Weak equivalence of local independence graphs

Søren Wengel Mogensen

Department of Automatic Control, Lund University

Abstract

Classical graphical modeling of multivariate random vectors uses graphs to encode conditional independence. In graphical modeling of multivariate stochastic processes, graphs may encode so-called local independence analogously. If some coordinate processes of the multivariate stochastic process are unobserved, the local independence graph of the observed coordinate processes is a directed mixed graph (DMG). Two DMGs may encode the same local independences in which case we say that they are Markov equivalent.

Markov equivalence is a central notion in graphical modeling. We show that deciding Markov equivalence of DMGs is coNP-complete, even under a sparsity assumption. As a remedy, we introduce a collection of equivalence relations on DMGs that are all less granular than Markov equivalence and we say that they are weak equivalence relations. This leads to feasible algorithms for naturally occurring computational problems related to weak equivalence of DMGs. The equivalence classes of a weak equivalence relation have attractive properties. In particular, each equivalence class has a greatest element which leads to a concise representation of an equivalence class. Moreover, these equivalence relations define a hierarchy of granularity in the graphical modeling which leads to simple and interpretable connections between equivalence relations corresponding to different levels of granularity.

1 Introduction

The distribution of a multivariate random vector, $(X^\alpha)_{\alpha \in V}$, induces an *independence model*, $\mathcal{I}$, which is simply the collection of triples, (A, B, C), such that X^A and X^B are conditionally independent given X^C. Graphs are often used as convenient representations of such independence models (Lauritzen, 1996; Maathuis et al., 2019). The graphical theory reflects the fact that conditional independence is *symmetric* in A and B, i.e., $(A, B, C) \in \mathcal{I}$ if and only if $(B, A, C) \in \mathcal{I}$. In graphical modeling of multivariate stochastic processes, it is useful to apply a notion of independence that distinguishes between past and present and for this purpose several authors have used *local independence*, analogously to how conditional independence is used in classical graphical modeling. However, local independence is not symmetric in the above sense and its graphical representation therefore requires a specialized framework. Local independence was first introduced by Schweder (1970) in composable Markov processes and later studied by Aalen (1987) in a broader class of stochastic processes. Didelez (2000, 2008) described graphical modeling of marked point processes based on local independence and Mogensen et al. (2018) extended this theory to Itô processes.

Graphs are said to be *Markov equivalent* if they represent the same independences, i.e., if they are indistinguishable when observing only the induced independences. Several characterizations of Markov equivalence are available in different classes of graphs representing classical conditional independence (Frydenberg, 1990; Verma and Pearl, 1990b; Spirtes and Verma, 1992; Andersson et al.,

1997a,b; Richardson, 1997; Andersson et al., 2001; Zhao et al., 2005; Zhang, 2007; Ali et al., 2009). Mogensen and Hansen (2020) used *directed mixed graphs* as representations of local independences in partially observed stochastic processes and they characterized Markov equivalence in this class of graphs by proving that each equivalence class contains a greatest element. Their equivalence result also provided a simple approach to visualizing and understanding an entire equivalence class. Mogensen and Hansen (2022) characterized Markov equivalence of *directed correlation graphs* representing local independence in the presence of correlated noise processes. Recent work studied local independence testing in point processes (Thams and Hansen, 2021) and Christgau et al. (2022) described nonparametric tests of local independence. It is worth noting that local independence is a continuous-time version of discrete-time Granger causality which has been used in graphical models of time series (Eichler and Didelez, 2007, 2010; Eichler, 2012, 2013). The graphical theory of directed mixed graphs and the results in this paper may be applied in both continuous-time and discrete-time stochastic processes (Mogensen and Hansen, 2020, supplementary material).

In graphs representing classical conditional independence, several characterizations of Markov equivalence lead to polynomial-time algorithms for deciding Markov equivalence (e.g., Richardson, 1997; Ali et al., 2009). In the local independence framework, Mogensen and Hansen (2022) proved that deciding Markov equivalence of two directed correlation graphs is coNP-complete which means that we should not expect to find a polynomial-time algorithm in this case. In this paper, we show that deciding Markov equivalence of directed mixed graphs is also coNP-complete. We further show that assuming sparsity of the directed mixed graphs does not generally remedy this. Our results imply that several computational problems that occur naturally when using directed mixed graphs are also computationally hard. For this reason, Markov equivalence in partially observed local independence graphs may not always be a practical notion. Instead, we introduce a class of *weak equivalence relations* between local independence graphs. We characterize the corresponding equivalence classes and show that they too contain a greatest element. Mogensen and Hansen (2020) argued that the existence of a greatest element leads to a straightforward Markov equivalence theory. We extend this theory to the more general weak equivalences studied in this paper. This allows a simple representation of weak equivalence classes. A subset of the weak equivalence relations may be understood as creating a *hierarchy* of equivalence relations in which a parameter, k, creates a trade-off between the size of the equivalence classes and the computational complexity, leading to a graphical theory which is both useful and practical. This hierarchy also illustrates interpretable connections between equivalence classes across different values of k.

The paper is structured in the following way. In Section 2, we introduce necessary terminology and notation. We also describe *global Markov properties* that connect so-called *μ-separation* in graphs to local independence and provide justification for using graphs as representations of local independence. Moreover, we give an example to illustrate the framework and purpose of the paper. In Section 3, we prove that deciding Markov equivalence of directed mixed graphs is computationally hard, even under sparsity restrictions, and we discuss the implications of this result. In Section 4, we introduce the notion of *weak equivalence* of graphs. We describe its properties and compare it with Markov equivalence. Section 5 proves that, under a regularity condition, every weak equivalence class has a greatest element. Using the main result from the previous section, Section 6 first describes a graph which concisely represents an entire equivalence class. It then describes a *hierarchy* of certain weak equivalence classes and how they represent different levels of granularity in their description of the underlying graphs. Section 7 discusses algorithmic aspects of weak equivalence, and in Section 8 we briefly outline how results from the previous sections relate to graphical structure learning. Section 9 provides a discussion of the results.

2 Local independence and graphs

The interest in μ-separation arises from its connection to local independence as formalized through various *global Markov properties*. We start by defining local independence following the exposition in Christgau et al. (2022). We will give the definition for counting processes, though, it can be extended to other classes of stochastic processes (Didelez, 2008; Mogensen et al., 2018; Mogensen and Hansen, 2022).

We consider a multivariate counting processes, $N_t = (N_t^1, \ldots, N_t^n)$, on a probability space, $(\Omega, \mathbb{F}, P)$, and we assume that N_t is observed over some interval $[0, T]$. We let V denote the set $\{1, 2, \ldots, n\}$. We use $\mathcal{F}_t^D$ to denote the right-continuous and complete filtration generated by $N_t^D = (N_t^\alpha : \alpha \in D)$. One can think of $\mathcal{F}_t^D$ as consisting of the information in the coordinate processes in $D \subseteq V$ up until time point t. For $\beta \in V$ and $C \subseteq V$, we assume that N_t^β has a $\mathcal{F}_t^C$-intensity, $\lambda_t^{\beta,C}$. The stochastic process $\lambda_t^{\beta,C}$ is $\mathcal{F}_t^C$-predictable and $N_t^\beta - \int_0^t \lambda_s^{\beta,C} \mathrm{d}s$ is a local $\mathcal{F}_t^C$-martingale.

Definition 2.1 (Local independence). *Let $\alpha, \beta \in V$ and let $C \subseteq V$. We say that N_t^β is* locally independent *of N_t^α given N_t^C (or simply, that β is locally independent of α given C) if the local $\mathcal{F}_t^C$-martingale as defined above is also a local $\mathcal{F}_t^{C \cup \{\alpha\}}$-martingale. For $A, B, C \subseteq V$, we say that B is locally independent of A given C if β is locally independent of α given C for all $\alpha \in A$ and $\beta \in B$, and we denote this by $A \not\rightarrow B \mid C$.*

Christgau et al. (2022) use the term *conditional local independence* instead of local independence which highlights the fact that Definition 2.1 is analogous to classical conditional independence of random variables. Intuitively, when β is locally independent of α given C, observation of the α-process over the interval $[0, t]$ does not provide additional information other than that contained in $\mathcal{F}_{t-}^C$ when trying to predict if there will be an event in process β in the interval $[t, t + \mathrm{d}t)$.

Local independence was first used by Schweder (1970) in composable Markov processes and later studied by Aalen (1987). Didelez (2000, 2008) described graphical modeling based on local independence. Other work on local independence Markov properties go into more detail (Didelez, 2000, 2008; Mogensen et al., 2018; Mogensen and Hansen, 2022).

Definition 2.2 (Local independence graph). *We consider a multivariate counting process, $N_t = (N_t^1, \ldots, N_t^n)$, as above, $V = \{1, \ldots, n\}$. Its* local independence graph *is the directed graph, $\mathcal{D}$, on nodes V such that*

$$\alpha \not\rightarrow \beta \text{ in } \mathcal{D} \Leftrightarrow \alpha \not\rightarrow \beta \mid V \setminus \{\alpha\}$$

for $\alpha, \beta \in V$ where $\alpha \not\rightarrow \beta$ indicates the absence of the directed edge from α to β.

The statement $\{\alpha\} \not\rightarrow \{\beta\} \mid V \setminus \{\alpha\}$ denotes that β is locally independent of α given $V \setminus \{\alpha\}$, and above we have simply written the singletons $\{\alpha\}$ and $\{\beta\}$ as α and β, respectively. The implication from left to right in Definition 2.2 is known as the *pairwise Markov property*. When this property holds, we see that the absence of an edge implies a local independence. The *global Markov property* allows one to read off more general local independences from a local independence graph using δ- or μ-separation (Definition 2.5). This is similar to other classes of graphical models (Maathuis et al., 2019). Several results state conditions for the equivalence of pairwise and global Markov properties (Didelez, 2008; Mogensen et al., 2018).

Local independence is a continuous-time analogue of Granger causality in discrete-time stochastic processes. The results of this paper also applies to Granger-causal graphs, see, e.g., the supplementary material of Mogensen and Hansen (2020) and Eichler (2007).
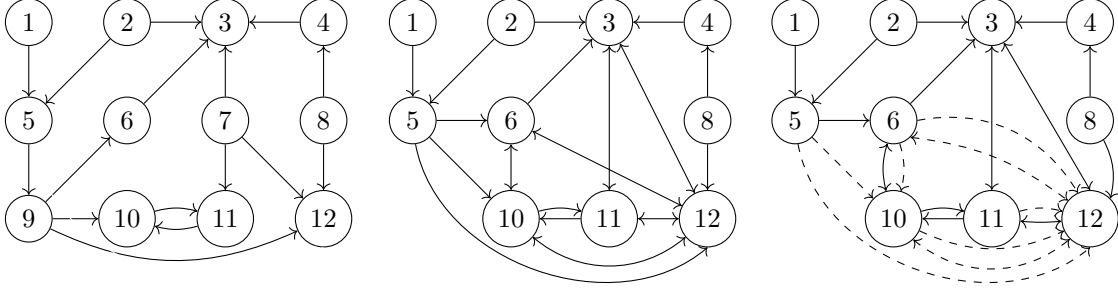
2.1 Alarm network

Figure 1: Graphs from the example in Subsection 2.1. Loops are omitted from the visualization. Left: underlying local independence graph (directed graph, $\mathcal{D}$,) representing the alarm network. Middle: latent projection, $\mathcal{G}$, of the graph $\mathcal{D}$ when processes 7 and 9 are unobserved. The graph $\mathcal{G}$ is a *directed mixed graph* and it represents the partially observed alarm network. The bidirected edges, e.g., $6 \leftrightarrow 12$, represent correlation mediated by unobserved nodes, e.g., $6 \leftarrow 9 \rightarrow 12$. Unobserved directed paths may create new edges, e.g., $5 \rightarrow 6$ in $\mathcal{G}$ corresponds to $5 \rightarrow 9 \rightarrow 6$ in $\mathcal{D}$. Right: the directed mixed equivalence graph of the Markov equivalence class containing $\mathcal{G}$.

We describe an example application based on modeling how alarms propagate through a complex industrial system. Example data is in Figure 2. In this industrial system, a number of *process variables* (e.g., temperatures and pressures) are measured repeatedly. Each process variable corresponds to an *alarm process*, and if a measured process is outside the normal range of operations an event occurs in the corresponding alarm process. The stochastic system is described by a 12-dimensional counting process, N_t^V,

$$V = \{\text{A1}, \text{A2}, \text{A3}, \text{A4}, \text{A5}, \text{A6}, \text{A7}, \text{A8}, \text{A9}, \text{A10}, \text{H}, \text{E}\},$$

observed over the interval $[0, 1]$. The coordinate processes in $V \setminus \{E\}$ are alarm processes. Process E represents exogenous events that feed into the system, e.g., changes in operating conditions, and this process is unobserved. Process H is an alarm process, but unavailable for some reason, and the observed processes are those in $V \setminus \{E, H\}$. We assume that $\mathcal{D}$ is a local independence graph in the sense of Definition 2.2. Under some regularity conditions, this implies that the global Markov property is satisfied in this graph (Didelez, 2008) and therefore μ-separation (Definition 2.5) in the graph implies local independence.

The graph $\mathcal{G}$ in Figure 1 (the *latent projection* of $\mathcal{D}$, see Section C) represents the observable local independences in the sense that for $A, B, C \subseteq V \setminus \{E, H\}$ it holds that B is μ-separated from A given C in $\mathcal{D}$ if and only if B is μ-separated from A given C in $\mathcal{G}$. The underlying graph of the full system, $\mathcal{D}$, is a *directed graph* while the latent projection is a *directed mixed graph*. In general, this larger class of graphs is needed to represent the local independences of partially observed multivariate stochastic processes.
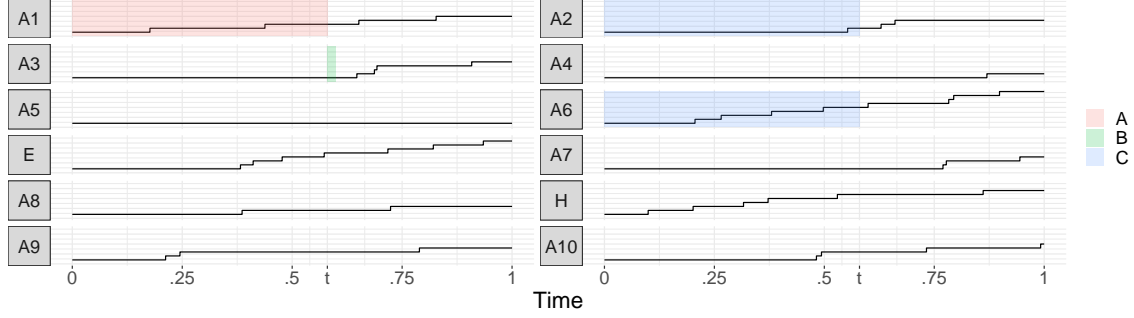
4

Figure 2: Example data from the alarm network in Subsection 2.1. Each jump of a step function corresponds to an event in the corresponding counting process. Under the global Markov property, the local independence graph $\mathcal{D}$ in Figure 1 implies that A3 (B) is locally independent of A1 (A) given $\{A2, A6\}$ (C). The meaning of this is illustrated in the above plot. We are trying to predict if there is an event in process A3 in the infinitesimal interval $[t, t + dt)$ (B, green), and we are asking if the information in the past of process A1 (A, red) adds anything when accounting for the past of processes $\{A2, A6\}$ (C, blue).

Local independence asks the following question. If we are to predict if processes B will have an event in the immediate future and we have the information in the past of processes C will the information in the past of proesses A add anything? This is illustrated visually in Figure 2 with $A = \{A1\}$, $B = \{A3\}$, and $C = \{A2, A6\}$. In this specific example, $\{A3\}$ is μ-separated from $\{A1\}$ given $\{A2, A6\}$ in $\mathcal{G}$ and under the global Markov property this implies that the corresponding local independence holds. Therefore, the information in the past of process $\{A1\}$ is superfluous when already accounting for the information in the past of processes $\{A2, A6\}$.

Several directed mixed graphs may induce the same μ-separations which means that they represent the same local independences. In this case, we say that they are *Markov equivalent*. The graph on the right in Figure 1 is the *directed mixed equivalence graph* of $\mathcal{G}$. It represents the entire Markov equivalence class by indicating if an edge is in every Markov equivalent graph (solid), in no Markov equivalent graph (absent), or in only some Markov equivalent graphs (dashed). This is a useful representation, but it may not be a practical one for all applications as it leads to computationally hard problems. In this paper, we trade away some of the expressive power of Markov equivalence to obtain a more feasible notion of equivalence and we show that weaker notions of equivalence remain easily interpretable.

2.2 Graphs

A *graph* is a pair (V, E) where V is a finite *node set* and E is an *edge set*. The edge set E is a disjoint union, $E = E_d \mathbin{\dot\cup} E_b$, where E_d is a set of ordered pairs, corresponding to directed edges, $\rightarrow$, and E_b is a set of unordered pairs, corresponding to bidirected edges, $\leftrightarrow$. We use $\alpha \leftrightarrow_{\mathcal{G}} \beta$ to denote that there is a bidirected edge between α and β in the graph $\mathcal{G}$, or just $\alpha \leftrightarrow \beta$ when it is clear from the context to which graph the statement refers, and we use $\alpha \rightarrow_{\mathcal{G}}$ and $\alpha \rightarrow \beta$ analogously. The definition of the node set implies that we allow multiple edges between a pair of nodes, however, the edges between two nodes α and β is always a subset of $\{\alpha \rightarrow \beta, \alpha \leftarrow \beta, \alpha \leftrightarrow \beta\}$. Moreover, $\alpha \leftrightarrow \beta$ and $\beta \leftrightarrow \alpha$ are equivalent while $\alpha \rightarrow \beta$ and $\alpha \leftarrow \beta$ are different edges. We emphasize that

the edge $\alpha \leftrightarrow \beta$ is not shorthand for the two edges $\alpha \to \beta$ and $\alpha \leftrightarrow \beta$, and the meaning of the bidirected edge is different from that of the two directed edges. This will be clear from subsequent definitions.

We use $\alpha \sim \beta$ to denote a generic edge of either type between α and β, and we say that α and β are *adjacent* in $\mathcal{G}$ when there exists an edge between them, $\alpha \sim \beta$. When there are multiple nodes on each side of the edge, $\alpha_1, \ldots, \alpha_k \sim \beta_1, \ldots, \beta_l$, this means that $\alpha_i \sim \beta_j$ for all $i = 1, \ldots, k$ and $j = 1, \ldots, l$. We separate such statements by semicolons, $\alpha_1, \ldots, \alpha_k \sim \beta_1, \ldots, \beta_l$; $\gamma_1, \ldots, \gamma_r \sim \delta_1, \ldots, \delta_s$. We use $\alpha \ast\to \beta$ to mean that $\alpha \to \beta$ or $\alpha \leftrightarrow \beta$. We say that edges $\alpha \to \beta$ and $\alpha \leftrightarrow \beta$ have a *head* at β, and that $\alpha \to \beta$ has a *tail* at α. If an edge e is between α and β and $\alpha = \beta$, we say that e is a *loop*.

We use V as a generic node set and let n denote the cardinality of V, $n = |V|$. The graphs described above are *directed mixed graphs* as formalized in the next definition.

Definition 2.3 (Directed mixed graph (DMG)). *We say that $\mathcal{G} = (V, E)$ is a* directed mixed graph *if its edge set, E, consists of directed and bidirected edges.*

We say that a DMG is a *directed graph* (DG) if it has no bidirected edges. A *walk* between γ_1 and γ_{l+1} is an alternating sequence of nodes, $\gamma_1, \ldots, \gamma_{l+1}$ and edges $\sim_1, \ldots, \sim_l$

$$\gamma_1 \sim_1 \gamma_2 \sim_2 \ldots \sim_l \gamma_{l+1}$$

such that for each $i = 1, \ldots, l$, $\sim_i$ is between γ_i and γ_{i+1}. Let e_i denote the edge $\sim_i$ above. We will sometimes write a walk as $(\gamma_1, e_1, \gamma_2, \ldots, e_l, \gamma_{l+1})$. A walk also specifies an *orientation* for each edge as one can otherwise not distingush between $\alpha \leftarrow \alpha$ and $\alpha \to \alpha$. We say that γ_i, $1 < i < l+1$, is a *collider* if $\sim_{i-1}$ and $\sim_i$ both have head at γ_i. Otherwise, we say that it is a *noncollider*. A node may be repeated on a walk, $\gamma_i = \gamma_j$, $i \neq j$, and may therefore occur both as a collider and as a noncollider on the same walk. Thus, the property of being a collider/noncollider pertains to the specific *instance* of the node on the walk. We say that γ_1 and γ_{l+1} are *endpoints* of the walk. Note that endpoints of a walk are neither colliders nor noncolliders. We say that a walk is *nontrivial* if it has at least one edge. A walk on which no node is repeated is a *path*.

Let $\mathcal{G} = (V, E)$. When e is an edge we use $\mathcal{G} + e$ to denote the graph $(V, E \cup \{e\})$, and we use $\mathcal{G} - e$ to denote the graph $(V, E \setminus \{e\})$. We say that $\mathcal{G}$ is *complete* if it contains $\alpha \to \beta$; $\alpha \leftarrow \beta$, and $\alpha \leftrightarrow \beta$ for all $\alpha, \beta \in V$, and we say that $\mathcal{G}$ is *empty* if $E = \emptyset$. We say that a walk between α and β is *directed* from α to β if every edge on the walk is directed and points towards (the last) β, $\alpha \to \ldots \to \beta$. We say that α is an *ancestor* of β in $\mathcal{G}$ if there exists a directed walk from α to β, and we allow this walk to be trivial (no edges) meaning that a node is always an ancestor of itself. We define $\mathrm{an}_{\mathcal{G}}(\alpha)$, or simply $\mathrm{an}(\alpha)$, to be the set of ancestors of α, and for $C \subseteq V$ we define $\mathrm{an}_{\mathcal{G}}(C) = \cup_{\alpha \in C} \mathrm{an}_{\mathcal{G}}(\alpha)$. Note that $C \subseteq \mathrm{an}_{\mathcal{G}}(C)$.

Definition 2.4 (μ-connecting walk). *We say that a nontrivial walk in a DMG, $\mathcal{G}$,*

$$\alpha \sim_1 \gamma_1 \sim_2 \ldots \sim_l \beta$$

is μ-connecting from α to β given C if $\alpha \notin C$, the edge $\sim_l$ has a head at β, every collider is in $\mathrm{an}(C)$ and no noncollider is in C.

The μ-connecting walks are used in the definition of μ-*separation* below which will help us connect DMGs to local independence. Mogensen et al. (2018) and Mogensen and Hansen (2020) defined μ-separation as an extension to δ-*separation* (Didelez, 2000, 2008). One can think of δ- and

μ-separation as analogous to d- and m-separation in DAG-based graphical models (Pearl, 2009; Richardson and Spirtes, 2002; Richardson, 2003).

Definition 2.5 (μ-separation)**.** *Let $\mathcal{G} = (V, E)$ and let $A, B, C \subseteq V$. We say that B is μ-separated from A given C in $\mathcal{G}$ if there is no μ-connecting walk from any $\alpha \in A$ to any $\beta \in B$ given C. We write this as $A \perp_\mu B \mid C \; [\mathcal{G}]$, or simply $A \perp_\mu B \mid C$. We say that C is a* conditioning set.

By definition, B is μ-separated from A given C if $A \subseteq C$. One should also note that μ-separation is not symmetric in A and B in that $A \perp_\mu B \mid C \; [\mathcal{G}]$ does not imply $A \perp_\mu B \mid C \; [\mathcal{G}]$, and neither is local independence. This lack of symmetry sets the graphical modeling of local independence apart from the classical graphical modeling of conditional independence (Lauritzen, 1996). In contrast to m-separation, μ-separation cannot be characterized using only paths (Mogensen and Hansen, 2020). It is, however, possible to obtain a characterization using only *routes* which are a finite subset of all possible walks (see Definition D.1 in Appendix D or Mogensen and Hansen (2020)). The next example illustrates the concept of μ-connecting walks and μ-separation in a DMG.

Example 2.6. *We consider the DMG, $\mathcal{G}$, in Figure 3. The walk $1 \leftrightarrow 2 \to 3$ is μ-connecting from 1 to 3 given $\emptyset$. It is not μ-connecting from 1 to 3 given $\{2\}$ as 2 is a noncollider. On the walk $1 \leftrightarrow 2 \leftarrow 2 \to 3$ the node 2 is a collider in its first instance and a noncollider in its second. The walk $3 \to 2 \leftrightarrow 1$ is μ-connecting from 3 to 1 given $\{2\}$, however, the reverse walk, $1 \leftrightarrow 2 \leftarrow 3$ is not μ-connecting from 1 to 3 given $\{2\}$.*

We see that 3 is μ-separated from 1 given $\{2,3\}$ in $\mathcal{G}$. On the other hand, 3 is not μ-separated from 1 given $\{2\}$ as the walk $1 \leftrightarrow 2 \leftarrow 3 \to 3$ is μ-connecting.
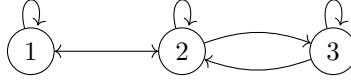

Figure 3: The graph $\mathcal{G}$ in Examples 2.6 and 2.8.

2.3 Independence models and Markov equivalence

For a fixed stochastic process, $X_t = (X_t^1, \ldots, X_t^n)^T$, and a DMG, $\mathcal{G} = (V, E)$, both local independence and μ-separation can be thought as *ternary relations* on a finite set $\mathbb{P}(V) \times \mathbb{P}(V) \times \mathbb{P}(V)$ where $V = \{1, 2, \ldots, n\}$ and $\mathbb{P}(\cdot)$ denotes power set. We use $\mathcal{P}$ to denote $\mathbb{P}(V) \times \mathbb{P}(V) \times \mathbb{P}(V) = \{(A, B, C) : A, B, C \subseteq V\}$ and we define an abstract independence model, $\mathcal{I}$, to be a subset of $\mathcal{P}$. Thus, $\mathcal{I}$ is a collection of triples (A, B, C) such that $A, B, C \subseteq V$. We say that $\mathcal{I}$ is an *independence model over V*. When A, B, or C are singletons, we will often omit the set notation and write, e.g., (α, β, C) instead of $(\{\alpha\}, \{\beta\}, C)$.

We use $\mathcal{I}(\mathcal{G})$ to denote the independence model induced by $\mathcal{G}$, that is, the set of μ-separations that are true in $\mathcal{G}$, $\mathcal{I}(\mathcal{G}) = \{(A, B, C) \in \mathcal{P} : A \perp_\mu B \mid C \; [\mathcal{G}]\}$. Similarly, an independence model can be defined as the set of local independences that hold in the distribution of a multivariate stochastic process. We say that an independence model, $\mathcal{I}$, is *graphical*, if there exist a DMG, $\mathcal{G}$, such that $\mathcal{I} = \mathcal{I}(\mathcal{G})$.

Definition 2.7 (Markov equivalence)**.** *Let $\mathcal{G}_1 = (V, E_1)$ and $\mathcal{G}_2 = (V, E_2)$ be DMGs. We say that $\mathcal{G}_1$ and $\mathcal{G}_2$ are* Markov equivalent *if for all $A, B, C \subseteq V$ it holds that B is μ-separated from A given*

C in $\mathcal{G}_1$ if and only if B is μ-separated from A given C in $\mathcal{G}_1$. Equivalently, $\mathcal{G}_1$ and $\mathcal{G}_2$ are Markov equivalent if $\mathcal{I}(\mathcal{G}_1) = \mathcal{I}(\mathcal{G}_2)$. We use $[\mathcal{G}_1]$ to denote the Markov equivalence class of $\mathcal{G}_1$.

Example 2.8. *We return to the graph, $\mathcal{G}$, in Figure 3. By definition, its independence model, $\mathcal{I}(\mathcal{G})$, consists of all triples (A, B, C) such that B is μ-separated from A given C in $\mathcal{G}$. It is enough to consider (A, B, C) such that A and B are singletons and $A \nsubseteq C$ as these characterize $\mathcal{I}(\mathcal{G})$ (Proposition 4.11). We see that 3 is μ-separated from 1 given $\{2, 3\}$, and this is the only μ-separation of this type in the graph.*

2.3.1 Extremal elements of sets of DMGs

Let $\mathbb{G} = \{\mathcal{G}_1 = (V, E_1), \ldots, \mathcal{G}_l = (V, E_l)\}$ be a set of DMGs on a common node set, V. If $E_i \subseteq E_j$, we write $\mathcal{G}_i \subseteq \mathcal{G}_j$, and we say that $\mathcal{G}_i$ is a *subgraph* of $\mathcal{G}_j$, and that $\mathcal{G}_j$ is a *supergraph* of $\mathcal{G}_i$. We write $\mathcal{G}_i \subsetneq \mathcal{G}_j$ when $E_i \subseteq E_j$ and $E_i \neq E_j$. The following definitions are common set-theoretic notions when considering the set $\mathbb{G}$ with the partial order, $\subseteq$.

Definition 2.9 (Maximal element, DMG). *We say that $\mathcal{G} \in \mathbb{G}$ is a* maximal element *of $\mathbb{G}$ if there is no $\bar{\mathcal{G}} \in \mathbb{G}$, $\bar{\mathcal{G}} \neq \mathcal{G}$, such that $\mathcal{G} \subseteq \bar{\mathcal{G}}$.*

Definition 2.10 (Greatest element, DMG). *We say that $\mathcal{G} \in \mathbb{G}$ is a* greatest element *of $\mathbb{G}$ if $\bar{\mathcal{G}} \subseteq \mathcal{G}$ for all $\bar{\mathcal{G}} \in \mathbb{G}$.*

When a greatest element exists, it is unique. It is also maximal, and it is the only maximal element. In this paper, we are mostly concerned with maximal and greatest elements, however, we also define *minimal* and *least* elements of sets of DMGs. We say that $\mathcal{G} \in \mathbb{G}$ is a *minimal element* of $\mathbb{G}$ if there is no $\bar{\mathcal{G}} \in \mathbb{G}$, $\bar{\mathcal{G}} \neq \mathcal{G}$, such that $\bar{\mathcal{G}} \subseteq \mathcal{G}$. We say that $\mathcal{G} \in \mathbb{G}$ is a *least element* of $\mathbb{G}$ if $\mathcal{G} \subseteq \bar{\mathcal{G}}$ for all $\bar{\mathcal{G}} \in \mathbb{G}$. The set $\mathbb{G}$ will most often be an equivalence class in our usage of the above terms, and we sometimes simply say that $\mathcal{G}$ is a maximal/minimal/greatest/least element when the equivalence class is understood from the context.

Example 2.11. *If we consider the set of graphs $\mathbb{G} = \{\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}\}$ in Figure 4, we see that graph $\mathbf{D}$ is the greatest element of $\mathbb{G}$ as every graph in $\mathbb{G}$ is a subgraph of $\mathbf{D}$, and therefore $\mathbf{D}$ is also the unique maximal element of $\mathbb{G}$. The smaller set $\bar{\mathbb{G}} = \{\mathbf{A}, \mathbf{B}, \mathbf{C}\}$ does not have a greatest element and graphs $\mathbf{B}$ and $\mathbf{C}$ are maximal elements of $\bar{\mathbb{G}}$.*

2.3.2 Representation of Markov equivalence classes

We introduce a central result from Mogensen and Hansen (2020). They show that every Markov equivalence class has a greatest element. Section 5 extends this theorem to weak equivalence relations.

Theorem 2.12 (Greatest element of a Markov equivalence class, (Mogensen and Hansen, 2020)). *Let $\mathcal{G}$ be a DMG, and let $[\mathcal{G}]$ be its Markov equivalence class. There exists $\mathcal{N} \in [\mathcal{G}]$ such that for all $\bar{\mathcal{G}} \in [\mathcal{G}]$ the edge set of $\bar{\mathcal{G}}$ is a subset of the edge set of $\mathcal{N}$.*

The next example illustrates the utility of this theorem.

Example 2.13. *Graphs $\mathbf{A}$-$\mathbf{D}$ in Figure 4 constitute a Markov equivalence class, $[\mathcal{G}]$ (for simplicity, we assume that all loops are present, and do not consider Markov equivalent graphs obtained by removing loops). Graph $\mathbf{D}$ is the greatest element of $[\mathcal{G}]$ in the sense that all Markov equivalent*

*graphs are subgraphs of graph **D**. In other words, if a graph in the Markov equivalence class contains the edge e, then e is also in the graph **D**. This means that we can represent the entire Markov equivalence class using graph **E**. The edges are the same as in the greatest element. Edges are solid in graph **E** if they are in every Markov equivalent graph and they are dashed if they are in some Markov equivalent graphs, but not in others. Absent edges are not in any graph in the Markov equivalence class. Therefore, graph **E** represents a summary of the information the Markov equivalence class provides on each edge. Moreover, Theorem 2.12 implies that every Markov equivalence class contains a greatest element, and therefore this is a general approach to representing and understanding Markov equivalence classes (Mogensen and Hansen, 2020).*
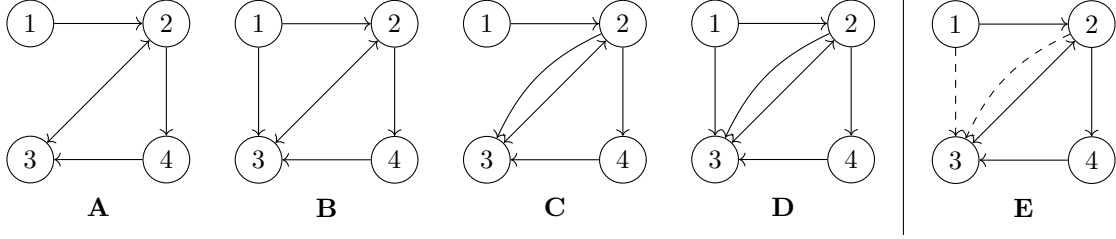

Figure 4: Graphs from Examples 2.11 and 2.13. All loops are present in the graphs but omitted from the visualization.

3   Hardness of marginalized local independence graphs

In this section, we argue that certain computational problems in relation to DMGs and Markov equivalence are hard. For this purpose, we give a very short introduction to the concepts from complexity theory that we will need. A decision problem is in **coNP** if no-instances have certificates which can be evaluated in polynomial time. For instance, if $\mathcal{G}_1$ and $\mathcal{G}_2$ are not Markov equivalent (they are a *no-instance* when deciding Markov equivalence) a triple (A, B, C) such that B is μ-separated from A given C in $\mathcal{G}_1$, but not in $\mathcal{G}_2$, may function as a certificate as one can check this specific separation in both graphs and conclude that they are not Markov equivalent. A decision problem is in **P** if it can be solved by a deterministic Turing machine in polynomial time. A decision problem is *coNP-hard* if it is at least as hard as any problem in **coNP**, and it is *coNP-complete* if it is coNP-hard and in **coNP**. It is generally believed that $\mathbf{P} \neq \mathbf{coNP}$ in which case there are no polynomial-time algorithm which can solve a coNP-hard problem. The *complement* of a decision problem arises from interchanging *yes* and *no*. A decision problem is in **coNP** if and only if its complement is in **NP**. We now introduce some decision problems relating to DMGs.

Decision problem 3.1 (Markov equivalence in DMGs)**.** *Let $\mathcal{G}_1 = (V, E)_1$ and $\mathcal{G}_2 = (V, E_2)$ be DMGs. Are $\mathcal{G}_1$ and $\mathcal{G}_2$ Markov equivalent?*

The development in this paper is partly motivated by the fact that the above decision problem is hard (Corollary 3.3). We can formulate a restricted version of the problem in which the pair of graphs for which to decide Markov equivalence only differ by a single (bidirected or directed) edge, as formalized in Decision problems A.1 (bidirected) and A.2 (directed). These problems are also hard and we prove this in Theorem 3.2. Corollary 3.3 follows immediately from this theorem.

Theorem 3.2. *Let $\mathcal{G}$ be a DMG and let e denote an edge. Deciding Markov equivalence of $\mathcal{G}$ and $\mathcal{G} + e$ is coNP-complete (Decision problems A.1 and A.2).*

Corollary 3.3. *Deciding Markov equivalence of DMGs is coNP-complete (Decision problem 3.1).*

Decision problem A.1 has been proven to be coNP-complete (PhD thesis, Mogensen (2020b)) and this was used to obtain the result in Corollary 3.3. We will give a slightly different proof to make the generalization to the proof in the sparse setting more transparent and to also prove that Decision problem A.2 is coNP-complete. The graphs $\mathcal{G}$, $\mathcal{G}_1$, and $\mathcal{G}_2$ used in the proof of Theorem 3.2 are clearly not sparse, that is, for the size of the node set going to infinity there are nodes with unbounded connectivity (formal definitions of node connectivity are in Subsection 3.1 and Section B). In the next section, we will show that the hardness results remain true under certain sparsity assumptions. We include the proof of the non-sparse result in Theorem 3.2 to illustrate the technique as the more general result can be proved using a similar approach, even if some additional ideas are needed.

Mogensen and Hansen (2022) showed that deciding μ-separation Markov equivalence of so-called *directed correlation graphs* (cDGs) is coNP-complete, though only in the non-sparse case. Their proof of coNP-hardness uses a reduction from 3DNF tautology as does the proof of Theorem 3.2. However, their proof is specific to cDGs as it uses a characterization of Markov equivalence which holds in cDGs, but not in DMGs (Mogensen and Hansen, 2022). While a DMG represents the local independences of a partially observed multivariate stochastic process, i.e., some coordinate processes are unobserved, a cDG represents a multivariate stochastic process driven by correlated noise. Mogensen and Hansen (2022) compared DMGs and cDGs further and showed that a Markov equivalence class of cDGs need not have a greatest element.

Proof. We consider n Boolean variables, $x_1, \ldots, x_n$, and a Boolean formula, H,

$$(z_1^1 \wedge z_2^1 \wedge z_3^1) \vee (z_1^2 \wedge z_2^2 \wedge z_3^2) \vee \ldots \vee (z_1^N \wedge z_2^N \wedge z_3^N)$$

such that z_i^k is a *literal* of a variable, that is, either x_l (a *positive* literal) or $\neg x_l$ (a *negative* literal). We assume H to be in 3DNF form (each conjunction has at most three literals). N is the number of conjunctions in the formula and n is the number of variables. We define n_j to be the number of factors in the j'th conjunction. Deciding whether H is a *tautology* (evaluates to true for all inputs) is known to be coNP-complete Garey and Johnson (1979) and we will use a reduction from this problem to show coNP-hardness of Decision problems A.1 and A.2.

We construct three graphs, $\mathcal{G} = (V, E)$, $\mathcal{G}_1 = (V, E_1)$, and $\mathcal{G}_2 = (V, E_2)$ from H such that $\mathcal{G}_1 = \mathcal{G} + e_b$ and $\mathcal{G}_2 = \mathcal{G} + e_d$ where e_b is a bidirected edge and e_d is a directed edge. We then show that $\mathcal{G}$ and $\mathcal{G}_1$ are Markov equivalent if and only if H is a tautology and that $\mathcal{G}_1$ and $\mathcal{G}_2$ are Markov equivalent if and only if H is a tautology.

First, we define the set V^-.

$$
\begin{aligned}
V^- = & \{\gamma, \bar{\gamma}, \delta, \bar{\delta}\} \\
& \cup \{\phi_i^k\}_{i=1,\ldots,n_k, k=1,\ldots,N} \\
& \cup \{\bar{\phi}_i^k\}_{i=1,\ldots,n_k, k=1,\ldots,N} \\
& \cup \{\chi_i, \lambda_i\}_{i=1,\ldots,n}.
\end{aligned}
$$

We define the node set $V = \{\alpha, \beta, \varepsilon, \phi\} \cup V^- \cup \{\nu_\beta^\rho, \nu_\varepsilon^\rho\}_{\rho \in V^-}$ and V is the node set of all three graphs $\mathcal{G} = (V, E)$, $\mathcal{G}_1 = (V, E_1)$, and $\mathcal{G}_2 = (V, E_2)$. Note that each literal, z_i^k, corresponds to two nodes, ϕ_i^k and $\bar{\phi}_i^k$.

We now define the edge set E. We add $\gamma \to \bar{\gamma}$; $\gamma \leftarrow \bar{\gamma}$; $\delta \to \bar{\delta}$; $\delta \leftarrow \bar{\delta}$. For each node $\rho \in V^-$, we add edges $\rho \to \nu_\varepsilon^\rho, \nu_\beta^\rho$ and $\rho \leftarrow \nu_\varepsilon^\rho, \nu_\beta^\rho$. We also add edges $\varepsilon \leftrightarrow \nu_\varepsilon^\rho$; $\beta \leftrightarrow \nu_\beta^\rho$. We add edges $\nu_\varepsilon^\rho \to \nu_\beta^\rho$ and $\nu_\varepsilon^\rho \to \nu_\beta^\rho$ for each $\rho \in V^-$. We also add all directed and bidirected loops, $\rho \sim \rho$, for all $\rho \in V$. We add edges $\alpha \leftrightarrow \gamma, \bar{\gamma}$; $\varepsilon \leftrightarrow \bar{\delta}$; $\beta \leftrightarrow \delta$, and $\varepsilon \to \beta$; $\varepsilon \leftarrow \beta$ as well as $\phi \leftrightarrow \varepsilon, \beta$. For each $k = 1, \ldots, N$, we add $\gamma \leftrightarrow \phi_1^k \leftrightarrow \ldots \leftrightarrow \phi_{n_k}^k \leftrightarrow \delta$ and $\bar{\gamma} \leftrightarrow \bar{\phi}_1^k \leftrightarrow \ldots \leftrightarrow \bar{\phi}_{n_k}^k \leftrightarrow \bar{\delta}$. We add $\bar{\gamma} \leftrightarrow \chi_1, \lambda_1$ and $\bar{\delta} \leftrightarrow \chi_n, \lambda_n$. For each $i = 1, \ldots, n-1$, we add $\chi_i, \lambda_i \leftrightarrow \chi_{i+1}, \lambda_{i+1}$. Finally, we add for each $l = 1, \ldots, n$ a directed cycle containing χ_l as well as every ϕ_i^k and $\bar{\phi}_i^k$ corresponding to a positive literal of the variable x_l, and we add a directed cycle containing λ_l as well as every ϕ_i^k and $\bar{\phi}_i^k$ corresponding to a negative literal of the variable x_l. This defines the edge set E, $\mathcal{G} = (V, E)$. We obtain $\mathcal{G}_1 = (V, E_1)$ from $\mathcal{G}$ by adding the edge $\varepsilon \leftrightarrow \beta$, that is, $E_1 = E \cup \{\varepsilon \leftrightarrow \beta\}$. Note that ρ_1 is an ancestor of ρ_2 in $\mathcal{G}$ if and only if ρ_1 is an ancestor of ρ_2 in $\mathcal{G}_1$. We obtain $\mathcal{G}_2 = (V, E_2)$ from $\mathcal{G}$ by adding the edge $\phi \to \varepsilon$, $E_2 = E \cup \{\phi \to \varepsilon\}$.

We will first argue that $\mathcal{G}$ and $\mathcal{G}_1$ are Markov equivalent if and only if H is a tautology. Assume first that H is a tautology and consider a μ-connecting walk in $\mathcal{G}_1$,

$$\rho_1 \sim \ldots \varepsilon \leftrightarrow \beta \ldots \sim \rho_m$$

Using the fact that all loops are included, we can always find a μ-connecting walk such that the edge $\varepsilon \leftrightarrow \beta$ occurs at most once and we assume that this is the case. We can assume that ρ_1 only occurs once on the walk. If $\rho_1 \neq \alpha$, there is a μ-connecting walk from ρ_1 to β with a head at β: If $\rho_1 \in V^-$, or $\rho_1 = \nu_\varepsilon^\rho$ for some $\rho \in V^-$, either $\rho_1 \to \nu_\beta^\rho \leftrightarrow \beta$ or $\rho_1 \leftarrow \nu_\beta^\rho \leftrightarrow \beta$ is connecting and can be composed with the subwalk from β to ρ_m to obtain a connecting walk in $\mathcal{G}$. If $\rho_1 = \varepsilon, \beta, \phi$ or $\rho_1 = \nu_\beta^\rho$ for some $\rho \in V^-$, then $\rho_1 \ast\!\!\to \beta$ is in $\mathcal{G}$. Assume instead that $\rho_1 = \alpha$,

$$\alpha \sim \ldots \varepsilon \leftrightarrow \beta \ldots \sim \rho_m$$

and consider the subwalk from α to ε, ω_1. If there is a noncollider on ω_1, say ψ, then $\psi \notin C$ and $\psi \in \mathrm{an}(C)$. We use this to argue that we can always find a walk from ψ to β such that when concatenated with the subwalk from α to ψ we obtain a μ-connecting walk from α to β. If $\psi \in V^-$, we can find a connecting walk from α to β with a head at β by concatenating the subwalk from α to ψ with $\psi \to \nu_\beta^\psi \leftrightarrow \beta$ if $\nu_\beta^\psi \in C$ and $\psi \leftarrow \nu_\beta^\psi \leftrightarrow \beta$ if $\nu_\beta^\psi \notin C$. If $\psi = \nu_\varepsilon^\rho$ for some ρ, we can concatenate with $\psi \to \nu_\beta^\rho \leftrightarrow \beta$ or $\psi \leftarrow \nu_\beta^\rho \leftrightarrow \beta$. If $\psi = \nu_\beta^\rho$ for some ρ, we can concatenate with $\psi \leftrightarrow \beta$. If $\psi = \varepsilon$, then we can replace $\varepsilon \leftrightarrow \beta$ with $\varepsilon \to \beta$ to obtain a connecting walk in $\mathcal{G}$. If $\psi = \beta$, we can concatenate with $\psi \to \beta$. If $\psi = \phi$, we can concatenate with $\psi \leftrightarrow \beta$. Finally, $\psi = \alpha$ is not possible as $\rho_1 = \alpha$ only occurs once on the original walk.

Assume now that ω_1 is a collider walk. If it goes through a $\bar{\phi}$-segment, then the corresponding ϕ-segment is open (note that γ and $\bar{\gamma}$ are in a directed cycle and so are δ and $\bar{\delta}$). If it goes through the χ-λ-segment, then for each $l = 1, \ldots, n$ either $\chi_l \in \mathrm{an}(C)$ or $\lambda_l \in \mathrm{an}(C)$. Let $x_l = 1$ if $\chi_l \in \mathrm{an}(C)$ and $x_l = 0$ otherwise. The formula H is a tautology and therefore it evaluates to 1 under this assignment of truth values. Thus, there exists k such that $z_i^k = 1$ for $i = 1, \ldots, n_k$. Assume first that z_i^k is a positive literal corresponding to the variable x_l. In this case, $x_l = 1$ and $\chi_l \in \mathrm{an}(C)$, and therefore $\phi_i^k \in \mathrm{an}(C)$. Assume instead that z_i^k is a negative literal corresponding to the variable x_l. In this case, $x_l = 0$ and $\chi_l \notin \mathrm{an}(C)$ which means that $\lambda_l \in \mathrm{an}(C)$ and $\phi_i^k \in \mathrm{an}(C)$.

This means that the walk $\alpha \leftrightarrow \gamma \leftrightarrow \phi_1^k \leftrightarrow \ldots \leftrightarrow \phi_{n_k}^k \leftrightarrow \delta \leftrightarrow \beta$ is open for some $k = 1, \ldots, N$ and this gives us a μ-connecting walk from α to ρ_m in $\mathcal{G}$ also in this case.

If instead

$$\rho_1 \sim \ldots \beta \leftrightarrow \varepsilon \ldots \sim \rho_m$$

then the same arguments hold.

On the other hand, say that H is not a tautology, and consider an assignment, A, of truth values such that H evaluates to false. Define the set

$$C = \mathrm{an}\Big(\{\chi_i : x_i = 1 \text{ in } A\} \cup \{\lambda_i : x_i = 0 \text{ in } A\} \cup \{\gamma, \delta, \varepsilon, \beta\}\Big).$$

In $\mathcal{G}_1$, there is an open, bidirected walk from α to β through the χ-λ segment, and we see that β is not μ-separated from α given C. On the other hand, consider a walk between α and β in $\mathcal{G}$. The first and last edges on a connecting walk from α to β given C must be bidirected and as $C = \mathrm{an}(C)$, this means that the walk must be a collider walk to be μ-connecting from α to β given C, and it must go through δ. If ϕ_i^k corresponds to a positive literal and it is open (i.e., in $\mathrm{an}(C)$) then the correspond variable is 1 in A and $z_i^k = 1$. If it corresponds to a negative literal and it is open, then the corresponding variable is 0 in A and $z_i^k = 1$. This means that each ϕ_i^k segment must be closed in at least one node as the assignment A evaluates to 0. Therefore, β is μ-separated from α given C in $\mathcal{G}$, and we conclude that $\mathcal{G}$ and $\mathcal{G}_1$ are Markov equivalent if and only if H is a tautology.

We now show that $\mathcal{G}_1$ and $\mathcal{G}_2$ are Markov equivalent. Take any μ-connecting walk in $\mathcal{G}_1$. Any occurrence of $\varepsilon \leftrightarrow \beta$ can be replaced by either $\beta \leftrightarrow \phi \rightarrow \varepsilon$ or $\beta \leftrightarrow \phi \leftrightarrow \varepsilon$, depending on whether $\phi \in C$. The resulting walk is present and connecting in $\mathcal{G}_2$. On the other hand, consider a μ-connecting walk from ρ_1 to ρ_m given C in $\mathcal{G}_2$. We start by removing all non-endpoint occurrences of ϕ. Say

$$\rho_1 \sim \ldots \sim \rho_i \sim \phi \rightarrow \varepsilon \sim \ldots \sim \rho_m.$$

If $\rho_i = \beta$, then $\rho_i \leftrightarrow \phi \rightarrow \varepsilon$ can be replaced by $\rho_i \leftrightarrow \varepsilon$. If $\rho_i = \phi$ or if $\rho_i = \varepsilon$, we can remove the cycle ($\varepsilon = \rho_m$ we may need to concatenate with $\varepsilon \rightarrow \varepsilon$ to obtain a μ-connecting walk after removing a cycle). If instead

$$\rho_1 \sim \ldots \sim \rho_i \sim \varepsilon \leftarrow \phi \sim \rho_j \sim \ldots \sim \rho_m$$

we do the same depending on ρ_j (if $\phi = \rho_m$ then we concatenate the subwalk from ρ_1 to ε with $\varepsilon \leftrightarrow \phi$). This gives us a μ-connecting walk in $\mathcal{G}_2$ such that ϕ is not a non-endpoint node. Finally, if $\phi \rightarrow \varepsilon$ is still on the walk ϕ, we must have $\rho_1 = \psi$ and this edge can be substituted by $\phi \leftrightarrow \varepsilon$. The resulting walk is present in $\mathcal{G}_1$. Every collider is different from ϕ and this means that it is in $\mathrm{an}_{\mathcal{G}_1}(C)$ as well. Therefore, this walk is μ-connecting in $\mathcal{G}_1$. It follows that $\mathcal{G}_1$ and $\mathcal{G}_2$ are Markov equivalent (regardless of whether H is a tautology). Therefore, H is a tautology if and only if $\mathcal{G}$ are $\mathcal{G}_2$ Markov equivalent.

The reduction from 3DNF tautology to Markov equivalence of $\mathcal{G}$ and $\mathcal{G}_1$ (or of $\mathcal{G}$ and $\mathcal{G}_2$) is done in polynomial time in the number of conjunctions and it follows that Decision problems A.1 and A.2 are coNP-hard. Given a triple (A, B, C), one can decide μ-separation in polynomial time. If two graphs are not Markov equivalent, then there exists a triple (A, B, C) such that μ-separation holds in one and not in the other. This is a polynomially-sized certificate, and this means that these problems are in **coNP**, thus, coNP-complete. $\qquad\square$
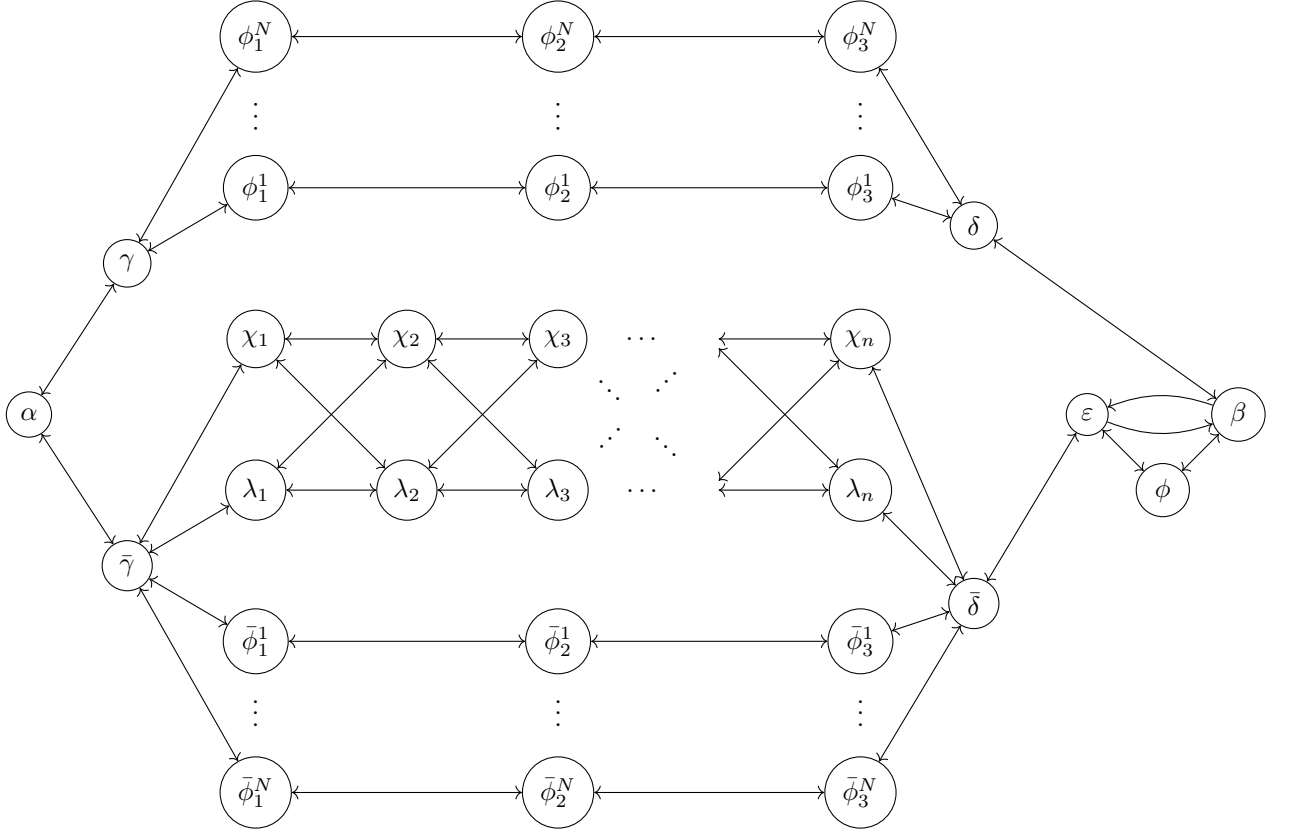
Figure 5: A subgraph of $\mathcal{G}$ in the proof of Theorem 3.2.

Theorem 3.2 shows that deciding Markov equivalence is not computationally feasible for large graphs which hurts the practical applicability of μ-separation DMGs. We discuss the implications further in Subsection 3.2. We now consider the analogous decision problems in a *sparse* setting.

3.1 Sparse DMGs

We may ask if the hardness results still apply if we fix the maximal connectivity of each node and let the size of the node set grow. As a formalization of this, we first define a notion of node connectivity based on *inseparability*. We say that β is *inseparable* from α in $\mathcal{I}(\mathcal{G})$ if there is no $C \subseteq V \setminus \{\alpha\}$ such that β is μ-separated from α given C in $\mathcal{G}$ (Mogensen et al., 2018). We let $\vec{u}(\beta, \mathcal{I}(\mathcal{G}))$ denote the set of nodes α such that β is inseparable from α in $\mathcal{G}$, and we let $\overleftarrow{u}(\beta, \mathcal{I}(\mathcal{G}))$ denote the set of nodes α such that α is inseparable from β.

Definition 3.4 (Node connectivity in DMG). *We define* $\mathrm{con}_{\mathcal{G}}^{\rightarrow}(\beta)$ *as the cardinality of the set* $\vec{u}(\beta, \mathcal{I}(\mathcal{G}))$ *and we define* $\mathrm{con}_{\mathcal{G}}^{\leftarrow}(\beta)$ *as the cardinality of the set* $\overleftarrow{u}(\beta, \mathcal{I}(\mathcal{G}))$. *We define* $\mathrm{con}_{\mathcal{G}}(\beta)$ *as the maximum of* $\mathrm{con}_{\mathcal{G}}^{\rightarrow}(\beta)$ *and* $\mathrm{con}_{\mathcal{G}}^{\leftarrow}(\beta)$.

We see that the above definitions are invariant under Markov equivalence, i.e., $\mathrm{con}_{\mathcal{G}_1}(\beta) =$
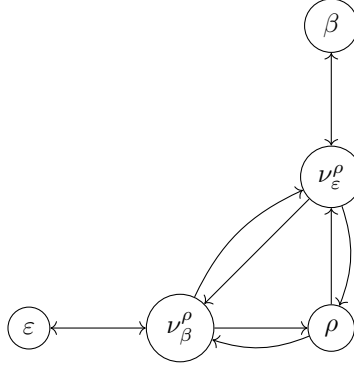
13

Figure 6: A subgraph of $\mathcal{G}$ in the proof of Theorem 3.2.

$\mathrm{con}_{\mathcal{G}_2}(\beta)$, $\mathrm{con}_{\mathcal{G}_1}^{\rightarrow}(\beta) = \mathrm{con}_{\mathcal{G}_2}^{\rightarrow}(\beta)$, and $\mathrm{con}_{\mathcal{G}_1}^{\leftarrow}(\beta) = \mathrm{con}_{\mathcal{G}_2}^{\leftarrow}(\beta)$ when $\mathcal{G}_1$ and $\mathcal{G}_2$ are Markov equivalent. One can define other notions of node connectivity in a DMG, in particular based on the edges directly, instead of using separability. However, a DMG in which every node is adjacent with only a small number of nodes may be Markov equivalent with the complete DMG (see Figure 7). Even in a maximal DMG, the lack of an edge between a pair of nodes does not generally imply separability (Appendix B), and therefore connectivity based on separability appears to be a more useful notion of connectivity. Moreover, the graphs are intended as representations of stochastic systems, thus functional sparsity (i.e., sparsity in the implied dependence structure) seems more useful than representational sparsity (sparsity in node adjacency). Appendix B provides more details and examples.

Figure 7: Loops are omitted from the visualization. This graph is Markov equivalent with the complete DMG on nodes $\{1, 2, \ldots, n\}$.

Definition 3.5 (m-sparsity). *Let $\mathcal{G}$ be a DMG. The* maximal connectivity *of $\mathcal{G}$ is defined as* $\max_{\alpha \in V}(\mathrm{con}_{\mathcal{G}}(\alpha))$. *We say that $\mathcal{G} = (V, E)$ is m-sparse if $\max_{\alpha \in V}(\mathrm{con}_{\mathcal{G}}(\alpha)) \leq m$.*

We now state a sparse version of Decision problem 3.1.

Decision problem 3.6 (Markov equivalence in m-sparse DMGs). *Let m be a nonnegative integer and let $\mathcal{G}_1 = (V, E_1)$ and $\mathcal{G}_2 = (V, E_2)$ be m-sparse DMGs. Are $\mathcal{G}_1$ and $\mathcal{G}_2$ Markov equivalent?*

The following are sparse versions of Theorem 3.2 and Corollary 3.3.

Theorem 3.7. *Let $m \geq 16$, let $\mathcal{G} = (V, E)$ be an m-sparse graph, and let e denote an edge. Deciding Markov equivalence of $\mathcal{G}$ and $\mathcal{G} + e$ is coNP-complete (Decision problems A.3 and A.4).*

14

Theorem 3.7 is a stronger version of Theorem 3.2 as it shows that the problem of deciding Markov equivalence of DMGs remains coNP-complete when restricting to sparse DMGs. We discuss the implications in Subsection 3.2.

Corollary 3.8. *Let $m \geq 16$. Deciding Markov equivalence of m-sparse DMGs is coNP-complete.*

The value $m = 16$ may not be what we expect from 'sparse' graphical models and two comments are in order. First, the adjacency sparsity (see Section B) of the graphs in the proof are only 8, also in the maximal Markov equivalent graphs of the graphs used in the proof. Second, the upshot of the corollary is that there exists a finite number such that deciding Markov equivalence of m-sparse DMGs is coNP-complete. This means that fixing the value of m does not generally lead to computational problems that scale as polynomials in the size of the graph. On the other hand, the so-called k-weak equivalences that are introduced in this paper provide polynomial-time algorithms for each fixed k (Section 7). Note that results analogous to those of Theorems 3.2 and 3.7 do not hold for ADMGs with m-separation. For those, polynomial-time algorithms for Markov equivalence are known, without making sparsity assumptions (Hu and Evans, 2020).

Proof. We consider a Boolean formula in 3DNF form as in the proof of Theorem 3.2 (see that proof for related notation and terminology). We will define three m-sparse graphs $\mathcal{G} = (V, E)$, $\mathcal{G}_1 = (V, E_1)$, and $\mathcal{G}_2 = (V, E_2)$ and show that $\mathcal{G}$ and $\mathcal{G}_1$ are Markov equivalent if and only if H is a tautology while $\mathcal{G}_1$ and $\mathcal{G}_2$ are always Markov equivalent.

We define M to be the smallest integer such that $2^{M-1} \geq N + 1$. We first define a number of sets that will be subsets of the node set V. Note that these sets are all pairwise disjoint.

$$\Gamma = \{\gamma_{ij}, i = 1, \ldots, M, j = 1 \ldots, 2^{i-1}\}$$
$$\bar{\Gamma} = \{\bar{\gamma}_{ij}, i = 1, \ldots, M, j = 1 \ldots, 2^{i-1}\}$$
$$\Delta = \{\delta_{ij}, i = 1, \ldots, M, j = 1 \ldots, 2^{i-1}\}$$
$$\bar{\Delta} = \{\bar{\delta}_{ij}, i = 1, \ldots, M, j = 1 \ldots, 2^{i-1}\}$$
$$\Phi = \{\phi_i^j, j = 1, \ldots, N, i = 1, \ldots, n_j\}$$
$$\bar{\Phi} = \{\bar{\phi}_i^j, j = 1, \ldots, N, i = 1, \ldots, n_j\}$$
$$\mathrm{X} = \{\chi_l, l = 1, \ldots, n\}$$
$$\Lambda = \{\lambda_l, l = 1, \ldots, n\}$$

The node χ_l corresponds to the Boolean variable x_l and the node λ_l corresponds to the negation of x_l. Nodes ϕ_i^j and $\bar{\phi}_i^j$ both correspond to the literal z_i^j (see also the proof of Theorem 3.2 for additional explanation). We define

$$V^- = \Gamma \cup \Delta \cup \Phi \cup \mathrm{X} \cup \Lambda$$
$$\bar{V}^- = \bar{\Gamma} \cup \bar{\Delta} \cup \bar{\Phi}$$
$$\mathrm{N}_\varepsilon = \{\nu_\varepsilon^\phi, \}_{\phi \in V^-}$$
$$\mathrm{N}_\beta = \{\nu_\beta^\phi\}_{\phi \in V^-}$$
$$\bar{\mathrm{N}}_\varepsilon = \{\bar{\nu}_\varepsilon^\phi, \}_{\phi \in \bar{V}^-}$$
$$\bar{\mathrm{N}}_\beta = \{\bar{\nu}_\beta^\phi\}_{\phi \in \bar{V}^-}.$$
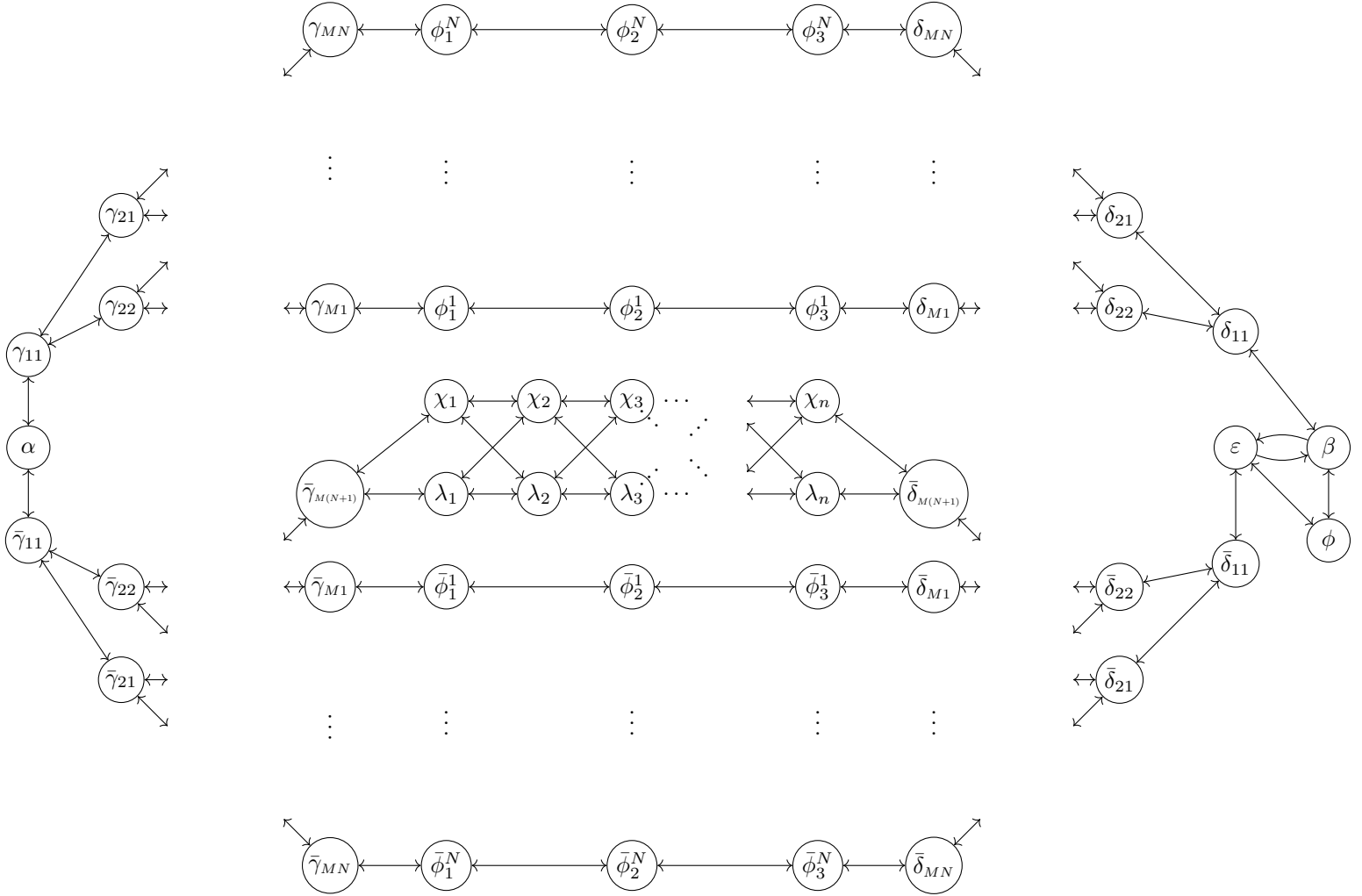
Figure 8: A subgraph of $\mathcal{G}$ in the proof of Theorem 3.7.

We now define the node set V as a disjoint union,

$$V = \{\alpha, \beta, \varepsilon, \phi\} \cup V^- \cup \bar{V}^- \cup N_\varepsilon \cup N_\beta \cup \bar{N}_\varepsilon \cup \bar{N}_\beta.$$

We add some intuition on the construction of the graph. The Γ- and Δ-nodes (and their barred versions) are 'triangular' in shape and help connect a single node to many more in a sparse manner (see Figure 8). The Φ- and $\bar{\Phi}$-nodes correspond to literals in the conjunctions of the Boolean formula, H. The elements of X correspond to variables in H, and the elements of Λ to their negation. The ν_ε- and ν_β-components will help connect every node to ε and to β and are copies of the V^- and $\bar{V}^-$ sets in the sense that $\rho \mapsto \nu_\varepsilon^\rho$ is a bijection from V^- to N_ε, $\rho \mapsto \nu_\beta^\rho$ is a bijection from V^- to N_β, $\rho \mapsto \bar{\nu}_\varepsilon^\rho$ is a bijection from $\bar{V}^-$ to $\bar{N}_\varepsilon$, and $\rho \mapsto \bar{\nu}_\beta^\rho$ is a bijection from $\bar{V}^-$ to $\bar{N}_\beta$, though the edges are not exact copies as explained below.

We now define the edge set of $\mathcal{G}$. We add bidirected edges $\gamma_{ij} \leftrightarrow \gamma_{(i+1)(2j)}, \gamma_{(i+1)(2j-1)}$ for $i = 1, \ldots, M-1$, and analogously for $\bar{\Gamma}$, Δ, and $\bar{\Delta}$ (see Figure 8). Moreover, we add $\gamma_{Mj} \leftrightarrow \phi_1^j$; $\bar{\gamma}_{Mj} \leftrightarrow \bar{\phi}_1^j$; $\delta_{Mj} \leftrightarrow \delta_{n_j}^j$; $\bar{\delta}_{Mj} \leftrightarrow \bar{\delta}_{n_j}^j$ for $j \leq N$. We also add $\bar{\gamma}_{M2^{M-1}} \leftrightarrow \chi_1, \lambda_1$; $\bar{\delta}_{M2^{M-1}} \leftrightarrow \chi_n, \lambda_n$. We add $\alpha \leftrightarrow \gamma_{11}, \bar{\gamma}_{11}$. We also add $\varepsilon \leftrightarrow \bar{\delta}_{11}$ and $\beta \leftrightarrow \delta_{11}$. We add $\varepsilon \to \beta$ and $\beta \to \varepsilon$ as well as $\phi \leftrightarrow \varepsilon, \beta$. We add for each $j = 1, \ldots, N$, $\phi_i^j \leftrightarrow \phi_{i+1}^j$ and $\bar{\phi}_i^j \leftrightarrow \bar{\phi}_{i+1}^j$ for $1 \leq i \leq n_j - 1$.

For $\phi_1, \phi_2 \in V^-$ such that $\phi_1 \notin \Phi$ or $\phi_2 \notin \Phi$, we add $\nu_\varepsilon^{\phi_1} \leftrightarrow \nu_\varepsilon^{\phi_2}$ and $\nu_\beta^{\phi_1} \leftrightarrow \nu_\beta^{\phi_2}$ if and only if $\phi_1 \leftrightarrow \phi_2$ was added above. For each j, we also add $\nu_\beta^{\gamma_{Mj}} \leftrightarrow \nu_\beta^{\phi_i^j} \leftrightarrow \nu_\beta^{\delta_{Mj}}$ and $\bar{\nu}_\varepsilon^{\gamma_{Mj}} \leftrightarrow \bar{\nu}_\varepsilon^{\phi_i^j} \leftrightarrow \bar{\nu}_\varepsilon^{\delta_{Mj}}$ for each $i = 1, \ldots, n_j$. We also add $\nu_\varepsilon^{\delta_{11}} \leftrightarrow \varepsilon$; $\nu_\beta^{\delta_{11}} \leftrightarrow \beta$; $\bar{\nu}_\varepsilon^{\bar{\delta}_{11}} \leftrightarrow \varepsilon$ and $\bar{\nu}_\beta^{\bar{\delta}_{11}} \leftrightarrow \beta$. Note that $\nu_\varepsilon^{\gamma_{11}}, \nu_\beta^{\gamma_{11}}, \bar{\nu}_\varepsilon^{\bar{\gamma}_{11}}, \bar{\nu}_\beta^{\bar{\gamma}_{11}}$ are not adjacent with α. For $\phi_1, \phi_2 \in \bar{V}^-$ such that $\phi_1 \notin \bar{\Phi}$ or $\phi_2 \notin \bar{\Phi}$, we add $\bar{\nu}_\varepsilon^{\phi_1} \leftrightarrow \bar{\nu}_\varepsilon^{\phi_2}$ and $\bar{\nu}_\beta^{\phi_1} \leftrightarrow \bar{\nu}_\beta^{\phi_2}$ if and only if $\phi_1 \leftrightarrow \phi_2$ was added above. For each j, we also add $\bar{\nu}_\beta^{\bar{\gamma}_{Mj}} \leftrightarrow \bar{\nu}_\beta^{\bar{\phi}_i^j} \leftrightarrow \bar{\nu}_\beta^{\bar{\delta}_{Mj}}$ and $\bar{\nu}_\varepsilon^{\bar{\gamma}_{Mj}} \leftrightarrow \bar{\nu}_\varepsilon^{\bar{\phi}_i^j} \leftrightarrow \bar{\nu}_\varepsilon^{\bar{\delta}_{Mj}}$ for each $i = 1, \ldots, n_j$.

In this proof, we will say that sets $V^-, \bar{V}^-, N_\varepsilon, N_\beta, \bar{N}_\varepsilon$, and $\bar{N}_\beta$ are *line segments*. We define

$$
\begin{aligned}
V^i &= \{\gamma_{ij}, \nu_\varepsilon^{\gamma_{ij}}, \nu_\beta^{\gamma_{ij}}, \bar{\gamma}_{ij}, \bar{\nu}_\varepsilon^{\bar{\gamma}_{ij}}, \bar{\nu}_\beta^{\bar{\gamma}_{ij}}, j = 1, \ldots, 2^{i-1}\}, && i = -M, \ldots, -1, \\
V^i &= \{\delta_{ij}, \nu_\varepsilon^{\delta_{ij}}, \nu_\beta^{\delta_{ij}}, \bar{\delta}_{ij}, \bar{\nu}_\varepsilon^{\bar{\delta}_{ij}}, \bar{\nu}_\beta^{\bar{\delta}_{ij}}, j = 1, \ldots, 2^{i-1}\}, && i = 1, \ldots, M, \\
V^0 &= \{\phi_i^j, \nu_\varepsilon^{\phi_i^j}, \nu_\beta^{\phi_i^j}, \bar{\phi}_i^j, \bar{\nu}_\varepsilon^{\bar{\phi}_i^j}, \bar{\nu}_\beta^{\bar{\phi}_i^j}, j = 1, \ldots, N, i = 1, \ldots, n_j\} \cup \\
&\quad \{\chi_i, \lambda_i, \nu_\varepsilon^{\chi_i}, \nu_\varepsilon^{\lambda_i}, \nu_\beta^{\chi_i}, \nu_\beta^{\lambda_i}\}, \\
V^{-(M+1)} &= \{\alpha\} \\
V^{M+1} &= \{\beta, \varepsilon, \phi\}
\end{aligned}
$$

and we say that V^i is a *vertical segment* for $i = -(M+1), M, \ldots, -1, 0, 1, \ldots, M, M+1$. 'Vertical' refers to the specific visualization of $\mathcal{G}$ used in Figure 8. The sets, V_j^i, defined above are disjoint and $\bigcup_{i=-(M+1)}^{M+1} V^i = V$.

We now add a number of directed edges. For every node $\phi \in V^-$, we add $\phi, \nu_\varepsilon^\phi, \nu_\beta^\phi \to \phi, \nu_\varepsilon^\phi, \nu_\beta^\phi$. For every node $\phi \in \bar{V}^-$, we add $\phi, \bar{\nu}_\varepsilon^\phi, \bar{\nu}_\beta^\phi \to \phi, \bar{\nu}_\varepsilon^\phi, \bar{\nu}_\beta^\phi$. For each $i = \pm 1, \ldots, \pm M$, we connect the nodes in the vertical segment V^i by a directed cycle (any will work). We add directed cycles

containing χ_k and all ϕ_i^j and $\bar{\phi}_i^j$ such that z_i^j is a positive literal of the variable x_k. We add directed cycles containing λ_k and all ϕ_i^j and $\bar{\phi}_i^j$ such that z_i^j is a negative literal of the variable x_k.

Finally, we add all directed and bidirected loops. The above defines the edge set E and we let $\mathcal{G} = (V, E)$. Note that the nodes in a vertical segment are connected by a directed cyclic walk for $i \neq -(M-1), 0, M+1$. We also define $\mathcal{G}_1 = (V, E_1)$ where $E_1 = E \cup \{\beta \leftrightarrow \varepsilon\}$ and $\mathcal{G}_2 = (V, E_2)$ where $E_2 = E \cup \{\phi \rightarrow \varepsilon\}$. Note that in all three graphs, if $\rho_1 \sim_e \rho_2$ and ρ_1 and ρ_2 are in different vertical segments, V^{i_1} and V^{i_2}, respectively, then e is bidirected and $i_1 - i_2 = \pm 1$.

We will first show that $\mathcal{G}$ and $\mathcal{G}_1$ are Markov equivalent if and only if H is a tautology. Assume first that H is a tautology and consider a μ-connecting walk from ρ_1 to ρ_m in $\mathcal{G}_1$,

$$\rho_1 \sim \ldots \sim \rho_m.$$

Every node has a self-loop, so it suffices to consider walks where e_1 (the edge $\varepsilon \leftrightarrow \beta$) only occurs once. If it does not occur at all the walk is present in $\mathcal{G}$ as well and connecting (ancestry is the same in $\mathcal{G}$ and $\mathcal{G}_1$). Say

$$\underbrace{\rho_1 \sim \ldots \sim \varepsilon}_{\omega_1} \leftrightarrow \underbrace{\beta \sim \ldots \sim \rho_m}_{\omega_2}.$$

If $\rho \in V^i$, then we say that i is the *order* of ρ.

Lemma 3.9. *Let $\rho \in V$ be of order j. If there is an open walk from ρ to β given C in $\mathcal{G}$ or in $\mathcal{G}_1$ then the k'th vertical segment , $j < k < M+1$, contains at least one node in C.*

Proof. If $j = M, M+1$ this is vacuously true as no vertical segment satisfies the condition, and we can assume that $\rho \neq \varepsilon, \beta, \phi$. Note that this walk must necessarily pass through a collider in each vertical segment V^k such that $k > j$ which gives the result. To see this, note that removing any vertical segment such that $k > j$ gives us a disconnected graph with ρ in one component and β in the other as a vertical segment, k, is only adjacent to vertical segments $k-1$ and $k+1$. When a walk contains a subwalk $\rho_1 \sim \rho_2$ such that ρ_1 is in V^{k-1} and ρ_2 is in V^k, then the connecting edge must be bidirected. If ρ_2 is a collider, we must have $\rho_2 \in \mathrm{an}_{\mathcal{G}}(C)$ and ρ_2 is only an ancestor of nodes in V^k. Otherwise, ρ_2 is an ancestor of a collider in V^k and the same argument applies. $\square$

Lemma 3.10. *Let $\rho \neq \alpha$ be a node in $\mathcal{G}$. If there exists an open walk from ρ to β in $\mathcal{G}_1$ with a head at β, then there exists an open walk $\rho \sim \nu_\beta^\rho \sim \ldots \sim \beta$ in $\mathcal{G}$ with a head at β such that every nonendpoint node equals ν_β^ρ for $\rho \in V^-$ or $\bar{\nu}_\beta^\rho$ for $\rho \in \bar{V}^-$.*

Proof. If $\rho = \beta, \varepsilon, \phi$, this is immediate. Assume instead that $\rho \in V^- \cup \bar{V}^-$. Choose first the edge $\rho \leftarrow \nu_\beta^\rho$ if $\nu_\beta^\rho \in C$, and otherwise $\rho \rightarrow \nu_\beta^\rho$. We concatenate this with the open bidirected path to β. Such a path exists as $\nu_\beta^{\gamma_{Mj}} \leftrightarrow \nu_\beta^{\delta_{Mj}}$ and $\bar{\nu}_\beta^{\bar{\gamma}_{Mj}} \leftrightarrow \bar{\nu}_\beta^{\bar{\delta}_{Mj}}$. This is open since all vertical segments between ρ and β must contain at least one node which is in C by Lemma 3.9.

If instead $\rho \in \mathrm{N}_\varepsilon \cup \bar{\mathrm{N}}_\varepsilon$ we can do as above as $\rho \leftarrow \nu_\beta^\rho$ and $\rho \rightarrow \nu_\beta^\rho$ are in the graph. If $\rho \in \mathrm{N}_\beta \cup \bar{\mathrm{N}}_\beta$, then there is an open bidirected path with a head at β between ρ and β. If $\rho = \varepsilon$ or $\rho = \beta$ it follows directly. $\square$

We split into cases depending on whether $\rho_1 = \alpha$.

$\rho_1 \neq \alpha$: There is an open walk (given C) from ρ_1 with a head at β (Lemma 3.10) that we can concatenate with ω_2 to obtain a connecting walk in $\mathcal{G}$.

If instead

$$\rho_1 \sim \ldots \sim \beta \leftrightarrow \varepsilon \sim \ldots \sim \rho_m$$

the same argument holds.

$\rho_1 = \alpha$: If we have a subwalk between α and β with a noncollider, then we can find a connecting path in the following way. Say we have

$$\rho_1 \sim \ldots \sim \psi_0 \sim \psi_1 \sim \psi_2 \sim \ldots \varepsilon \leftrightarrow \beta \sim \ldots \sim \rho_m$$

such that ψ_1 is a noncollider (note that, ignoring $\alpha \to \alpha$, α only has bidirected edges at it, so $\psi_1 \neq \alpha$ if we remove α-loops). There is necessarily a tail at ψ_1 on one of the adjacent edges, $\psi_1 \notin C$, and $\psi_1 \in \mathrm{an}(C)$. We concatenate the subwalk from α to ψ_1 with the open walk from ψ_1 to β that has a head at β. Lemma 3.10 gives the existence of this walk. This also holds if $\rho_1 = \psi_0$, $\psi_2 = \varepsilon$, or $\psi_1 = \varepsilon$.

On the other hand, if the subwalk between α and β has no noncolliders, then either it stays within a line segment or either α, β, or ε occur on the subwalk as a nonendpoint. We can assume that α is only an endpoint. If β occurs as a nonendpoint, then this β is a collider and this means that there is an open subwalk from α to β with a head at β which we can concatenate with ω_2. If ε is a collider (other than right before the final β), then we can remove the cycle from ε to ε from the walk. In any case, we can find a connecting collider walk in $\mathcal{G}_1$ (no noncolliders) such that α, β, and ε will each occur once. This means that the subwalk only contains nodes from a single line segment. This segment cannot be N_ε, N_β, $\bar{N}_\varepsilon$, nor $\bar{N}_\beta$ as α is not adjacent with any node in these line segments. If the walk only intersects the V^--line segment, then it must either go through Φ-nodes or the $X \cup \Lambda$-nodes, not both, as it has no noncolliders (or such a walk can be found). If it does not visit any χ- or λ-nodes, then there is an open walk in the $\bar{\Gamma} \cup \bar{\Phi} \cup \bar{\Delta}$-segment (the analogous walk through the barred versions). Finally, assume it does not visit any Φ-nodes. As H is a tautology, there is also a conjunction segment in $\bar{\Phi}$ which is open and connecting from α to β with a head at β. If instead the bidirected walk is in $\bar{V}^-$, the result follows, and if ε and β occur in the opposite order on the original μ-connecting walk, we can use similar arguments.

If the formula is not a tautology, let A be an assignment of values such that the formula evaluates to false. We then consider the set

$$C^- = \{\chi_l, \nu_\varepsilon^{\chi_l}, \nu_\beta^{\chi_l} : x_l = 1 \text{ in } A\} \cup \{\lambda_l, \nu_\varepsilon^{\lambda_l}, \nu_\beta^{\lambda_l} : x_l = 0 \text{ in } A\} \cup \Gamma \cup \Delta.$$

We also define $C = \mathrm{an}(C^-) \cup \{\beta, \delta\}$. We see immediately that β is not μ-separated from α given C in $\mathcal{G}_1$ as the $\chi - \lambda$-segment contains an open path from α to ε with a head at ε and furthermore $\varepsilon \leftrightarrow \beta$ is in the graph. On the other hand, consider a potential μ-connecting walk from α to β in $\mathcal{G}$. If ε is on the walk, it can only return to α. It cannot go between bidirected components because the directed cycles are either completely contained in C or in its complement. It cannot go through a ϕ-component because of the choice of A, and we conclude that it cannot be μ-connecting. In conclusion, $\mathcal{G}$ and $\mathcal{G}_1$ are Markov equivalent if and only if H is a tautology.

The arguments in the proof of Theorem 3.2 show that $\mathcal{G}_1$ and $\mathcal{G}_2$ are Markov equivalent. Arguments similar to those in the proof of Theorem 3.2 furthermore show that Decision problems A.3 and A.4 are coNP-complete.

19

Careful examination of the graphs reveals that all three are 16-sparse. □

One should note that the graphs in the proof of Theorem 3.7 could also be interpreted as δ-separation graphs (Didelez, 2008). In this case, the result also holds, i.e., determining δ-separation Markov equivalence of sparse DMGs is also coNP-complete. To see this one should simply note that μ-separation Markov equivalence implies δ-separation Markov equivalence and that the conditioning set used in the proof when H is not a tautology contains β. The hardness result in the δ-separation case then follows from the (A.1) property of the supplementary material of Mogensen and Hansen (2020) and from noting that the latent projection technique can also be used for δ-separation.

Richardson (1997) studied DGs under d-separation and gave an example of 'nonlocality' in this setting. The example consisted of a sequence of pairs of graphs, $\mathcal{D}_n^1$ and $\mathcal{D}_n^2$, such that $\mathcal{D}_n^1$ and $\mathcal{D}_n^2$ are not Markov equivalent, but the only separation on which the graphs disagree involves nodes that are arbitrarily far apart (for increasing values of n). Our setting is quite different, however, DMGs under μ-separation do exibit the same 'nonlocality' as seen from the proof of Theorem 3.7. Say that H is not a tautology, in which case $\mathcal{G}$ and $\mathcal{G}_1$ in the proof of Theorem 3.7 are not Markov equivalent. From the proof, it follows that the graphs only disagree on triples (A, B, C) such that $\alpha \in A$ and $\beta \in B$, and this means that the proof (for non-tautological H of increasing size) gives a sequence of pairs of graphs that only disagree on μ-separation of a pair of nodes, α and β, that are arbitrarily far from each other as measured by the shortest path between α and β. Note that this also holds in the maximal Markov equivalent graphs of $\mathcal{G}$ and $\mathcal{G}_1$, and it is therefore not due to non-maximality.

3.2 Implications of hardness results

The hardness results have several implications that we will outline in this section, in particular, we argue that several other computational problems are also hard in μ-separation DMGs.

Every Markov equivalence class has a greatest element (Mogensen and Hansen, 2020), and one can decide if two DMGs are Markov equivalent by computing the greatest Markov equivalent graph for each of them and compare. This means that finding such a greatest element is also hard. There are similar implications for *oracle* learning algorithms. A *(local independence) oracle* is an abstract function which a learning algorithm may query and which, when provided with a triple (A, B, C), outputs whether the corresponding local independence holds or not. The oracle gives the correct answer, but when using real data, the oracle has to be replaced by hypothesis tests of local independence, and the purpose of the oracle formalism is simply to separate the algorithmic aspects from the hypothesis testing. If we assume that there exists a constraint-based learning algorithm which can recover a unique representative of the Markov equivalence class (say the greatest element, or some other uniquely defined representative) of the true graph from when given access to a local independence oracle, then using this algorithm, one can decide Markov equivalence by querying the μ-separation models of the graphs. This is done by testing μ-separation in the graph and each test is done in polynomial time (Mogensen, 2020b). If only a polynomial number of queries are required we could also solve Markov equivalence in polynomial time by comparing the output for two graphs. Again, this means that such a learning algorithm would need an exponential number of tests.

3.2.1 Sparse DMGs

All of the above holds even if we are willing to assume that all graphs are somewhat sparse (m-sparse, $m \geq 16$). This means that a restriction to sparse graphs will not remedy this. This is also different from DAG-based models in the following sense. In partially observed DAGs, we may learn a graphical representation of the equivalence class using tests of conditional independence. If we fix m such that the node degree is less than m, this can be done in polynomial time (Claassen et al., 2013).

These hardness results motivate the second part of this paper. Instead of requiring sparsity of the DMGs, we will reinterpret them to obtain a weaker type of equivalence. Essentially, the DMGs are too expressive leading to the above infeasibility results in connection to their Markov equivalence classes. We can avoid this by considering a weaker type of equivalence. This leads to a simple and useful theory and to practical graph learning algorithms as we will see in subsequent sections.

4 Weak equivalence

In this section, we introduce a notion of *weak equivalence* and argue that it provides a computationally feasible notion of equivalence of DMGs. Under a regularity condition, the associated equivalence classes each have a greatest element and this leads to a simple graphical theory.

4.1 Classes of weak equivalence

We define three types of equivalence in this section and present them in decreasing order of generality. They each limit the set of triples, (A, B, C), that are used to distinguish between independence models represented by DMGs.

4.1.1 General weak equivalence

If $\mathcal{G}_1 = (V, E_1)$ and $\mathcal{G}_2 = (V, E_2)$ are Markov equivalent, then $(A, B, C) \in \mathcal{I}(\mathcal{G}_1)$ if and only if $(A, B, C) \in \mathcal{I}(\mathcal{G}_2)$ for all $A, B, C \subseteq V$. This means that Markov equivalence requires the independence models of $\mathcal{G}_1$ and $\mathcal{G}_2$ to agree on all triplets in the set $\mathcal{P} = \{(A, B, C) : A, B, C \subseteq V\}$. A very general approach to defining weaker notions of equivalence is to only compare the independence models on a subset of $\mathcal{P}$.

Definition 4.1 (General weak equivalence). *Let $\mathcal{J} \subseteq \{(A, B, C) : A, B, C \subseteq V\}$. We say that $\mathcal{G}_1 = (V, E_1)$ and $\mathcal{G}_2 = (V, E_2)$ are $\mathcal{J}$-weakly equivalent if*

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

We use $\mathcal{I}_{\mathcal{J}}(\mathcal{G}_1)$ to denote the $\mathcal{J}$-weak independence model induced by $\mathcal{G}_1$, $\mathcal{I}_{\mathcal{J}}(\mathcal{G}_1) = \mathcal{I}(\mathcal{G}_1) \cap \mathcal{J}$. We use $[\mathcal{G}_1]_{\mathcal{J}}$ to denote the $\mathcal{J}$-weak equivalence class of $\mathcal{G}_1$, that is, the set of graphs, $\mathcal{G} = (V, E)$, such that $\mathcal{I}_{\mathcal{J}}(\mathcal{G}) = \mathcal{I}_{\mathcal{J}}(\mathcal{G}_1)$.

Proposition 4.2. *Let $\mathcal{J} \subseteq \mathcal{P}$ and let V be a finite set. Definition 4.1 defines an equivalence relation on the set of DMGs with node set V.*

Proof. Let $\mathcal{G}$ be a DMG. We see that $\mathcal{G}$ is $\mathcal{J}$-weakly equivalent with itself such that the relation is reflexive. The relation is also symmetric and transitive. $\square$

The next statement follows directly from the definition of weak equivalence.

Proposition 4.3. *Let $\mathcal{J}_1 \subseteq \mathcal{J}_2 \subseteq \mathcal{P}$ and let $\mathcal{G}$ be a DMG. It holds that $\mathcal{I}_{\mathcal{J}_1}(\mathcal{G}) \subseteq \mathcal{I}_{\mathcal{J}_2}(\mathcal{G})$.*

A Markov equivalence class has a greatest element. However, a $\mathcal{J}$-weak equivalence class does not necessarily have a greatest element as illustrated by the following example.

Example 4.4. *We consider the graph, $\mathcal{G}$, in Figure 9 with all loops included as well. We define the set $\mathcal{J} \subseteq \mathcal{P}$,*

$$\mathcal{J} = \left[\bigcup_{\alpha, \beta \in V} (\alpha, \beta, \beta) \right] \cup \left\{ (1, 5, \{2, 3, 4, 5\}) \right\}.$$

We also define three other graphs from $\mathcal{G} = (V, E)$, $\mathcal{G}_i = (V, E_i)$, where $i = 1, 2, 3$, and

$$E_1 = E \cup \{2 \leftrightarrow 3\}, \qquad E_2 = E \cup \{3 \leftrightarrow 4\}, \qquad E_3 = E \cup \{2 \leftrightarrow 3, 3 \leftrightarrow 4\}.$$

Graphs $\mathcal{G}_1$ and $\mathcal{G}_2$ are both $\mathcal{J}$-weakly equivalent with $\mathcal{G}$ which can be seen from simply listing their $\mathcal{J}$-weak independence models.

We see that $(1, 5, \{2, 3, 4, 5\}) \in \mathcal{I}_{\mathcal{J}}(\mathcal{G})$, but $(1, 5, \{2, 3, 4, 5\}) \notin \mathcal{I}_{\mathcal{J}}(\mathcal{G}_3)$ which means that $\mathcal{G}$ and $\mathcal{G}_3$ are not $\mathcal{J}$-weakly equivalent. We have that $\mathcal{G}_1, \mathcal{G}_2 \in [\mathcal{G}]_{\mathcal{J}}$, and a greatest element of $[\mathcal{G}]_{\mathcal{J}}$ must be a supergraph of both $\mathcal{G}_1$ and $\mathcal{G}_2$, and therefore of $\mathcal{G}_3$. If $\mathcal{N}$ is a supergraph of $\mathcal{G}_3$, then $\mathcal{I}_{\mathcal{J}}(\mathcal{N}) \subseteq \mathcal{I}_{\mathcal{J}}(\mathcal{G}_3) \subsetneq \mathcal{I}_{\mathcal{J}}(\mathcal{G})$, and we conclude that the $\mathcal{J}$-weak equivalence class of $\mathcal{G}$ does not contain a greatest element.
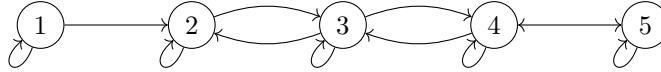

Figure 9: The graph $\mathcal{G}$ in Example 4.4.

Let $\mathcal{J} \subseteq \mathcal{P}$. If two graphs are Markov equivalent, they are of course also equivalent when restricting to comparisons on the set $\mathcal{J}$. Therefore, every graph is also weakly equivalent with the unique, maximal graph of its Markov equivalence class. However, the above example shows that when considering a general $\mathcal{J}$-weak equivalence, an equivalence class need not have a greatest element as the maximal Markov equivalent graph need not be a greatest element of the larger weak equivalence class. This leads us to introducing the notion of a *homogeneous* weak equivalence by imposing a regularity condition on the set $\mathcal{J}$. The equivalence classes of a homogeneous weak equivalence relation do indeed contain a greatest element (Section 5).

4.1.2 Homogeneous weak equivalence

We define *homogeneous* equivalence relation to obtain well-behaved equivalence classes.

Definition 4.5 (Homogeneous equivalence). *Consider some weak equivalence induced by $\mathcal{J} \subseteq \mathcal{P}$. We say that this equivalence is homogeneous if there exists a set $\mathcal{C}$, $\mathcal{C} \subseteq \{C : C \subseteq V\}$, such that*

$$\mathcal{J} = \{(A, B, C) \in \mathcal{P} : A, B \subseteq V, C \in \mathcal{C}\}.$$

In this case, we will also say that the set $\mathcal{J}$ is homogeneous *and we will say that $\mathcal{C}$ is the* collection of conditioning sets *of $\mathcal{J}$.*

In other words, a *homogeneous* equivalence relation is one that restricts only the set of conditioning sets, C. That is, if $\mathcal{J}$ is homogeneous, then $\mathcal{J}$-weak equivalence of $\mathcal{G}_1$ and $\mathcal{G}_2$ means that for all $A, B \subseteq V$ and $C \in \mathcal{C}$ we have $(A, B, C) \in \mathcal{I}(\mathcal{G}_1)$ if and only if $(A, B, C) \in \mathcal{I}(\mathcal{G}_2)$ where $\mathcal{C}$ is some collection of subsets of V. Therefore, the restriction of the independence model imposed by a homogeneous $\mathcal{J}$ only applies to the conditioning sets.

4.1.3 k-weak equivalence

We will now introduce a certain type of homogeneous equivalence which simply restricts the size of the conditioning sets.

Definition 4.6 (k-weak equivalence)**.** *Let $0 \leq k \leq n$. We say that $\mathcal{G}_1$ and $\mathcal{G}_2$ are k-weakly equivalent if for all C such that $|C| \leq k$, it holds that $(A, B, C) \in \mathcal{I}(\mathcal{G}_1)$ if and only if $(A, B, C) \in \mathcal{I}(\mathcal{G}_2)$.*

The above is formulated slightly differently than Definitions 4.1 and 4.5, however, k-weak equivalence is a homogeneous weak equivalence relation by using the set $\mathcal{C} = \{C \subseteq V : |C| \leq k\}$ in Definition 4.5. On the other hand, not all homogeneous equivalences correspond to a k-weak equivalence. We see that k-weak equivalence only compares graphs using 'small' conditioning sets of size less than k and that Markov equivalence is the same as n-weak equivalence.

For $\mathcal{G}_1 = (V, E_1)$, we use $\mathcal{I}_k(\mathcal{G}_1)$ to denote the k-weak independence model of $\mathcal{G}_1$, $\mathcal{I}_k(\mathcal{G}_1) = \{(A, B, C) \in \mathcal{I}(\mathcal{G}_1), |C| \leq k\}$. We let $[\mathcal{G}]_k$ denote the set of graphs on nodes V that are k-weakly equivalent with $\mathcal{G}$, and we say that $[\mathcal{G}]_k$ is the k-weak equivalence class of $\mathcal{G}$. When $k = n$, we also use $\mathcal{I}(\mathcal{G})$, that is, $\mathcal{I}(\mathcal{G}) = \mathcal{I}_n(\mathcal{G})$.

4.2 Properties of weak equivalence

This section describes some properties of weak equivalence and weak equivalence classes. Throughout the section $\mathcal{J}$ is a subset of $\mathcal{P} = \{(A, B, C) : A, B, C \subseteq V\}$. For Markov equivalence, it holds that $\mathcal{G}_1 \subseteq \mathcal{G}_2$ implies $\mathcal{I}(\mathcal{G}_2) \subseteq \mathcal{I}(\mathcal{G}_1)$ which follows from the definition of μ-separation. This is quite natural as a larger graph has more edges, therefore fewer independences. The same holds for weak equivalence classes as shown by the next proposition.

Proposition 4.7. *If $\mathcal{G}_1 \subseteq \mathcal{G}_2$, then $\mathcal{I}_{\mathcal{J}}(\mathcal{G}_2) \subseteq \mathcal{I}_{\mathcal{J}}(\mathcal{G}_1)$.*

Proof. If $(A, B, C) \in \mathcal{I}_{\mathcal{J}}(\mathcal{G}_2)$ then $(A, B, C) \in \mathcal{I}(\mathcal{G}_2)$ and $(A, B, C) \in \mathcal{J}$, and therefore $(A, B, C) \in \mathcal{I}(\mathcal{G}_1)$. This means that $(A, B, C) \in \mathcal{I}_{\mathcal{J}}(\mathcal{G}_1)$. $\qquad\square$

Proposition 4.8 (Well-ordered $\mathcal{J}$-classes)**.** *Let $\mathcal{J}_1 \subseteq \mathcal{J}_2 \subseteq \mathcal{P}$. If $\mathcal{G}_1$ and $\mathcal{G}_2$ are $\mathcal{J}_2$-weakly equivalent, then they are also $\mathcal{J}_1$-weakly equivalent.*

Proof. Let $(A, B, C) \in \mathcal{I}_{\mathcal{J}_1}(\mathcal{G}_1)$, then $(A, B, C) \in \mathcal{I}(\mathcal{G}_1)$ and $(A, B, C) \in \mathcal{J}_1$. Therefore $(A, B, C) \in \mathcal{J}_2$ and $(A, B, C) \in \mathcal{I}_{\mathcal{J}_2}(\mathcal{G}_1) = \mathcal{I}_{\mathcal{J}_2}(\mathcal{G}_2)$. It follows that $(A, B, C) \in \mathcal{I}(\mathcal{G}_2)$ and $(A, B, C) \in \mathcal{I}_{\mathcal{J}_1}(\mathcal{G}_2)$. Interchanging the roles of $\mathcal{G}_1$ and $\mathcal{G}_2$ and repeating the argument gives the result. $\qquad\square$

From the above, we also see that $\mathcal{J}_1 \subseteq \mathcal{J}_2$ implies $[\mathcal{G}]_{\mathcal{J}_2} \subseteq [\mathcal{G}]_{\mathcal{J}_1}$. The next corollary follows directly from the above proposition.

Corollary 4.9 (Well-ordered k-classes). *Let $0 \leq k_1 \leq k_2 \leq n$. If $\mathcal{G}_1$ and $\mathcal{G}_2$ are k_2-weakly equivalent, then they are also k_1-weakly equivalent.*

Definition 4.10. *We say that $\mathcal{J}$ is singleton stable if for all $A, B, C \subseteq V$, $(A, B, C) \in \mathcal{J}$ implies that $(\alpha, \beta, C) \in \mathcal{J}$ for all $\alpha \in A$ and $\beta \in B$.*

Note that the requirement is only on the A- and B-sets, not the C-set. If $\mathcal{J}$ is homogeneous and $(A, B, C) \in \mathcal{J}$, then $(\bar{A}, \bar{B}, C) \in \mathcal{J}$ for all $\bar{A}, \bar{B} \subseteq V$, thus a homogeneous $\mathcal{J}$ is also singleton stable. The following proposition shows, for a singleton stable $\mathcal{J}$, the independence model $\mathcal{I}_{\mathcal{J}}(\mathcal{G})$ is characterized by the independences (A, B, C) where A and B are singletons and A and C are disjoint. This proof uses the fact that μ-separation models satisfy so-called left and right composition as well as left and right decomposition which are *asymmetric graphoid properties* (Didelez, 2006; Mogensen et al., 2018). These are similar to classical graphoid properties (Lauritzen, 1996), but left and right version are needed due to the lack of symmetry.

Proposition 4.11. *Let $\mathcal{J}$ be singleton stable, let V be a finite set and let $\mathcal{S} = \{(A, B, C) \in \mathcal{P} : |A| = |B| = 1, A \cap C = \emptyset\}$. If $\mathcal{I}_{\mathcal{J}}(\mathcal{G}_1) \cap \mathcal{S} \subseteq \mathcal{I}_{\mathcal{J}}(\mathcal{G}_2) \cap \mathcal{S}$, then $\mathcal{I}_{\mathcal{J}}(\mathcal{G}_1) \subseteq \mathcal{I}_{\mathcal{J}}(\mathcal{G}_2)$.*

Without the assumption of singleton stability, this above statement is not true. For instance, if $\mathcal{J} \cap \mathcal{S} = \emptyset$, then $\mathcal{I}(\mathcal{G}_1)_{\mathcal{J}} \cap \mathcal{S} \subseteq \mathcal{I}(\mathcal{G}_2)_{\mathcal{J}} \cap \mathcal{S}$ is trivially true for any pair of graphs.

Proof. Let $(A, B, C) \in \mathcal{I}_{\mathcal{J}}(\mathcal{G}_1)$. If A or B is empty, then it follows immediately that $(A, B, C) \in \mathcal{I}_{\mathcal{J}}(\mathcal{G}_2)$. Assume that A and B are both nonempty. We can write $A = \{\alpha_1, \ldots, \alpha_{n_a}\}$ and $B = \{\beta_1, \ldots, \beta_{n_b}\}$. From the definition of μ-separation and using singleton stability of $\mathcal{J}$ it follows that $(\alpha_i, \beta_j, C) \in \mathcal{I}_{\mathcal{J}}(\mathcal{G}_1)$ for all $i = 1, \ldots, n_a$ and $j = 1, \ldots, n_b$. Therefore $(\alpha_i, \beta_j, C) \in \mathcal{I}_{\mathcal{J}}(\mathcal{G}_2)$ for all $i = 1, \ldots, n_a$ and $j = 1, \ldots, n_b$ (if $\alpha_i \in C$, then it holds trivially). From the definition of μ-separation, $(A, B, C) \in \mathcal{I}(\mathcal{G}_2)$ and therefore also $(A, B, C) \in \mathcal{I}_{\mathcal{J}}(\mathcal{G}_2)$. $\square$

Proposition 4.12 (Maximality). *The graph $\mathcal{G} = (V, E) \in [\mathcal{G}_1]_{\mathcal{J}}$ is maximal in $[\mathcal{G}_1]_{\mathcal{J}}$ if and only if it is complete or if $\mathcal{G} + e \notin [\mathcal{G}_1]_{\mathcal{J}}$ for all edges e such that $e \notin E$.*

When $\mathcal{G} = (V, E) \in [\mathcal{G}_1]_{\mathcal{J}}$ is maximal in $[\mathcal{G}_1]_{\mathcal{J}}$, then we also say that $\mathcal{G}$ is $\mathcal{J}$-maximal (the equivalence class is implicit as a graph can only be maximal in its own equivalence class). A graph is $\mathcal{J}$-maximal if the addition of any edge will change the $\mathcal{J}$-weak independence model.

Proof. If $\mathcal{G}$ is complete, then it is clearly maximal. If $\mathcal{G} \subsetneq \mathcal{G}_2$, then $\mathcal{G} \subsetneq \mathcal{G} + e \subseteq \mathcal{G}_2$ for some $e \notin E$. We have $\mathcal{I}_{\mathcal{J}}(\mathcal{G}_2) \subseteq \mathcal{I}_{\mathcal{J}}(\mathcal{G} + e)$ and $\mathcal{I}_{\mathcal{J}}(\mathcal{G} + e) \subsetneq \mathcal{I}_{\mathcal{J}}(\mathcal{G}) = \mathcal{I}_{\mathcal{J}}(\mathcal{G}_1)$ and therefore $\mathcal{G}_2 \notin [\mathcal{G}_1]_{\mathcal{J}}$.

On the other hand, assume that $\mathcal{G}$ is maximal, and that $\mathcal{G}$ is not complete. It follows from the definition of maximality that $\mathcal{G} + e \notin [\mathcal{G}]_{\mathcal{J}}$ for all $e \notin E$. $\square$

If $\mathcal{G}_1 \subseteq \mathcal{G}_2$ then $\mathcal{I}(\mathcal{G}_2) \subseteq \mathcal{I}(\mathcal{G}_1)$ (Proposition 4.7). One may ask if $\mathcal{I}(\mathcal{G}_2) \subseteq \mathcal{I}(\mathcal{G}_1)$ implies $\mathcal{G}_1 \subseteq \mathcal{G}_2$. The next example shows that this is not the case, also not for maximal graphs.

Example 4.13. *We consider two graphs, $\mathcal{G}_1$ and $\mathcal{G}_2$ as shown in Figure 10 (both graphs also have all directed and bidirected loops). Let $\mathcal{S} = \{(A, B, C) : |A| = |B| = 1, A \cap C = \emptyset\}$. Then $\mathcal{I}(\mathcal{G}_1) \cap \mathcal{S}$ equals*

$$\left\{ \Big(2,3,1\Big), \Big(2,3,\{1,3\}\Big), \Big(3,2,1\Big), \Big(3,2,\{1,2\}\Big), \Big(3,1,1\Big), \Big(3,1,\{1,2\}\Big) \right\}.$$

$\mathcal{I}(\mathcal{G}_2) \cap \mathcal{S}$ *equals*

$$\left\{ \Big(2,3,\{1,3\}\Big), \Big(3,1,1\Big) \right\}$$

and therefore it is a subset of $\mathcal{I}(\mathcal{G}_1) \cap \mathcal{S}$. Markov equivalence corresponds to $\mathcal{J}$-weak equivalence with $\mathcal{J} = \mathcal{P}$, and by Proposition 4.11, $\mathcal{I}(\mathcal{G}_2) \subseteq \mathcal{I}(\mathcal{G}_1)$. Both graphs are maximal which means that $2 \to 1$ cannot be added to $\mathcal{G}_2$ Markov equivalently. This illustrates that $\mathcal{I}(\mathcal{G}_2) \subseteq \mathcal{I}(\mathcal{G}_1)$ does not imply $\mathcal{G}_1 \subseteq \mathcal{G}_2$, not even if $\mathcal{G}_2$ is maximal.

Figure 10: Graphs $\mathcal{G}_1$ (left) and $\mathcal{G}_2$ (right) in Example 4.13. Loops are omitted from the visualization.

Proposition 4.14. *Let $\mathcal{J}_1 \subseteq \mathcal{J}_2$. If $\mathcal{G}$ is $\mathcal{J}_1$-maximal, then it is also $\mathcal{J}_2$-maximal.*

Proof. If $\mathcal{G}$ is complete, then it is also $\mathcal{J}_2$-maximal. Assume instead that $\mathcal{G} = (V, E)$ is not complete and $e \notin E$. $\mathcal{G}$ is $\mathcal{J}_1$-maximal, so $\mathcal{G} + e \notin [\mathcal{G}]_{\mathcal{J}_1}$ (Proposition 4.12). Using Proposition 4.7, there exist a triple (A, B, C) such that $(A, B, C) \in \mathcal{I}_{\mathcal{J}_1}(\mathcal{G})$ and $(A, B, C) \notin \mathcal{I}_{\mathcal{J}_1}(\mathcal{G} + e)$ and therefore $(A, B, C) \notin \mathcal{I}(\mathcal{G} + e)$. We see that $(A, B, C) \in \mathcal{I}_{\mathcal{J}_2}(\mathcal{G})$ and $(A, B, C) \notin \mathcal{I}_{\mathcal{J}_2}(\mathcal{G} + e)$. It follows that $\mathcal{G}$ is $\mathcal{J}_2$-maximal (Proposition 4.12). $\qquad\square$

We say that a graph, $\mathcal{G}$, is *k-maximal* if is $\mathcal{J}$-maximal for $\mathcal{J} = \{(A, B, C) \in \mathcal{P} : |C| \leq k\}$ which means that $\mathcal{J}$ induces a k-weak equivalence relation.

Corollary 4.15. *Let $0 \leq k_1 \leq k_2 \leq n$. If $\mathcal{G}$ is k_1-maximal, then it is also k_2-maximal.*

In particular, if a graph is k-maximal for some $k \leq n$, then it is also the unique maximal element in its Markov equivalence class.

Proposition 4.16 (Minimality)**.** *The graph $\mathcal{G} = (V, E) \in [\mathcal{G}_1]_{\mathcal{J}}$ is minimal in $[\mathcal{G}_1]_{\mathcal{J}}$ if and only if it is empty or if $\mathcal{G} - e \notin [\mathcal{G}_1]_{\mathcal{J}}$ for all edges such that $e \in E$.*

Proof. If it is empty, then it is clearly also minimal. Otherwise, let $\mathcal{G}_2 \subsetneq \mathcal{G}$. We have $\mathcal{I}_{\mathcal{J}}(\mathcal{G}) \subsetneq \mathcal{I}_{\mathcal{J}}(\mathcal{G} - e) \subseteq \mathcal{I}_{\mathcal{J}}(\mathcal{G}_2)$ for $e \in E$ (Proposition 4.7). Therefore, $\mathcal{G}_2 \notin [\mathcal{G}]_{\mathcal{J}}$.

If $\mathcal{G}$ is minimal in $[\mathcal{G}_1]_{\mathcal{J}}$, then it is either the empty graph, or for all $e \in E$, $\mathcal{G} - e \notin [\mathcal{G}_1]_{\mathcal{J}}$ by definition of minimality. $\qquad\square$

Proposition 4.17. *Let $\mathcal{J}_1 \subseteq \mathcal{J}_2$. If $\mathcal{G}$ is $\mathcal{J}_1$-minimal, then it is also $\mathcal{J}_2$-minimal.*

The proposition states that the property of being minimal is preserved when considering a larger set of independences. An equivalence class is finite and nonempty, hence, it always contains a maximal element and a minimal element. We will show later that it also contains a greatest element. However, a least element need not exist and Example 6.2 provides an example of this.

Proof. If $\mathcal{G} = (V, E)$ is empty, then it is also $\mathcal{J}_2$-minimal. Assume instead that $e \in E$. There exists a triple (A, B, C) such that $(A, B, C) \in \mathcal{I}_{\mathcal{J}_1}(\mathcal{G} - e)$ and $(A, B, C) \notin \mathcal{I}_{\mathcal{J}_1}(\mathcal{G})$. Then $(A, B, C) \in \mathcal{J}_1$ and therefore $(A, B, C) \in \mathcal{J}_2$. It follows that $(A, B, C) \in \mathcal{I}_{\mathcal{J}_2}(\mathcal{G} - e)$ and $(A, B, C) \notin \mathcal{I}_{\mathcal{J}_2}(\mathcal{G})$. As this holds for all $e \in E$, we see that $\mathcal{G}$ is $\mathcal{J}_2$-minimal (Proposition 4.16). □

4.2.1 Marginalization

We say that a class of graphs, $\mathbb{G}$, is *closed under marginalization* if for every $\mathcal{G} = (V, E) \in \mathbb{G}$ and every $O \subseteq V$ there exists $\mathcal{M} = (O, E_O) \in \mathbb{G}$ such that for every $A, B, C \subseteq O$,

$$(A, B, C) \in \mathcal{I}_\star(\mathcal{G}) \Leftrightarrow (A, B, C) \in \mathcal{I}_\star(\mathcal{M}) \tag{1}$$

where $\mathcal{I}_\star(\mathcal{G})$ is the independence model induced by $\mathcal{G}$. When $\mathbb{G}$ is the class of DMGs, $\mathcal{I}_\star(\cdot)$ could for instance be a $\mathcal{J}$-weak independence model. Appendix C shows that DMGs with weak equivalence are closed under marginalization. This follows directly from the analogous result in the case of Markov equivalence (Mogensen and Hansen, 2020) using a so-called *latent projection* (see also Verma and Pearl, 1990a; Richardson et al., 2017).

4.3 k-weak equivalence

In this subsection, we restrict our attention to k-weak equivalence relations. The following result shows that if $|V| = n$, then n-weak and $(n - 1)$-weak equivalence is the same. By convention, β is always μ-separated from α given C when $\alpha \in C$. If $|C| = n$, then $C = V$, and leads to a trivial separation.

Proposition 4.18. *Let $\mathcal{G}_1 = (V, E_1)$ and $\mathcal{G}_2 = (V, E_2)$ such that $|V| = n$. Graphs $\mathcal{G}_1$ and $\mathcal{G}_2$ are $(n - 1)$-weakly equivalent if and only if they are n-weakly equivalent.*

Proof. If $\mathcal{G}_1$ and $\mathcal{G}_2$ are n-weakly equivalent, then they are also $(n - 1)$-weakly equivalent.

On the other hand, assume that $\mathcal{G}_1$ and $\mathcal{G}_2$ are $(n - 1)$-weakly equivalent, and let $(\alpha, \beta, C) \in \mathcal{I}_n(\mathcal{G}_1)$ such that $\alpha, \beta \in V$, $C \subseteq V$, and $\alpha \notin C$. We must then have $|C| \leq n - 1$, and therefore $(\alpha, \beta, C) \in \mathcal{I}_n(\mathcal{G}_2)$ by $(n - 1)$-weak equivalence of $\mathcal{G}_1$ and $\mathcal{G}_2$. By Proposition 4.11, this implies $\mathcal{I}_n(\mathcal{G}_1) \subseteq \mathcal{I}_n(\mathcal{G}_2)$. Changing the roles of $\mathcal{G}_1$ and $\mathcal{G}_2$ completes the argument. □
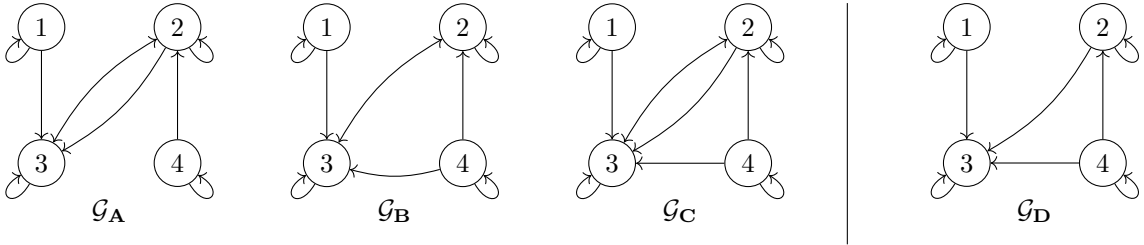

Figure 11: Graphs from Example 4.19. All bidirected loops are present in the graphs but omitted from the visualization.

Example 4.19 (Weak equivalence class)**.** *In this example, we restrict our attention to graphs with all loops included in which case graphs $\mathcal{G}_A$, $\mathcal{G}_B$, and $\mathcal{G}_C$ in Figure 4.19 constitute a 2-weak*

equivalence class and a 3-weak equivalence class. Graph $\mathcal{G}_C$ is the greatest element in both cases. We have that $[\mathcal{G}_C]_2 \subseteq [\mathcal{G}_C]_1$ (Corollary 4.9) and $[\mathcal{G}_C]_1 = \{\mathcal{G}_A, \mathcal{G}_B, \mathcal{G}_C, \mathcal{G}_D\}$. We see that $\mathcal{G}_C$ and $\mathcal{G}_D$ are not 2-weakly equivalent as 2 is μ-separated from 3 given $\{2,4\}$ in $\mathcal{G}_D$ while this is not the case in $\mathcal{G}_C$.

Example 4.20. *We give an example of how 'strong connectivity', that is, many similar paths, may lead to more edges in a k-weak graph than in the corresponding n-weak graph, $k \leq n$. For this purpose, we consider graphs $\mathcal{G}_1$ and $\mathcal{G}_2$ as shown in Figure 12. The graph $\mathcal{G}_2$ is 2-maximal and therefore it is k-maximal for all $k \geq 2$, including $k = n$ (Corollary 4.15). We construct a smaller graph, $\mathcal{G}_1$, by removing $1 \to 2$. The smaller graph is not Markov equivalent, but it is $(n-3)$-equivalent.*

In terms of interpretation, we see that in this class of graphs there are many directed paths from α to β and if there are more than k, then the edge $\alpha \to \beta$ can be added k-weakly equivalently. In a graphical sense, nodes α and β are 'strongly' connected as there are more than k disjoint, directed paths from α to β and they cannot all be blocked by conditioning on at most k nodes.
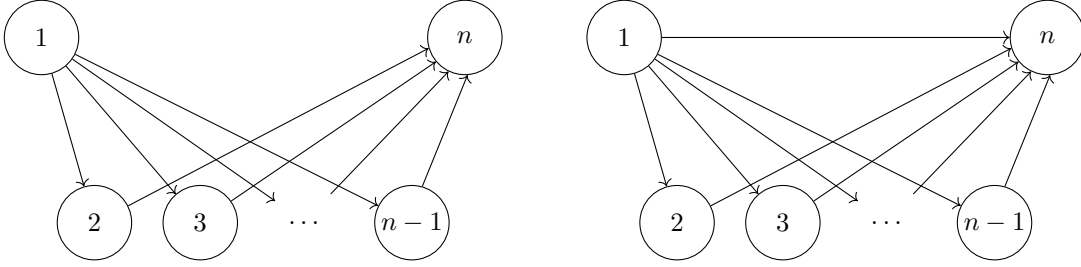

Figure 12: Graphs $\mathcal{G}_1$ (left) and $\mathcal{G}_2$ (right) from Example 4.20. All loops are present in the graphs, but not shown above.

We now define *treks* and *directed treks* (see also Foygel et al., 2012; Mogensen, 2020a). Foygel et al. (2012); Mogensen (2020a) used paths in their definitions of treks, however, we use walks such that treks between α and α are also allowed.

Definition 4.21 (Trek, directed trek). *Let ω be a nontrivial walk between α and β,*

$$\alpha \sim \ldots \sim_e \beta.$$

We say that ω is a trek *if it has no colliders. We say that a trek is* directed *from α to β if $\sim_e$ has a head at β.*

We let $\mathrm{dtr}_{\mathcal{G}}(\beta) \subseteq V$ denote the set of nodes, α, such that there exists a directed trek from α to β in $\mathcal{G}$.

Definition 4.22. *Let $\mathcal{G}_1$ and $\mathcal{G}_2$ be DMGs. We say that $\mathcal{G}_1$ and $\mathcal{G}_2$ are* trek equivalent *if for all $\beta \in V$, it holds that*

$$\mathrm{dtr}_{\mathcal{G}_1}(\beta) = \mathrm{dtr}_{\mathcal{G}_2}(\beta).$$
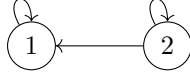
Figure 13: The above graph is trek equivalent with the complete graph, and therefore also 0-weakly equivalent with the complete graph (Corollary 4.23).
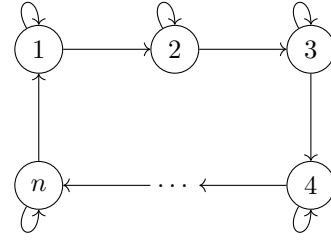

Figure 14: Directed cycle, see Proposition 4.24.

A walk is μ-connecting from α to β given $\emptyset$ if and only if it is a directed trek from α to β which is reflected in the next corollary.

Corollary 4.23. *Graphs $\mathcal{G}_1$ and $\mathcal{G}_2$ are 0-weakly equivalent if and only if they are trek equivalent.*

Proof. This follows from Corollary E.3. □

In Corollary 4.23, it is important to define treks using walks, not paths. For instance, the graph in Figure 13 is 0-weak equivalent with the complete graph, but the only directed treks from 1 to 2 is not are paths. Therefore, the result in Corollary 4.23 does not hold if directed treks are required to be paths. We say that a DMG $\mathcal{G} = (V, E)$, $V = \{1, 2, \ldots, n\}$, *contains a directed cycle* if there is some permutation of V, σ, such that $\sigma(1) \to \sigma(2) \to \ldots \to \sigma(n-1) \to \sigma(n) \to \sigma(1)$ in $\mathcal{G}$ (see an example in Figure 14).

Proposition 4.24. *Let $\mathcal{G} = (V, E)$ be a DMG, $V = \{1, 2, \ldots, n\}$, which contains a directed cycle. If every node has a loop, then the complete DMG on V is the greatest element of both $[\mathcal{G}]_0$ and $[\mathcal{G}]_1$.*

Proof. For $k = 0$, this follows from Corollary 4.23 as there is a directed trek between any ordered pair of nodes in $\mathcal{G}$. Let $k = 1$ and consider nodes α and β. We show that there is no separating set, C, such that $C \leq 1$. If $C = \emptyset$, this is clear. If $C = \{\gamma\}$, $\gamma \neq \alpha$, then either $\alpha \ast\!\to \ldots \to \beta$ is open, or $\gamma \neq \beta$ and $\alpha \leftarrow \ldots \leftarrow \beta \ast\!\to \beta$ is open. □

5 Greatest elements under homogeneous weak equivalences

In the rest of the paper, we assume every weak equivalence relation to be *homogeneous* (Definition 4.5) as this leads to the existence of a greatest element in each equivalence class which we will prove in Subsection 5.2. Mogensen and Hansen (2020) showed the analogous result in the case of Markov equivalence classes. The notions of *C-potential siblings* and *C-potential parents* are central to this proof and are introduced in the next subsection.

28

5.1 C-potential siblings and C-potential parents

The existence of a greatest element in each $\mathcal{J}$-weak equivalence class can be proven using C-*potential siblings* and C-*potential parents* as introduced in Definitions 5.1 and 5.2. We say that two graphs, $\mathcal{G}_1 = (V, E_1)$ and $\mathcal{G}_2 = (V, E_2)$, are C-*equivalent*, $C \subseteq V$, if for all $\gamma, \delta \in V$,

$$(\gamma, \delta, C) \in \mathcal{I}(\mathcal{G}_1) \Leftrightarrow (\gamma, \delta, C) \in \mathcal{I}(\mathcal{G}_2).$$

Let $\alpha, \beta \in V$ and let e be the edge $\alpha \leftrightarrow \beta$. The conditions (cs1)-(cs3) in Definition 5.1 are sufficient and necessary for $\mathcal{G}$ and $\mathcal{G} + e$ to be C-equivalent. When e is directed, the conditions (cp1)-(cp4) in Definition 5.2 are analogously necessary and sufficient for $\mathcal{G}$ and $\mathcal{G} + e$ to be C-equivalent. The sufficiency is proven in Lemmas D.2 and D.3 and the necessity follows from applying Propositions 5.5 and 5.6 to $\mathcal{G} + e$.

Definitions 5.1 and 5.2 use an abstract independence model, $\mathcal{I}$, while Propositions 5.3 and 5.4 describe the content of those definitions in the case of a graphical independence model, $\mathcal{I} = \mathcal{I}(\mathcal{G})$.

Definition 5.1 (C-potential sibling). *Let $\mathcal{I}$ be an independence model over V, let $\alpha, \beta \in V$, and let $C \subseteq V$. We say that α and β are C-potential siblings in $\mathcal{I}$ if (cs1)-(cs3) hold.*

(cs1) if $\alpha \notin C$: $(\alpha, \beta, C) \notin \mathcal{I}$, and
  if $\beta \notin C$: $(\beta, \alpha, C) \notin \mathcal{I}$

(cs2) if $\beta \in C$: for all $\gamma \in V$,
$$(\gamma, \alpha, C) \in \mathcal{I} \Rightarrow (\gamma, \beta, C) \in \mathcal{I}$$

(cs3) if $\alpha \in C$: for all $\gamma \in V$,
$$(\gamma, \beta, C) \in \mathcal{I} \Rightarrow (\gamma, \alpha, C) \in \mathcal{I}$$

Definition 5.2 (C-potential parent). *Let $\mathcal{I}$ be an independence model over V, let $\alpha, \beta \in V$, and let $C \subseteq V$. We say that α is a C-potential parent of β in $\mathcal{I}$ if (cp1)-(cp4) hold.*

(cp1) if $\alpha \notin C$: $(\alpha, \beta, C) \notin \mathcal{I}$

(cp2) if $\alpha \notin C$: for all $\gamma \in V$,
$$(\gamma, \beta, C) \in \mathcal{I} \Rightarrow (\gamma, \alpha, C) \in \mathcal{I}$$

(cp3) if $\alpha \notin C, \beta \in C$: for all $\gamma, \delta \in C$,
$$(\gamma, \delta, C) \in \mathcal{I} \Rightarrow (\gamma, \beta, C) \in \mathcal{I} \vee (\alpha, \delta, C) \in \mathcal{I}$$

(cp4) if $\alpha, \beta \notin C$: for all $\gamma \in V$,
$$(\beta, \gamma, C) \in \mathcal{I} \Rightarrow (\alpha, \gamma, C) \in \mathcal{I}$$

If $\mathcal{I}$ is graphical, $\mathcal{I} = \mathcal{I}(\mathcal{G})$, and α and β are C-potential siblings in $\mathcal{I}(\mathcal{G})$, we will say that $\alpha \leftrightarrow \beta$ is a C-*potential sibling edge* between α and β. Similarly, we will say that $\alpha \rightarrow \beta$ is a C-*potential parent edge* from α to β if α is a C-potential parent of β in $\mathcal{I}(\mathcal{G})$. The following two propositions simply rewrite Definitions 5.1 and 5.2 to explicitly use μ-connecting walks in the case of a graphical independence model. Their proofs follow directly from the definitions of μ-separation and the independence model $\mathcal{I}_{\mathcal{J}}(\mathcal{G})$.

29

Proposition 5.3 (Graphical version of C-potential siblings)**.** *Let $\mathcal{I}_{\mathcal{J}}(\mathcal{G})$ be the weak independence model induced by $\mathcal{G} = (V, E)$. Let $C \subseteq V$ and let $\mathcal{C}$ be the collection of conditioning sets of $\mathcal{J}$. Nodes α and β are C-potential siblings if and only if $C \notin \mathcal{C}$ or (gcs1)-(gcs3) holds.*

(gcs1) If $\alpha \notin C$, there exists a μ-connecting walk from α to β given C, and

if $\beta \notin C$, there exists a μ-connecting walk from β to α given C.

(gcs2) If $\beta \in C$, then for all $\gamma \in V$ such that there exists a μ-connecting walk from γ to β given C, there also exists a μ-connecting walk from γ to α given C.

(gcs3) If $\alpha \in C$, then for all $\gamma \in V$ such that there exists a μ-connecting walk from γ to α given C, there also exists a μ-connecting walk from γ to β given C.

Proposition 5.4 (Graphical version of C-potential parents)**.** *Let $\mathcal{I}_{\mathcal{J}}(\mathcal{G})$ be the weak independence model induced by $\mathcal{G} = (V, E)$. Let $C \subseteq V$ and let $\mathcal{C}$ be the collection of conditioning sets of $\mathcal{J}$. The node α is a C-potential parent of β if and only if $C \notin \mathcal{C}$ or (gcp1)-(gcp4) holds.*

(gcp1) If $\alpha \notin C$, there exists a μ-connecting walk from α to β given C.

(gcp2) If $\alpha \notin C$, then for all $\gamma \in V$ such that there exists a μ-connecting walk from γ to α given C, there also exists a μ-connecting walk from γ to β given C.

(gcp3) If $\alpha \notin C$ and $\beta \in C$, then for all $\gamma, \delta \in V$ such that there exists a μ-connecting walk from γ to β given C and a μ-connecting walk from α to δ given C, there also exists a μ-connecting walk from γ to δ given C.

(gcp4) If $\alpha, \beta \notin C$ then for all $\gamma \in V$ such that there exists a μ-connecting walk from α to γ given C, there also exists a μ-connecting walk from β to γ given C.

The next two propositions show that if $\alpha \leftrightarrow \beta$ ($\alpha \to \beta$) is in a graph, then α and β are C-potential siblings (α is a C-potential parent of β) in the independence model of the graph for all $C \subseteq V$. The edge e is therefore a C-potential sibling edge (C-potential parent edge) in $\mathcal{I}(\mathcal{G} + e)$, and if $\mathcal{G}$ and $\mathcal{G} + e$ are C-equivalent, then e is also a C-potential sibling edge (C-potential parent edge) in $\mathcal{I}(\mathcal{G})$. This means that $\mathcal{I}(\mathcal{G})$ satisfying the conditions in Definitions 5.1 and 5.2 is necessary for C-equivalence of $\mathcal{G}$ and $\mathcal{G} + e$.

Proposition 5.5. *Let $\mathcal{J}$ be homogeneous. If $\alpha \leftrightarrow \beta$ is in $\mathcal{G}$, then α and β are C-potential siblings in $\mathcal{I}_{\mathcal{J}}(\mathcal{G})$ for all $C \subseteq V$.*

Proof. If $C \notin \mathcal{C}$, then it follows immediately. We assume $C \in \mathcal{C}$ and prove (gcs1)-(gcs3). (gcs1) If $\alpha \notin C$, then $\alpha \leftrightarrow \beta$ is a μ-connecting walk in $\mathcal{G}$ given C. The proof of the other statement is analogous. (gcs2) Assume that $\beta \in C$ and let $\gamma \in V$ such that there exists a μ-connecting walk from γ to β given C. Composing this with $\beta \leftrightarrow \alpha$ gives a μ-connecting walk from γ to α given C as $\beta \in C$. (gcs3) This is shown similarly to (gcs2). $\qquad\square$

Proposition 5.6. *Let $\mathcal{J}$ be homogeneous. If $\alpha \to \beta$ is in $\mathcal{G}$, then α is a C-potential parent of β in $\mathcal{I}_{\mathcal{J}}(\mathcal{G})$ for all $C \subseteq V$.*

Proof. If $C \notin \mathcal{C}$, then this again follows immediately. We instead assume $C \in \mathcal{C}$ and prove (gcp1)-(gcp4). (gcp1) If $\alpha \notin C$, then $\alpha \to \beta$ is a μ-connecting walk given C. (gcp2) Assume that $\alpha \notin C$ and let $\gamma \in V$, and assume there is a μ-connecting walk from γ to α given C. Concatenating this with the edge $\alpha \to \beta$ gives a μ-connecting walk from γ to β given C as $\alpha \notin C$. (gcp3) Assume that $\alpha \notin C, \beta \in C$ and let $\gamma, \delta \in V$ such that there exist a μ-connecting walk from γ to β given C and a μ-connecting walk from α to δ given C. Concatenating them with the edge $\alpha \to \beta$ gives a μ-connecting walk from γ to δ given C as $\beta \in C$ and $\alpha \notin C$. (gcp4) Assume $\alpha, \beta \notin C$ and let $\gamma \in V$ such that there exists a μ-connecting walk from α to γ given C. Concatenating the edge $\alpha \to \beta$ with this walk gives a μ-connecting walk from β to γ given C as $\alpha, \beta \notin C$. $\qquad \square$

5.2 Existence of greatest elements

Markov equivalence classes of DMGs are known to contain a greatest element (Mogensen and Hansen, 2020). This means that for an equivalence class $[\mathcal{G}]$, there exists a graph $\bar{\mathcal{G}} \in [\mathcal{G}]$ such that $\bar{\mathcal{G}}$ is a supergraph of all graphs $\tilde{\mathcal{G}} \in [\mathcal{G}]$. This is a very convenient result as it allows a succinct representation of the entire Markov equivalence class as illustrated in Example 2.13. The main result of this section, Theorem 5.8, shows that $\mathcal{J}$-weak equivalence classes enjoy the same property when $\mathcal{J}$ is homogeneous. This means that we can represent weak equivalence classes in a similar way. Section 6 discusses this further and introduces a hierarchy of k-weak equivalence classes for different values of k.

Lemma 5.7. *Let $\mathcal{G}_1$ be a DMG. Let $\mathcal{J}$ be homogeneous and let $\mathcal{C}$ be the collection of conditioning sets of $\mathcal{J}$. If α and β are C-potential siblings for all $C \in \mathcal{C}$ and e denotes the edge $\alpha \leftrightarrow \beta$, then $\mathcal{I}_{\mathcal{J}}(\mathcal{G}) = \mathcal{I}_{\mathcal{J}}(\mathcal{G} + e)$. If α is a C-potential parent of β for all $C \in \mathcal{C}$ and e denotes the edge $\alpha \to \beta$, then $\mathcal{I}_{\mathcal{J}}(\mathcal{G}) = \mathcal{I}_{\mathcal{J}}(\mathcal{G} + e)$.*

Proof. The inclusion $\mathcal{I}_{\mathcal{J}}(\mathcal{G} + e) \subseteq \mathcal{I}_{\mathcal{J}}(\mathcal{G})$ follows from Proposition 4.7. We show the other inclusion by contraposition. Proposition 4.11 implies that it is enough to consider triples of the form (γ, δ, D), $\gamma, \delta \in V$, $D \subseteq V$, $\gamma \notin D$. Assume $(\gamma, \delta, D) \notin \mathcal{I}_{\mathcal{J}}(\mathcal{G} + e)$. If $(\gamma, \delta, D) \notin \mathcal{J}$, then $(\gamma, \delta, D) \notin \mathcal{I}_{\mathcal{J}}(\mathcal{G})$. If instead $(\gamma, \delta, D) \in \mathcal{J}$, then $(\gamma, \delta, D) \notin \mathcal{I}(\mathcal{G} + e)$ and $D \in \mathcal{C}$. In this case, there exist a μ-connecting walk from γ to δ given D in $\mathcal{G} + e$. Nodes α and β are C-potential siblings (or α is a C-potential parent of β) for all $C \in \mathcal{C}$, and therefore also for $D \in \mathcal{C}$. Lemma D.2 (Lemma D.3) gives the result. $\qquad \square$

Lemmas D.2 and D.3 that are used in the above proof are adaptations of lemmas in Mogensen and Hansen (2020). Appendix D describes how to make this generalization.

From an independence model $\mathcal{I}_{\mathcal{J}}(\mathcal{G})$ such that $\mathcal{J}$ is homogeneous we now define a graph on nodes V, $\mathcal{G} = (V, E)$. As $\mathcal{J}$ is homogeneous, we know that $\mathcal{J} = \{(A, B, C) \in \mathcal{P} : C \in \mathcal{C}\}$ for some $\mathcal{C} \subseteq \{C : C \subseteq V\}$. For all $\alpha, \beta \in V$, we include the directed edge $\alpha \to \beta$ if and only if α is a C-potential parent of β for all $C \in \mathcal{C}$. We include the bidirected edge $\alpha \leftrightarrow \beta$ if and only if α and β are C-potential siblings for all $C \in \mathcal{C}$. We denote the resulting graph by $\mathcal{N}$. We see that $\mathcal{N}$ is uniquely defined from the $\mathcal{J}$-independence model of $\mathcal{G}$, $\mathcal{I}_{\mathcal{J}}(\mathcal{G})$, and is therefore the same for all elements of the equivalence class $[\mathcal{G}]_{\mathcal{J}}$. The following shows that $\mathcal{N}$ is a unique maximal element, that is, a greatest element, in $[\mathcal{G}]_{\mathcal{J}}$.

Theorem 5.8. *Let $\mathcal{G}$ be a DMG and let $\mathcal{J}$ be homogeneous. The graph $\mathcal{N}$ defined above is $\mathcal{J}$-weakly equivalent with $\mathcal{G}$ and it is the unique maximal element in $[\mathcal{G}]_{\mathcal{J}}$.*

Proof. Let $\bar{\mathcal{G}} \in [\mathcal{G}]_{\mathcal{J}}$. If a directed edge, $\alpha \to \beta$, is in $\bar{\mathcal{G}}$, then α is a C-potential parent of β in $\mathcal{I}_{\mathcal{J}}(\bar{\mathcal{G}}) = \mathcal{I}_{\mathcal{J}}(\mathcal{G})$ for all C (Proposition 5.6). This means that the directed edge is in $\mathcal{N}$. Similarly, for bidirected edges (Proposition 5.5), and $\mathcal{N}$ is a supergraph of all graphs in $[\mathcal{G}]_{\mathcal{J}}$.

Every edge in $\mathcal{N}$ is a C-potential edge in $[\mathcal{G}]_{\mathcal{J}}$ for all $C \in \mathcal{C}$. We can construct a finite sequence of graphs starting from $\mathcal{G}$ and adding the edges that are in $\mathcal{N}$, but not in $\mathcal{G}$, sequentially. Lemma 5.7 shows that all graphs in this sequence are $\mathcal{J}$-weakly equivalent with $\mathcal{G}$, and therefore so is $\mathcal{N}$.

In conclusion, $\mathcal{N}$ is a greatest element of the equivalence class. $\qquad\square$

Theorem 5.8 is central in our development of graphical modeling based on weak equivalence as it provides a unique and interpretable representative of each equivalence class. We give examples of applications of this theorem in Section 6.

5.2.1 Comparison with Markov equivalence case

The above definitions and results are related to results in the case of Markov equivalence (Mogensen and Hansen, 2020). Definitions 5.1 and 5.2 can be thought of as C-specific versions of Definitions 5.1 and 5.5 in Mogensen and Hansen (2020). This leads to C-specific versions of Propositions 5.5 and 5.6 that are analogous to propositions in Mogensen and Hansen (2020).

Importantly, the potential parent conditions of Mogensen and Hansen (2020) use multiple conditioning sets and are therefore not amenable as a foundation for the proof of Theorem 5.8. The conditions in this paper use a single C which facilitates the generalization from Markov equivalence classes to weak equivalence classes. The reformulation of the definitions also entails an important change of perspective. Instead of describing conditions such that the addition of an edge does not change the independence model for *any* conditioning set (Markov equivalence), the above conditions describe conditions such that the addition of an edge does not change the independence model when restricted to a *specific* conditioning set. This allows us to aggregate these conditions for any set of conditioning sets as defined by a homogeneous $\mathcal{J}$, and from this we can prove the existence of a greatest element in this more general setting.

6 Representation of weak equivalence classes

The previous section proved the existence of a greatest element in each weak equivalence class when $\mathcal{J}$ is homogeneous. In Subsection 6.1, we first desribe how this leads to a simple and concise representation of an entire equivalence class, and Subsection 6.3 illustrates this representation using the alarm example. In Subsection 6.2, we restrict our attention to k-weak equivalence and describe a *hierarchy* of k-weak equivalence classes. Choosing a $k = 0, 1, \ldots, n-1$ leads to different notions of equivalence with different levels of granularity. The hierarchy in Subsection 6.2 provides a graphical representation of k-weak equivalence classes across different values of k which is meant to illuminate how equivalence classes change across different values of k.

6.1 Directed mixed equivalence graph

The following definition provides a graphical object representing an entire weak equivalence class. Mogensen and Hansen (2020) gave the same definition in the context of Markov equivalence as illustrated in Example 2.13.

Definition 6.1 (Directed mixed equivalence graph (DMEG)). *Let $\mathcal{J}$ be homogeneous and assume that $\mathcal{N} = (V, F)$ is $\mathcal{J}$-maximal and $\mathcal{N} \in [\mathcal{G}]_{\mathcal{J}}$. We define $\bar{F} \subseteq F$ such that $e \in \bar{F}$ if and only if $e \in F$ and there exists $\mathcal{G} = (V, E) \in [\mathcal{G}]_{\mathcal{J}}$ such that $e \notin E$. We define the directed mixed weak equivalence graph (DMEG) of $[\mathcal{G}]_{\mathcal{J}}$ as the triple $(V, F, \bar{F})$.*

We visualize a directed mixed weak equivalence graph by drawing the corresponding maximal graph and making all edges in $\bar{F}$ dashed (see the example in Figure 15). A DMEG summarizes the equivalence class in the following sense. Let $\mathcal{N}$ be a $\mathcal{J}$-maximal element such that $\mathcal{N} \in [\mathcal{G}]_{\mathcal{J}}$, that is, $\mathcal{N}$ is the greatest element of $[\mathcal{G}]_{\mathcal{J}}$, and let $\mathcal{N}'$ be the corresponding DMEG. If an edge is solid in $\mathcal{N}'$, then this edge is in every $\mathcal{G}_1 \in [\mathcal{G}]_{\mathcal{J}}$. If an edge is absent in $\mathcal{N}'$, then no $\mathcal{G}_1 \in [\mathcal{G}]_{\mathcal{J}}$ contains this edge. If an edge, e, is dashed in $\mathcal{N}'$, then there exists a $\mathcal{G}_1 = (V, E) \in [\mathcal{G}]_{\mathcal{J}}$ such that $e \notin E$. Clearly e is in $\mathcal{N} \in [\mathcal{G}]_{\mathcal{J}}$ and therefore e is in some elements of $[\mathcal{G}]_{\mathcal{J}}$, but not in others. One should note that removing multiple dashed edges from $\mathcal{N}'$ does not necessarily lead to a $\mathcal{J}$-weakly equivalent graph as removing an edge may impose restrictions on which other edges can be removed while maintaining $\mathcal{J}$-weak equivalence. This is related to the fact that a weak equivalence class need not contain a least element (see Figure 15).
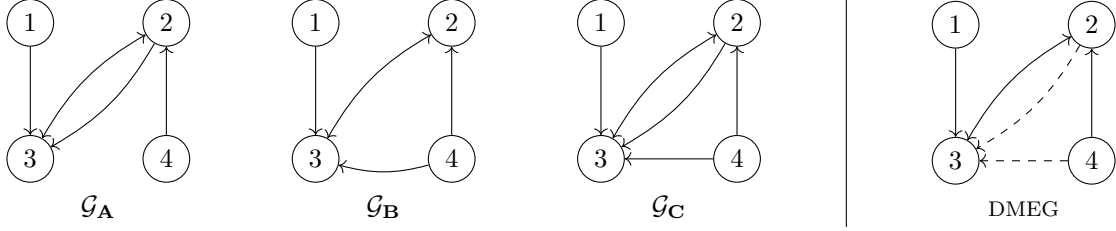

Figure 15: Graphs from Example 6.2. Loops are omitted from the visualization.

Example 6.2 (Directed mixed equivalence graph). *Graphs $\mathcal{G}_{\mathbf{A}}$, $\mathcal{G}_{\mathbf{B}}$, and $\mathcal{G}_{\mathbf{C}}$ in Figure 15 constitute a 2-weak and a 3-weak equivalence class when restricting to DMGs that have all loops present (for simplicity we make this assumption). The graph $\mathcal{G}_{\mathbf{C}}$ is the greatest element. The corresponding DMEG is also shown in Figure 15, see Definition 6.1. The 3-weak equivalence class (2-weak equivalence class) does not contain a least element as removing both $4 \to 3$ and $2 \to 3$ does not lead to a 3-weakly equivalent graph (2-weakly equivalent graph).*

Example 6.3. *This example describes a setting which leads to a weak equivalence with a homogeneous $\mathcal{J}$ which is not a k-weak equivalence. We consider a setting where a 5-dimensional process is observed, $V = \{1, 2, 3, 4, 5\}$, but not every coordinate process is observed simultaneously. This is essentially a setting with overlapping variable sets, see, e.g., Danks (2002); Danks et al. (2008); Triantafillou et al. (2010); Huang et al. (2020). We assume that data contains observations of X_t^R over an interval T_R for $R \in \mathcal{R}$,*

$$\Big\{ \{1, 2, 3, 4\}, \{1, 2, 3, 5\}, \{1, 2, 4, 5\}, \{3, 4, 5\} \Big\}.$$

The intervals are disjoint, $T_{R_1} \cap T_{R_2} = \emptyset$ for $R_1 \neq R_2$. We will approach this problem by restricting the local independences that can be tested using this data and require that there exists $R \in \mathcal{R}$ such that $A, B, C \subseteq R$ for us to be able to test the local independence (A, B, C).

We see that all local independences, (α, β, C), such that $\alpha, \beta \in V$ and $|C| \leq 1$ can be tested from this data as every triple, $\{\alpha, \beta, \gamma\}$, $\alpha, \beta, \gamma \in V$, is observed simultaneously (that is, $\alpha, \beta, \gamma \in R$ for some $R \in \mathcal{R}$). We can also test $(\alpha, \beta, \{1, 2\})$ for all $\alpha, \beta \in \{1, 2, 3, 4, 5\}$, but not $(4, 5, \{1, 3\})$. This means that we can model this using k-weak equivalence, but only for $k = 0$ or $k = 1$. We can obtain further information by defining

$$\mathcal{J} = \Big\{ (\alpha, \beta, C) : \alpha, \beta \in V, |C| \leq 1 \Big\} \bigcup \Big\{ (\alpha, \beta, C) : \alpha, \beta \in V, C = \{1, 2\} \Big\}.$$

This leads to a homogeneous weak equivalence relation which is not a k-weak equivalence.

6.2 Hierarchy of k-weak equivalence

The previous section describes a graph, the *directed mixed equivalence graph*, which can help us understand a single weak equivalence class for a fixed, homogeneous $\mathcal{J}$. In this section, we restrict our attention to k-weak equivalence relations and study a description of k-weak equivalence classes for varying values of k. We consider a fixed node set, V. For each value of k, the k-weak equivalence classes form a partition of the DMGs on node set V, with smaller k corresponding to more coarse partitions. Each weak equivalence class can be represented by its maximal element and there is an interpretable structure between k-weak equivalence classes for different values of k which can help us understand the connection between these different notions of equivalence. This section describes this *hierarchy* of k-weak equivalences.

6.2.1 Levels of granularity

Let $\mathcal{G}$ be a DMG, and let $k_1 < k_2$. Let $\mathcal{N}_1$ denote the greatest element of $[\mathcal{G}]_{k_1}$ and let $\mathcal{N}_2$ denote the greatest element of $[\mathcal{G}]_{k_2}$. We know that $[\mathcal{G}]_{k_2} \subseteq [\mathcal{G}]_{k_1}$ and it follows that $\mathcal{N}_2 \subseteq \mathcal{N}_1$. The graphs $\mathcal{N}_1$ and $\mathcal{N}_2$ are both representatives of $\mathcal{G}$, but at different levels of granularity. The k_2-equivalence class of $\mathcal{G}$ is smaller, thus k_2-weak equivalence is more expressive than k_1-weak equivalence. We may ask what 'approximation error' we make by using k_1-weak equivalence instead of k_2-weak equivalence. Let e be an edge in $\mathcal{N}_1$ which is not in $\mathcal{N}_2$. We know that $\mathcal{G}$ and $\mathcal{G} + e$ are k_1-weakly equivalent, so they can only differ on μ-separations with C such that $|C| > k_1$. The approximation error induced by including e is therefore restricted to 'large' conditioning sets. From a practical point of view, local independence tests with large conditioning sets are expected to perform poorly. This means that the loss of information when testing local independences from finite samples may be small.

6.2.2 Forest representation

We can provide a convenient representation of the k-weak equivalence hierarchy using *trees* and *forests*. A *tree*, $\mathcal{T} = (V_\mathcal{T}, E_\mathcal{T})$, is an undirected graph in which each pair of distinct nodes are connected by exactly one path. A *forest* is the disjoint union of a set of trees. We can construct a *forest* in the following way. For a fixed V, $|V| = n$, and $k \in K = \{0, 1, \ldots, n - 1\}$, we consider the set of k-weak equivalence classes of DMGs on node set V. We let n_k denote the number of such equivalence classes. The i'th k-weak equivalence class, $i = 1, \ldots, n_k$, contains a unique maximal element and we denote this graph by $\mathcal{G}_{k,i}$. We do this for every $k \in \{0, 1, \ldots, n - 1\}$ and define a node set

$$V_{\mathrm{g}} = \bigcup_{k \in K} \{(\mathcal{G}_{k,i}, k) : i = 1, 2, \ldots, n_k\}.$$

Note that we write this as a disjoint union as the same graph may be a maximal element for different k. Therefore, the set V_{g} contains pairs $(\mathcal{G}, k)$ such that $\mathcal{G}$ is k-maximal. For instance, if $\mathcal{G}$ is a maximal element of a k_1-weak equivalence class and of a k_2-weak equivalence class, then $(\mathcal{G}, k_1) \in V_{\mathrm{g}}$ and $(\mathcal{G}, k_2) \in V_{\mathrm{g}}$ and these are different nodes.

We now construct a forest with node set V_{g} in the following way. For each $(\mathcal{G}, k)$ such that $k > 0$, there exist a unique $(k-1)$-maximal graph, $\bar{\mathcal{G}}$, such that $\mathcal{G} \in [\bar{\mathcal{G}}]_{k-1}$, and we join $(\mathcal{G}, k)$ to $(\bar{\mathcal{G}}, k-1)$ by an undirected edge. We call the resulting graph the *weak equivalence hierarchy* over V and denote it by $\mathcal{H}_V$. For $k < n-1$, we will use $\mathrm{up}(\mathcal{G}, k))$ to denote the (nonempty) set of graphs $\bar{\mathcal{G}}$ such that $(\mathcal{G}, k)$ and $(\bar{\mathcal{G}}, \bar{k})$ are adjacent in $\mathcal{H}_V$ and such that $\bar{k} = k + 1$. For $k > 0$, we will use $\mathrm{down}(\mathcal{G}, k)$ to denote the unique graph $\bar{\mathcal{G}}$ such that $(\mathcal{G}, k)$ and $(\bar{\mathcal{G}}, \bar{k})$ are adjacent in $\mathcal{H}_V$ and such that $\bar{k} = k - 1$. Example 6.4 and Figure 16 describe (parts of) the weak hierarchy over $V = \{1, 2, 3, 4\}$.

Properties of $\mathcal{H}_V$ We first argue that $\mathcal{H}_V$ is a forest. The nodes $(\mathcal{G}_{0,i}, 0)$, $i = 1, \ldots, n_0$, must be in different connected components as for each node there is at most a single edge down in the hierarchy. Using induction on k and Corollary 4.9, we see that if $\mathcal{G}_{k,i} \in [\mathcal{G}_{0,j}]_0$, then there is a path between $(\mathcal{G}_{k,i}, k)$ and $(\mathcal{G}_{0,j}, 0)$, and $V_j = \{(\mathcal{G}_{k,i}, k) : \mathcal{G}_{k,i} \in [\mathcal{G}_{0,j}]_0\}$ is therefore a connected subset of V_{g}. It contains exactly $|V_j| - 1$ edges and is thus a tree. This means that $\mathcal{H}_V$ consists of n_0 disjoint trees, each tree rooted at $\mathcal{G}_{0,j}$ for some $j = 1, 2, \ldots, n_0$. Corollary 4.23 characterizes 0-weak equivalence.

When $i_1 \neq i_2$, $[\mathcal{G}]_{k_1, i_1}$ and $[\mathcal{G}]_{k_2, i_2}$ are disjoint when $k_1 = k_2$, but need not be when $k_1 \neq k_2$. For $k_2 \geq k_1$ and $i_1 = 1, \ldots, n_{k_1}$, there exist i_2 such that $\mathcal{G}_{k_1, i_1} = \mathcal{G}_{k_2, i_2}$ which is due to the fact that if a graph is k_1-maximal, then it is also k_2-maximal (Corollary 4.15). The leaves of the trees are the greatest elements of the Markov equivalence classes (Proposition 4.18).

The graph $\mathcal{H}_V$ represents the entire system of k-weak equivalence classes and can be conveniently drawn in levels such that the vertical placement is determined by k (see Figure 16). Let $[\mathcal{G}_{k,i}]_k$ be a k-weak equivalence class represented by its greatest element $\mathcal{G}_{k,i}$. If we move along the unique edge towards a $(k-1)$-maximal graph, we obtain the maximal element of the $(k-1)$-weak equivalence class containing graph the $\mathcal{G}_{k,i}$ by definition of $\mathcal{H}_V$. If we move to the $(k+1)$-level, one of the $(k+1)$-equivalence classes will be represented by $\mathcal{G}_{k,i}$ itself. Naturally, moving towards larger k in the hierarchy achieves smaller equivalence classes as if $\mathcal{G}$ is $(k-1)$-maximal, then $[\mathcal{G}]_{k-1} = \bigcup_{\bar{\mathcal{G}} \in \mathrm{up}(\mathcal{G}, k-1)} [\bar{\mathcal{G}}]_k$.

Dashed edges in the hierarchy   In $\mathcal{H}_V$, one may use DMEGs instead of the corresponding maximal DMGs, and in this paragraph we think of a node $(\mathcal{G}, k)$ in $\mathcal{H}_V$ as a pair consisting of a DMEG and an integer. In this case, there is also a certain structure in the dashed/solid status of edges across levels of k. If an edge $\alpha \sim \beta$ is solid in $(\mathcal{G}, k)$, then it is also solid in all graphs $\bar{\mathcal{G}} \in \mathrm{up}(\mathcal{G}, k)$. This is seen from the fact that if $\tilde{\mathcal{G}} \in [\bar{\mathcal{G}}]_{k+1}$ then $\tilde{\mathcal{G}} \in [\mathcal{G}]_k$ and every graph in this equivalence class contains $\alpha \sim \beta$ which is why all graphs in $[\bar{\mathcal{G}}]_{k+1}$ also contain it. If the edge $\alpha \sim \beta$ is dashed in $(\mathcal{G}, k)$, then it is also dashed in $\bar{\mathcal{G}} = \mathrm{down}(\mathcal{G}, k-1)$. This is because there exists a graph $\tilde{\mathcal{G}} \in [\mathcal{G}]_k$ without this edge, and $\tilde{\mathcal{G}} \in [\bar{\mathcal{G}}]_{k-1}$. On the other hand, the edge is in the maximal element of $[\mathcal{G}]_k$, thus the edge must be present in $\bar{\mathcal{G}}$ and dashed.

On the other hand, moving up (towards larger values of k) in the hierarchy a dashed edge may be removed, become solid, or remain dashed. Moving down (towards smaller values of k) in the hierarchy a solid edge may become dashed.

Example 6.4 (k-weak hierarchy over $V = \{1, 2, 3, 4\}$). *Figure 16 shows a subgraph of $\mathcal{H}_V$ for $V = \{1, 2, 3, 4\}$. A node in $\mathcal{H}_V$, $(\mathcal{G}, k)$, is shown as $\mathcal{G}$ (or rather, the corresponding DMEG), and k determines the vertical placement of the node. All loops are present in the maximal graphs, but omitted from the visualization for simplicity. We use the edge $\alpha \leftarrow\!\!\!\bullet\!\!\!\rightarrow \beta$ to indicate that all three possible edges between a pair of nodes, α and β, are present in the graph, that is, $\alpha \rightarrow \beta, \alpha \leftrightarrow \beta, \alpha \leftrightarrow \beta$. The letters (x, y), to the right of a graph index the graphs shown in the figure.*

Figure 16 shows two subtrees of trees in the hierarchy. We see that the two graphs shown on level $k = 0$, (a, a) and (e, a), are not 0-weak equivalent as there is no directed trek from 1 to 2 in (e, a) (see also Corollary 4.23).

In the figure, a red undirected edge indicates graph equality, for example, the edge between (a, c) and (a, d). As noted above, if $\mathcal{G}$ is k-maximal, then $\mathcal{G} \in \text{up}(\mathcal{G}, k)$ and when $\mathcal{G}$ is drawn both in levels k and $k + 1$, we indicate this by making the undirected edge connecting them red.

6.3 Alarm network

We return to the alarm example from Subsection 2.1. This is a network of moderate size with 10 observable coordinate processes. If we consider graphical modeling of this network using a k-weak equivalence relation, different values of $k \in \{0, 1, \ldots, 10\}$ lead to different levels of granularity as larger values of k will give us smaller equivalence classes. Let $\mathcal{G}$ denote the latent projection of the system (see Figure 1), and let $\mathcal{N}_k$ denote the greatest element of $[\mathcal{G}]_k$. Figure 17 shows the DMEGs of $\mathcal{N}_{10}$ and of $\mathcal{N}_3$. We know that $\mathcal{N}_{10} \subseteq \mathcal{N}_3$. In this example, we see that the only difference between the two DMEGs in Figure 17 is the bidirected edge between 3 and 10. This edge is necessarily dashed as $\mathcal{N}_{10} \in [\mathcal{G}]_3 = [\mathcal{N}_3]_3$. The added complexity of using $k = 10$ does therefore not provide much additional information in this example.

7 Algorithms for weak equivalence

The results in Section 3 imply that several computational tasks that occur naturally when using μ-separation and local independence for graphical modeling of stochastic processes are not feasible, even for a moderate number of coordinate processes. Section 4 introduces a more flexible notion of equivalence to circumvent these issues and Section 5 shows that the convenient theory of Markov equivalence classes translates seamlessly to the more general notion of weak equivalence. As a last component of this paper, we argue that this more general theory leads to algorithms that are in fact feasible from a computational point of view.

7.1 A parametrized hierarchy of graphical equivalence

We start this subsection by providing a formal definition of the weak equivalence decision problem.

Decision problem 7.1 (Weak Markov equivalence in DMGs). *Let $\mathcal{G}_1 = (V, E)_1$ and $\mathcal{G}_2 = (V, E_2)$ be DMGs. Are $\mathcal{G}_1$ and $\mathcal{G}_2$ $\mathcal{J}$-weakly equivalent?*
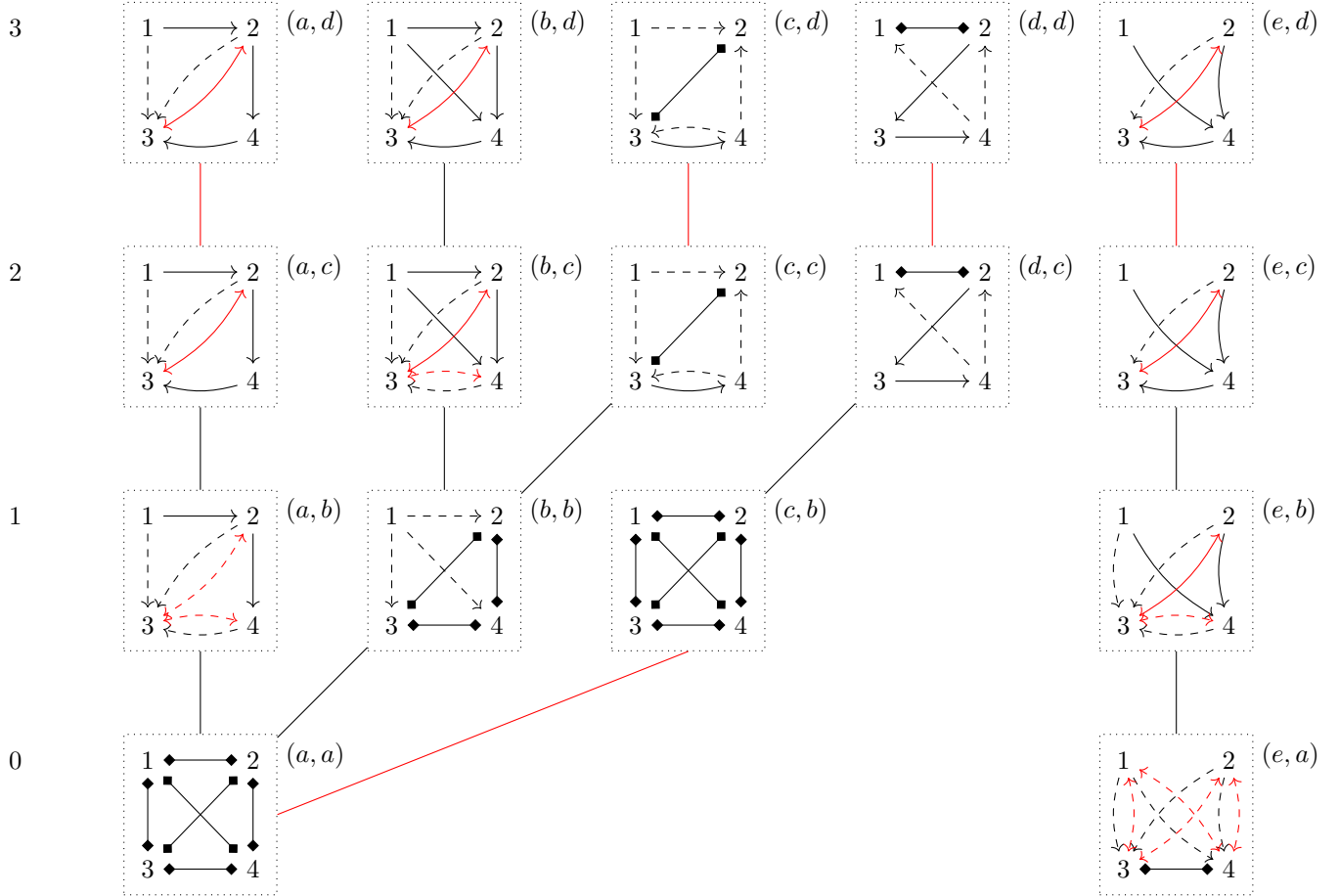
Figure 16: Subtrees from the k-weak equivalence hierarchy on $V = \{1, 2, 3, 4\}$. For simplicity, we use $\alpha \bullet\!\!-\!\!\bullet \beta$ to indicate that $\alpha \to \beta; \alpha \leftarrow \beta; \alpha \leftrightarrow \beta$ are all present in the graph. We use $\alpha \bullet\!\!-\!\!\bullet \beta$ regardless of whether some or all of these underlying edges are dashed. In this figure, bidirected edges are red for better legibility.
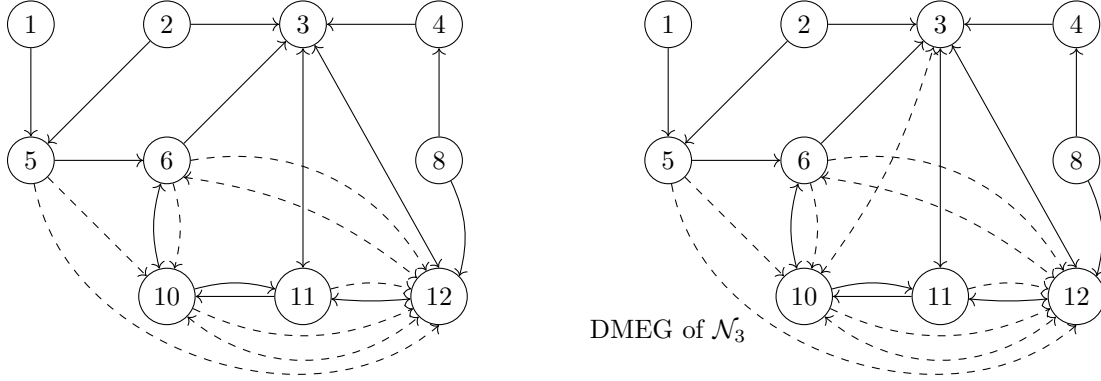
37

Figure 17: Graphs from Subsection 6.3. Loops are omitted from the visualization. Left: *directed mixed equivalence graph* of $\mathcal{N}_{10}$ which is the greatest element of $[\mathcal{G}]_k$ for $k = 4, 5, \ldots, 10$. Right: *directed mixed equivalence graph* of $\mathcal{N}_3$ which is the greatest element of $[\mathcal{G}]_k$ for $k = 2, 3$. The only difference between the two DMEGs is the dashed, bidirected edge betweeen 2 and 3. In $\mathcal{N}_3$, we see that $5 \to 10 \leftrightarrow 3$ is a μ-connecting walk from 5 to 3 given $\{2, 6, 10, 11\}$. In $\mathcal{N}_{10}$, there is no such connecting walk, and this illustrates that $\mathcal{N}_{10}$ and $\mathcal{N}_3$ are not 4-weakly equivalent.

Decision problem 7.1 is coNP-complete as it is a more general problem than Decision problem 3.1. We restrict this to k-weak equivalence and obtain a *parametrized* decision problem.

Decision problem 7.2 (Weak Markov equivalence in DMGs). *Let k be a nonnegative integer, and let $\mathcal{G}_1 = (V, E)_1$ and $\mathcal{G}_2 = (V, E_2)$ be DMGs. Are $\mathcal{G}_1$ and $\mathcal{G}_2$ k-weakly equivalent?*

A decision problem is said to be *slicewise polynomial* if there exists an algorithm which solves the problem in $\mathcal{O}(n^{g(k)})$ steps for a computable function g, input length n, and parameter k. For fixed k, we can decide k-weak equivalence of two DMGs by simply checking every possible triple $(\alpha, \beta, C), \alpha, \beta \in V, C \subseteq V$. This can be done in time bounded by $n^{g(k)}$ as the number of conditioning sets is bounded by n^k. This shows that parametrized k-weak equivalence is a *slicewise polynomial* problem, in that for a fixed k it is solvable by an algorithm which is polynomial in n. One should note that this is different from the m-sparse decision problems (e.g, Decision problem 3.6) as they remain hard for a fixed m whenever $m \geq 16$.

Intuitively, the unrestricted Markov equivalence problem is computationally hard as the maximal size of the conditioning sets also grows with n. On the other hand, if we consider k-weak equivalence for a fixed k then the maximal size of the conditioning sets is fixed, and the problem can be solved in time which scales polynomially in n.

7.2 Computing greatest elements and directed mixed equivalence graphs

As explained above, for a fixed k one can decide k-weak equivalence in polynomial time. The same applies to the related computational problems.

Assume we have a graph $\mathcal{G}$ and want to find the maximal element of $[\mathcal{G}]_k$. A simple algorithm checks for each edge if its addition violates any of the independences in $[\mathcal{G}]_k$ and adds the edge if

and only if this is not the case. For a fixed k, this is done in polynomial time.

When considering a weak equivalence class as represented by its greatest element, we are interested in computing the associated directed mixed equivalence graph (DMEG) as this graph represents the entire equivalence class concisely. We may remove a single edge at a time and decide Markov equivalence to obtain the corresponding DMEG from a greatest element.

8 Learning

There is a large literature on methods for recovering a graph from observational data Spirtes and Zhang (2018). In the case of DAG-based models, many methods use tests of conditional independence. Similarly, it is possible to learn local independence graphs using tests of local independence. In this section, we briefly discuss graphical structure learning based on tests of local independence as described by Meek (2014) and its connection to weak equivalence of DMGs. Mogensen et al. (2018) described a learning algorithm outputting the Markov equivalence DMEG from tests of local independence. Absar and Zhang (2021) implemented a PC-like algorithm based on μ-separation. Bhattacharjya et al. (2022) studied independence tests in proximal graphical event models and graphical structure learning based on tests of local independences. Other work described tests of local independence (Thams and Hansen (2021) and Christgau et al. (2022)) and good tests are of course a prerequisite for constraint-based structure learning. The learning problem has also been studied in the discrete-time processes (Eichler, 2013).

As argued in previous sections, constrained-based algorithms that learn the Markov equivalence class of a partially observed local independence graph and are correct in the oracle setting scale poorly with the size of the graph. Therefore, k-weak equivalence classes may constitute more reasonable targets for graphical structure learning. The oracle learning algorithm in Mogensen et al. (2018) leveraged the potential sibling and potential parent criteria to ensure correctness, though the number of these conditions also scales poorly with graph size, n. This naturally leads to the idea of using C-potential sibling and C-potential parent criteria directly for learning. In the oracle case this leads to a straightforward learning algorithm by starting from the complete DMG. For each pair of nodes, (α, β), one may test the C-potential parent criteria for all $|C| \leq k$. If one of these criteria is violated, one simply removes $\alpha \to \beta$, and similarly for the bidirected edges. For fixed k, this leads to a polynomial-time oracle learning algorithm which outputs the maximal k-weakly equivalent graph of the true graph. This is similar to early stopping in FCI (Spirtes, 2001) as it only uses tests with small conditioning sets C. While smaller values of k lead to less informative output (larger equivalence classes), the interpretation of a learned DMEG remains the same as when using $k = n$ as shown by the theory in previous sections.

Outside of the oracle setting, actual tests of local independence output a p-value. When learning local independence graphs, one may compute p-values from the local independence tests that comprise the C-potential parent/sibling criteria, $|C|$, and use these p-values to output a maximal graph which is in minimum violation with the data, see e.g. Hyttinen et al. (2014) for a similar idea in DAG-based graphical structure learning.

9 Discussion

The results in Section 3 show that deciding Markov equivalence is computationally hard, even under a sparsity constraint. This also implies that finding the unique maximal element of a Markov

equivalence class is hard and that constraint-based learning algorithms that are correct in oracle versions need exponentially many tests in the worst case.

The theory developed in this paper provides a new interpretation of μ-separation in directed mixed graphs as representations of local independence in partially observed stochastic processes. This leads to equivalence relations on directed mixed graphs that are weaker than Markov equivalence. Under a weak equivalence relation, each equivalence class of directed mixed graphs have a simple representation and interpretation using the existence of a greatest element. Importantly, they retain a clear interpretation and a convenient graphical representation of an entire k-weak equivalence class is available, just as in the case of Markov equivalence classes. The greatest element of an equivalence class also provides a feasible learning target, and one can give a constructive characterization of this element (the collection of C-potential sibling and C-potential parent conditions). The Markov equivalence class is often the learning target when trying to recover a graph from observational data, however, the complexity results in this paper imply that this target may be too expressive. The previous sections give the theoretical underpinning for feasible learning algorithms that output graphs that are less expressive than the Markov equivalence class.

A subset of the weak equivalence relations, k-weak equivalence relations, are naturally parametrized by a natural number k. Varying k, one obtains more or less granular graphical modeling, and a simple hierarchy of equivalence classes can be described across k. The parameter k specifies both the granularity of the equivalence class and the complexity of, e.g., finding a maximal graph. The work in this paper mostly focused on the k-weak equivalence, however, the central results hold more general weak equivalences, and one may find applications of other types of equivalence relations, e.g., with inspiration from specific applications.

10   Acknowledgments

References

Odd O. Aalen. Dynamic modelling and causality. *Scandinavian Actuarial Journal*, 1987(3-4): 177–190, 1987.

Saima Absar and Lu Zhang. Discovering time-invariant causal structure from temporal data. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, pages 2807–2811, 2021.

R. Ayesha Ali, Thomas S. Richardson, and Peter Spirtes. Markov equivalence for ancestral graphs. *The Annals of Statistics*, 37(5B):2808–2837, 2009.

Steen A Andersson, David Madigan, and Michael D Perlman. A characterization of Markov equivalence classes for acyclic digraphs. *The Annals of Statistics*, 25(2):505–541, 1997a.

Steen A Andersson, David Madigan, and Michael D Perlman. On the Markov equivalence of chain graphs, undirected graphs, and acyclic digraphs. *Scandinavian Journal of Statistics*, 24(1):81–102, 1997b.

Steen A Andersson, David Madigan, and Michael D Perlman. Alternative Markov properties for chain graphs. *Scandinavian journal of statistics*, 28(1):33–85, 2001.

Debarun Bhattacharjya, Karthikeyan Shanmugam, Tian Gao, and Dharmashankar Subramanian. Process independence testing in proximal graphical event models. In *Conference on Causal Learning and Reasoning*, pages 144–161. PMLR, 2022.

Alexander Mangulad Christgau, Lasse Petersen, and Niels Richard Hansen. Nonparametric conditional local independence testing. *arXiv preprint arXiv:2203.13559*, 2022.

Tom Claassen, Joris Mooij, and Tom Heskes. Learning sparse causal models is not NP-hard. In *Proceedings of the 29th Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 172–181, 2013.

David Danks. Learning the causal structure of overlapping variable sets. In *Discovery Science: 5th International Conference, DS 2002 Lübeck, Germany, November 24–26, 2002 Proceedings 5*, pages 178–191. Springer, 2002.

David Danks, Clark Glymour, and Robert Tillman. Integrating locally learned causal structures with overlapping variables. *Advances in Neural Information Processing Systems*, 21, 2008.

Vanessa Didelez. *Graphical Models for Event History Analysis based on Local Independence*. PhD thesis, Universität Dortmund, 2000.

Vanessa Didelez. Asymmetric separation for local independence graphs. In *Proceedings of the 22nd Conference on Uncertainty in Artificial Intelligence (UAI)*, 2006.

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*, 137:334–353, 2007.

Michael Eichler. Graphical modelling of multivariate time series. *Probability Theory and Related Fields*, 153(1):233–268, 2012.

Michael Eichler. Causal inference with multiple time series: Principles and problems. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371 (1997):20110613, 2013.

Michael Eichler and Vanessa Didelez. Causal reasoning in graphical time series models. In *Proceedings of the 23rd Conference on Uncertainty in Artificial Intelligence (UAI)*, 2007.

Michael Eichler and Vanessa Didelez. On Granger causality and the effect of interventions in time series. *Lifetime Data Analysis*, 16(1):3–32, 2010.

Rina Foygel, Jan Draisma, and Mathias Drton. Half-trek criterion for generic identifiability of linear structural equation models. *The Annals of Statistics*, 40(3):1682–1713, 2012.

Morten Frydenberg. The chain graph Markov property. *Scandinavian Journal of Statistics*, 17(4): 333–353, 1990.

Michael R. Garey and David S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. San Francisco: freeman, 1979.

Zhongyi Hu and Robin Evans. Faster algorithms for Markov equivalence. In *Conference on Uncertainty in Artificial Intelligence*, pages 739–748. PMLR, 2020.

Biwei Huang, Kun Zhang, Mingming Gong, and Clark Glymour. Causal discovery from multiple data sets with non-identical variable sets. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 10153–10161, 2020.

A. Hyttinen, F. Eberhardt, and M. Järvisalo. Constraint-based causal discovery: Conflict resolution with answer set programming. In *Proceedings of the 30th Conference on Uncertainty in Artificial Intelligence (UAI)*, 2014.

Jan T.A. Koster. On the validity of the Markov interpretation of path diagrams of Gaussian structural equations systems with correlated errors. *Scandinavian Journal of Statistics*, 26:413–431, 1999.

Steffen Lauritzen. *Graphical Models*. Oxford: Clarendon Press, 1996.

Steffen Lauritzen and Kayvan Sadeghi. Unifying Markov properties for graphical models. *Annals of Statistics*, 46(5):2251–2278, 2018.

Marloes Maathuis, Mathias Drton, Steffen Lauritzen, and Martin Wainwright, editors. *Handbook of graphical models*. Boca Raton, FL: CRC Press, 2019.

Christopher Meek. Toward learning graphical and causal process models. In *CI at UAI*, pages 43–48, 2014.

Søren Wengel Mogensen. Causal screening in dynamical systems. In *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI)*, volume 124, pages 310–319. PMLR, 2020a.

Søren Wengel Mogensen. *Graphical modeling in dynamical systems*. PhD thesis, University of Copenhagen, 2020b.

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.

Søren Wengel Mogensen, Daniel Malinsky, and Niels Richard Hansen. Causal learning for partially observed stochastic dynamical systems. In *Proceedings of the 34th Conference on Uncertainty in Artificial Intelligence (UAI)*, 2018.

Judea Pearl. *Causality*. Cambridge University Press, 2009.

Thomas S. Richardson. A characterization of Markov equivalence for directed cyclic graphs. *International Journal of Approximate Reasoning*, 17:107–162, 1997.

Thomas S. Richardson. Markov properties for acyclic directed mixed graphs. *Scandinavian Journal of Statistics*, 30(1):145–157, 2003.

Thomas S. Richardson and Peter Spirtes. Ancestral graph markov models. *The Annals of Statistics*, 30(4):962–1030, 2002.

Thomas S Richardson, Robin J Evans, James M Robins, and Ilya Shpitser. Nested Markov properties for acyclic directed mixed graphs. *arXiv:1701.06686*, 2017.

Kayvan Sadeghi. Stable mixed graphs. *Bernoulli*, 19(5B):2330–2358, 2013.

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

P. Spirtes and T. Verma. Equivalence of causal models with latent variables. Technical Report CMU-PHIL-33, Carnegie Mellon University, 1992.

Peter Spirtes. An anytime algorithm for causal inference. In *Proceedings of the Eighth International Workshop on Artificial Intelligence and Statistics*, volume R3, pages 278–285. PMLR, 2001.

Peter Spirtes and Kun Zhang. Search for causal models. In *Handbook of Graphical Models*, pages 439–470. CRC Press, 2018.

Nikolaj Thams and Niels Richard Hansen. Local independence testing for point processes. *arXiv preprint arXiv:2110.12709*, 2021.

Sofia Triantafillou, Ioannis Tsamardinos, and Ioannis Tollis. Learning causal structure from overlapping variable sets. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pages 860–867. JMLR Workshop and Conference Proceedings, 2010.

Thomas Verma and Judea Pearl. Equivalence and synthesis of causal models. Technical Report R-150, University of California, Los Angeles, 1990a.

Tom S. Verma and Judea Pearl. Equivalence and synthesis of causal models. In *Proceedings of the Sixth Conference on Uncertainty in Artificial Intelligence*, 1990b.

Jiji Zhang. A characterization of Markov equivalence classes for directed acyclic graphs with latent variables. In *Proceedings of the Twenty-Third Conference on Uncertainty in Artificial Intelligence (UAI)*, 2007.

Hui Zhao, Zhongguo Zheng, and Baijun Liu. On the Markov equivalence of maximal ancestral graphs. *Science in China, Series A: Mathematics*, 48(4):548–562, 2005.

A   Decision problems

We list the formal decision problems used in Section 3.

Decision problem A.1 (Add-1 bidirected Markov equivalence in DMGs)**.** *Let $\mathcal{G}_1 = (V, E_1)$ and $\mathcal{G}_2 = (V, E_2)$ be DMGs such that $E_2 = E_1 \cup \{e\}$ and e is bidirected edge. Are $\mathcal{G}_1$ and $\mathcal{G}_2$ Markov equivalent?*

Decision problem A.2 (Add-1 directed Markov equivalence in DMGs)**.** *Let $\mathcal{G}_1 = (V, E_1)$ and $\mathcal{G}_2 = (V, E_2)$ be DMGs such that $E_2 = E_1 \cup \{e\}$ and e is directed edge. Are $\mathcal{G}_1$ and $\mathcal{G}_2$ Markov equivalent?*

The next decision problems are sparse versions of Decision problems A.1 and A.2.

Decision problem A.3 (Add-1 birected Markov equivalence in sparse DMGs)**.** *Let m be a non-negative integer and let $\mathcal{G}_1 = (V, E_1)$ and $\mathcal{G}_2 = (V, E_2)$ be m-sparse DMGs such that $E_2 = E_1 \cup \{e\}$ and e is bidirected edge. Are $\mathcal{G}_1$ and $\mathcal{G}_2$ Markov equivalent?*

Decision problem A.4 (Add-1 directed Markov equivalence in sparse DMGs)**.** *Let m be a non-negative integer and let $\mathcal{G}_1 = (V, E_1)$ and $\mathcal{G}_2 = (V, E_2)$ be m-sparse DMGs such that $E_2 = E_1 \cup \{e\}$ and e is directed edge. Are $\mathcal{G}_1$ and $\mathcal{G}_2$ Markov equivalent?*

B   Node connectivity in DMGs

In this section, we elaborate on the discussion in Subsection 3.1 on different notions of node connectivity in a DMG. For a DMG, $\mathcal{G} = (V, E)$ and a node $\beta \in V$, we define β's *indegree*, $\mathrm{in}_\mathcal{G}(\beta)$, to be number of nodes, $\alpha \in V$, such that $\alpha \ast\!\!\to \beta$. Similarly, we define β's *outdegree*, $\mathrm{out}_\mathcal{G}(\beta)$, as the number of nodes, $\alpha \in V$, such that $\beta \ast\!\!\to \alpha$. This is an adaptation of the common definitions of in- and outdegree in DAGs. If $\alpha \ast\!\!\to \beta$ in $\mathcal{G}$, then $\alpha \in u(\beta, \mathcal{I}(\mathcal{G}))$, and it follows that the indegree of β is less than or equal to $\mathrm{con}^{\to}_{\mathcal{G}_1}(\beta)$. Similarly, the outdegree of β is less than or equal to $\mathrm{con}^{\leftarrow}_{\mathcal{G}_1}(\beta)$. It holds that $\sum_{\beta \in V} \mathrm{con}^{\to}_{\mathcal{G}_1}(\beta) = \sum_{\beta \in V} \mathrm{con}^{\leftarrow}_{\mathcal{G}_1}(\beta)$. However, as illustrated in Figure 18 it is possible for $\mathrm{con}^{\to}_{\mathcal{G}_1}(\beta)$ for some β to be large while $\mathrm{con}^{\leftarrow}_{\mathcal{G}_1}(\alpha)$ is small for all $\alpha \in V$.

Figure 18: Graphs $\mathcal{G}_n^1$ and $\mathcal{G}_n^2$. In $\mathcal{G}_n^1$ (left), $\mathrm{con}^{\leftarrow}_{\mathcal{G}_n^1}(1) = 5$ while $\mathrm{con}^{\to}_{\mathcal{G}_n^1}(\alpha) \leq 2$ for all nodes α. Similarly, in $\mathcal{G}_n^2$ above (right) $\mathrm{con}^{\to}_{\mathcal{G}_n^2}(1) = 5$ while $\mathrm{con}^{\leftarrow}_{\mathcal{G}_n^2}(\alpha) \leq 2$ for all nodes α. Therefore, we use a maximum over both measures of node connectivity in Section 3.1.

The indegree and outdegree of a node β need not equal the $\mathrm{con}^{\to}_{\mathcal{G}_1}(\beta)$ and $\mathrm{con}^{\leftarrow}_{\mathcal{G}_1}(\beta)$, respectively (see the example in Figure 19). Moreover, the indegree and outdegree need not be the same for Markov equivalent graphs (Figure 19).

The example in Figure 7 is exploiting non-maximality of the graph. In each Markov equivalence class, $[\mathcal{G}]$, there is a greatest element, $\mathcal{N}$ and one could define sparsity of the nodes in $\mathcal{G}$ by counting adjacencies in the $\mathcal{N}$ which is invariant under Markov equivalence. However, the in- and outdegree of β in $\mathcal{N}$ may still be strictly less than $\mathrm{con}^{\to}_{\mathcal{G}_1}(\beta)$ and $\mathrm{con}^{\leftarrow}_{\mathcal{G}_1}(\beta)$, respectively (Figure 19). In fact,

one can find a family of graphs, $\{\mathcal{G}_n = (V_n, E_n)\}$, and a node $\beta \in V_n$ for all n such that $\text{con}_{\mathcal{G}_1}^{\rightarrow}(\beta)$ is unbounded while the indegree and outdegree are fixed (see the example in Figure 21).

If α is inseparable into β and β is inseparable into α in a maximal DMG, they need not be adjacent (see the example in Figure 20).
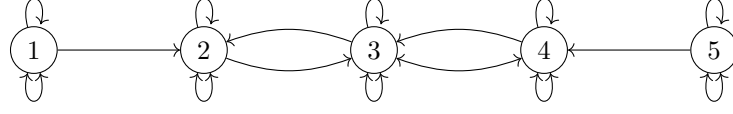

Figure 19: The graph above, $\mathcal{G} = (V, E)$, is Markov equivalent with the graph obtained by adding the edge $5 \rightarrow 3$, $\mathcal{N} = (V, F)$ where $F = E \cup \{5 \rightarrow 3\}$. This shows that in- and outdegrees can be different for two Markov equivalent graphs. The graph $\mathcal{N}$ is the greatest element of $\mathcal{G}$'s Markov equivalence class. We see that 4 cannot be separated from 2 by any subset of $\{1, 3, 4, 5\}$, and therefore $\text{in}_{\mathcal{N}}(4) < \text{con}_{\mathcal{G}_1}^{\rightarrow}(4)$ and $\text{out}_{\mathcal{N}}(2) < \text{con}_{\mathcal{G}_1}^{\leftarrow}(2)$, even though $\mathcal{N}$ is maximal.
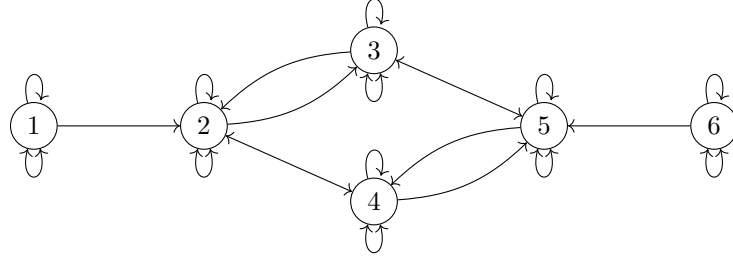

Figure 20: The graph above, $\mathcal{G} = (V, E)$, is maximal, however there is no set $C \subset \{1, 3, 4, 5, 6\}$ such that 5 is μ-separated from 2 by C, and there is also no set $C \subset \{1, 2, 3, 4, 6\}$ such that 2 is μ-separated from 5 by C.
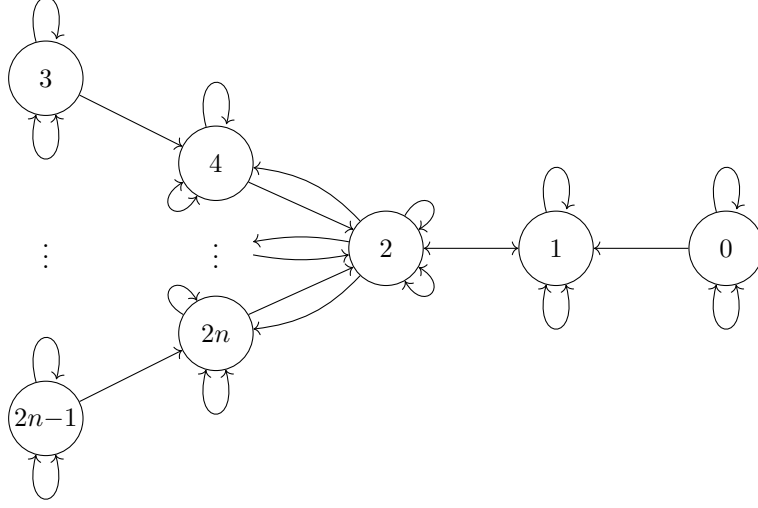
Figure 21: We consider a sequence of graphs, $\mathcal{G}_n$, $n \geq 3$, as illustrated above. The graph $\mathcal{G}_n$ has $2n+1$ nodes, and $\mathcal{G}_n$ is maximal for each n. For every n, the indegree of 1 in the graph $\mathcal{G}_n$ is three. On the other hand, $\mathrm{con}_{\vec{\mathcal{G}_n}}(1)$ equals $n+2$, thus is unbounded in this family of graphs.

C Marginalization

This section argues that the representation of weak equivalence is closed under marginalization in the sense that we can marginalize any graph, $\mathcal{G}$, onto a smaller node set, O, which represents the same independence model as the original graph when restricting independence statements to triples (A, B, C) such that $A, B, C \subseteq O$. This is formalized in Equation 1. A so-called *latent projection* of $\mathcal{G}$ satisfies this requirement. The latent projection was also used in Mogensen and Hansen (2020), and earlier in Verma and Pearl (1990a); Richardson et al. (2017).

Definition C.1 (Latent projection). *We denote the latent projection on $\mathcal{G}$ on O by $m(\mathcal{G}, O)$.*

The latent projection of a graph on a node set represents a marginalized version of the independence model, as formalized by the following corollary. Mogensen and Hansen (2020) proved this result in the case of $\mathcal{J} = \mathcal{P}$, that is, in the case of Markov equivalence (Mogensen and Hansen, 2020, Theorem 3.12). The general case follows directly from the Markov equivalence result.

Corollary C.2. *Let $\mathcal{G}(V, E)$, $O \subseteq V$, and let $\mathcal{M} = m(\mathcal{G}, O)$. For $A, B, C \subseteq O$, it holds that*

$$(A, B, C) \in \mathcal{I}_{\mathcal{J}}(\mathcal{G}) \Leftrightarrow (A, B, C) \in \mathcal{I}_{\mathcal{J}}(\mathcal{M}).$$

Proof. Theorem 3.12 of Mogensen and Hansen (2020) shows that

$$(A, B, C) \in \mathcal{I}(\mathcal{G}) \Leftrightarrow (A, B, C) \in \mathcal{I}(\mathcal{M}),$$

and the result follows immediately. □

Mogensen and Hansen (2020) stated an algorithm to output the latent projection of a DMG (Algorithm 1). This was similar to earlier algorithms in of other classes of graphs Koster (1999); Sadeghi (2013). The following proposition was proved by Mogensen and Hansen (2020).

Proposition C.3 (Mogensen and Hansen (2020)). *Let $\mathcal{G} = (V, E)$ be a DMG and $O \subseteq V$. Algorithm 1 outputs its latent projection, $m(\mathcal{G}, O)$.*

One should note that the marginalization of a (weakly) maximal graph need not be (weakly) maximal as illustrated in Figure 22.

Figure 22: Graphs $\mathcal{G}_1$ (left) and $\mathcal{G}_2$ (right) illustrating that marginalizations of (weakly) maximal graphs need not be (weakly) maximal. $\mathcal{G}_2$ is the marginalization of $\mathcal{G}_1$ over $O = \{1, 2, 3\}$. $\mathcal{G}_1$ is 2-maximal, and therefore maximal for all $k \geq 2$. $\mathcal{G}_2$ is a marginalization, but is not k-maximal for any k, $0 \leq k \leq 3$. Loops are omitted from the visualization.

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

Algorithm 1: Computing the latent projection of a DMG (Mogensen and Hansen, 2020). We consider a DMG, $\mathcal{G} = (V, E)$, and $M \subseteq V$ over which we will marginalize. We let $O = V \setminus M$. A *triroute* is a walk $\alpha \sim \gamma\beta$ such that $\gamma \neq \alpha, \beta$. This is similar to a *tripath* in Lauritzen and Sadeghi (2018), but we allow $\alpha = \beta$. We say that a triroute is *noncolliding* if γ is not a collider on the triroute. We say that heads and tails are *(edge) marks*. We say that two walks between α and β, $(\alpha, e_1^\alpha \ldots, e_1^\beta, \beta)$ and $(\alpha, e_2^\alpha \ldots, e_2^\beta, \beta)$, are *endpoint-identical* if e_1^α and e_2^α have the same mark at α and e_1^β and e_2^β have the same mark at β (note that this may depend on the orientation of directed edges on the walk). We say that a walk between α and β, ω, is endpoint-identical with an edge e between α and β if ω is endpoint-identical with the walk (α, e, β). The set $\Omega_M(\mathcal{G})$ is the set of noncolliding triroutes, $\alpha \sim m \sim \beta$, such that $m \in M$ and such that an endpoint-identical edge $\alpha \sim \beta$ is not in $\mathcal{G}$. We let $(\mathcal{G})_O$ denote the *induced subgraph* on node set O, that is, $(\mathcal{G})_O = (O, E_O)$ where E_O is the subset of E consisting of all edges, e, such that when e is between α and β then both of these edges are in O.

D Proofs and lemmas

The proofs of the following lemmas are adaptations of the proofs of Lemmas 5.4 and 5.8 in Mogensen and Hansen (2020). We include them for completeness to show how the appropriate changes are made. Lemmas 5.4 and 5.8 in Mogensen and Hansen (2020) did not use the C-specific conditions that are essential in obtaining the stronger results that we present in this paper.

Definition D.1 (Route). *We say that a walk, $\omega = (\gamma_1, e_1, \gamma_2, \ldots, e_l, \gamma_{l+1})$, is a* route *if the node γ_{l+1} occurs at most twice on ω and no other node occurs more than once on ω.*

Routes characterize μ-connections in DMGs (Mogensen and Hansen, 2020), and we use them in the next proofs. Note that the below lemmas are formulated using $\mathcal{I}(\mathcal{G})$, not the restricted version $\mathcal{I}_{\mathcal{J}}(\mathcal{G})$.

Lemma D.2. *Let $C \subseteq V$ and let e be a C-potential sibling edge between α and β in $\mathcal{I}(\mathcal{G})$. Let $\gamma, \delta \in V$. If there is a μ-connecting walk from γ to δ given C in $\mathcal{G} + e$, then there is a μ-connecting walk from γ to δ given C in $\mathcal{G}$.*

Proof. Consider any μ-connecting walk from γ to δ given C in $\mathcal{G} + e$. We can also find a μ-connecting route from γ to δ given C in $\mathcal{G} + e$ (Mogensen and Hansen, 2020), and we denote this route by ρ. If $\alpha \notin C$, then there exists a μ-connecting walk from α to β given C in $\mathcal{G}$ using (cs1) of Definition 5.1. If $\beta \notin C$, then there exists a μ-connecting walk from β to α given C in $\mathcal{G}$, also using (cs1). We denote these walks by ν_1 and ν_2, respectively, if they exist.

If e does not occur on ρ, then ρ is μ-connecting given C in $\mathcal{G}$. If e occurs twice, then either ρ contains a subroute $\alpha \leftrightarrow \beta \leftrightarrow \alpha$ and $\delta = \alpha$ or ρ contains a subroute $\beta \leftrightarrow \alpha \leftrightarrow \beta$ and $\delta = \beta$. Assume first the former. There is either a μ-connecting subroute from γ to α, or $\alpha \notin C$. If $\beta \in C$, then consider the subroute between γ and α. This subroute is either trivial or has a tail at α. In either case, composing it with ν_1 gives a μ-connecting walk from γ to β given C in $\mathcal{G}$, and using (cs2) there is also a μ-connecting walk from γ to α given C in $\mathcal{G}$. If $\beta \notin C$, then we can compose the subroute from γ to α with ν_1 and ν_2. The resulting walk will be μ-connecting as $\beta \in \mathrm{an}(C) \setminus C$. The argument is the same when $\beta \leftrightarrow \alpha \leftrightarrow \beta$ and $\delta = \beta$.

We now assume that e occurs only once on ρ and assume first that

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

If $\alpha \notin C$, then we can compose ρ_1, ν_1, and ρ_2 to obtain a μ-connecting walk given C. Note that this also holds if ρ_1 is trivial. If $\alpha \in C$, then ρ_1 is not trivial and it has a head at α. Using (cs3), there exists a μ-connecting walk from γ to β and composing it with ρ_2 gives the result. If instead

$$\gamma \sim \ldots \sim \beta \leftrightarrow \alpha \sim \ldots \ast\!\!\rightarrow \delta,$$

the same arguments work, now using (cs2). $\qquad\square$

Lemma D.3. *Let $C \subseteq V$ and let e be a C-potential parent edge from α to β in $\mathcal{I}(\mathcal{G})$. Let $\gamma, \delta \in V$. If there is a μ-connecting walk from γ to δ given C in $\mathcal{G} + e$, then there is a μ-connecting walk from γ to δ given C in $\mathcal{G}$.*

Proof. We consider a μ-connecting walk from γ to δ given C in $\mathcal{G}+e$. If $\alpha \notin C$, then by (cp1) there exists a μ-connecting walk from α to β given C, and we denote this walk by ν when it exists. We can find a μ-connecting route from γ to δ given C in $\mathcal{G}+e$, and we denote this route by ρ.

In this proof, we will say that a collider on a walk is *newly closed* if the collider is in $\mathrm{an}_{\mathcal{G}+e}(C)$, but not in $\mathrm{an}_{\mathcal{G}}(C)$. If there exists a newly closed collider, then $\alpha \notin C$ and $\beta \in \mathrm{an}_{\mathcal{G}}(C)$. We assume first that e occurs at most once on ρ. If there are newly closed colliders on ρ, the proof of Lemma 5.8 in Mogensen and Hansen (2020) shows that we can find a μ-connecting walk in $\mathcal{G}+e$ with no newly closed colliders such that e occurs at most once, and we denote this walk by $\tilde{\omega}$.

If $\tilde{\omega}$ does not contain e, then the result follows. If it does contain e, we split into two cases. Assume first that

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

We see that $\alpha \notin C$. If ρ_1 is trivial or if it has a tail at α, then composing ρ_1, ν, and ρ_2 gives a μ-connecting walk. If ρ_1 has a head at α, then (cp2) gives a μ-connecting walk from γ to β that we can compose with ρ_2. Assume instead that

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

If ρ_1 has a head at β and $\beta \in C$, then (cp3) gives the result. If $\beta \notin C$, we can find a walk in $\mathcal{G}+e$ with no newly closed colliders and only one occurrence of e of the type

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

where ρ_1 can be trivial, using the same argument as in the proof of Lemma 5.8 in Mogensen and Hansen (2020). We have $\alpha, \beta \notin C$ and there is a μ-connecting walk from α to δ. Using (cp4) there is also one from β to δ. Composing this with ρ_1 gives the result since ρ_1 is either trivial or has a tail at β.

Finally, if e occurs twice on ρ, we must have $\alpha \notin C$. We can use the same arguments as in the proof of Lemma 5.8 in Mogensen and Hansen (2020) using the walk ν and condition (cp2). $\square$

E   Additional results

When we count the number of colliders on a walk, we count them with multiplicity, that is, if

$$(\gamma_1, e_1, \gamma_2, e_2, \ldots, e_{l-1}, \gamma_l, e_l, \gamma_{l+1})$$

is a walk, ω, then the number of colliders on this walk equals the number of i, $2 \leq i \leq l$, such that e_{l-1} and e_i both have heads at γ_i on ω. Note that the endpoints, γ_1 and γ_{l+1} are not colliders by definition. The next lemma is useful for giving a characterization of k-weak equivalence in terms of μ-connecting walks.

49

Lemma E.1. *If there is a μ-connecting walk from α to β given C in $\mathcal{G}$, then there is a μ-connecting walk from α to β given C in $\mathcal{G}$ with at most $|C|$ colliders, all of which are in C.*

Proof. Let $\gamma_1, \gamma_2, \ldots, \gamma_l$ denote the colliders on the μ-connecting walk. We know that $\gamma_i \in \mathrm{an}(C)$ and therefore there exist a directed path $\gamma_i \to \delta_1 \to \ldots \to \delta_{l_i}$ such that $\delta_{l_i} \in C$ and such that $\delta_{l_i} \in C$ is the only node in C on this directed path. If $\gamma_i \in C$, then the path is trivial, that is, contains no edges and just a single node, γ_i. Adding $\gamma_i \to \delta_1 \to \ldots \to \delta_{l_i} \leftarrow \ldots \leftarrow \delta_1 \leftarrow \gamma_i$ for each i creates a walk which is μ-connecting from α to β given C such that every collider is in C. If a node occurs as a collider more than once, we can remove the loop. The resulting walk is also μ-connecting, also if β is a collider, and it has strictly fewer colliders. We can repeat this to find a μ-connecting walk with at most $|C|$ colliders. $\square$

Proposition E.2. *Let $\mathcal{G}$ be a DMG. Let $\alpha, \beta \in V$ and $C \subseteq V$ such that $|C| \leq k$. We have $(\alpha, \beta, C) \in \mathcal{I}_k(\mathcal{G})$ if and only if there is no μ-connecting walk from α to β given C in $\mathcal{G}$ with at most k colliders.*

Proof. If there is a μ-connecting walk given C, then clearly $(\alpha, \beta, C) \notin \mathcal{I}_k(\mathcal{G})$. On the other hand, if $(\alpha, \beta, C) \notin \mathcal{I}_k(\mathcal{G})$ then there is a μ-connecting walk from α to β given C and Lemma E.1 gives the result. $\square$

This means that the restriction of the independence models to k-weak equivalence ignores μ-connecting walks with more than k colliders.

Corollary E.3. *Graphs $\mathcal{G}_1$ and $\mathcal{G}_2$ are k-weak equivalent if and only if it holds for all $\alpha, \beta \in V$ and $C \subseteq V$ such that $|C| \leq k$ that there is a μ-connecting walk from α to β given C in $\mathcal{G}_1$ with at most k colliders if and only if there is a μ-connecting walk from α to β given C in $\mathcal{G}_2$ with at most k colliders.*

Proof. Assume first that $\mathcal{G}_1 \in [\mathcal{G}_2]_k$, and that ω is a μ-connecting walk from α to β given C, $|C| \leq C$, in $\mathcal{G}_1$ with at most k colliders. Proposition E.2 gives that $(\alpha, \beta, C) \notin \mathcal{I}_k(\mathcal{G}_1)$ and therefore $(\alpha, \beta, C) \notin \mathcal{I}_k(\mathcal{G}_2)$. Using Proposition E.2 again gives the result.

Assume instead that for all α, β, C such that $|C| \leq k$ it holds that there is a μ-connecting walk from α to β given C with at most k colliders in $\mathcal{G}_1$ if and only if there is one in $\mathcal{G}_2$. If $(\alpha, \beta, C) \in \mathcal{I}_k(\mathcal{G}_1)$, $\alpha \notin C$, then there is no μ-connecting walk from α to β given C in $\mathcal{G}_1$ and therefore also no μ-connecting walk with at most k colliders in $\mathcal{G}_2$, and Propositions 4.11 and E.2 give the result. $\square$