Università degli Studi di Modena e Reggio Emilia

On Block Ordering of Variables in Graphical Modelling

Alberto Roverato
Luca La Rocca

R 29-04 ottobre 2004

RESEARCH REPORT

Dipartimento di Scienze Sociali, Cognitive e Quantitative

Via G. Giglioli Valle, 9
42100 Reggio Emilia
Tel: +39 0522 522511
Fax: +39 0522 522512
Internet: http:://www.cei.unimore.it

On Block Ordering of Variables in Graphical Modelling

Alberto Roverato* and Luca La Rocca
University of Modena and Reggio Emilia, Italy
email: {roverato.alberto,larocca.luca}@unimore.it

October 19, 2004

Abstract

In graphical modelling, the existence of substantive background knowledge on block ordering of variables is used to perform structural learning within the family of chain graphs in which every block corresponds to an undirected graph and edges joining vertices in different blocks are directed in accordance with the ordering. We show that this practice may lead to an inappropriate restriction of the search space and introduce the concept of *labelled block ordering* $\mathcal{B}$ corresponding to a family of $\mathcal{B}$*-consistent chain graphs* in which every block may be either an undirected graph or a directed acyclic graph or, more generally, a chain graph. In this way we provide a flexible tool for specifying subsets of chain graphs, and we observe that the most relevant subsets of chain graphs considered in the literature are families of $\mathcal{B}$-consistent chain graphs for the appropriate choice of $\mathcal{B}$. Structural learning within a family of $\mathcal{B}$-consistent chain graphs requires to deal with Markov equivalence. We provide a graphical characterisation of equivalence classes of $\mathcal{B}$-consistent chain graphs, namely the $\mathcal{B}$*-essential graphs*, as well as a procedure to construct the $\mathcal{B}$-essential graph for any given equivalence class of $\mathcal{B}$-consistent chain graphs. Both largest chain graphs and essential graphs turn out to be special cases of $\mathcal{B}$-essential graphs.

Keywords: $\mathcal{B}$-essential graph; background knowledge; block ordering; chain graph; conditional independence; graphical model; labelled block ordering; Markov equivalence; Markov property; meta-arrow.

1 Introduction

A graphical Markov model for a random vector X_V is a family of probability distributions satisfying a collection of conditional independencies encoded by a graph with vertex set V. Every variable is associated with a vertex of the graph and the conditional independencies between variables are determined through a Markov property. A *chain graph* (CG) has both directed edges (arrows) and undirected edges (lines) but not semi-directed cycles. Two different Markov properties have been proposed for CGs: the Lauritzen-Wermuth-Frydenberg (LWF) Markov property (Lauritzen and Wermuth, 1989 and Frydenberg, 1990)

*Address for correspondence: Dipartimento di Scienze Sociali Cognitive e Quantitative, Via Giglioli Valle, 9, 42100 Reggio Emilia, Italy.

and the Andersson-Madigan-Perlman Markov property (Andersson *et al.*, 2001). Here, we consider the case in which the conditional independence structure of X_V is encoded by a CG $\mathcal{G}$ under the LWF-Markov property and $\mathcal{G}$ has to be learnt from data. Recent monographs on this subject include Pearl (1988, 2000), Whittaker (1990), Cox and Wermuth (1996), Lauritzen (1996), Jordan (1998), Cowell *et al.* (1999) and Edwards (2000).

The first step of any structural learning procedure is the specification of a search space, i.e. of a set of candidate structures, and this operation may present difficulties when using graphical models. A first problem derives from *Markov equivalence*. Let $\mathbb{H}_p \equiv \mathbb{H}$ denote the set of all CGs on $p = |V|$ vertices. Two different CGs in $\mathbb{H}$ may encode the same conditional independence structure and this induces a partition of $\mathbb{H}$ into equivalence classes each characterised by a *largest chain graph* (Frydenberg, 1990; Studený, 1997; Volf and Studený, 1999 and Roverato, 2005). The search space can thus be taken to be the set of all largest CGs on p vertices. However, it is common practice to perform structural learning on predefined subsets of $\mathbb{H}$ identified on the basis of *substantive* background knowledge on the problem. From a Bayesian perspective this amounts to giving zero prior probability to a subset of structures on the basis of subject matter knowledge.

The translation of the available background knowledge into a corresponding restriction of the search space represents a crucial step of any structural learning procedure. Indeed, it is important to exploit any background knowledge that may lead to a dimensionality reduction of the search space both for computational reasons and because, otherwise, the selected structure might be incompatible with expert knowledge on the problem. On the other hand, an erroneous interpretation of the existing background knowledge may lead to the exclusion from the search space of the "true" structure (or more realistically of a subset of relevant structures) thereby making any subsequent learning algorithm ineffective.

The two subsets of CGs most frequently recurring in the literature are the family of CGs with no undirected edges, called *directed acyclic graphs* (DAGs), and the family of CGs with no arrows, called *undirected graphs* (UGs); see, among others, Madigan *et al.* (1996), Cowell *et al.* (1999, Chapter 11), Giudici and Green (1999), Chickering (2002), Roverato (2002) and Drton and Perlman (2004). We recall that the family of all DAGs on p vertices can be partitioned into equivalence classes characterised by the so-called *essential graphs* (Andersson *et al.*, 1997a and Studený, 2004).

The restriction of the search space to either the family of essential graphs or the family of UGs may be justified by background knowledge on the *kind of association* between variables. On the other hand, the introduction of CG models with both directed and undirected edges was motivated by substantive background knowledge on the existence of a block ordering of variables (Wermuth and Lauritzen, 1990). Lauritzen and Richardson (2002) provided the following general instances of this framework:

(a) variables may be divided into *risk factors*, *diseases* and *symptoms*;

(b) in a longitudinal study variables may be grouped according to time;

(c) in a cross-sectional study, variables may be divided into purely explanatory variables, intermediate variables and responses (Cox and Wermuth, 1996).

Traditionally, when such a block ordering is available, structural learning is performed within the family of CGs whose edges joining vertices in the same block are undirected while edges joining vertices in different blocks are directed according to the ordering (Wermuth and Lauritzen, 1990; Whittaker, 1990; Cox and Wermuth, 1996). This is motivated by a "causal" interpretation of relationships between variables belonging to different blocks, while variables belonging to a same block are considered to be "on an equal footing"; see, for instance, Mohamed *et al.* (1998), Pigeot *et al.* (2000) and Caputo *et al.* (2003). However, as noticed by Cox and Wermuth (1993, 2000) and Lauritzen and Richardson (2002), undirected edges represent a very special kind of equal footing. In fact, as far as background knowledge is concerned, two variables joined by an arrow are on an equal footing whenever the direction of the arrow is not known in advance. Our standpoint is that the existence of a block ordering of variables is not sufficient to imply that edges within blocks are undirected, and that such a further restriction of the search space is justified only when additional knowledge on the kind of association allowed within blocks is available. Hence, we make an explicit distinction between prior knowledge on variable ordering and on the kind of association relating variables within blocks, and introduce the concept of *labelled block ordering of vertices*, denoted by $\mathcal{B}$, as a formal way to encode background knowledge of both types. Every specification of $\mathcal{B}$ is associated with a family of $\mathcal{B}$-consistent CGs, denoted by $\mathbb{H}(\mathcal{B})$, made up of all structures in $\mathbb{H}$ compatible with the background knowledge encoded by $\mathcal{B}$. We obtain in this way a flexible tool for specifying subsets of $\mathbb{H}$ and, in Section 3, we provide a list of relevant subsets of CGs considered in the literature that are families of $\mathcal{B}$-consistent CGs for the appropriate choice of $\mathcal{B}$. This includes the family of UGs, the family of DAGs and every family of "traditional" CGs with undirected edges within blocks. Structural learning within the set $\mathbb{H}(\mathcal{B})$ requires to deal with Markov equivalence of structures compatible with the available block ordering. We provide a graphical characterisation of equivalence classes of $\mathcal{B}$-consistent CGs, namely the $\mathcal{B}$-*essential graphs*, as well as a procedure for deriving the $\mathcal{B}$-essential graph associated with an equivalence class. This generalises several existing results on Markov equivalence and provides an unified view to the subject. In particular, $\mathcal{B}$-essential graphs coincide with largest CGs when $\mathbb{H}(\mathcal{B}) = \mathbb{H}$ and with essential graphs when $\mathbb{H}(\mathcal{B})$ is the family of all DAGs on p vertices.

The paper is organised as follows. Section 2 reviews the required graphical model theory. Section 3 introduces and motivates the family of $\mathcal{B}$-consistent CGs while Section 4 deals with the graphical characterisation of equivalence classes of $\mathcal{B}$-consistent CGs. Finally, Section 5 contains a brief discussion.

2 Preliminaries and Notation

Here we review *graph theory* and *Markov equivalence* as required in this paper. We introduce the notation we use as well as a few relevant concepts, but we omit the definitions of well established concepts such as *connected component* of a graph, *decomposable* undirected graph, *parent*, *path*, *skeleton* and *subgraph*; we refer to Cowell *et al.* (1999) for a full account of the theory of graphs and graphical models.

2.1 Graph theory

We denote an arbitrary graph by $\mathcal{G} = (V, E)$, where V is a finite set of *vertices* and $E \subseteq V \times V$ is a set of *edges*. We say that α and γ are joined by an *arrow* pointing at γ, and write $\alpha \rightarrow \gamma \in \mathcal{G}$, if $(\alpha, \gamma) \in E$ but $(\gamma, \alpha) \notin E$. We write $\alpha \,—\, \gamma \in \mathcal{G}$ if both $(\alpha, \gamma) \in E$ and $(\gamma, \alpha) \in E$ and say that there is an undirected edge, or *line*, between α and γ. For a subset $A \subseteq V$ we denote by $\mathcal{G}_A$ the *subgraph* induced by A and by $\text{pa}_{\mathcal{G}}(A)$ the parents of A in $\mathcal{G}$, or simply by $\text{pa}(A)$ when it is clear from the context which graph is being considered. A *semi-directed cycle* of length r is a sequence $\alpha_0, \alpha_1, \ldots, \alpha_r = \alpha_0$ of r different vertices such that either $\alpha_{i-1} \,—\, \alpha_i$ or $\alpha_{i-1} \rightarrow \alpha_i$ for all $i = 1, \ldots, r$ and $\alpha_{i-1} \rightarrow \alpha_i$ for at least one value of i.

A *chain graph* (CG) is a graph with both arrows and lines but no semi-directed cycles and we denote by $\mathbb{H}_p \equiv \mathbb{H}$ the set of all CGs on $|V| = p$ vertices. A CG with no arrows is an *undirected graph* (UG) whereas a CG with no lines is a *directed acyclic graph* (DAG). Hereafter, to stress that a graph is a DAG, we denote it by $\mathcal{D} = (V, E)$. For a pair of vertices $\alpha, \gamma \in V$ of $\mathcal{G} = (V, E)$, we write $\alpha \rightleftharpoons \gamma$ if either $\alpha = \gamma$ or there is an undirected path between α and γ. It is straightforward to see that $\rightleftharpoons$ is an equivalence relation that induces a partition of the vertex set V into equivalence classes called the *chain components* of $\mathcal{G}$. If $T \subseteq V$ is a chain component of $\mathcal{G}$ then $\mathcal{G}_T$ is and undirected graph and when $\mathcal{G}_T$ is decomposable we say that T is a decomposable chain component. Furthermore, by saying that a CG has decomposable chain components we mean that *all* its chain components are decomposable.

A triple (α, δ, γ) is an *immorality* of $\mathcal{G}$ if the subgraph of $\mathcal{G}$ induced by $\{\alpha, \delta, \gamma\}$ is $\alpha \rightarrow \delta \leftarrow \gamma$. A sequence of vertices $(\alpha, \delta_1, \ldots, \delta_k, \gamma)$ is a *minimal complex* in $\mathcal{G}$ if the subgraph induced by $\{\alpha, \delta_1, \ldots, \delta_k, \gamma\}$ looks like

$$\alpha \rightarrow \delta_1 \,—\, \delta_2 \,—\, \cdots \,—\, \delta_{k-1} \,—\, \delta_k \leftarrow \gamma \tag{1}$$

so that a minimal complex with $k = 1$ is an immorality.

Definition 1 Two CGs are called *complex-equivalent* if they have the same skeleton and the same minimal complexes. □

A triple (α, δ, γ) is a *flag* of $\mathcal{G}$ if the subgraph of $\mathcal{G}$ induced by $\{\alpha, \delta, \gamma\}$ is either $\alpha \rightarrow \delta \,—\, \gamma$ or $\alpha \,—\, \delta \leftarrow \gamma$. A *no flag chain graph* (NF-CG) is a CG containing no flags; note that both

UGs and DAGs are NF-CGs. Every minimal complex with $k > 1$ is associated with two flags: $\alpha \to \delta_1 — \delta_2$ and $\delta_{k-1} — \delta_k \leftarrow \gamma$. Therefore, an NF-CG has no minimal complexes other than immoralities and, consequently, two NF-CGs are complex-equivalent if and only if they have the same skeleton and the same immoralities.

For two graphs $\mathcal{G}_1 = (V, E_1)$ and $\mathcal{G}_2 = (V, E_2)$ with the same skeleton, we say that $\mathcal{G}_1$ is *larger* than $\mathcal{G}_2$, denoted by $\mathcal{G}_2 \subseteq \mathcal{G}_1$, if $E_2 \subseteq E_1$, i.e. $\mathcal{G}_1$ may have undirected edges where $\mathcal{G}_2$ has arrows. We write $\mathcal{G}_2 \subset \mathcal{G}_1$ if $E_2 \subset E_1$. The following result (Roverato, 2005, Proposition 1) is useful to check complex-equivalence of nested CGs.

Proposition 1 *Let $\mathcal{G} = (V, E)$, $\mathcal{G}' = (V, E')$ and $\mathcal{G}'' = (V, E'')$ be three CGs with the same skeleton. If $\mathcal{G} \subseteq \mathcal{G}' \subseteq \mathcal{G}''$ and $\mathcal{G}$ is complex-equivalent to $\mathcal{G}''$, then $\mathcal{G}'$ is complex-equivalent to $\mathcal{G}$ and $\mathcal{G}''$.*

The union of two CGs $\mathcal{G}_1 = (V, E_1)$ and $\mathcal{G}_2 = (V, E_2)$ is the smallest graph larger than $\mathcal{G}_1$ and $\mathcal{G}_2$; formally $\mathcal{G}_1 \cup \mathcal{G}_2 = (V, E_1 \cup E_2)$. It is clear that $\mathcal{G}_1 \cup \mathcal{G}_2$ may not be a CG and, following Frydenberg (1990, p. 347), we denote by $\mathcal{G}_1 \vee \mathcal{G}_2$ the smallest CG larger then $\mathcal{G}_1$ and $\mathcal{G}_2$, that is the CG obtained by changing into undirected edges all the arrows in $\mathcal{G}_1 \cup \mathcal{G}_2$ which are part of a semi-directed cycle (see also Consequence 2.5 of Volf and Studený, 1999). Frydenberg (1990, Proposition 5.4) proved the following result.

Theorem 2 *Two CGs $\mathcal{G}_1 = (V, E_1)$ and $\mathcal{G}_2 = (V, E_2)$ with the same skeleton are complex-equivalent if and only if they are both complex-equivalent to $\mathcal{G}_1 \vee \mathcal{G}_2$.*

For a family $\mathbb{K}$ of CGs with common vertex set V we denote by $\cup\mathbb{K} = \{\cup\mathcal{G} \mid \mathcal{G} \in \mathbb{K}\}$ the smallest graph larger than every element of $\mathbb{K}$ and by $\vee\mathbb{K} = \{\vee\mathcal{G} \mid \mathcal{G} \in \mathbb{K}\}$ the smallest CG larger than every element of $\mathbb{K}$. There are nontrivial families of CGs closed with respect to the $\vee$-union operation, as the following result shows.

Proposition 3 *If $\mathcal{G}_1 = (V, E_1)$ and $\mathcal{G}_2 = (V, E_2)$ are two equivalent NF-CGs then $\mathcal{G}_1 \vee \mathcal{G}_2$ is an NF-CG. Moreover, if $\mathcal{G}_1$ and $\mathcal{G}_2$ have decomposable chain components then $\mathcal{G}_1 \vee \mathcal{G}_2$ has decomposable chain components.*

Proof. The CG $\mathcal{G}_1 \vee \mathcal{G}_2$ is an NF-CG by Theorem 7 of Roverato (2005). Assume now that $\mathcal{G}_1$ and $\mathcal{G}_2$ have decomposable chain components. By Lemma 2 of Studený (2004) it holds that $\mathcal{G}_1$ and $\mathcal{G}_2$ are equivalent to some DAG $\mathcal{D}$. Then, it follows from Theorem 2 that $\mathcal{D}$ is also equivalent to $\mathcal{G}_1 \vee \mathcal{G}_2$ which therefore has decomposable chain components by Lemma 2 of Studený (2004). □

2.2 Markov equivalence

Let $X_V = (X_\alpha)_{\alpha \in V}$ be a collection of random variables taking values in the sample space $\mathcal{X} = \times_{\alpha \in V} \mathcal{X}_\alpha$. The spaces $\mathcal{X}_\alpha$, $\alpha \in V$, are supposed to be separable metric spaces endowed with Borel σ-algebras so that the existence of regular conditional probabilities is ensured.

A graphical Markov model uses a graph with vertex set V to specify a set of conditional independence relations, called a Markov property, among the components of X. Denote by $\mathcal{M}(\mathcal{G}, \mathcal{X})$ the set of probability distributions on $\mathcal{X}$ that satisfy the conditional independence relations associated with $\mathcal{G}$ by the LWF Markov property (Lauritzen and Wermuth, 1989 and Frydenberg, 1990). Two graphs $\mathcal{G}_1 = (V, E_1)$ and $\mathcal{G}_2 = (V, E_2)$ are said to be *Markov equivalent*, denoted by $\mathcal{G}_1 \sim \mathcal{G}_2$, if $\mathcal{M}(\mathcal{G}_1, \mathcal{X}) = \mathcal{M}(\mathcal{G}_2, \mathcal{X})$ for every product space $\mathcal{X}$ indexed by V. Markov equivalence is an equivalence relation and for a CG $\mathcal{G} = (V, E)$ we denote by $[\![\mathcal{G}]\!]$ the class of all CGs equivalent to $\mathcal{G}$. Similarly, for a DAG $\mathcal{D} = (V, E)$, we denote by $[\mathcal{D}]$ the class of all DAGs Markov equivalent to $\mathcal{D}$. Note that $[\mathcal{D}] \subseteq [\![\mathcal{D}]\!]$.

The skeleton and the minimal complexes of a CG $\mathcal{G} = (V, E)$ are sufficient to determine the set of conditional independencies encoded by $\mathcal{G}$ (Frydenberg, 1990, Theorem 5.6; Andersson *et al.*, 1997b, Theorem 3.1 and Verma and Pearl, 1991).

Theorem 4 *Two CGs are Markov equivalent if and only if they are complex-equivalent.*

As a consequence of Theorem 4, having adopted the LWF-Markov property, no formal distinction between Markov and complex equivalence is necessary and in the rest of this paper we shall simply say that two CGs are equivalent.

It is well known that structural learning procedures dealing with the space of CGs instead of the space of equivalence classes may face several problems concerning computational efficiency and the specification of prior distributions; see Chickering (2002) for a discussion on the difficulties deriving from Markov equivalence of DAGs. For this reason it is of interest to characterise every equivalence class by means of a single graph providing a suitable representation of the whole class. The class $[\![\mathcal{G}]\!]$ of all CGs equivalent to a CG $\mathcal{G}$ is naturally represented by means of the *largest chain graph* $\mathcal{G}^{\ddagger} = \vee[\![\mathcal{G}]\!] = \cup[\![\mathcal{G}]\!]$ (Frydenberg, 1990; Studený, 1996, 1997; Volf and Studený, 1999; Roverato, 2005). Similarly, the equivalence class $[\mathcal{D}]$ of a DAG $\mathcal{D}$ is naturally represented by the so-called *essential graph* $\mathcal{D}^* = \cup[\mathcal{D}]$, which Andersson *et al.* (1997a) showed to be a CG, that is $\mathcal{D}^* = \vee[\mathcal{D}]$. We recall that the essential graph is also known in the literature as completed pattern (Verma and Pearl, 1992), maximally oriented path for a pattern (Meek, 1995) and completed p-dag (Chickering, 2002).

The notion of *meta-arrow* plays an important role in the characterisation of equivalence classes of CGs (2004; Studený, 2004; Roverato and Studený; Roverato, 2005).

Definition 2 Let A and D be two chain components of a CG $\mathcal{G}$. The *meta-arrow* $A \rightrightarrows D$ is the set of all arrows in $\mathcal{G}$ pointing from A to D, that is formally $A \rightrightarrows D = \{\alpha \to \delta \in \mathcal{G} \mid \alpha \in A, \delta \in D\}$. □

The chain components of a nonempty meta-arrow can be merged to form a single chain component as follows.

Definition 3 Let $\mathcal{G} = (V, E)$ be a CG and $A \rightrightarrows D$ one of its meta-arrows. The graph obtained from $\mathcal{G}$ by *merging* (the chain components of) $A \rightrightarrows D$ is the graph obtained from $\mathcal{G}$ by replacing every arrow $\alpha \to \delta \in A \rightrightarrows D$ with the corresponding line $\alpha - \delta$. □

Interest in (merging) meta-arrows is due to the fact that if $\mathcal{G} \subseteq \mathcal{G}'$ are two CGs with the same skeleton and $A \rightrightarrows D$ is a meta-arrow of $\mathcal{G}$, then necessarily either *every* arrow or *no* arrow of $A \rightrightarrows D$ corresponds to a line in $\mathcal{G}'$. Indeed, if only some, but not all, the arrows in $A \rightrightarrows D$ were lines in $\mathcal{G}'$, then there would be a semi-directed cycle in $\mathcal{G}'$.

Roverato (2005) proposed a procedure for the construction of both largest CGs and essential graphs based on successive merging of meta-arrows. The existence of such a procedure is guaranteed by the following result (Roverato, 2005, Theorems 7 and 10).

Theorem 5 *Let $\mathcal{G} = (V, E)$ and $\mathcal{G}' = (V, E')$ be two equivalent CGs such that $\mathcal{G} \subset \mathcal{G}'$. Then, there exists a finite sequence $\mathcal{G} = \mathcal{G}_0 \subset \mathcal{G}_1 \subset \cdots \subset \mathcal{G}_r = \mathcal{G}'$, with $r \geq 1$, of equivalent CGs such that $\mathcal{G}_i$ can be obtained from $\mathcal{G}_{i-1}$ by merging exactly one of its meta-arrows, for all $i = 1, \ldots, r$. Furthermore, if $\mathcal{G}$ and $\mathcal{G}'$ are NF-CGs, it is possible to let $\mathcal{G}_1, \ldots, \mathcal{G}_r$ be NF-CGs.*

Note that this theorem can be applied to the construction of the essential graph $\mathcal{D}^*$ because it was shown by Studený (2004) that $\mathcal{D}^*$ is the largest NF-CG equivalent to $\mathcal{D}$. In order to practically take advantage of Theorem 5, it is necessary to characterise those meta-arrows of a (NF-)CG which, if merged, lead to an equivalent (NF-)CG. This requires the notion of (strongly) insubstantial arrowhead of a meta-arrow.

Definition 4 Let $A \rightrightarrows D$ be a meta-arrow of the CG $\mathcal{G} = (V, E)$. The arrowhead of $A \rightrightarrows D$ is *insubstantial* in $\mathcal{G}$ if the following conditions are satisfied:

(a) $\mathrm{pa}_{\mathcal{G}}(D) \cap A$ is complete;

(b) $\mathrm{pa}_{\mathcal{G}}(D) \setminus A \subseteq \mathrm{pa}_{\mathcal{G}}(\alpha)$, for all $\alpha \in \mathrm{pa}_{\mathcal{G}}(D) \cap A$.

Furthermore, if it also holds that

(c) $\mathrm{pa}_{\mathcal{G}}(D) \setminus A = \mathrm{pa}_{\mathcal{G}}(\alpha)$, for all $\alpha \in \mathrm{pa}_{\mathcal{G}}(D) \cap A$,

then the arrowhead of $A \rightrightarrows D$ is said to be *strongly* insubstantial. □

Roverato (2005, Theorems 8 and 11) provided a connection between Definition 4 and the operation of merging meta-arrows.

Theorem 6 *Let $A \rightrightarrows D$ be a meta-arrow of the CG $\mathcal{G} = (V, E)$. The CG $\mathcal{G}'$ obtained from $\mathcal{G}$ by merging the meta-arrow $A \rightrightarrows D$ is a CG equivalent to $\mathcal{G}$ if and only if the arrowhead of $A \rightrightarrows D$ is insubstantial in $\mathcal{G}$. Furthermore, if $\mathcal{G}$ is an NF-CG, then $\mathcal{G}'$ is an NF-CG if and only if the arrowhead of $A \rightrightarrows D$ is strongly insubstantial in $\mathcal{G}$.*

The results given in this section can be used to implement an efficient procedure for the construction of both largest CGs and essential graphs. Moreover, they were used by Roverato (2005) and Studený (2004) to provide a graphical characterisation of both largest CGs and essential graphs.

3 $\mathcal{B}$-consistent chain graphs

In this section we introduce the concept of *labelled block ordering* $\mathcal{B}$ of vertices and describe how $\mathcal{B}$ can be used to specify the appropriate subset $\mathbb{H}(\mathcal{B})$ of $\mathcal{B}$-consistent CGs. To clarify the meaning of $\mathcal{B}$ and eventually to reach a definition we proceed in steps.

Consider first the situation in which it is assumed that the independence structure of X_V is encoded by a CG but no further information is available, neither on the existence of a block ordering of variables nor on the kind of association between variables. In this case we write $\mathcal{B} = (V)$ and the appropriate search space for structural learning is the family of all largest CGs on p vertices. Assume now that, although no ordering of the variables can be specified, it is believed that the conditional independence structure of X_V is encoded by a DAG. Then the appropriate search space is the family of all essential graphs on p vertices. Another common situation is when the association between every pair of variables is believed to be symmetric. In this case structural learning is restricted to the subset of all UGs on p vertices; see Cox and Wermuth (2000) and Lauritzen and Richardson (2002) for a discussion on the interpretation of undirected edges. We write $\mathcal{B} = (V^d)$ to denote that background knowledge specifies no ordering of variables but indicates that the conditional independence structure of X_V is encoded by a DAG and, similarly, we write $\mathcal{B} = (V^u)$ for the case of UGs. A consistent notation for the unrestricted case is $\mathcal{B} = (V^g) \equiv (V)$.

We remark that (V^d), (V^u) and (V^g) represent three different types of background knowledge that lead to three different restrictions of the search space. Nevertheless, in all three cases, variables are regarded to be on an equal footing because every association is either asymmetric with unknown direction or symmetric. In the following we will refer to this situation as to the *one-block* case.

Consider now the situation in which background knowledge indicates that the set V is partitioned into ordered blocks $V_1, \ldots, V_k$ such that edges between vertices in different blocks are arrows pointing from blocks with lower numbering to blocks with higher numbering. If no further information is available about the relationship between variables in a same block, then we argue that the available background knowledge is compatible with all CGs consistent with the block ordering and such that the subgraphs $\mathcal{G}_{V_i}$, $i = 1, \ldots, k$, are themselves CGs. In our notation this is written as $\mathcal{B} = (V_1^g, \ldots, V_k^g)$ or, equivalently, as $\mathcal{B} = (V_1, \ldots, V_k)$. An instance of this situation is given in the following example adapted from Lauritzen and Richardson (2002).

Example 1 Let T represent a randomised treatment so that, for instance, $T = 1$ denotes that a patient has been treated with a given drug and $T = 0$ indicates placebo. Furthermore, let X and Y represent two responses of the experiment. Since the treatment is randomised, edges between $V_1 = \{T\}$ and $V_2 = \{X, Y\}$ correspond to causal effects, namely to arrows pointing from T to variables in V_2. It follows that, if the kind of relationship between X and Y is unknown, then background knowledge implies the block structure in Figure 1 (a) but,

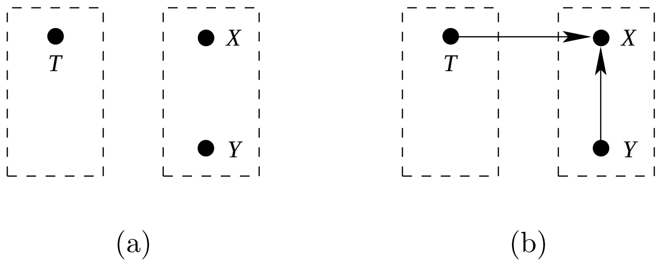

Figure 1: Block ordering of variables does not imply undirected edges within blocks. (a) block ordering of variables. (b) instance of CG compatible with the ordering.

at the same time, it is also compatible with the model given in Figure 1 (b). Nevertheless, a traditional CG modelling procedure would rule out the presence of an arrow between X and Y and, as a consequence, model (b) in Figure 1 from the search space. □

As well as for the one-block case, also in the multiple-bock case additional subject matter knowledge on the kind of association within blocks may indicate that for a given block V_i either only arrows or only lines are allowed. We incorporate this information in $\mathcal{B}$ by adding a label to V_i. In this way $\mathcal{B}$ specifies a block structure where every block may be an UG, a DAG or, more generally, a CG.

Definition 5 Let $V_1, \ldots, V_k$ be a partition of a set of vertices V. A *labelled block ordering* $\mathcal{B}$ of V is a sequence $(V_i^{\ell_i}, i = 1, \ldots, k)$ such that $\ell_i \in \{u, e, g\}$ and with the convention that $V_i = V_i^g$. □

Every labelled block ordering $\mathcal{B}$ of V identifies a subset of $\mathbb{H}$ made up of all CGs that satisfy the constraints imposed by $\mathcal{B}$.

Definition 6 Let $\mathcal{B} = (V_i^{\ell_i}, i = 1, \ldots, k)$ be a labelled block ordering of the vertex set V. We say that the CG $\mathcal{G} = (V, E)$ is $\mathcal{B}$-*consistent* if

(a) all edges joining vertices in different blocks of $\mathcal{B}$ are arrows pointing from blocks with lower numbering to blocks with higher numbering;

(b) for all i such that $\ell_i = u$, the subgraph $\mathcal{G}_{V_i}$ is an UG;

(c) for all i such that $\ell_i = d$, the subgraph $\mathcal{G}_{V_i}$ is a DAG.

□

Hereafter we use the shorthand $\mathcal{B}$-*CG* for "$\mathcal{B}$-consistent CG" and denote by $\mathbb{H}(\mathcal{B})$ the family of all $\mathcal{B}$-CGs on $|V|$ vertices.

Example 1 (continued). In this example no restriction can be imposed on the kind of association within blocks. Hence, we represent the existing background knowledge by the labelled block ordering $\mathcal{B} = (V_1^g, V_2^g)$ and, consequently, the space $\mathbb{H}(\mathcal{B})$ is made up of all structures in $\mathbb{H}$ such that every edge involving T is an arrow pointing to either X or Y. On the contrary, traditional CG modelling corresponds to the labelled block ordering $\mathcal{B}' = (V_1^u, V_2^u)$. It is easy to see that $\mathbb{H}(\mathcal{B}') \subset \mathbb{H}(\mathcal{B})$; in particular, the graph (b) of Figure 1 is contained in $\mathbb{H}(\mathcal{B})$ but not in $\mathbb{H}(\mathcal{B}')$. □

It is worth noting that $\mathcal{B}$-CGs are modular objects. They can be thought of as DAGs of boxes where every box is either an UG or a DAG, or more generally a CG. We remark that $\mathcal{B}$ cannot be used to encode any arbitrary background knowledge on the network structure; nevertheless, it is a versatile tool and, to our knowledge, every relevant subclass of CGs considered in the literature is a class of $\mathcal{B}$-CGs for the appropriate choice of $\mathcal{B}$. As shown above, the families of DAGs, UGs and CGs are all families of $\mathcal{B}$-CGs. When every block of $\mathcal{B}$ is made up of exactly one variable, then the class of $\mathcal{B}$-CGs is the class of all DAGs consistent with a given full ordering of variables. The family of CGs traditionally used in the literature is the class of $\mathcal{B}$-CGs with $\mathcal{B} = (V_i^u, i = 1, \ldots, k)$. If $\mathcal{B} = (V_1^u, V_2^d)$ then $\mathbb{H}(\mathcal{B})$ is a family of *recursive causal graphs* (Kiiveri *et al.*, 1984 and Richardson, 2001). Of special interest is also the family of $\mathcal{B}$-CGs with $\mathcal{B} = (V_i^d, i = 1, \ldots, k)$, namely the subset of all DAGs consistent with a given block ordering of variables. Structural learning within this family of graphs is implemented in the TETRAD program (Spirtes *et al.*, 2001 and Meek, 1995). Furthermore, the latter family of graphs is also implicitly used in the context of probabilistic expert systems involving both continuous and discrete variables where, for computational convenience, vertices are partitioned as $\mathcal{B} = (V_1^d, V_2^d)$ with V_1^d corresponding to the discrete variables and V_2^d to the continuous variables (see Cowell *et al.*, 1999, p. 136).

The implementation of structural learning algorithms with respect to the family $\mathbb{H}(\mathcal{B})$ requires to deal with Markov equivalence of $\mathcal{B}$-consistent CGs. In the next section we consider the characterisation of equivalence classes of $\mathcal{B}$-CGs. This requires the extension of the family of $\mathcal{B}$-CGs to the family of *weakly $\mathcal{B}$-consistent CGs*, shortly *w$\mathcal{B}$-CGs*, obtained by relaxing condition (c) of Definition 6 as follows:

(c′) for all i such that $\ell_i = d$, the subgraph $\mathcal{G}_{V_i}$ is a CG with decomposable chain components and there is no flag $\alpha \rightarrow \gamma - \delta \in \mathcal{G}$ such that $\gamma - \delta \in \mathcal{G}_{V_i}$.

We denote by $\mathbb{H}_+(\mathcal{B})$ the family of all $w\mathcal{B}$-CGs and remark that $\mathbb{H}(\mathcal{B}) \subseteq \mathbb{H}_+(\mathcal{B})$ because condition (c) of Definition 6 implies condition (c′) above. Note that, if $\mathcal{B} = (V^d)$, then $\mathbb{H}_+(\mathcal{B})$ is the family of NF-CGs with decomposable chain components.

4 Graphical characterisation of $\mathcal{B}$-CGs

In the graphical characterisation of equivalence classes of CGs one single graph replaces a whole class of CGs and, as a consequence, the amount of information carried by the characterising graph constitutes a central issue in the choice of the representative. For instance, in the characterisation of the class $[\![\mathcal{G}]\!]$ of all CGs equivalent to the CG $\mathcal{G} = (V, E)$, the largest CG $\mathcal{G}^{\ddagger} = \cup[\![\mathcal{G}]\!]$ constitutes a natural representative because:

(i) $\mathcal{G}^{\ddagger}$ is a CG equivalent to $\mathcal{G}$;

(ii) two vertices are joined by an arrow in $\mathcal{G}^{\ddagger}$ if and only if they are joined by an arrow with the same direction in every $\mathcal{G}' \in [\![\mathcal{G}]\!]$;

(iii) every $\mathcal{G}' \in [\![\mathcal{G}]\!]$ can be recovered from $\mathcal{G}^{\ddagger}$ by properly directing certain lines.

We remark that properties (ii) and (iii) are a consequence of the fact that $\mathcal{G}^{\ddagger}$ is the union of all CGs in $[\![\mathcal{G}]\!]$ and that property (ii) is especially important in causal discovery because it is used to identify the associations between variables that might have a causal interpretation (Spirtes *et al.*, 2001).

In the characterisation of the class $[\mathcal{D}]$ of all DAGs equivalent to the DAG $\mathcal{D}$ both the largest CG $\mathcal{D}^{\ddagger}$ and the essential graph $\mathcal{D}^{*} = \cup[\mathcal{D}]$ could be used, but the latter representative is preferred because it satisfies all three properties above (suitably rephrased by replacing $\mathcal{G}$ with $\mathcal{D}$, $\mathcal{G}^{\ddagger}$ with $\mathcal{D}^{*}$ and $[\![\mathcal{G}]\!]$ with $[\mathcal{D}]$). More precisely, properties (i) and (iii) are satisfied by both $\mathcal{D}^{\ddagger}$ and $\mathcal{D}^{*}$ whereas property (ii) is only satisfied by $\mathcal{D}^{*}$ because it is the smallest graph larger than every element of $[\mathcal{D}]$.

Let $\mathcal{G} \in \mathbb{H}(\mathcal{B})$ be a $\mathcal{B}$-CG. It seems natural, at this point, to represent the *$\mathcal{B}$-equivalence class* $[\mathcal{G}]^{\mathcal{B}} = \{\mathcal{G}' \in \mathbb{H}(\mathcal{B}) \mid \mathcal{G}' \sim \mathcal{G}\}$ of $\mathcal{G}$ by means of the smallest graph larger than every element of $[\mathcal{G}]^{\mathcal{B}}$, which we call the *$\mathcal{B}$-essential graph* of $\mathcal{G}$.

Definition 7 For a $\mathcal{B}$-CG $\mathcal{G}$, the *$\mathcal{B}$-essential graph* of $\mathcal{G}$ is the graph $\mathcal{G}^{\mathcal{B}} = \cup[\mathcal{G}]^{\mathcal{B}}$. □

The $\mathcal{B}$-essential graph is constructed so as to satisfy properties (ii) and (iii) above (suitably rephrased by replacing $\mathcal{G}^{\ddagger}$ with $\mathcal{G}^{\mathcal{B}}$ and $[\![\mathcal{G}]\!]$ with $[\mathcal{G}]^{\mathcal{B}}$) and in this section we show that property (i) also holds, i.e. that $\mathcal{G}^{\mathcal{B}}$ is a CG equivalent to $\mathcal{G}$. Furthermore, we provide a graphical characterisation of $\mathcal{G}^{\mathcal{B}}$ and a procedure to construct $\mathcal{B}$-essential graphs.

We remark that in general $\mathcal{G}^{\mathcal{B}} \notin [\mathcal{G}]^{\mathcal{B}}$, as well as $\mathcal{D}^{*} \notin [\mathcal{D}]$. Indeed, if $\mathcal{B} = (V^{d})$ then $[\mathcal{G}]^{\mathcal{B}} = [\mathcal{G}]$ and $\mathcal{G}^{\mathcal{B}} = \mathcal{G}^{*}$, while for $\mathcal{B} = (V^{g})$ it holds that $[\mathcal{G}]^{\mathcal{B}} = [\![\mathcal{G}]\!]$ and $\mathcal{G}^{\mathcal{B}} = \mathcal{G}^{\ddagger}$. Thus, the material in this section unifies and generalises some existing results for essential graphs and largest CGs.

4.1 $\mathcal{G}^{\mathcal{B}}$ is a CG equivalent to $\mathcal{G}$

Frydenberg (1990) showed that the $\vee$-union of CGs equivalent to a given CG $\mathcal{G}$ is a CG equivalent to $\mathcal{G}$ (see Theorem 2). Here we prove that $\mathcal{G}^{\mathcal{B}}$ is a CG equivalent to $\mathcal{G}$ by showing that it can be obtained as the $\vee$-union of CGs equivalent to $\mathcal{G}$.

It is easy to check that for two equivalent CGs $\mathcal{G}', \mathcal{G}'' \in [\mathcal{G}]^{\mathcal{B}}$ the $\vee$-union $\mathcal{G}' \vee \mathcal{G}''$ is not necessarily a $\mathcal{B}$-CG and thus may not belong to $[\mathcal{G}]^{\mathcal{B}}$. For this reason it is convenient to deal with the wider class of $w\mathcal{B}$-CGs introduced in Section 3. For a $w\mathcal{B}$-CG $\mathcal{G} \in \mathbb{H}_{+}(\mathcal{B})$ let

$[\mathcal{G}]^{\mathcal{B}}_{+} = \{\mathcal{G}' \in \mathbb{H}_{+}(\mathcal{B}) \mid \mathcal{G}' \sim \mathcal{G}\}$ be the corresponing *w$\mathcal{B}$-equivalence class* and let $\mathcal{G}^{\mathcal{B}}_{+} = \cup[\mathcal{G}]^{\mathcal{B}}_{+}$, so that if $\mathcal{G} \in \mathbb{H}(\mathcal{B})$ then $[\mathcal{G}]^{\mathcal{B}} \subseteq [\mathcal{G}]^{\mathcal{B}}_{+}$ and $\mathcal{G}^{\mathcal{B}} \subseteq \mathcal{G}^{\mathcal{B}}_{+}$. We now show that the class $[\mathcal{G}]^{\mathcal{B}}_{+}$ is closed with respect to the $\vee$-union operation.

Lemma 7 *For a labelled block ordering $\mathcal{B} = (V_i^{\ell_i}, i = 1, \dots, k)$ of a vertex set V, let $\mathcal{G} = (V, E)$ and $\mathcal{G}' = (V, E')$ be two w$\mathcal{B}$-CGs. Then*

(i) if $\alpha_0, \alpha_1, \dots, \alpha_r = \alpha_0$ is a semi-directed cycle in $\mathcal{G} \cup \mathcal{G}'$, then there exists a block V_i of $\mathcal{B}$ such that $\{\alpha_0, \alpha_1, \dots, \alpha_r\} \subseteq V_i$;

(ii) if the subset $U \subseteq V$ is the union of some blocks of $\mathcal{B}$, that is either $V_i \cap U = \emptyset$ or $V_i \cap U = V_i$ for all $i = 1, \dots, k$, then it holds that $(\mathcal{G} \vee \mathcal{G}')_U = \mathcal{G}_U \vee \mathcal{G}'_U$.

Proof. (i). For a vertex $\alpha \in V$ let $i(\alpha)$ denote the index of the block V_i of $\mathcal{B}$ such that $\alpha \in V_i$. Hence, to the semi-directed cycle $\alpha_0, \alpha_1, \dots, \alpha_r = \alpha_0$ it is associated the sequence $i(\alpha_0), i(\alpha_1), \dots, i(\alpha_r) = i(\alpha_0)$ and we prove point (i) by showing that this sequence is constant, i.e. that $i(\alpha_0) = i(\alpha_1) = \cdots = i(\alpha_r) = i$. We first note that $i(\alpha_{j-1}) \leq i(\alpha_j)$ for all $j = 1, \dots, r$. Indeed, as a consequence of the weak $\mathcal{B}$-consistency of $\mathcal{G}$ and $\mathcal{G}'$, $i(\alpha_{j-1}) > i(\alpha_j)$ would imply both $\alpha_{j-1} \leftarrow \alpha_j \in \mathcal{G}$ and $\alpha_{j-1} \leftarrow \alpha_j \in \mathcal{G}'$ and, consequently, $\alpha_{j-1} \leftarrow \alpha_j \in \mathcal{G} \cup \mathcal{G}'$ would be an arrow pointing against the direction of the cycle, which is not possible. We conclude that the sequence is non-decreasing and, since $i(\alpha_r) = i(\alpha_0)$, it can only be constant. Point (ii) follows directly from point (i) and the definition of the $\vee$-union operation. □

We remark that, as a consequence of point (ii) of Lemma 7, the $\vee$-union operation on w$\mathcal{B}$-CGs is local with respect to blocks; formally $(\mathcal{G} \vee \mathcal{G}')_{V_i} = \mathcal{G}_{V_i} \vee \mathcal{G}'_{V_i}$ for all $i = 1, \dots, k$.

Theorem 8 *Let $\mathcal{B}$ be a labelled block ordering of a vertex set V. If $\mathcal{G} = (V, E)$ and $\mathcal{G}' = (V, E')$ are two equivalent w$\mathcal{B}$-CGs, then $\mathcal{G} \vee \mathcal{G}'$ is a w$\mathcal{B}$-CG equivalent to $\mathcal{G}$ and $\mathcal{G}'$.*

Proof. The graph $\mathcal{G} \vee \mathcal{G}'$ is a CG by construction and it is equivalent to $\mathcal{G}$ by Theorem 2 (Frydenberg, 1990). We now show that $\mathcal{G} \vee \mathcal{G}'$ satisfies conditions (a), (b) and (c′) which define w$\mathcal{B}$-CGs. First, condition (a) holds since every between-block arrow of $\mathcal{G}$ and $\mathcal{G}'$ is also an arrow in $\mathcal{G} \cup \mathcal{G}'$ and Lemma 7 (i) guarantees that no between-block arrow of $\mathcal{G} \cup \mathcal{G}'$ can become a line in $\mathcal{G} \vee \mathcal{G}'$ as part of a semi-directed cycle in $\mathcal{G} \cup \mathcal{G}'$. Condition (b) is trivially true because $(\mathcal{G} \vee \mathcal{G}')_{V_i} = \mathcal{G}_{V_i} = \mathcal{G}'_{V_i}$ is an UG for all i such that $\ell_i = u$.

In order to prove condition (c′) it is sufficient to show that, for all i such that $\ell_i = d$, $(\mathcal{G} \vee \mathcal{G}')_{V_i}$ is an NF-CG with decomposable chain components and that, if there exists an arrow $\alpha \to \gamma \in \mathcal{G} \vee \mathcal{G}'$ with $\gamma \in V_i$ and $\alpha \in V_j$, so that $j < i$, then $(\mathcal{G} \vee \mathcal{G}')_{V_i \cup \{\alpha\}}$ is an NF-CG. By Lemma 7 (ii) $(\mathcal{G} \vee \mathcal{G}')_{V_i} = \mathcal{G}_{V_i} \vee \mathcal{G}'_{V_i}$ and thus, since both $\mathcal{G}_{V_i}$ and $\mathcal{G}'_{V_i}$ are NF-CGs with decomposable chain components, such is $(\mathcal{G} \vee \mathcal{G}')_{V_i}$ by Proposition 3. Assume now that $\alpha \to \gamma \in \mathcal{G} \vee \mathcal{G}'$ with $\gamma \in V_i$ and $\alpha \in V_j$, so that $j < i$. It is easy to see that, letting

$\tilde{V} = V_i \cup \{\alpha\}$, both $\mathcal{G}_{\tilde{V}}$ and $\mathcal{G}'_{\tilde{V}}$ are NF-CGs and that, if we set $\tilde{\mathcal{B}} = (\{\alpha\}, V_i^d)$, then $\mathcal{G}_{\tilde{V}}$ and $\mathcal{G}'_{\tilde{V}}$ are $w\tilde{\mathcal{B}}$-CGs. Clearly, both in $\mathcal{G}_{\tilde{V}} \vee \mathcal{G}'_{\tilde{V}}$ and in $(\mathcal{G} \vee \mathcal{G}')_{\tilde{V}}$ every edge involving α is an arrow pointing at a vertex in V_i. Furthermore, by Lemma 7, both $(\mathcal{G}_{\tilde{V}} \vee \mathcal{G}'_{\tilde{V}})_{V_i} = \mathcal{G}_{V_i} \vee \mathcal{G}'_{V_i}$ and $(\mathcal{G} \vee \mathcal{G}')_{V_i} = \mathcal{G}_{V_i} \vee \mathcal{G}'_{V_i}$. We can conclude that $(\mathcal{G} \vee \mathcal{G}')_{V_i \cup \{\alpha\}} = \mathcal{G}_{\tilde{V}} \vee \mathcal{G}'_{\tilde{V}}$ and the result follows because, by Proposition 3, $\mathcal{G}_{\tilde{V}} \vee \mathcal{G}'_{\tilde{V}}$ is an NF-CG. □

It follows from Theorem 8 that, unlike the class $[\mathcal{G}]^{\mathcal{B}}$, the class $[\mathcal{G}]^{\mathcal{B}}_{+}$ is closed with respect to the $\vee$-union operation so that there exists a largest CG in $[\mathcal{G}]^{\mathcal{B}}_{+}$.

Corollary 9 *For a labelled block ordering $\mathcal{B}$ of a vertex set V, let $\mathcal{G} = (V, E)$ be a $\mathcal{B}$-CG. Then it holds that*

(i) $\vee[\mathcal{G}]^{\mathcal{B}}_{+} \in [\mathcal{G}]^{\mathcal{B}}_{+}$;

(ii) $\mathcal{G}^{\mathcal{B}}_{+} = \vee[\mathcal{G}]^{\mathcal{B}}_{+}$.

Proof. Point (i) is an immediate consequence of Theorem 8 whereas point (ii) follows from the fact that the class $[\mathcal{G}]^{\mathcal{B}}_{+}$ contains a largest CG. □

The reason for introducing the graph $\mathcal{G}^{\mathcal{B}}_{+}$ is that, in fact, it coincides with $\mathcal{G}^{\mathcal{B}}$ for all $\mathcal{G} \in \mathbb{H}(\mathcal{B})$.

Theorem 10 *For a labelled block ordering $\mathcal{B}$ of a vertex set V let $\mathcal{G} = (V, E)$ be a $\mathcal{B}$-CG. Then, it holds that $\mathcal{G}^{\mathcal{B}} = \mathcal{G}^{\mathcal{B}}_{+}$.*

Proof. In this proof we make use of the following facts.

Fact 1. If $\mathcal{H}$ is an NF-CG with decomposable chain components, then it is possible to direct the undirected edges of $\mathcal{H}$ so as to obtain a DAG equivalent to $\mathcal{H}$. For a proof see Roverato (2005, Appendix A).

Fact 2. If $\mathcal{H}$ is an NF-CG with decomposable chain components and $\alpha — \gamma \in \mathcal{H}$, then the procedure described in Fact 1 can be used to construct two DAGs $\mathcal{D}'$ and $\mathcal{D}''$ both equivalent to $\mathcal{H}$ and such that $\alpha \rightarrow \gamma \in \mathcal{D}'$ while $\alpha \leftarrow \gamma \in \mathcal{D}''$. For a proof see Roverato (2005, Appendix A).

Fact 3. For a $w\mathcal{B}$-CG $\tilde{\mathcal{G}}$ let $\tilde{\mathcal{G}}'$ be any graph obtained from $\tilde{\mathcal{G}}$ by replacing every subgraph $\tilde{\mathcal{G}}_{V_i}$ such that $\ell_i = d$ with a DAG $\mathcal{D}_{V_i}$ equivalent to $\tilde{\mathcal{G}}_{V_i}$ constructed as described in Fact 1. Then $\tilde{\mathcal{G}}'$ is a $\mathcal{B}$-CG equivalent to $\tilde{\mathcal{G}}$.

It is straightforward to see that $\tilde{\mathcal{G}}'$ is a $\mathcal{B}$-CG and we now show that it is equivalent to $\tilde{\mathcal{G}}$. By construction $\tilde{\mathcal{G}}'$ has the same skeleton as $\tilde{\mathcal{G}}$ and, moreover, $\tilde{\mathcal{G}}_{V_i}$ has the same minimal complexes as $\tilde{\mathcal{G}}'_{V_i}$ for all $i = 1, \ldots, k$. Hence, we only have to check that $\tilde{\mathcal{G}}$ and $\tilde{\mathcal{G}}'$ have the same minimal complexes involving between-block arrows. More precisely, differences between the two CGs can only involve complexes which are immoralities $\alpha \to \gamma \leftarrow \delta$ with $\gamma \in V_i$, $\ell_i = d$, and, without loss of generality, $\alpha \notin V_i$. Note that if an immorality of this type is present in $\tilde{\mathcal{G}}'$ then it must also be present in $\tilde{\mathcal{G}}$ because otherwise there would be a flag $\alpha \to \gamma - \delta \in \tilde{\mathcal{G}}$ with $\gamma - \delta \in V_i$. Hence, every immorality of this type in $\tilde{\mathcal{G}}'$ is an immorality in $\tilde{\mathcal{G}}$, but the converse is also true because $\tilde{\mathcal{G}}' \subseteq \tilde{\mathcal{G}}$, and we can conclude that $\tilde{\mathcal{G}}'$ and $\tilde{\mathcal{G}}$ are equivalent.

It is clear that $\mathcal{G}^{\mathcal{B}} \subseteq \mathcal{G}^{\mathcal{B}}_{+}$ so that to prove that $\mathcal{G}^{\mathcal{B}} = \mathcal{G}^{\mathcal{B}}_{+}$ it is sufficient to show that $\mathcal{G}^{\mathcal{B}}_{+} \subseteq \mathcal{G}^{\mathcal{B}}$, i.e. that $\gamma - \delta \in \mathcal{G}^{\mathcal{B}}_{+}$ implies $\gamma - \delta \in \mathcal{G}^{\mathcal{B}}$. To this aim, let $\mathcal{G}'$ be a $\mathcal{B}$-CG obtained form $\mathcal{G}^{\mathcal{B}}_{+}$ as described in Fact 3. If γ and δ belong to a block V_i such that $\ell_i \neq d$, then $\gamma - \delta \in \mathcal{G}' \in [\mathcal{G}]^{\mathcal{B}}$ and therefore $\gamma - \delta \in \mathcal{G}^{\mathcal{B}}$. Otherwise, if γ and δ belong to a block V_i such that $\ell_i = d$, and $\gamma \to \delta \in \mathcal{G}'$, then by Fact 2 we can construct a second $\mathcal{B}$-CG $\mathcal{G}''$ equivalent to $\mathcal{G}$ such that $\delta \to \gamma \in \mathcal{G}''$ so that $\gamma - \delta \in \mathcal{G}' \cup \mathcal{G}'' \subseteq \mathcal{G}^{\mathcal{B}}$. □

We are now in a position to give the main result of this section.

Corollary 11 *For a labelled block ordering $\mathcal{B}$ of a vertex set V, let $G = (V, E)$ be a $\mathcal{B}$-CG. Then $\mathcal{G}^{\mathcal{B}}$ is a CG equivalent to $\mathcal{G}$.*

Proof. This is an immediate consequence of Theorem 10 and Corollary 9 above. □

4.2 Construction and characterisation of $\mathcal{G}^{\mathcal{B}}$

We have shown that $\mathcal{G}^{\mathcal{B}}$ is a $w\mathcal{B}$-CG and that it represents a natural characterising graph for the class $[\mathcal{G}]^{\mathcal{B}}$. We now provide a procedure to construct $\mathcal{G}^{\mathcal{B}}$ starting from an arbitrary $\mathcal{B}$-CG $\mathcal{G}$ and then characterise those CGs which are $\mathcal{B}$-essential graphs, that is the CGs that represent some $\mathcal{B}$-equivalence class.

Generalising the approach put forth by Roverato (2005), we propose a greedy strategy based on the operation of merging meta-arrows of $w\mathcal{B}$-CGs. The following theorem guarantees that our procedure is well defined.

Theorem 12 *For a labelled block ordering $\mathcal{B}$ of a vertex set V, let $\mathcal{G} = (V, E)$ and $\mathcal{G}' = (V, E')$ be two equivalent $w\mathcal{B}$-CGs such that $\mathcal{G} \subset \mathcal{G}'$. Then, there exists a finite sequence $\mathcal{G} = \mathcal{G}_0 \subset \mathcal{G}_1 \subset \cdots \subset \mathcal{G}_r = \mathcal{G}'$, with $r \geq 1$, of equivalent $w\mathcal{B}$-CGs such that $\mathcal{G}_j$ is obtained from $\mathcal{G}_{j-1}$ by merging exactly one of its meta-arrows, for all $j = 1, \ldots, r$.*

Proof. To prove the desired result it is sufficient to show that there exists a $w\mathcal{B}$-CG $\bar{\mathcal{G}}$ with $\mathcal{G} \subset \bar{\mathcal{G}} \subseteq \mathcal{G}'$ and such that $\bar{\mathcal{G}}$ can be obtained from $\mathcal{G}$ by merging exactly one of its meta-arrows; recall that in this case $\bar{\mathcal{G}}$ is necessarily equivalent to $\mathcal{G}$ and $\mathcal{G}'$ by Proposition 1. From the definition of $w\mathcal{B}$-CG it follows that the differences between $\mathcal{G}$ and $\mathcal{G}'$ may only involve subgraphs $\mathcal{G}_{V_i}$ and $\mathcal{G}'_{V_i}$ for $i = 1, \ldots, k$. More precisely, we can always find a block V_i of $\mathcal{B}$ such that $\mathcal{G}_{V_i} \subset \mathcal{G}'_{V_i}$ and either $\ell_i = g$ or $\ell_i = d$.

Consider first the case in which $\ell_i = g$. By Theorem 5 we can find a CG $\bar{\mathcal{G}}_{V_i}$ (equivalent to $\mathcal{G}_{V_i}$) with $\mathcal{G}_{V_i} \subset \bar{\mathcal{G}}_{V_i} \subseteq \mathcal{G}'_{V_i}$ and such that $\bar{\mathcal{G}}_{V_i}$ is obtained from $\mathcal{G}_{V_i}$ by merging exactly one meta-arrow. Let $\bar{\mathcal{G}}$ be the graph obtained by replacing $\mathcal{G}_{V_i}$ with $\bar{\mathcal{G}}_{V_i}$ in $\mathcal{G}$. It is not difficult to see that $\bar{\mathcal{G}}$ fulfills all the desired properties: it is a $w\mathcal{B}$-CG with $\mathcal{G} \subset \bar{\mathcal{G}} \subseteq \mathcal{G}'$ by construction and it can be obtained from $\mathcal{G}$ by merging one of its meta-arrows.

Consider now the case in which $\ell_i = d$. We can follow the same procedure used for $\ell_i = g$ with the only difference that we take $\bar{\mathcal{G}}_{V_i}$ in the family of NF-CGs. The existence of a suitable NF-CG is guaranteed by Theorem 5. In this way we construct a CG $\bar{\mathcal{G}}$ that can be obtained from $\mathcal{G}$ by merging one of its meta-arrows and such that $\mathcal{G} \subset \bar{\mathcal{G}} \subseteq \mathcal{G}'$, and we only have to show that $\bar{\mathcal{G}}$ is weakly $\mathcal{B}$-consistent. More precisely, since $\mathcal{G}$ and $\bar{\mathcal{G}}$ only differ for the subgraph corresponding to block V_i, and we have chosen $\bar{\mathcal{G}}_{V_i}$ so that it is a NF-CG equivalent to $\mathcal{G}_{V_i}$, it is sufficient to show that $\bar{\mathcal{G}}$ has no flags $\alpha \to \delta — \gamma$ with $\delta, \gamma \in V_i$ and $\alpha \notin V_i$. This follows by noticing that the presence of such a flag in $\bar{\mathcal{G}}$ would imply that $\alpha \to \delta — \gamma \in \mathcal{G}'$ because $\alpha \to \delta$ is a between-block arrow and, since $\bar{\mathcal{G}} \subseteq \mathcal{G}'$, every line of $\bar{\mathcal{G}}$ is a line in $\mathcal{G}'$. However, $\mathcal{G}'$ has no flags of this kind because it is weakly $\mathcal{B}$-consistent. □

By Theorem 12 we can find $\mathcal{G}^{\mathcal{B}} = \mathcal{G}^{\mathcal{B}}_{+}$ by successively merging meta-arrows of $w\mathcal{B}$-CGs, starting from any given $\mathcal{B}$-GC $\mathcal{G}$. However, in order to turn Theorem 12 into an efficient algorithm, we need a characterisation of those meta-arrows which can be merged obtaining an equivalent $w\mathcal{B}$-CG. This is given by Theorem 13 below, in terms of meta-arrows which have $\mathcal{B}$-*insubstantial* arrowhead.

Definition 8 For a labelled block ordering $\mathcal{B} = (V_i^{\ell_i}, i = 1, \ldots, k)$ of a vertex set V, let $\mathcal{G} = (V, E)$ a $w\mathcal{B}$-CG and $A \rightrightarrows D$ a meta-arrow of $\mathcal{G}$. We say that the arrowhead of $A \rightrightarrows D$ is $\mathcal{B}$-*insubstantial* if the following conditions hold:

(a) $A \cup D \subseteq V_i$ for some block V_i of $\mathcal{B}$;

(b) if $\ell_i = g$, the arrowhead of $A \rightrightarrows D$ is insubstantial in $\mathcal{G}$;

(c) if $\ell_i = d$, the arrowhead of $A \rightrightarrows D$ is strongly insubstantial in $\mathcal{G}$.

□

Note that, in the above definition, having fixed $A \rightrightarrows D$, only one among (b) and (c) is relevant. We also remark that if $\ell_i = d$ then condition (c) of Definition 4 simplifies to $\mathrm{pa}_{\mathcal{G}}(D) \setminus A = \mathrm{pa}_{\mathcal{G}}(A)$ (Roverato, 2005 Proposition 2).

Theorem 13 *For a labelled block ordering $\mathcal{B} = (V_i^{\ell_i}, i = 1, \ldots, k)$ of a vertex set V, let $\mathcal{G} = (V, E)$ a $w\mathcal{B}$-CG and $A \rightrightarrows D$ a meta-arrow of $\mathcal{G}$. Then, the CG $\mathcal{G}'$ obtained from $\mathcal{G}$ by merging the meta-arrow $A \rightrightarrows D$ is a $w\mathcal{B}$-CG equivalent to $\mathcal{G}$ if and only if the arrowhead of $A \rightrightarrows D$ is $\mathcal{B}$-insubstantial.*

Proof. Assume first that the CG $\mathcal{G}'$ obtained from $\mathcal{G}$ by merging the meta-arrow $A \rightrightarrows D$ is a $w\mathcal{B}$-CG equivalent to $\mathcal{G}$. Then, since $\mathcal{G}'$ is weakly $\mathcal{B}$-consistent, it holds that $A \cup D \subseteq V_i$

for some block V_i of $\mathcal{B}$ with either $\ell_i = g$ or $\ell_i = d$. By Theorem 6, the arrowhead of $A \rightrightarrows D$ is insubstantial and if $\ell_i = g$ we are finished. We now show that if $\ell_i = d$ then $\mathrm{pa}(D) \backslash A = \mathrm{pa}(A)$ in $\mathcal{G}$ so that the arrowhead of $A \rightrightarrows D$ is also strongly insubstantial. As $\mathcal{G}'$ is weakly $\mathcal{B}$-consistent, there are no flags in $\mathcal{G}'$ involving lines of $\mathcal{G}'_{V_i}$. Since every arrow of $A \rightrightarrows D$ is a line in $\mathcal{G}'$, this implies that $\mathrm{pa}(A) = \mathrm{pa}(D)$ in $\mathcal{G}'$. Furthermore, by construction $\mathrm{pa}_{\mathcal{G}}(D) \setminus A = \mathrm{pa}_{\mathcal{G}'}(D)$, and $\mathrm{pa}_{\mathcal{G}}(A) = \mathrm{pa}_{\mathcal{G}'}(A)$ so that $\mathrm{pa}(D) \backslash A = \mathrm{pa}(A)$ in $\mathcal{G}$ as required.

Conversely, assume that the arrowhead of $A \rightrightarrows D$ is $\mathcal{B}$-insubstantial in $\mathcal{G}$. In this case Theorem 6 implies that $\mathcal{G}'$ is a CG equivalent to $\mathcal{G}$ so that it remains to show that $\mathcal{G}'$ is weakly $\mathcal{B}$-consistent. By point (a) of Definition 8 we can find a block V_i, with either $\ell_i = g$ or $\ell_i = d$, such that $A \cup D \subseteq V_i$. This implies both that (a) of Definition 6 holds for $\mathcal{G}'$, and that weak $\mathcal{B}$-consistency of $\mathcal{G}'$ has to be checked only with respect to block V_i. If $\ell_i = g$ there is nothing to prove. If $\ell_i = d$ then we have to show that (i) $\mathcal{G}'$ contains no flag $\gamma \rightarrow \alpha - \delta$ such that $\alpha - \delta \in \mathcal{G}'_{V_i}$ and (ii) that the $\mathcal{G}'_{V_i}$ has decomposable chain components. Point (i) follows by noticing that $\gamma \rightarrow \alpha - \delta \in \mathcal{G}'$ implies $\gamma \rightarrow \alpha \in \mathcal{G}$ because $\mathcal{G} \subseteq \mathcal{G}'$ and, more precisely, that either $\gamma \rightarrow \alpha \leftarrow \delta \in \mathcal{G}$ or $\gamma \rightarrow \alpha \rightarrow \delta \in \mathcal{G}$ because $\gamma \rightarrow \alpha - \delta \notin \mathcal{G}$ by (c$'$) in the definition of weak $\mathcal{B}$-consistency. However, $\gamma \rightarrow \alpha \leftarrow \delta$ cannot belong to $\mathcal{G}$ because, in this case, $\mathcal{G}$ and $\mathcal{G}'$ would not be equivalent. Furthermore, $\gamma \rightarrow \alpha \rightarrow \delta$ cannot belong to $\mathcal{G}$ because $\gamma \in \mathrm{pa}_{\mathcal{G}}(\alpha)$ and $\gamma \notin \mathrm{pa}_{\mathcal{G}}(\delta)$ would imply that the arrowhead of the meta-arrow $A \rightrightarrows D$ to which $\alpha \rightarrow \beta$ belongs be not strongly $\mathcal{B}$-insubstantial. We conclude that $\gamma \rightarrow \alpha - \delta \notin \mathcal{G}'$. Point (ii) follows by noticing that if the arrowhead of $A \rightrightarrows D$ is $\mathcal{B}$-insubstantial in $\mathcal{G}$ then it is strongly insubstantial in $\mathcal{G}_{V_i}$, so that by Theorem 6 $\mathcal{G}'_{V_i}$ is a NF-CG equivalent to $\mathcal{G}_{V_i}$ and thus it has decomposable chain components. □

It follows from Theorem 13 that we can start from any $\mathcal{B}$-CG $\mathcal{G} = (V, E)$ and successively merge meta-arrows with $\mathcal{B}$-insubstantial arrowhead until we obtain the $\mathcal{B}$-essential graph $\mathcal{G}^{\mathcal{B}}$ in which no meta-arrow can be merged. We also remark that this procedure can be applied to every block of $\mathcal{B}$ independently of other blocks, i.e. blocks can be processed in parallel. Furthermore, if two $\mathcal{B}$-CGs $\mathcal{G}$ and $\tilde{\mathcal{G}}$ differ only with respect to the subgraph induced by the block V_i, then the corresponding $\mathcal{B}$-essential graphs $\mathcal{G}^{\mathcal{B}}$ and $\tilde{\mathcal{G}}^{\mathcal{B}}$ will also differ only with respect to the subgraphs $\mathcal{G}^{\mathcal{B}}_{V_i}$ and $\tilde{\mathcal{G}}^{\mathcal{B}}_{V_i}$; and the same is true if the difference involves more than one block. This property could be exploited to develop an efficient search procedure in the space of $\mathcal{B}$-essential graphs.

We conclude with the announced characterisation of those CGs which are $\mathcal{B}$-essential graphs and thus represent some $\mathcal{B}$-equivalence class.

Theorem 14 *A graph $\tilde{\mathcal{G}}$ is the $\mathcal{B}$-essential graph $\mathcal{G}^{\mathcal{B}}$ of a $\mathcal{B}$-equivalence class $[\mathcal{G}]^{\mathcal{B}}$ for some $\mathcal{B}$-CG $\mathcal{G}$ if and only if*

(i) $\tilde{\mathcal{G}}$ is a w$\mathcal{B}$-CG;

(ii) no meta-arrow of $\tilde{\mathcal{G}}$ has $\mathcal{B}$-insubstantial arrowhead.

Proof. If $\tilde{\mathcal{G}} = \mathcal{G}^{\mathcal{B}}$ for some $\mathcal{B}$-CG $\mathcal{G}$, then (i) follows from Theorem 10 and (ii) holds because otherwise, by Theorem 13, a larger $w\mathcal{B}$-CG could be found in $[\mathcal{G}]^{\mathcal{B}}_{+}$, but this is not possible because by Theorem 10 $\mathcal{G}^{\mathcal{B}}$ is the largest CG in $[\mathcal{G}]^{\mathcal{B}}_{+}$. Conversely, if (i) holds then $[\mathcal{G}']^{\mathcal{B}}_{+}$ is well defined and, by Theorem 12 and Theorem 13, point (ii) implies that $\tilde{\mathcal{G}}$ is the largest element of $[\tilde{\mathcal{G}}]^{\mathcal{B}}_{+}$. □

5 Discussion

As shown by Lauritzen and Richardson (2002), the interpretation of CGs is not as straightforward as it may appear. In particular, undirected edges represent a very special kind of symmetric association between variables. For this reason, constraining an edge between two vertices, if present, to be a line, cannot be simply motivated by the assumption that the two corresponding variables are on an equal footing and, in fact, it may constitute a much stronger assumption than believed. As well as "traditional" CG modelling, the framework of $\mathcal{B}$-CGs allows to implement prior knowledge on variable block ordering, but is also gives additional flexibility in the specification of associations within blocks.

A related issue concerns the confounding effect of latent variables. Lauritzen and Richardson (2002) showed that, due to the presence of latent variables, the most appropriate CG for a given problem may present arrows pointing against the causal ordering of the variables. In this case they suggest that a possible strategy is to ignore the block ordering when performing structural learning, which in our framework amounts to setting $\mathcal{B} = (V^g)$. However, it may also be known that latent variables are only connected with a proper, possibly small, subset of variables X_A with $A \subset V$. In this case $\mathcal{B}$-CGs allow the implementation of a mixed strategy in which the substantive labelled block ordering $\mathcal{B}$ is first specified, and then certain successive blocks are merged into a single block V_i with label $\ell_i = g$ in such a way that $A \subseteq V_i$.

We remark that $\mathcal{B}$-essential graphs are interesting objects. Both largest CGs and essential graphs are special cases of $\mathcal{B}$-essential graphs. More interestingly, however, $\mathcal{B}$-essential graphs may be thought of as DAGs of boxes where every box is an *undirected subgraph* for UG-boxes, an *essential subgraph* for DAG-boxes and a *largest chain subgraph* for CG-boxes. Indeed, it is easy to check that the characterisation of essential graphs provided by Studený (2004) and Roverato (2005) can also be used to characterise the DAG-boxes of $\mathcal{B}$-essential graphs if, in checking strong insubstantiality of meta arrows, also the parents outside the box are considered. In a similar way the characterisation of largest CGs given by Roverato (2005) can also be used to characterise the CG-boxes of $\mathcal{B}$-essential graphs.

Acknowledgements

The authors wish to thank Davide Cavallini for useful comments. Part of this paper was written when the first author was visiting the Center for Genomic Regulation, University

Pompeu Fabra, Spain, supported by a grant of Secretaría de Estado de Universitades e Investigatción, and the second author was visiting the Department of Mathematical Sciences, Aalborg University, Denmark, supported by a grant of Danish Natural Science Research Council. Financial support was also provided by MIUR, Rome (PRIN 03).

References

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

Andersson, S.A., Madigan, D. and Perlman, M.D. (1997b). On the Markov equivalence of chain graphs, undirected graphs, and acyclic digraphs. *Scand. J. Statist.* **24**, 81–102.

Andersson, S.A., Madigan, D. and Perlman, M.D. (2001). Alternative Markov properties for chain graphs. *Scand. J. Statist.*, **24**, 81–102.

Caputo, A., Foraita, F., Klasen, S. and Pigeot, I. (2003). Undernutrition in Benin — An analysis based on graphical models. *Social Science and Medicine*, **56**, 1677–1691.

Chickering, D.M. (2002). Learning equivalence classes of Bayesian-network structure. *J. Mach. Learn. Res.*, **2**, 445–498.

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

Cox, D.R. and Wermuth, N. (1993). Linear dependencies represented by chain graphs. *Statist. Science*, **8**, 3, 204–218.

Cox, D.R. and Wermuth, N. (1996). *Multivariate Dependencies: Models, Analysis and Interpretations.* Chapman & Hall, London.

Cox, D.R. and Wermuth, N. (2000). On the generation of the chordless four-cycle. *Biometrika*, **87**, 206–12.

Drton, M. and Perlman, M.D. (2004). Model selection for Gaussian concentration graphs. *Biometrika*, **91**, 591–602.

Edwards, D. (2000). *Introduction to Graphical Modelling* (2nd Edition). Springer, New York.

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

Giudici, P. and Green, P.J. (1999). Decomposable graphical Gaussian model determination. *Biometrika*, **86**, 785-801.

Jordan, M.I. (Ed.) (1998). *Learning in Graphical Models.* MIT Press.

Kiiveri, H., Speed, T. and Carlin, J. (1984). Recursive causal models. *J. Aust. Math. Soc. A*, **36**, 30–52.

Lauritzen, S.L. (1996). *Graphical Models.* Oxford University Press, Oxford.

Lauritzen, S.L. and Richardson, T.S. (2002). Chain graphs models and their causal interpretation (with discussion). *J. Roy. Statist. Soc. B*, **64**, 321–361.

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

Meek, C. (1995). Causal inference and causal explanation with background knowledge. In: *Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence* (eds P. Besnard & S. Hanks), 403–410. Morgan Kaufmann, San Francisco.

Madigan, D., Andersson, S.A., Perlman, M.D. and Volinsky, C.M. (1996). Bayesian model averaging and model selection for Markov equivalence classes of acyclic digraphs. *Commun. Statist. Theory Methods*, **25**, 2493–2519.

Mohamed, W.N., Diamond, I. and Smith, P.W.F. (1998). The determinants of infant mortality in Malaysia: A graphical chain modelling approach. *J. Roy. Statist. Soc. A*, **161**, 349–366.

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

Pearl, J. (2000). *Causality: models, reasoning and inference.* Cambridge University Press, Cambridge.

Pigeot, I., Heinicke, A., Caputo, A. and Brüderl, J. (2000). The professional career of sociologists: A graphical chain model reflecting early influences and associations. *Allgemeines Statistisches Archiv*, **84**, 3–21.

Richardson, T. (2001). *Chain graphs which are maximal ancestral graphs are recursive causal graphs.* Technical Report, Department of Statistics, Washington University, USA. Available at `http://www.stat.washington.edu/www/research/reports/`.

Roverato, A. (2002). Hyper inverse Wishart distribution for non-decomposable graphs and its application to Bayesian inference for Gaussian graphical models. *Scand. J. Statist.*, **29**, 391–411.

Roverato, A. (2005). A unified approach to the characterisation of equivalence classes of DAGs, chain graphs with no flags and chain graphs. *Scand. J. Statist.*, to appear.

Roverato, A. and Studený, M. (2004). *A Graphical Representation of Equivalence Classes of AMP Chain Graphs.* Technical Report. Dipartimento di Scienze Sociali Cognitive e Quantitative. Università di Modena e Reggio Emilia.

Spirtes, P., Glymour, C., and Scheines, R. (2001). *Causation, Prediction and Search* (2nd Edition). MIT Press.

Studený, M. (1996). On separation criterion and recovery algorithm for chain graphs. In *Uncertainty in Artificial Intelligence: Proceedings of the Twelfth Conference* (eds F. Jensen & E. Horvitz), 509–516. Morgan Kaufmann, San Francisco.

Studený, M. (1997). On recovery algorithm for chain graphs. *Int. J. Approx. Reasoning*, **17**, 265–293.

Studený, M. (2004). Characterisation of essential graphs by means of an operation of legal component merging. *Int. J. Uncertain. Fuzziness Knowledge-Based Systems* **12**, 43–62.

Verma, T. and Pearl, J. (1991). Equivalence and synthesis of causal models. In: *Uncertainty in Artificial Intelligence 6* (eds P.P. Bonissone, M. Henrion, L.N. Kanal & J.F. Lemmer), 255–268. North-Holland, Amsterdam.

Verma, T. and Pearl, J. (1992). An algorithm for deciding if a set of observed independencies has a causal explanation. In: *Uncertainty in Artificial Intelligence 8* (eds D. Dubois, M.P. Wellman, B. D'Ambrosio & P. Smets), 323–330. Morgan Kaufmann, San Francisco.

Volf, M. and Studený, M. (1999). A graphical characterisation of the largest chain graphs. *Int. J. Approx. Reasoning*, **20**, 209–236.

Wermuth, N. and Lauritzen, S.L. (1990). On substantive research hypotheses, conditional independence graphs and graphical chain models (with discussion). *J. Roy. Statist. Soc. B*, **52**, 21–72.

Whittaker, J. (1990). *Graphical Models in Applied Multivariate Statistics.* Wiley, Chichester.