## CS 2001 – Lecture 1

# Bayesian belief networks

**Milos Hauskrecht**
milos@cs.pitt.edu
5329 Sennott Square
X4-8845



## Milos' research interests

**Artificial Intelligence**

- Planning, reasoning and optimization in the presence of uncertainty
- Machine learning
- Applications:
  - Medicine
  - Finance and investments

**Main research focus:**

- Models of high dimensional stochastic problems and their efficient solutions

# KB for medical diagnosis.

We want to build a KB system for the **diagnosis of pneumonia**.

**Problem description:**

- **Disease:** pneumonia
- **Patient symptoms (findings, lab tests)**:
  - Fever, Cough, Paleness, WBC (white blood cells) count, Chest pain, etc.

**Representation of a patient case:**

- Statements that hold (are true) for that patient.

  E.g: Fever =*True*
  Cough =*False*
  WBCcount=*High*

**Diagnostic task:** we want to infer whether the patient suffers from the pneumonia or not given the symptoms



# Uncertainty

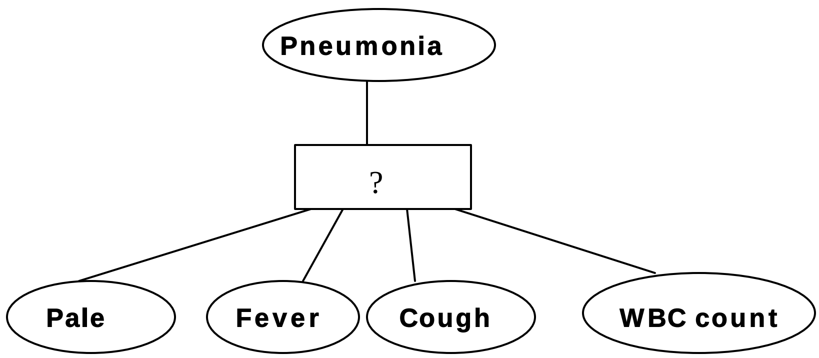
**To make diagnostic inference possible we need to represent rules or axioms that relate symptoms and diagnosis**

**Problem:** disease/symptoms relation is not deterministic (things may vary from patient to patient)

- **Disease → Symptoms uncertainty**
  - A patient suffering from pneumonia may not have fever all the times, may or may not have a cough, white blood cell test can be in a normal range.
- **Symptoms → Disease uncertainty**
  - High fever is typical for many diseases (e.g. bacterial diseases) and does not point specifically to pneumonia
  - Fever, cough, paleness, high WBC count combined do not always point to pneumonia

## Modeling the uncertainty.

- How to describe, represent the relations in the presence of uncertainty?
- How to manipulate such knowledge to make inferences?
  - Humans can reason with uncertainty.






## Methods for representing uncertainty

**KB systems** based on propositional and first-order logic often represent uncertain statements, axioms of the domain in terms of

- rules with various **certainty factors**

Very popular in 70-80s (MYCIN)

**If** 1. The stain of the organism is gram-positive, and
2. The morphology of the organism is coccus, and
3. The growth conformation of the organism is chains

**Then** **with certainty 0.7**
the identity of the organism is streptococcus

**Problems:**

- Chaining of multiple inference rules (propagation of uncertainty)
- Combinations of rules with the same conclusions
- After some number of combinations results not intuitive.

## Representing certainty factors

- **Facts** (propositional statements about the world) are assigned some certainty number reflecting the belief in that the statement is satisfied:

  $$CF(Pneumonia = True) = 0.7$$

- **Rules** incorporate tests on the certainty values

  $$(A \text{ in } [0.5,1]) \wedge (B \text{ in } [0.7,1]) \rightarrow C \text{ with } \mathrm{CF} = 0.8$$

- **Methods for combination of conclusions**

  $$(A \text{ in } [0.5,1]) \wedge (B \text{ in } [0.7,1]) \rightarrow C \text{ with } \mathrm{CF} = 0.8$$

  $$(E \text{ in } [0.8,1]) \wedge (D \text{ in } [0.9,1]) \rightarrow C \text{ with } \mathrm{CF} = 0.9$$

$$CF(C) = \max[0.9; 0.8] = 0.9$$
$$CF(C) = 0.9 * 0.8 = 0.72$$
$$CF(C) = 0.9 + 0.8 - 0.9 * 0.8 = 0.98$$

**?**



## Probability theory

a well-defined coherent theory for representing uncertainty and for reasoning with it

**Representation:**

**Proposition statements** – assignment of values to random variables

$$Pneumonia = True \qquad WBCcount = high$$

**Probabilities** over statements model the degree of belief in these statements

$$P(Pneumonia = True) = 0.001$$

$$P(WBCcount = high) = 0.005$$

$$P(Pneumonia = True, Fever = True) = 0.0009$$

$$P(Pneumonia = False, WBCcount = normal, Cough = False) = 0.97$$

## Joint probability distribution

**Joint probability distribution (for a set variables)**

- Defines probabilities for all possible assignments to values of variables in the set

$\mathbf{P}(pneumonia, WBCcount)$ $2 \times 3$ table

| | | *WBCcount* | | | $\mathbf{P}(Pneumonia)$ |
|---|---|---|---|---|---|
| | | *high* | *normal* | *low* | |
| *Pneumonia* | *True* | 0.0008 | 0.0001 | 0.0001 | 0.001 |
| | *False* | 0.0042 | 0.9929 | 0.0019 | 0.999 |
| $\mathbf{P}(WBCcount)$ | | 0.005 | 0.993 | 0.002 | |

**Marginalization** (summing of rows, or columns)
- summing out variables



## Conditional probability distribution

**Conditional probability distribution:**

- Probability distribution of A given (fixed B)

$$P(A \mid B) = \frac{P(A,B)}{P(B)}$$

- Conditional probability is defined in terms of joint probabilities
- Joint probabilities can be expressed in terms of conditional probabilities

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

- Conditional probability – is useful for **diagnostic reasoning**
  - the effect of a symptoms (findings) on the disease

$$P(Pneumonia = True \mid Fever = True, WBCcount = high, Cough = True)$$

# 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 reasoning

**Problems:**

- **Space complexity.** To store full joint distribution requires to remember $O(\mathrm{d}^n)$ numbers.

  *n* – number of random variables, *d* – number of values
- **Inference complexity.** To compute some queries requires $O(\mathrm{d}^n)$ 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

CS 2001 Bayesian belief networks

## Modeling uncertainty with probabilities

- **Knowledge based system era** (70s – early 80's)
  - Extensional non-probabilistic models
  - Probability techniques avoided because of space, time and acquisition bottlenecks in defining full joint distributions
  - Negative effect on the advancement of KB systems and AI in 80s in general
- Breakthrough (late 80s, beginning of 90s)
  - **Bayesian belief networks**
    - Give solutions to the space, acquisition bottlenecks
    - Significant improvements in the time cost of inferences



## Bayesian belief networks (BBNs)

**Bayesian belief networks.**

- Represent the full joint distribution more compactly with smaller number of parameters.
- Take advantage of conditional and marginal independences among components in the distribution

- **A and B are independent**

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

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

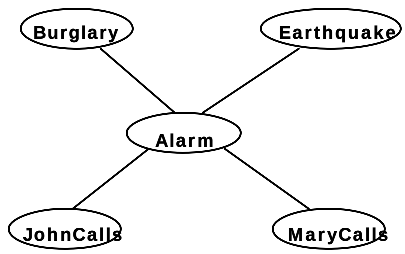
$$P(A,B \mid C) = P(A \mid C)P(B \mid C)$$

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

## Alarm system 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

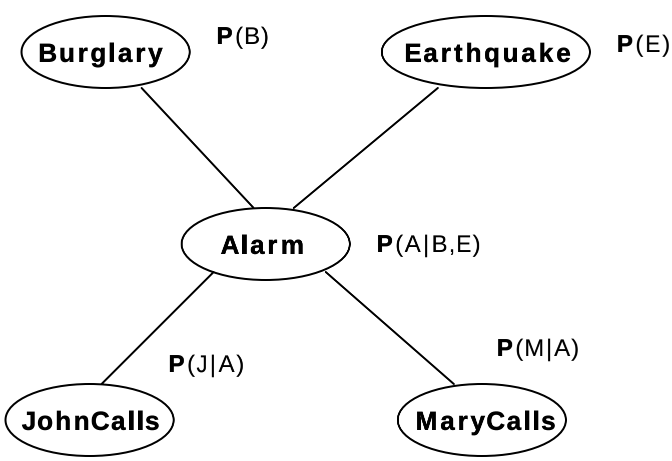
**Causal relations**



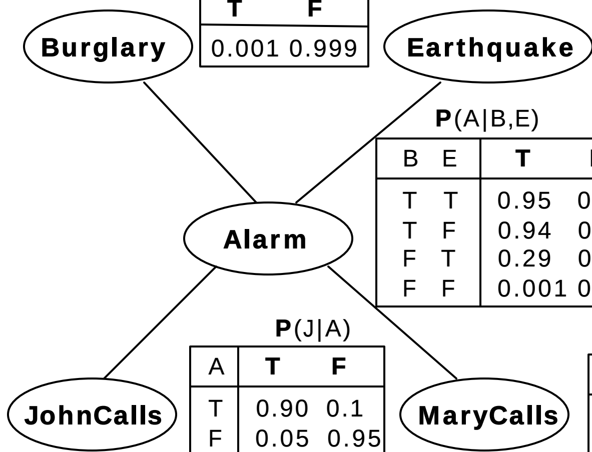



## Bayesian belief network.

1. Graph reflecting direct (causal) dependencies between variables
2. Local conditional distributions relating 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)

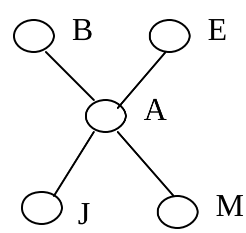
| A | T | F |
|---|---|---|
| T | 0.7 | 0.3 |
| F | 0.01 | 0.99 |



# Bayesian belief networks (general)

Two components: $B = (S, \Theta_S)$

- **Directed acyclic graph**
  - Nodes correspond to random variables
  - (Missing) links encode independences




- **Parameters**
  - Local conditional probability distributions for every variable-parent configuration

$$\mathbf{P}(X_i \mid pa(X_i))$$

Where:

$pa(X_i)$ - stand for parents of $X_i$

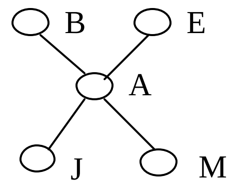
**P**(A|B,E)

| 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 |

## 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

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

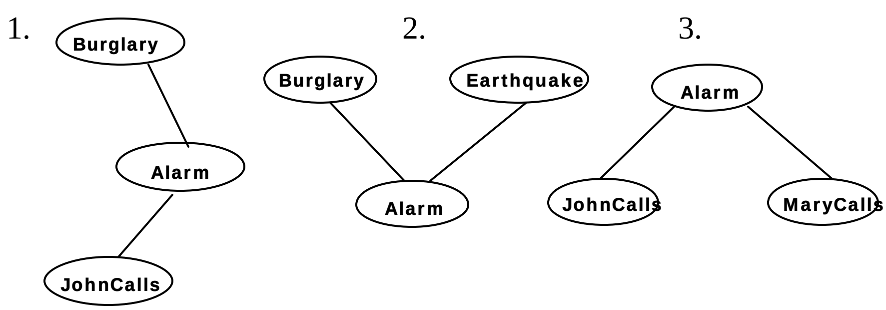
Then its probability is:

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



## Independences in BBNs

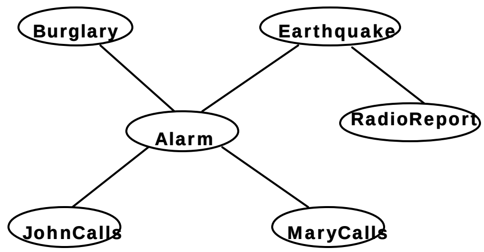
- **3 basic independence structures**




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

# Independences in BBNs

- Other dependences/independences in the network



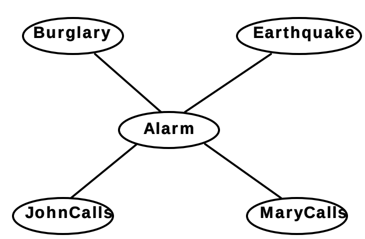

- Earthquake and Burglary are **dependent** given MaryCalls
- Burglary and MaryCalls **are dependent** (not knowing Alarm)
- Burglary and RadioReport **are independent** given Earthquake
- Burglary and RadioReport **are dependent** given MaryCalls



# Parameter complexity problem

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

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




**Parameters:**
**full joint:** $2^5 = 32$

**BBN:** $2^3 + 2(2^2) + 2(2) = 20$

**Parameters to be defined:**
**full joint:** $2^5 - 1 = 31$

**BBN:** $2^2 + 2(2) + 2(1) = 10$

# 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 random variables and their parents

- The complexity of local distributions is much smaller than the full joint
- Easier to estimate the probability parameters by consulting an expert or by learning them from data

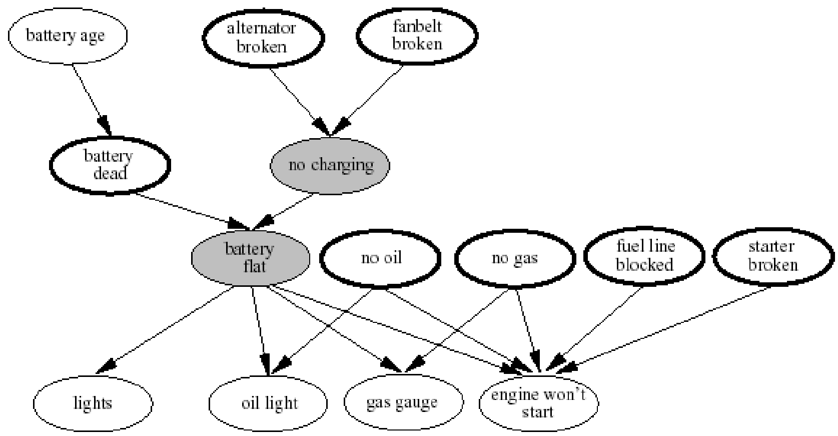


# BBNs built in practice

- **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

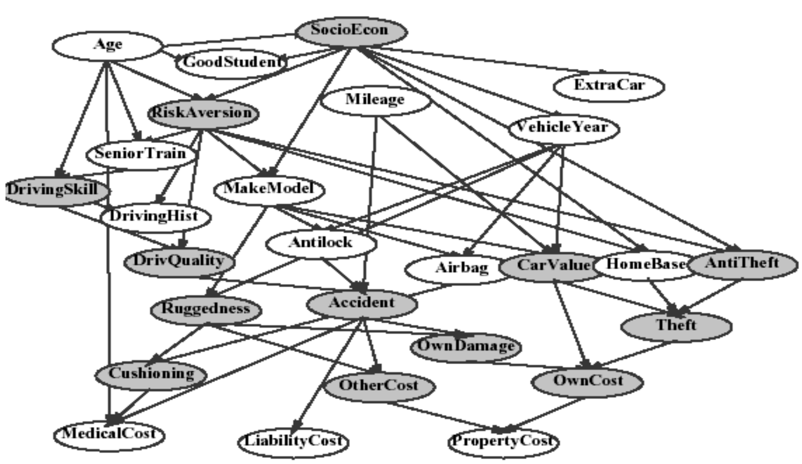
- Diagnose the engine start problem




CS 2001 Bayesian belief networks

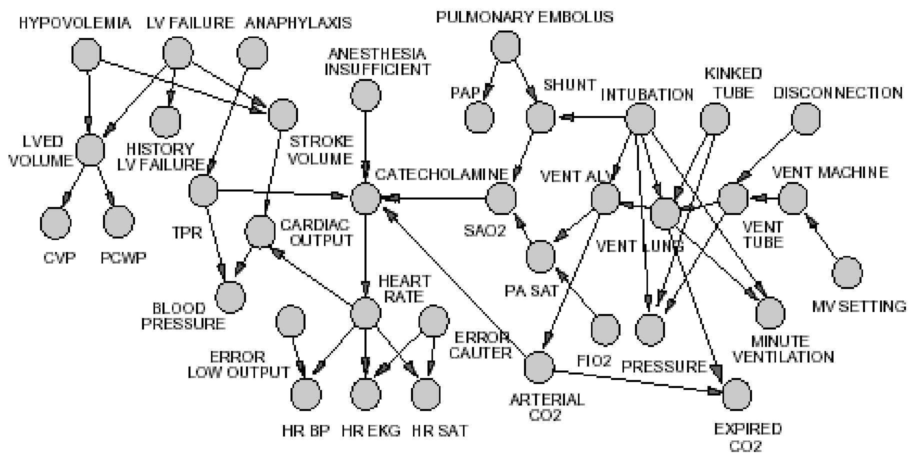
# Car insurance example

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




CS 2001 Bayesian belief networks

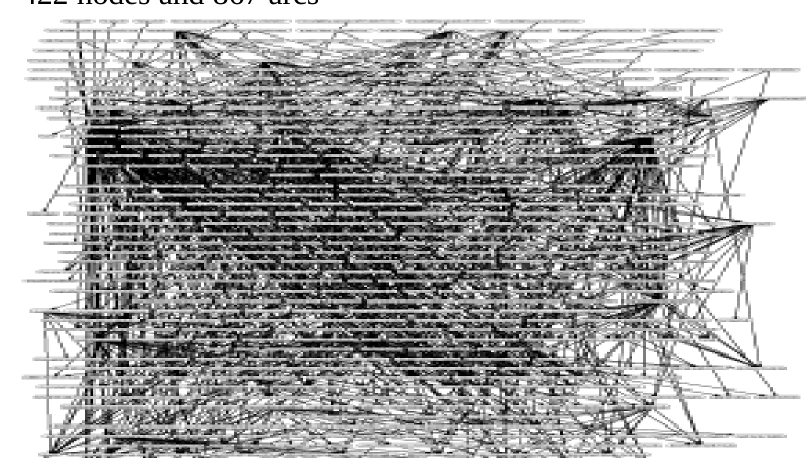
# (ICU) Alarm network






# 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