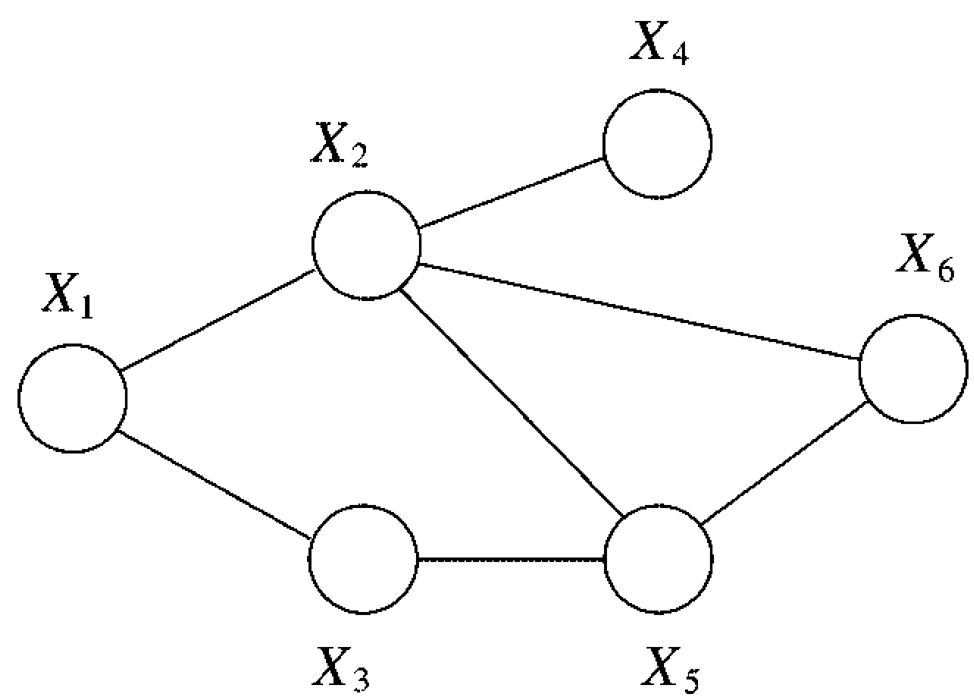
Artificial Intelligence
EECS 491

Undirected Graphical Models

Undirected graphical models: Markov Networks

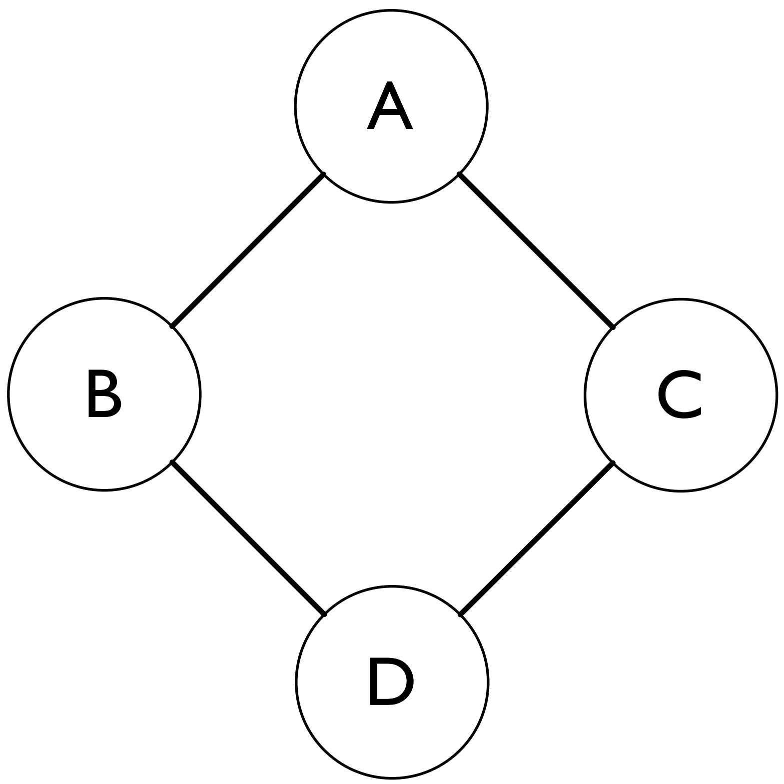
- *Undirected graphical models* are a different way of factorizing a joint probability density.
- Unlike Bayesian belief networks, undirected graphical models are useful when there is no clear con, e.g. relationships among pixels in images, language models.
- Absence of a links between nodes indicates *independence* of those two variables.
- This general approach is to represent the structure of a joint probability distribution in terms of independent factors represented by *potential functions*.



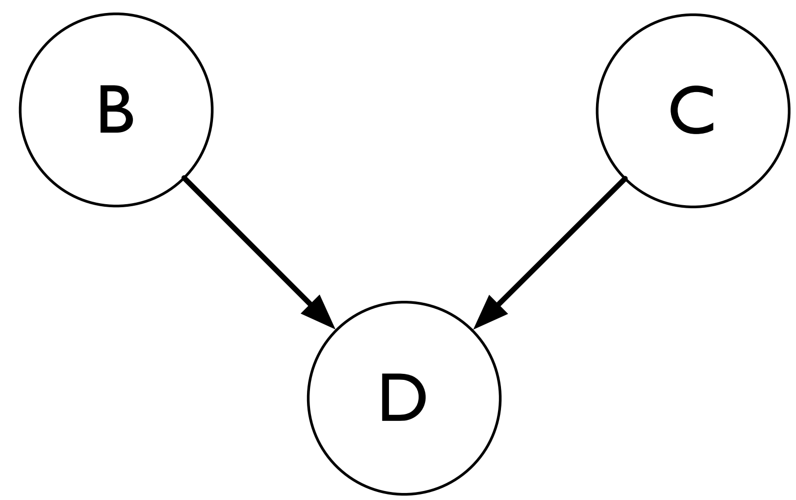
$$p(x_{\mathcal{V}}) = \frac{1}{Z}\psi(x_1, x_2)\psi(x_1, x_3)\psi(x_2, x_4)\psi(x_3, x_5)\psi(x_2, x_5, x_6)$$

Directed and undirected graphical models

- DGs and UGs define different dependence relationships.
- There are families of probability distributions that are captured by DGs but not any UG and vice versa.



No directed graph can represent only:
$A \perp D \mid \{B,C\}$
$B \perp C \mid \{A,D\}$



No undirected graph can represent only:
$B \perp C$

Why?

How do we parameterize the relationships?

- Why can't we use a simple conditional parameterization, where the joint probability is a product of the conditional probability of each node given its neighbors:

$$p(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \prod_i p(\mathbf{x}_i | \mathrm{neigh}(\mathbf{x}_i))$$

- This is a product of functions, which factorizes the distribution, but...
- Multiplying conditional densities does not, in general, yield valid joint probability distributions.
- What about products of marginals?

$$p(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \prod_i p(\mathbf{x}_i, \mathrm{neigh}(\mathbf{x}_i))$$

- This too does not yield valid probability distributions.
- Only directed graphs have this property because p(a,b) = p(a|b) p(b).
- Therefore we assume for undirected graphs that the joint distribution factorizes into arbitrarily defined *potential* functions, ψ(x).

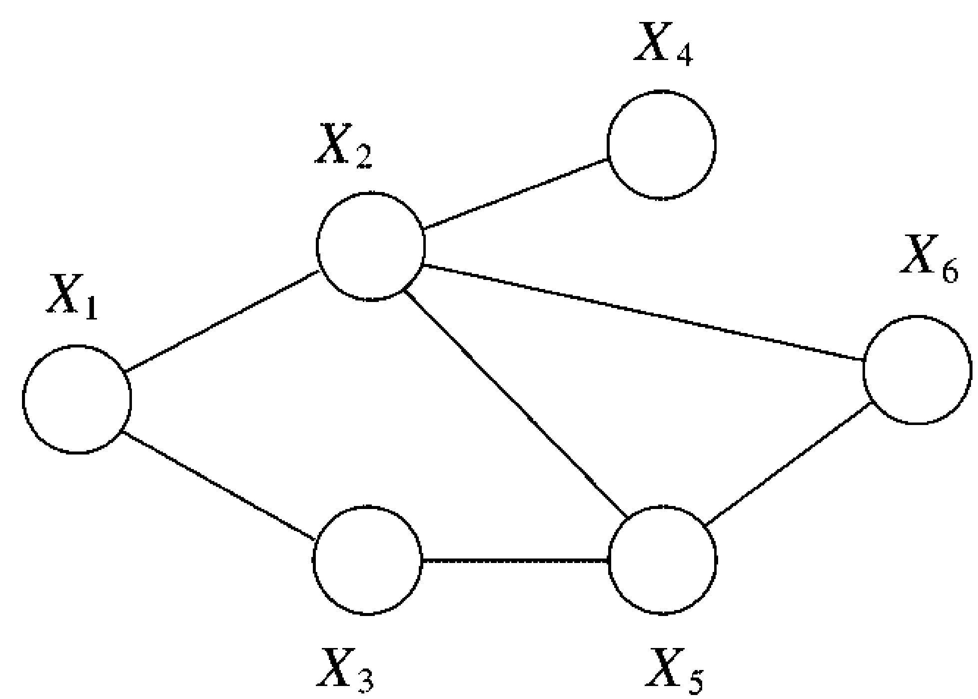
The joint pdf for a Markov net is a product of potential functions

- The joint probability is a product of *clique* potential functions:

$$p(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \frac{1}{Z} \prod_{\text{cliques } c} \psi_c(\mathbf{x}_c)$$

- Where each $\psi_c(\mathbf{x}_c)$ is an arbitrary positive function of it's arguments.
- The set of cliques is the set of maximal complete subgraphs.
- Z is a normalization constant that defines a valid joint pdf, and is sometimes called the *partition function*.

$$Z = \sum_{\mathbf{x}} \prod_{\text{cliques } c} \psi_c(\mathbf{x}_c)$$

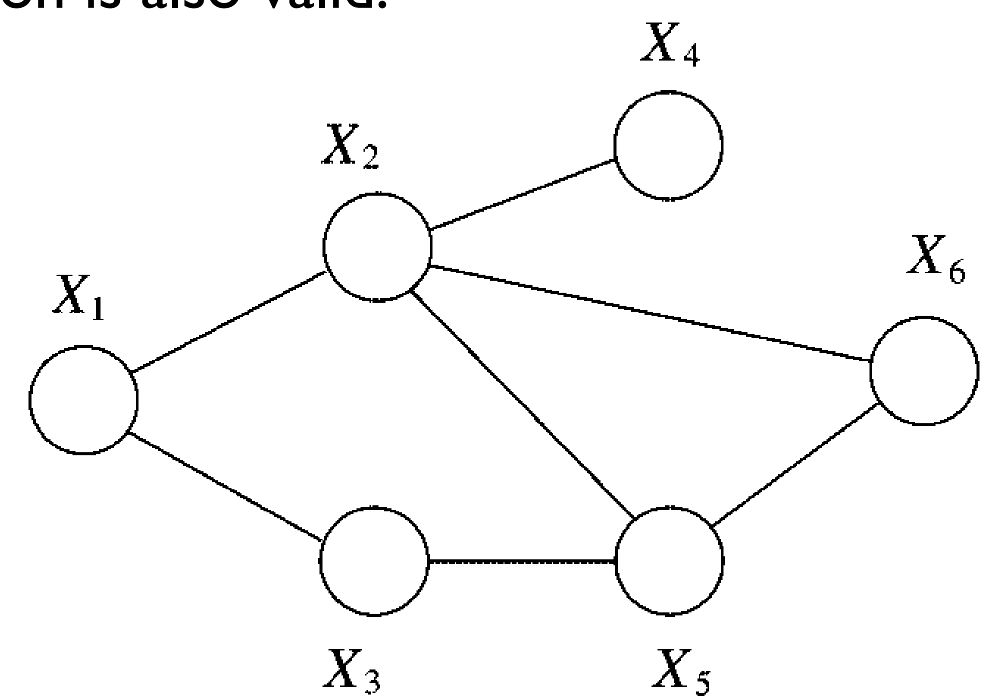


This is a greatly reduced representation.

$$p(x_{\mathcal{V}}) = \frac{1}{Z}\psi(x_1, x_2)\psi(x_1, x_3)\psi(x_2, x_4)\psi(x_3, x_5)\psi(x_2, x_5, x_6)$$

Multiple definitions are possible

- It is not necessary to restrict the definition to maximal cliques
- The following definition is also valid:



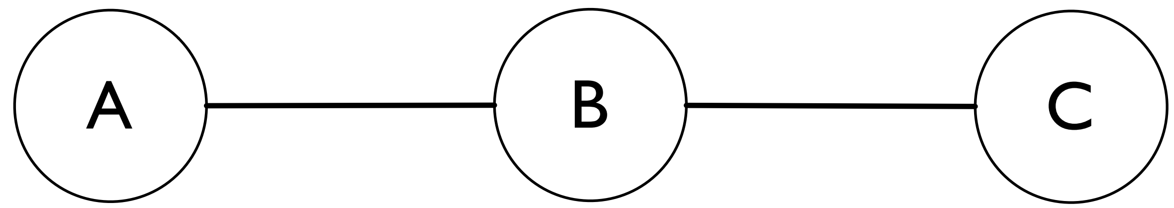
$$p(X) = \frac{1}{Z}\psi(x_1, x_2)\psi(x_1, x_3)\psi(x_2, x_4)\psi(x_3, x_5)\psi(x_2, x_5)\psi(x_2, x_6)\psi(x_5, x_6)$$

- Here, we have assumed a factorization in terms of 2D densities.
- Why can we do this?

This is equivalent to assuming that $\psi(x_2, x_5, x_6)$ factors.

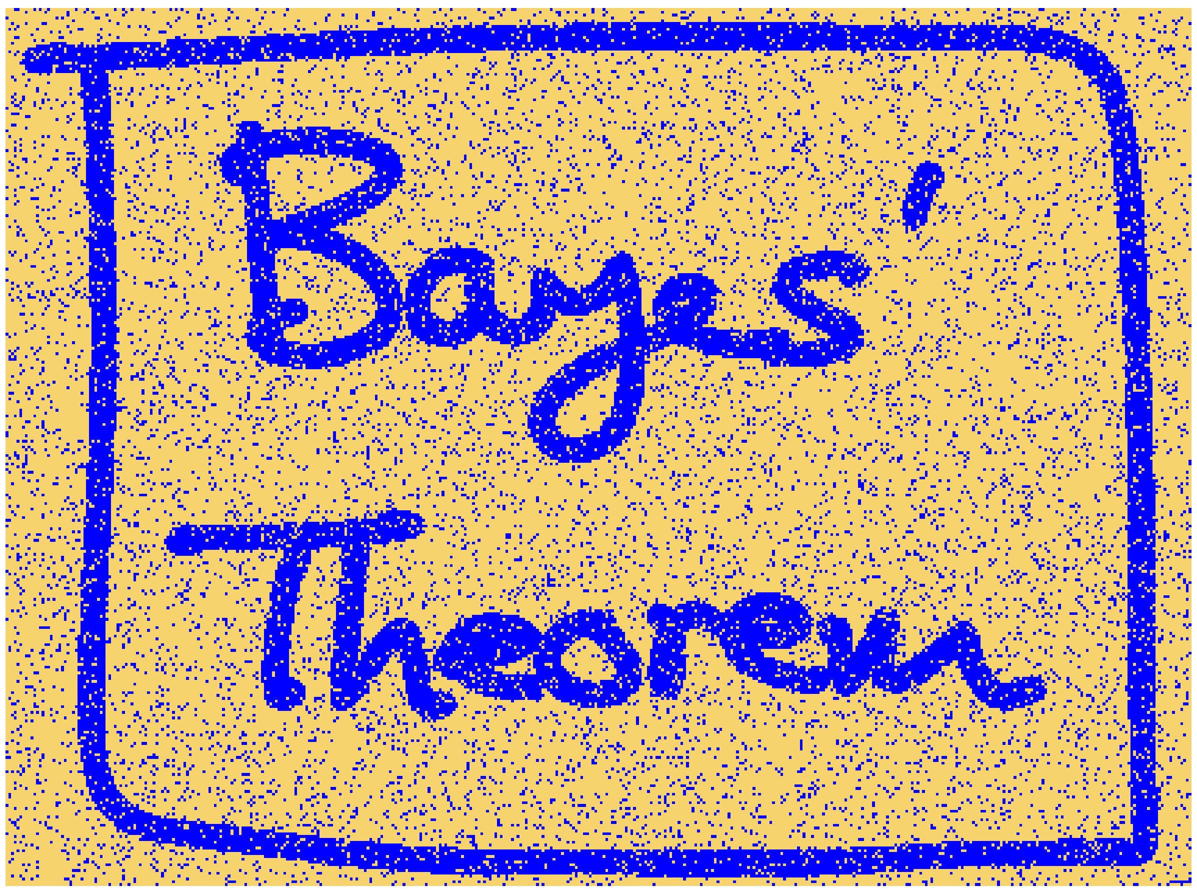
What are the clique potential functions?

- Consider the following model:

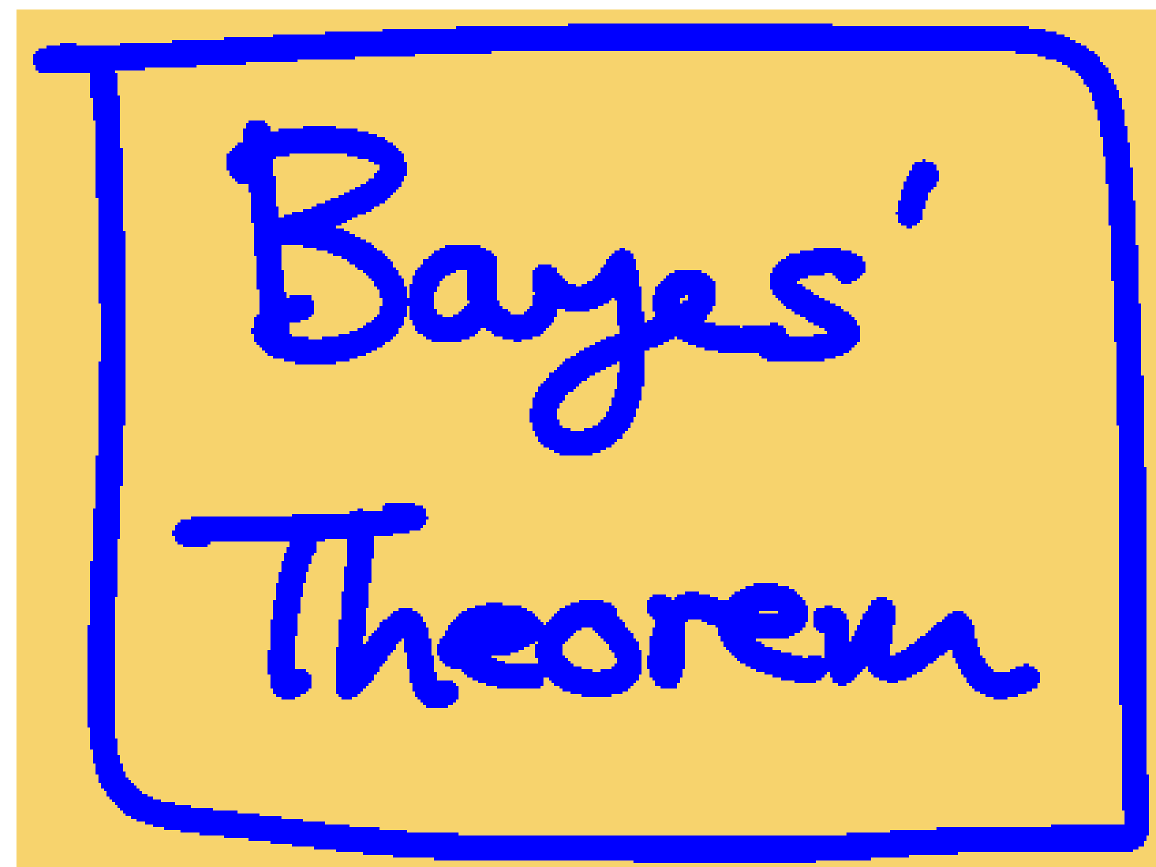


- The model specifies that $A \perp C \mid B$.
- The joint distribution can then be written as
  - $p(A,B,C) = p(B)\, p(A|B)\, p(C|B)$
- This can be written in two ways:
  - $p(A,B,C) = p(A,B)\, p(C|B) = \psi_1(A,B)\, \psi_2(B,C)$
  - $p(A,B,C) = p(A|B)\, p(B,C) = \psi_3(A,B)\, \psi_4(B,C)$
- This shows that the potential functions cannot both be marginals or both be conditionals.
- In general, the clique potential functions *do not* represent probability distributions. They are simply factors in the joint pdf.

How do we recover the original image?

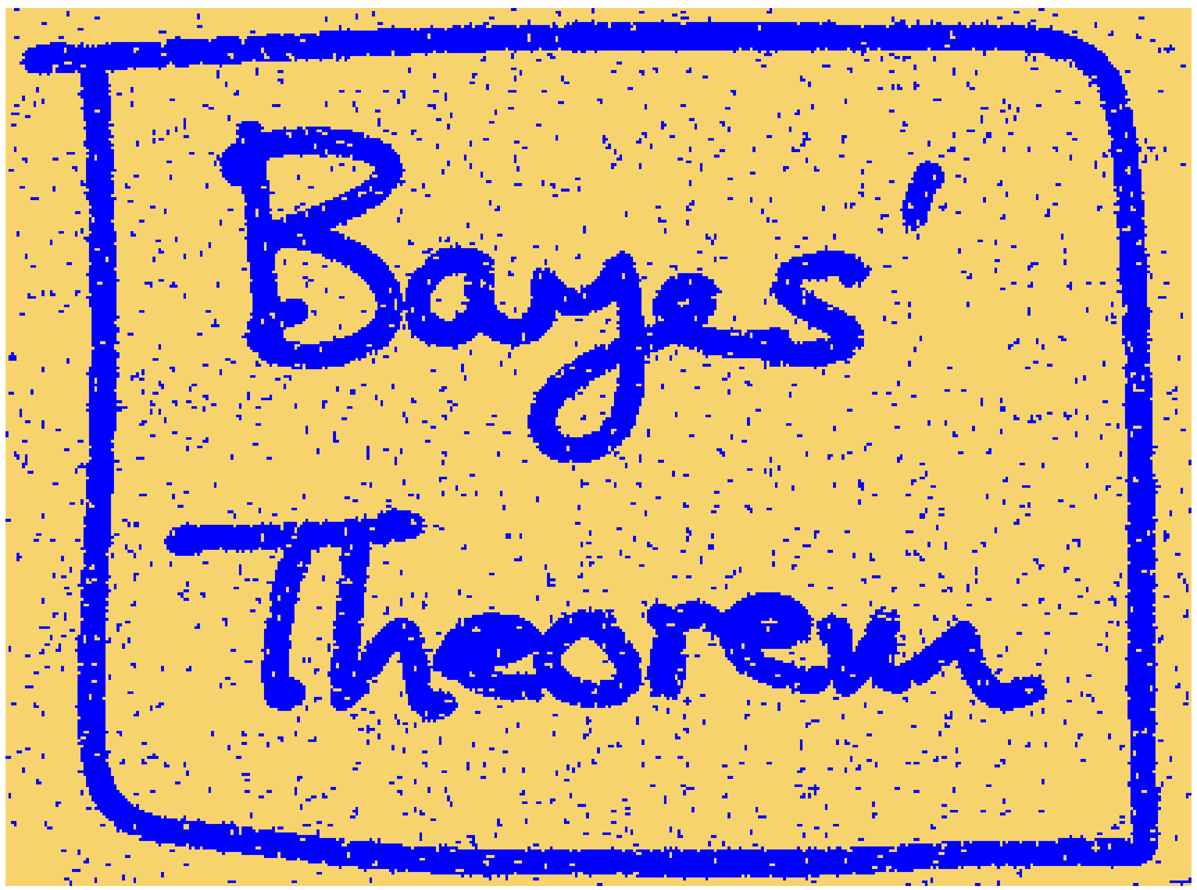


Noisy image

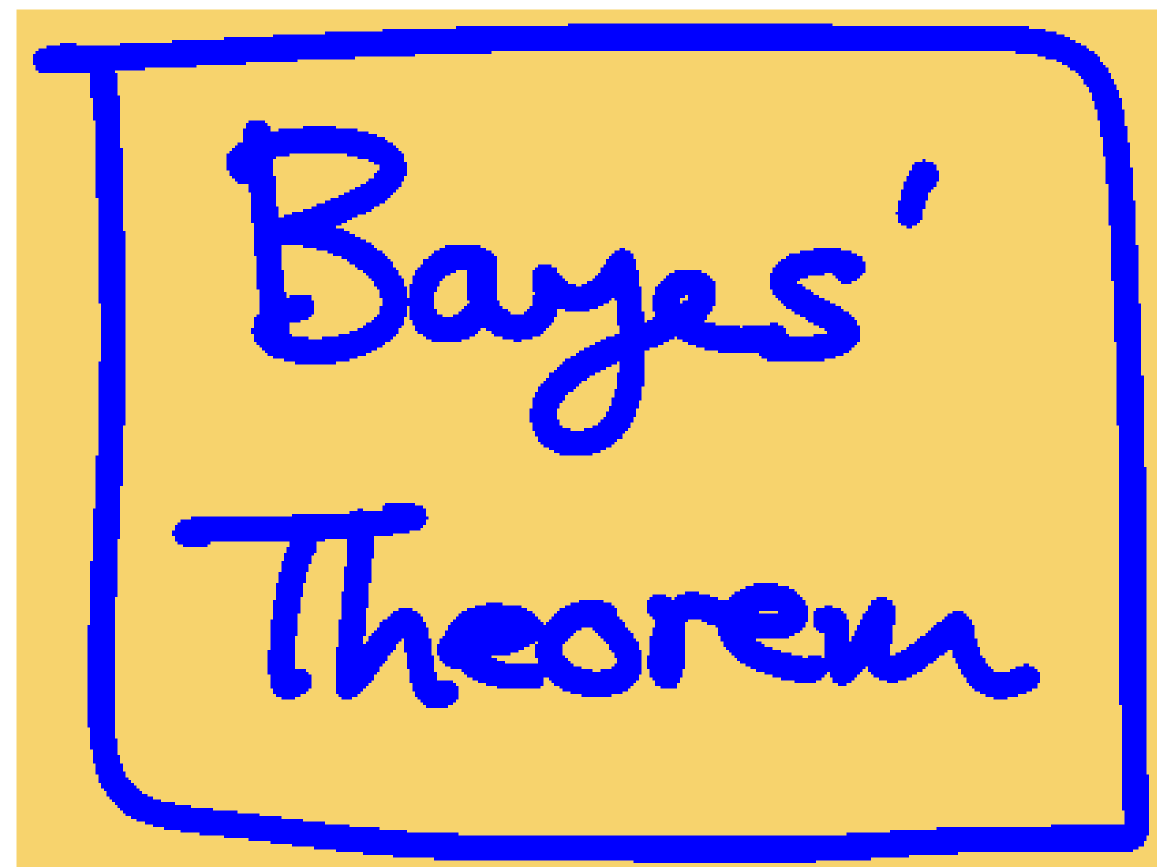


Original image

How do we recover the original image?



Recovered image



Original image

Factor graphs

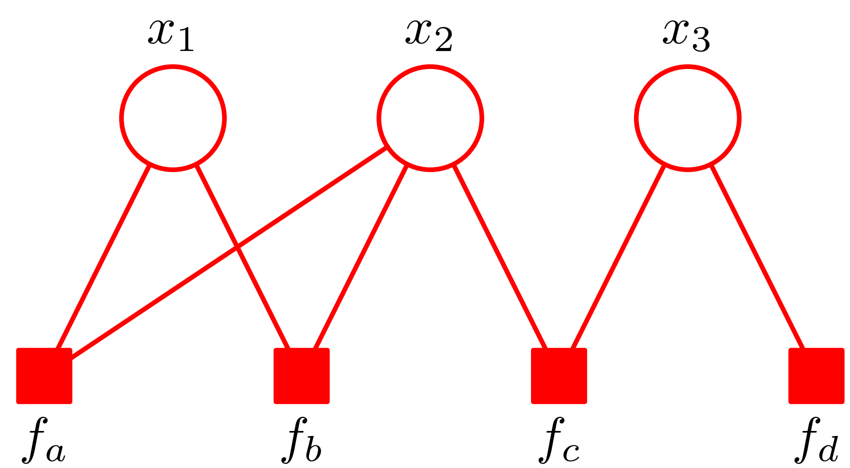
- Both directed and undirected graphs decompose the joint distribution in products.

- More generally, we can write a joint distribution as a product of factors:

$$p(\mathbf{x}) = \prod_s f_s(\mathbf{x}_s)$$

- where x_s denotes a subset of the variables.
- For:
  - directed graphs: fs = local conditional distribution
  - undirected graphs: fs = potential functions over maximal cliques
- Factor graphs keep all factors explicit.
  - a node for every variable
  - additional nodes for every factor

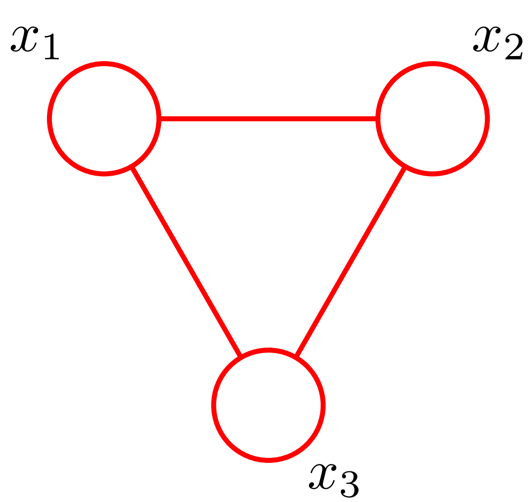
Factor graphs



- Variables are depicted by circles.
- Factors $f_s(x_s)$ are depicted by squares.
- The corresponding joint distribution for the above graph is

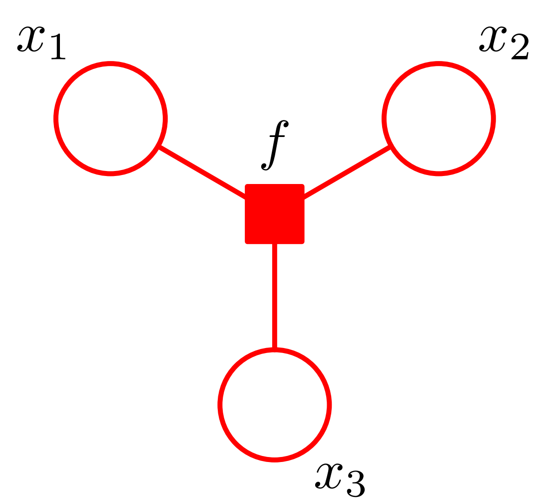
$$p(\mathbf{x}) = f_a(x_1, x_2) f_b(x_1, x_2) f_c(x_2, x_3) f_d(x_3).$$

- What is each joint distribution for the following graphs?



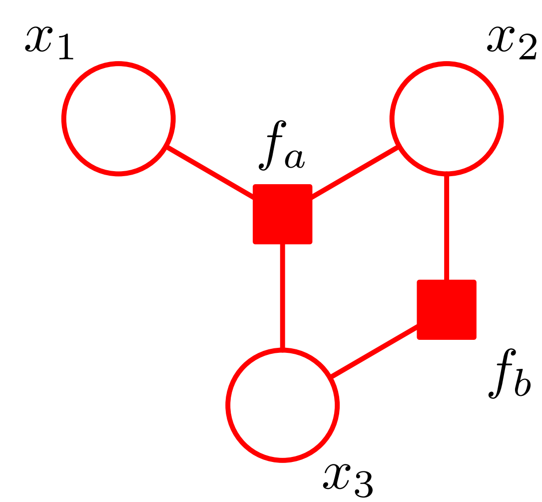
(a)

$\psi(x_1, x_2, x_3)$



(b)

$f(x_1, x_2, x_3) = \psi(x_1, x_2, x_3)$

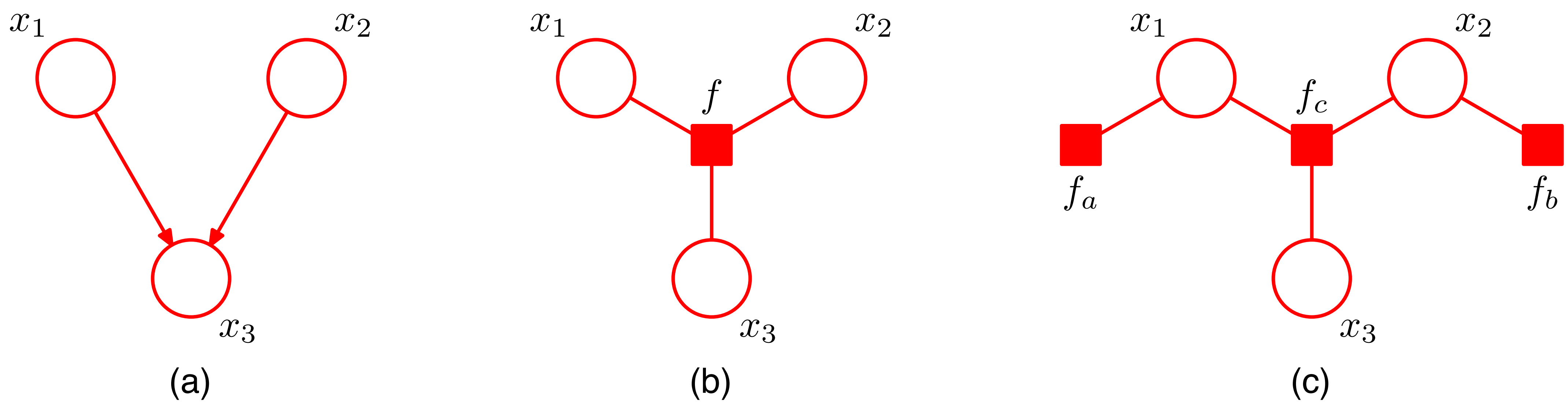


(c)

$f_a(x_1, x_2, x_3) f_b(x_1, x_2) = \psi(x_1, x_2, x_3)$

Factor graphs

- We can also represent directed graphical models with factor graphs:



$p(x_1)p(x_2)p(x_3|x_1, x_2)$

$$f(x_1, x_2, x_3) = p(x_1)p(x_2)p(x_3|x_1, x_2)$$

$$f_a(x_1) = p(x_1), \; f_b(x_2) = p(x_2)$$
$$f_c(x_1, x_2, x_3) = p(x_3|x_1, x_2)$$

- factor graphs are *bipartite* because the use two distinct kinds of nodes:
  - variable nodes and factor nodes
  - factor nodes only connect to variable nodes and vice versa