Local Dependence Graphs for Discrete Time Processes

Author names withheld

Editor: Under Review for CLeaR 2023

Abstract

Local dependence graphs for discrete time processes encapsulate information concerning the dependence relationships between the past of the multidimensional process and its present state and as such can represent feedback loops. Despite the apparent simplicity of the discrete time setting, some natural questions relating the conditional (in)dependence statements in the process to separation properties of the underlying local dependence graph seem to be missing from the literature. In this paper we give graphical characteristics for two kinds of conditional independencies: those occurring in Markov chains under the stationary regime and independencies between the past of one subprocess and the future of another given the past of the third subprocess.

Keywords: Local dependence graphs, dynamic Bayesian networks, composable Markov processes

1. Introduction

Probabilistic models of causality were originated in 1960s by Granger (1969) in the context of time series. Granger's definition explicitly and essentially involved ordering of events in time, assuming that a cause always precedes an effect. The pioneering work of Pearl (1985) dating from 1980s introduced graph-theoretic representation of causal relations. The Pearl's theory did not involve time. In its original form this theory was restricted to directed acyclic graphs (DAGs), and thus did not encompass the phenomena of feedback in cause-effect dependencies. Later work initiated in Pearl and Dechter (1996) generalised graphical models of causality allowing for cycles in directed graphs. Models based on "structural equations" have been intensively examined and extended. The monograph Peters et al. (2017) is an excellent overview. Recent developments, with focus on mixed graphs and marginalisation, can be found in Bongers et al. (2021). However, this approach leads to certain difficulties and paradoxes (c.f. Neal (2000)), if the cycles are present in the underlying graph. A priori it is not clear whether a given set of structural equations has a unique solution and this makes the task of building models rather hard.

On the other hand, a class of "composable Markov processes" (CMPs) introduced as early as in 1970 by Schweder (1970) provides natural and flexible examples of causality models which include time and use directed cyclic graphs (DCGs) to represent "local" (causal) dependencies. Research on CMPs was pursued by Nodelman et al. (2002) (who independently introduced these processes under the name of CTBNs) and by Didelez (2007). Local dependence graphs have been also defined and examined for other classes of structured stochastic processes, such as point processes (Didelez, 2008; Mogensen and Hansen, 2022) and diffusion processes (Mogensen and Hansen, 2020).

Presumably the simplest causal temporal models are those based on discrete time processes. Local dependence graphs for such processes were considered e.g. in Eichler and Didelez (2007). In a

© 2023 .

sense, investigation of a discrete time graphical model can be reduced to considering a space-time graph representing a dynamic Bayesian network (DBN). The structure of conditional independencies in such a model is widely regarded as perfectly understood. Nonetheless, in our paper we point out several facts that have been overlooked in the literature and we fill in certain gaps. We prove "Markov properties of the graph of local dependence" concerning some natural questions. Put differently, we translate probabilistic statements about conditional (in)dependence into separability properties of the underlying graph. In particular we examine dependence of the future of some variables on the past of other variables (Theorem 3.9) and conditional dependence of the whole sub-processes under the stationary regime (Theorem 3.4) and under the assumption of independent initialisation (Theorem 3.2). We conjecture that most our results can be generalised to the case of continuous time processes (under a suitable definition of local dependencies graph). In fact that we can prove that Theorem 3.2 remains true for also for CTBNs. Generalisations of Theorem 3.9 and Theorem 3.4 to continuous time setting remain open questions, as far as we know.

2. The setup

Let $\mathbf{X} = (X_v(t), v \in \mathcal{V}, t = 0, 1, \ldots)$ be a multivariate discrete time stochastic process with components indexed by elements of a finite set $\mathcal{V}$. Assume that random variables at node v, that is $X_v(t)$, take values in a measure space $\mathbb{S}_v$ (in most examples $\mathbb{S}_v$ is either $\mathbb{R}$ equipped with the Lebesgue measure or a finite set with the counting measure); moreover let $\mathbb{S} = \prod_{v\in\mathcal{V}} \mathbb{S}_v$. We will use the notations $\mathbf{X}_v = (X_v(t), t = 0, 1 \ldots)$ and $X(t) = (x_v(t), v \in \mathcal{V})$. Note that processes are denoted in bold, contrary to random variables with values in $\mathbb{S}$. For $\mathcal{C} \subseteq \mathcal{V}$ we will write $\mathbf{X}_\mathcal{C} = (\mathbf{X}_v : v \in \mathcal{C})$ and $X_\mathcal{C}(t) = (X_v(t) : v \in \mathcal{C})$.

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a directed graph with possible cycles. We write $v \to w$ or, equivalently, $w \leftarrow v$ if $(v, w) \in \mathcal{E}$. For $v \in \mathcal{V}$ we put $\mathrm{pa}(v) = \{w : w \to v\}$. Assume that all the vertices of $\mathcal{G}$ have self-loops ($v \to v$ for every $v \in \mathcal{V}$).[1]

Our standing assumption is the following. For arbitrary measurable subsets E_v of $\mathbb{S}_v$,

$$\mathbb{P}\left(\forall_{v\in\mathcal{V}}\, X_v(t) \in E_v \middle| X(s), s < t\right) = \prod_v \mathbb{P}(X_v(t) \in E_v | X_{\mathrm{pa}(v)}(s), s < t). \tag{1}$$

Thus, we assume that the "mechanism of generating" $X_v(t)$, given the past of the whole process, acts independently for every node v and depends only on parents of this node. This can be expressed as the following two conditional independence statements.

$$(X_v(t), v \in \mathcal{V}) \text{ are conditionally independent given } (X(s) : s < t), \tag{2}$$

$$X_v(t) \text{ is conditionally independent of } (X_{\mathcal{V}\setminus\mathrm{pa}(v)}(s) : s < t) \text{ given } (X_{\mathrm{pa}(v)}(s) : s < t). \tag{3}$$

Condition (3) formalises the relation between the graph $\mathcal{G}$ and the process $\mathbf{X}$. It says that only variables $X_w(s)$ with $w \to v$ and $s < t$ have potential causal influence on $X_v(t)$. Condition (2) excludes "contemporanuous local dependence" and basically boils down to the assumption that our model contains all relevant "confounding variables".[2]

1. This assumption is made mostly to simplify notation and is not essential so far as only discrete time processes are concerned. On the other hand it seems to be natural and needed for generalisations to continuous time processes.
2. If (2) is omitted then the relevant graph must include, apart from arrows, also another type of edges (c.f. undirected "dashed edges" in Eichler and Didelez (2007), "bidirected edges" and "blunt edges" in Mogensen and Hansen (2020)).

Definition 2.1 *If* (1) *holds then we say that* $\mathcal{G}$ *is a* ***local conditional dependence graph*** *for* $\mathbf{X}$.[3]

3. Conditional Independence Properties

We are going to examine conditional independence statements about process $\mathbf{X}$ and to find "equivalent" statements in terms of graph $\mathcal{G}$. Let us first note that we understand "equivalence" in the following (weak) sense:

- Property γ of $\mathcal{G}$ implies property ξ of $\mathbf{X}$ for every $\mathbf{X}$ which satisfies (1) wrt. $\mathcal{G}$.
- If γ is not true for $\mathcal{G}$ then there exists $\mathbf{X}$ which satisfies (1) wrt. $\mathcal{G}$, for which ξ fails.

We will have to consider several types of separability relations in graph $\mathcal{G}$. Some of them are well-known and some other presumably new. To begin with, we define a trail between $u \in \mathcal{V}$ and $w \in \mathcal{V}$ as a sequence

$$u = v_0, e_1, v_1, e_2, v_2, e_3, \ldots, e_{n-1}, v_{n-1}, e_n, v_n = w,$$

where $v_0, v_1, \ldots, v_n$ are *distinct* nodes and $e_1, \ldots, e_n$ are edges, where $e_i = (v_{i-1} \to v_i)$ or $e_i = (v_{i-1} \leftarrow v_i)$.[4] Let v_i be a non-end node in the trail, that is $i \neq 0$ and $i \neq n$. We say that

There is a *chain* connexion at v_i if we have $v_{i-1} \to v_i \to v_{i+1}$ or $v_{i-1} \leftarrow v_i \leftarrow v_{i+1}$.

There is a *fork* connexion at v_i if we have $v_{i-1} \leftarrow v_i \to v_{i+1}$.

There is a *collider* connexion at v_i if we have $v_{i-1} \to v_i \leftarrow v_{i+1}$.

A *directed path* from u to w is a trail such that all arrows are directed to the right, i.e. $e_i = (v_{i-1} \to v_i)$. We define the set of *ancestors* and the set of *descendants* as follows:

$$\mathrm{an}(v) = \{v\} \cup \{w : \text{ there exists a directed path from } w \text{ to } v\},$$
$$\mathrm{de}(v) = \{v\} \cup \{w : \text{ there exists a directed path from } v \text{ to } w\}.$$

Moreover, for $\mathcal{A} \subseteq \mathcal{V}$ we put $\mathrm{pa}(\mathcal{A}) = \bigcup_{v \in \mathcal{A}} \mathrm{pa}(v)$, $\mathrm{an}(\mathcal{A}) = \bigcup_{v \in \mathcal{A}} \mathrm{an}(v)$, $\mathrm{de}(\mathcal{A}) = \bigcup_{v \in \mathcal{A}} \mathrm{de}(v)$. Below $\mathcal{A}, \mathcal{B}, \mathcal{C}$ stand for any three disjoint subsets of $\mathcal{V}$ with $\mathcal{C}$ possible empty. A trail (directed path) from $\mathcal{A}$ to $\mathcal{B}$ is a trail (directed path) from an $a \in \mathcal{A}$ to a $b \in \mathcal{B}$.

Let us begin with the classical definition of d-separation (Pearl, 1985).

Definition 3.1 *Let us say that* $\mathcal{B}$ *is* d*-blocked from* $\mathcal{A}$ *by* $\mathcal{C}$ *if every trail from* $\mathcal{A}$ *to* $\mathcal{B}$

- *contains a chain* $\leftarrow c \leftarrow$ *a chain* $\to c \to$ *or a fork* $\leftarrow c \to$ *with* $c \in \mathcal{C}$ *or a collider* $\to v \leftarrow$ *with* $\mathrm{de}(v) \cap \mathcal{C} = \emptyset$.

We will then write $\mathcal{A} \perp_d \mathcal{B} | \mathcal{C}$.

3. The term "local" refers here to time adjacency and not to space adjacency. We follow the terminology introduced by Didelez (2007, 2008) and prevailing in the literature, despite the fact that this terminology might be misleading.
4. Definition of a *trail* requires some caution, because there are subtle differences between several definitions appearing in the literature.

Not only does Definition 3.1 apply to directed graphs with cycles but also a version of the seminal Pearl's result remains true in this more general setup. In the following theorem we consider a process with independent initial distribution,

$$\mathbb{P}\left(\forall_{v\in\mathcal{V}}\, X_v(0) \in E_v\right) = \prod_v \mathbb{P}(X_v(t) \in E_v). \tag{4}$$

Theorem 3.2

(a) Consider a discrete time process $\mathbf{X}$ *which satisfies* (4) *and* (1) *with respect to* $\mathcal{G}$*. If* $\mathcal{B}$ *is* d*-blocked from* $\mathcal{A}$ *by* $\mathcal{C}$ *then* $\mathbf{X}_{\mathcal{B}} \perp\!\!\!\perp \mathbf{X}_{\mathcal{A}} | \mathbf{X}_{\mathcal{C}}$*.*

(b) If $\mathcal{B}$ *is not* d*-blocked from* $\mathcal{A}$ *by* $\mathcal{C}$ *then there exists a process* $\mathbf{X}$ *satisfying* (4) *and* (1) *such that* $\mathbf{X}_{\mathcal{B}} \not\perp\!\!\!\perp \mathbf{X}_{\mathcal{A}} | \mathbf{X}_{\mathcal{C}}$*.*

Let us emphasise that in the above theorem $\mathbf{X}_{\mathcal{A}}$, $\mathbf{X}_{\mathcal{B}}$ and $\mathbf{X}_{\mathcal{C}}$ stand for the whole sub-processes, e.g. $\mathbf{X}_{\mathcal{C}} = (X_{\mathcal{C}}(t) : t = 0, 1, \ldots)$ and so on. Several results similar to Theorem 3.2 can be found in the literature and they date back to 1996 (Pearl and Dechter, 1996; Neal, 2000). Didelez (2007, Proposition 5) has an analogous result for continuous time Markov processes. However, these results are obtained in different setups under different sets of assumptions. Although Theorem 3.2 seems to be known, we were not able to find an exact reference. Therefore we provide an independent proof in the Appendix A.[5]

Note that condition (4) is essential in Theorem 3.2. It can be replaced by a weaker assumption but cannot be entirely omitted.[6] For example, to obtain conclusion (a), it is sufficient to assume that the probability distribution of $X(0)$ factorises along a DAG, say $\mathcal{G}_0$, obtained by removing some edges from $\mathcal{G}$.

More interesting is the case of $\mathbf{X}$ being a strictly stationary process. For simplicity assume that $\mathbf{X}$ is a homogeneous (in time) ergodic Markov chain starting from the stationary distribution π. It is hardly surprising that in this situation there are less conditional independencies than under independent initial distribution. We proceed to define a notion of c-separation which describes conditional independencies in the stationary regime.

Definition 3.3 *Let us say that* $\mathcal{B}$ *is* c*-blocked from* $\mathcal{A}$ *by* $\mathcal{C}$ *if every trail from* $\mathcal{A}$ *to* $\mathcal{B}$

- *contains a collider* $\rightarrow v \leftarrow$ *with* $\mathrm{de}(v) \cap \mathcal{C} = \emptyset$*.*

We will then write $\mathcal{A} \perp_c \mathcal{B} | \mathcal{C}$*.*

Theorem 3.4 *Consider a homogeneous Markov chain* $\mathbf{X} = (X_v(t))$ *satisfying* (1) *with respect to graph* $\mathcal{G}$*. Assume that* $\mathbf{X}$ *is ergodic, has unique stationary distribution* π*. If* $\mathcal{A}$ *is* c*-blocked from* $\mathcal{B}$ *by* $\mathcal{C}$ *then*

- *If* $X \sim \pi$ *then* $X_{\mathcal{A}} \perp\!\!\!\perp X_{\mathcal{B}} | X_{\mathcal{C}}$*.*
- *If* $X(0) \sim \pi$*, that is if the process* $\mathbf{X}$ *is strictly stationary, then* $\mathbf{X}_{\mathcal{A}} \perp\!\!\!\perp \mathbf{X}_{\mathcal{B}} | \mathbf{X}_{\mathcal{C}}$*.*

5. Our proof covers also the case of continuous time Markov processes.
6. The role of the initial distribution is sometimes overlooked, as e.g. in the cited above Proposition 5 in Didelez (2007).

Note that the first item above concerns a random variable X with values in $\mathbb{S}$ with law π, while the second item refers to the whole process $\mathbf{X}$.

Remark *We conjecture that c-separation is also necessary for conditional independence (at least in the weak sense explained before), that is if $\mathcal{A}$ is not c-blocked from $\mathcal{B}$ by $\mathcal{C}$ then there exist a stationary Markov chain satisfying* (1) *such that*

- $X_{\mathcal{A}}(0) \not\perp\!\!\!\perp X_{\mathcal{B}}(0)|X_{\mathcal{C}}(0)$,
- $\mathbf{X}_{\mathcal{A}} \not\perp\!\!\!\perp \mathbf{X}_{\mathcal{B}}|\mathbf{X}_{\mathcal{C}}$.

We proceed to examine the properties of "predictive conditional (in)dependence", that is structural dependence of the present and future of the process on its past. Let us write

$$\mathbf{X}(< t) = (X(s) : s < t), \quad \mathbf{X}(\geqslant t) = (X(s) : s \geqslant t),$$

and similarly define $\mathbf{X}_{\mathcal{C}}(< t)$ and $\mathbf{X}_{\mathcal{C}}(\geqslant t)$ for $\mathcal{C} \subseteq \mathcal{V}$. Note that the very notion of local dependence graph in Definition 2.1 is based on the idea of predictive independence.

Proposition 3.5 *Assume that process $\mathbf{X}$ satisfies condition* (1)*. Then $v \not\to w$ implies conditional independence $X_v(t) \perp\!\!\!\perp \mathbf{X}_w(< t)|\mathbf{X}_{\mathcal{V}\setminus w}(< t)$. More generally, if $\mathcal{A} \not\to \mathcal{B}$, that is $\mathrm{pa}(\mathcal{B}) \cap \mathcal{A} = \emptyset$, then $X_{\mathcal{B}}(t) \perp\!\!\!\perp \mathbf{X}_{\mathcal{A}}(< t)|\mathbf{X}_{\mathcal{V}\setminus\mathcal{A}}(< t)$.*

To find graph statements equivalent to more general predictive (in)dependence it is necessary to define *asymmetric* separability relations. Let us begin with a notion of δ-separation introduced by Didelez (2008, Definition 7).

Definition 3.6 *Let us say that $\mathcal{B}$ is δ-blocked from $\mathcal{A}$ by $\mathcal{C}$ if every trail from $\mathcal{A}$ to $\mathcal{B}$ which ends with an arrow $\to b \in \mathcal{B}$*

- *contains a chain $\leftarrow c \leftarrow$ a chain $\to c \to$ or a fork $\leftarrow c \to$ with $c \in \mathcal{C}$ or a collider $\to v \leftarrow$ with $\mathrm{de}(v) \cap \mathcal{C} = \emptyset$.*

We will then write $\mathcal{A} \not\to_{\delta} \mathcal{B}|\mathcal{C}$.

The relevance of the above definition is explained by the following result.

Theorem 3.7

(a) Consider a process $\mathbf{X} = (X_v(t))$ satisfying (1) *with respect to graph $\mathcal{G}$. If $\mathcal{B}$ is δ-blocked from $\mathcal{A}$ by $\mathcal{C}$ then $X_{\mathcal{B}}(t) \perp\!\!\!\perp \mathbf{X}_{\mathcal{A}}(< t)|\mathbf{X}_{\mathcal{B}\cup\mathcal{C}}(< t)$.*

(b) If $\mathcal{B}$ is not δ-blocked from $\mathcal{A}$ by $\mathcal{C}$ then there exists a process $\mathbf{X}$ satisfying and (1) *such that $X_{\mathcal{B}}(t) \perp\!\!\!\perp \mathbf{X}_{\mathcal{A}}(< t)|\mathbf{X}_{\mathcal{B}\cup\mathcal{C}}(< t)$.*

A result exactly analogous to Theorem 3.7 (a) in a setting of continuous time processes can be found in (Didelez, 2008, Theorem 1). Theorem 3.7 seems to belong to the folklore. Therefore we omit its proof, noting that it can also be obtained by a simple modification of the proof of the forthcoming Theorem 3.9.

The "local dependence" in the sense of Definition 2.1 considered in Theorem 3.7 deals with dependence of the *present* on the past. To describe dependence of the *future* on the past we need yet another notion of separability which presumably is new.

Definition 3.8 *Let us say that $\mathcal{B}$ is ε-blocked from $\mathcal{A}$ by $\mathcal{C}$ if every trail from $\mathcal{A}$ to $\mathcal{B}$ which ends with an arrow $\to b \in \mathcal{B}$*

- *contains a chain $\leftarrow c \leftarrow$ or a fork $\leftarrow c \to$ with $c \in \mathcal{C}$ or a collider $\to v \leftarrow$ such that $\mathrm{de}(v) \cap \mathcal{C} = \emptyset$,* ***or***
- *contains a chain $\to c \to$ with $c \in \mathcal{C}$ that occurs earlier than some collider $\to v \leftarrow$ with $\mathrm{de}(v) \cap \mathcal{C} \neq \emptyset$.*

We will then write $\mathcal{A} \not\to_{\varepsilon} \mathcal{B}|\mathcal{C}$.

In other words, we can traverse through chains, forks and colliders as in the usual sense of Pearl's d-separation ($\leftarrow x \leftarrow$, $\to x \to$, $\leftarrow x \to$ for $x \notin \mathcal{C}$ or $\to v \leftarrow$ for $\mathrm{de}(v) \cap \mathcal{C} \neq \emptyset$), moreover we can move through $\to c \to$ for $c \in \mathcal{C}$ but only after the last collider (this is an empty condition if there are no colliders).

To the best of our knowledge, the following result is new.

Theorem 3.9

(a) Consider a process $\mathbf{X} = (X_v(t))$ satisfying (1) with respect to graph $\mathcal{G}$. If $\mathcal{B}$ is ε-blocked from $\mathcal{A}$ by $\mathcal{C}$ then $X_{\mathcal{B}}(\geqslant t) \perp\!\!\!\perp \mathbf{X}_{\mathcal{A}}(< t) | \mathbf{X}_{\mathcal{B}\cup\mathcal{C}}(< t)$.

(b) If $\mathcal{B}$ is not ε-blocked from $\mathcal{A}$ by $\mathcal{C}$ then there exists a process $\mathbf{X}$ satisfying and (1) such that $X_{\mathcal{B}}(\geqslant t) \perp\!\!\!\perp \mathbf{X}_{\mathcal{A}}(< t) | \mathbf{X}_{\mathcal{B}\cup\mathcal{C}}(< t)$.

4. Proofs

Proof [Proof of Theorem 3.4]

We are going to define three disjoint sets $\mathcal{A}_2$, $\mathcal{B}_2$ and $\mathcal{C}_2$ such that

- No arrows enter $\mathcal{A}_2$ from $\mathcal{V}\backslash\mathcal{A}_2$, no arrows enter $\mathcal{B}_2$ from $\mathcal{V}\backslash\mathcal{B}_2$, no arrows enter $\mathcal{C}_2$ from $\mathcal{V}\backslash\mathcal{C}_2$.

To begin with, let

$$\mathcal{A}_1 = \{v \in \mathcal{V} : \text{there is a trail from } a \in \mathcal{A} \text{ to } v \text{ not blocked by a collider } c \notin \mathrm{an}(\mathcal{C})\}.$$

By convention we assume that $\mathcal{A} \subseteq \mathcal{A}_1$ (degenerate trail is not blocked) and also that $\mathcal{A}_1$ contains all nodes connected to some $a \in \mathcal{A}$ by a single edge (whatever its direction). Note that $\mathcal{A}_1$ can contain nodes in $\mathrm{an}(\mathcal{C})$. Now we put

$$\mathcal{A}_1' = \{v \in \mathcal{A}_1 : \text{there is a trail from } r \notin \mathcal{A}_1 \text{ to } v \text{ not blocked by a collider } c \notin \mathrm{an}(\mathcal{C})\},$$

and

$$\mathcal{A}_2 = \mathcal{A}_1 \backslash \mathcal{A}_1'.$$

To verify that there are no arrows entering $\mathcal{A}_2$ from outside, suppose the contrary, i.e. $w \to v$ with $v \in \mathcal{A}_2$ and $w \notin \mathcal{A}_2$. By definition of $\mathcal{A}_2$, we see that we cannot have $w \notin \mathcal{A}_1$, so $w \in \mathcal{A}_1'$. But the trail leading from some $r \notin \mathcal{A}_1$ to w can be augmented by adding the arrow $w \to v$. This trail from r to v is not blocked at w, just because w is not a collider. Thus $v \in \mathcal{A}_1'$, contrary to our assumption.

Sets $\mathcal{B}_1$, $\mathcal{B}_1'$ and $\mathcal{B}_2$ are defined analogously. Our next step is verifying that

- $A_2 \cap \mathcal{B}_2 = \emptyset$.

Indeed, if $v \in \mathcal{A}_2 \subseteq \mathcal{A}_1$ then there is an open trail from $a \in \mathcal{A}$ to v. If moreover we had $v \in \mathcal{B}_2 \subseteq \mathcal{B}_1$ then there would exist an open trail from $b \in \mathcal{B}$ to v. Since $b \notin \mathcal{A}_1$, we would obtain that $v \in \mathcal{A}_1'$ which is a contradiction.

Now we are ready to define

$$\mathcal{C}_2 = \mathrm{an}(\mathcal{C}) \backslash (\mathcal{A}_2 \cup \mathcal{B}_2)$$

and to verify that no arrows enter $\mathcal{C}_2$ from outside. Indeed, if we had an arrow $\mathcal{A}_2 \ni a \to c \in \mathrm{an}(\mathcal{C})$ then this would imply $c \in \mathcal{A}_1$ (because a in not a collider on a trail from $\mathcal{A}$ to c) and moreover $c \notin \mathcal{A}_1'$ for otherwise there would exist an open trail from $\mathcal{A}$ to c to $r \notin \mathcal{A}_1$ ($c \in \mathrm{an}(\mathcal{C})$ never blocks a trail). Therefore c must belong to $\mathcal{A}_2$. For the same reason, if $\mathcal{B}_2 \ni b \to c \in \mathrm{an}(\mathcal{C})$ then c must belong to $\mathcal{B}_2$. Finally if $\mathcal{V} \backslash (\mathcal{A}_2 \cup \mathcal{B}_2) \ni v \to c \in \mathrm{an}(\mathcal{C})$ then obviously $v \in \mathrm{an}(\mathcal{C})$.

The basic observation is that $\mathbf{X}_{\mathcal{A}_2}$, $\mathbf{X}_{\mathcal{B}_2}$ and $\mathbf{X}_{\mathcal{C}_2}$ are Markov chains. For readability, in the following we assume that all the transition probabilities and the stationary distribution have densities with respect to some reference measure. If we denote the set $\mathcal{V} \backslash (\mathcal{A}_2 \cup \mathcal{B}_2 \cup \mathcal{C}_2)$ by $\mathcal{R}$ then the transition probabilities (transition densities) of $\mathbf{X}$ have the following form:

$$P(x', x) = P(x'_{\mathcal{A}_2}, x_{\mathcal{A}_2}) P(x'_{\mathcal{B}_2}, x_{\mathcal{B}_2}) P(x'_{\mathcal{C}_2}, x_{\mathcal{C}_2}) P(x'_{\mathcal{V}}, x_{\mathcal{R}}).$$

The stationary probability distribution (density) π satisfies

$$\pi(x) = \int_{x'} \pi(x') P(x', x) \mathrm{d}x',$$

i.e.

$$\begin{aligned} &\pi(x_{\mathcal{A}_2}, x_{\mathcal{B}_2}, x_{\mathcal{C}_2}, x_{\mathcal{R}}) \\ &= \int_{x'} \pi(x'_{\mathcal{A}_2}, x'_{\mathcal{B}_2}, x'_{\mathcal{C}_2}, x'_{\mathcal{R}}) P(x'_{\mathcal{A}_2}, x_{\mathcal{A}_2}) P(x'_{\mathcal{B}_2}, x_{\mathcal{B}_2}) P(x'_{\mathcal{C}_2}, x_{\mathcal{C}_2}) P(x', x_{\mathcal{R}}) \mathrm{d}x'. \end{aligned}$$

Marginalising w.r.t. firstly $x_{\mathcal{R}}$ and then $x'_{\mathcal{R}}$ we get

$$\begin{aligned}&\pi(x_{\mathcal{A}_2}, x_{\mathcal{B}_2}, x_{\mathcal{C}_2})\\ &= \int_{x'_{\mathcal{A}_2}, x'_{\mathcal{B}_2}, x'_{\mathcal{C}_2}} \pi(x'_{\mathcal{A}_2}, x'_{\mathcal{B}_2}, x'_{\mathcal{C}_2}) P(x'_{\mathcal{A}_2}, x_{\mathcal{A}_2}) P(x'_{\mathcal{B}_2}, x_{\mathcal{B}_2}) P(x'_{\mathcal{C}_2}, x_{\mathcal{C}_2}) \mathrm{d}x'_{\mathcal{A}_2} \mathrm{d}x'_{\mathcal{B}_2} \mathrm{d}x'_{\mathcal{C}_2}.\end{aligned}$$

Iterating the above equation we obtain

$$\begin{aligned}&\pi(x_{\mathcal{A}_2}, x_{\mathcal{B}_2}, x_{\mathcal{C}_2})\\ &= \int_{x'_{\mathcal{A}_2}, x'_{\mathcal{B}_2}, x'_{\mathcal{C}_2}} \pi(x'_{\mathcal{A}_2}, x'_{\mathcal{B}_2}, x'_{\mathcal{C}_2}) P^n(x'_{\mathcal{A}_2}, x_{\mathcal{A}_2}) P^n(x'_{\mathcal{B}_2}, x_{\mathcal{B}_2}) P^n(x'_{\mathcal{C}_2}, x_{\mathcal{C}_2}) \mathrm{d}x'_{\mathcal{A}_2} \mathrm{d}x'_{\mathcal{B}_2} \mathrm{d}x'_{\mathcal{C}_2},\end{aligned}$$

where P^n denotes n-step transition probabilities. Since the chain is ergodic, we have $P^n(x', x) \to_{n\to\infty} \pi(x)$. We can first pass to the limit and then marginalise w.r.t. $x'_{\mathcal{A}_2}, x'_{\mathcal{B}_2}, x'_{\mathcal{C}_2}$ to obtain

$$\pi(x_{\mathcal{A}_2}, x_{\mathcal{B}_2}, x_{\mathcal{C}_2}) = \pi(x_{\mathcal{A}_2})\pi(x_{\mathcal{B}_2})\pi(x_{\mathcal{C}_2}).$$

Thus $X_{\mathcal{A}_2}$, $X_{\mathcal{A}_2}$ and $X_{\mathcal{C}_2}$ are π-independent. In particular, $(X_{\mathcal{A}}, X_{\mathcal{A}_2\cap\mathcal{C}})$, $(X_{\mathcal{B}}, X_{\mathcal{B}_2\cap\mathcal{C}})$ and $X_{\mathcal{C}_2\cap\mathcal{C}}$ are independent. It is clear that $\mathcal{C} = (\mathcal{A}_2 \cap \mathcal{C}) \cup (\mathcal{B}_2 \cap \mathcal{C}) \cup (\mathcal{C}_2 \cap \mathcal{C})$, hence

$$\pi(x_{\mathcal{A}}, x_{\mathcal{B}}, x_{\mathcal{C}}) = \pi(x_{\mathcal{A}}, x_{\mathcal{A}_2\cap\mathcal{C}})\pi(x_{\mathcal{B}}, x_{\mathcal{B}_2\cap\mathcal{C}})\pi(x_{\mathcal{C}_2\cap\mathcal{C}}).$$

It follows that $X_{\mathcal{A}}$ and $X_{\mathcal{B}}$ are conditionally independent given $X_{\mathcal{C}}$.

Analogous reasoning leads to the proof of the second conclusion (in the stationary regime we have conditional independence of the whole subprocesses, $\mathbf{X}_{\mathcal{A}} \perp\!\!\!\perp \mathbf{X}_{\mathcal{B}} | \mathbf{X}_{\mathcal{C}}$). ■

Proof [Proof of Theorem 3.9] It is clear that the joint probability distribution of $\mathbf{X}$ factorises along the graph $\widetilde{\mathcal{G}}$ with the set of vertices $\mathcal{V} \times \{0, 1, 2, \ldots\}$ and the set of edges of the form $(v, s) \to (w, t)$ where $s < t$ and $v \to w$ is an edge in $\mathcal{G}$. For $\mathcal{U} \subseteq \mathcal{V}$ we define $\mathcal{U}(\leqslant t) = \{(u, s) \colon u \in \mathcal{U}, s \leqslant t\}$. We similarly define $\mathcal{U}(< t)$, $\mathcal{U}(t)$ and $\mathcal{U}(\geqslant t)$.

Proof of part (a) Fix $t \in \mathbb{N}_0$ and take any trail

$$(a, t_0), e_1, (v_1, t_1), e_2, (v_2, t_2), e_3, \ldots, e_{n-1}, (v_{n-1}, t_{n-1}), e_n, (b, t_n)$$

in $\widetilde{G}$ from (a, t_0) to (b, t_n), where $a \in \mathcal{A}$, $b \in \mathcal{B}$, $t_0 < t \leqslant t_n$ and e_i is either $\to$ or $\leftarrow$ for each $i = 1, 2, \ldots, n$. Let us denote this trail by $\widetilde{\tau}$ and suppose it is **not** d-blocked by $(\mathcal{B} \cup \mathcal{C})(< t)$. Clearly

$$\begin{cases} t_{i+1} > t_i & \text{if } e_{i+1} = \to \\ t_{i+1} < t_i & \text{otherwise} \end{cases} \tag{5}$$

This implies that for any $1 \leqslant k \leqslant n$

$$\text{if } t_k \geqslant t \text{ then } e_k = \to. \tag{6}$$

Indeed, assume that $t_k \geqslant t$. As $t_k \geqslant t_0$, by (5) we have $e_i = \to$ for some $1 \leqslant i \leqslant k$. Let e_m be the furthest $\to$ to the left of v_k (i.e. $m = \max\{i \leqslant k \colon e_i = \to\}$). Then $m = k$, otherwise

$\rightarrow (v_m, t_m) \leftarrow$ would be a collider with neither the middle node, nor any of its descendants, not in $(\mathcal{B} \cup \mathcal{C})(\leqslant t)$ (as $t_m \geqslant t_k \geqslant t$). That would block $\widetilde{\tau}$.

In particular from (6) we obtain $e_n = \rightarrow$ and the trail τ defined by

$$a, e_1, v_1, e_2, v_2, e_3, \ldots, e_{n-1}, v_{n-1}, e_n, b,$$

is a trail in $\mathcal{G}$ that connects $\mathcal{A}$ and $\mathcal{B}$ and ends in an arrow $\rightarrow b$. Since $\mathcal{B}$ is ε-blocked from $\mathcal{A}$ by $\mathcal{C}$, this trail is ε-blocked. This implies one of the following

- There is a collider $\rightarrow v_k \leftarrow$ such that neither v_k nor any of its descendants in $\mathcal{G}$ are in $\mathcal{C}$. As a result, neither (v_k, t_k) nor any of its descendants are in $(\mathcal{B} \cup \mathcal{C})(< t)$. This means that $\rightarrow (v_k, t_k) \leftarrow$ d-blocks $\widetilde{\tau}$, which is a contradiction.
- There is a fork $\leftarrow v_k \rightarrow$ with $v_k \in \mathcal{C}$. Therefore $\leftarrow (v_k, t_k) \rightarrow$ is a fork in $\widetilde{\tau}$ and hence $t_k \geqslant t$ (otherwise $(v_k, t_k) \in \mathcal{C}(< t)$ would block $\widetilde{\tau}$). This contradicts (6).
- There is a chain $\leftarrow v_k \leftarrow$ with $v_k \in \mathcal{C}$. We obtain a contradiction in the same way, as in the case of a fork.
- There is a chain $\rightarrow v_k \rightarrow$ with $v_k \in \mathcal{C}$ and a collider $\rightarrow v_l \leftarrow$ with $\mathrm{de}(v_l) \cap \mathcal{C} \neq \varnothing$ that occurs afterwards ($k \leqslant l$). We can consider the earliest such collider (i.e. $e_i = \rightarrow$ for $i = k+1, \ldots, l$). This means that $t_k \leqslant t_l$. On the other hand, $\rightarrow (v_k, t_k) \rightarrow$ is open in $\widetilde{\mathcal{G}}$ which means that $t_k \geqslant t$ (otherwise $(v_k, t_k) \in (\mathcal{B} \cup \mathcal{C})(< t)$). Similarly, $t_l < t$ since the collider $\rightarrow (v_l, t_l) \leftarrow$ is open in $\widetilde{\mathcal{G}}$ and hence (v_l, t_l) or one of its descendants is in $(\mathcal{B} \cup \mathcal{C})(< t)$. We obtained $t_k \leqslant t_l$ and $t_k \geqslant t > t_l$, a contradiction.

All the possibilities implied by the assumption that $\widetilde{\tau}$ is not d-blocked by $(\mathcal{B} \cup \mathcal{C})(< t)$ have led us to a contradiction. This implies that $\mathcal{A}(< t)$ and $\mathcal{B}(\geqslant t)$ are d-blocked by $(\mathcal{B} \cup C)(< t)$, which by the classical theory of d-separation implies $\mathbf{X}_{\mathcal{B}}(\geqslant t) \perp\!\!\!\perp \mathbf{X}_{\mathcal{A}}(< t) | \mathbf{X}_{\mathcal{B} \cup \mathcal{C}}(< t)$.

Proof of part (b) Take any trail in $\mathcal{G}$ that connects $a \in \mathcal{A}$ to $b \in \mathcal{B}$ and is not ε-blocked by $\mathcal{C}$:

$$a, e_1, v_1, e_2, v_2, e_3, \ldots, e_{n-1}, v_{n-1}, e_n, b,$$

where $e_n = \rightarrow$. Let us denote this trail by τ and let $\widetilde{\tau}$ be the trail in $\widetilde{\mathcal{G}}$ given by

$$(a, t_0), e_1, (v_1, t_1), e_2, (v_2, t_2), e_3, \ldots, e_{n-1}, (v_{n-1}, t_{n-1}), e_n, (b, t_n)$$

where $t_0 = n$ and $t_1, \ldots, t_n$ are defined by $e_1, \ldots, e_n$ via

$$t_{i+1} = \begin{cases} t_i + 1 & \text{if } e_{i+1} = \rightarrow \\ t_i - 1 & \text{otherwise} \end{cases} \tag{7}$$

For any collider $\rightarrow (v_i, t_i) \leftarrow$ on $\widetilde{\tau}$ choose $w_i \in \mathrm{de}(v_i) \cap \mathcal{C}$ (such w_i exists as τ is ε-open) and let d_k be the length of the shortest directed path from v_i to w_i in $\mathcal{G}$. Set $t = t_0 + \max_{i:\, \rightarrow (v_i, t_i) \leftarrow} \{t_i + d_i\}$ (any t that is larger from t_0 and $\max_{i:\, \rightarrow (v_i, t_i) \leftarrow} \{t_k + d_k\}$ will do).

We now modify the trail $\widetilde{\tau}$ to obtain a new one. If there is a right-directed chain in τ with its middle node in $\mathcal{C}$, let r be the index of the earliest one. Otherwise let $r = n$. If $t_r < t$, replace $(v_{r-1}, t_{r-1}) \rightarrow (v_r, t_r)$ with

$$(v_{r-1}, t_{r-1}) \rightarrow (v_{r-1}, t_{r-1} + 1) \rightarrow \ldots \rightarrow (v_{r-1}, t - 1) \rightarrow (v_r, t)$$

and (v_l, t_l) with $(v_l, t_l + t - t_r)$ for $l > r$. It is clear that in this way we obtain a trail from (v_0, t_0) to (v_n, T) in $\widetilde{G}$, where $T = \max\{t_n, t_n + t - t_r\}$. Let us denote this trail by $\widehat{\tau}$.

Firstly we show that $\widehat{\tau}$ connects $\mathcal{A}(< t)$ to $\mathcal{B}(\geqslant t)$. By the definition of t we have $t_0 < t$ and $(a, t_0) \in \mathcal{A}(< t)$. Note that $e_l = \rightarrow$ for $l \geqslant r$; it follows from the fact that τ is ε-open and hence there can be no colliders to the right of v_r. This implies $t_k \geqslant t_r$ for $k \geqslant r$, in particular $t_n \geqslant t_r$, hence $T \geqslant t$ and (b, T) – which is the endpoint of $\hat{\tau}$ – belongs to $\mathcal{B}(\geqslant t)$.

Now we prove that $\hat{\tau}$ is not d-blocked by $(\mathcal{B} \cup \mathcal{C})(< t)$ Three consecutive nodes on $\widehat{\tau}$ can be in the one of the following forms:

(i) $(v_{k-1}, t_{k-1}), e_k, (v_k, t_k), e_{k+1}, (v_{k+1}, t_{k+1})$ for $k < r$
Note that $v_{k-1}, e_k, v_k, e_{k+1}, v_{k+1}$ is ε-open in $\mathcal{G}$. This means that

- $v_k \notin \mathcal{C}$ and $(e_k, e_{k+1}) \in \{(\rightarrow, \rightarrow), (\leftarrow, \rightarrow), (\leftarrow, \leftarrow)\}$. Then (i) is not d-blocked by $(\mathcal{B} \cup \mathcal{C})(< t)$ as $(v_k, t_k) \notin (\mathcal{B} \cup \mathcal{C})(< t)$
- $(e_k, e_{k+1}) = (\rightarrow, \leftarrow)$ and there exists $w_k \in \mathrm{de}(v_k) \cap \mathcal{C}$. By the definition of t we have $(w_k, t_k + d_k) \in \mathrm{de}((v_k, t_k)) \cap \mathcal{C}(< t)$ and again (i) is not d-blocked by $(\mathcal{B} \cap \mathcal{C})(< t)$.

(ii) $(v_{r-1}, t_{r-1} + i - 1) \rightarrow (v_{r-1}, t_{r-1} + i) \rightarrow (v_{k+1}, t_{r-1} + i + 1)$ for $i \geqslant t - t_r$
These are clearly d-open since by the definition of r we get $v_{r-1} \notin \mathcal{C}$.

(iii) $(v_{k-1}, t_{k-1} + t - t_r) \rightarrow (v_k, t_k + t - t_r) \rightarrow (v_{k+1}, t_{k+1} + t - t_r)$ for $k \geqslant r$
As $k \geqslant r$ we have $t_k \geqslant t_r$, hence $t_k + t - t_r \geqslant t$ and $(v_k, t_k + t - t_r) \notin (\mathcal{B} \cup \mathcal{C})(< t)$. This means that (iii) is not d-blocked by $(\mathcal{B} \cup \mathcal{C})(< t)$.

We proved that the trail $\widehat{\tau}$ joins $\mathcal{A}(< t)$ with $\mathcal{B}(\geqslant t)$ in $\widetilde{\mathcal{G}}$ and it is not d-blocked by $(\mathcal{B} \cup \mathcal{C})(< t)$. By the classical theory of d-separation this means that there is a probability distribution that factorises along $\widetilde{G}$ in which $\mathbf{X}_{\mathcal{B}}(\geqslant t) \not\perp\!\!\!\perp \mathbf{X}_{\mathcal{A}}(< t) | \mathbf{X}_{\mathcal{B} \cup \mathcal{C}}(< t)$; this probability distribution defines the relevant process. ■

5. Summary

We summarise our findings in a compact form of the following table that relates the well known separability relationships to our presumably novel observations.

	Separability property in $\mathcal{G}$	Symbol	Independence property of $\mathbf{X}$
(1)	Every trail between $\mathcal{A}$ and $\mathcal{B}$ is d-blocked by $\mathcal{C}$	$\mathcal{A} \perp_d \mathcal{B} \| \mathcal{C}$	$\mathbf{X}_{\mathcal{B}} \perp\!\!\!\perp \mathbf{X}_{\mathcal{A}} \| \mathbf{X}_{\mathcal{C}}$ (independent start)
(2)	Every trail between $\mathcal{A}$ and $\mathcal{B}$ is c-blocked by $\mathcal{C}$	$\mathcal{A} \perp_c \mathcal{B} \| \mathcal{C}$	$\mathbf{X}_{\mathcal{B}} \perp\!\!\!\perp \mathbf{X}_{\mathcal{A}} \| \mathbf{X}_{\mathcal{C}}$ (stationary)
(3)	No arrows from $\mathcal{A}$ to $\mathcal{B}$	$\mathcal{A} \not\rightarrow \mathcal{B}$	$X_{\mathcal{B}}(t) \perp\!\!\!\perp \mathbf{X}_{\mathcal{A}}(< t) \| \mathbf{X}_{\mathcal{V} \setminus \mathcal{A}}(< t)$
(4)	Every trail from $\mathcal{A}$ to $\mathcal{B}$ is δ-blocked by $\mathcal{C}$	$\mathcal{A} \not\rightarrow_\delta \mathcal{B} \| \mathcal{C}$	$X_{\mathcal{B}}(t) \perp\!\!\!\perp \mathbf{X}_{\mathcal{A}}(< t) \| \mathbf{X}_{\mathcal{B} \cup \mathcal{C}}(< t)$
(5)	No directed paths from $\mathcal{A}$ to $\mathcal{B}$	$\mathcal{A} \not\rightarrow \mathcal{B}$	$\mathbf{X}_{\mathcal{B}}(\geqslant t) \perp\!\!\!\perp \mathbf{X}_{\mathcal{A}}(< t) \| \mathbf{X}_{\mathcal{V} \setminus \mathcal{A}}(< t)$
(6)	Every trail from $\mathcal{A}$ to $\mathcal{B}$ is ε-blocked by $\mathcal{C}$	$\mathcal{A} \not\rightarrow_\varepsilon \mathcal{B} \| \mathcal{C}$	$\mathbf{X}_{\mathcal{B}}(\geqslant t) \perp\!\!\!\perp \mathbf{X}_{\mathcal{A}}(< t) \| \mathbf{X}_{\mathcal{B} \cup \mathcal{C}}(< t)$

We believe that the results presented should remain valid under the continuous time regime, which is the topic of our on-going research.

References

Stephan Bongers, Patrick Forré, Jonas Peters, and Joris M Mooij. Foundations of structural causal models with cycles and latent variables. *The Annals of Statistics*, 49(5):2885–2915, 2021.

Vanessa Didelez. Graphical models for composable finite Markov processes. *Scandinavian Journal of Statistics*, 34(1):169–185, 2007.

Vanessa Didelez. Graphical models for marked point processes based on local independence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(1):245–264, 2008.

Michael Eichler and Vanessa Didelez. Causal reasoning in graphical time series models. In *23rd Annual Conference on Uncertainty in Artifical Intelligence*, 2007.

Clive WJ Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, pages 424–438, 1969.

Søren Wengel Mogensen and Niels Richard Hansen. Markov equivalence of marginalized local independence graphs. *The Annals of Statistics*, 48(1):539–559, 2020.

Søren Wengel Mogensen and Niels Richard Hansen. Graphical modeling of stochastic processes driven by correlated noise. *Bernoulli*, 28(4):3023–3050, 2022.

Radford M Neal. On deducing conditional independence from d-separation in causal graphs with feedback (research note). *Journal of Artificial Intelligence Research*, 12:87–91, 2000.

Uri Nodelman, Christian R Shelton, and Daphne Koller. Continuous time Bayesian networks. In *Proceedings of the Eighteenth conference on Uncertainty in artificial intelligence*, pages 378–387, 2002.

Judea Pearl. Bayesian networks: A model of self-activated memory for evidential reasoning. In *Proceedings of the 7th conference of the Cognitive Science Society, University of California, Irvine, CA, USA*, pages 15–17, 1985.

Judea Pearl and Rina Dechter. Identifying independencies in causal graphs with feedback. In *Proceedings of the Twelfth international conference on Uncertainty in artificial intelligence*, pages 420–426, 1996.

Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. *Elements of causal inference: foundations and learning algorithms*. The MIT Press, 2017.

Tore Schweder. Composable Markov processes. *Journal of Applied Probability*, 7(2):400–410, 1970.

Appendix A. Proof of Theorem 3.2

Recall that $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is a directed graph with possible cycles. We assume that all nodes have self-loops ($v \in \mathrm{pa}(v)$ for every $v \in \mathcal{V}$). Definition of d-separation is exactly the same as in the acyclic case. Below $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ are disjoint sets of nodes, with $\mathcal{C}$ possible empty.

We say that $\mathcal{B}$ is d-blocked from $\mathcal{A}$ by $\mathcal{C}$ if every trail from $\mathcal{A}$ to $\mathcal{B}$

- *contains a chain $\leftarrow c \leftarrow$ a chain $\rightarrow c \rightarrow$ or a fork $\leftarrow c \rightarrow$ with $c \in \mathcal{C}$ or a collider $\rightarrow v \leftarrow$ with $\mathrm{de}(v) \cap \mathcal{C} = \emptyset$.*

(By our definition, a trail cannot contain any node more than once.)

We consider a discrete time stochastic process $\mathbf{X} = (X_v(t), v \in \mathcal{V}, t = 0, 1, \ldots)$. Random variables at node v, that is $X_v(t)$, take values in a measure space $\mathbb{S}_v$ equipped with some measure denoted by $\mathrm{d}x_v$. We assume that all probability distributions under consideration are absolutely continuous with respect to products of these reference measures. A generic notation for a density will be p.

Assume that $\mathbf{X}$ satisfies the following two conditions: for arbitrary measurable $E_v \subseteq \mathbb{S}_v$,

$$\mathbb{P}\left(\forall_{v\in\mathcal{V}}\, X_v(t) \in E_v \middle| X(s), s < t\right) = \prod_v \mathbb{P}(X_v(t) \in E_v | X_{\mathrm{pa}(v)}(s), s < t)$$

and

$$\mathbb{P}\left(\forall_{v\in\mathcal{V}}\, X_v(0) \in E_v\right) = \prod_v \mathbb{P}(X_v(0) \in E_v).$$

In terms of densities, our two basic assumptions can be rewritten as follows:

$$p\left(x(t)|x(s), s < t\right) = \prod_v p(x_v(t)|x_{\mathrm{pa}(v)}(s), s < t). \tag{A.1}$$

and

$$p\left(x(0)\right) = \prod_v p(x_v(0)). \tag{A.2}$$

Let us set a finite but arbitrary time horizon n. We will work with densities on the space of trajectories of a time-truncated process $(X_v(t), v \in \mathcal{V}, t = 0, 1, \ldots, n)$, that is $\mathbb{X} = \prod_{v\in\mathcal{V}} \mathbb{X}_v$, where $\mathbb{X}_v = \mathbb{S}_v^{n+1}$. It is clear that it is sufficient to prove Theorem 3.3 for this time-truncated process. From now on, to simplify notation, we just write $\mathbf{X} = (X_v(t), v \in \mathcal{V}, t = 0, 1, \ldots, n)$.

Theorem 3.2 asserts the following.

(a) Consider a discrete time process $\mathbf{X}$ which satisfies (A.2) and (A.1) with respect to $\mathcal{G}$. If $\mathcal{B}$ is d-blocked from $\mathcal{A}$ by $\mathcal{C}$ then $\mathbf{X}_\mathcal{B} \perp\!\!\!\perp \mathbf{X}_\mathcal{A} | \mathbf{X}_\mathcal{C}$.

(b) If $\mathcal{B}$ is not d-blocked from $\mathcal{A}$ by $\mathcal{C}$ then there exists a process $\mathbf{X}$ satisfying (A.2) and (A.1) such that $\mathbf{X}_\mathcal{B} \not\perp\!\!\!\perp \mathbf{X}_\mathcal{A} | \mathbf{X}_\mathcal{C}$.

Our proof is based on the following factorisation result for the joint density of $\mathbf{X}$. Define

$$p(\mathbf{x}_v \| \mathbf{x}_{\mathrm{pa}(v)}) = p(x_v(0)) \prod_{t=1}^{n} p(x_v(t)|x_{\mathrm{pa}(v)}(s) : s < t).$$

Proposition A.1 *(i) The joint density of* $\mathbf{X}$ *admits a factorisation*

$$p(\mathbf{x}) = \prod_{v \in \mathcal{V}} p(\mathbf{x}_v \| \mathbf{x}_{\mathrm{pa}(v)}).$$

(ii) Moreover, for every $\mathcal{A} \subset \mathcal{V}$ *and every fixed* $\mathbf{x}_{\mathrm{pa}(\mathcal{A})\setminus\mathcal{A}}$ *we have*

$$\int_{\mathbf{x}_{\mathcal{A}}} \prod_{v \in \mathcal{A}} p(\mathbf{x}_v \| \mathbf{x}_{\mathrm{pa}(v)}) \mathrm{d}\mathbf{x}_A = 1.$$

(We use the notations $\mathrm{d}\mathbf{x}_{\mathcal{A}} = \prod_{v \in \mathcal{A}} \mathrm{d}\mathbf{x}_v$*, where* $\mathrm{d}\mathbf{x}_v = \mathrm{d}x_v(0)\mathrm{d}x_v(1)\cdots\mathrm{d}x_v(n)$*. For emphasis we indicate the integrated variables under the symbol of integral.)*

Proof Using equations (A.1) and (A.2), we can express the joint density as

$$p(\mathbf{x}) = \prod_v p(x_v(0)) \prod_{t=1}^{n} \prod_v p(x_v(t) | x_{\mathrm{pa}(v)}(s), s < t).$$

Changing the order of the products yields the factorisation formula asserted in (i).

To verify (ii) we just take the integrals in the time-reversed order,

$$\int_{\mathbf{x}_{\mathcal{A}}} = \int_{x_{\mathcal{A}}(0)} \cdots \int_{x_{\mathcal{A}}(n-1)} \int_{x_{\mathcal{A}}(n)},$$

that is firstly with respect to $x_{\mathcal{A}}(n) = (x_v(n) : v \in \mathcal{A})$ then with respect to $x_{\mathcal{A}}(n-1)$ and down to $x_{\mathcal{A}}(0)$, repeatedly using the fact that

$$\int_{x_{\mathcal{A}}(t)} \prod_{v \in \mathcal{A}} p(x_v(t) | x_{\mathrm{pa}(v)}(s), s < t) \mathrm{d}x_{\mathcal{A}}(t) = 1.$$

The product over t telescopes. In the end we obtain that the integral over $\mathbf{x}_{\mathcal{A}}$ is equal to 1. ■

Remark *In spite of its simplicity, we think that Proposition A.1 is of independent interest and has an appealing interpretation. Assertion (ii) allows us to interpret* $\prod_{v \in \mathcal{A}} p(\mathbf{x}_v \| \mathbf{x}_{\mathrm{pa}(v)})$ *as a probability density on the space of trajectories of sub-process* $\mathbf{X}_{\mathcal{A}}$*. A moment of reflection reveals that this is the density of conditional-by-intervention probability distribution, given that* $\mathbf{X}_{\mathrm{pa}(\mathcal{A})\setminus\mathcal{A}}$ *is forced to assume value* $\mathbf{x}_{\mathrm{pa}(\mathcal{A})\setminus\mathcal{A}}$ *(we speak of intervention on the whole trajectories).*[7]

In the sequel we will use the following notation. Let $\mathcal{U}$ and $\mathcal{W}$ be disjoint subsets of nodes (in particular, they can be singletons). We write $\mathcal{U} — \mathcal{W}$ to indicate that $\mathcal{W}$ is not d-blocked from $\mathcal{U}$ and $\mathcal{U} \not— \mathcal{W}$ otherwise (with respect to the set $\mathcal{C}$ which is fixed and need not be explicitly indicated).

Proof [Proof of Theorem 3.3 (a)]

Our starting point is the factorisation formula for density $p(\mathbf{x})$ given in Proposition A.1 (i). We define the following subsets of $\mathcal{V}$. Firstly, $\mathcal{D} = \mathcal{V}\setminus(\mathcal{A} \cup \mathcal{B} \cup \mathcal{C})$. Now we put

7. Note that in the expression $\prod_{v \in \mathcal{A}} p(\mathbf{x}_v \| \mathbf{x}_{\mathrm{pa}(v)})$, components $\mathbf{x}_w$ with $w \in \mathcal{A}$ appear also *after* the symbol of conditioning "$\|$", in contrast with the usual *conditional-by-observation* density $p(\mathbf{x}_{\mathcal{A}} | \mathbf{x}_{\mathrm{pa}(\mathcal{A})\setminus\mathcal{A}})$.

- $\mathcal{D}_{\mathcal{AB}} = \{v \in \mathcal{D} : \mathcal{A} - v \text{ and } \mathcal{B} - v\}$.
- $\mathcal{D}_{\mathcal{A}} = \{v \in \mathcal{D} : \mathcal{A} - v \text{ and } \mathcal{B} \not- v\}$.
- $\mathcal{D}_{\mathcal{B}} = \{v \in \mathcal{D} : \mathcal{A} \not- v \text{ and } \mathcal{B} - v\}$.
- $\mathcal{D}_{\text{rest}} = \mathcal{D} \backslash (\mathcal{D}_{\mathcal{AB}} \cup \mathcal{D}_{\mathcal{A}} \cup \mathcal{D}_{\mathcal{B}})$.
- $\mathcal{C}_{\mathcal{A}} = \{v \in \mathcal{C} : \text{pa}(v) \cap (\mathcal{A} \cup \mathcal{D}_{\mathcal{A}}) \neq \varnothing\}$.
- $\mathcal{C}_{\mathcal{B}} = \{v \in \mathcal{C} : \text{pa}(v) \cap (\mathcal{B} \cup \mathcal{D}_{\mathcal{B}}) \neq \varnothing\}$.
- $\mathcal{C}_{\text{rest}} = \mathcal{C} \backslash (\mathcal{C}_{\mathcal{A}} \cup \mathcal{C}_{\mathcal{B}})$.

Sets $\mathcal{A}, \mathcal{B}, \mathcal{D}_{\mathcal{AB}}, \mathcal{D}_{\text{rest}}, \mathcal{D}_{\mathcal{A}}, \mathcal{D}_{\mathcal{B}}, \mathcal{C}_{\mathcal{A}}, \mathcal{C}_{\mathcal{B}}, \mathcal{C}_{\text{rest}}$ are disjoint ($\mathcal{C}_{\mathcal{A}} \cap \mathcal{C}_{\mathcal{B}} = \varnothing$ because $\mathcal{A} \not- \mathcal{B}$). We split $p(\mathbf{x}) = \prod_v p(\mathbf{x}_v \| \mathbf{x}_{\text{pa}(v)})$ into products over those subsets:

$$
\begin{aligned}
p(\mathbf{x}) = & \prod_{v \in \mathcal{D}_{\mathcal{AB}}} p(\mathbf{x}_v \| \mathbf{x}_{\text{pa}(v)}) \prod_{v \in \mathcal{D}_{\text{rest}}} p(\mathbf{x}_v \| \mathbf{x}_{\text{pa}(v)}) \prod_{v \in \mathcal{C}_{\text{rest}}} p(\mathbf{x}_v \| \mathbf{x}_{\text{pa}(v)}) \\
& \times \prod_{v \in \mathcal{A}} p(\mathbf{x}_v \| \mathbf{x}_{\text{pa}(v)}) \prod_{v \in \mathcal{C}_{\mathcal{A}}} p(\mathbf{x}_v \| \mathbf{x}_{\text{pa}(v)}) \prod_{v \in \mathcal{D}_{\mathcal{A}}} p(\mathbf{x}_v \| \mathbf{x}_{\text{pa}(v)}) \\
& \times \prod_{v \in \mathcal{B}} p(\mathbf{x}_v \| \mathbf{x}_{\text{pa}(v)}) \prod_{v \in \mathcal{C}_{\mathcal{B}}} p(\mathbf{x}_v \| \mathbf{x}_{\text{pa}(v)}) \prod_{v \in \mathcal{D}_{\mathcal{B}}} p(\mathbf{x}_v \| \mathbf{x}_{\text{pa}(v)}).
\end{aligned}
$$

Let us introduce the following notational convention: for subsets $\mathcal{W}, \mathcal{U} \subseteq \mathcal{V}$ we will write $p(\mathcal{W})$ instead of $p(\mathbf{x}_{\mathcal{W}})$ and

$$
p(\mathcal{W} \| \mathcal{U}) = \prod_{v \in \mathcal{W}} p(\mathbf{x}_v \| \mathbf{x}_{\text{pa}(v)}),
$$

whenever $\text{pa}(\mathcal{W}) \subseteq \mathcal{U}$, that is whenever all parents of any node in $\mathcal{W}$ belong to $\mathcal{U}$. It turns out that we can rewrite the joint probability distribution in a schematic form:

$$
\begin{aligned}
p(\mathcal{V}) = & \, p(\mathcal{D}_{\mathcal{AB}} \| \mathcal{V}) \, p(\mathcal{D}_{\text{rest}} \| \mathcal{D}_{\text{rest}} \cup \mathcal{C}) \, p(\mathcal{C}_{\text{rest}} \| \mathcal{D}_{\text{rest}} \cup \mathcal{C}) \\
& \times p(\mathcal{A} \| \mathcal{A} \cup \mathcal{D}_{\mathcal{A}} \cup \mathcal{C}) \, p(\mathcal{C}_{\mathcal{A}} \| \mathcal{A} \cup \mathcal{D}_{\mathcal{A}} \cup \mathcal{C}) \, p(\mathcal{D}_{\mathcal{A}} \| \mathcal{A} \cup \mathcal{D}_{\mathcal{A}} \cup \mathcal{D}_{\text{rest}} \cup \mathcal{C}) \\
& \times p(\mathcal{B} \| \mathcal{B} \cup \mathcal{D}_{\mathcal{B}} \cup \mathcal{C}) \, p(\mathcal{C}_{\mathcal{B}} \| \mathcal{B} \cup \mathcal{D}_{\mathcal{B}} \cup \mathcal{C}) \, p(\mathcal{D}_{\mathcal{B}} \| \mathcal{B} \cup \mathcal{D}_{\mathcal{B}} \cup \mathcal{D}_{\text{rest}} \cup \mathcal{C}).
\end{aligned} \tag{A.3}
$$

For better readability we postpone the justification of (A.3) to the end of this proof. Further notational convention concerns marginalization, i.e. "integrating out" a subset of variables. Let us write

$$
\int_{\mathcal{W}} [\cdots] = \int_{\mathbf{x}_{\mathcal{W}}} [\cdots] \mathrm{d}\mathbf{x}_{\mathcal{W}}.
$$

Now, we have that

$$
\begin{aligned}
p(\mathcal{A}, \mathcal{B}, \mathcal{C}) = & \int_{\mathcal{D}_{\mathcal{B}}} \int_{\mathcal{D}_{\mathcal{A}}} \int_{\mathcal{D}_{\text{rest}}} \int_{\mathcal{D}_{\mathcal{AB}}} p(\mathcal{D}_{\mathcal{AB}} \| \mathcal{V}) \, p(\mathcal{D}_{\text{rest}} \| \mathcal{D}_{\text{rest}} \cup \mathcal{C}) \, p(\mathcal{C}_{\text{rest}} \| \mathcal{D}_{\text{rest}} \cup \mathcal{C}) \\
& \times p(\mathcal{A} \| \mathcal{A} \cup \mathcal{D}_{\mathcal{A}} \cup \mathcal{C}) \, p(\mathcal{C}_{\mathcal{A}} \| \mathcal{A} \cup \mathcal{D}_{\mathcal{A}} \cup \mathcal{C}) \, p(\mathcal{D}_{\mathcal{A}} \| \mathcal{A} \cup \mathcal{D}_{\mathcal{A}} \cup \mathcal{D}_{\text{rest}} \cup \mathcal{C}) \\
& \times p(\mathcal{B} \| \mathcal{B} \cup \mathcal{D}_{\mathcal{B}} \cup \mathcal{C}) \, p(\mathcal{C}_{\mathcal{B}} \| \mathcal{B} \cup \mathcal{D}_{\mathcal{B}} \cup \mathcal{C}) \, p(\mathcal{D}_{\mathcal{B}} \| \mathcal{B} \cup \mathcal{D}_{\mathcal{B}} \cup \mathcal{D}_{\text{rest}} \cup \mathcal{C}).
\end{aligned}
$$

Since $\mathcal{D}_{\mathcal{AB}}$ appears only in the first factor and nowhere in the conditions, $\int_{\mathcal{D}_{\mathcal{AB}}} p(\mathcal{D}_{\mathcal{AB}} \| \mathcal{V}) = 1$ (here we have used assertion (ii) of Proposition A.1).

Now we assume that $\mathcal{D}_{\text{rest}} = \varnothing$. The two remaining integrals over $\mathcal{D}_{\mathcal{A}}$ and $\mathcal{D}_{\mathcal{B}}$ factorise. The integral over $\mathcal{D}_{\mathcal{A}}$ depends only on $\mathcal{A}$ and $\mathcal{C}$, the other one only on $\mathcal{B}$ and $\mathcal{C}$. Thus we see that

$$\begin{aligned} p(\mathcal{A}, \mathcal{B}, \mathcal{C}) &= \psi(\mathcal{C}) \\ &\quad \times \int_{\mathcal{D}_{\mathcal{A}}} p(\mathcal{A} \| \mathcal{A} \cup \mathcal{D}_{\mathcal{A}} \cup \mathcal{C})\, p(\mathcal{C}_{\mathcal{A}} \| \mathcal{A} \cup \mathcal{D}_{\mathcal{A}} \cup \mathcal{C})\, p(\mathcal{D}_{\mathcal{A}} \| \mathcal{A} \cup \mathcal{D}_{\mathcal{A}} \cup \mathcal{C}) \\ &\quad \times \int_{\mathcal{D}_{\mathcal{B}}} p(\mathcal{B} \| \mathcal{B} \cup \mathcal{D}_{\mathcal{B}} \cup \mathcal{C})\, p(\mathcal{C}_{\mathcal{B}} \| \mathcal{B} \cup \mathcal{D}_{\mathcal{B}} \cup \mathcal{C})\, p(\mathcal{D}_{\mathcal{B}} \| \mathcal{B} \cup \mathcal{D}_{\mathcal{B}} \cup \mathcal{C}) \\ &= \psi(\mathcal{C}) \times \psi_1(\mathcal{A}, \mathcal{C}) \times \psi_2(\mathcal{B}, \mathcal{C}). \end{aligned}$$

Thus we have shown that $\mathcal{A} \perp_d \mathcal{B} | \mathcal{C}$ and $\mathcal{D}_{\text{rest}} = \varnothing$ imply $\mathbf{X}_{\mathcal{B}} \perp\!\!\!\perp \mathbf{X}_{\mathcal{A}} | \mathbf{X}_{\mathcal{C}}$.

If $\mathcal{D}_{\text{rest}} \neq \varnothing$ let $\mathcal{B}' = \mathcal{B} \cup \mathcal{D}_{\text{rest}}$ and $\mathcal{D}'_{\text{rest}} = \{v \notin \mathcal{A} \cup \mathcal{B}' \cup \mathcal{C} \colon \mathcal{A} \not\perp v \text{ and } \mathcal{B}' \not\perp v\}$. Note that $\mathcal{D}'_{\text{rest}}$ is in fact an empty set; if $v \notin \mathcal{A} \cup \mathcal{B}' \cup \mathcal{C}$ then $v \notin \mathcal{A} \cup \mathcal{B} \cup \mathcal{C}$ and hence the conditions $\mathcal{A} \not\perp v$ and $\mathcal{B}' \not\perp v$ imply that $v \in \mathcal{D}_{\text{rest}}$ which contradicts $v \notin \mathcal{B}'$. By the definition of $\mathcal{D}_{\text{rest}}$ we have $\mathcal{A} \perp_d \mathcal{D}_{\text{rest}} | \mathcal{C}$ and hence $\mathcal{A} \perp_d \mathcal{B}' | \mathcal{C}$. As $\mathcal{D}'_{\text{rest}} = \varnothing$ our previous considerations lead to $\mathbf{X}_{\mathcal{B}'} \perp\!\!\!\perp \mathbf{X}_{\mathcal{A}} | \mathbf{X}_{\mathcal{C}}$, in particular $\mathbf{X}_{\mathcal{B}} \perp\!\!\!\perp \mathbf{X}_{\mathcal{A}} | \mathbf{X}_{\mathcal{C}}$.

What remains is to verify that (A.3) holds. We have to show that the following statements are true, provided that we have d-separation $\mathcal{A} \not\perp \mathcal{B}$:

1. Sets $\mathcal{A}, \mathcal{B}, \mathcal{D}_{\mathcal{AB}}, \mathcal{D}_{\mathcal{A}}, \mathcal{D}_{\mathcal{B}}, \mathcal{D}_{\text{rest}}, \mathcal{C}_{\mathcal{A}}, \mathcal{C}_{\mathcal{B}}, \mathcal{C}_{\text{rest}}$ are mutually disjoint.
2. $\text{pa}(\mathcal{D}_{\text{rest}}) \subseteq \mathcal{D}_{\text{rest}} \cup \mathcal{C}$,
3. $\text{pa}(\mathcal{C}_{\text{rest}}) \subseteq \mathcal{D}_{\text{rest}} \cup \mathcal{C}$,
4. $\text{pa}(\mathcal{A}) \subseteq \mathcal{A} \cup \mathcal{D}_{\mathcal{A}} \cup \mathcal{C}$ and $\text{pa}(\mathcal{B}) \subseteq \mathcal{B} \cup \mathcal{D}_{\mathcal{B}} \cup \mathcal{C}$
5. $\text{pa}(\mathcal{D}_{\mathcal{A}}) \subseteq \mathcal{A} \cup \mathcal{D}_{\mathcal{A}} \cup \mathcal{D}_{\text{rest}} \cup \mathcal{C}$ and $\text{pa}(\mathcal{D}_{\mathcal{B}}) \subseteq \mathcal{B} \cup \mathcal{D}_{\mathcal{B}} \cup \mathcal{D}_{\text{rest}} \cup \mathcal{C}$,
6. $\text{pa}(\mathcal{C}_{\mathcal{A}}) \subseteq \mathcal{A} \cup \mathcal{D}_{\mathcal{A}} \cup \mathcal{C}$ and $\text{pa}(\mathcal{C}_{\mathcal{B}}) \subseteq \mathcal{B} \cup \mathcal{D}_{\mathcal{B}} \cup \mathcal{C}$.

The argument goes as follows.

Firstly, we claim that $\text{de}(\mathcal{D}_{\mathcal{AB}}) \subseteq \mathcal{D}_{\mathcal{AB}}$. Indeed, let $v \in \mathcal{D}_{\mathcal{AB}}$. There are $\mathcal{C}$-active trails $\mathcal{A} - v$ and $v - \mathcal{B}$. Since $v \notin \mathcal{C}$ and v must block the trail between $\mathcal{A}$ and $\mathcal{B}$, we infer that the trails must create a *collider* at v and no descendant of v can belong to $\mathcal{C}$. All descendants of v are d-connected to both $\mathcal{A}$ and $\mathcal{B}$ and hence they cannot belong to $\mathcal{A}$ or $\mathcal{B}$ (as those sets are d-separated) or $\mathcal{D}_{\mathcal{A}}$ or $\mathcal{D}_{\mathcal{B}}$ (simply by their definition). Therefore all descendants of v are in $\mathcal{D}_{\mathcal{AB}}$. Consequently, no node in $\mathcal{D}_{\mathcal{AB}}$ can be a parent of a node $\notin \mathcal{D}_{\mathcal{AB}}$. This fact will be used below several times.

We now proceed to verification of conditions 1-7.

1. We only need to show that $\mathcal{C}_{\mathcal{A}} \cap \mathcal{C}_{\mathcal{B}} = \varnothing$. For any $v \in \mathcal{C}_{\mathcal{A}}$ there exists $w \in \text{pa}(v)$ such that $w \in \mathcal{A}$ or $w - \mathcal{A}$. If additionally we had $v \in \mathcal{C}_{\mathcal{B}}$ then there would exist $u \in \text{pa}(v)$ such that $u \in \mathcal{B}$ or $u - \mathcal{B}$. This would imply that a trail from $\mathcal{A}$ to $\mathcal{B}$ via v has a collider at $v \in \mathcal{C}$, which is impossible, because $\mathcal{A} \not\perp \mathcal{B}$.

2. If $v \in \mathcal{D}_{\text{rest}}$ and $w \in \text{pa}(v)$ then $w \notin \mathcal{A}$ and $w \notin \mathcal{D}_{\mathcal{A}}$, because otherwise there would be an $\mathcal{C}$-active trail v — A, contrary to the definition of $\mathcal{D}_{\text{rest}}$. Analogously, $w \notin \mathcal{B}$ and $w \notin \mathcal{D}_{\mathcal{B}}$. Therefore $w \in \mathcal{D}_{\text{rest}} \cup \mathcal{C}$.

3. Assume that $v \in \mathcal{C}_{\text{rest}}$ and $w \in \text{pa}(v)$. It is clear that $w \notin \mathcal{A} \cup \mathcal{D}_{\mathcal{A}}$ and $w \notin \mathcal{B} \cup \mathcal{D}_{\mathcal{B}}$, by definition of $\mathcal{C}_{\text{rest}}$. Therefore $w \in \mathcal{D}_{\text{rest}} \cup \mathcal{C}$.

4. If $v \in \mathcal{A}$ and $w \in \text{pa}(v)$ then clearly $w \notin \mathcal{B} \cup \mathcal{D}_{\mathcal{B}}$ because otherwise we would have an $\mathcal{C}$-active trail $\mathcal{A}$ — $\mathcal{B}$. Moreover we cannot have $w \in \mathcal{D}_{\text{rest}}$, for this would contradict the definition of $\mathcal{D}_{\text{rest}}$. Therefore $w \notin \mathcal{B} \cup \mathcal{D}_{\mathcal{B}} \cup \mathcal{D}_{\text{rest}}$ and we conclude that $w \in \mathcal{A} \cup \mathcal{D}_{\mathcal{A}} \cup \mathcal{C}$. The proof of $\text{pa}(\mathcal{B}) \subseteq \mathcal{B} \cup \mathcal{D}_{\mathcal{B}} \cup \mathcal{C}$ is analogous.

5. If $v \in \mathcal{D}_{\mathcal{A}}$ and $w \in \text{pa}(v)$ then clearly $w \notin \mathcal{B} \cup \mathcal{D}_{\mathcal{B}}$ because otherwise we would have v — $\mathcal{B}$, contrary to the definition of $\mathcal{D}_{\mathcal{A}}$. We conclude that $w \in \mathcal{A} \cup \mathcal{D}_{\mathcal{A}} \cup \mathcal{D}_{\text{rest}} \cup \mathcal{C}$. The proof of $\text{pa}(\mathcal{D}_{\mathcal{B}}) \subseteq \mathcal{B} \cup \mathcal{D}_{\mathcal{B}} \cup \mathcal{D}_{\text{rest}} \cup \mathcal{C}$ is analogous.

6. If $v \in \mathcal{C}_{\mathcal{A}}$ then $v \notin \mathcal{C}_{\mathcal{B}}$ by item 1, so no parent of v can belong to $\mathcal{B} \cup \mathcal{D}_{\mathcal{B}}$. We also know that no parent of v can belong to $\mathcal{D}_{\mathcal{AB}}$. No parent of v can belong to $\mathcal{D}_{\text{rest}}$, otherwise there would be an active trail $\mathcal{A}$ — $\mathcal{D}_{\text{rest}}$ (with a collider at $v \in \mathcal{C}$), contrary to the definition of $\mathcal{D}_{\text{rest}}$. The proof of $\text{pa}(\mathcal{C}_{\mathcal{B}}) \subseteq \mathcal{B} \cup \mathcal{D}_{\mathcal{B}} \cup \mathcal{C}$ is analogous.

We have shown that (A.3) is true and this concludes the proof. ■

Proof [Proof of Theorem 3.3 (b)] We adopt the notation from the proof of Theorem 3.9. Take any trail in $\mathcal{G}$ that connects $a \in \mathcal{A}$ to $b \in \mathcal{B}$ and is not d-blocked by $\mathcal{C}$:

$$a, e_1, v_1, e_2, v_2, e_3, \ldots, e_{n-1}, v_{n-1}, e_n, b.$$

Let us denote this trail by τ and let $\widetilde{\tau}$ be the trail in $\widetilde{\mathcal{G}}$ given by

$$(a, t_0), e_1, (v_1, t_1), e_2, (v_2, t_2), e_3, \ldots, e_{n-1}, (v_{n-1}, t_{n-1}), e_n, (b, t_n)$$

where $t_0 = n$ and $t_1, \ldots, t_n$ are defined by $e_1, \ldots, e_n$ via

$$t_{i+1} = \begin{cases} t_i + 1 & \text{if } e_{i+1} = \rightarrow \\ t_i - 1 & \text{otherwise} \end{cases} \tag{A.4}$$

For any collider $\rightarrow (v_i, t_i) \leftarrow$ on $\widetilde{\tau}$ choose $w_i \in \text{de}(v_i) \cap \mathcal{C}$ (such w_i exists as τ is d-open) and let d_k be the length of the shortest directed path from v_i to w_i in $\mathcal{G}$. Take any T larger than $\max_{i:\, \rightarrow (v_i, t_i) \leftarrow} \{t_i + d_i\}$ and $\max_{i \leqslant n} t_i$. We claim that $\widetilde{\tau}$ is an $\mathcal{C}(\leqslant T)$-active trail between $\mathcal{A}(\leqslant T)$ and $\mathcal{B}(\leqslant T)$ in $\widetilde{\mathcal{G}}$. It is straightforward to verify: if a fork, chain or collider connection is d-open in τ then its counterpart in $\widetilde{\tau}$ must be d-open in $\widetilde{\tau}$. By the classical theory of d-separation this means that there is a probability distribution that factorises along $\widetilde{G}$ in which $\mathbf{X}_{\mathcal{B}}(\leqslant T) \not\perp\!\!\!\perp \mathbf{X}_{\mathcal{A}}(\leqslant T) | \mathbf{X}_{\mathcal{C}}(\leqslant T)$; this probability distribution defines the desired process. ■

Remark *In our proof of Theorem 3.3 (a) we have not directly used the assumptions on the process* $\mathbf{X}$ *but we only have appealed to the properties stated in Proposition A.1. Consequently, the proof is valid also for those continuous time processes, for which the conclusion of Proposition A.1 holds true. An important example is the class of continuous time Bayesian networks (CTBNs).*