Bayesian Networks

3. Bayesian and Markov Networks

Lars Schmidt-Thieme

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

1. Complete Graphs, DAGs and Topological Orderings
2. Graph Representations of Ternary Relations
3. Markov Networks
4. Bayesian Networks
Complete (undirected) graphs

**Definition 1.** An undirected graph $G := (V, E)$ is called **complete**, if it contains all possible edges (i.e. if $E = \mathcal{P}^2(V)$).

![Undirected complete graph with 6 vertices.](image1)

**Definition 2.** Let $G := (V, E)$ be a directed graph. A bijective map

$$\sigma : \{1, \ldots, |V|\} \to V$$

is called an **ordering of (the vertices of) $G$**.

We can write an ordering as enumeration of $V$, i.e. as $v_1, v_2, \ldots, v_n$ with $V = \{v_1, \ldots, v_n\}$ and $v_i \neq v_j$ for $i \neq j$.

![Ordering of a directed graph.](image2)
**Definition 3.** An ordering $\sigma = (v_1, \ldots, v_n)$ is called **topological ordering** if

(i) all parents of a vertex have smaller numbers, i.e.
\[ \text{fanin}(v_i) \subseteq \{v_1, \ldots, v_{i-1}\}, \quad \forall i = 1, \ldots, n \]

or equivalently

(ii) all edges point from smaller to larger numbers
\[ (v, w) \in E \Rightarrow \sigma^{-1}(v) < \sigma^{-1}(w), \quad \forall v, w \in V \]

The reverse of a topological ordering – e.g. got by using the fanout instead of the fanin – is called **ancestral numbering**.

In general there are several topological orderings of a DAG.

---

**Lemma 1.** Let $G$ be a directed graph. Then

$G$ is acyclic (a DAG) $\iff$ $G$ has a topological ordering

---

**Exercise:** write an algorithm for checking if a given directed graph is acyclic.
**Definition 4.** A DAG $G := (V, E)$ is called complete, if

(i) it has a topological ordering $\sigma = (v_1, \ldots, v_n)$ with $\text{fanin}(v_i) = \{v_1, \ldots, v_{i-1}\}$, $\forall i = 1, \ldots, n$

or equivalently

(ii) it has exactly one topological ordering

or equivalently

(iii) every additional edge introduces a cycle.

![Figure 5: Complete DAG with 6 vertices. Its topological ordering is $\sigma = (A, B, C, D, E, F)$.](image)

---

## 1. Complete Graphs, DAGs and Topological Orderings

## 2. Graph Representations of Ternary Relations

## 3. Markov Networks

## 4. Bayesian Networks
Definition 5. Let $V$ be a set and $I$ a ternary relation on $\mathcal{P}(V)$ (i.e. $I \subseteq \mathcal{P}(V)^3$). In our context $I$ is often called an independence model.

Let $G$ be a graph on $V$ (undirected or DAG). $G$ is called a representation of $I$, if

$$I_G(X,Y|Z) \Rightarrow I(X,Y|Z) \quad \forall X,Y,Z \subseteq V$$

A representation $G$ of $I$ is called faithful, if

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

Representations are also called independency maps of $I$ or markov w.r.t. $I$, faithful representations are also called perfect maps of $I$.

Faithful representations

In $G$ also holds

$$I_G(B,\{A,C\}|D), \quad I_G(B, A|D), \quad I_G(B, C|D), \quad \ldots$$

so $G$ is not a representation of

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

at all. It is a representation of

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

and as all independency statements of $J$ hold in $G$, it is faithful.
Trivial representations

For a complete undirected graph or a complete DAG \( G := (V, E) \) there is
\[ I_G \equiv \text{false}, \]
i.e. there are no triples \( X, Y, Z \subseteq V \) with \( I_G(X,Y|Z) \). Therefore \( G \) represents any independency model \( I \) on \( V \) and is called trivial representation.

There are independency models without faithful representation.

Minimal representations

**Definition 6.** A representation \( G \) of \( I \) is called minimal, if none of its subgraphs omitting an edge is a representation of \( I \).

\[ I := \{(A, B|\{C, D\})\} \]

without faithful representation.
Lemma 2 (uniqueness of minimal undirected representation). An independency model $I$ has exactly one minimal undirected representation, if and only if it is

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

(ii) decomposable: $I(X,Y|Z) \Rightarrow I(X,Y'|Z)$ for any $Y' \subseteq Y$.

(iii) intersectable: $I(X,Y|Y' \cup Z)$ and $I(X,Y'|Y \cup Z) \Rightarrow I(X,Y \cup Y'|Z)$.

Then this representation is $G = (V,E)$ with

$$E := \{ \{x,y\} \in \mathcal{P}^2(V) | \text{not } I(x,y|V \setminus \{x,y\} ) \}$$

Example 1.

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

is symmetric, decomposable and intersectable.

Its unique minimal undirected representation is

If a faithful representation exists, obviously it is the unique minimal representation, and thus can be constructed by the rule in lemma 2.
Definition 7. 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.

$$I_G(X, Y \mid Z) \iff I_H(X, Y \mid Z), \quad \forall X, Y, Z \subseteq V$$

The notion of markov-equivalence for undirected graphs is uninteresting, as every undirected graph is markov-equivalent only to itself (corollary of uniqueness of minimal representation!).

1) + for decomposable JPDs.

There is provably no finite axiomatization of conditional independency of general JPDs.

It is still an open research problem, if there is a finite axiomatization of conditional independency for non-extreme JPDs.

Independency models that satisfy symmetry, decomposition, weak union, and contraction (as conditional independency of general JPDs) are called semi-graphoids. If they satisfy also intersection (as conditional independency of non-extreme JPDs), they are called graphoids.
Bayesian Networks / 2. Graph Representations of Ternary Relations

Properties of conditional independency / no composition

**Example 2** (example for composition in JPDs).

<table>
<thead>
<tr>
<th>$z$</th>
<th>$y_1$</th>
<th>$p(x \mid z, y_1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 0</td>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td>0 1</td>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td>1 0</td>
<td>0.75</td>
<td></td>
</tr>
<tr>
<td>1 1</td>
<td>0.75</td>
<td></td>
</tr>
</tbody>
</table>

$I(x, y_1 \mid z)$

<table>
<thead>
<tr>
<th>$z$</th>
<th>$y_1$</th>
<th>$p(x \mid z, y_1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 0</td>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td>0 1</td>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td>1 0</td>
<td>0.75</td>
<td></td>
</tr>
<tr>
<td>1 1</td>
<td>0.75</td>
<td></td>
</tr>
</tbody>
</table>

**Example 3** (counterexample for composition in JPDs).

<table>
<thead>
<tr>
<th>$z$</th>
<th>$y_1$</th>
<th>$p(x \mid z, y_1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 0</td>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td>0 1</td>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td>1 0</td>
<td>0.75</td>
<td></td>
</tr>
<tr>
<td>1 1</td>
<td>0.75</td>
<td></td>
</tr>
</tbody>
</table>

$I(x, y_1 \mid z)$

<table>
<thead>
<tr>
<th>$z$</th>
<th>$y_1$</th>
<th>$p(x \mid z, y_1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 0</td>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td>0 1</td>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td>1 0</td>
<td>0.75</td>
<td></td>
</tr>
<tr>
<td>1 1</td>
<td>0.75</td>
<td></td>
</tr>
</tbody>
</table>

$Lars Schmidt-Thieme, Information Systems and Machine Learning Lab (ISMLL), Institute of Computer Science, University of Hildesheim
Course on Bayesian Networks, winter term 2013/14
14/38
1. Complete Graphs, DAGs and Topological Orderings

2. Graph Representations of Ternary Relations

3. Markov Networks

4. Bayesian Networks

Definition 8. We say, a graph represents a JPD $p$, if it represents the conditional independency relation $I_p$ of $p$.

General JPDs may have several minimal undirected representations (as they may violate the intersection property).

Non-extreme JPDs have a unique minimal undirected representation.

To compute this representation we have to check $I_p(X, Y | V \setminus \{X, Y\})$ for all pairs of variables $X, Y \in V$, i.e.

$$p \cdot p^{\downarrow V \setminus \{X, Y\}} = p^{\downarrow V \setminus \{X\}} \cdot p^{\downarrow V \setminus \{Y\}}$$

Then the minimal representation is the complete graph on $V$ omitting the edges $\{X, Y\}$ for that $I_p(X, Y | V \setminus \{X, Y\})$ holds.
Representation of conditional independency

Example 4. Let \( p \) be the JPD on \( V := \{X, Y, Z\} \) given by:

\[
\begin{array}{ccc|c}
Z & X & Y & p(X, Y, Z) \\
0 & 0 & 0 & 0.024 \\
0 & 0 & 1 & 0.056 \\
0 & 1 & 0 & 0.036 \\
0 & 1 & 1 & 0.084 \\
1 & 0 & 0 & 0.096 \\
1 & 0 & 1 & 0.144 \\
1 & 1 & 0 & 0.224 \\
1 & 1 & 1 & 0.336 \\
\end{array}
\]

Checking \( p \cdot p_{V\setminus\{X,Y\}} = p_{V\setminus\{X\}} \cdot p_{V\setminus\{Y\}} \) one finds that the only independency relations of \( p \) are \( I_p(X, Y|Z) \) and \( I_p(Y, X|Z) \).

Its marginals are:

\[
\begin{array}{ccc|c}
Z & X & p(X, Z) \\
0 & 0 & 0.08 \\
0 & 1 & 0.12 \\
1 & 0 & 0.24 \\
1 & 1 & 0.56 \\
\end{array}
\quad
\begin{array}{ccc|c}
Z & Y & p(Y, Z) \\
0 & 0 & 0.06 \\
0 & 1 & 0.14 \\
1 & 0 & 0.32 \\
1 & 1 & 0.48 \\
\end{array}
\quad
\begin{array}{ccc|c}
X & Y & p(X, Y) \\
0 & 0 & 0.12 \\
0 & 1 & 0.2 \\
1 & 0 & 0.26 \\
1 & 1 & 0.42 \\
\end{array}
\quad
\begin{array}{ccc|c}
X & p(X) \\
0 & 0.32 \\
1 & 0.68 \\
\end{array}
\quad
\begin{array}{ccc|c}
Y & p(Y) \\
0 & 0.38 \\
1 & 0.62 \\
\end{array}
\quad
\begin{array}{ccc|c}
Z & p(Z) \\
0 & 0.2 \\
1 & 0.8 \\
\end{array}
\]

Thus, the graph

\[
\begin{array}{c}
(X) \\
\quad
\end{array}
\quad
\begin{array}{c}
(Y) \\
\quad
Y \\
\quad
Z
\end{array}
\]

represents \( p \), as its independency model is \( I_G := \{(X, Y|Z), (Y, X|Z)\} \).

As for \( p \) only \( I_p(X, Y|Z) \) and \( I_p(Y, X|Z) \) hold, \( G \) is a faithful representation.
**Definition 9.** Let \( p \) be a joint probability distribution of a set of variables \( V \). Let \( C \) be a cover of \( V \), i.e. \( C \subseteq \mathcal{P}(V) \) with \( \bigcup_{X \in C} X = V \).

\( p \) factorizes according to \( C \), if there are potentials

\[
\psi_X : \prod_{X \in X} X \to \mathbb{R}_0^+, \quad X \in C
\]

with

\[
p = \prod_{X \in C} \psi_X
\]

In general, the potentials are not unique and do not have a natural interpretation.

**Example 5.**

<table>
<thead>
<tr>
<th>( Z )</th>
<th>( X )</th>
<th>( Y )</th>
<th>( p(X, Y, Z) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.024</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.056</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.036</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.084</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.096</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0.144</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0.224</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.336</td>
</tr>
</tbody>
</table>

\( p \) factorizes according to \( C = \{\{X, Z\}, \{Y, Z\}\} \) as

\[
p = p(X, Z) \cdot p(Y | Z)
\]

**Definition 10.** Let \( G \) be an undirected graph. A maximal complete subgraph of \( G \) is called a **clique** of \( G \). \( C_G \) denotes the set of all cliques of \( G \).

\( p \) factorizes according to \( G \), if it factorizes according to its clique cover \( C_G \).

The factorization induced by the complete graph is trivial.

**Example 6.** The JPD \( p \) from last example factorized according to the graph

as it has cliques \( C = \{\{X, Z\}, \{Y, Z\}\} \).
**Lemma 3.** Let $p$ be a JPD of a set of variables $V$, $G$ be an undirected graph on $V$. Then

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

(ii) If $p > 0$ then

$p$ factorizes acc. to $G$ $\iff$ $G$ represents $p$.

(iii) If $p > 0$ then $p$ factorizes acc. to its (unique) minimal representation.

(iv) If $G$ is an undirected graph and $\psi_X$ for $X \in C_G$ are any potentials on its cliques, then $G$ represents the JPD

$$p := \left( \prod_{X \in C_G} \psi_X \right)^{\lambda}$$

---

**Definition 11.** Let $G$ be an undirected graph and $C_G$ be its cliques. A sequence $C_1, \ldots, C_n$ of cliques of $G$ is called **chain of cliques**, if

1. every clique occurs exactly once and
2. the **running intersection property** holds:

$$C_i \cap \bigcup_{j=1}^{i-1} C_j \subseteq C_k, \quad \forall i \exists k < i$$

---

**Figure 12:** A graph with chain of cliques $\{A, B, C\}$, $\{B, C, D, E\}$, $\{E, F, G\}$ and $\{E, G, H\}$.

**Figure 13:** A graph with cliques $\{A, B, C\}$, $\{B, D\}$, $\{C, E\}$, $\{D, E\}$, $\{E, F, G\}$ and $\{E, G, H\}$, but without chain of cliques.
**Definition 12.** Let $G$ be an undirected graph. $G$ is called **triangulated** (or **chordal**), if every cycle of length $\geq 4$ has a chord, i.e., it exists an additional edge in $G$ between non-successive vertices of the cycle.

**Lemma 4.** $G$ is chordal $\iff I_G$ is chordal.

**Perfect ordering**

**Definition 13.** Let $G$ be an undirected graph. An ordering $\sigma$ of (the vertices of) $G$ is called **perfect**, if

1. $\sigma(i)$ and its neighbors form a clique of the subgraph on $\sigma(\{1,\ldots,i\})$ or equivalently
2. the subgraph on $\text{fan}(\sigma(i)) \cap \sigma(\{1,\ldots,i-1\})$ is complete.

A perfect ordering is also called a **perfect numbering**. The reverse of a perfect ordering is also called **elimination** or **deletion sequence**.
Lemma 5. Let $G$ be an undirected graph. It is equivalent:

(i) $G$ is triangulated / chordal.
(ii) $G$ admits a perfect ordering.
(iii) $G$ admits a chain of cliques.

Figure 19: MCS finds the perfect ordering $(A, B, C, D, E, F, G, H)$.

```
1 perfect-ordering-MCS($G = (V, E)$)
2  for $i = 1, \ldots, |V|$ do
3    $\sigma(i) := v \in V \setminus \sigma(\{1, \ldots, i - 1\})$ with maximal $|\text{fan}_G(v) \cap \sigma(1, \ldots, i - 1)|$
4    breaking ties arbitrarily
5  od
6 return $\sigma$
```

Figure 20: Algorithm to find a perfect ordering of a triangulated graph by maximum cardinality search.

```
1 chain-of-cliques($G$) :
2  $C := \text{enumerate-cliques} (G)$
3  $\sigma := \text{perfect-ordering} (G)$
4  Order $C$ by ascending $\max(\sigma^{-1}(C))$ for $C \in C$
5    breaking ties arbitrarily
6  return $C$
```

Figure 21: Algorithm to find a chain of cliques of a triangulated graph.

Based on the perfect ordering $G, E, B, C, H, D, F$, the rank of the cliques is computed as $\{A, B, C\}$ (3), $\{B, C, D, E\}$ (5), $\{E, F, G\}$ (7) and $\{E, G, H\}$ (8). The algorithm outputs the chain of cliques $\{A, B, C\}$, $\{B, C, D, E\}$, $\{E, F, G\}$ and $\{E, G, H\}$. Based on the perfect ordering $A, B, C, D, E, F, G, H$, the rank of the cliques is computed as $\{A, B, C\}$ (8), $\{B, C, D, E\}$ (6), $\{E, F, G\}$ (7) and $\{E, G, H\}$ (5). The algorithm outputs the chain of cliques $\{E, G, H\}$, $\{B, C, D, E\}$, $\{E, F, G\}$ and $\{A, B, C\}$.
**Definition 14.** A joint probability distribution \( p \) is called **decomposable**, if its conditional independency relation \( I_p \) is chordal.

**Warning.** \( p \) being decomposable has nothing to do with \( I_p \) being decomposable!

**Definition 15.** Let \( G \) be a triangulated / chordal graph and \( C = C_1, \ldots, C_n \) a chain of cliques of \( G \). Then
\[
S_i := C_i \cap \bigcup_{j<i} C_j
\]
is called the \( i \)-th separator.

**Lemma 6.** Let \( p \) be a JPD of a set of variables \( V \), \( G \) be an undirected graph on \( V \). If \( G \) represents \( p \) and \( p \) is decomposable (i.e. \( G \) triangulated/chordal), let \( C = C_1, \ldots, C_n \) be a chain of cliques, and then
\[
p = \prod_{i=1}^n p \downarrow C_i \mid S_i
\]
i.e. \( p \) factorizes in the conditional probability distributions of the cliques given its separators.

---

**Markov networks**

**Definition 16.** A pair \((G, (\psi_C)_{C \in C_G})\) consisting of

(i) an undirected graph \( G \) on a set of variables \( V \) and

(ii) a set of potentials
\[
\psi_C : \prod_{X \in C} \text{dom}(X) \to \mathbb{R}_0^+, \quad C \in C_G
\]
on the cliques\(^1\) of \( G \) (called **clique potentials**)

is called a **markov network**.

Thus, a markov network encodes

(i) a joint probability distribution factorized as
\[
p = (\prod_{C \in C_G} \psi_C)^0
\]
and

(ii) conditional independency statements
\[
I_G(X, Y \mid Z) \Rightarrow I_p(X, Y \mid Z)
\]

\( G \) represents \( p \), but not necessarily faithfully.

Under some regularity conditions (not covered here), \( \psi_{C_i} \) can be choosen as conditional probabilities \( p \downarrow C_i \mid S_i \).

\(^1\) on the product of the domains of the variables of each clique.
Figure 23: Example for a markov network.
### Markov networks

<table>
<thead>
<tr>
<th><strong>structure</strong></th>
<th><strong>probability distribution</strong></th>
<th><strong>markov network</strong></th>
<th><strong>u-separation in graph</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>conditional independence $I_p$</td>
<td></td>
<td>representations exist always (e.g., trivial representation)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Sym+Dec+Int+SUn+STrans $\Leftrightarrow$ faithful (Lemma 2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>minimal representations exist always</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Sym+Dec+Int $\Rightarrow$ unique minimal (Lemma 3)</td>
</tr>
<tr>
<td></td>
<td>e.g. for $p$ non-extreme</td>
<td></td>
<td>different graphs give different representations (trivial markov-equivalence)</td>
</tr>
</tbody>
</table>

### Bayesian networks

<table>
<thead>
<tr>
<th><strong>structure</strong></th>
<th><strong>probability distribution</strong></th>
<th><strong>bayesian network</strong></th>
<th><strong>d-separation in graph</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>conditional independence $I_p$</td>
<td></td>
<td>representations exist always (e.g., trivial representation)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Sym+Dec+Comp+Contr+Int+WUn+WTrans+Chor $\Leftrightarrow$ faithful (Lemma)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>minimal representations exist always</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Sym+Dec+Contr+Int+WUn $\Rightarrow$ unique minimal up to ordering (Lemma)</td>
</tr>
<tr>
<td></td>
<td>e.g. for $p$ non-extreme</td>
<td></td>
<td>graphs with same DAG pattern give same representation (markov-equivalence)</td>
</tr>
</tbody>
</table>

- large probability table $p$ (i.e. $I_p$ chordal/triangulated) if $p$ is decomposable
- clique potentials $\phi$
- if $G$ is chordal/triangulated $\Rightarrow$ conditional clique probabilities $p(C_i|S_i)$ for a chain of cliques $C = (C_1, \ldots, C_n)$.
DAG-representations

Lemma 7 (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 $p \iff v$ and nondesc($v$) are conditionally independent given $\text{pa}(v)$ for all $v \in V$, i.e.,

$$I_p(\{v\}, \text{nondesc}(v)|\text{pa}(v)), \; \forall v \in V$$

It is still an open research problem, if there is a finite axiomatisation of faithful DAG-representability.
Example for a not faithfully DAG-representable independency model

Probability distributions may have no faithful DAG-representation.

Example 7. The independency model

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

does not have a faithful DAG-representation. [CGH97, p. 239]

Exercise: compute all minimal DAG-representations of \( I \) using lemma 9 and check if they are faithful.

---

**Lemma 9** (construction and uniqueness of minimal DAG-representation, [VP90]). Let \( I \) be an independence model of a JPD \( p \). 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 minimal set \( \pi_i \subseteq \{v_1, \ldots, v_{i-1}\} \) of \( \sigma \)-precurors of \( v_i \) with

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

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

\[ E := \{(w, v_i) \mid i = 1, \ldots, n, w \in \pi_i\} \]

is a minimal DAG-representation of \( p \).

(ii) If \( p \) also is non-extreme, then the minimal representation \( G \) is unique up to ordering \( \sigma \).
Minimal DAG-representations / example

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

Figure 25: Minimal DAG-representations of \( I \) [CGH97, p. 240].

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

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

<table>
<thead>
<tr>
<th></th>
<th>Markov network (undirected)</th>
<th>Bayesian network (directed)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>minimal</td>
<td>faithful</td>
</tr>
<tr>
<td>general JPD</td>
<td>may not be unique</td>
<td>may not exist</td>
</tr>
<tr>
<td>non-extreme JPD</td>
<td>unique</td>
<td>may not exist</td>
</tr>
</tbody>
</table>
Bayesian Network

Definition 17. 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
$p_X : \text{dom}(X) \times \prod_{Y \in \text{pa}(X)} \text{dom}(Y) \rightarrow \mathbb{R}^+_0$

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

and
(ii) conditional independency statements
$I_G(X, Y | Z) \Rightarrow I_p(X, Y | Z)$

$G$ represents $p$, but not necessarily faithfully.

Figure 26: Example for a bayesian network.

Bayesian Networks / 4. Bayesian Networks

Types of probabilistic networks

---

Figure 27: Types of probabilistic networks.
References
