Bernoulli **20**(2), 2014, 676–696
DOI: 10.3150/12-BEJ502

Markov properties for mixed graphs

KAYVAN SADEGHI[1] and STEFFEN LAURITZEN[2]

[1]*Department of Statistics, Baker Hall, Carnegie Mellon University, Pittsburgh, PA 15213, USA.*
E-mail: kayvans@andrew.cmu.edu
[2]*Department of Statistics, University of Oxford, 1 South Parks Road, Oxford, OX1 3TG, United Kingdom.*
E-mail: steffen@stats.ox.ac.uk

In this paper, we unify the Markov theory of a variety of different types of graphs used in graphical Markov models by introducing the class of loopless mixed graphs, and show that all independence models induced by m-separation on such graphs are compositional graphoids. We focus in particular on the subclass of ribbonless graphs which as special cases include undirected graphs, bidirected graphs, and directed acyclic graphs, as well as ancestral graphs and summary graphs. We define maximality of such graphs as well as a pairwise and a global Markov property. We prove that the global and pairwise Markov properties of a maximal ribbonless graph are equivalent for any independence model that is a compositional graphoid.

Keywords: composition property; global Markov property; graphoid; independence model; m-separation; maximality; pairwise Markov property

1. Introduction

1.1. Introduction and motivation

Graphical Markov models have become widely used in recent years. The models use graphs to represent conditional independence relations for systems of random variables, with nodes of the graph corresponding to random variables and edges representing dependencies. Several classes of graphs with various independence interpretations have been described in the literature. These range from undirected graphs with simple separation for derivation of independencies [19] to various forms of mixed graphs [18,24,30], including chain graphs with several different separation criteria [2,5,8,10,17].

In spite of the differences among these graphs, their structural similarities motivate an attempt to unify them. For this purpose, we introduce the class of loopless mixed graphs and let them entail independence models using the same separation criterion, m-separation. This unification covers many graphical independence models in the literature with some independence models for chain graphs forming a notable exception; see Section 4 for further details. We show that any independence model generated by m-separation in a loopless mixed graph is a compositional graphoid. This ensures that certain intuitive methods of reasoning are indeed valid for such graphs, as they in some sense behave as ordinary undirected graphs.

A common motivation for defining MC-graphs [18], summary graphs [30], and ancestral graphs [24], is to represent independence relations implied by marginalisation over and conditioning on sets of variables satisfying the Markov property of a directed acyclic graph (DAG). The focus of our study is on a subclass of loopless mixed graphs which we shall term *ribbonless*.

1350-7265 © 2014 ISI/BS

The class of ribbonless graphs is sufficiently rich to serve the same purpose: these graphs are obtained by a simple modification of MC graphs derived from a DAG after marginalisation and conditioning; and it contains summary graphs and ancestral graphs as special cases.

For ribbonless graphs, we define global and pairwise Markov properties, the latter being associated with interpreting missing edges in the graph as representing conditional independencies. We prove as our main result that a compositional graphoid independence model over a maximal ribbonless graph satisfies the global Markov property if and only if it satisfies the pairwise Markov property. This ensures that the independence models represented by such graphs are generated by their missing edges, which again supports the direct visual intuition.

1.2. Some early results on Markov properties

The concepts of pairwise and global Markov properties for undirected graphs were introduced in [13] in the context of random fields and shown to be equivalent for positive densities. Alternative proofs were later given independently by several authors, for example [3,12]; see also [4]. An abstract variant of this theorem was proven in [21] for independence models satisfying graphoid axioms as these are satisfied by probabilistic distributions with positive densities; see also [29] and [11]. Independence models for undirected graphs were discussed comprehensively in Chapter 3 of [19].

A global Markov property that uses the m-separation criterion and a pairwise Markov property were defined in [24] for maximal ancestral graphs without considering conditions under which they are equivalent. We use a generalisation of these Markov properties for maximal ribbonless graphs, which contains maximal ancestral graphs as a subclass, and prove their equivalence for compositional graphoids. This has been mentioned as a conjecture in [14].

1.3. Structure of the paper

In the next section, we introduce the basic concepts of graph theory, general and probabilistic independence models, and compositional graphoids.

In Section 3, we introduce the class of loopless mixed graphs and additional graph theoretical definitions special to mixed graphs. We also associate the m-separation criterion to this class, and prove for any loopless mixed graph that the independence model induced by m-separation is a compositional graphoid.

In Section 4, we introduce the class of ribbonless graphs and the concept of anterior graphs. We describe the relations between these as well as subclasses of loopless mixed graphs that have been discussed in the literature.

In Section 5, we introduce the concept of maximality by demanding that any additional edge will change the independence model. It is shown that ribbonless graphs are not necessarily maximal, and conditions for maximality are given.

In Section 6, we define a pairwise and a global Markov property for independence models for ribbonless graphs, and prove our main result: that pairwise and global Markov properties are equivalent for compositional graphoid independence models over maximal ribbonless graphs.

2. Basic definitions and concepts

In this section, we introduce basic definitions and notation for independence models, graphs, and compositional graphoids.

2.1. Basic graph theoretical definitions

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* and we write $i \sim j$. We say the 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 as a different object. Hence, for example, graph $i \frac{\quad}{\quad} j \frac{\quad}{\quad} k$ is not equal to $j \frac{\quad}{\quad} i \frac{\quad}{\quad} k$.

A *loop* is an edge with the same endpoints. *Multiple edges* are edges with the same pair of endpoints. A *simple graph* has neither loops nor multiple edges.

A *subgraph* of a graph G_1 is a 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. An *induced subgraph* by nodes $A \subseteq V$ is a subgraph that contains all and only nodes in A and all edges between two nodes in A. A subgraph induced by edges $B \subseteq E$ is a subgraph that contains all and only edges in B and all nodes that are endpoints of edges in B.

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 $1 \le i \le k$, the edge e_i has endpoints v_{i-1} and v_i. A *path* is a walk with no repeated node or edge. If the graph is simple then the path can be uniquely determined by an ordered sequence of node sets. Throughout this paper, we use node sequences to describe paths even in graphs with multiple edges, as it usually is apparent from the context which of multiple edges belong to the path. We say a path is *between* the first and the last nodes of the list in G. We call the first and the last nodes *endpoints* of the path and all other nodes *inner nodes*.

If $\pi_1 = \langle i = i_0, i_1, \dots, i_n, h\rangle$ and $\pi_2 = \langle h, j_m, j_{m-1}, \dots, j_0 = j\rangle$ are paths, their *combination* $\pi_{12} = \pi_1 \circ \pi_2$ is the path $\pi_{12} = \langle i, \dots, i_{p-1}, k, j_{q-1}, \dots, j\rangle$, where $k = i_p = j_q$ is the first node of π_1 which is on both paths. If $k = h$ then π_{12} is simply the concatenation of the two paths. In general, the concatenation of two paths will be a walk and not a path as the paths may intersect in more than one point.

A *subpath* of a path π is a path that can be considered a subgraph of π with the ordering associated with π. A *cycle* in a graph G is a simple subgraph whose nodes can be placed around a circle so that two nodes are adjacent if they appear consecutively along the circle.

2.2. Independence models

An *independence model* $\mathcal{J}$ over a set V is a set of triples $\langle X, Y \mid Z\rangle$ (called *independence statements*), where X, Y, and Z are disjoint subsets of V and Z can be empty, and $\langle \varnothing, Y \mid Z\rangle$ and

$\langle X, \varnothing \mid Z\rangle$ always being included in $\mathcal{J}$. The independence statement $\langle X, Y \mid Z\rangle$ is interpreted as "X is independent of Y given Z".

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

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

A semi-graphoid for which the reverse implication of the weak union property holds is said to be a *graphoid*, that is

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

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

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*).

Notice that simple separation in an undirected graph will trivially satisfy all of these properties, and hence compositional graphoids are direct generalisations of independence models given by separation in undirected graphs.

2.3. Probabilistic conditional independence models

The most common independence models are induced by probability distributions. Consider a set V and a collection of random variables $(X_\alpha)_{\alpha\in V}$ with state spaces $\mathcal{X}_\alpha, \alpha \in V$ 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 $A \perp\!\!\!\perp B \mid C$ to denote that X_A is *conditionally independent of* X_B *given* X_C [7,19], that is, that for any measurable $\Omega \subseteq \mathcal{X}_A$ and P-almost all x_B and x_C,

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

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

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

We say that an independence model $\mathcal{J}$ is *probabilistic* if there is a distribution P such that $\mathcal{J} = \mathcal{J}(P)$. We then also say that P is *faithful* to $\mathcal{J}$.

Probabilistic independence models are always semi-graphoids [21], whereas the converse is not necessarily true; see [29]. If P has strictly positive density, the induced independence model is also a graphoid; see, for example, Proposition 3.1 in [19]. If the distribution P is a regular multivariate Gaussian distribution, $\mathcal{J}(P)$ is a compositional graphoid. This follows from the fact that for such a distribution

$$A \perp\!\!\!\perp B \mid C \quad \Longleftrightarrow \quad k^{\alpha\beta}_{A\cup B\cup C} = 0 \qquad \text{for all } \alpha \in A, \beta \in B,$$

where $k^{\alpha\beta}_{A\cup B\cup C}$ is the $\alpha\beta$ entry in the concentration matrix of the distribution of $X_{A\cup B\cup C}$ and hence setwise conditional independence is directly determined by nodewise conditional independence.

Probabilistic independence models with positive densities are not in general compositional graphoids; this only holds for special types of multivariate distributions such as the Gaussian mentioned above and, say, the symmetric binary distributions used in [32].

3. Independence models for mixed graphs

3.1. Mixed graphs

A *mixed graph* is a graph containing three types of edges denoted by arrows, arcs (bi-directed edges), and lines (full lines). Notice that we allow multiple edges of the same type. A *loopless mixed graph* (LMG) is a mixed graph that does not contain any loops (a loop may be line, arrow, or arc). For an arrow $j \succ\!\!\!\longrightarrow i$, we say that the arrow is *from* j *to* i. We also call j a *parent* of i, i a *child* of j and we use the notation $\mathrm{pa}(i)$ for the set of all parents of i in the graph. In the cases of $i \longrightarrow\!\!\!\succ j$ or $i \prec\!\!\!\longrightarrow\!\!\!\succ j$, we say that there is *an arrowhead at* j or *pointing to* j.

A path $\langle i = i_0, i_1, \ldots, i_n = j\rangle$ is *direction-preserving* from i to j if all $i_k i_{k+1}$ edges are arrows pointing from i_k to i_{k+1}. If there is a direction-preserving path from j to i then j is an *ancestor* of i and i is a *descendant* of j. We denote the set of ancestors of i by $\mathrm{an}(i)$. Notice that we do not include i in its set of anteriors or descendants.

A *tripath* is a path with three nodes. Note that [26] used the term V-configuration for such a path. However, here we follow [16] and most texts by letting a V-configuration be a tripath with non-adjacent endpoints.

In a mixed graph the inner node of three tripaths $i \longrightarrow\!\!\!\succ t \prec\!\!\!\longrightarrow j$, $i \prec\!\!\!\longrightarrow\!\!\!\succ t \prec\!\!\!\longrightarrow j$, and $i \prec\!\!\!\longrightarrow\!\!\!\succ t \prec\!\!\!\longrightarrow\!\!\!\succ j$ is a *collider* (or a collider node) and the inner node of any other tripath is a *non-collider* (or a non-collider node) on the tripath or more generally on any path of which the tripath is a subpath. We shall also say that the tripath itself with inner collider or non-collider node is a *collider* or *non-collider*. We may speak of a collider or non-collider without mentioning the relevant tripath or path when this is apparent from the context. Notice that a node may be a collider on one tripath and a non-collider on another.

Two paths π_1 and π_2 (including tripaths or edges) between i and j are called *endpoint-identical* if there is an arrowhead pointing to i in π_1 if and only if there is an arrowhead pointing to i in π_2 and similarly for j. For example, the paths $i \longrightarrow\!\!\!\succ j$, $i \longrightarrow k \prec\!\!\!\longrightarrow\!\!\!\succ j$, and $i \longrightarrow\!\!\!\succ k \prec\!\!\!\longrightarrow l \prec\!\!\!\longrightarrow\!\!\!\succ j$ are all endpoint-identical as they have an arrowhead pointing to j but no arrowhead pointing to i on the paths.

3.2. Anterior graphs and sets

The *anterior graph* of a loopless mixed graph G, denoted by G^*, is the graph obtained from G by recursively removing arrowheads pointing to nodes that are the endpoints of a line, that is, by obtaining $\longrightarrow \circ \longrightarrow$ and $\prec\!\!\!\longrightarrow \circ \longrightarrow$ from $\longrightarrow\!\!\!\succ \circ \longrightarrow$ and $\prec\!\!\!\longrightarrow\!\!\!\succ \circ \longrightarrow$ respectively. Hence,

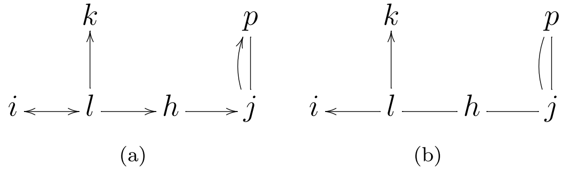


Figure 1. (a) A mixed graph G. (b) The anterior graph G^* of G.

it holds that $G = G^*$ if and only if there are no arrowheads pointing to lines in G. Notice also that since removing an arrowhead pointing to a line does not affect other arrowheads pointing to lines, it does not matter which arrowhead is removed first; therefore, the order of removing arrowheads pointing to lines does not affect the final graph obtained.

A path $\langle i = i_0, i_1, \dots, i_n = j\rangle$ from i to j ($i \neq j$) in G^* is an *anterior path* if it has the form $i - i_1 - \cdots - i_m \to i_{m+1} \to \cdots \to j$. Notice that this path may only contain lines or arrows. We shall say that i is *anterior* of j in G if there is an anterior path from i to j in G^*. Notice that although the anterior path is defined in G^* we may from time to time refer to an anterior path in G as the path corresponding to the anterior path in G^*.

We use the notation $\mathrm{ant}(i)$ for the set of all anteriors of i. Notice that, since ancestral graphs have no arrowheads pointing to lines, we have $G = G^*$ for an ancestral graph. Thus, our definition of anterior extends the notion of anterior used in [24] for ancestral graphs with the minor difference that we do not include a node in its anterior set. However, it is different from and inconsistent with the definition of anteriors in [10] and [1].

For example, in the graph G in Figure 1(a), $\mathrm{ant}(i) = \{l, h, j, p\}$ and $\mathrm{ant}(p) = \{l, h, j\}$. This can be seen by looking at the anterior paths $\langle p, j, h, l, i\rangle$ from p to i and $\langle l, h, j, p\rangle$ from l to p (as well as from p to l) in Figure 1(b).

We first show that transitivity holds for anteriors.

Lemma 1. *For any loopless mixed graph it holds that if* $i \in \mathrm{ant}(j)$ *and* $j \in \mathrm{ant}(k)$ *then* $i \in \mathrm{ant}(k)$.

Proof. If $i \in \mathrm{ant}(j)$ and $j \in \mathrm{ant}(k)$, G^* has anterior paths π_1 from i to j and π_2 from j to k. As no arrowhead meets a line in G^* their combination $\pi_1 \circ \pi_2$ is an anterior path from i to j in G^*. □

Here we also introduce a lemma that is used in several proofs of this paper.

Lemma 2. *Let G be a loopless mixed graph. If* $i \in \mathrm{ant}(j) \setminus \mathrm{an}(j)$, *then either i or a descendant of i is the endpoint of a line in G.*

Proof. The proof uses induction on the number of arrowheads removed from G to obtain G^*. For the base, if $G = G^*$ it follows immediately from the definition of an anterior path that i must be the endpoint of a line or we would have $i \in \mathrm{an}(j)$.

Next, suppose that G^* is obtained from G by removing $n+1$ arrowheads and let $\tilde{G}$ be obtained from G by removing a single arrowhead pointing to a line from G. Then G^* is also the anterior graph of $\tilde{G}$, but with only n arrowheads needing removal. Thus, if $i \in \text{ant}(j)$ in G, it is also anterior to j in $\tilde{G}$. Consider now two cases:

Case I. Assume i is an ancestor of j in $\tilde{G}$. Since i is not an ancestor of j in G, $\tilde{G}$ must have been obtained by turning an arc into an arrow. Say this arrowhead points to h. Then h is an endpoint of a line and it is a descendant of i in G.

Case II. If i is not an ancestor of j in $\tilde{G}$, the inductive hypothesis yields that i is either adjacent to a line ih in $\tilde{G}$ or has a descendant h in $\tilde{G}$ which is the endpoint of a line in $\tilde{G}$. Let h be the node adjacent to a line in $\tilde{G}$. If the arrowhead removed is not on the direction-preserving path π from i to h the conclusion obviously follows. Else, there must be node k on π which is adjacent to a line in G and can be used instead of h. □

3.3. The *m*-separation criterion

Here we define a separation criterion for LMGs. We use this criterion to induce independencies on LMGs and its subclasses defined in Section 3.

We first define an m-connecting path: Let C be a subset of the node set of an LMG. A path is *m-connecting* given C if all its collider nodes are in $C \cup \text{an}(C)$ and all its non-collider nodes are outside C. For two disjoint subsets of the node set A and B, we say that C m-separates A and B if there is no m-connecting path between A and B given C. In this case, we use the notation $A \perp_m B \mid C$. Notice that the m-separation criterion induces an independence model $\mathcal{J}_m(G)$ on G by $A \perp_m B \mid C \iff \langle A, B \mid C \rangle \in \mathcal{J}_m(G)$.

We note that m-separation is unaffected if we replace multiple edges of the same type with a single edge of that type. The m-separation criterion for LMGs is the same as the separation criterion defined in [24]. It is an extension of the d-separation criterion introduced in [21]. Clearly, m-separation is also an extension of simple separation in an undirected graph, as then all edges are lines.

For example, in graph G in Figure 2 it holds that $h \in \text{an}(l)$ and, thus, $\langle i, h, j \rangle$ is an m-connecting path given l. Therefore, $\langle i, j \mid l \rangle \notin \mathcal{J}_m(G)$. We now have the following theorem. A similar result for the induced independence model for MC graphs was given in Proposition 2.10 of [18].

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

Proof. For $G = (N, F)$ and disjoint subsets A, B, C, and D of N, we prove that $\perp_m$ satisfies the six compositional graphoid axioms:

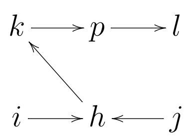


Figure 2. A loopless mixed graph G for which $\langle i, j \mid l \rangle \notin \mathcal{J}_m(G)$.

(1) *Symmetry*: If $A \perp_m B \mid C$, then $B \perp_m A \mid C$: If there is no m-connecting path between A and B given C, then there is no m-connecting path between B and A given C.

(2) *Decomposition*: If $A \perp_m (B \cup D) \mid C$, then $A \perp_m D \mid C$: If there is no m-connecting path between A and $B \cup D$ given C, then there is no m-connecting path between A and $D \subseteq (B \cup D)$ given C.

(3) *Weak union*: If $A \perp_m (B \cup D) \mid C$ then $A \perp_m B \mid (C \cup D)$: From (2) we know that $A \perp_m D \mid C$ and $A \perp_m B \mid C$. Suppose, for contradiction, that there exist m-connecting paths between A and B given $C \cup D$. Consider a shortest path of this type and call it π. If there is no inner collider node on π, then there is an m-connecting path between A and B given C, a contradiction. On π all collider nodes are in $(C \cup D) \cup \mathrm{an}(C \cup D)$. If all collider nodes are in $C \cup \mathrm{an}(C)$, then there is an m-connecting path between A and B given C, again a contradiction. Hence, consider the closest collider node $i \in (D \cup \mathrm{an}(D)) \setminus (C \cup \mathrm{an}(C))$ to A on π. Now since the nodes between A and i are not in $B \cup D$, there is an m-connecting path between A and i given C. If $i \in D$, then this is obviously a contradiction. Otherwise there is a node $k \in D$, for which $i \in \mathrm{an}(k)$ and thus an m-connecting path between A and k given C, a contradiction again. Therefore, there is no m-connecting path between A and B given $C \cup D$.

(4) *Contraction*: If $A \perp_m B \mid C$ and $A \perp_m D \mid (B \cup C)$, then $A \perp_m (B \cup D) \mid C$: Suppose, for contradiction, that there exists an m-connecting path between A and $B \cup D$ given C. Consider a shortest path of this type and call it π. The path π is either between A and B or between A and D. The path π being between A and B contradicts $A \perp_m B \mid C$. Therefore, π is between A and D. In addition, since all inner collider nodes on π are in $C \cup \mathrm{an}(C)$ and because $A \perp_m D \mid (B \cup C)$, an inner non-collider node should be in B. This contradicts the fact that π is a shortest m-connecting path between A and $B \cup D$ given C.

(5) *Intersection*: If $A \perp_m B \mid (C \cup D)$ and $A \perp_m D \mid (C \cup B)$, then $A \perp_m (B \cup D) \mid C$: Suppose, for contradiction, that there exists an m-connecting path between A and $B \cup D$ given C. Consider a shortest path of this type and call it π. The path π is either between A and B or between A and D. Because of symmetry between B and D in the formulation it is enough to suppose that π is between A and B. Since all inner collider nodes on π are in $C \cup \mathrm{an}(C)$ and because $A \perp_m B \mid (C \cup D)$, an inner non-collider node should be in D. This contradicts the fact that π is a shortest m-connecting path between A and $B \cup D$ given C.

(6) *Composition*: If $A \perp_m B \mid C$ and $A \perp_m D \mid C$, then $A \perp_m (B \cup D) \mid C$: Suppose, for contradiction, that there exist m-connecting paths between A and $B \cup D$ given C. Consider a path of this type and call it π. Path π is either between A and B or between A and D. Because of symmetry between B and D in the formula it is enough to suppose that π is between A and B. But this contradicts $A \perp_m B \mid C$. □

Theorem 1 implies that we can focus on establishing conditional independence for pairs of nodes, formulated in the corollary below.

Corollary 1. *For a loopless mixed graph G and disjoint subsets of its node set A, B, and C, it holds that $A \perp_m B \mid C$ if and only if $i \perp_m j \mid C$ for every nodes $i \in A$ and $j \in B$.*

Proof. The result follows from the fact that $\perp_m$ satisfies the decomposition and the composition properties. □

4. Subclasses of loopless mixed graphs

LMGs and their associated independence models induced by m-separation unify a variety of previously discussed graphical independence models.

4.1. Chain graphs

Important exceptions include certain independence models for chain graphs. Chain graphs themselves are LMGs, but at least four different Markov properties for chain graphs have been discussed in the literature. Drton [8] has classified them into (i) the *LWF* or *block concentration* Markov property, (ii) the *AMP* or *concentration regression* Markov property, (iii) a Markov property that is dual to the AMP Markov property, and (iv) and the *multivariate regression* Markov property. When the chain components consist entirely of arcs, the multivariate regression property is identical to the one induced by m-separation. However, the independence model induced by m-separation in a chain graph is typically different from any of the other chain graph interpretations; see also [22,25] and [20].

4.2. Ribbonless graphs

The class of MC graphs, defined in [18], contains line loops and uses a different separation criterion for inducing an independence model. However, a small modification of any MC graph that is derived from a DAG after marginalisation and conditioning yields a so-called ribbonless graph, which is loopless and induces the same independence model as the MC graph, but by m-separation [27]. Any ribbonless graph can be generated from a DAG by marginalisation and conditioning and ribbonless graphs are stable under these operations [26]. The remaining part of this paper deals with such graphs. We first give a formal definition of a ribbon.

A *ribbon* is a collider tripath $\langle h, i, j\rangle$ such that both of the following two conditions hold:

1. there is no endpoint-identical edge between h and j, that is, there is no hj-arc in the case of $h \leftrightarrow i \leftrightarrow j$; there is no hj-line in the case of $h \rightarrow i \leftarrow j$; and there is no arrow from h to j in the case of $h \rightarrow i \leftrightarrow j$;
2. i or a descendant of i is the endpoint of a line or is on a direction-preserving cycle.

If i or a descendant of i is the endpoint of a line, then we say the ribbon is *straight* and if they are on a direction-preserving cycle we say the ribbon is *cyclic*. A *ribbonless graph* (RG) is an LMG that has no ribbons as induced subgraphs. Figure 3 illustrates a straight ribbon $\langle h, i, j\rangle$ and the simplest cyclic ribbon.

Figure 4(a) illustrates a graph containing a straight ribbon $\langle h, i, j\rangle$ and Figure 4(b) illustrates a ribbonless graph. Notice that $\langle h, i, j\rangle$ is not a ribbon here since there is a line between h and j and this is an endpoint-identical edge. We proceed to establish that ribbonless graphs yield identical independence models to their anterior graphs and need the following lemma.

Lemma 3. *Let G be a ribbonless graph. If there is a collider tripath $\langle i, j, k\rangle$ in G that is non-collider in G^*, then G has an ik-edge that is endpoint-identical to $\langle i, j, k\rangle$.*

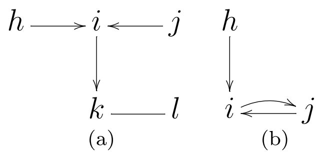


Figure 3. (a) A straight ribbon $\langle h, i, j\rangle$ with $\mathrm{ne}(i) = \varnothing$. (b) The simplest cyclic ribbon $\langle h, i, j\rangle$.

Proof. Suppose that $\langle G = G_0, G_1, \dots, G_n = G^*\rangle$ is a sequence of graphs, where each graph has been generated by removing one arrowhead pointing to a full line from the previous graph starting from G.

Consider the first intermediate graph G_{p+1} where $\langle i, j, k\rangle$ turns into a non-collider tripath. We prove by reverse induction that, for each $0 \le q \le p$, $\langle i, j, k\rangle$ is a straight ribbon unless there is an endpoint-identical ik-edge to $\langle i, j, k\rangle$.

In G_p, the node j is obviously the endpoint of a line and the result holds. Thus, we assume that the result holds for G_q. In G_{q-1}, it is easy to observe that if the line that makes the ribbon is an arrow pointing to another line or if an arrow on the direction-preserving cycle pointing to a line is an arc then j or a descendant of j is still the endpoint of a line. Therefore, the result holds in G_{q-1}. Therefore, by reverse induction, this result holds in G, and since G is ribbonless, in G there is an endpoint-identical ik-edge to $\langle i, j, k\rangle$. □

For the graph G in Figure 3(a), the anterior graph G^* is the graph where all edges become undirected. Clearly there is no endpoint-identical edge hj and the conclusion of Lemma 3 does not hold. This illustrates the role of a graph being ribbonless.

Proposition 1. *For a ribbonless graph G, it holds that $\mathcal{J}_m(G) = \mathcal{J}_m(G^*)$, that is, G and G^* are Markov equivalent.*

Proof. It is enough to prove that there is an m-connecting path between i and j given C in G if and only if there is an m-connecting path between i and j given C in G^*.

Suppose that there is an m-connecting path between i and j given C in G. All non-colliders on the path in G are preserved in G^*. In addition, by Lemma 3, a collider tripath $\langle i, j, k\rangle$ becomes non-collider if there is an endpoint-identical ik-edge to $\langle i, j, k\rangle$. In this case, the ik-edge can be used instead of $\langle i, j, k\rangle$ to establish an m-connecting path in G^*.

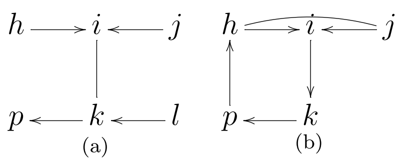


Figure 4. (a) A graph that is not ribbonless. (b) A ribbonless graph.

Conversely, suppose that there is an m-connecting path between i and j given C in G^*. Collider tripaths are collider tripaths in G, and if a non-collider tripath $\langle i, j, k\rangle$ has been collider in G then, by Lemma 3, one can again use the ik-edge instead of $\langle i, j, k\rangle$. Thus the only thing that remains to be proven is that a direction-preserving path pointing to a member of C in G remains direction-preserving in G^*.

In this case, by the same argument as in Lemma 3, if for the collider tripath $\langle i, j, k\rangle$, where $j \in \mathrm{an}(C)$, the arrowhead of an arrow on the direction-preserving path in G is taken away then $\langle i, j, k\rangle$ is a ribbon unless there is an endpoint-identical ik-edge to $\langle i, j, k\rangle$. Hence, we can use the ik-edge instead of $\langle i, j, k\rangle$ to establish an m-connecting path. □

Thus, the absence of ribbons ensures that the Markov property is unchanged by forming the anterior graph G^*. Again, as the anterior graph G^* of the graph G in Figure 3(a) is the graph with all edges becoming undirected, we have $h \perp_m j$ in G but not $h \perp_m j$ in G^*, illustrating that absence of ribbons is essential for the Markov equivalence of G and G^*.

Independence models induced by m-separation in a ribbonless graph can be induced by marginalisation over and conditioning on a DAG-independence model [26]. This implies that independence models corresponding to RGs are probabilistic, that is, any RG has a faithful probability distribution.

4.3. Other subclasses of loopless mixed graphs

Other subclasses of LMGs that use m-separation and have been discussed in the literature are *summary graphs* [30], *ancestral graphs* [24], *acyclic directed mixed graphs* [23,28], *undirected* or *concentration graphs* [6,19], *bidirected* or *covariance graphs* [5,9,15,31], and the class of directed acyclic graphs [11,16,21]. In papers on summary graphs and regression chain graphs, dashed undirected edges (without arrowheads) have often been used in place of bi-directed edges. Using the latter as we have done here makes the idea of a collider more immediate so m-separation can be used directly and the relation between the various types of graphs becomes transparent.

The use of some of the above graphs are motivated by representing independence models obtained by marginalisation over and conditioning on subsets of the node set of a DAG. For those graphs, arcs indicate marginalisation and lines indicate conditioning.

The diagram in Figure 5 illustrates the hierarchy of subclasses of LMGs and their associated independence models generated by m-separation. For example, it can be seen from the diagram that bidirected graphs are also ancestral graphs, since they form a subclass of multivariate regression chain graphs, which again form a subclass of ancestral graphs. Notice that the associated classes of independence models are all distinct except for ancestral, summary, and ribbonless graphs, which are alternative representations of the same class of independence models.

5. Maximal ribbonless graphs

Among the independence models over the node set V of a graph G, those that are of interest to us *conform* with G, meaning that $i \sim j$ in G implies $\langle i, j \mid C\rangle \notin \mathcal{J}$ for any $C \subseteq V \setminus \{i, j\}$. Henceforth, we assume that independence models $\mathcal{J}$ conform with G, unless otherwise stated.

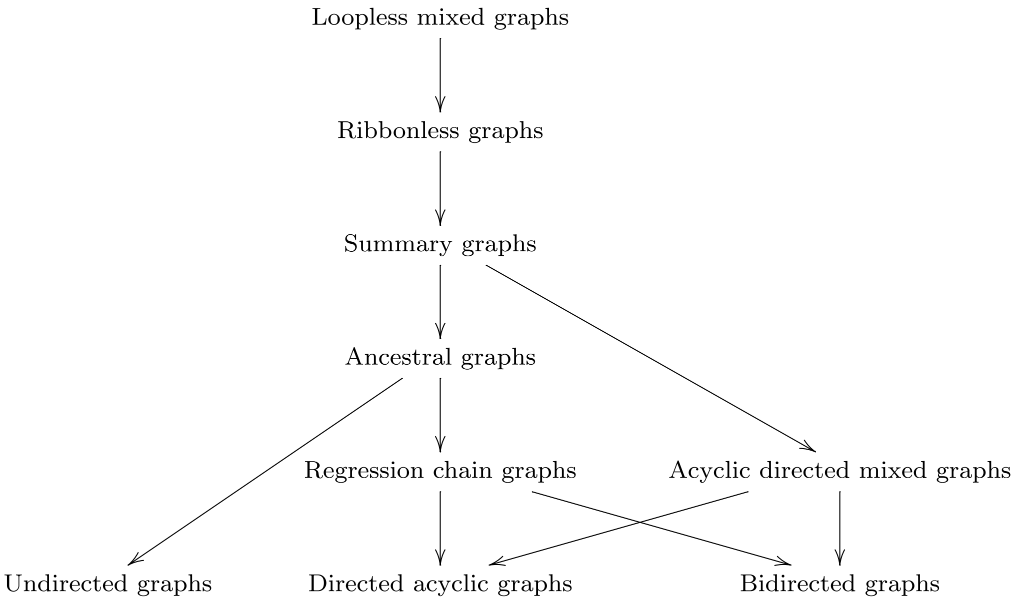


Figure 5. The hierarchy of subclasses of LMGs.

For example, the independence model $\mathcal{J} = \{\langle i, l \mid j\rangle, \langle i, k \mid \varnothing\rangle\}$ conforms with the graph G in Figure 6, whereas $\mathcal{J} = \{\langle i, l \mid j\rangle, \langle i, j \mid \varnothing\rangle\}$ does not conform with G because of the independence statement $\langle i, j \mid \varnothing\rangle$.

A ribbonless graph G is called *maximal* if by adding any edge to G, the independence model induced by m-separation changes. Note that in [30] a graph that is maximal is called an *independence graph*.

The independence models on RGs induced by m-separation conform with the graphs; hence for maximal graphs, adding an edge to the graph makes the independence model smaller. Therefore, we have the lemma below.

Lemma 4. *A graph $G = (V, E)$ is maximal if and only if for every pair of non-adjacent nodes i and j of V, there exists a subset C of $V \setminus \{i, j\}$ such that $i \perp_m j \mid C$.*

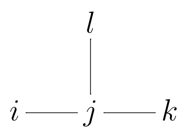


Figure 6. The independence model $\mathcal{J} = \{\langle i, l \mid j\rangle, \langle i, k \mid \varnothing\rangle\}$ conforms with G whereas $\mathcal{J} = \{\langle i, l \mid j\rangle, \langle i, j \mid \varnothing\rangle\}$ does not.

$$i \longrightarrow k \overset{\longrightarrow}{\longleftrightarrow} j$$

Figure 7. A non-maximal RG.

Proof. The result follows directly from the definition of maximality. □

RGs are not maximal in general. To see this consider the RG in Figure 7. There is no C such that $i \perp_m j \mid C$. This is because if $k \in C$, the path $i \longrightarrow k \longleftrightarrow j$ is m-connecting given C, and if $k \notin C$, $i \longrightarrow k \longrightarrow j$ is m-connecting given C.

To characterise maximal RGs, we need the following notion: A path $\langle j, q_1, q_2, \dots, q_p, i\rangle$ is a *primitive inducing* path between i and j if and only if for every n, $1 \le n \le p$,

(i) q_n is a collider on the path; and
(ii) $q_n \in \mathrm{an}(\{i\} \cup \{j\})$.

This definition is a trivial extension of a primitive inducing path as defined for ancestral graphs in [24]. Note in particular that we consider any edge between i and j to be a primitive inducing path. In Figure 7, $\langle i, k, j\rangle$ is a primitive inducing path.

Next, we need the following lemmas. These also establish a pairwise Markov property for maximal RGs.

Lemma 5. *A non-collider node k on a path π between i and j in a ribbonless graph G is either in* $\mathrm{ant}(i) \cup \mathrm{ant}(j)$ *or an anterior of a collider node h on π. Moreover, the relevant subpath of π between k and i, j or h is an anterior path in G^*.*

Proof. Let $k = i_m$ be a non-collider node on a path $\pi = \langle i = i_0, i_1, \dots, i_n = j\rangle$ in G. Then from at least one side (say from i_{m-1}) there is no arrowhead on π pointing to k. By moving towards i on the path as long as i_p, $1 \le p \le m-1$, is non-collider on the path, we obtain that $k \in \mathrm{ant}(i_{p-1})$. This implies that if no i_p is a collider then $k \in \mathrm{ant}(i)$ and hence the lemma follows. □

Lemma 6. *For nodes i and j in an RG that are not connected by any primitive inducing paths (and hence $i \not\sim j$), it holds that* $i \perp_m j \mid (\mathrm{ant}(i) \cup \mathrm{ant}(j)) \setminus \{i, j\}$.

Proof. Suppose, for contradiction, there is an m-connecting path between i and j given $(\mathrm{ant}(i) \cup \mathrm{ant}(j)) \setminus \{i, j\}$ and denote a shortest such path by π. If there is a non-collider node k on π then, by Lemma 5, k is either in $\mathrm{ant}(i) \cup \mathrm{ant}(j)$ or it is an anterior of a collider node on π. But since π is m-connecting given $(\mathrm{ant}(i) \cup \mathrm{ant}(j)) \setminus \{i, j\}$, collider nodes are in $\mathrm{ant}(i) \cup \mathrm{ant}(j)$ themselves. Hence, $k \in \mathrm{ant}(i) \cup \mathrm{ant}(j)$, which contradicts the fact that π is m-connecting. Therefore, all inner nodes of π must be colliders.

Now we know that all inner nodes of π are in $\mathrm{ant}(i) \cup \mathrm{ant}(j)$ and $i \not\sim j$. If, for a collider tripath $\langle r, l, s\rangle$ on π, $l \in (\mathrm{ant}(i) \cup \mathrm{ant}(j)) \setminus (\mathrm{an}(i) \cup \mathrm{an}(j))$ then, by Lemma 2 and since the graph is ribbonless, there is an endpoint-identical rs-edge to the tripath, which contradicts π being shortest. Therefore, $l \in \mathrm{an}(i) \cup \mathrm{an}(j)$, which implies that π is primitive inducing, again a contradiction. Therefore, there is no m-connecting path between i and j given $(\mathrm{ant}(i) \cup \mathrm{ant}(j)) \setminus \{i, j\}$, and hence $i \perp_m j \mid (\mathrm{ant}(i) \cup \mathrm{ant}(j)) \setminus \{i, j\}$. □

Next, in Theorem 2 we give a necessary and sufficient condition for an RG to be maximal. The analogous result for ancestral graphs was proved in Theorem 4.2 of [24].

Theorem 2. *A ribbonless graph G is maximal if and only if G does not contain any primitive inducing paths between non-adjacent nodes.*

Proof. Let $\pi = \langle i = i_0, i_1, \dots, i_n = j\rangle$ be a primitive inducing path between i and j in H, and let C be a subset $V \setminus \{i, j\}$, where V is the node set of H. We need to show that there is an m-connecting path between i and j given C.

This is immediate if each internal node, that is, each of $i_1, \dots, i_{n-1}$, is in $C \cup \mathrm{an}(C)$ by just using π, so assume that this is not the case. Thus there is an internal node of π not in $C \cup \mathrm{an}(C)$, and we may assume that there is one in $\mathrm{an}(i)$. Pick such a node i_q, $1 \le q < n$, as far along the path to j as possible. Consider a direction-preserving path from i_q to i, and let P_1 denote the reverse of this path. Note that no internal node in P_1 is in $C \cup \mathrm{an}(C)$. Let π_1 be the part of π from i_q to j. If each internal node in this path is in $C \cup \mathrm{an}(C)$ then we are done by taking the path P_1 followed by π_1 (note that no node can be repeated since each internal node in π_1 is in $C \cup \mathrm{an}(C)$ and each internal node in P_1 is outside $C \cup \mathrm{an}(C)$). So suppose not. Let i_p be the first node in π_1 that is not in $C \cup \mathrm{an}(C)$. Then $i_p \notin \mathrm{an}(i)$ (by the way i_q was chosen), so $i_p \in \mathrm{an}(j)$. Let π_2 be the part of π from i_q to i_p, and let P_2 be a direction-preserving path from i_p to j. Note that no internal node in P_2 is in $C \cup \mathrm{an}(C)$. If P_1 and P_2 have no intersection, then much as above we obtain an m-connecting path given C by taking P_1 followed by π_2, followed by P_2. If P_1 and P_2 do intersect, then we obtain an m-connecting path as required by following P_1 up to the first node on P_2 and then following P_2.

By letting $C = (\mathrm{ant}(i) \cup \mathrm{ant}(j)) \setminus \{i, j\}$ for every non-adjacent nodes i and j, the other direction follows from Lemmas 4 and 6. □

For other special types of graphs that are subclasses of RGs, the condition for maximality of RGs may get further simplified. Among the subclasses of RGs that have been mentioned in this paper, summary graphs, ancestral graphs, and acyclic directed mixed graphs are not necessarily maximal, while all others are maximal. This can be seen by checking whether primitive inducing paths are permissible in each subclass.

A Markov equivalent maximal graph can be generated from a non-maximal graph by adding endpoint-identical edges to a primitive inducing path between a pair of non-adjacent nodes. We refer the reader to [27] for details. The following lemma establishes that anterior graphs of maximal graphs are themselves maximal.

Lemma 7. *Let G be a ribbonless graph and G^* its anterior graph. Then if G is maximal, so is G^*.*

Proof. If, for contradiction, G^* is not maximal, then Theorem 2 implies that there is a primitive inducing path in G^* between non-adjacent nodes i and j. Consider a shortest primitive inducing path between i and j and denote it by π. We know that all inner nodes of π are colliders in G^*. This trivially implies that all inner nodes of π are colliders in G too. In addition, each inner node k on π is in $\mathrm{an}(\{i, j\})$ in G^*. In G, $k \in \mathrm{an}(\{i, j\})$ unless an arrow on the direction-preserving path from k to i or j is an arc turning into an arrow in G^*. In this case, k is an ancestor of a node that is the endpoint of a line. Hence the tripath $\langle h, k, l\rangle$ on π is a ribbon unless there is an endpoint-identical hl-edge to the tripath, which contradicts the fact that π is shortest. Therefore, π is a primitive inducing path in G, a contradiction. Hence, G^* is maximal. $\square$

6. Markov properties for ribbonless graphs

In this section, we give a precise definition of the global and pairwise Markov properties for an independence model $\mathcal{J}$ defined over the node set of a ribbonless graph. Further we show that these two Markov properties are equivalent for a maximal ribbonless graph if $\mathcal{J}$ is also a compositional graphoid. This result is a direct generalisation of the similar result of [21] for undirected graphs and graphoids.

6.1. Global and pairwise Markov properties

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

$$A \perp_m B \mid C \quad \Longrightarrow \quad \langle A, B \mid C\rangle \in \mathcal{J}.$$

Similarly, an independence model $\mathcal{J}$ defined over V satisfies the *pairwise Markov property* w.r.t. G if it holds for any nodes i and j that

$$i \not\sim j \quad \Longrightarrow \quad \big\langle i, j \mid \big(\mathrm{ant}(i) \cup \mathrm{ant}(j)\big) \setminus \{i, j\}\big\rangle \in \mathcal{J}.$$

For example, for the graph in Figure 8, the pairwise Markov property would imply that $\langle i, m \mid \{k, l, h\}\rangle$ as $\mathrm{ant}(i) = \{k, l, h, m\}$ and $\mathrm{ant}(m) = \{l, h\}$. It would also imply that $\langle l, p \mid \{h, m\}\rangle$.

Clearly, the independence model $\mathcal{J}_m(G)$ induced by m-separation always satisfies the global Markov property w.r.t. G. By Lemma 4, Lemma 6, and Theorem 2, $\mathcal{J}_m(G)$ *satisfies the pairwise Markov property if and only if G is maximal*.

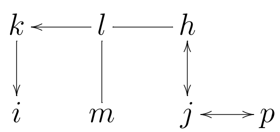


Figure 8. The pairwise Markov property for this RG implies, for example, $\langle i, m \mid \{k, l, h\}\rangle$. The global Markov property would for example imply $\langle \{i, k\}, j \mid l\rangle$.

6.2. Equivalence of pairwise and global Markov properties

Before establishing the main result of this section, we need two lemmas.

Lemma 8. *Let $\mathcal{J}$ be a compositional graphoid over a set V and M and C be disjoint subsets of V. It then holds that the marginal independence model*

$$\alpha(\mathcal{J}, M) = \big\{\langle A, B \mid C\rangle\colon \langle A, B \mid C\rangle \in \mathcal{J} \text{ and } (A \cup B \cup C) \cap M = \varnothing\big\},$$

which is defined over $V \setminus M$, is a compositional graphoid.

Proof. All the six compositional graphoid properties for $\alpha(\mathcal{J}, M)$ follow trivially from the facts that for A, B, and C such that $(A \cup B \cup C) \cap M = \varnothing$, $\langle A, B \mid C\rangle \in \alpha(\mathcal{J}, M)$ if and only if $\langle A, B \mid C\rangle \in \mathcal{J}$, and $\mathcal{J}$ satisfies the six properties. □

Notice that the notion of a marginal independence model $\alpha(\mathcal{J}, M)$ is identical to the notion formally defined in [24] with a different notation; it was also discussed in [26] with the same notation as in this paper.

The following lemma gives sufficient conditions for the combination of two m-connecting paths in anterior graphs to be m-connecting.

Lemma 9. *Let G^* be the anterior graph of a ribbonless graph G and suppose that there are paths $\pi_1 = \langle i = i_0, i_1, \dots, i_n, h\rangle$ between i and h and $\pi_2 = \langle h, j_m, j_{m-1}, \dots, j_0 = j\rangle$ between h and j which are m-connecting given C. The combination $\pi_{12} = \pi_1 \circ \pi_2$ is then an m-connecting path between i and j given C in each of the following mutually exclusive situations*:

- (a1) *$\langle i_n, h, j_m\rangle$ is a collider and $h \in C \cup \mathrm{an}(C)$;*
- (a2) *$i_n = j_m$ with an arrowhead pointing to h on the $i_n h$-edge and $h \in C \cup \mathrm{an}(C)$;*
- (b1) *$\langle i_n, h, j_m\rangle$ is a non-collider and $h \notin C$;*
- (b2) *$i_n = j_m$ with no arrowhead pointing to h on the $i_n h$-edge.*

Proof. Let $\pi_{12} = \pi_1 \circ \pi_2 = \langle i, \dots i_{p-1}, k, j_{q-1}, \dots, j\rangle$ be the combination of π_1 and π_2. If $k = h$ and either (a1) or (b1) holds then the conclusion is obvious. The cases (a2) or (b2) are only relevant when $k \neq h$.

Next consider the situation where $k \neq h$. Since π_1 and π_2 are m-connecting, for π_{12} to be m-connecting we only need to check the tripath $\langle i_{p-1}, k, j_{q-1}\rangle$. We have to deal with two cases:

Case 1: $\langle i_{p-1}, k, j_{q-1}\rangle$ is a non-collider.

In this case there is no arrowhead pointing to k from at least one of i_{p-1} or j_{q-1}. This means that $\langle i_{p-1}, k, i_{p+1}\rangle$ on π_1 or $\langle j_{q-1}, k, j_{q+1}\rangle$ on π_2 is a non-collider, and since π_1 and π_2 were both m-connecting we have $k \notin C$. Hence π_{12} is m-connecting.

Case 2: $\langle i_{p-1}, k, j_{q-1}\rangle$ is a collider. We need to consider the following two subcases:

Case 2.1. If $\langle i_{p-1}, k, j_{q-1}\rangle$ is a collider and any of $\langle i_{p-1}, k, i_{p+1}\rangle$ or $\langle j_{q-1}, k, j_{q+1}\rangle$ is also a collider then $k \in C \cup \mathrm{an}(C)$ and π_{12} is m-connecting.

Case 2.2. If $\langle i_{p-1}, k, j_{q-1}\rangle$ is a collider but $\langle i_{p-1}, k, i_{p+1}\rangle$ and $\langle j_{q-1}, k, j_{q+1}\rangle$ are both non-colliders then by Lemma 5, the subpath of π_1 from k to a collider node l_1 or to h is an anterior path and similarly for π_2, l_2, and h. However, since G^* is an anterior graph and there are arrowheads pointing to k, these anterior paths must be direction-preserving and thus $k \in \mathrm{an}(l_1) \cup \mathrm{an}(h)$ and $k \in \mathrm{an}(l_2) \cup \mathrm{an}(h)$. Now we have the two following further subcases:

Case 2.2.1: One of the subpaths of π_1, π_2 from k to l_1, l_2 is direction-preserving. Because π_1 and π_2 are m-connecting we must have l_1 or l_2 in $C \cup \mathrm{an}(C)$. Thus, $k \in \mathrm{an}(C)$ and π_{12} is m-connecting.

Case 2.2.2: Both subpaths of π_1 and π_2 from k to h are direction-preserving. Then $\langle i_n, h, j_m\rangle$ is collider or $i_n = j_m$ with an arrowhead pointing to h on the $i_n h$-edge and (b1) and (b2) are impossible. If (a1) or (a2) holds π_{12} is m-connecting since then $h \in C \cup \mathrm{an}(C)$. □

We are now ready to establish the main result of this paper.

Theorem 3. *Let G be a maximal ribbonless graph. If an independence model $\mathcal{J}$ over the node set of G is a compositional graphoid, then $\mathcal{J}$ satisfies the pairwise Markov property w.r.t. G if and only if it satisfies the global Markov property w.r.t. G.*

Proof. ($\Leftarrow$) If $\mathcal{J}$ is a compositional graphoid and satisfies the global Markov property it follows from Theorem 2 and Lemma 6 that it satisfies the pairwise Markov property.

($\Rightarrow$) Now suppose that $\mathcal{J}$ satisfies the pairwise Markov property and compositional graphoid axioms. For subsets A, B, and C of the node set of G, we should prove that $A \perp_m B \mid C$ implies $\langle A, B \mid C\rangle \in \mathcal{J}$. By composition, it is sufficient to show this when A and B are singletons, that is, that $i \perp_m j \mid C$ implies $\langle i, j \mid C\rangle \in \mathcal{J}$.

Further we observe that it is sufficient to establish the result in the case when $G = G^*$ is itself an anterior graph. Proposition 1 gives that $A \perp_m B \mid C$ in G, which implies $A \perp_m B \mid C$ in G^*. In addition, by Lemma 7, G^* is a maximal graph. Moreover, G and G^* have the same anterior sets, and therefore the same pairwise Markov property. Thus in the following, we assume that $G = G^*$ is an anterior graph.

We prove the result in two main parts. In part I, we prove the result for the case that $C \subseteq \mathrm{ant}(i) \cup \mathrm{ant}(j)$. In part II, we use the result of part I to establish the general case.

Part I. Suppose that $C \subseteq \mathrm{ant}(i) \cup \mathrm{ant}(j)$. We use induction on the number of nodes of the graph. The induction base for a graph with two nodes is trivial. Thus, suppose that the result holds for all anterior graphs with fewer than n nodes and assume that G^* has n nodes.

Let $D = \{i\} \cup \{j\} \cup \mathrm{ant}(i) \cup \mathrm{ant}(j)$ and $M = V \setminus D$, where V is the node set of the graph. First in case I.1 we suppose that $M \neq \varnothing$, and then in case I.2 we suppose that $M = \varnothing$.

Case I.1. Consider $G^*[D]$ to be the subgraph induced by D. Consider the marginal independence model $\alpha(\mathcal{J}, M) = \{\langle A, B \mid C\rangle\colon \langle A, B \mid C\rangle \in \mathcal{J} \text{ and } (A \cup B \cup C) \cap M = \varnothing\}$ defined over D. By Lemma 8, $\alpha(\mathcal{J}, M)$ is a compositional graphoid. In addition, it satisfies the pairwise Markov property: This is because two non-adjacent nodes l_1 and l_2 in $G^*[D]$ are non-adjacent in G^* and by the pairwise Markov property for $\mathcal{J}$, $\langle l_1, l_2 \mid (\mathrm{ant}_{G^*}(l_1) \cup \mathrm{ant}_{G^*}(l_2)) \setminus \{l_1, l_2\}\rangle \in \mathcal{J}$, where ant_{G^*} is the anterior set in G^*. We know that $\mathrm{ant}_{G^*}(l_1) \cup \mathrm{ant}_{G^*}(l_2) \subseteq D$ and hence $\mathrm{ant}_{G^*}(l_1) \cup \mathrm{ant}_{G^*}(l_2) \cap M = \varnothing$. In addition, for a node l in $G^*[D]$, $\mathrm{ant}_{G^*}(l) = \mathrm{ant}_{G^*[D]}(l)$. Therefore, $\langle l_1, l_2 \mid (\mathrm{ant}_{G^*[D]}(l_1) \cup \mathrm{ant}_{G^*[D]}(l_2)) \setminus \{l_1, l_2\}\rangle \in \alpha(\mathcal{J}, M)$.

We also know that $i \perp_m j \mid C$ in G^* implies $i \perp_m j \mid C$ in $G^*[D]$ since there is no m-connecting path between i and j given C in G^* and by removing nodes and edges from G^*

no new m-connecting paths are generated. Therefore, by the induction hypothesis $\langle i, j \mid C\rangle \in \alpha(\mathcal{J}, M)$. This implies that $\langle i, j \mid C\rangle \in \mathcal{J}$.

Case I.2. Now suppose that $M = \varnothing$ and thus the node set of G^* is $D = \{i\} \cup \{j\} \cup \mathrm{ant}(i) \cup \mathrm{ant}(j)$. We prove the result by reverse induction on $|C|$: For the base, $C = V \setminus \{i, j\} = (\mathrm{ant}(i) \cup \mathrm{ant}(j)) \setminus \{i, j\}$ and the result follows trivially from the pairwise Markov property.

For the inductive step, consider a node $h \notin C$. We want to show that h is not simultaneously m-connected to both i and j: Suppose, for contradiction, there are m-connecting paths $\pi_1 = \langle i, i_1, \dots, i_n, h\rangle$ and $\pi_2 = \langle h, j_m, j_{m-1}, \dots, j_0 = j\rangle$ given C. If (b1) or (b2) of Lemma 9 hold then i and j are m-connected given C which contradicts $i \perp_m j \mid C$. So we need only consider the cases where $\langle i_n, h, j_m\rangle$ is collider or $i_n = j_m$ with an arrowhead pointing to h on the $i_n h$-edge. However, we know that $h \in \mathrm{ant}(i)$ or $h \in \mathrm{ant}(j)$. Because of symmetry between i and j suppose that $h \in \mathrm{ant}(i)$. Since G^* is an anterior graph and there is an arrowhead pointing to h we have $h \in \mathrm{an}(i)$. Hence, there is a direction-preserving path π from h to i. If no node on π is in C then (b1) or (b2) of Lemma 9 implies that the combination of π and π_2 is an m-connecting path between i and j, again a contradiction. If there is a node on π that is in C then $h \in \mathrm{an}(C)$ and again, by (a1) and (a2) of Lemma 9, i and j are m-connected given C, again a contradiction.

We conclude that, given C, h is not m-connected to both i and j. By symmetry, suppose that $i \perp_m h \mid C$.

We also have that $i \perp_m j \mid C$. Since $\mathcal{J}_m(G^*)$ is a compositional graphoid (Theorem 1) the composition property gives that $i \perp_m \{j, h\} \mid C$. By weak union for $\perp_m$ we obtain $i \perp_m j \mid \{h\} \cup C$ and $i \perp_m h \mid \{j\} \cup C$. By the induction hypothesis, we obtain $\langle i, j \mid \{h\} \cup C\rangle \in \mathcal{J}$ and $\langle i, h \mid \{j\} \cup C\rangle \in \mathcal{J}$. By intersection, we get $\langle i, \{j, h\} \mid C\rangle \in \mathcal{J}$. By decomposition we finally obtain $\langle i, j \mid C\rangle \in \mathcal{J}$.

Part II. We now prove the result in the general case by induction on $|C|$. The base, that is, the case that $|C| = 0$, follows from part I. To prove the inductive step, we can assume that $C \nsubseteq \mathrm{ant}(i) \cup \mathrm{ant}(j)$, since otherwise part I implies the result.

We first show that if $C \nsubseteq \mathrm{ant}(i) \cup \mathrm{ant}(j)$ then there is a node l in C such that $i \perp_m j \mid C \setminus \{l\}$: Let first $l' \in C \setminus (\mathrm{ant}(i) \cup \mathrm{ant}(j))$ be arbitrary. If there is an $l'' \in C \setminus (\mathrm{ant}(i) \cup \mathrm{ant}(j))$ so that $l' \in \mathrm{ant}(l'')$ and $l'' \notin \mathrm{ant}(l')$ then replace l' by l'', and repeat this process until it terminates, the latter being ensured by transitivity of ant (Lemma 1) and the finiteness of C. Thus, we eventually obtain an l so that if $l \in \mathrm{ant}(\tilde{l})$ for $\tilde{l} \in C \setminus (\mathrm{ant}(i) \cup \mathrm{ant}(j))$ then we also have $\tilde{l} \in \mathrm{ant}(l)$.

Suppose, for contradiction, that there is a shortest m-connecting path π between i and j given $C \setminus \{l\}$. If l is not on π or is a collider on π then π is also m-connecting given C. Therefore, l is a non-collider on π. This, together with $l \notin \mathrm{ant}(i) \cup \mathrm{ant}(j)$, by using Lemma 5, implies that l is an anterior of a collider node p on π. Since π is m-connecting, $p \in C \cup \mathrm{an}(C)$. Thus, there is an $\tilde{l} \in C$ so that $p = \tilde{l}$ or $p \in \mathrm{an}(\tilde{l})$. Transitivity of anterior sets and the fact that $l \notin (\mathrm{ant}(i) \cup \mathrm{ant}(j))$ now imply that $\tilde{l} \in C \setminus (\mathrm{ant}(i) \cup \mathrm{ant}(j))$. The construction of l implies $\tilde{l} \in \mathrm{ant}(l)$ which again implies that $\tilde{l} \in \mathrm{an}(l)$ and $l \in \mathrm{an}(\tilde{l})$ and thus the collider tripath containing p is a cyclic ribbon unless its endpoints are adjacent with an endpoint-identical edge, which implies that π is not a shortest m-connecting path, a contradiction.

We now have that either $i \perp_m l \mid C \setminus \{l\}$ or $j \perp_m l \mid C \setminus \{l\}$ since otherwise, by Lemma 9 there is an m-connecting path between i and j given $C \setminus \{l\}$ in the case that l is a non-collider or given C in the case that l is a collider node. Because of symmetry suppose that $i \perp_m l \mid C \setminus \{l\}$. By the induction hypothesis, we have $\langle i, j \mid C \setminus \{l\}\rangle \in \mathcal{J}$ and $\langle i, l \mid C \setminus \{l\}\rangle \in \mathcal{J}$. By the composition property we get $\langle i, \{j, l\} \mid C \setminus \{l\}\rangle \in \mathcal{J}$. The weak union property implies $\langle i, j \mid C\rangle \in \mathcal{J}$. □

If we specialise Theorem 3 to the most common case of probabilistic independence models, we get the following corollary.

Corollary 2. *Let G be a maximal ribbonless graph. A probabilistic independence model that satisfies the intersection and composition axioms satisfies the pairwise Markov property w.r.t. G if and only if it satisfies the global Markov property w.r.t. G.*

6.3. Necessity of compositional graphoid axioms

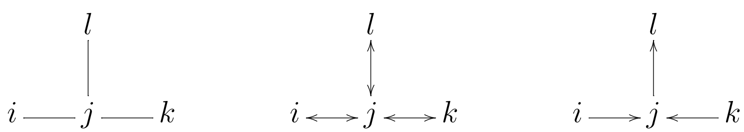
Theorem 3 states that, for equivalence of pairwise and global Markov properties, the six compositional graphoid axioms are sufficient. In fact, in general, for the mentioned equivalence, all six axioms are also necessary. The graphs in Figure 9 show that the intersection and composition properties are necessary for the equivalence of pairwise and global Markov properties.

For $G_1 = (V_1, E_1)$, if $\mathcal{J}_1$ defined over V_1 satisfies the pairwise Markov property, then $\langle i, k \mid \{j, l\}\rangle$, $\langle i, l \mid \{j, k\}\rangle$, and $\langle k, l \mid \{i, j\}\rangle$ are in $\mathcal{J}_1$. It can be seen that none of the compositional semi-graphoid axioms can be used to imply $\langle i, \{k, l\} \mid j\rangle \in \mathcal{J}_1$. The intersection property is the only axiom that implies the result.

For $G_2 = (V_2, E_2)$, if $\mathcal{J}_2$ defined over V_2 satisfies the pairwise Markov property then $\langle i, k \mid \varnothing\rangle$, $\langle i, l \mid \varnothing\rangle$, and $\langle k, l \mid \varnothing\rangle$ are in $\mathcal{J}_2$. It can be seen that none of the graphoid axioms can be used to imply $\langle i, \{k, l\} \mid \varnothing\rangle \in \mathcal{J}_2$. The composition property is the only axiom that implies the result.

For $G_3 = (V_3, E_3)$, if $\mathcal{J}_3$ defined over V_3 satisfies the pairwise Markov property then $\langle i, k \mid \varnothing\rangle$, $\langle i, l \mid \{j, k\}\rangle$, and $\langle k, l \mid \{i, j\}\rangle$ are in $\mathcal{J}_3$. It can be seen that none of the compositional semi-graphoid axioms can be used to imply $\langle l, \{i, k\} \mid j\rangle \in \mathcal{J}_3$. The intersection property is the only axiom that implies the result. See also for example, Example 3.26 of [19], showing that the pairwise Markov property does not imply the global Markov property for DAGs when intersection is violated.

It is known that, for undirected graphs, the five graphoid axioms are necessary and sufficient for equivalence of pairwise and global Markov properties; see [19]. For bidirected graphs, the independence statement associated with a missing edge between nodes i and j is $\langle i, j \mid \varnothing\rangle$ and only the five compositional semi-graphoid axioms are necessary for equivalence of pairwise and global Markov properties. This can be inferred from the proof of Theorem 3, since part I of the



(a) (b) (c)

Figure 9. For the equivalence of pairwise and global Markov properties, (a) an undirected graph G_1 that shows that the intersection property is necessary; (b) a bidirected graph G_2 that shows that the composition property is necessary; (c) a directed acyclic graph G_3 that shows that the intersection property is necessary.

proof is not relevant for bidirected graphs unless $C = \varnothing$ and the intersection property is not used in part II of the proof. We conclude by stating this as its own proposition.

Proposition 2. *Let $G = (V, E)$ be a bidirected graph. If an independence model $\mathcal{J}$ defined over V is a compositional semi-graphoid then $\mathcal{J}$ satisfies the pairwise Markov property w.r.t. G if and only if it satisfies the global Markov property w.r.t. G.*

Acknowledgements

We are grateful to Milan Studený, Nanny Wermuth, and anonymous referees for very helpful comments on earlier versions of this paper.

References

[1] Andersson, S.A., Madigan, D. and Perlman, M.D. (1997). A characterization of Markov equivalence classes for acyclic digraphs. *Ann. Statist.* **25** 505–541. MR1439312

[2] Andersson, S.A., Madigan, D. and Perlman, M.D. (2001). Alternative Markov properties for chain graphs. *Scand. J. Statist.* **28** 33–85. MR1844349

[3] Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *J. Roy. Statist. Soc. Ser. B* **36** 192–236. With discussion by D.R. Cox, A.G. Hawkes, P. Clifford, P. Whittle, K. Ord, R. Mead, J.M. Hammersley and M.S. Bartlett and with a reply by the author. MR0373208

[4] Clifford, P. (1990). Markov random fields in statistics. In *Disorder in Physical Systems. Oxford Sci. Publ.* 19–32. New York: Oxford Univ. Press. MR1064553

[5] Cox, D.R. and Wermuth, N. (1993). Linear dependencies represented by chain graphs. *Statist. Sci.* **8** 204–218, 247–283. With comments and a rejoinder by the authors. MR1243593

[6] Darroch, J.N., Lauritzen, S.L. and Speed, T.P. (1980). Markov fields and log-linear interaction models for contingency tables. *Ann. Statist.* **8** 522–539. MR0568718

[7] Dawid, A.P. (1979). Conditional independence in statistical theory. *J. Roy. Statist. Soc. Ser. B* **41** 1–31. MR0535541

[8] Drton, M. (2009). Discrete chain graph models. *Bernoulli* **15** 736–753. MR2555197

[9] Drton, M. and Richardson, T.S. (2008). Binary models for marginal independence. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **70** 287–309. MR2424754

[10] Frydenberg, M. (1990). The chain graph Markov property. *Scand. J. Statist.* **17** 333–353. MR1096723

[11] Geiger, D., Verma, T. and Pearl, J. (1990). Identifying independence in Bayesian networks. *Networks* **20** 507–534. MR1064736

[12] Grimmett, G.R. (1973). A theorem about random fields. *Bull. London Math. Soc.* **5** 81–84. MR0329039

[13] Hammersley, J.M. and Clifford, P. (1971) Markov fields on finite graphs and lattices. Unpublished manuscript.

[14] Kang, C. and Tian, J. (2009). Markov properties for linear causal models with correlated errors. *J. Mach. Learn. Res.* **10** 41–70.

[15] Kauermann, G. (1996). On a dualization of graphical Gaussian models. *Scand. J. Statist.* **23** 105–116. MR1380485

[16] Kiiveri, H., Speed, T.P. and Carlin, J.B. (1984). Recursive causal models. *J. Austral. Math. Soc. Ser. A* **36** 30–52. MR0719999

[17] Koster, J.T.A. (1997). Gibbs and Markov properties of graphs. *Ann. Math. Artificial Intelligence* **21** 13–26. MR1479006
[18] Koster, J.T.A. (2002). Marginalizing and conditioning in graphical models. *Bernoulli* **8** 817–840. MR1963663
[19] Lauritzen, S.L. (1996). *Graphical Models. Oxford Statistical Science Series* **17**. New York: Clarendon Press. MR1419991
[20] Lauritzen, S.L. and Richardson, T.S. (2002). Chain graph models and their causal interpretations. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **64** 321–361. MR1924296
[21] Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. The Morgan Kaufmann Series in Representation and Reasoning*. San Mateo, CA: Morgan Kaufmann. MR0965765
[22] Richardson, T. (2001) Chain graphs which are maximal ancestral graphs are recursive causal graphs. Technical Report 387, Dept. Statistics, Univ. Washington, Seattle, WA.
[23] Richardson, T. (2003). Markov properties for acyclic directed mixed graphs. *Scand. J. Statist.* **30** 145–157. MR1963898
[24] Richardson, T. and Spirtes, P. (2002). Ancestral graph Markov models. *Ann. Statist.* **30** 962–1030. MR1926166
[25] Richardson, T.S. (1998). Chain graphs and symmetric associations. In *Learning in Graphical Models* (M. Jordan, ed.) 231–260. Dordrecht, The Netherlands: Kluwer.
[26] Sadeghi, K. (2013). Stable mixed graphs. *Bernoulli* **19** 2330–2358.
[27] Sadeghi, K. (2012). Graphical representation of independence structures. Ph.D. thesis, Univ. Oxford.
[28] Spirtes, P., Richardson, T. and Meek, C. (1997). The dimensionality of mixed ancestral graphs. Technical Report CMU-PHIL-83, Dept. Philosophy, Carnegie–Mellon Univ., Pittsburgh, PA.
[29] Studený, M. (1989). Multi-information and the problem of characterization of conditional independence relations. *Problems Control Inform. Theory/Problemy Upravlen. Teor. Inform.* **18** 3–16. MR0986211
[30] Wermuth, N. (2011). Probability distributions with summary graph structure. *Bernoulli* **17** 845–879. MR2817608
[31] Wermuth, N. and Cox, D.R. (1998). On association models defined over independence graphs. *Bernoulli* **4** 477–495. MR1679794
[32] Wermuth, N., Marchetti, G.M. and Cox, D.R. (2009). Triangular systems for symmetric binary variables. *Electron. J. Stat.* **3** 932–955. MR2540847

Received May 2012 and revised October 2012