 Vol 32: 295–312, 2005

A Unified Approach to the Characterization of Equivalence Classes of DAGs, Chain Graphs with no Flags and Chain Graphs

ALBERTO ROVERATO

Department of Social Cognitive and Quantitative Sciences, University of Modena and Reggio Emilia

ABSTRACT. A Markov property associates a set of conditional independencies to a graph. Two alternative Markov properties are available for chain graphs (CGs), the Lauritzen–Wermuth–Frydenberg (LWF) and the Andersson–Madigan– Perlman (AMP) Markov properties, which are different in general but coincide for the subclass of *CGs with no flags*. Markov equivalence induces a partition of the class of CGs into equivalence classes and every equivalence class contains a, possibly empty, subclass of CGs with no flags itself containing a, possibly empty, subclass of directed acyclic graphs (DAGs). LWF-Markov equivalence classes of CGs can be naturally characterized by means of the so-called *largest CGs*, whereas a graphical characterization of equivalence classes of DAGs is provided by the *essential graphs*. In this paper, we show the existence of largest CGs with no flags that provide a natural characterization of equivalence classes of CGs of this kind, with respect to both the LWF- and the AMP-Markov properties. We propose a procedure for the construction of the largest CGs, the largest CGs with no flags and the essential graphs, thereby providing a unified approach to the problem. As by-products we obtain a characterization of graphs that are largest CGs with no flags and an alternative characterization of graphs which are largest CGs. Furthermore, a known characterization of the essential graphs is shown to be a special case of our more general framework. The three graphical characterizations have a common structure: they use two versions of a locally verifiable graphical rule. Moreover, in case of DAGs, an immediate comparison of three characterizing graphs is possible.

Key words: chain graph, chain graph with no flags, conditional independence, directed acyclic graph, essential graph, graphical model, Markov equivalence, Markov property, meta-arrow

1. Introduction

Graphical Markov models are a class of statistical models in which a graph is used to represent conditional independence relations among the variables of a probability distribution. The vertices represent the variables and the set of conditional independencies encoded by the graph is determined through a Markov property associated with the graph. Recent books on this subject include Pearl (1988, 2000), Whittaker (1990), Cox & Wermuth (1996), Lauritzen (1996), Cowell *et al*. (1999) and Edwards (2000).

Lauritzen & Wermuth (1989) introduced a class of models for *chain graphs* (CGs) which admit both directed edges (arrows) and undirected edges but not semi-directed cycles. In Lauritzen & Richardson (2002) the interpretation of these models is discussed whereas the mathematical theory of the corresponding Markov property, hereafter referred to as the *Lauritzen–Wermuth–Frydenberg (LWF)-Markov property*, can be found in Frydenberg (1990). More recently, an alternative *AMP-Markov property* for CGs has been proposed by Andersson *et al*. (2001).

This paper is concerned with *Markov equivalence* of CGs. Several different CGs, with common vertex set, may be equivalent with respect to a given Markov property, in the sense that they encode the same set of conditional independencies. Markov equivalence is an equivalence relation that induces a partition of the set of CGs into equivalence classes, and it is well known that model determination procedures that deal 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. On the other hand, the drawback of considering equivalence classes is that in this case most of the advantages deriving from the graphical representation of the models are lost, unless it is possible to characterize the equivalence class through a CG (Frydenberg, 1990; Verma & Pearl, 1992; Heckerman *et al.*, 1995; Chickering, 1995, 2002; Meek, 1995; Madigan *et al.*, 1996; Andersson *et al.*, 1997a, 1997b; Studený, 1996, 1997, 2004; Volf & Studený, 1999).

Frydenberg (1990, Theorem 5.6) gave a necessary and sufficient condition for the LWF-Markov equivalence of two CGs and showed that every equivalence class contains one largest CG whose arrows are common to every element of the class. Studený (1997) provided an algorithm for the construction of the largest CG and characterized such a graph on the basis of a property of its chain components. Successively, Volf & Studený (1999) gave an alternative graphical characterization of the largest CGs based on the concept of a *protected arrow*.

CGs with no undirected edges, called *directed acyclic graphs* (DAGs), play a central role in the theory of graphical models. For a given DAG, $\mathcal{D}$, it is of interest to consider the equivalence class of all DAGs Markov equivalent to $\mathcal{D}$, but in this case there is no DAG that provides a natural representative of the class. Typically, equivalence classes of DAGs are characterized by means of the smallest CG that is larger than every element of the class. This graph was called the *essential graph* by Andersson *et al.* (1997a) but it is also known in the literature as a completed pattern (Verma & Pearl, 1991), a maximally oriented path for a pattern (Meek, 1995) and a completed p-dag (Chickering, 2002). Andersson *et al.* (1997a) provided an algorithm for constructing the essential graph of an arbitrary DAG as well as a graphical characterization of the essential graph based on the concept of a *strongly protected arrow*. Recently, Studený (2004) introduced a construction procedure of essential graphs based on the operation of *legal component merging* corresponding to an alternative characterization of the essential graphs.

With respect to the AMP-Markov property, a necessary and sufficient condition for the Markov equivalence of CGs is given in Andersson *et al.* (2001, Theorem 5), but the characterization of equivalence classes is more difficult than in the LWF case because a largest CG AMP-Markov equivalent to every element of the class may not exist. In this context, it is important to distinguish between *CGs with no flags* (NF-CGs) and CGs with flags. A CG $\mathcal{G}$ is said to have no flags if the configuration $\alpha \to \beta - \delta$ does not occur as an induced subgraph of $\mathcal{G}$ and Andersson *et al.* (2001, Theorem 4) showed that the LWF- and the AMP-Markov properties are identical if and only if the CG has no flags. It follows that, if one's attention is restricted to the subclass of NF-CGs a distinction between the two Markov properties is not explicitly necessary. Although the subclass of NF-CGs arises in the comparison of the two Markov properties, it is worth pointing out that NF-CGs also occur in other relevant contexts. For instance, both DAGs and undirected graphs are NF-CGs. Moreover, the class of *recursive causal graphs* introduced by Kiiveri *et al.* (1984) is a subset of the class of NF-CGs, and Richardson (2001) showed that a CG is equivalent to some DAG under marginalizing and conditioning if and only if it is equivalent to a recursive causal graph.

For an NF-CG $\mathcal{G}$ both the respective LWF- and AMP-Markov equivalence classes may contain CGs with flags, and the intersection of the two equivalence classes appears to be the equivalence class of Markov equivalent NF-CGs, i.e. the class of CGs equivalent to $\mathcal{G}$ with respect to both Markov properties. However, the largest CG LWF-Markov equivalent to $\mathcal{G}$ does not provide a natural representative of the equivalence class of NF-CGs because it may contain some flag, and it is of interest to investigate the existence, and the successive characterization, of a largest NF-CG Markov equivalent to $\mathcal{G}$, as this would provide a natural representative of both the class of CGs LWF-Markov equivalent to $\mathcal{G}$ and the class of CGs AMP-Markov equivalent to $\mathcal{G}$.

In this paper we present a procedure that, for an arbitrary CG $\mathcal{G}$, constructs the largest CG LWF-Markov equivalent to $\mathcal{G}$. The most relevant aspect of the proposed procedure is its versatility: it can be modified both to deal with the subclass of NF-CGs and for the construction of essential graphs, thereby providing a unified approach to the problem. We show that for every equivalence class of NF-CGs there exists a unique largest NF-CG. The application of our procedure to an NF-CG $\mathcal{G}$ can return two outputs: (1) the largest CG Markov equivalent to $\mathcal{G}$ with respect to the LWF-Markov property, (2) the largest NF-CG Markov equivalent to $\mathcal{G}$, which is nothing but the largest CG equivalent to $\mathcal{G}$ with respect to both Markov properties (LWF and AMP). Furthermore, if $\mathcal{G}$ is a DAG, then our procedure for the construction of the largest NF-CG coincides with the procedure introduced by Studený (2004) for constructing the essential graph of $\mathcal{G}$. As by-products we obtain an alternative characterization of the largest CGs, a characterization of the largest NF-CGs and, finally, an alternative derivation of the characterization of essential graphs given by Studený (2004). The three characterizations are based on local properties of meta-arrows, where a meta-arrow is the collection of all arrows joining two chain components, and provide a common set of rules to deal efficiently with different characterization problems.

The general theory relating to graphical model theory and Markov equivalence is presented in section 2. In section 3, we show the existence of the largest NF-CG for an equivalence class of NF-CGs and describe its connection with the theory of essential graphs. The construction procedure and the characterizations of the largest CGs and of the largest NF-CGs are provided in sections 4 and 5, respectively.

2. Preliminaries

In this section, we review the graphical model theory required in this paper. We omit the definitions of well-established concepts such as *parent*, *child*, *path*, *skeleton*, *semi-directed cycle*, *perfect numbering*, *perfect directed version* and *decomposable graph*; we refer to Cowell *et al.* (1999) for a full account of the theory of graphs and graphical models.

2.1. Graph terminology

A *graph* is a pair $\mathcal{G} = (V, E)$ where V is a finite set of *vertices* and $E \subseteq V \times V$ is a set of *edges*. Two vertices joined by an edge are called *adjacent* and to denote that α and β are not adjacent we write $\alpha \not\sim \beta$. We say that α and β are joined by an arrow pointing at β, and write $\alpha \to \beta \in \mathcal{G}$, if $(\alpha, \beta) \in E$ but $(\beta, \alpha) \notin E$. We write $\alpha - \beta \in \mathcal{G}$ if both $(\alpha, \beta) \in E$ and $(\beta, \alpha) \in E$ and say that there is an undirected edge between α and β. For a subset $A \subseteq V$ we denote by $\mathrm{pa}_{\mathcal{G}}(A)$ the parents of A in $\mathcal{G}$, or simply $\mathrm{pa}(A)$ when it is clear from the context which graph is being considered.

A CG $\mathcal{G} = (V, E)$ is a graph that has no semi-directed cycles. For a pair of vertices $\alpha, \beta \in V$, we write $\alpha \leftrightharpoons \beta$ if either $\alpha = \beta$ or there is an undirected path between α and β. It is straightforward to see that $\leftrightharpoons$ is an equivalence relation that induces a partition of the vertex set V into equivalence classes. We denote by $\mathcal{T}(\mathcal{G})$ the corresponding set of equivalence classes and the elements $T \in \mathcal{T}(\mathcal{G})$ are called the *chain components* of $\mathcal{G}$. We say that a chain component T of $\mathcal{G}$ is decomposable if the subgraph of $\mathcal{G}$ induced by T, $\mathcal{G}_T$, is decomposable; furthermore when we say that a CG has decomposable chain components we mean that *all* its chain components are decomposable. A CG in which all the chain components are singletons is called a DAG. Hereafter, to stress that a graph is a DAG we write $\mathcal{D} = (V, E)$. We are mainly interested in chain components rather than single vertices of CGs and consequently a central role is played by the concept of the meta-arrow.

Definition 1

For a CG $\mathcal{G} = (V, E)$, the meta-arrow *between the chain components A and B of $\mathcal{G}$, denoted by $A \Rightarrow B$, is the set of all the arrows pointing from A to B; i.e.*

$$A \Rightarrow B := \{(\alpha, \beta) \in E | \alpha \in A, \beta \in B\}.$$

Note that in a DAG every meta-arrow is made up of at most one arrow and that, in general, $A \Rightarrow B \neq \emptyset$ implies $A \Leftarrow B = \emptyset$, because otherwise there would be a semi-directed cycle in $\mathcal{G}$.

An *immorality* in a graph $\mathcal{G}$ is an induced subgraph of the form $\alpha \rightarrow \beta \leftarrow \delta$ whereas a subgraph of the form $\alpha \rightarrow \beta - \delta$ is called a *flag* (in both cases α and δ are not adjacent). A triple (γ, S, δ) is called a *complex* of a CG $\mathcal{G} = (V, E)$ if S is a connected subset of a chain component and γ and δ are two non-adjacent vertices in pa(S). A complex (γ, S, δ) is *minimal* in $\mathcal{G}$ if no proper subset $S' \subset S$ forms a complex (γ, S', δ). Frydenberg (1990) noted that (γ, S, δ) is a minimal complex in $\mathcal{G}$ if and only if $\mathcal{G}_{S \cup \{\gamma, \delta\}}$ looks like:

$$\gamma \rightarrow v_1 - v_2 - \cdots - v_{k-1} - v_k \leftarrow \delta \tag{1}$$

where $S = \{v_1,\ldots,v_k\}$. The configuration (1) motivates the alternative notation of the minimal complex (γ, S, δ) through the ordered sequence $(\gamma, v_1,\ldots,v_k, \delta)$. Two different CGs may share the same set of complexes.

Definition 2

Two CGs are called complex-equivalent *if they have the same skeleton* (*i.e. undirected version*) *and the same minimal complexes.*

Note that for $k = 1$ the graph (1) simplifies to the immorality $\gamma \rightarrow v_1 \leftarrow \delta$ whereas to a minimal complex with $k > 1$ are associated two flags: $\gamma \rightarrow v_1 - v_2$ and $v_{k-1} - v_k \leftarrow \delta$.

A *triplex* in $\mathcal{G}$ is an ordered pair $(\{\alpha, \delta\}, \beta)$ such that the subgraph of $\mathcal{G}$ induced by $\{\alpha, \delta, \beta\}$ is either the immorality $\alpha \rightarrow \beta \leftarrow \delta$ or one of the flags $\alpha \rightarrow \beta - \delta$ and $\alpha - \beta \leftarrow \delta$.

Definition 3

Two CGs are called triplex-equivalent *if they have the same skeleton and the same triplexes.*

DAGs have no complexes other than immoralities as well as no triplexes other than immoralities and consequently two DAGs are complex-equivalent if and only if they are triplex-equivalent. More generally, however, such feature remains true for the larger class of *chain graphs with no flags* (NF-CGs) and, in the following, when two graphs are both complex- and triplex-equivalent we will simply say that they are *equivalent*. Note also that NF-CGs satisfy other useful properties some of which are summarized in appendix A.

2.2. Markov equivalence

Let $X = (X_\alpha)_{\alpha \in V}$ be a collection of random variables taking values in the sample space $\mathcal{X} = \times_{\alpha \in V} \mathcal{X}_\alpha$. The sample spaces 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. Let $\mathcal{M}(\mathcal{G}, \mathcal{X})$ denote the set of probability distributions on $\mathcal{X}$ that satisfy the conditional independence relations associated with $\mathcal{G}$. Two graphs $\mathcal{G}_1 = (V, E_1)$ and $\mathcal{G}_2 = (V, E_2)$ are said to be *Markov equivalent* if $\mathcal{M}(\mathcal{G}_1, \mathcal{X}) = \mathcal{M}(\mathcal{G}_2, \mathcal{X})$ for every product space $\mathcal{X}$ indexed by V.

Let $\mathcal{M}_{\mathrm{LWF}}(\mathcal{G}, \mathcal{X})$ denote the set of probability distributions that satisfy the LWF-Markov property relative to $\mathcal{G}$ on $\mathcal{X}$. Frydenberg (1990, Theorem 5.6), and in a more general context Andersson *et al.* (1997b, Theorem 3.1), gave the following necessary and sufficient condition for the LWF-Markov equivalence of two CGs.

Theorem 1
Two CGs are LWF-Markov equivalent if and only if they are complex-equivalent.

Note that theorem 1 implies that two DAGs, and more generally two NF-CGs, are LWF-Markov equivalent if and only if they have the same skeleton and the same immoralities (see also Verma & Pearl, 1991).

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 $E_2 \subset E_1$. For instance, in Fig. 1 graph (c) is larger than both (a) and (b) and graph (b) is larger than (a). The problem of checking the complex-equivalence of CGs is simplified if the graphs are nested.

Proposition 1
Let $\mathcal{G} = (V, E)$, $\bar{\mathcal{G}} = (\mathrm{V}, \bar{\mathrm{E}})$ and $\tilde{\mathcal{G}} = (\mathrm{V}, \tilde{\mathrm{E}})$ be three CGs with the same skeleton and such that $\mathcal{G} \subseteq \bar{\mathcal{G}} \subseteq \tilde{\mathcal{G}}$.

(i) $\mathcal{G}$ and $\bar{\mathcal{G}}$ are complex-equivalent if and only if every minimal complex of $\mathcal{G}$ is a minimal complex in $\bar{\mathcal{G}}$.
(ii) If $\mathcal{G}$ and $\tilde{\mathcal{G}}$ are complex-equivalent, then $\bar{\mathcal{G}}$ is complex-equivalent to both $\mathcal{G}$ and $\tilde{\mathcal{G}}$.

Proof. See the appendix.

The union of two CGs $\mathcal{G}_1 = (V, E_1)$ and $\mathcal{G}_2 = (V, E_2)$ is $\mathcal{G}_1 \cup \mathcal{G}_2 = (V, E_1 \cup E_2)$. 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 & Studený, 1999). Frydenberg (1990, Proposition 5.4) showed the following.

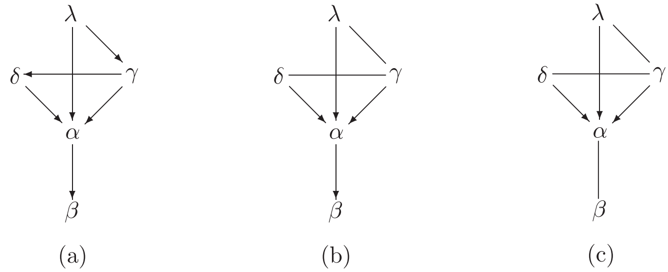


Fig. 1. (a) DAG $\mathcal{D}$; (b) its essential graph $\mathcal{D}^{[\vee]} = \mathcal{D}^{\langle\vee\rangle}$; (c) $\mathcal{D}^{[[\vee]]_c}$, the largest CG complex-equivalent to $\mathcal{D}$.

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 both have the same minimal complexes as $\mathcal{G}_1 \vee \mathcal{G}_2$.

Theorem 2 implies that for every set H of complex-equivalent CGs there is a unique smallest CG larger than, and complex-equivalent to, every element of H. This is given by $\vee(\mathcal{G}|\mathcal{G} \in H)$, the $\vee$-union of all the graphs $\mathcal{G} \in H$.

Complex-equivalence is an equivalence relation and, for a CG $\mathcal{G} = (V, E)$ we denote by $[\![\mathcal{G}]\!]_c$ the equivalence class made up of all CGs complex-equivalent to $\mathcal{G}$, and by $\mathcal{G}^{[\![\vee]\!]_c} = \vee(\mathcal{G}'|\mathcal{G}' \in [\![\mathcal{G}]\!]_c)$ the (unique) largest CG complex-equivalent to $\mathcal{G}$ (see also Frydenberg, 1990, Proposition 5.7). Since, by theorem 2, $\mathcal{G}^{[\![\vee]\!]_c} \in [\![\mathcal{G}]\!]_c$, it follows that

$$\mathcal{G}^{[\![\vee]\!]_c} = \cup(\mathcal{G}'|\mathcal{G}' \in [\![\mathcal{G}]\!]_c).$$

For an arbitrary CG $\mathcal{G}$, Volf & Studený (1999) provided a rule to identify the arrows of $\mathcal{G}$ that correspond to undirected edges in $\mathcal{G}^{[\![\vee]\!]_c}$ whereas Studený (1996, 1997) introduced a recovery algorithm for the construction of $\mathcal{G}^{[\![\vee]\!]_c}$. To the two procedures are associated two graphical characterizations of $\mathcal{G}^{[\![\vee]\!]_c}$, the former based on the concept of the *protected* arrow and the latter on a property of the chain components of $\mathcal{G}^{[\![\vee]\!]_c}$.

We now turn to the AMP-Markov property. Let $\mathcal{M}_{\mathrm{AMP}}(\mathcal{G}, \mathcal{X})$ denote the set of probability distributions that satisfy the AMP-Markov property relative to $\mathcal{G}$ on $\mathcal{X}$. Andersson *et al.* (2001, Theorem 5) gave the following condition for the AMP-Markov equivalence of two CGs.

Theorem 3

Two CGs are AMP-Markov equivalent if and only if they are triplex-equivalent.

As well as for the LWF-Markov equivalence, the AMP-Markov equivalence induces a partition of the set of CGs into equivalence classes. However, Andersson *et al.* (2001) showed that a largest CG AMP-Markov equivalent to every element of the equivalence class may not exist. For instance, the CGs AMP-Markov equivalent to $\alpha \rightarrow \beta - \delta$ are $\alpha - \beta \leftarrow \delta$ and $\alpha \rightarrow \beta \leftarrow \delta$. Clearly, no one of these CGs is larger than every other graph in the class. Furthermore, by taking the union of the three graphs one obtains the CG $\alpha - \beta - \delta$ that is not AMP-Markov equivalent to the elements of the equivalence class.

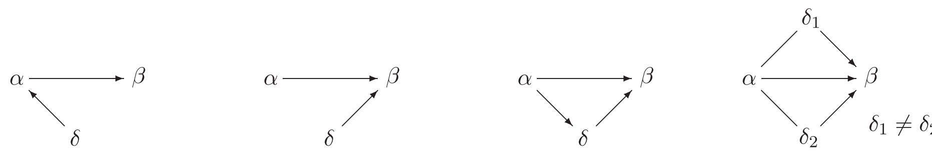
Andersson *et al.* (2001, Theorem 4) showed that a necessary and sufficient condition for the LWF- and AMP-Markov properties to coincide (formally, $\mathcal{M}_{\mathrm{LWF}}(\mathcal{G}, \mathcal{X}) = \mathcal{M}_{\mathrm{AMP}}(\mathcal{G}, \mathcal{X})$ for every product space $\mathcal{X}$ indexed by V) is that $\mathcal{G} = (V, E)$ is an NF-CG. As a consequence, for NF-CGs a distinction between the two Markov property is not necessary. However, for an NF-CG $\mathcal{G}$ the largest CG complex-equivalent to $\mathcal{G}$ may have some flag and, consequently, for this graph the two Markov properties may not coincide. For instance, for the DAG (a) in Fig. 1 the two Markov properties are identical, but the largest CG (c) has some flags and the two Markov properties differ for this graph. In particular, graphs (a) and (c) in Fig. 1 are LWF-Markov equivalent, but they are not AMP-Markov equivalent because they do not have the same triplexes; for instance, the pair $(\{\gamma, \beta\}, \alpha)$ is a triplex in (c) but not in (a). A possible solution to this problem can be provided by the identification of a characterizing graph belonging to the subclass of NF-CGs equivalent to $\mathcal{G}$. Such a graph would simultaneously characterize the equivalence class of CGs AMP-Markov equivalent to $\mathcal{G}$ and the class of CGs LWF-Markov equivalent to $\mathcal{G}$.

Every DAG $\mathcal{D} = (V, E)$ is an NF-CG so that for this subclass of graphs the two Markov properties coincide. For a DAG $\mathcal{D}$ let $[\mathcal{D}]$ be the set of all DAGs Markov equivalent to $\mathcal{D}$ and let $\mathcal{D}^{[\vee]} = \vee(\mathcal{D}'|\mathcal{D}' \in [\mathcal{D}])$ denote the smallest CG larger than every $\mathcal{D}' \in [\mathcal{D}]$. Clearly

$[\mathcal{D}] \subseteq [\![\mathcal{D}]\!]_c$, so that $\mathcal{D}^{[\vee]} \subseteq \mathcal{D}^{[\![\vee]\!]_c}$. Note also that $\mathcal{D}^{[\vee]} \notin [\mathcal{D}]$ unless $\mathcal{D}$ is the only member of $[\mathcal{D}]$. Andersson *et al*. (1997a) considered the graph $\mathcal{D}^*$ obtained as the union of all the DAGs in $[\mathcal{D}]$, formally $\mathcal{D}^* = \cup(\mathcal{D}' | \mathcal{D}' \in [\mathcal{D}])$, and called it the *essential graph* of $\mathcal{D}$. They showed that $\mathcal{D}^* = \mathcal{D}^{[\vee]}$, and provided the following characterization of the essential graph.

Theorem 4 (*Andersson et al., 1997a*)
A graph $\mathcal{G} = (V, E)$ is equal to $\mathcal{D}^{[\vee]}$ for some DAG $\mathcal{D}$ if and only if $\mathcal{G}$ satisfies the following conditions:

(i) $\mathcal{G}$ is an NF-CG with decomposable chain components;
(ii) every arrow $\alpha \rightarrow \beta$ is strongly protected in $\mathcal{G}$, i.e. it occurs in at least one of the following four configurations as an induced subgraph of $\mathcal{G}$:



(a) (b) (c) (d)

Studený (2004), starting from theorem 4, derived a procedure for the construction of essential graphs based on the operation of *legal component merging* and introduced a corresponding alternative characterization of the essential graph.

3. Chain graphs with no flags

For an NF-CG $\mathcal{G} = (V, E)$ the equivalence class of all NF-CGs equivalent to $\mathcal{G}$ corresponds to the intersection of the set of CGs LWF-Markov equivalent to $\mathcal{G}$ with the set of CGs AMP-Markov equivalent to $\mathcal{G}$ (this is an immediate consequence of theorem 4 of Andersson *et al*., 2001). Hence, a graphical characterization of such equivalence class would simultaneously provide a characterization of both larger equivalence classes. Hereafter, we write $\mathcal{K} = (V, E)$ to stress that a graph is an NF-CG and denote by $\langle\mathcal{K}\rangle$ the class of NF-CGs equivalent to $\mathcal{K}$. Formally, if $\mathcal{NF}$ denotes the class of all NF-CGs then $\langle\mathcal{K}\rangle = [\![\mathcal{K}]\!]_c \cap \mathcal{NF}$. Furthermore, if $[\![\mathcal{G}]\!]_t$ denotes the class of all CGs triplex-equivalent to $\mathcal{G}$, it holds that $\langle\mathcal{K}\rangle = [\![\mathcal{K}]\!]_t \cap \mathcal{NF}$ and that $\langle\mathcal{K}\rangle = [\![\mathcal{K}]\!]_c \cap [\![\mathcal{K}]\!]_t$.

In this section, we show that for every equivalence class of NF-CGs there exists a largest NF-CG, i.e. that the smallest CG larger than every element of $\langle\mathcal{K}\rangle$, denoted by $\mathcal{K}^{\langle\vee\rangle} = \vee(\mathcal{K}' | \mathcal{K}' \in \langle\mathcal{K}\rangle)$, is an NF-CG . Furthermore, if the equivalence class includes a DAG $\mathcal{D}$, then such a largest graph coincides with the essential graph of $\mathcal{D}$.

We first need the following result.

Lemma 1
Let $\mathcal{K}_1 = (V, E_1)$ and $\mathcal{K}_2 = (V, E_2)$ two equivalent NF-CGs and let $\beta = \beta_0, \beta_1, \ldots, \beta_{k-1}, \beta_k = \beta$ be a semi-directed cycle in $\mathcal{K}_1 \cup \mathcal{K}_2$. If $\alpha \rightarrow \beta \in \mathcal{K}_1 \vee \mathcal{K}_2$ then, for all $i = 1, \ldots, k$, $\alpha \rightarrow \beta_i \in \mathcal{K}_1 \vee \mathcal{K}_2$.

Proof. See the appendix.

We can now give the main result of this section.

Theorem 5
Let $\mathcal{K}_1 = (V, E_1)$ and $\mathcal{K}_2 = (V, E_2)$ be two equivalent NF-CGs, then $\mathcal{K}_1 \vee \mathcal{K}_2$ is an NF-CG.

Proof. $\mathcal{K}_1 \vee \mathcal{K}_2$ is a CG by construction so that it is sufficient to show that it has no flags. As $\mathcal{K}_1$ and $\mathcal{K}_2$ have no flags, a flag $\alpha \rightarrow \beta - \delta$ can be generated in $\mathcal{K}_1 \vee \mathcal{K}_2$ only if either (i) $\beta \leftarrow \delta \in \mathcal{K}_1$ and $\beta \rightarrow \delta \in \mathcal{K}_2$, without loss of generality; or (ii) the edge joining β and δ belongs to a semi-directed cycle in $\mathcal{K}_1 \cup \mathcal{K}_2$. However, neither of these situations can occur because in case (i) $\mathcal{K}_1$ and $\mathcal{K}_2$ do not have the same immoralities and, by lemma 1, in case (ii) $\alpha \rightarrow \delta \in \mathcal{K}_1 \vee \mathcal{K}_2$.

By theorem 5 the subclass $\langle\mathcal{K}\rangle$ is closed with respect to the $\vee$-union operation. It follows that $\mathcal{K}^{\langle\vee\rangle}$ is an NF-CG so that it is the largest CG equivalent to $\mathcal{K}$ with respect to both Markov properties. Furthermore, as well as $\mathcal{G}^{[\![\vee]\!]_c}$ and $\mathcal{D}^{[\vee]}$, it can be obtained as the simple union of all the graphs in the class.

Corollary 1
For any NF-CG $\mathcal{K} = (V, E)$ it holds that

(i) $\mathcal{K}^{\langle\vee\rangle} \in \langle\mathcal{K}\rangle$;
(ii) $\mathcal{K}^{\langle\vee\rangle} = \cup(\mathcal{K}'|\mathcal{K}' \in \langle\mathcal{K}\rangle)$.

Proof. This is an immediate consequence of the fact that $\mathcal{K}^{\langle\vee\rangle}$ is an NF-CG.

With respect to the LWF-Markov property, for a DAG $\mathcal{D} = (V, E)$ we can identify three equivalence classes: the set $[\mathcal{D}]$ of the DAGs equivalent to $\mathcal{D}$, the set $\langle\mathcal{D}\rangle$ of the NF-CGs equivalent to $\mathcal{D}$ and, finally, the set $[\![\mathcal{D}]\!]_c$ of the CGs complex-equivalent to $\mathcal{D}$; note that $[\mathcal{D}] \subseteq \langle\mathcal{D}\rangle \subseteq [\![\mathcal{D}]\!]_c$ and that it is easy to find examples where the two inclusions are strict. To each of these sets is associated a unique CG natural representative of the equivalence class: $\mathcal{D}^{[\vee]}$,$\mathcal{D}^{\langle\vee\rangle}$, and $\mathcal{D}^{[\![\vee]\!]_c}$ respectively, and it holds that $\mathcal{D}^{[\vee]} \subseteq \mathcal{D}^{\langle\vee\rangle} \subseteq \mathcal{D}^{[\![\vee]\!]_c}$. In Fig. 1 an example is given where $\mathcal{D}^{\langle\vee\rangle} \subset \mathcal{D}^{[\![\vee]\!]_c}$ but the first inclusion is in fact an identity. This result was first shown by Studený (2004, Consequence 3) and in the following we propose an alternative proof based on theorem 5.

Theorem 6
For every DAG $\mathcal{D} = (V, E)$ it holds that $\mathcal{D}^{[\vee]} = \mathcal{D}^{\langle\vee\rangle}$.

Proof. Recall that $\mathcal{D}^{[\vee]} = \cup(\mathcal{D}'|\mathcal{D}' \in [\mathcal{D}])$ and $\mathcal{D}^{\langle\vee\rangle} = \cup(\mathcal{K}|\mathcal{K} \in \langle\mathcal{D}\rangle)$. It is straightforward that $\mathcal{D}^{[\vee]} \subseteq \mathcal{D}^{\langle\vee\rangle}$ and it is sufficient to show that $\mathcal{D}^{\langle\vee\rangle} \subseteq \mathcal{D}^{[\vee]}$; i.e. that if $\alpha - \beta \in \mathcal{D}^{\langle\vee\rangle}$ then $\alpha - \beta \in \mathcal{D}^{[\vee]}$. By theorem 5, both $\mathcal{D}^{\langle\vee\rangle}$ and $\mathcal{D}^{[\vee]}$ are NF-CGs and, by *R.3* in appendix A, they have decomposable chain components. Thus, by *R.4* in appendix A, for every $\alpha - \beta \in \mathcal{D}^{\langle\vee\rangle}$ there exist two DAGs, $\mathcal{D}' \in [\mathcal{D}]$ and $\mathcal{D}'' \in [\mathcal{D}]$, such that $\alpha \rightarrow \beta \in \mathcal{D}'$ and $\alpha \leftarrow \beta \in \mathcal{D}''$. Consequently $\alpha - \beta \in \mathcal{D}' \cup \mathcal{D}'' \subseteq \mathcal{D}^{[\vee]}$.

Every DAG $\mathcal{D}$ is an NF-CG with decomposable chain components, furthermore an NF-CG $\mathcal{K}$ is Markov equivalent to some DAG if and only if it has decomposable chain components (see *R.2* and *R.3* in appendix A). As a consequence, by theorem 6 the problem of identifying the essential graph of an equivalence class of DAGs coincides with the problem of identifying the largest CG within a class of NF-CGs with decomposable chain components (see also Studený, 2004).

4. Construction and characterization of $\mathcal{G}^{[\![\vee]\!]_c}$

For a CG $\mathcal{G} = (V, E)$, the graph $\mathcal{G}^{[\![\vee]\!]_c}$ is the largest element of the equivalence class $[\![\mathcal{G}]\!]_c$: the arrows of $\mathcal{G}^{[\![\vee]\!]_c}$ are present with the same orientation in every $\mathcal{G}' \in [\![\mathcal{G}]\!]_c$, but to an arrow of $\mathcal{G}'$ might correspond an undirected edge in $\mathcal{G}^{[\![\vee]\!]_c}$. Here, we propose a stepwise procedure for the construction of $\mathcal{G}^{[\![\vee]\!]_c}$ in which, at every step, all the arrows of one meta-arrow are replaced by undirected edges. In this way we obtain an increasing sequence $\mathcal{G}_0,\dots, \mathcal{G}_r$ of complex-equivalent CGs such that $\mathcal{G}_0 = \mathcal{G}'$, $\mathcal{G}_r = \mathcal{G}^{[\![\vee]\!]_c}$ and $\mathcal{G}_{i-1} \subseteq \mathcal{G}_i$, for all $i = 1,\dots, r$. The existence of such a procedure is guaranteed by the following theorem.

Theorem 7

Let $\mathcal{G} = (V, E)$ and $\tilde{\mathcal{G}} = (\mathrm{V}, \tilde{\mathrm{E}})$ two complex-equivalent CGs such that $\mathcal{G} \subset \tilde{\mathcal{G}}$. Then there exists a finite sequence $\mathcal{G} = \mathcal{G}_0 \subset \mathcal{G}_1 \subset \cdots \subset \mathcal{G}_r = \tilde{\mathcal{G}}$, with $r \geq 1$, of complex-equivalent CGs such that, for all $i = 1,\dots,r$, $\mathcal{G}_i$ can be obtained from $\mathcal{G}_{i-1}$ by replacing the arrows of exactly one meta-arrow with undirected edges.

Proof. See the appendix.

We remark that meta-arrows constitute the natural building-blocks of any stepwise procedure such as the one proposed. Indeed, if $A \Rightarrow B$ is a non-empty meta-arrow of $\mathcal{G}$ and we remove the orientation of some, but not of all, the arrows in $A \Rightarrow B$ then we create a semi-directed cycle and the resulting (larger) graph is not a CG. On the other hand, the removal of the arrowhead of a meta-arrow is a valid step of the procedure if and only if the resulting graph is a CG with the same minimal complexes as $\mathcal{G}$.

Definition 4

For a CG $\mathcal{G} = (V, E)$ let $A \Rightarrow B \neq \emptyset$, with $A, B \in T(\mathcal{G})$, be a meta-arrow of $\mathcal{G}$. We say that the chain components of $A \Rightarrow B$ can be merged *in $\mathcal{G}$ (or in short that $A \Rightarrow B$ can be merged) if the graph $\tilde{\mathcal{G}}$ obtained by replacing every arrow $\alpha \rightarrow \beta \in A \Rightarrow B$ with an undirected edge is a CG complex-equivalent to $\mathcal{G}$.*

For a graph $\mathcal{G}' \in [\![\mathcal{G}]\!]_c\mathcal{G}$ theorem 7 implies that if $\mathcal{G}' \neq \mathcal{G}^{[\![\vee]\!]_c}$, that is $\mathcal{G}' \subset \mathcal{G}^{[\![\vee]\!]_c}$, at least one meta-arrow of $\mathcal{G}'$ can be merged. On the other hand, if some meta-arrow of $\mathcal{G}'$ can be merged, then in $[\![\mathcal{G}]\!]_c$ we can find a graph larger than $\mathcal{G}'$ so that $\mathcal{G}' \neq \mathcal{G}^{[\![\vee]\!]_c}$. We can conclude that $\mathcal{G}' = \mathcal{G}^{[\![\vee]\!]_c}$ if and only if $\mathcal{G}'$ has no meta-arrow whose chain components can be merged. Hence, a crucial point both for the efficiency of the construction procedure and for the successive characterization of $\mathcal{G}^{[\![\vee]\!]_c}$, is the formalization of a simple rule to check whether a meta-arrow can be merged.

We introduce our approach to the problem by deriving a rule to check whether a single arrow of a DAG can be replaced by an undirected edge. Chickering (1995) showed that if the arrow $\alpha \rightarrow \beta$ of a DAG $\mathcal{D}$ is reversed, then the resulting graph is a DAG equivalent to $\mathcal{D}$ if and only if $\mathrm{pa}(\beta)\backslash\{\alpha\} = \mathrm{pa}(\alpha)$ in $\mathcal{D}$. Following this rule, in the DAG (a) in Fig. 1 the only arrow that can be reversed is $\lambda \rightarrow \gamma$. Note that, if we replace such an arrow with the undirected edge $\lambda - \gamma$ we obtain a CG complex-equivalent to $\mathcal{D}$; on the other hand, the arrow $\alpha \rightarrow \beta$ cannot be reversed in $\mathcal{D}$ but it can be replaced by an undirected edge, and the same is true for $\gamma \rightarrow \delta$. We can conclude that the rule $\mathrm{pa}(\beta)\backslash\{\alpha\} = \mathrm{pa}(\alpha)$ is sufficient to assure that $\alpha \rightarrow \beta$ can be merged, but not necessary. It is not difficult to verify that a necessary and sufficient condition is given by the (weaker) rule $\mathrm{pa}(\beta)\backslash\{\alpha\} \subseteq \mathrm{pa}(\alpha)$ and in theorem 8 below it is shown that this rule can be extended to an arbitrary meta-arrow of a CG.

Definition 5
Let $A \Rightarrow B \neq \emptyset$ be a meta-arrow of a CG $\mathcal{G} = (V, E)$. We say that the arrowhead of $A \Rightarrow B$ is insubstantial *in $\mathcal{G}$ if the following conditions are satisfied:*

(*a*) $\mathrm{pa}(B) \cap A$ *is complete;*
(*b*) $\mathrm{pa}(B)\backslash A \subseteq \mathrm{pa}(\alpha)$ *for all* $\alpha \in \mathrm{pa}(B) \cap A$.

If the arrowhead of a meta-arrow is not insubstantial we say that it is *substantial*. We now provide a formal proof of the connection between definitions 4 and 5.

Theorem 8
Let $A \Rightarrow B \neq \emptyset$ be a meta-arrow of a CG $\mathcal{G} = (V, E)$. $A \Rightarrow B$ can be merged in $\mathcal{G}$ if and only if its arrowhead is insubstantial.

Proof. See the appendix.

It should be clear at this point that from any CG $\mathcal{G} = (V, E)$ we can recursively merge meta-arrows until we obtain the largest CG $\mathcal{G}^{[\![\vee]\!]_c}$ in which every meta-arrow has a substantial arrowhead. For instance, for the CG (a) in Fig. 1 we can merge the meta-arrow $\{\gamma\} \Rightarrow \{\delta\}$; from the resulting CG we can merge $\{\lambda\} \Rightarrow \{\delta, \gamma\}$ and, finally, $\{\alpha\} \Rightarrow \{\beta\}$. The resulting largest CG is graph (c) in Fig. 1 whose unique meta-arrow $\{\lambda, \gamma, \delta\} \Rightarrow \{\alpha, \beta\}$ has a substantial arrowhead.

To this construction procedure corresponds the following characterization of $\mathcal{G}^{[\![\vee]\!]_c}$.

Theorem 9
A graph $\tilde{\mathcal{G}}$ is equal to $\mathcal{G}^{[\![\vee]\!]_c}$ for some CG $\mathcal{G}$ if and only if

(i) $\tilde{\mathcal{G}}$ is a CG;
(ii) the arrowhead of every meta-arrow of $\tilde{\mathcal{G}}$ is substantial.

Proof. Necessity. Condition (i) is necessary because for every CG $\mathcal{G}$ the graph $\mathcal{G}^{[\![\vee]\!]_c}$ is a CG by Proposition 5.7 of Frydenberg (1990), and condition (ii) because otherwise, by theorem 8, an insubstantial arrowhead of $\tilde{\mathcal{G}}$ can be merged to produce a CG strictly larger than, and complex-equivalent to, $\tilde{\mathcal{G}}$.

Sufficiency. Condition (i) implies that $[\![\tilde{\mathcal{G}}]\!]_c$ is an equivalence class of CGs and condition (ii) that $\tilde{\mathcal{G}}^{[\![\vee]\!]_c} = \tilde{\mathcal{G}}$ because, by theorems 7 and 8, every $\mathcal{G} \in [\![\tilde{\mathcal{G}}]\!]_c$ such that $\mathcal{G} \subset \mathcal{G}^{[\![\vee]\!]_c}$ has at least one meta-arrow whose arrowhead is insubstantial.

We close this section by underlining that definition 5 provides a very efficient way to check the possible merging of the meta-arrows of a CG. First, such a condition does not require one to look for minimal complexes. A CG can be conveniently represented in a computer program by means of the cliques of its chain components and of the sets of parents of its vertices. In this case no other quantity has to be computed to check the possible merging of meta-arrows.

Secondly, it is *local* in the sense that it involves only the subgraph of $\mathcal{G}$ induced by the chain components A and B and their parents. It follows that, if $\mathcal{G}'$ is obtained from $\mathcal{G}$ by merging $A \Rightarrow B$, then $T(\mathcal{G}')$ can be constructed from $T(\mathcal{G})$ by replacing A and B with $A \cup B$. Furthermore, the meta-arrows of $\mathcal{G}$ that are not destroyed in $\mathcal{G}'$ do not change their status; formally, the arrowhead of a meta-arrow $C \Rightarrow D$ of $\mathcal{G}'$ with $C, D \neq A \cup B$ is substantial in $\mathcal{G}'$ if and only if it is substantial in $\mathcal{G}$. This also implies that we can make the procedure more efficient by simultaneously merging at every step all the insubstantial meta-arrows that have no chain component in common. For instance, in graph (a) in Fig. 1 we can simultaneously merge $\{\alpha\} \Rightarrow \{\beta\}$ and $\{\lambda\} \Rightarrow \{\gamma\}$ or alternatively $\{\alpha\} \Rightarrow \{\beta\}$ and $\{\gamma\} \Rightarrow \{\delta\}$ (in this case it is also possible to simultaneously merge $\{\lambda\} \Rightarrow \{\gamma\}$ and $\{\gamma\} \Rightarrow \{\delta\}$ but this is not true in general for a pair of meta-arrows with a common chain component).

5. Construction and characterization of $\mathcal{K}^{\langle\vee\rangle}$

In this section, we provide a procedure for the construction of $\mathcal{K}^{\langle\vee\rangle}$ and a corresponding characterization of $\mathcal{K}^{\langle\vee\rangle}$. The alternative characterization of essential graphs provided by Studený (2004) turns out to be a special case of our more general framework. The theory developed here follows closely that of section 4, and every result given for NF-CGs is a special case of a corresponding result for CGs. Thus, for an NF-CG $\mathcal{K} = (V, E)$, in parallel with section 4, we consider a procedure in which, at every step, one meta-arrow is merged, obtaining in this way an increasing sequence $\mathcal{K}_0, \ldots, \mathcal{K}_r$ of equivalent NF-CGs such that $\mathcal{K}_0 = \mathcal{K}$, $\mathcal{K}_r = \mathcal{K}^{\langle\vee\rangle}$ and $\mathcal{K}_{i-1} \subseteq \mathcal{K}_i$, for all $i = 1, \ldots, r$. The equivalent of theorem 7 for this case is the following.

Theorem 10
Let $\mathcal{K} = (V, E)$ and $\tilde{\mathcal{K}} = (\mathrm{V}, \tilde{\mathrm{E}})$ be two equivalent NF-CGs such that $\mathcal{K} \subset \tilde{\mathcal{K}}$. Then there exists a finite sequence $\mathcal{K} = \mathcal{K}_0 \subset \mathcal{K}_1 \subset \cdots \subset \mathcal{K}_r = \tilde{\mathcal{K}}$, with $r \geq 1$, of equivalent NF-CGs such that, for all $i = 1,\ldots, r$, $\mathcal{K}_i$ can be obtained from $\mathcal{K}_{i-1}$ by replacing the arrows of exactly one meta-arrow with undirected edges.

Proof. See the appendix.

For NF-CGs the definition of an insubstantial meta-arrow can be stated in a simpler form.

Proposition 2
Let $A \Rightarrow B \neq \emptyset$ be a meta-arrow of an NF-CG $\mathcal{K} = (V, E)$. The arrowhead of $A \Rightarrow B$ is insubstantial if and only if

(*a*) $\mathrm{pa}(B) \cap A$ *is complete;*
(*b*′) $\mathrm{pa}(B) \backslash A \subseteq \mathrm{pa}(A)$.

Proof. (b′) is identical to point (b) of definition 5 by *R.1* in appendix A.

The concept of an insubstantial arrowhead cannot be used in the construction of $\mathcal{K}^{\langle\vee\rangle}$ because if we merge an insubstantial meta-arrow of an NF-CG we can generate a flag. For instance, in graph (a) in Fig. 1 the arrowhead of $\{\alpha\} \Rightarrow \{\beta\}$ is insubstantial but merging its chain components generates a flag. However, the definition of an insubstantial arrowhead can be strengthened to deal with this special case.

Definition 6
Let $A \Rightarrow B \neq \emptyset$ be a meta-arrow of an NF-CG $\mathcal{K} = (V, E)$. We say that the arrowhead of $A \Rightarrow B$ is strongly insubstantial in $\mathcal{K}$ if the following conditions are satisfied:

(a) $\mathrm{pa}(B) \cap A$ *is complete;*
(b) $\mathrm{pa}(B) \backslash A = \mathrm{pa}(A)$.

We remark that definition 6 is, in a different terminology, identical to the definition of *legal merging* given in Studený (2004, Definition 1).

An arrowhead that is not strongly insubstantial is called *weakly substantial*. Clearly a substantial arrowhead is also weakly substantial, but a weakly substantial arrowhead might be insubstantial. We now show a result for NF-CGs equivalent to that given in theorem 8 for CGs (see also Studený, 2004, Statement 1).

Theorem 11
Let $A \Rightarrow B \neq \emptyset$ be a meta-arrow of an NF-CG $\mathcal{K} = (V, E)$ and let $\tilde{\mathcal{G}}$ the graph obtained from $\mathcal{K}$ by merging $A \Rightarrow B$. $\tilde{\mathcal{G}}$ is an NF-CG equivalent to $\mathcal{K}$ if and only if the arrowhead of $A \Rightarrow B$ is strongly insubstantial in $\mathcal{K}$.

Proof. See the appendix.

It follows that we can start from any NF-CG $\mathcal{K} = (V, E)$ and successively merge meta-arrows with strongly insubstantial arrowhead until we obtain $\mathcal{K}^{\langle\vee\rangle}$, in which every meta-arrow has a weakly substantial arrowhead. We can then go on merging meta-arrows with insubstantial arrowheads until we obtain $\mathcal{K}^{[\![\vee]\!]_c}$. For instance, in the NF-CG (a) in Fig. 1 we can first merge $\{\lambda\} \Rightarrow \{\gamma\}$ and then $\{\lambda, \gamma\} \Rightarrow \{\delta\}$ to obtain the NF-CG (b) in Fig. 1 in which all the meta-arrows have substantial arrowheads but for $\{\alpha\} \Rightarrow \{\beta\}$ whose arrowhead is only weakly substantial.

Similarly to theorem 9, the largest CG of an equivalence class of NF-CGs can be characterized as follows.

Theorem 12
A graph $\tilde{\mathcal{G}}$ is equal to $\mathcal{K}^{\langle\vee\rangle}$ for some NF-CG $\mathcal{K}$ if and only if

(i) $\tilde{\mathcal{G}}$ is an NF-CG;
(ii) the arrowhead of every meta-arrow of $\tilde{\mathcal{G}}$ is weakly substantial.

Proof. Necessity. Condition (i) is necessary because for every NF-CG $\mathcal{K}$ the graph $\mathcal{K}^{\langle\vee\rangle}$ is an NF-CG by corollary 1, and condition (ii) because otherwise, by theorem 11, merging a strongly insubstantial arrowhead of $\tilde{\mathcal{G}}$ would produce an NF-CG strictly larger than, and equivalent to, $\tilde{\mathcal{G}}$.

Sufficiency. Condition (i) implies that $\langle\tilde{\mathcal{G}}\rangle$ is an equivalence class of NF-CGs and condition (ii) that $\tilde{\mathcal{G}}^{\langle\vee\rangle} = \tilde{\mathcal{G}}$ because, by theorems 10 and 11, every $\mathcal{K} \in \langle\tilde{\mathcal{G}}\rangle$ such that $\mathcal{K} \subset \tilde{\mathcal{G}}^{\langle\vee\rangle}$ has at least one meta-arrow whose arrowhead is strongly insubstantial.

As a special case of theorem 12 we obtain the alternative characterization of the essential graph recently given in Studený (2004).

Theorem 13
A graph $\tilde{\mathcal{G}}$ is equal to the essential graph $\mathcal{D}^{[\vee]}$ of some DAG $\mathcal{D}$ if and only if

(i) $\tilde{\mathcal{G}}$ is an NF-CG with decomposable chain components;
(ii) the arrowhead of every meta-arrow of $\tilde{\mathcal{G}}$ is weakly substantial.

Proof. See the appendix.

The characterization described in theorem 13 has been (independently) achieved in Studený (2004), but it is worth underlining that our proof is self-contained in the sense that it does not rely on any previous characterizations of the essential graph. In particular, in this paper a central role is played by meta-arrows and single arrows are never explicitly considered.

6. Discussion

The concept of insubstantial arrowhead generalizes a well-known rule to check the reversibility of arrows in a DAG and constitutes a very useful tool to deal with the problem of Markov equivalence in graphical models. Such a concept is very simple to state and to verify because it is based on consolidated graphical model notions such as that of parent and of complete subset; it is very general because it can be applied to different subclasses of CGs and, finally, it is local and therefore efficiently implementable in computer programs.

For a CG $\mathcal{G}$ Andersson *et al*. (2001, Section 7) proposed to characterize the equivalence class of CGs AMP-Markov equivalent to $\mathcal{G}$ by means of a graph $\mathcal{G}^*$, where an arrow occurs in $\mathcal{G}^*$ if and only if it occurs with the same orientation in at least one CG AMP-Markov equivalent to $\mathcal{G}$ but with the opposite orientation in no CG AMP-Markov equivalent to $\mathcal{G}$. Neither a procedure for the construction of $\mathcal{G}^*$ nor a characterization of $\mathcal{G}^*$ is available, but Andersson *et al*. (2001) noticed that if $\mathcal{G}$ is AMP-Markov equivalent to some DAG $\mathcal{D}$, then $\mathcal{G}^* = \mathcal{D}^{[\vee]}$. If a CG $\mathcal{G}$ is AMP-Markov equivalent to some NF-CG $\mathcal{K}$, the largest NF-CG $\mathcal{K}^{\langle\vee\rangle}$ provides a graphical characterization of the class of CGs AMP-Markov equivalent to $\mathcal{G}$ and it is straightforward to see that if $\mathcal{K}$ has decomposable chain components then $\mathcal{K}^{\langle\vee\rangle} = \mathcal{G}^*$.

Acknowledgements

The author wishes to thank Milan Studený, Luca La Rocca and two Referees for useful comments. This paper was revised while the author was a fellow at the Center for Discrete Mathematics and Theoretical Computer Science (DIMACS), Rutgers University, New Jersey. I am grateful to this Institution for the excellent working environment it provides. Financial support was provided by MIUR, Rome (PRIN 03) and DIMACS.

References

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

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

Chickering, D. M. (1995). A transformation characterization of equivalent Bayesian network structures. In *Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence* (eds P. Besnard & S. Hanks), 87–98. Morgan Kaufmann, San Francisco, CA.

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. & Spiegelhalter, D. J. (1999). *Probabilistic networks and expert systems*. Springer-Verlag, New York.
Cox, D. R. & Wermuth, N. (1996). *Multivariate dependencies: models, analysis and interpretation*. Chapman & Hall, London.
Edwards, D. (2000). *Introduction to graphical modelling*, 2nd edn. Springer, New York.
Frydenberg, M. (1990). The chain graph Markov property. *Scand. J. Statist*. **17**, 333–353.
Heckerman, D., Geiger, D. & Chickering, D. M. (1995). Learning Bayesian networks: the combination of knowledge and statistical data. *Machine Learning* **20**, 197–243.
Kiiveri, H., Speed, T. & Carlin, J. (1984). Recursive causal models. *J. Aust. Math. Soc. Ser. A* **36**, 30–52.
Lauritzen, S. L. (1996). *Graphical models*. Oxford University Press, Oxford.
Lauritzen, S. L. & Richardson, T. S. (2002). Chain graphs models and their causal interpretation (with discussion). *J. R. Statist. Soc. Ser. B Stat. Methodol*. **64**, 321–361.
Lauritzen, S. L. & Wermuth, N. (1989). Graphical models for association between variables, some of which are qualitative and some quantitative. *Ann. Statist*. **17**, 31–57.
Madigan, D., Andersson, S. A., Perlman, M. D. & Volinsky, C. M. (1996). Bayesian model averaging and model selection for Markov equivalence classes of acyclic digraphs. *Comm. Statist. Theory Methods* **25**, 2493–2519.
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, CA.
Pearl, J. (1988). *Probabilistic reasoning in intelligent systems*. Morgan Kaufmann, San Mateo.
Pearl, J. (2000). *Causality: models, reasoning and inference*. Cambridge University Press, Cambridge.
Perlman, M. D. (2001). Graphical model search via essential graphs. Technical Report, Department of Statistics, Washington University, USA. Available at http://www.stat.washington.edu/www/research/reports/.
Richardson, T. (2001). *Chain graphs which are maximal ancestral graphs are recursive causal graphs*. Technical Report, Department of Statistics, Washington University, Washington, DC, USA. Available at http://www.stat.washington.edu/www/research/reports/.
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, CA.
Studený, M. (1997). On recovery algorithm for chain graphs. *Internat. J. Approx. Reasoning* **17**, 265–293.
Studený, M. (2004). Characterization of essential graphs by means of an operation of legal component merging. *Internat. J. Uncertain. Fuzziness Knowledge-Based Systems* **12**, 43–62.
Verma, T. & Pearl, J. (1991). Equivalence and synthesis of causal models. In *Proceedings of the 6th Conference on Uncertainty in Artificial Intelligence* (eds P. Bonissone, M. Henrion, L. N. Kanal & J. F. Lemmer), 220–227. North-Holland, Amsterdam.
Verma, T. & 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, SanFrancisco, CA.
Volf, M. & Studený, M. (1999). A graphical characterization of the largest chain graphs. *Int. J. Approx. Reasoning* **20**, 209–236.
Whittaker, J. (1990). *Graphical models in applied multivariate statistics*. Wiley, Chichester.

Received October 8, 2003, in final form June 28, 2004

Alberto Roverato, Department of Social Cognitive and Quantitative Sciences, University of Modena and Reggio Emilia, Via Giglioli Valle, 9, I-42100 Reggio Emilia, Italy.
E-mail: roverato@unimore.it

Appendix

Appendix A: Some properties of chain graphs with no flags

In this appendix, we summarize some useful properties of NF-CGs that are used in proofs throughout the paper.

R.1 Let T be a chain component of an NF-CG $\mathcal{K}$, then all the vertices in T have a common set of parents in $\mathcal{K}$. Formally, $\mathrm{pa}(\alpha) = \mathrm{pa}(T)$ for every $\alpha \in T$.
R.2 Let $\mathcal{K}$ be an NF-CG with decomposable chain components. The directed graph $\mathcal{D}$ obtained from $\mathcal{K}$ by orienting the edges of every subgraph $\mathcal{K}_T$, $T \in T(\mathcal{K})$, in any perfect way is acyclic and Markov equivalent to $\mathcal{K}$.
R.3 If $\mathcal{G}$ is a CG equivalent to some DAG $\mathcal{D}$, then every chain component $T \in T(\mathcal{G})$ is decomposable.
R.4 Let $\mathcal{K}$ be an NF-CG with decomposable chain components. For every $\alpha - \beta \in \mathcal{K}$ there exist two DAGs equivalent to $\mathcal{K}$, $\mathcal{D}'$ and $\mathcal{D}''$, such that $\alpha \rightarrow \beta \in \mathcal{D}'$ and $\alpha \leftarrow \beta \in \mathcal{D}''$.
R.5 If $\mathcal{K}$ is an NF-CG with decomposable chain components, then every $\mathcal{K}' \in \langle\mathcal{K}\rangle$ is an NF-CG with decomposable chain components.

R.1 is a straightforward consequence of the fact that $\mathcal{K}_T$ is connected and that $\mathcal{K}$ has no flags (see also Andersson *et al.*, 2001, p. 49; Perlman, 2001; Studený, 2004, Observation 1). *R.2* can be derived from the 'sufficiency' part of Proposition 4.2 of Andersson *et al.* (1997a). *R.3* is stated in Remark 4.2 of Andersson *et al.* (1997b). *R.4* can be shown by using *R.2* and recalling that every undirected edge $\alpha - \beta \in \mathcal{K}$ belongs to a unique decomposable chain component T and that $\mathcal{K}_T$ admits at least two perfect numberings with first vertex α and β, respectively (see Cowell *et al.*, 1999, p. 52). To these perfect numberings correspond two perfect directed versions, $\mathcal{D}'_T$ and $\mathcal{D}''_T$ of $\mathcal{K}_T$ (see Cowell *et al.*, 1999, p. 52), such that $\alpha \rightarrow \beta \in \mathcal{D}'_T$ and $\alpha \leftarrow \beta \in \mathcal{D}''_T$. *R.5* is an immediate consequence of *R.2* and *R.3*.

Appendix B: Proofs

Proof of proposition 1. (i) To prove this point it is sufficient to show that if there exists a minimal complex of $\bar{\mathcal{G}}$ that is not a minimal complex in $\mathcal{G}$, then there exists a minimal complex of $\mathcal{G}$ that is not a minimal complex in $\bar{\mathcal{G}}$. Assume that $(\gamma, v_1,\dots,v_k, \delta)$, with $S = \{v_1,\dots, v_k\}$, is a minimal complex of $\bar{\mathcal{G}}$ but not of $\mathcal{G}$, so that $\bar{\mathcal{G}}_{S\cup\{\gamma,\delta\}}$ looks like (1) and $\bar{\mathcal{G}}_{S\cup\{\gamma,\delta\}} \neq \mathcal{G}_{S\cup\{\gamma,\delta\}}$. From the facts that $\mathcal{G}_{S\cup\{\gamma,\delta\}} \subseteq \bar{\mathcal{G}}_{S\cup\{\gamma,\delta\}}$ (i.e. every arrow in $\bar{\mathcal{G}}_{S\cup\{\gamma,\delta\}}$ belongs also to $\mathcal{G}_{S\cup\{\gamma,\delta\}}$) and that $\bar{\mathcal{G}}$ and $\mathcal{G}$ have the same skeleton, it follows that $\bar{\mathcal{G}}_S \neq \mathcal{G}_S$. Thus, there exists an undirected edge $v_i - v_{i+1} \in \bar{\mathcal{G}}_S$, $i < k$, that is an arrow in $\mathcal{G}_S$. Let $v_i \leftarrow v_{i+1} \in \mathcal{G}_S$, without loss of generality, and consider the ordered sequence $(\gamma, v_1,\dots,v_i, v_{i+1})$. If either $i = 1$ or $v_j - v_{j+1} \in \mathcal{G}$ for all $j < i$, then such a sequence is a minimal complex in $\mathcal{G}$ that is not in $\bar{\mathcal{G}}$. Otherwise, there exists an undirected edge $v_j - v_{j+1} \in \bar{\mathcal{G}}_S$, $j < i$, that is an arrow in $\mathcal{G}_S$ and we can repeat the above procedure until we find a minimal complex in $\mathcal{G}$ that is not in $\bar{\mathcal{G}}$.

(ii) By (i) it is sufficient to show that every minimal complex of $\mathcal{G}$ is a minimal complex of $\bar{\mathcal{G}}$. If (γ, S, δ) is a minimal complex in $\mathcal{G}$ then it is also a minimal complex in $\tilde{\mathcal{G}}$ so that $\mathcal{G}_{S\cup\{\gamma,\delta\}} = \tilde{\mathcal{G}}_{S\cup\{\gamma,\delta\}}$ looks like (1). Since $\mathcal{G} \subseteq \bar{\mathcal{G}} \subseteq \tilde{\mathcal{G}}$, it follows that $\mathcal{G}_{S\cup\{\gamma,\delta\}} = \bar{\mathcal{G}}_{S\cup\{\gamma,\delta\}} = \tilde{\mathcal{G}}_{S\cup\{\gamma,\delta\}}$ and (γ, S, δ) is a minimal complex in $\bar{\mathcal{G}}$.

Proof of lemma 1. Since $\alpha \rightarrow \beta \in \mathcal{K}_1 \vee \mathcal{K}_2$ then $\alpha \rightarrow \beta \in \mathcal{K}_1 \cup \mathcal{K}_2$. Furthermore, it holds that $\alpha \neq \beta_i$ for all $i = 1\dots, k$, because it is easy to check that if $\alpha = \beta_i$ for some i, then $\alpha \rightarrow \beta$ belongs to a semi-directed cycle in $\mathcal{K}_1 \cup \mathcal{K}_2$ and, by construction, is undirected in $\mathcal{K}_1 \vee \mathcal{K}_2$. It follows that in $\mathcal{K}_1 \cup \mathcal{K}_2$ we have

$$\alpha \rightarrow \beta_0 \leftharpoondown \beta_1 \leftharpoondown \cdots \leftharpoondown \beta_{k-1} \leftharpoondown \beta_k \tag{2}$$

where $k > 2$ and $\beta_i \leftharpoondown \beta_{i+1}$ denotes either $\beta_i - \beta_{i+1}$ or $\beta_i \leftarrow \beta_{i+1}$. Note that, there is no loss of generality in (2) because $\beta_0 = \beta_k = \beta$. By assumption, $\alpha \rightarrow \beta_0$ belongs to $\mathcal{K}_1 \vee \mathcal{K}_2$ and therefore also to $\mathcal{K}_1$ and $\mathcal{K}_2$. With respect to the edge between β_0 and β_1 there are three possible cases: (i) $\beta_0 - \beta_1 \in \mathcal{K}_1$, so that α and β_1 are adjacent because $\mathcal{K}_1$ has no flags; (ii) $\beta_0 \leftarrow \beta_1 \in \mathcal{K}_1$, so that α and β_1 are adjacent because $\beta_0 - \beta_1 \in \mathcal{K}_1 \vee \mathcal{K}_2$ and, by theorem 2, $\mathcal{K}_1$ and $\mathcal{K}_1 \vee \mathcal{K}_2$ have the same immoralities; (iii) $\beta_0 \rightarrow \beta_1 \in \mathcal{K}_1$, but in this case it follows from (2) that either $\beta_0 \leftarrow \beta_1 \in \mathcal{K}_2$ or $\beta_0 - \beta_1 \in \mathcal{K}_2$ and, by applying to $\mathcal{K}_2$ the same arguments used in points (i) and (ii) with respect to $\mathcal{K}_1$, we obtain that α and β_1 are adjacent. We can conclude that α and β_1 are adjacent in $\mathcal{K}_1 \vee \mathcal{K}_2$ and, more precisely, it is only possible that $\alpha \rightarrow \beta_1 \in \mathcal{K}_1 \vee \mathcal{K}_2$ because both $\alpha \leftarrow \beta_1$ and $\alpha - \beta_1$ would produce a semi-directed cycle in $\mathcal{K}_1 \vee \mathcal{K}_2$. We have thus shown that $\alpha \rightarrow \beta_0 \in \mathcal{K}_1 \vee \mathcal{K}_2$ implies $\alpha \rightarrow \beta_1 \in \mathcal{K}_1 \vee \mathcal{K}_2$, and the same reasoning can be used iteratively to show that, for all $i = 1,\ldots, k$, $\alpha \rightarrow \beta_{i-1} \in \mathcal{K}_1 \vee \mathcal{K}_2$ implies $\alpha \rightarrow \beta_i \in \mathcal{K}_1 \vee \mathcal{K}_2$.

Proof of theorem 7. To prove the desired result it is sufficient to show that there exists a CG $\bar{\mathcal{G}}$, complex-equivalent to $\mathcal{G}$, with $\mathcal{G} \subseteq \bar{\mathcal{G}} \subset \tilde{\mathcal{G}}$ and such that $\tilde{\mathcal{G}}$ can be obtained from $\bar{\mathcal{G}}$ by merging one meta-arrow.

Since $\mathcal{G} \subset \tilde{\mathcal{G}}$, we can find a chain component T of $\tilde{\mathcal{G}}$ and a finite sequence $B_1,\ldots,B_k$, $k \geq 2$, of chain components of $\mathcal{G}$, such that $T = B_1 \cup \cdots \cup B_k$. Note that $\tilde{\mathcal{G}}_T$ is a connected undirected graph whereas $\mathcal{G}_T$ is a CG with chain components $B_1,\ldots, B_k$. Let $B = B_j$, for some $1 \leq j \leq k$, be a terminal chain component of $\mathcal{G}_T$, i.e. B has no children in $\mathcal{G}_T$. We can construct a graph $\bar{\mathcal{G}}$ by changing all the (undirected) edges between $T \backslash B$ and B in $\tilde{\mathcal{G}}$ into arrows pointing at B. We now show that the graph $\bar{\mathcal{G}}$ so constructed satisfies all the required properties:

1. $\mathcal{G} \subseteq \bar{\mathcal{G}} \subset \tilde{\mathcal{G}}$. It is straightforward that $\bar{\mathcal{G}} \subseteq \tilde{\mathcal{G}}$, and the strict inclusion $\bar{\mathcal{G}} \subset \tilde{\mathcal{G}}$ can be shown by noticing that $\tilde{\mathcal{G}}_T$ is connected so that at least one undirected edge of $\tilde{\mathcal{G}}$ is an arrow in $\bar{\mathcal{G}}$. The inclusion $\mathcal{G} \subseteq \bar{\mathcal{G}}$ is a consequence of the fact that every edge between $T \backslash B$ and B is directed in $\mathcal{G}$ and, since we have chosen B terminal in $\mathcal{G}_T$, no one of such arrows points at $T \backslash B$.
2. $\bar{\mathcal{G}}$ is a CG. Both $\bar{\mathcal{G}}_T$ and $\bar{\mathcal{G}}_{V \backslash T}$ are CGs: the former by construction and the latter because $\bar{\mathcal{G}}_{V \backslash T} = \tilde{\mathcal{G}}_{V \backslash T}$ is a subgraph of the CG $\tilde{\mathcal{G}}$. Thus a semi-directed cycle in $\bar{\mathcal{G}}$ has to involve at least one vertex in $V \backslash T$ and one vertex in T, but such a semi-directed cycle cannot occur in $\bar{\mathcal{G}}$ because it would be a semi-directed cycle in $\tilde{\mathcal{G}}$.
3. $\bar{\mathcal{G}}$ is complex-equivalent to both $\mathcal{G}$ and $\tilde{\mathcal{G}}$ by point 1 above and proposition 1.
4. $T \backslash B \Rightarrow B$ is a meta-arrow in $\bar{\mathcal{G}}$, i.e. $T \backslash B$ and B are chain components of $\bar{\mathcal{G}}$. It is straightforward that both $\bar{\mathcal{G}}_{T \backslash B}$ and $\bar{\mathcal{G}}_B$ are undirected graphs. Furthermore, $\bar{\mathcal{G}}_B$ is connected because B is a chain component of $\mathcal{G}$. We show that assuming $\bar{\mathcal{G}}_{T \backslash B}$ not connected leads to a contradiction. Assume that $T \backslash B = A_1 \cup A_2$, $A_1 \cap A_2 = \emptyset$, and that A_1 and A_2 are not connected in $\bar{\mathcal{G}}_{T \backslash B}$. Since T is a (connected) chain component of $\tilde{\mathcal{G}}$, then in $\tilde{\mathcal{G}}_T$ there is a path joining any vertex in A_1 to any vertex in A_2 and such a path has nonempty intersection with B. Thus in $\bar{\mathcal{G}}$ we can find two arrows $\alpha_1 \rightarrow \beta$ and $\alpha_2 \rightarrow \beta'$ such that $\alpha_1 \in A_1$, $\alpha_2 \in A_2$ and $\beta, \beta' \in B$. Since $\alpha_1 \not\sim \alpha_2$ such arrows belong to a minimal complex in $\bar{\mathcal{G}}$ that is not a minimal complex of $\tilde{\mathcal{G}}$, but this is not possible because $\tilde{\mathcal{G}}$ and $\bar{\mathcal{G}}$ are complex-equivalent.

Proof of theorem 8. Let $\tilde{\mathcal{G}}$ be the graph obtained from $\mathcal{G}$ by merging $A \Rightarrow B$. We first show that if $A \Rightarrow B$ cannot be merged then its arrowhead is substantial. Consider the case in which $A \Rightarrow B$ cannot be merged because $\tilde{\mathcal{G}}$ is not a CG, i.e. merging $A \Rightarrow B$ creates a semi-directed cycle in $\tilde{\mathcal{G}}$. Note that every semi-directed cycle in $\tilde{\mathcal{G}}$ must involve an undirected edge deriving from an arrow in $A \Rightarrow B$, because otherwise it would be a semi-directed cycle in $\mathcal{G}$. Without loss of generality, for a given semi-directed cycle in $\tilde{\mathcal{G}}$ it is possible to identify $r+2$, with $r \geq 1$, distinct chain components A, B, $T_1,\ldots, T_r$ of $\mathcal{G}$ and a numbering of $T_1,\ldots, T_r$ such that the following meta-arrows are non-empty in $\mathcal{G}$:

$$B \Leftarrow T_r \Leftarrow \cdots \Leftarrow T_1 \Leftarrow A \Rightarrow B. \tag{3}$$

From (3) we can deduce that $T_r \Rightarrow A = \emptyset$ in $\mathcal{G}$ because $T_r \Rightarrow A$ non-empty would imply the existence of a semi-directed cycle in $\mathcal{G}$. We can deduce that there exists a vertex $\delta \in T_r$ such that $\delta \in \mathrm{pa}_{\mathcal{G}}(B)\backslash A$ and $\delta \notin \mathrm{pa}_{\mathcal{G}}(\alpha)$ for all $\alpha \in A$. Hence condition (b) of definition 5 is not satisfied and the arrowhead of $A \Rightarrow B$ is substantial.

Assume that $\tilde{\mathcal{G}}$ is a CG but $A \Rightarrow B$ cannot be merged because $\tilde{\mathcal{G}}$ is not complex-equivalent to $\mathcal{G}$. More precisely, since $\tilde{\mathcal{G}}$ and $\mathcal{G}$ have the same skeleton, by proposition 1 $\mathcal{G}$ has a minimal complex $(\gamma, v_1,\dots, v_k, \delta)$, with $S = \{v_1,\dots, v_k\}$, that is not a minimal complex in $\tilde{\mathcal{G}}$. In other words, $\mathcal{G}_{S\cup\{\gamma,\delta\}}$ looks like (1) and $\mathcal{G}_{S\cup\{\gamma,\delta\}} \neq \tilde{\mathcal{G}}_{S\cup\{\gamma,\delta\}}$. Recall that $\tilde{\mathcal{G}}$ is larger then $\mathcal{G}$, so that every undirected edge of $\mathcal{G}$ is an undirected edge in $\tilde{\mathcal{G}}$, and in particular $\mathcal{G}_S = \tilde{\mathcal{G}}_S$. Consequently, it holds that $\gamma - v_1 \in \tilde{\mathcal{G}}$, without loss of generality, and this implies, in turn, that $\gamma \rightarrow v_1 \in A \Rightarrow B$, $\{v_1,\dots, v_k\} \subseteq B$, $\gamma \in \mathrm{pa}_{\mathcal{G}}(B) \cap A$ and $\delta \in \mathrm{pa}_{\mathcal{G}}(B)$. We can conclude that the arrowhead of $A \Rightarrow B$ is substantial because (i) if $\delta \in A$ then $\gamma, \delta \in \mathrm{pa}_{\mathcal{G}}(B) \cap A$ and condition (a) of definition 5 is not satisfied because $\gamma \not\sim \delta$; (ii) if $\delta \notin A$ then $\delta \in \mathrm{pa}_{\mathcal{G}}(B)\backslash A$ but $\delta \notin \mathrm{pa}_{\mathcal{G}}(\gamma)$ because $\gamma \not\sim \delta$, and condition (b) of definition 5 is not satisfied.

We now show that if the arrowhead of $A \Rightarrow B$ is substantial then $A \Rightarrow B$ cannot be merged. Assume that condition (a) of definition 5 is not satisfied; that is there exist two vertices $\gamma, \delta \in \mathrm{pa}(B) \cap A$ such that $\gamma \not\sim \delta$. Hence, (γ, B, δ) is a complex in $\mathcal{G}$ and it can be found a subset $B' \subseteq B$ such that (γ, B', δ) is a minimal complex in $\mathcal{G}$. However, this is not a complex in $\tilde{\mathcal{G}}$ because $\tilde{\mathcal{G}}_{B'\cup\{\gamma,\delta\}}$ is an undirected graph, so that either $\tilde{\mathcal{G}}$ is not a CG or $\tilde{\mathcal{G}}$ and $\mathcal{G}$ are not complex-equivalent. Assume now that condition (b) of definition 5 is not satisfied; i.e. there exist two vertices $\gamma \in \mathrm{pa}(B) \cap A$ and $\delta \in \mathrm{pa}(B)\backslash A$ such that $\delta \notin \mathrm{pa}(\gamma)$. There are two cases to consider: (i) $\gamma \not\sim \delta$ and (ii) $\gamma \rightarrow \delta \in \mathcal{G}$ (recall that $\gamma - \delta \notin \mathcal{G}$ because $\delta \notin A$). Since $\gamma, \delta \in \mathrm{pa}(B)$, in case (i) it holds that (γ, B, δ) is a complex in $\mathcal{G}$ and it can be found a subset $B' \subseteq B$ such that (γ, B', δ) is a minimal complex. However, the latter is not a complex in $\tilde{\mathcal{G}}$ because $\tilde{\mathcal{G}}_{B'\cup\{\gamma\}}$ is an undirected graph, so that either $\tilde{\mathcal{G}}$ is not a CG or $\mathcal{G}$ and $\tilde{\mathcal{G}}$ are not equivalent. In case (ii), δ is a child of A but also a parent of B so that merging $A \Rightarrow B$ creates a semi-directed cycle in $\tilde{\mathcal{G}}$. Thus merging a meta-arrow with a substantial arrowhead either generates a semi-directed cycle or destroys a minimal complex, and the proof is complete.

Proof of Theorem 10. Consider the graph $\bar{\mathcal{G}}$ given in the proof of theorem 7. To prove the desired result it is sufficient to show that if both $\mathcal{G} = \mathcal{K}$ and $\tilde{\mathcal{G}} = \tilde{\mathcal{K}}$ have no flags then $\bar{\mathcal{G}}$ has no flags. Assume that $\alpha \rightarrow \beta - \delta$ is a flag in $\bar{\mathcal{G}}$. From the fact that $\bar{\mathcal{G}} \subset \tilde{\mathcal{K}}$ and that $\tilde{\mathcal{K}}$ has no flags, it follows that $\alpha - \beta - \delta \in \tilde{\mathcal{K}}$. This implies that $\alpha \in T\backslash B$ and $\beta \in B$, and that α, β and δ belong to the same chain component in $\tilde{\mathcal{K}}$ and, more precisely, that $\alpha, \beta, \delta \in T$. From the fact that $\mathcal{K} \subseteq \bar{\mathcal{G}}$ and that $\mathcal{K}$ has the same immoralities as $\tilde{\mathcal{K}}$ and no flags, it follows that $\alpha \rightarrow \beta \rightarrow \delta \in \mathcal{K}$. We can conclude that δ is a child of $\beta \in B$ in $\mathcal{K}_T$, but this is not possible because we have chosen B to be terminal in $\mathcal{K}_T$.

Proof of theorem 11. By theorem 8 $\tilde{\mathcal{G}}$ is a CG equivalent to $\mathcal{K}$ if and only if the arrowhead of $A \Rightarrow B$ is insubstantial. Recall that if $\alpha \rightarrow \beta \in A \Rightarrow B$ then $\mathrm{pa}(\alpha) = \mathrm{pa}(A)$ and $\mathrm{pa}(\beta) = \mathrm{pa}(B)$ by *R.1*.

We first show that if the arrowhead of $A \Rightarrow B$ is strongly insubstantial then $\tilde{\mathcal{G}}$ has no flags. More precisely, we show that, for $\alpha \rightarrow \beta \in A \Rightarrow B$, assuming either $\delta \rightarrow \alpha - \beta \in \tilde{\mathcal{G}}$ or $\alpha - \beta \leftarrow \delta \in \tilde{\mathcal{G}}$ leads to the contradiction that the arrowhead of $A \Rightarrow B$ is weakly substantial. If $\delta \rightarrow \alpha - \beta \in \tilde{\mathcal{G}}$, then $\delta \in \mathrm{pa}(A)$ but $\delta \notin \mathrm{pa}(B)$ so that $\mathrm{pa}(B)\backslash A \neq \mathrm{pa}(A)$ and the arrowhead of $A \Rightarrow B$ is weakly substantial. Assume now that $\alpha - \beta \leftarrow \delta \in \tilde{\mathcal{G}}$. In this case

$\alpha \rightarrow \beta \leftarrow \delta \in \mathcal{K}$ so that $\mathcal{K}$ and $\tilde{\mathcal{G}}$ are not equivalent and the arrowhead of $A \Rightarrow B$ is substantial and therefore weakly substantial.

We now show that if the arrowhead of $A \Rightarrow B$ is insubstantial but not strongly insubstantial, i.e. $\mathrm{pa}(B)\backslash A \subset \mathrm{pa}(A)$ by proposition 2, then $\tilde{\mathcal{G}}$ has some flag. In this case there exists a vertex $\delta \notin \mathrm{pa}(B)\backslash A$ such that $\delta \in \mathrm{pa}(A)$. This implies that, for every $\alpha \rightarrow \beta \in A \Rightarrow B$, it holds that $\delta \rightarrow \alpha \rightarrow \beta \in \mathcal{K}$ with $\delta \not\sim \beta$ so that merging $\alpha \rightarrow \beta$ produces at least one flag.

Proof of theorem 13. By *R.5* in appendix A, if an NF-CG $\mathcal{K}$ has decomposable chain components then every element of $\langle\mathcal{K}\rangle$ has decomposable chain components. Conversely, if $\mathcal{K}$ has some non-decomposable chain components, then every element of $\langle\mathcal{K}\rangle$ has some non-decomposable chain component. As a consequence, by theorem 12, a graph $\tilde{\mathcal{G}}$ is equal to $\mathcal{K}^{\langle\vee\rangle}$ for some NF-CG with decomposable chain components $\mathcal{K}$ if and only if it satisfies (i) and (ii). Every DAG is an NF-CG with decomposable chain components and, by *R.2* in appendix A, every NF-CG with decomposable chain components is Markov equivalent to some DAG so that the previous sentence can be restated as: a graph $\tilde{\mathcal{G}}$ is equal to $\mathcal{D}^{\langle\vee\rangle}$ for some DAG $\mathcal{D}$ if and only if it satisfies (i) and (ii). The result follows because, by theorem 6, $\mathcal{D}^{\langle\vee\rangle} = \mathcal{D}^{[\vee]}$.