

On a Class of Tensor Markov Fields

Enrique Hernández-Lemus ^{1,2} 

¹ Computational Genomics Division, National Institute of Genomic Medicine, 14610 Mexico City, Mexico; ehernandez@inmegen.gob.mx; Tel.: +52-55-5350-1970

² Centro de Ciencias de la Complejidad, Universidad Nacional Autónoma de México, 04510 Mexico City, Mexico

Received: 6 March 2020; Accepted: 9 April 2020; Published: 16 April 2020

Abstract: Here, we introduce a class of Tensor Markov Fields intended as probabilistic graphical models from random variables spanned over multiplexed contexts. These fields are an extension of Markov Random Fields for tensor-valued random variables. By extending the results of Dobruschin, Hammersley and Clifford to such tensor valued fields, we proved that tensor Markov fields are indeed Gibbs fields, whenever strictly positive probability measures are considered. Hence, there is a direct relationship with many results from theoretical statistical mechanics. We showed how this class of Markov fields it can be built based on a statistical dependency structures inferred on information theoretical grounds over empirical data. Thus, aside from purely theoretical interest, the Tensor Markov Fields described here may be useful for mathematical modeling and data analysis due to their intrinsic simplicity and generality.

Keywords: Markov random fields; probabilistic graphical models; multilayer networks

1. General Definitions

Here, we introduce Tensor Markov Fields, i.e., Markov random fields [1,2] over tensor spaces. Tensor Markov Fields (TMFs) represent the joint probability distribution for a set of tensor-valued random variables.

Let $X = X_{\alpha}^{\beta}$ be one of such tensor-valued random variables. Here $X_i^j \in X$ may represent either a variable $i \in \alpha$, that may exist in a given context or layer $j \in \beta$ (giving rise to a class of so-called multilayer graphical models or multilayer networks) or a single tensor-valued quantity X_i^j . A TMF will be an undirected multilayer graph representing the statistical dependency structure of X as given by the joint probability distribution $\mathbb{P}(X)$.

As an extension of the case of Markov random fields, a TMF is a multilayer graph $\hat{G} = (V, E)$ formed by a set V of vertices or nodes (the X_i^j 's) and a set $E \subseteq V \times V$ of edges connecting the nodes, either on the same *layer* or through different layers (Figure 1). The set of edges represents a neighborhood law N stating which vertex is connected (dependent) to which other vertex in the multilayer graph. With this in mind, a TMF can be also represented (slightly abusing notation) as $\hat{G} = (V, N)$. The set of neighbors of a given point X_i^j will be denoted $N_{X_i^j}$.

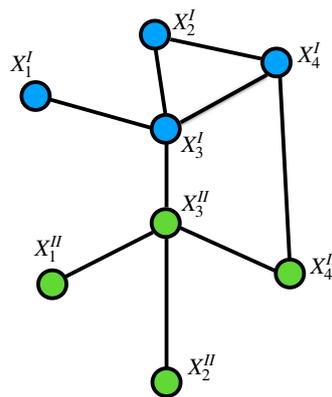


Figure 1. A Tensor Markov Field: represented as a multilayer graph spanning over X_i^j with $i = \{1, 2, 3, 4\}$ and $j = \{I, II\}$. To illustrate, layer I is colored in blue and layer II is colored green.

1.1. Configuration

It is possible to assign to each point in the multilayer graph, one of a finite set S of labels. Such assignment will be called a *configuration*. We will assign probability measures to the set Ω of all possible configurations ω . Hence, ω_A represents the configuration ω restricted to the subset A of V . It is possible to think of ω_A as a configuration on the smaller multilayer graph \hat{G}_A restricting V to points of A (Figure 2).

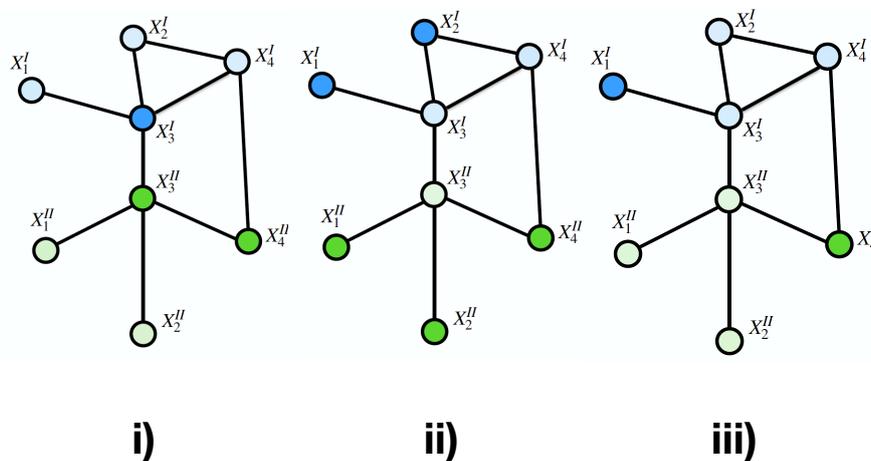


Figure 2. Three different configurations of a Tensor Markov Field panels (i), (ii) and (iii) present different configurations or *states* of the TMF. Labels are represented by color intensity.

1.2. Local Characteristics

It is also possible to extend the notion of *local characteristics* from MRFs. The local characteristics of a probability measure \mathbb{P} defined on Ω are the conditional probabilities of the form:

$$\mathbb{P}(\omega_t | \omega_{T \setminus t}) = \mathbb{P}(\omega_t | \omega_{N_t}) \tag{1}$$

i.e., the probability that the point t is assigned the value ω_t , given the values at all other points of the multilayer graph. In order to make explicit the tensorial nature of the multilayer graph \hat{G} , let us re-write Equation (1). Let us also recall the fact that the probability measure will define a tensor Markov random field (a TMF) if the local characteristics depend only of the knowledge of the outcomes at neighboring points, i.e., if for every ω

$$\mathbb{P}(\omega_{X_i^j} | \omega_{\hat{G} \setminus X_i^j}) = \mathbb{P}(\omega_{X_i^j} | \omega_{N_{X_i^j}}) \tag{2}$$

1.3. Cliques

Given an arbitrary graph (or in the present case a multilayer graph), we shall say that a set of points C is a *clique* if every pair of points in C are neighbors (see Figure 3). This definition includes the empty set as a clique. A clique is thus a set whose *induced subgraph* is complete, for this reason cliques are also called *complete induced subgraphs* or *maximal subgraphs* (although these latter term may be ambiguous).

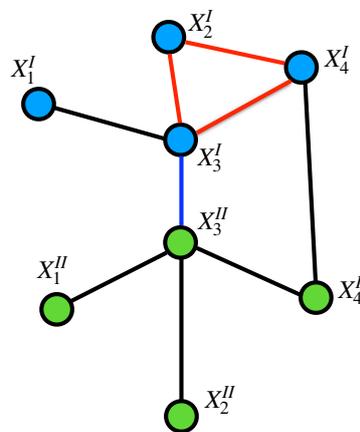


Figure 3. Cliques on a Tensor Markov Field: The set $\{X_2^I, X_3^I, X_4^I\}$ forms an intra-layer 2-clique (as marked by the red edges, all on layer I), the set $\{X_3^I, X_3^{II}\}$ forms an inter-layer 1-clique (marked by the blue edge connecting layers I and II). However, the set $\{X_3^I, X_3^{II}, X_4^I, X_4^{II}\}$ is not a clique since there are no edges between X_3^I and X_4^{II} nor between X_3^{II} and X_4^I .

1.4. Configuration Potentials

A *potential* η is a way to assign a number $\eta_A(\omega)$ to every subconfiguration ω_A of a configuration ω in the multilayer graph \hat{G} . Given a potential, we shall say that it defines (or better, induces) a *dimensionless energy* $U(\omega)$ on the set of all configurations ω by

$$U(\omega) = \sum_A \eta_A(\omega) \tag{3}$$

In the preceding expression, for fixed ω , the sum is taken over all subsets $A \subseteq V$ including the empty set. We can define a probability measure, called the *Gibbs measure induced by U* as

$$\mathbb{P}(\omega) = \frac{e^{-U(\omega)}}{Z} \tag{4}$$

with Z a normalization constant called the *partition function*.

$$Z = \sum_{\omega} e^{-U(\omega)} \tag{5}$$

In physics, the term *potential* is often used in connection with the so-called potential energies. Physicists often call η_A a *dimensionless potential energy*, and they call $\phi_A = e^{-\eta_A}$ a potential.

Equations (4) and (5) can be thus rewritten as:

$$\mathbb{P}(\omega) = \frac{\prod_A \phi_A(\omega)}{Z} \tag{6}$$

$$Z = \sum_{\omega} \prod_A \phi_A(\omega) \tag{7}$$

Since this latter use is more common in probability and graph theory, we will refer to Equations (6) and (7) as the definitions of Gibbs measure and partition function (respectively) unless otherwise stated.

1.5. Gibbs Fields

A potential is called a nearest neighbor Gibbs potential if $\phi_A(\omega) = 1$ whenever A is not a clique. It is customary to refer as a *Gibbs measure* to a measure induced by a nearest neighbor Gibbs potential. However, it is possible to define more general Gibbs measures by considering other types of potentials.

The inclusion of all cliques in the calculation of the Gibbs measure is necessary to establish the equivalence between Gibbs random fields and Markov random fields. Let us see how a nearest neighbor Gibbs measure on a multilayer graph determines a TMF.

Let $\mathbb{P}(\omega)$ be a probability measure determined on Ω by a nearest neighbor Gibbs potential ϕ :

$$\mathbb{P}(\omega) = \frac{\prod_C \phi_C(\omega)}{Z} \tag{8}$$

With the product taken over all cliques C on the multilayer graph \hat{G} . Then,

$$\mathbb{P}(\omega_{X_i^j} | \omega_{\hat{G} \setminus X_i^j}) = \frac{\mathbb{P}(\omega)}{\sum_{\omega'} \mathbb{P}(\omega')} \tag{9}$$

Here ω' is any configuration which agrees with ω at all points except X_i^j .

$$\mathbb{P}(\omega_{X_i^j} | \omega_{\hat{G} \setminus X_i^j}) = \frac{\prod_C \phi_C(\omega)}{\sum_{\omega'} \prod_C \phi_C(\omega')} \tag{10}$$

For any clique C that does not contain X_i^j , $\phi_C(\omega) = \phi_C(\omega')$, So that all the terms that correspond to cliques that do not contain the point X_i^j cancel both from the numerator and the denominator in Equation (10), therefore this probability depends only on the values x_i^j at X_i^j and its neighbors. \mathbb{P} defines thus a TMF.

A more general proof of this equivalence is given by Hammersley-Clifford theorem that will be presented in the following section.

2. Extended Hammersley Clifford Theorem

Here we will outline a proof for an extension of Hammersley-Clifford theorem for Tensor Markov Fields (i.e., we will show that a Tensor Markov Field is equivalent to a Tensor Gibbs Field).

Let $\hat{G} = (V, N)$ be a multilayer graph representing a TMF as defined in the previous section. With $V = X_{\alpha}^{\beta} = \{X_i^j\}$, a set of vertices over a tensor field and N a neighborhood law that connects vertices

over this tensor field. The field \hat{G} obeys the following neighborhood law given its Markov property (see Equation (2))

$$\mathbb{P}(X_i^j | X_{\hat{G} \setminus X_i^j}) = \mathbb{P}(X_i^j | X_{N_i^j}) \tag{11}$$

Here $X_{N_i^j}$ is any neighbor of X_i^j . The Hammersley-Clifford theorem states that a MRF is also a local Gibbs field. In the case of a TMF we have the following expression:

$$\mathbb{P}(X) = \frac{1}{Z} \prod_{c \in \mathcal{C}_{\hat{G}}} \phi_c(X_c) \tag{12}$$

In order to prove the equivalence of Equations (11) and (12), we will first built a deductive (backward direction) part of the proof to be complemented with a constructive (forward direction) part as presented in the following subsections.

2.1. Backward Direction

Let us consider Equation (11) at the light of Bayes' theorem:

$$\mathbb{P}(X_i^j | X_{\hat{G} \setminus X_i^j}) = \frac{\mathbb{P}(X_i^j, X_{N_i^j})}{\mathbb{P}(X_{N_i^j})} \tag{13}$$

Using a clique-approach to calculate the joint and marginal probabilities (see next subsection to support the following statement):

$$\mathbb{P}(X_i^j | X_{\hat{G} \setminus X_i^j}) = \frac{\sum_{\hat{G} \setminus D_i^j} \prod_{c \in \mathcal{C}_{\hat{G}}} \phi_c(X_c)}{\sum_{X_i^j} \sum_{\hat{G} \setminus D_i^j} \prod_{c \in \mathcal{C}_{\hat{G}}} \phi_c(X_c)} \tag{14}$$

Let us split the product $\prod_{c \in \mathcal{C}_{\hat{G}}} \phi_c(X_c)$ into two products, one over the set of cliques that contain X_i^j (let us call it C_i^j) and another set formed by cliques not containing X_i^j (let us call it R_i^j):

$$\mathbb{P}(X_i^j | X_{\hat{G} \setminus X_i^j}) = \frac{\sum_{\hat{G} \setminus D_i^j} \prod_{c \in C_i^j} \phi_c(X_c) \prod_{c \in R_i^j} \phi_c(X_c)}{\sum_{X_i^j} \sum_{\hat{G} \setminus D_i^j} \prod_{c \in C_i^j} \phi_c(X_c) \prod_{c \in R_i^j} \phi_c(X_c)} \tag{15}$$

Factoring out the terms depending on X_i^j (that do not contribute to cliques in the domain $\hat{G} \setminus X_i^j$):

$$\mathbb{P}(X_i^j | X_{\hat{G} \setminus X_i^j}) = \frac{\prod_{c \in C_i^j} \phi_c(X_c) \sum_{\hat{G} \setminus D_i^j} \prod_{c \in R_i^j} \phi_c(X_c)}{\sum_{X_i^j} \prod_{c \in C_i^j} \phi_c(X_c) \sum_{\hat{G} \setminus D_i^j} \prod_{c \in R_i^j} \phi_c(X_c)} \tag{16}$$

The term $\sum_{\hat{G} \setminus D_i^j} \prod_{c \in R_i^j} \phi_c(X_c)$ does not involve X_i^j (by construction) so, it can be factored out from the summation over X_i^j in the denominator.

$$\mathbb{P}(X_i^j | X_{\hat{G} \setminus X_i^j}) = \frac{\prod_{c \in C_i^j} \phi_c(X_c) \sum_{\hat{G} \setminus D_i^j} \prod_{c \in R_i^j} \phi_c(X_c)}{\sum_{\hat{G} \setminus D_i^j} \prod_{c \in R_i^j} \phi_c(X_c) \sum_{X_i^j} \prod_{c \in C_i^j} \phi_c(X_c)} \tag{17}$$

We can cancel the term in the numerator and denominator:

$$\mathbb{P}(X_i^j | X_{\hat{G} \setminus X_i^j}) = \frac{\prod_{c \in C_i^j} \phi_c(X_c)}{\sum_{X_i^j} \prod_{c \in C_i^j} \phi_c(X_c)} \tag{18}$$

Then we multiply by $\frac{\prod_{c \in R_i^j} \phi_c(X_c)}{\prod_{c \in R_i^j} \phi_c(X_c)}$

$$\mathbb{P}(X_i^j | X_{\hat{G} \setminus X_i^j}) = \frac{\prod_{c \in C_i^j} \phi_c(X_c) \prod_{c \in R_i^j} \phi_c(X_c)}{\sum_{X_i^j} \prod_{c \in C_i^j} \phi_c(X_c) \prod_{c \in R_i^j} \phi_c(X_c)} \tag{19}$$

Remembering that $C_i^j \cup R_i^j = C_{\hat{G}}$,

$$\mathbb{P}(X_i^j | X_{\hat{G} \setminus X_i^j}) = \frac{\prod_{c \in \hat{G}} \phi_c(X_c)}{\sum_{X_i^j} \prod_{c \in \hat{G}} \phi_c(X_c)} \tag{20}$$

Equation (20) is nothing but the definition of a local Gibbs Tensor Field (Equation (12)).

2.2. Forward Direction

In this subsection we will show how to express the clique potential functions $\phi_c(X_c)$, given the joint probability distribution over the tensor field and the Markov property.

Consider any subset $\sigma \subset \hat{G}$ of the multilayer graph \hat{G} . We define a candidate potential function (following Möbius inversion lemma) [3] as follows:

$$f_\sigma(X_\sigma = x_\sigma) = \prod_{\zeta \subset \sigma} \mathbb{P}(X_\zeta = x_\zeta, X_{\hat{G} \setminus \zeta} = 0)^{-1^{|\sigma| - |\zeta|}} \tag{21}$$

In order for f_σ to be a proper clique potential, it must satisfy the following two conditions:

- (i) $\prod_{\sigma \subset \hat{G}} f_\sigma(X_\sigma) = \mathbb{P}(X)$
- (ii) $f_\sigma(X_\sigma) = 1$ whenever σ is not a clique

To prove (i), we need to show that all factors in $f_\sigma(X_\sigma = x_\sigma)$ cancel out, except for $\mathbb{P}(X)$.

To do this, it will be useful to consider the following *combinatorial expansion of zero*:

$$0 = (1 - 1)^K = C_0^K - C_1^K + C_2^K + \dots + (-1)^K C_K^K \tag{22}$$

Here, of course C_B^A is the number of combinations of B elements from an A-element set.

Let us consider any subset ζ of \hat{G} . Let us consider a factor $\Delta = \mathbb{P}(X_\zeta = x_\zeta, X_{\hat{G} \setminus \zeta} = 0)$. For the case of $f_\zeta(X_\zeta)$ it occurs as $\Delta^{-1^0} = \Delta$. Such factor also occurs in subsets containing ζ and other additional elements. If it includes ζ and one additional element, there are $C_1^{|\hat{G}| - |\zeta|}$ such functions. The additional element creates an inverse factor $\Delta^{-1^1} = \Delta^{-1}$. The functions over subsets containing ζ and two additional elements contributes with a factor $\Delta^{-1^2} = \Delta^1 = \Delta$. If we continue this process and consider Equation (22), it is evident that all odd cardinality difference terms cancel out with all even cardinality difference terms so that the only remaining factor corresponds to $\zeta = \hat{G}$ equal to $\mathbb{P}(X)$ thus fulfilling condition (i).

In order to show how condition (ii) is fulfilled, we will need to use the Markov property of TMFs. Let us consider $\sigma^* \subset \hat{G}$ that is not a clique. Then it will be possible to find two nodes X_i^h and X_j^k in σ^* that are not connected to each other. Let us recall Equation (21):

$$f_{\sigma^*}(X_{\sigma^*} = x_{\sigma^*}) = \prod_{\zeta \subset \sigma^*} \mathbb{P}(X_\zeta = x_\zeta, X_{\hat{G} \setminus \zeta} = 0)^{-1^{|\sigma^*| - |\zeta|}} \tag{23}$$

An arbitrary subset ζ may belong to any of the following classes: (i) $\zeta = \omega$ a generic subset of σ ; (ii) $\zeta = \omega \cup \{X_i^h\}$; (iii) $\zeta = \omega \cup \{X_j^k\}$ or iv) $\zeta = \omega \cup \{X_i^h, X_j^k\}$. If we write down Equation (23) factored down to these contributions we get:

$$f_{\sigma}(X_{\sigma^*} = x_{\sigma^*}) = \prod_{\omega \subset \sigma^* \setminus \{X_i^h, X_j^k\}} \left[\frac{\mathbb{P}(X_{\omega}, X_{\hat{G} \setminus \omega} = 0) \mathbb{P}(X_{\omega \cup \{X_i^h, X_j^k\}}, X_{\hat{G} \setminus \omega \cup \{X_i^h, X_j^k\}} = 0)}{\mathbb{P}(X_{\omega \cup \{X_i^h\}}, X_{\hat{G} \setminus \omega \cup \{X_i^h\}} = 0) \mathbb{P}(X_{\omega \cup \{X_j^k\}}, X_{\hat{G} \setminus \omega \cup \{X_j^k\}} = 0)} \right]^{-1^{|\sigma^*| - |\zeta|}} \tag{24}$$

Let us consider two of the factors in Equation (24) at the light of Bayes' theorem:

$$\frac{\mathbb{P}(X_{\omega}, X_{\hat{G} \setminus \omega} = 0)}{\mathbb{P}(X_{\omega \cup \{X_i^h\}}, X_{\hat{G} \setminus \omega \cup \{X_i^h\}} = 0)} = \frac{\mathbb{P}(X_{\{X_i^h\}} = 0 | X_{\{X_j^k\}} = 0, X_{\omega}, X_{\hat{G} \setminus \omega \cup \{X_i^h, X_j^k\}} = 0) \mathbb{P}(X_{\{X_j^k\}} = 0, X_{\omega}, X_{\hat{G} \setminus \omega \cup \{X_i^h, X_j^k\}} = 0)}{\mathbb{P}(X_{\{X_i^h\}} | X_{\{X_j^k\}} = 0, X_{\omega}, X_{\hat{G} \setminus \omega \cup \{X_i^h, X_j^k\}} = 0) \mathbb{P}(X_{\{X_j^k\}} = 0, X_{\omega}, X_{\hat{G} \setminus \omega \cup \{X_i^h, X_j^k\}} = 0)} \tag{25}$$

We can notice that the priors in the numerator and denominator of Equation (25) are the same. We can then cancel them out. Since by definition X_i^h and X_j^k are conditionally independent given the rest of the multilayer graph, we can also replace the default value $X_j^k = 0$ for X_j^k instead.

$$\frac{\mathbb{P}(X_{\omega}, X_{\hat{G} \setminus \omega} = 0)}{\mathbb{P}(X_{\omega \cup \{X_i^h\}}, X_{\hat{G} \setminus \omega \cup \{X_i^h\}} = 0)} = \frac{\mathbb{P}(X_{\{X_i^h\}} = 0 | X_{\{X_j^k\}}, X_{\omega}, X_{\hat{G} \setminus \omega \cup \{X_i^h, X_j^k\}} = 0) \mathbb{P}(X_{\{X_j^k\}}, X_{\omega}, X_{\hat{G} \setminus \omega \cup \{X_i^h, X_j^k\}} = 0)}{\mathbb{P}(X_{\{X_i^h\}} | X_{\{X_j^k\}}, X_{\omega}, X_{\hat{G} \setminus \omega \cup \{X_i^h, X_j^k\}} = 0) \mathbb{P}(X_{\{X_j^k\}}, X_{\omega}, X_{\hat{G} \setminus \omega \cup \{X_i^h, X_j^k\}} = 0)} \tag{26}$$

Since X_i^h and X_j^k are conditionally independent given the rest of the multilayer graph, we can also replace the condition for X_j^k with any other, without affecting X_i^h . By adjusting this prior conveniently, we can write out:

$$\frac{\mathbb{P}(X_{\omega}, X_{\hat{G} \setminus \omega} = 0)}{\mathbb{P}(X_{\omega \cup \{X_i^h\}}, X_{\hat{G} \setminus \omega \cup \{X_i^h\}} = 0)} = \frac{\mathbb{P}(X_{\omega \cup \{X_j^k\}}, X_{\hat{G} \setminus \omega \cup \{X_j^k\}} = 0)}{\mathbb{P}(X_{\omega \cup \{X_i^h, X_j^k\}}, X_{\hat{G} \setminus \omega \cup \{X_i^h, X_j^k\}} = 0)} \tag{27}$$

By substituting Equation (27) into Equation (24) we get (condition (ii)):

$$f_{\sigma^*}(X_{\sigma^*}) = 1 \tag{28}$$

3. An Information-Theoretical Class of Tensor Markov Fields

Let us consider again the set of tensor-valued random variables $X = X_{\alpha}^{\beta}$. It is possible to calculate, for every duplex in X , the mutual information function $I(\cdot, \cdot)$ [4]:

$$I(X_i^h, X_j^k) = \sum_{\Omega} \sum_{\Omega'} p(X_i^h, X_j^k) \log \frac{p(X_i^h, X_j^k)}{p(X_i^h) p(X_j^k)} \tag{29}$$

Let us consider a multilayer graph scenario. From now on, the indices i, j will refer to the random variables, whereas h, k will be indices for the layers. Ω and Ω' are the respective sampling spaces (that may, of course, be equal). In order to discard self-information, let us define the *off-diagonal mutual information* as follows:

$$I^{\dagger}(X_i^h, X_j^k) = I(X_i^h, X_j^k) \times \left(1 - \delta_{X_i^h X_j^k} \right) \tag{30}$$

With the bi-delta function $\delta_{X_i^h X_j^k}$ defined as:

$$\delta_{X_i^h X_j^k} = \begin{cases} 1, & \text{if } i = j \text{ and } h = k \\ 0, & \text{otherwise} \end{cases} \tag{31}$$

By having the complete set of off-diagonal mutual information functions for all the random variables and layers, it is possible to define the following hyper-matrix elements:

$$A_{ij}^{hk} = \Theta \left[I^{\dagger}(X_i^h, X_j^k) - I_0 \right] \tag{32}$$

as well as:

$$W_{ij}^{hk} = A_{ij}^{hk} \circ I^{\dagger}(X_i^h, X_j^k) \tag{33}$$

Here $\Theta[\cdot]$ is Heavyside’s function and I_0 is a lower bound for mutual information (a threshold) to be considered *significant*.

We call A_{ij}^{hk} and W_{ij}^{hk} the *adjacency hypermatrix* and the *strength hypermatrix* respectively (notice that \circ in Equation (33) represents the product of a scalar times a hypermatrix). The adjacency hyper-matrix and the strength hyper-matrix define the (unweighted and weighted, respectively) neighborhood law of the associated TMF, hence the statistical dependency structure for the set of random variables and contexts (layers).

Although the adjacency and strength hypermatrices are indeed, proper representations of the undirected (unweighted and weighted) dependency structure of $\mathbb{P}(X)$, it has been considered advantageous to embed them into a tensor linear structure, in order to be able to work out some of the mathematical properties of such fields relying on the methods of tensor algebra. One relevant proposal in this regard, has been advanced by De Domenico and collaborators, in the context of multilayer networks.

Following the ideas of De Domenico and co-workers [5], we introduce the unweighted and weighted adjacency 4-tensors (respectively) as follows:

$$\mathbb{A} = \sum_{h,k=1}^L \sum_{i,j=1}^N A_{ij}^{hk} \otimes \zeta_{\beta\delta}^{\alpha\gamma} \tag{34}$$

$$\mathbb{W} = \sum_{h,k=1}^L \sum_{i,j=1}^N W_{ij}^{hk} \otimes \zeta_{\beta\delta}^{\alpha\gamma} \tag{35}$$

Here, $\zeta_{\beta\delta}^{\alpha\gamma} = \zeta_{\beta\delta}^{\alpha\gamma}[ijhk]$ is a unit four-tensor whose role is to provide the hypermatrices with the desired linear properties (projections, contractions, etc.). Square brackets indicate that the indices i, j, h and k belong to the α, β, γ and δ dimensions and \otimes represents a form of a *tensor matricization* product (i.e., the one producing a 4-tensor out of a 4-index hypermatrix times a unitary 4-tensor).

3.1. Conditional Independence in Tensor Markov Fields

In order to discuss the conditional independence structure induced by the present class of TMFs, let us analyze Equation (32). As already mentioned, the hyper-adjacency matrix A_{ij}^{hk} represents the neighborhood law (as given by the Markov property) on the multilayer graph \hat{G} (i.e., the TMF). Every non-zero entry on this hypermatrix represents a statistical dependence relation between two elements on X . The conditional dependence structure on TMFs inferred from mutual information measures via Equation (32) are related not only to the statistical independence conditions (as given by a zero mutual information measure between two elements), but also to the lower bound I_0 and in general to the dependency structure of the whole multilayer graph.

The definition of conditional independence (CI) for tensor random variables is as follows:

$$(X_i^h \perp\!\!\!\perp X_j^k) | X_l^m \iff \mathbb{F}_{X_i^h, X_j^k | X_l^m = X_l^{m*}}(X_i^{h*}, X_j^{k*}) = \mathbb{F}_{X_i^h | X_l^m = X_l^{m*}}(X_i^{h*}) \cdot \mathbb{F}_{X_j^k | X_l^m = X_l^{m*}}(X_j^{k*}) \tag{36}$$

$$\forall X_i^h, X_j^k, X_l^m \in X.$$

Here $\perp\!\!\!\perp$ represents conditional independence between two random variables, were $\mathbb{F}_{X_i^h, X_j^k | X_l^m = X_l^{m*}}(X_i^{h*}, X_j^{k*}) = Pr(X_i^h \leq X_i^{h*}, X_j^k \leq X_j^{k*} | X_l^m = X_l^{m*})$ is the joint conditional cumulative distribution of X_i^h and X_j^k given X_l^m and X_i^{h*}, X_j^{k*} and X_l^{m*} are realization events of the corresponding random variables.

In the case of MRFs (and by extension TMFs), CI is defined by means of (multi)graph separation: in this sense we say that $X_i^h \perp\!\!\!\perp_{\hat{G}} X_j^k | X_l^m$ iff X_l^m separates X_i^h from X_j^k in the multilayer graph \hat{G} . This means that if we remove node X_l^m there are no undirected paths from X_i^h to X_j^k in \hat{G} .

Conditional independence in random fields is often considered in terms of subsets of V . Let A, B and C be three subsets of V . The statement $X_A \perp\!\!\!\perp_{\hat{G}} X_B | X_C$, which holds only iff C separates A from B in the multilayer graph \hat{G} , meaning that if we remove all vertices in C there will be no paths connecting any vertex in A to any vertex in B is called the *global Markov property* of TMFs.

The smallest set of vertices that renders a vertex X_i^h conditionally independent of all other vertices in the multilayer graph is called its *Markov blanket*, denoted $mb(X_i^h)$. If we define the *closure* of a node X_i^h as $\mathcal{C}(X_i^h)$ then $X_i^h \perp\!\!\!\perp_{\hat{G}} \mathcal{C}(X_i^h) | mb(X_i^h)$.

It is possible to show that in a TMF, the Markov blanket of a vertex is its set of first neighbors. This is called the *undirected local Markov property*. Starting from the local Markov property it is possible to show that two vertices X_i^h and X_j^k are conditionally independent given the rest if there is no direct edge between them. This has been called the *pairwise Markov property*.

If we denote by $\hat{G}_{X_i^h \rightarrow X_j^k}$ the set of undirected paths in the multilayer graph \hat{G} connecting vertices X_i^h and X_j^k , then the pairwise Markov property of a TMF can be stated as:

$$X_i^h \perp\!\!\!\perp X_j^k | \hat{G} \setminus \{X_i^h, X_j^k\} \iff \hat{G}_{X_i^h \rightarrow X_j^k} = \emptyset \tag{37}$$

It is clear that the global Markov property implies the local Markov property which in turn implies the pairwise Markov property. For systems with positive definitivity probability densities, it has been probed (in the case of MRFs) that pairwise Markov actually implied global Markov (See [6] p. 119 for a proof). For the present extension this is important since it is easier to assess pairwise conditional independence statements.

3.2. Independence Maps

Let $I_{\hat{G}}$ denote the set of all conditional independence relations encoded by the multilayer graph \hat{G} (i.e., those CI relations given by the Global Markov property). Let $I_{\mathbb{P}}$ be the set of all CI relations implied by the probability distribution $\mathbb{P}(X_i^j)$. A multilayer graph \hat{G} will be called an *independence map (I-map)* for a probability distribution $\mathbb{P}(X_i^j)$, if all CI relations implied by \hat{G} hold for $\mathbb{P}(X_i^j)$, i.e., $I_{\hat{G}} \subseteq I_{\mathbb{P}}$ [6].

The converse statement is not necessarily true, i.e., there may be some CI relations implied by $\mathbb{P}(X_i^j)$ that are not encoded in the multilayer graph \hat{G} . We may be usually interested in *minimal I-maps*, i.e., I-maps from which none of the edges could be removed without destroying its CI properties.

Every distribution has a unique minimal I-map (and a given graph representation). Let $\mathbb{P}(X_i^j) > 0$. Let \hat{G}^{\dagger} be the multilayer graph obtained by introducing edges between all pairs of vertices X_i^h, X_j^k such that $X_i^h \perp\!\!\!\perp X_j^k | X \setminus \{X_i^h, X_j^k\}$, then \hat{G}^{\dagger} is the unique minimal I-map. We call \hat{G} a *perfect map* of \mathbb{P} when there is no dependencies \hat{G} which are not indicated by \mathbb{P} , i.e., $I_{\hat{G}} = I_{\mathbb{P}}$ [6].

3.3. Conditional Independence Tests

Conditional independence tests are useful to evaluate whether CI conditions apply either exactly or in the case of applications under a certain bounded error. In order to be able to write down expressions for C.I. tests let us introduce the following *conditional kernels* [7]:

$$\mathbb{C}_A(B) = \mathbb{P}(B|A) = \frac{\mathbb{P}(AB)}{\mathbb{P}(A)} \tag{38}$$

as well as their generalized recursive relations:

$$\mathbb{C}_{ABC}(D) = \mathbb{C}_{AB}(D|C) = \frac{\mathbb{C}_{AB}(CD)}{\mathbb{C}_{AB}(C)} \tag{39}$$

The conditional probability of X_h^k given X_i^j can be thus written as:

$$C_{X_i^j}(X_h^k) = \mathbb{P}(X_h^k | X_i^j) = \frac{\mathbb{P}(X_h^k, X_i^j)}{\mathbb{P}(X_i^j)} \tag{40}$$

We can then write down expressions for Markov conditional independence as follows:

$$X_i^j \perp\!\!\!\perp X_h^k | X_l^m \Rightarrow \mathbb{P}(X_i^j, X_h^k | X_l^m) = \mathbb{P}(X_i^j | X_l^m) \times \mathbb{P}(X_h^k | X_l^m) \tag{41}$$

Following Bayes’ theorem, CI conditions –in this case– will be of the form:

$$\mathbb{P}(X_i^j, X_h^k | X_l^m) = \frac{\mathbb{P}(X_i^j, X_l^m)}{\mathbb{P}(X_l^m)} \times \frac{\mathbb{P}(X_h^k, X_l^m)}{\mathbb{P}(X_l^m)} = \frac{\mathbb{P}(X_i^j, X_l^m) \times \mathbb{P}(X_h^k, X_l^m)}{\mathbb{P}(X_l^m)^2} \tag{42}$$

Equation (42) is useful since in large scale data applications is computationally cheaper to work with joint and marginal probabilities rather than conditionals.

Now let us consider the case of conditional independence given several conditional variables. The case for CI given two variables could be written—using conditional kernels—as follows:

$$X_i^j \perp\!\!\!\perp X_h^k | X_l^m, X_n^o \Rightarrow \mathbb{P}(X_i^j, X_h^k | X_l^m, X_n^o) = \mathbb{P}(X_i^j | X_l^m, X_n^o) \times \mathbb{P}(X_h^k | X_l^m, X_n^o) \tag{43}$$

Hence,

$$\mathbb{P}(X_i^j, X_h^k | X_l^m, X_n^o) = C_{X_l^m, X_n^o}(X_i^j) \times C_{X_l^m, X_n^o}(X_h^k) \tag{44}$$

Using Bayes’ theorem,

$$\mathbb{P}(X_i^j, X_h^k | X_l^m, X_n^o) = \frac{\mathbb{P}(X_i^j, X_l^m, X_n^o)}{\mathbb{P}(X_l^m, X_n^o)} \times \frac{\mathbb{P}(X_h^k, X_l^m, X_n^o)}{\mathbb{P}(X_l^m, X_n^o)} \tag{45}$$

$$\mathbb{P}(X_i^j, X_h^k | X_l^m, X_n^o) = \frac{\mathbb{P}(X_i^j, X_l^m, X_n^o) \times \mathbb{P}(X_h^k, X_l^m, X_n^o)}{\mathbb{P}(X_l^m, X_n^o)^2} \tag{46}$$

In order to generalize the previous results to CI relations given an arbitrary set of conditionals, let us consider the following *sigma-algebraic* approach:

Let Σ_{ih}^{jk} be the σ -algebra of all subsets of X that do not contain X_i^j or X_h^k . If we consider the contravariant index $i \in \alpha$ with $i = 1, 2, \dots, N$ and the covariant index $j \in \beta$ with $j = 1, 2, \dots, L$, then there are $\mathcal{M} = \frac{NL}{2}$ such σ -algebras in X (let us recall that TMFs are *undirected* graphical models).

A relevant problem for network reconstruction is that of establishing the more general Markov pairwise CI conditions, i.e., the CI relations for every edge not drawn in the graph. Two arbitrary nodes X_i^j and X_h^k are conditionally independent given the rest of the graph iff:

$$X_i^j \perp\!\!\!\perp X_h^k | \Sigma_{ih}^{jk} \Rightarrow \mathbb{P}(X_i^j, X_h^k | \Sigma_{ih}^{jk}) = \mathbb{P}(X_i^j | \Sigma_{ih}^{jk}) \times \mathbb{P}(X_h^k | \Sigma_{ih}^{jk}) \tag{47}$$

By using conditional kernels, the recursive relations and Bayes’ theorem it is possible to write down \mathcal{M} expressions of the form:

$$\mathbb{P}(X_i^j, X_h^k | \Sigma_{ih}^{jk}) = \frac{\mathbb{P}(X_i^j, \Sigma_{ih}^{jk}) \times \mathbb{P}(X_h^k, \Sigma_{ih}^{jk})}{\mathbb{P}(\Sigma_{ih}^{jk})^2} \tag{48}$$

The family of Equations (48) represent the CI relations for all the non-existing edges in the hypergraph \hat{G} , i.e., every pair of nodes X_i^j and X_h^k not-connected in \hat{G} must be conditionally

independent given the rest of the nodes in the graph. These expression may serve to implement exact tests or optimization strategies for *graph reconstruction* and/or *graph sparsification* in applications considering a mutual information threshold I_0 as in Equation (32).

In brief, for every node pair with a mutual information value lesser than I_0 , the presented graph reconstruction approach will not draw an edge, hence implying CI between the two nodes given the rest. Such CI condition may be tested on the data to see whether it holds or the threshold itself can be determined by resorting to optimization schemes (e.g. error bounds) in Equation (48).

4. Graph Theoretical Features and Multilinear Structure

Once the probabilistic properties of TMFs have been set, it may be fit to briefly present some of their graph theoretical features, as well as some preliminaries as to the reasons to embed hyperadjacency matrices into multilayer adjacency tensors. Given that TMFs are indeed PGMs, some of their graph characteristics will result relevant here.

Since the work by De Domenico and coworkers [5] covers in great detail how the multilinear structure of the multilayer adjacency tensor allows the calculation of these quantities—usually as projection operations—we will only mention connectivity degree vectors since these are related with the size of the TMF dependency neighborhoods.

Let us recall multilayer adjacency tensors, as defined in Equations (34) and (35). To ease presentation, we will work with the unweighted tensor $\mathbb{A}_{\beta\delta}^{\alpha\gamma}$ (Equation (34)). The *multidegree centrality vector* K^α which contains the connectivity degrees of the nodes spanning different layers can be written as follows:

$$K^\alpha = \mathbb{A}_{\beta\delta}^{\alpha\gamma} U_\gamma^\delta u^\beta \quad (49)$$

Here U_γ^δ is a rank 2 tensor that contains a 1 in every component and u^β is a rank 1 tensor that contains a 1 in every component—these quantities are called 1—tensors by De Domenico and coworkers [5]. It can be shown that K^α is indeed given by the sums of the *connectivity degree vectors* k^α corresponding to all different layers:

$$K^\alpha = \sum_{h=1}^L \sum_{k=1}^L k^\alpha(hk) \quad (50)$$

$k^\alpha(hk)$ is the vector of connections that nodes in the set $\alpha = 1, 2, \dots, N$ in layer h have to any other nodes in layer k . Whereas K^α is the vector with connections in all the layers. Appropriate projections will yield measures such as the size of the neighborhood to a given vertex $|N_{X_i^h}|$, the size of its Markov blanket $|mb(X_i^h)|$, or other similar quantities.

5. Specific Applications

After having considered some of the properties of this class of Tensor Markov Fields, it may become evident that aside from purely theoretical importance, there is a number of important applications that may arise as probabilistic graphical models in tensor valued problems, among the ones that are somewhat evident are the following:

- The analysis of multidimensional biomolecular networks such as the ones arising from multi-omic experiments (For a real-life example, see Figure 4) [8–10];
- Probabilistic graphical models in computer vision (especially 3D reconstructions and 4D [3D+time] rendering) [11];
- The study of fracture mechanics in continuous deformable media [12];
- Probabilistic network models for seismic dynamics [13];
- Boolean networks in control theory [14].

Some of these problems are being treated indeed as multiple instances of Markov fields or as multipartite graphs or hypergraphs. However, it may become evident that when random variables *across layers* are interdependent (which is often the case), the definitions of potentials, cliques and partition functions, as well as the conditional statistical independence features become manageable (and in some cases even meaningful) under the presented formalism of Tensor Markov Fields.

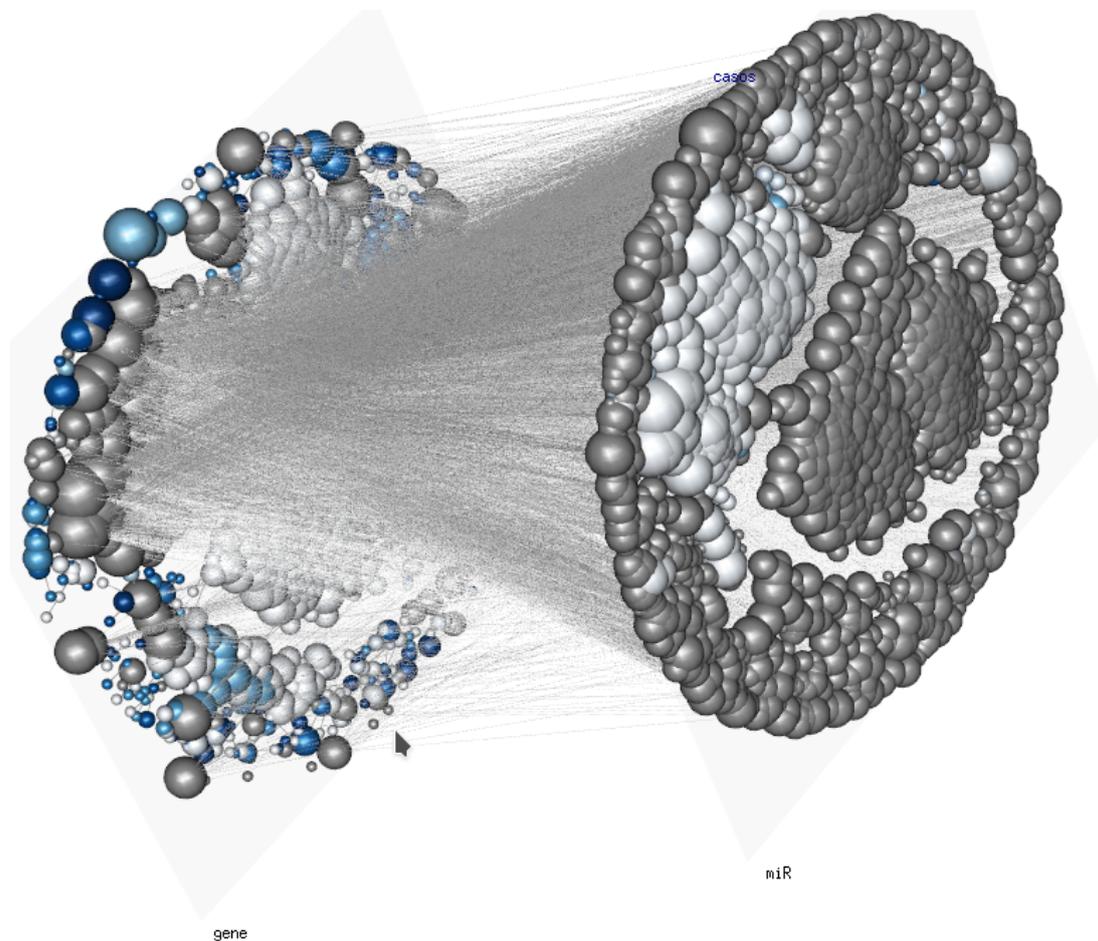


Figure 4. Gene and microRNA regulatory network: A Tensor Markov Field depicting the statistical dependence of genome wide gene and microRNA (miR) on a human phenotype. Edge width is given by the mutual information $I^{\dagger}(X_i^j, X_h^k)$ between expression levels of genes (layer j) and miRs (layer k) in a very large corpus of RNASeq samples, vertex size is proportional to the *degree*, i.e., the size of the node's neighborhood, $N_{X_i^j}$.

6. Conclusions

Here we have presented the definitions and fundamental properties of Tensor Markov Fields, i.e., random Markov fields over tensor spaces. We have proved –by extending the results of Dobruschin, Hammersley and Clifford to such tensor valued fields– that tensor Markov fields are indeed Gibbs fields whenever strictly positive probability measures are considered. We also introduced a class of tensor Markov fields obtained by using information theoretical statistical dependence measures inducing local and global Markov properties, and show how these can be used as probabilistic graphical models in multi-context environments much in the spirit of the so-called multilayer network approach. Finally, we discuss the convenience of embedding tensor Markov fields in the multilinear tensor representation of multilayer networks.

Funding: This research received no external funding.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Dobruschin, R.L. The description of a random field by means of conditional probabilities and conditions of its regularity. *Theory Probab. Appl.* **1968**, *13*, 197–224. [[CrossRef](#)]
2. Grimmett, G.R. A theorem about random fields. *Bull. Lond. Math. Soc.* **1973**, *5*, 81–84. [[CrossRef](#)]
3. Rota, G.C. On the foundations of combinatorial theory I: Theory of Möbius functions. *Probab. Theory Relat. Fields* **1964**, *2*, 340–368. [[CrossRef](#)]
4. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*; John Wiley & Sons: Hoboken, NJ, USA, 2012.
5. De Domenico, M.; Solé-Ribalta, A.; Cozzo, E.; Kivelä, M.; Moreno, Y.; Porter, M.A.; Gómez, S.; Arenas, A. Mathematical formulation of multilayer networks. *Phys. Rev. X* **2013**, *3*, 041022. [[CrossRef](#)]
6. Koller, D.; Friedman, N. Probabilistic Graphical Models: Principles and Techniques (Adaptive Computation and Machine Learning series). *Mit Press. Aug* **2009**, *31*, 2009.
7. Williams, D. *Probability with Martingales*; Cambridge University Press: Cambridge, UK, 1991.
8. Hernández-Lemus, E.; Espinal-Enríquez, J.; de Anda-Jáuregui, G. Probabilistic multilayer networks. *arXiv* **2018**, arXiv:1808.07857.
9. de Anda-Jauregui, G.; Hernandez-Lemus, E. Computational Oncology in the Multi-Omics Era: State of the Art. *Front. Oncol.* **2020**, *10*, 423. [[CrossRef](#)]
10. Hernández-Lemus, E.; Reyes-Gopar, H.; Espinal-Enríquez, J.; Ochoa, S. The Many Faces of Gene Regulation in Cancer: A Computational Oncogenomics Outlook. *Genes* **2019**, *10*, 865. [[CrossRef](#)] [[PubMed](#)]
11. McGee, F.; Ghoniem, M.; Melançon, G.; Oñ Jacques, B.; Pinaud, B. The state of the art in multilayer network visualization. In *Computer Graphics Forum*; John Wiley & Sons: Hoboken, NJ, USA, 2019; Volume 38, pp. 125–149.
12. Krejsa, M.; Koubová, L.; Flodr, J.; Protivínský, J.; Nguyen, Q.T. Probabilistic prediction of fatigue damage based on linear fracture mechanics. *Fract. Struct. Integr.* **2017**, *39*, 143–159. [[CrossRef](#)]
13. Abe, S.; Suzuki, N. Complex network of earthquakes. In *International Conference on Computational Science*; Springer: Berlin/Heidelberg, Germany, 2004; pp. 1046–1053.
14. Liu, F.; Cui, Y.; Wang, J.; Ji, D. Observability of probabilistic Boolean multiplex networks. *Asian J. Control.* **2020**, *1*, 1–8. doi:10.1002/asjc.2290. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).