# Probabilistic Reasoning

## *Independence Representation*

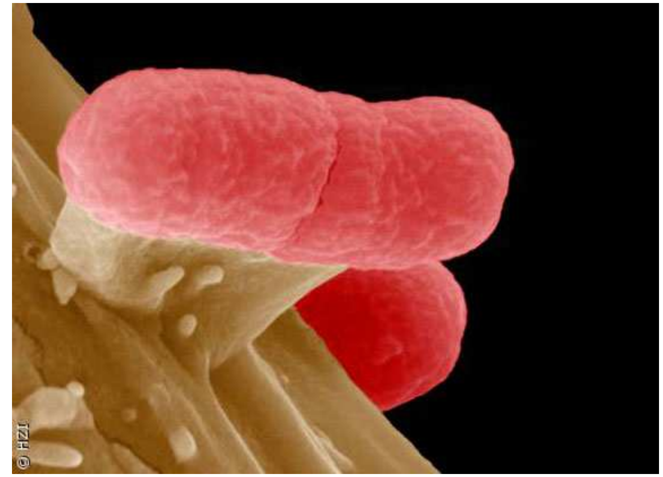
Peter Lucas

peterl@cs.ru.nl

Institute for Computing and Information Sciences

Radboud University Nijmegen

# Have you got Entero Hemorrhagic E. coli?

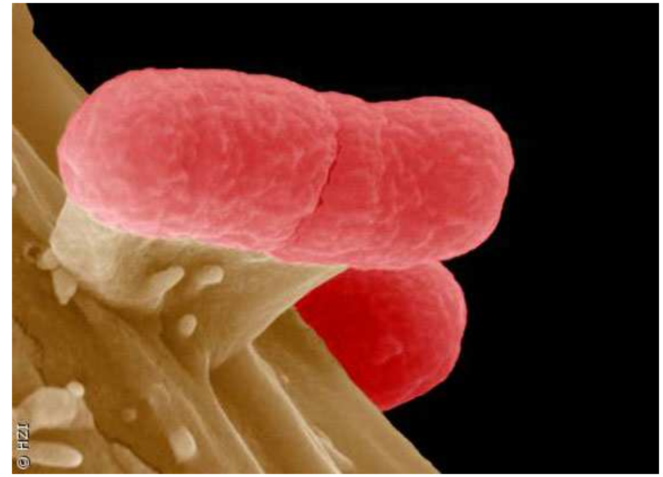



$$
\begin{aligned}
P(e, g, d) &= 0.009215 \\
P(e, \bar{g}, d) &= 0.000485 \\
P(e, g, \bar{d}) &= 0.000285 \\
P(e, \bar{g}, \bar{d}) &= 1.5 \cdot 10^{-5} \\
P(\bar{e}, g, d) &= 9.9 \cdot 10^{-6} \\
P(\bar{e}, \bar{g}, d) &= 0.0098901 \\
P(\bar{e}, g, \bar{d}) &= 0.0009801
\end{aligned}
$$

$$P(\bar{e}, \bar{g}, \bar{d}) = 0.97912$$

- $E$: EHEC; $G$: visited Northern Germany; $D$: diarrhea
- Enterohemorrhagic E. coli and visited Northern Germany?
- Probability of EHEC given visit to Northern Germany?

# Have you got Entero Hemorrhagic E. coli?


$$
\begin{aligned}
P(e, g, d) &= 0.009215 \\
P(e, \bar{g}, d) &= 0.000485 \\
P(e, g, \bar{d}) &= 0.000285 \\
P(e, \bar{g}, \bar{d}) &= 1.5 \cdot 10^{-5} \\
P(\bar{e}, g, d) &= 9.9 \cdot 10^{-6} \\
P(\bar{e}, \bar{g}, d) &= 0.0098901 \\
P(\bar{e}, g, \bar{d}) &= 0.0009801
\end{aligned}
$$

$$P(\bar{e}, \bar{g}, \bar{d}) = 0.97912$$

- $E$: EHEC; $G$: visited Northern Germany; $D$: diarrhea
- Enterohemorrhagic E. coli and visited Northern Germany? 0.0095
- Probability of EHEC given visit to Northern Germany? 0.906

# Probabilistic reasoning

Joint probability distribution $P(X) = P(X_1, X_2, \ldots, X_n)$

- marginalisation:

$$P(Y) = \sum_{Z} P(Y, Z), \text{ with } X = Y \cup Z$$

- conditional probabilities:

$$P(Y \mid Z) = \frac{P(Y, Z)}{P(Z)}$$

- Bayes' theorem:

$$P(Y \mid Z) = \frac{P(Z \mid Y)P(Y)}{P(Z)}$$

# Probabilistic reasoning (cont)

Examples:

$P(e, g) = P(e, g, d) + P(e, g, \bar{d}) = 0.009215 + 0.000285 = 0.0095$

$P(e \mid g) = P(e, g)/P(g) = 0.0095/0.01049 \approx 0.906$

Note that:

- Mainly interested in conditional probability distributions:

$$P(Z \mid \mathcal{E}) = P^{\mathcal{E}}(Z)$$

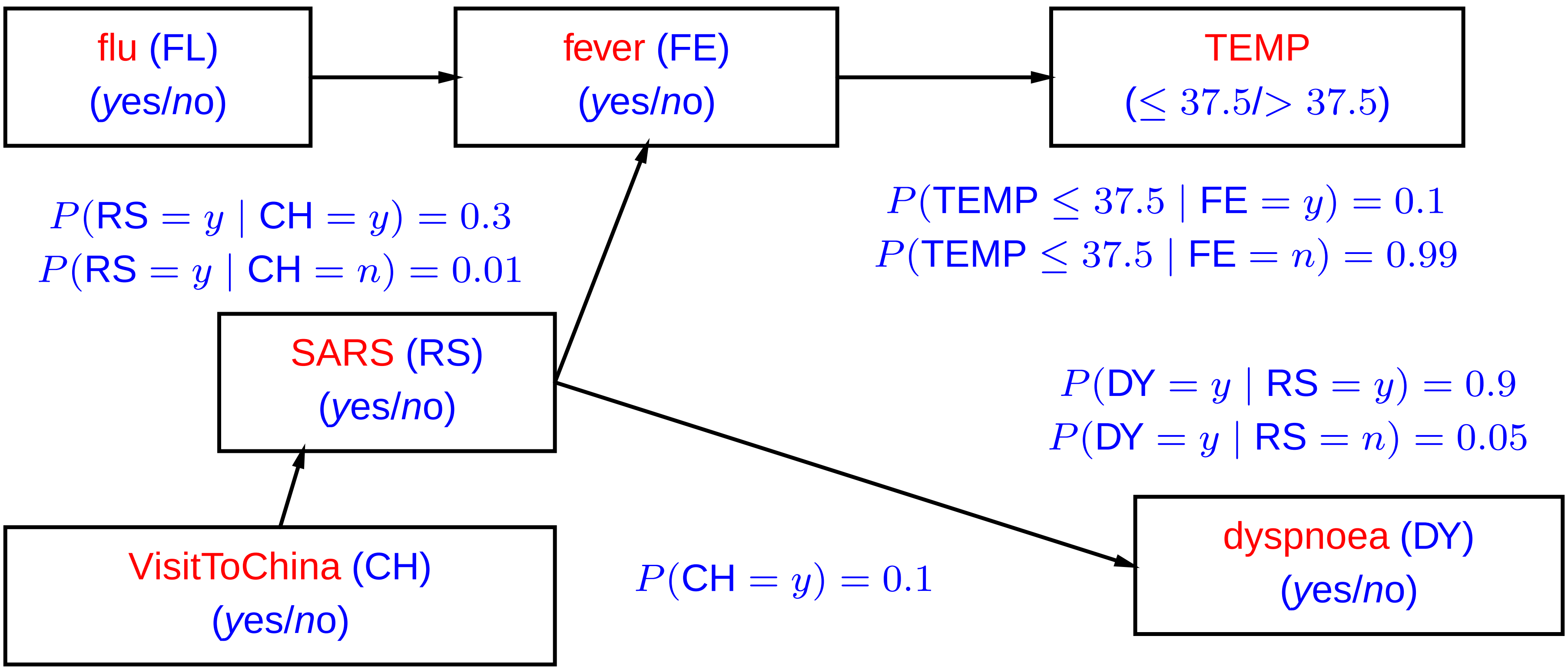
  for (possibly empty) evidence $\mathcal{E}$ (instantiated variables)
- Tendency to focus on conditional probability distributions of single variables
- Many efficient reasoning algorithms exist

# Bayesian networks

$P(\text{CH}, \text{FL}, \text{RS}, \text{DY}, \text{FE}, \text{TEMP})$

$P(\text{FE} = y \mid \text{FL} = y, \text{RS} = y) = 0.95$
$P(\text{FE} = y \mid \text{FL} = n, \text{RS} = y) = 0.80$
$P(\text{FE} = y \mid \text{FL} = y, \text{RS} = n) = 0.88$
$P(\text{FE} = y \mid \text{FL} = n, \text{RS} = n) = 0.001$

$P(\text{FL} = y) = 0.1$




$P(\text{RS} = y \mid \text{CH} = y) = 0.3$
$P(\text{RS} = y \mid \text{CH} = n) = 0.01$

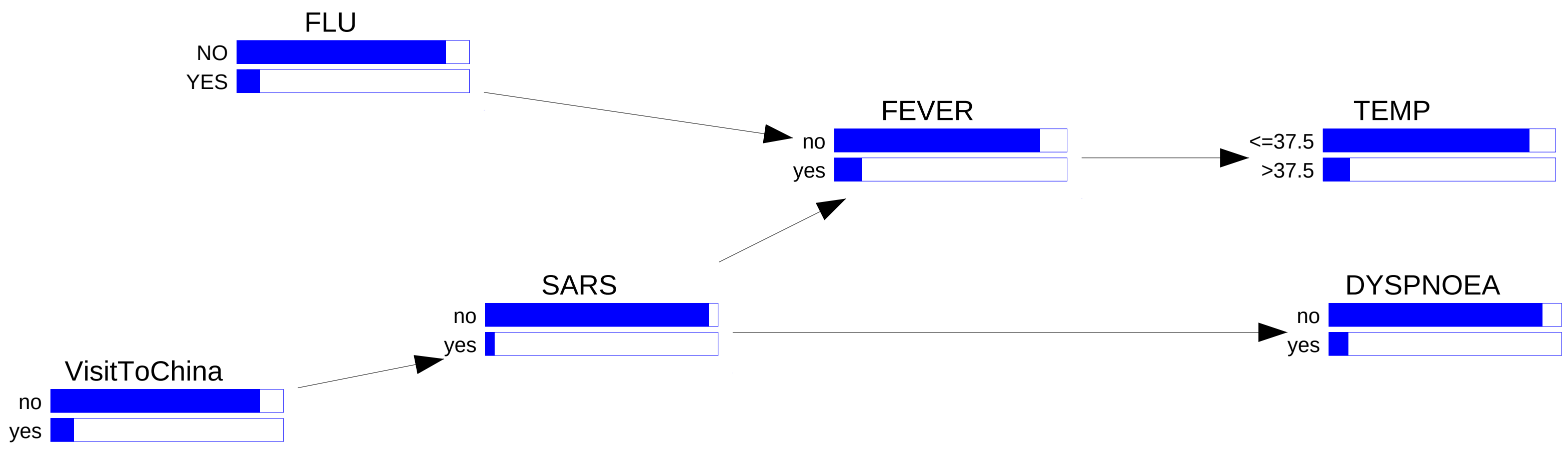
$P(\text{TEMP} \leq 37.5 \mid \text{FE} = y) = 0.1$
$P(\text{TEMP} \leq 37.5 \mid \text{FE} = n) = 0.99$

$P(\text{DY} = y \mid \text{RS} = y) = 0.9$
$P(\text{DY} = y \mid \text{RS} = n) = 0.05$

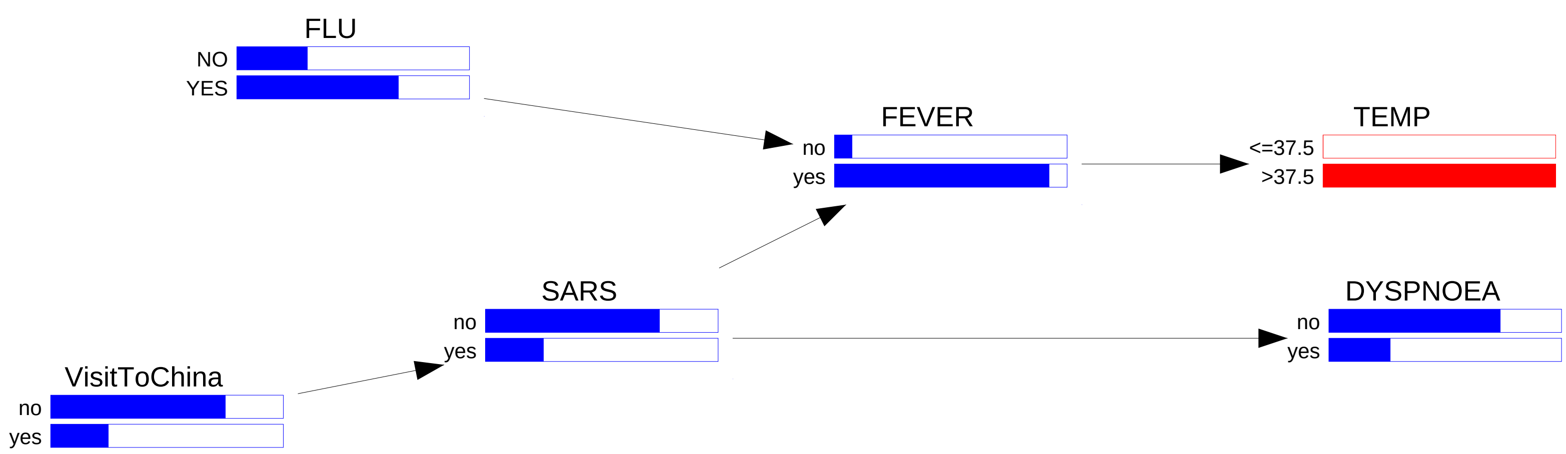
$P(\text{CH} = y) = 0.1$

# Reasoning: evidence propagation

- Nothing known:

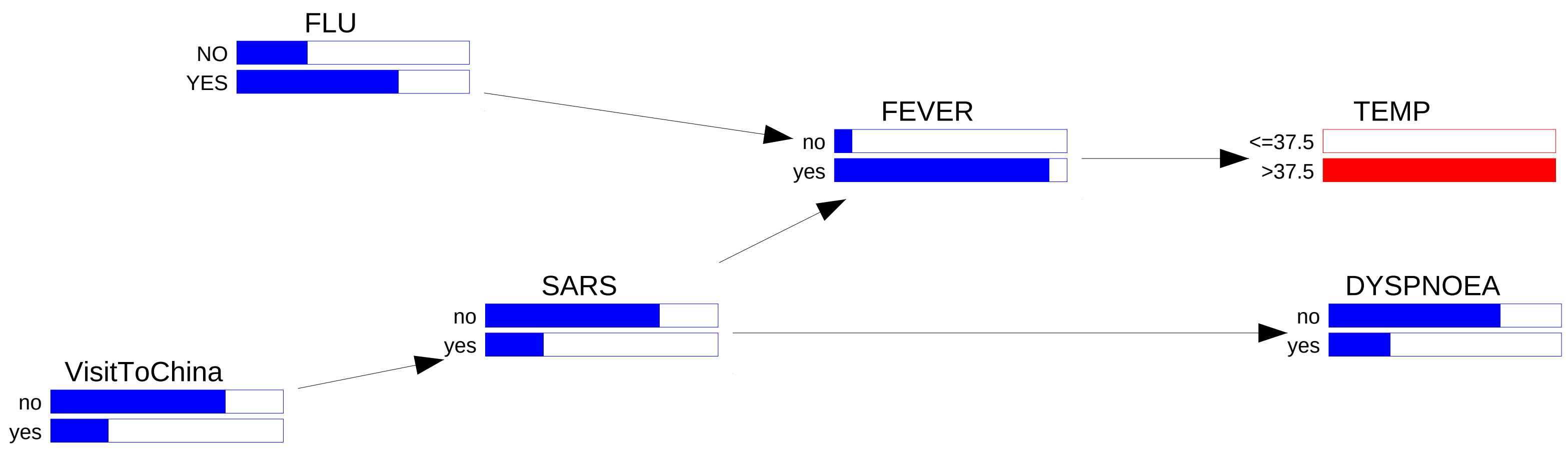



- Temperature $>37.5$ °C:

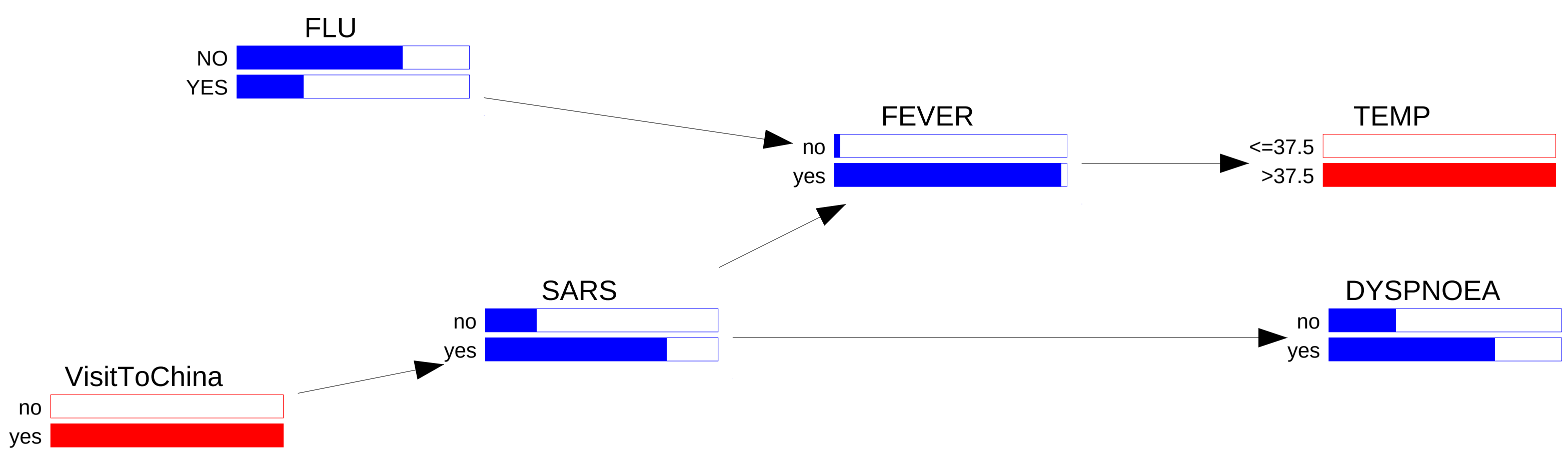


# Reasoning: evidence propagation

- Temperature $>37.5$ °C:




- I just returned from China:



# Definition Bayesian network

A Bayesian network $\mathcal{B}$ is a pair $\mathcal{B} = (G, P)$, where:

- (Qualitative part) $G = (V(G), A(G))$ is an acyclic directed graph, with
  - $V(G) = \{v_1, v_2, \ldots, v_n\}$, a set of vertices (nodes)
  - $A(G) \subseteq V(G) \times V(G)$ a set of arcs

- (Quantitative part) $P(X_{V(G)})$ is a joint probability distribution, such that

$$P(X_{V(G)}) = \prod_{v \in V(G)} P(X_v \mid X_{\pi(v)})$$

where $\pi(v)$ denotes the set of parents of vertex $v$ in $G$

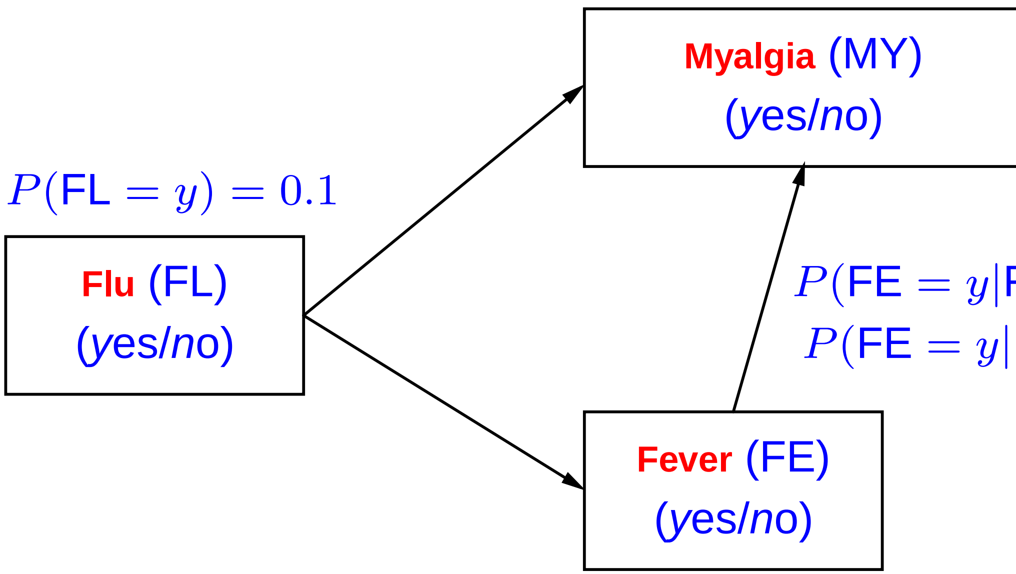
# Markov independence

# A Bayesian network

$\boxed{P(\text{FL}, \text{MY}, \text{FE})}$

$$P(\text{MY} = y|\text{FL} = y, \text{FE} = y) = 0.96$$
$$P(\text{MY} = y|\text{FL} = y, \text{FE} = n) = 0.96$$
$$P(\text{MY} = y|\text{FL} = n, \text{FE} = y) = 0.20$$
$$P(\text{MY} = y|\text{FL} = n, \text{FE} = n) = 0.20$$

$P(\text{FL} = y) = 0.1$

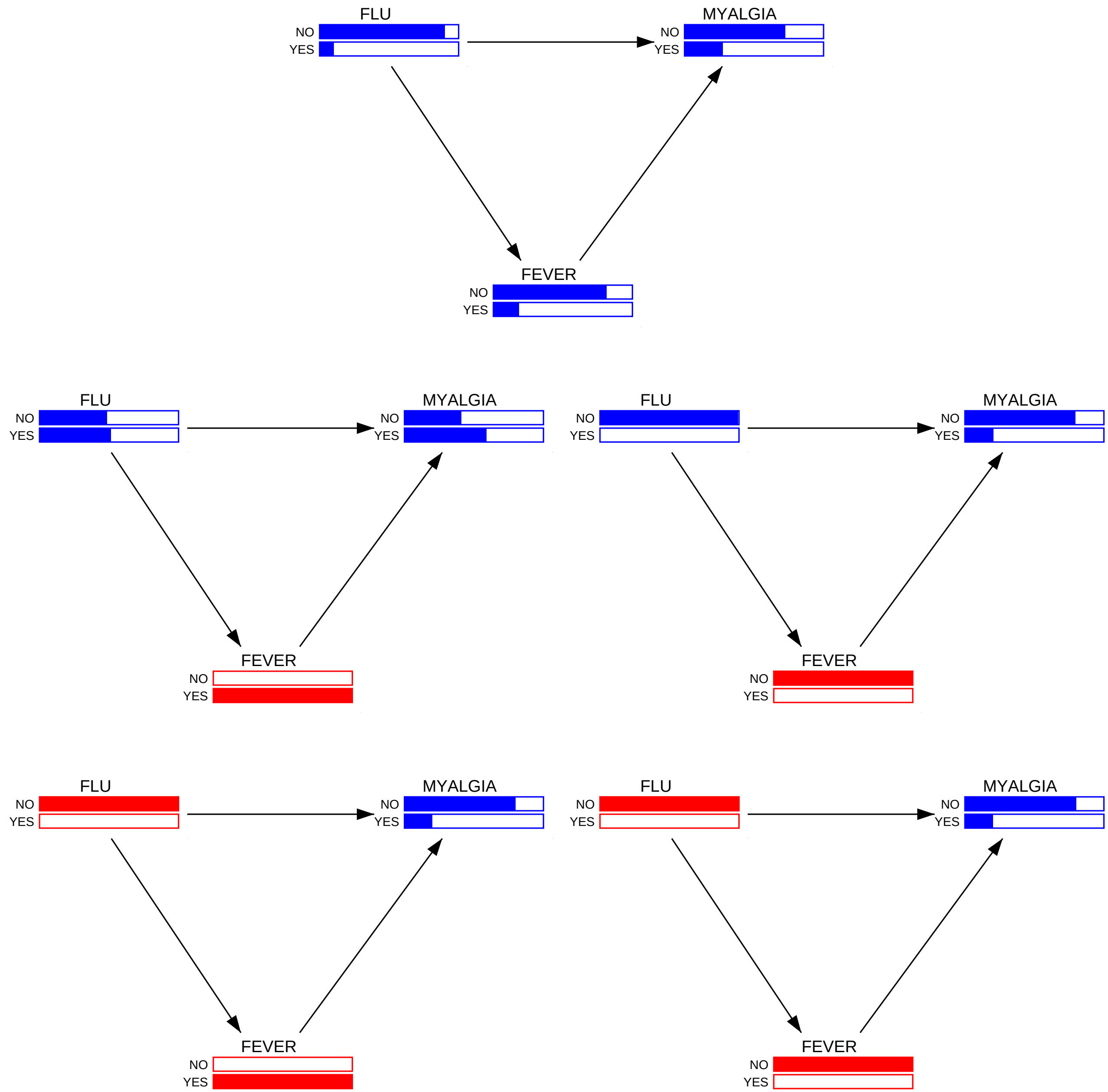



$$P(\text{FE} = y|\text{FL} = y) = 0.95$$
$$P(\text{FE} = y|\text{FL} = n) = 0.1$$

Thus: $P(\text{FL}, \text{MY}, \text{FE}) = P(\text{MY}|\text{FL}, \text{FE})P(\text{FE}|\text{FL})P(\text{FL})$

Example: $P(\neg \textit{fl}, \textit{my}, \textit{fe}) = 0.20 \cdot 0.1 \cdot 0.9 = 0.018$

# Independence and reasoning

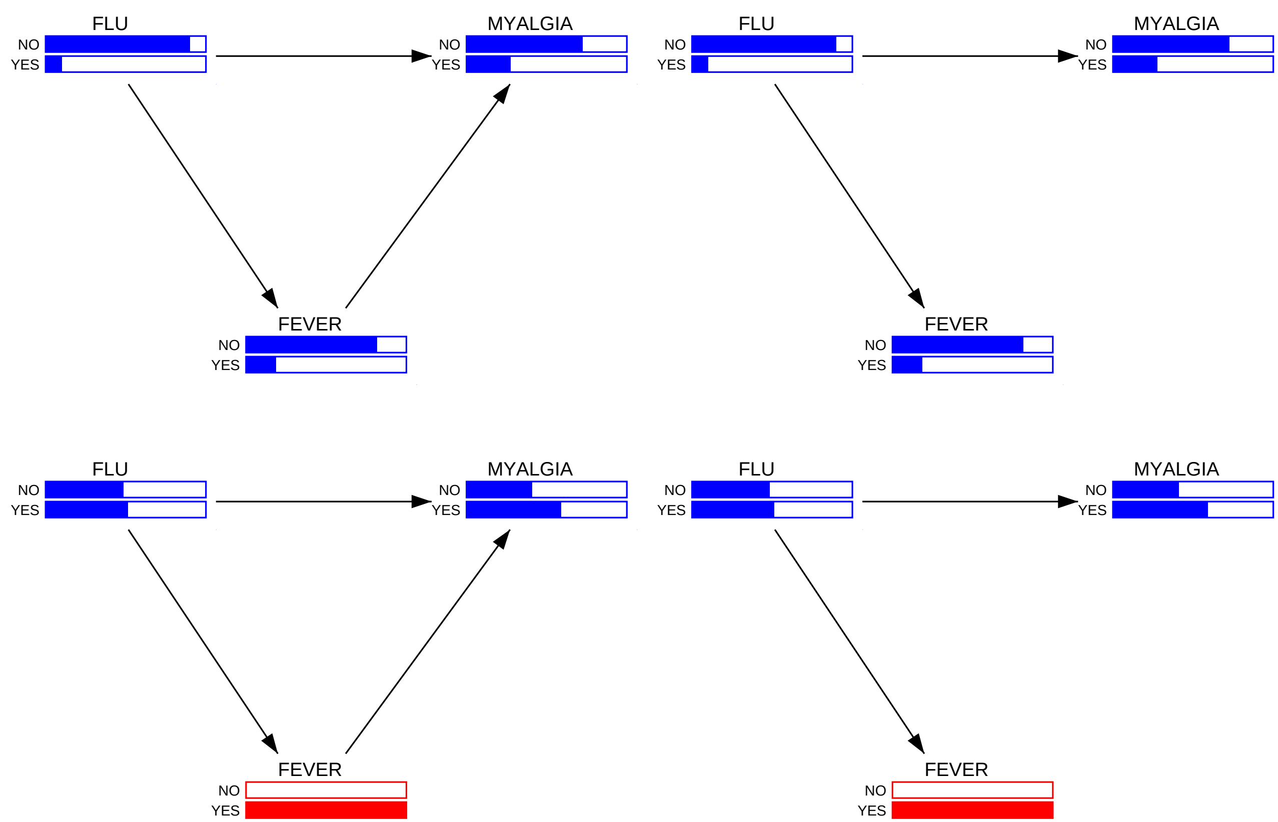


# Independence and reasoning

Conclusion: the arc from FEVER to MYALGIA can be removed, and hence only

$$P(\mathsf{MY} \mid \mathsf{FL}) \; (= P(\mathsf{MY} \mid \mathsf{FL}, \mathsf{FE}))$$

need be specified



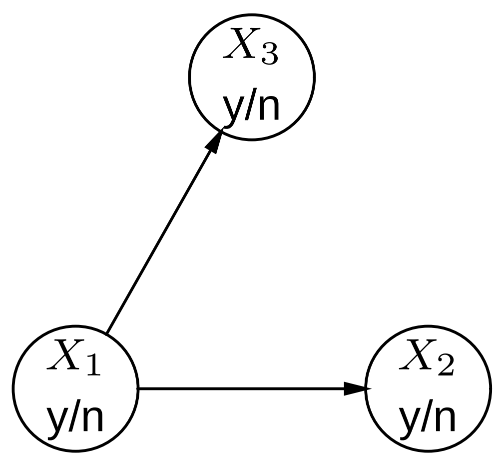
# Independence relation

Let $X, Y, Z \subseteq V$ be *sets of (random) variables*, and let $P$ be a probability distribution of $V$ then $X$ is called conditionally independent of $Y$ given $Z$, denoted as

$$X \perp\!\!\!\perp_P Y \mid Z, \quad \text{iff} \quad P(X \mid Y, Z) = P(X \mid Z)$$

Note: This relation is completely defined in terms of the probability distribution $P$, but there is *a relationship to graphs*, for example:

$$\{X_2\} \perp\!\!\!\perp_P \{X_3\} \mid \{X_1\}$$



# How to define an independence relation?

- List all the instances of $\perp\!\!\!\perp$
- List some of the instances of $\perp\!\!\!\perp$ and add axioms from which other instances can be derived
- Define a joint probability distribution $P$ and look into the numbers to see which instances of the independence relation $\perp\!\!\!\perp$ hold (this yields $\perp\!\!\!\perp_P$)
- Use a graph to encode $\perp\!\!\!\perp$, which yields $\perp\!\!\!\perp_G$ (so, what type of graph — directed, undirected, chain?)

# Explicit enumeration

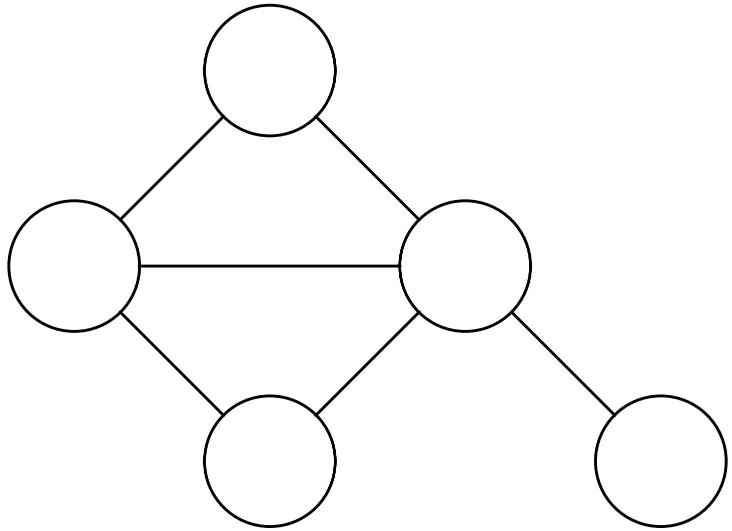
Consider $V = \{1,2,3,4\}$ and $\perp\!\!\!\perp$:

$\{1\} \perp\!\!\!\perp \{4\} \mid \varnothing$
$\{4\} \perp\!\!\!\perp \{3\} \mid \{1\}$
$\{4\} \perp\!\!\!\perp \{1\} \mid \varnothing$
$\{3\} \perp\!\!\!\perp \{4\} \mid \{2\}$
$\{1,2\} \perp\!\!\!\perp \{4\} \mid \varnothing$
$\{4\} \perp\!\!\!\perp \{3\} \mid \{2\}$
$\{4\} \perp\!\!\!\perp \{1,2\} \mid \varnothing$
$\{2\} \perp\!\!\!\perp \{4\} \mid \{3\}$
$\{1,2,3\} \perp\!\!\!\perp \{4\} \mid \varnothing$
$\{2\} \perp\!\!\!\perp \{1\} \mid \{4\}$
$\{2\} \perp\!\!\!\perp \{1\} \mid \varnothing$
$\{2\} \perp\!\!\!\perp \{4\} \mid \{1,3\}$
$\{2\} \perp\!\!\!\perp \{1,4\} \mid \varnothing$
$\{4\} \perp\!\!\!\perp \{1\} \mid \{2,3\}$
$\{3\} \perp\!\!\!\perp \{4\} \mid \{1\}$
$\{4\} \perp\!\!\!\perp \{2\} \mid \{3\}$

$\{4\} \perp\!\!\!\perp \{2\} \mid \{1\}$
$\{3\} \perp\!\!\!\perp \{4\} \mid \varnothing$
$\{1\} \perp\!\!\!\perp \{4\} \mid \{2\}$
$\{4\} \perp\!\!\!\perp \{3\} \mid \varnothing$
$\{4\} \perp\!\!\!\perp \{1\} \mid \{2\}$
$\{2,3\} \perp\!\!\!\perp \{4\} \mid \varnothing$
$\{1\} \perp\!\!\!\perp \{4\} \mid \{3\}$
$\{4\} \perp\!\!\!\perp \{2,3\} \mid \varnothing$
$\{1\} \perp\!\!\!\perp \{2\} \mid \{4\}$
$\{1\} \perp\!\!\!\perp \{2\} \mid \varnothing$
$\{4\} \perp\!\!\!\perp \{3\} \mid \{1,2\}$
$\{2,4\} \perp\!\!\!\perp \{1\} \mid \varnothing$
$\{1\} \perp\!\!\!\perp \{4\} \mid \{2,3\}$
$\{2\} \perp\!\!\!\perp \{4\} \mid \{1\}$
$\{4\} \perp\!\!\!\perp \{1\} \mid \{3\}$

$\{2\} \perp\!\!\!\perp \{4\} \mid \varnothing$
$\{4\} \perp\!\!\!\perp \{2,3\} \mid \{1\}$
$\{4\} \perp\!\!\!\perp \{2\} \mid \varnothing$
$\{1,3\} \perp\!\!\!\perp \{4\} \mid \{2\}$
$\{1,3\} \perp\!\!\!\perp \{4\} \mid \varnothing$
$\{4\} \perp\!\!\!\perp \{1,3\} \mid \{2\}$
$\{4\} \perp\!\!\!\perp \{1,3\} \mid \varnothing$
$\{1,2\} \perp\!\!\!\perp \{4\} \mid \{3\}$
$\{4\} \perp\!\!\!\perp \{1,2,3\} \mid \varnothing$
$\{3\} \perp\!\!\!\perp \{4\} \mid \{1,2\}$
$\{1,4\} \perp\!\!\!\perp \{2\} \mid \varnothing$
$\{4\} \perp\!\!\!\perp \{2\} \mid \{1,3\}$
$\{1\} \perp\!\!\!\perp \{2,4\} \mid \varnothing$
$\{4\} \perp\!\!\!\perp \{1,2\} \mid \{3\}$
$\{2,3\} \perp\!\!\!\perp \{4\} \mid \{1\}$

# As an undirected graph

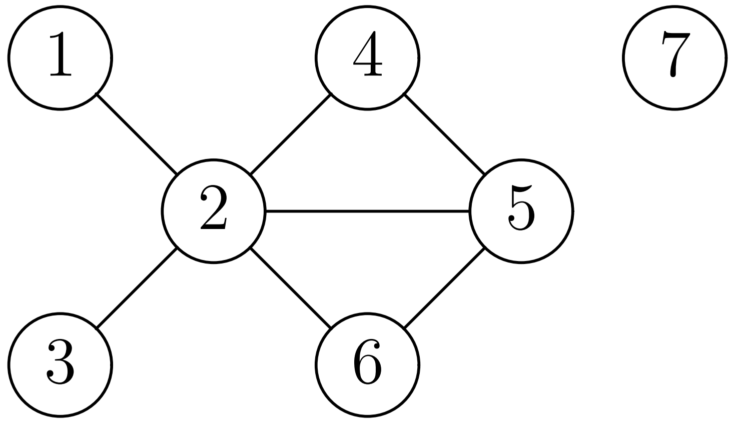


Basic idea:

- Each variable $V$ is represented as a vertex in an undirected graph $G = (V(G), E(G))$, with set of vertices $V(G)$ and set of edges $E(G)$
- the independence relation $\perp\!\!\!\perp_G$ is encoded as the absence of edges; a missing edge between vertices $u$ and $v$ indicates that random variables $X_u$ and $X_v$ are (conditionally) independent $=$ (u-)separation

# Example

Consider the following undirected graph $G$:




- $\{1\} \perp\!\!\!\perp_G \{3,6\} \mid \{2\}$
- $\{4\} \perp\!\!\!\perp_G \{6\} \mid \{2,5\}$
- $\{4\} \perp\!\!\!\perp_G \{6\} \mid \{1,2,3,5\}$
- $\{1\} \not\perp\!\!\!\perp_G \{5\} \mid \{4\}$, as the path $1-2-5$ does not contain $4$
- $\{1,5,6\} \perp\!\!\!\perp_G \{7\} \mid \varnothing$

# D-map and I-map for $\perp\!\!\!\perp_P$

Let $P$ be probability distribution of $X$. Let $G = (V(G), E(G))$ be an undirected graph, then for each $U, W, Z \subseteq V(G)$:

- $G$ is called an undirected dependence map, D-map for short, if

$$X_U \perp\!\!\!\perp_P X_W \mid X_Z \Rightarrow U \perp\!\!\!\perp_G W \mid Z$$

- $G$ is called an undirected independence map, I-map for short, if

$$U \perp\!\!\!\perp_G W \mid Z \Rightarrow X_U \perp\!\!\!\perp X_W \mid X_Z$$

- $G$ is called an undirected perfect map, or P-map for short, if $G$ is both a D-map and an I-map, or, equivalently

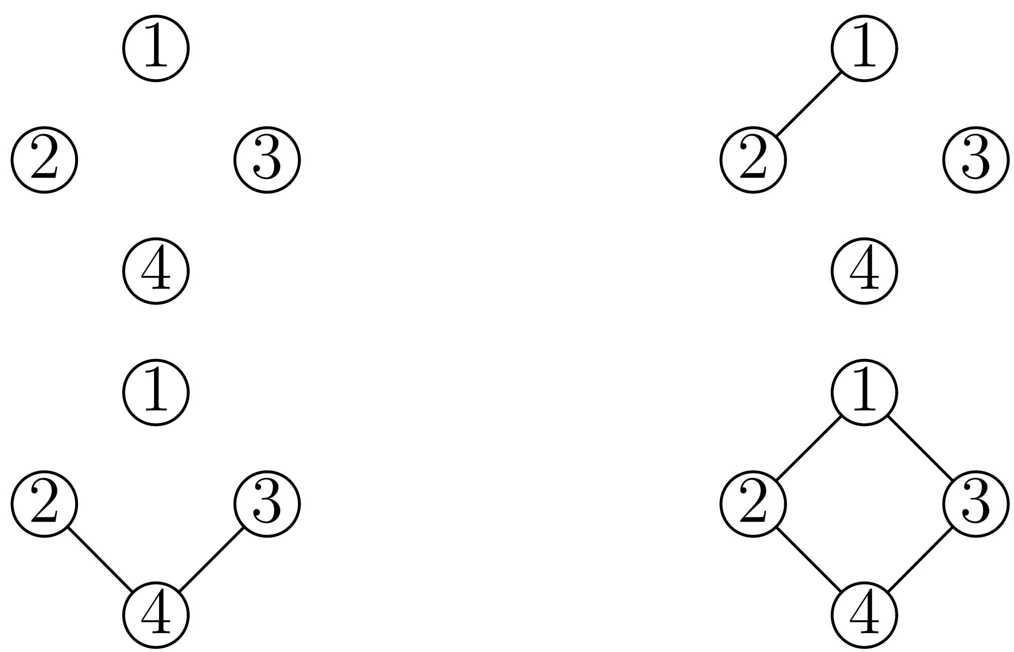
$$X_U \perp\!\!\!\perp_P X_W \mid X_Z \Longleftrightarrow U \perp\!\!\!\perp_G W \mid Z$$

# Examples D-maps

Let $V = \{1, 2, 3, 4\}$ be a set and $X_V$ the corresponding set of random variables, and consider the independence relation $\perp\!\!\!\perp_P$, defined by

$$\{X_1\} \perp\!\!\!\perp_P \{X_4\} \mid \{X_2, X_3\}$$
$$\{X_2\} \perp\!\!\!\perp_P \{X_3\} \mid \{X_1, X_4\}$$

The following undirected graphs are examples of D-maps:

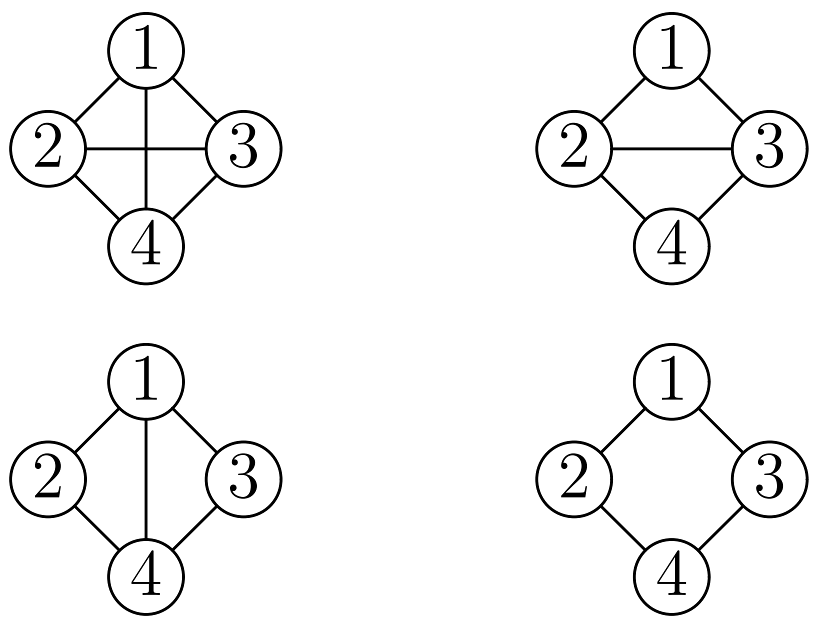


# Examples of I-maps

Let $V = \{1, 2, 3, 4\}$ be a set with random variables $X_V$, and consider the independence relation $\perp\!\!\!\perp_P$:

$$\{X_1\} \perp\!\!\!\perp_P \{X_4\} \mid \{X_2, X_3\}$$
$$\{X_2\} \perp\!\!\!\perp_P \{X_3\} \mid \{X_1, X_4\}$$

The following undirected graphs are examples of I-maps:




(So, what is the P-map?)

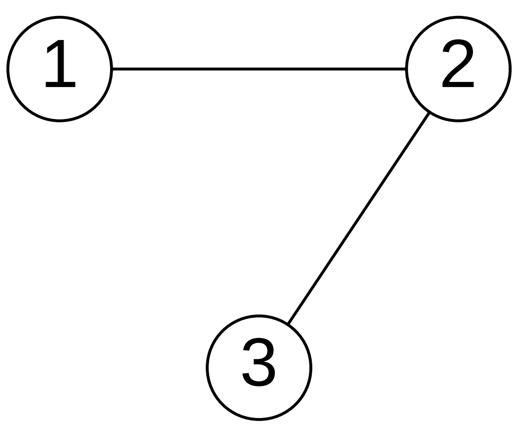
# Markov network

A pair $\mathcal{M} = (G, P)$, where

- $G = (V(G), E(G))$ is an *undirected* graph with set of vertices $V(G)$ and set of edges $E(G)$,
- $P$ is a joint probability distribution of $X_{V(G)}$, and
- $G$ is an *I-map* of $P$

is said to be a Markov network or Markov random field

Example $\mathcal{M} = (G, \phi) = (G, P)$:




Potential:

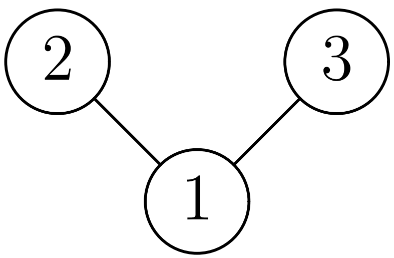
$$\phi(X_1, X_2, X_3) = \psi(X_1, X_2)\tau(X_2, X_3),$$

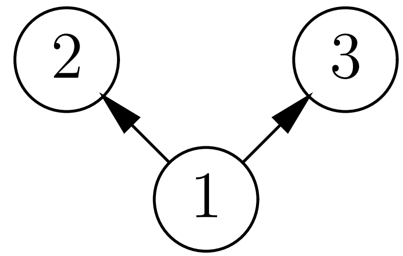
or joint probability distribution:

$$P(X_1, X_2, X_3) = \frac{P(X_1,X_2)P(X_2,X_3)}{P(X_2)}$$

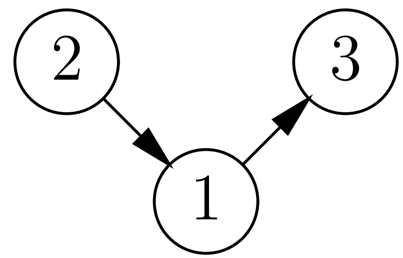
# Expressiveness: directed vs undirected

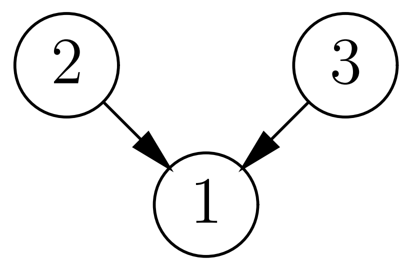
Directed graphs are more subtle when it comes to expressing independence information than undirected graphs




vs

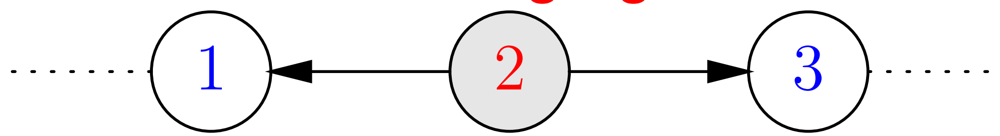


# d-Separation: 3 situations

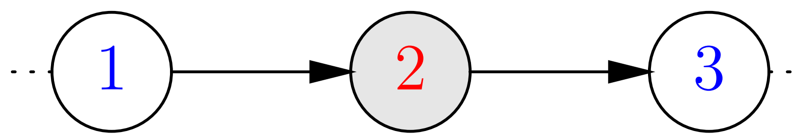
A chain $k$ (= path in undirected underlying graph) in an acyclic directed graph $G = (V(G), A(G))$ can be blocked:

Diverging



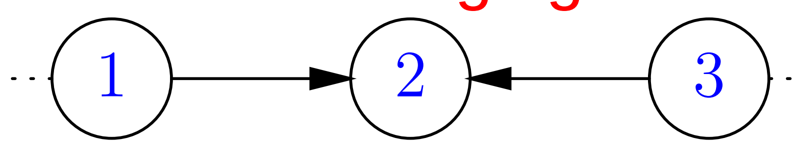

$2$ blocks (d-separates) $1$ and $3$: $\{1\} \perp\!\!\!\perp \{3\} \mid \{2\}$

Serial




$2$ blocks (d-separates) $1$ and $3$: $\{1\} \perp\!\!\!\perp \{3\} \mid \{2\}$

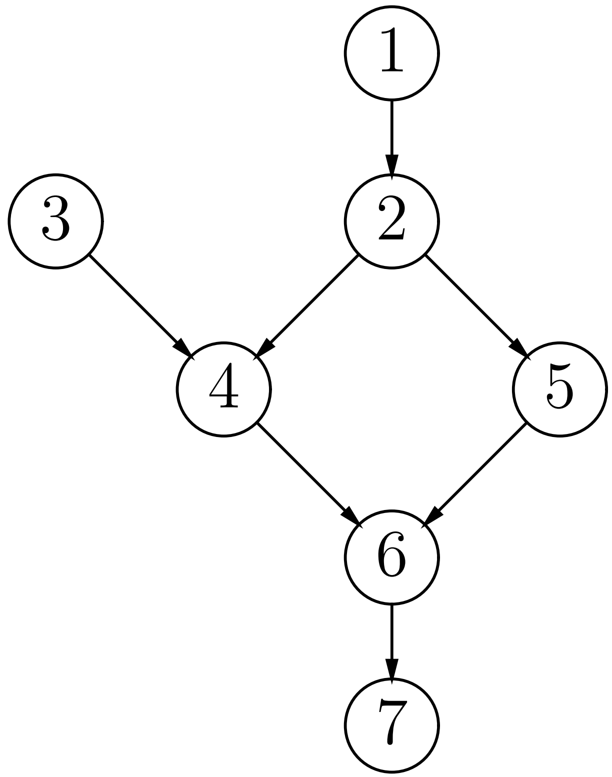
Converging




$2$ d-connects $1$ and $3$: $\{1\} \not\!\perp\!\!\!\perp \{3\} \mid \{2\}$

(same holds for successors of $2$); note $\{1\} \perp\!\!\!\perp \{3\} \mid \varnothing$

# Example blockage



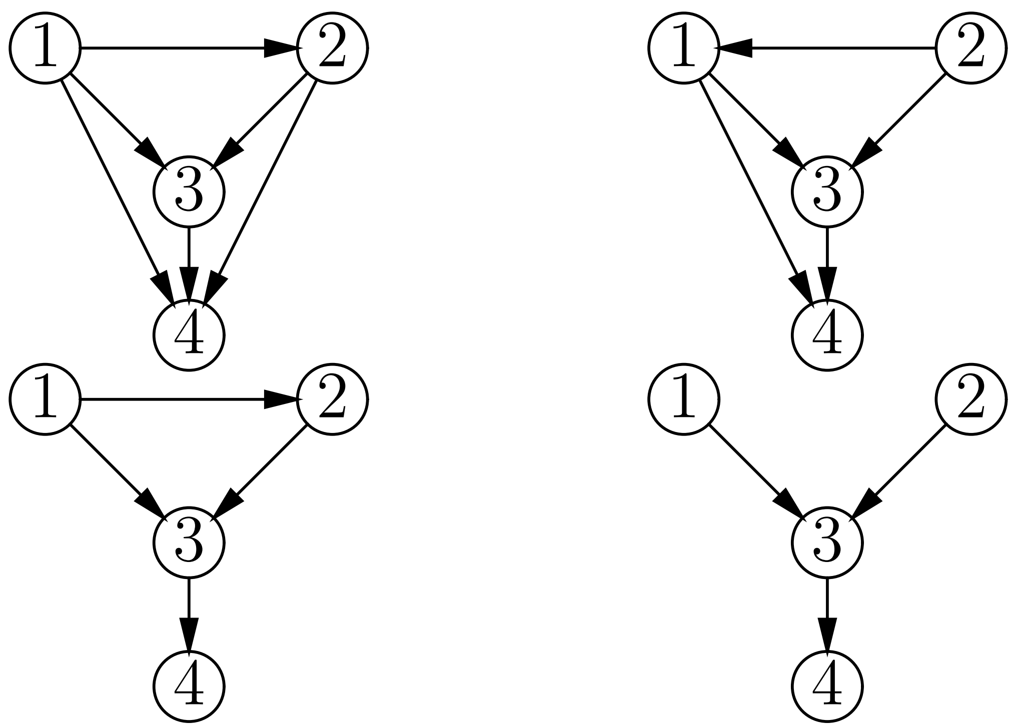

- The chain $4, 2, 5$ from $4$ to $5$ is blocked by $\{2\}$
- The chain $1, 2, 5, 6$ from $1$ to $6$ is blocked by $\{5\}$, and also by $\{2\}$ and $\{2, 5\}$
- The chain $3, 4, 6, 5$ from $3$ to $5$ is blocked by $\{4\}$ and $\{4, 6\}$, but *not* by $\{6\}$

# Examples directed I-maps

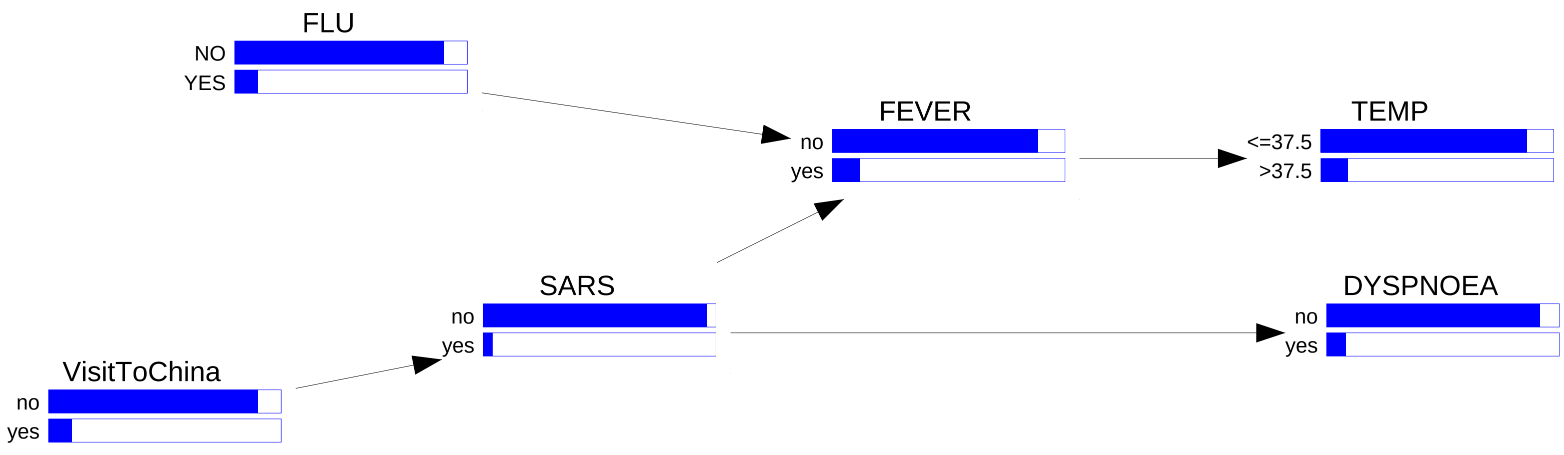
Consider the following independence relation $\perp\!\!\!\perp_P$:

$$
\begin{aligned}
\{X_1\} \quad &\perp\!\!\!\perp_P \quad \{X_2\} \mid \varnothing \\
\{X_1, X_2\} \quad &\perp\!\!\!\perp_P \quad \{X_4\} \mid \{X_3\}
\end{aligned}
$$

and the following directed I-maps of $P$:



# Find the independences



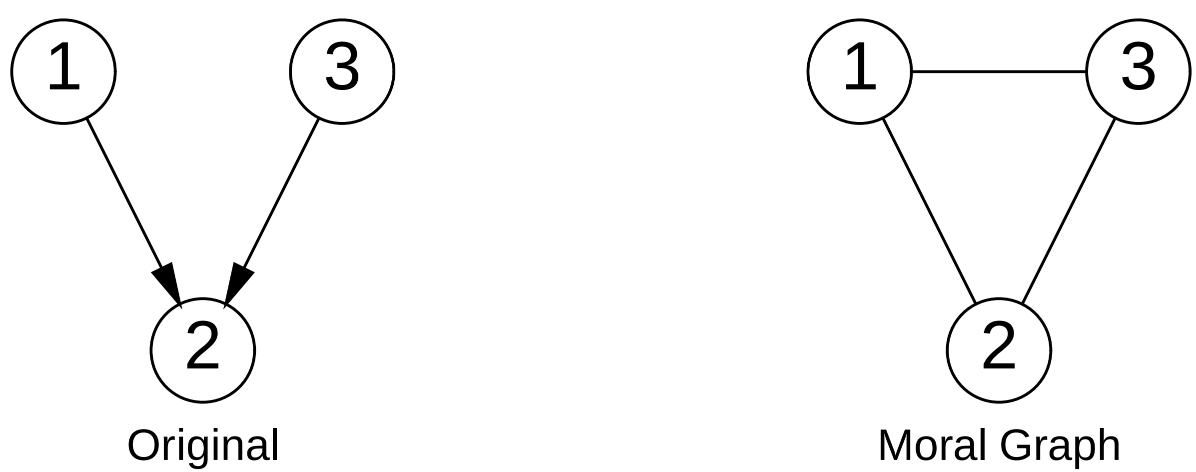

Examples:

- FLU $\perp\!\!\!\perp$ VisitToChina $\mid \emptyset$
- FLU $\perp\!\!\!\perp$ SARS $\mid \emptyset$
- 
- 

# Relationship directed and undirected graphs

- Directed graphs contain independences that become dependences after conditioning (instantiating variables)
- Undirected graphs do not have this property
- However, undirected subgraphs can be generated, by making potentially dependent parents of a child dependent

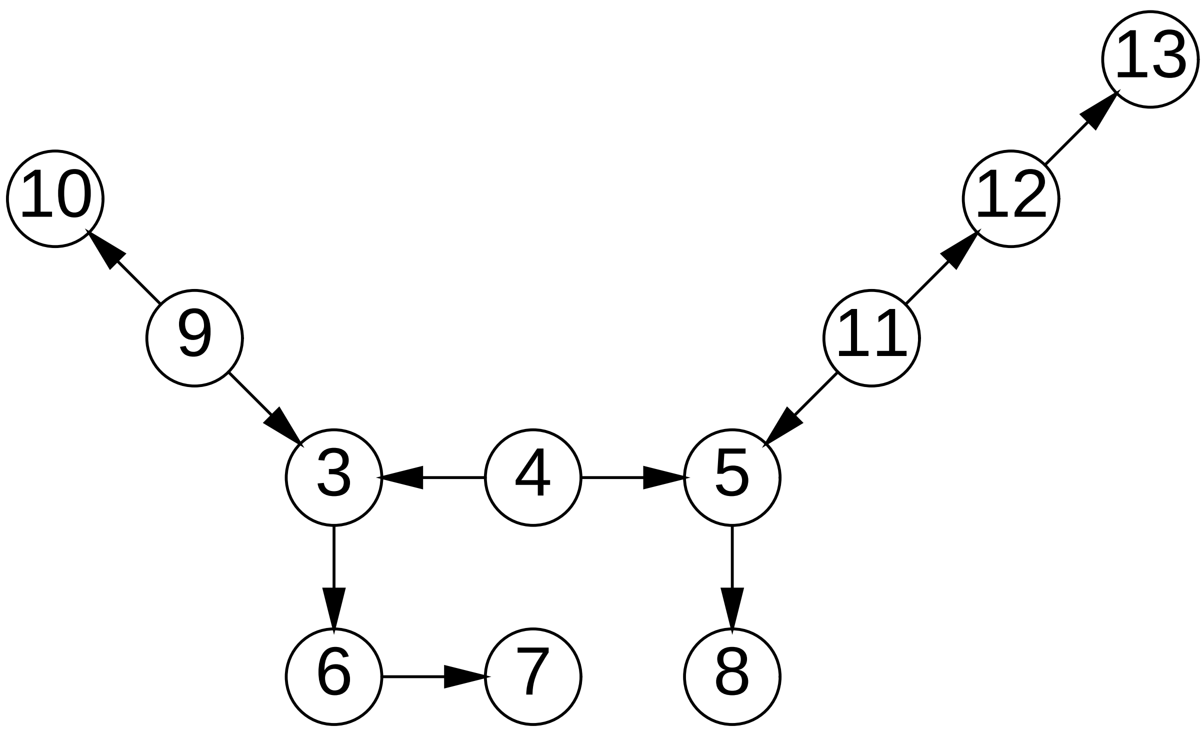
Example:



# Moralisation

Let $G$ be an acyclic directed graph; its associated undirected moral graph $G^m$ can be constructed by moralisation:

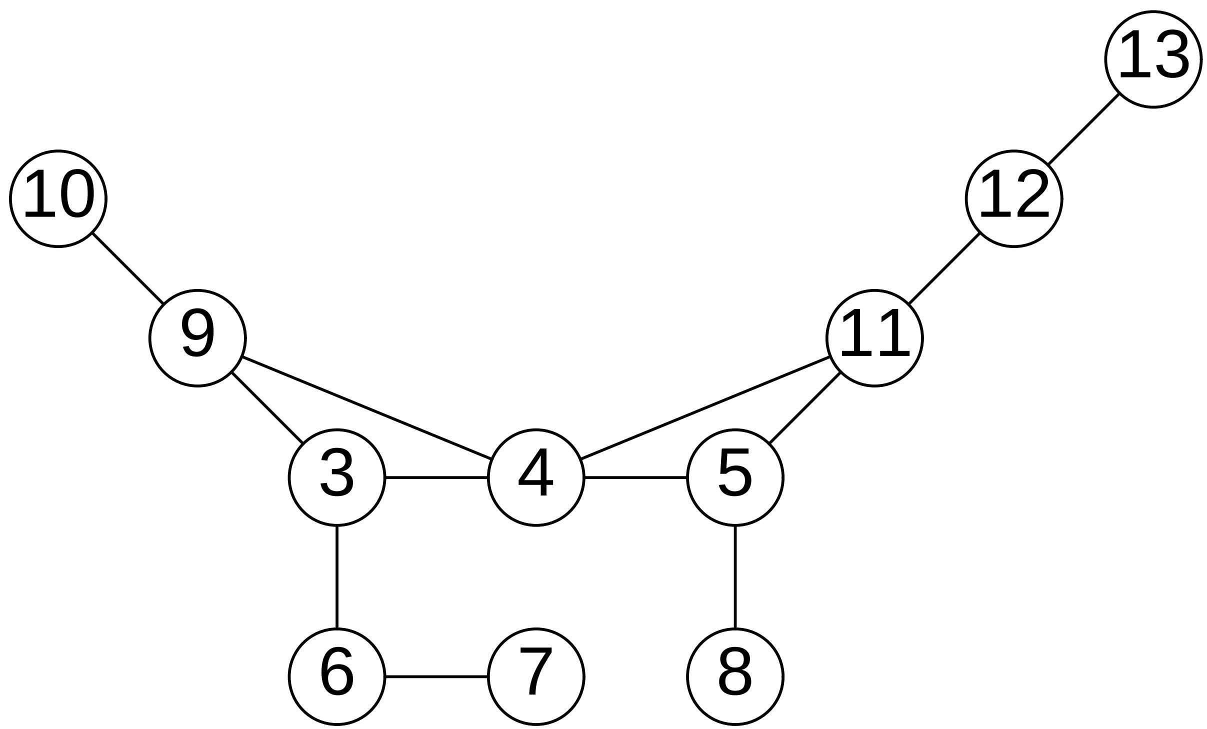
1. add lines to all non-connected vertices, which have a common child, or descendant of a common child, and
2. replace each arc with a line in the resulting graph



# Moralisation

Let $G$ be an acyclic directed graph; its associated undirected moral graph $G^m$ can be constructed by moralisation:

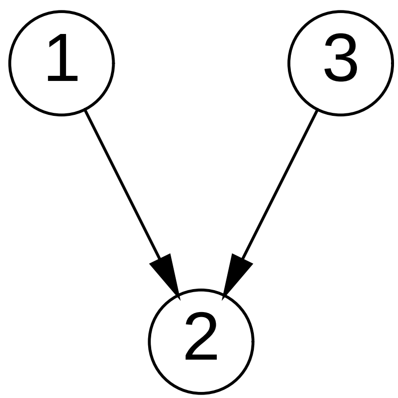
1. add lines to all non-connected vertices, which have a common child, or descendant of a common child, and
2. replace each arc with a line in the resulting graph



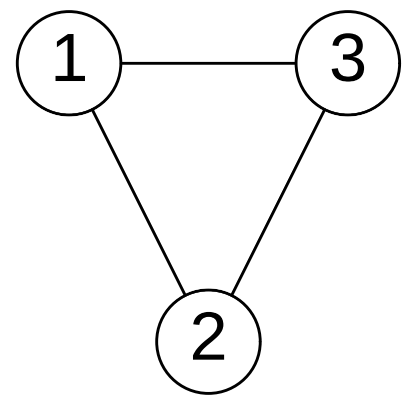
# Comments

- Resulting undirected (moral) graph is an I-map of the associated probability distribution
- However, it contains too many dependences!

Example: $\{1\} \perp\!\!\!\perp_G^d \{3\} \mid \varnothing$, whereas $\{1\} \not\perp\!\!\!\perp_{G^m} \{3\} \mid \varnothing$



Original

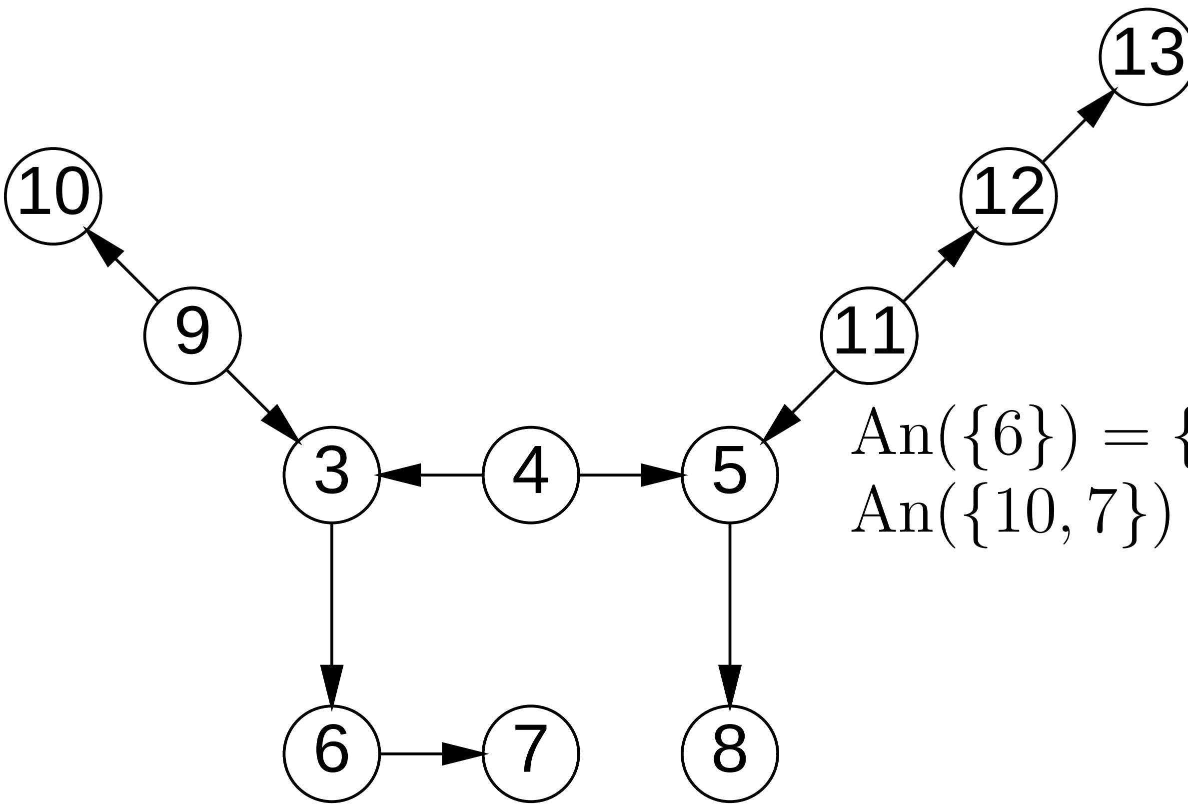


Moral Graph

- Conclusion: make moralisation 'dynamic' (i.e. a function of the set on which we condition)
- For this the notion of 'ancestral set' is required

# Ancestral set

Let $G = (V(G), A(G))$ be an acyclic directed graph, then if for $W \subseteq V(G)$ it holds that $\pi(v) \subseteq W$ for all $v \in W$, then $W$ is called an ancestral set of $W$. $\mathrm{An}(W)$ denotes the smallest ancestral set containing $W$




$$\mathrm{An}(\{6\}) = \{3, 4, 6, 9\}$$
$$\mathrm{An}(\{10, 7\}) = \{7, 6, 3, 4, 9, 10\}$$

# ‘Dynamic’ moralisation

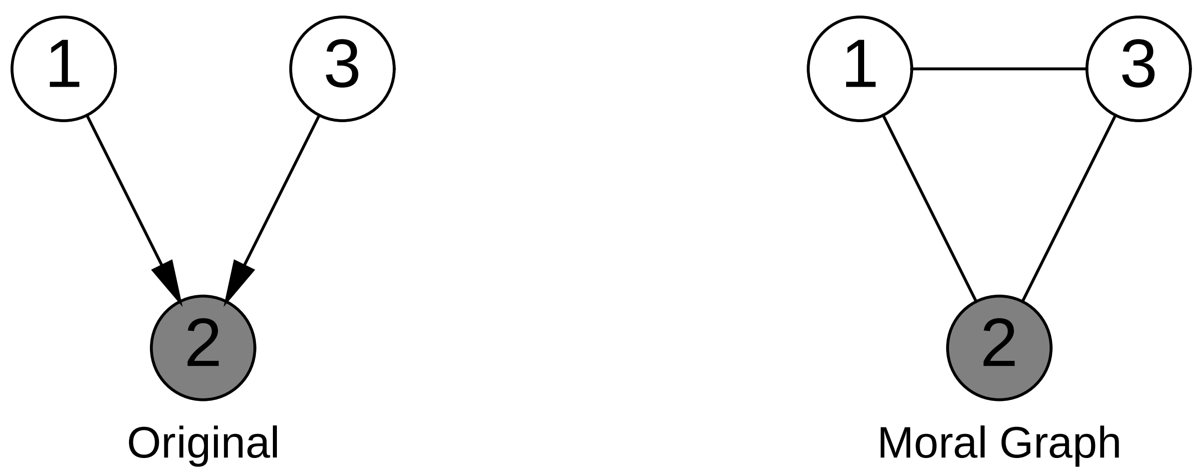
Let $P$ be a joint probability distribution of a Bayesian network $\mathcal{B} = (G, P)$, then

$$X_U \perp\!\!\!\perp_P X_V \mid X_W$$

holds iff $U$ and $V$ are (u-)separated by $W$ in the moral induced subgraph $G^m$ of $G$ with vertices $\mathrm{An}(U \cup V \cup W)$

Example:

$$X_1 \not\perp\!\!\!\perp_P X_3 \mid X_2; \quad \mathrm{An}(\{1,2,3\}) = \{1,2,3\}$$



# ‘Dynamic’ moralisation

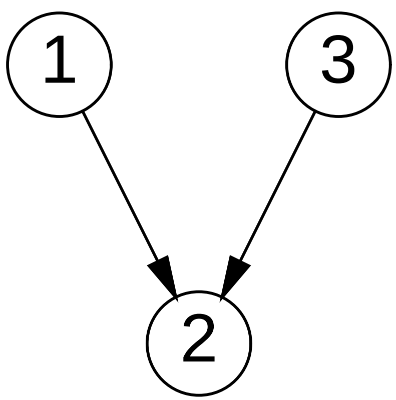
Let $P$ be a joint probability distribution of a Bayesian network $\mathcal{B} = (G, P)$, then

$$X_U \perp\!\!\!\perp_P X_V \mid X_W$$

holds iff $U$ and $V$ are (u-)separated by $W$ in the moral induced subgraph $G^m$ of $G$ with vertices $\mathrm{An}(U \cup V \cup W)$

Example:

$$X_1 \perp\!\!\!\perp_P X_3 \mid \varnothing; \quad \mathrm{An}(\{1,3\}) = \{1,3\}$$

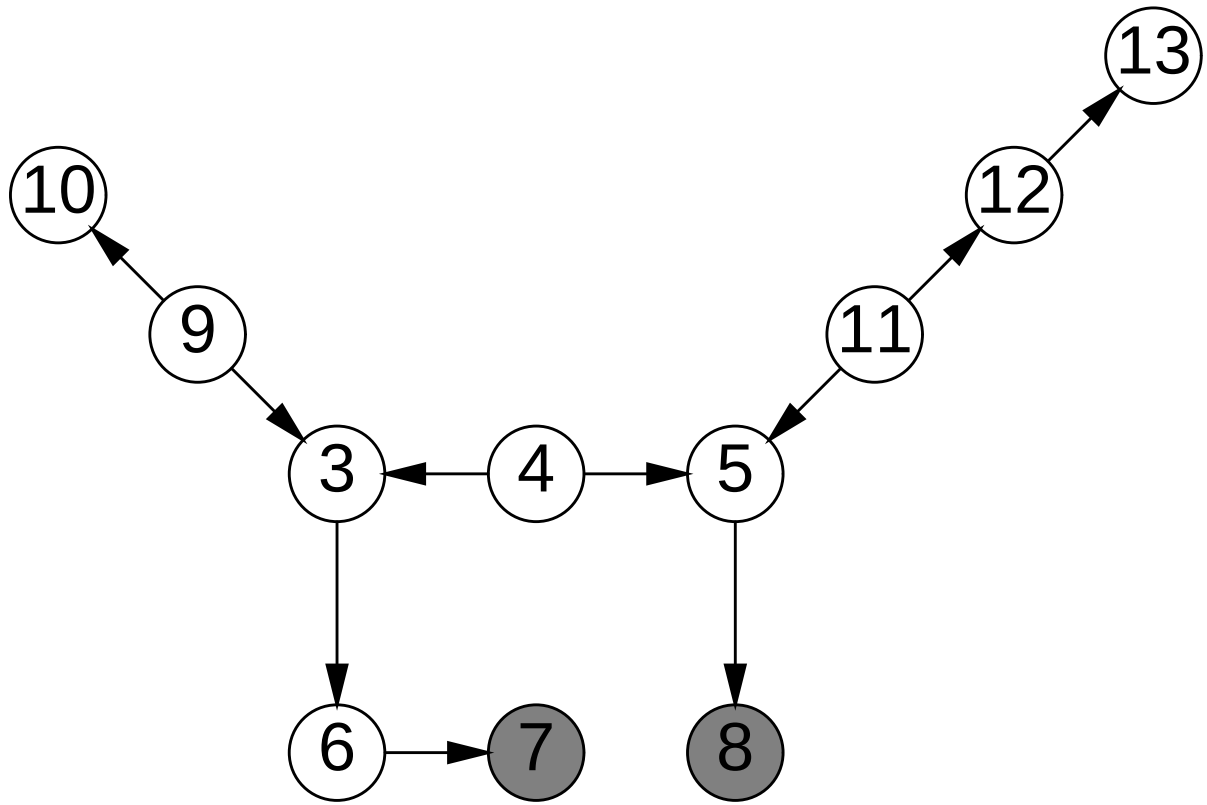


Original

Moral Graph

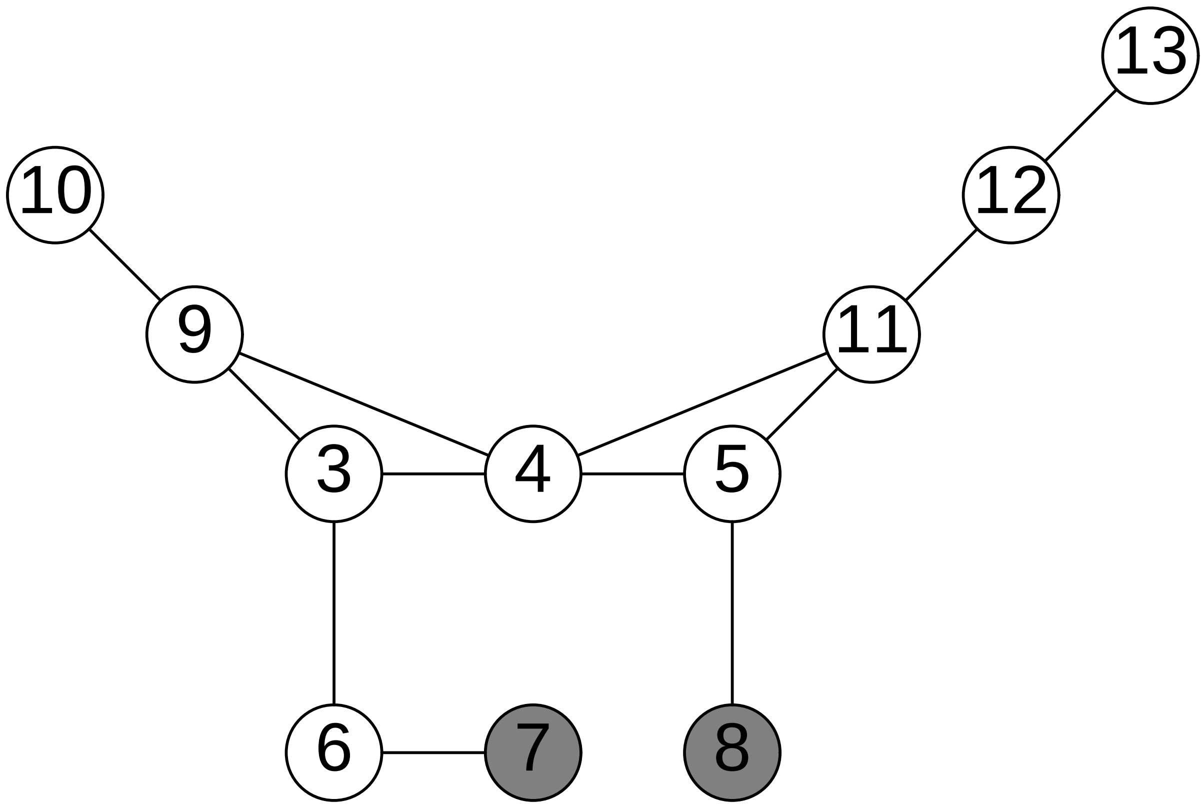
# Example (1)

$\{10\} \not\perp\!\!\!\perp_G^d \{13\} \mid \{7, 8\}$



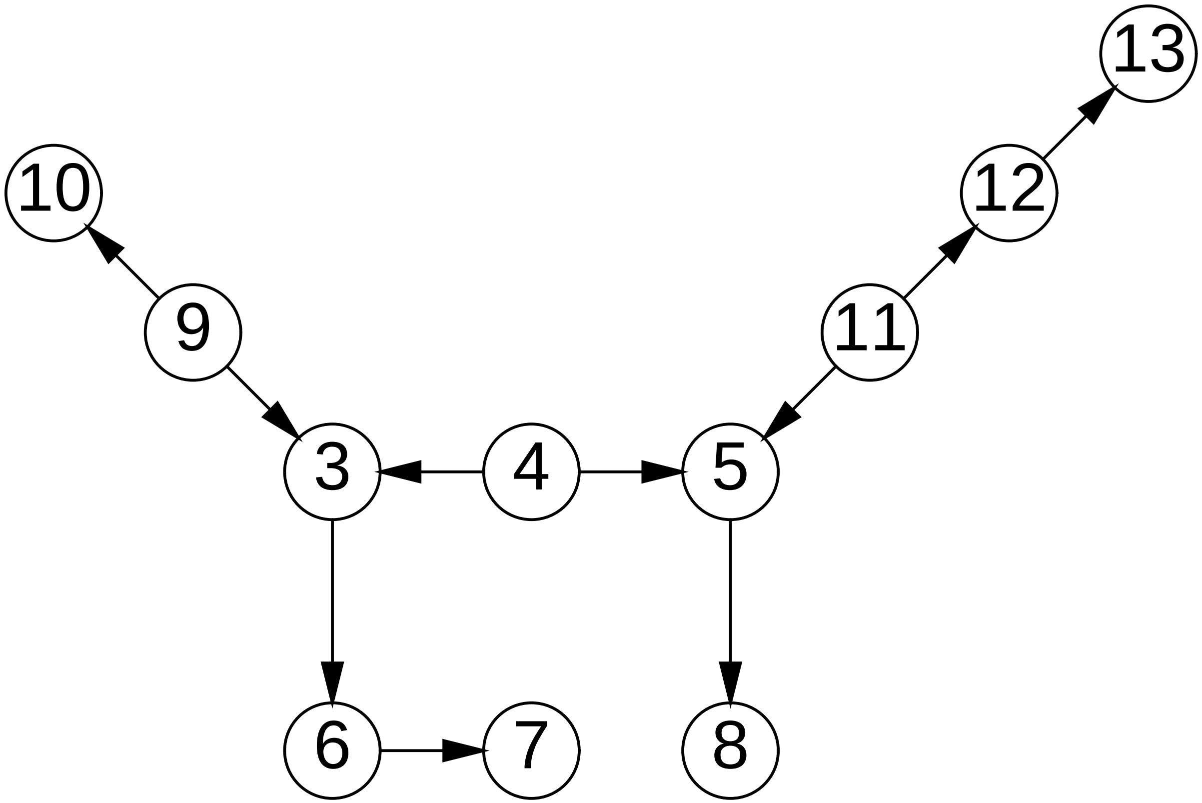
# Example (1)

$$\{10\} \not\perp\!\!\!\perp_{G^m_{\mathrm{An}(\{10,7,8,13\})}} \{13\} \mid \{7, 8\}$$



# Example (2)

$\{10\} \perp\!\!\!\perp_G^d \{13\} \mid \varnothing$



# Example (2)

$\{10\} \not\perp\!\!\!\perp_{G^m_{\mathrm{An}(\{10,13\})}} \{13\} \mid \varnothing$

# Conclusions

- Conditional independence is defined as a logic that supports:
  - symbolic reasoning about dependence and independence information
  - makes it possible to abstract away from the numerical detail of probability distributions
  - the process of assessing probability distributions
- Looking at graphs makes it easier to find probability distributions that are equivalent (important in learning)
- Conditional independence is currently being extended towards causal independence (a logic of causality) = maximal ancestral graphs