Markov Independence–Part I & II

Andrei A. Markov (1856 – 1922)

The focus of today ...

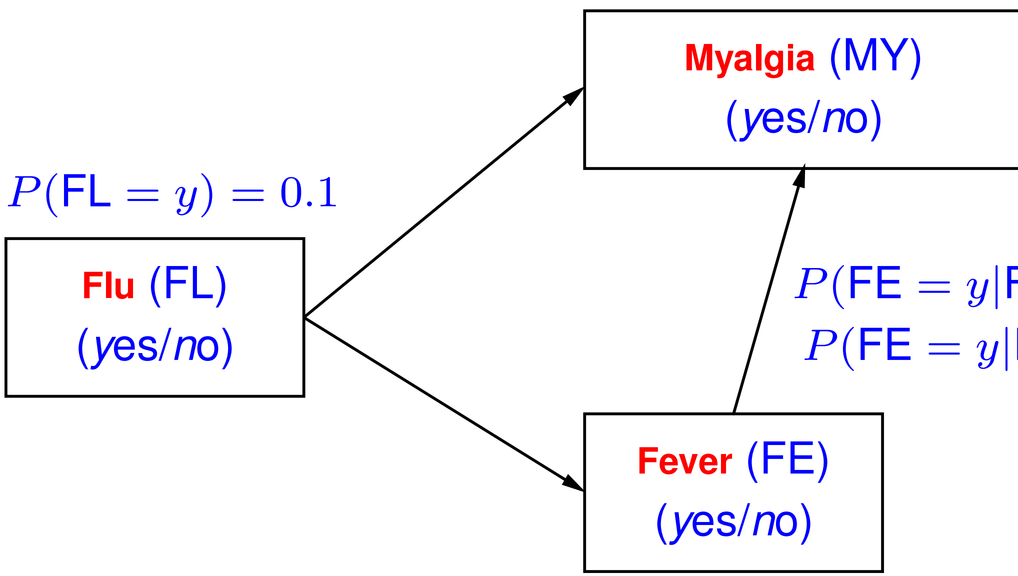
- Independence and probabilistic reasoning
- Why is representation of independence important?
  - To describe scientific results (in psychology, sociology, physics, biology, ...)
  - It is the foundation of statistical learning
- Bayes-ball algorithm
- Ways to represent independence information
- Properties of independence (axioms)

A Bayesian network

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

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

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

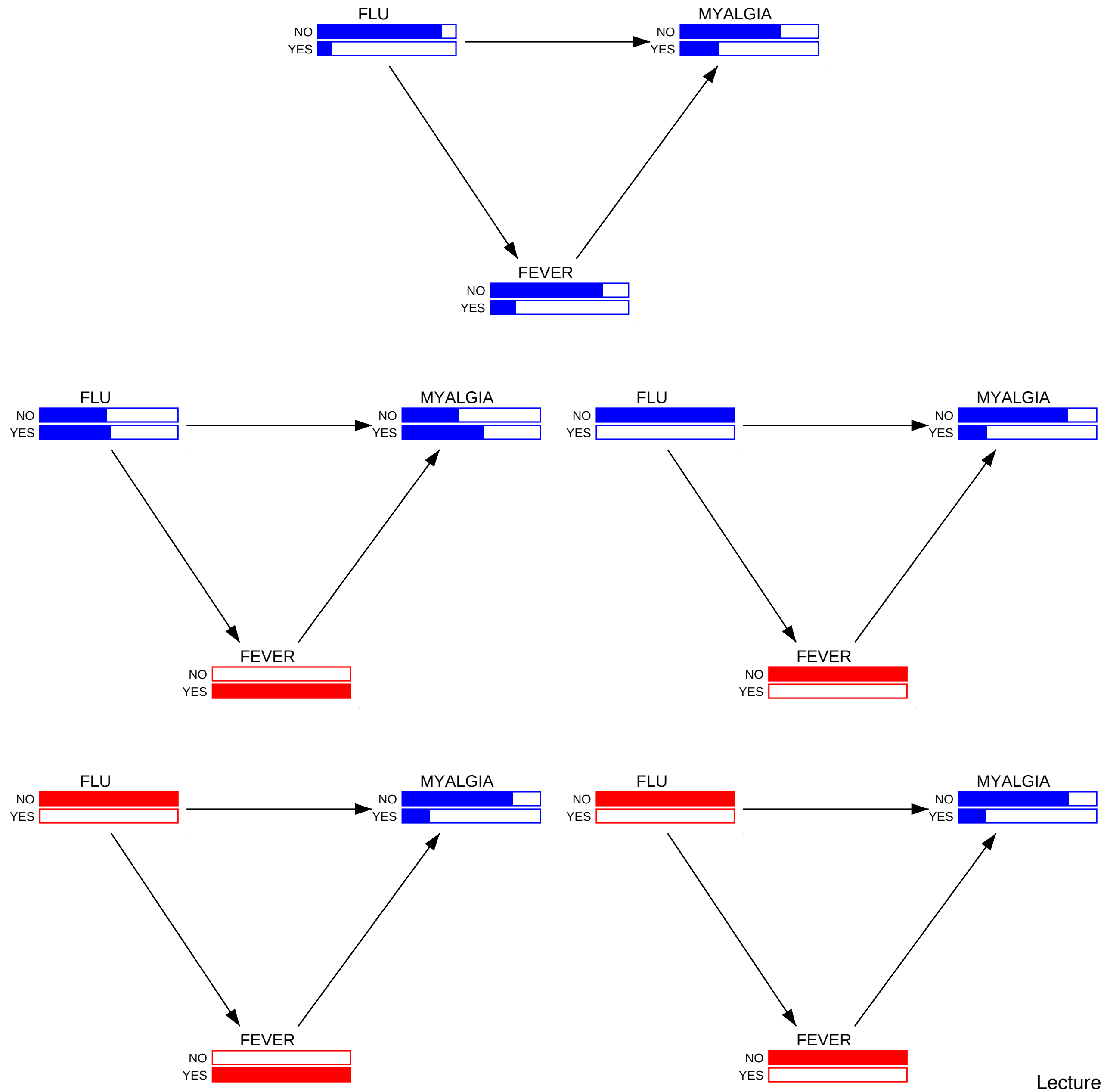


$$P(\text{FE} = y|\text{FL} = y) = 0.95$$
$$P(\text{FE} = y|\text{FL} = n) = 0.1$$

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

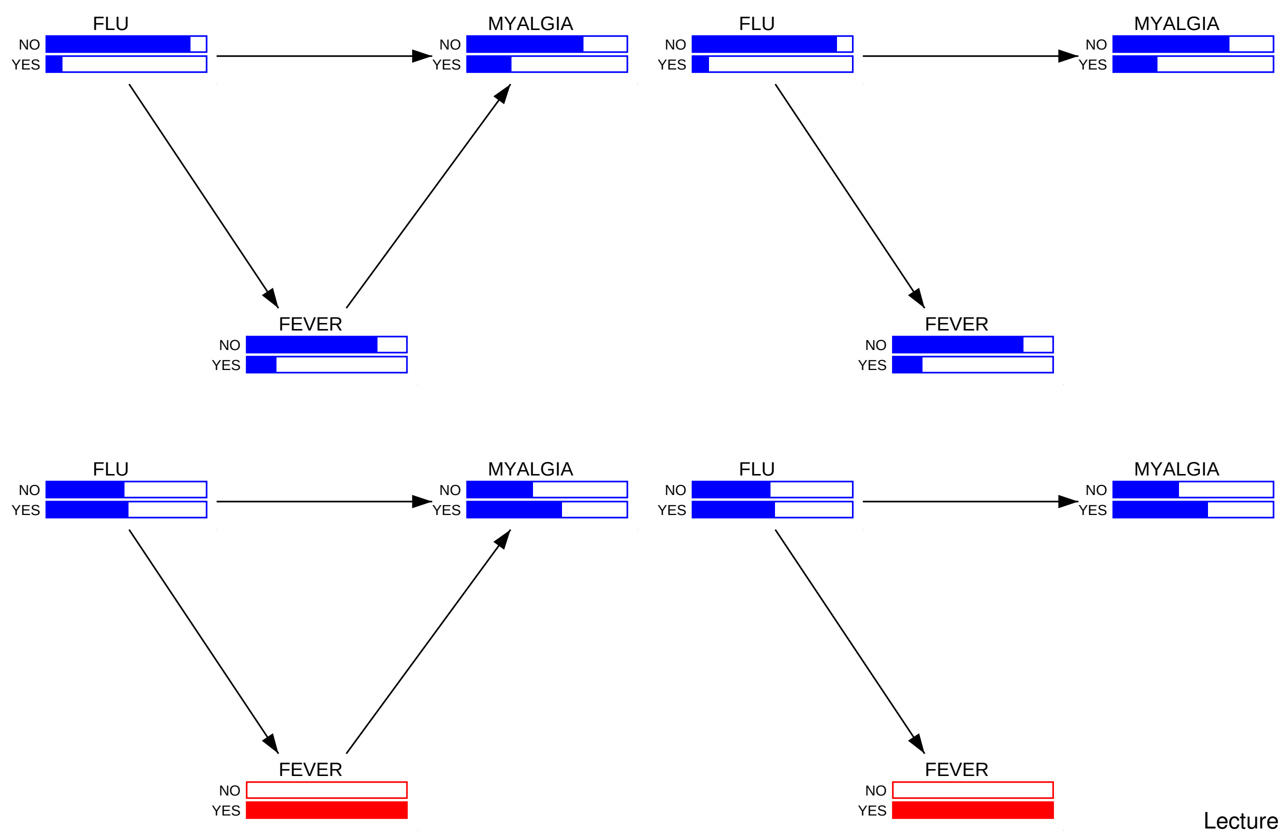


Independence and reasoning

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

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

need be specified



Importance of independence

- Compact knowledge *representation*
  - Simplify the model structure
  - Reduce parameter estimation
- Efficient *reasoning* (compute posterior probabilities) and *learning* of models
- Describe scientific results (Markov processes), e.g., in physics (Brownian motion), in economy (stock market fluctuations)

- Role of graphical models
  - Testing for conditional independence from a joint distribution is time consuming
  - Can be directly read off from the graphical model

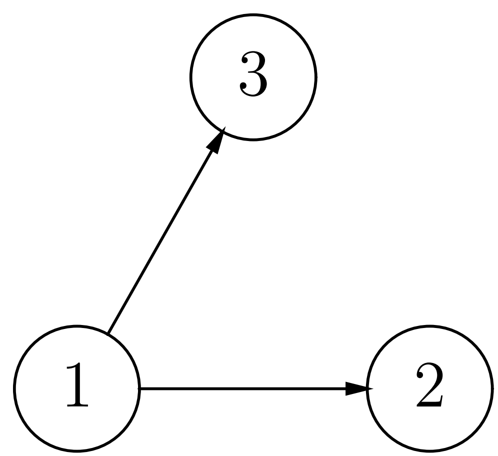
Independence relation

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:

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



Equivalences with indepedence

The following conditions are equivalent:

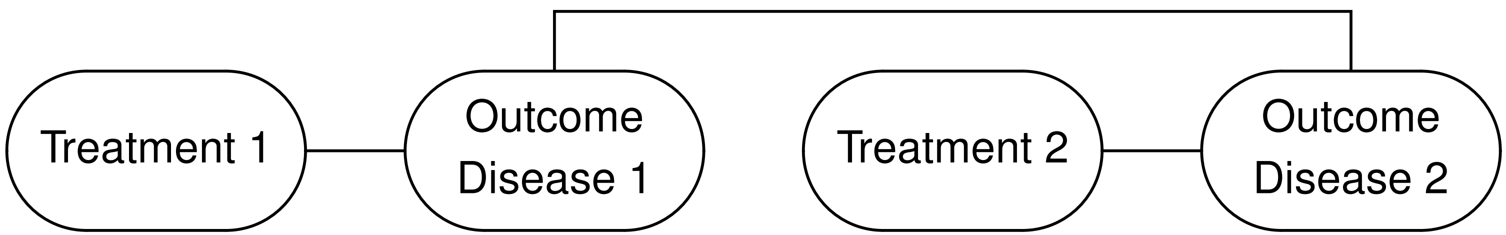
- $P(X \mid Y, Z) = P(X \mid Z)$ if $P(Y, Z) > 0$ (why?)
- $P(X, Y \mid Z) = P(X \mid Z)P(Y \mid Z)$ if $P(Y, Z) > 0$
- $P(X, Y, Z) = P(X \mid Z)P(Y \mid Z)P(Z)$
- $P(X, Y, Z) = P(X, Z)P(Y, Z)/P(Z)$ if $P(Z) > 0$

- $P(X \mid Y, Z)$ can be represented as the real function $\psi(X, Z)$, called a potential
- $P(X, Y \mid Z)$ can be written as $\phi(X, Z)\psi(Y, Z)$, with real potential functions ϕ and ψ

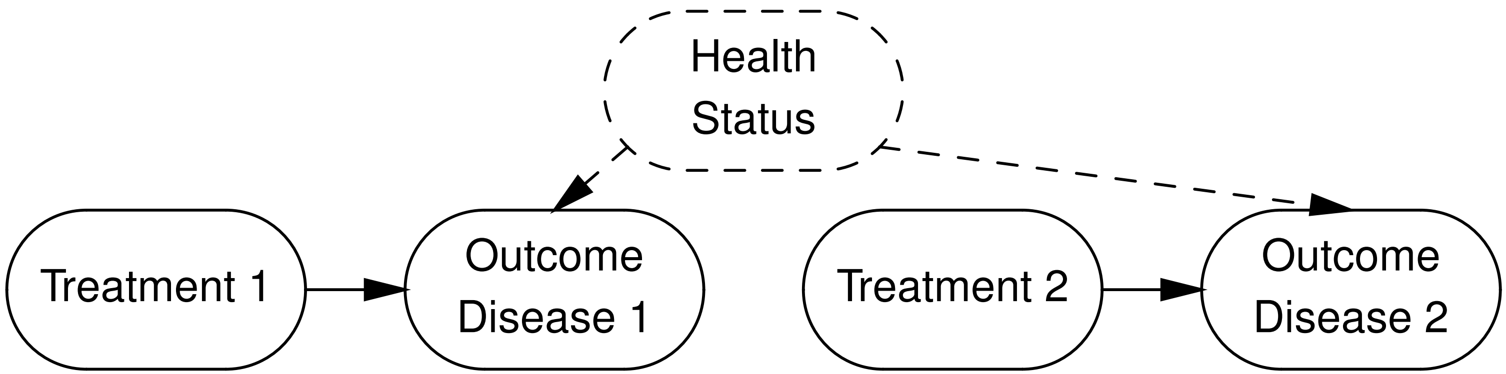
N.B. potentials are non-negative real functions, very similar to probability distributions, but they need not be normalised

Empirical sciences

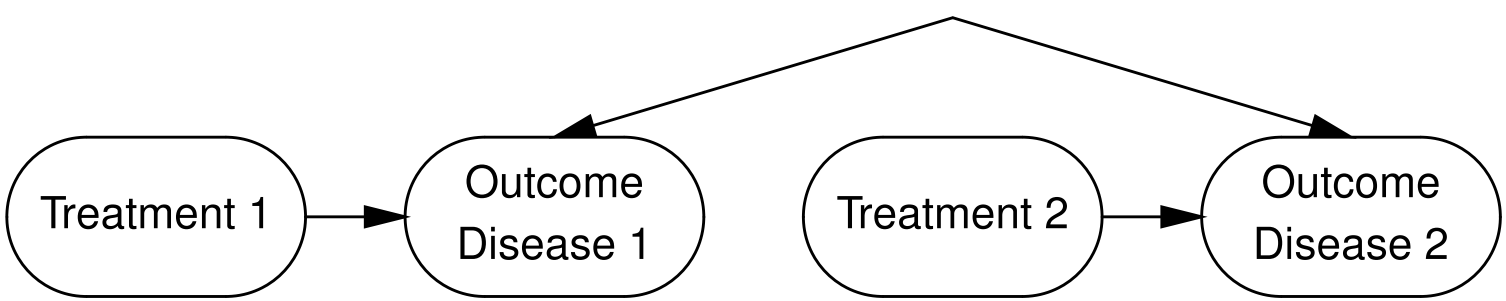
Result from conventional analysis:



Directed graph:



Mixed graph:



The $\perp\!\!\!\perp_P$ relation

The relation

$$X \perp\!\!\!\perp_P Y \mid Z$$

defines a ternary predicate

$$\perp\!\!\!\perp_P (X, Y, Z)$$

For this predicate particular properties hold, such as symmetry:

$$X \perp\!\!\!\perp_P Y \mid Z \Longleftrightarrow Y \perp\!\!\!\perp_P X \mid Z$$

These properties are in nature similar to properties as for equality $=$ (or some other relationship):

$$x = y \Longleftrightarrow y = x$$

(also called symmetry)

Properties of the $\perp\!\!\!\perp_P$ relation (1)

P1 Symmetry: If Y provides no new information about X given Z, then X provides no additional information about Y. Let $X, Y, Z \subseteq V$ be sets of variables, then:

$$X \perp\!\!\!\perp_P Y \mid Z \Longleftrightarrow Y \perp\!\!\!\perp_P X \mid Z$$

Proof:

$$X \perp\!\!\!\perp_P Y \mid Z \Leftrightarrow P(X \mid Y, Z) \stackrel{(1)}{=} P(X \mid Z)$$

$$\frac{P(X,Y,Z)}{P(Y,Z)} \stackrel{(1)}{=} \frac{P(X,Z)}{P(Z)}$$

$$\frac{P(X,Y,Z)}{P(X,Z)} \stackrel{(1)}{=} \frac{P(Y,Z)}{P(Z)}$$

$$P(Y \mid X, Z) \stackrel{(1)}{=} P(Y \mid Z) \Longleftrightarrow Y \perp\!\!\!\perp_P X \mid Z$$

Properties of the $\perp\!\!\!\perp_P$ relation (2)

P2 Decomposition: If both Y and W are irrelevant with regard to our knowledge of X given Z, then they are also irrelevant separately.
Let $X, Y, W, Z \subseteq V$ be disjoint sets of random variables:

$$X \perp\!\!\!\perp_P Y \cup W \mid Z \;\Rightarrow\; X \perp\!\!\!\perp_P Y \mid Z \quad \wedge \quad X \perp\!\!\!\perp_P W \mid Z$$

Proof:

$$\begin{aligned}
X \perp\!\!\!\perp_P Y \cup W \mid Z &\Leftrightarrow P(X \mid Y, W, Z) = P(X \mid Z) \quad (1) \\
P(X \mid Y, Z) &= \sum_W P(X \mid Y, W, Z) P(W \mid Y, Z) \\
&= \sum_W P(X \mid Z) P(W \mid Y, Z) \\
&= P(X \mid Z) \sum_W P(W \mid Y, Z) \\
&= P(X \mid Z) \cdot 1 = P(X \mid Z) \overset{(1)}{\Leftrightarrow} X \perp\!\!\!\perp_P Y \mid Z
\end{aligned}$$

Analogously we obtain the proof for $X \perp\!\!\!\perp_P W \mid Z$.

Properties of the $\perp\!\!\!\perp_P$ relation (3)

P3 Weak union: If both Y and W are irrelevant with regard to our knowledge of X given Z, then Y remains irrelevant for X given Z and W.
Let $X, Y, W, Z \subseteq V$ be disjoint sets of random variables:

$$X \perp\!\!\!\perp_P Y \cup W \mid Z \;\;\Rightarrow\;\; X \perp\!\!\!\perp_P Y \mid Z \cup W$$

Proof: $\cdots$ *DIY*

P4 Contraction: If Y is irrelevant to X given Z and if W is judged to be irrelevant to X after learning information about Y, then W must have been irrelevant prior to learning Y.
Let $X, Y, W, Z \subseteq V$ be disjoint sets of random variables of variables:

$$X \perp\!\!\!\perp_P Y \mid Z \;\;\wedge\;\; X \perp\!\!\!\perp_P W \mid Y \cup Z \;\;\Rightarrow\;\; X \perp\!\!\!\perp_P W \cup Y \mid Z$$

Proof: $\cdots$ *DIY*

Properties of the $\perp\!\!\!\perp_P$ relation (4)

P5 Intersection: Let Z be given. If Y is irrelevant to X after learning W, and W is irrelevant to X after learning Y, then neither Y, W nor their combination is relevant to X.
Let $X, Y, W, Z \subseteq V$ disjoint sets of random variables:

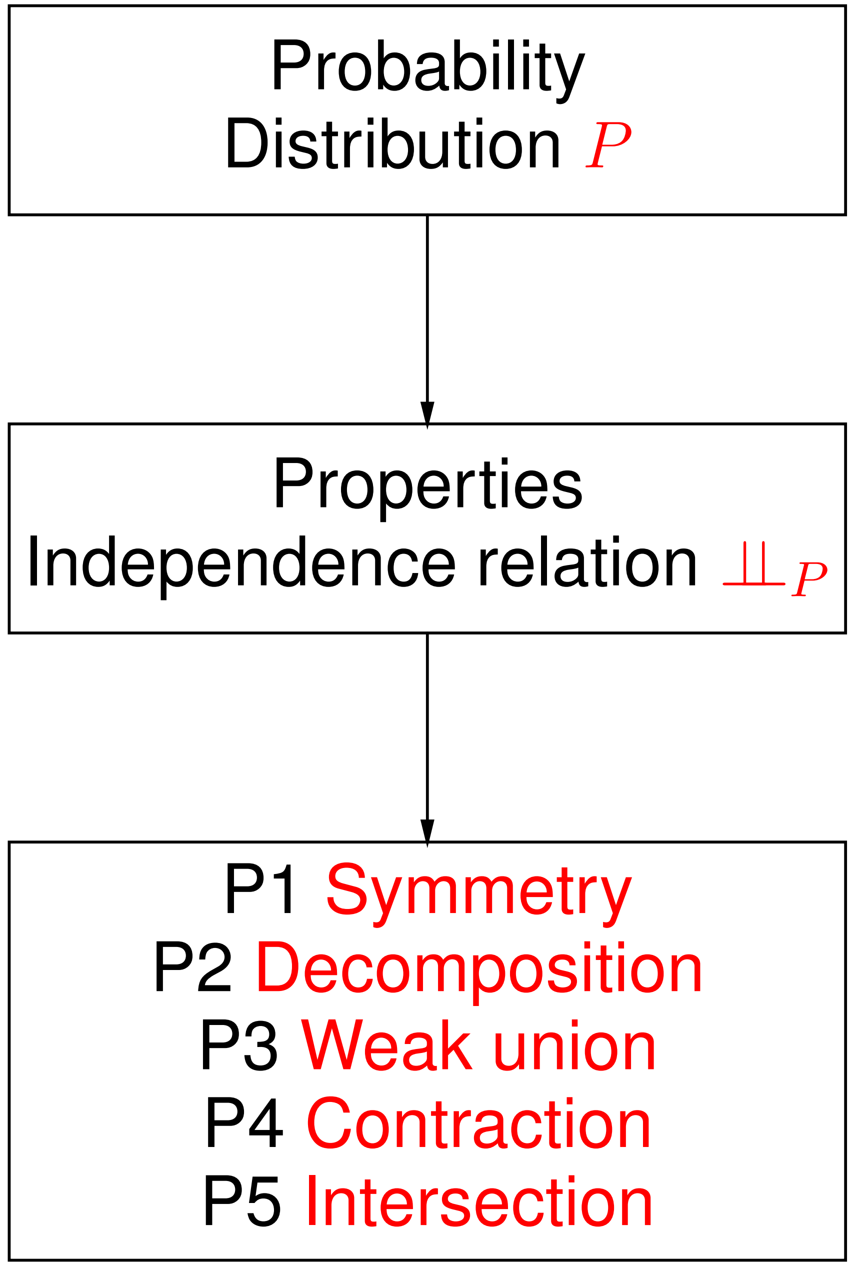
$$X \perp\!\!\!\perp_P Y \mid Z \cup W \quad \wedge \quad X \perp\!\!\!\perp_P W \mid Z \cup Y \quad \Rightarrow \quad X \perp\!\!\!\perp_P Y \cup W \mid Z$$

Proof: $\cdots$ *DIY*

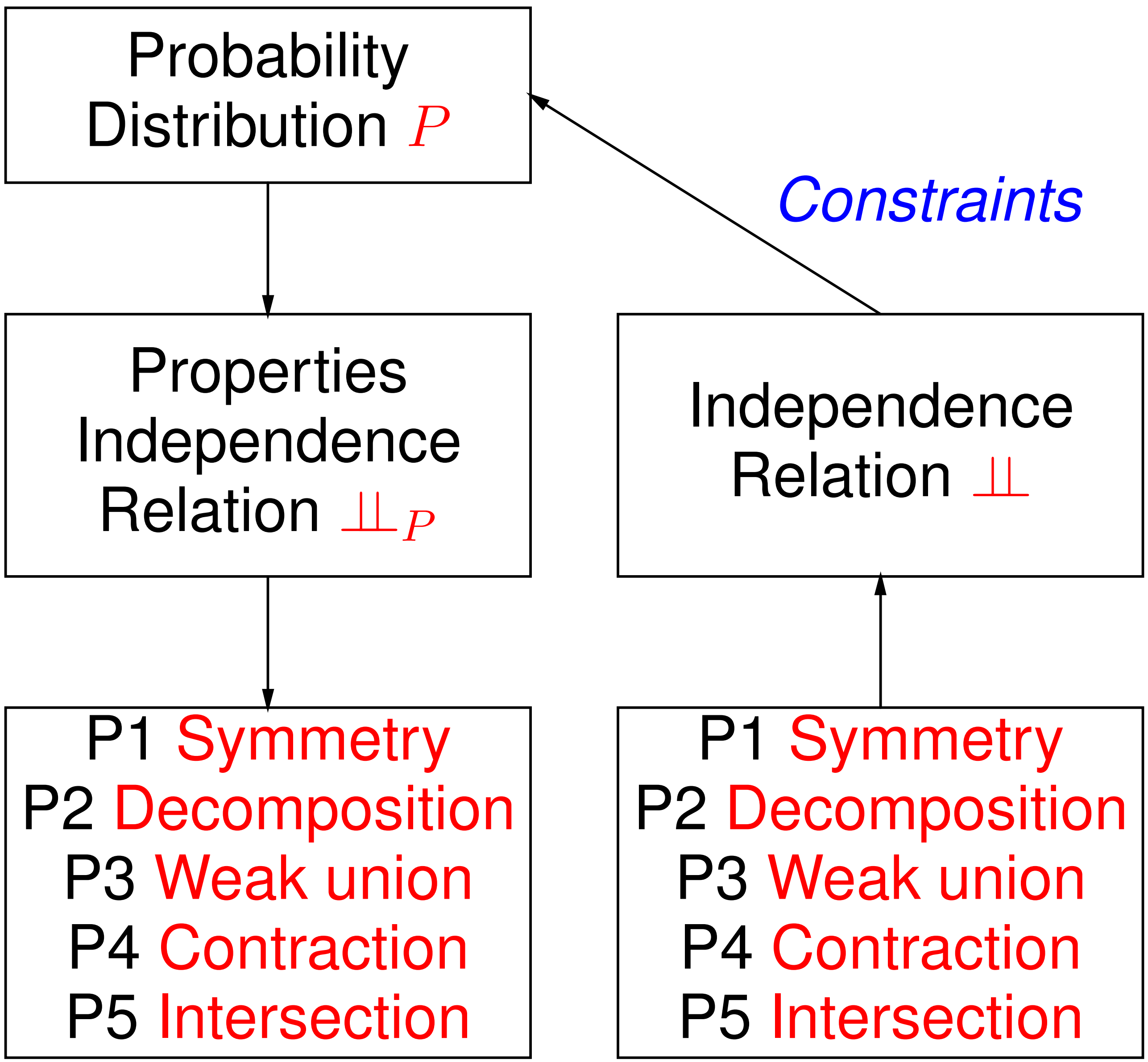
Note: This axiom only holds for *strictly positive* probability distributions, i.e. probability distributions that do **not** represent *logical relationships*.

- Semi-graphoid: Any model that satisfies axioms **P1–P4**
- Graphoid: Any model that satisfies axioms **P1–P5**

From probabilities to independence relation



Definition of an independence relation



Definition of an independence relation

Let $X, Y, Z, W \subseteq V$ be sets of objects. The independence relation $\perp\!\!\!\perp \subseteq \wp(V) \times \wp(V) \times \wp(V)$ is defined such that the following properties hold:

- Symmetry: $X \perp\!\!\!\perp Y \mid Z \Longleftrightarrow Y \perp\!\!\!\perp X \mid Z$
- Decomposition:
  $X \perp\!\!\!\perp Y \cup W \mid Z \;\; \Rightarrow \;\; X \perp\!\!\!\perp Y \mid Z \quad \wedge \quad X \perp\!\!\!\perp W \mid Z$
- Weak union: $X \perp\!\!\!\perp Y \cup W \mid Z \;\; \Rightarrow \;\; X \perp\!\!\!\perp Y \mid Z \cup W$
- Contraction:
  $X \perp\!\!\!\perp Y \mid Z \;\; \wedge \;\; X \perp\!\!\!\perp W \mid Y \cup Z \;\; \Rightarrow \;\; X \perp\!\!\!\perp W \cup Y \mid Z$

i.e. $\perp\!\!\!\perp$ defines a semi-graphoid. Note that the intersection property need not hold

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 $\perp\!\!\!\perp$:

$\{1\} \perp\!\!\!\perp \{4\} \mid \varnothing$

$\{4\} \perp\!\!\!\perp \{3\} \mid \{1\}$

$\{4\} \perp\!\!\!\perp \{1\} \mid \varnothing$

$\{3\} \perp\!\!\!\perp \{4\} \mid \{2\}$

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

$\{4\} \perp\!\!\!\perp \{3\} \mid \{2\}$

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

$\{2\} \perp\!\!\!\perp \{4\} \mid \{3\}$

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

$\{2\} \perp\!\!\!\perp \{1\} \mid \{4\}$

$\{2\} \perp\!\!\!\perp \{1\} \mid \varnothing$

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

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

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

$\{3\} \perp\!\!\!\perp \{4\} \mid \{1\}$

$\{4\} \perp\!\!\!\perp \{2\} \mid \{3\}$

$\{4\} \perp\!\!\!\perp \{2\} \mid \{1\}$

$\{3\} \perp\!\!\!\perp \{4\} \mid \varnothing$

$\{1\} \perp\!\!\!\perp \{4\} \mid \{2\}$

$\{4\} \perp\!\!\!\perp \{3\} \mid \varnothing$

$\{4\} \perp\!\!\!\perp \{1\} \mid \{2\}$

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

$\{1\} \perp\!\!\!\perp \{4\} \mid \{3\}$

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

$\{1\} \perp\!\!\!\perp \{2\} \mid \{4\}$

$\{1\} \perp\!\!\!\perp \{2\} \mid \varnothing$

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

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

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

$\{2\} \perp\!\!\!\perp \{4\} \mid \{1\}$

$\{4\} \perp\!\!\!\perp \{1\} \mid \{3\}$

$\{2\} \perp\!\!\!\perp \{4\} \mid \varnothing$

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

$\{4\} \perp\!\!\!\perp \{2\} \mid \varnothing$

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

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

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

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

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

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

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

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

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

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

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

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

Use of independence axioms

Lemma Let $X, Y, Z, W \subseteq V$ be sets of random variables:

$$X \perp\!\!\!\perp Y \mid Z \quad \wedge \quad X \cup Z \perp\!\!\!\perp W \mid Y \quad \Rightarrow \quad X \perp\!\!\!\perp W \mid Z$$

Proof: It holds that

$$X \cup Z \perp\!\!\!\perp W \mid Y \Rightarrow_{\text{symm}} W \perp\!\!\!\perp X \cup Z \mid Y$$
$$\Rightarrow_{\text{wu}} W \perp\!\!\!\perp X \mid Y \cup Z \Rightarrow_{\text{symm}} X \perp\!\!\!\perp W \mid Y \cup Z$$

From $X \perp\!\!\!\perp Y \mid Z$ and $X \perp\!\!\!\perp W \mid Y \cup Z$, using contraction, it follows that $X \perp\!\!\!\perp W \cup Y \mid Z$. Now, by using decomposition, it follows that $X \perp\!\!\!\perp W \mid Z$

Use of a joint probability distribution

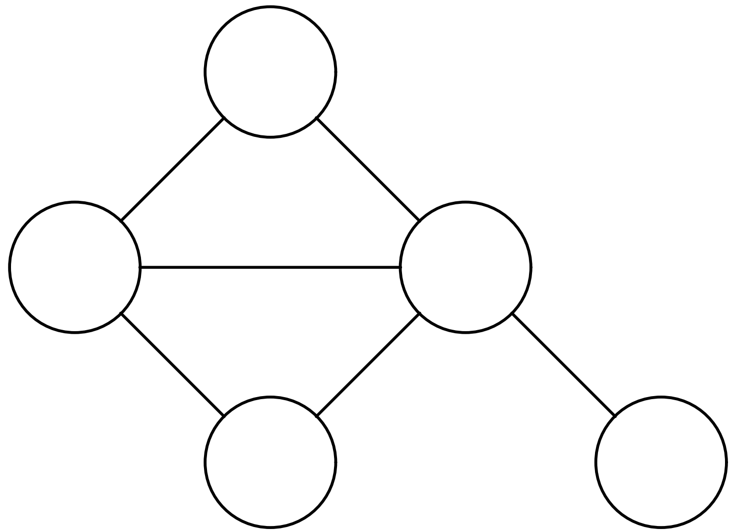
Let X, Y and Z be binary variables with the following joint distribution:

$$
\begin{aligned}
P(x, y, z) &= 0.00675 & P(\neg x, y, z) &= 0.01575 \\
P(x, y, \neg z) &= 0.002565 & P(\neg x, y, \neg z) &= 0.253935 \\
P(x, \neg y, z) &= 0.00825 & P(\neg x, \neg y, z) &= 0.01925 \\
P(x, \neg y, \neg z) &= 0.006935 & P(\neg x, \neg y, \neg z) &= 0.686565
\end{aligned}
$$

Check whether any of the following independence relations hold:

$$
\begin{aligned}
X \perp\!\!\!\perp Y \mid \varnothing &\Leftrightarrow P(X \mid Y) = P(X) & X \perp\!\!\!\perp Y \mid Z &\Leftrightarrow P(X \mid Y, Z) = P(X \mid Z) \\
X \perp\!\!\!\perp Z \mid \varnothing &\Leftrightarrow P(X \mid Z) = P(X) & X \perp\!\!\!\perp Z \mid Y &\Leftrightarrow P(X \mid Z, Y) = P(X \mid Y) \\
Y \perp\!\!\!\perp Z \mid \varnothing &\Leftrightarrow P(Y \mid Z) = P(Y) & Y \perp\!\!\!\perp Z \mid X &\Leftrightarrow P(Y \mid Z, X) = P(Y \mid X) \\
Y \perp\!\!\!\perp X \mid \varnothing &\Leftrightarrow P(Y \mid X) = P(Y) & Y \perp\!\!\!\perp X \mid Z &\Leftrightarrow P(Y \mid X, Z) = P(Y \mid Z) \\
Z \perp\!\!\!\perp X \mid \varnothing &\Leftrightarrow P(Z \mid X) = P(Z) & Z \perp\!\!\!\perp X \mid Y &\Leftrightarrow P(Z \mid X, Y) = P(Z \mid Y) \\
Z \perp\!\!\!\perp Y \mid \varnothing &\Leftrightarrow P(Z \mid Y) = P(Z) & Z \perp\!\!\!\perp Y \mid X &\Leftrightarrow P(Z \mid Y, X) = P(Z \mid X)
\end{aligned}
$$

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

Global Markov property – separation

Let $G = (V(G), E(G))$ be an undirected graph, and let $U, Z, W \subseteq V(G)$ be *sets of vertices* in G. The set W (u-)separates U and Z, denoted as

$$U \perp\!\!\!\perp_G Z \mid W$$

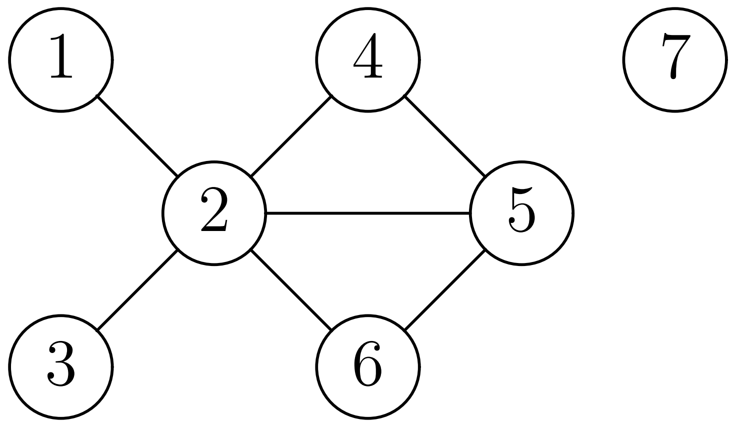
if every path from a vertex in U to a vertex in Z contains at least one vertex in W; otherwise these sets are (u-)connected

Remarks:

- This criterion is known as the global Markov property or (u-)separation criterion for undirected graphs
- Note that $\perp\!\!\!\perp_G$ indicates that the independence relation is defined in terms of G (cf. $\perp\!\!\!\perp_P$)
- If there are no paths between two vertices u and v, then $\{u\} \perp\!\!\!\perp_G \{v\} \mid \varnothing$

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

Let V be a set and let $\perp\!\!\!\perp$ be an independence relation defined on V. Let $G = (V(G), E(G))$ be an undirected graph with $V(G) = V$, then for each $X, Y, Z \subseteq V$:

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

$$X \perp\!\!\!\perp Y \mid Z \Rightarrow X \perp\!\!\!\perp_G Y \mid Z$$

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

$$X \perp\!\!\!\perp_G Y \mid Z \Rightarrow X \perp\!\!\!\perp Y \mid Z$$

- 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

$$X \perp\!\!\!\perp Y \mid Z \Longleftrightarrow X \perp\!\!\!\perp_G Y \mid Z$$

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

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

- 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

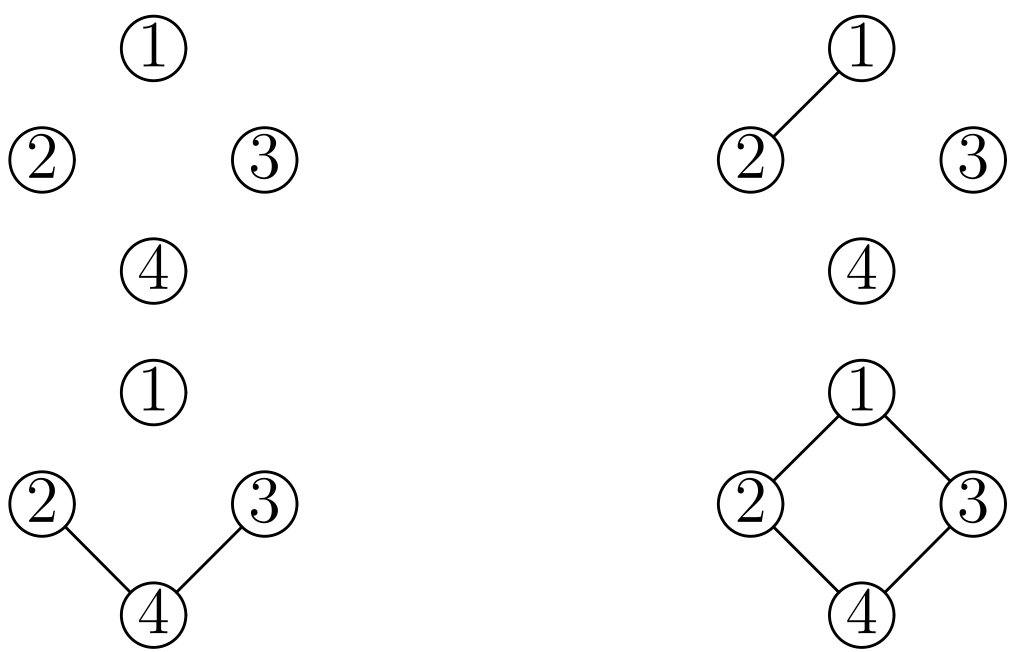
$$X_U \perp\!\!\!\perp_P X_W \mid X_Z \Longleftrightarrow U \perp\!\!\!\perp_G W \mid Z$$

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:



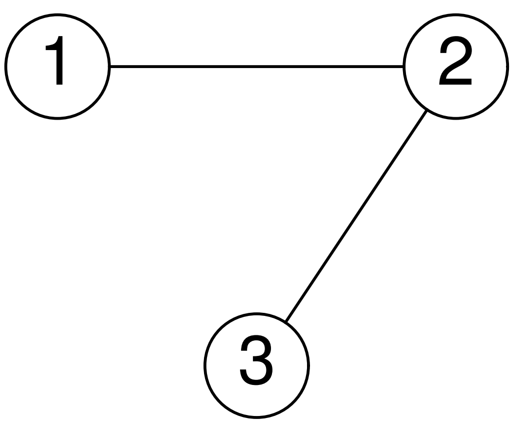
Markov network

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:

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

D-maps and I-maps again

Let $\perp\!\!\!\perp$ be an independence relation. D-maps and I-maps are limited in expressiveness in the following sense:

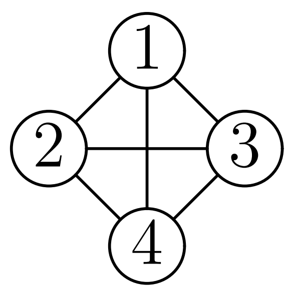
- A pair of neighbour vertices in a D-map for $\perp\!\!\!\perp$ are dependent. However, not all dependent variables are neighbours
- A pair of non-neighbour variables in an I-map for $\perp\!\!\!\perp$ corresponds to independent variables, but not each pair of independent variables in an I-map are non-neighbours

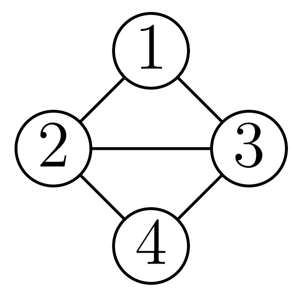
Examples of I-maps

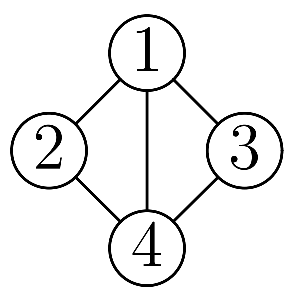
Let $V = \{1, 2, 3, 4\}$ be a set with random variables X_V, and consider the independence relation $\perp\!\!\!\perp_P$:

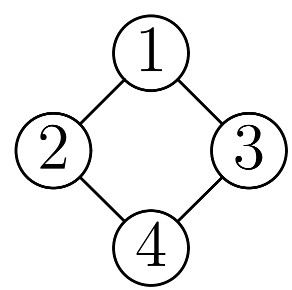
$$\{X_1\} \perp\!\!\!\perp_P \{X_4\} \mid \{X_2, X_3\}$$
$$\{X_2\} \perp\!\!\!\perp_P \{X_3\} \mid \{X_1, X_4\}$$

The following undirected graphs are examples of I-maps:









(So, what is the P-map?)

Obvious properties

Lemma For each independence relation $\perp\!\!\!\perp$ there exists an undirected D-map.

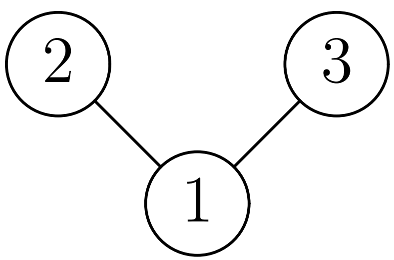
Proof:
The undirected graph $G = (V, \varnothing)$ is a D-map for $\perp\!\!\!\perp$

Lemma For each independence relation $\perp\!\!\!\perp$ there exists an undirected I-map.

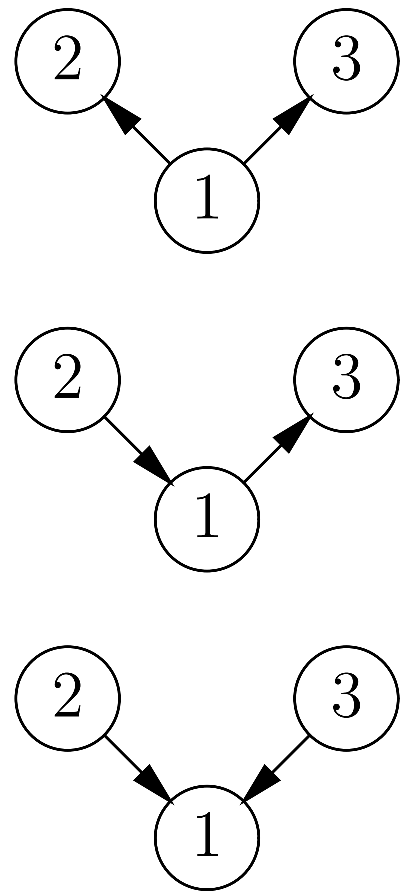
Proof:
The undirected graph $G = (V, V \times V)$ is an I-map for $\perp\!\!\!\perp$

Expressiveness: directed vs undirected

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



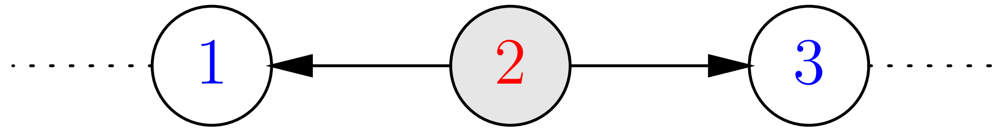
vs



d-Separation: *3* situations

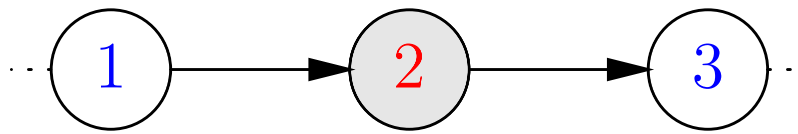
A chain k (= path in undirected underlying graph) in an acyclic directed graph $G = (V(G), A(G))$ can be blocked:

Diverging



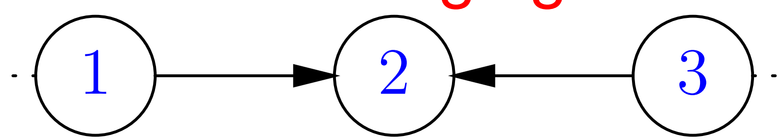
2 blocks (d-separates) 1 and 3: $\{1\} \perp\!\!\!\perp \{3\} \mid \{2\}$

Serial



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

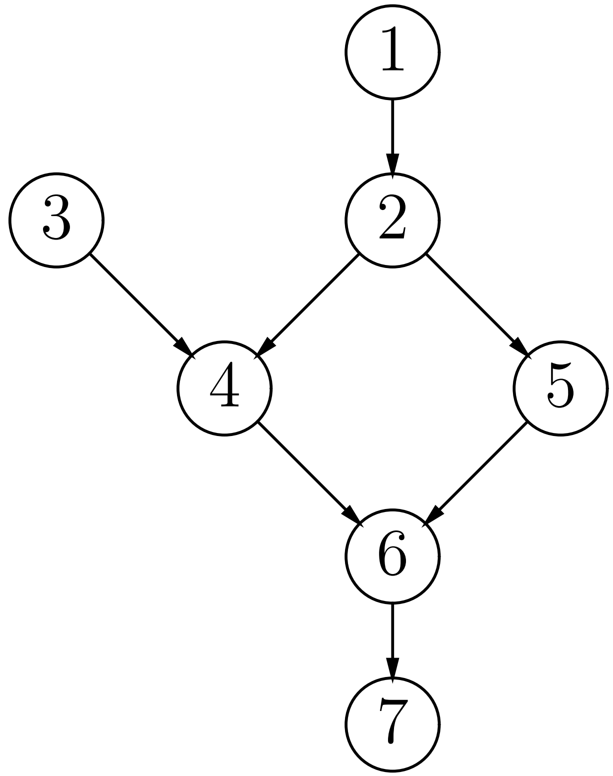
Converging



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 $\{6\}$

Directed global Markov property

Let $G = (V(G), A(G))$ be an acyclic directed graph, and let $U, W, Z \subseteq V(G)$ be sets of vertices in G. The set Z d-separates U and W, denoted as

$$U \perp\!\!\!\perp_G^d W \mid Z$$

if every chain from a vertex in U to a vertex in W is blocked by Z

Remarks

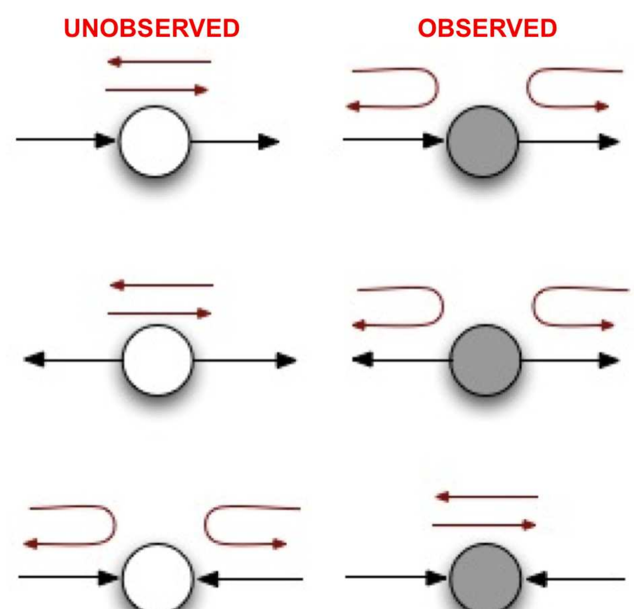
- This criterion is known as the global Markov property or d-separation criterion for acyclic directed graphs
- Note that $\perp\!\!\!\perp_G^d$ indicates that the independence relation is defined in terms of G (cf. $\perp\!\!\!\perp_P$)

Bayes-ball algorithm

Basic idea:

- simulate the transfer of probabilistic information by a bouncing ball
- if the ball is not allowed to pass through a vertex C from a vertex A to another vertex B, then these are conditionally independent given C

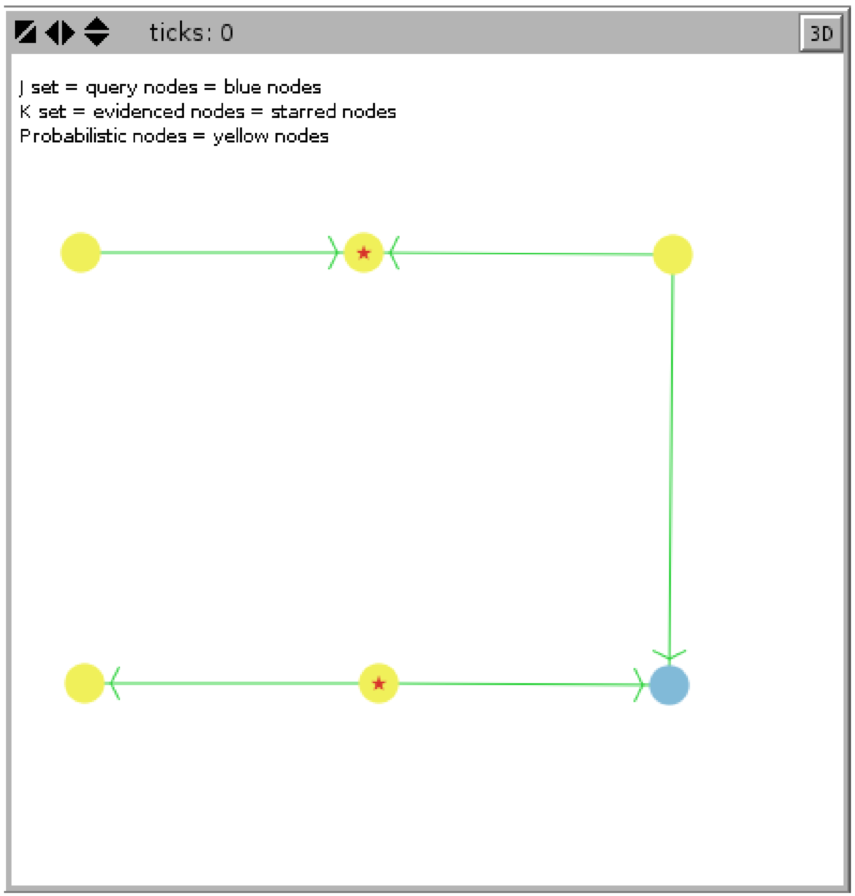
Principal operations:



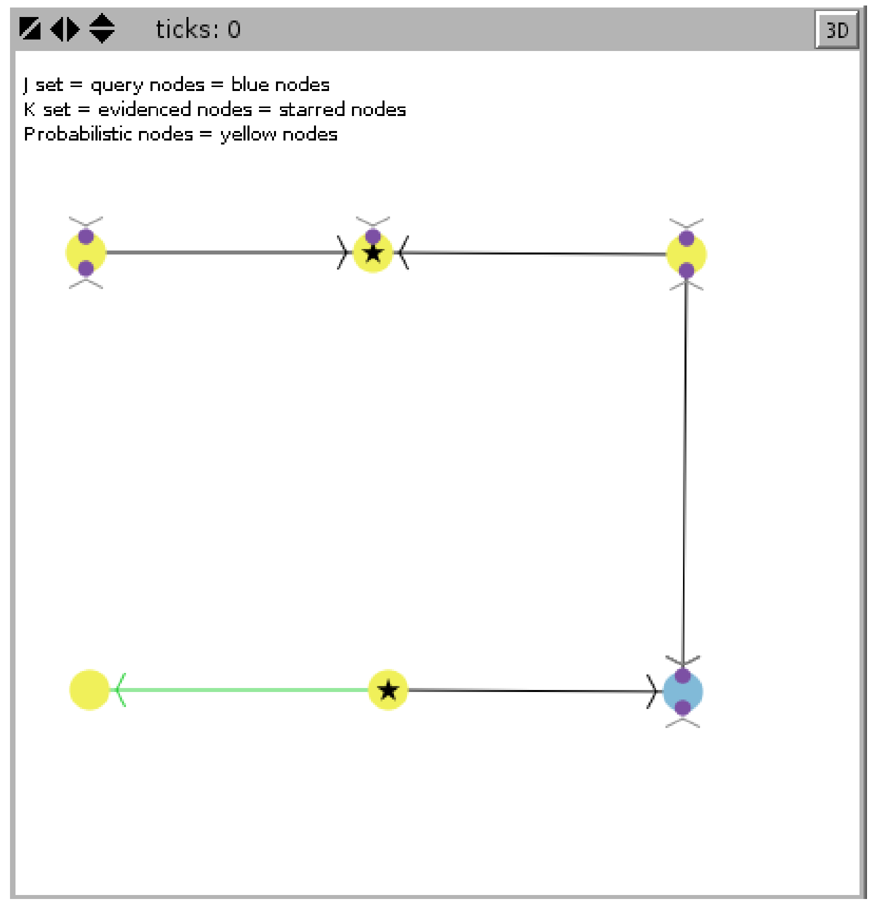
- an unobserved vertex passes balls through but also bounces balls back from children
- an observed vertex bounces balls back from parents, but blocks balls from children

Example

bayesball.nlogo (based on R.D. Shachter, “Bayes-Ball: The rational pastime for determining irrelevance and requisite information in belief networks and influence diagrams”)



Start



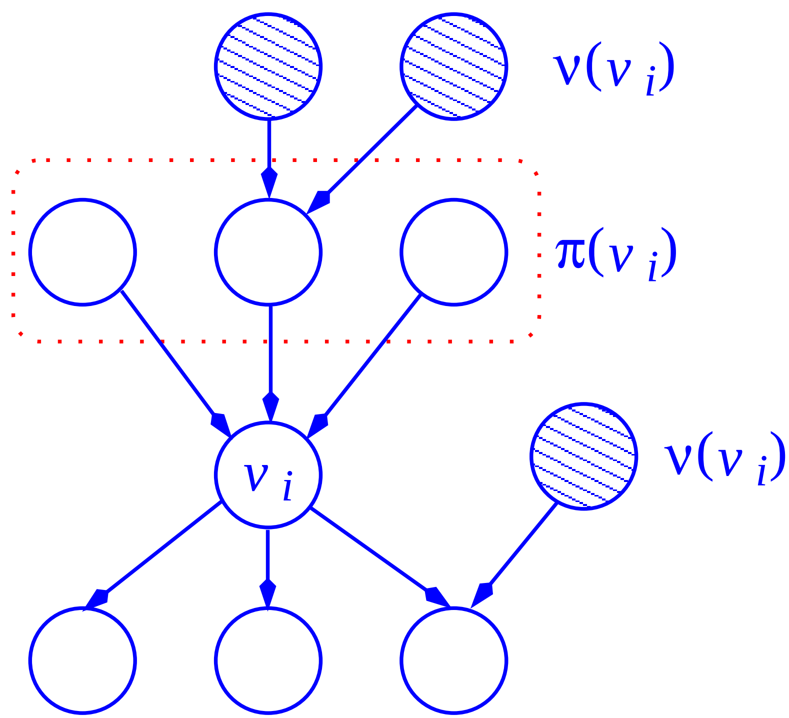
End

There is also a local Markov property

Let $G = (V(G), A(G))$ be an acyclic, directed graph, then the following local Markov property holds:

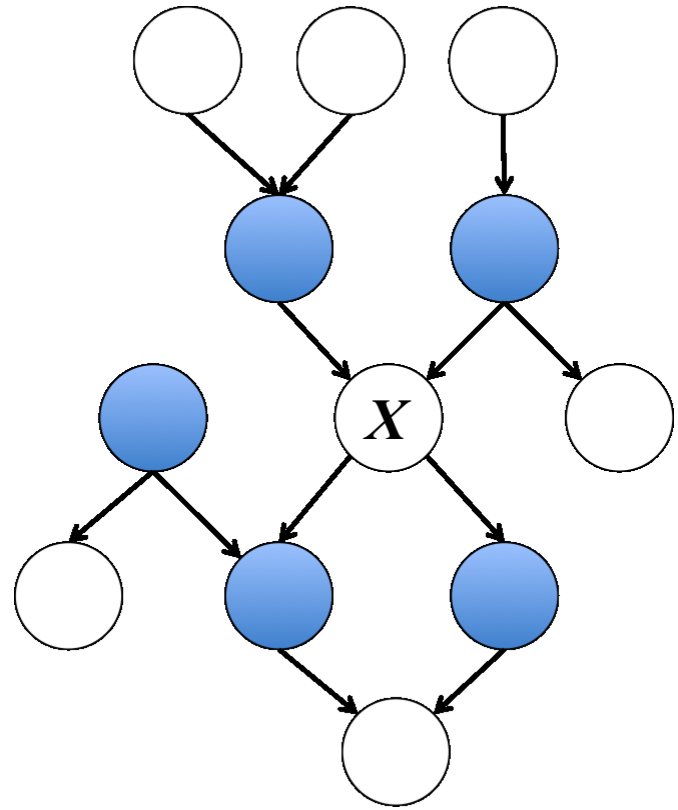
$$\{v_i\} \perp\!\!\!\perp_G^d \nu(v_i) \mid \pi(v_i)$$

with $\nu(v_i)$ *non-descendants* of vertex v_i, and $\pi(v_i)$ set of parents



Markov blanket

- Set of parents, children and co-parents of a node (for X these are the nodes in blue)
- The conditional distribution of X conditioned on all the other variables in the graph is dependent only on the variables in the Markov blanket



Directed D-map and I-map

Let V be a set and let $\perp\!\!\!\perp$ be an independence relation defined on V. Let $G = (V(G), A(G))$ be an acyclic directed graph, then for each $X, Y, Z \subseteq V$:

- G is called a directed dependence map, D-map for short, if

$$X \perp\!\!\!\perp Y \mid Z \Rightarrow X \perp\!\!\!\perp_G^d Y \mid Z$$

- G is called a directed independence map, I-map for short, if

$$X \perp\!\!\!\perp_G^d Y \mid Z \Rightarrow X \perp\!\!\!\perp Y \mid Z$$

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

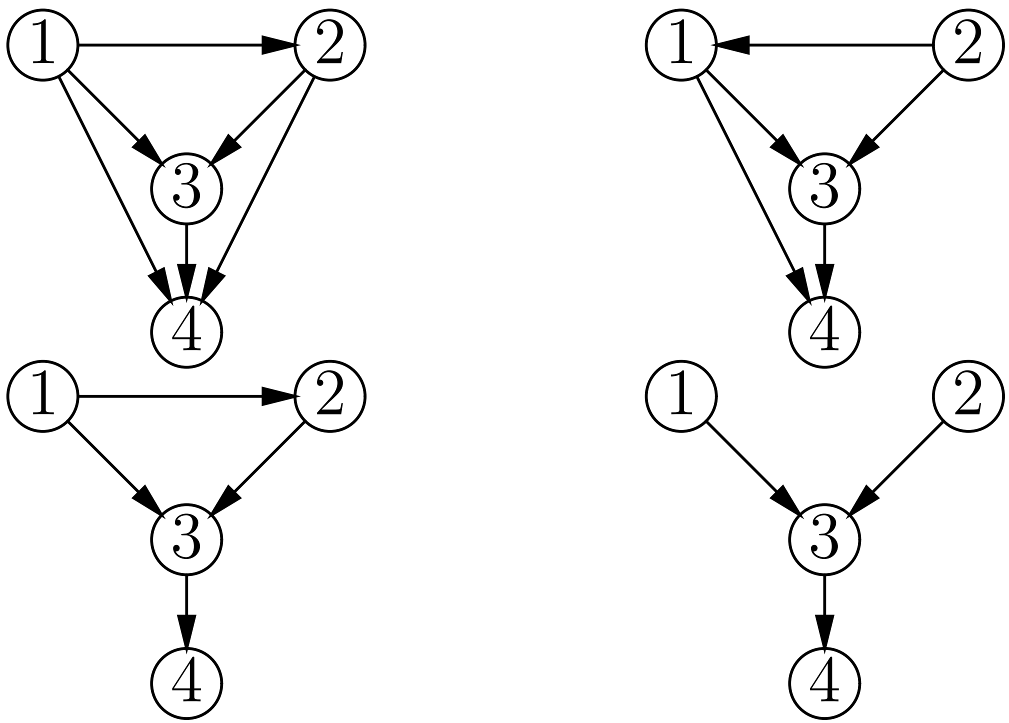
$$X \perp\!\!\!\perp Y \mid Z \Longleftrightarrow X \perp\!\!\!\perp_G^d Y \mid Z$$

Examples directed I-maps

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

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

and the following directed I-maps of P:



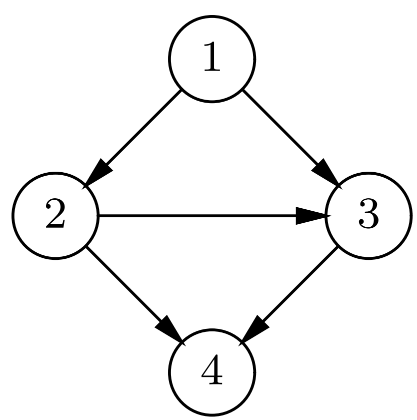
Minimal directed I-map

In the context of Bayesian networks, we are interested in I-maps that contain as few arcs as possible (makes probability tables smaller), i.e. minimal directed I-maps

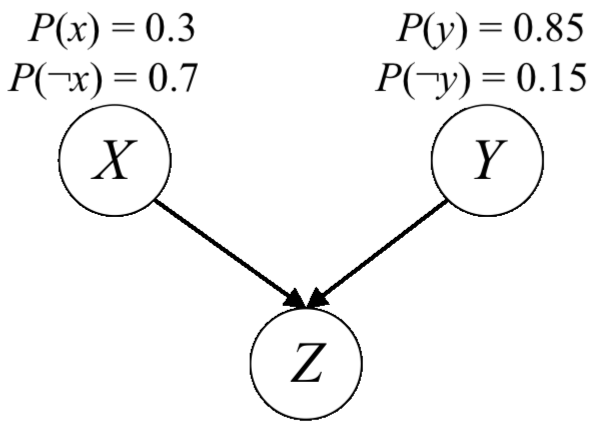
Let $G = V(G), A(G))$ be an acyclic directed graph and let $P(X_{V(G)})$ be a probability distribution of $X_{V(G)}$. G is said to be a minimal directed I-maps of P, if

- G is a directed I-map of P, and
- none of the subgraphs of G is a directed I-map of P

Example:



Example minimal directed I-map



X	Y	$P(z \mid X,Y)$	$P(\neg z \mid X,Y)$
x	y	0.60	0.40
x	$\neg y$	0.05	0.95
$\neg x$	y	0.22	0.78
$\neg x$	$\neg y$	0.35	0.65

So, $P(X,Y,Z) = P(Z \mid X,Y)P(X)P(Y)$:

$P(x,y,z) = 0.6 \cdot 0.3 \cdot 0.85 = 0.153$

$P(x,y,\neg z) = 0.102$

$P(\neg x,y,z) = 0.1309$

$\ldots$

Verify:

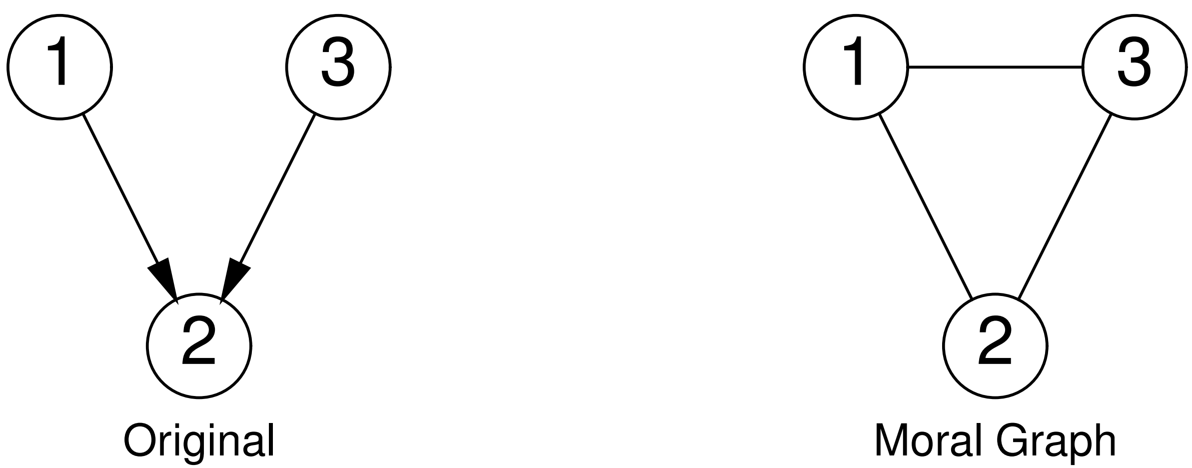
1) $X \perp\!\!\!\perp^d_G Y \mid \varnothing \Rightarrow P(X \mid Y) = P(X)$

2) $X \not\perp\!\!\!\perp^d_G Y \mid Z \Rightarrow P(X \mid Y,Z) \neq P(X \mid Z)$

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

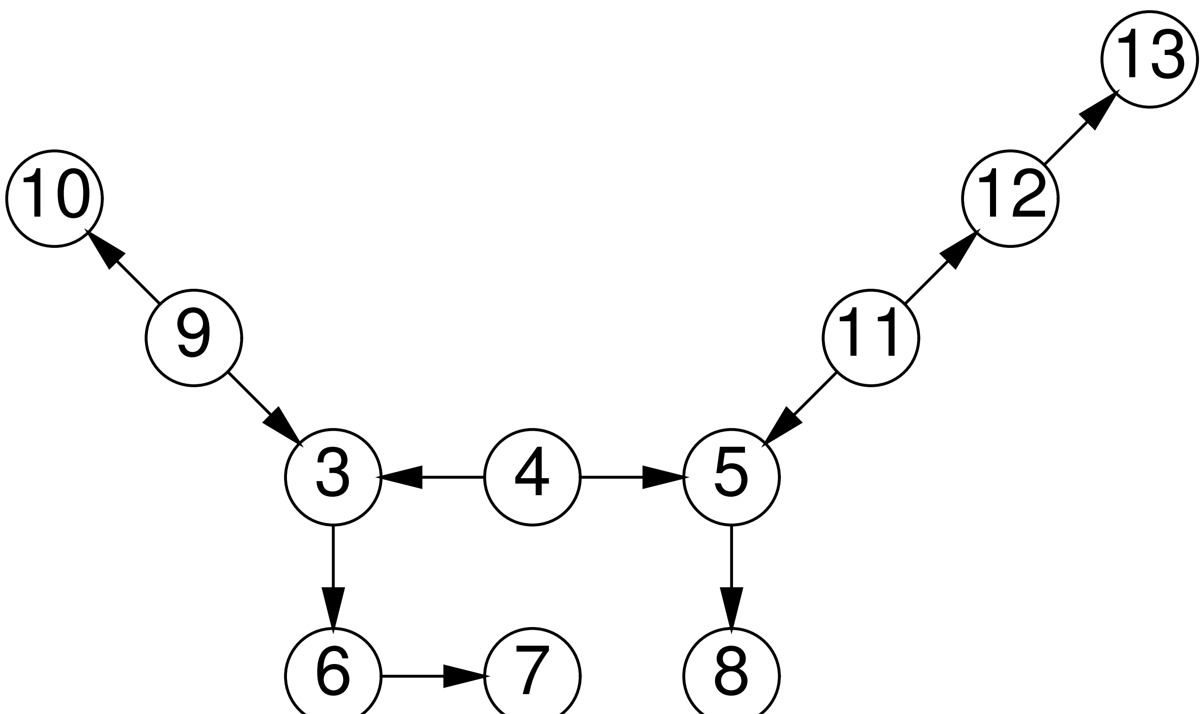
Example:



Moralisation

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

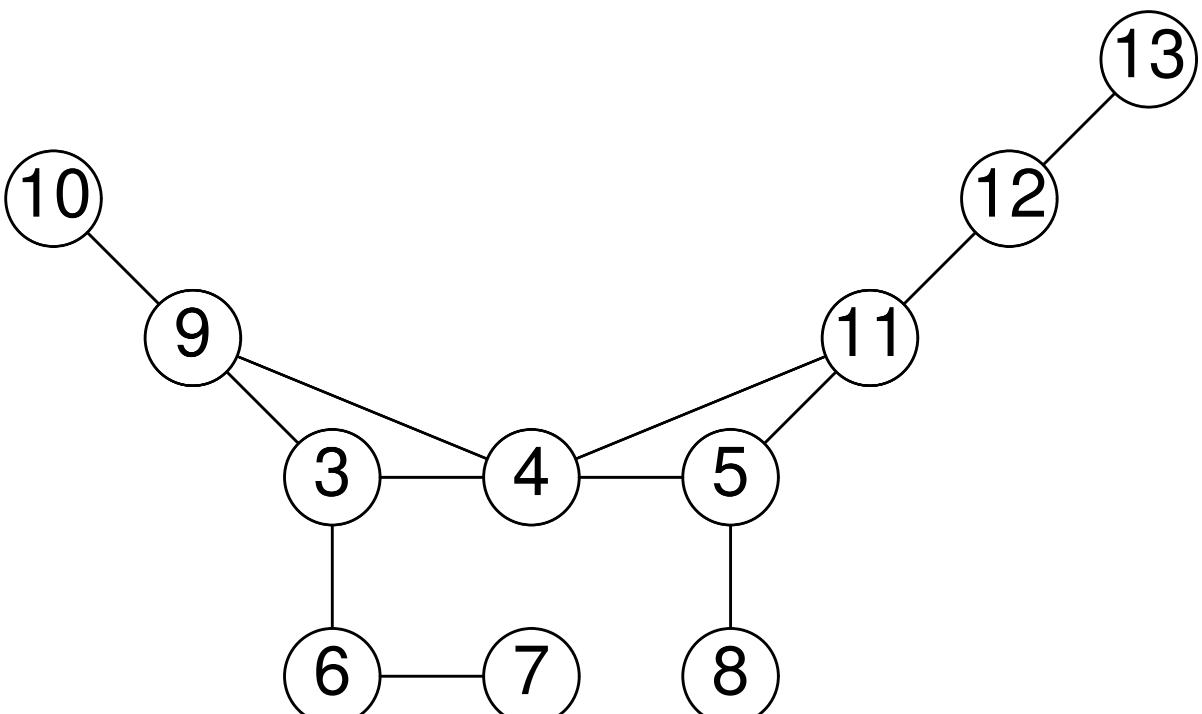
1. add lines to all non-connected vertices, which have a common child, or descendant of a common child, and
2. replace each arc with a line in the resulting graph



Moralisation

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

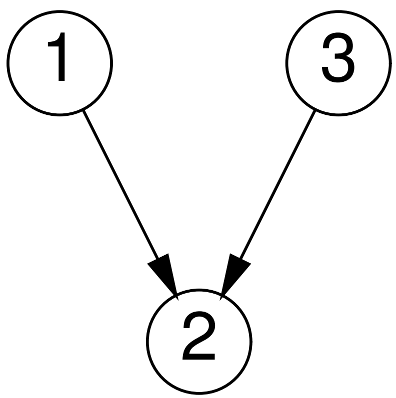
1. add lines to all non-connected vertices, which have a common child, or descendant of a common child, and
2. replace each arc with a line in the resulting graph



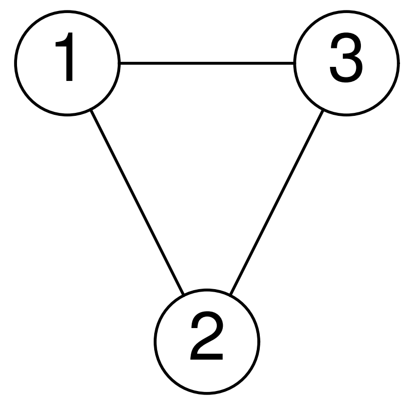
Comments

- 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

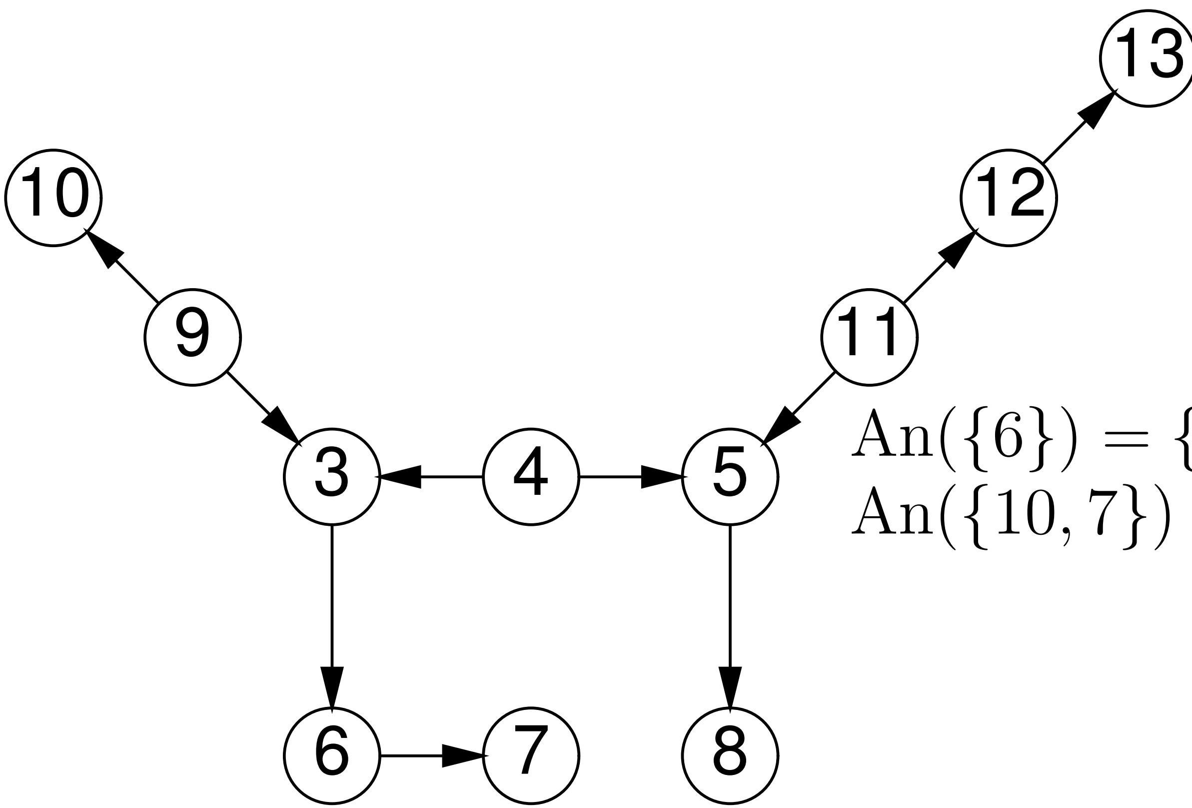


Moral Graph

- 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

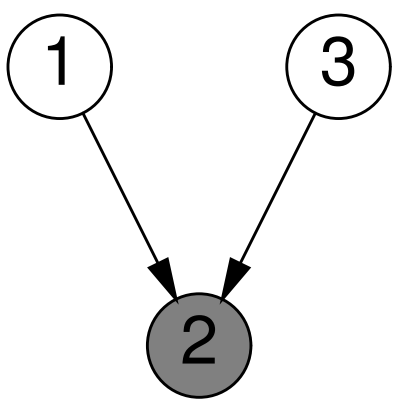
Let P be a joint probability distribution of a Bayesian network $\mathcal{B} = (G, P)$, then

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

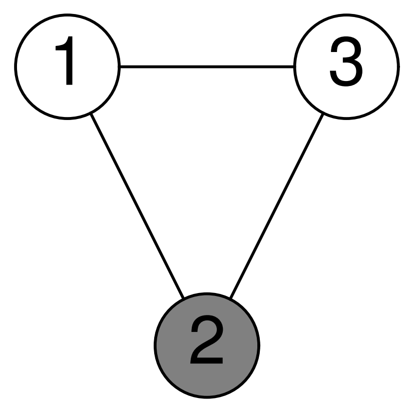
holds iff U and V are (u-)separated by W in the moral induced subgraph G^m of G with vertices $\mathrm{An}(U \cup V \cup W)$

Example:

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



Original



Moral Graph

‘Dynamic’ moralisation

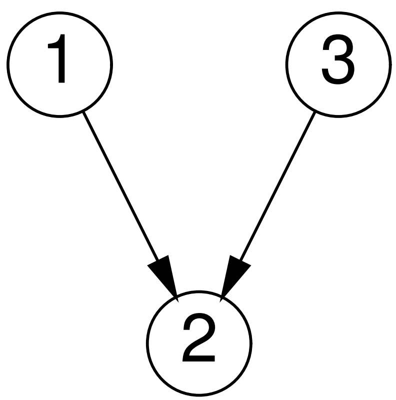
Let P be a joint probability distribution of a Bayesian network $\mathcal{B} = (G, P)$, then

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

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

Moralisation and d-separation

Let $G = (V(G), A(G))$ be an acyclic directed graph and let $U, W, S \subseteq V(G)$ be disjoint sets of vertices. Then, U and W are d-separated by S, i.e.

$$U \perp\!\!\!\perp_G^d W \mid S$$

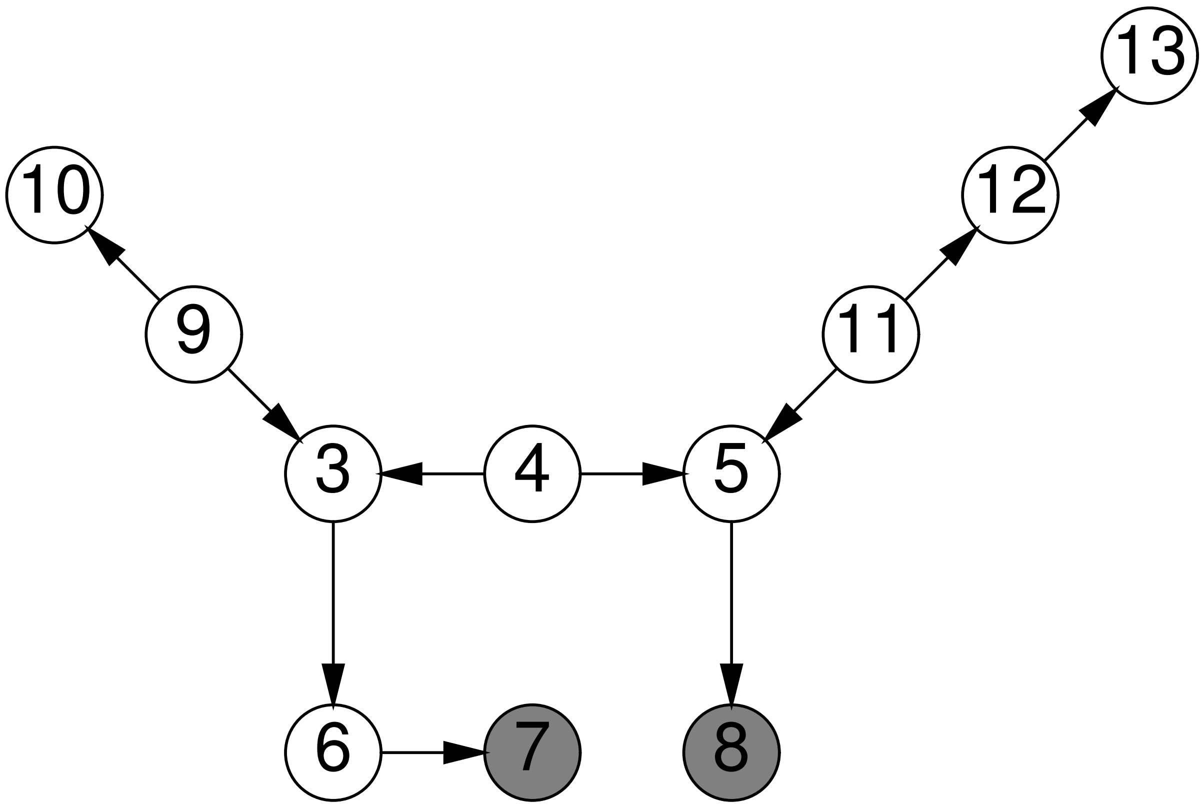
iff U and W are separated in the moral graph of the set of vertices $\mathrm{An}(U \cup W \cup S)$, i.e.

$$U \perp\!\!\!\perp_{G^m_{\mathrm{An}(U \cup W \cup S)}} W \mid S$$

Proof: Cowell et al, “Probabilistic Networks and Expert Systems”, 1999, Springer, New York, page 72

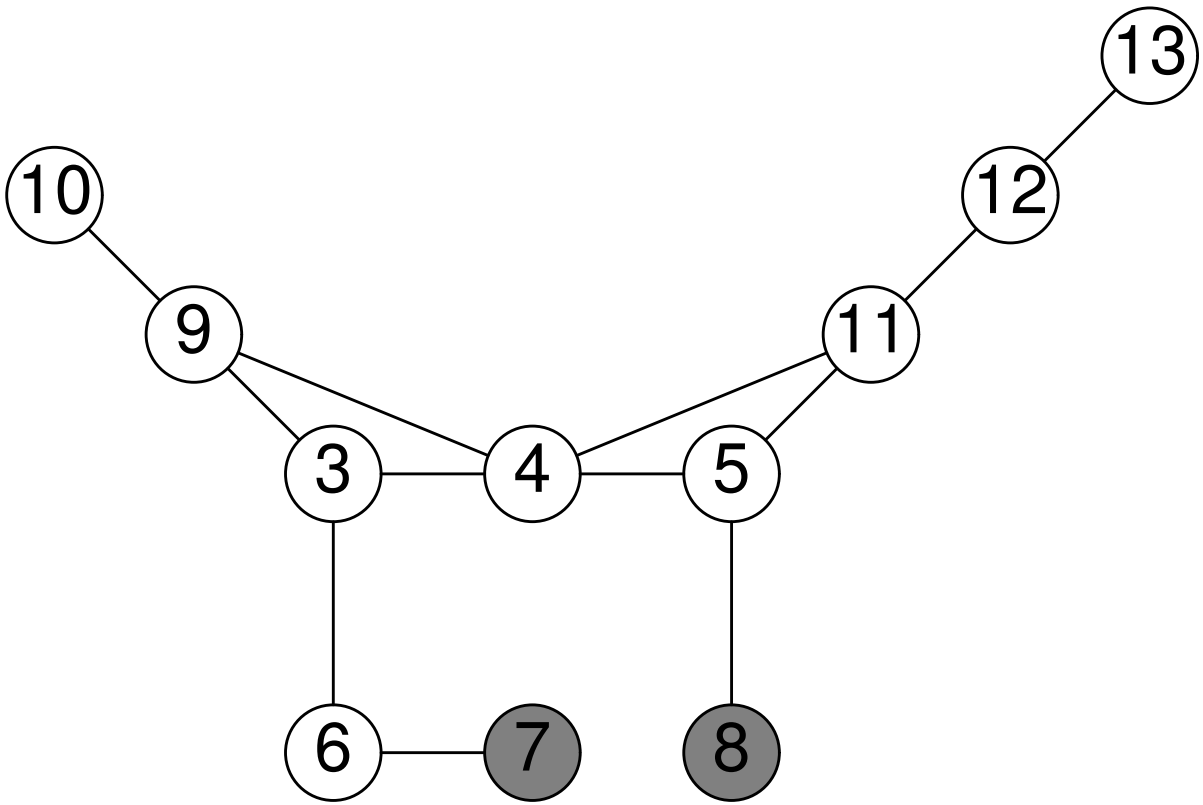
Example (1)

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



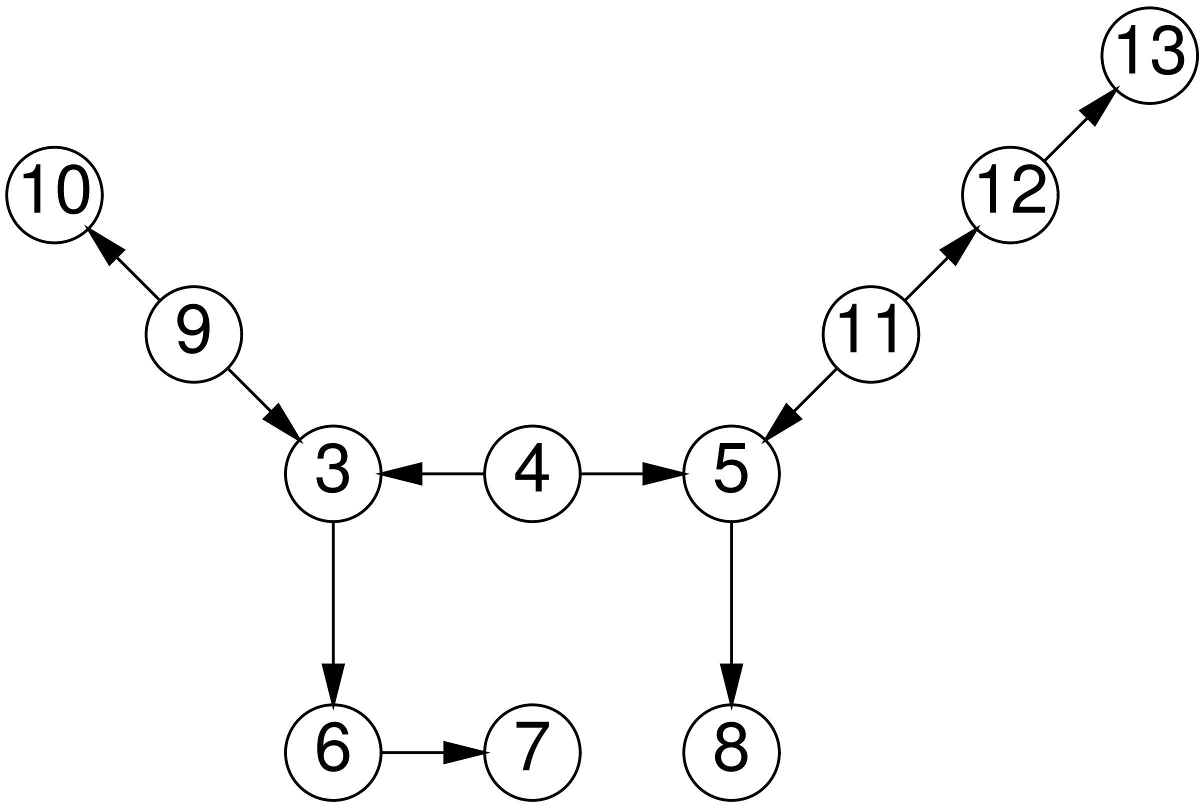
Example (1)

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



Example (2)

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



Example (2)

$\{10\} \perp\!\!\!\perp_{G^m_{\mathrm{An}(\{10,13\})}} \{13\} \mid \varnothing$

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