Bernoulli **8**(6), 2002, 817–840

Marginalizing and conditioning in graphical models

JAN T. A. KOSTER

Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands.
E-mail: koster@fsw.eur.nl

A class of graphs is introduced which is closed under marginalizing and conditioning. It is shown that these operations can be executed by performing in arbitrary order a sequence of simple, strictly local operations on the graph at hand. The results are based on a simplification of J. Pearl's notion of *d*-separation. As the simplification does not change the separation properties of graphs for which the original *d*-separation concept is applicable (e.g., directed graphs), it constitutes a true generalization of the latter concept to the present class of graphs.

Keywords: Gaussian linear equations system; graph; graphical Markov model; linear structural equations system; Markov property; Markov random field; MC graph; non-recursive causal model; separation

1. Introduction

The idea that 'conditional independence' is a central and unifying notion for many concepts and techniques in applied multivariate statistics was expressed explicitly in Dawid (1979). The full potential of this idea becomes apparent when graphs are used to represent the Markov properties of the random variables studied – that is, a graph is introduced whose vertices correspond to the variables at hand and whose separation properties reflect the presumed conditional independence relations between the variables (Lauritzen 1979; Darroch *et al*. 1980). The Markov properties are exploited, for example, in reducing the complexity encountered in estimating the joint distribution of a large number of random variables. At first the graphs defining graphical Markov models were undirected, but soon directed acyclic graphs (DAGs) and chain graphs were used to model systems of random variables in which either the flow of time or the presence of cause–effect relationships induces a fully or partially directed process (Pearl 1988; Lauritzen and Wermuth 1989; Frydenberg 1990). Thus for some time the scope of graphical Markov models remained restricted to recursive processes or to systems in which directionality was irrelevant. However, Spirtes (1995) and Koster (1996) showed independently that separation properties of graphs containing directed cycles can be defined while maintaining consistency with the use of such graphs in Gaussian structural equation modelling (where they are called path diagrams). As a consequence, it can now be said that graphical Markov models provide a consistent common framework for certain techniques belonging to applied multivariate statistics, such as covariance selection modelling, factor analysis, path analysis, log-linear analysis, (recursive and non-recursive)

1350–7265 © 2002 ISI/BS

Gaussian structural equation modelling, latent class analysis, etc. The essential feature of the graphical modelling approach to multivariate statistics is to merge ‘conditional independence’ with graphs. Let us briefly summarize some of the advantages of this approach.

First, a graphical model gives in a concise way a complete picture of the scope of the analysis as it shows explicitly which variables are studied (i.e., which marginal probability distribution is studied). Second, the graphical model allows one to separate Markovian probabilistic aspects (i.e., its Markov properties, entailed by the presence or absence in the graph of edges between variables and by the symmetric or asymmetric nature of these edges) from other statistical aspects of the model (e.g., assumptions concerning the parametric family, presumed linearity, and choice of link function). Third, when a positive density exists, the graphical model is equivalent to a certain factorization of the density. This so-called Gibbs factorization relative to the graph often enables one to perform the statistical analysis by splitting it up into a set of local analyses. Fourth, as graphical models are by their nature easy to visualize, they also provide a powerful tool for communication with the statistical layman – a practical advantage that cannot easily be overestimated. Finally, there are close similarities, both formally and graphically, with methods used in related fields, such as expert systems and decision analysis, where inference networks and influence diagrams are representations of Bayesian networks (Lauritzen and Spiegelhalter 1988; Pearl 1988; Jordan 1998; Cowell *et al*. 1999; see also Studený 1993 for a discussion of formal similarities with other frameworks). The common language of graphical models stimulates cross-fertilization between these different fields of research.

Even though the types of graph encountered in graphical Markov models – undirected graphs, DAGs, chain graphs and recursive or non-recursive path diagrams – look quite dissimilar, the way in which for any such graph the associated graphical Markov model is defined is essentially the same. For each type of graph a purely graph-theoretical concept of ‘separation’ is introduced, allowing one to state whether or not in a certain graph two given subsets of vertices are separated by a given third subset of vertices. Then the vertices of the graph are associated (in one-to-one fashion) with the random variables that are studied. Finally, the graphical Markov model is defined by stipulating that each valid separation statement pertaining to vertices of the graph be mirrored by a valid conditional independence statement pertaining to the associated random variables.

The main objective of the present paper is to show that the symbolism of graphs is rich enough to deal straightforwardly with problems of marginalization and conditioning – thereby strengthening the argument just given that graphical models constitute a convenient framework for applied multivariate statistics. In Wermuth *et al*. (1998; see also Cox and Wermuth 1996, Chapter 8) the problem of marginalizing and conditioning in graphical models (defined by a DAG) was phrased as follows:

> Suppose that we are given a [graphical model], defined over a set V of [variables]. We divide V into three nonoverlapping components (s, c and m) and consider the distribution of the variables in s, conditionally on those in c, i.e., marginalizing over those in m. [. . .] We look for a graphical representation of the new distribution deduced from [the graph defining the original model].

A more formal treatment based on the notion of a minor is given in Matúš (1997). In the present paper this problem is studied for models defined by graphs belonging to a very general class. It will be shown that analysis of the Markovian consequences of marginalizing and conditioning is possible by performing a sequence of elementary, local operations on the defining graph only. To this end we introduce a new type of graph (so-called MC graphs) and define for any such graph a concept of separation of subsets of vertices. Currently there are in the literature two main lines of approach in defining 'separation in graphs'. One approach, defined in Frydenberg (1990) for the class of chain graphs, is based on the operation of 'moralization' of subgraphs induced by certain subsets of vertices.[1] In Koster (1997) this is generalized to the class of so called 'general graphs', that is, graphs which may have, between each pair of distinct vertices i and j, any subset of $\{i \to j, i \leftarrow j, i - j\}$ as edges. The other approach, introduced by Pearl (1988), is based on the notion of d-separation and applies to DAGs. In d-separation two conditions are stated which determine whether a path in the graph is either blocked by a certain subset of vertices, or open given this subset of vertices. The relevant concepts can also be applied to directed cyclic graphs (Spirtes 1995) simply by extending their definition to such graphs. In the same manner their scope was enlarged still further in Koster (1999). Thus, d-separation became applicable to what in the present paper are called 'directional graphs' (see also Spirtes *et al*. 1998; Richardson 1999), that is, graphs which may have between each pair of distinct vertices i and j, any subset of $\{i \to j, i \leftarrow j, i \leftrightarrow j\}$ as edges. In Lauritzen *et al*. (1990) it is shown that both ways of defining separation in graphs are equivalent if the graph is a DAG; the proof of this result holds for directed cyclic graphs as well. The present paper is in the Pearl tradition, but, unlike in previous domain enlargements, this time 'generalization by domain extension' is no longer feasible. In addition, although the new class of graphs strictly contains all chain graphs (and indeed all 'general graphs'), the separation concept defined for it will, for certain chain graphs at least, differ from the Frydenberg (1990) separation concept which was based on the operation of moralization.[2]

The rest of the paper is organized as follows. In Section 2 we introduce the class of MC graphs. An MC graph can have up to four edges between each pair of its vertices i and j (say). More precisely, i and j can be connected by an arrow from i to j ($i \to j$), an arrow from j to i ($i \leftarrow j$), a two-headed arrow ($i \leftrightarrow j$), an undirected edge ($i - j$), or any subset of these four possibilities. Furthermore, at each vertex there can be an undirected self loop (e.g., $i - i$). The most important innovation of this section, however, consists of a simplification of the concept of an open path (by deleting the italicized words in Pearl's definition: 'its noncolliders are all outside the conditioning set, whereas its colliders are members of *or have a descendant in* the conditioning set'), which became possible after generalizing the concept of a path (by dropping the usual stipulation that all its vertices be different). It is a remarkable fact that, for graphs without undirected edges, these changes do not lead to an alteration of the ensuing separation properties of the graph.[3] In Section 3 we treat marginalization and conditioning, and show that the class of MC graphs is closed

[1]For chain graphs, a variant of this approach based on the operation of 'augmentation' of subgraphs leads to the so-called *alternative Markov property*. See Andersson *et al*. (2001) for details.
[2]It also differs from the augmentation-based separation concept of Andersson *et al*. (2001).
[3]Independently, Studený (1998) proposes the same modification of Pearl's d-separation concept and remarks that 'it is even simpler than the original' definition.

under these operations. In Section 4 the Markov property relative to an MC graph is defined and a final theorem is stated, essentially summarizing the paper. The results formulated in these sections enable one to consider in advance the effects upon the Markov properties of the model of the decision to study a certain subset of variables and either ignore completely or condition upon the value of all other variables. Readers will notice that most results are not in any way surprising and can be proved straightforwardly. In fact, while working on this paper the really surprising thing was that everything became very simple once the concepts 'path' and 'open path' were endowed with their new content, whereas some results remained totally unreachable without these two adjustments. Finally, in Section 5 relations with approaches based on summary graphs (Cox and Wermuth 1996; Wermuth *et al*. 1998) and on ancestral graphs (Richardson and Spirtes 2000) are discussed, and some concluding remarks are made.

2. MC graphs

An *MC graph* is an ordered pair $G = (V, E)$, where V is a finite set consisting of elements called *points* or *vertices*, and E is a collection of *edges*, that is, symbols denoted as $i \rightarrow j$, $i \leftarrow j$, $i \leftrightarrow j$, or $i - j$, where i and j are (not necessarily distinct) vertices of V. Here we make no distinction between $i \leftrightarrow j$ and $j \leftrightarrow i$, between $i - j$ and $j - i$, or between $i \rightarrow j$ and $j \leftarrow i$. However, if $i \neq j$ we *do* distinguish between each of $i \rightarrow j$, $i \leftarrow j$, $i \leftrightarrow j$, and $i - j$, so there can be up to four different edges between each two distinct vertices. The edges $i \rightarrow j$, $i \leftrightarrow j$ and $i - j$ are called, respectively, a (*one-headed*) *arrow* (or *directed edge*), an *arc* (or *two-headed arrow*, or *bidirected edge*) and a *line* (or *undirected edge*); the presence of an edge between i and j makes these vertices *adjacent*. If $i = j$, edges connecting i and j are called *self loops*. It will turn out later that the only self loops that are of interest to us are *undirected self loops*, that is, self loops of the type $i - i$, where $i \in V$. Note that an undirected graph is an MC graph without self loops in which all edges are lines; a directed graph is an MC graph without self loops in which all edges are one-headed arrows. An MC graph is called *directional* if it has no lines, that is, it may only have arrows and arcs. If $e \in \{i \rightarrow j, i \leftarrow j, i \leftrightarrow j, i - j\}$ is an edge of G, then i and j are called the *end-points* of e. In that case we define $[e] \equiv \{i, j\}$ as the set of its end-points (a singleton set if $i = j$). We will also write $e = i \cdots j$ to denote that e is some edge connecting the end-points i and j. If $e = i \rightarrow j$, i is the *tail* of e, while j is its *head*; if $e = i \leftrightarrow j$, both i and j are heads of e; if $e = i - j$, both i and j are tails of e. Henceforth, unless stated otherwise, by 'graph' we mean 'MC graph'.

Let $G \equiv (V, E)$ and $H \equiv (W, F)$ be graphs. Then H is called a *subgraph* of G, $H \subseteq G$, if $W \subseteq V$ and $F \subseteq E$. If $a \subseteq V$, the subgraph of G induced by a is $G_a := (a, E_a)$, where $E_a := \{e \in E | [e] \subseteq a\}$. Suppose i and j are (not necessarily different) vertices of the graph $G = (V, E)$. Let, for some $n \geqslant 1$, $\pi \in V \times (E \times V)^n$, say $\pi = (i_0, e_1, i_1, \ldots, e_n, i_n)$, where

[4]There is some redundancy in the notation for a path, as a path of length n is determined uniquely by its end-points and the n-tuple of its edges, that is, $(i_0, e_1, i_1, \ldots, e_n, i_n) = (j_0, f_1, j_1, \ldots, f_n, j_n)$ if and only if $(i_0, i_n) = (j_0, j_n)$ and $(e_1, \ldots, e_n) = (f_1, \ldots, f_n)$. Since this redundancy is convenient, it is maintained throughout.

$\{i_k\colon 0 \leqslant k \leqslant n\} \subseteq V$ and $\{e_1, \ldots, e_n\} \subseteq E$. Then π is called a *path of length n from i to j*, if $i_0 = i$, $i_n = j$ and $e_k = i_{k-1} \cdots i_k$, for $k = 1, \ldots, n$.[4] Sometimes, particularly in examples, the more convenient notation $(i_0 \cdots i_1 \cdots \ldots \cdots i_{n-1} \cdots i_n)$ will be used to denote π (where each instance of '$\cdots$' is replaced by one of $\to$, $\leftarrow$, $\leftrightarrow$, or $-$). The vertex i_0 (i_n) is called a *tail* or *head end-point* of π depending on it being a tail or head of e_1 (e_n); vertices in the set $\{i_k | 0 < k < n\}$ are called *intermediate points* of π (if $n = 1$ the path has no intermediate points). End-points and intermediate points together are the *points* of π (an end-point may also be an intermediate point on a given path), while $\{e_1, \ldots, e_n\}$ are its *edges*. Notice that it is *not* required that the points or edges of a path are all different. If all points of π are distinct, that is, $|\{i_0, \ldots, i_n\}| = n + 1$, then the path is *acyclic*, otherwise it is *cyclic*. It is called a *loop* if its end-points are the same vertex, i.e., $i_0 = i_n$. Suppose $\pi \equiv (i_0, e_1, i_1, \ldots, e_n, i_n)$ is a path from i_0 to i_n. Then π is called a *directed path* if $e_k = i_{k-1} \to i_k$, $1 \leqslant k \leqslant n$; in that case, if $0 \leqslant k < m \leqslant n$, i_k is called an *ancestor* of i_m, and i_m a *descendant* of i_k. If $a \subseteq V$, $\mathrm{an}(a) \equiv \mathrm{an}_G(a) := a \cup \bigcup_{i \in a}\{j \in V | j$ ancestor of $i\}$; note that $\mathrm{an}(a)$ contains a as a subset. Subsets $a \subseteq V$ which satisfy $an_G(a) = a$ are called *G-ancestral*. For a path $\pi = (i_0, e_1, i_1, \ldots, e_n, i_n)$, let $0 \leqslant k_0 < \ldots < k_m \leqslant n$ $(1 \leqslant m \leqslant n)$ be an increasing sequence of indices such that $i_{k_s+1} = i_{k_{s+1}}$, $0 \leqslant s < m$. Then $[e_{k_s+1}] = \{i_{k_s}, i_{k_s+1}\} = \{i_{k_s}, i_{k_{s+1}}\}$, so $\sigma \equiv (i_{k_0},\ e_{k_0+1},\ i_{k_1},\ \ldots,\ i_{k_{m-1}},\ e_{k_{m-1}+1},\ i_{k_m})$ is a path of length m from i_{k_0} to i_{k_m}; σ is called a *subpath* of π. The following will be clear:

Fact 2.1. *Let i and j be two different vertices of the graph G. If π is a path from i to j, then there exists a subpath of π from i to j which is acyclic.*

Suppose $\pi = (i_0, e_1, i_1, \ldots, e_n, i_n)$ is a path in $G = (V, E)$. The intermediate point j *occurs as a collider on* π (*at position* k, where $0 < k < n$) if $j = i_k$, $e_k \in \{i_{k-1} \to i_k, i_{k-1} \leftrightarrow i_k\}$ and $e_{k+1} \in \{i_k \leftarrow i_{k+1}, i_k \leftrightarrow i_{k+1}\}$. On the other hand, the intermediate point j *occurs as a non-collider on* π (*at position* k, $0 < k < n$) if $j = i_k$, and $e_k \in \{i_{k-1} \leftarrow i_k, i_{k-1} - i_k\}$ or $e_{k+1} \in \{i_k \to i_{k+1}, i_k - i_{k+1}\}$. Intuitively these notions correspond to the intermediate point j being or not being a point at which, along the path, two arrow heads meet. Notice that, trivially, in undirected graphs an intermediate point on a path can only occur as a non-collider. In general, however, due to the fact that one can have multiple occurrences on a path of the same intermediate point, it is possible that a certain intermediate point occurs both as a collider and as a non-collider on π. So the sets $C_\pi := \{j \in V | j$ occurs as a collider on $\pi\}$, and $N_\pi := \{j \in V | j$ occurs as a non-collider on $\pi\}$ need not be disjoint. In fact, since a path can have self loops as edges, an intermediate point can occur as a collider and as a non-collider on π at the same position. To give an example of this, consider the graph $G = (\{1, 2, 3\}, \{1 \to 2, 2 \to 2, 2 \to 3\})$. Now, let $\pi \equiv (1 \to 2 \to 2 \to 3)$, so π is a path of length 3 from 1 to 3. Then the intermediate point 2 occurs as a non-collider on π at position 1, as $e_1 = 1 \to 2$ and $e_2 = 2 \to 2$. But 2 also occurs as a collider on π at position 1, since $e_2 = 2 \leftarrow 2$. If the self loop $2 \to 2$ is deleted from π, the subpath $\sigma := (1 \to 2 \to 3)$ is obtained. Clearly, $N_\sigma \subseteq N_\pi$ and $C_\sigma \subseteq C_\pi$. This example brings us to the following general observation.

Fact 2.2. *Suppose $\pi = (i_0, e_1, i_1, \ldots, e_n, i_n)$ is a path from i_0 to i_n, and σ is the subpath of π obtained by deleting some self loops from π (more precisely, if $i_{k-1} = i_k$, then the pair (e_k, i_k) may be deleted from π). Then $N_\sigma \subseteq N_\pi$. Moreover, if no undirected self loops are deleted from π, then $C_\sigma \subseteq C_\pi$.*

If $\pi \equiv (i_0, e_1, i_1, \ldots, e_n, i_n)$ is a path from i_0 to i_n, then $\pi^- := (i_n, e_n, i_{n-1}, \ldots, e_1, i_0)$ is a path from i_n to i_0. Clearly, this reversal of the *order* of points and edges of the path has no effect on the collider status of the intermediate points, that is, $C_{\pi^-} = C_\pi$ and $N_{\pi^-} = N_\pi$. In situations where the difference between π and π^- (i.e., the order of points and edges) is immaterial, we will sometimes say that the path is *between* its end-points i_0 and i_n. A path π is called a *pure collision path* (*collisionless path*) if all its intermediate points occur as colliders (non-colliders) only, that is, $N_\pi = \emptyset$ ($C_\pi = \emptyset$). This is trivially the case if π has length 1.

Definition 2.3. *Separation for MC graphs. Suppose i and j are two distinct vertices of the MC graph $G = (V, E)$, and let $c \subseteq V \backslash \{i, j\}$. Let π be a path between i and j. Then π is* open given *c if $C_\pi \subseteq c$ and $N_\pi \subseteq V \backslash c$. If the path is not open given c, then it is* blocked by *c. Clearly, this means that either $C_\pi \cap V \backslash c \neq \emptyset$ or $N_\pi \cap c \neq \emptyset$. If all paths between i and j are blocked by c, then i and j are* separated by *c. Finally, let a, b and c be pairwise disjoint subsets of V. Then a and b are* separated by *c if, for all $i \in a$, $j \in b$, i and j are separated by c (it is understood that this holds trivially if $a = \emptyset$ or $b = \emptyset$).*

Remarks 2.4. (i) Clearly, if G is undirected, all intermediate points on a path can only occur as non-colliders. Hence, in this case the definition of 'separation' is equivalent to the more common definition which is as follows: subsets a and b are separated by c if all (acyclic) paths in G between a and b intersect c. (Note that acyclicity is irrelevant.)

(ii) Since a path is open given c if and only if all its colliders are in c and all its non-colliders are outside c, N_π and C_π partition the set of intermediate points if π is an open path.

(iii) If a, b and c partition V, then a *shortest open path* between (a vertex of) a and (a vertex of) b necessarily is a *shortest pure collision path* between a and b, and vice versa. Hence, in this situation, a and b are separated by c if and only if there are no pure collision paths between vertices of a and vertices of b.

Example 2.5. Consider the graph (V, E) depicted in Figure 1(a) which has $V = \{1, 2, 3, 4\}$, $E = \{1 \to 3,\ 2 \to 4,\ 3 \leftarrow 4,\ 3 - 4,\ 3 \to 4\}$. In this graph $\{1\}$ and $\{2\}$ are separated by $\{3, 4\}$. To see this, note that any path π from 1 to 2 must have $3 \in N_\pi$ or $4 \in N_\pi$ (or both), hence $N_\pi \cap \{3, 4\} \neq \emptyset$. On the other hand, for example $\{1\}$ and $\{2\}$ are *not* separated by $\{3\}$, for the path $\pi \equiv (1 \to 3 \leftarrow 4 \leftarrow 2)$ has $C_\pi = \{3\}$, $N_\pi = \{4\}$, so $C_\pi \subseteq \{3\}$ and $N_\pi \cap \{3\} = \emptyset$, thus π is open given $\{3\}$. It is easy to check that '$\{1\}$ and $\{2\}$ are separated by $\{3, 4\}$' is the *only* valid non-trivial separation statement for this graph. In contrast, the graph in Figure 1(c) which has the arrow $3 \leftarrow 4$ replaced by the arc $3 \leftrightarrow 4$ *only* satisfies the statement '$\{1\}$ and $\{2\}$ are separated by $\{3\}$'; for example, the path $\pi \equiv (1 \to 3 \leftrightarrow 4 \leftarrow 2)$ now is open given $\{3, 4\}$ (since $N_\pi = \emptyset$, $C_\pi = \{3, 4\}$), so 1 and 2 are no longer separated

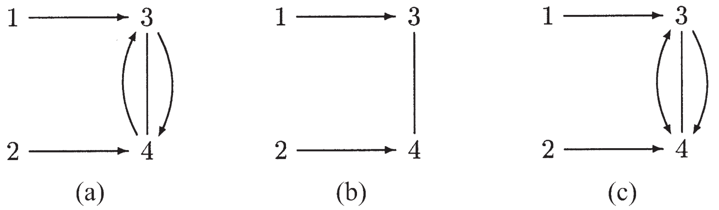

Figure 1. Three MC graphs defining distinct independence models: (a) $G(\{1, 2, 4\})$; (b) $G(\{4\})$; (c) $G(\{1, 3, 4\})$. See Example 2.8 for explanation.

by $\{3, 4\}$. The reader may wish to verify that the graph in Figure 1(b) satisfies '$\{1, 3\}$ are separated from $\{2\}$ by $\{4\}$', '$\{2, 4\}$ are separated from $\{1\}$ by $\{3\}$', and all statements which can be deduced from these by the properties of *decomposition* and *weak union* (cf. Proposition 2.10, below).

Our first lemma is a reduction result stating that self loops of type $i \to i$ or $i \leftrightarrow i$ may be deleted from a graph without changing its separation properties.

Lemma 2.6. *Suppose $G = (V, E)$ is a graph, and let $H = (V, F)$ be the subgraph of G obtained by removing all self loops of type $i \to i$ or $i \leftrightarrow i$, where $i \in V$, that is, $F = \{e \in E | \forall i \in V : e \neq i \to i, e \neq i \leftrightarrow i\}$. Let j and k be two distinct vertices, and let $c \subseteq V \backslash \{j, k\}$. Then j and k are separated by c in G if and only if j and k are separated by c in H.*

Proof. First note that any path in H is also a path in G. Necessity is now clear.

To prove sufficiency, let π be a path in G between j and k. Let σ be the subpath of π obtained by deleting all self loops of type $i \to i$ or $i \leftrightarrow i$ from π. Then σ is a path between j and k in H, so σ is blocked by c. Thus, $C_\sigma \cap V \backslash c \neq \varnothing$ or $N_\sigma \cap c \neq \varnothing$. By Fact 2.2 it now follows that $C_\pi \cap V \backslash c \neq \varnothing$ or $N_\pi \cap c \neq \varnothing$, that is, π is blocked by c in G. □

Convention 2.7. *Self loops of type $i \to i$ or $i \leftrightarrow i$, being irrelevant for a graph's separation properties by the preceding lemma, are henceforth ignored completely in all MC graphs. Alternatively, we assume that MC graphs do not contain such self loops. The self loop $i - i$ will also be denoted by i°. Occasionally, particularly in Section 3, it is convenient to denote the loop (i, i°, i) simply by its distinctive characteristic, which is the self loop i°, so the symbol i° is then considered as 'self loop plus end-points'.*

One may wonder if the possibility of having any subset of $\{i \to j, i \leftarrow j, i \leftrightarrow j, i - j\}$ as edges between two different vertices i and j really enlarges the number of independence models over the set of vertices which are defined by the separation properties of the graph. Here, if $G = (V, E)$ is a graph, then the independence model it defines over V is given by all triples (a, b, c) of pairwise disjoint subsets of V such that $a = \varnothing$, or $b = \varnothing$, or a and b are separated by c. The following example suggests a positive answer to the question.

Example 2.8. Put $V \equiv \{1, 2, 3, 4\}$, $e_1 \equiv 3 \to 4$, $e_2 \equiv 3 \leftarrow 4$, $e_3 \equiv 3 \leftrightarrow 4$, and $e_4 \equiv 3 - 4$. Let, for $a \subseteq \{1, 2, 3, 4\}$, $E(a) := \{1 \to 3, 2 \to 4\} \cup \{e_k | k \in a\}$ and $G(a) := (V, E(a))$. Also, let s_1 denote the statement '$\{1\}$ and $\{2\}$ are separated by $\{4\}$'; let s_2 denote the statement '$\{1\}$ and $\{2\}$ are separated by $\{3\}$'; let s_3 denote '$\{1\}$ and $\{2\}$ are separated by $\{3, 4\}$'; finally, let s_4 denote '$\{1\}$ and $\{2\}$ are separated by $\varnothing$'. Whenever $a \subseteq \{1, 2, 3, 4\}$, $S(a)$ will denote the statements s_k, $k \in a$. One may easily verify that, for all $a \subseteq \{1, 2, 3, 4\}$, the separation statements in $S(V \backslash a)$ hold for graph $G(a)$, whereas the statements in $S(a)$ do not hold for this graph. See Figure 1.

The next result entails that the separation concept defined above (Definition 2.3) is a proper generalization of Pearl's d-separation concept for DAGs to the class of MC graphs. Recall that a path is acyclic if all its points are distinct. In particular, acyclic paths cannot have self loops as edges. Also recall Pearl's definition of an open path (assumed acyclic) in a DAG, here rephrased slightly: 'An acyclic path between vertices i and j is open given a subset of vertices c if and only if all its non-colliders are outside c and all its colliders either are in c or have a descendant in c' (cf. Pearl 1988). This definition can be applied verbatim to directional graphs, that is, to MC graphs which do not contain undirected edges.

Proposition 2.9. *Suppose $G = (V, E)$ is a directional graph (that is, E contains no undirected edges). Let i and j be two distinct vertices, and let $c \subseteq V \backslash \{i, j\}$. Then i and j are separated by c if and only if i and j are d-separated given c (in the sense of Pearl).*

Proof. Note that, for $j \in V$, $c \subseteq V$, $\mathrm{an}(j) = \{j\} \cup \{i \in V | j$ is a descendant of $i\}$ and $\mathrm{an}(c) = \bigcup_{j \in c} \mathrm{an}(j)$. Pearl's condition for an acyclic path π to be open, is $C_\pi \subseteq \mathrm{an}(c)$ and $N_\pi \subseteq V \backslash c$. First suppose i and j are d-separated given c (in the sense of Pearl), but there exists in G an open path $\pi \equiv (i_0, e_1, i_1, \ldots, e_n, i_n)$ from $i = i_0$ to $j = i_n$ given c. We will show that there exists an acyclic subpath of π from i to j which is open given c (in the sense of Pearl), contradicting the hypothesis. Invoking Fact 2.1, denote by σ an acyclic subpath of π from i to j, say $\sigma \equiv (i_{k_0}, e_{k_0+1}, i_{k_1}, \ldots, i_{k_{m-1}}, e_{k_{m-1}+1}, i_{k_m})$, where $0 \leqslant k_0 < \ldots < k_m \leqslant n$ is an increasing sequence of indices such that $i = i_{k_0}$, $j = i_{k_m}$, and $i_{k_s+1} = i_{k_{s+1}}$, $0 \leqslant s < m$. We will first show that $N_\sigma \subseteq N_\pi$. Suppose $i_{k_s} \in N_\sigma$ for some $0 < s < m$. Then $e_{k_{s-1}+1} = i_{k_{s-1}} \leftarrow i_{k_s}$, or $e_{k_s+1} = i_{k_s} \to i_{k_{s+1}}$, that is, $e_{k_{s-1}+1} = i_{k_{s-1}} \leftarrow i_{k_{s-1}+1}$ or $e_{k_s+1} = i_{k_s} \to i_{k_s+1}$. It follows that $i_{k_{s-1}+1} = i_{k_s}$ occurs as a non-collider on π at position $k_{s-1} + 1$ or at position k_s, hence $i_{k_s} \in N_\pi$. To show that σ is open given c (in the sense of Pearl) it now suffices to show that $C_\sigma \subseteq \mathrm{an}(c)$, and since $C_\pi \subseteq c$ it is sufficient to show that $C_\sigma \backslash C_\pi \subseteq \mathrm{an}(c)$. So assume $i_{k_s} \in C_\sigma \backslash C_\pi$, that is, $e_{k_{s-1}+1} \in \{i_{k_{s-1}} \to i_{k_s}, i_{k_{s-1}} \leftrightarrow i_{k_s}\}$, $e_{k_s+1} \in \{i_{k_s} \leftarrow i_{k_{s+1}}, i_{k_s} \leftrightarrow i_{k_{s+1}}\}$, $e_{k_{s-1}+2} = i_{k_{s-1}+1} \to i_{k_{s-1}+2}$ (otherwise $i_{k_{s-1}+1} = i_{k_s}$ would occur as a collider on π at position $k_{s-1} + 1$) and $e_{k_s} = i_{k_s-1} \leftarrow i_{k_s}$ (otherwise i_{k_s} would occur as a collider on π at position k_s). So π has $(i_{k_{s-1}+1} \to i_{k_{s-1}+2} \cdots \ldots \cdots i_{k_s-1} \leftarrow i_{k_s})$ as a subpath (in fact, a loop). But then, since G has no undirected edges, at least one of the intermediate points i_r, $k_{s-1}+2 \leqslant r \leqslant k_s - 1$, must occur as a collider on π. Let r be the smallest index for which this is the case. Then $i_r \in C_\pi \subseteq c$, and as $(i_{k_{s-1}+1} \to i_{k_{s-1}+2} \to \ldots \to i_r)$ is a directed path from $i_{k_{s-1}+1} = i_{k_s}$ to i_r, $i_{k_s} \in \mathrm{an}(c)$.

To show the converse, suppose there exists an open (in the sense of Pearl) acyclic path $\sigma \equiv (i_0, e_1, i_1, \ldots, e_n, i_n)$ from $i = i_0$ to $j = i_n$ given c. The only 'problematic' intermediate points of σ (preventing σ from being an open path given c in our sense) are vertices i_k satisfying $i_k \in C_\sigma \cap \mathrm{an}(c) \backslash c$. However, for such vertices there exists a *minimum length* directed path $(i_k \to j_1 \to \ldots \to j_{m-1} \to j_m)$ such that $j_m \in c$. Clearly, if in σ all 'problematic' vertices i_k are substituted by the loop $(i_k \to j_1 \to \ldots \to j_{m-1} \to j_m \leftarrow j_{m-1} \leftarrow \ldots \leftarrow j_1 \leftarrow i_k)$, then the resulting path π (say) will be open since $\{i_k\} \cup \{j_r | 1 \leqslant r < m\} \subseteq N_\pi \cap V \backslash c$ and $j_m \in C_\pi \cap c$, hence $N_\pi \subseteq V \backslash c$ and $C_\pi \subseteq c$. □

Proposition 2.9 is false if the supposition that the graph has no lines is dropped. In Figure 2(a), all acyclic paths between 1 and 2 are blocked by $\varnothing$. However, the cyclic path $(1 \to 3 - 4 - 3 \leftarrow 2)$ is open given $\varnothing$, so it is not true that $\{1\}$ and $\{2\}$ are separated by $\varnothing$. On the other hand, $\{1\}$ and $\{2\}$ are separated by $\{4\}$. As this graph is a member of the class of chain graphs (Frydenberg 1990), this example also shows that the present separation concept is *not* equivalent with the separation concept defined for that particular class of graphs (Lauritzen and Wermuth 1989; Frydenberg 1990). Indeed, applying the latter separation concept to the graph gives that, for example, $\{1\}$ and $\{2\}$ are separated by $\varnothing$ but not by $\{4\}$. According to Proposition 2.9, for directional graphs the separation concept of the present paper is equivalent to Pearl's *d*-separation, although acyclic paths may be open in the sense of Pearl but closed in our sense. Consider the graph in Figure 2(b), which is a DAG. The acyclic path $(1 \to 3 \leftarrow 2)$ is open given $\{4\}$ in the sense of Pearl, but it is closed in our sense, since the collider 3 is not a member of the conditioning set $\{4\}$. However, the cyclic path $(1 \to 3 \to 4 \leftarrow 3 \leftarrow 2)$ is open given $\{4\}$ in our sense, and this 'trick' is precisely what makes the necessity part of the proof of Proposition 2.9 work.

The separation concept given in Definition 2.3 induces for each MC graph $G = (V, E)$ an *independence model* over the set of vertices which satisfies various properties. Here the independence model is defined as the set of all triples (a, b, c) of pairwise disjoint subsets of V such that a and b are separated by c. Some of these properties are summarized in the following proposition:

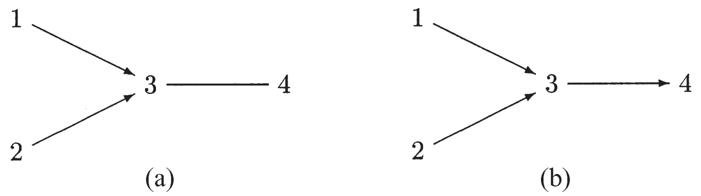

Figure 2. Two MC graphs: (a) chain graph; (b) directed acyclic graph.

Proposition 2.10. *Let $G = (V, E)$ be an MC graph. Then the following properties hold* (*in all cases a, b, c and d are pairwise disjoint subsets of V*)*:*

(i) *a and b are separated by $c \Leftrightarrow b$ and a are separated by c.* (*Symmetry.*)
(ii) *a and b are separated by c, and a and d are separated by $c \Leftrightarrow a$ and $b \cup d$ are separated by c.* (*Composition/decomposition.*)
(iii) *a and c are separated by d, and a and b are separated by $c \cup d \Leftrightarrow a$ and $b \cup c$ are separated by d.* (*Contraction.*)
(iv) *a and $b \cup c$ are separated by $d \Rightarrow a$ and b* are separated by *$c \cup d$.* (*Weak union.*)
(v) *a and b are separated by $c \cup d$, and a and c are separated by $b \cup d \Rightarrow a$ and $b \cup c$ are separated by d.* (*Intersection.*)

Proof. We will only show (iii) and (v), leaving the rest to the reader.

(iii) To show necessity, suppose a and c are separated by d, and a and b are separated by $c \cup d$. We must show that all paths from a to $b \cup c$ are blocked by d. Let $\pi \equiv (i_0, e_1, \ldots, e_n, i_n)$ be such a path. If π is from a to c there is nothing to prove, so assume π is from a to b. Since π is blocked by $c \cup d$, *either* $C_\pi \not\subseteq (c \cup d)$, hence $C_\pi \not\subseteq d$, so π is blocked by d, *or* $N_\pi \cap (c \cup d) \neq \varnothing$. In the latter case, if $N_\pi \cap c = \varnothing$, then $N_\pi \cap d \neq \varnothing$, hence π is blocked by d. Thus, assume $N_\pi \cap c \neq \varnothing$. Let k $(0 < k < n)$ be such that $i_k \in N_\pi \cap c$. Then $\sigma \equiv (i_0, e_1, \ldots, e_k, i_k)$ is a path from $i_0 \in a$ to $i_k \in c$, so by hypothesis σ is blocked by d. Since σ is a subpath of π, π is blocked by d as well.

To see sufficiency, assume a and $b \cup c$ are separated by d. Then clearly a and c are separated by d, so it suffices to show that all paths from a to b are blocked by $c \cup d$. Let $\pi \equiv (i_0, e_1, \ldots, e_n, i_n)$ be such a path. Since π is blocked by d, *either* $N_\pi \cap d \neq \varnothing$, hence $N_\pi \cap (c \cup d) \neq \varnothing$, so π is blocked by $c \cup d$, *or* $C_\pi \not\subseteq d$. In the latter case, if $C_\pi \cap c = \varnothing$, then $C_\pi \not\subseteq (c \cup d)$, hence π is blocked by $c \cup d$. Thus, assume $C_\pi \cap c \neq \varnothing$. Now, let k $(0 < k < n)$ be the smallest index such that $i_k \in C_\pi \cap c$. Then $\sigma \equiv (i_0, e_1, \ldots, e_k, i_k)$ is a path from $i_0 \in a$ to $i_k \in c$, so by hypothesis σ is blocked by d, that is, *either* $N_\sigma \cap d \neq \varnothing$, hence $N_\sigma \cap (c \cup d) \neq \varnothing$, *or* $C_\sigma \not\subseteq d$, hence $C_\sigma \not\subseteq (c \cup d)$ (since $C_\sigma \cap c = \varnothing$). So σ is blocked by $c \cup d$. As σ is a subpath of π, it follows that π is blocked by $c \cup d$ as well.

(v) Let $a \neq \varnothing$, $b \neq \varnothing$, $c \neq \varnothing$ and d be pairwise disjoint subsets of V, such that a and b are separated by $c \cup d$, and a and c are separated by $b \cup d$. We must show that a and $b \cup c$ are separated by d. To obtain a contradiction, suppose there exists an open path given d between a and $b \cup c$. Let $\pi \equiv (i_0, e_1, \ldots, e_n, i_n)$ be a *shortest* open path given d from a vertex of a to a vertex of $b \cup c$, say, without loss of generality, $i_0 \in a$ and $i_n \in b$. As π is open given d, $N_\pi \subseteq V \backslash d$ and $C_\pi \subseteq d$. Since π has minimum length, $\{i_1, \ldots, i_{n-1}\} \cap (b \cup c) = \varnothing$, hence $N_\pi \subseteq (V \backslash (b \cup c)) \cap (V \backslash d) \subseteq V \backslash (c \cup d)$. Also, $C_\pi \subseteq d \subseteq c \cup d$, so π is an open path given $c \cup d$ from $i_0 \in a$ to $i_n \in b$, contradicting the hypothesis. □

Proposition 2.11. *Suppose $G = (V, E)$ is a directional graph. Let a, b and c be pairwise disjoint subsets of V, and let $d \equiv \mathrm{an}(a \cup b \cup c)$. Then a and b are separated by c in G if and only if a and b are separated by c in G_d.*

Proof. Since any open path in $H \equiv G_d$ is an open path in G as well, necessity is immediate.

For sufficiency, suppose a and b are separated by c in H, whereas $\pi \equiv (i_0, e_1, \ldots, e_n, i_n)$ is an open path (given c) in G from $i_0 \in a$ to $i_n \in b$. Since $C_\pi \subseteq c \subseteq d$ and $N_\pi \subseteq V \backslash c$, to obtain a contradiction it suffices to show that $N_\pi \subseteq d$, as it follows from this that π is an open path in H. Let $i_r, i_s \in C_\pi$ be two successive colliders on π, that is, $i_t \in N_\pi$ for $r < t < s$. The subpath $\sigma \equiv (i_r, e_{r+1}, i_{r+1}, \ldots, e_{s-1}, i_s)$ of π has one of the following two generic forms: either $\sigma = (i_r \leftarrow \ldots \leftarrow i_t \rightarrow \ldots \rightarrow i_s)$ for some $r < t < s$, or $\sigma = (i_r \leftarrow \ldots \leftarrow i_t \leftrightarrow i_{t+1} \rightarrow \ldots \rightarrow i_s)$ for some $r \leqslant t < s$. Since d is G-ancestral, $N_\sigma \subseteq d$. A slight adaptation of this arguments (allowing for the fact that neither i_0 nor i_n needs to be a head end-point of π) shows that essentially the same conclusion regarding the generic form of σ holds when i_s is the first collider on π and $\sigma \equiv (i_0, e_1, i_1, \ldots, e_s, i_s)$, or when i_r is the last collider on π and $\sigma \equiv (i_r, e_{r+1}, i_{r+1}, \ldots, e_n, i_n)$, or when π is collisionless and $\sigma \equiv \pi$. But then $N_\pi = \cup_\sigma N_\sigma \subseteq d$. $\square$

Proposition 2.12. *Suppose $G = (V, E)$ is a directional graph. Let a, b and c be pairwise disjoint subsets of V, and let $d \equiv \mathrm{an}(a \cup b \cup c)$. Then a and b are separated by c in G if and only if there exist disjoint subsets $a' \subseteq V \backslash c$ and $b' \subseteq V \backslash c$ such that:* (*i*) *$a' \cup b' \cup c = d$;* (*ii*) *$a \subseteq a'$, $b \subseteq b'$; and* (*iii*) *a' and b' are separated by c in G_d. Furthermore, $b' := \{j \in d \backslash (a \cup c) | j$ is separated from a by $c\}$ and $a' := d \backslash (b' \cup c)$ satisfy the conditions of the proposition.*

Proof. Sufficiency is clear. For necessity, define a' and b' as in the final sentence of the proposition. As conditions (i) and (ii) are obvious, it suffices to show that a' and b' are separated by c in G. Using Proposition 2.11, we may assume without loss of generality that $\mathrm{an}(a \cup b \cup c) = V$. We will show that i and j are separated by c whenever $i \in a'$, $j \in b'$. If $i \in a$ this holds by the definition of b', so in order to obtain a contradiction, assume $i \in a' \backslash a$, $j \in b'$ are such that an open path π exists from j to i given c, say $\pi = (j_0, e_1, j_1, \ldots, e_n, j_n)$, where $j_0 = j$ and $j_n = i$. Since $i \notin b'$ there exists an open (given c) path σ from i to some $k \in a$, say $\sigma = (i_0, f_1, i_1, \ldots, f_m, i_m)$, $i_0 = i$ and $i_m = k$. Let ρ denote the concatenation of π and σ, that is, $\rho \equiv (j_0, e_1, j_1, \ldots, e_n, i, f_1, i_1, \ldots, f_m, i_m)$ is a path from $j_0 = j$ to $i_m = k$. As $j \in b'$ and $k \in a$, ρ is blocked by c. Since $N_\rho \subseteq N_\pi \cup N_\sigma \cup \{i\} \subseteq V \backslash c$ and $C_\rho \subseteq C_\pi \cup C_\sigma \cup \{i\} \subseteq c \cup \{i\}$, necessarily $i \in C_\rho \backslash c$. Now, $V = \mathrm{an}(a \cup b \cup c) = \mathrm{an}(a) \cup \mathrm{an}(b) \cup \mathrm{an}(c)$. Suppose $i \in \mathrm{an}(c)$. Let τ denote a *shortest* directed path from i to some $l \in c$, and let τ^- denote the reverse path (i.e., going from l to i). It is easily seen that the concatenation of π, τ, τ^- and σ now constitutes an open path (given c) from j to k, which is a contradiction. Hence $i \in \mathrm{an}(a) \backslash \mathrm{an}(c)$, or $i \in \mathrm{an}(b) \backslash \mathrm{an}(c)$. If $i \in \mathrm{an}(a) \backslash \mathrm{an}(c)$, then there exists a directed path τ (say) from i to some vertex $l \in a$. Clearly, all intermediate points of τ are non-colliders (as τ is a directed path), and none of the points of τ is in c (as otherwise $i \in \mathrm{an}(c)$). By concatenating the paths π and τ we now obtain an open path (given c) from $j \in b'$ to $l \in a$, which is a contradiction. Finally, if $i \in \mathrm{an}(b) \backslash \mathrm{an}(c)$, then there exists a directed path τ (say) from i to some vertex $l \in b$. Clearly, all intermediate points of τ are non-colliders (as τ is a directed path), and none of the points of τ is in c (as otherwise $i \in \mathrm{an}(c)$). By concatenating the paths σ^- (i.e., σ traversed in reverse order) and τ we now obtain an open path (given c) from $k \in a$ to $l \in b$, which again is a contradiction.□

3. Marginalizing and conditioning

In this section we will first show that the class of MC graphs is closed under marginalization, then that it is closed under conditioning. In Koster (1999) it is shown that the class of directional graphs is closed under marginalization. The class of undirected graphs possesses both closure properties (see Examples 3.1 and 3.7 below), but most other classes of graphs encountered in the literature do not share either of them (one notable exception is the class of *ancestral graphs*, which is discussed in Section 5). Finally, we will show that the order in which the operations of marginalization and conditioning are carried out is immaterial – as should be the case from a probabilistic point of view. We will first make explicit what is meant by 'closed under marginalization'.

3.1. Marginalizing

Let $\mathcal{G}$ denote a class of graphs such that for each member of $\mathcal{G}$ a concept of 'separation' is defined (applying to triples of subsets of vertices) whose associated independence model satisfies the properties stated in Proposition 2.10. Let $G = (V, E)$ be a member of $\mathcal{G}$, and suppose for all subsets $m \subseteq V$ the class $\mathcal{G}$ contains a (possibly not unique) graph $G^{m\cdot} \equiv (V \backslash m, F)$ with vertex set $V \backslash m$ and edge set $F = F(m)$ (say), such that, for all pairwise disjoint subsets a, b and c of $V \backslash m$, a and b are separated by c in G if and only if a and b are separated by c in $G^{m\cdot}$. Clearly, if $m = V$ or $m = \emptyset$ this is trivially satisfied by putting $G^{V\cdot} = (\emptyset, \emptyset)$, and $G^{\emptyset\cdot} = G$. The class $\mathcal{G}$ is *closed under marginalization* if the stated property holds for all its members G. Preferably, a simple algorithm (e.g., consisting of local operations only) should be available to construct $G^{m\cdot}$ from G and m. For the present class of graphs, as for the class of undirected graphs, this is the case, as will be shown below.

Example 3.1. Consider the undirected graph G in Figure 3(a). Vertex 5 is encircled since we intend to marginalize over this vertex, that is, we wish to obtain $G^{\{5\}\cdot}$. This can be done by

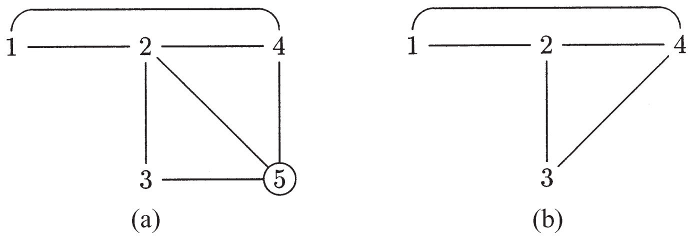

Figure 3. (a) Undirected graph G. (b) After marginalizing over vertex 5 the graph $G^{\{5\}\cdot}$ is obtained. See Example 3.1 for explanation.

applying the so-called *rubber band procedure*: for each pair of vertices (i, j) for which there is a path $(i - 5 - j)$, a new edge $i - j$ is added; then vertex 5 and all edges adjacent to it are wiped out. In this way, $G^{\{5\}\cdot}$ is obtained as the graph in Figure 3(b). It is easily verified that, for all pairwise disjoint subsets a, b and c of $\{1, 2, 3, 4\}$, a and b are separated by c in G if and only if a and b are separated by c in $G^{\{5\}\cdot}$.

Moving now to the general case, let $G = (V, E)$ be an MC graph. Suppose i, j and k are vertices of G, $k \notin \{i, j\}$, such that $f \in E$ and $g \in E$ satisfy $f = i \cdots k$ and $g = k \cdots j$. Then (i, f, k, g, j) is a path of length 2 from i to j with intermediate point k. Also, if $k^\circ \in E$, then $(i, f, k, k^\circ, k, g, j)$ is a path of length 3 from i to j with both intermediate points equal to k. (In fact, in that case the pair (k°, k) can be repeated arbitrarily often to obtain a path $(i, f, k, k^\circ, k, \ldots, k^\circ, k, g, j)$ of any finite length $n \geqslant 3$.) The latter type of path, which contains $(k, k^\circ, k, \ldots, k^\circ, k)$ as subpath, will be denoted by (i, f, k°, g, j). In that case we will say that the intermediate point k occurs as a self loop on the path (see Convention 2.7). Now define the edge $f \circ g = g \circ f$ by Table 1.

Table 1 essentially states that, if $f = i \cdots k$ and $g = k \cdots j$, then the edge $f \circ g = i \cdots j$ if and only if (i, f, k, g, j) or (i, f, k°, g, j) is a collisionless path in G from i to j. We define, for $k \in V$, the MC graph $G^{\{k\}\cdot}$ as the graph on $V \backslash \{k\}$ obtained after the $\circ$-operation of Table 1 is applied to all pairs of edges (f, g) satisfying $f = i \cdots k$ and $g = k \cdots j$, where $\{i, j\} \subseteq V \backslash \{k\}$. Formally, $G^{\{k\}\cdot} := (V \backslash \{k\}, E^{\{k\}\cdot})$, where the edge set $E^{\{k\}\cdot}$ is given by $E^{\{k\}\cdot} := \{e \in E | [e] \subseteq V \backslash \{k\}\} \cup \{f \circ g | f = i \cdots k, g = k \cdots j, \{i, j\} \subseteq V \backslash \{k\}\}$. It is understood here that Convention 2.7 applies, that is, if $i = j$, then the resulting self loop $f \circ g$ is void, unless it equals the undirected self loop

Table 1. Creation of edges by marginalizing over or conditioning on k. If $f = i \cdots k$ and $g = k \cdots j$, then the edge $f \circ g$ is created by marginalizing over k conforming to the entries of the table; for example, when $f = i \leftrightarrow k$ (third row) and $g = k - j$ (final column), then the edge $i \leftarrow j$ is produced. The four boxed entries of the table only apply when k occurs as a self loop, and are understood to be void otherwise. The edge $f \,\square\, g$ is created by conditioning on k conforming to the boxed entries of the table; for example, when $f = i \leftrightarrow k$ (third row) and $g = k \leftarrow j$ (first column), then the edge $i \leftarrow j$ is produced. Notice that the edge produced ($f \circ g$ or $f \,\square\, g$) inherits its 'edge ends' (i.e., the head or tail property of the end-points i and j) from the constituent edges f and g. Also, by Convention 2.7, if $i = j$ the resulting self loop is void, unless it equals the undirected self loop $i^\circ = i - i$

	g			
f	$k \leftarrow j$	$k \rightarrow j$	$k \leftrightarrow j$	$k - j$
$i \rightarrow k$	$\boxed{i - j}$	$i \rightarrow j$	$\boxed{i \rightarrow j}$	$i - j$
$i \leftarrow k$	$i \leftarrow j$	$i \leftrightarrow j$	$i \leftrightarrow j$	$i \leftarrow j$
$i \leftrightarrow k$	$\boxed{i \leftarrow j}$	$i \leftrightarrow j$	$\boxed{i \leftrightarrow j}$	$i \leftarrow j$
$i - k$	$i - j$	$i \rightarrow j$	$i \rightarrow j$	$i - j$

$i^\circ = i - i$. The next example may help clarifying the procedure of marginalizing over a single vertex $k \in V$ using the relevant entries of Table 1.

Example 3.2. Consider the MC graph G in Figure 4(a). Vertex 4 is encircled since we wish to marginalize over this vertex. Thus, when applying Table 1, $k = 4$. Now, if $f = 3 - 4$, $g = 4 \rightarrow 5$, then the edge $f \circ g = 3 \rightarrow 5$ is produced as indicated by the final row, second column of the table. Furthermore, putting $f = g = 3 - 4$ generates the self loop $f \circ g = 3 - 3 = 3^\circ$ (cf. final row, last column of the table). No other edges are produced by marginalizing over 4; for example, $f = g = 4 \rightarrow 5$ results in the self loop $f \circ g = 4 \leftrightarrow 5$ (second row, second column of the table) which, however, drops out by Convention 2.7. Thus the graph $G^{\{4\}\cdot}$ is obtained as Figure 4(b). Due to the presence of the self loop $3 - 3$ in $G^{\{4\}\cdot}$, 1 and 2 are not separated by $\varnothing$ (the path $(1 \rightarrow 3^\circ \leftarrow 2)$ is open given $\varnothing$).

Marginalizing over $m \equiv \{k_1, \ldots, k_s\} \subseteq V$ results in the graph $G^{m\cdot}$ which is defined recursively as $G^{m\cdot} := (G^{m\setminus\{k_s\}\cdot})^{\{k_s\}\cdot} = (V\setminus m, E^{m\cdot})$, where $E^{m\cdot} := (E^{m\setminus\{k_s\}\cdot})^{\{k_s\}\cdot}$. Clearly, it must be shown that this does not depend on the order $k_1, \ldots, k_s$ of the members of m. But this is an immediate consequence of the next proposition. Recall that a path π is called collisionless if all its intermediate points are non-colliders only, i.e., $C_\pi = \varnothing$.

Proposition 3.3. *Let $G^{m\cdot} = (V\setminus m, E^{m\cdot})$ be the graph obtained by marginalizing $G = (V, E)$ over the vertices $k_1, \ldots, k_s$ of $m \subseteq V$ (in some fixed order). Let $\{i, j\} \subseteq V\setminus m$. Then the following two statements are equivalent:*

(i) $\exists e \in E^{m\cdot}$: *i is a head (tail) of e, and j is a head (tail) of e.*
(ii) *There exists in G a collisionless path π between i and j with all intermediate points in m, such that i is a head (tail) end-point of π and j is a head (tail) end-point of π.*

If $i = j$ it is understood here that in both (i) and (ii) only 'tail' applies.

Corollary 3.4. *The graph $G^{m\cdot}$ is a well-defined MC graph.*

Proof of Proposition 3.3. The proof that (ii) $\Rightarrow$ (i) is by induction on $s \equiv |m|$. If $s = 1$ the result is immediate from Table 1, so assume $s > 1$. Let $\pi \equiv (i, e_1, i_1, \ldots, i_{n-1}, e_n, j)$ be a collisionless path from i to j having all its intermediate points in m. If

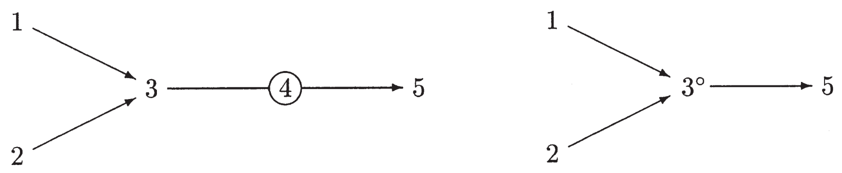

Figure 4. (a) MC graph G. (b) After marginalizing over vertex 4 the graph $G^{\{4\}\cdot}$ is obtained. The symbol 3° denotes the configuration of vertex 3 plus self loop $3 - 3$.

$\{i_r|0 < r < n\} \subseteq m\backslash\{k_s\}$, then the induction hypothesis entails that $e \in E^{m\backslash\{k_s\}\cdot}$, so $e \in E^{m\cdot}$ as well. If $k_s \in \{i_r|0 < r < n\}$, say $k_s = i_{r_t}$, $t = 1, \ldots, q$, where $r_1 < \ldots < r_q$, then there are two cases to consider.

First, $q = 1$, that is, k_s occurs only once as intermediate point on π. The induction hypothesis implies that there exist $f, g \in E^{m\backslash\{k_s\}\cdot}$ such that $f = i \cdots k_s$ and $g = k_s \cdots j$, so after marginalizing $G^{m\backslash\{k_s\}\cdot}$ over k_s (using the non-boxed entries of Table 1) the edge $e \equiv f \circ g$ will result which has the required properties.

Second, $q \geqslant 2$. By the induction hypothesis, there exist $f, g \in E^{m\backslash\{k_s\}\cdot}$ such that $f = i \cdots i_{r_1}$ and $g = k_{r_q} \cdots j$, so if k_s is a tail of either f or g, then marginalizing over k_s (again using the non-boxed entries of Table 1) will result in the edge $e \equiv f \circ g$ which has the required properties. On the other hand, if k_s is a head of both f and g (hence, of both e_{r_1} and e_{r_q+1}), then the edge $e \equiv f \circ g$ having the required properties will result by applying the boxed entries of Table 1 – if it can be shown that the graph $G^{m\backslash\{k_s\}\cdot}$ contains the self loop k_s°. But this follows from the induction hypothesis, since at least one of the loops $(k_s, e_{r_t+1}, i_{r_t+1}, \ldots, e_{r_t+1}, k_s)$, $t = 1, \ldots, q-1$ (each of them a subpath of π) must have both its end-points as tails (note that k_s is a tail of both e_{r_1+1} and e_{r_q}, since it would occur as a collider on π otherwise).

The induction proof (again on $s \equiv |m|$) that (i) $\Rightarrow$ (ii) is rather similar, and is left to the reader. □

If $G = (V, E)$ is a directional graph (that is, G has no lines or undirected self loops), then the intermediate points of a collisionless path between vertices i and j are ancestors of either i or j. Combining this observation with Proposition 3.3 immediately leads to the following.

Proposition 3.5. *Suppose $G = (V, E)$ is a directional graph. Let $a \subseteq V$ be a G-ancestral set, and let $m \equiv V\backslash a$. Then $G^{m\cdot} = G_a$.*

The next result entails that the class of MC graphs is indeed closed under marginalization.

Proposition 3.6. *Let $G = (V, E)$ be a graph, and let $m \subseteq V$. Then, for all pairwise disjoint subsets a, b and c of $V\backslash m$, a and b are separated by c in G if and only if a and b are separated by c in $G^{m\cdot}$.*

Proof. First note that the result is trivial if $m = \varnothing$. Clearly, due to the recursive definition of $G^{m\cdot}$ it suffices to prove the proposition for the case $m = \{k\}$, for $k \in V$ arbitrary. Suppose a, b and c are pairwise disjoint subsets of $V\backslash m$. We must show that there exists an open path from $i \in a$ to $j \in b$ given c in G if and only if there exists an open path from $i \in a$ to $j \in b$ given c in $G^{m\cdot}$. Let $\pi \equiv (i_0, e_1, i_1, \ldots, e_n, i_n)$ be an open path from $i \in a$ to $j \in b$ given c in G. Without loss of generality we may assume that the length n of π is minimal. If k is not an intermediate point of π, then π obviously is an open path given c in $G^{m\cdot}$ as well. If $k = i_r$ (say) is an intermediate point of π, then k is a non-collider on π, since $C_\pi \subseteq c$ and $k \notin c$ (this includes the possibility that k occurs as a self loop on π). Applying the $\circ$-operator to e_r,

e_{r+1} then has two possible consequences. It either results in the non-void edge $e_r \circ e_{r+1} \in E^{m\cdot}$, so in π the subpath $(i_{r-1}, e_r, i_r, e_{r+1}, i_{r+1})$ can be substituted by $(i_{r-1}, e_r \circ e_{r+1}, i_{r+1})$; repeating this for all occurences of k as an intermediate point of π eventually results in an open path from $i \in a$ to $j \in b$ given c in $G^{m\cdot}$. Alternatively, the edge $e_r \circ e_{r+1}$ is void (namely, when $i_{r-1} = i_{r+1}$, and $e_r \in \{i_{r-1} \leftarrow i_r, \, i_{r-1} \leftrightarrow i_r\}$ or $e_{r+1} \in \{i_r \rightarrow i_{r+1}, \, i_r \leftrightarrow i_{r+1}\}$). But as the length n of π is minimal, the latter case cannot happen. (Otherwise, the sequence $(e_r, i_r, e_{r+1}, i_{r+1})$ can be deleted from π to obtain a shorter open path; note that $N_\pi \cap C_\pi = \varnothing$, hence, if $i_{r-1} = i_{r+1}$ is a (non-)collider on π at position $r-1$, then it is also a (non-)collider at position $r+1$, and this collider status remains unchanged when the sequence $(e_r, i_r, e_{r+1}, i_{r+1})$ is deleted from π.) To prove the converse, let $\pi \equiv (i_0, e_1, i_1, \ldots, e_n, i_n)$ be an open path from $i \in a$ to $j \in b$ given c in $G^{m\cdot}$. Applying Proposition 3.1 to each edge of π immediately renders a path in G. This path will be open given c since, if k is one of its intermediate points, then k will be a non-collider, and $k \notin c$ by hypothesis. (Of course, the collider status of the other intermediate points remains invariant.) □

We will now show that the class of MC graphs is closed under conditioning as well. As was the case with closure under marginalization, the class of undirected graphs also has this closure property, but no other class of graphs discussed thus far in the literature shares it. We will first make explicit what we mean by ‘closure under conditioning’.

3.2. Conditioning

Let $\mathcal{G}$ denote a class of graphs such that for each member of $\mathcal{G}$ a concept of ‘separation’ is defined whose associated independence model satisfies the properties stated in Proposition 2.10. Let $G = (V, E)$ be a member of $\mathcal{G}$, and suppose that, for all subsets $c \subseteq V$, the class $\mathcal{G}$ contains a (possibly not unique) graph $G^{\cdot c} \equiv (V \backslash c, F)$ with vertex set $V \backslash c$ and edge set $F = F(c)$ (say), such that, for all pairwise disjoint subsets a, b and d of $V \backslash c$, a and b are separated by $c \cup d$ in G if and only if a and b are separated by d in $G^{\cdot c}$. Clearly, if $c = V$ of $c = \varnothing$ this is trivially satisfied by putting $G^{\cdot V} = (\varnothing, \varnothing)$, and $G^{\cdot \varnothing} = G$. The class $\mathcal{G}$ is *closed under conditioning* if the stated property holds for all its members G. As was the case with marginalization, it would be nice if a simple algorithm consisting of local operations only were available to construct $G^{\cdot c}$ from G and c. Fortunately, for the class of MC graphs (as for the class of undirected graphs) this is indeed the case, as will be shown below.

Example 3.7. For undirected graphs, the operation of conditioning is even simpler than the operation of marginalizing. Consider the undirected graph G in Figure 5(a). Vertex 4 is enclosed by a square to indicate conditioning upon this vertex, that is, we wish to obtain $G^{\cdot\{4\}}$. This can be done by simply wiping out vertex 4 and all edges adjacent to it. No new edges are to be added. In this way, $G^{\cdot\{4\}}$ is obtained as the graph in Figure 5(b). Clearly, for all pairwise disjoint subsets a, b and d of $\{1, 2, 3, 5\}$, a and b are separated by $d \cup \{5\}$ in G if and only if a and b are separated by d in $G^{\cdot\{5\}}$.

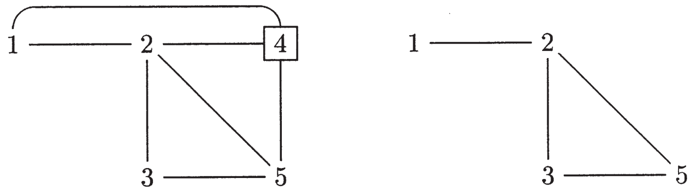

Figure 5. (a) Undirected graph G. (b) After conditioning on vertex 4 the graph $G^{\cdot\{4\}}$ is obtained. See Example 3.7 for explanation.

Let $G = (V, E)$ be an MC graph. Suppose i, j and k are vertices of G, $k \notin \{i, j\}$, such that $f \in E$ and $g \in E$ satisfy $f = i \cdots k$ and $g = k \cdots j$. Then (i, f, k, g, j) is a path of length 2 from i to j with intermediate point k. Now define the edge $f \,\Box\, g = g \,\Box\, f$ by the *boxed* entries of Table 1. Thus, if $i \neq j$, then the edge $f \,\Box\, g$ is non-void if and only if k is a collider on the path (i, f, k, g, j). If $i = j$, then $f \,\Box\, g$ is non-void if and only if $f = g = i \to k$ (and $f \,\Box\, g$ equals the undirected self loop i° in that case). We define, for $k \in V$, the MC graph $G^{\cdot\{k\}}$ as the graph on $V\backslash\{k\}$ obtained after the $\Box$-operation of Table 1 is applied to all pairs of edges (f, g) satisfying $f = i \cdots k$, and $g = k \cdots j$, where $\{i, j\} \subseteq V\backslash\{k\}$. Formally, $G^{\cdot\{k\}} := (V\backslash\{k\}, E^{\cdot\{k\}})$, where the edge set $E^{\cdot\{k\}}$ is given by $E^{\cdot\{k\}} := \{e \in E | [e] \subseteq V\backslash\{k\}\} \cup \{f \,\Box\, g | f = i \cdots k,\ g = k \cdots j,\ \{i, j\} \subseteq V\backslash\{k\}\}$. Again, it is understood that Convention 2.7 applies, that is, if $i = j$, then the resulting self loop $f \,\Box\, g$ is void, unless it equals the undirected self loop $i^\circ = i - i$. Conditioning over $c \equiv \{k_1, \ldots, k_s\} \subseteq V$ results in the graph $G^{\cdot c}$ which is defined recursively as $G^{\cdot c} := (G^{\cdot c\backslash\{k_s\}})^{\cdot\{k_s\}} = (V\backslash c, E^{\cdot c})$, where $E^{\cdot c} := (E^{\cdot c\backslash\{k_s\}})^{\cdot\{k_s\}}$. Clearly, it must be shown that this does not depend on the order $k_1, \ldots, k_s$ of the members of c. But this is an immediate consequence of the next proposition. Recall that a path π is a pure collision path if all its intermediate points are colliders only, that is, $N_\pi = \varnothing$.

Proposition 3.8. *Let $G^{\cdot c} = (V\backslash c, E^{\cdot c})$ be the graph obtained by conditioning $G = (V, E)$ over the vertices $k_1, \ldots, k_s$ of $c \subseteq V$ (in some fixed order). Let $\{i, j\} \subseteq V\backslash c$. Then the following two statements are equivalent:*

(i) *$\exists e \in E^{\cdot c}$: i is a head (tail) of e, and j is a head (tail) of e.*
(ii) *There exists in G a pure collision path π between i and j with all intermediate points in c, such that i is a head (tail) end-point of π and j is a head (tail) end-point of π.*

If $i = j$ it is understood here that in both (i) and (ii) only 'tail' applies.

Corollary 3.9. *The graph $G^{\cdot c}$ is a well-defined MC graph.*

Proof of Proposition 3.8. The proof that (ii) $\Rightarrow$ (i) is by induction on $s \equiv |c|$. If $s = 1$ the result is immediate from Table 1, so assume $s > 1$. Let $\pi \equiv (i, e_1, i_1, \ldots, i_{n-1}, e_n, j)$ be a pure collision path from i to j having all its intermediate points in c. If $\{i_r | 0 < r < n\} \subseteq c\backslash\{k_s\}$, then the induction hypothesis entails that $e \in E^{\cdot c\backslash\{k_s\}}$, so

$e \in E^{\cdot c}$ as well. If $k_s \in \{i_r | 0 < r < n\}$, say $k_s = i_{r_t}$, $t = 1, \ldots, q$, where $r_1 < \ldots < r_q$, then there are two cases to consider.

First, $q = 1$, that is, k_s occurs only once as intermediate point on π. The induction hypothesis implies that there exist $f, g \in E^{\cdot c\setminus\{k_s\}}$ such that $f = i \cdots k_s$ and $g = k_s \cdots j$, so after conditioning $G^{\cdot c\setminus\{k_s\}}$ upon k_s (using the boxed entries of Table 1) the edge $e \equiv f \,\square\, g$ will result which has the required properties.

Second, $q \geqslant 2$. By the induction hypothesis, there exist $f, g \in E^{\cdot c\setminus\{k_s\}}$ such that $f = i \cdots i_{r_1}$ and $g = i_{r_q} \cdots j$, so (as k_s is a head of both f and g), conditioning on k_s results in the edge $e \equiv f \,\square\, g$ which has the required properties.

The induction proof (again on $s \equiv |c|$) that (i) $\Rightarrow$ (ii) is similar, and is left to the reader. $\square$

Proposition 3.10. *Let $G = (V, E)$ be a graph, and let $c \subseteq V$. Then, for all pairwise disjoint subsets a, b and d of $V\setminus c$, a and b are separated by $c \cup d$ in G if and only if a and b are separated by d in $G^{\cdot c}$.*

Proof. First note that the result is trivial if one of a, b or c is empty, so assume otherwise. We must show that there exists an open path from $i \in a$ to $j \in b$ given $c \cup d$ in G if and only if there exists an open path from $i \in a$ to $j \in b$ given d in $G^{\cdot c}$. By the recursive definition of $G^{\cdot c}$, $G^{\cdot c} = (G^{\cdot c\setminus\{k\}})^{\cdot\{k\}}$ whenever $k \in c$. Thus, it suffices to prove the proposition for the case $c = \{k\}$, $k \in V$.

Let $\pi \equiv (i_0, e_1, i_1, \ldots, e_n, i_n)$ be an open path from $i \in a$ to $j \in b$ given $c \cup d$ in G, hence $C_\pi \subseteq c \cup d$ and $N_\pi \subseteq V\setminus(c \cup d)$. Without loss of generality, we may assume that the length n of π is minimal. If k is not an intermediate point of π, then $C_\pi \subseteq d$, and since $N_\pi \subseteq (V\setminus c)\setminus d$, π obviously is an open path given d in $G^{\cdot c}$. If k is an intermediate point of π, then k is a collider on π, since $k \in c \cup d \subseteq V\setminus N_\pi$. Since the length of π is minimal, k occurs only once as an intermediate point of π (otherwise, the loop between two occurrences of k can be deleted from π, thus resulting into a shorter path), say $k = i_r$. Applying the $\square$-operator to e_r, e_{r+1} then results in the non-void edge $e_r \,\square\, e_{r+1} \in E^{\cdot c}$, so the subpath $(i_{r-1}, e_r, i_r, e_{r+1}, i_{r+1})$ of π can be substituted by $(i_{r-1}, e_r \,\square\, e_{r+1}, i_{r+1})$ to obtain a path $\sigma \equiv (i_0, \ldots, i_{r-1}, e_r \,\square\, e_{r+1}, i_{r+1} \ldots, i_n)$ in $G^{\cdot c}$. Since $C_\sigma \subseteq (c \cup d)\setminus\{k\} = d$ and $N_\sigma = N_\pi \subseteq (V\setminus c)\setminus d$, σ is open given d in $G^{\cdot c}$.

To prove the converse, let $\sigma \equiv (i_0, e_1, i_1, \ldots, e_n, i_n)$ be an open path from $i \in a$ to $j \in b$ given d in $G^{\cdot c}$ (hence, $C_\sigma \subseteq d$, $N_\sigma \subseteq (V\setminus c)\setminus d$). Applying Proposition 3.8 to each edge of σ immediately renders a path π (say) in G. If k is one of the intermediate points of π, then it can only occur as a collider. Since the collider status of the other intermediate points remains unchanged, $C_\pi \subseteq d \cup \{k\} = c \cup d$ and $N_\pi = N_\sigma \subseteq V\setminus(c \cup d)$, so π is open given $c \cup d$ in G. $\square$

3.3. Marginalizing and conditioning

Let $G = (V, E)$ be a graph, and let m and c be disjoint subsets of V. As noted in Cox and Wermuth (1996, pp. 207–8), in order for the operations of marginalizing and conditioning in graphs to be consistent with standard probability theory the graphs $(G^{m\cdot})^{\cdot c}$, and $(G^{\cdot c})^{m\cdot}$ should define the same independence model over their vertex set $V\backslash(m \cup c)$. In fact, the next result shows that both graphs are identical.

Proposition 3.11. *Let $G = (V, E)$ be an MC graph, and suppose m and c are disjoint subsets of V. Let $(G^{m\cdot})^{\cdot c} = ((V\backslash m)\backslash c, (E^{m\cdot})^{\cdot c})$ be the graph obtained by first marginalizing over m and then conditioning on c. Let $(G^{\cdot c})^{m\cdot} = ((V\backslash c)\backslash m, (E^{\cdot c})^{m\cdot})$ be the graph obtained by first conditioning on c and then marginalizing over m. Then the following three statements are equivalent:*

(i) $\exists e \in (E^{m\cdot})^{\cdot c}$ *: i is a head (tail) of e, and j is a head (tail) of e.*
(ii) $\exists e \in (E^{\cdot c})^{m\cdot}$ *: i is a head (tail) of e, and j is a head (tail) of e.*
(iii) There exists in G a path π between i and j satisfying $N_\pi \subseteq m$, $C_\pi \subseteq c$, i is a head (tail) end-point of π and j is a head (tail) end-point of π.

In all cases, if $i = j$ it is understood that only 'tail' applies.

Proof. Using Propositions 3.3 and 3.8 it is immediately clear that both (i) and (ii) are equivalent to (iii). □

Henceforth we will denote $(G^{m\cdot})^{\cdot c} = (G^{\cdot c})^{m\cdot}$ simply by $G^{m\cdot c}$.

Corollary 3.12. *The graph $G^{m\cdot c} = (V\backslash(m \cup c), E^{m\cdot c})$ is a well-defined MC graph.*

Theorem 3.13. *Let $G = (V, E)$ be an MC graph, and suppose m and c are disjoint subsets of V. Then, for all pairwise disjoint subsets a, b and d of $V\backslash(m \cup c)$, a and b are separated by $c \cup d$ in G if and only if a and b are separated by d in $G^{m\cdot c}$.*

Proof. Just apply Propositions 3.6 and 3.10 (in either order). □

4. The Markov property for MC graphs

Thus far all results have been strictly graph-theoretic in nature. We will now introduce the Markov property for MC graphs, thereby giving the main results in the preceding section a probabilistic slant.

Let V be a finite set, and let, for $i \in V$, $(\mathrm{X}_i, \mathcal{X}_i)$ be a measurable space. Suppose P is a multivariate probability distribution defined on the product space $(\mathrm{X}, \mathcal{X}) \equiv (\times_{i\in V}\mathrm{X}_i, \times_{i\in V}\mathcal{X}_i)$. We will assume that each space $(\mathrm{X}_i, \mathcal{X}_i)$ is sufficiently regular to ensure the existence of *regular* conditional probability distributions. More specifically, we assume that all versions of conditional distributions encountered induce probability measures and two such versions define the same probability measures outside some P-negligible set. For $a \subseteq V$, define $(\mathrm{X}_a, \mathcal{X}_a) = (\times_{i\in a}\mathrm{X}_i, \times_{i\in a}\mathcal{X}_i)$ ($\mathrm{X}_\varnothing \equiv \mathrm{X}$, endowed with the trivial σ-algebra). We may assume P is the distribution of an X-valued random vector $X \equiv (X_i : i \in V)$, and for $a \subseteq V$, define $X_a = (X_i : i \in a)$; similarly, if $x \in \mathrm{X}$,

$x_a := (x_i : i \in a)$ ($X_\emptyset \equiv x_\emptyset \equiv$ constant). The marginal distribution of X_a on $(\mathrm{X}_a, \mathcal{X}_a)$ is denoted by P_a, so $P_a(\mathrm{A}) := P(X_a \in \mathrm{A}) = P(\mathrm{A} \times \mathrm{X}_{V\backslash a})$ whenever $\mathrm{A} \subseteq \mathrm{X}_a$ is a measurable subset. If $c \subseteq V$, $c \cap a = \emptyset$, the expression $P_{X_a|X_c=x_c}$ is used to denote the conditional distribution of X_a given $X_c = x_c$, i.e., $P_{X_a|X_c=x_c}(\mathrm{A}) \equiv P(X_a \in \mathrm{A}|X_c = x_c)$. (If $c = \emptyset$ this denotes the marginal distribution P_a of X_a.) For pairwise disjoint subsets a, b and c of V, let $a \perp\!\!\!\perp b|c[P]$ denote that X_a and X_b are conditionally independent given X_c under distribution P, that is, $P(X_a \in \mathrm{A}, X_b \in \mathrm{B}|X_c = x_c) = P(X_a \in \mathrm{A}|X_c = x_c) \quad P(X_b \in \mathrm{B}| X_c = x_c)$, for all measurable subsets $\mathrm{A} \subseteq \mathrm{X}_a$ and $\mathrm{B} \subseteq \mathrm{X}_b$, and for P_c-almost every $x_c \in \mathrm{X}_c$. Recall that, by the assumed regularity conditions, the left-hand side and both factors on the right-hand side denote regular versions of the conditional distribution of, respectively, $X_{a\cup b}$, X_a and X_b given X_c, that is, for P_c -a.e. $x_c \in \mathrm{X}_c$ they are all probability measures. Now suppose $G = (V, E)$ is an MC graph, hence that the components of $X \equiv (X_i : i \in V)$ are indexed by the vertices of G.

Definition 4.1. *Markov property for MC graphs. The distribution $P \equiv P_V$ is called a* Markov random field with respect to *G (G-Markov for short) if, for all pairwise disjoint subsets a, b and c of V, $a \perp\!\!\!\perp b|c[P]$ whenever a and b are separated by c.*

Theorem 4.2. *Let $G = (V, E)$ be an MC graph, and let P denote the probability distribution of $X \equiv (X_i : i \in V)$. Then P is G-Markov if and only if, for all disjoint subsets m and c of V, the conditional distribution $P_{X_s|X_c=x_c}$ is $G^{m\cdot c}$-Markov, P_c-a.e. $x_c \in \mathrm{X}_c$. Here $s \equiv V\backslash(m \cup c)$.*

Proof. Sufficiency is immediate by putting $m = c := \emptyset$ (hence $s = V$).

To see necessity, let a, b and d be pairwise disjoint subsets of s, and assume that a and b are separated by d in $G^{m\cdot c}$. By Theorem 3.13, a and b are separated by $c \cup d$ in G, so $a \perp\!\!\!\perp b|c \cup d[P_{s\cup c}]$. Equivalently, $a \perp\!\!\!\perp b|d[P_{X_s|X_c=x_c}]$, P_c-a.e. $x_c \in X_c$. Hence $P_{X_s|X_c=x_c}$ is $G^{m\cdot c}$-Markov, P_c-a.e. $x_c \in \mathrm{X}_c$. □

Example 4.3. In Besag (1974) a submodel of the following Gaussian linear structural equation system is discussed for random variables Y_{ij}, where the index (i, j) is an integer pair which varies over some finite rectangular lattice $A \subseteq \mathbb{Z} \times \mathbb{Z}$:

$$Y_{ij} = b_{i-1,j}Y_{i-1,j} + b'_{i+1,j}Y_{i+1,j} + c_{i,j-1}Y_{i,j-1} + c'_{i,j+1}Y_{i,j+1} + E_{ij}.$$

Here $(E_{ij} : (i, j) \in A) \sim \mathcal{N}_{|A|}(0, \Psi)$ (i.e., a multivariate normal distribution with expectation 0 and positive definite covariance matrix Ψ), where it is assumed that Ψ is a diagonal matrix. In Koster (1999) it is shown that the distribution of $Y_A \equiv (Y_{ij} : (i, j) \in A)$ is Markov relative to the path diagram associated with this system (Figure 6 shows a 5×5 subgraph of the path diagram.). Let us call the entire graph G; thus the distribution of Y_A is G-Markov. To give an example of what can be deduced from this, note that (i, j) is separated from $A\backslash A_{ij}$ by $A_{ij}\backslash\{(i, j)\}$, where $A_{ij} := \{(k, l) \in A : |i - k| + |j - l| \leqslant 2\}$, that is, A_{ij} is the subset inside the dashed line contour. Hence $Y_{ij} \perp\!\!\!\perp Y_{A\backslash A_{ij}}|Y_{A_{ij}\backslash\{(i,j)\}}$ is entailed by the Markov property. Now suppose the boundary of the system (i.e., the outer layer of the rectangular lattice), call it bd(A), is observed, say $Y_{\mathrm{bd}(A)} = y_{\mathrm{bd}(A)}$. Then by Theorem 4.2 the conditional distribution of $Y_{A\backslash \mathrm{bd}(A)}$ given $Y_{\mathrm{bd}(A)} = y_{\mathrm{bd}(A)}$ is Markov relative to the graph $G^{\emptyset\cdot \mathrm{bd}(A)}$. It is easy to see that

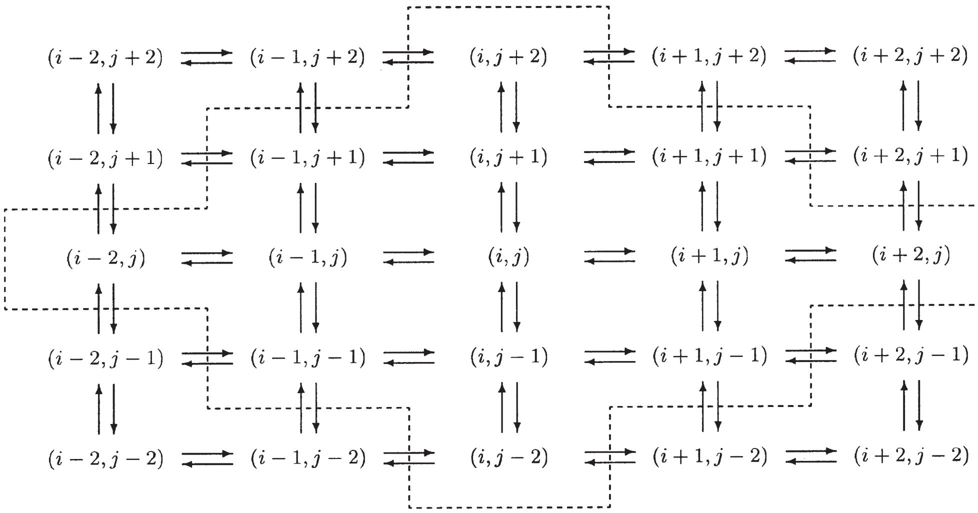

Figure 6. Part of path diagram of Gaussian linear structural equations system. See Example 4.3 for explanation.

the graph $G^{\varnothing \cdot \mathrm{bd}(A)}$ has the same structure as G, except for the fact that the vertices at its boundary (i.e., $\mathrm{bd}(B)$, where $B \equiv A \backslash \mathrm{bd}(A)$) become self loops. As this does not change the separation properties of the graph (in *this* example) we can conclude that observing the boundary of the system does not influence the Markov properties of the remainder of the system. On the other hand, if the boundary is 'neglected', then $G^{\mathrm{bd}(A) \cdot \varnothing}$ is the graph describing the Markov properties of the remaining variables Y_B. In this graph, each pair of distinct vertices in the boundary (say, i, $j \in \mathrm{bd}(B)$) are connected by each of the edges $i \rightarrow j$, $i \leftarrow j$ and $i \leftrightarrow j$. This strongly changes the Markov properties of variables at and near the boundary of the remaining system, although the Markov properties of deeper layers remain unchanged.

5. Concluding remarks

In this paper we have introduced the class of MC graphs and shown that it is closed under the operations of marginalizing over a certain subset of vertices m, or conditioning on a subset of vertices c. The operations on the graph G can be carried out 'locally' by applying Table 1 to each vertex in $m \cup c$ successively, rendering the well-defined MC graph $G^{m \cdot c}$. Proposition 3.11, however, shows that the edges of graph $G^{m \cdot c}$ can be characterized by a 'global' criterion as well. Key to these results is the deletion of the requirement that all vertices along a path be distinct, and the simplification of Pearl's d-separation concept which is made possible by this modification.

In the recent literature on graphical models, two other approaches to the problem of marginalizing and conditioning in graphical models can be found. The first is based on *summary graphs* (Cox and Wermuth 1996; Wermuth *et al*. 1998), and the second considers *ancestral graphs* (Richardson and Spirtes 2000). Both summary graphs and ancestral graphs are MC graphs, and the separation criteria involved are equivalent to Definition 2.3. However, the operations of marginalizing and conditioning are defined differently for each type of graph; thus, for example, after marginalizing a given DAG $G = (V, E)$ over a subset $m \subseteq V$ and conditioning over a subset $c \subseteq (V \backslash m)$, each of the three approaches may render a distinct (though separation-equivalent) 'graphical object'.

Pearl (1995; 2000) uses DAGs to represent causal theories formulated as (possibly nonlinear) recursive structural equation systems. Here each edge (arrow) has a substantive interpretation as a causal relation. Whenever a summary graph has a DAG as its point of departure, the independence model associated with it is said to be accounted for by some *data generating process*. It is shown in Richardson and Spirtes (2000) that the class of maximal ancestral graphs constitutes the smallest class of graphs that contains the class of DAGs, and which is closed under marginalizing and conditioning (in *their* sense). The class of independence models associated with maximal ancestral graphs therefore coincides with the class of models which can be accounted for by a data generating process. The present paper addresses the marginalization and conditioning problem for a (much) larger class of graphs, containing at least all undirected graphs and all directional graphs. Directional graphs are important, since they occur as path diagrams in Gaussian linear structural equation systems (Jöreskog and Sörbom 1989). Example 4.3 shows that our results can be applied straightforwardly to such systems. It is tempting to view the present paper, particularly Table 1, as offering a mode of data generation (in some generalized sense) for undirected edges (lines) and bidirected edges (arcs). Roughly speaking, lines are created by conditioning upon common effects, whereas arcs are created by marginalizing over common causes. However, it is not generally true that, for example, an MC graph can be obtained by marginalizing and conditioning starting from some directed (cyclic or acyclic) graph. A counter-example is given in Richardson and Spirtes (2000). Note that similar remarks hold for chain graphs when the Lauritzen–Wermuth–Frydenberg separation concept is employed (Lauritzen and Wermuth 1989; Frydenberg 1990). In Richardson (1998) it is shown that, in general, Lauritzen–Wermuth–Frydenberg chain graphs containing undirected edges cannot be obtained by marginalization or conditioning in DAGs. Therefore, in this sense, neither chain graphs endowed with Lauritzen–Wermuth–Frydenberg separation nor MC graphs generally have a straightforward substantive interpretation.

Acknowledgement

This paper was greatly inspired by an unpublished manuscript by Judea Pearl. I also thank Thomas Richardson for many helpful remarks, and two unknown referees for their critical remarks which led to improvements of the paper. The paper was presented at the Fifth World Congress of the Bernoulli Society for Mathematical Statistics and Probability and the 63rd Annual Meeting of the Institute of Mathematical Statistics, Guanajuato, Mexico, on 15–20 May, 2000. An earlier version was presented at the Joint Statistical Meeting of IMS/ASA/WNAR-ENAR/SSC, Dallas, Texas, on 9–13 August, 1998.

References

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

Besag, J.E. (1974) Spatial interaction and the statistical analysis of lattice systems (with discussion). *J. Roy. Statist. Soc. Ser. B*, **36**, 192–236.

Cowell, R.G., Dawid, A.P., Lauritzen, S.L. and Spiegelhalter, D.J. (1999) *Probabilistic Networks and Expert Systems.* New York: Springer-Verlag.

Cox, D.R. and Wermuth, N. (1996) *Multivariate Dependencies.* London: Chapmann & Hall.

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.

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

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

Jordan, M.I. (ed.) (1998) *Learning in Graphical Models.* Dordrecht: Kluwer Academic Publishers.

Jöreskog, K.G. and Sörbom, D. (1989) *LISREL 7. A Guide to the Program and Applications*, 2nd edn. Chicago: SPSS.

Koster, J.T.A. (1996) Markov properties of nonrecursive causal models. *Ann. Statist.*, **24**, 2148–2177.

Koster, J.T.A. (1997) Gibbs and Markov properties of graphs. *Ann. Math. Artificial Intelligence*, **21**, 13–26.

Koster, J.T.A. (1999) On the validity of the Markov interpretation of path diagrams of Gaussian structural equations systems with correlated errors. *Scand. J. Statist.*, **26**, 413–431.

Lauritzen, S.L. (1979) Lectures on Contingency Tables. Institute for Mathematical Statistics, University of Copenhagen.

Lauritzen, S.L. and Wermuth, N. (1989) Graphical models for association between variables, some of which are qualitative and some quantitative. *Ann. Statist.*, **17**, 31–57.

Lauritzen, S.L. and Spiegelhalter, D.J. (1988) Local computations with probabilities on graphical structures and their application to expert systems. *J. Roy. Statist. Soc. Ser. B*, **50**, 157–224.

Lauritzen, S.L., Dawid, A.P., Larsen, B.N. and Leimer, H.G. (1990) Independence properties of directed Markov fields. *Networks*, **20**, 491–505.

Matúš, F. (1997) Conditional independence structures examined via minors. *Ann. Math. Artificial Intelligence*, **21**, 99–128.

Pearl, J. (1988) *Probabilistic Inference in Intelligent Systems.* San Mateo, CA: Morgan Kaufmann.

Pearl, J. (1995) Causal diagrams for empirical research (with discussion). *Biometrika*, **82**, 669–710.

Pearl, J. (2000) *Causality: Models, Reasoning and Inference.* Cambridge: Cambridge University Press.

Richardson, T. (1998) Chain graphs and symmetric associations. In M.I. Jordan (ed.), *Learning in Graphical Models.* Dordrecht: Kluwer Academic Publishers.

Richardson, T. (1999) Markov properties for acyclic directed mixed graphs. Technical Report No. 353, Department of Statistics, University of Washington.

Richardson, T. and Spirtes, P. (2000) Ancestral graph Markov properties. Technical Report No. 375, Department of Statistics, University of Washington.

Spirtes, P. (1995) Directed cyclic graphical representation of feedback models. In P. Besnard and S. Hanks (eds), *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*. San Francisco: Morgan Kaufmann.

Spirtes, P., Richardson, T., Meek, C., Scheines, R. and Glymour, C. (1998) Using path diagrams as a structural equation modeling tool. *Sociological Methods Res.*, **27**, 182–225.

Studený, M. (1993) Formal properties of conditional independence in different calculi of AI. In

M. Clarke, R. Kruse and S. Moral (eds), *Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, Lecture Notes in Comput. Sci. 747, pp. 341–438. Berlin: Springer-Verlag.

Studený, M. (1998) Bayesian networks from the point of view of chain graphs. In G.F. Cooper and S. Moral (eds), *Uncertainty in Artificial Intelligence. Proceedings of the 14th Conference*, pp. 496–503. San Francisco: Morgan Kaufmann.

Wermuth, N., Cox, D.R. and Pearl, J. (1998) Explanations for multivariate structures derived from univariate recursive regressions. Report on Stochastics 94-1 (revised 98-1), University of Mainz, Germany.

Received June 2000 and revised February 2002