Proceedings of Machine Learning Research 275:1–15, 2025 4th Conference on Causal Learning and Reasoning

Causal Identification in Time Series Models

Erik Jahn EJAHN@CALTECH.EDU
Karthik Karnik SATHKAR2009@GMAIL.COM
Leonard J. Schulman SCHULMAN@CALTECH.EDU
Engineering and Applied Science, California Institute of Technology, Pasadena CA 91125, USA.

Editors: Biwei Huang and Mathias Drton

Abstract

In this paper, we analyze the applicability of the Causal Identification algorithm to causal time series graphs with latent confounders. Since these graphs extend over infinitely many time steps, deciding whether causal effects across arbitrary time intervals are identifiable appears to require computation on graph segments of unbounded size. Even for deciding the identifiability of intervention effects on variables that are close in time, no bound is known on how many time steps in the past need to be considered. We give a first bound of this kind that only depends on the number of variables per time step and the maximum time lag of any direct or latent causal effect. More generally, we show that applying the Causal Identification algorithm to a constant-size segment of the time series graph is sufficient to decide identifiability of causal effects, even across unbounded time intervals.

Keywords: Causal Graphs, Time Series, Causal Identification

1. Introduction

Causal inference on time series data has many important applications across fields such as economics, finance, earth and climate science (Runge et al., 2019a; Moraffah et al., 2021; Zhang et al., 2014). In these domains, conducting large-scale experiments to determine causal effects is often impractical or impossible, but large observational time-series data sets are frequently available. Developing statistical methods that leverage these datasets to infer causal relationships is thus an important problem. Causal time series graphs (also called Dynamic Bayesian Networks) provide a well-studied model for this purpose (Assaad et al., 2022). The vertices of these graphs represent a set of observable random variables recurring at each non-negative integer time, and the edges represent direct causal effects or latent confounding. The key assumption for these graphs is that the causal structure is time-invariant, that is, edges are invariant under time shifts. However, the joint distributions of sets of variables need not be time-invariant. Numerous methods have been explored for causal discovery and causal feature selection in these graphs, both in the setting without latent variables (Runge et al., 2019b; Pfister et al., 2019) and with latent variables (Malinsky and Spirtes, 2018; Mastakouri et al., 2021; Gerhardus and Runge, 2020).

In many cases, models without latent variables can be overly simplistic. Particularly for time series data, latent confounding may not only result from unobserved covariates, but can also arise when the time resolution of the observations is too coarse, even when all relevant variables are observed (Peters et al., 2017; Runge, 2018). In this paper we focus on causal time series graphs with latent variables. In this setting, even when a causal graph is discovered, the problem of causal identification remains. It is well-known that some causal effects in certain graphs may be unidentifiable (Pearl, 2009). For finite graphs the question of causal identifiability is answered precisely by the Causal ID algorithm (Tian and Pearl, 2002; Shpitser and Pearl, 2006; Huang and Valtorta, 2008;

© 2025 E. Jahn, K. Karnik & L.J. Schulman.

Shpitser and Pearl, 2008). However, in infinite periodic graphs representing a time series model, causal identification is less well-studied. The only existing work on causal identifiability in these graphs that we are aware of does not address the challenges described below (Blondel et al., 2016). Other work on causal identifiability in time series settings is based on Granger causality (Eichler and Didelez, 2007, 2009), which is known to have certain limitations due to violations of its assumptions (Peters et al., 2017). One might argue that the direction of time allows the standard ID algorithm to be applied even to infinite time series graphs. Indeed, one can show that identifying the causal effect of a variable X on Y requires considering only the section of the graph up to the time at which Y occurs, see Section 3. However, the following problems remain:

1. While the causal effect of X on Y is not affected by variables that lie in the future of Y, there may be chains of confounding variables in the past that obstruct identifiability (see Section 3 for an example). Hence, computing the identifiability of $P(Y|\operatorname{do} X)$ would naively require running the Causal ID algorithm on the entire past of the time series, from its initialization up to the occurrence of Y. Consider, for instance, a causal time series graph modeling Earth's climate system. Using the ID algorithm to check the identifiability of the causal effect between two variables in the near-present, would require computation scaling with the length of Earth's history, which is clearly infeasible. For causal discovery in such settings, one often just considers the past up to a maximum time interval $\tau_{\max}$. However, $\tau_{\max}$ is typically chosen empirically and this may distort causal effects (see Gerhardus and Runge (2020), Section 2.3). A priori, the effect of such a cut-off on causal identification is unclear.

2. Suppose we are interested in the causal effect between recurring variables X and Y but do not know the time span over which this effect might occur. A causal time series graph employed to investigate this effect should have the property that the causal effect of X_t on $Y_{t+\Delta}$ is identifiable for all time steps $\Delta \geq 0$ in the future. However, the Causal ID algorithm cannot decide in finite computing time whether a graph has this property. The only option seems to be to restrict the question of identifiability up to a maximum time step $\Delta_{\max}$, which would still require increasing computing time in $\Delta_{\max}$.

In this paper, we provide solutions to both problems through the following result.

Theorem 1 *Let $\mathcal{G}$ be a periodic causal graph and let $\mathbf{X}, \mathbf{Y} \subseteq \mathbf{V}(\mathcal{G})$ be disjoint subsets of variables. Let $t_{\min}$ be the smallest time index of a variable in $\mathbf{X}$ and $t_{\max}$ the largest time index of a variable in $\mathbf{Y}$. Then, there exists a constant C such that the following holds:*

1. *Running the Causal ID algorithm on the segment of $\mathcal{G}$ with time indices between $t_{\min} - C$ and $t_{\max}$ suffices to determine identifiability of the causal effect of $\mathbf{X}$ on $\mathbf{Y}$.*

2. *If the causal effect of $\mathbf{X}$ on $\mathbf{Y}$ is unidentifiable, then there is a time shift Δ such that the time difference between $\mathbf{X}$ and $\mathbf{Y}_{-\Delta}$ is at most C and the causal effect of $\mathbf{X}$ on $\mathbf{Y}_{-\Delta}$ is also unidentifiable.*

See Section 2 for formal definitions of the notation used above. Our result implies that identifiability of the causal effect between any two sets of variables can be decided in computing time that scales with C but not with the size of the time interval between these variables or the distance to the initialization time of the causal model, see Figure 1. Here, C depends only on the number of

variables per time step and the maximum time lag of any direct or latent causal effect. In Section 3 we show specific upper and lower bounds on the scaling of C. While our upper bound exhibits exponential growth, it is only a first generic bound, applicable to any periodic graph structure. For specific graphs, C could potentially be much smaller. In fact, we are only able to construct graphs that require a linear scaling of C in the number of variables per time step, and we present this construction together with the proofs of our results in Section 3.

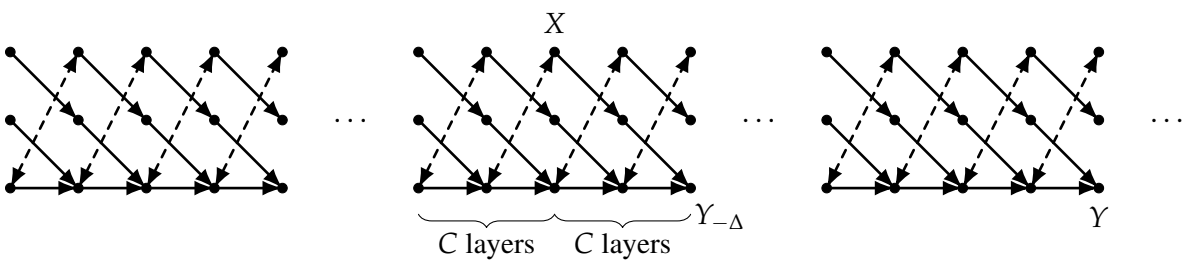

Figure 1: A priori, deciding if the causal effect of X on Y is identifiable requires running the causal ID algorithm on the entire part of the time series graph shown above - from its initialization up to the layer containing Y. Our results show that instead computation on a constant-size section around X (i.e. just the middle part) suffices, using a shifted version $Y_{-\Delta}$ of Y.

2. Preliminaries

First, we define notions related to the causal time series graphs that we consider and fix notation that we use throughout this paper. To clearly distinguish sets of variables from singletons, we use bold letters to denote sets. Graphs are denoted by calligraphic letters. All graphs in this paper are acyclic directed mixed graphs (ADMGs), i.e. graphs with both directed edges representing direct causal effects and bidirected edges representing confounding variables.

Definition 1 (periodic causal graph) *A periodic causal graph of width w is an acyclic directed mixed graph with vertices labeled by $X_{i,t}$ for $0 \leq i \leq w-1$ and $t \in \mathbb{N}$. We say, the vertex $X_{i,t}$ is in row i and column, or time, t and denote the set of vertices at time t by $\mathbf{X}_{*,t}$.*

The graph is periodic in that $(X_{i,t}, X_{i',t'})$ is a directed (resp. bidirected) edge if and only if the same is true of $(X_{i,t+1}, X_{i',t'+1})$ and directed edges do not go backwards in time, i.e. each directed edge $(X_{i,t}, X_{i',t'})$ must have $t \leq t'$ (allowing for contemporaneous edges).

Definition 2 (segments) *We denote the segment of a periodic causal graph $\mathcal{G}$ that is induced by the vertices $X_{i,t}$ with $0 \leq i \leq w-1$ and $T \leq t \leq T'$, by $\mathcal{G}[T, T']$.*

Definition 3 (latency) *A periodic causal graph is of latency L if for any directed or bidirected edge $(X_{i,t}, X_{i',t'})$, we have $|t - t'| \leq L$.*

Definition 4 (distance between sets) *Let $\mathcal{G}$ be a periodic causal graph with vertex set $\mathbf{V}(\mathcal{G})$. For a vertex $V \in \mathbf{V}(\mathcal{G})$, define $\mathrm{layer}(V) = t$, if $V = X_{i,t}$. For a set $\mathbf{X} \subseteq \mathbf{V}(\mathcal{G})$, we define its minimum*

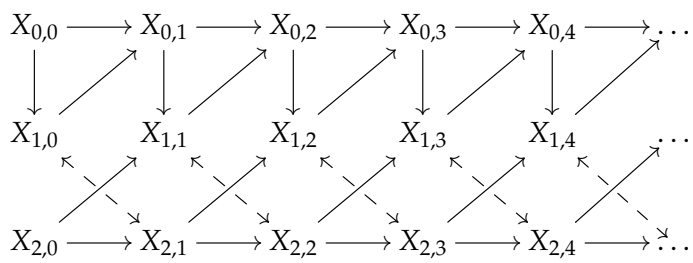

Figure 2: Periodic graph of width 3 and latency 1. Directed edges are solid, bidirected dashed.

layer and maximum layer by

$$\mathrm{t}_{\min}(\mathbf{X}) = \min\{\mathrm{layer}(V) | V \in \mathbf{X}\}$$
$$\mathrm{t}_{\max}(\mathbf{X}) = \max\{\mathrm{layer}(V) | V \in \mathbf{X}\}.$$

Finally, the distance between two sets $\mathbf{X}, \mathbf{Y} \subseteq \mathbf{V}(\mathcal{G})$ *is given by*

$$\mathrm{dist}(\mathbf{X}, \mathbf{Y}) = \min\{|\,\mathrm{layer}(V_1) - \mathrm{layer}(V_2)| \mid V_1 \in \mathbf{X}, V_2 \in \mathbf{Y}\}.$$

Definition 5 (time shifts) *Let* $\mathcal{G}$ *be a periodic causal graph and* $\mathbf{X} \subseteq \mathbf{V}(\mathcal{G})$ *with* $\mathrm{t}_{\min}(\mathbf{X}) \geq \Delta$ *for some* $\Delta \in \mathbb{N}$. *Then, we define the shifted sets*

$$\mathbf{X}_{+\Delta} = \{X_{i,t+\Delta} | X_{i,t} \in \mathbf{X}\},$$
$$\mathbf{X}_{-\Delta} = \{X_{i,t-\Delta} | X_{i,t} \in \mathbf{X}\}.$$

See Figure 2 for an example of a periodic causal graph. A periodic causal graph $\mathcal{G}$ represents a causal model, which is given by a set of equations

$$X_{i,t} = f_{i,t}(\mathrm{Pa}(X_{i,t}, \mathcal{G}), \varepsilon_{i,t}).$$

Here, $\mathrm{Pa}(X_{i,t}, \mathcal{G})$ denotes the parents of $X_{i,t}$ in $\mathcal{G}$ and the $\varepsilon_{i,t}$ are random noise variables that may be correlated for variables $X_{i,t}$ that are connected by a bi-directed edge, and are independent otherwise. Note that we do not require the functions $f_{i,t}$ to be the same for different time steps. For two sets of variables $\mathbf{X}, \mathbf{Y} \subseteq \mathbf{V}(\mathcal{G})$, $P(\mathbf{Y} | \operatorname{do} \mathbf{X} = \mathbf{x})$ denotes the probability distribution of the variables $\mathbf{Y}$, when the causal model is generated using the above equations, but with the variables in $\mathbf{X}$ set to values given by $\mathbf{x}$. This distribution quantifies the *causal effect* of setting $\mathbf{X} = \mathbf{x}$ on $\mathbf{Y}$. In what follows, we usually do not specify the assignment of values $\mathbf{x}$. We say that $P(\mathbf{Y} | \operatorname{do} \mathbf{X})$ is *identifiable* if one can compute it, knowing only the causal graph $\mathcal{G}$ and the total probability distribution of the variables $\mathbf{V}(\mathcal{G})$ but not the specific functions $f_{i,t}$ or noise terms $\varepsilon_{i,t}$ giving rise to the causal model. For a more detailed introduction into graphical causal models and do-calculus, see Pearl (2009), and see Blondel et al. (2016) for time series graphs specifically.

We will make use of the well-known characterization of identifiability for causal effects in finite causal graphs, developed by Tian and Pearl (2002); Shpitser and Pearl (2006); Huang and Valtorta (2008). The following definitions and results are according to Shpitser and Pearl (2008):

Definition 6 (C-component) *An ADMG $\mathcal{G}$ where any pair of vertices is connected by a path of bidirected edges is called a C-component (confounded component).*

Definition 7 (forest) *An ADMG $\mathcal{G}$ in which each vertex has at most one child is called a forest. The set of vertices $\mathbf{R} \subseteq \mathbf{V}(\mathcal{G})$ that do not have any children is called the set of roots of $\mathcal{G}$.*

Definition 8 (C-forest) *An ADMG $\mathcal{G}$ that is both a C-component and a forest is called a C-forest.*

Now, for a vertex set $\mathbf{X}$ in an ADMG $\mathcal{G}$, let $\mathcal{G}_{\overline{\mathbf{X}}}$ be the graph that is obtained from $\mathcal{G}$ by deleting all incoming directed edges to $\mathbf{X}$. We say that a vertex V is an ancestor of vertex W if there is a directed path from V to W (using only directed edges). The set of ancestors of W with respect to the graph $\mathcal{G}$ is denoted $\mathrm{An}(W, \mathcal{G})$.

Definition 9 (hedge) *Let $\mathcal{G}$ be an ADMG on a set of variables $\mathbf{V}(\mathcal{G})$. For two disjoint vertex sets $\mathbf{X}, \mathbf{Y} \subseteq \mathbf{V}(\mathcal{G})$, a hedge for $\mathbf{X}, \mathbf{Y}$ is given by two subgraphs $\mathcal{F}, \mathcal{F}' \subseteq \mathcal{G}$ such that*

(1) $\mathcal{F}'$ is a subgraph of $\mathcal{F}$;

(2) $\mathbf{X}$ only occurs in $\mathcal{F}$, i.e. $\mathbf{X} \subseteq \mathbf{V}(\mathcal{F} \setminus \mathcal{F}')$;

(3) $\mathcal{F}$ and $\mathcal{F}'$ are C-forests;

(4) $\mathcal{F}$ and $\mathcal{F}'$ have the same set $\mathbf{R}$ of roots, and $\mathbf{R} \subseteq \mathrm{An}(\mathbf{Y}, \mathcal{G}_{\overline{\mathbf{X}}})$.

Lemma 1 (Causal ID algorithm (Shpitser and Pearl, 2008)) *Given a causal graph $\mathcal{G}$ and disjoint vertex sets $\mathbf{X}, \mathbf{Y} \subseteq \mathbf{V}(\mathcal{G})$, then $P(\mathbf{Y} | \operatorname{do} \mathbf{X})$ is unidentifiable if and only if there exists a hedge for $\mathbf{X}, \mathbf{Y}$ in $\mathcal{G}$. Moreover, there is a polynomial-time algorithm, the Causal ID algorithm 1, that can decide whether or not a hedge exists for a given pair of sets $\mathbf{X}, \mathbf{Y}$.*

Algorithm 1 is a simplified version of the Causal ID algorithm, only giving a Boolean answer in case that the causal effect is identifiable. Here, $\mathcal{G}[\mathbf{A}]$ denotes the induced graph on a vertex set $\mathbf{A} \subseteq \mathbf{V}(\mathcal{G})$, and $C(\mathcal{G})$ denotes the set of C-components of $\mathcal{G}$.

Algorithm 1 Causal ID algorithm (adapted from Shpitser and Pearl (2008))

Input: ADMG $\mathcal{G}$ and disjoint sets $\mathbf{X}, \mathbf{Y} \subseteq \mathbf{V}(\mathcal{G})$.
Output: **True if** $P(\mathbf{Y} | \operatorname{do} \mathbf{X})$ is identifiable, hedge $(\mathcal{F}, \mathcal{F}')$ otherwise.
if $\mathbf{X} = \emptyset$, **return True**.
if $\mathbf{V}(\mathcal{G}) \setminus \mathrm{An}(\mathbf{Y}) \neq \emptyset$, **return** $\mathrm{ID}(\mathcal{G}[\mathrm{An}(\mathbf{Y})], \mathbf{X} \cap \mathrm{An}(\mathbf{Y}), \mathbf{Y})$.
let $\mathbf{W} = \mathbf{V}(\mathcal{G}) \setminus (\mathbf{X} \cup \mathrm{An}(\mathbf{Y}, \mathcal{G}_{\overline{\mathbf{X}}}))$; **if** $\mathbf{W} \neq \emptyset$, **return** $\mathrm{ID}(\mathcal{G}, \mathbf{X} \cup \mathbf{W}, \mathbf{Y})$.
if $C(\mathcal{G} \setminus \mathbf{X}) = \{\mathcal{S}_1, \ldots, \mathcal{S}_k\}$, **return** $\bigwedge_i \mathrm{ID}(\mathbf{V}(\mathcal{G}) \setminus \mathbf{V}(\mathcal{S}_i), \mathbf{V}(\mathcal{S}_i), \mathcal{G})$.
if $C(\mathcal{G} \setminus \mathbf{X}) = \{\mathcal{S}\}$:
    if $C(\mathcal{G}) = \{\mathcal{G}\}$, **return** the hedge $(\mathcal{G}, \mathcal{G} \cap \mathcal{S})$.
    if $\exists \mathcal{S} \subseteq \mathcal{S}'$, such that $\mathcal{S}' \in C(\mathcal{G})$, **return** $\mathrm{ID}(\mathcal{G}[\mathbf{V}(\mathcal{S}')], \mathbf{X} \cap \mathcal{S}', \mathbf{Y})$.
    if $\mathcal{S} \in C(\mathcal{G})$, **return True**.

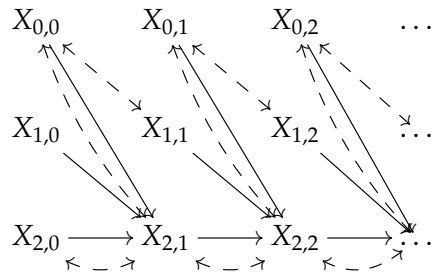

Figure 3: The effect $P(X_{2,2} | \operatorname{do} X_{1,1})$ is unidentifiable because of the unique hedge $(\mathcal{F}, \mathcal{F}')$ with $\mathbf{V}(\mathcal{F}) = \{X_{0,0}, X_{1,1}, X_{2,1}, X_{2,2}\}$ and $\mathbf{V}(\mathcal{F}') = \{X_{2,2}\}$. Note that this unidentifiability cannot be detected when only looking at the layers from $X_{1,1}$ onward.

3. Results

A priori, the Causal ID algorithm only works for causal graphs of finite size. However, note that in a periodic causal graph $\mathcal{G}$ corresponding to a time series, directed edges are only going forward in time. First of all, this implies that the causal effect from $X_{i,t}$ on $X_{i',t'}$ is trivially identifiable for $t > t'$ (as we have $P(X_{i',t'} | \operatorname{do} X_{i,t}) = P(X_{i',t'})$). Moreover, it is possible to compute identifiability of $P(X_{i',t'} | \operatorname{do} X_{i,t})$ for $t \leq t'$, by running the Causal ID algorithm on the segment of the graph $\mathcal{G}$ that includes all vertices up to time t'. To see this, note that any hedge for $X_{i,t}, X_{i',t'}$ can only consist of ancestors of $X_{i',t'}$. However, it is possible that there exists a hedge for $X_{i,t}, X_{i',t'}$ intersecting time layers strictly smaller than t, see Figure 3. This implies that even just identifying the causal effect between two single variables in the same layer t a priori requires running the Causal ID algorithm on $\mathcal{G}[0, t]$, which takes computing time that grows polynomially in wt. Moreover, using the Causal ID algorithm, one cannot answer the following question: Is the causal effect $P(X_{i',t'} | \operatorname{do} X_{i,t})$ identifiable for all $t' > t$?

Our results resolve both these problems. The following two statements are refinements of Theorem 1 that we stated in the introduction.

Proposition 1 *Let $\mathcal{G}$ be a periodic causal graph with width w and latency L and let $C = L \cdot 2^{Lw} \cdot (Lw+1)^{2Lw+2}$. For subsets $\mathbf{X}, \mathbf{Y} \subseteq \mathbf{V}(\mathcal{G})$, we have that $P(\mathbf{Y} | \operatorname{do} \mathbf{X})$ is unidentifiable if and only if there exists a hedge for $\mathbf{X}, \mathbf{Y}$ in $\mathcal{G}[\max(\mathsf{t}_{\min}(\mathbf{X}) - C, 0), \mathsf{t}_{\max}(\mathbf{Y})]$.*

Proposition 2 *Let $\mathcal{G}$ be a periodic causal graph with width w and latency L and let $C = L \cdot 2^{Lw} \cdot (Lw+1)^{2Lw+2}$. Consider subsets $\mathbf{X}, \mathbf{Y} \subseteq \mathbf{V}(\mathcal{G})$ such that $\mathsf{t}_{\max}(\mathbf{X}) \leq \mathsf{t}_{\min}(\mathbf{Y})$. If $P(\mathbf{Y} | \operatorname{do} \mathbf{X})$ is unidentifiable, then there exists a time shift $\Delta > \operatorname{dist}(\mathbf{X}, \mathbf{Y}) - C$, such that $P(\mathbf{Y}_{-\Delta} | \operatorname{do} \mathbf{X})$ is also unidentifiable.*

Note that part 1 of Theorem 1 follows from Proposition 1 and part 2 of Theorem 1 follows from repeatedly applying Proposition 2. In particular, these statements imply that whenever there are unidentifiable effects $P(\mathbf{Y} | \operatorname{do} \mathbf{X})$ in the causal graph $\mathcal{G}$, then they can be found by examining only a constant number of layers of $\mathcal{G}$, independent of the time distance between $\mathbf{X}$ and $\mathbf{Y}$ or the size of the past of the time series. Indeed, consider the following algorithm:

Algorithm 2 Decides identifiability of $P(\mathbf{Y}_{+\Delta} | \operatorname{do} \mathbf{X})$ for all $\Delta \geq 0$.

Input: Periodic causal graph $\mathcal{G}$ of width w and latency L, and disjoint sets $\mathbf{X}, \mathbf{Y} \subseteq \mathbf{V}(\mathcal{G})$ such that $\mathrm{t}_{\max}(\mathbf{X}) = \mathrm{t}_{\min}(\mathbf{Y})$.
Output: **True if** $P(\mathbf{Y}_{+\Delta} | \operatorname{do} \mathbf{X})$ is identifiable for all $\Delta \geq 0$, hedge $(\mathcal{F}, \mathcal{F}')$ for some $\mathbf{Y}_{+\Delta}, \mathbf{X}$ otherwise.
Let $C = L \cdot 2^{Lw} \cdot (Lw+1)^{2Lw+2}$.
For $\Delta = 0, \ldots, C-1$:
    Let $B = \mathrm{ID}(\mathcal{G}[\mathrm{t}_{\min}(\mathbf{X}) - C, \mathrm{t}_{\max}(\mathbf{Y}_{+\Delta})], \mathbf{X}, \mathbf{Y}_{+\Delta})$, **if** $B \neq$ **True**, **return** B.
return True.

Corollary 1 *Let $\mathcal{G}$ be a periodic causal graph and $\mathbf{X}, \mathbf{Y} \subseteq \mathbf{V}(\mathcal{G})$ such that $\mathrm{t}_{\max}(\mathbf{X}) = \mathrm{t}_{\min}(\mathbf{Y})$. Then, Algorithm 2 correctly computes whether $P(\mathbf{Y}_{+\Delta} | \operatorname{do} \mathbf{X})$ is identifiable for all $\Delta \in \mathbb{N}$.*

Corollary 1 follows directly from part 2 of Theorem 1. The value we are using for C in Algorithm 2 grows exponentially in the width w and the latency L of the causal graph. It would be of great interest to improve this dependence. The following statement shows that for some graphs, C has to grow at least linearly in w.

Theorem 2 *For infinitely many w, there exist periodic causal graphs $\mathcal{G}_w$ with width w and latency 1, such that $P(Y | \operatorname{do} X)$ is unidentifiable for some singletons $X, Y \in \mathbf{V}(\mathcal{G})$ but the causal effects $P(Z | \operatorname{do} X)$ are identifiable for all $Z \in \mathbf{V}(\mathcal{G})$ satisfying $\operatorname{dist}(X, Z) \leq \frac{w}{3} - 1$.*

4. Proofs

4.1. Cutting hedges to show Propositions 1, 2

The proof of Proposition 1 relies on the following idea: By Lemma 1 and time-directionality, we know that $P(\mathbf{Y} | \operatorname{do} \mathbf{X})$ is unidentifiable if and only if there is a hedge for $\mathbf{X}, \mathbf{Y}$ in $\mathcal{G}[0, \mathrm{t}_{\max}(\mathbf{Y})]$. Now if this hedge exits the section $\mathcal{G}[\mathrm{t}_{\min}(\mathbf{X}) - C, \mathrm{t}_{\max}(\mathbf{Y})]$ for a very large C, then it must be extremely stretched out. By periodicity of $\mathcal{G}$, there must then be two columns of $\mathcal{G}$ where the structure of the hedge is basically the same. We will show that cutting out the section of the graph between these two columns will result in a new hedge of smaller size. After doing this repeatedly, one eventually gets a hedge that is contained in $\mathcal{G}[\mathrm{t}_{\min}(\mathbf{X}) - C, \mathrm{t}_{\max}(\mathbf{Y})]$. Proposition 2 follows from the same idea, except that we are cutting out sections of the graph between $\mathbf{X}$ and $\mathbf{Y}$ as long as $\mathbf{X}$ and $\mathbf{Y}$ are too far away from each other. To formalize what we mean by the structure of the hedge in a given column, consider the following two definitions:

Definition 10 *Let $\mathcal{F} \subseteq \mathcal{G}$ be a subgraph of a periodic causal graph $\mathcal{G}$. We say that vertices $X_{i,t}$ and $X_{i',t}$ are left-connected in $\mathcal{F}$ if there exists a path of bidirected edges between them that goes only through vertices of $\mathcal{F}$ that are of the form $X_{j,t'}$, $t' \leq t$.*

Definition 11 *For two sets of vertices $\mathbf{X}, \mathbf{Y} \subseteq \mathcal{G}$, let $(\mathcal{F}, \mathcal{F}')$ be a hedge for $\mathbf{X}, \mathbf{Y}$. Let $\alpha_{\mathcal{F}}(t)$ be the ordered partition of $\{0, \ldots, w-1\}$ into $w+1$ blocks (some possibly empty), defined as follows: the first contains the row-indices of all vertices in $\mathbf{X}_{*,t} \setminus \mathbf{V}(\mathcal{F})$, while the subsequent blocks are the sets of row indices of the left-connected components of $\mathbf{X}_{*,t} \cap \mathbf{V}(\mathcal{F})$. If there are fewer than w such*

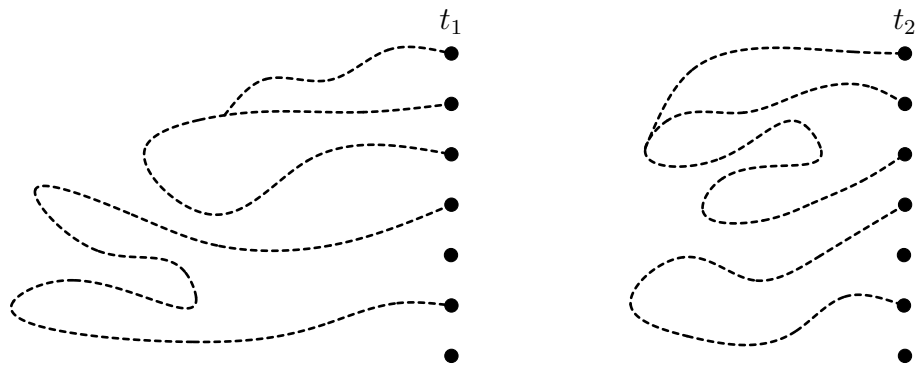

Figure 4: Two levels t_1 and t_2 with the same left-connected structure. Assuming that indices range from 0 to 6 from top to bottom, and all vertices and paths are in $\mathcal{F}$, we get $\alpha_{\mathcal{F}}(t_1) = \alpha_{\mathcal{F}}(t_2) = (\emptyset, \{1,2,3\}, \{4,6\}, \{5\}, \{7\})$.

blocks, then the final blocks of $\alpha_{\mathcal{F}}(t)$ are empty. The blocks which correspond to left-connected components are sorted by the smallest index i they contain. Likewise define $\alpha_{\mathcal{F}'}(t)$. Finally, define $\beta(t) = \mathbf{X}_{*,t} \cap \mathrm{An}(\mathbf{Y}, \mathcal{G}_{\overline{\mathbf{X}}})$.

See Figure 4 for an illustration of Definition 11. Next, we formalize what we mean by cutting out a segment of the graph.

Definition 12 *We define the map $\Phi_{b,\Delta} : \mathcal{G} \setminus \mathcal{G}[b+1, b+\Delta] \to \mathcal{G}$ to be the map that acts as the identity on $\mathcal{G}[0,b]$ and translates $\mathcal{G}[b+\Delta+1, \infty)$ by $-\Delta$ layers. That is,*

$$\Phi_{b,\Delta}(X_{i,t}) = \begin{cases} X_{i,t} \text{ if } t \leq b \\ X_{i,t-\Delta} \text{ if } t > b+\Delta. \end{cases}$$

The map Φ can be thought of as an operation that cuts the part $\mathcal{G}[b+1, b+\Delta]$ out of the graph $\mathcal{G}$ and then rewires the loose ends at layers b and $b+\Delta+1$ together according to the periodic graph structure of $\mathcal{G}$. The key property of Φ is the following: due to the periodic structure of $\mathcal{G}$, whenever (V, W) is a directed (resp. bidirected) edge for $V, W \in \mathbf{V}(\mathcal{G} \setminus \mathcal{G}[b+1, b+\Delta])$, then $(\Phi(V), \Phi(W))$ is also a directed (resp. bidirected) edge in $\mathcal{G}$ (and the converse almost holds, except if $V \in \mathbf{X}_{*,b}$ and $W \in \mathbf{X}_{*,b+\Delta+1}$). Here, we are assuming that $\mathcal{G}$ is of latency 1, and we will cover graphs with higher latency later. Our main technical result states that cutting at two layers on which the values of $\alpha_{\mathcal{F}}, \alpha_{\mathcal{F}'}$, and β align for a hedge $(\mathcal{F}, \mathcal{F}')$ creates a new (and smaller) hedge:

Lemma 2 *Let $\mathcal{G}$ be a periodic causal graph of latency $L = 1$, and let $(\mathcal{F}, \mathcal{F}')$ be a hedge for the vertex sets $\mathbf{X}, \mathbf{Y} \subseteq \mathcal{G}$. Assume $\mathcal{G}[b+1, b+\Delta] \cap (\mathbf{X} \cup \mathbf{Y}) = \emptyset$, and we have $\alpha_{\mathcal{F}}(b) = \alpha_{\mathcal{F}}(b+\Delta), \alpha_{\mathcal{F}'}(b) = \alpha_{\mathcal{F}'}(b+\Delta)$, and $\beta(b) = \beta(b+\Delta)$. Then, $(\Phi_{b,\Delta}(\mathcal{F} \setminus \mathcal{G}[b+1, b+\Delta]), \Phi_{b,\Delta}(\mathcal{F}' \setminus \mathcal{G}[b+1, b+\Delta]))$ is a hedge for $\Phi_{b,\Delta}(\mathbf{X}), \Phi_{b,\Delta}(\mathbf{Y})$.*

Proof Fix b and Δ satisfying the hypotheses of Lemma 2. To simplify notation, let

$$\Phi = \Phi_{b,\Delta},$$
$$\widetilde{\mathcal{F}} = \Phi(\mathcal{F} \setminus \mathcal{G}[b+1, b+\Delta]),$$
$$\widetilde{\mathcal{F}}' = \Phi(\mathcal{F}' \setminus \mathcal{G}[b+1, b+\Delta]).$$

Following Definition 9, it suffices to prove that the following statements hold:

(a) $\widetilde{\mathcal{F}}' \subseteq \widetilde{\mathcal{F}}$;

(b) $\Phi(\mathbf{X}) \subseteq \mathbf{V}(\widetilde{\mathcal{F}} \setminus \widetilde{\mathcal{F}}')$;

(c) $\widetilde{\mathcal{F}}$ and $\widetilde{\mathcal{F}}'$ are C-connected.

(d) Every vertex in $\widetilde{\mathcal{F}}$ has at most one child.

(e) $\widetilde{\mathcal{F}}$ and $\widetilde{\mathcal{F}}'$ have the same set $\widetilde{\mathbf{R}}$ of roots, and $\widetilde{\mathbf{R}} \subseteq \mathrm{An}(\Phi(\mathbf{Y}), \mathcal{G}_{\overline{\Phi(\mathcal{X})}})$.

Statements (a) and (b) follow straight from the definitions and the fact that Φ is injective. To show Statement (c), we first prove that any two vertices $X_{i,b}, X_{j,b} \in \mathbf{V}(\widetilde{\mathcal{F}})$ on layer b are C-connected in $\widetilde{\mathcal{F}}$. Since $\alpha_{\mathcal{F}}(b) = \alpha_{\mathcal{F}}(b+\Delta)$, we know that $X_{i,b+\Delta}, X_{j,b+\Delta} \in \mathbf{V}(\mathcal{F})$, so there must be a path of bidirected edges in $\mathcal{G}$ that connects these two vertices and only goes through vertices of $\mathcal{F}$. Suppose this path hits layer $b+\Delta$ exactly m times at the vertices $X_{k_1,b+\Delta}, \ldots, X_{k_m,b+\Delta}$ in this order, where $X_{k_1,b+\Delta} = X_{i,b+\Delta}$ and $X_{k_m,b+\Delta} = X_{j,b+\Delta}$. We claim that then there is a path of bidirectional edges connecting $X_{k_1,b}, \ldots, X_{k_m,b}$ in $\widetilde{\mathcal{F}}$. Observe that, since $\mathcal{G}$ has latency 1, for any $\ell \in \{1, \ldots, m-1\}$ the path from $X_{k_\ell,b+\Delta}$ to $X_{k_{\ell+1},b+\Delta}$ in $\mathcal{F}$ is either completely contained in $\mathcal{G}[0, b+\Delta]$ or in $\mathcal{G}[b+\Delta, \infty)$. In the first case, $X_{k_\ell,b+\Delta}$ and $X_{k_{\ell+1},b+\Delta}$ are left-connected in $\mathcal{F}$. Hence, $X_{k_\ell,b}$ and $X_{k_{\ell+1},b}$ are also left-connected in $\mathcal{F}$, but to the left of layer b, we have $\mathcal{F} \cap \mathcal{G}[0,b] = \widetilde{\mathcal{F}} \cap \mathcal{G}[0,b]$, so both vertices are also left-connected in $\widetilde{\mathcal{F}}$. In the second case, if the path from $X_{k_\ell,b+\Delta}$ to $X_{k_{\ell+1},b+\Delta}$ in $\mathcal{F}$ stays to the right of layer $b+\Delta$, then we can just translate it by $-\Delta$ layers to get a path from $X_{k_\ell,b}$ to $X_{k_{\ell+1},b}$ in $\widetilde{\mathcal{F}}$. Hence, $X_{i,b}$ and $X_{j,b}$ are C-connected in $\widetilde{\mathcal{F}}$. Now we show that any two vertices $\widetilde{A}, \widetilde{B} \in \mathbf{V}(\widetilde{\mathcal{F}})$ are C-connected in $\widetilde{\mathcal{F}}$. Consider the path π of bidirectional edges between $A = \Phi^{-1}(\widetilde{A})$ and $B = \Phi^{-1}(\widetilde{B})$ that goes through $\mathcal{F}$. If the path π never hits layer b or $b+\Delta$, then $\Phi(\pi)$ is a path between $\widetilde{A}$ and $\widetilde{B}$ (it either stays the same or it gets translated as a whole). Otherwise, let A' and B' be the first vertices that are on layer b or $b+\Delta$ and get hit by the path π when starting from A or B respectively. Let π_1 denote the path from A to A' and π_2 denote the path from B to B'. Then $\Phi(\pi_1)$ is a path from $\widetilde{A}$ to $\Phi(A')$ in $\widetilde{\mathcal{F}}$ and $\Phi(\pi_2)$ is a path from $\widetilde{B}$ to $\Phi(B')$ in $\widetilde{\mathcal{F}}$. Both, $\Phi(A'), \Phi(B') \in \mathbf{V}(\widetilde{\mathcal{F}})$ are in layer b, so they must also be C-connected in $\widetilde{\mathcal{F}}$. Hence, $\widetilde{A}$ and $\widetilde{B}$ are C-connected in $\widetilde{\mathcal{F}}$. This shows $\widetilde{\mathcal{F}}$ is C-connected and the same statement follows for $\widetilde{\mathcal{F}}'$ analogously.

To show Statement (d), assume that $\widetilde{A} \in \mathbf{V}(\widetilde{\mathcal{F}})$ has at least two children $\widetilde{B}, \widetilde{C}$ in $\widetilde{\mathcal{F}}$. Let $A = \Phi^{-1}(\widetilde{A})$, $B = \Phi^{-1}(\widetilde{B})$, and $C = \Phi^{-1}(\widetilde{C})$. We know that $A, B, C \in \mathbf{V}(\mathcal{F})$. Moreover, if A, B, C are fully contained in $\mathcal{G}[0,b]$ or A, B, C are fully contained in $\mathcal{G}[b+\Delta+1, \infty)$, then B, C are children of A, which is a contradiction to $\mathcal{F}$ being part of a hedge. Hence, we must have $A \in \mathbf{X}_{*,b}$ and $B, C \in \mathbf{X}_{*,b+\Delta+1}$. Let $A_{+\Delta} \in \mathbf{X}_{b+\Delta}$ be the vertex A shifted Δ layers to the right.

Then, B and C are children of $A_{+\Delta}$ but since $\alpha_{\mathcal{F}}(b) = \alpha_c F(b+\Delta)$, we have $A_{+\Delta} \in \mathbf{V}(\mathcal{F})$. This is again a contradiction to $\mathcal{F}$ being part of a valid hedge and shows the statement.

To prove Statement (e), observe first that one can show $\widetilde{\mathcal{F}}$ and $\widetilde{\mathcal{F}}'$ have the same set $\widetilde{\mathbf{R}}$ of roots using the same logic as the proof of Statement (d): Whenever there is a vertex $\widetilde{R}$, which is a root of $\widetilde{\mathcal{F}}$ but not of $\widetilde{\mathcal{F}}'$ or vice-versa, one can map it back (and potentially switch from layer b to $b+\Delta$) to find a vertex R, which is a root of $\mathcal{F}$ but not of $\mathcal{F}'$ or vice-versa, leading to a contradiction. Hence, it only remains to show that $\widetilde{\mathbf{R}} \subseteq \mathrm{An}(\Phi(\mathbf{Y}), \mathcal{G}_{\overline{\Phi(\mathbf{X})}})$. Consider a vertex $\widetilde{R} \in \widetilde{\mathbf{R}}$ and its preimage $R = \Phi^{-1}(\widetilde{R}) \in \mathbf{R}$, where $\mathbf{R}$ is the set of roots of $\mathcal{F}, \mathcal{F}'$. We know that there is a directed path π in $\mathcal{G}_{\overline{X}}$ from R to some vertex $Y \in \mathbf{Y}$. If π does not hit layer b, then it cannot hit the region $\mathcal{G}[b, b+\Delta]$ either, so $\Phi(\pi)$ is a directed path in $\mathcal{G}_{\overline{\Phi(\mathbf{X})}}$ from $\widetilde{R}$ to $\Phi(Y) \in \Phi(\mathbf{Y})$. Otherwise, let $X_{i,b}$ be the first vertex that π hits in layer b. Since $X_{i,b}$ is an ancestor of $\mathbf{Y}$ in $\mathcal{G}_{\overline{X}}$ and $\beta(b) = \beta(b+\Delta)$, we know that there is a directed path τ in $\mathcal{G}_{\overline{X}}$ from the vertex $X_{i,b+\Delta}$ to some vertex Y' in $\mathbf{Y}$. Hence, following the path $\Phi(\pi)$ from $\widetilde{R}$ to $X_{i,b}$ and the path $\Phi(\tau)$ from $X_{i,b}$ to $\Phi(Y')$ results in a path from $\widetilde{R}$ to $\Phi(\mathbf{Y})$ in $\mathcal{G}_{\overline{\Phi(\mathbf{X})}}$. This completes the proof of Lemma 2. ■

Now we show how Lemma 2 implies Proposition 1.

Proof (of Proposition 1). Let $\mathcal{G}$ be a periodic causal graph of width w and latency L, and $\mathbf{X}, \mathbf{Y} \subseteq \mathbf{V}(\mathcal{G})$. Wlog we may assume that $\mathrm{t}_{\min}(\mathbf{X}) \leq \mathrm{t}_{\min}(\mathbf{Y})$. By Lemma 1, it suffices to show that if there exists a hedge for $\mathbf{X}, \mathbf{Y}$, then there is also a hedge for $\mathbf{X}, \mathbf{Y}$ in $\mathcal{G}[\mathrm{t}_{\min}(\mathbf{X}) - C, \mathrm{t}_{\max}(\mathbf{Y})]$. Note that, by time-directionality, any hedge for $\mathbf{X}, \mathbf{Y}$ is contained in $\mathcal{G}[0, \mathrm{t}_{\max}(\mathbf{Y})]$. Now, to be able to apply Lemma 2, we transform $\mathcal{G}$ into a periodic causal graph $\mathcal{H}$ of width Lw and latency 1, by simply relabeling the vertices of $\mathcal{G}$ such that layer t of $\mathcal{H}$ contains the vertices of the layers $Lt, Lt+1, \ldots, Lt+t-1$ of $\mathcal{G}$ (that is, we aggregate each L consecutive time steps of $\mathcal{G}$). Let $\mathbf{X}', \mathbf{Y}'$ be the sets in $\mathbf{V}(\mathcal{H})$ corresponding to $\mathbf{X}, \mathbf{Y}$. Since the graphs $\mathcal{G}$ and $\mathcal{H}$ are isomorphic, if there exists a hedge for $\mathbf{X}, \mathbf{Y}$ in $\mathcal{G}$, then there exists a hedge for $\mathbf{X}', \mathbf{Y}'$ in $\mathcal{H}$. Let $(\mathcal{F}, \mathcal{F}')$ such a hedge with maximal $\mathrm{t}_{\min}(\mathcal{F})$. We claim that $\mathrm{t}_{\min}(\mathbf{X}') - \mathrm{t}_{\min}(\mathbf{V}(\mathcal{F})) \leq C/L = 2^{Lw}(Lw+1)^{2Lw+2}$.

Suppose not, and consider the functions $\alpha_{\mathcal{F}}(t)$, $\alpha_{\mathcal{F}'}(t)$ and $\beta(t)$. Since $\mathcal{H}$ has width Lw, both the partitions $\alpha_{\mathcal{F}}(t)$ and $\alpha_{\mathcal{F}'}(t)$ can attain at most $(Lw+1)^{Lw+1}$ values, while the function $\beta(t)$ can attain at most 2^{Lw} values. Hence, by the pigeonhole principle, there must exist some layers b and $b+\Delta$, such that $\mathrm{t}_{\min}(\mathcal{F}) \leq b < b+\Delta < \mathrm{t}_{\min}(\mathbf{X}')$ and $\alpha_{\mathcal{F}}(b) = \alpha_{\mathcal{F}}(b+\Delta)$, $\alpha_{\mathcal{F}'}(b) = \alpha_{\mathcal{F}'}(b+\Delta)$, and $\beta(b) = \beta(b+\Delta)$. Now, by Lemma 2, there exists a hedge $(\widetilde{\mathcal{F}}, \widetilde{\mathcal{F}}')$ for $\mathbf{X}'_{-\Delta}, \mathbf{Y}'_{-\Delta}$ that has $\mathrm{t}_{\min}(\widetilde{\mathcal{F}}) = \mathrm{t}_{\min}(\mathcal{F})$. After shifting this hedge by Δ layers to the right, we get another hedge for $\mathbf{X}', \mathbf{Y}'$ with strictly larger minimum layer than $\mathrm{t}_{\min}(\mathcal{F})$. This is a contradiction to our minimality assumption, and hence $(\mathcal{F}, \mathcal{F}')$ must have been contained in $\mathcal{H}[\mathrm{t}_{\min}(\mathbf{X}') - C/L, \mathrm{t}_{\max}(\mathbf{Y}')]$. After reversing the aggregation of layers, we find a corresponding hedge for $\mathbf{X}, \mathbf{Y}$ contained in $\mathcal{G}[\mathrm{t}_{\min}(\mathbf{X}) - C, \mathrm{t}_{\max}(\mathbf{Y})]$, which completes the proof.

■

The proof of Proposition 2 is analogous, except that we are using the pidgeonhole principle on the layers between $\mathrm{t}_{\max}(\mathbf{X})$ and $\mathrm{t}_{\min}(\mathbf{Y})$ to apply Lemma 2.

4.2. Constructing examples to show Theorem 2

The goal of this section is to construct periodic causal models $\mathcal{G}_w$ that satisfy the statement of Theorem 2. For this, it is enough to define an appropriate acyclic graph $\mathcal{A}_w$ on $\mathbf{X}_{*,0} \cup \mathbf{X}_{*,1}$. We

can extend $\mathcal{A}_w$ to a periodic, latency 1 graph $\mathcal{G}_w$ by including a directed (resp. bidirected) edge $(X_{i,t}, X_{i',t'})$ in $\mathcal{G}_w$ if $(X_{i,0}, X_{i',t'-t})$ is in $\mathcal{A}_w$.

Specifically, let $w = 3k + 1$ and define $\mathcal{G}_w$ to have (see Figure 5):

- Directed edges from every $X_{i,t}$ to $X_{i,t+1}$ and to $X_{i+3,t+1}$.
- Bidirected edges from every $X_{i,t}$ to $X_{i+1,t+1}$ and to $X_{i+2,t+1}$.
- In both cases the index i ranges from 0 to $w - 1$ and we use addition modulo w.

usetikzlibraryarrows.meta

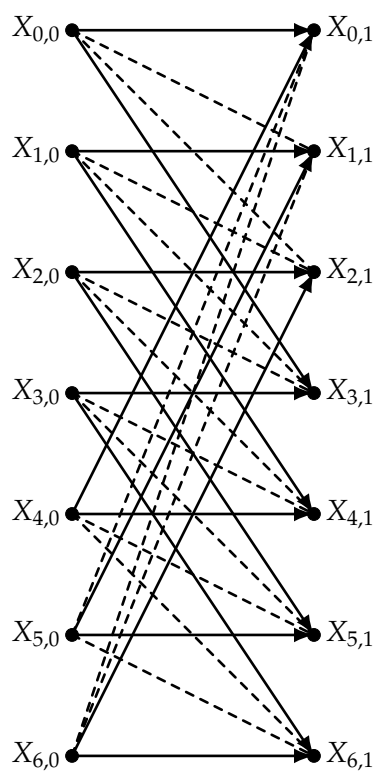

Figure 5: The graph $\mathcal{A}_7$.

First, we claim that in $\mathcal{G}_w$, the causal effect $P(X_{w-1,w-2} | \text{ do } X_{0,0})$ is unidentifiable. To show this, we construct a hedge for $X_{0,0}, X_{w-1,w-2}$. Our hedge consists of the following sets of vertices (see Figure 6)

$$\mathcal{F} = \{X_{0,0}, X_{0,1}, X_{1,1}, X_{w-3,0}, X_{w-2,0}, X_{w-1,0}, X_{w-1,1}\}$$
$$\mathcal{F}' = \{X_{0,1}, X_{1,1}, X_{w-3,0}, X_{w-2,0}, X_{w-1,0}, X_{w-1,1}\}$$

Both $\mathcal{F}$ and $\mathcal{F}'$ have the same set of roots $\mathbf{R} = \{X_{0,1}, X_{1,1}, X_{w-1,1}\}$. We now claim that there are directed paths from each of these roots to our effect vertex $X_{w-1,w-2}$. Clearly, since we have directed edges from $X_{i,j}$ to $X_{i,j+1}$, there is a ("horizontal") directed path from $X_{w-1,1}$ to $X_{w-1,w-2}$. Additionally, since w is of the form $3k + 1$, $X_{0,1}$ has a directed ("diagonal") path to the directed

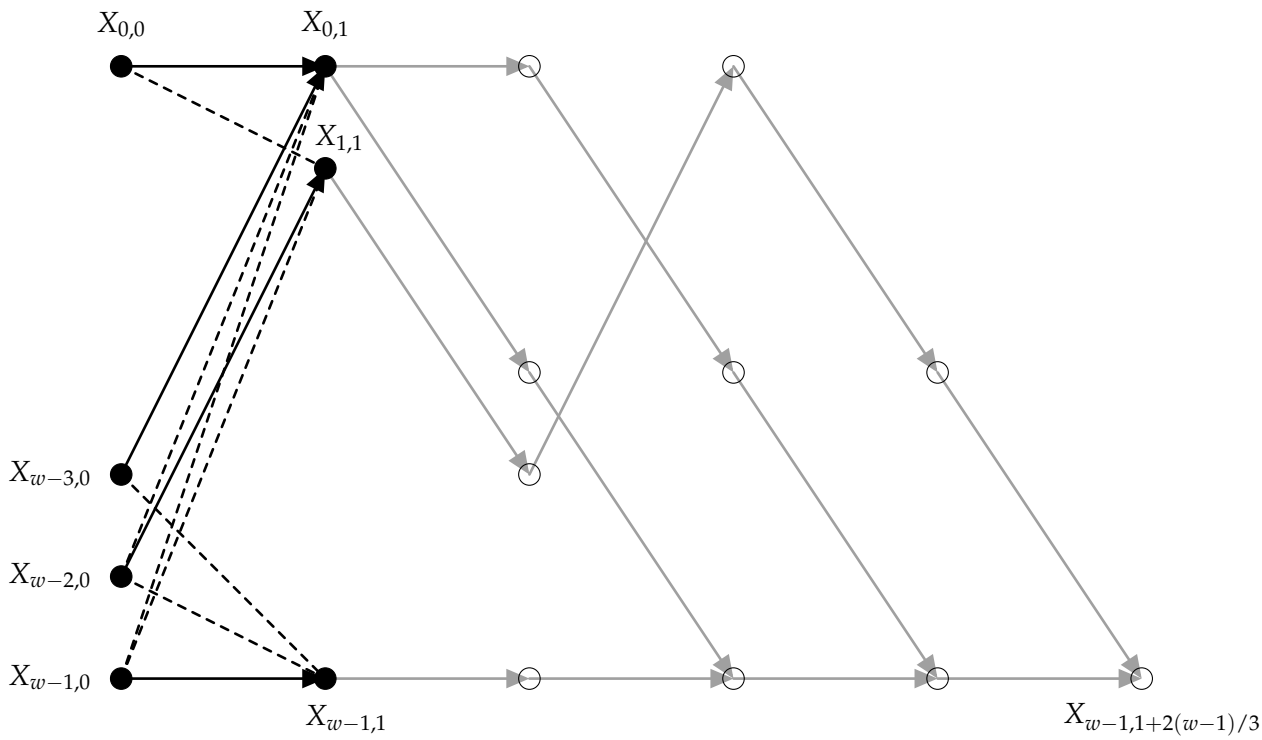

Figure 6: Hedge for $w = 7$

chain of vertices in row $w-1$ (meeting it at the vertex $X_{w-1,(w-1)/3+1}$), and therefore, it has a directed path to $X_{w-1,w-2}$. Finally, from $X_{1,1}$ there is a diagonal path of length $2(w-1)/3$ which wraps around (vertically) until it reaches the vertex $X_{w-1,2(w-1)/3+1}$, from which a path continues horizontally to $X_{w-1,w-2}$. (This last wrap-around path is key to the construction of the hedge. Indeed, if the graph were constructed with the same edge types but without vertical coordinates being taken modulo w, $\Pr(X_{j,t}|\text{ do } X_{0,0})$ would be identifiable for all $X_{j,t}$.) Finally, it is easy to verify that the bidirected edges make both $\mathcal{F}$ and $\mathcal{F}'$ C-connected; this is illustrated for $w = 7$ in Figure 6.

Due to the existence of this hedge, $\Pr(X_{w-1,w-2}|\text{ do } X_{0,0})$ is not identifiable. We now prove that for every $t \leq \frac{w}{3} - 1$ and every j, $P(X_{j,t}|\text{ do } X_{0,0})$ is identifiable.

First, assume j is not divisible by 3. In this case, there is no directed path from $X_{0,0}$ to $X_{j,t}$. This can be seen as all directed edges vertically descend by either 0 rows or 3 rows, so all descendants of $X_{0,0}$ have a first index that is divisible by 3 up to time $\frac{w-1}{3}$. We deduce that the causal effect $P(X_{j,t}|\text{ do } X_{0,0})$ is trivial, and hence identifiable.

It remains to consider the vertices $X_{j,t}$ with $3|j$. The hedge for such a vertex must consist solely of ancestors of $X_{j,t}$, which are vertices of the form $X_{j',t'}$ with $t' \leq t$, and j' satisfying either $0 \leq j' \leq j$, or else $w + j - 3(t - t') \leq j' \leq w - 1$.

As a consequence, there are two rows, namely the rows indexed $j+1$ and $j+2$, which cannot contain the vertices of a hedge for $X_{j,t}$. Therefore (after re-indexing the rows to start from $j+1$ and end at j), we can consider all directed and bidirected edges among hedge vertices to occur within a vertical segment, all directed edges skipping 0 or 3 rows, and all bidirected edges skipping 1 or 2

rows. The bidirected edges therefore can never connect two hedge vertices. The larger part of the hedge, $\mathcal{F}$, must contain more than one vertex. It therefore cannot be C-connected. Consequently, there is no hedge, so $P(X_{j,t} | \operatorname{do} X_{0,0})$ is identifiable. This proves Theorem 2.

5. Discussion

There remains an exponential gap between our upper and lower bounds on the number C of layers that are necessary to consider when deciding identifiability of a causal effect in a time series graph. We suspect that the lower bound is closer to the truth, but further work is needed to confirm this suspicion. If this can be done, it will make the result much more attractive computationally. Note that our upper bound on C is for graphs of generic periodic structure. It would be an interesting question of whether the bound can be improved given certain structural conditions on the periodic causal graph.

Our setting is also generic in that we do not impose time-invariance on the functional dependencies of the underlying causal system. This is why our results focus on deciding whether causal effects are identifiable rather than identifying them, as we do not expect a non-trivial speed-up for the latter (at least not in the sense of part 2 of Theorem 1). Efficient identification in time series graphs, potentially under the additional assumption of time-invariant dynamics, is therefore still a subject for further research.

A separate topic of interest is the condition number of the Causal ID mapping in periodic graphs (see Schulman and Srivastava (2016) for a definition). When one tries to translate an in-principle "perfect statistics" result into an actual application with empirical data, the sample size scales roughly proportionally to this condition number. It is known that, in general, the condition number can vary widely between different graphs: it can be exponential in the size of the graph (Schulman and Srivastava, 2016), but on the other hand is also known to scale only as $\exp(c \log c)$ if the C-components of the graph are of size bounded by c (Gordon et al., 2021). It seems quite likely that the periodic structure should impose further constraints that will help to upper bound the condition number, but this is at present an entirely open question.

Acknowledgments

This work was supported by NSF grant CCF-2321079.

References

Charles K. Assaad, Emilie Devijver, and Eric Gaussier. Survey and evaluation of causal discovery methods for time series. *J. Artif. Int. Res.*, 73, May 2022. ISSN 1076-9757. doi: 10.1613/jair.1.13428. URL https://doi.org/10.1613/jair.1.13428.

Gilles Blondel, Marta Arias, and Ricard Gavaldà. Identifiability and transportability in dynamic causal networks. *International Journal of Data Science and Analytics*, 3(2):131–147, Oct 2016. doi: 10.1007/s41060-016-0028-8.

Michael Eichler and Vanessa Didelez. Causal reasoning in graphical time series models. In *Proceedings of the Twenty-Third Conference on Uncertainty in Artificial Intelligence*, UAI'07, page 109–116, Arlington, Virginia, USA, 2007. AUAI Press. ISBN 0974903930.

Michael Eichler and Vanessa Didelez. On granger causality and the effect of interventions in time series. *Lifetime Data Analysis*, 16(1):3–32, Nov 2009. doi: 10.1007/s10985-009-9143-3.

Andreas Gerhardus and Jakob Runge. High-recall causal discovery for autocorrelated time series with latent confounders. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 12615–12625. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/94e70705efae423efda1088614128d0b-Paper.pdf.

Spencer L. Gordon, Vinayak M. Kumar, Leonard J. Schulman, and Piyush Srivastava. Condition number bounds for causal inference. In C. de Campos and M. H. Maathuis, editors, *Proc. Thirty-Seventh Conference on Uncertainty in Artificial Intelligence*, volume 161 of *PMLR*, pages 1948–1957, 2021. URL https://proceedings.mlr.press/v161/gordon21a.html.

Yimin Huang and Marco Valtorta. On the completeness of an identifiability algorithm for semi-Markovian models. *Ann. Math. Artif. Intell.*, 54(4):363–408, December 2008.

Daniel Malinsky and Peter Spirtes. Causal structure learning from multivariate time series in settings with unmeasured confounding. In Thuc Duy Le, Kun Zhang, Emre Kıcıman, Aapo Hyvärinen, and Lin Liu, editors, *Proceedings of 2018 ACM SIGKDD Workshop on Causal Disocvery*, volume 92 of *Proceedings of Machine Learning Research*, pages 23–47. PMLR, 20 Aug 2018. URL https://proceedings.mlr.press/v92/malinsky18a.html.

Atalanti A Mastakouri, Bernhard Schölkopf, and Dominik Janzing. Necessary and sufficient conditions for causal feature selection in time series with latent common causes. In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 7502–7511. PMLR, 18–24 Jul 2021. URL https://proceedings.mlr.press/v139/mastakouri21a.html.

Raha Moraffah, Paras Sheth, Mansooreh Karami, Anchit Bhattacharya, Qianru Wang, Anique Tahir, Adrienne Raglin, and Huan Liu. Causal inference for time series analysis: Problems, methods and evaluation, Dec 2021. URL https://dl.acm.org/doi/abs/10.1007/s10115-021-01621-0.

Judea Pearl. *Causality*. Cambridge, 2nd edition, 2009.

Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. *Elements of Causal Inference*. MIT Press, 2017.

Niklas Pfister, Peter Bühlmann, and Jonas Peters. Invariant causal prediction for sequential data. *Journal of the American Statistical Association*, 114(527):1264–1276, 2019. doi: 10.1080/01621459.2018.1491403. URL https://doi.org/10.1080/01621459.2018.1491403.

Jakob Runge. Causal network reconstruction from time series: From theoretical assumptions to practical estimation. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 28(7):075310, 07 2018. ISSN 1054-1500. doi: 10.1063/1.5025050. URL https://doi.org/10.1063/1.5025050.

Jakob Runge, Sebastian Bathiany, Erik Bollt, Gustau Camps-Valls, Dim Coumou, Ethan Deyle, Clark Glymour, Marlene Kretschmer, Miguel D. Mahecha, Jordi Muñoz-Marí, and et al. Inferring causation from time series in earth system sciences. *Nature Communications*, 10(1), Jun 2019a. doi: 10.1038/s41467-019-10105-3.

Jakob Runge, Peer Nowack, Marlene Kretschmer, Seth Flaxman, and Dino Sejdinovic. Detecting and quantifying causal associations in large nonlinear time series datasets. *Science Advances*, 5 (11):eaau4996, 2019b. doi: 10.1126/sciadv.aau4996. URL https://www.science.org/doi/abs/10.1126/sciadv.aau4996.

Leonard J. Schulman and Piyush Srivastava. Stability of causal inference. In *Proc. Thirty-Second Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 666–675. AUAI Press, 2016. URL http://dl.acm.org/citation.cfm?id=3020948.3021017.

Ilya Shpitser and Judea Pearl. Identification of joint interventional distributions in recursive semi-Markovian causal models. In *Proc. 20th AAAI Conference on Artifical Intelligence*, pages 1219–1226. AAAI Press, July 2006.

Ilya Shpitser and Judea Pearl. Complete identification methods for the causal hierarchy. *Journal of Machine Learning Research*, 9(64):1941–1979, 2008. URL http://jmlr.org/papers/v9/shpitser08a.html.

Jin Tian and Judea Pearl. A general identification condition for causal effects. In *Proc. 18th AAAI Conference on Artificial Intelligence*, pages 567–573. AAAI Press, 2002.

Xiangzhou Zhang, Yong Hu, Kang Xie, Shouyang Wang, E.W.T. Ngai, and Mei Liu. A causal feature selection algorithm for stock prediction modeling. *Neurocomputing*, 142:48–59, 2014. ISSN 0925-2312. doi: https://doi.org/10.1016/j.neucom.2014.01.057. URL https://www.sciencedirect.com/science/article/pii/S0925231214005359. SI Computational Intelligence Techniques for New Product Development.