

Milan Studený; Alberto Roverato; Šárka Štěpánová

Two operations of merging and splitting components in a chain graph

Kybernetika, Vol. 45 (2009), No. 2, 208--248

Persistent URL: <http://dml.cz/dmlcz/140074>

Terms of use:

© Institute of Information Theory and Automation AS CR, 2009

Institute of Mathematics of the Academy of Sciences of the Czech Republic provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these *Terms of use*.



This paper has been digitized, optimized for electronic delivery and stamped with digital signature within the project *DML-CZ: The Czech Digital Mathematics Library* <http://project.dml.cz>

TWO OPERATIONS OF MERGING AND SPLITTING COMPONENTS IN A CHAIN GRAPH

MILAN STUDENÝ, ALBERTO ROVERATO AND ŠÁRKA ŠTĚPÁNOVÁ

In this paper we study two operations of merging components in a chain graph, which appear to be elementary operations yielding an equivalent graph in the respective sense. At first, we recall basic results on the operation of *feasible merging* components, which is related to classic LWF (Lauritzen, Wermuth and Frydenberg) Markov equivalence of chain graphs. These results are used to get a graphical characterisation of *factorisation equivalence* of classic chain graphs. As another example of the use of this operation, we derive some important invariants of LWF Markov equivalence of chain graphs. Last, we recall analogous basic results on the operation of *legal merging* components. This operation is related to the so-called *strong equivalence* of chain graphs, which includes both classic LWF equivalence and alternative AMP (Andersson, Madigan and Perlman) Markov equivalence.

Keywords: chain graph, essential graph, factorisation equivalence, feasible merging components, legal merging components, strong equivalence

AMS Subject Classification: 62H05, 68T30, 05C90

1. INTRODUCTION

Graphical models are statistical models of conditional independence structure specified by graphs, whose nodes correspond to variables. Traditionally, undirected graphs and acyclic directed graphs have been used to describe probabilistic dependencies between variables [10]. In the mid-1980s, Lauritzen and Wermuth [7, 8] introduced the class of chain graphs which involves both these traditional classes of graphs. They also presented an original interpretation of chain graphs in terms of conditional independence. Basic mathematical results on chain graphs under this interpretation were achieved by Frydenberg [6]. He introduced graphical characterisation of Markov equivalent chain graphs, that is, graphs describing the same statistical model. Moreover, he showed that every equivalence class of chain graphs has a distinguished member, called the *largest chain graph*.

Two different graphical characterisations of graphs that are the largest chain graphs were given in the late 1990s [14, 19]. Another elegant characterisation of these graphs has recently been found in [11]. This characterisation has close connection to a certain elementary operation with chain graphs. This is a special operation of

merging components in a graph, which yields an equivalent graph. The operation seems to be a useful tool for proving results on chain graphs under the original interpretation. That's why we have decided to give it a special name, namely the operation of *feasible merging components*. Note that it appears to be equivalent to another operation with chain graphs, formerly introduced in [14], which was, however, formulated in a much more complicated way.

Another important operation with chain graphs was introduced in connection with the problem of representing an equivalence class of acyclic directed graphs. Graphical characterisation of equivalent acyclic directed graphs was presented by Verma and Pearl [18]. Since an equivalence class of these graphs cannot be represented by its distinguished member, the researchers in this area came up with the idea to describe it by a special chain graph, called the *essential graph*. This terminology is inspired by Andersson, Madigan and Perlman [2], who gave a graphical characterisation of graphs that are essential graphs. Recently, another graphical characterisation of essential graphs has independently been found in [15] and in [11]. This characterisation is closely related to the other elementary operation with chain graphs. That operation was originally introduced for a special class of chain graphs, namely, the class of chain graphs without flags. It is also an operation of merging components in a chain graph; we have decided to name it *legal merging components*.

The operation of legal merging has been extended in [12] to all chain graphs in connection with the problem of representing an equivalence class of chain graphs under an alternative interpretation. The idea of an alternative interpretation of chain graphs in terms of conditional independence was brought by Andersson, Madigan and Perlman [1]. To distinguish two different interpretations of chain graphs, they proposed to use the acronym LWF to indicate the original interpretation and the acronym AMP to indicate their alternative interpretation. They also gave a graphical characterisation of AMP Markov equivalent chain graphs [4], that is, graphs describing the same statistical model under the alternative interpretation. The question of representing an AMP equivalence class of chain graphs led to an analysis of its internal structure in [12]. It has been found that every AMP equivalence class breaks into smaller equivalence classes of special *strong equivalence* of chain graphs, every strong equivalence class has the largest graph, and there exists a distinguished strong equivalence class within an AMP equivalence class. In particular, the largest graph in the distinguished strong equivalence class, called the *largest deflagged graph*, can serve as a natural representative of the whole AMP equivalence class. An algorithm from [12], for getting the largest graph in a strong equivalence class of chain graphs, is based on the extended operation of legal merging components.

1.1. Structure of the paper

This paper has two aims. The first intention is to bring a unifying perspective on elementary operations of merging and splitting components in a chain graph which yield an equivalent graph (in the corresponding sense). However, the paper is not solely a review of former results. The second goal is to present some new results, namely the results whose proofs are substantially based on the idea of an elementary operation with a chain graph yielding an equivalent chain graph. This

hopefully demonstrates the applicability of those elementary operations.

Here is an overview of the paper. In Section 2, we recall various elementary concepts and well-known facts in this area. These are needed to understand the advanced concepts introduced in the sections that follow. Section 3 is devoted to the operation of feasible merging components in a chain graph. We recall basic results on this operation, show that the operation is equivalent to the application of the so-called pool-component rule from [14], and introduce an inverse operation of *feasible splitting a component* in a chain graph.

Section 4 deals with the concept of *factorisation* of density of a probability measure with respect to a chain graph and with the respective *factorisation equivalence* of chain graphs. First, in Section 4.1, we analyze the definition of factorisation from [9] and find out that it can be interpreted in two different ways. This leads to two different concepts of factorisation with respect to a chain graph. We argue that one of them is the right one and use it as the basis for our definition of factorisation equivalence of chain graphs. A by-product of our analysis is the observation that the validity of the recursive factorisation condition does not depend on the choice of versions of conditional densities. Second, in Section 4.2, a graphical characterisation of factorisation equivalence relative to non-trivial sample spaces is given. Actually, it coincides with the graphical characterisation of LWF Markov equivalence, and the proof of this result is substantially simplified by the idea of feasible merging components in a chain graph.

In Section 5, a few characteristics of chain graphs are defined. These characteristics, introduced in [16], are important from the point of view of an algebraic approach [17] to the description of Bayesian network models, that is, graphical models ascribed to acyclic directed graphs. Using the idea of the operation of feasible merging components, we prove that these characteristics are invariable within LWF Markov equivalence classes of chain graphs.

Section 6 is an overview of results on the operation of legal merging components. We recall the concept of strong equivalence of chain graphs and basic results on legal merging components. We also define an inverse operation of *legal splitting a component* in a chain graph, specialise the operation of legal merging components for chain graphs without flags, and relate this operation to the concept of an essential graph. The Appendix contains four technical proofs of important results from the main sections.

To help the reader to distinguish new contributions of the paper from former re-formulated results we list here the results whose proofs have not been published (in a reviewed paper) before (as far as we know). This is the minor observation in Proposition 8, then most of the statements in Section 4, in particular Proposition 10, Example 2, Proposition 15 and Theorem 16, and all results in Section 5. Section 6 recalls former results: its aim is to pinpoint the analogy with the results from Section 3 to help the reader to understand the shared features of both elementary operations with chain graphs.

2. BASIC CONCEPTS

In this section, the definitions of elementary concepts and some basic facts are recalled.

2.1. Graphical concepts

Graphs considered in this paper have a finite non-empty set N as the set of *nodes*. A *line* over N (= an undirected edge) is a subset of N of cardinality two, that is, an unordered pair $\{a, b\}$ of distinct nodes. An *arrow* over N (= a directed edge) is an ordered pair (a, b) of distinct nodes. By a *hybrid graph over N* we will understand a triplet $H = (N, \mathcal{L}(H), \mathcal{A}(H))$, where $\mathcal{L}(H)$ is a set of lines over N and $\mathcal{A}(H)$ is a set of arrows over N such that there is no multiple edge in H , which means that

$$\text{whenever } (a, b) \in \mathcal{A}(H) \text{ then } (b, a) \notin \mathcal{A}(H) \text{ and } \{a, b\} \notin \mathcal{L}(H).$$

We will write $a \rightarrow b$ in H or $b \leftarrow a$ in H to denote $(a, b) \in \mathcal{A}(H)$ and $a - b$ in H to denote $\{a, b\} \in \mathcal{L}(H)$. A pair $[a, b]$ of distinct nodes will be called an *edge in H* if either $a - b$ in H or $a \rightarrow b$ in H or $b \rightarrow a$ in H . Evidently, $[a, b]$ is an edge in H iff $[b, a]$ is an edge in H . If this is the case, we say that a and b are *adjacent* by an edge in H . If $a \rightarrow b$ in H then we also say that a is a *parent* of b in H or that b is a *child* of a in H . If $a - b$ in H then we say that a and b are *neighbours* in H . The set of parents of nodes in a set $C \subseteq N$ will be denoted by $\text{pa}_H(C)$, the set of its children by $\text{ch}_H(C)$, and set of its neighbours by $\text{ne}_H(C)$.

Given $\emptyset \neq A \subseteq N$, the *induced subgraph* of H for A , denoted by H_A , is the graph $(A, \mathcal{L}(H) \cap \mathcal{P}(A), \mathcal{A}(H) \cap (A \times A))$, where $\mathcal{P}(A) \equiv \{B; B \subseteq A\}$ is the power set of A . A set of nodes $C \subseteq N$ is *connected* in H if, for every $a, b \in C$, there exists an *undirected path* in C connecting them, that is, a sequence of distinct nodes $a = c_1, \dots, c_\ell = b$, $\ell \geq 1$ such that $c_i \in C$ and $c_i - c_{i+1}$ in H for $i = 1, \dots, \ell - 1$. A (connectivity) *component* in H is a maximal connected set in H with respect to set inclusion. Evidently, components in a hybrid graph are pairwise disjoint. A set $K \subseteq N$ is *complete* in H if, $\forall a, b \in K$ $a \neq b$, one has $a - b$ in H . Maximal complete sets in H with respect to set inclusion are called the *cliques* of H .

A *chain graph* is a hybrid graph H whose components can be ordered into a chain, that is, a sequence C_1, \dots, C_m , $m \geq 1$ such that

- if $a - b$ in H then $a, b \in C_i$ for some i ,
- if $a \rightarrow b$ in H then $a \in C_i, b \in C_j$ with $i < j$.

In particular, if there is an edge between two nodes of a component in a chain graph, then it is a line. Another consequence of this definition is that every chain graph has a *terminal component*, that is, a component C with $\text{ch}_H(C) = \emptyset$. The above definition is one of several possible definitions of a chain graph. Another equivalent definition is that a chain graph is a hybrid graph H without semi-directed cycles – see Lemma 2.1 in [14]. Recall that a semi-directed cycle is a sequence of nodes $d_1, \dots, d_\ell, d_{\ell+1} = d_1$, $\ell \geq 3$ such that d_1, \dots, d_ℓ are distinct, $d_1 \rightarrow d_2$ in H and $\forall i = 2, \dots, \ell$ either $d_i \rightarrow d_{i+1}$ in H or $d_i - d_{i+1}$ in H .

An important special case of a chain graph is an *undirected graph*, that is, a hybrid graph F without arrows: $\mathcal{A}(F) = \emptyset$. Of course, an undirected graph is determined by the collection of its cliques. The *underlying graph* of a chain graph H over N is an undirected graph H^u over N such that $a - b$ in H^u if and only if $[a, b]$ is an edge in H . Given an undirected graph F over N and $A, B, C \subseteq N$ three pairwise disjoint sets of nodes, we say that C *separates* A from B in F if every undirected path c_1, \dots, c_ℓ , $\ell \geq 1$ connecting a node $a = c_1$ in A with a node $b = c_\ell$ in B contains an internal node c_i , $1 < i < \ell$ in C . Given $A, B \subseteq N$ with $A \cup B = N$, we say that F *decomposes* into F_A and F_B if $A \cap B$ is complete in F and separates $A \setminus B$ from $B \setminus A$ in F ; that is, there is no edge in F connecting a node in $A \setminus B$ and a node in $B \setminus A$. The decomposition is *proper* if $A \setminus B \neq \emptyset \neq B \setminus A$.

An important subclass of the class of undirected graphs is the class of decomposable graphs. There is a number of equivalent definitions of this concept – for details see § 2.1 in [9]. The basic definition is recursive: the class of decomposable graphs over N is introduced by means of the classes of decomposable graphs over proper subsets of N . An undirected graph F over N is *decomposable* if either N is a complete set in F or there exists a proper decomposition of F into graphs F_A and F_B which are themselves decomposable. Another equivalent definition of a decomposable graph is the condition that the collection of its cliques can be ordered into a sequence K_1, \dots, K_r , $r \geq 1$ satisfying the *running intersection property*:

$$\forall i > 2 \quad \exists k < i \quad K_i \cap \left(\bigcup_{j < i} K_j \right) \subseteq K_k.$$

The sets $S_i = K_i \cap \left(\bigcup_{j < i} K_j \right)$, $i = 2, \dots, r$ are called *separators*. The point is that the class of separators of a decomposable graph does not depend on the choice of a sequence of its cliques satisfying the running intersection property. Moreover, the *multiplicity* of a separator, that is, the number of its occurrences in the sequence S_2, \dots, S_r , also does not depend on that choice – see Lemma 7.2 in [17]. Note that decomposable graphs are also known under another name: *triangulated graphs*. A useful observation concerning them is that every induced subgraph of a decomposable graph is again a decomposable graph.

Another common special case of a chain graph is an acyclic directed graph. A hybrid graph G is *directed* if it has no lines: $\mathcal{L}(G) = \emptyset$. A directed graph is *acyclic* if it has no directed cycle, that is, no sequence $d_1, \dots, d_\ell, d_{\ell+1} = d_1$, $\ell \geq 3$, such that d_1, \dots, d_ℓ are distinct and $d_i \rightarrow d_{i+1}$ in G for $i = 1, \dots, \ell$. A well-known fact is that a directed graph is acyclic iff there exists a total ordering a_1, \dots, a_n , $n \geq 1$ of all nodes of N which is *consistent* with the direction of arrows, that is, whenever $a_i \rightarrow a_j$ in G then $i < j$.

We also need some special concepts related to chain graphs. Let H be a chain graph and C a component in H . By the *closure graph* for C we will understand an undirected graph $\bar{H}(C)$ over $C \cup \text{pa}_H(C)$ such that $a - b$ in $\bar{H}(C)$ if either $[a, b]$ is an edge in $H_{C \cup \text{pa}_H(C)}$ or a, b are distinct nodes of $\text{pa}_H(C)$. An *immorality* in a chain graph H is an induced subgraph of H shown in the left-hand picture of Figure 1, that is, $a \rightarrow c \leftarrow b$ where a, b, c are distinct nodes and $[a, b]$ is not an

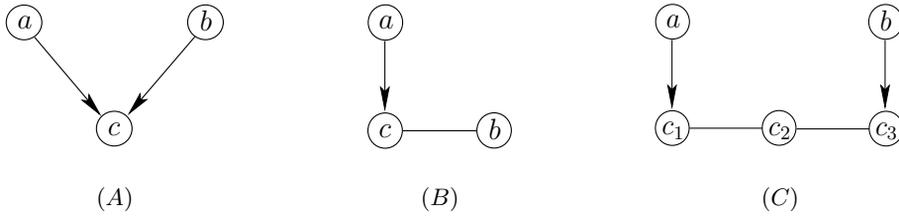


Fig. 1. An immorality, (A), a flag, (B), and a complex, (C).

edge in H . A *flag* in a chain graph H is an induced subgraph of H shown in the middle picture of Figure 1, that is, $a \rightarrow c - b$ where a, b, c are distinct nodes and $[a, b]$ is not an edge in H . A *complex* in H is an induced subgraph of H of the form $a \rightarrow c_1 - \dots - c_s \leftarrow b$, $s \geq 1$, which means that no other edges between nodes $\{a, b, c_1, \dots, c_s\}$ are present in H . An example of a complex is shown in the right-hand picture of Figure 1. Of course, an immorality in H is a special case of a complex in H , namely if $s = 1$. Given a chain graph H , we say that a node a is an *ancestor* of a node d in H , or that d is a *descendant* of a , if there exists a *descending path* in H from a to d , that is, a sequence of nodes $a = d_1, \dots, d_\ell = d$, $\ell \geq 1$ such that d_1, \dots, d_ℓ are distinct and $\forall i = 1, \dots, \ell - 1$ either $d_i \rightarrow d_{i+1}$ in H or $d_i - d_{i+1}$ in H . The set of ancestors of nodes in $A \subseteq N$ will be denoted by $\text{an}_H(A)$, the set of their descendants by $\text{ds}_H(A)$. Observe that $A \subseteq \text{an}_H(A)$ and $A \subseteq \text{ds}_H(A)$.

Let G and H be chain graphs with the same underlying graph. We say that H is *larger* than G if $a \rightarrow b$ in H implies $a \rightarrow b$ in G . It follows from this condition that, whenever $a - b$ in G , then $a - b$ in H ; in particular, H has at least as many lines as G . We will write $H \geq G$ to denote that H is larger than G .

2.2. Discrete probability measures

A discrete probability measure over N is specified as follows. For every $i \in N$, a non-empty finite *individual sample space* X_i is given. The *configuration* of values for a set $A \subseteq N$ is any list $[x_i]_{i \in A}$ where $x_i \in X_i$ for every $i \in A$. The set of all configurations for A will be denoted by the symbol X_A . In particular, the set of configurations for N is just the Cartesian product $\prod_{i \in N} X_i$, called the *joint sample space*, and the set of configurations for the empty set \emptyset is the singleton containing the empty list $[x_i]_{i \in \emptyset}$, denoted by \star . Given a configuration $x = [x_i]_{i \in A}$ for A and $B \subseteq A$, the *marginal configuration* (of x) for B , obtained by restricting the list x to items x_i that correspond to $i \in B$, will be denoted by x_B . Given disjoint sets $A, C \subseteq N$, a configuration y for A , and a configuration z for C , the symbol $[y, z]$ will denote the configuration for $A \cup C$ obtained by concatenating the lists y and z .

A discrete probability measure over N is a non-negative measure P on the joint sample space X_N , endowed with the σ -algebra $\mathcal{P}(X_N)$, such that $P(X_N) = 1$. It is

determined by its *density*, which is a function

$$p : X_N \rightarrow [0, 1] \quad \text{such that} \quad \sum \{p(x); x \in X_N\} = 1.$$

Given $A \subseteq N$, the respective *marginal density* $p_A : X_A \rightarrow [0, 1]$ is defined as follows:

$$p_A(y) = \sum \{p(x); x_A = y, x \in X_N\} \quad \text{for } y \in X_A.$$

Observe that $p_N = p$ and $p_\emptyset(\star) = 1$. In the pages that follow, we will often utilise the following *vanishing principle for marginal densities* of a discrete measure P over N :

$$\forall x \in X_N \quad \forall C \subseteq B \subseteq N \quad p_C(x_C) = 0 \Rightarrow p_B(x_B) = 0.$$

Another important concept is that of conditional density. Given disjoint sets $A, C \subseteq N$ and a density $p : X_N \rightarrow [0, 1]$, by a *version of conditional density* for A given C , we will understand any function of two variables $p_{A|C} : X_A \times X_C \rightarrow (-\infty, \infty)$ such that

$$\forall x \in X_N \quad p_{A \cup C}(x_{A \cup C}) = p_C(x_C) \cdot p_{A|C}(x_A|x_C). \tag{1}$$

Observe that we denote the value of $p_{A|C}$ for configurations $y \in X_A$ and $z \in X_C$ by the symbol $p_{A|C}(y|z)$. It may be the case that several versions of conditional density $p_{A|C}$ exist.¹ By a *zero-version* of conditional density for A given C , we will understand the function $p_{A|C}^0$ defined as follows:

$$p_{A|C}^0(y|z) = \begin{cases} \frac{p_{A \cup C}([y,z])}{p_C(z)} & \text{if } p_C(z) > 0, \\ 0 & \text{if } p_C(z) = 0, \end{cases} \quad \text{where } y \in X_A, z \in X_C.$$

Nevertheless, some authors prefer to restrict their attention to *regular versions* of conditional densities, that is, to functions $p_{A|C} : X_A \times X_C \rightarrow [0, 1]$ satisfying (1) such that, moreover, for every $z \in X_C$, one has

$$p_{A|C}(y|z) \geq 0 \quad \text{for every } y \in X_A, \quad \text{and} \quad \sum \{p_{A|C}(y|z); y \in X_A\} = 1.$$

Another relevant concept is that of conditional independence. Every conditional independence statement concerning a probability measure over N is ascribed to a *disjoint triplet over N* , by which we mean an ordered triplet of pairwise disjoint subsets $A, B, C \subseteq N$. We will use a special notation $\langle A, B|C \rangle$ to indicate a distinguished role of the third component C in the triplet, which is the conditioning set. The collection of all such triplets over N will be denoted by the symbol $\mathcal{T}(N)$. Given a probability measure P over N with a density p on X_N and a triplet $\langle A, B|C \rangle \in \mathcal{T}(N)$, we say that A is *conditionally independent* of B given C with respect to P , and write $A \perp\!\!\!\perp B | C [P]$ if

$$\forall x \in X_N \quad p_{A \cup B \cup C}(x_{A \cup B \cup C}) \cdot p_C(x_C) = p_{A \cup C}(x_{A \cup C}) \cdot p_{B \cup C}(x_{B \cup C}).$$

¹We show later in Section 4.1 that it is important to be aware of the distinction between versions of conditional density. This is because, in certain factorization conditions ascribed to a chain graph, it does matter which of the versions of conditional density are considered!

2.3. Classic interpretation of chain graphs

Every chain graph over N can be interpreted as a certain statistical model, that is, a class of probability measures over N . In this paper, we limit ourselves to the discrete case, that is, we only consider statistical models consisting of discrete probability measures. The basic statistical interpretation of chain graphs is in terms of conditional independence. Note that the other interpretation in terms of factorisation will be mentioned in Section 4 (and an alternative conditional independence interpretation in Section 2.5).

In the case of (both classic and alternative) conditional independence interpretation, every chain graph H over N defines through the respective graphical separation criterion a class of disjoint triplets over N , namely the class of triplets *represented in H* . More specifically, in the case of classic LWF interpretation of chain graphs one can use the *moralization criterion* from [9].

This criterion has three steps. Given a chain graph H and a triplet $\langle A, B|C \rangle \in \mathcal{T}(N)$, the first step is to restrict oneself to the set of *ancestors* of nodes in $A \cup B \cup C$. The result of the first step is the induced subgraph $G \equiv H_{\text{an}_H(A \cup B \cup C)}$. The second step is the replacement of the chain graph G with its *moral graph* G^{mor} . This is an undirected graph over the same set of nodes $\text{an}_H(A \cup B \cup C)$ such that $a - b$ in G^{mor} if either $[a, b]$ is an edge in G or there exists a component C of G such that a, b are distinct nodes of $\text{pa}_G(C)$. The third step is application of a well-known separation criterion for undirected graphs. We test whether C separates A from B in G^{mor} ; if this is the case then we say that the triplet $\langle A, B|C \rangle$ is *represented in H according to the moralization criterion*.

Given a joint sample space \mathbf{X}_N , the class of *LWF Markovian measures* with respect to a chain graph H over N is introduced as the class of probability measures on \mathbf{X}_N satisfying the conditional independence statements ascribed to triplets represented in H according to the moralization criterion. Thus, the corresponding statistical model is the class of LWF Markovian probability measures on \mathbf{X}_N .

A related concept is that of LWF Markov equivalence. Suppose that finite sample spaces \mathbf{X}_i , $i \in N$ are given. We say that chain graphs G and H over N are *LWF Markov equivalent* relative to $\mathbf{X}_N = \prod_{i \in N} \mathbf{X}_i$ if the class of LWF Markovian probability measures with respect to G on \mathbf{X}_N equals the class of LWF Markovian probability measures with respect to H on \mathbf{X}_N . This is closely related to another notion. We say that G and H over N are *LWF independence equivalent* if the class of triplets over N represented in G (according to the moralization criterion) coincides with the class of triplets over N represented in H . Of course, LWF independence equivalent chain graphs have to be LWF Markov equivalent, and the converse holds in the case of non-trivial individual sample spaces. This follows from a well-known result by Frydenberg [6], proved also in [3].

Theorem 1. Let \mathbf{X}_i , $i \in N$ be non-trivial finite sample spaces, that is, $|\mathbf{X}_i| \geq 2$ for every $i \in N$. Then two chain graphs over N are LWF Markov equivalent relative to \mathbf{X}_N iff they have the same underlying graph and complexes.

It can be derived from Theorem 1 that two chain graphs are LWF independence equivalent iff they have the same underlying graph and complexes. Nevertheless, perhaps even more important result of Frydenberg [6] is the following one, which says that every LWF independence equivalence class has a distinguished member.

Theorem 2. Every LWF independence equivalence class \mathcal{H} of chain graphs over N has the largest chain graph, that is, $H_\infty \in \mathcal{H}$ such that $H_\infty \geq H$ for every $H \in \mathcal{H}$.

Another valuable contribution of Frydenberg is the concept of factorisation of a strictly positive density $p : \mathcal{X}_N \rightarrow (0, 1]$ of a probability measure over N with respect to a chain graph over N . This concept can be extended to any density in the discrete case – see Section 4.1. Another basic result from [6] says that a probability measure P over N which has a strictly positive density p is LWF Markovian with respect to a chain graph H over N if and only if p factorises with respect to H .

2.4. Bayesian network (statistical) models

A special case of a chain graph model is that ascribed to an acyclic directed graph. The definition of a Markovian probability measure with respect to an acyclic directed graph is a special case of the definition of LWF Markovian measure from Section 2.3. One can also introduce the concepts of Markov equivalence and independence equivalence for acyclic directed graphs. Thus, the following result of Verma and Pearl [18] is a consequence of Theorem 1 and the fact that the only complexes in an acyclic directed graph are immoralities.

Corollary 3. Let $X_i, i \in N$ be non-trivial finite sample spaces. Then two acyclic directed graphs over N are Markov equivalent relative to \mathcal{X}_N iff they have the same underlying graph and immoralities.

There is no distinguished member in an equivalence class \mathcal{G} of acyclic directed graphs over N . For this reason, researchers in this area suggested one describes \mathcal{G} by a uniquely determined representative, which is possibly outside \mathcal{G} , namely by a special chain graph. The *essential graph* of \mathcal{G} is a hybrid graph G^* over N defined as follows:

- $a \rightarrow b$ in G^* iff $a \rightarrow b$ in G for every $G \in \mathcal{G}$,
- $a - b$ in G^* iff there exist $G_1, G_2 \in \mathcal{G}$ such that $a \rightarrow b$ in G_1 and $a \leftarrow b$ in G_2 .

It follows from the results of [2] that the essential graph G^* is a chain graph without flags. One can also observe that G^* is LWF independence equivalent to every $G \in \mathcal{G}$ [15].

Note that statistical models based on acyclic directed graphs are commonly used in the area of probabilistic reasoning [10], where another term, *Bayesian networks*, is widely accepted.

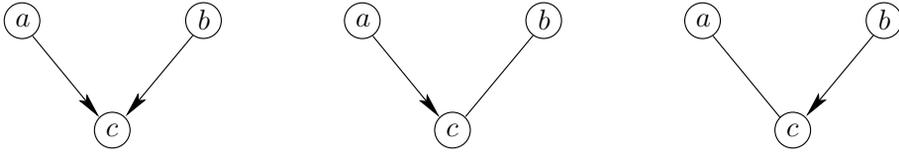


Fig. 2. Three forms of the triplex $(\{a, b\}, c)$.

2.5. Alternative interpretation of chain graphs

In the 1990s, Andersson, Madigan and Perlman [1] presented an alternative interpretation of chain graphs in terms of conditional independence. They later introduced the *augmentation criterion*, which is an analogue of the moralization criterion, defined the class of AMP Markovian measures and characterised in graphical terms the respective AMP Markov equivalence of chain graphs (for details see [4]). We say that chain graphs G and H over N are *AMP independence equivalent* if the class of triplets over N represented in G according to the augmentation criterion coincides with the class of triplets over N represented in H according to the augmentation criterion.

Here we do not give a detailed description of the augmentation criterion because it is quite technical and not very relevant to the particular topic of this paper. However, what is relevant is the graphical characterization of AMP independence equivalence [4]. A related graphical concept is that of a triplex. It is an induced subgraph for three nodes a, b, c , namely one of the graphs shown in Figure 2. Formally, a *triplex* in a chain graph H is a pair $(\{a, b\}, c)$, where $\{a, b\}$ is an unordered pair of distinct nodes, c is another node and one of the following three options occurs: either $a \rightarrow c \leftarrow b$ is an immorality in H , or $a \rightarrow c - b$ is a flag in H , or $b \rightarrow c - a$ is a flag in H . This was shown in [4]:

Theorem 4. Two chain graphs over N are AMP independence equivalent iff they have the same underlying graph and the same triplexes.

Note that the graphs shown in Figure 2 are AMP independence equivalent because all of them have the same triplex $(\{a, b\}, c)$. Actually, they form an AMP equivalence class. Observe that an AMP equivalence class need not have the largest graph.

To illustrate the difference in chain graph (conditional independence) interpretations consider the graph (B) in Figure 1. This is the simplest chain graph with different LWF and AMP interpretation. More specifically, the only non-trivial triplet represented in it according to the moralization criterion is $\langle a, b|c \rangle$, while the only non-trivial triplet represented in it according to the augmentation criterion from [4] is $\langle a, b|\emptyset \rangle$.

3. FEASIBLE MERGING COMPONENTS

This is an elementary operation with LWF independence equivalent chain graphs. It was first formulated in an elegant way in [11] under a different name “the removal of an insubstantial arrowhead of a meta-arrow”. Note that the operation appears to be equivalent to the application of the “pool-component rule” which was earlier introduced in [14] – this is shown in Section 3.2. Moreover, a careful reader of Frydenberg’s original paper [6] can find out that a special case of this operation was actually used to derive graphical characterisation of LWF Markov equivalence – see Lemma 5.1 of [6], namely, the condition saying that “a component τ is simplicial in the moral graph \tilde{G}^m ”.

This operation appears to be a very useful technical tool in proofs, as shown later in Sections 4 and 5. Thus, we feel that it deserves a special name.

Here is the formulation. Let H be a chain graph, and U, L a pair of its components which form a *meta-arrow* $U \rightrightarrows L$, by which we mean there exists at least one arrow $a \rightarrow b$ in H with $a \in U$ and $b \in L$. By *merging* of the *upper component* U and the *lower component* L (of the meta-arrow $U \rightrightarrows L$), we will understand the replacement of all arrows $a \rightarrow b$ in H , where $a \in U$ and $b \in L$, with lines. We then say that the resulting hybrid graph H' is obtained (from H) by merging (of) components U and L . Merging components will be called *feasible* if the following two conditions are satisfied:

- (i) $K \equiv \text{pa}_H(L) \cap U$ is a complete set in H ,
- (ii) for every $b \in K$ one has $\text{pa}_H(L) \setminus U \subseteq \text{pa}_H(b)$.

We also say that the resulting hybrid graph H' is obtained by the operation of *feasible merging* (of) components (U and L). Observe that the condition (ii) implies $\text{pa}_H(L) \setminus U \subseteq \text{pa}_H(U)$. Note that another formulation of (i)-(ii) is that $\text{pa}_H(L)$ is a complete set in the closure graph $\bar{H}(U)$ for the component U .

Example 1. We give here some instances of feasible and infeasible merging components with respect to the graph (A) in Figure 3.

1. For the components $U = C_1$ and $L = C_2$, condition (i) is not satisfied because $\text{pa}(L) \cap U = \{b, d\}$ is not complete; as a result, merging these components is not feasible and, in fact, such operation would destroy the complex $b \rightarrow e \leftarrow d$.
2. The components $U = C_1$ and $L = C_3$ satisfy condition (i) because $\text{pa}(L) \cap U = \{a\}$ is trivially complete, but condition (ii) is not satisfied because $\text{pa}(L) \setminus U = \{e\} \not\subseteq \text{pa}(a) = \emptyset$. Note that merging C_1 and C_2 would, for instance, create a semi-directed cycle $b \rightarrow e \rightarrow g - a - b$.
3. The components $U = C_2$ and $L = C_3$ satisfy condition (i) because $\text{pa}(L) \cap U = \{e\}$ is trivially complete but not condition (ii) because $\text{pa}(L) \setminus U = \{a\} \not\subseteq \text{pa}(e) = \{b, d\}$. Note that merging C_2 and C_3 would, for instance, create the complex $a \rightarrow g - e \leftarrow d$.

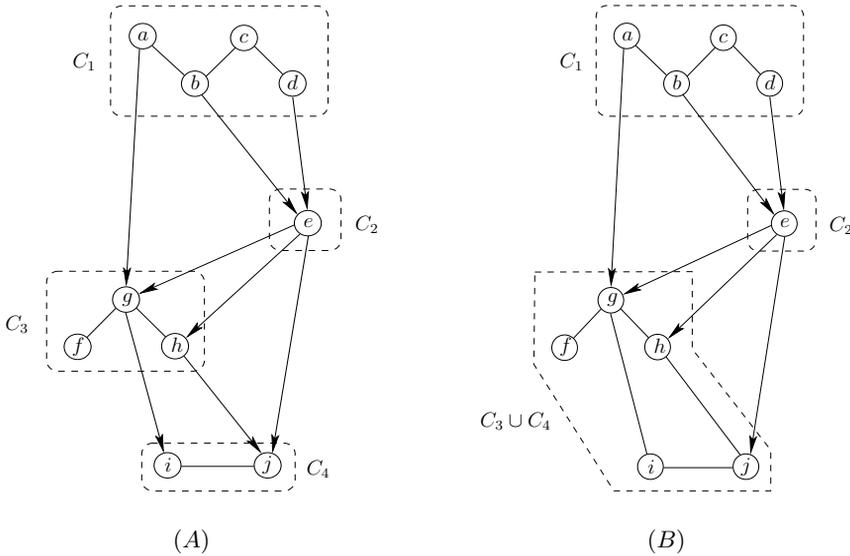


Fig. 3. The situation before, (A), and after, (B), a feasible merging components.

4. The components $U = C_3$ and $L = C_4$ satisfy condition (i) because $\text{pa}(L) \cap U = \{g, h\}$ is complete, and condition (ii) because both $\text{pa}(L) \setminus U = \{e\} \subseteq \text{pa}(h) = \{e\}$ and $\text{pa}(L) \setminus U = \{e\} \subseteq \text{pa}(g) = \{a, e\}$. Hence, merging C_2 and C_3 is feasible and the resulting chain graph is given in Figure 3 (B).

Consider now the chain graph (B) in Figure 3. In this case, feasible merging is not possible for any pair of components. Thus, it follows from Corollary 7 below that (B) is the largest chain graph.

3.1. Basic results on feasible merging

The meaning of the concept of feasible merging components in a chain graph is evident from the following characterisation, proved as Theorem 8 in [11].

Lemma 5. A graph obtained by merging components of a meta-arrow $U \rightrightarrows L$ in a chain graph H is a chain graph LWF independence equivalent to H iff the merging is feasible.

This result justifies our terminology “feasible merging”: since the outcome of this operation is again a graph in the same LWF equivalence class of chain graphs, the merging can be performed without leaving the class.

An important fact is that every larger LWF independence equivalent graph is attainable by a series of feasible merging operations. The following result follows from Theorems 7 and 8 in [11].

Theorem 6. If G and H are LWF independence equivalent chain graphs and $H \geq G$ then there exists a sequence of chain graphs H_1, \dots, H_r , $r \geq 1$ starting with $G = H_1$ and ending with $H = H_r$ such that $\forall i = 1, \dots, r - 1$ the graph H_{i+1} is obtained from H_i by the operation of feasible merging components (in H_i).

The above result has a simple consequence, which is the basis of our method to prove results in Sections 4 and 5.

Corollary 7. Let H be a chain graph, \mathcal{H} the LWF equivalence class of chain graphs containing H , and H_∞ the largest chain graph in \mathcal{H} . Then there exists a sequence of graphs $H = H_1, \dots, H_r = H_\infty$, $r \geq 1$ in which every next graph is obtained from the previous one by the operation of feasible merging components.

Thus, owing to Theorem 2, if we want to prove that some feature is shared by all chain graphs in an LWF independence equivalence class, it suffices to show that the feature is shared by any pair of chain graphs such that one of them is obtained from the other by feasible merging components. This observation simplifies some proofs.

3.2. Relation to the pool-component rule

The pool-component rule introduced in § 5 of [14] was intended as the basic step of a procedure for getting the largest chain graph in a LWF independence equivalence graph on the basis of any graph in the equivalence class. Its formulation, recalled below, is certainly more complicated than the formulation of the operation of feasible merging.

Here is the original formulation. Let H be a chain graph and C a component in H . Let $\mathcal{K}(C)$ denote the class of complexes $\kappa : a \rightarrow c_1 \text{ --- } \dots \text{ --- } c_s \leftarrow b$, $s \geq 1$ in H with $c_1, \dots, c_s \in C$. Given such a complex κ , the set of its “parents” $\{a, b\}$ will be denoted by $\text{par}(\kappa)$. The *pool-component rule* is applicable to H if there exists a component C in H such that the set $A = \text{pa}_H(C) \cap \bigcap_{\kappa \in \mathcal{K}(C)} \bigcap_{v \in \text{par}(\kappa)} \text{ch}_H(v)$ is non-empty. Note that a natural convention is accepted that the intersection of the empty class (of sets of nodes) is the largest possible set (of nodes). Thus, in case $\mathcal{K}(C) = \emptyset$ we have $\bigcap_{\kappa \in \mathcal{K}(C)} \bigcap_{v \in \text{par}(\kappa)} \text{ch}_H(v) = N$ and $A = \text{pa}_H(C)$.

Then we choose a terminal component B of H_A and replace all arrows from B to C with lines. It was shown in Lemma 4.6 of [14] that the obtained hybrid graph is a chain graph which is LWF independence equivalent to H . The point is that, despite formal differences, both operations are equivalent.

Proposition 8. Let H and H' be chain graphs over the same set of nodes. Then H' is obtained from H by the application of the pool-component rule iff it is obtained from H by the operation of feasible merging its components.

The proof is given in the Appendix (Section A1).

3.3. Feasible splitting a component

An operation inverse to merging components is splitting a component. Let H' be a chain graph and M a component in H' . Let M be partitioned into non-empty connected subsets U and L . By *splitting* (of) the component M into the *upper component* U and the *lower component* L we will understand the replacement of all lines between U and L with arrows from U to L . It is easy to see that there exists at least one line between U and L , and that the resulting graph H is also a chain graph in which $U \rightrightarrows L$ is a meta-arrow. We say that H is obtained (from H') by splitting (of) the component M (into the upper component U and the lower component L).

Splitting a component will be called *feasible* if the following two conditions are fulfilled:

- (i)' $K \equiv \text{ne}_{H'}(L) \cap U$ is a complete set in H' ,
- (ii)' for every $b \in K$ one has $\text{pa}_{H'}(L) \subseteq \text{pa}_{H'}(b)$.

Obviously, these two conditions are nothing but the requirements (i) and (ii) on the resulting graph H , formulated in terms of H' . They are also illustrated in Figure 3(B), where $M = L \cup U$ with $L = \{i, j\}$ and $U = \{f, g, h\}$. Thus, one can easily derive from Lemma 5:

Corollary 9. The chain graph obtained by splitting a component M in a chain graph H' is LWF independence equivalent to H' iff the splitting is feasible.

4. CHARACTERISATION OF FACTORISATION EQUIVALENCE

To motivate the results in this section let us put them in the context of statistical learning (graphical models). Learning procedures are usually based on the interpretation of graphs as statistical models. A chain graph can be interpreted as a statistical model either in terms of conditional independence or in terms of factorisation. Some people, however, prefer the interpretation in terms of factorisation, in particular in the case of undirected graphs and acyclic directed graphs. This is because the definition of a factorisable distribution is much simpler (than the definition of a Markovian distribution) in these two cases.

The factorisation condition with respect to a chain graph was shown by Frydenberg [6] to be equivalent to the LWF Markov property only for strictly positive densities. Later, Lauritzen, in his book [9], introduced factorisation with respect to a chain graph for any probability measure which has a density. Thus, the situation (in the discrete case) is as follows. One can assign to any chain graph G these statistical models (= classes of distributions): the class $\mathcal{M}_+(G)$ of Markovian distributions with strictly positive density, the class $\mathcal{M}(G)$ of (all discrete) Markovian distributions and the class $\mathcal{F}(G)$ of “factorizable” distributions. Their relation is as follows: $\mathcal{M}_+(G) \subseteq \mathcal{F}(G) \subseteq \mathcal{M}(G)$.

Now, because the aim of a learning procedure is to identify the statistical model an immediate mathematical question is when two (chain) graphs define the same statistical model. As mentioned in previous sections, Frydenberg [6] characterized this

equivalence in graphical terms in the case of Markovian and strictly positive Markovian interpretation. However, if one is interested in “factorisation” interpretation, then the natural question is to characterize (in graphical terms) the corresponding “factorisation” equivalence.

A certain complication is that Lauritzen’s [9] definition of factorisation with respect to a chain graph admits two different interpretations. We show below that one can assign, in the discrete case, to any chain graph G the class of factorisable distributions $\mathcal{F}(G)$ and the class of “strongly factorisable” distributions $\mathcal{F}_*(G)$ and their relation is $\mathcal{F}_*(G) \subseteq \mathcal{F}(G)$. In our view, the larger class $\mathcal{F}(G)$ is the right (= appropriate) interpretation. The problem of graphical characterisation of the corresponding factorisation equivalence was solved in [13] with the essential use of the operation of feasible merging components. In this paper, we publish this result.

4.1. The concept of factorisation

The aim of this section is to clarify the definition of factorisation of a discrete probability measure with respect to a chain graph. The point is that Lauritzen’s original definition (see § 3.2.3 of [9]) permits different interpretations, depending on how one formally defines the concept of recursive factorisation.

We have chosen one of those interpretations as the basis for our definition of factorisation. An example is given which shows that the other interpretation indeed leads to a different concept of factorisation with respect to a chain graph. We explain the reasons for our choice of the interpretation in a remark at the end of Section 4.1.

4.1.1. Component-wise factorisation

Factorisation with respect to a chain graph involves two requirements. The first one is a ‘global’ *component-wise recursive factorisation condition*. Let $p : \mathbf{X}_N \rightarrow [0, 1]$ be a density of a discrete probability measure P over N , and \mathcal{C}_H the class of components in a chain graph H over N . We have in mind the following condition:

$$\forall x \in \mathbf{X}_N \quad p(x) = \prod_{C \in \mathcal{C}_H} p_{C | \text{pa}_H(C)}(x_C | x_{\text{pa}_H(C)}). \tag{2}$$

What is perhaps unclear is which versions of conditional densities $p_{C | \text{pa}_H(C)}$, $C \in \mathcal{C}_H$ should be considered in (2). Fortunately, the validity of (2) does not depend on their choice. We show that this condition is actually equivalent to a certain Markovian condition, that is, to the validity of some conditional independence statements. Given a chain $\pi : C_1, \dots, C_m$, $m \geq 1$ of components in a chain graph H , the symbol $\text{pre}_\pi(C)$ will be used to denote the set of *predecessors* of a component $C \in \mathcal{C}_H$ in π , that is, if $C = C_i$ then $\text{pre}_\pi(C) = \bigcup_{j < i} C_j$.

Proposition 10. Let H be a chain graph over N , and π a chain of its components. Let $p : \mathbf{X}_N \rightarrow [0, 1]$ be the density of a discrete probability measure P over N . Then the following conditions are equivalent:

$$(\text{MC-}\pi) \quad \forall C \in \mathcal{C}_H \quad C \perp\!\!\!\perp \text{pre}_\pi(C) \setminus \text{pa}_H(C) \mid \text{pa}_H(C) [P],$$

(FC- \forall) the formula (2) holds for every collection $\{p_{C|\text{pa}_H(C)}; C \in \mathcal{C}_H\}$ of versions of conditional densities of P ,

(FC- \exists) there exists a collection $\{p_{C|\text{pa}_H(C)}; C \in \mathcal{C}_H\}$ of conditional density versions of P such that (2) holds.

A technical proof of this result appears in Section A2 of the Appendix. It follows from the result that the validity of (MC- π) does not depend on the choice of a chain π . Actually, we can formulate this condition without a reference to any chain.

Corollary 11. Under the assumptions of Proposition 10, the condition (MC- π) is equivalent to the condition

$$(MC) \quad \forall C \in \mathcal{C}_H \quad C \perp\!\!\!\perp N \setminus (\text{ds}_H(C) \cup \text{pa}_H(C)) \mid \text{pa}_H(C) [P].$$

Proof. The implication (MC) \Rightarrow (MC- π) easily follows from a well-known formal property of conditional independence, saying that $A \perp\!\!\!\perp B \mid C [P]$ and $B' \subseteq B$ implies $A \perp\!\!\!\perp B' \mid C [P]$.

To show (MC- π) \Rightarrow (MC), let us recall that Proposition 10 implies that the validity of (MC- π) does not depend on the choice of π . Therefore, it suffices to show that, for every $C \in \mathcal{C}_H$, there exists a chain π for H with $N \setminus \text{ds}_H(C) = \text{pre}_\pi(C)$. This follows easily from the observations that $\text{ds}_H(C) \setminus C$ is a union of components in H , and there is no arrow in H from $\text{ds}_H(C) \setminus C$ to its complement in N . \square

4.1.2. Clique-wise factorisation

The second part of factorisation with respect to a chain graph is a series of ‘local’ *clique-wise factorisation conditions*, one for each component. The simplest formulation is that the respective marginal density factorises according to the respective undirected graph, namely, the closure graph for the component.

Let $C \in \mathcal{C}_H$ be a component in a chain graph H over N , and let $\mathcal{K}_H(C)$ denote the collection of cliques of $\bar{H}(C)$. We have in mind the following requirement:

$$\begin{aligned} \forall C \in \mathcal{C}_H \quad \exists \varphi_K : \mathcal{X}_K \rightarrow [0, \infty), K \in \mathcal{K}_H(C) \quad \forall y \in \mathcal{X}_{C \cup \text{pa}_H(C)} \\ p_{C \cup \text{pa}_H(C)}(y) = \prod_{K \in \mathcal{K}_H(C)} \varphi_K(y_K). \end{aligned} \tag{3}$$

Of course, the collection of cliques $\mathcal{K}_H(C)$ in (3) can be replaced with the collection of complete sets in $\bar{H}(C)$.

Definition 1. We say that the density $p : \mathcal{X}_N \rightarrow [0, 1]$ of a discrete probability measure P over N *factorises with respect to a chain graph H over N* if the conditions (2) and (3) are fulfilled.

Nevertheless, the clique-wise factorisation condition can equivalently be expressed in terms of conditional densities.

Lemma 12. Let H be a chain graph over N , $C \in \mathcal{C}_H$ one of its components, and $\mathcal{K}_H(C)$ the collection of all cliques of the closure graph $\bar{H}(C)$. Then the following conditions are equivalent:

$$(MF) \exists \varphi_K : \mathbf{X}_K \rightarrow [0, \infty), K \in \mathcal{K}_H(C) \quad \forall y \in \mathbf{X}_{C \cup \text{pa}_H(C)}$$

$$p_{C \cup \text{pa}_H(C)}(y) = \prod_{K \in \mathcal{K}_H(C)} \varphi_K(y_K),$$

$$(CF-0) \exists \psi_K : \mathbf{X}_K \rightarrow [0, \infty), K \in \mathcal{K}_H(C) \quad \forall y \in \mathbf{X}_{C \cup \text{pa}_H(C)}$$

$$p_{C | \text{pa}_H(C)}^0(y_C | y_{\text{pa}_H(C)}) = \prod_{K \in \mathcal{K}_H(C)} \psi_K(y_K),$$

where $p_{C | \text{pa}_H(C)}^0$ denotes the zero-version of respective conditional density,

(CF- \exists) there exists a version of conditional density $p_{C | \text{pa}_H(C)}$ such that

$$\exists \psi_K : \mathbf{X}_K \rightarrow (-\infty, \infty), K \in \mathcal{K}_H(C) \quad \forall y \in \mathbf{X}_{C \cup \text{pa}_H(C)}$$

$$p_{C | \text{pa}_H(C)}(y_C | y_{\text{pa}_H(C)}) = \prod_{K \in \mathcal{K}_H(C)} \psi_K(y_K).$$

Proof. The implication (MF) \Rightarrow (CF-0) follows easily from the fact that

$$\forall y \in \mathbf{X}_{C \cup \text{pa}_H(C)} \quad p_{C | \text{pa}_H(C)}^0(y_C | y_{\text{pa}_H(C)}) = p_{C \cup \text{pa}_H(C)}(y) \cdot \tau(y_{\text{pa}_H(C)}),$$

where the potential $\tau : \mathbf{X}_{\text{pa}_H(C)} \rightarrow [0, \infty)$ is defined as follows:

$$\forall z \in \mathbf{X}_{\text{pa}_H(C)} \quad \tau(z) = \begin{cases} p_{\text{pa}_H(C)}^{-1}(z) & \text{if } p_{\text{pa}_H(C)}(z) > 0, \\ 0 & \text{if } p_{\text{pa}_H(C)}(z) = 0. \end{cases}$$

The implication (CF-0) \Rightarrow (CF- \exists) is evident. The implication (CF- \exists) \Rightarrow (MF) follows easily from the equality $p_{C \cup \text{pa}_H(C)} = |p_{C | \text{pa}_H(C)}| \cdot p_{\text{pa}_H(C)}$ and from the observation that ψ -potentials can be replaced with their absolute values. \square

An important warning concerning Lemma 12 should be mentioned. Unlike component-wise factorisation, clique-wise factorisation does depend on the choice of versions of conditional densities. More specifically, the condition (CF- \exists) is *not* equivalent to the condition:

(CF- \forall) for every version of conditional density $p_{C | \text{pa}_H(C)}$

$$\exists \psi_K : \mathbf{X}_K \rightarrow (-\infty, \infty), K \in \mathcal{K}_H(C) \quad \forall y \in \mathbf{X}_{C \cup \text{pa}_H(C)}$$

$$p_{C | \text{pa}_H(C)}(y_C | y_{\text{pa}_H(C)}) = \prod_{K \in \mathcal{K}_H(C)} \psi_K(y_K).$$

A counter-example is given in Example 2 in Section 4.1.3. Actually, the example shows that it may happen that the only version of conditional density which factorises according to the closure graph is the zero-version. Lemma 12 and Proposition 10 allow us to formulate factorisation with respect to a chain graph in the form of one two-stage factorisation condition. That means, the factors in the recursive factorisation can further factorise into subfactors.

Corollary 13. Let H be a chain graph over N and $p : \mathcal{X}_N \rightarrow [0, 1]$ the density of a discrete probability measure P over N . Then p factorises with respect to H iff there exists a collection of versions of conditional densities $\{p_{C|\text{pa}_H(C)}, C \in \mathcal{C}_H\}$ such that

- $\forall x \in \mathcal{X}_N \quad p(x) = \prod_{C \in \mathcal{C}_H} p_{C|\text{pa}_H(C)}(x_C|x_{\text{pa}_H(C)}),$ and
- $\forall C \in \mathcal{C}_H \quad \exists \psi_K^C : \mathcal{X}_K \rightarrow (-\infty, \infty), K \in \mathcal{K}_H(C)$ such that

$$\forall x \in \mathcal{X}_N \quad p_{C|\text{pa}_H(C)}(x_C|x_{\text{pa}_H(C)}) = \prod_{K \in \mathcal{K}_H(C)} \psi_K^C(x_K).$$

Moreover, if this is the case then the collection of zero-versions $\{p_{C|\text{pa}_H(C)}^0, C \in \mathcal{C}_H\}$ complies with these two requirements.

Note that the choice of zero-versions in Corollary 13 may be the only possible choice – see Example 2. A crucial fact is that factorisation implies the respective global Markov property.

Proposition 14. Let H be a chain graph over N and P a discrete probability measure over N . If the density of P factorises with respect to H then P is LWF Markovian with respect to H .

Proof. One has to show that $A \perp\!\!\!\perp B | C [P]$ whenever a disjoint triplet $\langle A, B | C \rangle$ over N is represented in H according to the moralization criterion. Consider the respective set $T = \text{an}_H(A \cup B \cup C)$. A basic step is to observe that the marginal density p_T factorises with respect to the induced chain graph H_T . To show that (2) holds for p_T , use Proposition 10, namely the fact that component-wise factorisation is equivalent to the condition (MC- π). Thus, it suffices to construct a chain of components $\pi : C_1, \dots, C_m, m \geq 1$ of H such that $T = \bigcup_{j \leq m} C_j$ for some $k \leq m$. The next step is to apply Corollary 13 to p_T and H_T , thus getting factorisation of p_T into factors depending on cliques of the closure graphs for components of H_T . These cliques are surely complete sets in the moral graph of H_T . In particular, p_T factorises according to the moral graph H_T^{mor} , which is an undirected graph. Thus, by Proposition 3.8 in [9], p_T is (globally) Markovian with respect to H_T^{mor} . As $\langle A, B | C \rangle$ is represented in H_T^{mor} , it implies $A \perp\!\!\!\perp B | C [P]$. \square

4.1.3. Strong factorisation

The condition in Corollary 13 was basically the way in which Lauritzen introduced factorisation with respect to a chain graph in § 3.2.3 of [9]. Nevertheless, his description of recursive factorisation raised some doubts. He wrote

“we first assume factorisation of the density as in the directed acyclic case”.

However, in § 3.2.2 of [9], where recursive factorisation with respect to an acyclic directed graph is defined, an additional technical assumption is mentioned, namely that the respective factors are kernels, that is, regular versions of conditional densities. Thus, strict formal interpretation of this idea leads to the following definition.

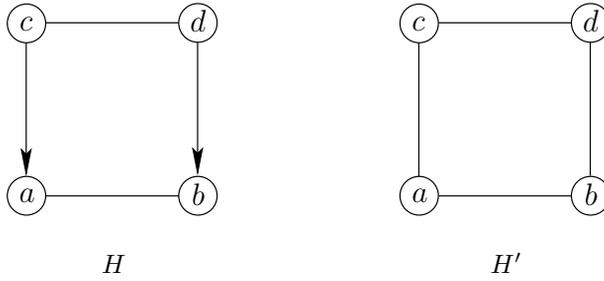


Fig. 4. Two chain graphs.

Definition 2. We say that the density $p : \mathcal{X}_N \rightarrow [0, 1]$ of a discrete probability measure P over N factorises strongly with respect to a chain graph H over N if there exists a collection of **regular** versions of conditional densities $\{p_{C|\text{pa}_H(C)}^{\text{reg}}, C \in \mathcal{C}_H\}$ such that

- $\forall x \in \mathcal{X}_N \quad p(x) = \prod_{C \in \mathcal{C}_H} p_{C|\text{pa}_H(C)}^{\text{reg}}(x_C | x_{\text{pa}_H(C)})$, and
- $\forall C \in \mathcal{C}_H \quad \exists \psi_K^C : \mathcal{X}_K \rightarrow (-\infty, \infty), K \in \mathcal{K}_H(C)$ such that

$$\forall x \in \mathcal{X}_N \quad p_{C|\text{pa}_H(C)}^{\text{reg}}(x_C | x_{\text{pa}_H(C)}) = \prod_{K \in \mathcal{K}_H(C)} \psi_K^C(x_K).$$

It follows from Corollary 13 that strong factorisation implies factorisation introduced in Definition 1. The converse implication holds for discrete probability measures with strictly positive density p , that is, with $p(x) > 0$ for every $x \in \mathcal{X}_N$. This is because in that case there is no flexibility in the choice of versions of conditional densities: $p_{C|\text{pa}_H(C)}^0$ is a regular version for every $C \in \mathcal{C}_H$. However, factorisation does not generally imply strong factorisation, as the following example shows.

Example 2. There exists a chain graph H over $N = \{a, b, c, d\}$ and the density p of a discrete probability measure P over N such that p factorises with respect to H but it does not factorise strongly with respect to H .

The graph is shown in the left-hand picture of Figure 4. Let us put $\mathcal{X}_a = \{a, \bar{a}\}$, $\mathcal{X}_b = \{b, \bar{b}\}$, $\mathcal{X}_c = \{c, \bar{c}\}$, $\mathcal{X}_d = \{d, \bar{d}\}$ and define p as a function on \mathcal{X}_N as follows:

$$\begin{aligned}
 p : \quad & (a, b, c, d) \rightarrow 1/5, \\
 & (a, \bar{b}, c, d) \rightarrow 1/5, \\
 & (\bar{a}, b, c, d) \rightarrow 1/5, \\
 & (\bar{a}, b, \bar{c}, d) \rightarrow 1/5, \\
 & (a, b, c, \bar{d}) \rightarrow 1/5, \\
 & \text{otherwise} \rightarrow 0.
 \end{aligned}$$

To show that p factorises with respect to H , one has to verify that $p \equiv p_{L \cup \text{pa}_H(L)}$ for the component $L = \{a, b\} \in \mathcal{C}_H$ factorises according to the graph H' shown in the right-hand picture of Figure 4. To this end, we put

$$\begin{aligned} \varphi_{ab}(\mathbf{ab}) &= \varphi_{ab}(\mathbf{a\bar{b}}) = \varphi_{ab}(\mathbf{\bar{a}b}) = 1, & \varphi_{ab}(\mathbf{\bar{a}\bar{b}}) &= 0, \\ \varphi_{ac}(\mathbf{ac}) &= \varphi_{ac}(\mathbf{\bar{a}c}) = \varphi_{ac}(\mathbf{a\bar{c}}) = 1, & \varphi_{ac}(\mathbf{a\bar{c}}) &= 0, \\ \varphi_{bd}(\mathbf{bd}) &= \varphi_{bd}(\mathbf{\bar{b}d}) = \varphi_{bd}(\mathbf{b\bar{d}}) = 1, & \varphi_{bd}(\mathbf{b\bar{d}}) &= 0, \end{aligned}$$

and $\varphi_{cd}(\hat{\mathbf{c}}\hat{\mathbf{d}}) = 1/5$ for any configuration $[\hat{\mathbf{c}}, \hat{\mathbf{d}}] \in \mathbf{X}_{\{c,d\}}$.

The next step is to show that the only version of conditional density $p_{L|\text{pa}_H(L)}$ which factorises according to H' is the zero-version. Thus, assume that, for some version of $p_{L|\text{pa}_H(L)}$, there are ψ -potentials such that

$$\forall x \in \mathbf{X}_N \quad p_{L|\text{pa}_H(L)}(x_L | x_{\text{pa}_H(L)}) = \psi_{ab}(x_{ab}) \cdot \psi_{ac}(x_{ac}) \cdot \psi_{bd}(x_{bd}) \cdot \psi_{cd}(x_{cd}).$$

First, we observe that some of the values of ψ -potentials are non-zero (for simplicity, we omit superfluous lower indices in the rest of the example):

$$\begin{aligned} 0 \neq 1/3 &= p(\mathbf{a}, \mathbf{b} | \mathbf{c}, \mathbf{d}) = \psi(\mathbf{ab}) \cdot \psi(\mathbf{ac}) \cdot \psi(\mathbf{bd}) \cdot \psi(\mathbf{cd}) \Rightarrow \psi(\mathbf{ab}) \neq 0, \psi(\mathbf{ac}) \neq 0, \\ 0 \neq 1 &= p(\mathbf{a}, \bar{\mathbf{b}} | \mathbf{c}, \bar{\mathbf{d}}) = \psi(\mathbf{a\bar{b}}) \cdot \psi(\mathbf{ac}) \cdot \psi(\bar{\mathbf{b}d}) \cdot \psi(\mathbf{c\bar{d}}) \Rightarrow \psi(\mathbf{c\bar{d}}) \neq 0, \psi(\mathbf{a\bar{b}}) \neq 0, \\ 0 \neq 1/3 &= p(\mathbf{a}, \bar{\mathbf{b}} | \bar{\mathbf{c}}, \mathbf{d}) = \psi(\mathbf{a\bar{b}}) \cdot \psi(\mathbf{ac}) \cdot \psi(\bar{\mathbf{b}d}) \cdot \psi(\mathbf{cd}) \Rightarrow \psi(\bar{\mathbf{b}d}) \neq 0, \\ 0 \neq 1 &= p(\bar{\mathbf{a}}, \bar{\mathbf{b}} | \bar{\mathbf{c}}, \mathbf{d}) = \psi(\bar{\mathbf{a}\bar{b}}) \cdot \psi(\bar{\mathbf{a}\bar{c}}) \cdot \psi(\bar{\mathbf{b}d}) \cdot \psi(\mathbf{c\bar{d}}) \Rightarrow \psi(\mathbf{c\bar{d}}) \neq 0, \\ 0 \neq 1/3 &= p(\bar{\mathbf{a}}, \bar{\mathbf{b}} | \mathbf{c}, \mathbf{d}) = \psi(\bar{\mathbf{a}\bar{b}}) \cdot \psi(\bar{\mathbf{a}\bar{c}}) \cdot \psi(\bar{\mathbf{b}d}) \cdot \psi(\mathbf{cd}) \Rightarrow \psi(\bar{\mathbf{a}\bar{c}}) \neq 0, \psi(\mathbf{cd}) \neq 0. \end{aligned}$$

Then, we use these facts to observe that some other values of ψ -potentials are zeros:

$$\begin{aligned} 0 &= p(\mathbf{a}, \mathbf{b} | \bar{\mathbf{c}}, \bar{\mathbf{d}}) = \psi(\mathbf{ab}) \cdot \psi(\mathbf{ac}) \cdot \psi(\bar{\mathbf{b}d}) \cdot \psi(\mathbf{c\bar{d}}) \Rightarrow \psi(\bar{\mathbf{b}d}) = 0, \\ 0 &= p(\mathbf{a}, \bar{\mathbf{b}} | \bar{\mathbf{c}}, \mathbf{d}) = \psi(\mathbf{a\bar{b}}) \cdot \psi(\bar{\mathbf{a}\bar{c}}) \cdot \psi(\bar{\mathbf{b}d}) \cdot \psi(\mathbf{cd}) \Rightarrow \psi(\bar{\mathbf{a}\bar{c}}) = 0, \\ 0 &= p(\bar{\mathbf{a}}, \bar{\mathbf{b}} | \mathbf{c}, \mathbf{d}) = \psi(\bar{\mathbf{a}\bar{b}}) \cdot \psi(\bar{\mathbf{a}\bar{c}}) \cdot \psi(\bar{\mathbf{b}d}) \cdot \psi(\mathbf{cd}) \Rightarrow \psi(\bar{\mathbf{a}\bar{b}}) = 0. \end{aligned}$$

These observations allow one to see that $p(\hat{\mathbf{a}}, \hat{\mathbf{b}} | \bar{\mathbf{c}}, \bar{\mathbf{d}}) = 0$ for any configuration $[\hat{\mathbf{a}}, \hat{\mathbf{b}}] \in \mathbf{X}_{\{a,b\}}$. Thus, the only version of conditional density $p_{L|\text{pa}_H(L)}$ which factorises according to H' is the zero-version. In particular, no regular version $p_{L|\text{pa}_H(L)}^{\text{reg}}$ satisfies this requirement, and p does not factorise strongly with respect to H .

Remark. We have learned that there are at least two distinct concepts of factorisation with respect to a chain graph. The reader may ask why we gave preference to one of them. The basis of our considerations is Proposition 10, which says that the recursive factorisation does not depend on the choice of versions of conditional densities. From this point of view, the requirement of their regularity accepted by Lauritzen in § 3.2.2 of his book [9] is superfluous. We think we understand his reasons for that additional requirement: it allows one to see immediately that factorisation is preserved if one restricts oneself to a set of nodes closed under ancestors. However, this fact can also be derived easily with the help of Proposition 10, as shown in the proof of Proposition 14.

Our conclusion is that the assumption of regularity of conditional densities should be omitted. This leads to the chosen interpretation of factorisation with respect to a chain graph. Moreover, by Corollary 13, this interpretation is consistent with Proposition 3.30 from [9], whose proof was left to the reader there. Let us note that Proposition 3.30 from [9] is not valid under the other interpretation, that is, it does not hold for “strong factorisation”.

Of course, one could accept the alternative interpretation of Lauritzen’s formulation in §3.2.3 of [9] and consider “strong factorisation” with respect to a chain graph. This may lead to an alternative concept of “strong factorisation equivalence”. Nevertheless, since we believe that the requirement of regularity of conditional densities is somehow artificial, we do not regard the concept of strong factorisation (and the respective concept of strong equivalence of chain graphs) as fruitful. Of course, other researchers can have a different opinion – see the remark concluding Section 4.2.

4.2. Factorisation equivalence

Suppose that finite sample spaces X_i , $i \in N$ are given. We say that chain graphs G and H over N are *factorisation equivalent* relative to $X_N = \prod_{i \in N} X_i$ if the class of densities on X_N which factorise with respect to G coincides with the class of densities on X_N which factorise with respect to H .

In this section we give graphical characterisation of factorisation equivalence relative to non-trivial joint sample spaces. Actually, the characterisation appears to be the same as in the case of LWF Markov equivalence. We repeat the proof from [13], which is based on the operation of feasible merging components. The basic step is as follows.

Proposition 15. Let H' be a chain graph obtained from a chain graph H over N by feasible merging its components. Then H' is factorisation equivalent to H relative to any $X_N = \prod_{i \in N} X_i$ with finite sample spaces X_i , $i \in N$.

The proof is given in the Appendix (Section A3). Now, the graphical characterisation of factorisation equivalence can be obtained.

Theorem 16. Let X_i , $i \in N$ be non-trivial finite sample spaces, that is, $|X_i| \geq 2$ for every $i \in N$. Two chain graphs over N are factorisation equivalent relative to X_N iff they have the same underlying graph and complexes, that is, iff they are LWF independence equivalent.

Note that to show the necessity of the graphical conditions, which is the easier part of the proof, we just repeat the arguments given already by Frydenberg [6]. The substantial simplification brought by the concept of feasible merging (of) components is in their sufficiency proof. Indeed, to show directly that two chain graphs with the same underlying graph and complexes are factorisation equivalent would have been very complicated because the graphs can differ quite a lot. On the other hand, if the graphs differ slightly, namely if one of them is obtained from the other by feasible merging of its components, then this is relatively easy – as shown in the proof of

Proposition 15. Nevertheless, owing to our former results on feasible merging from Section 3.1, this observation is enough to conclude what is needed.

Proof. The *necessity* of the graphical conditions can be derived as follows. Theorem 4.1 in [6] says that a strictly positive density of a probability measure P over N factorises with respect to a chain graph iff P is Markovian with respect to it. Thus, the assumption that chain graphs G and H are factorisation equivalent relative to X_N implies that they have the same class of Markovian probability measures on X_N with strictly positive densities. Therefore, one can repeat the constructions mentioned on pp. 347-348 of [6] to show that G and H have the same underlying graph and complexes. Note that, in this step, one actually needs the assumption that $|X_i| \geq 2$ for every $i \in N$.

To verify the *sufficiency* of the graphical conditions, realise that Proposition 15 says that two chain graphs are factorisation equivalent whenever one is obtained from the other by the operation of feasible merging its components. We also know (that it follows from Theorem 1) that two graphs the same underlying graph and complexes iff they are LWF independence equivalent. Thus, we can use Corollary 7 to show that every graph in an LWF independence equivalence class of chain graphs is factorisation equivalent to the respective largest chain graph. In particular, any two LWF independence equivalent chain graphs are factorisation equivalent to the respective largest chain graph, for which reason they must be factorisation equivalent. This implies what is needed. \square

Remark. The reader may ask if a result analogous to Theorem 16 holds for “strong factorisation equivalence” of chain graphs mentioned in the end of Section 4.1.3. The answer is negative. The graphs in Figure 4 are not strongly factorisation equivalent because the density p in Example 2 factorises strongly with respect to H' but not with respect to H . Thus, readers interested in causal interpretation of arrows in a chain graph perhaps appreciate that “strong factorisation equivalence” of chain graphs makes distinction between these two graphs. The question of graphical characterisation of this equivalence remains open.

5. INVARIANTS OF LWF INDEPENDENCE EQUIVALENCE

The operation of feasible merging components can also be used to show easily that some characteristics of chain graphs are invariant with respect to LWF independence equivalence of chain graphs. In this section, we introduce certain concepts invariable within equivalence classes of this equivalence. They have close connection to an algebraic approach to the description of Bayesian network models, described in Chapter 8 of [17]. More precisely, these concepts play an essential role in an algorithm from [16] for the transition between standard graphical and algebraic representatives of Bayesian network models. We hope that this algorithm can later be utilized in the area of (statistical) learning Bayesian network models.

5.1. Number of initial components

A component C of a chain graph H will be called *initial* if $\text{pa}_H(C) = \emptyset$. Actually, this is a counterpart of the concept of a terminal component. Although initial components themselves are not invariants of LWF equivalence, their number is invariant.

Proposition 17. The number of initial components is the same for LWF independence equivalent chain graphs.

Proof. First, we assume that a graph H' is obtained from a chain graph H by feasible merging components U and L . Then L is not initial in H and U is initial in H iff the merged component $M = U \cup L$ is initial in H' ; this is because $\text{pa}_H(L) \setminus U \subseteq \text{pa}_H(U)$. Since the sets $\text{pa}_H(C)$ for other components $C \in \mathcal{C}_H$ are saved in H' , the graphs have the same number of initial components.

Second, apply Corollary 7 to see that the largest chain graph H_∞ in an LWF equivalence class \mathcal{H} has the same number of initial components as any $H \in \mathcal{H}$. \square

5.2. The largest idle set

A set of nodes $I \subseteq N$ in a chain graph H over N will be called *idle* if the following two conditions hold:

- (a) $\forall y_1, y_2 \in I, y_1 \neq y_2 \quad [y_1, y_2]$ is an edge in H ,
- (b) $\forall x \in N \setminus I \quad \forall y \in I \quad x \rightarrow y$ in H .

Observe that the empty set of nodes is always idle. Another notable fact is that every non-empty idle set is the union of some components in the graph.

Lemma 18. Every two idle sets in a chain graph are in inclusion relation. In particular, every chain graph has a unique largest idle set, which is possibly empty.

Proof. Let I_1 and I_2 be two idle sets in a chain graph H ; for contradiction, suppose that both $y \in I_1 \setminus I_2$ and $z \in I_2 \setminus I_1$ exist. Then, by (b) for I_1 , derive $z \rightarrow y$ in H , and, by (b) for I_2 , observe $y \rightarrow z$ in H . These two facts are in contradiction, which means that one has either $I_1 \subseteq I_2$ or $I_2 \subseteq I_1$. The observation that the collection of idle sets in H is a finite nest implies the existence of the largest idle set. \square

Remark. The meaning of an idle set is as follows: there is no (non-trivial) conditional independence statement represented in the graph that involves variables in an idle set. For example, the only non-trivial triplet represented in the graph (A) in Figure 1 is $\langle a, b | \emptyset \rangle$. Thus, $\{c\}$ is the largest idle set in that graph. On the other hand, the largest idle set in the graph H in Figure 4 is empty since the triplet $\langle a, d | \{b, c\} \rangle$ is represented in the graph. Thus, the complement of the largest idle set, called the *core* of the graph in [16], is the least set involving all non-trivial conditional

independence statements. It can be interpreted as the set of “substantial” variables in the graphical model.

The largest idle set is a shared characteristic of LWF equivalent chain graphs.

Proposition 19. If G and H are LWF independence equivalent chain graphs then they have the same largest idle set.

Proof. Let H' be obtained from H by feasible merging components of a meta-arrow $U \rightrightarrows L$. Supposing $Y \subseteq N$ is an idle set in H' , observe that the merged component $M \equiv U \cup L$ is either a subset of Y or a subset of $N \setminus Y$. This implies that Y is an idle set in H .

Conversely, suppose that I is an idle set in H . We show that it is a subset of an idle set in H' . If $U, L \subseteq I$ or $U, L \subseteq N \setminus I$, then I is idle in H' . The case $[U \subseteq I \ \& \ L \subseteq N \setminus I]$ is excluded by the condition (b) for I . If $L \subseteq I$ and $U \subseteq N \setminus I$, then (b) implies $K \equiv \text{pa}_H(L) \cap U = U$. The condition (i) in the definition of feasible merging says $K = U$ is a complete set in H and this together with (a) and (b) allows one to see that every pair of distinct nodes of $U \cup I$ is an edge in H and, therefore, in H' . The condition (b) also implies $N \setminus (I \cup U) \subseteq \text{pa}_H(L) \setminus U$. Hence, by (ii) in the definition of feasible merging and (b), observe that $x \rightarrow y$ in H for every $x \in N \setminus (I \cup U)$ and $y \in I \cup U$. This is the condition (b) for $I \cup U$ and H' . In particular, $I \cup U$ is an idle set in H' .

The above observations allow one to show that H' and H have the same largest idle set. Then we can use Corollary 7 to get the desired claim. \square

5.3. Standard imset

An *imset* over N is an integer-valued function on the power set $\mathcal{P}(N) = \{A; A \subseteq N\}$ of N . Actually, it is nothing but a vector whose components are integers indexed by subsets of N . Given $A \subseteq N$, the symbol δ_A will denote an imset identifying the set A :

$$\delta_A(B) = \begin{cases} 1 & \text{if } B = A, \\ 0 & \text{otherwise.} \end{cases}$$

This notation allows us to define imsets over N as linear combinations of these basic units.

Let H be a chain graph over N which is LWF independence equivalent to an acyclic directed graph G over N . An important fact is that, for every component $C \in \mathcal{C}_H$, the respective closure graph $\bar{H}(C)$ is a decomposable graph – see Proposition 4.2 in [3]. Thus, the collection $\mathcal{K}_H(C)$ of cliques of $\bar{H}(C)$ can be ordered to satisfy the running intersection property. Let us denote by $\mathcal{S}_H(C)$ the class of *separators* in $\bar{H}(C)$ (see Section 2.1). Given $S \in \mathcal{S}_H(C)$ let $\nu_C(S)$ denote its multiplicity. The *standard imset* for H , denoted by u_H , is given by the following formula:

$$u_H = \delta_N - \delta_\emptyset + \sum_{C \in \mathcal{C}_H} \left\{ \delta_{\text{pa}_H(C)} - \sum_{K \in \mathcal{K}_H(C)} \delta_K + \sum_{S \in \mathcal{S}_H(C)} \nu_C(S) \cdot \delta_S \right\}. \quad (4)$$

A basic result on this concept is:

Proposition 20. Let G and H be LWF independence equivalent chain graphs over N such that there exists an acyclic directed graph over N LWF equivalent to them. Then $u_G = u_H$.

A technical proof of this result, based on the idea of feasible merging components, is presented in the Appendix (Section A4). Note the converse of Proposition 20 is also valid: if $u_G = u_H$ for chain graphs G and H that are (LWF independence) equivalent to acyclic directed graphs, then the graphs G and H are LWF independence equivalent. This can be derived from Proposition 20 and Corollary 7.1 in [17].

6. LEGAL MERGING COMPONENTS

This operation of merging components was at first introduced for chain graphs without flags, independently in [15] and in [11] under a different name "the removal of a strongly insubstantial arrowhead of a meta-arrow". This operation was later extended to general chain graphs in [12]. It can be viewed as an elementary operation with chain graphs attributed to a special strong equivalence of chain graphs. This equivalence, also introduced in [12], is a strengthening of both LWF independence equivalence and AMP independence equivalence.

6.1. Strong equivalence of chain graphs

Let G and H be chain graphs over N . We say that they are *strongly equivalent* if the following three conditions hold:

- G and H have the same underlying graph,
- an immorality $a \rightarrow c \leftarrow b$ ($[a, b]$ is not an edge) occurs in G iff it occurs in H ,
- a flag $a \rightarrow c - b$ ($[a, b]$ is not an edge) occurs in G iff it occurs in H .

It is evident (see Section 2.5) that if G and H are strongly equivalent then they are AMP independence equivalent. Moreover, it is easy to see that they also have to share complexes, for which reason they are LWF independence equivalent. On the other hand, two chain graphs that are both LWF and AMP independence equivalent need not be strongly equivalent.

A basic result on strong equivalence is analogous to Frydenberg's result [6] on LWF independence equivalence (reported in Theorem 2). The following result is proved as Corollary 5 in [12].

Theorem 21. Given an equivalence class \mathcal{H} of strong equivalence of chain graphs over N , there exists $H^\dagger \in \mathcal{H}$ which is the largest graph in \mathcal{H} .

6.2. The concept of legal merging components

Let H be a chain graph and L, U two of its components that form a meta-arrow $U \rightrightarrows L$. Merging (of) components of $U \rightrightarrows L$ will be called *legal* if the following three conditions hold:

- [i] $K \equiv \text{pa}_H(L) \cap U$ is a complete set in H ,
- [ii] for every $b \in K$ one has $\text{pa}_H(L) \setminus U = \text{pa}_H(b)$,
- [iii] for every $d \in L$ one has $\text{pa}_H(L) = \text{pa}_H(d)$.

We say that the hybrid graph H' resulting from the merging operation was obtained by *legal merging* (of) the (upper) component U and the (lower) component L .

Comparison with conditions (i) and (ii) in the definition of feasible merging (in Section 3) implies that every legal merging components is feasible. Indeed, [i] and (i) are identical and [ii] implies (ii). The converse implication does not hold: examples of feasible mergings which are not legal are given below.

Example 3. We give here some instances of legal and illegal merging components with respect to the graph (A) in Figure 5.

1. For the chain components $U = C_2$ and $L = C_3$, condition [i] is satisfied because $\text{pa}(L) \cap U = \{b, c\}$ is complete, condition [ii] is also satisfied because both $\text{pa}(L) \setminus U = \{a\} = \text{pa}(b)$ and $\text{pa}(L) \setminus U = \{a\} = \text{pa}(c)$ and, finally, condition [iii] is trivially true because $L = C_3$ has only one vertex. Consequently C_2 and C_3 can be legally merged and the resulting graph is given in Figure 5(B).
2. Merging components $U = C_1$ and $L = C_3$ is not feasible and, therefore, not legal.
3. Merging components $U = C_3$ and $L = C_4$ is feasible but not legal. More specifically, conditions [i] and (i) are satisfied because $\text{pa}(L) \cap U = \{e\}$ is trivially complete, condition (ii) is satisfied but not condition [ii] because $\text{pa}(L) \setminus U = \emptyset \subset \text{pa}(e) = \{a, b, c\}$. Also, condition [iii] is not satisfied because $\text{pa}(L) = \{e\} \neq \emptyset = \text{pa}(f)$. Note that merging components C_3 and C_4 would, for instance, create the flag $c \rightarrow e - g$ or destroy the flag $e \rightarrow g - f$.
4. Merging components $U = C_4$ and $L = C_5$ is also feasible but not legal. More specifically, conditions [i] and (i) are satisfied because $\text{pa}(L) \cap U = \{g, h\}$ is complete, condition (ii) and [ii] are satisfied because both $\text{pa}(L) \setminus U = \{e\} = \text{pa}(g)$ and $\text{pa}(L) \setminus U = \{e\} = \text{pa}(h)$ and, finally, condition [iii] is not satisfied because $\text{pa}(L) = \{e, g, h\} \neq \text{pa}(i) = \{g\}$. Note that merging the components C_4 and C_5 would, for instance, destroy the flag $g \rightarrow i - j$.

Consider now the chain graph (B) in Figure 5. In that graph, it is possible to legally merge components $U = C_1$ and $L = C_2 \cup C_3$, and the resulting graph is the largest graph in the strong equivalence class.

As every legal merging is feasible, it follows from Lemma 5 that the graph H' , obtained by legal merging components in a chain graph H , is also a chain graph. The below-mentioned results on legal merging are analogous to the results on feasible merging reported in Section 3.1.

The following lemma is proved as Proposition 6 in [12].

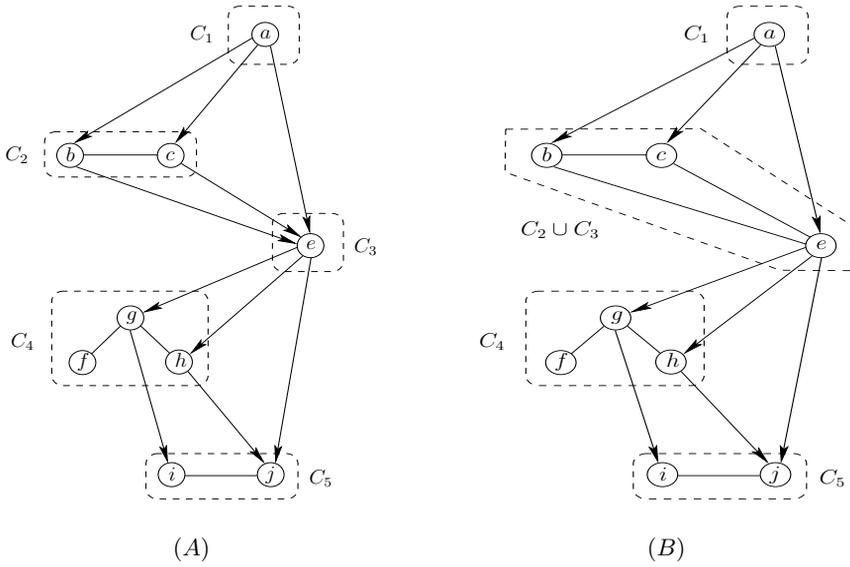


Fig. 5. The situation before, (A), and after, (B), a legal merging components.

Lemma 22. A graph obtained by merging components of a meta-arrow $U \rightrightarrows L$ in a chain graph H is a chain graph strongly equivalent to H iff the merging is legal.

This result justifies our terminology “legal merging”: the outcome of this operation is a graph in the same strong equivalence class of chain graphs. Actually, this justification is analogous to the case of feasible merging. To distinguish these two operations we use another adjective “legal” in the latter case. This reflects the relation of these two concepts: if something is legal, then it is usually feasible but not everything feasible is legal.

The following result is proved as Proposition 7 in [12].

Theorem 23. If G and H are strongly equivalent chain graphs and $H \geq G$ then there exists a sequence of chain graphs H_1, \dots, H_r , $r \geq 1$ starting with $G = H_1$ and ending with $H = H_r$ such that $\forall i = 1, \dots, r - 1$ the graph H_{i+1} is obtained from H_i by the operation of legal merging components (in H_i).

It follows easily from the above theorem:

Corollary 24. Let H be a chain graph, \mathcal{H} the equivalence class of strong equivalence containing H and H^\dagger the largest chain graph in \mathcal{H} . Then there exists a sequence $H = H_1, \dots, H_r = H^\dagger$, $r \geq 1$ in which each next graph is obtained from the previous one by the operation of legal merging components.

6.3. Legal splitting a component

Let M be a component in a chain graph H' over N partitioned into non-empty connected subsets U and L . Splitting M into the upper component U and the lower component L will be called *legal* if these three conditions are valid:

[i]' $K \equiv \text{ne}_{H'}(L) \cap U$ is a complete set in H' ,

[ii]' for every $b \in K$ one has $\text{pa}_{H'}(L) = \text{pa}_{H'}(b)$,

[iii]' for every $d \in L$ one has $\text{pa}_{H'}(d) = \text{pa}_{H'}(L)$ and $\text{ne}_{H'}(d) \cap U = K$.

Evidently, these conditions are the requirements [i]-[iii] on the graph H obtained from H' by the splitting operation, formulated in terms of H' . Therefore, it follows from Lemma 22:

Corollary 25. The chain graph obtained by splitting a component M of a chain graph H' is strongly equivalent to H' iff the splitting is legal.

6.4. Legal merging for chain graphs without flags

If H is a chain graph without flags then the operation of legal merging can be simplified. More specifically, the condition [iii] is automatically fulfilled and the condition [ii] takes a neat form. Thus, one only needs the following two conditions:

{i} $\text{pa}_H(L) \cap U$ is a complete set in H ,

{ii} $\text{pa}_H(L) \setminus U = \text{pa}_H(U)$.

Moreover, observe that if both components L and U are singletons, then {i} is automatically satisfied and {ii} turns into a well-known condition of Chickering's transformational characterisation [5] of equivalent acyclic directed graphs. In other words, legal merging components is closely related to the respective elementary operation of arrow converting with equivalent acyclic directed graphs. In this section, we interpret some former results on chain graphs without flags as consequences of the results reported in Sections 6.1 and 6.2.

A crucial observation is that if G and H are chain graphs without flags, then they are strongly equivalent iff they have the same underlying graph and immoralities. In particular, they are strongly equivalent iff they are LWF independence equivalent (cf. Theorem 1) and this occurs iff they are AMP independence equivalent (cf. Theorem 4). More concisely, we can say that strong equivalence, LWF independence equivalence and AMP independence equivalence coincide for chain graphs without flags. Thus, if one deals with equivalence of these graphs, no additional adjective is needed to indicate which type of equivalence is meant. Nevertheless, the strong equivalence appears to be the right equivalence for these graphs because every chain graph strongly equivalent to a chain graph without flags has to be a chain graph without flags. Of course, this is not true for LWF and AMP independence equivalences. In particular, Lemma 22 gives the following result, implied also by Theorem 1 in [15], respectively by Theorem 11 in [11].

Corollary 26. If H is a chain graph without flags then any graph H' which is obtained from H by the operation of legal merging components is also a chain graph without flags.

Another consequence of the fact that the strong equivalence class containing a chain graph without flags coincides with the respective equivalence class of chain graphs without flags can be derived from Theorem 21 (cf. Corollary 1 in [11]):

Corollary 27. Let \mathcal{H} be an equivalence class of chain graphs without flags. Then there exists a graph H^\ddagger which is the largest graph in \mathcal{H} .

Finally, Theorem 23 allows us to derive the following result, which alternatively follows from Theorems 10 and 11 in [11].

Corollary 28. If G and H are equivalent chain graphs without flags and $H \geq G$ then there exists a sequence of chain graphs H_1, \dots, H_r , $r \geq 1$ starting with $G = H_1$ and ending with $H = H_r$ such that $\forall i = 1, \dots, r - 1$ the graph H_{i+1} is obtained from H_i by the operation of legal merging components (in H_i). In particular, the largest chain graph H^\ddagger in an equivalence class \mathcal{H} of chain graphs without flags is attainable by the operation of legal merging components.

6.5. Essential graphs

The concept of an essential graph was introduced by Andersson, Madigan and Perlman [2] to describe uniquely in graphical terms an equivalence class of acyclic directed graphs. They also gave a graphical characterisation of these graphs, which implies that every essential graph is a chain graph. However, it has not been clear until recently whether the essential graph of an equivalence class of acyclic directed graphs can also be introduced as a distinguished member of the respective equivalence class of chain graphs. The following result has independently been achieved as Corollary 4 in [15] and as Theorem 6 in [11].

Proposition 29. If \mathcal{G} is an equivalence class of acyclic directed graphs and \mathcal{H} the equivalence class of chain graphs without flags such that $\mathcal{G} \subseteq \mathcal{H}$, then the essential graph G^* of \mathcal{G} is the largest chain graph in \mathcal{H} .

Thus, we can deduce from Corollary 28 the following consequence, which also implicitly follows from results in [11, 15].

Corollary 30. The essential graph G^* of an equivalence class \mathcal{G} of acyclic directed graphs is attainable by a series of operations of legal merging components from any equivalent chain graph without flags. In particular, it is attainable from any acyclic directed graph in \mathcal{G} .

The above observation is a basis for a special algorithm, described in § 6 of [15], which allows one to get the essential graph G^* of \mathcal{G} from any graph in \mathcal{G} as the result

of a series of legal merging operations determined uniquely by an ordering of nodes of the graph which is consistent with the direction of arrows. Another relevant fact is an alternative graphical characterisation of essential graphs obtained independently, as Theorem 2 in [15], and as Theorem 13 in [11].

Proposition 31. A graph H over N is the essential graph of an equivalence class of acyclic directed graphs over N iff it is a chain graph without flags such that the induced subgraph H_C is decomposable for every $C \in \mathcal{C}_H$ and no pair of components in H can be legally merged.

CONCLUSIONS

The paper was devoted to two elementary graphical operations with chain graphs (relative to the respective equivalence of graphs), namely to

- *feasible merging components*, which corresponds to LWF independence equivalence of chain graphs, and to
- *legal merging components*, which corresponds to a special strong equivalence of chain graphs.

The elementarity is meant in the sense that it is the simplest change (= graphical operation) which keeps the graphs equivalent in the respective sense. We have tried to convince the reader that the operations of this kind can be utilized to simplify substantially some proofs. The paper mainly deals with the operation of feasible merging and the LWF equivalence, but analogous ideas can be utilized in the case of legal merging and strong equivalence of chain graphs. Some of the achieved results, in particular, the results from Section 5, can be utilized in the area of learning Bayesian network models [16].

An analysis of the concept of factorisation with respect to a chain graph led to two different concepts of factorisation equivalence. One of those concepts, the one we consider to be natural, was shown to be identical with Markov equivalence. The graphical characterisation of the strong factorisation equivalence remains an open question for those who, unlike us, may consider this to be a fruitful concept.

APPENDIX: THE PROOFS

A1. Proof of Proposition 8

The first step is to modify the formulation of the pool-component rule (see Section 3.2) so that the proof will be easier. This is based on the following observation.

Fact 1. Let H be a chain graph, C its component, and a, b nodes in H . Then there exists $\kappa \in \mathcal{K}(C)$ with $\text{par}(\kappa) = \{a, b\}$ iff a, b are distinct nodes of $\text{pa}_H(C)$ that are not adjacent by an edge in H .

Proof. The necessity of the condition is evident. For the sufficiency proof, choose nodes $c \in \text{ch}_H(a) \cap C$, $d \in \text{ch}_H(b) \cap C$ and an undirected path $c = c_1, \dots, c_\ell = d$, $\ell \geq 1$ in H which cannot be shortened, that is, $[c_i, c_j]$ is not an edge in H whenever $|j - i| > 1$. Let us put $r = \max \{ i \leq \ell; [c_i, a] \text{ is an edge in } H \}$ and, moreover, $s = \min \{ j \geq r; [c_j, b] \text{ is an edge in } H \}$. It follows from the non-existence of semi-directed cycles in H that $a \rightarrow c_r \text{ --- } \dots \text{ --- } c_s \leftarrow b$ is a complex in H . \square

The previous observation allows us to simplify the definition of the set A from the formulation of the pool-component rule. Given a component C in a chain graph H , let us introduce:

$$\text{pa}_H^*(C) = \{ w \in \text{pa}_H(C); \exists v \in \text{pa}_H(C) \ v \neq w, [v, w] \text{ is not an edge in } H \}.$$

It follows from Fact 1 that the set $A = \text{pa}_H(C) \cap \bigcap_{\kappa \in \mathcal{K}(C)} \bigcap_{v \in \text{par}(\kappa)} \text{ch}_H(v)$ from the pool-component rule can be written as follows:

$$A = \text{pa}_H(C) \cap \bigcap_{w \in \text{pa}_H^*(C)} \text{ch}_H(w). \tag{5}$$

This will be utilised in the proof of Proposition 8.

Proof. We start with the *necessity proof*. Assume that a chain graph H' is obtained from a chain graph H by the application of the pool-component rule. The aim is to show that it can also be obtained from H by feasible merging.

More specifically, let C be a component in H such that the set A given by (5) is non-empty and B is a terminal component in H_A . The following two auxiliary observations are substantial.

$$\text{I. } b \in B, b \neq a \in \text{pa}_H(C) \Rightarrow [a, b] \text{ is an edge in } H.$$

This fact is easy to see by contradiction. If $[a, b]$ is not an edge in H then $b \in B \subseteq A \subseteq \text{pa}_H(C)$ implies $a \in \text{pa}_H^*(C)$. Hence, $b \in B \subseteq A \subseteq \bigcap_{w \in \text{pa}_H^*(C)} \text{ch}_H(w) \subseteq \text{ch}_H(a)$ gives a contradictory conclusion $a \rightarrow b$ in H .

$$\text{II. } b \in B, b \neq a \in \text{pa}_H(C) \cap \text{ds}_H(b) \Rightarrow a \in B.$$

Since $a \in \text{pa}_H(C)$, to show $a \in A$ it suffices to verify that $w \rightarrow a$ in H for any $w \in \text{pa}_H^*(C)$. The fact that $[a, w]$ is an edge in H can be shown by contradiction: otherwise observe $a \in \text{pa}_H^*(C)$ and hence $b \leftarrow a$ in H using (5) and the fact $b \in B \subseteq A$. Since H is a chain graph, this contradicts the assumption $a \in \text{ds}_H(b)$. Now, knowing that $[w, a]$ is an edge in H , the facts $b \in B \subseteq A \subseteq \text{ch}_H(w)$ and $a \in \text{ds}_H(b)$ imply $w \rightarrow a$ in H . Thus, the fact $a \in A$ has been verified.

By Step I, $[a, b]$ is an edge in H , and since $a \in \text{ds}_H(b)$, either $b \rightarrow a$ in H or $b \text{ --- } a$ in H . Provided that $b \rightarrow a$ in H , the fact $b \rightarrow a$ in H_A contradicts the assumption that B is a terminal component in H_A (cf. Section 2.1). Thus, $b \text{ --- } a$ in H_A , which means that a and b belong to the same component in H_A , that is, to B .

Recall that the application of the pool-component rule means that all arrows from B to C are replaced with lines. To show that this operation is equivalent to feasible

merging (of some) components, we put $L = C$ and define U as the component in H containing B . As $\emptyset \neq B \subseteq A \subseteq \text{pa}_H(C)$, there exists an arrow in H from U to $L = C$, that is, $U \rightrightarrows L$ is a meta-arrow. The conditions (i)-(ii) can be verified as follows.

$$\text{III. } B = \text{pa}_H(L) \cap U.$$

The inclusion $B \subseteq \text{pa}_H(L) \cap U$ is evident. For the converse inclusion, choose and fix $b \in B$. Given $a \in \text{pa}_H(L) \cap U$, $a \neq b$, the fact $a \in B$ follows from Step II because a and b are connected by an undirected path in H .

$$\text{IV. } B \text{ is a complete set in } H.$$

This follows from Step I and the fact B is a part of a component in a chain graph.

$$\text{V. } \forall b \in B \quad \text{pa}_H(L) \setminus U \subseteq \text{pa}_H(b).$$

Let us consider $a \in \text{pa}_H(L) \setminus U$. As $a \neq b$, $[a, b]$ is an edge in H , by Step I. By Step II, the hypothesis $a \in \text{ds}_H(b)$ leads to a contradictory conclusion $a \in B \subseteq U$. Therefore $a \rightarrow b$ in H , which was needed to show.

The facts above say that $B = K \equiv \text{pa}_H(L) \cap U$ and the conditions (i)–(ii) are fulfilled for $U \rightrightarrows L$. The application of the respective merging operation means that arrows from U to L are replaced with lines. These are the arrows from K to L and, as $K = B$, both operations are equivalent.

The second part is the *sufficiency proof*. Thus, assume that a chain graph H' is obtained from a chain graph H by feasible merging components of a meta-arrow $U \rightrightarrows L$. The aim is to show that it can also be obtained from H by the application of the pool-component rule. For this purpose, we choose a component C in H by putting $C = L$. One has to show that the set A given by (5) is non-empty. This follows from the next fact.

$$\text{VI. } b \in K \equiv \text{pa}_H(L) \cap U \Rightarrow b \in A.$$

Since $b \in \text{pa}_H(C)$, it suffices to show that $w \rightarrow b$ for every $w \in \text{pa}_H^*(C)$. Suppose for contradiction $w \in U$, that is, $w \in K$. We know that there exists $v \in \text{pa}_H(C)$, $w \neq v$, such that $[v, w]$ is not an edge in H . As $K \equiv U \cap \text{pa}_H(C)$ is complete in H by (i), $v \notin U$. The condition (ii) then implies a contradictory conclusion $v \in \text{pa}_H(L) \setminus U \subseteq \text{pa}_H(w)$. Thus, $w \notin U$, which implies by (ii) $w \in \text{pa}_H(L) \setminus U \subseteq \text{pa}_H(b)$, which was needed.

Step VI and the condition (i) imply that $K = \text{pa}_H(L) \cap U$ is a non-empty connected set in H_A . The next step is to show that K is a terminal component in H_A . This easily follows from the following fact.

$$\text{VII. } \forall b \in K \quad \forall a \in A \quad \text{either } a \in K \text{ or } a \rightarrow b \text{ in } H.$$

Of course, $a \in A \subseteq \text{pa}_H(L)$. If $a \in U$ then $a \in K$ by the definition. If $a \notin U$ then observe $a \in \text{pa}_H(L) \setminus U \subseteq \text{pa}_H(b)$, by (ii).

Now, Step VII implies both that K is a maximal connected subset in H_A and that there is no arrow from K in H_A . It suffices to put $B = K$ and observe that both operations are equivalent. \square

A2. Proof of Proposition 10

Two equivalent definitions of conditional independence in the discrete case will be used here. We leave the proof of the following fact to the reader.

Fact 2. Let $p : X_N \rightarrow [0, 1]$ be the density of a discrete probability measure P over N , $\langle A, B | C \rangle$ a triplet over N , and $p_{A|C}$ a version of conditional density of P . Then $A \perp\!\!\!\perp B | C [P]$ iff one of the following two conditions holds:

$$(a) \quad \forall x \in X_N \quad p_{A \cup B \cup C}(x_{A \cup B \cup C}) = p_{A|C}(x_A | x_C) \cdot p_{B \cup C}(x_{B \cup C}),$$

(b) $\forall w \in X_{A \cup B \cup C}$ with $p_{A \cup B \cup C}(w) > 0$ one has

$$p_{A \cup B \cup C}(w) = p_{B \cup C}(w_{B \cup C}) \cdot p_{A|C}(w_{A|C}) \cdot [p_C(w_C)]^{-1}.$$

Note that in the verification of Fact 2, as well as later, one repeatedly uses the vanishing principle for marginal densities of a discrete measure mentioned in Section 2.2. As concerns the implication (b) $\Rightarrow A \perp\!\!\!\perp B | C [P]$, the crucial fact is $\sum_{p_C > 0} p_{B \cup C} \cdot p_{A|C} \cdot p_C^{-1} = 1$, which allows one to observe that the desired equality holds even if $p_{A \cup B \cup C}(w) = 0 < p_C(w_C)$.

Let us recall the assumptions of Proposition 10. A chain of components $\pi : C_1, \dots, C_m$, $m \geq 1$ in a chain graph H over N is given. Discrete individual sample spaces X_i , $i \in N$ are fixed; the density p of a probability measure P is a function on $X_N \equiv \prod_{i \in N} X_i$. Let us introduce auxiliary notation: for $\ell = 1, \dots, m$, put $Z_\ell = \prod_{i \in C_\ell} X_i$, $A(\ell) = \bigcup_{j \leq \ell} C_j$, and $Y_\ell = \prod_{i \in A(\ell)} X_i \equiv \prod_{j \leq \ell} Z_j$. Let p_ℓ denote the marginal density of p on Y_ℓ .

Proof. We prove the following series of implications:

$$(\text{MC-}\pi) \Rightarrow (\text{FC-}\forall) \Rightarrow (\text{FC-}\exists) \Rightarrow (\text{MC-}\pi).$$

To show $(\text{MC-}\pi) \Rightarrow (\text{FC-}\forall)$, rewrite the assumption $(\text{MC-}\pi)$ as follows:

$$\forall \ell = 1, \dots, m \quad C_\ell \perp\!\!\!\perp (C_1 \cup \dots \cup C_{\ell-1}) \setminus \text{pa}_H(C_\ell) \mid \text{pa}_H(C_\ell) [P].$$

Having fixed a collection of versions of conditional densities $\{p_{C| \text{pa}_H(C)}; C \in \mathcal{C}_H\}$ in Fact 2(a), observe that

$$\forall \ell = 2, \dots, m \quad \forall x \in X_N \quad p_\ell(x_{A(\ell)}) = p_{C_\ell | \text{pa}_H(C_\ell)}(x_{C_\ell} | x_{\text{pa}_H(C_\ell)}) \cdot p_{\ell-1}(x_{A(\ell-1)}).$$

As $p_1 = p_{C_1 | \text{pa}_H(C_1)}$, by repeated substitution for $p_{\ell-1}$, a formula for $p = p_m$ is obtained: it is nothing but (2).

The implication $(\text{FC-}\forall) \Rightarrow (\text{FC-}\exists)$ is evident.

To show $(\text{FC-}\exists) \Rightarrow (\text{MC-}\pi)$, let us consider the corresponding collection of versions of conditional densities $\{p_{C| \text{pa}_H(C)}; C \in \mathcal{C}_H\}$ and replace every version with its absolute value, which is also a version of conditional density. That means we can

assume, without loss of generality, that $p_{C|\text{pa}_H(C)} \geq 0$ in (2) and introduce auxiliary non-negative partial products $f_\ell : \mathcal{Y}_\ell \rightarrow [0, \infty)$, $\ell = 1, \dots, m$ as follows:

$$\forall y \in \mathcal{Y}_\ell \equiv \mathcal{Y}_{\ell-1} \times \mathcal{Z}_\ell \quad f_\ell(y) = \prod_{i=1}^{\ell} p_{C_i|\text{pa}_H(C_i)}(y_{C_i}|y_{\text{pa}_H(C_i)}).$$

We prove the desired conclusion in three steps.

$$\text{I. } \forall \ell = 1, \dots, m \quad 0 \leq \sum_{y \in \mathcal{Y}_\ell, p_\ell(y) > 0} f_\ell(y) \leq 1.$$

The inequality $0 \leq \sum_{p_\ell > 0} f_\ell$ is ensured as explained above; since $f_1 \equiv p_1$ one has $\sum_{p_1 > 0} f_1 = 1$. It suffices to show $\sum_{p_\ell > 0} f_\ell \leq \sum_{p_{\ell-1} > 0} f_{\ell-1}$ for $\ell = 2, \dots, m$. More specifically, let us write, using the definition of f_ℓ :

$$\begin{aligned} \sum_{p_\ell > 0} f_\ell &= \sum_{(y,z) \in \mathcal{Y}_{\ell-1} \times \mathcal{Z}_\ell, p_\ell(y,z) > 0} f_\ell(y,z) = \sum_{y \in \mathcal{Y}_{\ell-1}, p_{\ell-1}(y) > 0} \sum_{z \in \mathcal{Z}_\ell, p_\ell(y,z) > 0} f_\ell(y,z) \\ &= \sum_{y \in \mathcal{Y}_{\ell-1}, p_{\ell-1}(y) > 0} \sum_{z \in \mathcal{Z}_\ell, p_\ell(y,z) > 0} f_{\ell-1}(y) \cdot p_{C_\ell|\text{pa}_H(C_\ell)}(z|y_{\text{pa}_H(C_\ell)}) \\ &= \sum_{y \in \mathcal{Y}_{\ell-1}, p_{\ell-1}(y) > 0} f_{\ell-1}(y) \cdot \sum_{z \in \mathcal{Z}_\ell, p_\ell(y,z) > 0} p_{C_\ell|\text{pa}_H(C_\ell)}(z|y_{\text{pa}_H(C_\ell)}). \end{aligned}$$

Now, having fixed $y \in \mathcal{Y}_{\ell-1}$ with $p_{\ell-1}(y) > 0$, by the vanishing principle for marginal densities, $p_{\text{pa}_H(C_\ell)}(y_{\text{pa}_H(C_\ell)}) > 0$, which implies $\sum_{z \in \mathcal{Z}_\ell} p_{C_\ell|\text{pa}_H(C_\ell)}(z|y_{\text{pa}_H(C_\ell)}) = 1$ by the definition of a version of conditional density. In particular, the internal sum in the last expression is a non-negative number less than or equal to 1. Since $f_{\ell-1} \geq 0$ it implies $\sum_{p_\ell > 0} f_\ell \leq \sum_{p_{\ell-1} > 0} f_{\ell-1}$.

$$\text{II. } \forall \ell = m, \dots, 1 \quad \forall y \in \mathcal{Y}_\ell \text{ if } p_\ell(y) > 0 \text{ then } p_\ell(y) = f_\ell(y).$$

We show this by reverse induction on ℓ . The assumption (FC- \exists) says $p_m(y) = f_m(y)$ for any $y \in \mathcal{Y}_m$. This implies the induction assumption for $\ell = m$. The induction step, saying that $[p_\ell = f_\ell \text{ whenever } p_\ell > 0] \Rightarrow [p_{\ell-1} = f_{\ell-1} \text{ whenever } p_{\ell-1} > 0]$ for $\ell = m, \dots, 2$, will be verified in two stages. First, we observe

$$\forall y \in \mathcal{Y}_{\ell-1} \quad p_{\ell-1}(y) > 0 \text{ implies } p_{\ell-1}(y) \leq f_{\ell-1}(y).$$

We can write this by the induction hypothesis and the definition of f_ℓ for a fixed $y \in \mathcal{Y}_{\ell-1}$ that satisfies $p_{\ell-1}(y) > 0$:

$$\begin{aligned} p_{\ell-1}(y) &= \sum_{z \in \mathcal{Z}_\ell, p_\ell(y,z) > 0} p_\ell(y,z) = \sum_{z \in \mathcal{Z}_\ell, p_\ell(y,z) > 0} f_\ell(y,z) \\ &= f_{\ell-1}(y) \cdot \sum_{z \in \mathcal{Z}_\ell, p_\ell(y,z) > 0} p_{C_\ell|\text{pa}_H(C_\ell)}(z|y_{\text{pa}_H(C_\ell)}). \end{aligned}$$

By the same arguments as in Step I, the last internal sum is a number between 0 and 1, which implies $p_{\ell-1}(y) \leq f_{\ell-1}(y)$. Second, we sum the obtained inequalities and use the inequality from Step I (for $\ell - 1$) to get

$$1 = \sum_{y \in \mathcal{Y}_{\ell-1}, p_{\ell-1}(y) > 0} p_{\ell-1}(y) \leq \sum_{y \in \mathcal{Y}_{\ell-1}, p_{\ell-1}(y) > 0} f_{\ell-1}(y) \leq 1.$$

As the upper and lower estimates above coincide, none of the summed inequalities is strict, which means $p_{\ell-1}(y) = f_{\ell-1}(y)$ whenever $p_{\ell-1}(y) > 0$.

III. $\forall \ell = 1, \dots, m \quad C_\ell \perp\!\!\!\perp (C_1 \cup \dots \cup C_{\ell-1}) \setminus \text{pa}_H(C_\ell) \mid \text{pa}_H(C_\ell) [P]$.

We verify this for $\ell \geq 2$ using Fact 2 (b). Indeed, having fixed $y \in Y_{\ell-1}$ and $z \in Z_\ell$ such that $p_\ell(y, z) > 0$, we can write by Step II and the definition of f_ℓ :

$$p_\ell(y, z) = f_\ell(y, z) = f_{\ell-1}(y) \cdot p_{C_\ell \mid \text{pa}_H(C_\ell)}(z \mid y_{\text{pa}_H(C_\ell)}).$$

Then we again observe that $p_{\text{pa}_H(C_\ell)}(y_{\text{pa}_H(C_\ell)}) > 0$, which allows us to write the value of the conditional density as the ratio of values of marginal densities. Moreover, by Step II, $f_{\ell-1}(y) = p_{\ell-1}(y)$, and we get

$$p_\ell(y, z) = p_{\ell-1}(y) \cdot p_{C_\ell \cup \text{pa}_H(C_\ell)}(z, y_{\text{pa}_H(C_\ell)}) \cdot [p_{\text{pa}_H(C_\ell)}(y_{\text{pa}_H(C_\ell)})]^{-1},$$

which is the required formula. □

A3. Proof of Proposition 15

The first step is the following lemma.

Lemma 32. Let H' be a chain graph obtained from a chain graph H over N by feasible merging components of a meta-arrow $U \rightrightarrows L$, and M denote the component in H' obtained by merging U and L . Then $\text{pa}_{H'}(L) \setminus U \subseteq \text{pa}_H(U)$, $\text{pa}_H(U) = \text{pa}_{H'}(M)$ and the closure graph $\bar{H}'(M)$ decomposes into the closure graphs $\bar{H}(U)$ and $\bar{H}(L)$, with a shared set of nodes $\text{pa}_H(L)$.

Proof. The condition (ii) in the definition of feasible merging gives $\text{pa}_H(L) \setminus U \subseteq \text{pa}_H(U)$, which implies both $\text{pa}_H(U) = \text{pa}_{H'}(M)$ and the fact that $\text{pa}_H(L)$ is the shared set of nodes of $\bar{H}(U)$ and $\bar{H}(L)$. We have to show that both $\bar{H}(U)$ and $\bar{H}(L)$ are induced subgraphs of $\bar{H}'(M)$, $\text{pa}_H(L)$ is a complete set in those graphs, and $\text{pa}_H(L)$ separates L from other nodes in $\bar{H}'(M)$.

The fact that H and H' have the same induced subgraph for $U \cup \text{pa}_H(U)$ and the fact $\text{pa}_H(U) = \text{pa}_{H'}(M)$ imply that $\bar{H}(U)$ is an induced subgraph of $\bar{H}'(M)$. To see that $\bar{H}(L)$ is an induced subgraph of $\bar{H}'(M)$, it suffices to verify that $\text{pa}_H(L)$ is a complete set in $\bar{H}'(M)$. However, this follows immediately from the conditions (i) and (ii) in the definition of feasible merging.

There is no arrow in H' from a node in L to a node in $M \cup \text{pa}_{H'}(M) = L \cup U \cup \text{pa}_H(U)$. In particular, since $L \cap \text{pa}_{H'}(M) = \emptyset$, it implies that the only nodes outside L adjacent with an edge in $\bar{H}'(M)$ to nodes in L are nodes of $\text{pa}_H(L)$. □

Recall that Proposition 15 claims that, under the assumptions of Lemma 32, the classes of factorisable densities on X_N with respect to H and H' coincide.

Proof. The first auxiliary fact is as follows.

- I. A disjoint triplet $\langle L, [U \cup \text{pa}_H(U)] \setminus \text{pa}_H(L) \mid \text{pa}_H(L) \rangle$ is represented in H according to the moralization criterion.

The respective set of nodes is $T = \text{an}_H(L \cup U \cup \text{pa}_H(U)) = \text{an}_H(L)$. Since L is a terminal component in the induced subgraph H_T , there is no arrow out of L in H_T . This also implies that none of possible ‘new’ edges in H_T^{mor} leads to a node in L . In particular, the only edges in H_T^{mor} connecting nodes in L with nodes outside L are the original edges in H_T between $\text{pa}_H(L)$ and L .

II. If the density p of a discrete probability measure P on X_N factorises with respect to H , then it factorises with respect to H' .

It follows from Proposition 14 that P is LWF Markovian with respect to H . Thus, by Step I,

$$L \perp\!\!\!\perp [U \cup \text{pa}_H(U)] \setminus \text{pa}_H(L) \mid \text{pa}_H(L) [P]. \tag{6}$$

This allows one to write, using set relations from Lemma 32, the definition of conditional independence and the vanishing principle for marginal densities:

$$\begin{aligned} \forall x \in X_N \quad p_{M \mid \text{pa}_{H'}(M)}^0(x_M \mid x_{\text{pa}_{H'}(M)}) \\ = p_{U \mid \text{pa}_H(U)}^0(x_U \mid x_{\text{pa}_H(U)}) \cdot p_{L \mid \text{pa}_H(L)}^0(x_L \mid x_{\text{pa}_H(L)}). \end{aligned}$$

Now, we can substitute this formula into the component-wise factorisation formula (2) with respect to H (with zero-versions of conditional densities) to get (2) with respect to H' .

As concerns the clique-wise factorisation for the merged component $M = U \cup L$ of H' , the fact (6) allows one to write using the convention $(0)^{-1} \equiv 0$:

$$\begin{aligned} p_{M \cup \text{pa}_{H'}(M)} &= p_{U \cup \text{pa}_H(U)} \cdot p_{L \cup \text{pa}_H(L)} \cdot (p_{\text{pa}_H(L)})^{-1} \\ &= \left(\prod_{K \in \mathcal{K}_H(U)} \varphi_K \right) \cdot \left(\prod_{K^* \in \mathcal{K}_H(L)} \varphi_{K^*}^* \right) \cdot (p_{\text{pa}_H(L)})^{-1}. \end{aligned}$$

The second line in the formula above follows from (3) for H . It follows from Lemma 32 that the collection of complete sets in $\bar{H}'(M)$ is the union of complete sets in $\bar{H}(U)$ and of complete sets in $\bar{H}(L)$. Hence, $p_{M \cup \text{pa}_{H'}(M)}$ factorises according to the graph $\bar{H}'(M)$. Thus, the condition (3) for H implies the same condition for H' .

III. If the density p of a discrete probability measure P on X_N factorises with respect to H' , then it factorises with respect to H .

It follows from Proposition 14 that P is LWF Markovian with respect to H' . By Lemma 5, H and H' are LWF independence equivalent, which implies that P is LWF Markovian with respect to H . Thus, (6) can be derived by Step I. Then we repeat the consideration in Step II to show that the component-wise factorisation (2) with respect to H' implies the same condition with respect to H .

Lemma 32 allows one to see that every complete set in $\bar{H}'(M)$ is either a subset of the set $U \cup \text{pa}_H(U)$ or a subset of the set $L \cup \text{pa}_H(L)$. That means it is either

a complete set in $\bar{H}(U)$ or a complete set in $\bar{H}(L)$. In particular, any $J \in \mathcal{K}_{H'}(M)$ disjoint with L belongs to $\mathcal{K}_H(U)$. Thus, it follows from the formula

$$p_{M \cup \text{pa}_{H'}(M)} = \prod_{J \in \mathcal{K}_{H'}(M)} \varphi_J = \prod_{J \in \mathcal{K}_{H'}(M), J \cap L \neq \emptyset} \varphi_J \cdot \prod_{J \in \mathcal{K}_{H'}(M), J \subseteq U \cup \text{pa}_H(U)} \varphi_J$$

by summing through configurations in \mathbf{X}_L that $p_{U \cup \text{pa}_H(U)}$ factorises according to $\bar{H}(U)$. Analogously, by summing through variables in $[U \cup \text{pa}_H(U)] \setminus \text{pa}_H(L)$, we can show that $p_{L \cup \text{pa}_H(L)}$ factorises according to $\bar{H}(L)$. Thus, (3) for H' implies (3) for H . \square

A4. Proof of Proposition 20

We need some auxiliary facts. Recall that an ordering a_1, \dots, a_n , $n \geq 1$ of nodes in an undirected graph F is called *perfect* if $\text{ne}_F(a_i) \cap \{a_1, \dots, a_{i-1}\}$ is a complete set in F for every $i = 2, \dots, n$.

Fact 3. Let F be a decomposable undirected graph over N , $A \subseteq N$ a complete set in F , and a_1, \dots, a_r , $r \geq 0$ an ordering of nodes in A . Then there exists a perfect ordering a_1, \dots, a_n , $n = |N|$ of nodes of F which starts by the section a_1, \dots, a_r .

Proof. The claim can be proved by induction on $|N|$. It is trivial if N is a complete set in F . If this is not the case then the well-known Dirac's lemma (see Lemma 2.9 in [9]) says that F has at least two non-adjacent simplicial nodes, that is, nodes a with complete $\text{ne}_F(a)$. Since A is complete, a simplicial node b of F exists in $N \setminus A$. The induced subgraph $F_{N \setminus \{b\}}$ is decomposable, and the respective ordering of its nodes a_1, \dots, a_{n-1} exists by the induction hypothesis. Put $a_n = b$ to get the ordering for F . \square

Lemma 33. Let H be a chain graph over N which is LWF independence equivalent to an acyclic directed graph, and $C \in \mathcal{C}_H$ be one of its components. Then there exists $c \in C$ such that $\text{pa}_H(c) = \text{pa}_H(C)$.

Proof. We already know, by Proposition 4.2 in [3], that the closure graph $F \equiv \bar{H}(C)$ is decomposable and that $A \equiv \text{pa}_H(C)$ is complete in F . Let us apply Fact 3 to find a perfect ordering a_1, \dots, a_n , $n \geq 1$ of nodes of F such that $A = \{a_i; i \leq r\}$ for $r = |A|$. In particular, $C = \{a_i; i \geq r + 1\}$. The first observation is as follows.

I. $\forall i, r + 1 < i \leq n \quad \exists j, r + 1 \leq j < i$ such that $a_i - a_j$ in F .

Since C is a component in H , it is a connected set in F . Thus, there exists a path in F_C connecting a_i and $B \equiv \{a_k; r + 1 \leq k < i\}$. Let us consider a path ρ of this kind which cannot be shortened. Then ρ does not involve nodes of $\{a_k; k > i\}$. This can be shown by contradiction: otherwise put $\ell = \max\{k; a_k \text{ is a node of } \rho\}$ and observe that, if $\ell > i$ then both the preceding node and the subsequent node of a_ℓ in ρ belongs to $\text{ne}_F(a_\ell) \cap \{a_1, \dots, a_{\ell-1}\}$. Since the ordering is perfect, these nodes form a line in F_C and ρ can be shortened, which contradicts the assumption. The fact that ρ consists of nodes of $B \cup \{a_i\}$ implies what is needed.

II. $ne_F(a_{r+1}) \cap A = A$.

A basic fact is that, if $a_i - a_j$ in F for $r + 1 \leq j < i$, then

$$ne_F(a_i) \cap A \subseteq ne_F(a_j) \cap A.$$

Indeed, both a_j and the nodes of $ne_F(a_i) \cap A$ belong to $ne_F(a_i) \cap \{a_1, \dots, a_{i-1}\}$. Since a_1, \dots, a_n is a perfect ordering, $ne_F(a_i) \cap \{a_1, \dots, a_{i-1}\}$ is a complete set in F , which implies that a_j is adjacent to all nodes of $ne_F(a_i) \cap A$. This basic fact together with Step I allows one to observe $ne_F(a_i) \cap A \subseteq ne_F(a_{r+1}) \cap A$ for every $i \geq r + 1$. On the other hand, the fact that for every $d \in A = pa_H(C)$ there exists $e \in C = \{a_i; i \geq r + 1\}$ with $d \rightarrow e$ in H , implies $A \subseteq \bigcup_{i \geq r+1} ne_F(a_i)$. Hence, one has $A \subseteq \bigcup_{i \geq r+1} ne_F(a_i) \cap A \subseteq ne_F(a_{r+1}) \cap A$.

It follows easily from Step II that $c = a_{r+1}$ is the node with $pa_H(c) = pa_H(C)$. \square

Now, the proof of Proposition 20 follows.

Proof. Suppose that H is a chain graph over N which is LWF independence equivalent to an acyclic directed graph. Let H' be a graph obtained from H be feasible merging components of a meta-arrow $U \rightrightarrows L$. Let $M \equiv U \cup L$ denote the merged component in H' . We are going to show that $u_H = u_{H'}$. Since H and H' have the same closure graphs for all components except for the closure graphs for U , L and M , it follows from (4) that the contribution in $u_{H'}$ corresponding to M must be shown to be the sum of contributions in u_H corresponding to U and L . Because $pa_{H'}(M) = pa_H(U)$ (see Lemma 32), it reduces to this formula:

$$\begin{aligned} & - \sum_{K \in \mathcal{K}_{H'}(M)} \delta_K + \sum_{S \in \mathcal{S}_{H'}(M)} \nu'_M(S) \cdot \delta_S \\ & = - \sum_{K \in \mathcal{K}_H(L)} \delta_K + \sum_{S \in \mathcal{S}_H(L)} \nu_L(S) \cdot \delta_S + \delta_{pa_H(L)} \quad (7) \\ & \quad - \sum_{K \in \mathcal{K}_H(U)} \delta_K + \sum_{S \in \mathcal{S}_H(U)} \nu_U(S) \cdot \delta_S, \end{aligned}$$

where $\nu'_M(S)$ denotes the multiplicity of a separator $S \in \mathcal{S}_{H'}(M)$ in $\bar{H}'(M)$. The first step is the following observation.

- I. A set $K \subseteq N$ is a clique of $\bar{H}'(M)$ iff it is either a clique of $\bar{H}(L)$ or a clique of $\bar{H}(U)$ distinct from $pa_H(L)$. There is no set which is simultaneously a clique of $\bar{H}(L)$ and a clique of $\bar{H}(U)$.

It follows from Lemma 32 that the collection of complete sets in $\bar{H}'(M)$ is the union of the class of complete sets in $\bar{H}(L)$ and the class of complete sets in $\bar{H}(U)$. Moreover, the intersection of these two classes is the class of subsets of $pa_H(L)$. Recall we assume that H is a chain graph LWF independence equivalent to an acyclic directed graph. Thus, one can apply Lemma 33, which says that there exists $c \in L$ such that $\{c\} \cup pa_H(L)$ is complete in $\bar{H}(L)$. In particular, $pa_H(L)$ is never a clique of $\bar{H}(L)$. On the other hand, $pa_H(L)$ could be a clique of $\bar{H}(U)$, but then it would not be a clique of $\bar{H}'(M)$. These observations imply what is claimed.

- II. We choose an ordering K_1^L, \dots, K_r^L , $r \geq 1$ of cliques of $\bar{H}(L)$ satisfying the running intersection property, and an ordering K_1^U, \dots, K_s^U , $s \geq 1$ of cliques of $\bar{H}(U)$ satisfying this property such that $\text{pa}_H(L) \subseteq K_1^U$.

Since both $\bar{H}(L)$ and $\bar{H}(U)$ are decomposable graphs, by Proposition 4.2 in [3], it is possible to use Lemma 2.18 in [9]. It says that, for every clique K of a decomposable graph F , there exists an ordering of cliques of F satisfying the running intersection property which starts with K .

- III. If $\text{pa}_H(L)$ is not a clique of $\bar{H}(U)$, then $K_1^L, \dots, K_r^L, K_1^U, \dots, K_s^U$ is an ordering of cliques of $\bar{H}(M)$ satisfying the running intersection property and the formula (7) holds.

Of course, the section K_1^L, \dots, K_r^L satisfies the respective requirements, and the corresponding separators $S_k^L = K_k^L \cap (\bigcup_{i < k} K_i^L)$, $r \geq k \geq 2$ remain unchanged. The basic observation is that, for any $j \geq 1$,

$$K_j^U \cap \left(\bigcup_{k \leq r} K_k^L \right) \subseteq [U \cup \text{pa}_H(U)] \cap [L \cup \text{pa}_H(L)] = \text{pa}_H(L) \subseteq K_1^U.$$

As concerns K_1^U , the facts $\text{pa}_H(L) \subseteq K_1^U$ and $\text{pa}_H(L) \subseteq L \cup \text{pa}_H(L) = \bigcup_{k \leq r} K_k^L$ allow one to see $K_1^U \cap (\bigcup_{k \leq r} K_k^L) = \text{pa}_H(L)$. Of course, there exists $i \leq r$ with $\text{pa}_H(L) \subseteq K_i^L$, and the respective separator is $S_1^U = \text{pa}_H(L)$.

As concerns $j \geq 2$, write $K_j^U \cap [\bigcup_{k \leq r} K_k^L \cup \bigcup_{i < j} K_i^U] = K_j^U \cap \bigcup_{i < j} K_i^U \equiv S_j^U$. That means, the section K_2^U, \dots, K_s^U also satisfies the respective requirements, and separators are preserved. The above-mentioned facts give (7).

- IV. If $\text{pa}_H(L)$ is a clique of $\bar{H}(U)$ then $K_1^L, \dots, K_r^L, K_2^U, \dots, K_s^U$ is an ordering of cliques of $\bar{H}(M)$ satisfying the running intersection property, and the formula (7) holds.

As in Step III, there is no problem with the section K_1^L, \dots, K_r^L . The fact that $K_1^U = \text{pa}_H(L)$ allows one to write for $j \geq 2$:

$$K_j^U \cap K_1^U = K_j^U \cap \text{pa}_H(L) \subseteq K_j^U \cap \left(\bigcup_{k \leq r} K_k^L \right).$$

The inverse inclusion $K_j^U \cap \bigcup_{k \leq r} K_k^L \subseteq K_j^U \cap \text{pa}_H(L) = K_j^U \cap K_1^U$ can be obtained, as in Step III. In particular, $K_j^U \cap [\bigcup_{k \leq r} K_k^L \cup \bigcup_{2 \leq i < j} K_i^U] = K_j^U \cap \bigcup_{1 \leq i < j} K_i^U \equiv S_j^U$, which means that the respective separators for K_2^U, \dots, K_s^U are preserved.

As concerns K_2^U , there exists $i \leq r$ with $\text{pa}_H(L) \subseteq K_i^L$, which implies $S_2^U \subseteq K_1^U = \text{pa}_H(L) \subseteq K_i^L$. Thus, the section K_2^U, \dots, K_s^U also satisfies the respective requirements. Thus, (7) is evident if one realizes that the term $+\delta_{\text{pa}_H(L)}$ is cancelled against $-\delta_{K_1^U}$.

Thus, we have thus shown $u_H = u_{H'}$ whenever H' is obtained from H (which is LWF independence equivalent to an acyclic directed graph) by feasible merging

components. It follows from Corollary 7 that, given an LWF independence equivalence class \mathcal{H} containing an acyclic directed graph, one has $u_H = u_{H_\infty}$ for any $H \in \mathcal{H}$. This implies what Proposition 20 says. \square

ACKNOWLEDGEMENT

This work was supported by the Czech Science Foundation through Grant 201/04/0393. The work of the second author was supported by Ministero dell'Università e della Ricerca, Rome, through Programma di Ricerca PRIN-2007AYHZWC.

(Received July 25, 2006.)

REFERENCES

- [1] S. A. Andersson, D. Madigan, and M. D. Perlman: An alternative Markov property for chain graphs. In: *Uncertainty in Artificial Intelligence, Proc. Twelfth Conference* (F. Jensen and E. Horvitz, eds.), Morgan Kaufmann, San Francisco 1996, pp. 40–48.
- [2] S. A. Andersson, D. Madigan, and M. D. Perlman: A characterization of Markov equivalence classes for acyclic digraphs. *Ann. Statist.* *25* (1997), 505–541.
- [3] S. A. Andersson, D. Madigan, and M. D. Perlman: On the Markov equivalence of chain graphs, undirected graphs and acyclic digraphs. *Scand. J. Statist.* *24* (1997), 81–102.
- [4] S. A. Andersson, D. Madigan, and M. D. Perlman: Alternative Markov properties for chain graphs. *Scand. J. Statist.* *28* (2001), 33–85.
- [5] D. M. Chickering: A transformational characterization of equivalent Bayesian network structures. In: *Uncertainty in Artificial Intelligence, Proc. Eleventh Conference* (P. Besnard and S. Hanks, eds.), Morgan Kaufmann, San Francisco 1995, pp. 87–98.
- [6] M. Frydenberg: The chain graph Markov property. *Scand. J. Statist.* *17* (1990), 333–353.
- [7] S. L. Lauritzen and N. Wermuth: Mixed Interaction Models. Research Report No. R-84-8, Inst. Elec. Sys., University of Aalborg 1984.
- [8] S. L. Lauritzen and N. Wermuth: Graphical models for association between variables, some of which are qualitative and some quantitative. *Ann. Statist.* *17* (1989), 31–57.
- [9] S. L. Lauritzen: *Graphical Models*. Clarendon Press, Oxford 1996.
- [10] J. Pearl: *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, San Mateo 1988.
- [11] A. Roverato: A unified approach to the characterisation of equivalence classes of DAGs, chain graphs with no flags and chain graphs. *Scand. J. Statist.* *32* (2005), 295–312.
- [12] A. Roverato and M. Studený: A graphical representation of equivalence classes of AMP chain graphs. *J. Machine Learning Research* *7* (2006), 1045–1078.
- [13] Š. Štěpánová: *Equivalence of Chain Graphs (in Czech)*. Diploma Thesis, Charles University, Prague 2003.
- [14] M. Studený: A recovery algorithm for chain graphs. *Internat. J. Approx. Reasoning* *17* (1997), 265–293.
- [15] M. Studený: Characterization of essential graphs by means of the operation of legal merging of components. *Internat. J. Uncertainty, Fuzziness and Knowledge-Based Systems* *12* (2004), 43–62.
- [16] M. Studený and J. Vomlel: Transition between graphical and algebraic representatives of Bayesian network models. In: *Proc. 2nd European Workshop on Probabilistic Graphical Models* (P. Lucas ed.), Leiden 2004, pp. 193–200.

- [17] M. Studený: Probabilistic Conditional Independence Structures. Springer-Verlag, London 2005.
- [18] T. Verma and J. Pearl: Equivalence and synthesis of causal models. In: Uncertainty in Artificial Intelligence, Proc. Sixth Conference (P. Bonissone, M. Henrion, L. N. Kanal, and J. F. Lemmer, eds.), North-Holland, Amsterdam 1991, pp. 255–270.
- [19] M. Volf and M. Studený: A graphical characterization of the largest chain graphs. *Internat. J. Approx. Reasoning* 20 (1999), 209–236.

*Milan Studený, Institute of Information Theory and Automation – Academy of Sciences of the Czech Republic, Pod Vodárenskou věží 4, 182 08 Praha 8. Czech Republic.
e-mail: studeny@utia.cas.cz*

*Alberto Roverato, Dipartimento di Scienze Statistiche, Università di Bologna, Via Belle Arti 41, 40126 Bologna. Italy.
e-mail: alberto.roverato@unibo.it*

*Šárka Štěpánová, Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics – Charles University, Sokolovská 83, 186 75 Praha 8. Czech Republic.
e-mail: sarka.stepanova@matfyz.cz*