COS513: FOUNDATIONS OF PROBABILISTIC MODELING

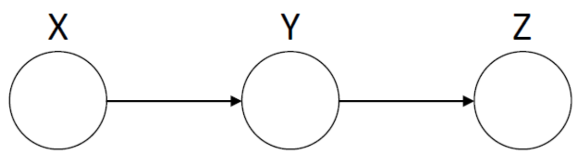
LECTURE 4

ERAN ELDAR

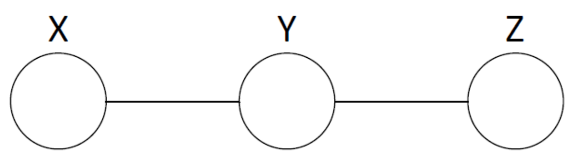
1. Directed and Undirected Graphs

Some directed graphs can be represented with an identical undirected graph.

For example, the following directed graph:



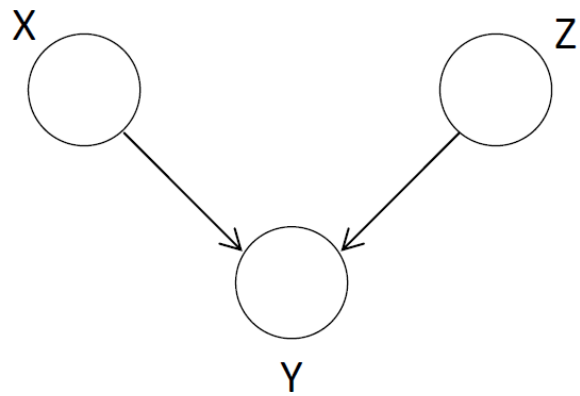
is identical to the following undirected graph:



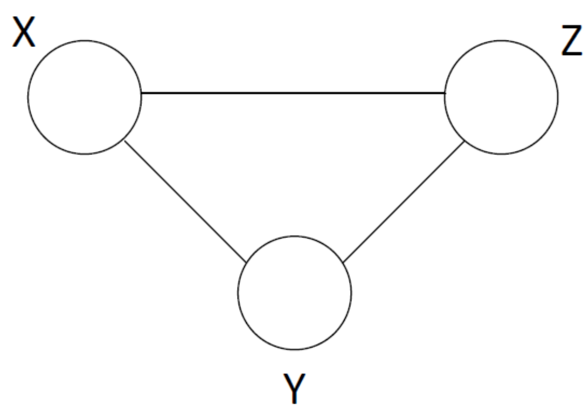
as both graphs imply:

$$X \perp\!\!\!\perp Z \mid Y$$
$$X \not\!\perp\!\!\!\perp Z$$

However, there is no undirected graph that is identical to the following directed graph:



When this is the case, we can use 'moralization' to construct an undirected graph that represents a family of probability distributions which includes (though is not identical to) the family of probability distributions that is represented by the undirected graph. 'Moralizing' is done by connecting the parents of each node. The result of 'moralizing' the graph above is:



2. Inference

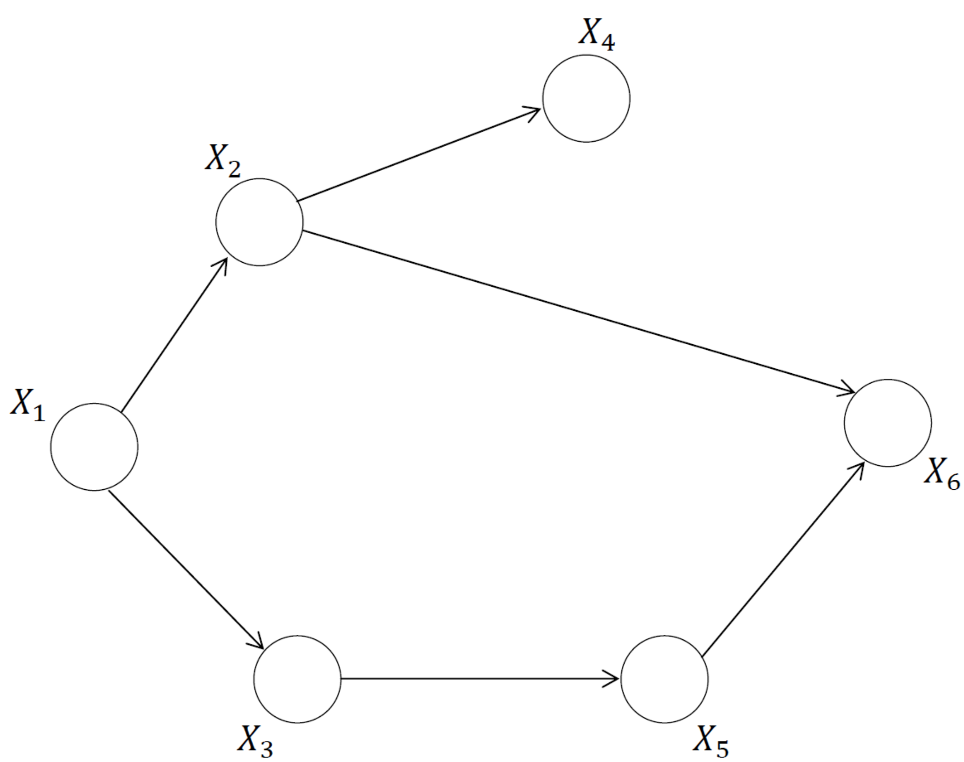
f - node index
E - set of 'evidence' nodes
R - remaining nodes

Goal: compute $P(X_f \mid X_e)$

Step 1: $P(X_f, X_E) = \sum_{X_R} P(X)$
Step 2: $P(X_E) = \sum_{X_f} P(X_f, X_E)$
Step 3: $P(X_f \mid X_e) = \frac{P(X_f, X_E)}{P(X_E)}$

Issue in step 1: if R contains many nodes, the inference is $O(K^{|R|})$, which is usually not pratical.

Let's return to our favorite graphical model:



Our goal is to compute $P(X_1 \mid X_6)$

Accordingly, we define:

$f = X_1$
$E = \{X_6\}$
$R = \{X_2, X_3, X_4, X_5\}$

$\overline{X}_6 = X_6$ clamped to the value we are conditioning on.

$$\begin{aligned}
P(X_1, \overline{X}_6) &= \sum_{X_2}\sum_{X_3}\sum_{X_4}\sum_{X_5}\sum_{X_6} P(X_1)P(X_2 \mid X_1)P(X_3 \mid X_1)P(X_4 \mid X_2) \\
&\quad P(X_5 \mid X_3)P(X_6 \mid X_2, X_5)\delta(X_6, \overline{X}_6) \\
&= P(X_1)\sum_{X_2} P(X_2 \mid X_1)\sum_{X_3} P(X_3 \mid X_1)\sum_{X_4} P(X_4 \mid X_2) \\
&\quad \sum_{X_5} P(X_5 \mid X_3)\sum_{X_6} P(X_6 \mid X_2, X_5)\delta(X_6, \overline{X}_6)
\end{aligned}$$

We then define $m_6(X_2, X_5) \triangleq \sum_{X_6} P(X_6 \mid X_2, X_5)\delta(X_6, \overline{X}_6)$ and get:

$$P(X_1, \overline{X}_6) = P(X_1)\sum_{X_2} P(X_2 \mid X_1)\sum_{X_3} P(X_3 \mid X_1)\sum_{X_4} P(X_4 \mid X_2)\sum_{X_5} P(X_5 \mid X_3)m_6(X_2, X_5)$$

We then define $m_5(X_2, X_3) \triangleq \sum_{X_5} P(X_5 \mid X_3)m_6(X_2, X_5)$ and get:

$$\begin{aligned}
P(X_1, \overline{X}_6) &= P(X_1)\sum_{X_2} P(X_2 \mid X_1)\sum_{X_3} P(X_3 \mid X_1)\sum_{X_4} P(X_4 \mid X_2)m_5(X_2, X_3) \\
&= P(X_1)\sum_{X_2} P(X_2 \mid X_1)\sum_{X_3} P(X_3 \mid X_1)m_5(X_2, X_3)\sum_{X_4} P(X_4 \mid X_2)
\end{aligned}$$

We then define $m_4(X_2) \triangleq \sum_{X_4} P(X_4 \mid X_2)$ which equals 1 and get:

$$\begin{aligned} P(X_1, \overline{X}_6) &= P(X_1) \sum_{X_2} P(X_2 \mid X_1) \sum_{X_3} P(X_3 \mid X_1) m_5(X_2, X_3) m_4(X_2) \\ &= P(X_1) \sum_{X_2} P(X_2 \mid X_1) \sum_{X_3} P(X_3 \mid X_1) m_5(X_2, X_3) \end{aligned}$$

We then define $m_3(X_1, X_2) \triangleq \sum_{X_3} P(X_3 \mid X_1) m_5(X_2, X_3)$ and get:

$$P(X_1, \overline{X}_6) = P(X_1) \sum_{X_2} P(X_2 \mid X_1) m_3(X_1, X_2)$$

We then define $m_2(X_1) \triangleq \sum_{X_2} P(X_2 \mid X_1) m_3(X_1, X_2)$ and get:

$$P(X_1, \overline{X}_6) = P(X_1) m_2(X_1)$$

Lastly, we compute:

1: $P(X_1, \overline{X}_6) = P(X_1) m_2(X_1)$

2: $P(\overline{X}_6) = \sum_{X_1} P(X_1) m_2(X_1)$

3: $P(X_1, \overline{X}_6) = \frac{P(X_1, \overline{X}_6)}{P(\overline{X}_6)}$

3. Elimination algorithm

At each step, sum over a product of functions:
- conditional probabilities $P(X_i \mid X_{\pi_i})$
- delta functions $\delta(X_i, \overline{X}_i)$
- internediate functions $m_i(X_{S_i})$, generated by previous steps of the algorithm.

I. Initialization

1. Choose an ordering of the variables such that X_f is last.
2. Place $P(X_i \mid X_{\pi_i})$ on an active list of functions.
3. Place $\delta(X_j, \overline{X}_j)$ on active list for all $j \in E$

II. Update

For each $i \in I$:
1. Remove all functions on the active list that contain i as an argument ($\phi_i(T_i)$ in the book).
2. Construct $m_i(S_i) = \sum_{X_i} \prod these functions.$

   Note: nothing before i in I can be in S_i or T_i.
3. Add $m_i(S_i)$ to the active list.

III. Normalize

$\phi_f(X_f) = P(X_f, \overline{X}_E)$

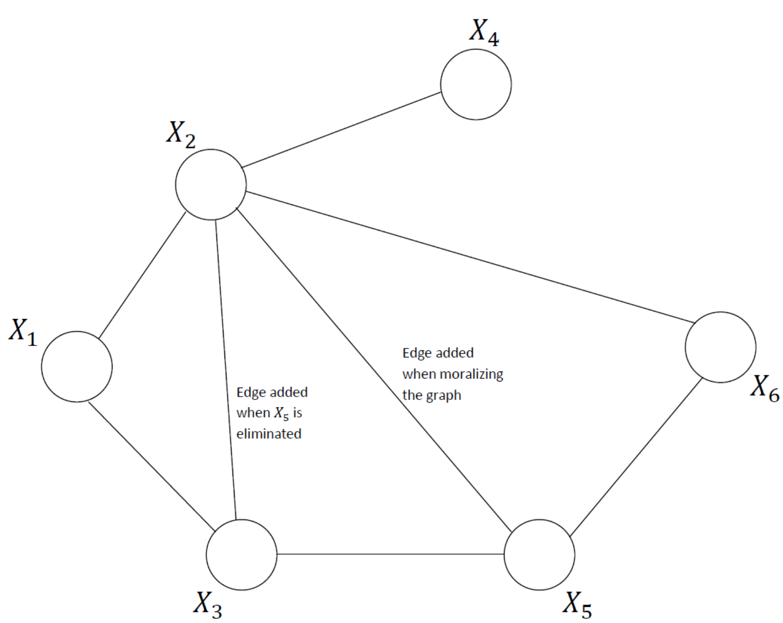
$P(X_f \mid \overline{X}_E) = \frac{\phi_f(X_f)}{\sum_{X_f} \phi_f(X_f)}$

Complexity

$I = \{6, 5, 4, 3, 2, 1\}$

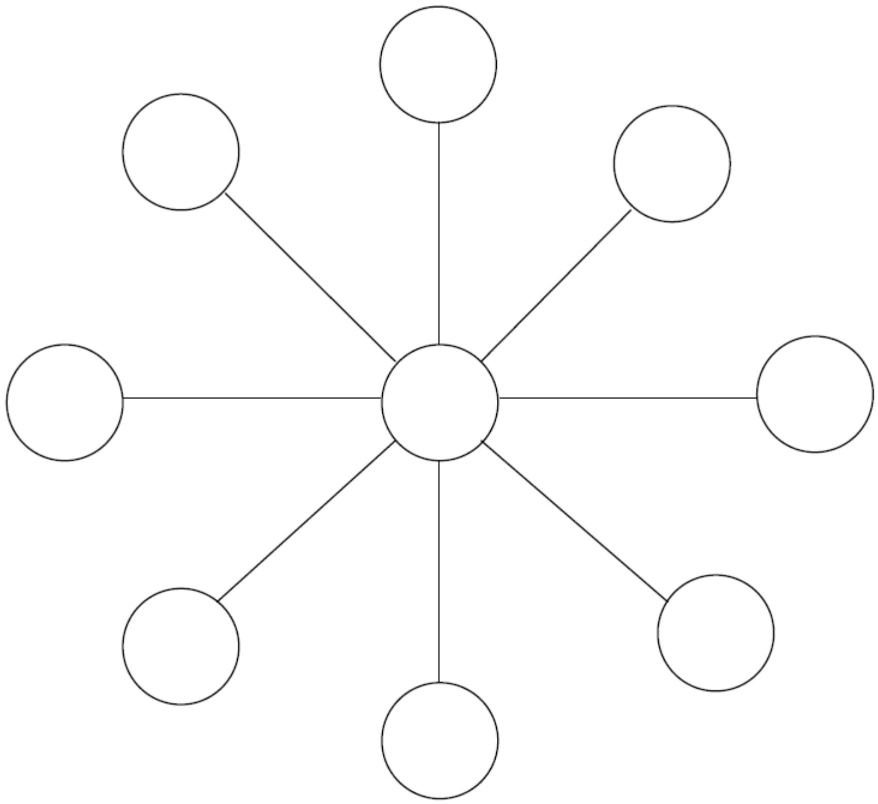
The complexity of Eliminate is governed by the number of arguments in the intermediate functions, which depends on the ordering I.

To determine the complexity for a certain ordering I, we first 'moralize' our graph and then repeatedly remove a node according to I and connect all of the nodes that it was connected to. The result, called the 'reconstituted' graph, is shown below for our example graphical model:



The complexity of using the Eliminate algorithm with the ordering I is exponential in the largest clique of the reconstituted graph.

For example, consider the following graph:



If the central node is removed first, all other nodes will need to be connected, forming a clique of size 8. This indicates that the complexity of Eliminate with this ordering will be relatively high (exponential in 8).

In contrast, if all of the leaf nodes are removed first, the largest clique that is formed at any point has size 2. With this ordering, the complexity of Eliminate will be lower (exponential in 2).