



Random Fields in Physics, Biology and Data Science

Enrique Hernández-Lemus^{1,2*}

¹Computational Genomics Division, National Institute of Genomic Medicine, Arenal Tepepan, Mexico, ²Centro de Ciencias de La Complejidad, Universidad Nacional Autónoma de México, Coyoacán, Mexico

A random field is the representation of the joint probability distribution for a set of random variables. Markov fields, in particular, have a long standing tradition as the theoretical foundation of many applications in statistical physics and probability. For strictly positive probability densities, a Markov random field is also a Gibbs field, i.e., a random field supplemented with a measure that implies the existence of a regular conditional distribution. Markov random fields have been used in statistical physics, dating back as far as the Ehrenfests. However, their measure theoretical foundations were developed much later by Dobruschin, Lanford and Ruelle, as well as by Hammersley and Clifford. Aside from its enormous theoretical relevance, due to its generality and simplicity, Markov random fields have been used in a broad range of applications in equilibrium and non-equilibrium statistical physics, in non-linear dynamics and ergodic theory. Also in computational molecular biology, ecology, structural biology, computer vision, control theory, complex networks and data science, to name but a few. Often these applications have been inspired by the original statistical physics approaches. Here, we will briefly present a modern introduction to the theory of random fields, later we will explore and discuss some of the recent applications of random fields in physics, biology and data science. Our aim is to highlight the relevance of this powerful theoretical aspect of statistical physics and its relation to the broad success of its many interdisciplinary applications.

Keywords: random fields, probabilistic graphical models, Gibbs fields, Markov fields, Gaussian random fields

OPEN ACCESS

Edited by:

Umberto Lucia,
Politecnico di Torino, Italy

Reviewed by:

Farrukh Mukhamedov,
United Arab Emirates University,
United Arab Emirates
Luca Martino,
Rey Juan Carlos University, Spain

*Correspondence:

Enrique Hernández-Lemus
ehernandez@inmegen.gob.mx

Specialty section:

This article was submitted to
Interdisciplinary Physics,
a section of the journal
Frontiers in Physics

Received: 15 December 2020

Accepted: 01 February 2021

Published: 15 April 2021

Citation:

Hernández-Lemus E (2021) Random
Fields in Physics, Biology and
Data Science.
Front. Phys. 9:641859.
doi: 10.3389/fphy.2021.641859

1 INTRODUCTION

The theory and applications of random fields born out of the fortunate marriage of two simple but deep lines of reasoning. On the one hand, physical intuition, strongly founded in the works of Boltzmann and the Ehrenfests, but also in other originators of the kinetic theory of matter, was that large scale, long range phenomena may originate from (a multitude of) local interactions. On the other hand, probabilistic reasoning induced us to think that such multitude of local interactions would be stochastic in nature. These two ideas, paramount to statistical mechanics, have been extensively explored and develop into a full theoretical subdiscipline, the theory of random fields. Perhaps the archetypal instance of a random field was laid out in the doctoral thesis of Ernst Ising, the Ising model of ferromagnetism [1]. However, although the physical ideas have been laid out mainly by physicists, much of the further mathematical development was made by the Russian school of probability. In particular, by the works of Averintsev [2, 3], which—along with the measure theoretical-inspired formalization of statistical mechanics by J.W. Gibbs—, was able to specify a general class of fields described only by pair potentials [4]. Theoretical advances were given by Stavskaya who studied random fields by measure theory considering them as invariant states for local

processes [5, 6], by Vasilyev who consider stationary measures as derived from local interactions in discrete mappings [7] and others.

The formal establishment of the theory of Markov-Gibbs random fields, however, is often attributed to the works of Dobruschin, Lanford and Ruelle [8, 9], in particular to their DLR equations for the probability measures. Also remarkable is the contribution of Hammersley and Clifford, who developed a proof of the equivalence of Gibbs random fields and Markov random fields, provided positive definite probabilities [10]. Although the authors never officially published this work, that they thought to be incomplete given the—now known to be essential—requirement of positive definite probabilities, several published works have been made on top of it and even alternative proofs have been published [11–13].

Aside from the extensive use of the Ising model and other random fields in statistical mechanics—too many contributions to mention here, but most of them comprehensively reviewed in the monographs by Baxter [14], Cibra [15], McCoy and Wu [16], Thompson [17] and in the simulation-oriented book by Adler [18]—; there has been also a deep interest in development in models in biophysics, computer science and other fields. The development of Hopfield networks as models of addressable memory in neurophysiology (and artificial neural networks) [19] is perhaps one of the earliest examples. Followed by the implementation of the so-called Boltzmann machines in artificial intelligence (AI) applications [20, 21] paved the way to a plethora of theoretical, computational and representational applications of random fields.

In the rest of this review paper, we will present some general grounds of the theory of Markov random fields to serve as a framework to elaborate on many of its relevant applications inside and outside physics. Our emphasis here will not be to be comprehensive but illustrative of some relevant features that have made this quintessential model of statistical physics so pervasive in our discipline and in many others (*Markov Random Fields: A Theoretical Framework*). We will also discuss how methodological and computational advances in these areas may be implemented to improve on the applications of random fields in physical models. We have chosen to focus on applications in Physics (*Markov Random Fields in Physics*), Biology (*Markov Random Fields in Biology*) and Data Science (*Markov Random Fields in Data Science and Machine Learning*). We are aware that by necessity (finiteness), we are leaving out contributions in fields such as sociology (Axelrod models, for instance), finance (volatility maps, Markov switching models, etc.) and others. However, we believe this panoramic view will make easier for the interested reader to look into these other applications. Finally, in *Concluding Remarks* we will outline some brief concluding remarks.

2 MARKOV RANDOM FIELDS: A THEORETICAL FRAMEWORK

Here we will define and describe Markov random fields [8, 12] (MRFs) as an appropriate theoretical framework useful for systematic probabilistic analysis in various settings. An MRF

represents, in this context, the joint probability distribution for a set (as large as desired) of real-valued random variables. There are several extensions of the general ideas presented here, that will be presented and briefly addressed as needed.

Let $X = X_\alpha$ be a vector of random variables (i.e., the features or characteristic functions used to describe a system of interest). An MRF may be represented as an undirected graph depicting the statistical dependency structure of X , as given by the joint probability distribution $\mathbb{P}(X)$ [22].

Let this graph be embodied in the form of a duplex $G = (V, E)$ consisting of a set V of vertices or nodes (the random variables X_i 's) and a set $E \subseteq V \times V$ of edges connecting the nodes (thus representing the statistical dependencies between random variables). E also represents a neighborhood law N stating which vertex is connected (i.e., dependent) to which other vertex in the graph. With this in mind, an MRF can be also represented as $G = (V, N)$. The set of neighbors of a given point X_i is denoted N_{X_i} .

2.1 Configuration

We can assign each point in the graph, one of a finite set S of labels. Such assignment, it is often called a *configuration*. We can then 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 . We may think of ω_A as a configuration on the subgraph G_A restricting V to points of A .

2.2 Local Characteristics

We can define *local characteristics* on MRFs. The local characteristics of a probability measure \mathbb{P} defined on Ω are the conditional probabilities:

$$\mathbb{P}(\omega_t | \omega_{T_t}) = \mathbb{P}(\omega_t | \omega_{N_t}) \quad (1)$$

This represents the probability that the point t is assigned the value ω_t , given the values at all other points of the graph. Let us rewrite **Eq. 1**. Since the probability measure will define an MRF if the local characteristics depend only on the outcomes at neighboring points, i.e., if for every ω

$$\mathbb{P}(\omega_{X_i} | \omega_{G \setminus X_i}) = \mathbb{P}(\omega_{X_i} | \omega_{N_{X_i}}) \quad (2)$$

2.3 Cliques

Given an arbitrary graph, we may refer to a set of points C , as a *clique*, if every pair of points in C are neighbors. This includes the empty set as a clique. A clique is then a set whose *induced subgraph* is complete. Cliques are also called *complete induced subgraphs* or *maximal subgraphs*.

2.4 Configuration Potentials

A *potential* η is an assignment of a number $\eta_A(\omega)$ to every subconfiguration ω_A of a configuration ω in the graph G . A given η , induces an *energy* $U(\omega)$ on the set of all configurations ω as follows:

$$U(\omega) = \sum_A \eta_A(\omega) \quad (3)$$

Here, for fixed ω , the sum is taken over all subsets $A \subseteq V$ including the empty set. It is possible to define a probability measure, called the *Gibbs measure induced by U* as

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

Z (taken from the German word *zustanssumme* or *sum over states*) is a normalization constant called the *partition function*. As it is known, explicit computation of the partition function is in many cases a very challenging endeavor. There is a great deal of work in the development of methods and approaches to overcome some (but not all) challenges in this regard. Some of these approximations will be discussed later on.

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

The term *potential* is often used in connection with potential energies. In this context η_A is commonly termed a *potential energy* in physics applications. $\phi_A = e^{-\eta_A}$ is then called a potential.

Equations 4, 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, and it is also used in theoretical physics, we will refer to Eqs. 6, 7 as the definitions of Gibbs measure and partition function (respectively) unless otherwise stated. This will also be justified given that Eq. 6 is a form of probability factorization (in this case a *clique factorization*) [11].

2.5 Gibbs Fields

A potential is termed a nearest neighbor Gibbs potential if $\phi_A(\omega) = 1$ whenever A is not a clique. We often call a *Gibbs measure* to any regular measure induced by a nearest neighbor Gibbs potential. However, we may define more general Gibbs measures by considering different classes of potentials.

The inclusion of all cliques in the calculation of the Gibbs measure is needed to establish the equivalence between Gibbs random fields and Markov random fields. A nearest neighbor Gibbs measure on a graph determines an MRF as follows [22]:

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 graph G . Then,

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

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

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

For any clique C that does not contain X_i , $\phi_C(\omega) = \phi_C(\omega')$. So that all the terms that correspond to the cliques that do not contain the point X_i cancel both from the numerator and the denominator in Eq. 10, therefore this probability depends only on the values x_i at X_i and its neighbors. \mathbb{P} defines thus an MRF. A more general proof of this equivalence was given by Hammersley-Clifford theorem (see for instance [11]).

In essence, we can state that among the general class of random fields, Markov random fields are defined by obeying the Markov neighborhood law. Gibbs fields are usually understood as Markov fields with strictly positive probability measures (in particular, a strictly positive joint probability density). These Markov-Gibbs fields are thus defined by the Markov property and the positive definite probabilities and are the ones that follow the Hammersley-Clifford theorem. More general Gibbs fields can be defined by other neighborhood laws than the Markov property [23], but these will not be addressed in the present work.

2.6 Conditional Independence in Markov Random Fields

To discuss the conditional independence structure induced by MRFs, let us consider the following: An adjacency matrix A_{ij} represents the neighborhood law (as given by the Markov property) on the graph G . Every non-zero entry in this matrix represents a statistical dependency relation between two elements on X . The conditional dependence structure on MRFs is related not only to the *local* statistical independence conditions, but also to the dependency structure of the whole graph [11, 24].

A definition of conditional independence (CI) for the set of random variables can be given as follows:

$$\begin{aligned} (X_i \perp\!\!\!\perp X_j) | X_l &\Leftrightarrow \mathbb{F}_{X_i, X_j | X_l = X_l^*}(X_i^*, X_j^*) \\ &= \mathbb{F}_{X_i | X_l = X_l^*}(X_i^*) \cdot \mathbb{F}_{X_j | X_l = X_l^*}(X_j^*) \end{aligned} \tag{11}$$

$\forall X_i, X_j, X_l \in X$

Here $\perp\!\!\!\perp$ refers to conditional independence between two random variables. $\mathbb{F}_{X_i, X_j | X_l = X_l^*}(X_i^*, X_j^*) = Pr(X_i \leq X_i^*, X_j \leq X_j^* | X_l = X_l^*)$ is the joint conditional cumulative distribution of X_i and X_j given X_l . X_i^* , X_j^* and X_l^* are realizations of the corresponding random variables.

In the case of MRFs, CI is defined by means of graph separation: Hence $X_i \perp\!\!\!\perp_G X_j | X_l$ iff X_l separates X_i from X_j in G . This means that if we remove node X_l there are no undirected paths from X_i to X_j in G .

Conditional independence in random fields can be considered in terms of subsets of V . Let A, B and C be subsets of V . The statement $X_A \perp\!\!\!\perp_G X_B | X_C$, which holds only iff C separates A from B in G , means that if we remove all vertices in C there will be no paths connecting any vertex in A to any vertex in B . This is customarily called the *global Markov property* of TMFs [11, 24].

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

In an MRF, the Markov blanket of a vertex is its set of first neighbors. This statement is the so-called *undirected local Markov property*. Starting from the local Markov property, it is possible to show that two vertices X_i and X_j are conditionally independent given the rest if there is no direct edge between them. This is the *pairwise Markov property*.

If we denote by $G_{X_i \rightarrow X_j}$ the set of undirected paths in the graph G connecting vertices X_i and X_j , then the pairwise Markov property of an MRF is given by:

$$X_i \perp\!\!\!\perp X_j | G \setminus \{X_i, X_j\} \Leftrightarrow G_{X_i \rightarrow X_j} = \emptyset \tag{12}$$

Hence the global Markov property implies the local Markov property which, in turn, implies the pairwise Markov property. For systems with positive definite probability densities, it has been proved that pairwise Markov actually implied global Markov (See [11] p. 119 for a proof). This is important for applications since it is easier to assess pairwise conditional independence statements.

2.6.1 Indepence Maps

Let I_G denote the set of all conditional independence relations encoded by the graph 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 $\mathbb{P}(X_i)$. A graph G will be called an *independence map (I-map)* for a probability distribution $\mathbb{P}(X_i)$, if all CI relations implied by G hold for $\mathbb{P}(X_i)$, i.e., $I_G \subseteq I_{\mathbb{P}}$ [11].

The converse statement is however not necessarily true, i.e., there may be some CI relations implied by $\mathbb{P}(X_i)$ that are not coded in the graph G . We may often be interested in the so-called *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 $\mathbb{P}(X_i) > 0$. Let G^\dagger be the graph obtained by introducing edges between all pairs of vertices X_i, X_j such that $X_i \perp\!\!\!\perp X_j | X \setminus \{X_i, X_j\}$, then G^\dagger is the unique minimal I-map. We call G a *perfect map* of \mathbb{P} when there are no dependencies G which are not indicated by \mathbb{P} , i.e., $I_G = I_{\mathbb{P}}$ [11].

2.6.2 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 [24]. In order to be able to write down expressions for C.I. tests let us introduce the following *conditional kernels* [25]:

$$C_A(B) = \mathbb{P}(B|A) = \frac{\mathbb{P}(AB)}{\mathbb{P}(A)} \tag{13}$$

As well as their generalized recursive relations:

$$C_{ABC}(D) = C_{AB}(D|C) = \frac{C_{AB}(CD)}{C_{AB}(C)} \tag{14}$$

The conditional probability of X_i given X_j can be thus written as:

$$C_{X_j}(X_i) = \mathbb{P}(X_i|X_j) = \frac{\mathbb{P}(X_i, X_j)}{\mathbb{P}(X_j)} \tag{15}$$

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

$$X_i \perp\!\!\!\perp X_j | X_l \Rightarrow \mathbb{P}(X_i, X_j | X_l) = \mathbb{P}(X_i | X_l) \times \mathbb{P}(X_j | X_l) \tag{16}$$

Following Bayes' theorem, CI conditions—in this case—will be of the form:

$$\mathbb{P}(X_i, X_j | X_l) = \frac{\mathbb{P}(X_i, X_l)}{\mathbb{P}(X_l)} \times \frac{\mathbb{P}(X_j, X_l)}{\mathbb{P}(X_l)} = \frac{\mathbb{P}(X_i, X_l) \times \mathbb{P}(X_j, X_l)}{\mathbb{P}(X_l)^2} \tag{17}$$

Equation 17 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:

$$X_i \perp\!\!\!\perp X_j | X_l, X_n \Rightarrow \mathbb{P}(X_i, X_j | X_l, X_n) = \mathbb{P}(X_i | X_l, X_n) \times \mathbb{P}(X_j | X_l, X_n) \tag{18}$$

Hence,

$$\mathbb{P}(X_i, X_j | X_l, X_n) = C_{X_l, X_n}(X_i) \times C_{X_l, X_n}(X_j) \tag{19}$$

Using Bayes' theorem,

$$\mathbb{P}(X_i, X_j | X_l, X_n) = \frac{\mathbb{P}(X_i, X_l, X_n)}{\mathbb{P}(X_l, X_n)} \times \frac{\mathbb{P}(X_j, X_l, X_n)}{\mathbb{P}(X_l, X_n)} \tag{20}$$

Or

$$\mathbb{P}(X_i, X_j | X_l, X_n) = \frac{\mathbb{P}(X_i, X_l, X_n) \times \mathbb{P}(X_j, X_l, X_n)}{\mathbb{P}(X_l, X_n)^2} \tag{21}$$

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_{ij} be the σ -algebra of all subsets of X that do not contain X_i or X_j . 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 on the graph. Two arbitrary nodes X_i and X_j are conditionally independent given the rest of the graph iff:

$$X_i \perp\!\!\!\perp X_j | \Sigma_{ij} \Rightarrow \mathbb{P}(X_i, X_j | \Sigma_{ij}) = \mathbb{P}(X_i | \Sigma_{ij}) \times \mathbb{P}(X_j | \Sigma_{ij}) \tag{22}$$

By using conditional kernels, the recursive relations and Bayes' theorem it is possible to write down:

$$\mathbb{P}(X_i, X_j | \Sigma_{ij}) = \frac{\mathbb{P}(X_i, \Sigma_{ij}) \times \mathbb{P}(X_j, \Sigma_{ij})}{\mathbb{P}(\Sigma_{ij})^2} \tag{23}$$

The family of Eq. 23 represent the CI relations for all the non-existing edges in the graph G , i.e., every pair of nodes X_i and X_j not-connected in G must be conditionally independent given the rest of the nodes in the graph. This is perhaps the most important features of MRFs in connection with potential applications as probabilistic graphical models. CI conditions often lead to simpler (or at least computationally tractable) ways to factorize the PDF or compute the partition function.

The algorithmic complexity of doing so *in general* (since the number of CI relations grows combinatorially with the size of the graph), makes it prohibitive in the case of a large number of variables/relationships, in spite of recent advances on optimizing large dimensional space CI testing for discrete distributions [26]. This is the biggest advantage of the present approach. As long as one deals with strictly positive probabilities (that one can often attain *via* regularization) and Hammersley-Clifford conditions apply, modeling with nearest neighbor Gibbs potentials ensure CI conditions in the graph (recall that global Markov property implies pairwise Markov property and vice versa).

Now that we have presented the fundamentals of MRFs at an introductory level, this may allow to discuss on how these features have impact on their wide range of applications, as the basis for probabilistic graphical models. Let us start by considering some recent applications in physics.

3 MARKOV RANDOM FIELDS IN PHYSICS

From the pioneering work of the Ehrenfests, to the foundational Ising models and its extensions (Potts, XY, etc.), MRFs have been thoroughly used and developed in many subdisciplines of physics, ranging from condensed matter and mathematical physics to geophysics, econophysics and more. There are numerous in-depth reviews and monographs summarizing research along these lines (see, for instance [27–30]). Since the main goal here is to present some of the characteristic features of the usefulness of MRFs as probabilistic graphical models, in terms of their mathematical properties and broad scope of applicability, both within and outside physics; our discussion will be somehow biased toward work showing one or more of such features.

3.1 MRFs in Statistical Mechanics and Mathematical Physics

Due to their intrinsic simplicity and generality, MRFs have attracted the attention of mathematical physicists and probability theorists looking to extend their associated theoretical foundations. Important work has been done, for instance, to incorporate geometrical properties and generalized embeddings to the theory of random fields. Extremely relevant in this regard is the monumental work presented in the monograph by Adler and Taylor [31]. There, the authors expand on the consideration of a random field as a stochastic process in a metric space (discrete, Euclidean, etc.) to consider random fields as stochastic mappings over manifolds. This extension is given *via* writing down differential geometry characterizations of the fields based on a measure-theoretic definition of probability.

Though this work may seem quite abstract, it was indeed born out of an idea for an application of random fields to neuroscience. Nurturing from similar ideas, recent work by Ganchev [32] has expanded the notion of locality of MRFs and assimilate it to the geometric features present in lattice quantum gauge theories, to generate a *gauge theory of Markov-Gibbs fields*. Again, even if the setting seems to be quite theoretical, an application to the modeling of trading networks in finance is given.

Other mathematical extensions of Markov random fields are related to the nature of the graphical model considered. In general, probabilistic graphical models may belong to one of two quite general classes: Markov networks (such as MRFs) which are *undirected* graphs or Bayesian networks which are *directed* graphs. The difference between undirected and directed graphical models impose consequences in the kind of fundamental mathematical objects of the theory: joint probabilities or conditional probabilities, loopy graphs or trees-directed acyclic graphs-, clique factorization vs. conditional probability factorization *via* the chain rule, etc. Whether the model is undirected or directed also has modeling and computational consequences. To be fair, both models have pros and cons.

Trying to overcome the limitations of both general approaches, Freno and Trentin [33] developed a more general approach to random fields termed *Hybrid random fields* (HRFs). The purpose of HRFs is to allow the systems to present a wider variety of conditional independence structures. As we will discuss later, allowing for a systematic incorporation of more general classes of conditional independence structures in indeed one of the current *hot topics* in computational intelligence and machine learning. Actually, even when HRFs are theoretical constructs (much alike MRFs) they were designed to be *learning machines*, i.e., to be supplemented with training algorithms to deal with high dimensional data. HRFs were developed for logical inference in the presence of partial information or noise. As in the case of MRFs and of their gauge extensions just mentioned, HRFs were developed to rely on a *principle of locality* which is an extension of the Markov property that allows for sparse stochastic matrix representations amenable for the computation on actual applications. Once a (graph) structure has been given (or inferred) HRFs are able (as is the case of MRFs) to learn the local (conditional or joint-partial) probability distributions from empirical data, a task commonly known in statistics as *parameter learning* [34]. Hence HRFs are theoretically founded, but developed thinking in applications. The scope of applicability of MRFs has also become broader by expanding its applicability to model tensor valued quantities [35], giving rise to the so-called multilayer graphical models, also called multilayer networks [36–39].

Aside from expanding the fundamental structure of MRFs, mathematical physics applications of Gibbs random fields are abundant. In particular, the so-called Random Field Ising model (RFIM) has gained a lot of attention in the recent years. By using the monotonicity properties of the associated stochastic field, Aizenmann and Peled [40] were able to prove that there is a power law upper bound on the correlations on a two-dimensional Ising model, supplemented with a quenched random magnetic

field. The fact that by combining random fields (the intrinsic Ising field and the quenched magnetic field), the nature of the phase transitions may drastically change has made the RFIM a current topic of discussion in mathematical statistical mechanics. The consequences of the induction of long range order in the RFIM, leading to the emergence of the so-called Imry-Ma phase or Imry-Ma states (named so since Imry and Ma were actually behind the first proposal of the RFIM [41]) have been the object of intense study recently. Berzin and co-workers [42] used MRFs to analyze the dynamic fluctuations of the order parameter in the Imry-Ma RFIM and its coupling with the static fluctuations of the structural random field (accounting for the defects). Interestingly, anisotropic coupling arises from two non-absolutely overlapping local fields [43]. The effects of the non-overlapping fields in anisotropy and disorder has been studied since several decades ago [44], but the actual relationship with non-locality was established relatively recently. For instance, it was until 2018 that Chatterjee was able to quantitatively describe the decay of correlations of the 2D RFIM [45] in a relevant paper that led Aizenmann to re-analyze his former, mostly qualitative proposal [40, 46].

Local stochastic phenomena in non-homogeneous and disordered media in the context of the RFIM has also attracted attention in relation to critical exponents and scaling. Trying to expand on the origins of long range order from local interactions, Fytas and coworkers have studied the 4D RFIM and its hyperscaling coefficients [47]. This is particularly interesting since it has been shown, *via* perturbative renormalization group calculations, that the critical exponents of the RFIM in D dimensions are the same as the exponents of the pure Ising model in $D - 2$ dimensions [48]. Related work has been carried out by Tarjus and Tissier, but they instead resort to the use of the so-called *functional renormalization group* approach in the multi-copy formalism setting [49]. Their work has extended the predictive capabilities of MRFs by incorporating ideas from symmetry breaking allowing to characterize not just long-range order (LRO) but also intermediate states characterized by quasi-long range order (QLRO). The fact that QLRO may be attained from purely Markov statistics (localized interactions) is in itself appealing for statistical physics. The fact that local dependencies may suffice to account for LRO and QLRO under certain conditions that do not violate the Markov property of the MRFs, will have relevant consequences for the applications of MRFs outside physics, such as in the case of image reconstruction and pattern recognition in machine learning. We will come back to these ideas later on.

Locality as depicted in MRFs can also have important consequences for the theory of fluctuations in fields of interacting particles. Reconstructing Boltzmann statistics from local Gibbs fields (that as we have repeatedly stated are formally equivalent to MRFs, provided strictly positive probability measures) imply that under central limit scales the fluctuation field of local functions can be represented instead as a function of the density fluctuation field, in what is known as the Boltzmann-Gibbs principle (BGP). It has been shown that the BGP induces a duality whose origins are purely probabilistic, i.e., is independent of the nature of the interactions provided their compliance with the tenets of MRFs [50].

It is worth noticing that these contemporary developments in the formal theory of MRFs are actually founded on seminal work by probability theorists and mathematical physicists such as Dobrushin, Ruelle, Gudder, Kessler and others. For instance, Dobrushin laid out the essential conditions of regularity that allow to make explicit the conditional probabilities in MRF models [8]. This work, further developed by Lanford and Ruelle [9] gives rise to the so called Dobrushin-Lanford-Ruelle (DLR) equations that established, in a formal way, the properties of general Gibbs measures. Later on, Dobrushin expanded on these ideas by applying perturbation methods to generalize Gibbs measures to even wider classes of interactions (i.e., to include other families of potentials) [51]. An application of these ideas in quantum field theory can be found in [52] within the context of (truncated) generalized Gibbs ensembles.

Aside from measure-theoretical and algebraic foundations of MRFs, important developments were made by considering explicit dependency structures. In particular, the introduction of strong independence properties led to the formal definition of Gaussian random fields by Gudder [53]. Much of this earlier work has been summarized in the monograph by Kindermann and Laurie Snell [22]. The fact that MRFs are characterized by Gibbs measures even for many-body interactions (under special conditions), and not only for paired-potentials, was already envisioned by Sherman [54], though it remained an unfinished task for decades. Many body effects have actually been reported in the context of localization in the random field Heisenberg chain [55]. One step ahead toward generalizing MRFs consisted in exploring the equivalence of some properties of random fields in terms of *sample functions*. In this regard, Starodubov [56] proved that there are random fields stochastically equivalent to an MRF, but defined on another probability triple whose sample functions belong to a map associated with the original MRF. The existence of such mappings has relevant implications for applications, in particular in cases in which explicit computation of the partition function is intractable.

3.2 MRFs in Condensed Matter Physics and Materials Science

Discrete and continuous versions of random fields have been applied to model systems in condensed matter physics and materials science (CMP/MS). The relevance of MRFs and its extensions relies on their suitability to describe the onset of spatio-temporal phenomena from localized interactions. Acar and Sundararaghavan [57] have used MRFs to model the spatio-temporal evolution of microstructures, such as grain growth in polychrystalline microstructures as captured by videomicroscopy experiments. Experimental data is the foundation for explicit calculations of the (empirical) conditional probability distributions.

Gaussian random fields have been used to model quenched random potentials in fluids *via* mode-coupling by Konincks and Krakoviack [58], and to model beta-distributed material properties by Liu and coworkers [59]. These and other extensions in CMP/MS made use of continuous, piecewise continuous or lattice fluid extensions of Gibbs random fields.

Such is also the case of the work of Chen and coworkers [60] who introduced stochastic harmonic potentials in random fields to account for the effects of local interactions on the properties of structured materials; of the work by Singh and Adhilari [61] on Brownian motion in confined active colloids and of the work of Yamazaki [62] on stochastic Hall magnetohydrodynamics. A semi-continuous approach (called smoothed particle hydrodynamics, SPH), using discrete MRFs and extension theorems, was used by Ullah and collaborators [63] in their density dependent *hydrodynamic* model for crowd coherency detection in active matter.

Extending the ideas of the classic RFIM, Tadic and collaborators [64] were able to describe critical Barkhausen avalanches in quasi-2D ferromagnets with an open boundary. The use of MRFs with disordered field components has also allowed to characterize embedded inhomogeneities in the spectral properties of Rayleigh waves with application to the study of the Earth's microseismic field [65]. Geoacoustic measurements and its MRF modeling allowed these researchers to estimate the mechanical and structural properties of the Earth's crust and upper mantle. Accurate estimates of these properties are foundational to develop seismic-resistant devices and structures.

3.3 Applications of MRFs in Other Areas of Physics

MRFs have also been applied in other areas of physics aside from statistical mechanics and condensed matter. MRFs were applied for instance, in geophysical models of marine climate patterns [66], to study reservoir lithology [67] and subsurface soil patterns [68] from remote sensing data. Aside from geophysics, optics and acoustics have also incorporated MRF applications. In acoustics, for instance, an MRF formalism can be used for the isolation of selected signals [69]; or for the segmentation of sonar pulses [70]. In chemical physics, MRFs are applied for the analysis of molecular structures [71], and in the implementation of quantum information algorithms for molecular physics modeling [72].

Disparate as the applications of MRF in the physical sciences just presented may be, these are neither a comprehensive nor even a representative list. However, we expect that some of the essential aspects of its wide range of applicability and the large room for theoretical development still available for these types of models were captured in the previous discussion. Moving on to applications and developments in other disciplines, such as Biology/Biomedicine and the Data Sciences, we will try to convey, not just the usefulness of a quintessential model in statistical physics in other realms—which is huge, indeed—. We also intend to show how some of the implementations and theoretical improvements in other disciplines, can be exported back to physics and may help to solve some of the many remaining conundrums of the theory and applications of random fields in the physical sciences.

4 MARKOV RANDOM FIELDS IN BIOLOGY

Biology and Biomedicine are also disciplines in which MRFs have flourished in applications and theoretical development. The

abundance of research problems and practical cases in which stochastic phenomena dependent in spatio-temporal localization is most surely behind. From the reconstruction of complex imaging patterns (not far from applications in geophysics/astrophysics imaging), to resolution of molecular maps in structural biology, to disentangling molecular interaction networks and ecological interactions; there are many outstanding advances involving random fields in biology. Again, we will discuss here just a few examples that will likely provide us with a panoramic view and perhaps spark interest and curiosity.

4.1 Applications of MRFs in Biomedical Imaging

One somehow natural application of MRFs is imaging de-noising or *segmentation*. This is a quite general problem in which one wishes to discern patterns from a *blurred* image. In particular an MRF is built to discern which points in imaging space (pixels, voxels) are locally correlated with each other, pointing out to their *membership* to the same object in the image. The Markov neighborhood structure of the MRF is hence used to *un-blur* patterns and being able to accurately interpret the images. Often MRFs (or its associated conditional Random fields) are used in conjunction with inference machines such as Convolutional Neural Networks (CNNs). This is the case of the work by Li and Ping [73] who used a neural conditional random field (NCRF) for metastasis detection from lymph node slide images. Their NCRF approach infers the spatial correlations among neighboring patches *via* a fully connected conditional MRF incorporated on top of a CNN feature extractor. Their modeling approach used a conditional distribution of an MRF with a Gibbs distribution. As is often the case the *energy function* (i.e., the *Hamiltonian*) consists of two terms, one summarizing the contributions from unary potentials characteristic for each patch, and the other one summing the pairwise potentials measuring the *cost* of jointly assigning two neighboring patches (i.e., the *interaction* potentials).

As is common in physics, estimating the marginals is an intractable problem. Li and Ping resorted to using a *mean-field* approach and then conditioning their results on this mean field calculations. In order to do this, they trained a CNN with the empirical data. CNN-MRF approaches have also been recently applied to successfully discern computerized tomography imaging (CT scans) [74] for prostate and other pelvic organs at risk. After processing the data with an encoder/decoder scheme, the output of CNN was used as the unary potential of the MRF. Then *via* a MRF block model based on local convolution layers, a global convolution layer, and a 3D max-pooling layer the authors were able to calculate the pairwise potential. The maximum likelihood optimization problem was then solved *via* an adaptive loss function.

A similar approach was followed by Fu and collaborators [75] to solve the retinal vessel segmentation problem, fundamental in the diagnostics and surgery of ophthalmological diseases, and, until quite recently *manually* performed by an ocular pathologist. The authors also used a two term energy function within a mean

field approach. To minimize the energy function subject to empirical constraints they used a recurrent neural network based on Gaussian kernels on the feature vectors applying standard gradient descent methods. Blood vessel segmentation was also studied using conditional MRFs by Orlando and coworkers [76]. However, instead of using a mean-field approach and inferring the marginals using neural networks, these authors chose to perform Maximum a Posteriori (MAP) labeling with likelihood functions optimized *via* Support Vector Machines (SVMs). Imaging segmentation *via* MRFs can be applied not only at the tisular level, but also on cellular (and even supramolecular) scales. Several blood diseases, for instance, are diagnosed by discerning the quantity, morphology and other aspects of leukocytes as well as their nuclear and cytoplasmic structure. To this end, Reta and coworkers used unsupervised binary MRFs (i.e., classical Ising-like fields) to study leukocyte segmentation [77]. A Markov neighborhood and clique potential approach was followed. This *classic* approach has been enough since from their high quality colored imaging data, it was possible to define an energy function based on *a priori* Gaussian-distributed probabilities, then applying a maximum likelihood approach to calculate the posterior probability. Related ideas were used to study microvasculature disorders in glioblastomas by the group of Kurz [78].

Application Box I: Metastasis Detection

General problem statement: Accurate detection of metastatic events is key to proper diagnostics in cancer patients. Pathologists often resort to the analysis of whole slide images (WSI). Computational histopathology aims for the automated modeling and classification of WSI to distinguish between normal and tumor cells, thus alleviating the heavy burden of manual image classification. Li and Ping [73] used Conditional Random Fields together with deep convolutional neural networks to approach this problem.

Theoretical/Methodological approach: The approach developed by the authors consisted in using a deep convolutional neural network (CNN) for the automated detection of the relevant variables (feature extraction or feature selection). Once these relevant variables have been determined, a conditional random field (CRF) was used to consider the spatial correlations between neighboring patches. The approach used to determine tumor and non-tumor regions is similar to the one used in statistical physics of condensed matter for the determination of ferromagnetic/anti-ferromagnetic domains.

Improvements/advantages: The use of CNNs to reduce the number of variables (and to find the optimal ones) is gaining relevance in computational biology and data analysis applications of random fields. It may result useful in any setting in which there are no *a priori* determined relevant variables. By conditioning these variables on the spatial location, the authors have turn the configuration problem into a classifier thus solving their problem.

Limitations: Though not an actual limitation for their particular problem, the authors resort to the use of a mean field approach to infer the marginals. This condition can be

strengthened by using approaches such as perturbative expansions or maximum entropy optimization with a suitable set of constraints.

MRFs have also been used in conjunction with deep learning approaches for the topographical reconstruction of colon structures from conventional endoscopy images. Since the colon is a deeply complex anatomical structure, accurately reconstructing its structure to detect anomalies related to, for instance, colorectal cancer is of paramount importance. Mahmmud and Durr [79] developed a deep convolutional neural network-conditional random field method, which uses a two-term energy function whose parameters are optimized *via* stochastic-descent back-propagation. Several convolution maps were used since their goal was also to estimate depth from photographic (2D) images *via* MAP (i.e., by an a posteriori maximum likelihood) optimization. This was actually possible since the authors trained their model with over 200,000 synthetic images of an anatomically realistic colon.

To improve the automated evaluation of mammography, Sari and coworkers [80] developed an MRF approach supplemented with simulated annealing optimization (MRF/SA). Improved performance was actually attained by using pre-processing filters leading to AUC/ROC of up to 0.84, which is considered quite high since mammograms have proved to be especially hard to interpret with computer aided diagnostics. MRFs have also helped improve the estimation of cardiac strain from magnetic resonance imaging data, a relatively non-invasive test to analyze cardiac muscle mechanics [81].

4.2 Applications of MRFs in Computational Biology and Bioinformatics

Computational biology and bioinformatics are also disciplines that have widely adopted the random field formalism as a relevant component of their toolkits. There are several instances in which MRFs can be adapted to solve problems in these domains: from structural biology problems in which the spatio-temporal locality is naturally mapped onto random fields, to molecular regulatory networks in which the graph structure of the MRFs mimic the underlying connectivity of the networks, to *semantic* and *linguistic* segmentation problems in genomic sequences or biomedical texts.

Regarding computational models in structural biology, Rosenberg-Johansen and his group [82] used a combination of deep neural networks and conditional random fields to improve predictions on the secondary structure of proteins (i.e., the three dimensional conformation of local protein segments, the formation of alpha helices, beta sheets and so on). The CRF approach was quite useful in this case (in general non-computationally tractable), since in protein secondary structure, there is a high degree of crosstalk between neighboring elements (residues), then the local dependency structure greatly shrinks the *search space*. Previously, Yanover and Fromer [83] applied an MRF formalism for the prediction of low energy, protein side configurations, a relevant problem from several aspects of structural biology such as *de novo* protein

folding, homology modeling and protein-protein docking. The different types of local interactions among amino acid residues: hydrophobic, hydrophilic, charged, polar, etc.) modeled as pairwise potentials let to semi-empirical expressions for the potential energies used in the MRF formalism. Once explicit expressions for the field have been written, the authors resort to a belief-propagation algorithm to find the optimal solution to the MRF problem given the constraints. Several improvements were actually applied to the message-passing algorithm that allow the authors to find a method to obtain the lowest energy amino acid chain configurations. This kind of approach may also be relevant to improve solving methods of random fields in statistical physics problems since it led to approximate explicit forms of the partition function.

Improving methods to discern the structural properties of proteins are also quite used in the context of protein homology, i.e., to investigate on the functions of proteins related to their structural similarity to other proteins, perhaps in different organisms. Local homology relationships can also be investigated by means of Markov random field methods. Xu and collaborators developed a method (or better, a *family* of methods) called MRAlign for protein homology detection based on the alignment of MRFs [84, 85]. Aside from purely *Ising* approaches, other methods of random fields of statistical mechanics have been adopted in the computational biology community. One of them is the Potts model. Recently, Wilburn and Eddy used a Potts model with latent variables for the prediction of *remote* protein homology (involving changes such as insertions and deletions) [86] importance sampling from extensive databases was used to perform MAP optimization as commonly done in computational biology and computer science.

A topic related to homology, but also involving space-dependent electrostatic interactions (protein-protein interactions, in particular) is protein function prediction. Networked models of protein prediction have been developed: primitive models can be used to associate a function to a given protein given the functions of proteins in their interaction neighborhood and probabilistic models may do this by weighting interactions with an associated probability. Gehrman and collaborator devised a CRF method for protein function prediction based on these premises [87]. To solve the CRF, they resort to a factor graph approach [88] to write down explicit contributions to the cliques [89] and then using an approximate Gibbs measure calculated from this clique factorization. The approximation is based on other relevant feature of Markov random fields, which we will discuss later in the context of statistics and computer science: the use of the so-called *Gibbs sampler* or Gibbs sampling algorithm [90]. The Gibbs sampler is a Markov chain Monte Carlo (MCMC) method used to obtain a sequence of observations—approximated from a specified multivariate probability distribution—, in those cases for which direct sampling is difficult or even impossible (e.g., NP-hard or super-combinatorial problems).

Perhaps not so well known as a relevant structural biology problem until recently, is the determination of three dimensional chromosome structure inside the cell's nucleus. Long range

chromosomal interactions are believed to be ultimately related to fundamental issues on global and local gene regulation phenomena. A recently devised experimental method for global *chromosome conformation capture* is known as Hi-C. Nuclear DNA is subject to formaldehyde treatment to enhance covalent interactions *glueing* chromosome segments that are three dimensionally adjacent. Then a battery of restriction enzymes is used to cut DNA into pieces. Such pieces are sequenced and the identity of the spatially adjacent regions are then discovered. The data is noisy and often incomplete. For these reasons, a team lead by Yun Li developed a hidden Markov random field method to analyze Hi-C data to detect long range chromosomal interactions [91]. This method combines ideas from MRFs, Bayesian networks and Hidden Markov models. In a nutshell, they assumed a mixture of negative binomials as an Ising prior [22] and supplemented it with Bayesian inference to calculate the joint probabilities *via* a Metropolis-Hastings pseudo-likelihood approach.

Application Box II: Prediction of Low Energy Protein Side Chain Configurations

General problem statement: The prediction of energetically favorable aminoacid chain configurations constrained on the three-dimensional structure of a protein principal chain is a relevant problem in structural biology. Accurate side configuration predictions are key to develop approaches to *de novo* protein folding, to model protein homology and to study protein-protein docking. Yanover and Fromer [83] used a Markov Random Field with pairwise energy interactions supplemented with a belief propagation algorithm to bypass the mean field approximation.

Theoretical/Methodological approach: The authors developed their approach by modeling energy levels (as obtained by simulation and calorimetric techniques) as the relevant variables in a pairwise Markov Random Field. Since local side chain configurations have inhomogeneous contributions to the global energy landscape, a mean field approach will not be accurate. In order to circumvent the other extreme of modeling all detailed molecular interactions, the authors used belief propagation algorithm (BPA), a class of message passing method that performs global optimization (in this case energy minimization) by iterative local calculations between neighboring sites.

Improvements/advantages: We can consider the use of the BPA on top of the MRF, as a compromise between mean field approach (not useful to solve the actual structural biology problem) and full-detail molecular interaction modeling (computationally intractable due to the large combinatorial search space involved).

Limitations: Protein side chain prediction may in many cases be affected by subtle angular variations in the rotamer side chains. The authors have discussed that, to improve the accuracy of their predictions in such cases, it may be useful to resort to continuous-valued (Gaussian) MRFs with their associated BPAs as an avenue for further improvement within the current theoretical framework.

The spatial configuration of proteins within protein assemblies such as membranes it is also relevant to understand the functions

of molecular machines in the cell. By applying a combination of deep recurring neural networks and CRFs, it was possible to predict transmembrane topology and three dimensional coupling in the important family of G-protein coupled receptors (GPCRs). These receptors are able to detect molecules outside the cell and activate cellular responses and are of paramount relevance in immune responses and intercellular signaling [92].

As we have mentioned molecular regulatory networks are models that may conceptually map random fields almost straight forward. They have a graph-theoretical structure already and their interactions are often so complex that modeling them as stochastic dependencies is somehow natural [93]. Depending on the nature of the regulatory interactions to be modeled, different approaches can be followed. Gitter and coworkers, for instance, used latent tree models combining a MRF with a set of hidden (or latent) variables, factorizing the joint probability on a Markov tree [94]. In this work, the action of transcription factor (TFs) was mapped to a set of latent variables and the MRF was used to establish the relationships of conditional independence of groups of neighboring genes, *via* their gene expression patterns obtained from experimental data. Zhong and colleagues [95] used a related approach to infer regulatory networks *via* a *directed* random field, giving rise to a tree structure known as a directed acyclic graph (DAG). In their work, all variables follow a pairwise Markov field with conditional dependencies following parametric Gaussian or multinomial distributions. Although they resorted to a DAG modeling due to its ability to work with *mixed* data (usually undepowered for common MRF approaches), the limitations of these studies to account for regulatory loops has to be considered.

Application Box III: Inference of Tissue-specific Transcriptional Regulatory Networks

General problem statement: Transcriptional regulatory programs determine how gene expression is regulated, thus determining cellular phenotypes and response to external stimuli. Such gene regulatory programs involve a complex network of interactions among gene regulatory elements, RNA polymerase enzymes, protein complexes such as mediator and cohesion machineries and sequence specific transcription factors. Ma and coworkers [96] used a Markov Random Field approach to construct tissue-specific transcriptional regulatory networks integrating gene expression and regulatory sites data from RNA-seq and DNAase-Seq experiments.

Theoretical/Methodological approach: The authors developed an MRF approach with unary (node functions) and binary (edge functions, i.e., pairwise interactions) potentials for transcriptional interaction within a cell line and across cell lines, respectively. With these two potential functions a joint probability distribution is written. To solve the problem, the JPD is mapped to a pseudo-energy optimization (PEO) test *via* logarithmic transformation. The PEO is in turn transformed into a network maximum flow problem and solved by a loopy BPA.

Improvements/advantages: An original contribution of this work is the use of belief propagation algorithms to solve for a quadratic pseudo-energy functions (with only unary and

pairwise potentials) representation and then using iterated conditional modes. This may open an interesting research path for other MRF applications.

Limitations: One possible shortcoming of this approach is the use of linear correlation measures (Pearson coefficients) and linear classifiers (Singular Value Decomposition) for a problem with strong non-linearities (complex biochemical kinetics associated with gene expression). The MRF structure will indeed allow for more general statistical dependency relationships, making the analysis even more robust.

Undirected graphical models in the form of usual MRFs, have been used to construct, tissue-specific transcriptional regulatory networks [96] in 110 cell lines and 13 different tissues, from an integrative analysis of RNASeq and DNAase-Seq data. The authors used a method to minimize the pseudo-energy function by converting the problem to a maximum flow in networks and solving the latter *via* a loopy belief propagation algorithm [97].

To improve on the modeling capabilities of MRFs to describe gene regulatory networks (GRNs) it is becoming customary to include several data sources as a means to partially disambiguate the statistical dependency structures. Banf and Rhee implemented a data integration strategy to their MRF modeling of GRNs in an algorithm called GRACE which exploits the energy function based on unary and binary terms that we previously described in the context of MRF modeling in biological imaging. Low confidence pairwise interactions were removed by mapping the problem to a classification task on imbalanced sets, and following the tenets of Ridge penalized regression [98].

A somehow related method was devised by Grimes, Potter and Datta, who integrate differential network analysis to their study of gene expression data [99]. Their study was based on the idea of using KEGG pathways to construct MRFs as a means to functionally improve differential expression profiling [100, 101]. A similar MRF method was used to improve transcriptome analysis in model (mouse) systems for biomedical research [102]. Data integration can be also used to incorporate biological function information (from metabolic and signaling pathways) to the modeling of statistical Genome Wide Association Studies (GWAS) *via* MRFs [103]. The MRF was then solved by a combination of parametric (inverse gamma) distributed priors and MAP techniques to find the posterior probabilities. This is relevant since the important results of GWAS research in biomedicine (statistical in nature and often poorly informative in the biological sense) can be contextualized *via* pathway interactions as devised *via* this MRF approach.

Though not properly a molecular interaction network study, Long, et al, developed a method combining graph convolutional networks with conditional random fields, to predict human microbe-drug associations [104]. Since there has been a growing emphasis on the ways in which the human microbiome may affect drug responses in the context of precision medicine [105], accurate methods to predict such associations are highly desirable for the design of tailor-made therapeutic interventions.

Since random fields are able to capture not only spatio-temporal and regulatory associations, but are also proper to represent semantic or *grammatical* relationships, they have been thoroughly used in text analysis in biology, being the subjacent texts genomic sequences or pieces of biomedical literature. The group led by Fariselli used hidden CRFs for the problem of biosequence labeling in the prediction of the topology of prokaryotic outer-membrane proteins. Their study was based on a grammatically restrained approach, using dynamic programming much in the tradition of the so-called Boltzmann machines in AI [106]. Poisson random fields over sequence spaces were studied by Zhang and coworkers to detect local genomic signals in large sequencing studies [107].

Moving on to data and literature mining methods based on MRFs, we can mention *passage relevance models* used for the integration of syntactic and semantic elements to analyze biomedical concepts and topics *via* a PGM. The semantic components such as topics, terms and document classes are represented as potential functions of an MRF [108]. Biomedical literature mining strategies using MRFs were also developed to study automated recognition of bacteria named entities [109] to curate experimental databases on microbial interactions. Related methods were previously used to identify gene and protein mentions in the literature using CRFs [110].

4.3 Applications of MRFs in Ecology and Other Areas of Biology

Other applications of random fields in biology include demography and selection to study weakly deleterious genetic variants in complex demographic environments [111] and for species clustering [112], in population genetics. MRFs have also been applied to understand species distribution patterns and endemism and to unveil [113] interactions between co-occurring species in processes governing community assembly [114]; as well as for spatially explicit community occupancy [115] in ecology.

Another group of disciplines in which MRFs have flourished is comprised of Data Science, Computer Science and Modern statistics. The next section will be devoted to presenting and discussing some developments of random fields in that setting.

5 MARKOV RANDOM FIELDS IN DATA SCIENCE AND MACHINE LEARNING

The term *Data Science* refers to a multidisciplinary field devoted to extracting knowledge and insight from structured and unstructured data. It shares commonalities and differences with its parent fields: statistics, computer and information sciences and engineering. However, much of the emphasis is on the extraction of *useful* knowledge from data, putting accuracy and usability above formal mathematical structure if needed. Naturally, Markov random fields as a theoretically powerful methodology that allows for the incorporation of *educated intuition* and has an intrinsic algorithmic nature has called the attention of data scientists. We will present here, but a handful of

the many uses and implementations of MRFs in data science and computational intelligence settings. As we will see, these studies share a lot of commonalities with the applications in statistical physics and computational biology while, at the same time, incorporating elements that may cross-fertilize to the modeling schemes in the natural sciences.

5.1 Applications of MRFs in Computer Vision and Image Classification

As we already mentioned in the context of applications of random field to biomedical imaging, segmentation and pattern identification to enhance the resolution of spatial and/or spatio-temporal maps is a common use of MRFs. From the many applications in the field of computerized image processing, we will discuss some that present peculiarities or distinctive features that may be of more general interest. For instance, to face the challenge of capturing three dimensional structure from two-dimensional images, the so-called *depth perception*, Kozik used an MRF-based methodology [116] in which the energy function was modeled *via* a polynomial regression model and a depth estimation algorithm with correlated uncertainties (a sort of twofold autoregressive model). By using these entries Kozik then solved an MAP problem to obtain the maximum likelihood solution to the MRF.

In the context of AI to enhance low-resolution images (the super-resolution problem), Stephenson and Chen devised an adaptive MRF method [117] based on passing-message optimization by a loopy propagation algorithm. Also in the context of AI approaches to image processing Li and Wand developed a combination of MRFs as generative models and deep CNNs to discriminate two-dimensional images to try to solve the so-called *image synthesis problem*, a relevant problem in computer vision with applications both to photo-editing and neuroscience [118]. A problem related to image synthesis is image classification, in which certain features of images are discerned and used to cluster images by similitudes in these feature spaces. Applications in image recognition in security, forensics and scientific microscopy and imaging among others abound. To improve the accuracy of image classification algorithms, Wen and coworkers developed a CRF method in which machine-learned feature functions took the place of the unary and binary terms in the potential energy [119], as in previous cases Gaussian priors and loopy belief propagation algorithms were used to solve the random field.

5.2 Applications of MRFs in Statistics and Geostatistics

Geostatistics and geographical information systems are also quite amenable to be modeled within the MRF paradigm due to their natural spatio-temporal dependency structures. In the context of prediction of environmental risks and the effects of limited sampling, Bohorquez and colleagues developed an approach based on multivariate functional random fields for the spatial prediction of functional features at unsampled locations by resorting to covariates [120]. As in the case of random field

hydrodynamics (mentioned in the physics section), an empirical approach based on continuous field estimators was chosen. Continuous spatio-temporal correlation structures *via* so-called Kriging methods extending the ideas of discrete random fields are commonly used in environmental analysis and risk assessment [121, 122].

Geological modeling is another field at the intersection of geostatistics and geophysics which has adopted the MRF formalism to deal with their problems. A segmentation approach was used for stochastic geological modeling with the use of hidden MRFs [123]. Using a methodological approximation similar to the one used in computer vision and biomedical imaging, latent variable MRFs are used to perform three-dimensional segmentation. The model is supplemented with finite Gaussian mixture models for the parameter calculations and a Gibbs sampling inference framework, following a similar approach to the one developed by the group of Li [124], based on the methods of Rue and Held [125] and by Solberg et al [126] and further developed by Toftaker and Tjelmeland [127]. More refined geostatistical methods have been based on a clever combination of several developments of Markov random field theory. Along these lines, the work by Reuschen, Xu and Nowak [128] is noteworthy, since they used Bayesian inversion (based on Markov conditional independence) to develop a random field approach to hierarchical geostatistical models and used Gibbs sampling MCMC to solve them.

The combined use of ideas from Markov and Gibbs random fields in statistical learning and other approaches in modern statistics has indeed become a fruitful line of research with important theoretical developments and a multitude of applications [24, 34, 129]. The use of MRFs and CRFs as tools for statistical learning has been used in a multitude of settings in both generative and discriminative models [33]. Aside Ising models and MRFs, perhaps the most widely used applications of the random fields are Gibbs sampling and Markov chain Monte Carlo methods that we already mentioned. Due to the generality and the relatively low computational complexity of these sampling/simulation methods, several methods have been developed based on them.

Gibbs sampling is a form of Markov chain Monte Carlo (MCMC) algorithm. MCMC methods are used to obtain a sequence of *observations* of a random experiment by an approximation from a given (specified) multivariate probability distribution when direct sampling is challenging (computationally or otherwise). The essence of the method is building a Markov chain whose equilibrium distribution is precisely the specified multivariate distribution. Then, a sample of such distribution is just a sequence of states of the Markov chain. The use of the Markov property of an MRF allows to use Gibbs sampling as an MCMC method, when the joint probability distribution is not known (or is very complex) but the conditional distributions are known (or easier). Due to this, by using the pairwise Markov property, Gibbs sampling is particularly fit to sampling the posterior distribution of Bayesian networks (understood as a collection of conditional distributions), a quite relevant problem in both,

statistical learning and in large computer simulation problems.

Aside from these basic issues, Gibbs sampling has been extensively enhanced over the years. One important improvement has been the incorporation of adaptive rejection sampling [130, 131], particularly useful for situations in which evaluation of the density distribution function is computationally expensive (e.g., non-conjugated Bayesian models). Adaptive rejection sampling can be even applied to modeling *via* non-linear mixed models [131]. To further minimize the computational burden of Gibbs sampling, Meyer and collaborators [132] developed an algorithm which samples *via* Lagrange interpolation polynomials, instead of exponential distributions. Convergence can be also improved by double-adaptive independent rejection sampling [133] which is based on a scheme of minimizing the correlation among samples. Gibbs sampling approaches also allow for the determination of dense distribution simulated sampling from sparse sampled data [134], even in high dimensional latent fields over large datasets [135].

Gao and Gormley implemented a Gibbs sampling scheme based on CRFs weighted *via* neural scoring factors (implemented as parameters in factor graphs) with applications to Natural Language Processing (NLP) [136]. MCMC has also been used, in the context of Gibbs random fields in data pre-processing, to reduce the computational burden of data intensive signal processing [137, 138]. Gibbs sampling can also be applied in parallel within the context of Gaussian MRFs on large grids or lattice models [139]. Parallel Gibbs sampling methods can also be developed in the context of sampling acceleration for structured graphs [140].

Markov random fields and its associated Gibbs measures can also be used to advance statistical methods in large deviation theory [141] and to develop methods of joint probability decomposition based on product measures [142]. Exact factorizability of joint probability distributions is a most relevant question in modern probability [143–146] with important applications in data analytics [147], applied mathematics [148], computational biology [149] and network science [150], among other fields. MRFs also have been applied to embed filtrations on high dimensional hyperparameter spaces. The main idea is using random fields as hierarchical models *projecting* the relevant hyper-parameter space to a lower dimensional filtration [135]. This general problem is closely related with the *feature selection problem* in computer science and data analytics. We will discuss applications of the MRF formalism in that context in the next subsection.

5.3 Applications of MRFs in Feature Selection and AI

Feature selection (FS) refers to a quite general class of problems in computer science, data analysis and AI. Feature selection aims to find the minimum number of maximal relevant features to characterize a high dimensional data set. One outstanding family of methods of feature selection is *regression methods* in which a set of *regression variables* is used to predict one (or a few) dependent variables *via* functional relationships (commonly

linear combinations with a distribution of weights). A subset of the whole set of regression variables is considered statistically significant, in that context those are the *selected features*. FS is a more general problem than linear, multivariate or even non-linear regression. MRF can be used to generalize regression procedures to more complex situations. One notable method was developed by Stoehr, Marin and Pudio [151] who used hidden Gibbs random fields to implement model selection *via* an information theoretical optimization criterion known as *Block likelihood information*. Cilla and coworkers [152] developed a FS method to be used in sequence classification based on hidden CRFs supplemented with a generalized Lasso group regularization method that instead of the colinearity condition employs L1-norm optimization of the parameters. The authors showed that FS outcomes with this method outperforms standard conditional random field approaches.

Feature selection efficacy of MRFs is closely related to the actual structure of the underlying adjacency matrices. Especially relevant is the issue of *separability*. Although non-trivial separability does not preclude the use of MRFs in large datasets, as long as the positive definite nature of the measures is ensured; there may be computational complexity limitations for practical uses. Recently, Sain and Furrer [153] discussed on some general properties of random fields (in particular for multivariate Gaussian MRFs) that need to be taken into account in the design of computationally efficient modeling strategies with such random fields. By designing FS schemes with MRFs based on the optimization of parameter estimation, for instance *via structured learning* it is possible to improve substantially on the computational complexity of such algorithms [154–158]. The graph structure of MRFs can also be optimized to enhance the FS capabilities of the algorithms [159–163]. More information along these lines can be found in the comprehensive review by Adams and Beling [164] and in the one by Vergara and Estevez [165].

As already mentioned the structure of MRF may result advantageous to solve *segmentation* problems or delimitation of statistical dependencies. These are problems that are extremely relevant in the context of computational linguistics and natural language processing applications. We will discuss these in the following subsection.

5.4 Applications of MRFs in Computational Linguistics and NLP

Automated textual identification and *meaning discernment* are extremely complex (and very useful) tasks in current artificial intelligence research and applications. The ability to detect text *patches* with semantic similarity is one of the founding steps in the ability to process natural language by a computer. By combing a deep learning approach (a convolutional neural network) with MRF models, Liu and collaborators [166] devised an effective algorithm for *semantic segmentation* [167], which they called a Deep Parsing Network (DPN). Within the DPN scheme, a CNN is used to calculate the unary terms of a two-term energy function, while the pairwise terms were approximated with a mean-field model. The mean field contributions were iteratively optimized

using a back-propagation algorithm able to generalize to higher order *perturbative contributions*. Although the original application of semantic segmentation has been applied to image segmentation, its applications to NLP are somehow straight forward [168, 169].

A similar method was developed earlier by Mai, Wu and Cui and applied to improve word segmentation disambiguation in the Chinese language [170]. Main and colleagues, however, decided to use a CRF on top of a bidirectional maximum matching algorithm. Parameter estimation for the CRF was performed *via* maximum likelihood estimates. These ideas were further advanced by Qiu, et al [171] who used CRFs for clinical entity recognition in Chinese. Speech tagging from voice recordings was performed using a CRF devised by Khan and collaborators [172]. Even computer assisted *fake news* detection [173] and headline prediction [174] can be achieved using CNNs and MRFs.

5.5 Applications of MRFs in the Analysis of Social Networks

Social network analysis, including online social networks, other forms of interpersonal interaction networks and even some social networks in non-human creatures, have become a relevant field of research in recent times (though the subject has been relevant in the contexts of sociology and animal behavior for decades) [175]. The analysis of social network *via* MRFs is becoming more and more common also. As an example, Jia and collaborators have used MRFs to infer attributes in online social network data [176]. Their model used the social network structure itself to develop a pairwise MRF. From empirical training data, the authors used the individual behaviors to learn a probability that each user has a given attribute. Then used that as an a priori probability, compute the posterior probabilities by a loopy belief propagation algorithm over the MRF, to, finally, optimizing the belief propagation algorithm by a *second neighbor* criteria that sparsifies the adjacency matrix. Further optimization of similar ideas was obtained by using graph convolutional networks, i.e., CNNs over CRFs [177]. Attribute inference in social network data *via* MRFs can also be used to improve cybersecurity algorithms [178], to learn consumer intentions [179], to study the epidemiology of depression [180] among other issues. Social networks as well as some classes of molecular interaction and ecological networks are also relevant to the development and improvement of MRF and CRF learning algorithms. This is so since often a sketch (sometimes a detailed one) of the network dependency structure is known a priori [181, 182]. This is yet another instance in which applications may nurture back the formal theory of random fields.

Application Box IV: Inference of User Attributes in Online Social Networks

General problem statement: The attribute inference problem (AIP), i.e., the discovery of personality traits from data on social networks, is a central question on computational social science. It is indeed an (unsupervised) extension of the personality analysis tests of classical psychology with important applications from sociological modeling to

commercial and political marketing, and even national security issues. Jia and collaborators [176] developed an approach to the AIP from public data on online social networks using an MRF with pairwise interactions.

Theoretical/Methodological approach: Given a training dataset, behaviors are used to learn the probabilities that each user (node) has a considered attribute, these are the prior probabilities. Based on the neighborhood structure of a pairwise Markov random field, posterior probabilities are computed *via* a loopy belief propagation algorithm. The MRF has a quadratic pseudo-energy function with node potentials (unary contributions) for each user and edge potentials (pairwise interactions) for every connected pair of nodes, as defined by node correlations. Edge potentials are defined as discrete-valued *spin-like* states $\lambda_{uv} = 1$ if nodes u and v have the same attribute state and $\lambda_{uv} = -1$ if they do not. This way, *homophily* in the social networks mimics spin-alignment in lattice models of magnetism.

Improvements/advantages: To optimize computational performance in large networks, the authors modified the BPA by using a loop renormalization strategy. Hence, circular node correlations are locally computed for each pair of nodes *prior to move to another edge* and then using a linear optimization approach. Thus, there is no need to allocate memory for all circular correlations (loops).

Limitations: More than a limitation itself, an avenue of predictive improvement may be given by extending their MRF approach to allow multi-categorical (or even continuous) state variables. Doing this will make possible to capture the fact that most behavioral attributes are not simply present/absent, but may occur over a range of possibilities.

5.6 Random Fields and Graph Signal Theory

Graph signal theory, also called graph signal processing (GSP) is a field of signal analytics that deals with signals whose domain (as identified by a graph) is irregular [183–185]. In the context of GSP, the vertices or nodes represent probes in which the signal has been evaluated or sensed and the edges are relationships between these vertices. Data processing of the signals exploits the structure of the associated graph. GSP is often seen as an intermediate step between single channel signal processing and spatio-temporal signal analysis. The nature of the edges is determined by the relationship (spatial, contextual, relational, etc.) between the vertices. Whenever edges are defined *via* a statistical dependence structure, GSP can be mapped to either an MRF or a CRF, thus allowing the use of all the tools of random field theory to perform GSP [186, 187]. The networked nature of the domain of signals embedded in a graph, allows the use of spectral graph theoretical methods for signal processing [188–190]. Conversely, correlations between features on the signals are also useful to identify the structure of the underlying graph [191, 192].

GSP has a number of relevant applications, from spatio-temporal analysis of brain data [193]; to analyze vulnerabilities in power grid data [194]; to topological data analysis [195], chemoinformatics [196] and single cell transcriptomic analysis [197], to mention but a few examples. Statistical learning

techniques have also being founded on a combination of MRFs and GSP [198, 199], taking advantage of both the networked structure, the statistical dependence relationships and the temporal correlations of the signals [200–202]. Random field approaches to GSP have also been applied in the context of deep convolutional networks [203, 204], often invoking features of the underlying joint conditional probability distributions such as ergodicity [205] and stationarity [206].

6 CONCLUDING REMARKS

As already known in statistical physics for decades, random fields are a quite powerful and versatile theoretical analytical framework. We have discussed here some fundamental ideas of the theory of Markov-Gibbs random fields, namely the notions of statistical dependency on neighborhoods, of potentials and local interactions, of conditional independence relationships and so on. After that, we discussed a handful of (mostly recent) advances and applications of Markov random fields in different physics subdisciplines, as well as in several areas of biology and the data sciences. The main goal of this presentation was not to be comprehensive but to be illustrative of the many ways in which research and applications of random field may be advancing both, inside and outside traditional statistical physics.

In the theoretical and conceptual advances side, we mentioned how random fields may be embedded in general manifolds, how by incorporating quenched fields (or somehow equivalently, by adding quenching potentials) to the usual Ising random field, a whole new phenomenology can be discovered in RFIMs. How Markov and Bayesian networks may be combined in HRFs and how gauge symmetries and other extended fields may broad the scope of MRFs.

By examining the applications in physics and in other disciplines, we discover (or often re-discover) methodological and computational improvements to the inference, analysis and solutions of problems within the MRF/GRF/CRF settings. In these regards, we can mention the use of CNNs as feature extractors on top of random fields, to refine hypotheses about marginals and (*via* convolution) to improve the accuracy of pairwise potential terms. We re-examined how to extend beyond mean-field approaches, either *via* MAP optimization, *via* higher order perturbations solved by neural networks or maximum likelihood approaches (depending on data availability). How, under certain circumstances (still dictated by physical intuition and data constrains) factorization of the partition function may be attained *via* clique potentials obtained from Gaussian (or other multivariate parametric distributions) or even from empirical distributions.

We also analyzed how simulations in random fields may be supplemented with well known methods—within the statistical physics community—, such as simulated annealing, Markov Chain Monte Carlo and importance sampling, but also from methods of wide use in other fields such as stochastic descent back-propagation, factor graph approaches, Gibbs sampling, pseudo-likelihood methods, latent models or loopy belief

propagation algorithms to name a few. And how, under some circumstances, parameter estimation (fundamental in applications involving non-trivial partition functions) can be reframed as a regression problem and benefit from the use of the Ridge and Lasso optimization techniques, dynamic programming and autoregressive modeling.

We want to highlight that, in spite of being a hundred-plus year developed formalism in statistical physics, the theory of Markov-Gibbs random fields is indeed a flourishing one, with many theoretical advances and applications within and outside physics.

AUTHOR CONTRIBUTIONS

EH performed research and wrote the manuscript.

REFERENCES

- Ising E. Beitrag zur theorie des ferromagnetismus. *Z Physik* (1925) 31:253–8. doi:10.1007/bf02980577
- Averintsev MB. Description of Markovian random fields by gibbsian conditional probabilities. *Theor Probab Appl* (1972) 17:20–33. doi:10.1137/1117002
- Averintsev M. Gibbsian distribution of random fields whose conditional probabilities may vanish. *Problemy Peredachi Informatsii* (1975) 11:86–96.
- Dobrushin RL, Kryukov V, and Toom AL. *Locally interacting systems and their application in biology*. Springer (1978).
- Stavskaya ON. Markov fields as invariant states for local processes. *Locally Interacting Systems and Their Application in Biology*. Springer (1978). 113–121. doi:10.1007/bfb0070088
- Stavskaya ON. Sufficient conditions for the uniqueness of a probability field and estimates for correlations. *Math Notes Acad Sci USSR* (1975) 18:950–6. doi:10.1007/bf01153051
- Vasilyev NB. Bernoulli and Markov stationary measures in discrete local interactions. *Locally interacting systems and their application in biology*. Springer (1978). 99–112. doi:10.1007/bfb0070087
- Dobruschin PL. The description of a random field by means of conditional probabilities and conditions of its regularity. *Theor Probab Appl* (1968) 13:197–224. doi:10.1137/1113026
- Lanford OE, and Ruelle D. Observables at infinity and states with short range correlations in statistical mechanics. *Commun Math Phys* (1969) 13:194–215. doi:10.1007/bf01645487
- Hammersley JM, and Clifford P. *Markov fields on finite graphs and lattices*. Unpublished manuscript (1971).
- Koller D, and Friedman N. *Probabilistic graphical models: principles and techniques (adaptive computation and machine learning series)*. MIT Press (2009).
- Grimmett GR. A theorem about random fields. *Bull Lond Math Soc* (1973) 5:81–84. doi:10.1112/blms/5.1.81
- Besag J. Spatial interaction and the statistical analysis of lattice systems. *J R Stat Soc Ser B (Methodological)* (1974) 36:192–225. doi:10.1111/j.2517-6161.1974.tb00999.x
- Baxter RJ. *Exactly solved models in statistical mechanics*. Elsevier (2016).
- Cipra BA. An introduction to the Ising model. *The Am Math Monthly* (1987) 94:937–59. doi:10.1080/00029890.1987.12000742
- McCoy BM, and Wu TT. *The two-dimensional Ising model*. North Chelmsford, MA: Courier Corporation (2014).
- Thompson CJ. *Mathematical statistical mechanics*. Princeton University Press (2015).
- Adler M. *Monte Carlo simulations of the Ising model*. Anchor Academic Publishing (2016).
- Hopfield JJ. Neural networks and physical systems with emergent collective computational abilities. *Proc Natl Acad Sci* (1982) 79:2554–8. doi:10.1073/pnas.79.8.2554
- Ackley DH, Hinton GE, and Sejnowski TJ. A learning algorithm for Boltzmann machines*. *Cogn Sci* (1985) 9:147–69. doi:10.1207/s15516709cog0901_7
- Salakhutdinov R, and Larochelle H. Efficient learning of deep Boltzmann machines. Proceedings of the thirteenth international conference on artificial intelligence and statistics. Sardinia, Italy: DBLP (2010) 693–700.
- Ross KJ, and Snell L. *Markov random fields and their applications*. American Mathematical Society (1980).
- Essler FHL, Mussardo G, and Panfil M. Generalized Gibbs ensembles for quantum field theories. *Phys Rev A* (2015) 91:051602. doi:10.1103/physreva.91.051602
- Murphy K. *Machine learning: a probabilistic perspective*. MIT Press (2012).
- Williams D. *Probability with martingales*. Cambridge University Press (1991).
- Canonon CL, Diakonikolas I, Kane DM, and Stewart A. *Testing conditional independence of discrete distributions*. Information Theory and Applications Workshop (ITA) (IEEE) (2018). 1–57.
- Schultz TD, Mattis DC, and Lieb EH. Two-dimensional Ising model as a solvable problem of many fermions. *Rev Mod Phys* (1964) 36:856. doi:10.1103/revmodphys.36.856
- Brush SG. History of the lenz-ising model. *Rev Mod Phys* (1967) 39:883. doi:10.1103/revmodphys.39.883
- Isichenko MB. Percolation, statistical topography, and transport in random media. *Rev Mod Phys* (1992) 64:961. doi:10.1103/revmodphys.64.961
- Sornette D. Physics and financial economics (1776-2014): puzzles, Ising and agent-based models. *Rep Prog Phys* (2014) 77:062001. doi:10.1088/0034-4885/77/6/062001
- Adler RJ, and Taylor JE. *Random fields and geometry*. Springer Science & Business Media (2009).
- Ganchev A. About Markov, Gibbs, . . . gauge theory. . . finance. *Quantum Theory And Symmetries*. Springer (2017) 403–12.
- Freno A, and Trentin E. *Hybrid random fields*. Springer (2011).
- Friedman J, Hastie T, and Tibshirani R. *The elements of statistical learning*. New York: Springer series in Statistics (2001).
- Hernández-Lemus E. On a class of tensor Markov fields. *Entropy* (2020) 22:451. doi:10.3390/e22040451
- Hernández-Lemus E, Espinal-Enriquez J, and de Anda-Jáuregui G. *Probabilistic multilayer networks*. Ithaca, NY: arXiv:1808.07857 (2018).
- De Domenico M, Solé-Ribalta A, Cozzo E, Kivela M, Moreno Y, Porter MA, et al. Mathematical formulation of multilayer networks. *Phys Rev X* (2013) 3:041022. doi:10.1103/physrevx.3.041022
- Kivela M, Arenas A, Barthelemy M, Gleeson JP, Moreno Y, and Porter MA. Multilayer networks. *J Complex Networks* (2014) 2:203–71. doi:10.1093/comnet/cnu016

FUNDING

This work was supported by the Consejo Nacional de Ciencia y Tecnología (SEP-CONACYT-2016-285544 and FRONTERAS-2017-2115), and the National Institute of Genomic Medicine, México. Additional support has been granted by the Laboratorio Nacional de Ciencias de la Complejidad, from the Universidad Nacional Autónoma de México. EH is recipient of the 2016 Marcos Moshinsky Fellowship in the Physical Sciences.

ACKNOWLEDGMENTS

The author is grateful to the lively and brilliant academic community that has been behind the Winter Meeting on Statistical Physics for five decades now.

39. Boccaletti S, Bianconi G, Criado R, Del Genio CI, Gómez-Gardeñes J, Romance M, et al. The structure and dynamics of multilayer networks. *Phys Rep* (2014) 544:1–122. doi:10.1016/j.physrep.2014.07.001
40. Aizenman M, and Peled R. A power-law upper bound on the correlations in the 2d random field Ising model. *Commun Math Phys* (2019) 372:865–92. doi:10.1007/s00220-019-03450-3
41. Imry Y, and Ma S-K. Random-field instability of the ordered state of continuous symmetry. *Phys Rev Lett* (1975) 35:1399. doi:10.1103/physrevlett.35.1399
42. Berzin AA, Morosov AI, and Sigov AS. Long-range order induced by random fields in two-dimensional O(n) models, and the imry-ma state. *Phys Solid State* (2020) 62:332–7. doi:10.1134/s1063783420020055
43. Berzin AA, Morosov AI, and Sigov AS. A mechanism of long-range order induced by random fields: effective anisotropy created by defects. *Phys Solid State* (2016) 58:1846–9. doi:10.1134/s1063783416090109
44. Bunde A, Havlin S, Roman HE, Schildt G, and Stanley HE. On the field dependence of random walks in the presence of random fields. *J Stat Phys* (1988) 50:1271–6. doi:10.1007/bf01019166
45. Chatterjee S. On the decay of correlations in the random field Ising model. *Commun Math Phys* (2018) 362:253–67. doi:10.1007/s00220-018-3085-0
46. Aizenman M, and Wehr J. Rounding of first-order phase transitions in systems with quenched disorder. *Phys Rev Lett* (1989) 62:2503. doi:10.1103/physrevlett.62.2503
47. Fytas NG, Martin-Mayor V, Picco M, and Sourlas N. Specific-heat exponent and modified hyperscaling in the 4d random-field Ising model. *J Stat Mech* (2017) 2017:033302. doi:10.1088/1742-5468/aa5dc3
48. Fytas NG, Martin-Mayor V, Picco M, and Sourlas N. Review of recent developments in the random-field Ising model. *J Stat Phys* (2018) 172:665–72. doi:10.1007/s10955-018-1955-7
49. Tarjus G, and Tissier M. Random-field Ising and o (n) models: theoretical description through the functional renormalization group. *The Eur Phys J B* (2020) 93:1–19. doi:10.1140/epjb/e2020-100489-1
50. Ayala M, Carinci G, and Redig F. Quantitative Boltzmann-gibbs principles via orthogonal polynomial duality. *J Stat Phys* (2018) 171:980–99. doi:10.1007/s10955-018-2060-7
51. Dobrushin RL. Perturbation methods of the theory of gibbsian fields. Lectures on probability theory and statistics. Springer (1996) 1–66. doi:10.1007/bfb0095674
52. Essler FHL, Mussardo G, and Panfil M. On truncated generalized Gibbs ensembles in the Ising field theory. *J Stat Mech* (2017) 2017:013103. doi:10.1088/1742-5468/aa53f4
53. Gudder SP. Gaussian random fields. *Found Phys* (1978) 8:295–302. doi:10.1007/bf00715214
54. Sherman S. Markov random fields and Gibbs random fields. *Isr J Math* (1973) 14:92–103. doi:10.1007/bf02761538
55. Luitz DJ, Laflorencie N, and Alet F. Many-body localization edge in the random-field heisenberg chain. *Phys Rev B* (2015) 91:081103. doi:10.1103/physrevb.91.081103
56. Starodubov SL. *A theorem on properties of sample functions of a random field and generalized random fields*. Moscow, Russia: Izvestiya Vysshikh Uchebnykh Zavedenii. Matematika (2011) 48–56.
57. Acar P, and Sundararaghavan V. A Markov random field approach for modeling spatio-temporal evolution of microstructures. *Model Simul Mater Sci Eng* (2016) 24:075005. doi:10.1088/0965-0393/24/7/075005
58. Koninckx T, and Krakoviack V. Dynamics of fluids in quenched-random potential energy landscapes: a mode-coupling theory approach. *Soft matter* (2017) 13:5283–97. doi:10.1039/c7sm00984d
59. Liu Y, Hu J, Wei H, and Saw A-L. A direct simulation algorithm for a class of beta random fields in modelling material properties. *Comput Methods Appl Mech Eng* (2017) 326:642–55. doi:10.1016/j.cma.2017.08.001
60. Chen J, He J, Ren X, and Li J. Stochastic harmonic function representation of random fields for material properties of structures. *J Eng Mech* (2018) 144:04018049. doi:10.1061/(asce)em.1943-7889.0001469
61. Singh R, and Adhikari R. Fluctuating hydrodynamics and the brownian motion of an active colloid near a wall. *Eur J Comput Mech* (2017) 26:78–97. doi:10.1080/17797179.2017.1294829
62. Yamazaki K. Stochastic hall-magneto-hydrodynamics system in three and two and a half dimensions. *J Stat Phys* (2017) 166:368–97. doi:10.1007/s10955-016-1683-9
63. Ullah H, Uzair M, Ullah M, Khan A, Ahmad A, and Khan W. Density independent hydrodynamics model for crowd coherency detection. *Neurocomputing* (2017) 242:28–39. doi:10.1016/j.neucom.2017.02.023
64. Tadić B, Mijatović S, Janičević S, Spasojević D, and Rodgers GJ. The critical barkhausen avalanches in thin random-field ferromagnets with an open boundary. *Scientific Rep* (2019) 9:1–13. doi:10.1038/s41598-019-42802-w
65. Tsukanov AA, and Gorbatnikov AV. Influence of embedded inhomogeneities on the spectral ratio of the horizontal components of a random field of Rayleigh waves. *Acoust Phys* (2018) 64:70–6. doi:10.1134/s1063771018010189
66. Shadaydeh M, Guaniche Y, and Denzler J. *Classification of spatiotemporal marine climate patterns using wavelet coherence and markov random field*. American Geophysical Union (2018). Fall Meeting 2018IN31C–0824.
67. Feng R, Luthi SM, Gisolf D, and Angerer E. Reservoir lithology determination by hidden Markov random fields based on a Gaussian mixture model. *IEEE Trans Geosci Remote Sensing* (2018) 56:6663–73. doi:10.1109/tgrs.2018.2841059
68. Wang H, Wellmann F, Verweij E, von Hebel C, and van der Kruk J. *Identification and simulation of subsurface soil patterns using hidden markov random fields and remote sensing and geophysical emi data sets*. Vienna, Austria: EGUGA (2017) 6530.
69. Ko GG, and Rutenbar RA. A case study of machine learning hardware: real-time source separation using Markov random fields via sampling-based inference. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. (IEEE) (2017) 2477–81.
70. Li J, Jiang P, and Zhu H. *A local region-based level set method with markov random field for side-scan sonar image multi-level segmentation*. *IEEE Sensors Journal* (2020).
71. Ziatdinov M, Maksov A, and Kalinin SV. Learning surface molecular structures via machine vision. *npj Comput Mater* (2017) 3:1–9. doi:10.1038/s41524-017-0038-7
72. Ciliberto C, Herberster M, Ialongo AD, Pontil M, Rocchetto A, Severini S, et al. Quantum machine learning: a classical perspective. *Proc R Soc A* (2018) 474:20170551. doi:10.1098/rspa.2017.0551
73. Li Y, and Ping W. *Cancer metastasis detection with neural conditional random field*. Ithaca, NY: arXiv:1806.07064 (2018).
74. Zhang Z, Zhao T, Gay H, Zhang W, and Sun B. Arpm-net: a novel cnn-based adversarial method with Markov random field enhancement for prostate and organs at risk segmentation in pelvic ct images. *Med Phys* (2020). doi:10.1002/mp.14580
75. Fu H, Xu Y, Lin S, Kee Wong DW, and Liu J. Deepvessel: retinal vessel segmentation via deep learning and conditional random field. *International conference on medical image computing and computer-assisted intervention*. Springer (2016) 132–9. doi:10.1007/978-3-319-46723-8_16
76. Orlando JJ, Prokofyeva E, and Blaschko MB. A discriminatively trained fully connected conditional random field model for blood vessel segmentation in fundus images. *IEEE Trans Biomed Eng* (2016) 64:16–27. doi:10.1109/TBME.2016.2535311
77. Reta C, Gonzalez J, Diaz R, and Guichard J. *Leukocytes segmentation using markov random fields*. *Software Tools and Algorithms for Biological Systems*. Springer (2011) 345–53.
78. Hahn A, Bode J, Krüwel T, Kampf T, Buschle LR, Sturm VJF, et al. Gibbs point field model quantifies disorder in microvasculature of u87-glioblastoma. *J Theor Biol* (2020) 494:110230. doi:10.1016/j.jtbi.2020.110230
79. Mahmood F, and Durr NJ. Deep learning and conditional random fields-based depth estimation and topographical reconstruction from conventional endoscopy. *Med Image Anal* (2018) 48:230–43. doi:10.1016/j.media.2018.06.005
80. Sari NLK, Prajitno P, Lubis LE, and Soejoko DS. Computer aided diagnosis (cad) for mammography with Markov random field method with simulated annealing optimization. *J Med Phys Biophys* (2017) 4:84–93.
81. Nitzken MJ, El-Baz AS, and Beache GM. Markov-gibbs random field model for improved full-cardiac cycle strain estimation from tagged cmr. *J Cardiovasc Magn Reson* (2012) 14:1–2. doi:10.1186/1532-429x-14-s1-p258
82. Johansen AR, Sønderby CK, Sønderby SK, and Winther O. Deep recurrent conditional random field network for protein secondary prediction. *Proceedings of the 8th ACM international conference on bioinformatics*,

- computational biology, and health informatics, Boston, MA: ACM-BCB '17 (2017) 73–8.
83. Yanover C, and Fromer M. Prediction of low energy protein side chain configurations using Markov random fields. *Bayesian Methods in Structural Bioinformatics*. Springer (2012) 255–84. doi:10.1007/978-3-642-27225-7_11
 84. Xu J, Wang S, and Ma J. *Protein homology detection through alignment of markov random fields: using MRAlign*. Springer (2015).
 85. Ma J, Wang S, Wang Z, and Xu J. Mrfalign: protein homology detection through alignment of Markov random fields. *Plos Comput Biol* (2014) 10: e1003500. doi:10.1371/journal.pcbi.1003500
 86. Wilburn GW, and Eddy SR. Remote homology search with hidden potts models. *Plos Comput Biol* (2020) 16:e1008085. doi:10.1371/journal.pcbi.1008085
 87. Gehrmann T, Loog M, Reinders MJT, and de Ridder D. Conditional random fields for protein function prediction. *IAPR International Conference on Pattern Recognition in Bioinformatics*. Springer (2013) 184–95. doi:10.1007/978-3-642-39159-0_17
 88. Loeliger H-A, Dauwels J, Hu J, Korl S, Ping L, and Kschischang FR. The factor graph approach to model-based signal processing. *Proc IEEE* (2007) 95: 1295–322. doi:10.1109/jproc.2007.896497
 89. Ray WC, Wolock SL, Callahan NW, Dong M, Li QQ, Liang C, et al. Addressing the unmet need for visualizing conditional random fields in biological data. *BMC bioinformatics* (2014) 15:202. doi:10.1186/1471-2105-15-202
 90. Geman S, and Geman D. *Stochastic relaxation, gibbs distributions, and the bayesian restoration of images*. *IEEE Transactions on pattern analysis and machine intelligence* (1984) 721–41.
 91. Xu Z, Zhang G, Jin F, Chen M, Furey TS, Sullivan PF, et al. A hidden Markov random field-based bayesian method for the detection of long-range chromosomal interactions in hi-c data. *Bioinformatics* (2016) 32:650–6. doi:10.1093/bioinformatics/btv650
 92. Wu H, Wang K, Lu L, Xue Y, Lyu Q, and Jiang M. Deep conditional random field approach to transmembrane topology prediction and application to gpcr three-dimensional structure modeling. *Ieee/acm Trans Comput Biol Bioinform* (2016) 14:1106–14. doi:10.1109/TCBB.2016.2602872
 93. Kordmahalleh MM, Sefidmazi MG, Harrison SH, and Homaifar A. Identifying time-delayed gene regulatory networks via an evolvable hierarchical recurrent neural network. *BioData mining* (2017) 10:29. doi:10.1186/s13040-017-0146-4
 94. Gitter A, Huang F, Valluvan R, Fraenkel E, and Anandkumar A. *Unsupervised learning of transcriptional regulatory networks via latent tree graphical models*. Ithaca, NY: arXiv:1609.06335 (2016).
 95. Zhong W, Dong L, Poston TB, Darville T, Spracklen CN, Wu D, et al. Inferring regulatory networks from mixed observational data using directed acyclic graphs. *Front Genet* (2020) 11:8. doi:10.3389/fgene.2020.00008
 96. Ma S, Jiang T, and Jiang R. Constructing tissue-specific transcriptional regulatory networks via a Markov random field. *BMC genomics* (2018) 19: 65–77. doi:10.1186/s12864-018-5277-6
 97. Kolmogorov V, and Zabih R. What energy functions can be minimized via graph cuts?. *IEEE Trans Pattern Anal Machine Intell* (2004) 26:147–59. doi:10.1109/tpami.2004.1262177
 98. Banf M, and Rhee SY. Enhancing gene regulatory network inference through data integration with Markov random fields. *Scientific Rep* (2017) 7:1–13. doi:10.1038/srep41174
 99. Grimes T, Potter SS, and Datta S. Integrating gene regulatory pathways into differential network analysis of gene expression data. *Scientific Rep* (2019) 9: 1–12. doi:10.1038/s41598-019-41918-3
 100. Wei Z, and Li H. A Markov random field model for network-based analysis of genomic data. *Bioinformatics* (2007) 23:1537–44. doi:10.1093/bioinformatics/btm129
 101. Gomez-Romero L, Lopez-Reyes K, and Hernandez-Lemus E. The large scale structure of human metabolism reveals resilience via extensive signaling crosstalk. *Front Physiol* (2020) 11:1667. doi:10.3389/fphys.2020.588012
 102. Lin Z, Li M, Sestan N, and Zhao H. A Markov random field-based approach for joint estimation of differentially expressed genes in mouse transcriptome data. *Stat Appl Genet Mol Biol* (2016) 15:139–50. doi:10.1515/sagmb-2015-0070
 103. Chen M, Cho J, and Zhao H. Incorporating biological pathways via a Markov random field model in genome-wide association studies. *Plos Genet* (2011) 7: e1001353. doi:10.1371/journal.pgen.1001353
 104. Long Y, Wu M, Kwok CK, Luo J, and Li X. Predicting human microbe-drug associations via graph convolutional network with conditional random field. *Bioinformatics* (2020) 36:4918–27. doi:10.1093/bioinformatics/btaa598
 105. Xu J, Yang P, Xue S, Sharma B, Sanchez-Martin M, Wang F, et al. Translating cancer genomics into precision medicine with artificial intelligence: applications, challenges and future perspectives. *Hum Genet* (2019) 138: 109–24. doi:10.1007/s00439-019-01970-5
 106. Fariselli P, Savojarado C, Martelli PL, and Casadio R. Grammatical-restrained hidden conditional random fields for bioinformatics applications. *Algorithms Mol Biol* (2009) 4:13. doi:10.1186/1748-7188-4-13
 107. Zhang NR, Yakir B, Xia LC, and Siegmund D. Scan statistics on Poisson random fields with applications in genomics. *Ann Appl Stat* (2016) 10: 726–55. doi:10.1214/15-aas892
 108. Urbain J, Frieder O, and Goharian N. Passage relevance models for genomics search. *Proceedings of the 2nd international workshop on Data and text mining in bioinformatics*. New York, NY: DTMBIO '08 (2008) 45–52.
 109. Wang X, Li Y, He T, Jiang X, and Hu X. Recognition of bacteria named entity using conditional random fields in spark. *BMC Syst Biol* (2018) 12:106. doi:10.1186/s12918-018-0625-3
 110. McDonald R, and Pereira F. Identifying gene and protein mentions in text using conditional random fields. *BMC bioinformatics* (2005) 6:S6. doi:10.1186/1471-2105-6-s1-s6
 111. Vecchyo OD, Marsden CD, and Lohmueller KE. Prefersim: fast simulation of demography and selection under the Poisson random field model. *Bioinformatics* (2016) 32:3516–8. doi:10.1093/bioinformatics/btw478
 112. François O, Ancelet S, and Guillot G. Bayesian clustering using hidden Markov random fields in spatial population genetics. *Genetics* (2006) 174: 805–16. doi:10.1534/genetics.106.059923
 113. Clark NJ, Wells K, and Lindberg O. Unravelling changing interspecific interactions across environmental gradients using Markov random fields. *Ecology* (2018) 99:1277–83. doi:10.1002/ecy.2221
 114. Salinas NR, and Wheeler WC. Statistical modeling of distribution patterns: a Markov random field implementation and its application on areas of endemism. *Syst Biol* (2020) 69:76–90. doi:10.1093/sysbio/syz033
 115. Shen Y, and Van Deelen TR. *Spatially explicit modeling of community occupancy using markov random field models with imperfect observation: mesocarnivores in apostle islands national lakeshore*. Cold Spring Harbor, NY: BioRxiv (2020).
 116. Koziak R. Improving depth map quality with Markov random fields. *Image Processing and Communications Challenges*. Springer (2011) 149–56. doi:10.1007/978-3-642-23154-4_17
 117. Stephenson TA, and Chen T. Adaptive Markov random fields for example-based super-resolution of faces. *EURASIP J Adv Signal Process* (2006) 2006: 031062. doi:10.1155/asp/2006/31062
 118. Li C, and Wand M. Combining Markov random fields and convolutional neural networks for image synthesis. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Ithaca, NY: arXiv:1601.04589 (2016) 2479–86.
 119. Wen M, Han H, Wang L, and Wang W. 2d conditional random fields for image classification. *International Conference on Intelligent Information Processing*. Springer (2006) 383–90.
 120. Bohorquez M, Giraldo R, and Mateu J. Multivariate functional random fields: prediction and optimal sampling. *Stoch Environ Res Risk Assess* (2017) 31: 53–70. doi:10.1007/s00477-016-1266-y
 121. Baca-Lopez K, Fresno C, Espinal-Enriquez J, Martinez-Garcia M, Camacho-Lopez MA, Flores-Merino MV, et al. Spatio-temporal representativeness of air quality monitoring stations in Mexico city: implications for public health. *Front Public Health* (2020) 8:849. doi:10.3389/fpubh.2020.536174
 122. Baca-Lopez K, Fresno C, Espinal-Enriquez J, Flores-Merino MV, Camacho-Lopez MA, and Hernandez-Lemus E. *Metropolitan age-specific mortality trends at borough and neighbourhood level: the case of Mexico city* (2020).
 123. Wang H, Wellmann JF, Li Z, Wang X, and Liang RY. A segmentation approach for stochastic geological modeling using hidden Markov random fields. *Math Geosci* (2017) 49:145–77. doi:10.1007/s11004-016-9663-9

124. Li Z, Wang X, Wang H, and Liang RY. Quantifying stratigraphic uncertainties by stochastic simulation techniques based on Markov random field. *Eng Geology* (2016) 201:106–22. doi:10.1016/j.enggeo.2015.12.017
125. Rue H, and Held L. *Gaussian Markov random fields: theory and applications*. Boca Raton, FL: CRC Press (2005).
126. Solberg AHS, Taxt T, and Jain AK. A Markov random field model for classification of multisource satellite imagery. *IEEE Trans Geosci Remote Sensing* (1996) 34:100–13. doi:10.1109/36.481897
127. Toftaker H, and Tjelmeland H. Construction of binary multi-grid Markov random field prior models from training images. *Math Geosci* (2013) 45: 383–409. doi:10.1007/s11004-013-9456-3
128. Reuschen S, Xu T, and Nowak W. Bayesian inversion of hierarchical geostatistical models using a parallel-tempering sequential Gibbs mcmc. *Adv Water Resour* (2020) 141:103614. doi:10.1016/j.advwatres.2020.103614
129. Sutton C, and McCallum A. An introduction to conditional random fields for relational learning. *Introduction Stat relational Learn* (2006) 2:93–128.
130. Gilks WR, and Wild P. Adaptive rejection sampling for Gibbs sampling. *Appl Stat* (1992) 41:337–48. doi:10.2307/2347565
131. Gilks WR, Best NG, and Tan KKC. Adaptive rejection metropolis sampling within Gibbs sampling. *Appl Stat* (1995) 44:455–72. doi:10.2307/2986138
132. Meyer R, Cai B, and Perron F. Adaptive rejection metropolis sampling using Lagrange interpolation polynomials of degree 2. *Comput Stat Data Anal* (2008) 52:3408–23. doi:10.1016/j.csda.2008.01.005
133. Martino L, Read J, and Luengo D. Independent doubly adaptive rejection metropolis sampling within Gibbs sampling. *IEEE Trans Signal Process* (2015) 63:3123–38. doi:10.1109/tsp.2015.2420537
134. Papanikolaou Y, Foulds JR, Rubin TN, and Tsoumakas G. Dense distributions from sparse samples: improved Gibbs sampling parameter estimators for lda. *J Machine Learn Res* (2017) 18:2058–115.
135. Norton RA, Christen JA, and Fox C. Sampling hyperparameters in hierarchical models: improving on Gibbs for high-dimensional latent fields and large datasets. *Commun Stat - Simulation Comput* (2018) 47: 2639–55. doi:10.1080/03610918.2017.1353618
136. Gao S, and Gormley MR. Training for Gibbs sampling on conditional random fields with neural scoring factors. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing. Punta Cana, Dominican Republic: EMNLP (2020) 4999–5011.
137. Boland A, Friel N, and Maire F. Efficient mcmc for Gibbs random fields using pre-computation. *Electron J Statist* (2018) 12:4138–79. doi:10.1214/18-ejs1504
138. Kaplan A, Kaiser MS, Lahiri SN, and Nordman DJ. Simulating Markov random fields with a conclave-based Gibbs sampler. *J Comput Graphical Stat* (2020) 29:286–96. doi:10.1080/10618600.2019.1668800
139. Marcotte D, and Allard D. Gibbs sampling on large lattice with gmrf. *Comput Geosciences* (2018) 111:190–9. doi:10.1016/j.cageo.2017.11.012
140. Ko GG, Chai Y, Rutenbar RA, Brooks D, and Wei GY. Flexgibbs: reconfigurable parallel Gibbs sampling accelerator for structured graphs. IEEE 27th Annual International Symposium on Field-Programmable Custom Computing Machines (FCCM). IEEE (2019) 334.
141. Liu W, and Wu L. Large deviations for empirical measures of mean-field Gibbs measures. *Stochastic Process their Appl* (2020) 130:503–20. doi:10.1016/j.spa.2019.01.008
142. Eldan R, and Gross R. Decomposition of mean-field Gibbs distributions into product measures. *Electron J Probab* (2018) 23. doi:10.1214/18-ejp159
143. Shafer GR, and Shenoy PP. Probability propagation. *Ann Math Artif Intell* (1990) 2:327–51. doi:10.1007/bf01531015
144. Zhang NL, and Poole D. Intercausal independence and heterogeneous factorization. Uncertainty Proceedings. Elsevier (1994) 606–14. doi:10.1016/b978-1-55860-332-5.50082-1
145. Kompass R. A generalized divergence measure for nonnegative matrix factorization. *Neural Comput* (2007) 19:780–91. doi:10.1162/neco.2007.19.3.780
146. Cichocki A, Lee H, Kim Y-D, and Choi S. Non-negative matrix factorization with α -divergence. *Pattern Recognition Lett* (2008) 29:1433–40. doi:10.1016/j.patrec.2008.02.016
147. Ding C, Li T, and Peng W. On the equivalence between non-negative matrix factorization and probabilistic latent semantic indexing. *Comput Stat Data Anal* (2008) 52:3913–27. doi:10.1016/j.csda.2008.01.011
148. Xie Y, and Berkowitz CM. The use of positive matrix factorization with conditional probability functions in air quality studies: an application to hydrocarbon emissions in houston, Texas. *Atmos Environ* (2006) 40:3070–91. doi:10.1016/j.atmosenv.2005.12.065
149. Xu J, Cai L, Liao B, Zhu W, Wang P, Meng Y, et al. Identifying potential mirna-disease associations with probability matrix factorization. *Front Genet* (2019) 10:1234. doi:10.3389/fgene.2019.01234
150. Wang Z, Liang J, and Li R. A fusion probability matrix factorization framework for link prediction. *Knowledge-Based Syst* (2018) 159:72–85. doi:10.1016/j.knsys.2018.06.005
151. Stoehr J, Marin J-M, and Pudlo P. Hidden Gibbs random fields model selection using block likelihood information criterion. *Stat* (2016) 5:158–72. doi:10.1002/sta4.112
152. Cilla R, Patricio MA, Berlanga A, and Molina JM. Model and feature selection in hidden conditional random fields with group regularization. International Conference on Hybrid Artificial Intelligence Systems. Springer (2013) 140–9. doi:10.1007/978-3-642-40846-5_15
153. Sain SR, and Furrer R. Comments on: some recent work on multivariate Gaussian Markov random fields. *Test* (2018) 27:545–8. doi:10.1007/s11749-018-0609-z
154. Zhu J, Lao N, and Xing E. Grafting-light: fast, incremental feature selection and structure learning of Markov random fields. Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining (2010), 303–12.
155. Liao L, Choudhury T, Fox D, and Kautz HA. Training conditional random fields using virtual evidence boosting. *Ijcai* (2007) 7:2530–5.
156. Lafferty J, Zhu X, and Liu Y. Kernel conditional random fields: representation and clique selection. Proceedings of the twenty-first international conference on Machine learning. New York, NY: ICML '04 (2004) 64.
157. Zhu J, Wang H, and Mao J. Sentiment classification using genetic algorithm and conditional random fields. IEEE international conference on information management and engineering. IEEE (2010) 193–6.
158. Metzler DA. Automatic feature selection in the Markov random field model for information retrieval. Proceedings of the sixteenth ACM conference on Conference on information and knowledge management. (2007) 253–62.
159. Aliferis CF, Statnikov A, Tsamardinos I, Mani S, and Koutsoukos XD. Local causal and Markov blanket induction for causal discovery and feature selection for classification part i: algorithms and empirical evaluation. *J Machine Learn Res* (2010) 11.
160. Adams S, Beling PA, and Cogill R. Feature selection for hidden Markov models and hidden semi-markov models. *IEEE Access* (2016) 4:1642–57. doi:10.1109/access.2016.2552478
161. Brownlee AEI, Regnier-Coudert O, McCall JAW, Massie S, and Stulajter S. An application of a ga with Markov network surrogate to feature selection. *Int J Syst Sci* (2013) 44:2039–56. doi:10.1080/00207721.2012.684449
162. Yu L, and Liu H. Efficient feature selection via analysis of relevance and redundancy. *J machine Learn Res* (2004) 5:1205–24.
163. Slawski M, zu Castell W, and Tutz G. Feature selection guided by structural information. *Ann Appl Stat* (2010) 4:1056–80. doi:10.1214/09-aos302
164. Adams S, and Beling PA. A survey of feature selection methods for Gaussian mixture models and hidden Markov models. *Artif Intell Rev* (2019) 52: 1739–79. doi:10.1007/s10462-017-9581-3
165. Vergara JR, and Estévez PA. A review of feature selection methods based on mutual information. *Neural Comput Applic* (2014) 24:175–86. doi:10.1007/s00521-013-1368-0
166. Liu Z, Li X, Luo P, Change Loy C, and Tang X. Deep learning Markov random field for semantic segmentation. *IEEE Trans Pattern Anal Mach Intell* (2017) 40:1814–28. doi:10.1109/TPAMI.2017.2737535
167. Hu R, Rohrbach M, and Darrell T. Segmentation from natural language expressions. European Conference on Computer Vision. Springer (2016) 108–24. doi:10.1007/978-3-319-46448-0_7
168. Guo J, He H, He T, Lausen L, Li M, Lin H, et al. Gluoncv and gluonnlp: deep learning in computer vision and natural language processing. *J Machine Learn Res* (2020) 21:1–7.
169. Zhang H, Zhang H, Wang C, and Xie J. Co-occurrent features in semantic segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2019), 548–57.

170. Mai F, Wu S, and Cui T. Improved Chinese word segmentation disambiguation model based on conditional random fields. Proceedings of the 4th International Conference on Computer Engineering and Networks. Springer (2015) 599–605. doi:10.1007/978-3-319-11104-9_70
171. Qiu J, Zhou Y, Wang Q, Ruan T, and Gao J. Chinese clinical named entity recognition using residual dilated convolutional neural network with conditional random field. *IEEE Trans.on Nanobioscience* (2019) 18: 306–15. doi:10.1109/tnb.2019.2908678
172. Khan W, Daud A, Nasir JA, Amjad T, Arafat S, Aljohani N, et al. Urdu part of speech tagging using conditional random fields. *Lang Resour Eval* (2019) 53: 331–62. doi:10.1007/s10579-018-9439-6
173. Nguyen DM, Do TH, Calderbank R, and Deligiannis N. Fake news detection using deep Markov random fields. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Minneapolis, MN: Long and Short Papers (2019) 1391–400.
174. Colmenares CA, Litvak M, Mantrach A, Silvestri F, and Rodríguez H. *Headline generation as a sequence prediction with conditional random fields*. Singapore City, Singapore: Multilingual Text Analysis: Challenges, Models, and Approaches (2019) 201.
175. Knoke D, and Yang S. *Social network analysis*. Perth, Australia: Sage Publications (2019).
176. Jia J, Wang B, Zhang L, and Gong NZ. Attriinfer: inferring user attributes in online social networks using Markov random fields. Proceedings of the 26th International Conference on World Wide Web (2017) 1561–9.
177. Jin D, Liu Z, Li W, He D, and Zhang W. Graph convolutional networks meet Markov random fields: semi-supervised community detection in attribute networks. *Aaai* (2019) 33:152–9. doi:10.1609/aaai.v33i01.3301152
178. Feng B, Li Q, Ji Y, Guo D, and Meng X. *Stopping the cyberattack in the early stage: assessing the security risks of social network users*. Security and Communication Networks (2019).
179. Zhou Q, Xu Z, and Yen NY. User sentiment analysis based on social network information and its application in consumer reconstruction intention. *Comput Hum Behav* (2019) 100:177–83. doi:10.1016/j.chb.2018.07.006
180. Yoon S, Kleinman M, Mertz J, and Brannick M. Is social network site usage related to depression? A meta-analysis of Facebook-depression relations. *J affective Disord* (2019) 248:65–72. doi:10.1016/j.jad.2019.01.026
181. Ö B, Alexander SM, Baggio J, Barnes ML, Berardo R, Cumming GS, et al. Improving network approaches to the study of complex social-ecological interdependencies. *Nat Sustainability* (2019) 2:551–9. doi:10.1038/s41893-019-0308-0
182. Bhattacharya R, Malinsky D, and Shpitser I. Causal inference under interference and network uncertainty. Uncertainty in Artificial Intelligence (PMLR). Ithaca, NY: arXiv:1907.00221 (2020) 1028–38.
183. Stanković L, Daković M, and Sejdčić E. Introduction to graph signal processing. *Vertex-Frequency Analysis of Graph Signals*. Springer (2019) 3–108.
184. Stankovic L, Mandic DP, Dakovic M, Kasil I, Sejdic E, and Constantinides AG. Understanding the basis of graph signal processing via an intuitive example-driven approach [lecture notes]. *IEEE Signal Process Mag* (2019) 36:133–45. doi:10.1109/msp.2019.2929832
185. Ortega A, Frossard P, Kovacevic J, Moura JMF, and Vandergheynst P. Graph signal processing: overview, challenges, and applications. *Proc IEEE* (2018) 106:808–28. doi:10.1109/jproc.2018.2820126
186. Gadde A, and Ortega A. A probabilistic interpretation of sampling theory of graph signals. IEEE international conference on Acoustics, Speech and Signal Processing (ICASSP). (IEEE) (2015) 3257–61.
187. Chen S, Sandryhaila A, and Kovačević J. Sampling theory for graph signals. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). (IEEE) (2015) 3392–6.
188. Stanković L, and Sejdčić E. *Vertex-frequency analysis of graph signals*. Springer (2019).
189. Pavez E, and Ortega A. Generalized laplacian precision matrix estimation for graph signal processing. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). (IEEE) (2016) 6350–4.
190. Sandryhaila A, and Moura JM. Discrete signal processing on graphs: graph fourier transform. IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE (2013) 6167–70.
191. Mateos G, Segarra S, Marques AG, and Ribeiro A. Connecting the dots: identifying network structure via graph signal processing. *IEEE Signal Process Mag* (2019) 36:16–43. doi:10.1109/msp.2018.2890143
192. Ji F, and Tay WP. A hilbert space theory of generalized graph signal processing. *IEEE Trans Signal Process* (2019) 67:6188–203. doi:10.1109/tsp.2019.2952055
193. Itani S, and Thanou D. A graph signal processing framework for the classification of temporal brain data. 28th European Signal Processing Conference (EUSIPCO). (IEEE) (2021) 1180–4.
194. Ramakrishna R, and Scaglione A. Detection of false data injection attack using graph signal processing for the power grid. IEEE Global Conference on Signal and Information Processing (GlobalSIP). (IEEE) (2019) 1–5.
195. Stankovic L, Mandic D, Dakovic M, Brajovic M, Scalzo B, Li S, et al. *Graph signal processing-part iii: machine learning on graphs, from graph topology to applications*. Ithaca, NY: arXiv:2001.00426 (2020).
196. Song X, Chai L, and Zhang J. *Graph signal processing approach to qsar/qspr model learning of compounds*. IEEE Transactions on Pattern Analysis and Machine Intelligence (2020).
197. Burkhardt DB, Stanley JS, Perdigoto AL, Gigante SA, Herold KC, Wolf G, et al. *Quantifying the effect of experimental perturbations in single-cell rna-sequencing data using graph signal processing*. Cold Spring Harbor, NY: bioRxiv (2019) 532846.
198. Colonnese S, Pagliari G, Biagi M, Cusani R, and Scarano G. Compound Markov random field model of signals on graph: an application to graph learning. 7th European Workshop on Visual Information Processing (EUVIP). (IEEE) (2018) 1–5.
199. Torkamani R, and Zayyani H. *Statistical graph signal recovery using variational bayes*. IEEE Transactions on Circuits and Systems II: Express Briefs (2020).
200. Ramezani-Mayami M, Hajimirsadeghi M, Skretting K, Blum RS, and Poor HV. *Graph topology learning and signal recovery via bayesian inference*. IEEE Data Science Workshop (DSW) (IEEE) (2019) 52–6.
201. Colonnese S, Lorenzo PD, Cattai T, Scarano G, and Fallani FDV. A joint Markov model for communities, connectivity and signals defined over graphs. *IEEE Signal Process Lett* (2020) 27:1160–4. doi:10.1109/lsp.2020.3005053
202. Dong X, Thanou D, Rabbat M, and Frossard P. Learning graphs from data: a signal representation perspective. *IEEE Signal Process Mag* (2019) 36:44–63. doi:10.1109/msp.2018.2887284
203. Cheung M, Shi J, Wright O, Jiang LY, Liu X, and Moura JMF. Graph signal processing and deep learning: convolution, pooling, and topology. *IEEE Signal Process Mag* (2020) 37:139–49. doi:10.1109/msp.2020.3014594
204. Jia J, and Benson AR. *A unifying generative model for graph learning algorithms: label propagation, graph convolutions, and combinations*. Ithaca, NY: arXiv:2101.07730 (2021).
205. Gama F, and Ribeiro A. Ergodicity in stationary graph processes: a weak law of large numbers. *IEEE Trans Signal Process* (2019) 67:2761–74. doi:10.1109/tsp.2019.2908909
206. Segarra S, Wang Y, Uhler C, and Marques AG. Joint inference of networks from stationary graph signals. 51st Asilomar Conference on Signals, Systems, and Computers. (IEEE) (2017) 975–979.

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Hernández-Lemus. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.