Graphical Models

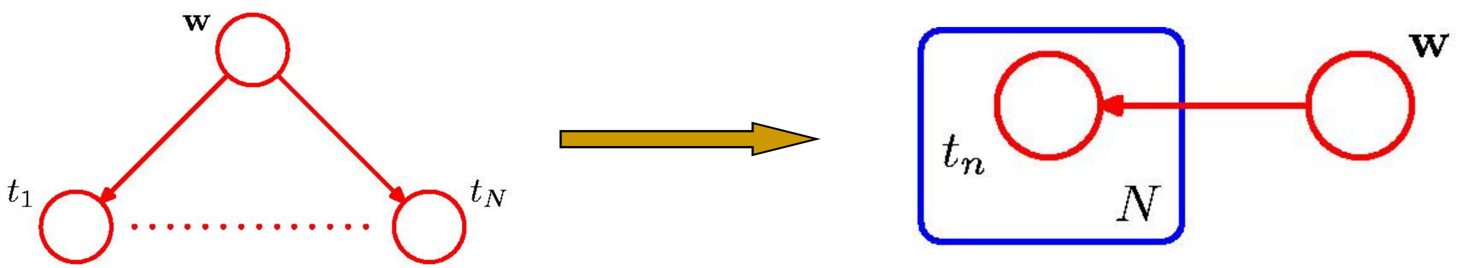
Lei Tang

Review of Graphical Models

- Directed Graph (DAG, Bayesian Network, Belief Network)
- Typically used to represent causal relationship
- Undirected Graph (Markov Random Field, Markov Network)
- Usually when the relationship between variables are not very clear.

Some rules(1)

- A graph to represent a regression problem
- Plate is used to represent repetition.

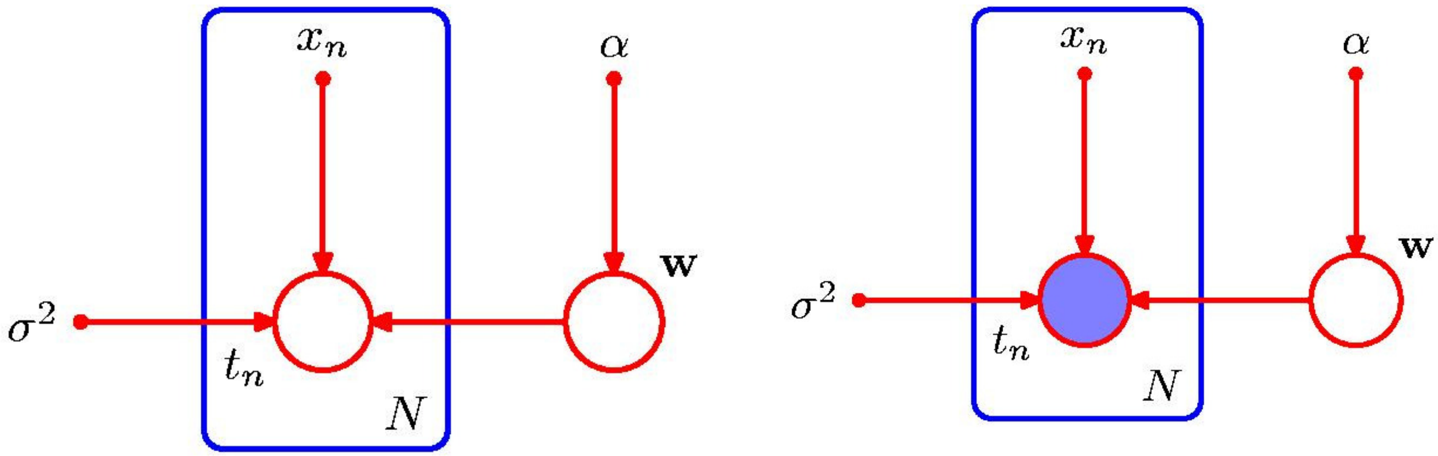


$$p(\mathbf{t}, \mathbf{w}) = p(\mathbf{w}) \prod_{n=1}^{N} p(t_n|\mathbf{w}).$$

Some rules(2)

- Suppose we have some parameters

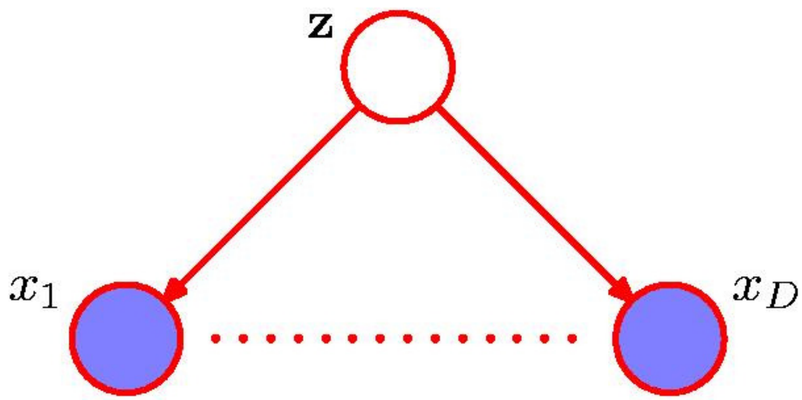
$$p(\mathbf{t}, \mathbf{w}|\mathbf{x}, \alpha, \sigma^2) = p(\mathbf{w}|\alpha) \prod_{n=1}^{N} p(t_n|\mathbf{w}, x_n, \sigma^2).$$



- Observations are shaded.

Model Representation (DAG)

- Usually, the higher-numbered variables corresponds to terminal nodes of the graph, representing the observations; Lower-numbered nodes are latent variables.
- A graph representing the naïve Bayes model.



Factorization

- For directed graph:

$$p(\mathbf{x}) = \prod_{k=1}^{K} p(x_k|\mathrm{pa}_k)$$

- (Ancestral Sampling)
- For undirected graph:

$$p(\mathbf{x}) = \frac{1}{Z}\prod_{C}\psi_C(\mathbf{x}_C).$$

Potential Function

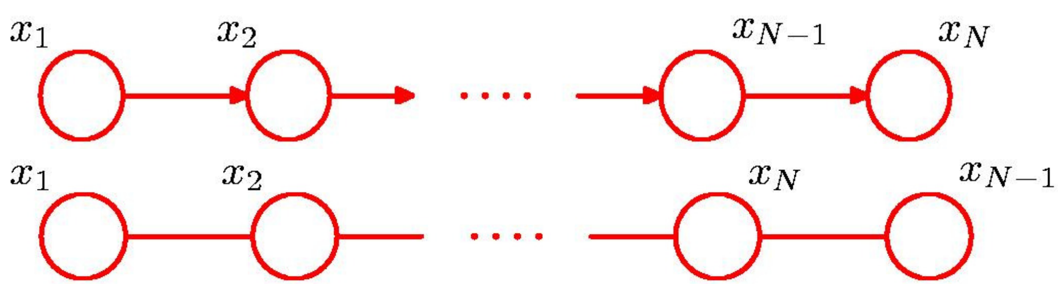
Partition Function

$$Z = \sum_{\mathbf{x}}\prod_{C}\psi_C(\mathbf{x}_C)$$

Energy Function

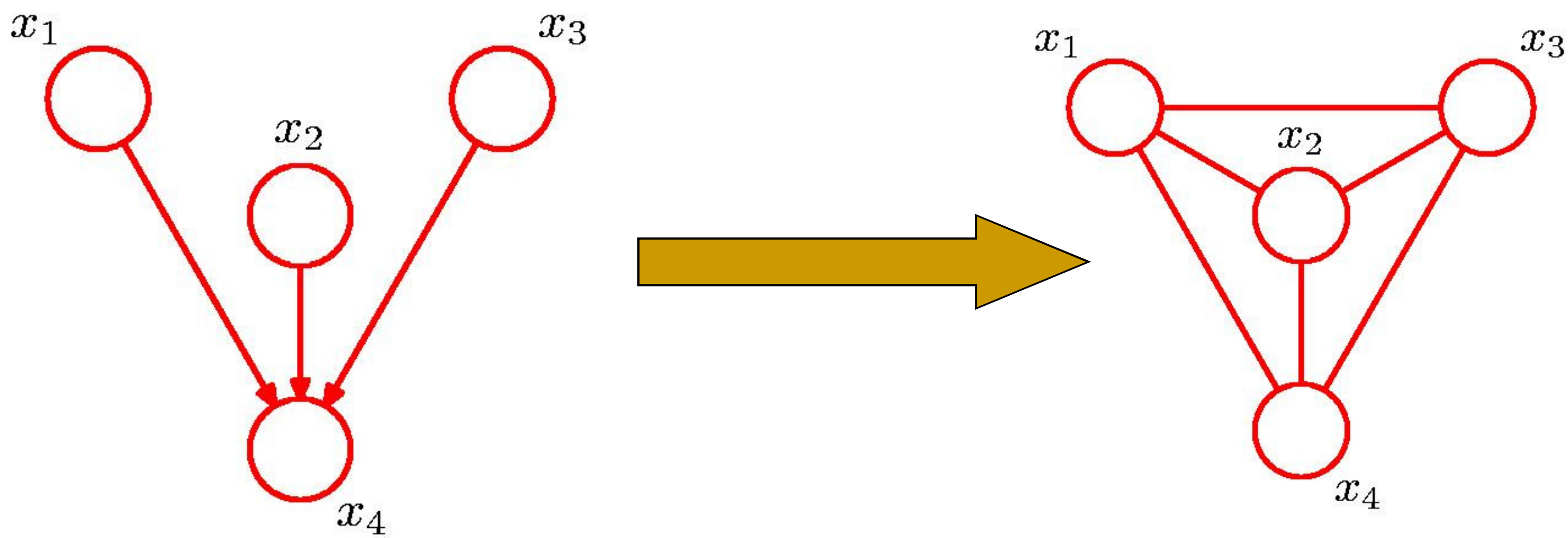
$$\psi_C(\mathbf{x}_C) = \exp\{-E(\mathbf{x}_C)\}$$

Directed-> Undirected Graph



$$\begin{aligned}
\psi_{1,2}(x_1, x_2) &= p(x_1)p(x_2|x_1) \\
\psi_{2,3}(x_2, x_3) &= p(x_3|x_2) \\
&\vdots \\
\psi_{N-1,N}(x_{N-1}, x_N) &= p(x_N|x_{N-1})
\end{aligned}$$

Moralization (marrying the parents)

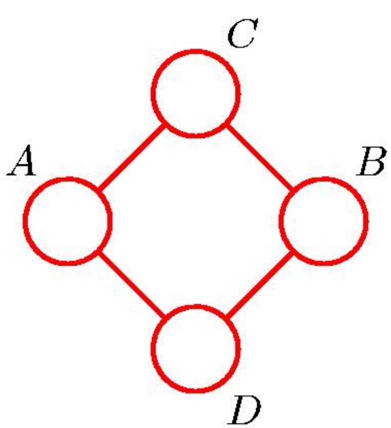


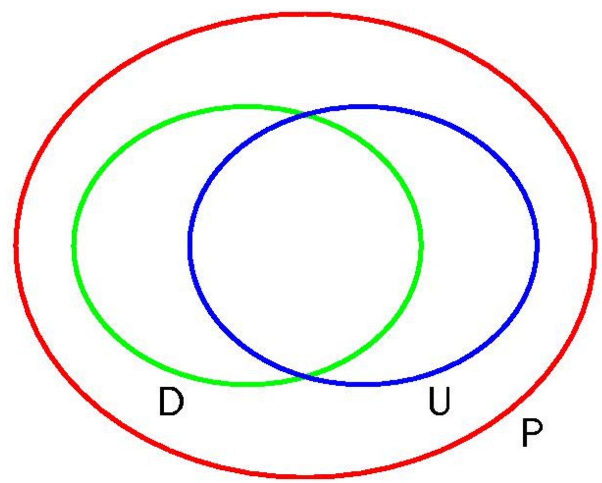
$$p(\mathbf{x}) = p(x_1)p(x_2)p(x_3)p(x_4|x_1, x_2, x_3).$$

Moralization adds the fewest extra links but remains the maximum number of independence properties.

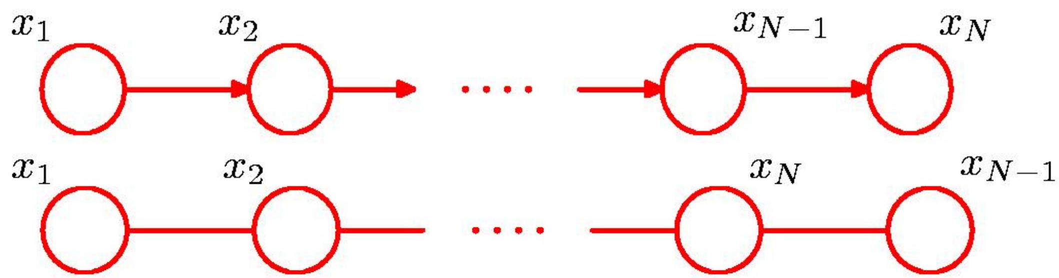
Perfect Map

- Every independence property of the distribution is reflected in the graph and vice versa, then the graph is a perfect map.
- Not all *directed graph* can not be represented as *undirected graph.* (As in previous example)
- Not all *undirected graph* can be represented as *directed graph.*





Inference on a Chain(1)



$$p(\mathbf{x}) = \frac{1}{Z}\psi_{1,2}(x_1, x_2)\psi_{2,3}(x_2, x_3)\cdots\psi_{N-1,N}(x_{N-1}, x_N).$$

$$p(x_n) = \sum_{x_1}\cdots\sum_{x_{n-1}}\sum_{x_{n+1}}\cdots\sum_{x_N} p(\mathbf{x}).$$

N variables, each one has K states, then O(K^(N-1))

Inference on a Chain(2)

$$p(x_n) = \frac{1}{Z}$$

$$\underbrace{\left[\sum_{x_{n-1}} \psi_{n-1,n}(x_{n-1}, x_n) \cdots \left[\sum_{x_2} \psi_{2,3}(x_2, x_3) \left[\sum_{x_1} \psi_{1,2}(x_1, x_2)\right]\right] \cdots \right]}_{\mu_\alpha(x_n)}$$

$$\underbrace{\left[\sum_{x_{n+1}} \psi_{n,n+1}(x_n, x_{n+1}) \cdots \left[\sum_{x_N} \psi_{N-1,N}(x_{N-1}, x_N)\right] \cdots \right]}_{\mu_\beta(x_n)}. \qquad (8.52)$$

$$p(x_n) = \frac{1}{Z}\mu_\alpha(x_n)\mu_\beta(x_n).$$

Complexity: O(KN)

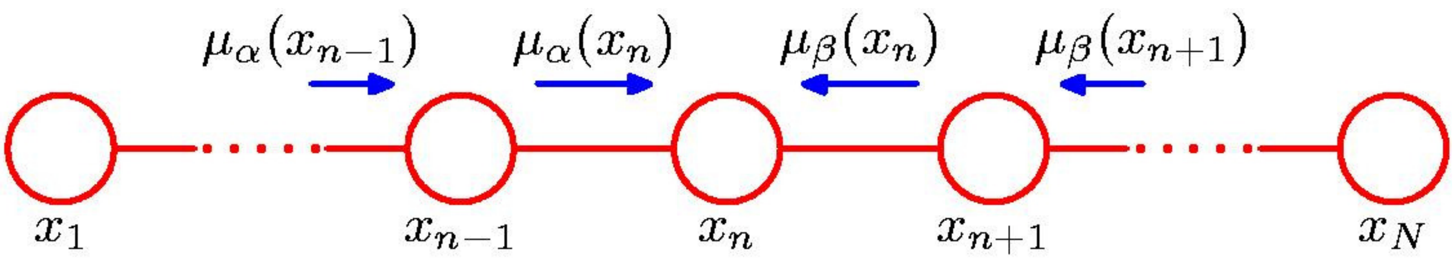
Inference on a Chain(3)

$$\mu_\alpha(x_n) = \sum_{x_{n-1}} \psi_{n-1,n}(x_{n-1}, x_n) \left[\sum_{x_{n-2}} \cdots \right]$$
$$= \sum_{x_{n-1}} \psi_{n-1,n}(x_{n-1}, x_n) \mu_\alpha(x_{n-1}).$$

Message Passed *forwards* along the chain

$$\mu_\beta(x_n) = \sum_{x_{n+1}} \psi_{n+1,n}(x_{n+1}, x_n) \left[\sum_{x_{n+2}} \cdots \right]$$
$$= \sum_{x_{n+1}} \psi_{n+1,n}(x_{n+1}, x_n) \mu_\beta(x_{n+1}).$$

Message Passed *backwards* along the chain



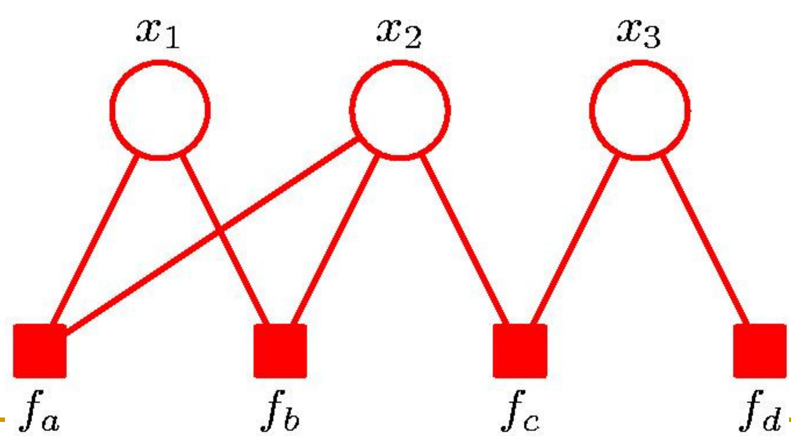
Inference on a Chain(4)

- This message passing is more efficient to find the marginal distributions of all variables.
- If some of the nodes in the graph are observed, then there is no summation for the corresponding variable.
- If some parameters are not observed, apply EM algorithm (discussed later)

Factor Graph

- We can apply similar strategy (message passing) to undirected/directed trees and polytrees as well.
- Polytree is a tree that one node has two or more parents.
- In a factor graph, a node (circle) represents a variable, and additional nodes (squares) represents a factor.

$$p(\mathbf{x}) = f_a(x_1, x_2) f_b(x_1, x_2) f_c(x_2, x_3) f_d(x_3).$$



Factor Graph is not unique

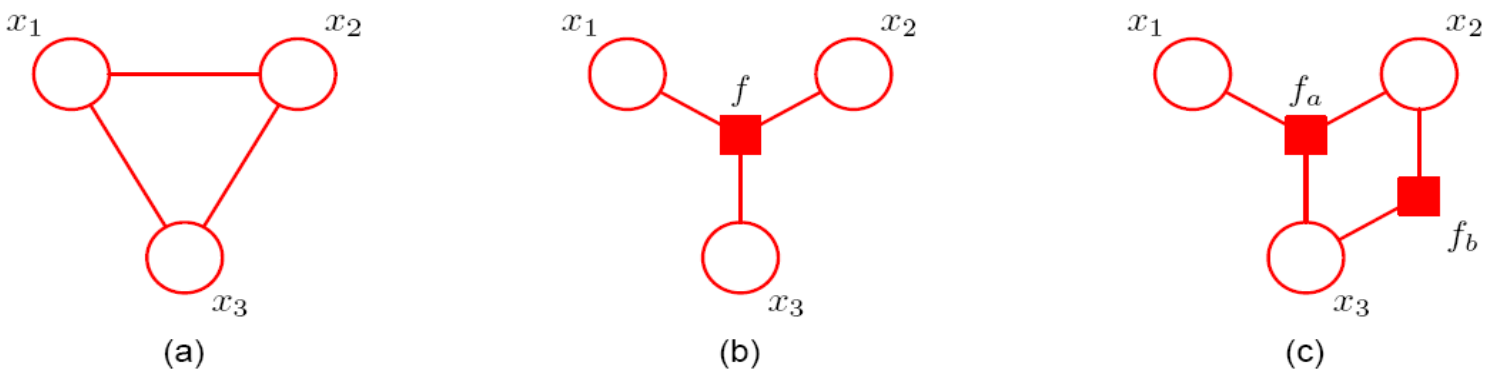


Figure 8.41 (a) An undirected graph with a single clique potential $\psi(x_1, x_2, x_3)$. (b) A factor graph with factor $f(x_1, x_2, x_3) = \psi(x_1, x_2, x_3)$ representing the same distribution as the undirected graph. (c) A different factor graph representing the same distribution, whose factors satisfy $f_a(x_1, x_2, x_3)f_b(x_1, x_2) = \psi(x_1, x_2, x_3)$.

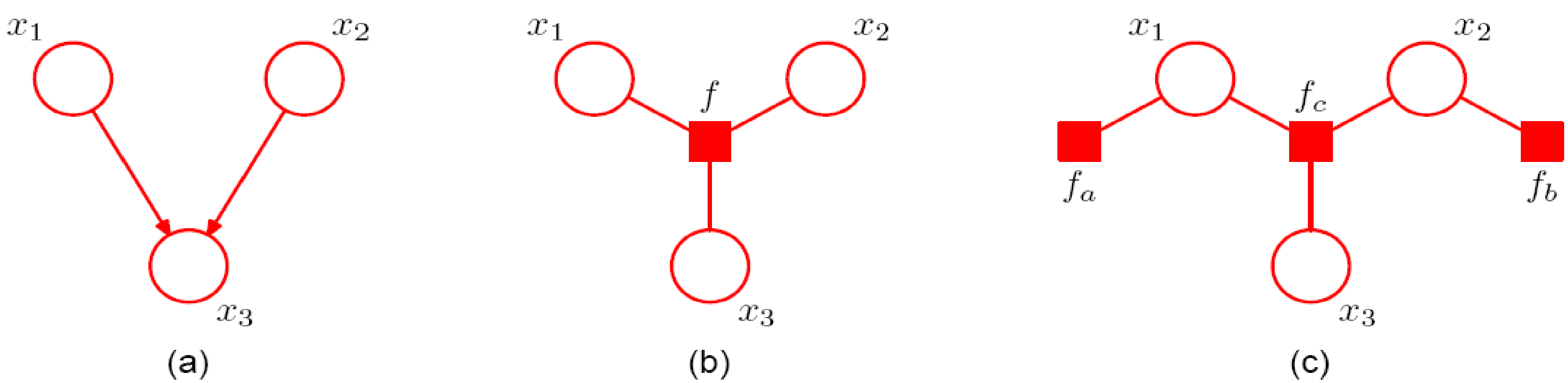


Figure 8.42 (a) A directed graph with the factorization $p(x_1)p(x_2)p(x_3|x_1, x_2)$. (b) A factor graph representing the same distribution as the directed graph, whose factor satisfies $f(x_1, x_2, x_3) = p(x_1)p(x_2)p(x_3|x_1, x_2)$. (c) A different factor graph representing the same distribution with factors $f_a(x_1) = p(x_1)$, $f_b(x_2) = p(x_2)$ and $f_c(x_1, x_2, x_3) = p(x_3|x_1, x_2)$.

A poly tree example

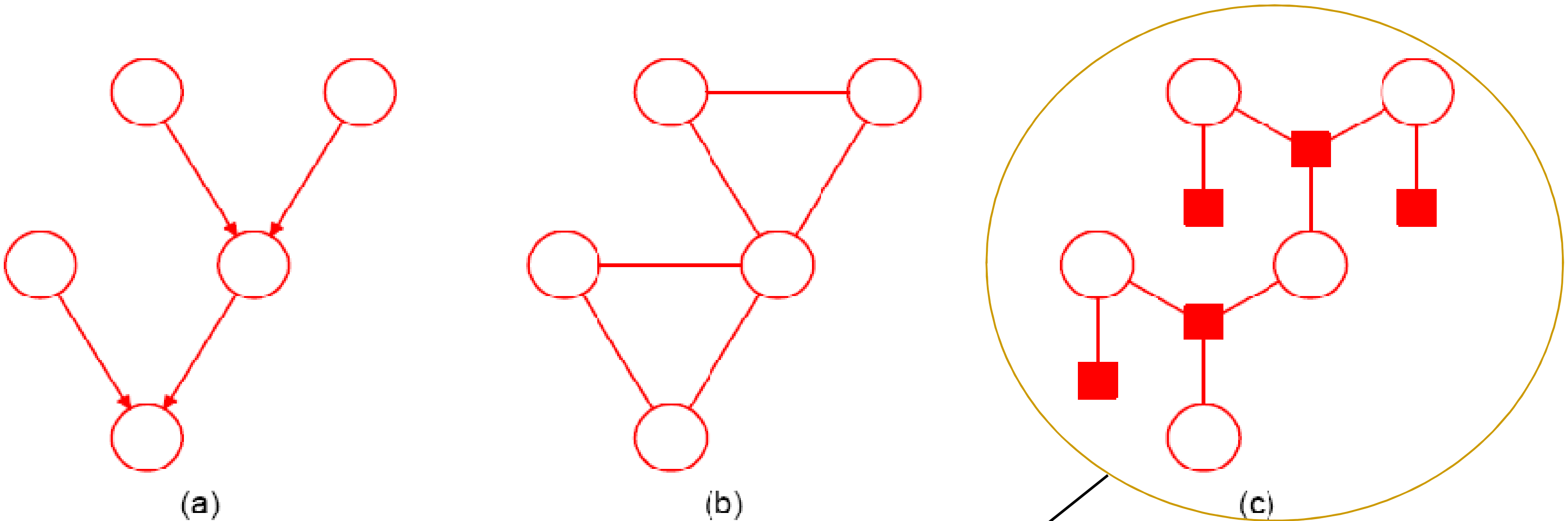


Figure 8.43 (a) A directed polytree. (b) The result of converting the polytree into an undirected graph showing the creation of loops. (c) The result of converting the polytree into a factor graph, which retains the tree structure.

It is still a tree without loops!!

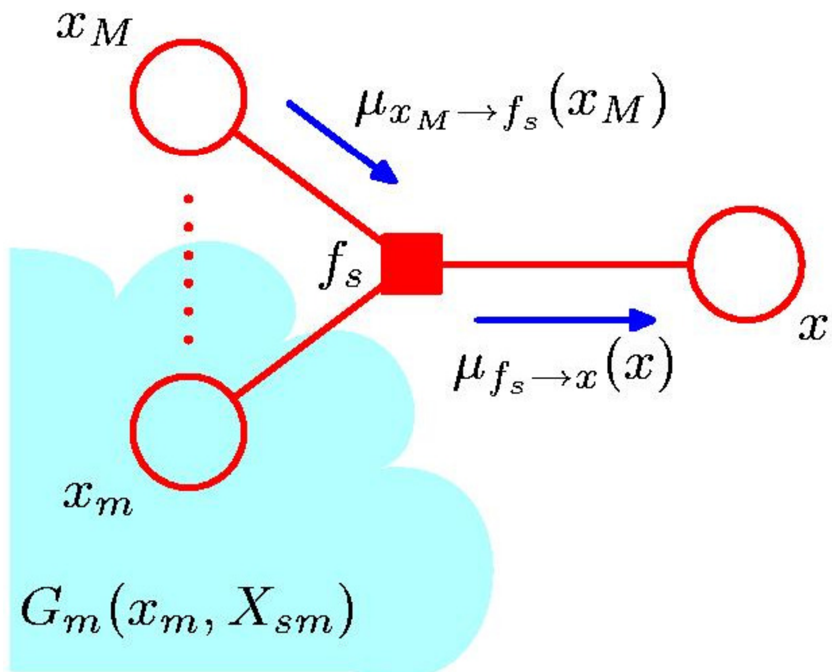
The sum-product algorithm

- This algorithm is the same as *belief propagation* which is proposed for directed graphs without loops.

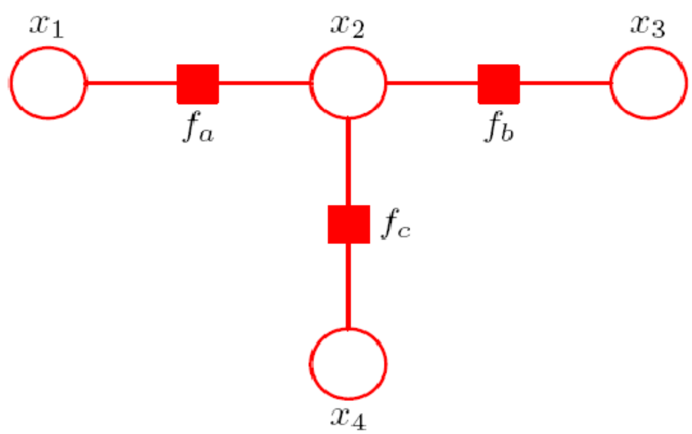
$$p(x) = \sum_{\mathbf{x} \setminus x} p(\mathbf{x})$$

$$p(\mathbf{x}) = \prod_{s \in \mathrm{ne}(x)} F_s(x, X_s)$$

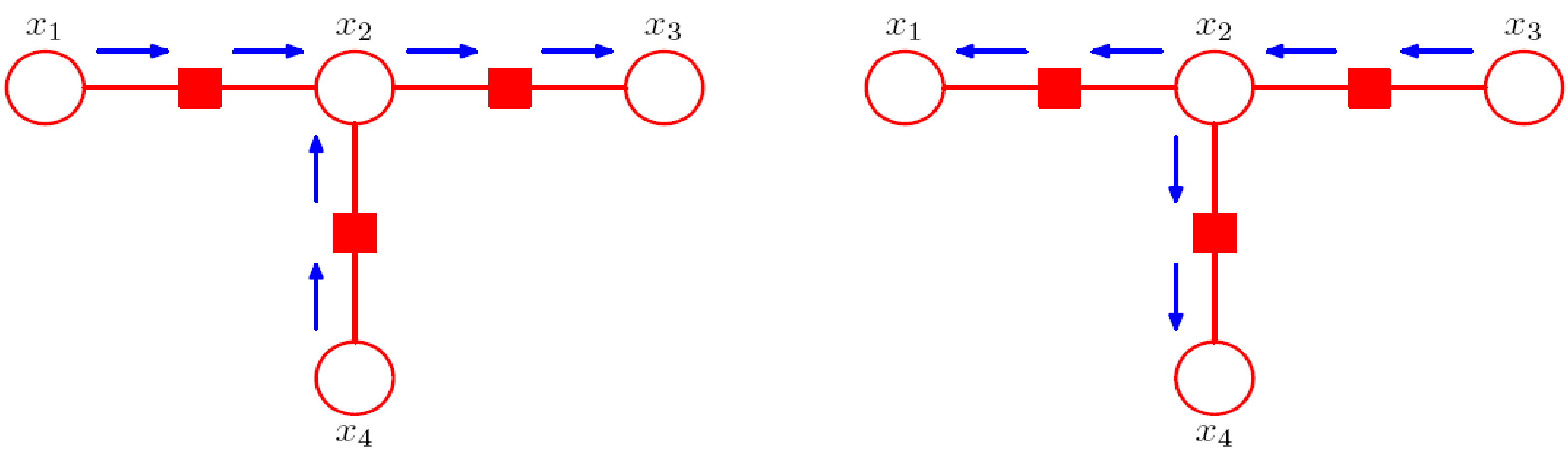
$$p(x) = \prod_{s \in \mathrm{ne}(x)} \left[\sum_{X_s} F_s(x, X_s) \right] = \prod_{s \in \mathrm{ne}(x)} \mu_{f_s \to x}(x).$$



An intuitive Example



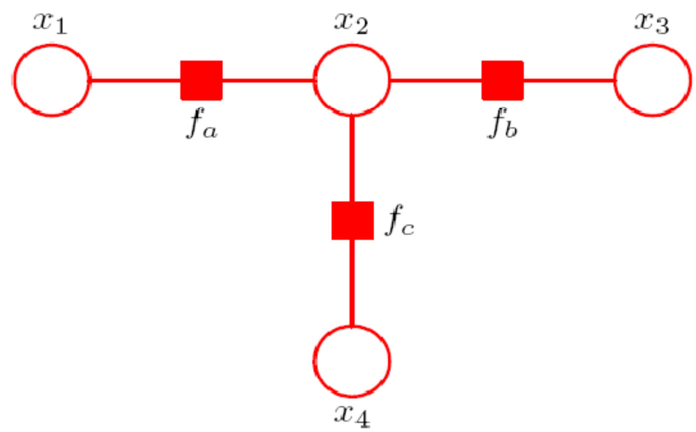
$$\widetilde{p}(\mathbf{x}) = f_a(x_1, x_2) f_b(x_2, x_3) f_c(x_2, x_4).$$



(a) (b)

$$\begin{aligned}
\mu_{x_1\to f_a}(x_1) &= 1 \\
\mu_{f_a\to x_2}(x_2) &= \sum_{x_1} f_a(x_1, x_2) \\
\mu_{x_4\to f_c}(x_4) &= 1 \\
\mu_{f_c\to x_2}(x_2) &= \sum_{x_4} f_c(x_2, x_4) \\
\mu_{x_2\to f_b}(x_2) &= \mu_{f_a\to x_2}(x_2)\mu_{f_c\to x_2}(x_2) \\
\mu_{f_b\to x_3}(x_3) &= \sum_{x_2} f_b(x_2, x_3)\mu_{x_2\to f_b}.
\end{aligned}$$

$$\begin{aligned}
\mu_{x_3\to f_b}(x_3) &= 1 \\
\mu_{f_b\to x_2}(x_2) &= \sum_{x_3} f_b(x_2, x_3) \\
\mu_{x_2\to f_a}(x_2) &= \mu_{f_b\to x_2}(x_2)\mu_{f_c\to x_2}(x_2) \\
\mu_{f_a\to x_1}(x_1) &= \sum_{x_2} f_a(x_1, x_2)\mu_{x_2\to f_a}(x_2) \\
\mu_{x_2\to f_c}(x_2) &= \mu_{f_a\to x_2}(x_2)\mu_{f_b\to x_2}(x_2) \\
\mu_{f_c\to x_4}(x_4) &= \sum_{x_2} f_c(x_2, x_4)\mu_{x_2\to f_c}(x_2).
\end{aligned}$$



$$
\begin{aligned}
\widetilde{p}(x_2) &= \mu_{f_a \rightarrow x_2}(x_2)\mu_{f_b \rightarrow x_2}(x_2)\mu_{f_c \rightarrow x_2}(x_2) \\
&= \left[\sum_{x_1} f_a(x_1, x_2)\right]\left[\sum_{x_3} f_b(x_2, x_3)\right]\left[\sum_{x_4} f_c(x_2, x_4)\right] \\
&= \sum_{x_1}\sum_{x_2}\sum_{x_4} f_a(x_1, x_2) f_b(x_2, x_3) f_c(x_2, x_4) \\
&= \sum_{x_1}\sum_{x_3}\sum_{x_4} \widetilde{p}(\mathbf{x})
\end{aligned}
$$

in General Graphs

- Exact Inference: Junction tree algorithm.
- Inexact inference:
- No closed form for the distribution.
- Dimensionality of latent space is too high.