

# Factor Graphs

Chris Williams  
(based on slides by Michael U. Gutmann)

Probabilistic Modelling and Reasoning (INFR11134)  
School of Informatics, The University of Edinburgh

Spring Semester 2024

# Recap

- ▶ Undirected and directed graphical models have complementary properties
- ▶ Both encode and (visually) represent statistical independencies (I-maps) and factorisations.
- ▶ For directed graphs with parent sets  $\text{pa}_i$

$$p(x_1, \dots, x_d) = \prod_{i=1}^d p(x_i | \text{pa}_i)$$

- ▶ For undirected graphs with maximal clique sets  $\mathcal{X}_c$

$$p(x_1, \dots, x_d) = \frac{1}{Z} \prod_c \phi_c(\mathcal{X}_c)$$

# Program

1. What are factor graphs?
2. Advantages over directed or undirected graphs?

# Program

## 1. What are factor graphs?

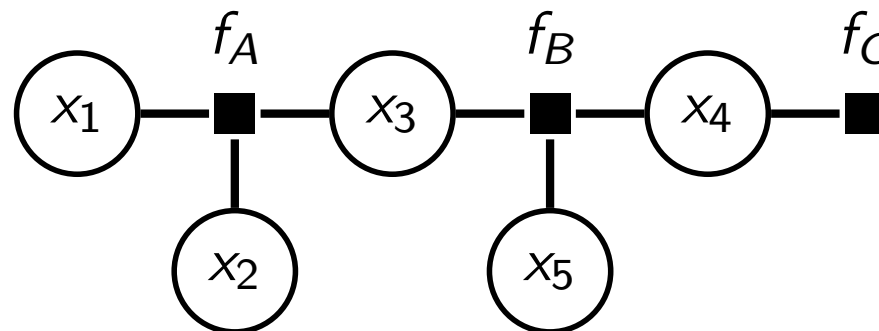
- Definition
- Visualising Gibbs distributions as factor graphs
- Visualising factors that are conditionals

## 2. Advantages over directed or undirected graphs?

# Definition of factor graphs

- ▶ A factor graph represents the factorisation of an arbitrary function (not necessarily related to pdfs/pmfs)
- ▶ Example:  $h(x_1, \dots, x_5) = f_A(x_1, x_2, x_3)f_B(x_3, x_4, x_5)f_C(x_4)$

Factor graph (FG):

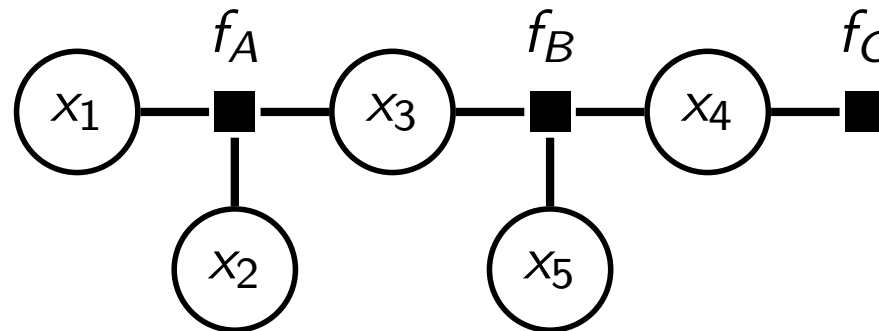


- ▶ Two types of nodes: factor and variable nodes
- ▶ Convention: squares for factors, circles for variables (other conventions are used too)

# Definition of factor graphs

- ▶ Example:  $h(x_1, \dots, x_5) = f_A(x_1, x_2, x_3)f_B(x_3, x_4, x_5)f_C(x_4)$

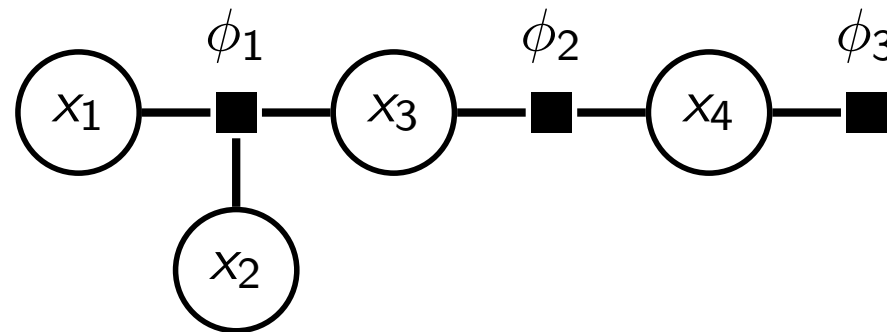
Factor graph (FG):



- ▶ Edge between variable  $x$  and factor  $f \Leftrightarrow x$  is an argument of  $f$
- ▶ Variable nodes are always connected to factor nodes; no direct links between factor or variable nodes (FGs are bipartite graphs)
- ▶ We can also use directed edges (to indicate conditionals)

# Visualising Gibbs distributions as factor graphs

- ▶ Example:  $p(x_1, x_2, x_3, x_4) = \frac{1}{Z} \phi_1(x_1, x_2, x_3) \phi_2(x_3, x_4) \phi_3(x_4)$



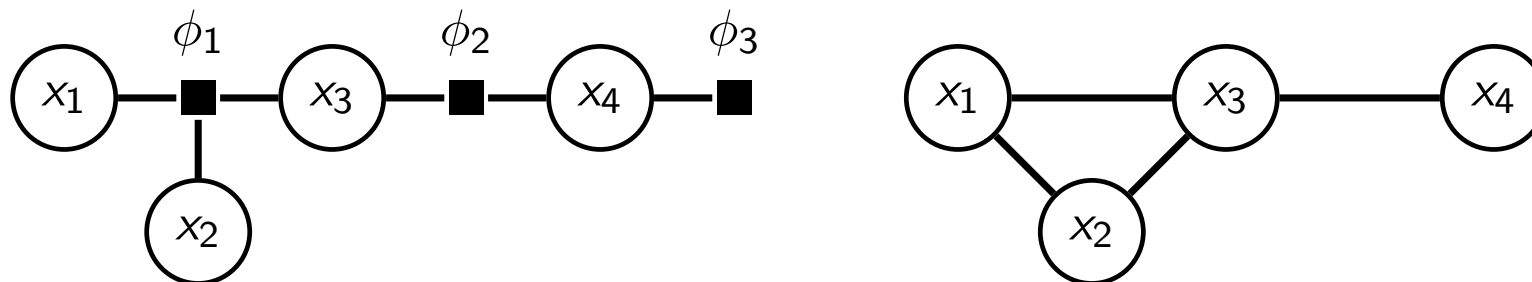
- ▶ General case:  $p(x_1, \dots, x_d) \propto \prod_c \phi_c(\mathcal{X}_c)$ 
  - ▶ Factor node for all  $\phi_c$
  - ▶ For all factors  $\phi_c$ :  
draw an undirected edge between  $\phi_c$  and all  $x_i \in \mathcal{X}_c$ .
- ▶ Can visualise any undirected graphical model as a factor graph.

# Visualising Gibbs distributions as factor graphs

Some differences to visualisation with undirected graph:

- ▶ Factors  $\phi_c$  are shown, which makes the graphs more informative (see next slide).
- ▶ Variables  $x_i$  are neighbours if they are connected to the same factor.

$$p(x_1, x_2, x_3, x_4) = \frac{1}{Z} \phi_1(x_1, x_2, x_3) \phi_2(x_3, x_4) \phi_3(x_4)$$



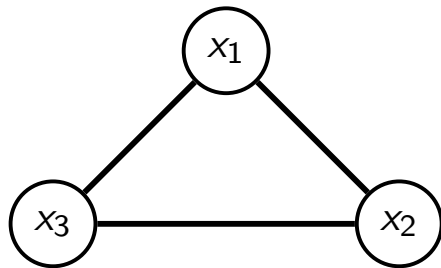


# More informative than undirected graphs

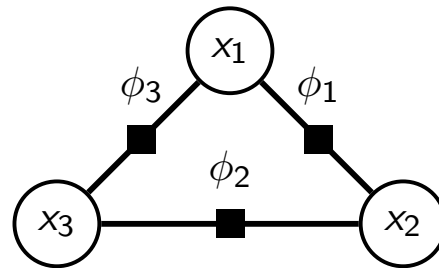
- ▶ Mapping from Gibbs distribution to undirected graph is many to one but one-to-one for factor graphs.
- ▶ Example

$$p_A(x_1, x_2, x_3) \propto \phi_1(x_1, x_2)\phi_2(x_2, x_3)\phi_3(x_3, x_1)$$

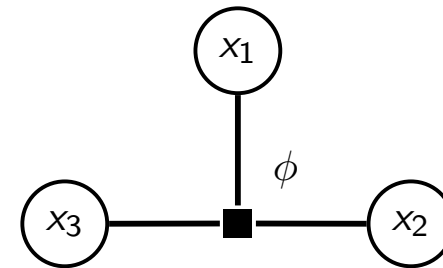
$$p_B(x_1, x_2, x_3) \propto \phi(x_1, x_2, x_3)$$



UG for  $p_A$  and  $p_B$



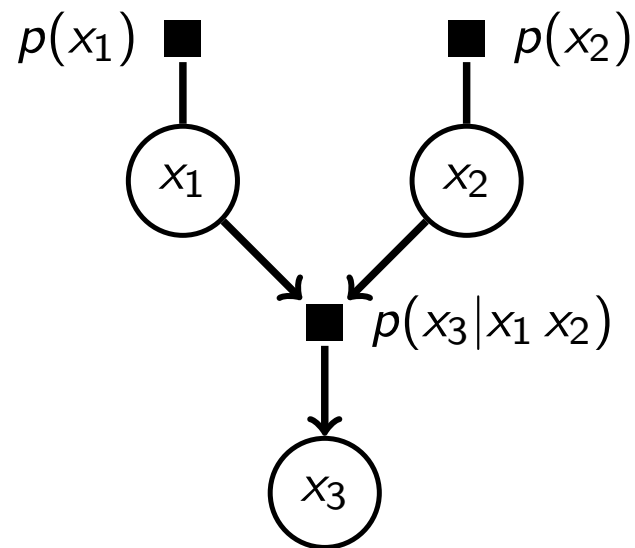
FG for  $p_A$



FG for  $p_B$

# Visualising factors that are conditionals

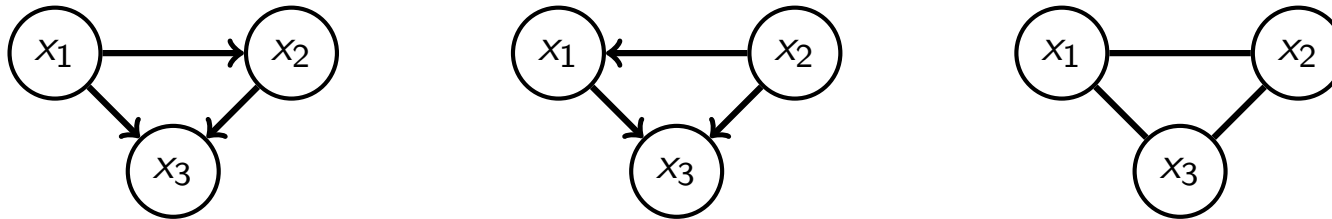
- ▶ For  $p(x_1, x_2, x_3) = p(x_1)p(x_2)p(x_3|x_1, x_2)$ , we may want to include the information that  $x_3$  is conditioned on  $x_1, x_2$
- ▶ Use arrows as in directed graphs.



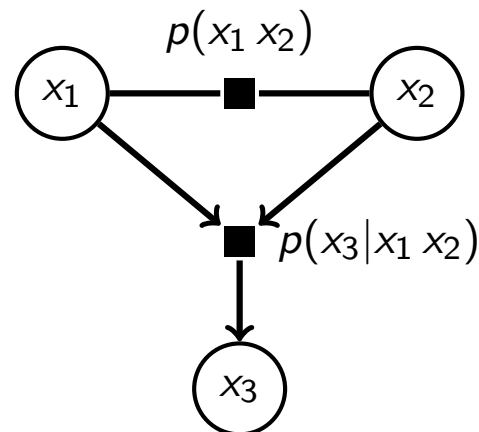
- ▶ Can visualise any directed graphical model as a factor graph.

# Mixed graphs

- ▶ Let  $p(x_1, x_2, x_3) = p(x_1, x_2)p(x_3|x_1, x_2)$ .
- ▶ Directed graphs forces ordering of the random variables; undirected graph does not show conditioning on  $x_1, x_2$



- ▶ Mixed FG to visualise the conditioning for  $p(x_3|x_1, x_2)$  without imposing an ordering on  $x_1$  and  $x_2$



# Program

1. What are factor graphs?
2. Advantages over directed or undirected graphs?
  - Computational advantages
  - Statistical advantages

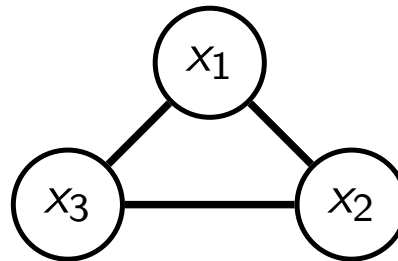
# Importance of factorisation

- ▶ Factorisation was central in the development so far
- ▶ But directed and undirected graphs are not able to fully represent arbitrary factorisations of pdfs/pmfs.

For example, same graph for

$$p(x_1, x_2, x_3) \propto \phi_1(x_1, x_2)\phi_2(x_2, x_3)\phi_3(x_3, x_1)$$

$$p(x_1, x_2, x_3) \propto \phi(x_1, x_2, x_3)$$



- ▶ We should expect that being able to better represent the factorisation has advantages.

# Example of computational advantages

Assume binary random variables  $x_i$ .

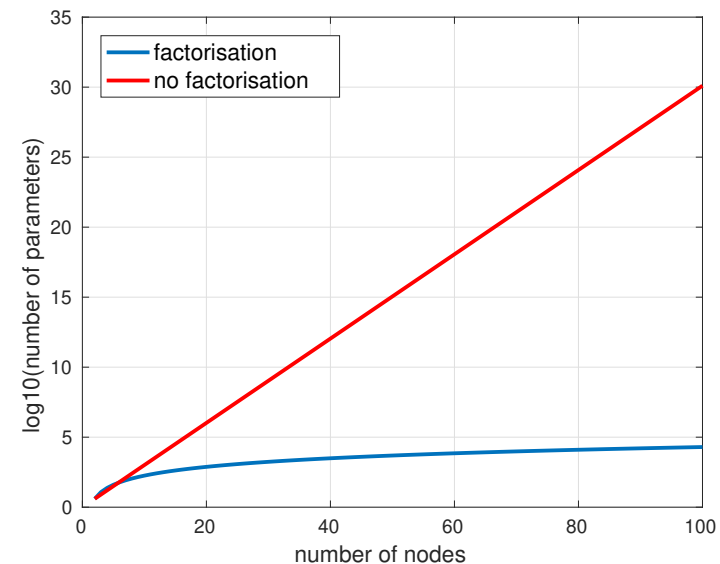
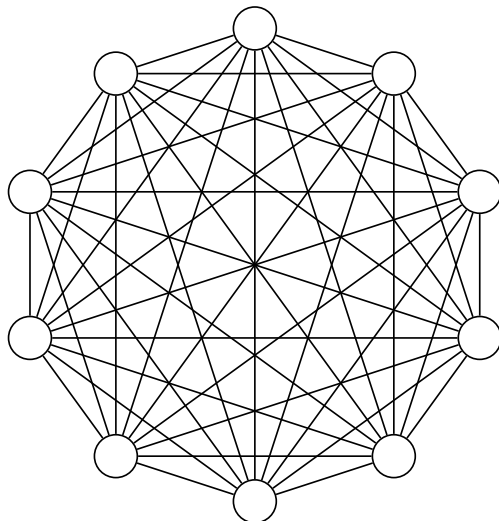
- ▶ Same undirected graph but

$p(x_1, \dots, x_d) \propto \phi(x_1, \dots, x_d)$  has  $2^d$  free parameters,

$p(x_1, \dots, x_d) \propto \prod_{i < j} \phi_{ij}(x_i, x_j)$  has  $\binom{d}{2} 2^2$  free parameters

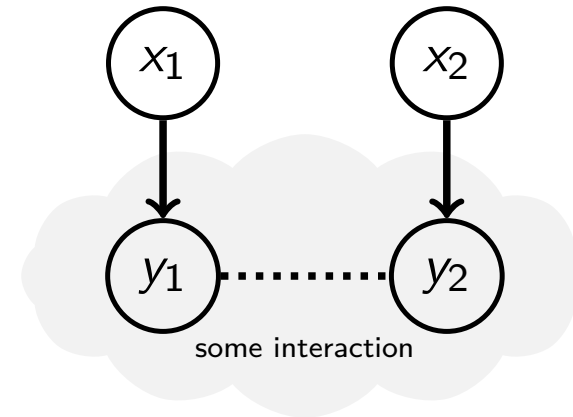
parameters  $\equiv$  entries to specify in a table representation

- ▶ The difference matters for learning and inference when the number of variables is large.



# Example of statistical advantages

- ▶ Let  $x_1$  and  $x_2$  be two inputs
- ▶  $x_1$  controls variable  $y_1$   
 $x_2$  controls  $y_2$
- ▶ Variables  $y_1$  and  $y_2$  influence each other



- ▶ Model:  $p(y_1, y_2, x_1, x_2) = p(y_1, y_2 | x_1, x_2) p(x_1) p(x_2)$   
(probabilistic modelling: pdf/pmf  $p(y_1, y_2 | x_1, x_2)$  captures uncertainty about how the  $x_i$  affect the  $y_i$  and about how the  $y_i$  interact)
- ▶ Choose  $p(y_1, y_2 | x_1, x_2)$  such that  $p(y_1, y_2, x_1, x_2)$  satisfies
  - ▶  $x_1 \perp\!\!\!\perp x_2$  (independence between control variables)
  - ▶  $x_1 \perp\!\!\!\perp y_2 \mid y_1, x_2$  ( $y_2$  is only directly influenced by  $y_1$  and  $x_2$ )
  - ▶  $x_2 \perp\!\!\!\perp y_1 \mid y_2, x_1$  ( $y_1$  is only directly influenced by  $y_2$  and  $x_1$ )

# Example of statistical advantages

- ▶ Three independencies are satisfied if  $p(y_1, y_2|x_1, x_2)$  factorises as

$$p(y_1, y_2|x_1, x_2) = p(y_1|x_1)p(y_2|x_2)\phi(y_1, y_2)n(x_1, x_2)$$

where  $n(x_1, x_2)$  ensures normalisation of  $p(y_1, y_2|x_1, x_2)$

$$n(x_1, x_2) = \left( \int p(y_1|x_1)p(y_2|x_2)\phi(y_1, y_2)dy_1dy_2 \right)^{-1}$$

(see exercises)

- ▶ Directed and undirected graphs cannot represent the independencies induced by factorisation of  $p(y_1, y_2|x_1, x_2)$  (see exercises).
- ▶ Factor graphs and chain graphs (see e.g. Barber, Section 4.3; Lauritzen, Section 3.2.3, not covered in PMR) can represent them.
- ▶ Factor graphs can represent independencies that DAGs or UGs cannot or do not represent (not covered in PMR).



# Program recap

## 1. What are factor graphs?

- Definition
- Visualising Gibbs distributions as factor graphs
- Visualising factors that are conditionals

## 2. Advantages over directed or undirected graphs?

- Computational advantages
- Statistical advantages

# Credits

These slides were produced by Michael Gutmann, made available under Creative Commons licence CC BY 4.0.

©Michael Gutmann, The University of Edinburgh 2018-2024  
CC BY 4.0 .