The Annals of Statistics
2020, Vol. 48, No. 1, 539–559
https://doi.org/10.1214/19-AOS1821

MARKOV EQUIVALENCE OF MARGINALIZED LOCAL INDEPENDENCE GRAPHS

BY SØREN WENGEL MOGENSEN* AND NIELS RICHARD HANSEN**

Department of Mathematical Sciences, University of Copenhagen, *swengel@math.ku.dk; **niels.r.hansen@math.ku.dk

Symmetric independence relations are often studied using graphical representations. Ancestral graphs or acyclic directed mixed graphs with m-separation provide classes of symmetric graphical independence models that are closed under marginalization. Asymmetric independence relations appear naturally for multivariate stochastic processes, for instance, in terms of local independence. However, no class of graphs representing such asymmetric independence relations, which is also closed under marginalization, has been developed. We develop the theory of directed mixed graphs with μ-separation and show that this provides a graphical independence model class which is closed under marginalization and which generalizes previously considered graphical representations of local independence.

Several graphs may encode the same set of independence relations and this means that in many cases only an equivalence class of graphs can be identified from observational data. For statistical applications, it is therefore pivotal to characterize graphs that induce the same independence relations. Our main result is that for directed mixed graphs with μ-separation each equivalence class contains a maximal element which can be constructed from the independence relations alone. Moreover, we introduce the directed mixed equivalence graph as the maximal graph with dashed and solid edges. This graph encodes all information about the edges that is identifiable from the independence relations, and furthermore it can be computed efficiently from the maximal graph.

1. Introduction. Graphs have long been used as a formal tool for reasoning with independence models. Most work has been concerned with symmetric independence models arising from standard probabilistic independence for discrete or real-valued random variables. However, when working with dynamical processes it is useful to have a notion of independence that can distinguish explicitly between the present and the past, and this is a key motivation for considering local independence.

The notion of local independence was introduced for composable Markov processes by Schweder [37] who also gave examples of graphs describing local independence structures. Aalen [1] discussed how one could extend the definition of local independence in the broad class of semimartingales using the Doob–Meyer decomposition. Several authors have since then used graphs to represent local independence structures of multivariate stochastic process models, in particular for point process models; see, for example, [4, 11–13, 35]. Local independence takes a dynamical point of view in the sense that it evaluates the dependence of the present on the past. This provides a natural link to statistical causality as cause must necessarily precede effect [1, 2, 28, 37]. Furthermore, recent work argues that for some applications it can be important to consider continuous-time models, rather than only cross-sectional models, when trying to infer causal effects [3].

Received February 2018; revised January 2019.
MSC2010 subject classifications. Primary 62M99; secondary 62A99.
Key words and phrases. Directed mixed graphs, independence model, local independence, local independence graph, Markov equivalence, μ-separation.

Local independence for point processes has been applied for data analysis (see, e.g., [2, 23, 44]), but in applications a direct causal interpretation may be invalid if only certain dynamical processes are observed while other processes of the system under study are unobserved. Allowing for such latent processes is important for valid causal inference, and this motivates our study of representations of marginalized local independence graphs.

Graphical representations of independence models have also been studied for time series [14–17]. In the time series context—using the notion of Granger causality—Eichler [15] gave an algorithm for learning a graphical representation of local independence. However, the equivalence class of graphs that yield the same local independences was not identified, and thus the learned graph does not have any clear causal interpretation. Related research has been concerned with inferring the graph structure from subsampled time series, but under the assumption of no latent processes; see, for example, [9, 22].

In this paper, we give a formal, graphical framework for handling the presence of unobserved processes and extend the work on graphical representations of local independence models by formalizing marginalization and giving results on the equivalence classes of such graphical representations. The graphical framework that we propose is a generalization of that of Didelez [11–13]. This development is analogous to work on marginalizations of graphical models using directed acyclic graphs, DAGs. Starting from a DAG, one can find graphs (e.g., maximal ancestral graphs or acyclic directed mixed graphs) that encode marginal independence models [8, 18, 19, 25, 33, 34, 36, 39]. One can then characterize the equivalence class of graphs that yield the same independence model [5, 45]—the so-called Markov equivalent graphs—and construct learning algorithms to find such an equivalence class from data. The purpose of this paper is to develop the necessary theoretical foundation for learning local independence graphs by developing a precise characterization of the learnable object: the class of Markov equivalent graphs.

The paper is structured as follows: in Section 2, we discuss abstract independence models, relevant graph-theoretical concepts and the notion of local independence and local independence graphs. In Section 3, we introduce μ-separation for directed mixed graphs, which will be used to represent marginalized local independence graphs, and we describe an algorithm to marginalize a given local independence graph. In Sections 4 and 5, we develop the theory of μ-separation for directed mixed graphs further, and we discuss, in particular, Markov equivalence of such graphs. All proofs of the main paper are given in the Supplementary Material [29]. Sections A to F are in the Supplementary Material.

2. Independence models and graph theory. Graphical separation criteria as well as probabilistic models give rise to abstract conditional independence statements. Graphical modeling is essentially about relating graphical separation to probabilistic independence. We will consider both as instances of abstract *independence models*.

Consider some set $\mathcal{S}$. An *independence model*, $\mathcal{I}$, on $\mathcal{S}$ is a set of triples (A, B, C) where $A, B, C \in \mathcal{S}$, that is, $\mathcal{I} \subseteq \mathcal{S} \times \mathcal{S} \times \mathcal{S}$. Mathematically, an independence model is a ternary relation. In this paper, we will consider independence models *over* a finite set V which means that $\mathcal{S} = \mathcal{P}(V)$, the power set of V. In this case, an independence model $\mathcal{I}$ is a subset of $\mathcal{P}(V) \times \mathcal{P}(V) \times \mathcal{P}(V)$. We will call an element $s \in \mathcal{P}(V) \times \mathcal{P}(V) \times \mathcal{P}(V)$ an *independence statement* and write s as $\langle A, B \mid C\rangle$ for $A, B, C \subseteq V$. This notation emphasizes that s is thought of as a statement about A and B conditionally on C.

Graphical and probabilistic independence models have been studied in very general settings, though mostly under the assumption of symmetry of the independence model, that is,

$$\langle A, B \mid C\rangle \in \mathcal{I} \quad \Rightarrow \quad \langle B, A \mid C\rangle \in \mathcal{I};$$

see, for example, [7, 10, 26] and references therein. These works take an abstract axiomatic approach by describing and working with a number of properties that hold in, for example, models of conditional independence. In this paper, we consider independence models that do not satisfy the symmetry property as will become evident when we introduce the notion of local independence.

2.1. *Local independence.* We consider a real-valued, multivariate stochastic process

$$X_t = (X_t^1, X_t^2, \dots, X_t^n), \quad t \in [0, T]$$

defined on a probability space $(\Omega, \mathcal{F}, P)$. In this section, the process is a continuous-time process indexed by a compact time interval. The case of a discrete time index, corresponding to $X = (X_t)$ being a time series, is treated in Section C in the Supplementary Material. We will later identify the coordinate processes of X with the nodes of a graph; hence, both are indexed by $V = \{1, 2, \dots, n\}$. As illustrated in Example 2.3 below, the index set may be chosen in a more meaningful way for a specific application. In that example, $X_t^I \geq 0$ is a price process, $X_t^L \in \mathbb{N}_0$ is a counting process of events, and the remaining four processes take values in $\{0, 1\}$ indicating if an individual at a given time is a regular user of a given substance. Figure 1 shows examples of sample paths for three individuals.

To avoid technical difficulties, irrelevant for the present paper, we restrict attention to right-continuous processes with coordinates of finite and integrable variation on the interval $[0, T]$. This includes most nonexplosive multivariate counting processes as an important special case, but also other interesting processes such as piecewise-deterministic Markov processes.

To define local independence below, we need a mathematical description of how the stochastic evolution of one coordinate process depends infinitesimally on its own past and the past of the other processes. To this end, let $\mathcal{F}_t^{C,0}$ denote the σ-algebra generated by $\{X_s^\alpha : s \leq t, \alpha \in C\}$ for $C \subseteq V$. For technical reasons, we need to enlarge this σ-algebra, and we define $\mathcal{F}_t^C$ to be the completion of $\bigcap_{s>t} \mathcal{F}_s^{C,0}$ w.r.t. P. Thus $(\mathcal{F}_t^C)$ is a right-continuous and complete filtration which represents the history of the processes indexed by $C \subseteq V$ until time t. Figure 2 illustrates, in the context of Example 2.3, the filtrations $\mathcal{F}_t^V$, $\mathcal{F}_t^{\{L,M,H\}}$ and $\mathcal{F}_t^{\{T,A,M,H\}}$.

For $\beta \in V$ and $C \subseteq V$, let $\Lambda^{C,\beta}$ denote an $\mathcal{F}_t^C$-predictable process of finite and integrable variation such that

$$E(X_t^\beta \mid \mathcal{F}_t^C) - \Lambda_t^{C,\beta}$$

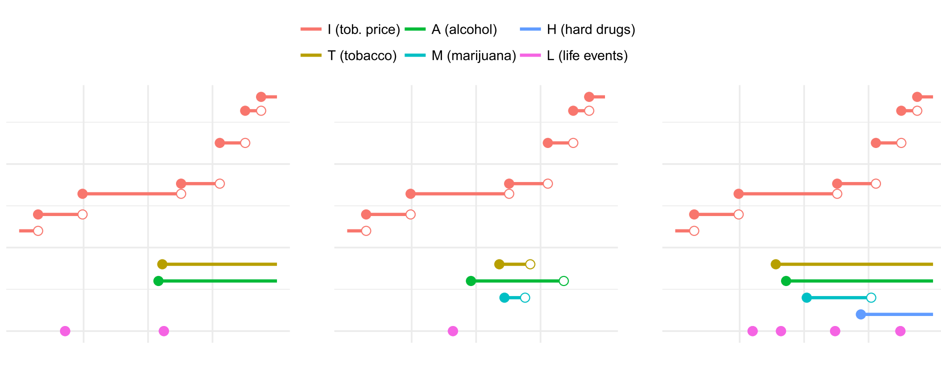

FIG. 1. *Sample paths for three individuals of the processes considered in Example 2.3. The price process (I) is a piecewise constant jump process and the life event process (L) is illustrated by the event times. The remaining four processes are illustrated by the segments of time where the individual is a regular user of the substance. The absence of a process, for example, the hard drug process (H) in the left and middle samples, means that the individual never used that substance.*

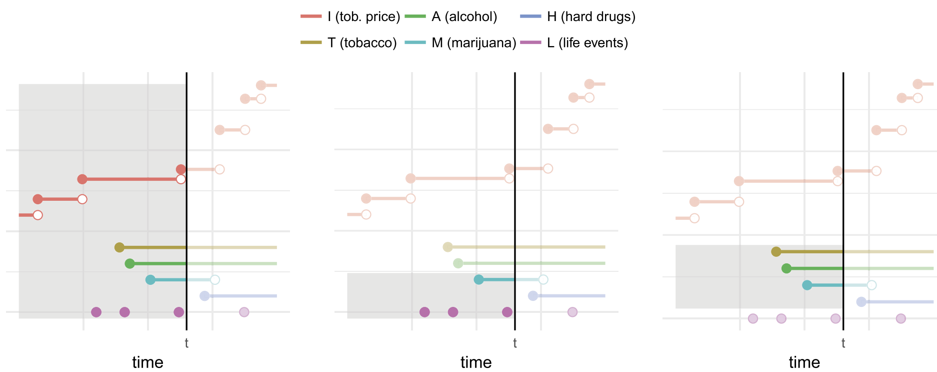

FIG. 2. *Illustration of the past at time t as captured by different filtrations for a single sample path of processes from Example 2.3. The filtration $\mathcal{F}_t^V$ (left) captures the past of all processes, while $\mathcal{F}_t^{\{L,M,H\}}$ (middle) captures the past of L, M and H only, and $\mathcal{F}_t^{\{T,A,M,H\}}$ (right) captures the past of T, A, M and H.*

is an $\mathcal{F}_t^C$ martingale. Such a process exists (see Section E for the technical details), and is usually called the compensator or the dual predictable projection of $E(X_t^\beta \mid \mathcal{F}_t^C)$. It is in general unique up to evanescence.

DEFINITION 2.1 (Local independence). Let $A, B, C \subseteq V$. We say that X^B is *locally independent* of X^A given X^C if there exists an $\mathcal{F}_t^C$-predictable version of $\Lambda^{A\cup C,\beta}$ for all $\beta \in B$. We use $A \not\rightarrow B \mid C$ to denote that X^B is locally independent of X^A given X^C.

In words, the process X^B is locally independent of X^A given X^C if, for each time point, the past up until time t of X^C gives us the same *predictable* information about $E(X_t^\beta \mid \mathcal{F}_t^{A\cup C})$ as the past of $X^{A\cup C}$ until time t. Note that when $\beta \in C$, $E(X_t^\beta \mid \mathcal{F}_t^C) = X_t^\beta$.

Local independence was introduced by Schweder [37] for composable Markov processes and extended by Aalen [1]. Local independence and graphical representations thereof were later considered by Didelez [11–13] and by Aalen et al. [4]. Didelez [12] also discussed local independence models of composable finite Markov processes under some specific types of marginalization. Commenges and Gégout-Petit [6, 21] discussed definitions of local independence in classes of semimartingales. Note that Definition 2.1 allows a process to be separated from itself by some conditioning set C, generalizing the definition used, for example, by Didelez [13].

Local independence defines the independence model

$$\mathcal{I} = \{\langle A, B \mid C\rangle \mid X^B \text{ is locally independent of } X^A \text{ given } X^C\}$$

such that the local independence statement $A \not\rightarrow B \mid C$ is equivalent to $\langle A, B \mid C\rangle \in \mathcal{I}$ in the abstract notation. We note that the local independence model is generally not symmetric. Using Definition 2.1, we introduce below an associated directed graph in which there is no directed edge from a node α to a node β if and only β is locally independent of α given $V \setminus \{\alpha\}$.

DEFINITION 2.2 (Local independence graph). For the local independence model determined by X, we define the *local independence graph* to be the directed graph, $\mathcal{D}$, with nodes V such that for $\alpha, \beta \in V$,

$$\alpha \not\rightarrow_{\mathcal{D}} \beta \quad \Leftrightarrow \quad \alpha \not\rightarrow \beta \mid V \setminus \{\alpha\}$$

where $\alpha \not\rightarrow_{\mathcal{D}} \beta$ denotes that there is no directed edge from α to β in the graph $\mathcal{D}$.

Didelez [11] gives almost the same definition of a local independence graph, however, in essence always assumes that there is a dependence of each process on its own past. See also Sections A and B.

The local independence graph induces an independence model by μ-separation as defined below. The main goal of the present paper is to provide a graphical representation of the induced independence model for a subset of coordinate processes corresponding to the case where some processes are unobserved. This is achieved by establishing a correspondence, which is preserved under marginalization, between directed mixed graphs and independence models induced via μ-separation. We emphasize that the correspondence only relates local independence to graphs when the local independence model satisfies the global Markov property with respect to a graph.

The local independence model satisfies the global Markov property with respect to the local independence graph if every μ-separation in the graph implies a local independence. This has been shown for point processes under some mild regularity conditions [13] using the slightly different notion of δ-separation. Section A discusses how δ-separation is related to μ-separation, and Section B shows how to translate the global Markov property of [13] into our framework. Moreover, general sufficient conditions for the global Markov property were given in [30] covering point processes as well as certain diffusion processes. Section C provides, in addition, a discussion of Markov properties in the context of time series.

To help develop a better understanding of local independence and its relevance for applications, we discuss an example of drug abuse progression.

EXAMPLE 2.3 (Gateway drugs). The theory of *gateway drugs* has been discussed for many years in the literature on substance abuse [24, 40]. In short, the theory posits that the use of "soft" and often licit drugs precedes (and possibly leads to) later use of "hard" drugs. Alcohol, tobacco and marijuana have all been discussed as candidate gateway drugs to "harder" drugs such as heroin.

We propose a hypothetical, dynamical model of transitions into abuse via a gateway drug, and more generally, a model of substance abuse progression. Substance abuse is known to be associated with social factors, genetics and other individual and environmental factors [43]. Substance abuse can evolve over time when an individual starts or stops using some drug. In this example, we consider substance processes Alcohol (A), Tobacco (T), Marijuana (M) and Hard drugs (H) modeled as zero-one processes, that is, stochastic processes that are piecewise constantly equal to zero (no substance use) or one (substance use). We also include L, a process describing life events, and a process I, which can be thought of as an exogenous process that influences the tobacco consumption of the individual, for example, the price of tobacco which may change due to changes in tobacco taxation. Let $V = \{A, T, M, H, L, I\}$.

We will visualize each process as a node in a graph and draw an arrow from one process to another if the first has a direct influence on the second. We will not go into a full discussion of how to formalize "influence" in terms of a continuous-time causal dynamical model as this would lead us astray; see instead [13, 27, 38]. The upshot is that for a (faithful) causal model, there is no direct influence if and only if $\alpha \nrightarrow \beta \mid V \setminus \{\alpha\}$, which identifies the "influence" graph with the local independence graph.

Several formalizations of the gateway drug question are possible. We will focus on the questions "is the use of hard drugs locally independent of use of alcohol for some conditioning set?" and "is the use of hard drugs locally independent of the use of tobacco for some conditioning set?" Using the dynamical nature of local independence, we are asking if, for example, the past alcohol usage changes the hard drug usage propensity when accounting for the past of all other processes in the model. This is one possible formalization of the gateway drug question as a negative answer would mean that there exist some gateway processes through which any influence of alcohol usage on hard drug usage is mediated. If the

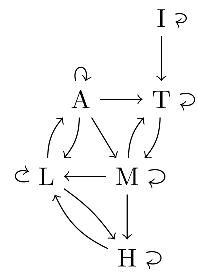

FIG. 3. *The directed graph of Example 2.3 illustrating a model where marijuana* (M) *potentially acts as a gateway drug*, *while alcohol* (A) *as well as tobacco* (T) *do not directly affect hard drug use.*

visualization in Figure 3 is indeed a local independence graph in the above sense we see that conditioning on all other processes, H is indeed locally independent of A and locally independent of T. In this hypothetical scenario, we could interpret this as marijuana in fact acting as a gateway drug to hard drugs. If the global Markov property holds, we can furthermore use μ-separation to obtain further local independences from the graph. We return to this example in Section 5.5 to illustrate how the main results of the paper can be applied. In particular, we are interested in what conclusions we can make when we do not observe all the processes but only a subset.

2.2. *Marginalization and separability.*

DEFINITION 2.4 (Marginalization). Given an independence model $\mathcal{I}$ over V, the marginal independence model over $O \subseteq V$ is defined as

$$\mathcal{I}^O = \{\langle A, B \mid C\rangle \mid \langle A, B \mid C\rangle \in \mathcal{I}; A, B, C \subseteq O\}.$$

Marginalization is defined abstractly above, though we are primarily interested in the marginalization of the independence model encoded by a local independence graph via μ-separation. The main objective is to obtain a graphical representation of such a marginalized independence model involving only the nodes O. To this end, we consider the notion of separability in an independence model.

DEFINITION 2.5 (Separability). Let $\mathcal{I}$ be an independence model over V. Let $\alpha, \beta \in V$. We say that β is *separable* from α if there exists $C \subseteq V \setminus \{\alpha\}$ such that $\langle \alpha, \beta \mid C\rangle \in \mathcal{I}$, and otherwise we say that β is *inseparable* from α. We define

$$s(\beta, \mathcal{I}) = \{\gamma \in V \mid \beta \text{ is separable from } \gamma\}.$$

We also define $u(\beta, \mathcal{I}) = V \setminus s(\beta, \mathcal{I})$.

We show in Proposition 3.6 that if $\mathcal{I}$ is the independence model induced by a directed graph via μ-separation, then $\alpha \in u(\beta, \mathcal{I})$ if and only if there is a directed edge from α to β. In this case, the graph is thus directly identifiable from separability properties of $\mathcal{I}$. That is, however, not true in general for a marginalization of $\mathcal{I}$, and this is the motivation for developing a theory of directed *mixed* graphs with μ-separation.

2.3. *Graph theory.* A graph, $\mathcal{G} = (V, E)$, is an ordered pair where V is a finite set of vertices (also called nodes) and E is a finite set of edges. Furthermore, there is a map that to each edge assigns a pair of nodes (not necessarily distinct). We say that the edge is *between* these two nodes. We consider graphs with two types of edges: directed ($\rightarrow$) and bidirected

($\leftrightarrow$). We can think of the edge set as a disjoint union, $E = E_d \dot{\cup} E_b$, where E_d is a set of ordered pairs of nodes (α, β) corresponding to directed edges, and E_b is a set of unordered pairs of nodes $\{\alpha, \beta\}$ corresponding to bidirected edges. This implies that the edge $\alpha \leftrightarrow \beta$ is identical to the edge $\beta \leftrightarrow \alpha$, but the edge $\alpha \to \beta$ is different from the edge $\beta \to \alpha$. It also implies that the graphs we consider can have multiple edges between a pair of nodes α and β, but they will always be a subset of the edges $\{\alpha \to \beta, \alpha \leftarrow \beta, \alpha \leftrightarrow \beta\}$.

DEFINITION 2.6 (DMG). A *directed mixed graph* (DMG), $\mathcal{G} = (V, E)$, is a graph with node set V and edge set E consisting of directed and bidirected edges as described above.

Throughout the paper, $\mathcal{G}$ will denote a DMG with node set V and edge set E. Occasionally, we will also use $\mathcal{D}$ and $\mathcal{M}$ to denote DMGs. We use $\mathcal{D}$ only when the DMG is also a directed graph, that is, has no bidirected edges. We use $\mathcal{M}$ to stress that some DMG is obtained as a marginalization of a DMG on a larger node set. We will use notation such as $\leftrightarrow_{\mathcal{G}}$ or $\to_{\mathcal{D}}$ to denote the specific graph that an edge belongs to.

If $\alpha \to \beta$, we say that the edge has a *tail* at α and a *head* at β. Jointly tails and heads are called (edge) *marks*. An edge $e \in E$ between nodes α and β is a *loop* if $\alpha = \beta$. We also say that the edge is *incident* with the node α and with the node β and that α and β are *adjacent*.

For $\alpha, \beta \in V$, we use the notation $\alpha \sim \beta$ to denote a generic edge of any type between α and β. We use the notation $\alpha \ast\!\!\to \beta$ to indicate an edge that has a head at β and may or may not have a head at α. Note that the presence of one edge, $\alpha \to \beta$, say, does not in general preclude the presence of other edges between these two nodes. Finally, $\alpha \ast\!\!\not\to_{\mathcal{G}} \beta$ means that there is no edge in $\mathcal{G}$ between α and β that has a head at β and $\alpha \not\to_{\mathcal{G}} \beta$ means that there is no directed edge from α to β. Note that $\alpha \not\to_{\mathcal{G}} \beta$ is a statement about the absence of an edge in the graph $\mathcal{G}$ and to avoid confusion with local independence, $\alpha \not\to \beta \mid C$, we always include the conditioning set when writing local independence statements, even if $C = \varnothing$ (see also Definition 2.2).

We say that α is a *parent* of β in the graph $\mathcal{G}$ if $\alpha \to \beta$ is present in $\mathcal{G}$ and that β is a *child* of α. We say that α is a *sibling* of β (and that β is a sibling of α) if $\alpha \leftrightarrow \beta$ is present in the graph. The motivation of the term sibling will be explained in Section 3. We use $\text{pa}(\alpha)$ to denote the set of parents of α.

A *walk* is an ordered, alternating sequence of vertices, γ_i, and edges, e_j, denoted $\omega = \langle \gamma_1, e_1, \ldots, e_n, \gamma_{n+1} \rangle$, such that each e_i is between γ_i and γ_{i+1}, along with an orientation of each directed loop along the walk (if e_i is a loop then we also know if e_i points in the direction of γ_1 or in the direction of γ_{n+1}). Without the orientation, for instance, the walks $\alpha \to \beta \to \beta \to \gamma$ and $\alpha \to \beta \leftarrow \beta \to \gamma$ would be indistinguishable. See Figure 4 for examples. We will often present the walk ω using the notation

$$\gamma_1 \overset{e_1}{\sim} \gamma_2 \overset{e_2}{\sim} \ldots \overset{e_n}{\sim} \gamma_{n+1},$$

where the loop orientation is explicit. We will omit the edge superscripts when they are not needed.

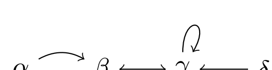

FIG. 4. *A directed mixed graph with node set $\{\alpha, \beta, \gamma, \delta\}$. Consider first the walk $\alpha \to \beta$. This is different from the walk $\beta \leftarrow \alpha$ as walks are ordered. Consider instead the two walks $\beta \leftrightarrow \gamma \leftarrow \gamma \leftarrow \delta$ and $\beta \leftrightarrow \gamma \to \gamma \leftarrow \delta$. These two walks have the same (ordered) sets of nodes and edges but are not equal as the loop at γ has different orientations between the two walks. Furthermore, one can note that for the first of the two walks, γ is a collider in the first instance, but not in the second. The walks $\alpha \to \beta \to \alpha$ and $\alpha \to \beta \leftarrow \alpha$ are both cycles, and the second is an example of the fact that the same edge can occur twice in a cycle.*

We say that the walk ω *contains* nodes γ_i and edges e_j. The *length* of the walk is n, the number of edges that it contains. We define a *trivial walk* to be a walk with no edges and, therefore, only a single node. Equivalently, a trivial walk can be defined as a walk of length zero. A *subwalk* of ω is either itself a walk of the form $\langle \gamma_k, e_k, \dots, e_{m-1}, \gamma_m \rangle$ where $1 \leq k < m \leq n+1$ or a trivial walk $\langle \gamma_k \rangle$, $1 \leq k \leq n+1$. A (nontrivial) walk is uniquely identified by its edges, and the ordering and orientation of these edges, hence the vertices can be omitted when describing the walk. At times, we will omit the edges to simplify notation, however, we will always have a specific, uniquely identified walk in mind even when the edges and/or their orientation is omitted. The first and last nodes of a walk are called *endpoint nodes* (these could be equal) or just endpoints, and we say that a walk is *between* its endpoints, or alternatively *from* its first node *to* its last node. We call the walk $\omega^{-1} = \langle \gamma_{n+1}, e_n, \dots, e_1, \gamma_1 \rangle$ the *inverse* walk of ω. Note that the orientation of directed loops is also reversed in the inverse walk such that they point toward γ_1 in the inverse if and only if they point toward γ_1 in the original walk. A *path* is a walk on which no node is repeated.

Consider a walk ω and a subwalk thereof, $\langle \alpha, e_1, \gamma, e_2, \beta \rangle$, where $\alpha, \beta, \gamma \in V$ and $e_1, e_2 \in E$. If e_1 and e_2 both have heads at γ, then γ is a *collider* on ω. If this is not the case, then γ is a noncollider. Note that an endpoint of a walk is neither a collider, nor a noncollider. We stress that the property of being a collider/noncollider is relative to a walk (see also Figure 4).

Let $\omega_1 = \langle \alpha, e_1^1, \gamma_1^1, \dots, \gamma_{n-1}^1, e_n^1, \beta \rangle$ and $\omega_2 = \langle \alpha, e_1^2, \gamma_1^2, \dots, \gamma_{m-1}^2, e_m^2, \beta \rangle$ be two (nontrivial) walks. We say that they are *endpoint-identical* if e_1^1 and e_1^2 have the same mark at α and e_n^1 and e_m^2 have the same mark at β. Note that this may depend on the orientation of directed edges in the two walks. Assume that some edge e is between α and β. We say that the (nontrivial) walk ω_1 is endpoint-identical to e if it is endpoint-identical to the walk $\langle \alpha, e, \beta \rangle$. If $\alpha = \beta$ and e is directed, this should hold for just one of the possible orientations of e.

Let ω_1 be a walk between α and γ, and ω_2 a walk between γ and β. The *composition* of ω_1 with ω_2 is the walk that starts at α, traverses every node and edge of ω_1, and afterwards every node and edge of ω_2, ending in β. We say that we *compose* ω_1 with ω_2.

A *directed path* from α to β is a path between α and β consisting of edges of type $\to$ only (possibly of length zero) such that they all point in the direction of β. A *cycle* is either a loop, or a (nontrivial) path from α to β composed with $\beta \sim \alpha$. This means that in a cycle of length 2, an edge can be repeated. A *directed cycle* is either a loop, $\alpha \to \alpha$, or a (nontrivial) directed path from α to β composed with $\beta \to \alpha$. For $\alpha \in V$, we let $\mathrm{An}(\alpha)$ denote the set of *ancestors*, that is,

$$\mathrm{An}(\alpha) = \{\gamma \in V \mid \text{there is a directed path from } \gamma \text{ to } \alpha\}.$$

This is generalized to nonsingleton sets $C \subseteq V$,

$$\mathrm{An}(C) = \bigcup_{\alpha \in C} \mathrm{An}(\alpha).$$

We stress that $C \subseteq \mathrm{An}(C)$ as we allow for trivial directed paths in the definition of an ancestor. We use the notation $\mathrm{An}_{\mathcal{G}}(C)$ if we wish to emphasize in which graph the ancestry is read, but omit the subscript when no ambiguity arises.

Let $\mathcal{G} = (V, E)$ be a graph, and let $O \subseteq V$. Define the *subgraph* induced by O to be the graph $\mathcal{G}_O = (O, E_O)$ where $E_O \subseteq E$ is the set of edges that are between nodes in O. If $\mathcal{G}_1 = (V, E_1)$ and $\mathcal{G}_2 = (V, E_2)$, we will write $\mathcal{G}_1 \subseteq \mathcal{G}_2$ to denote $E_1 \subseteq E_2$ and say that $\mathcal{G}_2$ is a *supergraph* of $\mathcal{G}_1$.

A *directed graph* (DG), $\mathcal{D} = (V, E)$, is a graph with only directed edges. Note that this also allows directed loops. Within a class of graphs, we define the *complete* graph to be the graph which is the supergraph of all graphs in the class when such a graph exists. For the class of DGs on node set V, the complete graph is the graph with edge set $E = \{(\alpha, \beta) \mid \alpha, \beta \in V\}$.

A *directed acyclic graph* (DAG) is a DG with no loops and no directed cycles. An *acyclic directed mixed graph* (ADMG) is a DMG with no loops and no directed cycles.

3. Directed mixed graphs and separation. In this section, we introduce μ-separation for DMGs which are then shown to be closed under marginalization. In particular, we obtain a DMG representing the independence model arising from a local independence graph via marginalization.

The class of DMGs contains as a subclass the ADMGs that have no directed cycles [19, 32]. ADMGs have been used to represent marginalized DAG models, analogously to how we will use DMGs to represent marginalized DGs. ADMGs come with the m-separation criterion which can be extended to DMGs, but this criterion differs in important ways from the μ-separation criterion introduced below. These differences also mean that our main result on Markov equivalence does not apply to, for example, DMGs with m-separation, and thus our theory of Markov equivalence hinges on the fact that we are considering DMGs using the asymmetric notion of μ-separation.

3.1. *μ-separation.* We define μ-separation as a generalization of δ-separation introduced by Didelez [11], analogously to how m-separation is a generalization of d-separation; see, for example, [33]. In Section A, we make the connection to Didelez's δ-separation exact and elaborate further on this in Section B.

DEFINITION 3.1 (μ-connecting walk). A nontrivial walk

$$\langle \alpha, e_1, \gamma_1, \dots, \gamma_{n-1}, e_n, \beta \rangle$$

in $\mathcal{G}$ is said to be μ-connecting (or simply *open*) from α to β given C if $\alpha \notin C$, every collider is in $\mathrm{An}(C)$, no noncollider is in C, and e_n has a head at β.

When a walk is not μ-connecting given C, we say that it is *closed* or *blocked* by C. One should note that if ω is a μ-connecting walk from α to β given C, the inverse walk, ω^{-1}, is not in general μ-connecting from β to α given C. The requirement that a μ-connecting walk be nontrivial, that is, of strictly positive length, leads to the possibility of a node being separated from itself by some set C when applying the following graph separation criterion to the class of DMGs.

DEFINITION 3.2 (μ-separation). Let $A, B, C \subseteq V$. We say that B is μ-separated from A given C if there is no μ-connecting walk from any $\alpha \in A$ to any $\beta \in B$ given C and write $A \perp_\mu B \mid C$, or write $A \perp_\mu B \mid C\ [\mathcal{G}]$ if we want to stress to what graph the separation statement applies.

The above notion of separation is given in terms of walks of which there are infinitely many in any DMG with a nonempty edge set. However, we will see that it is sufficient to consider a finite subset of walks from A to B (Proposition 3.5).

Given a DMG, $\mathcal{G} = (V, E)$, we define an independence model over V using μ-separation,

$$\mathcal{I}(\mathcal{G}) = \{\langle A, B \mid C \rangle \mid (A \perp_\mu B \mid C)\}.$$

Definition 3.1 implies $A \perp_\mu B \mid C$ whenever $A \subseteq C$ and, therefore, $\mathcal{I}(\mathcal{G}) \neq \varnothing$.

Below we state two propositions that essentially both give equivalent ways of defining μ-separation. The propositions are useful when proving results on μ-separation models.

PROPOSITION 3.3. *Let $\alpha, \beta \in V$, $C \subseteq V$. If there is a μ-connecting walk from α to β given C, then there is a μ-connecting walk from α to β that furthermore satisfies that every collider is in C.*

DEFINITION 3.4. A *route* from α to β is a walk from α to β such that no node different from β occurs more than once, and β occurs at most twice.

A route is always a path, a cycle or a composition of a path and a cycle that share no edge and only share the vertex β.

PROPOSITION 3.5. *Let $\alpha, \beta \in V$, $C \subseteq V$. If ω is a μ-connecting walk from α to β given C, then there is a μ-connecting route from α to β given C consisting of edges in ω.*

If there is a μ-connecting walk from A to B given C, it does not in general follow that we can also find a μ-connecting path or cycle from A to B given C. As an example of this, consider the following DMG on nodes $\{\alpha, \beta, \gamma\}$: $\alpha \leftarrow \beta \leftarrow \gamma$. There is a μ-connecting walk from α to β given $\varnothing$, and a μ-connecting route, but no μ-connecting path from α to β given $\varnothing$.

3.2. *Marginalization of DMGs.* Given a DG or a DMG, $\mathcal{G}$, we are interested in finding a graph that represents the marginal independence model over a node set $O \subseteq V$, that is, finding a graph $\mathcal{M}$ such that

$$\mathcal{I}(\mathcal{M}) = \big(\mathcal{I}(\mathcal{G})\big)^{O}. \tag{3.1}$$

It is well known that the class of DAGs with d-separation is not closed under marginalization, that is, for a DAG, $\mathcal{D} = (V, E)$, and $O \subsetneq V$, it is not in general possible to find a DAG with node set O that encodes the same independence model among the variables in O as did the original graph. Richardson and Spirtes [33] gave a concrete counterexample and in Example 3.7 we give a similar example to make the analogous point: DGs read with μ-separation are not closed under marginalization. In this example, we use the following proposition which gives a simple characterization of separability in DGs.

PROPOSITION 3.6. *Consider a DG, $\mathcal{D} = (V, E)$, and let $\alpha, \beta \in V$. Then β is μ-separable (see Definition 2.5) from α in $\mathcal{D}$ if and only if $\alpha \not\to_{\mathcal{D}} \beta$.*

EXAMPLE 3.7. Consider the directed graph, $\mathcal{G}$, in Figure 5. We wish to show that it is not possible to encode the μ-separations among nodes in $O = \{\alpha, \beta, \gamma, \delta\}$ using a DG on these nodes only. To obtain a contradiction, assume $\mathcal{D} = (O, E)$ is a DG such that

$$A \perp_{\mu} B \mid C\ [\mathcal{D}] \Leftrightarrow A \perp_{\mu} B \mid C\ [\mathcal{G}] \tag{3.2}$$

for $A, B, C \subseteq O$. There is no $C \subseteq O \setminus \{\alpha\}$ such that $\alpha \perp_{\mu} \beta \mid C\ [\mathcal{G}]$ and no $C \subseteq O \setminus \{\beta\}$ such that $\beta \perp_{\mu} \gamma \mid C\ [\mathcal{G}]$. If $\mathcal{D}$ has the property (3.2), then it follows from Proposition 3.6 that $\alpha \to_{\mathcal{D}} \beta$ and $\beta \to_{\mathcal{D}} \gamma$. However, then γ is not μ-separated from α given $\varnothing$ in $\mathcal{D}$. This shows that there exists no DG, $\mathcal{D}$, that satisfies (3.2).

We note that marginalization of a probability model does not only impose conditional independence constraints on the observed variables but also so-called equality and inequality constraints; see, for example, [18] and references therein. In this paper, we will only be

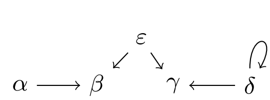

FIG. 5. *The directed graph of Example 3.7 which exemplifies that DGs are not closed under marginalization.*

concerned with the graphical representation of local independence constraints, and not with representing analogous equality or inequality constraints.

In the remainder of this section, we first introduce the *latent projection* of a graph (see also [41] and [34]), and then show that it provides a marginalized DMG in the sense of (3.1). At the end of the section, we give an algorithm for computing the latent projection of a DMG. This algorithm is an adapted version of one described by Sadeghi [36] for a different class of graphs. Koster [25] described a similar algorithm for ADMGs.

DEFINITION 3.8 (Latent projection). Let $\mathcal{G} = (V, E)$ be a DMG, $V = M \dot{\cup} O$. We define the *latent projection of* $\mathcal{G}$ *on* O to be the DMG (O, D) such that $\alpha \sim \beta \in D$ if and only if there exists an endpoint-identical (and nontrivial) walk between α and β in $\mathcal{G}$ with no colliders and such that every nonendpoint node is in M. Let $m(\mathcal{G}, O)$ denote the latent projection of $\mathcal{G}$ on O.

The definition of latent projection motivates the graphical term *sibling* for DMGs, as one way to obtain an edge $\alpha \leftrightarrow \beta$ is through a latent projection of a larger graph in which α and β share a parent.

To characterize the class of graphs obtainable from a DG via a latent projection, we introduce the *canonical DG of the DMG* $\mathcal{G}$, $\mathcal{C}(\mathcal{G})$, as follows: for each (unordered) pair of nodes $\{\alpha, \beta\} \subseteq V$ such that $\alpha \leftrightarrow_{\mathcal{G}} \beta$, add a distinct auxiliary node, $m_{\{\alpha,\beta\}}$, add edges $m_{\{\alpha,\beta\}} \to \alpha$, $m_{\{\alpha,\beta\}} \to \beta$ to E and then remove all bidirected edges from E. If $\mathcal{D}$ is a DG, then $\mathcal{M} = m(\mathcal{D}, O)$ will satisfy

$$\alpha \leftrightarrow_{\mathcal{M}} \beta \quad \Rightarrow \quad \alpha \leftrightarrow_{\mathcal{M}} \alpha \quad \text{for all } \alpha, \beta \in O \tag{3.3}$$

for all subsets of vertices O. Conversely, if $\mathcal{G} = (V, E)$ is a DMG that satisfies (3.3), then $\mathcal{G}$ is the latent projection of its canonical DG; $m(\mathcal{C}(\mathcal{G}), V) = \mathcal{G}$. The class of DMGs that satisfy (3.3) is closed under marginalization (Proposition 3.9) and has certain regularity properties (see, e.g., Proposition 3.10). These result provide the means for graphically representing marginals of local independence graphs. However, the theory that leads to our main results on Markov equivalence does not require the property (3.3) and, therefore, we develop it for general DMGs.

PROPOSITION 3.9. *Let* $O \subseteq V$. *The graph* $\mathcal{M} = m(\mathcal{G}, O)$ *is a DMG. If* $\mathcal{G}$ *satisfies* (3.3), *then* $\mathcal{M}$ *does as well.*

PROPOSITION 3.10. *Assume that* $\mathcal{G}$ *satisfies* (3.3) *and let* $\alpha \in V$. *Then* α *has no loops if and only if* $\alpha \perp_\mu \alpha \mid V \setminus \{\alpha\}$.

We also observe directly from the definition that the latent projection operation preserves ancestry and nonancestry in the following sense.

PROPOSITION 3.11. *Let* $O \subseteq V$, $\mathcal{M} = m(\mathcal{G}, O)$ *and* $\alpha, \beta \in O$. *Then* $\alpha \in \mathrm{An}_{\mathcal{G}}(\beta)$ *if and only if* $\alpha \in \mathrm{An}_{\mathcal{M}}(\beta)$.

The main result of this section is the following theorem, which states that the marginalization defined by the latent projection operation preserves the marginal independence model encoded by a DMG.

THEOREM 3.12. *Let* $O \subseteq V$, $\mathcal{M} = m(\mathcal{G}, O)$. *Assume* $A, B, C \subseteq O$. *Then*

$$A \perp_\mu B \mid C \; [\mathcal{G}] \quad \Leftrightarrow \quad A \perp_\mu B \mid C \; [\mathcal{M}].$$

input : a DMG, $\mathcal{G} = (V, E)$ a subset $M \subseteq V$ over which to marginalize
output : a graph $\mathcal{M} = (O, \bar{E})$, $O = V \setminus M$
Initialize $E_0 = E$, $\mathcal{M}_0 = (V, E_0)$, $k = 0$;
while $\Omega_M(\mathcal{M}_k) \neq \emptyset$ **do**
  Choose $\theta = {}_\theta(\alpha, m, \beta) \in \Omega_M(\mathcal{M}_k)$;
  Set e_{k+1} to be the edge $\alpha \sim \beta$ which is endpoint-identical to θ;
  Set $E_{k+1} = E_k \cup \{e_{k+1}\}$;
  Set $\mathcal{M}_{k+1} = (V, E_{k+1})$;
  Update $k = k + 1$
end
return $(\mathcal{M}_k)_O$

Algorithm 1: Computing the latent projection of a DMG

3.3. *A marginalization algorithm.* We describe an algorithm to compute the latent projection of a graph on some subset of nodes. For this purpose, we define a *triroute*, θ, to be a walk of length 2, $\langle \alpha, e_1, \gamma, e_2, \beta \rangle$, such that $\gamma \neq \alpha, \beta$. We suppress e_1 and e_2 from the notation and use ${}_\theta(\alpha, \gamma, \beta)$ to denote the triroute. We say that a triroute is *colliding* if γ is a collider on θ, and otherwise we say that it is *noncolliding*. This is analogous to the concept of a *tripath* (see, e.g., [26]), but allows for $\alpha = \beta$.

Define $\Omega_M(\mathcal{G})$ to be the set of noncolliding triroutes ${}_\theta(\alpha, m, \beta)$ such that $m \in M$ and such that an endpoint-identical edge $\alpha \sim \beta$ is not present in $\mathcal{G}$.

PROPOSITION 3.13. *Algorithm 1 outputs the latent projection of a DMG.*

4. Properties of DMGs.

DEFINITION 4.1 (Markov equivalence). Let $\mathcal{G}_1 = (V, E_1)$ and $\mathcal{G}_2 = (V, E_2)$ be DMGs. We say that $\mathcal{G}_1$ and $\mathcal{G}_2$ are *Markov equivalent* if $\mathcal{I}(\mathcal{G}_1) = \mathcal{I}(\mathcal{G}_2)$. This defines an equivalence relation and we let $[\mathcal{G}_1]$ denote the (Markov) equivalence class of $\mathcal{G}_1$.

EXAMPLE 4.2 (Markov equivalence in DGs). Let $\mathcal{D} = (V, E)$ be a DG. There is a directed edge from α to β if and only if β cannot be separated from α by any set $C \subseteq V \setminus \{\alpha\}$ (Proposition 3.6). This implies that two DGs are Markov equivalent if and only if they are equal. Thus, in the restricted class of DGs, every Markov equivalence class is a singleton and in this sense *identifiable* from its induced independence model. However, when considering Markov equivalence in the more general class of DMGs not every equivalence class of a DG is a singleton as the DG might be Markov equivalent to a DMG. As an example of this, consider the complete DG on a node set V which is Markov equivalent to the complete DMG on V.

DEFINITION 4.3 (Maximality of a DMG). We say that $\mathcal{G}$ is *maximal* if it is complete, or if any added edge changes the induced independence model $\mathcal{I}(\mathcal{G})$.

4.1. *Inducing paths.* Separability of nodes can be studied using the concept of an *inducing path* which has also been used in other classes of graphs [33, 41]. In the context of DMGs and μ-separation, it is natural to define several types of inducing paths due to the asymmetry of μ-separation and the possibility of directed cycles in DMGs.

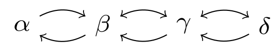

FIG. 6. *Examples of inducing paths in a DMG: the path $\beta \to \alpha$ is a unidirected inducing path from β to α, and also a directed inducing path. The path $\beta \leftrightarrow \gamma$ is a bidirected inducing path. The path $\beta \leftrightarrow \gamma \leftrightarrow \delta$ is a bidirected inducing path from β to δ (and by definition its inverse is a bidirected inducing path from δ to β). The path $\delta \to \gamma \leftrightarrow \beta$ is both a unidirected and a directed inducing path from δ to β, whereas the path $\alpha \to \beta \leftrightarrow \gamma$ is a unidirected inducing path from α to γ, but not a directed inducing path.*

DEFINITION 4.4 (Inducing path). An *inducing path from α to β* is a nontrivial path or cycle, $\pi = \langle\alpha, \dots, \beta\rangle$, which has a head at β and such that there are no noncolliders on π and every node is an ancestor of α or β. The inducing path π is *bidirected* if every edge on π is bidirected. If π is not bidirected, it has one of the forms $\alpha \to \beta$ or

$$\alpha \to \gamma_1 \leftrightarrow \cdots \leftrightarrow \gamma_n \leftrightarrow \beta.$$

and we say that it is *unidirected*. If, furthermore, $\gamma_i \in \mathrm{An}(\beta)$ for all $i = 1, \dots, n$ (or it is on the form $\alpha \to \beta$) then we say that it is *directed*.

Note that an inducing path is by definition either a path or a cycle. An inducing path is either bidirected or unidirected. Some unidirected inducing paths are also directed; see Figure 6 for examples. Propositions 4.7 and 4.8 show how bidirected and directed inducing paths in a certain sense correspond to bidirected and directed edges, respectively.

PROPOSITION 4.5. *Let ν be an inducing path from α to β. The following holds for any $C \subseteq V \setminus \{\alpha\}$. If $\alpha \neq \beta$, then there exists a μ-connecting path from α to β given C. If $\alpha = \beta$, then there exists a μ-connecting cycle from α to β given C. We call such a path or cycle a* ν-induced open path or cycle, *respectively, or simply a ν-induced open walk to cover both the case $\alpha = \beta$ and the case $\alpha \neq \beta$. If the inducing path is bidirected or directed, then the ν-induced open walk is endpoint-identical to the inducing path.*

The following corollary is a direct consequence of Proposition 4.5, showing that β is inseparable from α if there is an inducing path from α to β irrespectively of whether the nodes are adjacent.

COROLLARY 4.6. *Let $\alpha, \beta \in V$. If there exists an inducing path from α to β in $\mathcal{G}$, then β is not μ-separated from α given C for any $C \subseteq V \setminus \{\alpha\}$, that is, $\alpha \in u(\beta, \mathcal{I}(\mathcal{G}))$.*

The following two propositions show that for two of the three types of inducing paths there is a Markov equivalent supergraph in which the nodes are adjacent. This illustrates how one can easily find Markov equivalent DMGs that do not have the same adjacencies. Example 4.12 shows that for a unidirected inducing path it may not be possible to add an edge without changing the independence model.

PROPOSITION 4.7. *If there exists a bidirected inducing path from α to β in $\mathcal{G}$, then adding $\alpha \leftrightarrow \beta$ in $\mathcal{G}$ does not change the independence model.*

PROPOSITION 4.8. *If there exists a directed inducing path from α to β in $\mathcal{G}$, then adding $\alpha \to \beta$ in $\mathcal{G}$ does not change the independence model.*

We say that nodes α and β are *collider-connected* if there exists a nontrivial walk between α and β such that every nonendpoint node is a collider on the walk. We say that α is *directedly collider-connected* to β if α and β are collider-connected by a walk with a head at β.

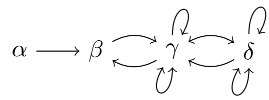

FIG. 7. *A maximal DMG in which δ is inseparable from β, though no edge is between the two. See Example 4.12. We will in general omit the bidirected loops from the visual presentations of DMGs; see also the discussion in Section 5.4.*

DEFINITION 4.9. Let $\alpha, \beta \in V$. We define the set

$$D(\alpha,\beta) = \{\gamma \in \mathrm{An}(\alpha,\beta) \mid \gamma \text{ is directedly collider-connected to } \beta\} \setminus \{\alpha\}.$$

Note that if $\alpha \not\to_{\mathcal{G}} \beta$, then $\mathrm{pa}(\beta) \subseteq D(\alpha,\beta)$, and if the graph is furthermore a directed graph then $\mathrm{pa}(\beta) = D(\alpha,\beta)$.

PROPOSITION 4.10. *If there is no inducing path from α to β in $\mathcal{G}$, then β is separated from α by $D(\alpha,\beta)$.*

EXAMPLE 4.11 (Inducing paths). Consider the DMG on nodes $\{\alpha,\gamma\}$ and with a single edge $\gamma \to \alpha$. In this case, there is no inducing path from α to α and α is μ-separated from α by $D(\alpha,\alpha) = \{\gamma\}$. Now add the edge $\alpha \leftrightarrow \gamma$. In this new DMG, there is an inducing path from α to α and therefore α is inseparable from itself.

EXAMPLE 4.12 (Nonadjacency of inseparable nodes in a maximal DMG). Consider the DMG in Figure 7. One can show that this DMG is maximal (Definition 4.3). There is an inducing path from β to δ making δ inseparable from β, yet no arrow can be added between β and δ without changing the independence model. This example illustrates that maximal DMGs do not have the property that inseparable nodes are adjacent. This is contrary to MAGs which form a subclass of ancestral graphs and have this exact property [33].

5. Markov equivalence of DMGs. The main result of this section is that each Markov equivalence class of DMGs has a greatest element, that is, an element which is a supergraph of all other elements. This fact is helpful for understanding and graphically representing such equivalence classes, and potentially also for constructing learning algorithms. We will prove this result by arguing that the independence model of a DMG, $\mathcal{G} = (V,E)$, defines for each node $\alpha \in V$ a set of *potential parents* and a set of *potential siblings*. We then construct the greatest element of $[\mathcal{G}]$ by simply using these sets, and argue that this is in fact a Markov equivalent supergraph. As we only use the independence model to define the sets of potential parents and siblings, the supergraph is identical for all members of $[\mathcal{G}]$, and thus a greatest element. Within the equivalence class, the greatest element is also the only maximal element, and we will refer to it as the maximal element of the equivalence class.

5.1. *Potential siblings.*

DEFINITION 5.1. Let $\mathcal{I}$ be an independence model over V and let $\alpha,\beta \in V$. We say that α and β are *potential siblings* in $\mathcal{I}$ if (s1)–(s3) hold:

(s1) $\beta \in u(\alpha,\mathcal{I})$ and $\alpha \in u(\beta,\mathcal{I})$,
(s2) for all $\gamma \in V$, $C \subseteq V$ such that $\beta \in C$,

$$\langle \gamma,\alpha \mid C\rangle \in \mathcal{I} \quad\Rightarrow\quad \langle \gamma,\beta \mid C\rangle \in \mathcal{I},$$

(s3) for all $\gamma \in V$, $C \subseteq V$ such that $\alpha \in C$,

$$\langle \gamma, \beta \mid C \rangle \in \mathcal{I} \quad \Rightarrow \quad \langle \gamma, \alpha \mid C \rangle \in \mathcal{I}.$$

Potential siblings are defined abstractly above in terms of the independence model only. The following proposition gives a useful characterization for graphical independence models by simply contraposing (s2) and (s3).

PROPOSITION 5.2. *Let $\mathcal{I}(\mathcal{G})$ be the independence model induced by $\mathcal{G}$. Then $\alpha, \beta \in V$ are potential siblings if and only if* (gs1)–(gs3) *hold*:

(gs1) $\beta \in u(\alpha, \mathcal{I}(\mathcal{G}))$ *and* $\alpha \in u(\beta, \mathcal{I}(\mathcal{G}))$,

(gs2) *for all $\gamma \in V$, $C \subseteq V$ such that $\beta \in C$: if there exists a μ-connecting walk from γ to β given C, then there exists a μ-connecting walk from γ to α given C*,

(gs3) *for all $\gamma \in V$, $C \subseteq V$ such that $\alpha \in C$: if there exists a μ-connecting walk from γ to α given C, then there exists a μ-connecting walk from γ to β given C.*

PROPOSITION 5.3. *Assume that $\alpha \leftrightarrow \beta$ is in $\mathcal{G}$. Then α and β are potential siblings in $\mathcal{I}(\mathcal{G})$.*

LEMMA 5.4. *Assume that α and β are potential siblings in $\mathcal{I}(\mathcal{G})$. Let $\mathcal{G}^+$ denote the DMG obtained from $\mathcal{G}$ by adding $\alpha \leftrightarrow \beta$. Then $\mathcal{I}(\mathcal{G}) = \mathcal{I}(\mathcal{G}^+)$.*

The above shows that if α and β are potential siblings in $\mathcal{I}(\mathcal{G})$ then there exists a supergraph, $\mathcal{G}^+$, which is Markov equivalent with $\mathcal{G}$, such that α and β are siblings in $\mathcal{G}^+$. This motivates the term *potential* siblings.

5.2. *Potential parents*. In this section, we will argue that also a set of *potential parents* are determined by the independence model. This case is slightly more involved for two reasons. First, the relation is asymmetric, as for each potential parent edge there is a parent node and a child node. Second, adding directed edges potentially changes the ancestry of the graph.

DEFINITION 5.5. Let $\mathcal{I}$ be an independence model over V and let $\alpha, \beta \in V$. We say that α is a *potential parent* of β in $\mathcal{I}$ if (p1)–(p4) hold:

(p1) $\alpha \in u(\beta, \mathcal{I})$,

(p2) for all $\gamma \in V$, $C \subseteq V$ such that $\alpha \notin C$,

$$\langle \gamma, \beta \mid C \rangle \in \mathcal{I} \quad \Rightarrow \quad \langle \gamma, \alpha \mid C \rangle \in \mathcal{I},$$

(p3) for all $\gamma, \delta \in V$, $C \subseteq V$ such that $\alpha \notin C$, $\beta \in C$,

$$\langle \gamma, \delta \mid C \rangle \in \mathcal{I} \quad \Rightarrow \quad \langle \gamma, \beta \mid C \rangle \in \mathcal{I} \vee \langle \alpha, \delta \mid C \rangle \in \mathcal{I},$$

(p4) for all $\gamma \in V$, $C \subseteq V$, such that $\alpha \notin C$,

$$\langle \beta, \gamma \mid C \rangle \in \mathcal{I} \quad \Rightarrow \quad \big\langle \beta, \gamma \mid C \cup \{\alpha\} \big\rangle \in \mathcal{I}.$$

PROPOSITION 5.6. *Let $\mathcal{I}(\mathcal{G})$ be the independence model induced by $\mathcal{G}$. Then $\alpha \in V$ is a potential parent of $\beta \in V$ if and only if* (*gp*1)–(*gp*4) *hold*:

(gp1) $\alpha \in u(\beta, \mathcal{I}(\mathcal{G}))$,

(gp2) *for all $\gamma \in V$, $C \subseteq V$ such that $\alpha \notin C$: if there exists a μ-connecting walk from γ to α given C, then there exists a μ-connecting walk from γ to β given C*, (gp3) *for all* $\gamma, \delta \in V$, $C \subseteq V$ *such that* $\alpha \notin C$, $\beta \in C$: *if there exists a* μ*-connecting walk from* γ *to* β *given* C *and a* μ*-connecting walk from* α *to* δ *given* C, *then there exists a* μ*-connecting walk from* γ *to* δ *given* C,

(gp4) *for all* $\gamma \in V$, $C \subseteq V$, *such that* $\alpha \notin C$: *if there exists a* μ*-connecting walk from* β *to* γ *given* $C \cup \{\alpha\}$, *then there exists a* μ*-connecting walk from* β *to* γ *given* C.

PROPOSITION 5.7. *Assume that* $\alpha \to \beta$ *is in* $\mathcal{G}$. *Then* α *is a potential parent of* β *in* $\mathcal{I}(\mathcal{G})$.

LEMMA 5.8. *Assume that* α *is a potential parent of* β *in* $\mathcal{I}(\mathcal{G})$. *Let* $\mathcal{G}^+$ *denote the DMG obtained from* $\mathcal{G}$ *by adding* $\alpha \to \beta$. *Then* $\mathcal{I}(\mathcal{G}) = \mathcal{I}(\mathcal{G}^+)$.

5.3. *A Markov equivalent supergraph.* Let $\mathcal{G} = (V, E)$ be a DMG. Define $\mathcal{N}(\mathcal{I}(\mathcal{G})) = (V, E^m)$ to be the DMG with edge set $E^m = E^d \cup E^b$ where E^d is a set of directed edges and E^b a set of bidirected edges such that the directed edge from α to β is in E^d if and only if α is a potential parent of β in $\mathcal{I}(\mathcal{G})$ and the bidirected edge between α and β is in E^b if and only if α and β are potential siblings in $\mathcal{I}(\mathcal{G})$.

THEOREM 5.9. *Let* $\mathcal{N} = \mathcal{N}(\mathcal{I}(\mathcal{G}))$. *Then* $\mathcal{N} \in [\mathcal{G}]$ *and* $\mathcal{N}$ *is a supergraph of all elements of* $[\mathcal{G}]$. *Furthermore, if we have a finite sequence of DMGs* $\mathcal{G}_0, \mathcal{G}_1, \dots, \mathcal{G}_m$, $\mathcal{G}_i = (V, E_i)$, *such that* $\mathcal{G}_0 = \mathcal{G}$, $\mathcal{G}_m = \mathcal{N}$, *and* $E_i \subseteq E_{i+1}$ *for all* $i = 0, \dots, m-1$, *then* $\mathcal{G}_i$ *is Markov equivalent with* $\mathcal{N}$ *for all* $i = 0, \dots, m-1$.

The graph $\mathcal{N}$ in the above theorem is a supergraph of every Markov equivalent DMG and, therefore, maximal. On the other hand, every maximal DMG is a representative of its equivalence class, and also a supergraph of all Markov equivalent DMGs. This means that we can use the class of maximal DMGs to obtain a unique representative for each DMG equivalence class.

Lemmas 5.4 and 5.8 show that conditions (gs1)–(gs3) and (gp1)–(gp4) are sufficient to Markov equivalently add a bidirected or a directed edge, respectively. The conditions are also necessary in the sense that for each condition one can find example graphs where only a single condition is violated and where the larger graph is not Markov equivalent to the smaller graph.

We can note that α is a potential parent and a potential sibling of α if and only if $\alpha \in u(\alpha, \mathcal{I}(\mathcal{G}))$. This means that in $\mathcal{N}(\mathcal{I}(\mathcal{G}))$ for each node either both a directed and a bidirected loop is present or no loop at all.

5.4. *Directed mixed equivalence graphs.* Theorem 5.9 suggests that one can represent an equivalence class of DMGs by displaying the maximal element and then simply indicate which edges are not present for all members of the equivalence class.

DEFINITION 5.10 (DMEG). Let $\mathcal{N} = (V, F)$ be a maximal DMG. Define $\bar{F} \subseteq F$ such that for $e \in F$ we let $e \in \bar{F}$ if and only if there exists a DMG $\mathcal{G} = (V, \tilde{F})$ such that $\mathcal{G} \in [\mathcal{N}]$ and $e \notin \tilde{F}$. We call $\mathcal{N}' = (V, F, \bar{F})$ a *directed mixed equivalence graph* (DMEG). When visualizing $\mathcal{N}'$, we draw $\mathcal{N}$, but use dashed edges for the set $\bar{F}$; see Figure 8.

Let $\mathcal{N}' = (V, F, \bar{F})$ be a DMEG. The DMG (V, F) is in the equivalence class represented by $\mathcal{N}'$. However, one cannot necessarily remove any subset of $\bar{F}$ and obtain a member of the Markov equivalence class (see Figure 8). Moreover, an equivalence class does not in general contain a least element, that is, an element which is a subgraph of all Markov equivalent graphs.

We will throughout this section let $\mathcal{N} = (V, F)$ be a maximal DMG. For $e \in F$, we will use $\mathcal{N} - e$ to denote the graph $(V, F \setminus \{e\})$. Assume that we have a maximal DMG from which we wish to derive the DMEG. Consider some edge $e \in F$. If $\mathcal{N} - e \in [\mathcal{N}]$, then $e \in \bar{F}$ as there exists a Markov equivalent subgraph of $\mathcal{N}$ in which e is not present. On the other hand, if $\mathcal{N} - e \notin [\mathcal{N}]$ then we note that $\mathcal{N} - e$ is the largest subgraph of $\mathcal{N}$ that does not contain e. Let $\mathcal{K}$ be a subgraph of $\mathcal{N}$ that does not contain e. Then $\mathcal{I}(\mathcal{N}) \subsetneq \mathcal{I}(\mathcal{N} - e) \subseteq \mathcal{I}(\mathcal{K})$. Using Theorem 5.9, we know that all $\mathcal{N}$-Markov equivalent DMGs are in fact subgraphs of $\mathcal{N}$, and using that $\mathcal{K}$ is not Markov equivalent to $\mathcal{N}$ we see that all graphs in $[\mathcal{N}]$ must contain e. This means that when $\mathcal{N} - e \notin [\mathcal{N}]$ then $e \notin \bar{F}$ as e must be present in all Markov equivalent DMGs.

Any loop should in principle be dashed when drawing a DMEG as for each node in a maximal DMG either both the directed and the bidirected loop are present or neither of them. However, we choose to not present them as dashed as if they are present in the maximal DMG, then at least one of them will be present in any Markov equivalent DMG satisfying (3.3), that is, for any DMG which is a marginalization of a DG. In addition, we only draw the directed loop to not overload the visualizations.

5.5. *Constructing a directed mixed equivalence graph.* When constructing a DMEG from $\mathcal{N}$, it suffices to consider the graphs $\mathcal{N} - e$ for each $e \in E$ and determine if they are Markov equivalent to $\mathcal{N}$ or not. A brute-force approach to doing so is to simply check all separation statements in both graphs. However, one can make a considerably more efficient algorithm.

PROPOSITION 5.11. *Assume $\alpha \xrightarrow{e}_{\mathcal{N}} \beta$. It holds that $\mathcal{N} - e \in [\mathcal{N}]$ if and only if $\alpha \in u(\beta, \mathcal{I}(\mathcal{N} - e))$.*

PROPOSITION 5.12. *Assume $\alpha \overset{e}{\leftrightarrow}_{\mathcal{N}} \beta$. Then $\mathcal{N} - e \in [\mathcal{N}]$ if and only if $\alpha \in u(\beta, \mathcal{I}(\mathcal{N} - e))$ and $\beta \in u(\alpha, \mathcal{I}(\mathcal{N} - e))$.*

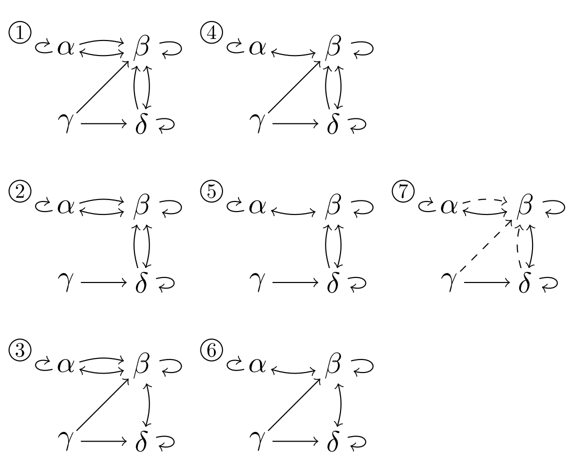

FIG. 8. *The DMG ① is maximal (the bidirected loops at α, β and δ have been omitted from the visual presentation). The DMGs ①–⑥ are the six elements of its Markov equivalence class (when ignoring Markov equivalent removal of loops). The graph ⑦ is the corresponding DMEG. In a DMEG, every solid edge is in every graph in the equivalence class, every absent edge is not in any graph, and every dashed edge is in some, but not in others. Note that every DMG in the above equivalence class contains the edge $\gamma \to \beta$ or the edge $\delta \to \beta$ even though both are dashed in the DMEG. This example shows that not every equivalence class contains a least element.*

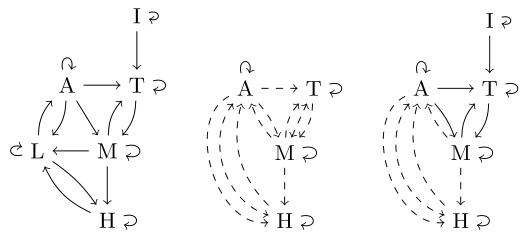

FIG. 9. *Left*: *Local independence graph of Example* 2.3. *Middle*: *DMEG for the marginalization over* L *and* I. *Right*: *DMEG for the marginalization over* L. *We have omitted the bidirected loops from the DMEGs and presented the directed loops as solid.*

We can now outline a two-step algorithm for constructing the DMEG from an arbitrary DMG, $\mathcal{G}$. We first construct the maximal Markov equivalent graph, $\mathcal{N}$. We know from Theorem 5.9 that one can simply check if each pair of nodes are potential siblings/parents in the independence model induced by $\mathcal{G}$ and construct the maximal Markov equivalent graph directly. This may, however, not be computationally efficient.

The above propositions show that given the maximal DMG, one can efficiently construct the DMEG by evaluating separability once for each directed edge and twice for each bidirected edge. Using Proposition 4.10, one can determine separability by testing a single separation statement, and this means that starting from $\mathcal{N}$, one can construct the corresponding DMEG in a way such that the number of separation statements to test scales linearly in the number of edges in $\mathcal{N}$.

EXAMPLE 5.13 (Gateway drugs, continued). We return to the model in Example 2.3 to consider what happens when it is only partially observed and to give an interpretation of the corresponding local independence model. The local independence graph is assumed to be as depicted on Figure 9, left.

Consider first the situation where L and I are unobserved. In this case, under the faithfulness assumption of the full model (Definition C.5) we can construct the DMEG, which is shown in the center panel of Figure 9, from the local independence model. The DMEG represents the Markov equivalence class which we can infer from the marginal local independence model (L and I are unobserved). Theoretically, the inference requires an oracle to provide us with local independence statements, which will in practice have to be approximated by statistical tests. What is noteworthy is that the DMEG can be inferred from the distribution of the observed variables only, and we do not need to know the local independences of the full model.

If we ignore which edges are dashed and which are not, the graph simply represents the local independence model of the marginal system as the maximal element in the Markov equivalence class. The dashed edges give us additional—and in some sense local—information. As an example, the directed edge from A to H is dashed and we cannot know if there exists a conditioning set that would render H locally independent of A in the full system. On the other hand, the directed edge from T to H is absent, and we can conclude that tobacco use is not directly affecting hard drug use.

Consider instead the situation where I is also observed. I serves as an analogue to an instrumental variable (see, e.g., [31] for an introduction to instrumental variables). The inclusion of this variable identifies some of the structure by removing some dashed edges and making others nondashed.

6. Discussion and conclusion. In this paper, we introduced a class of graphs to represent local independence structures of partially observed multivariate stochastic processes.

Previous work based on directed graphs, that allows for cycles and use the asymmetric δ-separation criterion, was extended to mixed directed graphs to account for latent processes and we introduced μ-separation in mixed directed graphs.

An important task is the characterization of equivalence classes of graphs and this has been studied, for example, in MAGs [5, 45]. In the case of MAGs, a key result is that every element in a Markov equivalence class has the same *skeleton*, that is, the same adjacencies [5]. As shown by Propositions 4.7 and 4.8, this is not the case for DMGs, and Example 4.12 shows that one cannot necessarily within a Markov equivalence class find an element such that two nodes are inseparable if and only if they are adjacent.

We proved instead a central maximality property which allowed us to propose the use of DMEGs to represent a Markov equivalence class of DMGs in a concise way. Given a maximal DMG, we furthermore argued that one can efficiently find the DMEG. Similar results are known for chain graphs, as one can also in a certain sense find a unique, largest graph representing a Markov equivalence class [20], though this graph is not a supergraph of all Markov equivalent graphs as in the case of DMGs. Volf and Studený [42] suggested to use this largest graph as a unique representative of the Markov equivalence class, and they provided an algorithm to construct it.

We emphasize that the characterization given of the maximal element of a Markov equivalence class of DMGs is constructive in the sense that it straightforwardly defines an algorithm for learning a maximal DMG from a local independence oracle. This learning algorithm may not be computationally efficient or even feasible for large graphs, and it is ongoing research to develop efficient learning algorithms and to develop the practical implementations of the tools needed for replacing the oracle by statistical tests.

Acknowledgments. The authors are grateful to Steffen Lauritzen for his helpful comments and suggestions. We also thank two referees and an area editor whose comments have helped improve this manuscript.

This work was supported by Villum Fonden Grant 13358.

SUPPLEMENTARY MATERIAL

Additional results and proofs (DOI: 10.1214/19-AOS1821SUPP; .pdf). The supplementary material consists of Sections A to F. In Sections A and B, we relate μ-separation to Didelez's δ-separation, and also relate our slightly different definitions of local independence. Section C describes how one can unroll a local independence graph and obtain a DAG. We use this to discuss Markov properties and faithfulness in the time series case. In Section D, we provide an augmentation criterion to determine μ-separation using an auxiliary undirected graph. In Section E, we discuss conditions for existence of compensators and elaborate on the definition of local independence. Section F contains the proofs of the main paper.

REFERENCES

[1] AALEN, O. O. (1987). Dynamic modelling and causality. *Scand. Actuar. J.* **3–4** 177–190. MR0943579 https://doi.org/10.1016/j.rser.2011.04.029

[2] AALEN, O. O., BORGAN, Ø., KEIDING, N. and THORMANN, J. (1980). Interaction between life history events. Nonparametric analysis for prospective and retrospective data in the presence of censoring. *Scand. J. Stat.* **7** 161–171. MR0605986

[3] AALEN, O. O., RØYSLAND, K., GRAN, J. M., KOUYOS, R. and LANGE, T. (2016). Can we believe the DAGs? A comment on the relationship between causal DAGs and mechanisms. *Stat. Methods Med. Res.* **25** 2294–2314. MR3553339 https://doi.org/10.1177/0962280213520436

[4] AALEN, O. O., RØYSLAND, K., GRAN, J. M. and LEDERGERBER, B. (2012). Causality, mediation and time: A dynamic viewpoint. *J. Roy. Statist. Soc. Ser. A* **175** 831–861. MR2993496 https://doi.org/10.1111/j.1467-985X.2011.01030.x

[5] ALI, R. A., RICHARDSON, T. S. and SPIRTES, P. (2009). Markov equivalence for ancestral graphs. *Ann. Statist.* **37** 2808–2837. MR2541448 https://doi.org/10.1214/08-AOS626

[6] COMMENGES, D. and GÉGOUT-PETIT, A. (2009). A general dynamical statistical model with causal interpretation. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 719–736. MR2749916 https://doi.org/10.1111/j.1467-9868.2009.00703.x

[7] CONSTANTINOU, P. and DAWID, A. P. (2017). Extended conditional independence and applications in causal inference. *Ann. Statist.* **45** 2618–2653. MR3737904 https://doi.org/10.1214/16-AOS1537

[8] COX, D. R. and WERMUTH, N. (1996). *Multivariate Dependencies: Models, Analysis and Interpretation. Monographs on Statistics and Applied Probability* **67**. CRC Press, London. MR1456990

[9] DANKS, D. and PLIS, S. (2013). Learning causal structure from undersampled time series. In *JMLR: Workshop and Conference Proceedings (NIPS Workshop on Causality)*.

[10] DAWID, A. P. (2001). Separoids: A mathematical framework for conditional independence and irrelevance. *Ann. Math. Artif. Intell.* **32** 335–372. MR1859870 https://doi.org/10.1023/A:1016734104787

[11] DIDELEZ, V. (2000). *Graphical models for event history analysis based on local independence*. Dissertation, Univ. Dortmund. Logos Verlag Berlin, Berlin. MR1864064

[12] DIDELEZ, V. (2007). Graphical models for composable finite Markov processes. *Scand. J. Stat.* **34** 169–185. MR2325249 https://doi.org/10.1111/j.1467-9469.2006.00528.x

[13] DIDELEZ, V. (2008). Graphical models for marked point processes based on local independence. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 245–264. MR2412641 https://doi.org/10.1111/j.1467-9868.2007.00634.x

[14] EICHLER, M. (2012). Graphical modelling of multivariate time series. *Probab. Theory Related Fields* **153** 233–268. MR2925574 https://doi.org/10.1007/s00440-011-0345-8

[15] EICHLER, M. (2013). Causal inference with multiple time series: Principles and problems. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **371** 20110613, 17. MR3081173 https://doi.org/10.1098/rsta.2011.0613

[16] EICHLER, M. and DIDELEZ, V. (2007). Causal reasoning in graphical time series models. In *Proceedings of the* 23*rd Conference on Uncertainty in Artificial Intelligence* 109–116.

[17] EICHLER, M. and DIDELEZ, V. (2010). On Granger causality and the effect of interventions in time series. *Lifetime Data Anal.* **16** 3–32. MR2575937 https://doi.org/10.1007/s10985-009-9143-3

[18] EVANS, R. J. (2016). Graphs for margins of Bayesian networks. *Scand. J. Stat.* **43** 625–648. MR3543314 https://doi.org/10.1111/sjos.12194

[19] EVANS, R. J. and RICHARDSON, T. S. (2014). Markovian acyclic directed mixed graphs for discrete data. *Ann. Statist.* **42** 1452–1482. MR3262457 https://doi.org/10.1214/14-AOS1206

[20] FRYDENBERG, M. (1990). The chain graph Markov property. *Scand. J. Stat.* **17** 333–353. MR1096723

[21] GÉGOUT-PETIT, A. and COMMENGES, D. (2010). A general definition of influence between stochastic processes. *Lifetime Data Anal.* **16** 33–44. MR2575938 https://doi.org/10.1007/s10985-009-9131-7

[22] HYTTINEN, A., PLIS, S., JÄRVISALO, M., EBERHARDT, F. and DANKS, D. (2016). Causal discovery from subsampled time series data by constraint optimization. In *Proceedings of the Eighth International Conference on Probabilistic Graphical Models* **52** 216–227.

[23] JENSEN, A.-M. and SCHWEDER, T. (1986). The engine of fertility—Influenced by interbirth employment? Discussion paper, Central Bureau of Statistics, Oslo.

[24] KANDEL, D. (1975). Stages in adolescent involvement in drug use. *Science* **190** 912–914.

[25] KOSTER, J. T. A. (1999). On the validity of the Markov interpretation of path diagrams of Gaussian structural equations systems with correlated errors. *Scand. J. Stat.* **26** 413–431. MR1712043 https://doi.org/10.1111/1467-9469.00157

[26] LAURITZEN, S. and SADEGHI, K. (2018). Unifying Markov properties for graphical models. *Ann. Statist.* **46** 2251–2278. MR3845017 https://doi.org/10.1214/17-AOS1618

[27] LU, J. (2015). *On Causal Inference for Ordinal Outcomes*. ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)–Harvard Univ. MR3664965

[28] MEEK, C. (2014). Toward learning graphical and causal process models. In *Proceedings of the UAI* 2014 *Workshop Causal Inference: Learning and Prediction*.

[29] MOGENSEN, S. W. and HANSEN, N. R. (2020). Supplement to “Markov equivalence of marginalized local independence graphs.” https://doi.org/10.1214/19-AOS1821SUPP.

[30] MOGENSEN, S. W., MALINSKY, D. and HANSEN, N. R. (2018). Causal learning for partially observed stochastic dynamical systems. In *Proceedings of the* 34*th Conference on Uncertainty in Artificial Intelligence*.

[31] PEARL, J. (2009). *Causality: Models, Reasoning, and Inference*, 2nd ed. Cambridge Univ. Press, Cambridge. MR2548166 https://doi.org/10.1017/CBO9780511803161

[32] RICHARDSON, T. (2003). Markov properties for acyclic directed mixed graphs. *Scand. J. Stat.* **30** 145–157. MR1963898 https://doi.org/10.1111/1467-9469.00323

[33] RICHARDSON, T. and SPIRTES, P. (2002). Ancestral graph Markov models. *Ann. Statist.* **30** 962–1030. MR1926166 https://doi.org/10.1214/aos/1031689015

[34] RICHARDSON, T. S., EVANS, R. J., ROBINS, J. M. and SHPITSER, I. (2017). Nested Markov properties for acyclic directed mixed graphs. https://arxiv.org/abs/1701.06686.

[35] RØYSLAND, K. (2012). Counterfactual analyses with graphical models based on local independence. *Ann. Statist.* **40** 2162–2194. MR3059080 https://doi.org/10.1214/12-AOS1031

[36] SADEGHI, K. (2013). Stable mixed graphs. *Bernoulli* **19** 2330–2358. MR3160556 https://doi.org/10.3150/12-BEJ454

[37] SCHWEDER, T. (1970). Composable Markov processes. *J. Appl. Probab.* **7** 400–410. MR0264755 https://doi.org/10.2307/3211973

[38] SOKOL, A. and HANSEN, N. R. (2014). Causal interpretation of stochastic differential equations. *Electron. J. Probab.* **19** no. 100, 24. MR3275852 https://doi.org/10.1214/ejp.v19-2891

[39] SPIRTES, P., RICHARDSON, T. S. and MEEK, C. (1997). The dimensionality of mixed ancestral graphs. Technical report. Philosophy Dept, Carnegie Mellon Univ., Pittsburgh, PA.

[40] VANYUKOV, M. M., TARTER, R. E., KIRILLOVA, G. P., KIRISCI, L., REYNOLDS, M. D., KREEK, M. J., CONWAY, K. P., MAHER, B. S., IACONO, W. G., BIERUT, L., NEALE, M. C., CLARK, D. B. and RIDENOUR, T. A. (2012). Common liability to addiction and "gateway hypothesis": Theoretical, empirical and evolutionary perspective. *Drug Alcohol Depend.* **123**.

[41] VERMA, T. and PEARL, J. (1991). Equivalence and synthesis of causal models Technical Report Univ. California, Los Angeles.

[42] VOLF, M. and STUDENÝ, M. (1999). A graphical characterization of the largest chain graphs. *Internat. J. Approx. Reason.* **20** 209–236. MR1685080 https://doi.org/10.1016/S0888-613X(99)00003-1

[43] WORLD HEALTH ORGANIZATION (2004). Neuroscience of psychoactive substance use and dependence Technical Report World Health Organization.

[44] XU, H., FARAJTABAR, M. and ZHA, H. (2016). Learning Granger causality for Hawkes processes. In *Proceedings of the* 33*rd International Conference on Machine Learning*.

[45] ZHAO, H., ZHENG, Z. and LIU, B. (2005). On the Markov equivalence of maximal ancestral graphs. *Sci. China Ser. A* **48** 548–562. MR2157690 https://doi.org/10.1360/04ys0023

SUPPLEMENTARY MATERIAL FOR MARKOV EQUIVALENCE OF MARGINALIZED LOCAL INDEPENDENCE GRAPHS

By Søren Wengel Mogensen and Niels Richard Hansen

University of Copenhagen

In this supplementary material we discuss relations between μ-separation and other asymmetric notions of graphical separation. We also compare our proposed definition of local independence to previous definitions to argue that ours is in fact a generalization. We furthermore relate μ-separation to m-separation. We provide, in particular, a detailed discussion of the local independence model for discrete-time stochastic processes (time series), and we show how to verify μ-separation via separation in an auxiliary undirected graph. We also discuss the existence of the compensators that are used in the definition of local independence for continuous-time stochastic process models. This supplementary material also contains proofs of the results of the main paper. A list of references can be found on the last page.

A. Relation to other asymmetric notions of graphical separation. In this section we relate μ-separation to δ-separation as introduced previously in the literature for directed graphs.

Definition A.1 (Bereaved graph). Let $\mathcal{G} = (V, E_d)$ be a DG, and let $B \subseteq V$. The B-*bereaved graph*, $\mathcal{G}^B$, is constructed from $\mathcal{G}$ by removing every directed edge with a tail at a node in B except loops. More precisely, $\mathcal{G}^B = (V, \bar{E}_d^B)$, where $\bar{E}_d^B = E_d \setminus \left(\bigcup_{\beta\in B}\{(\beta,\delta) \mid \delta \neq \beta\}\right)$.

Didelez [2] considered a DG, and for disjoint sets $A, B, C \subseteq V$ said that B is separated from A by C if there is no μ-connecting walk in $\mathcal{G}^B$, or equivalently, no μ-connecting path. This is called δ-separation. Note that the condition in Definitions 3.1 and 3.2 that a connecting walk be nontrivial makes no difference now due to A and B being disjoint. The condition that a μ-connecting walk ends with a head at $\beta \in B$ is also obsolete as we are evaluating separation in the bereaved graph $\mathcal{G}^B$. Didelez [2] always assumed that a process depended on its own past, and thus did not visualize loops in the DGs as a loop would always be present at every node.

Meek [9] generalized δ-separation to δ^*-separation in a DG (allowing for loops) by considering only nontrivial μ-connecting walks in $\mathcal{G}^B$ for sets

$A, B, C \subseteq V$ such that $A \cap C = \emptyset$ with the motivation that a node can be separated from itself using this notion of separation. However, if we consider the graph $\alpha \to \beta$, and sets $A = \{\alpha\}$, $B = \{\alpha, \beta\}$, $C = \emptyset$, then using δ^*-separation, B is separated from A given C, which runs counter to an intuitive understanding of separation. More importantly, δ^*-separation in the local independence graph will not generally imply local independence.

To establish an exact relationship between δ- and μ-separations and argue that we are indeed proposing a generalization of the former, assume that $\mathcal{G}$ is a DG and that $A, B, C \subseteq V$ are disjoint. We will argue that

$$A \perp_\mu B \mid C \cup B \; [\mathcal{G}] \Leftrightarrow A \perp_\delta B \mid C \; [\mathcal{G}]. \tag{A.1}$$

To see that this is the case, consider first a δ-connecting walk from $\alpha \in A$ to $\beta \in B$ given C in $\mathcal{G}^B$, ω. The subwalk from α to the first node on ω which is in B is also present and μ-connecting given $C \cup B$ in $\mathcal{G}$. On the other hand, assume that there exists a μ-connecting sequence, ω, in $\mathcal{G}$. We know that $A \cap B = \emptyset$, and because B is a subset of the conditioning set on the left hand side in (A.1), we must have that the first time the path enters B, it has a head at the node in B, and this implies that a subwalk of ω is δ-connecting, that is, present and connecting in $\mathcal{G}^B$. In Section B we will discuss why B is included in the conditioning set on the left side of (A.1).

B. Markov properties. The equivalence of pairwise and global Markov properties is pivotal in much of graphical modeling. In this section, we will show how our proposed graphical framework fits with known results on Markov properties in the case of point processes and argue that our graphical framework is a generalization of that of Didelez [3] to allow for non-disjoint sets and unobserved processes.

DEFINITION B.1 (The pairwise Markov property). Let $\mathcal{I}$ be an independence model over V. We say that $\mathcal{I}$ satisfies the *pairwise Markov property* with respect to the DG $\mathcal{D}$ if for all $\alpha, \beta \in V$,

$$\alpha \not\to_{\mathcal{D}} \beta \Rightarrow \langle \alpha, \beta \mid V \setminus \{\alpha\} \rangle \in \mathcal{I}.$$

DEFINITION B.2 (The global Markov property). Let $A, B, C \subseteq V$. Let $\mathcal{I}$ be an independence model over V. We say that $\mathcal{I}$ satisfies the *global Markov property* with respect to the DMG $\mathcal{G}$ if $\mathcal{I}(\mathcal{G}) \subseteq \mathcal{I}$, i.e., if

$$A \perp_\mu B \mid C \; [\mathcal{G}] \Rightarrow \langle A, B \mid C \rangle \in \mathcal{I}.$$

Didelez [3] only considered disjoint sets and gave a slighty different definition of local independence. For disjoint sets, Didelez [3] defined that B is locally independent of A given C if

$$A \not\to B \mid C \cup B,$$

and we will make the relation between the two definitions precise in this section. Consider sets $\mathcal{S}, \mathcal{S}_d \subseteq \mathcal{P}(V) \times \mathcal{P}(V) \times \mathcal{P}(V)$,

$$\begin{aligned} \mathcal{S}_d &= \{(A, B, C) \mid A, B, C \text{ disjoint}, A, B \text{ non-empty}\} \\ \mathcal{S} &= \{(A, B, C) \mid B \subseteq C, \; A, C \text{ disjoint}, A, B \text{ non-empty}\} \end{aligned}$$

and the bijection $s : \mathcal{S}_d \to \mathcal{S}$, $s((A, B, C)) = (A, B, C \cup B)$. We will in this section let $\mathcal{I}$ denote a subset of $\mathcal{S}$ and let $\mathcal{I}^d$ denote a subset of $\mathcal{S}_d$. In Section A we argued that for any directed graph $\mathcal{G}$ and $(A, B, C) \in \mathcal{S}_d$,

$$A \perp_\delta B \mid C \; [\mathcal{G}] \Leftrightarrow A \perp_\mu B \mid C \cup B \; [\mathcal{G}]$$

and therefore

$$\{(A, B, C) \in \mathcal{S}_d : A \perp_\delta B \mid C \, [\mathcal{G}]\} = s^{-1}\Big(\{(A, B, C) \in \mathcal{S} : A \perp_\mu B \mid C \, [\mathcal{G}]\}\Big).$$

For any local independence model defined by Didelez's definition, $\mathcal{I}^d$, and any local independence model defined by Definition 2.1, $\mathcal{I}$, it holds that

$$\begin{aligned} \langle A, B \mid C \rangle \in \mathcal{I}^d &\Leftrightarrow A \not\to B \mid C \cup B \\ &\Leftrightarrow \langle A, B \mid C \cup B \rangle \in \mathcal{I} \end{aligned}$$

so $\mathcal{I}^d = s^{-1}(\mathcal{I})$. Hence, there is a bijection between the two sets, and graphical and probabilistic independence models are preserved under the bijection. This means that we have equivalence of Markov properties between the two formulations. Thus, restricting our framework to $\mathcal{S}$, we get the equivalence of pairwise and global Markov property directly from the proof by Didelez in the case of point process models, and we see that our seemingly different definitions of local independence and graphical separation indeed give an extension of earlier work.

One can show that for two DMGs $\mathcal{G}_1$, $\mathcal{G}_2$, that both have all directed and bidirected loops it holds that

$$\mathcal{I}(\mathcal{G}_1) \cap \mathcal{S} = \mathcal{I}(\mathcal{G}_2) \cap \mathcal{S} \Leftrightarrow \mathcal{I}(\mathcal{G}_1) = \mathcal{I}(\mathcal{G}_2).$$

Let $\mathbb{G}$ denote the class of DMGs such that all directed and bidirected loops are present. Consider now some $\mathcal{G} \in \mathbb{G}$. By the above result we can identify the Markov equivalence class from the independence model restricted to $\mathcal{S}$. This equivalence class has a maximal element which is also in $\mathbb{G}$ and thus one can also in this case represent the Markov equivalence class using a DMEG.

C. Time series and unrolled graphs. In this section we first relate the cyclic DGs and DMGs to acyclic graphs and then use this to discuss Markov properties (see Definition B.2) and faithfulness of local independence models in the time series case.

DEFINITION C.1 (m-separation [10]). Let $\mathcal{G} = (V, E)$ be a DMG and let $\alpha, \beta \in V$. A path between α and β is said to be m-connecting if no noncollider on the path is in C and every collider on the path is in $An(C)$. For disjoint sets $A, B, C \subseteq V$, we say that A and B are m-separated by C if there is no m-connecting path between $\alpha \in A$ and $\beta \in B$. In this case, we write $A \perp_m B \mid C$.

The above m-separation is a generalization of the well-known d-separation in DAGs. In this section we will only consider m-separation for DAGs, and will thus use the d-separation terminology. In Section D we provide a more general relation between μ-separation and m-separation.

We first describe how to obtain a DAG from a DG such that the DAG, if read the right way, will give the same separation model as the DG. This can be useful in time series examples as well as when working with continuous-time models. Sokol and Hansen [13] studied solutions to stochastic differential equations and used a DAG in discrete time to approximate the continuous-time dynamics. Danks and Plis [1] and Hyttinen et al. [5] used similar translations between an *unrolled* graph in which time is discrete and explicit and a *rolled* graph in which time is implicit. Some authors use the term *unfolded* instead of unrolled. In a rolled graph each node represents a stochastic process whereas in an unrolled graph each node represents

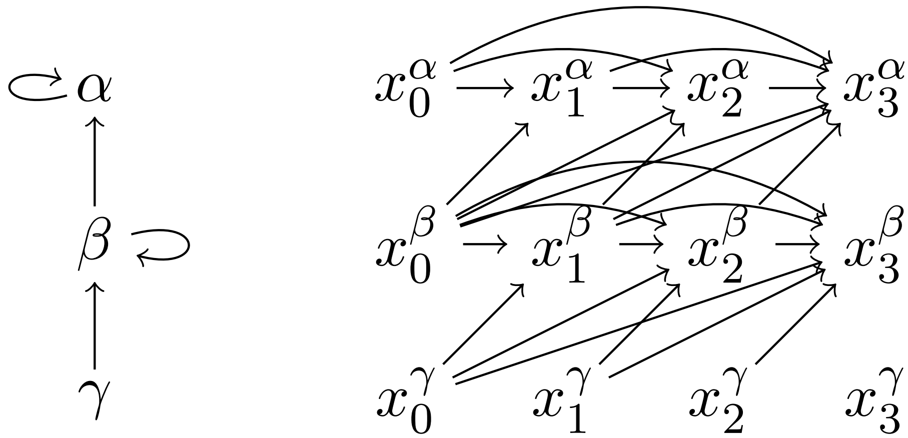

Fig 1. *A directed graph (left) and the corresponding unrolled version with four time points, $\mathcal{D}_3(\mathcal{G})$, (right). x_t^δ denotes the δ-coordinate process at time t for $\delta \in \{\alpha, \beta, \gamma\}$.*

a single random variable. Definition C.2 shows how to unroll a local independence graph and Lemma C.3 establishes a precise relationship between independence models in the rolled and unrolled graphs.

Definition C.2. Let $\mathcal{G} = (V, E)$ be a DG and let $T \in \mathbb{N}$. The unrolled version of $\mathcal{G}$, $\mathcal{D}_T(\mathcal{G}) = (\bar{V}, \bar{E})$, is the DAG on nodes

$$\bar{V} = \{x_t^\alpha \mid (t, \alpha) \in \{0, 1, \dots, T\} \times V\}$$

and with edges

$$\bar{E} = \{x_s^\alpha \to x_t^\beta \mid \alpha \to_{\mathcal{G}} \beta \text{ and } s < t\}.$$

Let $D \subseteq V$ and let $T \in \mathbb{N}$. We define $D_{0:T} = \{x_t^\alpha \in \bar{V} \mid \alpha \in D,\ t \leq T\}$ and $D_T = \{x_t^\alpha \in \bar{V} \mid \alpha \in D,\ t = T\}$.

Lemma C.3. Let $\mathcal{G} = (V, E)$ be a DG. If $A \perp_\mu B \mid C\ [\mathcal{G}]$ then $(A \setminus C)_{0:(T-1)} \perp_d B_T \mid C_{0:(T-1)}\ [\mathcal{D}_T(\mathcal{G})]$. For large enough values of T, the opposite implication holds as well.

Proof. Assume first that $\langle x_{s_0}^{\alpha_0}, e_1, x_{s_1}^{\alpha_1}, \dots, e_l, x_{s_l}^{\alpha_l} \rangle$ is a d-connecting path in $\mathcal{D}_T(\mathcal{G})$. This path has a head at $x_{s_l}^{\alpha_l} \in B_T$. Construct a walk in $\mathcal{G}$ by for each node, $x_{s_k}^{\alpha_k}$, taking the corresponding node, α_k, and for each edge

$x_{s_k}^{\alpha_k} \sim x_{s_{k+1}}^{\alpha_{k+1}}$ taking the corresponding, endpoint-identical edge $\alpha_k \sim \alpha_{k+1}$ in $\mathcal{G}$. On this walk, no noncollider is in C, and every collider is an ancestor of a node in C.

Assume instead that ω is a μ-connecting walk in $\mathcal{G}$ from A to B given C,

$$\alpha_1 \sim \ldots \sim \alpha_{l-1} \to \alpha_l$$

and let $T \geq 3(|E|+1)+1$. Using Proposition 3.5, we can assume that ω has length smaller than or equal to $|E|+1$. We construct a d-connecting walk in $\mathcal{D}_T(\mathcal{G})$ in the following way. Starting from $x_T^{\alpha_l}$, we choose the edge between $x_{|E|+1}^{\alpha_{l-1}}$ and $x_T^{\alpha_l}$. For the remaining edges, $\alpha_k \sim \alpha_{k+1}$, we choose the edge $x_{s_k-1}^{\alpha_k} \to x_{s_k}^{\alpha_{k+1}}$ if $\alpha_k \to \alpha_{k+1}$ in ω, and $x_{s_k}^{\alpha_{k+1}} \to x_{s_k+1}^{\alpha_k}$ if $\alpha_k \leftarrow \alpha_{k+1}$ in ω where s_k is determined by the endpoints of the previous edge. No noncollider on this walk will be in $C_{0:(T-1)}$. Every collider will be in $An_{\mathcal{D}_T(\mathcal{G})}(C_{0:(T-1)})$ as the collider will be in the time slices 0 to $2(|E|+1)$. This d-connecting walk can be trimmed down to a d-connecting path. □

We defined local independence for a class of continuous-time processes in Definition 2.1. In this section we define a similar notion for time series, as also introduced in [4]. Let $V = \{1, \ldots, n\}$. We consider a multivariate time series $(X_t)_{t \in \mathbb{N} \cup \{0\}}$, $X_t = (X_t^1, \ldots, X_t^n)$, of the form

$$X_t^\alpha = f_{\alpha t}(X_{s<t}, \varepsilon_t^\alpha),$$

where $X_{s<t} = \{X_u^\alpha \mid \alpha \in V, u < t\}$. The random variables $\{\varepsilon_t^\alpha\}$ are independent. For $S \subseteq \mathbb{N} \cup \{0\}$ and $D \subseteq V$ we let $X_S^D = \{X_s^\alpha \mid \alpha \in D, s \in S\}$ and $X^D = \{X^\alpha \mid \alpha \in D\}$. In the case of time series, a notable feature of local independence and local independence graphs is that they provide a simple representation in comparison with graphs in which each vertex represents a single time-point variable.

Definition C.4 (Local independence, time series). Let X be a multivariate time series. We say that X^B is *locally independent* of X^A given X^C if for all $t \in \mathbb{N}$, $\beta \in B$, $X_{s<t}^A$ and X_t^β are conditionally independent given $X_{s<t}^C$, that is,

$$X_{s<t}^A \perp\!\!\!\perp X_t^\beta \mid X_{s<t}^C$$

and write $A \not\to B \mid C$.

The above definition induces an independence model over V, which we will also refer to as the local independence model and denote $\mathcal{I}$ in the following. The main question that we address is whether this independence model is graphical. That is, we will construct a DG, consider the Markov and faithfulness properties of $\mathcal{I}$ and this DG, and relate them to Markov and faithfulness properties of the conditional independence model of finite distributions and unrolled versions of the DG.

DEFINITION C.5 (Faithfulness). Let $A, B, C \subseteq V$. Let $\mathcal{I}$ be an independence model on V and let $\mathcal{G}$ be a DMG. We say that $\mathcal{I}$ and $\mathcal{G}$ are *faithful* if $\mathcal{I} = \mathcal{I}(\mathcal{G})$, i.e., if

$$\langle A, B \mid C \rangle \in \mathcal{I} \Leftrightarrow A \perp_\mu B \mid C \; [\mathcal{G}].$$

One can give analogous definitions using other notions of graphical separation. Below we also consider faithfulness of a probability distribution and a DAG, implicitly using d-separation instead of μ-separation in the above definition.

Let $\mathcal{D}_T$ for $T \geq 1$ be the DAG on nodes $\{x_s^\alpha \mid s \in \{0, \ldots, T\}, \alpha \in V\}$ such that there is an edge $x_s^\alpha \to x_t^\beta$ if and only if $f_{\beta t}$ depends on the argument X_s^α. Let $D_S = \{x_s^\alpha \mid \alpha \in D, s \in S\}$. Let $\mathcal{G}$ denote the minimal DG such that its unrolled version, $\mathcal{D}_T(\mathcal{G})$, is a supergraph of $\mathcal{D}_T$ for all $T \in \mathbb{N}$.

For all $T \in \mathbb{N}$, the DAG $\mathcal{D}_T(\mathcal{G})$ and the distribution of $X_{s \leq T}$ satisfy

$$x_s^\alpha, x_t^\beta \text{ not adjacent } \Rightarrow X_s^\alpha \perp\!\!\!\perp X_t^\beta \mid (An(X_s^\alpha) \cup An(X_t^\beta)) \setminus \{X_s^\alpha, X_t^\beta\},$$

which is also known as the pairwise Markov property for DAGs. Assume equivalence of the pairwise and global Markov properties for this DAG and the finite-dimensional distribution (see e.g. [7] for necessary and sufficient conditions for this equivalence). Assume that B is μ-separated from A by C in the DG $\mathcal{G}$, $A \perp_\mu B \mid C \; [\mathcal{G}]$. By Lemma C.3, $(A \setminus C)_{s<T} \perp_m B_T \mid C_{s<T} \; [\mathcal{D}_T(\mathcal{G})]$, and by the global Markov property in this DAG, $X_{s<T}^{A \setminus C} \perp\!\!\!\perp X_T^B \mid X_{s<T}^C$. This holds for any T, and therefore $A \setminus C \not\rightarrow B \mid C$. It follows that $A \not\rightarrow B \mid C$. This means that $\mathcal{I}$ satisfies the global Markov property with respect to $\mathcal{G}$.

Assume furthermore that the distribution of X_T and the DAG $\mathcal{D}_T(\mathcal{G})$ for some $T \in \mathbb{N}$ are faithful and that $T \geq 3(|E| + 1) + 1$. Meek [8] studied faithfulness of DAGs and argued that faithful distributions exist for any DAG. If $A \not\rightarrow B \mid C$, then $A \setminus C \not\rightarrow B \mid C$ and $X_{s<T}^{A \setminus C} \perp\!\!\!\perp X_T^B \mid X_{s<T}^C$. By faithfulness of the distribution of X_T and the DAG $\mathcal{D}_T(\mathcal{G})$, we have $(A \setminus C)_{s<T} \perp_m B_T \mid C_{s<T} \; [\mathcal{D}_T(\mathcal{G})]$ and using Lemma C.3 this implies that $A \perp_\mu B \mid C \; [\mathcal{G}]$, giving us faithfulness of $\mathcal{I}$ and $\mathcal{G}$.

In summary, for every DG there exists a time series such that the local independence model induced by its distribution and the DG are faithful.

D. An augmentation criterion. In this section we present results that allow us to determine μ-separation from graphical separation in an undirected graph. An *undirected graph* is a graph, (V, E), with an edge set that consists of unordered pairs of nodes such that every edge is of the type $-$. Let A, B, and C be disjoint subsets of V. We say that A and B are separated by C if every path between $\alpha \in A$ and $\beta \in B$ contains a node in C.

When working with d-separation in DAGs, it is possible to give an equivalent separation criterion using a derived undirected graph, the *moral* graph [6]. Didelez [2] also gives both pathwise and so-called moral graph criteria for δ-separation. The augmented graph below is a generalization of the moral graph [10, 11] which allows one to give a criterion for m-separation based on an augmented graph. We use the similarity of μ-separation and m-separation to give an augmentation graph criterion for μ-separation. The first step in making a connection to m-separation is to explicate that each node of a DMG represents an entire stochastic process, and notably, both the past and the present of that process. We do that using graphs of the below type.

Definition D.1. Let $\mathcal{G} = (V, E)$ and let $B = \{\beta_1, \ldots, \beta_k\} \subseteq V$. The B-history version of $\mathcal{G}$, denoted by $\mathcal{G}(B)$, is the DMG with node set $V \,\dot\cup\, \{\beta_1^p, \ldots, \beta_k^p\}$ such that $\mathcal{G}(B)_V = \mathcal{G}$ and

- $\alpha \leftrightarrow_{\mathcal{G}(B)} \beta_i^p$ if $\alpha \leftrightarrow_{\mathcal{G}} \beta_i$ and $\alpha \in V, \beta_i \in B$,
- $\alpha \rightarrow_{\mathcal{G}(B)} \beta_i^p$ if $\alpha \rightarrow_{\mathcal{G}} \beta_i$ and $\alpha \in V, \beta_i \in B$.

$\mathcal{G}(B)$ is a graph such that every node $b \in B$ is simply split in two: one that represents the present and one that represents the past. We define $B^p = \{\beta_1^p, \ldots, \beta_k^p\}$.

Proposition D.2. Let $\mathcal{G} = (V, E)$ be a DMG, and let $A, B, C \subseteq V$. Then

$$A \perp_\mu B \mid C \ [\mathcal{G}] \Leftrightarrow A \setminus C \perp_m B^p \mid C \ [\mathcal{G}(B)].$$

Proof. Assume first that there is a μ-connecting walk from $\alpha \in A$ to $\beta \in B$ given C in $\mathcal{G}$. By definition $\alpha \in A \setminus C$. By Proposition 3.5 there is a μ-connecting route,

$$\alpha \sim \ldots \sim \beta \sim \ldots \gamma *\!\!\rightarrow \beta.$$

The subwalk from α to γ is also present in $\mathcal{G}(B)$ and composing it with $\gamma *\!\!\rightarrow_{\mathcal{G}(B)} \beta^p$ gives an m-connecting path between $A \setminus C$ and B^p which is open given C.

On the other hand, if there is an m-connecting path from $\alpha \in A \setminus C$ to $\beta^p \in B^p$ given C in $\mathcal{G}(B)$, then no non-endpoint node is in B^p,

$$\alpha \sim \ldots \gamma *\!\!\rightarrow \beta^p$$

The subpath from α to γ is present in $\mathcal{G}$ and can be composed with the edge $\gamma *\!\!\rightarrow \beta$ to obtain a μ-connecting walk from A to B given C in $\mathcal{G}$. □

Definition D.3. Let $\mathcal{G} = (V, E)$ be a DMG. We define the *augmented graph* of $\mathcal{G}$, $\mathcal{G}^a$, to be the undirected graph without loops and with node set V such that two distinct nodes are adjacent if and only if the two nodes are collider connected in $\mathcal{G}$.

Proposition D.4. Let $\mathcal{G} = (V, E)$ be a DMG, $A, B, C \subseteq V$. Then $A \perp_\mu B \mid C\ [\mathcal{G}]$ if and only if $A \setminus C$ and B^p are separated by C in the augmented graph of $\mathcal{G}(B)_{An(A\cup B^p \cup C)}$.

Proof. Using Proposition D.2 we have that $A \perp_\mu B \mid C\ [\mathcal{G}] \Leftrightarrow A\setminus C \perp_m B^p \mid C\ [\mathcal{G}(B)]$. Let $\mathcal{G}(B)'$ be the DMG obtained from $\mathcal{G}(B)$ by removing all loops. Then $A\setminus C \perp_m B^p \mid C\ [\mathcal{G}(B)]$ if and only if $A\setminus C \perp_m B^p \mid C\ [\mathcal{G}(B)']$. We can apply Theorem 1 of [10]. That theorem assumes an ADMG, however, as noted in the paper, acyclicity is not used in the proof which therefore also applies to $\mathcal{G}(B)'$, and we conclude that $A \setminus C \perp_m B^p \mid C\ [\mathcal{G}(B)']$ if and only if $A \setminus C$ and B^p are separated by C in $(\mathcal{G}(B)'_{An(A\cup B^p\cup C)})^a = (\mathcal{G}(B)_{An(A\cup B^p \cup C)})^a$.

□

E. Existence of compensators. Let $Z = (Z_t)$ denote a real-valued stochastic process defined on a probability space $(\Omega, \mathcal{F}, P)$, and let $(\mathcal{G}_t)$ denote a right-continuous and complete filtration w.r.t. P such that $\mathcal{G}_t \subseteq \mathcal{F}$. Note that Z is not assumed adapted w.r.t. the filtration. When Z is a right-continuous process of finite and integrable variation, it follows from Theorem VI.21.4 in [12] that there exists a predictable process of integrable variation, Z^p, such that ${}^{o}Z - Z^{\mathrm{p}}$ is a martingale. Here ${}^{o}Z$ denotes the *optional projection* of Z, which is a right-continuous version of the process $(E(Z_t \mid \mathcal{G}_t))$, cf. Theorem VI.7.1 and Lemma VI.7.8 in [12]. The process $\Lambda = Z^{\mathrm{p}}$

is called the dual predictable projection or compensator of the optional projection ${}^{\mathrm{o}}Z$ as well as of the process Z itself. It depends on the filtration $(\mathcal{G}_t)$.

If Z is adapted w.r.t. a (right-continuous and complete) filtration $(\mathcal{F}_t)$, it has a compensator $\tilde{\Lambda} = Z^{\mathrm{p}}$ such that $Z - \tilde{\Lambda}$ is an $\mathcal{F}_t$ martingale. When $\mathcal{G}_t \subseteq \mathcal{F}_t$ it may be of interest to understand the relation between Λ, as defined above w.r.t. $(\mathcal{G}_t)$, and $\tilde{\Lambda}$. If $\tilde{\Lambda}$ is continuous with $\tilde{\Lambda}_0 = 0$, say, we may ask if Λ equals the *predictable projection*, $E(\tilde{\Lambda}_t \mid \mathcal{G}_{t-})$. As $\tilde{\Lambda}$ is assumed continuous and is of finite variation,

$$\tilde{\Lambda}_t = \int_0^t \tilde{\lambda}_s \mathrm{d}s.$$

If $(\tilde{\lambda}_t)$ itself is an integrable right-continuous process, then its optional projection, $(E(\tilde{\lambda}_t \mid \mathcal{G}_t))$, is an integrable right-continuous process, and

$$E(\tilde{\Lambda}_t \mid \mathcal{G}_{t-}) = \int_0^t E(\tilde{\lambda}_s \mid \mathcal{G}_s)\mathrm{d}s$$

is a finite-variation, continuous version of the predictable projection of $\tilde{\Lambda}$. It is clear that

$$E(Z_t \mid \mathcal{G}_t) - \int_0^t E(\tilde{\lambda}_s \mid \mathcal{G}_s)\mathrm{d}s$$

is a $\mathcal{G}_t$ martingale, thus

$$\Lambda_t = \int_0^t E(\tilde{\lambda}_s \mid \mathcal{G}_s)\mathrm{d}s$$

is a compensator of Z w.r.t. the filtration $(\mathcal{G}_t)$.

We formulate the consequences of the discussion as a criterion for determining local independence via the computation of conditional expectations. The setup is as in Definition 2.1 in Section 2.1.

Proposition E.1. Assume that the process X^β for all $\beta \in V$ has a compensator w.r.t. the filtration $(\mathcal{F}_t^V)$ of the form

$$\Lambda_t^{V,\beta} = \Lambda_0^{V,\beta} + \int_0^t \lambda_s^\beta \mathrm{d}s$$

for an integrable right-continuous process (λ_t^β) and a deterministic constant $\Lambda_0^{V,\beta}$. Then X^β is locally independent of X^A given X^C for $A, C \subseteq V$ if the optional projection

$$E(\lambda_t^\beta \mid \mathcal{F}_t^{A\cup C})$$

has an $\mathcal{F}_t^C$ adapted version.

Another way to phrase the conclusion of the proposition is that if the optional projection $E(\lambda_t^\beta \mid \mathcal{F}_t^C)$ is indistinguishable from $E(\lambda_t^\beta \mid \mathcal{F}_t^{A\cup C})$, then $A \not\to \beta \mid C$, and it is a way of testing local independence via the computation of conditional expectations. It is a precise formulation of the innovation theorem stating how to compute compensators for one filtration via conditional expectations of compensators for a superfiltration.

F. Proofs. The following are proofs of the results from the main paper.

PROOF OF PROPOSITION 3.3. Let ω be a μ-connecting walk given C and let γ be a collider on the walk such that $\gamma \in An(C) \setminus C$. Then there exists a subwalk $\bar{\omega} = \alpha_1 \ast\!\!\to \gamma \leftarrow\!\!\ast\ \alpha_2$, and an open (given C), directed path from γ to $\delta \in C$, π. By composing $\alpha_1 \ast\!\!\to \gamma$ with π, π^{-1}, and $\gamma \leftarrow\!\!\ast\ \alpha_2$ we get an open walk which is endpoint-identical to $\bar{\omega}$ and with its only collider, δ, in C, and we can substitute $\bar{\omega}$ with this new walk. Making such a substitution for every collider in $An(C)\setminus C$ on ω, we obtain a μ-connecting walk on which every collider is in C. □

PROOF OF PROPOSITION 3.5. Assume that we start from α and continue along ω until some node, $\gamma \neq \beta$, is repeated. Remove the cycle from γ to γ to obtain another walk from α to β, $\bar{\omega}$. If $\gamma = \alpha$, then $\bar{\omega}$ is μ-connecting. Instead assume $\gamma \neq \alpha$. If this instance of γ is a noncollider on $\bar{\omega}$ then it must have been a noncollider in an instance on ω and thus $\gamma \notin C$. If on the other hand this instance of γ is a collider on $\bar{\omega}$ then either γ was a collider in an instance on ω or the ancestor of a collider on ω, and thus $\gamma \in An(C)$. In either case, we see that $\bar{\omega}$ is a μ-connecting walk. Repeating this argument, we can construct a μ-connecting walk where only β is potentially repeated. If there is $n > 2$ instances of β then we can remove at least $n - 2$ of them as above as long as we leave an edge with a head at the final β. □

PROOF OF PROPOSITION 3.6. Note first that a vertex can be a parent of itself. The result then follows from the fact that $\alpha \perp_\mu \beta \mid \mathrm{pa}(\beta)$. □

PROOF OF PROPOSITION 3.9. The first statement follows from the fact that no edge without heads (i.e. $-$) is ever added. Assume for the second statement that $\mathcal{G}$ satisfies (3.3). Let $M = V\setminus O$. Assume $\alpha \leftrightarrow_{\mathcal{M}} \beta$. By definition of the latent projection, we can find an endpoint-identical walk between α and β in $\mathcal{G}$ with no colliders and such that all non-endpoint nodes are in M. Either this walk has a bidirected edge at α in which case $\alpha \leftrightarrow_{\mathcal{G}} \alpha$ by (3.3) and therefore also $\alpha \leftrightarrow_{\mathcal{M}} \alpha$. Otherwise, there is a directed edge from some node $\gamma \in M$ such that $\gamma \to_{\mathcal{G}} \alpha$. Then the walk $\alpha \leftarrow \gamma \to \alpha$ is present in $\mathcal{G}$ and therefore $\alpha \leftrightarrow_{\mathcal{M}} \alpha$ because $\mathcal{M}$ is a latent projection. $\square$

PROOF OF PROPOSITION 3.10. Assume first that α has no loops. In this case, there are no bidirected edges between α and any node, and therefore the edges that have a head at α have a tail at the previous node. Any nontrivial walk between α and α is therefore blocked by $V \setminus \{\alpha\}$. Conversely, if α has a loop, then $\alpha \ast\!\!\to \alpha$ is a μ-connecting walk given $V \setminus \{\alpha\}$. $\square$

PROOF OF THEOREM 3.12. Let $M = V \setminus O$. Let first ω be a μ-connecting walk from $\alpha \in A$ to $\beta \in B$ given C in $\mathcal{G}$. Using Proposition 3.3, we can find a μ-connecting walk from $\alpha \in A$ to $\beta \in B$ given C in $\mathcal{G}$ such that all colliders are in C. Denote this walk by $\bar{\omega}$. Every node, m, on $\bar{\omega}$ which is in M is on a subwalk of $\bar{\omega}$, $\delta_1 \sim \ldots \sim m \sim \ldots \sim \delta_2$, such that $\delta_1, \delta_2 \in O$ and all other nodes on the subwalk are in M. There are no colliders on this subwalk and therefore there is an endpoint-identical edge $\delta_1 \sim \delta_2$ in $\mathcal{M}$. Substituting all such subwalks with their corresponding endpoint-identical edges gives a μ-connecting walk in $\mathcal{M}$.

On the other hand, let ω be a μ-connecting walk from A to B given C in $\mathcal{M}$. Consider some edge in ω which is not in $\mathcal{G}$. In $\mathcal{G}$ there is an endpoint-identical walk with no colliders and no non-endpoint nodes in C. Substituting each of these edges with such an endpoint-identical walk gives a μ-connecting walk in $\mathcal{G}$ using Proposition 3.11. $\square$

PROOF OF PROPOSITION 3.13. We first note that in Algorithm 1 adding an edge will never remove any triroutes. Therefore, Algorithm 1 returns the same output regardless of the order in which the algorithm adds edges.

Let $\mathcal{M}$ denote the output of Algorithm 1 which is clearly a DMG. The graphs $\mathcal{M}$ and $m(\mathcal{G}, O)$ have the same node set, thus it suffices to show that also the edge sets are equal. Assume first $\alpha \overset{e}{\sim}_{m(\mathcal{G},O)} \beta$. Then there exist an endpoint-identical walk in $\mathcal{G}$ that contains no colliders and such that all the non-endpoint nodes are in $M = V \setminus O$, $\alpha \sim \gamma_1 \sim \ldots \sim \gamma_n \sim \gamma_{n+1} = \beta$. Let e_l be the edge between α and γ_l which is endpoint-identical to the subwalk from α to γ_l. If e_l is present in $\mathcal{M}_k$ at some point during Algorithm 1, then edge e_{l+1} will also be added before the algorithm terminates, $l = 1, \ldots, n$. We see that e_1 is in $\mathcal{G}$, and this means that e is also present in $\mathcal{M}$.

On the other hand, assume that some edge e is in $\mathcal{M}$. If e is not in $\mathcal{G}$, then we can find a noncolliding, endpoint-identical triroute in the graph $\mathcal{M}_k$ (k has the value that it takes when the algorithm terminates) such that the noncollider is in M. By repeatedly using this argument, we can from any edge, e, in $\mathcal{M}$ construct an endpoint-identical walk in $\mathcal{G}$ that contains no colliders and such that every non-endpoint node is in M, and therefore e is also present in $m(\mathcal{G}, O)$. $\square$

Proof of Proposition 4.5. Let

$$\alpha \ast\!\!\rightarrow \gamma_1 \leftrightarrow \ldots \leftrightarrow \gamma_n \leftrightarrow \beta$$

be the inducing path, ν. Let γ_{n+1} denote β. If ν has length one, then it is directed or bidirected and itself a μ-connecting path/cycle regardless of C. Assume instead that the length of ν is strictly larger than one, and assume also first that $\alpha \neq \beta$. Let k be the maximal index in $\{1, \ldots, n\}$ such that there exists an open walk from α to γ_k given C which does not contain β and only contains α once. There is a μ-connecting walk from α to $\gamma_1 \neq \beta$ given C and therefore k is always well-defined.

Let ω be the open walk from α to γ_k. If $\gamma_k \in An(C)$, then the composition of ω with the edge $\gamma_k \leftrightarrow \gamma_{k+1}$ is open from α to γ_{k+1} given C. By maximality of k, we must have $k = n$, and the composition is therefore an open walk from α to β on which β only occurs once. We can reduce this to a μ-connecting path using arguments like those in the proof of Proposition 3.5. Assume instead that $\gamma_k \notin An(C)$. There is a directed path from γ_k to α or to β. Let π denote the subpath from γ_k to the first occurrence of either α or β on this directed path. If β occurs first, then the composition of ω with π gives an open walk from α to β. There is a head at β when moving from α to β and therefore the walk can be reduced to a μ-connecting path from α to β using the arguments in the proof of Proposition 3.5. If α occurs first, then the composition of π^{-1} and the edge $\gamma_k \leftrightarrow \gamma_{k+1}$ gives a μ-connecting walk and it follows that $k = n$ by maximality of k. This walk is a μ-connecting path.

To argue that the open path is endpoint-identical if ν is directed or bidirected, let instead k be the maximal index such that there exists a μ-connecting walk from α to γ_k with a head/tail at α. Using the same argument as above, we see that the μ-connecting path will be endpoint-identical to ν in this case. In the directed case, note that in the case $\gamma_k \notin An(C)$ one can find a directed path form γ_k to β, and if α occurs on this path one can simply choose the subpath from α to β.

In the case $\alpha = \beta$, analogous arguments can be made by assuming that k is the maximal index such that there exists a μ-inducing path from α to γ_k given C such that $\beta = \alpha$ only occurs once. $\square$

Proof of Propositions 4.7 and 4.8. For both propositions it suffices to argue that if there is a μ-connecting walk in the larger graph, then we can also find a μ-connecting walk in the smaller graph. Using Proposition 4.5 we can find endpoint-identical walks that are open given $C \setminus \{\alpha\}$ and replacing $\alpha \ast\!\!\rightarrow \beta$ with such a walk will give a walk which is open given C. For Proposition 4.8 one should note that adding the edge respects the ancestry of the nodes due to transitivity. $\square$

Proof of Proposition 4.10. Assume there is no inducing path from α to β and let ω be some walk from α to β with a head at β. Note that ω must have length at least 2.

$$\alpha = \gamma_0 \overset{e_0}{\sim} \gamma_1 \overset{e_1}{\sim} \dots \overset{e_{m-1}}{\sim} \gamma_m \overset{e_m}{\ast\!\!\rightarrow} \beta.$$

There must exist an $i \in \{0, 1, \dots, m\}$ such that γ_i is not directedly collider-connected to β along ω or such that $\gamma_i \notin An(\alpha, \beta)$. Let j be the largest such index. Note first that γ_m is always directedly collider-connected to β along ω and γ_0 is always in $An(\alpha, \beta)$. If $j \neq m$ and γ_j is not directedly collider-connected to β along ω, then γ_{j+1} is a noncollider and ω is closed in $\gamma_{j+1} \in D(\alpha, \beta)$ (note that $\alpha = \gamma_{j+1}$ is impossible as there would then be an inducing path from α to β). If $j \neq 0$ and $\gamma_j \notin An(\alpha, \beta)$ then there is some $k \in \{1, \dots, j\}$ such that γ_k is a collider and $\gamma_k \notin An(\alpha, \beta)$ and ω is therefore closed in this collider. $\square$

Proof of Proposition 5.3. We verify that (gs1)–(gs3) hold.
(gs1) The edge $\alpha \leftrightarrow \beta$ constitutes an inducing path in both directions.
(gs2-3) Let $\gamma \in V, C \subseteq V$ such that $\beta \in C$, and assume that there is a μ-connecting walk from γ to β given C in $\mathcal{G}$. This walk has a head at β and composing the walk with $\alpha \leftrightarrow \beta$ creates an μ-connecting walk from γ to α given C. $\square$

Proof of Lemma 5.4. Any μ-connecting walk in $\mathcal{G}$ is also present and μ-connecting in $\mathcal{G}^+$, hence $\mathcal{I}(\mathcal{G}^+) \subseteq \mathcal{I}(\mathcal{G})$.

Assume $\gamma, \delta \in V, C \subseteq V$ and assume that ρ is a μ-connecting route from γ to δ given C in $\mathcal{G}^+$. Let e denote the edge $\alpha \leftrightarrow \beta$. Using (gs1), there exist an inducing path from α to β in $\mathcal{G}$ and one from β to α. Denote these by ν_1 and ν_2. If e is not in ρ, then ρ is also in $\mathcal{G}$ and μ-connecting as the addition of the bidirected edge does not change the ancestry of $\mathcal{G}$.

If e occurs twice in ρ then it contains a subroute $\alpha \overset{e}{\leftrightarrow} \beta \overset{e}{\leftrightarrow} \alpha$ and $\alpha = \delta$ (or with the roles interchanged). Either one can find a μ-connecting subroute of ρ with no occurrences of e or $\alpha \notin C$. If $\beta \in C$, then compose the subroute of ρ from γ to the first occurrence of α (which is either trivial or can be assumed to have a tail at α) with the ν_1-induced open walk from α to β

using Proposition 4.5. This is a μ-connecting walk in $\mathcal{G}$ from γ to β and using (gs2) the result follows. If $\beta \notin C$, then the result follows from composing the subroute from γ to α with the ν_1-induced open walk from α to β and the ν_2-inducing open walk from β to α.

If e only occurs once on ρ, consider first a ρ of the form

$$\underbrace{\gamma \sim \ldots \sim \alpha}_{\rho_1} \overset{e}{\leftrightarrow} \underbrace{\beta \sim \ldots \ast\!\!\rightarrow \delta}_{\rho_2}.$$

Assume first that $\alpha \notin C$. Let π denote the ν_1-induced open walk from α to β and note that π has a head at β. If $\gamma = \alpha$ then π composed with ρ_2 is a μ-connecting walk from γ to δ in $\mathcal{G}$. If $\gamma \neq \alpha$ we can just replace e with π, and the resulting composition of the walks ρ_1, π and ρ_2 is a μ-connecting walk from γ to δ in $\mathcal{G}$. If instead $\alpha \in C$, then $\gamma \neq \alpha$ and α is a collider on ρ, and ρ_1 thus has a head at α and is μ-connecting from γ to α given C in $\mathcal{G}$. Using (gs3) we can find a μ-connecting walk from γ to β given C in $\mathcal{G}$. Composing this with ρ_2 gives a μ-connecting walk from γ to δ given C in $\mathcal{G}$.

If ρ instead has the form

$$\gamma \sim \ldots \sim \beta \overset{e}{\leftrightarrow} \alpha \sim \ldots \ast\!\!\rightarrow \delta,$$

a similar argument using (gs2) applies. In conclusion, $\mathcal{I}(\mathcal{G}) \subseteq \mathcal{I}(\mathcal{G}^+)$. □

Proof of Proposition 5.7. We verify that (gp1)–(gp4) hold.

(gp1) $\alpha \to \beta$ constitutes an inducing path from α to β.

(gp2) Let ω be a μ-connecting walk from γ to α given C, $\alpha \notin C$. Then ω composed with $\alpha \to \beta$ is μ-connecting from γ to β given C.

(gp3) Let ω_1 be a μ-connecting walk from γ to β given C, $\alpha \notin C, \beta \in C$, and let ω_2 be a μ-connecting walk from α to δ given C. The composition of ω_1, $\alpha \to \beta$, and ω_2 is μ-connecting.

(gp4) Let ω be a μ-connecting walk from β to γ given $C \cup \{\alpha\}$, $\alpha \notin C$. If this walk is closed given C, then there exists a collider on ω, which is an ancestor of α and not in $An(C)$. Let δ be the collider on ω with this property which is the closest to γ. Then we can find a directed and open path from δ to β and composing the inverse of this with the subwalk of ω from δ to γ gives us a connecting walk. □

Proof of Lemma 5.8. As $An_{\mathcal{G}}(C) \subseteq An_{\mathcal{G}^+}(C)$ for all $C \subseteq V$, any μ-connecting path in $\mathcal{G}$ is also μ-connecting in $\mathcal{G}^+$, and it therefore follows that $\mathcal{I}(\mathcal{G}^+) \subseteq \mathcal{I}(\mathcal{G})$.

We will prove the other inclusion by considering a μ-connecting walk from γ to δ given C in $\mathcal{G}^+$ and argue that we can find another μ-connecting walk in $\mathcal{G}^+$ that fits into cases (a) or (b) below. In both cases, we will use the potential parents properties to argue that there is also a μ-connecting walk from γ to δ given C in $\mathcal{G}$. Let e denote the edge $\alpha \to \beta$.

Let ν denote the inducing path from α to β in $\mathcal{G}$ which we know to exist by (gp1) and Proposition 4.10. Say we have a μ-connecting walk in $\mathcal{G}^+$, ω, from γ to δ given C. There can be two reasons why ω is not μ-connecting in $\mathcal{G}$: 1) e is in ω, 2) there exist colliders, $c_1, \ldots, c_k$, on ω, which are in $An_{\mathcal{G}^+}(C)$ but not in $An_{\mathcal{G}}(C)$. We will in this proof call such colliders *newly closed*. If there exists a newly closed collider on ω, c_i, then there exists in $\mathcal{G}$ a directed path from c_i to α on which no node is in C, and furthermore $\alpha \notin C$. Note that this path does not contain β, and the existence of a newly closed collider implies that $\beta \in An_{\mathcal{G}}(C)$.

Using Proposition 3.5, we can find a route, ρ, in $\mathcal{G}^+$ from γ to δ, which is μ-connecting in $\mathcal{G}^+$. Assume first that e occurs at most once on ρ. If there are newly closed colliders on ρ, we will argue that we can find a μ-connecting walk in $\mathcal{G}^+$ with no newly closed colliders and such that e occurs at most once. Assume that $c_1, \ldots, c_k$ are newly closed colliders, ordered by their occurrences on the route ρ. We allow for $k = 1$, in which case $c_1 = c_k$. We will divide the argument into three cases, and we use in all three cases that a μ-connecting walk in $\mathcal{G}$ is also present in $\mathcal{G}^+$ and has no newly closed colliders nor occurrences of e. We also use that $\alpha \notin C$ when applying (gp2).

(i) e is between γ and c_1 on ρ.
Consider the subwalk of ρ from γ to the first occurrence of α. If this subwalk has a tail at α (or is trivial) then we can compose it with the inverse of the path from c_k to α and the subwalk from c_k to δ. This walk is open. If there is a head at α, then using (gp2) we can find a μ-connecting walk from γ to β in $\mathcal{G}$, compose it with e, the inverse of the path from c_k to α and the subwalk from c_k to δ. This is open as $\beta \in An_{\mathcal{G}}(C)$ and $\alpha \notin C$ whenever there exist newly closed colliders.

(ii) e is between c_k and δ on ρ.
Consider the subwalk of ρ from γ to c_1, and compose it with the directed path from c_1 to α. This is μ-connecting in $\mathcal{G}$ and using (gp2) we can find a μ-connecting walk in $\mathcal{G}$ from γ to β. Composing this walk with the subwalk of ρ from β to δ gives a μ-connecting walk from γ to δ, noting that $\beta \in An_{\mathcal{G}}(C)$.

(iii) e is between c_1 and c_k on ρ or not on ρ at all.
Composing the subwalk from γ to c_1 with the directed path from c_1 to α gives a μ-connecting walk from γ to α given C in $\mathcal{G}$, and by (gp2) we can find a μ-connecting walk from γ to β in $\mathcal{G}$, thus there are no newly closed colliders on this walk and it does not contain e. Composing it

with e, the directed path from c_k to α and the subwalk from c_k to δ gives a μ-connecting walk in $\mathcal{G}^+$.

In all cases (i), (ii), and (iii) we have argued that there exists a μ-connecting walk from γ to δ in $\mathcal{G}^+$ that contains no newly closed colliders and that contains e at most once. Denote this walk by $\tilde{\omega}$. If $\tilde{\omega}$ does not contain e at all, then we are done. Otherwise, two cases remain, depending on the orientation of e in the μ-connecting walk $\tilde{\omega}$:

(a) Assume first we have a walk of the form

$$\gamma \sim \ldots \overset{e_\alpha}{\sim} \alpha \to \beta \sim \ldots \ast\!\!\to \delta,$$

If there is a tail on e_α at α, or if $\gamma = \alpha$, then we can substitute e with the open path between α and β induced by ν and obtain an open walk. Otherwise, assume a head on e_α at α. $\tilde{\omega}$ is μ-connecting in $\mathcal{G}^+$ and therefore $\alpha \notin C$. Using (gp2), there exists a μ-connecting walk from γ to β, and composing this walk with the (potentially trivial) subwalk from β to δ gives a μ-connecting walk from γ to δ given C in $\mathcal{G}$.

(b) Consider instead a walk of the form

$$\gamma \sim \ldots \overset{e_\beta}{\sim} \beta \leftarrow \alpha \sim \ldots \ast\!\!\to \delta.$$

If there is a head on e_β at β, β is a collider. If $\beta \in C$, then (gp3) directly gives a μ-connecting walk from γ to δ given C in $\mathcal{G}$. If instead $\beta \in An_{\mathcal{G}^+}(C) \setminus C$ then we can find a directed path, π, in $\mathcal{G}^+$ from β to $\varepsilon \in C$. The edge e is not present on π and therefore we can compose the subwalk from γ to β with π, π^{-1}, and the subwalk from β to δ to obtain an open walk from γ to δ without any newly closed colliders, only one occurrence of e, and such that there is a tail at β just before the occurence of e.

We have reduced this case to walks, $\tilde{\omega}$, of the form

$$\underbrace{\gamma \sim \ldots \leftarrow \beta}_{\tilde{\omega}_1} \leftarrow \underbrace{\alpha \sim \ldots \ast\!\!\to \delta}_{\tilde{\omega}_2},$$

where $\tilde{\omega}_1$ is potentially trivial. Let $\bar{\pi}$ denote the ν-induced open path or cycle from α to β in $\mathcal{G}$. Using Proposition 3.5 there is a μ-connecting route, $\bar{\rho}$, from α to δ given C in $\mathcal{G}$. If there is a tail at α on $\bar{\rho}$ or on $\bar{\pi}$ then the composition of $\tilde{\omega}_1$, $\bar{\pi}$ and $\bar{\rho}$ is μ-connecting. Otherwise, if $\alpha \neq \beta$, the composition of $\overline{\pi}$ and $\bar{\rho}$ is a μ-connecting walk from β to δ given $C \cup \{\alpha\}$ in $\mathcal{G}$ as α does not occur as a noncollider on this composition. Using (gp4) there is also one given C. As there is a tail at β on $\tilde{\omega}$ we can compose $\tilde{\omega}_1$ with this walk to obtain an open walk from γ to δ given C in $\mathcal{G}$. If $\alpha = \beta$ the composition of $\tilde{\omega}_1$ with $\tilde{\omega}_2$ is an open walk from γ to δ given C in $\mathcal{G}$.

Assume finally that e occurs twice on ρ. In this case ρ contains a subroute $\beta \overset{e}{\leftarrow} \alpha \overset{e}{\rightarrow} \beta$ and $\beta = \delta$. In this case $\alpha \notin C$. If there are any newly closed colliders, consider the one closest to γ, c. The subroute of ρ from γ to c composed with the directed path from c to α gives a μ-connecting path and (gp2) gives the result. Else if there is a head at α on the ν-induced open walk then (gp2) again gives the result. Otherwise, compose the subroute from γ to the first β, the inverse of the ν-induced open walk, and the ν-induced open walk to obtain an open walk in $\mathcal{G}$ from γ to $\beta = \delta$. □

Proof of Theorem 5.9. Propositions 5.3 and 5.7 show that $\mathcal{N}$ is in fact a supergraph of $\mathcal{G}$, and as E^m only depends on the independence model, it also shows that $\mathcal{N}$ is a supergraph of any element in $[\mathcal{G}]$. We can sequentially add the edges that are in $\mathcal{N}$ but not in $\mathcal{G}$, and Lemmas 5.4 and 5.8 show that this is done Markov equivalently, meaning that $\mathcal{N} \in [\mathcal{G}]$. □

Lemma F.1. Let $\alpha, \beta \in V$. If there is a directed edge, e, from α to β, and a unidirected inducing path from α to β of length at least two in $\mathcal{N}$, then there is a directed inducing path from α to β in $\mathcal{N} - e$.

Proof of Lemma F.1. Let ν denote the unidirected inducing path and $\gamma_1, \ldots, \gamma_n$ the non-endpoint nodes of ν. Then $\gamma_i \in An_{\mathcal{N}}(\{\alpha, \beta\})$ and also $\gamma_i \in An_{\mathcal{N}}(\beta)$ due to the directed edge from α to β. It follows that either $\gamma_i \in An_{\mathcal{N}}(\alpha)$ or $\gamma_i \in An_{(\mathcal{N}-e)}(\beta)$. If $\gamma_i \in An_{\mathcal{N}}(\alpha)$, let e_i denote the directed edge from γ_i to β, and let $\mathcal{N}^+ = (V, F \cup \{e_i\})$. We will argue that $\mathcal{N} = \mathcal{N}^+$ using the maximality of $\mathcal{N}$. Note first that the edge does not change the ancestry of the graph in the sense that $An_{\mathcal{N}}(\gamma) = An_{\mathcal{N}^+}(\gamma)$ for all $\gamma \in V$. Note also that there is a bidirected inducing path between γ_i and β in $\mathcal{N}$, and therefore $\gamma_i \leftrightarrow_{\mathcal{N}} \beta$. Assume that e_i is in a μ-connecting path in $\mathcal{N}^+$. There is a directed path from γ_i to α in $\mathcal{N}$ and therefore e_i can either be substituted with $\gamma_i \rightarrow \alpha_i \rightarrow \ldots \rightarrow \alpha_k \rightarrow \alpha \rightarrow \beta$ (if $\alpha_1, \ldots, \alpha_k, \alpha \notin C$), or with $\gamma_i \leftrightarrow \beta$ (otherwise), and we see that $\mathcal{I}(\mathcal{N}) = \mathcal{I}(\mathcal{N}^+)$. By maximality of $\mathcal{N}$ we have that $\mathcal{N} = \mathcal{N}^+$ which implies that $e_i \in F$. Thus $\gamma_i \in An_{(\mathcal{N}-e)}(\beta)$. This shows that ν is also a directed inducing path in $\mathcal{N} - e$. □

Lemma F.2. Let edges $\alpha \rightarrow \beta$, $\beta \rightarrow \alpha$ and $\alpha \leftrightarrow \beta$ be denoted by e_1, e_2, e_3, respectively. If $e_1, e_3 \in F$, then $\mathcal{N} - e_1 \in [\mathcal{N}]$. If $e_1, e_2, e_3 \in F$, then $\mathcal{N} - e_3 \in [\mathcal{N}]$.

PROOF OF LEMMA F.2. Note that if edges $\gamma \ast\!\!\rightarrow \alpha$, $\alpha \leftrightarrow \beta$, and $\alpha \rightarrow \beta$ are present in a maximal DMG, then so is $\gamma \ast\!\!\rightarrow \beta$ by Propositions 4.7 and 4.8. Assume $e_1, e_3 \in E$. Using the above observation, note that every vertex that is a parent of α in $\mathcal{N}$ is also a parent of β, thus $An_{\mathcal{N}}(\delta) \setminus \{\alpha\} = An_{(\mathcal{N}-e_1)}(\delta) \setminus \{\alpha\}$ for all $\delta \in V$. Consider a μ-connecting walk, ω, in $\mathcal{N}$ given C. Any collider different from α on this walk is in $An_{(\mathcal{N}-e_1)}(C)$. If $\alpha \notin An_{(\mathcal{N}-e_1)}(C)$ is a collider, then we can substitute the subwalk $\gamma_1 \ast\!\!\rightarrow \alpha \leftarrow\!\!\ast \gamma_2$ with $\gamma_1 \ast\!\!\rightarrow \beta \leftarrow\!\!\ast \gamma_2$. If e_1 is the first edge on ω and α the first node, then just substitute e_1 with e_3. Else, we need to consider two cases: in the first case there is a subwalk $\gamma \ast\!\!\rightarrow \alpha \rightarrow \beta$ (or $\beta \leftarrow \alpha \leftarrow\!\!\ast \gamma$) and therefore an edge $\gamma \ast\!\!\rightarrow \beta$ in $\mathcal{N} - e_1$ if $\gamma \neq \alpha$. If $\gamma = \alpha$, we can simply remove the loop, replacing e_1 with e_3 if γ was the final node on ω. In the second case, there is a subwalk $\gamma \leftarrow \alpha \rightarrow \beta$ (or $\beta \leftarrow \alpha \rightarrow \gamma$), and we can substitute e_1 with e_3 if $\beta \neq \gamma$. If $\beta = \gamma$, then we can substitute $\beta \leftarrow \alpha \rightarrow \beta$ with $\beta \leftrightarrow \beta$.

The proof of the other statement is similar. □

PROOF OF PROPOSITION 5.11. One implication is immediate by contraposition: if $\alpha \notin u(\beta, \mathcal{I}(\mathcal{N} - e))$, then $\mathcal{N} - e \notin [\mathcal{N}]$.

Assume $\alpha \in u(\beta, \mathcal{I}(\mathcal{N} - e))$. There exists an inducing path, ν, from α to β in $\mathcal{N} - e$. If ν is directed, then the conclusion follows from Proposition 4.8. If ν is unidirected and of length one, then it is also directed. If it is unidirected and has length at least two, it follows from Lemma F.1 that there also exists a directed inducing path in $\mathcal{N} - e$. Proposition 4.8 finishes the argument. Assume that ν is bidirected. Then $\alpha \leftrightarrow_{\mathcal{N}} \beta$ due to maximality and Proposition 4.7. Lemma F.2 gives the result. □

PROOF OF PROPOSITION 5.12. One implication follows by contraposition. Assume instead that $\alpha \in u(\beta, \mathcal{I}(\mathcal{N} - e))$ and $\beta \in u(\alpha, \mathcal{I}(\mathcal{N} - e))$. Then there is an inducing path from α to β and one from β to α in $\mathcal{N} - e$. Denote these by ν_1 and ν_2. If one of them is bidirected, then the conclusion follows. Assume instead that none of them are bidirected and assume first that both are a single edge. The conclusion then follows using Lemma F.2.

Assume now that ν_1 or ν_2 is an inducing path of length at least 2. Say that $\beta \rightarrow \gamma_1 \leftrightarrow \ldots \leftrightarrow \gamma_m \leftrightarrow \alpha$ is an inducing path. If ν_1 is the inducing path $\alpha \rightarrow_{\mathcal{N}} \beta$ of length one, then there is also a bidirected inducing path between γ_1 and β in $\mathcal{N}$, and there will also be a bidirected inducing path in $\mathcal{N} - e$ between α and β. If instead ν_1 is the inducing path $\alpha \rightarrow \phi_1 \leftrightarrow \ldots \leftrightarrow \phi_k \leftrightarrow \beta$ then $\gamma_1 \leftrightarrow_{\mathcal{N}} \phi_1$. In this case $\alpha \leftrightarrow \gamma_m \ldots \gamma_1 \leftrightarrow \phi_1 \ldots \phi_k \leftrightarrow \beta$ can be trimmed down to a bidirected inducing path in $\mathcal{N} - e$. □

REFERENCES

[1] DANKS, D. and PLIS, S. (2013). Learning causal structure from undersampled time series. In *JMLR: Workshop and Conference Proceedings (NIPS Workshop on Causality)*.

[2] DIDELEZ, V. (2000). Graphical models for event history analysis based on local independence, PhD thesis, Universität Dortmund.

[3] DIDELEZ, V. (2008). Graphical models for marked point processes based on local independence. *Journal of the Royal Statistical Society, Series B* **70** 245-264.

[4] EICHLER, M. and DIDELEZ, V. (2010). On Granger causality and the effect of interventions in time series. *Lifetime Data Analysis* **16** 3-32.

[5] HYTTINEN, A., PLIS, S., JÄRVISALO, M., EBERHARDT, F. and DANKS, D. (2016). Causal discovery from subsampled time series data by constraint optimization. In *Proceedings of the Eighth International Conference on Probabilistic Graphical Models* **52** 216-227.

[6] LAURITZEN, S. (1996). *Graphical models*. Oxford: Clarendon.

[7] LAURITZEN, S. and SADEGHI, K. (2018). Unifying Markov properties for graphical models. *Annals of Statistics* **46** 2251-2278.

[8] MEEK, C. (1995). Strong completeness and faithfulness in Bayesian networks. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence (UAI1995)*.

[9] MEEK, C. (2014). Toward learning graphical and causal process models. In *Proceedings of the UAI 2014 Workshop Causal Inference: Learning and Prediction*.

[10] RICHARDSON, T. (2003). Markov properties for acyclic directed mixed graphs. *Scandinavian Journal of Statistics* **30** 145-157.

[11] RICHARDSON, T. and SPIRTES, P. (2002). Ancestral graph Markov models. *The Annals of Statistics* **30** 962-1030.

[12] ROGERS, L. C. G. and WILLIAMS, D. (2000). *Diffusions, Markov processes, and martingales. Cambridge Mathematical Library* **2**. Cambridge University Press, Cambridge Itô calculus, Reprint of the second (1994) edition.

[13] SOKOL, A. and HANSEN, N. R. (2014). Causal interpretation of stochastic differential equations. *Electronic Journal of Probability* **19** 1-24.

DEPARTMENT OF MATHEMATICAL SCIENCES
UNIVERSITY OF COPENHAGEN
UNIVERSITETSPARKEN 5
2100 COPENHAGEN
DENMARK
E-MAIL: swengel@math.ku.dk

DEPARTMENT OF MATHEMATICAL SCIENCES
UNIVERSITY OF COPENHAGEN
UNIVERSITETSPARKEN 5
2100 COPENHAGEN
DENMARK
E-MAIL: niels.r.hansen@math.ku.dk