Rapporto n. 244

Graphical Model of type II: a smooth subclass

Federica Nicolussi- Roberto Colombi

Giugno 2014

Dipartimento di Statistica e Metodi Quantitativi
Università degli Studi di Milano Bicocca
Via Bicocca degli Arcimboldi, 8 - 20126 Milano - Italia
Tel +39/02/64483103/39 - Fax +39/2/64483105
Segreteria di redazione:
Andrea Bertolini

Graphical Model of type II: a smooth subclass

Federica Nicolussi Roberto Colombi

June 10, 2014

Abstract

The Probabilistic Graphical Models (GM) use graphs for representing the joint distribution of q variables. These models are useful for their ability to capture and represent the system of independences relationships between the variables involved, even when this is complex. This work concerns categorical variables and the possibility to represent symmetric and asymmetric dependences among categorical variables. At this aim we introduce the Chain Graphical Models proposed by Andersson, Madigan and Perlman (2001), also known as Chain Graphical Models of type II (GMs II). The GMs II allow for symmetric relationships typical of log-linear model and, at the same time, asymmetric dependences typical Graphical Models for Directed acyclic Graph. In general GMs II are not smooth, however this work provides a subclass of smooth GMs II by parameterizing the probability function through marginal log-linear models. Furthermore, we apply the proposed model to a data-set from the European Value Study (EVS), 2008.

1 Introduction

The increasingly use of graphical models is due to their ability to represent complex phenomena. The Probabilistic Graphical Models represent the joint probability function of q variables through a graph where each vertex of the graph corresponds to one variable and the arcs are indicators of dependences. Graphical Models for Chain graphs use both directed and undirected arcs, thus are able to represent simultaneously symmetrical or directional relationships. This topic is largely discussed in literature, see for instance Lauritzen, 1996.
With the purpose of representing simultaneous independence relationships between a collection of categorical variables we use the Graphical Models for Chain Graphs proposed by Andersson, Medigan and Perlman (2001) also known as Graphical Models of type II (**GM II**, see Drton, 2009). In GMs II the variables are partitioned into different sets. independences typical of log-linear models hold among the variables in the same set. On the other hand, asymmetrical independences typical of Graphical Model for DAG (Directed Acyclic Graph) hold among variables in different sets. It is important to observe that, the dependence between response and explicative variables is studied marginally with respect the other response variables. As it is shown in literature, a useful way to represent graphical models is through the marginal log-linear models introduced by Brgsma and Rudas (2002). This is not always possible for the GMs II because, as Drton (2009) showed, these models are not smooth in general. As the

parametric marginal models for categorical data have useful properties for the asymptotic theory of the ML estimators, we are interested to investigate which GMs of type II can be parameterized by a marginal models. In the past, the graphical models of type I, GMs I, proposed by Lauritzen and Wermuth (1989), and the *graphical models of type IV*, GMs IV, introduced by by Wermuth and Cox (2004) were preferred because all these models are smooth. It is common to think that the only GMs II, which are smooth, are equivalent to GMs I or GMs IV. In this work, not only we define a smooth subclass of GMs II, but we show that, in this class, there are models that can not be represented as GMs I or GMs IV.
The work is organized as follows. In Section 2 we introduce the marginal log-linear models with their important properties. Section 3 is dedicated to the GMs II, where, after an introduction on basic notation, new equivalent Markov properties will be proposed. Section 4 is reserved to describe the smooth subclass of GMs II which is representable with Marginal log-linear models. In particular, in subsection 4.1 the advantages of the parametrization proposed will be discussed. Finally, in section 5 we present an application on EVS data to illustrate the results obtained in this work.

2 Marginal Log-Linear Models

We consider q categorical variables $V = \{X_1, X_2, ..., X_q\}$, with levels $d_1, d_2, ..., d_q$ respectively. The *marginal log-linear models* are *models* where the interaction parameters are defined on marginal distributions. We will refer to these interactions, that are contrasts of logarithms of sum of probabilities, as *marginal parameters* (see Bergsma and Rudas, 2002). Thus, any parameter is distinguished by a pair of sets $(\mathcal{M}; \mathcal{L})$, where $\mathcal{L} \subseteq \mathcal{M} \subseteq V$. The so called *marginal set* $\mathcal{M}$ specifies the marginal distribution where the parameter is evaluated. The so called *interaction set* $\mathcal{L}$, s the set of the variables involved in the interaction. For each configuration of the variables in $\mathcal{L}$, the collection of the interactions defined in $\mathcal{M}$ with interaction set $\mathcal{L}$ are staked in the vector:

$$\boldsymbol{\eta}_{\mathcal{L}}^{\mathcal{M}} = \boldsymbol{C}_{\mathcal{L}}^{\mathcal{M}} \log \boldsymbol{M}_{\mathcal{L}}^{\mathcal{M}} \boldsymbol{\pi} \tag{1}$$

where, $\boldsymbol{C}_{\mathcal{L}}^{\mathcal{M}}$ and $\boldsymbol{M}_{\mathcal{L}}^{\mathcal{M}}$ are a *contrast* and a *marginalization* matrix while $\boldsymbol{\pi}$ is the vector of strictly positive joint probability of the q variables (Bartolucci *et al.*, 1977).
The whole set of parameters are collected in the vector $\boldsymbol{\eta}$ obtained stacking all the previous $\boldsymbol{\eta}_{\mathcal{L}}^{\mathcal{M}}$.

Definition 1. *A class $\mathcal{H} = \{\mathcal{M}_1, ..., \mathcal{M}_s\}$ of marginal sets, where $\mathcal{M}_s = V$ and $\mathcal{M}_i \nsubseteq \mathcal{M}_j$ if $j < i$, $\forall i, j = 1, ..., s$, is called hierarchical family of marginal set.*

Definition 2. *Given a hierarchical family of marginal sets, the vector of parameters $\boldsymbol{\eta}$ is complete if there is exactly one $\boldsymbol{\eta}_{\mathcal{L}}^{\mathcal{M}}$ for all $\mathcal{L} \subseteq V$; it is hierarchical if the marginal set $\mathcal{M}$ of $\boldsymbol{\eta}_{\mathcal{L}}^{\mathcal{M}}$ is the first in $\mathcal{H}$ which contains the variables in $\mathcal{L}$.*

A marginal model is characterized by a vector of hierarchical and complete parameters $\boldsymbol{\eta}$. Bergsma and Rudas showed that a complete and hierarchical $\boldsymbol{\eta}$ is a smooth parametrization of the probability distribution function of the variables in V (Theorem 2, Bergsma and Rudas, 2002).

Unfortunately, these marginal models are not able to represent all lists of conditional independences $\{A_i \perp B_i|C_i,\ i = 1, ..., k\}$. About this problem, Rudas et al. (2010) gave a sufficient condition so that a list of independences is representable by a marginal model. In theorem 1 we report this important result. Let us consider, for each independence $A_i \perp B_i|C_i$, the subclass D_i

$$D_i = \{\mathcal{L}_i : \mathcal{L}_i \in \mathcal{P}(A_i \cup B_i \cup C_i)\backslash(\mathcal{P}(A_i \cup C_i) \cup \mathcal{P}(B_i \cup C_i))\} \qquad (2)$$

of interaction sets containing at least one element of A_i, one element of B_i and possibly elements of C_i. Let $\mathcal{M}(\mathcal{L})$ be the first marginal set in the hierarchical class $\mathcal{H} = \{\mathcal{M}_1, ..., \mathcal{M}_m\}$ which contains the interaction $\mathcal{L}$, $\forall \mathcal{L} \subseteq V$.

Theorem 1. *Let us consider q variables, a hierarchical class of marginal sets $\mathcal{H}$ and the conditional independences system $\{A_i \perp B_i|C_i,\ i = 1, ..., k\}$. The class of probability distribution functions of the q variables that satisfies the previous conditional independences system is equivalent to the marginal model where*

$$\boldsymbol{\eta}_{\mathcal{L}}^{\mathcal{M}(\mathcal{L})} = 0, \quad \forall \mathcal{L} \in \cup_{i=1}^{k} D_i \quad \textit{and} \quad \mathcal{M}(\mathcal{L}) \in \mathcal{H} \qquad (3)$$

if the next condition is satisfied

$$C_i \subseteq \mathcal{M}(\mathcal{L}) \subseteq (A_i \cup B_i \cup C_i) \quad \forall \mathcal{L}_i \in D_i,\ i = 1, .., k \qquad (4)$$

In this work the marginal sets, the interaction sets and more generally the elements of classes of sets will be denoted by (...) instead of {...}.

Example 1. *Let consider a collection of four variables $V = \{V_1, V_2, V_3, V_4\}$ and the independences $\{V_1 \perp V_3|V_4;\ V_4 \perp V_1, V_2|V_3\}$. Let us take a hierarchical class of marginal set, for instance $\mathcal{H} = \{(V_1, V_3, V_4); (V_1, V_2, V_3, V_4)\}$. In order to verify the condition (4) we need all the classes D_i, $i = 1, 2$. Thus, from the independence $V_1 \perp V_3|V_4$, we have $D_1 = \{(V_1, V_3); (V_1, V_3, V_4)\}$, and from independence $V_4 \perp V_1, V_2|V_3$ we have $D_2 = \{(V_1, V_4); (V_2, V_4); (V_1, V_2, V_4); (V_1, V_3, V_4); (V_2, V_3, V_4); (V_1, V_2, V_3, V_4)\}$, see formula* (2)*. In the first case, the condition (4) becomes $V_4 \subseteq \mathcal{M}(\mathcal{L}) \subseteq V_1, V_3, V_4$; that always holds since $\mathcal{M}(\mathcal{L}) = V_1, V_3, V_4$ for all $\mathcal{L} \in D_1$. Thus the vectors of parameters to constrain to zero are $\boldsymbol{\eta}_{13}^{134}$, $\boldsymbol{\eta}_{134}^{134}$. According to the second independence the condition (4) becomes $V_3 \subseteq \mathcal{M}(\mathcal{L}) \subseteq V_1, V_2, V_3, V_4$. Even in this case the condition holds for all $\mathcal{L} \in D_2$ thus the second independence is represented by annulling the vectors $\boldsymbol{\eta}_{14}^{134}$,$\boldsymbol{\eta}_{134}^{134}$, $\boldsymbol{\eta}_{24}^{1234}$, $\boldsymbol{\eta}_{234}^{1234}$,$\boldsymbol{\eta}_{124}^{1234}$, $\boldsymbol{\eta}_{1234}^{1234}$.*

3 Graphical Models of type II

In Section 3.1 we will introduce the main notation and definitions on graph theory useful to understand the graphical models of type II described in Section 3.2; for a more general treatment see Lauritzen 1966. In Section 3.3 a new equivalent formulation of the Markov properties of the previous GMs is introduced.

3.1 Graph Theory

The graphs are mathematical objects defined by two sets $G = \{V, E\}$, where $V = \{V_1, .., V_q\}$ is the set of vertices and $E \subseteq V \times V$ is the set of edges or

arcs that can be both directed or undirected. Two vertices are *adjacent* if they are joined by an undirected edge. Given a subset A of V, the set of all vertices both not included in A and adjacent to at least one of the vertices in A, is called *set of neighbours* of A, $nb(A)$. The *neighbourhood* of A is defined by $Nb(A) = nb(A) \cup A$.
On the other hand, when there is a directed arc from a vertex V_i to a vertex V_j, the first vertex is called *parent* of V_j and V_j is called *child*. Given a subset A of V, the set $pa(A)$ of *parent* of A is the collection of all vertices with at least one child in A. We define the set $ch(A)$ of children of A as the set of all vertices with at least one parent included in A.
A *directed cycle* is an ordered sequence of vertices, all joined by direction preserving directed arcs (directed-path), starting and ending in the same vertex. A *semi-directed cycle* is an ordered sequence of vertices, joined by both direction preserving directed and undirected arcs (semi-directed path), which starts and ends in the same vertex.
A *Chain Graph* (CG) is a graph that can include both directed and undirected arcs without any directed or semi-directed cycle. A CG is decomposable into *Chain Components*, denoted by $T_1, ..., T_s$. Within these chain components there are only undirected arcs and between two components there are only directed arcs in the same direction.
Given a CG, the associated *Directed Graph* is a directed acyclic graph where the components $T_1, T_2, ...$ act for the vertices , and there is a direct arc linking T_j to T_h if at least an element of T_h is children of an element of T_j.The definition of parents and children also apply to the associated Directed Graph. Thus, for instance for a component T_h, the parent $pa_D(T_h)$ is the set of the chain components having as children T_h, $h = 1, , s$. The chain components are ordered in such way that if $j < h$ there is not any directed arc from T_h to T_j.

Example 2. *An example of chain graphs is represented in figure* 1 *where we may recognize three components:* $T_1 = 1$*,* $T_2 = 3, 4$ *and* $T_3 = 4, 5, 6$*. Within the component* T_3*, the vertices* 4 *and* 5 *are adjacent and the neighbours of* 5 *are* $nb(5) = 4, 6$*. The vertex* 2 *is child of* 1 *and parent of* 4*. The set of parents of* $4, 6$ *is* $pa(4, 6) = 2, 3$ *and the set of parent of the component* T_3 *is* $pa_D(T_3) = T_2 = 2, 3$*.*

Any component T_h can be partitioned in three subsets, respectively, the set CH_h of children, the set NC_h of vertices that are not children but are adjacent to them, and the set NA_h of vertices that are neither children nor adjacent to children. In the first set there are the vertices $V_j \in T_h$ such that $pa_G(V_j) \neq \emptyset$; in NC_h there are the $V_j \in T_h$ such that $(pa_G(V_j) = \emptyset)$ & $(nb(V_j) \cap CH_h \neq \emptyset)$; finally, in the set NA_h there are the remaining vertices of T_h.

Example 3. *(Continuation of Example* 2*) In the component* T_1*, the sets* CH_1 *and* NC_1 *are empty and the only vertex* 1 *defines the set* NA_1*. In the component* T_2*, since the vertices* 2 *and* 3 *are both children, the sets* NC_2 *and* NA_2 *are empty and* $CH_2 = \{2, 3\}$*. Finally, in the component* T_3 *we have* $CH_3 = \{4, 6\}$*,* $NC_3 = \{5\}$ *and* $NA_3 = \emptyset$*.*

A subset A of V is *complete* if every pair of vertices of A is adjacent. The class of the complete sets of the component T_h is denoted with $\mathcal{C}_h$. A complete subset A of V is a *clique* if it is maximal, that is if there are not complete sets

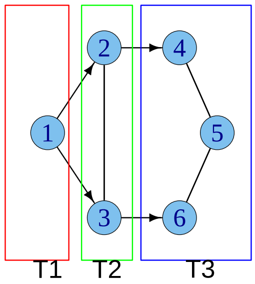

Figure 1: Chain graphs with set of vertex $V = \{1, 2, 3, 4, 5, 6\}$ and set of edges $E = \{(1, 2); (1, 3); (2, 3); (3, 2); (2, 4); (3, 6); (4, 5); (5, 4); (5, 6); (6, 5)\}$

containing it. We denote the *family of cliques* of a component T_h by $\mathcal{C}l_h$.A subset A of V is *connected* if every pair of vertices of A is linked by a path in A. Finally, we define $\mathcal{K}_h$ as the class of all *non-connected* sets of T_h.

3.2 Graphical Models of Type II

Graphical models GMs take advantage of graphs to represent multidimensional dependence structures among variables. There are four types of GMs for chain graph, listed in Drton (2009), each of which represents different dependences systems. The Drton's GMs use CG where the vertices act for the variables and the possible edges act for dependence relationships. When two vertices are connected by an undirected edge, we can guess that the two linked variables are "symmetrically" dependent. On the other hand, when there is a directed arc, we presume a dependence relationship among the linked variables. We study Graphical models of type II (GMs II), introduced by Anderson, Madigan and Perlman (2001), as generalization of both GMs for undirected graphs (UG) and GMs for directed acyclic graphs (DAG) (for details see Lauritzen,1996). These models are useful for different reasons. First, the grouping of variables in components allows to split the variables in "*purely explicative*" variables, "*purely response*" variables and "*intervening*" variables. Secondly, in the GMs II, the relationship among a variable and its explicative variables is considered marginally with respect to the variables in the same component. Finally, the GMs II model the association between the variables within the same component using a log-linear approach. The rules to read a list of conditional independences from a graph are called *Markov properties* and, for the GM II are the following three:

$$
\begin{array}{lll}
(C1) & T_h \perp \cup_{i<h} T_i \setminus pa_D(T_h) | pa_D(T_h) & \\
(C2a) & A \perp T_h \setminus Nb(A) | pa_D(T_h) \cup nb(A) & \forall A \subset T_h \\
(C3b) & A \perp pa_D(T_h) \setminus pa_G(A) | pa_G(A) & \forall A \subseteq T_h \\
 & \quad \forall h = 1, ..., s. &
\end{array}
\tag{5}
$$

The first Markov property, $(C1)$, describes the independences between the chain components; the second, $(C2a)$, reads the conditional independences within the components, and the third, $(C3b)$ interprets the lack of directed arcs between variables in different components.

Example 4. *Applying the Markov properties in formula* (5) *to the graph in figure* 1*, we get respectively, from the (C1)* $4, 5, 6 \perp 1|2, 3$*, from the (C2a)* $4 \perp 6|2, 3, 5$ *and from the (C3b)* $5 \perp 2, 3$*;* $4 \perp 3|2$ *and* $6 \perp 2|3$.

Unfortunately, for categorical variables these models are not always smooth, see Drton, 2009. This means that the probability function of the q variables under the constrains given by the removed arcs does not always belong to a curved exponential family. As the parametric marginal models introduced in Section 2 are always smooth, in Section 4 we will propose a subclass of smooth GMs II, that can be parameterized as marginal models. To do this, we need *non-redundant* list of independences. Since from the *(C2a)* and *(C3b)* we obtain redundant lists, in the next subsection we propose alternative Markov properties to the *(C2a)* and *(C3b)* which are not-redundant. These properties will be used to proof the main theorems.

3.3 Alternative Markov properties for GMs II

As mentioned above, conditions *(C2a)* and *(C3b)* are not a non-redundant list of independences. Below we propose two alternative conditions.

An alternative condition for *(C2a)*. We consider the family $\mathcal{C}l_h$ of the r cliques of the $h-$th component, and we split the elements C_i of Cl_h in two sets $C_i = B_{1i} \cup B_{2i}$, for $i = 1, ...r$, in such way that: $CH_h \cap B_{1i} \neq \emptyset$ and $V_j \in B_{1i}$ if and only if $Nb(V_j) = C_i$, while $B_{2i} = C_i \backslash B_{1i}$. Below is reported an example which shows how to decompose the elements C_i.

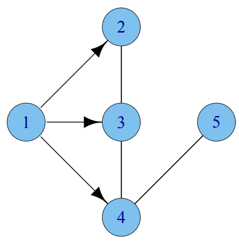

Figure 2: Chain graphs with set of vertex $V = \{1, 2, 3, 4, 5\}$ and set of edges $E = \{(1, 2); (1, 3); (1, 4); (2, 4); (3, 2); (3, 4); (4, 3); (4, 5); (5, 4)\}$

Example 5. *We consider the graph in figure* 2*. In the component* T_2 *we have the family of the cliques* $\mathcal{C}l_2 = \{(2, 3); (3, 4); (4, 5)\}$*. Now, we split the elements of* Cl_2 *according to the previous rules:*

C_i	B_{1i}	B_{2i}
$2, 3$	2	3
$3, 4$	$\emptyset$	$3, 4$
$4, 5$	$\emptyset$	$4, 5$

Definition 3. *The condition (C2*a) is described by the following list of independences:*

$$B_{1i} \perp T_h \backslash C_i | pa_D(T_h) \cup B_{2i} \tag{6}$$

$\forall i = 1, ..., r$ *and* $\forall h = 1, ..., s$.

$$\left(V_j \cup \mathbf{B}_{1,V_j}\right) \perp T_h \backslash Nb\left(V_j \cup \mathbf{B}_{1,V_j}\right) | pa_D(T_h) \cup nb\left(V_j \cup \mathbf{B}_{1,V_j}\right), \tag{7}$$

$\forall V_j \in \cup_{i=1}^{r} B_{2i}$ *and* $\forall h = 1, ..., s$ *where* $\mathbf{B}_{1,V_j} = nb(V_j) \cap (\cup_{i=1}^{r} B_{1i})$.

While the conditions 6 and 7 are clear consequence of the condition (C2a) in the formula 5 , the following theorem shows that these conditions are equivalent to the condition (C2a).

Theorem 2. *The (C2*a) yields a non-redundant list of independences that is equivalent to the list of independences given by (C2a).*

The proof of this theorem appears in the Appendix B.

Example 6. *(Continuation of Example* 5*) Regarding the graph in figure* (2)*, Applying the properties in formula* (6)*, we get* $2 \perp 4, 5|1, 3$*. Applying the formula* (7) *we get* $2, 3 \perp 5|1, 4$*;* $4 \perp 2|1, 3, 5$ *and* $5 \perp 2, 3|1, 4$.

An alternative condition for *(C3b)*. The condition *(C3b)* expresses the relationship between a vertex and its parents. On the other hand, the alternative condition, *(C3*b)*, focuses on the relationships between a vertex and its children. At this purpose, we define the class PA_h of sets composed by elements having the same children in T_h. Note that, the elements of PA_h are a partition of $pa_D(T_h)$.

Definition 4. *The class* PA_h *of elements with same children in* T_h*, is*

$$PA_h = \{\mathcal{A} : ch(V_i) \cap T_h = ch(V_j) \cap T_h, \forall V_i, V_j \in \mathcal{A}\} \tag{8}$$

We consider the elements of this class partially ordered according to the following rule: $\forall \mathcal{A}, \mathcal{B} \in PA_h$ *if* $|ch(\mathcal{B})| < |ch(\mathcal{A})|$ *then* $\mathcal{A} \prec \mathcal{B}$.

Definition 5. *The new condition (C3*b) is defined by the following list of conditional independences.*

$$\mathcal{A} \perp [T_h \backslash ch(\mathcal{A})] | (pa_D(T_h) \backslash \mathcal{A}), \qquad \forall \mathcal{A} \in PA_h \qquad \forall h = 1, ...s. \tag{9}$$

Theorem 3. *The (C3*b) yields a non-redundant list of independences that is equivalent to the list of independences given by (C3b).*

The proof of this theorem appears in the Appendix B.

Example 7. *Referring to the graph in figure* 1*, the* PA_1 *class is empty, since there are not parents of* T_1 *and the* PA_2 *is composed by the only vertex* 1*:* $PA_2 = \{1\}$*. Finally, for the third component, we have* $pa_D(T_3) = \{2, 3\}$*. Since* $ch(2) = 4 \neq ch(3) = 5$*, the class of parents with common children is* $PA_3 = \{(2); (3)\}$*. Note that the (C3b) and (C3*b) lead independences only between components* T_2 *and* T_3*. So, from the (C3b) we have* $4 \perp 3|2$*,* $5 \perp 2, 3$*,* $6 \perp 2|3$*,* $4, 5 \perp 3|2$ *and* $6, 5 \perp 2|3$*. Instead, from the (C3*b) we have* $2 \perp 5, 6|3$ *and* $3 \perp 4, 5$*. Through the properties of conditional independences we can prof the equivalence between the two lists.*

4 A smooth subclass of GMs II

In this section we will introduce a smooth subclass of GMs II. For this, we use the known smoothness property of marginal models. With the help of the theorem 1, we study which graphs yield lists of independences that can be represented with marginal models. At this purpose we follow the 4 steps:

1) define a class of hierarchical marginal sets $\mathcal{H}$;

2) define the list of hierarchical and complete marginal parameters associated to the previous marginal sets;

3) define the list of parameter to set equal to zero according to the formula 3 of Theorem 1;

4) verify when the condition 4 of Theorem 1 holds.

First, we define a hierarchical class of marginal sets. For every three sets A_i, B_i and C_i associated to an independence $A_i \perp B_i|C_i$ derived from *(C1)*, *(C2*a)* and *(C3*b)*, the hierarchical class of marginal sets must contain at least the elements $(A_i \cup B_i \cup C_i)$. Thus, according to *(C1)* for each component we introduce the marginal sets:

$$\mathcal{M}_h^1 = \cup_{j \leq h} T_j. \tag{10}$$

According to *(C2*a)*, for each component we define the marginal sets:

$$\mathcal{M}_h^{2^*a} = T_h \cup pa_D(T_h). \tag{11}$$

Finally, according to the third condition *(C3*b)*, for each component, we introduce the marginal sets:

$$\mathcal{M}_{h,A}^{3^*b} = pa_D(T_h) \cup (NC_h \cup NA_h) \cup A \qquad \forall A \in \{\mathcal{P}(CH_h) \backslash J_h\}, \tag{12}$$

where $\mathcal{P}(CH_h)$ denotes the power set of CH_h and J_h is the class of all subsets of T_h having parents equal to $pa_D(T_h)$, $J_h = \{A : pa(A) = pa_D(T_h), A \subseteq T_h\}$. Notice that, by definition, the following relationship always holds:

$$\mathcal{M}_{h,A}^{3^*b} \subseteq \mathcal{M}_h^{2^*a} \subseteq \mathcal{M}_h^1 \qquad \forall A \in \{\mathcal{P}(CH_h) \backslash J_h\}, \quad \forall h = 1, .., s. \tag{13}$$

Thus, the hierarchical class of marginal sets, for each h, first contains all sets $\mathcal{M}_{h,A}^{3^*b}$, sorted according to the hierarchical principle (see definition 1), then it contains the set $\mathcal{M}_h^{2^*a}$ and finally the set $\mathcal{M}_h^1$. Thus, for each component h, we have the following class:

$$\mathcal{H}_{II}^h = \{\{\mathcal{M}_{h,A}^{3^*b}, \forall A \in \mathcal{P}(CH_h)\backslash J_h\}, \mathcal{M}_h^{2^*a}, \mathcal{M}_h^1\} \tag{14}$$

At last, the family of all marginal sets of the chain graph is given by the following ordered collection:

$$\mathcal{H}_{II} = \{\mathcal{H}^h, h = 1, .., s\} \tag{15}$$

Note that, it may occur that some of the previous sets match. For example, for a given h and A, it could happen that $\mathcal{M}_h^{2^*a}$ is equal to $\mathcal{M}_{h,A}^{3^*b}$, or that $\mathcal{M}_h^{2^*a}$ is equal to $\mathcal{M}_h^1$.

Example 8. *We consider the graph in figure* 1*. From the (C1) we have the marginal sets* $\mathcal{M}_1^1 = (1)$*,* $\mathcal{M}_2^1 = (1, 2, 3)$ *and* $\mathcal{M}_3^1 = (1, 2, 3, 4, 5, 6)$*. From the (C2*a), we get the marginal sets* $\mathcal{M}_1^{2^*a} = (1)$*,* $\mathcal{M}_2^{2^*a} = (1, 2, 3)$ *and* $\mathcal{M}_3^{2^*a} = (2, 3, 4, 5, 6)$*. Finally, from the (C3*b) we get* $\mathcal{M}_{1,\emptyset}^{3^*b} = (1)$*,* $\mathcal{M}_{3,\emptyset}^{3^*b} = (2, 3, 5)$*,* $\mathcal{M}_{3,4}^{3^*b} = (2, 3, 4, 5)$ *and* $\mathcal{M}_{3,6}^{3^*b} = (2, 3, 5, 6)$*. The hierarchical class is* $\mathcal{H}_{\mathcal{II}} = \{(1); (1, 2, 3); (2, 3, 5); (2, 3, 4, 5); (2, 3, 5, 6); (2, 3, 4, 5, 6); (1, 2, 3, 4, 5, 6)\}$

Once the class of marginal sets is defined, we determine the hierarchical and complete parameters, $\boldsymbol{\eta}_{\mathcal{L}}^{\mathcal{M}(\mathcal{L})}$, $\forall \mathcal{L} \in \mathcal{P}(V)$, where $\mathcal{M}(\mathcal{L})$ is the first marginal set in $\mathcal{H}_{II}$ containing the set $\mathcal{L}$. At this point, we must select the parameters to constrain to zero. According Theorem 1, these parameters are $\boldsymbol{\eta}_{\mathcal{L}}^{\mathcal{M}(\mathcal{L})}$ where $\mathcal{L} \in D_i^1$ according to the *(C1)*, $\mathcal{L} \in D_i^{2^*a}$, according to the *(C2*a)* and $\mathcal{L} \in D_i^{3^*b}$ according to the *(C3*b)*. The classes D_i^1, $D_i^{2^*a}$ and $D_i^{3^*b}$ strictly depend from the alternative Markov properties defined in subsection 3.3. The definition of these classes are in the Appendix A and here we report only an example.

Example 9. *(Continuation Example 4 and* 8*) Applying the formula (2) to the the independences listed in the example* 4*, the resultant classes of interactions concerning null parameters are hereafter declared.* $D_1^1 = \{\emptyset\}$*,* $D_2^1 = \{\emptyset\}$ *and* $D_3^1 = \{(1, 4); (1, 5); (1, 6); (1, 4, 5); (1, 4, 6); (1, 5, 6); (1, 4, 5, 6)\}$*.* $D_1^{2^*a} = \{\emptyset\}$*,* $D_2^{2^*a} = \{\emptyset\}$*,* $D_3^{2^*a} = \{(4, 6); (4, 5, 6); (2, 4, 6); (2, 4, 5, 6); (3, 4, 6); (3, 4, 5, 6); (2, 3, 4, 6); (2, 3, 4, 5, 6)\}$*. Finally,* $D_1^{3^*b} = \{\emptyset\}$*,* $D_2^{3^*b} = \{\emptyset\}$ *and* $D_3^{3^*b} = \{(2, 5); (2, 6); (2, 5, 6); (3, 4); (3, 5); (3, 4, 5); (2, 3, 5); (2, 3, 6); (2, 3, 5, 6); (2, 3, 4); (2, 3, 5); (2, 3, 4, 5)\}$*.*

By implementing Theorem 1, we obtain the class of GMs II which is parametrizable with marginal models. Theorem 4 shows when condition 4 of Theorem 1 is satisfied, given the family $\mathcal{H}_{II}$ of marginal sets and the sets D_h^1, $D_h^{2^*a}$ and $D_h^{3^*b}$ previously defined.

Theorem 4. *A graphical model of type II is a marginal model with* $\{\boldsymbol{\eta}_{\mathcal{L}}^{\mathcal{M}} : \mathcal{L} \in \mathcal{P}(V)\backslash \cup_{h=1}^{s} \left(D_h^1 \cup D_h^{2^*a} \cup D_h^{3^*b}\right), \mathcal{M} \in \mathcal{H}_{II}\}$*, if, for all* $V_j \in CH_h$ *such that* $Nb(V_j) \notin \mathcal{C}_h$ *,* $\{K : K \in \mathcal{K}_h; K \cap nb(V_j) \neq \emptyset\} \subseteq J_h$*.*

This theorem shows that the smoothness problem concerns the vertices in the children set CH_h with non complete neighbourhood. For these vertices the smoothness property of the model is preserved if all non connected sets containing at least one neighbour of the problematic vertices, have parent set equal to the parent set of the component, $pa_D(T_h)$.The proof of this theorem is in the Appendix C.

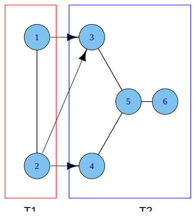

Figure 3:

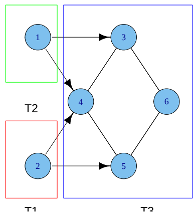

Figure 4:

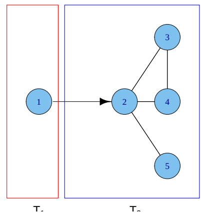

Figure 5:

Figure 6: *Three chain graphs corresponding respectively to two smooth chain graphical models and to a non-smooth chain graphical model*

Example 10. *The graph in figure* 3 *represents the following system of independences* $3 \perp 4, 6|1, 2, 5$*;* $4 \perp 3, 6|1, 2, 5$*;* $1 \perp 4, 5, 6|2$*;* $2 \perp 5, 6|1$*. The marginal class referring to this graph is* $\mathcal{H}_{II} = \{(1, 2); (1, 2, 5, 6); (1, 2, 4, 5, 6); (1, 2, 3, 4, 5, 6)\}$*. Since any vertex in* $CH_2 = 3; 4$ *has complete neighbourhood:* $Nb(3) \in \mathcal{C}_2$ *and* $Nb(4) \in \mathcal{C}_2$*, the theorem* 4 *holds.*

Example 11. *The structure of conditional independence represented in the graph in figure* 4 *can be explained by the statements* $1 \perp 2$*;* $3 \perp 5|1, 2, 4, 6$*;* $4 \perp 6|1, 2, 3, 5$*. The class of marginal sets is* $\mathcal{H}_{II} = \{(1, 2); (1, 2, 3, 6); (1, 2, 5, 6); (1, 2, 3, 4, 5, 6)\}$*. Note that in* CH_2 *there are the two vertices* 3 *and* 5*. Thus, it is necessary to take into account the family of not connected sets* $\mathcal{K}_2 = \{(4, 6); (3, 5)\}$*. Since* $(4, 6) \in \mathcal{J}_2$ *and* $(3, 5) \in \mathcal{J}_2$*the theorem 4 holds.*

Example 12. *The graph in figure* 5 *represents the following system of independences* $3, 4 \perp 5|1, 2$ *and* $1 \perp 3, 4, 5$*. The class of marginal sets referring to this graph is* $\mathcal{H}_{II} = \{(1); (1, 3, 4, 5);(1, 2, 3, 4, 5)\}$*. Since* $CH_2 = (2)$ *has a not complete neighbourhood (3,4,5), and since the non connected sets* $(3, 5)$*,* $(4, 5)$ *and* $(3, 4, 5)$ *of* $\mathcal{K}_2$ *have not* $pa_D(T_2)$ *as parent set, the conditions of the theorem* 4 *are not satisfied.*

4.1 Important Results

In the previous section we gave a subclass of GMs II parameterizable with Marginal Models. It is legitimate to ask whether other marginal parameterizations, characterized by different marginal sets, could describe different subclass of GMs II.
Also, it is worthwhile to consider whether the subclass defined in theorem 4 detects only the particular cases of GMs II equivalent to GMs I or GMs IV. In this section we answer to these questions.
As response to the first query, Theorem 5 shows that hierarchical classes, which differ from $\mathcal{H}_{II}$, in formula 15, lead to smaller subclass of GMs II parameterizable through marginal models. In other words, it is proved that no other GMs II could be expressed through marginal models.

Theorem 5. *All GMs II that can be expressed by a marginal model, can be characterized by the parametrization* $\{\boldsymbol{\eta}_{\mathcal{L}}^{\mathcal{M}} : \mathcal{L} \in \mathcal{P}(V) \setminus \cup_{h=1}^{s} (D_h^1 \cup D_h^{2^*a} \cup D_h^{3*b}), \mathcal{M} \in \mathcal{H}_{II}\}$*.*

From this theorem it derives that if a chain graph does not satisfy the condition of the theorem 4, then it can not be represented with a marginal model. As before, the proof of this theorem is postponed to the Appendix C.
Furthermore, as mentioned in the introduction, it is common to think that the only smooth GMs II are equivalent to GMs I or GMs IV. In this section, through the example 13 we will show that there are smooth GMs II that are neither GMs I or GMs IV.

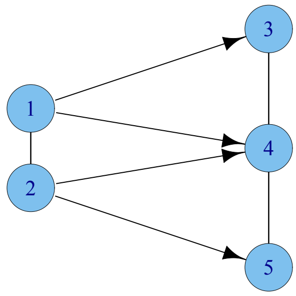

Figure 7: Chain graphs with set of vertices $V = \{1, 2, 3, 4, 5\}$ and set of edges $E = \{(1, 2); (2, 1); (1, 3); (1, 4); (2, 4); (2, 5); (3, 4); (4, 3); (4, 5); (5, 4)\}$

Example 13. *The GM II associated with the graphs in figure* 7 *represents the following system of independences:* $\{1 \perp 5|2;\ 2 \perp 4|1;\ 3 \perp 5|1, 2, 4\}$*. From the theorem* 4*, since the class of not-connected sets* $\mathcal{K}_2 = \{(3, 5)\}$ *is contained in* $J_h = \{(4), (3, 4), (3, 5), (4, 5), (3, 4, 5)\}$*, the GM II associated is a marginal model and therefore it is smooth. According to theorem 6 of Andersson, Medigan and Perlman, necessary and sufficient condition for the equivalence of GMs I and GMs II is the non presence of particular structure in the graph called bi-flags. The graph in figure* 7 *is an example of bi-flag, thus, according the aforementioned theorem does not exist any GM I which represents the same structure of relationships. Regarding the equivalence between GMs II and GMs IV, it is sufficient to note that the independence* $3 \perp 5|1, 2, 4$ *can be represented by no one GMs IV.*

5 GMs II applied to a real dataset

We chose a data-set from the European Values Study -EVS- (2008), in order to show the ability of the GMs II to represent a system of conditional independences of categorical variables. The EVS is a research project on human values in Europe. In particular, the research involves how Europeans think about *family*, *work*, *religion*, *politics* and *society*. We use the GMs II to highlight the dependence system of some variables classified as *gender* variable, *personal* variables and *opinion* variables. To this aim, we select six variables so described:

G: Gender ("Female", "Male");

E: Employed ("Yes", "No");

C: Children ("Yes", "No");

T: Trust in the people ("Yes", "No");

O: Opinion on Society ("High", "Mean", "Low");

W: Personal Perceived Well-being level ("High", "Low");

We divided the variables in three groups, each of one corresponding to one component in the chain graph. In the first group we placed only the Gender variable (G). In the second group there are variables about the status of the respondents (E= *employed*, C= *Children*). Finally, the last group regards the variables that consider the opinion of the respondents about some topics (W= *Personal Perceived Well-being level* , O= *Opinion about the society*, T=*Trust in people*). We represented each group of variables with a component in the chain graph. We fitted GMs II for different European Countries. The two most interesting case, concerning the Northern Islands (*Ireland*, the *United Kingdom* and the *Island*) and Italy are reported below. In both cases we fitted the saturated marginal model (unconstrained model) corresponding to the complete chain graphical model. We proceeded testing the GMs II obtained by removing the arcs one by one. Any model was tested using the Likelihood Ratio test which compares the saturated model with the chosen model. In both cases, we chose the simplest model, with fewer number of arcs, still able to representing the data. In the case for the Northern Islands we chose the graph in figure 8. The marginal model corresponding to this graph has Likelihood ratio test statistic $Gsq = 53.19302$ with 51 degree of freedom and the model displayed in figure 8 can be retained with a p-value of 0.38976. The second interesting case is the Italian case, well described by the graph in figure 9. In this case the statics test Gsq is 68.84138, with 55 degree of freedom and a p-value equal to 0.0935.

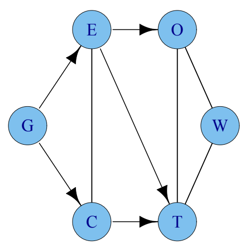

Figure 8: *Chain Graph representing the variables in the Northern Islands Case*

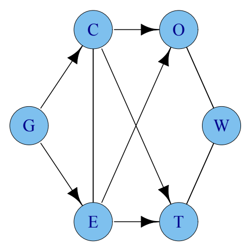

Figure 9: *Chain Graph representing the variables in the Italian Case*

Northern Islands	Italy
$G \perp T, W, O \mid C, E$	$G \perp T, W, O \mid C, E$
$W \perp C, E$	$W \perp C, E$
$O \perp C \mid E$	$O \perp T \mid W, C, E$

Table 1: List of non redundant conditional independences obtained from the condition *(C1)*, *(C2*a)* and *(C3*b)* for the two cases.

We can analyse the chosen models. Table 1 reports the non redundant lists of conditional independences obtained from the condition *(C1)*, *(C2*a)* and *(C3*b)* for the two models, while the table 5 reports the non null parameters describing the relationships among the variables.

In both models we observe the independence of the gender variable (G) from the opinion variables (T,O,W) given the personal variables (C,E).

Furthermore, in the Northern Islands Case (figure 8), we inspect that the variable referring to the children (C) is independent of both opinion on the society (O) and the personal perceived well-being level (W) given the employment (E). The last independence about this graph concerns the employment (E) that is conditional independent of personal perceived well-being level (W) given the children (C). On the contrary, the opinion in society (O) is not independent of the variable work (E) and the trust in people (T) is not independent of both the variable work (E) the children variable (C) . Further, looking at the non-null parameters in table 5 it is possible to say that the relationships between the trust in people (T) and both children (C) is stronger than the relationship between the trust in people (T) and employment (E).

In the Italian case we have the marginal independence between W and C,E and the variable O is independent of T given W, C and E.

Northern Islands			Italy		
$\mathcal{M}$	$\mathcal{L}$	$\boldsymbol{\eta}_{\mathcal{L}}^{\mathcal{M}}$	$\mathcal{M}$	$\mathcal{L}$	$\boldsymbol{\eta}_{\mathcal{L}}^{\mathcal{M}}$
G	G	-0.0449	G	G	0.0152
CEG	C	0.6361	CEG	C	0.6434
CEG	E	0.6491	CEG	E	0.7051
CEG	CE	0.0207	CEG	CE	-0.0410
CEG	CG	0.9245	CEG	CG	0.4120
CEG	EG	-0.1460	CEG	EG	-0.6676
CEG	CEG	-0.6237	CEG	CEG	-0.3566
WCE	W	1.5997	WCE	W	0.6695
$WOCE$	O	$[1.6709; -1.5237]$	$WOTCE$	O	$[1.4606; -3.1637]$
$WOCE$	WO	$[-0.0045; -0.1318]$	$WOTCE$	T	-0.6910
$WOCE$	OE	$[-0.7183; -0.2680]$	$WOTCE$	WO	$[0.6394; -0.4189]$
$WOCE$	WOC	$[0.3076; -0.0108]$	$WOTCE$	WT	-0.0988
$WOCE$	WOE	$[0.9925; -0.4202]$	$WOTCE$	OC	$[1.1911; 1.3128]$
$WOCE$	$WOCE$	$[-0.3958; 0.1132]$	$WOTCE$	TC	-0.6064
$WOTCE$	T	-1.5630	$WOTCE$	OE	$[-0.6684; 1.6665]$
$WOTCE$	WT	0.5808	$WOTCE$	TE	-1.1285
$WOTCE$	OT	$[0.3687; 0.5243]$	$WOTCE$	WOC	$[-0.3753; -0.1285]$
$WOTCE$	TC	-21.9010	$WOTCE$	WTC	0.0500
$WOTCE$	TE	-19.3440	$WOTCE$	WOE	$[0.4847; -0.6825]$
$WOTCE$	WOT	0.2704	$WOTCE$	WTE	1.3874
$WOTCE$	WTC	21.8659	$WOTCE$	OTC	$[0.7660; -0.9532]$
$WOTCE$	WTC	$21.8395; -0.3217$	$WOTCE$	OCE	$[-0.1567; -2.4855]$
$WOTCE$	WTE	20.0330	$WOTCE$	TCE	0.8888
$WOTCE$	OTE	$[1.5645; -1.4979]$	$WOTCE$	$WOTC$	$[0.9136; -0.4878]$
$WOTCE$	TCE	42.4563	$WOTCE$	$WOTE$	$[-0.0037; 1.6923]$
$WOTCE$	$WOTC$	$[-21.9093; 1.1838]$	$WOTCE$	$WOTCE$	-0.5678
$WOTCE$	$WOTC$	$[-20.3008; 20.3660]$			
$WOTCE$	$WTCE$	-42.6677			
$WOTCE$	$OTCE$	$[-42.5531; 21.0933]$			
$WOTCE$	$WOTCE$	$[43.0787; -21.6751]$			

Table 2: Non null parameters concerning the two marginal models

Appendix A

Family of interaction sets concerning null parameters

In this appendix we define for all the Markov properties *(C1)*, *(C2*a)* and *(C3*b)*, respectively the classes of interaction D_h^1, D_h^{2*a} and $D_h^{3^*b}$, $h = 1, ..., s$. Applying the formula 2 to the previous Markov properties we get the following classes:

$$D_h^1 = \left\{\mathcal{L} : \mathcal{L} \in \mathcal{P}\left(\cup_{j=1}^h T_j\right) \setminus \left(\mathcal{P}(T_h \cup pa_D(T_h)) \cup \mathcal{P}(\cup_{j=1}^{h-1}(T_j))\right)\right\}; \tag{16}$$

$$D_h^{2^*a} = \left(\cup_{i=1}^r D_{h,B_{1i}}^{2^*a}\right) \cup \left(\cup_{V_j \in \cup_{i=1}^r B_{2i}} D_{h,V_j}^{2^*a}\right), \forall h = 1, .., s \tag{17}$$

where $D_{h,B_{1i}}^{2^*a}$ is:

$$\{\mathcal{L} : \mathcal{L} \in \mathcal{P}(T_h \cup pa_D(T_h)) \setminus (\mathcal{P}(C_i \cup pa_D(T_h)) \cup \mathcal{P}(T_h \setminus B_{1i} \cup pa_D(T_h)))\} \tag{18}$$

and $D_{h,V_j}^{2^*a}$ is:

$$\begin{gathered}\{\mathcal{L} : \mathcal{L} \in \mathcal{P}(T_h \cup pa_D(T_h)) \setminus \\ \left(\mathcal{P}(Nb\left(V_j \cup \mathbf{B}_{1,V_j}\right) \cup pa_D(T_h)) \cup \mathcal{P}(T_h \setminus \left(V_j \cup \mathbf{B}_{1,V_j}\right) \cup pa_D(T_h))\right)\}\end{gathered} \tag{19}$$

and finally

$$D_h^{3^*b} = \cup_{\mathcal{A} \in PA_h} D_{h,\mathcal{A}}^{3^*b}, \forall h = 1, ..., s \tag{20}$$

where $D_{h,\mathcal{A}}^{3^*b}$ is:

$$\begin{gathered}\{\mathcal{L} : \mathcal{L} \in \mathcal{P}(T_h \setminus ch(\mathcal{A}) \cup pa_D(T_h)) \setminus \\ (\mathcal{P}(pa_D(T_h)) \cup \mathcal{P}(T_h \setminus ch(\mathcal{A}) \cup pa_D(T_h) \setminus \mathcal{A}))\}\end{gathered} \tag{21}$$

Appendix B

Proof of Markov equivalences

Proof of the Theorem 2 We prove that the independences *(C2*a)* follow from the the list of independences *(C2a)*. When $A = B_{1i}$, $B_{1i} \subset C_i \in Cl_h$, the set of neighbours of A is $nb(A) = nb(B_{1i}) = C_i \setminus B_{1i} = B_{2i}$. The neighbourhood of A is $Nb(A) = C_i$ and from *(C2a)* we get $B_{1i} \perp T_h \setminus C_i | pa_D(T_h) \cup B_{2i}$, that is the formula (6). On the other hand, when A is equal to $V_j \cup (\mathbf{B}_{1,V_j})$, where $\mathbf{B}_{1,V_j} = nb(V_j) \cap (\cup_{i=1}^r B_{1i})$, the set of neighbours of A is $nb(A) = nb(V_j \cup \mathbf{B}_{1,V_j})$ and the set of neighbourhood is $Nb(A) = Nb(V_j \cup (nb(V_j) \cap \mathbf{B}_{1,V_j}))$, so the *(C2a)* gives the formula (7).

Now we prove that from the *(C2*a)* we get the *(C2a)*. At this aim, we use the equivalence between the *(C2a)* and the following statement (See Lauritzen 1996):

$$V_j \perp T_h \setminus Nb(V_j) | pa_D(T_h) \cup nb(V_j) \qquad \forall V_j \in T_h \tag{22}$$

Thus, it is sufficient to prove the equivalence between the *(C2*a)* and the (22). From (6), applying the properties of the conditional independences (see Lauritzen, 1996), we have:

$$V_j \perp T_h \setminus C_i | pa_D(T_h) \cup B_2 \cup B_{1i} \setminus (V_j) \qquad \forall V_j \in B_{1i}$$

Since $\forall V_j \in B_{1i}$ by definition, $Nb(V_j) = C_i$, and $B_{2i} \cup B_{1i} \backslash V_j = C_i \backslash V_j = nb(V_j)$ the previous formula becomes:

$$V_j \perp T_h \backslash Nb(V_j) | pa_D(T_h) \cup nb(V_j) \qquad \forall V_j \in B_{1i}$$

that is the (22). The remaining vertices of T_H are considered in the statement (7) from which, using the same property we get:

$$V_j \perp T_h \backslash Nb(V_j \cup \mathbf{B}_{1,V_j}) | (pa_D(T_h) \cup nb(V_j \cup \mathbf{B}_{1,V_j}) \cup \mathbf{B}_{1,V_j}).$$

Note that, by definition, $\mathbf{B}_{1,V_j} \subset nb(V_j)$, thus $Nb(V_j \cup \mathbf{B}_{1,V_j}) = Nb(V_j)$. So, we get:

$$V_j \perp T_h \backslash Nb(V_j) | (pa_D(T_h) \cup nb(V_j))$$

that is the (22).

Proof of the Theorem 3 The list of independences from the *(C3b)* implies the list from the *(C3*b)*. Applying the statement *(C3b)* to $T_h \backslash ch(\mathcal{A})$, we get

$$T_h \backslash ch(\mathcal{A}) \perp pa_D(T_h) \backslash pa(T_h \backslash ch(\mathcal{A})) | pa(T_h \backslash ch(\mathcal{A})).$$

Since $\mathcal{A} \subseteq pa_D(T_h) \backslash pa(T_h \backslash ch(\mathcal{A}))$, from the properties of conditional independences it follows $T_h \backslash ch(\mathcal{A}) \perp \mathcal{A} | pa(T_h) \backslash \mathcal{A}$, that is the *(C3*b)*.
The *(C3*b)* implies the *(C3b)*. Given a set $A \subseteq T_h$, let $\{\mathcal{A}_1, \mathcal{A}_2, ..., \mathcal{A}_r\}$ be the collection of sets of PA_h having $A \subseteq T_h \backslash ch(\mathcal{A}_i)$, for $i = 1, .., r$. Now we consider the statements of independences related to this sets $\mathcal{A}_i$, $i = 1, .., r$:

$$\begin{cases} \mathcal{A}_1 \perp T_h \backslash ch(\mathcal{A}_1) | (pa_D(T_h) \backslash \mathcal{A}_1) \\ \qquad \qquad \cdots \\ \mathcal{A}_r \perp T_h \backslash ch(\mathcal{A}_r) | (pa_D(T_h) \backslash \mathcal{A}_r) \end{cases}$$

Since $A \subseteq T_h \backslash ch(\mathcal{A}_i)$, $\forall i = 1, ..., r$, with some passages we get:

$$\begin{cases} \mathcal{A}_1 \perp A | \left((pa_D(T_h) \backslash \left(\cup_{i=1}^r \mathcal{A}_i \right) \right) \cup \left(\cup_{i=1}^r \mathcal{A}_i \backslash \mathcal{A}_1 \right)) \\ \qquad \qquad \cdots \\ \mathcal{A}_r \perp A | \left((pa_D(T_h) \backslash \left(\cup_{i=1}^r \mathcal{A}_i \right) \right) \cup \left(\cup_{i=1}^r \mathcal{A}_i \backslash \mathcal{A}_r \right)) \end{cases}$$

Using the intersection property of conditional independence we obtain the *(C3b)*: $(\cup_{i=1}^r \mathcal{A}_i) \perp A | \left((pa_D(T_h) \backslash \left(\cup_{i=1}^r \mathcal{A}_i \right) \right))$.

Lemma 1. *Property of the sets B_{1i} and B_{2i}: We consider the family of the cliques Cl_h of a component T_h and its decomposition described in section 3.3. Then for all vertices $V_j \in \cup_{i=1}^r B_{2i}$, the set of vertices in $nb(V_j \cup \mathbf{B}_{1,V_j})$ is a subset of $\cup_{i=1}^r B_{2i}$.*

Proof. We consider $nb(V_j \cup \mathbf{B}_{1,V_j}) = \left(nb(V_j) \cup nb(\mathbf{B}_{1,V_j}) \right) \backslash \left(V_j \cup \mathbf{B}_{1,V_j} \right)$.
Since the set of neighbours of a set A do not contain the set itself, the right term can be rewritten as $\left(nb(V_j) \backslash \mathbf{B}_{1,V_j} \right) \cup \left(nb(\mathbf{B}_{1,V_j}) \backslash V_j \right)$.
By definition, $nb(B_{1i}) = B_2 i$, $\forall i = 1, .., r$, so $nb(\mathbf{B}_{1,V_j}) \subseteq \cup_{i=1}^r B_{2i}$ and consequently even $nb(\mathbf{B}_{1,V_j}) \backslash V_j \subseteq \cup_{i=1}^r B_{2i}$.
Since the set $\mathbf{B}_{1,V_j}$, by definition, is equal to $nb(V_j) \cap \cup_{i=1}^r B_{1i}$, then in the set $nb(V_j) \backslash \mathbf{B}_{1,V_j}$ there is not any vertex that is a subset of $\cup_{i=1}^r B_{1,i}$, thus $nb(V_j) \backslash \mathbf{B}_{1,V_j} \subseteq \cup_{i=1}^r B_{2,i}$. $\square$

Appendix C

Proof of the main results

Proof of the Theorem 4 In order to prove this theorem, we apply the theorem 1 to the parametrization $\{\boldsymbol{\eta}_{\mathcal{L}}^{\mathcal{M}} : \mathcal{L} \in \mathcal{P}(V), \mathcal{M} \in \mathcal{H}_{II}\}$.
According to the condition *(C1)*, the parameter to constrain to zero are $\boldsymbol{\eta}_{\mathcal{L}}^{\mathcal{M}(\mathcal{L})}$, when $\mathcal{L} \in \cup_{h=1}^{s} D_h^1$. Now, we check if the marginal $\mathcal{M}(\mathcal{L})$ satisfies the condition (4) of the theorem 1. It is easy to see that each element $\mathcal{L} \in D_h^1$ has at least one vertex in T_h and one vertex in $\cup_{j=1}^{h-1} T_j \backslash pa_D(T_h)$. Since for a given h the only marginal set containing these subsets of vertices is $\mathcal{M}_h^1 = \cup_{j \leq h} T_h$, the theorem 1 always holds for the independences following the *(C1)* property.
Regarding the *(C2*a)*, each vector of parameters $\boldsymbol{\eta}_{\mathcal{L}}^{\mathcal{M}}$, with $\mathcal{L} \in \cup_{i=1}^{s} D_h^{2^* a}$, must be constrained to zero. In this case, the condition (4) of theorem 1 holds if for every $\mathcal{L} \in \cup_{h=1}^{s} D_h^{2^* a}$, $\mathcal{M}(\mathcal{L})$ satisfies the conditions:

$$pa_D(T_h) \cup B_{2i} \subseteq \mathcal{M}(\mathcal{L}) \subseteq T_h \cup pa_D(T_h), \qquad \forall i = 1, ..., r \tag{23}$$

for the independence *(C2*a)* in (6) and the condition:

$$pa_D(T_h) \cup nb(V_j \cup \mathbf{B}_{1,V_j}) \subseteq \mathcal{M}(\mathcal{L}) \subseteq T_h \cup pa_D(T_h), \quad \forall V_j \in \cup_{i=1}^{r} B_{2i} \tag{24}$$

for the independence *(C2*a)* in (7) Obviously, when $\mathcal{M}(\mathcal{L})$ is equal to $\mathcal{M}_h^{2^* a}$, the conditions (23) and (24) are both satisfied and theorem 1 holds. But, in most of cases there will be a set $A \in \mathcal{P}(CH_h) \backslash J_h$, $A = \mathcal{L} \cap CH_h$, such that $\mathcal{M}(\mathcal{L}) = \mathcal{M}_{h,A}^{3^* b}$. When $\mathcal{M}(\mathcal{L}) = \mathcal{M}_{h,A}^{3^* b}$ the second inclusion in (23) and (24), still hold. For the first inclusion, it depends on the set A. From Lemma 1 it derives that, the conditional set $pa_D(T_h) \cup nb(V_j \cup \mathbf{B}_{1,V_j})$ is a subset of $pa_D(T_h) \cup (\cup_{i=1}^{r} B_{2i})$, thus if $\cup_{i=1}^{r} B_{2i}$ is contained in $NC_h \cup NA_h$, also the first inclusion in (23) and (24) are satisfied. This happens when in the CH_h sets there are only elements of B_{1i}, $\forall i = 1, ..., r$. From the definition of B_{1i}, it is sufficient that $Nb(V_j) \in \mathcal{C}_h$ or equivalently $nb(V_j) \in \mathcal{C}_h$.
On the other hand, if there is at least a vertex $V_j \in CH_h$ such that $nb(V_j) \notin \mathcal{C}_h$, so that $V_j \in B_{2i}$, for some i, then the vertex V_j occurs in a conditioning set of an independence of type $(C2 * a)$, that leads to constrain to zero some parameters $\boldsymbol{\eta}_{\mathcal{L}}^{\mathcal{M}(\mathcal{L})}$, where $\mathcal{L}$ is such that, $V_j \notin \mathcal{L}$, $\mathcal{L} \cap nb(V_j) \neq \emptyset$. If $\mathcal{L} \cap T_h \in J_h$, there is not any set A such that $\mathcal{M}(\mathcal{L}) = \mathcal{M}_{h,A}^{3^* b}$ and the smallest marginal containing these interactions is $\mathcal{M}_h^{2^* a}$ and the condition (4) of theorem 4 is still satisfied. If $\mathcal{L} \cap T_h \notin \mathcal{J}_h$, there is at least one set $A \in \mathcal{P}(CH)_h \backslash \mathcal{J}_h$ such that $A = \mathcal{L} \cap CH_h$ and $V_j \notin A$; thus $\mathcal{M}(\mathcal{L}) = \mathcal{M}_{h,A}^{3^* b}$ and the condition of the theorem 4 is violated.

This is not possible if all non-complete sub-sets C of T_h such that $C \cap nb(V_j) \neq \emptyset$, have $pa(C) = pa_D(T_h)$. This is equivalent to requiring that all not connected sets $K \in \mathcal{K}_h$, such that $nb(V_j) \cap K \neq \emptyset$, belong to J_h.
Finally, regarding the *(C3*b)*, the relationship $pa_D(T_h) \backslash \mathcal{A} \subseteq \mathcal{L} \subseteq T_h \backslash ch(\mathcal{A}) \cup pa_D(T_h) = \mathcal{M}_{h,A}^{3*b}$ must holds, where $A = T_h \backslash ch(\mathcal{A}) \cap CH_h$. Since each $\mathcal{M}_{h,A}^{3*b}$ contains $pa_D(T_h)$, the first inclusion is always verified. Even the second inclusion always holds, since there is a marginal set for every $A \in \mathcal{P}(CH_h) \backslash J_h$.

Proof of the Theorem 5 We will show that changing the marginal sets $\mathcal{M}_{h,A}^{3*b}$, the smooth sub-class of GMs II is reduced. From the proof of the previous Theorem it possible to see that, if we consider marginal sets $\mathcal{M}^{3*b}_{h,A}$ such that $(NC_h \cup NA_h) \notin \mathcal{M}^{3*b}_{h,A}$, not only the independences *(C2*a)*, having the children in the conditional set are potentially problematic, but every independences *(C2*a)*. Moreover if we consider a sub-class of the marginal sets $\mathcal{M}^{3*b}_{h,A}$, for some independences *(C*3b)* the second inclusion of condition 1 may be violated.

Reference

1 S. A. Andersson, D. Madigan, M. D. Perlman, Alternative Markov properties for chain graphs. Scandinavian Journal of Statistics 28(1): 33-85.(2001)

2 F. Bartolucci, R. Colombi, A. Forcina, An extended class of marginal link functions for modelling contingency tables by equality and inequality constraints, Statistica Sinica 17(2): 691.(2007)

3 W. P. Bergsma, T. Rudas, Marginal models for categorical data. The Annals of Statistics 30.1: 140-159.(2002)

4 R. Colombi, S. Giordano, M. Cazzaro, and the R Development Core Team. hmmm: Hierarchical Multinomial Marginal Models. R package version 1.0-1 (2013)

5 M. Drton, Discrete chain graph models, Bernoulli 15(3): 736-753.(2009)

6 Equal opportunity Committee- University of Milano- Bicocca. Survey on the issue of sexual harassment, as unwanted behaviors of sexual nature. (2012)

7 S. L. Lauritzen, Graphical Models. Oxford University Press, 1996.

8 F. Nicolussi, Marginal parameterizations for conditional independence models and graphical models for categorical data, PhD Thesis (2013).

9 T. Rudas, W. P. Bergsma, and R. Nmeth, Marginal log-linear parameterization of conditional independence models. Biometrika 97.4 : 1006-1012 (2010).