Available at WWW.MATHEMATICSWEB.ORG
POWERED BY SCIENCE DIRECT

Journal of Statistical Planning and Inference 115 (2003) 235–259

journal of statistical planning and inference

www.elsevier.com/locate/jspi

A characterization of moral transitive acyclic directed graph Markov models as labeled trees

Robert Castelo *, Arno Siebes

Department of Computer Science, Institute of Information and Computing Sciences, University of Utrecht, P.O. Box 80089, 3508 TB Utrecht, The Netherlands

Received 16 March 2001; received in revised form 13 November 2001; accepted 17 December 2001

Abstract

It follows from the known relationships among the different classes of graphical Markov models for conditional independence that the intersection of the classes of moral acyclic directed graph Markov models (or decomposable ≡ DEC Markov models), and transitive acyclic directed graph ≡ TDAG Markov models (or lattice conditional independence ≡ LCI Markov models) is non-empty. This paper shows that the conditional independence models in the intersection can be characterized as labeled trees. This fact leads to the definition of a specific Markov property for labeled trees and therefore to the introduction of labeled trees as part of the family of graphical Markov models. © 2002 Elsevier Science B.V. All rights reserved.

MSC: 62H99

Keywords: Graphical Markov model; Conditional independence; Multivariate distribution; Undirected graph Markov model; Acyclic directed graph Markov model; Transitive acyclic directed graph Markov model; Decomposable Markov model; Lattice conditional independence Markov model; Tree conditional independence Markov model; Finite distributive lattice; Poset; Labeled tree

1. Introduction

Graphical Markov models are a powerful tool for the representation and analysis of conditional independence among variables of a multivariate distribution. There are different classes of graphical Markov models. Each class is associated with a different type of graph, which embodies the structural (qualitative) information on the

* Corresponding author. Current address: Biomedical Informatics Group (GRIB), Pompeu Fabra University, Dr. Aiguader 80, E-08003 Barcelona, Spain. Tel.: +34-93-221-1009; fax: +34-93-221-3237.

E-mail addresses: rcastelo@imim.es (R. Castelo), siebes@cs.un.nl (A. Siebes).

PII: S0378-3758(02)00143-X

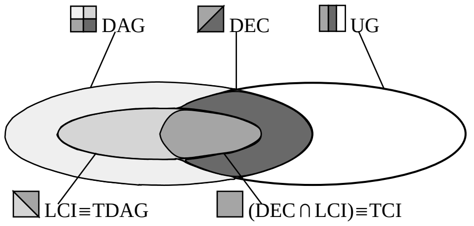


Fig. 1. Relation among the classes DAG, UG, DEC and LCI models.

relationships among the variables involved. More precisely, every vertex of the associated graph corresponds to a random variable of the multivariate distribution.

One of the most fascinating aspects is the algebraic structure that underlies the broad spectrum of different classes of graphical Markov models. This underlying algebraic structure is the foundation on which the present paper develops a particular characterization of the intersection of certain classes of graphical Markov models (and for which positivity or existence of joint densities is not required). The reader may find a guide to some different types of graphical Markov models in the books of Pearl (1988); Whittaker (1990); Cox and Wermuth (1996) and Lauritzen (1996).

In this paper we will deal with graphical Markov models defined by undirected graphs (UG models), acyclic directed graphs (DAG [1] models), chordal graphs (decomposable or DEC models), transitive directed acyclic graphs (TDAG models), and finite distributive lattices (lattice conditional independence or LCI models). In the next section the reader will find precise graph-theoretical definitions of these graphs.

LCI models were introduced by Andersson and Perlman (1993) in the context of the analysis of non-nested multivariate missing data patterns and non-nested dependent linear regression models. Later, Andersson et al. (1997b), Theorem 4.1 showed that the class of LCI models coincides with the class of TDAG models. Either of these terms, TDAG or LCI, will be used here depending on the algebraic context used at the moment.

Fig. 1 shows a picture that Andersson et al. (1995) devised in order to describe the location of LCI models within the scope of models represented by undirected and directed graphs. Although the class of LCI models appears on the picture as an isolated subclass, Andersson et al. (1995, p. 38) show that they are in fact interlaced through the class of DAG models. An important characterization also depicted in this figure corresponds to the definition of those DAG models that are equivalent to some UG model (Wermuth, 1980; Kiiveri et al., 1984). Thus, undirected and directed graphs in this intersecting class describe the same model of conditional independence. They are graphically determined by chordal ($\equiv$ decomposable) graphs and are known as DEC models. More concretely, one characterizes those DAG models equivalent to some UG models, as those determined by a DAG that does not contain immoralities, i.e. a

[1] Sometimes also referred as ADG.

moral DAG. In the graph-theoretic context, moral DAGs are known as *subtree acyclic digraphs* and were introduced by Harary et al. (1992).

With the exception of LCI models, graphical Markov models are usually determined by a graph (e.g. UG, DAG and DEC models) which is interpreted in terms of *separation*. Separation (see Lauritzen et al., 1990; Lauritzen, 1996) is a graphical notion that allows one to split the vertex set of the graph into a triplet that maps to the ternary relationship of conditional independence. The LCI models introduced by Andersson and Perlman (1993) are not graphical, thus they do not have a graph-separation interpretation, but another one that manipulates directly the lattice that determines the model. Andersson et al. (1997b) showed that in fact the LCI representation is equivalent to a graphical representation via TDAGs that does have a graph-separation interpretation.

In this paper we consider another special class of graphical Markov models, namely DEC $\cap$ LCI. Andersson et al. (1995) proved that DEC $\cap$ LCI $\neq \emptyset$ and for this particular class, we introduce an alternative graphical representation which is not interpreted in graph-separation terms. The new representation is more economical, in the sense that it affords an easier interpretation of the model of conditional independence, in contrast to its equivalent graphical counterpart in terms of either TDAGs or chordal graphs, or its equivalent non-graphical counterpart in terms of finite distributive lattices.

This new representation is based on a characterization of moral TDAGs as labeled trees, which will be presented first. Afterwards, a Markov property for labeled trees will be introduced. Finally, the relationship between this new Markov property and the rest of the existing Markov properties is investigated. From this study follows the new formalization of the graphical Markov models in DEC $\cap$ LCI. Because of the relation between trees and models for conditional independence, we will refer to DEC $\cap$ LCI models as *tree conditional independence* $\equiv$ TCI models.

The direct consequence of such a formalization is that it provides a different way to read the structural information ($\equiv$ the conditional independencies) contained in the model, by using the new associated Markov property.

The layout of the paper is as follows. In the next section we introduce the background concepts, terminology and notation, used throughout the paper, regarding graphs and lattices. Afterwards, in Section 3, we will obtain the characterization of moral TDAGs as labeled trees. In Section 4 we will introduce the new specific Markov property for labeled trees, and its relationship with respect to other Markov properties will be investigated. In Section 5, the notion of Markov equivalence in this setting is introduced and finally, in Section 6, the main issues of the paper are summarized. We have included an appendix that reviews the most relevant notions about graphical Markov models.

2. Background concepts, terminology and notation

The graphical terminology and notation has been mainly borrowed from Lauritzen (1996) and Andersson et al. (1995), whereas the concepts regarding finite distributive lattices have been taken from Grätzer (1978) and Davey and Priestley (1990). For more details, the reader is referred to these sources.

A graph G is a pair (V,E) where V is the set of vertices and E the set of edges. The set of vertices V indexes a vector of random variables $\mathbf{X}_V = \{X_1,\dots,X_n\}$ that form a multivariate distribution member of a family P. This family P of multivariate distributions is defined on some product space $\mathscr{X} = \times(\mathscr{X}_i | i \in V)$. For simplicity, we will refer to a random variable, or a set of them, by their indexes, i.e. X_i as i and $\mathbf{X}_A$ as A.

The set of edges E is a subset of the set of ordered pairs $\{V \times V\}$ such that it does not contain loops, i.e. $(a,a) \notin E$. An undirected edge $a - b$ in E implies that $(a,b) \in E$ and $(b,a) \in E$, whereas the directed version $a \to b$ implies that $(a,b) \in E$ and $(b,a) \notin E$. We use the standard terminology of Lauritzen (1996) regarding: *subgraph*, *adjacency*, *boundary* $\equiv \mathrm{bd}(v)$, *closure* $\equiv \mathrm{cl}(v)$, *complete graph*, *clique*,[2] *parent vertex*, *child vertex*, *parent set* $\equiv \mathrm{pa}(v)$, *immorality*, *moral graph*, *moralized graph* $\equiv G^{\mathrm{m}}$, *path*, *undirected cycle*, *undirected graph* $\equiv$ UG, *directed cycle*, *acyclic digraph* $\equiv$ DAG or ADG, *ancestor set* $\equiv \mathrm{an}(v)$, *descendant set* $\equiv \mathrm{de}(v)$, *non-descendant set* $\equiv \mathrm{nd}(v)$.

An *undirected cycle* is a path where $a = b$. A *tree* is a connected undirected graph without undirected cycles. In this case there is always a unique path between any two different vertices. A *rooted tree* is a tree in which a hierarchy among the vertices is created. One of the vertices of a rooted tree is the *root* and it is placed at the bottom of the hierarchy. The *leaves* of a rooted tree are those vertices connected to just one other vertex; they are placed at the top of the hierarchy. Under this convention we will say that the root is *below* the leaves, and the leaves are *above* the root. Given a tree $T = (V,E)$ and a vertex $u \in V$, a *subtree* rooted at u, denoted as T_u, is the pair $T_u = (U, E_U)$, where the vertex set $U \subseteq V$ contains all vertices involved in every path from u to the leaves above, and the edge set $E_U = E \cap (U \times U)$.

A DAG is said to be *transitive* $\equiv$ TDAG if for every vertex v, $\mathrm{pa}(v) = \mathrm{an}(v)$. A TDAG is *moral* if it contains no immoralities. For any moral TDAG and every vertex v, the induced subgraph $\{v\} \cup \mathrm{pa}(v)$ $(\{v\} \cup \mathrm{an}(v))$ is complete.

An undirected graph is *chordal*, or *decomposable* (DEC), iff it does not contain undirected cycles on more than three vertices without a chord. They are also known as *triangulated* graphs or *rigid circuit* graphs. In the introduction we already mentioned that DEC models correspond to the intersection of the classes of DAG and UG models, and therefore they characterize those UG models that are equivalent to some DAG model.

An important concept regarding directed graphs is *ancestral set*. Let $G = (V,E)$ be a DAG. Given a subset $A \subseteq V$, A is said to be *ancestral* iff for every vertex $v \in A$, $\mathrm{an}(v) \subseteq A$. Since the union and intersection of ancestral sets is again ancestral, all the different ancestral sets contained in a DAG $G = (V,E)$ form a ring of subsets of V, which is denoted as $\mathscr{A}(G)$. Further, given a subset of vertices $A \subset V$, the subset $\mathrm{An}(A) \subseteq V$ will denote the *smallest ancestral set* that contains A. To avoid confusion, note that $\mathrm{an}(v)$ refers to the set of vertices that are ancestors of the vertex v, while $\mathrm{An}(A)$ refers to the smallest subset $\mathrm{An}(A) \subseteq V$ that contains a given subset $A \subset V$ such that $\mathrm{An}(A)$ is ancestral in G.

[2] In the literature of graphical Markov models one finds the concept of *clique* referred to as *maximal clique* in the graph-theoretic literature.

A partially ordered set ($\equiv$ poset) $(S, \leqslant)$ is a set S equipped with an order relation [3] $\leqslant$. If the poset is *totally ordered*, i.e. $\forall a, b \in S$ $a \leqslant b$ *or* $b \leqslant a$, then it is a *chain*. A chain C in a poset S is called *maximal* iff, for any chain $D \in S$, $C \subseteq D$ implies that $C = D$. Let S be a poset and let $x, y \in S$. We say x is *covered* by y, and write $x \prec y$ if $x < y$ and $x \leqslant z < y \Rightarrow z = x$.

Grätzer (1978, p. 10) shows that this covering relation determines the partial ordering in a given poset in the following way. Let S be a finite poset. Then $a \leqslant b$ iff $a = b$ or there exists a finite sequence of elements $x_0, \dots, x_{n-1}$, such that $x_0 = a$, $x_{n-1} = b$, and $x_i \prec x_{i+1}$, for $0 \leqslant i < n-1$.

A poset $(S, \leqslant)$ has an associated undirected graph (V, E) in which $(x, y) \in E$ if $x \prec y$ (y covers x). This associated undirected graph is called the *covering graph* of the poset S. A *Hasse* diagram of a poset S is a representation of the covering graph of S in the plane such that if $x < y$, then x is below y in the plane.

An *envelope* [4] E of a poset $(S, \leqslant)$ is a subset $E \subseteq S$ such that for every $s \in S$, there exists $e \in E$ such that $s \leqslant e$. A *minimal envelope* (see footnote 4) E^* of a poset $(S, \leqslant)$ is an envelope of $(S, \leqslant)$ such that there is no subset $E \subseteq E^*$, that is an envelope of $(S, \leqslant)$ too.

Given a poset S, a subset $H \subseteq S$ and an element $a \in S$, it is said that a is an *upper bound* (*lower bound*) of H iff for every $h \in H$, $h \leqslant a$ ($h \geqslant a$). An upper bound (lower bound) a of H is the *least upper bound* (*greatest lower bound*) of H or *supremum* (*infimum*) of H iff, for any upper bound (lower bound) b of H, we have $a \leqslant b$ ($a \geqslant b$), and denote it by $a = \sup H$ ($a = \inf H$).

It is possible to define a *lattice* in different ways. We will introduce here just one of them, as follows. A poset L is a *lattice* iff $\sup H$ and $\inf H$ exist in L for any finite non-void subset H of L. Grätzer (1978) shows that the concept of a lattice as a poset is equivalent to the concept of a lattice as an algebra $\mathscr{K} \equiv \mathscr{K}(\wedge, \vee)$, where $\wedge$ and $\vee$ are binary operations on pairs of elements $a, b \in \mathscr{K}$, corresponding to $\inf\{a, b\}$ and $\sup\{a, b\}$, respectively. The operations $\wedge, \vee$ are idempotent, commutative and associative, and satisfy two absorption identities. It has been already mentioned that LCI models are determined by finite distributive lattices. Birkhoff (Grätzer, 1978, p. 62) characterized finite distributive lattices as those isomorphic to a ring of sets.

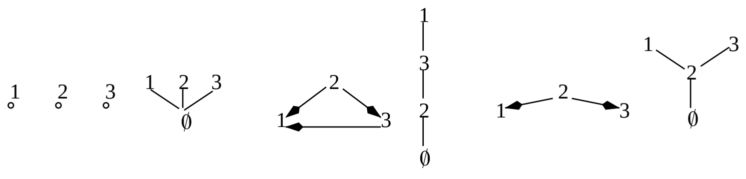
A finite distributive lattice $\mathscr{K}$ has a unique irredundant representation in terms of a finite poset $(J(\mathscr{K}), \leqslant)$, where $J(\mathscr{K}) \subseteq \mathscr{K}$ is the subset of *join-irreducible* elements (see Grätzer, 1978, p. 62). This poset is often substantially smaller than $\mathscr{K}$, and its elements are defined in the following way:

$$J(\mathscr{K}) = \{a \in \mathscr{K} \mid a \neq \emptyset, a = b \vee c \Rightarrow a = b \text{ or } a = c\}.$$

In this context, the lattice $\mathscr{K}$ can be constructed by unions ($\vee$) and intersections ($\wedge$) of the elements of the set of join-irreducible elements $J(\mathscr{K})$. Davey and Priestley (1990) characterize a join-irreducible element of a finite distributive lattice as an element which has exactly one lower cover, i.e. it covers exactly one other element. Note

[3] Reflexive, antisymmetric and transitive.

[4] Note that *envelope* and *minimal envelope* are concepts analogous to those of *cover* and *minimal cover* as defined in Hearne and Wagner (1973).



(a) (b) (c)

Fig. 2. From left to right, trees constructed from an empty moral TDAG, a complete moral TDAG and a moral TDAG where one vertex renders the two other non-adjacent vertices conditionally independent.

that the partial order $\leqslant$ of the finite poset $(J(\mathscr{K}), \leqslant)$ is inherited from $\mathscr{K}$: $a \leqslant b$ iff $a \wedge b = a$.

Analogous to the concept of an ancestral set in a DAG, one may define an *ancestral poset*. Let $(S, \leqslant)$ be a poset, a subset (which is again a poset) $A \subseteq S$ is ancestral in $(S, \leqslant)$ iff $\forall a \in A$ it follows that $b \in S$ and $b < a \Rightarrow b \in A$.

It is possible to establish a one-to-one correspondence between finite posets and TDAGs. Given the finite poset $(S, \leqslant)$ we can build a TDAG $G = (S, E^{<})$, where

$$E^{<} = \{(a, b) \in S \times S \mid a < b\}.$$

Given a TDAG $G = (V, E)$, for every pair of vertices $a, b \in V$, $a \in \mathrm{an}(b) \Leftrightarrow a < b$. Note that all ancestral subsets of a poset $(S, \leqslant)$ form a ring $\mathscr{A}((S, \leqslant))$ which is identical to the ancestral ring $\mathscr{A}((S, E^{<}))$ of the TDAG $G = (S, E^{<})$ defined before.

This correspondence between TDAGs and finite distributive lattices is used by Andersson et al. (1997b) to prove that TDAG models and LCI models coincide.

3. Moral TDAGs as labeled trees

In this section we build an isomorphism between moral TDAGs and labeled trees, which will allow us to represent any given moral TDAG with a unique corresponding labeled tree. In order to get a first intuition of such mapping, we may see in Fig. 2 an example of three simple moral TDAGs with their corresponding labeled tree representation.

In this section we do not yet discuss the Markov models, only purely graph-theoretic issues. The results in this section will be used later to introduce the new class of graphical Markov models based on labeled trees.

Lemma 3.1. *Let $G = (V, E)$ be a moral TDAG corresponding to the finite distributive lattice $\mathscr{K}$, which has set of join-irreducible elements $J(\mathscr{K})$. Let $\mathscr{A}(G)$ be the ring of ancestral subsets of V in G, which is identical to the ring of ancestral posets of $J(\mathscr{K})$, $\mathscr{A}((J(\mathscr{K}), \leqslant))$. The set of join-irreducible elements $J(\mathscr{K})$ is the collection of maximal chains*:

$$J(\mathscr{K}) = (H_i \mid H_i \subseteq \mathscr{A}(G) \wedge H_i \text{ maximal chain}). \tag{1}$$

Proof. A poset $(S, \leqslant)$ is a *tree poset* iff for $x, y, z \in (S, \leqslant)$, $x, y < z \Rightarrow x < y$ or $y < x$.

Recall from Section 2 that the relation between $G = (V, E)$ and its corresponding poset $(J(\mathscr{K}), \leqslant)$ is such that for every $a, b \in V$, $a \in \mathrm{an}(b) \Leftrightarrow a < b$ in $(J(\mathscr{K}), \leqslant)$. The fact that G is a moral TDAG implies that for $a, b, c \in V$ such that $a, b \in \mathrm{an}(c)$, either $a \in \mathrm{an}(b)$ or $b \in \mathrm{an}(a)$. Therefore, the poset $(J(\mathscr{K}), \leqslant)$ is a tree poset.

Consider a decomposition of a poset $(S, \leqslant)$ as a collection of smaller posets $H_1, H_2, \ldots, H_k$, $H_i \subseteq (S, \leqslant)$, $1 \leqslant i \leqslant k$, as follows. For every element x_i of the minimal envelope of $(S, \leqslant)$, create $H_i = \{y \in S\colon\ y < x_i \text{ in } (S, \leqslant)\} \cup \{x_i\}$.

Applying the previous decomposition to the poset $(J(\mathscr{K}), \leqslant)$ we will obtain a collection of k posets $H_i = \{y_1, \ldots, y_q, x_i\}$, $1 \leqslant i \leqslant k$. The tree poset condition of $(J(\mathscr{K}), \leqslant)$ and the fact that $y_j < x_i$, $1 \leqslant j \leqslant q$, implies that each H_i is a maximal chain. $\square$

Let $\mathscr{K}$ be a finite distributive lattice isomorphic to some moral TDAG G. Then the set of join-irreducible elements $J(\mathscr{K})$ is of form (1) and forms a poset $(J(\mathscr{K}), \leqslant)$. Let $\mathscr{L}$ denote the class of such finite distributive lattices. Consider a correspondence μ between the set of such posets $P(\mathscr{L}) = \{(J(\mathscr{K}), \leqslant) | \mathscr{K} \in \mathscr{L}\}$ and the set of labeled trees $T(\mathscr{L}) = \{(J(\mathscr{K}) \cup \{\emptyset\}, E^{\prec}) | \mathscr{K} \in \mathscr{L}\}$, defined as follows:

$$\begin{aligned} \mu\colon \qquad P(\mathscr{L}) &\longleftrightarrow T(\mathscr{L}), \\ (J(\mathscr{K}), \leqslant) &\longleftrightarrow (J(\mathscr{K}) \cup \{\emptyset\}, E^{\prec}), \end{aligned} \tag{2}$$

where for every labeled tree $t(\mathscr{K}) \equiv (J(\mathscr{K}) \cup \{\emptyset\}, E^{\prec}) \in T(\mathscr{L})$ the vertex set is formed by the elements in $J(\mathscr{K})$ plus an extra vertex labeled $\emptyset$, which acts as the root. Note that there is a one-to-one correspondence between $J(\mathscr{K})$ and the set of vertices from the equivalent moral TDAG. The set of edges of the labeled tree $t(\mathscr{K})$ is defined as follows:

$$E^{\prec} = \{(a, b) \in J(\mathscr{K}) \times J(\mathscr{K}) | a \prec b\} \cup \{(\emptyset, a) \in \{\emptyset\} \times J(\mathscr{K}) | \nexists b \in J(\mathscr{K}) b \prec a\},$$

where $\prec$ is the covering relation on the poset of join-irreducible elements $J(\mathscr{K})$. From the next three propositions it will follow that the correspondence μ is a bijection between moral TDAGs and labeled trees.

Proposition 3.1. *Let $\mathscr{K}$ be a finite distributive lattice that coincides with some moral TDAG. The graph $\mu(\mathscr{K})$ is a labeled tree.*

Proof. From Lemma 3.1 we can decompose $J(\mathscr{K})$ into its maximal chains H_{C}. Every $\mu(H_{\mathrm{C}})$ is a path in $\mu(\mathscr{K})$ from the root to a leaf and vice versa. For any two such chains H_{C_1} and H_{C_2}, $\mu(H_{\mathrm{C}_1}) \cap \mu(H_{\mathrm{C}_2})$ is a unique path from the root to a vertex. It follows directly that $\mu(\mathscr{K})$ has no cycles and therefore is a labeled tree. $\square$

Proposition 3.2. *The correspondence* (2) *is injective.*

Proof. Let $\mathscr{K}_1, \mathscr{K}_2$ be two finite distributive lattices that coincide with two moral TDAGs. If $\mu(\mathscr{K}_1) = \mu(\mathscr{K}_2)$, then for every path $h \in \mu(\mathscr{K}_1)$, for which we can create a maximal chain H, there exists a path $d \in \mu(\mathscr{K}_2)$, for which we can create a maximal chain D, such that $h = d$ and $H = D$. Since then the collection of all H will be the same as the collection of all D, it follows that $J(\mathscr{K}_1) = J(\mathscr{K}_2)$, and the correspondence (2) is injective. □

Proposition 3.3. *Correspondence* (2) *is surjective.*

Proof. For convenience, consider labeled trees T where the root is labeled as $\emptyset$ and the rest of the vertices using natural numbers $\{1,\dots,n\}$. From the fact that T is a labeled tree, there is always a unique path from the root to each of its leaves. For every path p of the tree T, such that $p = \{\emptyset, x_1, \dots, x_n\}$, take out the root $\emptyset$, and from the rest of the path $\{x_1,\dots,x_n\}$ construct a chain H_C such that $H_C = \{x_1, \{x_1, x_2\}, \dots, \{x_1,\dots,x_n\}\}$. From Lemma 3.1 we know that the collection of these chains H_C produces a set of join-irreducible elements $J(\mathscr{K})$ corresponding to a lattice $\mathscr{K}$ that coincides with a moral TDAG. □

Finally, we can establish the following result.

Theorem 3.1. *The class of moral TDAGs is isomorphic to the class of labeled trees.*

Proof. This follows directly from the fact that mapping (2) is a bijection between moral TDAGs and labeled trees. □

4. Moral TDAG models as tree conditional independence ≡ TCI models

This section introduces a new class of graphical Markov models called TCI models, based on labeled trees. Moreover, it is shown that TCI coincides with the class of DEC ∩ LCI graphical Markov models.

A graphical Markov model member of the class DEC∩LCI is determined by a TDAG with no immoralities (Andersson et al., 1995). By Theorem 3.1 we can represent a moral TDAG using a labeled tree.

One of the features that distinguishes tree structures from other types of graph, used in the context of graphical Markov models, is that they are connected. In this sense, they are quite similar to the Hasse diagrams used to represent lattices in LCI models. Thus, we may observe in Fig. 2a how a complete disconnected graph turns into a connected structure, a labeled tree, by using this new, artificially introduced, vertex labeled $\emptyset$.

The intuition behind the root node $\emptyset$ will become clear from the Markov property for labeled trees. To define this formally, we need two new concepts regarding labeled trees and the following proposition.

Proposition 4.1. *Let $T = (V \cup \{\emptyset\}, E)$ be a tree rooted at $\emptyset$. Given any two vertices $u, v \in V$ there is always at least one common vertex* (*the root $\emptyset$*) *in the two unique paths that lead from u and v to the root $\emptyset$.*

Proof. It follows directly from the fact that every vertex in a tree is reachable from the root by a unique path. □

Given the existence of at least one common element for every two paths from two given vertices to the root, consider the next two definitions.

Definition 4.1 (Meet). Let $T = (V \cup \{\emptyset\}, E)$ be a tree rooted at $\emptyset$. Let T_u, T_w be two subtrees of T, rooted at vertices u and w, respectively, such that neither is a subtree of the other. The *meet* is the first common vertex in the two unique paths from u, w to the root $\emptyset$. It will be denoted as $\varphi_{u,w}$.

Definition 4.2 (Meet path). Let $T = (V \cup \{\emptyset\}, E)$ be a tree rooted at $\emptyset$. Let $u, w \in V$ be two vertices inducing subtrees T_u, T_w, such that neither is a subtree of the other. Let $\varphi_{u,w}$ be their meet. The *meet path* is the set of vertices that forms the common path from the meet to the root, and denoted as $\mathrm{mp}(\varphi_{u,w}) = \{\varphi_{u,w}, \ldots, \emptyset\}$.

As we can see, *meet* and *meet path* are intuitive concepts that follow naturally from the definition of a tree. It is straightforward to identify the *meet* in a tree for two given vertices, even if this tree is large. Finally, the new Markov property can be introduced.

Definition 4.3 (Tree conditional independence Markov property (TCIMP)). Let $T = (V \cup \{\emptyset\}, E)$ be a tree rooted at $\emptyset$. A probability distribution P on $\mathscr{X}$ is said to satisfy the *tree conditional independence Markov property* (TCIMP) if, for every pair of vertices $u, w \in V$ inducing two subtrees $T_u = (U, E_U)$ and $T_w = (W, E_W)$ with meet $\varphi_{u,w}$ and meet path $\mathrm{mp}(\varphi_{u,w})$, P satisfies

$$U \perp\!\!\!\perp W \mid \mathrm{mp}(\varphi_{u,w}).$$

This Markov property leads to the following new type of graphical Markov model.

Definition 4.4 (TCI model). Let G be a labeled tree rooted at $\emptyset$. The set $\mathbf{T}_{\mathscr{X}}(G)$ of all probability distributions on $\mathscr{X}$ that satisfy the TCIMP relative to G is called the *TCI model* determined by G.

To illustrate, consider the three TCI models determined by the trees in Fig. 2. By the TCIMP, the tree in (a) renders the three vertices marginally independent: $1 \perp\!\!\!\perp 2 \perp\!\!\!\perp 3$. From the tree in (b) it is not possible to read off any conditional independencies, thus the set of restrictions of the model is empty. In (c) we may see that the vertex 2 is the meet of vertices 1 and 3, thus $1 \perp\!\!\!\perp 3 \mid 2$. These restrictions may be read from their corresponding moral TDAGs using the directed global Markov property (DGMP). Further examples are provided by Figs. 3 and 4.

In Fig. 3 we see the different graphical representations for three simple models of conditional independence, with the independencies as specified. In Fig. 4 we find a larger model which may help to understand the TCIMP. For instance, if we pick

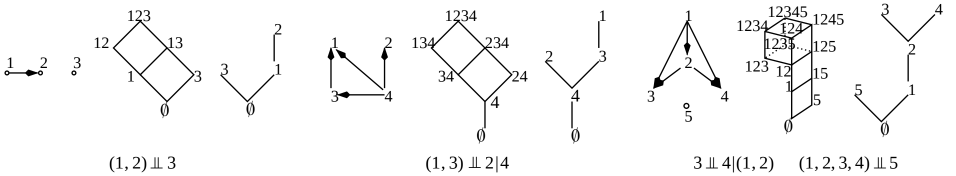


Fig. 3. Three different Markov models represented by a moral TDAG model, an LCI model and a TCI model.

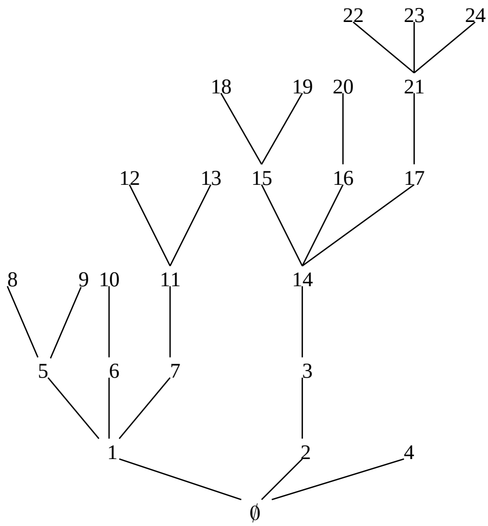


Fig. 4. Example of a TCI model for 24 variables.

the vertices 15 and 21, and apply the TCIMP, we see that the set $\{15, 18, 19\}$ is conditionally independent of $\{21, 22, 23, 24\}$ given $\{2, 3, 14\}$. While if we pick the vertices 12 and 13, the TCIMP renders the singletons $\{12\}$ and $\{13\}$ conditionally independent given $\{1, 7, 11\}$.

To show that DEC $\cap$ LCI coincides with TCI, we first have to investigate the relationship between TCIMP and the well-known Markov properties. To do this, we need some definitions.

Definition 4.5 (Moral ancestral set). Let $G = (V, E)$ be a DAG. Given a subset $A \subseteq V$, A is said to be *moral ancestral* iff for every vertex $v \in A$, $\mathrm{an}(v) \subseteq A$ and $\mathrm{an}(v) \cup \{v\}$ is complete in G.

Proposition 4.2. *Let* $G = (V, E)$ *be a* DAG. *Given two moral ancestral subsets* $A, B \subseteq V$, *the* union $A \cup B$ *is again moral ancestral in* G.

Proof. It is already known that the union of ancestral sets is ancestral. Thus, it is only necessary to determine whether the union of moral ancestral sets is moral.

Let $a, b, c \in A \cup B$ be such that $a \rightarrow c \leftarrow b$ where a and b are non-adjacent. Since $a, b \in \mathrm{an}(c)$, either $a, b, c \in A$ or $a, b, c \in B$, which would contradict the initial assumption that A and B are moral ancestral. □

Proposition 4.3. *Let* $G=(V,E)$ *be a* DAG. *Given two moral ancestral subsets* $A, B \subseteq V$, *the* intersection $A \cap B$ *is again moral ancestral in* G.

Proof. This follows firstly from the fact that the intersection of ancestral sets is again ancestral. And secondly, since $A \cap B \subseteq A$ and $A \cap B \subseteq B$, $A \cap B$ should be moral. Otherwise it would contradict the assumption of A, B being moral ancestral. □

From the previous propositions it follows that the moral ancestral sets contained in a DAG $G=(V,E)$ form a ring of subsets of V, which will be called the *moral ancestral ring* of G, and denoted as $\mathscr{A}^{\mathrm{m}}(G)$. The moral ancestral ring allow us to define the TCIMP in terms of DAGs.

Definition 4.6 (Directed tree conditional independence Markov property (DTCIMP)). Let $G=(V,E)$ be a DAG. A probability distribution P on $\mathscr{X}$ is said to satisfy the *directed tree conditional independence Markov property* (DTCIMP) if, for every pair of moral ancestral subsets $A, B \in \mathscr{A}^{\mathrm{m}}(G)$, P satisfies

$$A \perp\!\!\!\perp B | A \cap B.$$

Theorem 4.1. *Let* $\mathbf{D}_{\mathscr{X}}(G)$ *be a* DAG *model. For any probability distribution* P *on* $\mathscr{X}$,

$$\text{DGMP} \Rightarrow \text{LCIMP} \Rightarrow \text{DTCIMP} \Rightarrow \text{TCIMP}.$$

Proof. The first implication follows from Theorem A.1. For the second, let $A, B \in \mathscr{A}^{\mathrm{m}}(G)$. The LCIMP implies the DTCIMP if $A, B \in \mathscr{A}(G)$, and this follows because $\mathscr{A}^{\mathrm{m}}(G) \subseteq \mathscr{A}(G)$.

The third implication is proved as follows. For any pair $A, B \in \mathscr{A}^{\mathrm{m}}(G)$, the set $A \cup B$ induces a moral TDAG $G_{A \cup B}$ from G, such that it coincides with a tree $T_{A \cup B}$ by Theorem 3.1. The DTCIMP will imply the TCIMP if for each pair of vertices $a \in A \setminus B$ and $b \in B \setminus A$, $\mathrm{mp}(\varphi_{a,b}) = A \cap B$ in $T_{A \cup B}$. This equality follows from the fact that the meet path in $T_{A \cup B}$, for any pair of vertices $a \in A \setminus B$ and $b \in B \setminus A$, is formed by those vertices that are common to A and B, therefore $A \cap B$. □

Theorem 4.2. *Let* G *be a moral* TDAG. *For any probability distribution* P *on* $\mathscr{X}$, TCIMP $\Rightarrow$ DGMP. *Thus, for a moral* TDAG,

$$\text{TCIMP} \Leftrightarrow \text{DGMP} \Leftrightarrow \text{DLMP} \Leftrightarrow \text{LCIMP} \Leftrightarrow \text{DTCIMP}$$

and $\mathbf{D}_{\mathscr{X}}(G) = \mathbf{T}_{\mathscr{X}}(T)$, *for some tree* T *that coincides with the moral* TDAG G.

Proof. By Theorem 3.1, there is a unique labeled tree $T=(V,E)$ that coincides with the moral TDAG G. For any two vertices $u, w \in V$, that induce subtrees $T_u = (U, E_U)$,

$T_w = (W, E_W)$, the TCIMP in T implies the DGMP in G if U and W are separated by $\mathrm{mp}(\varphi_{u,w})$ in

$$(G_{\mathrm{An}(U \cup W \cup \mathrm{mp}(\varphi_{u,w}))})^{\mathrm{m}} = G_{\mathrm{An}(U \cup W \cup \mathrm{mp}(\varphi_{u,w}))}.$$

This equality follows since G is assumed to be moral. Now, we should find out which set separates U and W in the graph specified on the right hand of this equality. Consider a path between any two vertices $a \in U$ and $b \in W$. Since U, W were induced by vertices u, w, this path will intersect the sets $\mathrm{pa}(u)$ and $\mathrm{pa}(w)$, because of transitivity. More concretely, this path will always intersect those vertices $x \in \mathrm{pa}(u) \cap \mathrm{pa}(w)$. The set $\mathrm{pa}(u) \cap \mathrm{pa}(w)$ in a moral TDAG is equivalent to the definition of meet path, hence $\mathrm{mp}(\varphi_{u,w})$ separates U, W in $G_{\mathrm{An}(U \cup W \cup \mathrm{mp}(\varphi_{u,w}))}$. The second part of the theorem follows from Theorems A.2 and 4.1. □

Finally, we can establish the following theorem that determines the location of TCI models, within the family of graphical Markov models.

Theorem 4.3. *The class of* TCI *models coincides with the class of* DEC $\cap$ LCI *models.*

Proof. It follows from the fact that $\mathbf{D}_{\mathscr{X}}(G) = \mathbf{T}_{\mathscr{X}}(T)$, for some moral TDAG G and some labeled tree T, which is proved in Theorem 4.2. □

5. Markov equivalence among TCI models

DAG models are organized in classes of equivalence, such that two DAG models $\mathbf{D}_{\mathscr{X}}(G_1)$ and $\mathbf{D}_{\mathscr{X}}(G_2)$ determined by two different DAGs G_1 and G_2 may actually determine the same model of conditional independence, hence $\mathbf{D}_{\mathscr{X}}(G_1) = \mathbf{D}_{\mathscr{X}}(G_2)$. This situation can also occur in the case of TCI models: two different trees T_1, T_2 may determine the same TCI model $\mathbf{T}_{\mathscr{X}}(T_1) = \mathbf{T}_{\mathscr{X}}(T_2)$. We will investigate now the notion of Markov equivalence among TCI models. First, we review the notion of Markov equivalence for DAG models, which was given independently by Frydenberg (1990), and Verma and Pearl (1991).

Theorem 5.1. *Two* DAG *models are Markov equivalent if and only if they have the same skeleton and the same immoralities.*

It is possible to decide Markov equivalence for TCI models by simply creating the corresponding moral TDAG using Theorem 3.1 and applying the previous theorem. The notion specifically for TCI models is as follows.

Definition 5.1. Let $T = (V \cup \{\emptyset\}, E)$ be a labeled tree rooted at vertex $\emptyset$. Let $l \in V$ be a leaf in T. We define a *branch ending at* l as the set of vertices $\pi(l) = \{x_1, \ldots, x_n\}$, where $x_1 = \emptyset$ and $x_n = l$, present in the unique path between the root $\emptyset$ and the leaf l. The set of all branches of T will be denoted as $\lambda(T)$.

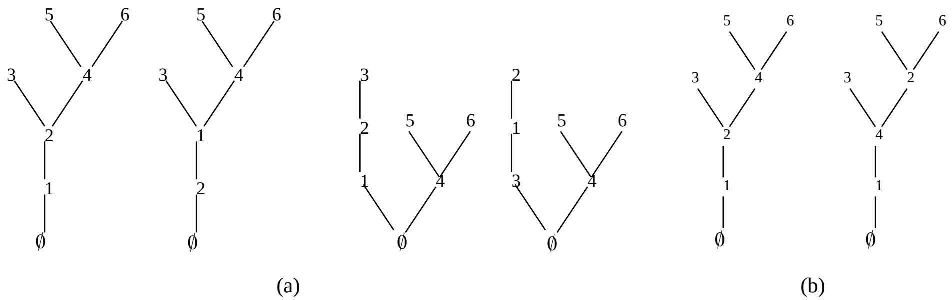


Fig. 5. Markov equivalence between TCI models. The two pairs on (a) are Markov equivalent, while the pair on (b) is not.

Theorem 5.2. *Two* TCI *models* $\mathbf{T}_{\mathscr{X}}(T)$ *and* $\mathbf{T}_{\mathscr{X}}(T')$ *are Markov equivalent, i.e.* $\mathbf{T}_{\mathscr{X}}(T) = \mathbf{T}_{\mathscr{X}}(T')$, *if and only if they have the same sets of branches, i.e.* $\lambda(T) = \lambda(T')$.

In order to illustrate the notion of Markov equivalence among trees, look at Fig. 5. The pairs of trees on part (a) are Markov equivalent because although two vertices are swapped between the trees, the paths from the leaves to the root remain the same. The two trees on part (b) are not Markov equivalent because given the swap of vertices 2 and 4, although it does not change the paths from vertices 5 and 6 to the root $\emptyset$, it does it from vertex 3 to the root $\emptyset$.

Proof of Theorem 5.2 (*Necessity*).
Assume that $\mathbf{T}_{\mathscr{X}}(T) = \mathbf{T}_{\mathscr{X}}(T')$. Then, any TCIMP read from T, holds also in T'. A TCIMP involves the vertex sets of two subtrees and the vertex set of their meet path. If a TCIMP holds in T and in T', then the two vertex sets of the two subtrees and the vertex set of their meet path in T, should be derived also from T'.

Let u, w be two vertices of the tree $T = (V, E)$, with meet $\varphi_{u,w}$, inducing subtrees $T_u = (U \cup \{\emptyset\}, E_W)$, $T_w = (W \cup \{\emptyset\}, E_W)$ such that none of them is subtree of the other. In order to find the same meet path $\mathrm{mp}(\varphi_{u,w})$ for u, w in both trees T and T', the paths from u, w to the root must intersect in the same vertices in T and T'. Because this must happen for every pair of vertices in T and T', it follows that the only way that the intersections of all the paths are the same in T and T', is when $\lambda(T) = \lambda(T')$.

(*Sufficiency*). Assume that $\lambda(T) = \lambda(T')$. This implies that every possible meet path from any given two subtrees in T must exist in T'. Because, if there were a meet path that differs in at least one vertex in T and T', there would exist some $\pi(l) \in \lambda(T)$ and $\pi'(l) \in \lambda(T')$ such that $\pi(l) \neq \pi'(l)$. Therefore, if for every two given subtrees in T, their meet path is the same in T', it follows from the TCIMP that the collection of Markov properties of T hold also in T', and vice versa. □

5.1. How many different Markov models?

The notion of Markov equivalence reveals the fact that two graphical Markov models determined by two different graphs may represent the same model of conditional independence. Very often, model selection on graphical Markov models is carried out over the space of graphs that determine the type of graphical Markov model we are selecting. While the number of different graphs that determine a given class of Markov models provides an estimation of the hardness of selecting a good set of models, the expressive power of a given class of graphical Markov models may be quantified by the number of different models of conditional independence that we can represent using this class. When the equivalence class of a given type of graphical Markov models has a precise graphical definition, one may try to use standard tools of graph theory and graphical enumeration to count how many models of conditional independence may be represented.

The most straightforward case is that of UG models, which are represented by undirected graphs, since there is a one-to-one correspondence between undirected graphs and models of conditional independence. Two UG models $\mathbf{U}_{\mathscr{X}}(G_1), \mathbf{U}_{\mathscr{X}}(G_2)$ are Markov equivalent, $\mathbf{U}_{\mathscr{X}}(G_1) = \mathbf{U}_{\mathscr{X}}(G_2)$, iff $G_1 = G_2$. Thus, there is the same number of different models of conditional independence, as different undirected graphs, i.e. $2^{\binom{n}{2}}$ for labeled graphs with n vertices.

The case of DEC models is also straightforward, since chordal graphs also have a one-to-one correspondence with Markov equivalence classes. Connected and disconnected chordal graphs were counted by Wormald (1985). The sum of these two quantities provides the number of all chordal graphs, i.e. the number of all different DEC models. We will treat in this section the derivation of the expression that allows us to compute the number of all chordal graphs from the numbers of connected chordal graphs. As we shall see, this expression is related to the computation of Markov equivalence classes of TCI models.

The case of DAG models is a difficult one. Markov equivalence classes of DAG models are represented by essential graphs (Andersson et al., 1997a), which are acyclic partially directed graphs with additional characterizing properties. An efficient way of enumerating such graphs is not yet known. In Andersson et al. (1997a) they were counted up to 5 vertices. Recently, Gillispie and Perlman (2001) developed a computer program which has calculated the number of essential graphs up to 10 vertices, by enumerating DAGs and taking into account the equivalence class to which each DAG belongs. We have taken the liberty to extrapolate these numbers up to 12 vertices, such that we can compare cardinalities of different Markov equivalence classes in Table 2.

The enumeration of Markov equivalence classes of TCI models provides insight into their nature such that, afterwards, it is easy to devise a canonical representation for a given equivalence class of TCI models (see Section 5.2).

The basic mathematical tool used in enumeration of graphs is that of *generating functions*. A generating function is a power series. The coefficients of the polynomial that forms these power series store the counts of the object we intend to enumerate. The exponents of this polynomial describe some structural feature associated to its attached coefficient, as for instance, the number of vertices of a graph. In the case of labeled enumeration, one uses an *exponential generating function* of the form $g(x) = \sum_{n=1} a_n(x^n/n!)$. For full insight into this subject the reader may consult the book of Harary and Palmer (1973).

Let $g(x)$ be the generating function for connected labeled chordal graphs. Then a_n corresponds to the number of such graphs with n vertices. Consider now another exponential generating function to count not only *connected* labeled chordal graphs, but *all* of them:

$$G(x) = \sum_{n=0} A_n \frac{x^n}{n!}. \tag{3}$$

In this generating function, the coefficient A_n is the number of *all* chordal graphs with n vertices, which corresponds to the number of Markov equivalence classes of DEC models. These two exponential generating functions are related through the following theorem.

Theorem 5.3 (Harary and Palmer, 1973, p. 8). *The exponential generating functions $G(x)$ and $g(x)$ for labeled graphs and labeled connected graphs satisfy the following relation*:

$$1 + G(x) = \mathrm{e}^{g(x)}.$$

Here the constant 1 refers to the null graph, i.e. the graph with no vertices. In the way we have expressed the generating function $G(x)$, the constant 1 is included in $G(x)$ since n starts on 0 vertices. Therefore we may discard the constant 1 in the previous expression. As we shall see now, by differentiating the previous equation and equating coefficients, it is possible to find a recurrence for both the number of all labeled chordal graphs A_n and labeled connected chordal graphs a_n. First, $g(x)$ is isolated, by taking logarithms on both sides, and afterwards we can differentiate the equation, which leads to the following relation:

$$\frac{\sum_{n=0}^{\infty} n(A_n/n!)x^{n-1}}{\sum_{n=0}^{\infty} (A_n/n!)x^n} = \sum_{n=1}^{\infty} n \frac{a_n}{n!} x^{n-1}.$$

Now multiply both sides by the polynomial at the bottom-left of the equation, to obtain

$$\sum_{n=0}^{\infty} n \frac{A_n}{n!} x^{n-1} = \sum_{n=0}^{\infty} \left(\sum_{k=0}^{\infty} (k+1) \frac{a_{k+1}}{(k+1)!} \frac{A_{n-k}}{(n-k)!} \right) x^n.$$

In order to equate coefficients, the exponents of both polynomials should match. Therefore we are going to move the running indexes on the right-hand side of the equation. First, move the index k of the inner sum, and then move the index n. Further, the first term of the sum on the left-hand side may be discarded since it cancels for $n = 0$. Thus we obtain

$$\sum_{n=0}^{\infty} n \frac{A_n}{n!} x^{n-1} = \sum_{n=0}^{\infty} \left(\sum_{k=1}^{n+1} k \frac{a_k}{k!} \frac{A_{n+1-k}}{(n+1-k)!} \right) x^n,$$

$$\sum_{n=1}^{\infty} n \frac{A_n}{n!} x^{n-1} = \sum_{n=1}^{\infty} \left(\sum_{k=1}^{n} k \frac{a_k}{k!} \frac{A_{n-k}}{(n-k)!} \right) x^{n-1}.$$

Table 1
Number of Markov equivalence classes of DEC models

A_n	n
1	1
2	2
8	3
61	4
822	5
18,154	6
617,675	7
30,888,596	8
2,192,816,760	9
215,488,096,587	10
28,791,414,081,916	11
5,165,908,492,061,926	12

We can now equate coefficients, and for our purposes, we will isolate from the sum the term for $k = n$. In this term, we can substitute afterwards $A_0 = 1$, since the null graph is unique, to obtain

$$n\,\frac{A_n}{n!} = n\,\frac{a_n}{n!}\,\frac{A_0}{0!} + \sum_{k=1}^{n-1} k\,\frac{a_k}{k!}\,\frac{A_{n-k}}{(n-k)!}.$$

Finally, by multiplying the whole expression by $n!$ and dividing it by n, we obtain the recurrence for all chordal graphs for n vertices:

$$A_n = a_n + \frac{1}{n}\left(\sum_{k=1}^{n-1} k \binom{n}{k} a_k A_{n-k}\right). \tag{4}$$

Wormald (1985) provides the numbers a_n for labeled connected chordal graphs; thus by using these and formula (4), we obtain the numbers of all labeled chordal graphs; which equals the number of Markov equivalence classes of DEC models, given in Table 1.

Next we count the Markov equivalence classes of TCI models. Andersson et al. (1997b) characterized these models as those DEC models $\mathbf{U}_{\mathscr{X}}(G)$ determined by a chordal graph G such that G does not contain the following induced undirected subgraph:

$$\circ\!\!-\!\!-\!\!-\!\!\circ\!\!-\!\!-\!\!-\!\!\circ\!\!-\!\!-\!\!-\!\!\circ$$

which is a path on four vertices and denoted as P_4. The term P_4-free is used in this context to denote the absence of such induced subgraph. In the graph theory literature Wolk (1962) gave a first characterization of these graphs, although it was Golumbic (1978) who described them later in the terms of forbidden subgraphs as above, hence the name "P_4-free chordal". From this characterization, Castelo and Wormald (2003) derived the following proposition.

Table 2
Numbers of Markov equivalence classes of DAGs, UGs, DECs and TCIs with n vertices

DAG (essential)	UDG (undirected)	DEC (chordal)	TCI (P_4-free chordal)	n
1	1	1	1	1
2	2	2	2	2
11	8	8	8	3
185	64	61	49	4
8,782	1,024	822	402	5
1,067,825	32,768	18,154	4,144	6
312,510,571	2,097,152	617,675	51,515	7
21,213,3402,500	268,435,456	30,888,596	750,348	8
326,266,056,291,213	68,719,476,736	2,192,816,760	12,537,204	9
1,118,902,054,495,975,141	35,184,372,088,832	215,488,096,587	236,424,087	10
$\approx 8.53\text{e}+21$	36,028,797,018,963,968	28,791,414,081,916	4,967,735,896	11
$\approx 1.40\text{e}+26$	73,786,976,294,838,206,464	5,165,908,492,061,926	115,102,258,660	12

Proposition 5.1 (Castelo and Wormald, 2003). *Let $G=(V,E)$ be a connected undirected graph. Let D be the set of vertices of degree $|V|-1$. Then G is P_4-free and chordal if and only if it is complete or $G-D$ is a disconnected P_4-free chordal graph.*

This proposition implicitly suggests the following two lemmas, where the first corresponds to the *diagonal property* of Wolk (1962), and the second corresponds to the concept of *central point* in Wolk (1965, p. 18) and implies the existence of the set of vertices $D\neq\emptyset$ of degree $|V|-1$ in Proposition 5.1.

Lemma 5.1 (Wolk, 1962, Lemma 1). *Let G be a connected P_4-free chordal graph. Let G have more than one clique. For every clique C of G, the intersections of all other cliques with C are nested.*

Lemma 5.2 (Wolk, 1965, Lemma in p. 18). *Let G be a connected P_4-free chordal graph. Let G have more than one clique. The intersection of all the cliques of G is non-empty.*

These properties allowed the authors in Castelo and Wormald (2003) to enumerate P_4-free chordal graphs, which in our context, correspond to Markov equivalence classes of TCI models. In particular, Castelo and Wormald (2003) provide the following recurrence for connected P_4-free chordal graphs:

$$a_n = 1 + \sum_{k=1}^{n-2} \binom{n}{k} (A_{n-k} - a_{n-k}). \tag{5}$$

In this recurrence, the term A_{n-k} refers to the number of all P_4-free chordal graphs with $n-k$ vertices. As shown in Castelo and Wormald (2003), since the generating functions for P_4-free chordal graphs are the same as for chordal graphs, A_{n-k} can be computed using the recurrence in (4). We may see in Table 2 the numbers for all P_4-free chordal

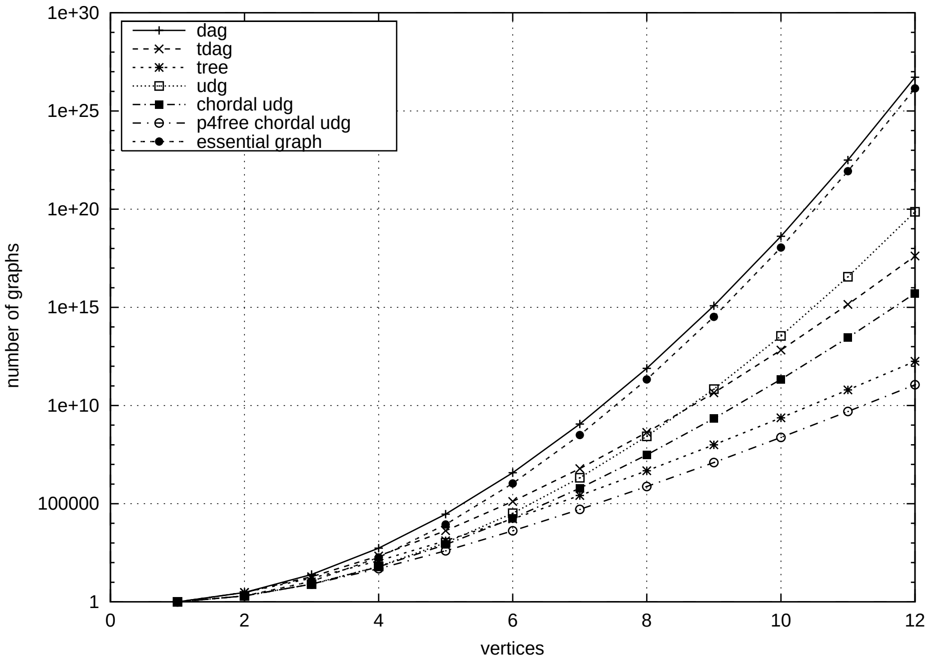


Fig. 6. Cardinalities of the types of graphs that determine different subclasses of Markov models.

graphs, which have a one-to-one correspondence with Markov equivalence classes of TCI models. They have been computed using the previously derived recurrences (4) and (5).

As noted above, TCI models may be defined as those graphical Markov models determined by P_4-free chordal graphs. From the characterization presented in the previous section of moral TDAGs as trees, it follows as well that there are $(n+1)^{n-1}$ different moral TDAGs on n vertices. As it has been already said, these quantities on different graphs may serve to quantify, roughly, difficulty of model selection and expressiveness of the graphical Markov model. Thus, it may be interesting to look at the plot of the cardinalities of the different graphs that determine in several forms different types of graphical Markov models, that we may find in Fig. 6.

5.2. A canonical representation of an equivalence class of TCI models

The P_4-free chordal graph characterization of TCI models suggests a canonical representation of Markov equivalence classes of TCI models. This representation will have again the form of a tree, but its nodes will contain, possibly, more than one single vertex (i.e. more than one single random variable). First we show how the cliques of a connected P_4-free chordal graph lead to a tree organization of their intersections. This allows a representation for the canonical element of an equivalence class of TCI

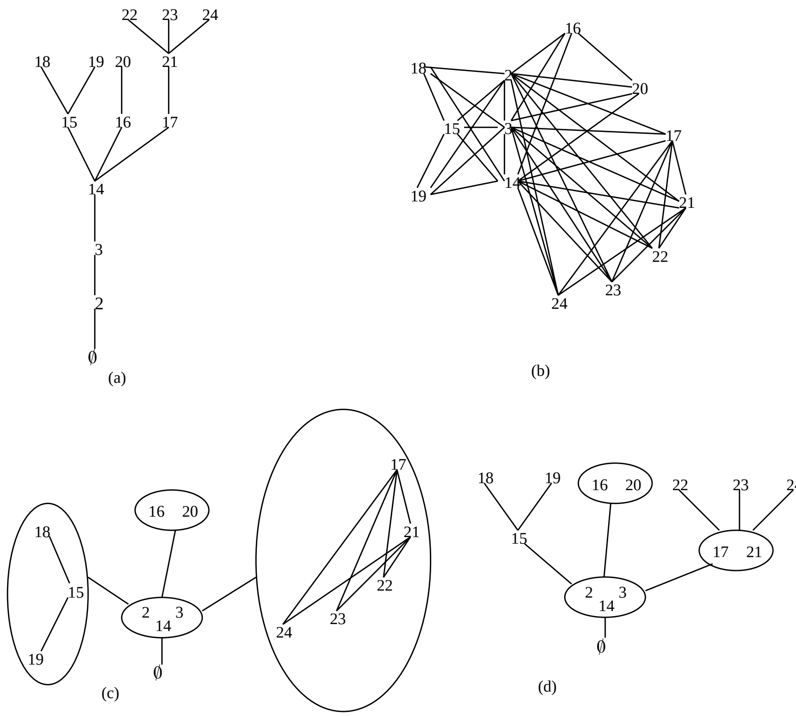


Fig. 7. (a) A TCI model. (b) Its corresponding P_4-free chordal graph. In (c) and (d) the two steps to obtain the canonical form of the equivalence class are shown.

models. Finally, it will be shown how to extract all the members of the equivalence class from this canonical representation.

By Lemma 5.2 a connected P_4-free chordal graph containing more than one clique has a non-empty subset of vertices D which correspond to the intersection of all cliques of the graph. By Proposition 5.1 the resulting graph, after the removal of D, will consist of $k > 1$ disconnected components that are again P_4-free chordal graphs. Now repeat the previous operation recursively until no disconnected component contains more than one clique. At each step of this operation, we will keep track of the different intersecting sets, and we will draw undirected edges from a given intersecting set, to those intersecting sets derived from the disconnected components that were created. It follows that such an undirected structure cannot have undirected cycles, thus has the form of a tree. In Fig. 7b we may see the P_4-free chordal graph corresponding to the TCI model of Fig. 7a, which corresponds to one of the subtrees of the TCI model of Fig. 4. In Fig. 7c we may see a first step of the procedure we just described, and in Fig. 7d we may see the second and last step, from which we have already obtained the canonical representation.

In graph-theoretic terms, the canonical representation of a TCI model, as for instance the one in Fig. 7d, corresponds to *homeomorphically irreducible trees* (Harary and

Palmer, 1973). Homeomorphically irreducible trees are those trees in which no vertex has degree of adjacency equal to two.

In order to find the members of the equivalence class, one only needs to perform all possible permutations on those nodes of the canonical element that contain more than one vertex. Then build a path for a given permutation, on which the vertices on the extremes of the path will connect to the adjacent nodes in the canonical element. For a given TCI model with more than one branch growing from the root $\emptyset$ one just applies this process to each of the branches separately, and plants the bottom roots (the first intersection set removed) on the root $\emptyset$. For a given canonical element with $s_1, \ldots, s_k$ nodes that contain more than one vertex, the number of trees on that equivalence class will amount to $|s_1|! \cdots |s_k|!$. The reason is obvious since it just corresponds to the number of possible permutations of those nodes that may be exchangeable on the tree.

6. Discussion

In this paper a new class of graphical Markov models, called TCI models, determined by labeled trees has been introduced. It is shown that the class of TCI models coincides with the intersection class of DEC$\cap$LCI. Moreover, a new Markov property, specific for trees, is introduced, and its relationship with the other Markov properties, is investigated.

We have also studied the notion of Markov equivalence among TCI models, which is based on a new concept also introduced in this paper: the concept of a *meet* of two subtrees, and their *meet path*. The one-to-one correspondence between Markov equivalence classes of TCI models and P_4-free chordal graphs allows the computation of the number of different Markov models contained in DEC $\cap$ LCI. In this way, we can compare the cardinalities of several interesting classes of graphical Markov models. Particular properties of P_4-free chordal graphs allow us to devise a canonical representation of an equivalence class of TCI models. This canonical representation also shows the correspondence between P_4-free chordal graphs and homeomorphically irreducible trees. Moreover, from Theorem 3.1 and from the number of labeled trees on n vertices, which is n^{n-2}, it follows that there are $(n+1)^{n-1}$ moral TDAGs on n vertices. In graph-theoretic terms one may say that there are $(n+1)^{n-1}$ transitive subtree acyclic digraphs. From the number of rooted labeled trees, which is n^{n-1}, it also follows directly that there are n^{n-1} connected moral TDAGs (or connected transitive subtree acyclic digraphs). These last few facts are interesting graph-theoretic results in their own right, and possibly have some consequence, in our context, from a model selection perspective.

TCI models expand the scope of possible formalisms for conditional independence within the superclass of graphical Markov models. Their expressiveness is smaller than any other subclass of graphical Markov models yet studied, but in turn, they provide much more clarity of representation. In contrast to most graphical Markov models, TCI models use a non-separation criterion to read off conditional independencies. This non-separation criterion, defined in terms of subtrees and meet path, provides a very easy identification of the conditioning sets.

Andersson et al. (1995, p. 38) claimed that since every conditional independence statement $A \perp\!\!\!\perp B|C$ is equivalent to a simple LCI model, any DAG model is the intersection of all LCI models that contain it. We can see further that every conditional independence statement $A \perp\!\!\!\perp B|C$ is equivalent to a simple TCI model, therefore any DAG model is the intersection of all TCI models that contain it. A remaining question is how TCI models can be combined graphically to determine the DAG structure of the intersection of TCI models.

Acknowledgements

Most of the work of this paper was carried out when the authors were working at the Research Center for Mathematics and Computer Science of Amsterdam (CWI). The first author also thanks the Fields Institute in Toronto for the support received during the three weeks in October 1999 in which he was visiting the institute, and during which he started the research that has led to the present paper.

We thank Nick Wormald and Ron Read for their valuable help in aspects related to graphical enumeration, Steven Gillispie and Tomáš Kočka for their information regarding essential graphs and Albrecht Schmidt for suggesting us the term *meet* for Definition 4.1.

Finally, we are grateful to the Advisory Editor, whose remarks and suggestions have substantially improved this paper.

Appendix. Graphical Markov models

This section gives an overview of graphical Markov models. We will survey very briefly UG, DAG, DEC, TDAG, and LCI models. The glue that binds the structural information of a graph, with the structural information of family of multivariate distributions P is its *Markov properties*. They make it possible to read conditional independencies from the graph. Moreover, there are relationships among the Markov properties that determine which ones are equivalent or which one is sharper than the other. For more insight into this discussion and the rest of the section, the reader may consult Lauritzen et al. (1990); Frydenberg (1990); Lauritzen (1996) and Andersson et al. (1997b).

Definition A.1 (Undirected global Markov property (UGMP)). Let $G=(V,E)$ be a UG, a probability distribution P on $\mathscr{X}$ is said to satisfy the *undirected global Markov property* (UGMP) if, for any triple of disjoint subsets of V such that S separates A from B in G, P satisfies

$$A \perp\!\!\!\perp B|S.$$

There are other two Markov properties [5] for undirected graphs, but the UGMP is the sharpest possible rule to read off conditional independencies from such graphs. The UGMP is used to define the UG Markov model in the following way.

Definition A.2 (UG Markov model). Let G be a UG. The set $\mathbf{U}_{\mathscr{X}}(G)$ of all probability distributions on $\mathscr{X}$ that satisfy the UGMP relative to G is called the UG *Markov model* determined by G.

Again, since all chordal graphs are undirected graphs, DEC models are defined in the same way as the set of all probability distributions that satisfy UGMP relative to G. We consider now DAG Markov models, which have a pairwise (DPMP), a local (DLMP) and a global (DGMP) Markov property. We provide here only the definition of the DGMP and refer the interested reader to the sources given at the beginning of this appendix.

Definition A.3 (Directed global Markov property (DGMP)). Let $G=(V,E)$ be a DAG, a probability distribution P on $\mathscr{X}$ is said to satisfy the *directed global Markov property* (DGMP) if, for any triple (A,B,S) of disjoint subsets of V such that S separates A from B in the moralized version of the smallest ancestral set of $A\cup B\cup S$, $(G_{\mathrm{An}(A\cup B\cup S)})^{\mathrm{m}}$, P satisfies

$$A \perp\!\!\!\perp B|S.$$

Lauritzen et al. (1990) show that the latter Markov property, the DGMP, is equivalent to the d-separation criterion from Pearl and Verma (1987). As in the undirected case, let us introduce now the formal definition of DAG Markov model.

Definition A.4 (DAG Markov model). Let G be a DAG. The set $\mathbf{D}_{\mathscr{X}}(G)$ of all probability distributions on $\mathscr{X}$ that satisfy the DGMP relative to G is called the DAG *Markov model* determined by G.

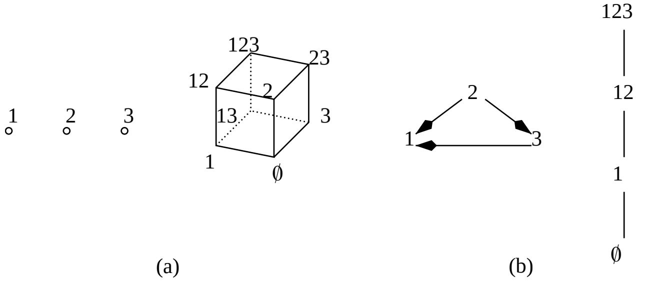
We discuss now briefly the relation between TDAG and LCI Markov models (cf. Andersson et al., 1997b).

Definition A.5 (Lattice conditional independence Markov property (LCIMP)). Let $G=(V,E)$ be a TDAG, a probability distribution P on $\mathscr{X}$ is said to satisfy the *lattice conditional independence Markov property* (LCIMP) if, for every pair of ancestral subsets $A,B\in\mathscr{A}(G)$, P satisfies

$$A \perp\!\!\!\perp B|A\cap B.$$

Andersson et al. (1997b) defined the LCIMP for the ancestral sets of a DAG. In this more general case, they prove:

[5] Pairwise ($\equiv$ UPMP) and local ($\equiv$ ULMP).



(a) (b)

Fig. 8. Comparison between DAG models and LCI models.

Theorem A.1 (Andersson et al., 1997b, Theorem 2.2, p. 32). *Let G be a* DAG. *For any probability distribution P on $\mathscr{X}$*, DGMP $\Rightarrow$ LCIMP.

The previous theorem is sharpened for TDAG Markov models as follows.

Theorem A.2 (Andersson et al., 1997b, Theorem 3.1, p. 33). *Let G be a* TDAG. *For any probability distribution P on $\mathscr{X}$*,

DGMP $\Leftrightarrow$ DLMP $\Leftrightarrow$ LCIMP.

The definition of an LCI model is the following.

Definition A.6 (LCI Markov model). Let G be a TDAG. The set $\mathbf{L}_{\mathscr{X}}(G)$ of all probability distributions on $\mathscr{X}$ that satisfy the LCIMP relative to G is called the LCI *Markov model* determined by G.

The fact that there is a one-to-one correspondence between TDAGs and finite distributive lattices, and the latter are isomorphic to a ring of sets, leads to an alternative reformulation of LCI Markov model in terms of posets. Consider a ring $\mathscr{K}$ of subsets of V, such that for every pair of subsets $L, M \in \mathscr{K}$, a probability distribution P satisfies

$$L \perp\!\!\!\perp M | L \cap M,$$

as in the LCIMP. The subsets L, M refer to subsets of random variables $\mathbf{X}_L, \mathbf{X}_M \subseteq \mathbf{X}_V$ that take values from a larger product space $\mathscr{X} = \times(\mathscr{X}_i | i \in V)$ and $L, M \subseteq V$. Over this product space, a family of probability distributions P underlies the LCIMP we rewrote before, and gives rise to an LCI model $\mathbf{L}_{\mathscr{X}}(\mathscr{K})$ that, as the notation suggests, is determined by a ring $\mathscr{K}$. For more details about LCI models determined by rings of subsets, the reader may consult Andersson and Perlman (1993) and Andersson et al. (1997b).

In Fig. 8a we may see an empty DAG, which represents the fully restricted DAG model, on the left, and its representation by a Hasse diagram on its right as the fully restricted LCI model. In Fig. 8b we may see a complete DAG, which represents the unrestricted or saturated DAG model, and its representation by a Hasse diagram on its

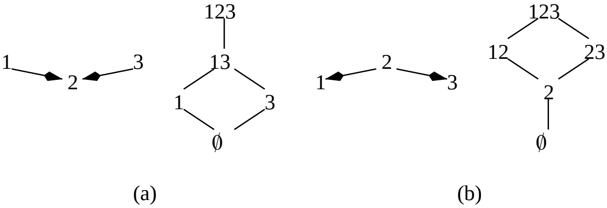


Fig. 9. On the left-hand side of (a) and (b), two DAGs representing $1 \perp\!\!\!\perp 3|\emptyset$ and $1 \perp\!\!\!\perp 3|2$, respectively, and on the right-hand side of (a) and (b) their corresponding Hasse diagrams.

right as the unrestricted LCI model. Let us note that for the LCI model in Fig. 8a, $J(\mathscr{K}_{\mathrm{a}}) = \{1, 2, 3\}$ and for the LCI model on Fig. 8b, $J(\mathscr{K}_{\mathrm{b}}) = \{1, 12, 123\}$.

While for the graphical Markov model in Fig. 8a the restrictions are characterized by all three vertices being marginally independent: $1 \perp\!\!\!\perp 2 \perp\!\!\!\perp 3$, the set of restrictions of the model in Fig. 8b is empty. In order to read conditional independencies from the Hasse diagram, we have to take into account that any two elements from this diagram are conditionally independent given their intersection (LCIMP). For instance, two trivial cases are those from Fig. 8.

In the next figure, we may see two more sophisticated models. The one in Fig. 9a corresponds to the immorality that induces the two non-adjacent vertices marginally independent, and the one in Fig. 9b makes the two non-adjacent vertices conditionally independent given the middle one. On the LCI model of Fig. 9a, $J(\mathscr{K}_{\mathrm{a}}) = \{1, 3, 123\}$ and on LCI model of Fig. 9b, $J(\mathscr{K}_{\mathrm{b}}) = \{2, 12, 23\}$ (recall that an element belongs to $J(\mathscr{K})$ iff it covers only one other element).

References

Andersson, S., Perlman, M., 1993. Lattice models for conditional independence in a multivariate normal distribution. Ann. Statist. 21, 1318–1358.

Andersson, S., Madigan, D., Perlman, M., Triggs, C., 1995. On the relation between conditional independence models determined by finite distributive lattices and by directed acyclic graphs. J. Statist. Plann. Inference 48, 25–46.

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

Andersson, S., Madigan, D., Perlman, M., Triggs, C., 1997b. A graphical characterization of lattice conditional independence models. Ann. Math. Artificial Intelligence 21, 27–50.

Castelo, R., Wormald, N., 2003. Enumeration of P_4-free chordal graphs. Technical Report UU-CS-2001-12, Institute for Computing and Information Sciences, University of Utrecht, The Netherlands, J. Graphs Combin., in press.

Cox, D., Wermuth, N., 1996. Multivariate Dependencies—Models, Analysis and Interpretation. Chapman & Hall, London.

Davey, B., Priestley, H., 1990. Introduction to Lattices and Order. Cambridge University Press, Cambridge.

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

Gillispie, S., Perlman, M., 2001. Enumerating Markov equivalence classes of acyclic digraph models. In: Breese, J., Kohler, D. (Eds.), Uncertainty in Artificial Intelligence, Morgan Kaufmann, Los Atlos, CA, pp. 171–177.

Golumbic, M., 1978. Trivially perfect graphs. Discrete Math. 24, 105–107.

Grätzer, G., 1978. General Lattice Theory. Birkhäuser, Basel.
Harary, F., Palmer, E., 1973. Graphical Enumeration. Academic Press, New York.
Harary, F., Kabell, J., McMorris, F., 1992. Subtree acyclic digraphs. Ars Combin. 34, 93–95.
Hearne, T., Wagner, C., 1973. Minimal covers of finite sets. Discrete Math. 5, 247–251.
Kiiveri, H., Speed, T., Carlin, J., 1984. Recursive causal models. J. Austral. Math. Soc. Ser. A 36, 30–52.
Lauritzen, S., 1996. Graphical Models. Oxford University Press, Oxford.
Lauritzen, S., Dawid, A., Larsen, B., Leimer, H., 1990. Independence properties of directed Markov fields. Networks 20, 491–505.
Pearl, J., 1988. Probabilistic Reasoning in Intelligent Systems. Morgan Kaufmann, Los Altos, CA.
Pearl, J., Verma, T., 1987. The logic of representing dependencies by directed graphs. In: Proceedings of the Conference of the American Association of Artificial Intelligence.
Verma, T., Pearl, J., 1991. Equivalence and synthesis of causal models. In: Bonissone, P., Henrion, M., Kanal, L., Lemmer, J. (Eds.), Uncertainty in Artificial Intelligence, Vol. 6. Elsevier, Amsterdam, pp. 255–268.
Wermuth, N., 1980. Linear recursive equations, covariance selection, and path analysis. J. Amer. Statist. Assoc. 75, 963–972.
Whittaker, J., 1990. Graphical Models in Applied Multivariate Statistics. Wiley, New York.
Wolk, E., 1962. The comparability graph of a tree. Proc. Amer. Math. Soc. 13, 789–795.
Wolk, E., 1965. A note on “the comparability graph of a tree”. Proc. Amer. Math. Soc. 16, 17–20.
Wormald, N., 1985. Counting labeled chordal graphs. Graphs Combin. 1, 193–200.