

Faithfulness of Probability Distributions and Graphs

Kayvan Sadeghi

*Statistical Laboratory
University of Cambridge
Centre for Mathematical Sciences, Wilberforce Road
Cambridge, CB3 0WB, United Kingdom*

K.SADEGHI@STATSLAB.CAM.AC.UK

Editor: Peter Spirtes

Abstract

A main question in graphical models and causal inference is whether, given a probability distribution P (which is usually an underlying distribution of data), there is a graph (or graphs) to which P is faithful. The main goal of this paper is to provide a theoretical answer to this problem. We work with general independence models, which contain probabilistic independence models as a special case. We exploit a generalization of ordering, called preordering, of the nodes of (mixed) graphs. This allows us to provide sufficient conditions for a given independence model to be Markov to a graph with the minimum possible number of edges, and more importantly, necessary and sufficient conditions for a given probability distribution to be faithful to a graph. We present our results for the general case of mixed graphs, but specialize the definitions and results to the better-known subclasses of undirected (concentration) and bidirected (covariance) graphs as well as directed acyclic graphs.

Keywords: causal discovery, compositional graphoid, directed acyclic graph, faithfulness, graphical model selection, independence model, Markov property, mixed graph, structural learning

1. Introduction

Graphs have been used in graphical models in order to capture the conditional independence structure of probability distributions. Generally speaking, nodes of the graph correspond to random variables and missing edges to conditional independencies (Lauritzen, 1996). The connection between graphs and probability distributions is usually established in the literature by the concept of Markov property (Clifford, 1990; Pearl, 1988; Studený, 1989), which ensures that if there is a specific type of separation between node subsets A and B of the graph “given the node subset C ” then random vectors X_A and X_B are conditionally independent given the random vector X_C in the probability distribution. However, the “ultimate” connection between probability distributions and graphs requires the other direction to hold, namely for every conditional independence in the probability distribution to correspond to a separation in the graph. This connection has been called faithfulness of the probability distribution and the graph in Spirtes et al. (2000), and the graph has been called the perfect map of such a distribution in Pearl (1988).

However, “given a probability distribution P whether there is a graph (or graphs) to which P is faithful” is an open problem, and consequently so is the problem of finding

these graphs. This problem can be raised for any type of graph existing in the literature of graphical models ranging from different types of mixed graphs with three types of edges (Richardson and Spirtes, 2002; Wermuth, 2011; Sadeghi, 2016) and different types of chain graphs (Lauritzen and Wermuth, 1989; Frydenberg, 1990; Andersson et al., 2001; Cox and Wermuth, 1993; Wermuth, 2011) to better-known classes of undirected (concentration) (Darroch et al., 1980) and bidirected (covariance) (Kauermann, 1996) graphs as well as directed acyclic graphs (Kiiveri et al., 1984; Pearl, 1988). Our goal is to provide an answer to this problem. A similar problem of “given a graph whether there is a family of distributions faithful to it” has been answered for very specific types of graphs and specific types of distributions; for example, for Gaussian distributions and undirected graphs in Lněnička and Matúš (2007), DAGs in Pearl (1988); Geiger et al. (1990); Spirtes et al. (2000), ancestral graphs in Richardson and Spirtes (2002), and LWF chain graphs in Peña (2011); and discrete distributions and DAGs in Geiger et al. (1990); Meek (1995) and LWF chain graphs in Peña (2009).

The concept of faithfulness was originally defined for the purpose of causal inference (Spirtes et al., 2000; Pearl, 1988), and the theory developed in this paper can be interpreted in causal language. A main approach to causal inference is based on graphical representations of causal structures, usually represented by *causal graphs* that are directed acyclic with nodes as random variables (that is a Bayesian network). Causal graphs are assumed to capture the true causal structure; see, for example, Neapolitan (2004). A main assumption made is called the *causal faithfulness condition* (CFC) stating that the true probability distribution is faithful to the true causal graph. Although the distribution being faithful to the “true underlying causal graph” is a much stronger assumption, it necessarily means that it must be faithful to “some graph” in order for one to be able to use graphical methods for causal inference. For an extensive discussion on the CFC, see Zhang and Spirtes (2008), and for related philosophical discussions, see, for example, Woodward (1998); Steel (2006). Although the results in this paper cover those of causal Bayesian networks, we have developed our theory to a much more general case of simultaneous representation of “direct effects”, “confounding”, and “non-causal symmetric dependence structures”; see, for example, Pearl (2009).

In this paper, we work with the general independence model \mathcal{J} . Independence models $\mathcal{J}(P)$ induced by probability distributions P are a special case. We provide necessary and sufficient conditions for \mathcal{J} to be graphical, that is to be faithful to a graph. In short, as proved in Theorem 17, \mathcal{J} is graphical if and only if it satisfies the so-called compositional graphoid axioms as well as singleton-transitivity, and what we call ordered upward- and downward-stability.

As apparent from their names, ordered upward- and downward-stability depend on a generalization of ordering of variables, and consequently the nodes of the graph (called pre-ordering). We provide our results for the most general case of mixed graphs, which contains almost all classes of graphs used in graphical models as subclasses with the exception of the chain graphs with the AMP Markov property (Andersson et al., 2001) and its generalizations (Peña, 2014). However, based on the preordering w.r.t. which ordered upward- and downward-stability are satisfied, one can deduce to what type (or types) of graph P is faithful.

Since many of the subclasses of mixed graphs are not as well-known or used as some simpler classes of graphs, we provide a specialization of definitions and the main results for undirected and bidirected graphs as well as directed acyclic graphs at the end of the paper. We advise the readers that are only familiar with or interested in these simpler subclasses to skip the general definitions and results for mixed graphs, and focus on the specialization.

The structure of the paper is as follows: In the next section, we provide the basic definitions for graphs as well as different classes of graphs in graphical models, and independence models. In Section 3, we define and provide basic results for Markovness and faithfulness of independence models and graphs. In Section 4, we define and exploit the preordering for (so-called anterital) graphs, and then we define the concept of upward- and downward-stability. In Section 5, we provide sufficient conditions for an independence model to be so-called minimally Markov to a graph, and then we provide necessary and sufficient conditions for an independence model to be faithful to a graph. In Section 6, we specialize the definitions and results to the classes of undirected and bidirected graphs as well as directed acyclic graphs. In Section 7, we show how other results in the literature concerning faithfulness of certain probability distributions and certain classes of graphs are corollaries of our results. This also provides a nice set of examples for the theory presented in the paper. In Section 8, we end the paper with a short summary, a discussion on the statistical implications of the results, and an outline of future work.

2. Definitions and Concepts

In this section, we provide the basic definitions and concepts needed for the paper.

2.1 Graph Theoretical Definitions and Concepts

A *graph* G is a triple consisting of a *node* set or *vertex* set V , an *edge* set E , and a relation that with each edge associates two nodes (not necessarily distinct), called its *endpoints*. When nodes i and j are the endpoints of an edge, they are *adjacent*. We call a node adjacent to the node i , a *neighbour* of i , and denote the set of all neighbours of i by $\text{ne}(i)$. We say that an edge is *between* its two endpoints. We usually refer to a graph as an ordered pair $G = (V, E)$. Graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ are called *equal* if $(V_1, E_1) = (V_2, E_2)$. In this case we write $G_1 = G_2$.

Notice that our graphs are *labeled*, that is every node is considered a different object. Hence, for example, graph $i — j — k$ is not equal to $j — i — k$.

A *loop* is an edge with the same endpoints. Here we only discuss graphs without loops. *Multiple edges* are edges with the same pair of endpoints. A *simple graph* has neither loops nor multiple edges.

Graphs in this paper are so-called *mixed graphs*, which contain three types of edges: *lines* ($i — j$), *arrows* ($i → j$), and *arcs* ($i ↗ j$), where in the two latter cases, we say that there is an *arrowhead* at j . If we remove all arrowheads from edges of the graph G , the obtained undirected graph is called the *skeleton* of G . The simple graph whose edge ij indicates whether there is an edge (or multiple edges) between i and j in G is called the *adjacency graph* of G . It is clear that if a graph is simple then its adjacency graph is the same as its skeleton. In this paper, only skeleton of anterital graphs or its subclasses are used, which, as will be defined later, are simple graphs; hence, we denote the simple

skeleton by $\text{sk}(G)$. We say that we *direct* the edges of a skeleton by putting arrowheads at the edges in order to obtain mixed graphs.

A *subgraph* of a graph G_1 is graph G_2 such that $V(G_2) \subseteq V(G_1)$ and $E(G_2) \subseteq E(G_1)$ and the assignment of endpoints to edges in G_2 is the same as in G_1 . We define a subgraph *induced by edges* $A \subseteq E$ of $G = (V, E)$ to be a subgraph that contains V as the node set and all and only edges in A as the edge set.

A *walk* is a list $\langle v_0, e_1, v_1, \dots, e_k, v_k \rangle$ of nodes and edges such that for $1 \leq i \leq k$, the edge e_i has endpoints v_{i-1} and v_i . A *path* is a walk with no repeated node or edge. A maximal set of nodes in a graph whose members are connected by some paths constitutes a *connected component* of the graph. A *cycle* is a walk with no repeated nodes or edges except for $v_0 = v_k$.

We say a walk is *between* the first and the last nodes of the list in G . We call the first and the last nodes *endpoints* of the walk and all other nodes *inner nodes*.

A *subwalk* of a walk $\omega = \langle i_0, e_1, i_1, \dots, e_n, i_n \rangle$ is a walk $\langle i_r, e_{r+1}, i_{r+1}, \dots, e_p, i_p \rangle$ that is a subsequence of ω between two occurrences of nodes $(i_r, i_p, 0 \leq r \leq p \leq n)$. If a subwalk forms a path then it is called a *subpath* of ω .

A walk $\omega = \langle i = i_0, i_1, \dots, i_n = j \rangle$ is *directed* from i to j if all edges $i_k i_{k+1}$, $0 \leq k \leq n-1$, are arrows pointing from i_k to i_{k+1} . If there is a directed walk from i to j then i is an *ancestor* of j and j is a *descendant* of i . We denote the set of ancestors of j by $\text{an}(j)$.

A walk $\omega = \langle i = i_0, i_1, \dots, i_n = j \rangle$ is *semi-directed* from i to j if it has at least one arrow, no arcs, and every arrow $i_k i_{k+1}$ is pointing from i_k to i_{k+1} . A walk between i and j is *anterior* from i to j if it is semi-directed from i to j or if it is undirected. If there is an anterior walk from i to j then we also say that i is an *anterior* of j . We use the notation $\text{ant}(j)$ for the set of all antecedents of j . For a set A , we define $\text{ant}(A) = \bigcup_{j \in A} \text{ant}(j) \setminus A$.

Notice that, unlike in most places in the literature (for example Richardson and Spirtes 2002), we use walks instead of paths to define ancestors and antecedents. Because of this and the fact that ancestral graphs have no arrowheads pointing to lines, our definition of anterior extends the notion of anterior for ancestral graphs in Richardson and Spirtes (2002) with the modification that in this paper, a node is not an anterior of itself. Using walks instead of paths is immaterial, as shown in Lauritzen and Sadeghi (2017).

A *section* ρ of a walk is a maximal subwalk consisting only of lines, meaning that there is no other subwalk that only consists of lines and includes ρ . Thus, any walk decomposes uniquely into sections; these are not necessarily edge-disjoint and sections may also be single nodes. A section ρ on a walk ω is called a *collider section* if one of the following walks is a subwalk of ω : $i \rightarrow \rho \leftarrow j$, $i \leftarrow \rho \leftarrow j$, $i \leftarrow \rho \leftrightarrow j$. All other sections on ω are called *non-collider* sections. Notice that a section may be a collider on one walk and a non-collider on another, but we may speak of collider or non-collider sections without mentioning the relevant walk when this is apparent from the context.

2.2 Different Classes of Graphs

All the subclasses of graphs included here are *acyclic* in the sense that they do not contain semi-directed cycles. The most general class of graphs discussed here is the class of *chain mixed graphs* (CMGs) (Sadeghi, 2016) that contains all mixed graphs without semi-directed cycles. They may have multiple edges containing an arc and a line or an arc and an

arrow but not a combination of an arrow and a line or arrows in opposite directions, as such combinations would constitute semi-directed cycles. In this paper, in using the term “graph” we mean a CMG unless otherwise stated.

A general class of graphs that plays an important role in this paper is the class of *anterial graphs* (AnGs) (Sadeghi, 2016). AnGs are CMGs in which an endpoint of an arc cannot be an anterior of the other endpoint.

CMGs include *summary graphs* (SGs) (Wermuth, 2011) and *acyclic directed mixed graphs* (ADMGs) (Richardson, 2003) as a subclass, but not include *AMP chain graphs* (Andersson et al., 2001), and AnGs include *ancestral graphs* (AGs) (Richardson and Spirtes, 2002) but not SGs or ADMGs. These have typically been introduced to describe independence structures obtained by marginalization and conditioning in DAG independence models; see for example Sadeghi (2013).

AnGs also contain *undirected and bidirected chain graphs* (CGs). A chain graph is an acyclic graph so that if we remove all arrows, all connected components of the resulting graph—called *chain components*—contain one type of edge only. If all chain components contain lines, the chain graph is an undirected chain graph (UCG) (known as *LWF chain graphs*); if all chain components contain arcs it is a bidirected chain graph (BCG) (known as *multivariate regression chain graphs*). AnGs also contain *Regression graphs* (Wermuth and Sadeghi, 2012), which are chain graphs consisting of lines and arcs (although dashed undirected edges have mostly been used instead of arcs in the literature), where there is no arrowhead pointing to lines.

These also contain graphs with only one type of edge; namely *undirected graphs* (UGs), containing only lines; *bidirected graphs* (BGs), containing only bidirected edges; and *directed acyclic graphs* (DAGs), containing only arrows and being acyclic. Clearly, a graph without arrows has no semi-directed cycles, and a semi-directed cycle in a graph with only arrows is a directed cycle. Cox and Wermuth (1993); Kauermann (1996); Wermuth and Cox (1998); Drton and Richardson (2008) used the terms *concentration graphs* and *covariance graphs* for UGs and BGs, referring to their independence interpretation associated with covariance and concentration matrices for Gaussian graphical models. DAGs have been particularly useful to describe causal Markov relations; see for example Kiiveri et al. (1984); Pearl (1988); Lauritzen and Spiegelhalter (1988); Geiger et al. (1990); Spirtes et al. (2000).

For an extensive discussion on the subclasses of acyclic graphs and their relationships and hierarchy, see Lauritzen and Sadeghi (2017).

2.3 Independence Models and Their Properties

An *independence model* \mathcal{J} over a finite set V is a set of triples $\langle X, Y | Z \rangle$ (called *independence statements*), where X , Y , and Z are disjoint subsets of V ; Z may be empty, but $\langle \emptyset, Y | Z \rangle$ and $\langle X, \emptyset | Z \rangle$ are always included in \mathcal{J} . The independence statement $\langle X, Y | Z \rangle$ is read as “ X is independent of Y given Z ”. Independence models may in general have a probabilistic interpretation, but not necessarily. Similarly, not all independence models can be easily represented by graphs. For further discussion on general independence models, see Studený (2005).

In order to define probabilistic independence models, consider a set V and a collection of random variables $\{X_\alpha\}_{\alpha \in V}$ with state spaces \mathcal{X}_α , $\alpha \in V$ and joint distribution P . We let

$X_A = \{X_v\}_{v \in A}$ etc. for each subset A of V . For disjoint subsets A , B , and C of V we use the short notation $A \perp\!\!\!\perp B | C$ to denote that X_A is *conditionally independent of X_B given X_C* (Dawid, 1979; Lauritzen, 1996), that is that for any measurable $\Omega \subseteq \mathcal{X}_A$ and P -almost all x_B and x_C ,

$$P(X_A \in \Omega | X_B = x_B, X_C = x_C) = P(X_A \in \Omega | X_C = x_C).$$

We can now induce an independence model $\mathcal{J}(P)$ by letting

$$\langle A, B | C \rangle \in \mathcal{J}(P) \text{ if and only if } A \perp\!\!\!\perp B | C \text{ w.r.t. } P.$$

Similarly we use the notation $A \not\perp\!\!\!\perp B | C$ for $\langle A, B | C \rangle \notin \mathcal{J}(P)$.

In order to use graphs to represent independence models, the notion of *separation* in a graph is fundamental. For three disjoint subsets A , B , and C , we use the notation $A \perp B | C$ if A and B are *separated* given C , and $A \not\perp B | C$ for A and B not separated given C .

The separation can be unified under one definition for all graphs by using walks instead of paths: A walk π is *connecting* given C if every collider section of π has a node in C and all non-collider sections are disjoint from C . For pairwise disjoint subsets (A, B, C) , $A \perp B | C$ if there are no connecting walks between A and B given C . This is in wording the same as the definition in Studený and Bouckaert (1998) for undirected (LWF) chain graphs (although it is in fact a generalization as collider sections are generalized), and a generalization of the m -*separation* used with different wordings in Richardson and Spirtes (2002); Wermuth (2011); Wermuth and Sadeghi (2012) for AGs and SGs, and of the d -*separation* of Pearl (1988). For UGs, the notion has a direct intuitive meaning, so that $A \perp B | C$ if all paths from A to B pass through C .

If A , B , or C has only one member $\{i\}$, $\{j\}$, or $\{k\}$, for better readability, we write $\langle i, j | k \rangle \in \mathcal{J}$ instead of $\langle \{i\}, \{j\} | \{k\} \rangle \in \mathcal{J}$; and similarly for $i \perp j | k$ and $i \not\perp j | k$. We also write $A \perp B$ when $C = \emptyset$; and similarly $A \perp\!\!\!\perp B$.

A graph G induces an independence model $\mathcal{J}(G)$ by separation, letting

$$\langle A, B | C \rangle \in \mathcal{J}(G) \iff A \perp B | C \text{ in } G.$$

An independence model \mathcal{J} over a set V is a *semi-graphoid* if it satisfies the four following properties for disjoint subsets A , B , C , and D of V :

1. $\langle A, B | C \rangle \in \mathcal{J}$ if and only if $\langle B, A | C \rangle \in \mathcal{J}$ (*symmetry*);
2. if $\langle A, B \cup D | C \rangle \in \mathcal{J}$ then $\langle A, B | C \rangle \in \mathcal{J}$ and $\langle A, D | C \rangle \in \mathcal{J}$ (*decomposition*);
3. if $\langle A, B \cup D | C \rangle \in \mathcal{J}$ then $\langle A, B | C \cup D \rangle \in \mathcal{J}$ and $\langle A, D | C \cup B \rangle \in \mathcal{J}$ (*weak union*);
4. if $\langle A, B | C \cup D \rangle \in \mathcal{J}$ and $\langle A, D | C \rangle \in \mathcal{J}$ then $\langle A, B \cup D | C \rangle \in \mathcal{J}$ (*contraction*).

Notice that the reverse implication of contraction clearly holds by decomposition and weak union. A semi-graphoid for which the reverse implication of the weak union property holds is said to be a *graphoid*; that is, it also satisfies

5. if $\langle A, B | C \cup D \rangle \in \mathcal{J}$ and $\langle A, D | C \cup B \rangle \in \mathcal{J}$ then $\langle A, B \cup D | C \rangle \in \mathcal{J}$ (*intersection*).

Furthermore, a graphoid or semi-graphoid for which the reverse implication of the decomposition property holds is said to be *compositional*, that is, it also satisfies

6. if $\langle A, B \mid C \rangle \in \mathcal{J}$ and $\langle A, D \mid C \rangle \in \mathcal{J}$ then $\langle A, B \cup D \mid C \rangle \in \mathcal{J}$ (*composition*).

Separation in graphs satisfies all these properties; see Theorem 1 in Lauritzen and Sadeghi (2017):

Proposition 1 *For any graph G , the independence model $\mathcal{J}(G)$ is a compositional graphoid.*

On the other hand, probabilistic independence models are always semi-graphoids (Pearl, 1988), whereas the converse is not necessarily true; see Studený (1989). If, for example, P has strictly positive density, the induced independence model is always a graphoid; see, for example, Proposition 3.1 in Lauritzen (1996). See also Peters (2015) for a necessary and sufficient condition for the intersection property to hold. If the distribution P is a regular multivariate Gaussian distribution, $\mathcal{J}(P)$ is a compositional graphoid; for example see Studený (2005). Probabilistic independence models with positive densities are not in general compositional; this only holds for special types of multivariate distributions such as, for example, Gaussian distributions and the symmetric binary distributions used in Wermuth et al. (2009).

Another important property that is satisfied by separation in all graphs, but not necessarily for probabilistic independence models, is *singleton-transitivity* (also called *weak transitivity* in Pearl (1988), where it is shown that for Gaussian and binary distributions P , $\mathcal{J}(P)$ always satisfies it). For i , j , and k , single elements in V ,

7. if $\langle i, j \mid C \rangle \in \mathcal{J}$ and $\langle i, j \mid C \cup \{k\} \rangle \in \mathcal{J}$ then $\langle i, k \mid C \rangle \in \mathcal{J}$ or $\langle j, k \mid C \rangle \in \mathcal{J}$ (*singleton-transitivity*).

A singleton-transitive compositional graphoid is equivalent to what is called a *Gaussoid* in Lněnička and Matúš (2007) with a rather different axiomatization. The name reminds one that these are the axioms satisfied by the independence model of a regular Gaussian distribution (Pearl, 1988).

Proposition 2 *For a graph G , $\mathcal{J}(G)$ satisfies singleton-transitivity.*

Proof If there is a walk between i and k given C and a walk between j and k given C then by connecting these two walks, we obtain a walk π between i and j . Except the section ρ that contains k , all collider sections on π have a node in C and all non-collider sections are outside C . Hence depending on ρ being a collider or non-collider, π is connecting given $C \cup \{k\}$ or C respectively. ■

3. Markov and Faithful Independence Models

In this section, we define and discuss the concepts of Markovness and faithfulness for probability distributions, independence models, and graphs.

3.1 Markov Properties

For a graph $G = (V, E)$, an independence model \mathcal{J} defined over V satisfies the *global Markov property* w.r.t. G , or is simply *Markov* to G , if for disjoint subsets A , B , and C of V it holds that

$$A \perp B \mid C \implies \langle A, B \mid C \rangle \in \mathcal{J},$$

or equivalently $\mathcal{J}(G) \subseteq \mathcal{J}$. In particular, the graphical independence $\mathcal{J}(G)$ is trivially Markov to its graph G .

For a probability distribution P , we simply say P is *Markov* to a graph G if $\mathcal{J}(P)$ is Markov to G , that is if $\mathcal{J}(G) \subseteq \mathcal{J}(P)$. Notice that every independence model over V is Markov to the complete graph with the node set V .

For a graph $G = (V, E)$, an independence model \mathcal{J} defined over V satisfies a *pairwise Markov property* w.r.t. a graph G , or is simply *pairwise Markov* to G , if for every non-adjacent pair of nodes i, j , it holds that $\langle i, j \mid C(i, j) \rangle \in \mathcal{J}$, for some $C(i, j)$, where $C(i, j)$ is the *conditioning set* of the pairwise Markov property and does not include i and j .

The independence model $\mathcal{J}(G)$ only satisfies pairwise Markov properties w.r.t. what is called *maximal graphs*, which are graphs where the lack of an edge between i and j corresponds to a conditional separation statement for i and j . For a probability distribution P , we say P is *pairwise Markov* to a graph G if $\mathcal{J}(P)$ is pairwise Markov to G .

For maximal graphs, a pairwise Markov property is defined by letting $C(i, j) = \text{ant}(i) \cup \text{ant}(j) \setminus \{i, j\}$ (Lauritzen and Sadeghi, 2017), which we henceforth use as “the” pairwise Markov property. The conditioning set of the pairwise Markov property simplifies for DAGs, $C(i, j) = \text{an}(i) \cup \text{an}(j) \setminus \{i, j\}$ (Richardson and Spirtes, 2002; Sadeghi and Lauritzen, 2014); for BGs, $C(i, j) = \emptyset$ (as defined in Wermuth and Cox 1998); and for connected UGs, $C(i, j) = V \setminus \{i, j\}$ (as defined in Lauritzen 1996).

This pairwise Markov property, is in fact equivalent to the global Markov property under compositional graphoids; see Theorem 4 of Lauritzen and Sadeghi (2017):

Proposition 3 *Let G be a maximal graph. If the independence model \mathcal{J} is a compositional graphoid, then \mathcal{J} is pairwise Markov to G if and only if \mathcal{J} is Markov to G .*

In particular, for UGs, the equivalence of the global and pairwise Markov properties holds under graphoids, and for BGs under compositional semi-graphoids.

Two undirected or two bidirected graphs G and H , where $G \neq H$, induce different independence models; that is $\mathcal{J}(G) \neq \mathcal{J}(H)$. This is not necessarily true for larger subclasses. We call two graphs G and H such that $\mathcal{J}(G) = \mathcal{J}(H)$ *Markov equivalent*. Conditions for Markov equivalence for most subclasses of graphs are known; see Verma and Pearl (1990); Ali et al. (2009); Wermuth and Sadeghi (2012). Notice that two Markov equivalent maximal graphs have the same skeleton.

3.2 Faithfulness and Minimal Markovness

We say that an independence model \mathcal{J} and a probability distribution P are *faithful* if $\mathcal{J} = \mathcal{J}(P)$. Similarly, we say that \mathcal{J} and a graph G are faithful if $\mathcal{J} = \mathcal{J}(G)$. We also say that P and G are faithful if $\mathcal{J}(P) = \mathcal{J}(G)$. If P and G are faithful then we may sometimes also say that P is *faithful to G* or vice versa, although in principle faithfulness is a symmetric relation; the same holds for saying that \mathcal{J} is faithful to G or P . Notice that

we are extending the definition of faithfulness to include the relation between independence models and graphs as well as independence models and probability distributions rather than only that between graphs and probability distributions as originally used in Spirtes et al. (2000).

Thus, if \mathcal{J} and G are faithful then \mathcal{J} is Markov to G , but it also requires every independence statement to correspond to a separation in G . We say that \mathcal{J} or G is *probabilistic* if there is a distribution P that is faithful to \mathcal{J} or G respectively. If there is a graph that is faithful to \mathcal{J} or P then we say that \mathcal{J} or P is *graphical*.

Our main goal is to characterize graphical probability distributions, and in addition, if existent, to provide graphs that are faithful to a given P . We solve this problem for the general case of independence models \mathcal{J} .

For a given independence model \mathcal{J} , we define the *skeleton* of \mathcal{J} , denoted by $\text{sk}(\mathcal{J})$, to be the undirected graph that is obtained from \mathcal{J} as follows: we define the node set of $\text{sk}(\mathcal{J})$ to be V , and for every pair of nodes i, j , we check whether $\langle i, j \mid C \rangle \in \mathcal{J}$ holds for some $C \subseteq V \setminus \{i, j\}$; if it does not then we draw an edge between i and j . One can similarly define $\text{sk}(P)$ for a probability distribution P by checking whether $i \perp j \mid C$. This is the same as the graph obtained by the first step of the SGS algorithm (Glymour et al., 1987). Notice that this is not necessarily the same as the undirected graph obtained by the pairwise Markov property for UGs; see Section 6.1.

For UGs and BGs, the skeleton of \mathcal{J} uniquely determines the graph, whereas for other subclasses, there are several graphs with the same skeleton $\text{sk}(\mathcal{J})$. As will be seen, the preordering, defined in the next section, enables us to direct the edges of the skeleton.

Proposition 4 *If an independence model \mathcal{J} is Markov to a graph G then $\text{sk}(\mathcal{J})$ is a subgraph of $\text{sk}(G)$.*

Proof Suppose that there is no edge between i and j in $\text{sk}(G)$. Since the global Markov property w.r.t. G implies the pairwise Markov property, it holds that $\langle i, j \mid C(i, j) \rangle \in \mathcal{J}$, which implies i is not adjacent to j in $\text{sk}(\mathcal{J})$. ■

Hence, if \mathcal{J} is Markov to a graph G such that $\text{sk}(G) = \text{sk}(\mathcal{J})$ then G has the fewest number of edges among those to which \mathcal{J} is Markov. We say that \mathcal{J} is *minimally Markov* to a graph G if \mathcal{J} is Markov to G and $\text{sk}(G) = \text{sk}(\mathcal{J})$. The same can be defined for probability distributions, and has been used in the literature under the name of *(causal) minimality assumption* (Pearl, 2009; Spirtes et al., 2000; Zhang and Spirtes, 2008; Neapolitan, 2004). Minimally Markov independence models to a graph are important since only these can also be faithful to the graph:

Proposition 5 *If an independence model \mathcal{J} and a graph G are faithful then \mathcal{J} is minimally Markov to G .*

Proof Since \mathcal{J} is Markov to G , we need to prove that $\text{sk}(G) = \text{sk}(\mathcal{J})$. By Proposition 4, $\text{sk}(\mathcal{J})$ is a subgraph of $\text{sk}(G)$. Now, suppose that there is no edge between i and j in $\text{sk}(\mathcal{J})$. By the construction of $\text{sk}(\mathcal{J})$, it holds that $\langle i, j \mid C \rangle \in \mathcal{J}$ for some C . Since \mathcal{J} and G are faithful, $i \perp j \mid C$ in G . This implies that i is not adjacent to j in G . ■

Hence, we need to discuss conditions under which \mathcal{J} is minimally Markov to a graph as well as conditions for faithfulness of minimally Markov independence models and graphs. These will be presented in Section 5.

4. Preordering in Graphs and Ordered Stabilities

In this section, we first define the preordering for sets and its validity for nodes of the graphs, and then use preordering to define some new properties of conditional independence.

4.1 Preordering of the Nodes of a Graph

Over a set V , a *partial order* is a binary relation \leq that satisfies the following properties for all members $a, b, c \in V$:

- $a \leq a$ (*reflexivity*);
- if $a \leq b$ and $b \leq a$ then $a = b$ (*antisymmetry*);
- if $a \leq b$ and $b \leq c$ then $a \leq c$ (*transitivity*).

If $a \leq b$ or $b \leq a$ then a and b are *comparable*; otherwise they are *incomparable*. A partial order under which every pair of elements is comparable is called a *total order* or a *linear order*.

A *preorder* \lesssim over V , on the other hand, is a binary relation that is only reflexive and transitive. Given a preorder \lesssim over V , one may define an equivalence relation \sim on V such that $a \sim b$ if and only if $a \lesssim b$ and $b \lesssim a$. This explains the use of the notation \lesssim . It then holds that

- if $a \lesssim b$ and $b \sim c$ then $a \lesssim c$; and
- if $a \lesssim b$ and $a \sim c$ then $c \lesssim b$.

We also use the notation $a < b$ indicating that $a \lesssim b$ and $a \not\sim b$.

We first provide the following result; see Section 5.2.1 of Schröder (2003):

Proposition 6 *Let \lesssim be a preorder over the set V , and \sim the equivalence relation on V as defined above. Let also Q be the set of all equivalence classes of V w.r.t. \sim . Then the relation \leq defined over Q by $[[a]] \leq [[b]] \iff a \lesssim b$ is a partial order over Q .*

We say that a graph $G = (V, E)$ admits a *valid preorder* \lesssim if, for nodes i and j of G , the following holds:

- if $i \rightarrow j$ then $i \sim j$;
- if $i \rightarrow j$ then $j < i$;
- if $i \not\rightarrow j$ then i and j are incomparable.

The global interpretation of a valid preorder on graphs is as follows:

Proposition 7 *Let \lesssim be a valid preorder for a graph G . It then holds for nodes i and j of G that if $i \in \text{ant}(j)$ then $j \lesssim i$; in particular,*

1. if there is a semi-directed path from i to j then $j < i$; and
2. if i and j are connected by a path consisting only of lines then $i \sim j$.

Proof The proof follows from transitivity of both preorders and anterior paths. \blacksquare

Corollary 8 Let \lesssim be a valid preorder for a graph G and G_- its subgraph induced by lines. The equivalence classes of G based on \lesssim are the connected components of G_- . In addition, \leq defines a partial order over the connected components of G_- by [if $i \rightarrow j$, $i \in \tau$ and $j \in \delta$ then $\tau > \delta$], for τ and δ two connected components of G_- .

Proof The results follows from Propositions 7 and 6. \blacksquare

For example, consider the graph in Fig. 1. The graph admits a valid preorder with the preorder provided over the nodes, or equivalently over the connected components of G_- (with the abuse of the notation for numbering). The notation implies that two nodes with label 2 are in the same equivalence class, $2 > 1$ and $2' > 1'$, but $a, b',$ and $c'', a, b, c \in \{1, 2\}$, are not comparable.

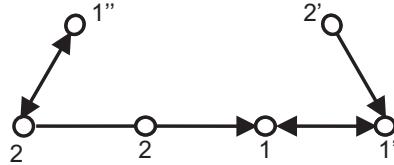


Figure 1: A preordered graph.

There may be many different preorders that are valid for a graph. If there is a valid preordering and we expect the other direction of Proposition 7 to hold then we obtain a unique preordering for the graph: Given a graph G , if $i \notin \text{ant}(j)$ and $j \notin \text{ant}(i)$ then set i and j to be incomparable. Otherwise, let $i \sim j$ when $i \rightarrow j$, and $i > j$ when $i \rightarrow j$. It is easy to see that this, in fact, is a preorder for the nodes of G . We call this preordering the *minimal preorder* for G since it gives the fewest possible comparable pairs of nodes. For example, the preordering in the graph of Fig. 1 is minimal. In this paper, we mostly deal with this type of preorderings for graphs. It is easy to observe the following:

Proposition 9 If G is anterial then the minimal preorder for G is a valid preorder for G .

In fact, in general, we have the following:

Proposition 10 A graph admits a valid preorder if and only if it is anterial.

Proof If a graph G admits a valid preorder then by Proposition 7, there cannot be a directed cycle, nor can there be an arc with one endpoint anterior of the other. The converse is Proposition 9. \blacksquare

Therefore, in addition to CMGs, SGs (and ADMGs) do not admit a valid preorder, but AGs do. However, notice that for every CMG, there exists a Markov equivalent AnG (see Sadeghi 2016); and the same for SGs and AG. Hence, the question of whether there is a CMG that is faithful to \mathcal{J} is the same as whether there is an AnG that is faithful to \mathcal{J} ; and the same holds for SG and AG.

As discussed above, there is a one-to-one correspondence between AnGs and the minimal preorder for the AnGs. When the skeleton of the concerned graph is known, we have the following trivial result:

Proposition 11 *Given the skeleton of a graph G and a preorder \lesssim of its nodes, G can be constructed uniquely by directing the edges of $\text{sk}(G)$. In addition, \lesssim is a valid preorder for G .*

Let \mathcal{J} be defined over V . We introduce preordering in this paper in order to stress that, in principle, a preordering of members of V (which are possibly random variables) is defined irrespective of graphs to which they are Markov or faithful. However, for our purposes, knowing the skeleton of such graphs, one can directly discover the directions of the edges of the graph, which correspond to the minimal preordering for the graph, rather than working directly on the set V .

Suppose that there exists a preorder \lesssim over V . Proposition 11 implies that we can direct the edges of $\text{sk}(\mathcal{J})$ based on \lesssim in order to define a *dependence graph* $G(\mathcal{J}, \lesssim)$ induced by \mathcal{J} and \lesssim . Notice that, by Proposition 10, $G(\mathcal{J}, \lesssim)$ is anterital. Similarly, for P , one can define $G(P, \lesssim) = G(\mathcal{J}(P), \lesssim)$.

As mentioned, we are only interested in preorderings that are minimal preorders of graphs. Given \mathcal{J} , we define a preordering \lesssim to be \mathcal{J} -compatible if \lesssim is the minimal preorder for $G(\mathcal{J}, \lesssim)$. Similarly, one can define a P -compatible preorder.

4.2 Ordered Upward- and Downward-stabilities

We now exploit the preordering for independence models in order to define two other properties in addition to the seven properties defined in Section 2.3 (namely singleton-transitive compositional graphoid axioms). We say that an independence model \mathcal{J} over the set V respectively satisfies *ordered upward-* and *downward-stability* w.r.t. a preorder \lesssim of V if the following hold:

8. if $\langle i, j | C \rangle \in \mathcal{J}$ then $\langle i, j | C \cup \{k\} \rangle \in \mathcal{J}$ for every $k \in V \setminus \{i, j\}$ such that $l \lesssim k$ for some $l \in \{i, j\}$ or $l \sim k$ for some $l \in C$ (ordered upward-stability);
9. if $\langle i, j | C \rangle \in \mathcal{J}$ then $\langle i, j | C \setminus \{k\} \rangle \in \mathcal{J}$ for every $k \in V \setminus \{i, j\}$ such that $l \not\lesssim k$ for every $l \in \{i, j\}$ and $l \not\sim k$ for every $l \in C \setminus \{k\}$ (ordered downward-stability).

Ordered upward-stability is a generalization of a modification of *upward stability*, defined in Fallat et al. (2017), and *strong union*, defined in Pearl and Paz (1985), for undirected graphs, with singletons instead of node subsets.

Proposition 12 *Let G be an AnG and \lesssim the minimal preorder for G . Then $\mathcal{J}(G)$ satisfies ordered upward- and downward-stability w.r.t. \lesssim .*

Proof First we prove that $\mathcal{J}(G)$ satisfies ordered upward-stability w.r.t. the minimal preorder: If $k \in C$, the result is obvious, thus suppose that $k \notin C$. Suppose that there is a connecting walk π between i and j given $C \cup \{k\}$. If k is not on π then we are done; otherwise, k is on a collider section with no other node in C on π . Notice that $l \lesssim k$ means here that $k \in \text{ant}(l)$, and $l \sim k$ that k and l are connected by a path consisting of lines.

If $l \in C$ then denote the path of lines between k and l by ω . The walk consisting of the subwalk of π from i to k , ω , ω in the reverse direction (from l to k), and the subwalk of π from k to j is connecting given C . This is because on this walk k and l are in the same collider section with l in C .

Now denote an anterior path from k to $l \in \{i, j\}$ by ϖ . Without loss of generality we can assume that $l = j$. We now consider two cases: 1) If ϖ has no nodes in C then the walk containing the subwalk of π between i and k and ϖ is connecting given C between i and j since k and all nodes on ϖ are on non-collider sections on this walk. 2) If ϖ has a node in C then consider the one closest to k and call it h . Then the walk consisting of the subwalk of π from i to k , the subpath of ϖ from k to h , the reverse of the subpath from h to k , and the subwalk of π from k to j is connecting given C as proven as follows: 1) if there is an arrow on ϖ then k is on a non-collider section in both instances in which it appears on this walk, and h is on a collider section and in C . 2) if ϖ only consists of lines then k and h are in the same collider section with a node (h) in C .

Now we prove that $\mathcal{J}(G)$ satisfies ordered downward-stability: If $k \notin C$ then the result is obvious, thus suppose that $k \in C$. Suppose that there is a connecting walk π between i and j given $C \setminus \{k\}$. If k is not on π then we are done. Otherwise, first it is clear that since π is connecting given $C \setminus \{k\}$, k cannot be the only node on a collider section that is in C .

If k is on the same section as another member of C then π is clearly connecting given C . The only case that is left is when k is on a non-collider section on π ; but this is impossible: there is no arrowhead at the section containing k from one side on π , say from the j side. By moving towards j from k , it can be implied that k is either an anterior of a node in a collider section that is in C , or an anterior of j , but both are impossible since the former implies $l < k$ for $l \in C$, and the latter implies $j \lesssim k$. ■

5. Characterization of Graphical Independence Models

In this section, we first provide sufficient conditions for minimal Markovness, and then necessary and sufficient conditions for faithfulness.

5.1 Sufficient Conditions for Minimal Markovness

Here we provide sufficient conditions for an independence model to be minimally Markov to a graph. First, we have the following trivial result:

Proposition 13 *If \mathcal{J} is minimally Markov to an anteriel graph G then $G = G(\mathcal{J}, \lesssim)$ for the minimal preorder \lesssim for G .*

By Proposition 5, the above statement also holds when \mathcal{J} and G are faithful.

Proposition 14 *Suppose that there exists a \mathcal{J} -compatible preorder \lesssim over V w.r.t. which \mathcal{J} satisfies ordered downward- and upward-stability. It then holds that \mathcal{J} is pairwise Markov to $G(\mathcal{J}, \lesssim)$.*

Proof Suppose that there are non-adjacent nodes i and j in $G(\mathcal{J}, \lesssim)$. These are non-adjacent in $\text{sk}(\mathcal{J})$ too. By definition of $\text{sk}(\mathcal{J})$, it holds that $\langle i, j | C \rangle \in \mathcal{J}$ for some $C \subseteq V \setminus \{i, j\}$.

Since \mathcal{J} satisfies ordered downward-stability and since the preorder is \mathcal{J} -compatible, as long as a node $k \in C$ is not an anterior of $\{i, j\}$ or there is no semi-directed path from it to $C \setminus \{k\}$, it can be removed from the conditioning set, that is we have that $\langle i, j | C \setminus k \rangle \in \mathcal{J}$. Since $G(\mathcal{J}, \lesssim)$ is acyclic, as long as $\text{ant}(\{i, j\}) \subset C$, such a k exists. Now consider a node $k' \in C \setminus \{k\}$ that is not an anterior of $\{i, j\}$ or there is no semi-directed path from it to $C \setminus \{k, k'\}$ and apply downward-stability again. By an inductive argument we imply that $\langle i, j | C \cap \text{ant}(\{i, j\}) \rangle \in \mathcal{J}$. Now since P satisfies ordered upward-stability, nodes outside C that are in $\text{ant}(\{i, j\})$ can be added to the conditioning set. Hence, it holds that $\langle i, j | \text{ant}(\{i, j\}) \setminus \{i, j\} \rangle \in \mathcal{J}$. This completes the proof. \blacksquare

In principle, a compatible preorder w.r.t. which \mathcal{J} satisfies ordered downward- and upward-stability can be found by going through all such preorderings. Finding such a preorder efficiently is a structural learning task. We believe we have an efficient way to do this when the skeleton of \mathcal{J} has been learned, but this is beyond the scope of this paper.

Theorem 15 *Suppose, for an independence model \mathcal{J} over V , that*

1. \mathcal{J} is a compositional graphoid; and
2. there exists a \mathcal{J} -compatible preorder \lesssim over V w.r.t. which \mathcal{J} satisfies ordered downward- and upward-stability.

It then holds that \mathcal{J} is minimally Markov to $G(\mathcal{J}, \lesssim)$.

Proof The proof follows from Proposition 14, the fact that $\text{sk}(\mathcal{J}) = \text{sk}(G(\mathcal{J}, \lesssim))$, and Proposition 3. \blacksquare

5.2 Conditions for Faithfulness

In this section, we present the main result of this paper, which provides necessary and sufficient conditions for faithfulness of independence models (and probability distributions) and a (not necessarily known) graph in the general case. In Section 6, we specialize the result to the more well-known subclasses as corollaries.

Proposition 16 *Let \mathcal{J} be an independence model over V . Suppose that \mathcal{J} is minimally Markov to an anteriel graph G . It then holds that \mathcal{J} and G are faithful if and only if*

1. \mathcal{J} satisfies singleton-transitivity; and
2. \mathcal{J} satisfies ordered downward- and upward-stability w.r.t. the minimal preorder for G .

Proof Suppose that \mathcal{J} and G are faithful. This is equivalent to $\mathcal{J} = \mathcal{J}(G)$. By Proposition 2, \mathcal{J} satisfies singleton-transitivity, and by Proposition 12, the result follows.

Conversely, suppose that \mathcal{J} satisfies singleton-transitivity as well as ordered downward- and upward-stability w.r.t. the minimal preorder \lesssim for G . By Proposition 13, $G = G(\mathcal{J}, \lesssim)$. We show that \mathcal{J} and G are faithful. We need to show that if $\langle A, B | C \rangle \in \mathcal{J}$ then $A \perp B | C$ in G . Consider $i \in A$ and $j \in B$. By decomposition, we have that $\langle i, j | C \rangle \in \mathcal{J}$, which implies that i and j are not adjacent in G .

If, for contradiction, C does not separate nodes i and j then there exists a connecting walk $\pi = \langle i = i_1, i_2, \dots, i_r = j \rangle$, on which all collider sections have a node in C and all non-collider sections are outside C . In addition, for every node i_q , $2 \leq q \leq r-1$, on π define C_{i_q} as follows: If $i_q \notin C$ then $C_{i_q} = \emptyset$; if $i_q \in C$ then C_{i_q} is the set of all $l \in C \setminus \pi$ such that there is a semi-directed path from i_q to l but not from any $i_p \in C$ to l , $p \in \{q+1, \dots, r-1\}$. Let $C_\pi = \bigcup_{q=2}^{r-1} C_{i_q}$.

We prove by induction along nodes of π that for every q , $2 \leq q \leq r$, it holds that $\langle i_1, i_q | (\bigcup_{p=2}^{q-1} C_{i_p}) \cup [C \setminus (\{i_q, \dots, i_r\} \cup C_\pi)] \rangle \notin \mathcal{J}$. The base for $q = 2$ is trivial since by the construction of $G(\mathcal{J}, \lesssim)$, adjacent nodes are dependent given anything. Suppose now that the result holds for q and we prove it for $q+1$.

First, again by the construction of $G(\mathcal{J}, \lesssim)$, it holds that $\langle i_q, i_{q+1} | (\bigcup_{p=2}^{q-1} C_{i_p}) \cup [C \setminus (\{i_q, \dots, i_r\} \cup C_\pi)] \rangle \notin \mathcal{J}$. By singleton-transitivity, this and the induction hypothesis imply (a): $\langle i_1, i_{q+1} | (\bigcup_{p=2}^{q-1} C_{i_p}) \cup [C \setminus (\{i_q, \dots, i_r\} \cup C_\pi)] \rangle \notin \mathcal{J}$; or (b): $\langle i_1, i_{q+1} | \{i_q\} \cup (\bigcup_{p=2}^{q-1} C_{i_p}) \cup [C \setminus (\{i_q, \dots, i_r\} \cup C_\pi)] \rangle \notin \mathcal{J}$. We now have two cases:

1) If i_q is on a collider section: The section containing i_q has a node in C . We again consider two subcases:

1.i) If i_q is in C : If (a) holds then first notice that i_q is not an anterior of $\{i_1, i_{q+1}\}$ nor is there a semi-directed path from i_q to $(\bigcup_{p=2}^{q-1} C_{i_p}) \cup [C \setminus (\{i_q, \dots, i_r\} \cup C_\pi)]$. We apply downward-stability to (a) to obtain (b). If $C_{i_q} \neq \emptyset$ then consider a highest order node l in C_{i_q} . Again by downward-stability, we obtain $\langle i_1, i_{q+1} | \{l\} \cup (\bigcup_{p=2}^{q-1} C_{i_p}) \cup [C \setminus (\{i_{q+1}, \dots, i_r\} \cup C_\pi)] \rangle \notin \mathcal{J}$. By an inductive argument along the members of C_{i_q} (by, at each step, choosing a highest order node in C_{i_q} that has not yet been chosen), we eventually obtain $\langle i_1, i_{q+1} | (\bigcup_{p=2}^q C_{i_p}) \cup [C \setminus (\{i_{q+1}, \dots, i_r\} \cup C_\pi)] \rangle \notin \mathcal{J}$. Notice that i_q is in the conditioning set of this dependency.

1.ii) If i_q is not in C : If (b) holds then observe that i_q is in the same equivalence class as that of either i_{q+1} or a node $i_s \in C$, $s < q$. Hence we can apply upward-stability to obtain (a). Since $C_{i_2} = \emptyset$, (a) is clearly the same as $\langle i_1, i_{q+1} | (\bigcup_{p=2}^q C_{i_p}) \cup [C \setminus (\{i_{q+1}, \dots, i_r\} \cup C_\pi)] \rangle \notin \mathcal{J}$. Notice that i_q is not in the conditioning set of this dependency.

2) If i_q is on a non-collider section: We have that $i_q \notin C$, and $i_q \in \text{ant}(\{i_1, i_{q+1}\})$. Hence, by upward-stability, from (b) we obtain (a). Again, since $C_{i_q} = \emptyset$, we obtain the result.

Therefore, we established that $\langle i_1, i_q | (\bigcup_{p=2}^{q-1} C_{i_p}) \cup [C \setminus (\{i_q, \dots, i_r\} \cup C_\pi)] \rangle \notin \mathcal{J}$, for all q , where depending on whether $i_q \in C$ or not, the conditioning set contains or does not contain i_q . By letting $q = r$, we obtain $\langle i_1, i_r | C \rangle \notin \mathcal{J}$, a contradiction. Therefore, it holds that $i \perp j | C$ in G . By the composition property for separation for graphs (Proposition 1), we obtain the result. \blacksquare

We follow the arguments in the proof for the following example.

Example 1 Consider the undirected (LWF) chain graph G in Fig. 2. Suppose that a probability distribution P is minimally Markov to G , and $\mathcal{J}(P)$ satisfies singleton-transitivity and ordered downward- and upward-stability w.r.t. the minimal preorder for G . In order to prove faithfulness, one needs to show that if two sets of nodes are not separated given a third set then they are not independent either.

For example, we have that $l \not\perp j | l_2$. Notice that this is a case where there is no path that connects l and j . In order to show that $l \not\perp j | l_2$, (as in the proof of the above theorem) we

start from $l = l_1$ and move towards $j = l_7$ via the connecting walk $\langle l_1, l_2, l_3, l_4, l_5, l_6, l_7 \rangle$ given l_2 .

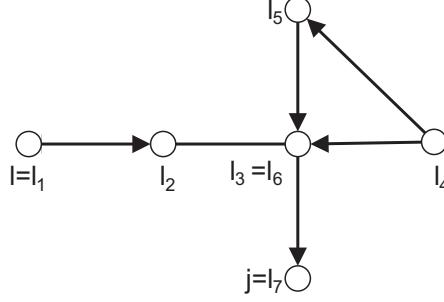


Figure 2: An undirected (LWF) chain graph.

We abbreviate singleton-transitivity, ordered upward-stability, and ordered downward stability by “st”, “ous”, and “ods” respectively. We write the arguments in a condensed from. For example, the first implication (on the left hand side) says that $l_1 \not\perp\!\!\!\perp l_2$ and $l_2 \not\perp\!\!\!\perp l_3$, using singleton-transitivity, imply that either $l_1 \not\perp\!\!\!\perp l_3$ or $l_1 \not\perp\!\!\!\perp l_3 | l_2$; but now ordered downward-stability implies that we must have $l_1 \not\perp\!\!\!\perp l_3 | l_2$.

$$\begin{array}{ccccccc}
 l_1 \not\perp\!\!\!\perp l_2 & l_1 \not\perp\!\!\!\perp l_3 & l_4 \not\perp\!\!\!\perp l_5 | l_2 & l_1 \not\perp\!\!\!\perp l_5 | \{l_2, l_4\} & l_6 \not\perp\!\!\!\perp l_7 | l_2 & l_1 \not\perp\!\!\!\perp l_7 | \{l_2, l_6\} \\
 \xrightarrow{\text{st}} \text{or} & \downarrow \text{ods} & \xrightarrow{\text{st}} & \text{or} & \xrightarrow{\text{st}} & \text{or} & \downarrow \text{ous} \\
 l_2 \not\perp\!\!\!\perp l_3 & l_1 \not\perp\!\!\!\perp l_3 | l_2 & l_1 \not\perp\!\!\!\perp l_4 | l_2 & l_1 \not\perp\!\!\!\perp l_5 | l_2 & l_1 \not\perp\!\!\!\perp l_6 | l_2 & l_1 \not\perp\!\!\!\perp l_7 | l_2. \\
 & \xrightarrow{\text{st}} & \text{or} & \uparrow \text{ous} & \xrightarrow{\text{st}} & \text{or} & \uparrow \text{ous} \\
 l_3 \not\perp\!\!\!\perp l_4 | l_2 & l_1 \not\perp\!\!\!\perp l_3 | \{l_2, l_3\} & l_5 \not\perp\!\!\!\perp l_6 | l_2 & l_1 \not\perp\!\!\!\perp l_6 | \{l_2, l_5\} & & &
 \end{array}$$

The following example shows how singleton-transitivity and ordered stabilities are necessary for faithfulness.

Example 2 Consider the DAG G in Fig. 3. If an independence model \mathcal{J} is minimally Markov to G then we must have $\langle h, k | \{j, l\} \rangle \in \mathcal{J}$ and $\langle j, l | k \rangle \in \mathcal{J}$. Now, in violation of faithfulness, if only $\langle h, k | \emptyset \rangle \in \mathcal{J}$ in addition to these holds then upward-stability is violated since for $j > h$, $\langle h, k | j \rangle \in \mathcal{J}$ must hold. If, in violation of faithfulness, only $\langle h, k | j \rangle \in \mathcal{J}$ in addition to the original statements holds then singleton-transitivity is violated since $\langle h, k | j \rangle \in \mathcal{J}$ and $\langle h, k | \{j, l\} \rangle \in \mathcal{J}$ must imply either $\langle h, l | j \rangle \in \mathcal{J}$ or $\langle k, l | j \rangle \in \mathcal{J}$, which are both impossible due to minimality since h, l and k, l are adjacent in G .

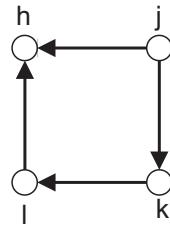


Figure 3: A DAG.

Without assuming the Markov assumption we have the following:

Theorem 17 Let \mathcal{J} be an independence model defined over V . It then holds that \mathcal{J} is graphical if and only if

1. \mathcal{J} is a singleton-transitive compositional graphoid; and
2. there exists a \mathcal{J} -compatible preorder \lesssim over V w.r.t. which \mathcal{J} satisfies ordered downward- and upward-stability.

In addition, if 1 and 2 hold then \mathcal{J} is faithful to $G(\mathcal{J}, \lesssim)$.

Proof (\Rightarrow) follows from Propositions 1, 2, and 12, and the fact that there is always an AnG that is Markov equivalent to a given CMG. To prove (\Leftarrow) and the last statement of the theorem, consider $G(\mathcal{J}, \lesssim)$. Theorem 15 implies that \mathcal{J} is minimally Markov to $G(\mathcal{J}, \lesssim)$. This together with the \mathcal{J} -compatibility of the preorder, using Proposition 16, implies that \mathcal{J} and $G(\mathcal{J}, \lesssim)$ are faithful. \blacksquare

Hence, for probability distributions, we have the following characterization:

Corollary 18 Let P be a probability distribution defined over $\{X_\alpha\}_{\alpha \in V}$. It then holds that P is graphical if and only if

1. $\mathcal{J}(P)$ satisfies intersection, composition, and singleton-transitivity; and
2. there exists a P -compatible preorder \lesssim over V w.r.t. which $\mathcal{J}(P)$ satisfies ordered downward- and upward-stability.

In addition, if 1 and 2 hold then P is faithful to $G(P, \lesssim)$.

In fact, we have shown in Theorem 17 that if \mathcal{J} is graphical then for every \mathcal{J} -compatible preorder w.r.t. which \mathcal{J} satisfies ordered downward- and upward-stability, $G(\mathcal{J}, \lesssim)$ is a graph that is faithful to \mathcal{J} . A consequence of Proposition 13 implies that such graphs (that is $G = G(\mathcal{J}, \lesssim)$ for some preorder \lesssim w.r.t. which \mathcal{J} satisfies ordered downward- and upward-stability) constitute the set of all graphs that are faithful to \mathcal{J} , that is a Markov equivalence class of graphs whose members are faithful to \mathcal{J} . Every member of the equivalence class corresponds to a different preorder w.r.t. which \mathcal{J} satisfies ordered downward- and upward-stability. This equivalence set can be represented by one member (a graph with the least number of arrowheads) similar to the idea of a CPDAG for the Markov equivalence class of DAGs (see Spirtes et al. 2000), but we do not go through the details of this here.

Of course, if the goal is to verify whether \mathcal{J} and a given G are faithful then we can use the following corollary:

Corollary 19 Let \mathcal{J} be an independence model defined over V , and G an AnG with node set V . It then holds that if

1. \mathcal{J} is a singleton-transitive compositional graphoid; and
 2. \mathcal{J} satisfies ordered downward- and upward-stability w.r.t. the minimal preorder for G ,
- then \mathcal{J} and G are faithful.

It is also important to note that singleton-transitivity and ordered upward or downward-stability are necessary in the sense that they are not implied by one another, although there are different ways to axiomatize singleton-transitive compositional graphoids that satisfy ordered upward- and downward-stability.

If P is Gaussian then we have the following:

Corollary 20 *Let P be regular Gaussian distribution. It then holds that P is graphical if and only if there exists a P -compatible preorder \lesssim w.r.t. which $\mathcal{J}(P)$ satisfies ordered downward- and upward-stability.*

Proof The proof follows from the fact that regular Gaussian distributions satisfy intersection, composition, and singleton-transitivity. \blacksquare

6. Specialization to Subclasses

The definitions and results presented in this paper can be specialized to any subclass of CMGs including ancestral graphs and LWF chain graphs. However, here we only focus on the three of the most used classes of undirected (concentration), bidirected (covariance), and directed acyclic graphs.

6.1 Specialization to Undirected and Bidirected Graphs

For connected UGs, every two nodes are in the same equivalence class, and for BGs every two nodes are incomparable; hence, the minimal (pre)ordering is trivial and uninteresting in these cases. More precisely, ordered upward- and downward-stabilities can be specialized for these two types of trivial preordering:

- 8a. if $\langle i, j \mid C \rangle \in \mathcal{J}$ then $\langle i, j \mid C \cup \{k\} \rangle \in \mathcal{J}$ for every $k \in V \setminus \{i, j\}$ (*upward-stability*);
- 9a. if $\langle i, j \mid C \rangle \in \mathcal{J}$ then $\langle i, j \mid C \setminus \{k\} \rangle \in \mathcal{J}$ for every $k \in V \setminus \{i, j\}$ (*downward-stability*).

Proposition 21 *An independence model \mathcal{J} satisfies ordered upward- and downward-stability (8) and (9) w.r.t. the minimal preorder if and only if,*

1. *for connected undirected graphs, it satisfies upward-stability (8a);*
2. *for bidirected graphs, it satisfies downward-stability (9a).*

In addition, if, for undirected graphs, \mathcal{J} satisfies (8a) then it satisfies (8).

Proof The proof of 1 and 2 are straightforward since in 1 all nodes are in the same equivalence class, and in 2 all nodes are incomparable. The second part is trivial. \blacksquare

Analogous to Proposition 12, it is also straightforward to show the following:

Proposition 22 *It holds that*

1. *for an undirected graph G , $\mathcal{J}(G)$ satisfies upward-stability;*

2. for a bidirected graph G , $\mathcal{J}(G)$ satisfies downward-stability.

For UGs and BGs, denote the unique and trivial valid preorders by \lesssim_u^* and \lesssim_b^* respectively. For UGs, $G_u^*(\mathcal{J}) := G(\mathcal{J}, \lesssim_u^*) = \text{sk}(\mathcal{J})$, whereas for BGs, $G_b^*(\mathcal{J}) := G(\mathcal{J}, \lesssim_b^*)$ is $\text{sk}(\mathcal{J})$ with all edges directed to be bidirected. Another way to construct induced UGs and BGs by \mathcal{J} is to construct them based on their corresponding pairwise Markov property, that is to connect every pair of nodes i and j if $\langle i, j | V \setminus \{i, j\} \rangle \notin \mathcal{J}$ for UGs, and if $\langle i, j | \emptyset \rangle \notin \mathcal{J}$ for BGs. Denote these graphs by $G_u(\mathcal{J})$ and $G_b(\mathcal{J})$ respectively. We then have the following:

Proposition 23 *It holds that,*

1. for undirected graphs, $G_u^*(\mathcal{J}) = G_u(\mathcal{J})$ if \mathcal{J} satisfies upward-stability;
2. for bidirected graphs, $G_b^*(\mathcal{J}) = G_b(\mathcal{J})$ if \mathcal{J} satisfies downward-stability.

For upward- and downward-stability, we also have the following observation:

Lemma 24 *For an independence model \mathcal{J} the following holds:*

1. If \mathcal{J} is a semi-graphoid and satisfies upward-stability then \mathcal{J} satisfies composition.
2. If \mathcal{J} is a semi-graphoid and satisfies downward-stability then \mathcal{J} satisfies intersection.

Proof 1. Suppose that $\langle A, B | C \rangle \in \mathcal{J}$ and $\langle A, D | C \rangle \in \mathcal{J}$. By decomposition, it holds, for $i \in A$, $j \in B$, and $k \in D$, that $\langle i, j | C \rangle \in \mathcal{J}$ and $\langle i, k | C \rangle \in \mathcal{J}$. By upward-stability, we obtain $\langle i, j | C \cup \{k\} \rangle \in \mathcal{J}$. Contraction implies $\langle i, \{j, k\} | C \rangle \in \mathcal{J}$. By an inductive argument, we obtain the result.

2. Suppose that $\langle A, B | C \cup D \rangle \in \mathcal{J}$ and $\langle A, D | C \cup B \rangle \in \mathcal{J}$. By decomposition, it holds, for $i \in A$, $j \in B$, and $k \in D$, that $\langle i, j | C \cup \{k\} \rangle \in \mathcal{J}$ and $\langle i, k | C \cup \{j\} \rangle \in \mathcal{J}$. By downward-stability, we obtain $\langle i, k | C \rangle \in \mathcal{J}$. Contraction implies $\langle i, \{j, k\} | C \rangle \in \mathcal{J}$. By an inductive argument, we obtain the result. \blacksquare

Lemma 25 *Let \mathcal{J} be an independence model, and suppose that \mathcal{J} is a graphoid. It then holds that \mathcal{J} satisfies ordered downward-stability (9) w.r.t. the minimal preorder for $G_u(\mathcal{J})$.*

Proof By definition, we know that \mathcal{J} is pairwise Markov to $G_u(\mathcal{J})$. It is known that under graphoid, this implies that \mathcal{J} is Markov to $G_u(\mathcal{J})$; see Pearl (1988). Now suppose that $\langle i, j | C \rangle \in \mathcal{J}$. The only applicable k in the definition of (9) is not in the same connected component as that of i or j . Hence, since $i \perp j | C \setminus \{k\}$, by the global Markov property, it holds that $\langle i, j | C \setminus \{k\} \rangle \in \mathcal{J}$. \blacksquare

As a consequence of Theorem 15, we have the following for UGs and BGs:

Corollary 26 *It holds that*

1. if \mathcal{J} is a graphoid and satisfies upward-stability then \mathcal{J} is minimally Markov to the undirected graph $G_u(\mathcal{J})$;
2. if \mathcal{J} is a compositional semi-graphoid and satisfies downward-stability then \mathcal{J} is minimally Markov to the bidirected graph $G_b(\mathcal{J})$.

Proof The proof follows from Propositions 23 and 21, and Lemmas 24 and 25. ■

For other classes of graphs, the preordering (or ordering) of nodes (and consequently random variables) is necessary since the conditioning set $C(i, j)$ for the pairwise Markov property depends on the preordering of the variables. This implies that in general one cannot construct more than the skeleton of graphs based on the pairwise Markov property if there is no given preordering.

Proposition 16 specializes to the following two corollaries:

Corollary 27 Suppose that an independence model \mathcal{J} is Markov to $G_u(\mathcal{J})$. Then \mathcal{J} and $G_u(\mathcal{J})$ are faithful if and only if \mathcal{J} satisfies singleton-transitivity and upward-stability.

Proof The proof follows from Propositions 21, 22, and 23, and Lemma 25. ■

Similarly, for BGs, we have the following:

Corollary 28 Suppose that an independence model \mathcal{J} is Markov to $G_b(\mathcal{J})$. Then \mathcal{J} and $G_b(\mathcal{J})$ are faithful if and only if \mathcal{J} satisfies singleton-transitivity and downward-stability.

The main conditions for an independence model to be faithful to a UG or BG (consequence of Theorem 17) are as follows; see also Theorem 3 of Pearl and Paz (1985) for a similar result for UGs:

Corollary 29 Let \mathcal{J} be an independence model defined over V . It then holds that \mathcal{J} is faithful to an undirected graph if and only if it is an upward-stable singleton-transitive graphoid. In addition, if the conditions hold then \mathcal{J} is faithful to $G_u(\mathcal{J})$.

Proof The proof follows from Propositions 21, 22, and 23, and Lemmas 24 and 25. ■

Corollary 30 Let \mathcal{J} be an independence model defined over V . It then holds that \mathcal{J} is faithful to a bidirected graph if and only if it is a downward-stable singleton-transitive compositional semi-graphoid. In addition, if the conditions hold then \mathcal{J} is faithful to $G_b(\mathcal{J})$.

Therefore, a probability distribution is faithful to an undirected graph if and only if it satisfies intersection, singleton-transitivity, and upward-stability; and to a bidirected graph if and only if it satisfies composition, singleton-transitivity, and downward-stability.

Finally, as a consequence of Corollary 20, we have the following:

Corollary 31 Let P be a regular Gaussian distribution. It then holds that P being faithful to an undirected graph is equivalent to it satisfying upward-stability; and P being faithful to a bidirected graph is equivalent to it satisfying downward-stability.

6.2 Specialization to Directed Acyclic Graphs

For DAGs, the valid preorder specializes to a valid (partial) order since there are no lines in the graph: If $i \rightarrow j$ then $i > j$. In the minimal order for a DAG, two nodes are incomparable if and only if neither is an ancestor of the other. Therefore, we define ordered stabilities w.r.t. ordering \leq for DAGs:

- 8b. if $\langle i, j \mid C \rangle \in \mathcal{J}$ then $\langle i, j \mid C \cup \{k\} \rangle \in \mathcal{J}$ for every $k \in V \setminus \{i, j\}$ such that $l < k$ for some $l \in \{i, j\}$ (*ordered upward-stability*);
- 9b. if $\langle i, j \mid C \rangle \in \mathcal{J}$ then $\langle i, j \mid C \setminus \{k\} \rangle \in \mathcal{J}$ for every $k \in V \setminus \{i, j\}$ such that $l \not< k$ for every $l \in \{i, j\} \cup C \setminus \{k\}$ (*ordered downward-stability*).

As a consequence of Proposition 12, we have the following:

Corollary 32 *For a directed acyclic graph G , the induced independence model by d -separation $\mathcal{J}(G)$ satisfies ordered upward- and downward-stability (8b) and (9b) w.r.t. the minimal order for G .*

As a consequence of Theorem 15, and for $G(\mathcal{J}, \leq)$, the DAG obtained by directing the edges of $\text{sk}(\mathcal{J})$ based on the order \leq , we have the following:

Corollary 33 *Suppose, for an independence model \mathcal{J} over V , that*

1. \mathcal{J} is a compositional graphoid; and
2. there exists a \mathcal{J} -compatible order \leq over V w.r.t. which \mathcal{J} satisfies ordered downward- and upward-stability, (8b) and (9b).

Then \mathcal{J} is minimally Markov to the directed acyclic graph $G(\mathcal{J}, \leq)$.

The main conditions for an independence model to be faithful to a DAG (consequence of Theorem 17) is as follows:

Corollary 34 *Let \mathcal{J} be an independence model defined over V . It then holds that \mathcal{J} is faithful to a directed acyclic graph G if and only if*

1. \mathcal{J} is a singleton-transitive compositional graphoid; and
2. there exists a \mathcal{J} -compatible order \leq over V w.r.t. which \mathcal{J} satisfies ordered downward- and upward-stability, (8b) and (9b).

In addition, if the conditions hold then \mathcal{J} is faithful to $G(\mathcal{J}, \leq)$.

For a probability distribution P , the first condition of the above corollary simplifies to that P satisfies intersection, composition, and singleton-transitivity.

7. Comparison to the Results in the Literature

All results in the literature concerning faithfulness of certain probability distributions and certain types of graphs can be considered corollaries (or examples) of Theorem 17. Here we verify this for some of the more well-known or interesting results. For brevity, we leave out some interesting results such as faithfulness of Gaussian and discrete distributions and LWF chain graphs; see Peña (2011) and Peña (2009); Studený and Bouckaert (1998) respectively.

7.1 Faithfulness of MTP₂ Distributions and Undirected Graphs

Multivariate totally positive of order 2 (MTP₂) distributions (Karlin and Rinott, 1980) are distributions whose density f is MTP₂, that is $f(x)f(y) \leq f(x \wedge y)f(x \vee y)$, for all vectors x, y , where $x \wedge y = (\min(x_1, y_1), \dots, \min(x_m, y_m))$ and $x \vee y = (\max(x_1, y_1), \dots, \max(x_m, y_m))$. In Fallat et al. (2017), it was shown that when an MTP₂ distribution P is a graphoid, it holds that P and what we denote here by $G_u(P)$ are faithful. It is known that MTP₂ distributions satisfy both singleton-transitivity and upward-stability, but not necessarily the intersection property; see Fallat et al. (2017). Hence, this result is a direct implication of Corollary 29.

7.2 Faithfulness of Gaussian Distributions and Undirected Graphs

Lněnička and Matúš (2007) provides regular Gaussian distributions that are faithful to any undirected graph; see Corollary 3 of the mentioned paper. In order to do so, given an undirected graph $G = (V, E)$, they provide the following matrix $A^{G,\epsilon}$, which acts as the covariance matrix of a Gaussian distribution that is faithful to G :

$$A^{G,\epsilon} = \begin{cases} 1 & \text{if } i = j, \\ \epsilon & \text{if } i \neq j \text{ and } i \text{ and } j \text{ are adjacent in } G, \\ 0 & \text{otherwise,} \end{cases}$$

where $i, j \in V$.

It is known that, for sufficiently small $\epsilon > 0$, the corresponding Gaussian distribution is regular, and by Corollary 31, we only need to show that it satisfies upward-stability. It is straightforward to show that for sufficiently small $\epsilon > 0$, the inverse of $A^{G,\epsilon}$, which is the concentration matrix of the Gaussian distribution, is an *M-matrix*, that is all its on-diagonal elements are positive and all its off-diagonal elements are non-positive. This is equivalent to the Gaussian distribution being MTP₂; see Karlin and Rinott (1983). Hence, upward-stability is satisfied as discussed above.

7.3 Unfaithfulness of Certain Gaussian Distributions and Undirected Graphs and DAGs

There are many examples in the literature for, and in fact it is easy to construct examples of, probability distributions that are not faithful to undirected graphs; for one of the first observations of this behaviour in a data set, see Example 4 of Cox and Wermuth (1993). As a constructed example, let us discuss Example 1 of Soh and Tatikonda (2014), which provided the Gaussian distribution for $X = (X_1, X_2, X_3, X_4)$ with zero mean and positive definite covariance matrix

$$\Sigma = \begin{bmatrix} 3 & 2 & 1 & 2 \\ 2 & 4 & 2 & 1 \\ 1 & 2 & 7 & 1 \\ 2 & 1 & 1 & 6 \end{bmatrix}.$$

It can be seen that the concentration matrix Σ^{-1} has no zero entries, but $X_1 \perp\!\!\!\perp X_3 \mid X_2$ since $\Sigma_{13} = \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{23}$. Hence, the distribution and its complete skeleton are not faithful. This is clearly a consequence of the violation of upward-stability.

For DAGs too, it is easy to construct or observe examples where faithfulness is violated. One of the simplest of such examples, is what is related to the so-called “transitivity of causation” in philosophical causality; see, for example, Ramsey et al. (2006), where the goal is to show how the PC algorithm (Spirtes et al., 2000) errs. Consider the graph $a \rightarrow b \rightarrow c$, where $a \perp\!\!\!\perp c$ and $a \perp\!\!\!\perp c | b$. This, in our language, is a clear violation of singleton-transitivity since the graph implies that a, b and b, c are always dependent.

7.4 Faithfulness of Gaussian and Discrete Distributions and DAGs

It was shown in Meek (1995) that, in certain measure-theoretic senses, almost all the regular Gaussian and discrete distributions that are Markov to a given DAG are faithful to it. It is known, for DAGs, that the global Markov property is equivalent to the *factorization property* (see Lauritzen 1996), which holds if the joint density can factorize as follows:

$$f_V(x) = \prod_{v \in V} f_V(x_v | x_{\text{pa}(v)}).$$

Since Gaussian and discrete distributions satisfy singleton-transitivity, in order to prove the result, one needs to show that ordered upward- and downward-stabilities, (8b) and (9b), are satisfied for almost all distributions that factorize w.r.t. the DAG. We refrain from showing this explicitly here due to technicality of the proof, but the general idea is similar to the proofs in Appendix B of Meek (1995): For (9b), we have that the probabilistic conditional independence holds for the marginal probability over (i, j, C, k) , and we need to exploit the factorization to show that the same conditional independence formula holds for the marginal probability over (i, j, C) when $k \notin \text{an}(\{i, j\} \cup C)$. In this case the corresponding resulting polynomials are much simpler than those of Meek (1995) since one needs to consider extra marginalization over only one variable k . For (8b), we will need to prove the other direction in a similar fashion.

7.5 Faithfulness of Gaussian Distributions and Ancestral Graphs

The main goal of Richardson and Spirtes (2002) is to capture the conditional independence structure of a DAG model after marginalization and conditioning; see also Sadeghi (2013). This was done by introducing the class of AGs, and it was shown that every AG is probabilistic. In fact, there is a Gaussian distribution faithful to any AG, which is indeed the Gaussian distribution faithful to a DAG after marginalization and conditioning; see Theorem 7.5 of this paper.

However, by using our results, without going through the standard theory of ancestral graphs, it is possible to show that distributions that are faithful to a DAG are graphical after marginalization and conditioning: In order to do so, we use the notation in Sadeghi (2013), for an independence model \mathcal{J} , to define the *independence model after marginalization over M and conditioning on C* , denoted by $\alpha(\mathcal{J}; M, C)$, for disjoint subsets M and C of V :

$$\alpha(\mathcal{J}; M, C) = \{\langle A, B | D \rangle : \langle A, B | D \cup C \rangle \in \mathcal{J} \text{ and } (A \cup B \cup D) \cap (M \cup C) = \emptyset\}.$$

First, we show that $\alpha(\mathcal{J}; M, C)$ satisfies condition 1 of Theorem 17.

Proposition 35 *If \mathcal{J} satisfies singleton-transitivity, intersection, and composition then so does $\alpha(\mathcal{J}; M, C)$.*

Proof For composition, assume that $\langle A, B \mid C' \rangle \in \alpha(\mathcal{J}; M, C)$ and $\langle A, D \mid C' \rangle \in \alpha(\mathcal{J}; M, C)$. By definition, we have $\langle A, B \mid C' \cup C \rangle \in \mathcal{J}$ and $\langle A, D \mid C' \cup C \rangle \in \mathcal{J}$. Composition for \mathcal{J} implies $\langle A, B \cup D \mid C' \cup C \rangle \in \mathcal{J}$. Therefore, $\langle A, B \cup D \mid C' \rangle \in \alpha(\mathcal{J}; M, C)$. Proof of the others is similar to this one and as straightforward. ■

Denote by $\alpha(G; M, C)$, the generated AG from the DAG G and M and C two disjoint subsets of its node set that are marginalized over and conditioned on respectively. Here we do not go through the generating process of AGs from DAGs; for more details, see Richardson and Spirtes (2002); Sadeghi (2013). Without explicitly using the theory of stability under marginalization and conditioning for AGs, we show the following.

Proposition 36 *If \mathcal{J} and a DAG G are faithful then $\alpha(\mathcal{J}; M, C)$ and the ancestral graph $\alpha(G; M, C)$ are faithful.*

Proof By Proposition 35, we only need to show that w.r.t. the ordering associated with $\alpha(G; M, C)$, condition 2 of Theorem 17 is also satisfied by $\alpha(\mathcal{J}; M, C)$.

To prove ordered upward-stability, assume that $\langle i, j \mid C' \cup \{k\} \rangle \notin \alpha(\mathcal{J}; M, C)$, where k is an anterior of a node l , $l \in \{i, j\}$ or connected by lines to l , $l \in C'$ in $H = \alpha(G; M, C)$. We have that $\langle i, j \mid C \cup C' \cup \{k\} \rangle \notin \mathcal{J}$. Since \mathcal{J} and G are faithful, there is a connecting walk π between i and j given $C \cup C' \cup \{k\}$. If k is not on π , π is connecting given $C \cup C'$. Otherwise, it holds that k must be a collider on π . By Lemma 2 in Sadeghi (2013), k is in ancestor of a node $l' \in \{i, j\} \cup C \cup C'$. By replacing the subwalk of π from l' to k by the directed path from k to l' if $l' \in \{i, j\}$, or by adding the directed path from k to l' plus the reverse from l' to k to π if $l' \in C \cup C'$, we conclude that $\langle i, j \mid C \cup C' \rangle \notin \mathcal{J}$. Hence, $\langle i, j \mid C' \rangle \notin \alpha(\mathcal{J}; M, C)$.

To prove ordered downward-stability, assume that $\langle i, j \mid C' \setminus \{k\} \rangle \notin \alpha(\mathcal{J}; M, C)$, where $k \in C'$ is not an anterior of any node $l \in \{i, j\} \cup C'$ in H . We have that $\langle i, j \mid C \cup C' \setminus \{k\} \rangle \notin \mathcal{J}$. Since \mathcal{J} and G are faithful, there is a connecting walk π between i and j given $C \cup C' \setminus \{k\}$. Node k cannot be on π since k must be a non-collider and consequently an ancestor of a node in $\{i, j\} \cup C \cup C'$ in G , which by Lemma 3 in Sadeghi (2013), implies that $k \in \text{an}(l)$ in H . Hence, k is not on π , and consequently π is connecting given $C \cup C'$. We conclude that $\langle i, j \mid C \cup C' \rangle \notin \mathcal{J}$. Hence, $\langle i, j \mid C' \rangle \notin \alpha(\mathcal{J}; M, C)$. ■

This shows that, for every AG, there is a distribution faithful to it since, for every AG, there is a DAG that, after marginalization and conditioning, results in the given AG.

8. Summary and Future Work

We provided sufficient and necessary conditions for an independence model, and consequently a probability distribution, to be faithful to a chain mixed graph. All the definitions and results concerning independence models are true for probabilistic independence models. The conditions can be divided into two main categories: 1) Those that ensure that the distribution is Markov to a graph whose skeleton is the same as the skeleton of the distribution—these properties are namely intersection, composition, and ordered upward- and downward-stabilities. If the distribution is already pairwise Markov to the graph then intersection and composition suffice. Ordered upward- and downward-stabilities direct the edges of the skeleton of the distribution so that the pairwise Markov property is satisfied.

- 2) Those that ensure that a Markov distribution is also faithful to the graph. These are singleton-transitivity and ordered upward- and downward-stabilities.

The type of preordering or preorderings w.r.t. which the distribution satisfies ordered upward- and downward-stabilities implies the type or types of graphs that are faithful to the distribution. In order to only deal with simpler classes of graphs, it is enough to search through a more refined set of preorderings determined by the conditions that graphs of a specific class must satisfy. It is indeed true that when the skeleton of the graph or distribution is known, the preordering is equivalent to different ways the edges of the skeleton can be directed, but in principle the preordering of the variables is unrelated to the skeleton of the distribution.

The conditions of faithfulness provided in this paper for probability distributions are sufficient and necessary. Thus, based on these conditions many families of distributions can be shown not to be suitable for graphical modeling. This also shows the importance of Gaussian distributions in graphical models as the provided conditions (other than ordered upward- and downward-stabilities) are clearly satisfied by the regular Gaussian distributions. In addition, these conditions help those who devise new parametric models to ensure that their models satisfy the provided conditions for faithfulness.

It is clearly important to study the characteristics of distributions that satisfy these conditions. The intersection property is well-studied and, in addition to positivity of the density that has been known to be a sufficient condition, necessary and sufficient conditions for a probabilistic independence to satisfy this property are known; see Peters (2015). For composition and singleton-transitivity, more studies are needed; in particular, it would be worthwhile to find simple sufficient conditions for probability distributions in order to satisfy these properties.

Another approach is to devise algorithms that test whether these conditions are satisfied given data or an independence model. Indeed one can argue for an alternative approach to determine whether a distribution is faithful to some graph: one first applies the structure learning algorithm (taking as input the distribution and outputting a graph) and then one tests whether the separation relationships in the found graph correspond to the conditional independencies in the distribution. However, testing our conditions are much more efficient. As an example, in the case of undirected graphs and Gaussian distributions, by our results, one only needs to test upward-stability for non-adjacent single elements i and j (starting with a minimal node cut-set as the conditioning set) rather than testing all global conditional independencies implied by the graph.

As another algorithmic consequence that the results in this paper may prompt, they lead to a *constraint-based* structural learning algorithm (as opposed to score-based, for example, Chickering 2003) most similar to the Fast Causal Inference (FCI) algorithm (Spirtes et al., 2000)—it finds the skeleton of the graph and then explores directions of the edges of the graph on the skeleton. Directing the edges of the skeleton is performed based on an exploitation of the preordering of the variables (and consequently the nodes of the graph) so that w.r.t. such a preordering ordered upward- and downward-stabilities are satisfied. This algorithm would have two advantages over the known constraint-based algorithms in the literature: Firstly, it is devised for a larger class of AnGs, which can be thought of as learning the LWF chain graphs with hidden and selection variables; see Sadeghi (2016). Secondly, the exploitation of ordered upward- and downward-stability for structural learn-

ing can be combined with actually testing whether these conditions are satisfied, and the theory in the paper ensures the faithfulness of the generated graphs. We plan on providing the details of this algorithm elsewhere.

When dealing with data, another issue is that obtaining a faithful graph is very sensitive to hypothesis testing errors for inferring the independence statements from data; see Uhler et al. (2013). The concept of *strong faithfulness* (Zhang and Spirtes, 2003) has been suggested for the Gaussian case in order to deal with this issue. Although strong faithfulness, by nature, does not seem to be characterizable in this way, it is an interesting problem to study which of the equivalent conditions to faithfulness are more prone to such sensitivity.

Acknowledgments

The original idea of this paper was inspired by discussions at the American Institute of Mathematics workshop “Positivity, Graphical Models, and the Modeling of Complex Multivariate Dependencies” in October 2014. The author thanks the AIM and NSF for their support of the workshop and is indebted to other participants of the discussion group, especially to Shaun Fallat, Steffen Lauritzen, Nanny Wermuth, and Piotr Zwiernik for follow-up discussions. The author is also grateful to Rasmus Petersen for valuable discussions. This work was partly supported by grant #FA9550-12-1-0392 from the U.S. Air Force Office of Scientific Research (AFOSR) and the Defense Advanced Research Projects Agency (DARPA).

References

- R. Ayesha Ali, Thomas S. Richardson, and Peter Spirtes. Markov equivalence for ancestral graphs. *Annals of Statistics*, 37(5B):2808–2837, 2009.
- Steen A. Andersson, David Madigan, and Michael D. Perlman. Alternative Markov properties for chain graphs. *Scandinavian Journal of Statistics*, 28:33–85, 2001.
- David Maxwell Chickering. Optimal structure identification with greedy search. *Journal of Machine Learning Research*, 3:507–554, 2003.
- Peter Clifford. Markov random fields in statistics. In G. R. Grimmett and D. J. A. Welsh, editors, *Disorder in Physical Systems: A Volume in Honour of John M. Hammersley*, pages 19–32. Oxford University Press, 1990.
- David R. Cox and Nanny Wermuth. Linear dependencies represented by chain graphs (with discussion). *Statistical Science*, 8(3):204–218; 247–277, 1993.
- John N. Darroch, Steffen L. Lauritzen, and Terry P. Speed. Markov fields and log-linear interaction models for contingency tables. *Annals of Statistics*, 8:522–539, 1980.
- A. Philip Dawid. Conditional independence in statistical theory (with discussion). *Journal of the Royal Statistical Society Series B*, 41(1):1–31, 1979.
- Mathias Drton and Thomas S. Richardson. Binary models for maginal independence. *Journal of the Royal Statistical Society Series B*, 41(2):287–309, 2008.

- Shaun Fallat, Steffen Lauritzen, Kayvan Sadeghi, Caroline Uhler, Nanny Wermuth, and Piotr Zwiernik. Total positivity in Markov structures. *Annals of Statistics.*, 45(3), 2017.
- Morten Frydenberg. The chain graph Markov property. *Scandinavian Journal of Statistics*, 17:333–353, 1990.
- Dan Geiger, Thomas S. Verma, and Judea Pearl. Identifying independence in Bayesian networks. *Networks*, 20:507–534, 1990.
- Clark Glymour, Richard Scheines, Peter Spirtes, and Kevin Kelly. *Discovering Causal Structure*. Academic Press, 1987.
- Samuel Karlin and Yosef Rinott. Classes of orderings of measures and related correlation inequalities. I. Multivariate totally positive distributions. *Journal of Multivariate Analysis*, 10(4):467–498, 1980.
- Samuel Karlin and Yosef Rinott. M-matrices as covariance matrices of multinormal distributions. *Linear Algebra and its Applications*, 52:419 – 438, 1983.
- Göran Kauermann. On a dualization of graphical Gaussian models. *Scandinavian Journal of Statistics*, 23(1):105–116, 1996.
- Harri Kiiveri, Terry P. Speed, and J. B. Carlin. Recursive causal models. *Journal of the Australian Mathematical Society, Series A*, 36:30–52, 1984.
- Steffen L. Lauritzen. *Graphical Models*. Clarendon Press, Oxford, United Kingdom, 1996.
- Steffen L. Lauritzen and Kayvan Sadeghi. Unifying Markov properties for graphical models. *Annals of Statistics*, To appear, 2017. URL <https://arxiv.org/abs/1608.05810>.
- Steffen L. Lauritzen and David J. Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society, Series B*, 50:157–224, 1988.
- Steffen L. Lauritzen and Nanny Wermuth. Graphical models for association between variables, some of which are qualitative and some quantitative. *Annals of Statistics*, 17:31–57, 1989.
- Radim Lněnička and František Matúš. On Gaussian conditional independence structures. *Kybernetika*, 43(3):327–342, 2007.
- Christopher Meek. Strong completeness and faithfulness in Bayesian networks. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, UAI’95, pages 411–418, San Francisco, CA, USA, 1995.
- Richard E. Neapolitan. *Learning Bayesian Networks*. Prentice Hall, Upper Saddle River, NJ, USA, 2004.
- Jose M. Peña. Faithfulness in chain graphs: The discrete case. *International Journal of Approximate Reasoning*, 50:1306–1313, 2009.

- Jose M. Peña. Faithfulness in chain graphs: The Gaussian case. In *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS 2011)*, volume 15, pages 588–599. JMLR.org, 2011.
- Jose M. Peña. Marginal AMP chain graphs. *International Journal of Approximate Reasoning*, 55:1185–1206, 2014.
- Judea Pearl. *Probabilistic Reasoning in Intelligent Systems : networks of plausible inference*. Morgan Kaufmann Publishers, San Mateo, CA, USA, 1988.
- Judea Pearl. *Causality: Models, Reasoning and Inference*. Cambridge University Press, New York, NY, USA, 2nd edition, 2009.
- Judea Pearl and Azaria Paz. *Graphoids: A Graph-based Logic for Reasoning about Relevance Relations*. University of California (Los Angeles). Computer Science Department, 1985.
- Jonas Peters. On the intersection property of conditional independence and its application to causal discovery. *Journal of Causal Inference*, 3:97–108, 2015.
- Joseph Ramsey, Jiji Zhang, and Peter Spirtes. Adjacency-faithfulness and conservative causal inference. In *UAI*. AUAI Press, 2006.
- Thomas Richardson. Markov properties for acyclic directed mixed graphs. *Scandinavian Journal of Statistics*, 30(1):145–157, 2003.
- Thomas S. Richardson and Peter Spirtes. Ancestral graph Markov models. *Annals of Statistics*, 30(4):962–1030, 2002.
- Kayvan Sadeghi. Stable mixed graphs. *Bernoulli*, 19(5(B)):2330–2358, 2013.
- Kayvan Sadeghi. Marginalization and conditioning for LWF chain graphs. *Annals of Statistics*, 44(4):1792–1816, 2016.
- Kayvan Sadeghi and Steffen L. Lauritzen. Markov properties for mixed graphs. *Bernoulli*, 20(2):676–696, 2014.
- Bernd S.W. Schröder. *Ordered Sets: An Introduction*. Birkhäuser, Boston, 2003.
- De W. Soh and Sekhar C. Tatikonda. Testing unfaithful Gaussian graphical models. In Z. Ghahramani, M. Welling, C. Cortes, N.d. Lawrence, and K.q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 2681–2689. Curran Associates, Inc., 2014.
- Peter Spirtes, Clark Glymour, and Richard Scheines. *Causation, Prediction, and Search*. MIT press, 2nd edition, 2000.
- Daniel Steel. Homogeneity, selection, and the faithfulness condition. *Minds and Machines*, 16(3):303–317, 2006.
- Milan Studený. Multiinformation and the problem of characterization of conditional independence relations. *Problems of Control and Information Theory*, 18:3–16, 1989.

- Milan Studený. *Probabilistic Conditional Independence Structures*. Springer-Verlag, London, United Kingdom, 2005.
- Milan Studený and Remco R. Bouckaert. On chain graph models for description of conditional independence structures. *Annals of Statistics*, 26(4):1434–1495, 1998.
- Caroline Uhler, Garvesh Raskutti, Peter Bühlmann, and Bin Yu. Geometry of the faithfulness assumption in causal inference. *Annals of Statistics*, 41(2):436–463, 2013.
- Thomas Verma and Judea Pearl. On the equivalence of causal models. In *Proceedings of the Sixth Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-90)*, pages 220–227, New York, NY, 1990.
- Nanny Wermuth. Probability distributions with summary graph structure. *Bernoulli*, 17(3):845–879, 2011.
- Nanny Wermuth and David R. Cox. On association models defined over independence graphs. *Bernoulli*, 4:477–495, 1998.
- Nanny Wermuth and Kayvan Sadeghi. Sequences of regressions and their independences. *TEST*, 21(2):215–252 and 274–279, 2012.
- Nanny Wermuth, Giovanni M. Marchetti, and David R. Cox. Triangular systems for symmetric binary variables. *Electronic Journal of Statistics*, 3:932–955, 2009.
- James Woodward. Causal independence and faithfulness. *Multivariate Behavioral Research*, 33(1):129–148, 1998.
- Jiji Zhang and Peter Spirtes. Strong faithfulness and uniform consistency in causal inference. In *Proceedings of the Nineteenth Conference on Uncertainty in Artificial Intelligence*, UAI'03, 2003.
- Jiji Zhang and Peter Spirtes. Detection of unfaithfulness and robust causal inference. *Minds and Machines*, 18(2):239–271, 2008.