Statistical Physics and Network Optimization Problems

Carlo Baldassi, Alfredo Braunstein, Abolfazl Ramezanpour, and Riccardo Zecchina

The scope of these lecture notes is to provide an introduction to modern statistical physics mean-field methods for the study of phase transitions and optimization problems over random structures. We first give a brief introduction to the field using as tutorial example the percolation problem in random graphs. Next we describe the so called cavity method and the related message-passing algorithms (Belief Propagation and variants) which can be used to analyze and solve optimization problems over random structures.

1 Statistical Physics and Optimization

Equilibrium statistical mechanics and combinatorial optimization have common roots. Phase transitions are mathematical phenomena which are not limited to physical systems but are typical of many combinatorial problems, one famous example being the percolation transition in random graphs. Similarly, the understanding of relevant physical problems, such as three dimensional lattice statistics

C. Baldassi (✉)
DISAT and Center for Computational Sciences, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy
e-mail: carlo.baldassi@polito.it

A. Braunstein • R. Zecchina
Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy

Human Genetics Foundation, Via Nizza 52, 10126 Torino, Italy

Collegio Carlo Alberto, Via Real Collegio 30, 10024 Moncalieri, Italy

A. Ramezanpour
Department of Physics, University of Neyshabur, P.O. Box 91136-899, Neyshabur, Iran

F. Fagnani et al. (eds.), *Mathematical Foundations of Complex Networked Information Systems*, Lecture Notes in Mathematics 2141,
DOI 10.1007/978-3-319-16967-5_2

or two dimensional quantum statistical mechanics problems, is strictly related to the question of purely combinatorial origin of solving counting problems over non planar lattices. Most of the tools and concepts which have allowed to solve problems in one field have a natural counterpart in the other. While the possibility of solving exactly physical models is related to the presence of algebraic properties which guarantee integrability, in the combinatorial approach the emphasis is more on algorithms that can be applied to specific problem instances.

Many interesting combinatorial optimization problems are NP-hard, implying that there is no known polynomial algorithm that is able to solve any instance of the problem. However, there is no a priori reason to assume that instances with real-world relevance will be specially hard, so the recent years have seen an upsurge of interest in the theory of typical-case complexity. In practice, hardness can still be present in random ensembles and one may try to identify those ensembles of optimization problems which are hard to solve, and the reason for this difficulty.

Combinatorial problems are usually written as Constraint Satisfaction Problems (CSP): N discrete variables are given which have to satisfy M constraints, all at the same time. Each constraint can take different forms depending on the problem under study. Well known examples are the K-Satisfiability (K-SAT) problem in which constraints are an 'OR' function of K variables in the ensemble (or their negations) and the Graph Q-coloring problem in which constraints simply enforce the condition that the end points of the edges in the graph must not have the same color (among the Q possible ones). A generic CSP can be written as the problem of finding a zero energy ground state of an appropriate energy function and its analysis amounts at performing a zero temperature statistical physics study. Hard combinatorial problems correspond to physical model systems with competing interaction, also called frustrated models.

In these lectures notes we will focus on two tutorial topics. The first one is to provide a brief introduction to statistical mechanics by showing how it can be used to study percolation in random graphs. The second is to give an introduction to the main statistical physics techniques which are used to study optimization problems over random structures.

2 Elements of Statistical Physics

The objective of statistical physics is to provide a probabilistic description of the macroscopic behaviour of a system at equilibrium from the knowledge of its microscopic components.

The implementation of this idea has required the introduction of revolutionary concepts and the interested reader can consult textbooks e.g. [1–3]. We shall adopt an operative approach and start from the following postulate.

A configuration C of the system, e.g. the specification of the N particle positions $\{\mathbf{r}_j\}$, has a probability $p(C)$ to be realized at any time when the system is in equilibrium. In other words, the probability of observing the system in configuration

C is given by[1]

$$p(C) = \frac{1}{Z} \exp\left(-\frac{1}{T} E(C)\right). \tag{1}$$

In the above expression, T is the temperature and E is the *energy* of the system (a real-valued function over the set of configurations). The *partition function* Z ensures the correct normalization of the probability distribution p,

$$Z = \sum_C \exp\left(-\frac{1}{T} E(C)\right). \tag{2}$$

The role of the temperature can be understood by considering two limiting cases:

1. *infinite temperature* $T = \infty$: the probability $p(C)$ becomes independent of C and all configurations are equiprobable. The system is said to be in a "disordered" phase (physically one may think to a gas or to a paramagnet).
2. *zero temperature* $T = 0$: the probability $p(C)$ is concentrated on the minimum of the energy function E, called the *ground state*.

The connection with combinatorics is easily understood by observing that the normalization factor Z is nothing but the *generating function* of the configuration energies.

We may rewrite the partition function (2) as

$$Z = \sum_E N(E) \exp(-\beta E), \tag{3}$$

where $N(E)$ is the number of configurations C having energies $E(C)$ precisely equal to E and $\beta = \frac{1}{T}$. If we set $x = \exp(-\beta)$ then $Z(x)$ becomes simply the generating function of the coefficients $N(E)$ as usually defined in combinatorics $Z = \sum_E N(E) x^E$.

For any configuration C, the *entropy* is defined as $\hat{S}(E(C)) = \log N(E(C))$. This quantity is in general very hard to compute, but in the large size limit $N \gg 1$, and in most cases of interest in statistical physics E is sharply peaked around its thermal average $\langle E \rangle_T$, and therefore one can compute the average entropy $S \simeq \sum_C \hat{S}(E(C)) p(C) \simeq \hat{S}(\langle E \rangle_T)$, which turns out to be equivalent to the *Shannon entropy* of the probability distribution p, as usually defined in information theory,

[1]Throughout this chapter, we will omit for simplicity the Boltzmann's constant k; this is always possible by choosing appropriate measurement units, such that $k = 1$.

and which is known as the *Gibbs entropy* in statistical physics:

$$S = -\sum_{C} p\left(C\right) \log p\left(C\right) = -\frac{1}{T}\left(F\left(T\right) - \langle E \rangle_{T}\right), \tag{4}$$

where

$$F\left(T\right) = -T \log Z\left(T\right) \tag{5}$$

is called the *free-energy* of the system. Therefore, it is usual to refer to S simply as the entropy of the system.

The entropy is an increasing function of temperature, which can be readily verified by computing the derivative of (4). At zero temperature, it corresponds to the logarithm of the number of absolute minima of the energy function $E\left(C\right)$.

The fact that $S \simeq \hat{S}\left(\langle E \rangle_{T}\right)$ is an instance of a more general phenomenon: for common statistical physics models, intensive statistical quantities such as the energy *density* E/N and the entropy *density* $\hat{S}/N$ become essentially fixed, when N is large, at any given value of T (except at most at phase transition boundaries, where discontinuities may appear): their probability distribution over the measure $p\left(C\right)$ tends exponentially to a Dirac delta distribution as $N \to \infty$. This property (known as *concentration of measure* in probability theory) allows to study e.g. the entropy as a function of the energy (by varying T), the phase transitions of the system, the structure of the ground states etc. A rigorous treatment of this subject can be found e.g. in [4].

3 Statistical Physics Approach to Percolation in Random Graphs

We shall attempt to familiarize with the statistical mechanics approach by describing its application to the analysis of the famous percolation problem in random graphs [5, 6]. Consider the complete graph K_N over N vertices. G_{N,N_L} is the set of graphs obtained by taking only $N_L = \gamma\ N/2$ among the $\binom{N}{2}$ edges of K_N in all possible different ways. A randomly chosen element of G_{N,N_L} with the flat measure is called a *random graph*.

The alternative procedure of deleting edges from K_N with probability $1 - \gamma/N$ lead to similar family of random graphs. In the large N limit, both families share common properties and we shall highlight the differences only when necessary.

The connected components of a given graph G are called *clusters*. Their size is the number of vertices they contain (e.g. isolated vertices are clusters of size one). We denote by $C\left(G\right)$ the number of connected components of G and by $c\left(G\right)$ their fractional number, $c\left(G\right) = \frac{C}{N}$. When c is small, the random graph G is

characterized by few large clusters whereas for c approaching unity there are many clusters of small size.

Percolation theory is concerned with the study of the relationship between the probability p of two vertices being connected with the typical value of c in the $N \to \infty$ limit.

In what follows we show how such a relationship can be analysed by the study of a statistical mechanics model, the so called Potts model, through a saddle-point technique.

We denote with $\mathcal{P}(G)$ the probability of generating a random graph G through the deletion process from the complete graph K_N. Given that the edge deletions are statistically independent, this probability depends on the number of edges N_L only, and factorizes as

$$\mathcal{P}(G) = p^{N_L(G)}(1-p)^{\frac{N(N-1)}{2}-N_L(G)}, \tag{6}$$

where $1 - p = 1 - \frac{\gamma}{N}$ is the probability of edge deletion. We want to study the probability density $\rho(c)$ of generating a random graph with c clusters,

$$\rho(c) = \sum_G \mathcal{P}(G)\ \delta(c - c(G)), \tag{7}$$

where δ indicates the Dirac distribution.

We can introduce a generating function of the cluster probability by

$$\begin{aligned} F(q) &= \int_0^1 dc\, \rho(c)\, q^{N c} \\ &= \int_0^1 dc\, q^{N c} \sum_{G \subseteq K_N} \mathcal{P}(G)\, \delta(c - c(G)) \\ &= \sum_{G \subseteq K_N} \mathcal{P}(G)\, q^{C(G)} = \sum_{G \subseteq K_N} p^{L(G)}(1-p)^{\frac{N(N-1)}{2}-L(G)}\, q^{C(G)}, \end{aligned} \tag{8}$$

with q being a formal parameter.

In the large size limit, $\rho(c)$ is expected to be highly concentrated around some value $c(\gamma)$ equal to the typical fraction of clusters per vertex and depending only the average degree of γ. Random graphs whose $c(G)$ differs enough from $c(\gamma)$ will be exponentially rare in N. Therefore, the quantity

$$\omega(c) = \lim_{N \to \infty} \frac{1}{N} \log \rho(c) \tag{9}$$

should vanish for $c = c(\gamma)$ and be strictly negative otherwise. In the following, we shall compute $\omega(c)$ and thus obtain information not only on the typical number of clusters but also on the large deviations.

Defining the logarithm $\tilde{f}(q)$ of the cluster generating function as

$$\tilde{f}(q) = \lim_{N\to\infty} \frac{1}{N} \log F(q), \tag{10}$$

we obtain from a saddle-point calculation on c, see (8), (9),

$$\tilde{f}(q) = \max_{0\le c\le 1} \Big[c \ \log q + \omega(c) \Big]. \tag{11}$$

$\tilde{f}$ and ω are conjugated Legendre transforms and it will turn out that a direct computation of $\tilde{f}$ is easier.

3.1 The Potts Model Representation

We proceed by computing the properties of random graphs by using a mapping of the generating function of the cluster probability to the so-called Potts model.

The Potts model [7] is defined in terms of an energy function which depends on N discrete variables σ_i (called *spins* in the physics jargon), one for each vertex of the complete graph K_N, which take q distinct values $\sigma_i = 0, 1, \ldots, q-1$. The energy function is written as

$$E[\{\sigma_i\}] = -\sum_{i<j} \delta(\sigma_i, \sigma_j), \tag{12}$$

where $\delta(a,b)$ is the Kronecker delta function: $\delta(a,b) = 1$ if $a = b$ and $\delta(a,b) = 0$ if $a \neq b$. The partition function of the Potts model can be written by summing over all q^N configurations

$$Z_{Potts} = \sum_{\{\sigma_i=0,\ldots,q-1\}} \exp\left[\beta \sum_{i<j} \delta(\sigma_i, \sigma_j)\right] \tag{13}$$

where $\beta = 1/T$ is the inverse temperature.

In order to identify the mapping we need to compare the expansion of Z_{Potts} to the definition of the cluster generating function $F(q)$ of the random graphs.

Following Kasteleyn and Fortuin [8], we start by rewriting Z_{Potts} as a dichromatic polynomial. Upon posing

$$v = e^{\beta} - 1, \tag{14}$$

one can easily check that (13) can be written in the form

$$Z_{Potts} = \sum_{\{\sigma_i\}} \prod_{i<j} \left[1 + v\delta\left(\sigma_i, \sigma_j\right)\right]. \tag{15}$$

When σ_i and σ_j take the same value there appears a factor $(1 + v)$ in the product (corresponding to a term e^{β} in (13)); on the contrary, whenever σ_i and σ_j are different the product remains unaltered. The expansion of the above product reads

$$Z_{Potts} = \sum_{\{\sigma_i\}} \left[1 + v \sum_{i<j} \delta\left(\sigma_i, \sigma_j\right) + v^2 \sum_{i<j,k<l/(i,j)\neq(k,l)} \delta\left(\sigma_i, \sigma_j\right)\delta\left(\sigma_k, \sigma_l\right) + \cdots \right]. \tag{16}$$

We obtain $2^{\frac{N(N-1)}{2}}$ terms each of which composed by two factors, the first one given by v raised to a power equal to the number of δs composing the second factor. It follows that each term corresponds to a possible subset of edges on K_N, each edge weighted by a factor v. There is a one-to-one correspondence between each term of the sum and the sub-graphs G of K_N. The edge structure of each sub-graph is encoded in the product of the δs. This fact allows to reinterpret the partition function as a sum over sub-graphs

$$Z_{Potts} = \sum_{\{\sigma_i\}} \sum_{G\subseteq K_N} \left[v^{L(G)} \prod_{k=0}^{L(G)} \delta\left(\sigma_{i_k}, \sigma_{j_k}\right) \right] \tag{17}$$

where $L(G)$ is the number of edges in the sub-graph G and i_k, j_k are the vertices connected by the kth edge of the sub-graph. The order of the summations can be exchanged to perform the sum over the configurations first. Given a sub-graph G with L links and C clusters (isolated vertices included), the sum over spins configurations will give zero unless all the σs belonging to a cluster of G have the same value (cfr. the δ functions). In such a cluster, one can set the σs to any of the q different values and hence the final form of the partition function reads

$$Z_{Potts} = \sum_{G\subseteq K_N} v^{L(G)} q^{C(G)}. \tag{18}$$

By making the identification $p = 1-e^{-\beta} = v/(1+v)$, we can rewrite the partition function as

$$\begin{aligned} Z_{Potts} &= \sum_{G \subseteq K_N} \left(\frac{p}{1-p}\right)^{L(G)} q^{C(G)} \\ &= (1-p)^{-\frac{N(N-1)}{2}} \sum_{G \subseteq K_N} p^{L(G)} (1-p)^{\frac{N(N-1)}{2} - L(G)} q^{C(G)}. \end{aligned} \tag{19}$$

By extracting the prefactor on the r.h.s. of (19), we thus find $Z_{Potts} = e^{\frac{N\gamma}{2}} F(q)$, at the leading exponential order in N. The large N behaviour of the cluster probability $\omega(c)$ is therefore related to the Potts free-energy,

$$f_{Potts}(q) = -\lim_{N\to\infty} \frac{1}{\beta N} \log Z_{Potts}, \tag{20}$$

through

$$-\frac{\gamma}{2} - f_{Potts}(q) = \max_{0\le c\le 1} (c \log q + \omega(c)). \tag{21}$$

We are interested in finding the value $c^\star(q)$ which maximizes the r.h.s. in (21); since

$$\left.\frac{d\omega(c)}{dc}\right|_{c^\star(q)} = -\log q \tag{22}$$

it follows that ω takes its maximum value for $q = 1$. Differentiating Eq. (21) with respect to q, we have

$$-\frac{df_{Potts}}{dq} = \frac{d}{dq}(c \log q + \omega(c)) = \frac{\partial}{\partial c}(c \log q + \omega(c)) \frac{\partial c}{\partial q} + \frac{c}{q}, \tag{23}$$

which, in virtue of Eq. (22) becomes:

$$c^\star(q) = -q \frac{df_{Potts}}{dq}(q). \tag{24}$$

We have thus shown that the typical fraction of clusters per site, $c^\star(q = 1)$, can be obtained, at a given connectivity γ, by computing the Potts free-energy in the vicinity of $q = 1$. Since the Potts model is originally defined for integer values of q, an analytic continuation to real values of q will be necessary.

A straightforward examination of the energy function (12) shows that the latter depends on the configurations only through the fractions $x(\sigma;\{\sigma_i\})$ of variables σ_i in the σth state ($\sigma = 0, 1, \ldots, q-1$) [9],

$$x(\sigma;\{\sigma_i\}) = \frac{1}{N}\sum_{i=1}^{N}\delta(\sigma_i,\sigma), \quad (\sigma = 0, 1, \ldots, q-1). \tag{25}$$

Of course, $\sum_\sigma x(\sigma;\{\sigma_i\}) = 1$.

Using these fractions, the energy (12) may be rewritten as

$$E[\{\sigma_i\}] = -\frac{N^2}{2}\sum_{\sigma=0}^{q-1} x(\sigma;\{\sigma_i\})^2 + \frac{N}{2}. \tag{26}$$

Note that the last term on the r.h.s. of (26) can be neglected with respect to the first term whose order of magnitude is $O\left(N^2\right)$.

The partition function (13) at inverse temperature $\beta = \gamma/N$ now becomes

$$\begin{aligned} Z_{Potts} &= \sum_{\{\sigma_i=0,1,\ldots,q-1\}} \exp\left(-\frac{\gamma}{2}\, N \sum_{\sigma=0}^{q-1} x(\sigma,\{\sigma_i\})^2\right) \\ &= \sum_{\{x_\sigma=0,1/N,\ldots,1\}}^{(\text{Norm})} \exp\left(\frac{\gamma}{2}\, N \sum_{\sigma=0}^{q-1} x(\sigma)^2\right) \frac{N!}{\Pi_{\sigma=0}^{q-1}[Nx(\sigma)]!} \\ &= \int_0^1 {}^{(\text{Norm})} \Pi_{\sigma=1}^{q-1}\, dx(\sigma)\ \exp\left(-Nf[\{x(\sigma)\}]\right) \end{aligned} \tag{27}$$

to the leading order in N. The superscript (Norm) indicates that the sum or the integral must be restricted to the normalized subspace $\sum_{\sigma=0}^{q-1} x(\sigma) = 1$. The "free-energy" density functional f appearing in (27) is

$$f[\{x(\sigma)\}] = \sum_{\sigma=0}^{q-1}\left\{-\frac{\gamma}{2}[x(\sigma)]^2 + x(\sigma)\log x(\sigma)\right\}. \tag{28}$$

In the limit of large N, the integral in (27) may be evaluated by the saddle-point method and the Potts free-energy (20) can be evaluated as

$$f_{Potts}(q) = \min_{\{x(\sigma)\}} f[\{x(\sigma)\}]. \tag{29}$$

From the definition of the problem, each possible value of σ plays the same role and f is invariant under the permutation symmetry of the different q values. However, we should keep in mind that such a symmetry could be broken in a given minimum. Depending on the value of the connectivity γ, the permutation symmetry

may or may not be broken, leading to a phase transition in the problem which coincides with the birth a giant component in the associated random graph.

3.1.1 Symmetric Saddle-Point

Consider first the symmetric candidate for an extremum of f,

$$x^{sym}(\sigma) = \frac{1}{q}, \qquad \forall \sigma = 0, \ldots, q-1. \tag{30}$$

We have

$$f^{sym}_{Potts}(q) = -\log q - \frac{\gamma}{2q}. \tag{31}$$

Taking the Legendre transform of this free-energy, see (21) and (24), we get for the logarithm of the cluster distribution density

$$\omega^{sym}(c) = -\frac{\gamma}{2} - (1-c)\left(1 + \log\gamma - \log\left[2(1-c)\right]\right). \tag{32}$$

$\omega^{sym}(c)$ is maximal and null at $c^{sym}(\gamma) = 1 - \frac{\gamma}{2}$, a result that cannot be true for connectivities larger than two and must break down somewhere below. Comparison with the rigorous derivation in random graph theory indicates that the symmetric result is exact as long as $\gamma \leq \gamma_c = 1$ and is false above the percolation threshold γ_c. The failure of the symmetric extremum in the presence of a giant component coincides with the appearance of symmetry broken saddle points.

To understand the mechanism responsible for the symmetry breaking, one may look at the local stability of the symmetric saddle-point (30) and compute the eigenvalues of the Hessian matrix. One finds a non degenerate eigenvalue $\lambda_0 = q(q-\gamma)$ and another eigenvalue $\lambda_1 = q - \gamma$ with multiplicity $q-2$. The analytic continuation of the eigenvalues to real $q \to 1$ lead to the single value $\lambda = 1 - \gamma$ which changes sign at the percolation threshold γ_c. Therefore, the symmetric saddle-point is not a local minimum of f above γ_c, showing that a more complicated saddle-point has to be found.

3.1.2 Symmetry Broken Saddle-Point

The simplest way to break the symmetry of the problem is to look for solutions in which one among the q values appears more frequently than the others. Therefore

one can look for a saddle-point of the form

$$x(0) = \frac{1}{q}\left[1 + (1-q)\, s\right]$$
$$x(\sigma) = \frac{1}{q}\left[1 - s\right], \qquad (\sigma = 1, \ldots, q-1). \tag{33}$$

The symmetric case can be recovered by setting $s = 0$. The free-energy of the Potts model is obtained by plugging the fractions (33) into (28). In the limit $q \to 1$ of interest,

$$f\left[\{x(\sigma)\}\right] = -\frac{\gamma}{2} + (q-1)\, f_{Potts}(s, \gamma) + O\left((q-1)^2\right) \tag{34}$$

with

$$f_{Potts}(s, \gamma) = \frac{\gamma}{2}\left(1 - \frac{1}{2}s^2\right) - 1 + s + (1-s)\log(1-s) \tag{35}$$

Minimization of $f_{Potts}(s, \gamma)$ with respect to the order parameter s shows that for $\gamma \leq 1$ the symmetric solution $s = 0$ is recovered, whereas for $\gamma > 1$ there exists a non vanishing optimal value $s^\star(\gamma)$ of s that is solution of the implicit equation

$$1 - s^\star = \exp(-\gamma\, s^\star). \tag{36}$$

The stability analysis shows that the solution is stable for any value of γ.

The interpretation of $s^\star(\gamma)$ is straightforward: $s^\star$ is the fraction of vertices belonging to the giant cluster. The average fraction of connected components $c(\gamma)$ equals $-f_{Potts}(s^\star(\gamma), \gamma)$, see (24), in perfect agreement with exact results by Erdős and Rényi.

Further results on the properties of random graphs can be extracted from the previous type of calculation, such as the scaling behaviour at the percolation point and large deviations. We refer to [10] for details.

4 Statistical Physics Methods for More Complex Problems

More advanced mean-field methods can be used to analyze and solve problems defined not only over K_N as in the previous example but on sparser graphs. Optimization and constraint satisfaction problems over finite connectivity random networks are the chief examples [11]. The functional mean field techniques which are used to study these problems are known as cavity methods with different levels of possible symmetry breaking. The story of these mathematical approaches is quite old [12, 13] and they have been rediscovered many times within different disciplines.

However it is only in the last decade that their algorithmic power has started to be fully appreciated.

In order to provide a concise presentation we introduce a notation which will used throughout the rest of this chapter for generic constraint satisfaction problems and Statistical physics models.

We will denote by $\mathbf{s}$ the set of N discrete variables over which the problem is defined, and use by convention the letters $i, j, k, \ldots$ to denote variable indices. Each variable s_i can in general take values in a different set X_i, thus $\mathbf{s} = \{s_i \in X_i | i = 1, \ldots, N\}$, but it is often the case that $X_i = X$ is common to all variables.

We will denote by C the set of M hard constraints which characterize the problem, and use by convention the letters $a, b, \ldots$ to denote hard constraint indices; therefore $C = \{C_a | a = 1, \ldots, M\}$. We also use the notation ∂a to indicate the set of all variable indices involved in the ath constraint (and by $s_{\partial a}$ the corresponding set of variables), and ∂i to denote the set of all constraint indices which involve the ith variable. So to each given constraint C_a we associate the indicator function $\mathbb{I}_a(s_{\partial a})$ which is 1 if the constraint is satisfied, 0 otherwise.

A constraint satisfaction problem can thus be mapped onto a bipartite graph (factor graph), with variable nodes representing the variables s_i and factor nodes representing the constraints C_a; the edges of the graph always connect nodes of different types and encode the structure of the problem (i.e. the sets ∂a, or equivalently ∂i). We denote with (ai) an edge from the factor node a to the variable node i. We denote the set of all edges of the graph by $\mathcal{E}$.

The problem is deemed *satisfiable* (SAT) as long as there is a solution satisfying all the hard constraints, otherwise it is *unsatisfiable* (UNSAT).

We also introduce soft constraints, associated with either node type, denoted as $E_i(s_i)$ for the variable nodes and $E_a(s_{\partial a})$ for the factor nodes. Together, they provide the energy function $E = \sum_i E_i(s_i) + \sum_a E_a(s_{\partial a})$, which allows to differentiate between valid configurations $\mathbf{s}$ by favouring those configurations which have the lowest energy. An external parameter β, which has the role of an inverse temperature, controls the relative weight of the valid configurations as a function of their energy; in the limit of $\beta \to 0$, all configurations are equally weighted, while in the zero-temperature limit $\beta \to \infty$ the soft constraints become equivalent to hard constraints and only the configurations which realize the minima of the energy (ground states) are allowed.

This formalism allows us to map any constraint satisfaction problem onto a statistical physics model defined by the following partition function:

$$Z = \sum_{\mathbf{s}} \prod_a \mathbb{I}_a(s_{\partial a}) \, e^{-\beta\left[\sum_i E_i(s_i) + \sum_a E_a(s_{\partial a})\right]} \tag{37}$$

which corresponds to the following probability measure:

$$\mu(\mathbf{s}) = \frac{1}{Z} \prod_a \mathbb{I}_a \left(s_{\partial a}\right) e^{-\beta\left[\sum_i E_i(s_i) + \sum_a E_a(s_{\partial a})\right]} \tag{38}$$

As we already noted in Sect. 2, in most common cases intensive statistical quantities such as the energy density are essentially fixed for almost any given value of the problem's parameters, when $N \gg 1$. Furthermore, CSPs and statistical physics models are often considered at the level of ensembles (families) of problems, where the parameters which describe each problem (the structure of the factor graph and of the constraints) are extracted from some probability distribution. A realization of the parameters constitutes the so-called *quenched disorder* of the system. When the probability distribution of some quantity, with respect to the quenched disorder (consider e.g. the thermal average of the energy density as a function of the model's parameters) is also peaked around its mean, the quantity is called *self-averaging*. This implies that any random instance of a problem is representative of the whole family, and, conversely, that studying the problem at the ensemble level is informative about almost all individual instances of the problem. The percolation transition discussed above is an example of this phenomenon. The self-averaging property of a given quantity may need to be assessed on a case-by-case basis, but in some important cases general results can be used (e.g. Talagrand's concentration inequality [14], which applies to Lipschitz-continuous functions).

Following are some examples of how some common problems can be mapped on this formalism:

- In a spin glass (a prototypical problem in statistical physics of disordered systems) the variables are binary, $s_i \in \{-1, +1\}$, and all constraints involve only two distinct variables, so if $\partial a = \{i, j\}$ we can make the identification $a \equiv (ij)$; hard constraints are moot, $\forall i, j : \mathbb{I}_{(ij)}\left(s_i, s_j\right) = 1$, and soft constraints are described by local fields h_i and interactions J_{ij}, extracted from some random distribution: $E = -\sum_i h_i s_i - \sum_{(ij)\in\mathcal{E}} J_{ij} s_i s_j$.
- In the q-coloring problem the interactions are again pair-wise, $a \equiv (ij)$, but the graph structure is typically non-trivial (e.g. may be random), each variable can take one of q states, $s_i \in \{0, \ldots, q-1\}$, and all constraints are hard: $\mathbb{I}_{(ij)}\left(s_i, s_j\right) = 1 - \delta_{s_i, s_j}$ (where δ is the Kronecker symbol).
- In the maximum weight independent set problem the variables are binary, $s_i \in \{0, 1\}$, the interaction are pair-wise, $a \equiv (ij)$, and defined on a non-trivial graph, and we have both hard and soft constraints: $\mathbb{I}_{(ij)}\left(s_i, s_j\right) = \delta_{s_i s_j, 0}$ and $E_i\left(s_i\right) = -s_i$.
- In the p-spin model (an extension of the spin glass model) variables are binary, $s_i \in \{-1, +1\}$, but each interaction involves p variables, $a \equiv \left(i_1, \ldots, i_p\right)$, and we have soft constraints $E = -\sum_{\left(i_1 \ldots i_p\right)\in\mathcal{E}} J_{i_1 \ldots i_p} \prod_{l=1}^{p} s_{i_l}$. When $\beta \to \infty$, it

becomes equivalent to the problem known as p-XORSAT in computer science, with each energy term corresponding to an XOR hard constraint enforcing the condition

$$\sum_{l=1}^{p} \frac{\left(1+s_{i_l}\right)}{2} = \frac{\left(1+\operatorname{sign}\left(J_{i_1 \ldots i_p}\right)\right)}{2} \quad \mod 2 \tag{39}$$

- In the K-SAT problem the variables are binary, $s_i \in \{-1,+1\}$, and each hard constraint C_a involves K variables: $\mathbb{I}_a\left(s_{\partial a}\right) = 1 - \prod_{l=1}^{K}\left(1 - J_{a i_l} s_{i_l}\right)$, with $J_{ai} \in \{-1,+1\}$.

5 Bethe Approximation and Message Passing Algorithms

5.1 Belief Propagation

Solving a model described by (37) is in general a hard problem. However, a general solution is possible when the underlying factor graph structure has no loops, in which case it is called a *tree*.[2]

5.1.1 Marginals

Let us define the local marginals

$$\mu_i\left(s_i\right) = \sum_{\{s_j\}_{j \neq i}} \mu\left(\mathbf{s}\right) \tag{40}$$

$$\mu_a\left(s_{\partial a}\right) = \sum_{\{s_j\}_{j \notin \partial a}} \mu\left(\mathbf{s}\right). \tag{41}$$

We will show that, on a tree, (38) can be written in terms of the above marginals, in one of the two equivalent forms:

$$\mu\left(\mathbf{s}\right) = \prod_i \mu_i\left(s_i\right) \prod_a \frac{\mu_a\left(s_{\partial a}\right)}{\prod_{i \in \partial a} \mu_i\left(s_i\right)} \tag{42}$$

$$= \prod_i \mu_i\left(s_i\right)^{1-|\partial i|} \prod_a \mu_a\left(s_{\partial a}\right). \tag{43}$$

[2] When the graph does not form a single connected component it is often called a *forest*, but this distinction is moot four our purposes.

The marginals (40) and (41) can be obtained by the so-called Belief Propagation (or Bethe-Peierls) equations. The procedure consists of writing marginals for a graph in which either a variable or some factor nodes are removed, as functions of analogous marginals. These marginals are called *cavity* marginals, and Belief Propagation is an example of a cavity method. The resulting system of equations is then solved by an iterative procedure, and the desired, non-cavity marginals are finally computed. We indicate with the subscript $i \to a$ the cavity marginals for variable i when the factor node a is removed from the graph, and with the subscript $a \to i$ the cavity marginal for variable i when all factor nodes $b \in \partial i \setminus a$ are removed.

The Belief Propagation (BP) equations for the cavity marginals are written as:

$$\mu_{i\to a}(s_i) = \frac{1}{z_{i\to a}} e^{-\beta E_i(s_i)} \prod_{b\in\partial i\setminus a} \mu_{b\to i}(s_i) \tag{44}$$

$$\mu_{a\to i}(s_i) = \frac{1}{z_{a\to i}} \sum_{s_{\partial a\setminus i}} \mathbb{I}_a(s_{\partial a})\, e^{-\beta E_a(s_{\partial a})} \prod_{j\in\partial a\setminus i} \mu_{j\to a}(s_j) \tag{45}$$

where $z_{i\to a}$ and $z_{a\to i}$ are normalization constants. These equations are exact on tree graphs, since, when removing a factor node, all the variables involved in that node become independent from each other, and their probability distribution factorizes. They can be solved iteratively by starting from the leaves of the graph (nodes of connectivity 0 or 1), where their expression becomes trivial, and iterating inwards; therefore, the time required is at worst of order MKq^K, where K is the maximum degree of the function nodes, and q is the maximum number of states which a variable can take. The term q^K can in some common cases be improved by exploiting the structure of the functions $\mathbb{I}_a$ and E_a.

The cavity marginals are also often called *messages*, and the iterative procedure *message passing*.

It is also useful to introduce the non-normalized message passing equations, which define the cavity partition functions:

$$Z_{i\to a}(s_i) = e^{-\beta E_i(s_i)} \prod_{b\in\partial i\setminus a} Z_{b\to i}(s_i) \tag{46}$$

$$Z_{a\to i}(s_i) = \sum_{s_{\partial a\setminus i}} \mathbb{I}_a(s_{\partial a})\, e^{-\beta E_a(s_{\partial a})} \prod_{j\in\partial a\setminus i} Z_{j\to a}(s_j)\,. \tag{47}$$

Using these, we can express the marginals (40) and (41) as:

$$\mu_i(s_i) = \frac{1}{Z} e^{-\beta E_i(s_i)} \prod_{a\in\partial i} Z_{a\to i}(s_i) \tag{48}$$

$$\mu_a(s_{\partial a}) = \frac{1}{Z} \mathbb{I}_a(s_{\partial a})\, e^{-\beta E_a(s_{\partial a})} \prod_{i\in\partial a} Z_{i\to a}(s_i)\,. \tag{49}$$

These allow us to write the constraints as functions of the marginals and the cavity partition functions; introducing them back into Eq. (38) allows us to prove the relation (42).

It is easy to see that the marginals (40) and (41) can also be written in terms of the cavity marginals as:

$$\mu_i(s_i) = \frac{1}{z_i} e^{-\beta E_i(s_i)} \prod_{a \in \partial i} \mu_{a \to i}(s_i) \tag{50}$$

$$\mu_a(s_{\partial a}) = \frac{1}{z_a} \mathbb{I}_a(s_{\partial a})\, e^{-\beta E_a(s_{\partial a})} \prod_{i \in \partial a} \mu_{i \to a}(s_i) \tag{51}$$

where again z_i and z_a are normalization constants.

Another useful relationship which immediately follows is:

$$\mu_i(s_i) \propto \mu_{i \to a}(s_i)\, \mu_{a \to i}(s_i) \quad \forall a \in \partial i. \tag{52}$$

5.1.2 Free Energy

The cavity partition functions also allow us to express the total free energy and partition function by:

$$e^{-\beta F} = Z = \sum_{s_i} e^{-\beta E_i(s_i)} \prod_{a \in \partial i} Z_{a \to i}(s_i) \quad \forall i. \tag{53}$$

We can decompose the total free energy as the sum of local contributions expressed in terms of the cavity messages. Let us define the cavity free energies and free energy shifts by:

$$e^{-\beta F_{i \to a}} = \sum_{s_i} Z_{i \to a}(s_i) \tag{54}$$

$$e^{-\beta F_{a \to i}} = \sum_{s_i} Z_{a \to i}(s_i) \tag{55}$$

$$e^{-\beta \Delta F_{i \to a}} = e^{-\beta\left(F_{i \to a} - \sum_{b \in \partial i \setminus a} F_{b \to i}\right)} = z_{i \to a} \tag{56}$$

$$e^{-\beta \Delta F_{a \to i}} = e^{-\beta\left(F_{a \to i} - \sum_{j \in \partial a \setminus i} F_{j \to a}\right)} = z_{a \to i} \tag{57}$$

and analogously for the non-cavity free energy shifts:

$$e^{-\beta \Delta F_i} = z_i \tag{58}$$

$$e^{-\beta \Delta F_a} = z_a. \tag{59}$$

Using the BP equations it is easy to see that:

$$\Delta F_{i\to a} = \Delta F_i - \Delta F_{ia} \tag{60}$$

$$\Delta F_{a\to i} = \Delta F_a - \Delta F_{ia} \tag{61}$$

with [cfr. Eq. (52)]

$$e^{-\beta \Delta F_{ia}} = \sum_{s_i} \mu_{i\to a}\left(s_i\right) \mu_{a\to i}\left(s_i\right) = \frac{z_i}{z_{i\to a}}. \tag{62}$$

With these, we can rewrite the free energy as:

$$F = \sum_i \Delta F_i + \sum_a \Delta F_a - \sum_{(ia)} \Delta F_{ia} \tag{63}$$

or as:

$$F = \sum_a \left(\Delta F_a + \sum_{i\in\partial a} \Delta F_{i\to a} \right) - \sum_i \left(|\partial i| - 1\right) \Delta F_i . \tag{64}$$

The free energy can also be written in terms of the local marginals $\mu_i\left(s_i\right)$ and $\mu_a\left(s_{\partial a}\right)$, as:

$$F = \sum_a F_a - \sum_i \left(|\partial i| - 1\right) F_i \tag{65}$$

where

$$F_a = \sum_{s_{\partial a}} E_a\left(s_{\partial a}\right) \mu_a\left(s_{\partial a}\right) - \frac{1}{\beta} \sum_{s_{\partial a}} \mu_a\left(s_{\partial a}\right) \ln \mu_a\left(s_{\partial a}\right) \tag{66}$$

$$F_i = \sum_{s_i} E_i\left(s_i\right) \mu_i\left(s_i\right) - \frac{1}{\beta} \sum_{s_i} \mu_i\left(s_i\right) \ln \mu_i\left(s_i\right). \tag{67}$$

In the above equations, we adopted the convention that $x \log x = 0$ if $x = 0$.

5.1.3 Graphs with Loops

The above equations are only exact if the factor graph is a tree. When that is not the case, expression (42), known as the Bethe approximation, can still provide good results in a wide range of cases, and even become asymptotically exact in the limit of large N.

Preliminarily, we shall note that, on graphs with loops, the procedure for solving the BP equations must be slightly modified: for example, one could initialize the cavity messages at random, or uniformly, and update them by using Eqs. (44) and (45) in random order, until a fixed point for the whole set is eventually found.

With regard to the approximation estimate, the crucial observation is that the property which was used in the derivation of the BP equations is that the cavity marginals factorize, i.e. that the connected correlations between any two variables involved in an interaction a vanish when the interaction a is removed: this is commonly called the *clustering property*.

In a graph with loops, the clustering property does not hold in general, but the correlations which one neglects by assuming that it does (i.e. that the joint probability of the variables $s_{\partial a}$ can be factorized if a is removed) can become asymptotically small as the problem size grows: one common example where this may happen is that of random graphs in which the typical size of the loops tends to diverge with the size of the graph. This kind of graphs are called *locally tree-like*: removing any factor node a means that the distance along the graph of the variables in $s_{\partial a}$ diverges, and we may expect that they become uncorrelated. This is often, but now always, the case, since there is one additional, more subtle, condition: in the above derivation, we assumed that the BP equations only have a single solution: this is always true on trees, but when loops are present there may be no solution at all, or more than one, even on locally tree-like graphs (see Sect. 5.4). In both cases, the observed effect is that the iterative procedure does not converge, or that some normalization constants $z_{i\to a}$ or $z_{a\to i}$ become 0 during the iteration.

In many cases, the BP equations provide reasonable estimates even when the clustering property doesn't hold; in particular, they can be exploited to identify "good" configurations (i.e. low-energy, SAT configurations) for a given instance of a problem, even when optimality is not strictly guaranteed, as we shall see in Sect. 5.3.

5.2 The $\beta \to \infty$ Limit: Minsum Algorithm

The BP equations (44) and (45) can be studied in the limit $\beta \to \infty$, assuming that the messages scale as $\mu_{i\to a}(s_i) \propto e^{-\beta M_{i\to a}(s_i)}$ and $\mu_{a\to i}(s_i) \propto e^{-\beta M_{a\to i}(s_i)}$. In this way we obtain:

$$M_{i\to a}(s_i) = E_i(s_i) + \sum_{b\in\partial i\setminus a} M_{b\to i}(s_i) - C_{i\to a} \tag{68}$$

$$M_{a\to i}(s_i) = \min_{s_{\partial a\setminus i}:\mathbb{I}_a(s_{\partial a})=1} \left\{ E_a(s_{\partial a}) + \sum_{j\in\partial a\setminus i} M_{j\to a}(s_j) \right\} - C_{a\to i}. \tag{69}$$

These are the so called minsum equations. The constants $C_{i\to a}$ and $C_{a\to i}$ enforce the conditions $\min_{s_i} M_{i\to a}(s_i) = \min_{s_i} M_{a\to i}(s_i) = 0$, which are the analogous to the normalization of the BP messages.

When applied to tree graphs, the resulting minsum algorithm can be ascribed to the family of dynamic programming algorithms.

The minsum messages can be used to obtain an estimate of the minimum energy by taking the limit $\beta \to \infty$ of the Bethe free energy (63):

$$E_0 = \lim_{\beta\to\infty} F = \sum_i \Delta E_i + \sum_a \Delta E_a - \sum_{(ia)} \Delta E_{ia} \tag{70}$$

where

$$\Delta E_i = \lim_{\beta\to\infty} \Delta F_i = \min_{s_i} \left\{ E_i(s_i) + \sum_{a\in\partial i} M_{a\to i}(s_i) \right\} \tag{71}$$

$$\Delta E_a = \lim_{\beta\to\infty} \Delta F_a = \min_{s_{\partial a}:\mathbb{I}_a(s_{\partial a})=1} \left\{ E_a(s_{\partial a}) + \sum_{i\in\partial a} M_{i\to a}(s_i) \right\} \tag{72}$$

$$\Delta E_{ia} = \lim_{\beta\to\infty} \Delta F_{ia} = \min_{s_i} \{ M_{a\to i}(s_i) + M_{i\to a}(s_i) \} \tag{73}$$

If the ground state is unique, it can be found from the expression of the marginal (50):

$$s_i^\star = \mathrm{argmin}_{s_i} \left(E_i(s_i) + \sum_{a\in\partial i} M_{a\to i}(s_i) \right). \tag{74}$$

However, as for the BP equations, this is only guaranteed to be correct on a tree. The condition of uniqueness of the ground state can be ensured by adding a small noise perturbation to single site energetic terms E_i. On graphs with loops, the iteration often fail to converge. A general strategy both to deal with this situation and to find solutions of CSP at non-zero temperature is presented in the next section.

5.3 *Finding a Solution: Decimation and Reinforcement Algorithms*

Suppose we want to find a solution of a constraint satisfaction problem described by (38), and that we are able to compute (perhaps approximately) the local marginals $\mu_i(s_i)$. We will introduce two heuristic approaches based on message passing: decimation and reinforcement.

5.3.1 Decimation

A decimation algorithm works as follows.

1. Compute the $\mu_i(s_i)$ (via BP, or minsum, or another algorithm).
2. Fix the most biased[3] variable $i_{\max}$ according to the local marginal: $i_{\max} = \operatorname{argmax}_i (\max_{s_i} \mu_i(s_i) - \operatorname{sndmax}_{s_i} \mu_i(s_i))$, where sndmax is the second maximum function: $\operatorname{sndmax}_x f(x) = \max_{x \neq \operatorname{argmax}_y f(y)} f(x)$. Fixing a variable reduces the problem to a new problem with $N-1$ variables.
3. Repeat the above steps until all variables are fixed.

Note that as long as we can compute the local marginals exactly the above decimation algorithm will end up with the optimal solution, if one solution exists. If the marginals are approximate, this is not guaranteed, and the actual results depend on the problem.

An easy modification which can speed up the algorithm is to fix a small fraction of the variables at a time in step 2, rather than only one.

5.3.2 Reinforcement

Another similar approach is to fix the variables smoothly, by introducing a reinforcement field which changes during the iterative message passing propagation. Let us consider the iterative BP algorithm where messages are initialized at random (or uniformly) and updated in succession according to Eqs. (44) and (45), and let us denote with τ the iterative step.[4] The reinforced BP equations read:

$$\mu_{i\to a}^{\tau+1}(s_i) = \frac{1}{z_{i\to a}^{\tau+1}} e^{h_i^\tau(s_i)} e^{-\beta E_i(s_i)} \prod_{b\in\partial i\setminus a} \mu_{b\to i}^{\tau}(s_i) \tag{75}$$

$$\mu_{a\to i}^{\tau}(s_i) = \frac{1}{z_{a\to i}^{\tau}} \sum_{s_{\partial a\setminus i}} \mathbb{I}_a(s_{\partial a})\, e^{-\beta E_a(s_{\partial a})} \prod_{j\in\partial a\setminus i} \mu_{j\to a}^{\tau}(s_j) \tag{76}$$

where we introduced the reinforcement field h_i^τ:

$$h_i^\tau(s_i) = r(\tau) \ln \mu_i^\tau(s_i) \tag{77}$$

$$\mu_i^{\tau+1}(s_i) = \frac{1}{z_i^{\tau+1}} e^{h_i^\tau(s_i)} e^{-\beta E_i(s_i)} \prod_{a\in\partial i} \mu_{a\to i}^{\tau}(s_i)\,. \tag{78}$$

[3] Using the most biased variable is a simple and reasonable heuristic which works well in practice, but other strategies may be considered.

[4] One step may correspond to one update of all messages (*synchronous* update scheme), or more often to the update of the messages associated to one randomly chosen variable or function node (*asynchronous* update scheme).

Here, $r(\tau)$ is the reinforcement parameter: it is initialized as $r(0) = 0$ and increased slowly, until the algorithm converges to a single solution of the problem, i.e. until $\mu_i^{\tau}(s_i) \simeq \delta_{s_i,s_i^{\star}}$. The underlying idea is to use the approximate marginals obtained after τ iterations, and use them to bias the problem in the following iterations: rather like decimation, but in an ongoing fashion. Compared to decimation, this reinforced iterative scheme has the advantage of being potentially quicker [this depends on the growth rate of $r(\tau)$], and of being applicable in some cases even when the standard BP equations don't admit a single fixed point.

The reinforcement scheme can be straightforwardly extended to the minsum algorithm of Sect. 5.2 by taking the limit $\beta \to \infty$.

5.4 Replica Symmetry Breaking and Higher Levels of BP

The multiple-BP-fixed-points situation mentioned at the end of Sect. 5.1 may occur e.g. when the space of solutions is fragmented in clusters, each one corresponding to a different fixed point of the BP equations (the clustering property is valid within each cluster, but not globally). In the context of the statistical physics of disordered systems, this phenomenon is known under the name of *replica symmetry breaking* (RSB), while the situation where the space of solutions forms a single connected cluster is indicated as *replica symmetric* (RS). These names originate from the studies on the typical structure of the phase space in disordered models via the so-called *replica method*; the computations are carried over by using the saddle point method, and the structure of said saddle point is described in terms of its level of symmetry breaking, with the symmetric solution being called RS and the successive levels 1RSB, 2RSB, etc., up to full-RSB. When the correct level of RSB is not taken into account properly, the estimates of the order parameters are only approximate. The underlying theory (see [12] for a full exposition) is not rigorously proven yet, but has been highly successful: as yet, its results on a wide class of models (called mean-field models) have always been in agreement with rigorous theoretical results (whenever available), and with numerical experiments. Note that the last statement does not contain a precise conjecture and is somehow tautological, as it is customary to define mean-field models as the ones that can be resolved by this theory.

The BP equations are apt to describe problems at the RS level, and thus fail to give correct results when RSB occurs. The cavity method approach is nonetheless still applicable: for example, a correct description at the 1RSB level can be obtained by propagating messages which describe probability distributions over BP messages. Higher levels of RSB can in principle be dealt with in the same way, by using distributions over distributions of messages, and so on. The detailed procedure can be in general described as follows: given a constraint satisfaction problem, write the corresponding BP equations; then, consider the BP messages as the variables of a new problem, and the BP equations as its constraints over such variables, and write

higher-level BP equations for this new meta-model.[5] Iterate this procedure for the desired level of symmetry breaking.

More formally, a 1RSB description is obtained by studying the model described by the following partition function:

$$Z_{1RSB} = \sum_{\boldsymbol{\mu}} e^{-m\beta F(\mu_{\partial i}, \mu_{\partial a})} \prod_i \mathbb{I}_i \left(\mu_{\partial i}\right) \prod_a \mathbb{I}_a \left(\mu_{\partial a}\right) \tag{79}$$

where

- $\mu_{\partial i} = \{\mu_{i \to a} | a \in \partial i\} \cup \{\mu_{a \to i} | a \in \partial i\}$, and analogous for $\mu_{\partial a}$;
- $\mathbb{I}_i$ and $\mathbb{I}_a$ enforce the BP equations (44) and (45);
- $F\left(\mu_{\partial i}, \mu_{\partial a}\right)$ is the Bethe free energy (63);
- m is the so-called *Parisi parameter*, which can be used to control how the different clusters of configurations are weighted with respect to each other. The RS solution corresponds to $m = 1$, while in the RSB phase $m < 1$. Setting $m = 0$ with β finite amounts at weighting all clusters equally. The limit $m \to 0$, $\beta \to \infty$ with $y = m\beta$ finite can be used to weight each cluster α according to its average energy E^α, as e^{-yE^α}.

The procedure as sketched above can of course have prohibitive computational costs, but not necessarily: for example, hard instances of the random K-SAT problem, where 1-step symmetry breaking occurs, can be efficiently solved by using the so-called Survey Propagation algorithm, which is equivalent to a simplified version of a 2-level BP, with a decimation procedure used to identify a solution (see [15, 16]).

References

1. R. Reif, *Fundamentals of Statistical and Thermal Physics* (McGraw-Hill, New York, 1965)
2. S. Ma, *Statistical Mechanics* (World Scientific, Singapore, 1985)
3. K. Huang, *Statistical Mechanics* (Wiley, New York, 1967)
4. R.S. Ellis, *Entropy, Large Deviations, and Statistical Mechanics* (Springer, New York, 1985)
5. P. Erdős, A. Rényi, On the evolution of random graphs. Publ. Math. Inst. Hung. Acad. Sci. **5**, 17 (1960)
6. B. Bollobás, *Random Graphs* (Academic, New York, 1985)
7. R. Potts, Proc. Camb. Philos. Soc. **48**, 106 (1952)
8. P. Kasteleyn, C. Fortuin, J. Phys. Soc. Jpn. Suppl. **26**, 1114 (1969)
9. F. Wu, The potts model. Rev. Mod. Phys. **54**, 235 (1982)
10. A. Engel, R. Monasson, A.K. Hartmann, On large-deviations properties of Erdős-Rényi random graphs. J. Stat. Phys. **117**, 387 (2004)

[5] This requires to extend the BP equations to models with continuous variables, which was omitted here for simplicity, but is rather straightforward.

11. O. Martin, R. Monasson, R. Zecchina, Statistical mechanics methods and phase transitions in optimization problems. Theor. Comput. Sci. **265**, 3 (2001)
12. M. Mézard, G. Parisi, M.A. Virasoro, *Spin-Glass Theory and Beyond*. Lecture Notes in Physics, vol. 9 (World Scientific, Singapore, 1987)
13. M. Mézard, A. Montanari, *Information, Physics, and Computation* (Oxford University Press, Oxford, 2009)
14. M. Talagrand, Concentration of measure and isoperimetric inequalities in product spaces. Publ. Math. de L'IHÉS **81**(1), 73–205 (1995)
15. M. Mézard, R. Zecchina, Random K-satisfiability: from an analytic solution to a new efficient algorithm. Phys. Rev. E **66**, 056126 (2002)
16. A. Braunstein, M. Mézard, R. Zecchina, Survey propagation: an algorithm for satisfiability. Random Struct. Algorithm **27**, 201–226 (2005)

http://www.springer.com/978-3-319-16966-8

Mathematical Foundations of Complex Networked Information Systems
Politecnico di Torino, Verrès, Italy 2009
Kumar, P.R.; Wainwright, M.J.; Zecchina, R. - Fagnani, F.; Fosson, S.M.; Ravazzi, C. (Eds.)
2015, VII, 135 p. 34 illus., 24 illus. in color., Softcover
ISBN: 978-3-319-16966-8