Bayesian Networks

I. Bayesian Networks / 1. Probabilistic Independence and Separation in Graphs

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

Outline of the Course

section	*key concepts*
I. Probabilistic Independence and Separation in Graphs	Prob. independence, separation in graphs, Markov and Bayesian Network
II. Inference	Exact inference, Approx. inference
III. Learning	Parameter Learning, Parameter Learning with missing values, Learning structure by Constrained-based Learning, Learning Structure by Local Search

1. Basic Probability Calculus

2. Separation in undirected graphs

Joint probability distributions

Pain	Y				N			
Weightloss	Y		N		Y		N	
Vomiting	Y	N	Y	N	Y	N	Y	N
Adeno Y	0.220	0.220	0.025	0.025	0.095	0.095	0.010	0.010
N	0.004	0.009	0.005	0.012	0.031	0.076	0.050	0.113

Figure 1: Joint probability distribution $p(P, W, V, A)$ of four random variables P (pain), W (weightloss), V (vomiting) and A (adeno).

Joint probability distributions

Discrete JPDs are described by

- nested tables,
- multi-dimensional arrays,
- data cubes, or
- tensors

having entries in $[0, 1]$ and summing to 1.

Marginal probability distributions

Definition 1. Let p be a the joint probability of the random variables $\mathcal{X} := \{X_1, \ldots, X_n\}$ and $\mathcal{Y} \subseteq \mathcal{X}$ a subset thereof. Then

$$p(\mathcal{Y} = y) := p^{\downarrow \mathcal{Y}}(y) := \sum_{x \in \operatorname{dom} \mathcal{X} \setminus \mathcal{Y}} p(\mathcal{X} \setminus \mathcal{Y} = x, \mathcal{Y} = y)$$

is a probability distribution of $\mathcal{Y}$ called **marginal probability distribution**.

Example 1. Marginal $p(V, A)$:

Vomiting	Y	N
Adeno Y	0.350	0.350
N	0.090	0.210

Pain	Y				N			
Weightloss	Y		N		Y		N	
Vomiting	Y	N	Y	N	Y	N	Y	N
Adeno Y	0.220	0.220	0.025	0.025	0.095	0.095	0.010	0.010
N	0.004	0.009	0.005	0.012	0.031	0.076	0.050	0.113

Figure 2: Joint probability distribution $p(P, W, V, A)$ of four random variables P (pain), W (weightloss), V (vomiting) and A (adeno).

Marginal probability distributions / example

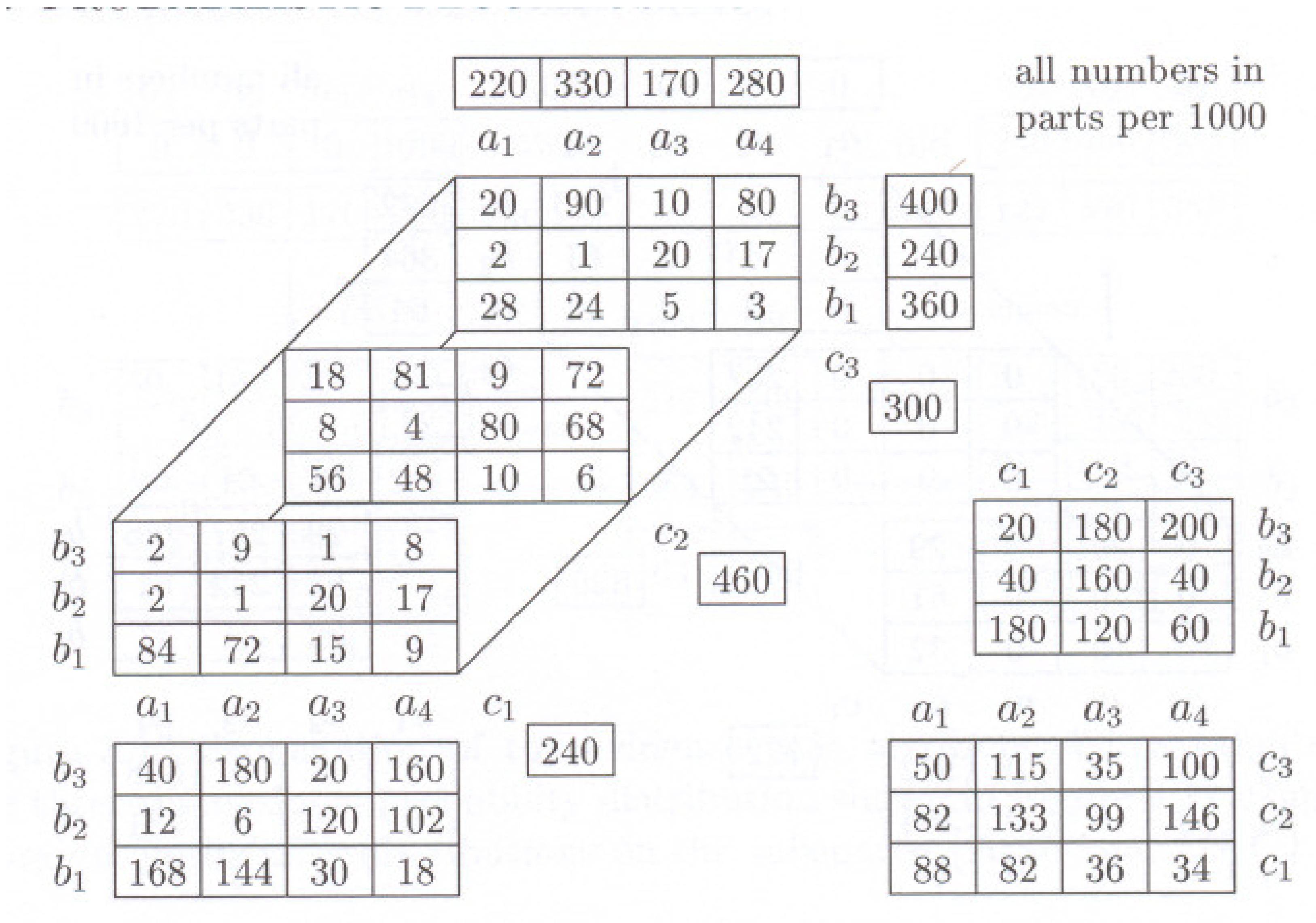


Figure 3: Joint probability distribution and all of its marginals [?, p. 75].

Extreme and non-extreme probability distributions

Definition 2. By $p > 0$ we mean

$$p(x) > 0, \quad \text{for all } x \in \prod \mathrm{dom}(p)$$

Then p is called **non-extreme**.

Example 2.

$$\begin{pmatrix} 0.4 & 0.0 \\ 0.3 & 0.3 \end{pmatrix}$$

$$\begin{pmatrix} 0.4 & 0.1 \\ 0.2 & 0.3 \end{pmatrix}$$

Conditional probability distributions

Definition 3. For a JPD p and a subset $\mathcal{Y} \subseteq \mathrm{dom}(p)$ of its variables with $p^{\downarrow\mathcal{Y}} > 0$ we define

$$p^{|\mathcal{Y}} := \frac{p}{p^{\downarrow\mathcal{Y}}}$$

as **conditional probability distribution of** p **w.r.t.** $\mathcal{Y}$.

A conditional probability distribution w.r.t. $\mathcal{Y}$ sums to 1 for all fixed values of $\mathcal{Y}$, i.e.,

$$(p^{|\mathcal{Y}})^{\downarrow\mathcal{Y}} \equiv 1$$

Conditional probability distributions / example

Example 3. Let p be the JPD

$$p := \begin{pmatrix} 0.4 & 0.1 \\ 0.2 & 0.3 \end{pmatrix}$$

on two variables R (rows) and C (columns) with the domains $\mathrm{dom}(R) = \mathrm{dom}(C) = \{1, 2\}$.

The conditional probability distribution w.r.t. C is

$$p^{|C} := \begin{pmatrix} 2/3 & 1/4 \\ 1/3 & 3/4 \end{pmatrix}$$

Chain rule

Lemma 1 (Chain rule)**.** *Let* p *be a JPD on variables* $X_1, X_2, \ldots, X_n$ *with* $p(X_1, \ldots, X_{n-1}) > 0$*. Then*

$$p(X_1, X_2, \ldots, X_n) = p(X_n | X_1, \ldots, X_{n-1}) \cdots p(X_2 | X_1) \cdot p(X_1)$$

The chain rule provides a **factorization** of the JPD in some of its conditional marginals.

The factorizations stemming from the chain rule are trivial as they have as many parameters as the original JPD:

$$\#\text{parameters} = 2^{n-1} + 2^{n-2} + \cdots + 2^1 + 2^0 = 2^n - 1$$

(example computation for all binary variables)

Bayes formula

Lemma 2 (Bayes Formula)**.** *Let* p *be a JPD and* $\mathcal{X}, \mathcal{Y}$ *be two disjoint sets of its variables. Let* $p(\mathcal{Y}) > 0$*. Then*

$$p(\mathcal{X} \mid \mathcal{Y}) = \frac{p(\mathcal{Y} \mid \mathcal{X}) \cdot p(\mathcal{X})}{p(\mathcal{Y})}$$

Thomas Bayes (1701/2–1761)

Independent variables

Definition 4. Two sets $\mathcal{X}, \mathcal{Y}$ of variables are called **independent**, when

$$p(\mathcal{X} = x, \mathcal{Y} = y) = p(\mathcal{X} = x) \cdot p(\mathcal{Y} = y)$$

for all x and y or equivalently

$$p(\mathcal{X} = x | \mathcal{Y} = y) = p(\mathcal{X} = x)$$

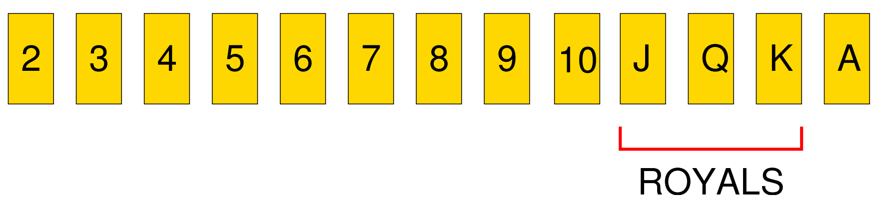
for y with $p(\mathcal{Y} = y) > 0$.

Independent variables / example

Example 4. Let Ω be the cards in an ordinary deck and

- $R =$ true, if a card is royal,
- $T =$ true, if a card is a ten or a jack,
- $S =$ true, if a card is spade.

Cards for a single color:



S	R	T	$p(R,T\|S)$
Y	Y	Y	1/13
		N	2/13
	N	Y	1/13
		N	9/13
N	Y	Y	3/39 = 1/13
		N	6/39 = 2/13
	N	Y	3/39 = 1/13
		N	27/39 = 9/13

R	T	$p(R,T)$
Y	Y	4/52 = 1/13
	N	8/52 = 2/13
N	Y	4/52 = 1/13
	N	36/52 = 9/13

Conditionally independent variables

Definition 5. Let $\mathcal{X}, \mathcal{Y}$, and $\mathcal{Z}$ be sets of variables.

$\mathcal{X}, \mathcal{Y}$ are called **conditionally independent given** $\mathcal{Z}$, when for all events $\mathcal{Z} = z$ with $p(\mathcal{Z} = z) > 0$ all pairs of events $\mathcal{X} = x$ and $\mathcal{Y} = y$ are conditionally independend given $\mathcal{Z} = z$, i.e.

$$p(\mathcal{X} = x, \mathcal{Y} = y, \mathcal{Z} = z) = \frac{p(\mathcal{X} = x, \mathcal{Z} = z) \cdot p(\mathcal{Y} = y, \mathcal{Z} = z)}{p(\mathcal{Z} = z)}$$

for all x, y and z (with $p(\mathcal{Z} = z) > 0$), or equivalently

$$p(\mathcal{X} = x | \mathcal{Y} = y, \mathcal{Z} = z) = p(\mathcal{X} = x | \mathcal{Z} = z)$$

We write $I_p(\mathcal{X}, \mathcal{Y} | \mathcal{Z})$ for the statement, that $\mathcal{X}$ and $\mathcal{Y}$ are conditionally independent given $\mathcal{Z}$.

Conditionally independent variables

Example 5. Assume S (shape), C (color), and L (label) be three random variables that are distributed as shown in figure 4.

We show $I_p(\{L\}, \{S\}|\{C\})$, i.e., that label and shape are conditionally independent given the color.

C	S	L	$p(L\|C,S)$
black	square	1	2/6 = 1/3
		2	4/6 = 2/3
	round	1	1/3
		2	2/3
white	square	1	1/2
		2	1/2
	round	1	1/2
		2	1/2

C	L	$p(L\|C)$
black	1	3/9 = 1/3
	2	6/9 = 2/3
white	1	2/4 = 1/2
	2	2/4 = 1/2

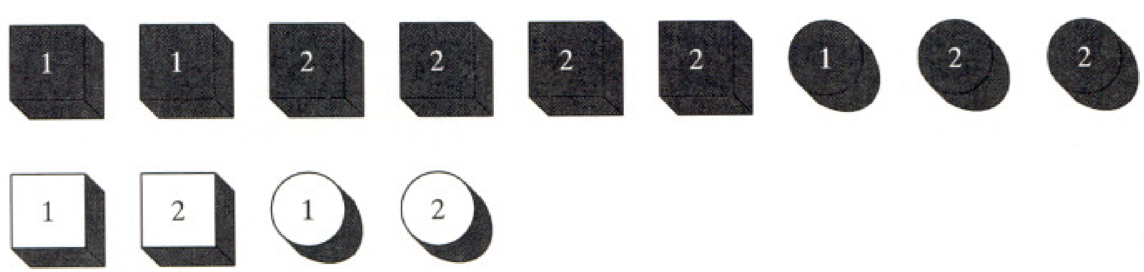


Figure 4: 13 objects with different shape, color, and label [**?**, p. 8].

1. Basic Probability Calculus

2. Separation in undirected graphs

Graphs

Definition 6. Let V be any set and

$$E \subseteq \mathcal{P}^2(V) := \{\{x, y\} \mid x, y \in V\}$$

be a subset of sets of unordered pairs of V. Then $G := (V, E)$ is called an **undirected graph**. The elements of V are called **vertices** or **nodes**, the elements of E **edges**.

Let $e = \{x, y\} \in E$ be an edge, then we call the vertices x, y **incident** to the edge e.

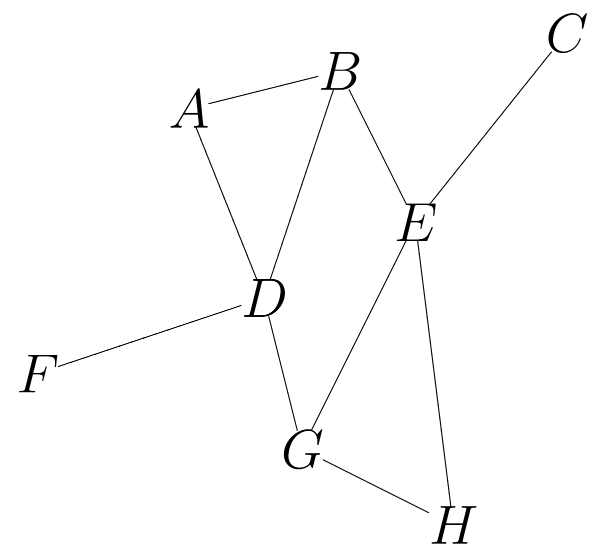


Figure 5: Example graph.

Graphs Representation

The most useful methods of representing graphs are:

- Symbolically as (V, E)
- Pictorially
- Numerically, using certain types of matrices

Characteristics of Undirected Graphs

Definition 7. We call two vertices $x, y \in V$ **adjacent**, or **neighbors** if there is an edge $\{x, y\} \in E$.

The set of all vertices adjacent with a given vertex $x \in V$ is called its **fan** or **boundary**:

$$\text{fan}(x) := \{y \in V \mid \{x, y\} \in E\}$$

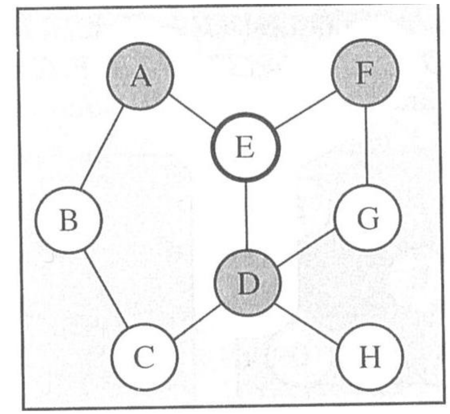


Figure 6: Neighbors of node E [**?**, p. 120].

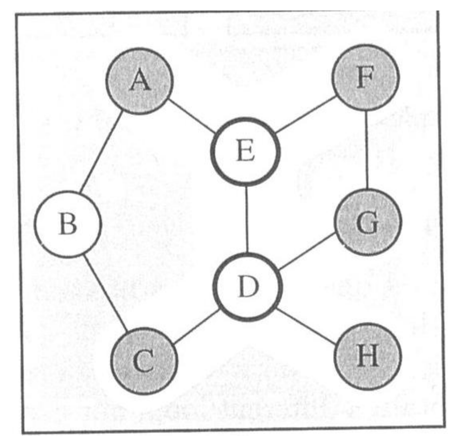


Figure 7: Boundary of the set $\{D, E\}$[**?**, p. 120].

Characteristics of Undirected Graphs

Definition 8. Let $G = (V, E)$ be an undirected graph. An undirected graph $G_X = (X, E_X)$ is called a **subgraph** of G iff $X \subseteq V$ and $E_X = (X \times X) \cap E$

An undirected graph is said to be **complete** iff its set of edges is complete, i.e. iff all possible edges are present, or formally iff $E = V \times V - \{(A, A) | A \in V\}$

A complete subgraph is called a **clique**. A clique is called **maximal** iff it is not a subgraph of a larger clique.

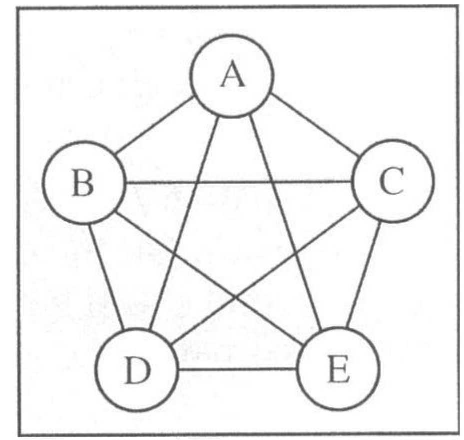


Figure 8: Example of complete graph [**?**, p. 118].

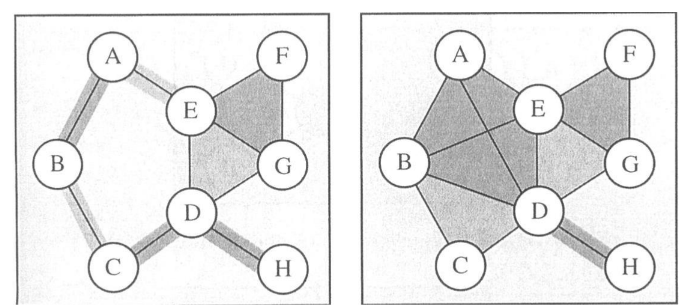


Figure 9: Example of cliques[**?**, p. 119].

Paths on graphs

Definition 9. Let V be a set. We call $V^* := \bigcup_{i\in\mathbb{N}} V^i$ the **set of finite sequences in** V. The length of a sequence $s \in V^*$ is denoted by $|s|$.

Let $G = (V, E)$ be a graph. We call

$$G^* := V^*_{|G} := \{p \in V^* \mid \{p_i, p_{i+1}\} \in E, \; i = 1, \dots, |p| - 1\}$$

the **set of paths on** G.

Any contiguous subsequence of a path $p \in G^*$ is called a **subpath of** p, i.e. any path $(p_i, p_{i+1}, \dots, p_j)$ with $1 \leq i \leq j \leq n$. The subpath $(p_2, p_3, \dots, p_{n-1})$ is called the **interior of** p. A path of length $|p| \geq 2$ is called **proper**.

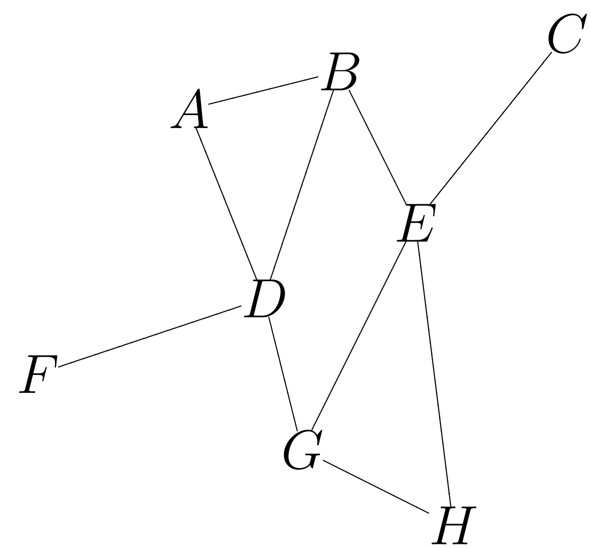


Figure 10: Example graph.

The sequences

$$(A, D, G, H)$$
$$(C, E, B, D)$$
$$(F)$$

are paths on G, but the sequences

$$(A, D, E, C)$$
$$(A, H, C, F)$$

are not.

Types of Undirected Graphs

Definition 10. Let $G = (V, E)$ be an undirected graph. Two distinct nodes $A, B \in V$ are called **connected** in G *iff* there exists at least one path between every two nodes.

A connected undirected graph is said to be a **tree** if for every pair of nodes there exists a unique path.

A connected undirected graph is called **multiply-connected** if it contains at least one pair of nodes that are joined by more than one path.

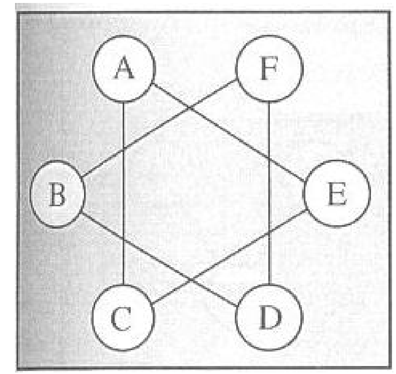


Figure 11: Disconnected graph [**?**, p. 121].

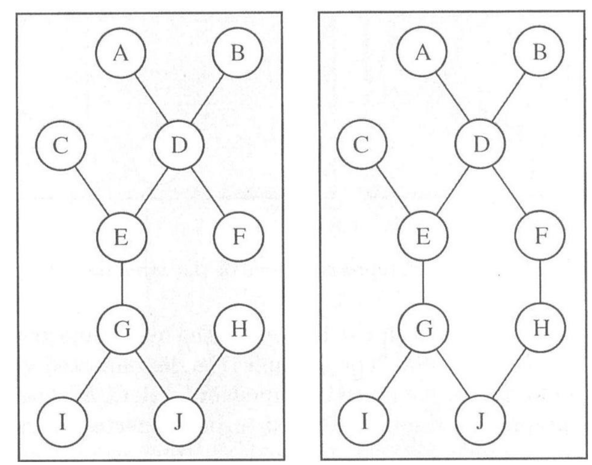


Figure 12: Examples of a tree and a multiply-connected graph [**?**, p. 122].

Separation in graphs (u-separation)

Definition 11. Let $G := (V, E)$ be a graph. Let $Z \subseteq V$ be a subset of vertices. We say, two vertices $x, y \in V$ are **u-separated by** Z **in** G, if every path from x to y contains some vertex of Z ($\forall p \in G^* : p_1 = x, p_{|p|} = y \Rightarrow \exists i \in \{1, \ldots, n\} : p_i \in Z$).

Let $X, Y, Z \subseteq V$ be three disjoint subsets of vertices. We say, the vertices X and Y are **u-separated by** Z **in** G, if every path from any vertex from X to any vertex from Y is separated by Z, i.e., contains some vertex of Z.

We write $I_G(X, Y|Z)$ for the statement, that X and Y are u-separated by Z in G.

I_G is called **u-separation relation in** G.

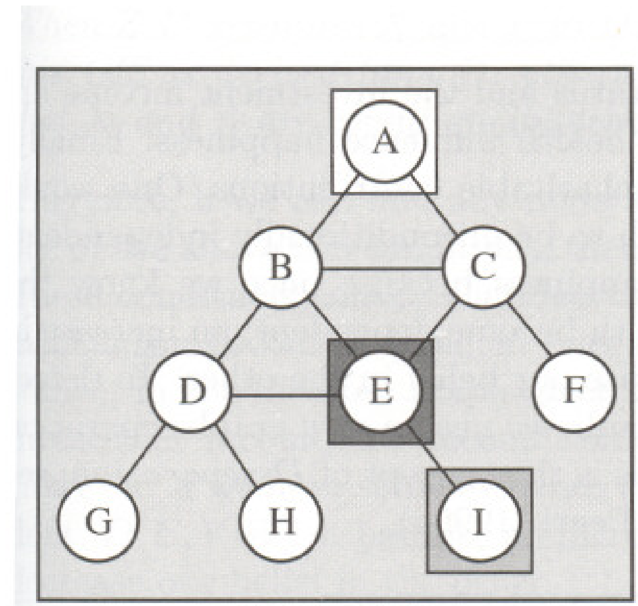


Figure 13: Example for u-separation [**?**, p. 179].

Separation in graphs (u-separation)

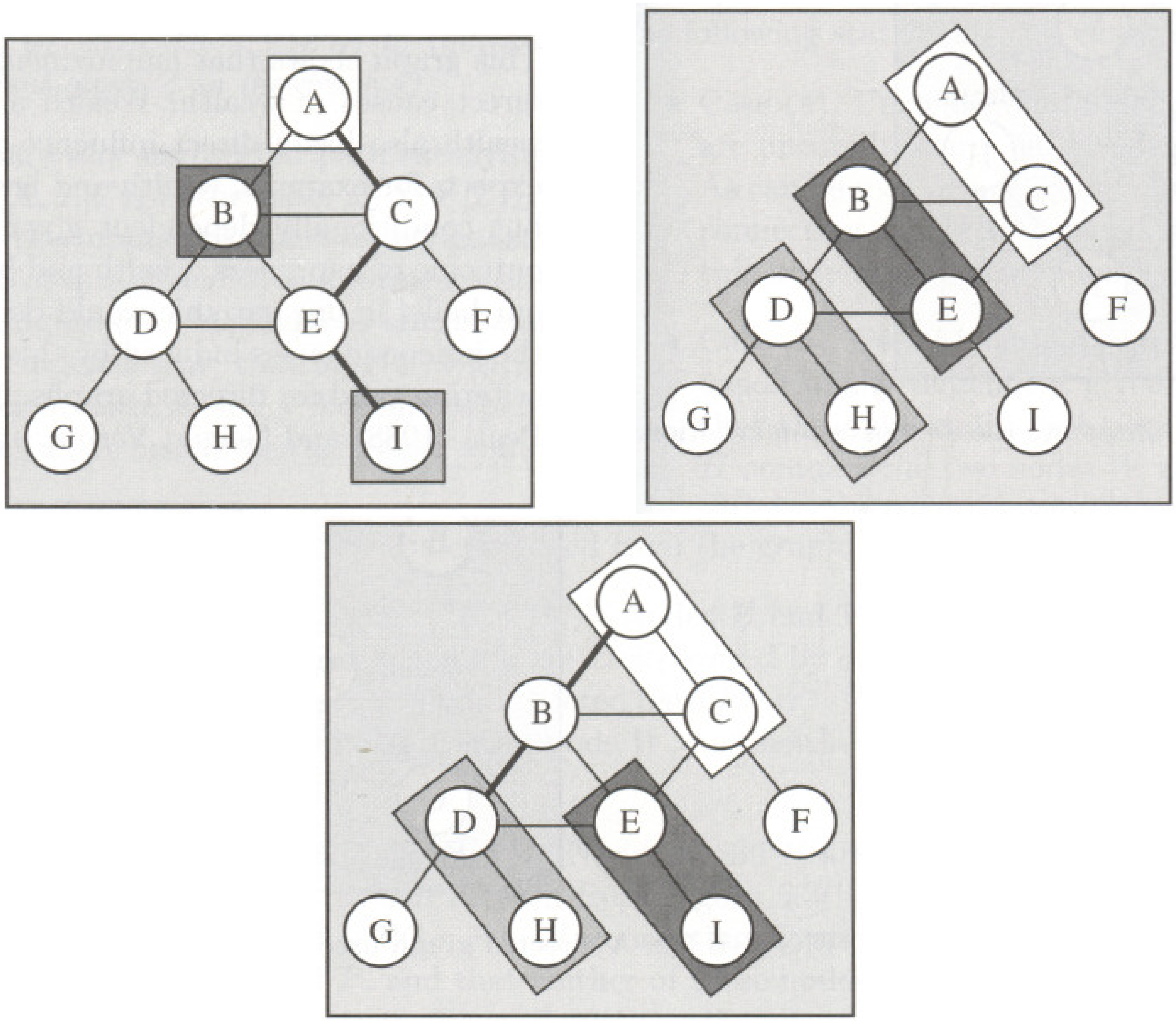


Figure 14: More examples for u-separation [**?**, p. 179].

Properties of ternary relations

Definition 12. Let V be any set and I a ternary relation on $\mathcal{P}(V)$, i.e., $I \subseteq (\mathcal{P}(V))^3$.

I is called **symmetric**, if

$$I(X,Y|Z) \Rightarrow I(Y,X|Z)$$

I is called **decomposable**, if

$$I(X,Y \cup W|Z) \Rightarrow I(X,Y|Z) \text{ and } I(X,W|Z)$$

I is called **composable**, if

$$I(X,Y|Z) \text{ and } I(X,W|Z) \Rightarrow I(X,Y \cup W|Z)$$

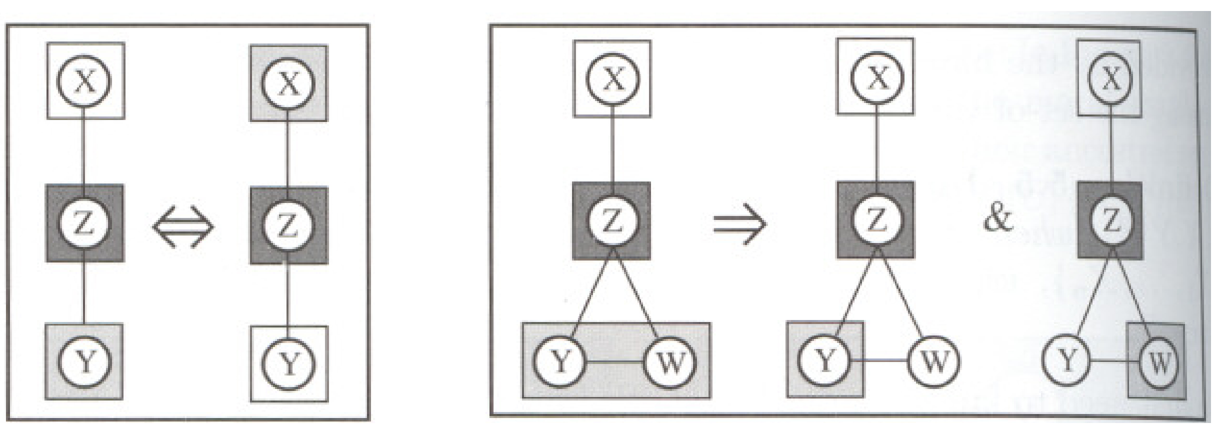


Figure 15: Examples for a) symmetry and b) decomposition [?, p. 186].

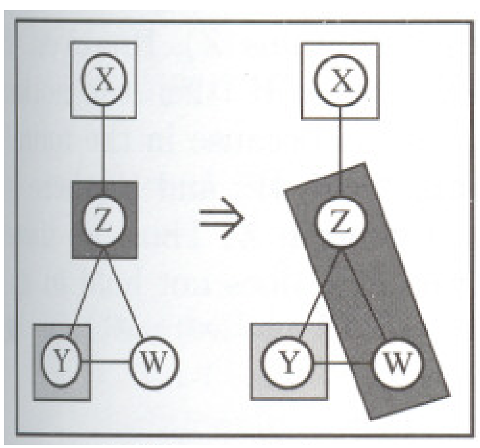
Properties of ternary relations

Definition 13. I is called **strongly unionable**, if

$$I(X, Y|Z) \Rightarrow I(X, Y|Z \cup W)$$

I is called **weakly unionable**, if

$$I(X, Y \cup W|Z) \Rightarrow I(X, W|Z \cup Y) \text{ and } I(X, Y|Z \cup W)$$



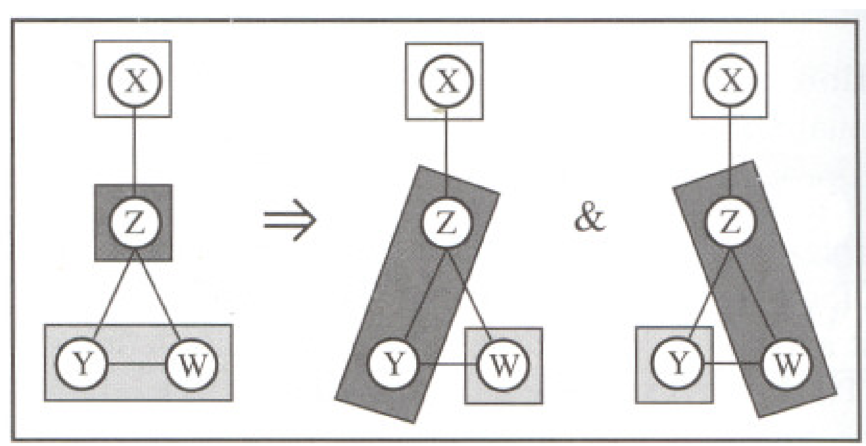


Figure 16: Examples for a) strong union and b) weak union [**?**, p. 186,189].

Properties of ternary relations

Definition 14. I is called **contractable**, if

$$I(X, W|Z \cup Y) \text{ and } I(X, Y|Z) \Rightarrow I(X, Y \cup W|Z)$$

I is called **intersectable**, if

$$I(X, W|Z \cup Y) \text{ and } I(X, Y|Z \cup W) \Rightarrow I(X, Y \cup W|Z)$$

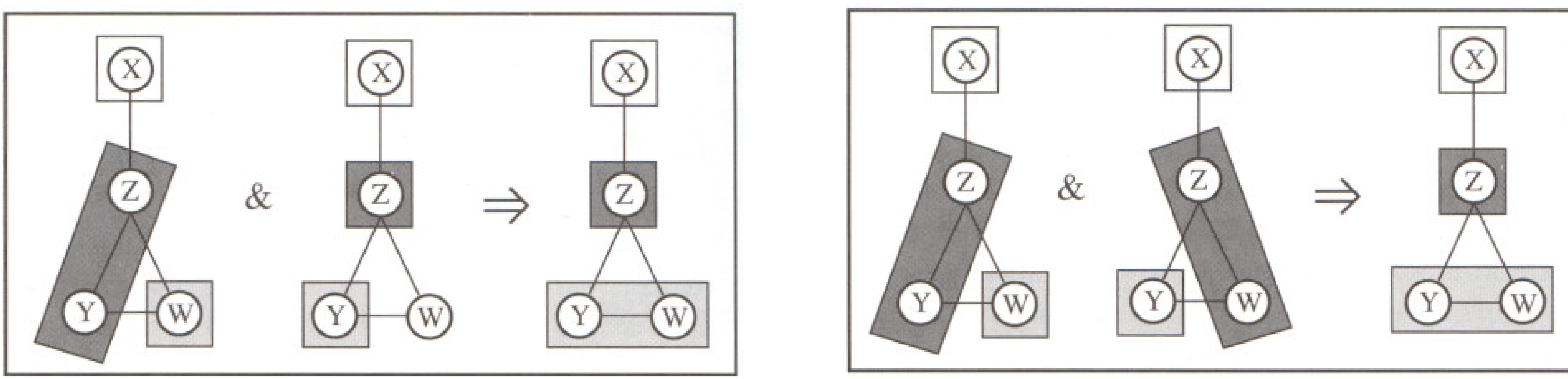


Figure 17: Examples for a) contraction and b) intersection [**?**, p. 186].

Properties of ternary relations

Definition 15. I is called **strongly transitive**, if

$$I(X, Y|Z) \Rightarrow I(X, \{v\}|Z) \text{ or } I(\{v\}, Y|Z) \quad \forall v \in V \setminus Z$$

I is called **weakly transitive**, if

$$I(X, Y|Z) \text{ and } I(X, Y|Z \cup \{v\}) \Rightarrow I(X, \{v\}|Z) \text{ or } I(\{v\}, Y|Z) \quad \forall v \in V \setminus Z$$

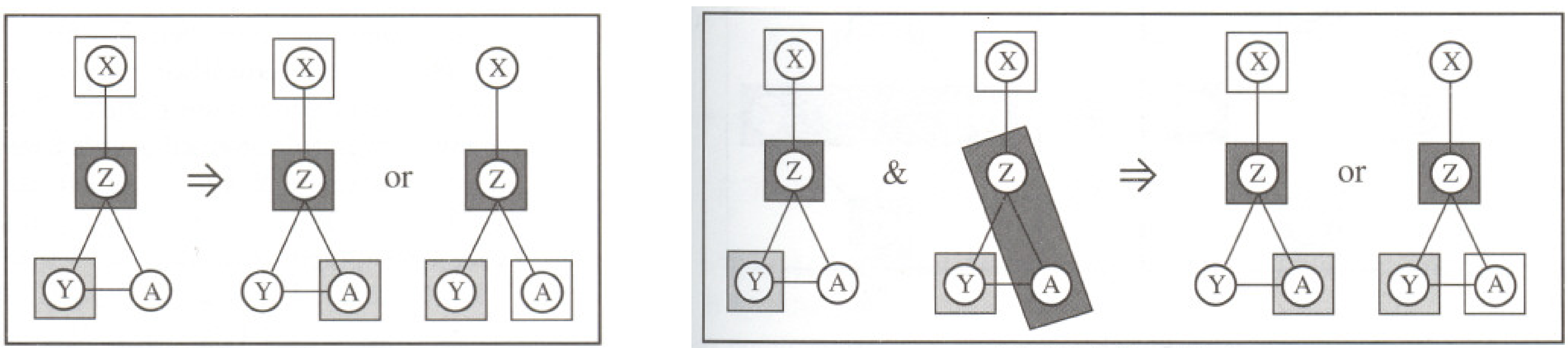


Figure 18: Examples for a) strong transitivity and b) weak transitivity. [**?**, p. 189]

Properties of ternary relations

Definition 16. I is called **chordal**, if

$$I(\{a\},\{c\}|\{b,d\}) \text{ and } I(\{b\},\{d\}|\{a,c\}) \Rightarrow I(\{a\},\{c\}|\{b\}) \text{ or } I(\{a\},\{c\}|\{d\})$$

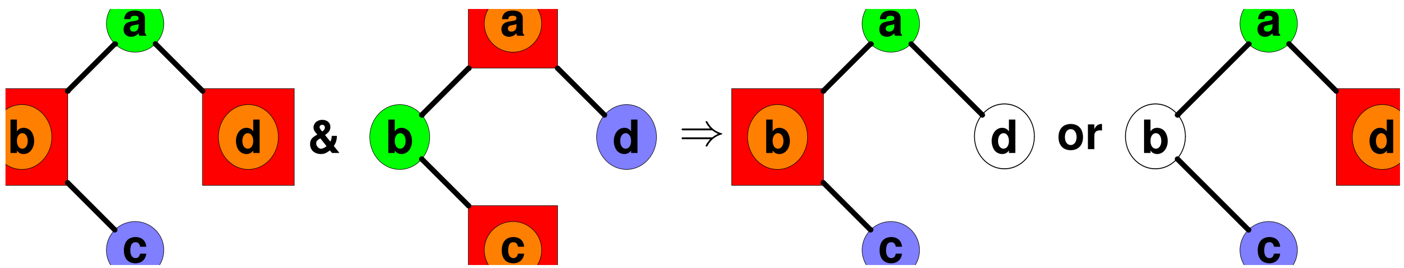


Figure 19: Example for chordality.

Properties of u-separation / no chordality

For u-separation the chordality property does not hold (in general).

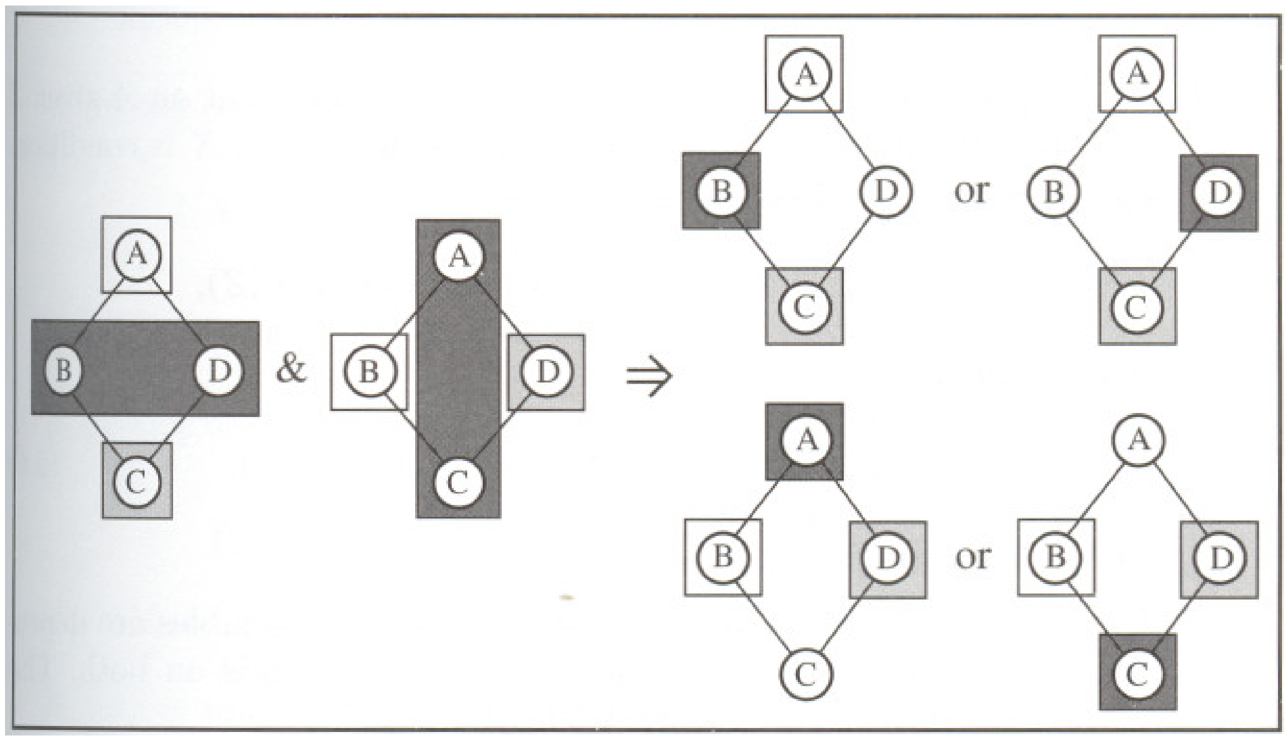


Figure 20: Counterexample for chordality in undirected graphs (u-separation) [**?**, p. 189].

Properties of u-separation

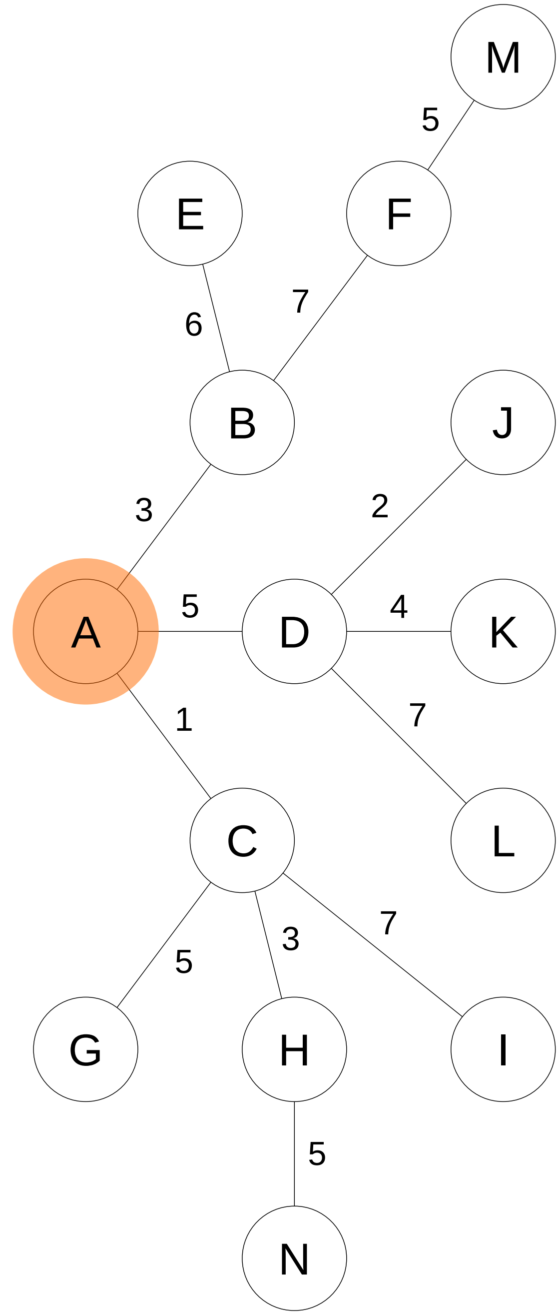
relation	symmetry	decomposition	composition	strong union	weak union	contraction	intersection	strong transitivity	weak transitivity	chordality
u-separation	+	+	+	+	+	+	+	+	+	−

Breadth-First Search

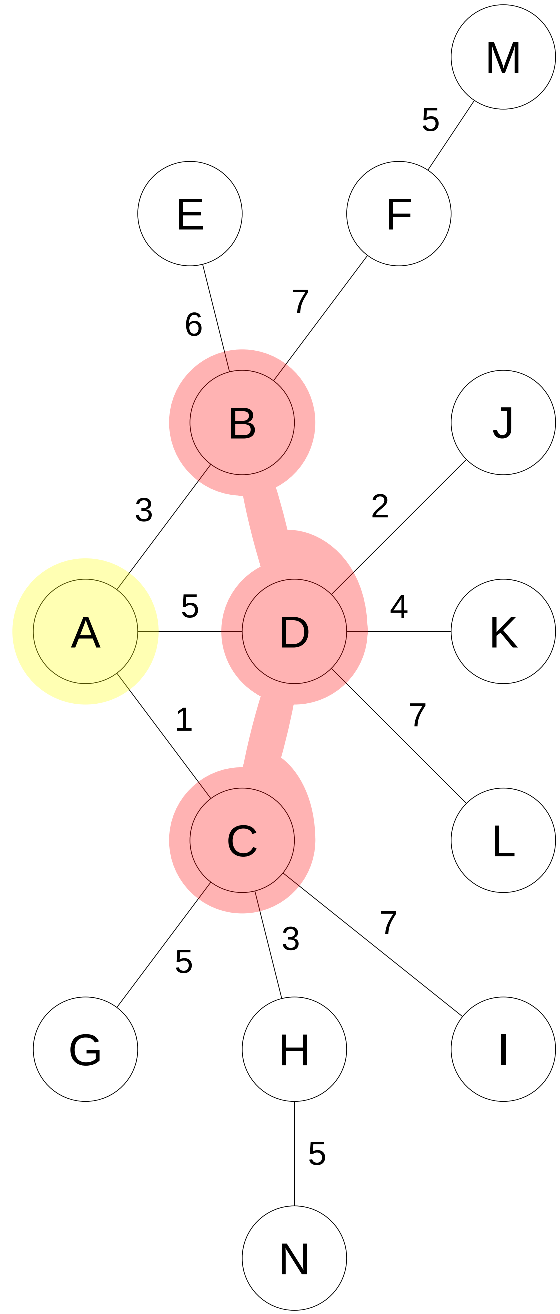
Idea:

- start with initial node as border.
- iteratively replace border by all nodes reachable from the old border.

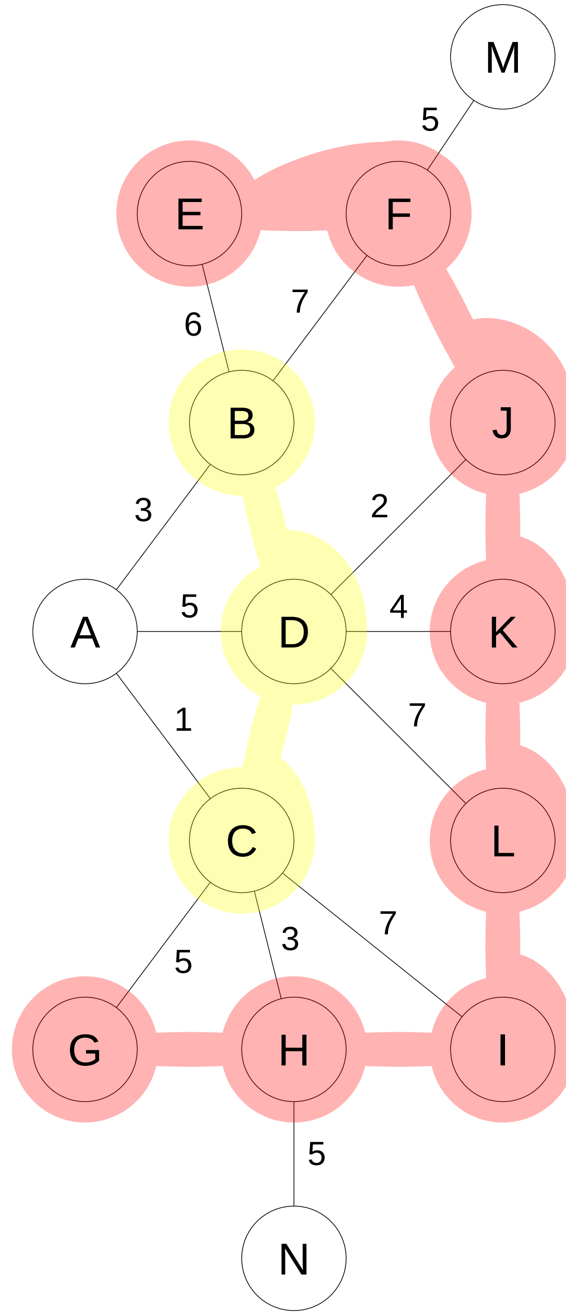
Breadth-First Search / Example



Breadth-First Search / Example



Breadth-First Search / Example



Checking u-separation

To test, if for a given graph $G = (V, E)$ two given sets $X, Y \subseteq V$ of vertices are u-separated by a third given set $Z \subseteq V$ of vertices, we may use standard breadth-first search to compute all vertices that can be reached from X (see, e.g., [**?**], [**?**]).

```
breadth-first search(G, X) :
border := X
reached := ∅
while border ≠ ∅ do
        reached := reached ∪ border
        border := fan_G(border) \ reached
od
return reached
```

Figure 24: Breadth-first search algorithm for enumerating all vertices reachable from X.

For checking u-separation we have to tweak the algorithm

1. not to add vertices from Z to the border and

2. to stop if a vertex of Y has been reached.

```
check-u-separation(G, X, Y, Z) :
border := X
reached := ∅
while border ≠ ∅ do
        reached := reached ∪ border
        border := fan_G(border) \ reached \ Z
        if border ∩ Y ≠ ∅
           return false
        fi
od
return true
```

Figure 25: Breadth-first search algorithm for checking u-separation of X and Y by Z.