Classifying causal structures: Ascertaining when classical correlations are constrained by inequalities

Shashaank Khanna,[1] Marina Maciel Ansanelli,[2,3] Matthew F. Pusey,[1] and Elie Wolfe [2,3]

[1]*Department of Mathematics, University of York, Heslington, York YO10 5DD, United Kingdom*
[2]*Perimeter Institute for Theoretical Physics, 31 Caroline Street North, Waterloo, Ontario, Canada N2L 2Y5*
[3]*Department of Physics and Astronomy, University of Waterloo, Waterloo, Ontario, Canada, N2L 3G1*

The classical causal relations between a set of variables, some observed and some latent, can induce both equality constraints (typically conditional independencies) as well as inequality constraints (Instrumental and Bell inequalities being prototypical examples) on their compatible distribution over the observed variables. Enumerating a causal structure's implied inequality constraints is generally far more difficult than enumerating its equalities. Furthermore, only inequality constraints ever admit violation by quantum correlations. For both those reasons, it is important to classify causal scenarios into those which impose inequality constraints versus those which do not. Here we develop methods for detecting such scenarios by appealing to *d* separation, *e* separation, and incompatible supports. Many (perhaps all?) scenarios with exclusively equality constraints can be detected via a condition articulated by Henson, Lal, and Pusey (HLP). Considering all scenarios with up to four observed variables, which number in the thousands, we are able to resolve all but three causal scenarios, providing evidence that the HLP condition is, in fact, exhaustive.

I. INTRODUCTION

Many times in science we are not only interested in the observed correlations between events, but also in the underlying causal explanations that govern such observed phenomena. In the mathematical framework for causal inference [1–5], the candidates for these causal explanations are represented by directed acyclic graphs (DAGs), where each node is associated with a variable and each edge represents a direct causal influence.

A DAG imposes causal compatibility constraints on the probability distributions that can be causally explained by it. For example, a probability distribution over variables $\{A, B, C\}$ where A and C remain correlated even after a value of B is conditioned upon *cannot* be causally explained by the DAG of Fig. 1. All of the distributions over $\{A, B, C\}$ that can be explained by Fig. 1 need to satisfy $P(AC|B) = P(A|B)P(C|B)$. [6]

The causal compatibility constraint described above, denoted by $A \perp\!\!\!\perp_{\mathrm{CI}} C|B$, is a *conditional independence relation*. That is, it says that one set of variables becomes independent of a second set of variables when conditioning on a third. In general, however, a DAG can impose more complicated types of constraints on the compatible probability distributions. This only happens for DAGs that have latent nodes, i.e., nodes

associated with variables that do not appear in the probability distributions of interest.

From now on, the variables that do appear in the probability distributions of interest will be called *observed variables*, while the ones that do not will be called *latent variables*. Nodes associated with observed variables will be called *observed nodes* and will be depicted by triangles, while nodes associated with latent variables will be called *latent nodes* and will be depicted by circles.

An example of a DAG that imposes more complicated types of constraints on the compatible distributions over observed variables is the Bell DAG, presented in Fig. 2. This DAG is of interest to physicists, as it encompasses the causal assumptions of Bell's theorem.

Bell's theorem [7,8] is central to the foundations of quantum mechanics [9–12], as it says that no locally causal hidden-variable theory in which the observers can choose their measurements independently of the source can ever be capable of reproducing all the operational predictions of quantum theory [13–17]. These assumptions are encoded in the Bell DAG, as there the settings X as Y are not causally connected to the source Λ, as well as the setting in one wing does not causally influence the outcome in the other wing.

As it turns out, the Bell DAG imposes causal compatibility constraints on the compatible distributions over $\{A, B, X, Y\}$ that take the form of inequalities, which are precisely the Bell inequalities. This means that all the distributions, which violate Bell's inequalities, including some of the quantum predictions for this scenario, *cannot* be causally explained by the Bell DAG.

With the goal of causally explaining the violation of Bell's inequalities without changing the causal assumptions
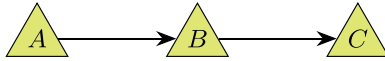
FIG. 1. Example of DAG.

embedded in the structure of the Bell DAG, Henson, Lal, and Pusey (HLP) [18] developed a generalization of Pearl's causal inference. In this generalized framework, the latent nodes can be associated with quantum or other generalized probabilistic theory (GPT) systems. HLP also proved that the conditional independence constraints remain the same independently of the theory that describes the latent nodes; other types of constraints, like Bell's inequalities, can be violated.

The Bell DAG is just one of the many causal structures for which inequality constraints can be violated when the latent nodes are associated with nonclassical systems [18–22]. In the Bell scenario, correlations that violate Bell inequalities have cryptographic applications precisely because of their nonclassicality, so it seems reasonable to hope that other scenarios allowing nonclassical correlations will have similar applications. Finding which causal structures imply inequalities, which potentially admit quantum violation is a critical step towards such potential applications.

In HLP, the concept of NON-ALGEBRAICNESS of a causal structure was defined. There, this concept was called *interestingness* [23]. A causal structure is said to be ALGEBRAIC when all of its causal compatibility constraints are of the form of conditional independence relations, even in the classical case. Conversely, if the causal structure imposes more complicated constraints on the compatible probability distributions, it is said to be NON-ALGEBRAIC. As proven in HLP, only the NON-ALGEBRAIC scenarios are passive of witnessing a difference between the sets of classically and quantumly achievable probability distributions.

In HLP, a sufficient condition for ALGEBRAICNESS of a causal structure was developed; it is referred to here as the *HLP criterion*. A central motivation behind this paper is that, at present, it is not known whether the HLP criterion is also necessary for ALGEBRAICNESS. That is, if a DAG cannot be proven ALGEBRAIC by virtue of the HLP criterion, is the DAG necessarily NON-ALGEBRAIC? The conjecture that the HLP criterion is indeed necessary for ALGEBRAICNESS will be referred to as the *HLP conjecture*.

How might we evaluate the HLP conjecture? To disprove it, we need to find only one DAG for which the HLP criterion does not apply, but which can be proven ALGEBRAIC by some other method. We did not pursue a search for such a counterexample, simply because we are unaware of any means to prove ALGEBRAICNESS when the HLP criterion does not apply. We therefore concentrate on providing evidence in support of the conjecture being true. Namely, we show that as one considers "larger and larger" DAGs, we can still certify the NON-ALGEBRAICNESS of (almost) all DAGs for which the HLP criterion does not apply.

To accomplish these goals we must clarify two preliminary questions. Firstly, how should we *enumerate* over DAGs? One enumeration style—the original enumeration choice employed by HLP—is to count DAGs by their total number of nodes. We can thus consider DAGs with five total nodes, with six total nodes, with seven total nodes, etc. While
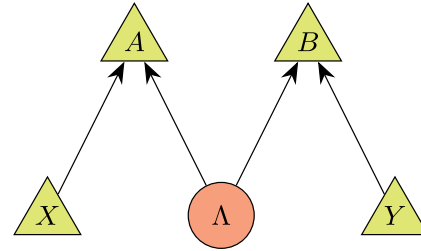

FIG. 2. The Bell DAG. It encompasses the assumptions of Bell's theorem for a Bell scenario where X and Y are the measurement settings of Alice and Bob, A and B are their outcomes and Λ is a classical hidden variable. The probability distributions that are classically compatible with this DAG are those that decompose as in Eq. (3).

this enumeration style has the advantage of simplicity, the arguably more natural enumeration, which will be used here is to count by the total number of observed nodes. We can thus consider DAGs with two observed nodes, three observed nodes, four observed nodes, etc. From naive structural considerations alone, however, one might imagine that there are infinitely many DAGs with any fixed number of observed nodes. Motivated in part by avoiding such infinities, when enumerating by the number of observed nodes we elect to work within Evans' framework of marginalized DAGs (mDAGs) [24], which will be explained in Sec. III A.

The second preliminary question is how to prove that a DAG for which the HLP criterion does not apply is indeed NON-ALGEBRAIC? Since we here seek to consider hundreds if not thousands of such DAGs, we are heavily invested in identifying *algorithmic* techniques for proving NON-ALGEBRAICNESS, within which we deprioritize computationally expensive methods. In stark contrast to the approach of HLP, we extensively leverage a new result due to Evans [25], who has shown that a DAG G is ALGEBRAIC if and only if it is observationally equivalent to *some* DAG that does not have latent variables—where two DAGs are said to be (classically) observationally equivalent when their sets of (classically) compatible probability distributions are the same. As such, we herein primarily exploit necessary conditions for a graph to be observationally *in*equivalent to every latent-free graph.

One such condition is that when the DAG is *nonmaximal*, i.e., when it has a pair of nodes that are not adjacent (not connected by an arrow nor by a shared latent common cause) but are nevertheless not d separated by any set of observed nodes, then the DAG is *not* observationally equivalent to any latent-free DAG. Another condition says that for two DAGs to be observationally equivalent, they must admit the same *e*-separation relations over their observed nodes. A third condition says that for two DAGs to be observationally equivalent, they must admit the same set of compatible *supports*. Although the latter condition subsumes the two former ones (as discussed in Sec. IV E), the former conditions can be evaluated more efficiently. The latter condition involving supports (remarkably!) can be assessed using Fraser's algorithm [26], which generally requires higher computational overhead. All of these conditions can be utilized to prove

TABLE I. A summary of our findings. Apart from the HLP criterion, which shows ALGEBRAICNESS, all of the other conditions listed show NON-ALGEBRAICNESS. Note that here we are counting by *unlabelled* DAGs. That is, two labeled DAGs. which are equivalent under a relabelling of the observed nodes and/or a relabelling of the hidden nodes are represented by a single unlabelled DAG in these enumerations.

Category	mDAGs with three observed nodes	mDAGs with four observed nodes
Total count	46	2809
remaining # for which the HLP criterion does not apply	5	996
remaining # for which our nonmaximality condition does not apply	1	186
remaining # for which our setwise nonmaximality condition does not apply	0	78
remaining # for which our d-separation condition does not apply	0	60
remaining # not resolved by our use of Fraser's algorithm	0	3

the NON-ALGEBRAICNESS of a given DAG, via proving the classical observational inequivalence of said DAG with *every* latent-free graph, which has the same number of observed nodes.

It is worth contrasting the tools we employ here to certify NON-ALGEBRAICNESS with those employed in prior literature by HLP and Pienaar [27,28]. HLP themselves attempted to explore all DAGs with six total nodes for which their sufficient condition for ALGEBRAICNESS did not apply. For all but five of these DAGs, HLP proved NON-ALGEBRAICNESS by means of deriving DAG-specific entropic inequalities and showing that those entropic inequalities could be violated by a DAG-specific distribution, which nevertheless satisfied the conditional independence constraints of the DAG. One of the remaining five cases is the Bell DAG, which is long-since established as NON-ALGEBRAIC. Another one of the five is the so-called triangle scenario, whose NON-ALGEBRAICNESS was proven using a one-off proof technique, which HLP did not generalize. The remaining three DAGs with six total nodes were eventually established as NON-ALGEBRAIC in a separate work by Pienaar [27], who employed a proof technique using fine-grained entropic inequalities.

At first glance, the algorithmic proofs of NON-ALGEBRAICNESS we employ here may seem unrelated to those utilized by HLP or Pienaar. However, our theorem relating e separation to NON-ALGEBRAICNESS turns out to recover all but one of the NON-ALGEBRAICNESS results that HLP achieved by appealing to entropic inequalities. Additionally, our application of Fraser's algorithm further witnesses the NON-ALGEBRAICNESS of every other DAG conjectured by HLP to be NON-ALGEBRAIC, including the three of which were only *proven* NON-ALGEBRAIC in the later study of Pienaar [27]. We are therefore confident that our plethora of techniques likely supersede those earlier employed by HLP and Pienaar [27], despite not having a formal proof yet. We have made the Python code implementing the filters for NON-ALGEBRAICNESS proposed throughout this manuscript publicly available [29], although without accompanying documentation at this time.

We summarize our ultimate findings in Table I. We interpret these results as evidence in favor of the HLP conjecture: among the thousands of analyzed mDAGs, there are only three potential counterexamples.

As discussed, this paper is of interest to quantum physicists because the NON-ALGEBRAIC DAGs are the possible candidates to explore quantum advantages in device

independent information processing protocols [30,31]. These DAGs are also the ones that should be looked into to compare quantum theory to more general probabilistic theories (GPTs) [32,33].

On the other hand, our problem is also of central interest for purely classical causal inference. The set of probability distributions, which are classically compatible with an ALGEBRAIC DAG is constrained only by conditional independence relations, which can be obtained from the d-separation relations of the DAG. Isolating a sufficient set of d-separation relations in a graph [34] is a well-studied problem. It is of paramount value to a classical data scientist, therefore, to know if the causal hypothesis encoded in a DAG may or may not be falsified by accounting for nontrivial inequality constraints. Such inequality constraints, when present, are often difficult to explicitly characterize.

Structure of the paper

In Sec. II A we present an introduction to the formalism used in causal inference, and proceed to explain what NON-ALGEBRAIC and ALGEBRAIC precisely mean in this context in Sec. III. We state and prove our e-separation condition in Sec. IV. In Sec. V we present our computational results and also discuss the methods we tried to confirm the NON-ALGEBRAICNESS of the three left mDAGs of four observed nodes. Our conclusions can be found in Sec. VI.

II. PRELIMINARIES

A. Causal Explanations of Observational Data

The area of causal inference is concerned with finding potential causal explanations for observed events. For example, imagine that we want to find out what is the causal relationship between three events: a cloudy sky, rain, and the floor being wet. Figure 3 depicts two possible causal structures between these three events; in 3(a) we hypothesize that the clouds cause the rain and the rain makes the floor wet. In 3(b), on the other hand, we hypothesize that the wet floor causes the rain, and the rain makes the sky cloudy.

There is an easy way we can check that Fig. 3(b) is not the correct causal hypotheses: if we pour water on the floor in a sunny day, it will not start raining.

Note that this method presupposes that we can *intervene* on our experiment, meaning that we can force one of the variables of the experiment (wet floor) to assume the value we
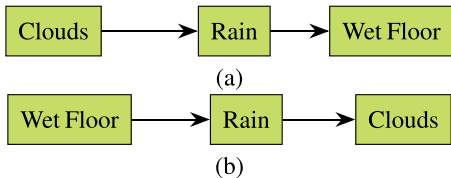
FIG. 3. Two hypothesis for the causal relationships between observing clouds, rain, and wet floor. By intervening on the experiment we can attest that (a) is the correct explanation, but we cannot attest that if we only passively look at the correlations between those events.

want. However, it is not always possible to do that; sometimes it is unethical or there are technical or fundamental limitations to do so.

If the experimenter cannot make interventions on the variables of interest, it is still possible for them to draw some conclusions from a passive observation of the correlations between the events of interest. As we will see, each causal structure imposes constraints on which probability distributions obtained from passive observations can be classically explained by it. These constraints can be tested against the observed probability distribution to see if the given causal structure is a valid causal hypothesis for the observed phenomena. In Fig. 3 it happens that both causal structures impose the same constraints on probability distributions, so they are not distinguishable by passive observations.

In this paper we will be mainly concerned with passive observations. The sets defined in Sec. III that are of relevance to us make reference to distributions obtained from passive observations only.

B. Directed Acyclic Graphs

In the framework we use here, the mathematical object that describes a causal structure is a directed acyclic graph (DAG). A directed graph G is a pair (A, E), where A is a set of nodes and $E \subseteq A \times A$ is a set of directed edges. A directed acyclic graph (DAG) is a directed graph that has no directed cycles. Below, we introduce some definitions regarding DAGs that will be useful later.

Definition 1 (Children, parents, descendants, ancestors). Let X be a node of a DAG G. If Y is another node of G such that there is a directed edge $X \to Y$, then Y is called a "child" of X, and X is called a "parent" of Y. The set of all children of X is denoted as $\mathrm{CH}_G(X)$, and the set of all parents of X is denoted as $\mathrm{PA}_G(X)$.

A directed path is a sequence of nodes $X^1, X^2, X^3 \ldots\ldots X^n$ such that $X^i \to X^{i+1}$ for $i = 1, \ldots, n$. The "descendants" of X in G are all the nodes in G that can be reached from X by a directed path. Conversely, all the nodes that have X as a descendant are called "ancestors" of X. The set of all ancestors of X is denoted as $\mathrm{ANC}_G(X)$, meanwhile the set of all descendants of X can be denoted as $\mathrm{DES}_G(X)$ [35].

For example, in the DAG of Fig. 4 we have that $\mathrm{PA}_G(B) = \{A, D\}$, and that E is a descendant of D, even if it is not its child.
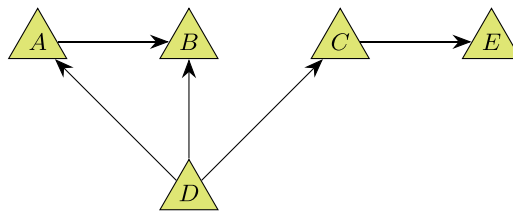

FIG. 4. Example of a directed acyclic graph (DAG). The probability distributions that are classically compatible with this DAG are those that can be decomposed as in Eq. (2).

C. DAGs as Causal Structures

When we associate each node of a DAG with an event of interest, the DAG is a representation of a causal structure: an edge $X \to Y$ shows a possibility of a direct causal influence of X on Y. Here, we will indicate the variable associated with a node by the same capital letter as the node itself. If all the events of interest are described by classical random variables, the idea that a probability distribution "can be causally explained" by a certain DAG is formalized through the *Markov condition*:

Definition 2 (Markov condition). Let G be a DAG with nodes A. A probability distribution P over the variables A is said to be *Markov with respect to G* if it can be factorized as

$$P(A) = \prod_i P(X^i | \mathrm{PA}_G(X^i)) \tag{1}$$

Where $A = \{X^1, \ldots, X^n\}$ and $\mathrm{PA}_G(X^i)$ is the set of parents of the node X^i in G.

If P is Markov with respect to G, we also say that it is "classically compatible" with G.

As an example of this definition, a joint probability distribution P_{ABCDE} over the random variables A, B, C, D, and E is Markov with respect to the DAG of Fig. 4 if it can be decomposed as

$$P(ABCDE) = P(A|D)P(B|A, D)P(C|D)P(D)P(E|C). \tag{2}$$

Therefore, through the Markov condition, each DAG imposes constraints on the probability distributions that are classically compatible with it. A DAG, which is classically compatible with *every* probability distribution, that is, a DAG that does not impose any constraint on the classically compatible distributions, is said to be a *saturated* DAG.

D. d Separation: A Graphical Criterion for Conditional Independence

If P is a probability distribution over a certain set of random variables A, we say that the variables of the subset $X \subseteq A$ are *conditionally independent* of the variables of the subset $Y \subseteq A$ given the subset $Z \subseteq A$ if P can be factorized as $P(X, Y|Z) = P(X|Z)P(Y|Z)$. This is denoted by $X \perp\!\!\!\perp_{\mathrm{CI}} Y|Z$.

Some of the constraints that a DAG imposes on the probability distributions that are classically compatible with it are of the form of conditional independence: If a probability distribution P *cannot* be factorized according to a certain conditional independence that is imposed by the DAG, then it is not classically compatible with the DAG. As it turns out, there exists a graphical algorithm to obtain all the conditional

independence constraints that are imposed by a DAG. This algorithm is called "d separation", and we will now describe it.

If G is a DAG with nodes A and $X, Y, Z \subseteq A$ are sets of nodes in G, the d-separation algorithm says whether X and Y are "d separated" by Z. This happens if all the undirected paths (paths that ignore the direction of the arrows) from X to Y are *blocked* by Z. A path is blocked if one or more of the following is true:

(1) There is a chain of nodes along the path: $i \rightarrow m \rightarrow j$ such that $m \in Z$.

(2) There is a fork along the path: $i \leftarrow m \rightarrow j$ such that $m \in Z$.

(3) There is a collider along the path: $i \rightarrow m \leftarrow j$ such that $m \notin Z$ and $d \notin Z$ for all the descendants d of m.

If X is d separated of Y given Z in the DAG under consideration, we denote it as $X \perp_d Y | Z$.

For example, for the DAG of Fig. 1 the d-separation criterion says that $A \perp_d C | B$. This means that the event A should become independent of the event C upon knowledge of B; if your distribution P *does not* satisfy the constraint $P(AC|B) = P(A|B)P(C|B)$, Fig. 1 is *not* a valid causal explanation for it. Interpreting the variables $\{A, B, C\}$ as respectively the clouds, rain and wet floor [such as in Fig. 3(a)], this d-separation relation says that the occurrence of clouds becomes independent of the floor being wet if we already know whether it is raining.

The following theorem, proven in Ref. [2], makes explicit the connection between d- separation relations and conditional independence relations:

Theorem 1 (d separation and conditional independence). Let G be a DAG with nodes A, and let $X \subseteq A$, $Y \subseteq A$ and $Z \subseteq A$ be three disjoint subsets of A. Then

(1) If G has the d-separation relation $X \perp_d Y | Z$, then all of the probability distributions over the variables A, which are Markov with respect to G need to satisfy $X \perp\!\!\!\perp_{CI} Y | Z$.

(2) If G *does not have* the d-separation relation $X \perp_d Y | Z$, then there exists some probability distribution over the variables A, which is Markov with respect to G and *does not* satisfy $X \perp\!\!\!\perp_{CI} Y | Z$.

When a DAG does not have latent nodes, the *only* constraints that it imposes on the compatible distributions are the conditional independence relations associated with the d-separation relations of the DAG [[39], Theorem 3]:

Theorem 2 (d separation is complete for DAGs without latent nodes). Let G be a DAG with nodes A. A probability distribution over the variables A is Markov with respect to G if and only if it satisfies all of the conditional independence relations associated with the d-separation relations of G.

Importantly, in general this is *not* true for DAGs that include latent nodes, as we will see now.

E. Latent-Permitting DAGs

As discussed in the introduction, sometimes we want to allow for causal explanations that include *latent variables*, i.e., variables that do not appear in the final probability distribution we are trying to explain. When our DAG of interest can have latent nodes, we call it a *latent-permitting DAG*, as opposed to the *latent-free DAGs* that only have observed nodes.

Let G be a DAG that has the set of nodes $A = V \cup L$, V and L disjoint, where V are the observed nodes and L are the

latent nodes. Mimicking the terminology used for the latent-free case, we will say that a probability distribution $P(V)$ over the observed variables is *classically compatible with G* if there exists some probability distribution $P(V, L)$ over $V \cup L$ such that

(i) $P(V, L)$ is Markov with respect to G.

(ii) The marginal of $P(V, L)$ over V is the original probability distribution that we are interested in, i.e., $\sum_L P(V, L) = \sum_L P(V|L)P(L) = P(V)$.

As an example, a probability distribution P_{ABXY} over the random variables $A, B, X,$ and Y is classically compatible with the Bell DAG (Fig. 2) if it can be decomposed as

$$P(ABXY) = \sum_\Lambda P(A|X, \Lambda)P(B|Y, \Lambda)P(X)P(Y)P(\Lambda). \quad (3)$$

When dealing with latent-permitting DAGs, we will call the conditional independence relations that only involve observed variables the *observed conditional independence relations*, and similarly the d-separation relations that only involve observed nodes will be called *observed d-separation relations*. From the definition of classical compatibility with a latent-permitting DAG, we can see that the conclusions of Theorem 1 are also valid for latent-permitting DAGs:

Theorem 3 (Observed d-separation in latent-permitting DAGs). Let G be a DAG with nodes $A = V \cup L$, where V are observed nodes and L are latent nodes. Let $X \subseteq V, Y \subseteq V$ and $Z \subseteq V$ be three disjoint sets of observed nodes of G. Then

(1) If G has the d-separation relation $X \perp_d Y | Z$, then all of the probability distributions over the variables V, which are classically compatible with G need to satisfy $X \perp\!\!\!\perp_{CI} Y | Z$.

(2) If G *does not have* the d-separation relation $X \perp_d Y | Z$, then there exists some probability distribution over the variables V, which is classically compatible with G and *does not* satisfy $X \perp\!\!\!\perp_{CI} Y | Z$.

As seen in Theorem 2, the only constraints that latent-free DAGs impose on the compatible probability distributions are the conditional independence relations, that can be obtained from d separation. Consequently, if a DAG G has nodes $A = V \cup L$, all the constraints that it imposes on the compatible joint probability distributions $P(A) = P(V, L)$ are the conditional independence relations. However, if we are interested only on distributions over the observed variables V, sometimes the conditional independence constraints that involve the latent variables L might induce more complicated extra constraints on the distributions over the observed variables V. If a probability distribution $P(V)$ over the observed variables satisfies both the observed conditional independence relations (obtained from d separation) and the extra constraints derived from the conditional independence relations that involve the latent variables L, then it is classically compatible with G.

Therefore, in principle one could find all the d-separation relations of a DAG (involving both observed and latent nodes), thus getting conditional independence constraints on $P(V, L)$, and from there infer the constraints on the probability distribution over the observed variables, $P(V)$. This process will be exemplified with the Bell DAG in the beginning of the next section. Inferring constraints on $P(V)$ from the conditional independence relations of $P(V, L)$, however, is in general very complicated.

III. THE PROBLEM: LOOKING FOR NON-ALGEBRAIC SCENARIOS

The goal of this paper is to classify DAGs as ALGEBRAIC or not, following the concept introduced in HLP. Before establishing this concept in full generality, we will explore the example of the Bell DAG (Fig. 2).

As discussed in Sec. II E, the d-separation criterion gives us all the classical compatibility constraints imposed on the probability distribution over *all the nodes*, $P(V, L)$. In the case of the Bell DAG, we have $P(V, L) = P(ABXY\Lambda)$, while $P(V) = P(ABXY)$.

The conditional independence relations of $P(ABXY\Lambda)$, that come from the d-separation relations of the Bell DAG, are

$$P(\Lambda|XY) = P(\Lambda), \tag{4}$$

$$P(AB|XY\Lambda) = P(A|X\Lambda)P(B|Y\Lambda), \tag{5}$$

$$P(A|XY) = P(A|X)$$
$$P(B|XY) = P(B|Y)$$
$$P(XY) = P(X)P(Y). \tag{6}$$

These equations will give rise to the constraints imposed by the Bell DAG on $P(ABXY)$ through the marginalization of Λ. Equations (6) that do not involve Λ, are automatically transported as observed conditional independence constraints imposed by the Bell DAG on $P(ABXY)$. Equations (4) and (5) that do involve the latent variable Λ, will give rise to another type of constraint on $P(ABXY)$: the Bell's inequality. In fact, Eq. (4) encodes the no-superdeterminism assumption and Eq. (5) encodes the local causality assumption that are used to derive Bell's inequality.

Therefore, to study the Bell DAG, it is *not enough* to just look at the conditional independence constraints on the compatible distributions $P(ABXY)$ [Eqs. (6)]. If one does that, they would miss important information that is encoded in Bell's inequality. This is so because the set of probability distributions $P(ABXY)$ that satisfy only the conditional independence relations of Eqs. (6) is *strictly larger* than the set of probability distributions that satisfies these conditional independence relations *and* Bell's inequality. In other words, Bell's inequality is not implied by Eqs. (6).

This is the core of the concept of NON-ALGEBRAICNESS: a NON-ALGEBRAIC DAG imposes nontrivial inequality constraints on the classically compatible distributions. A "nontrivial" inequality constraint is an inequality that is not implied by the observed conditional independence relations of the DAG, along with nonnegativity of all probabilities and normalization [40]. Note that an ALGEBRAIC DAG can impose *trivial* inequality constraints on the compatible distributions: For example, if the node Λ in the Bell DAG was treated as an observed node, then the Bell inequalities would still be satisfied by the compatible distributions $P(ABXY\Lambda)$. However, in this case the Bell inequalities would be trivial, because they would be implied by *observed* conditional independence relations [Eqs. (4) and (5)].

To formalize the idea of NON-ALGEBRAICNESS, we will introduce a few definitions:

Definition 3. Let G be a DAG. The sets $\mathcal{C}_G$ and and $\mathcal{I}_G$ of probability distributions over observed variables are defined as follows:

(1) $\mathcal{C}_G$: Set of probability distributions that are classically compatible with G.

(2) $\mathcal{I}_G$: Set of probability distributions that satisfy all the conditional independence constraints that follow from the observed d-separation relations of G.

For the case of the Bell DAG, $\mathcal{I}_{\text{Bell}}$ represents the set of distributions that obey the Eqs. (6). By contrast, $\mathcal{C}_{\text{Bell}}$ consists of a strict subset of $\mathcal{I}_{\text{Bell}}$, where we additionally restrict the conditional probabilities $P(AB|XY)$ to satisfy Bell's inequalities and thus lie in the local polytope.

Theorem 3 shows that $\mathcal{C}_G \subseteq \mathcal{I}_G$ for every DAG G. This is so because all the probability distributions that are classically compatible with G have to satisfy the conditional independence constraints that come from its observed d-separation relations. When the observed conditional independence relations are the only constraints imposed by a DAG on the compatible probability distributions over observed variables, the DAG is said to be ALGEBRAIC:

Definition 4 (ALGEBRAICNESS). Let G be a DAG. If $\mathcal{C}_G = \mathcal{I}_G$, then G is said to be ALGEBRAIC. Conversely, if $\mathcal{C}_G \subsetneq \mathcal{I}_G$, then G is said to be NON-ALGEBRAIC.

We borrow the terminology of NON-ALGEBRAIC and ALGEBRAIC from algebraic geometry: An algebraic set is defined by polynomial equalities (or more generally, by some finite union of sets each of which is defined by polynomial equalities). Semialgebraic sets, by contrast, are characterized by both polynomial equalities and polynomial inequalities. To emphasise that a DAG's set of (classically) compatible distributions is defined by *more* that just the conditional independence (notably, equality) constraints, we therefore elect to speak of such a DAG as NON-ALGEBRAIC.

As proven by HLP, the observed conditional independence constraints imposed by a DAG on the compatible probability distributions do not change when the latent variables of the DAG are associated with quantum systems or other GPT systems. Therefore, if one is interested in studying causal structures that provide any quantum or GPT observational advantage, then there is only hope among the NON-ALGEBRAIC scenarios. If a DAG is ALGEBRAIC, then *all* the probability distributions that exhibit the conditional independence relations associated with its observed d-separation relations can be explained by this DAG *classically*.

Theorem 2 implies that every latent-free DAG is ALGEBRAIC. In HLP, a stronger sufficient condition for ALGEBRAICNESS is provided, together with an algorithmic strategy to check it. This condition, called the *HLP criterion*, will be discussed in Sec. III B. It is still not known whether the HLP criterion is also necessary for ALGEBRAICNESS; its possible outcomes for a given DAG are either that it is ALGEBRAIC or that it is "unresolved". The unresolved DAGs thus need to be assessed by some other method.

Based on the HLP criterion and certain types of entropic inequalities, HLP and Pienaar [27] classified the ALGEBRAICNESS or NON-ALGEBRAICNESS of all DAGs of up to six *total* nodes (observed and latent), thus leaving no DAGs of six total nodes with inconclusive status. In this paper, however,
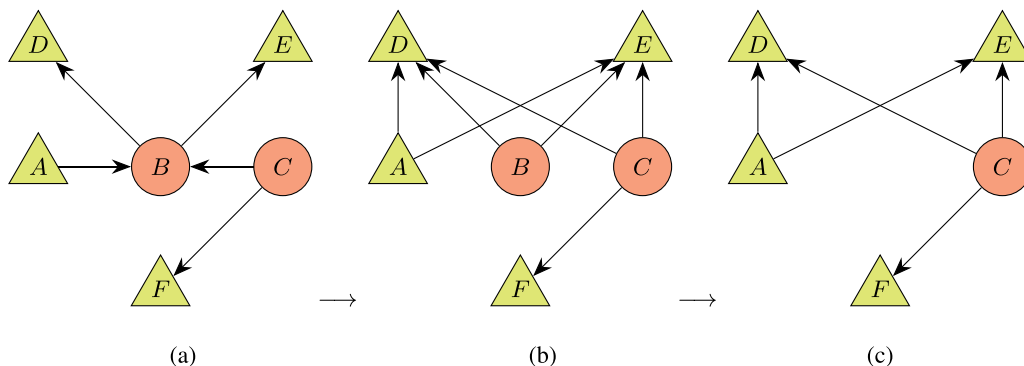
FIG. 5. The DAGs (a) and (b) are classically observationally equivalent to the mDAG of (c). The step (a)→(b) can be shown by Lemma 1, and the step (b)→(c) can be shown by Lemma 2.

we elect to count DAGs not by their *total* node count but rather by their number of *observed* nodes. It turns out that HLP 's complete classification of DAGs with *total* node count up to six meant that they resolved *all* DAGs with three observed nodes, a *few* DAGs with four observed nodes, and *no* DAGs with five or more observed nodes. Here we attempt to tackle the ALGEBRAICNESS classification of all causal structures with up to fpir observed nodes [41]. To do so, we will utilize the *mDAG formalism* introduced by Evans in Ref. [24].

A. Simplifying the Problem by Using mDAG Formalism

Two DAGs G and H are said to be *classically observationally equivalent* when $\mathcal{C}_G = \mathcal{C}_H$. Note that $\mathcal{C}_G = \mathcal{C}_H$ implies that $\mathcal{I}_G = \mathcal{I}_H$: by Theorem 3, if a certain d-separation relation is *not* present in a DAG, it is always possible to find a probability distribution that violates the conditional independence corresponding to this d-separation relation and is classically compatible with the DAG. In other words, if two DAGs are classically compatible with the same sets of distributions, they have to present the same d-separation relations [42].

In short,

$$\mathcal{C}_G = \mathcal{C}_H \Rightarrow \mathcal{I}_G = \mathcal{I}_H, \tag{7}$$

with contrapositive

$$\mathcal{I}_G \neq \mathcal{I}_H \Rightarrow \mathcal{C}_G \neq \mathcal{C}_H. \tag{8}$$

In particular, this means that if a DAG G is ALGEBRAIC (NON-ALGEBRAIC), then all of the DAGs H that are compatible with it are also ALGEBRAIC (NON-ALGEBRAIC).

In this section, we will present two results of [24] that prove classical observational equivalence, thus reducing the number of DAGs that have to be examined for ALGEBRAICNESS. After presenting the two results, we will show that they allow for a definition of a new object, called mDAG, that encompasses this simplification.

To do so, we start with the definition of *exogenization*. It might be easier to understand this definition by following Fig. 5, where DAG 5(b) is obtained from DAG 5(a) by exogenizing node B.

Definition 5 (Exogenized DAG). Let G be a DAG and let λ be a latent variable of G. We define the *exogenized DAG* $\mathcal{E}(G, \lambda)$ as follows: take the vertices and edges of G and (1)

add an edge $m \rightarrow n$ from every $m \in \text{PA}_G(\lambda)$ to every $n \in CH_G(\lambda)$ and (2) delete edges $m \rightarrow \lambda$ for every $l \in \text{PA}_G(\lambda)$. All other edges remain the same.

With this definition at hand, we state the Lemma 3.7 of [24]:

Lemma 1 (Exogenization). Let G be DAG with observed nodes V and latent nodes L, and let $\lambda \in L$ be a latent node of G. Furthermore, let $\tilde{G} = \mathcal{E}(G, \lambda)$. Then, $\mathcal{C}_G = \mathcal{C}_{\tilde{G}}$.

Now, we state the Lemma 3.8 of [24]. This Lemma is also illustrated in Fig. 5, where it is applied to go from 5(b) to 5(c).

Lemma 2 (No redundant latents). Let G be a DAG with observed nodes V and latent nodes L. Let $\lambda \in L$ and $\mu \in L$ be latent nodes of G such that $\lambda \neq \mu$, $\text{PA}_G(\lambda) = \text{PA}_G(\mu) = \emptyset$ and $CH_G(\lambda) \subseteq CH_G(\mu)$. In this case, we say that the node λ is "redundant". Let G' be the DAG obtained after deleting the node λ. Then, $\mathcal{C}_G = \mathcal{C}_{G'}$.

Like Lemma 1, Lemma 2 also shows conditions under which proving the NON-ALGEBRAICNESS of one DAG automatically gives you the NON-ALGEBRAICNESS of another.

For example, Lemmas 1 and 2 show that all the three DAGs of Fig. 5 have the same sets $\mathcal{C}$ and $\mathcal{I}$. Thus, we need only to examine one DAG out of the these. It makes sense to pick the DAG of Fig. 5(c), as it is the simpler.

Following this same idea, we can work with the concept of an mDAG, first defined in Ref. [24]. The definition below is different than, but equivalent to the one presented in Ref. [24].

Definition 6 (mDAG). An mDAG is a DAG where none of the latent nodes is redundant (as defined in Lemma 2) nor has any parents.

For example, the DAG of Fig. 5(c) is an mDAG. For a fixed number of observed nodes, there is a finite number of mDAGs. In particular, for three observed nodes there are 46 mDAGs, while for four observed nodes there are 2809 mDAGs. Lemmas 1 and 2 show that the mDAG encodes all the necessary information of the DAG if you only want to talk about the sets $\mathcal{C}$ and $\mathcal{I}$.

The Lemmas here presented also give an argument in favor of counting the causal structures in terms of the number of observed nodes instead of in terms of the total number of nodes; Lemma 2 shows that DAGs 5(a) and 5(b), that have six total nodes, actually do not have to be analyzed; their NON-ALGEBRAICNESS can be examined by looking at 5(c), that has five total nodes.
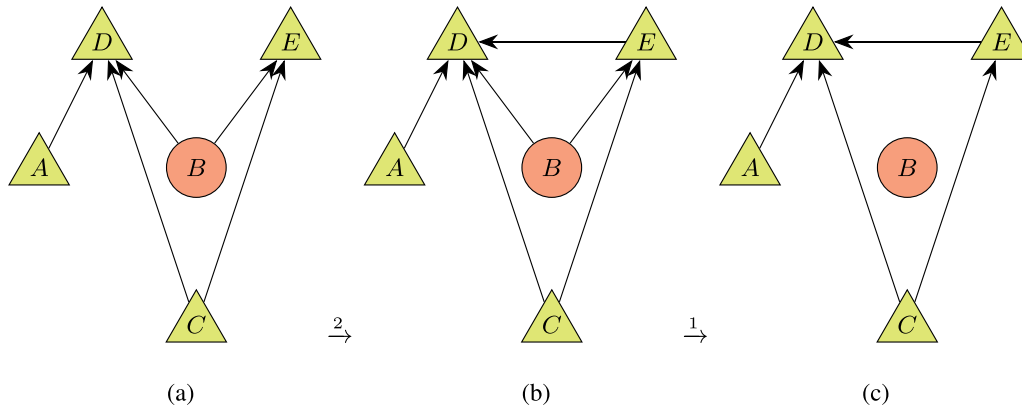
FIG. 6. An example of the application of the HLP criterion. All of these three mDAGs have the same set of d-separation relations: $A \perp_\mathrm{d} E$, $A \perp_\mathrm{d} E | C$, $A \perp_\mathrm{d} C$ and $A \perp_\mathrm{d} C | D$. Since (c) is latent-free, we can conclude that (a) is ALGEBRAIC.

B. The HLP Criterion

In this section, we describe the sufficient criterion for AL-GEBRAICNESS that was developed in HLP. The criterion gives some transformations that take a DAG G to another DAG H such that $\mathcal{C}_H \subseteq \mathcal{C}_G$ while $\mathcal{I}_H = \mathcal{I}_G$. If the final DAG H is known to be ALGEBRAIC (for example, by being latent-free), then the original DAG G is also ALGEBRAIC.

These transformations, adjusted to the language of mDAGs, are presented in Theorem 4.

Theorem 4. Let G and H be two mDAGs. Suppose that H can be obtained by starting from G and applying one or more of the following transformations:

(1) Removal of an edge.

(2) Addition of an edge $X \rightarrow Y$, where previously $PA(X) \subseteq PA(Y)$ and $PA(X)$ contained at least one latent node.

Then, $\mathcal{C}_H \subseteq \mathcal{C}_G$.

Now we state the *HLP criterion* as a corollary:

Corollary 1 (HLP criterion). Let G be an mDAG. Suppose that by a sequence of the transformations defined in Theorem 4 it is possible to start from G and reach another mDAG H such that

(1) H does not have latent nodes.

(2) The set of observed d-separation relations of H and G is the same, i.e., $\mathcal{I}_H = \mathcal{I}_G$.

Then, the original mDAG G is ALGEBRAIC.

Proof. From Theorem 4, $\mathcal{C}_H \subseteq \mathcal{C}_G$. Since H is latent-free, by Theorem 2 it is ALGEBRAIC, i.e., $\mathcal{C}_H = \mathcal{I}_H$. Therefore, we have $\mathcal{I}_G = \mathcal{I}_H = \mathcal{C}_H \subseteq \mathcal{C}_G$. As noted before, $\mathcal{C} \subseteq \mathcal{I}$ is valid for any DAG, due to Theorem 3. Therefore, $\mathcal{C}_G = \mathcal{I}_G$, meaning that G is ALGEBRAIC. ∎

Figure 6(a) exemplifies an mDAG that can be shown ALGEBRAIC by the HLP criterion: by a sequence of transformations defined in Theorem 4 we can obtain the mDAG shown in Fig. 6(c), that obeys the conditions presented in Corollary 1.

The natural question to ask here is whether the HLP criterion is also necessary for an mDAG to be ALGEBRAIC. The conjecture that the HLP criterion is also necessary will be called the *HLP conjecture*:

Conjecture 1 (HLP conjecture). Let G be an ALGEBRAIC mDAG. Then, by a sequence of transformations defined in Theorem 4, it is possible to start from G and reach another mDAG H such that

(1) H does not have latent nodes.

(2) The set of observed d-separation relations of H and G is the same, i.e., $\mathcal{I}_H = \mathcal{I}_G$.

As we will see, our current results do not prove this conjecture, but give hints towards its validity.

In Ref. [25], Evans has shown that:

Theorem 5. A latent-permitting DAG G is ALGEBRAIC if and only if it is classically observationally equivalent to a latent-free DAG H.

This result allows us to restate the HLP conjecture in a different manner:

Conjecture 2 (Reformulation of the HLP conjecture). Let G be an mDAG. G is classically observationally equivalent to a latent-free mDAG H if and only if

(1) The set of observed d-separation relations of H and G is the same, i.e., $\mathcal{I}_H = \mathcal{I}_G$.

(2) Through the transformations defined in Theorem 4, it is possible to go from G to some latent-free mDAG H', which has the same set of observed d-separation relations as G and H, i.e., $\mathcal{I}_{H'} = \mathcal{I}_H = \mathcal{I}_G$.

Therefore, proving the HLP conjecture would also be of relevance to the problem of classifying causal structures into classical observational equivalence classes.

IV. METHODS TO DETERMINE Non-Algebraicness

In this section we discuss the methods we used to prove the NON-ALGEBRAICNESS of a large number of DAGs that do not respect the HLP criterion.

A. Using Nonmaximality to Prove Non-Algebraicness

The first method to show ALGEBRAICNESS that we will present relies on the concept of *maximality* [25,43]. To define this, we will first define what are *adjacent* and *d-separable* pairs of nodes.

Definition 7 (Adjacency). Let G be an mDAG, and let A and B be a pair of observed nodes of G. We say that A and B are *adjacent* in G if A is a parent of B, or B is a parent of A, or A and B share a common latent parent [44].

Definition 8 (d-(un)separable pair of observed nodes). Let G be a DAG with nodes $A = V \cup L$, where V are observed nodes and L are latent nodes. A pair of observed
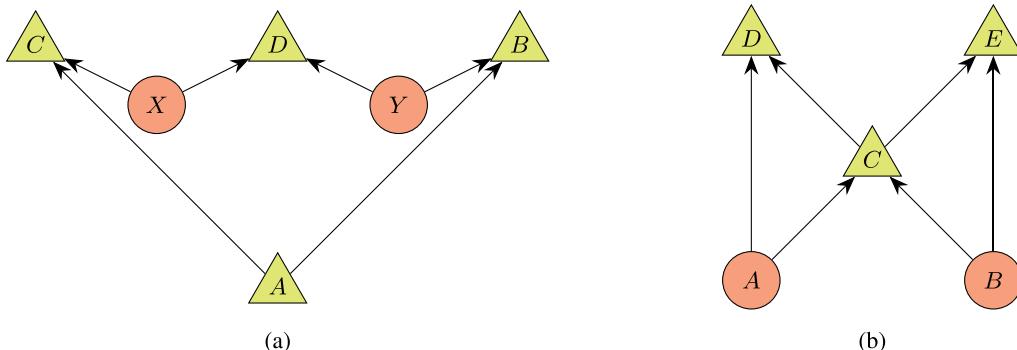
FIG. 7. (a) An example of maximal DAG, which so happens to be NON-ALGEBRAIC. (b) An example of an nonmaximal DAG, namely, The Unrelated Confounders scenario introduced in Ref. [45]. It is not maximal because nodes D and E are not d separable, but are nevertheless not adjacent. Like *all* nonmaximal DAGs, the Unrelated Confounders scenario is NON-ALGEBRAIC.

nodes $\{A, B\}$ for $A \in \mathbf{V}$, $B \in \mathbf{V}$ is said to be *d separable* if there is some subset $\mathbf{Z} \subseteq \mathbf{V}$ of the remaining observed nodes such that $(A \perp_{\mathrm{d}} B | \mathbf{Z})$, otherwise, A and B are said to be *d unseparable*.

These two definitions are related, in that they are criteria with relate to the compatibility of a particular distribution, namely, perfect correlation between one pair of nodes while all other nodes are point distributed. Consider the following distribution:

$$P(A, B, C, D \ldots) = \tfrac{1}{2}(\delta_{A,0}\delta_{B,0} + \delta_{A,1}\delta_{B,1})\delta_{C,0}\delta_{D,0} \ldots \quad (9)$$

Equation (9) describes a probability distribution in which A and B are random but perfectly correlated, and every other observed node is point distributed at the value 0. Then,

Proposition 1 (Adjacency $\Leftrightarrow P_{(9)} \in \mathcal{C}_G$). The distribution in Eq. (9) is classically compatible with the graph G if and only if A and B are adjacent nodes in G.

Proof. See Refs. [24,45]. Clearly $P_{(9)} \in \mathcal{C}_G$ when A and B *are* adjacent. Whenever A and B are *not* adjacent, then evidently A and B are *e* separated [46] upon removal of $\mathbf{V} \setminus \{A, B\}$, which also means that $P_{(9)}$ would violate the entropic inequality in Theorem 5 of Ref. [47]. ∎

Proposition 2 (d unseparability $\Leftrightarrow P_{(9)} \in \mathcal{I}_G$). The distribution in Eq. (9) satisfies all the conditional independence constraints that follow from the observed d-separation relations relations of graph G if and only if A and B are d unseparable in G.

Proof. The only conditional independence relations that the distribution in Eq. (9) *fails* to satisfy are those of the form $A \perp\!\!\!\perp_{\mathrm{CI}} B | \mathbf{Z}$. Such conditional independence relations follow from the observed d separation relations relations of graph G if and only if A and B are d separable in G. ∎

Putting Propositions 1 and 2 together, we find that $P_{(9)} \in \mathcal{I}_G$ but $P_{(9)} \notin \mathcal{C}_G$ whenever a graph has A and B nonadjacent but also d unseparable. This leads us to the concept of *maximality* [25,43]:

Definition 9 (Maximal DAG). Let G be a DAG. If all of the pairs of nodes of G, which are *not d* separable are also adjacent then G is said to be *maximal*, otherwise G is said to be *nonmaximal*.

Figure 7 shows an example of a maximal DAG [7(a)] and an example of a nonmaximal DAG [7(b)].

Theorem 6 [Nonmaximal]. Every ALGEBRAIC DAG is maximal, that is, every nonmaximal DAG is NON-ALGEBRAIC.

Proof. As we have seen from Propositions 1 and 2, if a graph G is nonmaximal, then there is some pair of observed nodes such that the distribution given by "those two nodes are random and perfectly correlated while all other observed nodes are point distributed" lies in some *gap* between $\mathcal{I}_G$ and $\mathcal{C}_G$. Alternatively, note that nonmaximality implies NON-ALGEBRAICNESS also follows from the fact that all latent-free graphs are maximal [[43], Prop. 3.19]. We then simply note that a nonmaximal graph cannot be observationally equivalent to any maximal graph: It follows from Eq. 8 that every pair of observationally equivalent DAGs must agree on their sets of d-(un)separable observed-node pairs. Furthermore, pursuant to Lemma 10 (discussed in Appendix B), agreement with respect to adjacency structure is *also* a prerequisite for observational equivalence [24,45]. These facts imply that if a DAG G is nonmaximal, then it is *not* going to be observationally equivalent to any latent-free DAG (and will thus be NON-ALGEBRAIC by Theorem 5). ∎

Figure 8 shows an example of a four-observed-nodes mDAG whose NON-ALGEBRAICNESS may be shown by Theorem 6.
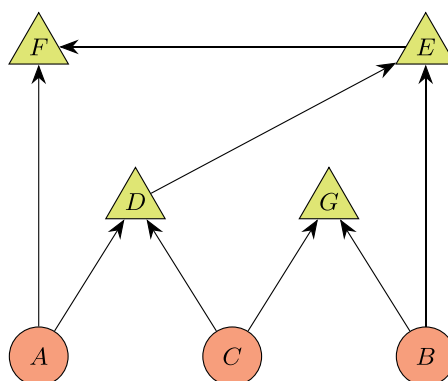

FIG. 8. A DAG with four observed nodes that is shown to be NON-ALGEBRAIC per Theorem 6. This can be seen because G and F not d separable, i.e., none of the d-separation relations $(G \perp_{\mathrm{d}} F | E)$, $(G \perp_{\mathrm{d}} F | D)$ or $(G \perp_{\mathrm{d}} F | D, E)$ hold, but they are not adjacent.

To find out which sets of pairs of observed nodes are *d* separable in *G* and then check whether it is maximal, is it necessary to first obtain *all* the *d*-separation relations of *G*? The following accessory lemma shows that this is not the case, thus simplifying the application of Theorem 6 in practice.

Lemma 3 (Rapid test for d separability of pairs). A pair of node subsets $\boldsymbol{A}$ and $\boldsymbol{B}$ of a DAG *G* are *d* separable by *some* set of observed nodes if and only if they are *d* separable by a *particular* set of observed nodes, namely, the set of all and only those observed nodes, which are ancestors of either $\boldsymbol{A}$ or of $\boldsymbol{B}$. In other words, $\boldsymbol{A} \perp_{\mathrm{d}} \boldsymbol{B} \mid \mathrm{ANC}_G(\boldsymbol{A}) \cup \mathrm{ANC}_G(\boldsymbol{B})$ whenever there exists some set $\boldsymbol{Z}$ such that $\boldsymbol{A} \perp_{\mathrm{d}} \boldsymbol{B}|\boldsymbol{Z}$.

Proof. This follows from Theorem 6 of Ref. [48]. In particular, it follows from the use of Ref. [48]'s Algorithm 3 to solve Ref. [48]'s Problem 4. ∎

In their Appendix E, HLP effectively utilized Shannon-type inequalities to certify the NON-ALGEBRAICNESS of many DAGs with five or six total nodes. In particular, they found that the Shannon-type inequalities for those DAGs could by violated by distributions that satisfy all of the conditional independence relations imposed by the DAG: those distributions were such that two particular variables are perfectly correlated and all other observable variables are point distributed, identical in nature to the construction of Eq. (9). From this, note that Proposition 1 implies that the nodes corresponding to these perfectly correlated variables are *not* adjacent, and Proposition 2 implies that they are not *d* separable. Thus, our Theorem 6 also certifies the NON-ALGEBRAICNESS of all those examples. Conversely, *every* pair of nodes, which are not *d* separable but also not adjacent generate a distribution from Eq. (9) that violates a Shannon-type inequality; this is a consequence Theorem 5 of Ref. [47], that says that every *e*-separation relation (which will be defined in Sec. IV D) implies on a Shannon-type inequality. This result was already used in the proof of Proposition 1.

B. Using Setwise Nonmaximality to Prove NON-ALGEBRAICNESS

In the previous subsection we found that that membership in $\mathcal{I}_G$ or $\mathcal{C}_G$ of the distribution in which *pair* of observed nodes is perfectly correlated while all other observed nodes are point distributed is directly related to the concepts of *d* unseparability and adjacency, respectively. Here we formulate analogous criteria in order to assess perfect correlation of three-or-more variables when all other variables in the distribution are point-distributed.

Definition 10 (Setwise adjacency). Let *G* be a DAG with nodes $\boldsymbol{A} = \boldsymbol{V} \cup \boldsymbol{L}$, where $\boldsymbol{V}$ are observed nodes and $\boldsymbol{L}$ are latent nodes. Then, the subset of observed nodes $\{V_1...V_k\}$ is setwise adjacent in *G* if and only if there is some node *X* (possibly but not necessarily within $\{V_1...V_k\}$) such that *X* is an ancestor of *every* node in $\{V_1...V_k\}$ not only in the DAG *G* but also in the *subgraph* of *G* formed by deleting all nodes $\boldsymbol{V} \setminus \{V_1...V_k\}$ from *G* [49].

Definition 11 (Setwise d unrestriction). Let $\boldsymbol{V}$ be the set of all observed nodes in some DAG *G*, and let $\boldsymbol{S}$ be some subset of $\boldsymbol{V}$. Then, the nodes $\boldsymbol{S}$ are setwise *d* unrestricted in *G* if and only if there does *not* exist any pair of nodes $\{S_i, S_j\} \subset \boldsymbol{S}$ along with some (possibly empty) set of observed nodes $\boldsymbol{Z} \subset \boldsymbol{V} \setminus \boldsymbol{S}$ such that $(S_i \perp_{\mathrm{d}} S_j|\boldsymbol{Z})$.

We next show that these definitions for setwise adjacency and setwise *d* unrestriction have the desired properties. Consider the following distribution:

$$P(\boldsymbol{V}) = \tfrac{1}{2}\big(\delta_{V_1,0}\delta_{V_2,0}...\delta_{V_k,0} + \delta_{V_1,1}\delta_{V_2,1}...\delta_{V_k,1}\big)\delta_{V_{k+1},0}\delta_{V_{k+2},0}...$$

(10)

Equation (10) describes a probability distribution in which the first *k* observed variables are random but perfectly correlated while all other observed variables are point distributed at the value 0. Then,

Proposition 3 (Setwise adjacency $\Leftrightarrow P_{(10)} \in \mathcal{C}_G$). The distribution in Eq. (10) is classically compatible with graph *G* if and only if $\{V_1...V_k\}$ are setwise adjacent in *G*.

Proposition 3 follows from the two accessory lemmas below.

Lemma 4 (Partial point distribution $\Rightarrow$ subgraph compatibility). Suppose $P(\boldsymbol{V})$ is some distribution wherein the variables $\boldsymbol{V} \setminus \{V_1...V_k\}$ are point distributed and moreover all the variables in $\{V_1...V_k\}$ have finite cardinality. Then, $P(\boldsymbol{V})$ is classically compatible with *G* if and only if $P(\{V_1...V_k\})$ is compatible with the *subgraph* of *G* formed be deleting all nodes $\boldsymbol{V} \setminus \{V_1...V_k\}$ from *G*.

Proof. This lemma is an immediate consequence of the *e*-separation theorem central in Ref. [45]. ∎

Lemma 5 (Setwise correlation $\Rightarrow$ common ancestor). Let $P_{\mathrm{perfect\ correlation}}(\boldsymbol{A})$ be the distribution in which all variables in $\boldsymbol{A}$ are random but perfectly correlated with each other. Then, $P_{\mathrm{perfect\ correlation}}(\boldsymbol{A})$ is compatible with a DAG *G* if and only if all the nodes in $\boldsymbol{A}$ share some common ancestor in *G*.

Proof. The "if" direction is trivial; the "only if" direction follows from Ref. [[37], Theorem 2]. ∎

Note that Proposition 3 implies Proposition 1 as a special case: A pair of observed nodes share a common ancestor upon removing all other observed nodes from a DAG *G* if and and only they are adjacent in *G*.

We likewise highlight the utility of the definition of a setwise *d*-unrestricted set:

Proposition 4 (Setwise d unrestriction $\Leftrightarrow P_{(10)} \in \mathcal{I}_G$). The distribution in Eq. (10) satisfies all the conditional independence constraints that follow from the observed *d*-separation relations relations of graph *G* if and only if $\{V_1...V_k\}$ are setwise *d* unrestricted in *G*.

Proof. Suppose that $\{V_1...V_k\}$ are *not* setwise *d*-unrestricted in *G*. Then, there exists a pair of nodes $\{S_i, S_j\} \subseteq \{V_1...V_k\}$ such that *G* exhibits the *d*-separation relation $(S_i \perp_{\mathrm{d}} S_j|\boldsymbol{Z})$. In this case, the distribution [50]

$$P(\boldsymbol{V}) = \tfrac{1}{2}(\delta_{\{V_1...V_k\},\vec{0}^k} + \delta_{\{V_1...V_k\},\vec{1}^k})$$

(11)

$$\delta_{\boldsymbol{V} \setminus \{V_1...V_k\},\vec{0}^{(|\boldsymbol{V}|-k)}}$$

(12)

violates the conditional independence relation $S_i \perp\!\!\!\perp_{\mathrm{CI}} S_j|\boldsymbol{Z}$, and therefore $P_{(10)} \notin \mathcal{I}_G$. For the other direction, note that if $(S_i \not\perp_{\mathrm{d}} S_j|\boldsymbol{Z})$ then so too $(\boldsymbol{S}_i \not\perp_{\mathrm{d}} \boldsymbol{S}_j|\boldsymbol{Z})$ for any disjoint sets $\boldsymbol{S}_i$ and $\boldsymbol{S}_j$ wherein $S_i \in \boldsymbol{S}_i$ and $S_j \in \boldsymbol{S}_j$. Consequently, when $\{V_1...V_k\}$ are setwise *d* unrestricted in *G*, there is no way to *d* separate *any* subset of $\{V_1...V_k\}$ from any other subset of $\{V_1...V_k\}$ by any subset of the observed nodes outside of $\{V_1...V_k\}$, which are the *only d* separation relations whose corresponding conditional independence relations would exclude
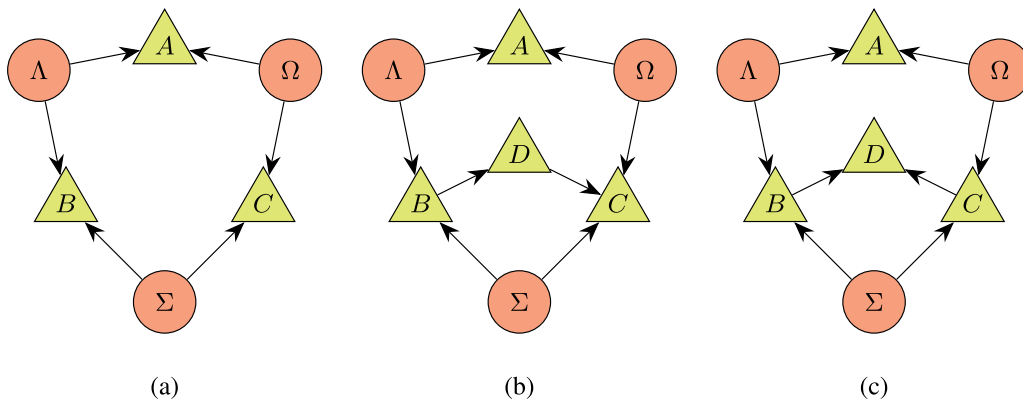
FIG. 9. Examples of maximal DAGs, which are setwise-nonmaximal. (a) is the Triangle scenario, where the set $\{A, B, C\}$ (all visible nodes) is setwise d unrestricted but not setwise adjacent. In (b), the set $\{A, B, C\}$ is setwise d unrestricted (in fact, there are no d-separation relations between observed nodes) but not setwise adjacent; by contrast, the larger set $\{A, B, C, D\}$ is *both* setwise d unrestricted *and* setwise adjacent, despite (b) exhibiting the d-separation relation ($A \perp_d D|B$), as that d-separation relation involves conditioning on a node *within the set*. In (c), both of the sets $\{A, B, C\}$ and $\{A, B, C, D\}$ are d unrestricted but not setwise adjacent. Note that $\{A, B, C, D\}$ is d unrestricted in the DAG (c) despite that DAG exhibiting the d-separation relation ($A \perp_d D|B, C$), as that d-separation relation involves conditioning on nodes *within the set*.

the distribution of Eq. (11). This means that, in this case, $P_{(10)} \in \mathcal{I}_G$. ∎

Putting Propositions 3 and 4 together lead us to a natural generalization of maximality, which we now define and employ in a theorem.

Definition 12 (Setwise maximal DAG). Let G be DAG. If every subset of the observed nodes of G, which is setwise d-unrestricted is also setwise adjacent then G is said to be *setwise maximal*, otherwise G is said to be *setwise nonmaximal*.

Theorem 7 [Setwise Nonmaximal]. Every ALGEBRAIC DAG is setwise-maximal, that is, every setwise-nonmaximal DAG is NON-ALGEBRAIC.

Proof. Theorem 7 follows immediately from Propositions 3 and 4. ∎

The DAGs of Fig. 9 are setwise-nonmaximal, so they are shown NON-ALGEBRAIC by Theorem 7 (even if they are *not* shown NON-ALGEBRAIC by Theorem 6). On the other hand, the mDAG of Fig. 7(a) is both maximal and setwise maximal, as all maximal DAGs are also setwise maximal.

Note that there are precisely five mDAGs with three observed nodes, which are NON-ALGEBRAIC (the Triangle scenario, the unrelated confounders scenario and three observationally equivalent versions of the Instrumental scenario). Theorem 7 certifies the NON-ALGEBRAICNESS of *all five*. Indeed, Theorem 7 turns out to be an extremely powerful filter for recognizing NON-ALGEBRAIC mDAGs with four or five observed nodes as well, as discussed in Sec. V.

C. Using d Separation to Prove NON-ALGEBRAICNESS

With Eq. (8), it was noted that any two DAGs that have different sets of observed d-separation relations are *not* classically observationally equivalent. In other words, imposing the same observed conditional independence constraints on the compatible distributions is a necessary condition for two DAGs to be classically observationally equivalent.

This fact can be used to establish the NON-ALGEBRAICNESS of some DAGs: If we can prove that the set of observed d-separation relations of a latent-permitting DAG *does not* match the set of d-separation relations of *any* latent-free DAG, then our latent-permitting DAG is classically observationally *inequivalent* to all latent-free DAGs. Via Evans' [25] Theorem 5, then, we can conclude that our latent-permitting DAG is NON-ALGEBRAIC.

This type of reasoning, that says that when a certain property of a DAG G is unmatched by all latent-free DAGs then G is NON-ALGEBRAIC, will be used a few times in this subsection and in the two next ones. As such, it will be useful to define the auxiliary term *Not Achievable in Latent-Free (NALF)*:

Definition 13 (NALF property of a DAG). Let G be a latent-permitting DAG. If G has a certain property that *does not match* that same property of any latent-free DAG, we say that this property of G is NALF*(not achievable in latent-free)*.

For example, we can have a DAG G whose set of observed d-separation relations is NALF. If there is a proof that a DAG is observationally inequivalent to all latent-free DAGs whenever certain property of the DAG is NALF, then this can be used to show NON-ALGEBRAICNESS. As discussed, this is the case for d separation.

Theorem 8 [NALF d-sep]. Let G be a DAG. Suppose that the set of observed d-separation relations of G is NALF (as per Definition 13). Then, G is NON-ALGEBRAIC.

Proof. From Eq. (8), G is not classically observationally equivalent to any latent-free DAG. Thus, via Theorem 5, G is NON-ALGEBRAIC. ∎

The mDAG presented in Fig. 7(a), which was not shown NON-ALGEBRAIC by Theorem 7 due to being setwise maximal, *can* be shown NON-ALGEBRAIC by Theorem 8: its set of observed d-separation relations, $A \perp_d D|\varnothing$ and $B \perp_d C|A$, is not matched by any latent-free DAG. This mDAG can be considered a special case of the bilocality scenario [51] restricted to have the same setting employed at both the extreme wings. Remarkably, it can support non-classical correlations

even though the "setting" for the extreme wings is the same (in stark contradiction with the Bell scenario).

On the other hand, note that the Evans scenario [Fig. 7(b)], which was shown NON-ALGEBRAIC by Theorem 6, cannot be shown NON-ALGEBRAIC via Theorem 8: it does not have any d-separation relation, just like the saturated latent-free DAGs. Therefore, Theorems 6 and 8 are not redundant to each other. Similarly, Theorems 7 and 8 are not redundant to each other.

There are a variety of different practical methods to ascertain whether or not Theorem 8 is satisfied for a given latent-permitting DAG G. Naively, one could construct *all* the latent-free DAGs with the same number of observed nodes as G, and one-by-one check whether any of them have observed d-separation relations matching those of G. Alternatively, one could envision employing a constraint-based causal discovery algorithm where the input to the algorithm is precisely the observed d-separation relations of G: If the output of the causal discovery algorithm fails to include *any* latent-free DAG as a viable explanation given the input constraints, then evidently G is NON-ALGEBRAIC per Theorem 8. While such "brute force" approaches are viable for a small number of observed nodes, as the number of observed nodes increases one would need to contend with potential combinatorial explosion. As it turns out, however, when the DAG is maximal (and thus not shown NON-ALGEBRAIC by Theorem 6) there is an efficient way to check whether or not its set of observed d-separation relations is NALF; the efficient algorithm is presented in Appendix C.

As well as our previous methods, Theorem 8 is only a *sufficient* condition for NON-ALGEBRAICNESS, and not necessary. If there *exists* a latent-free DAG H, which has the same observed d-separation relations as G (i.e., $\mathcal{I}_H = \mathcal{I}_G$), this *does not* imply that $\mathcal{C}_G = \mathcal{C}_H$, and as such we cannot conclude anything about the relation between $\mathcal{C}_G$ and $\mathcal{I}_G$. We will now discuss other sufficient conditions for NON-ALGEBRAICNESS that can be used when Theorems 6 and 8 fail.

D. Using e Separation to Prove NON-ALGEBRAICNESS

Just as the mismatch of d-separation relations is a witness of classical observational inequivalence, we can show that the mismatch of e-separation relations, a concept that will be defined below, also witnesses classical observational inequivalence. As such, we can mimic the same logic used in the last section, thus showing that e-separation relations can be used to attest NON-ALGEBRAICNESS. An e-separation relation is defined as

Definition 14 (e separation). Let G be a DAG, and let X, Y, Z, and W be four disjoint sets of nodes of G. Let G_{del_W} be the DAG obtained by starting from G and deleting the nodes of W. The sets X and Y are said to be *e separated by Z upon deletion of W* in G, denoted $(X \perp_e Y | Z)_{\mathrm{del}_W}$, if $X \perp_d Y | Z$ holds in G_{del_W}.

Note that if $W = \emptyset$, the concept of e separation reduces to d separation. As well as for the case of d separation, we will say that an e-separation relation is an *observed e-separation relation* when the sets X, Y, Z, and W only involve observed nodes. Matching observed e-separation relations is a prerequisite for observational equivalence:

Lemma 6. (e-separation condition for observational equivalence) Let G and H be two DAGs. If they are classically observationally equivalent (i.e., $\mathcal{C}_G = \mathcal{C}_H$), then their sets of observed e-separation relations must be identical.

Proof. In Ref. [47] it is shown that e-separation relations imply in inequalities that must be satisfied by the compatible probability distributions. In Appendix E of that same reference, it is further shown that if a DAG *does not* exhibit an e-separation relation, then there must exist a compatible probability distribution, which violates the inequality associated with that e-separation relation. This implies that, if a DAG G exhibits an e-separation relation, which is not exhibited by another DAG H, then it is possible to find a probability distribution that is compatible with H but not with G. ∎

Using this lemma, we can derive the analog of Theorem 8 for the case of e separation,

Theorem 9 [NALF e-sep)]. Let G be a DAG. Suppose that the set of observed e-separation relations of G is NALF (as per Definition 13). Then, G is NON-ALGEBRAIC.

Proof. Follows directly from Lemma 6 and Theorem 5. ∎

To use Theorem 9 in practice, one might enumerate the e-separation relations exhibited by every latent-free DAG with the same number of observed nodes of G and compare them to the observed e-separation relations of G. Far more efficiently, one need only check against any *one* latent-free graph, which matches the observed d-separation relations of G. (Such a latent-free DAG must exist if G is not already certified as NON-ALGEBRAIC by Theorem 8 [52]). After all, if two latent-free graphs share the same d-separation relations, they will also share the same e-separation relations, per Theorem 2 and Lemma 6. We thus advise verifying that a DAG G in question is not already certified as NON-ALGEBRAIC by Theorem 8 before invoking Theorem 9.

It is clear that every DAG that can be shown NON-ALGEBRAIC by Theorem 8 can also be shown NON-ALGEBRAIC by Theorem 9, since d separation is a special case of e separation. A little less trivial, every DAG that can be shown NON-ALGEBRAIC by Theorem 6 can also be shown NON-ALGEBRAIC by Theorem 9: all nonmaximal DAGs have a set of e-separation relations, which is NALF.

Proposition 5 (Theorem 9 subsumes Theorem 6). Let G be a nonmaximal DAG. Then, the set of e-separation relations of G is NALF.

Proof. If G is nonmaximal, then there are at least two nodes A and B, which are not d separable, but are also not adjacent in G. If they are not adjacent in G, it is clear that A is e separated from B by deletion of *every other* observed node of G. ∎

Let us make a proof by contradiction: Suppose that the set of e-separation relations of G *is not* NALF. Then, there is a latent-free DAG H, which has exactly the same set of e-separation relations as G. This means that A should be e separated from B by deletion of every other node of H, which implies that A and B are *not* adjacent in H. In a latent-free graph, two nodes are *nonadjacent* if and only if said pair of nodes are d separable [[43], Prop. 3.19]. However, if A and B are d separable in H but not in G, then their sets of d-separation relations do not coincide, which is a contradiction.

It is still an open question whether the conjunction of Theorems 6 and 8 is as good as Theorem 9 to show

NON-ALGEBRAICNESS. We did not find any DAG that was shown NON-ALGEBRAIC by Theorem 9 but not by any of these two previous methods. (See Appendix E for further discussion of this question.)

Note, however, that Theorem 9 *does not* subsume Theorem 7. For example, the triangle scenario [Fig. 9(a)], that is shown NON-ALGEBRAIC by Theorem 7, does *not* have any *e*-separation relation (just like the saturated latent-free DAGs).

Appendices A and B relate Theorems 6 and 9 to results of prior literature. In particular, in Appendix A we show that a version of Theorem 6 in terms of *e* separation that was presented in Ref. [27] is incorrect.

E. Using Incompatible Supports to Prove NON-ALGEBRAICNESS

The final method we used to prove NON-ALGEBRAICNESS is based on the classical feasibility of supports.

Given a set of random variables $\{X_1, \ldots, X_n\}$, a specific set $\{X_1 = x_1, \ldots, X_n = x_n\}$ of values that these random variables can take is called an *event*. The *support* $\mathcal{S}(P_{X_1,\ldots,X_n})$ of a probability distribution $P_{X_1,\ldots,X_n}$ over the variables $\{X_1, \ldots, X_n\}$ is the set of events that have nonzero probability,

$$\mathcal{S}(P_{X_1,\ldots,X_n}) = \{\{x_1, \ldots, x_n\} \mid P_{X_1,\ldots,X_n}(x_1, \ldots, x_n) > 0\}. \quad (13)$$

We previously defined what it means for a probability distribution to be classically compatible with a DAG. Here, we define what it means for a support to be classically compatible with a DAG:

Definition 15 (Compatibility of a support with a DAG). Let G be a DAG with observed nodes $\boldsymbol{A} = \boldsymbol{V} \cup \boldsymbol{L}$, where $\boldsymbol{V}$ are observed nodes and $\boldsymbol{L}$ are latent nodes. Let $\mathcal{S}$ be a set of events over the variables $\boldsymbol{V}$. We say that $\mathcal{S}$ is a support *classically compatible with G* if there exists a probability distribution P_V over $\boldsymbol{V}$ that is classically compatible with G (i.e., $P_V \in \mathcal{C}_G$) and whose support is $\mathcal{S}(P_V) = \mathcal{S}$. We say that $\mathcal{S}$ is a support *compatible-up-to-CI* with G if there exists a probability distribution P_V over $\boldsymbol{V}$ such that $P_V \in \mathcal{I}_G$ and whose support is $\mathcal{S}(P_V) = \mathcal{S}$.

As an example, the following support is *not* compatible with the Bell DAG (Fig. 2), as it corresponds to the Popescu-Rohrlich box [53,54],

$$\mathcal{S}_{\text{Bell}} = \begin{cases} \{X = 0, Y = 0, A = 0, B = 0\} \\ \{X = 0, Y = 0, A = 1, B = 1\} \\ \{X = 0, Y = 1, A = 0, B = 0\} \\ \{X = 0, Y = 1, A = 1, B = 1\} \\ \{X = 1, Y = 0, A = 0, B = 0\} \\ \{X = 1, Y = 0, A = 1, B = 1\} \\ \{X = 1, Y = 1, A = 1, B = 0\} \\ \{X = 1, Y = 1, A = 0, B = 1\} \end{cases}. \quad (14)$$

Note that if a support *is* compatible with DAG G, that does *not* mean that every distribution with that support will be compatible with G. There are countless counterexamples, but let us simply note that the full support (the one where all events have positive probability) is compatible with any DAG, but at the same time we know of many incompatible distributions, which nevertheless have full support.

Naturally, admitting the same set of compatible supports is a prerequisite for two DAGs to admit the same set of compatible distributions:

Lemma 7 (Supports condition for observational equivalence). Let G and H be two DAGs. If they are classically observationally equivalent (i.e., $\mathcal{C}_G = \mathcal{C}_H$), then their sets of classically compatible supports must be identical.

It remains an open question whether the condition of Lemma 7 is also necessary for observational equivalence. In particular, it is not known whether or not there exists a DAG for which some distributions are incompatible (due to inequalities) but for which all *supports* are compatible.

As before, this necessary condition for observational equivalence immediately translates into a method for proving NON-ALGEBRAICNESS:

Theorem 10 [NALF Supports]. Let G be a DAG. Suppose the set of classically compatible supports of G is NALF (as per Definition 13). Then, G is NON-ALGEBRAIC.

To exploit Theorem 10 in practice, we need an an algorithm capable of assessing whether or not a given support is compatible with a given DAG. Such algorithm was developed in Ref. [26], and is referred to here as *Fraser's algorithm*. We have implemented Fraser's algorithm in Python and scripted it to yield *all* the supports that are classically incompatible with a given DAG (for a certain assignment of the cardinalities of the observed variables).

In general, Fraser's algorithm is much more computationally expensive than simply assessing whether or not a graph exhibits some *d*-separation or *e*-separation relation. Consequently, we consider Theorem 10 a method of last resort to show NON-ALGEBRAICNESS.

1. Rapidly testing supports (without comparing to any latent-free graph)

As well as for the case of *e* separation, Theorem 10 has the downside that it requires one to find the supports compatible with the DAG G and then check the compatible supports of all the latent-free DAGs with the same number of observed nodes (or alternatively to find the latent-free H that has the same set of *d*-separation relations as G, and then check which supports are compatible with H). Since Fraser's algorithm is computationally expensive, doing this in practice can be cumbersome.

Luckily, it is possible to develop a rapid supports test where it is not even necessary to find *all* of the supports compatible with G. The idea of the rapid supports test comes from noting that sometimes we can prove the *incompatibility* of a given support with a DAG G by recognizing that the given support conflicts with a *d*-separation relation exhibited by G.

Suppose, for instance, that a DAG G has the (unconditional) *d*-separation relation $A \perp_d B$. Then, the support in Eq. (15) is clearly incompatible with G, since any probability distribution with that support will must have $P_{AB}(1, 1) = 0 \neq P_A(1)P_B(1) > 0$, contradicting $A \perp_{\text{CI}} B$,

$$\mathcal{S}_{10} = \begin{cases} \{A = 0, B = 0\} \\ \{A = 0, B = 1\} \\ \{A = 1, B = 0\} \end{cases}. \quad (15)$$

Indeed, we can formally categorize all such "trivial" proofs of support incompatibility through the following two definitions:

Definition 16 (Support conflicting with a conditional independence relation). Let $\mathcal{S}$ be a support over a set $\boldsymbol{V}$ of variables, and let $\boldsymbol{A} \subseteq \boldsymbol{V}$, $\boldsymbol{B} \subseteq \boldsymbol{V}$, and $\boldsymbol{C} \subseteq \boldsymbol{V}$ be three disjoint subsets of $\boldsymbol{V}$. We say that $\mathcal{S}$ *conflicts with the conditional independence relation* $\boldsymbol{A} \perp\!\!\!\perp_{\text{CI}} \boldsymbol{B}|\boldsymbol{C}$ if there exists a set $\{\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}\}$ of values of the variables in $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$ such that the events $\{\boldsymbol{A} = \boldsymbol{a}, \boldsymbol{C} = \boldsymbol{c}\}$ and $\{\boldsymbol{B} = \boldsymbol{b}, \boldsymbol{C} = \boldsymbol{c}\}$ occur in $\mathcal{S}$, but the event $\{\boldsymbol{A} = \boldsymbol{a}, \boldsymbol{B} = \boldsymbol{b}, \boldsymbol{C} = \boldsymbol{c}\}$ does not occur in $\mathcal{S}$.

For example, the support of Eq. (15) conflicts with the conditional independence relation $A \perp\!\!\!\perp_{\text{CI}} B$: both the events $A = 1$ and $B = 1$ occur in the support, but the event $\{A = 1, B = 1\}$ does not.

If a support $\mathcal{S}$ conflicts with a conditional independence relation $A \perp\!\!\!\perp_{\text{CI}} \boldsymbol{B}|\boldsymbol{C}$, then there is no probability distribution with support $\mathcal{S}$ that obeys $A \perp\!\!\!\perp_{\text{CI}} \boldsymbol{B}|\boldsymbol{C}$. This will be seen explicitly in the proof of Lemma 8.

Definition 17 (Triviality of support incompatibility). A support $\mathcal{S}$ is said to be *trivially incompatible* with a given DAG whenever the DAG exhibits some d-separation relation whose associated conditional independence relation conflicts with $\mathcal{S}$ (as in Definition 16).

By generalizing the discussion made around Eq. (15), we see that this definition indeed implies in classical incompatibility of the support with the DAG:

Lemma 8 (Trivial incompatibility implies incompatibility). If a support $\mathcal{S}$ is trivially incompatible with a DAG G, then it is classically incompatible with G.

Proof. Let $\boldsymbol{A} \perp\!\!\!\perp_{\text{d}} \boldsymbol{B}|\boldsymbol{C}$ be a d-separation relation of G, which is in conflict with $\mathcal{S}$. Furthermore, let $\{\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}\}$ be a set of values of the variables in $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$ that witnesses this conflict. This means that all of the probability distributions, which have the support $\mathcal{S}$ must have $P_{\boldsymbol{AB}|\boldsymbol{C}}(\boldsymbol{ab}|\boldsymbol{c}) = 0 \neq P_{\boldsymbol{A}|\boldsymbol{C}}(\boldsymbol{a}|\boldsymbol{c})P_{\boldsymbol{B}|\boldsymbol{C}}(\boldsymbol{b}|\boldsymbol{c}) > 0$. Therefore, $\mathcal{S}$ is not classically compatible with G. ∎

The key insight, which allows us to accelerate the application of Theorem 10 is that for a latent-free DAG, the *only* supports incompatible with it are those which are *trivially* incompatible with it.

Lemma 9 (Latent-free support compatibility). Let H be a latent-free DAG. If $\mathcal{S}$ is a support, which is not trivially incompatible with H as per Definition 17, then $\mathcal{S}$ is classically compatible with H as per Definition 15 [55].

Proof. Let $X_1, \ldots, X_n$ be the nodes of H in some topological order, i.e., an order where X_i is a nondescendant of X_{i+1}. We will make this proof by explicitly constructing a probability distribution $P(X_1, \ldots, X_n)$, which is classically compatible with H and has the support $\mathcal{S}$.

The explicit construction is given by $P(X_1, \ldots, X_n) = \prod_i P(X_i|\text{PA}_H(X_i))$, where each $P(X_i|\text{PA}_H(X_i))$ is uniformly distributed over the values of X_i which occur along with the given values of $\text{PA}_H(X_i)$ (and any value of the remaining variables) in $\mathcal{S}$. It is clear that this distribution is classically compatible with H, since it takes the form of the Markov condition. Now, we will show that its support is in fact $\mathcal{S}$.

We will make a proof by induction on m that this distribution has the correct support on $X_1, \ldots, X_m$. For $m = 1$

we simply have $P(X_1)$ uniformly distributed on values of X_1 that are possible under $\mathcal{S}$, so the basis case is immediately satisfied. Now, we assume that $X_1, \ldots, X_k$ has the correct support, and we will prove that $X_1, \ldots, X_{k+1}$ also does.

First, consider some $(x_1, \ldots, x_k, x_{k+1})$ that occurs in $\mathcal{S}$. Since we assumed that $X_1, \ldots, X_k$ has the correct support, we know that the corresponding $P(X_1, \ldots, X_k)$ is nonzero. All of the parents of X_{k+1} are elements of $\{X_1, \ldots, X_k\}$; therefore, by definition, the corresponding $P(X_{k+1}|\text{PA}_H(X_{k+1}))$ is also nonzero. Therefore, $P(X_1, \ldots, X_{k+1}) = P(X_1, \ldots, X_k)P(X_{k+1}|\text{PA}_H(X_{k+1}))$ is nonzero as required.

Now, consider some $(x_1, \ldots, x_k, x_{k+1})$ that *does not* occur in $\mathcal{S}$. There are two possibilities: the set of values $(x_1, \ldots, x_k)$ could occur or not occur in $\mathcal{S}$. If $(x_1, \ldots, x_k)$ also does not occur in $\mathcal{S}$, the proof is simple: by the inductive hypothesis, the corresponding $P(X_1, \ldots, X_k)$ is zero, so $P(X_1, \ldots, X_{k+1})$ has to be zero as required.

Suppose now that $(x_1, \ldots, x_k, x_{k+1})$ does not occur in $\mathcal{S}$ but $(x_1, \ldots, x_k)$ *occurs*. Let $C_1, \ldots, C_p$ denote the parents of X_{k+1} and $B_1, \ldots, B_{k-p}$ the remaining variables among $X_1, \ldots, X_k$, which we know are nondescendants of X_{k+1} in H. The DAG H must have the d-separation relation $X_{k+1} \perp\!\!\!\perp_{\text{d}} \{B_1, \ldots, B_{k-p}\}|\{C_1, \ldots, C_p\}$; since $\mathcal{S}$ is *not* trivially incompatible with H, then it must *not* be in conflict with the associated conditional independence relation. Since $(x_1, \ldots, x_k) = (b_1, \ldots, b_{k-p}, c_1, \ldots, c_p)$ occurs in $\mathcal{S}$ but $(x_1, \ldots, x_k, x_{k+1}) = (b_1, \ldots, b_{k-p}, c_1, \ldots, c_p, x_k + 1)$ does not occur, we can then conclude that $(x_{k+1}, c_1, \ldots, c_p)$ *must not occur* in $\mathcal{S}$. Therefore, by definition the associated $P(X_{k+1}|\text{PA}_H(X_{k+1})) = P(X_{k+1}|C_1, \ldots, C_p)$ is zero and hence so is $P(X_1, \ldots, X_k, X_{k+1})$, as required. ∎

Accordingly, we have the following upgrade to Theorem 10:

Theorem 11 (Rapid Supports). Let G be a DAG. If there is a support $\mathcal{S}$, which is incompatible with G despite *not* being trivially incompatible with G, then G is NON-ALGEBRAIC.

Proof. We start by noting that if G was ALGEBRAIC, then it would be classically observationally equivalent to some latent-free DAG H, per Theorem 5. If so, then H and G would (at least!) share the same d-separation relations, and hence a support would only *not* be trivially incompatible with G if was not trivially incompatible with H. But by Lemma 9, if a support is not trivially incompatible with H then it must be compatible with H. Since our starting premise is that this support is not compatible with G, then by Lemma 7 it follows that $\mathcal{C}_G \neq \mathcal{C}_H$, and hence G is NON-ALGEBRAIC. ∎

Note that Theorem 11 allows us to leverage tools distinct from Fraser's algorithm for assessing support incompatibility. Fraser's algorithm is a *necessary and sufficient* test for support compatibility. For the purposes of Theorem 11, however, we can instead consider variant algorithms related to Inflation [56], which cannot certify support compatibility but which can often efficiently detect support incompatibility. Some such algorithms are discussed in Ref [57], for example.

It is clear that the application of Theorem 11 is much more efficient than the application of Theorem 10. We can also show that both theorems are equally powerful:

Proposition 6. Let G be a NON-ALGEBRAIC DAG. If the NON-ALGEBRAICNESS of G can be shown via Theorem 10, then it can also be shown via Theorem 11.

Proof. First, suppose that there exists a latent-free DAG H such that $\mathcal{I}_H = \mathcal{I}_G$. If G is NON-ALGEBRAIC, then $C_G \subsetneq C_H$, and hence every support, which is incompatible with G must also be incompatible with H. If Theorem 10 shows the NON-ALGEBRAICNESS of G, then there must be a support, which is *compatible* with H but not G. Since the set of supports incompatible with H is exactly the set of trivially incompatible supports as per Lemmas 8 and 9, it follows that the support, which is incompatible with G but not H must *not* be trivially incompatible with G. Therefore, the NON-ALGEBRAICNESS of G can be shown by Theorem 11.

Now, suppose that there is *no* latent-free DAG that has the same set of d-separation relations as G, i.e., that the set of d-separation relations of G is NALF. If G is not maximal, in the proof of Theorem 6 we showed a distribution, which is not compatible with G [Eq. (9)]. In reality, *all* the distributions, which have the same support as this one will be incompatible with G. Furthermore, since Eq. (9) is an element of $\mathcal{I}_G$, this support is not trivially incompatible with G. Therefore, the support of the distribution of Eq. (9) is an incompatible support, which is not trivially incompatible.

If the set of d-separation relations of G is NALF and G *is* maximal, per Theorem 17 we know that G has one of the eighteen mDAGs of Fig. 12 as a subgraph. One can explicitly check that these eighteen mDAGs *have* incompatible supports that are not trivially incompatible; therefore, G itself also must have incompatible supports that are not trivially incompatible, namely, by taking all its observed variables outside of the pertinent subgraph to be point distributed. Therefore, this case also falls under the scope of Theorem 11. ∎

It is also worth noting that, when there *is* a latent-free DAG H with the same set of d-separation relations as G, Theorem 11 is *constructive*: The distribution P constructed for H in the proof of Lemma 9 is such that $P \in \mathcal{I}_G$ yet $P \notin C_G$.

An example of a support, which is incompatible—but not *trivially* incompatible—with the Evans scenario [Fig. 7(b)] is the following:

$$\mathcal{S}_{\text{Evans}} = \left\{ \begin{array}{l} \{C = 0, D = 0, E = 0\} \\ \{C = 0, D = 1, E = 1\} \end{array} \right\}. \qquad (16)$$

The Evans scenario does not have any d-separation relation. Nevertheless, the support $\mathcal{S}_{\text{Evans}}$ of Eq. (16) is not compatible with it. This can be seen by noting that the variable C in $\mathcal{S}$ is associated with a point distribution: it always takes the value 0. Since in the Evans scenario all the correlation between D and E is established through C, it is impossible to have perfect correlation between D and E while C takes a point distribution.

An example of DAG whose NON-ALGEBRAICNESS was first certified in Ref. [26] via the discovery of an incompatible support is presented in Fig. 10. By means of his eponymous algorithm for compatible supports, Fraser showed that the following support is classically incompatible with the DAG
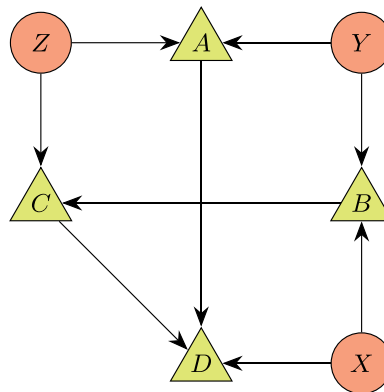

FIG. 10. A DAG with four observed nodes and seven total nodes whose NON-ALGEBRAICNESS can be shown by Fraser's algorithm for compatible supports.

of Fig. 10[58],

$$\mathcal{S}_{10} = \left\{ \begin{array}{l} \{A = 0, B = 0, C = 0, D = 0\} \\ \{A = 0, B = 0, C = 0, D = 1\} \\ \{A = 0, B = 1, C = 0, D = 0\} \\ \{A = 1, B = 0, C = 1, D = 0\} \end{array} \right\}. \qquad (17)$$

When using Theorem 11 to attest NON-ALGEBRAICNESS, we start by checking supports at binary cardinalities of the observed variables. If such an incompatible but not trivially incompatible support is found, we can try to search for such a support at higher cardinalities of the observed variables.
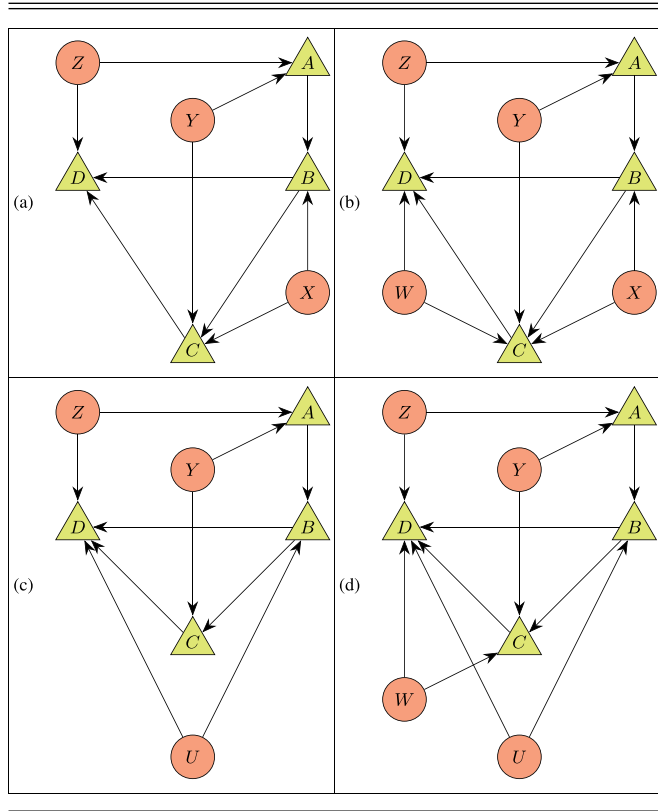
Remark 1. We anticipated that if a DAG has *any* incompatible support (for *any* cardinality), then it seemed likely that we should expect to find *some* incompatible support where all variables have binary cardinality. Indeed, prior to this study, we are not aware of any counterexample. Even the three challenging DAGs identified in Fig. 14 of Ref. [24] were eventually found to have incompatible supports with merely binary variables. However, this intuition turns out to be misplaced: We identified four mDAGs for which the high-cardinality support of Eq. (18) is identified as incompatible, but where nevertheless *every* support over binary variables is provably compatible. These mDAGs are depicted in Table II.

$$\mathcal{S}_{\text{for Table II}} = \left\{ \begin{array}{l} \{A = 0, B = 0, C = 0, D = 0\} \\ \{A = 0, B = 0, C = 1, D = 0\} \\ \{A = 0, B = 1, C = 0, D = 0\} \\ \{A = 1, B = 0, C = 0, D = 0\} \\ \{A = 1, B = 1, C = 0, D = 0\} \\ \{A = 2, B = 0, C = 0, D = 1\} \\ \{A = 2, B = 1, C = 1, D = 0\} \end{array} \right\}. \qquad (18)$$

2. Incompatible supports subsumes all other methods

The methods to show NON-ALGEBRAICNESS that we presented here are not independent of each other. For example, it is clear that Theorem 9 subsumes Theorem 8, and Proposition

TABLE II. The only four mDAGs with four observed nodes for which every support over binary observed variables is classically compatible but which are nevertheless provably Non-Algebraic by virtue of the higher-cardinality support given in Eq. (18) being incompatible with all four DAGs here.
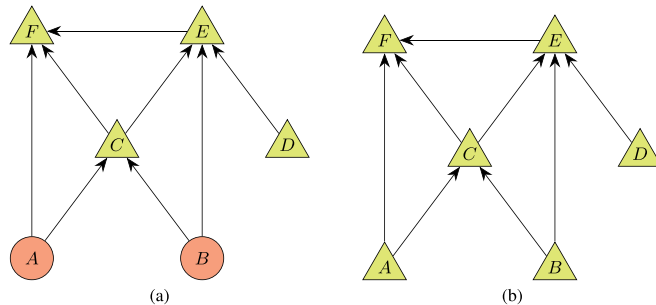

FIG. 11. (a) Depicts a scenario characterized correctly by Pienaar's theorem as Non-Algebraic, whereas (b) depicts an Algebraic scenario which Pienaar's theorem incorrectly deems Non-Algebraic.

as *G*, then there is also no latent-free DAG, which presents the same set of classically compatible supports as *G*. In other words, if *G* can be shown Non-Algebraic by Theorem 9, then it can also be shown Non-Algebraic by Theorem 10.

Since Proposition 6 shows that the rapid supports test (Theorem 11) can show Non-Algebraicness in all the cases covered by Theorem 10, from Corollary 2 we can conclude that Theorem 11 is the most powerful tool we have so far to show Non-Algebraicness. It is still an open question whether detection by Theorem 11 is a necessary condition for Non-Algebraicness.

Note that Theorem 11 also subsumes Theorem 7, as all distributions *with same support* of that in Eq. (10) are in $\mathcal{I}$ but not $\mathcal{C}$, and hence every DAG, which is provably Non-Algebraic via Theorem 11 will also be provably Non-Algebraic via Theorem 11. In particular, merely by considering compatible versus incompatible supports *over two events*, Theorem 11 already subsumes Theorem 7.

V. COMPUTATIONAL RESULTS

We consider the problem of certifying the Non-Algebraicness of those mDAGs of four observed nodes, which are not shown Algebraic by the HLP criterion (Corollary 1). We start by the method described in Sec. IV A, that says that nonmaximal DAGs are Non-Algebraic. Among the 996 mDAGs, which are left as *potentially* interesting after applying the HLP criterion, we find that 810 are nonmaximal. Therefore, nonmaximality seems to be a powerful tool to show Non-Algebraicness; at this stage, we are left with only 186 unresolved mDAGs after this preliminary filtering.

We then exploit the *d*-separation test for Non-Algebraicness per Sec. IV C. That is, among those 186 remaining mDAGs we filter out any mDAGs, which possess observed *d*-separation relations *not* matching those of some latent-free DAG (NALF *d*-separation relations). We find that 168 mDAGs remain as-yet unresolved—the 18 mDAGs that are shown Non-Algebraic at this stage are the ones presented in Fig. 12.

The *e*-separation test for Non-Algebraicness presented in Sec. IV D does not resolve *any* of these 168 unresolved cases. As mentioned, it is still an open problem whether the nonmaximality test and the *d*-separation test together will always subsume the *e*-separation test. We then turn to the

5 shows that Theorem 9 also subsumes Theorem 6. However, it is more efficient to start by checking maximality of the DAG of interest or the *d*-separation relations of latent-free DAGs instead of directly checking all of their *e*-separation relations; this is the reason why we presented these results separately.

As it turns out, we can also show that Theorem 10 subsumes Theorem 9: every DAG that can be shown Non-Algebraic via its *e*-separation relations (through Theorem 9) can also be shown Non-Algebraic via its incompatible supports (through Theorem 10). This result will be presented here as Corollary 2. All of the discussion about *e* separation was made here because checking *e*-separation relations is in general much faster than checking incompatible supports.

Theorem 12 (Supports subsumes e separation for observational inequivalence). Let *G* and *H* be two DAGs. If their sets of observed *e*-separation relations are different, then it is possible to find a support *over binary variables,* which is compatible with one of them but incompatible with the other.

Proof. Presented in Appendix D. ∎

Note that Theorem 12 is a generic result about observational equivalence between two DAGs, not restricted to the comparison with latent-free DAGs. A direct corollary is

Corollary 2 (Supports subsumes e separation for Non-Algebraicness). Let *G* be a DAG. If there is no latent-free DAG, which presents the same set of *e*-separation relations
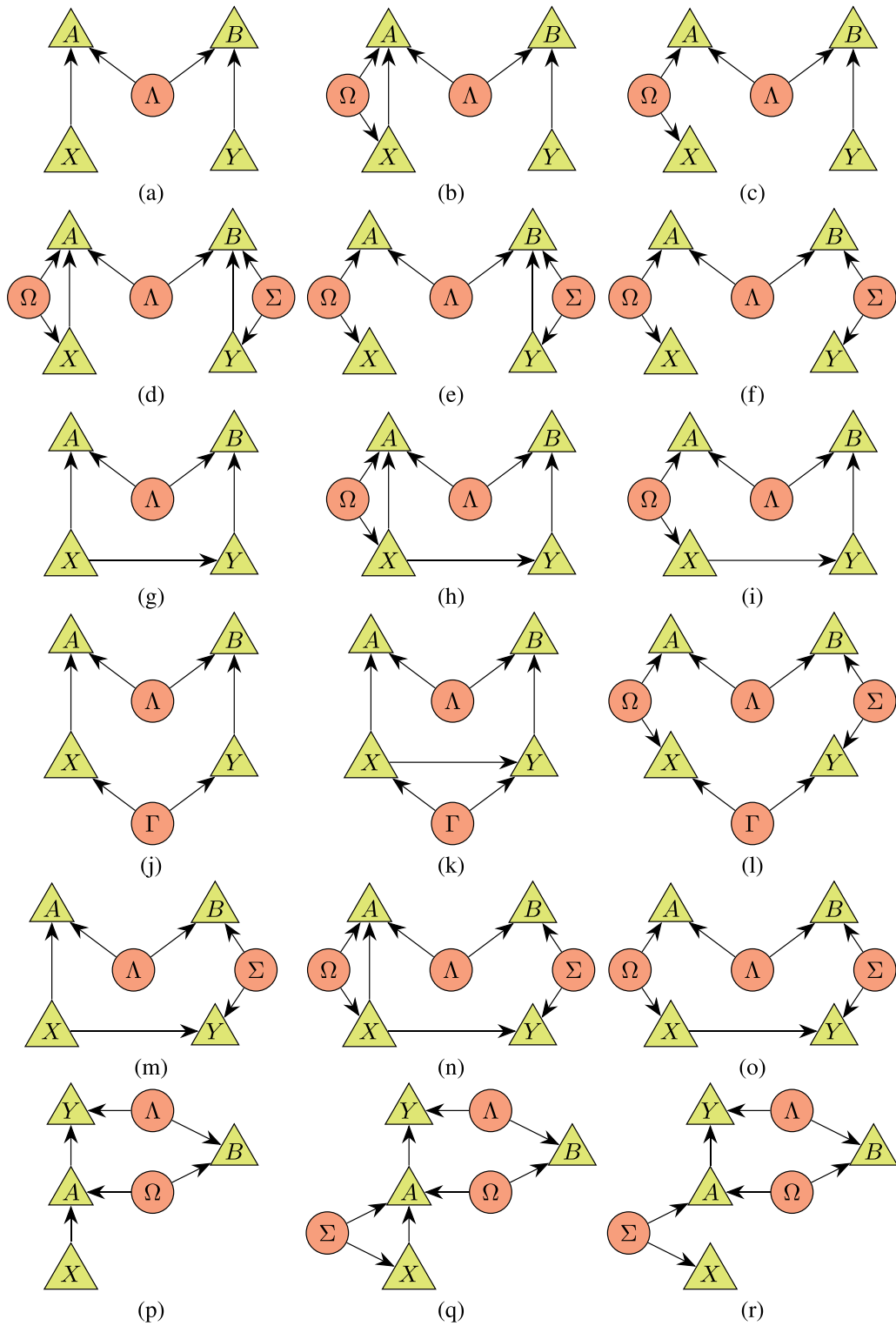
FIG. 12. A maximal DAG has a set of *d*-separation relations unmatched by any latent-free DAG (and is thus Non-Algebraic by Theorem 8) if and only if it has one of these eighteen patterns as a subgraph.

method of supports analysis per Sec. IV E, which ultimately leaves us with only three remaining unresolved mDAGs.

More specifically, by considering supports with *binary* cardinalities of the observed variables we were able to certify the Non-Algebraicness of 161 out of the 168

remaining mDAGs. We could not, however, find *any* classically incompatible supports for the remaining seven mDAGs when only considering binary cardinality variables. But by increasing the cardinality of variable *A* to be three we were able to identify a support that is incompatible—but not

trivially incompatible—with the four mDAGs in Table II, hence certifying their NON-ALGEBRAICNESS. Said support is explicitly reproduced in Eq. (18); it is obviously not trivially incompatible with any of the mDAGs in Table II, since none of those four mDAGs exhibits any *d*-separation relations over its observed nodes. The remaining three mDAGs, which we were *unable* to resolve as NON-ALGEBRAIC via supports considerations—up to computational tractability limits—are depicted in Table III.

The exact number of mDAGs that we are able to characterize as NON-ALGEBRAIC or ALGEBRAIC at each stage is summarized in Table I.

On the Potential NON-ALGEBRAICNESS of the Remaining three mDAGs

The three as-yet unresolved mDAGs with four observed nodes are depicted in Table III. For these three mDAGs we could not find *any* incompatible supports [29], at least up to the small cardinalities of the observed nodes that we checked. Searching for incompatible supports at higher cardinalities of the observed nodes using Fraser's algorithm is computationally expensive, as the algorithm's complexity increases significantly on increasing the cardinalities. Perhaps future acceleration of Fraser's algorithm may allow us to probe supports for higher cardinalities. For the present paper, however, we considered one final attempt to prove [59] the NON-ALGEBRAICNESS of these three mDAGs, namely, by exploring entropic inequalities. For a more comprehensive introduction to entropic inequalities and Shannon cones see Refs. [60,61].

In particular, we attempted to isolate some Shannon-type inequalities that constrain $\mathcal{C}_G$, but are not Shannon-type inequalities for $\mathcal{I}_G$. This analysis is also computationally expensive—often intractable—as generating the Shannon-type inequalities corresponding to $\mathcal{C}$ is accomplished via linear quantifier elimination. The complexity of the most common algorithm for performing linear quantum elimination is doubly exponential in the number of eliminated variables, although alternative algorithms have different complexities [62].

We can nevertheless certify that the Shannon cone corresponding to $\mathcal{C}_G$ and $\mathcal{I}_G$ are indistinguishable for these three remaining mDAGs *without* explicitly constructing the Shannon cone for $\mathcal{C}_G$. We do so as follows:

(1) From all the Shannon type inequalities corresponding to $\mathcal{I}_G$, generate the extremal rays of this cone.

(2) Check whether each of those extremal rays is *implicitly* contained in the Shannon cone corresponding to $\mathcal{C}_G$ by asking if the Shannon-type inequalities over *all* the variables (thus not using linear quantifier elimination) corresponding to $\mathcal{C}$ are *satisfiable* by the given extremal ray. If every extremal ray of the Shannon cone of $\mathcal{I}_G$ is contained in the Shannon cone corresponding to $\mathcal{C}_G$, then those two Shannon cones coincide.

For the three mDAGs in Table III we find that the Shannon cones corresponding to $\mathcal{C}_G$ and $\mathcal{I}_G$ are the same. That is, we were unable to find any valid Shannon type inequality for $\mathcal{C}_G$ that is not also a Shannon type inequality for $\mathcal{I}_G$ for these three mDAGs. Thus, entropic methods are incapable of proving the NON-ALGEBRAICNESS of these three mDAGs,

TABLE III. The mDAGs of four observed nodes whose NON-ALGEBRAICNESS remains unresolved.
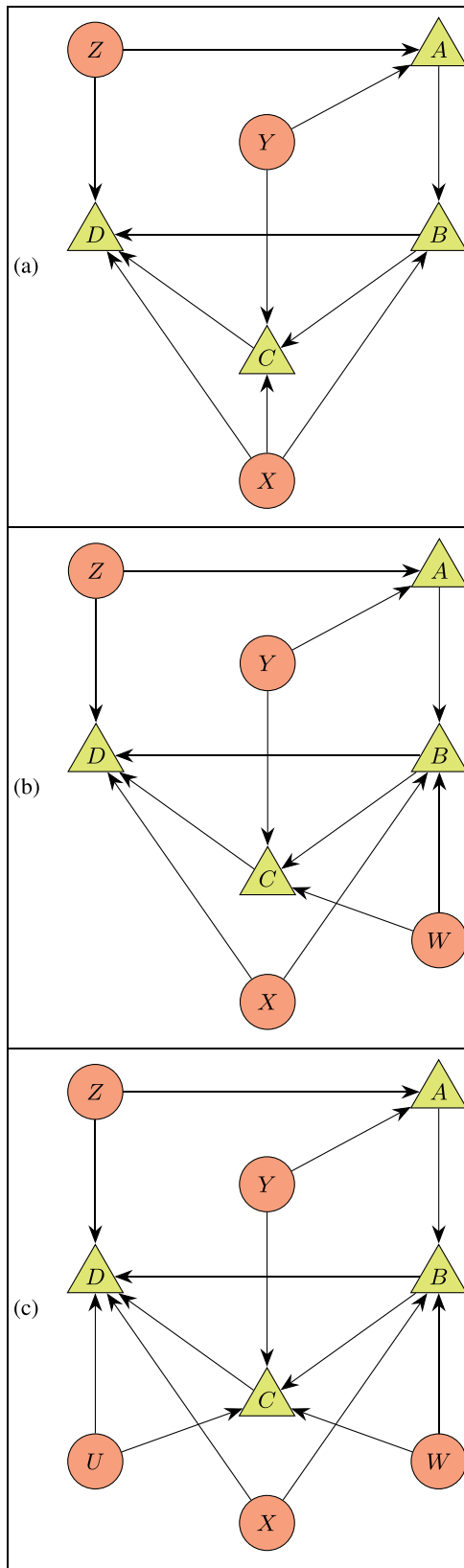
unless perhaps we explore non-Shannon-type inequalities or entropic inequalities involving non-Shannon entropies.

VI. CONCLUSIONS

In this paper, we contributed to causal investigation by categorizing which causal structures of four observed nodes present inequality constraints or not. To do so, we developed a plethora of techniques to prove that a causal structure is NON-ALGEBRAIC (has inequality constraints), while we used one single technique to prove that a causal structure is ALGEBRAIC: the HLP criterion.

As can be seen from Table I, out of the 2809 mDAGs with four observed nodes, the HLP criterion shows that that 1813 are ALGEBRAIC. Out of the remaining 996 mDAGs, our techniques showed 993 of them to be NON-ALGEBRAIC, while we are still uncertain about the status of three mDAGs (presented in Table III). While these three remaining mDAGs are still potential counter-examples to the HLP conjecture (which says that the HLP criterion is necessary and sufficient for ALGEBRAICNESS), we believe that our numerical results are a hint towards the validity of this conjecture. A truly thorough analysis of all mDAGs with five observed nodes proved to be quite computationally demanding. Nevertheless, in Appendix E we show that—among those five-node mDAGs, which the HLP criterion fails to certify as NON-ALGEBRAIC—*at least* 99% are NON-ALGEBRAIC, which we again elect to interpret as at least *consistent* with the HLP conjecture.

It is also interesting to note that all of our techniques to show NON-ALGEBRAICNESS give explicit constructions of distributions, which are in $\mathcal{I}_G$ but not in $\mathcal{C}_G$, i.e., respect the conditional independence constraints of DAG G but not its inequality constraints. Theorem 6 is related to the construction of Eq. 9, as well as Theorem 7 is related to the construction of Eq. 10. Theorem 11 is constructive whenever there *is* a latent-free DAG H with the same set of d-separation relations as G (such construction can then be found in the proof of Lemma 9). If G is maximal but there is *no* latent-free DAG H with the same set of d-separation relations as G, then Theorem 17 says that G has one of the eighteen DAGs of Fig. 12 as a subgraph. In the end of Appendix C, an explicit distribution, which is in $\mathcal{I}_G$ but not in $\mathcal{C}_G$ for these eighteen DAGs is presented: it is the uniform distribution over the events in the Popescu-Rohrlich support presented in Eq. (14).

In particular, our Theorem 6, which showed itself to be very powerful in proving NON-ALGEBRAICNESS, is a corrected version of the e-separation theorem of [27] (as discussed in Appendix A).

By showing practical tools to attest that a causal structure presents inequality constraints, this paper simultaneously contributes to purely classical causal inference and advances the question of which causal scenarios might exhibit quantum or post-quantum advantage.

APPENDIX A: PIENAAR'S e-SEPARATION THEOREM IS INCORRECT

In Ref. [[27], Lemma 2] Pienaar presented the following claim, which we demonstrate to be incorrect:

Theorem 13 (Pienaar's incorrect e-separation theorem). Let G be a DAG, and let X, Y, Z and W be disjoint sets of nodes of G such that none of the nodes of Z is a descendant of a node in W and $(X \perp_e Y | Z)_{\text{del}_W}$. Then G is NON-ALGEBRAIC *if and only if* the d-separation relations of G *do not* include any relations of the form $(X \perp_d Y | ZS)$, where S is a subset of W.

If this theorem were true, then the problem of classifying NON-ALGEBRAICNESS would be completely solved whenever a suitable e separation is present, as Theorem 13 claims to provides a necessary and sufficient condition for such DAGs that is simple to check. However, the condition turns out to be *neither* necessary *nor* sufficient.

Firstly, note that there are plenty of well-known graphs, which are NON-ALGEBRAIC despite not satisfying the conditions of Theorem 13. Examples include the Bell DAG of Fig. 2, as well as the mDAG in Fig. 7(a), among many others.

Pienaar notably included similar examples in his own study [27], and therefore we believe the inclusion of the "only if" language in Theorem 13 was an oversight, in that Pienaar himself never actually intended to communicate that, but rather only that the e-separation relation $(X \perp_e Y | Z)_{\text{del}_W}$ would automatically follow from the d-separation relation $(X \perp_d Y | ZS)$ if that d-separation relation was present in the graph and where S is a subset of W.

Regardless, the "if" direction in Theorem 13 *also* turns out to be invalid, although it is a bit more subtle. Consider the DAG in Fig. 11(a); Pienaar [27] uses it as an example of a DAG that is deemed (in this case, *correctly* deemed) as NON-ALGEBRAIC pursuant to Theorem 13. In this DAG, $(F \perp_e D | C)_{\text{del}_E}$ holds, and node C is not a descendant of node E, and neither $(F \perp_d D | C)$ nor $(F \perp_d D | CE)$ holds true. Therefore, Theorem 13 classifies the DAG of Fig. 11(a) as NON-ALGEBRAIC. But now consider the DAG of Fig. 11(b), which has the same *structure* as the DAG in Fig. 11(a), with the only difference being that the DAG in Fig. 11(b) has no latent variables. All the conditions of Theorem 13 are again met, however, when applied to Fig. 11. Again, $(F \perp_e D | C)_{\text{del}_E}$ holds, and node C is not a descendant of node E, and neither $(F \perp_d D | C)$ nor $(F \perp_d D | CE)$ hold true. So this DAG is again—but this time, wrongly—characterized as NON-ALGEBRAIC by Pienaar's theorem. As previously discussed, all latent-free DAGs are ALGEBRAIC!

In fact, *every* DAG for which the conditions of Theorem 13 hold can be converted into a different latent-free DAG for which the conditions of the theorem would

continue to hold by making all the nodes observed. That is, for every DAG, which is correctly classified as NON-ALGEBRAIC by the (invalid) condition formulated as Theorem 13 one can construct a latent-free counterexample to Theorem 13.

The flaw with Pienaar's proof of the "if" direction is that it invokes a probability distribution that may not actually be in $\mathcal{I}$, in that the constructed distribution may be inconsistent with some d-separation relation *not mentioned in the statement of Theorem 13*. For example, in Fig. 11, Pienaar's proof would invoke a distribution, which posits perfect correlation between D and F while all other variables are point-distributed. While such a distribution is consistent with *all* the observable d-separation relations of Fig. 11(a), it is nevertheless *inconsistent* with the d-separation relation $(F \perp_d D|BCE)$ exhibited by Fig. 11(b).

This loophole in Pienaar's proof can be closed by *strengthening the conditions of Pienaar's theorem*, namely, to exclude $(F \perp_d D|S)$ for *any* subset S of the remaining observed nodes. In other words,

Theorem 14 (A corrected version of Pienaar's e-separation theorem). Let G be a DAG, and let X, Y, Z and W be disjoint sets of nodes of G such that none of the nodes of Z is a descendant of a node in W and $(X \perp_e Y|Z)_{\text{del}_W}$. Then G is NON-ALGEBRAIC if the d-separation relations of G *do not* include any relations of the form $(X \perp_d Y|S)$, where S may be any subset of the observed nodes of G other than $X \cup Y$.

However, a graph can only exclude relations of the form $X \perp_d Y|S$ for any S if for every pair of singleton nodes $\{X, Y\}$ such that $X \in X$ and $Y \in Y$ it holds that $(X \perp_d Y|S)$. On the other hand, a pair of nodes $\{X, Y\}$ can only be e separated by Z upon the removal of W if X and Y are not *adjacent*. Consequently, a DAG can *only* be certified as NON-ALGEBRAIC by Theorem 14 if there exists a pair of d *unseparable* but nevertheless non-adjacent nodes. Which is to say, Theorem 14 is ultimately *equivalent* [63] to Theorem 6 of the main text, albeit obfuscated.

APPENDIX B: REVISITING THE SKELETON CONDITION

In the main test we noted (Theorem 6) that every nonmaximal DAG is NON-ALGEBRAIC. Here we revisit that finding, and contrast is with a prior result due to Pienaar [27].

Firstly, we recall that the set of adjacent node pairs in a DAG is conventionally referred to as the DAG's *skeleton*.

Definition 18 (Skeleton of a DAG). The skeleton of a DAG G is the undirected graph wherein A and B are adjacent in G's skeleton—denoted $A \bullet\!\!-\!\!\bullet B$—if and only if A and B are adjacent in G.

Notably, the skeleton of a DAG can be used to witness observational inequivalence. In particular, Evans has shown that if a pair of nodes is adjacent in some mDAG G but not in some other mDAG H, then $\mathcal{C}_G \neq \mathcal{C}_H$ [24,45]. We formally express this idea here as:

Lemma 10 (Skeleton condition for observational equivalence and dominance). For any pair of DAGs G and H, $\mathcal{C}_G = \mathcal{C}_H$ only if the skeletons of G and H are the same, i.e., they agree on all observable node (non)adjacencies. Additionally,

$\mathcal{C}_G \subset \mathcal{C}_H$ only if every pair of adjacent nodes in G is also adjacent in H.

As with any necessary conditional for observational equivalence, Lemma 10 can then be utilized to formulate a sufficient condition for NON-ALGEBRAICNESS. Pienaar did exactly that when he formulated the following Theorem in Ref. [[27], Theorem 1]:

Theorem 15 (Pienaar's skeleton condition for. NON-ALGEBRAICNESS) Let G be the DAG we need to check for NON-ALGEBRAICNESS. Suppose that one can find some other DAG H with the same observed conditional independencies as G and which is certainly known to be ALGEBRAIC, i.e., $\mathcal{I}_G = \mathcal{I}_H$ and $\mathcal{I}_H = \mathcal{C}_H$. Then, if the skeletons of G and H are different, evidently $\mathcal{C}_G \neq \mathcal{C}_H$ per Lemma 10, and hence $\mathcal{C}_G \neq \mathcal{I}_G$, i.e., DAG G is NON-ALGEBRAIC.

Theorem 15 is both correct and useful. Note, however, that it is *superseded* by the Theorem 6 presented in the main text. Suppose that there *does* exists some latent-free DAG H that agrees with the observed d-separation relations of G. Then, among other things, G and H evidently agree on d unseparable observed node pairs. Since the skeleton of H is defined by H's d-unseparable node pairs [[43], Prop. 3.19], the only way for the skeletons of G and H to differ is if G contains some pair of *nonadjacent d* unseparable observed nodes. Consequently, Theorem 15 can only certify a DAG as NON-ALGEBRAIC when the DAG's NON-ALGEBRAICNESS is also certifiable by Theorem 6.

Thus, *both* of Pienaar's conditions for NON-ALGEBRAICNESS are subsumed by Theorem 6 here, at least after adjusting Pienaar's condition based on e separation as per Appendix A.

We know that there are DAGs for which one cannot find any DAG H with the required properties to make use of Theorem 15; see Appendix C for examples. In light of Theorem 5 we would want to reformulate Theorem 15 in a manner that is clearly (strictly) more powerful, which removes the caveat about finding such an H.

Theorem 16 (Improved skeleton condition for. NON-ALGEBRAICNESS) Let G be the DAG we need to check for NON-ALGEBRAICNESS. Consider all latent-free graphs, which share the same skeleton as that of G. If no latent-free graph within that set furthermore matches the observed d separations of G, then G is NON-ALGEBRAIC.

Proof. Theorem 16 follows directly from combining Lemma 10 and Eq. (8) with Theorem 5. ∎

As it turns out, however, Theorem 16 is *equivalent* to the conjunction of Theorems 6 and 8. First, we can see that Theorem 6 is a special case of Theorem 16. If G is nonmaximal, then it has at least one non-adjacent pair of observed nodes, which is d unseparable. This implies that all the latent-free DAGs, which have the same skeleton (same adjacency structure) as G have a different set of d-separable pairs of nodes than that of G, because all latent-free DAGs are maximal [[43], Prop. 3.19]. This then implies that they have a different set of d-separation relations than those of G, so Theorem 16 witnesses all nonmaximal DAGs as NON-ALGEBRAIC. Secondly, it is easy to see that Theorem 8 is also a special case of Theorem 16: if there are no latent-free DAGs that match the set of d-separation relations of G, then in particular there

are no latent-free DAGs that match the set of d-separation relations *and* the skeleton of G.

Next, we prove the inverse, that Theorems 6 and 8 *together* subsume Theorem 16. Let G be a DAG, which is shown NON-ALGEBRAIC by Theorem 16. If G is nonmaximal, then it is also shown NON-ALGEBRAIC by Theorem 6. What remains, then, is to show that there are no *maximal* DAGs, which *cannot* be proven NON-ALGEBRAIC via Theorem 8 but *can* be seen as NON-ALGEBRAIC via Theorem 16. But if the DAG G is both maximal *and* shares the same d-separation relations as some latent-free DAG H, it *automatically follows* that G and H agree on their skeletons as well. After all, G and G are *both* maximal, which means that their skeletons are dictated by their respective d-unseparable node pairs, which are identical.

Note that Theorem 16 is also subsumed by Theorem 9, since agreeing on e-separation relations implies agreeing on adjacencies (i.e., skeleton) as well as on d-separation relations.

APPENDIX C: RAPID d-SEPARATION TEST

In Sec. IV C, we described a method to show the NON-ALGEBRAICNESS of a DAG when its set of observed d-separation relations is NALF, i.e., it does not match those of any latent-free DAG; this is encoded in Theorem 8. To apply this method in practice, we indicated that one can construct all of the latent-free DAGs that have the same number of observed nodes as the DAG in question, and then compare their sets of d-separation relations. However, when the DAG is *maximal* (as per Definition 9) there is a faster way to apply this method, which will be described now. When the DAG is *not* maximal, its NON-ALGEBRAICNESS is automatically attested by Theorem 6.

When the DAG is maximal, to see whether its set of d-separation relations is NALF one just needs to check whether it has one of the DAGs of Fig. 12 as a subgraph. This was recognized in Evans' own proof of Theorem 5 [25], as we argue in the proof of the following theorem:

Theorem 17 (rapid d-separation condition for. NON-ALGEBRAICNESS) Let G be an mDAG. If G is not maximal, then it is NON-ALGEBRAIC by Theorem 6. If G is maximal, then it has a set of observed d-separation relations unmatched by any latent-free DAG (NALF)—and is therefore NON-ALGEBRAIC pursuant to Theorem 8—if and only if it contains one of the eighteen graph patterns (up to relabelling the nodes) presented in Fig. 12 as a subgraph.

Proof. Evans [25] has shown that if a latent-permitting DAG G has a NALF set of observed d-separation relations, then either its associated PAG (partial ancestral graph) contains a so-called "locally unshielded collider path" of length 3, and/or it contains a so-called "discriminating path" of length 3 [64]. This implies that the PAG associated with G will contain a sub-PAG matching one of the four-node PAGs depicted within Figs. 4 and 5(i) of [25] whenever the observed d-separation relations of G are NALF.

A PAG is an abstract graphical representation of a set of d-separation relations. In particular, two nodes are *adjacent* in a PAG if and only the two nodes are d unseparable in the original DAG. A "path of length 3" in a PAG refers to a set
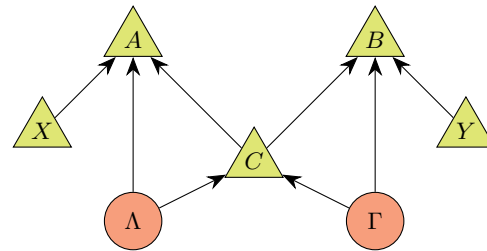

FIG. 13. Nonmaximal DAG. It does not have any of the DAGs of Fig. 12 as a subgraph, but it nevertheless has a d-separation pattern that does not correspond to any latent-free DAG.

of four nodes $\{X, A, B, Y\}$ such that $\{X, A\}$, $\{A, B\}$, $\{B, Y\}$ all constitute d-unseparable pairs.

The adjacencies of a DAG G and the adjacencies of the PAG associated with G can, in general, differ. For *maximal* DAGs, however, they must coincide: for those, adjacency is equivalent to d unseparability. The PAG associated with a *maximal* DAG G will exhibit an unshielded collider path or a discriminating path if and only if in the original G we can find four nodes $\{X, A, B, Y\}$ such that $\{X, A\}$, $\{A, B\}$, $\{B, Y\}$ represent adjacent pairs and such that the d-separation relations pertaining exclusively to $\{X, A, B, Y\}$ are of one of "unshielded collider path" type or "discriminating path" type. We do not define these two types for brevity, but we exhibit all such four-node mDAGs in Fig. 12. ∎

The set of d-separation relations of each one of the DAGs in Fig. 12 is

(1) $(A \perp_d Y | \emptyset)$ and $(A \perp_d Y | X)$ and $(B \perp_d X | \emptyset)$ and $(B \perp_d X | Y)$ and $(X \perp_d Y)$ and $(X \perp_d Y | A)$ and $(X \perp_d Y | B)$ [Figs. 12(a)–12(f)]

(2) $(A \perp_d Y | X)$ and $(B \perp_d X | Y)$ [Figs. 12(g)–12(k)]

(3) $(A \perp_d Y | \emptyset)$ and $(B \perp_d X | \emptyset)$ [Fig. 12(l)]

(4) $(B \perp_d X | \emptyset)$ and $(A \perp_d Y | X)$[Figs. 12(m)–12(o)]

(5) $(B \perp_d X | \emptyset)$ and $(X \perp_d Y | A)$ [Figs. 12(p)–12(r)]

By checking the d separation of all the four-observed-node mDAGs, we know that these are the only mDAGs that present these sets of d-separation relations. While searching for subgraphs is computationally easy, an alternative is to consider all four-node subsets of a given large mDAG and ask if the d-separation relations *pertaining exclusively to some four nodes* contains all and only one of the patterns listed above, up to relabelling. If yes, and if the large mDAG G is maximal, then G certainly contains one of the patterns of Fig. 12 as a subgraph.

Figure 13 shows an example of a DAG that does not have any of the DAGs of Fig. 12 as a subgraph, but nevertheless has a NALF set of d-separation relations. However, this is not a problem: this DAG is *not* maximal (X and Y are not d separable but are e separable by the empty set), so its NON-ALGEBRAICNESS follows from Theorem 6.

Finally, we will show an explicit construction of a distribution, which is in $\mathcal{I}_G$ but not in $\mathcal{C}_G$ for the maximal mDAGs, which are shown NON-ALGEBRAIC by Theorem 6. First, we note that the Popescu-Rohrlich box support shown in Eq. (14) is not compatible with any of the mDAGs of Fig. 12, as can be shown using Fraser's algorithm. This implies that the *uniform* distribution over the events of the Popescu-Rohrlich box

support is not classically compatible with any of the mDAGs of Fig. 12, and is thus not an element of $\mathcal{C}_G$ for these mDAGs. On the other hand, it is well known that said distribution obeys all of the conditional independence relations that come from the d-separation relations of the Bell DAG. As seen above, all the d-separation relations of the mDAGs of Fig. 12 are included in the d-separation relations of the Bell DAG, which implies that such distribution is an element of $\mathcal{I}_G$ for all the mDAGs of Fig. 12.

APPENDIX D: SUPPORTS SUBSUMES e-SEPARATION AS A TEST OF OBSERVATIONAL INEQUIVALENCE

Here we provide a full proof of Theorem 12, that says that if DAGs G and H can be shown inequivalent by virtue of having different sets of e-separation relations, then they can be shown inequivalent by virtue of having different sets of compatible supports at *binary* variables. Before proving this more general result, we will prove the case for d-separation relations as an auxiliary lemma:

Lemma 11 (d separation and compatible supports). Let G be an DAG with observed nodes V, with $Z \subseteq V$ being a subset of V and $X \in V$ and $Y \in V$ being two observed nodes. If the DAG G *does not* exhibit the d-separation relation $(X \perp_d Y | Z)$, then there is at least one support over binary variables, which is compatible with G but at the same time is in conflict with $(X \perp\!\!\!\perp_{CI} Y | Z)$, i.e., is trivially incompatible with every other DAG for which $(X \perp_d Y | Z)$.

Proof. We explicitly construct a support such that the marginal events $\{X=0, Z=1\}$ and $\{Y=1, Z=1\}$ both occur, but such that the event $\{X=0, Y=1, Z=1\}$ does not occur. Plainly such a support is *trivially incompatible* with any DAG wherein $(X \perp_d Y | Z)$ per Lemma 8. Here we show that a distribution with this support *can* arise in every DAG wherein $(X \not\perp_d Y | Z)$. We note that this construction is identical to that which appears in Appendix E of [47].

If $(X \not\perp_d Y | Z)$ in G, then there exists a path from X to Y in G, which is unblocked by Z. Every node of the path has either no parents in the path (in which case it is the base of a fork), one parent in the path (if it is the middle node of a chain or an end node of the path) or two parents in the path (if it is a collider). The support is constructed by assigning the following functional dependencies to the nodes of the path:

(i) A node $F \in \boldsymbol{F}$ ("F" for "Fork") has 0 parents in the path:

$$F = \begin{cases} 0 \text{ with probability } 1/2 \\ 1 \text{ with probability } 1/2 \end{cases}$$

(ii) A node $M \in \boldsymbol{M}$ ("M" for "Mediary") has 1 parent P_M in the path:

$$M = P_M \text{ with unity probability}$$

(iii) Node $C \in \boldsymbol{C}$ ("C" for "Collider") has 2 parents $P_{C,1}$ and $P_{C,2}$ in the path:

$$C = \begin{cases} 0 \text{ if } P_{C,1} \neq P_{C,2} \\ 1 \text{ if } P_{C,1} = P_{C,2} \end{cases}$$

Since the path is unblocked $C \in \boldsymbol{Z}$ while $\boldsymbol{MF} \notin \boldsymbol{Z}$.

This construction leads to a probability distribution compatible with G wherein $P(\boldsymbol{M} = \boldsymbol{F} = 0 | Z=1) = 1/2$ and

$P(\boldsymbol{M} = \boldsymbol{F} = 1 | Z=1) = 1/2$. Since $\{X, Y\} \subset \boldsymbol{MF}$, we confirm that the marginal event $\{X=0, \boldsymbol{Z} = 1\}$ occurs, and such that the marginal event $\{Y=2, \boldsymbol{Z} = 1\}$ occurs, yet that the event $\{X=0, Y=1, \boldsymbol{Z} = 1\}$ does not occur.

Using Lemma 11, we can now proceed to proof Theorem 12.

Theorem 12. (Supports subsumes e separation for observational inequivalence). Let G and H be two DAGs. If their sets of observed e-separation relations are different, then it is possible to find a support over binary variables, which is compatible with one of them but incompatible with the other.

Proof. Suppose that G has an e-separation relation $(A \perp_e B | \boldsymbol{C})_{\text{del}_{\boldsymbol{D}}}$ that H does not have, where $A, B \in V$ are observed nodes and $\boldsymbol{C}, \boldsymbol{D} \subseteq V$ are sets of observed nodes.

Proof. By the definition of e separation, we know that $G_{V \setminus \boldsymbol{D}}$, the DAG obtained by deleting the nodes $\boldsymbol{D}$ from G, exhibits the d-separation relation $A \perp_d B | \boldsymbol{C}$ while $H_{V \setminus \boldsymbol{D}}$ does not. This implies that *none* of the supports compatible with $G_{V \setminus \boldsymbol{D}}$ are in conflict with $A \perp\!\!\!\perp_{CI} B | \boldsymbol{C}$, while from Lemma 11 we know that there is at least one support over binary variables, which is compatible with $H_{V \setminus \boldsymbol{D}}$ that is in conflict with $A \perp\!\!\!\perp_{CI} B | \boldsymbol{C}$. This means that it is possible to find a support $\mathcal{S}_{V \setminus \boldsymbol{D}}$ over binary variables, which is compatible with $H_{V \setminus \boldsymbol{D}}$ but not compatible with $G_{V \setminus \boldsymbol{D}}$.

Let $\mathcal{S}_V$ be the support over V, which coincides with $\mathcal{S}_{V \setminus \boldsymbol{D}}$ for the variables in $V \setminus \boldsymbol{D}$ and where the variables in $\boldsymbol{D}$ are set to a point value. This support is incompatible with G, because setting the variables in $\boldsymbol{D}$ to a point value has the same effect on its children as deleting the respective nodes (and thus reaching $G_{V \setminus \boldsymbol{D}}$). Furthermore, $\mathcal{S}_V$ is *compatible* with H: since $\mathcal{S}_{V \setminus \boldsymbol{D}}$ is compatible with $H_{V \setminus \boldsymbol{D}}$, we can obtain $\mathcal{S}_V$ by functional models where the children of $\boldsymbol{D}$ in G ignore the parents in $\boldsymbol{D}$, as indicated in Lemma 13 of [47].

Therefore, $\mathcal{S}_V$ is a support over binary variables that is compatible with H but incompatible with G. ∎

APPENDIX E: RESULTS FOR MDAGS OF Five OBSERVED NODES

When trying to certify the NON-ALGEBRAICNESS of mDAGs with five or more observed nodes, a simple but important consideration should be taken into account. Namely,

Proposition 7. Consider a DAG G, as well as the subgraph $G_{\text{del} S}$ of G, where $G_{\text{del} S}$ is formed by deleting a strict subset of G's observed nodes from G. If $G_{\text{del} S}$ is NON-ALGEBRAIC, then G is NON-ALGEBRAIC as well.

Proof. Proposition 7 is an immediate consequence of Lemma 4 from the main text. That is, Lemma 4 ensure that $P(V \setminus S) \notin \mathcal{C}_{G_{\text{del} S}}$ then $P(V) \notin \mathcal{C}_G$ where $P(V) := P(V \setminus S) \delta_{S, 0^{|S|}}$. At the same time, if $P(V \setminus S) \in \mathcal{I}_{G_{\text{del} S}}$ then $P(V) \in \mathcal{I}_G$ where again $P(V) := P(V \setminus S) \delta_{S, 0^{|S|}}$. That $P(V) \in \mathcal{I}_G$ follows from the fact that no *new* d-separation relations are induced on $V \setminus S$ by embedding the DAG $G_{\text{del} S}$ as a subgraph of a larger DAG, namely G. ∎

The reader might ask why we did not employ Proposition 7 in assessing the NON-ALGEBRAICNESS of mDAGs with *4* observed nodes. After all, there are five mDAGs with *three* observed nodes. However, any four-observed-nodes mDAGs, which would be certifiable as NON-ALGEBRAIC via Proposition 7 would already be picked up by Theorem 7, since

TABLE IV. Results for mDAGs of five observed nodes.

Category	mDAGs with five observed nodes
Total Count	1 718 596
remaining # for which the HLP criterion does not apply	1 009 961
remaining # for which our nonmaximality condition does not apply	278 964
remaining # for which our setwise nonmaximality condition does not apply	118 278
remaining # which do not contain an NON-ALGEBRAIC four-observed-nodes subgraph	12 834
remaining # for which our d-separation condition does not apply	12 834
remaining # for which our e-separation condition does not apply	12 834

that theorem detects 100% of the NON-ALGEBRAIC three-observed-nodes mDAGs by itself.

The application of the conditions for NON-ALGEBRAICNESS presented here to mDAGs of five observed nodes gives the results shown in Table IV. Although the application of Fraser's algorithm on the remaining 12 834 mDAGs was computationally infeasible, nevertheless, the application of all the other techniques provides a similar success percentage as compared to mDAGs of four and three observed nodes. Precisely, for mDAGs of five observed nodes all the other techniques (apart from supports) reduce the number of unresolved mDAGs of five observed nodes by 99.15%, while this percentage is 99.89% and 97.8% for mDAGs of four and three observed nodes respectively. This result is consistent with the HLP conjecture, though whether or not it can be considered *evidence* in favor of the conjecture is debatable.

Unsurprisingly, our d-separation condition as articulated in Theorem 8 is now effectively redundant to the conjuction of Proposition 7 and Theorem 6, since a *maximal* mDAG with five+ observed nodes will have a set of observed d-separation relations inequivalent to any latent-free DAG only if the large mDAG contains one of 18 particular four-observed-nodes mDAGs, as discussed extensively in Appendix C.

The fact that the Theorem 9 does not resolve any further mDAGs as NON-ALGEBRAIC is evidence in favor of a conjecture that Theorem 9 is perhaps subsumed by the conjunction of Theorems 6 and 8.

[1] J. Pearl, *Causality*, 2nd ed. (Cambridge University Press, Cambridge, 2009).

[2] T. Verma and J. Pearl, Causal networks: Semantics and expressiveness, in *Machine Intelligence and Pattern Recognition* (Elsevier, 1988), Vol. 9, pp. 69–76.

[3] J. Pearl, On the testability of causal models with latent and instrumental variables, in *Proceedings of the Eleventh Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-95)* (Morgan Kaufmann Publishers Inc., 1995), pp. 435–443.

[4] P. Spirtes, C. Glymour, and R. Scheines, Causation, Prediction, and Search (MIT Press Books, Boston, 2001).

[5] S. L. Lauritzen, *Graphical Models* (Oxford University Press, Oxford, 1996).

[6] We will denote probabilities by the concise notation $P(XY) \equiv P_{XY}(X = x, Y = y)$.

[7] J. S. Bell, On the Einstein-Podolsky-Rosen paradox, Phys. Phys. Fiz. **1**, 195 (1964).

[8] J. S. Bell, M. A. Horne, and A. Zeilinger, Speakable and unspeakable in quantum mechanics, Am. J. Phys. **57**, 567 (1989).

[9] T. Norsen, *Foundations of Quantum Mechanics* (Springer International Publishing, New York, 2017).

[10] J. Bricmont, *Making Sense of Quantum Mechanics* (Springer International Publishing, New York, 2016).

[11] D. Dürr and D. Lazarovici, *Understanding Quantum Mechanics* (Springer International Publishing, New York, 2020).

[12] A. Einstein, B. Podolsky, and N. Rosen, Can quantum-mechanical description of physical reality be considered complete? Phys. Rev. **47**, 777 (1935).

[13] T. Norsen, John S. Bell's concept of local causality, Am. J. Phys. **79**, 1261 (2011).

[14] T. Norsen, Local causality and completeness: Bell vs. Jarrett, Found. Phys. **39**, 273 (2009).

[15] S. Goldstein, T. Norsen, D. V. Tausk, and N. Zanghi, Bell's theorem, Scholarpedia **6**, 8378 (2011).

[16] T. Norsen, Against 'Realism', Found. Phys. **37**, 311 (2007).

[17] T. Norsen, Einstein's boxes, Am. J. Phys. **73**, 164 (2005).

[18] J. Henson, R. Lal, and M. F. Pusey, Theory-independent limits on correlations from generalized Bayesian networks, New J. Phys. **16**, 113043 (2014).

[19] T. Fritz, Beyond Bell's theorem: correlation scenarios, New J. Phys. **14**, 103001 (2012).

[20] T. Fritz, Beyond Bell's Theorem II: Scenarios with Arbitrary Causal Structure, Commun. Math. Phys. **341**, 391 (2016).

[21] R. Chaves, L. Luft, and D. Gross, Causal structures from entropic information: Geometry and novel scenarios, New J. Phys. **16**, 043001 (2014).

[22] M.-O. Renou, E. Bäumer, S. Boreiri, N. Brunner, N. Gisin, and S. Beigi, Genuine quantum nonlocality in the triangle network, Phys. Rev. Lett. **123**, 140401 (2019).

[23] Our use of ALGEBRAIC and NON-ALGEBRAIC is equivalent to the distinction between BORING and INTERESTING causal structures in the terminology of Ref. [18]. We recognize that, in general, renaming existing technical jargon is counterproductive. However, we feel that our terminology of ALGEBRAIC and NON-ALGEBRAIC has semantic connotations so much closer to their underlying meanings as to warrant the renaming. The authors acknowledge Robert W. Spekkens for suggesting these terms.

[24] R. J. Evans, Graphs for margins of Bayesian networks, Scandinavian J. Stat. **43**, 625 (2016).

[25] R. J. Evans, Latent-free equivalent mDAGs, Alge. Stat. 14, 3 (2023).

[26] T. C. Fraser, A combinatorial solution to causal compatibility, J. Causal Inference **8**, 22 (2020).

[27] J. Pienaar, Which causal structures might support a quantum–classical gap? New J. Phys. **19**, 043021 (2017).

[28] In doing so, we expose a critical error in one of the theorem's in Ref. [27], which we then rectify here (Happily, the handful of explicit DAGs, which were declared as NON-ALGEBRAIC pursuant to the fallacious theorem in Ref. [27] are ultimately indeed NON-ALGEBRAIC nevertheless, and moreover their NON-ALGEBRAICNESS follows from our replacement nonmaximality theorem here). This discussion is made in Appendix A.

[29] Note that this repository has not been optimized for public use. We nevertheless include the link in the spirit of transparency, https://github.com/eliewolfe/mDAG-analysis.

[30] A. Acín, N. Brunner, N. Gisin, S. Massar, S. Pironio, and V. Scarani, Device-independent security of quantum cryptography against collective attacks, Phys. Rev. Lett. **98**, 230501 (2007).

[31] U. Vazirani and T. Vidick, Fully device independent quantum key distribution, Commun. ACM **62**, 133 (2019).

[32] J. Barrett, Information processing in generalized probabilistic theories, Phys. Rev. A **75**, 032304 (2007).

[33] M. Plávala, General probabilistic theories: An introduction, Phys. Rep. **1033**, 1 (2023).

[34] A set of d-separation relations is said to be sufficient for a DAG if all other d-separation relation in the DAG can be inferred from the sufficient subset by application of semigraphoid axioms.

[35] In this article we follow the convention of Refs. [36–38] and others, namely, the convention in which $X \in \text{ANC}_G(X)$ and $X \in \text{DEC}_G(X)$ but $X \notin \text{PA}_G(X)$ and $X \notin \text{CH}_G(X)$. That is, a node is considered it own ancestor and its own descendant, but not its own parent or its own child.

[36] T. Richardson, A characterization of Markov equivalence for directed cyclic graphs, Int. J. Approximate Reasoning **17**, 107 (1997).

[37] B. Steudel and N. Ay, Information-theoretic inference of common ancestors, Entropy **17**, 2304 (2015), note that the journal version and the arXiv version of this citation differ significantly in terms of the enumeration of theorems and examples. See Theorem 2 and Example 2 of the *journal* version.

[38] B. van der Zander, M. Liśkiewicz, and J. Textor, Separators and adjustment sets in causal graphs: Complete criteria and an algorithmic framework, Artif. Intell. **270**, 1 (2019).

[39] D. Geiger and J. Pearl, On the logic of causal models, *Machine Intelligence and Pattern Recognition* (Elsevier, 1990), Vol. 9, pp. 3–14.

[40] There is another type of equality constraint that a DAG can impose on the compatible distributions, apart from conditional independence relations: the "nested Markov constraints". In Ref. [25] it was proven that every DAG that presents nested Markov equality constraints also present nontrivial inequalities. Consequently, NON-ALGEBRAICNESS may be equivalently defined relative to *all implied equality constraints* or relative to *all implied conditional independence relations*.

[41] Note that the set of all DAGs with four observed nodes *includes* the set of all DAGs with seven total nodes, which persist under HLP 's reduction techniques. Thus, this paper can *also* be considered an extension of HLP 's classification from up-to-six to up-to-seven total nodes. As noted in Sec. V, we ultimately

resolve the ALGEBRAICNESS or NON-ALGEBRAICNESS of *all but three* DAGs of four observed nodes. Only *one* of those three has seven total nodes. Thus, we ultimately resolve *all but one* DAG with seven total nodes.

[42] On the other hand, if DAGs G and H are such that $\mathcal{I}_G = \mathcal{I}_H$, this does *not* imply that $\mathcal{C}_G = \mathcal{C}_H$.

[43] T. Richardson and P. Spirtes, Ancestral graph Markov models, Ann. Stat. **30**, 962 (2002).

[44] For DAGs, which are not exogenized (see Definition 5) we additionally say that A and B are adjacent if there is some *latent treck* from on to the other. Which is to say, A and B are said to be *adjacent* in a nonexogenized DAG G iff they are adjacent in the mDAG resulting from exogenizing G per see Definition 5.

[45] R. J. Evans, Graphical methods for inequality constraints in marginalized DAGs, in *2012 IEEE International Workshop on Machine Learning for Signal Processing* (IEEE, 2012), pp. 1–6.

[46] The concept of e separation will be elaborated upon in Sec. IV D here.

[47] N. Finkelstein, B. Zjawin, E. Wolfe, I. Shpitser, and R. W. Spekkens, Entropic inequality constraints from e-separation relations in directed acyclic graphs with hidden variables, in *Uncertainty in Artificial Intelligence* (PMLR, 2021), pp. 1045–1055.

[48] J. Tian, A. Paz, and J. Pearl, *Finding Minimal D-Separators* (Computer Science Department, University of California, 1998).

[49] Note that we are using the convention that a node counts as its own ancestor, a la Refs. [36–38].

[50] Here $|\mathbf{X}|$ denotes the number of elements of the set $\mathbf{X}$, and $\vec{0}^n$ represents the vector with n entries equaling 0.

[51] C. Branciard, D. Rosset, N. Gisin, and S. Pironio, Bilocal versus nonbilocal correlations in entanglement-swapping experiments, Phys. Rev. A **85**, 032119 (2012).

[52] We do not discuss *how* to find a latent-free DAG with the same d-separation relations as G, as ultimately we conclude that there is apparently no advantage in doing so, as discussed in Appendix C.

[53] L. Khalfin and B. Tsirelson, Steklov Mathematical Institute, Leningrad D-11, USSR, in *Symposium on the Foundations of Modern Physics: 50 Years of the Einstein-Podolsky-Rosen Gedankenexperiment, Joensuu, Finland, 16-20 June 1985* (World Scientific, Singapore, 1985), p. 441.

[54] S. Popescu and D. Rohrlich, Quantum nonlocality as an axiom, Found. Phys. **24**, 379 (1994).

[55] More generally, we conjecture that for *any* DAG G, if $\mathcal{S}$ is a support, which is not trivially incompatible with G as per Definition 17, then $\mathcal{S}$ is compatible *up-to-CI* with G. We did not prove this, however, except for the special case when G is latent-free.

[56] E. Wolfe, R. W. Spekkens, and T. Fritz, The inflation technique for causal inference with latent variables, J. Caus. Inf. **7**, 20170020 (2019).

[57] A. Restivo, N. Brunner, and D. Rosset, Possibilistic approach to network nonlocality, arXiv:2208.13526.

[58] Note that the support is not *trivially* incompatible, since the DAG in question does not exhibit any d-separation relations involving observed nodes

[59] The entropic technique is predicated on trying to prove that the Shannon-type entropic inequalities constraining $\mathcal{C}_G$ define

a hypercone strictly in the interior of the hypercone given by the Shannon-type entropic inequalities constraining $\mathcal{I}_G$. Even when such a finding can be established, however, it need not mean that $\mathcal{C}_G \subsetneq \mathcal{I}_G$, though that is *almost certainly* the case. The loophole—however implausible—is that perhaps the *true* entropy cone constraining $\mathcal{I}_G$ actually coincides with (or is interior to) the projection of the Shannon cone associated with $\mathcal{C}_G$. See Appendix E of Ref. [18].

[60] M. Weilenmann and R. Colbeck, Analysing causal structures with entropy, Proc. R. Soc. A **473**, 20170483 (2017).

[61] N. Miklin, A. A. Abbott, C. Branciard, R. Chaves, and C. Budroni, The entropic approach to causal correlations, New J. Phys. **19**, 113041 (2017).

[62] T. Gläßle, D. Gross, and R. Chaves, Computational tools for solving a marginal problem with applications in Bell non-locality and causal modeling, J. Phys. A: Math. Theor. **51**, 484002 (2018).

[63] Plainly Theorem 6 can be seen as a special case of Theorem 14, by taking $\boldsymbol{Z}$ to be empty, choosing $\boldsymbol{X}$ and $\boldsymbol{Y}$ to be singleton-sized sets, and letting $\boldsymbol{W}$ be the set of all observed nodes other than $\boldsymbol{X}$ and $\boldsymbol{Y}$.

[64] The phrasing presented in the proof of Proposition 4.2 of Ref. [25] says that either the associated PAG has a locally unshielded collider path of length exactly 3, all options of which are presented in Fig. 4 of Ref. [25], or a discriminating path of length *at least* 3. However, Proposition B.1 of [25] further shows that all discriminating paths of length at least 3 either contain a locally unshielded collider path of length 3 and/or a discriminating path of length exactly 3, which is presented in Fig. 5(i) of [25].