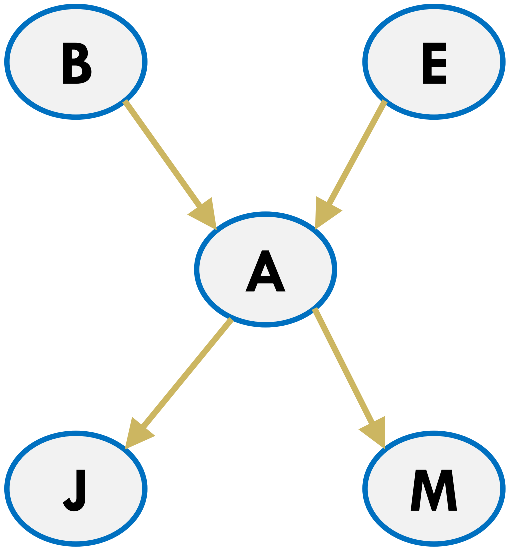
CS 4104
APPLIED MACHINE LEARNING

Dr. Hashim Yasin

National University of Computer and Emerging Sciences,

Faisalabad, Pakistan.



BAYESIAN BELIEF NETWORKS

Bayesian Belief Networks

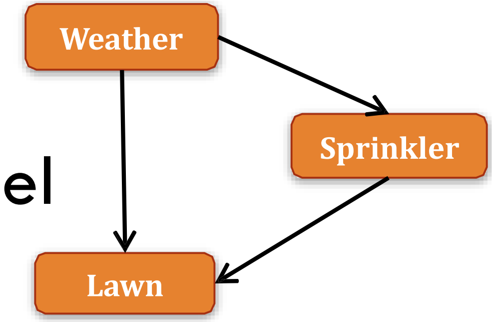
Naive Bayes classifier

- which assumes that ***all*** the variables are conditionally independent given the value of the target variable,

Bayesian belief networks

- which allows stating conditional independence assumptions that apply to ***subsets*** of the variables.
- It provide an *intermediate approach that is less constraining* than the global assumption of conditional independence made by the naive Bayes classifier.

Bayesian Belief Networks



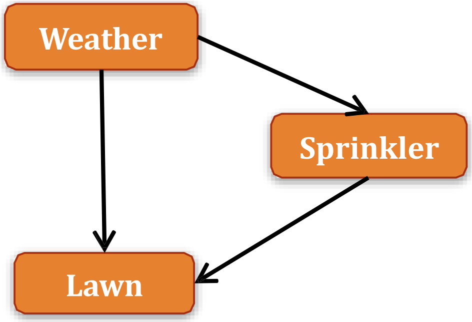
Bayesian Belief Networks (BBN)

- Graphical (Directed Acyclic Graph) Model
- **Nodes** are the features:
  - Each has a set of possible parameters/values/ states
  - Weather = {sunny, cloudy, rainy}; Sprinkler = {off, on}; Lawn = {dry, wet}
- **Edges / Links** represent relations between features
- BBN is a probabilistic graphical model (PGM)
- Each node/feature is a random variable

Bayesian Belief Networks

Bayesian Belief Networks

- We call these probabilities of occurring states - **Beliefs**
  - Example: our belief in the state {*coin*=‘*head*’} is 50%
- All beliefs of all possible states of a node are gathered in a single CPT - Conditional Probability Table
- *BBN is a 2-component model:*
  - Graph
  - CPTs



Bayesian Belief Networks

Network represents joint probability distribution over all variables

- In general,

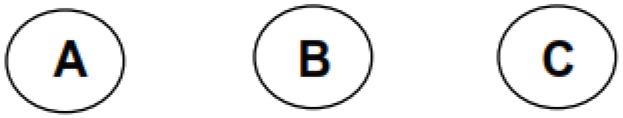
$$\boxed{P(y_1, \ldots, y_n) = \prod_{i=1}^{n} P(y_i | Parents(Y_i))}$$

  where $Parents(Y_i)$ denotes immediate predecessors of Y_i in graph

- So joint distribution is fully defined by graph, plus the $P(y_i | Parents(Y_i))$

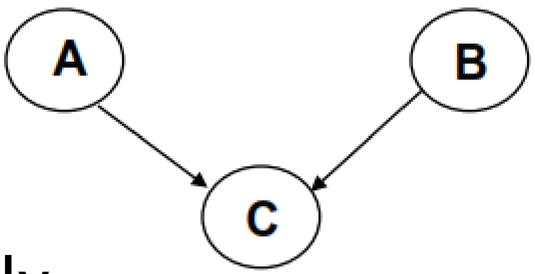
Bayesian Belief Networks

Marginal Independence:



- $P(A, B, C) \ = P(A)\ P(B)\ P(C)$

Independent Causes:

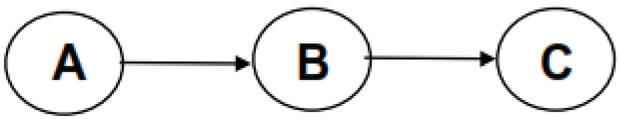


- $P(A, B, C) \ = \ P(C|A, B) P(A) P(B)$

"Explaining away" effect:

- Given C, observing A makes B less likely
- A and B are (marginally) independent but become dependent once C is known

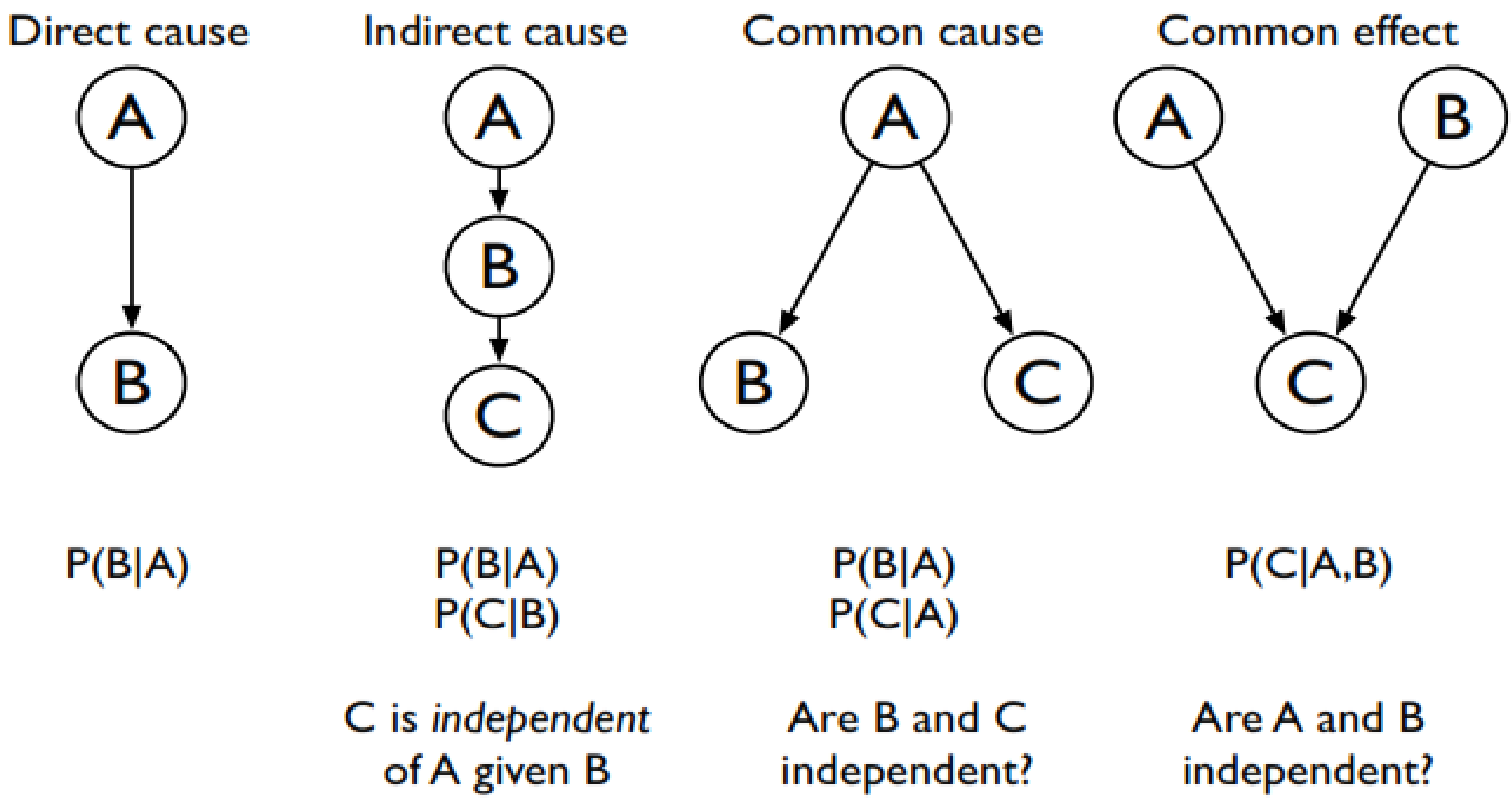
Markov dependence:

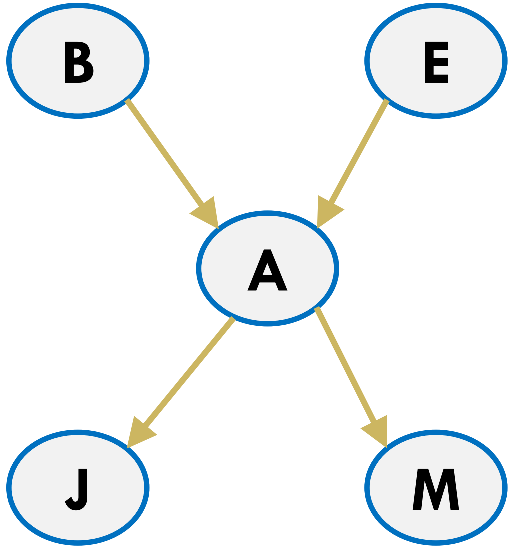


- $P(A, B, C) \ = \ P(C|B)\ P(B|A) P(A)$

Bayesian Belief Networks

- **Types of probabilistic relationships**





BAYESIAN BELIEF NETWORKS EXAMPLE

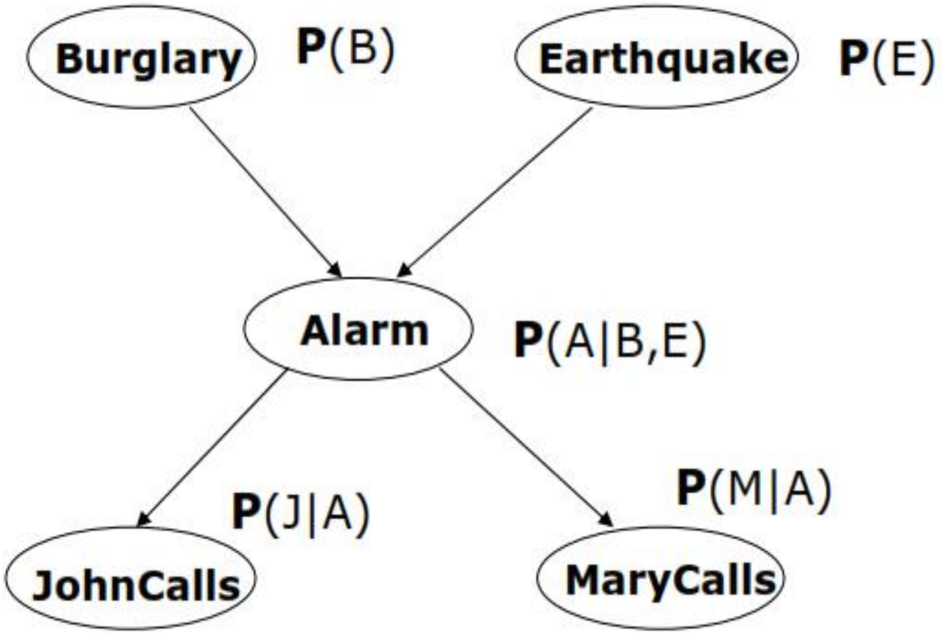
Example

- Assume your house has an **alarm system** against **burglary**. You live in the seismically active area and the alarm system can get occasionally set off by an **earthquake.**
- You have two neighbors, **Mary** and **John**, who do not know each other. *If they hear the alarm they call you, but this is not guaranteed*.
- We want to represent the probability distribution of events:
  - Burglary, Earthquake, Alarm, Mary calls and John calls

Example

Directed acyclic graph

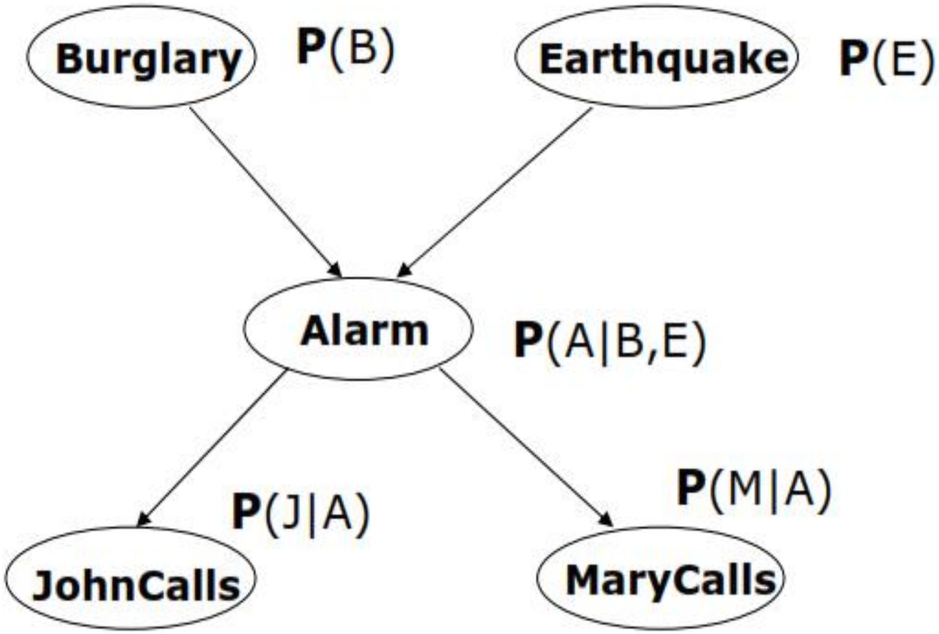
- **Nodes** = random variables (Burglary, Earthquake, Alarm, Mary calls and John calls)



Example

Directed acyclic graph

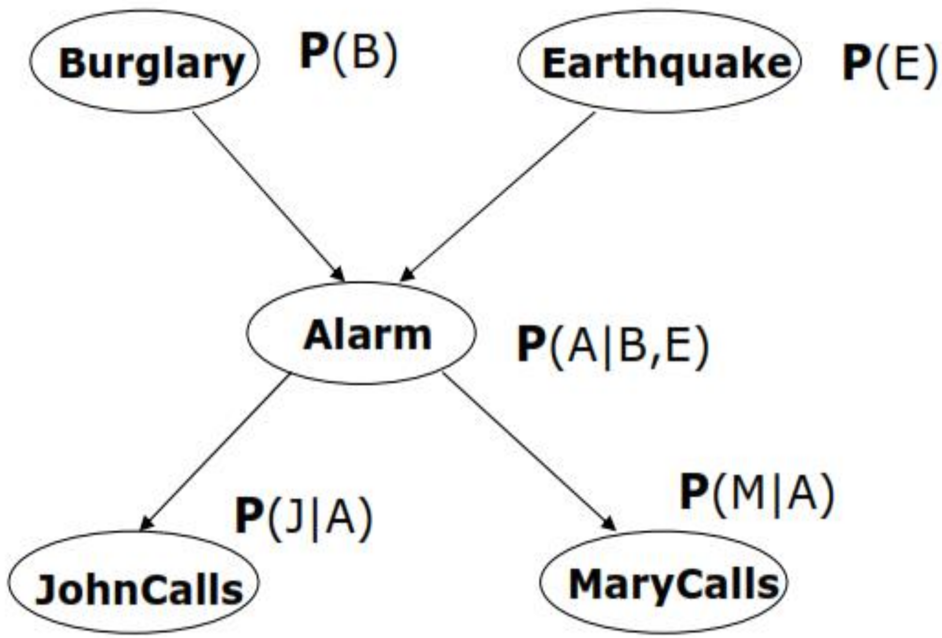
- **Links** = direct (causal) dependencies between variables.



Example

Directed acyclic graph

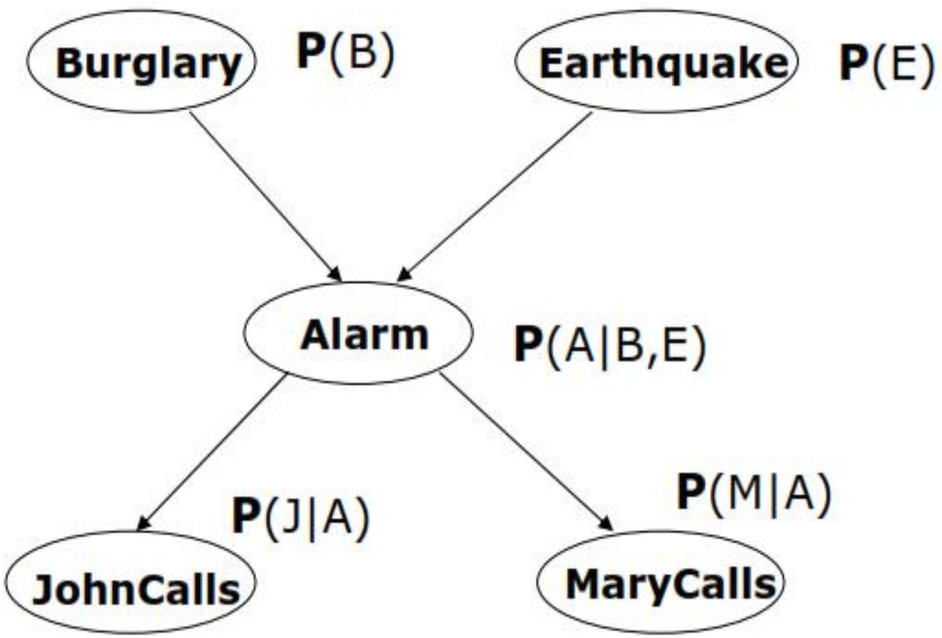
- The chance of Alarm is influenced by Earthquake, The chance of John calling is affected by the Alarm.



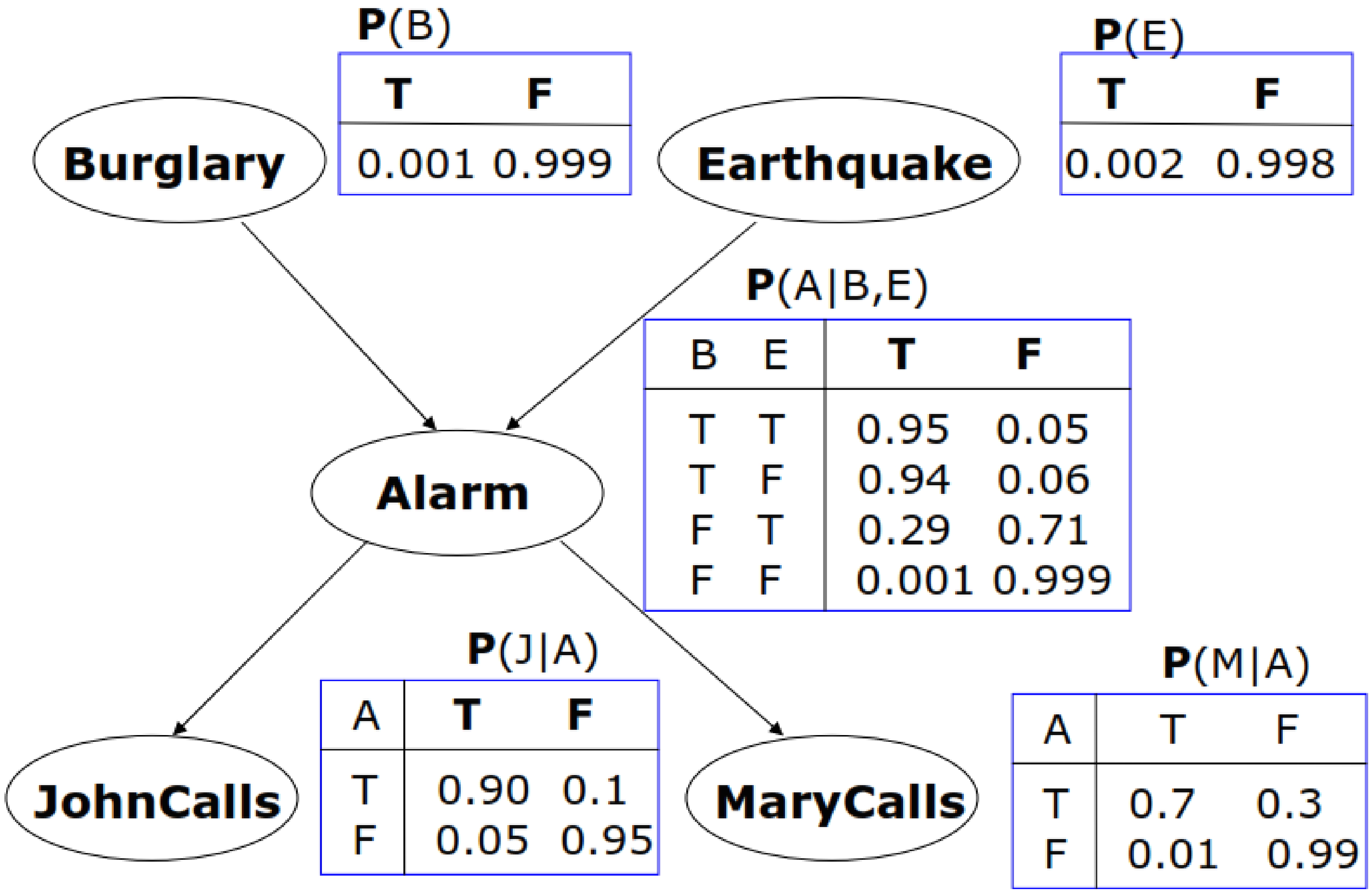
Example

Local conditional distributions

- relate variables and their parents.



Example



P(B)

T	F
0.001	0.999

P(E)

T	F
0.002	0.998

P(A|B,E)

B	E	T	F
T	T	0.95	0.05
T	F	0.94	0.06
F	T	0.29	0.71
F	F	0.001	0.999

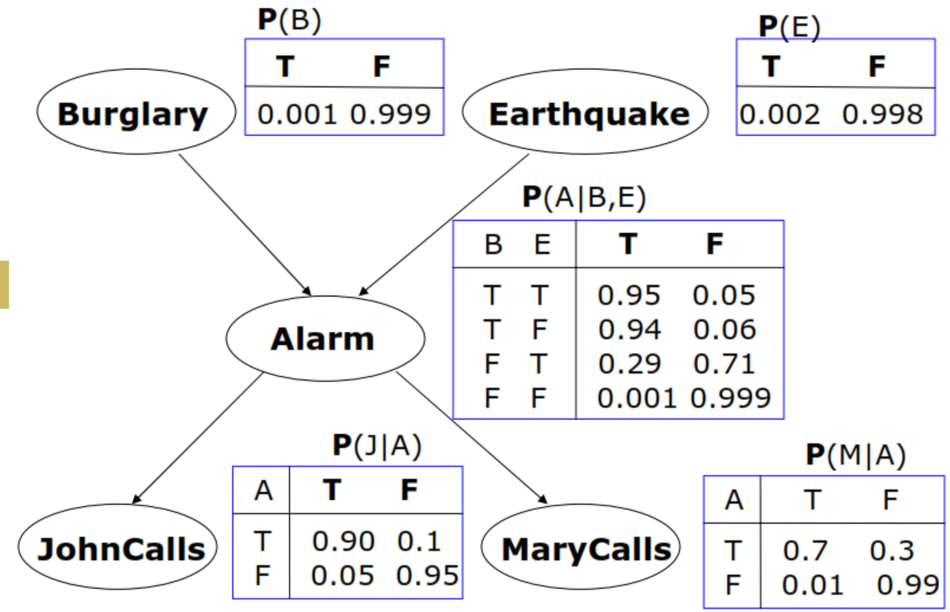
P(J|A)

A	T	F
T	0.90	0.1
F	0.05	0.95

P(M|A)

A	T	F
T	0.7	0.3
F	0.01	0.99

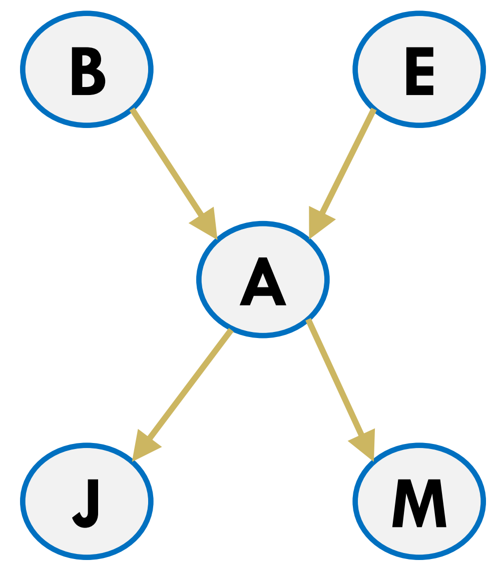
Example



<u>Example 2:</u>

$\underline{P(J, M, A, \neg B, \neg E)}$

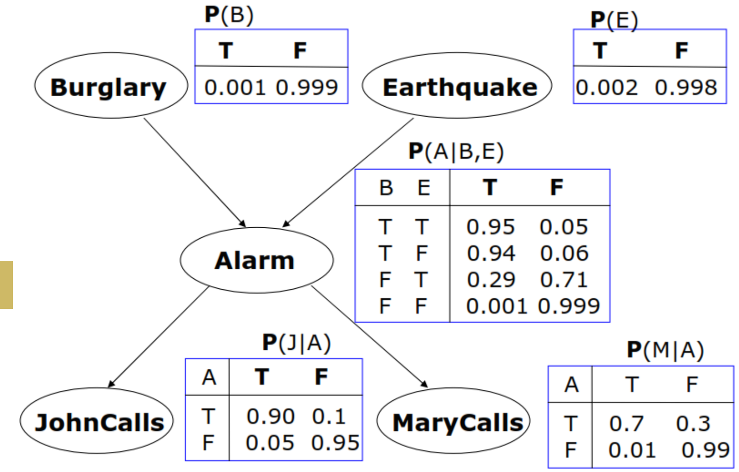
$$P(J, M, A, \neg B, \neg E) =$$
$$= P(J \mid A) P(M \mid A) P(A \mid \neg B, \neg E) P(\neg B) P(\neg E)$$
$$= 0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998 = 0.00062$$



Example

Example 3:

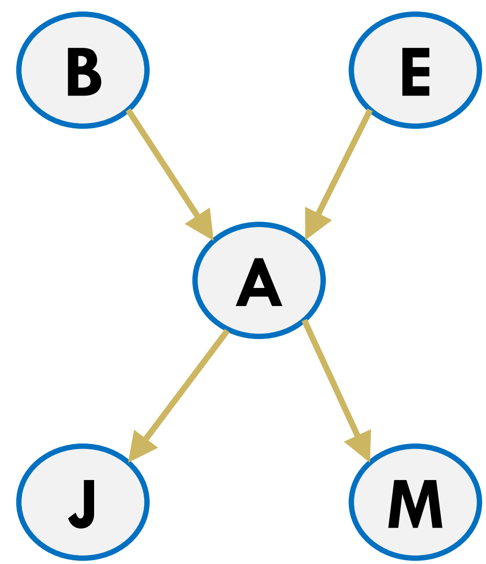
- $P(J, M, A, E, B) = P(J \mid A)\, P(M \mid A)\, P(A \mid E, B)\, P(E)\, P(B)$



- There are 3 conditional probability tables (CPTs) to be determined:
  - $P(J \mid A), P(M \mid A), P(A \mid E, B)$
  - Requiring $4 + 4 + 8 = 16$ probabilities

- And 2 marginal probabilities $P(E), P(B)$
  - Requiring $2 + 2 = 4$ probabilities

- Total: $4 + 16 = 20$ probabilities



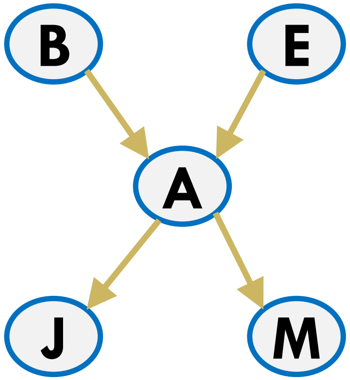
Example

- **Example 4:** What's the probability of a burglary if both Mary and John call,
  P(burglary | johhcalls, marycalls)?

Example: *P(burglary | johncalls, marycalls)*? (Abbrev. $P(b \mid j, m)$)

$$
\begin{aligned}
&P(b \mid j, m) \\
&= \alpha P(b, j, m) \\
&= \alpha \sum_a \sum_e P(b, j, m, a, e) \\
&= \alpha \sum_a \sum_e P(j, m, a, b, e) \\
&= \alpha\big(P(j,m,a,b,e) + P(j,m,\neg a,b,e) + P(j,m,a,b,\neg e) + P(j,m,\neg a,b,\neg e)\big)
\end{aligned}
$$

$\alpha = 480$



Example

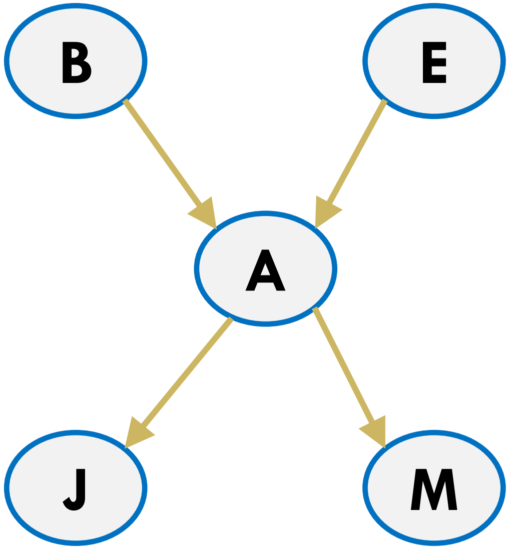
- What's the probability of a burglary if both Mary and John call, *P(burglary | johhcalls, marycalls)?*

$\alpha = 480$

$$P(b \mid j,m) = \alpha \textstyle\sum_a \sum_e P(j,m,a,b,e)$$
$$P(\neg b \mid j,m) = \alpha \textstyle\sum_a \sum_e P(j,m,a,\neg b,e)$$

$$P(b \mid j,m) = \alpha P(b) \textstyle\sum_a P(j|a)\, P(m|a) \sum_e P(a|b,e)P(e) = \ldots = \alpha * 0.00059$$
$$P(\neg b \mid j,m) = \alpha P(\neg b) \textstyle\sum_a P(j|a)\, P(m|a) \sum_e P(a|\neg b,e)P(e) = \ldots = \alpha * 0.0015$$

$$\mathbf{P}(B \mid j,m) = \alpha \langle 0.00059, 0.0015 \rangle = \mathbf{\langle 0.28, 0.72 \rangle}.$$



CURSE OF DIMENSIONALITY

Curse of Dimensionality

Network Size = number of parameters

Sunny
Rainy

Weather

On
Off

Sprinkler

Lawn

Wet
Dry

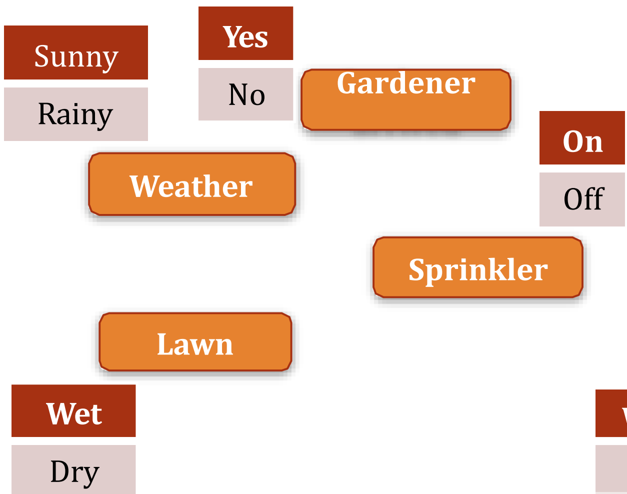
Weather
Sunny
Rainy

weather	sprinkler
Sunny	On
Sunny	Off
Rainy	On
Rainy	Off

Weather	Sprinkler	Lawn
Sunny	On	Wet
Sunny	On	Dry
Sunny	Off	Wet
Sunny	Off	Dry
Rainy	On	Wet
Rainy	On	Dry
Rainy	Off	Wet
Rainy	Off	Dry

Curse of Dimensionality

Network Size = number of parameters



Weather
Sunny
Rainy

weather	sprinkler
Sunny	On
Sunny	Off
Rainy	On
Rainy	Off

Weather	Sprinkler	Lawn
Sunny	On	Wet
Sunny	On	Dry
Sunny	Off	Wet
Sunny	Off	Dry
Rainy	On	Wet
Rainy	On	Dry
Rainy	Off	Wet
Rainy	Off	Dry

Weather	Sprinkler	Lawn	Gardener Arrived
Sunny	On	Wet	Yes
Sunny	On	Wet	No
Sunny	On	Dry	Yes
Sunny	On	Dry	No
Sunny	Off	Wet	Yes
Sunny	Off	Wet	No
Sunny	Off	Dry	Yes
Sunny	Off	Dry	No
Rainy	On	Wet	Yes
Rainy	On	Wet	No
Rainy	On	Dry	Yes
Rainy	On	Dry	No
Rainy	Off	Wet	Yes
Rainy	Off	Wet	No
Rainy	Off	Dry	Yes
Rainy	Off	Dry	No

Curse of Dimensionality

Network Size = number of parameters

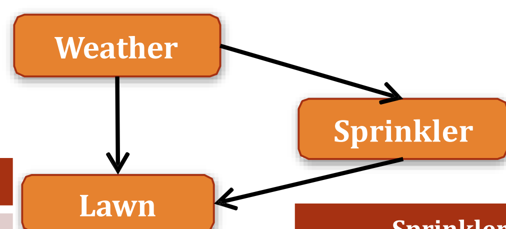
Network grows exponentially with number of nodes ~ 2^N

Each additional node doubles the size of the network!

A network with 100 nodes → 2^{100} parameters! → Impractical!

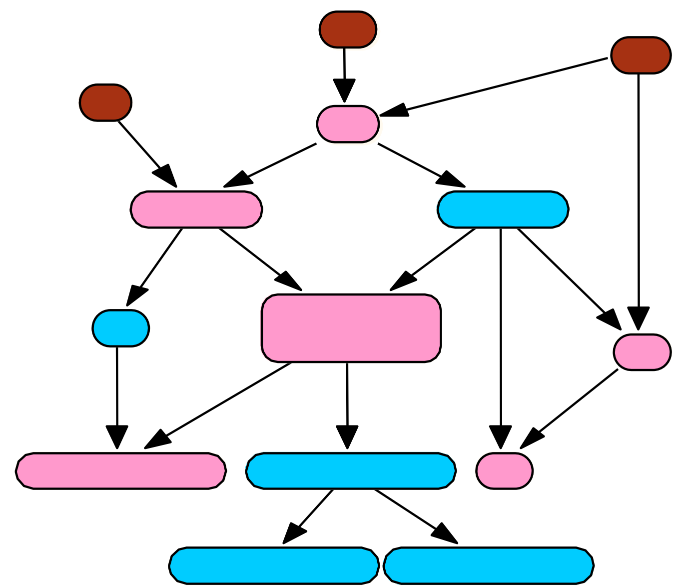
BBN – that reduce dimensionality

Weather	
Sunny	
Rainy	



Lawn			
Weather	Sprinkler	Wet	Dry
Sunny	On		
Sunny	Off		
Rainy	On		
Rainy	Off		

Sprinkler		
Weather	On	Off
Sunny		
Rainy		



Joint size = 2^{14} = 16K

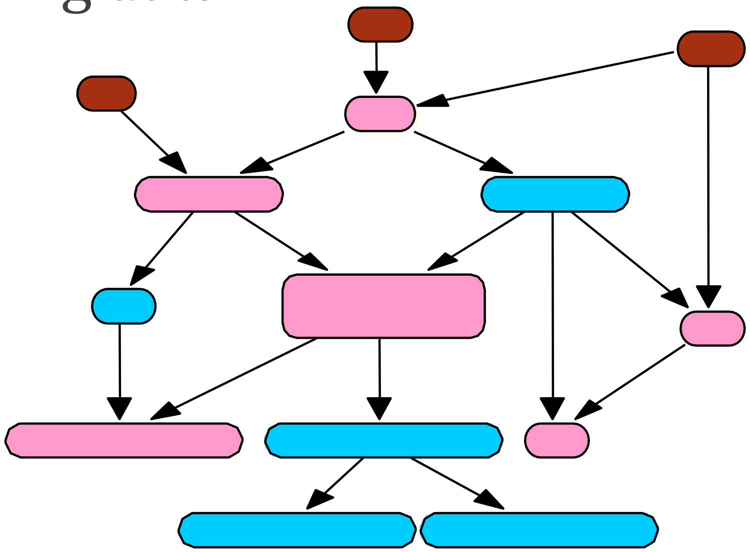
BBN size = 3*2 + 5*4 + 6*8 = 74

Curse of Dimensionality

BBN battles the curse of dimensionality

One of the most powerful properties of BBN

For estimating 74 parameters instead of 16K you need much less training data



Joint size = 2^{14} = 16K

BBN size = 3*2 + 5*4 + 6*8 = 74

Acknowledgement

Tom Mitchel, Russel & Norvig, Wolfram Burgard, Maren Bennewitz, Marco Ragni