Building Bayesian Networks

Are Bayesian networks always suitable?

- Problem (modelling) objective, e.g., for function approximation or pure numeric prediction without a need to explain the results a “black box” model such as neural networks can be sufficient
- Sufficient knowledge about the problem (domain experts, data)
- Complexity of the problem e.g., is it decomposable

Problem solving

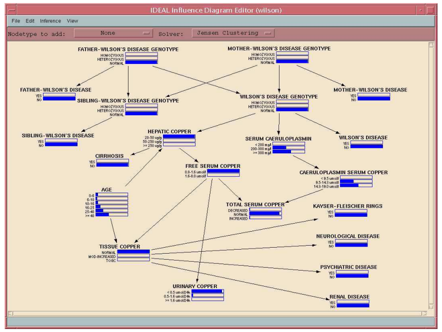
Bayesian networks: a declarative knowledge-representation formalism, i.e.,:

- mathematical basis
- problem to be solved determined by (1) entered evidence $\mathcal{E}$ (including potential decisions); (2) given hypothesis $H : P(H \mid \mathcal{E})$

Examples:

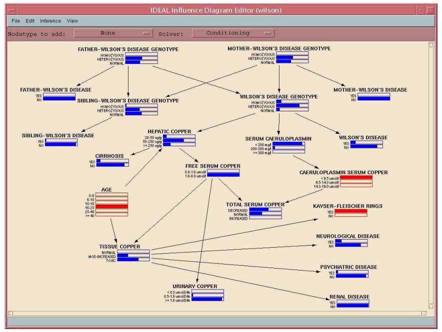
- Description of population
- Classification and diagnosis: $D = \arg\max_H P(H \mid \mathcal{E})$ i.e. D is the hypothesis with maximum $P(H \mid \mathcal{E})$
- Prediction
- Decision making based on *what-if* scenario’s

Prior information



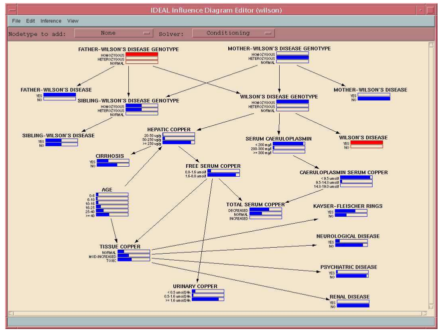
- Marginal probabilities $P(V)$ for every vertex V, e.g., $P($WILSON’S DISEASE $=$ *yes*$)$
- Gives description of the population on which the assessed probabilities are based

Diagnostic problem solving



- Marginal probabilities $P^*(V) = P(V \mid \mathcal{E})$ for every vertex V, e.g., P(WILSON'S DISEASE $=$ *yes* $\mid \mathcal{E})$ for entered evidence $\mathcal{E}$ (red vertices, with probability for one value equal to 1)
- Gives description of the *sub*population of the original population or individual cases

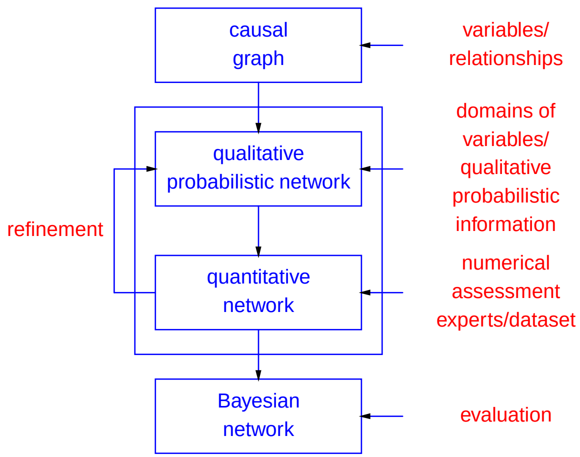
Prediction of associated findings



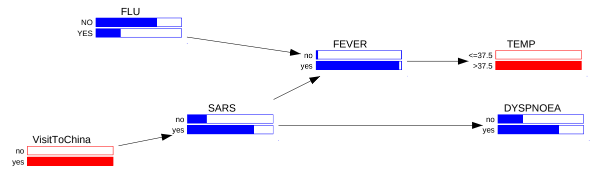
- Marginal probabilities $P^*(V) = P(V \mid \mathcal{E})$ for every vertex V, e.g., P(Kayer-Fleischer Rings $=$ *yes* $\mid \mathcal{E})$ with $\mathcal{E}$ evidence
- Gives description of the findings associated with a given class or category, such as Wilson's disease

Design of Bayesian network

- Current design principle: start modelling qualitatively (different from traditional knowledge-based systems)

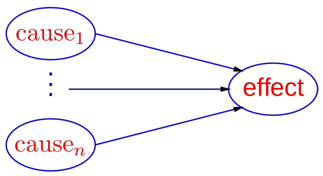


Terminology



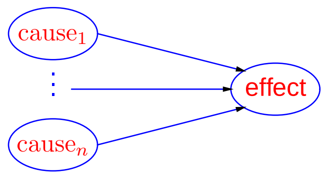
- Parent SARS of Child FEVER
- SARS is Ancestor of TEMP
- $\mathrm{DYSPNOEA}$ is Descendant of $\mathrm{VisitToChina}$
- Query node, e.g., FEVER
- Evidence, e.g., $\mathrm{VisitToChina}$ and TEMP
- Markov blanket, e.g.,
  for SARS: $\{\mathrm{VisitToChina}, \mathrm{DYSPNOEA}, \mathrm{FEVER}, \mathrm{FLU}\}$

Causal graph: Topology (structure)



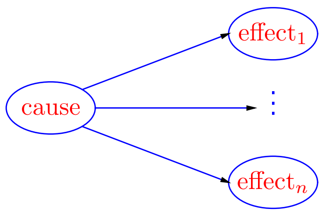
- Identify factors that are relevant
- Determine how those factors are causally related to each other
- The arc $\text{cause} \rightarrow \text{effect}$ does mean that cause is a factor involved in causing effect

Causal graph: Common effects



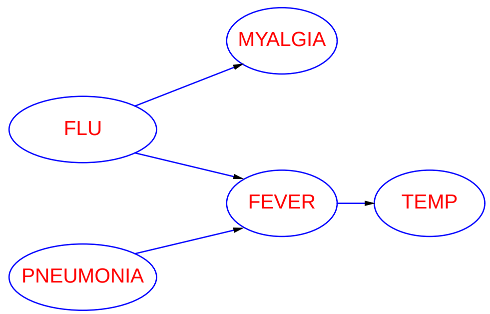
- An effect that has two or more ingoing arcs from other vertices is a common effect of those causes
- Kinds of causal interaction
  - Synergy: $\text{POLUTION} \longrightarrow \text{CANCER} \longleftarrow \text{SMOKING}$
  - Prevention: $\text{VACCINE} \longrightarrow \text{DEATH} \longleftarrow \text{SMALLPOX}$
  - XOR: $\text{ALKALI} \longrightarrow \text{DEATH} \longleftarrow \text{ACID}$

Causal graph: Common causes



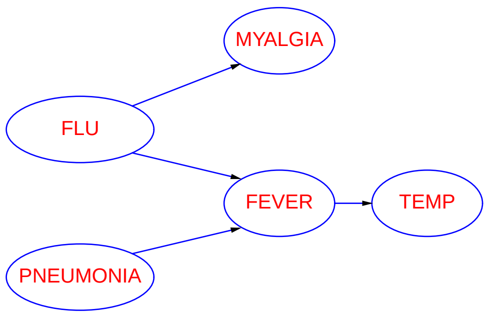
- A cause that has two or more outgoing arcs to other vertices is a common cause (factor) of those effects
- The effects of a common cause are usually observables (e.g. manifestations of failure of a device or symptoms in a disease)

Causal graph: Example



- FEVER and PNEUMONIA are two alternative causes of fever (but may enhance each other)
- FLU has two common effects: MYALGIA and FEVER
- High body TEMPerature is an indirect effect of FLU and PNEUMONIA, caused by FEVER

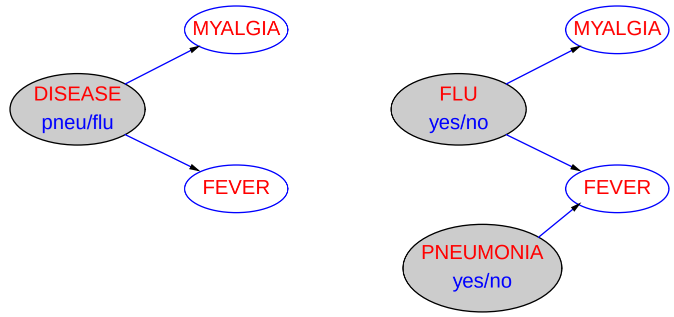
Check independence relationship



- Conditional independence: $X \perp\!\!\!\perp Y \mid Z$
  - $\{\mathrm{FLU}\} \perp\!\!\!\perp \{\mathrm{TEMP}\} \mid \{\mathrm{FEVER}\}$
  - $\{\mathrm{FEVER}\} \perp\!\!\!\perp \{\mathrm{MYALGIA}\} \mid \{\mathrm{FLU}\}$
  - $\{\mathrm{PNEUMONIA}\} \perp\!\!\!\perp \{\mathrm{FLU}\} \mid \varnothing$
  - $\{\mathrm{PNEUMONIA} \not\perp\!\!\!\perp \{\mathrm{FLU}\} \mid \{\mathrm{FEVER}\}$

Choose variables

- Factors are mutually exclusive (cannot occur together with absolute certainty): put as values in the same variable, or
- Factors may co-occur: multiple variables

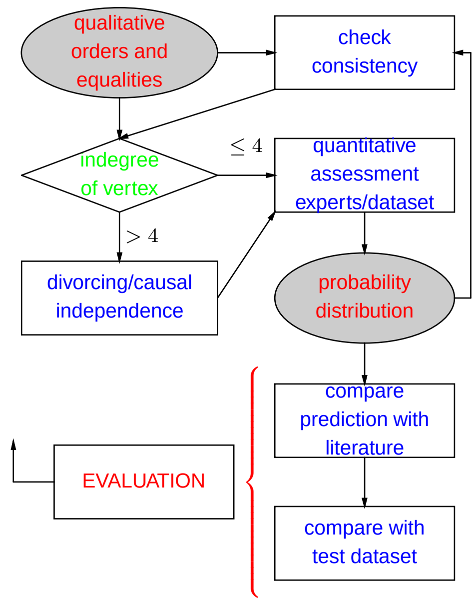


(a) Single variable (b) Multiple variables

Choose values

- Discrete values
  - Mutually exclusive and exhaustive
  - Types:
    - binary, e.g., $\mathrm{FLU} =$ *yes/no, true/false, 0/1*
    - ordinal, e.g., $\mathrm{INCOME} =$ *low, medium, high*
    - nominal, e.g., $\mathrm{COLOR} =$ *brown, green, red*
    - integral, e.g., $\mathrm{AGE} = \{1, \ldots, 120\}$
- Continuous values
- Discretization (of continuous and integral values)
  - Example for TEMP:
    $[-50, +5) \rightarrow$ *cold*
    $[+5, +20) \rightarrow$ *mild*
    $[+20, +50] \rightarrow$ *hot*

Probability assessment



Expert judgements

- Qualitative probabilities:
  - *Qualitative orders*:

AGE	$P($General Health Status $\mid \mathrm{AGE})$
10-69	good $>$ average $>$ poor
70-79	average $>$ good $>$ poor
80-89	average $>$ poor $>$ good
$\geq$ 90	poor $>$ average $>$ good

  - *Equalities*:

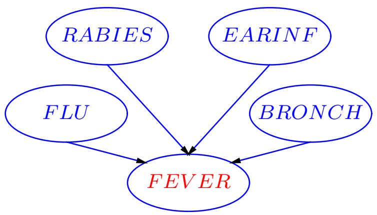
$$P(\mathrm{CANCER} = \mathrm{T1} \mid \mathrm{AGE} = 15-29) =$$
$$P(\mathrm{CANCER} = \mathrm{T2} \mid \mathrm{AGE} = 15-29)$$

Expert judgements (cont.)

- Quantitative, subjective probabilities:

	$P($GHS $\mid$ AGE$)$		
AGE	good	average	poor
10-69	0.99	0.008	0.002
70-79	0.3	0.5	0.2
80-89	0.1	0.5	0.4
$\geq$ 90	0.1	0.3	0.6

A bottleneck in Bayesian networks



- The number of parameters for the effect given n causes grows exponentially: $\geq 2^n$ (for binary causes)
- Unlikely evidence combination: $P(fever \mid flu, rabies, ear\_infection) = ?$

Problem: for many BNs too many probabilities have to be assessed

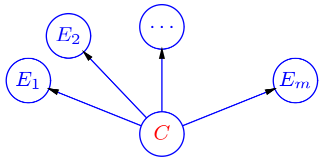
Special form Bayesian networks

Solution: use simpler probabilistic model, such that either

- the structure becomes simpler, e.g.,
  - *naive* (independent) form BN
  - *T*ree-*A*ugmented Bayesian *N*etwork (TAN)

  or,
- the assessment of the conditional probabilities becomes simpler (even though the structure is still complex), e.g.,
  - *parent divorcing*
  - *causal independence* BN

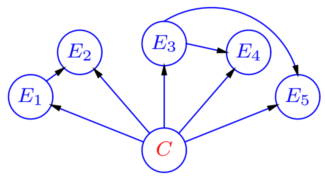
Independent (Naive) form BN



- C is a class variable
- E_i are evidence variables and $\mathcal{E} \subseteq \{E_1, \ldots, E_m\}$. We have $E_i \perp\!\!\!\perp E_j \mid C$, for $i \neq j$. Hence, using Bayes' rule:

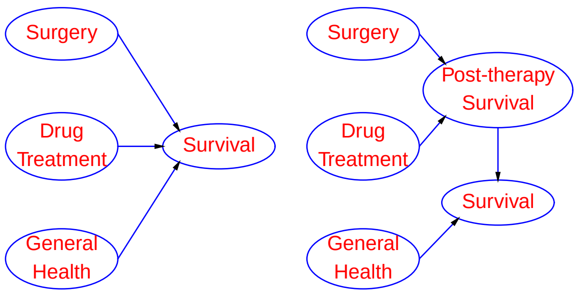
$$
\begin{array}{rcll}
P(C \mid \mathcal{E}) & = & \dfrac{P(\mathcal{E} \mid C)P(C)}{P(\mathcal{E})} & \text{with:} \\
P(\mathcal{E} \mid C) & = & \displaystyle\prod_{E \in \mathcal{E}} P(E \mid C) & \text{by cond. ind.} \\
P(\mathcal{E}) & = & \displaystyle\sum_{C} P(\mathcal{E} \mid C)P(C) & \text{marg. \& cond.}
\end{array}
$$

Tree-Augmented BN (TAN)



- Extension of Naive Bayes: reduce the number of independent assumptions
- Each node has at most two parents (one is the class node)

Divorcing multiple parents

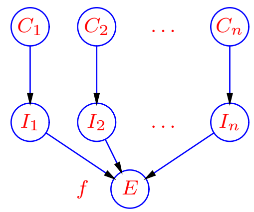


(a) Original network (b) Divorced network

Reduction in number of probabilities to assess:

- Identify a potential common effect of two or more parent vertices of a vertex
- Introduce a new variable into the network, representing the common effect

Causal Independence

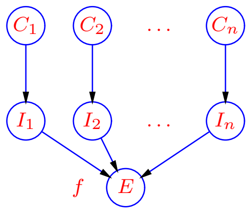


with:

- cause variables C_j, intermediate variables I_j, and the effect variable E
- $P(E \mid I_1, \ldots, I_n) \in \{0, 1\}$
- interaction function f, defined such that

$$
f(I_1, \ldots, I_n) = \begin{cases} e & \text{if } P(e \mid I_1, \ldots, I_n) = 1 \\ \neg e & \text{if } P(e \mid I_1, \ldots, I_n) = 0 \end{cases}
$$

Causal Independence: BN

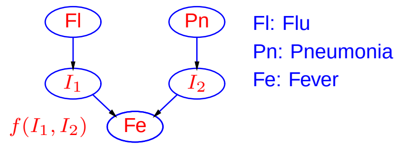


$$
\begin{aligned}
P(e \mid C_1, \ldots, C_n) &= \sum_{I_1,\ldots,I_n} P(e \mid I_1, \ldots, I_n) P(I_1, \ldots, I_n \mid C_1, \ldots, C_n) \\
&= \sum_{f(I_1,\ldots,I_n)=e} P(e \mid I_1, \ldots, I_n) P(I_1, \ldots, I_n \mid C_1, \ldots, C_n)
\end{aligned}
$$

Note that as $I_i \perp\!\!\!\perp I_j \mid \varnothing$, and $I_i \perp\!\!\!\perp C_j \mid C_i$, for $i \neq j$, it holds that:

$$P(I_1, \ldots, I_n \mid C_1, \ldots, C_n) = \prod_{k=1}^{n} P(I_k \mid C_k)$$

Example of causal independence



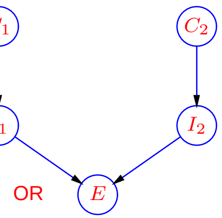
$$
\begin{aligned}
P(fe|\text{Fl}, \text{Pn}) &= \sum_{I_1,I_2} P(fe, I_1, I_2|\text{Fl}, \text{Pn}) \sum_{I_1,I_2} P(fe|I_1, I_2, \text{Fl}, \text{Pn}) P(I_1, I_2|\text{Fl}, \text{Pn}) \\
&= \sum_{I_1,I_2} P(fe|I_1, I_2) P(I_1|\text{Fl}) P(I_2|\text{Pn})
\end{aligned}
$$

Select *only* those $P(I_1|\text{Fl})P(I_2|\text{Pn})$ for which $f(I_1, I_2) = fe$, i.e. $P(fe|I_1, I_2) = 1$; do not select if $f(I_1, I_2) = \neg fe$, i.e. $P(fe|I_1, I_2) = 0$

Result:

$$P(fe|\text{Fl}, \text{Pn}) = \sum_{f(I_1,I_2)=fe} P(I_1|\text{Fl}) P(I_2|\text{Pn})$$

Causal independence: Noisy OR

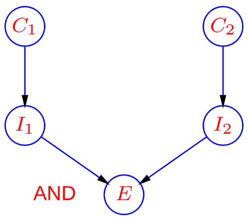


- Interactions among causes, as represented by the function f and $P(E \mid I_1, I_2)$, is a logical OR
- Meaning: presence of any one of the causes C_i with absolute certainty will cause the effect e (i.e. $E = true$)

$$
\begin{aligned}
P(e|C_1, C_2) &= \sum_{I_1 \vee I_2 = e} P(e|I_1, I_2) \prod_{k=1,2} P(I_k|C_k) \\
&= P(i_1|C_1)P(i_2|C_2) + P(\neg i_1|C_1)P(i_2|C_2) \\
&\quad + P(i_1|C_1)P(\neg i_2|C_2)
\end{aligned}
$$

Conclusion: assessment of $P(I \mid C)$ instead of

Causal independence: Noisy AND



- Interactions among causes, as represented by the function f and $P(E \mid I_1, I_2)$, is a logical AND
- Meaning: presence of all causes C_i with absolute certainty will cause the effect e (i.e. $E = true$); otherwise, $\neg e$

$$
\begin{aligned}
P(e|C_1, C_2) &= \sum_{I_1 \wedge I_2 = e} P(e|I_1, I_2) \prod_{k=1,2} P(I_k|C_k) \\
&= P(i_1|C_1)P(i_2|C_2)
\end{aligned}
$$

Refining causal graphs

"*Essentially, all models are wrong but some are useful*"

G.Box, N.Draper (1987)

So, model refinement is necessary.

- How:
  - Manual
  - Automatic
- What:
  - Probability adjustment
  - Removing irrelevant factors
  - Adding previously hidden, unknown factors
  - Causal relationships adjustment, e.g., including, removing independence relations