Bayesian Networks

3. Bayesian and Markov Networks

Lars Schmidt-Thieme

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

Complete (undirected) graphs

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

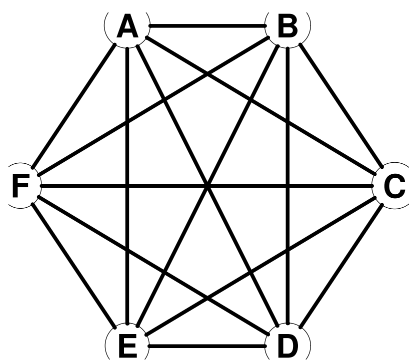


Figure 1: Undirected complete graph with 6 vertices.

Orderings (of a directed graph)

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 $i \neq j$.

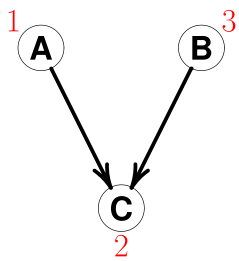


Figure 2: Ordering of a directed graph.

Topological orderings

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.

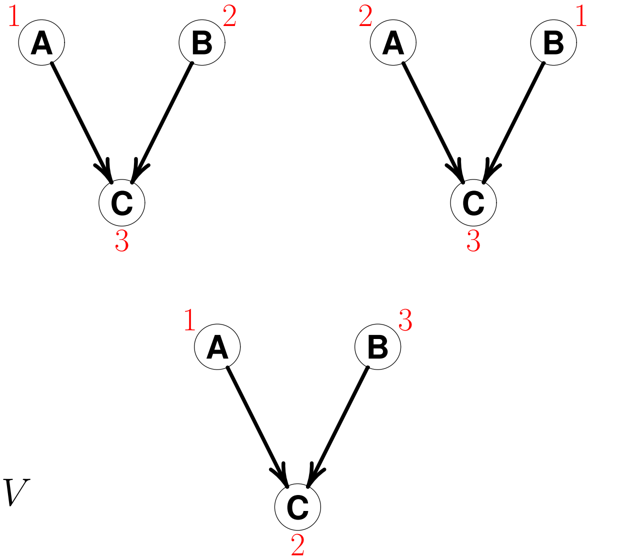


Figure 3: DAG with different topological orderings: $\sigma_1 = (A, B, C)$ and $\sigma_2 = (B, A, C)$. The ordering $\sigma_3 = (A, C, B)$ is not topological.

Topological orderings and DAGs

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

$$G \text{ is acyclic (a DAG)} \Leftrightarrow G \text{ has a topological ordering}$$

topological-ordering$(G = (V, E))$:
choose $v \in V$ with $\text{fanout}(v) = \emptyset$
$\sigma(|V|) := v$
$\sigma|_{\{1,\ldots,|V|-1\}} :=$ topological-ordering$(G \setminus \{v\})$
return σ

Figure 4: Algorithm to compute a topologcial ordering of a DAG.

Exercise: write an algorithm for checking if a given directed graph is acyclic.

Complete DAGs

Definition 4. A DAG $G := (V, E)$ is called complete, if

(i) it has a topological ordering $\sigma = (v_1, \ldots, v_n)$ with

$$\text{fanin}(v_i) = \{v_1, \ldots, v_{i-1}\}, \quad \forall i = 1, \ldots, n$$

or equivalently

(ii) it has exactly one topological ordering

or equivalently

(iii) every additional edge introduces a cycle.

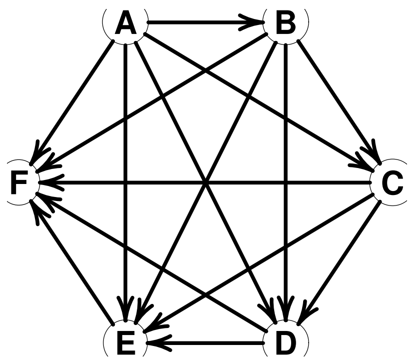


Figure 5: Complete DAG with 6 vertices. Its topological ordering is $\sigma = (A, B, C, D, E, F)$.

Graph representations of ternary relations on $\mathcal{P}(V)$

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 **independency 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.

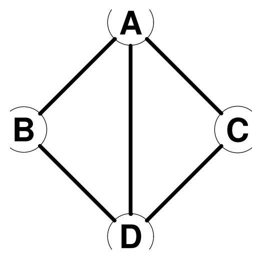


Figure 6: Non-faithful representation of

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

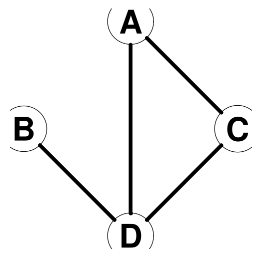


Figure 7: Faithful representation of I. Which I?

Faithful representations

In G also holds

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

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

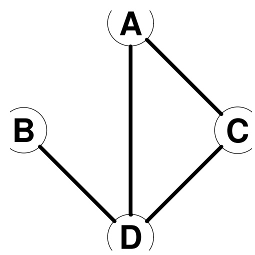


Figure 8: Faithful representation of J.

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

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.

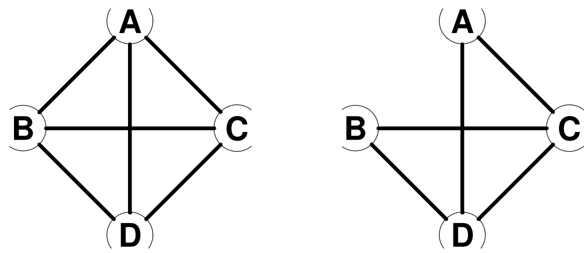


Figure 9: Independency model

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

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.

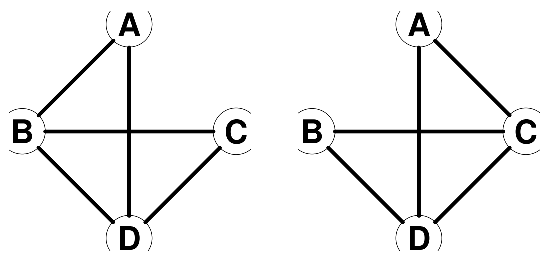


Figure 10: Different minimal undirected representations of the independency model

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

Minimal representations

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*

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

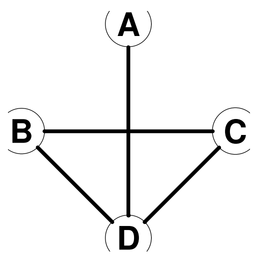
Minimal representations (2/2)

Example 1.

$$\begin{aligned} I := \{&(A, B|\{C, D\}), (A, C|\{B, D\}), (A, \{B, C\}|D), (A, B|D), (A, C|D),\\ &(B, A|\{C, D\}), (C, A|\{B, D\}), (\{B, C\}, A|D), (B, A|D), (C, A|D)\} \end{aligned}$$

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.

Markov-equivalence

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.

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

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

Properties of conditional independency

relation	symmetry	decomposition	composition	strong union	weak union	contraction	intersection	strong transitivity	weak transitivity	chordality
u-separation	+	+	+	+	+	+	+	+	+	−
d-separation	+	+	+	−	+		+	−	+	+
cond. ind. in general JPD	+	+	−	−	+	+	−	−	−	−[1)]
cond. ind. in non-extreme JPD	+	+	−	−	+	+	+	−	−	−[1)]

[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**.

Properties of conditional independency / no composition

Example 2 (example for composition in JPDs).

z	y_1	$p(x\|z, y_1)$
0	0	0.2
0	1	0.2
1	0	0.75
1	1	0.75

$I(x, y_1|z)$

z	y_2	$p(x\|z, y_2)$
0	0	0.2
0	1	0.2
1	0	0.75
1	1	0.75

$I(x, y_2|z)$

z	x	y_1	y_2	$p(x, y_1, y_2, z)$
0	0	0	0	0.04
0	0	0	1	0.04
0	0	1	0	0.04
0	0	1	1	0.04
0	1	0	0	0.01
0	1	0	1	0.01
0	1	1	0	0.01
0	1	1	1	0.01
1	0	0	0	0.05
1	0	0	1	0.05
1	0	1	0	0.05
1	0	1	1	0.05
1	1	0	0	0.15
1	1	0	1	0.15
1	1	1	0	0.15
1	1	1	1	0.15

z	y_1	y_2	$p(x\|z, y_1, y_2)$
0	0	0	0.2
0	0	1	0.2
0	1	0	0.2
0	1	1	0.2
1	0	0	0.75
1	0	1	0.75
1	1	0	0.75
1	1	1	0.75

$I(x, \{y_1, y_2\}|z)$

z	$p(z)$
0	0.2
1	0.8

Properties of conditional independency / no composition

Example 3 (counterexample for composition in JPDs).

z	y_1	$p(x\|z, y_1)$
0	0	0.2
0	1	0.2
1	0	0.75
1	1	0.75

$I(x, y_1|z)$

z	y_2	$p(x\|z, y_2)$
0	0	0.2
0	1	0.2
1	0	0.75
1	1	0.75

$I(x, y_2|z)$

z	x	y_1	y_2	$p(x, y_1, y_2, z)$
0	0	0	0	~~0.04~~ 0.03
0	0	0	1	~~0.04~~ 0.05
0	0	1	0	~~0.04~~ 0.05
0	0	1	1	~~0.04~~ 0.03
0	1	0	0	0.01
0	1	0	1	0.01
0	1	1	0	0.01
0	1	1	1	0.01
1	0	0	0	0.05
1	0	0	1	0.05
1	0	1	0	0.05
1	0	1	1	0.05
1	1	0	0	0.15
1	1	0	1	0.15
1	1	1	0	0.15
1	1	1	1	0.15

z	y_1	y_2	$p(x\|z, y_1, y_2)$
0	0	0	~~0.2~~ 0.25
0	0	1	~~0.2~~ 0.17
0	1	0	~~0.2~~ 0.17
0	1	1	~~0.2~~ 0.25
1	0	0	0.75
1	0	1	0.75
1	1	0	0.75
1	1	1	0.75

$\neg I(x, \{y_1, y_2\}|z)!$

z	$p(z)$
0	0.2
1	0.8

Representation of conditional independency

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:

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

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

Its marginals are:

Z	X	$p(X, Z)$
0	0	0.08
0	1	0.12
1	0	0.24
1	1	0.56

Z	Y	$p(Y, Z)$
0	0	0.06
0	1	0.14
1	0	0.32
1	1	0.48

X	Y	$p(X, Y)$
0	0	0.12
0	1	0.2
1	0	0.26
1	1	0.42

X	$p(X)$
0	0.32
1	0.68

Y	$p(Y)$
0	0.38
1	0.62

Z	$p(Z)$
0	0.2
1	0.8

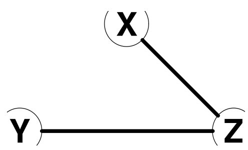
Representation of conditional independency

Example 4 (cont.).

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

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

Thus, the graph



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.

Factorization of a JPD according to a graph

Definition 9. Let p be a joint probability distribution of a set of variables V. Let $\mathcal{C}$ be a cover of V, i.e. $\mathcal{C} \subseteq \mathcal{P}(V)$ with $\bigcup_{\mathcal{X}\in\mathcal{C}} \mathcal{X} = V$.
p **factorizes according to** $\mathcal{C}$, if there are potentials

$$\psi_{\mathcal{X}} : \prod_{X\in\mathcal{X}} X \rightarrow \mathbb{R}_0^+, \quad \mathcal{X} \in \mathcal{C}$$

with

$$p = \prod_{\mathcal{X}\in\mathcal{C}} \psi_{\mathcal{X}}$$

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

Example 5.

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

Z	X	$p(X,Z)$
0	0	0.08
0	1	0.12
1	0	0.24
1	1	0.56

Z	Y	$p(Y,Z)$	$p(Y\|Z)$
0	0	0.06	0.3
0	1	0.14	0.7
1	0	0.32	0.4
1	1	0.48	0.6

p factorizes according to $\mathcal{C} = \{\{X,Z\},\{Y,Z\}\}$ as

$$p = p(X,Z) \cdot p(Y|Z)$$

Factorization of a JPD according to a graph

Definition 10. Let G be an undirected graph. A maximal complete subgraph of G is called a **clique of** G. $\mathcal{C}_G$ denotes the set of all cliques of G.

p **factorizes according to** G, if it factorizes according to its clique cover $\mathcal{C}_G$.

The factorization induced by the complete graph is trivial.

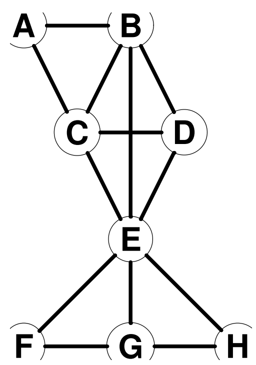
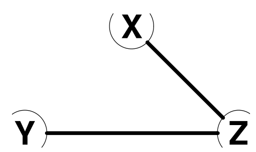


Figure 11: A graph with cliques $\{A,B,C\}$, $\{B,C,D,E\}$, $\{E,F,G\}$ and $\{E,G,H\}$.

Example 6. The JPD p from last example factorized according to the graph



as it has cliques $\mathcal{C} = \{\{X,Z\},\{Y,Z\}\}$

Factorization and representation

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 $G \Rightarrow G$ represents p.

(ii) If $p > 0$ then
p factorizes acc. to $G \Leftrightarrow G$ represents p.

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

(iv) If G is an undirected graph and $\psi_{\mathcal{X}}$ for $\mathcal{X} \in \mathcal{C}_G$ are any potentials on its cliques, then G represents the JPD

$$p := (\prod_{\mathcal{X} \in \mathcal{C}_G} \psi_{\mathcal{X}})^{|\emptyset}$$

Chain of cliques

Definition 11. Let G be an undirected graph and $\mathcal{C}_G$ 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:

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

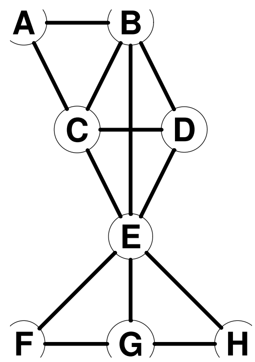


Figure 12: A graph with chain of cliques $\{A, B, C\}$, $\{B, C, D, E\}$, $\{E, F, G\}$ and $\{E, G, H\}$.

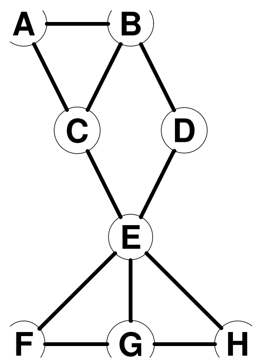


Figure 13: A graph with cliques $\{A, B, C\}$, $\{B, D\}$, $\{C, E\}$, $\{D, E\}$, $\{E, F, G\}$ and $\{E, G, H\}$, but without chain of cliques.

Triangulated/chordal graphs

Definition 12. Let G be an undirected graph.
G is called **triangulated** (or **chordal**), if every cycle of length ≥ 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 $\Leftrightarrow$ I_G is chordal.*

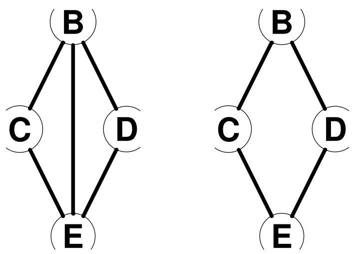


Figure 14: Cycle with chord and cycle without chord.

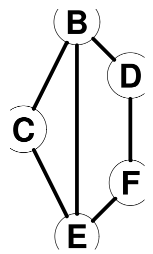


Figure 15: Chordal or non-chordal graph?

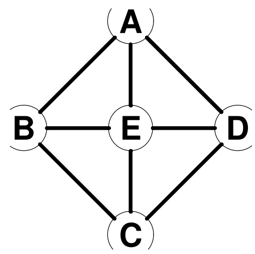


Figure 16: Chordal or non-chordal graph?

Perfect ordering

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

(i) $\sigma(i)$ and its neighbors form a clique of the subgraph on $\sigma(\{1,\ldots,i\})$

or equivalently

(ii) the subgraph on

$$\operatorname{fan}(\sigma(i)) \cap \sigma(\{1,\ldots,i-1\})$$

is complete.

A perfect ordering is also called a **perfect numbering**. The reverse of a perfect ordering is also called **elimination** or **deletion sequence**.

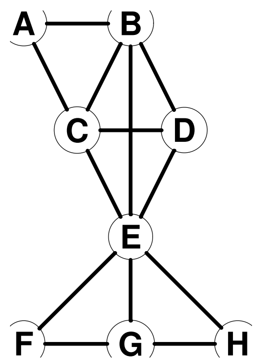


Figure 17: There are several perfect orderings of this graph, e.g., H, G, E, F, D, C, B, A and G, E, B, C, H, D, F, A.

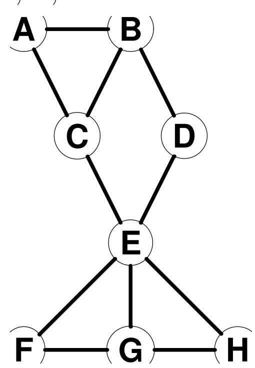


Figure 18: A graph without perfect ordering.

Triangulation, perfect ordering, and chain of cliques

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.

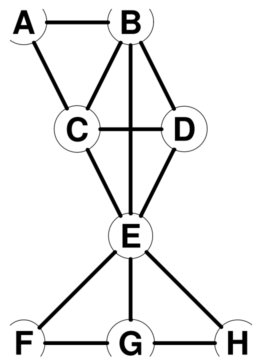


Figure 19: MCS finds the perfect ordering (A, B, C, D, E, F, G, H).

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

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

Triangulation, perfect ordering, and chain of cliques

chain-of-cliques(G) :
$\mathcal{C} :=$ *enumerate-cliques*(G)
$\sigma :=$ *perfect-ordering*(G)
Order $\mathcal{C}$ by ascending $\max(\sigma^{-1}(C))$ for $C \in \mathcal{C}$
  breaking ties arbitrarily
return $\mathcal{C}$

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

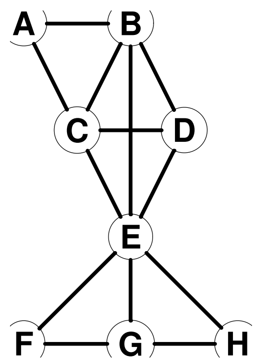


Figure 22: Based on the perfect ordering (A, B, C, D, E, F, G, H) the rank of the cliques is computed as $\{A, B, C\}$ (3) $\{B, C, D, E\}$ (5), $\{E, F, G\}$ (7) and $\{E, G, H\}$ (8). The algorithm outputs the chain of cliques $\{A, B, C\}$, $\{B, C, D, E\}$, $\{E, F, G\}$ and $\{E, G, H\}$.
Based on the perfect ordering $G, E, B, C, H, D, F,$ 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 $\{A, B, C\}$.

Factorization and representation (2/2)

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 $\mathcal{C} = C_1, \ldots, C_n$ 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 $\mathcal{C} = C_1, \ldots, C_n$ be a chain of cliques, and then*

$$p = \prod_{i=1}^{n} p^{\downarrow C_i | S_i}$$

i.e. p factorizes in the conditional probability distributions of the cliques given its separators.

Markov networks

Definition 16. A pair $(G, (\psi_C)_{C \in \mathcal{C}_G})$ consisting of

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

(ii) a set of potentials

$$\psi_C : \prod_{X \in C} \mathrm{dom}(X) \to \mathbb{R}_0^+, \quad C \in \mathcal{C}_G$$

on the cliques[1)] of G (called **clique potentials**)

is called a **markov network.**

[1)] on the product of the domains of the variables of each clique.

Thus, a markov network encodes

(i) a joint probability distribution factorized as

$$p = (\prod_{C \in \mathcal{C}_G} \psi_C)^{|\emptyset}$$

and

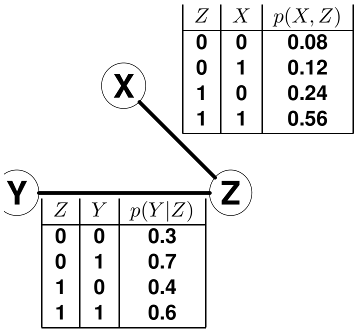
(ii) conditional independency statements

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

G represents p, but not necessarily faithfully.

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

Markov networks / examples



Z	X	$p(X,Z)$
0	0	0.08
0	1	0.12
1	0	0.24
1	1	0.56

Z	Y	$p(Y\|Z)$
0	0	0.3
0	1	0.7
1	0	0.4
1	1	0.6

Figure 23: Example for a markov network.

Markov networks

	probability distribution	markov network
structure	conditional independence I_p	u-separation in graph
	representations exist always (e.g., trivial representation) Sym+Dec+Int+SUn+STrans $\Leftrightarrow$ faithful (Lemma 2)	
	minimal representations exist always Sym+Dec+Int $\Rightarrow$ unique minimal (Lemma 3) e.g. for p non-extreme	
		different graphs give different representations (trivial markov-equivalence)
parameters	large probability table p	clique potentials ϕ
	if p is decomposable (i.e. I_p chordal/triangulated)	if G is chordal/triangulated $\Rightarrow$ conditional clique probabilities $p(C_i\|S_i)$ for a chain of cliques $\mathcal{C} = (C_1, \ldots, C_n)$.

Bayesian networks

	probability distribution	bayesian network
structure	conditional independence I_p	d-separation in graph
	representations exist always (e.g., trivial representation) Sym+Dec+Comp+Contr+Int+WUn+WTrans+Chor $\Leftarrow$ faithful (Lemma .	
	minimal representations exist always Sym+Dec+Contr+Int+WUn $\Rightarrow$ unique minimal up to ordering (Lemma e.g. for p non-extreme	
		graphs with same DAG pattern give same representation (markov-equivalence)
parameters	large probability table p	conditional vertex probabilities $p(v\|\,\mathrm{pa}(v))$

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

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

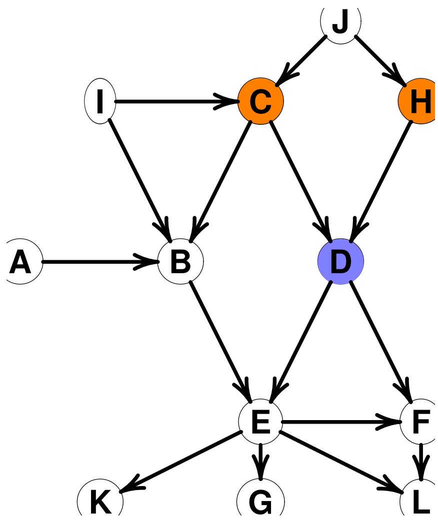


Figure 24: Parents of a vertex (orange).

Faithful DAG-representations

Lemma 8 (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) decomposable: $I(X, Y|Z) \Rightarrow I(X, Y'|Z)$ for any $Y' \subseteq Y$

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

(iv) contractable: $I(X, Y|Z)$ and $I(X, Y'|Y \cup Z) \Rightarrow I(X, Y \cup Y'|Z)$

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

(vi) weakly unionable: $I(X, Y|Z) \Rightarrow I(X, Y'|(Y \setminus Y') \cup Z)$ for any $Y' \subseteq Y$

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

(viii) 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 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.

Minimal DAG-representations

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 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 p also is non-extreme, then the minimal representation G is unique up to ordering σ.

Minimal DAG-representations / example

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

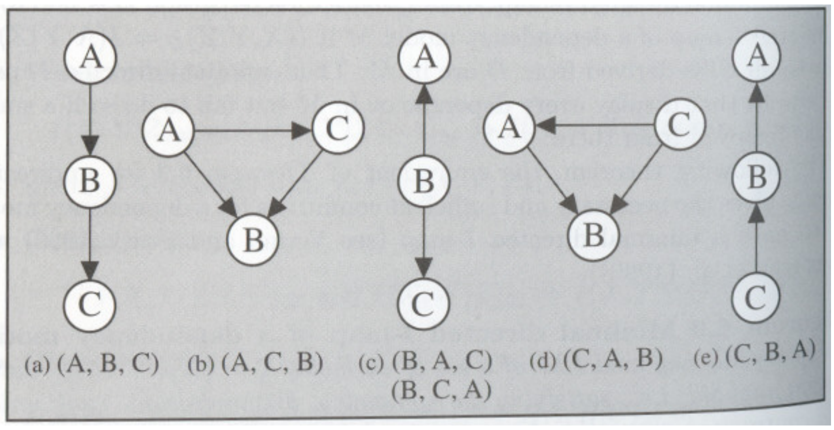


Figure 25: Minimal DAG-representations of I [CGH97, p. 240].

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

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

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

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

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

and

(ii) conditional independency statements

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

G represents p, but not necessarily faithfully.

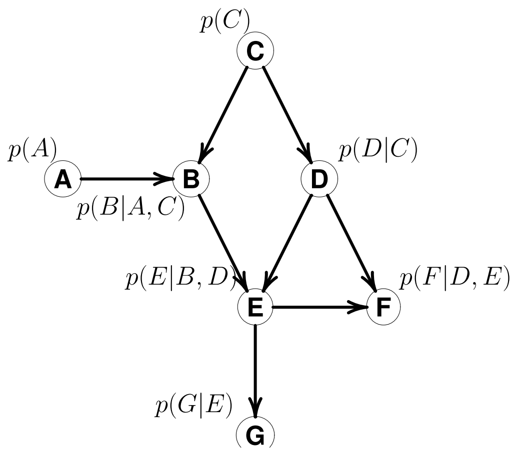


Figure 26: Example for a bayesian network.

Types of probabilistic networks

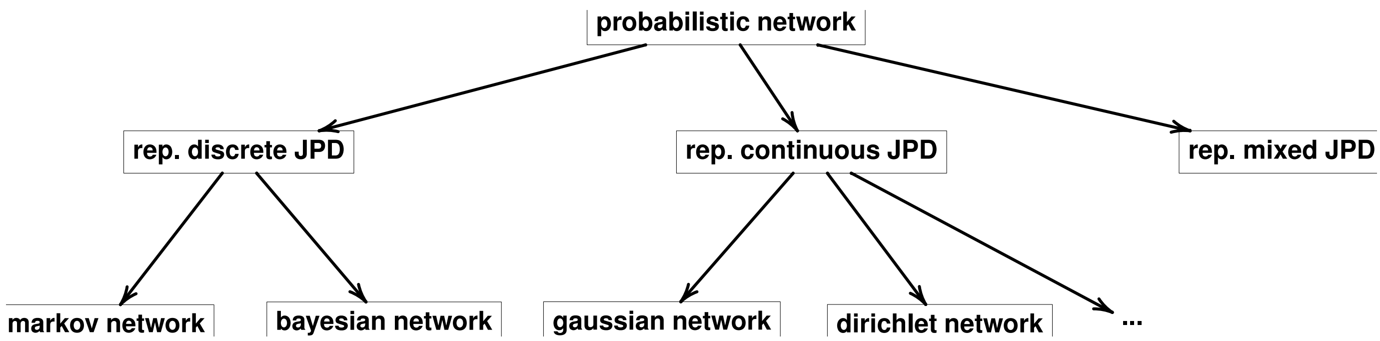


Figure 27: Types of probabilistic networks.

References

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

[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.