
CAUSAL HOMOTOPY

A PREPRINT

Sridhar Mahadevan

Adobe Research, 345 Park Avenue, San Jose, CA 95110
smahadev@adobe.com

September 20, 2021

ABSTRACT

We characterize homotopical equivalences between causal DAG models, exploiting the close connections between partially ordered set representations of DAGs (posets) and finite Alexandroff topologies. Alexandroff spaces yield a directional topological space: the topology is defined by a unique minimal basis defined by open sets U_x for each variable x , specified as the intersection of all open sets containing x . Alexandroff spaces induce a (reflexive, transitive) preorder: a variable $x \leq y$ if $x \in U_y$. Alexandroff spaces satisfying the Kolmogorov T_0 separation criterion, where open sets distinguish variables, converts the preordering into a partial ordering. Our approach broadly is to construct a topological representation of posets from data, and then use the poset representation to build a conventional DAG-oriented causal model. We illustrate our framework by showing how it unifies disparate algorithms and case studies proposed previously. Topology plays two key roles in causal discovery. First, topological separability constraints on datasets have been used in several previous approaches to infer causal structure from observations and interventions. Second, a diverse range of graphical models used to represent causal structures can be represented in a unified way in terms of a topological representation of the induced poset structure. We show that the homotopy theory of Alexandroff spaces can be exploited to significantly efficiently reduce the number of possible DAG structures, reducing the search space by several orders of magnitude.

1 Introduction

Topology (Munkres, 1984a) has found extensive use in many areas in AI, machine learning and optimization. The Hahn-Banach theorem is a topological result concerning separation of points from convex sets by hyperplanes, and the entire framework of Lagrange duality can be derived from this topological insight (Luenberger, 1997). The Hahn-Banach theorem is also the basis for the universal representation theorem in deep neural networks (Cybenko, 1989). Topological data analysis techniques, such as persistent homology, are playing an increasingly important role in different areas of machine learning (Edelsbrunner, 2007; Zomorodian and Carlsson, 2005).

Graphical models have been extensively studied in artificial intelligence (AI), causal reasoning, machine learning (ML), physics, statistics, and many other fields (Koller and Friedman, 2009; Lauritzen, 1996; Pearl, 1989, 2019). Causal discovery (Spirtes et al., 2000) involves the construction of a causal model, for example a DAG graphical model structure and the specification of a probability model, from observational or experimental data. A broad family of models, ranging from Bayes networks (Pearl, 1989) on directed acyclic graphs (DAGs) to more recent variants, such as directed acyclic mixed graphs (ADMG) (Richardson, 2009), marginalized DAGs (mDAGs) (Evans, 2018) and hyperedge-directed graphical models (HEDGs) (Forre and Mooij, 2017), can be represented by finite Alexandroff spaces with different topological properties. For example, a DAG model imposes a T_0 topology on a finite Alexandroff space, which induces a partial ordering σ on the variables V in the model so that function f_i determining the value of variable X_{σ_i} in the model is measurable given the values of the previous variables $X_{\sigma_1}, \dots, X_{\sigma_{i-1}}$. A directed acyclic mixed graph (ADMG) (Richardson, 2009) and chain graphs (Andersson et al., 1996), on the other hand, have both undirected and directed edges, which induce only a preordering on the set of variables. Marginalized DAGs (MDAGs) (Evans, 2018) and HEDG (Forre and Mooij, 2017) models allow hyperedges between nodes, representing the effect

of latent variables. These can be represented using topological constructions, such as non-Hausdorff cones $\mathbb{C}(X)$ or non-Hausdorff suspensions $\mathbb{S}(X)$ (Barmak, 2011).

We propose a novel topological framework for causal inference, building an initial topological representation of a partially ordered set (poset) from data, prior to building a probabilistic graphical model from the poset (see Figure 1). To that end, we represent posets using the algebraic topology of finite Alexandroff spaces (Alexandroff, 1956; Barmak, 2011; May). Representing posets as topological spaces confers many computational advantages, such as the ability to combine multiple posets into a joint poset, and to use algebraic homotopy theory (Barmak, 2011; May) of finite topological spaces to significantly reduce the search space of possible structures. We show how a wide variety of graphical models, from chain graphs (Lauritzen and Richardson, 2002) to DAGs (Pearl, 1989), can be topologically represented as posets in a finite Alexandroff space. We illustrate our approach using a real-world dataset of pancreatic cancer (Beerenwinkel and Sullivant, 2009; Diaz-Uriarte, 2017). Our primary goal is to illustrate how algebraic topology provides some powerful tools to design more scalable algorithms for structure discovery, which otherwise presents an intractable combinatorial search space. We also relate our approach to previous studies of causal discovery showing how to classify them using the concept of *intervention topology* (Pearl, 2009; Spirtes et al., 2000; Eberhardt, 2008; Hauser and Bühlmann, 2012a; Kocaoglu et al., 2017; Mao-cheng, 1984; Tadepalli and Russell, 2021). (Bernstein et al., 2020) develop a greedy poset-based algorithm for learning DAG models, but do not exploit poset topological properties.

Our approach builds on a key technical innovation of using a topological representation of partially ordered sets, in between the original datasets and the final probabilistic or statistical graphical model. The principal reasons for explicitly modeling posets as topological spaces is that it allows us to exploit the rich algebraic theory of finite spaces (Barmak, 2011; May; McCord, 1966; Stong, 1966) to our computational advantage. For example, in modeling cancer disease progression, in addition to obtaining crucial timing information about mutations from clinical datasets on tumors and their associated genotypes, we can also exploit the disease pathways (Jones et al., 2008) that are known from medical research. Each source of information results in a different poset, which can be combined together using the algebraic topology theory of finite spaces. In addition, algebraic topology gives powerful tools for reducing the combinatorial search space of possible structures. We characterize homeomorphic equivalences among minimal poset models, and show homotopical equivalences sharply reduce the number of structures that need to be examined during structure discovery. In the case of pancreatic cancer, for example, we can reduce the search space of possible structures by three orders of magnitude.

The topology of posets of graphical models based on the algebraic topology of finite Alexandroff spaces (Alexandroff, 1956). We show how a wide variety of graphical models, from chain graphs (Lauritzen and Richardson, 2002) to DAGs (Pearl, 1989), can be topologically embedded in a finite Alexandroff space. For a DAG \mathcal{G} , simply compute the unique transitive closure graph \mathcal{G}_{tc} , and define the open sets of the induced topological model (\mathcal{M}, \leq) by defining the open sets $U_x \subset \mathcal{T}$ as the ancestors of a node x (including itself) in \mathcal{G}_{tc} . To revert from a topological model $\mathcal{M} = (X, \mathcal{T})$ to its DAG representation, form the Hasse diagram of the partial order defined by $x \leq y$ if $x \in U_y$. We present a novel algorithmic paradigm for structure discovery as iterating between searching among topologically distinct structures and causally faithful structures. We show how this paradigm can be used to characterize many previous studies of causal discovery in terms of the concept of *intervention topology*, collections of subsets intervened on to determine directionality (Pearl, 2009; Spirtes et al., 2000; Eberhardt, 2008; Hauser and Bühlmann, 2012a; Kocaoglu et al., 2017; Mao-cheng, 1984; Tadepalli and Russell, 2021). This connection immediately suggests topological generalizations of these previous algorithms.

2 Representing Causal DAGs as Finite Topological Spaces

We begin with brief review of basic point-set topology, and then give a succinct characterization of finite space topologies. Topology (Munkres, 1984a) characterizes the abstract properties of arbitrary spaces that are equivalent under smooth deformations, usually represented as continuous invertible mappings called homeomorphisms. Formally, a general topological space (X, \mathcal{U}) is characterized by a base space X , along with a collection of “open” sets $\mathcal{O}_i \subseteq \mathcal{U}$ closed under arbitrary union and *finite* intersection. Note the asymmetry in these restrictions. For example, if we define $X = \mathbb{R}$ to be the real line, and consider the open sets $\mathcal{O}_i = (-\frac{1}{i}, \frac{1}{i})$ to be the open intervals around 0 for $i = 1, \dots, \infty$, then $\bigcap_{i=1}^{\infty} (-\frac{1}{i}, \frac{1}{i}) = \{0\}$, which is not an open set!

Alexandroff (Alexandroff, 1956) pioneered the study of the subclass of topological spaces that are closed under both arbitrary union and intersection. While our framework can be potentially be extended to the non-finite case, for simplicity, we will restrict our presentation in this paper to the case of finite Alexandroff spaces (Barmak, 2011; May; McCord, 1966; Stong, 1966). It is obvious to note that finite topological spaces (X, \mathcal{U}) are trivially closed under arbitrary unions and intersections, because there are only a finite number of open sets $\mathcal{O} \in \mathcal{U}$. However, what turns out

to be surprising is that the particular construction used by Alexandroff in defining open sets applies even in the finite case, and results in spaces with surprising topological richness, even though they are finite.

Definition 1. A finite Alexandroff topological space (or simply, finite space, in the remainder of the paper) (X, \mathcal{U}) is a finite set X and a collection \mathcal{U} of “open” sets, namely subsets of X , such that (i) \emptyset (the empty set) and X are in \mathcal{U} (ii) Any union of sets in \mathcal{U} is in \mathcal{U} (iii) Any intersection of subsets in \mathcal{U} is in \mathcal{U} as well.

We will often refer to a finite space simply by its elements X , where the topology is left implicit, unless its character is important, when we will clarify it. The most common topologies on X will be the *discrete* topology, where the collection of open sets \mathcal{U} is just the powerset 2^X , and the *trivial* topology $\mathcal{U} = \{\emptyset, X\}$.

The major contribution of this paper is the use of a specific topological representation of partially ordered sets based on Alexandroff spaces (Alexandroff, 1956), who showed that finite topological spaces naturally defined preordered and partially ordered sets. Two classic papers by McCord (McCord, 1966) and Stong (Stong, 1966) laid the foundations for much of the subsequent study of finite topological spaces. Detailed proofs of all the main theorems on finite topological spaces in this paper can be found in (Barmak, 2011; May; McCord, 1966; Stong, 1966).

Remarkably, a key idea that is implicit in many causal discovery algorithms is the topological notion of *separability* (Beerenwinkel et al., 2007; Kocaoglu et al., 2017; Acharya et al., 2018), which intimately relates to the topology of the finite space. In order to construct a poset model from data, (Beerenwinkel et al., 2007) assume that the dataset $u : \mathcal{G} \rightarrow \mathbb{N}$ specifies the number of observations of each genotype g . For example, Table 5 specifies the number of tumors that contain a specific set of gene mutations g . The *support* $\mathcal{S}(u)$ is the non-zero coordinates of u , namely the genotypes that occur in the data. A crucial assumption here is that the dataset u *separates the events* e and f if there exists some genotype $g \in \mathcal{S}(u)$ such that $g \cap \{e, f\} \neq \emptyset$. Viewed more abstractly, this notion of separability implies that the underlying space has the T_0 Kolmogorov topology. (Kocaoglu et al., 2017) assume a separating set, which is essentially a restricted type of finite topological space.

Definition 2. The neighborhood of an element x in a finite space X is a subset $V \subset X$ such that $x \in U$ for some open set $U \subset V$.

- X is a Kolmogorov (or T_0) finite space X if each pair of points $x, y \in X$ is distinguishable in the space, namely for each $x, y \in X$, there is an open set $U \in \mathcal{U}$ such that $x \in U$ and $y \notin U$. Alternatively, if $x \in U$ if and only if $y \in U$, $\forall U \in \mathcal{U}$ implies that $x = y$.
- X is a T_1 finite space if element $x \in X$ defines a closed set $\{x\}$.
- X is a T_2 finite space or a Hausdorff space if any two points have distinct neighborhoods.

It turns out that T_1 finite spaces are not interesting since the only topology defined on them is the discrete (powerset) topology. The most interesting finite spaces are those equipped with the T_0 topology.

Lemma 1. (May) If X is a T_2 space, then it is a T_1 space. If X is a T_1 space, then it is a T_0 space.

The key concept that gives finite (Alexandroff) spaces its power is the definition of the minimal open basis. First, we introduce the concept of a basis in a topological space.

Definition 3. A basis for the topological space X is a collection \mathcal{B} of subsets of X such that

- For each $x \in X$, there is at least one $B \in \mathcal{B}$ such that $x \in B$.
- If $x \in B' \cap B''$, where $B, B'' \in \mathcal{B}$, then there is at least one $B \in \mathcal{B}$ such that $x \in B \subset B' \cap B''$.

The topology \mathcal{U} generated by the basis \mathcal{B} is the set of subsets U such that for every $x \in U$, there is a $B \in \mathcal{B}$ such that $x \in B \subset U$. In other words, $U \in \mathcal{U}$ if and only if U can be generated by taking unions of the sets in the basis \mathcal{B} . Now, we turn to giving the most important definition in Alexandroff spaces, namely the *unique minimal basis*.

Lemma 2. (May) Let X be a finite Alexandroff space. For each $x \in X$, define the open set U_x to be the intersection of all open sets that contain x . Define the relationship \leq on X by $x \leq y$ if $x \in U_y$, or equivalently, $U_x \subset U_y$ (where $x < y$ if the inclusion is strict). The open sets U_x constitute a **unique minimal basis** \mathcal{B} for X in that if \mathcal{C} is another basis for X , then $\mathcal{B} \subset \mathcal{C}$. Alternatively, define the closed sets $F_x = \{y \mid y \geq x\}$, which provide an equivalent characterization of finite Alexandroff spaces.¹

Note that the relation \leq defined above is a *preorder* because it is reflexive (clearly, $x \in U_x$) and transitive (if $x \in U_y$, and $y \in U_z$, then $x \in U_z$). However, in the special case where the finite space X has a T_0 topology, then the relation \leq

¹The minimal basic closed sets in a T_0 finite Alexandroff space correspond to the *ancestral sets* in a DAG graphical model.

becomes a partial ordering. This gives a topological way to model DAG models, which will play a crucial role in our framework.

Lemma 3. *A function $f : X \rightarrow Y$ from one finite space to another is continuous if and only if $f^{-1}(U)$ is open in X if U is open in Y .*

Lemma 4. *If X is a T_2 space, then it is a T_1 space. If X is a T_1 space, then it is a T_0 space.*

Lemma 5. *A function $f : X \rightarrow Y$ between two finite spaces is continuous if and only if it is order-preserving, meaning if $x \leq x'$ for $x, x' \in X$, this implies $f(x) \leq f(x')$.*

Lemma 6. *Let x, y be two comparable points in a finite space X . Then, there exists a path from x to y in X , that is, a continuous map $\alpha : (0, 1) \rightarrow X$ such that $\alpha(0) = x$ and $\alpha(1) = y$.*

Theorem 1. *If X is a finite topological space containing a point y such that the only open (or closed) subset of X containing y is X itself, then X is contractible. In particular, the non-Hausdorff cone $\mathbb{C}(X)$ is contractible for any X .*

Proof: Let $Y = \{*\}$ denote the space with a single element, $*$. Define the retraction mapping $r : X \rightarrow *$ by $r(x) = *$ for all $x \in X$, and define the inclusion mapping $i : Y \rightarrow X$ by $i(*) = y$. Clearly, $r \circ i = id_Y$. Define the homotopy $h : X \rightarrow X$ by $h(x, t) = x$ if $t < 1$, and $h(x, 1) = y$. Then, h is continuous, because for any open set U in X , if $y \in U$, then clearly $U = X$ (as X is the only open set containing y), and hence $h^{-1}(U) = X \times I$, which is open. If on the other hand, $y \notin U$, then $h^{-1}(U) = U \times [0, 1)$. It follows that h is a homotopy $h \cong id_X = i \circ r$. \square

Lemma 7. *If X is an finite Alexandroff space, then U_x is contractible. In particular, if X has a unique maximal point or unique minimal point, then X is contractible.*

2.1 Open sets induced by Causal DAGs

We use a motivational example of causal discovery for treatment of patients for the COVID pandemic using vaccines (Greinacher et al., 2021; Schultz et al., 2021). In the COVID vaccination causal discovery problem, we are given a finite set of 5 variables $X = \{\text{AZV}, \text{PF4}, \text{H}, \text{BC}, \text{VITT}\}$, defined as follows:

- **AZV:** This variable represents the administration of the AstraZeneca vaccine.
- **PF4:** A number of patients suffering from vaccine-induced abnormal blood clotting tested positive for heparin-induced platelet factor 4 (PF4).
- **Gender:** Many of the patients who exhibited adverse effects to the Covid vaccine were disproportionately women, so gender may be a causal factor.
- **HIT:** Heparin is a blood thinner used to prevent blood clots. Triggered by the immune system in response to heparin, HIT causes a low platelet count (thrombocytopenia).
- **VITT:** This variable denotes whether patients suffered from this rare vaccine-related variant of spontaneous heparin-induced thrombocytopenia that the authors of these studies referred to as vaccine-induced immune thrombotic thrombocytopenia.

Figure 1 illustrates a simple causal model for the COVID problem, represented both conventionally as a DAG, as well as two alternative representations of a finite topological space (X, \mathcal{T}) , where X is represented by the two variables shown, and \mathcal{T} is either a set of open sets, defined as the descendants of a node (including itself), or a set of closed sets comprised of the ancestors of a node (including itself). In the open set parameterization, there is an arrow from node x to node y whenever $x \in O_y$, that is, when the node x is in the open set corresponding to node y .

Figure 2 shows how latent variables are typically modeled in causal reasoning using acyclic mixed directed graphs (ADMGs), where the latent variable is represented by a dashed undirected edge connecting the two observable variables. Such latent variables can be captured in our finite space topological framework by the use of the non-Hausdorff cone construction, which is one of several ways of connecting two topological spaces. Recall the non-Hausdorff cone merging of topological space X with $\{*\}$ yields the new space $\mathbb{C}(X)$, whose open sets are now $O_{\mathbb{C}(X)} = O_X \cup \{X \cup \{*\}\}$.

To take a real-world example, recently two studies were published in the New England Journal of Medicine that described patients in Austria, Germany and Norway who developed an unexpected blood clotting disorder in reaction to their first dose of the AstraZeneca/Oxford COVID-19 vaccine (Greinacher et al., 2021; Schultz et al., 2021). Understanding causal pathways in such problems requires modeling the effects of tens of thousands of discrete and continuous variables, from the administration of the vaccine, heparin-induced platelet factors like PF4, thrombocytopenia (blood clots) and its various causes, and the entire previous medical history of the patient. Clinicians have to juggle through all these factors in describing a potential treatment (e.g., should heparin be given to a patient?).

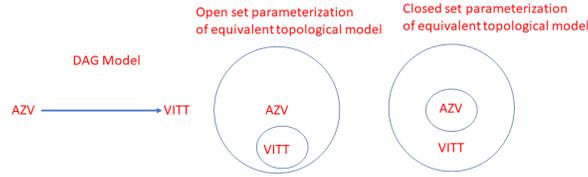


Figure 1: A simple example of a causal relationship between **AZV**, the administering of the AstraZeneca vaccine, and **VITT**, the potentially vaccine-induced clotting of blood. On the left is a traditional DAG model of this causal relationship. In the middle, we give one characterization of this DAG as a finite space topological model, where the open sets of the model correspond to descendants of each node (including the node itself). On the right is an equivalent characterization in terms of closed sets, which correspond to ancestors of a node (including the node itself).

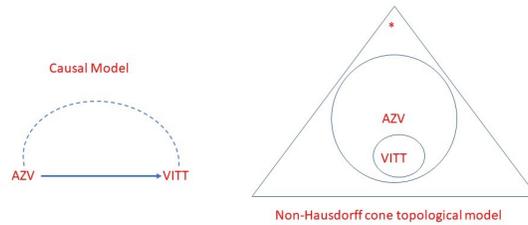


Figure 2: A simple causal model with a latent variable correlating **AZV**, the administering of the AstraZeneca vaccine, and **VITT**, the potentially vaccine-induced clotting of blood. On the left is a traditional acyclic mixed directed graphical model (ADMG) of this causal relationship. On the right is the equivalent finite space topological model, where the open sets of the model correspond to descendants of each node (including the node itself), and the latent unobserved variable is represented by the element $*$ joined to the remainder of the model using a non-Hausdorff cone.

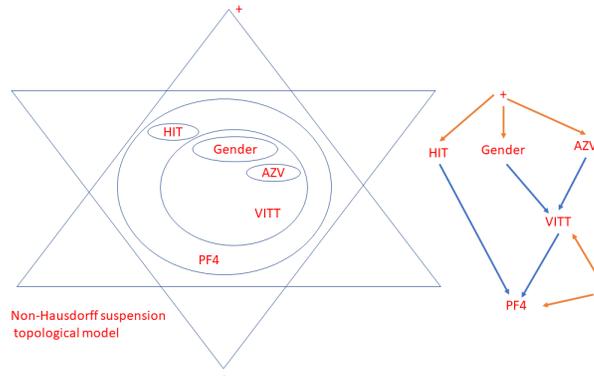


Figure 3: A directed graph with hyper-edges (HEDG) represented as a finite topological space. Open sets associated with variables are shown as circles or ellipsoids. The $+$ and $-$ nodes are two latent factors connected by hyper-edges to observables. The *non-Hausdorff suspension* \mathbb{S} defined in Section 2 permits composing the hyper-edge latent variable model with the topological model induced by the DAG over observables.

2.2 Connectivity in Topological Spaces

As mentioned above, every concept in a topological space must be defined in terms of the open (or closed) set topology, and that includes (path) connectivity. The crucial idea here is that connectivity is defined in terms of a continuous mapping from the unit interval $I = (0, 1)$ to a topological space X . Remarkably, the upshot of this construction is that every graph-theoretic concept in causal models, e.g. separation and conditional independence, can be translated into properties of open sets in the topology.

Lemma 8. A function $f : X \rightarrow Y$ between two finite spaces is continuous if and only if it is order-preserving, meaning if $x \leq x'$ for $x, x' \in X$, this implies $f(x) \leq f(x')$.

Definition 4. We call two points $x, y \in X$ **comparable** if there is a sequence of elements x_0, \dots, x_n , where $x_0 = x, x_n = y$ and for each pair x_i, x_{i+1} either $x_i \leq x_{i+1}$ or $x_i \geq x_{i+1}$. A **fence** in X is a sequence x_0, x_1, \dots, x_n of elements such that any two consecutive elements are comparable. X is **order connected** if for any two elements $x, y \in X$, there exists a fence starting in x and ending in y .

Lemma 9. Let x, y be two comparable points in a finite space X . Then, there exists a path from x to y in X , that is, a continuous map $\alpha : (0, 1) \rightarrow X$ such that $\alpha(0) = x$ and $\alpha(1) = y$.

Lemma 10. Let X be a finite space. The following are equivalent: (i) X is a connected topological space. (ii) X is an order-connected topological space (iii) X is a path-connected topological space.

To illustrate the notion of connectivity, Table 4 gives examples of connected and disconnected finite space topologies for a small three element space.

2.3 Separation and Conditional Independence in Topological Spaces

We now give a purely topological characterization of separation and conditional independence in finite topological spaces, which draw upon equivalent notions in graphical models (Lauritzen and Richardson, 2002; Pearl, 1989), but are defined with respect to the open sets of the topology.

Definition 5. Given a connected finite space X , with an induced (pre,partial) ordering \leq , and subsets $U, V, Z \subset X$, the subset U is **topologically blocked or d-separated** from V given Z , if for every fence from an element $u \in U$ to an element $v \in V$, the following conditions hold:

1. The fence x_0, x_1, \dots, x_n , where $x_0 = u$ and $x_n = v$ is such that every consecutive pair of elements is of the form $x_i \leq x_{i+1}$ or $x_{i+1} \leq x_i$, and some element $x_k \in Z$ (this condition is equivalent to stating that all the edges are of the form $x_i \rightarrow x_{i+1}$ or $x_i \leftarrow x_{i+1}$).
2. The fence x_0, x_1, \dots, x_n , where $x_0 = u$ and $x_n = v$ is such that for every **collider** in the fence, namely a triple of elements x_i, x_{i+1}, x_{i+2} is such that $x_i \leq x_{i+1}$, and $x_{i+2} \leq x_{i+1}$, it holds that $U_{x_{i+1}} \cap Z = \emptyset$. This condition is the topological restatement of the standard collider condition in graphical models, where for a path to be blocked, no collider or any of its descendants can be in the conditioning set.

Definition 6. Given a finite space X , and subsets $U, V, Z \subset X$, U is **topologically conditionally independent (TCI)** of V given Z if and only if every fence from an element $u \in U$ to an element $v \in V$ is topologically blocked with respect to the conditioning set Z .

3 Stable and Solvable Causal Models over Finite Topological Spaces

We now precisely define *stable* and *solvable* causal models, including both their Alexandroff topological structure, and a decomposable product probability measure constituting the parameters of the model. We impose the condition that the product probability measure respect the underlying Alexandroff topology, namely its open (or closed) sets, and requirement of a particular factorization is translated into a requirement of a particular topology. Our formulation is related to the *intrinsic* model of decision making (Witsenhausen, 1975), which was recently adapted to causal inference (Heymann et al., 2020). Neither of these investigated Alexandroff topologies.

Definition 7. A **finite space causal model** is defined as $\mathcal{M} = (U_\alpha, \mathcal{F}_\alpha, \mathcal{I}_\alpha, (\Omega, \mathcal{B}, P))$, where $\alpha \in X$, a finite Alexandroff space topology. U_α is a non-empty set that defines the range of values that variable α can take. \mathcal{F}_α is a sigma field (or algebra) of measurable sets for variable α . The triple (Ω, \mathcal{B}, P) is a probability space, where \mathcal{B} is a sigma field of measurable subsets of sample space Ω . The **information field** $\mathcal{I}_\alpha \subset \mathcal{F}$ represents the “receptive field” of an element $\alpha \in X$, namely the set of other elements $\beta \in X$ whose values α must consult in determining its own value. We impose the restriction that the information field \mathcal{I}_α respect the Alexandroff topology on X , so that $\mathcal{I}_\alpha \subset \mathcal{F}(U_\alpha)$, where U_α is the minimal basic open set associated with element $\alpha \in X$.

Following structural causal models (Pearl, 1989), we can decompose the elements of the topological space into disjoint subsets $X = U \sqcup V$, where U represents “exogenous” variables that have no parents, namely α is exogenous precisely when $\mathcal{I}_\alpha \subset \mathcal{F}(\emptyset)$, and V are “endogenous” variables whose values are defined by measurable functions over exogenous and endogenous variables. Note that the probability space can be defined over the “exogenous” variables $\alpha \in U$, in which case it is convenient to attach a local probability space $(\Omega_\alpha, \mathcal{B}_\alpha, P)$ to each exogenous variable, where $\mathcal{B}_\alpha \subset \mathcal{B}$. We define conditional independence with respect to the induced information fields over the open sets of the Alexandroff space.

Definition 8. Given the induced probability space over information fields in a topological finite space, a stochastic basis is a sequence of information fields $\mathcal{G} = \mathcal{I}_1, \dots, \mathcal{I}_n$ such that for $1 \leq i \leq n-1$, $\mathcal{I}_i \subset \mathcal{I}_{i+1}$, and $\bigcup_{i=1}^n \mathcal{I}_i = \mathcal{F}$. Two such sequences \mathcal{G}_1 and \mathcal{G}_2 are **conditionally independent** given the base sigma algebra \mathcal{F} , if for all subsets $A \in \mathcal{G}_1$, $B \in \mathcal{G}_2$, it follows that $P(A|B|\mathcal{F}) = P(A|\mathcal{F})P(B|\mathcal{F})$.

Definition 9. The **decision field** $U = \prod_{\alpha \in X} U_\alpha$ defines the space of all possible values of the variables in the finite space causal model, where the cartesian product is interpreted as a map $u : X \rightarrow \bigcup_{\alpha \in X} U_\alpha$ such that $u(\alpha) \equiv u_\alpha \in U_\alpha$.

Definition 10. For any subset of elements $B \in X$, let P_B denote the projection of the product $\prod_{\alpha \in X} U_\alpha$ upon the product $\prod_{\beta \in B} U_\beta$, that is $P_B(u)$ is simply the restriction of u to the domain B .

Definition 11. The **product sigma field** is defined as $\prod_{\alpha \in B} \mathcal{F}_\alpha$ over $\prod_{\alpha \in B} U_\alpha$, where $\mathcal{F}(B)$ is the smallest sigma-field such that P_B is measurable. Note that if $B_1 \subset B_2$, then $\mathcal{F}_{B_1} \subset \mathcal{F}_{B_2}$. The finest sigma-field $\mathcal{F}(X) = \prod_{\alpha \in X} \mathcal{F}_\alpha$.

Definition 12. A finite space causal model \mathcal{M} is **causally faithful** with respect to the probability distribution P over \mathcal{M} if every conditional independence in the topology, as defined in Section 2.3, is satisfied by the distribution P , and vice-versa, every conditional independence property of the P is satisfied by the topology.

We can now formally define what it means to “solve” a causal finite space model \mathcal{M} . We impose the requirement that each variable $\alpha \in X$ must compute its value using a function measurable on its own information field.

Definition 13. Let the policy function f_α of each element $\alpha \in X$ be constrained so that $f_\alpha : U \times \Omega \rightarrow U_\alpha$ is measurable on the product sigma field $\mathcal{I}_\alpha \times \mathcal{B}_\alpha$, namely $f_\alpha^{-1}(\mathcal{F}_\alpha) \subset \mathcal{I}_\alpha \times \mathcal{B}_\alpha$.

Definition 14. The finite space causal model $\mathcal{M} = (X, U_\alpha, \mathcal{F}_\alpha, \mathcal{I}_\alpha)$ is **measurably solvable** if for every $\omega \in \Omega$, the closed loop equations $P_\alpha(u) = f_\alpha(u, \omega)$ have a unique solution for all $\alpha \in X$, where for a fixed $\omega \in \Omega$, the induced map $\mathcal{M}^\omega : \Omega \rightarrow U$ is a measurable function from the measurable space (Ω, \mathcal{B}) into (U, \mathcal{F}) .

Definition 15. The finite space causal model $\mathcal{M} = (X, U_\alpha, \mathcal{F}_\alpha, \mathcal{I}_\alpha)$ is **stable** if for every $\omega \in \Omega$, the closed loop equations $P_\alpha(u) = f_\alpha(u, \omega)$ are solvable by a fixed constant ordering Ξ that does not depend on $\omega \in \Omega$.

Measurably solvable models capture the corresponding property in a structural causal model (U, V, F, P) , which states that for any fixed probability distribution P defined over the exogenous variables U , each function f_i computes the value of variable $x_i \in V$, given the value of its parents $Pa(x_i)$ uniquely as a function of $u \in U$. This allows defining the induced distribution $P_u(V)$ over exogenous variables in a unique functional manner depending on some particular instantiation of the random exogenous variables U . Stable models are those where the ordering of variables is fixed. We now extend the notion of *recursive* causal models in DAGs (Pearl, 2009) to finite topological spaces.

Definition 16. The finite space causal model $\mathcal{M} = (X, U_\alpha, \mathcal{F}_\alpha, \mathcal{I}_\alpha)$ is a **recursively causal model** if there exists an ordering function $\psi : X \rightarrow \Xi_n$, where Ξ_n is the set of all injective (1-1) mappings of $(1, \dots, n)$ to the set X , such that for any $1 \leq k \leq n$, the information field of variable α_k in the ordering Ξ_n is contained in the joint information fields of the variables preceding it:

$$\mathcal{I}_{\alpha_k} \subset \mathcal{F}(\alpha_1, \dots, \alpha_{k-1}) \quad (1)$$

Definition 17. A **causal intervention** $do(\beta=u_\beta)$ in a finite space topological model $\mathcal{M} = (X, U_\alpha, \mathcal{F}_\alpha, \mathcal{I}_\alpha)$ is defined as the submodel \mathcal{M}_β whose information fields \mathcal{I}_α are exactly the same as in \mathcal{M} for all elements $\alpha \neq \beta$, and the information field of the intervened element β is defined to be $\mathcal{I}_\beta \subset \mathcal{F}(\emptyset) \times \mathcal{B}_\beta$. Note that since the only measurable function on $\mathcal{F}(\emptyset)$ is the constant function, whose value depends on a random sample space element $\omega \in \Omega_\beta$, this generalizes the notion of causal intervention in DAGs, where an intervened node has all its incoming edges deleted.²

3.1 Embedding Causal Graphical Models into Finite Topological Spaces

We now explain how to construct faithful topological embeddings of causal graphical models. The following lemma plays a fundamental role in constructing topologically faithful embeddings of graphical models.

Lemma 11. (Barmak, 2011; May) A preorder $\{X, \leq\}$ determines a topology \mathcal{U} on space X with the basis given by the collection of open sets $U_x = \{y \mid y \leq x\}$. It is referred to as the **order topology** on X . The space (X, \mathcal{U}) is a T_0 space if and only if (X, \leq) is a partially ordered set (poset). As before, we can alternatively characterize finite space topologies by the closed sets $F_x = \{y \mid y \geq x\}$.

The unique minimal basis gives us a way of characterizing whether or not a finite space has T_0 topology.

Lemma 12. (May) Two elements $x, y \in X$ have the same neighborhoods if and only if $U_x = U_y$. Thus, a finite space X has T_0 topology if and only if $U_x = U_y$.

²Our definition of causal intervention differs from that proposed in causal information fields (Heymann et al., 2020), where additional intervention nodes were added to the model.

Table 1: Examples of finite Alexandroff spaces on $X = \{a, b, c\}$, and their equivalent graphical models. A proper open set is any set other than \emptyset or X (which are in any topology). P_n is a topology on a set of size n with only one proper open set. D_n is the discrete topology over n elements. $P_{m,n}$ are topologies where the proper open sets are all non-empty subsets of a subset of size m . The \cong equivalence relation is homotopy equivalence. See text for explanation.

Proper Open Sets	Name	T_0 ?	Connected?	Equivalent graphical model
All	D_3	yes	no	HEDG (hyper-edge over (a,b,c)) (Forre and Mooij, 2017)
b, c		yes	yes	DAG $b \rightarrow a, c \rightarrow a$ (collider over a)
$a, b, (a, b)$	$P_{2,3} \cong \mathbb{C}D_2$	yes	yes	Chain graph: $a \rightarrow c, b \rightarrow c, a - b$
$a, b, (a, b), (b, c)$	$D_1 \sqcup P_2$	yes	no	DAG with node a disconnected, $b \rightarrow c$
a	P_3	no	yes	Chain graph: $a \rightarrow b, a \rightarrow c$, and $b - c$

Proof: If x and y have the same neighborhoods, then trivially $U_x = U_y$. Conversely, if $U_x = U_y$, then if $x \in U$ for some open set U , then $U_y = U_x \subset U$ (recall that U_x is the intersection of all sets that contain x), and hence $y \in U$. Similarly, if $y \in U$, the same argument shows $x \in U$. Thus, x and y have the same neighborhoods. \square

We state three theorems that show how to reduce several popular causal models into their faithful topological embedding. The same construction can be followed for all the other models in the literature as well. We focus on embedding the topology of a graphical model, leaving aside the parametric specification of a probability measure on the model (which we discuss in more depth in the appendix).

Theorem 2. *Every causal DAG graphical structure $G = (V, E)$ defines a finite T_0 Alexandroff topological space with a partial ordering.*

Proof: Define the elements of the topology $X = V$, the vertices representing the variables of the DAG $G = (V, E)$. Construct the transitive closure G_{tc} of the DAG G . Define the partial ordering $x \leq y$ in the topological space if the variable x is a descendant of y in G_{tc} . Define the open sets of X as $U_y = \{x | x \leq y\}$. \square

Theorem 3. *Every chain graph (Lauritzen and Richardson, 2002) structure $G = (V, E)$ defines a finite Alexandroff topological space X with a \leq preordering.*

Proof: Once again, define X as the variables in the chain graph. Recall that in a chain graph $G = (V, E)$, two nodes x and y are connected by an edge that is either directed, so $x \leftarrow y$ or $x \rightarrow y$, or there is an undirected edge $x - y$ between them. Define the ordering $x \leq y$ on the topology X if and only if there exists a path from x to y such that every comparable pair of nodes on this path is either of the form $x_k \leftarrow x_{k+1}$ or alternatively $x_k - x_{k+1}$. This ordering \leq on X is a preordering, and hence defines a general Alexandroff finite space topology. Define the open sets of X as $U_x = \{y | y \leq x\}$. \square

Theorem 4. *Every mDAG graphical model (Evans, 2018) or HEDG hyper-edge directed graphical model (Forre and Mooij, 2017) $G = (V, E, H)$, where H is a set of hyper-edges, represented by an abstract simplicial complex, can be represented by a finite Alexandroff topological space X with a \leq preordering.*

Proof: Define the space X by the variables in the graphical model $G = (V, E, H)$. For the observable edges represented by E , we follow the same construction as in DAG models described above. Note the hyper-edges $h \in H$ in effect represent an abstract simplicial complex. For example, in Table 4, the D_3 discrete topology on $X = \{a, b, c\}$ can be represented as an mDAG model where the three observable variables a, b, c are connected only through one latent variable, whose effect on the observable variables is manifested by the hyper-edge that constitutes a simplicial complex C defined by the non-empty power set of X . This simplicial complex C can be modeled as a non-Hausdorff cone \mathbb{C} between the latent variable and the open set topology of the observable variables (see Section 3.2 below). \square

3.2 Combining Poset Models

A crucial strength of our topological framework is the ability to combine two topological spaces X and Y into a new space, which can generate a rich panoply of models. Here are a few of the myriad ways in which topological spaces can be combined (Munkres, 1984a). Table 1 illustrates some of these ways of combining spaces for a small finite space X comprised of just three elements.

- Subspaces: The **subspace** topology on $A \subset X$ is defined by the set of all intersections $A \cap U$ for open sets U over X .
- Quotient topology: The **quotient topology** on U defined by a surjective mapping $q : X \rightarrow Y$ is the set of subsets U such that $q^{-1}(U)$ is open on X .
- Union: The **topology of the union** of two spaces X and Y is given by their disjoint union $X \sqcup Y$, which has as its open sets the unions of the open sets of X and that of Y .

Table 2: Enumeration of Alexandroff finite space topologies (see (May)).

n	Distinct	Distinct T_0	Inequivalent	Inequivalent T_0
1	1	1	1	1
2	4	3	3	2
3	29	19	9	5
4	355	219	33	16
5	6942	4231	139	63
6	209,527	130,023	718	318
7	9,535,241	6,129,859	4,535	2,045
8	642,779,354	431,723,379	35,979	16,999
9	63,260,289,423	44,511,042,511	363,083	183,231
10	8,977,053,873,043	6,611,065,248,783	4,717,687	2,567,284

- Product of two spaces: The **product topology** on the cartesian product $X \times Y$ is the topology with basis the "rectangles" $U \times V$ of an open set U in X with an open set V in Y .
- Wedge sum of two spaces: The *wedge sum* is the "one point" union of two "pointed" spaces (X, x_0) with (Y, y_0) , defined by $X \vee Y/x_0 \sim y_0$, the quotient space of the disjoint union of X and Y , where x_0 and y_0 are identified.
- Smash product: The *smash product* topology is defined as the quotient topology $X \wedge Y = X \times Y/X \vee Y$.
- Non-Hausdorff cone: The **non-Hausdorff cone** of topological space X with $Y = \{*\}$ yields the new space $\mathbb{C}(X)$, whose open sets are now $\mathcal{O}_{\mathbb{C}(X)} = \mathcal{O}_X \cup \{X \cup \{*\}\}$.
- Non-Hausdorff suspension: The **non-Hausdorff suspension** of topological space X with $Y = \{+, -\}$ yields the new space $\mathbb{S}(X)$, whose open sets are now $\mathcal{O}_{\mathbb{S}(X)} = \mathcal{O}_X \cup \{X \cup \{+, -\}\}$.

3.3 Homeomorphisms and Homotopical Equivalences

Like the number of possible DAG structures, the number of possible finite space topologies grows extremely rapidly. However, there are powerful tools in algebraic topology, such as homotopies, which characterize equivalences among spaces (Munkres, 1984a). In particular, Table 2 shows that exploiting homotopical inequivalences, we can save over three orders of magnitude in searching for an appropriate poset model over naive search. Given that evolutionary processes, such as pancreatic cancer, may potentially involve multiple thousands of elements (genes), the savings may be very significant. Of course, it is crucial to combine the savings from domain knowledge, as provided in Table 5 and Table 4 with that provided by efficient enumeration of poset topologies under homomorphie equivalences.

Definition 18. A topological space X is **contractible** if the identity map $id_X : X \rightarrow X$ is homotopically equivalent to the constant map $f(x) = c$ for some $c \in X$.

For example, any convex subset $A \subset \mathbb{R}^n$ is contractible. Let $f(x) = c, c \in A$ be the constant map. Define the homotopy $H : A \times I \rightarrow X$ as equal to $H(x, t) = tc + (1 - t)x$. Note that at $t = 0$, we have $H(x, 0) = x$, and that at $t = 1$, we have $H(x, 1) = c$, and since A is a convex subset, the convex combination $tc + (1 - t)x \in A$ for any $t \in [0, 1]$.

Theorem 5. If X is a finite topological space containing a point y such that the only open (or closed) subset of X containing y is X itself, then X is contractible. In particular, the non-Hausdorff cone $\mathbb{C}(X)$ is contractible for any X .

Proof: Let $Y = \{*\}$ denote the space with a single element, $*$. Define the retraction mapping $r : X \rightarrow *$ by $r(x) = *$ for all $x \in X$, and define the *inclusion* mapping $i : Y \rightarrow X$ by $i(*) = y$. Clearly, $r \circ i = id_Y$. Define the homotopy $h : X \rightarrow I \rightarrow X$ by $h(x, t) = x$ if $t < 1$, and $h(x, 1) = y$. Then, h is continuous, because for any open set U in X , if $y \in U$, then clearly $U = X$ (as X is the only open set containing y), and hence $h^{-1}(U) = X \times I$, which is open. If on the other hand, $y \notin U$, then $h^{-1}(U) = U \times [0, 1)$. It follows that h is a homotopy $h \cong id_X = i \circ r$. \square

The following lemma is of crucial importance in Section 4, where we will define *beat points*, elements of a topological space that can be removed, reducing model size.

Definition 19. A point x in a finite Alexandroff topological space X is **maximal** if there is no $y > x$, and **minimal** if there is no $y < x$.

Lemma 13. If X is an finite Alexandroff space, then U_x is contractible. In particular, if X has a unique maximal point or unique minimal point, then X is contractible.

Definition 20. Let $f, g : X \rightarrow Y$ be two continuous maps between finite space topologies X and Y . We say f is homotopic to g , denoted as $f \cong g$ if there exists a continuous map $h : X \times [0, 1] \rightarrow Y$ such that $h(x, 0) = f(x)$ and

$h(x, 1) = g(x)$. In other words, there is a smooth “deformation” between f and g , so we can visualize f being slowly warped into g . Note that \cong is an equivalence relation, since $f \cong f$ (reflexivity), and if $f \cong g$, then $g \cong f$ (symmetry), and finally $f \cong g, g \cong h \implies f \cong h$ (transitivity).

Definition 21. A map $f : X \rightarrow Y$ is a homotopy equivalence if there exists another map $g : Y \rightarrow X$ such that $g \circ f \cong id_X$ and $f \circ g \cong id_Y$, where id_X and id_Y are the identity mappings on X and Y , respectively.

4 Algorithms for Learning Causal Posets

In this section, we describe a number of algorithms for constructing causal poset models. Our approach is intended to highlight the important role played by topological constraints, which were implicit in many previous studies. We use the domain of cancer genomics to illustrate how topological constraints on datasets makes it possible to efficiently learn the structure and parameters of a causal model from observational data (Beerenwinkel and Sullivan, 2009; Beerenwinkel et al., 2007, 2006; Gerstung et al., 2011). A greedy algorithm for learning causal poset models is described in (Bernstein et al., 2020), but it does not fully exploit the algebraic topology of posets for efficient enumeration. We show that the space of causal structures can be significantly pruned by exploiting the algebraic topology of finite spaces, in particular using homeomorphisms among topologically equivalent finite space models.

4.1 Intervention Topologies

First, we show that many previous studies of causal discovery from interventions, including the *conservative* family of intervention targets (Hauser and Bühlmann, 2012a), path queries (Bello and Honorio, 2018), and *separating systems* of finite sets or graphs. (Eberhardt, 2008; Hauser and Bühlmann, 2012a; Kocaoglu et al., 2017; Katona, 1966; Mao-cheng, 1984) can all be viewed as imposing an intervention topology. Table 3 classifies a few previous studies in terms of the induced intervention topology.

If no experiments are allowed, the intervention topology is simply the trivial topology. If $X = I \sqcup X \setminus I$, where I is the intervention target (Eberhardt, 2008), then the intervention topology is disconnected. (Shpitser and Tchetgen, 2016) study single node interventions, which can be viewed as a T_1 intervention topology where singleton sets are closed. (Hauser and Bühlmann, 2012a) introduce the idea of *conservative* family \mathcal{I} of intervention targets, meaning a family of (open) subsets of variables in a causal model such that for every variable $x \in X$, there exists an $I \subset \mathcal{I}$ such that $x \notin I$. This is closely related to the idea of Alexandroff topologies where elements have distinguishable neighborhoods, and each intervention target $I \in \mathcal{I}$ defines a neighborhood. A very related notion is that of *separating systems* of finite sets as intervention targets (Eberhardt, 2008; Hauser and Bühlmann, 2012a; Kocaoglu et al., 2017). or separating systems of graphs (Mao-cheng, 1984; Hauser and Bühlmann, 2012b). (Kocaoglu et al., 2017) used *antichains*, a partitioning of a poset into subsets of non-comparable elements. (Bello and Honorio, 2018) use path queries, which can be viewed as chains. Finally, (Tadepalli and Russell, 2021) used leaf queries on tree structures, where none of the interior nodes can be intervened on.

We first introduce the notion of a *separating system*, which is a special case of the T_0 topology separation axiom of finite Alexandroff spaces.

Definition 22. A separating system on a finite set X is a collection of subsets $\{\mathcal{U}_1, \dots, \mathcal{U}_m\}$ such that for every pair of elements $x, y \in X$, there is a set \mathcal{U}_k such that either $x \in \mathcal{U}_k, y \notin \mathcal{U}_k$ or alternatively, $x \notin \mathcal{U}_k, y \in \mathcal{U}_k$. An (m, n) strongly separating system is a pair of sets $\mathcal{U}_i, \mathcal{U}_j$ such that $x \in \mathcal{U}_i, y \notin \mathcal{U}_j$ and $x \notin \mathcal{U}_i, y \in \mathcal{U}_j$.

Definition 23. Given a finite space Alexandroff topology (X, \mathcal{T}) , the T_0 -topogenous matrix A (Shiraki, 1969) associated with it is defined as the $m \times n$ binary matrix defined as $A(i, j) = 1$ if $x_j \in U_i$, and $A(i, j) = 0$ otherwise. In words, each row defines the open sets that element j belongs to, and each column defines the elements that are contained in open set U_i .

Table 3: Intervention topologies of some previous causal discovery methods.

Intervention Sets	Topology	Intervention Class	Causal Structure	Reference
$\{\emptyset, X\}$	Trivial	Observations	Conjunctive Bayes Network	(Beerenwinkel et al., 2006)
$I \sqcup X \setminus I$	Disconnected	All subsets	DAG	(Eberhardt, 2008)
$F_x = \{x\}$	T_1	Single-node	Tree graphs	(Shpitser and Tchetgen, 2016)
Strong separating sets	T_2	Restricted open sets	Latent DAG	(Kocaoglu et al., 2017)
$\{x\}$	T_1	Path queries	Transitive DAG	(Bello and Honorio, 2018)
Minimal elements	Restricted open sets	Leaf queries	Tree Graphs	(Tadepalli and Russell, 2021)

Algorithm 1: Learn Causal Poset using Interventions and Conditional Independence Oracle.

Input: Dataset $\mathcal{D} = (\mathcal{E}, \mathcal{T})$ of a finite set of events \mathcal{E} and their associated *open sets* $U \in \mathcal{T}$, represented as a topological space of open sets $g \subset \mathcal{E}$. A conditional independence oracle is also assumed.

Output: Causally faithful poset \mathcal{P} that is consistent with conditional independences in data.

begin

Set the basic closed sets $F_e \leftarrow \emptyset$

repeat

Select an open separating set $g \in \mathcal{T}$ and intervene on g .

for $e \in g, f \notin g$ **do**

Use samples and the CI oracle to test if $(e \perp\!\!\!\perp f)_{M_g}$ on dataset \mathcal{D} .

If CI test fails, then set $F_e \leftarrow F_e \cup \{f\}$ because f is an ancestor of e .

end

until convergence;

Define the relation $e \leq f$ if $f \in F_e$, for all $e, f \in \mathcal{E}$, and compute its transitive closure.

Return the poset $\mathcal{P} = (\mathcal{E}, \leq)$, where \leq is the induced relation on the poset \mathcal{P} .

end

Theorem 6. Given a finite space Alexandroff topology (X, \mathcal{T}) , the T_0 -topogenous matrix A associated with it defines a separating set.

Proof: Note that in a T_0 finite space Alexandroff topology, each element is distinguished by a unique neighborhood. Consequently, the open sets U_x and U_y associated with $x, y \in X$ must be distinct, and if $U_x = U_y$, then trivially $x = y$. Consequently, the topogenous matrix defines a separating set for the topology (X, \mathcal{T}) . Similarly, the closed sets F_x in Algorithm 1 also define separating sets. \square

4.2 Learning Causal Poset from Interventions

Algorithm 2: Given a causally faithful T_0 topological model $(X, \leq, \mathcal{T}$ and access to a conditional independence testing oracle, compute causal (observable) DAG model $G = (V, E)$.

Input: T_0 causally faithful topology $\mathcal{M} = (X, \leq, \mathcal{T})$ with \leq defining a transitively closed partial ordering sets.

Output: Causal DAG model $G = (V, E)$ on observable nodes.

begin

Compute the *antichain* sets T_i induced by partial ordering \leq . The set of open (or closed) sets in \mathcal{M} are in 1–correspondence with the antichain sets.

Set $E \leftarrow \emptyset$.

repeat

Intervene on an antichain set T_i .

for $x \in T_i, y \notin U$ **do**

Use samples and the CI oracle to test if $(x \perp\!\!\!\perp y)_{M_{T_i}}$ on a dataset.

If CI test fails, then set $E \leftarrow E \cup (x, y)$.

end

until convergence;

Set $V = X$, and return observable causal DAG $G = (V, E)$

end

Theorem 7. Algorithm 1 requires only $|\mathcal{T}|$ interventions and conditional independence tests on samples obtained from each post-interventional distribution, to find a statistically consistent T_0 topological model. If there are $O(\log n)$ separating sets, the algorithm requires $O(\log n)$ interventions.

Proof: If we intervene on the separating open set U and find an element $y \notin F_x$ that is statistically dependent on x , then we change F_x to include y (since x needs to “consult” y in determining its value). The bound $O(\log n)$ in (Kocaoglu et al., 2017) assumes that there are up to $2 \log n$ sets in the original separating system. It has been argued in (Bello and Honorio, 2018) that interventions on multiple variables, such as used here and in the previous work (Kocaoglu et al., 2017) can potentially require an exponential number of experiments, if for example a separating set has $O(n)$ distinct elements, and each node is a binary variable, which requires two experiments (setting it to both 0 and 1). There is an inherent trade off between the size of each separating set, and the number of separating sets. \square

Algorithm 2 is a generalization of Algorithm 1 in the recent paper by (Kocaoglu et al., 2017), who construct a DAG by doing interventions on the *antichains* of posets, and extend this approach to discover causal models with latent variables as well. (Acharya et al., 2018) propose a related approach for inferring causal models, which does not require conditional independence testing, but uses a sample efficient testing methodology based on squared Hellinger distances.

Algorithm 2 constructs the observable DAG model, based on *antichain* sets, namely the set of incomparable elements at each level of the partial ordering. Antichain sets can be shown to be in bijective correspondence with the open sets on an Alexandroff T_0 topology.

Theorem 8. Mirsky’s theorem (*mir, 1971*): *The height of a T_0 topology causal model (X, \leq, \mathcal{T}) is defined to be the maximum cardinality of a chain, a totally ordered subset of the given partial order. For every partially ordered T_0 causal model (X, \leq, \mathcal{T}) , the height also equals the minimum number of **antichains**, namely subsets in which no pair of elements are ordered, into which the set may be partitioned.*

Theorem 9. *Algorithm 2 requires $O(h)$ interventions and conditional independence tests on samples obtained from the post-interventional distributions, where h is the height h of a T_0 topology causal model (X, \leq, \mathcal{T}) .*

Note that Algorithms 1 and 2 are topological generalizations of Algorithm 1 and 2 in (Kocaoglu et al., 2017). The remaining Algorithms 3 and 4 in (Kocaoglu et al., 2017) on learning a latent variable DAG model can be generalized as well (see the appendix).

4.3 Efficient Enumeration of Homeomorphically Distinct Posets

Next, we turn to the fundamental problem of how to efficiently enumerate posets, which is a key requirement for scaling many causal discovery algorithms (Kocaoglu et al., 2017; Acharya et al., 2018; Bernstein et al., 2020; Beerenwinkel et al., 2006).

Definition 24. *For every T_0 finite space model M with a partial ordering \leq , define its associated **Hasse diagram** H_M as a directed graph which captures all the relevant order information of M . More precisely, the vertices of H_M are the elements of M , and the edges of H_M are such that there is a directed edge from x to y whenever $y \leq x$, but there is no other vertex z such that $y \leq z \leq x$.*

General pre-ordered finite spaces can be reduced to partially ordered T_0 topologies up to homeomorphic equivalence.

Theorem 10. (*Stong, 1966*) *Let (X, \mathcal{T}) be an arbitrary finite space model with an associated preordering \leq . Let X_0 represent the quotient topological space X/\sim , where $x \sim y$ if $x \leq y$ and $y \leq x$. Then X_0 is a homotopically equivalent topological model with T_0 separability, and the quotient map $q : X \rightarrow X_0$ is a homotopy equivalence. Furthermore, X_0 induces a partial ordering on the elements $x \in X_0$.*

A key idea in the enumeration is to assume that each element in the Hasse diagram of the poset does not have an in-degree or out-degree of 1.

Definition 25. (*Stong, 1966*) *An element $x \in X$ in a finite T_0 space X is a **down beat point** if x covers one and only one element of X . Alternatively, the set $\hat{U}_x = U_x \setminus \{x\}$ has a (unique) maximum. Similarly, $x \in X$ is an **up beat point** if x is covered by a unique element, or equivalently if $\hat{F}_x = F_x \setminus \{x\}$ has a (unique) minimum. A **beat point** is either a down beat or up beat point.*

Theorem 11. (*Stong, 1966*) *Let X be a finite T_0 topological model, and let $x \in X$ be a (down, up) beat point. Then the reduced model $X \setminus \{x\}$ is a **strong deformation retract** of X .³ A point x in a finite space M is an **upbeat point** if and only if it has in-degree one in the associated Hasse diagram H_M , i.e., it has only one incoming edge). Similarly, x is **downbeat** if and only if it has out-degree one (it has only one outgoing edge).*

Definition 26. *A finite T_0 topological space is a **minimal model** if it has no beat points. A **core** of a finite topological space X is a strong deformation retract, which is a minimal finite space. The **minimal graph** of a minimal model is its equivalent Hasse diagram.*

Theorem 12. (*Stong, 1966*) **Classification Theorem:** *A homotopy equivalence between minimal finite space topological models is a homeomorphism. In particular, the core of a finite space model is unique up to homeomorphism and two finite spaces are homotopy equivalent if and only if they have homeomorphic cores.*

Figure 4 illustrates the process of removing beat points to construct the minimal poset. b is an up beat point of X , c is an upbeat point of $X \setminus \{b\}$, and e is an up beat point of $X \setminus \{b, c\}$. Similarly, points c and e are removed, resulting in the minimal poset. The figure also shows that homeomorphic equivalences greatly reduces the search space of possible structures. Note the plot is on log scale. For example, for 12 variables, the number of minimal posets is $< 0.1\%$ of the number of possible posets, a savings of three orders of magnitude.

³In algebraic topology, a subspace $A \subset X$ is called a strong deformation retract of X if there is a homotopy $F : X \times [0, 1] \rightarrow A$ such that $F(x, 0) = x$, $F(x, 1) \in A$, $F(a, t) = a$ for all $x \in X, t \in [0, 1], a \in A$. (Munkres, 1984b).

Figure 4: Left: Constructing minimal posets by removing beat points (Barmak, 2011; Stong, 1966). Right: Efficiently enumerating minimal posets (Fix and Partias; Brinkmann and McKay, 2002).



Algorithm 3: Find Topologically Minimal T_0 Causal Model.

Input: General pre-ordered causal model, such as a chain graph $G = (V, E)$ that is causally faithful to a dataset.
Output: Minimal T_0 causal model homotopically equivalent to original non- T_0 model (e.g., from a chain graph G).
 The algorithm uses homotopy theory to find the *core* T_0 model of a general chain graph.

```

begin
    Define the topological model  $(X, \mathcal{U})$  where  $X = V$  and the open sets in  $\mathcal{U}$  are constructed from the induced
    pre-order  $\leq$  from  $G$ . Define the minimal model  $(X_0, \mathcal{U}')$ , and set  $X_0 = X$ .
    repeat
        for  $x, y \in X_0$  s.t.  $x \leq y, y \leq x$  do
            Remove  $x, y$  from  $X_0$ , and replace them with a new variable  $z = x \sim y$ .
            Set  $X_0 \leftarrow X_0 \setminus \{x, y\} \cup \{z\}$ .  $z$  represents the equivalence class that includes  $x$  and  $y$ .
        end
        for  $x \in X_0$  do
            Remove down beat points: If  $\hat{U}_x = U_x \setminus \{x\}$  has a maximum, then  $X_0 \leftarrow X_0 \setminus \{x\}$ .
            Remove up beat points: If  $\hat{F}_x = F \setminus \{x\}$  has a minimum, then  $X_0 \leftarrow X_0 \setminus \{x\}$ .
        end
    until convergence;
    Define the open sets  $U_x \in \mathcal{U}'$  as  $U_x = \{y \mid y \leq x\}$  for  $x \in X_0$ .
end
    
```

Algorithm 3 determines a quotient T_0 topology that is homotopically equivalent to the general non- T_0 topology defined by a chain graphical models. Second, the algorithm further reduces the model to its core by removing *beat points* (Barmak, 2011; May; McCord, 1966; Stong, 1966).

5 Bioinformatics application

Table 5 shows a small fragment of a dataset for pancreatic cancer (Jones et al., 2008). Like many cancers, it is marked by a particular partial ordering of mutations in some specific genes, such as **KRAS**, **TP53**, and so on. In order to understand how to model and treat this deadly disease, it is crucial to understand the inherent partial ordering in the mutations of such genes. Pancreatic cancer remains one of the most prevalent and deadly forms of cancer. Roughly half a million humans contract the disease each year, most of whom succumb to it within a few years.⁴ Figure 5 shows the roughly 20 most common genes that undergo mutations during the progression of this disease. The most common gene, the KRAS gene, provides instructions for making a protein called K-Ras that is part of a signaling pathway known as the RAS/MAPK pathway. The protein relays signals from outside the cell to the cell’s nucleus. The second most common mutation occurs in the TP53 gene, which makes the p53 protein that normally acts as the supervisor in the cell as the body tries to repair damaged DNA. Like many cancers, pancreatic cancers occur as the normal reproductive machinery of the body is taken over by the cancer.

In the pancreatic cancer problem, for example, the topological space X is comprised of the significant events that mark the progression of the disease, as shown in Table 5. In particular, the table shows that specific genes are mutated at

⁴Sadly, this disease killed the much admired and long-time host of Jeopardy, Alex Trebek, last year.

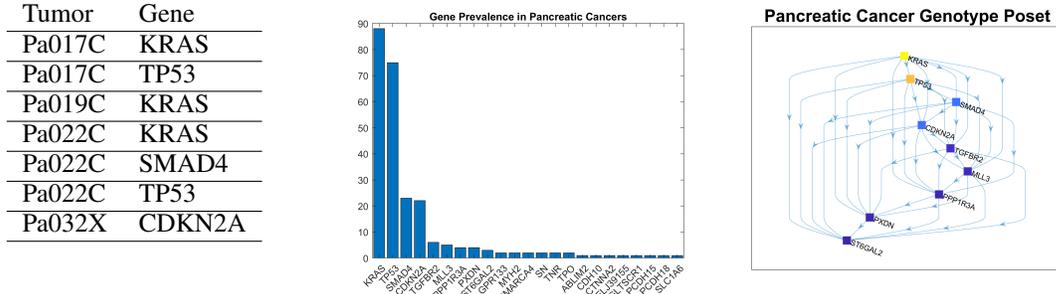


Figure 5: Left: Genetic mutations in pancreatic cancer (Jones et al., 2008). Middle: histogram of genes sorted by mutation frequencies. Right: Poset learned from dataset.

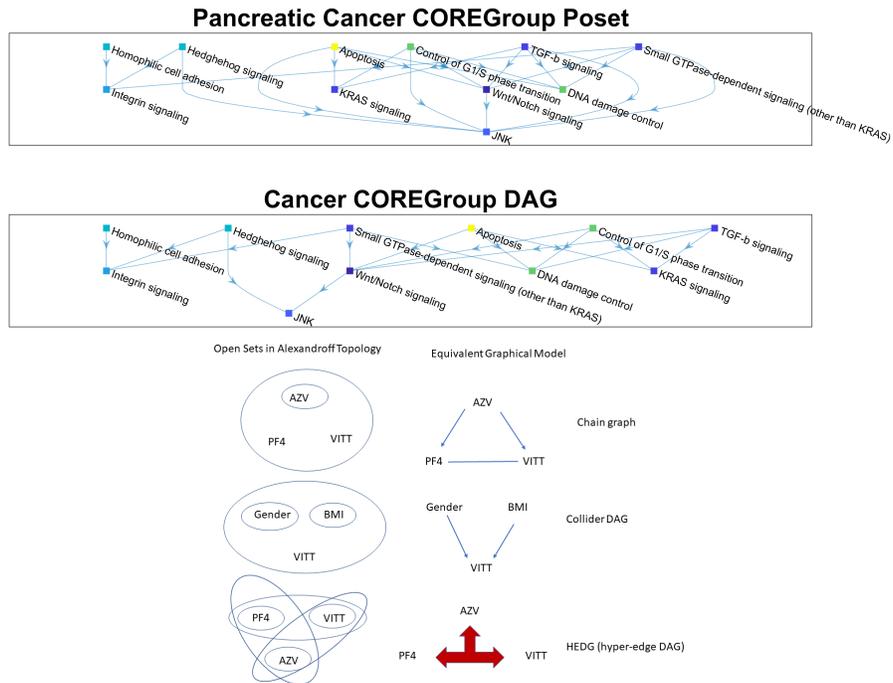


Figure 6: Top: causal poset and DAG model of pathways in pancreatic cancer learned from a real-world dataset (Jones et al., 2008), showing genetic mutations occur along distinct pathways. Bottom: Topological representation of causal poset and DAG models for COVID-19 AZV (AstraZeneca vaccine), VITT (vaccine-induced clotting of blood), and other factors (Greinacher et al., 2021; Schultz et al., 2021).

specific locations by the change of an amino acid, causing the gene to malfunction. We can model a tumor in terms of its *genotype*, namely the subset of X , the gene events, that characterize the tumor. For example, the table shows the tumor **Pa022C** can be characterized by the genotype **KRAS**, **SMAD4**, and **TP53**. In general, a finite space topology is just the elements of the space (e.g. genetic events), and the subspaces (e.g., genomes) that define the topology.

We illustrate our framework using the problem of inferring topological causal models for cancer (Beerenwinkel and Sullivant, 2009; Beerenwinkel et al., 2007, 2006; Gerstung et al., 2011). The progression of many types of cancer are marked by mutations of key genes whose normal reproductive machinery is subverted by the cancer (Jones et al., 2008). Often, viruses such as HIV and COVID-19 are constantly mutating to combat the pressure of interventions such as drugs, and successful treatment requires understanding the partial ordering of mutations. A number of past approaches use topological separability constraints on the data, assuming observed genotypes separate events, which as we will show, is abstractly a separability constraint on the underlying topological space.

A key computational level in making model discovery tractable in evolutionary processes, such as pancreatic cancer, is that multiple sources of information are available that guide the discovery of the underlying poset model. In particular, for pancreatic cancer (Jones et al., 2008), in addition to the tumor genotype information show in Table 5, it is also

Table 4: Core signaling pathways and processes genetically altered in most pancreatic cancers (Jones et al., 2008).

Regulatory pathway	% altered genes	Tumors	Representative altered genes
Apoptosis	9	100%	CASP10, VCP, CAD, HIP1
DNA damage control	9	83%	ERCC4, ERCC6, EP300, TP53
G1/S phase transition	19	100%	CDKN2A, FBXW7, CHD1, APC2
Hedgehog signaling	19	100%	TBX5, SOX3, LRP2, GLI1, GLI3
Homophilic cell adhesion	30	79%	CDH1, CDH10, CDH2, CDH7, FAT
Integrin signaling	24	67%	ITGA4, ITGA9, ITGA11, LAMA1
c-Jun N-terminal kinase	9	96%	MAP4K3, TNF, ATF2, NFATC3
KRAS signaling	5	100%	KRAS, MAP2K4, RASGRP3
Regulation of invasion	46	92%	ADAM11, ADAM12, ADAM19
Small GTPase-dependent	33	79%	AGHGEF7, ARHGEF9, CDC42BPA
TGF- β signaling	37	100%	TGFBR2, BMPR2, SMAD4, SMAD3
Wnt/Notch signaling	29	100%	MYC, PPP2R3A, WNT9A

Algorithm 4: Application of Poset Discovery Algorithm to Bioinformatics.

Input: Dataset $\mathcal{D} = (\mathcal{E}, \mathcal{T})$ of a finite set of events \mathcal{E} and their associated *genotypes* $U \in \mathcal{T}$, represented as a topological space of open sets $g \subset \mathcal{E}$. Here, it is assumed that each genome is an intervention target, whose size will affect the complexity of each causal experiment. A conditional independence oracle is also assumed.

Output: Causally faithful poset \mathcal{P} that is consistent with conditional independences in data.

begin

Set the basic closed sets $F_e \leftarrow \emptyset$

repeat

Select an open separating set $g \in \mathcal{T}$ and intervene on g .

for $e \in g, f \notin g$ **do**

Use samples and the CI oracle to test if $(e \perp\!\!\!\perp f)_{M_g}$ on dataset \mathcal{D} .

If CI test fails, then set $F_e \leftarrow F_e \cup \{f\}$ because f is an ancestor of e .

end

until *convergence*;

Define the relation $e \leq f$ if $f \in F_e$, for all $e, f \in \mathcal{E}$, and compute its transitive closure.

Return the poset $\mathcal{P} = (\mathcal{E}, \leq)$, where \leq is the induced relation on the poset \mathcal{P} .

end

known that the disease follows certain pathways, as shown in Table 4. This type of information from multiple sources gives the ability to construct multiple posets that reflect different event constraints (Beerenwinkel et al., 2006).

Algorithm 4 is a generalization of past algorithms that infer conjunctive Bayesian networks (CBN) from a dataset of events (e.g., tumors or signaling pathways) and their associated genotypes (e.g., sets of genes) (Beerenwinkel et al., 2007, 2006) The pathway poset and DAG shown in Figure 1 and the poset in Figure 5 were learned using Algorithm 4 using the pancreatic cancer dataset published in (Jones et al., 2008).

5.1 Limitations and Future Work

We proposed a topological framework for causal discovery, building on the key relationship between posets and finite Alexandroff topologies. In the supplementary material, we elaborate on additional details. We gave some examples from the domain of cancer genomics. A growing body of work in causal discovery has implicitly used topological constraints to constrain search. Our paper uses insights from algebraic topology of finite spaces into developing more scalable algorithms. Our paper has a number of significant limitations. We did not discuss building poset models over latent variables, which is important in many applications. Furthermore, a deeper study of the empirical performance of the algorithms proposed here is necessary to fully evaluate the promise of the proposed framework.

References

- James R. Munkres. *Elements of algebraic topology*. Addison-Wesley, 1984a. ISBN 978-0-201-04586-4.
- David Luenberger. *Optimization in Vector Spaces*. John Wiley, 1997.
- George Cybenko. Approximation by superpositions of a sigmoidal function. *Math. Control. Signals Syst.*, 2(4):303–314, 1989. doi:10.1007/BF02551274. URL <https://doi.org/10.1007/BF02551274>.
- Herbert Edelsbrunner. An introduction to persistent homology. In Bruno Lévy and Dinesh Manocha, editors, *Proceedings of the 2007 ACM Symposium on Solid and Physical Modeling, Beijing, China, June 4-6, 2007*, page 9. ACM, 2007. doi:10.1145/1236246.1236249. URL <https://doi.org/10.1145/1236246.1236249>.
- Afra Zomorodian and Gunnar E. Carlsson. Computing persistent homology. *Discret. Comput. Geom.*, 33(2):249–274, 2005. doi:10.1007/s00454-004-1146-y. URL <https://doi.org/10.1007/s00454-004-1146-y>.
- Daphne Koller and Nir Friedman. *Probabilistic Graphical Models - Principles and Techniques*. MIT Press, 2009. ISBN 978-0-262-01319-2. URL <http://mitpress.mit.edu/catalog/item/default.asp?ttype=2&tid=11886>.
- S. Lauritzen. *Graphical Models*. Oxford University Press, 1996.
- Judea Pearl. *Probabilistic reasoning in intelligent systems - networks of plausible inference*. Morgan Kaufmann series in representation and reasoning. Morgan Kaufmann, 1989.
- Judea Pearl. The seven tools of causal inference, with reflections on machine learning. *Commun. ACM*, 62(3):54–60, 2019. doi:10.1145/3241036. URL <https://doi.org/10.1145/3241036>.
- Peter Spirtes, Clark Glymour, and Richard Scheines. *Causation, Prediction, and Search, Second Edition*. Adaptive computation and machine learning. MIT Press, 2000. ISBN 978-0-262-19440-2.
- Thomas S. Richardson. A factorization criterion for acyclic directed mixed graphs. In Jeff A. Bilmes and Andrew Y. Ng, editors, *UAI 2009, Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence, Montreal, QC, Canada, June 18-21, 2009*, pages 462–470. AUAI Press, 2009. URL https://dslpitt.org/uai/displayArticleDetails.jsp?mmnu=1&smnu=2&article_id=1659&proceeding_id=25.
- Robin J. Evans. Margins of discrete Bayesian networks. *The Annals of Statistics*, 46(6A):2623 – 2656, 2018. doi:10.1214/17-AOS1631. URL <https://doi.org/10.1214/17-AOS1631>.
- Patrick Forre and Joris M. Mooij. Markov properties for graphical models with cycles and latent variables, 2017.
- Steen A. Andersson, David Madigan, and Michael D. Perlman. An alternative markov property for chain graphs. In Eric Horvitz and Finn Verner Jensen, editors, *UAI '96: Proceedings of the Twelfth Annual Conference on Uncertainty in Artificial Intelligence, Reed College, Portland, Oregon, USA, August 1-4, 1996*, pages 40–48. Morgan Kaufmann, 1996. URL https://dslpitt.org/uai/displayArticleDetails.jsp?mmnu=1&smnu=2&article_id=350&proceeding_id=12.
- Jonathan A. Barmak. *Algebraic topology of finite topological spaces and applications*. Lecture notes in mathematics <Berlin>. Springer, Heidelberg ; Berlin u.a., 2011. URL <http://deposit.d-nb.de/cgi-bin/dokserv?id=3826587&prov=M&dok%5Fvar=1&dok%5Fext=htm>.
- P. S. Alexandroff. *Diskrete Räume. Rec. Math. [Mat. Sbornik] N.S.*, 2:501–518.
- P. S. Alexandroff. *Combinatorial topology. Vol. 1*. Graylock Press, 1956.
- J. P. May. Finite spaces and larger contexts. <https://math.uchicago.edu/~may/FINITE/FINITEBOOK/FINITEBOOKCollatedDraft.pdf>.
- 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. doi:<https://doi.org/10.1111/1467-9868.00340>. URL <https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/1467-9868.00340>.
- N. Beerenwinkel and S. Sullivant. Markov models for accumulating mutations. *Biometrika*, 96(3):645–661, 06 2009. ISSN 0006-3444. doi:10.1093/biomet/asp023. URL <https://doi.org/10.1093/biomet/asp023>.
- Ramon Diaz-Uriarte. Cancer progression models and fitness landscapes: a many-to-many relationship. *Bioinformatics*, 34(5):836–844, 10 2017. ISSN 1367-4803. doi:10.1093/bioinformatics/btx663. URL <https://doi.org/10.1093/bioinformatics/btx663>.
- Judea Pearl. *Causality: Models, Reasoning and Inference*. Cambridge University Press, USA, 2nd edition, 2009. ISBN 052189560X.
- Frederick Eberhardt. Almost optimal intervention sets for causal discovery. In David A. McAllester and Petri Myllymäki, editors, *UAI 2008, Proceedings of the 24th Conference in Uncertainty in Artificial Intelligence, Helsinki, Finland, July 9-12, 2008*, pages 161–168. AUAI Press, 2008. URL https://dslpitt.org/uai/displayArticleDetails.jsp?mmnu=1&smnu=2&article_id=1948&proceeding_id=24.

- Alain Hauser and Peter Bühlmann. Characterization and greedy learning of interventional markov equivalence classes of directed acyclic graphs. *J. Mach. Learn. Res.*, 13:2409–2464, 2012a. URL <http://dl.acm.org/citation.cfm?id=2503320>.
- Murat Kocaoglu, Karthikeyan Shanmugam, and Elias Bareinboim. Experimental design for learning causal graphs with latent variables. In Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pages 7018–7028, 2017. URL <https://proceedings.neurips.cc/paper/2017/hash/291d43c696d8c3704cbe0a72ade5f6c-Abstract.html>.
- CAI Mao-cheng. On separating systems of graphs. *Discrete Mathematics*, 49(1):15–20, 1984. ISSN 0012-365X. doi:[https://doi.org/10.1016/0012-365X\(84\)90146-8](https://doi.org/10.1016/0012-365X(84)90146-8). URL <https://www.sciencedirect.com/science/article/pii/S0012365X84901468>.
- Prasad Tadepalli and Stuart Russell. PAC learning of causal trees with latent variables. In *AAAI*, 2021.
- Daniel Bernstein, Basil Saeed, Chandler Squires, and Caroline Uhler. Ordering-based causal structure learning in the presence of latent variables. In Silvia Chiappa and Roberto Calandra, editors, *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pages 4098–4108. PMLR, 26–28 Aug 2020. URL <http://proceedings.mlr.press/v108/bernstein20a.html>.
- Michael C. McCord. Singular homology groups and homotopy groups of finite topological spaces. *Duke Mathematical Journal*, 33(3):465 – 474, 1966. doi:10.1215/S0012-7094-66-03352-7. URL <https://doi.org/10.1215/S0012-7094-66-03352-7>.
- R. E. Stong. Finite topological spaces. *Trans. Amer. Math. Soc.*, 123:325–340, 1966.
- Siân Jones, Xiaosong Zhang, D. Williams Parsons, Jimmy Cheng-Ho Lin, Rebecca J. Leary, Philipp Angenendt, Parminder Mankoo, Hannah Carter, Hirohiko Kamiyama, Antonio Jimeno, Seung-Mo Hong, Baojin Fu, Ming-Tseh Lin, Eric S. Calhoun, Mihoko Kamiyama, Kimberly Walter, Tatiana Nikolskaya, Yuri Nikolsky, James Hartigan, Douglas R. Smith, Manuel Hidalgo, Steven D. Leach, Alison P. Klein, Elizabeth M. Jaffee, Michael Goggins, Anirban Maitra, Christine Iacobuzio-Donahue, James R. Eshleman, Scott E. Kern, Ralph H. Hruban, Rachel Karchin, Nickolas Papadopoulos, Giovanni Parmigiani, Bert Vogelstein, Victor E. Velculescu, and Kenneth W. Kinzler. Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science*, 321(5897):1801–1806, 2008. ISSN 0036-8075. doi:10.1126/science.1164368. URL <https://science.sciencemag.org/content/321/5897/1801>.
- Niko Beerenwinkel, Nicholas Eriksson, and Bernd Sturmfels. Conjunctive Bayesian networks. *Bernoulli*, 13(4):893 – 909, 2007. doi:10.3150/07-BEJ6133. URL <https://doi.org/10.3150/07-BEJ6133>.
- Jayadev Acharya, Arnab Bhattacharyya, Constantinos Daskalakis, and Saravanan Kandasamy. Learning and testing causal models with interventions. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL <https://proceedings.neurips.cc/paper/2018/file/78631a4bb5303be54fa1cfdcb958c00a-Paper.pdf>.
- Andreas Greinacher, Thomas Thiele, Theodore E. Warkentin, Karin Weisser, Paul A. Kyrle, and Sabine Eichinger. Thrombotic thrombocytopenia after chadox1 ncov-19 vaccination. *New England Journal of Medicine*, 2021. doi:10.1056/NEJMoa2104840. URL <https://doi.org/10.1056/NEJMoa2104840>.
- Nina H. Schultz, Ingvild H. Sorvoll, Annika E. Michelsen, Ludvig A. Munthe, Fridtjof Lund-Johansen, Maria T. Ahlen, Markus Wiedmann, Anne-Hege Aamodt, Thor H. Skattor, Geir E. Tjonnfjord, and Pal A. Holme. Thrombosis and thrombocytopenia after chadox1 ncov-19 vaccination. *New England Journal of Medicine*, 2021. doi:10.1056/NEJMoa2104882. URL <https://doi.org/10.1056/NEJMoa2104882>.
- H. S. Witsenhausen. The intrinsic model for discrete stochastic control: Some open problems. In A. Bensoussan and J. L. Lions, editors, *Control Theory, Numerical Methods and Computer Systems Modelling*, pages 322–335, Berlin, Heidelberg, 1975. Springer Berlin Heidelberg. ISBN 978-3-642-46317-4.
- Benjamin Heymann, Michel de Lara, and Jean-Philippe Chancelier. Causal information with information fields. In *Neural Information Processing Systems Workshop on Causal Discovery and Causality-inspired Machine Learning*, 2020.
- Niko Beerenwinkel, Nicholas Eriksson, and Bernd Sturmfels. Evolution on distributive lattices. *Journal of Theoretical Biology*, 242(2):409–420, 2006. ISSN 0022-5193. doi:<https://doi.org/10.1016/j.jtbi.2006.03.013>. URL <https://www.sciencedirect.com/science/article/pii/S0022519306001159>.

- Moritz Gerstung, Nicholas Eriksson, Jimmy Lin, Bert Vogelstein, and Niko Beerenwinkel. The temporal order of genetic and pathway alterations in tumorigenesis. *PLOS ONE*, 6(11):1–9, 11 2011. doi:10.1371/journal.pone.0027136. URL <https://doi.org/10.1371/journal.pone.0027136>.
- Kevin Bello and Jean Honorio. Computationally and statistically efficient learning of causal bayes nets using path queries. In Samy Bengio, Hanna M. Wallach, Hugo Larochelle, Kristen Grauman, Nicolò Cesa-Bianchi, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada*, pages 10954–10964, 2018. URL <https://proceedings.neurips.cc/paper/2018/hash/a0b45d1bb84fe1bedbb8449764c4d5d5-Abstract.html>.
- Gyula Katona. On separating systems of a finite set. *Journal of Combinatorial Theory*, 2(1):174–194, 1966.
- Ilya Shpitser and Eric Tchetgen Tchetgen. Causal inference with a graphical hierarchy of interventions. *Annals of statistics*, 44(6):2433–2466, 2016.
- Alain Hauser and Peter Bühlmann. Two optimal strategies for active learning of causal models from interventions. *CoRR*, abs/1205.4174, 2012b. URL <http://arxiv.org/abs/1205.4174>.
- Mitsobu Shiraki. On Finite Topological Spaces II. *Math. Physics Chemistry*, 46(2):1–15, 1969.
- A dual of Dilworth’s decomposition theorem. *The American Mathematical Monthly*, 78(8):876–77, 1971.
- James R. Munkres. *Elements of algebraic topology*. Addison-Wesley, 1984b. ISBN 978-0-201-04586-4.
- Alex Fix and Stephen Partias. Enumeration of homotopy classes of finite t0 topological spaces. <http://math.uchicago.edu/~may/FINITE/REUPapers/Fix.pdf>.
- Gunnar Brinkmann and Brendan D. McKay. Posets on up to 16 points. *Order*, 19:147–179, 2002.