Almost Optimal Universal Lower Bound for Learning Causal DAGs with Atomic Interventions

Vibhor Porwal
Adobe Research
viporwal@adobe.com

Piyush Srivastava
Tata Institute of Fundamental Research
piyush.srivastava@tifr.res.in

Gaurav Sinha
Adobe Research
gasinha@adobe.com

Abstract

A well-studied challenge that arises in the structure learning problem of causal directed acyclic graphs (DAG) is that using observational data, one can only learn the graph up to a "Markov equivalence class" (MEC). The remaining undirected edges have to be oriented using interventions, which can be very expensive to perform in applications. Thus, the problem of minimizing the number of interventions needed to fully orient the MEC has received a lot of recent attention, and is also the focus of this work. We prove two main results. The first is a new universal lower bound on the number of atomic interventions that any algorithm (whether active or passive) would need to perform in order to orient a given MEC. Our second result shows that this bound is, in fact, within a factor of two of the size of the smallest set of atomic interventions that can orient the MEC. Our lower bound is provably better than previously known lower bounds. The proof of our lower bound is based on the new notion of clique-block shared-parents (CBSP) orderings, which are topological orderings of DAGs without v-structures and satisfy certain special properties. Further, using simulations on synthetic graphs and by giving examples of special graph families, we show that our bound is often significantly better.

1 INTRODUCTION

Causal Bayesian Networks (CBN) provide a very convenient framework for modeling causal relationships between a collection of random variables (Pearl, 2009). A CBN is fully specified by (a) a directed acyclic graph (DAG), whose nodes model random variables of interest, and whose edges depict immediate causal relationships between the nodes, and (b) a conditional probability distribution (CPD) of each variable given its parent variables (in the DAG) such that the joint distribution of all variables factorizes as a product of these conditionals. The generality of the framework has led to CBN becoming a popular tool for the modeling of causal relationships in a variety of fields, with health science (Shen et al., 2020), molecular cell biology (Friedman, 2004), and computational advertising (Bottou et al., 2013) being a few examples.

It is well known that the underlying DAG of a CBN is not uniquely determined by the joint distribution of its nodes. In fact, the joint distribution only determines the DAG up to its Markov Equivalence Class (MEC), which is represented as a partially directed graph with well-defined combinatorial properties (Verma and Pearl, 1990; Chickering, 1995; Meek, 1995; Andersson et al., 1997). Information about which nodes are *adjacent* is encoded in the MEC, but the *direction* of several edges remains undetermined. Thus, learning algorithms based only on the observed joint distribution (Glymour et al., 2019) cannot direct these remaining edges. As a result, algorithms which use additional *interventional* distributions were developed (Squires et al. (2020) and references therein). In addition to the joint distribution, these algorithms also assume access to interventional distributions generated as a result of randomizing some *target* vertices in the original CBN (a process called *intervention*) and thereby breaking their dependence on any of their ancestors. A natural and well-motivated (Eberhardt et al., 2005) question, therefore, is to find the minimum number of interventions required to fully resolve the orientations of the undirected edges in the MEC.

Interventions, especially on a large set of nodes, however, can be expensive to perform (Kocaoglu et al., 2017). In this respect, the setting of *atomic interven-*

Proceedings of the 25$^{\text{th}}$ International Conference on Artificial Intelligence and Statistics (AISTATS) 2022, Valencia, Spain. PMLR: Volume 151.

tions, where each intervention is on a single node, is already very interesting and finding the smallest number of atomic interventions that can orient the MEC is well studied (Squires et al., 2020). A long line of work, including those cited above, has considered in various settings the problem of designing methods for finding the smallest set of atomic interventions that would fully orient all edges of a given MEC. An important distinction between such methods is whether they are *active* (He and Geng, 2008), i.e., where the directions obtained via the current intervention are available before one decides which further interventions to perform; or *passive*, where all the interventions to be performed have to be specified beforehand. Methods can also differ in whether or not randomness is used in selecting the targets of the interventions. An important question, therefore, is to understand how many interventions must be performed by any given method to fully orient an MEC.

Universal Lower Bounds While several works have reported lower bounds (on minimum number of atomic interventions required to orient an MEC) in different settings, a very satisfying solution concept for such lower bounds, called *universal lower bounds*, was proposed by Squires et al. (2020). A universal lower bound of L atomic interventions for orienting a given MEC means that if a set of atomic interventions is of size less than L, then for *every* ground-truth DAG D in the MEC, the set S will fail to fully orient the MEC. Thus, a universal lower bound has two universality properties. First, the value of a universal lower bound depends only upon the MEC, and applies to *every* DAG in the MEC. Second, the lower bound applies to *every* set of interventions that would fully orient the MEC, without regards to the method by which the intervention set was produced.

In this work, we address the problem of obtaining tight universal lower bounds. The goal is to find a universal lower bound such that for any DAG D in the MEC, the smallest set of atomic interventions that can orient the MEC into D has size bounded above by a constant factor of the universal lower bound. Similar to Squires et al. (2020), we work in the setting of *causally sufficient* models, i.e. there are no hidden confounders, selection bias or feedback. To the best of our knowledge, this is the first work that addresses the problem of tight (up to a constant factor) universal lower bounds. We note that the best known universal lower bounds (Squires et al., 2020) so far are not tight and provide concrete examples of graph families that illustrate this in Section 3.2.

1.1 Our Contributions

We prove a new universal lower bound on the size of any set of atomic interventions that can orient a given MEC, improving upon previous work (Squires et al., 2020). We further prove that our lower bound is optimal within a factor of 2 in the class of universal lower bounds: we show that for any DAG D in the MEC, there is a set of atomic interventions of size at most twice our lower bound, that would fully orient the MEC if the unknown ground-truth DAG were D.

We also compare our new lower bound with the one obtained previously by Squires et al. (2020). We prove analytically that our lower bound is at least as good as the one given by Squires et al. (2020). We further give examples of graph classes where our bound is significantly better (in fact, it is apparent from our proof that the graphs in which the two lower bounds are close must have very special properties). We then supplement these theoretical findings with simulation results comparing our lower bound with the "true" optimal answer and with the lower bound in previous work.

Our lower bound is based on elementary combinatorial arguments drawing upon the theory of chordal graphs, and centers around a notion of certain special topological orderings of DAGs without v-structures, which we call *clique-block shared-parents (CBSP) orderings* (Definition 3.2). This is in contrast to the earlier work of Squires et al. (2020), where they had to develop sophisticated notions of directed clique trees and residuals in order to prove their lower bound. We expect that the notion of CBSP orderings may also be of interest in the design of optimal intervention sets.

1.2 Related Work

The theoretical underpinning for many works dealing with the use of interventions for orienting an MEC can be said to be the notion of "interventional" Markov equivalence (Hauser and Bühlmann, 2012), which, roughly speaking, says that given a collection $\mathcal{I}$ of sets of targets for interventions, two DAGs D_1 and D_2 are *$\mathcal{I}$-Markov equivalent* if and only if for all $S \in \mathcal{I}$, the DAGs obtained by removing from D_1 and D_2 the incoming edges of all vertices in S are in the same MEC (Hauser and Bühlmann, 2012, Theorem 10). Thus, interventions have the capability of distinguishing between DAGs in the same Markov Equivalence class, and in particular, "interventional" Markov equivalence classes can be finer than MECs (Hauser and Bühlmann, 2012, see also Theorem 2.1 below).

As described above, the problem of learning the orientations of a CBN using interventions has been studied in a wide variety of settings. Lower bounds and

algorithms for the problem have been obtained in the setting of interventions of arbitrary sizes and with various cost models (Eberhardt, 2008; Shanmugam et al., 2015; Kocaoglu et al., 2017), in the setting when the underlying model is allowed to contain feedback loops (and is therefore not a CBN in the usual sense) (Hyttinen et al., 2013b,a), in settings where hidden variables are present (Addanki et al., 2020, 2021), and in interventional "sample efficiency" settings (Agrawal et al., 2019; Greenewald et al., 2019). The related notion of orienting the maximum possible number of edges given a fixed budget on the number or cost of interventions has also been studied (Hauser and Bühlmann, 2014; Ghassami et al., 2018; AhmadiTeshnizi et al., 2020). However, to the best of our knowledge, the work of Squires et al. (2020) was the first to isolate the notion of a universal lower bound, and prove a lower bound in that setting. We discuss lower bounds in some other settings in Section 3.3.

2 PRELIMINARIES

Graphs A *partially directed graph* (or just *graph*) $G = (V, E)$ consists of a set V of *nodes* or *vertices* and a set E of *adjacencies*. Each adjacency in E is of the form $a - b$ or $a \to b$, where $a, b \in V$ are *distinct* vertices, with the condition that for any $a, b \in V$, at most one of $a - b$, $a \to b$ and $b \leftarrow a$ is present in E. If there is an adjacency in E containing both a and b, then we say that a and b are *adjacent* in G, or that there is an *edge* between a and b in G. If $a - b \in E$, then we say that the edge between a and b in G is undirected, while if $a \to b \in E$ then we say that the edge between a and b is *directed* in G. G is said to be *undirected* if all its adjacencies are undirected, and *directed* if all its adjacencies are directed. Given a directed graph G, and a vertex v in G, we denote by $\mathrm{pa}_G(v)$ the set of nodes u in G such that $u \to v$ is present in G. A vertex v in G is said to be a *child* of u if $u \in \mathrm{pa}_G(v)$. An *induced subgraph* of G is a graph whose vertices are some subset S of V, and whose adjacencies $E[S]$ are all those adjacencies in E both of whose elements are in S. This induced subgraph of G is denoted as $G[S]$. The *skeleton* of G, denoted $\mathrm{skeleton}(G)$, is an undirected graph with nodes V and adjacencies $a - b$ whenever a, b are adjacent in G.

A *cycle* in a graph G is a sequence of vertices $v_1, v_2, v_3, \ldots, v_{n+1} = v_1$ (with $n \geq 3$) such that for each $1 \leq i \leq n$, either $v_i \to v_{i+1}$ or $v_i - v_{i+1}$ is present in E. The *length* of the cycle is n, and the cycle is said to be *simple* if $v_1, v_2, \ldots, v_n$ are distinct. The cycle is said to have a *chord* if two non-consecutive vertices in the cycle are adjacent in G, i.e., if there exist $1 \leq i < j \leq n$ such that $j - i \neq \pm 1 \pmod{n}$ and such that v_i and v_j are adjacent in G. The cycle is said to be *directed* if for some $1 \leq i \leq n$, $v_i \to v_{i+1}$ is present in G. A graph G is said to be a *chain graph* if it has no directed cycles. The *chain components* of a chain graph G are the connected components left after removing all the directed edges from G. A *directed acyclic graph* or DAG is a directed graph without directed cycles. Note that both DAGs and undirected graphs are chain graphs. An undirected graph G is said to be *chordal* if any simple cycle in G of length at least 4 has a chord.

A *clique* C in a graph $G = (V, E)$ is a subset of nodes of G such that any two distinct u and v in C are adjacent in G. The clique C is *maximal* if for all $v \in V \setminus C$, the set $C \cup \{v\}$ is not a clique.

A *perfect elimination ordering* (PEO), $\sigma = (v_1, \ldots, v_n)$ of a graph G is an ordering of the nodes of G such that $\forall i \in [n]$, $ne_G(v_i) \cap \{v_1, \ldots, v_{i-1}\}$ is a clique in G, where $ne_G(v_i)$ is the set of nodes adjacent to v_i.[1] A graph is chordal if and only if it has a perfect elimination ordering (Blair and Peyton, 1993). A *topological ordering*, σ of a DAG D is an ordering of the nodes of D such that $\sigma(a) < \sigma(b)$ whenever $a \in \mathrm{pa}_D(b)$, where $\sigma(u)$ denotes the index of u in σ. We say that D is oriented according to an ordering σ to mean that D has a topological ordering σ.

A *v-structure* in a graph G is an induced subgraph of the form $b \to a \leftarrow c$ (v-structures are also known as *unshielded colliders*). It follows easily from the definitions that by orienting the edges of a chordal graph according to a perfect elimination ordering, we get a DAG without v-structures, and that the skeleton of a DAG without v-structures is chordal (see Proposition 1 of Hauser and Bühlmann (2014)). In fact, any topological ordering of a DAG D without v-structures is a perfect elimination ordering of $\mathrm{skeleton}(D)$.

Interventions An *intervention* I on a partially directed graph $G = (V, E)$ is specified as a subset of *target* vertices of G. Operationally, an intervention at I is interpreted as completely randomizing the distributions of the random variables corresponding to the vertices in I. We work here in the "infinite sample" setting, where, under standard assumptions, performing the intervention I reveals at least the directions of all edges between vertices in I and $V \setminus I$ (see Theorem 2.1 below for a more formal statement of the extent to which a set of interventions orients the edges of a partially directed graph). An *intervention set* is a set of interventions. In this paper, we make the standard assumption that the "empty" intervention, in which no vertices are intervened upon, is *always* included in any intervention set we consider:

[1] Our definition of a PEO uses the same ordering convention as Hauser and Bühlmann (2014).

this corresponds to assuming that information from purely observational data is always available (see, e.g., the discussion surrounding Definition 6 of Hauser and Bühlmann (2012)). The *size* of an intervention set $\mathcal{I}$ is the number of interventions in $\mathcal{I}$, not counting the empty intervention.

$\mathcal{I}$ is a set of *atomic interventions* if $|I| = 1$ for all non-empty $I \in \mathcal{I}$. With a slight abuse of notation, we denote a set of atomic interventions $\mathcal{I} = \{\varnothing, \{v_1\}, \ldots, \{v_k\}\}$ as just the set $I = \{v_1, \ldots, v_k\}$ when it is clear from the context that we are talking about a set of atomic interventions. Given an intervention set $\mathcal{I}$ and a DAG D, we denote, following Hauser and Bühlmann (2012), by $\mathcal{E}_{\mathcal{I}}(D)$ the partially directed graph representing the set of all DAGs that are $\mathcal{I}$-Markov equivalent to D. $\mathcal{E}_{\mathcal{I}}(D)$ is also known as the $\mathcal{I}$-essential graph of D (see Fig. 2 of Hauser and Bühlmann (2014) for an example). For a formal definition of $\mathcal{I}$-Markov equivalence, we refer to Definitions 7 and 9 of Hauser and Bühlmann (2012); we use instead the following equivalent characterization developed in the same paper.

Theorem 2.1 (**Characterization of $\mathcal{I}$-essential graphs, Definition 14 and Theorem 18 of Hauser and Bühlmann (2012)**). *Let D be a DAG and $\mathcal{I}$ an intervention set. A graph H is an $\mathcal{I}$-essential graph of D if and only if H has the same skeleton as D, all directed edges of H are directed in the same direction as in D, all v-structures of D are directed in H, and*

1. *H is a chain graph with chordal chain components.*
2. *For any three vertices a, b, c of H, the subgraph of H induced by a, b and c is not $a \to b - c$.*
3. *If $a \to b$ in D (so that a, b are adjacent in H) and there is an intervention $J \in \mathcal{I}$ such that $|J \cap \{a, b\}| = 1$, then $a \to b$ is directed in H.*
4. *Every directed edge $a \to b$ in H is strongly $\mathcal{I}$-protected. An edge $a \to b$ in H is said to be strongly $\mathcal{I}$-protected if either (a) there is an intervention $J \in \mathcal{I}$ such that $|J \cap \{a, b\}| = 1$, or (b) at least one of the four graphs in Figure 1 appears as an induced subgraph of H, and $a \to b$ appears in that induced subgraph in the configuration indicated in the figure.*

3 UNIVERSAL LOWER BOUND

In this section, we establish our main technical result (Theorem 3.3). Our new lower bound (Theorems 3.6 and 3.7) then follows easily from this combinatorial result, without having to resort to the sophisticated machinery of residuals and directed clique trees developed in previous work (Squires et al., 2020).

We begin with a definition that isolates two important properties of certain topological orderings of DAGs without v-structures. Given a DAG D without v-structures, and a maximal clique C of skeleton(D), we denote by $\text{sink}_D(C)$ any vertex in D such that $C = \text{pa}_D(\text{sink}_D(C)) \cup \{\text{sink}_D(C)\}$. The fact that $\text{sink}_D(C)$ is uniquely defined, and that $\text{sink}_D(C_1) \neq \text{sink}_D(C_2)$ when C_1 and C_2 are distinct maximal cliques of skeleton(D) is guaranteed by the following observation. (The standard proof of this is deferred to supplementary material Section B.1.) This observation also makes it clear that $\text{sink}_D(C)$ is the unique vertex with out-degree 0 in the induced subgraph $D[C]$.

Observation 3.1. *Let D be a DAG without v-structures. Then, for every maximal clique C of* skeleton(D)*, there is a unique vertex v of D, denoted* $\text{sink}_D(C)$*, such that $C = \text{pa}_D(v) \cup \{v\}$. Further, for any two distinct maximal cliques C_1 and C_2 in* skeleton(D)*, we have* $\text{sink}_D(C_1) \neq \text{sink}_D(C_2)$.

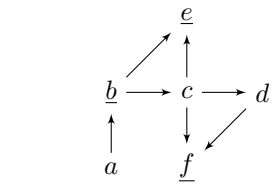
We refer to each vertex v of D that is equal to $\text{sink}_D(C)$ for some maximal clique of skeleton(D) as a *maximal-clique-sink* vertex of the DAG D. We now present a definition of clique-block shared-parents orderings (see Figure 2 for an example).

Definition 3.2 (**Clique-Block Shared-Parents (CBSP) ordering**). *Let σ be a topological ordering of a DAG D without v-structures. Let $s_1, s_2, s_3, \ldots, s_r$ be the maximal-clique-sink vertices of D indexed so that $\sigma(s_i) < \sigma(s_j)$ when $i < j$. (Here r is the number of maximal cliques in* skeleton(D)*.) Then, σ is said to be a* clique-block shared-parents (CBSP) ordering *of D if it satisfies the following two properties:*

1. ***P1: Clique block property*** *Define $L_1(\sigma)$ to be the set of nodes u which occur before or at the same position as s_1 in σ i.e., $\sigma(u) \leq \sigma(s_1)$. Similarly, for $2 \leq i \leq r$, define $L_i(\sigma)$ to be the set of nodes which occur in σ before or at the same position as s_i, but strictly after s_{i-1} (i.e., $\sigma(s_{i-1}) < \sigma(u) \leq \sigma(s_i)$). Then, for each $1 \leq i \leq r$ the subgraph induced by $L_i(\sigma)$ in* skeleton(D) *is a (not necessarily maximal) clique.*
2. ***P2: Shared parents property*** *If vertices a and b in D are consecutive in σ (i.e., $\sigma(b) = \sigma(a)+1$), and also lie in the same $L_i(\sigma)$ for some $1 \leq i \leq r$, then all parents of a are also parents of b in D.*

We illustrate the definition with an example in Figure 2. In the figure, vertices $\underline{b}$, $\underline{e}$, and $\underline{f}$ are the

Figure 1: Strong Protection (Andersson et al., 1997; Hauser and Bühlmann, 2012)



A DAG D without v-structures

$$\sigma' := \underbrace{a\,\underline{b}}_{L_1(\sigma')} \quad \underbrace{c\,d\,\underline{e}}_{L_2(\sigma')} \quad \underbrace{\underline{f}}_{L_3(\sigma')}$$

$$\sigma := \underbrace{a\,\underline{b}}_{L_1(\sigma)} \quad \underbrace{c\,d\,\underline{f}}_{L_2(\sigma)} \quad \underbrace{\underline{e}}_{L_3(\sigma)}$$

$$\tau := \underbrace{a\,\underline{b}}_{L_1(\tau)} \quad \underbrace{c\,\underline{e}}_{L_2(\tau)} \quad \underbrace{d\,\underline{f}}_{L_3(\tau)}$$

Figure 2: Clique-Block Shared-Parents Topological Orderings: τ Satisfies both P1 and P2; σ Satisfies only P1

maximal-clique-sink vertices of D, and are highlighted with an underbar. The orderings σ', σ and τ in the figure are valid topological orderings of D. However, σ' does not satisfy P1 of Definition 3.2 (since $L_2(\sigma')$ is not a clique), while σ satisfies P1 of Definition 3.2, but does not satisfy P2, because c, d in $L_2(\sigma)$ are consecutive in σ, but b is a parent only of c and not of d. Finally, τ satisfies both P1 and P2 and hence is a CBSP ordering.

Our main technical result is that for any DAG D that has no v-structures, there exists a CBSP ordering σ of D, and the new lower bound is an easy corollary of this result. Further, the proof of this result uses only standard notions from the theory of chordal graphs.

Theorem 3.3. *If D is a DAG without v-structures, then D has a CBSP ordering.*

Towards the proof of this theorem, we note first that the existence of a topological ordering σ satisfying just P1 can be established using ideas from the analysis of, e.g., the "maximum cardinality search" algorithm for chordal graphs (Tarjan and Yannakakis (1984), see also Corollary 2 of Wienöbst et al. (2021)). We state this here as a lemma, and provide the proof in supplementary material Section A.

Lemma 3.4. *If D is a DAG without v-structures, then D has a topological ordering σ satisfying P1 of Definition 3.2.*

We now prove the theorem.

Proof of Theorem 3.3. Let $\mathcal{O}$ be the set of topological orderings of D which satisfy P1 of Definition 3.2. By Lemma 3.4, $\mathcal{O}$ is non-empty. If there is a $\sigma \in \mathcal{O}$ which also satisfies P2 of Definition 3.2, then we are done.

We now proceed to show by contradiction that such a σ must indeed exist. So, suppose for the sake of contradiction that for each σ in $\mathcal{O}$, P2 is violated. Then, for each $\sigma \in \mathcal{O}$, there exist vertices a, b and an index i such that $a, b \in L_i(\sigma)$, $\sigma(b) = \sigma(a) + 1$, and there exists a parent of a in D that is not a parent of b. For any given $\sigma \in \mathcal{O}$, we choose a, b as above so that $\sigma(a)$ is as small as possible. With such a choice of a for each $\sigma \in \mathcal{O}$, we then define a function $f : \mathcal{O} \to [n-1]$ by defining $f(\sigma) = \sigma(a)$. Note that by the assumption that P2 is violated by each σ in $\mathcal{O}$, f is defined for each σ in $\mathcal{O}$. But then, since $\mathcal{O}$ is a finite set, there must be some $\sigma \in \mathcal{O}$ for which $f(\sigma)$ attains its maximum value. We obtain a contradiction by exhibiting another $\tau \in \mathcal{O}$ for which $f(\tau)$ is *strictly* larger than $f(\sigma)$. We first describe the construction of τ from σ, and then prove that τ so constructed is in $\mathcal{O}$ and has $f(\tau) > f(\sigma)$.

Construction. Given $\sigma \in \mathcal{O}$, let $f(\sigma) = j \in [n-1]$. Let $\sigma(a) = j, \sigma(b) = j+1$ and suppose that $a, b \in L_i(\sigma)$. By the definition of f, there is a parent of a that is not a parent of b. Define C^a to be the set $\{a\} \cup \mathrm{pa}_D(a)$. Since D has no v-structures, C^a is a clique in skeleton (D). Define $S^a := \{z | C^a \subseteq \mathrm{pa}_D(z)\}$. Let Y_a be the set of nodes occurring after a in σ that are not in S^a (note that $b \in Y_a$, and, in general, $z \in Y_a$ if and only if $\sigma(z) > \sigma(a)$ and there is an $x \in C^a$ that is not adjacent to z).

We now note the following easy to verify properties of the sets S^a and Y_a (the proof is provided in supple-

mentary material Section B.2).

Proposition 3.5. *1. If $y \in Y_a$ and z is a child of y in D, then $z \in Y_a$.*

2. Suppose that $x \in \{a\} \cup S^a$ is not a maximal-clique-sink node of D. Then there exists $y \in S^a$ such that $x \in \mathrm{pa}_D(y)$ and such that y is a maximal-clique-sink node in D. In particular, S^a is non-empty.

3. Suppose that x is a maximal-clique-sink node in the induced DAG $H := D[S^a]$. Then x is also a maximal-clique-sink node in D.

The ordering τ is now defined as follows: the first j nodes in τ are the same as σ. After this, the nodes of S^a appear according to some topological ordering γ of the induced DAG $H := D[S^a]$ that satisfies P1 of Definition 3.2 in H (such a γ exists because of Lemma 3.4 applied to the induced DAG H, which also cannot have any v-structures). Finally, the nodes of Y_a appear according to their ordering in σ.

Proof that $\tau \in \mathcal{O}$ and $f(\tau) > f(\sigma)$. Note first that τ is a topological ordering of D: if not, then there must exist $u \in Y_a$ and $v \in S^a$ such that the edge $u \to v$ is present in D, but this cannot happen by item 1 of Proposition 3.5 above.

To show that $\tau \in \mathcal{O}$ (i.e., that τ satisfies P1), the following notation will be useful. For each maximal-clique-sink vertex s in D, denote by $\lambda(s)$ the unique $L_\alpha(\sigma)$ such that $s \in L_\alpha(\sigma)$. Similarly, denote by $\mu(s)$ the unique $L_\beta(\tau)$ such that $s \in L_\beta(\tau)$. Since $\sigma \in \mathcal{O}$, we already know that $\lambda(s)$ is a clique for each maximal-clique-sink node s of D. In order to show that $\tau \in \mathcal{O}$, all we need to show is that $\mu(s)$ is also a clique for each maximal-clique-sink node s of D.

Let J be the set of maximal-clique-sink vertices of D present in S^a. By item 2 of Proposition 3.5, the last vertex in γ must be an element of J. Note also that $s_i \notin J$ since $b \in Y_a$ and the edge $b \to s_i$ in D together imply that $s_i \in Y_a$ by item 1 of Proposition 3.5. We also observe that $\lambda(s) \subseteq S^a$, for all $s \in J$. For if there exists $u \in \lambda(s) \cap Y_a$ then the edge $u \to s$ in D implies that $s \in Y_a$, contradicting that $s \in S^a$. From the construction of τ, we already have $\mu(s_j) = \lambda(s_j)$ for all maximal-clique-sink vertices s_j that precede s_i in σ. From the fact that all vertices in S^a precede Y_a in the ordering τ, and from the observations above that (i) the sink node $s_i \in Y_a$, and (ii) $\lambda(s) \subseteq S^a$, for all sink nodes $s \in S^a$, we also get that for any maximal-clique-sink node s of D such that $s \in Y_a$, $\mu(s) \subseteq \lambda(s)$. Thus, when s is a maximal-clique-sink node of D that is not in $J \subset S^a$, we have that $\mu(s)$ is a clique in skeleton (D), since $\mu(s) \subseteq \lambda(s)$, and $\lambda(s)$ is a clique in skeleton (D). It remains to show that $\mu(s)$ is a clique when $s \in J$.

Let $t_1, t_2, \ldots, t_k$ be the maximal-clique-sink nodes of $H = D[S^a]$, arranged in increasing order by γ. Since γ satisfies P1 in H, each $L_i(\gamma)$, $1 \le i \le k$, is a clique in H (and thus also in D). Now consider a maximal-clique-sink node $s \in J \subseteq S^a$. Since the t_i are maximal-clique-sink nodes of D (from item 3 of Proposition 3.5), it follows that $\mu(s) \subseteq L_i(\gamma)$ (if $s \in L_i(\gamma)$ for $i \ge 2$) or $\mu(s) \subseteq L_1(\gamma) \cup C^a$ (if $s \in L_1(\gamma)$). In the former case, $\mu(s)$ is automatically a clique, since $L_i(\gamma)$ is a clique in D. In the latter case also $\mu(s)$ is a clique since $L_1(\gamma) \subseteq S^a$, so that $L_1(\gamma) \cup C^a$ is a clique in D (since (i) $L_1(\gamma)$ and C^a are cliques, and (ii) by definition of S^a, every node of S^a is adjacent to every node in C^a).

Thus, we get that τ also satisfies P1, so that $\tau \in \mathcal{O}$. Consider the node $b' := \tau(j+1)$ next to a in τ. Since S^a is non-empty, the construction of τ implies $b' \in S^a$, so that b' is adjacent to all parents of a. Since σ and τ agree on the ordering of all vertices up to a, we thus have $f(\tau) \ge \tau(b') = j+1 > j = f(\sigma)$. This gives the desired contradiction to σ being chosen as a maximum of f. Thus, there must exist some ordering in $\mathcal{O}$ which satisfies P2. □

Theorem 3.6. *Let D be a DAG without v-structures with n nodes. Then, any set I of atomic interventions that fully orients* skeleton (D) *when the ground-truth DAG is D must be of size at least $\lceil \frac{n-r}{2} \rceil$ (*skeleton (D) *is also the MEC of D, since D has no v-structures). Here r is the number of distinct maximal cliques in* skeleton (D)*. In other words, if I is a set of atomic interventions such that $\mathcal{E}_I(D) = D$, then $|I| \ge \lceil \frac{n-r}{2} \rceil$.*

Proof. Consider a CBSP ordering σ of nodes in D. Let a, b be two nodes such that $a, b \in L_i(\sigma)$ for some $i \in [r]$ and such that $\sigma(b) = \sigma(a) + 1$. Consider a set I of atomic interventions such that $I \cap \{a, b\} = \varnothing$. We show now that the edge $a - b$ is not directed in the I-essential graph $\mathcal{E}_I(D)$ of D.

Suppose, for the sake of contradiction, that $a \to b$ is directed in $\mathcal{E}_I(D)$. Then, by item 4 of Theorem 2.1, $a \to b$ must be strongly I-protected in $\mathcal{E}_I(D)$. Since $I \cap \{a, b\} = \varnothing$, one of the graphs in Figure 1 must appear as an induced subgraph of $\mathcal{E}_I(D)$.

We now show that none of these subgraphs can appear as an induced subgraph of $\mathcal{E}_I(D)$. First, subgraphs (ii) and (iv) cannot be induced subgraphs of $\mathcal{E}_I(D)$ since they have a v-structure at b while D (and therefore also $\mathcal{E}_I(D)$) has no v-structures. For subgraph (iii) to appear as an induced subgraph, the vertex c must lie between a and b in any topological ordering of D, which contradicts the fact that a and b are consecutive in the topological ordering σ. For subgraph (i)

to appear, we must have a parent c of a that is not adjacent to b. However, since σ is a CBSP ordering, it satisfies property P2 of Definition 3.2, so that, since a, b are consecutive in σ and belong to the same $L_i(\sigma)$, any parent of a must also be a parent of b. We thus conclude that $a \to b$ cannot be strongly I-protected in $\mathcal{E}_I(D)$, and hence is not directed in it.

The above argument implies that any set I of atomic interventions that fully orients D starting with skeleton (D) (i.e., for which $\mathcal{E}_I(D) = D$) must contain at least one node of each pair of consecutive nodes (in σ) of $L_i(\sigma)$, for each $i \in [r]$. Thus, for each $i \in [r]$, I must contain at least $\lceil (|L_i(\sigma)| - 1)/2 \rceil$ nodes of $L_i(\sigma)$. We therefore have,

$$|I| \geq \sum_{i=1}^{r} \left\lceil \frac{|L_i(\sigma)| - 1}{2} \right\rceil \geq \left\lceil \sum_{i=1}^{r} \frac{|L_i(\sigma)| - 1}{2} \right\rceil$$
$$= \left\lceil \frac{\sum_{i=1}^{r} |L_i(\sigma)|}{2} - \frac{r}{2} \right\rceil = \left\lceil \frac{n-r}{2} \right\rceil. \qquad \square$$

The following corollary for general DAGs (those that may have v-structures) follows from the previous result about DAGs without v-structures in a manner identical to previous work (Squires et al., 2020), using the fact that it is necessary and sufficient to separately orient each chordal chain component of an MEC in order to fully orient an MEC (Hauser and Bühlmann, 2014, Lemma 1). We defer the standard proof to supplementary material Section B.3.

Theorem 3.7. *Let D be an arbitrary DAG and let $\mathcal{E}(D)$ be the chain graph with chordal chain components representing the MEC of D. Let CC denote the set of chain components of $\mathcal{E}(D)$, and $r(S)$ the number of maximal cliques in the chain component $S \in CC$. Then, any set of atomic interventions which fully orients $\mathcal{E}(D)$ must be of size at least*

$$\sum_{S \in CC} \left\lceil \frac{|S| - r(S)}{2} \right\rceil \geq \left\lceil \frac{n-r}{2} \right\rceil,$$

where n is the number of nodes in D, and r is the total number of maximal cliques in the chordal chain components of $\mathcal{E}(D)$ (including chain components consisting of singleton vertices).

3.1 Tightness of Universal Lower Bound

We now show that our universal lower bound is tight up to a factor of 2: for any DAG D, there is a set of atomic interventions of size at most twice the lower bound that fully orients the MEC of D. In fact, as the proof of the theorem below (see supplementary material Section B.4) shows, when D has no v-structures, this intervention set can be taken to be the set of nodes of D that are *not* maximal-clique-sink nodes of D.

Theorem 3.8. *Let D be a DAG without v-structures with n nodes, and let r be the number of distinct maximal cliques in* skeleton (D)*. Then, there exists a set I of atomic interventions of size at most $n - r$ such that I fully orients* skeleton (D) *(i.e., $\mathcal{E}_I(D) = D$).*

Using again the fact that it is necessary and sufficient to separately orient each of the chordal chain components of an MEC in order to fully orient an MEC, the following result for general DAGs follows immediately from Theorem 3.8, and implies that the lower bound for general DAGs is also tight up to a factor of 2 (the proof is provided in supplementary material Section B.5).

Theorem 3.9. *Let D be an arbitrary DAG on n nodes and let $\mathcal{E}(D)$ and r be as in the notation of Theorem 3.7. Then, there is a set of atomic interventions of size at most $n - r$ that fully orients $\mathcal{E}(D)$.*

3.2 Comparison with Known Lower Bounds

To compare our universal lower bound with the universal lower bound of Squires et al. (2020), we start with the following combinatorial lemma, whose proof can be found in supplementary material Section B.6.

Lemma 3.10. *Let G be an undirected chordal graph on n nodes in which the size of the largest clique is ω. Then, $n - |\mathcal{C}| \geq \omega - 1$, where $\mathcal{C}$ is the set of maximal cliques of G.*

Lemma 3.10 implies that $\left\lceil \frac{n-|\mathcal{C}|}{2} \right\rceil \geq \left\lceil \frac{\omega-1}{2} \right\rceil = \left\lfloor \frac{\omega}{2} \right\rfloor$ in chordal graphs which shows that our universal lower bound is always equal to or better than the one by Squires et al. (2020). The proof of Lemma 3.10 makes it apparent that two bounds are close only in very special circumstances. (*Split graphs* and *k-trees* are some special families of chordal graphs for which $\left\lceil \frac{n-|\mathcal{C}|}{2} \right\rceil = \left\lfloor \frac{\omega}{2} \right\rfloor$). We further strengthen this intuition through theoretical analysis of special classes of graphs and via simulations.

Examples where our Lower Bound is Significantly Better We provide two constructions of special classes of chordal graphs in which our universal lower bound is $\Theta(k)$ times the $\left\lfloor \frac{\omega}{2} \right\rfloor$ lower bound of Squires et al. (2020) for any $k \in \mathbb{N}$. Further discussion of such examples can be found in supplementary material Section C.

Construction 1. First, we provide a construction by Shanmugam et al. (2015) for graphs that require about k times more number of interventions than their lower bound, where k is size of the maximum independent set of the graph. This construction of a chordal graph G starts with a line L consisting of vertices $1, \ldots, 2k$

such that each node $1 < i < 2k$ is connected to $i-1$ and $i+1$. For each $1 \leq p \leq k$, G has a clique C_p of size ω which has exactly two nodes $2p-1, 2p$ from the line L. Maximum clique size of G is ω, number of nodes, $n = k\omega$, and number of maximal cliques, $|\mathcal{C}| = 2k-1$. Thus, for G, we have, $n - |\mathcal{C}| = k(\omega-2)+1$ which implies $\left\lceil \frac{n-|\mathcal{C}|}{2} \right\rceil = \Theta(k) \left\lfloor \frac{\omega}{2} \right\rfloor$ for $\omega > 2$.

Construction 2. G has k cliques of size ω, with every pair of cliques intersecting at a unique node v. The number of nodes in G is $k(\omega-1)+1$, maximum clique size is ω, and number of maximal cliques is k, thus, $n-|\mathcal{C}| = k(\omega-2)+1$ which implies $\left\lceil \frac{n-|\mathcal{C}|}{2} \right\rceil = \Theta(k) \left\lfloor \frac{\omega}{2} \right\rfloor$ for $\omega > 2$.

3.3 Lower Bounds in Other Settings

Recall from the introduction that a universal lower bound must apply (i) no matter how the intervention set was generated, and (ii) for every ground-truth DAG D in the MEC. When further conditions are imposed on the process generating the intervention set, one can prove better (i.e., larger) lower bounds: Lindgren et al. (2018) show that the number of atomic interventions required to fully orient a chordal MEC, when the set of interventions is generated by a non-adaptive (or *passive*) algorithm, is at least the size of the minimum vertex cover of the chordal MEC. Similarly, a worst-case ground-truth DAG in the MEC may require a larger set of atomic interventions than that required by the universal lower bound: Shanmugam et al. (2015) show that for any active or passive algorithm, there exists a ground-truth DAG without v-structures for which at least $\frac{\omega}{k} \log_{\frac{\omega e}{k}} \omega$ interventions of size at most k are required in order to fully learn the DAG starting from its MEC (here, ω is the size of the largest clique in the chordal MEC of the DAG).

Lower bounds can also be much higher in the presence of *latent variables*, a setting not considered in this paper: Addanki et al. (2020) construct families of causal graph with latent variables such that for any causal graph $\mathcal{G}$ in the family, any passive algorithm requires $\Omega(n)$ atomic interventions to recover the observable graph of $\mathcal{G}$, where n is the number of nodes in the observable graph.

4 EMPIRICAL EXPLORATIONS

In this section, we report the results of two experiments on synthetic data. In *Experiment* 1, we compare our lower bound with the *optimal intervention size* for a large number of randomly generated DAGs. Optimal intervention size for a DAG D is defined as the size of the smallest set of atomic interventions I such that $\mathcal{E}_I(D) = D$. Next, in *Experiment* 2, we compare our universal lower bound with the one in the work of Squires et al. (2020) for randomly generated DAGs with small cliques. These experiments provide empirical evidence that strengthens our result about the tightness of our universal lower bound (Theorem 3.9) and the constructions presented in Section 3.2. The experiments use the open source `causaldag` (Squires, 2018) and `networkx` (Hagberg et al., 2008) packages. Further details about the experimental setup for both experiments are given in supplementary material Section D.

Experiment 1 For this experiment, we generate 1000 graphs from Erdős-Rényi graph model $G(n,p)$: for each of these graphs, the number of nodes n is a random integer in $[5, 25]$ and the connection probability p is a random value in $[0.1, 0.3)$. These graphs are then converted to DAGs without v-structures by imposing a random topological ordering and adding extra edges if needed. To compute the optimal intervention size, we check if a subset of nodes, I of a DAG D is such that $\mathcal{E}_{\mathcal{I}}(D) = D$, in increasing order of the size of such subsets. Next, we compute the universal lower bound value for each of these DAGs as given in Theorem 3.7. In Figure 3, we plot the optimal intervention size and our lower bound for each of the generated DAGs. Thickness of the points is proportional to the number of points landing at a coordinate. Notice that, all points lie between lines $y = x$ and $y = 2x$, as implied by our theoretical results. Further, we can see that, a large fraction of points are closer to the line $y = x$ compared to the line $y = 2x$, suggesting that our lower bound is even tighter for many graphs.

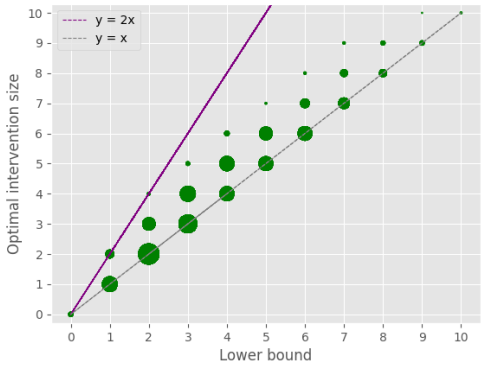


Figure 3: Comparison of the Optimal Intervention Set Size with our Universal Lower Bound

Experiment 2 For this experiment, we generate 1000 random DAGs for each size in

$\{10, 20, 30, 40, 50, 60\}$ by fixing a perfect-elimination ordering of the nodes and then adding edges (which are oriented according to the perfect-elimination ordering) to the DAG making sure that there are no v-structures, while trying to keep the size of each clique below 5. For each DAG, we compute the ratio of the two lower bounds. In Figure 4, we plot each of these ratios in a scatter plot with the x-axis representing the number of nodes of the DAG. Thickness of the points is proportional to the number of DAGs having a particular value of the ratio described above. We also plot the average of the ratios for each different value of the number of nodes. We see that our lower bound can sometimes be ~ 5 times of the lower bound of Squires et al. (2020). Moreover, the average ratio has an increasing trend suggesting that our lower bound is much better for this class of randomly generated DAGs.

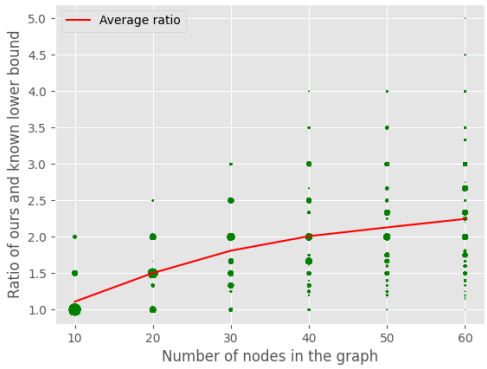


Figure 4: Comparison of our Universal Lower Bound with that of Squires et al. (2020)

5 CONCLUSION

We prove a strong universal lower bound on the minimum number of atomic interventions required to fully learn the orientation of a DAG starting from its MEC. For any DAG D, by constructing an explicit set of atomic interventions that learns D completely (starting with the MEC of D) and has size at most twice of our lower bound for the MEC of D, we show that our universal lower bound is tight up to a factor of two. We prove that our lower bound is better than the best previously known universal lower bound (Squires et al., 2020) and also construct explicit graph families where it is significantly better. We then provide empirical evidence that our lower bound may be stronger than what we are able to prove about it: by conducting experiments on randomly generated graphs, we demonstrate that our lower bound is often tighter (than what we have proved), and also that it is often significantly better than the previous universal lower bound (Squires et al., 2020). An interesting direction for future work is to design intervention sets of sizes close to our universal lower bound. Another direction for future work is to better understand the power of non-atomic interventions, especially in practical interventional cost models.

We note that in contrast to the earlier work of Squires et al. (2020), whose lower bound proofs were based on new sophisticated constructions, our proof is based on the simpler notion of a CBSP ordering, which in turn is inspired from elementary ideas in the theory of chordal graphs. We expect that the notion of CBSP orderings may also play an important role in future work on designing optimal intervention sets.

Acknowledgments

We thank the reviewers and meta-reviewer for their helpful suggestions to improve our paper. Piyush Srivastava acknowledges support from the Department of Atomic Energy, Government of India, under project no. RTI4001, from the Ramanujan Fellowship of SERB, from the Infosys foundation, through its support for the Infosys-Chandrasekharan virtual center for Random Geometry, and from Adobe Systems Incorporated via a gift to TIFR. The contents of this paper do not necessarily reflect the views of the funding agencies listed above.

References

Addanki, R., Kasiviswanathan, S. P., McGregor, A., and Musco, C. (2020). Efficient Intervention Design for Causal Discovery with Latents. *Proceedings of the 37th International Conference on Machine Learning (ICML 2020), PMLR*, 119:63–73. arXiv:2005.11736.

Addanki, R., McGregor, A., and Musco, C. (2021). Intervention Efficient Algorithms for Approximate Learning of Causal Graphs. *Proceedings of the 32nd International Conference on Algorithmic Learning Theory (ALT 2021), PMLR*, 132:151–184. arXiv:2012.13976.

Agrawal, R., Squires, C., Yang, K. D., Shanmugam, K., and Uhler, C. (2019). ABCD-Strategy: Budgeted Experimental Design for Targeted Causal Structure Discovery. *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics (AISTATS 2019), PMLR*, 89:3400–3409.

AhmadiTeshnizi, A., Salehkaleybar, S., and Kiyavash, N. (2020). LazyIter: A Fast Algorithm for Counting Markov Equivalent DAGs and Designing Experiments. *Proceedings of the 37th International Con-*

ference on Machine Learning (ICML 2020), PMLR, 119:125–133.

Andersson, S. A., Madigan, D., and Perlman, M. D. (1997). A Characterization of Markov Equivalence Classes for Acyclic Digraphs. *Annals of Statistics*, 25(2):505–541.

Blair, J. R. S. and Peyton, B. (1993). An Introduction to Chordal Graphs and Clique Trees. In *Graph Theory and Sparse Matrix Computation*, volume 56 of *The IMA Volumes in Mathematics and its Applications*, pages 1–29. Springer.

Bottou, L., Peters, J., Quiñonero-Candela, J., Charles, D. X., Chickering, D. M., Portugaly, E., Ray, D., Simard, P., and Snelson, E. (2013). Counterfactual Reasoning and Learning Systems: The Example of Computational Advertising. *Journal of Machine Learning Research*, 14(65):3207–3260.

Chickering, D. M. (1995). A Transformational Characterization of Equivalent Bayesian Network Structures. *Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence (UAI 1995)*, pages 87–98. arXiv:1302.4938.

Eberhardt, F. (2008). Almost Optimal Intervention Sets for Causal Discovery. *Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence (UAI 2008)*. arXiv:1206.3250.

Eberhardt, F., Glymour, C., and Scheines, R. (2005). On the Number of Experiments Sufficient and in the Worst Case Necessary to Identify All Causal Relations Among N Variables. *Proceedings of the 21st Conference on Uncertainty in Artificial Intelligence (UAI 2005)*. arXiv:1207.1389.

Friedman, N. (2004). Inferring Cellular Networks Using Probabilistic Graphical Models. *Science*, 303(5659):799–805.

Ghassami, A., Salehkaleybar, S., Kiyavash, N., and Bareinboim, E. (2018). Budgeted Experiment Design for Causal Structure Learning. *Proceedings of the 35th International Conference on Machine Learning, (ICML 2018), PMLR*, 80:1719–1728.

Glymour, C., Zhang, K., and Spirtes, P. (2019). Review of Causal Discovery Methods Based on Graphical Models. *Frontiers in Genetics*, 10:524.

Greenewald, K., Katz, D., Shanmugam, K., Magliacane, S., Kocaoglu, M., Boix Adsera, E., and Bresler, G. (2019). Sample Efficient Active Learning of Causal Trees. *Proceedings of 33rd Annual Conference on Neural Information Processing Systems (NeurIPS 2019)*.

Hagberg, A. A., Schult, D. A., and Swart, P. J. (2008). Exploring Network Structure, Dynamics, and Function using NetworkX. *Proceedings of the 7th Python in Science Conference (SciPy 2008)*, pages 11–15.

Hauser, A. and Bühlmann, P. (2012). Characterization and Greedy Learning of Interventional Markov Equivalence Classes of Directed Acyclic Graphs. *Journal of Machine Learning Research*, 13:2409–2464.

Hauser, A. and Bühlmann, P. (2014). Two Optimal Strategies for Active Learning of Causal Models from Interventional Data. *International Journal of Approximate Reasoning*, 55(4):926–939.

He, Y.-B. and Geng, Z. (2008). Active Learning of Causal Networks with Intervention Experiments and Optimal Designs. *Journal of Machine Learning Research*, 9(84):2523–2547.

Hyttinen, A., Eberhardt, F., and Hoyer, P. O. (2013a). Experiment Selection for Causal Discovery. *Journal of Machine Learning Research*, 14(57):3041–3071.

Hyttinen, A., Hoyer, P. O., Eberhardt, F., and Jarvisalo, M. (2013b). Discovering Cyclic Causal Models with Latent Variables: A General SAT-Based Procedure. *Proceedings of the 29th Confrence on Uncertainty in Artificial Intelligence (UAI 2013)*. arXiv:1309.6836.

Kocaoglu, M., Dimakis, A., and Vishwanath, S. (2017). Cost-Optimal Learning of Causal Graphs. *Proceedings of the 34th International Conference on Machine Learning (ICML 2017), PMLR*, 70:1875–1884. arXiv:1703.02645.

Lindgren, E. M., Kocaoglu, M., Dimakis, A. G., and Vishwanath, S. (2018). Experimental Design for Cost-Aware Learning of Causal Graphs. *arXiv preprint arXiv:1810.11867*.

Meek, C. (1995). Causal Inference and Causal Explanation with Background Knowledge. *Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence (UAI 1995)*, pages 403–410. arXiv:1302.4972.

Pearl, J. (2009). *Causality: Models, Reasoning and Inference*. Cambridge University Press, USA, 2nd edition.

Shanmugam, K., Kocaoglu, M., Dimakis, A. G., and Vishwanath, S. (2015). Learning Causal Graphs with Small Interventions. *Proceedings of the 29th Annual Conference on Neural Information Processing Systems (NeurIPS 2015)*, pages 3195–3203. arXiv:1511.00041.

Shen, X., Ma, S., Vemuri, P., Simon, G., and The Alzheimer's Disease Neuroimaging Initiative (2020). Challenges and Opportunities with Causal Discovery Algorithms: Application to Alzheimer's Pathophysiology. *Scientific Reports*, 10:2975.

Squires, C. (2018). `Causaldag` Python Package. BSD License. Available from https://github.com/uhlerlab/causaldag.

Squires, C., Magliacane, S., Greenewald, K. H., Katz, D., Kocaoglu, M., and Shanmugam, K. (2020). Active Structure Learning of Causal DAGs via Directed Clique Trees. *Proceedings of 34th Annual Conference on Neural Information Processing Systems (NeurIPS 2020)*. arXiv:2011.00641.

Tarjan, R. E. and Yannakakis, M. (1984). Simple Linear-Time Algorithms to Test Chordality of Graphs, Test Acyclicity of Hypergraphs, and Selectively Reduce Acyclic Hypergraphs. *SIAM Journal on Computing*, 13(3):566–579.

Verma, T. S. and Pearl, J. (1990). Equivalence and Synthesis of Causal Models. *Proceedings of the 6th Conference on Uncertainty in Artificial Intelligence (UAI 1990)*, pages 220–227. arXiv:1304.1108.

Wienöbst, M., Bannach, M., and Liśkiewicz, M. (2021). Polynomial-Time Algorithms for Counting and Sampling Markov Equivalent DAGs. *Proceedings of the 35th AAAI Conference on Artificial Intelligence (AAAI 2021)*, pages 12198–12206. arXiv:2012.09679.

Supplementary Material: Almost Optimal Universal Lower Bound for Learning Causal DAGs with Atomic Interventions

A PROOF OF LEMMA 3.4 OF THE MAIN PAPER

Here, we provide the proof of Lemma 3.4 of the main paper. As stated there, the lemma follows from well-known ideas in the theory of chordal graphs. The following generalization of the definition of the clique block property P1 of Definition 3.2 will be useful in the proof.

Definition A.1 (*A*-**clique block ordering**). *Let D be a DAG and A a subset of vertices of D. Let σ be a topological ordering of D. Let the elements of A be $a_1, a_2, \ldots, a_k$, arranged so that $\sigma(a_i) < \sigma(a_j)$ whenever $i < j$. Define $L_1^A(\sigma)$ to be the set of nodes u which occur before or at the same position as a_1 in σ i.e., $\sigma(u) \leq \sigma(a_1)$. Similarly, for $2 \leq i \leq k$, define $L_i^A(\sigma)$ to be the set of nodes which occur in σ before or at the same position as a_i, but strictly after a_{i-1} (i.e., $\sigma(a_{i-1}) < \sigma(u) \leq \sigma(a_i)$). Then, σ is said to be an* A-clique block ordering *of D if $\cup_{i=1}^{k} L_i^A(\sigma)$ is the set of all vertices of D, and for each $1 \leq i \leq k$, $L_i^A(\sigma)$ is a (not necessarily maximal) clique in* skeleton (D).

The following observation is immediate with this definition.

Observation A.2. *Let D be a DAG without v-structures, and let A be the set of maximal-clique-sink vertices of D. Then, a topological ordering σ of D satisfies property P1 of Definition 3.2 if and only if σ is an A-clique block ordering of D.*

Proof. The "if" direction follows from the definition. For the "only if" direction, we note that since every vertex of D must be contained in some maximal clique C of skeleton (D), and since $C = \{s\} \cup \mathrm{pa}_D(s)$ for some maximal-clique-sink vertex s, it follows that every vertex of D must lie in some $L_i(\sigma)$ if σ satisfies the clique block property P1. □

We also note the following simple property of A-clique block orderings.

Observation A.3. *Let D be a DAG without v-structures, and let σ be a topological ordering of D. Let A and B be subsets of vertices of D such that $A \subseteq B$. If σ is an A-clique block ordering of D, then it is also a B-clique block ordering of D.*

Proof. When $A = B$, there is nothing to prove. Thus, we can assume that there must exist a $b \in B \setminus A$. We consider the case when $B = A \cup \{b\}$. The general case then follows by straightforward induction on the size of $|B \setminus A|$.

For $1 \leq i \leq |A|$, let $L_i^A(\sigma)$ be as in the definition of the A-clique block orderings. Let i be the unique index such that $b \in L_i^A(\sigma)$. Now, for $j < i$, define $L_j^B(\sigma) := L_j^A(\sigma)$, and for $j \geq i+1$, define $L_{j+1}^B(\sigma) := L_j^A(\sigma)$. By construction, for $1 \leq j \leq i-1$ and $i+2 \leq j \leq |A|+1$, the $L_j^B(\sigma)$ are cliques in skeleton (D). Further define

$$L_i^B(\sigma) := \left\{u \in L_i^A(\sigma) \mid \sigma(u) \leq \sigma(b)\right\}, \text{ and}$$
$$L_{i+1}^B(\sigma) := L_i^A(\sigma) \setminus L_i^B(\sigma).$$

Again, $L_i^B(\sigma)$ and $L_{i+1}^B(\sigma)$ are also cliques in skeleton (D) since they are subsets of the clique $L_i(\sigma)$. Further, by construction, $\cup_{j=1}^{|B|} L_j^B(\sigma) = \cup_{j=1}^{|A|} L_j^A(\sigma)$. This shows that σ is also a B-clique block ordering. □

We can now state the main technical lemma required for the proof of Lemma 3.4 of the main paper.

Lemma A.4. *Let D be a DAG without v-structures. Let S be the set of maximal-clique-sink vertices of D. Then, there exists a maximal clique C of* skeleton (D) *with the following two properties:*

1. *If $u \in C$ then $\mathrm{pa}_D(u) \subseteq C$. That is, if $u \in C$ and $v \notin C$, then the edge $v \to u$ is not present in D.*
2. *Let S' be the set of maximal-clique-sink nodes of the induced DAG $D[V \setminus C]$, where V is the set of nodes of D. Then S' is a subset of $S \setminus \{\mathrm{sink}_D(C)\}$.*

Proof. As already alluded to in the main paper, the proof of item 1 uses ideas that are very similar to the "maximal cardinality search" algorithm for chordal graphs (Tarjan and Yannakakis (1984), see also Corollary 2 of Wienöbst et al. (2021)). Fix an arbitrary topological ordering τ of D, and let $v_1 = \tau(1)$ be the top vertex in τ. Note that v_1 has no parents in D, so any vertices adjacent to v_1 in D are children of v_1. Let C' be the set of these children of v_1 in D. If C' is empty, then v_1 is isolated in D and we are done with the proof of item 1 after taking $C = v_1$. So, assume that C' is not empty, and let its elements be $c_1, c_2, \ldots c_k$, arranged so that $\tau(c_i) < \tau(c_j)$ whenever $i < j$. Now, define the sets C'_i, where $1 \leq i \leq k$ as follows. First, $C'_1 := \{v_1, c_1\}$. For $2 \leq i \leq k$,

$$C'_i := \begin{cases} C'_{i-1} \cup \{c_i\} & \text{if } C'_{i-1} \subseteq \mathrm{pa}_D(c_i), \\ C'_{i-1} & \text{otherwise.} \end{cases} \tag{1}$$

Define $C := C'_k$. Note that by construction, C is a clique. Note also the following property of this construction: for any i, $c_i \notin C$ if and only if there exists $1 \leq j < i$ such that $c_j \in C$ and c_j is not adjacent to c_i in D.

We now claim that C is also a maximal clique. For, if not, let $u \notin C$ be such that u is adjacent to every vertex in C. Then, we must have $u = c_i$ for some i (since $v_1 \in C$, and only the children of v_1 are adjacent to v_1 in D). But then, since $u = c_i \notin C$, there must exist some c_j, $j < i$, such that $c_j \in C$ is not adjacent to $u = c_i$, which is a contradiction to the assumption of u being adjacent to every vertex of C.

We now claim that if $v \notin C$, then for all $u \in C$, the edge $v \to u$ is not present in D. Suppose, if possible, that there exist $v \notin C$ and $u \in C$ such that $v \to u$ is present in D. By the choice of v_1 as a top vertex in a topological order, we must have $u \neq v_1$. Thus, u must be a child of v_1 in D. Suppose $u = c_i$, for some $i \in [k]$. Then, v must also be a child of v_1, for otherwise $v \to c_i \leftarrow v_1$ would be a v-structure in D. Thus, $v = c_j$ for some $j < i$. Since $c_j = v \notin C$, there exists some $\ell < j$ such that $c_\ell \in C$ and c_ℓ and $v = c_j$ are not adjacent. But then $c_\ell \to c_i \leftarrow c_j$ is a v-structure in D, so we again get a contradiction. This proves item 1 of the lemma for the clique C.

Item 2 of the lemma trivially follows if $V \setminus C$ is empty, therefore, we are interested in the case when $V \setminus C$ is non-empty. Now consider the induced DAG $H := D[V \setminus C]$. Since D has no v-structures, neither does H. Thus, by Observation 3.1 of the main paper, the maximal-clique-sink nodes of H and the maximal cliques of skeleton(H) are in one-to-one correspondence: for each maximal clique C' of skeleton(H), there is a unique vertex $\mathrm{sink}_H(C')$ of H such that $C' = \{\mathrm{sink}_H(C')\} \cup \mathrm{pa}_H(\mathrm{sink}_H(C'))$.

Consider now a maximal-clique-sink vertex s' of H. There exists then a maximal clique C' of skeleton(H) such that $s' = \mathrm{sink}_H(C')$. Also, since H is an induced subgraph of D, there must exist a maximal clique C'' of skeleton(D) such that $C'' \supseteq C'$. In fact, we must further have $C'' \cap (V \setminus C) = C'$, for otherwise C' would not be a maximal clique of H. Let $t := \mathrm{sink}_D(C'')$. We will show that $t = s'$. Note first that we cannot have $t \in C$, for then, by item 1, $C'' = \{t\} \cup \mathrm{pa}_D(t)$ would be contained in C, and would not therefore contain $C' \subseteq V \setminus C$. Thus, t must be a node in $V \setminus C$. But then $C'' \cap (V \setminus C) = C'$ implies that t must in fact be in C', and must therefore be equal to $s' = \mathrm{sink}_H(C')$. We thus see that any maximal-clique-sink vertex s' of H is also a maximal-clique-sink vertex of D. Item 2 of the lemma then follows by noting that $\mathrm{sink}_D(C)$ is the only maximal-clique-sink vertex of D not contained in $V \setminus C$. □

We are now ready to prove Lemma 3.4 of the main paper.

Proof of Lemma 3.4 of the main paper. We prove this claim by induction on the number of nodes in D. The claim of the lemma is trivially true when D has only one node. Now, fix $n > 1$, and assume the induction hypothesis that every DAG without v-structures which has at most $n-1$ nodes admits a topological ordering that satisfies the clique block property P1 of Definition 3.2. We will complete the induction by showing that if D is a DAG without v-structures which has n nodes, then D also admits a topological ordering that satisfies the clique block property P1 of Definition 3.2.

Let the maximal clique C of D be as guaranteed by Lemma A.4 above. If all the nodes of D are contained in C, then the total ordering on the vertices of the clique C in D trivially satisfies the clique block property. Therefore,

we assume henceforth that $V \setminus C$ is non-empty. Thus, the induced DAG $H := D[V \setminus C]$ is a DAG on at most $n-1$ nodes. Let S' be the set of maximal-clique-sink nodes of H, and let S be the set of maximal-clique-sink nodes of D. By the induction hypothesis, H has a topological ordering τ which satisfies the clique block property. Equivalently, by Observation A.2, τ is an S'-clique block ordering of H.

Consider now the ordering σ of D obtained by listing first the vertices of the clique C in the total order imposed on them by the DAG D, followed by the vertices of $V \setminus C$ in the order specified by τ. By item 1 of Lemma A.4, there is no directed edge in D from a vertex in $V \setminus C$ to a vertex in C, so we get that σ is in fact a topological ordering of D.

Define $T = S' \cup \text{sink}_D(C)$. We now observe that σ is a T-clique block ordering of D, with $L_1^T(\sigma) = C$ and $L_{i+1}^T(\sigma) = L_i^{S'}(\tau)$, for $1 \leq i \leq |S'|$. By item 2 of Lemma A.4, we have $T \subseteq S$. Thus, by Observation A.3, σ is also an S-clique block ordering of D, and therefore (by Observation A.2) satisfies the clique block property P1 of Definition 3.2. □

B OTHER OMITTED PROOFS

B.1 Proof of Observation 3.1

Proof of Observation 3.1 of main paper. Let C be a maximal clique of skeleton(D). Since the induced subgraph $D[C]$ is a DAG, there is at least one node s in $D[C]$ with out-degree 0. Thus, for all $v \in C, v \neq s$ we have, $v \to s$, which implies that $C \setminus \{s\} \subseteq \text{pa}_D(s)$. Now, $\text{pa}_D(s) \cup \{s\}$ must be a clique as D does not contain v-structures. Thus, we must indeed have $\text{pa}_D(s) = C \setminus \{s\}$ since C is maximal. We thus see that there is a unique $s \in C$ such that $C = \text{pa}_D(s) \cup \{s\}$.

Now, suppose, if possible that there exist distinct maximal cliques C_1 and C_2 of skeleton(D) such that $\text{sink}_D(C_1) = \text{sink}_D(C_2) = s$. Since C_1 and C_2 are distinct maximal cliques, there must exist $a \in C_1 \setminus C_2, b \in C_2 \setminus C_1$ such that a is not adjacent to b. But then, since we have $a \in \text{pa}_D(s)$ and $b \in \text{pa}_D(s)$, we would have a v-structure $a \to s \leftarrow b$, which is a contradiction to the hypothesis that D has no v-structures. □

B.2 Proof of Proposition 3.5

Proof of Proposition 3.5 from the main paper. We use the same notation as in the proof of Theorem 3.3.

1. Since $y \in Y_a$, there exists $u \in C^a$ such that u is not adjacent to y in D. But then, if $z \in S^a$, we get the v-structure $u \to z \leftarrow y$, which is a contradiction to D not having any v-structures. This proves item 1 of the proposition.

2. Consider the clique $C := \{x\} \cup \text{pa}_D(x)$. There exists a maximal clique C' in skeleton(D) such that $C' \supsetneq C$, since x is not a maximal-clique-sink node. Set $y := \text{sink}_D(C')$. Since x is not a maximal-clique-sink node in D, we thus have $x \in C \subseteq \text{pa}_D(y)$, so that y is a child of x. We also have $y \in S^a$ since $C^a \subseteq \{x\} \cup \text{pa}_D(x) = C \subseteq \text{pa}_D(y)$, where the first inclusion comes from the assumption that $x \in \{a\} \cup S^a$. The fact that S^a is non-empty follows by applying the item with $x = a$, and noticing that, by construction, a is not a maximal-clique-sink vertex in D. This follows since $a \in L_i(\sigma)$ has a child (namely, b) in $L_i(\sigma)$, while by the definition of the L_i, only the last vertex in $L_i(\sigma)$ is a maximal-clique-sink node of D. This proves item 2 of the proposition.

3. Since x is a maximal-clique-sink node in $H = D[S^a]$, $C := \{x\} \cup \text{pa}_H(x)$ is a maximal clique in skeleton(H), and thereby a clique in skeleton(D). However, note that $C' := C \cup C^a$ is also a clique in skeleton(D), since both C and C^a are cliques in skeleton(D), and as $C \subseteq S^a$, every vertex in C is adjacent to every vertex in C^a. Thus, there exists a maximal clique C'' in skeleton(D) such that $C' \subseteq C''$.

   Consider $y := \text{sink}_D(C'')$. Suppose, if possible, that $x \neq y$. Then we must have $x \in \text{pa}_D(y)$ (since $x \in C''$), and also that $C^a \subseteq \text{pa}_D(y)$ (as $C^a \subseteq C''$). Thus, we must have $y \in S^a$. But then, we get that $\{y\} \cup C \subseteq C'' \cap S^a$, which contradicts the assumption that C is a maximal clique in H. Thus, we must have $x = y$, so that x is a maximal-clique-sink node in D. This proves item 3 of the proposition. □

B.3 Proof of Theorem 3.7

For use in this subsection and the next, we recall the characterization of $\mathcal{I}$-essential graphs (Theorem 2.1) from the main paper. Recall that in the main paper, this characterization was only used in the setting of DAGs without v-structures, in the proof of Theorem 3.6. Here, we will need to use it in the setting of general graphs. (Figure 1 in the statement of the theorem can be found in the main paper). Recall also that we always assume that every intervention set contains the empty intervention, but the empty intervention is not counted in the size of an intervention set.

Theorem 2.1 (Characterization of $\mathcal{I}$-essential graphs, Definition 14 and Theorem 18 of Hauser and Bühlmann (2012)). *Let D be a DAG and $\mathcal{I}$ an intervention set. A graph H is an $\mathcal{I}$-essential graph of D if and only if H has the same skeleton as D, all directed edges of H are directed in the same direction as in D, all v-structures of D are directed in H, and*

1. *H is a chain graph with chordal chain components.*
2. *For any three vertices a, b, c of H, the subgraph of H induced by a, b and c is not $a \to b - c$.*
3. *If $a \to b$ in D (so that a, b are adjacent in H) and there is an intervention $J \in \mathcal{I}$ such that $|J \cap \{a, b\}| = 1$, then $a \to b$ is directed in H.*
4. *Every directed edge $a \to b$ in H is strongly $\mathcal{I}$-protected. An edge $a \to b$ in H is said to be strongly $\mathcal{I}$-protected if either (a) there is an intervention $J \in \mathcal{I}$ such that $|J \cap \{a, b\}| = 1$, or (b) at least one of the four graphs in Figure 1 appears as an induced subgraph of H, and $a \to b$ appears in that induced subgraph in the configuration indicated in the figure.*

Remark B.1. *Strictly speaking, Theorem 18 of Hauser and Bühlmann (2012) only identifies the class of all $\mathcal{I}$-essential graphs. However, it is well known, and follows easily from their results that H satisfies all the four items in the statement of Theorem 2.1 along with the additional conditions in the theorem (i.e., H has the same skeleton as D, all directed edges of H are directed in the same direction as in D, and all v-structures of D are directed in H), if and only if $H = \mathcal{E}_{\mathcal{I}}(D)$.*

For completeness, we provide a proof of the above folklore remark in Section E. Here, we proceed to the following easy and folklore corollary of the characterization Theorem 2.1 that has a proof similar to the proof of Lemma 1 of Hauser and Bühlmann (2014). For completeness, we provide this proof in Section F.

Corollary B.2. *Let D be an arbitrary DAG, and let $\mathcal{I}$ be any intervention set containing the empty set. Let CC be the set of chordal chain components of the essential graph $\mathcal{E}(D) = \mathcal{E}_{\{\varnothing\}}(D)$ of D. For each $S \in CC$, define $\mathcal{I}_S := \{I \cap S \mid I \in \mathcal{I}\}$ to be the projection of the intervention set $\mathcal{I}$ to S. Consider vertices a and b that are adjacent in D. Then, the edge between a and b is directed in the $\mathcal{I}$-essential graph $\mathcal{E}_{\mathcal{I}}(D)$ if and only if one of the following conditions are true:*

1. *a and b are elements of distinct chain components of the observational essential graph $\mathcal{E}(D)$ (so that the edge between a and b is already directed in $\mathcal{E}(D)$).*
2. *a and b are in the same chain components $S \in CC$ of $\mathcal{E}(D)$ and the edge between a and b is directed in the $\mathcal{I}_S$-essential graph $\mathcal{E}_{\mathcal{I}_s}(D[S])$ of the induced DAG $D[S]$.*

In particular, $\mathcal{E}_{\mathcal{I}}(D) = D$ if and only if $\mathcal{E}_{\mathcal{I}_S}(D[S]) = D[S]$ for every $S \in CC$.

We now use the corollary to prove Theorem 3.7.

Proof of Theorem 3.7 from the main paper. Corollary B.2 says that an intervention set $\mathcal{I}$ learns D starting with $\mathcal{E}(D)$ if and only if $\mathcal{E}_{\mathcal{I}_S}(D[S]) = D[S]$ for every $S \in CC$. If $\mathcal{I}$ is a set of atomic interventions, then for each $I \in \mathcal{I}$, $|I \cap S| = 0$ for all but one of the $S \in CC$. This means that if an intervention set $\mathcal{I}$ of atomic interventions is such that $\mathcal{E}_{\mathcal{I}}(D) = D$, then $|\mathcal{I}| = \sum_{S \in CC} |\mathcal{I}_{\mathcal{S}}|$, and $\mathcal{E}_{\mathcal{I}_S}(D[S]) = D[S]$ for every $S \in CC$, where $\mathcal{I}_{\mathcal{S}}$ is a set of atomic interventions defined as in Corollary B.2. Since $D[S]$ is a DAG without v-structures, by Theorem 3.6 we

have, $|\mathcal{I}_S| \geq \left\lceil \frac{|S|-r(S)}{2} \right\rceil$ which implies,

$$|\mathcal{I}| \geq \sum_{S \in CC} \left\lceil \frac{|S|-r(S)}{2} \right\rceil \geq \left\lceil \sum_{S \in CC} \frac{|S|-r(S)}{2} \right\rceil = \left\lceil \frac{n-r}{2} \right\rceil.$$

This completes the proof. □

B.4 Proof of Theorem 3.8

Proof of Theorem 3.8 from the main paper. Fix any topological ordering σ of D. Let the maximal cliques of D be $C_1, \ldots, C_r$, and let $s_i := \text{sink}_D(C_i)$, for $i \in [r]$. Observation 3.1 implies that each node of $S = \{s_1, \ldots, s_r\}$ is distinct. We re-index these nodes according to the ordering σ, i.e. $\sigma(s_i) < \sigma(s_j)$ when $i < j$. Consider the set $I := V \setminus S$ of atomic interventions (note that $|I| = n - r$). We show that $\mathcal{E}_I(D) = D$. Note that every edge of D, except those which have both end-points in S, has a single end-point in one of the interventions in I, and hence is directed in $\mathcal{E}_I(D)$ (by item 3 of Theorem 2.1). We show now that all edges with both end-points in S are also oriented in $\mathcal{E}_I(D)$.

Suppose, if possible, that there exist $s_i, s_j \in S$, with $i < j$ such that s_i and s_j are adjacent in skeleton(D), so that the edge $s_i \to s_j$ is present in D, but for which $s_i - s_j$ is not directed in $\mathcal{E}_I(D)$. We derive a contradiction to this supposition. To start, choose an s_i, s_j as above with the smallest possible value of i. In particular, this choice implies that every edge of the form $u \to s_i$ in D is directed in $\mathcal{E}_I(D)$.

Note that, by Observation 3.1, $C_i = \{s_i\} \cup \text{pa}_D(s_i)$ and $C_j = \{s_j\} \cup \text{pa}_D(s_j)$ are distinct maximal cliques in skeleton(D). Thus, there must exist an $x \in C_i$ that is not a parent of s_j in D. Further, since $\sigma(s_i) < \sigma(s_j)$, all vertices of C_i appear before s_j in σ. Thus, $x \in C_i$ that is not a parent of s_j in D is also not adjacent to s_j in skeleton(D). Further, by the choice of i, the edge $x \to s_i$ is directed in $\mathcal{E}_I(D)$. Thus, we have the induced subgraph $x \to s_i - s_j$ in $\mathcal{E}_I(D)$. However, according to item 2 of Theorem 2.1, such a graph cannot appear as an induced subgraph of an I-essential graph $\mathcal{E}_I(D)$, and we have therefore reached the desired contradiction. It follows that $\mathcal{E}_I(D)$ has no undirected edges, and is therefore the same as D. □

B.5 Proof of Theorem 3.9

Here we restate Theorem 3.9 and provide its proof.

Theorem B.3 (Restatement of Theorem 3.9 from the main paper)**.** *Let D be an arbitrary DAG and let $\mathcal{E}(D)$ be the chain graph with chordal chain components representing the MEC of D. Let CC denote the set of chain components of $\mathcal{E}(D)$, and $r(S)$ the number of maximal cliques in the chain component $S \in CC$. Then, there exists a set I of atomic interventions of size at most $\sum_{S \in CC}(|S| - r(S)) = n - r$, such that I fully orients $\mathcal{E}(D)$ (i.e., $\mathcal{E}_I(D) = D$), where n is the number of nodes in D, and r is the total number of maximal cliques in the chordal chain components of $\mathcal{E}(D)$ (including chain components consisting of singleton vertices).*

Proof. Theorem 3.8 implies that for each $S \in CC$ there is a set $\mathcal{I}_S$ of atomic interventions such that $|\mathcal{I}_S| \leq |S| - r(S)$ and $\mathcal{E}_{\mathcal{I}_S}(D[S]) = D[S]$. Now, let $\mathcal{I} = \cup_{S \in CC} \mathcal{I}_S$. $\mathcal{E}_{\mathcal{I}}(D) = D$ by Corollary B.2, and $|\mathcal{I}| = \sum_{S \in CC} |\mathcal{I}_S|$, which means $|\mathcal{I}| \leq \sum_{S \in CC}(|S| - r(S)) = n - r$. This shows that there is a set of atomic interventions of size at most $n - r$ which fully orients $\mathcal{E}(D)$. □

B.6 Proof of Lemma 3.10

Proof of Lemma 3.10 from the main paper. Let C be a (necessarily maximal) clique of G of size ω. Since C is a maximal clique of the chordal graph G, there exists a perfect elimination ordering σ of G that starts with C (this is a consequence of the structure of the lexicographic breadth-first-search algorithm used to find perfect elimination orderings of chordal graphs: see, e.g., the paragraph before Proposition 1 of Hauser and Bühlmann (2014) and Algorithm 6 of Hauser and Bühlmann (2012) for a proof).

Now, let D be the DAG obtained by orienting the edges of G according to σ (i.e., the edge $u - v \in G$ is directed as $u \to v$ in D if and only if $\sigma(u) < \sigma(v)$). Suppose that $\text{sink}_D(C) = s$. Note that C cannot contain the node $\text{sink}_D(C')$ for any other maximal clique C' since, as σ starts with C, this would imply $C' \subseteq C$ and

would contradict the maximality of C'. Thus, there are $|\mathcal{C}| - 1$ maximal-clique-sink nodes of D other than s by Observation 3.1, and, by the above observation, they occur in σ after C. Thus, $n \geq |C| + |\mathcal{C}| - 1$, which gives $n - |\mathcal{C}| \geq \omega - 1$ as $|C| = \omega$. □

C VARIOUS EXAMPLE GRAPHS

In Section 3.2 of the main paper, we proved that our universal lower bound is always at least as good as the previous best universal lower bound given by Squires et al. (2020), and also gave examples of graph families where our bound is significantly better. We also pointed out that our lower bound and the lower bound of Squires et al. (2020) are close only in certain special circumstances. We now make give more details of these special cases.

We work with the same notation as that used in Lemma 3.10: G is an undirected chordal graph, n is the number of nodes in G, ω is the size of its largest clique, and $\mathcal{C}$ is the set of its maximal cliques. From Lemma 3.10, it follows that for our lower bound of $\left\lceil \frac{n-|\mathcal{C}|}{2} \right\rceil$ and the lower bound of $\left\lfloor \frac{\omega}{2} \right\rfloor = \left\lceil \frac{\omega-1}{2} \right\rceil$ of Squires et al. (2020) to be equal, one of the following conditions must be true: either (i) $n - |\mathcal{C}| = \omega - 1$, or (ii) ω is even and $n - |\mathcal{C}| = \omega$.

Now consider the perfect elimination ordering σ of G used in the proof of Lemma 3.10, and let D be the DAG with skeleton G constructed by orienting the edges of G in accordance with σ. Note that by the construction of σ, the ω vertices of a largest clique C of G are the first ω vertices in σ. Note also that $n - |\mathcal{C}|$ is the number of vertices in G that are *not* maximal-clique-sink nodes of D (by Observation 3.1).

Thus, it follows that condition (i) above for the two lower bounds to be equal can hold only when D is such that *all* nodes of G outside the largest clique C of G are maximal-clique-sink nodes of D. In other words, σ is a clique block ordering, in the sense of P1 of Definition 3.2 of a CBSP orderings, in which the first clique block $L_1(\sigma)$ consists of the largest clique C, while *all* other clique blocks $L_i(\sigma), i \geq 2$ are of size exactly 1. Similarly, condition (ii) above for the two lower bounds to be equal can hold only when D is such that *all but one* of the nodes of G outside the largest clique C of G are maximal-clique-sink nodes of G.

We now give examples of two special families of chordal graphs where the above conditions for the equality of the two lower bounds hold: Split graphs and k-trees. Here, $\mathcal{C}(G)$ will denote the set of maximal cliques of graph G.

Split graphs. G is a split graph if its vertices can be partitioned into a clique C and an independent set Z. For such a G, one of the following possibilities must be true.

1. $\exists x \in Z$ such that $C \cup \{x\}$ is complete. In this case, $C \cup \{x\}$ is a maximum clique and Z is a maximum independent set.

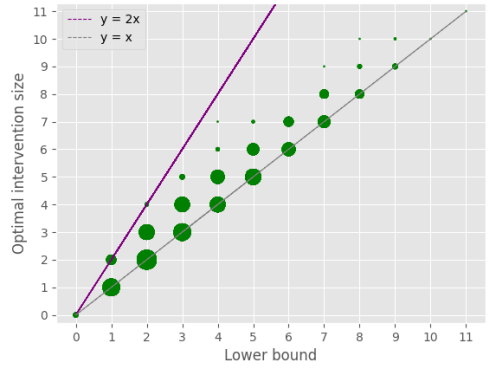
2. $\exists x \in C$ such that $Z \cup \{x\}$ is independent. In this case, $Z \cup \{x\}$ is a maximum independent set and C is a maximum clique.

3. C is a maximal clique and Z is a maximal independent set. In this case, C must also be a *maximum* clique and Z a *maximum* independent set.

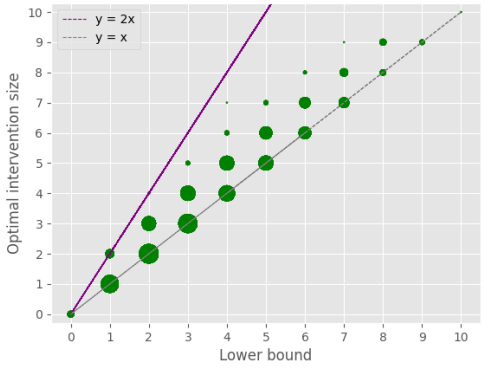
For each of these cases, we have $n - |\mathcal{C}(G)| = \omega - 1$, which implies $\left\lceil \frac{n-|\mathcal{C}(G)|}{2} \right\rceil = \left\lfloor \frac{\omega}{2} \right\rfloor$.

k-trees. A k-tree is formed by starting with K_{k+1} (complete graph with $k+1$ vertices) and repeatedly adding vertices in such a way that each added vertex v has exactly k neighbors, and such that these neighbors along with v form a clique. Thus, each added vertex creates exactly one clique of size $k+1$. In particular, in a k-tree, all maximal cliques are of size $k+1$. So, in a k-tree G with $n = k + 1 + r$ nodes, we have, $|\mathcal{C}(G)| = 1 + r$ and $\omega = k + 1$, which implies $n - |\mathcal{C}(G)| = \omega - 1$. Thus, $\left\lceil \frac{n-|\mathcal{C}(G)|}{2} \right\rceil = \left\lceil \frac{\omega-1}{2} \right\rceil = \left\lfloor \frac{\omega}{2} \right\rfloor$.

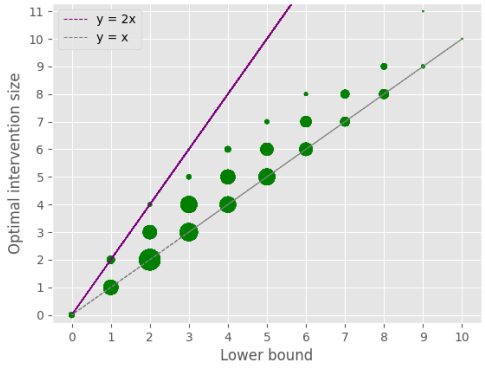
In contrast to the above two families, *block graphs* are an example family of chordal graphs where our lower bound can be significantly better. Construction 1 and 2 presented in Section 3.2 of the main paper are examples of block graphs, and as discussed there, our lower bound can be $\Theta(k)$ times the previous best universal lower bound for block graphs, where k can be as large as $\Theta(n)$ (where n is the number of nodes in the graph).



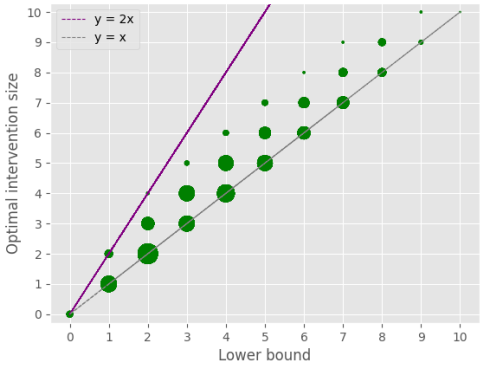
(i) Run 1



(ii) Run 2



(iii) Run 3



(iv) Run 4

Figure 5: Experiment 1 Runs with Varying Seeds

D DETAILS OF EXPERIMENTAL SETUP

Experiment 1 For this experiment, we generate 1000 graphs from Erdős-Rényi graph model $G(n, p)$: for each of these graphs, the number of nodes n is a random integer in $[5, 25]$ and the connection probability p is a random value in $[0.1, 0.3)$. Each of these graphs is then converted to a DAG without v-structures, using the following procedure. First, the edges of G are oriented according to a topological ordering σ which is a random permutation of the nodes of G: this converts G into a DAG D (possibly with v-structures). Now, the nodes of D are processed in a *reverse* order according to σ (i.e., nodes coming later in σ are processed first) and whenever we find two non-adjacent parents, a and b of the current node u being processed, we add an edge $a \to b$ in D if $\sigma(a) < \sigma(b)$, and $b \to a$ in D if $\sigma(b) < \sigma(a)$. Since nodes are processed in an order that is a reversal of σ, this procedure ensures that the resulting DAG D has no v-structures.

In Figure 5, we provide plots from four further runs of Experiment 1. These plots use exactly the same set-up and procedure as the plot given in Figure 3 in the main paper, and differ only in the initial seed provided to the underlying pseudo-random number generator. These seeds are used for generation of random graphs as well as for generating n and p. To avoid any post selection bias, the seeds for these plots were formed using the decimal expansion of π after skipping first 1015 digits[2] in the decimal expansion, and then taking the next 10 digits as the first seed, 10 consecutive digits after that as the second seed, and so on. Our interpretation and inferences from these further runs remain the same as that reported in the main paper for the run underlying Figure 3.

[2]The number "1015" corresponds to the submission deadline date for the conference, after removing the year part from the ISO standard date format.

Experiment 2 For this experiment, we generate 1000 random DAGs without v-structures for each size in $\{10, 20, 30, 40, 50, 60\}$. We now describe the procedure for generating a DAG D (without v-structures) with n nodes, other than n, this procedure takes two more inputs, *min_clique_size* and *max_clique_size*. If *min_clique_size* $= X$ and *max_clique_size* $= Y$, we try to keep the size of all cliques of D in $[X, Y]$. First, we initialize a DAG D with nodes $0, \ldots, n-1$, and no edges. We take $\sigma = (0, \ldots, n-1)$ to be a perfect elimination ordering of D. We then process the nodes of D in reverse order of σ. When node u is being processed, we first compute the number of parents that u already has in D. Now we compute lower and upper bounds $\ell_1 \geq 0$ and $\ell_2 \geq 0$ on the number of parents that could be added to the set of parents of u while still keeping the total number of parents below $Y - 1$, and at least $X - 1$. (Note that $\ell_1, \ell_2 \leq |\{0, \ldots, u-1\} \setminus \mathrm{pa}_D(u)|$, since the latter is the number of currently available vertices that could be added to the parent set of u). We now choose an integer ℓ uniformly at random from $[\ell_1, \ell_2]$: this will be the number of new parents to be added to the set of parents of u. Note that it may happen that $\ell = 0$, for example when u already has $Y - 1$ or more parents, so that $\ell_2 = 0$. Next, we sample a set Z of ℓ nodes (without replacement) from $(\{0, \ldots, u-1\} \setminus \mathrm{pa}_D(u))$, and add the edges $z \to u$ to D for each $z \in Z$. Further, for any two non-adjacent parents of u, we add an edge $a \to b$ to D if $\sigma(a) < \sigma(b)$, and $b \to a$ if $\sigma(b) < \sigma(a)$. This makes sure that there are no v-structures in D. Note that, as described in the main paper, the procedure used here only tries to keep the maximum clique size bounded above by Y, but it can overshoot and produce a graph with a clique of size larger than Y as well. In our experiments, we take *min_clique_size* $= 2$, *max_clique_size* $= 4$.

In Figure 6, we provide plots from four further runs of Experiment 2. These plots use exactly the same set-up and procedure as the plot given in Figure 4 in the main paper, and differ only in the initial seed provided to the underlying pseudo-random number generator. Again, to avoid post-selection bias, we use seeds given by the procedure given for Experiment 1 above. Our interpretation and inferences from these further runs remain the same as that reported in the main paper for the run underlying Figure 4.

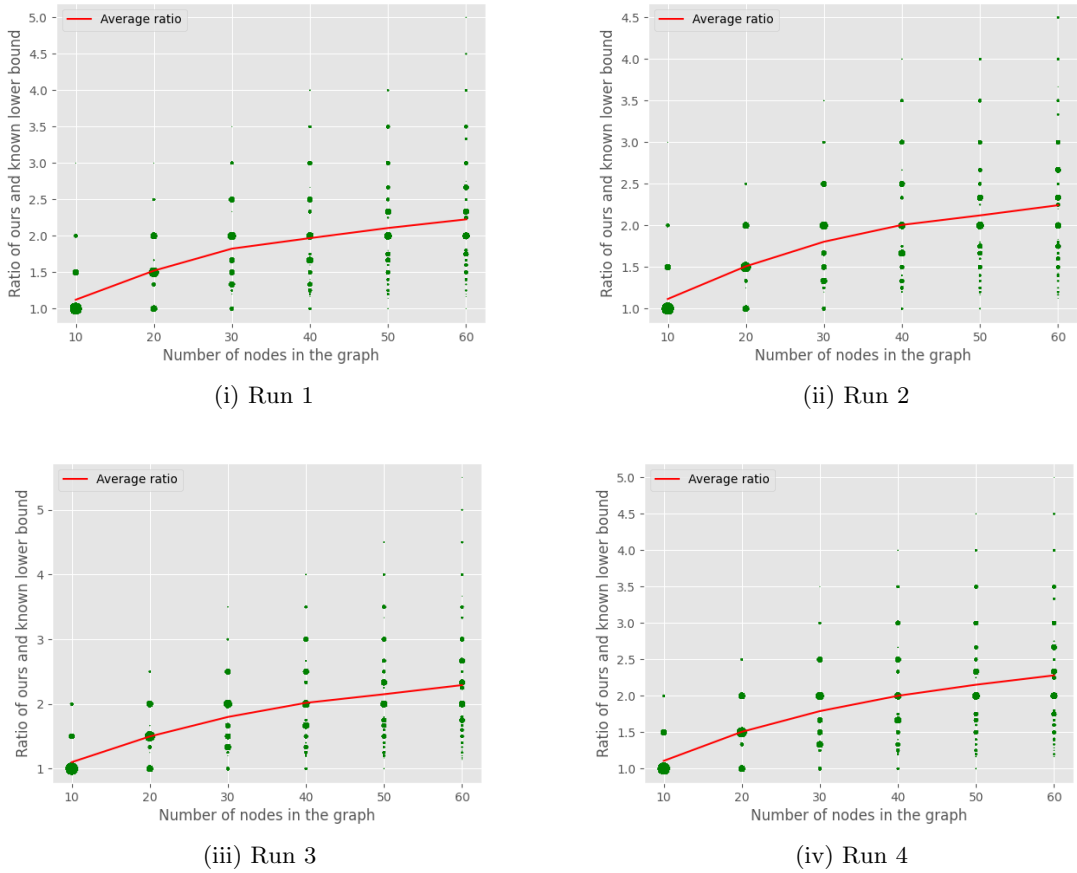
E PROOF OF REMARK B.1

In this section, we supply the proof of Remark B.1 for completeness.

Proof of Remark B.1. By Theorem 10(iv) of Hauser and Bühlmann (2012), it follows that $\mathcal{E}_{\mathcal{I}}(D)$ must have the same skeleton and the same v-structures as D, and must also have all its directed edges directed in the same direction as in D. This proves that $\mathcal{E}_{\mathcal{I}}(D)$ satisfies all the conditions of Theorem 2.1. To complete the proof, we now show that it is the only graph showing all the conditions of the theorem.

For if not, then let let G and H be two different graphs satisfying all the conditions of Theorem 2.1. Thus, G and H have the same skeleton as D, all their directed edges are in the same direction as in D, and further, all v-structures of D are directed in both G and H. If $G \neq H$, the set E' of edges that are directed in G but not in H is therefore non-empty (possibly after interchanging the labels G and H). Fix a topological ordering σ of D, and let $a \to b \in E'$ be such that $\sigma(b) \leq \sigma(b')$ for all $a' \to b' \in E'$. Since $a \to b$ in G, $a \to b$ must be strongly $\mathcal{I}$-protected in G. Now, there cannot exist $J \in \mathcal{I}$ such that $|J \cap \{a, b\}| = 1$, since in that case $a \to b$ would be directed in H as well (by item 3 of Theorem 2.1). Thus, at least one of the four graphs in Figure 1 must appear as an induced subgraph of G, with $a \to b$ appearing in that induced subgraph in the cofiguration indicated in the figure. If subgraph (i) appears as an induced subgraph of G, then we must have $c \to a$ in H since $\sigma(a) < \sigma(b)$, but this means that $c \to a - b$ would be an induced subgraph of H, contradicting item 2 of Theorem 2.1. Similarly, $a \to b$ cannot be in the configuration indicated in subgraph (ii), since any v-structure in G is directed in H, so that $a \to b$ would be directed in H as well. If subgraph (iii) appears as an induced subgraph of G, $a \to c$ must be directed in H, as $\sigma(c) < \sigma(b)$, but this would mean that H contains a directed cycle a, c, b, a (since $a - b$ is undirected in H), and this contradicts that H is a chain graph.[3] If subgraph (iv) appears as an induced subgraph of G, then $c_1 \to b \leftarrow c_2$ appears in H as well since any v-structure of G must also be directed in H. Further, at least one of the following four configurations must appear in H: (a) $a \to c_1$ (b) $a \to c_2$ (c) $a - c_1$ (d) $a - c_2$ (for if not, then $c_1 \to a \leftarrow c_2$ would be a v-structure in H that is not directed in G, contradicting that all v-structures of D are directed in both G and H). However, if any of the four configuration appears in H, we get a directed cycle in H (since $a - b$ is undirected in H), which contradicts the fact that H is a chain graph.

[3]Recall that a directed cycle in a general graph is a cycle in which all directed edges point in the same direction, and in which at least one edge is directed. The formal definition is given in Section 2.



(i) Run 1

(ii) Run 2

(iii) Run 3

(iv) Run 4

Figure 6: Experiment 2 Runs with Varying Seeds

We conclude therefore that E' must in fact be empty and hence $G = H$. Thus, given a DAG D, the unique graph satisfying all conditions of Theorem 2.1 is the $\mathcal{I}$-essential graph $\mathcal{E}_{\mathcal{I}}(D)$ of D. □

F PROOF OF COROLLARY B.2

In this section, we supply the proof of the folklore Corollary B.2, for completeness.

Proof of Corollary B.2. Let H be the graph with the same skeleton as D in which exactly the edges satisfying one of the two conditions of the corollary are directed. We prove that H satisfies all the conditions of Theorem 2.1, and must therefore be the same as $\mathcal{E}_{\mathcal{I}}(D)$ (see also Remark B.1). This will complete the proof of the corollary.

Recall that by construction, any edge of H is directed if and only if

1. the endpoints of the edge are in different chain components of $\mathcal{E}(D)$, so that it is already directed in $\mathcal{E}(D)$, or
2. the endpoints of the edge lie in the same chain component S of $\mathcal{E}(D)$, and the edge is directed in $\mathcal{E}_{\mathcal{I}_S}(D_S)$.

In particular, item 1 implies that any edge that is directed in $\mathcal{E}(D)$ is also directed in H (since all directed edges of a chain graph have their endpoints in different chain components).

We now verify that H satisfies all the conditions of Theorem 2.1. By construction, H has the same skeleton as D, and all its directed edges are directed in the same direction as D. Further, all the v-structures of D are directed in H, since these are directed in $\mathcal{E}(D)$.

Any directed cycle C in H would be a directed cycle either in $\mathcal{E}(D)$ (in case C includes vertices from at least two different chain components of $\mathcal{E}(D)$), or in $\mathcal{E}_{\mathcal{I}_S}(D_S)$ for some chain component S of $\mathcal{E}(D)$ (in case C is contained within a single chain component S of $\mathcal{E}(D)$). Since both $\mathcal{E}(D)$ and $\mathcal{E}_{\mathcal{I}_S}(D)$ are chain graphs (from Theorem 2.1), they do not have any directed cycles. It therefore follows that H cannot have a directed cycle either, and hence is a chain graph. Further, the chain components of H are induced subgraphs of the chain components of $\mathcal{E}(D)$. Since the chain components of $\mathcal{E}(D)$ are chordal (again from Theorem 2.1), it follows that the chain components of H are also chordal. Thus, H satisfies item 1 of Theorem 2.1.

Suppose now that, if possible, H has an induced subgraph of the form $a \to b - c$. Thus, the edge $b - c$ must be undirected in $\mathcal{E}(D)$ as well, so that b and c are in the same chain component S of $\mathcal{E}(D)$. If a is also in S, then $a \to b - c$ would be an induced sub-graph of the interventional essential graph $\mathcal{E}_{\mathcal{I}_S}(D_S)$, which would contradict item 2 of Theorem 2.1. Similarly, if a is not in S, then $a \to b$ would be directed in $\mathcal{E}(D)$, so that $a \to b - c$ would be an induced sub-graph of the essential graph $\mathcal{E}(D) = \mathcal{E}_{\{\varnothing\}}(D)$, again contradicting item 2 of Theorem 2.1. We conclude that an induced subgraph of the form $a \to b - c$ cannot occur in H. Thus, H satisfies item 2 of Theorem 2.1.

To verify item 3, consider any two adjacent vertices a and b in H such that $|I \cap \{a, b\}| = 1$ for some $I \in \mathcal{I}$. If a and b are in different chain components of $\mathcal{E}(D)$, then the edge between them is directed in $\mathcal{E}(D)$ and hence also in H. On the other hand, if a and b are in the same chain component S of $\mathcal{E}(D)$, then we have $|(I \cap S) \cap \{a, b\}| = |I \cap \{a, b\}| = 1$ for $I \cap S \in \mathcal{I}_S$, so that the edge between a and b is directed in $\mathcal{E}_{\mathcal{I}_S}(D_S)$ (by item 3 of Theorem 2.1) and hence also in H. It thus follows that H satisfies item 3 of Theorem 2.1.

Finally, we show that any directed edge in H is $\mathcal{I}$-strongly protected. Consider first a directed edge $a \to b$ in H where a and b belong to the same chain component S of $\mathcal{E}(D)$. Then, since $a \to b$ is directed also in $\mathcal{E}_{\mathcal{I}_S}(D_S)$, it must be $\mathcal{I}_S$-strongly protected in $\mathcal{E}_{\mathcal{I}_S}(D_S)$. It follows directly from the definition of interventional strong protection and the construction of H then that $a \to b$ is $\mathcal{I}$-strongly protected in H (since any of the configurations of Figure 1, if present as an induced subgraph of $\mathcal{E}_{\mathcal{I}_S}(D_S)$, is also present as an induced subgraph in H).

Consider now a directed edge $a \to b$ in H when a and b are in different chain components of $\mathcal{E}(D)$. Then $a \to b$ must be directed, and hence also $\{\varnothing\}$-strongly protected, in $\mathcal{E}(D)$. Now, if $a \to b$ appears as part of an induced subgraph of $\mathcal{E}(D)$ of the forms (i), (ii) or (iii) of Figure 1, then the same configurarion also appears as an induced subgraph of H (since all directed edges of $\mathcal{E}(D)$ are directed in H), so that $a \to b$ is also $\mathcal{I}$-strongly protected in H. Suppose then that $a \to b$ appears as part of an induced subgraph of the form (iv) of Figure 1. Then, the vertices a, c_1 and c_2 appearing in the configuration must be in the same chain component S of $\mathcal{E}(D)$ (since they are in a connected component of undirected edges). It follows that the configurations $c_1 \to a - c_2$ and $c_2 \to a - c_1$ cannot appear in H. For, if they did, then they would also appear in the $\mathcal{I}_S$ essential graph $\mathcal{E}_{\mathcal{I}_S}(D_S)$, thereby contradicting item 2 of Theorem 2.1 (when applied to the interventional essential graph $\mathcal{E}_{\mathcal{I}_S}(D_S)$). The configuration $c_1 \to a \leftarrow c_2$ also cannot occur in H, since otherwise, the v-structure $c_1 \to a \leftarrow c_2$ of D could not have remained undirected in $\mathcal{E}(D)$. It follows that at least one of the following three configurations must appear in H: $a \to c_1$, $a \to c_2$ or $c_1 - a - c_2$. In the last case, $a \to b$ is $\mathcal{I}$-strongly protected in H as configuration (iv) of Figure 1 appears as an induced subgraph of H (exactly as it does in $\mathcal{E}(D)$)). In the first two cases, $a \to b$ is $\mathcal{I}$-strongly protected in H as configurations (iii) of Figure 1 appears as an induced subgraph of H (with the role of the vertex c in that configuration played by either c_1 or c_2, as the edges $c_1 \to b$ and $c_2 \to b$ are both directed in H, since they are directed in $\mathcal{E}(D)$). Thus, we see that every directed edge in H is $\mathcal{I}$-strongly protected in H, and hence H satisfies item 4 of Theorem 2.1 as well.

It follows from Theorem 2.1 therefore that $H = \mathcal{E}_{\mathcal{I}}(D)$. As discussed at the beginning of the proof, this completes the proof of the corollary. □