

Recovering Causal Structures from Low-Order Conditional Independencies

Marcel Wienöbst, Maciej Liśkiewicz

Institute of Theoretical Computer Science, University of Lübeck, Germany
 {wienoebst, liskiewi}@tcs.uni-luebeck.de

Abstract

One of the common obstacles for learning causal models from data is that high-order conditional independence (CI) relationships between random variables are difficult to estimate. Since CI tests with conditioning sets of low order can be performed accurately even for a small number of observations, a reasonable approach to determine casual structures is to base merely on the low-order CIs. Recent research has confirmed that, e.g. in the case of sparse true causal models, structures learned even from zero- and first-order conditional independencies yield good approximations of the models. However, a challenging task here is to provide methods that faithfully explain a given set of low-order CIs. In this paper, we propose an algorithm which, for a given set of conditional independencies of order less or equal to k , where k is a small fixed number, computes a faithful graphical representation of the given set. Our results complete and generalize the previous work on learning from pairwise marginal independencies. Moreover, they enable to improve upon the 0-1 graph model which, e.g. is heavily used in the estimation of genome networks.

1 Introduction

Graphical models, as e.g. directed acyclic graphs (DAGs), allow an intuitive and mathematically sound approach to analyze complex causal mechanisms (Lauritzen 1996; Pearl 2009). Generally, they encode the causal links between variables of interests based on conditional independence (CI) statements between the variables (Spirtes, Glymour, and Scheines 2000). Hence, the accuracy of estimate of the CIs plays a key role in learning graphical models and consequently in causal inference from observational data.

CI testing is a challenging task, particularly in the presence of high-order independencies, when the number of variables far exceeds the number of observations (Wille and Bühlmann 2006). In such cases, estimations of CIs are usually inaccurate, potentially resulting in incorrect links between variables in the graphical model. On the other hand, CI tests with conditioning sets of low dimension can be performed accurately even for relatively small observed data sets. Thus, a natural task is to approximate the true causal

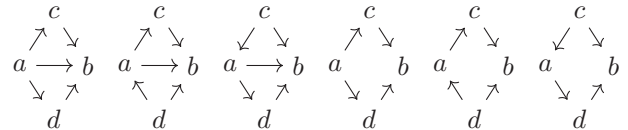


Figure 1: All 1-faithful DAGs for the vertex set $\{a, b, c, d\}$ and the single CI statement $(c \perp d | a)$.

model using merely low-order CIs. Recent research in inferring genetic networks has confirmed the effectiveness of this approach when basing only on zero- and first-order independencies (Wille et al. 2004; Magwene and Kim 2004).

In this paper, we systematically study the problem to extract as much “causal knowledge” as possible from CI statements of order at most k , where $k \geq 0$ is a (typically small) integer. More precisely, we investigate the following task: For a set of variables V and a given set \mathcal{I} of CI statements of the form $(a \perp b | Z)$, with $a, b \in V$, $Z \subseteq V$, and $|Z| \leq k$, find all DAGs D which encode up to order k exactly the CIs in \mathcal{I} , i.e., such that for all a, b, Z , with $|Z| \leq k$, it is true that a and b are d -separated by Z in D if and only if $(a \perp b | Z)$ is in \mathcal{I} . We will call such DAGs k -faithful to \mathcal{I} (for formal definitions, see Section 3). Figure 1 illustrates all DAGs which are 1-faithful to a single CI statement $(c \perp d | a)$ for the vertex set $V = \{a, b, c, d\}$.

We observe that this is a generalization of several problems already studied in the literature. For the simplest case $k = 0$, the CI statements are marginal independencies and the 0-faithful DAGs are called *faithful to pairwise marginal independencies*. The problem of deciding if a 0-faithful DAG exists for a given set of CIs of order zero, represented as an undirected graph, has been studied in (Pearl and Verma 1994; Textor, Idelberger, and Liśkiewicz 2015).

Next, with n denoting the cardinality of V , the problem for $k = n - 2$ was first investigated by Verma and Pearl (1992). They called such k -faithful DAGs just *faithful* ones and presented an algorithm which, for a given \mathcal{I} , tests for the existence of a DAG faithful to \mathcal{I} and produces a representation of all such DAGs encoded in form of a completed partially directed acyclic graph (CPDAG) (we recall all used graphical notions in Section 2).

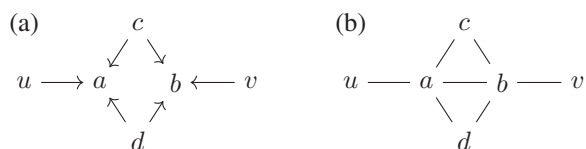


Figure 2: In (a) the underlying true DAG D is displayed. In (b) we show the skeleton of the CPDAG computed by the PC algorithm restricted to CI tests of order zero and one.

Note the important difference between the k -faithful and faithful DAGs. Even if \mathcal{I} consists of CI statements of order $\leq k$, these two notions differ considerably. E.g., for the CI statements \mathcal{I} of order zero and one induced by the underlying DAG D shown in Fig. 2(a) the only 1-faithful DAG is D itself, while no faithful DAG to such \mathcal{I} exists. This is because for a 1-faithful DAG, the CIs of order > 1 are irrelevant, while a faithful DAG takes that $(x \not\perp\!\!\!\perp y \mid Z)$, for all x, y , and Z , with $|Z| > 1$.

We also notice that one cannot construct a k -faithful DAG just using a constraint based structure learning algorithm, as the SGS or the PC algorithm (Spirtes, Glymour, and Scheines 2000; Kalisch and Bühlmann 2007), restricting the CI tests to independencies of order $\leq k$. For example, for the underlying true DAG shown in Fig. 2(a) such an approach returns a structure with the skeleton given in Fig. 2(b). It is analyzed in detail in Section 4 why the superfluous edge $a - b$ is included in the result of classical causal structure learning algorithms and through which rule we are able to remove it.

Previous Work. Pearl and Wermuth (1994) investigated the problem whether a set of marginal independencies \mathcal{I} has a causal interpretation – meaning a DAG faithful to \mathcal{I} . Moreover, they proposed an algorithm to construct a faithful DAG, but in their paper they did not give proofs for their theorems. Textor, Idelberger, and Liškiewicz (2015) further considered the stated problem, characterizing the DAG-representable sets by graph theoretical properties of the marginal independence graphs (these are undirected graphs with an edge between a and b iff¹ a and b are marginally dependent). Additionally, they proposed an algorithm which is based on the construction by Pearl and Wermuth (1994). However, they did not provide the missing proofs.

Other works have considered the more general setting which includes conditional independencies with a singleton conditioning set on top of marginal independencies. In this context, de Campos and Huete (2000) introduced the notion of a 0-1 graph. This is an *undirected* graph which contains an edge $a - b$ iff $(a \not\perp\!\!\!\perp b) \wedge [\forall c : (a \not\perp\!\!\!\perp b \mid c)]$. In other words, we obtain the graph by removing all edges between nodes for which we find an independence of order zero or one.

Wille and Bühlmann (2006) showed that – in the case of graphical Gaussian models – the 0-1 graphs are good estimators of sparse graphical models and relevant in biological applications. In particular, they have been used

¹We use *iff* as shorthand for *if and only if*.

to model genome networks (De la Fuente et al. 2005; Magwene and Kim 2004; Wille et al. 2004). Later, Castelo and Roverato (2006) generalized the 0-1 graph and the covariance graphs (Cox and Wermuth 1993) to so called q -partial graphs.

Our Results. We provide a constructive solution to the problem of deciding if, for a given set \mathcal{I} of CIs of order less or equal to k , there exists a DAG which is k -faithful to \mathcal{I} . We propose an algorithm called LOCI (Low-Order Causal Inference) which – in case a k -faithful DAG exists – outputs all such DAGs encoded in form of a CPDAG. This extends and generalizes previously known results by Pearl and Wermuth (1994) as well as by Textor, Idelberger, and Liškiewicz (2015) who provided solutions only for sets of marginal independencies, i.e. for $k = 0$. Moreover, the analysis for the correctness of the construction given in this paper, fills the gaps in the proofs by Pearl and Wermuth, and by Textor, Idelberger, and Liškiewicz.

The proposed approach also improves some other methods known in the literature to learn DAGs from CIs up to a fixed order k . In particular, it improves the algorithm by De Campos and Huete (2000) that presupposes knowledge of the topological sorting of nodes in the underlying DAG. In contrast, no such knowledge is assumed in our algorithm.

Structure of the paper. In the following section we introduce all preliminary definitions. Afterwards, in Section 3, we formally define what faithfulness to a set of CIs means. In Sections 4 and 5 we derive an algorithm for finding a compact and faithful representation of a set of low-order independencies. We experimentally compare this algorithm to previous approaches in Section 6. Finally, we discuss our results in Section 7.

2 Preliminaries

We consider directed and partially directed graphs $G = (V, E)$ with $|V| = n$. In the latter case, a graph has both directed $a \rightarrow b$ and undirected $c - d$ edges. Two nodes a and b are called *adjacent* if there is an edge between them (directed or undirected). The *degree* of a node a is the number of nodes adjacent to a . For an edge $a \rightarrow b$ we call a the *parent* of b and b the *child* of a . A *way* is a sequence p_0, \dots, p_t of nodes so that for all i , with $0 \leq i < t$, there is an edge connecting p_i and p_{i+1} . Such a sequence is called a *path* if $p_i \neq p_j$ holds for all i, j , with $0 \leq i < j \leq t$. A path from p_0 to p_t is called *causal* if every edge on the path is directed from p_i towards p_{i+1} . A node b is called an *ancestor* of a if there is a causal path from b to a . A node b is called a *descendant* of a if there is a causal path from a to b . $\text{An}_G(a)$ is the set of all ancestors of a in graph G , $\text{De}_G(a)$ is the set of all descendants of a in G . We use small letters for nodes and values, and capital letters for sets and random variables.

Of special importance are directed acyclic graphs (DAGs) containing only directed edges and no directed cycles, and partially directed acyclic graphs (PDAGs) that may contain both directed and undirected edges but no directed cycles. Every DAG is a PDAG. The *skeleton* of a PDAG G is the

undirected graph where every edge in G is substituted by an undirected edge.

Let P be a joint probability distribution over random variables X_i , with $i \in V$, and X, Y and Z stand for any subsets of variables. We use the notation $(X \perp\!\!\!\perp Y \mid Z)_P$ to state that X is independent of Y given Z in P . A distribution P and a DAG $D = (V, E)$ are called *compatible* if D factorizes P as $\prod_{i \in V} P(x_i \mid pa_i)$ over all realizations x_i of X_i and pa_i of variables corresponding to the parents of i in D . It is possible to read CIs over X_i , with $i \in V$, off a compatible DAG through the notion of d -separation. Recall, a path π is said to be d -separated (or *blocked*) by a set of nodes Z iff (1.) π contains a chain $u \rightarrow v \rightarrow w$ or $u \leftarrow v \leftarrow w$ or a fork $u \leftarrow v \rightarrow w$ such that the middle node v is in Z , or (2.) π contains an inverted fork (or *collider*) $u \rightarrow v \leftarrow w$ such that the middle node v is not in Z and such that no descendant of v is in Z . A set Z is said to d -separate a from b iff Z blocks every path from a to b . We write $(a \perp\!\!\!\perp b \mid Z)_D$ when a and b are d -separated by Z in D . Whenever G and P are compatible, it holds for all $a, b \in V$, and $Z \subseteq V$, that if $(a \perp\!\!\!\perp b \mid Z)_D$ then $(X_a \perp\!\!\!\perp X_b \mid \{X_i : i \in Z\})_P$.

An inverted fork $u \rightarrow v \leftarrow w$ is called a v -structure if u and w are not adjacent. A *pattern* of a DAG D is the PDAG which has the same skeleton as D and which has an oriented edge $a \rightarrow b$ iff there is a vertex c , which is not adjacent to a , such that $c \rightarrow b$ is an edge in D , too. Essentially, in the pattern of D , the only directed edges are the ones which are part of a v -structure in D .

A special case of PDAGs are the so called CPDAGs (Andersson, Madigan, and Perlman 1997) or completed partially directed graphs. They represent Markov equivalence classes. If two DAGs are Markov equivalent, it means that every probability distribution that is compatible with one of the DAGs is also compatible with the other (Pearl 2009). As shown by Verma and Pearl (1990) two DAGs are Markov equivalent *iff* they have the same skeleton and the same v -structures.

Given a DAG $D = (V, E)$, the class of Markov equivalent graphs to D , denoted as $[D]$, is defined as $[D] = \{D' \mid D' \text{ is Markov equivalent to } D\}$. The graph representing $[D]$ is called a CPDAG and is denoted as $D^* = (V, E^*)$, with the set of edges defined as follows: $a \rightarrow b$ is in E^* if $a \rightarrow b$ belongs to every $D' \in [D]$ and $a - b$ is in E^* if there exist $D', D'' \in [D]$ so that $a \rightarrow b$ is an edge of D' and $a \leftarrow b$ is an edge of D'' . A partially directed graph G is called a CPDAG if $G = D^*$ for some DAG D .

Given a partially directed graph G , a DAG D is an extension of G iff G and D have the same skeleton and if $a \rightarrow b$ is in G , then $a \rightarrow b$ is in D . An extension is called *consistent* if additionally G and D have the same v -structures. Due to Meek (1995, Theorem 3), we know that when starting with a pattern G of some DAG D and repeatedly executing the following three rules until none of them applies, we obtain a CPDAG D^* representing the Markov equivalent DAGs:

1. Orient $b - c$ into $b \rightarrow c$ if there is $a \rightarrow b$ such that a and c are nonadjacent.
2. Orient $a - c$ into $a \rightarrow c$ if there is a chain $a \rightarrow b \rightarrow c$.
3. Orient $a - b$ into $a \rightarrow b$ if there are two chains $a - c \rightarrow b$ and $a - d \rightarrow b$ such that c and d are nonadjacent.

We will call these three rules the *Meek rules*.

We note that one obtains the CPDAG D^* by applying the rules not only when starting with the pattern of a DAG D but also, more generally, when the initial graph G is any PDAG whose consistent extensions form a Markov equivalence class $[D]$. We will use this property in the correctness proof of the LOCI algorithm (Algorithm 1).

3 Models Faithful to CI Statements

In this section, we give a formal definition for a k -faithful DAG and – for the sake of completeness – we recall the definitions of a faithful and a k -partial graph. Next, we propose a definition for a compact representation of all k -faithful DAGs in terms of PDAGs and show that it yields a CPDAG.

Let V represent the set of variables and $k \geq 0$ be a fixed integer. Let \mathcal{I}_V be a set of CI statements over variables X_i , with $i \in V$, given as $(a \perp\!\!\!\perp b \mid Z)$, with $a, b \in V$ and $Z \subseteq V$. Analogously, let \mathcal{I}_V^k be a set of CI statements of order $\leq k$, i.e. such that $|Z| \leq k$. For example, the set \mathcal{I}_V^0 solely contains marginal independencies. For a more consistent notation we write $(a \perp\!\!\!\perp b \mid Z)_{\mathcal{I}_V^k}$ instead of $(a \perp\!\!\!\perp b \mid Z) \in \mathcal{I}_V^k$, and respectively, $(a \not\perp\!\!\!\perp b \mid Z)_{\mathcal{I}_V^k}$ for $(a \perp\!\!\!\perp b \mid Z) \notin \mathcal{I}_V^k$. We use an analogous notation for \mathcal{I}_V . Additionally, in statements like e.g. $(a \perp\!\!\!\perp b \mid \{c, d\})$, we omit the brackets and write $(a \perp\!\!\!\perp b \mid c, d)$.

Definition 1 (Faithful Graph (Verma and Pearl 1990)). *For a set \mathcal{I}_V of CIs, a DAG $D = (V, E)$ is called faithful to \mathcal{I}_V if*

$$\forall(a, b, Z) \quad [(a \perp\!\!\!\perp b \mid Z)_{\mathcal{I}_V} \Leftrightarrow (a \perp\!\!\!\perp b \mid Z)_D].$$

Definition 2 (k -Partial Graph (Castelo and Roverato 2006)). *For a set \mathcal{I}_V^k of CIs of order $\leq k$, an undirected graph $G = (V, E)$ is called a k -partial graph with respect to \mathcal{I}_V^k if*

$$(\forall a, b, Z, |Z| \leq k) \quad [(a \perp\!\!\!\perp b \mid Z)_{\mathcal{I}_V^k} \Leftrightarrow (a - b \notin E)].$$

We will call k -partial graphs with $k = 1$ also 0-1 graphs, as proposed by Wille and Bühlmann (2006) who considered such structures in the context of graphical Gaussian models.

Definition 3 (k -Faithful Graph). *For a set \mathcal{I}_V^k of CIs of order $\leq k$, a DAG $D = (V, E)$ is called k -faithful to \mathcal{I}_V^k if*

$$(\forall a, b, Z, |Z| \leq k) \quad [(a \perp\!\!\!\perp b \mid Z)_{\mathcal{I}_V^k} \Leftrightarrow (a \perp\!\!\!\perp b \mid Z)_D].$$

Due to Verma and Pearl (1990), we know that, for a given set \mathcal{I}_V , all DAGs faithful to \mathcal{I}_V can be represented as a CPDAG over V . A representation of a k -partial graph follows straightforwardly from the definition. On the other hand, note that it is not obvious how to represent all DAGs which are k -faithful to \mathcal{I}_V^k , like e.g. those shown in Fig. 1.

Definition 4. *A set \mathcal{I}_V^k of CI statements will be termed DAG-representable if there is a DAG which is k -faithful to it. We call a DAG D , which is k -faithful to \mathcal{I}_V^k , edge maximal if there is no k -faithful DAG whose edge set is a superset of D . Moreover, we denote by $\mathcal{F}(\mathcal{I}_V^k)$ the set of all k -faithful DAGs to \mathcal{I}_V^k .*

For example, for $\mathcal{I}_V^1 = \{(c \perp\!\!\!\perp d \mid a)\}$ with $V = \{a, b, c, d\}$, Fig. 1 shows all DAGs in $\mathcal{F}(\mathcal{I}_V^1)$.

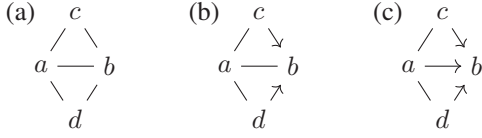


Figure 3: For the example from Fig. 1 we show the k -partial graph (part (a) on the left), the pattern of the edge maximal DAGs (part (b) in the middle) and the PDAG representing $\mathcal{F}(\mathcal{I}_V^1)$, with $\mathcal{I}_V^1 = \{(c \perp\!\!\!\perp d \mid a)\}$ (part (c) on the right).

Below, we define a representation of a set $\mathcal{F}(\mathcal{I}_V^k)$ as a PDAG. Using our definition, the set of k -faithful DAGs from Fig. 1 is represented by the PDAG shown in part (c) of Fig. 3.

We say that a PDAG $G = (V, E)$ contains a set of DAGs $\{D_i = (V, E_i) : i = 1, \dots, t\}$ if for every DAG $D_i = (V, E_i)$ it is true that $E_i \subseteq E$. Here, we assume that an undirected edge $a - b$ in G is encoded by two directed edges $a \rightarrow b$ and $b \rightarrow a$. Obviously, a complete undirected graph over V contains every set $\mathcal{F}(\mathcal{I}_V^k)$. From a causal structure learning perspective, our goal is to extract from \mathcal{I}_V^k as much causal knowledge as possible. We formalize this goal as to find the minimal PDAG which contains every DAG k -faithful to \mathcal{I}_V^k . In this setting, minimality is considered in regard to the inclusion relation between the sets of edges.

Definition 5. A PDAG G represents the set $\mathcal{F}(\mathcal{I}_V^k)$ if G is a minimal graph that contains every graph in $\mathcal{F}(\mathcal{I}_V^k)$.

It is easy to see, that, according to this definition, the PDAG in part (c) of Fig. 3 represents the set of k -faithful DAGs from Fig. 1.

We note that a PDAG G representing a set $\mathcal{F}(\mathcal{I}_V^k)$ fulfills the following conditions:

1. There is an edge $a - b$ in G iff DAGs $D, D' \in \mathcal{F}(\mathcal{I}_V^k)$ exist such that there is an edge $a \rightarrow b$ in D and an edge $a \leftarrow b$ in D' .
2. There is an edge $a \rightarrow b$ in G iff a DAG $D \in \mathcal{F}(\mathcal{I}_V^k)$ exists which contains the edge $a \rightarrow b$ and no DAG in $\mathcal{F}(\mathcal{I}_V^k)$ contains the edge $a \leftarrow b$.
3. There is no edge between a and b in G iff no DAG in $\mathcal{F}(\mathcal{I}_V^k)$ contains an edge between a and b .

From this perspective one can already view the representation G as a generalization of the notion of a CPDAG that is used to represent Markov equivalent DAGs of the same skeleton. Note that DAGs in $\mathcal{F}(\mathcal{I}_V^k)$ can have different skeletons. Interestingly, we prove that the PDAG representing the set of k -faithful graphs is still a CPDAG.

Proposition 1. For a given set \mathcal{I}_V^k of CIs, the representation G of all k -faithful DAGs $\mathcal{F}(\mathcal{I}_V^k)$ is a CPDAG. Moreover, any consistent extension of G is a DAG k -faithful to \mathcal{I}_V^k .

In particular, this means that the representation G is itself a faithful model of all CIs up to order k .

4 Determining the Skeleton

For a given set \mathcal{I}_V^k of conditional independence statements up to order k , our goal is to find the representation of the set

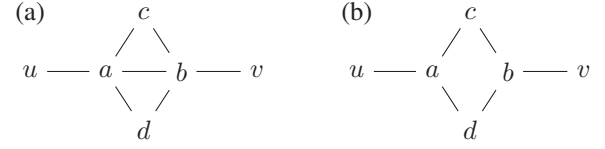


Figure 4: Left (a): The 0-1 graph for the CI statements \mathcal{I}_V^1 induced by the underlying DAG D shown in Fig. 2(a). It contains the edge $a - b$, as there is no independence of order zero or one between these nodes. Right (b): The skeleton of the graph computed by our algorithm (presented in the next section). Nodes a and b are not incident since they are incompatible (according to our definition). The justification is that, as seen in Fig. 2(a), there is an independence of order two ($a \perp\!\!\!\perp b \mid c, d$).

of k -faithful DAGs $\mathcal{F}(\mathcal{I}_V^k)$. By definition, this is the minimal graph which contains every k -faithful DAG. Thus, our strategy is the following. Starting with the complete graph, we want to remove all edges which do not belong to any k -faithful DAG and, vice versa, keep all edges which are in at least one k -faithful DAG. This is in line with the paradigm of *constraint-based* causal structure learning.

In this section, we characterize all pairs of nodes which are nonadjacent in *every* k -faithful DAG. These pairs of nodes are exactly the ones which are nonadjacent in the representation as well. This means that, by finding them, we can construct the skeleton of the representation. We will explore how edges are oriented in the subsequent section.

One setting in which two nodes have to be nonadjacent is quite obvious. If we have a statement $(a \perp\!\!\!\perp b \mid Z)_{\mathcal{I}_V^k}$, it follows trivially that there cannot be an edge between a and b in any k -faithful DAG. However, as we will see, though this condition is necessary, it is not sufficient for the non-adjacency of vertices in k -faithful DAGs. As the main result of this section, we provide a property between two nodes (we call it *incompatibility*) and using this property we formulate a criterion for non-adjacency which is both necessary and sufficient (Proposition 2). The incompatibility between two nodes a and b expresses some higher order conditional independencies which can be derived from CI statements up to order k .

Derivation of Higher-Order CI Statements

When having access to *all* conditional independencies without a restriction on the order, removing edges corresponding to these known CIs is sufficient for learning the skeleton of the underlying causal structure. For example, the SGS and the PC algorithm (Spirtes, Glymour, and Scheines 2000) work exactly in this fashion. However, only removing these edges is not sufficient even for obtaining the skeleton of the representation (or the skeleton of a k -faithful DAG) when we consider *order-bounded* sets of independencies. We will now investigate why this is the case and show how this obstacle can be overcome.

The outlined problem is illustrated in Fig. 4, for the CI statements of order 0 or 1 induced by the underlying DAG

D shown in Fig. 2(a), i.e. for the set $\mathcal{I}_V^1 = \{(u \perp\!\!\!\perp c), (u \perp\!\!\!\perp d), (u \perp\!\!\!\perp b), (u \perp\!\!\!\perp v), (a \perp\!\!\!\perp v), (c \perp\!\!\!\perp v), (d \perp\!\!\!\perp v), (c \perp\!\!\!\perp d), (u \perp\!\!\!\perp c \mid v), \dots\}$ of all zero- and first-order independencies found in this DAG. Choosing the value $k = 1$ allows us a comparison with 0-1 graphs, but such an example can be constructed for all $0 \leq k < n - 2$. In part (a) we show the corresponding 0-1 graph. This graph is constructed using the simple strategy of removing an edge if a zero- or first-order independence is present. We see that the nodes a and b are adjacent in this graph because no independence $(a \perp\!\!\!\perp b \mid Z)$ of order zero or one exists. However, in the underlying DAG D the nodes a and b are nonadjacent. Moreover, it is impossible to find a k -faithful DAG which contains the edge $a \rightarrow b$ or $a \leftarrow b$. In fact, D is the only k -faithful DAG. Essentially, this is the case because the edge between a and b (if present) would need to be in two conflicting v -structures, namely $u \rightarrow a \leftarrow b$ and $a \rightarrow b \leftarrow v$, to make sure that u and b as well as a and v are marginally independent. This is clearly impossible. From the fact that no k -faithful DAG contains an edge between a and b , we can infer that there has to be a higher-order CI between a and b . Here, this higher-order CI is $(a \perp\!\!\!\perp b \mid c, d)$.

It should be noted that de Campos and Huete (2000) already discovered that it is possible to remove further edges from 0-1 graphs (they considered a similar example in Fig. 2 of their paper). However, their method for deleting such edges relied on the topological ordering in the underlying DAG and they did not classify these edges. Requiring the topological ordering is a large obstacle for practical applications. Our proposed methods do not rely on the topological sorting as we give a simple classification of the edges that have to be removed from the 0-1 graph in order to obtain the skeleton of the representation and, by that, the skeleton of a k -faithful DAG.

We will now formalize the situation just described in the following definition and thereby introduce the so called *incompatible nodes*:

Definition 6. Let \mathcal{I}_V^k be a set of CIs of order $\leq k$. Then two nodes a and b are called *incompatible* iff there exist u, v, S, T such that the following two conditions hold:

1. $(u \perp\!\!\!\perp b \mid S)_{\mathcal{I}_V^k} \wedge (u \not\perp\!\!\!\perp a \mid S)_{\mathcal{I}_V^k} \wedge (a \not\perp\!\!\!\perp b \mid S)_{\mathcal{I}_V^k} \wedge a \notin S$,
2. $(v \perp\!\!\!\perp a \mid T)_{\mathcal{I}_V^k} \wedge (v \not\perp\!\!\!\perp b \mid T)_{\mathcal{I}_V^k} \wedge (b \not\perp\!\!\!\perp a \mid T)_{\mathcal{I}_V^k} \wedge b \notin T$.

We can see that the nodes a and b in the example in Fig. 4 are incompatible because $(u \perp\!\!\!\perp b)_{\mathcal{I}_V^1}$, $(u \not\perp\!\!\!\perp a)_{\mathcal{I}_V^1}$, $(a \not\perp\!\!\!\perp b)_{\mathcal{I}_V^1}$, $(v \perp\!\!\!\perp a)_{\mathcal{I}_V^1}$, and $(v \not\perp\!\!\!\perp b)_{\mathcal{I}_V^1}$ hold. In this case, S and T are both the empty set. It follows immediately that $a \notin S$ and $b \notin T$ are satisfied. Moreover, $(b \not\perp\!\!\!\perp a)_{\mathcal{I}_V^1}$ follows by symmetry from $(a \not\perp\!\!\!\perp b)_{\mathcal{I}_V^1}$.

We now prove formally that if the nodes a and b are incompatible, there cannot be an edge between a and b in any k -faithful DAG. Firstly, we show the following:

Lemma 1. Let \mathcal{I}_V^k be a set of CIs of order $\leq k$. If we have $(u \perp\!\!\!\perp b \mid Z)_{\mathcal{I}_V^k}$, $(u \not\perp\!\!\!\perp a \mid Z)_{\mathcal{I}_V^k}$ and $a \notin Z$, then no DAG k -faithful to \mathcal{I}_V^k contains the edge $a \rightarrow b$.

Proof. Assume, there is an edge $a \rightarrow b$ in a k -faithful DAG D . In this DAG, $(u \not\perp\!\!\!\perp a \mid Z)_{\mathcal{I}_V^k}^k$ has to hold. This means that

there is a path between u and a which is not blocked by Z . But as we have the edge $a \rightarrow b$ in G , there will also be a path between u and b which is not blocked by Z (note that $a \notin Z$). A contradiction. \square

We immediately conclude that incompatible nodes cannot be adjacent in any k -faithful DAG:

Corollary 1. Let \mathcal{I}_V^k be a set of CIs of order $\leq k$. If the nodes a and b are incompatible, they are nonadjacent in every DAG k -faithful to \mathcal{I}_V^k .

Due to the conditions stated in the definition of incompatible nodes (Definition 6), it follows from Lemma 1 that neither the edge $a \rightarrow b$ nor $a \leftarrow b$ can be in any k -faithful DAG.

A Complete Criterion for Adjacency

The following proposition underlines the importance of the notion of incompatible nodes by showing that making such nodes nonadjacent in a 0-1 graph is not only necessary, but also sufficient in order to obtain the skeleton of the representation:

Proposition 2. In the representation of $\mathcal{F}(\mathcal{I}_V^k)$ two nodes a and b are adjacent if and only if

- (i) there is no CI $(a \perp\!\!\!\perp b \mid Z)_{\mathcal{I}_V^k}$ for $Z \subseteq V$, $|Z| \leq k$, and
- (ii) the nodes a and b are not incompatible.

This result stems from the correctness proof of the LOCI algorithm presented in the following section (Algorithm 1). There, we complete the construction of the representation by showing how edges can be oriented.

5 Determining the Faithful Model

Now we are ready to discuss how to find the representation which k -faithfully models the CIs given in the set \mathcal{I}_V^k . This will also enable us to decide if \mathcal{I}_V^k even has a causal explanation. To answer this question, we attempt to construct the representation and if this fails, conclude that there can be no k -faithful DAG.

The LOCI (Low-Order Causal Inference) algorithm for constructing the representation is presented as Algorithm 1. We note that it works for arbitrary values k , in particular for $k = 0$ for which the CI statements represent marginal independencies. The algorithm can be divided into three stages.

In the first stage (line 1), the k -partial graph is generated which can be constructed by removing, from the complete undirected graph, edges corresponding to a CI in \mathcal{I}_V^k . We remark that, in general, one does not obtain this graph by executing e.g. the ‘‘skeleton phase’’ of the PC algorithm (Spirtes, Glymour, and Scheines 2000) up to order k . Here, only separating sets formed by the adjacent nodes are considered. Therefore, some separating sets of order $\leq k$ can be overlooked. Instead, it is necessary to consider all possible separating sets Z up to order k .

In the second stage (lines 2 to 6), directed edges are removed according to the rule in Lemma 1. Recall that an undirected edge $u - v$ is represented as a pair $u \rightarrow v$ and $u \leftarrow v$. Thus, removing only the edge $u \rightarrow v$ means the orientation of $u - v$ into $u \leftarrow v$. Obviously, removing both

input : Vertex set V , DAG-representable set \mathcal{I}_V^k of CIs with order $\leq k$

output: CPDAG G representing $\mathcal{F}(\mathcal{I}_V^k)$

- 1 Form the graph G on the vertex set V which has an undirected edge $a - b$ if for every subset Z of V , with $|Z| \leq k$, it is true $(a \not\perp\!\!\!\perp b \mid Z)_{\mathcal{I}_V^k}$.
- 2 **foreach** CI $(a \perp\!\!\!\perp b \mid Z)$ in \mathcal{I}_V^k and every $c \in V \setminus \{a, b\}$ **do**
- 3 **if** $(a \not\perp\!\!\!\perp c \mid Z)_{\mathcal{I}_V^k}$, $(c \not\perp\!\!\!\perp b \mid Z)_{\mathcal{I}_V^k}$ and $c \notin Z$ **then**
- 4 Remove $a \leftarrow c$ and $c \rightarrow b$ from G .
- 5 **end**
- 6 **end**
- 7 **repeat** the Meek rules until no rule can be applied.
- 8 1. $a \rightarrow b - c \Rightarrow a \rightarrow b \rightarrow c$
- 9 2. $a \rightarrow b \rightarrow c \Rightarrow a \rightarrow b \rightarrow c$
- 10 3. $a \begin{array}{c} \leftarrow c \\ \leftarrow d \end{array} \rightarrow b \Rightarrow a \begin{array}{c} \leftarrow c \\ \leftarrow d \end{array} \rightarrow b$
- 11 **end**

Algorithm 1: The LOCI algorithm computes the representation G for a DAG-representable set of CIs up to order k . Note that we represent an undirected edge $a - b$ as a pair $a \rightarrow b$ and $a \leftarrow b$.

directed edges denotes the deletion of the edge $u - v$. The aim of the second stage is (1) to remove the remaining undirected edges which do not satisfy the criterion in Proposition 2, i.e. the edges $u - v$ between incompatible nodes, and (2) to determine all v-structures. We note that in this stage, the algorithm also orients some further edges, which are not involved in v-structures.

To prove the correctness, we make use of the fact that we can always apply Lemma 1 to triples a, b, c (used in lines 2 to 6) and through this delete two directed edges $a \leftarrow c$ and $c \rightarrow b$ at the same time. This step ensures that all incompatible nodes are nonadjacent. In particular, nodes a and c are incompatible iff the edges $a \leftarrow c$ and $a \rightarrow c$ are removed at the different steps of the iteration corresponding to triples a, b, c and \hat{a}, c, a in lines 2 to 6: Indeed, the two conditions (cf. Definition 6)

1. $(a \perp\!\!\!\perp b \mid Z)_{\mathcal{I}_V^k} \wedge (a \not\perp\!\!\!\perp c \mid Z)_{\mathcal{I}_V^k} \wedge (c \not\perp\!\!\!\perp b \mid Z)_{\mathcal{I}_V^k} \wedge c \notin Z$,
 2. $(\hat{a} \perp\!\!\!\perp c \mid \hat{Z})_{\mathcal{I}_V^k} \wedge (\hat{a} \not\perp\!\!\!\perp a \mid \hat{Z})_{\mathcal{I}_V^k} \wedge (a \not\perp\!\!\!\perp c \mid \hat{Z})_{\mathcal{I}_V^k} \wedge a \notin \hat{Z}$
- are true iff the algorithm removes $a \leftarrow c$ and $a \rightarrow c$ in line 4. If, however, only the edge $a \leftarrow c$ is removed from the undirected edge $a - c$, the edge $a \rightarrow c$ remains, meaning the orientation of $a - c$ into $a \rightarrow c$.

Moreover, we show that in stage two all v-structures of the representation are oriented. Note that, in order to make sure all v-structures $x \rightarrow y \leftarrow z$ are correctly oriented, even if x and z are incompatible, it is necessary to consider all triples of nodes a, b, c and not only chains $a - c - b$ as in common causal structure learning algorithms like the PC-algorithm (Spirtes, Glymour, and Scheines 2000).

Finally, in the third stage (line 7 to 11), the algorithm orients further undirected edges through the Meek rules. The

graph obtained after completing stage two already characterizes a Markov equivalence class, as the skeleton and the v-structures are determined. In order to obtain the representation, we have to maximally extend it into a CPDAG. This is why we are able to apply the Meek rules.

Before stating the main results, we illustrate how the LOCI algorithm works using as an example instance the zero- and first-order independencies $\mathcal{I}_V^1 = \{(c \perp\!\!\!\perp d \mid a)\}$ over $V = \{a, b, c, d\}$, that have been discussed in Fig. 1 and 3. In (a), (b) and (c) of Fig. 3 the graph G is shown after completing stage one, two, and three, respectively. Thus, in (a) there is no edge between c and d as we have the independence $(c \perp\!\!\!\perp d \mid a)$ in \mathcal{I}_V^1 , while all other edges are present. In (b) we see that the edges $c \rightarrow b$ and $d \rightarrow b$ are oriented. Essentially, there can be no edge $c \leftarrow b$ (or $d \leftarrow b$) as in that case $(c \perp\!\!\!\perp d \mid a)_{\mathcal{I}_V^1}$ cannot hold without a collider at node b . In this regard, stage two is similar to the orientation of v-structures in the SGS or PC algorithm (Spirtes, Glymour, and Scheines 2000). The difference is, however, as emphasized before, that in the LOCI algorithm further nodes can be separated during this stage. An example for this are the incompatible nodes a and b of the example in Fig. 2 and 4. We, moreover, remark that, while all v-structures are detected, the result is not always a pattern, as it is possible that even further edges are already oriented. Finally, in part (c) the resulting graph G is shown. Here, the edge $a - b$ has been oriented into $a \rightarrow b$ due to the third Meek rule. As seen in Fig. 1, there are six DAGs which are k -faithful to \mathcal{I}_V^1 . Three of them contain the edge $a \rightarrow b$ and in the other three a and b are nonadjacent. However, the edge $a \leftarrow b$ is in no k -faithful DAG which is why the orientation $a \rightarrow b$ is correct.

We now state the main result of this paper that the LOCI algorithm produces the required representation:

Theorem 1. *The graph G resulting from the LOCI algorithm (Algorithm 1) is the representation of the set $\mathcal{F}(\mathcal{I}_V^k)$, if \mathcal{I}_V^k is DAG-representable.*

Some ingredients of the proof of this theorem have already been stated in this and the previous section. The complete proof can be found in the full version of this paper.

The result enables us to *decide* whether a given set \mathcal{I}_V^k has a causal explanation. This is possible through the following approach: We can apply the LOCI algorithm to \mathcal{I}_V^k and check whether the resulting graph is a k -faithful CPDAG. If it is, clearly there is a causal explanation of \mathcal{I}_V^k , namely the produced graph (Proposition 1). If it is not, then \mathcal{I}_V^k cannot have such a causal explanation as, if this were the case, G would be the representation (Theorem 1) and, therefore, as argued above, a faithful model. Thus, we conclude:

Proposition 3. *There exists an algorithm which for a given set \mathcal{I}_V^k of CIs tests if the set is DAG-representable.*

6 Experimental Analysis

The representation G of a set \mathcal{I}_V^k is in itself a very useful graph as it faithfully models the CIs of order $\leq k$. But apart from this, it can also be used as an approximation of the true underlying causal structure. It can even be argued that it is the best approximation obtained through the given conditional independence information. Because of the minimality

DAG		Number of edges		
n	d	0-1	skel. G	skel. D
20	2	27.21	25.43	19.81
20	3	57.88	51.07	29.79
20	4	96.57	87.47	40.03
20	5	126.73	119.49	50.21
60	2	77.85	69.69	58.97
60	3	226.03	160.43	88.55
60	4	512.89	346.06	118.29
60	5	820.69	579.67	148.08
100	2	125.87	113.37	99.40
100	3	413.68	266.92	149.42
100	4	1,061.39	598.11	199.34
100	5	1,905.20	1,118.60	248.92

Table 1: We consider random DAGs with n nodes and expected node degree d . This means each edge is present with probability $d/(n-1)$. We present the number of edges in the 0-1 graph, the skeleton of representation G and the skeleton of the true DAG D . All values are the means of 100 independent trials.

of the representation, removing a further edge from G would mean that some DAG k -faithful to \mathcal{I}_V^k is not contained in it anymore.

Thus, we investigate in this section how well the representation G of a set of low-order CIs is able to capture the underlying true causal structure. We do this experimentally by generating a sparse DAG which we then try to recover with the LOCI algorithm. We confine our analysis to the case $k = 1$ which allows us a comparison with the 0-1 graph model. For this, we compare the number of adjacencies (meaning the number of edges in the skeleton) in the 0-1 graph, the CPDAG G and the true DAG. This enables us, in particular, to estimate the influence of removing edges between incompatible nodes. Additionally, we investigate how many v-structures from the true DAG can already be found in the CPDAG G , giving us an indication how well the edge orientations are captured in the representation.

We begin by explaining how we generated the set of independencies \mathcal{I}_V^1 . An undirected graph with n nodes is drawn randomly. More precisely, each edge is present with probability $d/(n-1)$, meaning every node has expected degree d . Afterwards, a topological ordering of the nodes is randomly chosen in order to obtain a DAG D from the generated graph. From this DAG we can read off all zero- and first-order independencies through the notion of d-separation and thereby produce the set \mathcal{I}_V^1 needed for the LOCI algorithm.

The representation G can be obtained by performing the LOCI algorithm on \mathcal{I}_V^1 and the 0-1 graph can be easily obtained as well by removing edges which correspond to independencies in \mathcal{I}_V^1 . Through the generation procedure we also have access to the underlying true DAG D . First, we look at the number of adjacencies in the different graphs (see Table 1). The displayed numbers are the means of 100 independent trials and we consider graphs with 20, 60 and 100

DAG		Number of v-s per node		
n	d	v-s in G	v-s in D	v-s in both
20	2	1.427	0.561	0.552
20	3	5.276	1.190	1.111
20	4	10.784	2.024	1.649
20	5	13.556	2.938	2.031
60	2	1.445	0.614	0.612
60	3	10.904	1.343	1.317
60	4	42.979	2.369	2.200
60	5	90.279	3.650	3.119
100	2	1.383	0.654	0.653
100	3	12.982	1.441	1.424
100	4	59.859	2.521	2.419
100	5	161.390	3.872	3.504

Table 2: In the same setting as in Table 1 we present the number of v-structures (v-s, for short) in G , in D and those in both graphs.

nodes and expected node degree 2, 3, 4 and 5. Clearly, the numbers are nonincreasing from left to right. To be more precise, it holds that $A_D \subseteq A_G \subseteq A_{0-1}$ where A is the set of all adjacencies. This is due to the fact that every k -faithful DAG is contained in G and that G is constructed by removing edges from the 0-1 graph.

We begin the analysis by exploring how close G is to the true causal structure. It can be seen that, in particular for larger graphs, we are only able to reasonably estimate the underlying structure up to expected degree 3. For example for $n = 100$ and $d = 4$, the representation G contains almost three times as many adjacencies as D . For $n = 100$ and $d = 2$, the estimation is very close to the true DAG and even for $d = 3$ the ratio between the number of adjacencies in G and D is quite reasonable, being well below two. Notably, in the latter setting the improvement over the 0-1 graphs is significant. Actually, the difference in the number of adjacencies is larger between the 0-1 graph and G than between G and D . More generally, we see that for larger graphs the gap between the 0-1 graph and G is substantial, meaning there is a great number of incompatible nodes. This underlines the importance of removing edges between such nodes in order to find a graphical model which is k -faithful to a set of independencies \mathcal{I}_V^1 . We can conclude that it is possible to estimate the true causal structure reasonably well, given that it is sparse. Moreover, it is crucial to remove the edges between incompatible nodes. But, apart from the adjacencies (or in other words the skeleton), the representation also contains directed edges and, thus, also v-structures. Therefore, it is interesting to investigate how many v-structures from the true DAG can already be found in G . These numbers are presented in Table 2. Here, we show the number of v-structures in D , in G and those which are in both D and G . For better readability, the numbers are normalized by the number of nodes n . We consider the same setting as above.

We investigate first how many v-structures are in both G and D compared to the number of v-structures in D . This

shows how many of the v -structures of the true underlying DAG the LOCI algorithm is able to detect. We can see that almost all v -structures are found even for larger expected node degrees 4 or 5. E.g. for $n = 100$ and $d = 5$, the LOCI algorithm discovers 3.504 out of 3.872 v -structures per node.

While the LOCI algorithm finds most of the v -structures in D , we can see that there are many more additional v -structures in the representation G . While this is in reasonable limits for sparse graphs (for $d = 2$ we see roughly a doubling of the number of v -structures), the difference is much more extreme in denser graphs. In particular, for $n = 100$ and $d = 5$ there are 161.39 v -structures per node in G and only 3.872 in D . This is due to the fact that, as we have seen in Table 1, there are many more edges in G . At first glance, however, the increase in v -structures is much more extreme (a factor more than forty) than the increase in edges (a factor slightly less than five). But recall that these additional edges have an important property. As we know that G is a k -faithful CPDAG, both G and D contain the same zero- and first-order CIs. Therefore, all additional edges in G lead to *no* further dependencies of order zero or one. It is reasonable to assume that these additional edges are, thus, part of a disproportionate number of v -structures as they do not create new paths and thereby new dependencies.

7 Discussion

This paper has investigated the problems of determining how, for a given set of CI statements of order up to k , all DAGs k -faithful to the set can be represented and how such a representation can be computed. We solve both problems showing that such faithful DAGs can be represented in a compact way as a CPDAG G and then proving that the representation G can be computed efficiently.

The experimental results show that, for small values of k , this graphical representation is also useful as a good estimator of the underlying true causal structure in case of sparse models. It is considerably better than the k -partial graph because further edges are removed due to the concept of incompatible nodes which allows us to infer the existence of higher-order independencies. An additional advantage over k -partial graphs is that we also obtain edge orientations and can, through this, recover a large portion of the v -structures in the true DAG.

Our experiments are conducted in the oracle model where we assume all CI statements up to order k are known. This has the reason that, in this model, we are able to estimate best how many incompatible edges are removed. In future work, it would be interesting to analyze how the proposed algorithm performs if one would use statistical tests to find the independence statements. Another interesting topic for future research is to extend our algorithmic technique to compute the k -faithful representation, or a good approximation of it, by asking conditional independence queries in such a way that the number of queries is significantly smaller than the number of all CI statements of order up to k . This would be interesting both in the oracle model and when using statistical tests to estimate independencies.

Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) grant LI 634/4-2.

References

- Andersson, S. A.; Madigan, D.; and Perlman, M. D. 1997. A characterization of Markov equivalence classes for acyclic digraphs. *The Annals of Statistics* 25(2):505–541.
- Castelo, R., and Roverato, A. 2006. A robust procedure for Gaussian graphical model search from microarray data with p larger than n . *Journal of Machine Learning Research* 7(Dec):2621–2650.
- Cox, D. R., and Wermuth, N. 1993. Linear dependencies represented by chain graphs. *Statistical Science* 8(3):204–283.
- De Campos, L. M., and Huete, J. F. 2000. A new approach for learning belief networks using independence criteria. *International Journal of Approximate Reasoning* 24(1):11–37.
- De la Fuente, A.; Bing, N.; Hoeschele, I.; and Mendes, P. 2005. Discovery of meaningful associations in genomic data using partial correlation coefficients. *Bioinformatics* 20:3565–74.
- Kalisch, M., and Bühlmann, P. 2007. Estimating high-dimensional directed acyclic graphs with the PC-Algorithm. *Journal of Machine Learning Research* 8:613–636.
- Lauritzen, S. L. 1996. *Graphical models*. Clarendon Press.
- Magwene, P. M., and Kim, J. 2004. Estimating genomic co-expression networks using first-order conditional independence. *Genome biology* 5:R100.
- Meek, C. 1995. Causal inference and causal explanation with background knowledge. In *Proc. of UAI 1995*, 403–410. MK Publishers Inc.
- Pearl, J., and Wermuth, N. 1994. When can association graphs admit a causal interpretation? In *Selecting Models from Data*, 205–214. Springer.
- Pearl, J. 2009. *Causality: Models, Reasoning and Inference*. Cambridge University Press, 2nd edition.
- Spirites, P.; Glymour, C.; and Scheines, R. 2000. *Causation, Prediction, and Search*. MIT press, 2nd edition.
- Textor, J.; Idelberger, A.; and Liškiewicz, M. 2015. Learning from pairwise marginal independencies. In *Proc. of UAI 2015*, 882–891. AUAI Press.
- Verma, T., and Pearl, J. 1990. Equivalence and synthesis of causal models. In *Proc. of UAI 1990*, 255–270. Elsevier.
- Verma, T., and Pearl, J. 1992. An algorithm for deciding if a set of observed independencies has a causal explanation. In *Proc. of UAI 1992*, 323–330. MK Publishers Inc.
- Wille, A., and Bühlmann, P. 2006. Low-order conditional independence graphs for inferring genetic networks. *Statistical applications in genetics and molecular biology* 5(1).
- Wille, A.; Zimmermann, P.; Vranová, E.; Fürholz, A.; Laule, O.; Bleuler, S.; Hennig, L.; Prelic, A.; von Rohr, P.; Thiele, L.; Zitzler, E.; Gruissem, W.; and Bühlmann, P. 2004. Sparse graphical Gaussian modeling of the isoprenoid gene network in *arabidopsis thaliana*. *Genome biology* 5:R92.