Revisiting the General Identifiability Problem
Appendix

Yaroslav Kivva[1] **Ehsan Mokhtarian**[1] **Jalal Etesami**[1] **Negar Kiyavash**[1,2]

[1]School of Computer and Communication Sciences, EPFL, Lausanne, Switzerland
[2]College of Management of Technology, EPFL, Lausanne, Switzerland

1 ON THE POSITIVITY ASSUMPTION

We first present some definitions and notations from [Lee et al., 2019] including their illustrations using the causal graph $\mathcal{G}$ from Example 2 in the main text.

1.1 NOTATION

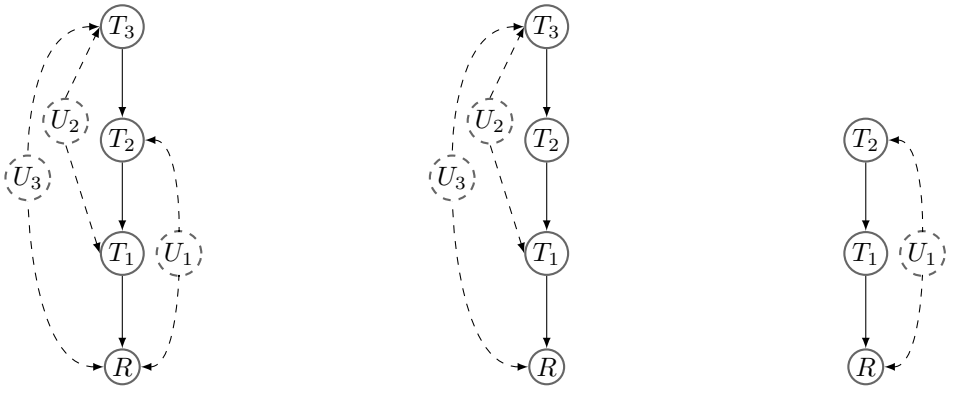
Definition 1 ([Lee et al., 2019])**.** *Assume that* $\mathbf{R}$ *is a subset of observed variables* $\mathbf{V}$*. A hedge is a pair of* $\mathbf{R}$*-rooted c-forests* $\langle \mathcal{F}, \mathcal{F}' \rangle$ *such that* $\mathcal{F}'$ *is a subgraph of* $\mathcal{F}$*.*

In Figure 2 of the main text: Subgraphs $\mathcal{F} = \mathcal{G}[\{R, T_1, T_2, T_3\}]$ and $\mathcal{F}' = \mathcal{G}[\{R\}]$ form a hedge $\langle \mathcal{F}, \mathcal{F}' \rangle$.

Denote by $\mathcal{C}(\mathcal{G}) = \{\mathbf{W}_i\}_{i=1}^{k}$, the set of c-components that partition observed variables in $\mathcal{G}$ such that each $\mathbf{W}_i$ is a maximal c-component. Maximal in the sense of number of nodes that is there is no $\mathbf{W} \in \mathbf{V}$ such that $\mathbf{W}_i \subsetneq \mathbf{W}$ and $\mathbf{W}$ is a c-component in $\mathcal{G}$. Assume that $\mathbf{T}$ is the set of all observed variables in $\mathcal{F}$ but not in $\mathcal{F}'$. We define $\mathcal{F}'' := \mathcal{F}[\mathbf{T}]$.

In Figure 2 of the main text: $\mathcal{C}(\mathcal{G}[\{T_1, T_2, T_3\}]) = \{\{T_1, T_3\}, \{T_2\}\}$. Additionally, $\mathcal{F}'' = \mathcal{G}[\{T_1, T_2, T_3\}]$ for the hedge constructed before.

Definition 2 (Lee et al. [2019])**.** *Given a hedge* $\langle \mathcal{F}, \mathcal{F}' \rangle$*. Denote by* $\mathbf{V}'$ *a set of all observed variables of* $\mathcal{F}'$*. The hedgelet decomposition of a hedge* $\langle \mathcal{F}, \mathcal{F}' \rangle$ *is a collection of hedgelets* $\{\mathcal{F}(\mathbf{W})\}_{\mathbf{W} \in \mathcal{C}(\mathcal{F}'')}$ *where each hedgelet* $\mathcal{F}(\mathbf{W})$ *is a subgraph*



(a) Thicket $\mathcal{J}$ (b) Hedgelet $\mathcal{H}_1$ (c) Hedgelet $\mathcal{H}_2$

Figure 1: (a) Thicket is formed for the causal effect of $\{T_1, T_2, T_3\}$ on $\{R\}$ in Example 2; (b) and (c) are the hedgelets formed by the thicket $\mathcal{J}$

Accepted for the 38th Conference on Uncertainty in Artificial Intelligence (UAI 2022).

of $\mathcal{F}$ made of (i) $\mathcal{F}[\mathbf{W} \cup \mathbf{V}']$ and (ii) $\mathcal{F}[De_{\mathcal{F}}(\mathbf{W})]$ without bidirected edges, that is all observed descendants of $\mathbf{W}$ and all directed edges between them. Let $\mathbb{H}_{\mathcal{F}} := \{\mathcal{F}(\mathbf{W})\}_{\mathbf{W} \in \mathcal{C}(\mathcal{F}'')}$ be the set of hedgelets of $\langle \mathcal{F}, \mathcal{F}' \rangle$.

In Figure 2 of the main text: For the hedge $\langle \mathcal{F}, \mathcal{F}' \rangle$, where $\mathcal{F} = \mathcal{G}[\{R, T_1, T_2, T_3\}]$ and $\mathcal{F}' = \mathcal{G}[\{R\}]$, there are two hedgelets $\mathcal{H}_1, \mathcal{H}_2$ displayed in Figures (1b)-(1c). Moreover, we have $\mathbb{H}_{\mathcal{F}} = \{\mathcal{H}_1, \mathcal{H}_2\}$.

Definition 3 ([Lee et al., 2019])**.** *Let $\mathbf{R}$ be a non-empty set of variables and $\mathbb{Z}$ be a collection of sets of variables in $\mathcal{G}$. A thicket $\mathcal{J}$ is a subgraph of $\mathcal{G}$ which is an $\mathbf{R}$-rooted c-component consisting of a minimal c-component over $\mathbf{R}$ and hedges*

$$\mathbb{F}_{\mathcal{J}} := \{\langle \mathcal{F}_{\mathbf{Z}}, \mathcal{J}[\mathbf{R}] \rangle \mid \mathcal{F}_{\mathbf{Z}} \subseteq \mathcal{G}[\mathbf{V} \setminus \mathbf{Z}], \mathbf{Z} \cap \mathbf{R} = \varnothing\}_{\mathbf{Z} \in \mathbb{Z}}.$$

Let $\mathbf{X}$ and $\mathbf{Y}$ be disjoint sets of observed variables in $\mathcal{G}$. A thicket $\mathcal{J}$ is said to be formed for $P_{\mathbf{x}}(\mathbf{y})$ in $\mathcal{G}$ with respect to $\mathbb{Z}$ if $\mathbf{R} \subseteq Anc_{\mathcal{G}[\mathbf{V} \setminus \mathbf{X}]}(\mathbf{Y})$ and every hedgelet of each hedge $\langle \mathcal{F}_{\mathbf{Z}}, \mathcal{J}[\mathbf{R}] \rangle$ intersects with $\mathbf{X}$.

In Figure 2 of the main text: This graph is a thicket, also displayed in Figure 1a. Let $\mathbb{F}_{\mathcal{J}}$ be

$$\mathbb{F}_{\mathcal{J}} = \{\langle \mathcal{F}, \mathcal{F}' \rangle\},$$

where $\mathcal{F} = \mathcal{G}[\{R, T_1, T_2, T_3\}]$ and $\mathcal{F}' = \mathcal{G}[\{R\}]$. One can observe that thicket $\mathcal{J}$ is formed for the causal effect $\mathbf{X} = \{T_1, T_2, T_3\}$ on $\mathbf{Y} = \{R\}$.

Denote by $\mathbf{T}$ all observed variables in thicket $\mathcal{J}$ outside of subgraph $\mathcal{J}[\mathbf{R}]$. Let $\mathbb{H} = \bigcup_{\{\langle \mathcal{F}, \mathcal{F}' \rangle\} \in \mathbb{F}_{\mathcal{J}}} \mathbb{H}_{\mathcal{F}}$, that is, a collection of all hedgelets induced by the hedges of $\mathcal{J}$.

In Figure 2 of the main text: $\mathbf{T} = \{T_1, T_2, T_3\}$ and $\mathbb{H} = \{\mathcal{H}_1, \mathcal{H}_2\}$.

1.2 ON THE POSITIVITY ASSUMPTION

Given the above definitions, we can state Lemma 3 in [Lee et al., 2019].

Lemma. *Let $\mathbf{T}' \subsetneq \mathbf{T}$ such that there exists a hedgelet $\mathcal{H} \in \mathbb{H} \setminus \mathbb{H}(\mathbf{T}')$, where $\mathbb{H}(\mathbf{T}')$ is a set of hedgelets from $\mathbb{H}$ which contain at least one variable from $\boldsymbol{T}'$. Then, under the intervention $do(\mathbf{t}')$, there exists $R \in \mathbf{R}$, for any instantiation of $\mathbf{U}$, such that $r = 0$ in both models.*

Note that by the construction in [Lee et al., 2019], R in the above Lemma is a binary random variable. In the above Lemma, let $\mathbf{T}' = \varnothing$. Based on this Lemma, for any instantiation of unobserved variables $\mathbf{U}$, $P(\mathbf{V} = \mathbf{v}) = 0$, where $\mathbf{v}$ is a realization for observed variables in which $r = 1$. This clearly shows that the constructed models in [Lee et al., 2019] violate the positivity assumption.

1.3 ON THE RELAXED POSITIVITY ASSUMPTION

Herein, we study Figure 2 of the main text in more details and show that the models in [Lee et al., 2019] violate the relaxed positivity assumption. To this end, we present the models $\mathcal{M}_1$ and $\mathcal{M}_2$ constructed in [Lee et al., 2019] for the thicket $\mathcal{J}$ which is defined for this case in Appendix 1.1. By the construction, each variable from $\{U_1, U_2, U_3, T_3\}$ is a binary number, i.e., $\{0, 1\}$ and each variable from $\{T_1, T_2\}$ is a vector of length two, because each variable from $\{U_1, U_2, U_3, T_3\}$ appears in only one hedgelet and each variables in $\{T_1, T_2\}$ appears in exactly two different hedgelets. Thus, $T_1 = (T_{1,1}, T_{1,2})$ and $T_2 = (T_{2,1}, T_{2,2})$, where $T_{1,1}, T_{1,2}, T_{2,1}, T_{2,2}$ are binary numbers. The first coordinate captures some properties of the hedgelet $\mathcal{H}_1$ while the second coordinate captures some properties of the hedgelet $\mathcal{H}_2$. Lee et al. [2019] define both models $\mathcal{M}_1, \mathcal{M}_2$ for the hedgelet $\mathcal{H}_1$ as

$$T_3 = U_2 \oplus U_3, \quad T_{2,1} = T_3, \quad T_{1,1} = T_{2,1} \oplus U_2,$$

and for the hedgelet $\mathcal{H}_2$ as

$$T_{2,2} = U_1, \quad T_{1,2} = T_{2,2}, \quad T_{2,2} = U_1.$$

Additionally, in model $\mathcal{M}_1$, variable R is defined by

$$R = \mathbb{1}_{T_{1,1}=0} \wedge \mathbb{1}_{T_{1,2}=0} \wedge \mathbb{1}_{U_3=1} \wedge \mathbb{1}_{U_1=1},$$

and in model $\mathcal{M}_2$, it is defined to be zero, i.e., $R = 0$.

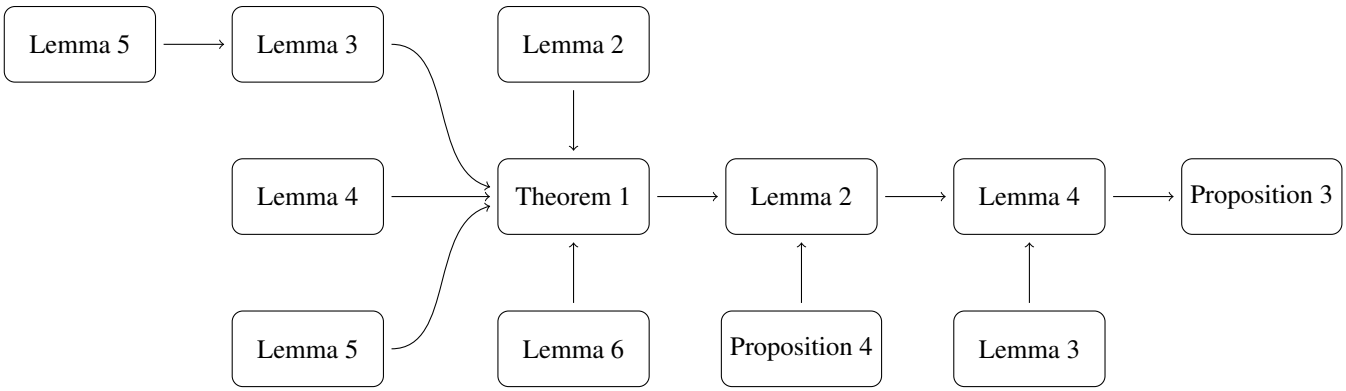


Figure 2: Logical order of our proofs.

2 TECHNICAL PROOFS

In this section, we first present some technical lemmas which we use throughout our proofs. The proofs of the lemmas and propositions within the main text are provided in Subsections 2.2 and 2.3.

The logical order of our proofs is depicted in Figure 2. For instance, we use Theorem 1 to prove Lemma 2. Also note that the proof of Theorem 1 is provided in the main text using Lemmas 2, 3, 4, 5, and 6.

Definition 4 (Ancestral)**.** *We say a subset* $\mathbf{X}$ *of observed variables* $\mathbf{V}$ *is ancestral in* $\mathcal{G}$*, if* $\mathbf{X} = Anc_{\mathcal{G}_\mathbf{V}}(\mathbf{X})$.

2.1 TECHNICAL LEMMAS

Lemma 1 ([Tian and Pearl, 2003])**.** *Let* $\mathbf{W} \subseteq \mathbf{C} \subseteq \mathbf{V}$*,* $\mathbf{T} = \mathbf{C} \setminus \mathbf{W}$*,* $\mathbf{S} = \mathbf{V} \setminus \mathbf{T}$*. If* $\mathbf{W}$ *is an ancestral set in* $\mathcal{G}[\mathbf{C}]$*, then:*

$$Q[\mathbf{W}] = \sum_{\mathbf{C}\setminus\mathbf{W}} Q[\mathbf{C}].$$

Lemma 2. *Consider a causal graph* $\mathcal{G}$ *with observed variables* $\mathbf{V}$*. Suppose* $\mathbf{X} \subseteq \mathbf{V}$ *and* $e := (X_1, Z)$ *is a directed edge such that* $X_1 \in \mathbf{X}$*.* $Q[\mathbf{X}]$ *is g-identifiable from* $(\mathbb{A}, \mathcal{G})$ *if and only if* $Q[\mathbf{X}]$ *is g-identifiable from* $(\mathbb{A}, \mathcal{H})$*, where* $\mathcal{H}$ *is the graph obtained by deleting* e *from* $\mathcal{G}$*.*

Proof. $\mathbf{X}$ has the same c-components in $\mathcal{G}$ and $\mathcal{H}$ since $\mathcal{G}_\mathbf{V}$ and $\mathcal{H}_\mathbf{V}$ have the same undirected edges. Let $\mathbf{X}_1, \cdots, \mathbf{X}_l$ be the c-components of $\mathbf{X}$. For any $i \in [1:l]$ and $\mathbf{A} \in \mathbb{A}$ such that $\mathbf{X}_i \subseteq \mathbf{A}$, Huang and Valtorta [2008] showed that $Q[\mathbf{X}_i]$ is identifiable from $\mathcal{G}[\mathbf{A}]$ if and only if $Q[\mathbf{X}_i]$ is identifiable from $\mathcal{H}[\mathbf{A}]$. Hence, Theorem 1 implies that $Q[\mathbf{X}_i]$ is g-identifiable from $(\mathbb{A}, \mathcal{G})$ if and only if $Q[\mathbf{X}_i]$ is g-identifiable from $(\mathbb{A}, \mathcal{H})$. In this case, Proposition 4 implies that $Q[\mathbf{X}]$ is g-identifiable from $(\mathbb{A}, \mathcal{G})$ if and only if $Q[\mathbf{X}]$ is g-identifiable from $(\mathbb{A}, \mathcal{H})$. □

Lemma 3. *Suppose that* $\mathbf{X}$ *and* $\mathbf{Y}$ *are disjoint subsets of* $\mathbf{V}$*. Let* (Y_1, Y_2) *(i.e.,* $Y_1 \to Y_2$*) denotes a directed edge in* $\mathcal{G}$*, where* $Y_1, Y_2 \in \mathbf{Y}$*. Let* $\mathcal{G}'$ *denotes the resulting graph after removing edge* (Y_1, Y_2) *from* $\mathcal{G}$*. If the causal effect of* $\mathbf{X}$ *on* $\mathbf{Y}$ *is not g-identifiable from* $(\mathbb{A}, \mathcal{G}')$*, then the causal effect of* $\mathbf{X}$ *on* $\mathbf{Y} \setminus \{Y_1\}$ *is not g-identifiable from* $(\mathbb{A}, \mathcal{G})$*.*

Proof. Herein, we provide a proof that is similar to one of the proofs in [Huang and Valtorta, 2008].

Using Markov factorization property in graph $\mathcal{G}'$, $P_\mathbf{x}(\mathbf{y})$ is given by

$$P_\mathbf{x}(\mathbf{y}) = \sum_{\mathbf{V}\setminus(\mathbf{X}\cup\mathbf{Y})} \sum_{\mathbf{U}} \prod_{W\in\mathbf{V}\setminus\mathbf{X}} P(w \mid Pa_{\mathcal{G}'}(W)) \prod_{U\in\mathbf{U}} P(u).$$

Similarly, in graph $\mathcal{G}$ we have

$$P_\mathbf{x}(\mathbf{y} \setminus \{Y_1\}) = \sum_{\{Y_1\}\cup(\mathbf{V}\setminus(\mathbf{X}\cup\mathbf{Y}))} \sum_{\mathbf{U}} \prod_{W\in\mathbf{V}\setminus\mathbf{X}} P(w \mid Pa_{\mathcal{G}}(W)) \prod_{U\in\mathbf{U}} P(u).$$

Since the causal effect of $\mathbf{X}$ on $\mathbf{Y}$ is not g-identifiable from $(\mathbb{A}, \mathcal{G}')$, there exists $\mathcal{M}_1$ and $\mathcal{M}_2$ in $\mathbb{M}^+(\mathcal{G}')$ such that:

$$Q^{\mathcal{M}_1}[\mathbf{A}_i](\mathbf{v}) = Q^{\mathcal{M}_2}[\mathbf{A}_i](\mathbf{v}),\ \forall \mathbf{v} \in \mathfrak{X}_{\mathbf{V}},\ \forall i \in [0:m],$$

$$P_{\mathbf{x}}^{\mathcal{M}_1}(\mathbf{y}) \neq P_{\mathbf{x}}^{\mathcal{M}_2}(\mathbf{y}),\ \exists \mathbf{x} \in \mathfrak{X}_{\mathbf{X}},\ \exists \mathbf{y} \in \mathfrak{X}_{\mathbf{Y}}.$$

Using $\mathcal{M}_1$ and $\mathcal{M}_2$, we construct two SEMs $\mathcal{M}_1'$ and $\mathcal{M}_2'$ in $\mathbb{M}^+(\mathcal{G})$. Define a surjective function $F\colon\ \mathfrak{X}_{Y_1} \to \{0,1\}$ and a function $\Psi\colon\ \{0,1\} \times \mathfrak{X}_{Y_1} \to (0,1)$ such that $\Psi(0,y_1) + \Psi(1,y_1) = 1$ for each $y_1 \in \mathfrak{X}_{Y_1}$. We will later assume some constraints for these functions, but for now lets assume they are arbitrary.

For any node S which is either unobserved or in $\mathbf{V} \setminus (\{Y_2\} \cup \mathit{Ch}_{\mathcal{G}}(Y_2))$, we define

$$P^{\mathcal{M}_i'}(s|\mathit{Pa}_{\mathcal{G}}(S)) = P^{\mathcal{M}_i}(s|\mathit{Pa}_{\mathcal{G}'}(S)),$$

where $i \in \{1,2\}$. The domain of Y_2 in $\mathcal{M}_i'$ is defined as $\mathfrak{X}_{Y_2}^{\mathcal{M}} \times \{0,1\}$, where $\mathfrak{X}_{Y_2}^{\mathcal{M}}$ is the domain of Y_2 in $\mathcal{M}_i$. For $y_2 \in \mathfrak{X}_{Y_2}^{\mathcal{M}}$, $i \in \{0,1\}$, and $k \in \{0,1\}$ we define:

$$P^{\mathcal{M}_i'}((y_2,k) \mid \mathit{Pa}_{\mathcal{G}'}(Y_2), y_1) = P^{\mathcal{M}_i}(y_2 \mid \mathit{Pa}_{\mathcal{G}'}(Y_2))\Psi(F(y_1) \oplus k, y_1).$$

Note that $\mathit{Pa}_{\mathcal{G}'}(Y_2) \cup \{Y_1\} = \mathit{Pa}_{\mathcal{G}}(Y_2)$. Moreover, for a fixed realization $(\mathit{Pa}_{\mathcal{G}'}(Y_2), y_1)$, we have:

$$\sum_{k \in \{0,1\}} \sum_{y_2 \in \mathfrak{X}_{Y_2}^{\mathcal{M}}} P^{\mathcal{M}_i'}((y_2,k)|pa(Y_2), y_1) = 1.$$

For each $S \in \mathit{Ch}_{\mathcal{G}}(Y_2)$, we define:

$$P^{\mathcal{M}_i'}(s \mid \mathit{Pa}_{\mathcal{G}}(S) \setminus \{Y_2\}, (y_2,k)) = P^{\mathcal{M}_i}(s \mid \mathit{Pa}_{\mathcal{G}}(S) \setminus \{Y_2\}, y_2).$$

Next, we show that $Q^{\mathcal{M}_1'}[\mathbf{A}_i](\mathbf{v}) = Q^{\mathcal{M}_2'}[\mathbf{A}_i](\mathbf{v})$ for each $\mathbf{v} \in \mathfrak{X}_{\mathbf{V}}$ and $i \in [0:m]$. Suppose $\mathbf{v}$ is a realization of $\mathbf{V}$ in $\mathcal{M}_1'$ with realizations y_1 and (y_2,k) for Y_1 and Y_2, respectively. Consider two cases:

- If $Y_2 \notin \mathbf{A}_i$:

$$\begin{aligned}
Q^{\mathcal{M}_1'}[\mathbf{A}_i](\mathbf{v}) &= \sum_{\mathbf{U}} \prod_{A \in \mathbf{A}_i} P^{\mathcal{M}_1'}(a \mid \mathit{Pa}_{\mathcal{G}}(A)) \prod_{U \in \mathbf{U}} P^{\mathcal{M}_1'}(u) \\
&= \sum_{\mathbf{U}} \prod_{A \in \mathbf{A}_i} P^{\mathcal{M}_1}(a \mid \mathit{Pa}_{\mathcal{G}'}(A)) \prod_{U \in \mathbf{U}} P^{\mathcal{M}_1}(u) = Q^{\mathcal{M}_1}[\mathbf{A}_i](\mathbf{v}) = Q^{\mathcal{M}_2}[\mathbf{A}_i](\mathbf{v}) \\
&= \sum_{\mathbf{U}} \prod_{A \in \mathbf{A}_i} P^{\mathcal{M}_2}(a \mid \mathit{Pa}_{\mathcal{G}'}(A)) \prod_{U \in \mathbf{U}} P^{\mathcal{M}_2}(u) \\
&= \sum_{\mathbf{U}} \prod_{A \in \mathbf{A}_i} P^{\mathcal{M}_2'}(a \mid \mathit{Pa}_{\mathcal{G}}(A)) \prod_{U \in \mathbf{U}} P^{\mathcal{M}_2'}(u) \\
&= Q^{\mathcal{M}_2'}[\mathbf{A}_i](\mathbf{v}).
\end{aligned}$$

- If $Y_2 \in \mathbf{A}_i$:

$$\begin{aligned}
Q^{\mathcal{M}_1'}[\mathbf{A}_i](\mathbf{v}) &= \sum_{\mathbf{U}} \prod_{A \in \mathbf{A}_i} P^{\mathcal{M}_1'}(a \mid \mathit{Pa}_{\mathcal{G}}(A)) \prod_{U \in \mathbf{U}} P^{\mathcal{M}_1'}(u) \\
&= \Psi\left(F(y_1) \oplus k, y_1\right) \sum_{\mathbf{U}} \prod_{A \in \mathbf{A}_i} P^{\mathcal{M}_1}(a \mid \mathit{Pa}_{\mathcal{G}'}(A)) \prod_{U \in \mathbf{U}} P^{\mathcal{M}_1}(u) \\
&= \Psi(F(y_1) \oplus k, y_1) Q^{\mathcal{M}_1}[\mathbf{A}_i](\mathbf{v}) = \Psi(F(y_1) \oplus k, y_1) Q^{\mathcal{M}_2}[\mathbf{A}_i](\mathbf{v}) \\
&= \Psi(F(y_1) \oplus k, y_1) \sum_{\mathbf{U}} \prod_{A \in \mathbf{A}_i} P^{M_2}(a \mid \mathit{Pa}_{\mathcal{G}'}(A)) \prod_{U \in \mathbf{U}} P^{M_2}(u) \\
&= \sum_{\mathbf{U}} \prod_{A \in \mathbf{A}_i} P^{\mathcal{M}_2'}(a \mid \mathit{Pa}_{\mathcal{G}}(A))) \prod_{U \in \mathbf{U}} P^{\mathcal{M}_2'}(u) \\
&= Q^{\mathcal{M}_2'}[\mathbf{A}_i](\mathbf{v}).
\end{aligned}$$

Therefore, $Q^{\mathcal{M}_1'}[\mathbf{A}_i](\mathbf{v}) = Q^{\mathcal{M}_2'}[\mathbf{A}_i](\mathbf{v})$ for each $\mathbf{v} \in \mathfrak{X}_{\mathbf{V}}$ and $i \in [0:m]$.

We know that there exists $\hat{\mathbf{x}} \in \mathfrak{X}_{\mathbf{X}}^{\mathcal{M}}$ and $\hat{\mathbf{y}} \in \mathfrak{X}_{\mathbf{Y}}^{\mathcal{M}}$ such that $P_{\hat{\mathbf{x}}}^{\mathcal{M}_1}(\hat{\mathbf{y}}) \neq P_{\hat{\mathbf{x}}}^{\mathcal{M}_2}(\hat{\mathbf{y}})$. Denote by $\hat{y}_1$ and $\hat{y}_2$ the realizations of Y_1 and Y_2 in the realization $\hat{\mathbf{y}}$, respectively. Assume that $P_{\hat{\mathbf{x}}}^{\mathcal{M}_1}(\hat{\mathbf{y}}) = d_1 > P_{\hat{\mathbf{x}}}^{\mathcal{M}_2}(\hat{\mathbf{y}}) = d_2$. Assume that $\Psi(F(\hat{y}_1) \oplus 0, \hat{y}_1) = 0.5$ and $\Psi(F(y) \oplus 0, y) = \frac{d_1 - d_2}{4}$ for all $y \in \mathfrak{X}_{Y_1} \setminus \{\hat{y}_1\}$. Then we have:

$$\begin{aligned}
P_{\hat{\mathbf{x}}}^{\mathcal{M}_1'}(\hat{\mathbf{y}} \setminus \{\hat{y}_1\}) &= \sum_{y_1 \in \mathfrak{X}_{Y_1}} \sum_{\mathbf{V}\setminus(\mathbf{X}\cup\mathbf{Y})} \sum_{\mathbf{U}} \prod_{Z \in \mathbf{V}\setminus\mathbf{X}} P^{\mathcal{M}_1'}(z \mid Pa_{\mathcal{G}}(Z)) \prod_{U \in \mathbf{U}} P(u) \\
&> \sum_{y_1 = \hat{y}_1} \sum_{\mathbf{V}\setminus(\mathbf{X}\cup\mathbf{Y})} \sum_{\mathbf{U}} \prod_{Z \in \mathbf{V}\setminus\mathbf{X}} P^{\mathcal{M}_1'}(z \mid Pa_{\mathcal{G}}(Z)) \prod_{U \in \mathbf{U}} P(u) \\
&= P_{\hat{\mathbf{x}}}^{\mathcal{M}_1}(\hat{\mathbf{y}}) \Psi(F(\hat{y}_1) \oplus 0, \hat{y}_1) = 0.5 d_1.
\end{aligned}$$

but,

$$\begin{aligned}
P_{\hat{\mathbf{x}}}^{\mathcal{M}_2'}(\hat{\mathbf{y}} \setminus \{\hat{y}_1\}) &= \sum_{y_1 \in \mathfrak{X}_{Y_1}} \sum_{\mathbf{V}\setminus(\mathbf{X}\cup\mathbf{Y})} \sum_{\mathbf{U}} \prod_{Z \in \mathbf{V}\setminus\mathbf{X}} P^{\mathcal{M}_1'}(z \mid Pa_{\mathcal{G}}(Z)) \prod_{U \in \mathbf{U}} P(u) \\
&= \sum_{y_1 = \hat{y}_1} \sum_{\mathbf{V}\setminus(\mathbf{X}\cup\mathbf{Y})} \sum_{\mathbf{U}} \prod_{Z \in \mathbf{V}\setminus\mathbf{X}} P^{\mathcal{M}_1'}(z \mid Pa_{\mathcal{G}}(Z)) \prod_{U \in \mathbf{U}} P(u) \\
&+ \sum_{y_1 \in \mathfrak{X}_{Y_1} \setminus \{\hat{y}_1\}} \sum_{\mathbf{V}\setminus(\mathbf{X}\cup\mathbf{Y})} \sum_{\mathbf{U}} \prod_{Z \in \mathbf{V}\setminus\mathbf{X}} P^{M_1'}(z \mid Pa_{\mathcal{G}}(Z)) \prod_{U \in \mathbf{U}} P(u) \\
&\leq P_{\hat{\mathbf{x}}}^{\mathcal{M}_2}(\hat{\mathbf{y}}) \Psi(F(\hat{y}_1) \oplus 0, \hat{y}_1) + P_{\hat{\mathbf{x}}}^{\mathcal{M}_2}(\hat{\mathbf{y}} \setminus \{\hat{y}_1\}) \Psi(F(Y_1 \neq y_1) \oplus 0, Y_1 \neq y_1) \\
&= 0.5 d_2 + \frac{d_1 - d_2}{4} < 0.5 d_1.
\end{aligned}$$

This implies that $P_{\hat{\mathbf{x}}}^{\mathcal{M}_1'}(\hat{\mathbf{y}} \setminus \{\hat{y}_1\}) \neq P_{\hat{\mathbf{x}}}^{\mathcal{M}_2'}(\hat{\mathbf{y}} \setminus \{\hat{y}_1\})$ which concludes the proof. □

Lemma 4. *Assume $\mathbf{Y} \subset \mathbf{W} \subset \mathbf{V}$ such that for each $W \in \mathbf{W} \setminus \mathbf{Y}$, there exists a directed path in $\mathcal{G}[\mathbf{W}]$ from W to a variable in $\mathbf{Y}$. Then, the causal effect of $\mathbf{V} \setminus \mathbf{W}$ on $\mathbf{Y}$ is g-identifiable from $(\mathbb{A}, \mathcal{G})$ if and only if $Q[\mathbf{W}]$ is g-identifiable from $(\mathbb{A}, \mathcal{G})$.*

Proof. Let $\mathbf{X} := \mathbf{V} \setminus \mathbf{W}$.

Sufficient part: Suppose $Q[\mathbf{W}]$ is g-identifiable from $(\mathbb{A}, \mathcal{G})$. Since $Q[\mathbf{W}] = P_{\mathbf{x}}(\mathbf{W})$, we have

$$P_{\mathbf{x}}(\mathbf{y}) = \sum_{\mathbf{W}\setminus\mathbf{Y}} Q[\mathbf{W}].$$

Hence, $P_{\mathbf{x}}(\mathbf{y})$ is uniquely computed and the causal effect of $\mathbf{X}$ on $\mathbf{Y}$ is g-identifiable from $(\mathbb{A}, \mathcal{G})$.

Necessary part: Suppose $Q[\mathbf{W}]$ is not g-identifiable from $(\mathbb{A}, \mathcal{G})$, we will show that $P_{\mathbf{x}}(\mathbf{y})$ is also not g-identifiable. To this end, first, we order the nodes in $\mathbf{W} \setminus \mathbf{Y}$, say $(W_1, W_2, \cdots, W_n)$, such that for each $1 \leq i \leq n$, W_i is a parent of at least one node in $\mathbf{Y} \cup \{W_1, W_2, \ldots, W_{i-1}\}$. Assume that e_i is the directed edge from W_i to its child in $\mathbf{Y} \cup \{W_1, W_2, \ldots, W_{i-1}\}$. We also define $\mathcal{G}'$ to be the graph obtained by deleting all the edges $\{e_i\}_{i=1}^n$ from $\mathcal{G}$. Applying Lemma 2 repeatedly n times imply that $Q[\mathbf{W}]$ is not g-identifiable from $(\mathbb{A}, \mathcal{G}')$.

Let $\mathcal{G}_n := \mathcal{G}$ and for $0 \leq i \leq n-1$, we define $\mathcal{G}_i$ to be the graph obtained by removing e_{i+1} from $\mathcal{G}_{i+1}$. From Lemma 3, we know that if $Q[\mathbf{W}]$ is not g-identifiable from $(\mathbb{A}, \mathcal{G}')$, then adding edge e_1 will make the causal effect of $\mathbf{X}$ on $\mathbf{W} \setminus \{W_1\}$ not g-identifiable from $(\mathbb{A}, \mathcal{G}_1)$. Note that $\mathcal{G}_1$ is obtained from $\mathcal{G}'$ by adding edge e_1. Using this lemma again implies that the causal effect of $\mathbf{X}$ on $\mathbf{W} \setminus \{W_1, W_2\}$ is not g-identifiable from $(\mathbb{A}, \mathcal{G}_2)$. Repeating this procedure yields that the causal effect of $\mathbf{X}$ on $\mathbf{W} \setminus \{W_1, \ldots W_n\} = \mathbf{Y}$ is not g-identifiable from $(\mathbb{A}, \mathcal{G}_n)$. Since $\mathcal{G}_n = \mathcal{G}$, the causal effect of $\mathbf{X}$ on $\mathbf{Y}$ is not g-identifiable from $(\mathbb{A}, \mathcal{G})$. □

Lemma 5. *Consider a set of vectors $\{c_i\}_{i=1}^n$, where $c_i \in \mathbb{R}^d$. Assume $c \in \mathbb{R}^d$ is a vector that is linearly independent of $\{c_i\}_{i=1}^n$, then there is a vector $b \in \mathbb{R}^d$ such that*

$$\begin{aligned}
&\langle c_i, b \rangle = 0, \quad \forall i \in [1:n], \\
&\langle c, b \rangle \neq 0.
\end{aligned}$$

Proof. Denote by $\{\phi_i\}_{i=1}^{l}$ a subset of $\{c_i\}_{i=1}^{n}$ which forms a basis for the vectors in $\{c_i\}_{i=1}^{n}$. Clearly, we have $l < d$. Now, consider the following system of linear equations with respect to b:

$$\begin{aligned} &\langle \phi_i, b\rangle = 0, \quad \forall i \in [1:l], \\ &\langle c, b\rangle = 13 \neq 0. \end{aligned} \tag{1}$$

By the assumption, vectors in $\{\phi_i\}_{i=1}^{l} \cup \{c\}$ are linearly independent, thus there exists a solution to (1). □

2.2 PROOFS OF SECTION 4

Proposition 3. *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be two disjoint subsets of* $\mathbf{V}$. *The causal effect of* $\mathbf{X}$ *on* $\mathbf{Y}$ *is g-identifiable from* $(\mathbb{A}, \mathcal{G})$ *if and only if* $Q[Anc_{\mathcal{G}_{\mathbf{V}\setminus\mathbf{X}}}(\mathbf{Y})]$ *is g-identifiable from* $(\mathbb{A}, \mathcal{G})$.

Proof. Let $\mathbf{W} := Anc_{\mathcal{G}_{\mathbf{V}\setminus\mathbf{X}}}(\mathbf{Y})$. Since $Q[\mathbf{V}\setminus\mathbf{X}] = P_{\mathbf{x}}(\mathbf{V}\setminus\mathbf{X})$, using marginalization, we obtain

$$P_{\mathbf{x}}(\mathbf{y}) = \sum_{\mathbf{V}\setminus(\mathbf{X}\cup\mathbf{Y})} Q[\mathbf{V}\setminus\mathbf{X}] = \sum_{\mathbf{W}\setminus\mathbf{Y}} \sum_{\mathbf{V}\setminus(\mathbf{W}\cup\mathbf{X})} Q[\mathbf{V}\setminus\mathbf{X}]. \tag{2}$$

Since $\mathbf{W}$ is an ancestral set in $\mathcal{G}[\mathbf{V}\setminus\mathbf{X}]$, Lemma 1 implies

$$\sum_{\mathbf{V}\setminus(\mathbf{W}\cup\mathbf{X})} Q[\mathbf{V}\setminus\mathbf{X}] = Q[\mathbf{W}].$$

Substituting the above equation into (2) implies

$$P_{\mathbf{x}}(\mathbf{y}) = \sum_{\mathbf{W}\setminus\mathbf{Y}} Q[\mathbf{W}] = P_{\mathbf{v}\setminus\mathbf{w}}(\mathbf{y}). \tag{3}$$

Sufficient part: Suppose $Q[\mathbf{W}]$ is g-identifiable from $(\mathbb{A}, \mathcal{G})$. Equation (3) implies that $P_{\mathbf{x}}(\mathbf{y})$ is uniquely computable from $Q[\mathbf{W}]$, and therefore, the causal effect of $\mathbf{X}$ on $\mathbf{Y}$ is g-identifiable from $(\mathbb{A}, \mathcal{G})$.

Necessary part: Suppose $Q[\mathbf{W}]$ is not g-identifiable from $(\mathbb{A}, \mathcal{G})$. For each $W \in \mathbf{W}\setminus\mathbf{Y}$, there exists a directed path in $\mathcal{G}[\mathbf{W}]$ from W to a variable in $\mathbf{Y}$. Hence, Lemma 4 implies that the causal effect of $\mathbf{V}\setminus\mathbf{W}$ on $\mathbf{Y}$ is not g-identifiable from $(\mathbb{A}, \mathcal{G})$. Hence, Equation (3) implies that $P_{\mathbf{x}}(\mathbf{y})$ cannot be uniquely computed and the causal effect of $\mathbf{X}$ on $\mathbf{Y}$ is not g-identifiable from $(\mathbb{A}, \mathcal{G})$. □

Proposition 4. *Suppose* $\mathbf{S} \subseteq \mathbf{V}$ *and* $\mathbf{S}_1, \cdots, \mathbf{S}_l$ *are the c-components of* $\mathbf{S}$. $Q[\mathbf{S}]$ *is g-identifiable from* $(\mathbb{A}, \mathcal{G})$ *if and only if* $Q[\mathbf{S}_i]$ *is g-identifiable from* $(\mathbb{A}, \mathcal{G})$ *for each* $i \in [1:l]$.

Proof. Sufficient part: Suppose $Q[\mathbf{S}_i]$ is g-identifiable from $(\mathbb{A}, \mathcal{G})$ for each $i \in [1:l]$. Tian and Pearl [2003] showed that

$$Q[\mathbf{S}] = \prod_{i=1}^{l} Q[\mathbf{S}_i].$$

Hence, $Q[\mathbf{S}]$ is uniquely computable and therefore, g-identifiable from $(\mathbb{A}, \mathcal{G})$.

Necessary part: Suppose $Q[\mathbf{S}]$ is g-identifiable from $(\mathbb{A}, \mathcal{G})$. For $i \in [1:l]$, Tian and Pearl [2003] provided a formula for computing $Q[\mathbf{S}_i]$ from $Q[\mathbf{S}]$ (Lemma 4, Equations (71) and (72) in [Tian and Pearl, 2003]). Hence, for each $i \in [1:l]$, $Q[\mathbf{S}]$ is uniquely computable and therefore, g-identifiable from $(\mathbb{A}, \mathcal{G})$. □

2.3 PROOFS OF SECTION 5

Lemma 2. *If $Q[\mathbf{S}]$ is not g-identifiable from $(\mathbb{A}', \mathcal{G}')$, then $Q[\mathbf{S}]$ is not g-identifiable from $(\mathbb{A}, \mathcal{G})$.*

Proof. If $Q[\mathbf{S}]$ is not g-identifiable from $(\mathbb{A}', \mathcal{G}')$, then there exists two models $\mathcal{M}'_1$ and $\mathcal{M}'_2$ in $\mathbb{M}^+(\mathcal{G}')$ such that for each $i \in [0:m]$ and any $\mathbf{v} \in \mathfrak{X}_{\mathbf{V}'}$,

$$Q^{\mathcal{M}'_1}[\mathbf{A}'_i](\mathbf{v}) = Q^{\mathcal{M}'_2}[\mathbf{A}'_i](\mathbf{v}),$$

and there exists $\mathbf{v}_0 \in \mathfrak{X}_{\mathbf{V}'}$ such that

$$Q^{\mathcal{M}'_1}[\mathbf{S}](\mathbf{v}_0) \neq Q^{\mathcal{M}'_2}[\mathbf{S}](\mathbf{v}_0).$$

Next, we will construct two models $\mathcal{M}_1$ and $\mathcal{M}_2$ in $\mathbb{M}^+(\mathcal{G})$ to prove that $Q[\mathbf{S}]$ is not g-identifiable from $(\mathbb{A}, \mathcal{G})$. We define the domains of variables in $\mathbf{V}'$ in the model $\mathcal{M}_i$ similar to model $\mathcal{M}'_i$, for $i \in \{1, 2\}$. Since for each node $V \in \mathbf{V}'$, we have $Pa_{\mathcal{G}'}(V) \subseteq Pa_{\mathcal{G}}(V)$, then for all $V \in \mathbf{V}'$ and $i \in \{1, 2\}$, we can define:

$$P^{M_i}(V|Pa_{\mathcal{G}}(V)) := P^{M'_i}(V \mid Pa_{\mathcal{G}'}(V)).$$

And for $V \in \mathbf{V} \setminus \mathbf{V}'$, we define:

$$\mathfrak{X}_V = \{0\}, \quad P(V = 0) = 1.$$

Because variable $V \in \mathbf{V} \setminus \mathbf{V}'$ can only take value 0 with probability one, then $Q^{\mathcal{M}_j}[\mathbf{A}_i](\mathbf{v}) = Q^{\mathcal{M}'_j}[\mathbf{A}'_i](\mathbf{v})$ for all i and $Q^{\mathcal{M}_j}[\mathbf{S}](\mathbf{v}_0) = Q^{\mathcal{M}'_j}[\mathbf{S}](\mathbf{v}_0)$ for $j \in \{1, 2\}$. Thus, we have

$$\begin{aligned} &Q^{\mathcal{M}_1}[\mathbf{A}_i](\mathbf{v}) = Q^{\mathcal{M}_2}[\mathbf{A}_i](\mathbf{v}), \quad i \in [0:m], \\ &Q^{\mathcal{M}_1}[\mathbf{S}](\mathbf{v}_0) \neq Q^{\mathcal{M}_2}[\mathbf{S}](\mathbf{v}_0). \end{aligned}$$

This shows that $Q[\mathbf{S}]$ is not g-identifiable from $(\mathbb{A}, \mathcal{G})$. □

Lemma 3. *Consider the following set of vectors in $\mathbb{R}^d$*

$$\mathbf{\Omega} := \{\theta_i(\mathbf{v}) : \ i \in [0:m], \mathbf{v} \in \mathfrak{X}_{\mathbf{V}}\} \cup \mathbb{1}_d, \tag{4}$$

where $\mathbb{1}_d$ denotes the all-ones vector in $\mathbb{R}^d$. If there exists $\mathbf{v}_0 \in \mathfrak{X}_{\mathbf{V}}$ such that $\eta(\mathbf{v}_0)$ is linearly independent from all the vectors in $\mathbf{\Omega}$, then the system of linear equations in (11) *admits a solution.*

Proof. This is a direct consequence of Lemma 5 with $\{c_i\}$ to be $\mathbf{\Omega}$ and c to be $\eta(\mathbf{v}_0)$. □

Lemma 4. *The SEM constructed above belongs to $\mathbb{M}^+(\mathcal{G}')$.*

Proof. By the construction, it is clear that the model belongs to $\mathbb{M}(\mathcal{G}')$. Hence, we need to show that $P(\mathbf{v}) > 0$ for any $\mathbf{v} \in \mathfrak{X}_{\mathbf{V}'}$. To this end, it is enough to show that for any realization $\mathbf{v} \in \mathfrak{X}'_{\mathbf{V}}$, there exists a realization $\hat{\mathbf{u}} \in \mathfrak{X}_{\mathbf{U}'}$ such that $P(\mathbf{v}, \hat{\mathbf{u}}) > 0$, because in this case we have

$$P(\mathbf{v}) = \sum_{\mathbf{u}=\mathfrak{X}_{\mathbf{U}'}} P(\mathbf{v}, \mathbf{u}) \geq P(\mathbf{v}, \hat{\mathbf{u}}) > 0.$$

Let $\mathbf{v}$ be a fixed realization in $\mathfrak{X}_{\mathbf{V}'}$. For the rest of the proof, we assume all the realizations for $\mathbf{V}'$ are consistent with $\mathbf{v}$.

By Markov factorization property, for any $\mathbf{u} \in \mathfrak{X}_{\mathbf{U}'}$ we have

$$P(\mathbf{v}, \mathbf{u}) = \prod_{V \in \mathbf{V}'} P(v \mid Pa_{\mathcal{G}'}(V)) \prod_{U \in \mathbf{U}'} P(u). \tag{5}$$

By the construction of our model, we have $P(u) > 0$ for any $U \in \mathbf{U}'$ and $u \in \mathfrak{X}_U$. Moreover, for any $X \in \mathbf{S}$ and any realization for $Pa_{\mathcal{G}'}(X) \cap \mathbf{U}'$ we have $P(x \mid Pa_{\mathcal{G}'}(X)) > 0$. Hence, it is enough to show that there exists $\hat{\mathbf{u}} \in \mathfrak{X}_{\mathbf{U}'}$ such that $P(x \mid Pa_{\mathcal{G}'}(X)) > 0$ for each $X \in \mathbf{T}$.

Recall that for each $X \in \mathbf{T}$, we have $X = (X[i_1], \cdots, X[i_{\alpha(X)}])$, where X belongs to $\mathcal{F}_{i_1}, \cdots, \mathcal{F}_{i_{\alpha(X)}}$ and

$$X[i_j] \equiv \left(\sum_{Y \in Pa_{\mathcal{F}_{i_j}}(X)} Y[i_j] \right) \pmod 2.$$

By the construction, we define the entries corresponding to each $\mathcal{F}_i$ separately. For each $i \in [0:k]$, let $\mathbf{U}_i$ to be the set of unobserved variables in $\mathbf{U^T}$ that are in $\mathcal{F}_i$.

Let us fix an $i \in [0:k]$. To finish the prove, we will introduce a method to determine $\hat{u}[i]$ for each $U \in \mathbf{U}_i$ such that

$$x[i] \equiv \left(\sum_{Y \in Pa_{\mathcal{F}_i}(X)} y[i] \right) \pmod 2, \tag{6}$$

for each $X \in \mathbf{T} \cap \mathbf{B}_i$.

Lets start with an arbitrary set of values for $\{\hat{u}[i] \colon U \in \mathbf{U}_i\}$ which are either 0 or 1. Suppose $X \in \mathbf{T} \cap \mathbf{B}_i$. We introduce a trick such that $x[i]$ will be replaced by $1 - x[i]$ while for all $Y \in \mathbf{T} \cap \mathbf{B}_i$, $y[i]$ remains the same:
By the construction of $\mathcal{F}_i$, there exists a path $(X = X_1, U_1, X_2, \cdots, X_l, U_l, Z = X_{l+1})$ from X to a variable $Z \in \mathbf{S}$ such that $\{U_1, \cdots, U_l\} \subseteq \mathbf{U}_i$, $\{X_1, \cdots, X_l\} \subseteq \mathbf{B}_i \cap \mathbf{T}$, and $Ch_{\mathcal{F}_i}(U_j) = \{X_j, X_{j+1}\}$ for each $j \in [1:l]$. Now for each $j \in [1:l]$, we replace $\hat{u}_j[i]$ by $1 - \hat{u}_j[i]$. Since Equation (6) is in mod 2, the value of $x_j[i]$ will be the same for each $j \in [2:l]$ while $x[i]$ will be replaced by $1 - x[i]$. Note that $X_{l+1} = Z \notin \mathbf{T}$.

With the trick described above, we can construct any realization for the i-th bit of the variables in $\mathbf{T} \cap \mathbf{B}_i$. Hence, we can construct $\hat{\mathbf{u}} \in \mathfrak{X}_{\mathbf{U}'}$ such that $P(x \mid Pa_{\mathcal{G}'}(X)) > 0$ for each $X \in \mathbf{T}$. □

Lemma 5. *For any* $\mathbf{v} \in \mathfrak{X}_{\mathbf{V}'}$ *and* $i \in [0:m]$,

$$\theta_{i,j_1}(\mathbf{v}) = \theta_{i,j_2}(\mathbf{v}) = \cdots = \theta_{i,j_{\frac{\kappa+1}{2}}}(\mathbf{v}).$$

Proof. Lets fix a realization $\mathbf{v}$ for the observed variables $\mathbf{V}'$. Suppose that l_1 and l_2 are two integers such that

$$\begin{aligned} \gamma_{l_1} &= (2x, 0, \ldots, 0)), \\ \gamma_{l_2} &= (2x + 2 \pmod{\kappa + 1}, 0, \ldots, 0), \end{aligned}$$

where x is any fixed integer in $[0 : \frac{\kappa-1}{2}]$. To show the result, we will prove that $\theta_{i,l_1}(\mathbf{v}) = \theta_{i,l_2}(\mathbf{v})$. Let

$$\begin{aligned} f_{i,j}(\mathbf{v}, \mathbf{u^T}) &:= \sum_{\mathbf{u} \in \mathbf{U^S}} \prod_{V \in \mathbf{A}'_i} P(v \mid Pa_{\mathcal{G}'}(V)) \prod_{U \in \mathbf{U}' \setminus \{U_0\}} P(u) \\ &= \prod_{V \in \mathbf{A}'_i \setminus \mathbf{B}_i} P(v \mid Pa_{\mathcal{G}'}(V)) \prod_{V \in \mathbf{B}_i \setminus \mathbf{S}} P(v \mid Pa_{\mathcal{G}'}(V)) \sum_{\mathbf{u} \in \mathbf{U^S}} \prod_{V \in \mathbf{S}} P(v \mid Pa_{\mathcal{G}'}(V)) \prod_{U \in \mathbf{U}' \setminus \{U_0\}} P(u). \end{aligned}$$

where index j indicates $U_0 = \gamma_j$. Note that variable U_0 may appear in the parent set of some observed variables. Using the above definition, we have

$$\theta_{i,j}(\mathbf{v}) = \sum_{\mathbf{u^T} \in \mathbf{U^T}} f_{i,j}(\mathbf{v}, \mathbf{u^T}).$$

Hence, if we show $f_{i,l_1}(\mathbf{v}, \mathbf{u^T}) = f_{i,l_2}(\mathbf{v}, \mathbf{u^T})$ for any fixed realization $\mathbf{u^T}$, the above equation implies $\theta_{i,l_1}(\mathbf{v}) = \theta_{i,l_2}(\mathbf{v})$.

When $T \in \mathbf{A}'_i \setminus \mathbf{B}_i$, then for fixed realizations of $\mathbf{u^T}$, $P(t|Pa_{\mathcal{G}'}(T))$ is the same for both realizations γ_{l_1} and γ_{l_2} since $\gamma_{l_1} \equiv \gamma_{l_2} \bmod 2$.

When $T \in \mathbf{B}_i \setminus \mathbf{S}$, unobserved variables in $Pa_{\mathcal{G}'}(T)$ are a subset of $\mathbf{U^T} \cup \{U_0\}$. Note that in the definition of $f_{i,j}(\mathbf{v}, \mathbf{u^T})$, all such unobserved variables are fixed. Thus, if there exists $T \in \mathbf{B}_i \setminus \mathbf{S}$, such that $P(t|Pa_{\mathcal{G}'}(T)) = 0$, then

$$f_{i,l_1}(\mathbf{v}, \mathbf{u^T}) = f_{i,l_2}(\mathbf{v}, \mathbf{u^T}) = 0.$$

When $P(t|Pa_{\mathcal{G}'}(T)) = 1$ for all $T \in \mathbf{B}_i \setminus \mathbf{S}$, to prove $f_{i,l_1}(\mathbf{v}, \mathbf{u}^{\mathbf{T}}) = f_{i,l_2}(\mathbf{v}, \mathbf{u}^{\mathbf{S}})$, we show that for any realization $(\mathbf{u}_1, \gamma_{l_1})$ of $(\mathbf{U}^{\mathbf{S}}, U_0)$, there is a realization $(\mathbf{u}_2, \gamma_{l_2})$ of $(\mathbf{U}^{\mathbf{S}}, U_0)$ such that

$$\prod_{V \in \mathbf{S}} P(v \mid Pa_{\mathcal{G}'}(V))\Big|_{(\mathbf{U}^{\mathbf{S}}, U_0) = (\mathbf{u}_1, \gamma_{l_1})} = \prod_{V \in \mathbf{S}} P(v \mid Pa_{\mathcal{G}'}(V))\Big|_{(\mathbf{U}^{\mathbf{S}}, U_0) = (\mathbf{u}_2, \gamma_{l_2})},$$

where $P(v \mid Pa_{\mathcal{G}'}(V))\Big|_{(\mathbf{U}^{\mathbf{S}}, U_0) = (\mathbf{u}_1, \gamma_{l_1})}$ denotes the conditional probability of v given its parents in which the unobserved variables $(\mathbf{U}^{\mathbf{S}}, U_0)$ are fixed to be $(\mathbf{u}_1, \gamma_{l_1})$. To this end, we consider two cases depending on i.

First case, when $i \in [0 : k]$: In this case, we have

$$t[i] = \left(\sum_{Y \in Pa_{\mathcal{F}_i}(T)} y[i] \right) \pmod 2. \tag{7}$$

Consider the set $\mathbf{\Lambda} := Pa_{\mathcal{F}_i}(\mathbf{S}) \setminus Pa_{\mathcal{F}_i[\mathbf{S}]}(\mathbf{S})$, that is the set of all parents of nodes in $\mathbf{S}$ that are outside of $\mathbf{S}$. By the construction of our models, summation of the values of the observed and unobserved nodes in $\mathbf{\Lambda}$ are the same, i.e.,

$$\sum_{W \in \mathbf{\Lambda} \cap \mathbf{B}_i} w[i] \equiv \sum_{W \in \mathbf{\Lambda} \cap \mathbf{U}'} w[i] \pmod 2,$$

or equivalently

$$\sum_{W \in \mathbf{\Lambda}} w[i] \equiv 0 \pmod 2. \tag{8}$$

This is because, in graph $\mathcal{F}_i$, each observed variable outside of $\mathbf{S}$ has at most one child outside of $\mathbf{S}$, and each unobserved node has either one or two children outside of $\mathbf{S}$. According to (7), those unobserved nodes with two children outside of $\mathbf{S}$ do not belong to $\mathbf{\Lambda} \cap \mathbf{U}'$. Such unobserved nodes have exactly two observed descendants in $\mathbf{\Lambda} \cap \mathbf{B}_i$, and because both descendants appear in (8), their summation is zero mod 2. On the other hand, the unobserved nodes with only one child outside of $\mathbf{S}$ belong to $\mathbf{\Lambda} \cap \mathbf{U}'$ and have exactly one observed descendant in $\mathbf{\Lambda} \cap \mathbf{B}_i$. Thus, the summation of such unobserved variables and their observed descendant is again zero mod 2 in (8).

If $\mathbb{I}(S) = 0$ for all $S \in \mathbf{S}$, then by our model construction, for any variable $W \in \mathbf{\Lambda} \setminus \{T_i\}$, $w[i]$ is an even number but T_i takes value 1 with probability one. Hence, the summation in (8) cannot be an even number. Therefore, there exists at least a variable $S \in \mathbf{S}$ such that $\mathbb{I}(S) = 1$. In this case, the value of $P(S|Pa_{\mathcal{G}'}(S))$ does not depend on the realizations of variables in $\mathbf{U}^{\mathbf{S}}$. Next, we show that for any realization $\mathbf{u}_1$ of $\mathbf{U}^{\mathbf{S}}$, there is a realization $\mathbf{u}_2$ such that

$$P(s|Pa_{\mathcal{G}'}(S))\Big|_{(\mathbf{U}^{\mathbf{S}}, U_0) = (\mathbf{u}_1, \gamma_{l_1})} = P(s|Pa_{\mathcal{G}'}(S))\Big|_{(\mathbf{U}^{\mathbf{S}}, U_0) = (\mathbf{u}_2, \gamma_{l_2})}. \tag{9}$$

Since $\mathcal{G}'_{\mathbf{S}}$ is a c-component, there exists a sequence of variables $U_0, \hat{S}_1, \hat{U}_1, \hat{S}_2, \hat{U}_2, \ldots, \hat{U}_l, S$, such that U_0 is a parent of $\hat{S}_1$, S is a children of $\hat{U}_l$ and $\hat{U}_j$ is a parent of variables $\hat{S}_j$ and $\hat{S}_{j+1}$ for $j \in [1 : l-1]$. Let $\hat{\mathbf{U}} := \{\hat{U}_1, \ldots, \hat{U}_l\}$. For realization $\mathbf{u}_1$, we define $\mathbf{u}_2$ by

$$\begin{aligned} &u_{2,\hat{U}_j} := u_{1,\hat{U}_j} + 2(-1)^j \pmod{\kappa + 1}, \quad j \in [1 : l], \\ &u_{2,U} := u_{1,U}, \quad \forall U \in \mathbf{U}' \setminus (\hat{\mathbf{U}} \cup \{U_0\}), \end{aligned} \tag{10}$$

where $u_{2,U}$ denotes the realization for variable U in $\mathbf{u}_2$. It is straightforward to see that this mapping is a bijection between $\mathbf{u}_1$ and $\mathbf{u}_2$ and (9) holds.

Second case, when $i \in [k+1 : m]$: In this case, $\mathbf{S} \setminus \mathbf{A}'_i \neq \varnothing$. Since $\mathcal{G}'_{\mathbf{S}}$ is a c-component, there exists a sequence of variables $U_0, \hat{S}_1, \hat{U}_1, \hat{S}_2, \hat{U}_2, \ldots, \hat{U}_l, S$, such that U_0 is a parent of $\hat{S}_1$, $S \in \mathbf{S} \setminus \mathbf{A}'_i$ is a children of $\hat{U}_l$ and $\hat{U}_j$ is a parent of variables $\hat{S}_j$ and $\hat{S}_{j+1}$ for $j \in [1 : l-1]$. Let $\hat{\mathbf{U}} := \{\hat{U}_1, \ldots, \hat{U}_l\}$. Similar to the previous case, for a given realization $\mathbf{u}_1$ of $\mathbf{U}^{\mathbf{S}}$, we define $\mathbf{u}_2 \in \mathfrak{X}_{\mathbf{U}^{\mathbf{S}}}$ by

$$\begin{aligned} &u_{2,\hat{U}_j} := u_{1,\hat{U}_j} + 2(-1)^j \pmod{\kappa + 1}, \quad j \in [1 : l], \\ &u_{2,U} := u_{1,U}, \quad \forall U \in \mathbf{U}' \setminus (\hat{\mathbf{U}} \cup \{U_0\}), \end{aligned} \tag{11}$$

where $u_{2,\hat{U}}$ denotes the realization for variable U in $\mathbf{u}_2$. Analogous to the previous setting, we have (9).

Herein, we proved that $\theta_{i,l_1}(\mathbf{v}) = \theta_{i,l_2}(\mathbf{v})$. By varying x within $[0 : \frac{\kappa-1}{2}]$ in the definition of γ_{l_1} and γ_{l_2}, we conclude the lemma. □

Lemma 6. *There exists* $0 < \epsilon < \frac{1}{\kappa}$ *such that there exists* $\mathbf{v}_0 \in \mathfrak{X}_{\mathbf{V}'}$ *and* $1 \leq r < t \leq \frac{\kappa+1}{2}$ *such that*

$$\eta_{j_r}(\mathbf{v}_0) \neq \eta_{j_t}(\mathbf{v}_0).$$

Proof. Lets consider r and t such that $\gamma_r = (0, 0, \ldots, 0)$ and $\gamma_t = (2, 0, \ldots, 0)$. Recall that:

$$\eta_r(\mathbf{v}) := \sum_{\mathbf{U}' \setminus \{U_0\}} \prod_{X \in \mathbf{S}} P(x \mid Pa_{\mathcal{G}'}(X))\Big|_{U_0=\gamma_r} \prod_{U \in \mathbf{U}' \setminus \{U_0\}} P(u), \tag{12}$$

$$\eta_t(\mathbf{v}) := \sum_{\mathbf{U}' \setminus \{U_0\}} \prod_{X \in \mathbf{S}} P(x \mid Pa_{\mathcal{G}'}(X))\Big|_{U_0=\gamma_t} \prod_{U \in \mathbf{U}' \setminus \{U_0\}} P(u). \tag{13}$$

We choose $\mathbf{v}_0$ as follows: set all variables in $\mathbf{S}$ to be zero and select a realization for variables in $\mathbf{V}' \setminus \mathbf{S}$ such that $\mathbb{I}(S) = 0$ for all $S \in \mathbf{S}$. Denote by S_0 a child of U_0 in $\mathbf{S}$.

Note that there is a term in the summation of the right side of equation (12) that is $(1 - \kappa\epsilon)^{|\mathbf{S}|}$. For instance, this occurs when all realizations of unobserved variables in $\mathbf{U}^{\mathbf{S}}$ are zero.

Next, we prove that there is no realization of unobserved variables $\mathbf{U}^{\mathbf{S}}$ such that $P(S|Pa_{\mathcal{G}'}(S)) = 1 - \epsilon\kappa$ for all $S \in \mathbf{S}$ and $U_0 = \gamma_t$. In other words, each term in the summation of (13) has at least a term ϵ. To do so, it suffices to show that there is no realization of $\mathbf{U}^{\mathbf{S}}$ such that:

$$s = \sum_{W \in Pa_{\mathcal{G}'[\mathbf{S}]}(S)} w, \quad S \in \mathbf{S} \setminus \{S_0\},$$

$$s_0 = u_0[0] + \sum_{W \in Pa_{\mathcal{G}'[\mathbf{S}]}(S)} w.$$

Suppose there is a realization of $\mathbf{U}^{\mathbf{S}}$ such that the above equations hold. In this case, since $\mathcal{G}'_{\mathbf{S}}$ is a tree, we can color its nodes with two colors, red and black, such that connected nodes by biderected edges have different colors. Suppose that $\mathbf{S}_1$ is the set of black variables and $\mathbf{S}_2$ is the set of red variables which (without loss of generality) contains $S_0 \in \mathbf{S}_1$. Then:

$$\sum_{W \in \mathbf{S}_1} w \equiv u_0[0] + \sum_{U \in \mathbf{U}^{\mathbf{S}}} u \pmod{\kappa + 1},$$

$$\sum_{W \in \mathbf{S}_2} w \equiv \sum_{U \in \mathbf{U}^{\mathbf{S}}} u \pmod{\kappa + 1}.$$

The left-hand sides of both above equations are zero because of our choice of $\mathbf{v}_0$. However, the right-hand sides cannot be the same since $u_0[0] = 2$. Hence, in Equation (13), there exists a term in the summation with probability ϵ. Therefore, in extreme case, when $\epsilon = 0$, $\eta_t(\mathbf{v}') = 0$. However, $\eta_r(\mathbf{v}') \geq (1 - \kappa\epsilon)^{|\mathbf{S}|} \prod_{U \in \mathbf{U}' \setminus \{U_0\}} P(u) > 0$. Since $\eta_r(v)$ and $\eta_t(v)$ are polynomial functions of ϵ and they are not equal at $\epsilon = 0$, then there exists a small enough $0 < \epsilon < \frac{1}{\kappa}$ such that $\eta_r(\mathbf{v}') \neq \eta_t(\mathbf{v}')$. □

3 A SPECIAL CASE IN THE PROOF OF THEOREM 1

In this section, we provide our proof for the necessary part of Theorem 1 when $\mathbf{S} \not\subseteq \mathbf{A}'_i$ for all $i \in [0 : m]$.

We define $\mathcal{F}^{\mathbf{S}}$ to be a minimal (in terms of edges) spanning subgraph of $\mathcal{G}[\mathbf{S}]$ such that $\mathcal{F}^{\mathbf{S}}_{\mathbf{S}}$ is a single c-component. In this case, we can assume $\mathbf{V}' = \mathbf{S}$, $\mathcal{G}'$ is $\mathcal{F}^{\mathbf{S}}$, and $\mathbb{A}' = \{\mathbf{A}'_i := \mathbf{A}_i \cap \mathbf{V}'\}_{i=0}^{m}$. For each $i \in [0 : m]$, we have $\mathbf{A}'_i \subsetneq \mathbf{V}'$. Note that Lemma 2 holds for this case. Hence, it is enough to show that $Q[\mathbf{S}]$ is not g-identifiable from $(\mathbb{A}', \mathcal{G}')$.

Recall that our assumptions and goal in this section are as follows:
$\mathcal{G}'$ is a DAG with observed variables $\mathbf{V}'$ and unobserved variables $\mathbf{U}'$ such that $\mathcal{G}'_{\mathbf{V}'}$ has no directed edges and its bidirected

edges form a spanning tree over $\mathbf{V}'$. $\mathbb{A}' = \{\mathbf{A}'_i\}_{i=0}^m$ is a collection of subsets such that $\mathbf{A}'_i \subsetneq \mathbf{V}'$. The goal is to show that $Q[\mathbf{V}']$ is not g-identifiable from $(\mathbb{A}', \mathcal{G}')$.

For this case we will define two model $\mathcal{M}_1$ and $\mathcal{M}_2$ such that for each $i \in [0:m]$ and any $\mathbf{v} \in \mathfrak{X}_{\mathbf{V}'}$,

$$Q^{\mathcal{M}_1}[\mathbf{A}'_i](\mathbf{v}) = Q^{\mathcal{M}_2}[\mathbf{A}'_i](\mathbf{v}),$$

but there exists $\mathbf{v}_0 \in \mathfrak{X}_{\mathbf{V}'}$ such that

$$Q^{\mathcal{M}_1}[\mathbf{S}](\mathbf{v}_0) \neq Q^{\mathcal{M}_2}[\mathbf{S}](\mathbf{v}_0).$$

For both models $\mathcal{M}_1$ and $\mathcal{M}_2$ we define each observed and unobserved variable to be binary, i.e $\mathfrak{X}_W = \{0, 1\}$ for all $W \in \mathbf{V}' \cup \mathbf{U}'$. Next, we define the equation of the variables in each model.

Model 1: For $V \in \mathbf{V}'$:

$$V = \begin{cases} \bigoplus Pa_{\mathcal{G}'}(V), & \text{with probability } 1-\epsilon, \\ 1, & \text{with probability } \frac{\epsilon}{2}, \\ 0, & \text{with probability } \frac{\epsilon}{2}, \end{cases} \tag{14}$$

and for $U \in \mathbf{U}'$:

$$P(U=0) = P(U=1) = 0.5.$$

Model 2: Suppose V_1 is a fixed observed variable in $\mathbf{V}'$. Then, for all V in $\mathbf{V}' \setminus \{V_1\}$ we define:

$$V = \begin{cases} \bigoplus Pa_{\mathcal{G}'}(V), & \text{with probability } 1-\epsilon \\ 1, & \text{with probability } \frac{\epsilon}{2}, \\ 0, & \text{with probability } \frac{\epsilon}{2}, \end{cases} \tag{15}$$

and for V_1:

$$V_1 = \begin{cases} \neg \bigoplus Pa_{\mathcal{G}'}(V_1), & \text{with probability } 1-\epsilon \\ 1, & \text{with probability } \frac{\epsilon}{2}, \\ 0, & \text{with probability } \frac{\epsilon}{2}, \end{cases} \tag{16}$$

where $\neg$ denotes the logical not. Similar to the first mode, for each unobserved variables $U \in \mathbf{U}'$,

$$P(U=0) = P(U=1) = 0.5.$$

Lemma 11. *Let $i \in [0, m]$ and denote the cardinality of $\mathbf{A}'_i$ by n, i.e. $|\mathbf{A}'_i| = n$. Then for any realization $\mathbf{v} \in \mathfrak{X}_{\mathbf{V}'}$:*

$$Q^{\mathcal{M}_1}[\mathbf{A}'_i](\mathbf{v}) = Q^{\mathcal{M}_2}[\mathbf{A}'_i](\mathbf{v}) = \frac{1}{2^n}.$$

Proof. Suppose $\mathbf{A}'_i := \{A_1, A_2, \ldots, A_n\}$. Since $\mathbf{A}'_i \subsetneq \mathbf{V}'$, there are distinct unobserved variables $U_1, U_2, \ldots, U_n$, such that U_j is a parent of the A_j for $j \in [1:n]$. Denote by $\mathcal{M}$ any of the model $\mathcal{M}_1$ or $\mathcal{M}_2$.

Assume that for some realization of observed and unobserved variables, exactly $t \in [0, n]$ variables in $\mathbf{A}'_i$ are defined by the XOR or $\neg XOR$ of their parents. Without loss of generality, assume that these variables are $\{A_1, A_2, \ldots, A_t\}$. If we know all unobserved variables $\mathbf{U}'$ except $\{U_1, U_2, \ldots, U_t\}$, then we can determine uniquely the values of $\{U_1, U_2, \ldots, U_t\}$ from the following equations:

$$A_i = \bigotimes Pa_{\mathcal{G}'}(A_i), \quad i \in [1:t],$$

where $\bigotimes$ denotes the corresponding equation, either XOR or $\neg XOR$, for variable A_i in model $\mathcal{M}$. Thus, by considering all possible realizations of unobserved variables that lead to a realization $\mathbf{v} \in \mathfrak{X}_{\mathbf{V}'}$, we obtain

$$Q[\mathbf{A}'_i] = \sum_{j=0}^{n} C_n^j (1-\epsilon)^j \left(\frac{\epsilon}{2}\right)^{n-j} \left(\frac{1}{2}\right)^j = \left(\frac{1}{2}\right)^n,$$

where C_n^j is the number of different ways to choose j variables out of n, such that with probability $(1-\epsilon)$ their values are determined by either XOR or $\neg XOR$ equation. All other $n-j$ variables are equal to either 0 or 1 with probability $\frac{\epsilon}{2}$. $\square$

Lemma 12. *Let* $\mathbf{v} = \mathbf{0}$ *be the realization of* $\mathbf{V}'$ *such that all observed variables are equal to* 0*. Then* $Q^{\mathcal{M}_1}[\mathbf{V}'](\mathbf{v}) \neq Q^{\mathcal{M}_2}[\mathbf{V}'](\mathbf{v})$.

Proof. Define $n = |\mathbf{V}'|$ and $\mathbf{V}' = \{V_1, V_2, \ldots, V_n\}$. Firstly, we will prove that for any $\mathbf{v} \in \mathfrak{X}_{\mathbf{V}'}$, the value of $Q^{\mathcal{M}_2}[\mathbf{V}'](\mathbf{v})$ does not depend on the position of V_1 in graph $\mathcal{G}'$. Denote by V_2 an observed variable which is connected to the V_1 by a bidirected edge in $\mathcal{G}'_{\mathbf{V}'}$. Let U denotes the unobserved variable (corresponding to the bidirected edge) which is a parent of V_1 and V_2. Next, we define a new model $\mathcal{M}'_2$ in which all variables in $\mathbf{V}'$ are defined similarly as they are defined in model $\mathcal{M}_2$ except for variables V_1 and V_2. In $\mathcal{M}'_2$, we define V_2 in the same way as V_1 is defined in $\mathcal{M}_2$. We also define V_1 in $\mathcal{M}'_2$ in the same way as V_2 is defined in $\mathcal{M}_2$. Then, we have

$$\prod_{i=1}^{n} P^{\mathcal{M}_2}(v_i|Pa_{\mathcal{G}'}(V_i)) = P^{\mathcal{M}_2}(v_1|Pa_{\mathcal{G}'}(V_1))P^{\mathcal{M}_2}(v_2|Pa_{\mathcal{G}'}(V_2))\prod_{i=3}^{n} P(v_i|Pa_{\mathcal{G}'}(V_i))$$
$$= P^{\mathcal{M}'_2}(v_1|Pa_{\mathcal{G}'}(V_1) \setminus \{U\}, u \oplus 1)P^{\mathcal{M}'_2}(v_2|Pa_{\mathcal{G}'}(V_2) \setminus \{U\}, u \oplus 1)\prod_{i=3}^{n} P(v_i|Pa_{\mathcal{G}'}(V_i)).$$

This implies that substituting V_1 by V_2 does not change the value of $Q^{\mathcal{M}_2}[\mathbf{V}'](\mathbf{v})$.

Without loss of generality, suppose that V_1 is a leaf in $\mathcal{G}'$ and U_1 is a parent of V_1. Note that there are exactly $n-1$ unobserved variables in graph $\mathcal{G}'$. This is because $\mathcal{G}'_{\mathbf{V}'}$ is a tree with bidirected edges over $\mathbf{V}'$. Therefore, we have

$$2^{n-1}Q^{\mathcal{M}_1}[\mathbf{V}'](\mathbf{0}) = P^{\mathcal{M}_1}(V_1 = 0|U_1 = 0)\sum_{\mathbf{U}'\setminus\{U_1\}}\prod_{j>1} P(v_j|Pa_{\mathcal{G}'}(V_j)) + P^{\mathcal{M}_1}(V_1 = 0|U_1 = 1)\sum_{\mathbf{U}'\setminus\{U_1\}}\prod_{j>1} P(v_j|Pa_{\mathcal{G}'}(V_j)),$$
$$2^{n-1}Q^{\mathcal{M}_2}[\mathbf{V}'](\mathbf{0}) = P^{\mathcal{M}_2}(V_1 = 0|U_1 = 0)\sum_{\mathbf{U}'\setminus\{U_1\}}\prod_{j>1} P(v_j|Pa_{\mathcal{G}'}(V_j)) + P^{\mathcal{M}_2}(V_1 = 0|U_1 = 1)\sum_{\mathbf{U}'\setminus\{U_1\}}\prod_{j>1} P(v_j|Pa_{\mathcal{G}'}(V_j)).$$

Note that:

$$P^{\mathcal{M}_1}(V_1 = 0|U_1 = 0) = 1 - \frac{\epsilon}{2}$$
$$P^{\mathcal{M}_1}(V_1 = 0|U_1 = 1) = \frac{\epsilon}{2}$$
$$P^{\mathcal{M}_2}(V_1 = 0|U_1 = 0) = \frac{\epsilon}{2}$$
$$P^{\mathcal{M}_2}(V_1 = 0|U_1 = 1) = 1 - \frac{\epsilon}{2}$$

More over, we have

$$\sum_{U_1=0,\mathbf{U}'\setminus\{U_1\}}\prod_{j>1} P(v_j|Pa_{\mathcal{G}'}(V_j)) + \sum_{U_1=1,\mathbf{U}'\setminus\{U_1\}}\prod_{j>1} P(v_j|Pa_{\mathcal{G}'}(V_j)) = Q[\mathbf{V}' \setminus \{V_1\}] = \left(\frac{1}{2}\right)^{n-1}$$

This yields

$$2^{n-1}Q^{\mathcal{M}_1}[\mathbf{V}'](\mathbf{0}) = \left(1 - \frac{\epsilon}{2}\right)a + \frac{\epsilon}{2}b,$$
$$2^{n-1}Q^{\mathcal{M}_2}[\mathbf{V}'](\mathbf{0}) = \left(1 - \frac{\epsilon}{2}\right)b + \frac{\epsilon}{2}a,$$

where

$$a = \sum_{U_1=0,\mathbf{U}'\setminus\{U_1\}}\prod_{j>1} P(V_j = 0|Pa_{\mathcal{G}'}(V_j)),$$
$$b = \sum_{U_1=1,\mathbf{U}'\setminus\{U_1\}}\prod_{j>1} P(V_j = 0|Pa_{\mathcal{G}'}(V_j)).$$

To prove that $Q^{\mathcal{M}_1}[\mathbf{V}'](\mathbf{0}) \neq Q^{\mathcal{M}_2}[\mathbf{V}'](\mathbf{0})$, it is enough to show that $a \neq b$.

Denote by S_n an observed variable connected to the V_1 by a bidirect edge in $\mathcal{G}'_{\mathbf{V}'}$. We define $\mathbf{V}'_{n-1} := \mathbf{V}' \setminus \{V_1\}$, $\mathbf{U}'_{n-1} := \mathbf{U}' \setminus \{U_1\}$ and $\mathcal{G}_{n-1} := \mathcal{G}'[\mathbf{V}' \setminus \{V_1\}]$. We also define models $\mathcal{M}_1^{(n-1)}$ and $\mathcal{M}_2^{(n-1)}$ as follows:

New model $\mathcal{M}_1^{(n-1)}$: For $V \in \mathbf{V}'_{n-1}$:

$$V = \begin{cases} \bigoplus Pa_{\mathcal{G}_{n-1}}(V), & \text{with probability } 1-\epsilon, \\ 1, & \text{with probability } \frac{\epsilon}{2}, \\ 0, & \text{with probability } \frac{\epsilon}{2}, \end{cases} \tag{17}$$

and for $U \in \mathbf{U}'_{n-1}$:

$$P(U=0) = P(U=1) = 0.5.$$

Model $\mathcal{M}_2^{(n-1)}$: For all V in $\mathbf{V}'_{n-1} \setminus \{S_n\}$:

$$V = \begin{cases} \bigoplus Pa_{\mathcal{G}_{n-1}}(V), & \text{with probability } 1-\epsilon \\ 1, & \text{with probability } \frac{\epsilon}{2}, \\ 0, & \text{with probability } \frac{\epsilon}{2}, \end{cases} \tag{18}$$

and for S_n:

$$S_n = \begin{cases} \neg \bigoplus Pa_{\mathcal{G}_{n-1}}(S_n), & \text{with probability } 1-\epsilon \\ 1, & \text{with probability } \frac{\epsilon}{2}, \\ 0, & \text{with probability } \frac{\epsilon}{2}. \end{cases} \tag{19}$$

Similar to the first model, for each unobserved variables $U \in \mathbf{U}'_{n-1}$, we define

$$P(U=0) = P(U=1) = 0.5.$$

Note that:

$$\left(\frac{1}{2}\right)^{n-2} \sum_{U_1=0, \mathbf{U}'_{n-1}} \prod_{j>1} P(V_j | Pa_{\mathcal{G}'}(V_j)) = \left(\frac{1}{2}\right)^{n-2} a = Q^{\mathcal{M}_1^{(n-1)}}[\mathbf{V}'_{n-1}](\mathbf{0}),$$
$$\left(\frac{1}{2}\right)^{n-2} \sum_{U_1=1, \mathbf{U}'_{n-1}} \prod_{j>1} P(V_j | Pa_{\mathcal{G}'}(V_j)) = \left(\frac{1}{2}\right)^{n-2} b = Q^{\mathcal{M}_2^{(n-1)}}[\mathbf{V}'_{n-1}](\mathbf{0}).$$

It remains to show $Q^{\mathcal{M}_1^{(n-1)}}[\mathbf{V}'_{n-1}](\mathbf{0}) \neq Q^{\mathcal{M}_2^{(n-1)}}[\mathbf{V}'_{n-1}](\mathbf{0})$. Note that if this holds, then by our construction, $Q^{\mathcal{M}_1}[\mathbf{V}'](\mathbf{0}) \neq Q^{\mathcal{M}_2}[\mathbf{V}'](\mathbf{0})$. In other words, we could reduce the size of the graph while keeping the same problem. Thus, by continuing this procedure, we eventually reach graph $\mathcal{G}_2$ that consists of only two observed nodes and showing $Q^{\mathcal{M}_1^{(2)}}[\mathbf{V}'_2](\mathbf{0}) \neq Q^{\mathcal{M}_2^{(2)}}[\mathbf{V}'_2](\mathbf{0})$ in that graph will conclude the result. For graph $\mathcal{G}_2$, we have

$$Q^{\mathcal{M}_1^{(2)}}[\mathbf{V}'_2](\mathbf{0}) = \left(\frac{\epsilon}{2}\right)^2 + 2\frac{\epsilon}{2}(1-\epsilon)\frac{1}{2} + (1-\epsilon)^2\frac{1}{2},$$
$$Q^{\mathcal{M}_2^{(2)}}[\mathbf{V}'_2](\mathbf{0}) = \left(\frac{\epsilon}{2}\right)^2 + 2\frac{\epsilon}{2}(1-\epsilon)\frac{1}{2}.$$

This clearly shows that $Q^{\mathcal{M}_1^{(2)}}[\mathbf{V}'_2](\mathbf{0}) \neq Q^{\mathcal{M}_2^{(2)}}[\mathbf{V}'_2](\mathbf{0})$. □