Building Bayesian Networks

The focus today . . .

- Problem solving by Bayesian networks
- Designing Bayesian networks
  - Qualitative part (structure)
  - Quantitative part (probability assessment)
- Simplified Bayesian networks
  - In structure: Naïve Bayes, Tree-Augmented Networks
  - In probability assessment: Parent divorcing, Causal Independence

Problem solving

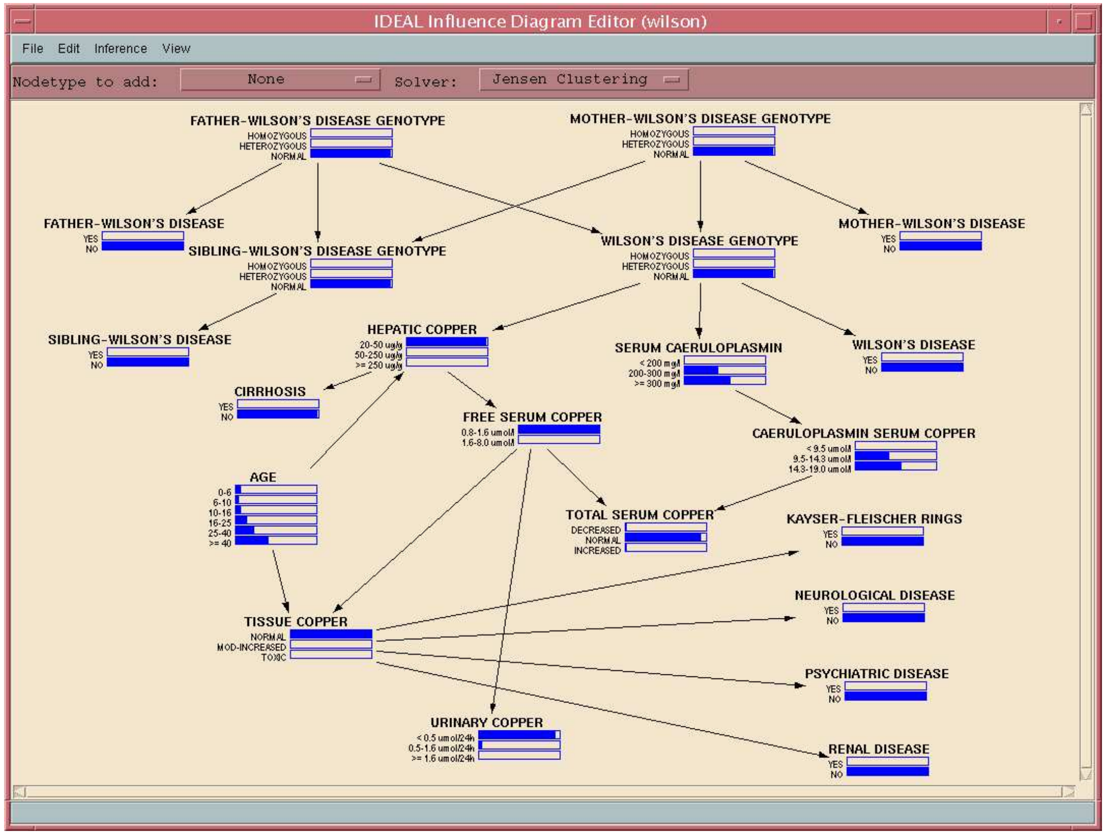
Bayesian networks: a declarative (*knowing what*) knowledge-representation formalism, i.e.,:

- 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:

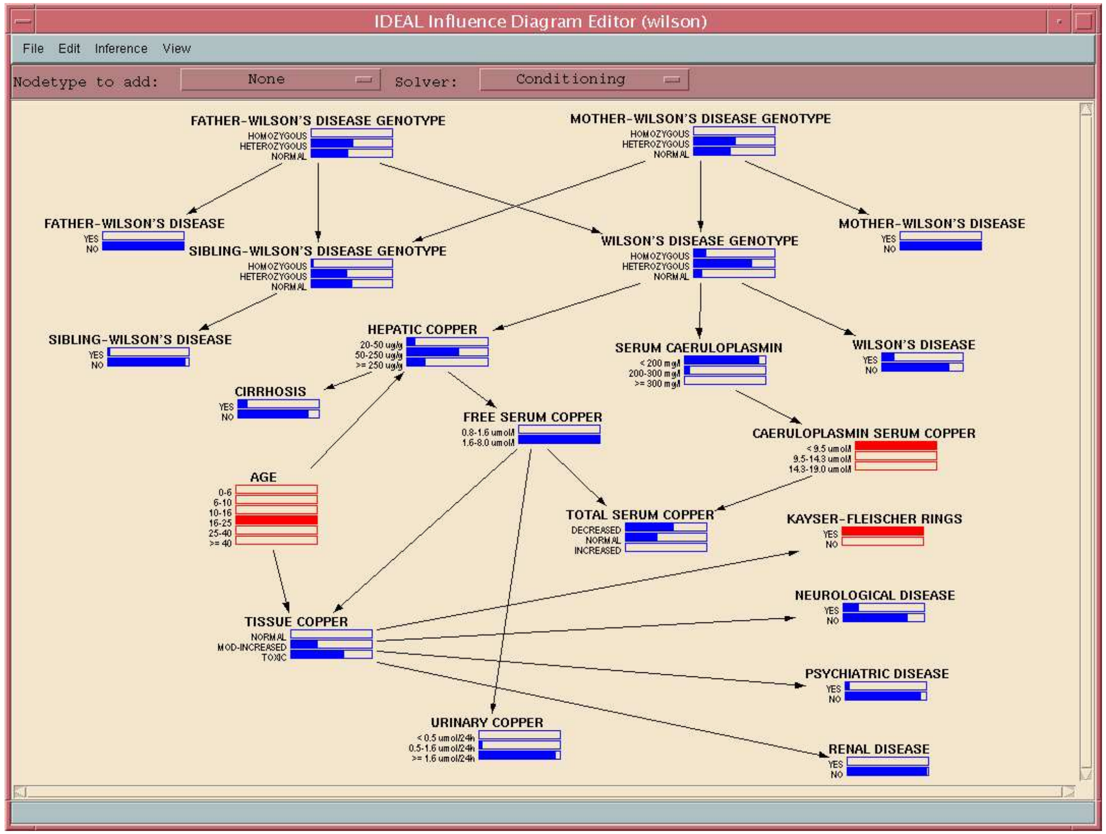
- Description of population (or prior information)
- Classification and diagnosis: $D = \arg\max_H P(H \mid \mathcal{E})$ i.e. D is the hypothesis with maximum $P(H \mid \mathcal{E})$
- Prediction
- Decision making based on *what-if* scenario's

Prior information



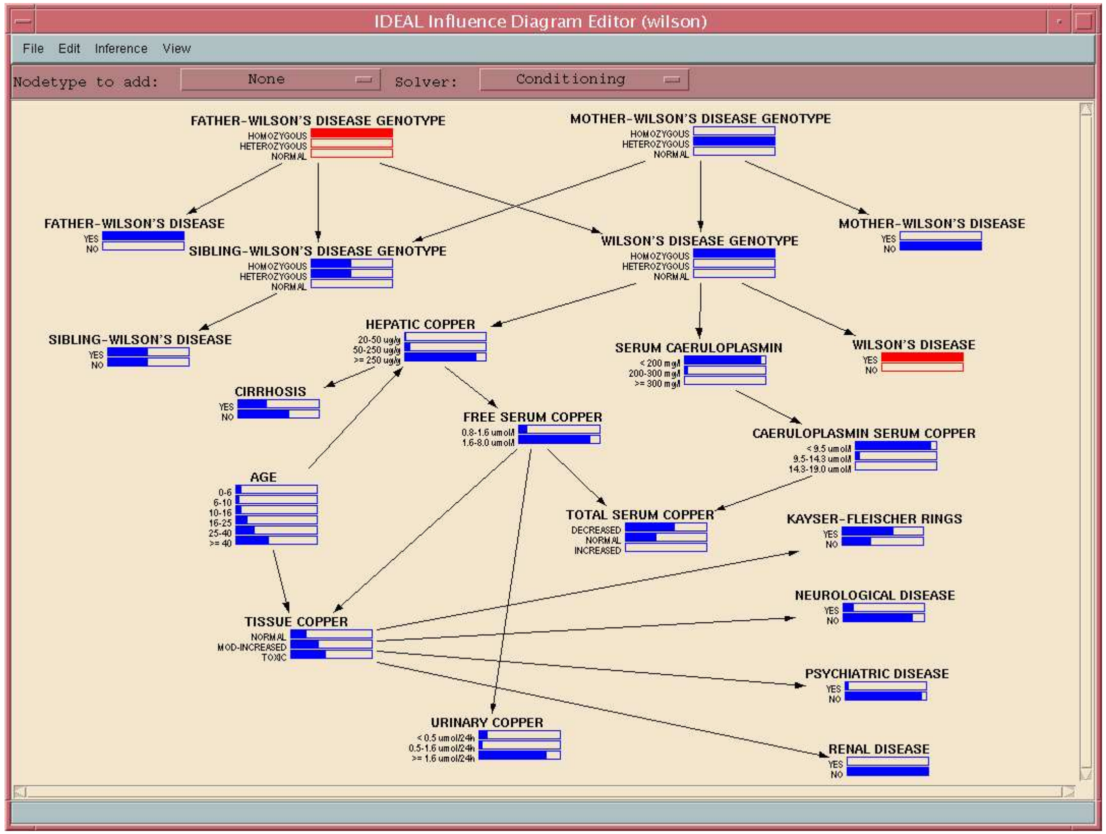
- Gives description of the population on which the assessed probabilities are based, i.e., the original probabilities before new evidence is uncovered
- Marginal probabilities $P(V)$ for every vertex V, e.g., $P(\text{WILSON'S DISEASE} = \textit{yes})$

Diagnostic problem solving



- Gives description of the *sub*population of the original population or individual cases
- 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)

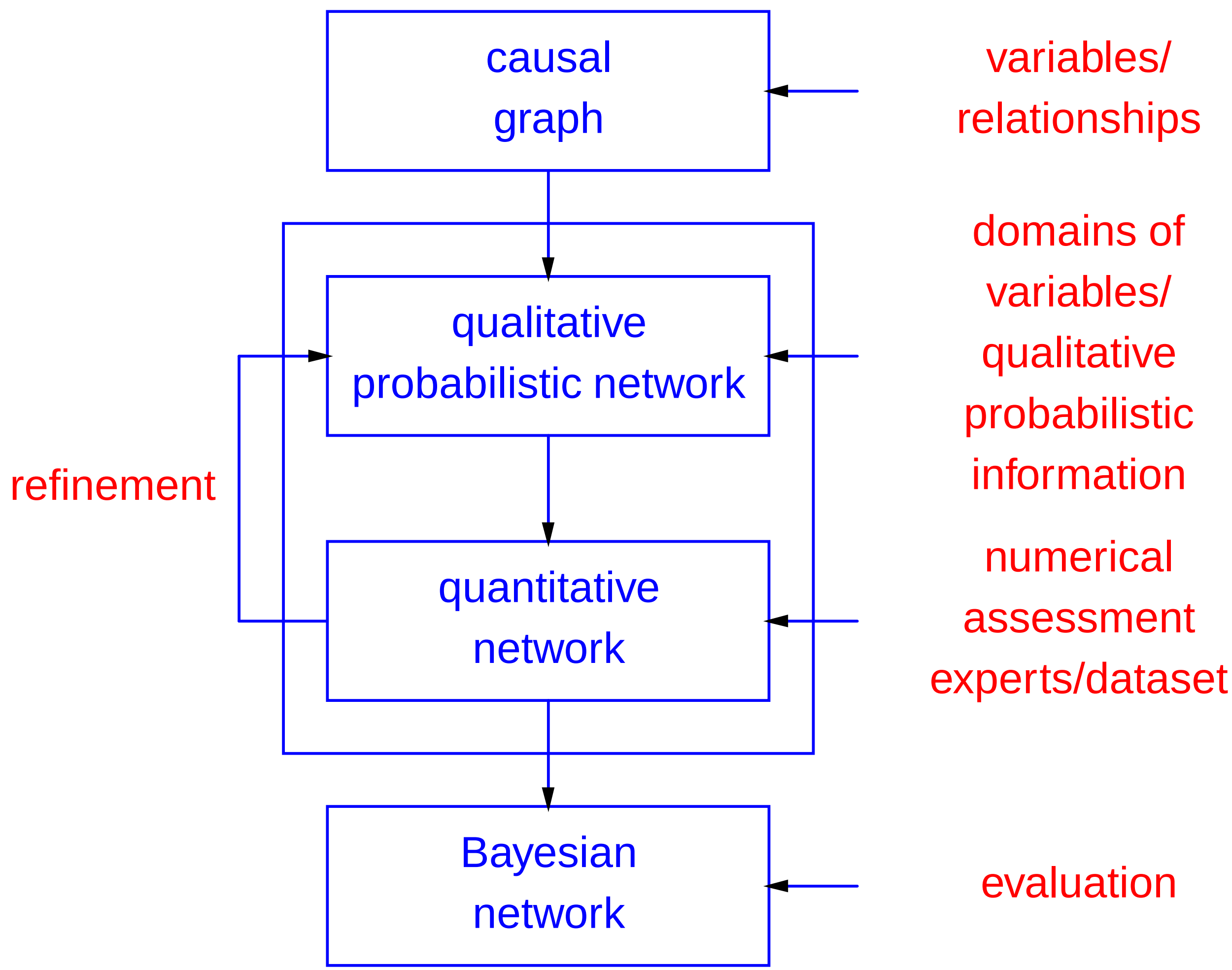
Prediction of associated findings



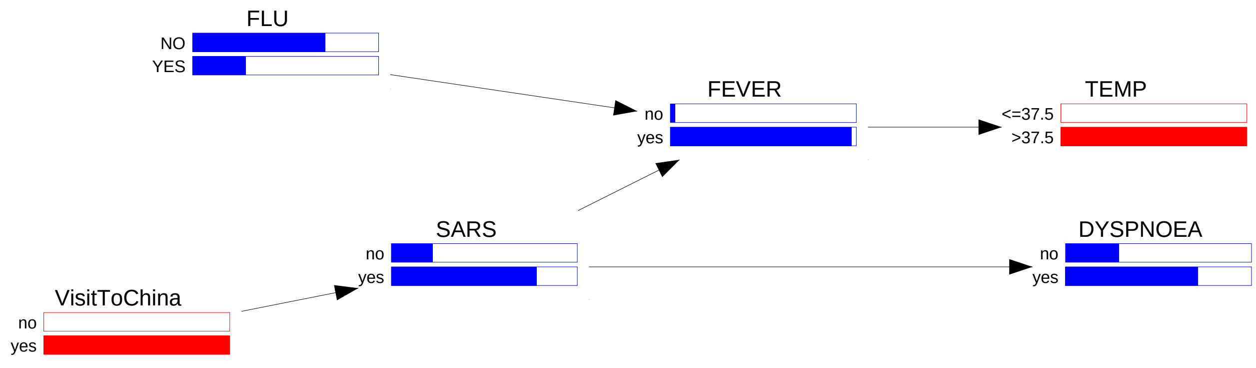
- Gives description of the findings associated with a given class or category, such as Wilson's disease
- Marginal probabilities $P^*(V) = P(V \mid \mathcal{E})$ for every vertex V, e.g., $P($Kayser-Fleischer Rings $=$ *yes* $\mid \mathcal{E})$ with $\mathcal{E}$ evidence

Design of Bayesian network

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

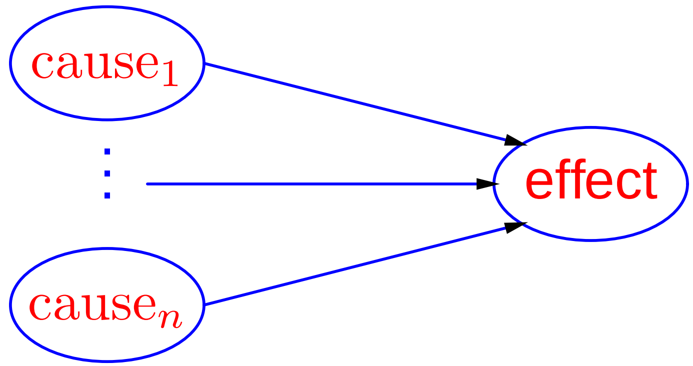


Terminology



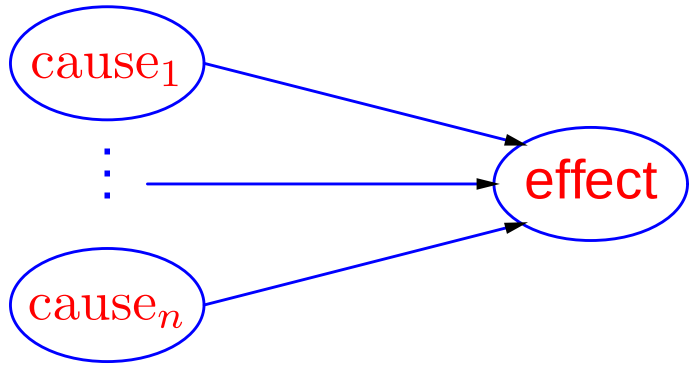
- Parent SARS of Child FEVER
- SARS is Ancestor of TEMP
- $\mathrm{DYSPNOEA}$ is Descendant of $\mathrm{VisitToChina}$
- Query node, e.g., FEVER
- Evidence, e.g., $\mathrm{VisitToChina}$ and TEMP
- Markov blanket, e.g.,
  for SARS: $\{\mathrm{VisitToChina}, \mathrm{DYSPNOEA}, \mathrm{FEVER}, \mathrm{FLU}\}$

Causal graph: Topology (structure)



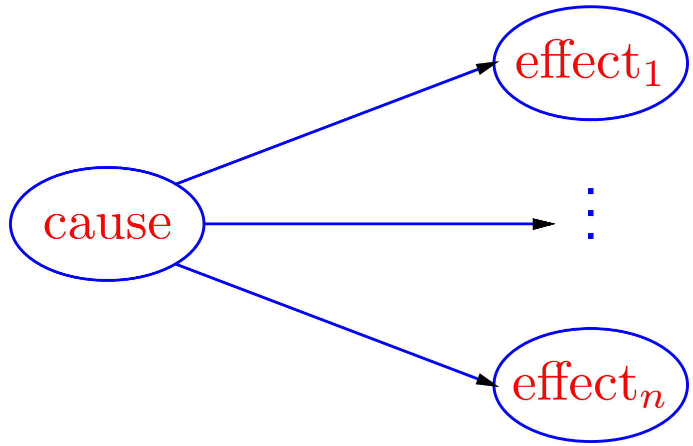
- Identify factors that are relevant
- Determine how those factors are causally related to each other
- The arc $\mathrm{cause} \rightarrow \mathrm{effect}$ does mean that cause is a factor involved in causing effect

Causal graph: Common effect



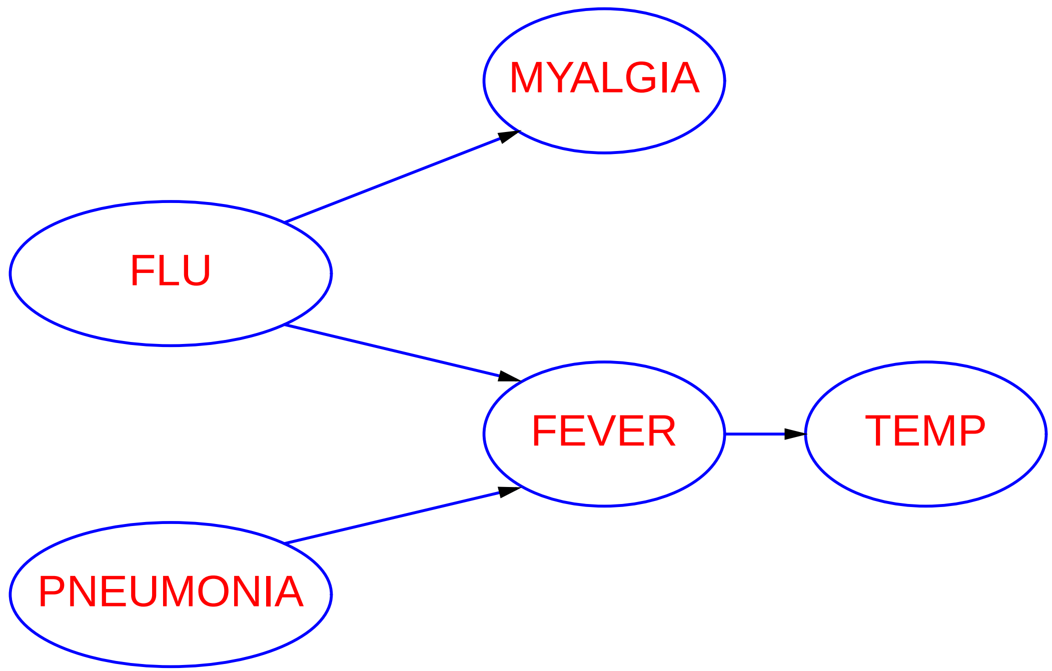
- An effect that has two or more ingoing arcs from other vertices is a common effect of those causes
- Kinds of causal interaction
  - Synergy: POLUTION $\longrightarrow$ CANCER $\longleftarrow$ SMOKING
  - Prevention: VACCINE $\longrightarrow$ DEATH $\longleftarrow$ SMALLPOX
  - XOR: ALKALI $\longrightarrow$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

Causal graph: Common cause



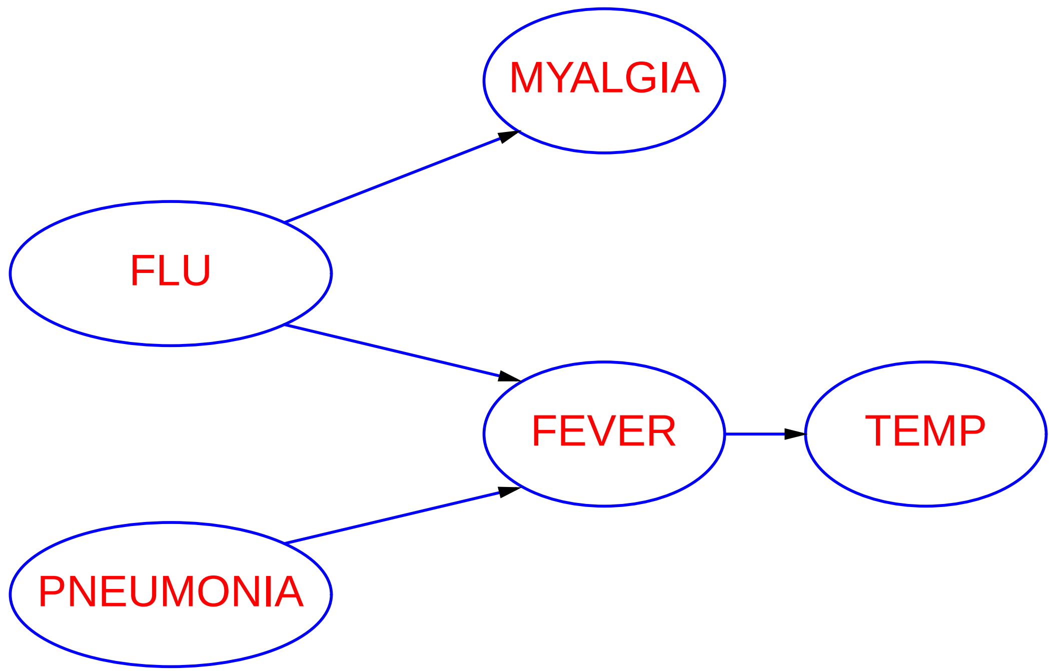
- A cause that has two or more outgoing arcs to other vertices is a common cause (factor) of those effects
- The effects of a common cause are usually observables (e.g. manifestations of failure of a device or symptoms in a disease)

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

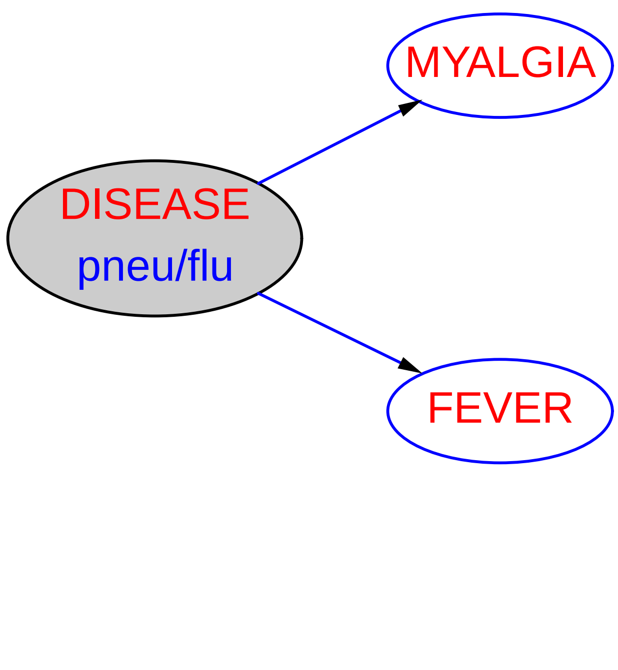
Check independence relationship



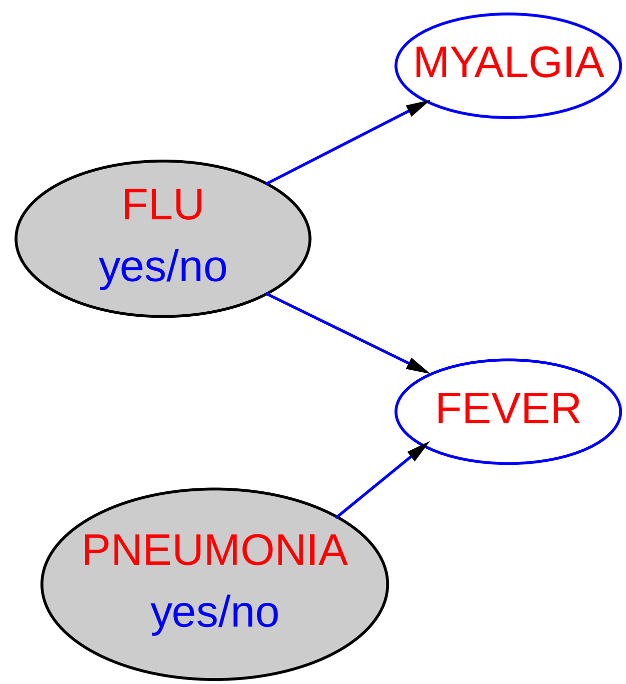
- Conditional independence: $X \perp\!\!\!\perp Y \mid Z$
  - $\{\mathrm{FEVER}\} \perp\!\!\!\perp \{\mathrm{MYALGIA}\} \mid \{\mathrm{FLU}\}$
  - **?** $\mid \{\mathrm{FEVER}\}$
  - $\{\mathrm{PNEUMONIA}\} \perp\!\!\!\perp \{\mathrm{FLU}\} \mid$ **?**
  - $\{\mathrm{PNEUMONIA}\} \not\perp\!\!\!\perp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

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

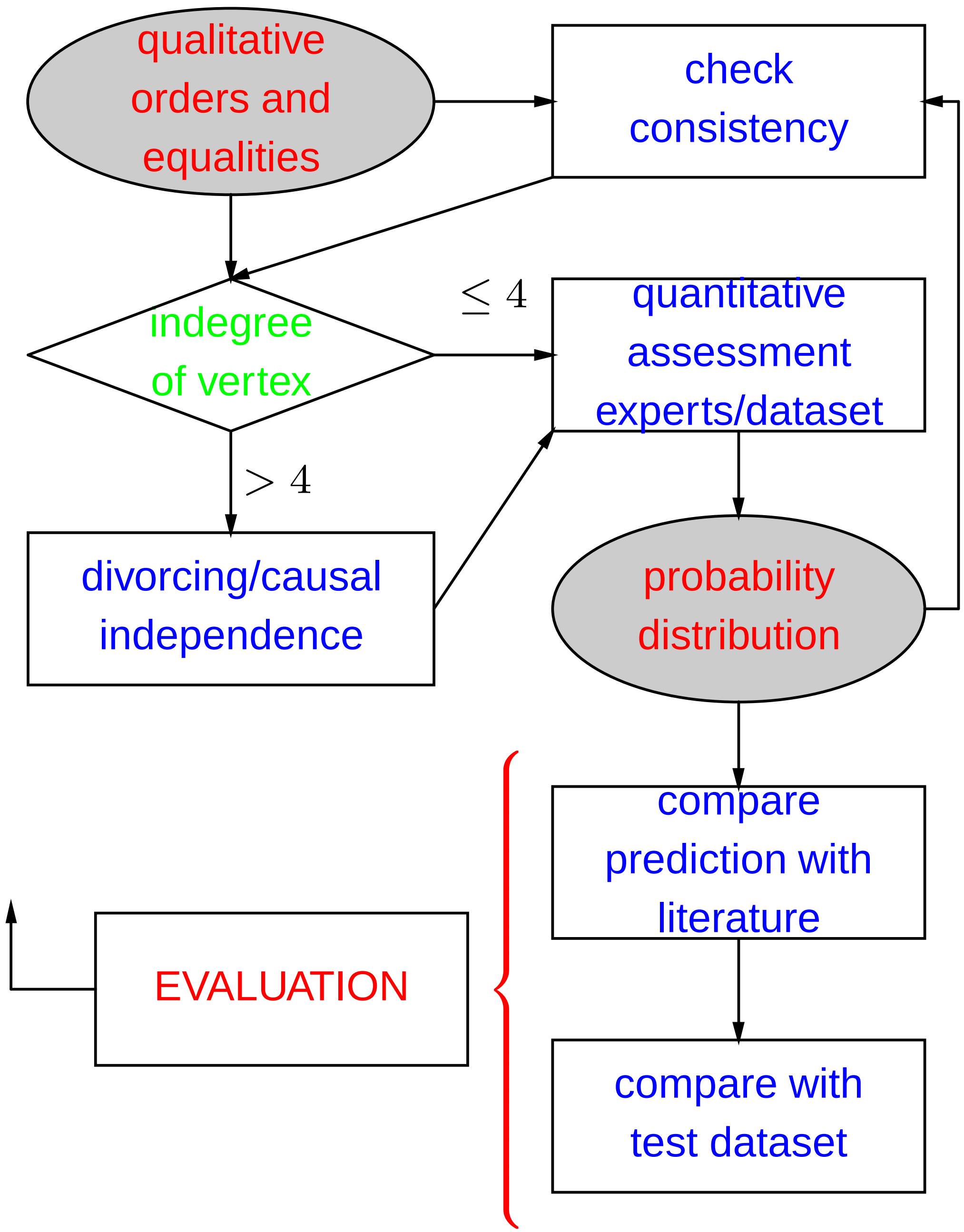
- Discrete values
  - Mutually exclusive and exhaustive
  - Types:
    - binary, e.g., $\mathrm{FLU} =$ *yes/no, true/false, 0/1*
    - ordinal, e.g., $\mathrm{INCOME} =$ *low, medium, high*
    - nominal, e.g., $\mathrm{COLOR} =$ *brown, green, red*
    - integral, e.g., $\mathrm{AGE} = \{1, \ldots, 120\}$
- Continuous values
- Discretization (of continuous 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*:

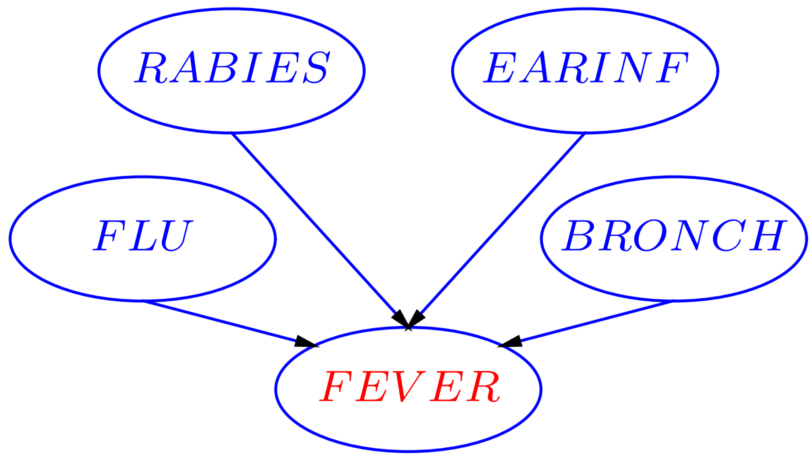
$$P(\mathrm{CANCER} = \mathrm{T1} | \mathrm{AGE} = 15 - 29) =$$
$$P(\mathrm{CANCER} = \mathrm{T2} | \mathrm{AGE} = 15 - 29)$$

Expert judgements (cont.)

- Quantitative, subjective probabilities:

	$P(\text{GHS} \mid \text{AGE})$		
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: $P(fever|flu, rabies, ear\_infection) = ?$

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

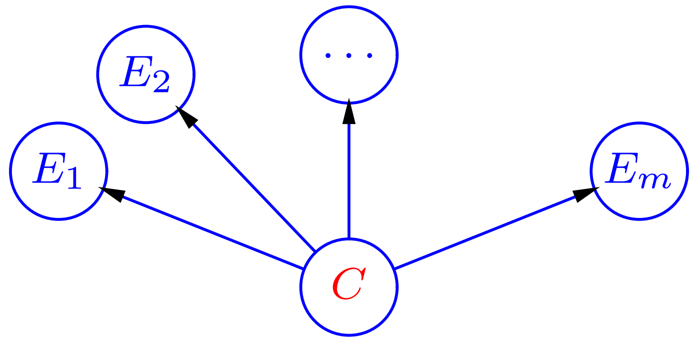
Special form Bayesian networks

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:

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

Example of Naive Bayes

PlayTennis: training examples

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Example of Naive Bayes (1)

- Learning phase

Outlook	Play=*Yes*	Play=*No*
Sunny	2/9	3/5
Overcast	4/9	0/5
Rain	3/9	2/5

Temperature	Play=*Yes*	Play=*No*
Hot	2/9	2/5
Mild	4/9	2/5
Cool	3/9	1/5

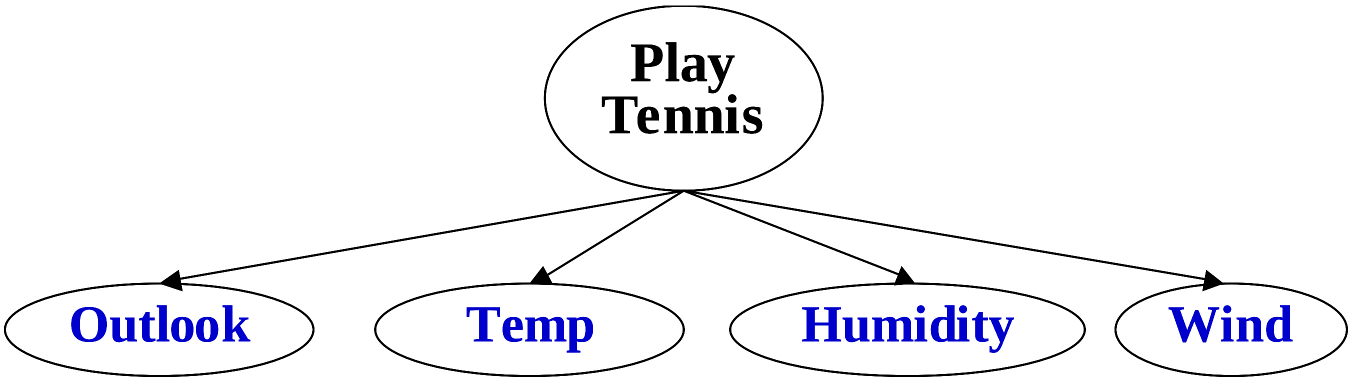
Humidity	Play=*Yes*	Play=N*o*
High	3/9	4/5
Normal	6/9	1/5

Wind	Play=*Yes*	Play=*No*
Strong	3/9	3/5
Weak	6/9	2/5

$$P(\text{Play}=Yes) = 9/14 \qquad P(\text{Play}=No) = 5/14$$

Example of Naive Bayes (2)

- Testing phase (inference)



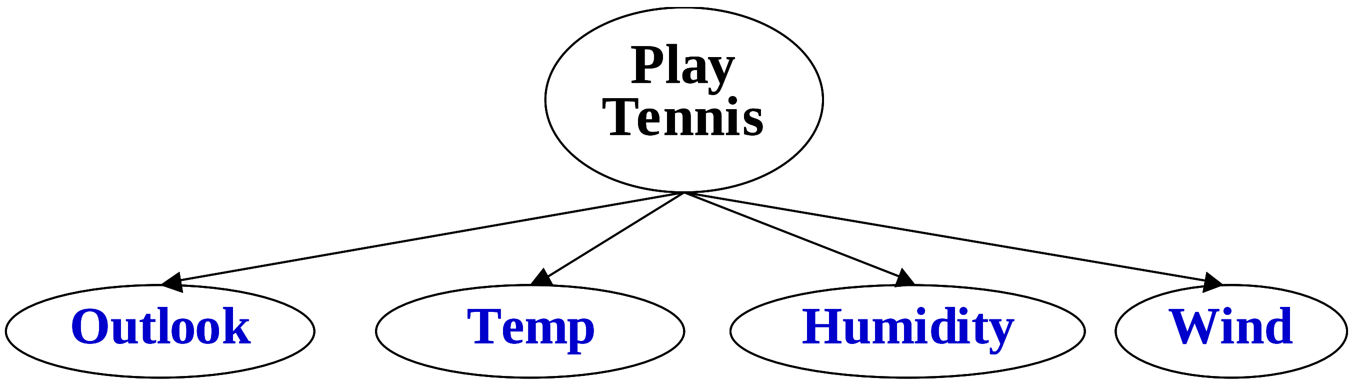
Evidence:

x = (**Outlook**=*Sunny*, **Temperature**=*Cool*, **Humidity**=*High*, **Wind**=*Strong*)

Then given **x**, PlayTennis=?

Example of Naive Bayes (3)

- Testing phase (inference)



$P(Yes|\mathbf{x}) = P(\mathbf{x}|\text{PlayTennis}=Yes)*P(\text{PlayTennis}=Yes)/\text{P}(\mathbf{x}) \propto$

$\propto [\text{P}(Sunny|\text{Yes})\text{P}(Cool|Yes)\text{P}(High|\text{Yes})\text{P}(Strong|Yes)]\text{P}(\text{Play}=Yes) =$

$= 2/9 * 3/9 * 3/9 * 3/9 * 9/14 = \mathbf{0.0053}$

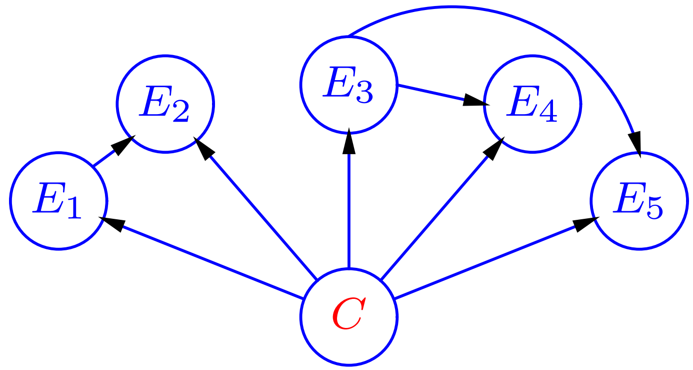
$\text{P}(No|\mathbf{x}) \propto [\text{P}(Sunny|\text{No})\ \text{P}(Cool|\text{No})\text{P}(High|No)\text{P}(Strong|No)]\text{P}(\text{Play}=No) = \mathbf{0.0206}$

Given that $P(Yes|\mathbf{x}') < P(No|\mathbf{x}')$, then for $\mathbf{x}$ we label **PlayTennis** = *No*

Note: to get probabilities we need to normalise

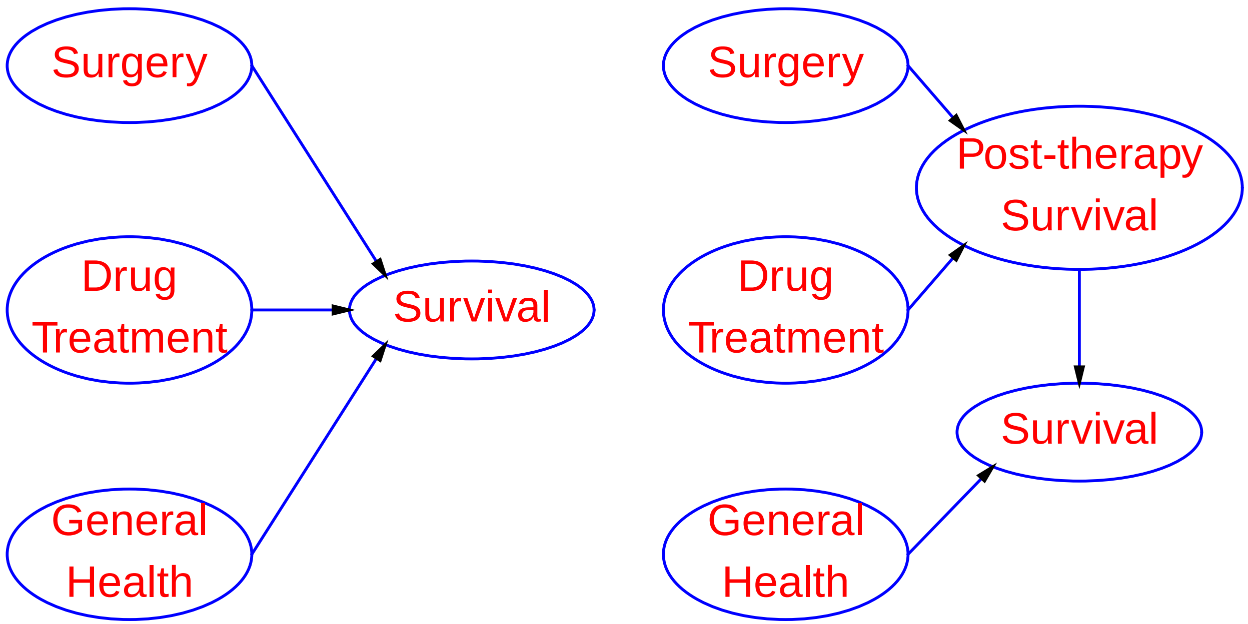
$\boldsymbol{P}(Yes|\mathbf{x}) = P(Yes|\mathbf{x}) \,/\, (P(Yes|\mathbf{x}) + P(No|\mathbf{x})) = 0.0053/(0.0053+0.0206) = 0.20$

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

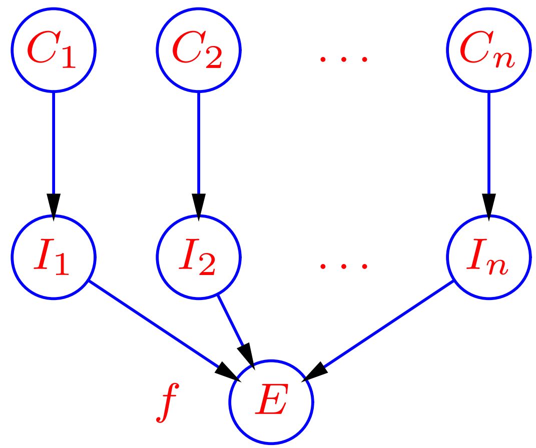


(a) Original network (b) Divorced network

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

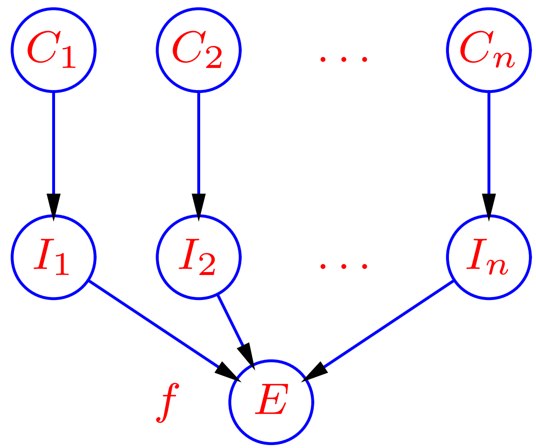


with:

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

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

Causal Independence: BN



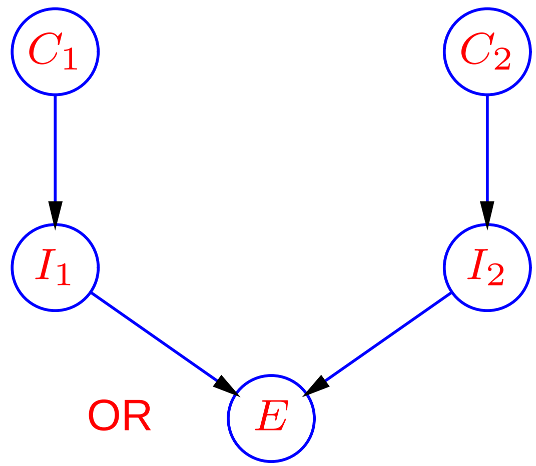
$$
\begin{aligned}
P(e \mid C_1, \ldots, C_n) &= \sum_{I_1, \ldots, I_n} P(e \mid I_1, \ldots, I_n) P(I_1, \ldots, I_n \mid C_1, \ldots, C_n) \\
&= \sum_{f(I_1, \ldots, I_n) = e} P(e \mid I_1, \ldots, I_n) \underline{P(I_1, \ldots, I_n \mid C_1, \ldots, C_n)}
\end{aligned}
$$

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:

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

Conclusion: assessment of $P(I_i|C_i)$ instead of $P(E|C_1, \ldots, C_n)$, i.e., $2n$ vs. 2^n probabilities

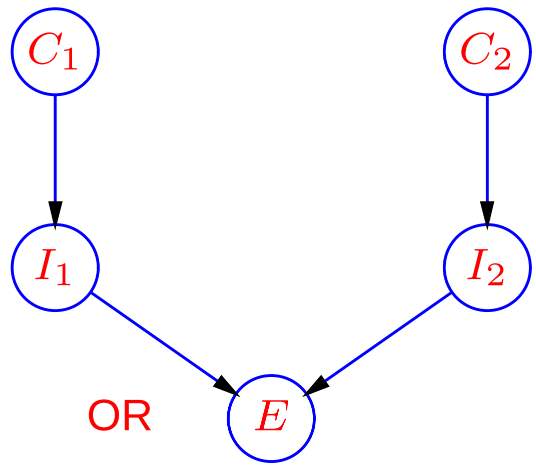
Causal independence: Noisy OR



- 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. $E = true$)

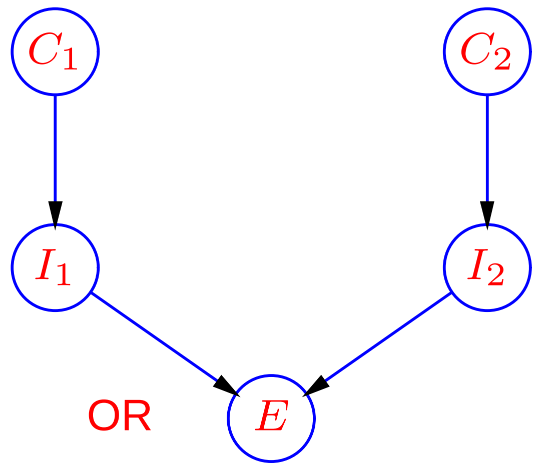
$$P(e|C_1, C_2) = \textbf{?}$$

Causal independence: Noisy OR (cont.)



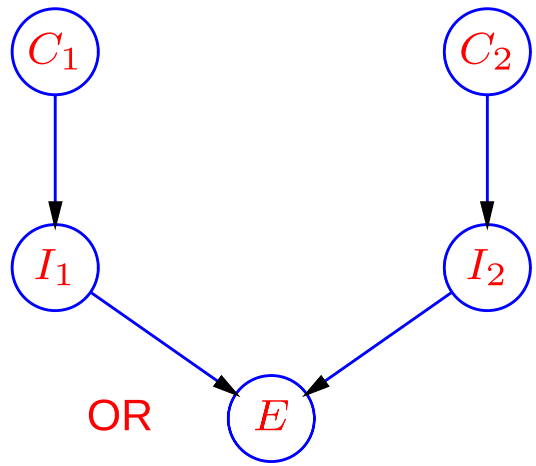
$$
\begin{aligned}
P(e|C_1, C_2) &= \sum_{I_1, I_2} P(e|I_1, I_2, C_1, C_2) P(I_1, I_2|C_1, C_2) \\
&= \textbf{?}
\end{aligned}
$$

Causal independence: Noisy OR (cont.)



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

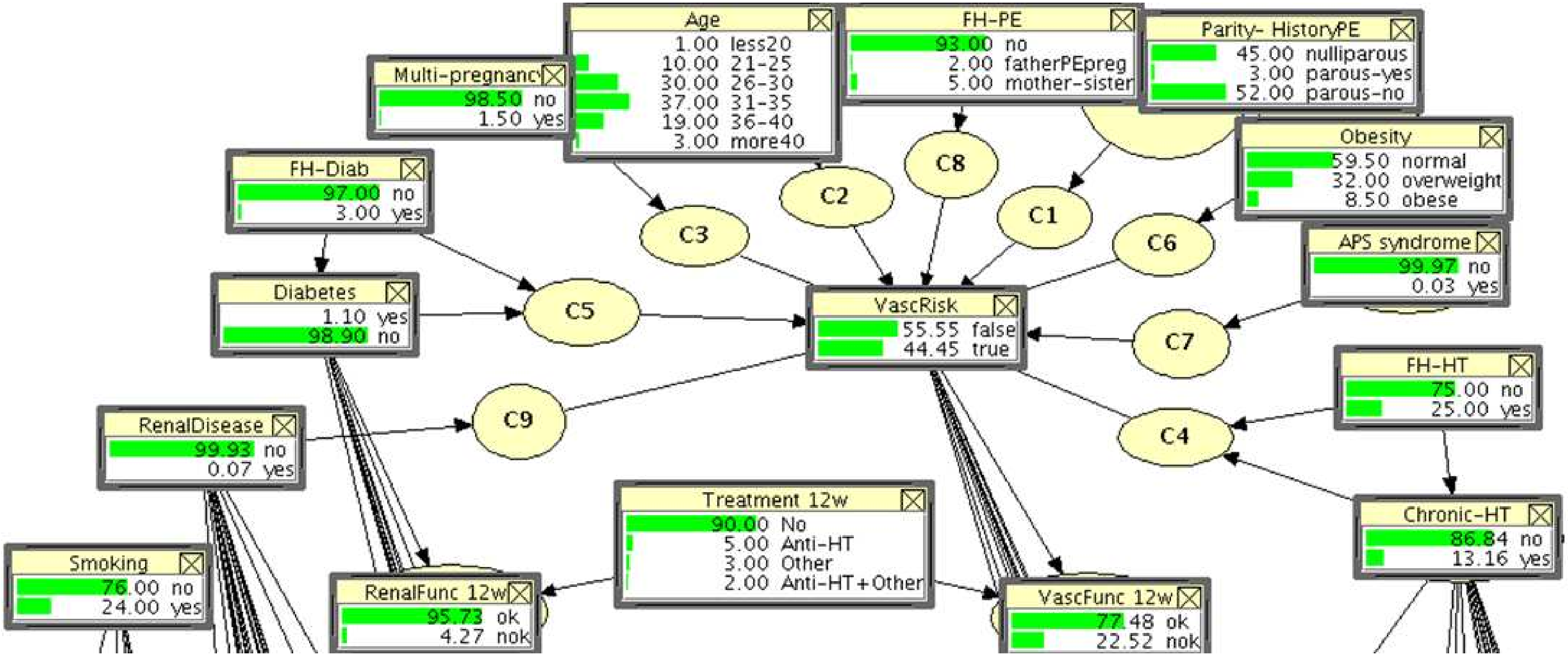
Causal independence: Noisy OR (cont.)



I_1	I_2	$P(e \mid I_1, I_2)$	$f(I_1, I_2) = I_1 \vee I_2$
0	0	0	$\neg e$
0	1	1	$\Rightarrow\ e$
1	0	1	$\Rightarrow\ e$
1	1	1	$\Rightarrow\ e$

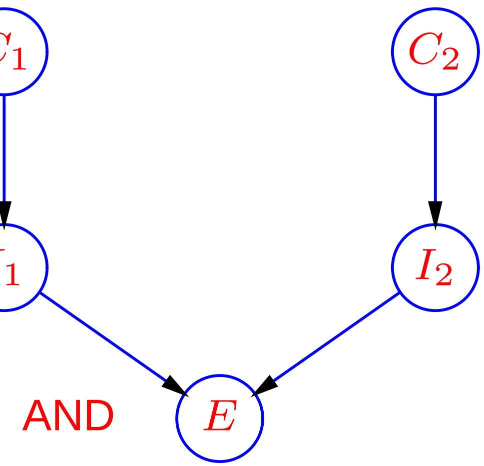
$$\begin{aligned} P(e|C_1, C_2) &= \sum_{f(I_1, I_2)=e} P(e|I_1, I_2) \prod_{k=1,2} P(I_k|C_k) \\ &= P(i_1|C_1)P(i_2|C_2) + P(\neg i_1|C_1)P(i_2|C_2) + P(i_1|C_1)P(\neg i_2|C_2) \end{aligned}$$

Noisy OR: Real-world example



- Dynamic Bayesian network for predicting the development of hypertensive disorders during pregnancy
- `VascRisk` (Vascular risk) has 11 causes and its original CPT requires the estimation of 20736!!! entries. Practically impossible!
- Solution: use noisy OR to simplify it

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, $\neg e$

$$P(e|C_1, C_2) = \textbf{?}$$

Are Bayesian networks always suitable?

“*Essentially, all models are wrong but some are useful*”

George Box, Norman Draper (1987),
Empirical Model-Building and Response Surfaces, Wiley

- 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, literature, data)
- Complexity of the problem e.g., is it decomposable

Refining causal graphs

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