Harnessing the Bethe Free Energy*.

Victor Bapst and Amin Coja-Oghlan

Mathematics Institute, Goethe University
10 Robert Mayer St, Frankfurt 60325, Germany
{bapst, acoghlan}@math.uni-frankfurt.de

Abstract

Gibbs measures induced by random factor graphs play a prominent role in computer science, combinatorics and physics. A key problem is to calculate the typical value of the partition function. According to the "replica symmetric cavity method", a heuristic that rests on non-rigorous considerations from statistical mechanics, in many cases this problem can be tackled by way of maximising a functional called the "Bethe free energy". In this paper we prove that the Bethe free energy upper-bounds the partition function in a broad class of models. Additionally, we provide a sufficient condition for this upper bound to be tight.

1998 ACM Subject Classification G.2.1 Combinatorics, G.3 Probability and Statistics, F.2.2 Nonnumerical Algorithms and Problems

Keywords and phrases Belief Propagation, free energy, Gibbs measure, partition function

Digital Object Identifier 10.4230/LIPIcs.APPROX-RANDOM.2015.467

1 Introduction

Many problems in combinatorics, computer science and physics can be cast along the following lines [2, 23]. There are a (large) number of *variables*, each of them ranging over a finite domain Ω. The variables interact through *constraints* that each bind a few variables. Every constraint comes with a "weight function" that either encourages or discourages certain value combinations of the incident variables. The interactions can be described naturally by a *factor graph*, whose vertices are the variables and the constraints. A constraint is adjacent to the variables that it binds. The weight of an *assignment* σ that maps each variable to a value from Ω is the product of all the weights of the constraints. The obvious questions is: how many assignments of a specific total weight exist?

In this paper we are concerned with models where the factor graph is random. An excellent example is the *random k-SAT model:* there are n Boolean variables $x_1, \ldots, x_n$ and m clauses $a_1, \ldots, a_m$. Each clause binds k variables, which are chosen independently and uniformly from $x_1, \ldots, x_n$, and discourages them from taking one of the 2^k possible value combinations. This value combination is chosen uniformly and independently for each clause. The key quantity associated with the random k-SAT instance $\mathbf{\Phi}$ is its *partition function*, defined as

$$Z_{\beta,\mathbf{\Phi}} = \sum_{\sigma \in \{0,1\}^n} \prod_{i=1}^{m} \exp(-\beta \mathbf{1}\{\sigma \text{ violates } a_i\}) \qquad (\beta > 0). \tag{1}$$

* The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n. 278857–PTCC

18th Int'l Workshop on Approximation Algorithms for Combinatorial Optimization Problems (APPROX'15) / 19th Int'l Workshop on Randomization and Computation (RANDOM'15).
Editors: Naveen Garg, Klaus Jansen, Anup Rao, and José D. P. Rolim; pp. 467–480
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

In words, we sum the weights of all 2^n possible truth assignments σ. Each σ incurs a "penalty factor" of $\exp(-\beta)$ for every violated clause. It is not difficult to see that the random variable $Z_{\beta,\mathbf{\Phi}}$ incorporates key characteristics of the model. For instance, the maximum number of clauses that can be satisfied simultaneously equals

$$m + \lim_{\beta\to\infty} \frac{\partial}{\partial\beta} \ln Z_{\beta,\mathbf{\Phi}}.$$

Apart from random k-SAT, there are a host of other models of a similar nature. Prominent examples include the random graph colouring problem, LDPC codes or the so-called "mean-field" models of statistical mechanics [23].

Over the past decade the *second moment method* has emerged as the principal tool for the analysis of such models [1, 2, 15]. Its "vanilla" version works as follows. If the partition function Z of the model satisfies the bound $\mathbb{E}[Z^2] \leq O(\mathbb{E}[Z]^2)$ in the limit as the number n of variables tends to infinity, then $n^{-1}\ln(Z/\mathbb{E}[Z])$ converges to 0 in probability. Since $\mathbb{E}[Z]$ is normally easy to compute, we thus obtain the exponential order of Z. In fact, by calculating $\mathbb{E}[Z^2]/\mathbb{E}[Z]^2$ accurately enough it is sometimes possible to infer the limiting distribution of Z [20].

However, in many examples the use of the second moment method is precluded by large deviations phenomena. The random k-SAT model with $m = \lceil \alpha n \rceil$ clauses is a case in point as $n^{-1}\ln(Z_{\beta,\mathbf{\Phi}}/\mathbb{E}[Z_{\beta,\mathbf{\Phi}}])$ does *not* converge to 0 as $n \to \infty$ for any $\alpha, \beta > 0$. The reason is that the first moment $\mathbb{E}[Z_{\beta,\mathbf{\Phi}}]$ is driven up by a "lottery effect": there are a tiny minority of formulas with an abundance of "good" assignments [1, 3, 4]. Of course, this implies that $\mathbb{E}[Z_{\beta,\mathbf{\Phi}}^2] \geq \exp(\Omega(n))\mathbb{E}[Z_{\beta,\mathbf{\Phi}}]^2$. Thus, the second moment method fails rather spectacularly.

The obvious remedy is to condition such lottery effects away. That is, we ought to condition on an event $\mathcal{U}$ that pins down those parameters of the model whose large deviations drive $\mathbb{E}[Z]$ up. But even if we manage to identify the relevant parameters, the necessary conditioning on $\mathcal{U}$ may be so complicated as to render a second moment computation at best unpleasant and at worst infeasible. Indeed, the recent history of the random k-SAT problem illustrates how conditioning turns a second moment computation into a formidable task [7, 12].

A completely different but non-rigorous method for calculating Z, the *replica symmetric cavity method*, has been suggested on the basis of ideas from statistical physics [23]. According to the cavity method, under certain assumptions the asymptotic value of $n^{-1}\ln Z$ can be calculated by maximising a functional called the *Bethe free energy*. Furthermore, the physics recipe for solving this maximisation problem is to iterate a message passing algorithm called *Belief Propagation* on the factor graph until convergence. This recipe is somewhat plausible due to the (rigorous) fact that the stationary points of the Bethe free energy are in one-to-one correspondence with the Belief Propagation fixed points [32]. However, in general there are several fixed points and non-trivial insights are necessary to steer Belief Propagation toward the "correct" one. Even worse, in general the maximum value of the Bethe free energy may or may not approximate $n^{-1}\ln Z$ well.[1]

The purpose of this paper is to provide a rigorous foundation for the idea of using Belief Propagation to calculate the free energy. We establish two main results. First, that under mild assumptions the maximum of the Bethe free energy provides an upper bound on the typical value of $n^{-1}\ln Z$ on a random factor graph (Theorem 3). The proof of this is based

[1] The quantity $n^{-1}\ln Z$ is called the *free energy* of the factor graph. We do not use this term to avoid confusion with the *Bethe free energy*.

on a physics-enhanced version of the classical "first moment method". Along the way we derive several general results on Gibbs distributions that should be of independent interest (e.g., Theorem 6). Second, we propose a corresponding refined "second moment method" (Theorem 14). More specifically, we prove that if the maximum of the Bethe free energy on a certain auxiliary model is upper-bounded by a term that corresponds to the square of the first moment and if certain additional (reasonable) assumptions hold, then the free energy converges in probability to the value predicted by the cavity method.

2 Related work

Belief Propagation has been re-discovered several times in varying degrees of generality [5, 17, 29]. On finite acyclic factor graphs Belief Propagation has a unique fixed point and the corresponding Bethe free energy equals $n^{-1} \ln Z$. (e.g., [23, Chapter 14]). To what extent this is true in the presence of cycles is a long-standing problem.

The results of the present paper are most relevant in cases where the local structure of the factor graph is not perfectly "uniform". For instance, we are going to be interested in the case that different variable nodes may have different degrees. More subtly, different variable nodes may have different marginals under the Gibbs distribution that the factor graph induces, see (3) below. The case of uniform models is conceptually simpler and has been treated before [9]. In fact, in the uniform case the computation of $n^{-1} \ln Z$ can essentially be transformed into the problem of maximising the Bethe free energy of a "tensorised" model on the d-regular tree [9, 13, 11]. This fact has played a key role in recent work on the hardness of counting problems [16, 26, 30]. Although we use a similar tensor construction in our second moment argument as well (cf. Proposition 13), non-uniformity makes matters far more intricate, as witnessed by recent work on random k-SAT [7, 12]. Thus, the main point of the present work is to establish a connection between the Bethe free energy and $|V|^{-1} \ln Z$ even in the non-uniform case.

That said, if the model enjoys certain spatial mixing properties (such as "Gibbs uniqueness"), then the Bethe free energy is known to yield the correct value of $n^{-1} \ln Z$ even in the non-uniform case [10, 25]. However, the necessary spatial mixing properties are quite strong and they cease to be satisfied, e.g., in the random k-SAT model from (1) for large β for clause/variable ratios as low as about $\ln k/k$ [25]. By comparison, the k-SAT threshold is about $2^k \ln 2$ [12].

The "interpolation method" provides a different approach to calculating or at least upper-bounding $n^{-1} \ln Z$ [14, 19]. In particular, the upper bound comes in a variational form [28]. For example, this can be used to obtain a tight upper bound on the k-SAT threshold [12]. Generally speaking, the interpolation method is great if it works, but it comes with certain (convexity-type) assumptions that are not always satisfied. Furthermore, it seems difficult to use the interpolation method directly to carry out a second moment argument in order to lower-bound the partition function. By contrast, Theorems 3 and 14 do not require such assumptions.

The physicists' cavity method comes in several instalments; for a detailed discussion we refer to [23]. In this paper we are chiefly concerned with the simplest, "replica symmetric" variant. This version does not always provide the correct value of $n^{-1} \ln Z$ [8]. It seems that one reason for this is that models such as random k-SAT undergo a so-called "condensation phase transition" [21]. The more complex "1-step replica symmetry breaking (1RSB)" version of the cavity method [24] is expected to yield the correct value of $n^{-1} \ln Z$ some way beyond condensation. However, another phase transition called *full replica symmetry breaking* seems to spell doom on even the 1RSB cavity method (see [23] for details). In summary, we do not hope for an unconditional result that vindicates either the replica symmetric or the 1RSB version of the cavity method.

3 Random factor graphs

In this section we explain the class of models that we deal with. Throughout, $\Delta > 0$ is an integer, Ω, Θ are finite sets and $\Psi = \{\psi_1, \ldots, \psi_l\}$ is a finite set of maps $\psi_i : \Omega^{h_i} \to (0, \infty)$, where $1 \le h_i \le \Delta$. The following abstract definition encompasses a multitude of concrete examples.

▶ **Definition 1.** A $(\Delta, \Omega, \Psi, \Theta)$-*model* $\mathcal{M} = (V, F, d, t, \psi)$ consists of
M1 a countable set V of ***variable nodes***,
M2 a countable set F of ***constraint nodes***,
M3 a map $d : V \cup F \to [\Delta]$ such that $\sum_{x \in V} d(x) = \sum_{a \in F} d(a)$,
M4 a map $t : C_V \cup C_F \to \Theta$, where we let

$$C_V = \bigcup_{x \in V} \{x\} \times [d(x)], \qquad\qquad C_F = \bigcup_{a \in F} \{a\} \times [d(a)],$$

such that $\left|t^{-1}(\theta) \cap C_V\right| = \left|t^{-1}(\theta) \cap C_F\right|$ for each $\theta \in \Theta$,
M5 a map $F \to \Psi$, $a \mapsto \psi_a$ such that $\psi_a : \Omega^{d(a)} \to (0, \infty)$ for all $a \in F$.
The ***size*** of the model is defined as $\#\mathcal{M} = |V|$. Furthermore, a ***$\mathcal{M}$-factor graph*** is a bijection

$$G : C_V \to C_F,\ (x, i) \mapsto G_{x,i} \qquad \text{such that} \qquad t(G_{x,i}) = t(x, i) \text{ for all } (x, i) \in C_V.$$

Of course, the equalities in **M3** and **M4** require that either both quantities are infinite or both are finite, in which case they have to coincide.

The semantics is that the map d prescribes the degree of each variable and constraint node (i.e., their number of neighbours in any $\mathcal{M}$-factor graph). Just like in the “configuration model” of graphs with a given degree sequence we create $d(v)$ “clones” of each node v. The sets C_V, C_F contain the clones of the variable and constraint nodes, respectively. Additionally, the map t assigns each clone a “type” from the set Θ. Moreover, each constraint node a comes with a “weight function” ψ_a from the set Ψ.

Like in the “configuration model” a $\mathcal{M}$-factor graph is a type-preserving matching G of the variable and constraint clones. Let $\mathcal{G}(\mathcal{M})$ be the set of all $\mathcal{M}$-factor graphs and let $\boldsymbol{G}(\mathcal{M})$ denote a uniformly random sample from $\mathcal{G}(\mathcal{M})$. We usually think of $G \in \mathcal{G}(\mathcal{M})$ as the (multi-)graph obtained by contracting the clones of each node. Clearly, this yields a bipartite graph with $|V|$ variable nodes and $|F|$ constraint nodes. For a node $x \in V$ we denote by $\partial_G x$ the set of neighbours of x in this multi-graph, i.e., the set of all $a \in F$ such that there exist $i \in [d(x)]$, $j \in [d(a)]$ such that $G_{x,i} = (a, j)$. Analogously, for $a \in F$ and $j \in [d(a)]$ we write $\partial_G(a, j) = x$ if there is $i \in [d(x)]$ such that $G_{x,i} = (a, j)$. Moreover, $\partial_G a = \{\partial_G(a, j) : j \in [d(a)]\}$. Finally, we denote the inverse image of a clone $(a, j) \in C_F$ under the bijection G simply by $G_{a,j}$.

A ***$\mathcal{M}$-assignment*** is a map $\sigma : V \to \Omega$. Let $\mathcal{C}_{\mathcal{M}}$ be the set of all $\mathcal{M}$-assignments. Further, define the ***partition function*** of $G \in \mathcal{G}(\mathcal{M})$ as

$$Z_G = \sum_{\sigma \in \mathcal{C}_{\mathcal{M}}} \prod_{a \in F} \psi_a\big(\sigma(\partial_G(a, 1)), \ldots, \sigma(\partial_G(a, d(a)))\big). \tag{2}$$

It is closely intertwined with the ***Gibbs distribution*** of G, which is the distribution on $\mathcal{C}_{\mathcal{M}}$ defined by

$$\mu_G(\sigma) = Z_G^{-1} \prod_{a \in F} \psi_a\big(\sigma(\partial_G(a,1)), \ldots, \sigma(\partial_G(a,d(a)))\big). \tag{3}$$

Our key object of study is the random variable $|V|^{-1} \ln Z_{\boldsymbol{G}(\mathcal{M})}$.

▶ **Example 2** (The random k-SAT model)**.** Let $\Omega = \{0,1\}$. Given some $\beta \geq 0$ let Ψ contain the 2^k weight functions

$$\psi^{(\tau)} : \Omega^k \to (0,\infty), \qquad \sigma \mapsto \exp(-\beta \mathbf{1}\{\sigma = \tau\}) \qquad \text{for } \tau \in \Omega^k.$$

Let $\Delta > 0$ be a positive integer and let $\Theta = \{\star\}$. We obtain a $(\Delta, \Omega, \Psi, \Theta)$-model $\mathcal{M}_{\text{SAT}}$ by letting $V = \{x_1, \ldots, x_n\}$ and $F = \{a_1, \ldots, a_m\}$. Pick any degree sequence $d : V \to [\Delta]$ such that $\sum_{x \in V} d(x) = km$ and let $d(a) = k$ for all $a \in F$. Further, pick some $\psi_a \in \Psi$ for each $a \in F$, thereby prescribing a "sign pattern" for each "clause" a. Finally, let $t : C_V \cup C_F \to \Theta$ be the trivial (constant) map. Then $\boldsymbol{G}(\mathcal{M}_{\text{SAT}})$ corresponds to choosing a random k-SAT formula with the given degree sequence and sign patterns. Moreover, $n^{-1} \ln Z_{\boldsymbol{G}(\mathcal{M}_{\text{SAT}})}$ accounts for weighted truth assignments (cf. (1)) [18].

In Example 2 we did not actually use the types in a non-trivial way. They could be used to prescribe not merely the degree of each variable but also how many times each Boolean variable appears positively or negatively.

While Definition 1 encompasses a many problems of interest, there are two restrictions. They arise because we are going to be interested in *sequences* $(\mathcal{M}_n)_n$ of $(\Delta, \Omega, \Psi, \Theta)$-models of sizes $\#\mathcal{M}_n = n$. That is, the size of the model tends to infinity while $\Delta, \Omega, \Psi, \Theta$ remain fixed. In effect, the maximum degree remains bounded as $n \to \infty$. This is not quite the case in, e.g., the "standard" random k-SAT model where clauses are chosen uniformly and independently and where consequently the variable degrees are asymptotically Poisson. However, in such examples the free energy can by means of standards arguments be approximated arbitrarily well by truncating the degrees at a large enough Δ.

The second restriction is that the weight functions $\psi \in \Psi$ are assumed to be strictly positive. This condition precludes *hard* constraints such as "no single clause must be violated". Although most of our proofs extend to the case of hard constraints, we chose to exclude them from the general statement of the results for the sake of clarity. For instance, the positivity assumption ensures that $Z_G > 0$ for all $G \in \mathcal{G}(\mathcal{M}_n)$ and hence that the random variable $n^{-1} \ln Z_{\boldsymbol{G}(\mathcal{M}_n)}$ has a finite mean. Furthermore, the case of hard constraints can be handled by introducing an "inverse temperature" parameter $\beta > 0$ like in Example 2 and ultimately taking the limit $\beta \to \infty$ (cf. [25]), although some additional work is needed.

In Section 4 we will prove that the "Bethe free energy" provides an upper bound on $|V|^{-1} \ln Z_{\boldsymbol{G}(\mathcal{M})}$. Further, in Section 5 we are going to provide a sufficient condition under which this upper bound is asymptotically tight.

Preliminaries

Throughout the paper we always let $\Delta \geq 1$ be an integer, Ω, Θ finite sets, and Ψ a finite set of functions as in Definition 1. We let $\#\psi$ be the arity of $\psi \in \Psi$, i.e., $\psi : \Omega^{\#\psi} \to (0,\infty)$.

For a finite set $\mathcal{X} \neq \emptyset$ we denote by $\mathcal{P}(\mathcal{X})$ the set of probability measures on $\mathcal{X}$, which we identify with the $|\mathcal{X}|$-simplex. For $\mu \in \mathcal{P}(\mathcal{X})$ we denote by $H(\mu) = -\sum_{x \in \mathcal{X}} \mu(x) \ln \mu(x)$

the entropy of μ (with the convention $0\ln 0 = 0$). Further, if $\mu, \nu : \mathcal{X} \to [0, \infty)$ are such that $\nu(x) > 0$ only if $\mu(x) > 0$, then

$$D(\nu\|\mu) = \sum_{x\in\mathcal{X}} \nu(x)\ln(\nu(x)/\mu(x))$$

signifies the Kullback-Leibler divergence. Moreover, for integers $k > 0$, $j \in [k]$ and $\mu \in \mathcal{P}(\mathcal{X}^k)$ we let $\mu_{\downarrow j} \in \mathcal{P}(\mathcal{X})$ be the marginal of the jth component.

For $\mu \in \mathcal{P}(\mathcal{X})$ we write $\boldsymbol{\sigma}_\mu$ for an random element of $\mathcal{X}$ chosen according to μ. Where μ is apparent from the context we drop the index. Further, if $X : \mathcal{X} \to \mathbb{R}$ is a random variable we write $\langle X \rangle_\mu = \sum_{\sigma\in\mathcal{X}} X(\sigma)\mu(\sigma)$ for the expectation of X with respect to μ. For the sake of brevity we normally write $\langle \cdot \rangle_G$ instead of $\langle \cdot \rangle_{\mu_G}$ for $G \in \mathcal{G}(\mathcal{M})$.

Further, if S is a subset of the set V of variable nodes of $\mathcal{M}$, $\sigma : V \to \Omega$ and $\omega \in \Omega$ we write

$$\sigma[\omega|S] = \frac{1}{|S|}\sum_{x\in S} \mathbf{1}\{\sigma(x) = \omega\}.$$

Thus, $\sigma[\,\cdot\,|S] \in \mathcal{P}(\Omega)$ is the empirical distribution of σ on S. Analogously, if $G \in \mathcal{G}(\mathcal{M})$ is a factor graph and $A \neq \emptyset$ is a set of factor nodes such that all $a \in A$ have degree $d(a) = l$ for some $l > 0$, then we let

$$\sigma[\omega_1, \dots, \omega_l|A] = \frac{1}{|A|}\sum_{a\in A}\prod_{j=1}^{l} \mathbf{1}\{\sigma(\partial_G(a,j)) = \omega_j\}.$$

Thus, $\sigma[\,\cdot\,|A] \in \mathcal{P}(\Omega^l)$ is the joint empirical distribution of the value combinations induced by σ on $a \in A$.

4 The upper bound

Let $\mathcal{M} = (V, F, d, t, (\psi_a)_{a\in F})$ be a $(\Delta, \Omega, \Psi, \Theta)$-model of finite size $n = |V|$. Let $\boldsymbol{G} = \boldsymbol{G}(\mathcal{M})$ for brevity.

4.1 The Bethe free energy

The aim in this section is to show that the "Bethe free energy", a concept that hails from the cavity method, provides an upper bound on the partition function. To formulate the result we need the following definition [23, Chapter 14]. Let $G \in \mathcal{G}(\mathcal{M})$. A ***marginal sequence*** of G is a family $\nu = (\nu_x, \nu_a)_{x\in V, a\in F}$ such that $\nu_x \in \mathcal{P}(\Omega)$ for each $x \in V$, $\nu_a \in \mathcal{P}(\Omega^{d(a)})$ for each $a \in F$ and if $G_{x,i} = (a, j)$ entails that $\nu_x = \nu_{a\downarrow j}$. Thus, if a variable x occurs in the jth position of a constraint a, then the jth marginal of ν_a coincides with ν_x. The ***Bethe free energy***[2] of (G, ν) is

$$\mathcal{B}_{\mathcal{M}}(G, \nu) = -\frac{1}{n}\left[\sum_{a\in F} D(\nu_a\|\psi_a) + \sum_{x\in V}(d(x) - 1)H(\nu_x)\right].$$

Additionally, the ***Bethe free energy of*** G is

$$\mathcal{B}_{\mathcal{M}}(G) = \max\{\mathcal{B}_{\mathcal{M}}(G, \nu) : \nu \text{ is a marginal sequence of } G\}.$$

[2] For a detailed derivation of the Bethe free energy in the context of the cavity method see [23, Chapter 14].

▶ **Theorem 3.** *For any $\Delta, \Omega, \Psi, \Theta$ and any $\varepsilon > 0$ there exists $n_0 > 0$ such that the following is true. Suppose that $\mathcal{M}$ is a finite $(\Delta, \Omega, \Psi, \Theta)$-model of size $n > n_0$. Moreover, let $\emptyset \neq \mathcal{U} \subset \mathcal{G}(\mathcal{M})$ be an event. Then*

$$n^{-1} \ln \mathbb{E}[Z_{\boldsymbol{G}} \mathbf{1}\{\boldsymbol{G} \in \mathcal{U}\}] \leq \max\{\mathcal{B}_{\mathcal{M}}(G) : G \in \mathcal{U}\} + \varepsilon.$$

Thus, there exists a number n_0 that depends *only* on the basic parameters $\Delta, \Omega, \Psi, \Theta$ and the desired accuracy ε such that for any model of size $n \geq n_0$ the Bethe free upper bounds on the expectation of $Z_{\boldsymbol{G}}$ on $\mathcal{U}$. The following corollary provides a handy way to apply Theorem 3.

▶ **Corollary 4.** *Let $(\mathcal{M}_n)_n$ be a sequence of $(\Delta, \Omega, \Psi, \Theta)$-models such that $\#\mathcal{M}_n = n$. Assume that $b > 0$ is such that the event $\mathcal{U}_n = \{\mathcal{B}_{\mathcal{M}_n}(\boldsymbol{G}(\mathcal{M}_n)) \leq b\}$ satisfies $\lim_{n\to\infty} \mathbb{P}[\mathcal{U}_n] = 1$. Then*

$$\limsup_{n\to\infty} n^{-1} \ln \mathbb{E}\left[Z_{\boldsymbol{G}(\mathcal{M}_n)} | \mathcal{U}_n\right] \leq b.$$

By Markov's and Jensen's inequalities, the bound $\limsup_{n\to\infty} n^{-1} \ln \mathbb{E}\left[Z_{\boldsymbol{G}(\mathcal{M}_n)} | \mathcal{U}_n\right] \leq b$ entails that

$$\lim_{\varepsilon \searrow 0} \lim_{n\to\infty} \mathbb{P}\left[n^{-1} \ln Z_{\boldsymbol{G}(\mathcal{M}_n)} \leq b + \varepsilon\right] = 1.$$

In other words, if the Bethe free energy is bounded by b with high probability, then $n^{-1} \ln Z_{\boldsymbol{G}(\mathcal{M}_n)} \leq b + o(1)$ with high probability.

The proof of Theorem 3 contains several concepts that we deem to be of independent interest. The most important one is that of a "state". More specifically, we prove Theorem 3 by showing that the lion's share of $\mathbb{E}[Z_{\boldsymbol{G}} \mathbf{1}\{\boldsymbol{G} \in \mathcal{U}\}]$ comes from a set Γ of factor graph/assignment pairs (G, σ) such that certain key parameters of all pairs $(G, \sigma) \in \Gamma$ approximately coincide. For instance, for (almost) any ψ and value combination $\omega = (\omega_1, \ldots, \omega_{\#\psi})$, about the same number of constraint nodes a with $\psi_a = \psi$ display the value combination ω. Theorem 3 will follow because the contribution of any single state s to $\mathbb{E}[Z_{\boldsymbol{G}} \mathbf{1}\{\boldsymbol{G} \in \mathcal{U}\}]$ can be cast as the Bethe free energy of a marginal sequence induced by s. We proceed with the precise definition of states.

4.2 States

For an integer $N \geq 1$ we write

$$\Psi[N] = \{(\psi, h_1, \ldots, h_{\#\psi}) : h_1, \ldots, h_{\#\psi} \in [N]\}.$$

▶ **Definition 5.** A $\mathcal{M}$-***state*** of size $N \geq 1$ consists of

ST1 a map $s : V \to [N]$ such that $s(x) = s(y)$ only if $d(x) = d(y)$ and $t(x, i) = t(y, i)$ for all $i \in d(x)$,

ST2 a probability distribution $\bar{s} = (\bar{s}_{\psi,h})_{\psi,h}$ on $\Psi[N]$,

ST3 a set $\hat{s} \subset \Psi[N]$,

ST4 a sequence $(\tilde{s}_h)_{h\in[N]}$ of probability distributions on Ω,

ST5 for any $(\psi, h) \in \Psi[N]$ a distribution $\tilde{s}_{\psi,h} \in \mathcal{P}(\Omega^{\#\psi})$ such that $\tilde{s}_{\psi,h\downarrow j} = \tilde{s}_{h_j}$ for all $j \in [\#\psi]$.

Normally we denote an $\mathcal{M}$-state simply by s and we write $\#s$ for its size. Moreover, let

$$V_h^s = s^{-1}(h) \quad \text{for } h \in [\#s], \qquad V_{\psi,h}^s = \prod_{j\in[\#\psi]} V_{h_j}^s \quad \text{for } (\psi, h) \in \Psi[N].$$

In addition, if $G \in \mathcal{G}(\mathcal{M})$ and $(\psi, h) \in \Psi\,[N]$ we let

$$\partial_{G,s}(\psi, h) = \left\{a \in F : \psi_a = \psi, \partial_G(a) \in V^s_{\psi,h}\right\}.$$

Thus, a state induces a partition $V_1^s, \ldots, V_N^s$ of the set of variable nodes. Condition **ST1** ensures that this partition respects the degrees and the types. Let us call $G \in \mathcal{G}(\mathcal{M})$ ε***-compatible*** with s for some $\varepsilon > 0$ ("$G \models_\varepsilon s$") if

$$\sum_{(\psi,h)\in\Psi[N]} \left|\frac{|\partial_{G,s}(\psi,h)|}{|F|} - \bar{s}_{\psi,h}\right| < \varepsilon, \qquad\qquad \sum_{(\psi,h)\in\Psi[N]} \mathbf{1}\left\{(\psi,h) \in \hat{s}\right\} \bar{s}_{\psi,h} < \varepsilon.$$

Thus, for any (ψ, h) there are about $\bar{s}_{\psi,h}|F|$ constraint nodes a with $\psi_a = \psi$ that join variable nodes from the classes $V^s_{h_1}, \ldots, V^s_{h_{\#\psi}}$. And no more than an ε fraction of all constraint nodes belong to the "rogue classes" $(\psi, h) \in \hat{s}$.

Further, suppose that $G \models_\varepsilon s$ and $\sigma \in \mathcal{C}_\mathcal{M}$. We say that (G, σ) is ε***-judicious*** with respect to s (in symbols: $(G, \sigma) \models_\varepsilon s$) if

J1 for all $h \in [N]$ we have $\|\tilde{s}_h - \sigma[\,\cdot\,|V_h^s]\|_{\mathrm{TV}} < \varepsilon$,

J2 for all $(\psi, h) \in \Psi\,[N] \setminus \hat{s}$ such that $\partial_{G,s}(\psi, h) \neq \emptyset$ we have $\|\tilde{s}_{\psi,h} - \sigma[\,\cdot\,|\partial_{G,s}(\psi, h)]\|_{\mathrm{TV}} < \varepsilon$.

Hence, the empirical distributions $\sigma[\,\cdot\,|V_h^s]$ do not deviate by more than ε from $\tilde{s}_h$. Similarly, for a "non-rogue" (ψ, h) the empirical distribution $\sigma[\,\cdot\,|\partial_{G,s}(\psi, h)]$ of the ψ-factors that connect variables in $V^s_{\psi,h}$ is within ε of $\tilde{s}_{\psi,h}$. The following theorem provides the key fact about states. It should be of interest in its own right.

► **Theorem 6.** *For any $\Delta, \Omega, \Psi, \Theta$ and any $\varepsilon > 0$ there exists $\eta > 0$ such that the following is true. Let $\mathcal{M}$ be a finite $(\Delta, \Omega, \Psi, \Theta)$-model of size $\#\mathcal{M} \geq 1/\eta$ and let $G \in \mathcal{G}(\mathcal{M})$. Then there exists a $\mathcal{M}$-state s of size $\#s \leq 1/\eta$ such that $G \models_\varepsilon s$ and $\langle \mathbf{1}\left\{(G, \boldsymbol{\sigma}) \models_\varepsilon s\right\}\rangle_G \geq \eta$.*

Crucially, the number η promised by Theorem 6 depends on ε and the basic parameters $\Delta, \Omega, \Psi, \Theta$ only. It is independent of the model and its size. Hence, for *any* large enough $\mathcal{M}$ and *any* $G \in \mathcal{G}(\mathcal{M})$ there is a single "dominant state" s that captures a constant fraction of the mass of the Gibbs distribution μ_G.

Theorem 6 sits well with the *replica symmetry breaking* picture drafted by the cavity method. According to this prediction, there are three possible shapes that the Gibbs distribution can take. Roughly speaking, in the case of *replica symmetry* the joint distribution of any two variable nodes that are far apart (say, at distance at least $\ln\ln n$) in the factor graph is close to a product distribution. The state corresponding to this scenario simply partitions the variable nodes according to their Gibbs marginals. In the second scenario, called *1-step replica symmetry breaking*, the Gibbs distribution is mixture of a bounded number of distributions, i.e.,

$$\left\|\mu_G - \sum_{i=1}^{K} w_i \mu_{G,i}\right\|_{\mathrm{TV}} < \varepsilon \qquad \text{where } (w_1, \ldots, w_K) \in \mathcal{P}([K]), \quad \mu_{G,1}, \ldots, \mu_{G,K} \in \mathcal{P}(\mathcal{C}_\mathcal{M}).$$

Each $\mu_{G,i}$ corresponds to a "cluster" of assignments and is such that the joint distribution of far apart variables factorises. In this case, we obtain a state by partitioning the variables according to their $\mu_{G,i}$-marginals for some i with $w_i \geq \eta$. In the third case, called *full replica symmetry breaking*, the $\mu_{G,i}$ themselves are mixtures of distributions $\mu_{G,i,j}$. Further, each of the $\mu_{G,i,j}$ decomposes into clusters etc., yielding an infinite cascade. A dominant state would truncate the cascade after a finite number of steps (depending on ε) and home in on one of the sub-clusters.

The key concept behind the proof of Theorem 6 is the following.

▶ **Definition 7.** Let Ω be a finite set, let $\varepsilon > 0$, let n be an integer and let μ be a probability measure on Ω^n. A partition $\boldsymbol{V} = (V_1, \ldots, V_N)$ of $[n]$ is called ***ε-homogeneous with respect to*** μ if there is a set $J \subset [N]$ such that $\sum_{i \in [N] \setminus J} |V_i| < \varepsilon n$ and such that for all $j \in J$ the following is true.

For any subset $S \subset V_j$ of size $|S| \geq \varepsilon |V_j|$ we have $\left\langle \|\boldsymbol{\sigma}[\,\cdot\,|S] - \boldsymbol{\sigma}[\,\cdot\,|V_j]\|_{\mathrm{TV}} \right\rangle_\mu < \varepsilon$.

If $\boldsymbol{V} = (V_1, \ldots, V_k)$, then we call $\#\boldsymbol{V} = k$ the ***size*** of $\boldsymbol{V}$. Furthermore, a partition $\boldsymbol{W} = (W_1, \ldots, W_l)$ ***refines*** another partition $\boldsymbol{V} = (V_1, \ldots, V_k)$ if for each $i \in [l]$ there is $j \in [k]$ such that $W_i \subset V_j$. The proof of Theorem 6 builds upon

▶ **Theorem 8.** *Let Ω be a finite set. For any $\varepsilon > 0$ there exists $N = N(\varepsilon, \Omega)$ such that for $n > N$ and any probability measure μ on Ω^n the following is true. Let $\boldsymbol{V}_0$ be a partition of $[n]$ such that $\#\boldsymbol{V}_0 \leq 1/\varepsilon$. Then $\boldsymbol{V}_0$ has a refinement $\boldsymbol{V}$ of size $\#\boldsymbol{V} \leq N$ that is ε-homogeneous with respect to μ.*

Theorem 8 and its proof are inspired by the proof of Szemerédi's regularity lemma [31].

Theorem 6 produces a "dominant state" for each individual factor graph. In combination with a compactness argument this entails that a single state suffices to approximate $\frac{1}{n} \ln \mathbb{E}[Z_{\boldsymbol{G}} \mathbf{1}\{\boldsymbol{G} \in \mathcal{U}\}]$ for a given event $\mathcal{U}$.

▶ **Corollary 9.** *For any $\varepsilon > 0$ and any $\Delta, \Omega, \Psi, \Theta$ there exist $\gamma > 0, n_0 > 0$ such that the following is true. Suppose that $\mathcal{M}$ is a finite $(\Delta, \Omega, \Psi, \Theta)$-model of size $\#\mathcal{M} \geq n_0$ and that $\emptyset \neq \mathcal{U} \subset \mathcal{G}(\mathcal{M})$. Then there exist a $\mathcal{M}$-state s and $G_0 \in \mathcal{U}$ such that $G_0 \models_\gamma s$ and*

$$n^{-1} \ln \mathbb{E}[Z_{\boldsymbol{G}} \mathbf{1}\{\boldsymbol{G} \in \mathcal{U}\}] \leq \varepsilon + n^{-1} \ln \mathbb{E}\left[Z_{\boldsymbol{G}} \left\langle (\boldsymbol{G}, \boldsymbol{\sigma}) \models_\gamma s \right\rangle_{\boldsymbol{G}} \middle| \boldsymbol{G} \models_\gamma s\right].$$

Finally, it is not difficult to derive Theorem 3. Indeed,

$$n^{-1} \ln \mathbb{E}\left[Z_{\boldsymbol{G}} \left\langle (\boldsymbol{G}, \boldsymbol{\sigma}) \models_\gamma s \right\rangle_{\boldsymbol{G}} \middle| \boldsymbol{G} \models_\gamma s\right]$$

can be cast as the Bethe free energy of the marginal sequence induced by s: let $\nu_x = \tilde{s}_{s(x)}$ for $x \in V$ and $\nu_a = \tilde{s}_{\psi,h}$ for all $a \in \partial_{G_0,s}(\psi, h)$. But how we can get a handle on the Bethe free energy $\mathcal{B}_{\mathcal{M}}(G)$?

4.3 Belief Propagation

The Bethe free energy of a given factor graph G can be calculated by analysing the Belief Propagation message passing algorithm. Belief Propagation can be viewed as an operator acting on the ***message space*** $\mathrm{Mes}_{\mathcal{M}}(G)$ of G, which we define as the set of all maps $\hat{\nu} : C_V \cup C_F \to \mathcal{P}(\Omega)$, $(v, j) \mapsto \hat{\nu}_{v,j}$. The ***Belief Propagation operator*** $\mathrm{BP} : \mathrm{Mes}_{\mathcal{M}}(G) \to \mathrm{Mes}_{\mathcal{M}}(G)$ maps $\hat{\nu} \in \mathrm{Mes}_{\mathcal{M}}(G)$ to $\tilde{\nu} = \mathrm{BP}(\hat{\nu})$ defined by[3]

$$\tilde{\nu}_{x,i}(\omega_i) \propto \prod_{h \in [d(x)] \setminus \{i\}} \hat{\nu}_{G_{x,h}}(\omega_i) \tag{4}$$

for $(x, i) \in C_V, \omega_i \in \Omega$ and

$$\tilde{\nu}_{a,j}(\omega_j) \propto \sum_{(\omega_h)_{h \in [d(a)] \setminus \{j\}}} \psi_a(\omega_1, \ldots, \omega_{d(a)}) \prod_{h \in [d(a)] \setminus \{j\}} \hat{\nu}_{G_{a,h}}(\omega_h) \tag{5}$$

[3] As per common practice, we use the $\propto$ symbol to define probability distributions on a finite set $\mathcal{X}$ as follows. If $f : \mathcal{X} \to [0, \infty)$, then $p \propto f$ means that $p(\omega) = f(\omega) / \sum_{x \in \mathcal{X}} f(x)$ unless $\sum_{x \in \mathcal{X}} f(x) = 0$, in which case p is the uniform distribution.

for $(a,j) \in C_F, \omega_1, \ldots, \omega_{d(a)} \in \Omega$. Let $\mathrm{Fix}_{\mathcal{M}}(G)$ be the set of all Belief Propagation fixed points, i.e., all $\hat{\nu} \in \mathrm{Mes}_{\mathcal{M}}(G)$ such that $\mathrm{BP}(\hat{\nu}) = \hat{\nu}$. Any point $\hat{\nu} \in \mathrm{Fix}_{\mathcal{M}}(G)$ gives rise to a marginal sequence, namely

$$\hat{\hat{\nu}}_x(\omega) \propto \prod_{h \in [d(x)]} \hat{\nu}_{G_{x,h}}(\omega) \tag{6}$$

for $x \in V, \omega \in \Omega$ and

$$\hat{\hat{\nu}}_a(\omega_1, \ldots, \omega_{d(a)}) \propto \psi_a(\omega_1, \ldots, \omega_{d(a)}) \prod_{h \in [d(a)]} \hat{\nu}_{G_{a,h}}(\omega_h) \tag{7}$$

for $a \in F, \omega_1, \ldots, \omega_{d(a)} \in \Omega$.

▶ **Proposition 10.** *We have* $\mathcal{B}_{\mathcal{M}}(G) = \max\left\{\mathcal{B}_{\mathcal{M}}(G, \hat{\hat{\nu}}) : \hat{\nu} \in \mathrm{Fix}_{\mathcal{M}}(G)\right\}$.

Proof. The set M of marginal sequences is compact. Because the functions $\psi \in \Psi$ are strictly positive and as the derivative of the entropy diverges as ν approaches the boundary ∂M, $\mathcal{B}_{\mathcal{M}}(G, \cdot\,)$ does not attain its global maximum on ∂M. Furthermore, for any stationary point $\nu \in M$ of the Bethe free energy $\mathcal{B}_{\mathcal{M}}(G, \cdot\,)$ there exists $\hat{\nu} \in \mathrm{Fix}_{\mathcal{M}}(G)$ such that $\mathcal{B}_{\mathcal{M}}(G, \nu) = \mathcal{B}_{\mathcal{M}}(G, \hat{\hat{\nu}})$ [23, Proposition 14.6]. ◀

Theorem 3 shows that the Bethe free energy provides an upper bound on $n^{-1} \ln Z_{\boldsymbol{G}(\mathcal{M})}$. Furthermore, Proposition 10 reduces the problem of calculating the Bethe free energy to the task of determining the "dominant fixed point" of Belief Propagation, i.e., the task of analysing an algorithm on a random graph.

5 The lower bound

In this section we consider a sequence $(\mathcal{M}(n) = (V_n, F_n, d_n, t_n, (\psi_{n,a}))_n$ *of* $(\Delta, \Omega, \Psi, \Theta)$*-models such that* $\#\mathcal{M}(n) = n$. *Let* $\boldsymbol{G}(n) = \boldsymbol{G}(\mathcal{M}(n))$ *and* $\mathcal{G}(n) = \mathcal{G}(\mathcal{M}(n))$.

5.1 A Bethe-enhanced second moment method

The cavity method provides a "recipe" for calculating a number ϕ such that $(n^{-1} \ln Z_{\boldsymbol{G}(n)})_n$ is deemed to converge to ϕ in probability. This number is determined by applying Belief Propagation and the Bethe free energy to the "limit" of the typical local structure of $\boldsymbol{G}(n)$ as $n \to \infty$. The aim in this section is to develop a version of the second moment method that allows us to prove such a claim rigorously. But first we need to formalise the "limiting local structure". To this end we adapt the concept of *local weak convergence* of graph sequences [22, Part 4] to our current setup, which can be viewed as a generalisation of the one from [10].

▶ **Definition 11.** A $(\Delta, \Omega, \Theta, \Psi)$-***template*** consists of a $(\Delta, \Omega, \Psi, \Theta)$-model $\mathcal{M}$, a connected factor graph $H \in \mathcal{G}(\mathcal{M})$ and a ***root*** (r_H, i_H), which is a variable or factor clone. We denote the template by H. Its ***size*** is $\#H = \#\mathcal{M}$.

Two templates H, H' with models $\mathcal{M} = (V, F, d, t, (\psi_a), \sigma_*)$, $\mathcal{M}' = (V', F', d', t', (\psi'_a), \sigma'_*)$ are ***isomorphic*** if there exists a bijection $\pi : V \cup F \to V' \cup F'$ such that the following conditions are satisfied.

ISM1 $\pi(x) \in V'$ for all $x \in V$ and $\pi(a) \in F'$ for all $a \in F$,
ISM2 if $r_H = (x_H, i_H)$ and $r_{H'} = (x_{H'}, i_{H'})$, then $\pi(x_H) = x_{H'}$ and $i_H = i_{H'}$,
ISM3 $d(v) = d'(\pi(v))$, $\sigma_*(v) = \sigma'_*(\pi(v))$ for all $v \in V \cup F$ and $t(v,i) = t'(\pi(v,i))$ for all $(v,i) \in C_V \cup C_F$,

ISM4 $\psi_a = \psi_{\pi(a)}$ for all $a \in F$,

ISM5 for all $(v,i) \in C_V$ we have $H_{v,i} = (a,j)$ iff $H'_{\pi(v),i} = (\pi(a), j)$.

We denote the isomorphism class of a template H by $[H]$. Let $\mathfrak{G} = \mathfrak{G}(\Delta, \Omega, \Theta, \Psi)$ be the set of all isomorphism classes. Further, let $\mathfrak{T} \subset \mathfrak{G}$ be the set of all isomorphism classes of acyclic templates. For each $[H] \in \mathfrak{G}$ and $\ell \geq 1$ let $\partial^\ell[H]$ be the isomorphism class of the template obtained by removing all vertices at a distance greater than ℓ from the root if the root is a variable clone and $\ell + 1$ if the root is a factor node. We endow $\mathfrak{G}$ with the coarsest topology that makes all the functions $\Gamma \in \mathfrak{G} \mapsto \mathbf{1}\{\partial^\ell[\Gamma] = \partial^\ell[\Gamma_0]\} \in \{0,1\}$ for $\ell \geq 1, \Gamma_0 \in \mathfrak{G}$ continuous. Moreover, the space $\mathcal{P}(\mathfrak{G})$ of probability measures on $\mathfrak{G}$ carries the weak topology. So does the space $\mathcal{P}^2(\mathfrak{G})$ of probability measures on $\mathcal{P}(\mathfrak{G})$. For $\Gamma \in \mathfrak{G}$ and $\lambda \in \mathcal{P}(\mathfrak{G})$ we write $\delta_\Gamma \in \mathcal{P}(\mathfrak{G})$ and $\delta_\lambda \in \mathcal{P}^2(\mathfrak{G})$ for the Dirac measure that puts mass one on Γ resp. λ.

For a factor graph $G \in \mathcal{G}(n)$ and a clone (v,i) we write $[G,v,i]$ for the isomorphism class of the connected component of (v,i) in G rooted at (v,i). Each $G \in \mathcal{G}(n)$ gives rise to the empirical distribution

$$\lambda_G = \frac{1}{|C_{V_n}| + |C_{F_n}|} \sum_{(v,i) \in C_{V_n} \cup C_{F_n}} \delta_{[G,v,i]} \in \mathcal{P}(\mathfrak{G}).$$

Let $\Lambda_n = \mathbb{E}[\delta_{\lambda_{\boldsymbol{G}(n)}}] \in \mathcal{P}^2(\mathfrak{G})$. We say that $(\mathcal{M}(n))_n$ ***converges locally*** to $\vartheta \in \mathcal{P}(\mathfrak{T})$ if $\lim_{n\to\infty} \Lambda_n = \delta_\vartheta$.

Additionally, to exclude some pathological cases we need the following assumption. Let us call a factor graph G ***ℓ-acyclic*** if it does not contain a cycle of length at most ℓ. We say that the sequence $(\mathcal{M}(n))_n$ of models ***has high girth*** if for any $\ell > 0$ we have

$$\liminf_{n\to\infty} \mathbb{P}[\boldsymbol{G}(n) \text{ is } \ell\text{-acyclic}] > 0.$$

The key prediction of the "replica symmetric cavity method" can be cast as follows: $(n^{-1} \ln Z_{\boldsymbol{G}(n)})_n$ converges in probability to the Bethe free energy of a "Belief Propagation fixed point" on the (possibly infinite) trees in the support of ϑ [23]. To formalise this, let $\boldsymbol{T}_\vartheta \in \mathfrak{T}$ be a sample from $\vartheta \in \mathcal{P}(\mathfrak{T})$. Further, let $\mathcal{V}$ be the event that the root of $\boldsymbol{T}_\vartheta$ is a variable clone and let $\mathcal{F}$ be the event that the root is a constraint clone. For $T \in \mathfrak{T}$ let d_T denote the degree of the root of T. Moreover, for $j \in [d_T]$ let $T \uparrow j \in \mathfrak{T}$ denote the tree pending on the jth child of the root of T. Finally, if the root is the clone of a constraint node we let ψ_T be its associated function.

▶ **Definition 12.** A measurable map $p : \mathfrak{T} \to \mathcal{P}(\Omega)$, $T \mapsto p_T$ is called a ***ϑ-Belief Propagation fixed point*** if the following conditions are satisfied ϑ-almost surely.

1. if the root of T is a variable clone (x,i), then

$$p_T(\omega) \propto \prod_{j \in [d_T] \setminus \{i\}} p_{T\uparrow j}(\omega).$$

2. if the root of T is a factor clone (a,i) with associated factor $\psi \in \Psi$, then

$$p_T(\omega_i) \propto \sum_{(\omega_j)_{j \in [d_T] \setminus \{i\}}} \psi(\omega_1, \ldots, \omega_{d_T}) \prod_{j \in [d_T] \setminus \{i\}} p_{T\uparrow j}(\omega_j).$$

Further, we need to define the Bethe free energy of a ϑ-Belief Propagation fixed point p. To this end, we turn p into a map that assigns each tree a "marginal". More precisely, we let

$$\hat{p}_T(\omega) \propto \prod_{j\in[d_T]} p_{T\uparrow j}(\omega) \qquad \text{if } T\in\mathcal{V},\ \omega\in\Omega,$$

$$\hat{p}_T(\omega_1,\ldots,\omega_{\#\psi_T}) \propto \psi_T(\omega_1,\ldots,\omega_{\#\psi_T}) \prod_{j\in[\#\psi_T]} p_{T\uparrow j}(\omega_j) \quad \text{if } T\in\mathcal{F},\ \omega_1,\ldots,\omega_{\#\psi_T}\in\Omega.$$

The ***Bethe free energy*** of p with respect to ϑ is

$$\mathcal{B}_\vartheta(p) = \left(\mathbb{E}\left[d_{\boldsymbol{T}_\vartheta}^{-1}(1-d_{\boldsymbol{T}_\vartheta})H(\hat{p}_{\boldsymbol{T}_\vartheta})|\mathcal{V}\right] - \mathbb{E}\left[d_{\boldsymbol{T}_\vartheta}^{-1}D\left(\hat{p}_{\boldsymbol{T}_\vartheta}\|\psi_{\boldsymbol{T}_\vartheta}(\sigma)\right)|\mathcal{F}\right]\right)\mathbb{E}[d_{\boldsymbol{T}_\vartheta}|\mathcal{V}]$$

Finally, to obtain a sufficient condition for the convergence $n^{-1}\ln Z_{\boldsymbol{G}(n)} \to \mathcal{B}_\vartheta(p)$ we are going to apply Theorem 3 to upper-bound the second moment of $Z_{\boldsymbol{G}(n)}$. The necessary construction, reminscent of those used in [9, 13, 11, 16, 26, 30], is as follows.

▶ **Proposition 13.** *For any $\varepsilon>0$ there exists $\eta>0$ such that the following is true. Suppose that $\mathcal{M}$ is a $(\Delta,\Omega,\Psi,\Theta)$-model of size $n=\#\mathcal{M}\geq 1/\eta$. There exists a finite set of functions $\Psi^\otimes$ and a $(\Delta,\Omega\times\Omega,\Psi^\otimes,\Theta)$-model $\mathcal{M}^\otimes$ with the following properties.*

(i) *There is a bijection $\mathcal{G}(\mathcal{M})\to\mathcal{G}(\mathcal{M}^\otimes)$, $G\mapsto G^\otimes$.*

(ii) *Let $\mathcal{U}\subset\mathcal{G}(\mathcal{M})$ be an event such that $\mathbb{P}[\boldsymbol{G}\in\mathcal{U}]>\varepsilon$. Then*

$$n^{-1}\ln\mathbb{E}[Z_{\boldsymbol{G}}^2|\mathcal{U}] \leq \max\left\{\mathcal{B}_{\mathcal{M}^\otimes}(G^\otimes) : G\in\mathcal{U}\right\}+\varepsilon.$$

Proof. Let $\Omega^\otimes=\Omega\times\Omega$ and denote $(\omega,\omega')\in\Omega^\otimes$ by $\omega\otimes\omega'$. For $\psi\in\Psi$ let

$$\psi^\otimes : (\Omega^\otimes)^{\#\psi}\to(0,\infty), \qquad (\omega_1\otimes\omega_1',\ldots,\omega_{\#\psi}\otimes\omega_{\#\psi}')\mapsto\psi(\omega_1,\ldots,\omega_{\#\psi})\cdot\psi(\omega_1',\ldots,\omega_{\#\psi}').$$

Then $\mathcal{M}^\otimes=(V,F,d,t,(\psi_a^\otimes)_{a\in F})$ satisfies the requirements. ◀

▶ **Theorem 14.** *Suppose that $(\mathcal{M}(n))_{n\geq1}$ has high girth and converges locally to $\vartheta\in\mathcal{P}(\mathfrak{T})$. Furthermore, assume that there is a ϑ-Belief Propagation fixed point p such that for any $\varepsilon>0$ we have*

$$\lim_{n\to\infty}\mathbb{P}\left[\mathcal{B}_{\mathcal{M}^\otimes(n)}(\boldsymbol{G}^\otimes(n))\leq 2\mathcal{B}_\vartheta(p)+\varepsilon\right]=1 \quad \textit{and} \tag{8}$$

$$\lim_{\ell\to\infty}\lim_{n\to\infty}\frac{1}{n}\mathbb{E}\left[\ln\frac{\mathbb{E}[Z_{\boldsymbol{G}(n)}\mathbf{1}\{\mathcal{B}_{\mathcal{M}^\otimes(n)}(\boldsymbol{G}^\otimes(n))\leq 2\mathcal{B}_\vartheta(p)+\varepsilon\}|\mathcal{T}_\ell]}{\mathbb{E}[Z_{\boldsymbol{G}(n)}|\mathcal{T}_\ell]}\right]=0. \tag{9}$$

Then $\frac{1}{n}\ln Z_{\boldsymbol{G}(n)}$ converges to $\mathcal{B}_\vartheta(p)$ in probability.

For a given ϑ the construction or, at least, identification of the ϑ-Belief Propagation fixed point p in Theorem 14 is similar to the computations done in the physics literature. However, to apply Theorem 14 it will generally be necessary to perform these calculations more thoroughly, e.g., by means of the contraction method [27]. Further, to verify condition (8) we need to study the Bethe free energy of the models $\mathcal{M}_n^\otimes$, which will typically be done by way of analysing Belief Propagation on the random factor graph $\boldsymbol{G}^\otimes(n)$. This task may be far from trivial, but at least it is a well-defined combinatorial problem.

Finally, (9) provides that given that the local structure up to depth ℓ is "typical", conditioning on the event that the second moment Bethe free energy is bounded by $2\mathcal{B}_\vartheta(p)+\varepsilon$ does not cause a substantial drop in the first moment. This is a technical condition that can be verified by studying an auxiliary probability space, namely a variant of the "planted model" with a given local structure. Technically, this task can be tackled via a generalised "configuration model" as put forward in [6].

References

1 Dimitris Achlioptas and Cristopher Moore. Random k-sat: Two moments suffice to cross a sharp threshold. *SIAM J. Comput.*, 36(3):740–762, September 2006.

2 Dimitris Achlioptas, Assaf Naor, and Yuval Peres. Rigorous location of phase transitions in hard optimization problems. *Nature*, 435(7043):759–764, 06 2005.

3 Dimitris Achlioptas, Assaf Naor, and Yuval Peres. On the maximum satisfiability of random formulas. *J. ACM*, 54(2), April 2007.

4 Dimitris Achlioptas and Yuval Peres. The threshold for random k-sat is $2k(\ln 2 - o(k))$. In *Proceedings of the Thirty-fifth Annual ACM Symposium on Theory of Computing*, STOC'03, pages 223–231, New York, NY, USA, 2003. ACM.

5 H. A. Bethe. Statistical theory of superlattices. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 150(871):552–575, 1935.

6 Charles Bordenave and Pietro Caputo. Large deviations of empirical neighborhood distribution in sparse random graphs. *Probability Theory and Related Fields*, pages 1–73, 2014.

7 Amin Coja-Oghlan and Konstantinos Panagiotou. The asymptotic k-SAT threshold. *arXiv:1310.2728*, 2014.

8 Amin Coja-Oghlan and Lenka Zdeborová. The condensation transition in random hypergraph 2-coloring. In *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2012, Kyoto, Japan, January 17-19, 2012*, pages 241–250, 2012.

9 Amir Dembo, Andrea Montanari, Allan Sly, and Nike Sun. The replica symmetric solution for potts models on d-regular graphs. *Communications in Mathematical Physics*, 327(2):551–575, 2014.

10 Amir Dembo, Andrea Montanari, and Nike Sun. Factor models on locally tree-like graphs. *Ann. Probab.*, 41(6):4162–4213, 11 2013.

11 Jian Ding, Allan Sly, and Nike Sun. Maximum independent sets on random regular graphs. *arXiv:1310.4787*, 2013.

12 Jian Ding, Allan Sly, and Nike Sun. Proof of the satisfiability conjecture for large k. *arXiv:1411.0650*, 2014.

13 Jian Ding, Allan Sly, and Nike Sun. Satisfiability threshold for random regular NAE-SAT. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, STOC'14, pages 814–822, New York, NY, USA, 2014. ACM.

14 Silvio Franz and Michele Leone. Replica bounds for optimization problems and diluted spin systems. *J. Stat. Phys.*, 111(3-4):535–564, 2003.

15 Alan Frieze and Nicholas C. Wormald. Random k-sat: A tight threshold for moderately growing k. In *Proceedings of the Fifth International Symposium on Theory and Applications of Satisfiability Testing*, pages 1–6, 2002.

16 Andreas Galanis, Daniel Štefankovič, and Eric Vigoda. Inapproximability for antiferromagnetic spin systems in the tree non-uniqueness region. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, STOC'14, pages 823–831, New York, NY, USA, 2014. ACM.

17 R.G. Gallager. *Low-density Parity-check Codes.* M.I.T. Press research monographs. M.I.T. Press, 1963.

18 H.O. Georgii. *Gibbs Measures and Phase Transitions.* De Gruyter studies in mathematics. De Gruyter, 2011.

19 Francesco Guerra. Broken replica symmetry bounds in the mean field spin glass model. *Communications in Mathematical Physics*, 233(1):1–12, 2003.

20 Svante Janson. Random regular graphs: Asymptotic distributions and contiguity. *Combinatorics, Probability and Computing*, 4:369–405, 12 1995.

21 Florent Krzakala, Andrea Montanari, Federico Ricci-Tersenghi, Guilhem Semerjian, and Lenka Zdeborová. Gibbs states and the set of solutions of random constraint satisfaction problems. *P. Natl. Acad. Sci. USA*, 104(25):10318–10323, 2007.

22 L. Lovász. *Large Networks and Graph Limits*. American Mathematical Society colloquium publications. American Mathematical Society, 2012.

23 M. Mézard and A. Montanari. *Information, Physics and Computation*. Oxford University Press, 2009.

24 M. Mézard, G. Parisi, and R. Zecchina. Analytic and algorithmic solution of random satisfiability problems. *Science*, 297:812–815, 2002.

25 Andrea Montanari and Devavrat Shah. Counting good truth assignments of random k-sat formulae. In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA'07, pages 1255–1264, Philadelphia, PA, USA, 2007.

26 Elchanan Mossel, Dror Weitz, and Nicholas Wormald. On the hardness of sampling independent sets beyond the tree threshold. *Probability Theory and Related Fields*, 143(3-4):401–439, 2009.

27 Ralph Neininger and Ludger Rüschendorf. A general limit theorem for recursive algorithms and combinatorial structures. *Ann. Appl. Probab.*, 14(1):378–418, 02 2004.

28 Dmitry Panchenko and Michel Talagrand. Bounds for diluted mean-fields spin glass models. *Probab. Theory Relat. Fields*, 130(3):319–336, 2004.

29 Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1988.

30 Allan Sly and Nike Sun. The computational hardness of counting in two-spin models on d-regular graphs. *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, 0:361–369, 2012.

31 Endre Szemerédi. Regular partitions of graphs. In *Colloq. Internat. CNRS*, volume 260, pages 399–401, 1978.

32 Jonathan S. Yedidia, W.T. Freeman, and Y. Weiss. Constructing free-energy approximations and generalized belief propagation algorithms. *Information Theory, IEEE Transactions on*, 51(7):2282–2312, July 2005.