

# Undirected Graphical Models: Markov Random Fields

*Probabilistic Graphical Models*

Tavassolipour

Bayesian Network:

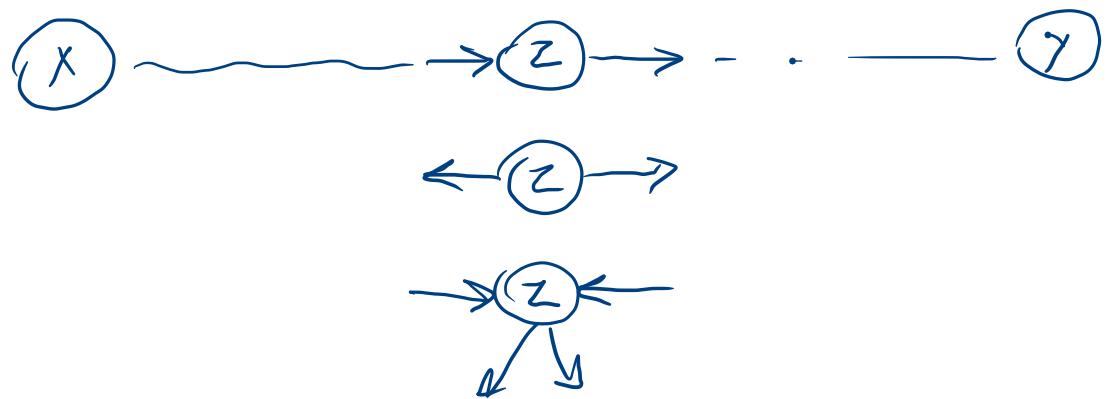
$$p(x_1, \dots, x_n) = \prod_{i=1}^n p(x_i | \text{Pa}(x_i))$$

$$\underline{I(G)} \subseteq I(p)$$

D-separation

$$x \perp y | z$$

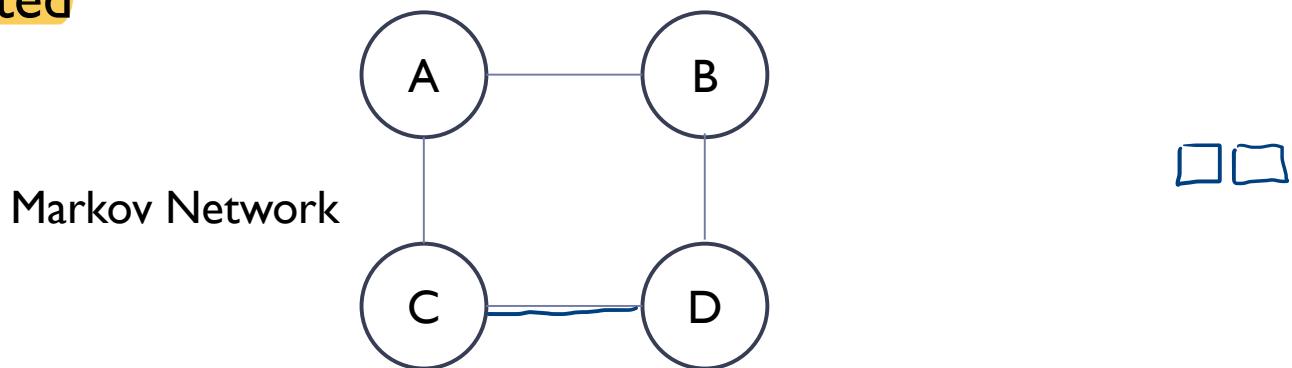
Directed





## 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



# MRF: Joint distribution

- ▶ Factor  $\phi(X_1, \dots, X_k)$

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

- ▶ Gibbs Distribution

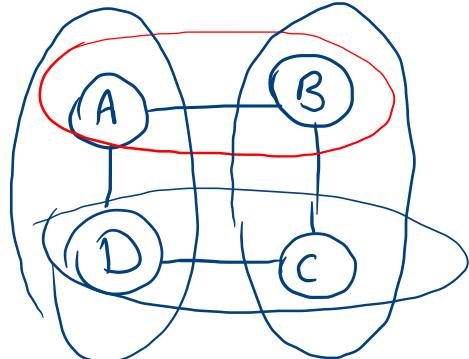
Joint distribution is parameterized by factors  $\Phi = \{\phi_1(\mathbf{D}_1), \dots, \phi_K(\mathbf{D}_K)\}$ :

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

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

$$Z = \sum_X \prod_k \phi_k(\mathbf{D}_k)$$

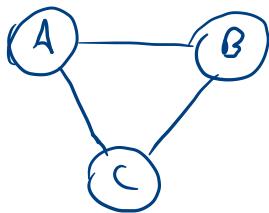
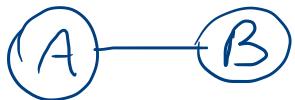
Z: normalization constant called **partition function**



$$P(A, B, C, D) = \frac{1}{Z} \underbrace{\phi_1(A, B)}_{\text{potential}} \underbrace{\phi_2(B, C)}_{\text{function}} \underbrace{\phi_3(C, D)}_{\text{ }} \underbrace{\phi_4(D, A)}_{\text{ }}$$

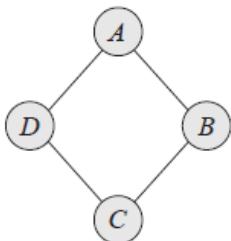
normalization  
constant

clique



$$P(A, B, C, D) = \frac{1}{Z} 30 \times 100 \times 1 \times 100$$

## Misconception example



[Koller & Friedman]

$\phi_1(A, B)$	$\phi_2(B, C)$	$\phi_3(C, D)$	$\phi_4(D, A)$
$a^0 b^0$	$b^0 c^0$	$c^0 d^0$	$d^0 a^0$
$a^0 b^1$	$b^0 c^1$	$c^1 d^0$	$d^0 a^1$
$a^1 b^0$	$b^1 c^0$	$c^0 d^1$	$d^1 a^0$
$a^1 b^1$	$b^1 c^1$	$c^1 d^1$	$d^1 a^1$

$c^0 d^0$	$d^0 a^0$	1
$c^0 d^1$	$d^0 a^1$	1
$c^1 d^0$	$d^1 a^0$	1
$c^1 d^1$	$d^1 a^1$	100

A	B	C	D	$\propto$	
0	0	0	0	300000	
0	0	0	1		
⋮	⋮	⋮	⋮		
1	1	1	1		

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

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

## MRF Factorization: clique

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

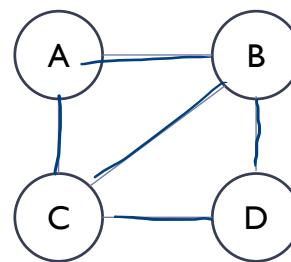
**Maximal clique:** where no superset of the nodes in a clique are also composed a clique, the clique is maximal

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}



▶ 5



1

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

$$P(A, B, C, D) = \frac{1}{Z} \underbrace{\phi_1(A, B)}_{\uparrow} \underbrace{\phi_2(B, D)}_{\uparrow} \underbrace{\phi_3(C, D)}_{\downarrow} \underbrace{\phi_4(A, C)}_{\uparrow} \phi_5(B, C)$$

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

# MRF Factorization and pairwise independencies

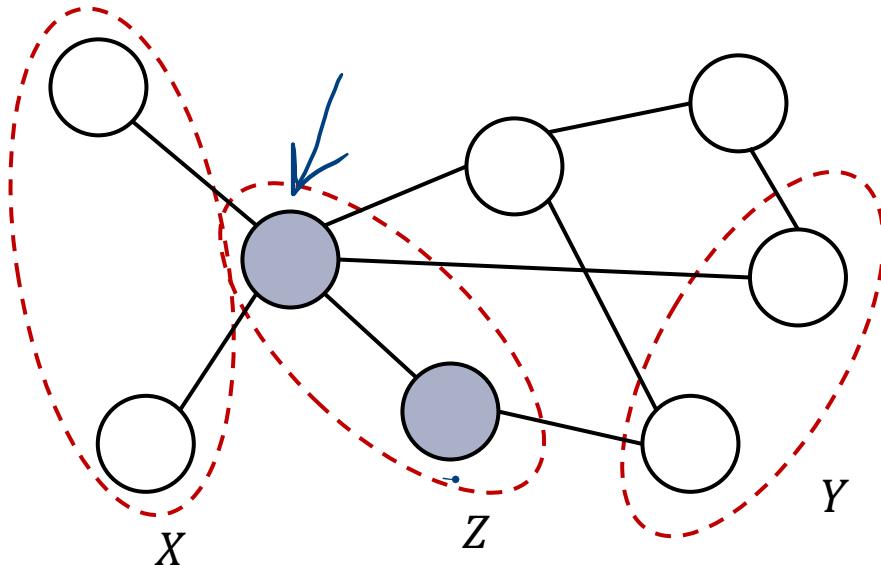
---

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

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

► 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 (soundness of separation criterion)
- ▶ **Theorem:** If  $P$  factorizes over  $H$ , and  $\text{sep}_H(X, Y|Z)$  then  $P$  satisfies  $X \perp Y|Z$  (i.e.,  $H$  is an I-map of  $P$ )
- ▶  $I(H) \subseteq I(P)$

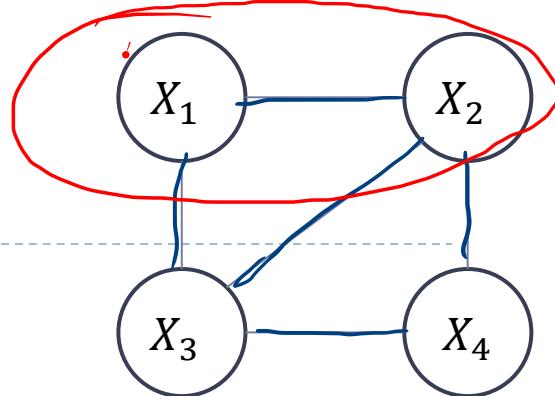


# Interpretation of clique potentials

---

- ▶ 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

# Different factorizations



- Maximal cliques:

→ ▶  $P_{\Phi}(X_1, X_2, X_3, X_4) = \frac{1}{Z} \phi_{123}(X_1, X_2, X_3) \phi_{234}(X_2, X_3, X_4)$

▶  $Z = \sum_{X_1, X_2, X_3, X_4} \phi_{123}(X_1, X_2, X_3) \phi_{234}(X_2, X_3, X_4)$

- Sub-cliques:

→ ▶  $P_{\Phi'}(X_1, X_2, X_3, X_4) =$   
 $\frac{1}{Z} \underbrace{\phi_{12}(X_1, X_2)}_{\text{red}} \underbrace{\phi_{23}(X_2, X_3)}_{\text{red}} \underbrace{\phi_{13}(X_1, X_3)}_{\text{red}} \phi_{24}(X_2, X_4) \phi_{34}(X_3, X_4)$

▶  $Z = \sum_{X_1, X_2, X_3, X_4} \phi_{12}(X_1, X_2) \phi_{23}(X_2, X_3) \phi_{13}(X_1, X_3) \phi_{24}(X_2, X_4) \phi_{34}(X_3, X_4)$

- Canonical representation

▶  $P_{\Phi'}(X_1, X_2, X_3, X_4) =$   
 $\left\{ \begin{array}{l} \frac{1}{Z} \phi_{123}(X_1, X_2, X_3) \phi_{234}(X_2, X_3, X_4) \phi_{12}(X_1, X_2) \phi_{23}(X_2, X_3) \phi_{13}(X_1, X_3) \times \\ \phi_{24}(X_2, X_4) \phi_{34}(X_3, X_4) \phi_1(X_1) \phi_2(X_2) \phi_3(X_3) \phi_4(X_4) \end{array} \right.$

▶  $Z = \sum_{X_1, X_2, X_3, X_4} \phi_{123}(X_1, X_2, X_3) \phi_{234}(X_2, X_3, X_4) \phi_{12}(X_1, X_2) \phi_{23}(X_2, X_3) \times$   
 $\phi_{13}(X_1, X_3) \phi_{24}(X_2, X_4) \phi_{34}(X_3, X_4) \phi_1(X_1) \phi_2(X_2) \phi_3(X_3) \phi_4(X_4)$



## Pairwise MRF

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

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

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



## Ising model

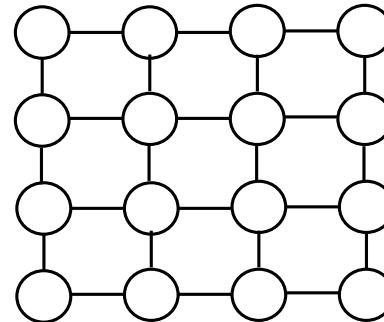
- ▶  $X_i \in \{-1, 1\}$

$$P(x) = \frac{1}{Z} e^{\sum_i \theta_i x_i + \sum_{i,j \in E} \alpha_{ij} x_i x_j}$$

$$= \frac{1}{Z} e^{\theta_1 x_1} e^{\theta_2 x_2} \dots e^{\theta_n x_n} e^{\alpha_{1,2} x_1 x_2} e^{\alpha_{1,3} x_1 x_3} \dots$$

$\Phi(x_1, x_3)$

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

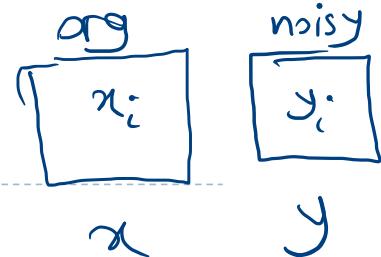


$E$

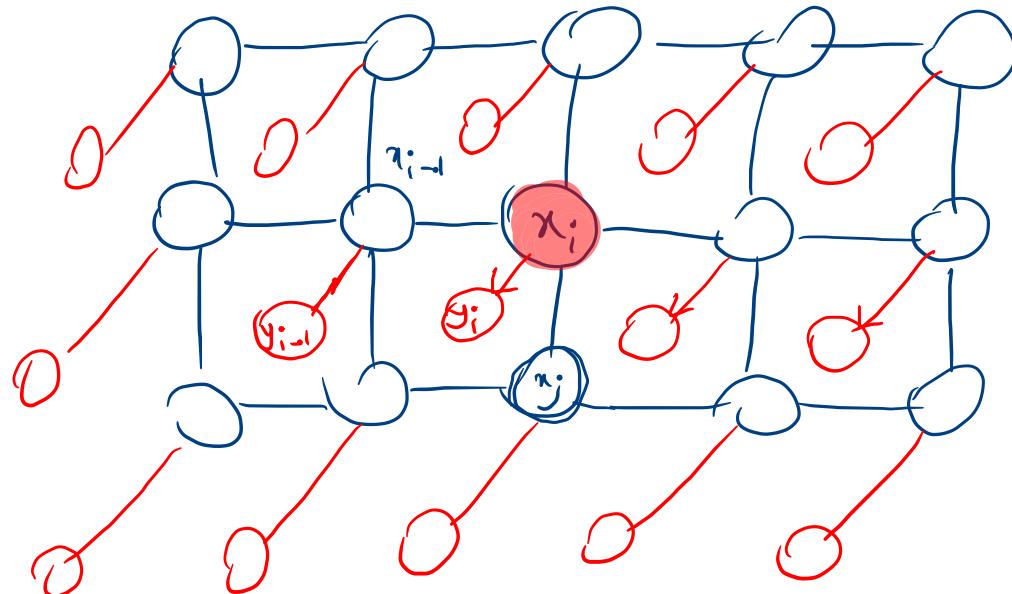
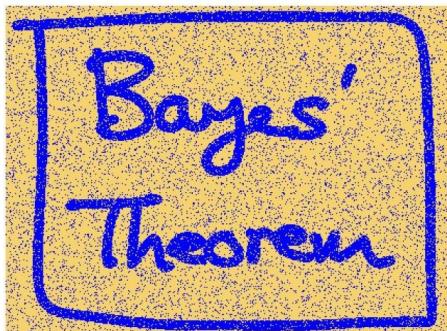
## Grid model

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

# Binary Image Denoising



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



► 14

$$\phi(x_i, x_j)$$

$$(x_i, x_j) \in E$$

$$\phi(x_i, y_i)$$

$$\phi(x_i)$$

$$\phi(y_i)$$

$$P = \frac{1}{2} \quad \text{noise}$$

$$P = 0.5$$

$$P(y=1 \mid x=1) = \frac{1}{2}$$

$$P(y=-1 \mid x=1) = \frac{1}{2}$$

$$P(y=1 \mid x=-1) = \frac{1}{2}$$

$$P(y=-1 \mid x=-1) = \frac{1}{2}$$

$$P(y|x) = P(y) = \begin{cases} \frac{1}{2} \\ \frac{1}{2} \end{cases}$$

$$\phi(x_i, x_j) = x_i x_j \quad \times$$

$$= 1 + x_i x_j = 0 \quad \times$$

$$= 2 + x_i x_j \quad \checkmark$$

$$= e^{x_i x_j}$$

$$= e^{-(x_i - x_j)^2}$$

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

$$\begin{aligned} P &\approx 0.01 \\ P &\approx 0.9 \end{aligned}$$

$$e^{-x_i y_i} = e^{x_i (-y_i)}$$

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

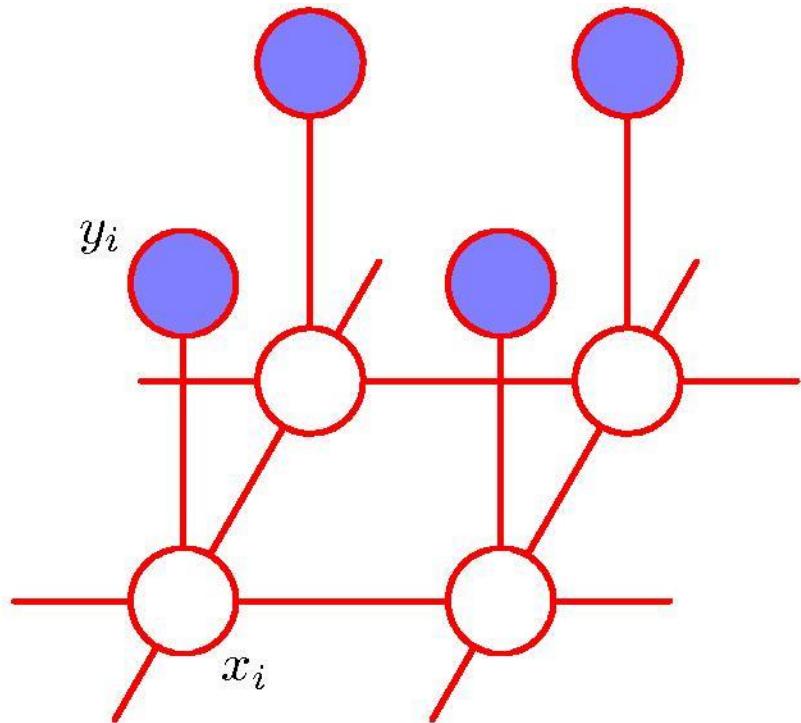
$$= e^{-x_i}$$

# Image denoising

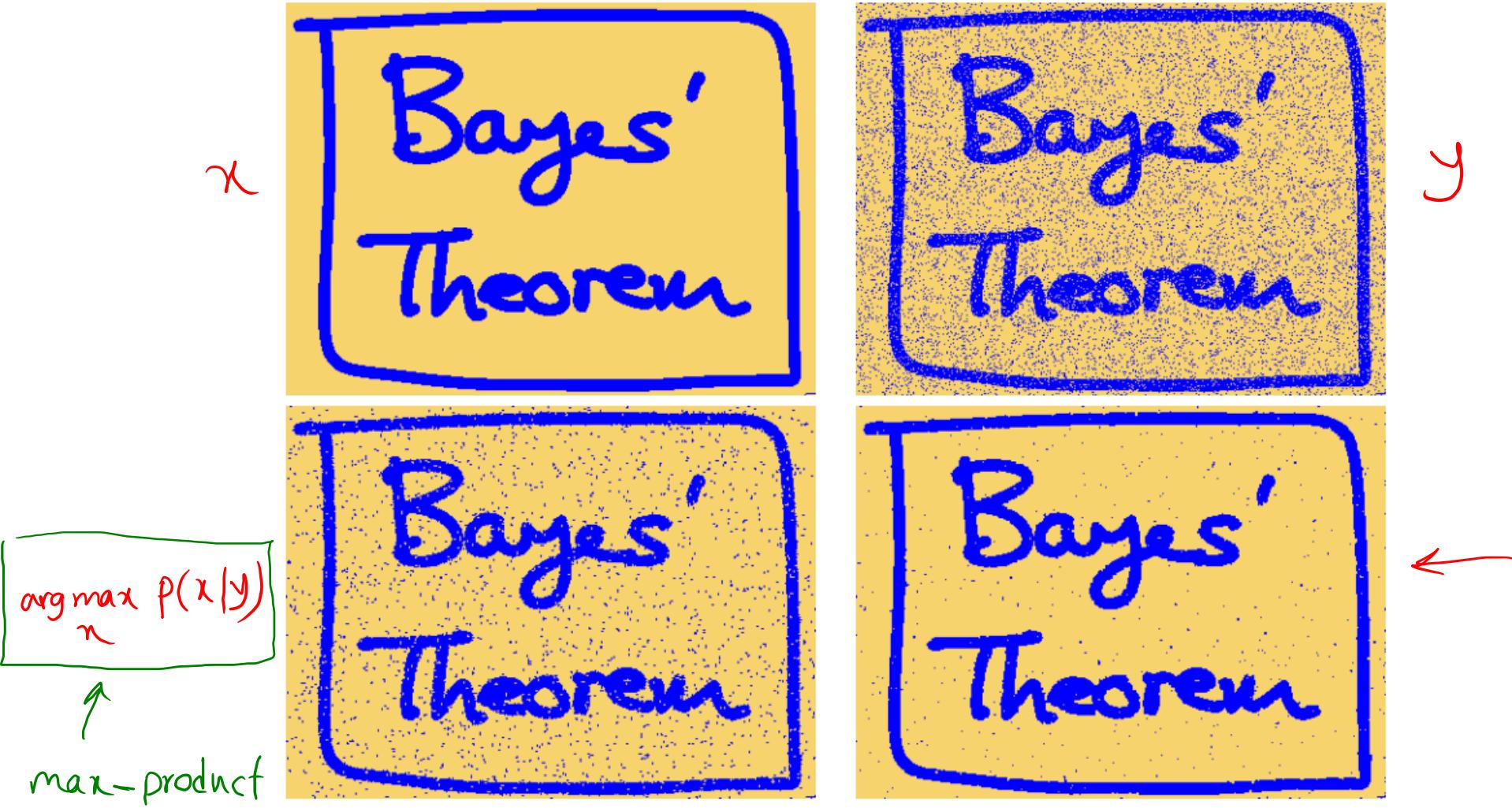
$$P(x|y)$$

Bayes' Theorem

$$\hat{x} = \arg \max_a P(a|y)$$



$$P(\underline{x}, \underline{y}) = \frac{1}{Z} e^{\alpha \sum_{x_i, x_j \in E} x_i x_j + \beta \sum_i x_i y_i + \gamma \sum_i x_i}$$



**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%.

## Bayesian Network

$$P(x_1, \dots, x_n) = \prod_{i=1}^n P(x_i | \text{Pa}(x_i))$$

D-separation

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

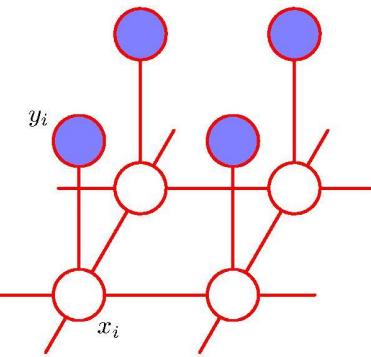
---

## Markov Random Fields (MRFs, MN)

$$P(x_1, \dots, x_n) = \frac{1}{Z} \prod_{i=1}^K \phi_i(D_i)$$

separation

# Image denoising



$$\begin{aligned}
 P(x, y) &= \frac{1}{Z} \prod_i \underbrace{\exp\{\gamma x_i y_i\}}_{\text{Evidence term}} \prod_i \underbrace{\exp\{\beta x_i\}}_{\text{Prior term}} \prod_{i,j \in H} \underbrace{\exp\{\alpha x_i x_j\}}_{\text{Smoothness term}} \\
 &= \frac{1}{Z} \exp \left\{ \sum_i \gamma x_i y_i + \sum_i \beta x_i + \sum_{i,j \in H} \alpha x_i x_j \right\}
 \end{aligned}$$

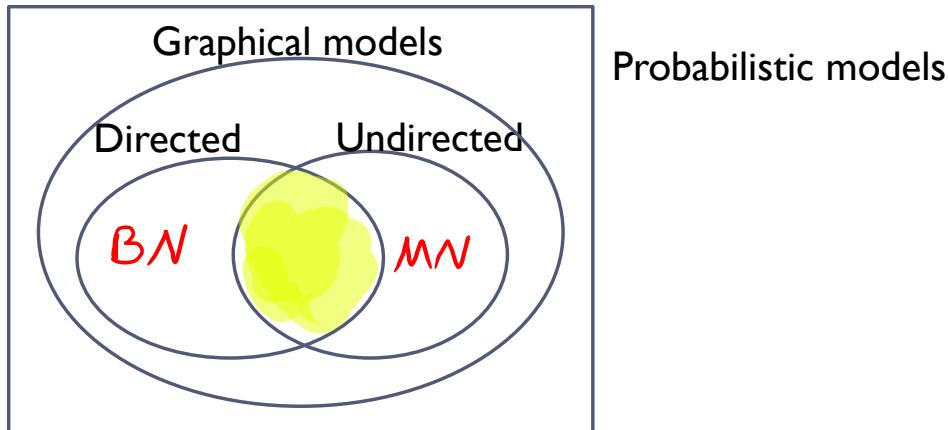
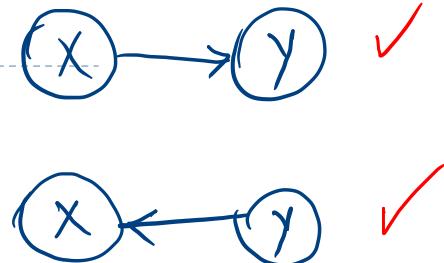
MPA: Most probable assignment of  $x$  variables  
given an evidence  $y$

$$\boxed{\hat{x} = \underset{x}{\operatorname{argmax}} P(x|y)} = \underset{x}{\operatorname{argmax}} \frac{P(x, y)}{P(y)}$$

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

# Perfect map of a distribution

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



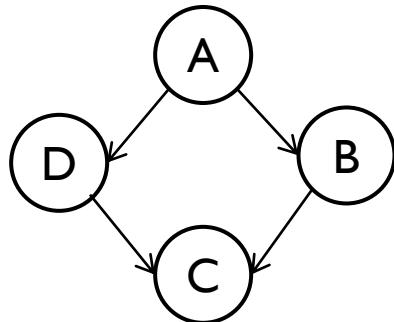
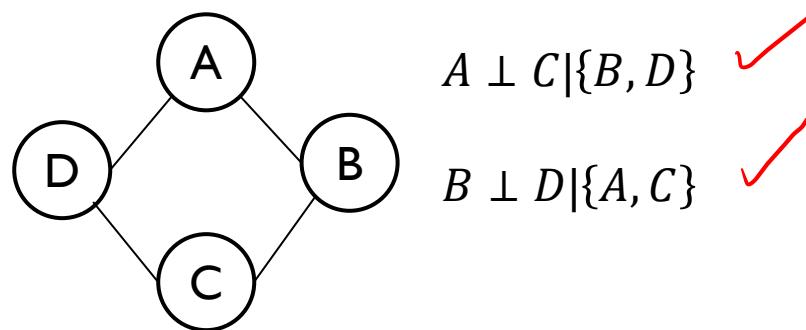
# Minimal I-map

---

- ▶ 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  $G$

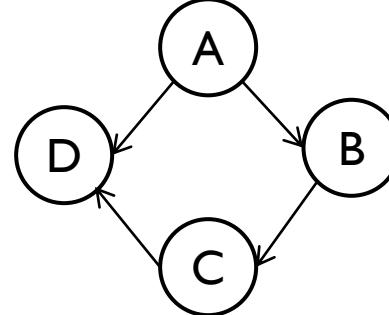
Loop of at least 4 nodes without chord has no equivalent in BNs

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



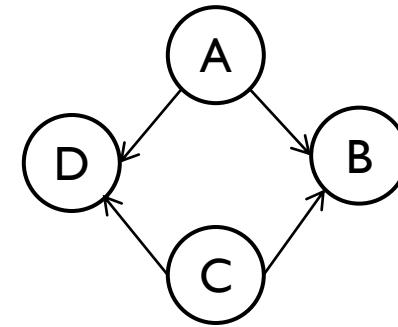
$$A \perp C | \{B, D\}$$

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



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

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

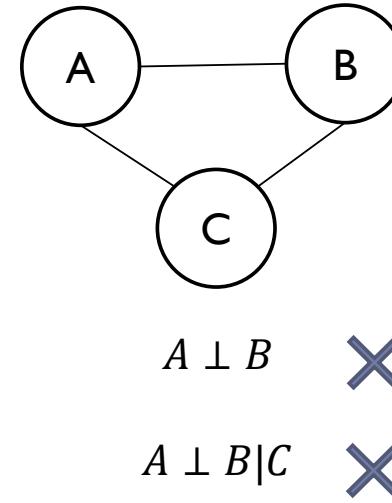
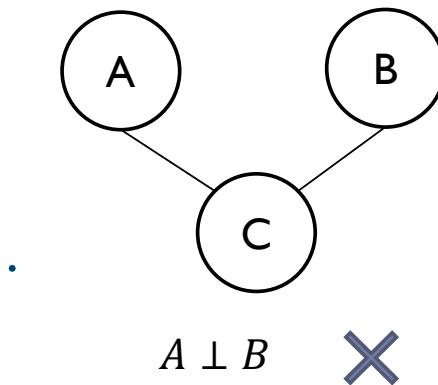
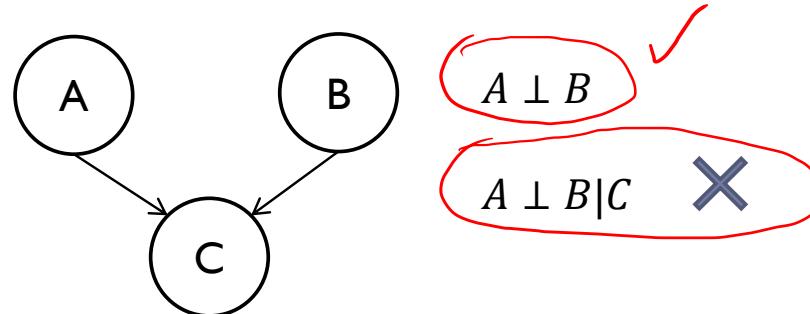
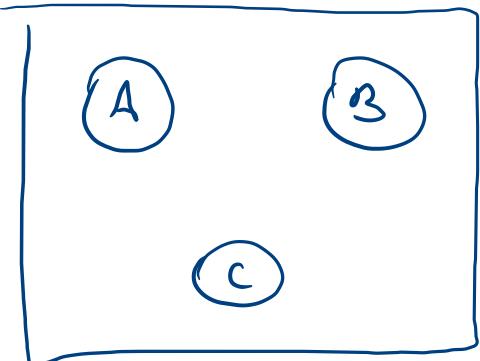


$$B \perp D | \{A, C\}$$

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

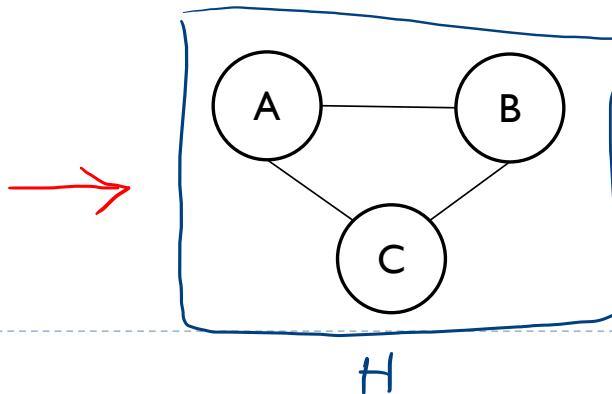
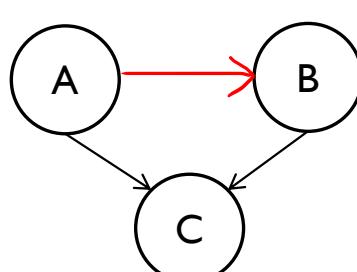
# V-structure has no equivalent in MNs

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



## 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



$$\mathcal{I}(H) \subseteq \mathcal{I}(G)$$

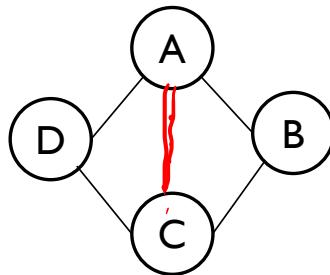
# Minimal I-maps: from DAGs to MNs

- ▶ **Theorem:** The moral graph  $\underline{M(G)}$  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  $G$ .

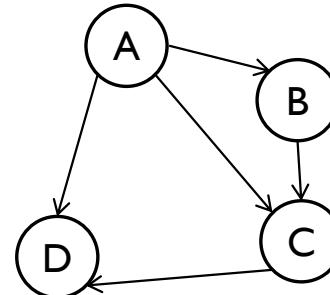
$$I(G) \geq I(H)$$

## 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

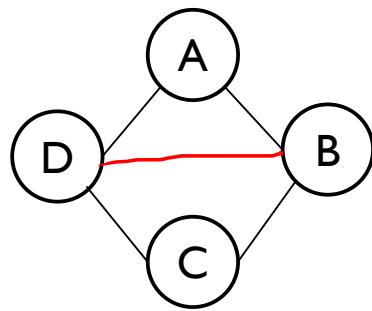


$G$  is a minimal I-map of the left MN

$$I(G) \subseteq I(H)$$

## Perfect I-map

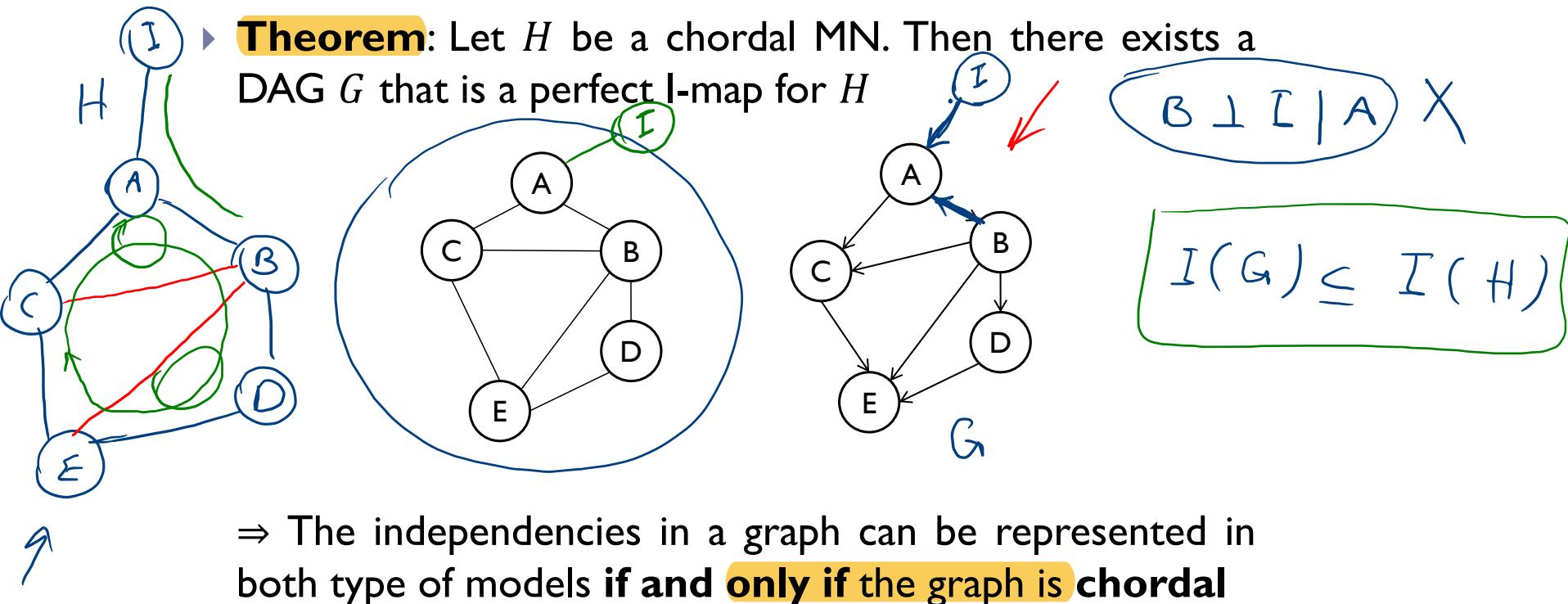
- ▶ **Theorem:** Let  $H$  be a non-chordal MN. Then there is no BN that is a perfect I-map for  $H$ .



$$I(G) = I(P)$$

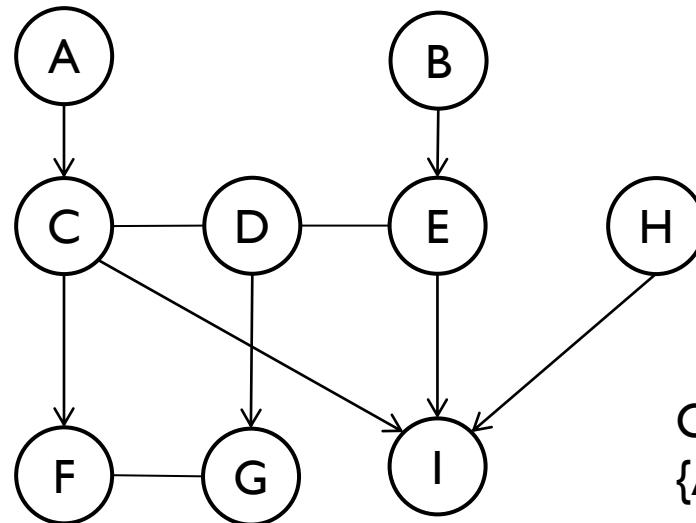
⇒ If the independencies in an MN can be exactly represented via a BN then the MN graph is **chordal**

## Perfect I-map



# Partially Directed Acyclic Graphs (PDAGs)

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



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

# 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

Markov Blanket

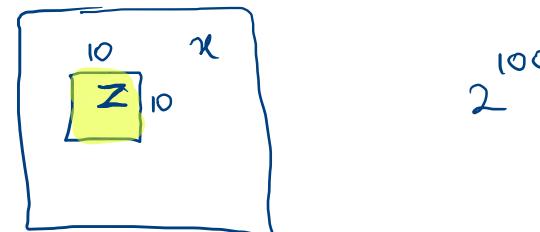
$\downarrow$   
 $MB(x) = \text{neighbors of } X$

## Variational Inference

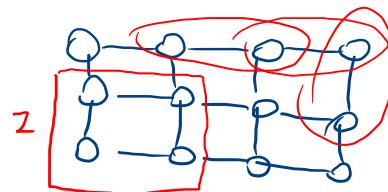
$$P(x_1, x_2, \dots, x_n)$$

$$\begin{cases} \rightarrow P(x_1 | x_2, \dots, x_n) = ? \\ \rightarrow P(x_1 | x_2, x_4) = ? \end{cases}$$

## Image Inpainting



$$P(z|x)$$



$$P(x, z) = \frac{1}{C} \prod_{i=1}^K \Phi_i(D_i)$$

$$P(x, z) \rightarrow \begin{aligned} &P(z|x) && \text{hard} \\ &P(x|z) \\ &P(x) \\ &P(z) \end{aligned}$$

$$P(z|x) = \frac{P(x,z)}{P(x)} = \frac{P(x,z)}{\sum_z P(x,z)}$$

$$S = \alpha \cup z$$

observed



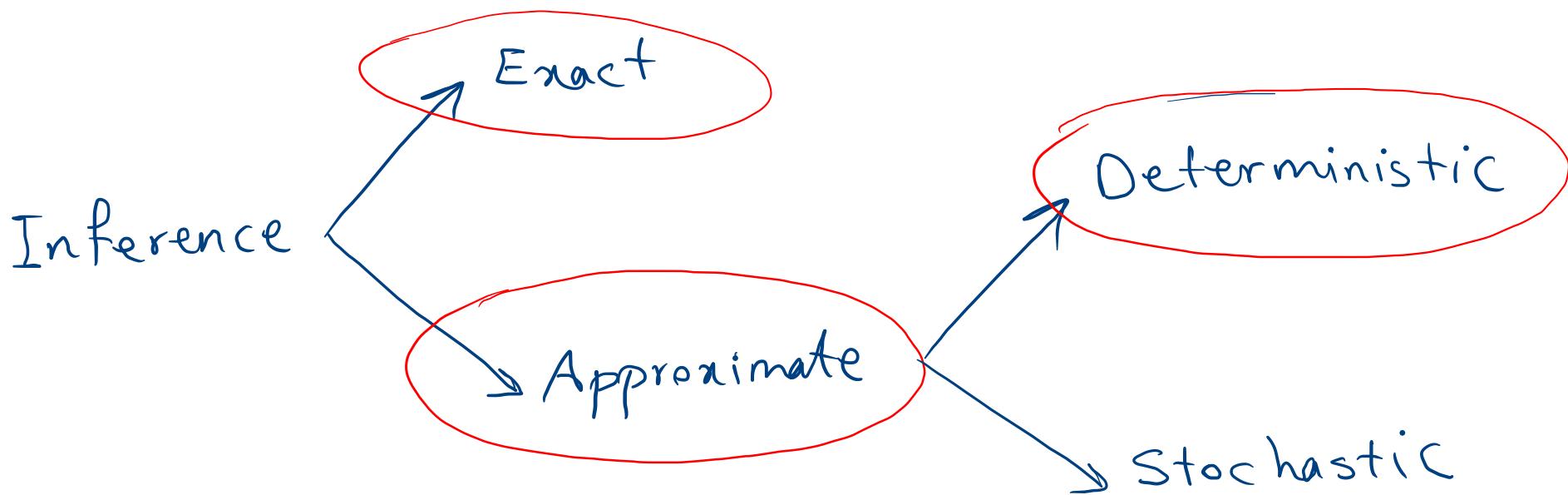
unobserved / Latent / hidden

Inference:

$$P(z|x)$$



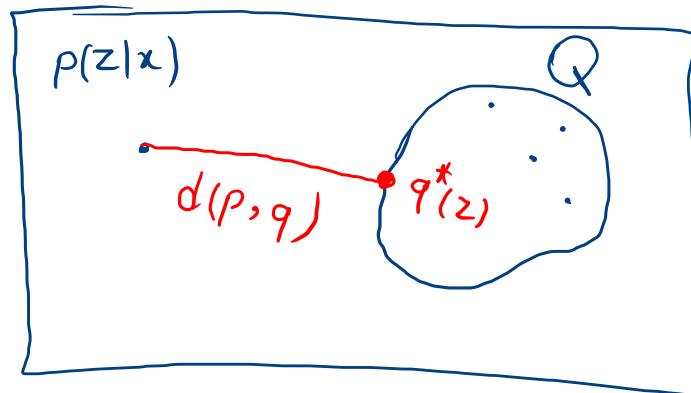
$$\underline{P(z|x)} \simeq q(z)$$



## Variational Inference

$$\underline{p(z|x)} \simeq q(z) \quad q(z|x)$$

variation calculus



$$q^*(z) = \arg \min_{q(z) \in Q} d(q(z), p(z|x))$$

V.I.

$$\min_x f(x)$$

Functional

$$\underline{H(P)} = - \sum_x P(x) \log P(x)$$

varitional calculus

$$p(x) \quad q(x)$$

$$\textcircled{1} \quad d(p, q) = \sum_x |p(x) - q(x)|$$

$$\textcircled{2} \quad d(p, q) = \sum_x (p(x) - q(x))^2$$

$$\textcircled{3} \quad d(p, q) = \sum_x p(x) |p(x) - q(x)|$$

$$\textcircled{4} \quad d(p, q) = \sum_x q(x) |p(x) - q(x)|$$

x

$$p(x) \gg q(x)$$

$$p(0.2) = 0.9$$

$$p(0.2) \approx 0.001$$

$$q(0.2) = 0.001$$

$$q(0.2) \approx 0.9$$

$$d(p, q) = \sum_n p(n) q(n) |p(n) - q(n)|$$

$$p(n) = \epsilon$$

$$p(n) = 1 - \epsilon$$

$$q(n) = 2\epsilon$$

$$q(n) = 1 - 2\epsilon$$

$d(x, y)$

①  $d(x, y) \geq 0$  ✓ Jensen-inequality

②  $x = y \iff d(x, y) = 0$  ✓

③  $d(x, y) = d(y, x)$  X

④  $d(x, y) + d(y, z) \geq d(x, z)$  X

$$\sum_x \log \frac{p(x)}{q(x)}$$

$$p(x) > 0$$

$$q(x) > 0$$

KL Divergence

$$KL(p \parallel q) = \sum_x p(x) \log \frac{p(x)}{q(x)}$$

$$KL(q \parallel p) = \sum_x q(x) \log \frac{q(x)}{p(x)}$$