# Undirected Graphical Models: Markov Random Fields

*Probabilistic Graphical Models*

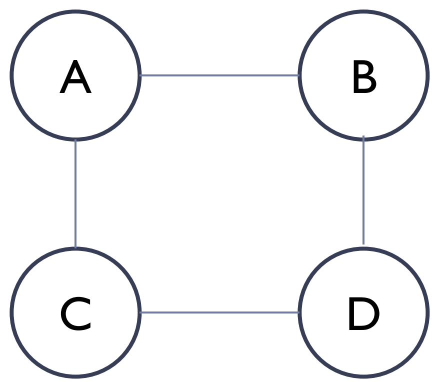
Tavassolipour

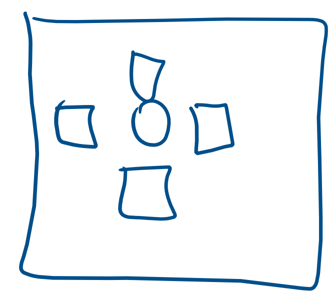
Slides from Dr. Soleymani's PGM course, Sharif University of Technology

# Markov Network

- Structure: ***undirected graph***
- Undirected edges show correlations (non-causal relationships) between variables
  - e.g., Spatial image analysis: intensity of neighboring pixels are correlated

Markov Network



# MRF: Joint distribution

- Factor $\phi(X_1, \dots, X_k)$
  - $\phi: Val(X_1, \dots, X_k) \to \mathbb{R}$
  - Scope: $\{X_1, \dots, X_k\}$

- Gibbs Distribution

Joint distribution is parameterized by factors $\boldsymbol{\Phi} = \{\phi_1(\boldsymbol{D}_1), \dots, \phi_K(\boldsymbol{D}_K)\}$:

$$P(X_1, \dots, X_N) = \frac{1}{Z} \prod_k \phi_k(\boldsymbol{D}_k)$$

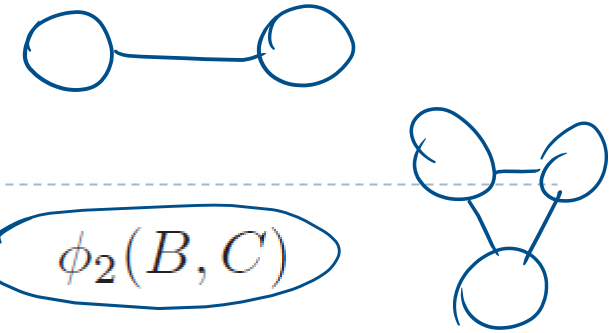
$\boldsymbol{D}_k$: the set of variables in the k-th factor

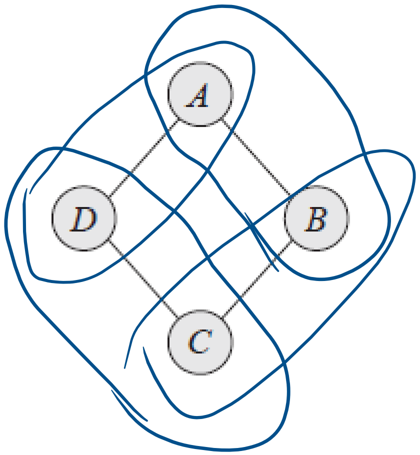
$$Z = \sum_{\boldsymbol{X}} \prod_k \phi_k(\boldsymbol{D}_k)$$

$Z$: normalization constant called **partition function**

$30 \times 100 \times 1 \times 100$

# Misconception example






[Koller & Friedman]

a b c d

0 0 0 0

$A = 0$

$\phi_1(A, B)$

| | | |
|---|---|---|
| $a^0$ | $b^0$ | 30 |
| $a^0$ | $b^1$ | 5 |
| $a^1$ | $b^0$ | 1 |
| $a^1$ | $b^1$ | 10 |

$\phi_2(B, C)$

| | | |
|---|---|---|
| $b^0$ | $c^0$ | 100 |
| $b^0$ | $c^1$ | 1 |
| $b^1$ | $c^0$ | 1 |
| $b^1$ | $c^1$ | 100 |

$\phi_3(C, D)$

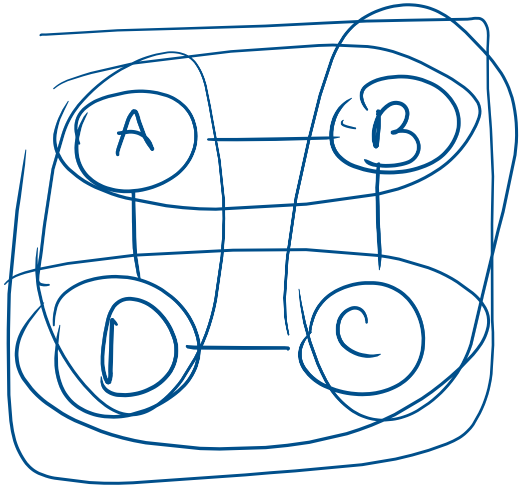
| | | |
|---|---|---|
| $c^0$ | $d^0$ | 1 |
| $c^0$ | $d^1$ | 100 |
| $c^1$ | $d^0$ | 100 |
| $c^1$ | $d^1$ | 1 |

$\phi_4(D, A)$

| | | |
|---|---|---|
| $d^0$ | $a^0$ | 100 |
| $d^0$ | $a^1$ | 1 |
| $d^1$ | $a^0$ | 1 |
| $d^1$ | $a^1$ | 100 |

Factors show "compatibilities" between different values of the variables in their scope

A factor is only one contribution to the overall joint distribution.




$$\boxed{P(A,B,C,D)}$$

$$P(A,B,C,D) = \frac{1}{Z}\,\phi_1(A,B) \times \phi_2(B,C) \times \phi_3(C,D) \times \phi_4(A,D)$$

Normalization Constant

$$\frac{1}{}$$

$$P(x_1, x_2, \ldots, x_n) = \frac{1}{Z} \phi_1 \phi_2 \cdots \phi_K$$

# Relation between factorization and independencies

$X \perp Y | Z \Rightarrow P(X,Y,Z) = f(X,Z)\,g(Y,Z)$

- **Theorem:** $P(X,Y,Z) = P(Z)\,P(X,Y|Z)$

  Let $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}$ be three disjoint sets of variables:

  $P \vDash (\boldsymbol{X} \perp \boldsymbol{Y} | \boldsymbol{Z})$ iff $P(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}) = f(\boldsymbol{X}, \boldsymbol{Z})\,g(\boldsymbol{Y}, \boldsymbol{Z})$

  $$= \underbrace{P(Z)\,P(X|Z)}_{f(X,Z)}\;\underbrace{P(Y|Z)}_{g(Y,Z)}$$

  To hold conditional independence property, $X_i$ and $X_j$ that are not directly connected must not appear in the same factor in the distributions belonging to the graph

$$P(A,B,C,D) = \frac{1}{Z}\Phi_1(A,B,C)\,\Phi_2(B,C,D)$$

# MRF Factorization: clique

$$\Phi_5(C,D)$$

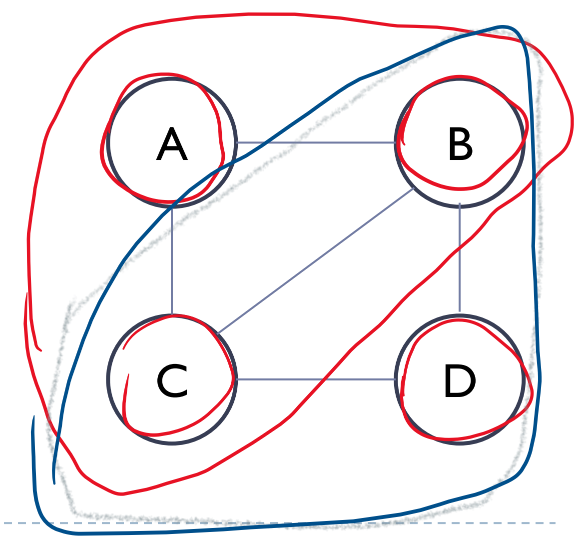
$$P(A,B,C,D) = \frac{1}{Z}\Phi_1(A,B)\,\Phi_2(A,C)\,\Phi_3(B,C)\,\Phi_4(B,D)$$

**Clique**: subsets of nodes in the graph that are fully connected (complete subgraph)

**Maximal clique**: where no superset of the nodes in a clique are also compose a clique, the clique is maximal

Cliques:
{A,B,C}, {B,C,D}, {A,B}, {A,C}, {B,C}, {B,D}, {C,D}, {A}, {B}, {C}, {D}

Max-cliques:
{A,B,C}, {B,C,D}



# MRF Factorization and pairwise independencies

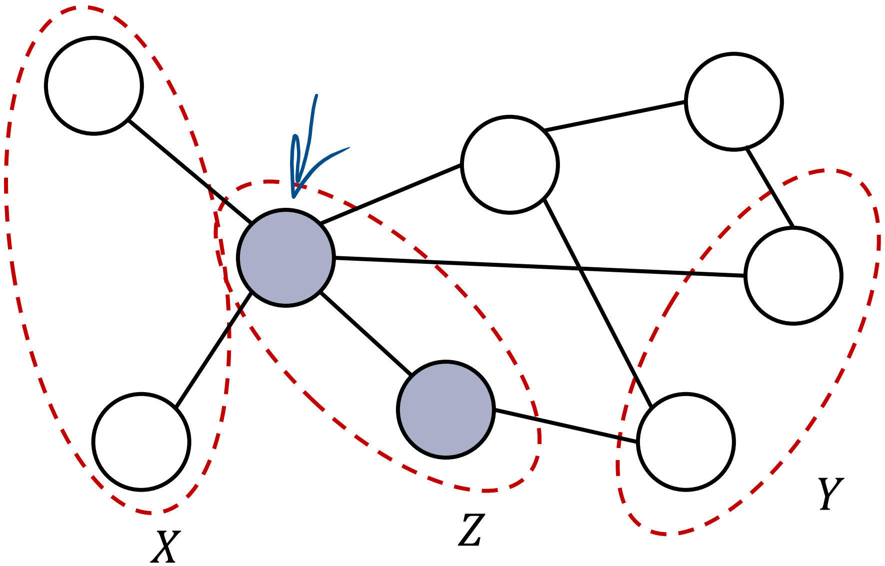
- A distribution $P_{\mathbf{\Phi}}$ with $\mathbf{\Phi} = \{\phi_1(\boldsymbol{D}_1), \dots, \phi_K(\boldsymbol{D}_K)\}$ **factorizes** over an MRF $H$ if each $\boldsymbol{D}_k$ is a **complete subgraph** of $H$

- $\overbrace{\textbf{Potential functions}}^{\phi}$ and **cliques** in the graph completely determine the **joint** distribution.

# MRFs: independencies

$X \perp Y | Z$

**Separation** in the undirected graph:




A path is active given $Z$ if no node in it is in $Z$

$X$ and $Y$ are separated given $Z$ if there is no active path between $X$ and $Y$ given $Z$

$$\mathrm{sep}_H(X, Y|Z)$$

- Global independencies for any disjoint sets A, B, C:
  - $A \perp B|C$

If all paths that connect a node in $A$ to a node in $B$ pass through one or more nodes in set $C$

# MRF: independencies

- Determining conditional independencies in undirected models is much easier than in directed ones

- Conditioning in undirected models can only eliminate dependencies while in directed ones observations can create new dependencies (v-structure)

# Factorization & Independence

- Factorization $\Rightarrow$ Independence (<u>soundness of separation criterion</u>)
  - **Theorem:** If $P$ factorizes over $H$, and $\text{sep}_H(\boldsymbol{X}, \boldsymbol{Y} | \boldsymbol{Z})$ then $P$ satisfies $\boldsymbol{X} \perp \boldsymbol{Y} | \boldsymbol{Z}$ (i.e., $H$ is an I-map of $P$)
  - $I(H) \subseteq I(P)$

- Independence $\Rightarrow$ Factorization
  - **Theorem** (Hammersley Clifford): For a positive distribution $P$, if $P$ satisfies $I(H) = \{(\boldsymbol{X} \perp \boldsymbol{Y} | \boldsymbol{Z}) : \text{sep}_H(\boldsymbol{X}, \boldsymbol{Y} | \boldsymbol{Z})\}$ then $P$ factorizes over $H$

# Factorization & Independence

- **Theorem**: Two equivalent views of graph structure for **positive distributions**:
  - If $P$ satisfies all independencies held in $H$, then it can be factorized on cliques of $H$
  - If $P$ factorizes over a graph $H$, we can read from the graph structure, independencies that must hold in $P$

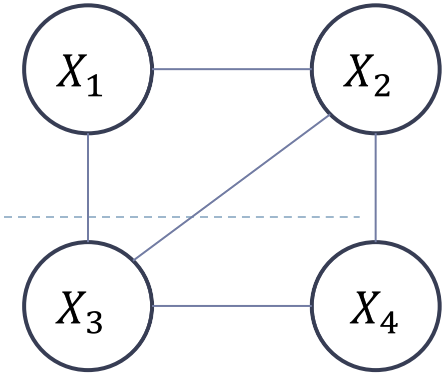
# Interpretation of clique potentials

- Potentials cannot all be marginal or conditional distributions

- A positive clique potential can be considered as general compatibility or goodness measure over values of the variables in its scope

$\phi$

# Different factorizations




- Maximal cliques:
  - $P_{\boldsymbol{\Phi}}(X_1, X_2, X_3, X_4) = \frac{1}{Z}\phi_{123}(X_1, X_2, X_3)\phi_{234}(X_2, X_3, X_4)$
  - $Z = \sum_{X_1,X_2,X_3,X_4} \phi_{123}(X_1, X_2, X_3)\phi_{234}(X_2, X_3, X_4)$
- Sub-cliques:
  - $P_{\boldsymbol{\Phi}'}(X_1, X_2, X_3, X_4) = \frac{1}{Z}\phi_{12}(X_1, X_2)\phi_{23}(X_2, X_3)\phi_{13}(X_1, X_3)\phi_{24}(X_2, X_4)\phi_{34}(X_3, X_4)$
  - $Z = \sum_{X_1,X_2,X_3,X_4} \phi_{12}(X_1, X_2)\phi_{23}(X_2, X_3)\phi_{13}(X_1, X_3)\phi_{24}(X_2, X_4)\phi_{34}(X_3, X_4)$
- Canonical representation
  - $P_{\boldsymbol{\Phi}'}(X_1, X_2, X_3, X_4) = \frac{1}{Z}\phi_{123}(X_1, X_2, X_3)\phi_{234}(X_2, X_3, X_4)\phi_{12}(X_1, X_2)\phi_{23}(X_2, X_3)\phi_{13}(X_1, X_3) \times \phi_{24}(X_2, X_4)\phi_{34}(X_3, X_4)\phi_1(X_1)\phi_2(X_2)\phi_3(X_3)\phi_4(X_4)$
  - $Z = \sum_{X_1,X_2,X_3,X_4} \phi_{123}(X_1, X_2, X_3)\phi_{234}(X_2, X_3, X_4)\phi_{12}(X_1, X_2)\phi_{23}(X_2, X_3) \times \phi_{13}(X_1, X_3)\phi_{24}(X_2, X_4)\phi_{34}(X_3, X_4)\phi_1(X_1)\phi_2(X_2)\phi_3(X_3)\phi_4(X_4)$

# Pairwise MRF

- All of the factors on single variables or pair of variables $(X_i, X_j)$:

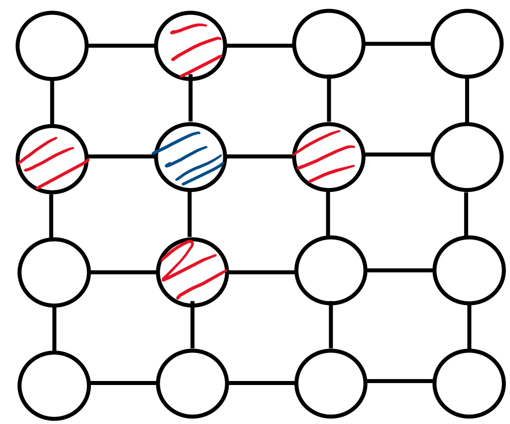
$$P(\boldsymbol{X}) = \frac{1}{Z} \prod_{(X_i, X_j) \in H} \phi_{ij}(X_i, X_j) \prod_i \phi_i(X_i)$$

- Pairwise MRFs are popular (simple special case of general MRFs)

# Ising model

- $X_i \in \{-1,1\}$

$$P(\boldsymbol{x}) = \frac{1}{Z}\exp\left\{\sum_i \theta_i x_i + \sum_{i,j\in E} w_{ij}x_i x_j\right\}$$



- Grid model
  - Image processing, lattice physics, etc.
  - The states of adjacent nodes are related

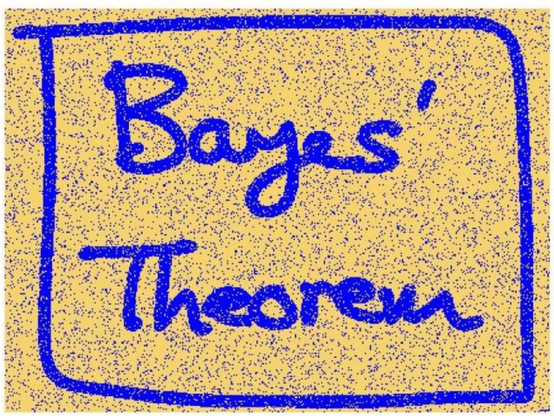
# Binary Image Denoising

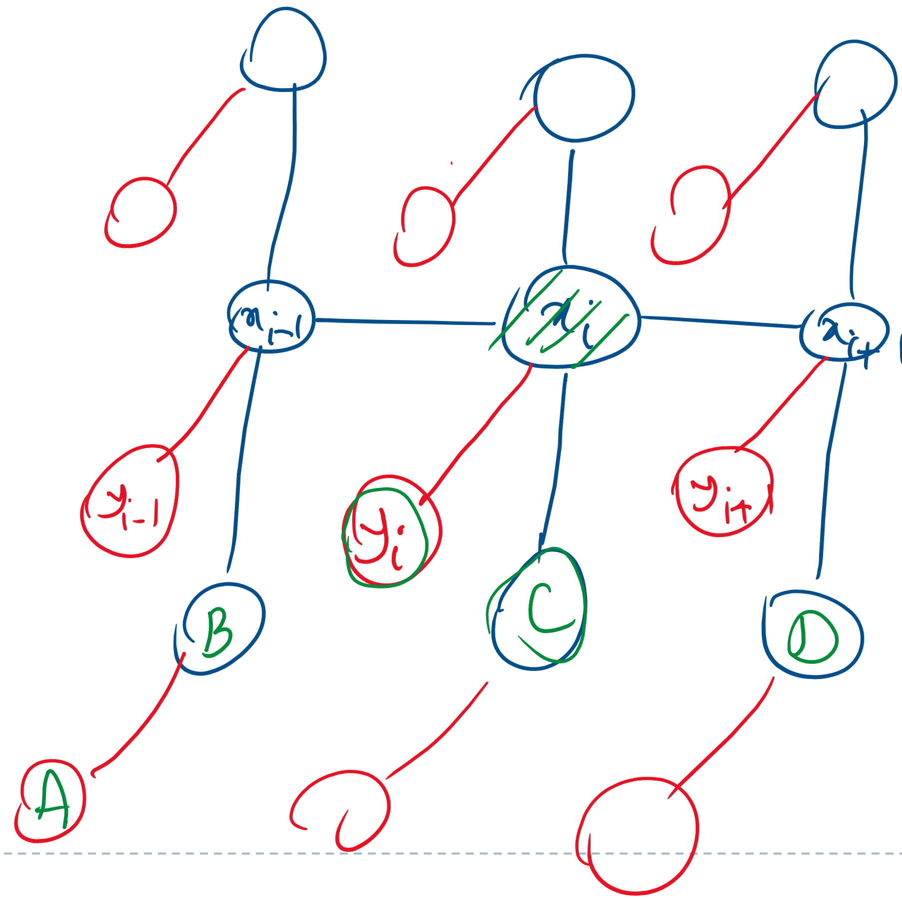
org

$x_i$ $y_i$

$x$ $y$

- $y_i \in \{-1,1\}$, array of observed noisy pixels
- $x_i \in \{-1,1\}$, noise free image



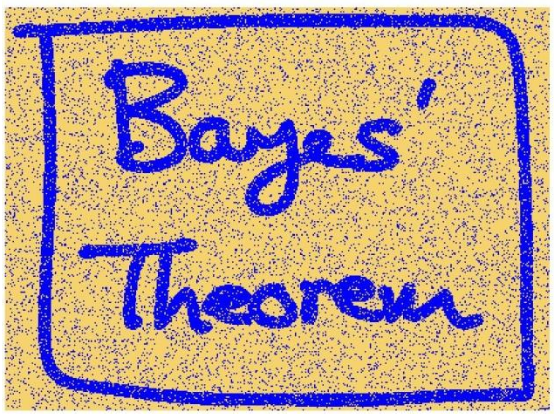

$\phi(y_i, x_i)$

$\phi(x_i, x_j)$

$\boxed{(x_i, x_j) \in E}$

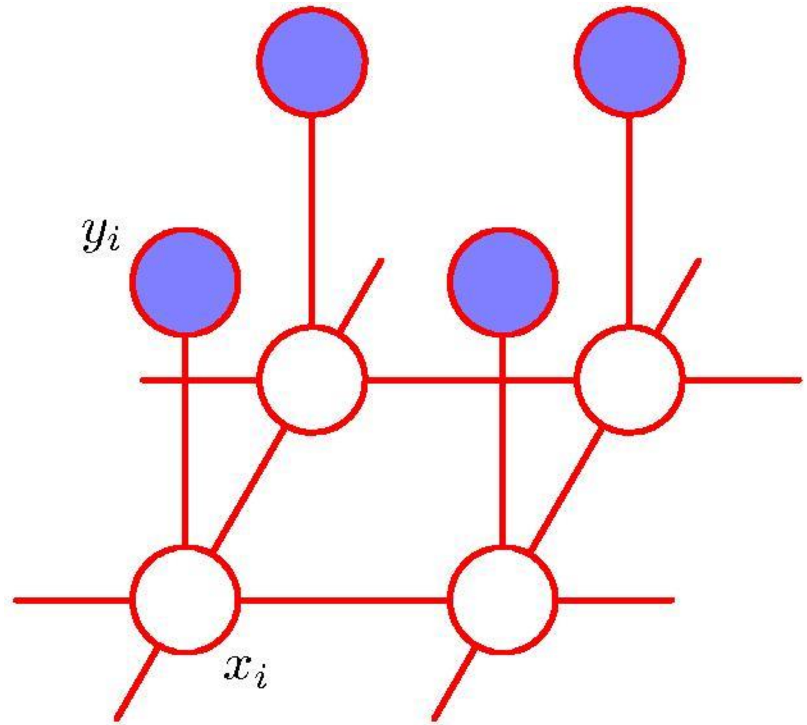
# Image denoising

$$\boxed{\max_{x} P(x|y)}$$




$$\phi(x_i, x_j) = e^{x_i x_j}$$

$$\phi(x_i, y_i) = e^{x_i y_i}$$






[Bishop]

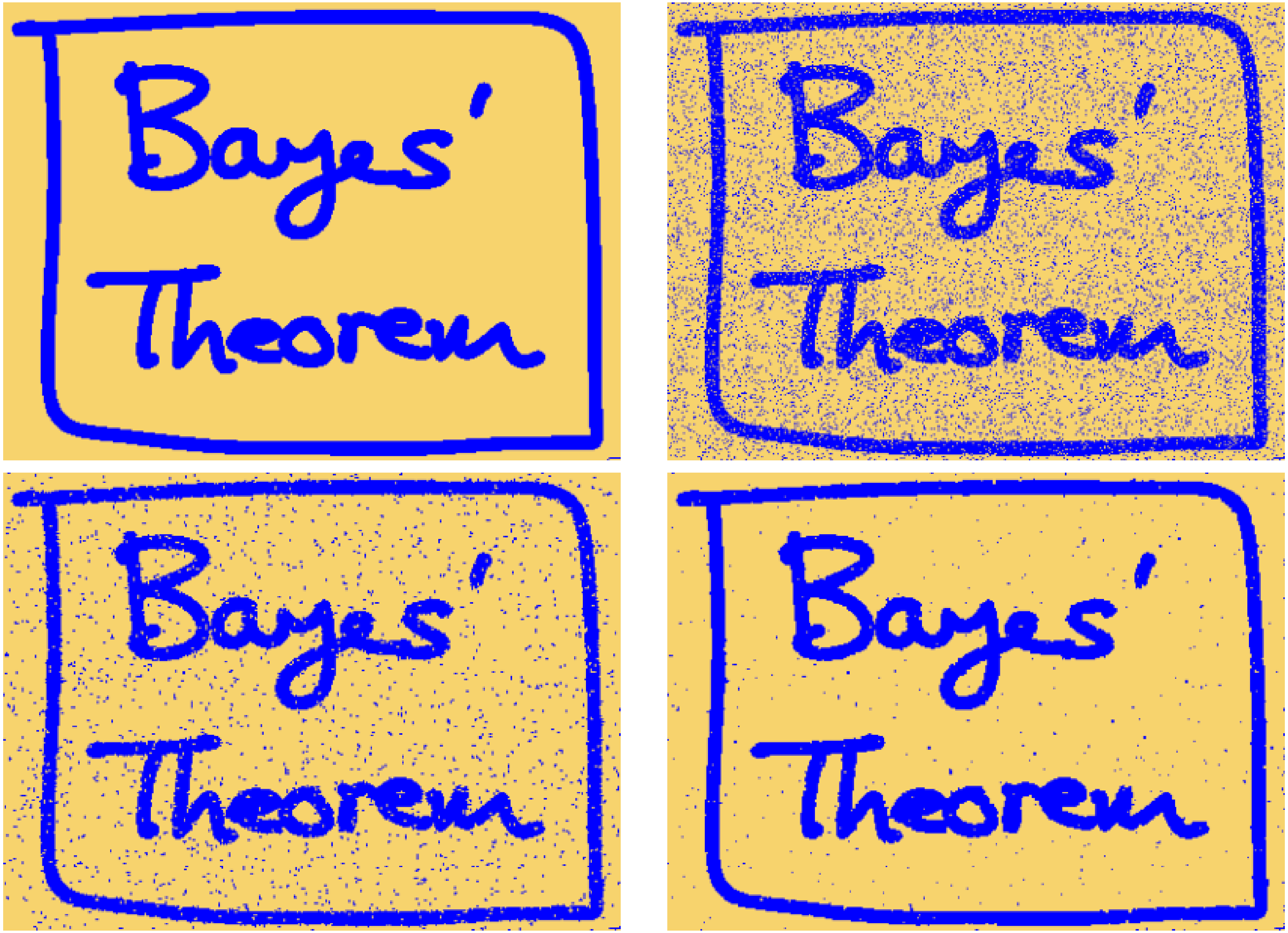
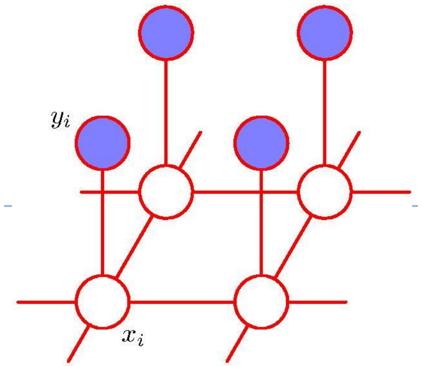



**Figure 8.30** Illustration of image de-noising using a Markov random field. The top row shows the original binary image on the left and the corrupted image after randomly changing 10% of the pixels on the right. The bottom row shows the restored images obtained using iterated conditional models (ICM) on the left and using the graph-cut algorithm on the right. ICM produces an image where 96% of the pixels agree with the original image, whereas the corresponding number for graph-cut is 99%.

# Image denoising




$$P(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{Z} \prod_i \exp\{\gamma x_i y_i\} \prod_i \exp\{\beta x_i\} \prod_{i,j \in H} \exp\{\alpha x_i x_j\}$$

$$= \frac{1}{Z} \exp\left\{ \sum_i \gamma x_i y_i + \sum_i \beta x_i + \sum_{i,j \in H} \alpha x_i x_j \right\}$$

MPA: Most probable assignment of $\boldsymbol{x}$ variables given an evidence $\boldsymbol{y}$

$$\widehat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\operatorname{argmax}}\, P(\boldsymbol{x}|\boldsymbol{y})$$

# MRF: Markov Blanket

- A variable is conditionally independent of every other variables conditioned only on its neighboring nodes

$$X_i \perp \boldsymbol{X} - \{X_i\} - MB(X_i) \mid MB(X_i)$$

$$MB(X_i) = \{X' \in \boldsymbol{X} | (X_i, X') \in edges\}$$

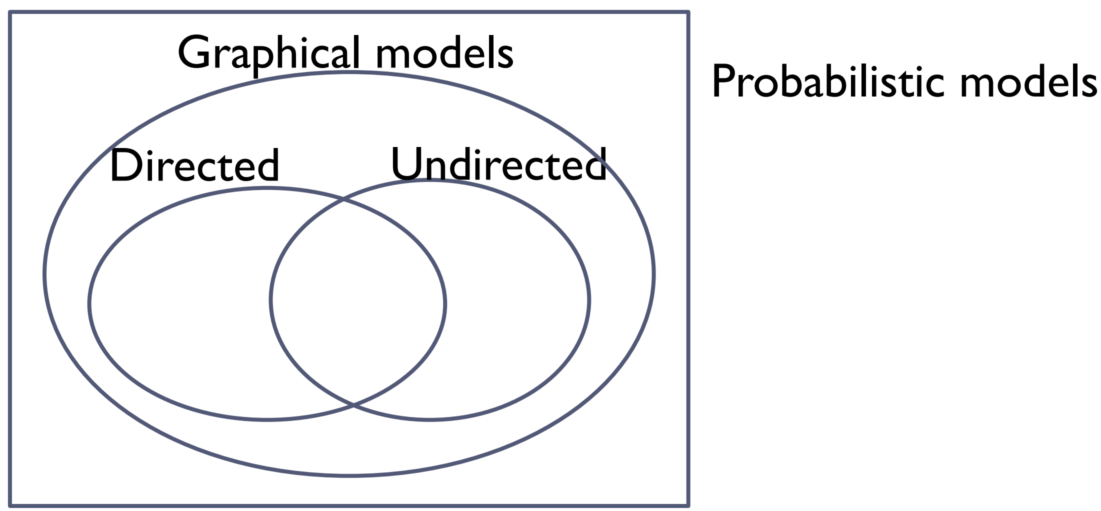
# Minimal I-map

- Since we may not find a Markov Network (MN) that is a perfect map of a BN and vice versa, we study the minimal I-map property

- $H$ is a minimal I-map for $G$ if
  - $I(H) \subseteq I(G)$
  - Removal of a single edge in $H$ renders it is not an I-map of $G$

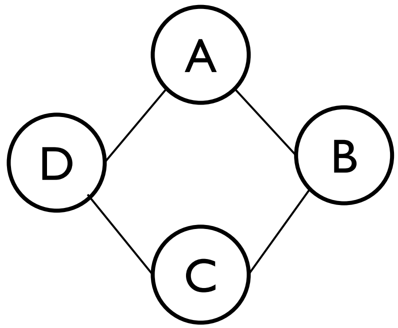
# Perfect map of a distribution

- Not every distribution has a MN perfect map
- Not every distribution has a BN perfect map



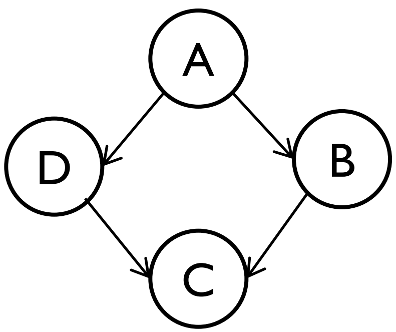
# Loop of at least 4 nodes without chord has no equivalent in BNs

- Is there a BN that is a perfect map for this MN?



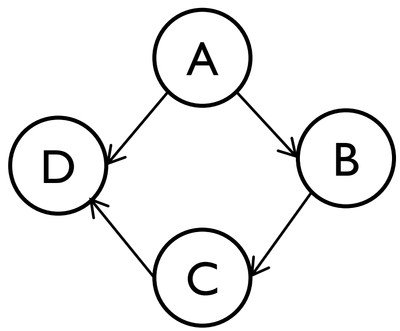

$A \perp C | \{B, D\}$

$B \perp D | \{A, C\}$



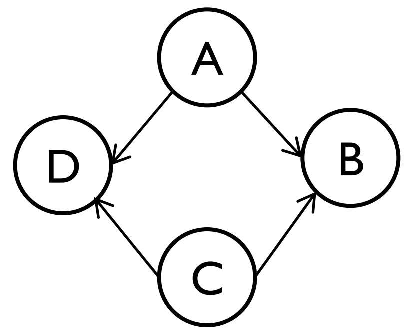

$A \perp C | \{B, D\}$

$B \perp D | \{A, C\}$ ✗




$B \perp D | \{A, C\}$

$A \perp C | \{B, D\}$ ✗

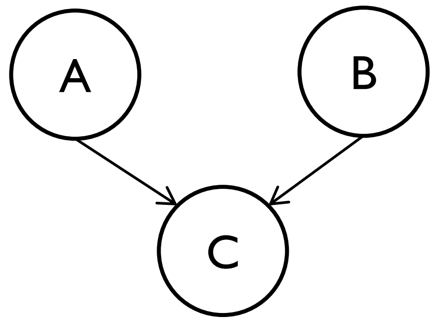



$B \perp D | \{A, C\}$

$A \perp C | \{B, D\}$ ✗

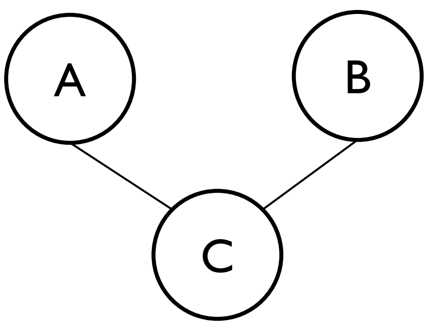
# V-structure has no equivalent in MNs

- Is there an MN that is a perfect I-map of this BN?



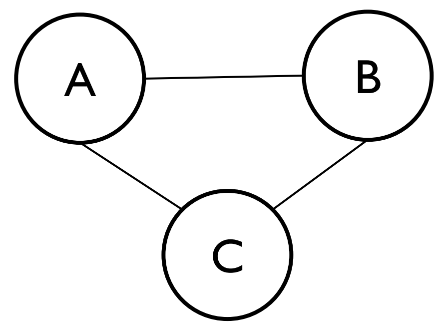

$A \perp B$

$A \perp B|C$ ✗




$A \perp B$ ✗

$A \perp B|C$




$A \perp B$ ✗

$A \perp B|C$ ✗

# Minimal I-maps: from DAGs to MNs

- The **moral graph** $M(G)$ of a DAG $G$ is an undirected graph that contains an undirected edge between $X$ and $Y$ if:
  - there is a directed edge between them in either direction
  - $X$ and $Y$ are parents of the same node

- Moralization turns a node and its parent into a fully connected sub-graph

# Minimal I-maps: from DAGs to MNs

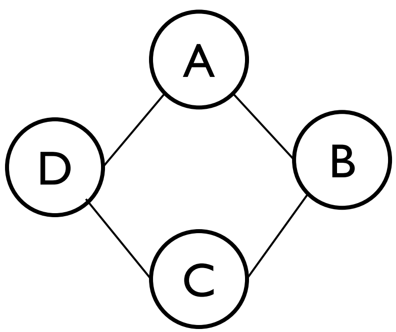
- **Theorem**: The moral graph $M(G)$ of a DAG $G$ is a minimal I-map for $G$
  - The moral graph loses some independence information
  - But, we cannot remove any edge from it without appearing new independencies that are not in $G$
    - all independencies in the moral graph are also satisfied in $G$

- **Theorem**: If a DAG $G$ is "moral", then its moralized graph $M(G)$ is a perfect I-map of $G$.

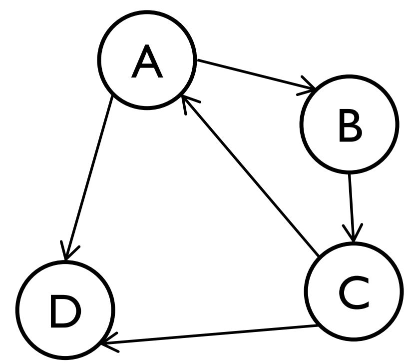
# Minimal I-maps: from MNs to DAGs

- **Theorem**: If $G$ is a BN that is minimal I-map for an MN, then $G$ cannot have immoralities.

- **Corollary**: If $G$ is a minimal I-map for an MN then it is **chordal**
  - Any BN that is I-map for an MN must add triangulating edges into the graph



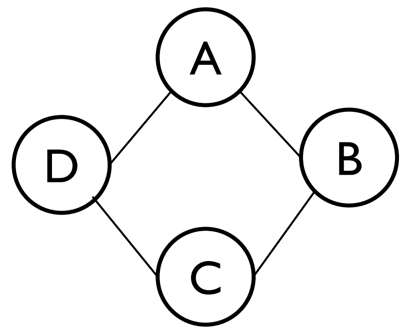

An undirected graph is chordal if any loop with more than three nodes has a chord




$G$ is a minimal I-map of the left MN

# Perfect I-map

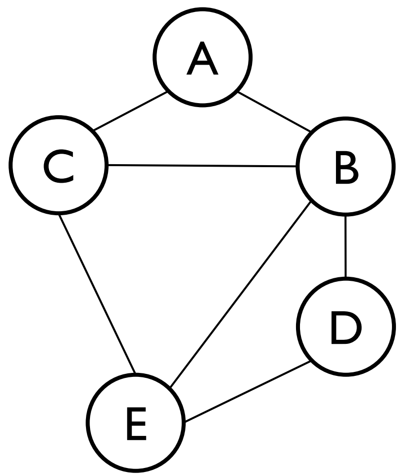
- **Theorem**: Let $H$ be a non-chordal MN. Then there is no BN that is a perfect I-map for $H$.

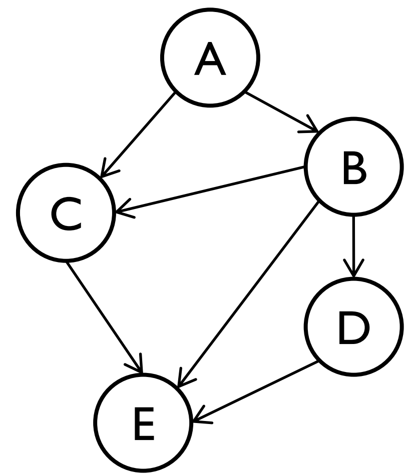



$\Rightarrow$ If the independencies in an MN can be exactly represented via a BN then the MN graph is **chordal**

# Perfect I-map

- **Theorem**: Let $H$ be a chordal MN. Then there exists a DAG $G$ that is a perfect I-map for $H$

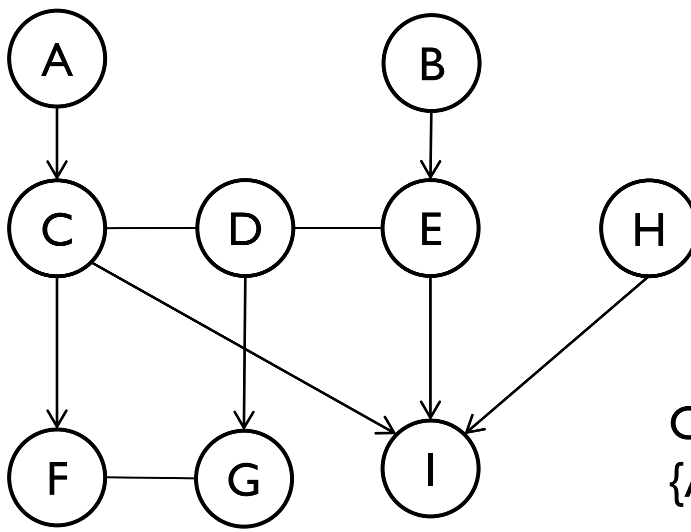



$\Rightarrow$ The independencies in a graph can be represented in both type of models **if and only if** the graph is **chordal**

# Partially Directed Acyclic Graphs (PDAGs)

- Superset of both directed and undirected graphs
- PDAGs are also called **chain graphs**




Chain components:
{A}, {B}, {C,D,E},{F,G},{H}, {I}



[Koller, PGM book]

# Relationship between BNs and MNs: summary

- Directed and undirected models represent different families of independence assumptions
  - Chordal graphs can be represented in both BNs and MNs

- For inference, we can use a single representation for both types of these models
  - simpler design and analysis of the inference algorithm