1. Consider a Boltzman machine model parameterized as

$$Pr(x_1, \cdots, x_n) = \frac{1}{Z} \exp(\sum_{(i,j \in E)} w_{i,j} x_i x_j - \sum_{i \in V} u_i, x_i) \tag{1}$$

and an Ising model parameterized as

$$Pr(y_1, \cdots, y_n) = \frac{1}{C} \exp(\sum_{(i,j \in E)} a_{i,j} y_i y_j - \sum_{i \in V} b_i, y)_i \tag{2}$$

We will construct w, u, Z such that $Pr(x_1, \cdots, x_n) = Pr(y_1, \cdots, y_n)$ for $y_i = 2x_i - 1$.

Consider

$$\begin{aligned} g(x_1, \cdots, x_n) = & \frac{1}{C} \exp(\sum_{(i,j) \in E} a_{i,j} y_i y_j - \sum_{i \in V} b_i y_i) \\ = & \frac{1}{C} \exp(\sum_{(i,j) \in E} a_{i,j} (2x_i - 1)(2x_j - 1) - \sum_{i \in V} b_i (2x_i - 1)) \\ = & \frac{1}{C} \exp(\sum_{(i,j) \in E} a_{i,j} (4x_i x_j - 2x_i - 2x_j + 1) - \sum_{i \in V} b_i (2x_i - 1)) \end{aligned} \tag{3}$$

Note that

$$\begin{aligned} \sum_{(i,j) \in E} a_{i,j} (4x_i x_j - 2x_i - 2x_j + 1) = \\ 4 \sum_{(i,j) \in E} a_{i,j} x_i x_j - 2 \sum_{(i,j) \in E} a_{i,j} x_i - 2 \sum_{(i,j) \in E} a_{i,j} x_j + \sum_{(i,j) \in E} a_{i,j} \end{aligned} \tag{4}$$

Since we only consider each edge in one direction when doing our summations, we can write

$$\begin{aligned} \sum_{(i,j) \in E} a_{i,j} x_i + \sum_{(i,j) \in E} a_{i,j} x_j = & \sum_{i \in V} \sum_{j \in N_G(i)} a_{i,j} x_i \\ = & \sum_{i \in V} x_i \sum_{j \in N_G(i)} a_{i,j} \\ = & \sum_{i \in V} \alpha_i x_i \end{aligned} \tag{5}$$

for

$$\alpha_i = \sum_{j \in N_G(i)} a_{i,j} \tag{6}$$

Then

$$\begin{aligned} &\sum_{(i,j)\in E} a_{i,j}(4x_ix_j - 2x_i - 2x_j + 1) - \sum_{i\in V} b_i(2x_i - 1) \\ = 4\sum_{(i,j)\in E} a_{i,j}x_ix_j - 2\sum_{i\in V}\alpha_i x_i - 2\sum_{i\in V} b_i x_i + &\sum_{(i,j)\in E} a_{i,j} + \sum_{i\in V} b_i \\ = \sum_{(i,j\in E)} w_{i,j}x_ix_j - &\sum_{i\in V} u_i x_i + D \end{aligned} \tag{7}$$

for $D = \sum_{(i,j)\in E} a_{i,j} + \sum_{i\in V} b_i$, $w_{i,j} = 4a_{i,j}$ and $u_i = \alpha_i + b_i$.
So

$$\begin{aligned} g(x_1, \cdots, x_n) = &\ \frac{1}{C}\exp(\sum_{(i,j\in E)} w_{i,j}x_ix_j - \sum_{i\in V} u_ix_i + D) \\ = &\ \frac{1}{C\exp(D)}\exp(\sum_{(i,j\in E)} w_{i,j}x_ix_j - \sum_{i\in V} u_ix_i) \\ = &\ \frac{1}{Z}\exp(\sum_{(i,j\in E)} w_{i,j}x_ix_j - \sum_{i\in V} u_ix_i) \\ = &\ Pr(x_1, \cdots, x_n) \end{aligned} \tag{8}$$

for $Z = C\exp(D)$.

2. (a) i. Note that

$$\begin{aligned} P_N(x;\mu, I) = &\ \frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}(x-\mu)^T(x-\mu)) \\ = &\ \frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}\sum_i x_i^2 - 2\mu_ix_i + \mu_i^2) \\ = &\ \frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}x^Tx)\exp(\mu^Tx - \ln(\exp(\frac{1}{2}\mu^T\mu))) \\ = &\ h(x)\exp(\eta f(x) - \ln Z(\eta)) \end{aligned} \tag{9}$$

for $h(x) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}x^Tx)$, $\eta = \mu$, $f(x) = x$ and $Z(\eta) = \exp(\frac{1}{2}\mu^T\mu)$.

ii.

$$\begin{aligned} Pr(x) = &\ \frac{1}{B(\alpha)}\prod_i x_i^{\alpha_i - 1} \\ = &\ \exp((\alpha - 1)^T\ln(x) - \ln B(\alpha)) \\ = &\ h(x)\exp(\eta f(x) - \ln Z(\eta) \end{aligned} \tag{10}$$

for $h(x) = 1$, $\eta = \alpha - 1$, $f(x) = \ln(x)$ and $Z(\eta) = B(\eta + 1)$.

iii.

$$\begin{aligned} Pr(x) = & \frac{1}{x\sigma\sqrt{2\pi}}\exp(\frac{\ln(x)^2}{2\sigma^2}) \\ = & \frac{1}{x\sqrt{2\pi}}\exp(\frac{1}{2\sigma^2}\ln(x)^2 - \ln(\sigma)) \\ = & h(x)\exp(\eta f(x) - \ln Z(\eta)) \end{aligned} \tag{11}$$

for $H(x) = \frac{1}{x\sqrt{2\pi}}$, $\eta = \frac{1}{2\sigma^2}$, $f(x) = \ln(x)^2$ and $V(\eta) = 2\eta^{-\frac{1}{2}} = \sigma$.

iv. Let $\eta = (u, w)$ and $f(x) = (f_v(x), f_e(x))$ where $f_{v_i}(x) = x_i$ and $f_{e_{i,j}} = x_i x_j$. Note that

$$\begin{aligned} Pr(x) \propto & \exp(\sum_i u_i x_i + \sum_{(i,j)\in E} w_{i,j} x_i x_j) \\ = & \exp(\eta^T f(x)) \end{aligned} \tag{12}$$

So $Pr(x) = h(x)\exp(\eta^T f(x) - \ln(Z(\eta)))$ for $h(x) = 1$ and

$$Z(\eta) = \sum_{x\in\{0,1\}^n} \exp(\eta^T f(x)) \tag{13}$$

(b) Let

$$g_i(x) = \begin{cases} 1 & \text{if } i = 0 \\ x_i & \text{otherwise} \end{cases} \tag{14}$$

$$f_i(x, y) = g_i(x)(1 - y) \tag{15}$$

$$\eta = -\alpha \tag{16}$$

Then

$$\begin{aligned} Pr(Y = y|x;\alpha) = & \frac{\exp((1-y)(-\alpha_0 - \sum_{i=1}^n \alpha_i x_i))}{1 + \exp(-\alpha_0 - \sum_{i=1}^n \alpha_i x_i)} \\ = & \frac{\exp(\eta^T f(x,y))}{1 + \eta^T g(x)} \\ = & \exp(\eta^T f(x,y) - \ln(1 + \eta^T g(x)) \\ = & h(x,y)\exp(\eta^T f(x,y) - \ln(Z(\eta, x))) \end{aligned} \tag{17}$$

for $h(x, y) = 1$ and $Z(\eta, x) = 1 + \eta^T g(x)$.

3. Choose $r \in V$ and let T' be the edges of T directed away from r. Now choose $A, B, C \subseteq V$. Suppose that $A \perp B|C$ in the Markov Random Field defined by T. Then every path from a node in A to a node in C contains a node in B.

   All edges in T' are directed away from r. Thus, there are no v-structures in T'. Hence, we can apply the Bayes Ball algorithm to show that B d-separates A and C in T'. So $A \perp B|C$ in the Bayesian Network defined by T'. The Bayesian Network defined by T' has the same conditional independence structure as the MRF defined over T. Thus, any probability distribution p which factorizes over the BN T' also factorizes over the MRF T.

Let p be such a probability distribution. Then

$$\begin{aligned} p(x) = & \quad p(x_r) \prod_{i \to j \in T'} p(x_j|x_i) \\ = & \quad p(x_r) \prod_{i \to j \in T'} \frac{p(x_i, x_j)}{p(x_i)} \\ = & \quad p(x_r) \prod_{i \to j \in T'} \frac{p(x_i, x_j)p(x_j)}{p(x_i)p(x_j)} \end{aligned} \tag{18}$$

By construction, if $k \in V$ and $k \neq r$ then there exists a unique edge $u \to w \in T'$ such that $w = k$. Then

$$p(x_r) \prod_{i \to j \in T'} p(x_j) = p(x_r) \prod_{i \in V \setminus \{r\}} p(x_i) = \prod_{i \in V} p(x_j) \tag{19}$$

and

$$\begin{aligned} p(x) = & \prod_{i \to j \in T'} \frac{p(x_i, x_j)}{p(x_i)p(x_j)} \prod_{j \in V} p(x_j) \\ = & \prod_{(i,j) \in T} \frac{p(x_i, x_j)}{p(x_i)p(x_j)} \prod_{j \in V} p(x_j) \end{aligned} \tag{20}$$

due to the one-to-one correspondence between edges in T and directed edges in T'.

4. (a) Choose i, j and suppose that $(i, j) \notin E$. Consider $p(x_i, x_j|x_k : k \neq i, j)$. Since X_i, X_j obey the conditional independence properties of G and $(i, j) \notin E$ we know that $X_i \perp X_j|\{X_k|k \neq i, j\}$. Thus, by definition,

$$p(x_i, x_j|x_k : k \neq i, j) = p(x_i|x_k : k \neq i, j)p(x_j|x_k : k \neq i, j) \tag{21}$$

Recall that

$$p(x_i, x_j|x_k : k \neq i, j\}) = \frac{p(x_1, \cdots, x_n)}{p(x_k : k \neq i, j)} \tag{22}$$

It follows directly from Murphy Chapter 4.3 that

$$p(x_1, \cdots, x_n) \propto \exp(\sum_{i,j} \Theta_{ij} x_i x_j) \tag{23}$$

and that the marginal

$$p(X_k = x_k : k \neq i, j) \propto \exp(\sum_{i,j \neq k} \Theta_{ij} x_i x_j) \tag{24}$$

Thus,

$$\begin{aligned} p(x_i, x_j | x_k : k \neq i,j) &= \frac{p(x_1, \cdots, x_n)}{p(x_k : k \neq i,j)} \\ &\propto \frac{\exp(\sum_{i,j} \Theta_{ij} x_i x_j)}{\exp(\sum_{i,j \neq k} \Theta_{ij} x_i x_j)} \\ &= \exp(\Theta_{ii} x_i^2 + \Theta_{jj} x_j^2 + \Theta_{ij} x_i x_j + \sum_{k \neq i,j} \Theta_{ik} x_i x_k + \sum_{k \neq i,j} \Theta_{jk} x_j x_k) \end{aligned} \tag{25}$$

Which marginalizing over X_j and X_i respectively gives us

$$p(x_i | x_k : k \neq i,j) \propto \exp(\Theta_{ii} x_i^2 + \sum_{k \neq i,j} \Theta_{ik} x_i x_k) \tag{26}$$

and

$$p(x_j | x_k : k \neq i,j) \propto \exp(\Theta_{jj} x_j^2 + \sum_{k \neq i,j} \Theta_{jk} x_j x_k) \tag{27}$$

Then, since

$$p(x_i, x_j | x_k : k \neq i,j) = p(x_i | x_k : k \neq i,j) p(x_j | x_k : k \neq i,j) \tag{28}$$

we know that

$$\begin{aligned} \exp(\Theta_{ii} x_i^2 + \Theta_{jj} x_j^2 + \Theta_{ij} x_i x_j + \sum_{k \neq i,j} \Theta_{ik} x_i x_k + \sum_{k \neq i,j} \Theta_{jk} x_j x_k) \propto \\ \exp(\Theta_{ii} x_i^2 + \sum_{k \neq i,j} \Theta_{ik} x_i x_k) \exp(\Theta_{jj} x_j^2 + \sum_{k \neq i,j} \Theta_{jk} x_j x_k) = \\ \exp(\Theta_{ii} x_i^2 + \Theta_{jj} x_j^2 + \sum_{k \neq i,j} \Theta_{ik} x_i x_k + \sum_{k \neq i,j} \Theta_{jk} x_j x_k) \end{aligned} \tag{29}$$

Then

$$\exp(\Theta_{ij} x_i x_j) \propto 1 \tag{30}$$

which implies that

$$\Theta_{ij} x_i x_j \propto 0 \tag{31}$$

which holds only if $\Theta_{ij} = 0$.
Now suppose that $\Theta_{ij} = 0$. Then

$$\begin{aligned} p(x_i, x_j | x_k : k \neq i,j) &\propto \exp(\Theta_{ii} x_i^2 + \Theta_{jj} x_j^2 + \sum_{k \neq i,j} \Theta_{ik} x_i x_k + \sum_{k \neq i,j} \Theta_{jk} x_j x_k) \\ &\propto \exp(\Theta_{ii} x_i^2 + \sum_{k \neq i,j} \Theta_{ik} x_i x_k \Theta_{jj} x_j^2 + \sum_{k \neq i,j} \Theta_{jk} x_j x_k) \\ &\propto \exp(\Theta_{ii} x_i^2 + \sum_{k \neq i,j} \Theta_{ik} x_i x_k) \exp(\Theta_{jj} x_j^2 + \sum_{k \neq i,j} \Theta_{jk} x_j x_k) \\ &\propto p(x_i | x_k : k \neq i,j) p(x_j | x_k : k \neq i,j) \end{aligned} \tag{32}$$

Thus,

$$p(x_i, x_j | x_k : k \neq i, j) = p(x_i | x_k : k \neq i, j) p(x_j | x_k : k \neq i, j) \tag{33}$$

and so X_i and X_j are conditionally independent. Therefore $(i, j) \notin E$.

(b) From the result above we can conclude that $\Theta_{ij} = 0$ if and only if there is a cut set that separates i and j. Equivalently, $\Theta_{ij} = 0$ if and only if X_i and X_j are conditionally independent.

5. To show that r is a valid probability distribution we need to show that

$$\sum_{x_1, \cdots, x_n} r(x_1, \cdots, x_n) = 1 \tag{34}$$

Let $m_i = 1 - |N_G(i)|$. Then

$$\begin{aligned} r(x_1, \cdots, x_n) = & \prod_{i=1}^{n} [\mu_i(x_i)]^{m_i} \prod_{(i,j) \in E} \mu_{ij}(x_i, j) \\ = & \prod_{i=1}^{n} \mu_i(x_i) \prod_{(i,j) \in E} \frac{\mu_{ij}(x_i, x_j)}{\mu_i(x_i) \mu_j(x_j)} \end{aligned} \tag{35}$$

Choose $k \leq n$. Define $G_k = (V_k, E_k)$ with $V_k = \{v \in V | v \leq k\}$ and $E_k = \{(i, j) \in E | i, j \leq k$ for any $k \leq n$. Without loss of generality, we can label G such that (i) k is a leaf node of G_k for all $2 \leq k \leq n$ and (ii) G_k is connected. For all such G, consider

$$r_k(x_1, \cdots, x_k) = \prod_{i=1}^{k} \mu_i(x_i) \prod_{(i,j) \in E_k} \frac{\mu_{ij}(x_i, x_j)}{\mu_i(x_i) \mu_j(x_j)} \tag{36}$$

for all $k \leq n$. We will show by induction that

$$\sum_{x_1, \cdots, x_k} r(x_1, \cdots, x_k) = 1 \tag{37}$$

for all $2 \leq k \leq n$ demonstrating that

$$\sum_{x_1, \cdots, x_n} r(x_1, \cdots, x_n) = 1 \tag{38}$$

First consider the base case $k = 2$. Since G_2 is connected, $(1, 2) \in E_2$. So

$$r(x_1, x_2) = \mu_{12}(x_1, x_2) \tag{39}$$

Hence

$$\begin{aligned}\sum_{x_1,x_2} r(x_1,x_2) = & \sum_{x_1,x_2} \mu_1(x_1,x_2) \\ = & \sum_{x_1}\sum_{x_2} \mu_1(x_1,x_2) \\ = & \sum_{x_1} \mu_1(x_1) \\ = & 1\end{aligned} \tag{40}$$

thus proving the base case.

Now choose $2 < k \leq n$. Suppose that $\sum_{x_1,\cdots,x_k} r(x_1,\cdots,x_k) = 1$. To prove the induction step, we will show that

$$\sum_{x_1,\cdots,x_{k+1}} r(x_1,\cdots,x_{k+1}) = 1 \tag{41}$$

Consider

$$r_{k+1}(x_1,\cdots,x_{k+1}) = \prod_{i=1}^{k+1} \mu_i(x_i) \prod_{(i,j)\in E_{k+1}} \frac{\mu_{ij}(x_i,x_j)}{\mu_i(x_i)\mu_j(x_j)} \tag{42}$$

Note that

$$\prod_{i=1}^{k+1} \mu_i(x_i) = \mu_{k+1}(x_{k+1}) \prod_{i=1}^{k} \mu_i(x_i) \tag{43}$$

and

$$\prod_{(i,j)\in E_{k+1}} \frac{\mu_{ij}(x_i,x_j)}{\mu_i(x_i)\mu_j(x_j)} = \prod_{(i,j)\in E_k} \frac{\mu_{ij}(x_i,x_j)}{\mu_i(x_i)\mu_j(x_j)} \prod_{i\in N_{G_{k+1}}(k+1)} \frac{\mu_{(k+1)i}(x_{k+1},x_i)}{\mu_{k+1}(x_{k+1})\mu_i(x_i)} \tag{44}$$

Since $k+1$ is a leaf node in G_{k+1}, $N_{G_{k+1}}(k+1) = \{v\}$ for some $1 \leq v \leq k$. Thus,

$$\prod_{i\in N_{G_{k+1}}(k+1)} \frac{\mu_{(k+1)i}(x_{k+1},x_i)}{\mu_{k+1}(x_{k+1})\mu_i(x_i)} = \frac{\mu_{(k+1)v}(x_{k+1},x_v)}{\mu_{k+1}(x_{k+1})\mu_v(x_v)} \tag{45}$$

So

$$\begin{aligned} r_{k+1}(x_1,\cdots,x_{k+1}) = & \; r_k(x_1,\cdots,x_k)\mu_{k+1}(x_{k+1}) \frac{\mu_{(k+1)v}(x_{k+1},x_v)}{\mu_{k+1}(x_{k+1})\mu_v(x_v)} \\ = & \; r_k(x_1,\cdots,x_k) \frac{\mu_{(k+1)v}(x_{k+1},x_v)}{\mu_v(x_v)} \end{aligned} \tag{46}$$

Thus,

$$\begin{aligned}
\sum_{x_1,\cdots,x_{k+1}} r_{k+1}(x_1,\cdots,x_{k+1}) &= \sum_{x_1,\cdots,x_{k+1}} r_k(x_1,\cdots,x_k)\frac{\mu_{(k+1)v}(x_{k+1},x_v)}{\mu_v(x_v)} \\
&= \sum_{x_1,\cdots,x_k} r_k(x_1,\cdots,x_k)\sum_{x_{k+1}}\frac{\mu_{(k+1)v}(x_{k+1},x_v)}{\mu_v(x_v)} \\
&= \sum_{x_1,\cdots,x_k} r_k(x_1,\cdots,x_k) \\
&= 1
\end{aligned} \tag{47}$$

Thus, by induction,

$$\sum_{x_1,\cdots,x_k} r(x_1,\cdots,x_k) = 1 \tag{48}$$

for all $2 \le k \le n$. So

$$\sum_{x_1,\cdots,x_n} r(x_1,\cdots,x_n) = 1 \tag{49}$$

and therefore r is a valid probability distribution for $m_i = |N_G(i)| - 1$.