
Building Bayesian Networks

Kick-off Assignment I

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 $H : P(H | \mathcal{E})$

Examples:

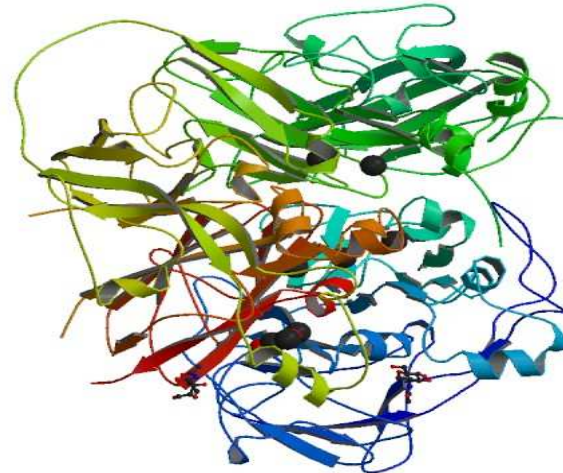
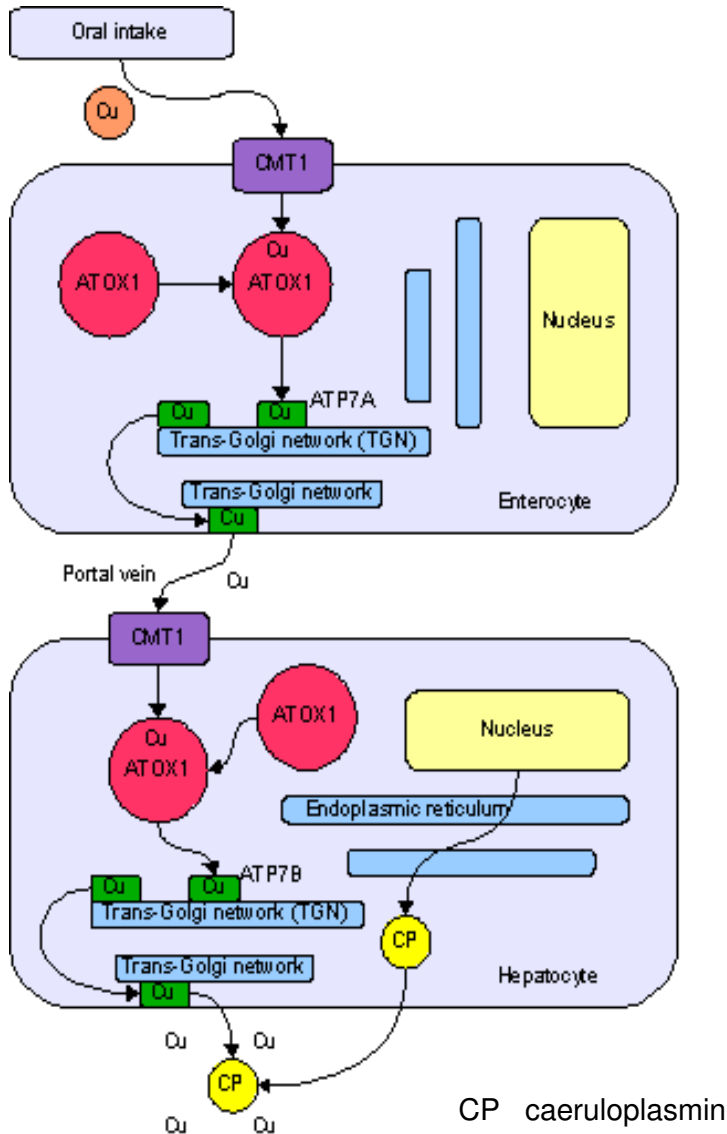
- Description of population (or prior information)
- Classification and diagnosis: $D = \arg \max_H P(H | \mathcal{E})$ i.e. D is the hypothesis with maximum $P(H | \mathcal{E})$
- Prediction
- Decision making based on *what-if* scenario's

Example: Symptomate



- **Huge** Bayesian network covering a large part of medicine (1000s of diseases)
- Diagnostic problem solving
- Designed by hand!
- <https://www.symptomate.com> (also Apps for Android and iOS)

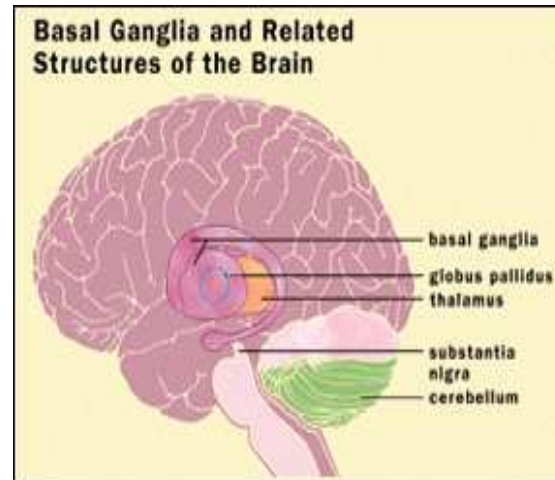
Example: Wilson's disease



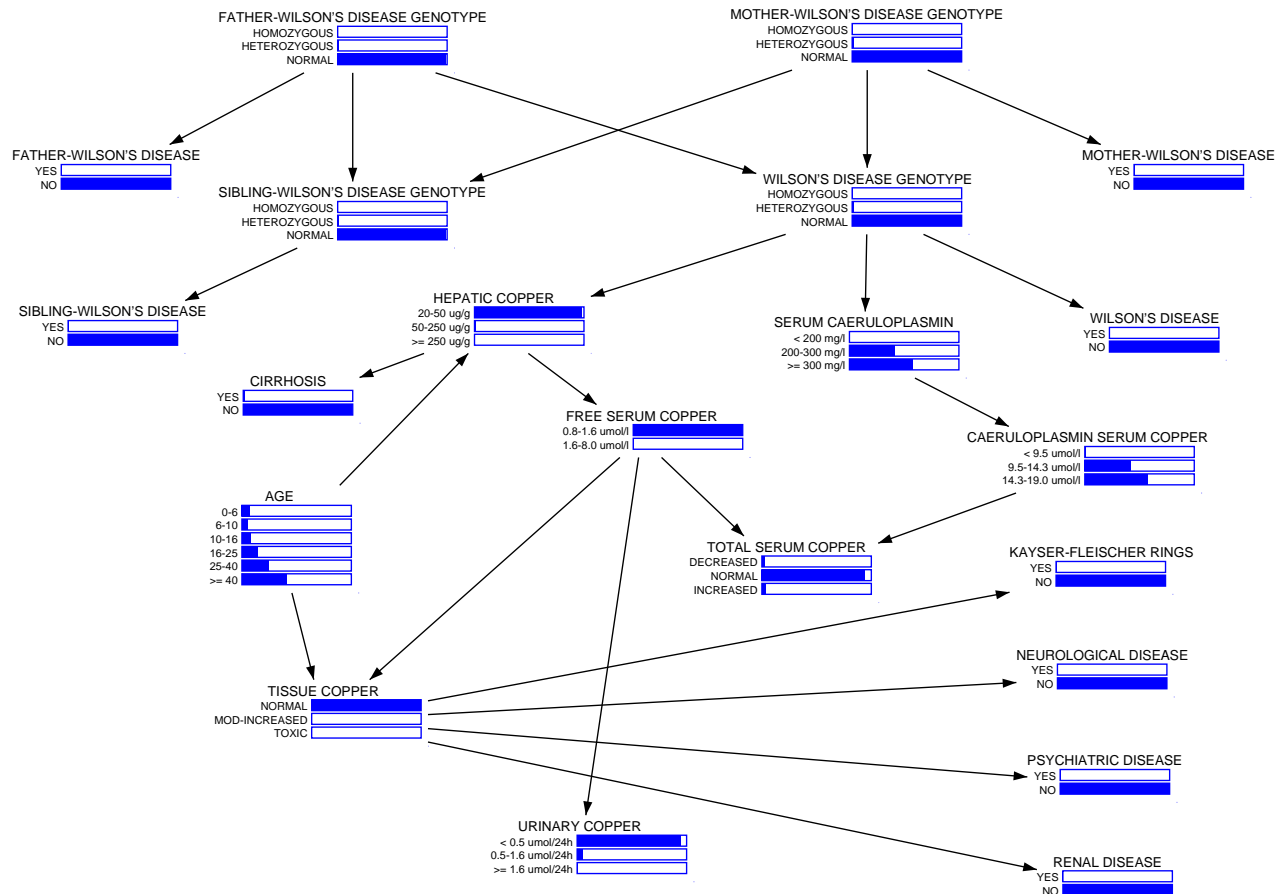
Wilson's disease protein ATP7B



Kayser Fleischer rings

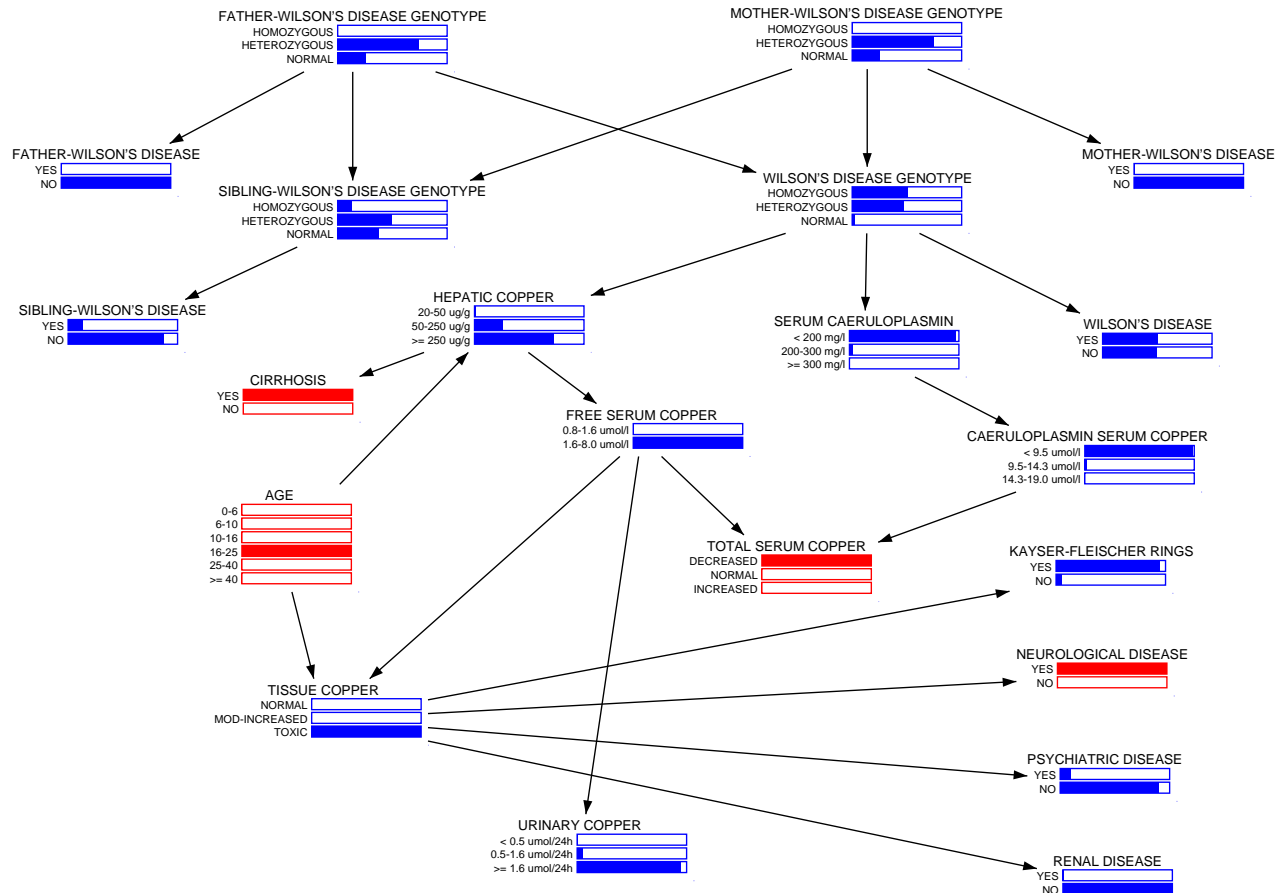


Prior information



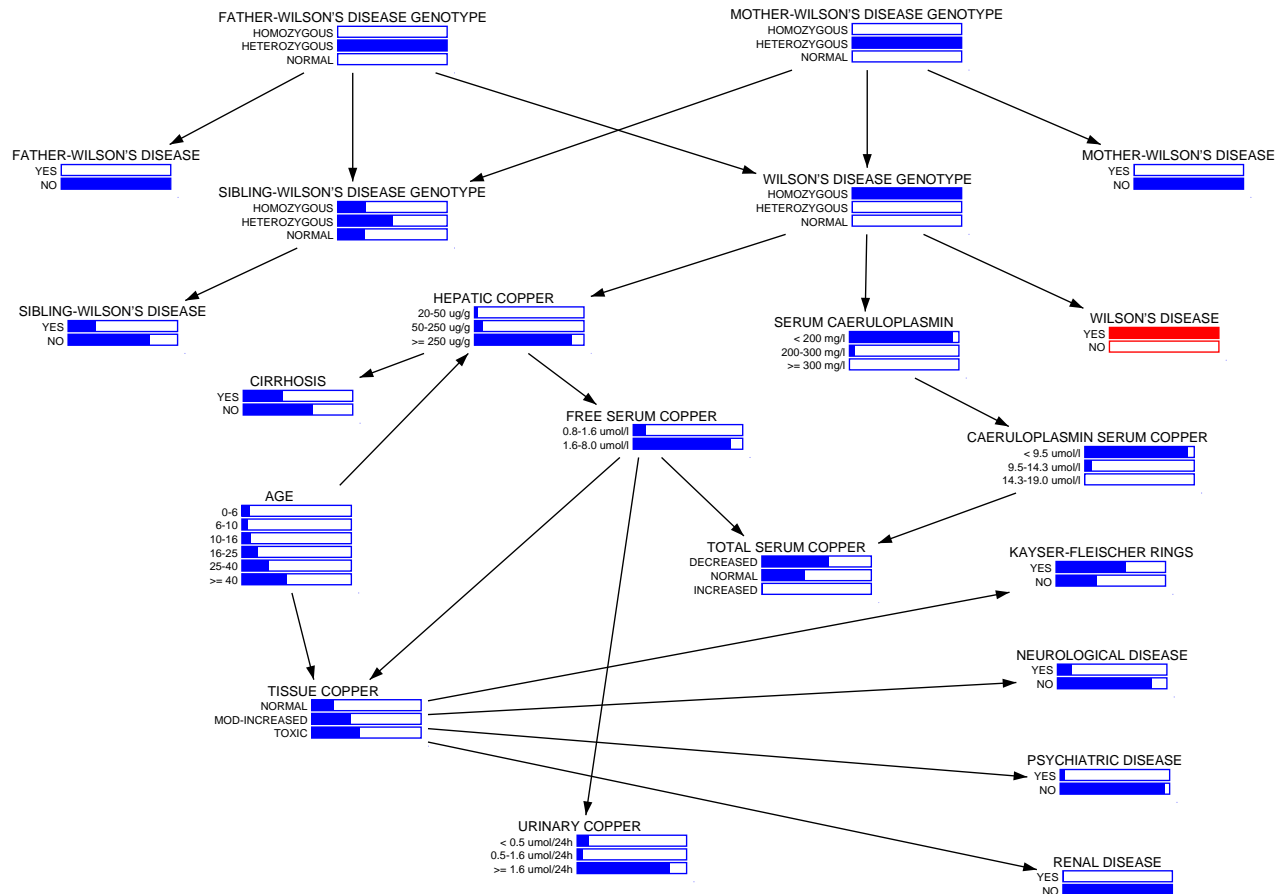
- Gives **description of the population** on which the assessed probabilities are based, i.e., the original probabilities before new evidence is uncovered
- Marginal probabilities $P(V)$ for every vertex V , e.g., $P(\text{WILSON'S DISEASE} = \text{yes})$

Diagnostic problem solving



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

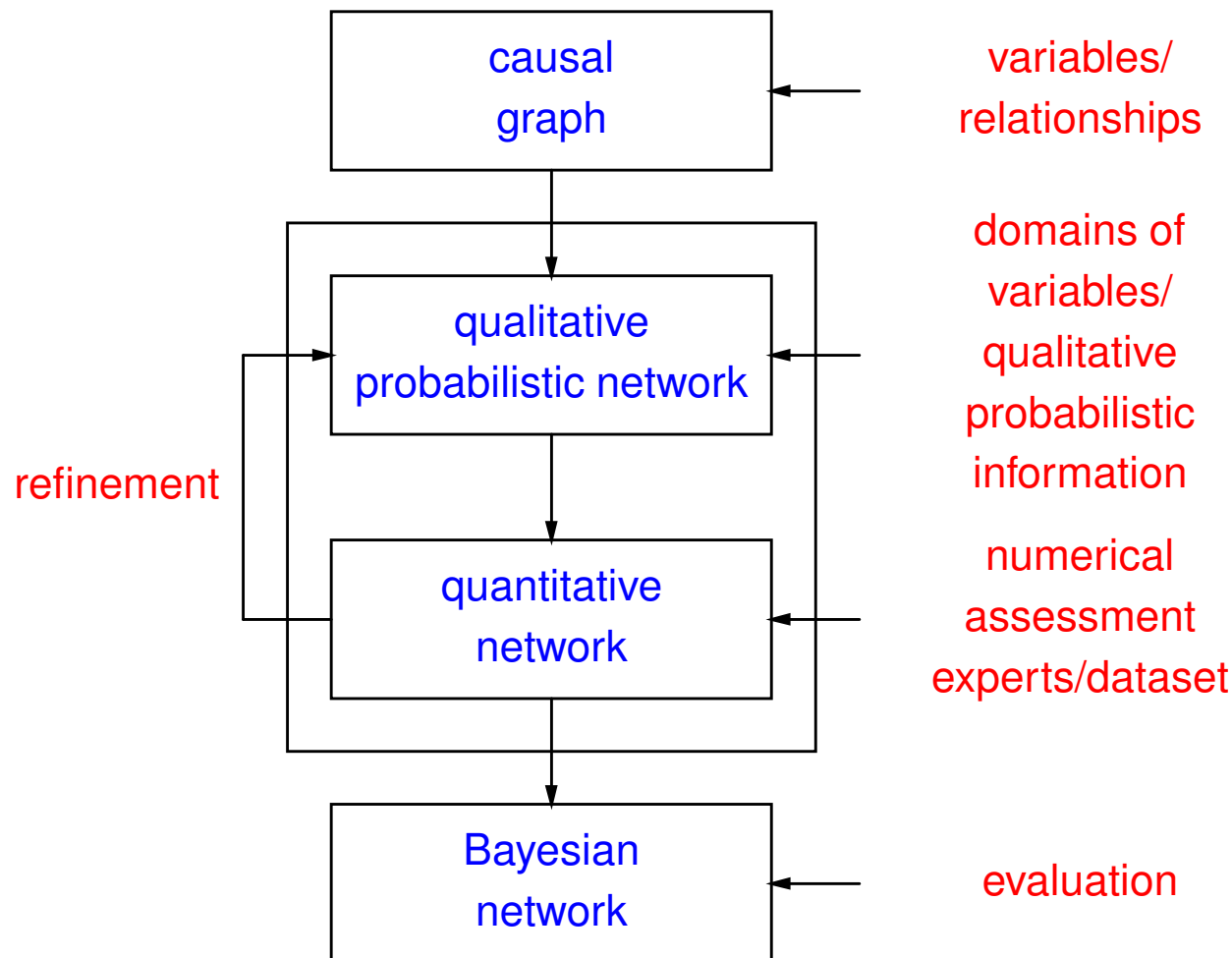
Prediction of associated findings



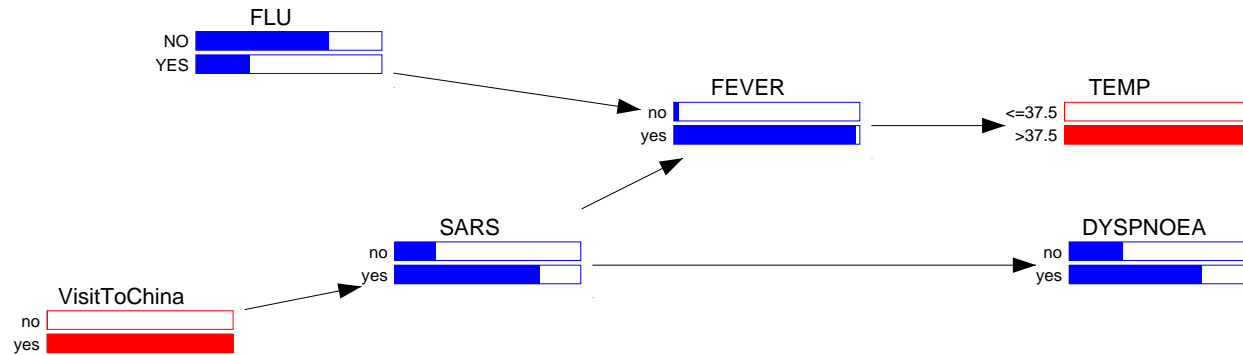
- Gives **description of the findings associated with a given class or category**, such as Wilson's disease
- Marginal probabilities $P^*(V) = P(V | \mathcal{E})$ for every vertex V , e.g., $P(\text{Kayser-Fleischer Rings} = \text{yes} | \mathcal{E})$ with \mathcal{E} evidence

Design of Bayesian network

- Design principle: start modelling qualitatively (as with traditional knowledge-based systems)

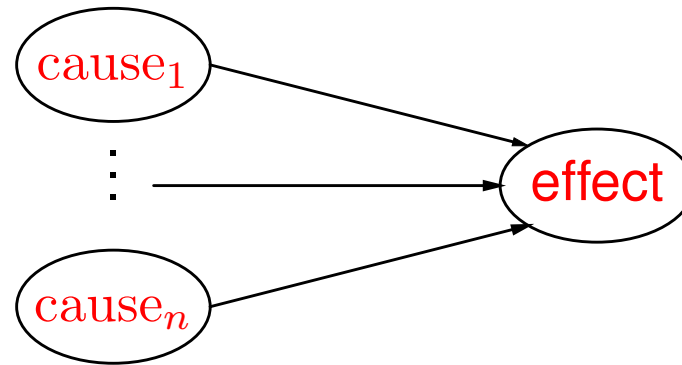


Terminology



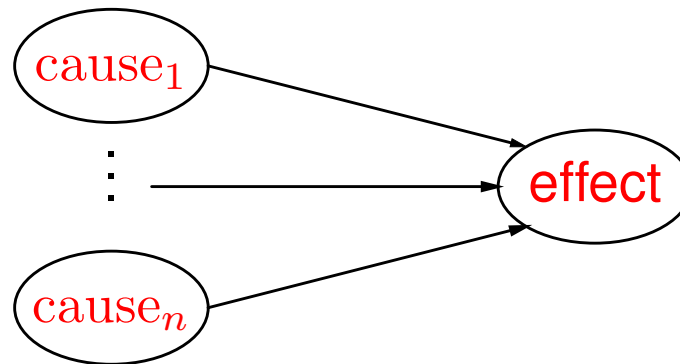
- **Parent SARS of Child FEVER**
- SARS is **Ancestor** of TEMP
- DYSPTNOEA is **Descendant** of VisitToChina
- **Query node**, e.g., FEVER
- **Evidence**, e.g., VisitToChina and TEMP
- **Markov blanket**, e.g.,
for SARS: {VisitToChina, DYSPTNOEA, FEVER, FLU}

Causal graph: Topology (structure)



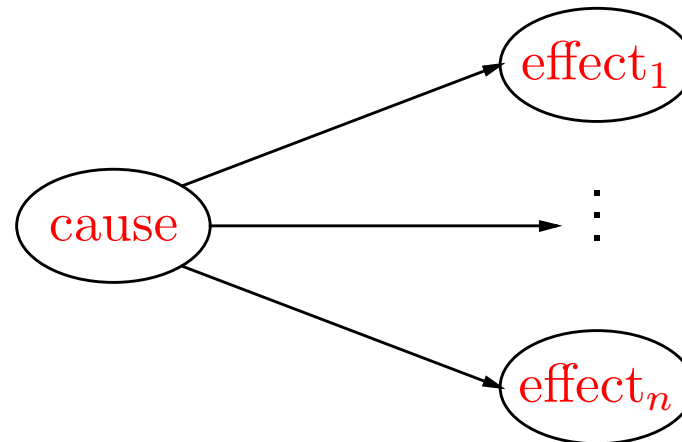
- Identify factors that are relevant
- Determine how those factors are causally related to each other
- The arc $cause \rightarrow effect$ means that 'cause' is a factor involved in causing 'effect'

Causal graph: Common effect



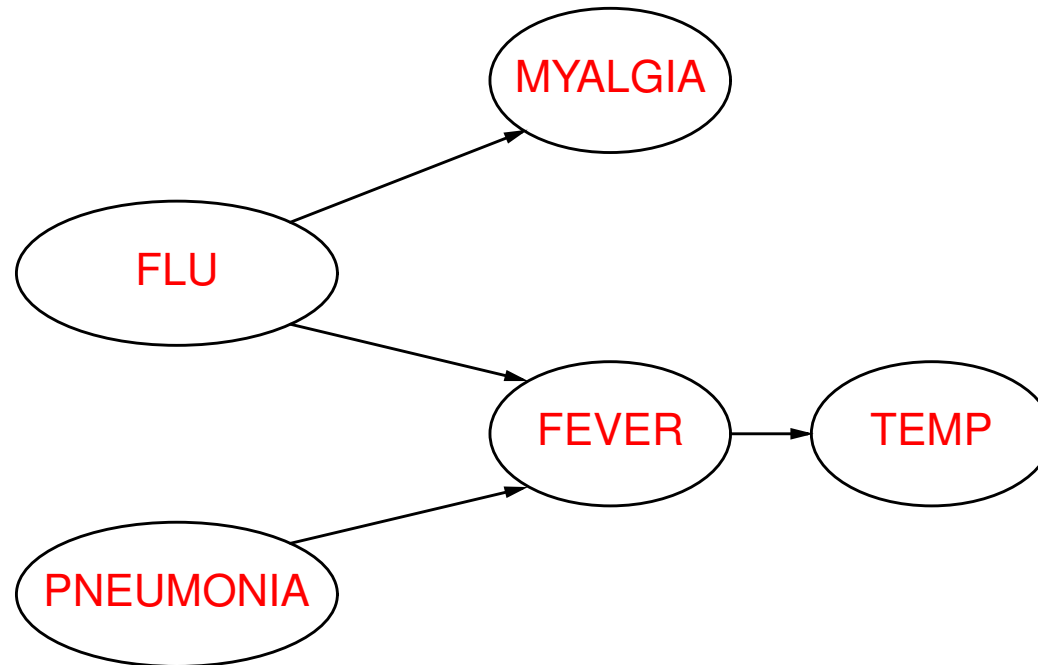
- An effect that has two or more ingoing arcs from other vertices is a **common effect** of those causes
- Kinds of causal interaction
 - **(Positive) synergy:**
POLLUTION → CANCER ← SMOKING
 - **Prevention (negative synergy):**
VACCINE → DEATH ← SMALLPOX
 - **XOR:** ALKALI → DEATH ← ACID

Causal graph: Common cause



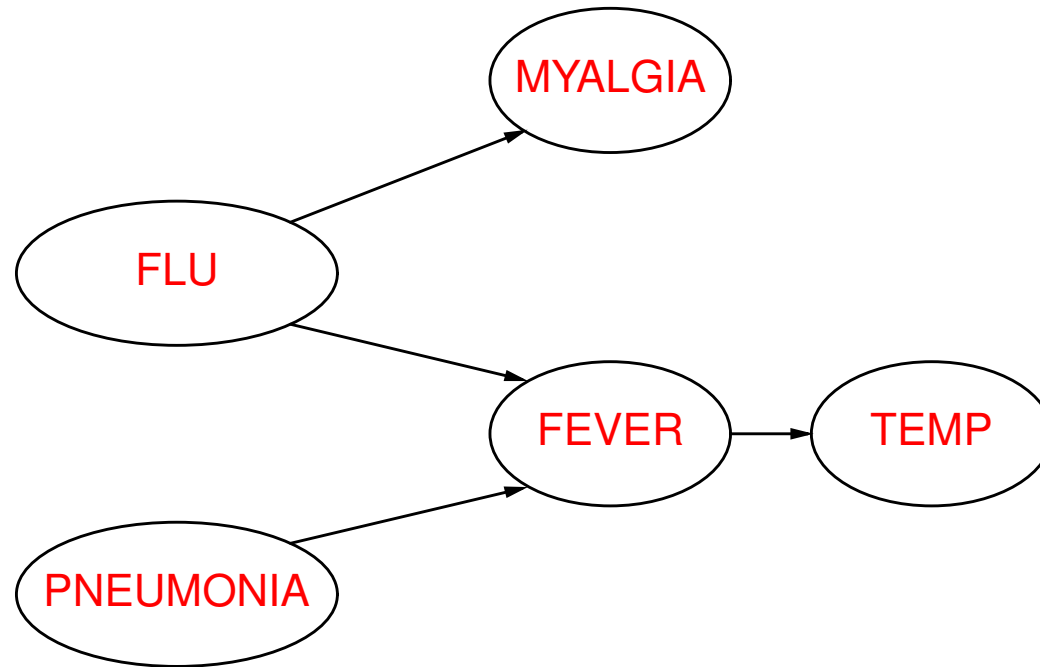
- A cause that has two or more outgoing arcs to other vertices is a **common cause (factor)** of those effects
- The effects of a common cause are usually observables (e.g. manifestations of failure of a device or symptoms in a disease)

Causal graph: Example



- FLU 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

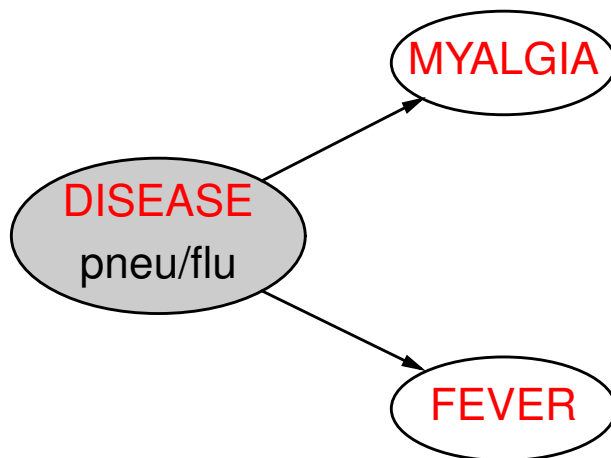
Check independence relationship



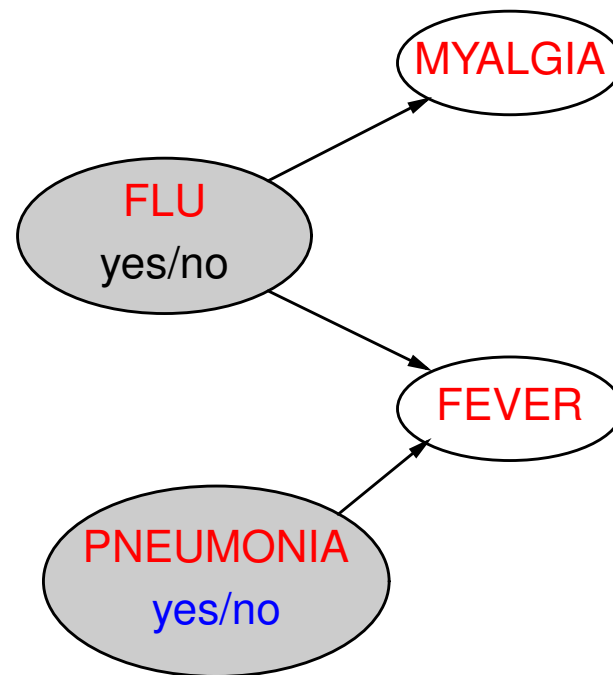
- **Conditional independence:** $X \perp\!\!\!\perp Y \mid Z$
 - $\{\text{FEVER}\} \perp\!\!\!\perp \{\text{MYALGIA}\} \mid \{\text{FLU}\}$
 - $? \mid \{\text{FEVER}\}$
 - $\{\text{PNEUMONIA}\} \perp\!\!\!\perp \{\text{FLU}\} \mid ?$
 - $\{\text{PNEUMONIA}\} \not\perp\!\!\!\perp \{\text{FLU}\} \mid \{\text{FEVER}\}$

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

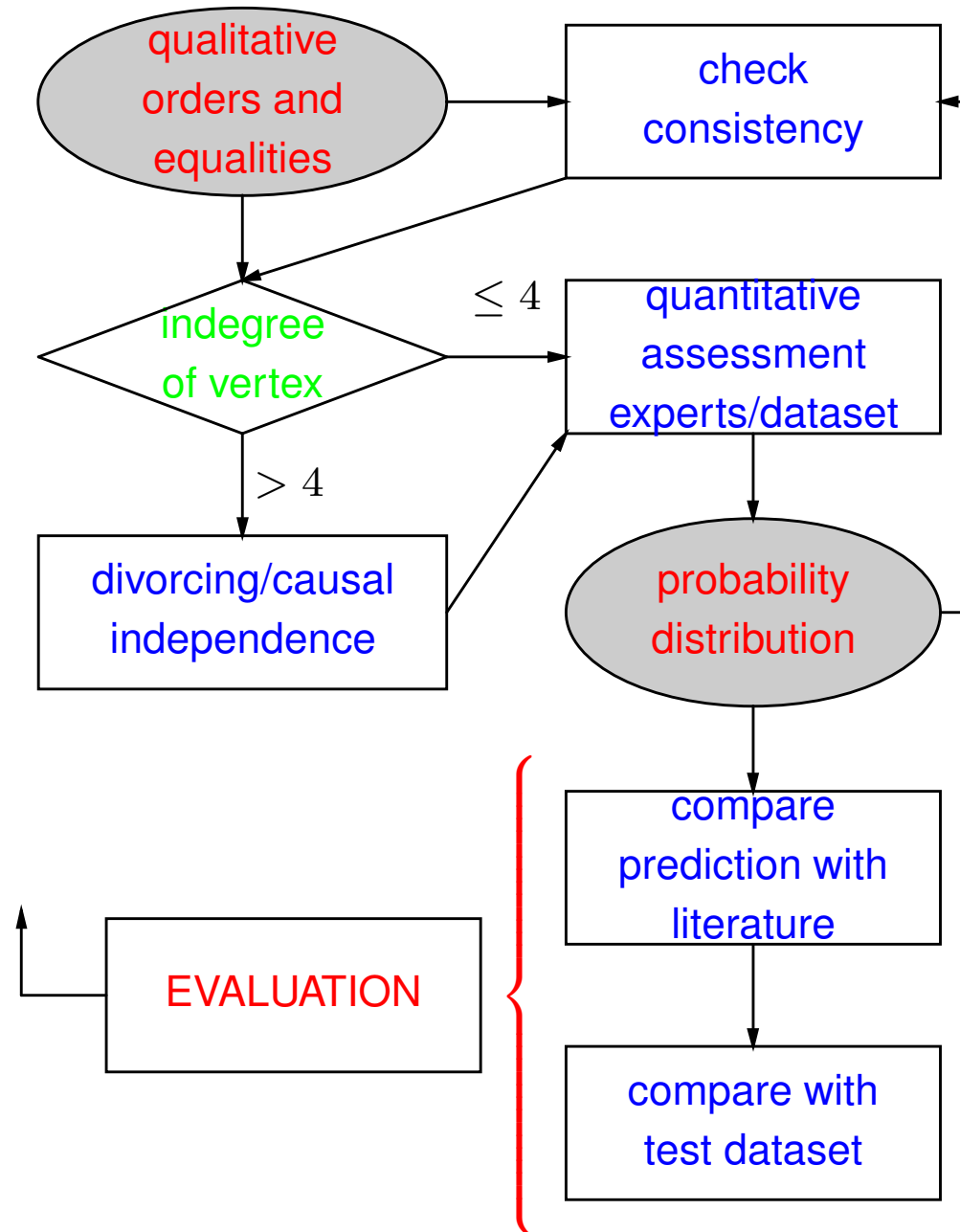


(b) Multiple variables

Choose values

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

Probability assessment



Expert judgements

- **Qualitative probabilities:**
 - *Qualitative orders:*

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

- *Equalities:*

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

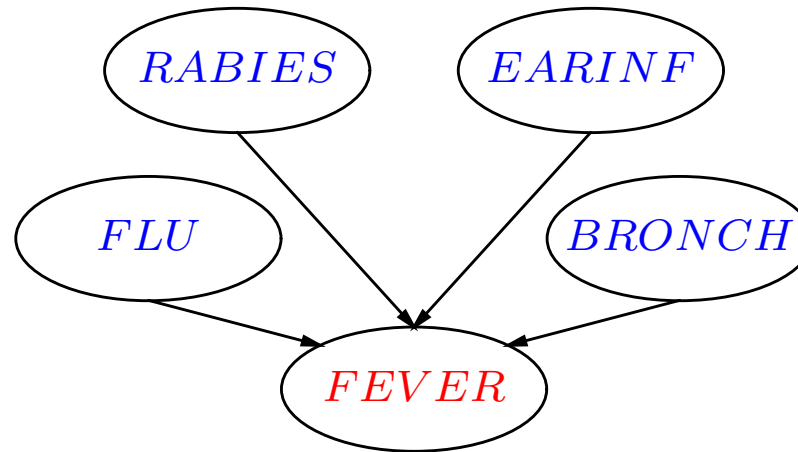
Expert judgements (cont.)

- 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
≥ 90	0.1	0.3	0.6

- Order information can be used as **inequality constraints** in learning (see Y. Zhou, N. Fenton, M. Neil: Bayesian network approach to multinomial parameter learning using data and expert judgments, Int. Journal of Approximate Reasoning, 55 (2014) 1252–1268)

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:

$$P(\text{fever} \mid \text{flu}, \text{rabies}, \text{ear_infection}, \text{bronchitis}) = ?$$

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

- *Tree-Augmented Bayesian Network* (TAN)

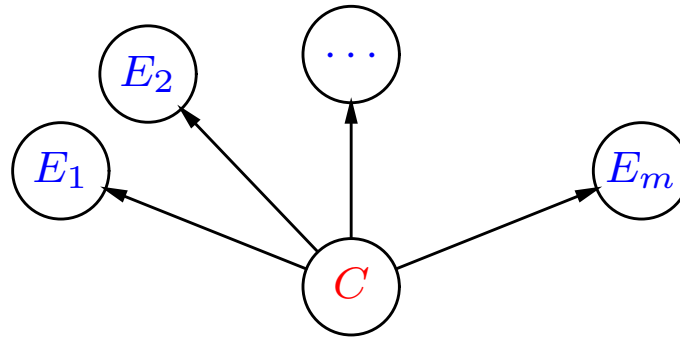
or,

- 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 $\mathcal{E} \subseteq \{E_1, \dots, E_m\}$. We have $E_i \perp\!\!\!\perp E_j \mid C$, for $i \neq j$. Hence, using Bayes' rule:

$$P(C \mid \mathcal{E}) = \frac{P(\mathcal{E} \mid C)P(C)}{P(\mathcal{E})} \quad \text{with:}$$
$$P(\mathcal{E} \mid C) = \prod_{E \in \mathcal{E}} P(E \mid C) \quad \text{by cond. ind.}$$
$$P(\mathcal{E}) = \sum_C P(\mathcal{E} \mid C)P(C) \quad \text{marg. \& cond.}$$

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
<i>Sunny</i>	2/9	3/5
<i>Overcast</i>	4/9	0/5
<i>Rain</i>	3/9	2/5

Temperature	Play=Yes	Play=No
<i>Hot</i>	2/9	2/5
<i>Mild</i>	4/9	2/5
<i>Cool</i>	3/9	1/5

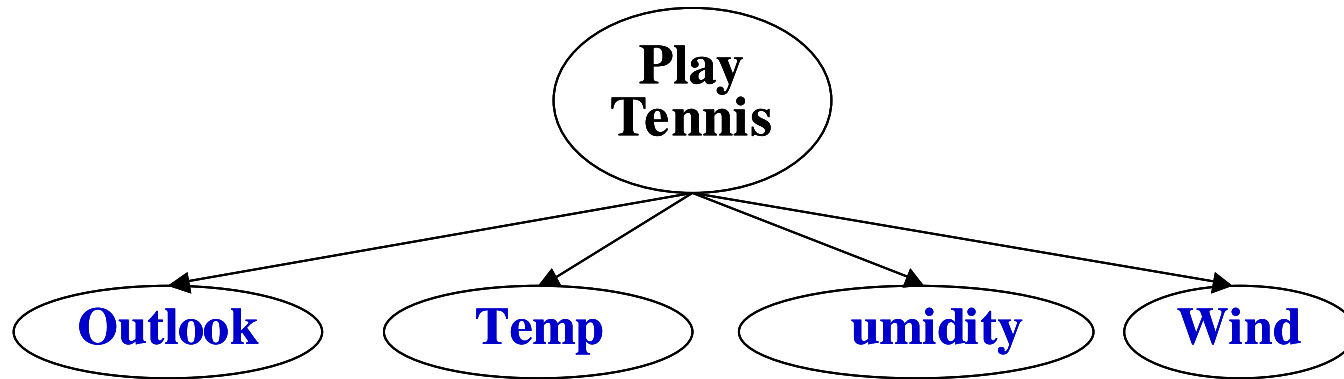
Humidity	Play=Yes	Play=No
<i>High</i>	3/9	4/5
<i>Normal</i>	6/9	1/5

Wind	Play=Yes	Play=No
<i>Strong</i>	3/9	3/5
<i>Weak</i>	6/9	2/5

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

Example of Naive Bayes (2)

- Testing phase (inference)



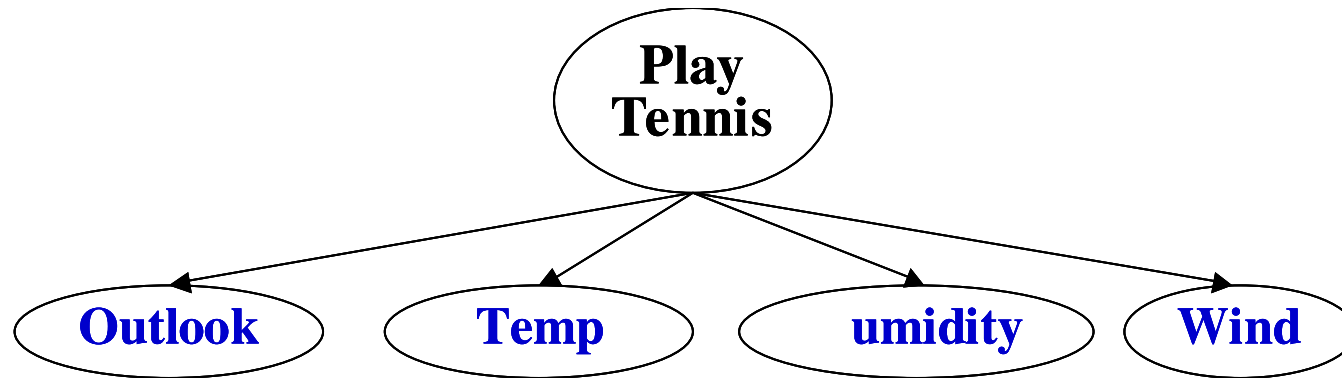
Evidence:

$x = (\text{Outlook}=\textit{Sunny}, \text{Temperature}=\textit{Cool}, \text{Humidity}=\textit{High}, \text{Wind}=\textit{Strong})$

Then given x , $\text{PlayTennis}=?$

Example of Naive Bayes (3)

- Testing phase (inference)



$$\begin{aligned} P(\text{Yes}|\mathbf{x}) &= P(\mathbf{x}|\text{PlayTennis}=\text{Yes})P(\text{PlayTennis}=\text{Yes})/P(\mathbf{x}) \propto \\ &\propto [P(\text{Sunny}|\text{Yes})P(\text{Cool}|\text{Yes})P(\text{High}|\text{Yes})P(\text{Strong}|\text{Yes})]P(\text{Play}=\text{Yes}) = \\ &= 2/9 * 3/9 * 3/9 * 3/9 * 9/14 = \mathbf{0.0053} \end{aligned}$$

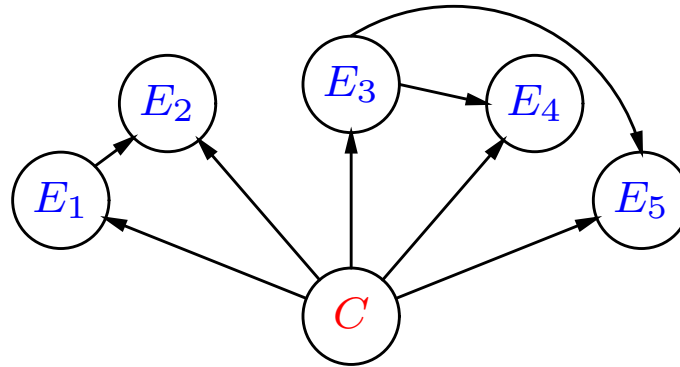
$$P(\text{No}|\mathbf{x}) \propto [P(\text{Sunny}|\text{No}) P(\text{Cool}|\text{No})P(\text{High}|\text{No})P(\text{Strong}|\text{No})]P(\text{Play}=\text{No}) = \mathbf{0.0206}$$

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

Note: to get probabilities we need to normalise

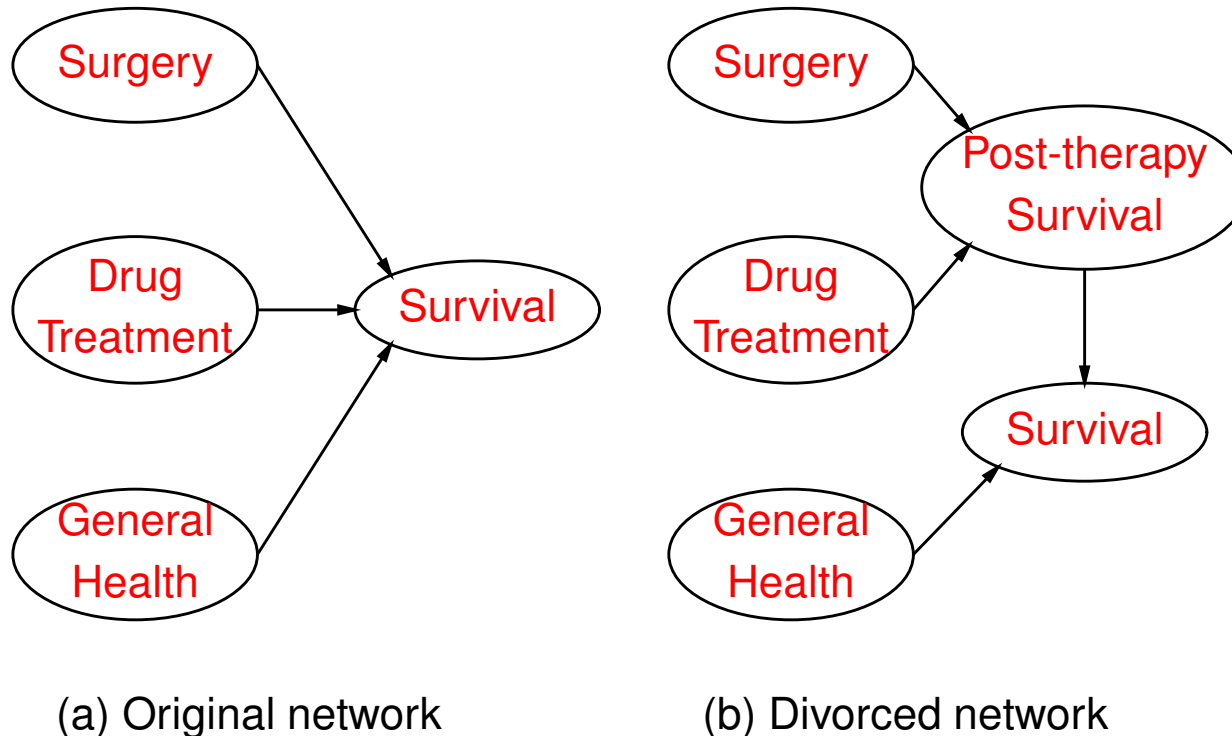
$$P(\text{Yes}|\mathbf{x}) = P(\text{Yes}|\mathbf{x}) / (P(\text{Yes}|\mathbf{x}) + P(\text{No}|\mathbf{x})) = 0.0053/(0.0053+0.0206) = 0.20$$

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)

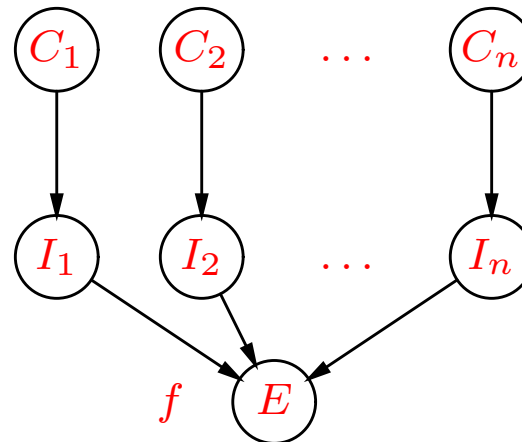
Divorcing multiple parents



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

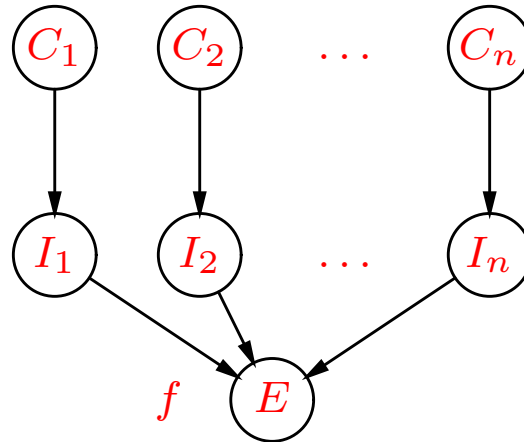


with:

- **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



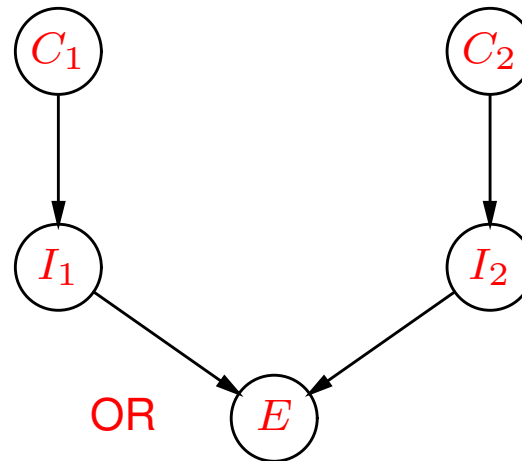
$$\begin{aligned} P(e | C_1, \dots, C_n) &= \sum_{I_1, \dots, I_n} P(e | I_1, \dots, I_n) P(I_1, \dots, I_n | C_1, \dots, C_n) \\ &= \sum_{f(I_1, \dots, I_n) = e} P(e | I_1, \dots, I_n) \underline{P(I_1, \dots, I_n | C_1, \dots, C_n)} \end{aligned}$$

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

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

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

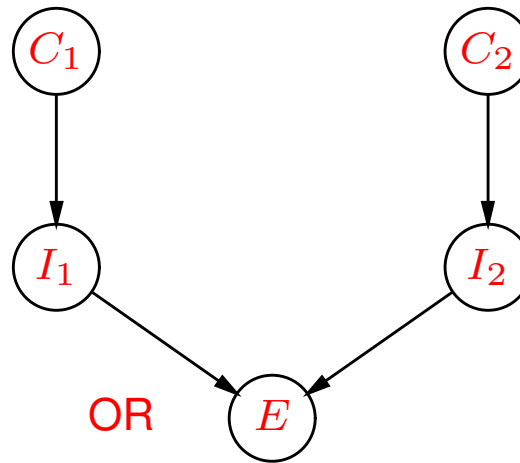
Causal independence: Noisy OR



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

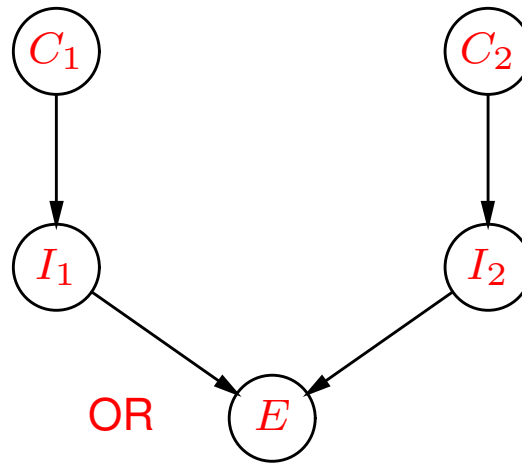
$$P(e | C_1, C_2) = ?$$

Causal independence: Noisy OR (cont.)



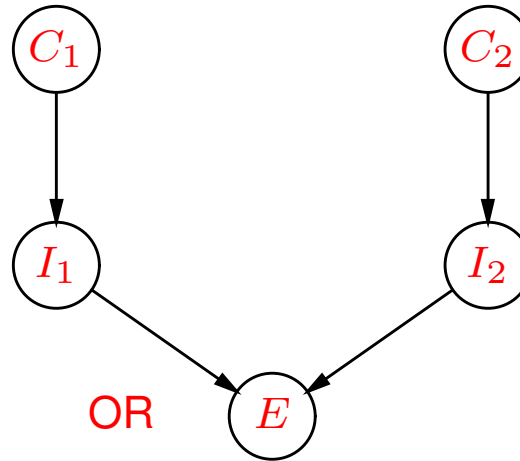
$$\begin{aligned} P(e \mid C_1, C_2) &= \sum_{I_1, I_2} P(e \mid I_1, I_2, C_1, C_2) P(I_1, I_2 \mid C_1, C_2) \\ &= ? \end{aligned}$$

Causal independence: Noisy OR (cont.)



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

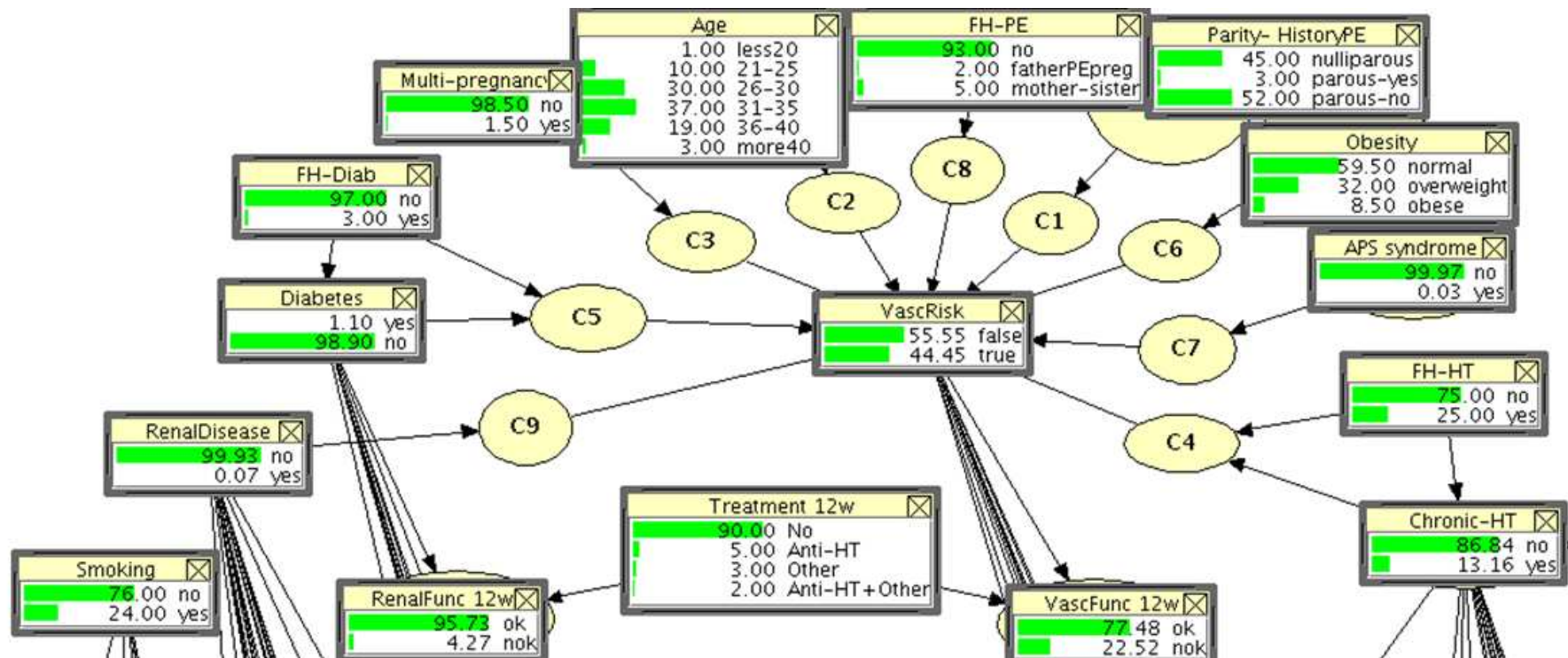
Causal independence: Noisy OR (cont.)



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

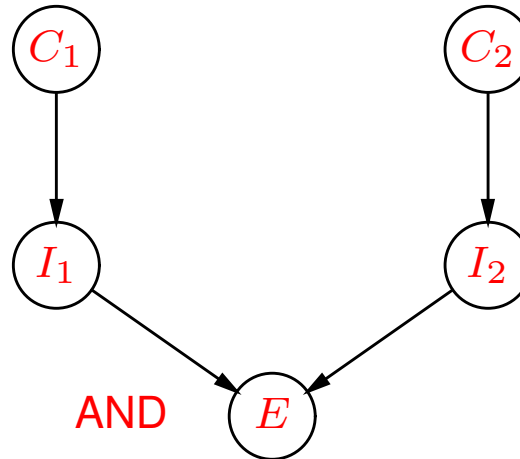
$$\begin{aligned} P(e | C_1, C_2) &= \sum_{f(I_1, I_2)=e} P(e | I_1, I_2) \prod_{k=1,2} P(I_k | C_k) \\ &= P(i_1 | C_1)P(i_2 | C_2) + P(\neg i_1 | C_1)P(i_2 | C_2) + P(i_1 | C_1)P(\neg i_2 | C_2) \end{aligned}$$

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 $P(E | I_1, I_2)$, is a logical AND
- Meaning: presence of all causes C_i with absolute certainty will cause the effect e (i.e. $E = true$); otherwise, $\neg e$

$$P(e | C_1, C_2) = ?$$

Are Bayesian networks always suitable?

OF COURSE NOT !

- **Uncertainty** must be one of the essential features of the problem
- Sufficient **knowledge about the problem** (domain experts, literature, data)
- Decomposable representation because of **independence** information

However,

- a “black box” model such as neural networks is sometimes sufficient and will often give the best results of image analysis problems

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