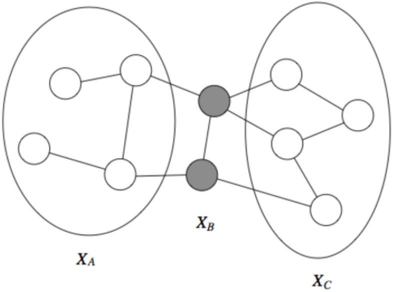
# Undirected graphical models (Markov random fields)

- Some distributions cannot be perfectly represented by a DAG
- More natural representation without directionality
- Parallel to the directed case, we will discuss
  - factorised parametrisations
  - conditional independence based on graph separation
- Comparison with DAG

# Undirected graphical models

Conditional independence

- Undirected graph $G(V, E)$, $V$ is a set of nodes, each corresponding to a random variable, $E$ is a set of undirected edges.
- $X_A \perp\!\!\!\perp X_C | X_B$ if $X_B$ separates the nodes $X_A$ from the nodes $X_C$.



# Parametrisation for undirected graphical models

- Express the joint PMF as the product of “local” functions. (In the directed case, local means $\{i, \pi_i\}$)
- If $X_i$ and $X_j$ are not directly connected, they are conditionally independent given all the other nodes $\Rightarrow$ $X_i$ and $X_j$ should not appear in the same local function
- A clique is a fully-connected subset of nodes.
- Define local functions on cliques

# Parametrisation for undirected graphical models

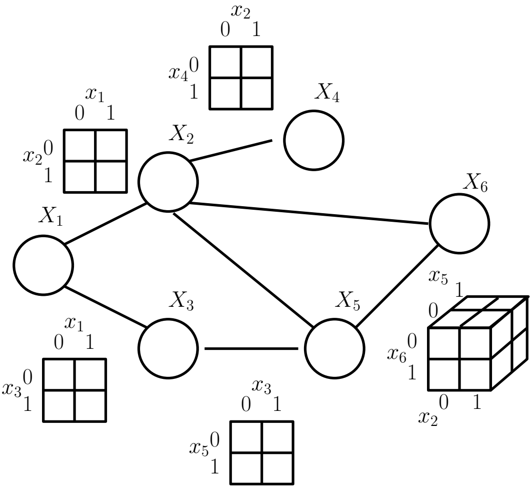
- A maximal clique is a clique that cannot be extended without losing the property of being fully connected.
- Let $C$ be a set of indices of a maximal clique in an undirected graph $G(V, E)$, and let $\mathcal{C}$ be the set of all such $C$.
- A potential function $\psi_{X_c}(x_C)$ is a function on the possible realisations $x_C$ of the maximal clique $X_C$. $\psi_C$ is nonnegative, real-valued, but otherwise arbitrary.
- Define joint PMF

$$p(x) = \frac{1}{Z} \prod_{C \in \mathcal{C}} \psi_{X_c}(x_C),$$

where $Z = \sum_x \prod_{C \in \mathcal{C}} \psi_{X_c}(x_C)$

# Parametrisation for undirected graphical models

Example.



# Two equivalent characterisations

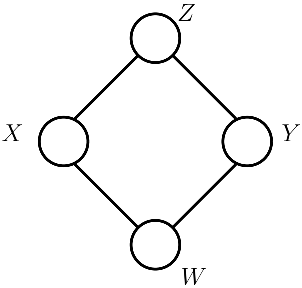
Given an undirected graph $G$,

- Generate a family of distributions $\mathcal{U}_1$ as follows:
  - factorisation in terms of potential functions
  - range over all possible choices of nonnegative potential functions on the maximal cliques of $G$
- Generate a family of distributions $\mathcal{U}_2$ as follows:
  - find all conditional independences $X_A \perp\!\!\!\perp X_B | X_C$ by assessing whether the subset of nodes in $X_A$ is separated from $X_B$ when the nodes $X_C$ are removed
  - consider all possible joint distributions
  - test each against the list of conditional independences; keep the distribution if all satisfied

$\mathcal{U}_1$ and $\mathcal{U}_2$ are identical by the Hammersley-Clifford theorem.

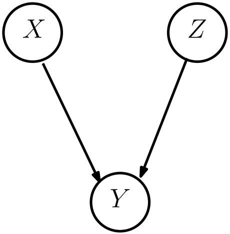
# Some comparisons

Voting preferences among $X, Y, Z, W$,
$(X, W), (W, Y), (Z, Y), (X, Z)$ are friends.




$X \perp\!\!\!\perp Y | \{W, Z\}$, $W \perp\!\!\!\perp Z | \{X, Y\}$. Can we represent this as a directed graph?

# Some comparisons




$X \perp\!\!\!\perp Z$, but $X$ and $Z$ are not independent given $Y$. Can we represent this as an undirected graph?

# Some comparisons

Advantages of undirected graphs

- Can be applied to a wider range of problems in which there is no natural directionality associated with variable dependencies.
- Can succinctly express certain dependencies that DAGs cannot easily describe.

Drawbacks

- Computing the normalisation constant $Z$ can be difficult.
- May be difficult to interpret.
- Easier to generate data from DAGs.

# Inference on graphical models

Given a probabilistic model, how do we obtain answers to relevant questions about the model?

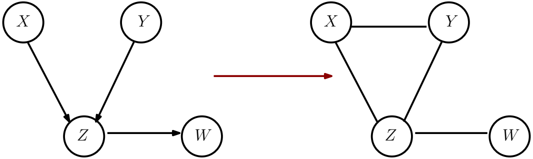
- Marginal / conditional inference: compute $p(x_A)$ or $p(x_A|x_B)$
- Maximum a posteriori (MAP) inference : find the most likely assignment

$$\max_{x_1,\ldots,x_n} p(x_1, \ldots, x_n)$$

Whether inference is tractable depends on the graph structure. If intractable, we can still obtain useful answers via approximate inference methods.

# Moralisation

Converting DAGs to undirected graphs



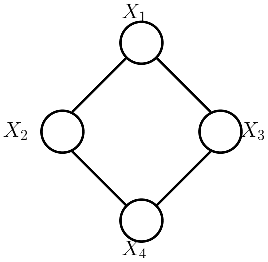

Redefine $p(z|x)p(z|y)$ as $\psi(x, y, z)$.
In general, take a DAG and add edges to all parents of a given node and remove all directionalities

# Variable elimination

**Goal**: find marginal probability $p(x_i) = \sum_{x_j : j \neq i} p(x_1, \ldots, x_n)$

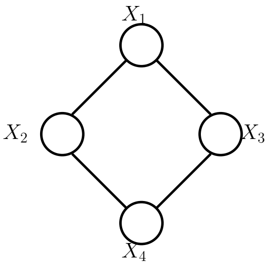
If each variable has $d$ possible values, naive calculation suggests we need $d^{n-1}$ operations. The key is to use the factorisation form given by the graph.

# Variable elimination - an example




$$
\begin{aligned}
p(x_1) &= \sum_{x_2,x_3,x_4} p(x_1,x_2,x_3,x_4) \\
p(x_1) &\propto \sum_{x_2,x_3,x_4} \psi_{1,2}(x_1,x_2)\psi_{1,3}(x_1,x_3)\psi_{2,4}(x_2,x_4)\psi_{3,4}(x_3,x_4) \\
&= \sum_{x_2,x_3} \psi_{1,2}(x_1,x_2)\psi_{1,3}(x_1,x_3) \sum_{x_4} \psi_{2,4}(x_2,x_4)\psi_{3,4}(x_3,x_4) \\
&= \sum_{x_2,x_3} \psi_{1,2}(x_1,x_2)\psi_{1,3}(x_1,x_3) m_4(x_2,x_3)
\end{aligned}
$$

# Variable elimination - example 1

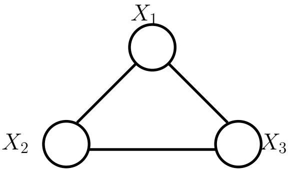



$$
\begin{aligned}
p(x_1) &\propto \sum_{x_2} \psi_{1,2}(x_1, x_2) \sum_{x_3} \psi_{1,3}(x_1, x_3) m_4(x_2, x_3) \\
&= \sum_{x_2} \psi_{1,2}(x_1, x_2) m_3(x_1, x_2) \\
&= m_2(x_1).
\end{aligned}
$$

The normalising constant is given by $\sum_{x_1} m_2(x_1)$.

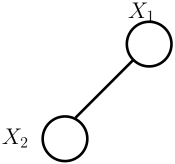
# Variable elimination - example 1

$$\begin{aligned}
p(x_1) &\propto \sum_{x_2,x_3,x_4} \psi_{1,2}(x_1,x_2)\psi_{1,3}(x_1,x_3)\psi_{2,4}(x_2,x_4)\psi_{3,4}(x_3,x_4) \\
&= \sum_{x_2,x_3} \psi_{1,2}(x_1,x_2)\psi_{1,3}(x_1,x_3) \sum_{x_4} \psi_{2,4}(x_2,x_4)\psi_{3,4}(x_3,x_4) \\
&= \sum_{x_2,x_3} \psi_{1,2}(x_1,x_2)\psi_{1,3}(x_1,x_3) m_4(x_2,x_3) \\
&= \sum_{x_2} \psi_{1,2}(x_1,x_2) m_3(x_1,x_2) \\
&= m_2(x_1).
\end{aligned}$$



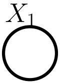
# Variable elimination - example 1

$$
\begin{aligned}
p(x_1) &\propto \sum_{x_2,x_3,x_4} \psi_{1,2}(x_1,x_2)\psi_{1,3}(x_1,x_3)\psi_{2,4}(x_2,x_4)\psi_{3,4}(x_3,x_4) \\
&= \sum_{x_2,x_3} \psi_{1,2}(x_1,x_2)\psi_{1,3}(x_1,x_3) \sum_{x_4} \psi_{2,4}(x_2,x_4)\psi_{3,4}(x_3,x_4) \\
&= \sum_{x_2,x_3} \psi_{1,2}(x_1,x_2)\psi_{1,3}(x_1,x_3) m_4(x_2,x_3) \\
&= \sum_{x_2} \psi_{1,2}(x_1,x_2) m_3(x_1,x_2) \\
&= m_2(x_1).
\end{aligned}
$$

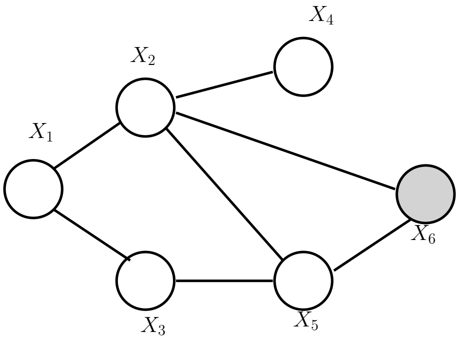


# Variable elimination - example 1

$$
\begin{aligned}
p(x_1) &\propto \sum_{x_2,x_3,x_4} \psi_{1,2}(x_1,x_2)\psi_{1,3}(x_1,x_3)\psi_{2,4}(x_2,x_4)\psi_{3,4}(x_3,x_4) \\
&= \sum_{x_2,x_3} \psi_{1,2}(x_1,x_2)\psi_{1,3}(x_1,x_3) \sum_{x_4} \psi_{2,4}(x_2,x_4)\psi_{3,4}(x_3,x_4) \\
&= \sum_{x_2,x_3} \psi_{1,2}(x_1,x_2)\psi_{1,3}(x_1,x_3) m_4(x_2,x_3) \\
&= \sum_{x_2} \psi_{1,2}(x_1,x_2) m_3(x_1,x_2) \\
&= m_2(x_1).
\end{aligned}
$$



# Variable elimination - example 2




Compute $p(x_1|\bar{x}_6)$. $X_1$ query node, $X_6$ evidence node. Define evidence potential

$$\delta(x_6, \bar{x}_6) = \begin{cases} 1 & \text{if } x_6 = \bar{x}_6 \\ 0 & \text{otherwise} \end{cases}$$

# Variable elimination - example 2

$$
\begin{aligned}
& p(x_1, \bar{x}_6) \\
=& \frac{1}{Z} \sum_{x_2, x_3, x_4, x_5, x_6} \psi(x_1, x_2)\psi(x_1, x_3)\psi(x_2, x_4)\psi(x_3, x_5)\psi(x_2, x_5, x_6)\delta(x_6, \bar{x}_6) \\
=& \frac{1}{Z} \sum_{x_2, x_3, x_4, x_5} \psi(x_1, x_2)\psi(x_1, x_3)\psi(x_2, x_4)\psi(x_3, x_5) m_6(x_2, x_5) \\
=& \frac{1}{Z} \sum_{x_2, x_3, x_4} \psi(x_1, x_2)\psi(x_1, x_3)\psi(x_2, x_4) m_5(x_2, x_3) \\
=& \frac{1}{Z} \sum_{x_2, x_3} \psi(x_1, x_2)\psi(x_1, x_3) m_5(x_2, x_3) m_4(x_2) \\
=& \frac{1}{Z} \sum_{x_2} \psi(x_1, x_2) m_4(x_2) m_3(x_1, x_2) \\
=& \frac{1}{Z} m_2(x_1)
\end{aligned}
$$

Elimination order $\{6, 5, 4, 3, 2\}$

# Graph elimination

- When eliminating a variable, link the neighbours of the node being summed over and remove the node from the graph
- The computational complexity is determined by the size of largest elimination clique: when removing node $X_i$, record the collection of nodes that are neighbours of $X_i$ at that moment, including $X_i$ itself.
- Finding the optimal elimination order is computationally intractable in general.