CS 2750 Machine Learning
Lecture 17

Bayesian belief networks.

Milos Hauskrecht
milos@cs.pitt.edu
5329 Sennott Square

Modeling uncertainty with probabilities

- **Full joint distribution:** joint distribution over all random variables defining the domain
  - it is sufficient to represent the complete domain and to do any type of probabilistic inferences

Problems:

- **Space complexity.** To store full joint distribution requires to remember $O(d^n)$ numbers.
  n – number of random variables, d – number of values
- **Inference complexity.** To compute some queries requires $O(d^n)$ steps.
- **Acquisition problem.** Who is going to define all of the probability entries?

Pneumonia example. Complexities.

- **Space complexity**.
  - Pneumonia (2 values: T,F), Fever (2: T,F), Cough (2: T,F), WBCcount (3: high, normal, low), paleness (2: T,F)
  - Number of assignments: 2*2*2*3*2=48
  - We need to define at least 47 probabilities.
- **Time complexity**.
  - Assume we need to compute the probability of Pneumonia=T from the full joint

$$P(Pneumonia = T) =$$
$$= \sum_{i \in T,F} \sum_{j \in T,F} \sum_{k=h,n,l} \sum_{u \in T,F} P(Fever = i, Cough = j, WBCcount = k, Pale = u)$$

  - Sum over 2*2*3*2=24 combinations

Bayesian belief networks (BBNs)

Bayesian belief networks.

- Represent the full joint distribution over the variables more compactly with a **smaller number of parameters**.
- Take advantage of **conditional and marginal independences** among random variables

- **A and B are independent**

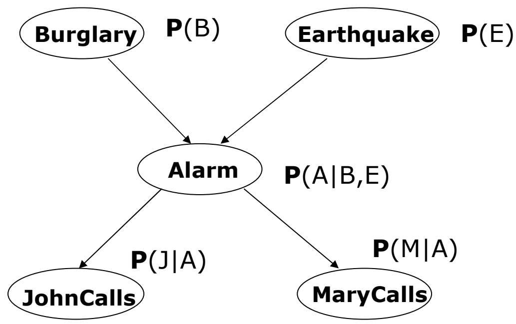
$$P(A, B) = P(A)P(B)$$

- **A and B are conditionally independent given C**

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

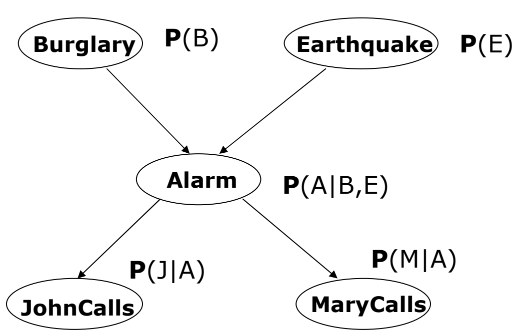
Bayesian belief network.

1. **Directed acyclic graph**
   - **Nodes** = random variables
   - **Links** = missing links encode independences.

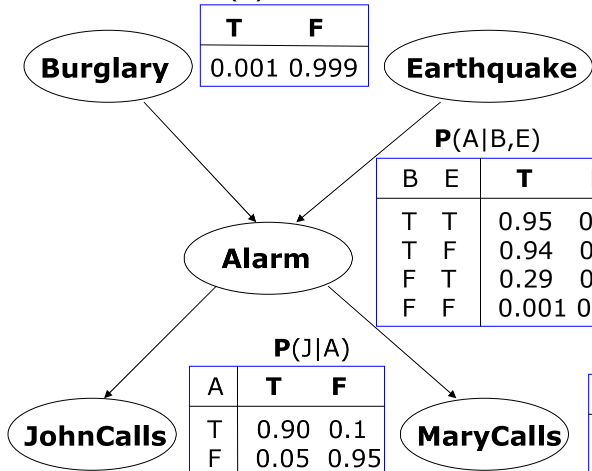


Bayesian belief network.

2. **Local conditional distributions**
   - relate variables and their parents



Bayesian belief network



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)

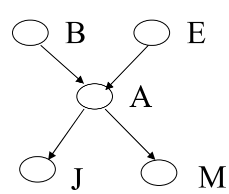
A	T	F
T	0.7	0.3
F	0.01	0.99

Full joint distribution in BBNs

Full joint distribution is defined in terms of local conditional distributions (obtained via the chain rule):

$$\mathbf{P}(X_1, X_2, .., X_n) = \prod_{i=1,..n} \mathbf{P}(X_i \mid pa(X_i))$$

Example:



Assume the following assignment of values to random variables

$B=T, E=T, A=T, J=T, M=F$

Then its probability is:

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

Bayesian belief networks (BBNs)

Bayesian belief networks
- Represent the full joint distribution over the variables more compactly using the product of local conditionals.
- **But how did we get to local parameterizations?**

Answer:
- **Graphical structure** encodes **conditional and marginal independences** among random variables
- **A and B are independent** $P(A,B) = P(A)P(B)$
- **A and B are conditionally independent given C**

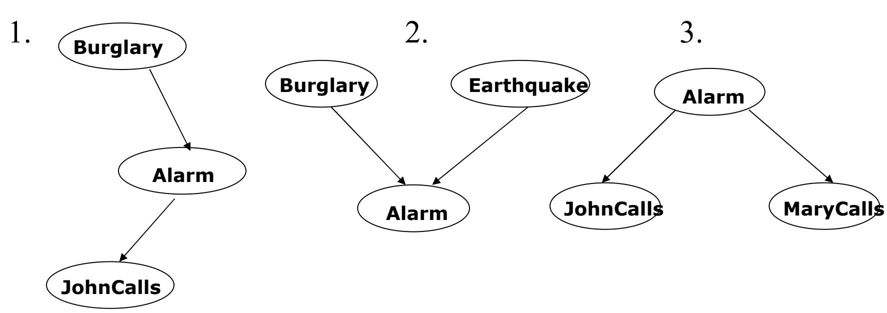
$$P(A \mid C, B) = P(A \mid C)$$

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

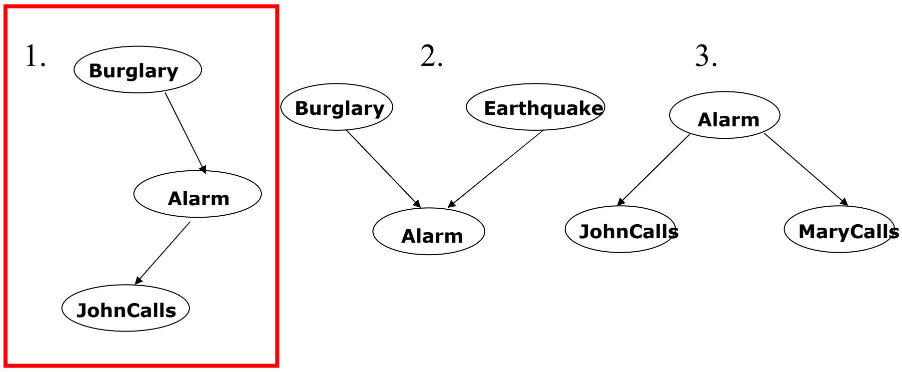
- **The graph structure implies the decomposition !!!**

Independences in BBNs

3 basic independence structures:



Independences in BBNs

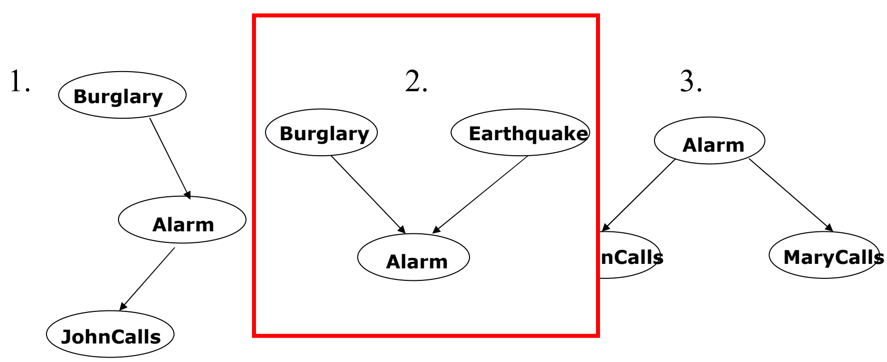


1. JohnCalls **is independent** of Burglary given Alarm

$$P(J \mid A, B) = P(J \mid A)$$

$$P(J, B \mid A) = P(J \mid A) P(B \mid A)$$

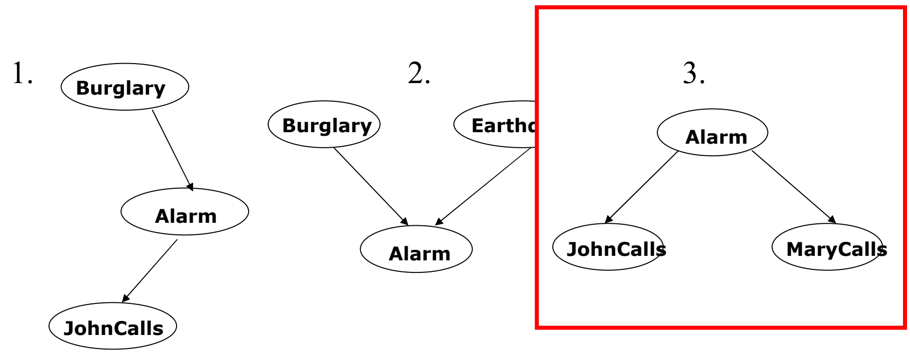
Independences in BBNs



2. Burglary **is independent** of Earthquake (not knowing Alarm)
   Burglary and Earthquake **become dependent** given Alarm !!

$$P(B, E) = P(B) P(E)$$

Independences in BBNs



3. MaryCalls **is independent** of JohnCalls given Alarm

$$P(J \mid A, M) = P(J \mid A)$$

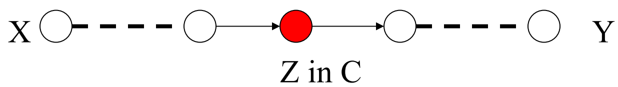
$$P(J, M \mid A) = P(J \mid A) P(M \mid A)$$

Independences in BBN

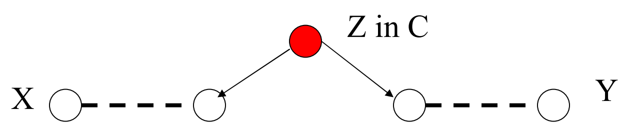
- BBN distribution models many conditional independence relations relating distant variables and sets
- These are defined in terms of the graphical criterion called d-separation
- **D-separation in the graph**
  - Let X,Y and Z be three sets of nodes
  - If X and Y are d-separated by Z then X and Y are conditionally independent given Z
- **D-separation :**
  - **A is d-separated from B given C** if every undirected path between them is **blocked**
- **Path blocking**
  - 3 cases that expand on three basic independence structures

Undirected path blocking

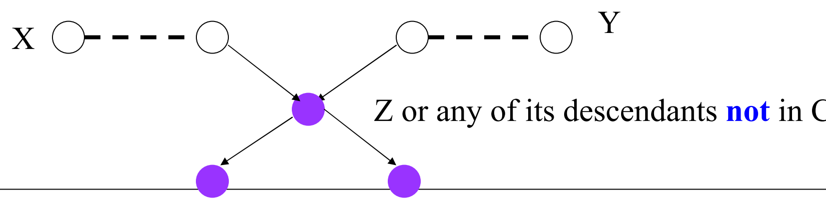
- 1. With linear substructure



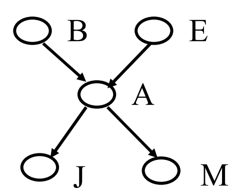
- 2. With wedge substructure



- 3. With vee substructure



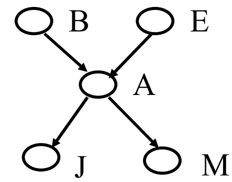
Full joint distribution in BBNs



Rewrite the full joint probability using the product rule:

$$P(B=T,E=T,A=T,J=T,M=F)=$$

Full joint distribution in BBNs

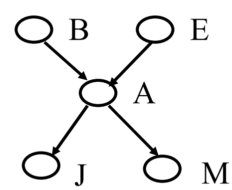


Rewrite the full joint probability using the product rule:

$$P(B=T,E=T,A=T,J=T,M=F)=$$

$$=P(J=T \mid B=T,E=T,A=T,M=F)P(B=T,E=T,A=T,M=F)$$
$$=\underline{P(J=T \mid A=T)}P(B=T,E=T,A=T,M=F)$$

Full joint distribution in BBNs



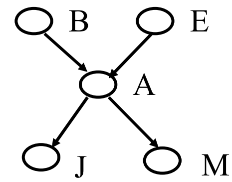
Rewrite the full joint probability using the product rule:

$$P(B=T,E=T,A=T,J=T,M=F)=$$

$$=P(J=T \mid B=T,E=T,A=T,M=F)P(B=T,E=T,A=T,M=F)$$
$$=\underline{P(J=T \mid A=T)}P(B=T,E=T,A=T,M=F)$$
$$P(M=F \mid B=T,E=T,A=T)P(B=T,E=T,A=T)$$
$$\underline{P(M=F \mid A=T)}P(B=T,E=T,A=T)$$

Full joint distribution in BBNs

Rewrite the full joint probability using the product rule:



$$P(B=T,E=T,A=T,J=T,M=F)=$$

$$=P(J=T\,|\,B=T,E=T,A=T,M=F)P(B=T,E=T,A=T,M=F)$$

$$=\underline{P(J=T\,|\,A=T)}P(B=T,E=T,A=T,M=F)$$

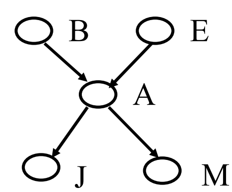
$$P(M=F\,|\,B=T,E=T,A=T)P(B=T,E=T,A=T)$$

$$\underline{P(M=F\,|\,A=T)}P(B=T,E=T,A=T)$$

$$\underline{P(A=T\,|\,B=T,E=T)}P(B=T,E=T)$$

Full joint distribution in BBNs

Rewrite the full joint probability using the product rule:



$$P(B=T,E=T,A=T,J=T,M=F)=$$

$$=P(J=T\,|\,B=T,E=T,A=T,M=F)P(B=T,E=T,A=T,M=F)$$

$$=\underline{P(J=T\,|\,A=T)}P(B=T,E=T,A=T,M=F)$$

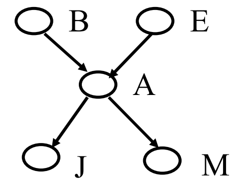
$$P(M=F\,|\,B=T,E=T,A=T)P(B=T,E=T,A=T)$$

$$\underline{P(M=F\,|\,A=T)}P(B=T,E=T,A=T)$$

$$\underline{P(A=T\,|\,B=T,E=T)}P(B=T,E=T)$$

$$P(B=T)P(E=T)$$

Full joint distribution in BBNs



Rewrite the full joint probability using the product rule:

$$P(B=T,E=T,A=T,J=T,M=F)=$$

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

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

$$P(M=F \mid B=T,E=T,A=T)P(B=T,E=T,A=T)$$

$$P(M=F \mid A=T)P(B=T,E=T,A=T)$$

$$P(A=T \mid B=T,E=T)P(B=T,E=T)$$

$$P(B=T)P(E=T)$$

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

Parameter complexity problem

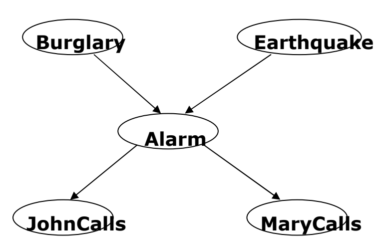
- In the BBN the full joint distribution is expressed as a product of conditionals (of smaller) complexity

$$\mathbf{P}(X_1, X_2, \ldots, X_n) = \prod_{i=1,..n} \mathbf{P}(X_i \mid pa(X_i))$$

Parameters:

full joint: ?

BBN: ?



Parameter complexity problem

- In the BBN the **full joint distribution** is defined as:

$$\mathbf{P}(X_1, X_2, \ldots, X_n) = \prod_{i=1,..n} \mathbf{P}(X_i \mid pa(X_i))$$

- **What did we save?**

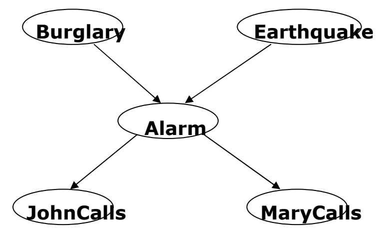
Alarm example: 5 binary (True, False) variables

# of parameters of the full joint:

$$2^5 = 32$$

One parameter is for free:

$$2^5 - 1 = 31$$



Parameter complexity problem

- In the BBN the **full joint distribution** is defined as:

$$\mathbf{P}(X_1, X_2, \ldots, X_n) = \prod_{i=1,..n} \mathbf{P}(X_i \mid pa(X_i))$$

- **What did we save?**

Alarm example: 5 binary (True, False) variables

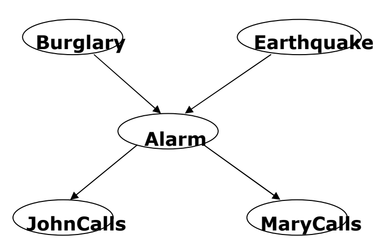
# of parameters of the full joint:

$$2^5 = 32$$

One parameter is for free:

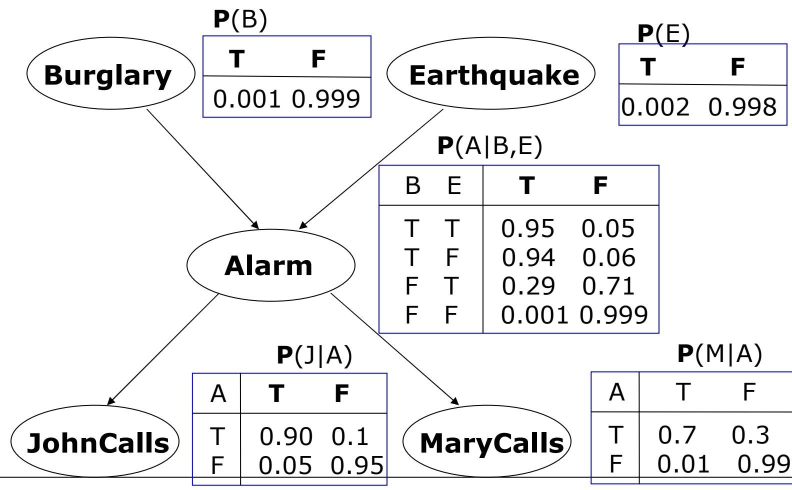
$$2^5 - 1 = 31$$

# of parameters of the BBN: ?



Bayesian belief network.

- In the BBN the **full joint distribution** is expressed using a set of local conditional distributions



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)

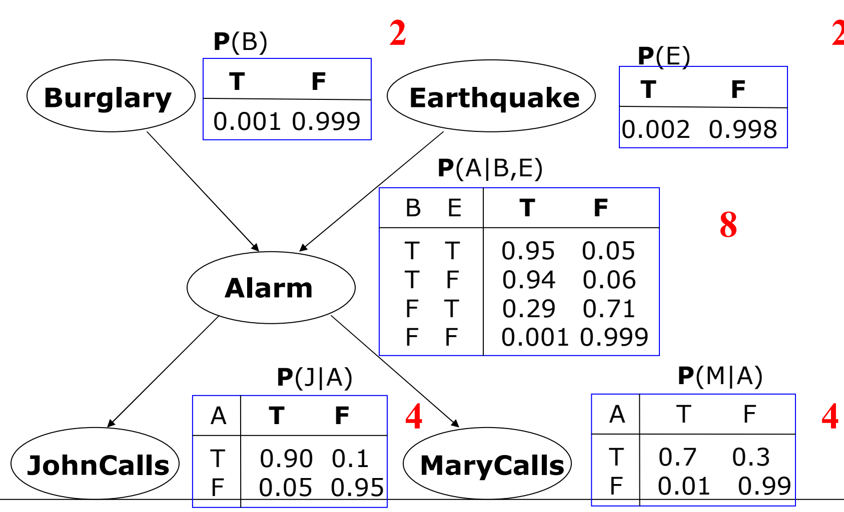
A	**T**	**F**
T	0.90	0.1
F	0.05	0.95

P(M|A)

A	T	F
T	0.7	0.3
F	0.01	0.99

Bayesian belief network.

- In the BBN the **full joint distribution** is expressed using a set of local conditional distributions



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

Parameter complexity problem

- In the BBN the **full joint distribution** is defined as:

$$\mathbf{P}(X_1, X_2, \ldots, X_n) = \prod_{i=1,..n} \mathbf{P}(X_i \mid pa(X_i))$$

- **What did we save?**

Alarm example: 5 binary (True, False) variables

# of parameters of the full joint:

$$2^5 = 32$$

One parameter is for free:

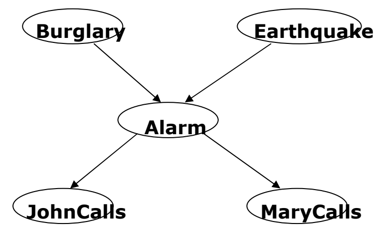
$$2^5 - 1 = 31$$

# of parameters of the BBN:

$$2^3 + 2(2^2) + 2(2) = 20$$

One parameter in every conditional is for free:

?



Parameter complexity problem

- In the BBN the **full joint distribution** is defined as:

$$\mathbf{P}(X_1, X_2, \ldots, X_n) = \prod_{i=1,..n} \mathbf{P}(X_i \mid pa(X_i))$$

- **What did we save?**

Alarm example: 5 binary (True, False) variables

# of parameters of the full joint:

$$2^5 = 32$$

One parameter is for free:

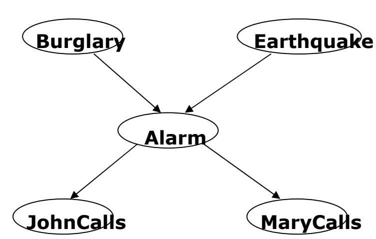
$$2^5 - 1 = 31$$

# of parameters of the BBN:

$$2^3 + 2(2^2) + 2(2) = 20$$

One parameter in every conditional is for free:

$$2^2 + 2(2) + 2(1) = 10$$



Model acquisition problem

The structure of the BBN typically reflects causal relations

- BBNs are also sometime referred to as **causal networks**
- Causal structure is very intuitive in many applications domain and it is relatively easy to obtain from the domain expert

Probability parameters of BBN correspond to conditional distributions relating a random variable and its parents only

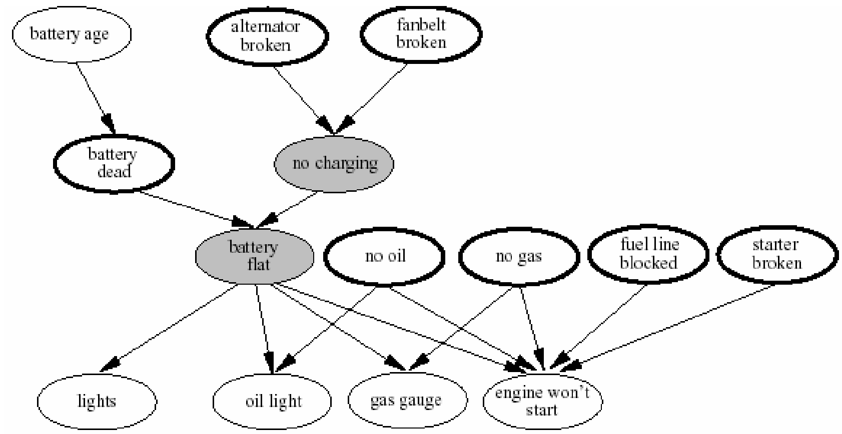
- Their complexity much smaller than the full joint
- Easier to come up (estimate) the probabilities from expert or automatically by learning from data

BBNs built in practice

- **In various areas:**
  - Intelligent user interfaces (Microsoft)
  - Troubleshooting, diagnosis of a technical device
  - Medical diagnosis:
    - Pathfinder (Intellipath)
    - CPSC
    - Munin
    - QMR-DT
  - Collaborative filtering
  - Military applications
  - Insurance, credit applications

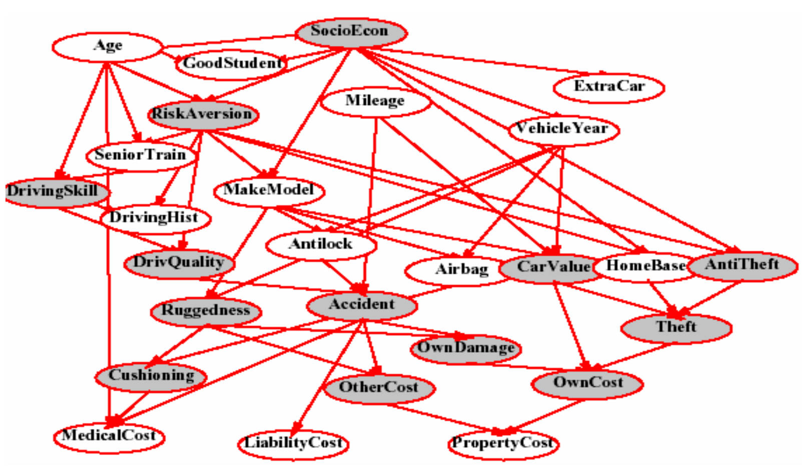
Diagnosis of car engine

- Diagnose the engine start problem

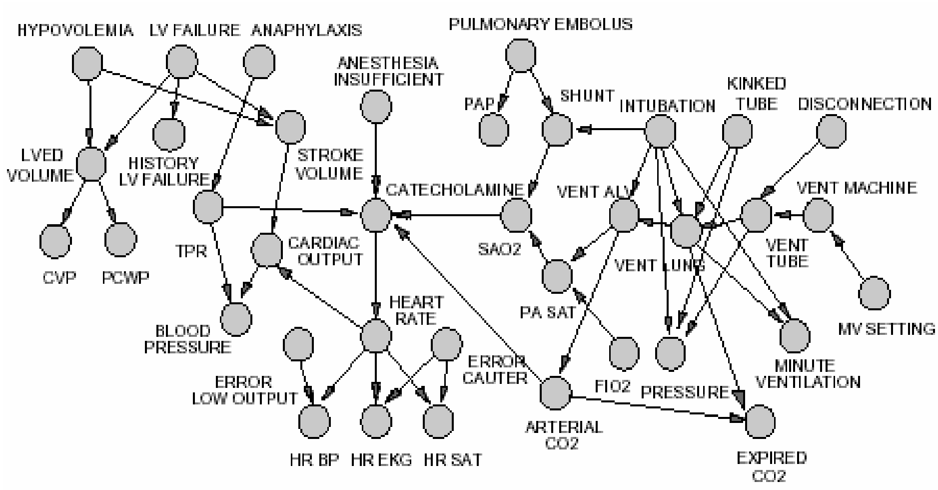


Car insurance example

- Predict claim costs (medical, liability) based on application data



(ICU) Alarm network



CPCS

- **C**omputer-based **P**atient **C**ase **S**imulation system (CPCS-PM) developed by Parker and Miller (at University of Pittsburgh)
- 422 nodes and 867 arcs

QMR-DT

- **Medical diagnosis in internal medicine**
  - Bipartite network of disease/findings relations
  - Derived from the Internist-1/QMR knowledge base