Chapter 20. Exact inference for graphical models

20.1.
a. The entire sequence of variable elimination of the given elimination order goes as follows:

$$\sum_{x_6}\sum_{x_5}\sum_{x_4}\sum_{x_3}\sum_{x_2}[\sum_{x_1}\phi_{12}(x_1,x_2)\phi_{13}(x_1,x_3)]\phi_{24}(x_2,x_4)\phi_{34}(x_3,x_4)\phi_{45}(x_4,x_5)\phi_{56}(x_5,x_6)$$
$$=\sum_{x_6}\sum_{x_5}\sum_{x_4}\sum_{x_3}[\sum_{x_2}\tau_1(x_2,x_3)\phi_{24}(x_2,x_4)]\phi_{34}(x_3,x_4)\phi_{45}(x_4,x_5)\phi_{56}(x_5,x_6)$$
$$=\sum_{x_6}\sum_{x_5}\sum_{x_4}[\sum_{x_3}\tau_2(x_3,x_4)\phi_{34}(x_3,x_4)]\phi_{45}(x_4,x_5)\phi_{56}(x_5,x_6)$$
$$=\sum_{x_6}\sum_{x_5}[\sum_{x_4}\tau_3(x_4)\phi_{45}(x_4,x_5)]\phi_{56}(x_5,x_6)=\sum_{x_6}\sum_{x_5}\tau_4(x_5)\phi_{56}(x_5,x_6)$$
$$=\sum_{x_6}\tau_5(x_6).$$

There are two largest intermediate factors: $\tau_1(x_2,x_3)=\sum_{x_1}\phi_{12}(x_1,x_2)\phi_{13}(x_1,x_3)$ and $\tau_2(x_3,x_4)=\sum_{x_2}\tau_1(x_2,x_3)\phi_{24}(x_2,x_4)$.
b. There is only one fill-in edge $2-3$. The largest maximal cliques are $\{1,2,3\}$ and $\{2,3,4\}$. Each of these corresponds to the largest intermediate factors.
c. The entire sequence of variable elimination of the given elimination order goes as follows:

$$\sum_{x_6}\sum_{x_5}\sum_{x_3}\sum_{x_2}\sum_{x_1}[\sum_{x_4}\phi_{24}(x_2,x_4)\phi_{34}(x_3,x_4)\phi_{45}(x_4,x_5)]\phi_{12}(x_1,x_2)\phi_{13}(x_1,x_3)\phi_{56}(x_5,x_6)$$

$$= \sum_{x_6}\sum_{x_5}\sum_{x_3}\sum_{x_2}[\sum_{x_1}\phi_{12}(x_1,x_2)\phi_{13}(x_1,x_3)]\tau_1(x_2,x_3,x_5)\phi_{56}(x_5,x_6)$$

$$= \sum_{x_6}\sum_{x_5}\sum_{x_3}[\sum_{x_2}\tau_2(x_2,x_3)\tau_1(x_2,x_3,x_5)]\phi_{56}(x_5,x_6)$$

$$= \sum_{x_6}\sum_{x_5}[\sum_{x_3}\tau_3(x_3,x_5)]\phi_{56}(x_5,x_6) = \sum_{x_6}[\sum_{x_5}\tau_4(x_5)\phi_{56}(x_5,x_6)]$$

$$= \sum_{x_6}\tau_5(x_6).$$

There is only one largest intermediate factor:
$\tau_1(x_2,x_3,x_5) = \sum_{x_4}\phi_{24}(x_2,x_4)\phi_{34}(x_3,x_4)\phi_{45}(x_4,x_5)$.
d. There are three fill-in edges: $2-3$, $3-5$, $2-5$. The largest maximal clique is $\{2,3,4,5\}$, which corresponds to the largest intermediate factor.

20.2.

$$\mathcal{N}(x|\mu_1,\lambda_1^{-1})\mathcal{N}(x|\mu_2,\lambda_2^{-1}) = \frac{\sqrt{\lambda_1\lambda_2}}{2\pi}e^{-\frac{\lambda_1}{2}(x-\mu_1)^2-\frac{\lambda_2}{2}(x-\mu_2)^2}$$

$$= \frac{\sqrt{\lambda_1\lambda_2}}{2\pi}e^{-\frac{1}{2}(\lambda_1+\lambda_2)(x-\frac{\lambda_1\mu_1+\lambda_2\mu_2}{\lambda_1+\lambda_2})^2-\frac{1}{2}(\lambda_1\mu_1^2+\lambda_2\mu_2^2-\frac{(\lambda_1\mu_1+\lambda_2+\mu_2)^2}{\lambda_1+\lambda_2})}$$

$$= C\mathcal{N}(x|\mu,\lambda^{-1})$$

where:

$$C = \sqrt{\frac{\lambda}{\lambda_1\lambda_2}}e^{\frac{1}{2}(\lambda_1\mu_1^2(\frac{\lambda_1}{\lambda}-1)+\lambda_2\mu_2^2(\frac{\lambda_2}{\lambda}-1)+\frac{2\lambda_1\lambda_2\mu_1\mu_2}{\lambda})}$$

$$\lambda = \lambda_1+\lambda_2, \mu = \frac{\lambda_1\mu_1+\lambda_2\mu_2}{\lambda}.$$

20.3.
a.

$$p(G_1|X_2=50) = \frac{p(G_1,X_2=50)}{p(X_2=50)} = \frac{p(G_1)\sum_{G_2}p(G_2|G_1)p(X_2=50|G_2)}{\sum_{G_1}\sum_{G_2}p(G_1)p(G_2|G_1)p(X_2=50|G_2)}.$$

$$p(G_1=1)\sum_{G_2}p(G_2|G_1=1)p(X_2=50|G_2) = 0.5\cdot(0.9\cdot\frac{1}{\sqrt{20\pi}}+0.1\cdot\frac{1}{\sqrt{20\pi}}e^{-5}) \simeq 0.05681$$

$$p(G_1=2)\sum_{G_2} p(G_2|G_1=2)p(X_2=50|G_2) = 0.5\cdot(0.1\cdot\frac{1}{\sqrt{20\pi}}+0.9\cdot\frac{1}{\sqrt{20\pi}}e^{-5}) \simeq 0.00669$$

$$\Rightarrow p(G_1|X_2=50) \simeq [\frac{0.05681}{0.05681+0.00669}, \frac{0.00669}{0.05681+0.00669}] \simeq [0.895, 0.105].$$

b.

$$p(G_1|X_2=50, X_3=50) = \frac{p(G_1, X_2=50, X_3=50)}{p(X_2=50, X_3=50)}$$
$$= \frac{p(G_1)\sum_{G_2} p(G_2|G_1)p(X_2=50|G_2)p(X_3=50|G_2)}{\sum_{G_1} p(G_1)p(G_2|G_1)p(X_2=50|G_2)p(X_3=50|G_2)}.$$

$$p(G_1=1)\sum_{G_2} p(G_2|G_1=1)p(X_2=50|G_2)p(X_3=50|G_2)$$
$$= 0.5\cdot(0.9\cdot\frac{1}{20\pi}+0.1\cdot\frac{1}{20\pi}e^{-10}) \simeq 0.00716$$
$$p(G_1=2)\sum_{G_2} p(G_2|G_1=2)p(X_2=50|G_2)p(X_3=50|G_2)$$
$$= 0.5\cdot(0.1\cdot\frac{1}{20\pi}+0.9\cdot\frac{1}{20\pi}e^{-10}) \simeq 0.00080$$

$$\Rightarrow p(G_1|X_2=50, X_3=50) \simeq [\frac{0.00716}{0.00716+0.00080}, \frac{0.00080}{0.00716+0.00080}] \simeq [0.899, 0.101].$$

More observation toward $G_i = 1$ gives more posterior belief of $G_1 = 1$.
c. Similar computation gives

$$p(G_1=1)\sum_{G_2} p(G_2|G_1=1)p(X_2=60|G_2)p(X_3=60|G_2)$$
$$= 0.5\cdot(0.9\cdot\frac{1}{20\pi}e^{-10}+0.1\cdot\frac{1}{20\pi}) \simeq 0.00080$$
$$p(G_1=2)\sum_{G_2} p(G_2|G_1=2)p(X_2=60|G_2)p(X_3=60|G_2)$$
$$= 0.5\cdot(0.1\cdot\frac{1}{20\pi}e^{-10}+0.9\cdot\frac{1}{20\pi}) \simeq 0.00716$$
$$\Rightarrow p(G_1|X_2=60, X_3=60) \simeq [0.101, 0.899].$$

Now observed evidences suggest that $G_i = 2$, the posterior belief leands toward $G_1 = 2$.
d.

$$p(G_1|X_1, X_2 = 50) = p(G_1|X_2 = 50) \simeq [0.895, 0.105],$$

as calculated in part a, because X_1 is unobserved and we marginalized X_3 out.

20.4.
a. We use the variable elimination algorithm in the following elimination order:

$$X_{11} \to X_{21} \to \cdots \to X_{m1} \to X_{21} \to X_{22} \to \cdots X_{m2} \to \cdots X_{mn}.$$

Used the fact that the treewidth of this graphical model is m.
b. It is $O(K^{m+1})$, because the size of the largest intermediate factor is $m+1$.
c. Since the assignments are conditionally independent given observations, and each has known conditional density that depends on only the local evidence, the likelihood function may be rewritten as:

$$p(\mathbf{y}|\mathbf{x}) = \prod_i p(y_i|x_i) = \prod_i p(y_i|x_i = 1)^{x_i} p(y_i|x_i = 0)^{1-x_i}.$$

The prior distribution $p(\mathbf{x})$ can be modelled as a pairwise Markov random field of the form

$$p(\mathbf{x}) \propto e^{\frac{1}{2}\sum_i \sum_j \lambda_{ij}(x_i x_j + (1-x_i)(1-x_j))}$$

where $\lambda_{ij} > 0$ if i and j are neighbors and $\lambda_{ij} = 0$ otherwise.
Thus, if we set $\alpha_i = \ln f(y_i|x_i = 1) - \ln f(y_i|x_i = 0)$, we get

$$\ln p(\mathbf{x}|\mathbf{y}) = \sum_i \alpha_i + \frac{1}{2}\sum_i \sum_j \lambda_{ij}(x_i x_j + (1-x_i)(1-x_j)).$$

The MAP estimate is that $\mathbf{x}$ which maximizes above formula.
Now, consider a network with $n^2 + 2$ vertices, added a source s and a sink t, and the n^2 original points. If $\alpha_i > 0$, set a directed edge $s \to i$ with weight α_i, otherwise, set a directed edge $i \to t$ with weight $-\alpha_i$. For two lattices that are neighbors, set an undirected edge $i \to j$ with weight λ_{ij}.
Now define $A = \{s\} \cup \{i : x_i = 1\}$ and $B = \{t\} \cup \{i : x_i = 0\}$. Then the set of edges between A and B is called a cut of the graph, and we see that maximizing the log posterior probability is equivalent to minimizing $\sum_{a\in A}\sum_{b\in B} w_{ab}$ where w_{ab} is the weight of edge $a - b$. This problem is called

a min-cut problem.
Edmonds-Karp algorithm is guaranteed to solve this min-cut problem in $O(n^4)$ time.