Solutions to Week 4 Problems

Compiled by Ernesto Brau and Anurag Katiyar

Note: most of these solutions are solutions by students.

1. The first half of chapter 8 is about diagrammatic representations of probability distributions known as graphical models. It talks about how joint probability distribution over random variables can be broken down in to factors. The graphs help us to infer the concept of conditional independence. The end the first half talks about the potential functions, partitions, markov blanket and markov random fields. The second half of chapter 8 (from 8.3.4 on) is mostly about inference on graphical models. We first need to decide what we mean by inference, in terms of graphical models: calculating marginal and posterior distributions given a joint distribution. The author first discusses an efficient way to do inference (to calculate the marginal probability of all of the variables, given the joint distribution) when the graph is a chain. From there, we can generalize the result to work trees and polytrees, using the new concept of 'factor graph'. On these factor graphs, we can apply two algorithms – sum-product and max-sum – to infer marginal probabilities. The chapter ends with a brief description on how to solve inference problems on general graphs.

2. For ease of reading, I will define $\psi_{i,j} = \psi_{i,j}(x_i, x_j)$. If we fill in the steps, we have

$$
\begin{aligned}
p(x_n) &= \sum_{x_1} \cdots \sum_{x_{n-1}} \sum_{x_{n+1}} \cdots \sum_{x_N} p(\mathbf{x}) \\
&= \sum_{x_1} \cdots \sum_{x_N} \psi_{N-1,N} \cdots \psi_{1,2} \\
&= \sum_{x_1} \cdots \sum_{x_{N-1}} \left[\psi_{N-2,N-1} \cdots \psi_{1,2} \sum_{x_N} \psi_{N-1,N} \right] \\
&\ \ \vdots \\
&= \sum_{x_1} \cdots \sum_{x_{n+1}} \left[\psi_{n,n+1} \cdots \psi_{1,2} \sum_{x_{n+2}} \left[\psi_{n+1,n+2} \cdots \sum_{x_N} \psi_{N-1,N} \right] \cdots \right] \\
&= \sum_{x_1} \cdots \sum_{x_{n-1}} \left[\psi_{n-1,n} \cdots \psi_{1,2} \sum_{x_{n+1}} \left[\psi_{n,n+1} \cdots \sum_{x_N} \psi_{N-1,N} \right] \cdots \right] \\
&\ \ \vdots \\
&= \sum_{x_1} \left[\psi_{1,2} \cdots \sum_{x_{n-1}} \left[\psi_{n-1,n} \sum_{x_{n+1}} \left[\psi_{n,n+1} \cdots \sum_{x_N} \psi_{N-1,N} \right] \right] \cdots \right].
\end{aligned}
$$

Up to now, we have only used the fact that the sum over each x_k only affects two potential functions, and the rest can be taken out of the summation. We can do the same thing one more time – noting that everything to the right of the summation over x_{n+1} is independent of everything to the left of it – to get

$$\begin{aligned} p(x_n) &= \sum_{x_1} \left[\psi_{1,2} \cdots \sum_{x_{n-1}} \left[\psi_{n-1,n} \sum_{n+1} \left[\psi_{n,n+1} \cdots \sum_{x_N} \psi_{N-1,N} \right] \right] \cdots \right] \\ &= \left[\sum_{n+1} \left[\psi_{n,n+1} \cdots \sum_{x_N} \psi_{N-1,N} \right] \right] \left[\sum_{x_1} \left[\psi_{1,2} \cdots \sum_{x_{n-1}} \psi_{n-1,n} \right] \right]. \end{aligned}$$

3. (a) A standard directed tree is moral because every node has one unique parent. Hence the moral graph could be achieved by simply dropping the arrows.

   (b) No it is not moral because its a polytree where each node has more than one parent. The graph that converts this directed graph to a moral graph is at the end of this pdf document.

   (c) In this case, the potential functions are just products of the conditional probabilities that make up the joint distribution. Remembering that each clique has a potential function, we have

$$\begin{aligned} \psi_{1,3,4,5} &= p(x_1)p(x_3)p(x_5|x_1,x_3) \\ \psi_{4,5,7} &= p(x_7|x_5,x_4) \\ \psi_{4,6} &= p(x_6|x_4) \\ \psi_{1,2,3,4} &= p(x_2)p(x_4|x_1,x_2,x_3) \end{aligned}$$

   (d) Graphs included at the end of the pdf.

   (e) To construct the factor graph we first reprsent the probability distribution given by the graph

$$p(x_2)p(x_1)p(x_3)p(x_4|x_1,x_2,x_3)p(x_5|x_1,x_3)p(x_7|x_5,x_4)p(x_6|x_4).$$

   We now create variable nodes in the factor graph corresponding to the nodes of the directed graph and then create factor nodes corresponding to the conditional distribution and finally add appropriate nodes. See the figure at the end of this pdf. Each factor will be equal to the confitional probability of the nodes it neighbors in the graph. If done correctly, each factor node will only neighbor nodes that are dependant on each other in some way.

4. If we use the fact that $p(\mathbf{x}) = \prod_{s\in\text{ne}(x_n)} F_s(x, X_s)$, for any $x \in X_s$, we have

$$
\begin{aligned}
p(x) &= \sum_{\mathbf{x}\backslash x} \prod_{s\in\text{ne}(x_n)} F_s(x, X_s) \\
&= \sum_{X_1}\sum_{X_2}\cdots\sum_{X_M} [F_1(x, X_1)F_2(x, F_2)\cdots F_M(x, X_M)] \\
&= \left[\sum_{X_1} F_1(x, X_1)\right]\left[\sum_{X_2} F_2(x, X_2)\right]\cdots\left[\sum_{X_M} F_M(x, X_M)\right] \\
&= \prod_{s\in\text{ne}(x_n)} \left[\sum_{X_s} F_s(x, X_s)\right].
\end{aligned}
$$

In order to better understand this, we must remember that $X_1 \cup X_2 \cup \cdots \cup X_M = \mathbf{x}\backslash x$, where $X_1, \cdots, X_M$ are the neighbors of x.

5. Page 409 shows an illustration of the sum-product algorithm applied to a particular tree in Figure 8.51. Our goal is to get the marginal distributions of the variable nodes. For that we designate one of the nodes as the root, x_3 in this case and locate the leaf nodes which are x_1 and x_4 here. Then we compute all the messages from factor to variable nodes and variable to factor nodes starting from the root to the leaves and from leaves to root. In this example there are 3 variable to factor messages and 3 factor to variable messages in leaf-to-root direction; 3 variable to factor and 3 factor to variable messages in root-to-leaves direction. These messages are computed recursively and stored. Using these we can get the marginal distributions of all the variable nodes, which are

$$
\begin{aligned}
p(x_1) &= \mu_{f_a\to x_1}(x_1) \\
p(x_2) &= \mu_{f_a\to x_1}(x_2)\mu_{f_b\to x_1}(x_2)\mu_{f_c\to x_1}(x_2) \\
p(x_3) &= \mu_{f_b\to x_3}(x_3).
\end{aligned}
$$

6. The required joint probablity distribution can be given by Table 1. As can be seen, $\widehat{x}$ that maximizes the marginal $p(x)$ is given by $x = 1$ and $\widehat{y}$ that maximizes the marginal $p(y)$ is given by $y = 1$, and they together have probability zero under the joint distribution.

7. (a) The max sum algorithm helps to find the value of the variables which give the maximum probability of the distribution (joint or marginal).

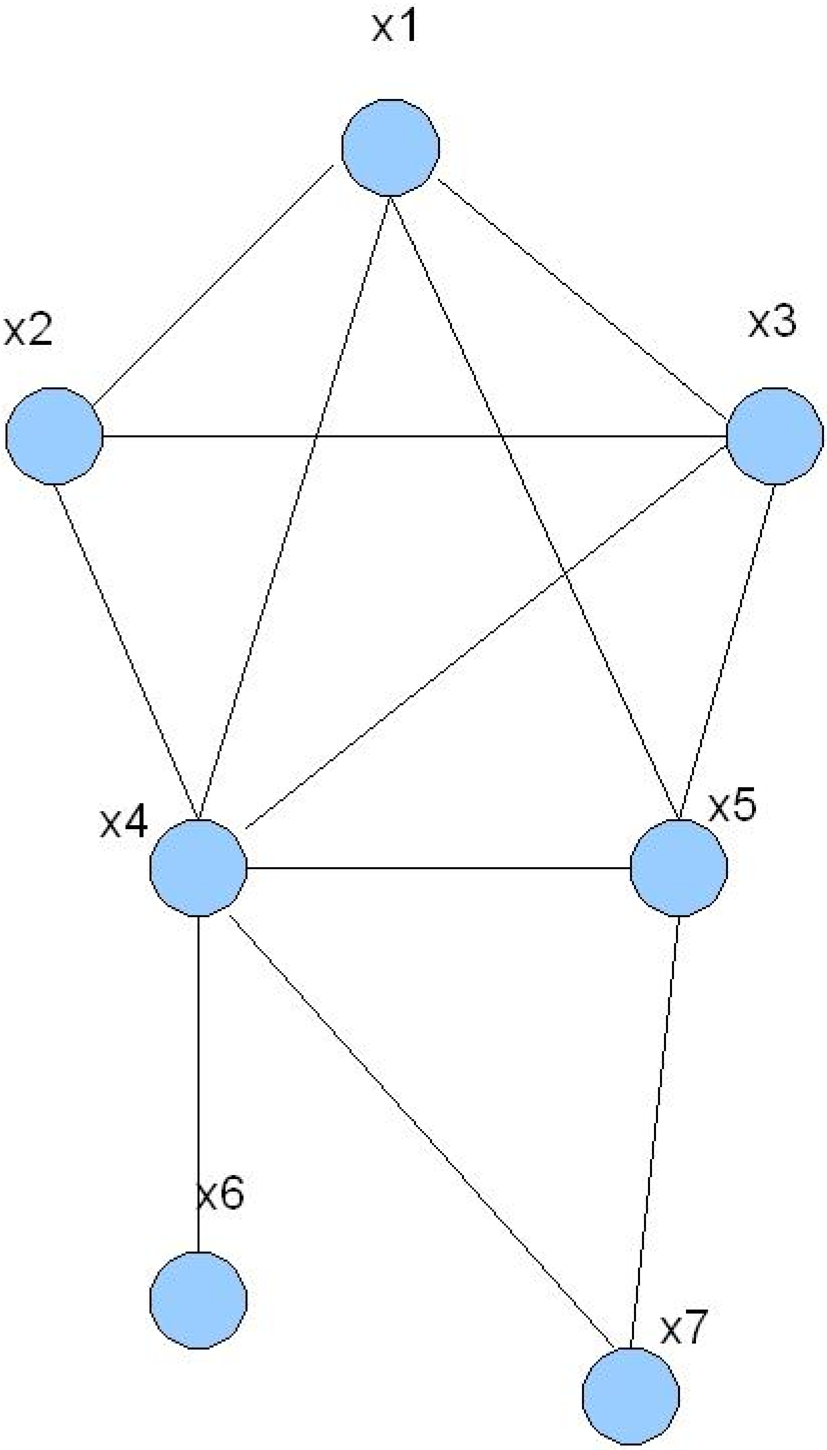
   (b) The modified version of the algorithm computes the most probable value of the node x_N. We calculate the configuration of

Table 1: Joint probability distribution of two variables x and y each having 3 states.

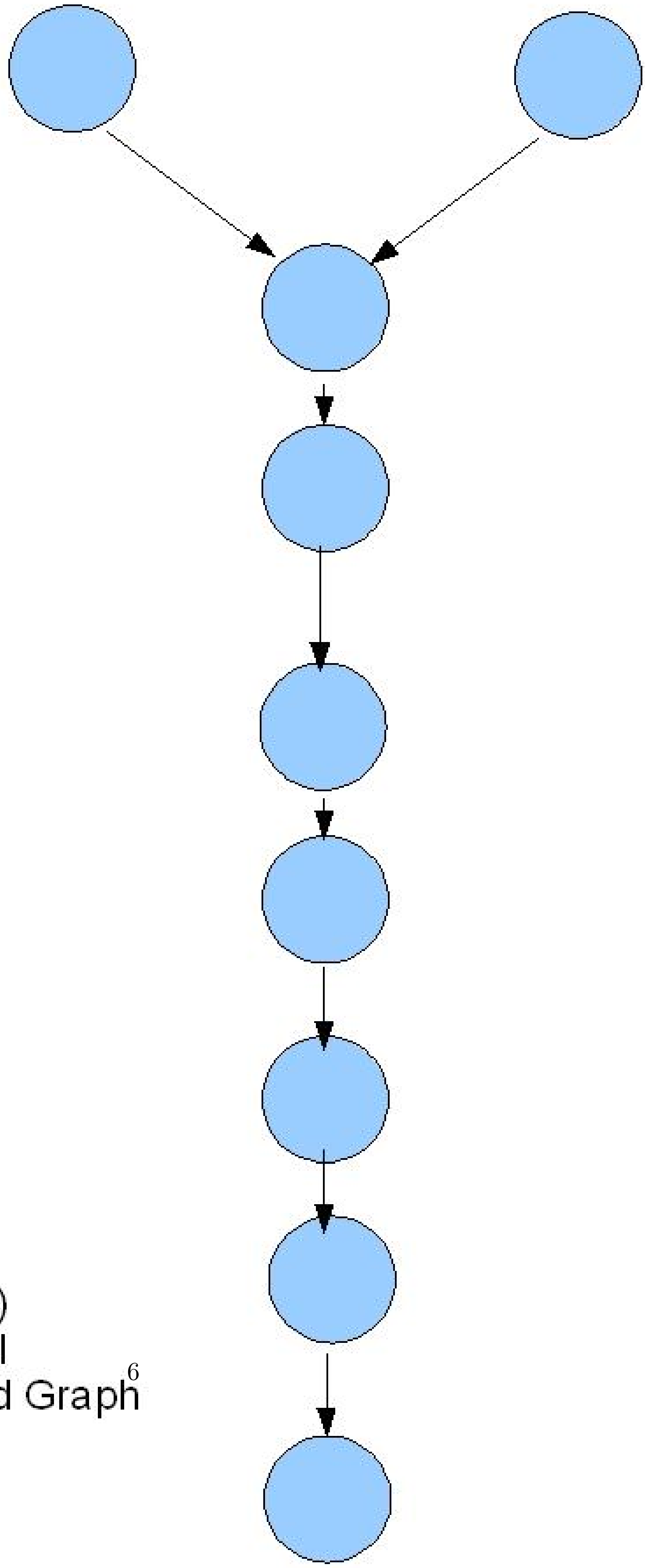
State	y=0	y=1	y=2
x=0	0	0.25	0
x=1	0.25	0	0.25
x=2	0	0.25	0

variables which correspond to the maximum of the joint probability distribution. This helps to calculate the global maximum of the joint probability distribution.

(c) This modification is necessary because we want to calculate the configuration values of the variables that correspond to the global maximum of the joint probability distribution. There could be multiple configuration of the variables that give rise to the maximum value of the probability. It is possible for the individual variable values obtained by maximizing the product of messages at each node to belong to different maximizing configurations giving an overall configuration that no longer corresponds to a maximum.

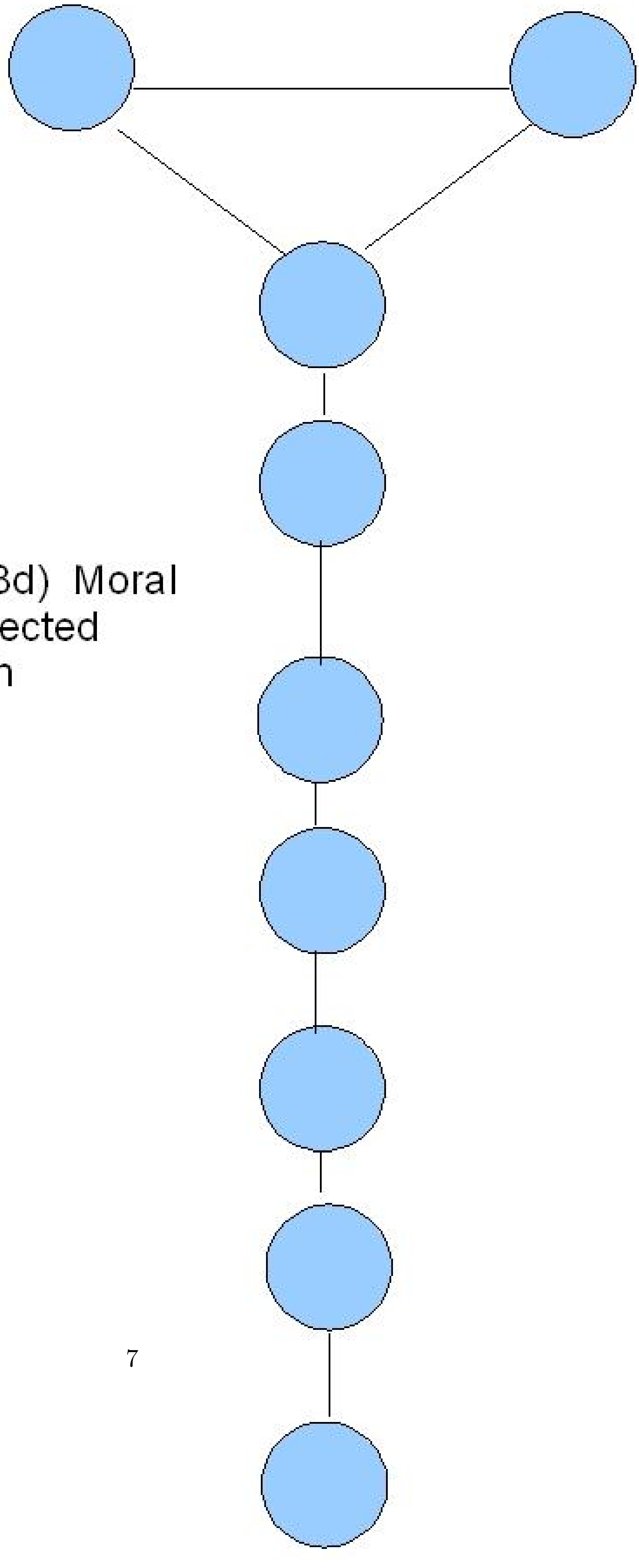


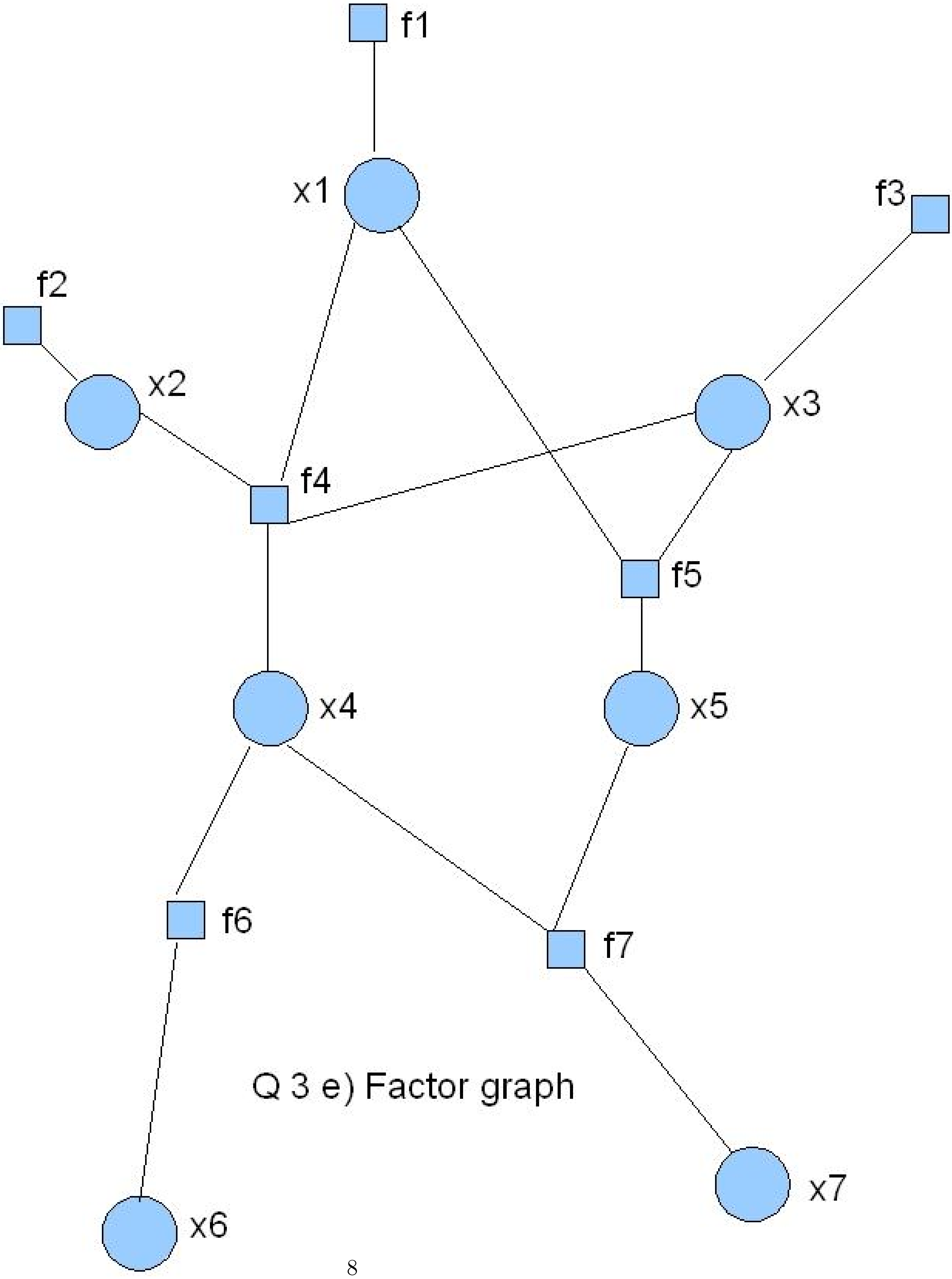
Q 3d : Undirected Graph



Q 3d)
Immoral
Directed Graph

Q 3d) Moral Undirected Graph





Q 3 e) Factor graph