

## Research Article

Jonas Wahl\*, Urmi Ninad, and Jakob Runge

# Foundations of causal discovery on groups of variables

<https://doi.org/10.1515/jci-2023-0041>

received June 12, 2023; accepted May 10, 2024

**Abstract:** Discovering causal relationships from observational data is a challenging task that relies on assumptions connecting statistical quantities to graphical or algebraic causal models. In this work, we focus on widely employed assumptions for causal discovery when objects of interest are (multivariate) groups of random variables rather than individual (univariate) random variables, as is the case in a variety of problems in scientific domains such as climate science or neuroscience. If the group level causal models are derived from partitioning a micro-level model into groups, we explore the relationship between micro- and group level causal discovery assumptions. We investigate the conditions under which assumptions like causal faithfulness hold or fail to hold. Our analysis encompasses graphical causal models that contain cycles and bidirected edges. We also discuss grouped time series causal graphs and variants thereof as special cases of our general theoretical framework. Thereby, we aim to provide researchers with a solid theoretical foundation for the development and application of causal discovery methods for variable groups.

**Keywords:** causality, causal discovery, graphical models, Markov property, faithfulness, time series

**MSC 2020:** 62D20

## 1 Introduction

Inferring causal relationships from observational data and estimating their strength is an ubiquitous task in many research domains for which a multitude of tools [1–7] have been developed throughout the last decades. While the underlying assumptions on the data-generating process differ from method to method, the majority of approaches have in common that the causal objects of interest are one-dimensional random variables. However, in some applications, the relevant causal entities can be multivariate groups of variables, such as spatial regions of measurements, or collections of random variables that together describe or approximate a phenomenon of interest, such as the phase and amplitude of an oscillation. For instance, neuroscientists may be interested in causal interactions between brain regions rather than in interactions between individual neurons [8,9], while climate scientists would like to improve their understanding of spatio-temporal climate modes that extend across large regions on the globe [10–12] and interact across long distances. Similarly, economists may want to approximate the economic activity of a given country by a range of different indicators rather than a single univariate index [13].

At present, domain experts typically address such problems by employing the group mean of a variable group as a stand-in for the group as a whole or by means of more elaborate standard dimension reduction

---

\* **Corresponding author: Jonas Wahl**, Institute of Computer Engineering and Microelectronics, TU Berlin, Berlin, Germany; DLR Institute for Data Science, Jena, Germany, e-mail: [wahl@tu-berlin.de](mailto:wahl@tu-berlin.de)

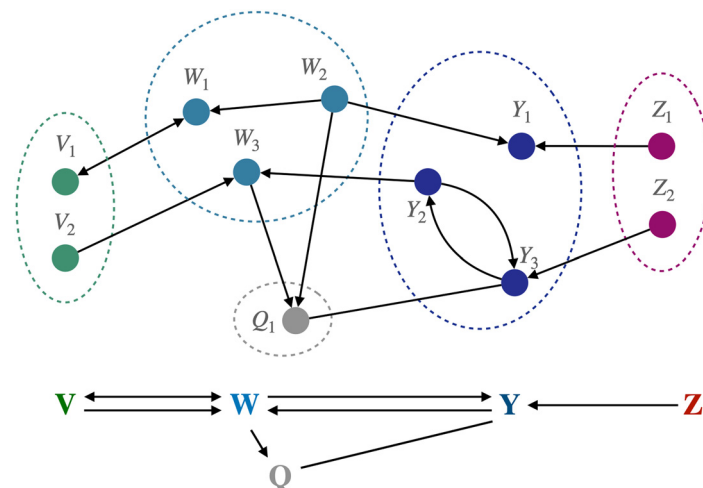
**Urmi Ninad:** Institute of Computer Engineering and Microelectronics, TU Berlin, Berlin, Germany; DLR Institute for Data Science, Jena, Germany, e-mail: [urmi.ninad@tu-berlin.de](mailto:urmi.ninad@tu-berlin.de)

**Jakob Runge:** DLR Institute for Data Science, Jena, Germany; Technische Universität Berlin, Berlin, Germany; Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI) Dresden/Leipzig, TU Dresden, Dresden, Germany, e-mail: [runge@tu-berlin.de](mailto:runge@tu-berlin.de)

techniques such as principal component analysis (PCA). For instance, in climate science, the *El Niño Southern Oscillation (ENSO)* is often represented as either a regional average of sea surface temperatures or as a principal component in a PCA [14]. Unfortunately, if some of the causal processes at hand happen at a smaller scale than averages or principal components can capture, relevant causal information may be lost. As an example, the group mean of two variable groups  $\mathbf{W}$  and  $\mathbf{Y}$  may be conditionally dependent given the group mean of a third group  $\mathbf{Z}$ , while the groups, considered as a whole, satisfy the conditional independence  $\mathbf{W} \perp\!\!\!\perp \mathbf{Y}|\mathbf{Z}$  [2,15]. Causal inference methods based on conditional independence testing such as the PC algorithm might therefore infer different causal structures depending on whether they use group means or the full variable groups as their basic causal objects. Moreover, the dominant mode of *internal* variability of a variable group  $\mathbf{Y}$  as recovered by PCA may not be the causally relevant driver of its effect on another group  $\mathbf{Z}$ , which could for instance be captured more accurately by a higher order principal component. If only the dominant component is consequently used in a causal analysis, then the causal effect of  $\mathbf{Y}$  on  $\mathbf{Z}$  may be diluted or disappear completely. A practical example of this, again from climate science, that deals with the effect of ENSO on the *North Atlantic Oscillation (NAO)* can be found in the study by Zhang et al. [16].

A second approach to causal discovery for variable groups is to run causal discovery algorithms on the totality of all micro-variables and then deduce group level relationships from the inferred micro-graph. Such an approach will inevitably need to unravel micro-relations of little interest to the group level problem at hand. For example, one is typically not interested in causal relations between individual grid locations of satellite measurements of temperature data but between different spatial temperature fields as a whole [7]. In addition, to be sound, a micro-level causal discovery method may require strong technical assumptions on micro-relations that are again of no relevance to the between-group interactions, and it can quickly become computationally inefficient and statistically frail, see, e.g. [17] for empirical evidence of this for two variable groups. We will return to causal discovery with dimension reduction and full microlevel causal discovery in the final section of this article, Section 8, where we will discuss their strengths and weaknesses in more detail.

An alternative approach to the group level causal discovery problem is thus to consider variable groups as a whole as the basic causal entities on which to apply available causal discovery methods [18]. For instance, approaches based on conditional independence testing such as the PC-algorithm do not make any assumptions on the dimensionality of its node variables per se and can still be executed provided that its conditional independence tests are adapted to the multivariate setting [19–22]. However, such constraint-based methods rely on two fundamental assumptions, the *causal Markov property* and *causal faithfulness*, or variants thereof, which now have to be assumed directly on the group level for the methods to be sound. Thus, the following question arises: If causally interacting micro-variables are partitioned into variable groups, see Figure 1, do



**Figure 1:** A mixed graph over micro-variables is coarsened to a mixed graph of variable groups, see Definition 5. In graphical causal modelling, directed edges represent direct causal influences, bidirected edges represent confounding by a hidden variable, and undirected edges indicate the presence of a selection variable that has been conditioned on.

causal discovery assumptions on the micro-level transfer to the group level and if not, what else is required for these group level causal discovery assumptions to be valid?

To answer this question, in this work, we provide a thorough theoretical analysis of the relationship between micro- and macro-level causal models with a view on causal discovery assumptions. We do so for causal models that exhibit cyclic as well as acyclic behaviour. Parallel questions on causal effect estimation on directed acyclic graphs over variable groups have been addressed recently by Anand et al. [23]. Anand et al. [23] also present general rules of graphical calculus for acyclic graphs of groups, which we recall and adapt to our setting in Section 3. In order to discuss our main results, we now recall that the Markov property and causal faithfulness relate the graphical structure of the model, the *causal graph*, or more precisely its *d*- or  $\sigma$ -separations, to the observational distribution of the involved variables.

The Markov property states that two variables that can be separated graphically by a separating set  $S$  are conditionally independent given that set, or for short that *d*- (or  $\sigma$ -)separation implies conditional independence. The assumption of causal faithfulness on the other hand requires that also the converse implication is true, i.e. that conditional independence implies *d*- (or  $\sigma$ -)separation. Taken together, both properties thus state that graphical separations and conditional independencies are in exact correspondence to each other.

While the Markov property is a given in almost every causal inference method, causal faithfulness is more controversial and its validity has been discussed in various places [1,24]. As a consequence, weaker versions of causal faithfulness have been developed, most notably *adjacency* and *orientation faithfulness* [5,25]. We study under which conditions the causal Markov property, faithfulness and some of its relatives do and do not carry over from a fine grained micro-level causal graphical model to a more coarse grained macro-level graph in which the micro-level variables are partitioned into groups, see Figure 1. In order to do so, we additionally study the relationship between micro-level and coarse grained group level causal graphs on a purely graphical level, see Section 3.

As our main results, we show that the Causal Markov property does transfer from the micro- to the group level (Theorems 2 and 3) relatively straightforwardly, but that this is no longer true for causal faithfulness, a fact that was already noted in empirical simulations in the study by Parviainen and Kaski [18]. We point out that in some sense when dealing with variable groups, the faithfulness assumption is more complicated than was already known: not only does faithfulness fail to transfer to the macro-level can even be violated even though its weaker relatives *adjacency* and *orientation faithfulness* [5] are both satisfied on the macro-level, see Section 5 and Figure 9. We are not aware of this type of faithfulness violations (i.e., faithfulness being violated but adjacency and orientation faithfulness holding) in other settings and call them *non-local faithfulness violations*.

On the other hand, we also provide two criteria that *do guarantee* macro-level causal faithfulness whenever the variables are *sufficiently well-connected internally*, either through cycles (Theorem 4) or through directed or bidirected paths (Theorem 5). This may justify the assumption of causal faithfulness in some settings, as often variable groups are chosen the way they are, exactly because of their internal coherence or their strong internal interactions. Nevertheless, considered in entirety, our discussion shows that faithfulness, already controversial in the univariate case, can be a strong assumption for causal graphs over variable groups and practitioners are advised to proceed with care when assuming it.

We also demonstrate that graphs over variable groups need to be interpreted carefully with respect to their causal meaning as we will discuss in Section 7. In addition, we point out that the weaker notion of adjacency faithfulness *does transfer* from the micro-level to macro-level (Lemma 9). Therefore, when developing causal discovery tools for variable groups, proceeding in line with methods such as the conservative PC-algorithm of Ramsey et al. [5], which only rely on adjacency faithfulness, may be advisable if there are no domain-specific reasons to believe that faithfulness is a valid assumption.

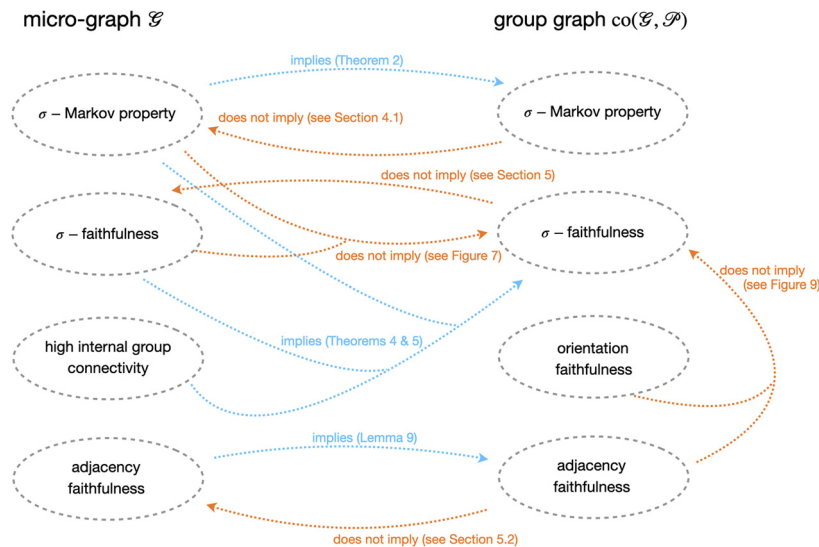
We end with a discussion on causal discovery for time series, and generalize the widely employed notion of the *time series summary graph* [10] to the notion of *time series summary graphs of groups*. We show that, under a dynamical systems inspired condition that we dub *causal mixing*, stronger causal conclusions can be derived from grouped time series summary graphs. Thus, while causal conclusions on the time-resolved level need to be interpreted carefully, global interactions between groups of processes may be more robust with respect to the standard assumptions of causal inference. To summarize, our main contributions are as follows:

- We extend the theoretical framework of Anand et al. [23] for graphical causal reasoning between variable groups to  $\sigma$ -separation and discuss the relationship between micro- and macro-level versions of fundamental graphical properties, such as acyclicity and acyclification (Theorem 1).
- We discuss Markov properties for  $m$ - and  $\sigma$ -separation and show that they transfer from micro- to group level directed mixed graphs (DMGs), see Theorems 2 and 3.
- We discuss different failure modes of causal faithfulness for graphs over variable groups including an example of a non-local faithfulness violation (Section 5).
- We provide two criteria (Theorems 4 and 5) that ensure faithfulness on the group level after coarsening a micro-graph. We also discuss the role of adjacency faithfulness (Lemma 9) and an example addressing the applicability of Meek's orientation rules [26] that was brought forward by Parviainen and Kaski [18] (Subsection 5.3).
- We show how time series causal graphs fit into our framework (Section 6).
- We elaborate on the difference between apparent and true causation in group DMGs and time series group DMGs (Section 7).
- We discuss strengths and failure modes of causal discovery for variable groups through dimension reduction and full microlevel causal discovery and contrast this to an approach that proceeds directly on the group level (Section 8).

We summarize our main results on faithfulness and Markov properties in Figure 2. We hope that this work will provide a solid theoretical footing for the development and empirical validation of group level causal discovery algorithms in the future.

## 1.1 Related work

The compatibility of averaging across variable and causal inference has been discussed by Rubenstein et al. [15], which also provides some toy examples. Arguably, the studies by Parviainen and Kaski, and Anand et al. [18,23] are closest to our work. Parviainen and Kaski [18] discussed several causal discovery methods for



**Figure 2:** A diagrammatic summary of the relationship between the  $\sigma$ -Markov property and different types of faithfulness on a micro-graph  $\mathcal{G}$  and its graph of groups with respect to a partition  $\mathcal{P}$ , see Definition 4 for details. A blue arrow indicates that all properties from which the arrow emerges imply the target property. An orange arrow indicates that the properties from which the arrow emerges are not sufficient to guarantee the target property.

variable groups, introduced the notion of groupwise faithfulness and provide a first analysis of this property, including some empirical experiments with discrete micro-variables. We expand upon the theoretical analysis of Parviainen and Kaski [18] in several directions, e.g. by including cyclic structures, addressing Markov properties as well as by providing new sufficient criteria for groupwise faithfulness, new examples of faithfulness violations and results on time series. Parviainen and Kaski [18] provide an example in which groupwise faithfulness with respect to  $d$ -separation is deemed insufficient to ensure that the Meek orientation rules [26], a fundamental part of the PC-algorithm [2], still hold. However, we will point out in Section 5.3 that this is no longer true if group level cycles in the example of the study by [18] are properly accounted for by replacing  $d$ -faithfulness with  $\sigma$ -faithfulness. Anand et al. [23] present a graphical calculus for  $d$ -separation over graph of groups, which we will adapt to  $\sigma$ -separation below, and use this calculus to discuss causal effect estimation for directed acyclic graphs over variable groups, therein called cluster DAGs. Zscheischler et al. [27] and Wahl et al. [17] present ways of inferring cause–effect relationships when only two groups of variables are involved. Constraint-based causal discovery methods for variable groups require conditional independence testing for multivariate random vectors which are discussed in various places [19–22]. Causal discovery for time series is treated in many works, see e.g. [7,28,29] for discussions on state-of-the-art methods.

## 2 Preliminaries on (directed) mixed graphs

To account for latent confounding and selection bias, many concepts of causal inference have been extended to mixed graphs [30]. Cyclic causal relationships have also been incorporated successfully into causal graphical modelling [31–33] although, for the most part, these works do not deal with undirected edges.

A *mixed graph* (MG) is a tuple  $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{B}, \mathcal{U})$  of a set of nodes  $\mathcal{V}$ , a set of directed edges  $\mathcal{E}$ , a set of bidirected edges  $\mathcal{B}$  and a set of undirected edges  $\mathcal{U}$ . All these sets are assumed to be countable. Directed edges will be depicted by one-sided arrows  $A \rightarrow B$  or  $B \leftarrow A$ , bidirected edges by two-sided arrows  $A \leftrightarrow B$  and undirected edges by simple lines  $A - B$ . We will assume that graphs considered in this work do not admit *self-edges* of any type, i.e. both nodes of an edge are not allowed to coincide. A DMG is a mixed graph without undirected edges, and in this case, we will always suppress the (empty) set  $\mathcal{U}$  from the notation. Finally, a *directed graph* (DG) is a DMG without bidirected edges, and again we will suppress the (empty) set  $\mathcal{B}$  from the notation. A *walk*  $\pi$  from  $A \in \mathcal{V}$  to  $B \in \mathcal{V}$  on a mixed graph  $\mathcal{G}$  is a finite alternating tuple  $\pi = (\pi(1), e_1, \pi(2), e_2, \dots, e_{m-1}, \pi(m))$ ,  $\pi(1) = A$ ,  $\pi(m) = B$  of nodes  $\pi(i) \in \mathcal{V}$  and edges  $e_i \in \mathcal{E} \cup \mathcal{B} \cup \mathcal{U}$  such that  $e_i$  connects  $\pi(i)$  and  $\pi(i+1)$ , i.e.  $e_i \in \{\pi(i) \rightarrow \pi(i+1), \pi(i) \leftarrow \pi(i+1), \pi(i) \leftrightarrow \pi(i+1), \pi(i) - \pi(i+1)\}$ . A *path* is a walk whose nodes  $\pi(1), \dots, \pi(m)$  are all (pairwisely) different. A trivial walk (path) is a walk (path) that consists of only one node and no edges. A walk (path) is called *right-directed* if it is of the form  $\pi(1) \rightarrow \pi(2) \rightarrow \dots \rightarrow \pi(m)$ , *left-directed* if it is of the form  $\pi(1) \leftarrow \pi(2) \leftarrow \dots \leftarrow \pi(m)$  and *directed* if it is left- or right-directed. A *cycle* on  $\mathcal{G}$  is a directed walk  $\pi = (\pi(1), e_1, \pi(2), e_2, \dots, e_{m-1}, \pi(m))$  such that  $\pi(1) = \pi(m)$ , and a graph is said to be *acyclic* if it does not admit any cycles. As is common practice, *directed acyclic graphs* will be abbreviated as *DAGs*. A subset of nodes  $\mathcal{W} \subset \mathcal{V}$  of a mixed graph is *strongly connected* if for any two nodes  $A, B \in \mathcal{W}$  there is a directed path from  $A$  to  $B$ . In particular, there is a cycle between any two nodes in a strongly connected subset. The *strongly connected components* of  $\mathcal{G}$  are the maximal strongly connected subsets of  $\mathcal{V}$ , i.e. those that cannot be enlarged without losing their strong connectivity. For any node  $A$ , the unique strongly connected component that contains  $A$  will be written as  $\text{sc}(A)$ . The strongly connected components of a  $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{B})$  form a partition of  $\mathcal{V}$ , i.e.  $\mathcal{V}$  is a disjoint union of its strongly connected components. We also use the common conventions that  $A \in \mathcal{V}$  is called a *parent* of  $B \in \mathcal{V}$  if there is a directed edge  $A \rightarrow B$ , and an *ancestor* of  $B$  if there is a directed path from  $A$  to  $B$ . Conversely, in the first case,  $B$  is called a *child* of  $A$ , and in the latter case,  $B$  is called a *proper descendant* of  $A$ . A *descendant* of  $A$  is a node that is either  $A$  itself or a proper descendant of  $A$ . A *collider* of a walk  $\pi = (\pi(1), e_1, \pi(2), e_2, \dots, e_{m-1}, \pi(m))$  is an inner node  $\pi(i)$ ,  $1 < i < m$  of  $\pi$  such that both its adjacent edges point into  $\pi(i)$ . Any inner node of  $\pi$  that is not a collider on  $\pi$  is consequently called a *non-collider* of  $\pi$ .



For the purpose of encoding conditional independencies efficiently when modelling causal relationships of random variables graphically, different notions of graphical *separation* have been introduced for different types of graphs.

**Definition 1.** (*m*-separation, see [30]) Let  $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{B}, \mathcal{U})$  be a mixed graph and let  $S \subset \mathcal{V}$  be a set of nodes. A walk  $\pi$  between nodes  $A = \pi(1)$  and  $B = \pi(m)$  is said to be *m-blocked* by  $S$  if one of the following holds:

- (1) its first node  $A$  or its last node  $B$  lie in  $S$ ;
- (2) there is a collider of  $\pi$  that does not have any descendants in  $S$ ;
- (3)  $S$  contains a non-collider of  $\pi$ .

If all walks (or, equivalently, all paths) between  $A$  and  $B$  are *m-blocked* by  $S$ , we say that  $A$  and  $B$  are *m-separated* by  $S$  and write  $A \not\propto_{\mathcal{G}}^m B|S$ . If  $A$  and  $B$  are not *m-separated* by  $S$ , we say that they are *m-connected* by  $S$ .

If the graph  $\mathcal{G}$  is a DAG, *m*-separation is known under the more familiar name *d*-separation. Since *m*-separation can be inadequate to deal with cyclic relationships (see the study by Bongers et al. [33] for a detailed explanation of why this is the case), another type of separation dubbed  $\sigma$ -separation was introduced by Forré and Mooij [31] and studied by Mooij and Claassen and Bongers et al. [32,33]. We have only found the definition of  $\sigma$ -separation for DMGs in the literature, but it is easily adapted to general mixed graphs as well.  $\sigma$ -separation also reduces to the more familiar notion of *d*-separation in the case of directed acyclic graphs.

**Definition 2.** ( $\sigma$ -separation [31]) Let  $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{B}, \mathcal{U})$  be a mixed graph and let  $S$  be a set of nodes. A walk  $\pi$  from  $A = \pi(1)$  and  $B = \pi(m)$  is said to be  $\sigma$ -blocked by a subset  $S \subset \mathcal{V}$  if one of the following holds:

- (1) its first node  $A$  or its last node  $B$  lie in  $S$ ;
- (2) there is a collider of  $\pi$  that does not have any descendants in  $S$ ;
- (3)  $S$  contains a non-collider  $\pi(i)$  that has a neighbor  $\pi(j)$ ,  $j \in \{i-1, i+1\}$  such that
  - $\pi(j) \notin \text{sc}(\pi(i))$  and
  - the edge of  $\pi$  between  $\pi(i)$  and  $\pi(j)$  is of the form  $\pi(i) \rightarrow \pi(j)$  or  $\pi(i) - \pi(j)$ .

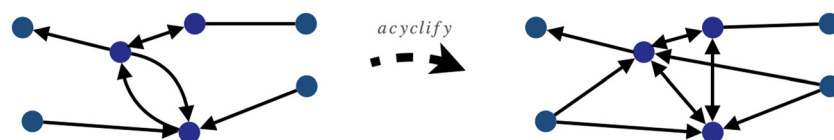
If all walks (or, equivalently, all paths) between  $A$  and  $B$  are  $\sigma$ -blocked by  $S$ , we say that  $A$  and  $B$  are  $\sigma$ -separated by  $S$  and write  $A \not\propto_{\mathcal{G}}^{\sigma} B|S$ . If  $A$  and  $B$  are not  $\sigma$ -separated by  $S$ , we say that they are  $\sigma$ -connected by  $S$ .

A convenient way of linking the usual notion of *d*-separation on DAGs and  $\sigma$ -separation is through *acyclification* [33].

**Definition 3.** (Acyclification of a MG, see [33]) Let  $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{B}, \mathcal{U})$  be a mixed graph. The *acyclification* of  $\mathcal{G}$  is the graph  $\mathcal{G}^{\text{acy}} = (\mathcal{V}, \mathcal{E}^{\text{acy}}, \mathcal{B}^{\text{acy}}, \mathcal{U}^{\text{acy}})$  defined as follows:

- there is a directed edge  $A \rightarrow B \in \mathcal{E}^{\text{acy}}$  if and only if  $A \in \text{pa}_{\mathcal{G}}(\text{sc}_{\mathcal{G}}(B)) \setminus \text{sc}_{\mathcal{G}}(B)$ ;
- there is an undirected edge  $A - B \in \mathcal{U}^{\text{acy}}$  if and only if  $A \notin \text{sc}_{\mathcal{G}}(B)$  and  $A - B \in \mathcal{U}$ ;
- there is a bidirected edge  $A \leftrightarrow B \in \mathcal{B}^{\text{acy}}$  if and only if  $\text{sc}_{\mathcal{G}}(A) = \text{sc}_{\mathcal{G}}(B)$  or there exist  $A' \in \text{sc}_{\mathcal{G}}(A)$ ,  $B' \in \text{sc}_{\mathcal{G}}(B)$  with  $A' \leftrightarrow B' \in \mathcal{B}$ .

The following result is a straightforward generalization of [33, Supplement, Proposition A.19]. It states that  $\sigma$ -separation on a mixed graph can alternatively be understood as *d*-separation on its acyclification (Figure 3).



**Figure 3:** Acyclification of a cyclic mixed graph to an acyclic mixed graph.

**Proposition 1.** Let  $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{B}, \mathcal{U})$  be a mixed graph with acyclification  $\mathcal{G}^{\text{acy}}$ , let  $A, B \in \mathcal{V}$  and let  $S \subset \mathcal{V}$  be a subset of nodes. Then

$$A \bowtie_{\mathcal{G}}^{\sigma} B \mid S \iff A \bowtie_{\mathcal{G}^{\text{acy}}}^d B \mid S.$$

### 3 Group (D)MGs

We will now move to the setting where nodes of graphs are no longer supposed to correspond to scalar random variables but to groups of random variables. If the graphs of groups under investigation are assumed acyclic and directed, they appear in the literature under the name *Group DAGs* [18] or *Cluster DAGs* [23]. We will adopt the former terminology. Even for directed acyclic graphs, many of the following results including those of Sections 4 and 5 are new. Proofs of the results of this section are either provided immediately or have been moved to Appendix B.1.

From now on, we will reserve the bold letter  $\mathbf{X}$  for a given countable set  $\mathbf{X} = \{X_1, X_2, \dots\}$  of *micro-nodes*. Although all results of this section are still purely graphical, we will also sometimes freely refer to the micro-nodes as *micro-variables* as they will correspond to random variables later on. A *partition* of  $\mathbf{X}$  is a set  $\mathcal{P}$  of pairwise disjoint subsets of  $\mathbf{X}$  such that  $\cup_{Y \in \mathcal{P}} Y = \mathbf{X}$ . Partitions will always be assumed finite and its elements will be called *variable groups* and will be denoted by bold letters other than  $\mathbf{X}$ , e.g.  $\mathbf{W}, \mathbf{Y}, \mathbf{Z}$ .

**Definition 4.** Let  $\mathbf{X}$  be the set of micro-nodes, and let  $\mathcal{P}$  be a partition of  $\mathbf{X}$  into finitely many subsets. A *DMG of groups* or *group (D)MG* is a (directed) mixed graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{B}, \mathcal{U})$  whose nodes are the elements of  $\mathcal{P}$ , i.e.  $\mathcal{V} = \mathcal{P}$ . If  $\mathcal{G}$  is an acyclic directed graph, we speak of a *Group DAG*.

To clearly distinguish the usual setting from the group setting, we will speak of a micro (D)MG and a micro DAG if all groups are of size one. There are two natural ways of deriving a group MG: one can (a) coarsen a MG over micro-nodes to a group MG or (b) use a structural causal model over random vectors to induce a group MG directly. The former approach is the main focus of this work, while the latter will be defined and shortly discussed in Section A of the appendix.

#### 3.1 From micro-variable graphs to graphs of groups

If we start out with a mixed graph  $\mathcal{G}$  over (the micro nodes in)  $\mathbf{X}$  and a partition  $\mathcal{P}$  of  $\mathbf{X}$ , there is a straightforward way to obtain a group MG over  $\mathcal{P}$  by “coarsening” the graph  $\mathcal{G}$ . The resulting graph is the quotient of the  $\mathcal{G}$  with respect to  $\mathcal{P}$  (in the category-theoretical sense) and is therefore referred to as the *quotient graph* of  $\mathcal{G}$  (with respect to  $\mathcal{P}$ ) in graph theory [34]. In the context of causal inference, quotient graphs of (bi)directed graphs were first introduced by Anand et al. [23, Definition 1] under the name *cluster DAGs*.

**Definition 5.** (see [23]) Let  $\mathcal{G}$  be a mixed graph over  $\mathbf{X}$ , and let  $\mathcal{P}$  be a partition of  $\mathbf{X}$ . The *coarse graph* or *quotient graph*  $\text{co}(\mathcal{G}, \mathcal{P})$  is the mixed graph with nodes  $\mathbf{Y} \in \mathcal{P}$  obtained by

- (i) drawing a directed edge  $\mathbf{Y} \rightarrow \mathbf{Z}$  if and only if  $\mathbf{Y} \neq \mathbf{Z}$ , and there is a directed edge  $Y \rightarrow Z$  on  $\mathcal{G}$  with  $Y \in \mathbf{Y}$  and  $Z \in \mathbf{Z}$ ;
- (ii) drawing a bidirected edge  $\mathbf{Y} \leftrightarrow \mathbf{Z}$  if and only if  $\mathbf{Y} \neq \mathbf{Z}$ , and there is a bidirected edge  $Y \leftrightarrow Z$  on  $\mathcal{G}$  with  $Y \in \mathbf{Y}$  and  $Z \in \mathbf{Z}$ ;
- (iii) drawing an undirected edge  $\mathbf{Y} - \mathbf{Z}$ , if and only if  $\mathbf{Y} \neq \mathbf{Z}$ , and there is an undirected edge  $Y - Z$  on  $\mathcal{G}$  with  $Y \in \mathbf{Y}$  and  $Z \in \mathbf{Z}$ .

Note that we do not allow self-edges on  $\text{co}(\mathcal{G}, \mathcal{P})$  but that multiple edges, each of a different type, are possible between two nodes of  $\text{co}(\mathcal{G}, \mathcal{P})$ .

Clearly, in this generality, the newly defined graph  $\text{co}(\mathcal{G}, \mathcal{P})$  need not be acyclic even if the underlying micro graph  $\mathcal{G}$  is a DAG. On the other hand, the coarse graph *can* be acyclic even if the micro graph  $\mathcal{G}$  does have cycles, see Figure 4 for illustrations of both of these statements.

To discuss separation on directed graphs of groups, it is useful to introduce *walk (path) segments* and *coarse paths*.

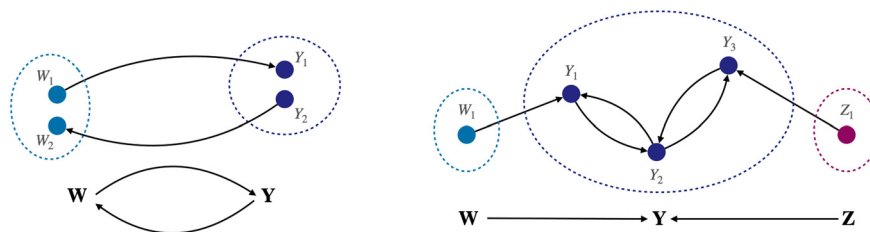
**Definition 6.** Let  $\mathcal{G}$  be a micro MG over  $\mathbf{X}$ , and let  $\mathcal{P}$  be a partition of  $\mathbf{X}$ . Moreover, let  $\pi = (\pi(1), e_1, \pi(2), \dots, e_{m-1}, \pi(m))$  be a walk on  $\mathcal{G}$ . A subwalk  $\pi(i, j) = (\pi(i), e_i, \dots, e_{j-1}, \pi(j))$ ,  $i \leq j$  of  $\pi$  is called a  $\mathcal{P}$ -segment of  $\pi$  if there exists a group  $\mathbf{Y} \in \mathcal{P}$  such that  $\pi(l) \in \mathbf{Y}$  for all  $i \leq l \leq j$  and  $\pi(i-1), \pi(j+1) \notin \mathbf{Y}$ . If  $i = 1$  or  $j = m$ , we only require the respective one-sided condition.

We can thus represent any walk  $\pi = (\pi(1), e_1, \pi(2), \dots, e_{m-1}, \pi(m))$  on a mixed graph as a sequence  $(\pi(i_0, i_1), e_{i_1}, \pi(i_1, i_2), e_{i_2}, \dots, \pi(i_{s-1}, i_s))$ ,  $i_0 = 1$ ,  $i_s = m$  where  $\pi(i_l, i_{l+1})$  are the  $\mathcal{P}$ -segments of  $\pi$  and  $e_{i_l}$  are edges that connect nodes that belong to different groups of  $\mathcal{P}$ . We call this representation the  $\mathcal{P}$ -segment representation of  $\pi$ , see Figure 5.

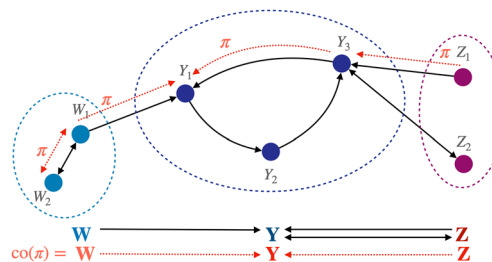
**Definition 7.** Let  $\mathcal{G}$  be a micro MG over  $\mathbf{X}$ , let  $\mathcal{P}$  be a partition of  $\mathbf{X}$  and let  $\mathbf{Y}, \mathbf{Z} \in \mathcal{P}$ . Consider a walk  $\pi$  from  $\mathbf{Y} \in \mathcal{P}$  to  $\mathbf{Z} \in \mathcal{P}$  on  $\mathcal{G}$  with  $\mathcal{P}$ -segment representation  $(\pi(i_0, i_1), e_{i_1}, \pi(i_1, i_2), e_{i_2}, \dots, \pi(i_{s-1}, i_s))$ ,  $i_0 = 1$ ,  $i_s = m$ . The *coarse walk (path)*  $\text{co}(\pi) = (\text{co}(\pi)(1), \tilde{e}_1, \dots, \tilde{e}_{s-1}, \text{co}(\pi)(u))$  of  $\pi$  is the walk on  $\text{co}(\mathcal{G}, \mathcal{P})$  defined as follows:

- (i)  $\text{co}(\pi)(l)$  is the unique  $\mathbf{W} \in \mathcal{P}$  containing the nodes of the  $\mathcal{P}$ -segment  $\pi(i_{l-1}, i_l)$ ;
- (ii)  $\tilde{e}_l$  connects  $\text{co}(\pi)(l)$  and  $\text{co}(\pi)(l+1)$  and is of the same type (directed, bidirected, undirected) as  $e_{i_l}$ .

**Remark.** If  $\pi$  in Definition 7 is a path, then  $\text{co}(\pi)$  need not be a path as well. For instance, if  $\pi$  is of the form  $\pi = W_1 \rightarrow Y_1 \rightarrow W_2 \rightarrow Y_2$  and the micro nodes are grouped as  $\mathbf{W} = \{W_1, W_2\}$ ,  $\mathbf{Y} = \{Y_1, Y_2\}$ , then  $\text{co}(\pi) = \mathbf{W} \rightarrow \mathbf{Y} \rightarrow \mathbf{W} \rightarrow \mathbf{Y}$  is no longer a path. On the other hand, a micro-walk that is not a path can coarsen to a macro-path if micro-node repetitions only happen within  $\mathcal{P}$ -segments. Note also, that if  $\pi$  is a directed walk, then  $\text{co}(\pi)$  is directed as well, see [23, Supplement, Proposition 2].

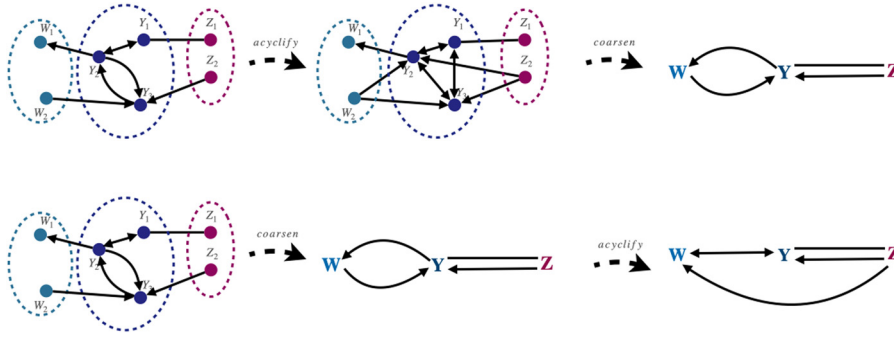


**Figure 4:** Left: A DAG partitioned such that the resulting group DMG is cyclic [23, Figure 1(d)]. Right: A cyclic micro DMG partitioned such that the resulting group DMG is acyclic.



**Figure 5:** The micro path  $\pi$  (in red) from  $W_2$  to  $Z_1$  is coarsened to the path  $\text{co}(\pi)$  in the group DMG  $\text{co}(\mathcal{G}, \mathcal{P})$ ,  $\mathcal{P} = \{\mathbf{W}, \mathbf{Y}, \mathbf{Z}\}$ . The three  $\mathcal{P}$ -segments of  $\pi$  are  $W_2 \leftrightarrow W_1$ ,  $Y_1 \leftrightarrow Y_3$  and  $Z_1$ .





**Figure 6:** Illustration of acyclification and coarsening. In general, these operations do not commute with each other.

**Lemma 1.** Let  $\mathcal{G}$  be a micro MG over  $\mathbf{X}$ , and let  $\mathcal{P}$  be a partition of  $\mathbf{X}$ .

- (i) If  $\text{co}(\mathcal{G}, \mathcal{P})$  is acyclic, then for any strongly connected component  $\mathcal{W}$  of  $\mathcal{G}$ , there is  $\mathbf{Y} \in \mathcal{P}$  such that  $\mathcal{W} \subset \mathbf{Y}$ .
- (ii) The converse of (i) is not true.
- (iii) If the elements of  $\mathcal{P}$  are exactly the strongly connected components of  $\mathcal{G}$ , then  $\text{co}(\mathcal{G}, \mathcal{P})$  is acyclic.<sup>1</sup>

**Definition 8.** We will call a partition  $\mathcal{P}$  of  $\mathbf{X}$

- (i) *acyclic* with respect to the micro MG  $\mathcal{G}$  if the coarse graph  $\text{co}(\mathcal{G}, \mathcal{P})$  is acyclic.
- (ii) *maximally acyclic* if  $\mathcal{P}$  is the partition of  $\mathcal{G}$  into its strongly connected components.

In particular, acyclicity of a partition  $\mathcal{P}$  entails *unidirectionality*, that is, all directed edges between micro nodes  $Y \in \mathbf{Y}$  and  $Z \in \mathbf{Z}$ ,  $\mathbf{Y} \neq \mathbf{Z}$  on the micro graph  $\mathcal{G}$  must point in the same direction, e.g. from the elements of  $\mathbf{Y}$  to the elements of  $\mathbf{Z}$ .

It was pointed out by Anand et al. [23] that coarsening micro DAGs to group DAGs induces an equivalence relation on the set of DAGs over  $\mathbf{X}$ , and this observation carries through when the acyclicity assumption on the micro DAGs is dropped.

**Definition 9.** Given a partition  $\mathcal{P}$ , we will call two micro MGs  $\mathcal{G}$  and  $\mathcal{G}'$   $\mathcal{P}$ -equivalent if their coarse graphs with respect to  $\mathcal{P}$  are the same, i.e. if  $\text{co}(\mathcal{G}, \mathcal{P}) = \text{co}(\mathcal{G}', \mathcal{P})$ .

The two operations of acyclification in the sense of Definition 3 and coarsening in the sense of Definition 5 do not commute in general, see Figure 6. However, if the partition for coarsening is acyclic with respect to the micro MG, then acyclification of the micro MG has no effect on coarsening.

**Theorem 1.** Let  $\mathcal{G}$  be a mixed graph and let  $\mathcal{P}$  be a partition of its nodes. If  $\mathcal{P}$  is acyclic with respect to  $\mathcal{G}$ , then

$$\text{co}(\mathcal{G}^{\text{acy}}, \mathcal{P}) = \text{co}(\mathcal{G}, \mathcal{P}).$$

**Lemma 2.** Let  $\mathcal{G}$  be a mixed graph and let  $\mathcal{P}$  be a partition of its nodes. If there exist  $Y \in \mathbf{Y}$  and  $Z \in \mathbf{Z}$  such that  $\text{sc}_{\mathcal{G}}(Y) = \text{sc}_{\mathcal{G}}(Z)$ , then  $\text{sc}_{\text{co}(\mathcal{G}, \mathcal{P})}(\mathbf{Y}) = \text{sc}_{\text{co}(\mathcal{G}, \mathcal{P})}(\mathbf{Z})$ .

**Proof.** This result directly follows from the following fact: if there is a directed path from  $Y$  to  $Z$  (respectively  $Z$  to  $Y$ ), then the induced coarse path is a directed path from  $\mathbf{Y}$  to  $\mathbf{Z}$  (respectively from  $\mathbf{Z}$  to  $\mathbf{Y}$ ).  $\square$

The following result clarifies the relationship between  $\sigma$ -separation on the micro- and the group level. It generalizes [23, Theorem 1] to  $\sigma$ -separation in DMGs and also demonstrates that said theorem does not generalize to arbitrary mixed graphs in which undirected edges are present.

<sup>1</sup> In this case,  $\text{co}(\mathcal{G}, \mathcal{P})$  is also sometimes referred to as the *condensation* of  $\mathcal{G}$  in the graph theory literature.

**Lemma 3.** Let  $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{B})$  be a DMG, and let  $\mathcal{P}$  be a partition of its nodes. Consider a micro walk  $\pi$  on  $\mathcal{G}$  and denote its induced coarse walk on  $\text{co}(\mathcal{G}, \mathcal{P})$  by  $\text{co}(\pi)$ .

- (i) If  $\text{co}(\pi)$  is  $\sigma$ -blocked by a set  $S \subset \mathcal{P}$  of nodes of  $\text{co}(\mathcal{G}, \mathcal{P})$ , then  $\pi$  is  $\sigma$ -blocked by  $\mathcal{T} = \cup_{W \in S} W$ .
- (ii) The converse of (i) is not true.
- (iii) If  $S \subset \mathcal{P}$  is a set of nodes of  $\text{co}(\mathcal{G}, \mathcal{P})$  that  $\sigma$ -separates  $\mathbf{Y}, \mathbf{Z}$  in  $\tilde{\mathcal{G}}$ , then  $\mathcal{T} = \cup_{W \in S} W$   $\sigma$ -separates any pair of micro nodes  $Y \in \mathbf{Y}, Z \in \mathbf{Z}$  in  $\mathcal{G}$ .
- (iv) (i) and (iii) are no longer true in arbitrary mixed graphs.

We also record the analogue of Lemma 3 for  $m$ -separation for the sake of completeness.

**Lemma 4.** Lemma 3 remains true if  $\sigma$ -separation is replaced by  $m$ -separation.

The proofs of Lemma 3 (i)–(iii) and of Lemma 4 only require straightforward adjustments of the proof of [23, Theorem 1] to  $\sigma$ -separation ( $m$ -separation), and to the fact that we need to deal with walks instead of paths. We include these proofs in Appendix B.1 for the convenience of the reader.

## 4 Markov properties for group (D)MGs

In this section, we will quickly recap the different types of Markov properties that relate  $m$ -separation, respectively  $\sigma$ -separation, to conditional independence statements for scalar node variables. Then we will discuss the transferral of Markov properties from micro graphs to graphs of groups under coarsening. The results of this section are thus no longer purely graphical and micro nodes will always correspond to univariate random variables while nodes of group MGs will consequently always correspond to groups of variables respectively random vectors.

If  $\mathcal{G}$  is a mixed graph over a set of node variables  $\mathbf{X}$  with joint distribution  $P_{\mathbf{X}}$ , then we recall that the pair  $(\mathcal{G}, P_{\mathbf{X}})$  is said to have the  $m$ -Markov property (or to be  $m$ -Markovian) if every valid  $m$ -separation statement on  $\mathcal{G}$  implies the corresponding conditional independence statement, i.e. for  $A, B \in \mathbf{X}$  and  $S \subset \mathbf{X}$

$$A \bowtie_{\mathcal{G}}^m B \mid S \Rightarrow A \perp\!\!\!\perp B \mid S.$$

If the converse implication also holds, i.e.

$$A \perp\!\!\!\perp B \mid S \Rightarrow A \bowtie_{\mathcal{G}}^m B \mid S,$$

then  $(\mathcal{G}, P_{\mathbf{X}})$  is said to be  $m$ -faithful. Similar properties can also be defined for  $\sigma$ - instead of  $m$ -separation:  $(\mathcal{G}, P_{\mathbf{X}})$  is said to have the  $\sigma$ -Markov property (or to be  $\sigma$ -Markovian) if for  $A, B \in \mathbf{X}$  and  $S \subset \mathbf{X}$

$$A \bowtie_{\mathcal{G}}^{\sigma} B \mid S \Rightarrow A \perp\!\!\!\perp B \mid S.$$

and is  $\sigma$ -faithful if the converse implication also holds, i.e.

$$A \perp\!\!\!\perp B \mid S \Rightarrow A \bowtie_{\mathcal{G}}^{\sigma} B \mid S.$$

To introduce analogous properties for mixed graphs of groups, the first observation is that there are now two possible notions of conditional independence that can be considered: pairwise conditional independence and mutual conditional independence. For convenience, we will assume that all distributions have positive densities.

**Definition 10.** (Mutual and pairwise independence) Two groups of random variables  $\mathbf{Y} = \{Y_1, Y_2, \dots\}$  and  $\mathbf{Z} = \{Z_1, Z_2, \dots\}$  are called

- (i) *mutually conditionally independent* given a third group  $\mathbf{W} = \{W_1, W_2, \dots\}$  (written  $\mathbf{Y} \perp\!\!\!\perp \mathbf{Z} \mid \mathbf{W}$ ) if their joint conditional density almost surely factorizes as  $p(\mathbf{y}, \mathbf{z} \mid \mathbf{w}) = p(\mathbf{y} \mid \mathbf{w})p(\mathbf{z} \mid \mathbf{w})$ ;
- (ii) *pairwise (conditionally) independent* given a third group  $\mathbf{W} = \{W_1, W_2, \dots\}$  (written  $\mathbf{Y} \perp\!\!\!\perp^{pw} \mathbf{Z} \mid \mathbf{W}$ ) if for all  $Y \in \mathbf{Y}$  and all  $Z \in \mathbf{Z}$ , we have  $Y \perp\!\!\!\perp Z \mid \mathbf{W}$ .

**Lemma 5.** Consider groups of random variables  $\mathbf{Y}, \mathbf{Z}$ , and  $\mathbf{W}$  and let  $\mathbf{Z}' \subset \mathbf{Z}$  be a non-empty subset. The following are equivalent:

- (i)  $\mathbf{Y}$  and  $\mathbf{Z}$  are mutually conditionally independent given  $\mathbf{W}$ .
- (ii) We have  $\mathbf{Y} \perp\!\!\!\perp \mathbf{Z}' \mid \mathbf{W}$  and  $\mathbf{Y} \perp\!\!\!\perp \mathbf{Z} \setminus \mathbf{Z}' \mid \mathbf{Z}', \mathbf{W}$ .

The situation is more convenient in graphical models in which the  $\sigma$ -Markov property and  $\sigma$ -faithfulness hold on the micro-level. In this case, mutual and pairwise conditional independence turn out to be the same in the sense of the following lemma.

### 4.1 $\sigma$ -Markov properties

**Definition 11.** ( $\sigma$ -Markov properties) Let  $\mathbf{X}$  be a set of scalar random variables with joint distribution  $P_{\mathbf{X}}$  and let  $\mathcal{P}$  be a partition of  $\mathbf{X}$ . Let  $\mathcal{G}'$  be a mixed graph with node set  $\mathcal{P}$ . We say that  $(\mathcal{G}', P_{\mathbf{X}})$  has the

(i)  $\sigma$ -Markov property (or is  $\sigma$ -Markovian) if for  $\mathbf{Y}, \mathbf{Z} \in \mathcal{P}$  and  $S \subset \mathcal{P}$ , we have

(ii) *weak  $\sigma$ -Markov property* (or is *weakly  $\sigma$ -Markovian*) if for  $\mathbf{Y}, \mathbf{Z} \in \mathcal{P}$  and  $S \subset \mathcal{P}$ , we have

**Remark.** In the previous definition,  $S$  is a set of sets and therefore, to be precise, we should have written  $\bigcup_{W \in S} W$  instead of  $S$  in the independence statements. However, whenever the context is clear, we prefer to use  $S$  instead to keep the notation more simple.

**Theorem 2.** *Let  $\mathcal{G}$  be a micro DMG over the micro-variables  $\mathbf{X}$  and suppose that the pair  $(\mathcal{G}, P_{\mathbf{X}})$  is  $\sigma$ -Markovian. Let  $\mathcal{P}$  be a partition of  $\mathbf{X}$  into finite sets, with coarse graph  $\text{co}(\mathcal{G}, \mathcal{P})$ . Then  $(\text{co}(\mathcal{G}, \mathcal{P}), P_{\mathbf{X}})$  is  $\sigma$ -Markovian and consequently weakly  $\sigma$ -Markovian.*

**Remark.** It is worthwhile to remark here that while being sufficient, the  $\sigma$ -Markov property of  $(\mathcal{G}, P_{\mathbf{X}})$  is certainly not necessary for the  $\sigma$ -Markov property of  $(\text{co}(\mathcal{G}, \mathcal{P}), P_{\mathbf{X}})$  as the latter does not care about non-Markovianity *strictly within* variable groups. For instance, if  $\mathbf{Y} = \{Y_1, Y_2\}$ ,  $\mathbf{Z} = \{Z_1\}$  are two variable groups with only one micro-edge  $Y_2 \rightarrow Z_1$ , then the coarse graph  $\mathbf{Y} \rightarrow \mathbf{Z}$  is  $\sigma$ -Markovian with respect to any distribution

as there are no  $\sigma$ -separations. In particular, it is  $\sigma$ -Markovian with respect to distributions in which  $Y_1$  and  $Y_2$  are not independent, that is for distributions that are not  $\sigma$ -Markovian on the micro-graph  $\mathcal{G}$ .

## 4.2 $m$ -Markov properties

For good measure, we provide analogues of Definition 11 and Theorem 2 for  $m$ -separation.

**Definition 12.** ( $m$ -Markov properties) Let  $\mathbf{X}$  be a set of scalar random variables with joint distribution  $P_{\mathbf{X}}$  and let  $\mathcal{P}$  be a partition of  $\mathbf{X}$ . Let  $\mathcal{G}'$  be a mixed graph with node set  $\mathcal{P}$ . We say that  $(\mathcal{G}', P_{\mathbf{X}})$  has the

(i)  $m$ -Markov property (or is  $m$ -Markovian) if for  $\mathbf{Y}, \mathbf{Z} \in \mathcal{P}$  and  $S \subset \mathcal{P} \setminus \{\mathbf{Y}, \mathbf{Z}\}$ , we have

$$\mathbf{Y} \nsim_{\mathcal{G}'}^m \mathbf{Z} \mid S \Rightarrow \mathbf{Y} \perp\!\!\!\perp \mathbf{Z} \mid S.$$

(ii) weak  $m$ -Markov property (or is weakly  $m$ -Markovian) if for  $\mathbf{Y}, \mathbf{Z} \in \mathcal{P}$  and  $S \subset \mathcal{P}$ , we have

$$\mathbf{Y} \nsim_{\mathcal{G}'}^m \mathbf{Z} \mid S \Rightarrow \mathbf{Y} \perp\!\!\!\perp^{pw} \mathbf{Z} \mid S.$$

The analogue of Theorem 2 is as follows. Again, see Appendix B.2 for a proof.

**Theorem 3.** Let  $\mathcal{G}$  be a micro DMG over  $\mathbf{X}$  and suppose that the pair  $(\mathcal{G}, P_{\mathbf{X}})$  is  $m$ -Markovian. Let  $\mathcal{P}$  be a partition of  $\mathbf{X}$  into finite sets, with coarse graph  $\text{co}(\mathcal{G}, \mathcal{P})$ . Then  $(\text{co}(\mathcal{G}, \mathcal{P}), P_{\mathbf{X}})$  is  $m$ -Markovian and in particular weakly  $m$ -Markovian.

## 5 Types of faithfulness for group (D)MGs

In this section, we discuss how different notions of faithfulness on scalar mixed graphs relate to faithfulness on a coarsened graph. As we will see, faithfulness is often not preserved under coarsening. However, we will provide sufficient criteria for faithfulness to hold in both the cyclic and the acyclic setting. We discuss when the strong assumptions that are needed to guarantee faithfulness on the macro-level might be realistic and continue with a discussion on weaker notions of faithfulness. Proofs of the results of this section are provided in Appendix B.3.

As mentioned by Parviainen and Kaski [18], where coarsening a scalar DAGs  $\mathcal{G}$  to a group DAG  $\text{co}(\mathcal{G}, \mathcal{P})$  by means of a partition  $\mathcal{P}$ ,  $d$ -faithfulness, i.e. faithfulness with respect to  $d$ -separation, need not be preserved. Since DAGs are special cases of mixed graphs, and  $m$ -separation/respective  $\sigma$ -separation collapse to  $d$ -separation on DAGs, this conclusion does not change when either of these separations is considered instead. Figure 7 shows simple examples of  $\sigma/m$ -faithfulness violations for  $(\text{co}(\mathcal{G}, \mathcal{P}), P_{\mathbf{X}})$  that occur even if  $\sigma/m$ -Markovianity and  $\sigma/m$ -faithfulness of a the pair  $(\mathcal{G}, P_{\mathbf{X}})$  is assumed. This observation seriously challenges the most naive approach to causal discovery between groups of variables, namely, running the standard PC-algorithm with



**Figure 7:** Two simple examples of  $d$ -faithfulness violations. In the first figure, faithfulness is violated due to the internal disconnectness of  $\mathbf{Y}$ . In the second figure, conditioning on  $\mathbf{Y}$  will open the macro path from  $\mathbf{W}$  to  $\mathbf{Z}$  but closes the micro path, see Figure 6(i) in the study by Anand *et al.* [23, Supplement].

multivariate conditional independence tests or any adaption thereof that relies on the causal faithfulness condition. We also observe that, conversely,  $\sigma/m$ -faithfulness of  $(\text{co}(\mathcal{G}, \mathcal{P}), P_X)$  need not imply  $\sigma/m$ -faithfulness of  $(\mathcal{G}, P_X)$ . This is because any  $\sigma/m$ -faithfulness violation for  $(\mathcal{G}, P_X)$  that is confined within a variable group will not affect  $\sigma/m$ -faithfulness of  $(\text{co}(\mathcal{G}, \mathcal{P}), P_X)$ . As a concrete example, if  $(\mathcal{G}, P_X)$  is not  $\sigma/m$ -faithful and  $\mathcal{P}$  collects all variables in one group, then  $(\text{co}(\mathcal{G}, \mathcal{P}), P_X)$  is always  $\sigma/m$ -faithful for the trivial reason that only one node is present.

## 5.1 Faithfulness criteria for coarse graphs

In this subsection, we will work towards two  $\sigma$ -faithfulness criteria for group DMGs that are obtained from coarsening a micro-DMG. We will start with the following simple characterization of  $\sigma$ -faithfulness.

**Lemma 8.** *Let  $\mathcal{G}$  be a scalar DMG over the micro-variables  $\mathbf{X}$  and suppose that the pair  $(\mathcal{G}, P_X)$  is  $\sigma$ -Markovian and  $\sigma$ -faithful. Let  $\mathcal{P}$  be a partition with coarse graph  $\text{co}(\mathcal{G}, \mathcal{P})$ . Then  $(\text{co}(\mathcal{G}, \mathcal{P}), P_X)$  is  $\sigma$ -faithful if and only if the following holds: whenever  $\mathbf{Y}$  and  $\mathbf{Z}$  are  $\sigma$ -connected by a set  $S \subset \mathcal{P}$ , then there exist  $Y \in \mathbf{Y}$  and  $Z \in \mathbf{Z}$  that are  $\sigma$ -connected by  $\mathcal{T} = \bigcup_{W \in S} W$ .*

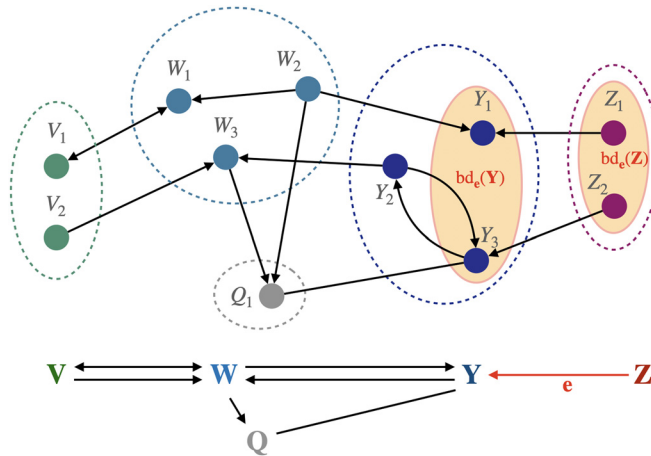
**Corollary 1.** *Let  $\mathcal{G}$  be a scalar DMG over the micro-variables  $\mathbf{X}$  and suppose that the pair  $(\mathcal{G}, P_X)$  is  $\sigma$ -Markovian and  $\sigma$ -faithful. Let  $\mathcal{P}$  be a partition with coarse graph  $\text{co}(\mathcal{G}, \mathcal{P})$ . Assume that for any path  $\Pi$  on  $\text{co}(\mathcal{G}, \mathcal{P})$ , there exists a path  $\pi$  on  $\mathcal{G}$  such that*

- (i)  $\text{co}(\pi) = \Pi$  and
- (ii) *whenever  $\Pi$  is  $\sigma$ -unblocked by a set  $S \subset \mathcal{P}$ , then  $\pi$  is  $\sigma$ -unblocked by  $\mathcal{T} = \bigcup_{W \in S} W$ .*

*Then  $(\text{co}(\mathcal{G}, \mathcal{P}), P_X)$  is  $\sigma$ -faithful.*

**Proof.** This is a direct consequence of Lemma 8. □

In Theorem 4, we will now derive a simple sufficient condition that guarantees  $\sigma$ -faithfulness on a coarsened graph. In a nutshell, it shows that  $\sigma$ -faithfulness *does hold* if variable groups are sufficiently connected internally. Before formulating Theorem 4, we need to introduce some additional definitions (see Figure 8 for a graphical illustration).



**Figure 8:** The two boundaries of the edge  $e$  that is marked in red in the macro graph.

**Definition 13.** Let  $\mathcal{G}$  be a mixed graph with edge sets  $\mathcal{E}$ ,  $\mathcal{B}$ ,  $\mathcal{U}$ , and let  $\mathcal{P}$  be a partition of its nodes. Moreover, let  $\mathbf{e}$  be an edge on  $\text{co}(\mathcal{G}, \mathcal{P})$ .

- If  $\mathbf{e} = \mathbf{Z} \rightarrow \mathbf{Y}$  is right-directed, define the set of  $\mathbf{e}$ -micro edges as follows:

$$\text{mic}(\mathbf{e}) = \{e \in \mathcal{E}; e = Z \rightarrow Y \text{ with } Z \in \mathbf{Z}, Y \in \mathbf{Y}\}.$$

- If  $\mathbf{e} = \mathbf{Z} \leftarrow \mathbf{Y}$  is left-directed, define the set of  $\mathbf{e}$ -micro edges as follows:

$$\text{mic}(\mathbf{e}) = \{e \in \mathcal{E}; e = Z \leftarrow Y \text{ with } Z \in \mathbf{Z}, Y \in \mathbf{Y}\}.$$

- If  $\mathbf{e} = \mathbf{Z} \leftrightarrow \mathbf{Y}$  is bidirected, define the set of  $\mathbf{e}$ -micro edges as follows:

$$\text{mic}(\mathbf{e}) = \{e \in \mathcal{B}; e = Z \leftrightarrow Y \text{ with } Z \in \mathbf{Z}, Y \in \mathbf{Y}\}.$$

- If  $\mathbf{e} = \mathbf{Z} - \mathbf{Y}$  is undirected, define the set of  $\mathbf{e}$ -micro edges as follows:

$$\text{mic}(\mathbf{e}) = \{e \in \mathcal{U}; e = Z - Y \text{ with } Z \in \mathbf{Z}, Y \in \mathbf{Y}\}.$$

Given an arbitrary edge  $\mathbf{e} = (\mathbf{Z}, \mathbf{Y}) \in \mathcal{E} \cup \mathcal{B} \cup \mathcal{U}$ , the  $\mathbf{e}$ -boundary of  $\mathbf{Z}$  is then the projection of  $\text{mic}(\mathbf{e})$  to its source node, i.e.

$$\text{bd}_{\mathbf{e}}(\mathbf{Z}) = \{Z \in \mathbf{Z}; \text{there is } Y \in \mathbf{Y} \text{ such that } (Z, Y) \in \text{mic}(\mathbf{e})\} \subset \mathbf{Z}.$$

Similarly, the  $\mathbf{e}$ -boundary of  $\mathbf{e} = (\mathbf{Z}, \mathbf{Y})$  is defined as follows:

$$\text{bd}_{\mathbf{e}}(\mathbf{Y}) = \{Y \in \mathbf{Y}; \text{there is } Z \in \mathbf{Z} \text{ such that } (Z, Y) \in \text{mic}(\mathbf{e})\} \subset \mathbf{Y}.$$

**Theorem 4.** (Faithfulness criterion 1) Let  $\mathcal{G}$  be a DMG over the micro-variables  $\mathbf{X}$  with distribution  $P_{\mathbf{X}}$ , and let  $\mathcal{P}$  be a partition of its nodes. Assume the following:

- The pair  $(\mathcal{G}, P_{\mathbf{X}})$  is  $\sigma$ -Markovian and  $\sigma$ -faithful.
- For any strongly connected component  $\mathcal{W}$  of  $\mathcal{G}$ , there exists  $\mathbf{W} \in \mathcal{P}$ , with  $\mathcal{W} \subset \mathbf{W}$ .
- For any adjacent pair of edges  $\mathbf{e} = (\mathbf{W}, \mathbf{Y})$ ,  $\mathbf{e}' = (\mathbf{Y}, \mathbf{Z})$ , and any  $Y \in \text{bd}_{\mathbf{e}}(\mathbf{Y})$ , there exists  $Y' \in \text{bd}_{\mathbf{e}'}(\mathbf{Y})$  such that  $\text{sc}_{\mathcal{G}}(Y) = \text{sc}_{\mathcal{G}}(Y')$ .

Then,  $(\text{co}(\mathcal{G}, \mathcal{P}), P_{\mathbf{X}})$  is  $\sigma$ -faithful and  $\text{co}(\mathcal{G}, \mathcal{P})$  is acyclic.

**Corollary 2.** Let  $\mathcal{G}$  be a DMG over the micro-variables  $\mathbf{X}$  with distribution  $P_{\mathbf{X}}$ , and let  $\mathcal{P}$  be the partition into the strongly connected components of  $\mathcal{G}$ . If the pair  $(\mathcal{G}, P_{\mathbf{X}})$  is  $\sigma$ -Markovian and  $\sigma$ -faithful, then  $(\text{co}(\mathcal{G}, \mathcal{P}), P_{\mathbf{X}})$  is  $\sigma$ -faithful.

Corollary 2 is no longer true if  $\sigma$ -separation is replaced by  $m$ -separation. The graph on the right of Figure 4 provides a counterexample, as every micro-path between groups  $\mathbf{W}$  and  $\mathbf{Z}$  is  $m$ -blocked by  $\mathbf{Y}$  but  $\sigma$ -unblocked by  $\mathbf{Y}$ . This example serves as another illustration that the notions of separation entail different consequences, see the study by Bongers *et al.* [33] for more.

The previous results show that if cyclic relationships are present *internal* to the variable groups, this can be an advantage for causal discovery rather than a disadvantage. Assuming a variable group to be well connected internally to achieve  $\sigma$ -faithfulness on the group level is to some degree at odds with assuming acyclicity on the micrograph as acyclicity disallows paths to be present if they induce a cycle. However, if one zooms in on the proof of Theorem 4, it becomes clear that condition (iii) can be replaced by weaker sufficient conditions that still guarantee  $\sigma$ -faithfulness, even if the micrograph is acyclic. These conditions need to be formulated separately for (almost) mediators, confounders and colliders and are therefore more technical to formulate. Here, by an almost mediator, we mean a motive of the form  $A \leftrightarrow B \rightarrow C$  (right-directed almost mediator) or  $A \leftrightarrow B \leftarrow C$  (left-directed almost mediator). For (almost) mediators, condition (iii) can be replaced by

- For any adjacent pair of edges  $\mathbf{e} = \mathbf{W} \rightarrow \mathbf{Y}$ ,  $\mathbf{e}' = \mathbf{Y} \rightarrow \mathbf{Z}$  (or  $\mathbf{e} = \mathbf{W} \leftrightarrow \mathbf{Y}$ ,  $\mathbf{e}' = \mathbf{Y} \rightarrow \mathbf{Z}$ ), and any  $Y \in \text{bd}_{\mathbf{e}}(\mathbf{Y})$  there exists  $Y' \in \text{bd}_{\mathbf{e}'}(\mathbf{Y})$  and a right-directed (possibly trivial) path  $Y \rightarrow \cdots \rightarrow Y'$  that does not leave  $\mathbf{Y}$ .
- For any adjacent pair of edges  $\mathbf{e} = \mathbf{W} \leftarrow \mathbf{Y}$ ,  $\mathbf{e}' = \mathbf{Y} \leftarrow \mathbf{Z}$  or  $(\mathbf{e} = \mathbf{W} \leftarrow \mathbf{Y}, \mathbf{e}' = \mathbf{Y} \leftrightarrow \mathbf{Z})$ , and any  $Y' \in \text{bd}_{\mathbf{e}'}(\mathbf{Y})$  there exists  $Y \in \text{bd}_{\mathbf{e}}(\mathbf{Y})$  and a left-directed (possibly trivial) path  $Y \leftarrow \cdots \leftarrow Y'$  that does not leave  $\mathbf{Y}$ .



For confounders the corresponding condition becomes

- (iii-c) For any adjacent pair of edges  $\mathbf{e} = \mathbf{W} \leftarrow \mathbf{Y}$ ,  $\mathbf{e}' = \mathbf{Y} \rightarrow \mathbf{Z}$ , and any  $Y \in \text{bd}_{\mathbf{e}}(\mathbf{Y})$  there exists  $Y' \in \text{bd}_{\mathbf{e}'}(\mathbf{Y})$  and a confounding path  $Y \leftarrow \dots \leftarrow Y'' \rightarrow \dots \rightarrow Y'$  that does not leave  $\mathbf{Y}$ .

Finding an appropriate condition for colliders is a bit less straightforward, as faithfulness violations may arise by conditioning on a collider  $\mathbf{Y}$ , e.g.  $\mathbf{W} \rightarrow \mathbf{Y} \leftarrow \mathbf{Z}$  in such a way that while a micro-collider inside  $\mathbf{Y}$  is unblocked, a non-collider in  $\mathbf{Y}$  is blocked again, see e.g. the second example in Figure 7. In Lemma 4, this was avoided by enforcing these non-colliders to only point to neighbours in the same strong connected component and condition (iii) in the Definition of  $\sigma$ -separation, Definition 2. The following condition, although strong, will do the job.

- (iii-d) For any adjacent pair of colliding edges  $\mathbf{e} = (\mathbf{W}, \mathbf{Y})$ ,  $\mathbf{e}' = (\mathbf{Y}, \mathbf{Z})$ , and any  $\text{bd}_{\mathbf{e}}(\mathbf{Y}) \cap \text{bd}_{\mathbf{e}'}(\mathbf{Y}) \neq \emptyset$ , i.e. there exist colliding edges  $(W, Y)$ ,  $(Y, Z)$  with  $(W, Y) \in \text{mic}(\mathbf{e})$  and  $(Y, Z) \in \text{mic}(\mathbf{e}')$ .

Thus, we have the following  $\sigma$ -faithfulness criterion that is more meaningful when a micro DMG is acyclic, i.e. an ADMG. Note that in this case, condition (ii) of the following theorem is trivially satisfied. In addition, perhaps surprisingly, it does not enforce the coarse graph  $\text{co}(\mathcal{G}, \mathcal{P})$  to be acyclic as did Theorem 4.

**Theorem 5.** (Faithfulness criterion 2) *Let  $\mathcal{G}$  be a DMG over the micro-variables  $\mathbf{X}$  with distribution  $P_{\mathbf{X}}$ , and let  $\mathcal{P}$  be a partition of its nodes. Assume the following:*

- (i) *The pair  $(\mathcal{G}, P_{\mathbf{X}})$  is  $\sigma$ -Markovian and  $\sigma$ -faithful.*
- (ii) *For any strongly connected component  $\mathcal{W}$  of  $\mathcal{G}$ , there exists  $\mathbf{W} \in \mathcal{P}$ , with  $\mathcal{W} \subset \mathbf{W}$ .*
- (iii) *An adjacent pair of edges  $\mathbf{e} = (\mathbf{W}, \mathbf{Y})$ ,  $\mathbf{e}' = (\mathbf{Y}, \mathbf{Z})$ , satisfies conditions (iii-a), (iii-b), (iii-c), or (iii-d) depending on whether it is a right-directed (almost) mediator, a left-directed almost mediator, a confounder or a collider, respectively.*

*Then,  $(\text{co}(\mathcal{G}, \mathcal{P}), P_{\mathbf{X}})$  is  $\sigma$ -faithful.*

The discussion in this section also shows the importance of choosing variable groups carefully if one wants to guarantee  $\sigma$ -faithfulness which may be a non-trivial task in real-world applications. Parviainen and Kaski [18] tested empirically how often group level faithfulness would be violated in Erdős-Rényi random DAGs with groups of small sizes. They found that such violations were likely to appear in sparse graphs but unlikely to appear in dense random graphs. This matches the theoretical results of this section that internally well-connected groups help to ensure group level faithfulness.

## 5.2 Adjacency and orientation faithfulness

We will therefore consider the two weaker notions of *adjacency faithfulness* and *orientation faithfulness*. The former is at the base of the conservative PC-algorithm [5] and does transfer from the micro-variable to the group level.

**Definition 14.** (Adjacency faithfulness) A pair  $(\mathcal{G}, P)$  of a mixed graph  $\mathcal{G}$  and a distribution  $P$  over its node variables is *adjacency faithful* if any two nodes  $X, Y$  that are independent given some conditioning set  $S$  are not adjacent, i.e. they do not share an edge.

Note that adjacency faithfulness only makes reference to the skeleton of the graph  $\mathcal{G}$  and not to any specific type of separation.

**Lemma 9.** *Let  $\mathcal{G}$  be a mixed graph over the variables  $\mathbf{X}$  with distribution  $P_{\mathbf{X}}$ , and let  $\mathcal{P}$  be a partition that induces the coarse graph  $\text{co}(\mathcal{G}, \mathcal{P})$ . If the pair  $(\mathcal{G}, P_{\mathbf{X}})$  is adjacency faithful on  $\mathcal{G}$ , then the pair  $(\text{co}(\mathcal{G}, \mathcal{P}), P_{\mathbf{X}})$  is adjacency faithful as well.*

**Proof.** Suppose that  $Y \perp\!\!\!\perp Z|S$  for some  $S \subset \mathcal{P} \setminus \{Y, Z\}$ . Because mutual conditional independence implies pairwise conditional independence, it follows by adjacency faithfulness on  $\mathcal{G}$  that  $Y$  and  $Z$  do not share an edge for all  $Y \in \mathbf{Y}, Z \in \mathbf{Z}$ . By definition of  $\text{co}(\mathcal{G}, \mathcal{P})$ ,  $\mathbf{Y}$  and  $\mathbf{Z}$  do not share an edge.  $\square$

**Remark.** Lemma 9 does not use the full strength of adjacency faithfulness on the micro-level: in fact, it suffices to assume that  $X$  and  $Y$  that belong to different variable groups do not share an edge if they are conditionally independent given a conditioning set  $S$ . In other words, adjacency faithfulness violations *within a group* do not matter for adjacency faithfulness on the macro-level.

Combining Theorem 2 with Lemma 9, we see that if  $(\mathcal{G}, P_{\mathbf{X}})$  is a  $\sigma$ -Markovian and adjacency faithful pair of a DMG  $\mathcal{G}$  and a distribution of micro-variables  $P_{\mathbf{X}}$ , then for a given partition  $\mathcal{P}$ , the pair  $(\text{co}(\mathcal{G}, \mathcal{P}), P_{\mathbf{X}})$  is strongly  $\sigma$ -Markovian and adjacency faithful as well. If the graph  $\text{co}(\mathcal{G}, \mathcal{P})$  is moreover a DAG, these are exactly the assumptions that the conservative PC algorithm of [5] requires to be sound. To our knowledge, soundness of conservative PC has not been discussed beyond the acyclic case, but we believe it to hold as well. This is because soundness of the PC algorithm is not affected by allowing cycles and working with  $\sigma$ -separation as demonstrated by Mooij and Claassen [32]. Recall that the conservative PC algorithm takes the observational distribution as an input and outputs a so-called *e-pattern*, see the study by Ramsey et al. [5] for an exact definition.

**Corollary 3.** Let  $(\mathcal{G}, P_{\mathbf{X}})$  be a  $\sigma$ -Markovian and  $\sigma$ -faithful pair of a DMG  $\mathcal{G}$  and a distribution of micro-variables  $P_{\mathbf{X}}$ . Let  $\mathcal{P}$  be a partition of  $\mathbf{X}$  such that  $\text{co}(\mathcal{G}, \mathcal{P})$  is a DAG. Then the conservative PC algorithm with vector-valued (oracle) conditional independence tests is sound for  $\text{co}(\mathcal{G}, \mathcal{P})$  in that it outputs an *e-pattern* that represents  $\text{co}(\mathcal{G}, \mathcal{P})$ .

In an *e-pattern*, specific violations of faithfulness, namely, violations of *orientation faithfulness* can be singled out and are marked by a \*. To recap the definition of orientation faithfulness for DAGs, we recall that a triple of nodes  $(X, Y, Z)$  in a DAG is called *unshielded* if there is an edge between  $X$  and  $Y$  and an edge between  $Y$  and  $Z$  but none between  $X$  and  $Z$ .

**Definition 15.** (Orientation faithfulness) Let  $\mathcal{G}$  be a DAG over a set of variables  $\mathbf{X}$  with distribution  $P_{\mathbf{X}}$ . The pair  $(\mathcal{G}, P_{\mathbf{X}})$  is called *orientation faithful* if for any unshielded triple  $(X, Y, Z)$  the following holds.

- (O1) If  $(X, Y, Z)$  is a collider, then  $X$  and  $Z$  are dependent given any subset of  $\mathbf{X} \setminus \{X, Z\}$  that contains  $Y$ ;
- (O2) If  $(X, Y, Z)$  is a non-collider, then  $X$  and  $Z$  are dependent given any subset of  $\mathbf{X} \setminus \{X, Z\}$  that does not contain  $Y$ ;

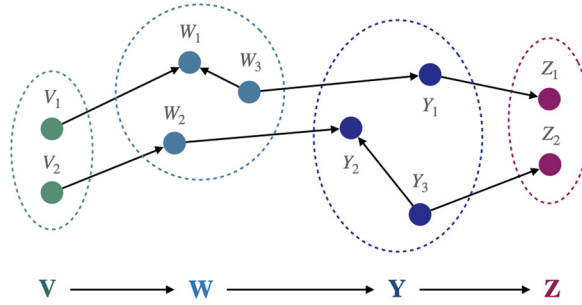
For DMGs with potential cycles, orientation faithfulness is more tricky to define, as the absence of an edge between two nodes  $X, Y$  does no longer mean that they can be  $\sigma$ -separated. To deal with this, we will rather introduce the following notion of *local faithfulness* for DMGs which agrees with orientation faithfulness if the graph is a DAG.

**Definition 16.** (Local faithfulness) Let  $\mathcal{G}$  be a DMG over a set of variables  $\mathbf{X}$  with distribution  $P_{\mathbf{X}}$ . A *local faithfulness violation* is a short path  $(X, e_1, Y, e_2, Z)$  such that there exists a set  $S \subset \mathbf{X} \setminus \{X, Y, Z\}$  with  $X \perp\!\!\!\perp Z|S$  and  $X \perp\!\!\!\perp Y|S$ .

The pair  $(\mathcal{G}, P_{\mathbf{X}})$  is called *locally faithful* if there are no local faithfulness violations.

**Lemma 10.** [5] If  $G$  is a DAG, a pair  $(\mathcal{G}, P_{\mathbf{X}})$  is locally faithful if it is orientation faithful.

Examples of faithfulness violations in the literature are typically either violations of adjacency or orientation faithfulness. Figure 9 shows that if the nodes correspond to variable groups, there are faithfulness violations that are non-local. In other words, both orientation and adjacency faithfulness are satisfied, still



**Figure 9:** An example of a non-local  $\sigma$ -faithfulness violation (resp.  $d$ -faithfulness violation as there are no cycles). If the joint distribution is  $\sigma$ -Markovian and  $\sigma$ -faithful to the micro graph, the group DAG does not contain any orientation faithfulness violations. At the same time, all micro paths between the groups  $\mathbf{V}$  and  $\mathbf{Z}$  are  $\sigma$ -blocked, while the only macro path is  $\sigma$ -open.

( $\sigma$ - or  $d$ -)faithfulness is violated. In particular, such non-local violations would not be marked in the output of the conservative PC algorithm.

### 5.3 Faithfulness and Meek's orientation rules revisited

Constrained-based algorithms for causal discovery such as the PC-algorithm [2] infer the directionality of arrows in a DAG by first identifying  $v$ -structures and then applying Meek's orientation rules<sup>2</sup> [26]. In this subsection, only the first of these rules will be relevant. It states that an edge  $X - Y$  is to be oriented as  $X \rightarrow Y$  if there is an edge  $Z \rightarrow X$  such that  $Z$  and  $Y$  are non-adjacent. Parviainen and Kaski [18] discuss the validity of Meek's orientation rules for group DAGs using the example depicted in Figure 10. Translated to our terminology, their example consists of a micro-variable DAG  $\mathcal{G}$ , a partition  $\mathcal{P} = \{\mathbf{V}, \mathbf{W}, \mathbf{Y}, \mathbf{Z}\}$  of the micro-variables and a group DAG  $\mathcal{G}'$  with nodes  $\mathbf{V}, \mathbf{W}, \mathbf{Y}, \mathbf{Z}$  such that

- the micro-level pair  $(\mathcal{G}, P)$  is causally Markovian and  $d$ -faithful, where  $P$  is the micro-variable distribution;
- the macro-level pair  $(\mathcal{G}', P)$  is causally Markovian and  $d$ -faithful;
- $\mathcal{G}' \neq \text{co}(\mathcal{G}, P)$  and in particular  $\mathbf{Y} \leftarrow \mathbf{V}$  in  $\text{co}(\mathcal{G}, P)$  and  $\mathbf{Y} \rightarrow \mathbf{V}$  in  $\mathcal{G}'$ .

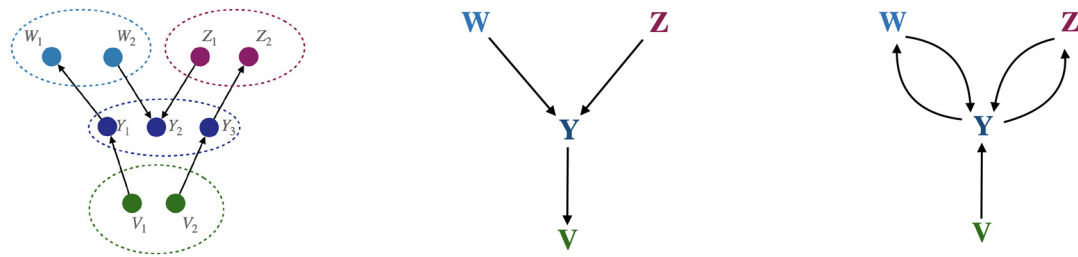
As the mentioned orientation rule implies the orientation  $\mathbf{Y} \rightarrow \mathbf{V}$  of  $\mathcal{G}'$  instead of the correct orientation  $\mathbf{Y} \leftarrow \mathbf{V}$  in the ground truth group DMG  $\text{co}(\mathcal{G}, P)$ , Parviainen and Kaski [18] argue that Meek's orientation rules are no longer valid for group DAGs even if  $d$ -faithfulness on the group level does hold. However, we argue that faithfulness should refer to the cyclic *ground truth graph*  $\text{co}(\mathcal{G}, P)$ , and the pair  $(\text{co}(\mathcal{G}, P), P)$  *does* violate  $\sigma$ -faithfulness: the groups  $\mathbf{W}$  and  $\mathbf{Z}$  are not  $\sigma$ -separated in  $\text{co}(\mathcal{G}, P)$  but are independent. In fact, by the study by Mooij and Claassen [32, Corollary 1] which does not make assumptions on the dimensionality of the node variables, the PC-algorithm (and thus the Meek rules for DAGs) is sound if the ground truth graph of groups is directed and acyclic, and if this DAG and the joint distribution of the variables are assumed  $d$ -faithful<sup>3</sup> to each other. To summarize, in the example given in the study by Parviainen and Kaski [18], the Meek rules lead to a wrong orientation, because the graph of groups is incorrectly assumed to be acyclic.

## 6 Grouped time series graphs

When using graphical modes to model causation for time evolving processes, there are several common modeling choices that are discussed in the literature that can all be adapted to the group setting. The arguably most common notion is that of a (stationary) time series DMG (ts-DMG for short)  $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{B})$  in

<sup>2</sup> Note that these rules pertain to DAGs, not to general DMGs.

<sup>3</sup> Recall that  $d$ -faithfulness and  $\sigma$ -faithfulness are equivalent for DAGs.



**Figure 10:** Left: The micro DAG  $\mathcal{G}$  presented in the study by Parviainen and Kaski [18]. Middle: The macro DAG presented in the study by Parviainen and Kaski [18]. Right: The coarse DMG  $\text{co}(\mathcal{G}, \mathcal{P})$  with respect to the indicated partition.

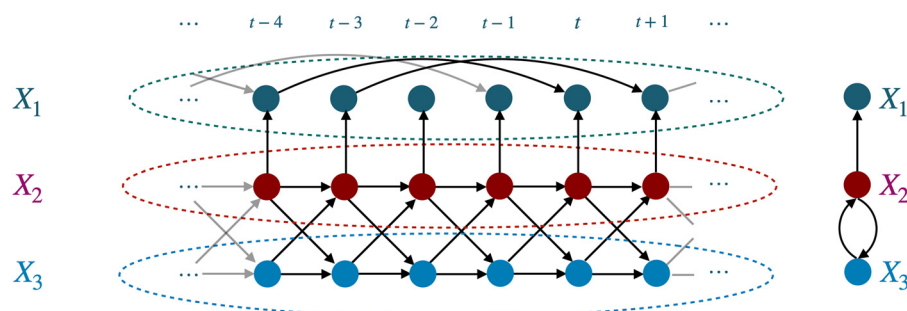
which the processes are unrolled in time and discretized. That is, the processes are modelled as univariate infinite time series  $X_i = (X_i(t))_{t \in \mathbb{Z}}$ ,  $i \in I = \{1, \dots, n\}$  and the nodes of the ts-DMG correspond to the indices  $(i, t) \in \mathcal{V} = I \times \mathbb{Z}$ . As usual, we freely identify an index  $(i, t)$  with a variable  $X_i(t)$  as long as there is no danger of confusion. In other words, there is a node in the causal graph for every time instance of every process. In addition, directed edges are not allowed to point into the past, i.e.  $X_i(s) \rightarrow X_j(t)$  implies  $s \leq t$ . Finally, the stationarity assumption means that the presence of edges only depends on the time lag between nodes and not the actual time instances. More precisely, if there is a directed or bidirected edge  $(X_i(s), X_j(t))$ , then there is an edge  $(X_i(s+u), X_j(t+u))$  of the same type for any  $u \in \mathbb{Z}$ . A coarser representation of causal interactions between time series is that of a *time series summary DMG* or *process DMG*  $\mathcal{G}^{\text{sum}} = (\mathcal{V}^{\text{sum}}, \mathcal{E}^{\text{sum}}, \mathcal{B}^{\text{sum}})$  in which a node corresponds to a process  $X_i$  as a whole, i.e.  $\mathcal{V}^{\text{sum}} = I$ . Such graphs thus express whether processes causally influence each other but hold no information on the time lag of the interaction. Depending on the convention, self-edges  $(X_i, X_i)$  are allowed or not allowed, and we stick to the latter (no self-edges) in this work. While some causal discovery methods [10,36,37] aim to infer the time unrolled ts-DMG, others such as Granger causality [38] infer the process graph. Clearly, any ts-DMG can be projected to a process DMG by ignoring the time component and adding a (bi)directed edge  $(X_i, X_j)$ ,  $i \neq j$  if and only if there is a (bi)directed edge  $(X_i(s), X_j(t))$  for some  $s, t \in \mathbb{Z}$ . Note that this is nothing but a special instance of our coarsening operation in the case where micrographs have infinite nodes, see Figure 11.

**Lemma 11.** If  $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{B})$ ,  $\mathcal{V} = I \times \mathbb{Z}$  is a time series DMG, then its summary DMG is  $\text{co}(\mathcal{G}, \mathcal{P})$  for the partition  $\mathcal{P} = \{\{X_i\} \times \mathbb{Z}\}_{i \in I} \cong I$ .

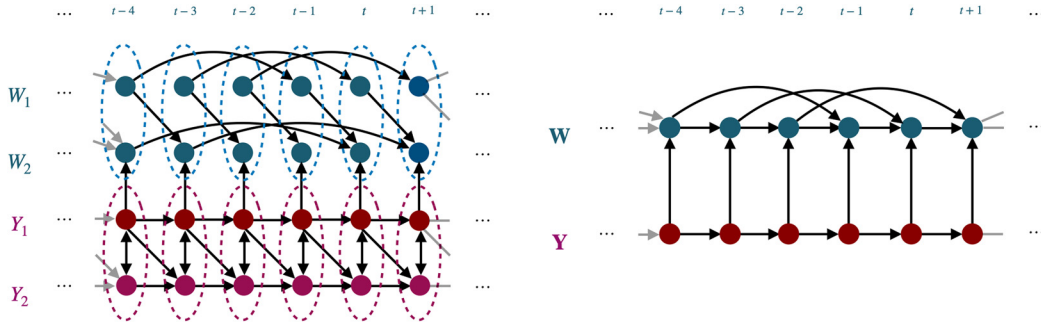
Of course, there is no formal reason to disallow more general partitions of  $\mathcal{V} = I \times \mathbb{Z}$ . For instance, when  $Q$  is a partition of the set of processes  $\{X_1, \dots, X_n\} \cong I$ , we can define the *grouped ts-DMG* of  $\mathcal{G}$  as  $\text{co}(\mathcal{G}, Q')$  where  $Q' = \{\mathbf{Y} \times \{t\}; \mathbf{Y} \in Q, t \in \mathbb{Z}\}$  is the *contemporaneous partition* of  $Q$ , see Figure 12. We can coarsen the grouped ts-DMG further to obtain the *grouped summary DMG* or *grouped process DMG*

$$\text{co}(\mathcal{G}, Q')^{\text{sum}} = \text{co}(\mathcal{G}, Q'')$$

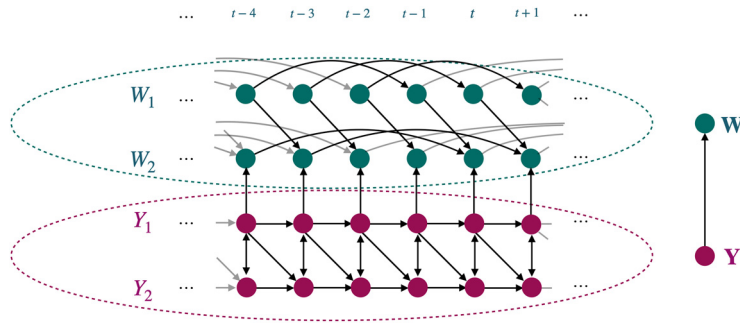
where  $Q'' = \{\mathbf{Y} \times \mathbb{Z}; \mathbf{Y} \in Q\} \cong Q$  is the *full process partition* of  $Q$ , see Figure 13.



**Figure 11:** The summary graph viewed as a coarsened group DMG of the unrolled time series DMG.



**Figure 12:** Left: A partition  $\mathcal{P}$  of an unrolled time series DMG into contemporaneous groups. Right: The grouped ts-DMG  $\text{co}(\mathcal{G}, \mathcal{P})$  with respect to the partition  $\mathcal{P}$ .



**Figure 13:** Left: A process grouping of a ts DMG. Right: The corresponding grouped summary DMG.

## 6.1 Faithfulness in grouped time series graphs

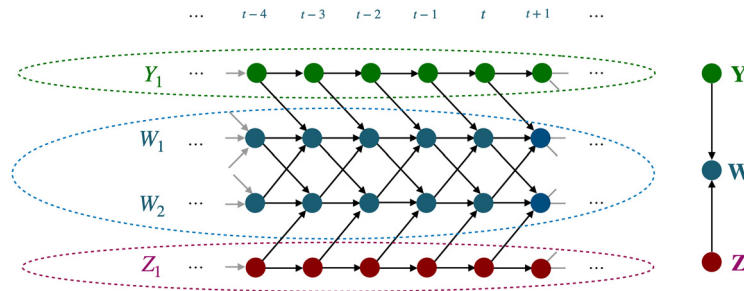
Given that grouped ts-DMGs and grouped summary DMGs are special cases of coarsened graphs, the criteria of Theorems 4 and 5 are still sufficient to ensure  $\sigma$ -faithfulness.

**Corollary 4.** *Let  $\mathcal{G}$  be time series DMG and let  $Q$  be a partition of the set of processes  $\{X_1, \dots, X_n\}$  with contemporaneous partition  $Q'$  and full process partition  $Q''$ . Moreover, let  $P_X$  be the joint distribution of  $\{X_i(t)\}_{i \in I, t \in \mathbb{Z}}$ . If the assumptions of Theorems 4 or 5 are satisfied with respect to  $Q'$  (respectively  $Q''$ ), then the pair  $(\text{co}(\mathcal{G}, Q'), P_X)$  (respectively  $(\text{co}(\mathcal{G}, Q''), P_X)$ ) is  $\sigma$ -faithful.*

At the same time, if these criteria are not assumed to hold, violations of  $\sigma$ -faithfulness are still easily constructed even if there are no contemporaneous edges and all micro-processes are autocorrelated, see Figure 14 for a faithfulness violation on the grouped summary DMG. In addition, in micro-level ts-DMGs, cycles can only appear in the contemporaneous part of the graph as directed edges cannot point backwards in time. Cycles will thus only be included in the grouped time series DMG if the time resolution of the analyzed data is not fine enough to resolve all feedback loops. If the time resolution is believed to be fine enough, all cycles are resolved which renders Theorem 4 useless in the ts-domain.

## 7 Interpretation of causation in group (D)MGs

Many of the examples presented in this work, see Figures 7 and 9, show that group DMGs have to be carefully interpreted when associating a causal meaning to paths in the graph; a point that has already been made by Parviainen and Kaski [18]. They formulate a notion of *potential* and *actual causation* in terms of interventions that can be mirrored in our graphical language.



**Figure 14:** A  $d$ -faithfulness violation on the grouped summary DMG. Conditioning on the process  $W$  blocks all micro paths between the processes  $Y$  and  $Z$ .

**Definition 17.** (Apparent and true causes) Let  $\mathcal{G}$  be a DMG over a set of micro-variables  $X$  and let  $\mathcal{P}$  be a partition of  $X$  inducing the group DMG  $\text{co}(\mathcal{G}, \mathcal{P})$ . We say that  $Y \in \mathcal{P}$  is an *apparent cause* of  $Z \in \mathcal{P}$  if there exists a directed path  $Y \rightarrow \cdots \rightarrow Z$  on  $\text{co}(\mathcal{G}, \mathcal{P})$ .  $Y$  is called a *true cause* of  $Z$  if there is a directed path  $Y \rightarrow \cdots \rightarrow Z$  on  $\mathcal{G}$  for some  $Y \in Y$  and  $Z \in Z$ .

In other words, directed paths on group DMGs may not be regarded as truly causal in general as corresponding micro-paths might be absent. In particular, intervening on a potential cause  $Y$  of  $Z$  might not change the distribution of the effect group  $Z$ . We record the following result for good measure.

**Lemma 12.** Let  $\mathcal{G}$  be a DMG over a set of micro-variables  $X$  and let  $\mathcal{P}$  be a partition of  $X$  inducing the group DMG  $\text{co}(\mathcal{G}, \mathcal{P})$ .

- If  $Y \rightarrow Z$  is a directed edge, then  $Y$  is a true cause of  $Z$ .
- If the conditions (ii) and (iii) of Theorem 4 are satisfied, then any apparent cause of a group  $Z \in \mathcal{P}$  is a true cause of  $Z$ .
- If the condition (ii) and (iii-a) of Theorem 5 are satisfied, then any apparent cause of a group  $Z \in \mathcal{P}$  is a true cause of  $Z$ .

**Proof.** The first claim of the lemma follows directly from the definition of  $\text{co}(\mathcal{G}, \mathcal{P})$ . The second and third claim follows directly from the proof of Lemmas 4 and 5 where for a given directed path  $\Pi = Y \rightarrow \cdots \rightarrow Z$  on  $\text{co}(\mathcal{G}, \mathcal{P})$ , we constructed a connecting directed micro-path  $\pi = Y \rightarrow \cdots \rightarrow Z$  on  $\mathcal{G}$  for some  $Y \in Y$  and  $Z \in Z$  such that  $\text{co}(\pi) = \Pi$ .  $\square$

## 7.1 Causation in grouped time series graphs

We now turn to the question whether any apparent cause in a grouped ts-DMGs or a grouped summary DMG is a true cause. For grouped ts-DMGs, the answer is no for the same reason as for usual group DMGs. On the level of the grouped summary graph, however, apparent causation implying true causation may be more realistic, at least if the grouped processes are believed to be *causally mixing*, a notion inspired by the common assumption of mixing in dynamical systems.

**Definition 18.** Consider a ts-DMG  $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{B})$  over micro processes  $X_1, \dots, X_n$ ,  $X_i = (X_i(t))_{t \in \mathbb{Z}}$ . Let  $Q$  be a partition of  $\{X_1, \dots, X_n\}$  and consider the induced grouped ts-DMG  $\text{co}(\mathcal{G}, Q)$  where  $Q' = \{Y \times \{t\}; Y \in Q, t \in \mathbb{Z}\}$ .

Then, the pair  $(\mathcal{G}, Q)$  is called *causally mixing* if for any  $Y \in Q$  and any pair of micro-processes  $X_i, X_k \in Y$  the following holds:

- for any  $s \in \mathbb{Z}$ , there exists  $t > s$  and a directed path  $X_i(s) \rightarrow X_{i_1}(s+1) \rightarrow X_{i_2}(s+2) \rightarrow \cdots \rightarrow X_{i_m}(t-1) \rightarrow X_k(t)$  such that  $X_{i_\alpha} \in Y$  for all  $\alpha = 1, \dots, m$ .



Causal mixing means that after a sufficient amount of time has passed, causal information has fully spread throughout any process group. We will see now that causal mixing ensures that, at least at the level of the grouped summary graph, directed causal paths can be understood in the usual sense as any apparent cause is a true cause. However, causal mixing does not ensure  $\sigma$ -faithfulness on the grouped summary DMG as the example in Figure 14 demonstrates.

**Lemma 13.** *Consider a stationary ts-DMG  $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{B})$  over micro processes  $X_1, \dots, X_n$ ,  $X_i = (X_i(t))_{t \in \mathbb{Z}}$ . Let  $Q$  be a partition of  $\{X_1, \dots, X_n\}$  and consider the induced grouped summary DMG  $\tilde{\mathcal{G}} = \text{co}(\mathcal{G}, Q'')$ , where  $Q'' = \mathbf{Y} \times \mathbf{Z}$ ;  $\mathbf{Y} \in Q$ . If  $(\mathcal{G}, Q)$  is causally mixing, then every apparent cause in  $\tilde{\mathcal{G}}$  is a true cause in  $\tilde{\mathcal{G}}$ .*

**Proof.** For this proof, recall that we can identify elements of  $Q$  and  $Q''$  through the map  $\mathbf{Y} \mapsto \mathbf{Y} \times \mathbf{Z}$ . Consider two process groups  $\mathbf{Y}, \mathbf{Z} \in Q$ . Moreover, let  $\Pi \times \mathbf{Z} = (\Pi(1) \times \mathbf{Z}, \mathbf{e}_1, \dots, \mathbf{e}_{r-1}, \Pi(r) \times \mathbf{Z})$  be a directed path on from  $\Pi(1) \times \mathbf{Z} = \mathbf{Y} \times \mathbf{Z}$  to  $\Pi(r) \times \mathbf{Z} = \mathbf{Z} \times \mathbf{Z}$  in the group summary DMG  $\tilde{\mathcal{G}} = \text{co}(\mathcal{G}, Q'')$ . We need to show that there exists a micro-path  $\pi$  in  $\mathcal{G}$  from  $Y(s)$  to  $Z(t)$ ,  $s \leq t$ , for some micro-processes  $Y \in \mathbf{Y}$  and  $Z \in \mathbf{Z}$ . We construct  $\pi$  inductively as follows. First choose a directed micro-edge  $e_1 = Y(s) \rightarrow W(s_1) \in \text{mic}(\mathbf{e}_1)$  for some  $s \leq s_1$ . Then, consider  $\Pi(i)$ ,  $1 < i < r$  and assume that a directed micro-path  $\pi_i$  that ends in  $W(s_{i-1}) \in \text{bd}_{\mathbf{e}_{i-1}}(\Pi(i) \times \mathbf{Z})$  has already been constructed. Choose a micro-process  $W' \in \Pi(i)$  such that  $W'(t') \in \text{bd}_{\mathbf{e}_i}(\Pi(i) \times \mathbf{Z})$  for some  $t' \in \mathbb{Z}$ . By causal mixing there is a directed path  $\xi_i$  from  $W(s_{i-1})$  to  $W'(t_i)$  for some  $t_i > s_{i-1}$  that does not leave  $\Pi(i) \times \mathbf{Z}$ . Stationarity of  $\mathcal{G}$  and  $W'(t') \in \text{bd}_{\mathbf{e}_i}(\Pi(i) \times \mathbf{Z})$  imply that also  $W'(t_i) \in \text{bd}_{\mathbf{e}_i}(\Pi(i) \times \mathbf{Z})$  so we can find a micro-edge  $e_{i+1} \in \text{mic}(\mathbf{e}_{i+1})$  whose source node is  $W'(t_i)$ . After concatenating  $\pi_{i+1} = \pi_i \circ \xi_i \circ e_{i+1}$ , we have obtained the micro-path  $\pi_{i+1}$  to  $\Pi(i+1) \times \mathbf{Z}$  and we continue inductively until we reach  $\mathbf{Z} \times \mathbf{Z}$ .  $\square$

## 8 Further discussions and outlook

In this section, we will zoom out from the technical results of the previous sections and turn towards a high-level discussion on variable groupings and dimension reduction.

### 8.1 Choosing variable groups

In this work, we have operated under the standing assumption that the partition  $\mathcal{P}$  of all micro-variables into variable groups is fixed. We have then studied the transferal of causal discovery assumptions from the micro- to the group level given this fixed partition  $\mathcal{P}$ . While in many problems, practitioners may have clear ideas on which micro-variables should be grouped together or not, in others there might be more than one plausible choice of partition. When the goal is to make this choice in such a way that faithfulness is a realistic assumption on the group level, Theorems 4 and 5 at least provide a heuristic: there should be sufficient causal interactions internal to the variable groups. In particular, grouping together micro-variables that seem to be unrelated causally, appears to be problematic. This seems to be in line with our intuition. After all, why would one group together variables that seem unrelated in the first place? Beyond these heuristic considerations, learning pairs  $(\mathcal{P}, \mathcal{G}(\mathcal{P}))$  of a partition  $\mathcal{P}$  and a graph  $\mathcal{G}(\mathcal{P})$  over its constituents from data under appropriate optimality constraints may be an interesting, although challenging problem for future research.

### 8.2 Dimension reduction and causal discovery

As alluded to in Section 1, in observation-based analyses of causal interactions, the common alternative to working with variable groups in their entirety, is to reduce them to a single univariate variable, or, if they

evolve dynamically, to a single index time series. While some form of dimensionality reduction is unavoidable in high-dimensional settings, the goal of this paragraph is to point out the pitfalls of applying a causal discovery method to dimensionally-reduced proxies, at least if dimension reduction is applied naively. In the subsequent paragraph, we will carry out a similar analysis for a second naive approach, namely, using all available micro-variables as the input of a constraint-based causal discovery method. We contrast this to constraint-based group level discovery, that is the application of a constraint-based method such as the PC algorithm to groups of random variables in which only multivariate conditional independence test between groups are employed as a whole.

### 8.2.1 Applying causal discovery to dimensionally-reduced variables

The most common dimension reduction approach to causal discovery on variable groups  $\mathbf{X}_1, \dots, \mathbf{X}_r$  proceeds as follows:

- Reduce  $\mathbf{X}_i$  to a univariate random variable  $X_i$ , for instance by setting  $X_i = m(\mathbf{X}_i)$  to be the group mean or the first principal component in a PCA on  $\mathbf{X}_i$ .
- Apply a causal discovery algorithm to  $X_1, \dots, X_r$ .

This procedure is appealing to domain researchers for several reasons. First, dimension reduction techniques can be carried out quickly, they counter the *curse of dimensionality*, and the resulting quantities can often be interpreted easily. Moreover, as per the law of large numbers, averaging can help to reduce observational noise, at least if noise terms of different members of a given variable group are believed to be weakly correlated. For instance, if every member  $X_i^j$  of group  $\mathbf{X}_i$  is believed to be produced by a common driver and purely observational noise, i.e.  $X_i^j = \hat{X}_i + \eta_{i,j}$  and the noise terms  $\eta_{i,j}$  have mean zero and are weakly or uncorrelated across the  $j$  index, then in the large group limit, the group mean  $\hat{X}_i = m(\mathbf{X}_i)$  will recover  $\hat{X}_i$ . Thus, if the causal dynamics are modelled by structural equations on the  $\hat{X}_i$  such as  $\hat{X}_i = f_i(\text{pa}(\hat{X}_i), \eta_{\hat{X}_i})$  with  $\text{pa}(\hat{X}_i) \subset \{\hat{X}_1, \dots, \hat{X}_r\}$  and the groups arise as  $\mathbf{X}_i = (\hat{X}_i, \dots, \hat{X}_i)^T + \boldsymbol{\eta}_i$  with zero-mean noise vectors that are mutually independent across the  $i$  index and whose components are weakly or uncorrelated, then the group mean will be an appropriate choice of aggregation technique to recover the causal dynamics.

On the other hand, if different parts of a given cause group  $\mathbf{Y}$  have opposing causal effects on a target group  $\mathbf{Z}$  that roughly cancel each other, the effect of the group mean of  $\mathbf{Y}$  on the group mean of  $\mathbf{Z}$  may be zero, and neither the causal effect nor the dependence  $\mathbf{Y} \perp\!\!\!\perp \mathbf{Z}$  can be recovered from the averaged data. An often invoked real-world example of this are the opposite-sign effects of two different types of blood cholesterol, low-density lipoprotein (LDL) and high-density lipoprotein (HDL), on heart disease [15]. Consequently, research on the effect of total blood cholesterol (LDL + HDL) on heart disease has come to contradictory conclusions.

In a similar vein, conditioning on the mean value  $m(\mathbf{W})$  of a variable group  $\mathbf{W}$  may not suffice to recover a conditional independence  $\mathbf{Y} \perp\!\!\!\perp \mathbf{Z} | \mathbf{W}$ . For instance, consider a structural causal model

$$\begin{aligned} W_1 &= \eta_{W_1} \\ W_2 &= \eta_{W_2} \\ Y &= W_1 + 2W_2 + \eta_Y \\ Z &= W_1 + 2W_2 + \eta_Z, \end{aligned}$$

with variable partition  $\mathbf{Y} = \{Y\}$ ,  $\mathbf{Z} = \{Z\}$ ,  $\mathbf{W} = \{W_1, W_2\}$ , and with independent noise terms  $\eta_{W_1}, \eta_{W_2}, \eta_Y, \eta_Z$ . Then we have  $\mathbf{Y} \perp\!\!\!\perp \mathbf{Z} | \mathbf{W}$  but  $m(\mathbf{Y}) \not\perp\!\!\!\perp m(\mathbf{Z}) | m(\mathbf{W})$ , where again  $m(\cdot)$  denotes the group mean. The latter relation becomes apparent when rewriting  $Y = 2m(\mathbf{W}) + W_2 + \eta_Y$  and  $Z = 2m(\mathbf{W}) + W_2 + \eta_Z$ , so that after conditioning on  $m(\mathbf{W})$ ,  $Y$  and  $Z$  still share the common random component  $W_2$  which is not fully determined by  $m(\mathbf{W})$ . Thus, causal discovery approaches that invoke conditional independence tests on aggregated quantities may come to wrong conclusions. However, this example also illustrates that the primary reason for such faulty inferences is that dimension reduction and inference were conducted *independently of each other*. In fact, in the example

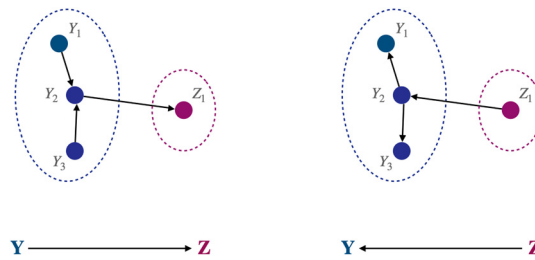
above, there is an aggregation of  $\mathbf{W}$  that does preserve the independence  $\mathbf{Y} \perp\!\!\!\perp \mathbf{Z}|\mathbf{W}$ : if  $m'(\mathbf{W}) = W_1 + 2W_2$ , then  $\mathbf{Y} \perp\!\!\!\perp \mathbf{Z} | m'(\mathbf{W})$ . Research on how variable aggregation and inference can be combined in such a way that they inform each other, is still relatively scarce, and we refer to previous studies [15,39,40] for interesting ideas and further discussions.

### 8.2.2 Micro-level causal discovery

A second straightforward approach to causal discovery on variable groups  $\mathbf{X}_1, \dots, \mathbf{X}_r$  roughly works as follows:

- (1) apply a given causal discovery method to the totality of all micro-variables. This will output a graph over all micro-variables containing edges of different types.
- (2a) Then coarsen this micro-graph as in Definition 4, that is draw an edge of a specific type between groups  $\mathbf{Y}$  and  $\mathbf{Z}$  if there exists an edge of this type between two members  $Y \in \mathbf{Y}$  and  $Z \in \mathbf{Z}$  of these groups.
- (2b) Alternatively, if only one edge is to be allowed between groups, decide on the type of this edge by a majority rule, e.g. draw a directed edge  $\mathbf{Y} \rightarrow \mathbf{Z}$  if the majority of edges between members  $Y \in \mathbf{Y}$  and  $Z \in \mathbf{Z}$  are directed as  $Y \rightarrow Z$ .

As constraint-based causal discovery algorithms such as PC typically come with soundness and completeness guarantees under method-specific assumptions [1,41], in theory, the micro-graph (and therefore the macro-graph derived from it) can be inferred to an optimal degree, that is up to a certain type of equivalence. Still, in practice, there are some obvious drawbacks of such an approach. First, as the number of micro-variables within groups can be very high, the computational effort can be massive while much of the inferred micro-level information, namely, all interaction *internal to variable groups* is of little relevance to the actual task of inferring the interactions *between variable groups*. This issue is particularly problematic if the variable groups happen to be very dense, i.e. if there are many micro-edges within groups. This is because this case falls firmly into the computational worst case scenario for constraint-based causal inference in which computing time grows exponentially with the number of variables [41]. At the same time, one can argue that typically variable groups are chosen the way they are exactly because their members are highly correlated or have strong causal interactions. From a statistical perspective, running many conditional independence tests on the micro-level that are irrelevant to the actual inference task tends to be detrimental to the method's success, see Wahl et al. [17] for some toy experiments with two variable groups and continuous data. In addition, the well-known finite sample guarantees of Kalisch and Bühlmann [41] for the PC algorithm again rely on sparsity conditions that may not be met on the micro-graph if the variable groups are very dense, while they might be met on the coarse group DMG.



**Figure 15:** Left: Running the PC-algorithm with perfect independence tests on the micro-variables will infer the full micro-structure and will therefore also be able to orient the group level edge  $\mathbf{Y} \rightarrow \mathbf{Z}$ . group level PC will not be able to infer this orientation. Right: If, due to a wrong statistical test result or due to a faithfulness violation, the micro-level PC-algorithm mistakenly judges  $Y_1 \perp\!\!\!\perp Y_3$ , it has found a separating set for  $Y_1$  and  $Y_3$  that does not contain  $Y_2$  and will thus orient the unshielded triple  $Y_1 - Y_2 - Y_3$  as a collider  $Y_1 \rightarrow Y_2 \leftarrow Y_3$ . If the remaining tests return the true (in)dependencies  $Y_1 \perp\!\!\!\perp Z|Y_2$ ,  $Y_3 \perp\!\!\!\perp Z|Y_2$ , then PC's orientation rules will imply the edge orientation  $Y_2 \rightarrow Z$ . Hence, the PC-algorithm will again infer the micro-structure on the left and the wrong group level orientation  $\mathbf{Y} \rightarrow \mathbf{Z}$ . Note that the wrong test only involves micro-variables that belong to group  $\mathbf{Y}$ . group level PC will never run this wrong test and will not orient the edge  $\mathbf{Y} - \mathbf{Z}$ , neither correctly nor wrongly.

**Table 1:** Strengths and weaknesses of the three fundamental approaches to causal discovery for variable groups: causal discovery after dimension reduction, micro-level causal discovery, and group level causal discovery. Approaches that integrate dimension reduction and inference, while perhaps retaining reduced variable groups of smaller size might be a fruitful middle ground

	Dimension reduction + CD	Micro-level CD	group level CD
Strengths	Computationally most efficient approach; noise-removal.	Good for small groups; empirically superior to group level CD on discrete data.	Fewer CI tests than micro-level CD; robust to within-group confounding and other violations.
Weaknesses	May change conditional independencies and causal conclusions fundamentally.	Computationally inefficient; vulnerable to within-group assumption violations.	Assumptions and interpretation of output must be evaluated carefully; multivariate CI testing less developed; computationally less efficient than dimension reduction + CD.

On the other hand, full micro-variable causal discovery can sometimes orient edges between groups that a group level approach cannot orient, see Figure 15. This can be both a blessing and a curse: While additional orientations are a plus whenever they are correct, a wrong statistical test result of an independence test that only involves micro-variables *within the same group* can lead to a wrongly oriented edge *between* variable groups, see Figure 15. Therefore, group level causal discovery can be considered more conservative than full micro-level causal discovery in the sense that it might provide fewer orientations while being more robust to testing errors. Finally, if the causal discovery algorithm at hand assumes the absence of hidden confounders, it will suffer if hidden confounding is actually present in the data. Hence, if hidden confounders only affect micro-variables within the same group, then micro-level causal discovery will be challenged while group level causal discovery will only be affected by confounders between different groups, see again the discussion in Section 4. Nevertheless, in the case of discrete data, conditional independence tests are particularly challenged by large conditioning sets as every state of the conditioning variables has to be considered separately. In this case, the empirical experiments conducted by Parviainen and Kaski [18] suggest that the micro-level causal discovery approach which employs more tests but has smaller conditioning sets than the group level approach outperforms the latter.

We summarize strengths and pitfalls of dimension reduction causal discovery, micro-level causal discovery as well as group level causal discovery in Table 1.

## 9 Summary

In this work, we have provided a thorough discussion of assumptions for causal discovery on groups of random variables. In particular, we have shown that causal faithfulness is easily violated in generic settings so that faithfulness-based causal discovery methods need to be applied with care. On the other hand, we have presented two criteria (Theorems 4 and 5) on the internal connectivity of variable groups that do guarantee  $\sigma$ -faithfulness. It will be important to develop and evaluate more elaborate group level causal discovery techniques and to compare them to the baseline methods presented in Section 8 empirically, in particular for continuous data. On the theoretical side, it would be worthwhile to study the compatibility of statistical dimension reduction and causal modelling in greater detail, for instance following the ideas laid out in [15,39,40].

**Acknowledgements:** The authors thank Sofia Faltenbacher for designing the layout used for many figures in this work, as well as the reviewers for their valuable suggestions.

**Funding information:** J.W., U.N., and J.R. received funding from the European Research Council (ERC) Starting Grant CausalEarth under the European Union's Horizon 2020 research and innovation program (Grant Agreement No. 948112).

**Author contributions:** All authors have accepted responsibility for the entire content of this manuscript and consented to its submission to the journal, reviewed all the results and approved its final version. J.W. wrote the manuscript, and formalized and proved the mathematical results of this work. J.R. proposed the initial idea to investigate causal discovery on variable groups. All three authors discussed and advanced the main ideas of this work on several occasions. U.N. and J.R. proofread the manuscripts and suggested valuable improvements.

**Conflict of interest:** The authors declare no conflict of interest.

## References

- [1] Pearl J. Causality: Models, Reasoning and Inference. 2nd ed. USA: Cambridge University Press; 2009.
- [2] Spirtes P, Glymour C, Scheines R. Causation, prediction, and search. vol. 81 of Lecture Notes in Statistics. New York, NY: Springer; 1993. <http://link.springer.com/10.1007/978-1-4612-2748-9>.
- [3] Spirtes P. An anytime algorithm for causal inference. In: International Workshop on Artificial Intelligence and Statistics. PMLR; 2001. p. 278–85. <https://proceedings.mlr.press/r3/spirtes01a.html>.
- [4] Peters J, Janzing D, Schölkopf B. Elements of causal inference - foundations and learning algorithms. Adaptive Computation and Machine Learning Series. Cambridge, MA, USA: The MIT Press; 2017.
- [5] Ramsey J, Spirtes P, Zhang J. Adjacency-faithfulness and conservative causal inference. In: Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence. UAI'06. Arlington, Virginia, USA: AUAI Press; 2006. p. 401–8.
- [6] Shimizu S, Hoyer PO, Hyvärinen A, Kerminen A. A linear non-Gaussian acyclic model for causal discovery. J Mach Learn Res. 2006 Dec;7:2003–30.
- [7] Runge J, Bathiany S, Bollt E, Camps-Valls G, Coumou D, Deyle E, et al. Inferring causation from time series in earth system sciences. Nature Commun. 2019 Jun;10(1):2553. Number: 1 Publisher: Nature Publishing Group. <https://www.nature.com/articles/s41467-019-10105-3>.
- [8] Semedo JD, Gokcen E, Machens CK, Kohn A, Yu BM. Statistical methods for dissecting interactions between brain areas. Current Opinion Neurobiol. 2020 Dec;65:59–69. <https://www.sciencedirect.com/science/article/pii/S0959438820301367>.
- [9] Perich MG, Rajan K. Rethinking brain-wide interactions through multi-region “Network of Networks” models. Current Opinion Neurobiol. 2020 Dec;65:146–51. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7822595/>.
- [10] Runge J, Nowack P, Kretschmer M, Flaxman S, Sejdinovic D. Detecting and quantifying causal associations in large nonlinear time series datasets. Sci Adv. 2019;5(11):eaau4996. <https://www.science.org/doi/abs/10.1126/sciadv.aau4996>.
- [11] Runge J, Petoukhov V, Donges JF, Hlinka J, Jajcay N, Vejmelka M, et al. Identifying causal gateways and mediators in complex spatio-temporal systems. Nature Commun. 2015;6(1):1–10.
- [12] Wang C. Three-ocean interactions and climate variability: a review and perspective. Climate Dynamics. 2019 Oct;53(7):5119–36.
- [13] Costanza R, Kubiszewski I, Giovannini E, Lovins H, McGlade J, Pickett KE, et al. Development: time to leave GDP behind. Nature. 2014 Jan;505(7483):283–5. Number: 7483 Publisher: Nature Publishing Group. <https://www.nature.com/articles/505283a>.
- [14] Timmermann A, An SI, Kug JS, Jin FF, Cai W, Capotondi A, et al. El Niño–Southern oscillation complexity. Nature. 2018 Jul;559(7715):535–45. Number: 7715 Publisher: Nature Publishing Group. <https://www.nature.com/articles/s41586-018-0252-6>.
- [15] Rubenstein PK, Weichwald S, Bongers S, Mooij JM, Janzing D, Grosse-Wentrup M, et al. Causal consistency of structural equation models. In: Proceedings of the 33rd Conference on Uncertainty in Artificial Intelligence (UAI); 2017. p. ID 11. <http://auai.org/uai2017/proceedings/papers/11.pdf>.
- [16] Zhang W, Wang Z, Stuecker MF, Turner AG, Jin FF, Geng X. Impact of ENSO longitudinal position on teleconnections to the NAO. Climate Dynamics. 2019 Jan;52(1):257–74. doi: 10.1007/s00382-018-4135-1.
- [17] Wahl J, Ninad U, Runge J. Vector causal inference between two groups of variables. Proc AAAI Conference Artif Intelligence. 2023 Jun;37(10):12305–12. <https://ojs.aaai.org/index.php/AAAI/article/view/26450>.
- [18] Parviainen P, Kaski S. Learning structures of Bayesian networks for variable groups. Int J Approx Reasoning. 2017;88:110–27. <https://www.sciencedirect.com/science/article/pii/S0888613X17303134>.
- [19] Shah RD, Peters J. The hardness of conditional independence testing and the generalised covariance measure. Ann Stat. 2020;48(3):1514–38.
- [20] Josse J, Holmes SP. Measuring multivariate association and beyond. Stat Surveys. 2016;10:132–67.
- [21] Chatterjee S. A survey of some recent developments in measures of association. 2022. <https://arxiv.org/abs/2211.04702>.
- [22] Hochsprung T, Wahl J, Gerhardus A, Ninad U, Runge J. Increasing effect sizes of pairwise conditional independence tests between random vectors. In: Evans RJ, Shpitser I, editors. Proceedings of the Thirty-Ninth Conference on Uncertainty in Artificial Intelligence. vol. 216 of Proceedings of Machine Learning Research. PMLR; 2023. p. 879–89. <https://proceedings.mlr.press/v216/hochsprung23a.html>.
- [23] Anand TV, Ribeiro AH, Tian J, Bareinboim E. Causal effect identification in cluster DAGs. Proceedings of the AAAI Conference on Artificial Intelligence. 2023 Jun;37(10):12172–9. <https://ojs.aaai.org/index.php/AAAI/article/view/26435>.

- [24] Weinberger N. Faithfulness, coordination and causal coincidences. *Erkenntnis*. 2018 Apr;83(2):113–33.
- [25] Marx A, Gretton A, Mooij JM. A weaker faithfulness assumption based on triple interactions. In: *Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence*. vol. 161 of *Proceedings of Machine Learning Research*. PMLR; 2021. p. 451–60. <https://proceedings.mlr.press/v161/marx21a.html>.
- [26] Meek C. Causal inference and causal explanation with background knowledge. In: *Proceedings of the Eleventh conference on Uncertainty in Artificial Intelligence*. UAI'95. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.; 1995. p. 403–10. <https://dl.acm.org/doi/10.5555/2074158.2074204>.
- [27] Zscheischler J, Janzing D, Zhang K. Testing whether linear equations are causal: a free probability theory approach. *Proceedings of the 27th Conference on Uncertainty in Artificial Intelligence, UAI 2011*. 2012 Feb:839–48. doi: <https://dl.acm.org/doi/abs/10.5555/3020548.3020645>.
- [28] Runge J, Gerhardus A, Varando G, Eyring V, Camps-Valls G. Causal inference for time series. *Nature Reviews Earth Environ*. 2023;4:487–505. <https://www.nature.com/articles/s43017-023-00431-y>.
- [29] Glymour C, Zhang K, Spirtes P. Review of causal discovery methods based on graphical models. *Frontiers Genetics*. 2019;10:524. <https://www.frontiersin.org/articles/10.3389/fgene.2019.00524>.
- [30] Zhang J. Causal reasoning with ancestral graphs. *J Machine Learn Res*. 2008;9(47):1437–74. <http://jmlr.org/papers/v9/zhang08a.html>.
- [31] Forré P, Mooij JM. Markov properties for graphical models with cycles and latent variables. 2017. <https://arxiv.org/abs/1710.08775>.
- [32] Mooij JM, Claassen T. Constraint-based causal discovery using partial ancestral graphs in the presence of cycles. In: *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI)*. PMLR; 2020. p. 1159–68. <https://proceedings.mlr.press/v124/m-mooij20a.html>.
- [33] Bongers S, Foré P, Peters J, Mooij JM. Foundations of structural causal models with cycles and latent variables. *Ann Stat*. 2021;49(5):2885–915.
- [34] McConnell RM, De Montgolfier F. Linear-time modular decomposition of directed graphs. *Discrete Appl Math*. 2005;145(2):198–209.
- [35] Dawid AP. Conditional independence in statistical theory. *J R Stat Soc Ser B (Methodological)*. 1979;41(1):1–31. <http://www.jstor.org/stable/2984718>.
- [36] Runge J. Discovering contemporaneous and lagged causal relations in autocorrelated nonlinear time series datasets. In: *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI)*. vol. 124 of *Proceedings of Machine Learning Research*. PMLR; 2020. p. 1388–97. <https://proceedings.mlr.press/v124/runge20a.html>.
- [37] Gerhardus A, Runge J. High-recall causal discovery for autocorrelated time series with latent confounders. In: *Advances in neural information processing systems*. vol. 33. Curran Associates, Inc.; 2020. p. 12615–25.
- [38] Granger CWJ. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*. 1969;37(3):424–38. Publisher: [Wiley, Econometric Society]. <https://www.jstor.org/stable/1912791>.
- [39] Chalupka K, Eberhardt F, Perona P. Multi-level cause-effect systems. In: *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*. vol. 51 of *Proceedings of Machine Learning Research*. Cadiz, Spain: PMLR; 2016. p. 361–9. <https://proceedings.mlr.press/v51/chalupka16.html>.
- [40] Chalupka K, Eberhardt F, Perona P. Causal feature learning: an overview. *Behaviormetrika*. 2017;44(1):137–64.
- [41] Kalisch M, Bühlmann P. Estimating high-dimensional directed acyclic graphs with the PC-algorithm. *J Machine Learn Res*. 2007;8(22):613–36. <http://jmlr.org/papers/v8/kalisch07a.html>.



## Appendix

### A Group DMGs from group-valued SCMs

In this appendix, we will shortly discuss another way of obtaining a group DMG that is distinct from coarsening a graph of micro-variables, namely, by defining a model directly through structural equations. For a discussion of counterfactual distributions in vector-valued SCMs, see [23, Supplement, Theorem 7].

**Definition A1.** (Vector-valued SCMs) A *vector-valued structural causal model (vSCM)*  $\mathfrak{M} = (\mathfrak{S}, P_E)$  over a partition  $\mathcal{P}$  of a set of random variables  $\mathbf{X}$  into random vectors  $\mathbf{X}^1, \dots, \mathbf{X}^r$  is a collection of structural assignments

$$\mathbf{X}_i = \mathbf{f}^i(\text{pa}(\mathbf{X}^i), \mathbf{E}^i)$$

with  $\text{pa}(\mathbf{X}^i) \subset \{\mathbf{X}^1, \dots, \mathbf{X}^r\} \setminus \{\mathbf{X}^i\}$  and multivariate noise vectors  $\mathbf{E}^1, \dots, \mathbf{E}^r$  with  $\dim(\mathbf{X}^i) = \dim(\mathbf{E}^i)$  that have joint distribution  $P_E$ . The causal graph  $\mathcal{G}(\mathfrak{M})$  of  $\mathfrak{M}$  is the DMG with nodes  $\mathbf{X}^1, \dots, \mathbf{X}^r$  where a directed edge  $\mathbf{X}^j \rightarrow \mathbf{X}^i$  is drawn if  $\mathbf{X}^j \in \text{pa}(\mathbf{X}^i)$  and a bidirected edge  $\mathbf{X}^i \leftrightarrow \mathbf{X}^j$  is drawn if  $\mathbf{E}^i \perp\!\!\!\perp \mathbf{E}^j$ .<sup>4</sup>

While group DMGs derived by coarsening micro-variable graphs assume a causal structure on the level of the micro-variables and is then “forgotten” after coarsening, in a vector-valued SCM any causal meaning in the form of a graph is only defined on the group level. The internal relationships among the entries of a vector  $\mathbf{X}^i$  that are not due to external influences are modelled only by the distribution  $P_{\mathbf{E}^i}$  and are thus of a probabilistic nature. This seems reasonable for many practical applications where the micro-variables may not be considered causal entities (for instance, imagine  $\mathbf{X}^i$  to be a field of surface temperature measurements in some spatial region). On the other hand, vector-valued SCMs make it hard to *derive* faithfulness results from properties of the micro-variables, as no notion of faithfulness is purely distributional. Instead faithfulness can only be postulated as an assumption on the group level directly.

## B Proofs

### B.1 Proofs of the Results in Section 3

#### Proof of Lemma 1.

- (i) Assume first that  $\text{co}(\mathcal{G}, \mathcal{P})$  is acyclic and let  $\mathcal{W}$  be a strongly connected component. If there were  $W_1, W_2 \in \mathcal{W}$  that belonged to different groups of the partition  $\mathcal{P}$ , say  $W_1 \in \mathbf{Y}$  and  $W_2 \in \mathbf{Z}$ , then on  $\mathcal{G}$  we could find directed paths  $\pi_1$  from  $W_1$  to  $W_2$  and  $\pi_2$  from  $W_2$  to  $W_1$ . Then the induced coarse path  $\text{co}(\pi_1)$  would constitute a directed path from  $\mathbf{Y}$  to  $\mathbf{Z}$  and the induced coarse path  $\text{co}(\pi_2)$  would constitute a directed path from  $\mathbf{Z}$  to  $\mathbf{Y}$ . Concatenating both paths, we would obtain a cycle which contradicts our assumption.
- (ii) The converse is already wrong for coarsenings of micro DAGs in which the strongly connected components correspond to the nodes of the graph, see e.g. Figure 4.
- (iii) Let  $\mathcal{P}$  be the partition of  $\mathcal{G}$  into strongly connected component and let  $\tilde{\pi}$  be a directed path from  $\mathbf{Y}$  to  $\mathbf{Z}$ . Then, we argue first that for any two node  $Y \in \mathbf{Y}$ ,  $Z \in \mathbf{Z}$ , there is a directed micro path  $\pi$  from  $Y$  to  $Z$  on  $\mathcal{G}$ . Indeed, if  $\tilde{\pi}$  just consists of an edge  $\mathbf{Y} \rightarrow \mathbf{Z}$ , then there must be  $Y' \in \mathbf{Y}$  and  $Z' \in \mathbf{Z}$  that are connected by a micro edge  $Y' \rightarrow Z'$ . By the definition of strongly connected components, there must also be directed

<sup>4</sup> If the causal graph  $\mathcal{G}(\mathfrak{M})$  has cycles, it is not always true that random vectors  $\mathbf{X}^1, \dots, \mathbf{X}^r, \mathbf{E}^1, \dots, \mathbf{E}^r$  obeying the SCM actually exist. Solvability of cyclic SCMs is thoroughly discussed in Bongers et al. [33] for univariate node variables and the results straightforwardly transfer to the multivariate setting. The same is true for their analysis on Markov properties if mutual independence is used as an independence model.

paths from  $Y$  to  $Y'$  and from  $Z'$  to  $Z$ , so we have found the desired micro path. If  $\tilde{\pi}$  has more than one edge, we can proceed similarly by noting that for any motive  $\mathbf{W} \rightarrow \mathbf{Y} \rightarrow \mathbf{Z}$  there are micro edges  $W \rightarrow Y$ ,  $Y' \rightarrow Z$  with  $W \in \mathbf{W}$ ,  $Y, Y' \in \mathbf{Y}$ ,  $Z \in \mathbf{Z}$  and either  $Y = Y'$  or there is a directed path from  $Y$  to  $Y'$  as  $\mathbf{Y}$  is strongly connected. Concatenating all edges and paths found this way, we obtain the desired micro path. Finally, we conclude by observing that any cycle on  $\text{co}(\mathcal{G}, \mathcal{P})$  must thus induce a cycle on the micro  $\text{MG } \mathcal{G}$ . Indeed, a cycle on  $\text{co}(\mathcal{G}, \mathcal{P})$  could be decomposed into directed paths  $\tilde{\pi}_1$  and  $\tilde{\pi}_2$  one from say  $\mathbf{Y}$  to  $\mathbf{Z}$  and one from  $\mathbf{Z}$  to  $\mathbf{Y}$  to which we then apply the argument above.  $\square$

**Proof of Theorem 1.** Write  $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{B}, \mathcal{U})$  and  $\mathcal{G}^{\text{acy}} = (\hat{\mathcal{V}}, \hat{\mathcal{E}}, \hat{\mathcal{B}}, \hat{\mathcal{U}})$  and  $\mathcal{P} = \{\mathbf{X}^1, \dots, \mathbf{X}^r\}$ . We have to show that  $\text{co}(\mathcal{G}^{\text{acy}}, \mathcal{P})$  and  $\text{co}(\mathcal{G}, \mathcal{P})$  have the same directed, bidirected and undirected edges.

First, let  $\mathbf{X}^i \rightarrow \mathbf{X}^j$  be a directed edge on  $\text{co}(\mathcal{G}, \mathcal{P})$  so that there must exist a directed edge  $A \rightarrow B \in \mathcal{E}$  with  $A \in \mathbf{X}^i$  and  $B \in \mathbf{X}^j$ . Hence  $A \in \text{pa}_{\mathcal{G}}(B) \subset \text{pa}_{\mathcal{G}}(\text{sc}_{\mathcal{G}}(B))$  and we also see that  $A \notin \text{sc}_{\mathcal{G}}(B)$  by part (a) of Lemma 1 as  $\mathcal{P}$  was assumed acyclic with respect to  $\mathcal{G}$ . So by definition of acyclification, we get  $A \rightarrow B \in \hat{\mathcal{E}}$  and thus the edge  $\mathbf{X}^i \rightarrow \mathbf{X}^j$  is present on  $\text{co}(\mathcal{G}^{\text{acy}}, \mathcal{P})$ . On the other hand, if  $\mathbf{X}^i \rightarrow \mathbf{X}^j$  is a directed edge on  $\text{co}(\mathcal{G}^{\text{acy}}, \mathcal{P})$ , then there must be an edge  $A \rightarrow B \in \hat{\mathcal{E}}$  with  $A \in \mathbf{X}^i$  and  $B \in \mathbf{X}^j$ . Therefore,  $A \in \text{pa}_{\mathcal{G}}(\text{sc}_{\mathcal{G}}(B)) \setminus \text{sc}_{\mathcal{G}}(B)$ , so there must be a node  $C \in \text{sc}_{\mathcal{G}}(B)$  and an edge  $A \rightarrow C \in \mathcal{E}$ . By Lemma 1 (a), we obtain  $\text{sc}_{\mathcal{G}}(B) \subset \mathbf{X}^j$  so that there must be an edge  $\mathbf{X}^i \rightarrow \mathbf{X}^j$  on  $\text{co}(\mathcal{G}, \mathcal{P})$ .

We now turn to bidirected edges. If  $\mathbf{X}^i \leftrightarrow \mathbf{X}^j$  is a bidirected edge on  $\text{co}(\mathcal{G}, \mathcal{P})$ , then there exists a bidirected edge  $A \leftrightarrow B \in \mathcal{B}$  with  $A \in \mathbf{X}^i$  and  $B \in \mathbf{X}^j$ . By definition of acyclification, we also have  $A \leftrightarrow B \in \hat{\mathcal{B}}$ , so  $\mathbf{X}^i \leftrightarrow \mathbf{X}^j$  is a bidirected edge on  $\text{co}(\mathcal{G}^{\text{acy}}, \mathcal{P})$  as well. Finally assume that  $\mathbf{X}^i \leftrightarrow \mathbf{X}^j$  is a bidirected edge on  $\text{co}(\mathcal{G}^{\text{acy}}, \mathcal{P})$ , so that there exists a bidirected edge  $A \leftrightarrow B \in \hat{\mathcal{B}}$  with  $A \in \mathbf{X}^i$  and  $B \in \mathbf{X}^j$ . By acyclicity and Lemma 1A and  $B$  must lie in different strongly connected components of  $\mathcal{G}$ . Therefore, there must be  $A' \in \text{sc}_{\mathcal{G}}(A) \subset \mathbf{X}^i$  and  $B' \in \text{sc}_{\mathcal{G}}(B) \subset \mathbf{X}^j$  connected by a bidirected edge  $A' \leftrightarrow B' \in \mathcal{B}$ . We conclude that  $\mathbf{X}^i \leftrightarrow \mathbf{X}^j$  must be a bidirected edge on  $\text{co}(\mathcal{G}, \mathcal{P})$ . Finally, we discuss undirected edges. Thus assume first that  $\mathbf{X}^i - \mathbf{X}^j$  is an undirected edge on  $\text{co}(\mathcal{G}, \mathcal{P})$ , so that there must exist an undirected edge  $A - B \in \mathcal{U}$  with  $A \in \mathbf{X}^i$  and  $B \in \mathbf{X}^j$ . Since  $\mathcal{P}$  was assumed acyclic, we see that  $A \notin \text{sc}_{\mathcal{G}}(B)$  by part (a) of Lemma 1, so that there must be an undirected edge  $A - B \in \hat{\mathcal{U}}$ . Thus,  $\mathbf{X}^i - \mathbf{X}^j$  must be an undirected edge of  $\text{co}(\mathcal{G}^{\text{acy}}, \mathcal{P})$ . Conversely if  $\mathbf{X}^i - \mathbf{X}^j$  is assumed to be an undirected edge of  $\text{co}(\mathcal{G}^{\text{acy}}, \mathcal{P})$ , there must be an undirected edge  $A - B \in \hat{\mathcal{U}}$  with  $A \in \mathbf{X}^i$ ,  $B \in \mathbf{X}^j$ . By definition of acyclification, we must have  $A - B \in \mathcal{U}$  and thus,  $\mathbf{X}^i - \mathbf{X}^j$  must be an undirected edge of  $\text{co}(\mathcal{G}, \mathcal{P})$ . This finishes the proof.  $\square$

**Proof of Lemma 3.** We will only discuss the case where  $\text{co}(\pi)$  (and thus  $\pi$ ) is a non-trivial walk.

- (i) If  $\text{co}(\pi)$  is  $\sigma$ -blocked by  $S$ , then there are three options.
  - (1) If the first (or last) node of  $\text{co}(\pi)$  is in  $S$ , then  $\mathcal{T}$  must contain the first (or last) node of  $\pi$  and thus  $\sigma$ -blocks  $\pi$ .
  - (2) There is a collider  $\mathbf{W}$  on  $\text{co}(\pi)$  with  $S \cap \text{des}(\mathbf{W}) = \emptyset$ . We argue that in this case  $\mathbf{W}$  must contain a collider  $W$  of the micro walk  $\pi$ . Indeed, if  $\pi$  passes through only one node of  $\mathbf{W}$ , this follows directly. If  $\pi$  passes through more than one node,  $\pi$  must enter  $\mathbf{W}$  at a micro node  $\pi(i)$  with an edge pointing to  $\pi(i)$  (either bidirected or directed) and leave  $\mathbf{W}$  at a micro node  $\pi(j)$ ,  $j > i$ , again with an edge pointing to  $\pi(j)$  (either bidirected or directed). Thus at some point in the path segment  $\pi(i, j)$  the directionality of the arrows must oppose each other, that is to say that path segment must contain a collider, say  $\pi(l)$ . Any descendant  $D$  of  $\pi(l)$  must lie in  $\mathbf{W}$  itself or in a proper descendant of  $\mathbf{W}$ , say  $D \in \mathbf{D}$  as the directed path  $\pi(l) \rightarrow \dots \rightarrow D$  induces a coarse path  $\mathbf{W} \rightarrow \dots \rightarrow \mathbf{D}$ . As both  $\mathbf{W}$  and its proper descendants do not lie in  $S$ ,  $\text{des}(\pi(l)) \cap \mathcal{T} = \emptyset$ .
  - (3) There is a non-collider  $\text{co}(\pi)(k)$  on  $\text{co}(\pi)$  that is contained in  $S$  and an edge  $\text{co}(\pi)(k) \rightarrow \text{co}(\pi)(l)$ ,  $l \in \{k-1, k+1\}$  with  $\text{sc}_{\text{co}(\mathcal{G}, \mathcal{P})}(\text{co}(\pi)(k)) \neq \text{sc}_{\text{co}(\mathcal{G}, \mathcal{P})}(\text{co}(\pi)(l))$ . Therefore, on  $\pi$ , there must be an edge  $\pi(i) \rightarrow \pi(j)$ ,  $j \in \{i-1, i+1\}$  with  $\pi(i) \in \text{co}(\pi)(k)$  and  $\pi(j) \in \text{co}(\pi)(l)$ . Since  $\pi(i)$  has an outgoing edge it is a non-collider and by the contraposition of Lemma 2  $\text{sc}_{\mathcal{G}}(\pi(i)) \neq \text{sc}_{\mathcal{G}}(\pi(j))$ . Therefore  $\mathcal{T}$ , which contains  $\pi(i)$   $\sigma$ -blocks  $\pi$ .

- (ii) We use the following counterexample to show that the converse of (i) is not true. Let  $\mathcal{G}$  be given by  $W \rightarrow Y_1 \rightarrow Y_2 \leftarrow Y_3 \rightarrow Z$  partitioned as  $\mathbf{W} = \{W\}$ ,  $\mathbf{Y} = \{Y_1, Y_2, Y_3\}$ ,  $\mathbf{Z} = \{Z\}$ . Then the path from  $W$  to  $Z$  is closed since it contains the collider  $Y_2$  while the coarse path  $\mathbf{W} \rightarrow \mathbf{Y} \rightarrow \mathbf{Z}$  is open.
- (iii) If  $\pi$  is an arbitrary walk between  $Y \in \mathbf{Y}$  and  $Z \in \mathbf{Z}$ , then  $\text{co}(\pi)$  is a walk between  $\mathbf{Y}$  and  $\mathbf{Z}$  and thus  $\sigma$ -blocked by  $\mathcal{S}$ . Hence, by assertion (i),  $\pi$  is  $\sigma$ -blocked by  $\mathcal{T}$ .
- (iv) Let  $\mathcal{G}$  be given by  $W \rightarrow Y_1 - Y_2 - Y_3 \leftarrow Z$  partitioned as  $\mathbf{W} = \{W\}$ ,  $\mathbf{Y} = \{Y_1, Y_2, Y_3\}$ ,  $\mathbf{Z} = \{Z\}$ . Then the path  $\mathbf{W} \rightarrow \mathbf{Y} \leftarrow \mathbf{Z}$  is  $\sigma$ -blocked while the micro path from  $W$  to  $Z$  is  $\sigma$ -open as it does not contain any colliders.  $\square$

**Proof of Lemma 4.** For part (i), it suffices to note that the only difference between the two types of separation lies in their definition for non-colliders, so only part (i)(3) of the proof of Lemma 3 slightly differs. When a coarse walk  $\text{co}(\pi)$  has a non-collider, say  $\text{co}(\pi)(k)$ , then  $\pi$  must have a non-collider  $\pi(j) \in \text{co}(\pi)(k)$  as well. So if  $\text{co}(\pi)(k) \in \mathcal{S}$ , then  $\pi(j) \in \mathcal{T}$  and  $\pi$  is  $m$ -blocked. Part (iii) follows directly from (i) and the counterexamples of parts (ii) and (iv) do not involve cycles and are equally valid for  $m$ -separation.  $\square$

## B.2 Proofs of the Results in Section 4

**Proof of Lemma 6.** We prove Lemma 6 by induction over  $n = |\mathbf{Y} \cup \mathbf{Z}|$ . For  $n = 2$ , we must have  $\mathbf{Y} = \{Y\}$ ,  $\mathbf{Z} = \{Z\}$  and the result follows by choosing the subset  $\mathcal{M} = \emptyset$ . Now assume that the result has been shown for some arbitrary but fixed  $n \geq 2$  and let  $|\mathbf{Y} \cup \mathbf{Z}| = n + 1$ . W.l.o.g. we can assume that  $|\mathbf{Z}| > 1$ . Let  $Y \in \mathbf{Y}$  and  $Z \in \mathbf{Z}$  be arbitrary. Choosing  $\mathcal{M} = \mathbf{Z} \setminus \{Z\}$ , by assumption we have  $Y \perp\!\!\!\perp Z | \mathbf{W}, \mathbf{Z} \setminus \{Z\}$ . According to Lemma 5, we are done if we can show that also  $Y \perp\!\!\!\perp \mathbf{Z} \setminus \{Z\} | \mathbf{W}$ . To prove this, we observe first that  $|\mathbf{Y} \cup \mathbf{Z} \setminus \{Z\}| = n$ . Moreover, for any  $Y \in \mathbf{Y}$ ,  $Z' \in \mathbf{Z} \setminus \{Z\}$  and  $\mathcal{M}' \subset \mathbf{Y} \cup \mathbf{Z} \setminus \{Z\} \setminus \{Y, Z'\}$ , we have  $Y \perp\!\!\!\perp Z' | \mathbf{W}, \mathcal{M}'$ . Thus, the induction hypothesis implies that  $Y \perp\!\!\!\perp \mathbf{Z} \setminus \{Z\} | \mathbf{W}$  and in particular  $Y \perp\!\!\!\perp \mathbf{Z} \setminus \{Z\} | \mathbf{W}$  as desired.  $\square$

**Proof of Lemma 7.** We only have to prove that conditional pairwise independence implies conditional mutual independence. So let  $\mathbf{Y}, \mathbf{Z}$  be pairwise independent given  $\mathcal{T}$ . Iterating part (ii) of Lemma 5, mutual independence follows if we can show that for any pair  $Y \in \mathbf{Y}, Z \in \mathbf{Z}$  and any subset  $\mathcal{M} \subset \mathbf{Y} \cup \mathbf{Z}$ , we also have  $Y \perp\!\!\!\perp \mathbf{Z} | \mathcal{M} \cup \mathcal{T}$ . By  $\sigma$ -faithfulness of  $(\mathcal{G}, P_{\mathbf{X}})$ , any pair  $Y \in \mathbf{Y}, Z \in \mathbf{Z}$  is  $\sigma$ -separated by  $\mathcal{T}$ , that is all micro-paths leading from the group  $\mathbf{Y}$  to the group  $\mathbf{Z}$  are  $\sigma$ -blocked by  $\mathcal{T}$ . If we can show that all micro-paths are still  $\sigma$ -blocked by  $\mathcal{M} \cup \mathcal{T}$ , the result follows by the  $\sigma$ -Markov property of  $(\mathcal{G}, P_{\mathbf{X}})$ . Since all such micro-paths are  $\sigma$ -blocked by  $\mathcal{T}$ , we only need to make sure that none of these paths is opened again by adding  $\mathcal{M}$  to the separating set. Suppose there was such a path  $\pi$  starting at  $\pi(1) \in \mathbf{Y}$  and ending at  $\pi(r) \in \mathbf{Z}$ , that is  $\sigma$ -blocked by  $\mathcal{T}$  but  $\sigma$ -unblocked by  $\mathcal{T} \cup \mathcal{M}$ . Let  $\pi(k)$  be the last node of  $\pi$  in  $\mathbf{Y}$  and let  $\pi(l), k < l$  be the first node of  $\pi$  in  $\mathbf{Z}$  after  $k$ . Since  $\mathcal{T} \subset \mathbf{X} \setminus \{\mathbf{Y}, \mathbf{Z}\}$ , the subpath  $\pi'$  of  $\pi$  starting at  $\pi'(1) = \pi(k)$  and ending at  $\pi'(s) = \pi(l)$  must still be  $\sigma$ -blocked by  $\mathcal{T}$ . On the other hand, since  $\mathcal{M} \cup \mathcal{T}$   $\sigma$ -unblocks  $\pi$ , it must  $\sigma$ -unblock  $\pi'$ . Thus, there must be at least one collider on  $\pi'$  that has a descendant in  $\mathcal{M}$ . Let  $\pi'(i)$  be the last such collider on  $\pi'$  with descendant  $D \in \mathcal{M}$ . If  $\pi'(i) = D$ , then  $\pi'(i) \in \mathbf{Y} \cup \mathbf{Z}$  contradicting the fact that  $\pi'$  does not have inner nodes in  $\mathbf{Y} \cup \mathbf{Z}$ . Therefore,  $D$  must be a proper descendant of  $\pi'(i)$ . As  $D \in \mathcal{M}$ , we have in particular  $D \in \mathbf{Y} \cup \mathbf{Z}$ , and we will assume w.l.o.g.  $D \in \mathbf{Y}$ . Let  $\pi'' = \pi(i) \rightarrow \dots \rightarrow D$  be the descending micro-path and assume that  $\pi''(j)$  is the first node that belongs to  $\mathcal{M}$ . But then, the concatenation of  $\pi''(j) \leftarrow \dots \leftarrow \pi'(i)$  and  $\pi'(i, s)$  leads from  $\mathbf{Y}$  to  $\mathbf{Z}$  and is  $\sigma$ -unblocked by  $\mathcal{T}$  contradicting our assumption that all such micro-paths must be  $\sigma$ -blocked by  $\mathcal{T}$ .  $\square$

**Proof of Theorem 2.** We assume that  $P_{\mathbf{X}}$  does not have the strong  $\sigma$ -Markov property and show that this leads to a contradiction. Since  $(\text{co}(\mathcal{G}, \mathcal{P}), P_{\mathbf{X}})$  does not have the  $\sigma$ -Markov property, there must be vectors  $\mathbf{Y}, \mathbf{Z}$  that are  $\sigma$ -separated by a set of vectors  $\mathcal{S}$  but not mutually conditionally independent given  $\mathcal{S}$ . By Lemma 3, for every pair  $Y \in \mathbf{Y}, Z \in \mathbf{Z}$ , all paths on  $\mathcal{G}$  between  $Y$  and  $Z$  are  $\sigma$ -blocked by  $\mathcal{T}$ . On the other hand, using Lemma 6, we see that there must be  $Y' \in \mathbf{Y}, Z' \in \mathbf{Z}$  and a subset  $\mathcal{M} \subset \mathbf{Y} \cup \mathbf{Z}$  such that

$$Y' \not\perp\!\!\!\perp Z' | \mathcal{M}, \mathcal{T}.$$

By the  $\sigma$ -Markov property on the micro DMG  $\mathcal{G}$  this means that there must be a path  $\pi$  between  $Y'$  and  $Z'$  on  $\mathcal{G}$  that is not  $\sigma$ -blocked by  $\mathcal{M}, \mathcal{T}$ , but is  $\sigma$ -blocked by  $\mathcal{T}$ . Therefore,  $Y \notin S$  and  $Z \notin S$  as otherwise we would have  $Y' \in \mathcal{T}$  or  $Z' \in \mathcal{T}$ . Moreover,  $\mathcal{T}$  cannot contain any non-collider  $\pi(l)$  of  $\pi$  pointing to a neighbor  $\pi(l \pm 1)$  in a different strongly connected component,  $\pi$  must have at least one collider and any collider on  $\pi$  must have a descendant in  $\mathcal{M}$ . Our goal is now to construct a path  $\tilde{\pi}$  on  $\text{co}(\mathcal{G}, \mathcal{P})$  that is not  $\sigma$ -blocked by  $S$  resulting in a contradiction. Consider first the coarsened path  $\text{co}(\pi)$  of  $\pi$ .  $S$  can not contain any non-colliders of  $\text{co}(\pi)$  pointing to a neighbor in a different strongly connected component of  $\text{co}(\mathcal{G}, \mathcal{P})$  by Lemma 2. If  $\text{co}(\pi)$  does not have any colliders it must be  $\sigma$ -unblocked by  $S$  and we are done. Therefore, assume that  $\text{co}(\pi)$  does contain colliders. If all such colliders would have a descendant in  $S$ , again the path would be  $\sigma$ -unblocked by  $S$  and the desired contradiction would be obtained. Thus, we can assume that at least one collider on  $\text{co}(\pi)$  does not have any descendants in  $S$ . Any collider  $C$  of  $\text{co}(\pi)$  must contain a micro-collider  $C$  of the micro-path  $\pi$  and by the considerations above there must be a directed micro-path  $C \rightarrow \dots \rightarrow M$  for some  $M \in \mathcal{M} \subset Y \cup Z$ . Coarsening this micro-path we see that for any collider  $C$  of  $\text{co}(\pi)$  there must be a directed macro-path from  $C$  to  $Y$  or to  $Z$ . Writing out  $\text{co}(\pi) = (\text{co}(\pi)(1), \dots, \text{co}(\pi)(r))$  where  $\text{co}(\pi)(1) = Y$  and  $\text{co}(\pi)(r) = Z$ , we define the sets

$$U = \{k \mid \text{co}(\pi)(k) \in \text{col}(\text{co}(\pi)), S \cap \text{des}(\text{co}(\pi)(k)) = \emptyset, \text{ and } Z \in \text{des}(\text{co}(\pi)(k))\}$$

and

$$U' = \{k \mid \text{co}(\pi)(k) \in \text{col}(\text{co}(\pi)), \text{des}(\text{co}(\pi)(k)) = \emptyset, \text{ and } Y \in \text{des}(\text{co}(\pi)(k))\}.$$

By the aforementioned considerations, at least one of these sets must be non-empty. If  $U$  is non-empty, let  $k'$  be its minimum, so that  $\text{co}(\pi)(k)$  is the collider closest to  $Y$ . Thus, the subpath  $\text{path}(\text{co}(\pi)(1), \dots, \text{co}(\pi)(k'))$  must be right-directed, i.e.  $\text{co}(\pi)(1) \rightarrow \dots \rightarrow \text{co}(\pi)(k')$ . Since  $k' \in U$ , we can join it with a directed path from  $\text{co}(\pi)(k')$  to  $Z$  yielding a path  $Y \rightarrow \dots \rightarrow Z$ . This path can not be  $\sigma$ -blocked by  $S$  as all of its nodes are either non-colliders of  $\text{co}(\pi)$  or descendants of  $\text{co}(\pi)(k')$ . Thus we have found the desired path. If  $U'$  is non-empty, the argument is analogous with  $k'' = \max U'$  instead of  $k'$  and left-directed instead of right-directed paths.  $\square$

**Proof of Theorem 3.** Up to a few subtleties, the proof is similar to Theorem 2. We assume that  $P_X$  that  $(\text{co}(\mathcal{G}, \mathcal{P}), P_X)$  does not have the  $m$ -Markov property. We show that this leads to a contradiction. Since  $(\text{co}(\mathcal{G}, \mathcal{P}), P_X)$  does not have the  $m$ -Markov property, there must be vectors  $Y, Z$  that are  $m$ -separated by a set of vectors  $S$  but not mutually conditionally independent given  $S$ . Hence, by Lemma 4, for every pair  $Y \in Y, Z \in Z$ , all paths on  $\mathcal{G}$  between  $Y$  and  $Z$  are  $m$ -blocked by  $\mathcal{T}$ . On the other hand, using Lemma 6, we see that there must be  $Y' \in Y, Z' \in Z$  and a subset  $\mathcal{M} \subset Y \cup Z$  such that

$$Y' \not\perp\!\!\!\perp Z' \mid \mathcal{M}, \mathcal{T}.$$

By the  $m$ -Markov property on  $\mathcal{G}$  this means that there must be a path  $\pi$  between  $Y'$  and  $Z'$  on  $\mathcal{G}$  that is not  $m$ -blocked by  $\mathcal{M}, \mathcal{T}$ , but is  $m$ -blocked by  $\mathcal{T}$ . Therefore,  $\mathcal{T}$  cannot contain any non-colliders of  $\pi$ ,  $\pi$  must have at least one collider and any collider on  $\pi$  must have a descendant in  $\mathcal{M}$ . Our goal is now to construct a path  $\tilde{\pi}$  on  $\text{co}(\mathcal{G}, \mathcal{P})$  that is not  $m$ -blocked by  $S$  resulting in a contradiction. Consider first the coarsened path  $\text{co}(\pi)$  of  $\pi$ .  $S$  can not contain any non-colliders of  $\text{co}(\pi)$  as otherwise  $\mathcal{T}$  would contain a non-collider of  $\pi$ . So if  $\text{co}(\pi)$  does not have any colliders it must be  $m$ -open given  $S$  and we are done. So assume that  $\text{co}(\pi)$  does contain colliders. If all such colliders would have a descendant in  $S$ , again the path would be  $m$ -opened by  $S$  and the desired contradiction would be obtained. Thus, assume that at least one collider on  $\text{co}(\pi)$  does not have any descendants in  $S$ . Any collider  $C$  of  $\text{co}(\pi)$  must contain a micro-collider  $C$  of the micro-path  $\pi$  and by the considerations above there must be a directed micro-path  $C \rightarrow \dots \rightarrow M$  for some  $M \in \mathcal{M} \subset Y \cup Z$ . Coarsening this micro-path we see that for any collider  $C$  of  $\text{co}(\pi)$  there must be a directed macro-path from  $C$  to  $Y$  or to  $Z$ . Writing out  $\text{co}(\pi) = (\text{co}(\pi)(1), \dots, \text{co}(\pi)(r))$  where  $\text{co}(\pi)(1) = Y$  and  $\text{co}(\pi)(r) = Z$ , we define the sets

$$U = \{k \mid \text{co}(\pi)(k) \in \text{col}(\text{co}(\pi)), S \cap \text{des}(\text{co}(\pi)(k)) = \emptyset, \text{ and } Z \in \text{des}(\text{co}(\pi)(k))\}$$

and

$$U' = \{k \mid \text{co}(\pi)(k) \in \text{col}(\text{co}(\pi)), \text{des}(\text{co}(\pi)(k)) = \emptyset, \text{ and } Y \in \text{des}(\text{co}(\pi)(k))\}.$$

By the aforementioned considerations, at least one of these sets must be non-empty. If  $U$  is non-empty, let  $k'$  be its minimum, so that  $\text{co}(\pi)(k)$  is the collider closest to  $\mathbf{Y}$ . Thus, the subpath  $\text{co}(\pi)(1), \dots, \text{co}(\pi)(k')$  must be right-directed, i.e.  $\text{co}(\pi)(1) \rightarrow \dots \rightarrow \text{co}(\pi)(k')$ . Since  $k' \in U$ , we can join it with a directed path from  $\text{co}(\pi)(k')$  to  $\mathbf{Z}$  yielding a path  $\mathbf{Y} \rightarrow \dots \rightarrow \mathbf{Z}$ . This path can not be  $m$ -blocked by  $S$  as all of its nodes are either non-colliders of  $\text{co}(\pi)$  or descendants of  $\text{co}(\pi)(k')$ . Thus, we have found the desired path. If  $U'$  is non-empty, the argument is analogous with  $k'' = \max U'$  instead of  $k'$  and left-directed instead of right-directed paths.  $\square$

### B.3 Proofs of the Results in Section 5

**Proof of Lemma 8.** Assume first that  $(\text{co}(\mathcal{G}, \mathcal{P}), P_X)$  is  $\sigma$ -faithful and let be  $\mathbf{Y}$  and  $\mathbf{Z}$   $\sigma$ -connected by a set  $S \subset \mathcal{P}$ . Therefore by assumption  $\mathbf{Y} \not\perp\!\!\!\perp \mathbf{Z} | \mathcal{T}$ . By Lemma 7 and our assumptions, conditional pairwise independence of groups implies conditional mutual independence. Applying the logical contraposition, we thus obtain  $\mathbf{Y} \not\perp\!\!\!\perp^{pw} \mathbf{Z} | \mathcal{T}$ . Thus there must be  $Y \in \mathbf{Y}$  and  $Z \in \mathbf{Z}$  such that  $Y \not\perp\!\!\!\perp Z | \mathcal{T}$ . By  $\sigma$ -Markovianity of  $(\mathcal{G}, P_X)$ ,  $Y$  and  $Z$  must be  $\sigma$ -connected by  $\mathcal{T}$ .

Conversely, assume that whenever  $\mathbf{Y}$  and  $\mathbf{Z}$  are  $\sigma$ -connected by a set  $S \subset \mathcal{P}$ . Then there exist  $Y \in \mathbf{Y}$  and  $Z \in \mathbf{Z}$  that are  $\sigma$ -connected by  $\mathcal{T} = \bigcup_{W \in S} W$ . By  $\sigma$ -faithfulness of  $(\mathcal{G}, P_X)$  it follows that  $Y \not\perp\!\!\!\perp Z | \mathcal{T}$  and thus  $\mathbf{Y} \not\perp\!\!\!\perp \mathbf{Z} | \mathcal{T}$ . Thus,  $(\text{co}(\mathcal{G}, \mathcal{P}), P_X)$  is  $\sigma$ -faithful.  $\square$

**Proof of Theorem 4.** Let  $\mathbf{W}, \mathbf{Y} \in \mathcal{P}$  and assume that  $\Pi = (\Pi(1), \mathbf{e}_1, \Pi(2), \mathbf{e}_2, \dots, \mathbf{e}_{n-1}, \Pi(n))$  is a path on  $\text{co}(\mathcal{G}, \mathcal{P})$  with  $\Pi(1) = \mathbf{W}$ ,  $\Pi(n) = \mathbf{Y}$  that is  $\sigma$ -unblocked by  $S \subset \mathcal{P}$ . Let  $\mathcal{T} = \bigcup_{Z \in S} \mathbf{Z}$ . According to Corollary 1, we need to construct a path  $\pi$  that coarsens to  $\Pi$  and that is  $\sigma$ -unblocked by  $\mathcal{T}$ . Consider the edge  $\mathbf{e}_1$  and choose  $e_1 \in \text{mic}(\mathbf{e}_1)$  whose nodes we will immediately denote by  $\pi(1) \in \Pi(1)$  and  $\pi(2) \in \Pi(2)$ , i.e.  $e_1 = (\pi(1), \pi(2))$ . For  $i = 2, \dots, n-1$ , we proceed inductively as follows. Assume that we have already defined a path  $(\pi(1), e_1, \dots, e_{s-1}, \pi(s))$  such that  $\pi(s-1) \in \Pi(i-1)$ ,  $\pi(s) \in \Pi(i)$  and  $e_{s-1} \in \text{mic}(\mathbf{e}_{i-1})$ . By assumption (iii) of the theorem, we can find a node  $Y \in \text{bd}_{\mathbf{e}_i}(\Pi(i)) \subset \Pi(i)$  such that  $\text{sc}(\pi(s)) = \text{sc}(Y)$ . If  $\mathbf{e}_{i-1}$  is left-directed and  $\mathbf{e}_i$  is left-directed or bidirected, choose a left-directed path  $\xi_i = (\xi_i(1) = \pi(s), \tilde{e}_1, \dots, \tilde{e}_m, \xi_i(m) = Y)$ , in all other cases, choose a right-directed path  $\xi_i = (\xi_i(1) = \pi(s), \tilde{e}_1, \dots, \tilde{e}_m, \xi_i(m) = Y)$ . Note that by condition (ii), all nodes of  $\xi_i$  must remain in  $\Pi(i)$ . Concatenate  $\xi_i$  with  $\pi$ , i.e. set  $e_{s-1+j} = \tilde{e}_j$ ,  $\pi(s-1+j) = \xi_i(j)$ ,  $j = 1, \dots, m$ . Finally since  $Y \in \text{bd}_{\mathbf{e}_i}(\Pi(i))$  we can choose an edge  $e_{s+m} \in \text{mic}(\mathbf{e}_i)$  connecting  $Y = \pi(s+m)$  to some  $\pi(s+m+1) \in \Pi(i+1)$ . If  $i+1 = n$ , we have finished the construction of our micro-path  $\pi$  and by construction  $\text{co}(\pi) = \Pi$ . We need to show now that  $\mathcal{T} \sigma$ -unblocks  $\pi$ . There are different cases to check.

- Assume that  $\pi(1) \in \mathcal{T}$  (respectively  $\pi(n) \in \mathcal{T}$ ). In this case, we must have  $\Pi(1) \in S$  (or  $\Pi(n) \in S$ ) which would  $\sigma$ -block  $\Pi$ , contrary to our assumption. Thus,  $\pi(1), \pi(n) \notin \mathcal{T}$ .
- Assume that  $\pi(k)$  is a collider on  $\pi$  for some  $1 < k < \text{len}(\pi)$  and that  $\pi(k) \in \Pi(i)$ .
  - The first case to discuss here is  $|\Pi(i)| = 1$ , i.e.  $\Pi(i) = \{\pi(k)\}$ . In this case, either  $\Pi(i) \in S$  in which case  $\pi(k) \in \mathcal{T}$  or  $\Pi(i)$  must have a proper descendant  $S \in S$ . Then similar to the aforementioned construction of  $\pi$ , using (ii) and (iii), we can also construct a descending path  $\pi(k) \rightarrow \dots \rightarrow S$  for some  $S \in S$ . Thus,  $\pi(k)$  has a descendant in  $\mathcal{T}$  and again the collider  $\pi(k)$  is  $\sigma$ -unblocked.
  - The second case is  $|\Pi(i)| > 1$ . Because of our choice of the internal path  $\xi_i$  as directed,  $\pi(k) \in \text{bd}_{\mathbf{e}_i}$  or  $\pi(k) \in \text{bd}_{\mathbf{e}_{i-1}}$ . We will only discuss the first case  $\pi(k) \in \text{bd}_{\mathbf{e}_i}$  as the second one is completely analogous. If  $\pi(k) \in \text{bd}_{\mathbf{e}_i}$ , then  $\pi(k+1) \in \Pi(i+1)$  and the edge  $e_k = (\pi(k), \pi(k+1))$  must be left- or bidirected as  $\pi(k)$  is a collider. Again because of the way we chose  $\xi_i$ , the unique edge on  $\pi$  in  $\text{mic}(\mathbf{e}_{i-1})$  must be right- or bidirected. Thus, both  $\mathbf{e}_{i-1}$  and  $\mathbf{e}_i$  must have an arrowhead towards  $\Pi(i)$ , i.e.  $\Pi(i)$  is a collider on  $\Pi$ . Thus,  $\Pi(i)$  must have a descendant in  $S$ . Suppose first that this descendant is  $\Pi(i)$  itself. Then the collider  $\pi(k)$  is  $\sigma$ -unblocked as it is contained in  $\mathcal{T}$ . The other nodes on  $\pi$  that are part of  $\Pi(i)$  are non-collider but do not point to neighbors on  $\pi$  that are part of a different strongly connected component. Thus, by condition (3) in the definition of  $\sigma$ -separation (Definition 2) including them in  $\mathcal{T}$  does not  $\sigma$ -block  $\pi$ . Next, suppose that the descendant of  $\Pi(i)$  in  $S$  is a proper descendant. Once again, by using (ii) and (iii), we can construct a descending path  $\pi(k) \rightarrow \dots \rightarrow S$  for some  $S \in S$  so that the collider  $\pi(k)$  is unblocked.



- Assume that  $\pi(k)$  is a non-collider on  $\pi$  for some  $1 < k < \text{len}(\pi)$  and that  $\pi(k) \in \Pi(i)$ . As  $S\sigma$ -unblocks  $\Pi$ , we must be in one of the following situations. Either (I)  $\Pi(i) \notin S$  or (II)  $\Pi(i) \in S$  but if  $\mathbf{e}_{i-1} = \Pi(i-1) \leftarrow \Pi(i)$  or  $\mathbf{e}_i = \Pi(i) \rightarrow \Pi(i+1)$  then  $\text{sc}(\Pi(i-1)) = \text{sc}(\Pi(i))$ , respectively  $\text{sc}(\Pi(i)) = \text{sc}(\Pi(i+1))$ .
  - (I) In this case,  $\pi(k) \notin \mathcal{T}$  thus the non-collider  $\pi(k)$  is  $\sigma$ -unblocked by  $\mathcal{T}$ .
  - (II) If  $\Pi(i) \in S$  suppose that  $\mathbf{e}_i = \Pi(i) \rightarrow \Pi(i+1)$  has a tail at  $\Pi(i)$ . As stated above, the fact that  $S\sigma$ -unblocks  $\Pi$  means that we must have  $\text{sc}(\Pi(i+1)) = \text{sc}(\Pi(i))$ . If  $\pi(k+1)$  is also an element of  $\Pi(i)$ , then by construction of  $\pi$ , it has a right-directed edge  $\pi(k) \rightarrow \pi(k+1)$  and  $\text{sc}(\pi(k)) = \text{sc}(\pi(k+1))$ . Thus, even though  $\pi(k) \in \mathcal{T}$ , it is still  $\sigma$ -unblocked by  $\mathcal{T}$ . Thus we can assume that  $\pi(k+1) \in \Pi(i+1)$  which means in particular that  $\pi(k) \in \text{bd}_{\mathbf{e}_i}(\Pi(i))$  and  $\pi(k+1) \in \text{bd}_{\mathbf{e}_i}(\Pi(i+1))$ . As  $\text{sc}(\Pi(i+1)) = \text{sc}(\Pi(i))$ , there exists a directed path  $\Gamma$  on  $\text{co}(\mathcal{G}, \mathcal{P})$  starting at  $\Pi(i+1)$  and ending at  $\Pi(i)$ . As with the construction of  $\pi$ , because of the boundary connection condition (iii), we can once again construct a micro-path  $\gamma$  from  $\pi(k+1)$  to  $\pi(k)$  so that  $\text{sc}(\pi(k)) = \text{sc}(\pi(k+1))$ . So, once again even though  $\pi(k) \in \mathcal{T}$ , it is still  $\sigma$ -unblocked by  $\mathcal{T}$ . The final case is that  $\mathbf{e}_i$  does not have a tail at  $\Pi(i)$  which means that  $\mathbf{e}_{i-1} = \Pi(i-1) \leftarrow \Pi(i)$  must have one. The argument that  $\text{sc}(\pi(k)) = \text{sc}(\pi(k-1))$  is then completely parallel to the discussion for tailed  $\mathbf{e}_i$ , taking into account that the path segment  $\xi_i$  of  $\pi$  that is internal to  $\Pi(i)$  is left-directed (or trivial) by construction. Therefore, also in this case, even though  $\pi(k) \in \mathcal{T}$ , it is still  $\sigma$ -unblocked by  $\mathcal{T}$ .

We have shown earlier that every collider of  $\pi$  has a descendant in  $\mathcal{T}$  and that every non-collider is either not part of  $\mathcal{T}$  or points exclusively to neighbors in the same connected component. In summary,  $\mathcal{T}$   $\sigma$ -unblocks  $\pi$ , so  $\sigma$ -faithfulness is proven.

To show that  $\text{co}(\mathcal{G}, \mathcal{P})$  is acyclic, assume that there exists a right-directed cycle  $\Pi$  with  $\Pi(1) = \Pi(\text{len}(\Pi))$ . W.l.o.g., we can assume that  $\Pi$  is irreducible. Let  $\mathcal{P}'$  be the partition of  $\mathcal{G}$  into strongly connected components and  $\mathcal{H} = \text{co}(\mathcal{G}, \mathcal{P}')$ , which is always acyclic. Then condition (iii) implies that  $\Pi$  induces a right-directed walk  $I_0$  on  $\mathcal{H}$  with no repeating middle vertices such that  $I_0(1), I_0(\text{len}(I_0)) \subset \Pi(1)$ . If  $I_0(1) = I_0(\text{len}(I_0))$ , we have found a cycle in  $\mathcal{H}$  and thus a contradiction. If not, we can again use condition (iii) to construct a right-directed walk  $I'$  with  $I'(1) = I_0(\text{len}(I_0))$  and  $I'(\text{len}(I')) \subset \Pi(1)$ . Concatenating  $I_0$  and  $I'$  to  $I_1 = I_0 \circ I'$ , we have found two walks now that start in the same strongly connected component  $I_0(1)$  and end in  $\mathcal{P}' \cap 2^{\Pi(1)}$ . Again if  $I_1(\text{len}(I_1)) = I_1(1)$ , we are done, otherwise we continue to construct walks  $I_k$  on  $\mathcal{H}$  in this manner. Since the set  $\mathcal{P}' \cap 2^{\Pi(1)}$  is finite, at some point, the condition  $I_k(\text{len}(I_k)) = I_k(1)$  must be met and we arrive at a contradiction.  $\square$

**Proof of Theorem 5.** Let  $\mathbf{W}, \mathbf{Y} \in \mathcal{P}$  and assume that  $\Pi = (\Pi(1), \mathbf{e}_1, \Pi(2), \mathbf{e}_2, \dots, \mathbf{e}_{n-1}, \Pi(n))$  is a path on  $\text{co}(\mathcal{G}, \mathcal{P})$  with  $\Pi(1) = \mathbf{W}$ ,  $\Pi(n) = \mathbf{Y}$  that is  $\sigma$ -unblocked by  $S \subset \mathcal{P}$ . Let  $\mathcal{T} = \bigcup_{Z \in S} \mathbf{Z}$ . According to Corollary 1, we need to construct a path  $\pi$  that coarsens to  $\Pi$  and that is  $\sigma$ -unblocked by  $\mathcal{T}$ . We proceed as in the proof of Lemma 4 and for non-colliders  $(\mathbf{e}_i, \Pi(i+1), \mathbf{e}_{i+1})$ , the construction of  $\pi$  completely parallels that proof. For a collider  $(\mathbf{e}_i, \Pi(i+1), \mathbf{e}_{i+1})$ , we simply can choose the micro path  $\xi_i$  in the argument to be a micro-collider  $(\xi_i(1), e_1^i, \xi_i(2), e_2^i, \xi_i(3))$  with  $\xi_i(1) \in \Pi(i-1)$ ,  $\xi_i(2) \in \Pi(i)$ ,  $\xi_i(3) \in \Pi(i+1)$ . Showing that  $\mathcal{T}$  unblocks  $\pi$  again completely parallels the the proof of Lemma 4, noting that if  $\pi(k)$  is a non-collider on  $\pi$  it must lie in a group  $\Pi(i)$  that is a non-collider of  $\Pi$ . If  $\pi(k)$  is a collider on  $\pi$ , then  $\pi(k) \in \Pi(i)$  and  $\Pi(i)$  must be a collider on  $\Pi$ . Since  $S\sigma$ -unblocks  $\Pi(i)$ , we must either have  $\Pi(i) \in S$  or  $\Pi(i)$  must have a proper descendant in  $S$ , say  $\mathbf{W}$ . In the first case, it follows that  $\pi(k) \in \mathcal{T}$ . In the second case, consider the descending path  $\Gamma = (\Pi(i), \tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_s, \mathbf{W})$ . We can again construct a descending micro-path  $\gamma$  starting at some  $\gamma(1) \in \text{bd}_{\tilde{\mathbf{e}}_1}(\Pi(i))$  to some  $W \in \mathbf{W}$ . Since the pair  $\mathbf{e}_i, \tilde{\mathbf{e}}_1$  is an (almost) mediator, we can use condition (iii-a) to find a directed micro-path from  $\pi(k)$  to  $\gamma(1)$  and hence to  $W$ . Thus,  $\pi(k)$  has a descendant in  $\mathcal{T}$  as desired.  $\square$