Graphical Models: Inference

Pradeep Ravikumar

Co-Instructor: Ziv Bar-Joseph

Machine Learning 10-701

Topics in Graphical Models

- Representation
  - Which joint probability distributions does a graphical model represent?
- Inference
  - How to answer questions about the joint probability distribution?
    - Marginal distribution of a node variable
    - Most likely assignment of node variables
- Learning
  - How to learn the parameters and structure of a graphical model?

Topics in Graphical Models

- Representation
  - Which joint probability distributions does a graphical model represent?

- Inference
  - How to answer questions about the joint probability distribution?
    - Marginal distribution of a node variable
    - Most likely assignment of node variables

- Learning
  - How to learn the parameters and structure of a graphical model?

Inference

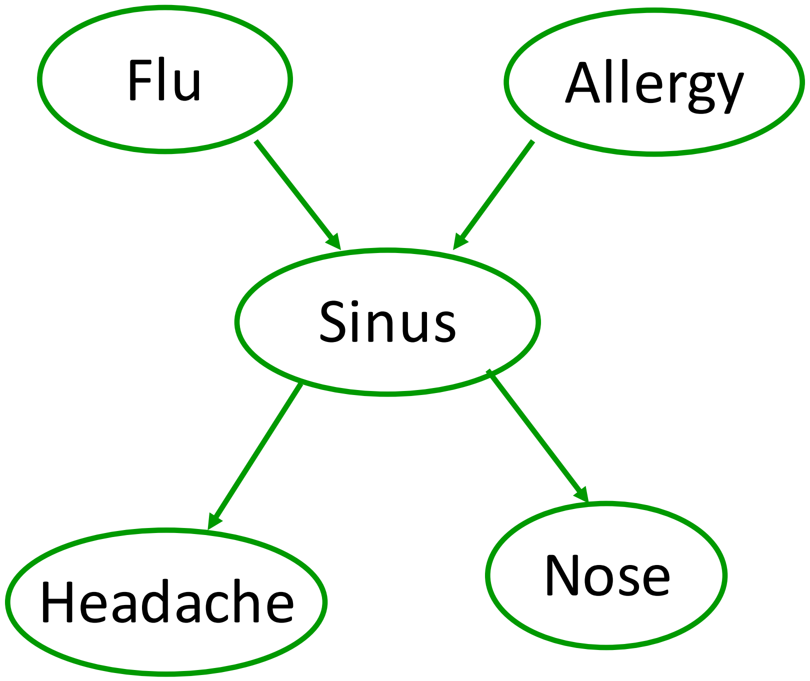
- Possible queries:

1) Marginal distribution e.g. $P(S)$

   Posterior distribution e.g. $P(F|H=1)$

2) Most likely assignment of nodes

   $$\arg\max_{f,a,s,n} P(F=f,A=a,S=s,N=n|H=1)$$

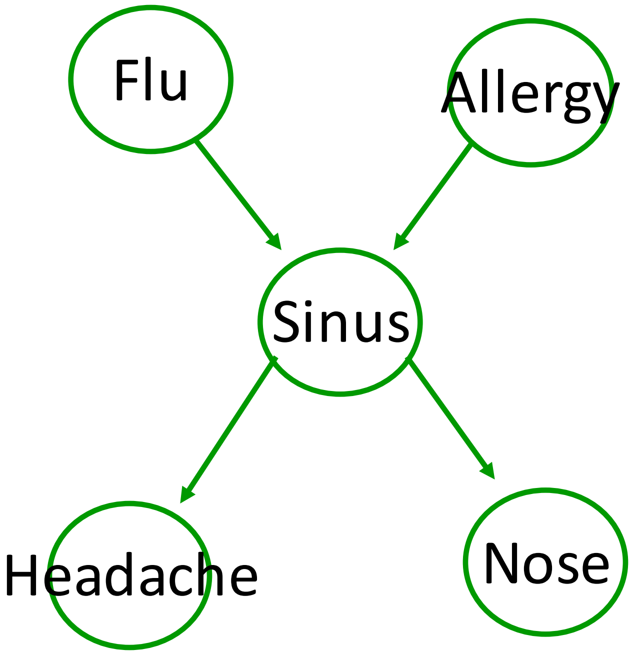


Inference

- Possible queries:

1) Marginal distribution e.g. P(S)

   Posterior distribution e.g. P(F|H=1)



$$P(F|H=1)\ ?$$

$$P(F|H=1) = \frac{P(F, H=1)}{P(H=1)}$$

$$= \frac{P(F, H=1)}{\sum_f P(F=f, H=1)}$$

$$\propto P(F, H=1)$$

will focus on computing this, posterior will follow with only constant times more effort

Marginalization

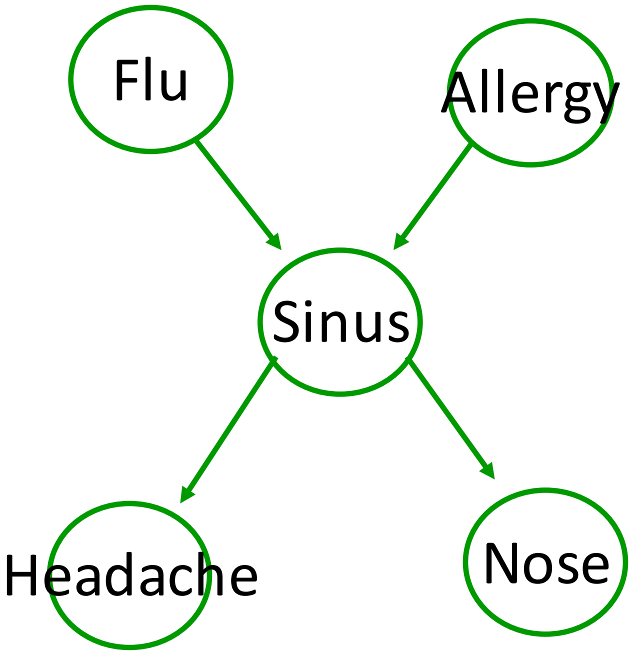
Need to marginalize over other vars

$$P(S) = \sum_{f,a,n,h} P(f,a,S,n,h)$$

$$P(F,H=1) = \underbrace{\sum_{a,s,n}}_{2^3 \text{ terms}} P(F,a,s,n,H=1)$$

To marginalize out n binary variables,

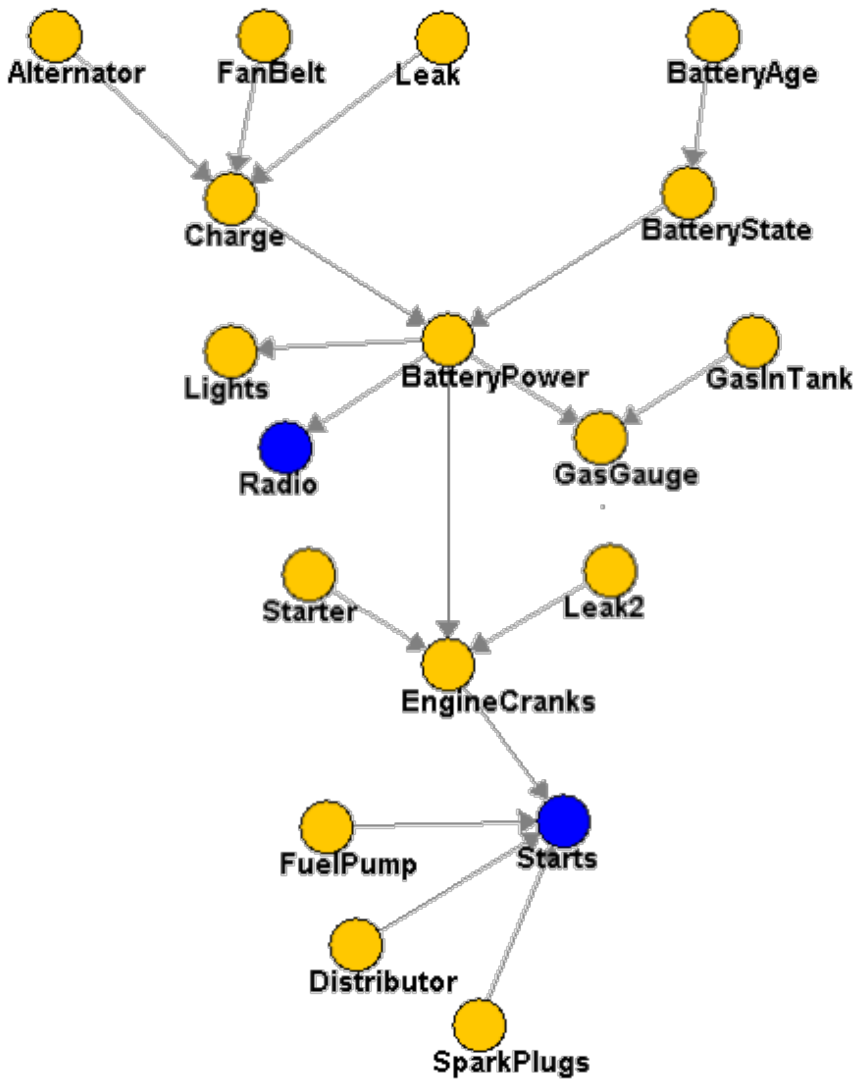
need to sum over 2^n terms



Inference seems exponential in number of variables!

Actually, inference in graphical models is NP-hard ☹

Bayesian Networks Example



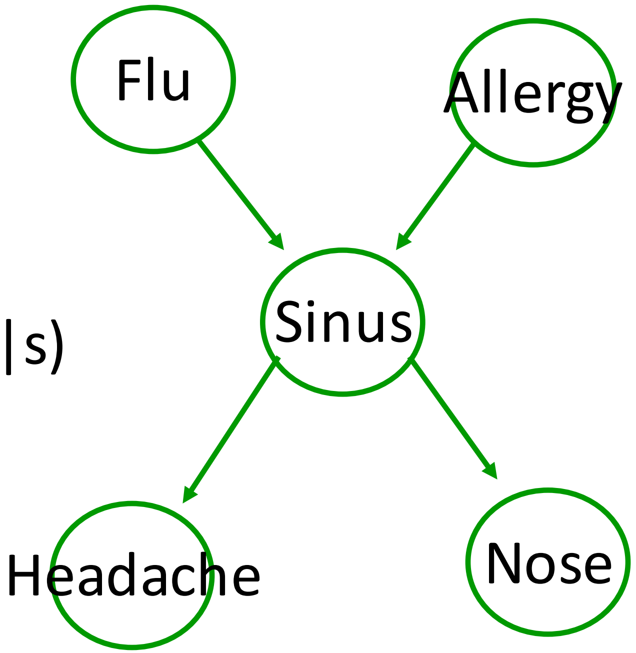
- 18 binary attributes
- Inference
  - P(BatteryAge|Starts=f)
- need to sum over 2^{16} terms!
- Not impressed?
  - HailFinder BN – more than 3^{54} = 58149737003040059690390169 terms

Fast Probabilistic Inference

$$P(F,H=1) = \sum_{a,s,n} P(F,a,s,n,H=1)$$

$$= \sum_{a,s,n} P(F)P(a)P(s|F,a)P(n|s)P(H=1|s)$$

$$= P(F) \sum_{a} P(a) \sum_{s} P(s|F,a)P(H=1|s) \sum_{n} P(n|s)$$



Push sums in as far as possible

Distributive property: $x_1 z + x_2 z = z(x_1+x_2)$

2 multiply 1 mulitply

Fast Probabilistic Inference

$$P(F,H=1) = \sum_{a,s,n} P(F,a,s,n,H=1)$$

8 values x 4 multiplies

$$= \sum_{a,s,n} P(F)P(a)P(s|F,a)P(n|s)P(H=1|s)$$

$$= P(F) \sum_{a} P(a) \sum_{s} P(s|F,a)P(H=1|s) \cancelto{1}{\sum_{n} P(n|s)}$$

$$= P(F) \sum_{a} P(a) \sum_{s} P(s|F,a)P(H=1|s)$$

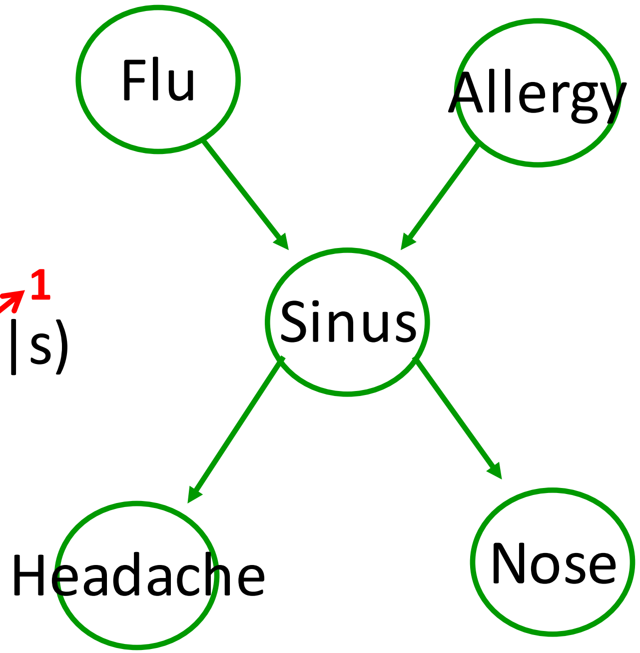
4 values x 1 multiply

$$= P(F) \sum_{a} P(a)\, g_1(F,a)$$

2 values x 1 multiply

$$= P(F)\, g_2(F)$$

1 multiply



32 multiplies vs. 7 multiplies

2^p vs. $p\ 2^k$

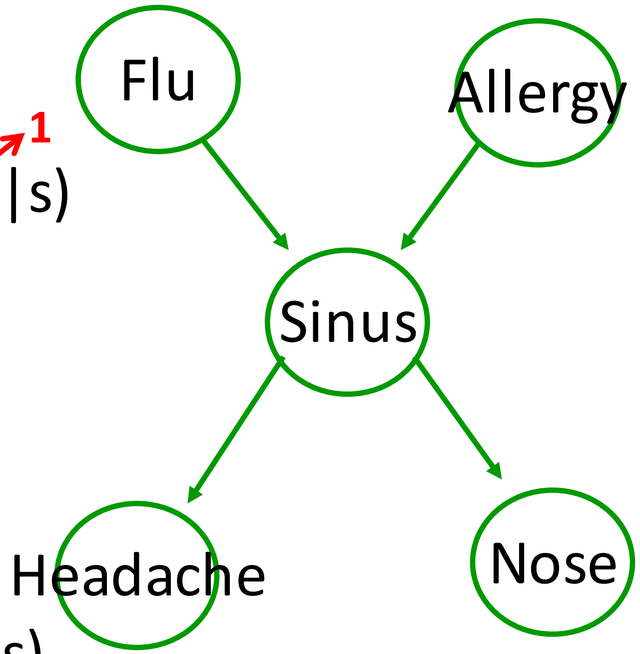
k – scope of largest factor

(Potential for) exponential reduction in computation!

Variable Elimination – Order can make a HUGE difference

$$P(F,H=1) = \sum_{a,s,n} P(F)P(a)P(s|F,a)P(n|s)P(H=1|s)$$

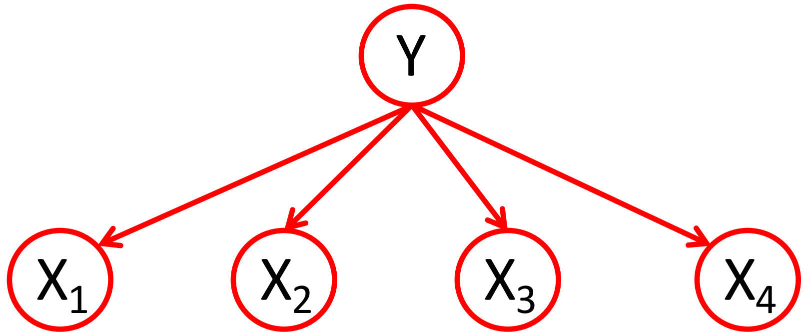
$$= \underbrace{P(F)\sum_a \underbrace{P(a)\sum_s \underbrace{P(s|F,a)P(H=1|s)}_{g_1(F,a,s)}}_{g_2(F,a)}}_{g_3(F)} \cancelto{1}{\sum_n P(n|s)}$$



$$P(F,H=1) = P(F)\sum_a P(a)\sum_n\sum_s \underbrace{P(s|F,a)P(n|s)P(H=1|s)}_{g(F,s,a,n)}$$

3 – scope of largest factor

Variable Elimination – Order can make a HUGE difference



$$P(X_1) = \sum_{Y,X_2,\dots,X_n} P(Y)P(X_1|Y)\prod_{i=2}^{n} P(X_i|Y)$$

$$= \sum_{Y,X_3,\dots,X_n} P(Y)P(X_1|Y)\prod_{i=3}^{n} P(X_i|Y)\underbrace{\sum_{X_2} P(X_2|Y)}_{g(Y)}$$

1 – scope of largest factor

$$= \sum_{X_2,\dots,X_n}\sum_{Y}\underbrace{P(Y)P(X_1|Y)\prod_{i=2}^{n} P(X_i|Y)}_{g(Y,\, X_1,X_2,..,X_n)}$$

n+1 – scope of largest factor

Variable Elimination Algorithm

- Given BN – set initial factors $p(x_i|pa_i)$ for $i=1,..,n$)
- Given Query $P(X|e) \equiv P(X,e)$ X – set of variables
- Instantiate evidence e e.g. set H=1 in previous example
- Choose an ordering on the variables e.g., $X_1, \ldots, X_n$
- For i = 1 to n, If $X_i \notin \{X,e\}$
  - Collect factors $g_1,\ldots,g_k$ that include X_i
  - Generate a new factor by eliminating X_i from these factors

$$g = \sum_{X_i} \prod_{j=1}^{k} g_j$$

  - Variable X_i has been eliminated!
  - Remove $g_1,\ldots,g_k$ from set of factors but add g
- Normalize $P(X, e)$ to obtain $P(X|e)$

Inference

- Possible queries:

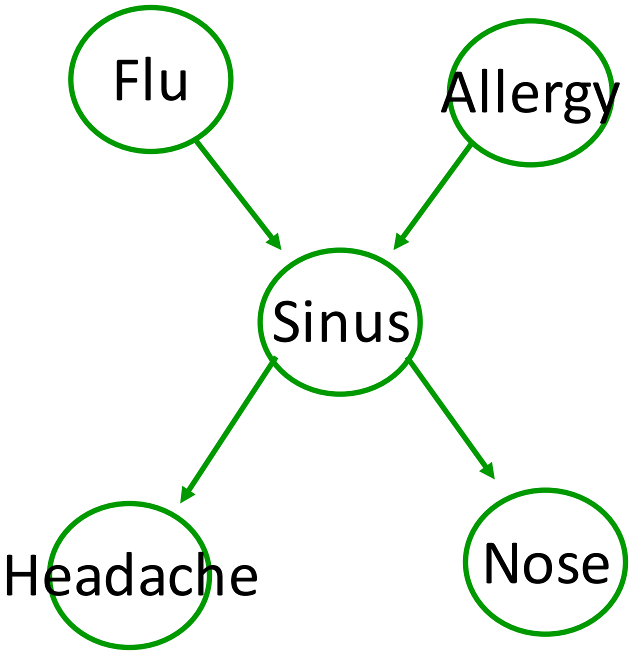
2) Most likely assignment of nodes

$$\arg\max_{f,a,s,n} P(F=f, A=a, S=s, N=n \mid H=1)$$

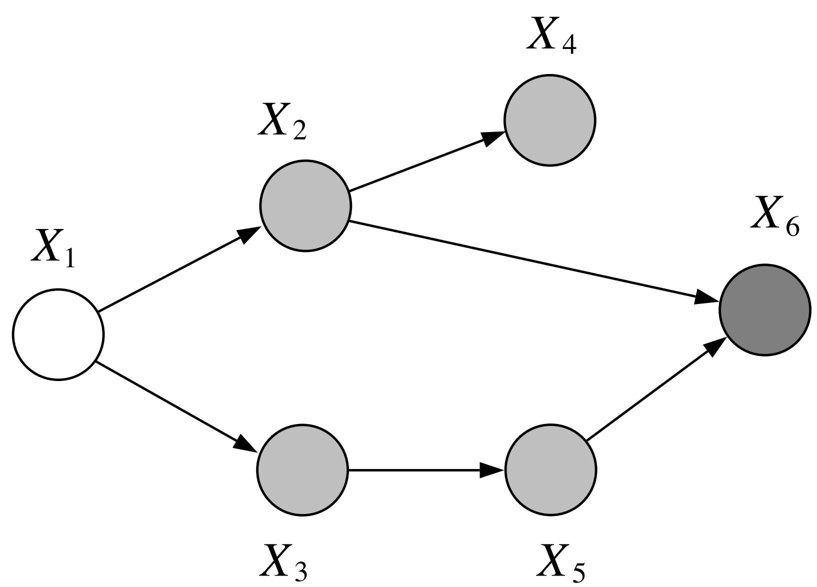
Use Distributive property:

$$\max(x_1 z,\ x_2 z) = z \max(x_1, x_2)$$

2 multiply 1 mulitply

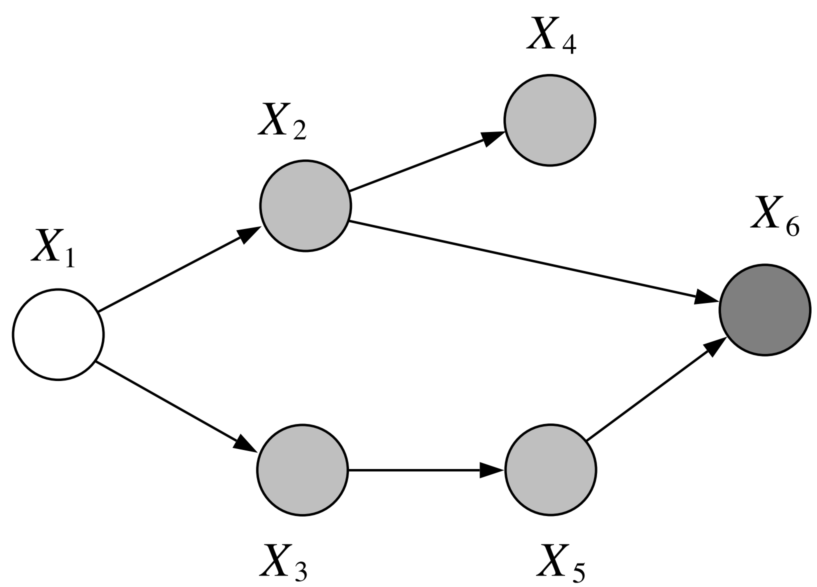


Variable Elimination: Directed Graphs



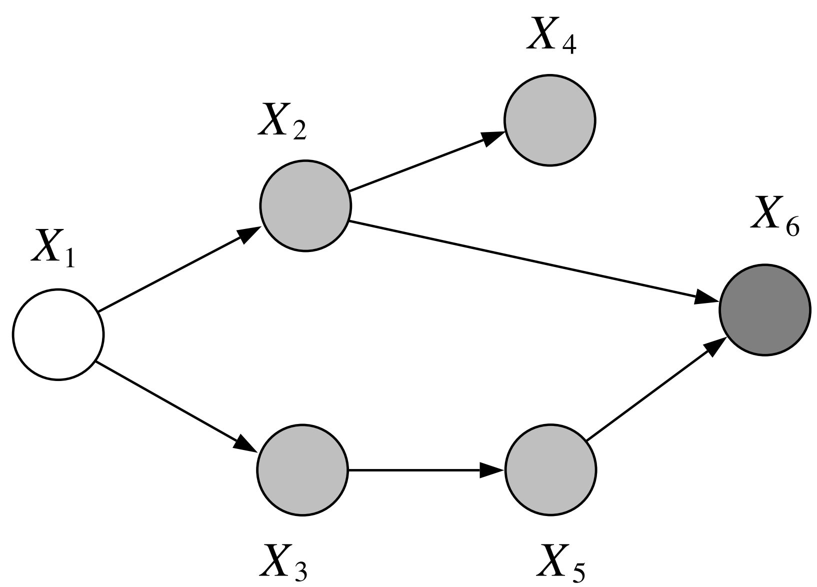
$$p(x_1, x_2, \ldots, x_5) \quad = \quad \sum_{x_6} p(x_1) p(x_2 \,|\, x_1) p(x_3 \,|\, x_1) p(x_4 \,|\, x_2) p(x_5 \,|\, x_3) p(x_6 \,|\, x_2, x_5)$$

Variable Elimination: Directed Graphs



$$
\begin{aligned}
p(x_1, x_2, \ldots, x_5) &= \sum_{x_6} p(x_1)p(x_2 \,|\, x_1)p(x_3 \,|\, x_1)p(x_4 \,|\, x_2)p(x_5 \,|\, x_3)p(x_6 \,|\, x_2, x_5) \\
&= p(x_1)p(x_2 \,|\, x_1)p(x_3 \,|\, x_1)p(x_4 \,|\, x_2)p(x_5 \,|\, x_3) \sum_{x_6} p(x_6 \,|\, x_2, x_5).
\end{aligned}
$$

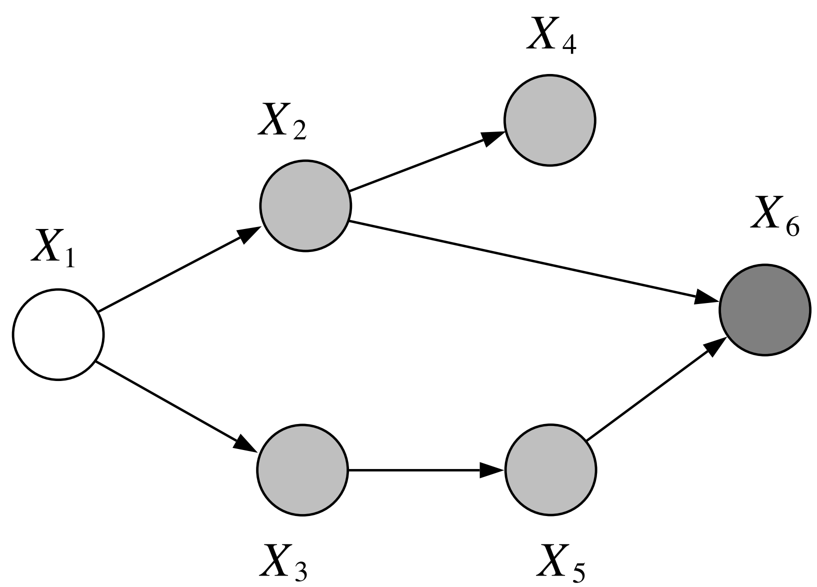
Variable Elimination: Directed Graphs



$$
\begin{aligned}
p(x_1, x_2, \ldots, x_5) &= \sum_{x_6} p(x_1)p(x_2 \mid x_1)p(x_3 \mid x_1)p(x_4 \mid x_2)p(x_5 \mid x_3)p(x_6 \mid x_2, x_5) \\
&= p(x_1)p(x_2 \mid x_1)p(x_3 \mid x_1)p(x_4 \mid x_2)p(x_5 \mid x_3) \sum_{x_6} p(x_6 \mid x_2, x_5).
\end{aligned}
$$

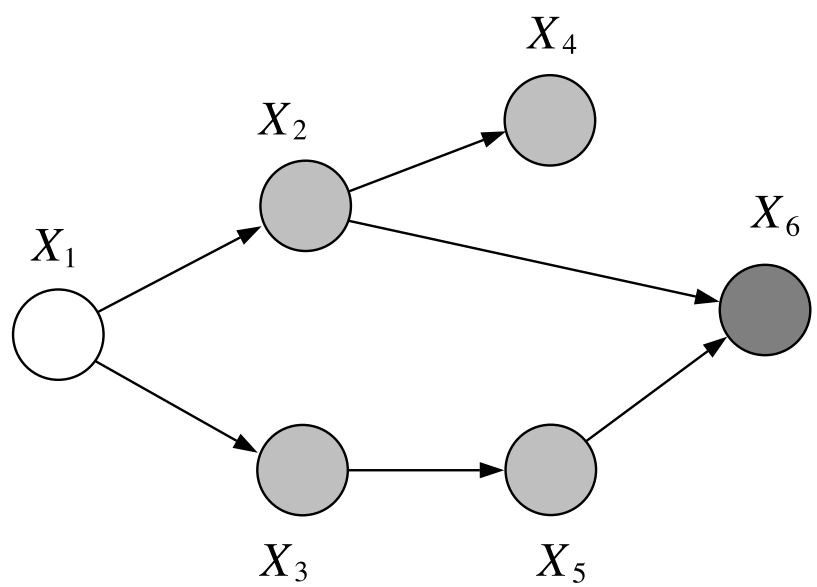
Reduced the count from $O(k^6)$ to $O(k^3)$ (actually we know the sum here is equal to one, but assume we didn't know that)

Variable Elimination: Directed Graphs



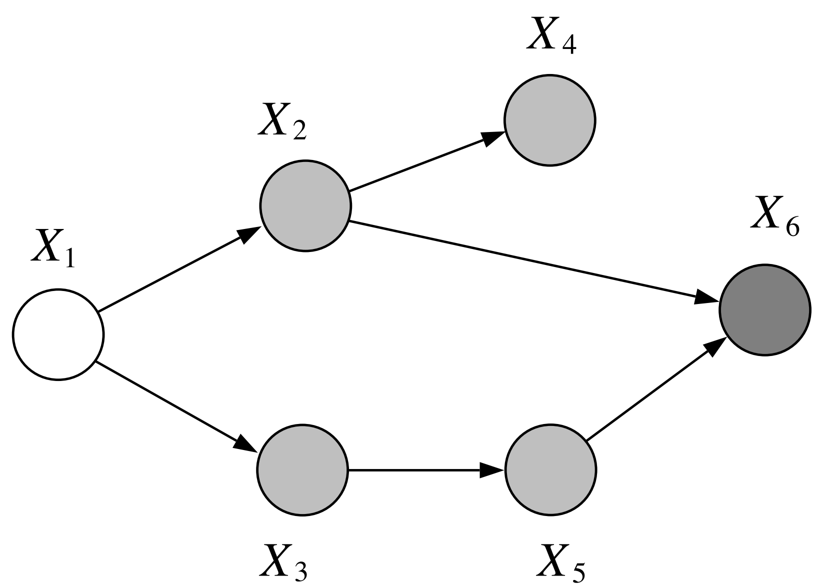
$$p(x_1, \bar{x}_6) \quad = \quad \sum_{x_2} \sum_{x_3} \sum_{x_4} \sum_{x_5} p(x_1) p(x_2 \,|\, x_1) p(x_3 \,|\, x_1) p(x_4 \,|\, x_2) p(x_5 \,|\, x_3) p(\bar{x}_6 \,|\, x_2, x_5)$$

Variable Elimination: Directed Graphs



$$\begin{aligned} p(x_1, \bar{x}_6) &= \sum_{x_2} \sum_{x_3} \sum_{x_4} \sum_{x_5} p(x_1) p(x_2 \,|\, x_1) p(x_3 \,|\, x_1) p(x_4 \,|\, x_2) p(x_5 \,|\, x_3) p(\bar{x}_6 \,|\, x_2, x_5) \\ &= p(x_1) \sum_{x_2} p(x_2 \,|\, x_1) \sum_{x_3} p(x_3 \,|\, x_1) \sum_{x_4} p(x_4 \,|\, x_2) \sum_{x_5} p(x_5 \,|\, x_3) p(\bar{x}_6 \,|\, x_2, x_5) \end{aligned}$$

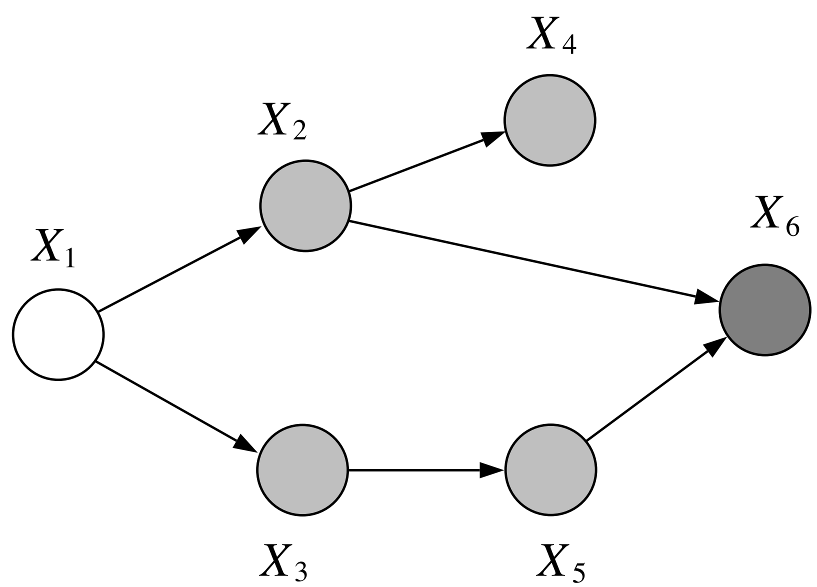
Variable Elimination: Directed Graphs



$$
\begin{aligned}
p(x_1, \bar{x}_6) &= \sum_{x_2} \sum_{x_3} \sum_{x_4} \sum_{x_5} p(x_1) p(x_2 \mid x_1) p(x_3 \mid x_1) p(x_4 \mid x_2) p(x_5 \mid x_3) p(\bar{x}_6 \mid x_2, x_5) \\
&= p(x_1) \sum_{x_2} p(x_2 \mid x_1) \sum_{x_3} p(x_3 \mid x_1) \sum_{x_4} p(x_4 \mid x_2) \sum_{x_5} p(x_5 \mid x_3) p(\bar{x}_6 \mid x_2, x_5) \\
&= p(x_1) \sum_{x_2} p(x_2 \mid x_1) \sum_{x_3} p(x_3 \mid x_1) \sum_{x_4} p(x_4 \mid x_2) m_5(x_2, x_3)
\end{aligned}
$$

where we define $m_5(x_2, x_3) \triangleq \sum_{x_5} p(x_5 \mid x_3) p(\bar{x}_6 \mid x_2, x_5)$.

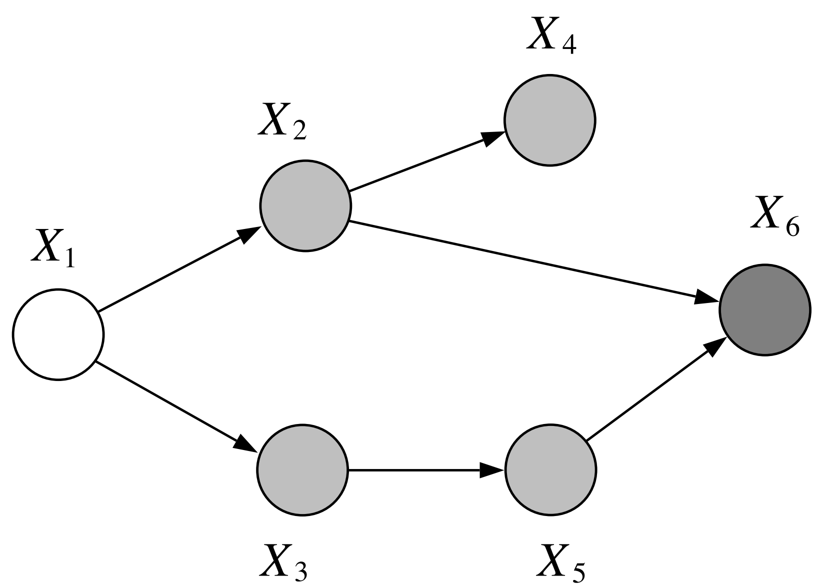
Variable Elimination: Directed Graphs



$$
\begin{aligned}
p(x_1, \bar{x}_6) &= p(x_1) \sum_{x_2} p(x_2 \mid x_1) \sum_{x_3} p(x_3 \mid x_1) m_5(x_2, x_3) \sum_{x_4} p(x_4 \mid x_2) \\
&= p(x_1) \sum_{x_2} p(x_2 \mid x_1) m_4(x_2) \sum_{x_3} p(x_3 \mid x_1) m_5(x_2, x_3).
\end{aligned}
$$

$$m_4(x_2) \triangleq \sum_{x_4} p(x_4 \mid x_2)$$

Variable Elimination: Directed Graphs

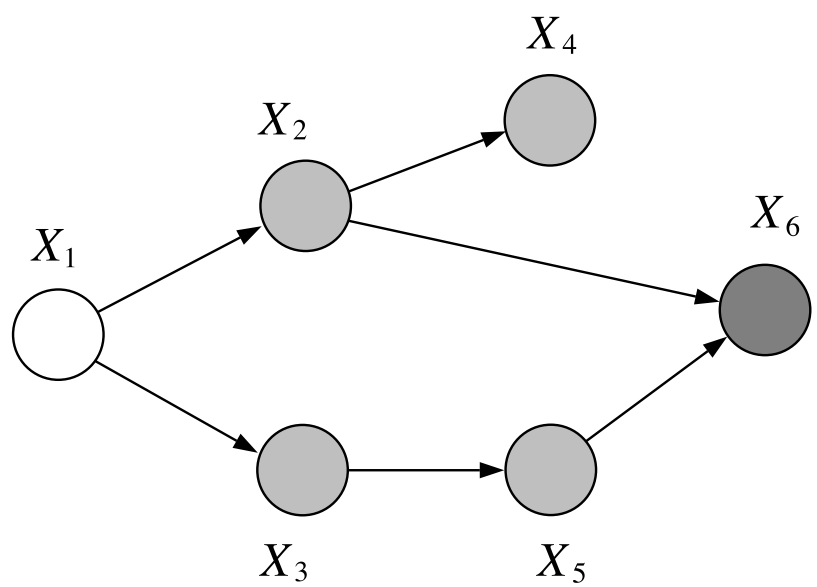


$$
\begin{aligned}
p(x_1, \bar{x}_6) &= p(x_1) \sum_{x_2} p(x_2 \,|\, x_1) \sum_{x_3} p(x_3 \,|\, x_1) m_5(x_2, x_3) \sum_{x_4} p(x_4 \,|\, x_2) \\
&= p(x_1) \sum_{x_2} p(x_2 \,|\, x_1) m_4(x_2) \sum_{x_3} p(x_3 \,|\, x_1) m_5(x_2, x_3).
\end{aligned}
$$

$$m_4(x_2) \triangleq \textstyle\sum_{x_4} p(x_4 \,|\, x_2)$$

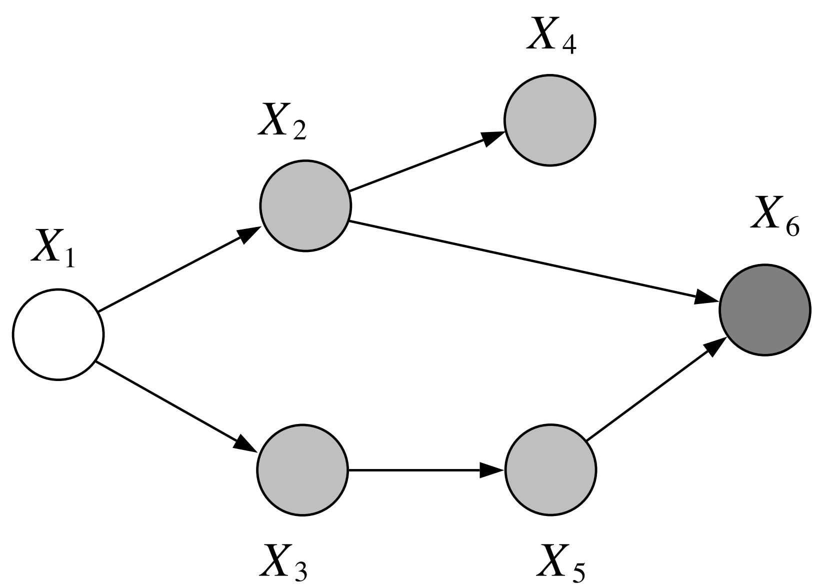
We denote by $m_i(S_i)$ the expression after computing $\sum_{x_i}$ with S_i the index of variables, other than i that appear in the summand

Variable Elimination: Directed Graphs



$$\begin{aligned} p(x_1, \bar{x}_6) &= p(x_1) \sum_{x_2} p(x_2 \,|\, x_1) m_4(x_2) m_3(x_1, x_2) \\ &= p(x_1) m_2(x_1). \end{aligned}$$

Variable Elimination: Directed Graphs

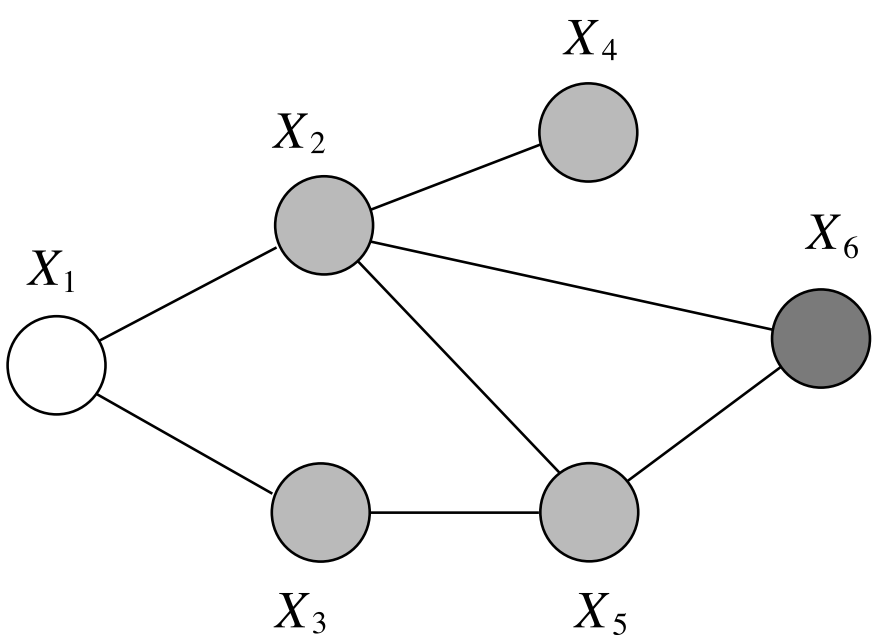


$$\begin{aligned} p(x_1, \bar{x}_6) &= p(x_1) \sum_{x_2} p(x_2 \,|\, x_1) m_4(x_2) m_3(x_1, x_2) \\ &= p(x_1) m_2(x_1). \end{aligned}$$

$$p(\bar{x}_6) = \sum_{x_1} p(x_1) m_2(x_1),$$

$$p(x_1 \,|\, \bar{x}_6) = \frac{p(x_1) m_2(x_1)}{\sum_{x_1} p(x_1) m_2(x_1)}.$$

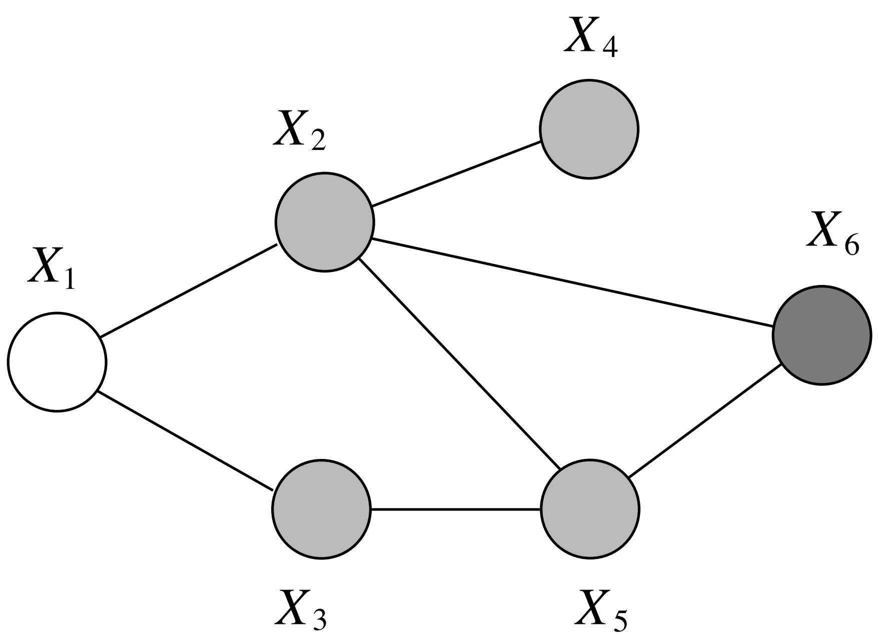
Variable Elimination: undirected graphs



- Potentials $\{\psi_C(x_C)\}$ on the cliques $\{X_1, X_2\}, \{X_1, X_3\}, \{X_2, X_4\}, \{X_3, X_5\}$, and $\{X_2, X_5, X_6\}$.

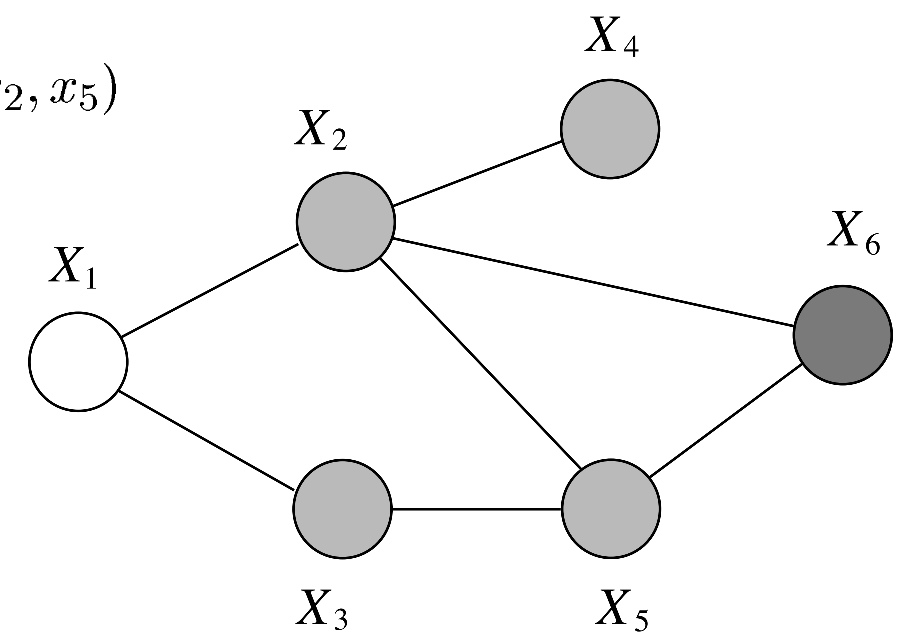
Elimination; undirected graphs

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



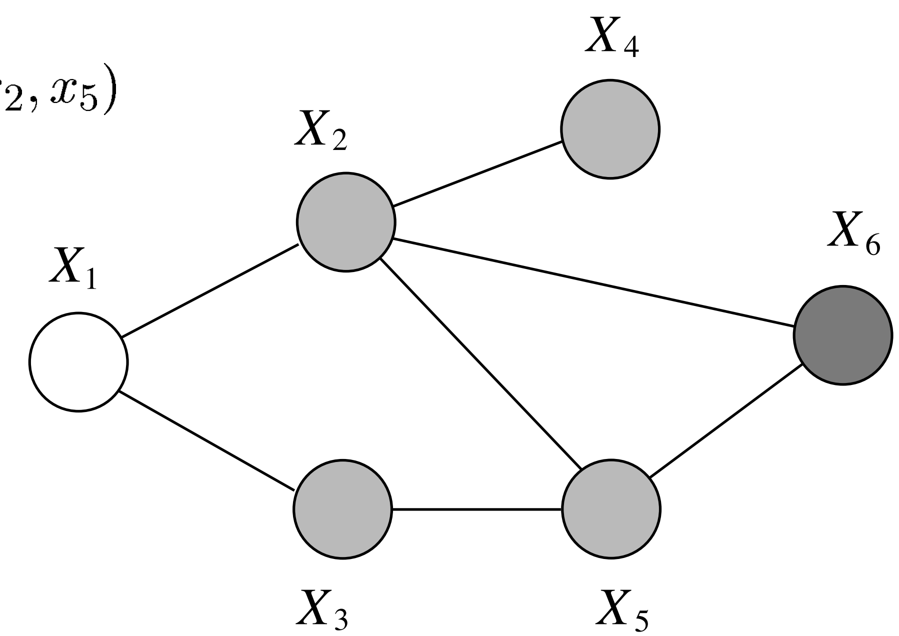
Elimination; undirected graphs

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



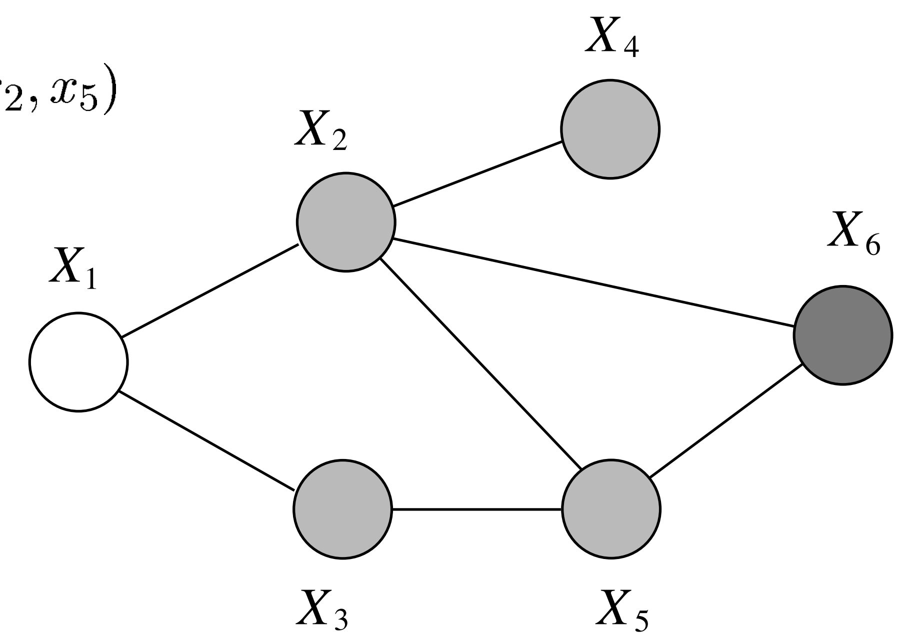
Elimination; undirected graphs

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



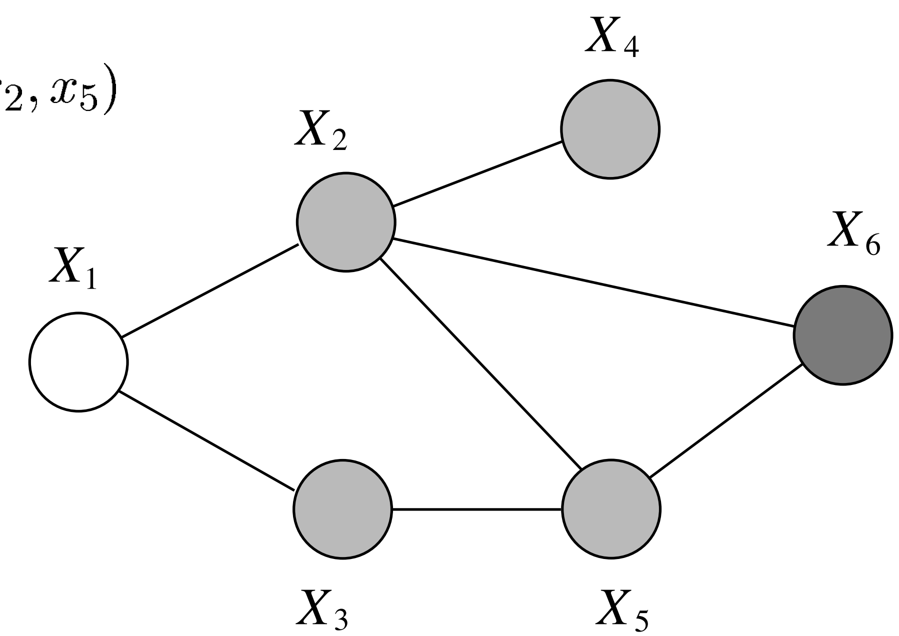
Elimination; undirected graphs

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



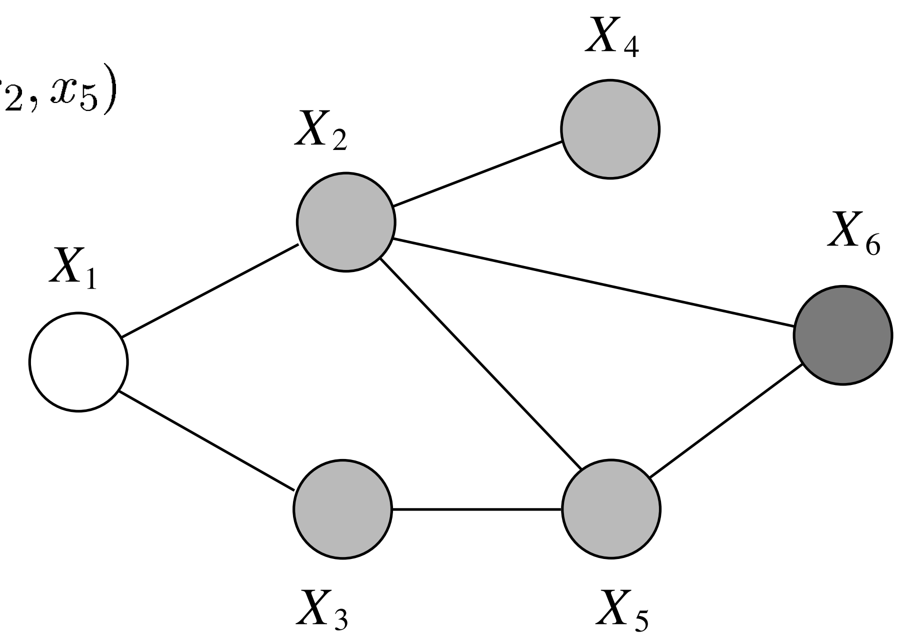
Elimination; undirected graphs

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



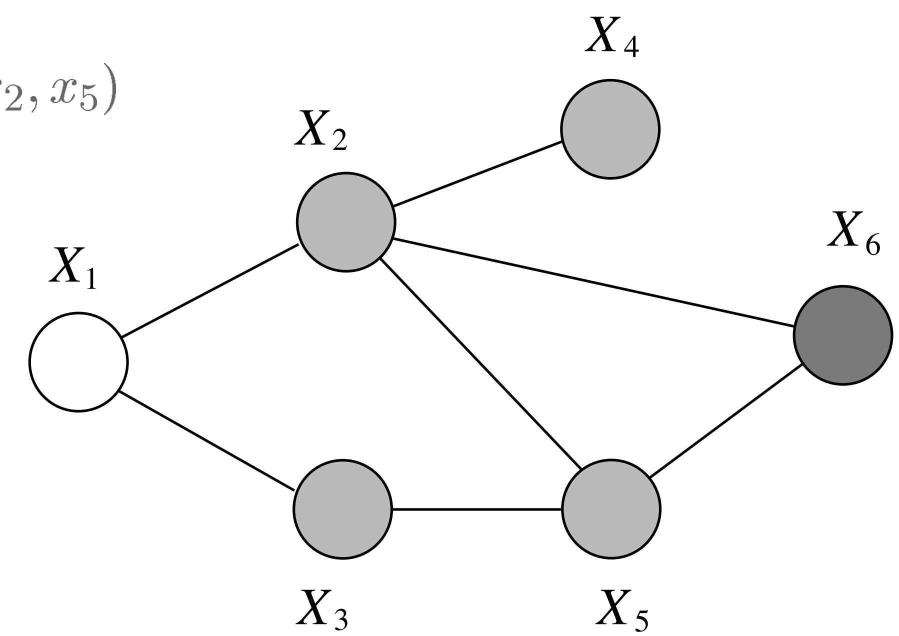
Elimination; undirected graphs

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



Elimination; undirected graphs

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



$$
p(\bar{x}_6) = \frac{1}{Z} \sum_{x_1} m_2(x_1),
$$

$$
p(x_1 \mid \bar{x}_6) = \frac{m_2(x_1)}{\sum_{x_1} m_2(x_1)},
$$

<--- No Z!

A graph-theoretic view of elimination

- Till now, we have seen an algebraic view of probabilistic inference, using factorizations to simplify calculations
- How does this play out graph-theoretically?

A graph-theoretic view of elimination

UndirectedGraphEliminate($\mathcal{G}, I$)

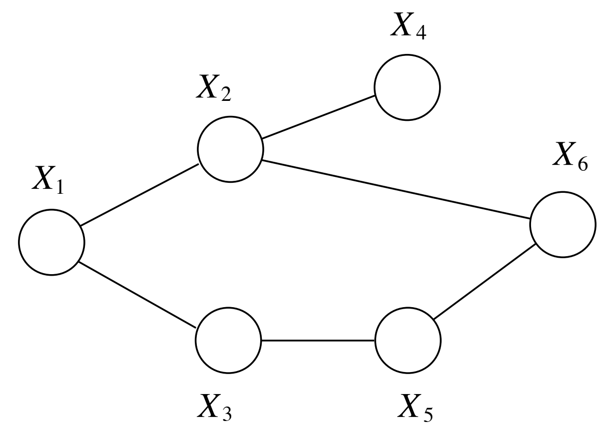
  for each node X_i in I

    connect all of the remaining neighbors of X_i

    remove X_i from the graph

  end

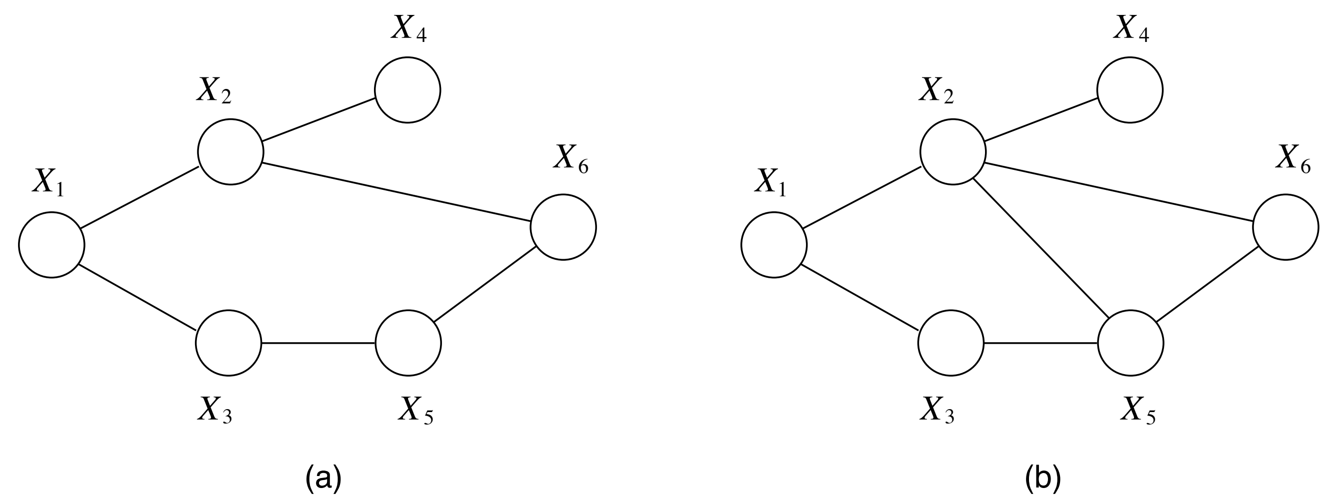
A graph-theoretic view of elimination



(a)

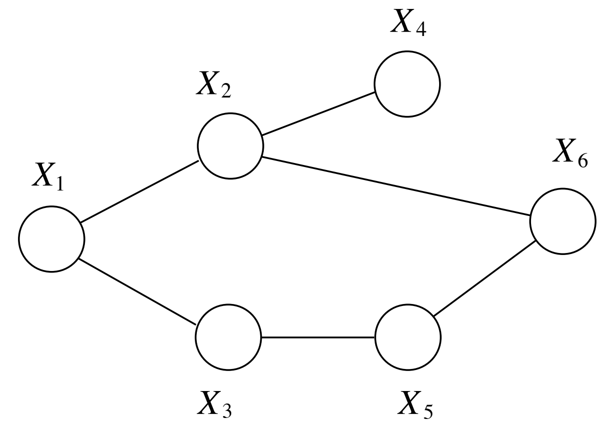
$\textsc{UndirectedGraphEliminate}(\mathcal{G}, I)$
for each node X_i in I
connect all of the remaining neighbors of X_i
remove X_i from the graph
end

A graph-theoretic view of elimination

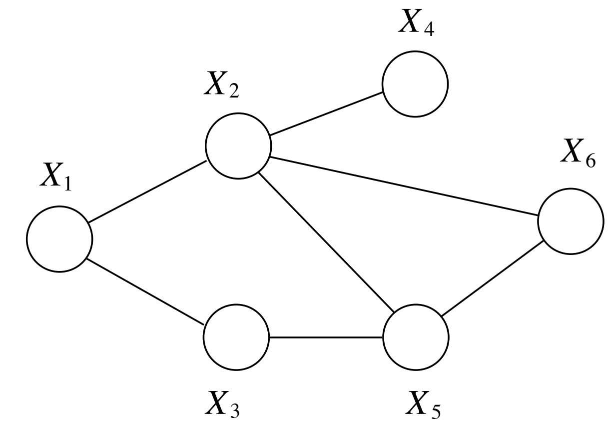


UndirectedGraphEliminate($\mathcal{G}, I$)
 for each node X_i in I
  connect all of the remaining neighbors of X_i
  remove X_i from the graph
 end

A graph-theoretic view of elimination



(a)



(b)

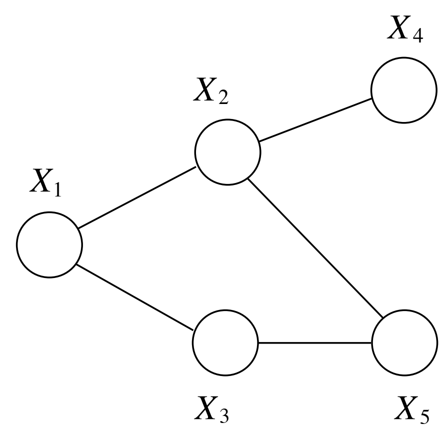
UNDIRECTEDGRAPHELIMINATE($\mathcal{G}, I$)

    for each node X_i in I

        connect all of the remaining neighbors of X_i

        remove X_i from the graph

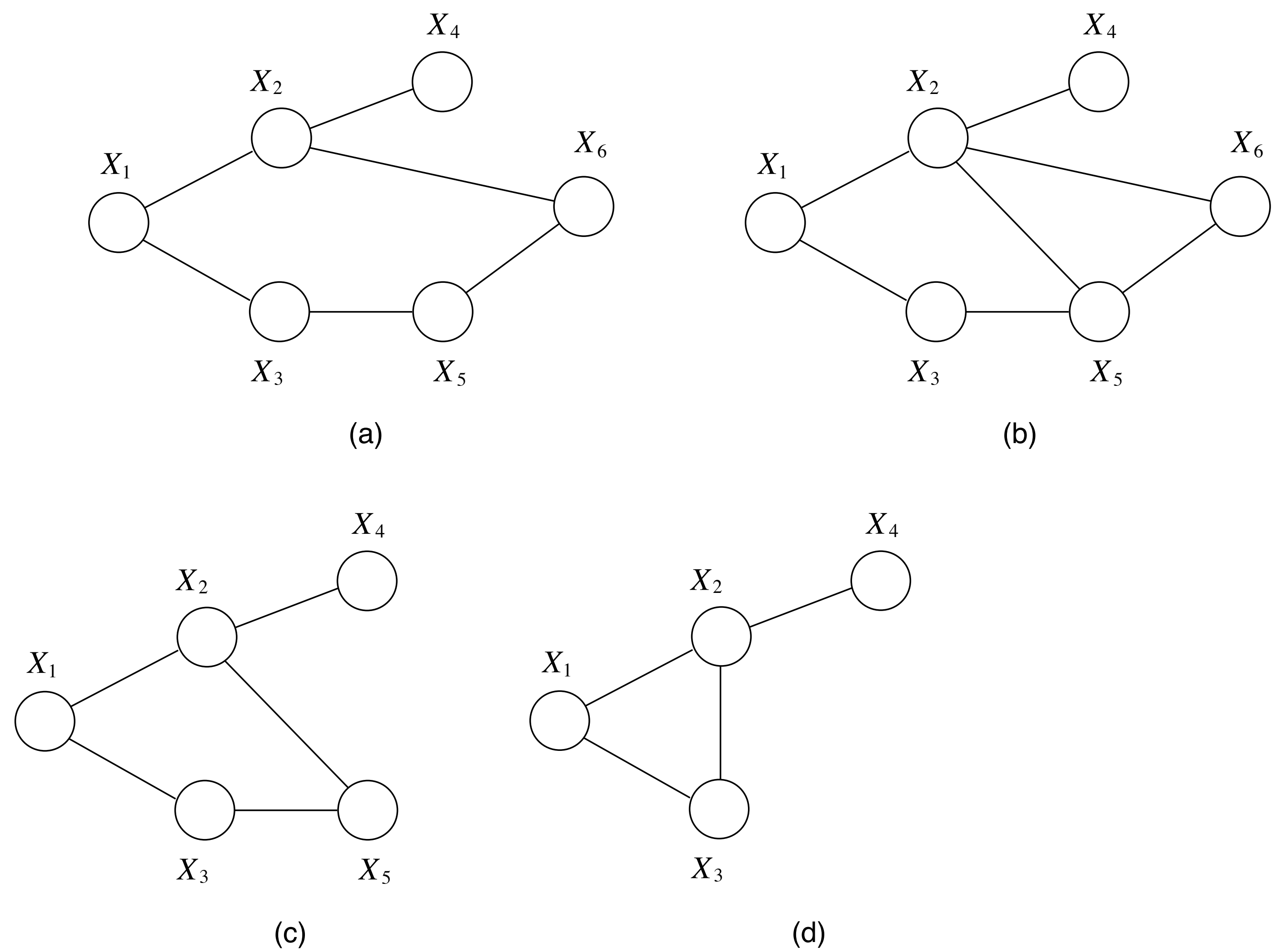
    end



(c)

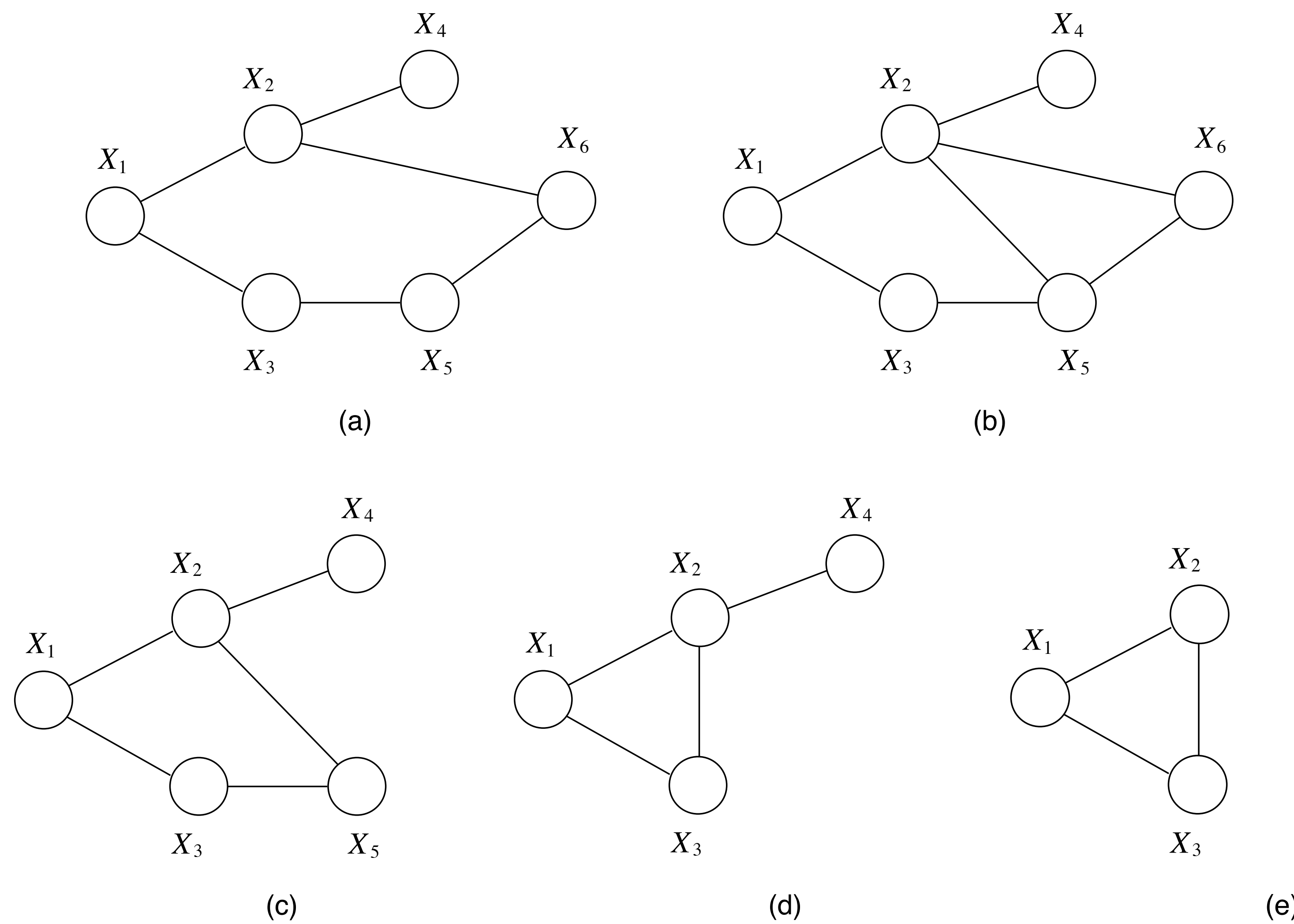
A graph-theoretic view of elimination

UNDIRECTEDGRAPHELIMINATE($\mathcal{G}, I$)
    for each node X_i in I
        connect all of the remaining neighbors of X_i
        remove X_i from the graph
    end



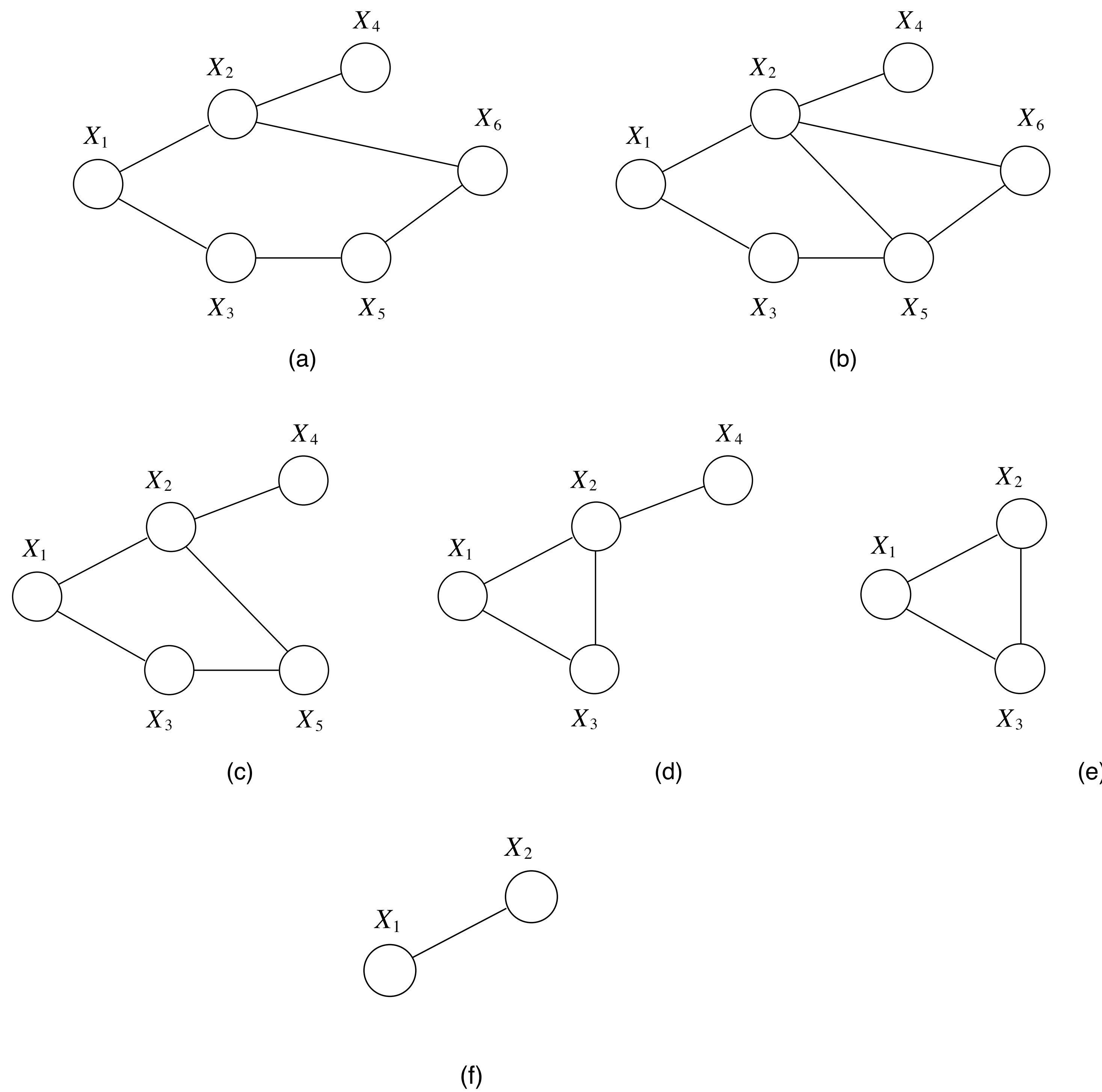
A graph-theoretic view of elimination

UndirectedGraphEliminate($\mathcal{G}, I$)
- **for** each node X_i in I
  - connect all of the remaining neighbors of X_i
  - remove X_i from the graph
- **end**



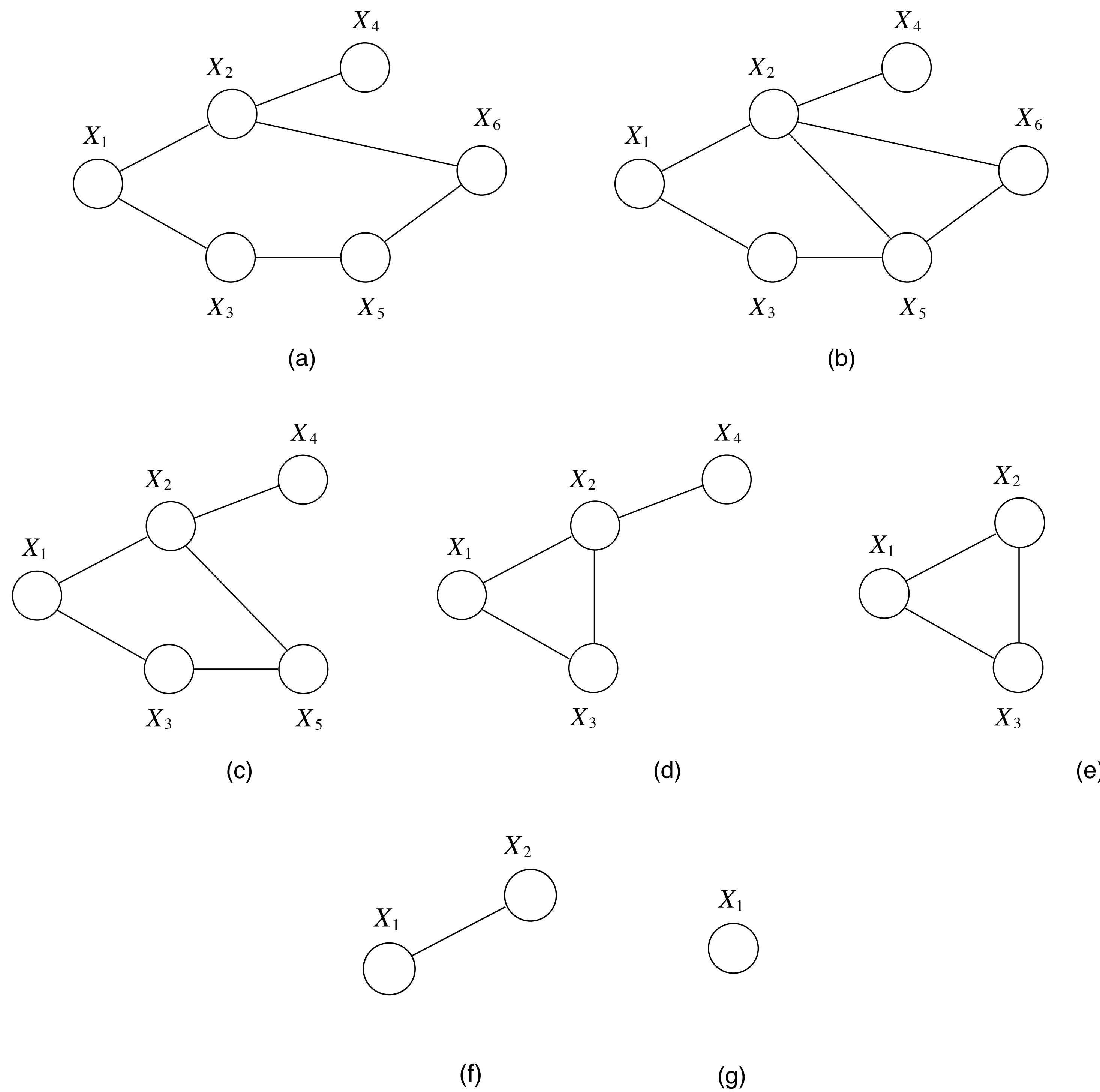
A graph-theoretic view of elimination

UNDIRECTEDGRAPHELIMINATE($\mathcal{G}, I$)
for each node X_i in I
connect all of the remaining neighbors of X_i
remove X_i from the graph
end

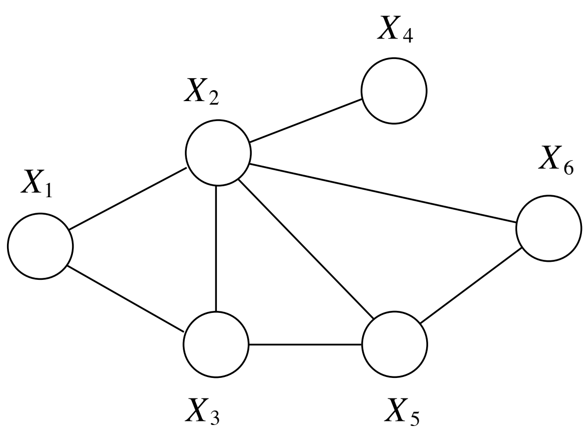


A graph-theoretic view of elimination

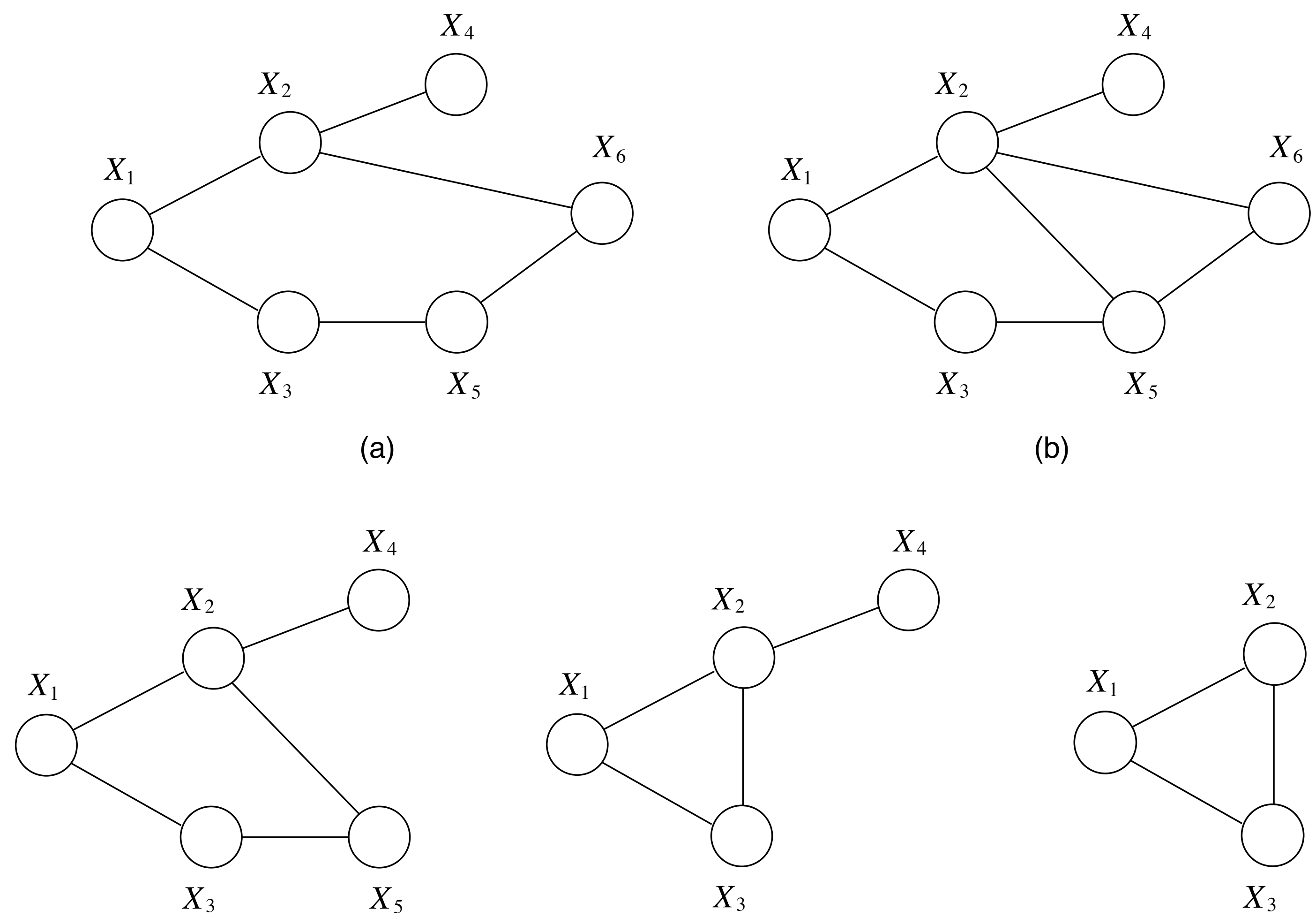
$\textsc{UndirectedGraphEliminate}(\mathcal{G}, I)$
  for each node X_i in I
    connect all of the remaining neighbors of X_i
    remove X_i from the graph
  end



A graph-theoretic view of elimination



Reconstituted Graph



(a) (b) (c) (d)

- Elimination Cliques: when you remove a node, the set of nodes neighboring it, including node itself; denote by T_i
- Example: $T_6 = \{2, 5, 6\}$, and $T_5 = \{2, 3, 5\}$

Graph Elimination and Marginalization

Proposition: Elimination Cliques in UndirectedGraphEliminate correspond to sets of variables on which summations operate in Eliminate.

Computational Complexity

- Given previous proposition, computational complexity depends on a purely graph-theoretic quantity: the size of the largest elimination clique created by UNDIRECTEDGRAPHELIMINATE

Computational Complexity

- Given previous proposition, computational complexity depends on a purely graph-theoretic quantity: the size of the largest elimination clique created by UNDIRECTEDGRAPHELIMINATE

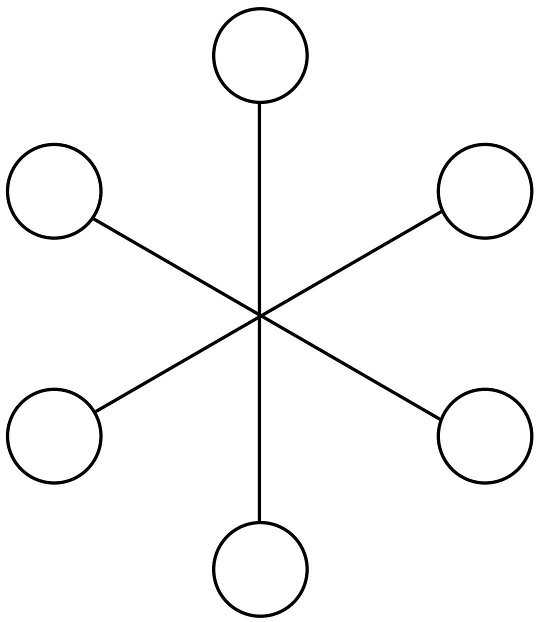
- Note that the largest such clique depends on the elimination ordering; we want the minimum over all possible orderings (since the ordering is under our control)

  - A well-studied problem in graph-theory

  - Tree-width: one minus the size of the smallest achievable largest elimination clique (ranging over all elimination orderings)

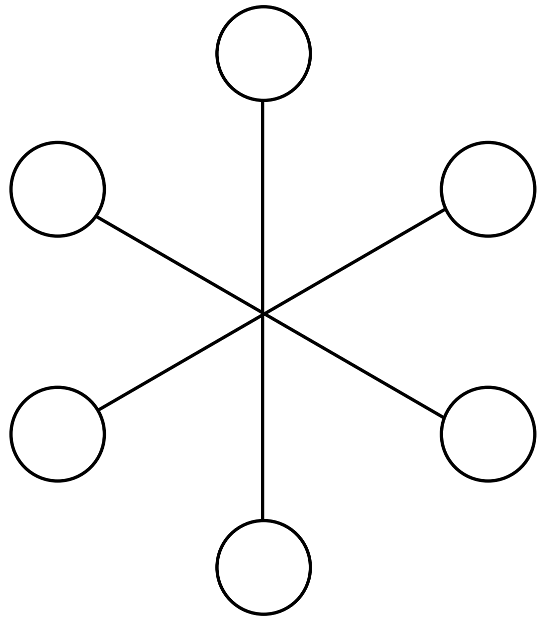
  - But NP-hard to find this best possible elimination ordering

Treewidth

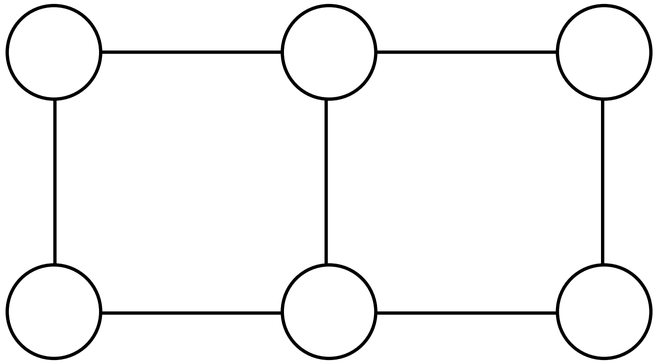


One!

Treewidth



One!



Two

Graph Elimination; Directed Graphs

$\textsc{DirectedGraphEliminate}(G, I)$

$G^m = \textsc{Moralize}(G)$

$\textsc{UndirectedGraphEliminate}(G^m, I)$

$\textsc{Moralize}(G)$

for each node X_i in I

connect all of the parents of X_i

end

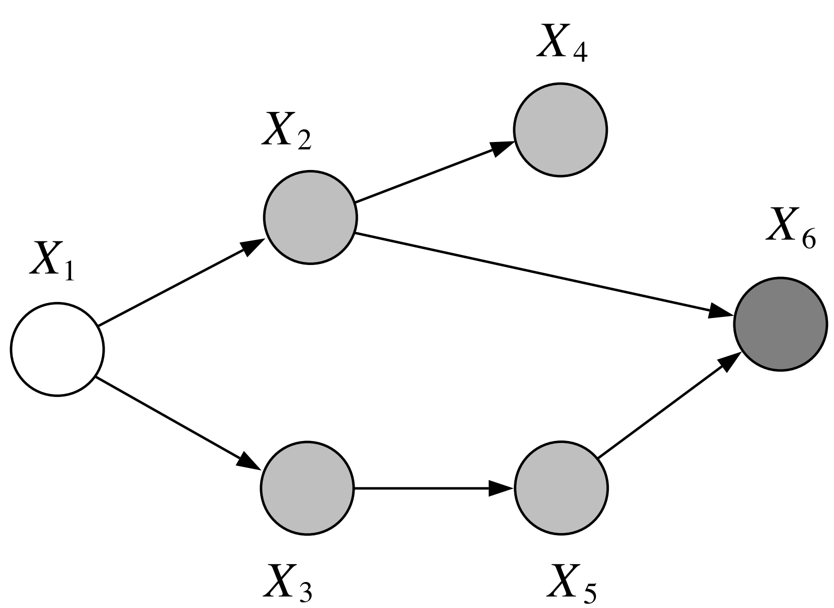
drop the orientation of all edges

return G

Graph Elimination; Directed Graphs

DirectedGraphEliminate(G, I)
    G^m = Moralize(G)
    UndirectedGraphEliminate(G^m, I)

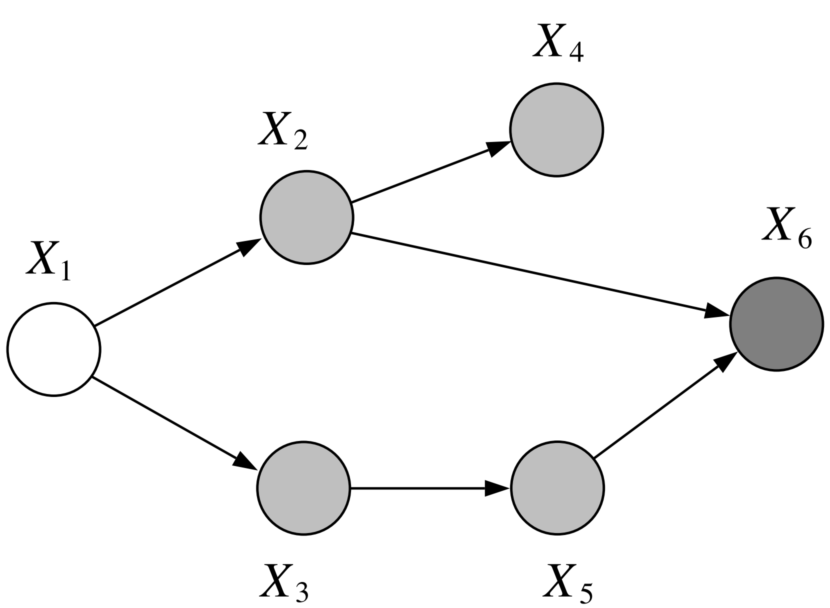
Moralize(G)
    for each node X_i in I
        connect all of the parents of X_i
    end
    drop the orientation of all edges
    return G

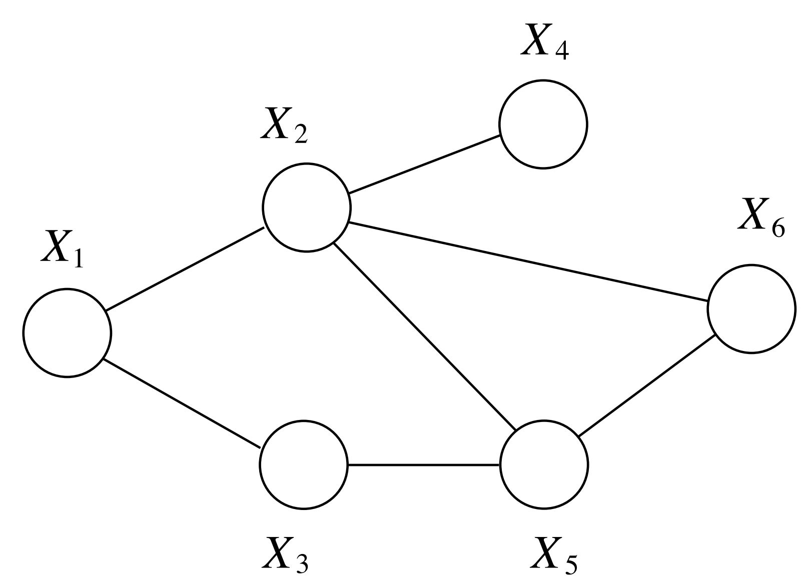


Graph Elimination; Directed Graphs

DIRECTEDGRAPHELIMINATE(G, I)
    G^m = MORALIZE(G)
    UNDIRECTEDGRAPHELIMINATE(G^m, I)

MORALIZE(G)
    for each node X_i in I
        connect all of the parents of X_i
    end
    drop the orientation of all edges
    return G





Moralized Graph