Proceedings of Machine Learning Research vol 213:1–19, 2023 2nd Conference on Causal Learning and Reasoning

Local Dependence Graphs for Discrete Time Processes

Wojciech Niemiro WNIEM@MIMUW.EDU.PL and **Łukasz Rajkowski** L.RAJKOWSKI@UW.EDU.PL
Faculty of Mathematics, Informatics and Mechanics, University of Warsaw

Editors: Mihaela van der Schaar, Dominik Janzing and Cheng Zhang

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. Even in the discrete time setting some natural questions relating the conditional (in)dependence statements in the stochastic process to separation properties of the underlying local dependence graph are scattered throughout the literature. We provide an unifying view and fill in certain gaps. In this paper we examine graphical characteristics for two kinds of conditional independences: those occurring in Markov chains under the stationary regime and independences 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 dependences. 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) dependences. Research on CMPs was pursued by Nodelman et al. (2002) (who independently introduced these processes under the name of CTBNs) and by Didelez (2007a). 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, as in Eichler and Didelez (2010) and Eichler (2007). Local dependence graphs for such processes were considered e.g. in Eichler and Didelez (2007) and Eichler and Didelez (2010). In a sense, investigation of a discrete time graphical model can be reduced to considering a space-time graph representing a dynamic Bayesian network (DBN), cf. Figure 1. The structure of conditional independences in such a model is widely regarded as perfectly understood. Nonetheless, in our paper we point out several facts that are scattered throughout the literature and we provide an unifying view. We give simple independent proofs of "Markov properties of the graph of local dependence" concerning some natural questions. Put differently, Markov properties translate probabilistic statements about conditional (in)dependence into separability characteristics of the underlying graph. In particular we describe conditional dependence of the whole sub-processes under the assumption of independent initialisation (Theorem 3.2) and the dependence of the future of some variables on the past of other variables (Theorem 3.9). These results are already present in the literature (e.g. Eichler (2007, Theorem 4.5)) but usually without the 'only if' part. Moreover the role of the initial distribution is often overlooked. We show that the assumption of stationary initialisation leads to a different graphical characterisations of the independences of whole subprocesses (Theorem 3.4). To our knowledge this last result is new.

We conjecture that most our results can be generalised to the case of continuous time processes (under a suitable definition of the local dependence graph). For example, 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.

The results presented in this paper can be used in designing the tests of local independence that are crucial for structure learning algorithms.

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 $(v, w) \in \mathcal{E}$; we write $v \not\to w$ otherwise. 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 $\mathbb{S}_v$ and $t > 0$:

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

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 some continuous time processes (e.g. Continuous Time Bayesian Networks).

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.

$$(X_v(t), v \in \mathcal{V}) \text{ are conditionally independent given } (X(s) : s < t) \text{ for } t > 0, \quad (2)$$

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

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 "contemporaneous local dependence" and basically boils down to the assumption that our model contains all relevant "confounding variables".[2]

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

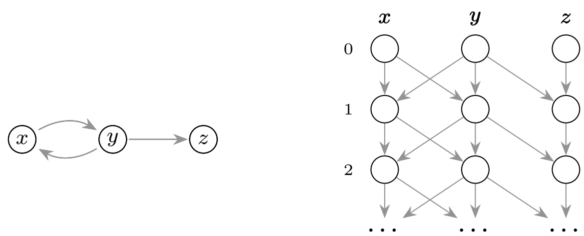


Figure 1: A very simple local dependence graph on the set $\mathcal{V} = \{x, y, z\}$ (left panel) and its *space-time* graph with the set of vertices $\mathcal{V} \times \{0, 1, 2, \ldots\}$ and the set of edges of the form $(v, s) \to (w, t)$ where $s < t$ and $v \to w$ is an edge of the local dependence graph. For better readability only edges $(v, t-1) \to (w, t)$ are shown and self-loops in the local dependence graph are not indicated.

3. Conditional Independence Properties

We are going to examine conditional independence statements about process $\mathbf{X}$ and to find corresponding statements in terms of graph $\mathcal{G}$.

We will have to consider several types of separability relations in graph $\mathcal{G}$. Some of them are well-known, whereas the notion of c-separation (Definition 3.3) is new. To begin with, we define a *trail*[4] 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 $e_i = (v_{i-1} \leftarrow v_i)$.[5] Let v_i be a non-end node in the trail, that is $i \neq 0$ and $i \neq n$. We say that

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)).
3. The term "local" refers here to time adjacency and not to space adjacency. We follow the terminology introduced by Didelez (2007a, 2008) and prevailing in the literature, despite the fact that this terminology might be misleading.
4. What we call a trail is sometimes called a *path* in the literature, but in our opinion this last name is better suited for directed connections.
5. Definition of a *trail* requires some caution, because there are subtle differences between several definitions appearing in the literature.

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:

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

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 $\mathrm{de}(v) \cap \mathcal{C} = \varnothing$.*

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

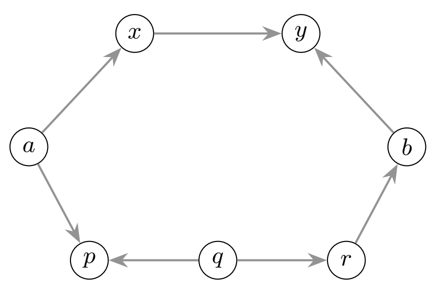


Figure 2

Example *In the graph shown on Figure 2 $\{b\}$ is d-blocked from $\{a\}$ by $\varnothing$ or $\{x, y, p, r\}$. It is not d-blocked by $\{y\}$.*

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

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

Theorem 3.2

(a) *Consider a discrete time process* $\mathbf{X}$ *which satisfies* (1) *and* (4) *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* (1) *and* (4) *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 (2007a, 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 and can be easily deduced from an analogous result for DAGs by analysing the space-time graph, we were not able to find an exact reference. Therefore we provide a self-contained proof in the Appendix A (it is essentially identical to the proof given in (Koski and Noble, 2009, Theorem 2.2) for the case of DAGs). This proof covers also the case of continuous time Markov processes (cf. the last Remark in the Appedix A).

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 independences than under independent initial distribution. We proceed to define a notion of c-separation which describes conditional independences 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* $\to v \leftarrow$ *with* $\mathrm{de}(v) \cap \mathcal{C} = \varnothing$.

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

Example *In the graph shown on Figure 2* $\{b\}$ *is* c*-blocked from* $\{a\}$ *by any subset of* $\{x, q, r\}$. *It is not* c*-blocked by* $\{x, y, p, r\}$ *(note that it is* d*-blocked by this set).*

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 *and* $X(0) \sim \pi$, *that is the process* $\mathbf{X}$ *is strictly stationary. If* $\mathcal{A}$ *is* c*-blocked from* $\mathcal{B}$ *by* $\mathcal{C}$ *then*

- $X_\mathcal{A}(t) \perp\!\!\!\perp X_\mathcal{B}(t) | X_\mathcal{C}(t)$ *for all* $t \geqslant 0$.
- $\mathbf{X}_\mathcal{A} \perp\!\!\!\perp \mathbf{X}_\mathcal{B} | \mathbf{X}_\mathcal{C}$.

6. The role of the initial distribution is sometimes overlooked, as e.g. in the cited above Proposition 5 in Didelez (2007a).

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 in the following weak sense: 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 describe 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 conditions* (1) *and* (4)*. 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 $\mathrm{pa}(\mathcal{B}) \cap \mathcal{A} = \varnothing$, then $X_{\mathcal{B}}(t) \perp\!\!\!\perp \mathbf{X}_{\mathcal{A}}(< t) | \mathbf{X}_{\mathcal{V}\setminus\mathcal{A}}(< t)$.*

To find graph statements corresponding 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 $\mathrm{de}(v) \cap \mathcal{C} = \varnothing$.*

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

Example *In our running example from Figure 2 $\{b\}$ is δ-blocked from $\{a\}$ by $\{y\}$ (whereas it is not d-blocked by this set). The only non-δ-blocking set is $\{p\}$.*

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) *and* (4) *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 (1) *and* (4) *such that $X_{\mathcal{B}}(t) \not\perp\!\!\!\perp \mathbf{X}_{\mathcal{A}}(< t) | \mathbf{X}_{\mathcal{B}\cup\mathcal{C}}(< t)$.*

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 appears (without a name) in (Eichler and Didelez, 2007, Theorem 4.8).

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

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

Example *In the graph shown on Figure 2 $\{b\}$ is ε-blocked from $\{a\}$ by $\{y\}$ (but it is not d-blocked by this set). It is not ε-blocked by $\{p, r\}$ (although it is d-blocked).*

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 $\mathrm{de}(v) \cap \mathcal{C} \neq \varnothing$), 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).

Note that $\mathcal{A} \not\rightarrow_{\varepsilon} \mathcal{B}|\mathcal{C}$ is equivalent to $\mathcal{A} \not\rightarrow_{\delta} \mathrm{an}(\mathcal{B})|\mathcal{C}$. Part (a) of the following Theorem can be easily recognized as (Eichler and Didelez, 2007, Theorem 4.8), part (b) is new.

Theorem 3.9

(a) Consider a process $\mathbf{X} = (X_v(t))$ satisfying (1) and (4) 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 (1) and (4) such that $X_{\mathcal{B}}(\geqslant t) \not\perp\!\!\!\perp \mathbf{X}_{\mathcal{A}}(< t)|\mathbf{X}_{\mathcal{B}\cup\mathcal{C}}(< t)$.

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 $\mathcal{V}\backslash\mathcal{C}_2$.

In this proof the phrases 'open'/'blocked' always refer to Definition 3.3.

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

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

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

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

Iterating the above equation we obtain

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

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 $X_{\mathcal{A}_2}$, $X_{\mathcal{B}_2}$ 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 *space-time* 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 $\mathcal{G}$ (cf. Figure 1). 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 $t \in \mathbb{N}_0$. For now on we refer to times less that t as 'the past' and to the remaining times as 'the future'. 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 $(\mathcal{B} \cup \mathcal{C})(< t)$.

We show that this leads to a contradiction, which by the classical theory of d-separation implies $\mathbf{X}_{\mathcal{B}}(\geqslant t) \perp\!\!\!\perp \mathbf{X}_{\mathcal{A}}(< t) | \mathbf{X}_{\mathcal{B}\cup\mathcal{C}}(< t)$.

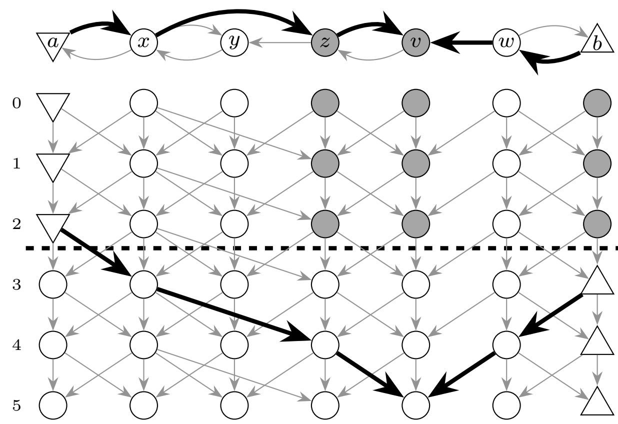
Clearly

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

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

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

Figure 3: Local dependence graph on 7 nodes (the first line) and its space-time graph (for better readability only edges $(v, s) \to (w, s+1)$ are included). Nodes in gray are observed up to time $t = 3$ (excluded). The lower path cannot be a d-open path between $a(< t)$ and $b(\geqslant t)$ since it has an arrow $\leftarrow$ in the future (e.g. $(w, 4) \leftarrow (b, 3)$) which inevitably leads to a collider to the left of it ($(z, 4) \to (v, 5) \leftarrow (w, 4)$) whose center is by necessity unobserved with all its descendands (as they belong to the future) and hence it is d-closed.



(cf. Figure 3). 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 $\to (v_m, t_m) \leftarrow$ 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 d-block $\widetilde{\tau}$.

In particular from (6) we obtain $e_n =\to$ 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 $\to b$. 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 $\to v_k \leftarrow$ 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 $\to (v_k, t_k) \leftarrow$ d-blocks $\widetilde{\tau}$, which is a contradiction. For example, in Figure 3 no connection of the form $(x, s) \to (y, s+1) \leftarrow (z, s)$ can be d-open since neither $(y, s+1)$ nor any of its descendants is in $(\mathcal{B} \cup \mathcal{C})(< t)$.

- There is a fork $\leftarrow v_k \to$ with $v_k \in \mathcal{C}$. Therefore $\leftarrow (v_k, t_k) \to$ 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).

  For instance, in Figure 3 if $(y, s) \leftarrow (z, s-1) \to (v, s)$ is d-open then it necessarily happens in the non-observed future, but it impossible in the view of (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 $\to v_k \to$ with $v_k \in \mathcal{C}$ and a collider $\to v_l \leftarrow$ 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. $e_i =\to$ for $i = k+1, \ldots, l$). This means that $t_k \leqslant t_l$. On the other hand, $\to (v_k, t_k) \to$ 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 $\to (v_l, t_l) \leftarrow$ 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.

> E.g. in Figure 3 the connection $(x, s) \to (z, s+1) \to (v, s+2) \leftarrow (w, s+1)$ cannot be d-open since that would mean that the center of the chain connection is in the future and the center of the collider connection is in the past.

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 finishes the proof.

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 $e_n = \to$. 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

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

For any collider $\to (v_i, t_i) \leftarrow$ 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 $t = t_0 + \max_{i:\, \to (v_i, t_i) \leftarrow} \{t_i + d_i\}$ (any t that is larger from t_0 and $\max_{i:\, \to(v_i,t_i)\leftarrow} \{t_k + d_k\}$ will do).

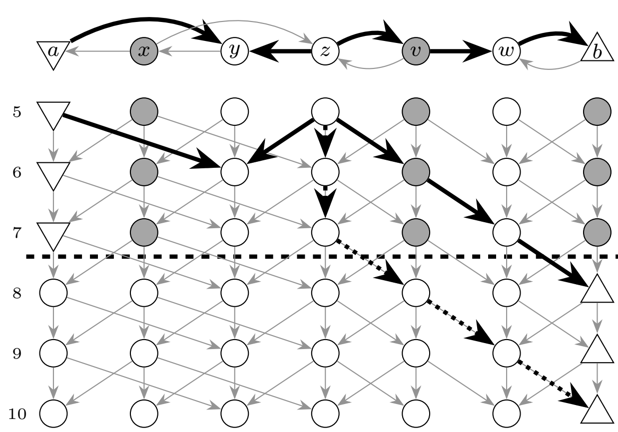


Figure 4: An exemplary ε-open path τ in local dependence graph (upper diagram) and its counterpart $\widetilde{\tau}$ in the space-time graph. The start time $t_0 = 5$ is chosen so that $\widetilde{\tau}$ has no chance to stop at negative time. The cut-off time $t = 8$ is chosen so that the descendant $(x, 7)$ of the collider $(y, 6)$ is observed. Despite the fact that the collider is d-open, the path $\widetilde{\tau}$ is d-closed because of the chain $(z, 5) \to (v, 6) \to (w, 7)$ because it has an observed center. This is why we 'bypass it' (dotted arrows) to obtain a d-open path $\widehat{\tau}$ from $a(< t)$ to $b(\geqslant t)$

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 $(v_{r-1}, t_{r-1}) \to (v_r, t_r)$ with

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

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 $e_l = \to$ 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 $\hat{\tau}$ 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 $\hat{\tau}$ 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 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 $\mathcal{C}$	$\mathcal{A} \perp_d \mathcal{B} \vert \mathcal{C}$	$\mathbf{X}_{\mathcal{B}} \perp\!\!\!\perp \mathbf{X}_{\mathcal{A}} \vert \mathbf{X}_{\mathcal{C}}$ (independent start)
(2)	Every trail between $\mathcal{A}$ and $\mathcal{B}$ is c-blocked by $\mathcal{C}$	$\mathcal{A} \perp_c \mathcal{B} \vert \mathcal{C}$	$\mathbf{X}_{\mathcal{B}} \perp\!\!\!\perp \mathbf{X}_{\mathcal{A}} \vert \mathbf{X}_{\mathcal{C}}$ (stationary)
(3)	No arrows from $\mathcal{A}$ to $\mathcal{B}$	$\mathcal{A} \not\rightarrow \mathcal{B}$	$X_{\mathcal{B}}(t) \perp\!\!\!\perp \mathbf{X}_{\mathcal{A}}(< t) \vert \mathbf{X}_{\mathcal{V} \setminus \mathcal{A}}(< t)$
(4)	Every trail from $\mathcal{A}$ to $\mathcal{B}$ is δ-blocked by $\mathcal{C}$	$\mathcal{A} \not\rightarrow_\delta \mathcal{B} \vert \mathcal{C}$	$X_{\mathcal{B}}(t) \perp\!\!\!\perp \mathbf{X}_{\mathcal{A}}(< t) \vert \mathbf{X}_{\mathcal{B} \cup \mathcal{C}}(< t)$
(5)	No directed paths from $\mathcal{A}$ to $\mathcal{B}$	$\mathcal{A} \not\rightarrow \mathcal{B}$	$\mathbf{X}_{\mathcal{B}}(\geqslant t) \perp\!\!\!\perp \mathbf{X}_{\mathcal{A}}(< t) \vert \mathbf{X}_{\mathcal{V} \setminus \mathcal{A}}(< t)$
(6)	Every trail from $\mathcal{A}$ to $\mathcal{B}$ is ε-blocked by $\mathcal{C}$	$\mathcal{A} \not\rightarrow_\varepsilon \mathcal{B} \vert \mathcal{C}$	$\mathbf{X}_{\mathcal{B}}(\geqslant t) \perp\!\!\!\perp \mathbf{X}_{\mathcal{A}}(< t) \vert \mathbf{X}_{\mathcal{B} \cup \mathcal{C}}(< t)$

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

Note that among presented separation types some imply others. It is an easy exercise to show the implications presented on the following diagram.

$$\begin{array}{ccc} d\text{-separation} & \Leftarrow & c\text{-separation} \\ \Downarrow & & \Downarrow \\ \delta\text{-separation} & \Leftarrow & \epsilon\text{-separation} \end{array}$$

No other implication is true. This leaves us with only six possible configurations of these separation types. All of them are represented in the table below. The last row contains blocking sets $\mathcal{C}$ that result in consecutive configurations in the graph from Figure 2 (where $\mathcal{A} = \{a\}$ and $\mathcal{B} = \{b\}$).

d	✓	✓	✓	✗	✗	✗
c	✓	✗	✗	✗	✗	✗
δ	✓	✓	✓	✓	✓	✗
ε	✓	✓	✗	✓	✗	✗
	$\varnothing$	$\{x, y\}$	$\{p, r\}$	$\{y\}$	$\{p, r, y\}$	$\{p\}$

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, 2007a.

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

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. Granger causality and path diagrams for multivariate time series. *Journal of Econometrics*, 137(2):334–353, 2007.

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

Michael Eichler and Vanessa Didelez. On Granger causality and the effect of interventions in time series. *Lifetime data analysis*, 16:3–32, 2010.

Yu Fan, Jing Xu, and Christian R Shelton. Importance sampling for continuous time Bayesian networks. *The Journal of Machine Learning Research*, 11:2115–2140, 2010.

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

Timo Koski and John Noble. *Bayesian networks: an Introduction*. John Wiley & Sons, 2009.

Błażej Miasojedow and Wojciech Niemiro. Geometric ergodicity of Rao and Teh's algorithm for markov jump processes and CTBNs. *Electronic Journal of Statistics*, 11:4629–4648, 2017.

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 $\mathrm{de}(v) \cap \mathcal{C} = \varnothing$.*

(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 3.2 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). Note that $\mathcal{U} \not— \mathcal{W}$ is equivalent to $\mathcal{U} \perp_d \mathcal{W} | \mathcal{C}$ but in our opinion the former is more readable in the proof.

Proof [Proof of Theorem 3.2 (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

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

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 $p(\mathbf{x}_{\mathcal{A}} | \mathbf{x}_{\mathrm{pa}(\mathcal{A}) \setminus \mathcal{A}})$.

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 $\mathcal{A} \not\to \mathcal{B}$). 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

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

Thus we have shown that $\mathcal{A} \not\leftrightarrow \mathcal{B}$ 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 $\mathcal{D}'_{\text{rest}} = \{v \notin \mathcal{A} \cup \mathcal{B}' \cup \mathcal{C} \colon \mathcal{A} \not\leftrightarrow v \text{ and } \mathcal{B}' \not\leftrightarrow v\}$. 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 $\mathcal{A} \not\leftrightarrow v$ and $\mathcal{B}' \not\leftrightarrow v$ imply that $v \in \mathcal{D}_{\text{rest}}$ which contradicts $v \notin \mathcal{B}'$. By the definition of $\mathcal{D}_{\text{rest}}$ we have $\mathcal{A} \not\leftrightarrow \mathcal{D}_{\text{rest}}$ and hence $\mathcal{A} \not\leftrightarrow \mathcal{B}'$. 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 $\mathcal{A} \not\leftrightarrow \mathcal{B}$:

1. Sets $\mathcal{A}, \mathcal{B}, \mathcal{D}_{\mathcal{A}\mathcal{B}}, \mathcal{D}_{\mathcal{A}}, \mathcal{D}_{\mathcal{B}}, \mathcal{D}_{\text{rest}}, \mathcal{C}_{\mathcal{A}}, \mathcal{C}_{\mathcal{B}}, \mathcal{C}_{\text{rest}}$ 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 $\text{de}(\mathcal{D}_{\mathcal{A}\mathcal{B}}) \subseteq \mathcal{D}_{\mathcal{A}\mathcal{B}}$. Indeed, let $v \in \mathcal{D}_{\mathcal{A}\mathcal{B}}$. There are $\mathcal{C}$-active trails $\mathcal{A} \longrightarrow v$ and $v \longrightarrow \mathcal{B}$. 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 $\mathcal{D}_{\mathcal{A}\mathcal{B}}$. Consequently, no node in $\mathcal{D}_{\mathcal{A}\mathcal{B}}$ can be a parent of a node $\notin \mathcal{D}_{\mathcal{A}\mathcal{B}}$. 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 $w \longrightarrow \mathcal{A}$. 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 $u \longrightarrow \mathcal{B}$. 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 $\mathcal{A} \not\leftrightarrow \mathcal{B}$.
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 $v \longrightarrow A$, 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 $\mathcal{A} \longrightarrow \mathcal{B}$. 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 $w \in \mathrm{pa}(v)$ 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 $\mathrm{pa}(\mathcal{D}_\mathcal{B}) \subseteq \mathcal{B} \cup \mathcal{D}_\mathcal{B} \cup \mathcal{D}_{\text{rest}} \cup \mathcal{C}$ 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 $\mathrm{pa}(\mathcal{C}_\mathcal{B}) \subseteq \mathcal{B} \cup \mathcal{D}_\mathcal{B} \cup \mathcal{C}$ is analogous.

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

Proof [Proof of 3.2 (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 $\tilde{\tau}$ be the trail in $\tilde{\mathcal{G}}$ 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 $\tilde{\tau}$ choose $w_i \in \mathrm{de}(v_i) \cap \mathcal{C}$ (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 $\tilde{\tau}$ is an $\mathcal{C}(\leqslant T)$-active trail between $\mathcal{A}(\leqslant T)$ and $\mathcal{B}(\leqslant T)$ in $\tilde{\mathcal{G}}$. It is straightforward to verify: if a fork, chain or collider connection is d-open in τ then its counterpart in $\tilde{\tau}$ must be d-open in $\tilde{\tau}$. By the classical theory of d-separation this means that there is a probability distribution that factorises along $\tilde{G}$ 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 3.2 (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). The factorisation formula from Proposition A.1 (i) for CTBNs can be found e.g. in Nodelman et al. (2002), Fan et al. (2010) and (Miasojedow and Niemiro, 2017, eq. (12)). A version of such formula is also in (Didelez, 2007b, p. 183, formula (9)). One should be cautious though, because 'likelihood' is defined up to multiplicative constants and with initial distribution being ignored. Thus factorisation for* $p(\mathbf{x})$ *in Proposition A.1 (i) follows directly but Proposition A.1 (ii) requires careful examination. We think this second part of Proposition A.1 is needed in a proof of an analogue of Theorem 3.2 for CTBNs and this point seems to be overlooked in the literature.*