A characterization of moral transitive directed acyclic graph Markov models as trees and its properties

Robert Castelo Arno Siebes
roberto@cs.uu.nl siebes@cs.uu.nl
Department of Computer Science
University of Utrecht
P.O. Box 80089, 3508TB Utrecht
The Netherlands

AMS 1991 Subject Classification: 62H99
Keywords: Graphical Markov model; Conditional independence; Multivariate distribution; Undirected graph model; Directed acyclic graph model; Transitive directed acyclic graph model; Decomposable model; Lattice conditional independence model; Tree conditional independence model; Finite distributive lattice; Poset; Labeled tree.

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 directed acyclic graph models (or decomposable –DEC models), and transitive directed acyclic graph –TDAG models (or lattice conditional independence –LCI models) is non-empty. This paper shows that the conditional independence models in the intersection can be characterized as labeled trees, where every vertex on the tree corresponds to a single random variable. This fact leads to the definition of a specific Markov property for trees and therefore to the introduction of trees as part of the family of graphical Markov Models.

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 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 comprehensive guide to the 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 (UDG models), directed acyclic 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, Madigan, Perlman, and Triggs (1997, 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.

Figure 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

[1]Sometimes also referred as acyclic directed graph –ADG

of those UDG models that are equivalent to certain type of DAG models (Wermuth, 1980; Kiiveri, Speed, & Carlin, 1984). Thus, undirected and directed graphs members of this intersecting class describe the same model of conditional independence. They are graphically defined as chordal graphs and are known as DEC models.

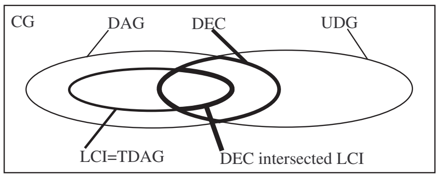

Figure 1: Relation among the classes of chain graph models (CG), directed acyclic graph models (DAG), undirected graph models (UDG), decomposable models (DEC) and lattice conditional independence models (LCI)

As has been mentioned already, the intersection between the classes of DEC and LCI is non-empty, as proven by Andersson et al. (1995). In this paper, a new formalization of the graphical Markov models in of DEC$\cap$LCI is presented. In the first place, this new formalization is based on a characterization of moral TDAGs as labeled trees. Then, a Markov property for trees is 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* –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 some graph-theoretic definitions and notation will be introduced. In section 3 an overview of Graphical Markov models will be given, and it will serve to introduce section 4, where we will find the characterization of moral TDAG models as trees, as well as the definition of its specific Markov property. In section 5, the notion Markov equivalence in this setting will be investigated. Finally, on section 6, the main issues of the paper will be summarized.

2 Background concepts, terminology and notation

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

A graph is a pair $G = (V, E)$ where V is the set of vertices and E is the set of edges. In the present context of graphical Markov models, the set of vertices V represents the set of random variables of the multivariate distribution that underlies the model. This multivariate distribution is of a family of probability distributions P defined on a product space $\mathcal{X} = \times(\mathcal{X}_i | i \in V)$. For simplicity, it is convenient to refer to a random variable x_i as i, and a set of random variables $x_A = \{x_i | i \in A, A \subseteq V\}$, as A. Therefore, a statement of conditional independence regarding three subsets of random variables may be here specified as $A \perp B | S[P]$, where $A, B, S \subseteq V$, A, B are non-empty. It claims that the random variables in x_A are conditionally independent of the random variables in x_B given the random variables in x_S under P. In the rest of this paper every statement of conditional independence is asserted under P, thus it will be dropped from the notation.

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., $\forall a \in V \ (a, a) \notin E$. For a given pair of vertices $a, b \in V \ a \neq b$, a solid line in the graph joining them a–b will represent an undirected edge, i.e., it means that $(a, b) \in E$ and $(b, a) \in E$. An arrow $a \rightarrow b$ between these two vertices will represent a directed edge, and it means that $(a, b) \in E$ and $(b, a) \notin E$. A *subgraph* $G_S = (S, E_S)$ is given by a subset $S \subseteq V$, and the induced edge set $E_S = E \cap (S \times S)$.

When two vertices are joined by an (un)directed edge, these two vertices are regarded as *adjacent*. Given a vertex $v \in V$, the set of those vertices that are adjacent to it are known as the *boundary* of v, denoted by $bd(v)$. Further, the *closure* of a vertex v is defined as $cl(v) = bd(v) \cup \{v\}$. A graph $G = (V, E)$ is said to be *complete* iff $(x, y) \notin E \Rightarrow (y, x) \in E$, or in other words, every possible pair of vertices is adjacent. A *subset* is *complete* if it induces a complete subgraph. A *clique* is a complete subset that is maximal with respect to $\subseteq$, i.e., G has no larger complete subgraph that contains it.

For a directed edge $a \rightarrow b$ we distinguish between the two joined vertices by specifying that a is the *parent* of b, and that b is the *child* of a. Those parent vertices that have a common child, will be considered as the *parent set* of this child vertex, and it will be noted as $pa(v)$, being v the child vertex. An important concept regarding directed graphs in the context of conditional independence is the concept of *immorality*. An *immorality* is formed by two non-adjacent vertices with a common child. A directed graph without immoralities is called a *moral* graph.

A directed graph (V, E) can be *moralized* by *marrying* non-adjacent parents (joining them with an undirected edge) and dropping directions on the edges in E. Given a directed graph $G = (V, E)$, its moralized version will be noted as G^m. An immorality is also known as a *sink-oriented* V-configuration. Cox and Wermuth (1996) define a *V-configuration* as a triplet of vertices (a, b, c) such that two of them are adjacent with the third one but they are not adjacent themselves. Therefore a sink-oriented V-configuration (an immorality) for the previous three vertices would be, for instance, $a \rightarrow b \leftarrow c$, and the vertex b is called a *collision* vertex. Following the same terminology, other two types of V-configurations are the *source-oriented* V-configuration, e.g. $a \leftarrow b \rightarrow c$, and the *transition-oriented* V-configuration, e.g. $a \rightarrow b \rightarrow c$.

In a directed or undirected graph $G = (V, E)$, a *path* from a to b is a sequence $a = a_0, \ldots, a_n = b$ of distinct vertices such that $n > 0$ and either $(a_{i-1}, a_i) \in E$ or $(a_i, a_{i-1}) \in E$ for $i = 1, \ldots, n$. Given three subsets of vertices $A, B, S \subseteq V$, it is said that S *separates* A from B in an undirected graph iff every path between vertices in A and B, intersects S.

An *undirected cycle* is a path where $a = b$. A *tree* is a connected undirected graph without undirected cycles such that there is always only a unique path between any two different vertices from the graph. 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 considered at the top of the hierarchy. The leaves of a rooted tree are those vertices connected to just one other vertex and they are considered at the bottom of the hierarchy. Under this convention we will say that the root is *above* the leaves, and the leaves are *below* the root. Given a tree $T = (V, E)$ and a vertex $u \in V$, a *subtree* rooted at u, and noted 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 below, and the edge set $E_U = E \cap (U \times U)$.

In a directed graph, a *directed path* is a *direction-preserving* path, that means all its edges point towards the same direction. A given vertex a is called the *ancestor* of b if there is a directed path from a to b. A *directed cycle* is a directed path where the first vertex coincides with the last.

A directed acyclic graph (DAG) is a directed graph without directed cycles. For every vertex v, one may consider the set of those vertices that are ancestors of v, which it will be called the *ancestor set* of v, and noted $an(v)$. From the definition of ancestor set, it follows that $pa(v) \subseteq an(v)$. In the same manner, a vertex b is called the *descendant* of a if there is a directed path from a to b (i.e. a is ancestor of b), and all the vertices reachable from a by directed paths will form the *descendant set* of a. Given a vertex v the descendant set will be noted as $de(v)$ and the *non-descendant* set of v is defined as $nd(v) = V \backslash (de(v) \cup \{v\})$. A DAG is said to be *transitive* –TDAG if for every vertex v, $pa(v) = an(v)$.

An undirected graph is *chordal*, or *decomposable* (DEC), iff it does not contain undirected cycles of length greater than three 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 UDG models, and therefore they characterize those UDG models that are equivalent to DAG models. In the same vein, it is possible to characterize those DAG models equivalent to UDG models, as those determined by a DAG that does not contain immoralities (sink-oriented V-configurations), i.e. a moral DAG.

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 $v \in A$, $an(v) \subseteq A$. 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 noted as $\mathcal{A}(G)$. Further, given a subset of vertices $A \subset V$, the smallest ancestral subset that contains A is called the *smallest ancestral set* of A and denoted $An(A)$. To avoid confusion, let's remark the difference between $an(v)$ and $An(A)$. The former one refers to the set of vertices that are ancestors of the vertex v, while the latter refers to the smallest subset $An(A) \subseteq V$ that contains a given subset $A \subset V$ such that $An(A)$ is ancestral in G.

A set S equipped with an order relation[2] is called a *partially ordered set*, or *poset*. If linearity holds[3], then the set is fully ordered, or 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 $x \leq z < y \Rightarrow z = x$.

[2] reflexive, antisymmetric and transitive

[3] $\forall a, b \in S \;\; a \leq b \;\; or \;\; b \leq a$

Gratzer (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 $a \leq b$ iff $a = b$ or there exists a finite sequence of elements $x_0, \ldots, x_{n-1}$, such that $x_0 = a, x_{n-1} = b$, and $x_i \prec x_{i+1}$, for $0 \leq i < n-1$.

A poset S 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 on the plane.

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 $h \in H$, $h \leq a$ ($h \geq a$). 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 $a \leq b$ ($a \geq b$), and note it $a = \sup H$ ($a = \inf H$).

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 for any finite nonvoid subset H of L. Gratzer (1978) shows that the concept of a lattice as a poset is equivalent to the concept of a lattice as an algebra $L(\wedge, \vee)$. Where $\wedge$ and $\vee$ are binary operations on pairs of elements $a, b \in L$, 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. Gratzer (1978, p. 62) characterizes finite distributive lattices as those isomorphic to a ring of sets.

A finite distributive lattice L has a unique irredundant representation as a finite poset $J(L) \subseteq L$, known as set of *join-irreducible* elements (Gratzer, 1978, p. 62). This poset is substantially smaller than L, and its elements are defined in the following way.

$$J(L) = \{a \in L | a \neq \emptyset, a = b \vee c \Rightarrow a = b \ \ or \ \ a = c\}$$

In this context, the lattice L can be constructed by unions ($\vee$) and intersections ($\wedge$) of the elements of the set of join-irreducible elements $J(L)$. 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 is covering exactly one other single element.

Analogously to the concept of ancestral set for vertices of a DAG, that we have seen before, one may define *ancestral poset*. Let $J(L)$ be a poset, a subset (which is again a poset) $A \subseteq J(L)$ is ancestral in $J(L)$ iff $\forall a \in A, b \in J(L)$ it follows that $b < a \Rightarrow b \in A$.

It is possible to establish a mapping between finite posets and TDAGs if we consider that $a < b \Leftrightarrow a \in an(b)$ for any $a, b \in V$ from a TDAG $G = (V, E)$. Conversely, given the finite poset $J(L)$ we can build a TDAG $G = (J(L), E^{<})$, where

$$E^{<} = \{(a, b) \in J(L) \times J(L) | a < b\}$$

This mapping between TDAGs and finite distributive lattices is used by Andersson, Madigan, Perlman, and Triggs (1997) to prove that TDAG models and LCI models coincide.

3 Graphical Markov models

This section gives an overview of graphical Markov models. Attention will be paid to UDG, DAG, DEC, TDAG models, and specially to TDAG models in its equivalent form of LCI models, since this latter class is, in a large part, the basis of the main contribution of this paper.

The glue that binds the structural information of a graph, with the structural information of a multivariate distribution P, is *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, Madigan, Perlman, and Triggs (1997).

Definition 3.1. *Undirected pairwise Markov property (UPMP)*
Let $G = (V, E)$ be a UDG, a probability distribution P on $\mathcal{X}$ is said to satisfy the undirected pairwise Markov property *(UPMP) if, for any pair $u, v \in V$ of non-adjacent vertices, P satisfies*

$$u \perp v | V \setminus \{u, v\}$$

Definition 3.2. *Undirected local Markov property (ULMP)*
Let $G = (V, E)$ be a UDG, a probability distribution P on $\mathcal{X}$ is said to satisfy the undirected local Markov property *(ULMP) if, for any vertex $v \in V$, P satisfies*

$$v \perp (V \setminus cl(v)) | bd(v)$$

Definition 3.3. *Undirected global Markov property (UGMP)*
Let $G = (V, E)$ be a UDG, a probability distribution P on $\mathcal{X}$ 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*

$$A \perp B | S$$

These three Markov properties are defined for DEC models as well, since DEC models are determined by chordal graphs and all chordal graphs are undirected graphs. These properties are related in the following way (Lauritzen et al., 1990),

$$UGMP \Rightarrow ULMP \Rightarrow UPMP$$

meaning that the UGMP is the sharpest possible rule to read off conditional independencies from a UDG (or DEC) model. At this point, it is possible to introduce the formal definition of a specific type of graphical Markov model, the UDG model.

Definition 3.4. *UDG model*
Let G be a UDG. The set $\mathbf{U}_{\mathcal{X}}(G)$ of all probability distributions on $\mathcal{X}$ that satisfy the UGMP relative to G is called the Markov model *determined by G, or more specifically, the* UDG 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. The next set of graphs to consider are DAGs. The Markov properties of DAGs are defined as follows.

Definition 3.5. *Directed pairwise Markov property (DPMP)*
Let $G = (V, E)$ be a DAG, a probability distribution P on $\mathcal{X}$ is said to satisfy the directed pairwise Markov property *(DPMP) if, for any pair u, v of non-adjacent vertices such that $v \in nd(u)$, P satisfies*

$$u \perp v | nd(u) \backslash \{v\}$$

Definition 3.6. *Directed local Markov property (DLMP)*
Let $G = (V, E)$ be a DAG, a probability distribution P on $\mathcal{X}$ is said to satisfy the directed local Markov property *(DLMP) if, for any vertex $v \in V$, P satisfies*

$$v \perp (nd(v) \backslash pa(v)) | pa(v)$$

Definition 3.7. *Directed global Markov property (DGMP)*
Let $G = (V, E)$ be a DAG, a probability distribution P on $\mathcal{X}$ 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 $A \cup B \cup S$, $(G_{An(A \cup B \cup S)})^m$, P satisfies*

$$A \perp B | S$$

Lauritzen et al. (1990) show that the latter Markov property, the DGMP, is equivalent to the separation criteria from Pearl and Verma (1987). They also prove that the three Markov properties for DAG models are equivalent, that is,

$$DGMP \Leftrightarrow DLMP \Leftrightarrow DPMP$$

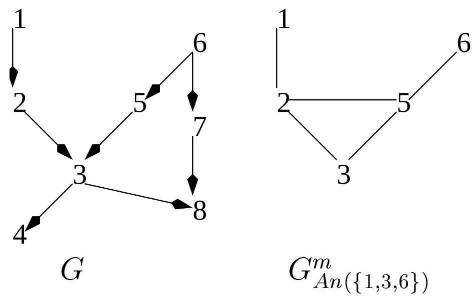

Figure 2: DAG on the left, and on the right its moralized version over the smallest ancestral set of $\{1, 3, 6\}$.

We may see more clearly the intuition behind this Markov property in figure 2. Given the DAG G on the lefthand side of the figure, let's try to find out whether the DGMP holds for $1 \perp 6 | 3$ in G. The smallest ancestral set of 1,6 and 3 is $An(\{1, 3, 6\}) = \{1, 2, 3, 5, 6\}$. The moralized version of the graph induced by this subset of vertices is on the righthand side of the figure. Since the subgraph over $An(\{1, 3, 6\})$ contains the immorality $2 \rightarrow 3 \leftarrow 5$, 2 and 5 become adjacent at the moment we moralize the graph, creating therefore, a new path between 1 and 6, which does not intersect the conditioning set $\{3\}$, thus $1 \perp 6 | 3$ *does not* hold. In order to find 1 separated from 6 one should add 2 or 5 to the conditioning set, or remove 3. As in the undirected case, let's introduce now the formal definition of DAG model.

Definition 3.8. *DAG model*
Let G be a DAG. The set $\mathbf{D}_{\mathcal{X}}(G)$ of all probability distributions on $\mathcal{X}$ that

satisfy the DGMP relative to G is called the Markov model *determined by G, or more specifically, the* DAG model *determined by G.*

Finally, let's move to the case of TDAG, and LCI, models. This is the only one new Markov property attached to this class of models, and below we will find how its related to other Markov properties within the class of TDAG models.

Definition 3.9. *Lattice conditional independence Markov property (LCIMP) Let $G = (V, E)$ be a TDAG, a probability distribution P on $\mathcal{X}$ is said to satisfy the* lattice conditional independence Markov property *(LCIMP) if, for every pair of ancestral subsets $A, B \in \mathcal{A}(G)$, P satisfies*

$$A \perp B | A \cap B$$

Theorem 3.1. *Andersson, Madigan, Perlman, and Triggs (1997, theorem 3.1,p. 33)*
Let G be a TDAG. For any probability distribution P on $\mathcal{X}$,

$$DGMP \Leftrightarrow DLMP \Leftrightarrow DPMP \Leftrightarrow LCIMP$$

Andersson, Madigan, Perlman, and Triggs (1997) defined the LCIMP for the ancestral sets of a DAG. In this more general case, they prove:

Theorem 3.2. *Andersson, Madigan, Perlman, and Triggs (1997, theorem 2.2,p. 32)*
Let G be a DAG. For any probability distribution P on $\mathcal{X}$, DGMP$\Rightarrow$LCIMP.

The definition of an LCI model is then as follows.

Definition 3.10. *LCI model*
Let G be a TDAG. The set $\mathbf{L}_{\mathcal{X}}(G)$ of all probability distributions on $\mathcal{X}$ that satisfy the LCIMP relative to G is called the Markov model *determined by G, or more specifically, the* LCI model *determined by G.*

So far, we have been dealing with graphs as the graphical counterpart of graphical Markov models. LCI models are special in this aspect, since they can be specified not only in terms of TDAGs as we have just seen, but also as rings of subsets. This comes from the fact that TDAGs are the same mathematical objects than finite distributive lattices. In these terms, let V be an index set in which each element represents a random variable from the multivariate distribution x_V. One may consider a ring $\mathcal{K}$ of subsets of V, such that for every pair of subsets $L, M \in \mathcal{K}$, a probability distribution P satisfies

$$L \perp M | L \cap M$$

as in the LCIMP, and the subsets L, M refer to subsets of random variables $x_L, x_M \subseteq X_V$ that take values from a larger product space $\mathcal{X} = \times(\mathcal{X}_i | x_i \in x_V)$ 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 $\mathbf{L}_{\mathcal{X}}(\mathcal{K})$ that, as the notation suggests, is determined by a ring $\mathcal{K}$. For more details about LCI models determined by rings of subsets, the reader may consult Andersson and Perlman (1993), Andersson, Madigan, Perlman, and Triggs (1997).

In figure 3a 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 figure 3b we may see a complete DAG, which represents the unrestricted or saturated DAG model, and its representation by a hasse diagram on its right as the unrestricted LCI model. Let's note that for the LCI model on 3a, $J(\mathcal{K}_a) = \{1, 2, 3\}$ and for the LCI model on 3b, $J(\mathcal{K}_b) = \{1, 12, 123\}$.

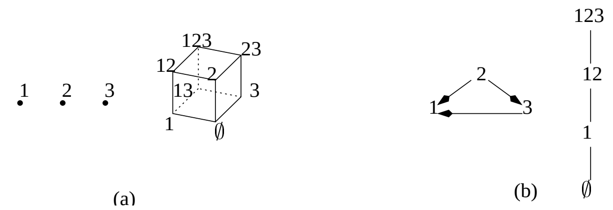

Figure 3: Comparison between DAG models and LCI models

While for the graphical Markov model in 3a the restrictions are characterized by being all three vertices marginally independent $1 \perp 2 \perp 3$, the set of restrictions of the model in 3b 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). So, for instance two trivial cases are those from figure 3.

On the left hand side of (a) and (b), the fully restricted DAG model and the unrestricted DAG model, respectively represented by DAGs. On the right hand side of (a) and (b), the fully restricted LCI model and the unrestricted LCI model, respectively represented by hasse diagrams.

In the next figure, we may see two more sophisticated models. The one on 4a corresponds to the immorality that induces the two non-adjancent vertices marginally independent, and the one on 4b corresponds to the source-oriented V-configuration that makes the two non-adjacent vertices conditionally independent. On the LCI model of 4a, $J(\mathcal{K}_a) = \{1, 3, 123\}$ and on LCI model of 4b, $J(\mathcal{K}_b) = \{2, 12, 23\}$ (recall that an element belongs to $J(\mathcal{K})$ iff it covers only one other element).

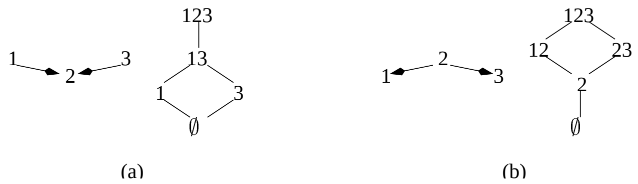

Figure 4: On the left hand side of (a) and (b), two DAGs representing $1 \perp 3|\emptyset$ and $1 \perp 3|2$ respectively, and on the right hand side of (a) and (b) their corresponding Hasse diagrams

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.

Recall that a DAG is transitive, i.e. a TDAG, if for every vertex v, $pa(v) = an(v)$. A Markov model member of the class DEC∩LCI is determined by a TDAG with no immoralities (Andersson et al., 1995). It inmediately follows that

Proposition 4.1. *Let* $\mathbf{L}_{\mathcal{X}}(G)$ *be a LCI model member of the intersection class DEC∩LCI, where* $G = (V, E)$ *is a moral TDAG. For every* $v \in V$*, the set* $\{x\} \cup pa(x)$ *is a complete set in G.*

Proof. If $\{v\} \cup pa(v)$ was not complete, then at least two nodes $u, w \in pa(v)$ would not be adjacent, creating an immorality. In such case G would not be a moral TDAG and therefore $\mathbf{L}_{\mathcal{X}}(G) \notin \text{DEC} \cap \text{LCI}$, which contradicts the first assumption of the proposition. □

Given a TDAG G that determines a graphical Markov model $\mathbf{L}_{\mathcal{X}}(G) \in \text{DEC}\cap\text{LCI}$, the corresponding finite distributive lattice $\mathcal{K}$ will have a set $J(\mathcal{K})$ of join-irreducible elements which, it may be characterized as a composition of several maximal chains. These chains are, in fact, induced by the *cliques* of the TDAG G.

Lemma 4.1. *Let G be a TDAG, such that $\mathbf{L}_{\mathcal{X}}(G) \in$DEC$\cap$LCI. Let $\mathcal{C}$ be the set of cliques of G. Let G_C be the TDAG that corresponds to a given clique $C \in \mathcal{C}$. The TDAG $G_C = (V_C, E_C)$ induces a complete order among the vertices of V_C such that it coincides with a maximal chain H_C derived from the vertex set V_C. Let $\mathcal{K}$ be the finite distributive lattice that coincides with G. The set of join-irreducible elements $J(\mathcal{K})$ that represents the lattice $\mathcal{K}$ may be defined as*

$$J(\mathcal{K}) = \bigcup_{C \in \mathcal{C}} H_C$$

Proof. We are going to show that (a) every poset defined that way is a join-irreducible set and (b) that every join-irreducible set of a LCI model from DEC$\cap$LCI corresponds to the union of maximal chains induced by the cliques of a TDAG that defines a graphical Markov model member of DEC$\cap$LCI.

(a) Recall that the lattice $\mathcal{K}$ is generated by unions and intersections from its set of join-irreducible elements $J(\mathcal{K})$. To prove that the union of the maximal chains H_C forms the join-irreducible set $J(\mathcal{K})$, we need to see that no element, generated by unions and intersections of the elements in $\bigcup H_C$, is a second lower cover for any element in $\bigcup H_C$.

    Clearly, the union of two elements from $\bigcup H_C$ will not create a second lower cover for any element in $\bigcup H_C$. However, intersection of two elements of $\bigcup H_C$ needs to be carefully examined.

    If the vertex sets V_C of every clique $C \in \mathcal{C}$ are disjoint, the intersection of any two elements of $\bigcup H_C$ is empty. Hence $\bigcup H_C$ is join-irreducible.

Since the TDAG $G = (V, E)$ is moral and for every $v \in V$, $pa(v) = an(v)$, the vertex set V_C of every clique $C \in \mathcal{C}$ is ancestral in G. Therefore, the non-empty intersection of the two vertex sets of any given two cliques of G, is ancestral too. This means that any of two such maximal chains H_{C_1} and H_{C_2} will be of the form $H_{C_1} = \{x_1, \ldots, x_l, y_1, \ldots, y_n\}$ and $H_{C_2} = \{x_1, \ldots, x_l, z_1, \ldots, z_m\}$, where $x_i < y_j$, $x_i < z_k$, $y_j \not< z_k$ and $z_k \not< y_j$ for $1 \leq i \leq l$, $1 \leq j \leq n$ and $1 \leq k \leq m$.

For any two incomparable elements y_j and z_k such that $y_j \cap z_k = w$, obviously $w < y_j, z_k$ and $w \in H_{C_1}$, $w \in H_{C_2}$. Therefore, $w \in \{x_1, \ldots, x_l\}$, and since $x_i < y_j$ and $x_i < z_k$ it follows that $w = x_l$, so the set $\bigcup H_C$ remains join-irreducible.

(b) Let's consider that the mapping between the poset $J(\mathcal{K})$ and the TDAG G is of the form $G = (J(\mathcal{K}), E^<)$ where

$$E^< = \{(a, b) \in J(\mathcal{K}) \times J(\mathcal{K}) | a < b\}$$

Let H be a maximal chain in $J(\mathcal{K})$. Let $c \in H$ and $a \in an(c)$, such that then $a < c$. Let $b \in H$ and without loss of generality $b < c$. From $a < c$ and $b < c$ it follows that $(a, c) \in E^<$ and $(b, c) \in E^<$. Since the TDAG G induced by $E^<$ is moral, either $(a, b) \in E^<$ or $(b, a) \in E^<$, thus $a < b$ or $b < a$. Therefore $a, b, c \in H$ are all comparable so $a \in H$ and H is an ancestral poset.

Let $C_1 = (V_1, E_1)$ and $C_2 = (V_2, E_2)$ be two cliques created from two maximal chains H_1 and H_2 in $J(\mathcal{K})$. If the vertex sets V_1 and V_2 are disjoint ($V_1 \cap V_2 = \emptyset$), the graph that results from the union of these two cliques $G_{12} = (V_1 \cup V_2, E_1 \cup E_2)$ is a moral TDAG. If $V_1 \cap V_2 \neq \emptyset$, then it goes as follows. Because H_1 and H_2 were ancestral posets, V_1 and V_2 are ancestral in G_{12}. Therefore, $an(v) \subseteq V_1$ (or $an(v) \subseteq V_2$) and since in a TDAG $pa(v) = an(v)$ the parent set $pa(v)$ induces a complete subgraph and the resulting TDAG is moral. By doing the union of all the cliques derived from all the maximal chains in $J(\mathcal{K})$ we will obtain moral TDAG G which will determine a graphical Markov model $\mathbf{L}_{\mathcal{X}}(G) \in$DEC$\cap$LCI.

□

Let's consider the following mapping between a set $J(\mathcal{K})$ of join-irreducible elements of a finite distributive lattice $\mathcal{K}$, and a labeled tree T.

$$\mu : J(\mathcal{K}) \longrightarrow T = (J(\mathcal{K}) \cup \{\emptyset\}, E^{\prec})$$

The finite distributive lattice $\mathcal{K}$ coincides with some moral TDAG $G = (V, E)$, and the labeled tree T has a vertex set formed by the elements of $J(\mathcal{K})$ plus an extra vertex labeled $\emptyset$, that acts as the root, and the set of edges

$$E^{\prec} = \{(a, b) \in J(\mathcal{K}) \times J(\mathcal{K}) | b \prec a\} \cup \{(\emptyset, a) \in \{\emptyset\} \times J(\mathcal{K}) | \not\exists\, b \in J(\mathcal{K})\ b \prec a\}$$

Where $\prec$ is the covering relation taken over the set of join-irreducible elements $J(\mathcal{K})$. Note that there is a one to one correspondence between $J(\mathcal{K})$ and the set of vertices V from the corresponding TDAG G. From the next three propositions it will follow that the mapping μ is a bijection between moral TDAGs and labeled trees.

Proposition 4.2. *Let $\mathcal{K}$ be a finite distributive lattice that coincides with some moral TDAG. The graph $\mu(J(\mathcal{K}))$ is a labeled tree.*

Proof. From lemma 4.1 we can decompose $J(\mathcal{K})$ in several maximal chains H_C. Every $\mu(H_C)$ is a path in $\mu(J(\mathcal{K}))$ from the root to a leaf an viceversa. For any of such two chains H_{C_1} and H_{C_2}, $\mu(H_{C_1}) \cap \mu(H_{C_2})$ is a unique path from the root to a vertex. It follows directly that $\mu(J(\mathcal{K}))$ has no cycles and therefore is a labeled tree. □

Proposition 4.3. *The mapping μ is injective, for a finite distributive lattice $\mathcal{K}$ that coincides with a moral TDAG.*

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

Proposition 4.4. *The mapping μ is surjective, for a finite distributive lattice $\mathcal{K}$ that coincides with a moral TDAG.*

Proof. Let's consider labeled trees where the root is labeled as $\emptyset$ and the rest of the vertices using natural numbers $\{1, \ldots, n\}$. From the fact that T is a tree, there is always a unique path from the root to every of its leaves. For every path p of the tree T, such that $p = \{\emptyset, x_1, \ldots, x_n\}$, let's take out the root label $\emptyset$, and from the rest of the path $\{x_1, \ldots, x_n\}$ let's build a chain H_C such that $H_C = \{x_1, \{x_1, x_2\}, \ldots, \{x_1, \ldots, x_n\}\}$. From lemma 4.1 we know that the union of these chains H_C produces a set of join-irreducible elements $J(\mathcal{K})$ corresponding to a lattice $\mathcal{K}$ that coincides with a moral TDAG. □

Finally, we can establish the following result.

Theorem 4.1. *Moral TDAGs coincide with labeled trees.*

Proof. It follows directly from the fact that the mapping μ is a bijection between moral TDAGs and labeled trees. □

The bijection between moral TDAGs and labeled trees shows the way to construct a labeled tree that corresponds *uniquely* to a given moral TDAG. Let's remark that we are not talking yet in terms of Markov models, but just from a pure graph-theoretic perspective. To illustrate this construction, we may find in figures 5a, 5b and 5c, the trees corresponding to the moral TDAGs that appear on figures 3a, 3b and 4b, respectively.

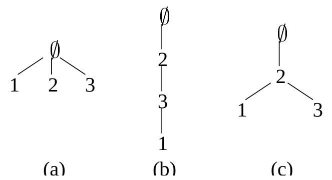

Figure 5: From left to right, trees constructed from an empty moral TDAG, a complete moral TDAG and a moral TDAG formed by a single source-oriented V-configuration

One of the features that distinguishes the trees from other types of graph used in the context of graphical Markov models is that it is a connected structure. In this sense, they are quite similar to the hasse diagrams used to represent lattices. Thus, we may observe in between figures 3a and 5a how a complete disconnected graph turns into a connected structure, a 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 trees. To define this formally, we need two new concepts regarding trees and the following proposition.

Proposition 4.5. *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 in the two unique paths that lead from u and v to the root $\emptyset$. The minimal element is the root of T, $\emptyset$.*

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 a common minimal element for every two paths from two given vertices to the root, let's consider the next two definitions.

Definition 4.1. *Meet*
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 none of them is 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 noted 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 none of them is 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 noted $mp(\varphi_{u,w}) = \{\varphi_{u,w}, \ldots, \emptyset\}$.*

As we may 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 $\mathcal{X}$ 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 $mp(\varphi_{u,w})$, P satisfies*

$$U \perp W | mp(\varphi_{u,w})$$

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

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

Let's consider three TCI models determined by the trees in figure 5. By the TCIMP the tree on (a) renders the three vertices marginally independent $1 \perp 2 \perp 3$, from the tree on (b) is not possible to read off any conditional independency, thus the set of restrictions of the model is empty and on (c) we may see that the vertex 2 is the meet of vertices 1 and 3, thus $1 \perp 3 | 2$. These restrictions were already read from the TDAG models from which these trees have been built, at the end of the previous section. Further examples are provided by figures 6 and 7.

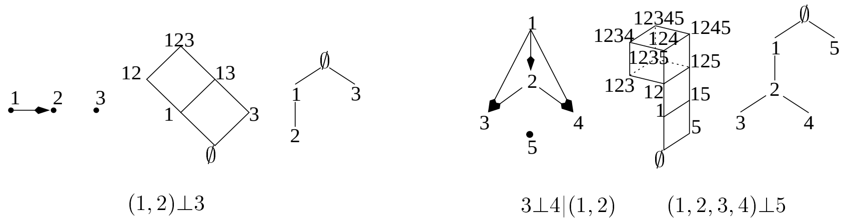

Figure 6: Two different Markov models represented by a TDAG model, a LCI model and a TCI model

In figure 6 we see the different graphical representations for two simple models of conditional independence, with the independencies as specified. In figure 7 we find a larger model which may help understanding the TCIMP. For instance, if we pick 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 first.

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 $v \in A$, $an(v) \subseteq A$ and $an(v) \cup \{v\}$ is complete in G.*

Proposition 4.6. *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 find out whether the union of moral ancestral sets is moral.

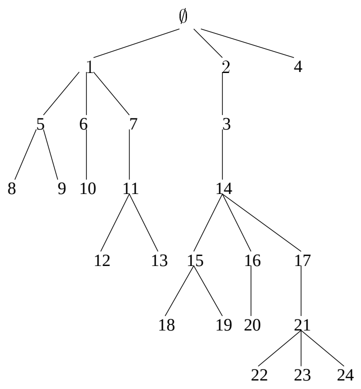

Figure 7: Example of a TCI model for 24 variables

Let $a, b, c \in A \cup B$ such that $a \rightarrow c \leftarrow b$. Since $a, b \in an(c)$, 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. $\square$

Proposition 4.7. *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. It follows directly from the fact that, the intersection of ancestral sets is again ancestral, and since $A \cap B \subseteq A$ and $A \cap B \subseteq B$ then $A \cap B$ should be moral otherwise it would contradict the assumption of A, B being moral ancestral. $\square$

From the above propositions it follows that all the different moral ancestral sets contained in a DAG $G = (V, E)$ form a ring of subsets of V, which will be called as the moral ancestral ring of G, and noted $\mathcal{A}^m(G)$. 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 $\mathcal{X}$ is said to satisfy the directed tree conditional independence Markov property *(DTCIMP) if, for every pair of moral ancestral subsets $A, B \in \mathcal{A}^m(G)$, P satisfies*

$$A \perp B | A \cap B$$

Theorem 4.2. *Let $\mathbf{D}_{\mathcal{X}}(G)$ be a DAG model. For any probability distribution P on $\mathcal{X}$,*

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

Proof. From left to right. The first implication follows from theorem 3.2. Let $A, B \in \mathcal{A}^m(G)$. The LCIMP implies the DTCIMP if $A, B \in \mathcal{A}(G)$, and this follows from $\mathcal{A}^m(G) \subseteq \mathcal{A}(G)$.

The third implication may be proved as follows. For any pair $A, B \in \mathcal{A}^m(G)$, 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 4.1. The DTCIMP will imply the TCIMP if for each pair of vertices $a \in A \backslash B$ and $b \in B \backslash A$, $mp(\varphi_{a,b}) = A \cap B$ 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 $a \in A\backslash B$ and $b \in B \backslash A$, is formed by those vertices that are common to A and B, therefore $A \cap B$. □

Theorem 4.3. *Let G be a moral TDAG. For any probability distribution P on $\mathcal{X}$, TCIMP⇒DGMP. Thus, for a moral TDAG,*

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

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

Proof. By theorem 4.1, there is a unique 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 $mp(\varphi_{u,w})$ in

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

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. Let's 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 traverse at some point the sets $pa(u)$ and $pa(w)$, because of transitivity. And more concretely, this path will always traverse those vertices $x \in pa(u) \cap pa(w)$. The set $pa(u) \cap pa(w)$ is equivalent to the definition of meet path, hence $mp(\varphi_{u,w})$ separates U, W in $G_{An(U\cup W\cup mp(\varphi_{u,w}))}$. The second part of the theorem follows from theorems 3.1 and 4.2. □

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

Theorem 4.4. *The class of TCI models coincides with the class of DEC$\cap$LCI models.*

Proof. It follows from the fact that $\mathbf{D}_{\mathcal{X}}(G) = \mathbf{T}_{\mathcal{X}}(T)$, for some moral TDAG G and some tree T, which is proved on theorem 4.3. $\square$

5 Markov equivalence among TCI models

DAG models are organized in classes of equivalence, such that two DAG models $\mathbf{D}_{\mathcal{X}}(G_1)$ and $\mathbf{D}_{\mathcal{X}}(G_2)$ determined by two different DAGs G_1 and G_2 may actually infer the same model of conditional independence, hence $\mathbf{D}_{\mathcal{X}}(G_1) \equiv \mathbf{D}_{\mathcal{X}}(G_2)$. In this context we use $\equiv$ to denote that two graphical Markov models are *structurally* equivalent. This situation is reproduced as well in the case of TCI models, so for two different trees T_1, T_2, they may determine the same TCI model $\mathbf{T}_{\mathcal{X}}(T_1) \equiv \mathbf{T}_{\mathcal{X}}(T_2)$. We will investigate now the notion of Markov equivalence among TCI models. First, let's review the notion of Markov equivalence for DAG models, which was given independently by Frydenberg (1990), 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 TDAG using theorem 4.1 and applying the previous theorem. The notion specifically for TCI models is as follows.

Theorem 5.2. *Two TCI models are Markov equivalent if and only if they have the same subsets of vertices induced from every path starting at each of its leaves to the root $\emptyset$.*

Proof. (Necessity). Let's assume that two given TCI models T and T' are Markov equivalent. 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 a tree $T = (V, E)$, with meet $\varphi_{u,v}$, inducing subtrees $T_u = (U, E_U), T_w = (W, E_U)$ such that none of them is subtree of the other. In order to find the meet path $mp(\varphi_{u,v})$ for u, v in both trees T

and T', the paths from u, v to the root must intersect in the same vertices in T and T'. Provide that 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 to be the same in T and T', is when all paths from the leaves to the root $\emptyset$ in T and T' involve the same subsets of vertices.

(Sufficiency). Let's assume that two given TCI models T and T' have the same subsets of vertices induced from every path from each of its leaves to the root $\emptyset$. This implies that every possible meet path from any given two subtrees, in T must exist in T'. Because, if there would be a meet path that differs in at least one vertex for T and T', there would be subsets of vertices from one or more leaves to the root $\emptyset$ that differ in T and T'. Therefore, if every meet path for any two given subtrees in T and T', exists in T and T', it follows from the TCIMP that the collection of Markov properties of T hold also in T'. $\square$

In order to illustrate the notion of Markov equivalence among trees, let's look at figure 8. The pairs of trees on part (a) are Markov equivalent because although two vertices are swapped for every pair, 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 $\emptyset$.

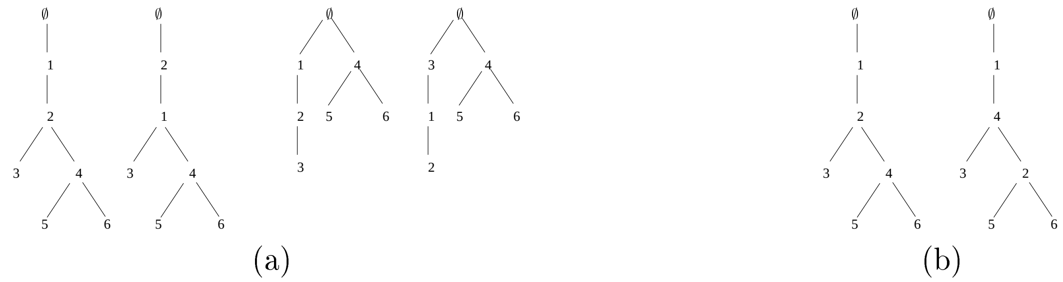

Figure 8: Markov equivalence between TCI models. The two pairs on (a) are Markov equivalent, while the pair on (b) is not.

5.1 How many different Markov models

The notion of Markov equivalence uncovers the fact that two graphical Markov models determined by two different graphs, 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 certain 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 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 UDG models, which are represented by undirected graphs, since there is a one to one correspondence between undirected graphs and models of conditional independence. Two UDG models $\mathbf{U}_{\mathcal{X}}(G_1), \mathbf{U}_{\mathcal{X}}(G_2)$ are Markov equivalent, $\mathbf{U}_{\mathcal{X}}(G_1) \equiv \mathbf{U}_{\mathcal{X}}(G_2)$, iff $G_1 = G_2$. Thus, there is the same number of different models of conditional independence, as different undirected graphs, i.e., $2^{\binom{n}{2}}$.

The case of DEC models is also straightforward, since chordal graphs also have a one to one correspondence with Markov equivalence classes. Their enumeration is slightly more difficult and it will be treated in this section, firstly since it is related to the enumeration of Markov equivalence classes of TCI models and secondly DEC models have not been counted yet[4].

The case of DAG models is a difficult one. Markov equivalence classes of DAG models are represented by essential graphs (Andersson, Madigan, & Perlman, 1997), which are partially directed acyclic graphs. An efficient way of enumerating such graphs is not yet known, and they have been counted only up to five vertices. For comparison, in table 2 their number is given together with those of DEC and TCI models.

The enumeration of Markov equivalence classes of TCI models, provides insight into their nature such that, afterwards, it will be easy to devise a canonical representation for a given equivalence class of TCI models.

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

[4]to the best knowledge of the authors

number of vertices of a graph. In the case of labeled enumeration, one uses an *exponential generating function* as the one in (1). For full insight into this subject the reader should consult the book of Harary and Palmer (1973).

$$g(x) = \sum_{n=1} a_n \frac{x^n}{n!} \tag{1}$$

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. Let's consider now another exponential generating function to count, not only *connected* labeled chordal graphs, but *all* of them.

$$G(x) = \sum_{n=0} A_n \frac{x^n}{n!} \tag{2}$$

In this generating function, the coefficients A_n equal the number of *all* chordal graphs with n vertices, which correspond 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 come to terms in the following relation

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

Where 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. In such case one may remove the constant 1 of the previous expression. As we shall see now, by differenciating 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 differenciate the equation, which leads to the following form.

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

The polynomial at the right-hand side of the equation should be multiplied by the polynomial at the bottom-left of the equation.

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

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

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

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

We can now equate coefficients, and for our purposes, we will write out of the sum the term for $k = n$. In this term, we can substitute afterwards $A_0 = 1$ since the null-graph is unique.

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

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

$$A_n = a_n + \frac{1}{n}\left(\sum_{k=1}^{n-1}k\binom{n}{k}a_kA_{n-k}\right) \tag{3}$$

Wormald (1985) provides the numbers a_n for labeled connected chordal graphs, thus by using these and the formula 3, we obtain the numbers of all labeled chordal graphs; which equals to the number of Markov equivalence classes of DEC models, of table 1.

Next we count the Markov equivalence classes of TCI models. Andersson, Madigan, Perlman, and Triggs (1997) characterized these models as those DEC models $\mathbf{U}_{\mathcal{X}}(G)$ determined by a chordal graph G such that G does not contain the following induced undirected subgraph,

A_n	n
1	1
2	2
8	3
61	4
822	5
18154	6
617675	7
30888596	8
2192816760	9
215488096587	10
28791414081916	11
5165908492061926	12

Table 1: Number of Markov equivalence classes of DEC models

which is a path on four vertices. In the graph theory literature Golumbic (1978) characterized in such way these graphs and called them P_4-free chordal graphs. From this characterization Wormald and Castelo (2000) derived the following two propositions.

Proposition 5.1. *Wormald and Castelo (2000)*
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.

Proposition 5.2. *Wormald and Castelo (2000)*
Let $G = (V, E)$ be a connected P_4-free chordal graph, such that G is not complete. Let $D \subset V$ be the non-empty set of vertices that form the intersection of all the cliques of G. Let $G' = (V', E')$ be the graph that results of removing the set of vertices D from V in G, thus $V' = V \backslash D$ and $E' = E \cap \{V' \times V'\}$. G' will consist of a number of k components which are again connected P_4-free chordal graphs and $k > 1$, i.e. G' will be disconnected.

These two propositions allowed the authors in (Wormald & Castelo, 2000) to enumerate P_4-free chordal graphs, which in our context, correspond to Markov equivalence classes of TCI models. In particular, Wormald and Castelo (2000) provide the following recurrence for connected P_4-free chordal graphs.

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

In this recurrence, the term A_{n-k} refers to the number of all P_4-free chordal graphs. As shown in (Wormald & Castelo, 2000), 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 3. We may see in table 2 the numbers for connected P_4-free chordal graphs (a_n) and all P_4-free chordal graphs (A_n). The latter A_n corresponds to Markov equivalence classes of TCI models.

a_n	A_n	n
1	1	1
1	2	2
4	8	3
23	49	4
181	402	5
1812	4144	6
22037	51515	7
315569	750348	8
5201602	12537204	9
97009833	236424087	10
2019669961	4967735896	11
46432870222	115102258660	12

essential graph	UDG	DEC	DEC∩LCI (TCI)	n
1	1	1	1	1
2	2	2	2	2
11	8	8	8	3
185	64	61	49	4
8782	1024	822	402	5

Table 2: On the left: counts of A_n correspond to the number of Markov equivalence classes of TCI models. On the right: comparison of Markov equivalence classes of DAGs (essential graphs), UDGs, DECs and TCIs (DEC∩LCI)

Hence, 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, hardness 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 figure 9.

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

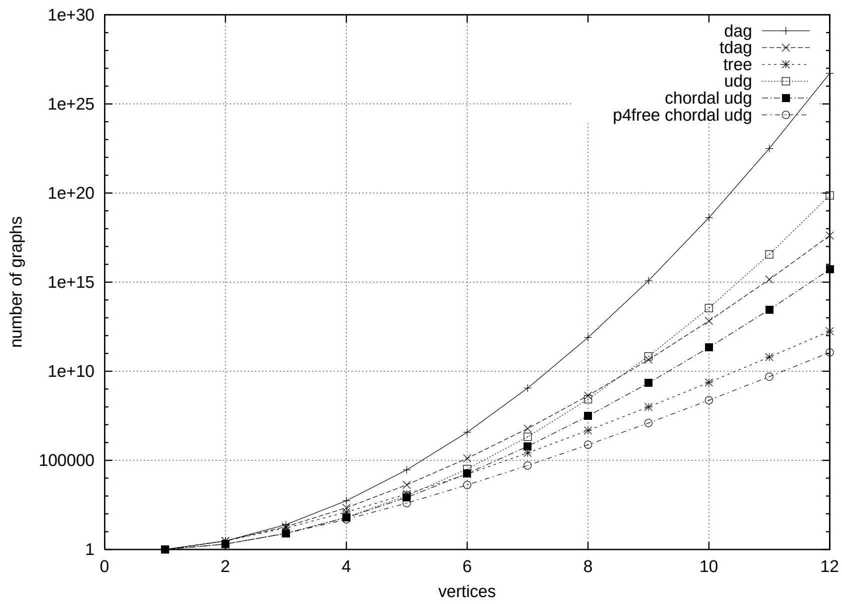

Figure 9: Cardinalities of the types of graphs that determine the different subclasses of chain graph Markov models

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

By proposition 5.1 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 5.2 the resulting graph of the removal of D, will consist of $k > 1$ disconnected components that are again P_4-free chordal graphs. Let's 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 formed upon the disconnected components that were created. It follows directly, that such an undirected structure cannot have undirected cycles, thus has the form of a tree. In figure 10b we may see the P_4-free chordal graph corresponding to the TCI model of figure 10a, which corresponds to one of the branches of the TCI model of figure 7. In figure 10c we may see a first step of the procedure we just described, and in 10d we may see the second and last step, from which we already obtain the canonical representation.

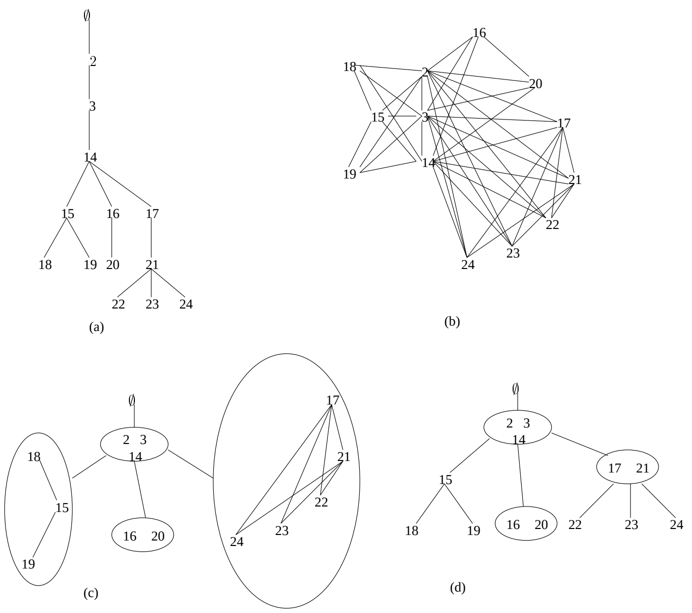

Figure 10: On (a) a TCI model. In (b) its corresponding P_4-free chordal graph. In (c) and (d) the two steps to obtain the canonical form of the equivalence class

In graph-theoretic terms, the canonical representation of a TCI model, as for instance the one in figure 10d, corresponds to *homeomorphically irreducible trees* (Harary & 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, and 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 hanging from the root $\emptyset$ one just applies all this process to each of the branches separately, and hang the top roots (the first intersection set removed) from the root $\emptyset$. For a given canonical element with $s_1, \ldots, s_k$ nodes that contain more than one vertex, the amount of trees on that equivalence class will amount to $|s_1|! \ldots |s_k|!$. 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∩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, allowed the computation of the number of different Markov models contained in DEC∩LCI. In this way, we could compare the cardinalities of all the different subclasses of chain graph –CG Markov models. Particular properties of P_4-free chordal graphs have lead 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 4.1 it follows that there are $(n+1)^{n-1}$ moral TDAGs on n vertices. These last two facts are interesting graph-theoretic results in its own right, and possibly could have some consequence, in our context, from a model selection perspective.

TCI models complete the scene of possible formalisms for conditional independence within the superclass of CG Markov models. Their expressiveness is smaller than any other subclass of CG Markov models, but in turn, they provide much more clarity of representation. Also the graphical structure is very close to the Markov properties represented as a graphical Markov model. This is because of the correspondence between such an intuitive concept as the meet path with a conditioning set, and a subtree with a separated set in the TCI model.

Andersson et al. (1995, p. 38) claimed that since every conditional independence statement $A \perp B | C$ is equivalent to a simple LCI model, then any DAG model is the intersection of all LCI models that contain it. We can see further that every conditional independence statement $A \perp B | C$ 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.

Acknowledgments

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 lead to the present paper.

We also would like to thank Nick Wormald and Ron Read for their valuable help in aspects related to graphical enumeration and Albrecht Schmidt for suggesting us the term *meet* for definition 4.1.

References

Andersson, S., Madigan, D., & Perlman, M. (1997). A characterization of markov equivalence classes for acyclic digraphs. *Annals of Statistics*, *25*, 505–541.

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. Planning and Inference*, *48*, 25–46.

Andersson, S., Madigan, D., Perlman, M., & Triggs, C. (1997). A graphical characterization of lattice conditional independence models. *Annals of Mathematics and Artificial Intelligence*, *21*, 27–50.

Andersson, S. & Perlman, M. (1993). Lattice models for conditional independence in a multivariate normal distribution. *Annals of Statistics*, *21*, 1318–1358.

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. *Scandinavian Journal of Statistics*, *17*, 333–353.

Golumbic, M. (1978). Trivially perfect graphs. *Discrete Mathematics*, *24*, 105–107.

Gratzer, G. (1978). *General Lattice Theory*. Birkhäuser Verlag, Basel.

Harary, F. & Palmer, E. (1973). *Graphical Enumeration*. Academic Press, New York.

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.

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, pp. 255–268. Elsevier Science Publishers B.V.

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.

Wormald, N. (1985). Counting labeled chordal graphs. *Graphs and Combinatorics*, *1*, 193–200.

Wormald, N. & Castelo, R. (2000). Enumeration of P_4-free chordal graphs. *In preparation.*