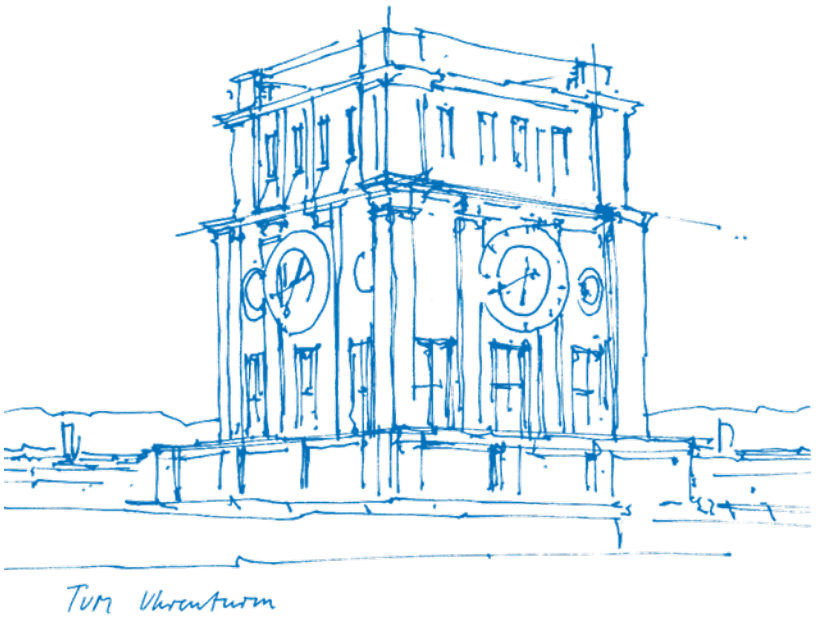
Computer Vision & Artificial Intelligence
Department of Informatics
Technical University of Munich

II : Graphical Model Representation

Tao Wu, Yuesong Shen, Zhenzhang Ye

Computer Vision & Artificial Intelligence
Technical University of Munich



Outline of the Chapter

- Bayesian network (directed graphical model).
- Markov random field (undirected graphical model).
- Independence assumption, representation power, parameterization, etc.

Bayesian Network

Bayesian Network

A **Bayesian network** (BN) is a *directed acyclic graph* $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ together with:

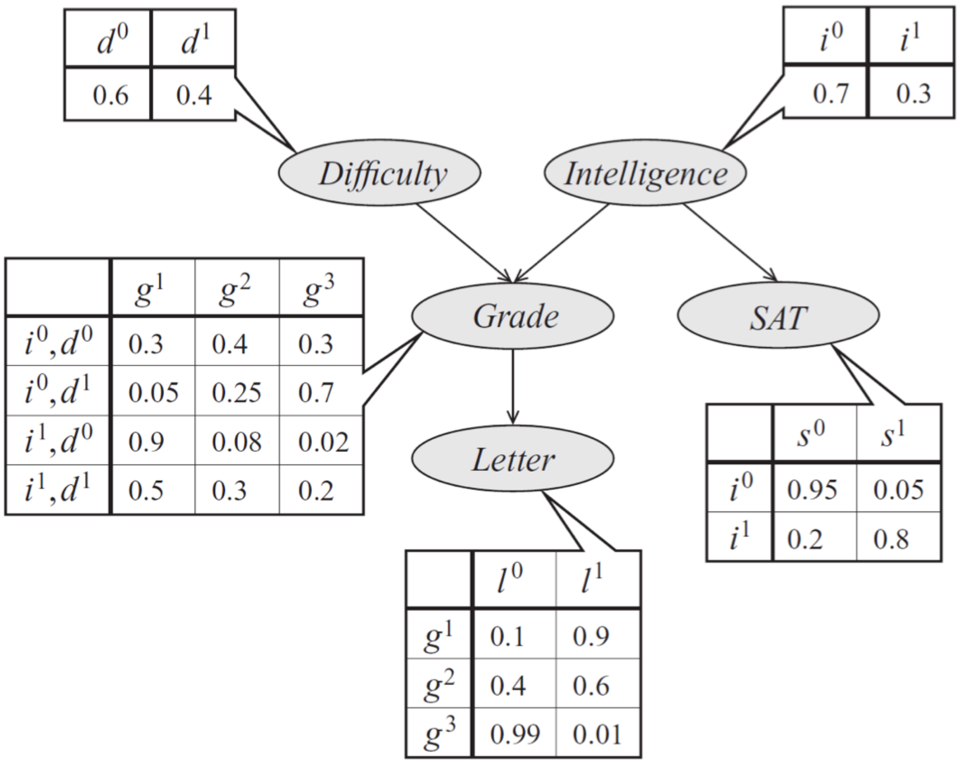
- Random variables $X = (X_i)_{i \in \mathcal{V}}$ over $\mathcal{V}$;
- A (joint probability) distribution P factorized as a product of conditional probability distributions (CPDs):

$$p(x) = \prod_{i \in \mathcal{V}} p(x_i | (x_j)_{j \in \mathrm{Pa}_{\mathcal{G}}(i)}),$$

where $\mathrm{Pa}_{\mathcal{G}}(i) = \{j \in \mathcal{V} : (j, i) \in \mathcal{E}\}$ consists of parents of i in $\mathcal{G}$.

Example "Student"

$$P(D, I, G, S, L) = P(D)P(I)P(G|D, I)P(S|I)P(L|G).$$



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

	g^1	g^2	g^3
i^0, d^0	0.3	0.4	0.3
i^0, d^1	0.05	0.25	0.7
i^1, d^0	0.9	0.08	0.02
i^1, d^1	0.5	0.3	0.2

	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

	l^0	l^1
g^1	0.1	0.9
g^2	0.4	0.6
g^3	0.99	0.01

Figure: Bayesian network represented in probability tables.

Model Complexity

Consider BN representation for RVs $(X_i)_{i=1}^n$.

- If each RV X_i takes at most d outcomes and has at most k parents, then representation of
$$p(x_i | (x_j)_{j \in \mathrm{Pa}_{\mathcal{G}}(i)})$$
requires $O(d^{k+1})$ free parameters.

- Since the joint distribution for $(X_i)_{i=1}^n$ is a product of n CPDs, the overall model complexity for BN is $O(nd^{k+1})$.

- Compared to a naive representation for the joint distribution which requires $O(d^n)$ parameters (typically $n \gg k$).

The reduction of complexity is due to the underlying independence assumptions.

Independence in BN

- For a distribution P for RVs (X_i), we denote by $\mathcal{I}(P)$ the set of all **independence assumptions (assertions)** that hold in P:

$$\mathcal{I}(P) = \big\{(X_i \perp X_j \mid X_k)\big\}.$$

Recall conditional independence: $X_i \perp X_j \mid X_k$ iff

$$p(x_i, x_j | x_k) = p(x_i | x_k) p(x_j | x_k).$$

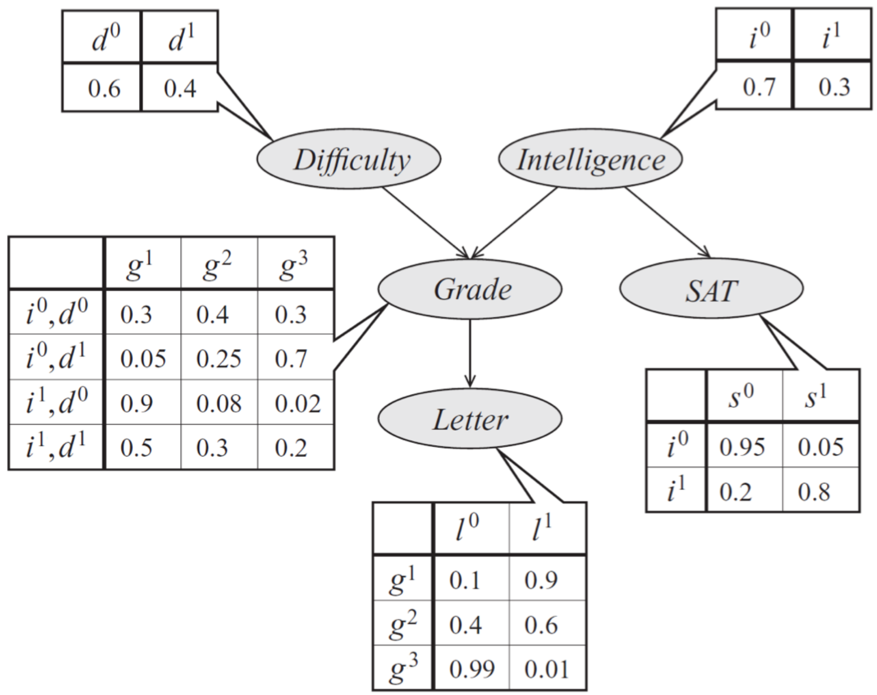
- BN $\mathcal{G}$ implies **local independencies**:

$$\mathcal{I}_\ell(\mathcal{G}) = \Big\{\Big(X_i \perp (X_j)_{j \in \mathrm{NonDes}_\mathcal{G}(i)} \mid (X_k)_{k \in \mathrm{Pa}_\mathcal{G}(i)}\Big)\Big\},$$

where $\mathrm{NonDes}_\mathcal{G}(i)$ denotes the non-descendants of i (including i itself) in $\mathcal{G}$.

Example "Student"

$$\mathcal{I}_\ell(\mathcal{G}) = \left\{ \left(X_i \perp (X_j)_{j \in \mathrm{NonDes}_{\mathcal{G}}(i)} \,\middle|\, (X_k)_{k \in \mathrm{Pa}_{\mathcal{G}}(i)} \right) \right\}.$$



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

	g^1	g^2	g^3
i^0, d^0	0.3	0.4	0.3
i^0, d^1	0.05	0.25	0.7
i^1, d^0	0.9	0.08	0.02
i^1, d^1	0.5	0.3	0.2

	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

	l^0	l^1
g^1	0.1	0.9
g^2	0.4	0.6
g^3	0.99	0.01

In the above example, we have: $(L \perp \{I, D, S\} \mid G)$, $(G \perp S \mid \{I, D\}) \in \mathcal{I}_\ell(\mathcal{G})$.

Beyond Local Independence

- Does $\mathcal{G}$ encode other independence assertions besides $\mathcal{I}_\ell(\mathcal{G})$? (Yes.)
- How to identify whether a specific independence assertion holds in $\mathcal{G}$?

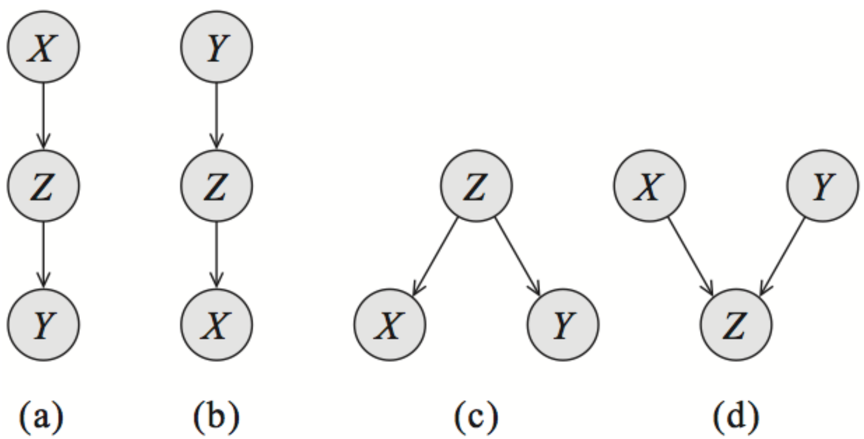
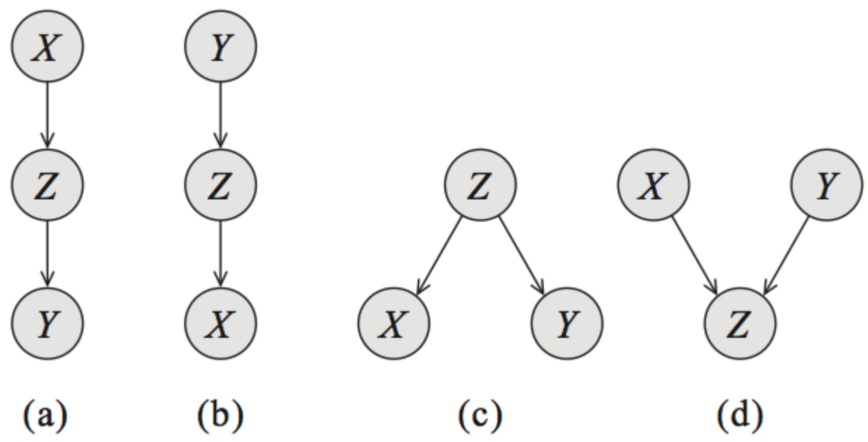


Figure: Two-edge trails from X to Y via Z. (d) is called the **v-structure**.

In the above figure, information/dependence flows from X to Y if the **trail** $X \leftrightarrow Z \leftrightarrow Y$ is **active**. This is the case if:

- In (a)–(c), Z is unobserved. (In contrast, $X \perp Y \mid Z$.)
- In (d), Z or one of its descendants is observed. (In contrast, $X \perp Y$ o.w.)

Active Trail



Let $X_1 \leftrightarrow X_2 \leftrightarrow \ldots \leftrightarrow X_n$ be a trail in a BN $\mathcal{G}$, and Z be a set of observed nodes (RVs). The trail is **active** given Z if

- Whenever there is a v-structure (case (d)) in the trail $X_{i-1} \leftrightarrow X_i \leftrightarrow X_{i+1}$, then X_i or one of its descendants are in Z.
- No other node along the trail is in Z.

Intuitively, information/dependence flows from X_1 to X_n (and vice versa) through the active trail $X_1 \leftrightarrow X_2 \leftrightarrow \ldots \leftrightarrow X_n$.

D-separation, Global Independence

Let X, Y, Z be three sets of nodes in a BN $\mathcal{G}$. If there is no active trail between any node in X and Y given Z, we say X and Y are **d-separated** given Z.

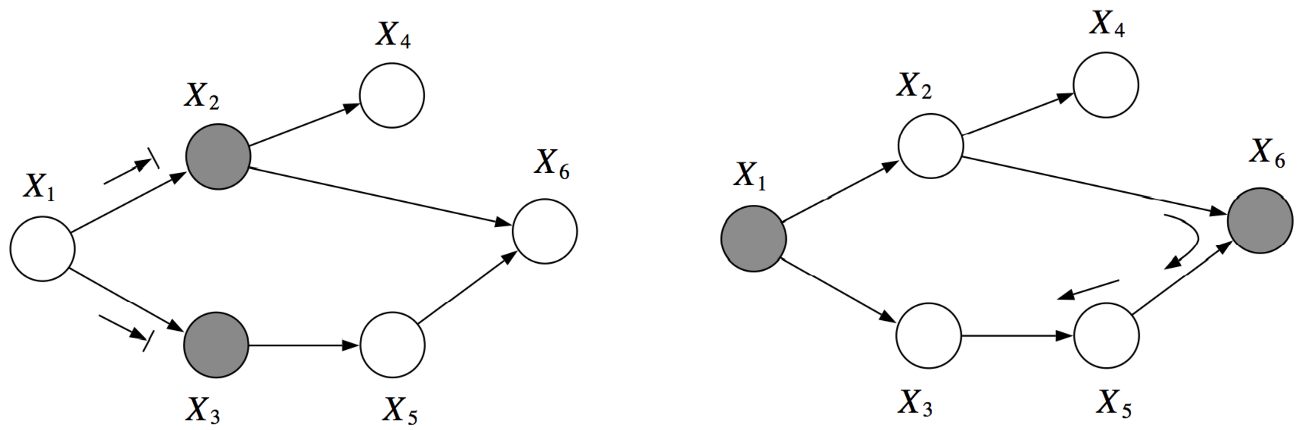


Figure: (left) X_1 and X_6 are d-sep. given $\{X_2, X_3\}$; (right) X_2 and X_3 are not d-sep. given $\{X_1, X_6\}$.

We denote by $\mathcal{I}(\mathcal{G})$ the set of **global Markov independencies**:

$$\mathcal{I}(\mathcal{G}) = \{(X \perp Y \mid Z) : X \text{ and } Y \text{ are d-separated given } Z\}.$$

Facts about D-separation

F1. (**soundness**) If a distribution P factorizes according to $\mathcal{G}$, then $\mathcal{I}(\mathcal{G}) \subset \mathcal{I}(P)$. In this case, we call $\mathcal{G}$ an **I-map** for P.

F2. (**sharpness**) If nodes X and Y are not d-separated given Z in $\mathcal{G}$, then X and Y are dependent given Z in some distribution P that factorizes over $\mathcal{G}$.

F3. (**completeness**) When a distribution P factorizes according to $\mathcal{G}$, $\mathcal{I}(\mathcal{G}) = \mathcal{I}(P)$ does not necessarily holds. Obviously, one can add superfluous edges to $\mathcal{G}$ s.t. $\mathcal{I}(\mathcal{G}) \subsetneq \mathcal{I}(P)$.

$p(b \mid a)$	b_0	b_1
a_0	0.4	0.6
a_1	0.4	0.6

Figure: A possible I-map for P is $A \to B$, but $\emptyset = \mathcal{I}(\mathcal{G}) \subsetneq \mathcal{I}(P)$. Such cases almost surely do not happen.

Remark: For almost all P (except for a set of measure zero in the space of CPD parameterizations) for which $\mathcal{G}$ is an I-map, we have $\mathcal{I}(\mathcal{G}) = \mathcal{I}(P)$.

I-equivalence

- Two BNs $\mathcal{G}_1$ and $\mathcal{G}_2$ are said to be **I-equivalent** if $\mathcal{I}(\mathcal{G}_1) = \mathcal{I}(\mathcal{G}_2)$.
- The **skeleton** of a BN $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is an undirected graph $(\mathcal{V}, \mathcal{E}')$ such that $\{X, Y\} \in \mathcal{E}'$ whenever $(X, Y) \in \mathcal{E}$.
- Fact: If two BNs have the same skeleton and the same set of v-structures, then they are I-equivalent.

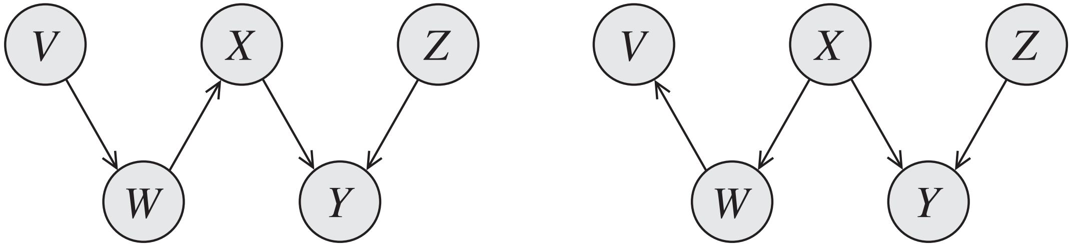


Figure: Example of two I-equivalent BNs.

Perfect Map and Counterexamples

- We say a BN $\mathcal{G}$ is a **perfect map** for a distribution P if $\mathcal{I}(\mathcal{G}) = \mathcal{I}(P)$.
- Certain independencies cannot be expressed perfectly by BN.

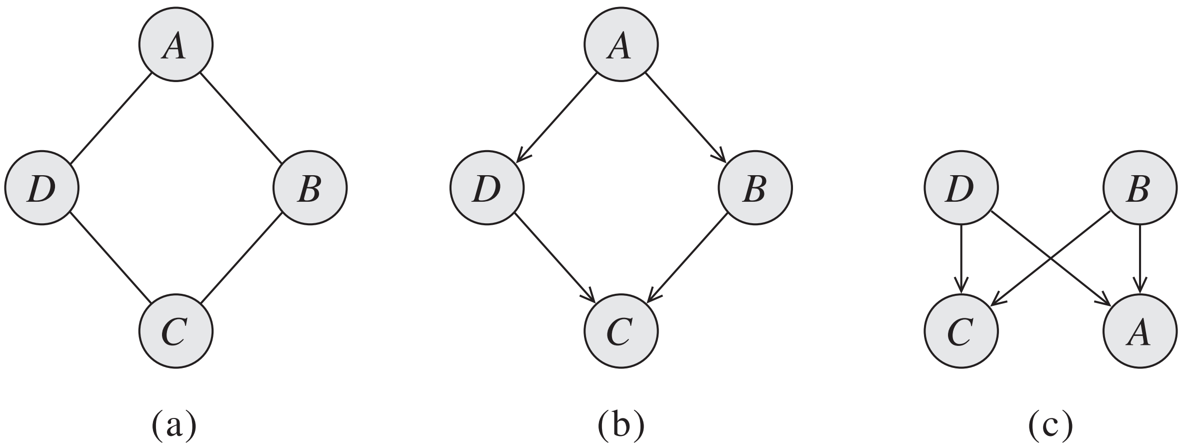


Figure: A counterexample where a perfect map does not exist.

(a) Desired independence assertions: $A \perp C \mid \{B, D\}$, $B \perp D \mid \{A, C\}$.

(b) In this BN: $(A \perp C \mid \{B, D\}) \in \mathcal{I}(\mathcal{G})$, but $(B \perp D \mid \{A, C\}) \notin \mathcal{I}(\mathcal{G})$.

(c) Again, $(A \perp C \mid \{B, D\}) \in \mathcal{I}(\mathcal{G})$, but $(B \perp D \mid \{A, C\}) \notin \mathcal{I}(\mathcal{G})$.

Topics which are not covered here ...

- Algorithm for detecting d-separation in a BN $\mathcal{G}$.
- Algorithm for finding minimal I-map $\mathcal{G}$ for a given distribution P.
- Algorithm for finding perfect map $\mathcal{G}$ (if exists) for a given distribution P.
- Further reading: Koller & Friedman, Chapter 3.