Proceedings of Machine Learning Research 323:1318–1344, 2026 5th Conference on Causal Learning and Reasoning

Coarsening Causal DAG Models

Francisco Madaleno
Department of Technology, Management and Economics, Danish Technical University

Pratik Misra
Department of Mathematics and Statistics, Binghamton University, State University of New York

Alex Markham ALEX.MARKHAM@CAUSAL.DEV
Department of Mathematical Sciences, University of Copenhagen

Editors: Bijan Mazaheri and Niels Richard Hansen

Abstract

Directed acyclic graphical (DAG) models are a powerful tool for representing causal relationships among jointly distributed random variables, especially concerning data from across different experimental settings. However, it is not always practical or desirable to estimate a causal model at the granularity of given features in a particular dataset. There is a growing body of research on *causal abstraction* to address such problems. We contribute to this line of research by (i) providing novel graphical identifiability results for practically-relevant interventional settings, (ii) proposing an efficient, provably consistent algorithm for directly learning abstract causal graphs from interventional data with unknown intervention targets, and (iii) uncovering theoretical insights about the lattice structure of the underlying search space, with connections to the field of causal discovery more generally. As proof of concept, we apply our algorithm on synthetic and real datasets with known ground truths, including measurements from a controlled physical system with interacting light intensity and polarization.

Keywords: causal discovery; causal abstraction; cluster DAGs; interventional data; partition refinement lattice.

1. Introduction

Discovering causal relationships is one of the fundamental goals of scientific research. Nevertheless, the causal relationships of interest are not always between features of a given dataset. For example, in neuroscience, one may have data describing the interactions of groups of (or even individual) neurons and wish to learn from this data a causal model over cognitive or behavioral states (Grosse-Wentrup et al., 2024). Similarly, in genomics, one may observe gene expression profiles for individual cells, while the goal is to learn a causal model over clusters of cells with similar regulatory states (Squires et al., 2022).

This is a very general and difficult problem, so we restrict our interest here to particular cases where there is data from some assumed low-level or fine-grained causal DAG model, and we offer a formalization of what it means for a DAG with fewer nodes to be a high-level or coarse-grained abstraction of this underlying causal DAG model. Our approach is based on the intuition that variables with similar causes and similar effects (graphically, similar ancestors and descendants) are in some sense redundant and can be abstracted away by clustering them together while retaining only the salient causal relationships.

There is a growing number of similarly motivated works on consistent transformations of causal models (Rubenstein et al., 2017) and causal abstractions (Beckers and Halpern, 2019; Beckers et al.,

2020; Beckers, 2021; Otsuka and Saigo, 2022) that lay a groundwork for formally describing such problems. There is also some work on learning abstractions (Massidda et al., 2024), on directly learning a reduced DAG according to its marginal independences (Deligeorgaki et al., 2023), on using partitions to simplify learning large graphs (Gu and Zhou, 2020), and on clustering nodes in a DAG while preserving the identifiability of specified causal estimands (Tikka et al., 2023). Other approaches focus on causal layerings (Feigenbaum et al., 2024), on learning abstractions in the unsupervised setting (Zhu et al., 2024), and on targeted causal reduction for extracting compact high-level explanations of simulator phenomena and of reinforcement learning policy behavior (Kekić et al., 2024; Kekić et al., 2025). Similarly, for complex decision settings, Dyer et al. (2024) develop interventionally consistent surrogates for expensive simulators, while Dyer et al. (2025) accelerate decision making in causal bandit problems by enabling transfer across levels of granularity. A broader discussion of related literature is provided in Appendix A.

In contrast to existing work, our main contribution is to develop a general framework for understanding causal abstractions graphically, which we call coarsening. In particular, we study the space of all coarsenings, proving that it has "nice" mathematical structure. This allows us: (i) to provide novel graphical identifiability results, in particular for practically-relevant interventional coarsenings, also suggesting new directions for future work on abstractions; (ii) to propose an efficient, provably consistent algorithm for learning interventional coarsenings directly from data—a coarsening is a simpler mathematical object than a fine-grained DAG model, and our algorithm leverages this; and (iii) to establish the fundamental mathematical properties of the coarsening space, connecting it to the search space of other causal discovery algorithms.

The paper proceeds as follows: Section 2 formalizes the general problem of graphical causal abstraction, establishes the lattice structure of the coarsening space, establishes identifiability of interventional coarsenings, and draws connections to causal discovery more broadly. Section 3 develops a flexible learning algorithm for interventional coarsenings, with formal consistency and complexity guarantees. All proofs are deferred to Appendix B. For empirical evaluation, Section 4 presents an open-source implementation of the algorithm, evaluating it on synthetic data to empirically test the theoretical guarantees and on real data, compared to baselines from the literature, to demonstrate practical relevance. Section 5 concludes with a discussion of our approach, its limitations, and future directions.

2. Coarsening Causal DAGs

The following section relies on basic ideas from graphical models (Lauritzen, 1996), order theory (Davey and Priestley, 2002), and combinatorics (Stanley, 2011).

For a DAG $G = (V, E)$, in addition to the familiar terminology of parents $\mathrm{pa}_G(v)$ and children $\mathrm{ch}_G(v)$, we denote *ancestors* as $\mathrm{an}_G(v) := \{w \in V \mid \text{there exists a directed path from } w \text{ to } v\}$ and analogously-defined *descendants* as $\mathrm{de}_G(v)$—note that every node is its own ancestor and descendant. These all extend naturally to sets: the parents of a set of nodes is the union of the parent sets of the nodes, and so on. A distribution is *Markov* to a DAG if the distribution's conditional independence statements are implied by the DAG's d-separation statements—the set of all such distributions for a given DAG is denoted $\mathcal{M}(G)$.

A *poset* (partially ordered set) is a set with a reflexive, transitive, and antisymmetric order relation $\preceq$; a *lattice* is a poset where every pair of elements has a unique meet (greatest lower bound) and join (least upper bound).

A *partition* of V is a collection $\Pi = \{\pi_1, \ldots, \pi_k\}$ of disjoint non-empty sets, called *parts*, whose union is V. A partition Π *refines* another Π' (written $\Pi \preceq \Pi'$) if every part of Π is contained in some part of Π'—equivalently, Π' is said to *coarsen* Π. The collection of all partitions of V, ordered by refinement, forms the *partition refinement lattice*, where the meet of two partitions is their finest common refinement and the join is their coarsest common coarsening.

2.1. General Coarsenings

We begin with what we argue are necessary but not sufficient graphical conditions for a "reasonable" causal abstraction:

Definition 1 *Given a DAG $G = (V, E)$, a* coarsening *is a DAG $G' = (V', E')$ for which there exists a surjection $\chi : V \to V'$ such that*

$$E' = \{\chi(v) \to \chi(w) \mid v \to w \in E, \chi(v) \neq \chi(w)\}.$$

For example, this precludes coarsenings that reverse causal edges as well as coarsenings that group $\{u, w\}$ together apart from $\{v\}$ in graph structures like $u \to v \to w$. More generally, it ensures model containment:

Lemma 2 *Let G be a DAG and G' be one of its coarsenings. Every distribution Markov to G is also Markov to G', that is, $\mathcal{M}(G) \subseteq \mathcal{M}(G')$.*

This containment induces a natural partial order over coarsenings, forming a lattice:

Theorem 3 *For any DAG $G = ([d], E)$, the poset of its coarsenings is a lattice, more specifically, a sublattice of the partition refinement lattice of $[d]$.*

To make this more concrete, consider the following example:

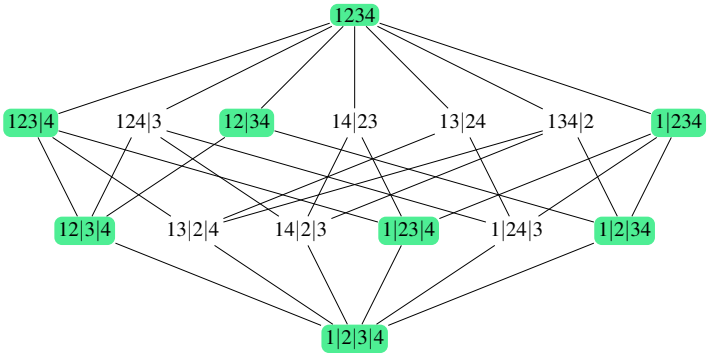


Figure 1: The partition refinement lattice on 4 nodes, with an example coarsening sublattice highlighted.

Example 4 *Let $G = (V, E)$ be the DAG $1 \to 2 \to 3 \to 4$ and consider the partition lattice of $[4]$ shown in Figure 1. The lattice is organized by the number of parts: level i (for $i \in \{0, 1, 2, 3\}$) contains all partitions of $[d]$ (here, $d = 4$) into $i + 1$ parts, counted by the* Stirling numbers of the second kind *$S(d, i + 1)$. These satisfy the recurrence $S(n, k) = k \cdot S(n - 1, k) + S(n - 1, k - 1)$. For $d = 4$, this gives: $S(4, 1) = 1$, $S(4, 2) = 7$, $S(4, 3) = 6$, and $S(4, 4) = 1$. The total number of partitions is the* Bell number *$B_n = \sum_{k=1}^{n} S(n, k)$.*

Not every partition corresponds to a valid coarsening of G. Consider the partition $1|2|34$, represented by coarse nodes $V' = \{1, 2, \{3, 4\}\}$[1]*. This is a valid coarsening: the surjection χ*

1. To keep notation light, we conflate singleton sets and their elements, e.g., writing 1 instead of $\{1\}$, when it is clear from context.

Algorithm 1: Recursive Partition Refinement for DAG Learning

Input: Coarsening $G = (\Pi, E)$
Output: Refinement G^*

1 **Function** `RePaRe` (G) **:**
  // Attempt split
2 $(\pi^*, \{\pi_a, \pi_b\}) \coloneqq$ `Refine` (Π)
3 **if** $\pi^* = \emptyset$ **then return** G
  // Define new node set
4 $\Pi' \coloneqq (\Pi \setminus \{\pi^*\}) \cup \{\pi_a, \pi_b\}$
  // Define new edge set
5 $E' \coloneqq \{u \to v \in E \mid u \neq \pi^* \text{ and } v \neq \pi^*\}$
  $\cup\ \{u \to v \mid \{u, v\} = \{\pi_a, \pi_b\}, \texttt{IsEdge}(u, v)\}$
  $\cup\ \{\pi \to \pi_{\text{new}} \mid \pi \in \text{pa}_G(\pi^*), \pi_{\text{new}} \in \{\pi_a, \pi_b\}, \texttt{IsEdge}(\pi, \pi_{\text{new}})\}$
  $\cup\ \{\pi_{\text{new}} \to \pi \mid \pi \in \text{ch}_G(\pi^*), \pi_{\text{new}} \in \{\pi_a, \pi_b\}, \texttt{IsEdge}(\pi_{\text{new}}, \pi)\}$
  // Construct and Recurse
6 $G' \coloneqq (\Pi', E')$
7 **return** `RePaRe` (G')

defined by $\chi(1) = 1, \chi(2) = 2, \chi(3) = \chi(4) = \{3, 4\}$ *induces the coarsened DAG* G' *with* $1 \to 2 \to \{3, 4\}$. *In contrast, the partition* $1|3|24$ *(corresponding to* $V'' = \{1, 3, \{2, 4\}\}$*) does* not *correspond to a valid coarsening. Although a surjection to* V'' *exists, the edges it induces do not form a DAG:* $2 \xrightarrow{G} 3$ *requires* $\chi(2) \xrightarrow{G''} \chi(3)$ *while* $3 \xrightarrow{G} 4$ *requires* $\chi(3) \xrightarrow{G''} \chi(4) = \chi(2)$*, forming a cycle.*

The lattice structure means that the space of coarsenings is a "nice" space to work with in the sense that we can traverse it easily with something like Algorithm 1. This gives us a framework for identifying *any* valid coarsening by choosing suitable oracles for `Refine` (Line 2) and `IsEdge` (Line 5) functions.

Definition 5 *Given an underlying DAG* $\underline{G} = ([d], \underline{E})$ *and a coarsening* $G = (\Pi, E)$*, a* `Refine`-oracle *is any function that takes in* Π *and returns a part* $\pi^* \in \Pi$ *that itself is partitioned into* π_a, π_b *satisfying:*

1. acyclicity preservation*: there does not exist a directed cycle in* G *between the sets* π_a *and* π_b*, i.e., either* $\text{an}_{\underline{G}}(\pi_a) \cap \pi_b = \emptyset$ *or* $\text{an}_{\underline{G}}(\pi_b) \cap \pi_a = \emptyset$.

If no such π^* *exists, the oracle returns* $(\emptyset, \emptyset)$*. Furthermore, an* `IsEdge`-oracle *is any function taking in parts* $u, v \in \Pi'$ *(Line 4) and returning* `True` *if and only if* u, v *satisfy:*

2. parent consistency*:* $\text{pa}_{\underline{G}}(v) \cap u \neq \emptyset$.

Theorem 6 (Completeness of `RePaRe`) *Let* $\underline{G} = (\underline{V}, \underline{E})$ *be a DAG and* G^* *be any valid coarsening of* $\underline{G}$*. There exist* `Refine`*- and* `IsEdge`*-oracles (Definition 5) such that Algorithm 1, with the trivial coarsening* $\overline{G} = (\{V\}, \emptyset)$ *as input, terminates and outputs* G^*.

The proof constructs explicit oracles by leveraging the lattice structure of valid coarsenings. Whenever $\Pi^* \prec \Pi$, we can find a part $\pi \in \Pi$ that must be split in any refinement chain from Π

to Π^*. The `Refine`-oracle identifies such a part and splits it according to the structure imposed by Π^*, while the `IsEdge`-oracle determines edges in the refined coarsening by querying the true parental relations of $\underline{G}$. This greedy refinement is guaranteed to converge to Π^* because every step moves strictly closer in the lattice order.

In the next subsections, we show how this general notion of coarsening can be narrowed down into other (perhaps more causally- or domain-relevant) notions of coarsening, but first we look at an example of how the lattice is already interesting for understanding existing—and perhaps developing future—causal discovery algorithms:

Example 7 *Let G be the 4-node path $1 \to 2 \to 3 \to 4$. A topological order $\sigma = (1, 2, 3, 4)$ induces a natural refinement chain in the coarsening lattice: $1234 \succ 1|234 \succ 1|2|34 \succ 1|2|3|4$. This illustrates how causal discovery methods that estimate a topological order can be interpreted as walking down the lattice. For example, under linear non-Gaussian assumptions, DirectLiNGAM (Shimizu et al., 2011) identifies a causal ordering from observational data and thereby a compatible DAG—its search corresponds to such a walk.*

2.2. Interventional Coarsening

We now narrow down the general notion of coarsening from Definition 1, giving it a more rigorous causal grounding in interventions and demonstrating how the general framework can be adapted to answer meaningful practical questions. Across scientific domains, a natural question is: *given a collection of post-intervention samples, which of the measured features are affected similarly, and what is the salient causal structure among these clusters?*[2]

Thus, we begin by formalizing the idea of "measured features being affected similarly". Relative to a set of (possibly soft) interventions $\mathcal{I}$, denote a node's *intervened ancestors* as $\mathcal{I}\text{-an}_G(v) := \{w \in \text{an}_G(v) \mid \exists I \in \mathcal{I} \text{ such that } w \in I\}$. This naturally leads to interventional coarsening:

Definition 8 *Given a DAG $G = (V, E)$ and a set of interventions $\mathcal{I}$, define the* interventional coarsening*, denoted $G^{\mathcal{I}}$, to be the coarsening whose partition surjection χ satisfies*

$$\textit{for all } v, w \in V, \quad \chi(v) = \chi(w) \iff \mathcal{I}\text{-an}_G(v) = \mathcal{I}\text{-an}_G(w)\,.$$

Framed in this way, it is clear that $G^{\mathcal{I}}$ is a valid coarsening and thus one that is identifiable according to the completeness result of Theorem 6 by specifying appropriate oracles. In this case, the oracles need to be able to work with distributions rather than merely graphs as well as to be able to determine which nodes are descendants of which interventions—formally:

Assumption 9 *Given a ground-truth DAG $G = (V, E)$ and intervention set $\mathcal{I}$ (which includes $\emptyset$, representing the observational setting), we make the following assumptions about the associated distributions:*

2. This question mirrors common practice in high-dimensional interventional settings, where one summarizes responses by clusters and studies relations between them. In genomics, CWGCNA inserts a mediation-based causal inference step into WGCNA to estimate causal relationships among phenotypes, gene modules, and module features (Liu, 2024). In neuroscience, TMS-EEG analyses track how stimulation-evoked activity propagates through brain networks (Xiao et al., 2025). In simulation modeling, Targeted Causal Reduction learns a concise set of causal factors explaining a specified target phenomenon from interventional simulation data (Kekić et al., 2024), and in robotics SCALE uses simulated interventions to discover causally relevant context variables associated with distinct manipulation modes (Lee et al., 2023).

1. coarse Markov: *the distributions factorize according to $G^{\mathcal{I}}$, i.e., for all $I \in \mathcal{I}$:*

$$f^I(X_V) = \prod_{u \in \chi(V) \setminus \chi(I)} f^{\emptyset}\big(X_{\chi^{-1}(u)} \mid X_{\chi^{-1}(\mathrm{pa}_{G^{\mathcal{I}}}(u))}\big) \prod_{u \in \chi(I)} f^I\big(X_{\chi^{-1}(u)} \mid X_{\chi^{-1}(\mathrm{pa}_{G^{\mathcal{I}}}(u))}\big).$$

2. coarse faithfulness: *the only conditional independences among coarse nodes $\chi(V)$ are those implied by the Markov factorization above.*
3. interventional soundness: *interventions induce changes in marginal distributions compared to the observational distribution, that is:*

$$\textit{for all } I \in \mathcal{I} \setminus \emptyset \textit{ and } v \in V, \quad v \in \mathrm{de}_G(I) \implies f^I(X_v) \neq f^{\emptyset}(X_v).$$

Note that, when $\chi(V) \neq V$, these are strictly weaker than the usual Markov and faithfulness assumptions—they allow violations at the V-level as long as those do not induce violations at the $\chi(V)$-level. Making use of these assumptions, we can specify oracles that plug into Algorithm 1 to enable us to prove identifiability of $G^{\mathcal{I}}$:

Theorem 10 *Under Assumption 9, the interventional coarsening $G^{\mathcal{I}}$ is identifiable.*

As we will see in Section 3, replacing the oracles used to prove Theorem 10 with valid statistical hypothesis tests allows us to turn Algorithm 1 into a consistent estimator of $G^{\mathcal{I}}$. However, we first spend the rest of this section building up theoretical intuitions with the coarsening framework.

Example 11 *Let G be the DAG on $V = \{1, 2, 3, 4\}$ with edges $1 \to 3$, $2 \to 3$, and $3 \to 4$. Consider the intervention set $\mathcal{I} = \{\emptyset, \{1\}, \{2\}\}$, where $\emptyset$ denotes the observational regime and $\{1\}, \{2\}$ are single-target interventions.*

For each node v, $\mathcal{I}\text{-an}_G(v)$ collects the intervened ancestors of v. This yields

$$\mathcal{I}\text{-an}_G(1) = \{1\}, \quad \mathcal{I}\text{-an}_G(2) = \{2\}, \quad \mathcal{I}\text{-an}_G(3) = \{1, 2\}, \quad \mathcal{I}\text{-an}_G(4) = \{1, 2\}$$

Thus the interventional coarsening groups nodes that are affected by the same intervention targets, yielding the partition $\Pi^{\mathcal{I}} = \big\{\{1\}, \{2\}, \{3, 4\}\big\}$. The induced coarsened DAG has two edges $\{1\} \to \{3, 4\}$ and $\{2\} \to \{3, 4\}$.

This example highlights how $\mathcal{I}$ determines the resolution *of the learnable coarse structure: descendants that are indistinguishable with respect to the available interventions are merged. This intervention-driven notion of abstraction is closely aligned with the role of distributional changes across environments in causal discovery with unknown targets, as in UT-IGSP (Squires et al., 2020) and GnIES (Gamella et al., 2022).*

2.3. Other Coarsenings

In contrast to the general and interventional coarsenings above, we now present two specific observational coarsenings, one related to Markov equivalence classes (Andersson et al., 1997) and one related to unconditional equivalence classes (Deligeorgaki et al., 2023).

2.3.1. ESSENTIAL COARSENINGS

Classical identifiability for causal discovery in the observational setting extends only to Markov equivalence classes (MECs) of DAGs. Andersson et al. (1997) represent these equivalence classes with *essential graphs* (CPDAGs), whose nodes can be partitioned into chain components (which we denote $\mathrm{cc}_G(v)$), and whose edges can be partitioned into directed and undirected sets. The directed edges represent all the direct causal relations that are identifiable without interventions or further assumptions, and the undirected edges encode uncertainty about the causal direction; an edge is directed if and only if it is between two different chain components. Our general notion of coarsening (Definition 1) can be narrowed down to capture much of the same information:

Definition 12 *Given a DAG G, its* essential coarsening *is the coarsening with partition surjection χ defined by:*

$$\text{for all } v, w \in V^G, \quad \chi(v) = \chi(w) \iff \mathrm{cc}_G(v) = \mathrm{cc}_G(w)\,.$$

Example 13 *Let G be the DAG on $V = \{1,2,3,4,5\}$ with edges $1 \to 2$, $2 \to 3$, $4 \to 3$, and $3 \to 5$. The MEC of G is represented by a CPDAG whose edges are $1 - 2$ (undirected), $2 \to 3 \leftarrow 4$ (directed) and $3 \to 5$ (directed), since the directed edges incident to 3 are strongly protected while the one between 1 and 2 is not (Andersson et al., 1997).*

The chain components of the CPDAG are $\{1,2\}, \{3\}, \{4\}, \{5\}$, and the essential coarsening is given by the partition $\Pi^{\text{ess}} = \big\{\{1,2\}, \{3\}, \{4\}, \{5\}\big\}$ with induced coarsened edges $\{1,2\} \to \{3\}$, $\{3\} \to \{5\}$, and $\{4\} \to \{3\}$. Hence, the essential coarsening preserves the chain components and the partial order between but not the undirected structure within each chain component.

From this perspective, a purely observational algorithm analogous to PC (Kalisch and Bühlmann, 2007) or GES (Chickering, 2002) could use conditional independences to move down the coarsening lattice until arriving at the essential coarsening, at which point no further refinement is justified by observational data. Including background knowledge (Bang and Didelez, 2023) could substantially reduce the search space within the lattice.

2.3.2. MARGINAL COARSENINGS

Unconditional equivalence classes (UECs) provide a coarser notion of equivalence than MECs, capturing less causal information—still identifying all v-structures and possible source nodes—but at a much cheaper cost: a UEC can be learned via quadratically-many pairwise *marginal* independence tests rather than exponentially-many conditional independence tests. UECs are also relevant for latent causal DAGs (Markham and Grosse-Wentrup, 2020; Markham et al., 2023; Jiang and Aragam, 2023) as well as causal clustering (Markham et al., 2022a). Markham et al. (2022b, Theorem 1 (2)) provide a characterization of UECs that relies on grouping nodes together when they share the same set of maximal ancestors (denoted $\mathrm{ma}_G(v)$). Using this, our general notion of coarsening (Definition 1) can be narrowed down to encode exactly the information captured by a UEC:

Definition 14 *Given a DAG G, its* marginal coarsening *is the coarsening with partition surjection χ defined by:*

$$\text{for all } v, w \in V^G, \quad \chi(v) = \chi(w) \iff \mathrm{ma}_G(v) = \mathrm{ma}_G(w)\,.$$

Algorithm 2: Precompute intervention descendants

Input: Data $\mathcal{D} = \{X^{(I)}\}_{I \in \mathcal{I}}$, Threshold α
Output: Intervention descendant indicator matrix $M \in \{0,1\}^{|V| \times |\mathcal{I}|}$

Function `RefineAux` $(\mathcal{D}, \alpha)$ **:**
  for $(v, I) \in V \times \mathcal{I}$ ***in parallel*** **do**
    $p_{v,I} \coloneqq \text{Welch}_t\left(X_v^{(I)}, X_v^{(\emptyset)}\right)$
    $M_{v,I} \coloneqq \mathbb{1}[p_{v,I} < \alpha]$
  end
  return M

Algorithm 3: Refine via intervention descendant patterns

Input: Partition Π, Descendant patterns M
Output: Pair $(\pi^*, \{\pi_a, \pi_b\})$ or $(\emptyset, \emptyset)$

Function `RefineTest` (Π, M) **:**
  for ***each*** $\pi \in \Pi$ ***with*** $|\pi| > 1$ **do**
    Pick $u \in \pi$
    $\pi_a \coloneqq \{v \in \pi \mid M_{v,\cdot} = M_{u,\cdot}\}$
    $\pi_b \coloneqq \pi \setminus \pi_a$
    if $\pi_b \neq \emptyset$ **then return** $(\pi, \{\pi_a, \pi_b\})$
  end
  return $(\emptyset, \emptyset)$

Example 15 *For the DAG G from Example 13, its UEC is represented by the DAG-reduction $\{1,2\} \to \{3,5\} \leftarrow \{4\}$ (Deligeorgaki et al., 2023, Definition 5.1.2), corresponding to the partition $12|4|35$. This is indeed a marginal coarsening: nodes 3 and 5 have the same maximal ancestors $\text{ma}_G(3) = \text{ma}_G(5) = \{1,4\}$, while $\text{ma}_G(4) = \{4\}$ and $\text{ma}_G(1) = \text{ma}_G(2) = \{1\}$. Thus, the reduced DAG can be obtained by traversing the coarsening lattice downward via marginal independence tests.*

3. Learning Interventional Coarsenings

In place of the oracles used for identifiability in Theorem 10, we now present practical algorithms that turn Algorithm 1 from a theoretical tool into a method for learning interventional coarsenings from data. For simplicity, we assume Gaussianity and employ two standard statistical tests.[3]

First, `RefineAux` (Algorithm 2) precomputes an *intervention descendant indicator matrix* M using a t-test (Welch, 1947) to recover which nodes respond to each intervention. Then, `RefineTest` (Algorithm 3) greedily selects the first part containing nodes with different intervention descendant patterns and splits it. Finally, `IsEdgeTest`[4] (Algorithm 4) applies canonical correlation analy-

3. Note that the Gaussianity assumption can easily be relaxed and the test statistics replaced with something more general, like energy distance (Székely and Rizzo, 2013) or maximum mean discrepancy (Gretton et al., 2012) for the two-sample test and distance correlation (Székely et al., 2007) or the Hilbert-Schmidt independence criterion (Gretton et al., 2007) for the independence test.
4. See the proof of Theorem 10 for details about constructing the conditioning sets Z.

Algorithm 4: IsEdge via conditional CCA and Wilks' Lambda

Input: Parts $U, W \subseteq V$, Observational data X, Conditioning set $Z \subseteq V$, Threshold α
Output: `True` or `False`

Function `IsEdgeTest` (U, W, X, Z, α) **:**
  $\tilde{X}_U \coloneqq X_U - \mathbb{E}[X_U \mid X_Z]$, $\tilde{X}_W \coloneqq X_W - \mathbb{E}[X_W \mid X_Z]$ *// Regression residuals*
  $\Lambda \coloneqq \text{Wilks}_\lambda\left(\tilde{X}_U, \tilde{X}_W\right)$ *// CCA-based test statistic*
  $p \coloneqq p\text{-value}(\Lambda)$ *// Under* $H_0 : X_U \perp\!\!\!\perp X_W \mid X_Z$
  return $p < \alpha$

sis (CCA) and computes a likelihood ratio test (Hotelling, 1936; Rencher and Christensen, 2012, Chapter 11) to determine directed edges between coarsened nodes. Plugging in statistical tests in this way, the identifiability in Theorem 10 leads straightforwardly to consistency (Corollary 16) and allows us to analyze the time complexity of learning interventional coarsenings (Theorem 17).

Corollary 16 *Given Gaussian distributed X_V satisfying Assumption 9,* `RePaRe` *using* `RefineTest` *and* `IsEdgeTest` *learns the correct coarsening in the large sample limit.*

Theorem 17 *Let $d = |V|$ be the number of nodes in the underlying DAG, $e = |\mathcal{I}|$ the number of interventions, and $n = \max_{I \in \mathcal{I}} |X^{(I)}|$ the maximum sample size across datasets. Let $k = |\Pi^{\mathcal{I}}|$ be the number of parts in the true coarsening and $p = \max_{\pi \in \Pi^{\mathcal{I}}} |\pi|$ be the maximum size of any part.* `RePaRe`*, using* `RefineTest` *and* `IsEdgeTest`*, and assuming the statistical regime where $d, p \leq n$, runs in worst-case time $O\big(den + k^2 p^2 n\big)$.*

`RePaRe` decouples the dependence on underlying graph size d from dependence on coarsened graph complexity. The precomputation and refinement phases incur $O(den)$ cost (inherent to processing d nodes), but crucially, the expensive edge-finding phase depends only on the coarsened graph size k and part sizes p, costing $O(k^2 p^2 n)$. The algorithm pays for large d once (in refinement), then scales with the simpler coarsened structure. This makes `RePaRe` particularly suited for learning causal structure at "human scale"—for instance, coarsened graphs with only tens of nodes—remaining efficient regardless of the size of d.

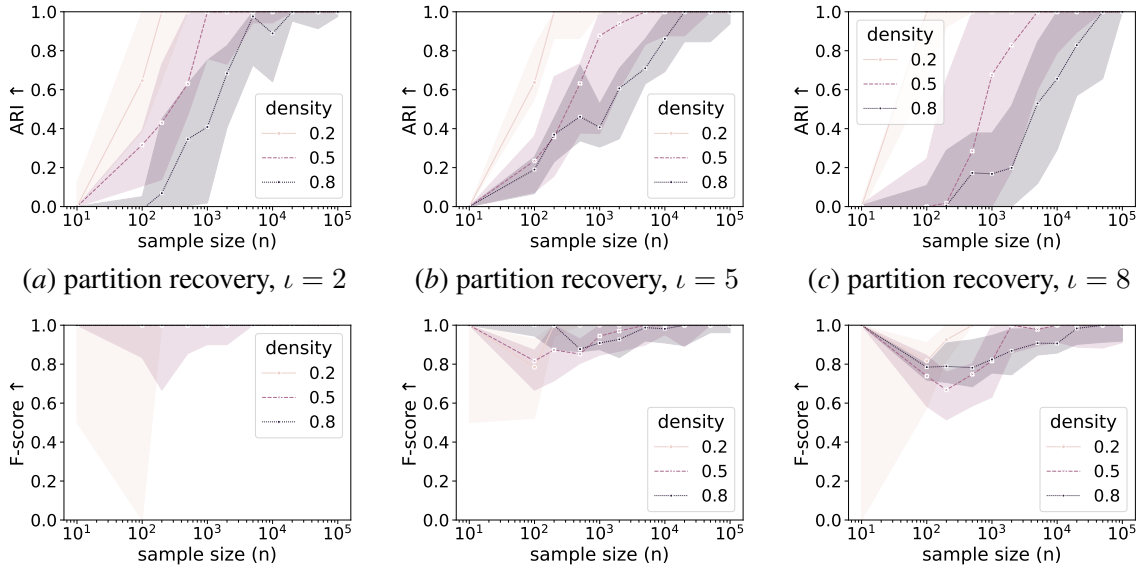
4. Experimental Results

Across synthetic and real data, we evaluate whether `RePaRe` recovers the ground-truth interventional coarsening, measuring partition recovery with the adjusted Rand index (ARI) (Hubert and Arabie, 1985, Eq (5)) and edge recovery with the F-score (Manning et al., 2008, Eq (8.6)). In all experiments, we run `RePaRe` with the Section 3 algorithms: `RefineTest` for partition refinement, and `IsEdgeTest` for edge recovery.[5]

4.1. Synthetic Data

We generate linear Gaussian SCMs on $d = 10$ nodes using `sempler` (Gamella et al., 2022) with varying densities (i.e., edge probabilities). For each SCM, we generate one observational dataset and

5. An open source implementation as well as scripts for reproducing all of the following experiments can be found at https://github.com/Alex-Markham/repare/tree/v0.2.0.



(*a*) partition recovery, $\iota = 2$ (*b*) partition recovery, $\iota = 5$ (*c*) partition recovery, $\iota = 8$

(*d*) edge recovery, $\iota = 2$ (*e*) edge recovery, $\iota = 5$ (*f*) edge recovery, $\iota = 8$

Figure 2: Evaluation on synthetic data, averaged over 10 seeds, as sample size per intervention increases, across different intervention budgets and ground-truth graph densities. **Top**: ARI for evaluating partition recover. **Bottom**: F-score for evaluating edge recovery.

$\iota \in \{2, 5, 8\}$ interventional datasets, with per-environment sample size n ranging from 10 to 10^5; this is replicated across 10 random seeds. See Appendix D.1 for full details as well as a scale-free rather than Erdős-Rényi version.

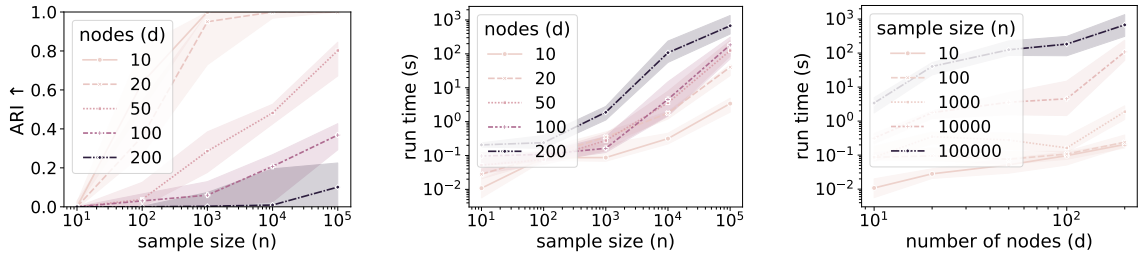
Figure 2 shows both the ARI (top row) and F-score (bottom row) reliably increasing with n across densities and intervention budgets, empirically corroborating the consistency result in Corollary 16.

For a fixed n and density, Figure 2 shows that the performance decreases with increased intervention budget. This aligns with theory: more interventions means a finer coarsening, which is harder to learn, with the full DAG being the extreme case.

Exploring this further in Figure 3 (see Appendix D.2 for more details), Figure 3(*a*) demonstrates that, though performance steadily increases with data availability, the performance drops as d increases for a fixed n—in other words, larger values of d require more data to maintain performance, with $d > 50$ requiring $n > 10\,000$. Nevertheless, the Figures 3(*b*) and 3(*c*) demonstrate good scalability in terms of run time as d and n increase. Hence, practical limitations in applying `RePaRe` stem primarily from statistical power of hypothesis testing and data scarcity rather than the lattice traversal approach itself.

4.2. Causal Chambers Light Tunnel

We apply `RePaRe` to the *light-tunnel* system from Causal Chambers (Gamella et al., 2025), learning an interventional coarsening over a subset of $d = 20$ actuator and sensor variables—see Appendix D.3 for full details. We use interventional data from RGB light intensity perturbations and



(*a*) ARI for partition recovery (*b*) Run time vs. sample size (*c*) Run time vs. num. nodes

Figure 3: Scalability on synthetic data, averaged over 10 seeds, with density 0.2 and $\iota = 5$.

polarizer-angle perturbations. We report two variants: *grouped* (pool by intervention type) and *ungrouped* (treat each intervention separately), which ablates intervention granularity.

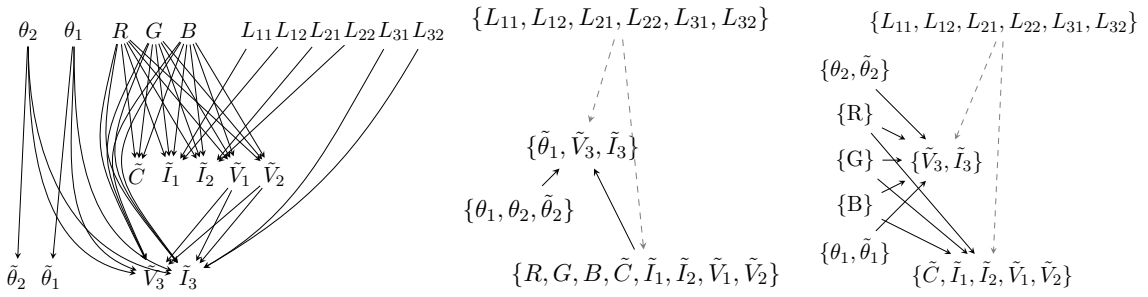
For hyperparameter selection, `RePaRe` has two significance thresholds $(\alpha_{\text{ref}}, \alpha_{\text{edge}})$ used by `RefineTest` and `IsEdgeTest`. We perform grid search over these thresholds, selecting via a maximum likelihood estimate (MLE) heuristic score applied to the learned coarsening. Importantly, this heuristic does not use the ground truth and performs comparably to optimal (given the ground truth) hyperparameter selection.

We compare against three baselines—GIES (Hauser and Bühlmann, 2012), GnIES (Gamella et al., 2022), and UT-IGSP (Squires et al., 2020)—which learn interventional equivalence classes of fine-grained DAGs over the original variables. Though the baselines solve a different task than `RePaRe` (recovering the refined equivalence class versus learning a coarsening), the results in Table 1 are nevertheless informative: `RePaRe` achieves close to perfect partition recovery, competitive edge recovery, and orders of magnitude speedups.

Method (Selection)	ARI	Precision	Recall	F-score	Run time (s)
GIES (score)	–	0.632	0.615	0.623	7.227
GnIES (score)	–	0.426	0.513	0.465	1966.916.804
UT-IGSP (score)	–	0.333	0.385	0.357	1.117
`RePaRe` (grouped, score)	0.932	1.000	0.500	0.667	0.063
`RePaRe` (grouped, optimal)	1.000	1.000	0.500	0.667	0.141
`RePaRe` (ungrouped, score)	1.000	1.000	0.800	0.889	0.753
`RePaRe` (ungrouped, optimal)	1.000	1.000	0.800	0.889	0.753

Table 1: Light-tunnel real data results, `RePaRe` compared to baselines.

Table 1 shows near-perfect to perfect partition recovery for all variants. While the RePaRe (grouped, score) variant achieves a high ARI of 0.932, the other variants achieve an ARI of 1.0, confirming that the interventions largely recover the intended abstraction level. The ungrouped variants achieve significantly higher edge recall (0.800) and F-scores (0.889) than the grouped variants, reflecting that finer intervention granularity reveals additional causal relations that are otherwise obscured in finite samples.The learned coarsenings in Figure 4 remain interpretable and capture the salient causal structure of the ground truth: color perturbations (R, G, B) are grouped separately from polarization perturbations (θ_1, θ_2), while the lenses $(L_{11}, L_{12}, L_{21}, L_{22}, L_{31}, L_{32})$ are shown



(a) Ground-truth light-tunnel DAG.

(b) Coarsening learned by `RePaRe` (score) on grouped interventions.

(c) Coarsening learned by `RePaRe` (score) on ungrouped interventions.

Figure 4: `RePaRe` on the light-tunnel data. Panels show (a) the full ground-truth DAG, (b) the coarsened DAG under grouped interventions, and (c) the coarsened DAG under finer (ungrouped) interventions. False negatives in dashed gray; no false positives.

as independent sources; and the mediators $(\tilde{V}_1, \tilde{V}_2)$ of color perturbations are separated from the common downstream effects $(\tilde{V}_3, \tilde{I}_3)$ of both color and polarization perturbations.

5. Discussion and Limitations

We develop a graphical framework for causal abstraction via coarsening: we characterize when a map from fine- to coarse-grained variables preserves causal structure (Definition 1) and show that the resulting set of coarsenings forms a sublattice of the partition refinement lattice (Theorem 3). This yields a reusable foundation: any notion of abstraction that enforces the basic coarsening properties of Definition 1 corresponds to restricting attention to a principled subspace of this lattice.

`RePaRe` is one concrete realization of this view. We focus on the interventional coarsening $G^{\mathcal{I}}$ (Definition 8) because it has a clear causal interpretation—clusters are precisely the variables that are indistinguishable in terms of intervened-ancestor sets—and because it enables favorable worst-case complexity for learning the abstract graph (Theorem 17), with large-sample correctness under our assumptions (Corollary 16). The primary limitation is statistical: two-sample (refinement) and conditional independence (edge recovery) tests require sufficient sample sizes per intervention; in practice, performance degrades when sample size n does not grow sufficiently with the number of features d. Results also depend on the intervention set (which determines the attainable abstraction resolution) and can be affected by finite-sample error accumulation from repeated testing.

Future work could attempt to address these limitations with score- rather than constraint-based search. Another direction could be to explore different concrete realizations of our general framework, as hinted at in our essential and marginal coarsening examples.

Acknowledgements The work by AM was supported by Novo Nordisk Foundation grant number NNF20OC0062897. The work of FM was supported by Novo Nordisk Foundation grant number NNF23OC0085356. PM received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 883818). Part of this research was performed while AM and PM were visiting the Institute for Mathematical and Statistical Innovation (IMSI), which is supported by the National Science Foundation (Grant

No. DMS-1929348). The authors thank the organizers of the IMSI long program *Algebraic Statistics and Our Changing World* for helping start the collaboration as well as Benjamin Hollering, Joseph Johnson, Liam Solus, and Gaetano Tedesco for helpful discussions and feedback on earlier drafts.

References

Tara V. Anand, Adele H. Ribeiro, Jin Tian, and Elias Bareinboim. Causal effect identification in cluster DAGs. In *AAAI Conference on Artificial Intelligence*, 2023.

Tara Vafai Anand, Adele H. Ribeiro, Jin Tian, George Hripcsak, and Elias Bareinboim. Causal discovery over clusters of variables in Markovian systems. Technical Report R-128, Columbia CausalAI Laboratory, 2025. URL https://www.causalai.net/r128.pdf.

S. Andersson, D. Madigan, and M. Perlman. A characterization of Markov equivalence classes for acyclic digraphs. *The Annals of Statistics*, 25(2):505–541, 1997.

Christine W. Bang and Vanessa Didelez. Do we become wiser with time? On causal equivalence with tiered background knowledge. In *Uncertainty in Artificial Intelligence*, pages 119–129. PMLR, 2023.

Sander Beckers. Equivalent causal models. In *Proceedings of the Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI-21)*, volume 35, pages 6202–6209, 2021. doi: 10.1609/aaai.v35i7.17310.

Sander Beckers and Joseph Y. Halpern. Abstracting causal models. In *The Thirty-Third AAAI Conference on Artificial Intelligence*, pages 2678–2685, 2019. doi: 10.1609/aaai.v33i01.33012678. arXiv:1812.03789 [cs.AI].

Sander Beckers, Frederick Eberhardt, and Joseph Y. Halpern. Approximate causal abstractions. In *Uncertainty in Artificial Intelligence*, pages 606–615. PMLR, 2020.

Simon Bing, Jonas Wahl, and Jakob Runge. Structural causal bottleneck models. *arXiv preprint arXiv:2603.08682*, 2026.

Krzysztof Chalupka, Pietro Perona, and Frederick Eberhardt. Multi-level cause-effect systems. In *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*, volume 51 of *Proceedings of Machine Learning Research*, pages 361–369. PMLR, 2016.

David Maxwell Chickering. Optimal structure identification with greedy search. *Journal of Machine Learning Research*, 3(Nov):507–554, 2002.

Gabriele D'Acunto, Fabio Massimo Zennaro, Yorgos Felekis, and Paolo Di Lorenzo. Causal abstraction learning based on the semantic embedding principle. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*. PMLR, 2025.

Brian A. Davey and Hilary A. Priestley. *Introduction to lattices and order*. Cambridge University Press, 2002.

Danai Deligeorgaki, Alex Markham, Pratik Misra, and Liam Solus. Combinatorial and algebraic perspectives on the marginal independence structure of Bayesian networks. *Algebraic Statistics*, 14(2):233–286, 2023. doi: 10.2140/astat.2023.14.233.

Mathias Drton and Michael Eichler. Maximum likelihood estimation in Gaussian chain graph models under the alternative Markov property. *Scandinavian Journal of Statistics*, 33(2):247–257, 2006.

Joel Dyer, Nicholas Bishop, Yorgos Felekis, Fabio Massimo Zennaro, Anisoara Calinescu, Theodoros Damoulas, and Michael Wooldridge. Interventionally consistent surrogates for complex simulation models. In *Advances in Neural Information Processing Systems*, volume 37, pages 21814–21841. Neural Information Processing Systems Foundation, 2024.

Joel Dyer, Nicholas Bishop, Anisoara Calinescu, Michael Wooldridge, and Fabio Massimo Zennaro. Using causal abstractions to accelerate decision-making in complex bandit problems. arXiv preprint arXiv:2509.04296, 2025.

Doris Entner and Patrik O. Hoyer. Estimating a causal order among groups of variables in linear models. In *Artificial Neural Networks and Machine Learning – ICANN*, pages 84–91, 2012.

Itai Feigenbaum, Devansh Arpit, Shelby Heinecke, Juan Carlos Niebles, Weiran Yao, Caiming Xiong, Silvio Savarese, and Huan Wang. Causal layering via conditional entropy. In *Causal Learning and Reasoning*, pages 1176–1191. PMLR, 2024.

Yorgos Felekis, Fabio Massimo Zennaro, Nicola Branchini, and Theodoros Damoulas. Causal optimal transport of abstractions. In *Proceedings of the Third Conference on Causal Learning and Reasoning*, volume 236 of *Proceedings of Machine Learning Research*, pages 462–498. PMLR, 2024.

Yorgos Felekis, Theodoros Damoulas, and Paris Giampouras. Distributionally robust causal abstractions. *arXiv preprint arXiv:2510.04842*, 2025.

Juan L. Gamella, Armeen Taeb, Christina Heinze-Deml, and Peter Bühlmann. Characterization and greedy learning of Gaussian structural causal models under unknown interventions. arXiv preprint arXiv:2211.14897, 2022.

Juan L. Gamella, Jonas Peters, and Peter Bühlmann. Causal chambers as a real-world physical testbed for AI methodology. *Nature Machine Intelligence*, 7(1):107–118, January 2025. ISSN 2522-5839. doi: 10.1038/s42256-024-00964-x.

Konstantin Göbler, Tobias Windisch, and Mathias Drton. Nonlinear causal discovery for grouped data. arXiv preprint arXiv:2506.05120, 2025.

Arthur Gretton, Kenji Fukumizu, Choon Teo, Le Song, Bernhard Schölkopf, and Alexander Smola. A kernel statistical test of independence. *Advances in Neural Information Processing Systems*, 20, 2007.

Arthur Gretton, Karsten M. Borgwardt, Malte J. Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *The Journal of Machine Learning Research*, 13(1):723–773, 2012.

Moritz Grosse-Wentrup, Akshey Kumar, Anja Meunier, and Manuel Zimmer. Neuro-cognitive multilevel causal modeling: A framework that bridges the explanatory gap between neuronal activity and cognition. *PLOS Computational Biology*, 20(12):e1012674, 2024.

Jiaying Gu and Qing Zhou. Learning big gaussian bayesian networks: Partition, estimation and fusion. *Journal of Machine Learning Research*, 21(158):1–31, 2020. URL http://jmlr.org/papers/v21/19-318.html.

Alain Hauser and Peter Bühlmann. Characterization and greedy learning of interventional Markov equivalence classes of directed acyclic graphs. *The Journal of Machine Learning Research*, 13 (1):2409–2464, 2012.

Harold Hotelling. Relations between two sets of variates. *Biometrika*, 1936.

Lawrence Hubert and Phipps Arabie. Comparing partitions. *Journal of Classification*, 2:193–218, 1985.

Yibo Jiang and Bryon Aragam. Learning nonparametric latent causal graphs with unknown interventions. *Advances in Neural Information Processing Systems*, 36:60468–60513, 2023.

Markus Kalisch and Peter Bühlmann. Estimating high-dimensional directed acyclic graphs with the PC-algorithm. *Journal of Machine Learning Research*, 8(3), 2007.

Armin Kekić, Bernhard Schölkopf, and Michel Besserve. Targeted reduction of causal models. In Negar Kiyavash and Joris M. Mooij, editors, *Proceedings of the Fortieth Conference on Uncertainty in Artificial Intelligence*, volume 244 of *Proceedings of Machine Learning Research*, pages 1953–1980. PMLR, 15–19 Jul 2024. URL https://proceedings.mlr.press/v244/kekic24a.html.

Armin Kekić, Jan Schneider, Dieter Büchler, Bernhard Schölkopf, and Michel Besserve. Learning nonlinear causal reductions to explain reinforcement learning policies. arXiv preprint arXiv:2507.14901, 2025.

Steffen L. Lauritzen. *Graphical Models*. Oxford University Press, 05 1996. ISBN 9780198522195. doi: 10.1093/oso/9780198522195.001.0001. URL https://doi.org/10.1093/oso/9780198522195.001.0001.

Steffen L. Lauritzen and Thomas S. Richardson. Chain graph models and their causal interpretations. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 64(3):321–348, 2002.

Tabitha Edith Lee, Shivam Vats, Siddharth Girdhar, and Oliver Kroemer. SCALE: Causal learning and discovery of robot manipulation skills using simulation. In *Proceedings of The 7th Conference on Robot Learning*, volume 229 of *Proceedings of Machine Learning Research*, pages 2229–2256. PMLR, 2023.

Yu Liu. CWGCNA: An R package to perform causal inference from the wgcna framework. *NAR Genomics and Bioinformatics*, 6(2):lqae042, April 2024. doi: 10.1093/nargab/lqae042.

Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. *Introduction to Information Retrieval*. Cambridge University Press, 2008.

Alex Markham and Moritz Grosse-Wentrup. Measurement dependence inducing latent causal models. In Jonas Peters and David Sontag, editors, *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI)*, volume 124 of *Proceedings of Machine Learning Research*, pages 590–599. PMLR, 03–06 Aug 2020. URL https://proceedings.mlr.press/v124/markham20a.html.

Alex Markham, Richeek Das, and Moritz Grosse-Wentrup. A distance covariance-based kernel for nonlinear causal clustering in heterogeneous populations. In Bernhard Schölkopf, Caroline Uhler, and Kun Zhang, editors, *Proceedings of the First Conference on Causal Learning and Reasoning*, volume 177 of *Proceedings of Machine Learning Research*, pages 542–558. PMLR, 11–13 Apr 2022a. URL https://proceedings.mlr.press/v177/markham22a.html.

Alex Markham, Danai Deligeorgaki, Pratik Misra, and Liam Solus. A transformational characterization of unconditionally equivalent Bayesian networks. In Antonio Salmerón and Rafael Rumí, editors, *Proceedings of The 11th International Conference on Probabilistic Graphical Models*, volume 186 of *Proceedings of Machine Learning Research*, pages 109–120. PMLR, 05–07 Oct 2022b. URL https://proceedings.mlr.press/v186/markham22a.html.

Alex Markham, Mingyu Liu, Bryon Aragam, and Liam Solus. Neuro-causal factor analysis, 2023. URL https://arxiv.org/abs/2305.19802.

Riccardo Massidda, Sara Magliacane, and Davide Bacciu. Learning causal abstractions of linear structural causal models. In Negar Kiyavash and Joris M. Mooij, editors, *Proceedings of the Fortieth Conference on Uncertainty in Artificial Intelligence*, volume 244 of *Proceedings of Machine Learning Research*, pages 2486–2515. PMLR, 15–19 Jul 2024. URL https://proceedings.mlr.press/v244/massidda24a.html.

Henri Meyniel. Une condition suffisante d'existence d'un circuit hamiltonien dans un graphe oriente. *Journal of Combinatorial Theory, Series B*, 14(2):137–147, 1973.

Jun Otsuka and Hayato Saigo. On the equivalence of causal models: A category-theoretic approach. In *Conference on Causal Learning and Reasoning*, pages 634–646. PMLR, 2022.

Alvin C. Rencher and William F. Christensen. *Methods of Multivariate Analysis*. Wiley, 2 edition, 2012. doi: 10.1002/9781118391686.

Paul K. Rubenstein, Sebastian Weichwald, Stephan Bongers, Joris M. Mooij, Dominik Janzing, Moritz Grosse-Wentrup, and Bernhard Schölkopf. Causal consistency of structural equation models. In *Conference on Uncertainty in Artificial Intelligence (UAI)*, 2017.

Pedro Sanchez, Xiao Liu, Alison Q. O'Neil, and Sotirios A. Tsaftaris. Diffusion models for causal discovery via topological ordering. In *International Conference on Learning Representations*, 2023.

Willem Schooltink and Fabio Massimo Zennaro. Aligning graphical and functional causal abstractions. In *Proceedings of the Fourth Conference on Causal Learning and Reasoning*, volume 275 of *Proceedings of Machine Learning Research*. PMLR, 2025.

Shohei Shimizu, Takanori Inazumi, Yasuhiro Sogawa, Aapo Hyvärinen, Yoshinobu Kawahara, Takashi Washio, Patrik O. Hoyer, and Kenneth Bollen. DirectLiNGAM: A direct method for learning a linear non-Gaussian structural equation model. *Journal of Machine Learning Research*, 12:1225–1248, 2011.

Chandler Squires, Yuhao Wang, and Caroline Uhler. Permutation-based causal structure learning with unknown intervention targets. In *Conference on Uncertainty in Artificial Intelligence*, pages 1039–1048. PMLR, 2020.

Chandler Squires, Annie Yun, Eshaan Nichani, Raj Agrawal, and Caroline Uhler. Causal structure discovery between clusters of nodes induced by latent factors. In *Conference on Causal Learning and Reasoning*, pages 669–687. PMLR, 2022.

Richard P. Stanley. *Enumerative Combinatorics, Volume 1*. Cambridge Studies in Advanced Mathematics, second edition, 2011.

Gábor J. Székely and Maria L. Rizzo. Energy statistics: A class of statistics based on distances. *Journal of Statistical Planning and Inference*, 143(8):1249–1272, 2013.

Gábor J. Székely, Maria L. Rizzo, and Nail K. Bakirov. Measuring and testing dependence by correlation of distances. *The Annals of Statistics*, 35(6):2769–2794, 2007.

Santtu Tikka, Jouni Helske, and Juha Karvanen. Clustering and structural robustness in causal diagrams. *Journal of Machine Learning Research*, 24(195):1–32, 2023.

Jonas Wahl, Urmi Ninad, and Jakob Runge. Vector causal inference between two groups of variables. In *AAAI Conference on Artificial Intelligence*, 2023.

Jonas Wahl, Urmi Ninad, and Jakob Runge. Foundations of causal discovery on groups of variables. *Journal of Causal Inference*, 12(1), 2024.

Yuhao Wang, Liam Solus, Karren Dai Yang, and Caroline Uhler. Permutation-based causal inference algorithms with interventions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 5824–5833, 2017. ISBN 9781510860964.

Bernard L. Welch. The generalization of Student's problem when several different population variances are involved. *Biometrika*, 34(1-2):28–35, 1947.

Kevin Xia and Elias Bareinboim. Neural causal abstractions. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 2024. doi: 10.1609/aaai.v38i18.30044.

Jinming Xiao, Qing Yin, Lei Li, Yao Meng, Xiaobo Liu, Wanrou Hu, Xinyue Huang, Yu Feng, Xiaolong Shan, Weixing Zhao, Peng Wang, Xiaotian Wang, Youyi Li, Huafu Chen, and Xujun Duan. Tracking causal pathways in TMS-evoked brain responses. *PLOS Computational Biology*, 21(7):e1013316, July 2025. doi: 10.1371/journal.pcbi.1013316.

Sascha Xu, Sarah Mameche, and Jilles Vreeken. Information-theoretic causal discovery in topological order. In *Proceedings of the 28th International Conference on Artificial Intelligence and Statistics*, Proceedings of Machine Learning Research, pages 2008–2016, 2025.

Fabio Massimo Zennaro, Máté Drávucz, Geanina Apachitei, W. Dhammika Widanage, and Theodoros Damoulas. Jointly learning consistent causal abstractions over multiple interventional distributions. In *Proceedings of the Second Conference on Causal Learning and Reasoning*, volume 213 of *Proceedings of Machine Learning Research*, pages 88–121. PMLR, 2023.

Yuchen Zhu, Sergio Hernan Garrido Mejia, Bernhard Schölkopf, and Michel Besserve. Unsupervised causal abstraction. In *Causal Representation Learning Workshop*, 2024.

Supplementary Material

Appendix A. Additional Related Work

Here we briefly elaborate on connections to related work, focusing on three perspectives: (i) causal abstraction learning, (ii) causal discovery methods that recover a DAG or an interventional equivalence class over the original variables, and (iii) approaches to causal discovery with grouped or multivariate nodes when the grouping is assumed known. This appendix highlights the key distinction of our setting: we aim to *discover* the grouping (a valid partition) and its induced abstract DAG directly from data. See Table 2 for a summarized comparison to related work.

A.1. Extensions in Causal Abstraction Learning

Building on early foundational work that defines macro-variables through observational and interventional equivalence (Chalupka et al., 2016), recent literature has significantly expanded the methodological toolkit for learning causal abstractions directly from data. Several approaches parameterize the abstraction map using deep learning architectures. For instance, Xia and Bareinboim (2024) introduce neural causal abstractions to flexibly learn high-level causal representations, while Bing et al. (2026) enforce structural causal constraints at the macro level through the use of causal bottleneck models. Taking a different representational approach, D'Acunto et al. (2025) propose learning these abstractions guided by the semantic embedding principle.

Another crucial direction in this space focuses on the robustness, consistency, and alignment of abstractions across different environments or functional forms. Zennaro et al. (2023) address the challenge of jointly learning abstractions that remain consistent across multiple interventional distributions. To handle complex distributional mappings between micro and macro levels, optimal transport has been leveraged to align abstractions (Felekis et al., 2024), an approach that was subsequently extended to ensure distributionally robust causal abstractions (Felekis et al., 2025). Furthermore, recent theoretical work formalizes the necessary alignment between graphical abstractions (operations on the DAG itself) and functional abstractions (operations on the underlying structural equations) (Schooltink and Zennaro, 2025).

Our framework can be formalized through the lens of Causal Abstraction (CA) (Rubenstein et al., 2017; Beckers and Halpern, 2019). In standard CA, an abstraction can be characterized by a tuple (τ, ω), where τ maps low-level states to high-level states and ω maps low-level interventions to high-level interventions. In the context of Coarsening Causal Models, the abstraction map τ is restricted to a surjective variable partition, denoted as $\chi : V \to V'$, which groups fine-grained variables into coarse clusters. Crucially, rather than learning an arbitrary intervention map ω, our framework enforces a canonical *constructive* intervention: an intervention on a coarse node $C \in V'$ is defined as the simultaneous intervention on the set of constituent micro-variables $\chi^{-1}(C)$. Therefore, the learning objective simplifies from finding a general pair (τ, ω) to identifying a valid partition χ such that this canonical ω preserves causal consistency (i.e., ensuring micro-variables within a cluster share interventional ancestors). This constraint structures the hypothesis space as a lattice, enabling efficient search via the RePaRe algorithm.

Reference	Setting	Partition	Edges	Code	Key Distinction
Structure learning given a known clustering					
Entner and Hoyer (2012)	Obs	Given	Learned	—	Causal order via linear non-Gaussianity
Wahl et al. (2023)	Obs	Given	Learned	Yes	Pairwise causal direction between two groups
Anand et al. (2025)	Obs	Given	Learned	—	Cluster-level discovery in Markovian systems
Göbler et al. (2025)	Obs	Given	Learned	Yes	Nonlinear additive noise for grouped variables
Wahl et al. (2024)	Obs	Given	—	—	Faithfulness transfer from micro to group level
Causal inference given a known cluster DAG					
Anand et al. (2023)	Obs	Given	Given	—	Causal effect identification in cluster DAGs
Jointly learning abstractions and low-/high-level transformations					
Chalupka et al. (2016)†	Both	Learned	Learned	—	Macro-variables via obs./int. equivalence
Zennaro et al. (2023)†	Int	Partially Learned	Given	Yes	Consistency of abstractions across environments
Xia and Bareinboim (2024)†	Obs	Given/Learned	Learned	Yes	Neural parameterization of abstraction maps
Felekis et al. (2024)†	Int	Learned	Given	Yes	Optimal-transport-based abstraction alignment
Massidda et al. (2024)†	Obs	Learned	Learned	Yes	Learning causal abstractions from data
Felekis et al. (2025)†	Int	Learned	Learned	Yes	Distributionally robust causal abstractions
D'Acunto et al. (2025)†	Obs	Learned	—	Yes	Semantic-embedding-guided abstractions
Bing et al. (2026)†	Obs/Given	Learned	Given	Yes	Structural constraints via causal bottlenecks
Learning clusters that preserve causal effect identifiability					
Tikka et al. (2023)	—	Learned	Given	Yes	Clustering that preserves effect identifiability
Learning interventional coarsenings (partition-based)					
`RePaRe` (ours)	Int	Learned	Learned	Yes	Lattice structure; consistency and complexity guarantees

Table 2: Practical comparison of related approaches to coarse-grained causal modeling. *Setting*: whether the method uses observational (Obs), interventional (Int), or both types of data. *Partition*: whether the variable grouping is assumed given or learned. *Edges*: whether the coarse-grained causal structure is assumed given, learned, or not applicable (—). *Code*: whether public open-source code is readily available (linked if Yes). †These methods learn a general abstraction map, not restricted to a variable partition.

A.2. Causal Discovery over Original Variables

Most causal discovery algorithms operate at the finest-grained level, finding a DAG or an equivalence class over the original nodes. Methods such as GES and its interventional variant GIES search over DAGs or Markov equivalence classes in observational and interventional settings (Hauser and Bühlmann, 2012) or under unknown interventions like GnIES (Gamella et al., 2022). Permutation-based procedures such as IGSP (Wang et al., 2017) or under unknown interventions UT-IGSP (Squires et al., 2020) target interventional Markov equivalence classes. Other approaches leverage additional structural assumptions: DirectLiNGAM learns a linear non-Gaussian DAG from observational data (Shimizu et al., 2011), and recent information-theoretic methods (Xu et al., 2025) and diffusion-model approaches (Sanchez et al., 2023) recover the causal structure via a topological order. While powerful, these methods typically do not address the case where the target causal relationships live at a coarser level than the observed features.

A.3. Causal Discovery with Grouped Nodes

A related line of work studies causal discovery *with grouped or multivariate nodes*. Early work extends linear non-Gaussian identifiability to infer a causal order among *known* groups (Entner and Hoyer, 2012). Two-group methods then decide direction between *pre-specified* multivariate sets (Wahl et al., 2023). More recently, Göbler et al. (2025) extend nonlinear additive noise models to

random vectors and propose a two-step approach that learns a group-level order before selecting a compatible graph. Cluster DAGs provide a coarse graphical representation over *given* clusters and support effect identification and cluster-level discovery in Markovian systems (Anand et al., 2023, 2025). Complementarily, recent theory examines when faithfulness transfers from micro-level graphs to graphs over groups, revealing subtle failure modes (Wahl et al., 2024). Our setting differs in a key aspect: we do *not* assume the grouping is given, but instead aim to *discover* a valid partition and its induced abstract DAG.

Appendix B. Proofs for Results in the Main Text

Proof [Lemma 2] Recall that $\mathcal{M}(G) \subseteq \mathcal{M}(G')$ is equivalent to stating that for any disjoint sets $A', B', C' \subseteq V'$, if $A' \perp_{G'} B' \mid C'$, then $\chi^{-1}(A') \perp_G \chi^{-1}(B') \mid \chi^{-1}(C')$. Contrapositively, we show that if the pre-images are d-connected in G, their images are d-connected in G'.

Let $A = \chi^{-1}(A')$, $B = \chi^{-1}(B')$, and $C = \chi^{-1}(C')$. Suppose $A \not\perp_G B \mid C$. Then there exists an active path τ in G between some $a \in A$ and $b \in B$. Consider the sequence of nodes τ' in G' formed by applying χ to the nodes in τ, removing consecutive duplicates. We verify that τ' constitutes an active path in G' relative to C':

- For every edge $u \to v$ in τ, Definition 1 implies that either $\chi(u) \to \chi(v)$ is an edge in G' or $\chi(u) = \chi(v)$. In either case, the nodes lie on τ'.
- If v_i is a non-collider on τ, then $v_i \notin C$. Since $C = \chi^{-1}(C')$, we have $\chi(v_i) \notin C'$. Thus, the non-collider condition is satisfied in G'.
- If v_j is a collider on τ, then $v_j \in C$ or some descendant v_d of v_j is in C. Notice that by Definition 1,$v_d \in \mathrm{de}_G(v_j)$ implies $\chi(v_d) \in \mathrm{de}_{G'}(\chi(v_j))$. Thus, $\chi(v_j)$ or one of its descendants in G' is in $\chi(C) = C'$, so the collider condition is satisfied in G'.

■

Proof [Theorem 3] Let $\mathcal{L}$ be the set of all valid coarsenings of G. We equip $\mathcal{L}$ with the standard refinement ordering $\preceq$.

First, we establish the existence of the meet. Consider $A, B \in \mathcal{L}$ with partitions Π_A and Π_B. Their meet in the partition lattice, $\Pi_{A \wedge B}$, consists of the non-empty intersections $\{U \cap V \mid U \in \Pi_A, V \in \Pi_B\}$.

Next, we verify that $\Pi_{A \wedge B}$ corresponds to a valid coarsening. Following Definition 1, let $\chi : V \to \Pi_{A \wedge B}$ be the surjection mapping each node to its part and $E' = \{\chi(u) \to \chi(v) \mid u \to v \in E, \chi(u) \neq \chi(v)\}$ be the induced edges. Suppose for contradiction that E' contains a cycle $W_1 \to W_2 \to \cdots \to W_k \to W_1$. For each edge, there exist $u_i \in W_i, u_{i+1} \in W_{i+1}$ with $u_i \to u_{i+1} \in E$. Writing $W_i = U_i \cap V_i$ with $U_i \in \Pi_A, V_i \in \Pi_B$, we have $U_i \neq U_{i+1}$ or $V_i \neq V_{i+1}$. Whenever $U_i \neq U_{i+1}$, the edge $u_i \to u_{i+1}$ induces $U_i \to U_{i+1}$ in A's induced graph. Following the cycle, at least one of Π_A or Π_B admits a cycle in its induced edges, contradicting that A and B are DAGs. Thus E' is acyclic and $\Pi_{A \wedge B}$ is a valid coarsening.

Finally, notice that the coarsest common coarsening is the trivial partition $\mathbf{1} = \{V\}$, which is always a valid coarsening. Thus, $\mathcal{L}$ is a finite meet-semilattice with a top element, so (Stanley, 2011, Proposition 3.3.1) guarantees that it is a lattice. Since $\mathcal{L}$ is a subset of partitions of V ordered by refinement, it is a sublattice of the partition refinement lattice. ■

Proof [Theorem 6] Let $\underline{G}$ be the underlying DAG and $G^* = (\Pi^*, E^*)$ be the target valid coarsening. We explicitly construct the oracles as follows:

- `Refine`-oracle: Given a current partition Π, if $\Pi = \Pi^*$, return $\emptyset$. Otherwise, if $\Pi^* \prec \Pi$, there must exist a part $\pi \in \Pi$ that is not a part in Π^*. Since Π^* refines Π, π is the union of some subset of parts from Π^*, say $\pi = \bigcup_{j=1}^{m} \rho_j$ where $\rho_j \in \Pi^*$. We define the split π_a, π_b by partitioning this set of sub-parts. Because G^* is a valid coarsening, it is a DAG. Let $G^*[\pi]$ denote the subgraph of G^* induced by the node set $\{\rho_1, \ldots, \rho_m\}$. This induced subgraph must also be acyclic, so it contains at least one source node. Let ρ_1 be such a source node. We set $\pi_a = \rho_1$ and $\pi_b = \bigcup_{j=2}^{m} \rho_j$.

  We now verify Condition 1 (acyclicity preservation). Suppose for contradiction that there is a cycle between π_a and π_b in $\underline{G}$. This means there exist $u \in \pi_a = \rho_1$ and $v \in \pi_b$ such that $v \in \mathrm{an}_{\underline{G}}(u)$. By Definition 1, since (v, u) lies on a directed path in $\underline{G}$, the coarsening property ensures that $\chi(v)$ is an ancestor of $\chi(u)$ in G^*. However, $\chi(u) = \rho_1$ and $\chi(v) \in \{\rho_2, \ldots, \rho_m\}$, so there exists an edge from some part in $\{\rho_2, \ldots, \rho_m\}$ to ρ_1 in $G^*[\pi]$. This contradicts the choice of ρ_1 as a source node in the acyclic graph $G^*[\pi]$. Therefore, $\mathrm{an}_{\underline{G}}(\pi_a) \cap \pi_b = \emptyset$, satisfying Condition 1.
- `IsEdge`-oracle: The oracle returns `True` for parts u, v if and only if $\mathrm{pa}_{\underline{G}}(v) \cap u \neq \emptyset$. This satisfies Condition 2 (parent consistency) of Definition 5 directly. By Definition 1, an edge $u \to v$ appears in G^* if and only if there exist $x \in u, y \in v$ with $x \in \mathrm{pa}_{\underline{G}}(y)$ and $u \neq v$. Our `IsEdge`-oracle reconstructs exactly this: it returns `True` whenever $\mathrm{pa}_{\underline{G}}(v) \cap u \neq \emptyset$ (for $u \neq v$), which is precisely the condition for an edge in the coarsening.

Initialized with the trivial partition $\Pi_0 = \{V\}$, the algorithm iteratively applies these oracles. By construction, if $\Pi^* \prec \Pi_t$ and a split occurs, the new partition Π_{t+1} is formed by splitting a part π into unions of parts from Π^*, ensuring $\Pi^* \preceq \Pi_{t+1}$. Since the partition becomes strictly finer at each step and the lattice is finite, the algorithm must terminate. Termination occurs if and only if the `Refine`-oracle returns $\{\emptyset, \emptyset\}$, which happens precisely when $\Pi_t = \Pi^*$. Finally, the edge set is constructed at each step by the `IsEdge`-oracle according to the definition of a valid coarsening, ensuring the final output is exactly G^*. ■

Proof [Theorem 10] We show that there exist `Refine`- and `IsEdge`-oracles satisfying Definition 5 that can be implemented using only the family of interventional distributions $\{f^I\}_{I \in \mathcal{I}}$ and Assumption 9. Together with Theorem 6, this implies that $G^{\mathcal{I}}$ is identifiable.

For each node $v \in V$ and each intervention $I \in \mathcal{I} \setminus \emptyset$, interventional soundness (Assumption 9.3) along with its converse (which follows from Assumption 9.1) gives

$$v \in \mathrm{de}_G(I) \quad \Longleftrightarrow \quad f^I(X_v) \neq f^{\emptyset}(X_v).$$

Thus, from the distributions alone one can decide for every pair (v, I) whether v is a descendant of I in G.

Define the (observable) intervention signature of a node v as

$$\sigma(v) := \{I \in \mathcal{I} \setminus \{\emptyset\} \mid f^I(X_v) \neq f^{\emptyset}(X_v)\}.$$

This $\sigma(v)$ is precisely the set of interventions whose targets have a directed path to v. Since $\sigma(v)$ depends only on $\mathcal{I}$-$\mathrm{an}_G(v)$, we have $\sigma(v) = \sigma(w)$ whenever $\mathcal{I}$-$\mathrm{an}_G(v) = \mathcal{I}$-$\mathrm{an}_G(w)$.

By Definition 8, the true partition $\Pi^{\mathcal{I}}$ is exactly the partition of V into signature equivalence classes.

We construct a `Refine`-oracle that uses only the signatures $\sigma(v)$ and satisfies acyclicity preservation in Definition 5. Given the current partition Π, if all $v \in \pi$ share the same signature $\sigma(v)$ for

every $\pi \in \Pi$, the oracle returns $\pi^* = \emptyset$. Otherwise, there exists some $\pi \in \Pi$ containing nodes with at least two distinct signatures. Fix such a π and choose some $u \in \pi$. Define

$$\pi_a := \{v \in \pi \mid \sigma(v) = \sigma(u)\}, \qquad \pi_b := \pi \setminus \pi_a,$$

and set $\pi^* = \pi$.

By construction, every node in π_a lies in the same part of $\Pi^{\mathcal{I}}$ as u, whereas every node in π_b lies in a different part of $\Pi^{\mathcal{I}}$. In particular, in the true interventional coarsening $G^{\mathcal{I}}$ there is no directed cycle between the coarse nodes containing π_a and π_b.

Suppose for contradiction that there exists a directed cycle in G alternating between nodes in π_a and π_b. Since $\Pi^{\mathcal{I}}$ is a valid coarsening of G (by Definition 8), this cycle would project to a directed cycle in $G^{\mathcal{I}}$ via the surjection χ. However, $G^{\mathcal{I}}$ is a DAG by assumption, a contradiction. Therefore, no such cycle exists, which means either $\mathrm{an}_G(\pi_a) \cap \pi_b = \emptyset$ or $\mathrm{an}_G(\pi_b) \cap \pi_a = \emptyset$. Thus the split (π_a, π_b) satisfies acyclicity preservation as required by Definition 5. Hence this construction defines a valid `Refine`-oracle which, starting from the trivial partition, refines until it reaches exactly the partition into signature classes, that is, $\Pi^{\mathcal{I}}$.

As the partition is iteratively refined toward $\Pi^{\mathcal{I}}$, edges are incrementally identified among parts using the `IsEdge`-oracle. By Assumption 9.1–2, the family $\{f^I\}_{I \in \mathcal{I}}$ is Markov and faithful with respect to the coarse DAG $G^{\mathcal{I}}$.

When each part π^* is refined into π_a and π_b (corresponding to refining G into G'), the oracle constructs edges (following Line 5 of Algorithm 1), leveraging the partial order $\preceq_G$ accumulated during refinement:

1. between refined parts: The refinement process ensures $\pi_a \preceq_{G'} \pi_b$ (by signature structure). The oracle tests $X_{\pi_a} \perp\!\!\!\perp X_{\pi_b} \mid X_{\mathrm{pa}_G(\pi^*)}$ to determine if there is an edge from π_a to π_b.
2. parents of refined parts: For each part $\pi \in \mathrm{pa}_G(\pi^*)$ the oracle tests $X_{\pi} \perp\!\!\!\perp X_{\pi_a} \mid X_{\mathrm{pa}_G(\pi^*)\setminus\{\pi\}}$. For π_b, the oracle additionally conditions on π_a if it was determined to be a parent in item 1 above: $X_{\pi} \perp\!\!\!\perp X_{\pi_b} \mid X_{(\mathrm{pa}_G(\pi^*)\setminus\{\pi\})\cup\{\pi_a\}}$.
3. children of refined parts: For each part $\pi \in \mathrm{ch}_G(\pi^*)$ and each $\pi_{\mathrm{new}} \in \{\pi_a, \pi_b\}$, let π_{sib} be the sibling part $\{\pi_a, \pi_b\} \setminus \{\pi_{\mathrm{new}}\}$. The oracle tests $X_{\pi_{\mathrm{new}}} \perp\!\!\!\perp X_{\pi} \mid X_{(\mathrm{pa}_G(\pi)\setminus\{\pi^*\})\cup\{\pi_{\mathrm{sib}}\}}$.

In line with the ordered Markov property, conditioning on any superset of the true parents (excluding the candidate parent) is sufficient for these tests, provided it does not include descendants. The conditioning sets in items 1–3 are chosen to satisfy this: they include known predecessors in the partial order (specifically including the sibling part, which guarantees the set contains all true parents of $G^{\mathcal{I}}$ other than the candidate) without including descendants. By coarse faithfulness, conditional independence among coarse variables correctly identifies edges in G'. The algorithm terminates when $\Pi = \Pi^{\mathcal{I}}$, at which point all edges have been identified.

We have exhibited `Refine`- and `IsEdge`-oracles that depend only on the distributions $\{f^I\}_{I \in \mathcal{I}}$ and satisfy the conditions of Definition 5. By Theorem 6, running `RePaRe` with these oracles terminates and outputs exactly the interventional coarsening $G^{\mathcal{I}}$. Therefore $G^{\mathcal{I}}$ is identifiable under Assumption 9. ■

Proof [Proof of Theorem 17] We break the algorithm into three phases.

In Phase 1, `RefineAux` computes the intervention descendant matrix M via Welch t-tests for each $(v, I) \in V \times \mathcal{I}$. Each test takes $O(n)$ time, giving total time $O(edn)$.

In Phase 2, `RefineTest` refines the partition by repeatedly splitting parts that contain nodes with different intervention descendant patterns. Starting from one part and ending at k parts requires

exactly $k-1$ splits. At each split, checking the intervention patterns for all nodes takes $O(de)$ time. Thus, Phase 2 takes $O(dek) = O(d^2e)$ (since $k \leq d$).

In Phase 3, `IsEdgeTest` applies CCA-based CI tests to pairs of parts. Each test takes $O(p^2n + p^3)$ time. At iteration i (for $i \in \{1, \ldots, k-1\}$), we split a part π_i^* with in-degree $|\operatorname{pa}_G(\pi_i^*)|$ and out-degree $|\operatorname{ch}_G(\pi_i^*)|$. Since the partition has i parts, π_i^* has at most $i-1$ parents and at most $i-1$ children, so $r_i := |\operatorname{pa}_G(\pi_i^*)| + |\operatorname{ch}_G(\pi_i^*)| \leq 2(i-1) = O(i)$. When splitting π_i^* into π_a and π_b, the oracle performs: one test between the refined parts, for each in-parent two tests, and for each out-child two tests. This gives $O(1 + 2r_i) = O(r_i)$ tests per split. Summing across all $k-1$ splits:

$$\sum_{i=1}^{k-1} r_i \leq \sum_{i=1}^{k-1} O(i) = O(k^2).$$

Therefore, Phase 3 performs $O(k^2)$ total CI tests, each costing $O(p^2n + p^3)$, giving Phase 3 time $O(k^2(p^2n + p^3))$.

Summing the three phases, we have $O(edn) + O(d^2e) + O(k^2(p^2n + p^3))$. When $d \leq n$ (the statistical regime), the second term is dominated by the first, and p^2n dominates p^3, leaving $O\big((de + k^2p^2)n\big)$. ■

Appendix C. Additional Theoretical Results

Theorem 18 *A coarsening poset of a DAG $G = ([d], E)$ is a distributive lattice only if the DAG has a directed $(d-1)$-path.*

Proof Suppose the longest directed path in G is $p = (v_1, v_2, \ldots, v_m)$ with $m < d$. Extend the partial order induced by G to a total order $t = (t_1 = v_1, t_2, \ldots, t_d = v_m)$. Consider a vertex $v^* = t_i \notin p$, and let $v^- = \max_{j<i}\{t_j \in p\}$ while $v^+ = \min_{j>i}\{t_j \in p\}$. Then $\{v, v'\}, \{v, v^*\}, \{v', v^*\}$ and $\{v, v', v^*\}$ are all valid partitions. This sublattice does not satisfy distributivity (it is just the partition refinement lattice on 3 nodes). ■

Theorem 18 tells us that DAGs that are sparse (in the sense of not connecting all nodes by at least a single long path, which can perhaps be related more explicitly to average degree (Meyniel, 1973)) do not result in distributive coarsening lattices. This is a somewhat negative result, because distributive lattices have nice algebraic properties (e.g., a relation to Gröbner bases) that would allow use of computational algebraic tools.

Appendix D. Additional Simulation Results

See `src/expt/workflow/scripts/evaluate.py` in the repo for implementation details of the F-score and ARI evaluation metrics.

D.1. Synthetic Data

D.1.1. Data Generation Details

For each sampled ground-truth DAG, intervention targets are selected uniformly at random without replacement. Interventions are soft, shift operations as implemented in `sempler` (Gamella et al., 2022).

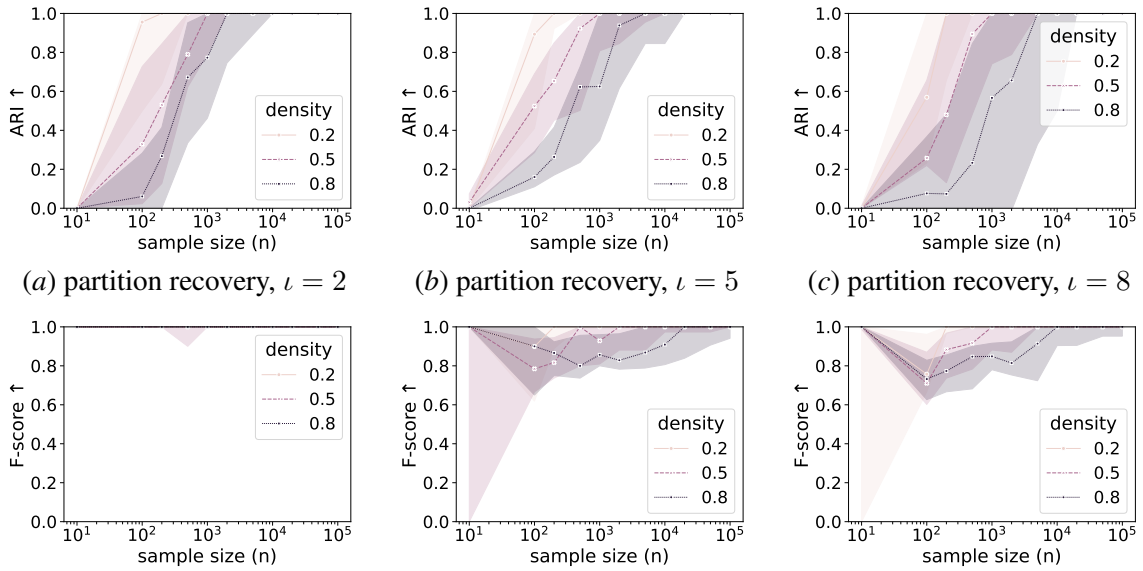
All variables are standardized before learning.

Edge F-scores are computed by comparing the learned coarsening $G = (\Pi, E)$ to the ground-truth DAG *coarsened under the same learned partition* Π. Concretely, we apply the surjection induced by Π to the ground-truth DAG to get ground-truth directed edges between the estimated parts, and then we compare the estimated edges to these ground-truth edges.

D.1.2. SCALE-FREE EXPERIMENTS

For the scale-free setting, we use the same linear Gaussian additive noise model (LGANM) framework implemented in `sempler` (Gamella et al., 2022), but replace the Erdős–Rényi topology by Barabási–Albert (BA) graphs. We fix $d = 10$ and, for each target density, translate it into an attachment parameter $m = \mathrm{round}(\max(\deg/2, 1))$. We then sample an undirected BA graph with `barabasi_albert_graph` from `networkx` library and orient its edges according to a random topological ordering to obtain a DAG with a heavy-tailed degree distribution. Edge weights are drawn uniformly from $[0.5, 2.0]$ with independent random signs, and we instantiate an LGANM by sampling node-wise means from $[-2, 2]$ and noise variances from $[0.5, 2]$.

For each ground-truth DAG, we generate one observational dataset and $\iota \in \{2, 5, 8\}$ interventional datasets. Intervention targets are selected uniformly at random without replacement, and soft interventions that shift the target mean by 2 with a variance of 1. As in the Erdős–Rényi experiments, we vary the per-environment sample size $n \in [10, 10^5]$ and repeat each $(\text{density}, n, \iota)$ configuration over 10 random seeds. All variables are standardized before learning.



(*a*) partition recovery, $\iota = 2$ (*b*) partition recovery, $\iota = 5$ (*c*) partition recovery, $\iota = 8$

(*d*) edge recovery, $\iota = 2$ (*e*) edge recovery, $\iota = 5$ (*f*) edge recovery, $\iota = 8$

Figure 5: Evaluation on synthetic data from scale-free DAG models, averaged over 10 seeds, as sample size per intervention increases, across different intervention budgets and ground-truth graph densities. **Top row**: ARI for evaluating partition recover. **Bottom row**: F-score for evaluating edge recovery.

D.2. Empirical Scalability

For the empirical computational evaluation, we again use LGANMs, with Erdős–Rényi. For each configuration we sample a DAG on $d \in \{10, 20, 50, 100, 200\}$ nodes by including each possible edge independently with probability $p = 0.2$. Edge weights are drawn uniformly from $[0.5, 2.0]$ with independent random signs, and we instantiate an LGANM by sampling node-wise means from $[-2, 2]$ and noise variances from $[0.5, 2]$.

For every ground-truth DAG, we generate one observational dataset and $\iota = 5$ interventional datasets with single-target interventions. Targets are selected uniformly at random without replacement, and soft interventions that shift the target mean by 2 with a variance of 1, as in the main Erdős–Rényi experiments. We vary the per-environment sample size $n \in \{10^2, 10^3, 10^4, 10^5\}$ and repeat each (d, n) configuration over 10 random seeds. All variables are standardized before learning.

Figure 3 summarizes these results. For each d, the median ARI increases reliably with n but larger graphs require substantially more samples to reach the same level of partition recovery. This is expected, since the number of possible coarsenings and adjacency relations grows quickly with number of nodes. At the same time, median runtime grows with both n and d, reflecting the fact that `RePaRe` repeatedly runs multivariate tests over parts whose size and number both increase. Thus, the regime of good statistical performance is constrained by computational cost in exactly the way suggested by the theory. Nevertheless, the curves also show that for moderate graph sizes the method is practically usable: for instance, around $d = 50$ we already obtain high ARI at sample sizes where wall-clock run time remains well below ten seconds, indicating that `RePaRe` can reliably recover informative coarsenings on nontrivial graphs within reasonable computational budgets.

D.3. Light-tunnel Data

D.3.1. LIGHT-TUNNEL VARIABLE SUBSET

We learn a coarsened DAG over the following $d = 20$ numeric variables (actuators and sensor readouts):

$$R,\ G,\ B,\ \tilde{C},\ \tilde{I}_1,\ \tilde{I}_2,\ \tilde{I}_3,\ \tilde{V}_1,\ \tilde{V}_2,\ \tilde{V}_3,$$
$$\theta_1,\ \theta_2,\ \tilde{\theta}_1,\ \tilde{\theta}_2,\ L_{11},\ L_{12},\ L_{21},\ L_{22},\ L_{31},\ L_{32}.$$

D.3.2. LIGHT-TUNNEL DATASETS AND GROUPED VS. UNGROUPED CONSTRUCTION

We use the observational dataset and five interventional datasets provided for the standard light-tunnel configuration.[6] From each experiment we extract a data matrix over the 20 selected variables.

In the grouped setting, we pool runs of the same intervention type, yielding three datasets:

1. observational: X^{obs} from `uniform_reference` (10,000 samples);
2. RGB pooled: X^{rgb} by concatenating `uniform_red_strong`, `uniform_green_strong`, `uniform_blue_strong` (3,000 samples total);
3. polarizer pooled: X^{pol} by concatenating `uniform_pol_1_strong`, `uniform_pol_2_strong` (2,000 samples total).

6. The data and documentation are available as the `lt_interventions_standard_v1` dataset at https://github.com/juangamella/causal-chamber.

In the ungrouped setting, we treat each of the five intervention experiments as a separate environment rather than pooling by type. In both cases we assume Gaussianity and instantiate partition refinement via `RefineAux` +`RefineTest` (based on a t-test) and adjacency via `IsEdgeTest` (Wilks' λ CCA test), as in Section 3.

D.3.3. LIGHT-TUNNEL BASELINES (FINE-GRAINED METHODS)

All baseline methods estimate directed structure over the 20 fine-grained variables. *GIES* (Hauser and Bühlmann, 2012) is a score-based greedy search method that returns a CPDAG and is in general not consistent (Wang et al., 2017). *UT-IGSP* is a permutation-based procedure that returns a DAG up to the same interventional Markov equivalence class (Squires et al., 2020). Finally, *GnIES* recovers an interventional equivalence class in a linear Gaussian model with unknown intervention targets and returns an interventional essential graph (Gamella et al., 2022).

D.3.4. MODEL SELECTION HEURISTIC FOR REPARE

We select hyperparameters without access to the ground truth by ranking candidates using a likelihood heuristic. For each candidate pair of thresholds $(\alpha_{\text{ref}}, \alpha_{\text{edge}})$, `RePaRe` learns an interventional coarsening $G = (\Pi, E)$ using `RefineAux` +`RefineTest` at level α_{ref} and `IsEdgeTest` at level α_{edge}.

To score this, we form an expansion $\widetilde{G}$ from the paritition back the original variables V. In $\widetilde{G}$, we retain all between-part directed edges from E. Within each part $\pi \in \Pi$, we assume a fully connected DAG (consistent with a fixed topological ordering $\leq_\sim$), which effectively allows for an arbitrary covariance structure within each part. Formally,

$$\widetilde{G} := \big(V, \{v \to w \mid \chi(v) \to \chi(w) \in G^{\mathcal{I}} \text{ or } (\chi(v) = \chi(w) \text{ and } v \leq_\sim w)\}\big).$$

Let $\mathcal{I}$ be the set of intervention targets (including $\emptyset$ for the observational setting), and let $\mathbf{X}^I \in \mathbb{R}^{n_I \times d}$ be the data matrix for intervention $I \in \mathcal{I}$ with sample size n_I. Following Gamella et al. (2022)[7], for a weight matrix B compatible with $\widetilde{G}$, we define the precision matrix implied by the graph for intervention I as:

$$K^I := [(\mathrm{Id} - B)^{-1}\Omega^I(\mathrm{Id} - B)^{-T}]^{-1},$$

where Id denotes the identity matrix. The log-likelihood under a Gaussian model is:

$$\log P(\mathbf{X}^I; B, \Omega^I) = -n_I \ln\det(K^I) - n_I \operatorname{tr}(K^I \hat{\Sigma}^I),$$

where $\hat{\Sigma}^I$ is the sample covariance for intervention I. We rank candidate coarsenings using the maximized interventional likelihood without penalization:

$$S(G, \mathcal{I}) = \max_{\substack{B \sim \widetilde{G} \\ \text{diag. pos. def. } \Omega^I I \in \mathcal{I} \\ \text{s.t. } \mathbb{I}(\Omega^I I \in \mathcal{I}) \subseteq \mathcal{I}}} \sum_{I \in \mathcal{I}} \log P(\mathbf{X}^I; B, \Omega^I)$$

Lastly, a coarsening $G = (\Pi, E)$ can be interpreted as a chain graph whose chain components are the parts in Π. This connects the heuristic above to Gaussian chain graph MLE theory and intervention semantics (Drton and Eichler, 2006; Lauritzen and Richardson, 2002).

7. We directly use the scoring code in https://github.com/juangamella/gnies with no penalization term.