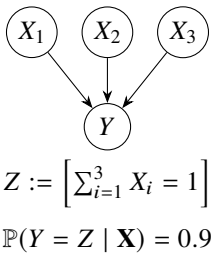
High-Order Markov Blanket Discovery via a k-Order Relaxation of the Faithfulness Assumption

Loong Kuan Lee[1] **Ragavi Krishnamoorthy**[1] **Nico Piatkowski**[1]

[1]Hybrid Intelligence, Fraunhofer IAIS, Sankt Augustin, Germany,

Abstract

The problem of learning the graphical Markov blanket (MB) of a variable from data has applications in many areas such as structure learning for Bayesian networks and Markov random fields, causal discovery, and feature selection. However, a common assumption most methods make is that the conditional independencies in the distribution imply the same separation in the graphical structure—also known as the faithfulness assumption. Unfortunately, this assumption can be violated by higher-order dependencies such as XOR and parity-type relations, and—on finite samples—by empirical violations that, in extreme cases, even induce spurious dependencies absent from the true distribution. Therefore, in this paper we propose a "k-order" relaxation of the faithfulness assumption that captures parity type relationships between k+2 variables. We then propose a proof of concept algorithm called k-order Markov blanket (kOMB) that uses this relaxation for MB discovery. Finally, we empirically show how kOMB can recover the MB of a variable under both true and empirical violations of faithfulness. Code available at: `https://github.com/lklee9/k-order-Markov-blanket`.



Recall (%)

Method	Sample Size 100	1000
Grow and Shrink	28.33	100
0-OMB (ours)	28.33	100
1-OMB (ours)	77.50	100
2-OMB (ours)	**100.0**	100

Figure 1: Recovering the MB of variable Y when Y is the output of a noisy Boolean function that returns 1 when only one of X_1, X_2, and X_3 is 1. Our algorithm that assumes a 2-order faithfulness relaxation is able to fully recover the MB of each variable—even with just 100 samples.

1 INTRODUCTION

The goal of the Markov blanket (MB) discovery problem is to infer the MB of some variable Y in the graph $\mathcal{G}$ given a dataset $\mathcal{D}$ sampled from the joint distribution $\mathbb{P}$. Originating from the Bayesian network and causal learning literature [Margaritis and Thrun, 1999, Tsamardinos et al., 2003a, Ling et al., 2021, Hu and Zheng, 2024], the MB discovery problem has also been discussed and utilised in the feature selection literature as well [Tsamardinos and Aliferis, 2003, Wang et al., 2020]. More recently, MB discovery has also been employed as a subroutine within distributed, divide-and-conquer approaches to large-scale BN structure learning [Dong et al., 2025]. That said, we are interested in inferring MBs in both Bayesian network (BN) and Markov random field (MRF). Therefore, whenever possible we will try to keep our notation and exposition agnostic to the graph type of $\mathcal{G}$.

Most existing approaches to MB discovery operate under two main assumptions. The first being that any separations in $\mathcal{G}$ imply the same conditional independence in $\mathbb{P}$. The second assumption—known as the faithfulness assumption—goes the opposite direction, that is any conditional independencies in $\mathbb{P}$ imply the same separation in $\mathcal{G}$.

Theoretically, BNs with faithfulness violations have Lebesgue measure zero [Meek, 1995, Spirtes et al., 2001, Boeken et al., 2025]; however these "typicality" results depend on somewhat arbitrary choices, like the choice of σ-ideal [Boeken et al., 2025]. There are well-known examples of some—very simple—relations that violate faithfulness. The most famous of which is the (noisy) XOR function for violations involving three variables [Inazumi et al., 2011, Marx et al., 2021]; which is a special case of the (noisy) parity function whose higher-order faithfulness violations we detail in Example 1.

Proceedings of the 42nd Conference on Uncertainty in Artificial Intelligence (UAI 2026), PMLR 337:3338–3354.

That said, even if unfaithful BNs and MRFs are truly almost non-existent in nature, a lack of sufficient samples can still cause an empirical faithfulness violation [Uhler et al., 2013, Lemeire et al., 2012, Boeken et al., 2025]. For example, the small example in Figure 1 is one such case where the true BN is faithful, but due to a finite sample, empirical faithfulness violations occur—which can be overcome by the algorithm we will propose later in Section 5. A full walkthrough of empirical faithfulness violation can be found later in Example 2 under Section 3. Furthermore, in extreme cases, a limited sample size can cause MB discovery algorithms to falsely discover spurious dependencies absent from the true distribution. We later observe this phenomenon in Figure 4 under Section 6.1 with the low precision exhibited by the different MB discovery methods at low sample sizes.

There are several weaker assumptions compared to faithfulness such as adjacency faithfulness [Spirtes et al., 2001], Pearl-minimality [Pearl, 1988, 2009], SGS-minimality [Spirtes et al., 2001], frugality [Forster et al., 2018], and more recently 2-adjacency faithfulness [Marx et al., 2021]. 2-adjacency faithfulness in particular is quite practical as by limiting faithfulness violations to "Unfaithful Triples", it yields an algorithm with a much smaller search space compared to frugality. However, although the 2-adjacency faithfulness can find XOR-type relations, it is unable to find higher-order parity relationships between more than 3 variables.

Contributions Therefore, motivated by the considerations above, in Section 4 we propose a relaxation of faithfulness for parity-like relationships between $k + 2$ variables: the k-order faithfulness assumption. This relaxation of faithfulness allows us to infer—from sample data—the graphical Markov blanket of a variable under mild assumptions. We show this in Section 5 by implementing a proof-of-concept algorithm called the k-order Markov blanket (kOMB). We then empirically show in Section 6 that kOMB can overcome both true and empirical faithfulness violations and that it outperforms some existing constraint-based methods on well-known benchmark datasets—where we find merit in exploring relationships between 4 variables at a time, and not just 3.

2 PROBABILISTIC GRAPHICAL MODELS AND INDEPENDENCIES

Consider the set of n variables $\boldsymbol{V} = \{V_1, \ldots, V_n\}$ with no other hidden or latent variables. The conditional independencies between these variables—$\mathbf{Y} \perp\!\!\!\perp \mathbf{X} \mid \mathbf{Z}$ for disjoint variables sets $\mathbf{Y}, \mathbf{X}, \mathbf{Z} \subseteq \boldsymbol{V}$—can be represented by either a directed acyclic graph (DAG) [Pearl, 1988] or an undirected graph (UG) Lauritzen [1996]. That said, we will denote both graphs with the tuple $\mathcal{G} = (\boldsymbol{V}, \boldsymbol{E})$ since we are agnostic on the type of graph used for representing these graphical conditional independencies $\mathcal{J}(\mathcal{G})$.

Regardless of the graph type of $\mathcal{G}$, for disjoint subsets $\mathbf{X}, \mathbf{Y}, \mathbf{S} \subseteq \boldsymbol{V}$, we will denote that $\mathbf{X}$ and $\mathbf{Y}$ are separated in $\mathcal{G}$ as, $\mathbf{X} \perp\!\!\!\perp_{\mathcal{G}} \mathbf{Y} \mid \mathbf{S}$. Furthermore, we denote the set of conditional independencies encoded by the graph $\mathcal{G}$—i.e. the independence model of $\mathcal{G}$—as,

$$\langle \mathbf{X}, \mathbf{Y} \mid \mathbf{S} \rangle \in \mathcal{J}(\mathcal{G}) \iff \mathbf{X} \perp\!\!\!\perp_{\mathcal{G}} \mathbf{Y} \mid \mathbf{S}, \tag{1}$$

where $\langle \mathbf{X}, \mathbf{Y} \mid \mathbf{S} \rangle$ is known as an independence statement [Sadeghi, 2017].

In addition to the graphical structure $\mathcal{G}$, we also consider a joint distribution $\mathbb{P}$ over $\boldsymbol{V}$ that might encode its own set of conditional independencies between the variables in $\boldsymbol{V}$,

$$\langle \mathbf{X}, \mathbf{Y} \mid \mathbf{S} \rangle \in \mathcal{J}(\mathbb{P}) \iff \mathbf{X} \perp\!\!\!\perp_{\mathbb{P}} \mathbf{Y} \mid \mathbf{S}, \tag{2}$$

where $\mathbf{X} \perp\!\!\!\perp_{\mathbb{P}} \mathbf{Y} \mid \mathbf{S}$ denotes $\mathbf{X}$ and $\mathbf{Y}$ are probabilistically conditionally independent given $\mathbf{S}$.

So far, $\mathcal{J}(\mathcal{G})$ and $\mathcal{J}(\mathbb{P})$ are two separate collections of conditional independencies in the graph $\mathcal{G}$ and distribution $\mathbb{P}$ respectively. Our base assumption throughout this paper, the *Global Markov Property*, links the two by requiring that every conditional independence encoded by the graph $\mathcal{G}$ also hold in the distribution $\mathbb{P}$. This is a common assumption in the Markov blanket discovery literature Schlüter [2014], Yu et al. [2020], Marx et al. [2021].

Assumption 1 (Global Markov Property)**.** Given a PGM $\mathcal{M} = (\mathcal{G}, \mathbb{P})$, we say that the distribution $\mathbb{P}$ is *Markov* to the graph $\mathcal{G}$ when for all disjoint subsets $\mathbf{X}, \mathbf{Y}, \mathbf{S} \subseteq \boldsymbol{V}$ [Hammersley and Clifford, 1971, Lauritzen and Sadeghi, 2018],

$$\mathbf{X} \perp\!\!\!\perp_{\mathcal{G}} \mathbf{Y} \mid \mathbf{S} \implies \mathbf{X} \perp\!\!\!\perp_{\mathbb{P}} \mathbf{Y} \mid \mathbf{S}. \tag{3}$$

Furthermore, by Equations (1) and (2), the relation in Equation (3) is equivalent to $\mathcal{J}(\mathcal{G}) \subseteq \mathcal{J}(\mathbb{P})$.

Despite their differences, the independence models of $\mathcal{G}$ and $\mathbb{P}$ are both known to obey a set of axioms called the (semi)-graphoid axioms [Pearl et al., 1989] (see Definition 6). As mentioned in Section 1, we are interested in learning the graphical Markov blanket of some $Y \in \boldsymbol{V}$ defined as follows.

Definition 1 (Markov Blanket and Boundary)**.** For a variable $Y \in \boldsymbol{V}$, the MB of Y in $\mathcal{G}$ is any subset $\mathbf{S} \subseteq \mathbf{V} \setminus Y$[1] s.t.

$$Y \perp\!\!\!\perp_{\mathcal{G}} \boldsymbol{V} \setminus (Y \cup \mathbf{S}) \mid \mathbf{S}, \tag{4}$$

or in other words, every other variable in $\boldsymbol{V} \setminus (Y \cup \mathbf{S})$ is graphically separated from Y by $\mathbf{S}$. The Markov boundary of the variable Y—henceforth denoted as $\mathrm{MB}_{\mathcal{G}}(Y)$—is then defined as the minimal Markov blanket of Y [Pearl, 1988].

[1] We allow set operations between variable sets and single variables to implicitly mean the operation between the set and the atomic set of the single variable—for example: $\mathbf{V} \setminus Y := \mathbf{V} \setminus \{Y\}$.

An equivalent definition for the Markov blanket and boundary of Y in $\mathbb{P}$ can be made by just substituting the graph separation term in Equation (4) for conditional independence in $\mathbb{P}$.

Therefore, it is more accurate to say that we wish to learn the Markov boundary of Y and not just its Markov blanket. However, since colloquially the term Markov blanket is commonly used to implicitly refer to the Markov boundary, we shall use the term Markov blanket in such a fashion as well.

Several methods exist for discovering the Markov blanket (MB) of a target variable $Y \in \boldsymbol{V}$ in $\mathcal{G}$ with one of the larger class of methods being constraint-based approaches [Margaritis and Thrun, 1999, Tsamardinos et al., 2003a, Schlüter, 2014, Ling et al., 2019]. These approaches iteratively test conditional independence statements—using data sampled from $\mathbb{P}$—to grow and shrink a candidate MB of Y. However, the MB of Y in the joint distribution $\mathbb{P}$ might not be the same as its MB in the graph $\mathcal{G}$. Following the Global Markov Property in Assumption 1, we at know that the MB of Y in $\mathbb{P}$ is a subset of the MB in $\mathcal{G}$—which this might prevent us from fully recovering the MB in $\mathcal{G}$. Therefore, in order to ensure that the Markov blankets of Y in both $\mathbb{P}$ and $\mathcal{G}$ are equal, a common assumption constraint-based approaches make is that $\mathcal{G}$ is *faithful* to the joint distribution $\mathbb{P}$ as defined in Definition 2 [Spirtes et al., 2001].

Definition 2 (Faithfulness)**.** Given a PGM $\mathcal{M} = (\mathcal{G}, \mathbb{P})$, the graph $\mathcal{G}$ is *faithful* to the joint distribution $\mathbb{P}$ if for all disjoint subsets $\mathbf{X}, \mathbf{Y}, \mathbf{S} \subseteq \boldsymbol{V}$ [Spirtes et al., 2001],

$$\mathbf{X} \perp\!\!\!\perp_{\mathbb{P}} \mathbf{Y} \mid \mathbf{S} \implies \mathbf{X} \perp\!\!\!\perp_{\mathcal{G}} \mathbf{Y} \mid \mathbf{S} \tag{5}$$

which is basically the reverse direction of the global Markov property in Assumption 1.

Therefore, by assuming both the Global Markov Property faithfulness, the conditional independencies of both $\mathcal{G}$ and $\mathbb{P}$ then coincide, $\mathcal{J}(\mathcal{G}) = \mathcal{J}(\mathbb{P})$. However faithfulness can be violated in many ways and we will explore some of them in the following section.

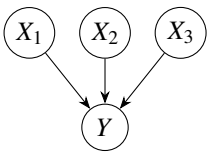
3 FAITHFULNESS VIOLATIONS AND EXISTING RELAXATIONS

Probably the most well-known example of a faithfulness violation is the (noisy) XOR function $Y := X_1 \oplus X_2$ with the graphical structure $X_1 \to Y \leftarrow X_2$ [Inazumi et al., 2011, Marx et al., 2021][2]. A recent faithfulness relaxation for tackling XOR-type relations that is of particular interest to us is 2-adjacency faithfulness [Marx et al., 2021].

Definition 3 (2-adjacency faithfulness)**.** Given the PGM $\mathcal{M} = (\mathcal{G}, \mathbb{P})$, $\mathcal{G}$ is 2-adjacent faithful to $\mathbb{P}$ if for all adjacent variables $(X, Y) \in \boldsymbol{E}$, $\exists \mathbf{X} \subseteq \mathrm{MB}_{\mathcal{G}}(Y)$ where $X \in \mathbf{X}$ s.t. $\forall X \in \mathbf{X} : X \not\perp\!\!\!\perp_{\mathbb{P}} Y \mid \mathbf{X} \setminus X$ and if $|\mathbf{X}| = 2$, $X_1 \not\perp\!\!\!\perp_{\mathbb{P}} X_2 \mid Y$.

[2]The (noisy) XOR function is a special case of the (noisy) parity function whose faithfulness violation we detail in Example 1.

X_1	X_2	X_3	$\mathbb{P}_Y(1)$
0	0	0	0.1
0	0	1	0.9
0	1	0	0.9
0	1	1	0.1
1	0	0	0.9
1	0	1	0.1
1	1	0	0.1
1	1	1	0.9



$\forall i \in [3]$	
$\mathbb{P}_{X_i}(0)$	$\mathbb{P}_{X_i}(1)$
0.5	0.5

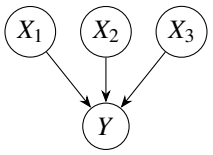
Figure 2: Example of a noisy parity BN over 4 variables.

That said XOR-type relationships only involve three variables. Since we are interested in higher-order unfaithfulness, we will use the noisy parity function as a primary example of the type of faithfulness violation we are interested in.

Example 1 (Noisy Parity Function). Let $\mathcal{M} = (\mathcal{G}, \mathbb{P})$ be a PGM where $\mathbf{X} = \{X_1, X_2, X_3\}$, $\boldsymbol{V} = Y \cup \mathbf{X}$, and $\mathcal{G}$ has the structure of the DAG in Figure 2. Each $X \in \mathbf{X}$ are independent of each other and has the distribution of a fair coin. The distribution of Y conditioned on $\mathbf{X}$ is then $\mathbb{P}(Y = f(\mathbf{x}) \mid \mathbf{X} = \mathbf{x}) = 0.9$, where, $f(\mathbf{x}) \equiv \sum_{i=1}^{n} x_n \pmod 2$. For any random variable $X \in \mathbf{X}$, we know that $\mathbb{P}(Y, X) = 0.25 = 0.5 \times 0.5 = \mathbb{P}(Y)\mathbb{P}(X)$ for any value Y and X takes. Therefore it must the case that $Y \perp\!\!\!\perp_{\mathbb{P}} X$, but since Y and X are not separable in $\mathcal{G}$, $Y \not\perp\!\!\!\perp_{\mathcal{G}} X$; $\mathcal{G}$ is not faithful to $\mathbb{P}$. Additionally for any distinct random variables $X_i, X_j \in \mathbf{X}$, $\mathbb{P}(Y, X_i, X_j) = 1/8 = 0.5^3 = \mathbb{P}(Y)\mathbb{P}(X_i)\mathbb{P}(X_j)$ for any values of Y, X_i, X_j. Therefore we know that $Y \perp\!\!\!\perp_{\mathbb{P}} \{X_i, X_j\}$, $X_i \perp\!\!\!\perp_{\mathbb{P}} \{Y, X_j\}$, and $X_j \perp\!\!\!\perp_{\mathbb{P}} \{X_i, Y\}$. Since Y is adjacent to all $X \in \mathbf{X}$, but there is no 2-cardinality subsets $\mathbf{X}' \subset \mathbf{X}$ that makes $Y \not\perp\!\!\!\perp_{\mathbb{P}} \mathbf{X}'$; $\mathcal{G}$ is not 2-adjacent faithful to $\mathbb{P}$ following the definition in Marx et al. [2021, Definition 7].

Although we now know of a BN that violates both faithfulness and 2-adjacency faithfulness, these types of relationships between variables might very well not be common in nature. In fact, it has been shown that *unfaithful* BNs has Lebesgue measure zero [Meek, 1995, Spirtes et al., 2001, Boeken et al., 2025]. However, empirical violations of faithfulness due to limited samples can be quite common [Uhler et al., 2013, Lemeire et al., 2012, Boeken et al., 2025]. Therefore in Example 2, we will present a scenario where, even though the true BN is faithful; due to sampling error, the DAG of the BN is not faithful to the empirical distribution obtained from the sample.

Example 2. Similar to Example 1, let $\mathcal{M} = (\mathcal{G}, \mathbb{P})$ be a PGM with random variables $\mathbf{X}$ and Y with the DAG structure in Figure 3. The distribution of Y conditioned on $\mathbf{X}$ is then $\mathbb{P}\big(Y = \big[\sum_{X \in \mathbf{X}} X = 1\big] \;\big|\; \mathbf{X}\big) = 0.9$, where $[\cdot]$ is the Iverson bracket notation for the indicator function. In other words,



$$g(\mathbf{X}) := \left[\sum_{X \in \mathbf{X}} X = 1\right]$$

$$\mathbb{P}\big(Y = g(\mathbf{X}) \mid \mathbf{X}\big) = 0.9$$

Pathological $\mathcal{D}$

X_1	X_2	X_3	Y
0	0	0	0
0	0	1	1
0	1	0	1
0	1	1	0
1	0	0	1
1	0	1	0
1	1	0	0
1	0	0	1

Figure 3: Example of a noisy exactly-1 BN over 4 variables with a pathological sample that causes the DAG to be unfaithful to the empirical distribution.

$Y = 1$ has a probability of 0.9 when only one of the variables in $\mathbf{X}$ is 1. Otherwise, $Y = 0$ has a probability of 0.9.

The marginal distribution of Y is $\mathbb{P}(Y = 1) = 3/8$ since only 3 value-combinations of $\mathbf{X}$ has just a single 1 among them. We then know that $\forall X \in \mathbf{X} : \mathbb{P}(Y = 1, X = 1) = 1/8 \neq (3/8) \times (4/8) = \mathbb{P}(Y = 1)\mathbb{P}(X = 1)$. Therefore, $\forall X \in \mathbf{X} : Y \not\perp\!\!\!\perp_{\mathbb{P}} X$ and consequently, $\mathcal{G}$ *is* faithful to $\mathbb{P}$.

However, it is possible for a sample $\mathcal{D}$ of $\mathbb{P}$ to be pathological in the sense that, the empirical distribution from $\mathcal{D}$ induces a independence between Y and some $X \in \mathbf{X}$. For example, the sample in Figure 3 causes the marginal distribution of Y to be $\hat{\mathbb{P}}(Y) = 1/2$ and the marginal distribution over Y and X_1 to be $\hat{\mathbb{P}}(Y, X_1) = 1/4 = (1/2) \times (1/2) = \hat{\mathbb{P}}(Y)\hat{\mathbb{P}}(X_1)$ for all possible values of Y and X_1. Therefore under this empirical distribution $\hat{\mathbb{P}}$, $Y \perp\!\!\!\perp_{\hat{\mathbb{P}}} X_1$—which empirically violates the faithfulness assumption.

Fortunately in Example 2, the DAG in Figure 3 is still 2-adjacent faithful to the empirical distribution, therefore the full MB is still recoverable by the algorithm outlined in Marx et al. [2021]. However, this is not always the case and as we will see in Section 6.1, sometimes considering higher-order relationships between variables can help better overcome empirical faithfulness violations.

4 THE K-ORDER FAITHFULNESS RELAXATION

In this section we will introduce our generalisation of the faithfulness assumption in Definition 2 for finding conditional dependencies between two variables in $\boldsymbol{V}$ that require considering k additional variables to be found. But before that, we first define in Definition 4 what it means for two distinct variables $Y, X \in \boldsymbol{V}$ to only be conditionally dependent given some k-cardinality subset $\mathbf{Z} \subseteq \boldsymbol{V} \setminus \{Y, X\}, |\mathbf{Z}| = k$.

Definition 4 (k-Order Dependence)**.** Given a PGM $\mathcal{M} = (\mathcal{G}, \mathbb{P})$, let $Y, X \in \mathbf{V}$ be two distinct random variables and $\mathbf{S}, \mathbf{Z} \subseteq \boldsymbol{V} \setminus \{Y, X\}$ be two disjoint subsets. Then we denote the absence of any subset of $\mathbf{S}' \subseteq \mathbf{S}$ that renders Y and X conditionally independent given $\mathbf{Z} \cup \mathbf{S}'$ as

$$asc_{\mathbf{S}}(Y, X \mid \mathbf{Z}) := \forall \mathbf{S}' \subseteq \mathbf{S} : Y \not\perp\!\!\!\perp_{\mathbb{P}} X \mid \mathbf{Z} \cup \mathbf{S}'. \tag{6}$$

Here we say that Y and X are $|\mathbf{Z}|$-order associated w.r.t. $\mathbf{Z}$ over the "separating" set $\mathbf{S}$. Of course under this definition there might be some proper subset $\mathbf{Z}' \subset \mathbf{Z}$ where $asc_{\mathbf{S}}(Y, X \mid \mathbf{Z}')$ still holds. Therefore we say Y and X are $|\mathbf{Z}|$-order *dependent* on $\mathbf{Z}$ only if this is not the case,

$$\begin{aligned} &dep_{\mathbf{S}}(Y, X \mid \mathbf{Z}) \\ &:= asc_{\mathbf{S}}(Y, X \mid \mathbf{Z}) \wedge \forall \mathbf{Z}' \subset \mathbf{Z} : \neg\, asc_{\mathbf{S}}(Y, X \mid \mathbf{Z}'). \end{aligned} \tag{7}$$

With this definition of k-order dependence, we are able to capture parity-like relations over $k + 2$ variables—such as the one in Example 1. Specifically we know that the noisy-parity relationship in Example 1 is 2-order dependent since for distinct variables $A, B \in \{X_1, X_2, X_3, Y\}$, $\mathbf{Z}$ must always contain the other two variables for $asc_{\emptyset}(A, B \mid \mathbf{Z})$ to be true.

Further note that when $k = 1$, our notion of k-order dependence— similar to 2-*association* in Marx et al. [2021, Definition 6]—is able to capture the relationships between *Unfaithful Triples* which includes XOR-type relationships.

There are two aspects of Definition 4 whose purpose might not immediately jump out. The first being that in Equation (6), we require $Y \not\perp\!\!\!\perp_{\mathbb{P}} X \mid \mathbf{Z} \cup \mathbf{S}'$ to be true for all subsets $\mathbf{S}' \subseteq \mathbf{S}$—instead of just $Y \not\perp\!\!\!\perp_{\mathbb{P}} X \mid \mathbf{Z} \cup \mathbf{S}$. This requirement ensures that Y and X are dependent as long as they are conditioned on $\mathbf{Z}$—forcing a distinction between the possible dependants $\mathbf{Z}$ and the possible separators $\mathbf{S}$. The other odd aspect of Definition 4 is that we define both k-order association and dependence over some arbitrary subset $\mathbf{S} \subseteq \boldsymbol{V} \setminus \{Y, X\}$ instead of just the entirety of $\boldsymbol{V} \setminus \{Y, X\}$. This allows us to reason about associations relative to the candidate Markov blanket maintained by our algorithm, which is central to proving in Section 5 that *kOMB* correctly recovers the Markov boundary of some variable $Y \in \boldsymbol{V}$.

Since henceforth we need to reason about subsets with a restricted cardinality; before continuing we will first define some notational shorthand for dealing with such sets.

Definition 5 (Power Sets of Limited Cardinality)**.** Let $\boldsymbol{V}$ be a set of elements and $k \in \mathbb{N} : k \leq |\boldsymbol{V}|$ be some natural number[3] smaller than the cardinality of $\boldsymbol{V}$. Then we use,

$$\mathcal{P}_{=k}(\boldsymbol{V}) := \{\mathbf{Z} \subseteq \boldsymbol{V} : |\mathbf{Z}| = k\}, \tag{8}$$

to denote the set of all subsets of $\boldsymbol{V}$ with cardinality k and,

$$\mathcal{P}_{\leq k}(\boldsymbol{V}) := \{\mathbf{Z} \subseteq \boldsymbol{V} : |\mathbf{Z}| \leq k\}, \tag{9}$$

to denote the set of all the subsets of $\boldsymbol{V}$ with cardinality less than or equal to k.

[3]We adopt the definition of the natural numbers in which 0 is included.

Before introducing our generalisation of the faithfulness assumption to k-order dependencies, we first show in Lemma 1 that k-order associations w.r.t. $\mathbf{Z}$ implies the existence of a k-order dependence w.r.t. some subset of $\mathbf{Z}$.

Lemma 1 (k-Order Dependence Arises from Association)**.** *Given a PGM $\mathcal{M} = (\mathcal{G}, \mathbb{P})$, let $Y, X \in \mathbf{V}$ be distinct random variables and $\mathbf{S}, \mathbf{Z} \subset \boldsymbol{V} \setminus \{Y, X\}$ be disjoint subsets. Then the following holds for all $Y, X, \mathbf{S}, \mathbf{Z}$:*

$$asc_{\mathbf{S}}(Y, X \mid \mathbf{Z}) \implies \exists \mathbf{Z}' \subseteq \mathbf{Z} : dep_{\mathbf{S}}(Y, X \mid \mathbf{Z}'). \quad (10)$$

The proof of Lemma 1 and all other Lemmas in the paper can be found in Appendix B.

Assumption 2 (k-Order Faithfulness)**.** Given the PGM $\mathcal{M} = (\mathcal{G}, \mathbb{P})$, $\mathcal{G}$ is k-order faithful to $\mathbb{P}$ if $\forall \mathbf{S} \subseteq \boldsymbol{V} \setminus \{Y, X\}$,

$$\begin{aligned} &\forall \mathbf{Z} \in \mathcal{P}_{\leq k}(\boldsymbol{V} \setminus \{Y, X\}) \;\; \exists \mathbf{S}' \subseteq \mathbf{S} : Y \perp\!\!\!\perp_{\mathbb{P}} X \mid \mathbf{Z} \cup \mathbf{S}' \\ &\implies Y \perp\!\!\!\perp_{\mathcal{G}} X \mid \mathbf{S}. \end{aligned} \quad (11)$$

Its contrapositive then states that $\forall \mathbf{S} \subseteq \boldsymbol{V} \setminus \{Y, X\}$,

$$\begin{aligned} &Y \not\perp\!\!\!\perp_{\mathcal{G}} X \mid \mathbf{S} \\ &\implies \exists \mathbf{Z} \in \mathcal{P}_{\leq k}(\boldsymbol{V} \setminus \{Y, X\}) : asc_{\mathbf{S}}(Y, X \mid \mathbf{Z}). \end{aligned} \quad (12)$$

However, the definition in Equation (12) allows k to be larger than it needs to be—i.e. k is not strict. Therefore, we say that $\mathcal{G}$ is *strictly* k-order faithful to $\mathbb{P}$ if $\forall \mathbf{S} \subseteq \boldsymbol{V} \setminus \{Y, X\}$,

$$\begin{aligned} &Y \not\perp\!\!\!\perp_{\mathcal{G}} X \mid \mathbf{S} \\ &\implies \exists \mathbf{Z} \in \mathcal{P}_{=k'}(\boldsymbol{V} \setminus \{Y, X\}) : dep_{\mathbf{S}}(Y, X \mid \mathbf{Z}), \end{aligned} \quad (13)$$

where $k' \leq k$. This definition in Equation (13) logically follows from Equation (12) as a result of Lemma 1.

With this new generalisation of faithfulness, we will now present an initial algorithm, *kOMB*, that makes use of our k-order faithfulness assumption for MBs discovery.

5 KOMB: K-ORDER MARKOV BLANKET

Our algorithm for Markov blanket (MB) discovery, *kOMB*, is a modified version of the Grow and Shrink (GS) algorithm by Margaritis and Thrun [1999]. In GS, a candidate MB $\mathbf{S}$ is iteratively grown by testing the conditional independence between Y and variables that are not yet in $\mathbf{S}$—conditioned the current candidate MB $\mathbf{S}$. Therefore, the separating set used in the conditional independence tests change as the GS algorithm progresses.

In *kOMB*, something similar occurs where we will only consider subsets of the candidate MB when trying to find subsets that successfully cause Y and X to be conditionally independent. In other words, instead of considering the dependence between Y and X over all the variables in $\boldsymbol{V}$, $dep_{\boldsymbol{V} \setminus \{Y,X\}}(Y, X \mid \mathbf{Z})$; we will instead consider the association between Y and X over just the current candidate MB $\mathbf{S}$, $asc_{\mathbf{S}}(Y, X \mid \mathbf{Z})$. As a first step, we show that a k-order dependence between Y and X with no separating variables to consider implies that Y is k-order associated with every one of the involved variables.

Lemma 2 (Marginal k-Order Dependence Implies Inter-Association)**.** *Given a PGM $\mathcal{M} = (\mathcal{G}, \mathbb{P})$, let $X, Y \in \boldsymbol{V}$ be distinct variables and $\mathbf{Z} \subseteq \boldsymbol{V} \setminus \{Y, X\}$. Then if Y and X are $|\mathbf{Z}|$-order dependent w.r.t. $\mathbf{Z}$ over the empty separating set, Y is $|\mathbf{Z}|$-order associated with every variable in $\mathbf{Z} \cup X$,*

$$dep_{\emptyset}(Y, X \mid \mathbf{Z}) \implies asc_{\emptyset}(Y, \mathbf{Z} \cup X), \quad (14)$$

where

$$asc_{\mathbf{S}}(Y, \mathbf{W}) := \forall W \in \mathbf{W} : asc_{\mathbf{S}}(Y, W \mid \mathbf{W} \setminus W). \quad (15)$$

When Equation (14) *is true, we say that Y is $|\mathbf{Z}|$-order* inter-associated *with $X \cup \mathbf{Z}$.*

Lemma 2 requires no faithfulness assumption—it holds for every distribution via the semi-graphoid axioms—and generalises the 2-association property of unfaithful triples in Marx et al. [2021, Definition 6] to arbitrary orders; the noisy-parity Example 1 is exactly this case with $\mathbf{Z} = \{X_2, X_3\}$. It motivates the *make_strict* routine of Algorithm 2, which prunes a found dependence set towards an inter-associated one.

The last consideration we need to make for *kOMB* to be remotely feasible—especially in cases where the true MB of $Y \in \boldsymbol{V}$ in $\mathcal{G}$ is large—is that, we need to somehow limit the size of the conditioning set when conducting the conditional independence (CI) tests in *kOMB*. When testing if Y and X are k-order associated, we will generally use CI tests of the form, $\forall \mathbf{S}' \subseteq \mathbf{S} : Y \perp\!\!\!\perp_{\mathbb{P}} X \mid \mathbf{Z} \cup \mathbf{S}'$, where $\mathbf{Z}$ is the "dependence" set that renders Y and X conditionally dependent. Fortunately, assuming k-order faithfulness, we know that the size of $\mathbf{Z}$ is bounded by k.

However the same cannot be said for the set $\mathbf{S}'$ as it can be as large as the candidate MB $\mathbf{S}$–which can be arbitrarily large depending on $\mathcal{G}$. Therefore, the third and last assumption we will make is that if Y and X are conditionally independent given $\mathbf{Z} \cup \mathbf{S}$, then there must exist some subset with size less than l, $\mathbf{S}' \in \mathcal{P}_{\leq l}(\mathbf{S})$, such that Y and X are still conditionally independent given $\mathbf{Z} \cup \mathbf{S}'$.

Assumption 3 (l-Bounded Separator Assumption)**.** For a PGM $\mathcal{M} = (\mathcal{G}, \mathbb{P})$, distinct variables $Y, X \in \boldsymbol{V}$, and $\boldsymbol{V}' := \boldsymbol{V} \setminus \{Y, X\}$, let us first recall the k-order faithfulness assumption as defined in Equation (11) $\forall \mathbf{S} \subseteq \boldsymbol{V} \setminus \{Y, X\}$,

$$\begin{aligned} &\forall \mathbf{Z} \in \mathcal{P}_{\leq k}(\boldsymbol{V}') \;\; \exists \mathbf{S}' \subseteq \mathbf{S} : Y \perp\!\!\!\perp_{\mathbb{P}} X \mid \mathbf{Z} \cup \mathbf{S}' \\ &\implies Y \perp\!\!\!\perp_{\mathcal{G}} X \mid \mathbf{S}. \end{aligned} \quad (11)$$

Then, for some natural number $l \in \mathbb{N}$, we define an l-bounded separator version of the k-order faithfulness assumption

$\forall \mathbf{S} \subseteq \boldsymbol{V} \setminus \{Y, X\}$,

$$\begin{aligned} &\forall \mathbf{Z} \in \mathcal{P}_{\leq k}(\boldsymbol{V}') \;\; \exists \mathbf{S}' \in \mathcal{P}_{\leq l}(\mathbf{S}) : Y \perp\!\!\!\perp_{\mathbb{P}} X \mid \mathbf{Z} \cup \mathbf{S}' \\ &\quad \iff \forall \mathbf{Z} \in \mathcal{P}_{\leq k}(\boldsymbol{V}') \;\; \exists \mathbf{S}' \subseteq \mathbf{S} : Y \perp\!\!\!\perp_{\mathbb{P}} X \mid \mathbf{Z} \cup \mathbf{S}' \\ &\quad \implies Y \perp\!\!\!\perp_{\mathcal{G}} X \mid \mathbf{S}, \end{aligned} \tag{16}$$

which essentially states that as long as Y and X are conditionally independent given $\mathbf{Z}$ and every $\leq l$-cardinality subset of $\mathbf{S}$, then Y and X are conditionally independent given $\mathbf{Z}$ and every possible subset of $\mathbf{S}$—regardless of cardinality.

The contrapositive of this assumption is then

$$\begin{aligned} &Y \not\perp\!\!\!\perp_{\mathcal{G}} X \mid \mathbf{S} \\ &\quad \implies \exists \mathbf{Z} \in \mathcal{P}_{\leq k}(\boldsymbol{V}') \;\; \forall \mathbf{S}' \subseteq \mathbf{S} : Y \not\perp\!\!\!\perp_{\mathbb{P}} X \mid \mathbf{Z} \cup \mathbf{S}' \\ &\quad \iff \exists \mathbf{Z} \in \mathcal{P}_{\leq k}(\boldsymbol{V}') : asc_{\mathbf{S}}(Y, X \mid \mathbf{Z}) \\ &\quad \iff \exists \mathbf{Z} \in \mathcal{P}_{\leq k}(\boldsymbol{V}') : asc^{(l)}_{\mathbf{S}}(Y, X \mid \mathbf{Z}), \end{aligned} \tag{17}$$

where $asc^{(l)}_{\mathbf{S}}(Y, X \mid \mathbf{Z}) \coloneqq \forall \mathbf{S}' \in \mathcal{P}_{\leq l}(\mathbf{S}) : Y \not\perp\!\!\!\perp_{\mathbb{P}} X \mid \mathbf{Z} \cup \mathbf{S}'$.

The concept of bounding the maximum size of the separators in a graph has been explored before. For instance, Soh and Tatikonda [2019] defined an undirected graph $\mathcal{G}$ to be weakly K-separable if for any two non-adjacent vertices in $\mathcal{G}$, there exists a set of vertices with cardinality $\leq K$ that separates the two vertices in $\mathcal{G}$. However, instead of focusing on separation in the graph $\mathcal{G}$, our l-bounded separator assumption focuses on the cardinality of "separating" sets that causes conditional independencies between variables in $\boldsymbol{V}$.

The l-bounded separator assumption is also analogous to assumptions that existing constraint-based MB discovery methods make about the size of the conditioning sets needed to decide conditional independence reliably from finite samples [Margaritis and Thrun, 1999, Tsamardinos et al., 2003a, Aliferis et al., 2003, Tsamardinos et al., 2003b]. Assumption 3 simply makes such a bound explicit. Furthermore, as shown later in Appendix F, the bound is mild on the networks used in Section 6.2.

5.1 THEORETICAL ALGORITHM

As a proof-of-concept, in this section we propose *kOMB*—a modification of the Grow and Shrink (GS) by Margaritis and Thrun [1999] to discover Markov blankets with higher-order relationships under the following assumptions: 1) the Global Markov Property from Assumption 1, 2) k-order faithfulness as in Assumption 2, and 3) the l-bounded separator assumption in Assumption 3.

The original GS algorithm starts with an empty MB $\mathbf{S} = \emptyset$ and iteratively adds a variable X into $\mathbf{S}$ if the conditional independence (CI) test $Y \perp\!\!\!\perp_{\mathbb{P}} X \mid \mathbf{S}$ indicates that Y is conditionally independent to X given $\mathbf{S}$. The main difference between *kOMB* and the GS algorithm then is: 1) considering the addition of entire sets of variables $X \cup \mathbf{X} : \mathbf{X} \in \mathcal{P}_{\leq k}(\boldsymbol{V} \setminus Y), X \in \boldsymbol{V} \setminus (Y \cup \mathbf{S} \cup \mathbf{X})$ at each iteration to find higher-order parity-type relationships between variables, and 2) conducting CI tests using just subsets of the current candidate MB $\mathbf{S}$ in the condition to ensure that these CI tests remain feasible even when the cardinality of $\mathbf{S}$ is large. As such, we break up *kOMB* into 3 sub-algorithms:

Algorithm 1: kOMB

Input: $\mathcal{D}, \mathcal{M}, Y, \alpha, k, l$
Output: $\text{MB}_{\mathcal{G}}(Y)$

```
S ← ∅;        // Candidate Markov Blanket
repeat
    V' ← V \ (Y ∪ S);
    X' ← find_inter_dep(Y, ∅, S, V', k, l);
    S ← S ∪ X'
until X' = ∅;
while ∃X ∈ S : find_cond(Y, X, ∅, S \ X, k, l) = ⊥ do
    S ← S \ X;
return S;
```

Algorithm 1 is just the entry-point function that implements the overall logic of the GS algorithm, with the main modifications and additions located in the other two algorithms,

Algorithm 2 is overall responsible for finding some subset of the variables not already in the candidate MB, $\mathbf{X} \in \mathcal{P}_{\leq k}(\boldsymbol{V} \setminus (X \cup Y \cup \mathbf{S}))$, that might cause Y and X to be conditionally dependent given $\mathbf{X}$ and some subset $\mathbf{X}' \in \mathcal{P}_{\leq k-|\mathbf{X}|}(\mathbf{S})$ of the current candidate MB, finally,

Algorithm 3 is then tasked with finding such a $\mathbf{X}' \in \mathcal{P}_{\leq k-|\mathbf{X}|}(\mathbf{S})$ that causes $Y \not\perp\!\!\!\perp_{\mathbb{P}} Y \mid \mathbf{X} \cup \mathbf{X}'$ and checking if there are any l-bounded subsets of the current candidate MB, $\mathbf{S}' \in \mathcal{P}_{\leq l}(\mathbf{S})$, such that $Y \perp\!\!\!\perp_{\mathbb{P}} Y \mid \mathbf{X} \cup \mathbf{X}' \cup \mathbf{S}'$.

Keeping this outline in mind, we now present the full algorithm for *kOMB* and show that it correctly discovers the MB of some $Y \in \boldsymbol{V}$ assuming that Assumptions 1 to 3 hold. The proofs for any subsequent Theorems can be found in Appendix C.

Theorem 1 (Correctness of Algorithm 3)**.** *Given an estimated distribution $\hat{\mathbb{P}}$, distinct variables $Y, X \in \boldsymbol{V}$, disjoint subsets $\mathbf{X}, \mathbf{S} \subset \boldsymbol{V} \setminus \{Y, X\}$, and the natural numbers $k, l \in \mathbb{N}$; the function find_cond in Algorithm 3 returns the set $\mathbf{X}' \cup \mathbf{X}$ if and only if $\exists \mathbf{X}' \in \mathcal{P}_{\leq k-|\mathbf{X}|}(\mathbf{S})$ where,*

$$\forall \mathbf{S}' \in \mathcal{P}_{\leq l}(\mathbf{S} \setminus \mathbf{X}') : Y \not\perp\!\!\!\perp_{\hat{\mathbb{P}}} X \mid \mathbf{X} \cup \mathbf{X}' \cup \mathbf{S}'. \tag{18}$$

Otherwise find_cond returns the failure value $\perp$, distinct from the empty set, which it may return on success when $\mathbf{X}' = \mathbf{X} = \emptyset$.

Theorem 2 (Correctness of Algorithm 2)**.** *Given an estimated distribution $\hat{\mathbb{P}}$, the target variable $Y \in \boldsymbol{V}$, the current candidate MB $\mathbf{S} \subset \boldsymbol{V} \setminus Y$, the remaining variables*

Algorithm 2: Association Mining

Function *make_strict(Y, X,* $\mathbf{Z}$*,* $\mathbf{S}$*, k, l)*:
  for $Z \in \mathbf{Z}$ **do**
    if $\neg$ *no_seps(Y, Z,* $X \cup \mathbf{Z} \setminus Z$*,* $\mathbf{S}$*, l)* **then**
      remove Z from $\mathbf{Z}$;
  return $X \cup \mathbf{Z}$;
Function *find_inter_dep(Y,* $\mathbf{X}$*,* $\mathbf{S}$*,* $\mathbf{V}'$*, k, l)*:
  if $|\mathbf{X}| > k$ **then**
    return $\emptyset$;
  for $X \in \mathbf{V} \setminus (Y \cup \mathbf{S} \cup \mathbf{X})$ **do**
    $\mathbf{Z} \leftarrow$*find_cond* $(Y, X, \mathbf{X}, \mathbf{S}, k, l)$;
    if $\mathbf{Z} \neq \perp$ **then**
      return *make_strict(Y, X,* $\mathbf{Z}$*,* $\mathbf{S}$*, k, l)*;
  for $X \in \mathbf{V}' \setminus (Y \cup \mathbf{S} \cup \mathbf{X})$ **do**
    $\mathbf{V}' \leftarrow \mathbf{V}' \setminus X$;
    $\mathbf{X}' \leftarrow$ find_inter_dep$(Y, \mathbf{X} \cup X, \mathbf{S}, \mathbf{V}', k, l)$;
    if $\mathbf{X}' \neq \emptyset$ **then**
      return $\mathbf{X}'$;
  return $\emptyset$;

Algorithm 3: Practical Conditional Independence Test

Function *no_seps(Y, X,* $\mathbf{Z}$*,* $\mathbf{S}$*, l)*:
  for $\mathbf{S}' \in \mathcal{P}_{\leq l}(\mathbf{S} \setminus \mathbf{Z})$ **do**
    if $Y \perp\!\!\!\perp X \mid \mathbf{Z} \cup \mathbf{S}'$ **then**
      return *False*;
  return *True*;
Function *find_cond(Y, X,* $\mathbf{X}$*,* $\mathbf{S}$*, k, l)*:
  for $\mathbf{X}' \in \mathcal{P}_{\leq k-|\mathbf{X}|}(\mathbf{S})$ **do**
    $\mathbf{Z} \leftarrow \mathbf{X}' \cup \mathbf{X}$;
    if $Y \not\perp\!\!\!\perp X \mid \mathbf{Z}$ **then**
      if *no_seps(Y, X,* $\mathbf{Z}$*,* $\mathbf{S}$*, l)* **then**
        return $\mathbf{Z}$;
  return $\perp$;

$\boldsymbol{V}' = \boldsymbol{V} \setminus (Y \cup \mathbf{S})$*, and the natural numbers* $k, l \in \mathbb{N}$*; the call* $\textit{find_inter_dep}(Y, \emptyset, \mathbf{S}, \boldsymbol{V}', k, l)$ *returns the empty set if and only if*

$$\forall X \in \boldsymbol{V}' \;\; \forall \mathbf{Z} \in \mathcal{P}_{\leq k}(\boldsymbol{V} \setminus \{Y, X\}) \;\; \exists \mathbf{S}' \in \mathcal{P}_{\leq l}(\mathbf{S} \setminus \mathbf{Z}) : \quad Y \perp\!\!\!\perp_{\hat{\mathbb{P}}} X \mid \mathbf{Z} \cup \mathbf{S}'. \tag{19}$$

Otherwise it returns a nonempty set $X \cup \mathbf{Z}_R$ *with* $X \in \boldsymbol{V}'$ *and* $\mathbf{Z}_R \subseteq \mathbf{Z}$ *for some* $\mathbf{Z} \in \mathcal{P}_{\leq k}(\boldsymbol{V} \setminus \{Y, X\})$ *satisfying the estimated* l*-bounded association* $\forall \mathbf{S}' \in \mathcal{P}_{\leq l}(\mathbf{S} \setminus \mathbf{Z}) : Y \not\perp\!\!\!\perp_{\hat{\mathbb{P}}} X \mid \mathbf{Z} \cup \mathbf{S}'$.

Theorem 3 (Correctness of Algorithm 1)**.** *Given an estimated distribution* $\hat{\mathbb{P}}$ *over variables* $\boldsymbol{V}$*, the target variable whose graphical Markov blanket we want to discover* $Y \in \boldsymbol{V}$*, and the natural numbers* $k, l \in \mathbb{N}$*; the grow phase of* kOMB *then grows the candidate MB* $\mathbf{S}$ *such that,*

$$MB_{\mathcal{G}}(Y) \subseteq \mathbf{S}. \tag{20}$$

If, in addition, every true blanket member admits an in-blanket dependence witness—i.e. for every $X \in MB_{\mathcal{G}}(Y)$ *and every* $\mathbf{S}$ *with* $MB_{\mathcal{G}}(Y) \subseteq \mathbf{S} \subseteq \boldsymbol{V} \setminus Y$ *there is a witness* $\mathbf{Z} \in \mathcal{P}_{\leq k}(MB_{\mathcal{G}}(Y) \setminus X)$ *with* $asc^{(l)}_{\mathbf{S} \setminus (X \cup \mathbf{Z})}(Y, X \mid \mathbf{Z})$*—then the shrink phase of* kOMB *removes variables from* $\mathbf{S}$ *until,*

$$MB_{\mathcal{G}}(Y) = \mathbf{S}, \tag{21}$$

where it subsequently terminates and returns $\mathbf{S}$*.*

Beyond Assumptions 1 to 3, the recovery guarantee of Theorem 3 assumes the CI decisions are *sufficiently accurate*—i.e. they reflect the true conditional independencies of $\mathbb{P}$.

5.2 COMPUTATIONAL COMPLEXITY OF *KOMB*

As *kOMB* is a proof-of-concept, we prioritised correctness and clarity over efficiency. It is nonetheless instructive to characterise its worst-case time complexity, which we state in Corollary 1 with proof given in Appendix D.

Corollary 1 (Computational Complexity of *kOMB*)**.** *Let* $n = |\boldsymbol{V}|$ *denote the number of variables, N the sample size, and K the maximum size of any candidate MB* $\mathbf{S}$ *encountered by Algorithm 1 during execution. Furthermore assume each CI test runs in* $O(N)$ *time. Then, for a given order k and separator bound l,* kOMB *(Algorithm 1) terminates in worst-case time*

$$O\left(K\, n \sum_{i=0}^{k} \binom{n}{i} \left[\sum_{i'=0}^{k-i} \binom{K}{i'} \sum_{j=0}^{l} \binom{K - i'}{j} N \right] \right), \tag{22}$$

where we adopt the convention that $\binom{n}{m} = 0$ *whenever* $n < m$*.*

Treating k and l as constants, Equation (22) is bounded by $O(n^{k+1} K^{k+l+1} N)$—polynomial in n, K, and N, and for $k = 0$ it reduces to a scan that is linear in n, recovering the original GS algorithm. Crucially, the search for a dependence-exposing subset and its separator is governed by K rather than n, a direct consequence of the l-bounded separator assumption (Assumption 3). The remaining exponential dependence on k and l motivates keeping them small. The ablation study in Appendix G showcases how differing values of k and l affects Algorithm 1 runtime and its ability to find high-order dependencies. Generally, from the results in Appendix G for the benchmark datasets used, we found that setting $k = l$ and having $k \leq 2$ provides a decent trade-off between runtime and finding high-order dependencies.

Table 1: Results on Synthetically Generated Data

Method	and	ex-1	ex-2	or	parity
GS	0.410	0.313	0.288	0.365	0.025
	±0.405	±0.419	±0.407	±0.374	±0.110
kOMB	0.410	0.283	0.268	0.350	0.025
$(k = 0, l = 3)$	±0.405	±0.372	±0.374	±0.351	±0.110
kOMB	**0.618**	0.675	0.675	0.560	0.050
$(k = 1, l = 3)$	±0.374	±0.474	±0.474	±0.435	±0.221
kOMB	**0.618**	**1.000**	**1.000**	**0.710**	**1.000**
$(k = 2, l = 3)$	±0.374	±0.000	±0.000	±0.384	±0.000

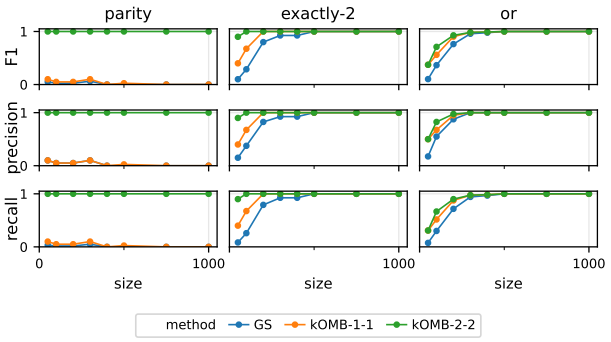


Figure 4: Precision, recall, and F1 score of GS and (*kOMB-k-l*) when recovering the MBs of the synthetic BNs in Section 6.1, over sample size. Each row corresponds to a metric and each column to a BN.

6 EXPERIMENTS

In order to empirically test *kOMB* we implemented the algorithm as a `Python` package written in `Rust`. The implementation uses the G-Test with $\alpha = 0.01$ as its conditional independence test of choice when possible and falls back to the SCI test [Marx and Vreeken, 2019] when G-test is too weak for the test being conducted—i.e. when the average frequency for each cell is < 5. A public repo for our implementation of *kOMB* and code for the experiments can be found at `https://github.com/lklee9/k-order-Markov-blanket`. All experiments were run on an Apple M4 Mac mini.

Throughout this section we will compare *kOMB* against existing methods for Markov blanket (MB) discovery. Specifically we will compare against the following eight MB discovery methods: GS [Margaritis and Thrun, 1999], IAMB [Tsamardinos et al., 2003a], HITON [Aliferis et al., 2003], MMMB [Tsamardinos et al., 2003b], PCMB [Peña et al., 2007], LRH [Liu and Liu, 2016], STMB [Gao and Ji, 2017], BAMB [Ling et al., 2019]. We will use the implementation of these methods found in the `pyCausalFS` python library [Yu et al., 2020] with a significance level of 0.01 for any statistical tests as well.

When comparing between different MB discovery methods, we will use the F1 score defined as: $F1 = 2 \times precision \times recall/(precision + recall)$, which is the harmonic mean of *precision* and *recall*. Here *precision* denotes the ratio between the number of true positives in the inferred MB and the size of the inferred MB. On the other hand, *Recall* denotes the ratio between the number of true positives in the inferred MB and the size of the true MB. Therefore $F1 = 1$ represents the perfect precision and recall.

6.1 EXPERIMENTS WITH SYNTHETIC BAYESIAN NETWORKS

To systematically test how *kOMB* behaves in the face of true and empirical faithfulness violations, we will task *kOMB* at learning the MB of every variable for five different basic toy Bayesian networks. These problems will have 4 variables—$\mathbf{X} = \{X_1, X_2, X_3\}$ and Y—with a DAG that has edges $\forall i \in [3] : X_i \to Y$. Each variable $X \in \mathbf{X}$ is an independent Bernoulli random variable with $p = 0.5$. The conditional distribution of Y is then $\mathbb{P}(Y = f(\mathbf{X}) \mid \mathbf{X}) = 0.9$ where $f(\mathbf{X})$ is one of 5 different Boolean functions, one for each toy problem which we describe in Appendix E.

We then sample 100 samples from the 5 Bayesian networks 10 different times, leading to 10×5 samples of size 100—hence the size of each sample is much larger than the cardinality of the domain over all variables, $100 >> 2^4$. We then tasked each method to infer—for each Bayesian network—the MB of all 4 variables from the 10 different samples separately. We then averaged the F1 score over the 10×4 runs and variables for each Bayesian network, resulting in Table 1. Note that we only include the results for *kOMB* and Grow and Shrink (GS) because GS was the best performing method out of all the previous methods mentioned in Section 6 for these synthetic datasets.

From the results in Table 1, we can observe that *kOMB* consistently outperforms the other approaches at MB discovery. More interestingly—even for the problems whose Bayesian networks are faithful—as we increase k and look at higher order relationships, *kOMB* achieves a better F1 score and therefore is capable of better overcoming empirical unfaithfulness. Furthermore the MB for the parity problem is only recoverable when $k = 2$, which confirms *kOMB*'s ability to find these higher-order parity-like relations.

In order to better understand the source of these empirical unfaithfulness in the low sample size regime, Figure 4 traces the precision, recall, and F1 score as the sample size grows from 50 to 100. As expected only *kOMB*-2-2 is capable of recovering the higher-order dependence of the parity network regardless of sample size. From the other faithful networks, we can observe that at small sample size, all methods exhibit lower recall and precision which indicates dependencies missed by the MB discovery methods and spurious depend-

Table 2: F1 score and runtime (in seconds) on benchmark datasets. Best F1 and fastest runtime per dataset in bold.

	F1 Score				Runtime (s)			
Method	Alarm1	Barley	Insurance	Mildew	Alarm1	Barley	Insurance	Mildew
BAMB	0.6742 ±0.1771	0.3385 ±0.1226	0.6588 ±0.0797	0.4945 ±0.1084	3.517 ±4.305	1.149 ±0.681	5.615 ±5.358	0.797 ±0.392
GS	0.3434 ±0.1503	0.2113 ±0.0807	0.4615 ±0.1477	0.2867 ±0.1581	4.875 ±5.183	1.575 ±0.606	4.433 ±3.792	1.035 ±0.341
HITON_MB	0.7636 ±0.1861	0.3400 ±0.1284	0.6746 ±0.1594	0.4918 ±0.1487	**2.560** ±1.277	**1.027** ±0.298	3.514 ±2.525	0.604 ±0.218
MMMB	0.7692 ±0.1874	0.3298 ±0.1274	0.6978 ±0.1405	0.4861 ±0.1810	2.724 ±1.278	1.090 ±0.374	**3.508** ±2.525	**0.553** ±0.128
kOMB ($k=1, l=1$)	0.7804 ±0.1145	**0.5158** ±0.1964	0.7136 ±0.1328	**0.6220** ±0.1595	5.258 ±2.921	58.190 ±31.619	4.856 ±2.041	17.698 ±10.205
kOMB ($k=2, l=2$)	**0.8209** ±0.1914	0.3128 ±0.1582	**0.7336** ±0.1624	0.4360 ±0.1394	18.144 ±7.566	335.821 ±135.966	14.493 ±7.280	121.909 ±33.153

encies erroneously detected by said methods respectively. That said, we can also observe that for *kOMB*, the higher the value of k used, the greater its performance in the low sample regime and the quicker it recovers the true MBs for all variables.

6.2 EXPERIMENTS WITH BENCHMARK BAYESIAN NETWORKS

To further explore how *kOMB* behaves in more realistic scenarios, we then repeated the same procedure as Section 6.1, but with some well-known benchmark datasets[4] in the MB discovery literature. The main difference in the experiments in this section is that we used samples of 5000 and we only tasked each method at finding the MB for the 10 variables with the largest neighborhoods in the Bayesian network. Furthermore, we also omit some of the poorer performing methods in Table 2 to conserve space.

From the results in Table 2 we can observe that *kOMB* does generally outperform the other existing approaches to MB discovery. Whether a higher order helps, however, depends on the network's cardinality. On the lower-cardinality Alarm1 and Insurance networks *kOMB* with order 2 improves on order 1, whereas on the high-cardinality Barley and Mildew networks order 1 is best and order 2 is substantially worse—there the larger conditioning sets leave the higher-order CI tests underpowered at this sample size. Runtime, in turn, grows with both the order k and the cardinality of the network. Specifically, the baselines finish within roughly 6 s per dataset, while *kOMB* ranges from comparable at order 1 on the lower-cardinality networks (Alarm1, Insurance) to a few hundred times slower at order 2 on the high-cardinality networks (Barley, Mildew). The full results and runtimes for every method and every (k, l) configuration are given in Tables 5 and 6 under Appendix G.

7 CONCLUSION

We introduced a k-order relaxation of faithfulness that explicitly accounts for parity-type dependencies among $k+2$ variables, and used it to derive a proof-of-concept Markov blanket (MB) discovery algorithm, *kOMB*; with theoretical guarantees that *kOMB* recovers the graphical MB under the proposed assumptions. Empirically, *kOMB* overcomes both true and empirical faithfulness violations on synthetic problems, and it performs competitively on commonly used benchmark datasets.

As *kOMB* is a proof-of-concept demonstrating that exploiting higher-order dependencies can improve MB discovery, an important direction for future work is the development of *approximate* algorithms that uncover these higher-order dependencies without *kOMB*'s worst-case cost—for example, through more intricate pruning rules that make *kOMB* scale better with larger values of k. Secondly, although our ablation (Appendix G) shows that $l = k$ with $k \leq 2$ is a robust default, principled *a priori* selection of k and l without expert knowledge remains open. A practical alternative is to run *kOMB* for several values of k and l while memoising results across runs to speed up the total runtime. Finally, our evaluation compares against established constraint-based MB discovery methods using the G-test (with an SCI fallback) for conditional independence. Therefore, utilising more recent alternative CI tests is a further avenue for future work.

[4]The datasets can be accessed at: `https://pages.mtu.edu/~lebrown/supplements/mmhc_paper/mmhc_index.html`

Acknowledgements

This research has been funded by the Federal Ministry of Research, Technology and Space of Germany and the state of North Rhine-Westphalia as part of the Lamarr Institute for Machine Learning and Artificial Intelligence.

References

C. F. Aliferis, I. Tsamardinos, and A. Statnikov. HITON: a novel Markov Blanket algorithm for optimal variable selection. *AMIA ... Annual Symposium proceedings. AMIA Symposium*, 2003:21–25, 2003. ISSN 1942-597X.

Philip Boeken, Patrick Forré, and Joris M. Mooij. Are Bayesian networks typically faithful? *arXiv:2410.16004 [math]*, January 2025. doi: 10.48550/arXiv.2410.16004.

A. Philip Dawid. Conditional independence in statistical theory. *Journal of the Royal Statistical Society. Series B (Methodological)*, 41(1):1–31, 1979. ISSN 0035-9246.

Shuyu Dong, Michele Sebag, Kento Uemura, Akito Fujii, Shuang Chang, Yusuke Koyanagi, and Koji Maruhashi. DCILP: A Distributed Approach for Large-Scale Causal Structure Learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(15):16345–16353, April 2025. ISSN 2374-3468. doi: 10.1609/aaai.v39i15.33795.

Malcolm Forster, Garvesh Raskutti, Reuben Stern, and Naftali Weinberger. The Frugal Inference of Causal Relations. *The British Journal for the Philosophy of Science*, 69(3): 821–848, 2018. ISSN 0007-0882.

Tian Gao and Qiang Ji. Efficient Markov Blanket Discovery and Its Application. *IEEE Transactions on Cybernetics*, 47(5):1169–1179, May 2017. ISSN 2168-2275. doi: 10.1109/TCYB.2016.2539338.

John Michael Hammersley and Peter Clifford. Markov fields on finite graphs and lattices. *Unpublished manuscript*, 1971.

Lijuan Hu and Zhuoyuan Zheng. Causal Feature Selection Algorithm Based on Maximizing Neighbourhood Mutual Information. *Proceedings of the 2024 8th International Conference on Computer Science and Artificial Intelligence*, pages 482–490, December 2024. doi: 10.1145/3709026.3709031.

Takanori Inazumi, T. Washio, Shohei Shimizu, J. Suzuki, Akihiro Yamamoto, and Y. Kawahara. Discovering causal structures in binary exclusive-or skew acyclic models. In *Conference on Uncertainty in Artificial Intelligence*, July 2011.

Steffen Lauritzen and Kayvan Sadeghi. Unifying Markov Properties for Graphical Models. *The Annals of Statistics*, 46(5):2251–2278, 2018. ISSN 0090-5364.

Steffen L. Lauritzen. *Graphical models*. Number 17 in Oxford statistical science series. Clarendon Press ; Oxford University Press, Oxford : New York, 1996. ISBN 978-0-19-852219-5.

Jan Lemeire, Stijn Meganck, Francesco Cartella, and Tingting Liu. Conservative independence-based causal structure learning in absence of adjacency faithfulness. *International Journal of Approximate Reasoning*, 53(9): 1305–1325, December 2012. ISSN 0888-613X. doi: 10.1016/j.ijar.2012.06.004.

Zhaolong Ling, Kui Yu, Hao Wang, Lin Liu, Wei Ding, and Xindong Wu. BAMB: A Balanced Markov Blanket Discovery Approach to Feature Selection. *ACM Trans. Intell. Syst. Technol.*, 10(5):52:1–52:25, October 2019. ISSN 2157-6904. doi: 10.1145/3335676.

Zhaolong Ling, Kui Yu, Hao Wang, Lei Li, and Xindong Wu. Using Feature Selection for Local Causal Structure Learning. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 5(4):530–540, August 2021. ISSN 2471-285X. doi: 10.1109/TETCI.2020.2978238.

Xuqing Liu and Xinsheng Liu. Swamping and masking in Markov boundary discovery. *Machine Learning*, 104(1): 25–54, July 2016. ISSN 1573-0565. doi: 10.1007/s10994-016-5545-0.

Dimitris Margaritis and Sebastian Thrun. Bayesian Network Induction via Local Neighborhoods. In *Advances in Neural Information Processing Systems*, volume 12. MIT Press, 1999.

Alexander Marx and Jilles Vreeken. Testing Conditional Independence on Discrete Data using Stochastic Complexity. In *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, pages 496–505. PMLR, April 2019.

Alexander Marx, Arthur Gretton, and Joris M. Mooij. A weaker faithfulness assumption based on triple interactions. In *Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence*, pages 451–460. PMLR, December 2021.

Christopher Meek. Strong completeness and faithfulness in Bayesian networks. In *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*, UAI'95, pages 411–418, San Francisco, CA, USA, August 1995. Morgan Kaufmann Publishers Inc. ISBN 978-1-55860-385-1.

J. Pearl and Thomas Verma. The Logic of Representing Dependencies by Directed Graphs. In *AAAI Conference on Artificial Intelligence*, July 1987.

Judea Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1988. ISBN 978-1-55860-479-7.

Judea Pearl. *Causality: models, reasoning, and inference*. Cambridge University Press, Cambridge New York, NY Port Melbourne New Delhi Singapore, second edition, reprinted with corrections edition, 2009. ISBN 978-0-521-89560-6.

Judea Pearl and Azaria Paz. GRAPHOIDS: a graph-based logic for reasoning about relevance relations. Technical Report CSD-850038, UCLA Computer Science Department, 1985.

Judea Pearl, Dan Geiger, and Thomas Verma. Conditional independence and its representations. *Kybernetika*, 25(7): 33–44, 1989.

Jose M. Peña, Roland Nilsson, Johan Björkegren, and Jesper Tegnér. Towards scalable and data efficient learning of Markov boundaries. *International Journal of Approximate Reasoning*, 45(2):211–232, July 2007. ISSN 0888-613X. doi: 10.1016/j.ijar.2006.06.008.

Kayvan Sadeghi. Faithfulness of Probability Distributions and Graphs. *Journal of Machine Learning Research*, 18 (148):1–29, 2017. ISSN 1533-7928.

Federico Schlüter. A survey on independence-based Markov networks learning. *Artificial Intelligence Review*, 42(4): 1069–1093, December 2014. ISSN 1573-7462. doi: 10.1007/s10462-012-9346-y.

De Wen Soh and Sekhar Tatikonda. Learning Unfaithful K-separable Gaussian Graphical Models. *Journal of Machine Learning Research*, 20(109):1–30, 2019. ISSN 1533-7928.

Peter Spirtes, Clark Glymour, and Richard Scheines. *Causation, Prediction, and Search*. The MIT Press, 2nd edition, January 2001. ISBN 978-0-262-28415-8. doi: 10.7551/mitpress/1754.001.0001.

I. Tsamardinos, C. Aliferis, and A. Statnikov. Algorithms for Large Scale Markov Blanket Discovery. In *The Florida AI Research Society*, 2003a.

Ioannis Tsamardinos and Constantin F. Aliferis. Towards Principled Feature Selection: Relevancy, Filters and Wrappers. In *International Workshop on Artificial Intelligence and Statistics*, pages 300–307. PMLR, January 2003.

Ioannis Tsamardinos, Constantin F. Aliferis, and Alexander Statnikov. Time and sample efficient discovery of Markov blankets and direct causal relations. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '03, pages 673–678, New York, NY, USA, August 2003b. Association for Computing Machinery. ISBN 978-1-58113-737-8. doi: 10.1145/956750.956838.

Caroline Uhler, Garvesh Raskutti, Peter Bühlmann, and Bin Yu. Geometry of the Faithfulness Assumption in Causal Inference. *The Annals of Statistics*, 41(2):436–463, 2013. ISSN 0090-5364.

Hao Wang, Zhaolong Ling, Kui Yu, and Xindong Wu. Towards efficient and effective discovery of Markov blankets for feature selection. *Information Sciences*, 509:227–242, January 2020. ISSN 0020-0255. doi: 10.1016/j.ins.2019.09.010.

Kui Yu, Xianjie Guo, Lin Liu, Jiuyong Li, Hao Wang, Zhaolong Ling, and Xindong Wu. Causality-based Feature Selection: Methods and Evaluations. *ACM Comput. Surv.*, 53(5):111:1–111:36, September 2020. ISSN 0360-0300. doi: 10.1145/3409382.

High-Order Markov Blanket Discovery via a k-Order Relaxation of the Faithfulness Assumption (Supplementary Material)

Loong Kuan Lee[1] **Ragavi Krishnamoorthy**[1] **Nico Piatkowski**[1]

[1]Hybrid Intelligence, Fraunhofer IAIS, Sankt Augustin, Germany,

A GRAPHOID AXIOMS

Definition 6 (Graphoid Axioms)**.** For disjoint subsets $\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbf{W} \subseteq \boldsymbol{V}$, a semi-graphoid independence model $\mathcal{J}(\cdot)$ obeys the following properties [Dawid, 1979, Pearl and Verma, 1987]:

1. Symmetry: $\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z} \implies \mathbf{Y} \perp\!\!\!\perp \mathbf{X} \mid \mathbf{Z}$
2. Decomposition: $\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \cup \mathbf{W} \mid \mathbf{Z} \implies \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z}$
3. Weak Union: $\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \cup \mathbf{W} \mid \mathbf{Z} \implies \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{W} \cup \mathbf{Z}$
4. Contraction: $(\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{W} \cup \mathbf{Z}) \wedge (\mathbf{X} \perp\!\!\!\perp \mathbf{W} \mid \mathbf{Z}) \implies \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \cup \mathbf{W} \mid \mathbf{Z}$

The independence model is called a graphoid if it also obeys the following additional property [Pearl and Paz, 1985]:

5. Intersection: $(\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z} \cup \mathbf{W}) \wedge (\mathbf{X} \perp\!\!\!\perp \mathbf{W} \mid \mathbf{Z} \cup \mathbf{Y}) \implies \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \cup \mathbf{W} \mid \mathbf{Z}$

All graph independence models $\mathcal{J}(\mathcal{G})$ are guaranteed to obey all the graphoid axioms [Lauritzen and Sadeghi, 2018]. However, probabilistic independence models $\mathcal{J}(\mathbb{P})$ are only guaranteed to obey the semi-graphoid axioms. However if the joint distribution is strictly positive, then $\mathcal{J}(\mathbb{P})$ is guaranteed to be a graphoid as well [Pearl, 1988, Pearl et al., 1989].

B PROOFS

Lemma 1 (k-Order Dependence Arises from Association)**.** *Given a PGM $\mathcal{M} = (\mathcal{G}, \mathbb{P})$, let $Y, X \in \mathbf{V}$ be distinct random variables and $\mathbf{S}, \mathbf{Z} \subset \boldsymbol{V} \setminus \{Y, X\}$ be disjoint subsets. Then the following holds for all $Y, X, \mathbf{S}, \mathbf{Z}$:*

$$asc_{\mathbf{S}}(Y, X \mid \mathbf{Z}) \implies \exists \mathbf{Z}' \subseteq \mathbf{Z} : dep_{\mathbf{S}}(Y, X \mid \mathbf{Z}'). \quad (10)$$

Proof. Consider the family of association witnesses contained in $\mathbf{Z}$,

$$\mathcal{W} := \{\mathbf{W} \subseteq \mathbf{Z} : asc_{\mathbf{S}}(Y, X \mid \mathbf{W})\}.$$

By the hypothesis $asc_{\mathbf{S}}(Y, X \mid \mathbf{Z})$ we have $\mathbf{Z} \in \mathcal{W}$, so $\mathcal{W}$ is nonempty; and since $\mathbf{Z}$ is finite, so is $\mathcal{W}$. Hence $\mathcal{W}$ contains an element $\mathbf{Z}'$ of minimum cardinality. Every proper subset $\mathbf{Z}^* \subsetneq \mathbf{Z}'$ satisfies $|\mathbf{Z}^*| < |\mathbf{Z}'|$ and $\mathbf{Z}^* \subseteq \mathbf{Z}$, so by the minimality of $\mathbf{Z}'$ we have $\mathbf{Z}^* \notin \mathcal{W}$, i.e. $\neg\, asc_{\mathbf{S}}(Y, X \mid \mathbf{Z}^*)$. Together with $asc_{\mathbf{S}}(Y, X \mid \mathbf{Z}')$, this is by Equation (7) exactly $dep_{\mathbf{S}}(Y, X \mid \mathbf{Z}')$. As $\mathbf{Z}' \subseteq \mathbf{Z}$, this establishes Equation (10). □

Lemma 2 (Marginal k-Order Dependence Implies Inter-Association)**.** *Given a PGM $\mathcal{M} = (\mathcal{G}, \mathbb{P})$, let $X, Y \in \boldsymbol{V}$ be distinct variables and $\mathbf{Z} \subseteq \boldsymbol{V} \setminus \{Y, X\}$. Then if Y and X are $|\mathbf{Z}|$-order dependent w.r.t. $\mathbf{Z}$ over the empty separating set, Y is $|\mathbf{Z}|$-order associated with every variable in $\mathbf{Z} \cup X$,*

$$dep_{\emptyset}(Y, X \mid \mathbf{Z}) \implies asc_{\emptyset}(Y, \mathbf{Z} \cup X), \quad (14)$$

where

$$asc_{\mathbf{S}}(Y, \mathbf{W}) := \forall W \in \mathbf{W} : asc_{\mathbf{S}}(Y, W \mid \mathbf{W} \setminus W). \quad (15)$$

When Equation (14) *is true, we say that Y is $|\mathbf{Z}|$-order* inter-associated *with $X \cup \mathbf{Z}$.*

Proof. Assume $dep_{\emptyset}(Y, X \mid \mathbf{Z})$. For the variable X, the inter-association clause is $asc_{\emptyset}(Y, X \mid \mathbf{Z})$, i.e. $Y \not\perp\!\!\!\perp_{\mathbb{P}} X \mid \mathbf{Z}$, which is exactly the first conjunct of Equation (7) and hence immediate.

Now suppose, towards a contradiction, that the clause fails for some $Z \in \mathbf{Z}$, i.e. $Y \perp\!\!\!\perp_{\mathbb{P}} Z \mid X \cup (\mathbf{Z} \setminus Z)$. Since $\mathbf{Z} \setminus Z$ is a proper subset of $\mathbf{Z}$ and the separating set is empty, the second conjunct of Equation (7) gives $Y \perp\!\!\!\perp_{\mathbb{P}} X \mid \mathbf{Z} \setminus Z$. Applying the contraction and then weak union graphoid axioms of Definition 6,

$$\begin{aligned}&(Y \perp\!\!\!\perp_{\mathbb{P}} Z \mid X \cup (\mathbf{Z} \setminus Z)) \wedge (Y \perp\!\!\!\perp_{\mathbb{P}} X \mid \mathbf{Z} \setminus Z)\\ &\quad\implies Y \perp\!\!\!\perp_{\mathbb{P}} \{X, Z\} \mid \mathbf{Z} \setminus Z \implies Y \perp\!\!\!\perp_{\mathbb{P}} X \mid \mathbf{Z},\end{aligned} \quad (23)$$

contradicting $Y \not\perp\!\!\!\perp_{\mathbb{P}} X \mid \mathbf{Z}$. Hence $Y \not\perp\!\!\!\perp_{\mathbb{P}} Z \mid X \cup (\mathbf{Z} \setminus Z)$ for every $Z \in \mathbf{Z}$, which together with $Y \not\perp\!\!\!\perp_{\mathbb{P}} X \mid \mathbf{Z}$ is precisely $asc_{\emptyset}(Y, \mathbf{Z} \cup X)$. □

C CORRECTNESS OF *KOMB*

Theorem 1 (Correctness of Algorithm 3)**.** *Given an estimated distribution $\hat{\mathbb{P}}$, distinct variables $Y, X \in \boldsymbol{V}$, disjoint subsets $\mathbf{X}, \mathbf{S} \subset \boldsymbol{V} \setminus \{Y, X\}$, and the natural numbers $k, l \in \mathbb{N}$; the function find_cond in Algorithm 3 returns the set $\mathbf{X}' \cup \mathbf{X}$ if and only if $\exists \mathbf{X}' \in \mathcal{P}_{\leq k-|\mathbf{X}|}(\mathbf{S})$ where,*

$$\forall \mathbf{S}' \in \mathcal{P}_{\leq l}(\mathbf{S} \setminus \mathbf{X}') : Y \not\perp\!\!\!\perp_{\hat{\mathbb{P}}} X \mid \mathbf{X} \cup \mathbf{X}' \cup \mathbf{S}'. \tag{18}$$

Otherwise find_cond returns the failure value $\perp$, distinct from the empty set, which it may return on success when $\mathbf{X}' = \mathbf{X} = \emptyset$.

Proof. The function *find_cond* first iterates through all the subsets $\mathbf{X}' \in \mathcal{P}_{\leq k-|\mathbf{X}|}(\mathbf{S})$ and for each $\mathbf{X}'$ where $Y \not\perp\!\!\!\perp_{\hat{\mathbb{P}}} X \mid \mathbf{X} \cup \mathbf{X}'$ (line 9 in *find_cond*), it passes $\mathbf{X}'$ to the function *no_seps* (line 10 in *find_cond*) to determine if there are any subsets $\mathbf{S}' \in \mathcal{P}_{\leq l}(\mathbf{S} \setminus \mathbf{X}')$ that results in $Y \perp\!\!\!\perp_{\hat{\mathbb{P}}} X \mid \mathbf{X} \cup \mathbf{X}' \cup \mathbf{S}'$ (line 4 in *no_seps*). As soon as *no_seps* finds such a $\mathbf{S}'$, it returns `False` (line 4 in *no_seps*), and *find_cond* will need to go to the next iteration. Otherwise, if *no_seps* is unable to find any such subset $\mathbf{S}'$, it returns `True` (line 5 in *no_seps*), and *find_cond* then immediately returns $\mathbf{Z} = \mathbf{X}' \cup \mathbf{X}$ (line 11 in *find_cond*) as there are no l-bounded subsets $\mathbf{S}' \in \mathcal{P}_{\leq l}(\mathbf{S} \setminus \mathbf{Z})$ that render Y and X conditionally independent given $\mathbf{Z} \cup \mathbf{S}'$. Here we use that the external set $\mathbf{X}$ is disjoint from $\mathbf{S}$, so $\mathbf{S} \setminus \mathbf{Z} = \mathbf{S} \setminus \mathbf{X}'$; and since $\emptyset \in \mathcal{P}_{\leq l}(\mathbf{S} \setminus \mathbf{X}')$, the dependence check $Y \not\perp\!\!\!\perp_{\hat{\mathbb{P}}} X \mid \mathbf{Z}$ on line 9 is precisely the $\mathbf{S}' = \emptyset$ instance tested by *no_seps*, so a returned $\mathbf{Z}$ satisfies Equation (18) in full. However, if *find_cond* completes all its iteration without returning early, then we know that all subset $\mathbf{X}'$ has some l-bounded subset $\mathbf{S}'$ that will cause $Y \perp\!\!\!\perp_{\hat{\mathbb{P}}} X \mid \mathbf{X}' \cup \mathbf{X} \cup \mathbf{S}'$. Therefore *find_cond* returns the failure value $\perp$ in this case (line 12 in *find_cond*). □

Theorem 2 (Correctness of Algorithm 2)**.** *Given an estimated distribution $\hat{\mathbb{P}}$, the target variable $Y \in \boldsymbol{V}$, the current candidate MB $\mathbf{S} \subset \boldsymbol{V} \setminus Y$, the remaining variables $\boldsymbol{V}' = \boldsymbol{V} \setminus (Y \cup \mathbf{S})$, and the natural numbers $k, l \in \mathbb{N}$; the call find_inter_dep$(Y, \emptyset, \mathbf{S}, \boldsymbol{V}', k, l)$ returns the empty set if and only if*

$$\forall X \in \boldsymbol{V}' \ \forall \mathbf{Z} \in \mathcal{P}_{\leq k}(\boldsymbol{V} \setminus \{Y, X\}) \ \exists \mathbf{S}' \in \mathcal{P}_{\leq l}(\mathbf{S} \setminus \mathbf{Z}) : \\ Y \perp\!\!\!\perp_{\hat{\mathbb{P}}} X \mid \mathbf{Z} \cup \mathbf{S}'. \tag{19}$$

Otherwise it returns a nonempty set $X \cup \mathbf{Z}_R$ with $X \in \boldsymbol{V}'$ and $\mathbf{Z}_R \subseteq \mathbf{Z}$ for some $\mathbf{Z} \in \mathcal{P}_{\leq k}(\boldsymbol{V} \setminus \{Y, X\})$ satisfying the estimated l-bounded association $\forall \mathbf{S}' \in \mathcal{P}_{\leq l}(\mathbf{S} \setminus \mathbf{Z}) : Y \not\perp\!\!\!\perp_{\hat{\mathbb{P}}} X \mid \mathbf{Z} \cup \mathbf{S}'$.

Proof. First consider a call that returns a nonempty set. *find_inter_dep* does so only by returning *make_strict*$(Y, X, \mathbf{Z}, \mathbf{S}, k, l)$ after a call *find_cond*$(Y, X, \mathbf{X}, \mathbf{S}, k, l)$ has returned some $\mathbf{Z} \neq \perp$, for some $X \in \boldsymbol{V} \setminus (Y \cup \mathbf{S} \cup \mathbf{X})$ (as $\mathbf{S}$ is passed unchanged through every recursive call, any such X satisfies $X \in \boldsymbol{V} \setminus (Y \cup \mathbf{S}) = \boldsymbol{V}'$, matching the theorem's candidate set). By Theorem 1, $\mathbf{Z} = \mathbf{X} \cup \mathbf{X}'$ for some $\mathbf{X}' \in \mathcal{P}_{\leq k-|\mathbf{X}|}(\mathbf{S})$ such that $Y \not\perp\!\!\!\perp_{\hat{\mathbb{P}}} X \mid \mathbf{Z} \cup \mathbf{S}'$ holds for all $\mathbf{S}' \in \mathcal{P}_{\leq l}(\mathbf{S} \setminus \mathbf{X}')$. As the dependence set $\mathbf{X}$ is disjoint from $\mathbf{S}$, we have $\mathbf{S} \setminus \mathbf{X}' = \mathbf{S} \setminus \mathbf{Z}$, so $Y \not\perp\!\!\!\perp_{\hat{\mathbb{P}}} X \mid \mathbf{Z} \cup \mathbf{S}'$ holds for all $\mathbf{S}' \in \mathcal{P}_{\leq l}(\mathbf{S} \setminus \mathbf{Z})$ with $\mathbf{Z} \in \mathcal{P}_{\leq k}(\boldsymbol{V} \setminus \{Y, X\})$—the estimated l-bounded association claimed in the theorem. Since *make_strict* only removes variables from $\mathbf{Z}$, the returned set is $X \cup \mathbf{Z}_R$ for some $\mathbf{Z}_R \subseteq \mathbf{Z}$, as claimed.

Next consider a call that returns the empty set. Here every internal *find_cond* call must have failed, so *find_inter_dep* traverses the full enumeration tree of the dependence sets $\mathbf{X} \in \mathcal{P}_{\leq k}(\boldsymbol{V} \setminus (Y \cup \mathbf{S}))$ (lines 13–17 of *find_inter_dep*), and for each visits every candidate $X \in \boldsymbol{V} \setminus (Y \cup \mathbf{S} \cup \mathbf{X})$ through *find_cond*$(Y, X, \mathbf{X}, \mathbf{S}, k, l)$, which by Theorem 1 searches all $\mathbf{X}' \in \mathcal{P}_{\leq k-|\mathbf{X}|}(\mathbf{S})$. Every conditioning set $\mathbf{Z} \in \mathcal{P}_{\leq k}(\boldsymbol{V} \setminus \{Y, X\})$ decomposes uniquely into its part outside and its part inside the candidate blanket, $\mathbf{Z} = (\mathbf{Z} \setminus \mathbf{S}) \cup (\mathbf{Z} \cap \mathbf{S}) = \mathbf{X} \cup \mathbf{X}'$, so this enumeration covers every pair $(X, \mathbf{Z})$ with $X \in \boldsymbol{V}'$ and $\mathbf{Z} \in \mathcal{P}_{\leq k}(\boldsymbol{V} \setminus \{Y, X\})$. By Theorem 1, the call for $(X, \mathbf{X})$ returns $\perp$ exactly when every such $\mathbf{X}'$ admits a separator $\mathbf{S}' \in \mathcal{P}_{\leq l}(\mathbf{S} \setminus \mathbf{X}')$ with $Y \perp\!\!\!\perp_{\hat{\mathbb{P}}} X \mid \mathbf{X} \cup \mathbf{X}' \cup \mathbf{S}'$. Therefore *find_inter_dep* returns the empty set if and only if Equation (19) holds. The distinguished value $\perp$ lets the callers tell such a failure from a success that returns the empty conditioning set. □

Theorem 3 (Correctness of Algorithm 1)**.** *Given an estimated distribution $\hat{\mathbb{P}}$ over variables $\boldsymbol{V}$, the target variable whose graphical Markov blanket we want to discover $Y \in \boldsymbol{V}$, and the natural numbers $k, l \in \mathbb{N}$; the grow phase of* kOMB *then grows the candidate MB $\mathbf{S}$ such that,*

$$MB_{\mathcal{G}}(Y) \subseteq \mathbf{S}. \tag{20}$$

If, in addition, every true blanket member admits an in-blanket dependence witness—i.e. for every $X \in MB_{\mathcal{G}}(Y)$ and every $\mathbf{S}$ with $MB_{\mathcal{G}}(Y) \subseteq \mathbf{S} \subseteq \boldsymbol{V} \setminus Y$ there is a witness $\mathbf{Z} \in \mathcal{P}_{\leq k}(MB_{\mathcal{G}}(Y) \setminus X)$ with $asc^{(l)}_{\mathbf{S} \setminus (X \cup \mathbf{Z})}(Y, X \mid \mathbf{Z})$—then the shrink phase of kOMB *removes variables from $\mathbf{S}$ until,*

$$MB_{\mathcal{G}}(Y) = \mathbf{S}, \tag{21}$$

where it subsequently terminates and returns $\mathbf{S}$.

Proof. Throughout we assume that the CI decisions are accurate, so $\hat{\mathbb{P}}$ and $\mathbb{P}$ agree on every tested statement.

In the grow phase, every nonempty set returned by *find_inter_dep* contains a variable $X \notin \mathbf{S}$ by Theorem 2, so each iteration of the repeat loop strictly grows $\mathbf{S}$ and the phase terminates. Upon termination *find_inter_dep* returned the empty set, so by Theorem 2 the condition in Equation (19) holds in $\hat{\mathbb{P}}$, and by accuracy in $\mathbb{P}$. Fix any $X \in \boldsymbol{V} \setminus (\mathbf{S} \cup Y)$. For every $\mathbf{Z} \in \mathcal{P}_{\leq k}(\boldsymbol{V} \setminus \{Y, X\})$, Equation (19) supplies an $\mathbf{S}' \in \mathcal{P}_{\leq l}(\mathbf{S} \setminus \mathbf{Z}) \subseteq \mathcal{P}_{\leq l}(\mathbf{S})$ with $Y \perp\!\!\!\perp_{\mathbb{P}} X \mid \mathbf{Z} \cup \mathbf{S}'$; the l-bounded separator assumption (Assumption 3) then lifts

this, for every such $\mathbf{Z}$, to the unbounded statement that some $\mathbf{S}' \subseteq \mathbf{S}$ gives $Y \perp\!\!\!\perp_{\mathbb{P}} X \mid \mathbf{Z} \cup \mathbf{S}'$, which is the antecedent of k-order faithfulness (Equation (11)). Hence $Y \perp\!\!\!\perp_{\mathcal{G}} X \mid \mathbf{S}$ for every $X \in \boldsymbol{V} \setminus (\mathbf{S} \cup Y)$. Since the parents and children of Y—and, for an undirected $\mathcal{G}$, its neighbours—can never be separated from Y, they must all lie in $\mathbf{S}$. Any spouse X shares a child $C \in \mathrm{ch}(Y) \subseteq \mathbf{S}$, so the collider path $Y \to C \leftarrow X$ is active given $\mathbf{S}$; hence $Y \not\perp\!\!\!\perp_{\mathcal{G}} X \mid \mathbf{S}$, which forces $X \in \mathbf{S}$ as well. Therefore $\mathrm{MB}_{\mathcal{G}}(Y) \subseteq \mathbf{S}$, establishing Equation (20).

During the shrink phase we maintain the invariant $\mathrm{MB}_{\mathcal{G}}(Y) \subseteq \mathbf{S}$, which holds on entry by Equation (20). First, we show that no true member is ever removed. For $X \in \mathrm{MB}_{\mathcal{G}}(Y)$, the in-blanket witness hypothesis of Theorem 3 supplies a set $\mathbf{Z} \in \mathcal{P}_{\leq k}(\mathrm{MB}_{\mathcal{G}}(Y) \setminus X) \subseteq \mathcal{P}_{\leq k}(\mathbf{S} \setminus X)$ such that $Y \not\perp\!\!\!\perp_{\mathbb{P}} X \mid \mathbf{Z} \cup \mathbf{S}'$ holds for all $\mathbf{S}' \in \mathcal{P}_{\leq l}((\mathbf{S} \setminus X) \setminus \mathbf{Z})$. By accuracy and Theorem 1 (with candidate blanket $\mathbf{S} \setminus X$, external dependence set $\mathbf{X} = \emptyset$, and $\mathbf{X}' = \mathbf{Z}$), the call *find_cond*$(Y, X, \emptyset, \mathbf{S} \setminus X, k, l)$ does not return $\perp$, so X is never selected for removal (a member whose only witness is $\mathbf{Z} = \emptyset$ makes *find_cond* return the empty set rather than $\perp$).

Next, we show that every non-member is removed once selected. Let $X \in \mathbf{S} \setminus \mathrm{MB}_{\mathcal{G}}(Y)$; by the invariant $\mathrm{MB}_{\mathcal{G}}(Y) \subseteq \mathbf{S} \setminus X$. We first observe that $\mathrm{MB}_{\mathcal{G}}(Y)$ separates Y from X even after conditioning on any further set $\mathbf{Z} \subseteq \boldsymbol{V} \setminus \{Y, X\}$. Since $X \notin \mathrm{MB}_{\mathcal{G}}(Y)$, X is not adjacent to Y, so every path between them has at least one interior vertex; write such a path as $Y = N_0 \sim N_1 \sim \cdots \sim N_m = X$ with $m \geq 2$. Its first interior vertex N_1 is adjacent to Y and hence lies in $\mathrm{MB}_{\mathcal{G}}(Y)$. If N_1 is a non-collider on the path, it blocks the path. If instead N_1 is a collider $Y \to N_1 \leftarrow N_2$, then N_1 is a common child of Y and its successor N_2 on the path, so N_2 is a spouse of Y and hence $N_2 \in \mathrm{MB}_{\mathcal{G}}(Y)$. In particular $N_2 \neq X$ (as $X \notin \mathrm{MB}_{\mathcal{G}}(Y)$), so N_2 is a genuine interior vertex; and since the edge $N_2 \to N_1$ points out of N_2, it is a non-collider on the path and blocks it. For an undirected $\mathcal{G}$, N_1 is instead a neighbour of Y and blocks the path directly. As each path is thus blocked at a conditioned vertex of $\mathrm{MB}_{\mathcal{G}}(Y)$ irrespective of $\mathbf{Z}$ [Pearl, 1988, Lauritzen, 1996], $Y \perp\!\!\!\perp_{\mathcal{G}} X \mid \mathrm{MB}_{\mathcal{G}}(Y) \cup \mathbf{Z}$ for every $\mathbf{Z} \in \mathcal{P}_{\leq k}(\boldsymbol{V} \setminus \{Y, X\})$.

The Global Markov Property transfers this to $\mathbb{P}$ with the separator $\mathbf{S}' := \mathrm{MB}_{\mathcal{G}}(Y) \setminus \mathbf{Z} \subseteq \mathbf{S} \setminus X$, so the unbounded side of Assumption 3—at the separating set $\mathbf{S} \setminus X$—holds for every such $\mathbf{Z}$. The biconditional then yields some $\mathbf{S}' \in \mathcal{P}_{\leq l}(\mathbf{S} \setminus X)$; discarding any overlap with $\mathbf{Z}$ leaves $\mathbf{Z} \cup \mathbf{S}'$ unchanged and gives $\mathbf{S}' \in \mathcal{P}_{\leq l}((\mathbf{S} \setminus X) \setminus \mathbf{Z})$ with $Y \perp\!\!\!\perp_{\mathbb{P}} X \mid \mathbf{Z} \cup \mathbf{S}'$. By accuracy and Theorem 1, *find_cond*$(Y, X, \emptyset, \mathbf{S} \setminus X, k, l)$ returns $\perp$, so X is removed once selected.

Each iteration of the while loop removes one variable, so the phase terminates; it exits only when no $X \in \mathbf{S}$ yields $\perp$, whence $\mathbf{S} \setminus \mathrm{MB}_{\mathcal{G}}(Y) = \emptyset$ and, with the invariant, $\mathbf{S} = \mathrm{MB}_{\mathcal{G}}(Y)$, establishing Equation (21). □

D COMPLEXITY OF *KOMB*

Corollary 1 (Computational Complexity of *kOMB*)**.** *Let $n = |\boldsymbol{V}|$ denote the number of variables, N the sample size, and K the maximum size of any candidate MB $\mathbf{S}$ encountered by Algorithm 1 during execution. Furthermore assume each CI test runs in $O(N)$ time. Then, for a given order k and separator bound l,* kOMB *(Algorithm 1) terminates in worst-case time*

$$O\left(K\, n \sum_{i=0}^{k} \binom{n}{i} \left[\sum_{i'=0}^{k-i} \binom{K}{i'} \sum_{j=0}^{l} \binom{K-i'}{j} N\right]\right), \tag{22}$$

where we adopt the convention that $\binom{n}{m} = 0$ whenever $n < m$.

Proof. We charge $O(N)$ to each CI test and bound the running time by analysing the three sub-algorithms of Section 5 from the innermost outwards.

First consider *find_cond* (Algorithm 3) for a fixed candidate variable X and external dependence set $\mathbf{X}$ with $|\mathbf{X}| = i$. It iterates over every subset $\mathbf{X}' \in \mathcal{P}_{\leq k-i}(\mathbf{S})$, of which there are at most $\sum_{i'=0}^{k-i} \binom{K}{i'}$ since $|\mathbf{S}| \leq K$. For each $\mathbf{X}'$ it performs one CI test and, whenever that test indicates a dependence, calls *no_seps*, which iterates over every separator $\mathbf{S}' \in \mathcal{P}_{\leq l}(\mathbf{S} \setminus \mathbf{X}')$ and performs one further CI test per separator. Writing $|\mathbf{X}'| = i'$, there are at most $\sum_{j=0}^{l} \binom{K-i'}{j}$ such separators, so a single call to *find_cond* costs

$$O\left(\sum_{i'=0}^{k-i} \binom{K}{i'} \sum_{j=0}^{l} \binom{K-i'}{j} N\right), \tag{24}$$

where the convention $\binom{n}{m} = 0$ for $n < m$ accounts for the separator set $\mathbf{S} \setminus \mathbf{X}'$ being exhausted.

Next consider *find_inter_dep* (Algorithm 2). It explores the enumeration tree of $\boldsymbol{V} \setminus Y$ up to subsets of cardinality k (lines 13–17 of *find_inter_dep*), and hence visits at most $\sum_{i=0}^{k} \binom{n}{i}$ dependence sets $\mathbf{X} \in \mathcal{P}_{\leq k}(\boldsymbol{V} \setminus Y)$ (a loose bound on the exact count $\sum_{i=0}^{k} \binom{n-1}{i}$, as $|\boldsymbol{V} \setminus Y| = n - 1$). For every such $\mathbf{X}$ it loops over the $O(n)$ candidate variables $X \in \boldsymbol{V} \setminus (Y \cup \mathbf{S} \cup \mathbf{X})$ and invokes *find_cond*. Multiplying these two factors by Equation (24) bounds the cost of a single association-mining sweep by

$$O\left(n \sum_{i=0}^{k} \binom{n}{i} \left[\sum_{i'=0}^{k-i} \binom{K}{i'} \sum_{j=0}^{l} \binom{K-i'}{j} N\right]\right). \tag{25}$$

Each successful sweep additionally invokes *make_strict* exactly once; it iterates over the found dependence set $\mathbf{Z}$—of size at most $\min(k, n-2)$, since $\mathbf{Z} \subseteq \boldsymbol{V} \setminus \{Y, X\}$ and $|\mathbf{Z}| \leq k$—and performs a single *no_seps* call per element, at cost $O\left(\min(k, n-2) \sum_{j=0}^{l} \binom{K}{j} N\right)$. This is dominated by the per-sweep *find_cond* total already counted in Equation (25)—whose bracketed factor already contains the

term $\sum_{j=0}^{l} \binom{K}{j}$ (at $i' = 0$) multiplied by the outer factor $n \sum_{i=0}^{k} \binom{n}{i} \geq n \geq \min(k, n-2)$—and is therefore absorbed into the bound.

Finally, consider *kOMB* itself (Algorithm 1). Each successful call to *find_inter_dep* in the grow phase returns a nonempty set $X \cup \mathbf{Z}$ that is added to $\mathbf{S}$, so the candidate Markov blanket grows by at least one variable per successful call; as $|\mathbf{S}| \leq K$ throughout, the grow phase performs at most $O(K)$ association-mining sweeps. The shrink phase re-evaluates its guard after each removal, calling *find_cond* at most $O(K^2)$ times over the at most K removals; as $K \leq n-1$, this cost is dominated by that of the grow phase. Multiplying Equation (25) by these $O(K)$ sweeps yields the bound in Equation (22), completing the proof. □

E BOOLEAN FUNCTIONS FOR THE SYNTHETIC BAYESIAN NETWORKS

$$\begin{aligned}
\textbf{parity:}\quad & f(X_1, \ldots, X_3) = (\textstyle\sum_{i=1}^{3} X_i) \pmod 2 \\
\textbf{ex-}s\textbf{:}\quad & f(X_1, \ldots, X_3; s) = \left[(\textstyle\sum_{i=1}^{3} X_i) = s\right] \\
\textbf{and:}\quad & f(X_1, \ldots, X_3) = \left[(\textstyle\sum_{i=1}^{3} X_i) = 3\right] \\
\textbf{or:}\quad & f(X_1, \ldots, X_3) = \left[(\textstyle\sum_{i=1}^{3} X_i) \geq 1\right]
\end{aligned}$$

where $s \in \{1, 2\}$.

F SEPARATOR SIZES IN THE BENCHMARK DATASETS

To gauge how restrictive the l-bounded separator assumption (Assumption 3) is in practice, we examine the benchmark networks of Section 6.2. We use every variable in turn as a target and, for every pair formed by a target and a variable outside its Markov blanket, compute a minimal d-separator drawn from the true Markov blanket. Table 3 reports the mean and worst-case (maximum) separator size and the fraction of pairs whose separator has size at most three. Separators are small across all four networks—the mean never exceeds 2.19 and size-≤ 3 separators cover at least 83% of pairs—and no pair had to be skipped, i.e. a separator within the true Markov blanket always existed. The bound is loosest on Barley, where the largest separator reaches size 9; this is also the network on which *kOMB* is the most expensive and least accurate (Section 6.2), illustrating that the difficulty of MB discovery tracks the separator sizes the assumption must accommodate. Overall, a modest separator bound l already suffices to capture most of the conditional independencies needed to recover the Markov blanket on these networks.

Table 3: Minimal d-separator sizes restricted to the true Markov blanket, computed using every variable of each benchmark network as a target. "Targets" is the number of targets (all variables) and "Pairs" the number of target–non-blanket pairs; "Mean sep. size" and "Max" are the mean and maximum separator size; and "Cover ≤ 3" is the fraction whose minimal separator has size at most three.

Network	Targets	Pairs	Mean sep. size	Max	Cover ≤ 3
Alarm1	37	1202	0.621	3	1.000
Barley	48	2004	1.739	9	0.835
Insurance	27	562	2.190	7	0.925
Mildew	35	1030	1.065	6	0.976

G ABLATION

For completeness, we report the full results of the experiments in Section 6.2 for every method and every (k, l) configuration of *kOMB*. Table 5 gives the performance of all methods, while Table 6 gives their runtimes in seconds. In both tables, "–" denotes a configuration that failed to complete within the allotted time budget.

Table 4 summarises the F1 score of *kOMB* across all synthetic and benchmark datasets for every (k, l) configuration. F1 generally improves from $k = 0$ to $k = 2$—higher orders expose dependencies that lower orders miss—but the gains flatten or reverse beyond $k = 2$ as the larger conditioning sets incur both higher runtime (Table 6) and more error-prone CI tests. Setting $l = k$ with $k \leq 2$ therefore offers a robust default, the heuristic we adopt in Section 5.2. For a fixed k, raising l beyond k does not consistently improve F1 (Table 4), which motivates simply setting $l = k$ rather than tuning l separately.

Table 4: F1 score across datasets for each (k, l) configuration of *kOMB*.

Method	parity	ex-1	ex-2	and	or	Alarm1	Barley	Insurance	Mildew
kOMB	0.0250	0.2825	0.2675	0.4100	0.3500	–	–	–	–
($k = 0, l = 1$)	±0.1104	±0.3720	±0.3737	±0.4050	±0.3508	–	–	–	–
kOMB	0.0250	0.2825	0.2675	0.4100	0.3500	–	–	–	–
($k = 0, l = 2$)	±0.1104	±0.3720	±0.3737	±0.4050	±0.3508	–	–	–	–
kOMB	0.0250	0.2825	0.2675	0.4100	0.3500	–	–	–	–
($k = 0, l = 3$)	±0.1104	±0.3720	±0.3737	±0.4050	±0.3508	–	–	–	–
kOMB	0.0500	0.6750	0.6750	**0.6175**	0.5600	0.7804	**0.5158**	0.7136	**0.6220**
($k = 1, l = 1$)	±0.2207	±0.4743	±0.4743	±0.3741	±0.4349	±0.1145	±0.1964	±0.1328	±0.1595
kOMB	0.0500	0.6750	0.6750	**0.6175**	0.5600	0.8187	0.3128	**0.7336**	0.4360
($k = 1, l = 2$)	±0.2207	±0.4743	±0.4743	±0.3741	±0.4349	±0.1944	±0.1582	±0.1624	±0.1394
kOMB	0.0500	0.6750	0.6750	**0.6175**	0.5600	0.7923	0.2498	0.7076	0.3726
($k = 1, l = 3$)	±0.2207	±0.4743	±0.4743	±0.3741	±0.4349	±0.1885	±0.1140	±0.1304	±0.1558
kOMB	**1.0000**	**1.0000**	**1.0000**	**0.6175**	**0.7100**	0.7775	0.4634	0.7249	0.5059
($k = 2, l = 1$)	±0.0000	±0.0000	±0.0000	±0.3741	±0.3842	±0.1511	±0.1774	±0.1464	±0.1458
kOMB	**1.0000**	**1.0000**	**1.0000**	**0.6175**	**0.7100**	**0.8209**	0.3128	**0.7336**	0.4360
($k = 2, l = 2$)	±0.0000	±0.0000	±0.0000	±0.3741	±0.3842	±0.1914	±0.1582	±0.1624	±0.1394
kOMB	**1.0000**	**1.0000**	**1.0000**	**0.6175**	**0.7100**	0.7967	0.2498	0.7058	0.3726
($k = 2, l = 3$)	±0.0000	±0.0000	±0.0000	±0.3741	±0.3842	±0.1828	±0.1140	±0.1311	±0.1558
kOMB	–	–	–	–	–	0.7721	–	0.7227	0.5243
($k = 3, l = 1$)	–	–	–	–	–	±0.1599	–	±0.1552	±0.1075
kOMB	–	–	–	–	–	0.8001	–	0.7288	0.4194
($k = 3, l = 2$)	–	–	–	–	–	±0.1948	–	±0.1471	±0.1414
kOMB	–	–	–	–	–	0.7967	–	0.7058	0.3726
($k = 3, l = 3$)	–	–	–	–	–	±0.1828	–	±0.1311	±0.1558

Table 5: Full results on benchmark datasets.

Method	Alarm1	Barley	Insurance	Mildew
BAMB	0.6742	0.3385	0.6588	0.4945
	±0.1771	±0.1226	±0.0797	±0.1084
GS	0.3434	0.2113	0.4615	0.2867
	±0.1503	±0.0807	±0.1477	±0.1581
HITON_MB	0.7636	0.3400	0.6746	0.4918
	±0.1861	±0.1284	±0.1594	±0.1487
IAMB	0.6931	0.3373	0.6293	0.4677
	±0.1827	±0.1223	±0.0778	±0.1224
LRH	0.6813	0.3373	0.6281	0.4674
	±0.1777	±0.1223	±0.0782	±0.1230
MMMB	0.7692	0.3298	0.6978	0.4861
	±0.1874	±0.1274	±0.1405	±0.1810
PCMB	0.7371	0.2284	0.6299	0.4156
	±0.2127	±0.1199	±0.1291	±0.1445
STMB	0.6598	0.2145	0.5648	0.2734
	±0.1627	±0.1281	±0.1324	±0.1765
kOMB	0.7804	**0.5158**	0.7136	**0.6220**
$(k = 1, l = 1)$	±0.1145	±0.1964	±0.1328	±0.1595
kOMB	0.8187	0.3128	**0.7336**	0.4360
$(k = 1, l = 2)$	±0.1944	±0.1582	±0.1624	±0.1394
kOMB	0.7923	0.2498	0.7076	0.3726
$(k = 1, l = 3)$	±0.1885	±0.1140	±0.1304	±0.1558
kOMB	0.7775	0.4588	0.7249	0.5059
$(k = 2, l = 1)$	±0.1511	±0.1825	±0.1464	±0.1458
kOMB	**0.8209**	0.3128	**0.7336**	0.4360
$(k = 2, l = 2)$	±0.1914	±0.1582	±0.1624	±0.1394
kOMB	0.7967	0.2498	0.7058	0.3726
$(k = 2, l = 3)$	±0.1828	±0.1140	±0.1311	±0.1558
kOMB	0.7721	0.0000	0.7227	0.0419
$(k = 3, l = 1)$	±0.1599	±0.0000	±0.1552	±0.1458
kOMB	0.8001	0.0000	0.7288	0.3607
$(k = 3, l = 2)$	±0.1948	±0.0000	±0.1471	±0.1964
kOMB	0.7967	0.0000	0.7058	0.3726
$(k = 3, l = 3)$	±0.1828	±0.0000	±0.1311	±0.1558

Table 6: Full runtimes (in seconds) on benchmark datasets.

Method	Alarm1	Barley	Insurance	Mildew
BAMB	3.517	1.149	5.615	0.797
	±4.305	±0.681	±5.358	±0.392
GS	4.875	1.575	4.433	1.035
	±5.183	±0.606	±3.792	±0.341
HITON_MB	2.560	1.027	3.514	0.604
	±1.277	±0.298	±2.525	±0.218
IAMB	2.527	1.104	1.650	0.943
	±1.710	±0.496	±0.967	±0.420
LRH	89.877	5.071	45.401	1.833
	±93.285	±2.496	±24.653	±0.651
MMMB	2.724	1.090	3.508	0.553
	±1.278	±0.374	±2.525	±0.128
PCMB	15.515	4.982	27.213	3.066
	±10.203	±1.872	±21.695	±1.486
STMB	5.305	1.895	14.658	0.561
	±5.855	±2.367	±33.396	±0.524
kOMB	5.258	58.190	4.856	17.698
$(k = 1, l = 1)$	±2.921	±31.619	±2.041	±10.205
kOMB	3.974	12.369	3.828	7.562
$(k = 1, l = 2)$	±2.468	±7.709	±2.235	±2.301
kOMB	4.111	9.639	3.171	7.384
$(k = 1, l = 3)$	±2.791	±3.794	±1.707	±2.127
kOMB	26.092	788.531	19.901	311.324
$(k = 2, l = 1)$	±14.668	±453.059	±7.226	±167.079
kOMB	18.144	335.821	14.493	121.909
$(k = 2, l = 2)$	±7.566	±135.966	±7.280	±33.153
kOMB	18.016	290.512	10.424	121.146
$(k = 2, l = 3)$	±9.135	±85.391	±4.290	±32.239
kOMB	174.900	1800.000	96.149	1761.246
$(k = 3, l = 1)$	±83.504	±0.000	±46.913	±188.357
kOMB	132.619	1800.000	58.350	1534.465
$(k = 3, l = 2)$	±46.566	±0.000	±21.745	±300.944
kOMB	127.861	1800.000	53.452	1356.514
$(k = 3, l = 3)$	±49.652	±0.000	±18.285	±304.778