Bayesian Networks

IV. Probabilistic Independence and Separation in Graphs (part 4)

Prof. Dr. Lars Schmidt-Thieme, L. B. Marinho, K. Buza

Information Systems and Machine Learning Lab (ISMLL)
Institute of Economics and Information Systems
& Institute of Computer Science
University of Hildesheim
http://www.ismll.uni-hildesheim.de

1. Basic Probability Calculus

2. Separation in undirected graphs

3. Separation in directed graphs

4. Markov networks

5. Bayesian networks

DAG-representations

Lemma 1 (criterion for DAG-representation)**.** *Let* p *be a joint probability distribution of the variables* V *and* G *be a graph on the vertices* V*. Then:*

G *represents* $p \Leftrightarrow v$ *and* $\mathrm{nondesc}(v)$ *are conditionally independent given* $\mathrm{pa}(v)$ *for all* $v \in V$*, i.e.,*

$$I_p(\{v\}, \mathrm{nondesc}(v) | \mathrm{pa}(v)), \quad \forall v \in V$$

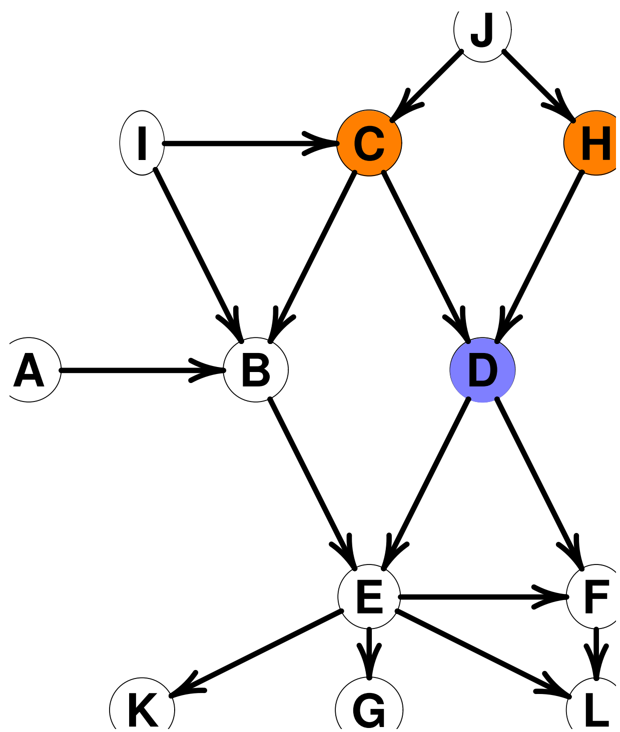


Figure 1: Parents of a vertex (orange).

Faithful DAG-representations

Lemma 2 (necessary conditions for faithful DAG-representability). *An independency model I has a faithful DAG representation, only if it is*

(i) symmetric: $I(X,Y|Z) \Rightarrow I(Y,X|Z)$.

(ii) (de)composable: $I(X,Y \cup W|Z) \Leftrightarrow I(X,Y|Z)$ *and* $I(X,W|Z)$

(iii) contractable: $I(X,Y|Z \cup W)$ *and* $I(X,W|Z) \Rightarrow I(X,Y \cup W|Z)$

(iv) intersectable: $I(X,W|Z \cup Y)$ *and* $I(X,Y|W \cup Z) \Rightarrow I(X,Y \cup W|Z)$

(v) weakly unionable: $I(X,Y \cup Z|W) \Rightarrow I(X,Y|W \cup Z)$

(vi) weakly transitive: $I(X,Y|Z)$ *and* $I(X,Y|Z \cup \{v\}) \Rightarrow I(X,\{v\}|Z)$ *or* $I(\{v\},Y|Z)$
$V \setminus Z$

(vii) chordal: $I(\{a\},\{c\}|\{b,d\})$ *and* $I(\{b\},\{d\}|\{a,c\}) \Rightarrow I(\{a\},\{c\}|\{b\})$ *or* $I(\{a\},\{c\}|\{$

It is still an open research problem, if there is a finite axiomatisation of faithful DAG-representability.

Example for a not faithfully DAG-representable dependency model

Example 1. The independency model

$$I := \{I(x, y|z), I(y, x|z), I(x, y|w), I(y, x|w)\}$$

satisfies all conditions of lemma 2, but it does not have a faithful DAG-representation. [CGH97, p. 239]

Exercise: compute all minimal DAG-representations of I using lemma 3 and check if they are faithful.

Probability distributions can violate composition, weak transitivity and chordality, and may thus have no faithful DAG-representation.

Minimal DAG-representations

Lemma 3 (construction and uniqueness of minimal DAG-representation, [VP90]). *Let I be an independence model on a set V that satisfies symmetry, decomposition, weak union, and contraction. Then:*

(i) A minimal DAG-representation can be constructed as follows: Choose an arbitrary ordering $\sigma := (v_1, \ldots, v_n)$ of V. Choose a maximal set $\pi_i \subseteq \{v_1, \ldots, v_{i-1}\}$ of σ-precursors of v_i with

$$I(v_i, \{v_1, \ldots, v_{i-1}\} \setminus \pi_i | \pi_i)$$

Then $G := (V, E)$ with

$$E := \{(w, v_i) \,|\, i = 1, \ldots, n, w \in \pi_i\}$$

is a minimal DAG-representation of p.

(ii) If I also satisfies intersection, then the minimal representation G is unique up to ordering σ.

The lemma provides an algorithm for the construction of a minimal DAG-representation for any joint probability distribution p. Furthermore it assures uniqueness up to ordering for non-extreme JPDs.

Minimal DAG-representations / example

$$I := \{(A, C|B), (C, A|B)\}$$

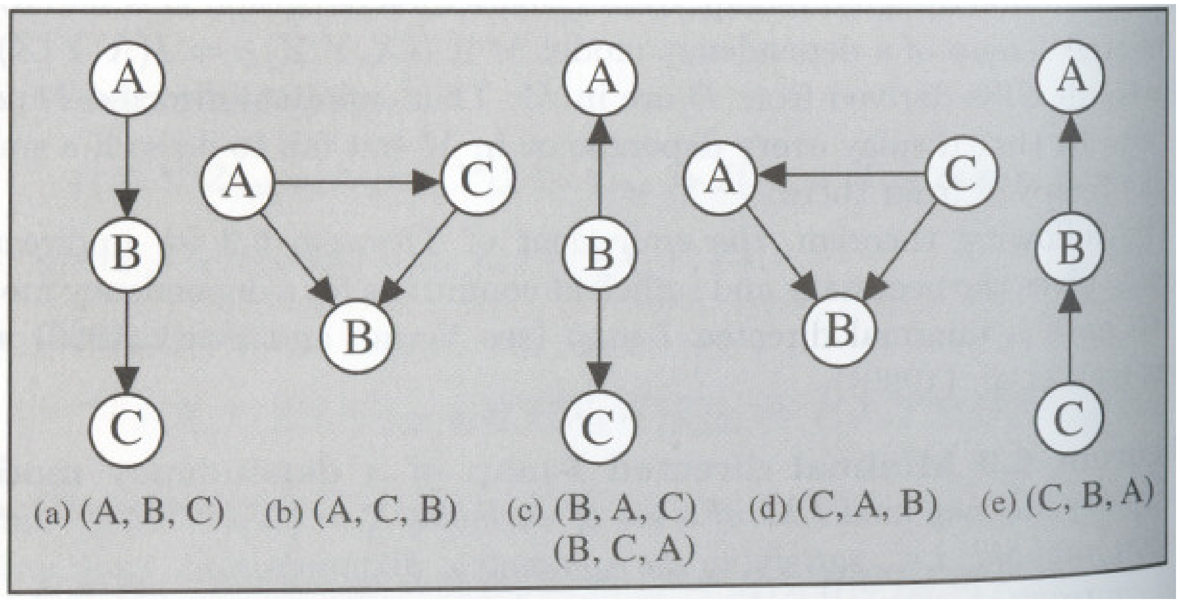


Figure 2: Minimal DAG-representations of I [CGH97, p. 240].

Minimal representations / conclusion

Representations always exist (e.g., trivial).

Minimal representations always exist
(e.g., start with trivial and drop edges successively).

	Markov network (undirected)		Bayesian network (directed)	
	minimal	faithful	minimal	faithful
general JPD	may not be unique	may not exist	may not be unique	may not exist
non-extreme JPD	unique	may not exist	unique up to ordering	may not exist

factorization of a JPD according to a DAG

Definition 1. Let p be a joint probability distribution of a set of variables V. Let G be a DAG.

p **factorizes according to** G, if

$$p = \prod_{v \in V} p(v | \operatorname{pa}(v))$$

Lemma 4. *Let p be a JPD and G a DAG on V. Then:*

(i) p factorizes acc. to $G \Leftrightarrow G$ represents p

(ii) If G is a DAG and p_v are conditional probability distributions on $\{v\} \cup \operatorname{pa}(v)$ conditioned on $\operatorname{pa}(v)$ for each of its vertices $v \in V$, then G represents the JPD

$$p = \prod_{v \in V} p_v$$

bayesian network

Definition 2. A pair $(G := (V, E), (p_v)_{v\in V})$ consisting of

(i) a directed graph G on a set of variables V and

(ii) a set of conditional probability distributions

$$p_X : \mathrm{dom}(X) \times \prod_{Y\in \mathrm{pa}(X)} \mathrm{dom}(Y) \to \mathbb{R}_0^+$$

at the vertices $X \in V$ conditioned on its parents (called **(conditional) vertex probability distributions**)

is called a **bayesian network**.

Thus, a bayesian network encodes

(i) a joint probability distribution factorized as

$$p = \prod_{X\in V} p(X|\,\mathrm{pa}(X))$$

and

(ii) conditional independency statements

$$I_G(X, Y|Z) \Rightarrow I_p(X, Y|Z)$$

G represents p, but not necessarily faithfully.

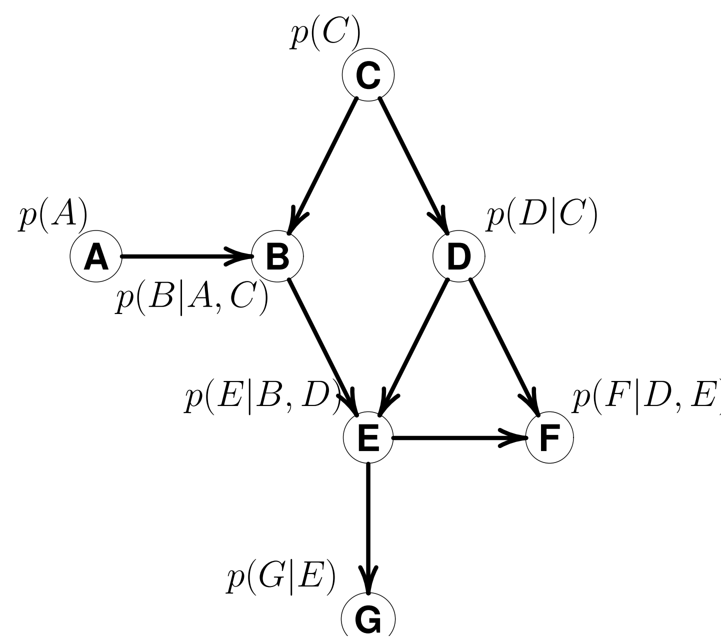


Figure 3: Example for a bayesian network.

Markov-equivalence

Definition 3. Let G, H be two graphs on a set V (undirected or DAGs).
G and H are called **markov-equivalent**, if they have the same independency model, i.e.

$$I_G(X, Y|Z) \Leftrightarrow I_H(X, Y|Z), \quad \forall X, Y, Z \subseteq V$$

The notion of markov-equivalence for undirected graphs is uninteresting, as every undirected graph is markov-equivalent only to itself. However, different DAGs can lead to the same (in)dependency model.

Markov-equivalence (DAGs)

Definition 4. Let G be a directed graph. We call a chain

$$p_1 - p_2 - p_3$$

uncoupled if there is no edge between p_1 and p_3.

Lemma 5 (markov-equivalence criterion, [PGV90]). *Let G and H be two DAGs on the vertices V.*
G and H are markov-equivalent if and only if

(i) G and H have the same links ($u(G) = u(H)$) and

(ii) G and H have the same uncoupled head-to-head meetings.

The set of uncoupled head-to-head meetings is also denoted as **V-structure** of G.

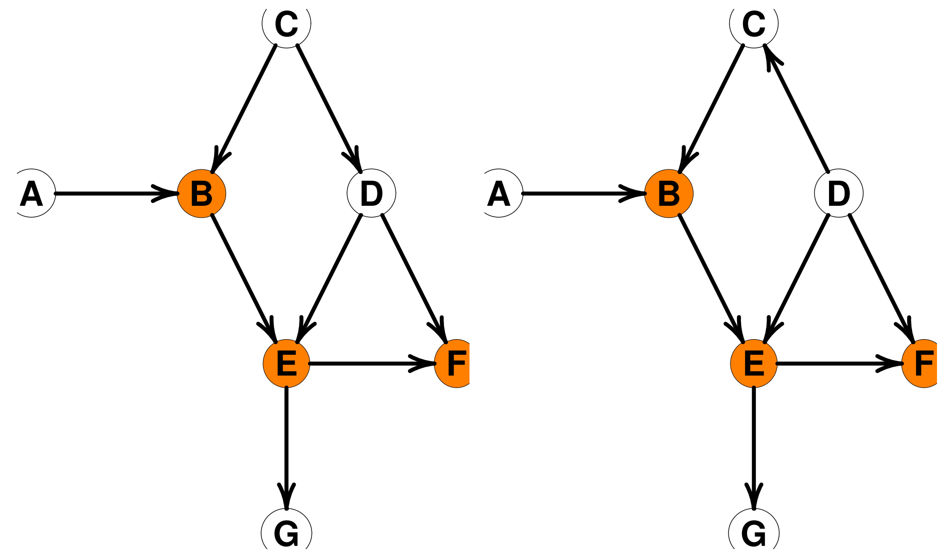


Figure 4: Example for markov-equivalent DAGs.

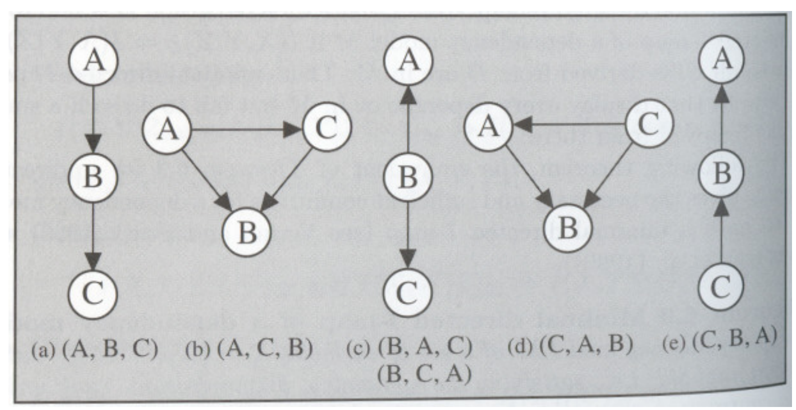


Figure 5: Which minimal DAG-representations of I are equivalent? [CGH97, p. 240]

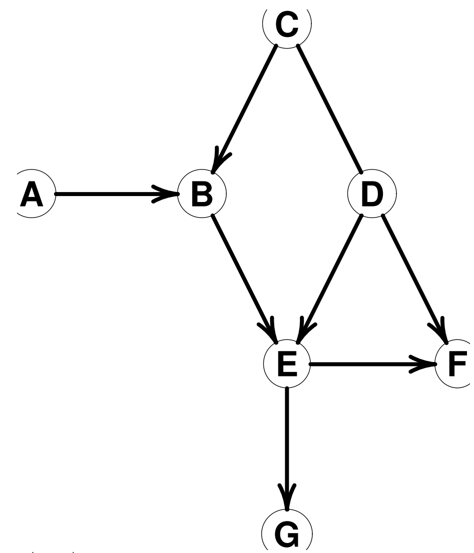
DAG patterns represent markov equivalence classes

Lemma 6. *Each markov equivalence class corresponds uniquely to a DAG pattern G:*

(i) The markov equivalence class consists of all DAGs that G is a pattern of, i.e., that give G by dropping the directions of some edges.

(ii) The DAG pattern contains a directed edge (v, w), if all representatives of the markov equivalence class contain this directed edge, otherwise (i.e. if some representatives have (v, w), some others (w, v)) the DAG pattern contains the undirected edge $\{v, w\}$.

The directed edges of the DAG pattern are also called **irreversible** or **compelled**, the undirected edges are also called **reversible**.



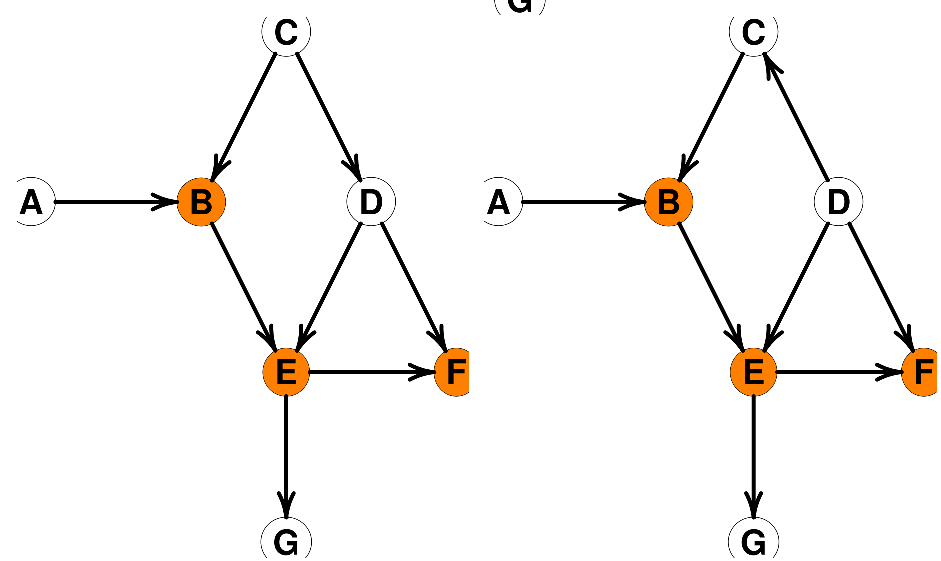


Figure 6: DAG pattern and its markov equivalence class representatives.

Toplogical edge ordering

Definition 5. Let $G := (V, E)$ be a directed graph.
A bijective map

$$\tau : \{1, \ldots, |E|\} \rightarrow E$$

is called an **ordering of the edges of** G.

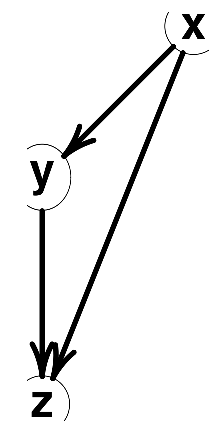
An edge ordering τ is called **topological edge ordering** if

(i) numbers increase on all paths, i.e.

$$\tau^{-1}(x, y) < \tau^{-1}(y, z)$$

for paths $x \rightarrow y \rightarrow z$ and

(ii) shortcuts have larger numbers, i.e. for x, y, z with



it is

$$\tau^{-1}(x, y) < \tau^{-1}(y, z) \overset{!}{<} \tau^{-1}(x, z)$$

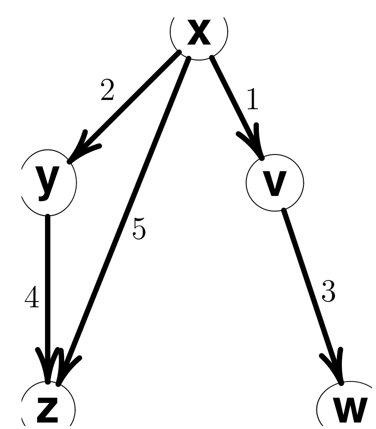


Figure 7: Example for a topological edge ordering.

Toplogical edge ordering

topological-edge-ordering($G = (V, E)$) :
$\sigma :=$ *topological-ordering*(G)
$E' := E$
for $i = 1, \ldots, |E|$ **do**
    Let $(v, w) \in E'$ with $\sigma^{-1}(w)$ minimal and then with $\sigma^{-1}(v)$ maximal
    $\tau(i) := (v, w)$
    $E' := E' \setminus \{(v, w)\}$
od
return τ

Figure 8: Algorithm for computing a topological edge ordering of a DAG.

dag-pattern$(G = (V, E))$:
$\tau :=$ *topological-edge-ordering*(G)
$E_{\text{irr}} := \emptyset$
$E_{\text{rev}} := \emptyset$
$E_{\text{rest}} := E$
while $E_{\text{rest}} \neq \emptyset$ **do**
    Let $(y, z) \in E_{\text{rest}}$ with $\tau^{-1}(y, z)$ minimal
    [label $\text{pa}(z)$:]
    if $\exists (x, y) \in E_{\text{irr}}$ with $(x, z) \notin E$
      $E_{\text{irr}} := E_{\text{irr}} \cup \{(x', z) \mid x' \in \text{pa}(z)\}$
    else
        $E_{\text{irr}} := E_{\text{irr}} \cup \{(x', z) \mid (x', y) \in E_{\text{irr}}\}$
        if $\exists (x, z) \in E$ with $x \notin \{y\} \cup \text{pa}(y)$
          $E_{\text{irr}} := E_{\text{irr}} \cup \{(x', z) \mid (x', z) \in E_{\text{rest}}\}$
        else
            $E_{\text{rev}} := E_{\text{rev}} \cup \{(x', z) \mid (x', z) \in E_{\text{rest}}\}$
        fi
    fi
    $E_{\text{rest}} := E \setminus E_{\text{irr}} \setminus E_{\text{rev}}$
od
return $\bar{G} := (V, E_{\text{irr}} \cup \{\{v, w\} | (v, w) \in E_{\text{rev}}\})$

Figure 9: Algorithm for computing the DAG pattern representing the markov equivalence class of a DAG G. [Chi95]

Types of probabilistic networks

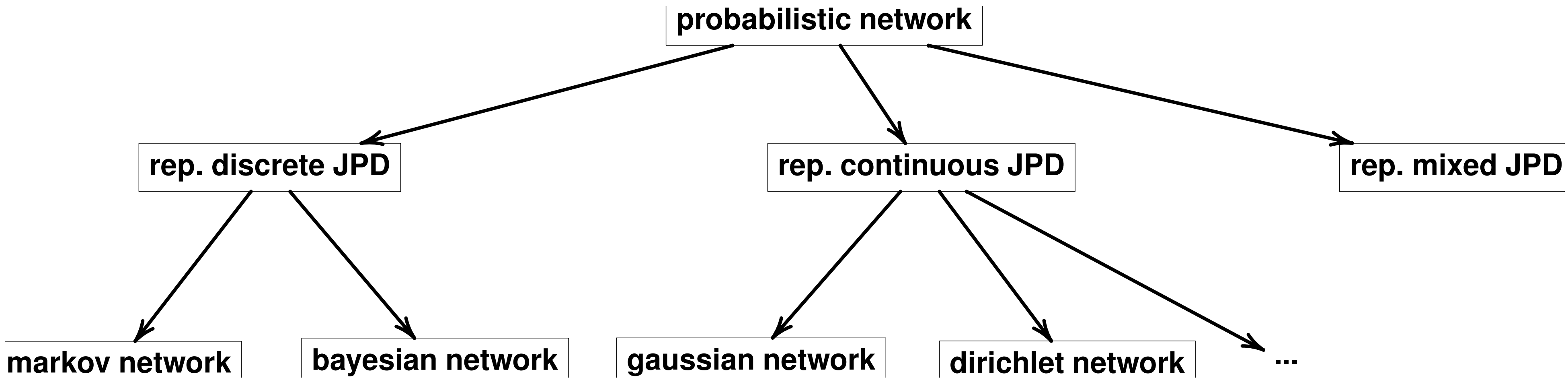


Figure 10: Types of probabilistic networks.

References

[CGH97] Enrique Castillo, José Manuel Gutiérrez, and Ali S. Hadi. *Expert Systems and Probabilistic Network Models*. Springer, New York, 1997.

[Chi95] D. Chickering. A transformational characterization of equivalent bayesian network structures. In Philippe Besnard and Steve Hanks, editors, *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, pages 87–98. Morgan Kaufmann, 1995.

[PGV90] J. Pearl, D. Geiger, and T. S. Verma. The logic of influence diagrams. In R. M. Oliver and J. Q. Smith, editors, *Influence Diagrams, Belief Networks and Decision Analysis*. Wiley, Sussex, 1990. (a shorter version originally appeared in Kybernetica, Vol. 25, No. 2, 1989).

[VP90] Thomas Verma and Judea Pearl. Causal networks: semantics and expressiveness. In Ross D. Shachter, Tod S. Levitt, Laveen N. Kanal, and John F. Lemmer, editors, *Uncertainty in Artificial Intelligence 4*, pages 69–76. North-Holland, Amsterdam, 1990.