# Directed Graphical Models: Bayesian Networks

*Probabilistic Graphical Models*

Tavassolipour

Slides from Dr. Soleymani's PGM course, Sharif University of Technology

# Basics

- Multivariate distributions with large number of variables

- Independency assumptions are useful
  - Independence and conditional independence relationships simplify representation and alleviate inference complexities

# Conditional and marginal independence

- $X$ and $Y$ are **conditionally independent** given $Z$ if:

$X \perp Y|Z$

$$P(X,Y|Z) = P(X|Z)P(Y|Z) \longleftrightarrow \begin{array}{l} P(X|Y,Z) = P(X|Z) \\ P(Y|X,Z) = P(Y|Z) \end{array}$$

- $X$ and $Y$ are **marginal independent** if:

$X \perp Y|\emptyset$

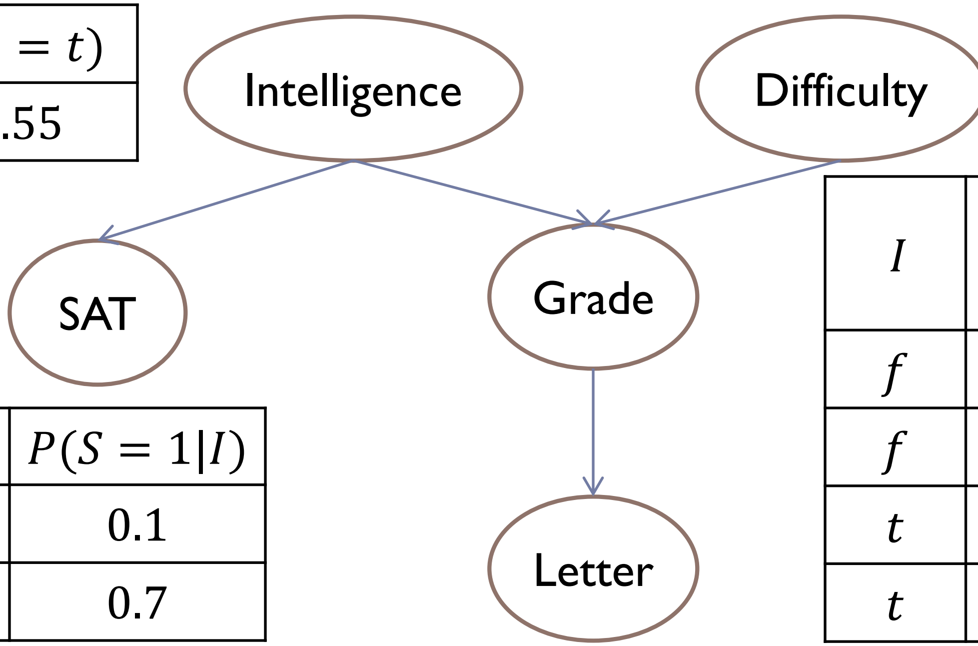
$$P(X,Y) = P(X)P(Y) \longleftrightarrow \begin{array}{l} P(X|Y) = P(X) \\ P(Y|X) = P(Y) \end{array}$$

# Example

- Random variables:
  - Course difficulty
  - Quality of recommendation letter
  - Intelligence
  - Grade
  - SAT score

# Example

| $P(I = t)$ |
|---|
| 0.55 |




| $P(D = t)$ |
|---|
| 0.65 |

| $I$ | $D$ | $P(G \mid I, D)$ | | |
|---|---|---|---|---|
| | | $G = 1$ | $G = 2$ | $G = 3$ |
| $f$ | $f$ | 0.3 | 0.4 | 0.3 |
| $f$ | $t$ | 0.05 | 0.25 | 0.7 |
| $t$ | $f$ | 0.9 | 0.08 | 0.02 |
| $t$ | $t$ | 0.5 | 0.3 | 0.2 |

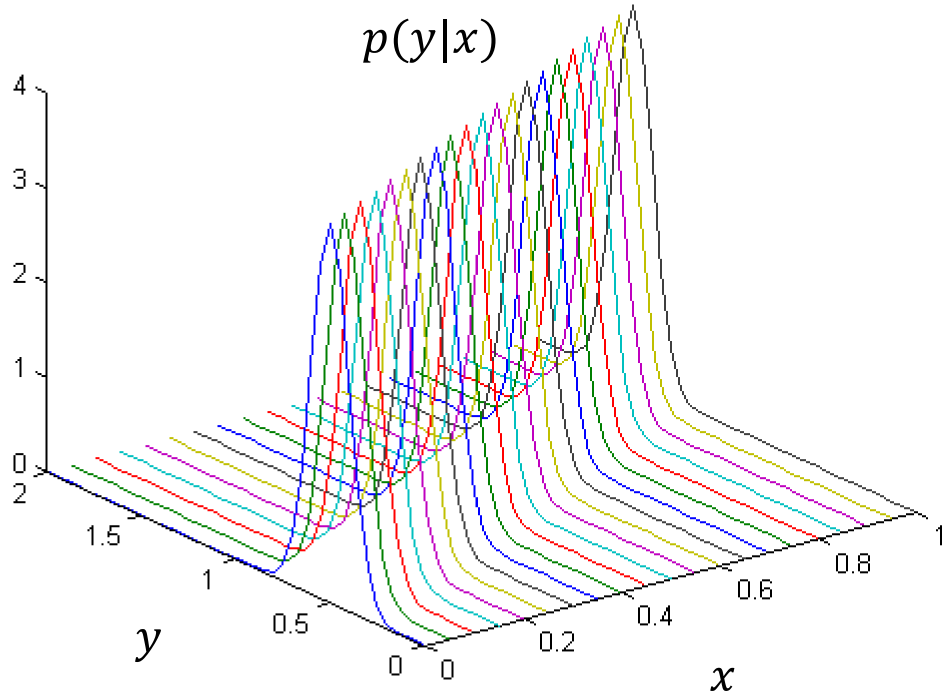
| $I$ | $P(S = 1 \mid I)$ |
|---|---|
| $f$ | 0.1 |
| $t$ | 0.7 |

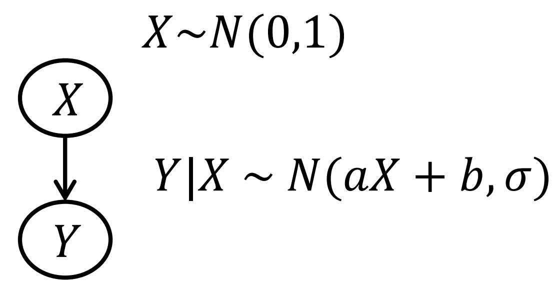
| $G$ | $P(L = t \mid G)$ |
|---|---|
| 1 | 0.9 |
| 2 | 0.5 |
| 3 | 0.05 |

# Continuous example: Linear regression

# Continuous variables example

- Linear Gaussian



# Missing edges

- Chain Rule

- Missing edges imply conditional independencies.

- The more the sparse DAG, the more conditional independencies.

# Compact representation

- A BN for a Boolean variables with <u>*k* Boolean parents</u>

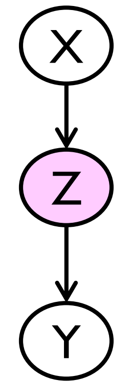
# Factorization & independence

- Let $G$ be a graph over $X_1, \dots, X_n$, distribution $P$ **factorizes** over $G$ if:

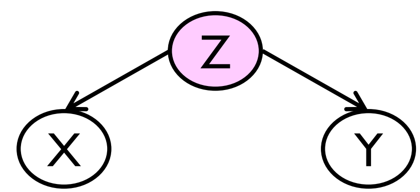
$$P(X_1, \dots, X_n) = \prod_{i=1}^{n} P\left(X_i \mid Pa(X_i)\right)$$

# Basic structures

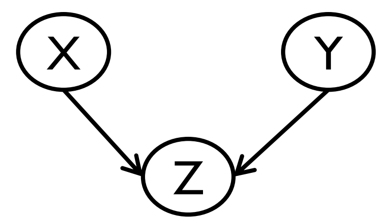
- $X \perp Y | Z$




- $X \perp Y | Z$



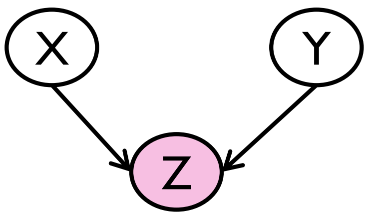

- $X \perp Y$




Explaining away

# Explaining away

- When we condition on $Z$ are $X$ and $Y$ are independent?




$$P(X, Y, Z) = P(X)P(Y)P(Z|X, Y)$$

- $X$ and $Y$ are marginally independent but given $Z$ they are conditionally dependent

- This is called explaining away

- Two coins example

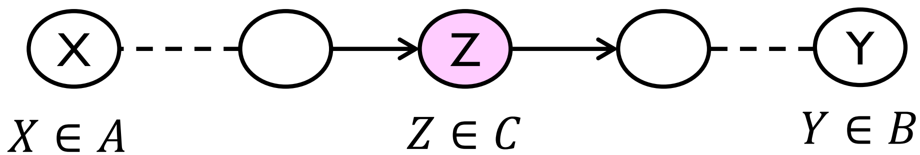
# D-separation

- Let $A, B, C$ denote three disjoint sets of nodes, $A$ is **d-separated** from $B$ by $C$ then $\boldsymbol{A \perp B | C}$

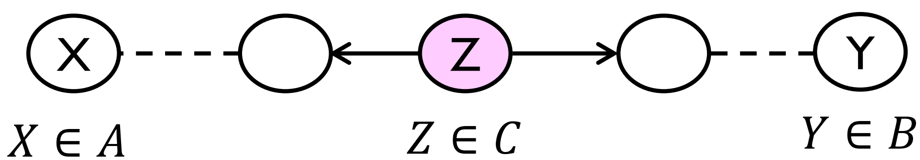
- $A$ is **d-separated** from $B$ by $C$ if all undirected paths between $A$ and $B$ are **blocked** by $C$

# Undirected path blocking

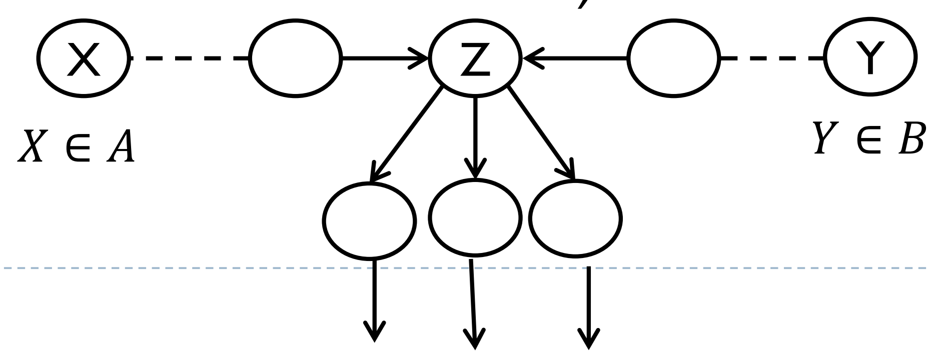
- Head-to-tail at a node $Z \in C$




- Tail-to-tail at a node $Z \in C$



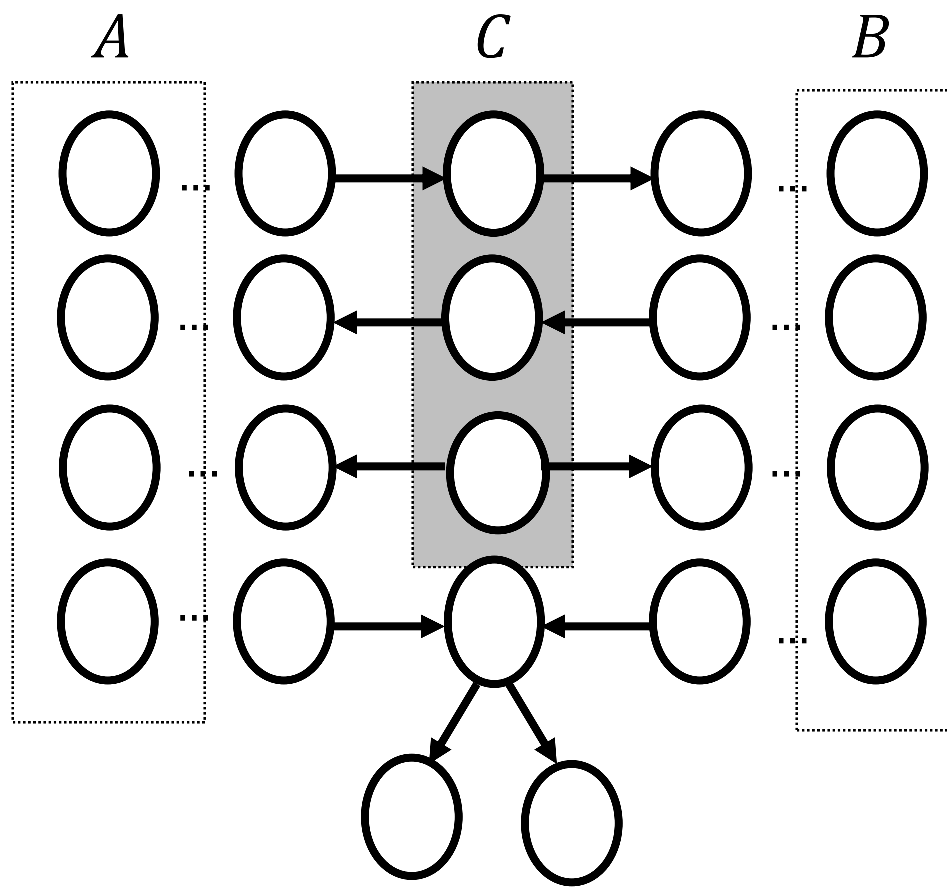

- Head-to-head (i.e., v-structure) at a node $Z$ ($Z \notin C$ & none of its descendants are in $C$)



# Undirected path blocking

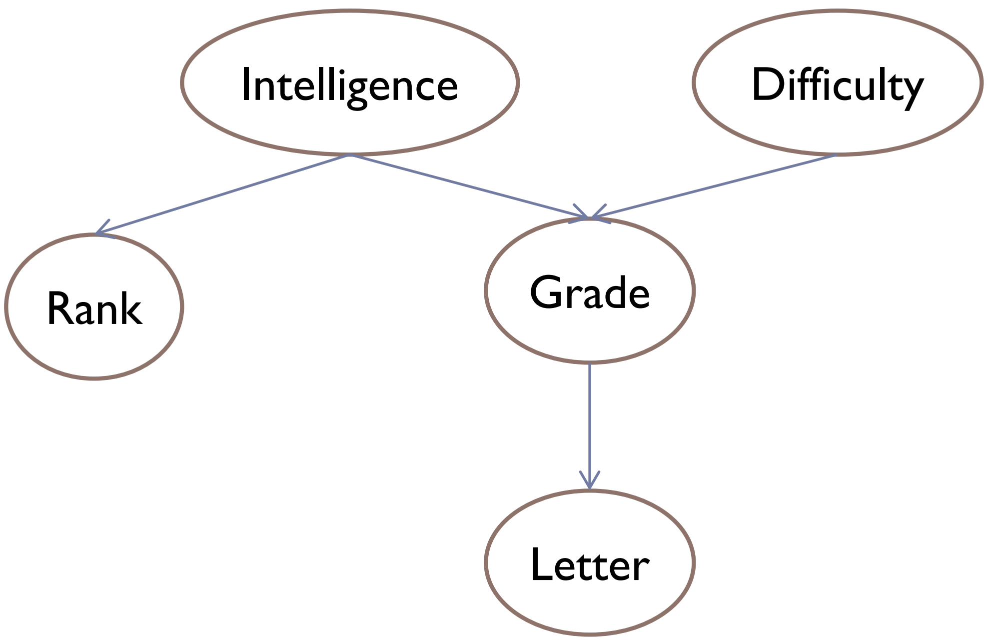
In all trails (undirected paths) between A and B:

- A node in the path is in $C$ and the path at the node do not meet head-to-head.
- Or a head-to-head node in the path, and neither the node, nor any of its descendants, is in C




$A \perp B | C$

# D-separation: example




$R \perp G | I$

$R \perp D | I$

$R \perp D | G$ ✗

$R \perp D | L$ ✗

$R \perp L | G$

$D \perp L | G$

# Markov Blanket in Bayesian Network

- A variable is <u>conditionally independent of all other variables given its Markov blanket</u>

- Markov blanket of a node:
  - All parents
  - Children
  - Co-parents of children

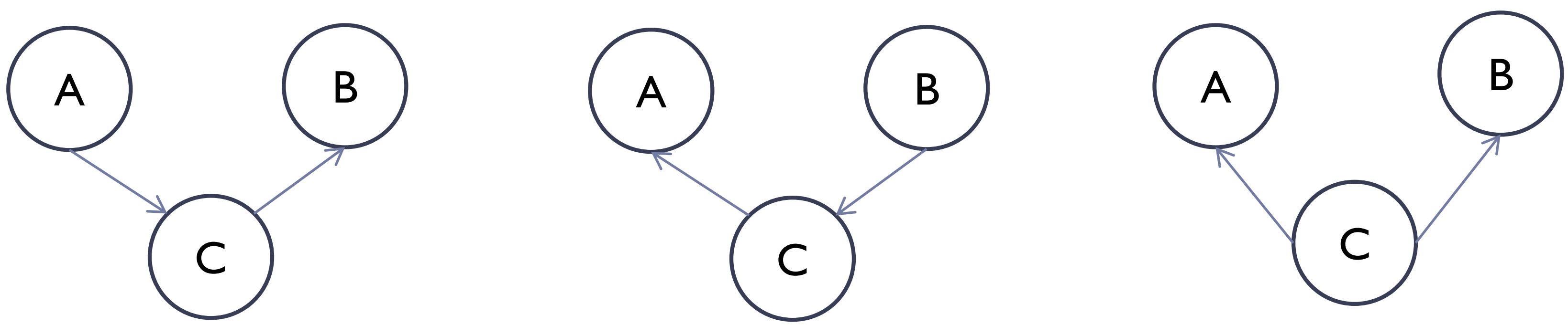
# D-Separation: soundness & completeness

- **Soundness**: Any conditional independence properties that we can derive from $G$ should hold for the probability distribution that factorize over $G$
  - **Theorem**: If $P$ factorizes over $G$, and d-sep$_G(\boldsymbol{X}, \boldsymbol{Y}|\boldsymbol{Z})$ then $P$ satisfies $\boldsymbol{X} \perp \boldsymbol{Y}|\boldsymbol{Z}$

- Weak **completeness**:
  - For almost all distributions $P$ that factorize over $G$, if $\boldsymbol{X} \perp \boldsymbol{Y}|\boldsymbol{Z}$ in $P$ then $\boldsymbol{X}$ and $\boldsymbol{Y}$ are d-separated given $\boldsymbol{Z}$ in the graph $G$
    - There can be independencies in $P$ that are not found by conditional independence properties of $G$

# I-equivalence

- Definition: Two graphs $G_1$ and $G_2$ over a set of variables are I-equivalent if $I(G_1) = I(G_2)$




- Most graphs have many I-equivalent variants

# I-map

- $I(G) = \{(\boldsymbol{X} \perp \boldsymbol{Y}|\boldsymbol{Z}) : \text{d-sepG}(\boldsymbol{X}, \boldsymbol{Y}|\boldsymbol{Z})\}$

- $I(G) \subseteq I(P)$

# Minimal I-map

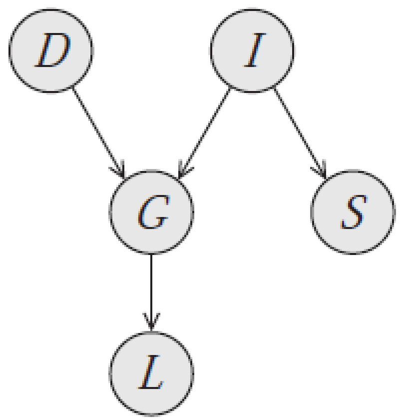
- When more independence relations exist in the graph
  - ⇒ sparser representation (fewer parameters)
  - ⇒ more informative or intuitive representation

- We want a graph that captures as much of the structure (conditional independence relations) in $P$ as possible

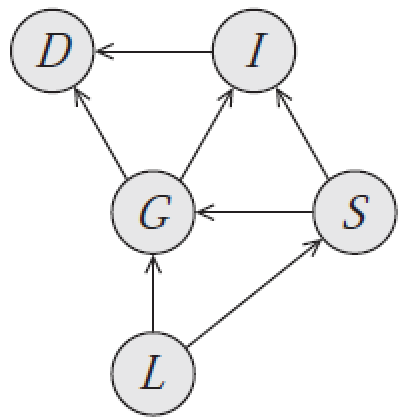
- $G$ is a **minimal I-map** for $P$ if it is an I-map for $P$, and also the removal of each edge from $G$ renders it not an I-map.

# Minimal I-map

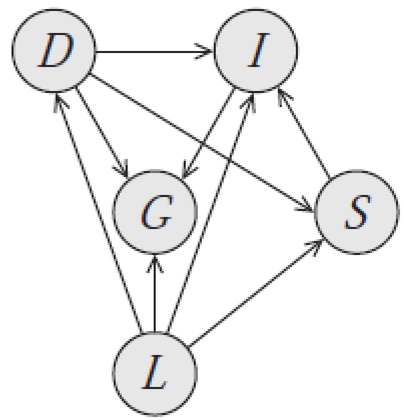
- The fact that G is a minimal I-map for P is far from a guarantee that G captures the independence structure in P




Perfect map of a distribution P




Minimal I-map of P



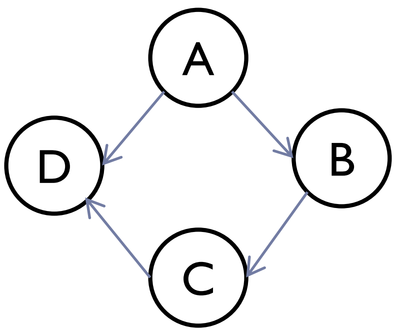

Minimal I-map of P

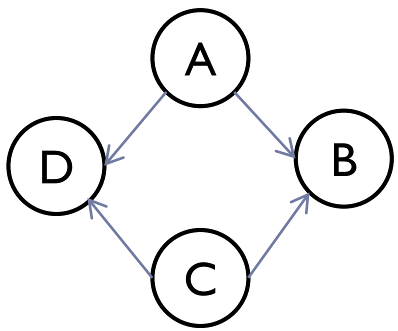


[Koller and Friedman Book]

# Perfect map

- Theorem: not every distribution has a perfect map as a DAG.
  - A distribution $P$ with the independencies
    $I(P) = \{A \perp C | \{B, D\},\ B \perp D | \{A, C\}\}$
    cannot be represented by any Bayesian network.



# Bayesian networks: summary

- *Bayesian network* is a pair ($G$, CPDs) where $G$ is a DAG and CPDs can be used to find a joint distribution $P$ that factorizes over $G$
  - Each CPD is the conditional distribution $P(X_i|Pa(X_i))$ associated to the graph node $X_i$.

- We can show “causality”, “generative schemes”, “asymmetric influences”, etc., between variables via a Bayesian network

- We can find conditional independencies from the graph structure via d-separation criteria.

# Reference


- D. Koller and N. Friedman, "Probabilistic Graphical Models: Principles and Techniques", MIT Press, 2009 [Chapter 3].