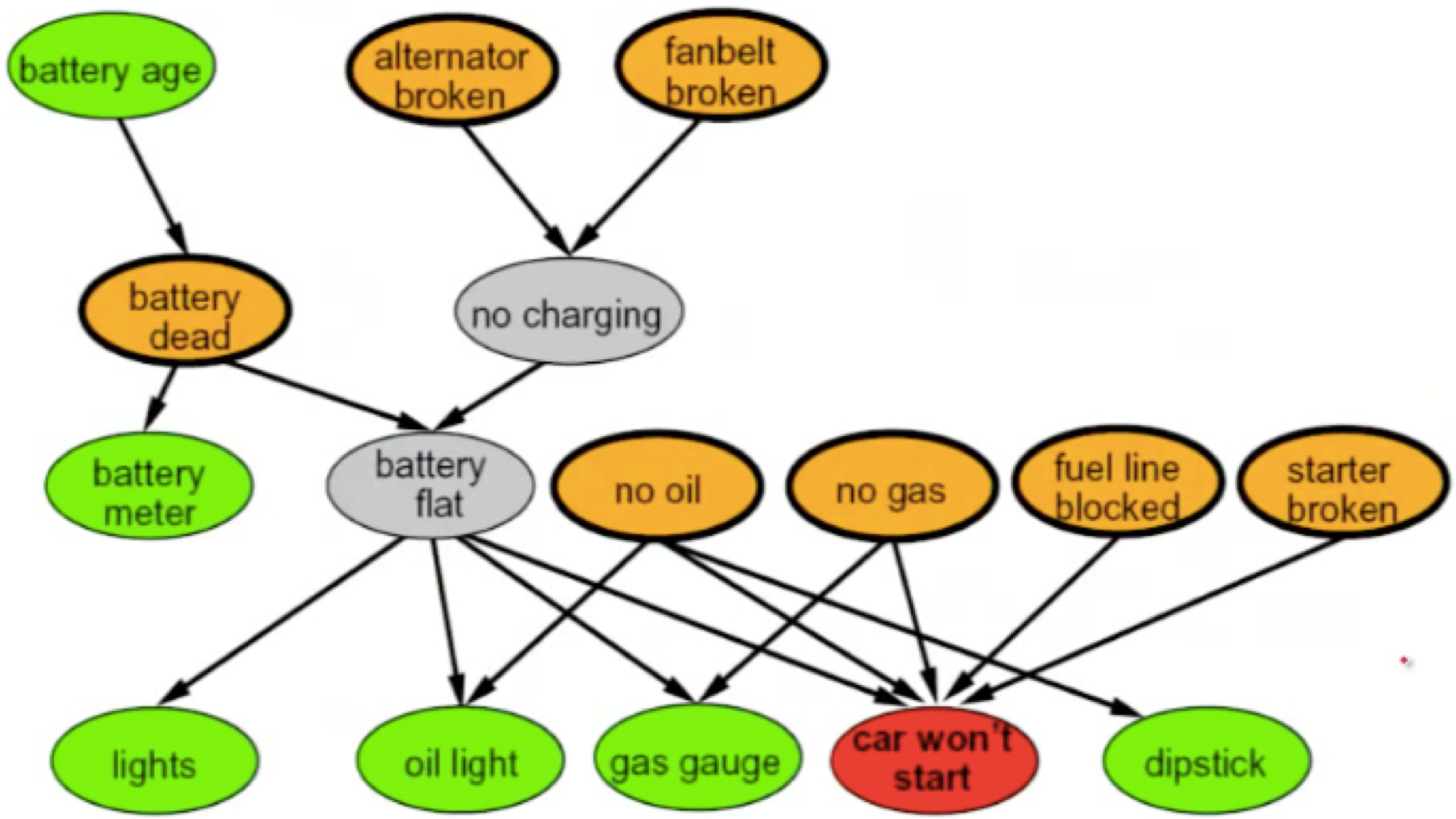
- Try calculate the CPT size (number of free parameters) for the following
  - Assume: $|A| = 2, |B| = 2, |C| = 2$, they are all Boolean (binary) variables

- Example: direct cause
  - $|A| - 1 + |A| \times (|B| - 1) = 2 - 1 + 2 \times 1 = 3$
- Other cases?
  - Indirect cause: $|A| - 1 + |A|(|B| - 1) + |B|(|C| - 1) = 2 - 1 + 2 \times 1 + 2 \times 1 = 5$
  - Common cause: $|A| - 1 + |A|(|B| - 1) + |A|(|C| - 1) = 2 - 1 + 2 \times 1 + 2 \times 1 = 5$
  - Common effect: $|A| - 1 + |B| - 1 + |A||B|(|C| - 1) = 2 - 1 + 2 - 1 + 2 \times 2 \times 1 = 6$



Large BN Example

- It can be quite tricky to build a BN
- Can build in different ways, but the CPT size can be quite different



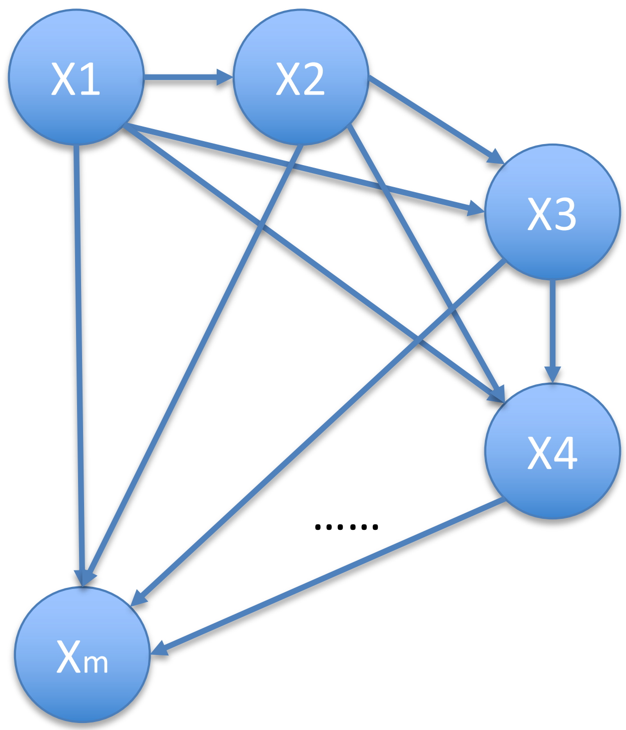
Building a BN

- Pearl's Network Construction Algorithm (A way):
  1. Choose a set of relevant variables that describe the domain
  2. Choose an **order** for the variables
  3. While there are variables left
     - add the next variable X_i to the network
     - 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: $P(X_i \mid X'_1, \ldots, X'_m) = P(X_i \mid Parents(X_i))$, where $X'_1, \ldots, X'_m$ are all the variables preceding X_i
     - Define the conditional probability table for X_i

Building a BN

- If we have variables $X_1, \dots, X_m$
- We don't know which are causes, which are effects
- We only have joint probability $P(X_1, \dots, X_m)$
- **Chain rule** (repeatedly use product rule):
  - $P(X_1, \dots, X_m) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2) \dots P(X_m|X_1, \dots, X_{m-1})$

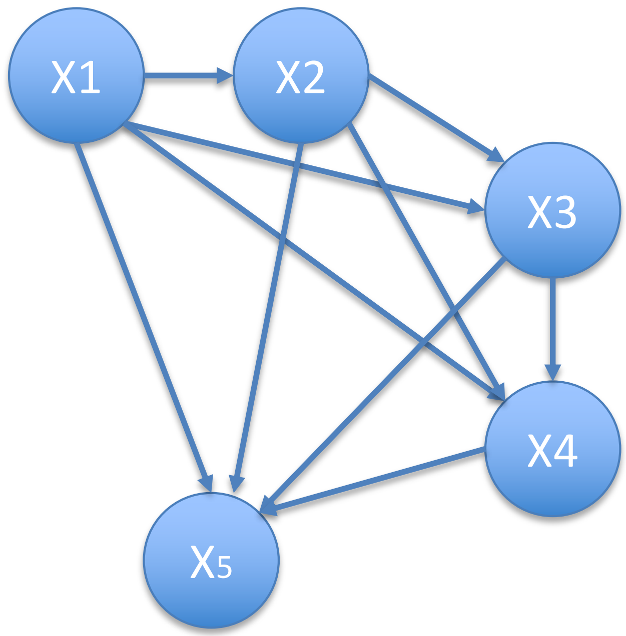
- CPT size (number of free parameters):
  - X_1: $|X_1| - 1$
  - X_2: $|X_1| \times (|X_2| - 1)$
  - …
  - X_m: $|X_1| \times |X_2| \times \cdots \times |X_{m-1}| \times (|X_m| - 1)$
  - Add together: $|X_1| \times |X_2| \times \cdots \times |X_{m-1}||X_m| - 1$



- The CPT size does not depend on order

Building a BN

- But we can **use domain knowledge**
  - E.g. given the murderer, whether to confess is independent of the weapon and time
  - $P(X_k \,|X_1, \ldots, X_{k-1}) = P(X_k \mid Parents(X_k))$
- Example:
  - X_5 is conditional independent of $\{X_2, X_4\}$ given $\{X_1, X_3\}$
  - $P(X_5|X_1, X_2, X_3, X_4) = P(X_5|X_1, X_3)$



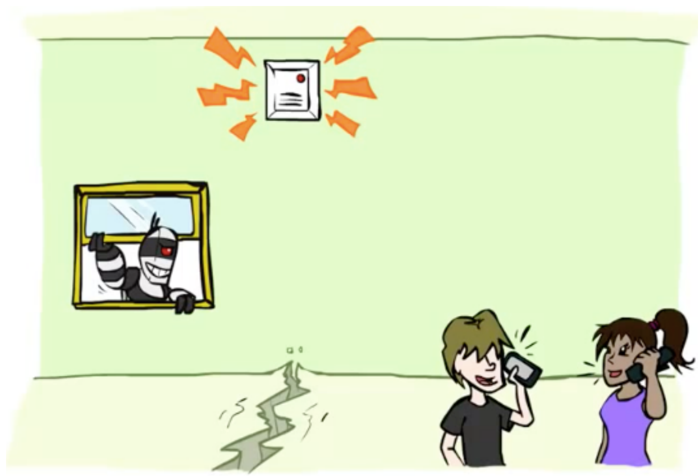
- Fewer parents leads to smaller CPT size
- BN to represent joint probability (factorisation)
  - Chain rule (always true): $P(X_1, \ldots, X_m) = \prod_i P(X_i|X_1, \ldots, X_{i-1})$
  - BN structure: $P(X_1, \ldots, X_m) = \prod_i P(X_i|Parents(X_i))$

Building a BN: Example

- Alarm network
  - I'm at work
  - John calls to say that my house alarm went off
  - but Mary did not call
  - The alarm will usually be set off by burglars
  - but sometimes it may also go off because of minor earthquakes

  - **Variables:**
    - Burglary, Earthquake, Alarm, JohnCalls, MaryCalls

  - Network topology reflects **causal knowledge** (given):
    - 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

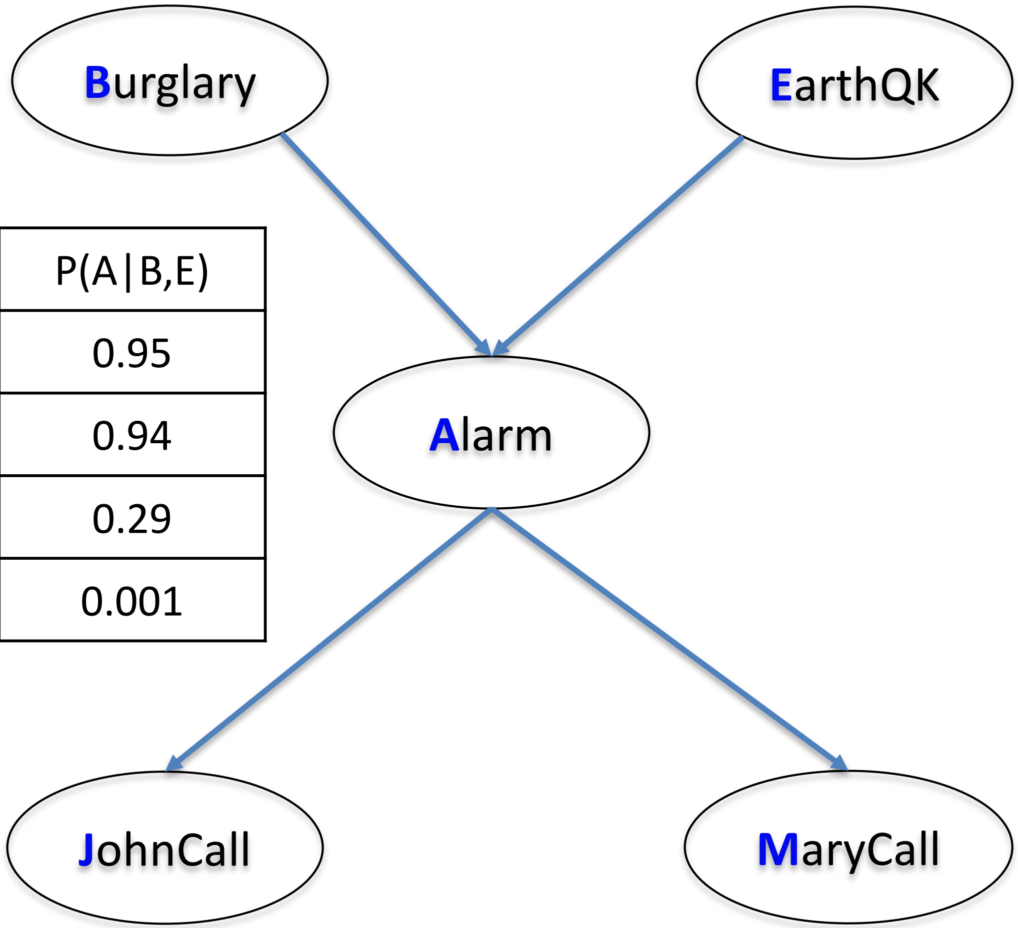


Building a BN: Example

- Factorisation of a BN (tell parents of each node):
  - P(B, E, A, J, M) = P(B)*P(E)*P(A | B,E)*P(J | A)*P(M | A)

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

Compactness and Node Ordering

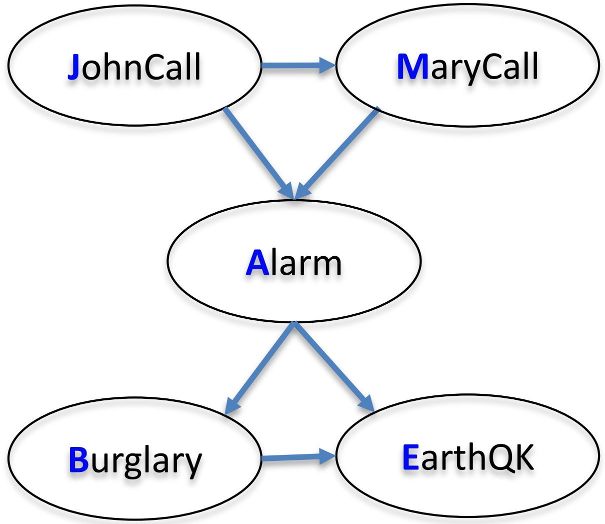
- Compactness: 紧凑性
  - The more compact the model is, the smaller the CPT size
  - Less computer memory, more computationally efficient 计算效率
  - Over dense networks fail to represent independencies explicitly
  - Over dense networks fail to represent the causal dependencies in the domain ---过密网络无法明确表示独立性 -- 过密网络无法表示域中的因果依赖关系

- 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.

  紧凑性取决于节点的顺序是否"正确"。
  最佳的顺序是先添加（root causes）根本原因，
  然后再添加（variables they influence directly）它们直接影响的变量，
  一直到达到叶子

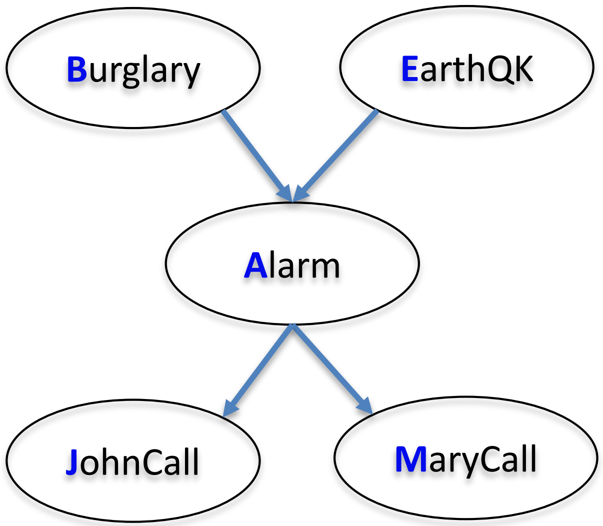
Building BN

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

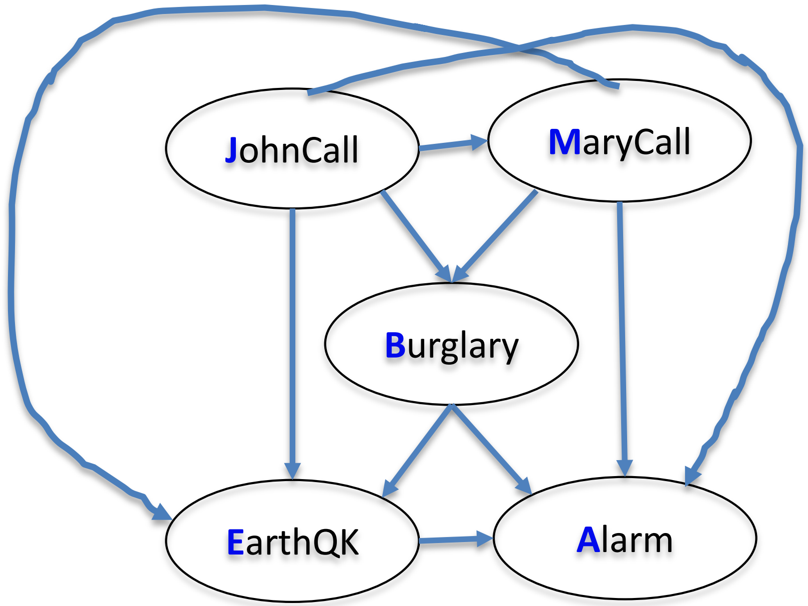


- **Step 1**: Add node J
- **Step 2**: Add node M
  - $P(M \mid J) = P(M)$? No, J -> M
- **Step 3**: Add node A
  - $P(A \mid M, J) = P(A)$? No
  - $P(A \mid M, J) = P(A \mid J)$? No
  - $P(A \mid M, J) = P(A \mid M)$? No, M -> A and J -> A
- **Step 4**: Add node B
  - $P(B \mid M, J, A) = P(B)$? No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

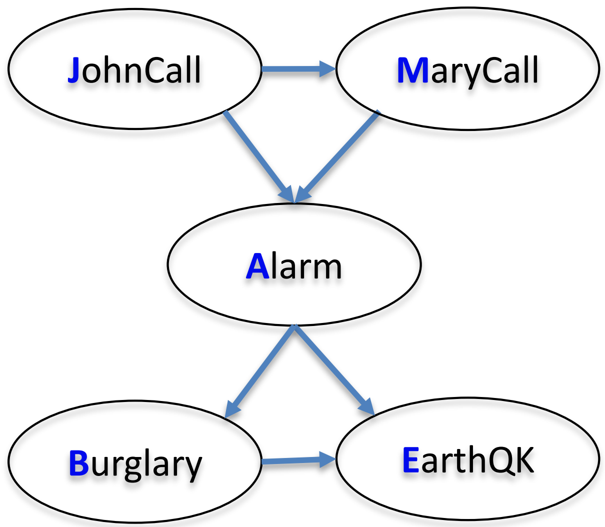
Ordering and Compactness



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



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



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

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)