Undirected Graphical Models II

Independencies

Michael U. Gutmann

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

Spring Semester 2022

Recap

- We have seen that we can visualise factorised pdfs/pmfs $p(\mathbf{x})$ without imposing an ordering or directionality of interaction between the random variables by using an undirected graph.
- When we defined the graph for a pdf/pmf $p(\mathbf{x})$ the numerical values of the factors did not matter; we only used its arguments (inputs).
- This led us to defining a set of probability distributions based on an undirected graph, i.e. an undirected graphical model.

Program

1. Graph separation and the undirected global Markov property

2. Further methods to determine independencies

Program

1. Graph separation and the undirected global Markov property
   - Link between conditioning, graph structure, factorisation, and independencies
   - Graph separation to determine independencies
   - Examples

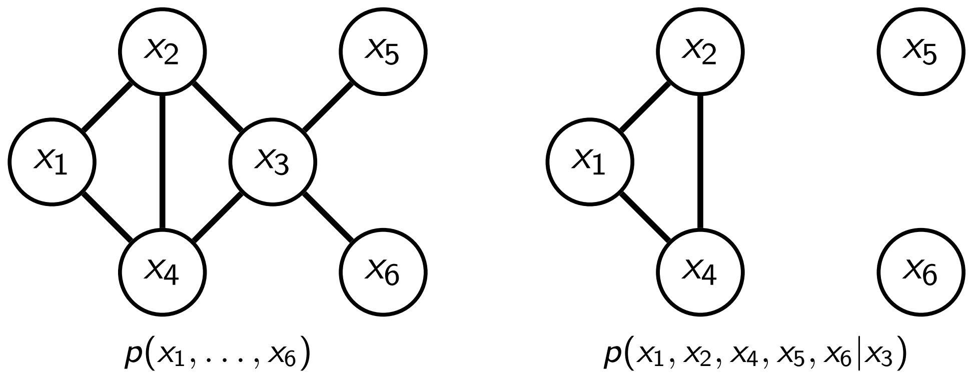
2. Further methods to determine independencies

Motivating the graph separation criterion

- Given an undirected graph H, we defined the undirected graphical model (UGM) to be the set of pdfs/pmfs that factorise as

$$p(x_1, \ldots, x_d) = \frac{1}{Z} \prod_c \phi_c(\mathcal{X}_c), \quad \phi_c \geq 0$$

  where the $\mathcal{X}_c$ correspond to the maximal cliques in the graph.
- We have seen that conditioning on variables corresponds to removing them from the graph (and redefining some factors).
- Combine this with $\mathbf{x} \perp\!\!\!\perp \mathbf{y} \mid \mathbf{z} \iff p(\mathbf{x}, \mathbf{y}, \mathbf{z}) \propto \phi_A(\mathbf{x}, \mathbf{z})\phi_B(\mathbf{y}, \mathbf{z})$

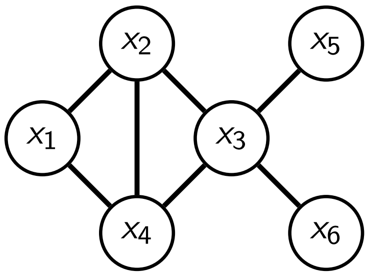


Motivating the graph separation criterion

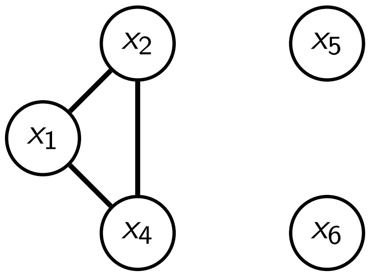
- Example:

$$p(x_1, \ldots, x_6) \propto \underbrace{\phi_1(x_1, x_2, x_4)\phi_2(x_2, x_3, x_4)}_{\phi_A(x_1, x_2, x_4, {\color{red}x_3})} \underbrace{\phi_3(x_3, x_5)\phi_4(x_3, x_6)}_{\phi_B(x_5, x_6, {\color{red}x_3})}$$

- We thus have $(x_1, x_2, x_4) \perp\!\!\!\perp (x_5, x_6) \mid x_3$
- Removing x_3 from the graph blocks all trails between x_5 and x_6, and to all other variables.
- Let us build on this link between conditioning, blocking of trails in the graph, factorisation, and independencies.



$p(x_1, \ldots, x_6)$

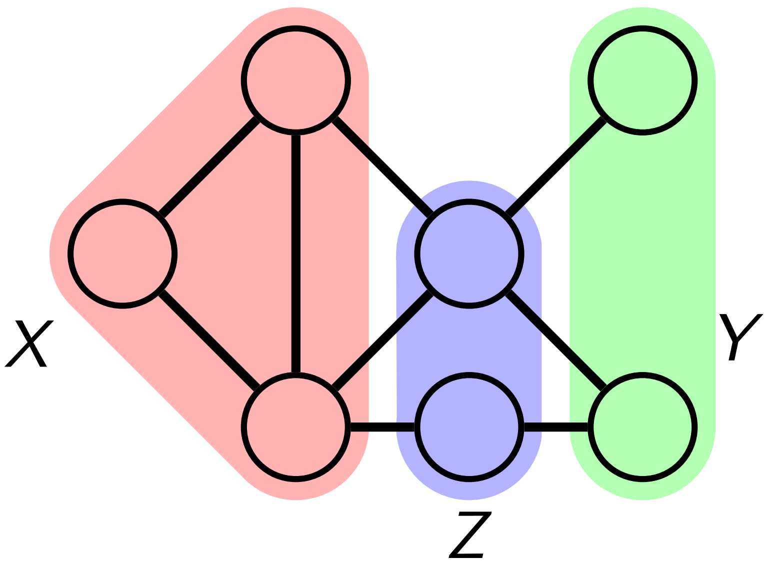


$p(x_1, x_2, x_4, x_5, x_6 | x_3)$

Graph separation

Let X, Y, Z be three disjoint set of nodes in an undirected graph.

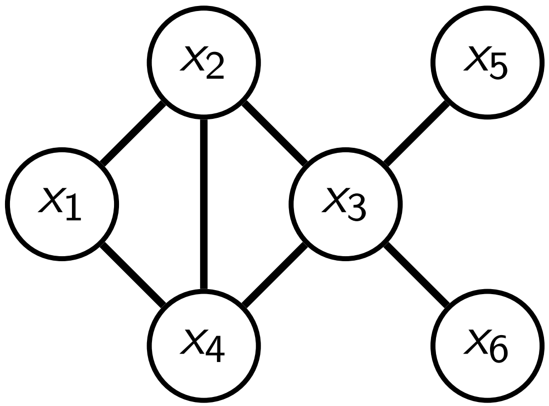
- *Definition* X and Y are separated by Z if every trail from any node in X to any node in Y passes through at least one node of Z.
- In other words:
  - all trails from X to Y are blocked by Z
  - removing Z from the graph leaves X and Y disconnected.
  - Nodes are valves; open by default but closed when part of Z.



Example

In the previous example:

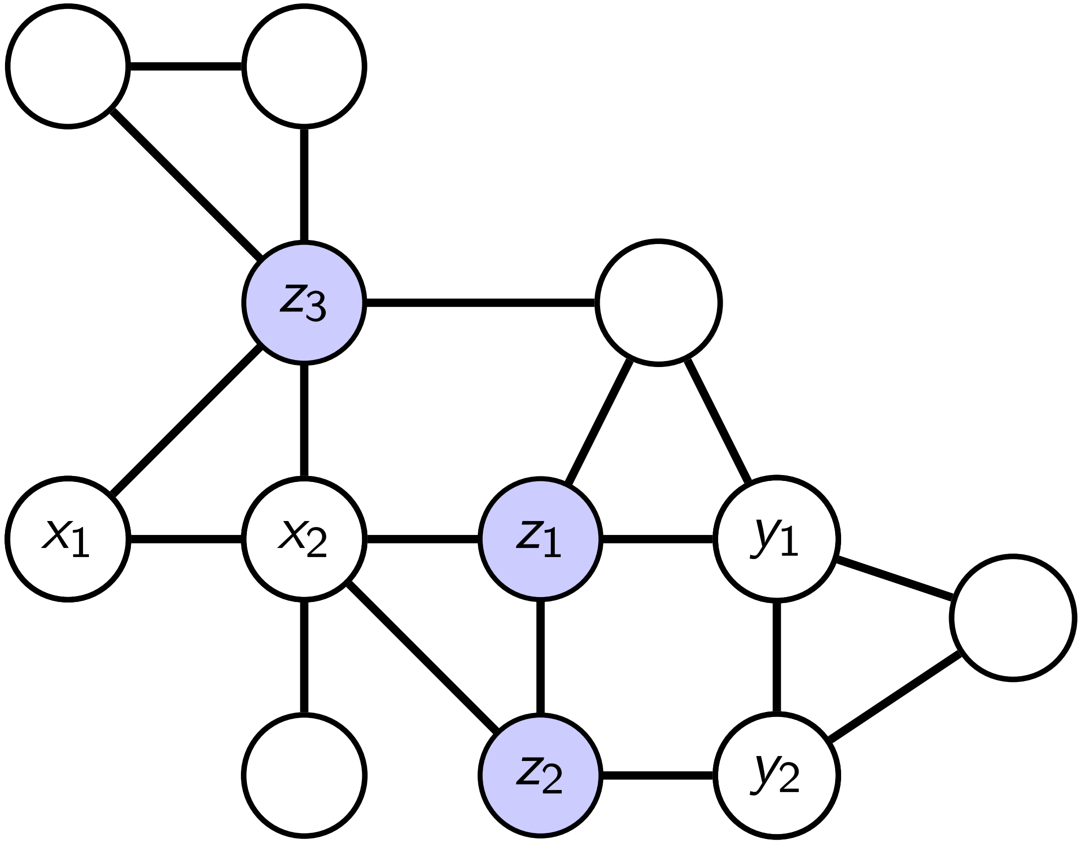
- x_3 separates (x_1, x_2, x_4) from (x_5, x_6)
- x_3 separates x_5 from x_6.
- However, it does e.g. not separate x_2 from x_4.



Deriving the graph separation criterion

Without loss of generality, consider the graph below and assume that $p(x_1, \ldots, x_d) \propto \prod_c \phi_c(\mathcal{X}_c)$, with $\mathcal{X}_c \subset \{x_1, \ldots, x_d\}$, factorises over it.

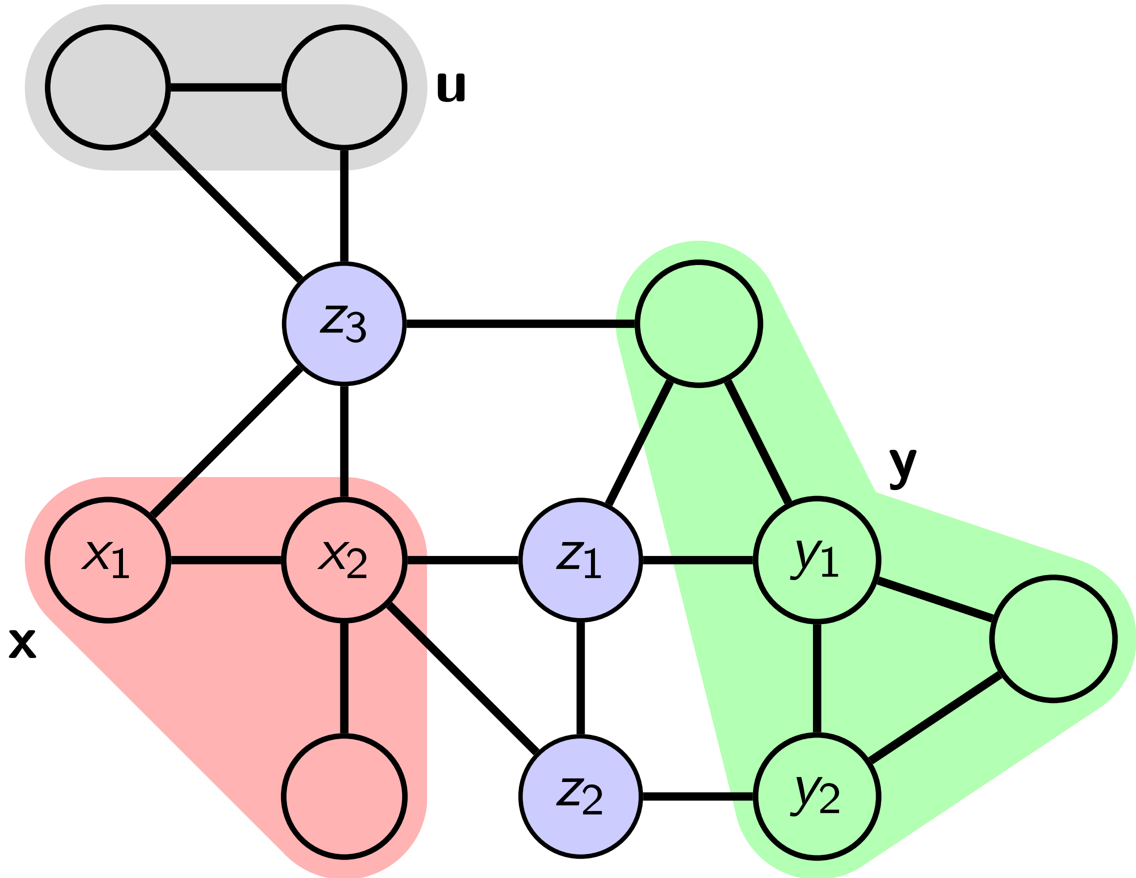
Do we have $x_1, x_2 \perp\!\!\!\perp y_1, y_2 \mid z_1, z_2, z_3$?



Deriving the graph separation criterion

Without loss of generality, consider the graph below and assume that $p(x_1, \ldots, x_d) \propto \prod_c \phi_c(\mathcal{X}_c)$, with $\mathcal{X}_c \subset \{x_1, \ldots, x_d\}$, factorises over it.

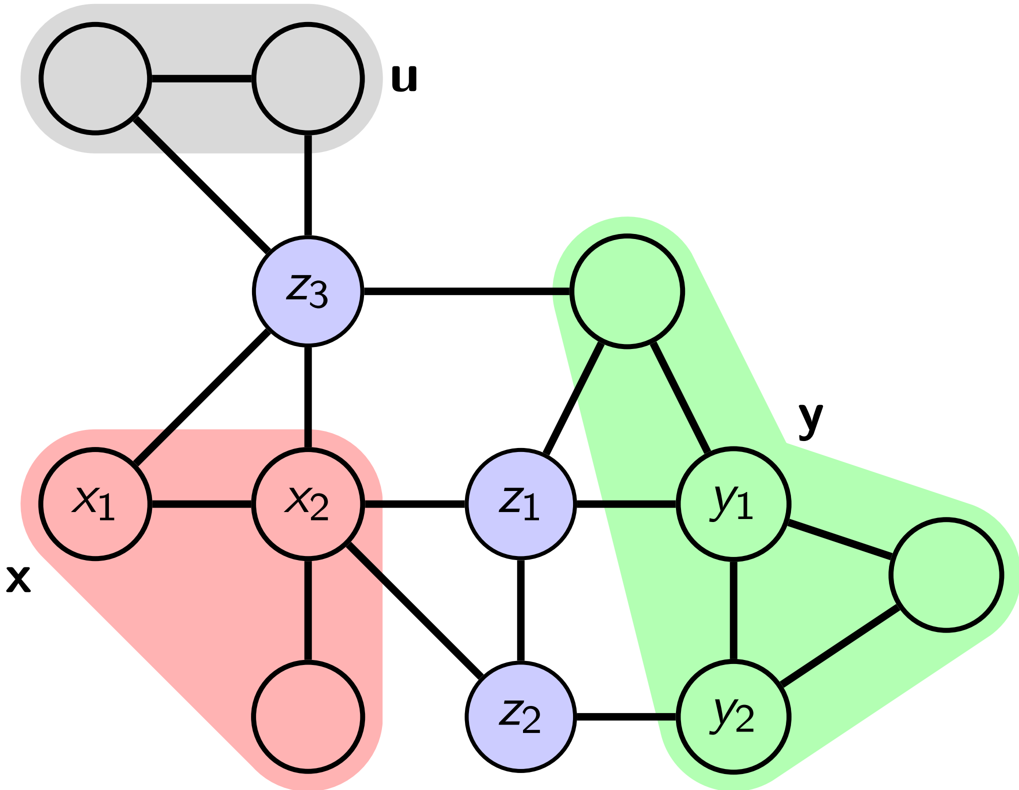
Do we have $\mathbf{x} \perp\!\!\!\perp \mathbf{y} \mid z_1, z_2, z_3$?



Deriving the graph separation criterion

- With $\mathbf{z} = (z_1, z_2, z_3)$, all x_i belong to one of the $\mathbf{x}, \mathbf{y}, \mathbf{z}$, or $\mathbf{u}$.
- We thus have $p(x_1, \ldots, x_d) = p(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u})$ and we can group the factors ϕ_c together so that

$$p(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}) \propto \phi_1(\mathbf{x}, \mathbf{z})\phi_2(\mathbf{y}, \mathbf{z})\phi_3(\mathbf{u}, \mathbf{z})$$



Deriving the graph separation criterion

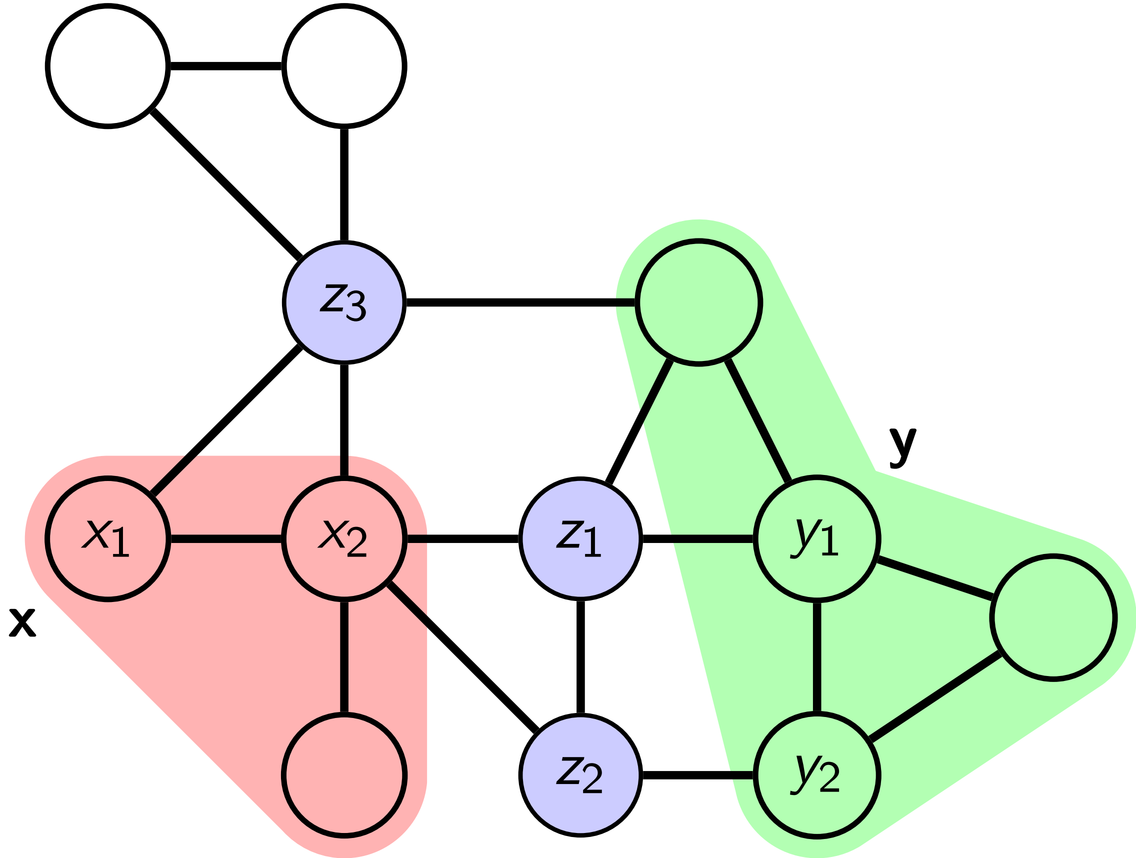
- Integrating (summing) out **u** gives

$$
\begin{aligned}
p(\mathbf{x}, \mathbf{y}, \mathbf{z}) &= \sum_{\mathbf{u}} p(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}) && (1)\\
&\propto \sum_{\mathbf{u}} \phi_1(\mathbf{x}, \mathbf{z})\phi_2(\mathbf{y}, \mathbf{z})\phi_3(\mathbf{u}, \mathbf{z}) && (2)\\
\text{(distributive law)} \quad &\propto \phi_1(\mathbf{x}, \mathbf{z})\phi_2(\mathbf{y}, \mathbf{z}) \sum_{\mathbf{u}} \phi_3(\mathbf{u}, \mathbf{z}) && (3)\\
&\propto \phi_1(\mathbf{x}, \mathbf{z})\phi_2(\mathbf{y}, \mathbf{z})\tilde{\phi}(\mathbf{z}) && (4)\\
&\propto \phi_A(\mathbf{x}, \mathbf{z})\phi_B(\mathbf{y}, \mathbf{z}) && (5)
\end{aligned}
$$

- And $p(\mathbf{x}, \mathbf{y}, \mathbf{z}) \propto \phi_A(\mathbf{x}, \mathbf{z})\phi_B(\mathbf{y}, \mathbf{z})$ means $\mathbf{x} \perp\!\!\!\perp \mathbf{y} \mid \mathbf{z}$

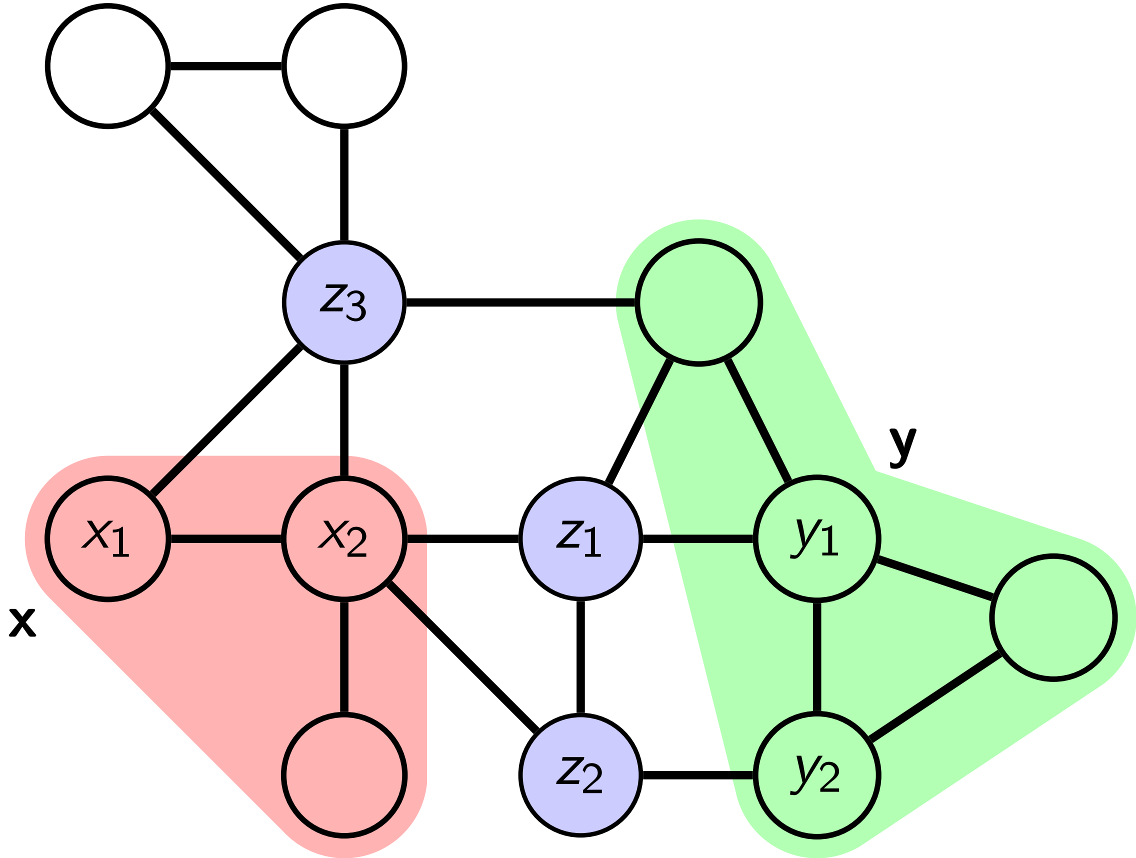
Deriving the graph separation criterion

We have shown that if $\mathbf{x}$ and $\mathbf{y}$ are separated by $\mathbf{z}$, then $\mathbf{x} \perp\!\!\!\perp \mathbf{y} \mid \mathbf{z}$.



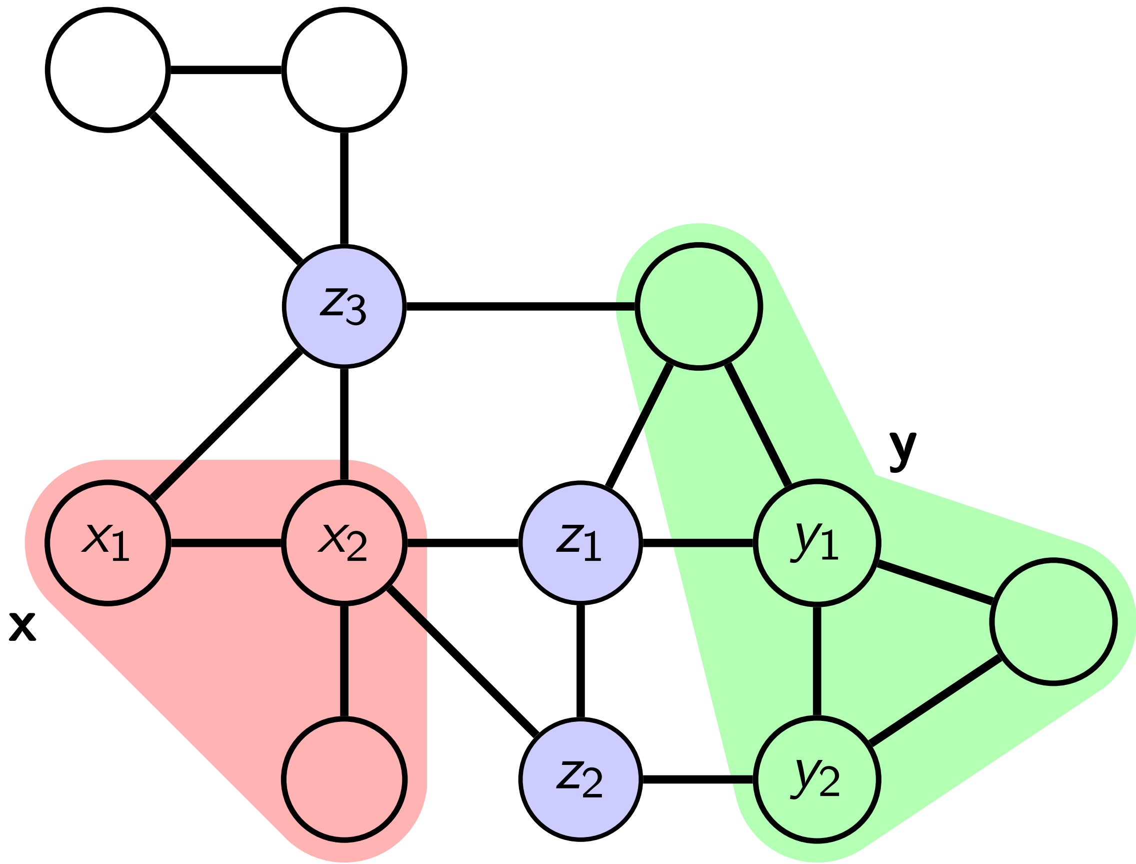
Deriving the graph separation criterion

So do we have $x_1, x_2 \perp\!\!\!\perp y_1, y_2 \mid z_1, z_2, z_3$?



Deriving the graph separation criterion

- From exercises: $x \perp\!\!\!\perp \{y, w\} \mid z$ implies $x \perp\!\!\!\perp y \mid z$
- Hence $\mathbf{x} \perp\!\!\!\perp \mathbf{y} \mid z_1, z_2, z_3$ implies $x_1, x_2 \perp\!\!\!\perp y_1, y_2 \mid z_1, z_2, z_3$.



Graph separation and conditional independence

Theorem:
Let H be an undirected graph and X, Y, Z three disjoint subsets of its nodes. If X and Y are separated by Z, then $X \perp\!\!\!\perp Y \mid Z$ for all probability distributions that factorise over the graph.

Important because:

1. the theorem allows us to read out (conditional) independencies from the undirected graph
2. no restriction on the sets X, Y, Z
3. the independencies detected by graph separation are "true positives" ("soundness" of the independence assertions made by the graph separation criterion).
   (not a "if and only if" statement. Consider e.g. the example that we used to illustrate that d-connected variables may be independent)

Global Markov property $M_g(H)$

- Distributions $p(\mathbf{x})$ are said to satisfy the global Markov property with respect to the undirected graph H, or $M_g(H)$, if for any triple X, Y, Z of disjoint subsets of nodes such that Z separates X and Y in H, we have $X \perp\!\!\!\perp Y \mid Z$.
- *Global* Markov property because we do not restrict the sets X, Y, Z.
- The theorem says that $F(H) \Longrightarrow M_g(H)$.
- Undirected analogue to d-separation and the directed global Markov property.

What if two sets of nodes are not graph separated?

Theorem: If X and Y are not separated by Z in the undirected graph H then $X \not\perp\!\!\!\perp Y \mid Z$ in some probability distributions that factorise over H.

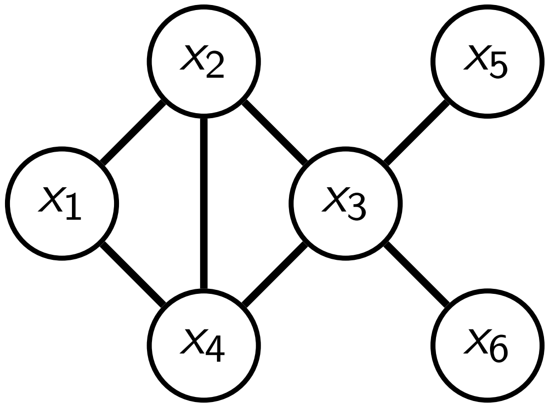
Optional, for those interested: A proof sketch can be found in Section 4.3.1.2 of *Probabilistic Graphical Models* by Koller and Friedman.

Remarks:

- The theorem implies that for some distributions, we may have $X \perp\!\!\!\perp Y \mid Z$ even though X and Y are not separated by Z. The separation criterion is not “complete” (“recall-rate” is not guaranteed to be 100%).
- Same caveat as for d-separation.

Example

Undirected graph:

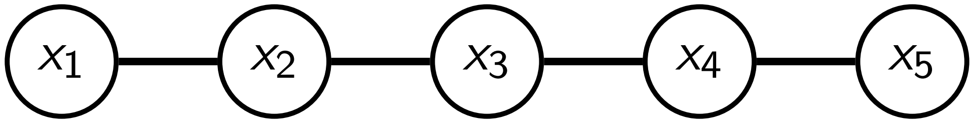


All models defined by the undirected graph satisfy:

$x_1 \perp\!\!\!\perp \{x_3, x_5, x_6\} \mid x_2, x_4 \qquad x_2 \perp\!\!\!\perp x_6 \mid x_3 \qquad x_5 \perp\!\!\!\perp x_6 \mid x_3$

Example: Markov chain

Undirected graph:



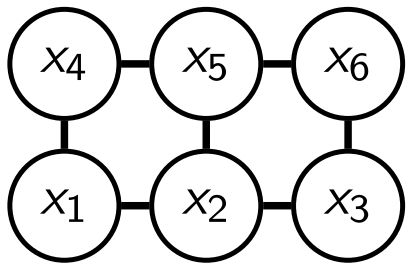
All models defined by the undirected graph satisfy:

$$x_1, \dots x_{i-1} \perp\!\!\!\perp x_{i+1}, \dots, x_5 \mid x_i$$

(past and future are independent given the present)

Example: pairwise Markov network

Undirected graph:



All models defined by the undirected graph satisfy:

$$x_1, x_4 \perp\!\!\!\perp x_3, x_6 \mid x_2, x_5$$

$$x_1 \perp\!\!\!\perp x_5, x_6, x_3 \mid x_4, x_2 \qquad x_1 \perp\!\!\!\perp x_6 \mid x_2, x_3, x_4, x_5$$

(Last two are examples of the "local Markov property" and the "pairwise Markov property" relative to the undirected graph.)

Program

1. Graph separation and the undirected global Markov property
   - Link between conditioning, graph structure, factorisation, and independencies
   - Graph separation to determine independencies
   - Examples
2. Further methods to determine independencies

Program

1. Graph separation and the undirected global Markov property

2. Further methods to determine independencies
   - Local and pairwise Markov property
   - Equivalences
   - Markov blanket

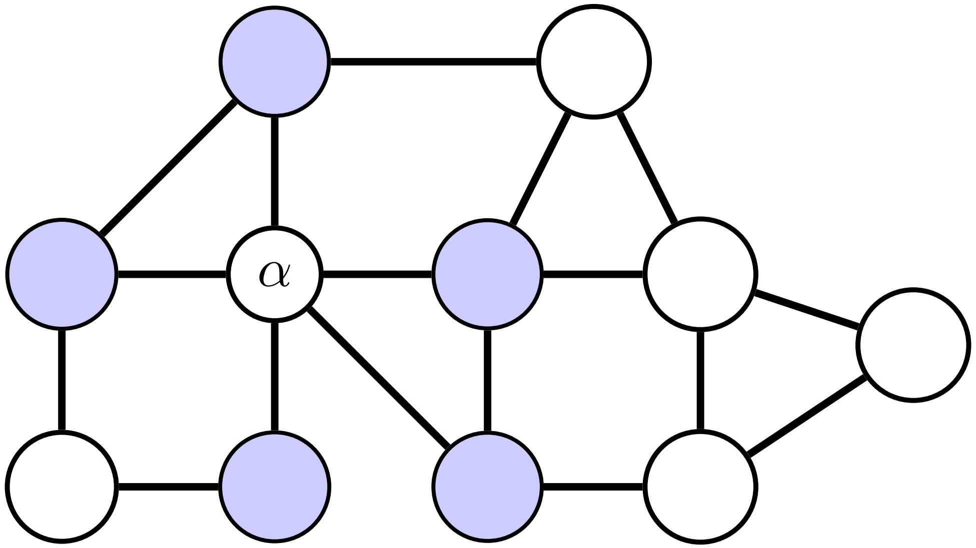
Local Markov property

Denote the set of all nodes by X and the neighbours of a node α by $\mathrm{ne}(\alpha)$.

- A probability distribution is said to satisfy the local Markov property $M_l(H)$ relative to an undirected graph H if

$$\alpha \perp\!\!\!\perp X \setminus (\alpha \cup \mathrm{ne}(\alpha)) \mid \mathrm{ne}(\alpha) \quad \text{for all nodes } \alpha \in X$$

- If p satisfies the global Markov property, then it satisfies the local Markov property. This is because $\mathrm{ne}(\alpha)$ blocks all trails to remaining nodes.



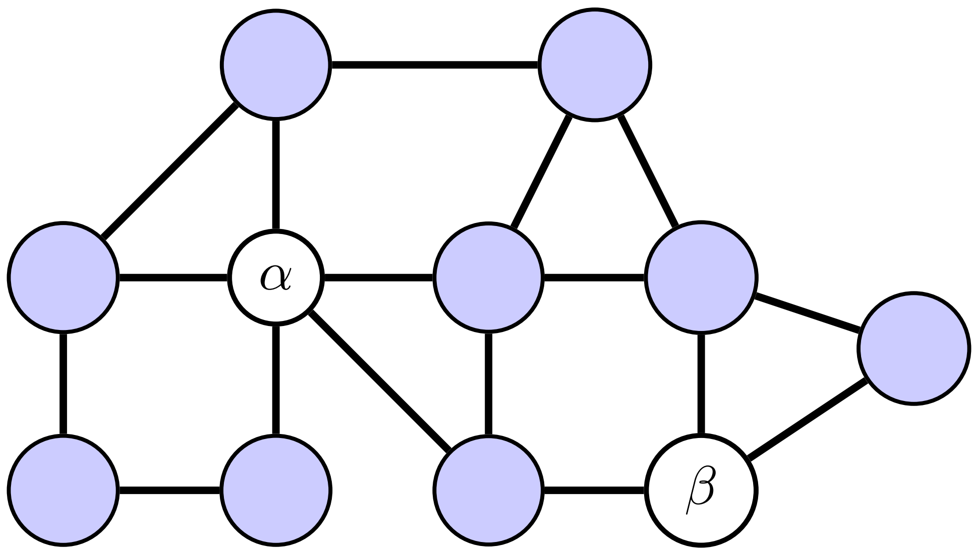
Pairwise Markov property

Denote the set of all nodes by X.

- A probability distribution is said to satisfy the pairwise Markov property $M_p(H)$ relative to an undirected graph H if

$$\alpha \perp\!\!\!\perp \beta \mid X \setminus \{\alpha, \beta\} \quad \text{for all non-neighbouring } \alpha, \beta \in X$$

- If p satisfies the local Markov property, then it satisfies the pairwise Markov property.



Summary

Consider an undirected graph H and the undirected graphical model defined by it.

$$p \text{ satisfies } F(H) \text{ (it factorises over } H)$$
$$\Downarrow$$
$$p \text{ satisfies the global Markov property } M_g(H)$$
$$\Downarrow$$
$$p \text{ satisfies the local Markov property } M_l(H)$$
$$\Downarrow$$
$$p \text{ satisfies the pairwise Markov property } M_p(H)$$

Do we have an equivalence?

- In directed graphical models, we had an equivalence of
  - factorisation,
  - ordered Markov property,
  - local directed Markov property, and
  - global directed Markov property.
- Do we have a similar equivalence for undirected graphical models?

Yes, under some mild condition

From pairwise to global Markov property and factorisation

- Theorem: Assume $p(\mathbf{x}) > 0$ for all $\mathbf{x}$ in its domain (excludes deterministic relationships). If p satisfies the pairwise Markov property with respect to an undirected graph H then p factorises over H.

  (For a proof and weaker conditions, see e.g. Lauritzen, 1996, Section 3.2.)

- Hence: equivalence of factorisation and the global, local, and pairwise Markov properties for positive distributions.
- Equivalence known as Hammersely-Clifford theorem.
- Important e.g. for learning because prior knowledge may come in form of conditional independencies (the graph), which we can incorporate by specifying models that factorise accordingly.

Summary of the equivalences

For a undirected graph H with nodes (random variables) x_i and maximal cliques $\mathcal{X}_c$, we have the following equivalences:

$p(\mathbf{x})$ satisfies $F(H)$ — $p(x_1, \ldots, x_d) = \frac{1}{Z} \prod_c \phi_c(\mathcal{X}_c), \quad \phi_c(\mathcal{X}_c) > 0$

$\Updownarrow$

$p(\mathbf{x})$ satisfies $M_p(H)$ — $\alpha \perp\!\!\!\perp \beta \mid \{x_1, \ldots, x_d\} \setminus \{\alpha, \beta\}$ for all non-neighbouring α, β

$\Updownarrow$

$p(\mathbf{x})$ satisfies $M_l(H)$ — $\alpha \perp\!\!\!\perp \{x_1, \ldots, x_d\} \setminus (\alpha \cup \mathrm{ne}(\alpha)) \mid \mathrm{ne}(\alpha)$ for all nodes α

$\Updownarrow$

$p(\mathbf{x})$ satisfies $M_g(H)$ — all independencies asserted by graph separation

F: factorisation property, M_l: pairwise MP, M_l: local MP, M_g: global MP
(MP: Markov property)

Broadly speaking, the graph serves two related purposes:

1. it tells us how distributions factorise
2. it represents the independence assumptions made

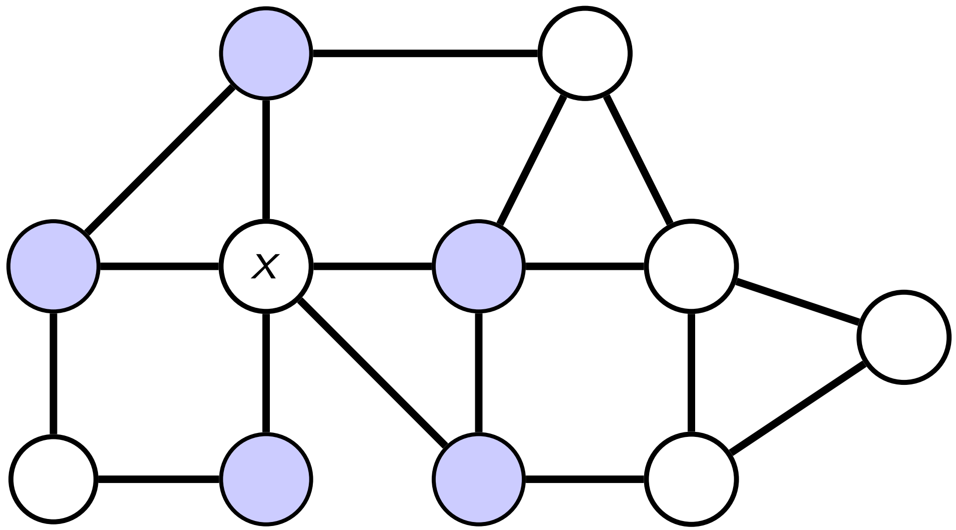
Markov blanket

What is the minimal set of variables such that knowing their values makes x independent from the rest?

From local Markov property: $\mathrm{MB}(x) = \mathrm{ne}(x)$:

$$x \perp\!\!\!\perp \{\text{all variables} \setminus (x \cup \mathrm{ne}(x))\} \mid \mathrm{ne}(x)$$

(Same set of nodes that we get by connecting x to all other variables in factors ϕ_c that contain x, see visualisation of Gibbs distributions).)



What can we do with the equivalences?

- The main things that we have covered:
  - If we know the factorisation of a $p(\mathbf{x})$, we can build a graph H such that $p(\mathbf{x})$ satisfies $F(H)$ and then use the graph to determine independencies that $p(\mathbf{x})$ satisfies.
  - Relatedly, if we know the Markov blanket for each variable, we can build an undirected graph H such that $p(\mathbf{x})$ satisfies $M_l(H)$.
  - We can start with the graph and check which independencies it implies, and, when happy, define a set of pdfs/pdfs that all satisfy the specified independencies.
- What we haven't covered:
  - How to determine an undirected graph from an arbitrary set of independencies.
  - How to learn an undirected graph from samples from $p(\mathbf{x})$ (structure learning).

Program recap

1. Graph separation and the undirected global Markov property
   - Link between conditioning, graph structure, factorisation, and independencies
   - Graph separation to determine independencies
   - Examples

2. Further methods to determine independencies
   - Local and pairwise Markov property
   - Equivalences
   - Markov blanket