A Proofs

Proof of Theorem 1. Consider a node X with parents $\mathbf{U}_X$. We first show the following, where Equation (1) below follows since the state-abstraction is harmless.

$$
\begin{aligned}
& Pr(x|\bigvee_{\mathbf{u}_X \in \mathbf{u}^{\mathsf{a}}_X} \mathbf{u}_X) \\
= \; & \frac{Pr(x \wedge \bigvee_{\mathbf{u}_X \in \mathbf{u}^{\mathsf{a}}_X} \mathbf{u}_X)}{Pr(\bigvee_{\mathbf{u}_X \in \mathbf{u}^{\mathsf{a}}_X} \mathbf{u}_X)} \\
= \; & \frac{\sum_{\mathbf{u}_X \in \mathbf{u}^{\mathsf{a}}_X} Pr(x, \mathbf{u}_X)}{\sum_{\mathbf{u}_X \in \mathbf{u}^{\mathsf{a}}_X} Pr(\mathbf{u}_X)} \\
= \; & \frac{\sum_{\mathbf{u}_X \in \mathbf{u}^{\mathsf{a}}_X} Pr(x|\mathbf{u}_X) Pr(\mathbf{u}_X)}{\sum_{\mathbf{u}_X \in \mathbf{u}^{\mathsf{a}}_X} Pr(\mathbf{u}_X)} \\
= \; & \frac{Pr(x|\mathbf{u}^{\star}_X) \sum_{\mathbf{u}_X \in \mathbf{u}^{\mathsf{a}}_X} Pr(\mathbf{u}_X)}{\sum_{\mathbf{u}_X \in \mathbf{u}^{\mathsf{a}}_X} Pr(\mathbf{u}_X)} \text{ for any elementary state } \mathbf{u}^{\star}_X \in \mathbf{u}^{\mathsf{a}}_X \qquad (1) \\
= \; & Pr(x|\mathbf{u}^{\star}_X)
\end{aligned}
$$

To simplify notation, we will define $Pr(x|\mathbf{u}^{\mathsf{a}}_X) \triangleq Pr(x|\bigvee_{\mathbf{u}_X \in \mathbf{u}^{\mathsf{a}}_X} \mathbf{u}_X)$. We now have:

$$
Pr(x|\mathbf{u}^{\mathsf{a}}_X) = Pr(x|\mathbf{u}_X) \text{ for any elementary state } \mathbf{u}_X \text{ in superstate } \mathbf{u}^{\mathsf{a}}_X. \qquad (2)
$$

We define the CPT of a node X with parents $\mathbf{U}_X$ in the abstracted BN $(G^{\mathsf{a}}, \Theta^{\mathsf{a}})$ as follows:

$$
\theta^{\mathsf{a}}(x^{\mathsf{a}}|\mathbf{u}^{\mathsf{a}}_X) \triangleq \sum_{x \in x^{\mathsf{a}}} Pr(x|\mathbf{u}^{\mathsf{a}}_X). \qquad (3)
$$

We now show that distributions $Pr(\mathbf{Z})$ and $Pr^{\mathsf{a}}(\mathbf{Z})$ are consistent, where $\mathbf{Z}$ are all variables:

$$
\begin{aligned}
\sum_{\mathbf{z} \in \mathbf{z}^{\mathsf{a}}} Pr(\mathbf{z}) &= \sum_{\mathbf{z} \in \mathbf{z}^{\mathsf{a}}} \prod_{x, \mathbf{u}_X \in \mathbf{z}} Pr(x|\mathbf{u}_X) \\
&= \sum_{\mathbf{z} \in \mathbf{z}^{\mathsf{a}}} \prod_{x \in \mathbf{z}} Pr(x|\mathbf{u}^{\mathsf{a}}_X) \text{ by (2) where } \mathbf{u}^{\mathsf{a}}_X \in \mathbf{z}^{\mathsf{a}} \\
&= \prod_{x^{\mathsf{a}}, \mathbf{u}^{\mathsf{a}}_X \in \mathbf{z}^{\mathsf{a}}} \sum_{x \in x^{\mathsf{a}}} Pr(x|\mathbf{u}^{\mathsf{a}}_X) \text{ see justification below} \qquad (4) \\
&= \prod_{x^{\mathsf{a}}, \mathbf{u}^{\mathsf{a}}_X \in \mathbf{z}^{\mathsf{a}}} \theta^{\mathsf{a}}(x^{\mathsf{a}}|\mathbf{u}^{\mathsf{a}}_X), \text{ by (3)} \\
&= Pr^{\mathsf{a}}(\mathbf{z}^{\mathsf{a}})
\end{aligned}
$$

Equation (4) is justified as follows. The state $\mathbf{u}^{\mathsf{a}}_X \in \mathbf{z}^{\mathsf{a}}$ is determined by variable X and instantiation $\mathbf{z}^{\mathsf{a}}$ so it does not depend on instantiation $\mathbf{z}$. Hence, $Pr(x|\mathbf{u}^{\mathsf{a}}_X)$ depends only on the state of variable X in $\mathbf{z}$ so factors $Pr(X|\mathbf{u}^{\mathsf{a}}_X)$ are over disjoint variables. By properties of variable elimination, we can commute addition and multiplication to get Equation (4). $\square$

Lemma 1. *A high-order edge $\mathcal{E}_1 : U_1 \rightarrow X_1$ in G remains a high-order edge after another high-order edge $\mathcal{E}_2 : U_2 \rightarrow X_2$ is harmlessly deleted from G.*

Proof. Suppose that $\mathcal{P}_1$ is the alternative directed path of $\mathcal{E}_1$ in G, and $\mathcal{P}_2$ is the alternative directed path of $\mathcal{E}_2$ in G. If edge $\mathcal{E}_2$ is not on path $\mathcal{P}_1$, then trivially $\mathcal{P}_1$ remains a directed path from U_1 to X_1 after $\mathcal{E}_2$ is deleted. If $\mathcal{E}_2$ is indeed on path $\mathcal{P}_1$, then after deleting $\mathcal{E}_2$, we can always construct another directed path from U_1 to X_1 other than $\mathcal{P}_1$ by going through path $\mathcal{P}_2$ instead of edge $\mathcal{E}_2$. Hence, $\mathcal{E}_1$ remains a high-order edge after $\mathcal{E}_2$ is deleted. $\square$

Lemma 2. *Let Pr^1, Pr^2, and Pr^3 be three distributions over variables $\mathbf{V}$ where Pr^2 abstracts the state-space of Pr^1 and Pr^3 further abstracts the state-space of Pr^2. If Pr^1 and Pr^2 are consistent, Pr^2 and Pr^3 are consistent, then Pr^1 and Pr^3 are consistent.*

Proof. Let $\mathbf{v}^1, \mathbf{v}^2$ and $\mathbf{v}^3$ denote the instantiations of $\mathbf{V}$ in Pr^1, Pr^2 and Pr^3. Then:

$$\begin{aligned}
\sum_{\mathbf{v}^1 \in \mathbf{v}^3} Pr^1(\mathbf{v}^1) &= \sum_{\mathbf{v}^2 \in \mathbf{v}^3} \sum_{\mathbf{v}^1 \in \mathbf{v}^2} Pr^1(\mathbf{v}^1) \\
&= \sum_{\mathbf{v}^2 \in \mathbf{v}^3} Pr^2(\mathbf{v}^2) \text{ since } Pr^1 \text{ and } Pr^2 \text{ are consistent} \\
&= Pr^3(\mathbf{v}^3) \text{ since } Pr^2 \text{ and } Pr^3 \text{ are consistent}
\end{aligned}$$

Hence, Pr^1 and Pr^3 are consistent. $\square$

Proof of Theorem 2. Given Lemmas 1 and 2, we will prove the theorem while assuming that only a single high-order edge is harmlessly deleted (we can repeat the proof to account for multiple deleted edges). Suppose BN (G, Θ) is obtained by harmlessly deleting high-order edge $U \rightarrow X$ from BN (G', Θ'). Let $U \rightarrow Z_1 \rightarrow \cdots \rightarrow Z_n \rightarrow X$ be the alternative directed path from U to X that is selected when implementing harmless edge deletion. DAG G' can be viewed as the result of adding edge $U \rightarrow X$ to G and abstracting nodes Z_i. We will use z_i to denote the elementary states of a node Z_i in G and z'_i to denote its abstracted states in G'. By definition of harmless edge deletion, nodes Z_i have compound states $z_i = (z'_i, u)$ in G, where u is a state of node U. Moreover, $sup(z_i = (z'_i, u)) = z'_i$.

To show that G' is reducible to G, we will show that for every parametrization Θ' of G', there exists a parametrization Θ of G such that distributions Pr' and Pr are consistent. We first introduce some notation, then show how to construct parameterization Θ based on parameterization Θ' and finally prove that distributions Pr' and Pr are consistent.

- $\mathbf{W}$ denotes all nodes, partitioned X, $\mathbf{Z} = \{Z_1, \ldots, Z_n\}$ and $\mathbf{Y}$ (hence $U \in \mathbf{Y}$).
- Since node X and nodes $Y \in \mathbf{Y}$ are not abstracted, we have $sup(x) = x$ and $sup(y) = y$ so we use x and y to denote both their elementary and superstates.

- For an elementary state (instantiation) $\mathbf{v}$ of nodes $\mathbf{V} \subseteq \mathbf{W}$ and state u of node U, we say that $\mathbf{v}$ is *congruent* with u iff no state of U other than u appears in $\mathbf{v}$.
- For an elementary state $\mathbf{v}$ of nodes $\mathbf{V} \subseteq \mathbf{W}$, we use $\mathbf{v}^{\uparrow}$ to denote the superstate which results from replacing every compound state (z'_i, u) in $\mathbf{v}$ with z'_i.
- We use $\sim$ to indicate compatibility between instantiations (agreement on the states of common variables).

We next show how to construct parametrization Θ based on parameterization Θ'. We have three cases for nodes X, $Z_i \in \mathbf{Z}$ and $Y \in \mathbf{Y}$ which have parents $\mathbf{V}_X$, $\mathbf{V}_i$ and $\mathbf{V}_Y$ in DAG G. Note that U is a parent of X in G' but not in G as we deleted edge $U \to X$ so $U \notin \mathbf{V}_X$.

$$\theta(z_i|\mathbf{v}_i) \triangleq \begin{cases} \theta'(z'_i|\mathbf{v}_i^{\uparrow}) & \text{if } z_i = (z'_i, u) \text{ and } \mathbf{v}_i \text{ is congruent with } u \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

$$\theta(x|\mathbf{v}_X) \triangleq \begin{cases} \theta'(x|\mathbf{v}_X^{\uparrow}, u) & \text{if } \mathbf{v}_X \text{ is congruent with } u \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

$$\theta(y|\mathbf{v}_Y) \triangleq \theta'(y|\mathbf{v}_Y^{\uparrow}) \tag{7}$$

To show that distributions Pr' and Pr are consistent, we first need to show the following: If u is the state of node U in instantiation $\mathbf{w}$ and $Pr(\mathbf{w}) \neq 0$, then $\mathbf{w}$ must be congruent with u (that is, u will appear in every compound state z_i in $\mathbf{w}$ for $i = 1, \ldots, n$). Suppose u is the state of node U in instantiation $\mathbf{w}$ and $Pr(\mathbf{w}) \neq 0$. Since $Pr(\mathbf{w}) \neq 0$, we have $\theta(z_i|\mathbf{v}_i) \neq 0$ for every $z_i, \mathbf{v}_i \sim \mathbf{w}$; by Equation 5, $\mathbf{v}_i$ must be congruent with some state of node U and that state must appear in z_i. We prove the above result by induction. For $i = 1$ (base case), we have $U \in \mathbf{V}_1$, so $\mathbf{v}_1$ must be congruent with u and u must appear in z_1. Assume u appears in z_{i-1} for $i \geq 2$ (induction hypthesis). Since $Z_{i-1} \in \mathbf{V}_i$, $\mathbf{v}_i$ must be congruent with u and u must appear in z_i. Hence, u appears in all compound states of $\mathbf{w}$.

In the next derivation, we will sum only over elementary states $x\mathbf{yz} \in \mathbf{w}'$ that are congruent with u, where u is the state of node U in superstate $\mathbf{w}'$ (otherwise, $Pr(x, \mathbf{y}, \mathbf{z}) = 0$). We also have $\mathbf{w}' = x\mathbf{yz}^{\uparrow}$ since X and $\mathbf{Y}$ are not abstracted.

$$\begin{aligned} \sum_{x\mathbf{yz} \in \mathbf{w}'} Pr(x, \mathbf{y}, \mathbf{z}) &= \sum_{x\mathbf{yz} \in \mathbf{w}'} \prod_{\mathbf{v}_X \sim x\mathbf{yz}} \theta(x|\mathbf{v}_X) \prod_{z_i \mathbf{v}_i \sim x\mathbf{yz}} \theta(z_i|\mathbf{v}_i) \prod_{y\mathbf{v}_Y \sim x\mathbf{yz}} \theta(y|\mathbf{v}_Y) \\ &= \sum_{x\mathbf{yz} \in \mathbf{w}'} \prod_{\mathbf{v}_X \sim x\mathbf{yz}} \theta'(x|\mathbf{v}_X^{\uparrow}, u) \prod_{z_i \mathbf{v}_i \sim x\mathbf{yz}, z_i = (z'_i, u)} \theta'(z'_i|\mathbf{v}_i^{\uparrow}) \prod_{y\mathbf{v}_Y \sim x\mathbf{yz}} \theta'(y|\mathbf{v}_Y^{\uparrow}) \\ &= \prod_{\mathbf{v}'_X \sim \mathbf{w}'} \theta'(x|\mathbf{v}'_X, u) \prod_{z'_i \mathbf{v}'_i \sim \mathbf{w}'} \theta'(z'_i|\mathbf{v}'_i) \prod_{y\mathbf{v}'_Y \sim \mathbf{w}'} \theta'(y|\mathbf{v}_Y) \\ &= Pr'(\mathbf{w}') \end{aligned}$$

Hence, distributions Pr' and Pr are consistent. □

Proof of Theorem 3. To prove this theorem, we show that if we use the belief propagation algorithm to compute $Pr^{\mathsf{a}}(Q, \mathbf{e})$ on the abstracted polytree described by the theorem, then we get $Pr^{\mathsf{a}}(Q, \mathbf{e}) = Pr(Q, \mathbf{e})$. We show this by re-deriving the belief propagation algorithm in Section B while using the modified CPTs and assuming that we can only access superstates of abstracted nodes but not their elementary states. The derivation shows that the meaning of some messages changes yet the computed marginal $Pr^{\mathsf{a}}(Q, \mathbf{e})$ equals $Pr(Q, \mathbf{e})$. The reason why some messages attain a new meaning is that we loose some independences due to abstracting nodes. For example, in a polytree, the children of a node X are independent given X; that is, given an elementary state of X. However, these children are not independent given a superstate of X. As a result, some simplifications are no longer possible which changes the meaning of some messages but maintains the final result. $\square$

B Belief Propagation on Polytrees with Abstracted Nodes

Given a polytree BN (G, Θ) with distribution Pr and a state-space abstraction a, let $(G^{\mathsf{a}}, \Theta^{\mathsf{a}})$ be the abstracted polytree BN described by Theorem 3. We will next show that if we run the belief propagation on BN $(G^{\mathsf{a}}, \Theta^{\mathsf{a}})$ to compute the marginal $Pr^{\mathsf{a}}(Q, \mathbf{e})$, then the result we get equals $Pr(Q, \mathbf{e})$.

As stated by Theorem 3, we have the following:

1. Query node Q and evidence nodes $\mathbf{E}$ are not abstracted.
2. Abstracted nodes must be ancestors of query node Q.
3. We have an ordering π of polytree nodes which places Q-children after their siblings.
4. For a node X with parents $\mathbf{U}$ that are not abstracted, $\theta^{\mathsf{a}}(x^{\mathsf{a}}|\mathbf{u}^{\mathsf{a}}) = Pr(x^{\mathsf{a}}|\mathbf{u}^{\mathsf{a}})$.
5. Consider an abstracted node U with children $X_1, \cdots, X_k$ that respect ordering π, and let $\mathbf{U}_i$ be the other parents of child X_i. The abstracted CPT for X_i is as follows:

$$\theta^{\mathsf{a}}(x_i^{\mathsf{a}} \mid u^{\mathsf{a}}, \mathbf{u}_i^{\mathsf{a}}) = Pr(x_i^{\mathsf{a}} \mid u^{\mathsf{a}}, \mathbf{u}_i^{\mathsf{a}}, \mathbf{e}_{X_i}^{+} \setminus \{\mathbf{e}_{UX_j}^{-}\}_{j>i})$$

To simplify notation, we will drop the superscript a while keeping track of which nodes are abstracted. We will also use $\theta_{\mathbf{e}}^{\pi}$ instead of θ^{a} to highlight that the CPTs we are using in the abstracted polytree depend on evidence $\mathbf{e}$ and ordering π.

The belief propagation algorithm can be used to compute marginals over all nodes. This is done by passing two messages across each edge $U \rightarrow X$, a message $\pi_X(u)$ from parent U to child X and a message $\lambda_X(u)$ from child X to parent U. However, we are only interested in the marginal for query node Q which entails sending only one of the two messages that is directed towards node Q. Moreover, if node U is abstracted, then u is an abstracted state and hence messages $\pi_X(u)$ and $\lambda_X(u)$ are over the abstracted states.

B.1 Notation and Observations

We first settle some notation and state some observations that will be used in the proof.

We use capital letters (e.g., X) to denote variables (nodes) and small letters (e.g., x) to denote states of variables. When node X is abstracted, then x denotes a superstate (the algorithm will not access elmentary states of abstracted nodes). As customary, we assume that evidence on a node X is modeled using a message $\lambda_{\mathbf{e}}(x)$ sent from a virtual child of X, so only (virtual) leaf nodes can have evidence. This is defined as $\lambda_{\mathbf{e}}(x) = 1$ if state x is compatible with evidence $\mathbf{e}$ and $\lambda_{\mathbf{e}}(x) = 0$ otherwise.

We use the classical notation for partitioning polytree evidence $\mathbf{e}$ based on edge $U \rightarrow X$.

$\mathbf{e}^-_X$: evidence on the descendants of X
$\mathbf{e}^+_X$: evidence on the non-descendants of X, $\mathbf{e} \setminus \mathbf{e}^-_X$
$\mathbf{e}^+_{UX}$: evidence on the U-side of edge $U \rightarrow X$
$\mathbf{e}^-_{UX}$: evidence on the X-side of edge $U \rightarrow X$

The following are some implications of (a) abstracted nodes must be ancestors of query node Q and (b) all messages are directed towards query node Q. We will use these implications in the proof of correctness for the algorithm.

- If node X is an ancestor of query node Q, then exactly one child of X will be either Q or an ancestor of Q. This is what we called the Q-child of node X.
- An abstracted node X will never send a λ-message.
- The one message sent by an abstracted node will be a π-message to its Q-child.
- If a node has two abstracted parents, it must be the Q-child of both.
- If node X sends a λ-message towards its parent U, then no other parent of X can be abstracted.
- A node that sends a π-message and has an abstracted parent must be the Q-child of that parent.

B.2 The Belief Propagation Algorithm

The belief propagation algorithm is specified by the following equations, which show how messages are computed and then used to compute $BEL(Q, \mathbf{e})$, the marginal for node Q.

These equations assume that node X has parents $U_1, \cdots, U_m$ and children $Y_1, \cdots, Y_n$.

$$\begin{aligned}
BEL(q, \mathbf{e}) &= \pi(q)\lambda(q) && (8)\\
\pi(x) &= \sum_{u_1,\cdots,u_m} \theta(x \mid u_1, \cdots, u_m) \prod_{i=1}^{m} \pi_X(u_i) && (9)\\
\lambda(x) &= \lambda_{\mathbf{e}}(x) \prod_{i=1}^{n} \lambda_{Y_i}(x) && (10)\\
\pi_{Y_i}(x) &= \pi(x) \prod_{j \neq i} \lambda_{Y_j}(x) && (11)\\
\lambda_X(u_i) &= \sum_{x,u_1,\cdots,u_{i-1},u_{i+1},\ldots,u_m} \theta(x \mid u_1, \cdots, u_m)\lambda(x) \prod_{j \neq i} \pi_X(u_j) && (12)
\end{aligned}$$

When applying this algorithm to the abstracted polytree, we will use abstracted CPTs $\theta^{\pi}_{\mathbf{e}}$ as indicated earlier. Moreover, if node X is abstracted, then x is a superstate. Similarly, if a parent U_i is abstracted, then u_i is a superstate.

B.3 Proof Structure

Our proof is based on showing the following. Let Pr be the distribution induced by the original polytree BN, which is used to compute the abstracted CPTs $\theta^{\pi}_{\mathbf{e}}$. When applying the belief propagation algorithm to the abstracted polytree, Equations 9-12 compute the following quantities:

$$\begin{aligned}
\pi(x) &= Pr(x, \mathbf{e}^+_X) && (13)\\
\lambda(x) &= Pr(\mathbf{e}^-_X \mid x) && (14)\\
\pi_X(u_i) &= Pr(u_i, \mathbf{e}^+_{U_i X}) && (15)\\
\lambda_{Y_i}(x) &= Pr(\mathbf{e}^-_{XY_i} \mid x) \text{ when } X \text{ is not abstracted} && (16)\\
\lambda_{Y_i}(x) &= Pr(\mathbf{e}^-_{XY_i} \mid x, \mathbf{e}^+_X, \mathbf{e}^-_{XY_1}, \cdots, \mathbf{e}^-_{XY_{i-1}}) \text{ when } X \text{ is abstracted} && (17)\\
& \quad \text{where } Y_1, \cdots, Y_i \text{ respects the ordering } \pi.
\end{aligned}$$

Given these equalities, Equation 8 establishes the desired result.

$$\begin{aligned}
BEL(q, \mathbf{e}) &= \pi(q)\lambda(q)\\
&= Pr(q, \mathbf{e}^+_Q) Pr(\mathbf{e}^-_Q \mid q)\\
&= Pr(q, \mathbf{e}^+_Q) Pr(\mathbf{e}^-_Q \mid q, \mathbf{e}^+_Q)\\
&= Pr(q, \mathbf{e}^+_Q, \mathbf{e}^-_Q)\\
&= Pr(q, \mathbf{e})
\end{aligned}$$

That is, the marginal for query node Q as computed by the belief propagation algorithm on the abstracted polytree BN is nothing but $Pr(q, \mathbf{e})$ (the marginal it would have computed if applied to the original polytree BN).

Equations 13-16 are known to hold for the belief propagation algorithm. However, Equation 16 does not hold if node X is abstracted and is replaced by Equation 17. To see why, suppose that node X is not abstracted. Then $\mathbf{E}^-_{XY_i}$ is independent of $\mathbf{E}^-_{XY_j}$ given X for $i \neq j$. Moreover, $\mathbf{E}^-_{XY_i}$ is independent of $\mathbf{E}^+_X$ given X for all i. Therefore, $Pr(\mathbf{e}^-_{XY_i} \mid x, \mathbf{e}^+_X, \mathbf{e}^-_{XY_1}, \cdots, \mathbf{e}^-_{XY_{i-1}}) = Pr(x, \mathbf{e}^+_{U_iX})$ so Equation 17 reduces to Equation 16. These independences do not hold though when X is abstracted since conditioning on a superstate of X does not fix the elementary state of X. The loss of such independence when conditioning on superstates is the main subtlety of the upcoming proofs as it requires one to keep track of conditionings on abstracted nodes. Aside from this, the derivations below are mostly standard derivations used when justifying the standard belief propagation algorithm.

The proof is by induction on the structure of the polytree. The base cases are nodes with a single neighbor: root nodes with a single child and leaf nodes with a single parent.

B.4 Base Cases

We first note that the query node Q does not send messages.

Suppose $X \neq Q$ is a root node that has a single child Y. Then X will send a message $\pi_Y(x)$ to its child Y. Moreover, $\pi_Y(x) = \pi(x) = Pr(x) = Pr(x, \mathbf{e}^+_X)$ since $\mathbf{e}^+_X$ is empty.

Suppose now that $X \neq Q$ is a leaf node that has a single parent U. Then X will send a message $\lambda_X(u)$ to its parent U. Hence, U cannot be abstracted since query node Q must be on the U-side of edge $U \to X$ so U cannot be an ancestor of Q. Hence, $\lambda(x) = \lambda_{\mathbf{e}}(x)$ and $\lambda_X(u) = \sum_x \lambda_{\mathbf{e}}(x) Pr(x \mid u, \mathbf{e}^+_X)$. Since $Pr(x \mid u, \mathbf{e}^+_X) = Pr(x \mid u)$ we have $\lambda_X(u) = \sum_x \lambda_{\mathbf{e}}(x) Pr(x \mid u) = Pr(\mathbf{e}^-_{UX} \mid u)$ since $\mathbf{e}^-_{UX}$ is the evidence on node X.

We will next prove Equations 13-17 while assuming that the messages received by a node have the correct meanings.

B.5 π-Support and λ-Support: Proving Equations 13 and 14

The support $\pi(x)$ is computed only if $X = Q$ or X is sending a π-message to its Q-child.

If X has no abstracted parents, then it is independent of its non-descendants given its parents and we have $\theta^\pi_{\mathbf{e}}(x \mid u_1, \cdots, u_m) = Pr(x \mid u_1, \cdots, u_m) = Pr(x \mid u_1, \cdots, u_m, \mathbf{e}^+_X)$. If X has some abstracted parents, then it is their Q-child so it is ordered last by π among its siblings and we have $\theta^{\pi}_{\mathbf{e}}(x \mid u_1, \cdots, u_m) = Pr(x \mid u_1, \cdots, u_m, \mathbf{e}^+_X)$.

$$\begin{aligned}
\pi(x) &= \sum_{u_1,\cdots,u_m} \theta^{\pi}_{\mathbf{e}}(x \mid u_1, \cdots, u_m) \prod_{i=1}^{m} \pi_X(u_i) \\
&= \sum_{u_1,\cdots,u_m} \theta^{\pi}_{\mathbf{e}}(x \mid u_1, \cdots, u_m) \prod_{i=1}^{m} Pr(u_i, \mathbf{e}^+_{U_iX}) \\
&= \sum_{u_1,\cdots,u_m} \theta^{\pi}_{\mathbf{e}}(x \mid u_1, \cdots, u_m) Pr(\{u_i, \mathbf{e}^+_{U_iX}\}_{i=1}^{m}) \\
&= \sum_{u_1,\cdots,u_m} \theta^{\pi}_{\mathbf{e}}(x \mid u_1, \cdots, u_m) Pr(u_1, \cdots, u_m, \mathbf{e}^+_X) \\
&= \sum_{u_1,\cdots,u_m} Pr(x \mid u_1, \cdots, u_m, \mathbf{e}^+_X) Pr(u_1, \cdots, u_m, \mathbf{e}^+_X) \\
&= \sum_{u_1,\cdots,u_m} Pr(x, u_1, \cdots, u_m, \mathbf{e}^+_X) \\
&= Pr(x, \mathbf{e}^+_X)
\end{aligned}$$

The support $\lambda(x)$ is computed only if $X = Q$ or X is sending a λ-message to a parent.

The query node Q cannot be abstracted and abstracted nodes do not send λ-messages so X cannot be abstracted. Hence the evidence connected to its children, $\mathbf{e}^-_{XY_j}$, are independent given X. Moreover, the λ-messages it receives are classical (Equation 16).

$$\begin{aligned}
\lambda(x) &= \lambda_{\mathbf{e}}(x) \prod_{i=1}^{n} \lambda_{Y_i}(x) \\
&= \lambda_{\mathbf{e}}(x) \prod_{i=1}^{n} Pr(\mathbf{e}^-_{XY_i} \mid x) \\
&= \lambda_{\mathbf{e}}(x) Pr(\{\mathbf{e}^-_{XY_i}\}_{i=1}^{n} \mid x) \\
&= Pr(\mathbf{e}^-_X \mid x)
\end{aligned}$$

B.6 π-Messages: Proving Equation 15

Suppose node X is not abstracted. Then the evidence connected to its children, $\mathbf{e}^-_{XY_j}$, are independent given X. Moreover, the λ-messages it receives are classical (Equation 16).

$$
\begin{aligned}
\pi_{Y_i}(x) &= \pi(x) \prod_{j \neq i} \lambda_{Y_j}(x) \\
&= Pr(x, \mathbf{e}^+_X) \prod_{j \neq i} Pr(\mathbf{e}^-_{XY_j} \mid x) \\
&= Pr(x, \mathbf{e}^+_X) Pr(\{\mathbf{e}^-_{XY_j}\}_{j \neq i} \mid x) \\
&= Pr(x, \mathbf{e}^+_X) Pr(\{\mathbf{e}^-_{XY_j}\}_{j \neq i} \mid x, \mathbf{e}^+_X) \\
&= Pr(x, \mathbf{e}^+_X, \{\mathbf{e}^-_{XY_j}\}_{j \neq i}) \\
&= Pr(x, \mathbf{e}^+_{XY_i})
\end{aligned}
$$

Suppose node X is abstracted. Then the λ-messages it receives are non-classical (Equation 17) and it will only send a π-message to its Q-child Y_n.

$$
\begin{aligned}
\pi_{Y_n}(x) &= \pi(x) \prod_{i=1}^{n-1} \lambda_{Y_i}(x) \\
&= Pr(x, \mathbf{e}^+_X) \prod_{i=1}^{n-1} Pr(\mathbf{e}^-_{XY_i} \mid x, \mathbf{e}^+_X, \mathbf{e}^-_{XY_1}, \cdots, \mathbf{e}^-_{XY_{i-1}}) \\
&= Pr(x, \mathbf{e}^+_X) Pr(\mathbf{e}^-_{XY_1} \mid x, \mathbf{e}^+_X) Pr(\mathbf{e}^-_{XY_2} \mid x, \mathbf{e}^+_X, \mathbf{e}^-_{XY_1}) \cdots Pr(\mathbf{e}^-_{XY_{n-1}} \mid x, \mathbf{e}^+_X, \mathbf{e}^-_{XY_1}, \cdots, \mathbf{e}^-_{XY_{n-2}}) \\
&= Pr(x, \mathbf{e}^+_X) Pr(\mathbf{e}^-_{XY_1}, \cdots, \mathbf{e}^-_{XY_{n-1}} \mid x, \mathbf{e}^+_X) \\
&= Pr(x, \mathbf{e}^+_X, \mathbf{e}^-_{XY_1}, \cdots, \mathbf{e}^-_{XY_{n-1}}) \\
&= Pr(x, \mathbf{e}^+_{XY_n})
\end{aligned}
$$

B.7 λ-Messages: Proving Equations 16 and 17

Consider the message $\lambda_X(u_i)$ from X to its parent U_i. Node X cannot be abstracted and no other parent U_j of X can be abstracted. We consider two cases.

Parent U_i is not abstracted. All parents of X are not abstracted so X is independent of its non-descendants given its parents and we have $\theta^\pi_{\mathbf{e}}(x \mid u_1, \cdots, u_m) = Pr(x \mid u_1, \cdots, u_m) =$

$Pr(x \mid u_1, \cdots, u_m, \mathbf{e}_X^+)$.

$$\begin{aligned}
\lambda_X(u_i) &= \sum_{x,u_1,\cdots,u_{i-1},u_{i+1},\ldots,u_m} Pr(x \mid u_1, \cdots, u_m, \mathbf{e}_X^+)\lambda(x)\prod_{j\neq i}\pi_X(u_j) \\
&= \sum_{x,u_1,\cdots,u_{i-1},u_{i+1},\ldots,u_m} Pr(x \mid u_1, \cdots, u_m, \mathbf{e}_X^+)Pr(\mathbf{e}_X^- \mid x)\prod_{j\neq i} Pr(u_j, \mathbf{e}_{U_jX}^+) \\
&= \sum_{x,u_1,\cdots,u_{i-1},u_{i+1},\ldots,u_m} Pr(x \mid u_1, \cdots, u_m, \mathbf{e}_X^+)Pr(\mathbf{e}_X^- \mid x)Pr(\{u_j, \mathbf{e}_{U_jX}^+\}_{j\neq i}) \\
&= \sum_x Pr(\mathbf{e}_X^- \mid x) \sum_{u_1,\cdots,u_{i-1},u_{i+1},\ldots,u_m} Pr(x \mid u_1, \cdots, u_m, \mathbf{e}_X^+)Pr(\{u_j, \mathbf{e}_{U_jX}^+\}_{j\neq i}) \\
&= \sum_x Pr(\mathbf{e}_X^- \mid x) \sum_{u_1,\cdots,u_{i-1},u_{i+1},\ldots,u_m} Pr(x \mid \{u_j, \mathbf{e}_{U_jX}^+\}_j)Pr(\{u_j, \mathbf{e}_{U_jX}^+\}_{j\neq i}) \\
&= \sum_x Pr(\mathbf{e}_X^- \mid x) \sum_{u_1,\cdots,u_{i-1},u_{i+1},\ldots,u_m} Pr(x, \{u_j, \mathbf{e}_{U_jX}^+\}_{j\neq i} \mid u_i, \mathbf{e}_{U_iX}^+) \\
&= \\
&= \sum_x Pr(\mathbf{e}_X^- \mid x)Pr(x, \{\mathbf{e}_{U_jX}^+\}_{j\neq i} \mid u_i, \mathbf{e}_{U_iX}^+) \\
&= \sum_x Pr(\mathbf{e}_X^- \mid x)Pr(x, \{\mathbf{e}_{U_jX}^+\}_{j\neq i} \mid u_i) \text{ (requires } U_i \text{ non-abstracted)} \\
&= \sum_x Pr(\mathbf{e}_X^- \mid x, u_i, \{\mathbf{e}_{U_jX}^+\}_{j\neq i})Pr(x, \{\mathbf{e}_{U_jX}^+\}_{j\neq i} \mid u_i) \text{ (requires } X \text{ non-abstracted)} \\
&= \sum_x Pr(x, \mathbf{e}_X^-, \{\mathbf{e}_{U_jX}^+\}_{j\neq i} \mid u_i) \\
&= Pr(\mathbf{e}_X^-, \{\mathbf{e}_{U_jX}^+\}_{j\neq i} \mid u_i) \\
&= Pr(\mathbf{e}_{U_iX}^- \mid u_i)
\end{aligned}$$

Parent U_i is abstracted. The meaning of message $\lambda_X(u_i)$ depends on the order of X within the children of U_i. We will index these children $X_1, \cdots, X_o$ according to order π so X_o is the Q-child of U_i. Only messages $\lambda_{X_j}(u_i)$ for $j = 1, \cdots, o-1$ will be sent. Recall again that X cannot be abstracted and parents U_k, $k \neq i$, cannot be abstracted either. Moreover,

$$\mathbf{e}_{X_j}^+ \setminus \{\mathbf{e}_{U_iX_k}^-\}_{k>j} = \mathbf{e}_{U_i}^+, \{\mathbf{e}_{U_iX_k}^-\}_{k<j}, \{\mathbf{e}_{U_kX_j}^+\}_{k\neq i}$$

and hence

$$\theta_{\mathbf{e}}^{\pi}(x_j \mid u_1, \cdots, u_m) = Pr(x \mid u_1, \cdots, u_m, \mathbf{e}_{U_i}^+, \{\mathbf{e}_{U_iX_k}^-\}_{k<j}, \{\mathbf{e}_{U_kX_j}^+\}_{k\neq i}).$$

Since parents U_k, $k \neq i$, are not abstracted we have that $u_i, \mathbf{e}_{U_i}^+, \{\mathbf{e}_{U_iX_k}^-\}_{k<j}$ is independent of $\{u_k, \mathbf{e}_{U_kX_j}^+\}_{k\neq i}$. This is needed for the following equality, which will be the basis for

showing the meaning of message $\lambda_{X_j}(u_i)$.

$$
\begin{aligned}
&\theta^{\pi}_{\mathbf{e}}(x_j \mid u_1, \cdots, u_m) Pr(\{u_k, \mathbf{e}^+_{U_k X}\}_{k \neq i}) \\
&\quad = \quad Pr(x_j \mid u_1, \cdots, u_m, \mathbf{e}^+_{U_i}, \{\mathbf{e}^-_{U_i X_k}\}_{k<j}, \{\mathbf{e}^+_{U_k X_j}\}_{k \neq i}) Pr(\{u_k, \mathbf{e}^+_{U_k X}\}_{k \neq i}) \\
&\quad = \quad Pr(x_j, \{u_k, \mathbf{e}^+_{U_k X}\}_{k \neq i} \mid u_i, \mathbf{e}^+_{U_i}, \{\mathbf{e}^-_{U_i X_k}\}_{k<j})
\end{aligned}
$$

The last step follows since $Pr(x_j|\alpha, \beta)Pr(\beta) = Pr(x_j, \beta|\alpha)$ when α and β are independent. Here, $\alpha = u_i, \mathbf{e}^+_{U_i}, \{\mathbf{e}^-_{U_i X_k}\}_{k<j}$ and $\beta = \{u_k, \mathbf{e}^+_{U_k X}\}_{k \neq i}$.

We now have

$$
\begin{aligned}
\lambda_{X_j}(u_i) &= \sum_{x_j, u_1, \cdots, u_{i-1}, u_{i+1}, \ldots, u_m} \theta^{\pi}_{\mathbf{e}}(x_j \mid u_1, \cdots, u_m) \lambda(x_j) \prod_{k \neq i} \pi_{X_j}(u_k) \\
&= \sum_{x_j} Pr(\mathbf{e}^-_{X_j} \mid x_j) \sum_{u_1, \cdots, u_{i-1}, u_{i+1}, \ldots, u_m} \theta^{\pi}_{\mathbf{e}}(x_j \mid u_1, \cdots, u_m) Pr(\{u_k, \mathbf{e}^+_{U_k X}\}_{k \neq i}) \\
&= \sum_{x_j} Pr(\mathbf{e}^-_{X_j} \mid x_j) \sum_{u_1, \cdots, u_{i-1}, u_{i+1}, \ldots, u_m} Pr(x_j, \{u_k, \mathbf{e}^+_{U_k X}\}_{k \neq i} \mid u_i, \mathbf{e}^+_{U_i}, \{\mathbf{e}^-_{U_i X_k}\}_{k<j}) \\
&= \sum_{x_j} Pr(\mathbf{e}^-_{X_j} \mid x_j) Pr(x_j, \{\mathbf{e}^+_{U_k X}\}_{k \neq i} \mid u_i, \mathbf{e}^+_{U_i}, \{\mathbf{e}^-_{U_i X_k}\}_{k<j}) \\
&= \sum_{x_j} Pr(\mathbf{e}^-_{X_j} \mid x_j, u_i, \mathbf{e}^+_{U_i}, \{\mathbf{e}^-_{U_i X_k}\}_{k<j}, \{\mathbf{e}^+_{U_k X}\}_{k \neq i}) Pr(x_j, \{\mathbf{e}^+_{U_k X}\}_{k \neq i} \mid u_i, \mathbf{e}^+_{U_i}, \{\mathbf{e}^-_{U_i X_k}\}_{k<j}) \\
&= \sum_{x_j} Pr(\mathbf{e}^-_{X_j}, x_j, \{\mathbf{e}^+_{U_k X}\}_{k \neq i} \mid u_i, \mathbf{e}^+_{U_i}, \{\mathbf{e}^-_{U_i X_k}\}_{k<j}) \\
&= Pr(\mathbf{e}^-_{X_j}, \{\mathbf{e}^+_{U_k X}\}_{k \neq i} \mid u_i, \mathbf{e}^+_{U_i}, \{\mathbf{e}^-_{U_i X_k}\}_{k<j}) \\
&= Pr(\mathbf{e}^-_{U_i X_j} \mid u_i, \mathbf{e}^+_{U_i}, \mathbf{e}^-_{U_i X_1}, \cdots, \mathbf{e}^-_{U_i X_{j-1}})
\end{aligned}
$$