Structure of the High-Order Boltzmann Machine from Independence Maps

F. Xabier Albizuri, Alicia d'Anjou, Manuel Graña, *Member, IEEE,* and Pedro Larrañaga, *Member, IEEE*

***Abstract*— In this paper we consider the determination of the structure of the high-order Boltzmann machine (HOBM), a stochastic recurrent network for approximating probability distributions. We obtain the structure of the HOBM, the hypergraph of connections, from conditional independences of the probability distribution to model. We assume that an expert provides these conditional independences and from them we build independence maps, Markov and Bayesian networks, which represent conditional independences through undirected graphs and directed acyclic graphs respectively. From these independence maps we construct the HOBM hypergraph. The central aim of this paper is to obtain a minimal hypergraph. Given that different orderings of the variables provide in general different Bayesian networks, we define their intersection hypergraph. We prove that the intersection hypergraph of all the Bayesian networks $(N!)$ of the distribution is contained by the hypergraph of the Markov network, it is more simple, and we give a procedure to determine a subset of the Bayesian networks that verifies this property. We also prove that the Markov network graph establishes a minimum connectivity for the hypergraphs from Bayesian networks.**

***Index Terms*— Bayesian networks, Boltzmann machines, independence maps, graphical models, log-linear models, neural networks.**

I. Introduction

The conventional Boltzmann machine (BM) [1], [9], as well as the high-order Boltzmann machine (HOBM) [15], [3], is a technique whose purpose is, in its fundamental formulation, to describe and model probability distributions defined on a set of binary random variables.

The BM approximates a distribution with a model where the probability function is defined as the normalized exponential of a consensus function. The learning algorithm is a steepest descent of the Kullback–Leibler divergence between the distribution to learn and the approximation distribution. In the conventional BM there are hidden units, the consensus function is formed by first and second-degree terms on the variables, i.e., connections up to order two between units, and the approximation distribution is the marginal distribution on the visible units. The HOBM is a variation of the conventional BM where we consider higher order connections and do not use hidden units. The HOBM allows us to undertake the problem considered in this paper, that is the determination of the structure of the BM.

The structure of the HOBM is given by the set of connections. We will call this connection set the hypergraph of the HOBM. We study the determination of the hypergraph from independence maps [5], [18]. The books by Pearl [13], Whittaker [19], and Lauritzen [11] are good comprehensive works on the subject of independence maps. An independence map is a graph with a vertex separation criterion for inferring conditional independences between random variables, the vertices of the graph, and it represents part of the independences of a probability distribution. There are basically two kinds of independence maps: Markov networks, which are undirected graphs, and Bayesian networks, which are directed acyclic graphs. Some attempts to link independence maps and BM's have been carried out [6], [10], [12]. In this paper we give a systematic solution to the problem of determining the structure of the HOBM from conditional independences.

Our approach assumes that an expert provides us conditional independences of the probability distribution to learn. Markov and Bayesian networks are built from these conditional independences. The structure of the HOBM will be fixed according to the probability function factorizations associated with these independence maps. In this paper we will show how to construct the hypergraph of connections of the HOBM from Markov and Bayesian networks. Given a probability distribution, i.e., its conditional independences, the Markov network is unique, but we have in general different Bayesian networks for different orderings of the variables. We will introduce the notion of intersection hypergraph of the Bayesian networks corresponding to different orderings of the variables. The central aim of our paper is to get a minimal hypergraph, and we have investigated whether the Markov network provides this minimal hypergraph, or we can optimize it by means of Bayesian networks, searching for suitable variable orderings. In this study we will prove that we can determine a set of Bayesian networks such that the intersection hypergraph is contained by the hypergraph of the Markov network, it is more simple. We will establish this result through chordal independence maps, which will be an intermediate step between the Markov network and the Bayesian networks that we search for. We will also prove that the Markov network determines a minimum of connectivity for the hypergraphs constructed from Bayesian networks.

In this paper we will show how we can define the structure of the HOBM, i.e., the set of weighted connections, from the qualitative information provided by conditional indepen-

Manuscript received July 29, 1995; revised May 16, 1996, January 28, 1997, and July 2, 1997. This work was supported by Grants PI94-78 from the Department of Education of the Basque Government, OF93/96 from Gipuzkoako Foru Aldundia, and 140.226-EA080/96 from UPV/EHU.

The authors are with the Department of Computer Science and Artificial Intelligence, University of the Basque Country, Informatika Fakultatea, Donostia, Spain.

Publisher Item Identifier S 1045-9227(97)07527-9.

dences. Once the structure is defined, the learning algorithm of the HOBM can be applied to a set of samples from the distribution to learn, obtaining the connection weights that minimize the divergence between the approximation distribution and the distribution of the sample set [2].

This paper is organized as follows. In Section II we introduce the HOBM. In Section III we study the determination of the structure of the HOBM, the hypergraph of connections, from factorizations of the probability function to learn. In Sections IV and V we get hypergraphs from Markov and Bayesian networks respectively. In Section VI we present the results that relate the hypergraphs obtained from these independence maps. In Section VII we end with the conclusions.

II. The High-Order Boltzmann Machine

A BM is a stochastic recurrent network with a local learning algorithm. The configuration of the HOBM with N units is defined by $\mathbf{u} \in \{0,1\}^N$ and the state of the unit $i \in \{1, \cdots, N\}$ by x_i. A connection $\lambda = \{i_1, \cdots, i_m\}$ is a nonempty subset of $[1,N] = \{1, \cdots, N\}$, that is $\lambda \in \mathcal{P}^*([1,N])$, and $|\lambda|$ denotes the order m of the connection λ, the number of its ends. Some connections have an associated weight, a real number w_λ, modified by the learning algorithm. We define the *consensus function*

$$C^*(\mathbf{u}) = \sum_{\lambda \in L^*} w_\lambda a(\lambda \mid \mathbf{u})$$

where $L^* \subseteq \mathcal{P}^*([1,N])$ is the set of weighted connections. The function $a(\lambda \mid \mathbf{u})$ is defined as

$$a(\lambda \mid \mathbf{u}) = \prod_{i \in \lambda} x_i$$

so $a(\lambda \mid \mathbf{u}) = 1$ if every end of λ takes value one (the connection is activated) and $a(\lambda \mid \mathbf{u}) = 0$ otherwise.

The stochastic transition law is the following: given a configuration $\mathbf{u}$ of the HOBM, we choose at random an unit j and change its state x_j to $x_j' = 1 - x_j$ with probability

$$\frac{1}{1 + \exp(-\Delta C^*(\mathbf{u}))}$$

where

$$\Delta C^*(\mathbf{u}) = (1 - 2x_j) \sum_{\lambda \in L^*/j \in \lambda} w_\lambda a(\lambda - \{j\} \mid \mathbf{u}).$$

The dynamics defined corresponds to a Markov chain where the stationary probability distribution is the Boltzmann–Gibbs distribution

$$P^*(\mathbf{u}) = \frac{1}{Z} \exp C^*(\mathbf{u})$$

where $Z = \sum_{\mathbf{u}} \exp C^*(\mathbf{u})$.

The purpose of the HOBM is to approximate a positive probability distribution $P(\mathbf{u})$ on $\{0,1\}^N$, usually given by the frequency distribution of a set of samples, with the distribution $P^*(\mathbf{u})$. The learning algorithm is a steepest descent of the Kullback–Leibler divergence

$$D = \sum_{\mathbf{u}} P(\mathbf{u}) \ln \frac{P(\mathbf{u})}{P^*(\mathbf{u})}. \tag{1}$$

The weights $\{w_\lambda / \lambda \in L^*\}$ are modified by the iterative rule

$$w_\lambda^{k+1} = w_\lambda^k - \alpha \cdot (p_\lambda^* - p_\lambda)$$

where

$$p_\lambda^* = \sum_{\mathbf{u}} P^*(\mathbf{u}) a(\lambda \mid \mathbf{u}) \qquad p_\lambda = \sum_{\mathbf{u}} P(\mathbf{u}) a(\lambda \mid \mathbf{u})$$

are the activation probabilities of the connection λ under the approximation distribution and under the distribution to learn.

The learning algorithm of the HOBM converges to the strict global minimum of the divergence (1), which corresponds to the maximum likelihood estimate of the connection weights [2].

III. Structure of the HOBM from Conditional Independences

We will determine the structure of the HOBM, the hypergraph of weighted connections L^*, from conditional independences. We assume that an expert provides the conditional independences of the probability distribution to learn $P(\mathbf{u})$ and from these independences we will establish the structure of the HOBM. Once the structure is determined the learning algorithm of the HOBM is applied to a set of samples from $P(\mathbf{u})$ obtaining the distribution $P^*(\mathbf{u})$, the estimation of the distribution to learn $P(\mathbf{u})$.

Given the conditional independences, the structure of the HOBM is obtained through the factorizations of the probability function $P(\mathbf{u})$ provided by these independences. In order to determine the structure from factorizations of the distribution to learn we begin defining the *factorization hypergraphs*.

We start considering the factorization of a probability function. Let $P(\mathbf{u})$ be a positive probability function on $\{0,1\}^N$ that admits a factorization

$$P(\mathbf{u}) = \prod_{i=1}^{m} Q_i(\mathbf{u}_i) \tag{2}$$

where[1] $\mathbf{U}_i \subseteq \mathbf{U} = \{X_1, \cdots, X_N\}$. It can be shown that the probability function $P(\mathbf{u})$ can be written as

$$P(\mathbf{u}) = \frac{1}{Z} \exp C(\mathbf{u})$$

through a consensus function

$$C(\mathbf{u}) = \sum_{\lambda \in L} w_\lambda a(\lambda \mid \mathbf{u})$$

where the weights w_λ are determined and the set of weighted connections is

$$L = \bigcup_{i=1}^{m} \{\lambda \in \mathcal{P}^*([1,N]) / \lambda \subseteq I(\mathbf{U}_i)\}. \tag{3}$$

We name L the *hypergraph of the factorization* (2). In this notation $I(\mathbf{U}_i)$ is the set of indexes of the variables in $\mathbf{U}_i$. A detailed proof is in Appendix A.

[1] We use capital letters for variables, lower-case letters for values taken by variables, boldfaced capital letters for (ordered) sets of variables, and boldfaced lower-case letters for assignments of values to the variables in these sets. So a boldfaced lower-case letter, with or without subscript, is a vector with dimension the number of variables in the set.

So, if a probability function $P(\mathbf{u})$ admits a factorization (2) it can be written as the normalized exponential of a consensus function whose terms correspond with the connections in (3), that is the connections in L correspond to the subsets of the variables in the argument of each function $Q_i(\mathbf{u}_i)$.

If $P(\mathbf{u})$ admits various factorizations, $L_1, \cdots, L_p$ being the corresponding factorization hypergraphs, then $P(\mathbf{u})$ is the normalized exponential of a consensus function whose terms correspond with the connections in the *intersection hypergraph*

$$L = \bigcap_{j=1}^{p} L_j.$$

This is clear since given $P(\mathbf{u})$ the w_λ are determined.

Establishing the structure of the HOBM from factorization hypergraphs is justified as follows. Let L be the intersection hypergraph of some factorizations of the distribution to learn $P(\mathbf{u})$ and let $\mathcal{S}$ be a set of samples whose frequency distribution is $P(\mathbf{u})$. We define the set of weighted connections $L^* = L$. The learning algorithm of the HOBM converges to the global minimum of the divergence (1). The divergence D is nonnegative [19], and $D = 0$ if and only if $P(\mathbf{u}) = P^*(\mathbf{u})$. Since $L^* = L$ there exist connection weights such that $P^*(\mathbf{u}) = P(\mathbf{u})$. Therefore the learning algorithm of the HOBM converges to the connection weights for which $P^*(\mathbf{u}) = P(\mathbf{u})$.

In general we have a set of samples from the distribution to learn $P(\mathbf{u})$, we define $L^* = L$ and the learning algorithm converges to the maximum likelihood estimate of the connection weights. So we obtain, among the distributions that can be written through the hypergraph $L^* = L$ provided by the factorizations, the distribution $P^*(\mathbf{u})$ that best fits the sample set.

We consider now the factorization of a probability function $P(\mathbf{u})$ from conditional independences. Let $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ be three disjoint subsets of $\mathbf{U}$, being $\mathbf{X}, \mathbf{Y} \neq \emptyset$. The subsets of variables $\mathbf{X}$ and $\mathbf{Y}$ are *conditionally independent given* $\mathbf{Z}$, we write $\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z}$, if

$$P(\mathbf{x}, \mathbf{y} \mid \mathbf{z}) = P(\mathbf{x} \mid \mathbf{z}) P(\mathbf{y} \mid \mathbf{z})$$

whenever $P(\mathbf{z}) > 0$. Equivalently $\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z}$ if $\mathbf{P}(\mathbf{x} \mid \mathbf{y}, \mathbf{z}) = P(\mathbf{x} \mid \mathbf{z})$ whenever $P(\mathbf{y}, \mathbf{z}) > 0$. We will study the factorizations provided by independence maps, Markov and Bayesian networks that represent graphically (part of the) conditional independences of a probability distribution. The Markov networks are undirected graphs and the Bayesian networks are directed acyclic graphs (DAG's). A separation criterion allows us to infer the conditional independences from the graph.

IV. Hypergraph from the Markov Network

Let $P(\mathbf{u})$ be a probability distribution on a set of variables $\mathbf{U}$. An undirected graph $G = (\mathbf{U}, \mathbf{E})$ can be used to represent (part of) its conditional independences, where the set of vertices is $\mathbf{U}$ and $\mathbf{E}$ is the set of edges between vertices. If a subset $\mathbf{Z}$ of vertices intercepts every path[2] between the vertices of $\mathbf{X}$ and those of $\mathbf{Y}$ then we write $\langle \mathbf{X} \mid \mathbf{Z} \mid \mathbf{Y} \rangle$, that is $\mathbf{Z}$ separates $\mathbf{X}$ and $\mathbf{Y}$. An undirected graph G is an *independence map* or *I-map* of $P(\mathbf{u})$ if

$$\langle \mathbf{X} \mid \mathbf{Z} \mid \mathbf{Y} \rangle \Rightarrow \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z}.$$

A *Markov network* of $P(\mathbf{u})$ is a minimal I-map of $P(\mathbf{u})$. It is constructed according to the following result proved by Pearl and Paz [14], [13]: every positive distribution $P(\mathbf{u})$ has a unique Markov network $G_0 = (\mathbf{U}, \mathbf{E}_0)$, where

$$(\alpha, \beta) \notin \mathbf{E}_0 \text{ iff } \alpha \perp\!\!\!\perp \beta \mid \mathbf{U} - \alpha - \beta.$$

It follows that the graph $G = (\mathbf{U}, \mathbf{E})$ is an I-map of $P(\mathbf{u})$ if and only if the Markov network of $P(\mathbf{u})$, $G_0 = (\mathbf{U}, \mathbf{E}_0)$, is a partial graph of G, i.e., $\mathbf{E}_0 \subseteq \mathbf{E}$.

We are interested in the factorizations obtained from the conditional independences represented in the Markov network of a distribution. Hammersley and Clifford [7] showed that a graph $G = (\mathbf{U}, \mathbf{E})$ is an I-map of $P(\mathbf{u})$ if and only if $P(\mathbf{u})$ is a normalized product of nonnegative functions on the cliques[3] of G (also proved in [11]). Therefore given the Markov network $G_0 = (\mathbf{U}, \mathbf{E}_0)$ of a positive distribution $P(\mathbf{u})$, it admits a factorization

$$P(\mathbf{u}) = \prod_{i=1}^{m} Q_i(\mathbf{c}_i)$$

where $\mathbf{C}_1, \cdots, \mathbf{C}_m$ are the cliques of G_0. The hypergraph of this factorization is

$$L_{G_0} = \bigcup_{i=1}^{m} \{\lambda \in \mathcal{P}^*([1, N]) / \lambda \subseteq I(\mathbf{C}_i)\}. \tag{4}$$

The connections in L_{G_0} correspond to the subsets of the cliques. We will name it the *hypergraph of the Markov network* of $P(\mathbf{u})$.

We present an example where the structure of the HOBM will be determined from the Markov network. Let X, Y, Z, U, V, and W be variables corresponding to units that are fixed according to a probability law defined as follows. The values zero or one for X, Y, and Z are fixed randomly and independently. When $X = Y$ we fix $U = 1$ with probability 0.9 ($U = 0$ with probability 0.1) and when $X \neq Y$ we fix $U = 0$ with probability 0.9. Likewise, if $Y = Z$ the variable V tends to one and otherwise to zero, and if $X = Z$ the variable W tends to one and otherwise to zero.

In Fig. 1 we show the Markov network G_0 of this probability distribution. For instance, knowing Y, Z, V, and W, the variables X and U are not independent, then the edge (X, U) is in G_0. Knowing Z, U, V, and W, the variables X and Y are not independent (we know U), then (X, Y) is in G_0. Knowing Y, Z, U, and W, the variables X and V are independent, (X, V) is not in G_0. The cliques of G_0 are $\{X, Y, Z\}$, $\{X, Y, U\}$, etc. and therefore the hypergraph of the Markov network is L_{G_0}.

[2] We follow generally the terminology of [8].

[3] Given an undirected graph, a subset of vertices is complete if all its vertices are adjacent to each other. A clique is a maximal complete subset.

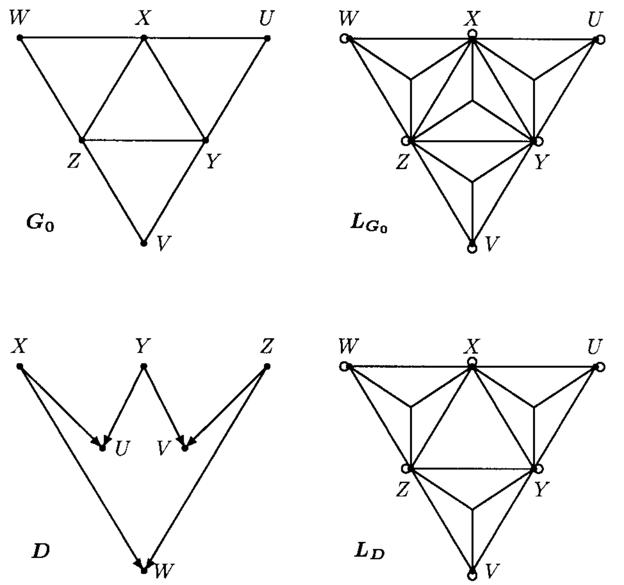


Fig. 1. In our example G_0 is the Markov network and L_{G_0} its hypergraph. For the ordering $\{X, Y, Z, U, V, W\}$ we have the Bayesian network D and its hypergraph L_D, D being a Bayesian network on $G_0 = G^*$. The intersection hypergraph of all the Bayesian networks is L_D. We note that $L_D \subset L_{G_0}$, and $G(L_D) = G_0$. The architecture of the HOBM would be given by L_D.

V. Hypergraphs from Bayesian Networks

Besides the undirected graphs, with their straightforward separation criterion for inferring conditional independences, DAG's can be used to represent conditional independences of a probability distribution. The *d-separation* criterion in a DAG is as follows. Given a DAG $D = (\mathbf{U}, \mathbf{E})$, where $\mathbf{U}$ is the vertex set and $\mathbf{E}$ the arrow set, the vertex subset $\mathbf{Z}$ is said to activate a path[4] between a vertex of $\mathbf{X}$ and a vertex of $\mathbf{Y}$ if, for vertices in this path:

1) every vertex with two incident arrows (of the path) is in $\mathbf{Z}$ or has some descendant[5] in $\mathbf{Z}$;
2) the remaining vertices are not in $\mathbf{Z}$.

The subset $\mathbf{Z}$ is said to d-separate $\mathbf{X}$ and $\mathbf{Y}$, $\langle \mathbf{X} \mid \mathbf{Z} \mid \mathbf{Y} \rangle_D$, if there are no paths between vertices in $\mathbf{X}$ and in $\mathbf{Y}$ activated by $\mathbf{Z}$, i.e., every path is *blocked* by $\mathbf{Z}$. A DAG $D = (\mathbf{U}, \mathbf{E})$ is an I-map of $P(\mathbf{u})$ if $\langle \mathbf{X} \mid \mathbf{Z} \mid \mathbf{Y} \rangle_D$ implies $\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z}$.

A *Bayesian network* of $P(\mathbf{u})$ is a DAG that is a minimal I-map. It is constructed as follows. Given $P(\mathbf{u})$, let us consider an ordering of the variables $\mathbf{U} = \{X_1, X_2, \cdots, X_N\}$. Let $\mathbf{F}_{X_i}$ be a minimal subset of $\mathbf{U}_{X_i} = \{X_1, X_2, \cdots, X_{i-1}\}$ such that

$$X_i \perp\!\!\!\perp \mathbf{U}_{X_i} - \mathbf{F}_{X_i} \mid \mathbf{F}_{X_i} \tag{5}$$

(true if $\mathbf{U}_{X_i} - \mathbf{F}_{X_i} = \emptyset$). The *boundary DAG* of $P(\mathbf{u})$ relative to the ordering of $\mathbf{U}$ is the DAG obtained assigning the vertices of $\mathbf{F}_{X_i}$ as the parents of X_i, for $i = 1, \cdots, N$. It is unique given an ordering if $P(\mathbf{u})$ is positive [13], and[6] we will obtain $\mathbf{F}_{X_i}$ starting with $\mathbf{F}^*_{X_i} = \mathbf{U}_{X_i}$ and eliminating successively vertices so that the conditional independence (5) holds. Verma [17] proved that if D is a boundary DAG of $P(\mathbf{u})$, then D is a Bayesian network of $P(\mathbf{u})$. Conversely, given a DAG D that is I-map of $P(\mathbf{u})$, from d-separation, every variable X is independent of all its nondescendants given its parents $\mathbf{F}_X$.

So given $P(\mathbf{u})$ we must fix an ordering of the variables in $\mathbf{U}$ to obtain the Bayesian network, constructing the boundary DAG. We have in general different Bayesian networks of $P(\mathbf{u})$ for different orderings.

We consider now the factorization provided by a Bayesian network. Let D be a Bayesian network of a positive distribution $P(\mathbf{u})$ on $\{0,1\}^N$ and $\mathbf{U} = \{X_1, X_2, \cdots, X_N\}$ an ordering consistent with the DAG D, that is the parents $\mathbf{F}_{X_i}$ of a variable X_i come before X_i in the ordering (such ordering always exists). Writing the chain rule

$$\begin{aligned} &P(x_1, x_2, \cdots, x_N) \\ &\quad = P(x_N \mid x_{N-1}, \cdots, x_1) P(x_{N-1} \mid x_{N-2}, \cdots, x_1) \cdots \\ &\qquad P(x_3 \mid x_2, x_1) P(x_2 \mid x_1) P(x_1) \end{aligned}$$

we have that each vertex is independent of its predecessors in the ordering given its parents, therefore

$$P(\mathbf{u}) = \prod_{i=1}^{N} P(x_i \mid \mathbf{f}_{X_i}).$$

The hypergraph of this factorization is

$$L = \bigcup_{i=1}^{N} \{\lambda \in \mathcal{P}^*([1, N]) / \lambda \subseteq I(\{X_i\} \cup \mathbf{F}_{X_i})\}.$$

The connections in L correspond to the subsets of the parents of X_i, together with X_i, for each vertex. We will name it the *hypergraph of the Bayesian network* D of $P(\mathbf{u})$. In general, we have different Bayesian networks calculating the boundary DAG for each ordering of $\mathbf{U}$ ($N!$ possible orderings) and therefore different hypergraphs.

We continue with the example introduced in Section IV and shown in Fig. 1. For the variable ordering $\{X, Y, Z, U, V, W\}$ the Bayesian network is D. For instance Z is independent of X and Y, it has not any parents. Knowing X and Z, the variable W is independent of Y, U, and V, the parents of W are X and Z. The hypergraph of this Bayesian network is L_D. For other orderings of the variables, as in $\{U, V, W, X, Y, Z\}$, more complex Bayesian networks and hypergraphs are obtained.

VI. Relation Between the Markov Network Hypergraph and Intersection Hypergraphs from the Bayesian Networks

Given the conditional independences of the distribution $P(\mathbf{u})$ to learn we can establish the structure of the HOBM through the hypergraph of the Markov network or through the intersection hypergraph of Bayesian networks corresponding to different orderings of $\mathbf{U}$. The question is which hypergraphs are more simple. We will develop some theoretical results that relate these hypergraphs.

In the Section VI-A we will prove that, given the conditional independences of a distribution P, we can determine a set

[4] A path is a list of different adjacent vertices, without considering the direction of the arrows.

[5] β is said to be a descendant of α if there is a directed path from α to β.

[6] According to the weak union axiom [13]: if X and Y ∪ W are independent given Z then X and Y are independent given Z ∪ W.

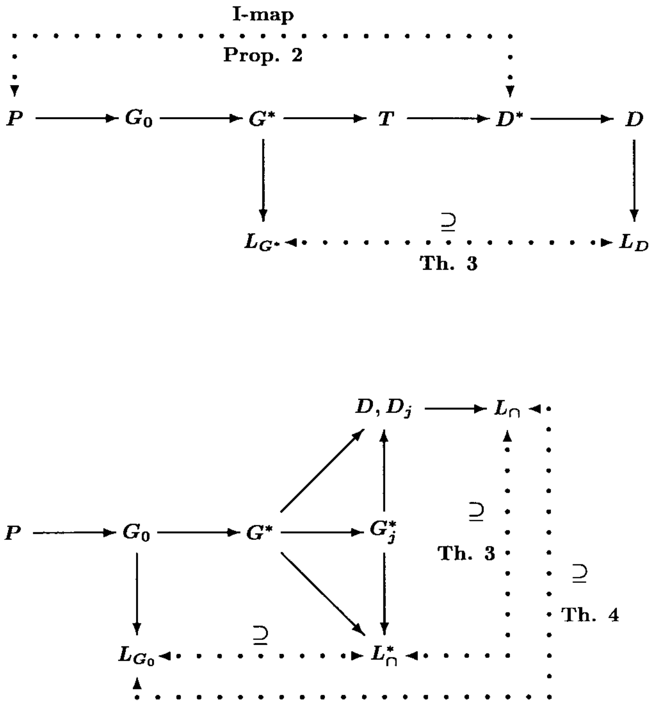


Fig. 2. From the independences of P we have the Markov network G_0 and from it a chordal I-map G^*. Through a join tree T we define D^*. By Proposition 2, D^* is an I-map. From D^* we get a Bayesian network D, Theorem 3 stating that $L_D \subseteq L_{G^*}$. From G^* we define chordal I-maps G_j^* such that, together with G^*, $L_\cap^* \subseteq L_{G_0}$. D, D_j are Bayesian networks on G^*, G_j^*, and by Theorem 3, $L_\cap \subseteq L_\cap^*$, so $L_\cap \subseteq L_{G_0}$, obtaining Theorem 4.

of Bayesian networks such that the intersection hypergraph is contained by the hypergraph of the Markov network, therefore the intersection hypergraph of all the Bayesian networks is more simple than the Markov network hypergraph. In order to obtain this result we will start from the Markov network and we will construct a determined set of chordal I-maps triangulating it. The intersection hypergraph of Bayesian networks on these chordal I-maps will minimize the Markov network hypergraph.

The structure of our argument is illustrated in Fig. 2. From the Markov network G_0 a chordal I-map G^* can be obtained. By means of a join tree T associated with G^* we will define a DAG D^* on G^*. The DAG D^* will be an I-map (Proposition 2). From D^* we will get a Bayesian network D on G^*, its hypergraph L_D being contained by the hypergraph of the chordal I-map, L_{G^*} (Theorem 3). We will continue with our discussion as follows. From the chordal I-map G^* we will define a set of chordal I-maps G_j^* such that, together with G^*, the intersection hypergraph $L_\cap^*$ is contained by the hypergraph of the Markov network L_{G_0}. Considering Bayesian networks D, D_j on G^*, G_j^*, we will have (from Theorem 3) that their intersection hypergraph $L_\cap$ is contained by $L_\cap^*$, therefore $L_\cap$ is contained by L_{G_0}. So we will obtain Theorem 4, the intersection hypergraph of the Bayesian networks is contained in the hypergraph of the Markov network.

We will end this study with Theorem 5 and Corollary 6 in the Section VI-B, proving that although we can get from Bayesian networks a hypergraph more simple than the hypergraph of the Markov network, the Markov network establishes a minimum of connectivity for the hypergraphs of Bayesian networks.

A. Minimizing the Markov Network Hypergraph

We will begin with the *chordal* I-maps. An undirected graph $G = (\mathbf{U}, \mathbf{E})$ is chordal if every cycle of length four or more has an edge joining two nonconsecutive vertices. We have the following characterization of chordal graphs due to Beeri *et al.* [4].

Theorem 1: An undirected graph $G = (\mathbf{U}, \mathbf{E})$ is chordal if and only if there exists a tree[7] T (join tree) with the cliques of G as vertices, such that for every vertex α of G, any two cliques containing α are either adjacent in T or connected by a path made of cliques that contain α.

If the Markov network G_0 of $P(\mathbf{u})$ is not chordal we will add edges to get a chordal I-map of $P(\mathbf{u})$, which is not unique in general. An algorithm for triangulating a graph and getting a join tree is described in [13] and [16]. We will prove that given a chordal I-map G^* of $P(\mathbf{u})$ there exists a Bayesian network whose hypergraph is contained in the hypergraph of G^* (defined as in (4), by means of its cliques).

If G^* is a chordal graph and T a join tree of G^*, we direct acyclicly the edges of G^* as follows. Let us consider a directing of the tree T and an ordering of the vertices $\mathbf{C}$ of T consistent with T directed, i.e., if $\mathbf{C}_{f(i)}$ is the parent of $\mathbf{C}_i$ then $f(i) < i$. Following the ordering of the cliques, we number the vertices (in the first clique where they appear), numbering arbitrarily if in a clique several vertices appear the first time. Defining as parents of a vertex the adjacent vertices with smaller index, we get a *DAG on* G^* *consistent with* T.

Proposition 2: If D^* is a DAG on a chordal I-map G^* of $P(\mathbf{u})$ consistent with a join tree T of G^*, then D^* is an I-map of $P(\mathbf{u})$.

A detailed proof of this Proposition is found in Appendix B.

Let D^* be a DAG on a chordal I-map G^* of $P(\mathbf{u})$ consistent with a join tree T of G^*. From Proposition 2, we have that D^* is an I-map of $P(\mathbf{u})$, then $X_i \perp\!\!\!\perp \{X_1, X_2, \cdots, X_{i-1}\} - \mathbf{F}_{X_i}^* \mid \mathbf{F}_{X_i}^*$ for every X_i, therefore we can get a boundary DAG D such that $\mathbf{F}_{X_i} \subseteq \mathbf{F}_{X_i}^*$, $\mathbf{F}_{X_i}$ and $\mathbf{F}_{X_i}^*$ being the parents of X_i in D and D^*, respectively. We say that D is a *Bayesian network on* G^* *consistent with* T.

Theorem 3: Let $P(\mathbf{u})$ be a positive distribution on $\{0,1\}^N$. Given a chordal I-map G^*, the hypergraph of any Bayesian network on G^* consistent with a join tree T of G^* is contained in the hypergraph of G^*.

A detailed proof of this Theorem is found in Appendix B.

We have proved that given a chordal I-map of a distribution, obtained triangulating its Markov network, we can define an ordering of variables such that the corresponding Bayesian network has a hypergraph contained in the hypergraph of the chordal I-map. Now we will prove that the intersection hypergraph of (some of) the Bayesian networks is contained in the hypergraph of the Markov network.

Let G_0 be the Markov network of $P(\mathbf{u})$ and G^* a chordal I-map obtained adding the edges $l_1, \cdots, l_q$ to G_0. We define G_j^*, for $j = 1, \cdots, q$, as follows. Let G_j be the graph obtained

[7] An undirected graph T is a tree if it is connected and acyclic.

from G^* eliminating the edge l_j. We define G_j^* as a chordal graph that we get adding edges different from l_j to G_j (this is always possible).

The intersection hypergraph $L_\cap^*$ of $G_1^*, \cdots, G_q^*, G^*$ is contained in the hypergraph L_{G_0} of the Markov network. Effectively, if $\lambda = \{i_1, \cdots, i_p\} \in L_\cap^*$, the set of vertices $\mathbf{U}_\lambda = \{X_{i_1}, \cdots, X_{i_p}\}$ is complete in $G_1^*, \cdots, G_q^*$, then in $\mathbf{U}_\lambda$ there is at most one end of l_j, for $j = 1, \cdots, q$; but $\mathbf{U}_\lambda$ is complete in G^*, which has not more edges than G_0 save $l_1, \cdots, l_q$, then $\mathbf{U}_\lambda$ is complete in G_0, $\lambda \in L_{G_0}$. Therefore $L_\cap^* \subseteq L_{G_0}$ (in fact $L_\cap^* = L_{G_0}$).

Let $D_1, \cdots, D_q, D$ be, respectively, Bayesian networks on $G_1^*, \cdots, G_q^*, G^*$ consistent with respective join trees. From the preceding result and Theorem 3 we have that the intersection hypergraph $L_\cap$ of $D_1, \cdots, D_q, D$ is contained in L_{G_0}. Consequently we can state this theorem.

Theorem 4: Let $P(\mathbf{u})$ be a positive distribution on $\{0,1\}^N$. The intersection hypergraph of the Bayesian networks is contained in the hypergraph of the Markov network.

It is not necessary to consider all ($N!$) the Bayesian networks in order to obtain a hypergraph more simple than the hypergraph of the Markov network. We can obtain such hypergraph following the procedure described above.

B. Minimum Connectivity of an Intersection Hypergraph of Bayesian Networks

We will terminate this study showing that the Markov network of a distribution establishes a minimum of connectivity for the hypergraphs of Bayesian networks. Let L be the intersection hypergraph of some Bayesian networks. We define the (undirected) graph associated with L, $G(L)$, saying that the adjacency in[8] L and in $G(L)$ coincide.

Theorem 5: Let L be the intersection hypergraph of some Bayesian networks of a positive distribution $P(\mathbf{u})$ on $\{0,1\}^N$. The Markov network G_0 is a partial graph of $G(L)$.

Corollary 6: If L is contained in the hypergraph of the Markov network G_0, then $G(L) = G_0$.

Let us prove the theorem. If L is the intersection hypergraph of some Bayesian networks of $P(\mathbf{u})$, then $P(\mathbf{u}) = Z^{-1} \exp C(\mathbf{u})$, where

$$C(\mathbf{u}) = \sum_{\lambda \in L} w_\lambda a(\lambda \mid \mathbf{u}). \tag{6}$$

Each $\lambda \in L$ is included in some clique of $G(L)$ so that we can group the terms in (6) according to the cliques of $G(L)$ (there will be several possible groupings). Therefore $P(\mathbf{u})$ is a product of functions on the cliques of $G(L)$, then by Section IV, $G(L)$ is an I-map of $P(\mathbf{u})$, i.e., G_0 is a partial graph of $G(L)$. The corollary follows directly from the theorem.

Consequently, given a probability distribution, if L is the intersection hypergraph of all the Bayesian networks or the intersection hypergraph of the subset of Bayesian networks provided by the procedure described above, we know that L is more simple than the Markov network hypergraph $L \subseteq L_{G_0}$, but the Markov network G_0 establishes a minimum connectivity for L in the sense that we have defined, that is $G(L) = G_0$.

We conclude with the example studied in Sections IV and V, in Fig. 1. The Markov network G_0 is chordal, $G_0 = G^*$. A (directed) join tree is formed by the cliques $\mathbf{C}_1 = \{X, Y, Z\}$, $\mathbf{C}_2 = \{X, Y, U\}$, $\mathbf{C}_3 = \{Y, Z, V\}$, and $\mathbf{C}_4 = \{X, Z, W\}$, the clique $\mathbf{C}_1$ being the parent of $\mathbf{C}_2$, $\mathbf{C}_3$ and $\mathbf{C}_4$. Ordering the vertices as in $\{X, Y, Z, U, V, W\}$ we have a DAG D^* on $G_0 = G^*$ like D but X being parent of Y and Z, and Y parent of Z. From it we get D, a Bayesian network on $G_0 = G^*$, L_D being the corresponding hypergraph. Besides, the Bayesian networks obtained for other orderings of the variables do not simplify the connections in L_D, therefore L_D is the intersection hypergraph of all the Bayesian networks. The hypergraph L_D is more simple than L_{G_0}, that is $L_D \subset L_{G_0}$. And $G(L_D) = G_0$. The structure of the HOBM would be given by L_D.

VII. Conclusion

In this paper we have studied the determination of the structure of the HOBM. We have started observing that if a factorization of the probability function to learn is known, this function can be written through a consensus function whose terms correspond with a determined hypergraph of connections. If various factorizations are known, these terms correspond with the intersection of the respective hypergraphs. Assuming that an expert provides conditional independences of the probability distribution to learn, factorizations are obtained from these conditional independences, and from the factorizations we determine the structure of the HOBM. Next, the learning algorithm of the HOBM can be applied to a set of samples from the distribution to learn, obtaining the approximation distribution that best fits the sample set.

We have considered independence maps, Markov and Bayesian networks that represent conditional independences through undirected graphs and directed acyclic graphs respectively. We have got from these independence maps the hypergraphs that determine the structure of the HOBM.

Given a positive distribution $P(\mathbf{u})$ its Markov network is unique, so the corresponding hypergraph too. In order to define the Bayesian network we have to give an ordering of the variables in $\mathbf{U}$. Therefore we have different Bayesian networks and different hypergraphs. We have established a link between these hypergraphs. We have proved that the intersection hypergraph of all the Bayesian networks is contained by the hypergraph of the Markov network, i.e., it is more simple. The number of Bayesian networks is nonpolynomical ($N!$) in the number of variables, and if the computation of the $N!$ Bayesian networks is not feasible the procedure described in Section VI for obtaining the Bayesian networks $D_1, \cdots, D_q, D$ permits us to define a subset of Bayesian networks whose intersection hypergraph is contained in the hypergraph of the Markov network. Finally, we have proved that although we can get from Bayesian networks a hypergraph more simple than the hypergraph of the Markov network, the Markov network

[8] Two vertices are adjacent in the hypergraph if there is a connection of some order between them.

establishes a minimum of connectivity for the hypergraphs of Bayesian networks.

In practice, the choice between the Bayesian construction and the Markov construction depends on each problem. The intersection hypergraph of the $N!$ Bayesian networks is the minimal one but, if the computation of all the Bayesian networks is not feasible we can determine an adequate subset with the procedure described in our work, or we can simply use the hypergraph of the Markov network. The complexity of the hypergraph provided by each method depends on the problem.

In our study several questions remain open for further investigation. The set of Bayesian networks $D_1, \cdots, D_q, D$ provided by the procedure of Section VI is not unique. A line of research is to get a procedure that obtains a subset of Bayesian networks such that the intersection hypergraph is the minimal one. Another line of research is to establish the conditions for which the hypergraph given by the procedure described in Section VI coincides with the intersection hypergraph of all the $N!$ Bayesian networks, the minimal hypergraph.

APPENDIX A
FACTORIZATION HYPERGRAPHS

The hypergraph (3) is obtained from this lemma (proved, e.g., in [11]).

Lemma 7 (Möbius Inversion): Let V and Φ be real functions defined on the set of all subsets of a finite set $\mathbf{U}$. Then the following statements are equivalent:

1) $\forall \mathbf{A} \subseteq \mathbf{U} : V(\mathbf{A}) = \sum_{\mathrm{B} \subseteq \mathrm{A}} \Phi(\mathbf{B})$

2) $\forall \mathbf{B} \subseteq \mathbf{U} : \Phi(\mathbf{B}) = \sum_{\mathrm{A} \subseteq \mathrm{B}} (-1)^{|\mathrm{B}-\mathrm{A}|} V(\mathbf{A})$

Given a positive distribution $P(\mathbf{u})$ on $\{0,1\}^N$ and defining for $\lambda \in \mathcal{P}^*([1,N])$

$$w_\lambda = \sum_{\mathbf{u} \in D(\lambda)} (-1)^{|\lambda|-|\mathbf{u}|} \ln \frac{P(\mathbf{u})}{P(\mathbf{0})} \tag{7}$$

where $D(\lambda) = \{\mathbf{u}/j \notin \lambda \Rightarrow x_j = 0\}$ and $|\mathbf{u}|$ is the number of variables with value 1, we have from Lemma 7 that

$$\ln \frac{P(\mathbf{u})}{P(\mathbf{0})} = \sum_{\lambda \in \mathcal{P}^*([1,N])} w_\lambda a(\lambda \mid \mathbf{u})$$

that is $P(\mathbf{u}) = Z^{-1} \exp C(\mathbf{u})$ where $C(\mathbf{u}) = \sum_\lambda w_\lambda a(\lambda \mid \mathbf{u})$ and $Z = P(\mathbf{0})^{-1} = \sum_{\mathbf{u}} \exp C(\mathbf{u})$. Therefore any positive distribution can be written as the normalized exponential of a consensus function. Conversely, if a distribution is the normalized exponential of a consensus function then the connection weights are given by (7), according to Lemma 7.

If $P(\mathbf{u})$ admits the factorization (2), by Lemma 7, defining for $\lambda \subseteq I(\mathbf{U}_i)$

$$w_\lambda^i = \sum_{\mathbf{u}_i \in D_i(\lambda)} (-1)^{|\lambda|-|\mathbf{u}_i|} \ln Q_i(\mathbf{u}_i)$$

where $D_i(\lambda) = \{\mathbf{u}_i/j \notin \lambda \Rightarrow x_j = 0, \forall X_j \in \mathbf{U}_i\}$ and $|\mathbf{u}_i|$ is the number of variables in $\mathbf{U}_i$ with value one, we have

$$\ln Q_i(\mathbf{u}_i) = \sum_{\lambda \subseteq I(\mathrm{U}_i)} w_\lambda^i a(\lambda \mid \mathbf{u}_i)$$

being $a(\lambda \mid \mathbf{u}_i) = \prod_{j \in \lambda} x_j$ (1 if $\lambda = \emptyset$). Then $P(\mathbf{u}) = Z^{-1} \exp C(\mathbf{u})$, where $Z = P(\mathbf{0})^{-1} = \sum_{\mathbf{u}} \exp C(\mathbf{u})$ and $C(\mathbf{u}) = \sum_{\lambda \in L} w_\lambda a(\lambda \mid \mathbf{u})$, where the connection weights are

$$w_\lambda = \sum_{i \in \{1,\cdots,m\}/\lambda \subseteq I(\mathrm{U}_i)} w_\lambda^i$$

and L is given by (3).

APPENDIX B
BAYESIAN NETWORKS ON CHORDAL INDEPENDENCE MAPS

We prove now Proposition 2 and Theorem 3. First we establish the following lemma.

Lemma 8: If D^* is a DAG on a chordal graph G^* consistent with a join tree T of G^*, then every pair of arrows incident on a vertex emanate from two adjacent vertices.

Proof: We consider the ordering of vertices of T and G^* used to build D^*. If two arrows incident on γ emanate from α and β, the indexes of α and β are smaller than the index of γ. Let $\mathbf{C}_i$ be the first clique where both α and γ appear. If $\beta \in \mathbf{C}_i$ then α and β are adjacent. Suppose $\beta \notin \mathbf{C}_i$. Then $i \neq 1$ since the index of β is smaller than the index of γ. Let $\mathbf{U}_1$ and $\mathbf{U}_2$ be, respectively, the union of the cliques of the two subtrees of T obtained removing the edge between $\mathbf{C}_{f(i)}$ and $\mathbf{C}_i$, being $\mathbf{C}_{f(i)} \subseteq \mathbf{U}_1$ and $\mathbf{C}_i \subseteq \mathbf{U}_2$. β belongs to some clique with smaller index than i, then $\beta \in \mathbf{U}_1$. From Theorem 1 we have that $\mathbf{U}_1 \cap \mathbf{U}_2 = \mathbf{C}_{f(i)} \cap \mathbf{C}_i$, then $\beta \notin \mathbf{U}_2$, therefore as β and γ are adjacent, $\gamma \in \mathbf{U}_1$, $\gamma \in \mathbf{C}_{f(i)}$. Since the index of α is smaller than the index of γ, $\alpha \in \mathbf{U}_1$, $\alpha \in \mathbf{U}_1 \cap \mathbf{U}_2$, $\alpha \in \mathbf{C}_{f(i)}$, $\alpha, \gamma \in \mathbf{C}_{f(i)}$ being contradictory to the definition of $\mathbf{C}_i$. Then α and β are adjacent. $\square$

We prove Proposition 2: If D^* is a DAG on a chordal I-map G^* *of* $P(\mathbf{u})$ consistent with a join tree T of G^*, then D^* is an I-map of $P(\mathbf{u})$.

Proof: Suppose that $\langle \mathbf{X} \mid \mathbf{Z} \mid \mathbf{Y} \rangle_{D^*}$ and that $\alpha \in \mathbf{X}$ and $\beta \in \mathbf{Y}$. Let $\mu(\alpha, \beta)$ be a path between α and β. If there is no vertices of $\mathbf{Z}$ in $\mu(\alpha, \beta)$, since it is blocked, there are vertices in $\mu(\alpha, \beta)$ with two incident arrows. From Lemma 8, removing the vertices with two incident vertices of $\mu(\alpha, \beta)$, we get a shorter path $\mu'(\alpha, \beta)$. $\mu'(\alpha, \beta)$ has no vertices of $\mathbf{Z}$ and it is blocked. Repeating the process, we will conclude that there is an arrow between α and β, against the hypothesis. Therefore, there exists some vertex of $\mathbf{Z}$ in $\mu(\alpha, \beta)$. Then $\langle \mathbf{X} \mid \mathbf{Z} \mid \mathbf{Y} \rangle_{G^*}$, so $\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z}$. $\square$

Finally, we prove Theorem 3: Let $P(\mathbf{u})$ be a positive distribution on $\{0,1\}^N$. Given a chordal I-map G^*, the hypergraph of any Bayesian network on G^*consistent with a join tree T of G^* is contained in the hypergraph of G^*.

Proof: Let D^* be a DAG on G^* consistent with T and D a Bayesian network consistent with T, $\mathbf{F}_{X_i} \subseteq \mathbf{F}^*_{X_i}$. Given a vertex X_i, let $\mathbf{C}_k$ be the first clique where it appears. If $k = 1$, $\{X_i\} \cup \mathbf{F}^*_{X_i} \subseteq \mathbf{C}_k$. Suppose $k \neq 1$. $X_i \notin \mathbf{C}_{f(k)}$. Let $\mathbf{U}_1$ and $\mathbf{U}_2$ be the unions of the cliques of the two subtrees of

T obtained eliminating the edge between $\mathbf{C}_{f(k)}$ and $\mathbf{C}_k$, being $\mathbf{C}_{f(k)} \subseteq \mathbf{U}_1$ and $\mathbf{C}_k \subseteq \mathbf{U}_2$. From Theorem 1 we have that $\mathbf{U}_1 \cap \mathbf{U}_2 = \mathbf{C}_{f(k)} \cap \mathbf{C}_k$. For $X_j \in \mathbf{F}^*_{X_i}$, suppose $X_j \notin \mathbf{C}_k$. So $X_j \in \mathbf{U}_1$ and $X_j \notin \mathbf{U}_2$. As X_j and X_i are adjacent $X_i \in \mathbf{U}_1$, then $X_i \in \mathbf{U}_1 \cap \mathbf{U}_2$, $X_i \in \mathbf{C}_{f(k)}$, this being contradictory to the definition of $\mathbf{C}_k$. Therefore $X_j \in \mathbf{C}_k$. So $\{X_i\} \cup \mathbf{F}^*_{X_i} \subseteq \mathbf{C}_k$. Therefore the hypergraph of the Bayesian network D is contained by the hypergraph of G^*. □

References

[1] D. H. Ackley, G. E. Hinton, and T. J. Sejnowski, "A learning algorithm for Boltzmann machines," *Cognitive Sci.*, vol. 9, pp. 147–169, 1985.

[2] F. X. Albizuri, A. d'Anjou, M. Grana, and J. A. Lozano, "Convergence properties of high-order Boltzmann machines," *Neural Networks*, vol. 9, no, 9, pp. 1561–1567, 1996.

[3] F. X. Albizuri, A. d'Anjou, M. Grana, F. J. Torrealdea, and M. C. Hernandez, "The high-order Boltzmann machine: Learned distribution and topology," *IEEE Trans. Neural Networks*, vol. 6, pp. 767–770, 1995.

[4] C. Beeri, R. Fagin, D. Maier, and M. Yannakakis, "On the desirability of acyclic database schemes," *J. Assoc. Computing Machinery*, vol. 3, no. 3, pp. 479–513, 1983.

[5] J. N. Darroch, S. L. Lauritzen, and T. P. Speed, "Markov fields and log linear interaction models for contingency tables," *Ann. Statist.*, vol. 8, pp. 522–539, 1980.

[6] H. Geffner and J. Pearl, "On the probabilistic semantics of connectionist networks," in *Proc. 1st IEEE Int. Conf. Neural Networks*, San Diego, CA, 1987, pp. 187–195.

[7] J. M. Hammersley and P. Clifford, "Markov fields on finite graphs and lattices," Unpublished manuscript cited in V. Isham 1981 "An introduction to spatial point processes and Markov random fields," *Int. Statist. Rev.*, vol. 49, pp. 21–43, 1971.

[8] F. Harary, *Graph Theory*. Reading, MA: Addison-Wesley, 1969.

[9] G. E. Hinton and T. J. Sejnowski, "Learning and relearning in Boltzmann machines," in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, D. E. Rumelhart and J. L. McClelland, Eds., vol. 1. Cambridge, MA: MIT Press, 1986, pp. 282–317.

[10] T. Hrycej, "Common features of neural-network models of high- and low-level human information processing," in *Artificial Neural Networks*, T. Kohonen, K. Makisara, O. Simula, and J. Kangas, Eds. Amsterdam, The Netherlands: Elsevier, 1991, pp. 861–866.

[11] S. L. Lauritzen, *Graphical Models*. Oxford, U.K.: Clarendon Press, 1996.

[12] P. Myllymaki, "Bayesian reasoning by stochastic neural networks," Dept. Computer Sci., Univ. Helsinki, Finland, Tech. Rep. C-67, 1993.

[13] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA: Morgan Kaufmann, 1988.

[14] J. Pearl and A. Paz, "Graphoids: A graph-based logic for reasoning about relevance relations," Cognitive Syst. Lab., Univ. California, Los Angeles, CA, Tech. Rep. 850038 (R-53-L), 1985.

[15] T. J. Sejnowski, "Higher-order Boltzmann machines," in *AIP Cof. Proc. 151, Neural Networks for Computing*, J. S. Denker, Ed., Snowbird, UT, 1986, pp. 398–403.

[16] R. E. Tarjan and M. Yannakakis, "Simple linear-time algorithms to test chordality of graphs, test acyclicity of hypergraphs, and selectively reduce acyclic hypergraphs," *SIAM J. Computing*, vol. 13, no. 3, pp. 566–579, 1984.

[17] T. S. Verma, "Causal networks: Semantics and expressiveness," Cognitive Syst. Lab., Univ. California, Los Angeles, CA, Tech. Rep. R-65, 1986.

[18] N. Wermuth and S. L. Lauritzen, "Graphical and recursive models for contingency tables," *Biometrika*, vol. 70, pp. 537–552, 1983.

[19] J. Whittaker, *Graphical Models in Applied Multivariate Statistics*. New York: Wiley, 1990.

F. Xabier Albizuri received the M.Sc. degree in physics in 1987 and the Ph.D. degree in computer science in 1995 from the University of the Basque Country, Donositia, Spain.

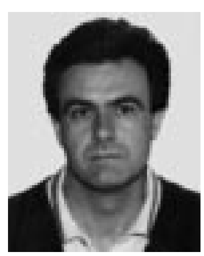
He is a Professor of the Department of Computer Science and Artificial Intelligence at the University of the Basque Country. His current research interest is on neural networks, pattern recognition, and graphical models.

Alicia d'Anjou received the M.Sc. and the Ph.D. degrees in physics from the University of Navarra, Spain, in 1974 and 1978, respectively.

She is a Professor of the Department of Computer Science and Artificial Intelligence at the University of the Basque Country, Donositia, Spain. Her current interests include adaptive algorithms and neural networks.

Manuel Graña (M'94) received the M.Sc. and Ph.D. degrees in computer science from the University of the Basque Country, Donositia, Spain, in 1982 and 1989, respectively.

He is a Professor of the Department of Computer Science and Artificial Intelligence at the University of the Basque Country. His current interests include neural networks for digital signal processing.



Pedro Larrañaga (M'97) received the M.Sc. degree in mathematics (mentions: statistics) from the University of Valladolid, Spain, in 1981, and the Ph.D. degree in computer science from the University of the Basque Country, Donositia, Spain, in 1995.

He is a Professor of statistics and artificial intelligence at the Department of Computer Science and Artificial Intelligence of the University of the Basque Country. His current research interests are in the fields of Bayesian networks, combinatorial optimization and data analysis with applications to medicine, molecular biology, and finance.