HARNESSING THE BETHE FREE ENERGY

VICTOR BAPST∗, AMIN COJA-OGLHAN∗

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 handy sufficient condition for this upper bound to be tight.

Mathematics Subject Classification: 05C80, 82B44

1. INTRODUCTION

Many problems in combinatorics, computer science and physics can be cast along the following lines [3, 25]. There are a (large) number of variables, each of them ranging over a finite domain \( \Omega \). 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 \( \sigma \) that maps each variable to a value from \( \Omega \) 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 \( \Phi \) is its partition function, defined as

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

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

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

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 [25].

Over the past decade the second moment method has emerged as the principal tool for the analysis of such models [2, 3, 17]. Its “vanilla” version works as follows. If the partition function \( Z \) of the model satisfies the bound \( \mathbb{E}[Z^2] \leq \mathcal{O}(\mathbb{E}[Z]^2) \) in the limit as the number \( n \) of variables tends to infinity, then \( n^{-1} \ln(\mathbb{E}[Z]) \) 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 \( Z \) [24].

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 \( n^{-1} \ln(Z_{\beta, \Phi}/\mathbb{E}[Z_{\beta, \Phi}]) \) does not converge to 0 as \( n \to \infty \) for any \( \alpha, \beta > 0 \). The reason is that the first moment \( \mathbb{E}[Z_{\beta, \Phi}] \) is driven up by a “lottery effect”:

\[\text{Date: April 16, 2015.}\]

∗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.
there are a tiny minority of formulas with an abundance of “good” assignments \([2, 4, 5]\). Of course, this implies that \(\mathbb{E}[Z^2_{\beta, \Phi}] \geq \exp(\Omega(n))\mathbb{E}[Z_{\beta, \Phi}]^2\). Thus, the second moment method fails rather spectacularly.

The obvious remedy 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. \([8, 15]\).

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 \([24]\). 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 local maxima of the Bethe free energy are in one-to-one correspondence with the Belief Propagation fixed points \([34]\). 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.

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 3.1). The proof of this is based 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 4.4). Second, we propose a corresponding refined “second moment method” (Theorem 5.5). 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 \([6, 19, 31]\). 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., \([25]\) 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 \((5.2)\) below. The case of uniform models is conceptually simpler and has been treated before \([12]\). 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 \([12, 13, 14]\). This fact has played a key role in recent work on the hardness of counting problems \([18, 28, 32]\). Although we use a similar tensor construction in our second moment argument as well (cf. Proposition 5.4), non-uniformity makes matters far more intricate, as witnessed by recent work on random \(k\)-SAT \([8, 15]\). 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 \([11, 27]\). However, the necessary spatial mixing properties are quite strong and they cease to be satisfied, e.g., in the random \(k\)-SAT model from \([1, 1]\) for large \(\beta\) for clause/variable ratios as low as \(\ln k/k\) \([27]\). By comparison, the \(k\)-SAT threshold is about \(2^k \ln 2\) \([15]\).

The “interpolation method” provides a different approach to calculating or at least upper-bounding \(n^{-1} \ln Z\) \([16, 21]\). In particular, the upper bound comes in a variational form \([30]\). For example, this can be used to obtain a tight upper bound on the \(k\)-SAT threshold \([15]\). 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 5.4 and 5.5 do not require such assumptions.

---

\(^1\)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.
The physicists’ cavity method comes in several instalments; for a detailed discussion we refer to [25]. In this paper we are chiefly concerned with the simplest, “replica symmetric” variant. This version does not always provide the correct value of $n^{-1} \ln Z$ [9]. It seems that one reason for this is that models such as random $k$-SAT undergo a so-called “condensation phase transition” [23]. The more complex “1-step replica symmetry breaking (1RSB)” version of the cavity method [26] 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 [25] 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, $\alpha$ is a positive real, $\Delta > 0$ is an integer, $\Omega$, $\Theta$ are finite sets and $\Psi = \{\psi_1, \ldots, \psi_t\}$ is a finite set of maps $\psi_i : \Omega^{h_i} \to (0, \infty)$, where $1 \leq h_i \leq \Delta$. The following abstract definition encompasses a multitude of concrete examples.

**Definition 3.1.** A $(\alpha, \Delta, \Omega, \Psi, \Theta)$-model $M = (V, F, d, t, \psi, \sigma_*)$ 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
  $$ C_V = \bigcup_{x \in V} \{x\} \times [d(x)], \quad C_F = \bigcup_{a \in F} \{a\} \times [d(a)], $$
  such that $|t^{-1}(\theta) \cap C_V| = |t^{-1}(\theta) \cap C_F|$ 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 $a \in F$,
- **M6**: a map $\sigma_* : V \to \Omega$.

The size of the model is defined as $\#M = |V|$. Furthermore, a $M$-*factor graph* is a bijection

$$ G : C_V \to C_F, \quad (x, i) \mapsto G_{x,i} $$

such that $t(G_{x,i}) = t(x, i)$ for all $(x, i) \in C_V$.

Of course, the equations **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 $M$-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 $\Theta$. Moreover, each constraint node $a$ comes with a “weight function” $\psi_a$ from the set $\Psi$. Finally, $\sigma_*$ is a distinguished “reference assignment”.

Like in the “configuration model” a $M$-factor graph is a type-preserving matching $G$ of the variable and constraint clones. Let $G(M)$ be the set of all $M$-factor graphs and let $G(M)$ denote a uniformly random sample from $G(M)$. We usually think of $G \in G(M)$ 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 $G_{a,j}$.

In the case where $|V|$ is finite, a $M$-*assignment* is a map $\sigma : V \to \Omega$ such that $|\{x \in V : \sigma(x) \neq \sigma_*(x)\}| \leq \alpha|V|$. Let $C_M$ be the set of all $M$-assignments. Further, define the *partition function* of $G \in G(M)$ as

$$ Z_G = \sum_{\sigma \in C_M} \prod_{a \in F} \psi_a(\sigma(\partial_G(a, 1)), \ldots, \sigma(\partial_G(a, d(a)))) \tag{3.1} $$

It is closely intertwined with the *Gibbs distribution* of $G$, which is the distribution on $C_M$ defined by

$$ \mu_G(\sigma) = Z_G^{-1} \prod_{a \in F} \psi_a(\sigma(\partial_G(a, 1)), \ldots, \sigma(\partial_G(a, d(a)))) \tag{3.2} $$

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

Example 3.2 (the random $k$-SAT model). Let $\Omega = \{0, 1\}$. Given some $\beta \geq 0$ let $\Psi$ contain the $2^k$ weight functions
\[\psi(\tau) : \Omega^k \to (0, \infty), \quad \sigma \mapsto \exp(-\beta 1(\sigma = \tau)) \quad \text{for } \tau \in \Omega^k.\]

Let $\Delta > 0$ be a positive integer, let $\alpha = 1$ and let $\Theta = \{\ast\}$. We obtain a $(\alpha, \Delta, \Omega, \Psi, \Theta)$-model $M_{\text{SAT}}$ 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, choose any reference assignment $\sigma_r$ and let $t : C_V \cup C_F \to \Theta$ be the trivial (constant) map. Then $G(M_{\text{SAT}})$ corresponds to choosing a random $k$-SAT formula with the given degree sequence and sign patterns. Moreover, $n^{-1} \ln Z_{G(M_{\text{SAT}})}$ accounts for weighted truth assignments (cf. (7.1) [20]).

In Example 3.2 we did not actually use the types or the reference assignment $\sigma_r$ in a non-trivial way. The types could be used to prescribe not merely the degree of each variable but also how many times each Boolean variable appears positively or negatively. Additionally, the reference assignment and the “radius” $\alpha$ can be employed to study, e.g., the “planted $k$-SAT model”. This model plays an important role in understanding the geometry of the space of solutions and the performance of local search algorithms [1].

While Definition 3.1 encompasses a many problems of interest, there are two restrictions. They arise because we are going to be interested in sequences $(M_n)_n$ of $(\alpha, \Delta, \Omega, \Psi, \Theta)$-models of sizes $\# M_n = n$. That is, the size of the model tends to infinity while $\alpha, \Delta, \Omega, \Psi, \Theta$ 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 $\Delta$.

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 instance, the positivity assumption ensures that $Z_{G} > 0$ for all $G \in G(M_n)$ and hence that the random variable $n^{-1} \ln Z_{G(M_n)}$ has a finite mean. Furthermore, the case of hard constraints can be handled by introducing an “inverse temperature” parameter $\beta > 0$ like in Example 3.2 and ultimately taking the limit $\beta \to \infty$ (cf. [27]).

In Section 4 we will prove that the “Bethe free energy” provides an upper bound on $|V|^{-1} \ln Z_{G(M)}$. 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 $\alpha > 0$ be a real, $\Delta \geq 1$ an integer, $\Omega$, $\Theta$ finite sets, and $\Psi$ a finite set of functions as in Definition 3.1. 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 $\mu$. Here and throughout we use the convention that $0 \ln 0 = 0$. 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 $\mu_{i,j}$ be the marginal of the $j$th component, i.e., $\mu_{i,j}(x_j) = \sum_{(x_1, \ldots, x_j) \in \mathcal{X}^{k-1}} \mu(x_1, \ldots, x_k)$.

Fact 3.3. The map $\mathcal{P}(\mathcal{X}^k) \to \mathbb{R}_{\geq 0}$, $p \mapsto D \left( p \| \otimes_{i \in [k]} p_{i,i} \right)$ is uniformly continuous.

Fact 3.4. Let $\mathcal{X} \neq \emptyset$ be a finite set and let $k \geq 2$. For any $\delta > 0$ there is $\varepsilon = \varepsilon(\delta, \mathcal{X}, k) > 0$ such that the following is true. Assume that $q_1, \ldots, q_k \in \mathcal{P}(\mathcal{X})$ satisfy $q_j(\omega) \geq \delta$ for all $j \in [k], \omega \in \mathcal{X}$. Moreover, suppose that $p \in \mathcal{P}(\mathcal{X}^k)$ is such that $\|p_{i,j} - q_j\|_{\text{TV}} < \varepsilon$ for all $j \in [k]$. Then there is $p' \in \mathcal{P}(\mathcal{X}^k)$ with $p'_{i,j} = q_j$ for all $j$ such that $\|p - p'\|_{\text{TV}} < \delta$.

For $\mu \in \mathcal{P}(\mathcal{X})$ we write $\sigma_{\mu}$ for a random element of $\mathcal{X}$ chosen according to $\mu$. Where $\mu$ is apparent from the context we drop the index. Further, if $X : \mathcal{X} \to \mathbb{R}$ is a random variable we write $\langle X \rangle_\mu = \sum_{x \in \mathcal{X}} X(x) \mu(x)$ for the expectation of $X$ with respect to $\mu$. We use this notation because we are going to deal with two levels of randomness. First, the choice of a random factor graph $G(M)$, for which we use the common $\mathbb{E}[\cdot]$ notation. Second, the Gibbs distribution $\mu_{G(M)}$ given $G(M)$, for which we use $\langle \cdot \rangle_G$ notation. The sake of brevity we normally write $\langle \cdot \rangle_G$ instead of $\langle \cdot \rangle_{G(M)}$ for $G \in G(M)$.

Further, if $S$ is a subset of the set $V$ of variable nodes of $M$, $\sigma : V \to \Omega$ and $\omega \in \Omega$ we write
\[\sigma|_S(\omega) = \frac{1}{|S|} \sum_{x \in S} 1 \{\sigma(x) = \omega\}.\]
Thus, $\sigma \cdot |S| \in \mathcal{P}(\Omega)$ is the empirical distribution of $\sigma$ on $S$. If $S = \{x\}$ for some $x \in V$ we let $\sigma[\omega|x] = \sigma[\omega|\{x\}]$. Analogously, if $G \in G(\mathcal{M})$ 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

$$\sigma[\omega_1, \ldots, \omega_l|A] = \frac{1}{|A|} \prod_{a \in A} \prod_{j=1}^l 1 \{\sigma(\partial G(a, j)) = \omega_j\}.$$ 

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

Further, we are going to need the following consequence of [10, Theorem 2.4.1].

**Fact 3.5.** For any $\varepsilon > 0$ there is $n_0 > 0$ such that for all $n > n_0$ the following is true. Let $S_1, \ldots, S_k$ be sets of the size $|S_i| = n$ and let $\sigma_i : S_i \rightarrow \Omega$ be maps. Let $p_i = \sigma_i \cdot |S_i| \in \mathcal{P}(\Omega)$. Moreover, let $p \in \mathcal{P}(\Omega^k)$ be such that $p_{i,j} = p_i$ for all $i \in [k]$. Finally, let $(\pi_j)_{j \in [k]}$ be a family of independent uniformly random permutations $[n] \rightarrow [n]$ and let

$$q(\omega_1, \ldots, \omega_k) = \frac{1}{n} \sum_{i=1}^n \prod_{j=1}^k 1 \{\pi_j(i) = \omega_j\}.$$ 

Then $|D(p||p_1 \otimes \cdots \otimes p_k) + \frac{1}{n} \ln \mathbb{P}[q = p]| < \varepsilon$.

Finally, we use the $\propto$ symbol to define probability distributions on a finite set $\mathcal{X}$. If $f : \mathcal{X} \rightarrow [0, \infty)$, 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.

4. **The upper bound**

In this section we let $\mathcal{M} = (V, F, d, t, (\psi_a)_{a \in F}, \sigma_*)$ be a $(\alpha, \Delta, \Omega, \Psi, \Theta)$-model of finite size $n = |V|$. We write $G = G(\mathcal{M})$ for the sake of 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. Let $G \in G(\mathcal{M})$. 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

- **MS1:** if $G_{x,i} = (a, j)$, then $\nu_x = \nu_{a,j}$,

- **MS2:** $\sum_{x \in V} \nu_x(\sigma_*(x)) \geq (1 - \alpha)n$.

Thus, if a variable $x$ occurs in the $j$th position of a constraint $a$, then the $j$th marginal of $\nu_a$ coincides with $\nu_x$. Further, if we draw $\sigma(x)$ from $\nu_x$ for each $x \in V$, then the expected “overlap” of $\sigma$ and $\sigma_*$ is at least $(1 - \alpha)n$. Condition MS1 is identical to the “consistent marginal sequence” of [25, Section 14.4.1]. Clearly, any probability distribution $\nu^*$ on $C_\mathcal{M}$ gives rise to a marginal sequence. Simply let $\nu^*_{x}$ be the $x$-marginal for each $x \in V$ and let $\nu^*_a$ be the joint distribution of the variable nodes adjacent to $a \in F$.

The **Bethe free energy** is

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

(for a detailed derivation of the Bethe free energy in the context of the cavity method see [25, Chapter 14]). Additionally, we define the **Bethe free energy of $G$** as $\mathcal{B}_\mathcal{M}(G) = \max \{\mathcal{B}_\mathcal{M}(G, \nu) : \nu$ is a marginal sequence of $G\}$.

**Theorem 4.1.** For any $\alpha, \Delta, \Omega, \Psi, \Theta$ and any $\varepsilon > 0$ there exists $n_0 > 0$ such that the following is true. Suppose that $\mathcal{M}$ is a finite $(\alpha, \Delta, \Omega, \Psi, \Theta)$-model of size $n > n_0$. Moreover, let $\emptyset \neq \mathcal{U} \subset G(\mathcal{M})$ be an event. Then

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

Thus, there exists a number $n_0$ that depends only on the basic parameters $\alpha, \Delta, \Omega, \Psi, \Theta$ and the desired accuracy $\varepsilon$ such that for any model of size $n \geq n_0$ the Bethe free upper bounds on the expectation of $Z_G$ on $\mathcal{U}$. The following corollary provides a handy way to apply Theorem [4.1]

**Corollary 4.2.** Let $(\mathcal{M}_n)_n$ be a sequence of $(\alpha, \Delta, \Omega, \Psi, \Theta)$-models such that $\# \mathcal{M}_n = n$. Assume that $b > 0$ is such that the event $\mathcal{U}_n = \{\mathcal{B}_\mathcal{M}_n(G(\mathcal{M}_n)) \leq b\}$ satisfies $\lim_{n \rightarrow \infty} \mathbb{P}[\mathcal{U}_n] = 1$. Then $\lim \sup_{n \rightarrow \infty} n^{-1} \ln \mathbb{E}[Z_G(\mathcal{M}_n) | \mathcal{U}_n] \leq b$.
By Markov’s and Jensen’s inequality, the bound \(\lim_{n \to \infty} \sup_{n} n^{-1} \ln E[Z_{G(M_n)}] \leq b\) entails that
\[
\lim_{\varepsilon \to 0} \lim_{n \to \infty} P\left[n^{-1} \ln Z_{G(M_n)} \leq b + \varepsilon\right] = 1.
\]
In other words, if the Bethe free energy is bounded by \(b\), “with high probability”, then \(n^{-1} \ln Z_{G(M_n)} \leq b + o(1)\) “with high probability”.

The proof of Theorem 4.1, which the rest of this section deals with, 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 4.1 by showing Theorem 4.4 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

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 4.3.** A \(M\)-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 \(\bar{s} \subseteq \Psi[N]\),

**ST4:** a sequence \((\tilde{s}_h)_{h \in [N]}\) of probability distributions on \(\Omega\) such that \(\sum_{x \in V} \tilde{s}_{s(x)}(\sigma_{s}(x)) \geq (1 - \alpha)n\),

**ST5:** for any \((\psi, h) \in \Psi[N]\) a probability distribution \(\bar{s}_{\psi,h} \) on \(\Omega^{\#\psi}\) such that \(\bar{s}_{\psi,h,j} = \bar{s}_{h_j}\) for all \(j \in [\#\psi]\).

Normally we denote an \(M\)-state simply by \(s\) and we write \(\#s\) for its size. Moreover, let
\[
V^s_h = s^{-1}(h) \quad \text{for } h \in [\#s],
\]
\[
V^s_{\psi,h} = \prod_{j \in [\#\psi]} V^s_{h_j} \quad \text{for } (\psi, h) \in \Psi[N].
\]

In addition, if \(G \in G(M)\) and \((\psi, h) \in \Psi[N]\) we let
\[
\partial_{G,s}(\psi, h) = \{ a \in F : \psi_a = \psi, \partial_G(a) \in V^s_{\psi,h} \}.
\]

Thus, a state induces a partition \(V^s_{h_1}, \ldots, V^s_{h_{\#\psi}}\) of the set of variable nodes. Condition **ST1** ensures that this partition respects the degrees and the types. Let us call \(G \in G(M)\) \(\varepsilon\)-compatible with \(s\) for some \(\varepsilon > 0\) (“\(G \models_{\varepsilon} s\)”) if
\[
\sum_{(\psi, h) \in \Psi[N]} \frac{\|\partial_{G,s}(\psi, h)\|}{|F|} \leq \bar{s}_{\psi,h} < \varepsilon,
\]
\[
\sum_{(\psi, h) \in \Psi[N]} 1\{ (\psi, h) \in \bar{s} \} \bar{s}_{\psi,h} < \varepsilon.
\]

Thus, for any \((\psi, h)\) 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 \(\varepsilon\) fraction of all constraint nodes belong to the “rogue classes” \((\psi, h) \in \bar{s}\).

Further, suppose that \(G \models_{\varepsilon} s\) and \(\sigma \in C_M\). We say that \((G, \sigma)\) is \(\varepsilon\)-judicious with respect to \(s\) (“\((G, \sigma) \models_{\varepsilon} s\)”) if
\[
\text{J1: for all } h \in [N] \text{ we have } \|\bar{s}_{h} - \sigma[|V^s_{h}]\|_{TV} < \varepsilon,
\]
\[
\text{J2: for all } (\psi, h) \in \Psi[N] \backslash \bar{s} \text{ such that } \partial_{G,s}(\psi, h) = \emptyset \text{ we have } \|\bar{s}_{\psi,h} - \sigma[|\partial_{G,s}(\psi, h)]\|_{TV} < \varepsilon.
\]

In words, the empirical distributions \(\sigma[|V^s_{h}]\) do not deviate by more than \(\varepsilon\) from \(\bar{s}_h\). Similarly, for a “non-rogue” \((\psi, h)\) the empirical distribution \(\sigma[|\partial_{G,s}(\psi, h)]\) of the \(\psi\)-factors that connect variables in \(V^s_{\psi,h}\) is within \(\varepsilon\) of \(\bar{s}_{\psi,h}\).

The following theorem provides the key fact about states. It should be of interest in its own right.

**Theorem 4.4.** For any \(\alpha, \Delta, \Omega, \Psi, \Theta\) and any \(\varepsilon > 0\) there exists \(\eta > 0\) such that the following is true. Let \(M\) be a finite \((\alpha, \Delta, \Omega, \Psi, \Theta)\)-model of size \#\#\# \(\geq 1/\eta\) and let \(G \in G(M)\). Then there exists a \(M\)-state \(s\) of size \#\#\# \(\leq 1/\eta\) such that \(G \models_{\varepsilon} s\) and \(\{ (G, \sigma) \models_{\varepsilon} s\} \cap \{ s \} \geq \eta\).

Crucially, the number \(\eta\) promised by Theorem 4.4 depends on \(\varepsilon\) and the basic parameters \(\alpha, \Delta, \Omega, \Psi, \Theta\) only. It is independent of the model and its size. Hence, for any large enough \(M\) and any \(G \in G(M)\) there is a single “dominant state” \(s\) that captures a constant fraction of the mass of the Gibbs distribution \(\mu_G\).

Theorem 4.4 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.,

$$\left\| \mu_G - \sum_{i=1}^{K} w_i \mu_{G,i} \right\|_{TV} < \varepsilon$$

where $(w_1, \ldots, w_K) \in \mathcal{P}([K])$, $\mu_{G,1}, \ldots, \mu_{G,K} \in \mathcal{P}(\mathcal{C}_M)$.

Each $\mu_{G,i}$ corresponds to a “cluster” of assignments and is such that the joint distribution of far apart variable factors is.

In Section 4.4 we will use that result to derive Theorem 4.4. Further, in Section 4.5 we will show how Theorem 4.1 follows from Theorem 4.4. Finally, in Section 4.7 we discuss how the problem of calculating the Bethe free energy is

Corollary 4.5. For any $\varepsilon > 0$ and any $\alpha$, $\Delta$, $\Omega$, $\Psi$, $\Theta$ there exist $\gamma > 0$, $n_0 > 0$ such that the following is true. Suppose that $\mathcal{M}$ is a finite $(\alpha, \Delta, \Omega, \Psi, \Theta)$-model of size $\# \mathcal{M} \geq n_0$ and that $\emptyset \neq U \subseteq G(\mathcal{M})$. Then there exist a $\mathcal{M}$-state $s$ and $G_0 \in \mathcal{U}$ such that $G_0 \mid= \gamma$ and $n^{-1} \ln \mathbb{E}[Z_G \mathcal{1}\{G \in \mathcal{U}\}] \leq \varepsilon + n^{-1} \ln \mathbb{E}[Z_G \mathcal{1}\{G \mid= \gamma\}]$.

Finally, it is not difficult to derive Theorem 4.4. Indeed, $n^{-1} \ln \mathbb{E}[Z_G \mathcal{1}\{G \mid= \gamma\} \mathcal{1}\{G \mid= \gamma\}]$ can be cast as the Bethe free energy of the marginal sequence induced by $s$: let $\nu_x = \bar{s}_{s(x)}$ for $x \in V$ and $\nu_a = \bar{s}_{s \psi, h}$ for all $a \in \partial_{G_0, s}(\psi, h)$.

We proceed to prove Theorem 4.4. Section 4.3 contains the key ingredient, namely a general theorem about approximating probability measures on a cube $\Omega^n$ by measures that resemble product distributions (Theorem 4.7 below). In Section 4.4 we will use that result to derive Theorem 4.4. Further, in Section 4.5 we will show how Theorem 4.1 follows from Theorem 4.4. Finally, in Section 4.7 we discuss how the problem of calculating the Bethe free energy is related to the Belief Propagation message passing algorithm.

**Homogeneity.** The aim in this section is to establish a general statement about the structure of probability measures on a cube $\Omega^n$ for some finite set $\Omega$ and a large dimension $n$. To get started, we consider the following example of a probability measure $\mu$ on $\{0, 1\}^n$ for a large even number $n$: let

$$\mu_0(\sigma) = 2^{-\beta} \prod_{i=n/2}^{n} \left( \frac{1}{3} \right)^{\sigma(i)} \left( \frac{2}{3} \right)^{1-\sigma(i)}, \mu_1(\sigma) = 2^{-\beta} \prod_{i=n/2}^{n} \left( \frac{2}{3} \right)^{\sigma(i)} \left( \frac{1}{3} \right)^{1-\sigma(i)}, \mu(\sigma) = \frac{1}{2}(\mu_0(\sigma) + \mu_1(\sigma)).$$

Thus, $\mu_0$ and $\mu_1$ are product distributions. In both cases the first $n/2$ bits are chosen independently from the Bernoulli distribution $\text{Be}(1/2)$. But the last $n/2$ bits are chosen independently from $\text{Be}(1/3)$ and $\text{Be}(2/3)$, respectively. Furthermore, $\mu$ is the mixture of $\mu_0, \mu_1$. By construction, $(\sigma(i))_{\mu} = \frac{1}{3}$ for all $i \in [n]$, i.e., the $\mu$-marginal of every coordinate is equal to $1/2$. But of course $\mu$ is not a product distribution.

Clearly, if we decompose the set $[n]$ into the two subsets $V_1 = \{1, \ldots, n/2\}$ and $V_2 = \{n/2 + 1, \ldots, n\}$, then the projection of $\mu$ onto the first subset $V_1$ is a product distribution. In particular, it is very likely that the vector $(\sigma(1), \ldots, \sigma(n/2))$ has “low discrepancy”. For instance, most likely the average $\sigma[\cdot | V_1]$ of the entries is close to $1/2$. Furthermore, if we fix a number $\varepsilon > 0$ and assume that $n > n_0(\varepsilon)$ is sufficiently large, then with probability at least $1 - \varepsilon$ for any “extensive subsets” $S \subseteq V_1$, $|S| \geq \varepsilon n/2$, the average $\sigma[\cdot | |S|$ should be close to $1/2$ (by the Chernoff bound). Hence,

$$\mathbb{E}[\sigma[1|S] - \sigma[1|V_1]]_{\mu} < \varepsilon. \quad (4.1)$$

Moreover, the projection onto the second subset $V_2$, while not a product distribution itself, resembles a product distribution once we condition on the average $\sigma[1|V_2]$. Indeed, if $n$ is large and $\sigma[1|V_2] < 1/2$, then most likely $\sigma$ is a sample from $\mu_0$, and vice versa. Thus, the distribution of $\sigma[i]_{i \geq n/2}$ should be close to a product of $\text{Be}(1/3)$ variables if $\sigma[1|V_2] < 1/2$, resp. a product of $\text{Be}(2/3)$ variables if $\sigma[1|V_2] > 1/2$. Indeed, the Chernoff bound implies easily that for large enough $n$ the analogue of (4.1) holds: if $S \subseteq V_2$ has size $|S| \geq \varepsilon n$, then $\mathbb{E}[\sigma[1|S] - \sigma[1|V_2]]_{\mu} < \varepsilon$. Guided by this example, we introduce the following concept.
Definition 4.6. Let $\Omega$ be a finite set, let $\varepsilon > 0$, let $n$ be an integer and let $\mu$ be a probability measure on $\Omega^n$. A partition $V = (V_1, \ldots, V_N)$ of $[n]$ is called $\varepsilon$-homogeneous with respect to $\mu$ 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 $\langle \|\sigma[\cdot; S] - \sigma[\cdot; V_j]\|_{TV} \rangle_{\mu} < \varepsilon$.

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

Theorem 4.7. Let $\Omega$ be a finite set. For any $\varepsilon > 0$ there exists $N = N(\varepsilon, \Omega)$ such that for $n > N$ and any probability measure $\mu$ on $\Omega^n$ the following is true. Let $V_0$ be a partition of $[n]$ such that $\#V_0 \leq 1/\varepsilon$. Then $V_0$ has a refinement $V$ of size $\#V \leq N$ that is $\varepsilon$-homogeneous with respect to $\mu$.

Theorem 4.7 and its proof are inspired by the proof of Szemerédi’s regularity lemma [33]. The key feature of Theorem 4.7 is that the number $N$ depends only on $\varepsilon$, $\Omega$, but not on $n$ or $\mu$. To prove Theorem 4.7 fix $\Omega \neq \emptyset$ and $0 < \varepsilon < 0.01/|\Omega|$, assume that $n \geq N(\varepsilon, \Omega)$ is large enough and consider a measure $\mu$ on $\Omega^n$. Let us define the $\mu$-index of a partition $V = (V_1, \ldots, V_k)$ as

$$I_\mu(V) = \frac{1}{|\Omega|^n} \sum_{j \in [k]} \sum_{x \in V_j} \sum_{\omega \in \Omega} (\|\sigma[\cdot; x] - \sigma[\cdot; V_j]\|_{TV})_{\mu}.$$ 

Lemma 4.8. For any partition $V$ we have $I_\mu(V) \leq 1$.

Proof. Since $(\|\sigma[\cdot; x] - \sigma[\cdot; V_j]\|_{TV})_{\mu} \leq 1$, this is immediate from the definition. \qed

Lemma 4.9. If $V$ fails to be $\varepsilon$-homogeneous with respect to $\mu$, then there is a refinement $W$ of $V$ such that $\#W \leq 2\#V$ and $I_\mu(W) \leq I_\mu(V) - \varepsilon^4/(4|\Omega|^2)$.

Proof. Let $J$ be the set of all indices $j \in \#V$ such that there exists $S \subset V_j$ of size $|S| \geq \varepsilon |V_j|$ such that $\langle \|\sigma[\cdot; S] - \sigma[\cdot; V_j]\|_{TV} \rangle_{\mu} \geq \varepsilon$.

(4.2)

Since $V$ fails to be $\varepsilon$-homogeneous we have

$$\sum_{j \in J} |V_j| \geq \varepsilon n.$$

(4.3)

Let $j \in J$ and let $S_j \subset V_j$, $|S_j| \geq \varepsilon |V_j|$ be a set such that $\langle \|\sigma[\cdot; x] - \sigma[\cdot; S_j]\|_{TV} \rangle_{\mu} \geq \varepsilon$. Then there exists $\omega_j \in \Omega$ such that

$$\langle \|\sigma[\omega_j; S_j] - \sigma[\omega_j; V_j]\|_{TV} \rangle_{\mu} \geq \varepsilon/|\Omega|.$$ (4.4)

We may assume that $|S_j| \leq |V_j|/2$; for otherwise there exists a subset $S_j' \subset S_j$ of size $|V_j| \leq |S_j'| \leq |S_j|/2$ such that

$$\langle \|\sigma[\omega_j; S_j'] - \sigma[\omega_j; V_j]\|_{TV} \rangle_{\mu} \geq \langle \|\sigma[\omega_j; S_j] - \sigma[\omega_j; V_j]\|_{TV} \rangle_{\mu}$$

and we simply replace $S_j$ by $S_j'$. Hence, we have

$$\varepsilon |V_j| \leq |S_j| \leq |V_j|/2.$$ (4.5)

Let $\bar{S}_j = V_j \setminus S_j$ and

$$\Delta_j(\omega) = \sum_{x \in V_j} \langle \|\sigma[\omega; x] - \sigma[\omega; V_j]\|_{TV} \rangle_{\mu} - \sum_{x \in S_j} \langle \|\sigma[\omega; x] - \sigma[\omega; S_j]\|_{TV} \rangle_{\mu} - \sum_{x \in \bar{S}_j} \langle \|\sigma[\omega; x] - \sigma[\omega; \bar{S}_j]\|_{TV} \rangle_{\mu}.$$ 

Since

$$\sum_{x \in V_j} \langle \|\sigma[\omega; x] - \sigma[\omega; V_j]\|_{TV} \rangle_{\mu} = \sum_{x \in V_j} \left[ \langle \|\sigma[\omega; x]\|_{TV} \rangle_{\mu} - \langle \|\sigma[\omega; V_j]\|_{TV} \rangle_{\mu} \right],$$

$$\sum_{x \in S_j} \langle \|\sigma[\omega; x] - \sigma[\omega; S_j]\|_{TV} \rangle_{\mu} + \sum_{x \in \bar{S}_j} \langle \|\sigma[\omega; x] - \sigma[\omega; \bar{S}_j]\|_{TV} \rangle_{\mu} = -|S_j| \langle \sigma[\omega; S_j]^2 \rangle_{\mu} - |\bar{S}_j| \langle \sigma[\omega; \bar{S}_j]^2 \rangle_{\mu} + \sum_{x \in V_j} \langle \sigma[\omega; x]^2 \rangle_{\mu},$$
Furthermore, combining (4.4), (4.5) and (4.6), we obtain

\[ \Delta_j(\omega) = |V_j| \left( \frac{\langle \sigma[\omega|V_j]\rangle^2}{|V_j|} - \frac{|S_j|}{|V_j|} \langle \sigma[\omega|S_j]\rangle^2 - \frac{|\bar{S}_j|}{|V_j|} \langle \sigma(\omega|\bar{S}_j)\rangle^2 \right). \]  

(4.6)

Hence, by Jensen’s inequality

\[ \Delta_j(\omega) \leq 0 \quad \text{for all } \omega \in \Omega. \]  

(4.7)

Furthermore, combining (4.4), (4.5) and (4.6), we obtain

\[ \Delta_j(\omega) \leq -\frac{\varepsilon^3}{4|\Omega|^2}. \]  

(4.8)

Finally, let \( W \) be the refinement of \( V \) comprising of the classes \( S_j, \bar{S}_j \) for \( j \in J \) and \( V_j \) for \( j \in \#V \setminus J \). Clearly, \( \#W \leq \#V \) and (4.3), (4.7) and (4.8) imply that

\[ I_\mu(W) - I_\mu(V) = \frac{1}{n} \sum_{j \in J, \omega \in \Omega} \Delta_j \leq \frac{1}{n} \sum_{j \in J} \Delta_j(\omega_j) \leq -\frac{\varepsilon^3}{4|\Omega|^2} \sum_{j \in J} |V_j| \leq -\frac{\varepsilon^4}{4|\Omega|^2}, \]

as desired.

\[ \Box \]

**Proof of Theorem 4.7.** Starting from \( V_0 \), we construct a sequence \( V_1, V_2, \ldots \) of refined partitions as follows. If \( V_t \) is \( \varepsilon \)-homogeneous with respect to \( \mu \) for \( t \geq 1 \), then we let \( T = t \) and stop. Otherwise, we apply Lemma 4.9 to obtain a refinement \( V_{t+1} \) of \( V_t \) such that \( I_\mu(V_{t+1}) \leq I_\mu(V_t) - \varepsilon^4/(4|\Omega|^2) \) and \( \#V_{t+1} \leq \#V_t \). The assertion follows because Lemma 4.8 ensures that \( T \leq 4\varepsilon^{-4}|\Omega|^2 \).

\[ \Box \]

**4.4. Proof of Theorem 4.4.** We keep the notation and the assumptions of Theorem 4.4. We may assume that \( \varepsilon < \varepsilon_0 \) for a small enough number \( \varepsilon_0 = \varepsilon_0(\alpha, \Delta, \Omega, \Psi, \Theta) \). Let \( \hat{\varepsilon} = \hat{\varepsilon}(\alpha, \Delta, \Omega, \Psi, \Theta, \varepsilon) > 0 \) a small enough and let us assume that \( \#M > n_0 \) for some large \( n_0 = n_0(\alpha, \Delta, \Omega, \Psi, \Theta, \varepsilon, \hat{\varepsilon}) \). Let \( W \) be the coarsest partition of the set \( V \) of variable nodes such that any two variable nodes \( v, w \) that belong to the same class satisfy \( d(v) = d(w) \) and \( t(v, i) = t(w, i) \) for all \( i \in [d(v)] \). By Theorem 4.7 there exists \( N = N(\alpha, \Delta, \Omega, \Psi, \Theta, \varepsilon) \) such that \( W \) has a refinement \( V \) of size \( \#V \leq N \) such that \( V \) is \( \hat{\varepsilon} \)-homogeneous with respect to \( \mu_G \). Define \( s : \hat{V} \to [N] \) by letting \( s(v) = i \) for all \( v \in V_i \). Further, letting \( m = |F| \), we set

\[ s_{\psi, h} = \frac{|\partial_{G,s}(\psi, h)|}{m} \quad \text{for all } (\psi, h) \in \Psi \cdot [N]. \]

To define \( \tilde{s}, \tilde{s}_{\psi, h} \) and \( \tilde{s}_{\psi, h} \) as required by Definition 4.3, it is convenient to use the notation

\[ \partial_{G,s}(\psi, h, j) = \bigcup \{ \partial_G(a, j) : a \in \partial_{G,s}(\psi, h) \} \quad \text{for } (\psi, h) \in \Psi \cdot [N], j \in [\#\psi]. \]

In words, this is the set of all variable nodes that appear in the \( j \)th position of a \( \psi \)-factor \( a \) that binds variables from the classes \( V_{h_1}, \ldots, V_{h_{\#\psi}} \). In particular, \( \partial_{G,s}(\psi, h, j) \subset V_{h_j} \).

**Lemma 4.10.** We have

\[ \sum_{(\psi, h) \in \Psi \cdot [N]} \sum_{j = 1}^{\#\psi} |\tilde{s}_{\psi, h} \partial_{G,s}(\psi, h, j)| \left\| \sigma[s \cdot |V^*_h]|_G \right\| \leq \hat{\varepsilon}^{31}. \]

**Proof.** As \( V \) is \( \hat{\varepsilon} \)-homogeneous there is \( I \subset \#V \) such that

\[ \sum_{i \in I} |V^*_i| < \hat{\varepsilon}^{32} n \]  

(4.9)

and such that for any \( i \in I \) we have

\[ \forall S \subset V_i, |S| \geq \hat{\varepsilon}^{32} |V_i| : \left\| \sigma[s \cdot |S|] - \sigma[s \cdot |V^*_i|] \right\|_G < \hat{\varepsilon}^{32}. \]  

(4.10)

For \( i \in I \) and \( j \in [\Delta] \) let \( H_{ij} \) be the set of all pairs \( (\psi, h) \in \Psi \cdot [N] \) such that \( h_j = i \). Moreover, for \( \omega \in \Omega \) consider

\[ S_{ij}(\omega) = \{ (\psi, h) \in H_{ij} : \langle \sigma[\omega|\partial_{G,s}(\psi, h, j)] - \sigma[\omega|V^*_i] \rangle_G > \hat{\varepsilon}^{32} \}, \]

\[ T_{ij}(\omega) = \{ (\psi, h) \in H_{ij} : \langle \sigma[\omega|\partial_{G,s}(\psi, h, j)] - \sigma[\omega|V^*_i] \rangle_G < -\hat{\varepsilon}^{32} \}. \]

Then by (4.10) and because the degrees of the variable nodes are bounded by \( \Delta \),

\[ \sum_{(\psi, h) \in S_{ij}(\omega)} \tilde{s}_{\psi, h} + \sum_{(\psi, h) \in T_{ij}(\omega)} \tilde{s}_{\psi, h} \leq 4\hat{\varepsilon}^{32} \Delta |V_i|/m. \]
Consequently,
\[
\sum_{(\psi,h)\in H_{ij}} \tilde{s}_{\psi,h} \left\langle \mathbf{\sigma}[|\partial_{G,s}(\psi,h,j)] - \mathbf{\sigma}[^{\ast}V_i] \right\rangle_G \leq \frac{4\tilde{\varepsilon}^{32} \Delta |V_i|}{m} + \tilde{\varepsilon}^{32} \sum_{(\psi,h)\in H_{ij}} \tilde{s}_{\psi,h} \leq \frac{5\tilde{\varepsilon}^{32} \Delta |V_i|}{m}.
\] (4.11)

Observe that \(m \geq n/\Delta\) because \(d(x) > 0\) for all \(x \in V\). Hence, summing (4.11) over all \(\omega \in \Omega\), we obtain
\[
\sum_{(\psi,h)\in H_{ij}} \tilde{s}_{\psi,h} \left\langle \mathbf{\sigma}[|\partial_{G,s}(\psi,h,j)] - \mathbf{\sigma}[^{\ast}V_i] \right\rangle_{TV}/G \leq \frac{5\tilde{\varepsilon}^{32} \Delta^2 |\Omega| |V_i|}{n}.
\] (4.12)

Further, let \(H = \bigcup_{i,j \in [\Delta]} H_{ij} \times \{j\}\). Summing (4.12) over \(i, j \in [\Delta]\) yields
\[
\sum_{(\psi,h,j)\in H} \tilde{s}_{\psi,h} \left\langle \mathbf{\sigma}[|\partial_{G,s}(\psi,h,j)] - \mathbf{\sigma}[^{\ast}V_i] \right\rangle_{TV}/G \leq \frac{5\tilde{\varepsilon}^{32} \Delta^2 |\Omega|}{n}.
\] (4.13)

To complete the proof, let \(\bar{H} = \{(\psi,h,j) : (\psi,h) \in \Psi[N], j \in [\#\psi]\} \setminus H\). Then the fact that the degrees of the variable nodes are bounded by \(\Delta\) implies that
\[
\sum_{(\psi,h,j)\in \bar{H}} \tilde{s}_{\psi,h} \left\langle \mathbf{\sigma}[|\partial_{G,s}(\psi,h,j)] - \mathbf{\sigma}[^{\ast}V_i] \right\rangle_{TV}/G \leq \sum_{(\psi,h,j)\in \bar{H}} \tilde{s}_{\psi,h} \leq \Delta \sum_{i \notin I} |V_i|/m \leq \tilde{\varepsilon}^{32} \Delta n/m \leq \tilde{\varepsilon}^{32} \Delta^2.
\] (4.14)

Combining (4.13) and (4.14) and using the assumption \(\tilde{\varepsilon} < 1/(6\Delta^2 |\Omega|)\) completes the proof. \(\square\)

To proceed, let \(\delta > 0, (\psi,h) \in \Phi[N]\) and \(j \in [\#\psi]\). We say that \(\sigma \in C_M\) is \(\delta\)-homogeneous w.r.t. \((\psi,h,j)\) if
\[
\left\| \mathbf{\sigma}[|\partial_{G,s}(\psi,h,j)] - \mathbf{\sigma}[^{\ast}V_i] \right\|_{TV} < \delta.
\]
Let \(H(\delta, \sigma)\) be the set of all \((\psi,h,j)\) with respect to which \(\sigma\) fails to be \(\delta\)-homogeneous.

**Lemma 4.11.** We have \(\sum_{(\psi,h,j)\in H(\tilde{\varepsilon}^7, \sigma)} \tilde{s}_{\psi,h} \left\langle G < \tilde{\varepsilon}^7.\right\rangle\)

**Proof.** Let \(H\) be the set of all triples \((\psi,h,j)\) such that \((\psi,h) \in \Psi[N]\) and \(j \in [\#\psi]\). Let
\[
H_0 = \left\{(\psi,h,j) \in H : \left\langle \mathbf{\sigma}[|\partial_{G,s}(\psi,h,j)] - \mathbf{\sigma}[^{\ast}V_i] \right\rangle_{TV}/G \leq \tilde{\varepsilon}^{15} \right\}.
\] (4.15)

Then Lemma 4.10 implies that
\[
\sum_{(\psi,h,j)\in H_0} \tilde{s}_{\psi,h} \leq \tilde{\varepsilon}^{16}.
\] (4.16)

Of course, (4.15) implies that \(\sum_{(\psi,h,j)\in H(\tilde{\varepsilon}^7, \sigma)} \tilde{s}_{\psi,h} \leq \tilde{\varepsilon}^8\) for all \((\psi,h,j)\in H_0\). Consequently,
\[
\sum_{(\psi,h,j)\in H(\tilde{\varepsilon}^7, \sigma)\cap H_0} \tilde{s}_{\psi,h} \left\langle G \leq \sum_{(\psi,h,j)\in H_0} \tilde{s}_{\psi,h} \left\langle \mathbf{\sigma}[\Psi[N]] G \right\rangle < \tilde{\varepsilon}^8.
\]

Thus, the assertion follows from (4.16). \(\square\)

**Proof of Theorem 4.2** Let \(S\) be the set of all \(\tilde{s} \subset \Psi[N]\) such that
\[
\sum_{(\psi,h)\in \Psi[N]} \mathbf{1}(\tilde{s}) \tilde{s}_{\psi,h} < \tilde{\varepsilon}.
\] (4.17)

Further, for \(\tilde{s} \in S\) let \(Q(\tilde{s})\) be the set of all \(\sigma \in C_M\) that are \(\tilde{\varepsilon}^7\)-homogeneous with respect to \((\psi,h,j)\) for all \((\psi,h) \in \Psi[N]\) \(\setminus \tilde{s}\) and all \(j \in [\#\psi]\). Since \(\tilde{s}_{\psi,h} = |\partial_{G,s}(\psi,h)|/m\), Lemma 4.11 implies \(\sum_{\sigma \in Q(\tilde{s})} \delta N^2 - 1 \geq 1/2\). Consequently, there exists \(\tilde{s} \in S\) such that
\[
\mathbf{1}(\sigma \in Q(\tilde{s})) \geq 2^{-1}|\Psi[N]|^{-1}.
\] (4.18)
Let \( P = \mathcal{P}(\Omega)^N \times \prod_{\psi,h} \mathcal{P}(\Omega^\#\psi) \) be the set of all \( p = (p_k, p_{\psi,h})_{k \in [N], (\psi,h) \in \Psi^\# \setminus \hat{s}} \) such that \( p_k \in \mathcal{P}(\Omega) \) for all \( k \in [N] \) and \( p_{\psi,h} \in \mathcal{P}(\Omega^\#\psi) \) for all \( (\psi,h) \notin \hat{s} \). For \( p, p' \in P \) we define
\[
\mathcal{D}(p, p') = \sum_{k \in [N]} \| p_k - p_k' \|_{TV} + \sum_{(\psi,h) \in \Psi^\# \setminus \hat{s}} \| p_{\psi,h} - p'_{\psi,h} \|_{TV}.
\]
Because \( P \) is compact there exists a number \( N^* = N^*(\hat{\varepsilon}, \Delta, \Omega, \Psi, \Theta) > 0 \) and a set \( P^* \subset P \) of size \( |P^*| \leq N^* \) such that for any \( p \in P \) there is \( p^* \in P^* \) such that \( \mathcal{D}(p, p^*) < \hat{\varepsilon}^4 \).

Any \( \sigma \in \mathcal{C}_M \) gives rise to the empirical distributions
\[
\hat{\sigma}_h = \sigma[\cdot|V_h^\sigma] \quad (h \in [N]),
\]
\[
\hat{\sigma}_{\psi,h} = \sigma[\cdot|\partial G, s(\psi,h)] \quad ((\psi,h) \in \Psi[N] \setminus \hat{s}, j \in [\#\psi])
\]
and thus to a point \( \hat{\sigma} = (\hat{\sigma}_h, \hat{\sigma}_{\psi,h}) \in P \). Consequently, (4.18) implies that there exists \( p^* \in P^* \) such that the event
\[
S^* = Q(\hat{s}) \cap \{ \sigma \in \mathcal{C}_M : \mathcal{D}(\hat{\sigma}, p^*) < \hat{\varepsilon}^4 \}
\]
satisfies
\[
(1 \{ \sigma \in S^* \})_G \geq \frac{2^{-|\Psi[N]|/N^*}}{N^*}.
\]

In particular, (4.19) implies that there exists \( \sigma \in Q(\hat{s}) \) such that \( \mathcal{D}(\hat{\sigma}, p^*) < \hat{\varepsilon}^4 \). Hence, \( \sigma \) is \( \hat{\varepsilon}^7 \)-homogeneous with respect to any triple \((\psi,h,j)\) such that \((\psi,h) \in \Psi[N] \setminus \hat{s} \) and \( j \in [\#\psi] \). Therefore, for any \((\psi,h) \notin \hat{s} \) and any \( j \in [\#\psi] \) the jth marginal distribution of \( p_{\psi,h}^\sigma \) satisfies
\[
\| p_{\psi,h}^\sigma_{h,j} - p_{h,j}^\sigma \|_{TV} \leq \| \hat{\sigma}_{\psi,h} - \hat{\sigma}_{h,j} \|_{TV} + \| \sigma[\cdot|\partial G, s(\psi,h,j)] - \sigma[\cdot|V_h^\sigma] \|_{TV} + \| \hat{\sigma}_{h,j} - p_{h,j}^\sigma \|_{TV} \leq 3\hat{\varepsilon}^4. \tag{4.20}
\]
In addition, because any \( \sigma \in \mathcal{C}_M \) has Hamming distance no more than \( \alpha n \) from \( \sigma_s \), we find
\[
\sum_{x \in V} 1 - p_{s(x)}^\sigma(\sigma_s(x)) \leq \alpha n + \mathcal{D}(\hat{\sigma}, p^*) \leq \alpha n + \hat{\varepsilon}^4.
\]

Further, let \( u \) be the uniform distribution on \( \Omega \) and obtain \( p'' \) from \( p^* \) by letting
\[
p''_k = (1 - \hat{\varepsilon}^4)p_k^* + \hat{\varepsilon}^5 u, \quad p''_{\psi,h} = (1 - \hat{\varepsilon}^4)p_{\psi,h}^* + \hat{\varepsilon}^5 u^{\#\psi} \quad \text{for } k \in [N], (\psi,h) \in \Psi[N].
\]
Then (4.20) ensures that for all \((\psi,h) \in \Psi[N], j \in [\#\psi],
\[
\| p''_{h,j} - p_{h,j}^\sigma \|_{TV} \leq 3\hat{\varepsilon}^4, \quad \text{while } p''_k(\omega) \geq \hat{\varepsilon}^5/\Omega \quad \text{for all } k \in [N], \omega \in \Omega.
\]
Hence, provided that \( \hat{\varepsilon} \) is chosen small enough, Fact 3.4 implies that there is \( p' \in P \) such that for all \( k \in [N], (\psi,h) \in \Psi[N], \),
\[
\| p_k - p_k' \|_{TV} < \hat{\varepsilon}^3, \quad \| p_{\psi,h} - p_{\psi,h}' \|_{TV} < \hat{\varepsilon}^3, \quad \sum_{x \in V} 1 - p_{s(x)}^\sigma(\sigma_s(x)) \leq \alpha n, \quad \text{and } p_{\psi,h}^\sigma_{h,j} = p_{h,j}^\sigma.
\]

To complete the construction of the \( \mathcal{M} \)-state \( s \), define \( \hat{s}_h = p_{h}^\sigma \) for all \( h \in [N] \) and
\[
\hat{s}_{\psi,h} = \begin{cases} \hat{s}_h \otimes \cdots \otimes \hat{s}_{h \# \psi} & \text{if } (\psi,h) \in \hat{s}, \\ p_{\psi,h}^\sigma & \text{otherwise.} \end{cases}
\]
Then (4.21) implies that \( s \) satisfies all the requirements of Definition 4.3. Moreover, (4.17) ensures that \( G \models \varepsilon \) s. Finally, (4.19) and (4.21) yield (4.1).

4.5. Proof of Theorem 4.1

Let \( \varepsilon > 0 \) and let \( \mathcal{U} \subset \mathcal{G}(\mathcal{M}) \) be an event such that \( \mathbb{P}[G \in \mathcal{U}] \geq \varepsilon \). The aim in this section is to prove Theorem 4.1. We assume that \( \varepsilon < \varepsilon_0 \) for some small \( \varepsilon_0 = \varepsilon_0(\alpha, \Delta, \Omega, \Psi, \Theta) > 0 \), that \( \gamma = \gamma(\varepsilon, \alpha, \Delta, \Omega, \Psi, \Theta) > 0 \) is small enough, and that \( n > n_0 \) for some large enough \( n_0 = n_0(\alpha, \Delta, \Omega, \Psi, \Theta, \varepsilon, \gamma) \).

The proof of Theorem 4.1 consists of two parts. The following lemma shows that the Bethe free entropy yields the first moment of \( Z_G \) restricted to \( \sigma \) that “belong to” a given state \( s \). More precisely for \( \delta > 0 \) we let
\[
Z_G(s, \delta) = \sum_{\sigma \in \mathcal{C}_M: (G, \sigma) = s} \prod_{a \in F} \psi_a(\sigma(\partial G(a,1)), \ldots, \sigma(\partial G(a,d(a)))) = Z_G \cdot (1 \{ (G, \sigma) = s \})_G.
\]
Moreover, for an $\mathcal{M}$-state $s$ we let $\tilde{s}_h = \frac{1}{m} |V_h'|$ for $h \in \#s$. In addition, we write $m = |F|$ and
\[
B_{\mathcal{M}}(s) = \sum_{h=1}^{\#s} \tilde{s}_h H(\tilde{s}_h) - \frac{m}{n} \sum_{\psi,h} \tilde{s}_{\psi,h} \left[ D \left( \tilde{s}_{\psi,h} \| \tilde{s}_{\psi,h} \right) - \sum_{\omega \in \Omega^{|s|}} \tilde{s}_{\psi,h} \ln \psi(\omega) \right]. \tag{4.22}
\]

**Lemma 4.12.** For any $\varepsilon > 0$, $N > 0$ there exist $\gamma > 0$, $n_0 > 0$ such that the following is true. Let $\mathcal{M}$ be a finite model of size $n \geq n_0$ and let $s$ be a $\mathcal{M}$-state of size $N$. Then $|B_{\mathcal{M}}(s) - n^{-1} \ln \mathbb{E}[Z_G(s,\gamma)|G \models \gamma, s]| < \varepsilon$.

We defer the proof of Lemma 4.12 to Section 4.6.

In the second step of the proof of Theorem 4.1 we are going to use Theorem 4.4 to identify a “dominant” $\mathcal{M}$-state $s$. More specifically, $s$ will be such that $\mathbb{E}[Z_G1(\{G \in \mathcal{U}\})]$ is not much greater than $\mathbb{E}[Z_G(s,\gamma)|G \models \gamma, s]$. To find such a state we need a bit of notation. Let $\Pi_{\mathcal{M}}$ be the set of all permutations $\pi : V \to V$ such that $\pi(x) = y$ implies that $d(x) = d(y)$ and $t(x, i) = t(y, i)$ for all $i \in \{d(x)\}$. Two $\mathcal{M}$-states $s, s'$ are isomorphic (in symbols: $s \cong s'$) if $\#s = \#s'$ and if there exists a bijection $\pi \in \Pi_{\mathcal{M}}$ such that $s' = s \circ \pi$, $\tilde{s} = \tilde{s}'$, $\tilde{s}_h = \tilde{s}_h'$ for all $h \in \#s$, and $\tilde{s}_{\psi,h} = \tilde{s}'_{\psi,h}$ and $\tilde{s}_{\psi,h} = \tilde{s}'_{\psi,h}$ for all $(\psi, h) \in \Psi\#s$. Let $\text{Ism}(s, s')$ be the set of all such isomorphisms. Furthermore, let $\text{Aut}(s) = \text{Ism}(s, s)$ and let $[s]$ denote the isomorphism class of $s$. Additionally, for $G \in \mathcal{G}(\mathcal{M})$ and $\pi \in \Pi_{\mathcal{M}}$ we let $G^\pi \in \mathcal{G}(\mathcal{M})$ be the “permuted” factor graph defined by
\[
G^\pi_{x,i} = G_{\pi(x),i} \quad ((x, i) \in C_V).
\]

**Lemma 4.13.** There exist a $\mathcal{M}$-state $s$ and a factor graph $G_0 \in \mathcal{U}$ such that the following is true. Let $\Gamma = \{G \in \mathcal{G}(\mathcal{M}) : G \models \gamma, s\}$. Then $G_0 \in \Gamma$ and $\ln \mathbb{E}[Z_G1(\{G \in \Gamma\})] < c n + \ln \mathbb{E}[Z_G(s,\gamma)|\Gamma]$.

**Proof.** Let $\eta = \eta(\varepsilon, \alpha, \Delta, \Psi, \Theta) > 0$ be the number promised by Theorem 4.4 and set $N = [1/\eta]$. Let $\Sigma$ be the set of all $\mathcal{M}$-states of size at most $N$. Then Theorem 4.4 yields a map $\mathcal{U} \to \Sigma$, $G \mapsto s'_G$, such that for some $c_1 = c_1(\varepsilon, \alpha, \Delta, \Psi, \Theta) > 0$ we have
\[
Z_G(s'_G,\gamma/2) \geq c_1 Z_G \quad \text{for all } G \in \mathcal{U}. \tag{4.23}
\]

The set $[\Sigma]$ of all isomorphic classes is contained in the product $A \times K$ of a set $A$ of size $|A| \leq N^2 N^n N$ and the compact set $K = \{0, 1\}^{|\Psi(N)|} \times \mathcal{P}(\Omega)^N \times \mathcal{P}(\Omega^{\Delta})^{\Psi(N)|}$. Therefore, there exists a set $\Sigma_* \subset [\Sigma]$ whose size is bounded by $c_2 N^2 N^n N$ for some $c_2 = c_2(\varepsilon, \alpha, \Delta, \Psi, \Theta) > 0$ with the following property. For any $r \in \Sigma$ there exists $s \in \Sigma$ such that $[s] \in \Sigma_*$, $\#s = \#s$, $r(x) = s(x)$ for all $x \in V$, $\tilde{r} = \tilde{s}$ and
\[
\sum_{h \in \#s} \|\tilde{s}_h - \tilde{r}_h\|_{\text{TV}} + \sum_{(\psi, h) \in \Psi\#s} \|\tilde{s}_{\psi,h} - \tilde{r}_{\psi,h}\|_{\text{TV}} < \gamma/2.
\]

Hence, by (4.23) and the pigeonhole principle there exists an $\mathcal{M}$-state $s \in \Sigma_*$ such that
\[
\mathbb{E}[Z_G1(\{G \in \mathcal{U}\})] \leq c_1^{-1} \mathbb{E}[Z_G(s'_G,\gamma/2)1(\{G \in \mathcal{U}\})] \leq \frac{c_2 N (2n)^N}{N c_1} \mathbb{E}[Z_G(s'_G,\gamma)1(\{G \in \mathcal{U}\})1(s'_G \cong s)]. \tag{4.24}
\]

Since we are assuming that $n > n_0$ for some large $n_0$, there exists $c_3 = c_3(\varepsilon, \alpha, \Delta, \Psi, \Theta)$ such that $2 c_2 N (2n)^N / c_1 < n^{c_3}$. Now, consider the event $\Gamma' = \{G \in \mathcal{G}(\mathcal{M}) : \exists r \in \Sigma : r \cong s, G \models \gamma, r\}$. There exists a map $\Gamma' \to \Sigma$, $G \mapsto s_G$ such that the following three conditions hold for all $G \in \Gamma'$:
\[
G \models \gamma, s_G, \quad s_G \cong s, \quad s'_G \cong s \Rightarrow s_G = s'_G.
\]

Then (4.24) entails
\[
\mathbb{E}[Z_G1(\{G \in \mathcal{U}\})] \leq n^{c_3} \mathbb{E}[Z_G(s_G,\gamma)1(\{G \in \Gamma'\})] \leq n^{c_3} \mathbb{E}[Z_G(s_G,\gamma)|\Gamma'].
\]

To complete the proof, let $\Gamma = \{G \in \Gamma' : G \models \gamma, s\}$. We are going to show that
\[
\mathbb{E}[Z_G(s_G,\gamma)|\Gamma'] = \mathbb{E}[Z_G(s,\gamma)|\Gamma]. \tag{4.26}
\]

Indeed, let $\Pi$ be the set of all pairs $(G, \pi)$ with $G \in \Gamma'$ and $\pi \in \text{Ism}(s_G, s)$. There are two canonical maps
\[
\varphi : \Pi \to \Gamma', (G, \pi) \mapsto G, \quad \hat{\varphi} : \Pi \to \Gamma, (G, \pi) \mapsto \tilde{G}.
\]
Clearly, if \( r \equiv s \), then \(|\text{Ism}(r,s)| = |\text{Aut}(s)|\). Therefore, \( \phi \) maps the uniform distribution on \( \Pi \) to the uniform distribution on \( \Gamma' \). Similarly, \( \hat{\phi} \) maps the uniform distribution on \( \Pi \) to the uniform distribution on \( \Gamma \) because for any \( G \in \Gamma \) we have \(|\hat{\phi}^{-1}(G)| = |\text{Aut}(s)|\). Consequently,

\[
\mathbb{E}[Z_G(s_G,\gamma)|\Gamma'] = \frac{1}{|\Pi|} \sum_{(G,\pi) \in \Pi} Z_G(s_G,\gamma) = \frac{1}{|\Pi|} \sum_{(G,\pi) \in \Pi} Z_G^*(s,\gamma) = \mathbb{E}[Z_G(s,\gamma)|\Gamma'],
\]

whence (4.26) follows. Finally, combining (4.25) and (4.26) yields the assertion. \( \square \)

**Proof of Theorem 4.1.**

Further, let us write

\[
\text{To proceed, fix any } \lambda \in \Lambda. \text{ Let } \hat{\phi} \text{ be the set of all integer vectors } \lambda = (\lambda_{\psi,h}(\psi,h))_{\psi,h} \text{ such that for all } \lambda \in \Lambda \text{ we have }
\]

\[
\sum_{(\psi,h) \in \Psi[N]} |\lambda_{\psi,h} - m\hat{s}_{\psi,h}| \leq \gamma m.
\]

Further, for \( \lambda \in \Lambda \) and \( G \in \mathcal{G}(\mathcal{M}) \) we write \( G \models \gamma (s,\lambda) \) if \( G \models \gamma s \) and \( |\partial G,s(\psi,h)| = \lambda_{\psi,h} \) for all \( (\psi,h) \in \Psi[N] \).

In addition, for \( \lambda \in \Lambda \) and \( G \in \mathcal{G}(\mathcal{M}) \) we write \( G \models \gamma (s,\lambda) \) if \( G \models \gamma s \) and \( |\partial G,s(\psi,h)| = \lambda_{\psi,h} \) for all \( (\psi,h) \in \Psi[N] \) such that for all \( k, (\psi,h) \) the following conditions are satisfied:

(i) \( q_k \in \mathcal{P}(\Omega) \) is a probability distribution such that \(|V_k|q_k \in \mathcal{R}^\Omega \) is an integer vector,

(ii) \( q_{\psi,h} \in \mathcal{P}(\Omega^{\#\psi}) \) is such that \( \lambda_{\psi,h} q_{\psi,h} \) is an integer vector.

(iii) \( \sum_{x \in \mathcal{V}_k} q_k(x) = \sum_{(\psi,h) \in \Psi[N]} |\#\psi|:k_{\psi,h} = k_{\psi,h} \psi_{\psi,h} h_{\psi,h} \downarrow j \),

(iv) \( \sum_{x \in \mathcal{V}} q_{\psi,h}(\sigma_\psi(x)) \geq (1 - \alpha)n \),

(v) \( \|q_k - \hat{s}_{\psi,h}\|_{TV} < \gamma \),

(vi) \( \|q_{\psi,h} - \hat{s}_{\psi,h}\|_{TV} < \gamma \) unless \( (\psi,h) \in \delta \).

Further, let us write \( \hat{\sigma} = q \) if \( q_{\psi,h} = \sigma \cdot [V_{\psi,h}] \) for all \( h \in N \) and if for all \( (\psi,h) \in \Psi[N] \) such that \( \partial G,s(\psi,h) \neq 0 \) we have \( q_{\psi,h} = \sigma \cdot [\partial G,s(\psi,h)] \).

\[ Z_G(q) = Z_G(1_{\{\sigma \models q\}})_G. \]

Assuming that \( \gamma = \gamma(\varepsilon,\mathcal{N},\alpha,\Delta,\Omega,\Psi,\Theta) > 0 \) is sufficiently small, we claim that for any \( \lambda \in \Lambda, q \in Q(\lambda) \),

\[
|\ln \mathbb{E}[Z_G(q)|G \models \gamma (s,\lambda)] - B(\lambda, q) - D(\lambda, q)| < \varepsilon n/3,
\]

where

\[
B(\lambda, q) = n \sum_{h=1}^{N} s_h H(q_h) - \sum_{(\psi,h) \in \Psi[N]} \lambda_{\psi,h} \left[ D \left( q_{\psi,h} \| \bigotimes_{j=1}^{\#\psi} q_{\psi,h} \right) - \sum_{\omega \in \Omega^{\#\psi}} q_{\psi,h} (\omega) \ln \psi(\omega) \right],
\]

\[
D(\lambda, q) = - \sum_{(\psi,h) \in \Psi[N]} \lambda_{\psi,h} \sum_{j=1}^{\#\psi} D \left( q_{\psi,h} \| q_{\psi,h,j} \right). \]

Indeed, let \( \Sigma'(q) \) be the set of all \( \sigma \in \Omega^V \) such that \( q_{\psi,h} = \sigma \cdot [V_{\psi,h}] \) for all \( h \in N \) and let \( \Sigma(q) = \Sigma'(q) \cap \mathcal{C}_\mathcal{M} \). Then

\[
|\Sigma'(q)| = \prod_{h \in N} \left( s_h n \right). \]

Furthermore, the condition \( \sum_{x \in \mathcal{V}} q_{\psi,h}(\sigma_\psi(x)) \geq (1 - \alpha)n \) ensures that for \( n > n_0 \) large enough we have \(|\Sigma(q)| \geq \frac{1}{\lambda} |\Sigma(q)'| \). Hence, by Stirling’s formula and once more due to our assumption that \( n > n_0 \),

\[
|\ln |\Sigma(q)| - n \sum_{h=1}^{N} s_h H(q_h) | \leq \varepsilon n/9.
\]

To proceed, fix any \( \sigma \in \Sigma(q) \). Let \( \mathcal{E}(\sigma) \) be the event that \( \sigma \cdot [\partial G,s(\psi,h)] \|_{ij} = q_{\psi,h} \|_{ij} \) for all \( (\psi,h) \in \Psi[N] \) and all \( j \in [\#\psi] \). Given \( G \models \gamma (s,\lambda) \), the random factor graph \( G \) is nothing but a random perfect matching between \( CV \) and
Proposition 4.14. Let $\nu$ be a distribution such that $t(x,i) = t(a,j)$ and such that $|\partial G_s(\psi, h)| = \lambda_{\psi,h}$. Therefore, by Fact 3.5 and because $n > n_0$ we have

$$\left| \ln P[\mathcal{E}(\sigma)|G = \gamma(s,\lambda)] + \sum_{(\psi,h) \in \Psi[N]} \lambda_{\psi,h} \sum_{j=1}^{#\psi} D(q_{\psi,h,j}||q_{\psi,j}) \right| < \varepsilon n/9. \quad (4.30)$$

Finally, let $\mathcal{F}(\sigma)$ be the event that $\sigma[\cdot,|\partial G_s(\psi, h)]] = q_{\psi,h}$ for all $(\psi, h) \in \Psi[N]$. Then by Fact 3.5 we have

$$\left| \ln P[\mathcal{F}(\sigma)|G = \gamma(s,\lambda), \mathcal{E}(\sigma)] + \sum_{(\psi,h) \in \Psi[N]} \lambda_{\psi,h} D(q_{\psi,h}||\bigotimes_{j=1}^{#\psi} q_{\psi,j}) \right| < \varepsilon n/9. \quad (4.31)$$

Hence, (4.28) follows from (4.29)–(4.31) and the linearity of expectation.

Because $s_{\psi,h,j} = \tilde{s}_h$, the conditions $\|q_k - \tilde{s}_k\|_{TV} < \gamma$ and $\|q_{\psi,h} - \tilde{s}_{\psi,h}\|_{TV} < \gamma$ for $(\psi, h) \notin \hat{s}$ imply that

$$\|q_{\psi,h,j} - q_j\|_{TV} < \gamma \quad \text{for all } (\psi, h) \notin \hat{s}, j \in \#\psi. \quad (4.32)$$

Furthermore, (4.27) ensures that $\sum_{(\psi,h) \in \xi} \lambda_{\psi,h} < \gamma m$. Hence, condition (iv), (4.32) and the continuity of $D(\cdot||\cdot)$ imply that $|D(\lambda, q)| < \varepsilon/3$, provided that $\gamma$ is small enough. Consequently, (4.28) yields

$$\left| \ln \mathbb{E}[Z_G(q)|G = \gamma(s,\lambda)] - B(\lambda, q) \right| < 2\varepsilon n/3. \quad (4.33)$$

Assuming that $\gamma > 0$ is small enough, we obtain from Fact 3.5 that

$$|B_M(s) - n^{-1}B(\lambda, q)| < \varepsilon/3 \quad \text{for all } \lambda \in \Lambda, q \in Q(\lambda). \quad (4.34)$$

Moreover, if the event $G = \gamma(s,\lambda)$ occurs, then there exists $\lambda \in \Lambda$ such that $G = \gamma(s,\lambda)$ and $Z_G(\gamma, s) = \sum_{q \in Q(\lambda)} Z_G(q)$. Therefore, (4.33), (4.34) and Bayes’ formula yield

$$n^{-1} \ln \mathbb{E}[Z_G(\gamma, s)|G = \gamma(s,\lambda)] \geq B_M(s) - \varepsilon. \quad (4.35)$$

Conversely, since $|Q(\lambda)| \leq (n^2\Delta)^{\Omega||\Psi(N)||\Delta|} \leq \exp(\varepsilon n)$ for $n > n_0$ large enough, we find

$$n^{-1} \ln \mathbb{E}[Z_G(\gamma, s)|G = \gamma(s,\lambda)] \leq B_M(s) + 2\varepsilon. \quad (4.36)$$

Finally, the assertion follows from (4.35) and (4.36).

4.7. Belief Propagation. If the “maximum distance” parameter satisfies $\alpha \geq 1$, then 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 $\text{Mes}_M(G)$ of $G$, which we define as the set of all maps $\hat{\nu} : C_V \cup C_F \to P(\Omega)$, $(v,j) \mapsto \hat{\nu}_{v,j}$. The Belief Propagation operator $BP : \text{Mes}_M(G) \to \text{Mes}_M(G)$ maps $\hat{\nu} \in \text{Mes}_M(G)$ to $\hat{\nu} = BP(\hat{\nu})$ defined by

$$\hat{\nu}_{x,i}(\omega_i) \propto \prod_{h \in [d(x)] \setminus \{i\}} \hat{\nu}_{G_{x,h}}(\omega_i) \quad \text{for } (x,i) \in C_V, \omega_i \in \Omega,$n\left|

$$\hat{\nu}_{a,j}(\omega_j) \propto \sum_{(\omega_h)_{h \in [d(a)] \setminus \{j\}}} \psi_a(\omega_1, \ldots, \omega_{d(a)}) \prod_{h \in [d(a)] \setminus \{j\}} \hat{\nu}_{G_{a,h}}(\omega_h) \quad \text{for } (a,j) \in C_F, \omega_j \in \Omega. \quad (\alpha/j)$$

Let $\text{Fix}_M(G)$ be the set of all Belief Propagation fixed points, i.e., all $\hat{\nu} \in \text{Mes}_M(G)$ such that $BP(\hat{\nu}) = \hat{\nu}$. Any point $\hat{\nu} \in \text{Fix}_M(G)$ gives rise to probability distributions

$$\hat{\nu}_x(\omega) \propto \prod_{h \in [d(x)]} \hat{\nu}_{G_{x,h}}(\omega) \quad \text{for } x \in V, \omega \in \Omega,$n\left|

$$\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) \quad \text{for } a \in F, \omega_1, \ldots, \omega_{d(a)} \in \Omega.$n\left|

These distributions satisfy condition MS1, while condition MS2 is void in the case $\alpha \geq 1$.

Proposition 4.14. If $\alpha \geq 1$, then $B_M(G) = \max \left\{ B_M(G, \hat{\nu}) : \hat{\nu} \in \text{Fix}_M(G) \right\}$. 
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 $\nu$ approaches the boundary $\partial M$, $B_M(G, \cdot)$ does not attain its global maximum on $\partial M$. Furthermore, for any stationary point $\nu \in M$ of the Bethe free energy $B_M(G, \cdot)$ there exists $\hat{\nu} \in \text{Fix}_M(G)$ such that $B_M(G, \nu) = B_M(G, \hat{\nu})$. \qed

Theorem 4.14 shows that the Bethe free energy provides an upper bound on $n^{-1} \ln Z_{G(M)}$. Furthermore, Proposition 4.15 reduces the problem of calculating the Bethe free energy to the task of determining the “dominant fixed point” of Belief Propagation.

**Remark 4.15.** In the case $\alpha < 1$ the Bethe free energy may attain its maximum on the boundary of the set $M$ of marginal sequences. Specifically, the maximum may occur on the face $\partial M$. Furthermore, for any stationary point $\nu \in M$ of the Bethe free energy $B_M(G, \cdot)$ there exists $\hat{\nu} \in \text{Fix}_M(G)$ such that $B_M(G, \nu) = B_M(G, \hat{\nu})$. \qed

## 5. THE LOWER BOUND

Let $(M(n)) = (V_n, F_n, d_n, t_n, (\psi_{n,a}), (\sigma_{n,a}))$ be a sequence of $(\alpha, \Delta, \Omega, \Psi, \Theta)$-models such that $\# M(n) = n$. Let $G(n) = G(M(n))$ and $\mathcal{G}(n) = \mathcal{G}(M(n))$.

### 5.1. A Bethe-enhanced second moment method.

The cavity method provides a “recipe” for calculating a number $\phi$ such that $(n^{-1} \ln Z_{G(n)})_n$ is deemed to converge to $\phi$ in probability. This number is determined by applying Belief Propagation and the Bethe free energy to the “limit” of the typical local structure of $G(n)$ 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 [24, Part 4] to our current setup and also define an appropriate formulation of “Belief Propagation fixed points” in the limiting object. This can be viewed as a generalisation of the framework developed in [11].

**Definition 5.1.** A $(\alpha, \Delta, \Omega, \Theta, \Psi)$-template consists of a $(\alpha, \Delta, \Omega, \Psi, \Theta)$-model $M$, a connected factor graph $H \in \mathcal{G}(M)$ and a root $(r_H, i_H)$, which is a variable or factor clone. We denote the template by $T_H$. Its size is $\#H = \#M$.

**Definition 5.2.** Two templates $H, H'$ with models $M = (V, F, d, t, (\psi_a), (\sigma_a)), M' = (V', F', d', t', (\psi'_a), (\sigma'_a))$ are isomorphic if there exists a bijection $\pi : V \cup F \to V' \cup F'$ such that the following conditions are satisfied.

1. **ISM1:** $\pi(x) \in V'$ for all $x \in V$ and $\pi(a) \in F'$ for all $a \in F$,
2. **ISM2:** if $r_H = (x_H, i_H)$ and $r_H' = (x_H', i_H')$, then $\pi(x_H) = x_H'$ and $i_H = i_H'$,
3. **ISM3:** $d(v) = d'(\pi(v))$, $\sigma_a(v) = \sigma'_a(\pi(v))$ for all $v \in V \cup F$ and $t(v, i) = t'(\pi(v), i)$ for all $(v, i) \in C_V \cup C_F$,
4. **ISM4:** $\psi_a = \psi_{\pi(a)}$ for all $a \in F$,
5. **ISM5:** for all $(v, i) \in C_V$ we have $H_{\pi(v), i} = (a, j)$ if $H'_{\pi(v), i} = (a, j)$.

We denote the isomorphism class of a template $H$ by $[H]$. Let $\mathfrak{S} = \mathfrak{S}(\alpha, \Delta, \Theta, \Psi)$ be the set of all isomorphism classes of acyclic templates. For each $[H] \in \mathfrak{S}$ 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 $\ell$ from the root if the root is a variable clone and $\ell + 1$ if the root is a factor node. We endow $\mathfrak{S}$ with the coarsest topology that makes all the functions $\Gamma \in \mathfrak{S} \to 1\{\partial^\ell \Gamma = \partial^\ell \Gamma_0\} \in \{0, 1\}$ for $\ell \geq 1, \Gamma_0 \in \mathfrak{S}$ continuous. Moreover, the space $\mathcal{P}(\mathfrak{S})$ of probability measures on $\mathfrak{S}$ carries the weak topology. So does the space $\mathcal{P}(\mathfrak{S})$ of probability measures on $\mathcal{P}(\mathfrak{S})$. For $\Gamma \in \mathfrak{S}$ and $\lambda \in \mathcal{P}(\mathfrak{S})$ we write $\delta_\Gamma \in \mathcal{P}(\mathfrak{S})$ and $\delta_\lambda \in \mathcal{P}(\mathfrak{S})$ for the Dirac measure.

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 $(v, i)$. Let $T_G$ be the $\sigma$-algebra on $\mathcal{G}(n)$ generated by the functions $(G, v, i) \mapsto \partial^\ell[G, v, i]$. Each $G \in \mathcal{G}(n)$ gives rise to the empirical distribution

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

Let $\Lambda_n = E[\delta_{\lambda_G(n)}] \in \mathcal{P}(\mathfrak{S})$. We say that $(\mathcal{M}(n))_n$ converges locally to $\theta \in \mathcal{P}(\mathfrak{S})$ if $\lim_{n \to \infty} \Lambda_n = \delta_\theta$.

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

$$
\lim inf_{n \to \infty} \mathbb{P}[G(n) is \ell-acyclic] > 0.
$$
In this setup, the key prediction of the “replica symmetric cavity method” can be cast as follows: \((n^{-1} \ln Z_{G(n)})_n\) converges in probability to the Bethe free energy of a “Belief Propagation fixed point” on the (possibly infinite) trees in the support of \(\vartheta\); see \([25]\) for more background. To formalise this, let \(T_\vartheta \in \mathcal{T}\) be a sample from \(\vartheta \in \mathcal{P}(\mathcal{T})\). Further, let \(\mathcal{V}\) be the event that the root of \(T_\vartheta\) is a variable clone and let \(\mathcal{F}\) be the event that the root is a constraint clone. For \(T \in \mathcal{T}\) let \(d_T\) denote the degree of the root of \(T\). Moreover, for \(j \in [d_T]\) let \(T \uparrow j \in \mathcal{T}\) denote the tree pending on the \(j\)th child of the root of \(T\). Additionally, let \(\sigma_\ast(T)\) be the value assigned to the root if it is a variable clone. Finally, if the root is the clone of a constraint node we let \(\psi_T\) be its associated function.

**Definition 5.3.** A measurable map \(p : \mathcal{T} \to \mathcal{P}(\Omega)\), \(T \mapsto p_T\) is called a \(\vartheta\)-**Belief Propagation fixed point** if the following conditions are satisfied \(\vartheta\)-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
   \[
   p_T(\omega_i) \propto \sum_{(\omega_j) \in [d_T] \setminus \{i\}} \psi(\omega_1, \ldots, \omega_{d_T}) \prod_{j \in [d_T] \setminus \{i\}} p_{T \uparrow j}(\omega_j).
   \]

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

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

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

The **Bethe free energy** of \(p\) with respect to \(\vartheta\) is

\[
\mathcal{B}_\vartheta(p) = \mathbb{E}[d_T^{-1}(1 - d_T\vartheta)H(\hat{p}_T(\omega)|\mathcal{V}) - \mathbb{E}[d_T^{-1}D(\hat{p}_T(\omega)|\mathcal{F})]|\mathcal{F}]\mathbb{E}[d_T\vartheta|\mathcal{F}].
\]

Finally, to obtain a sufficient condition for the convergence \(n^{-1} \ln Z_{G(n)} \to \mathcal{B}_\vartheta(p)\) we are going to apply Theorem 41 to upper-bound the second moment of \(Z_{G(n)}\). The necessary construction, reminiscent of those used in \([12, 13, 14, 18, 28, 32]\), is as follows.

**Proposition 5.4.** For any \(\alpha, \Delta, \Omega, \Psi, \Theta\) and any \(\varepsilon > 0\) there exists \(\eta > 0\) such that the following is true. Suppose that \(\mathcal{M}\) is a \((\alpha, \Delta, \Omega, \Psi, \Theta)\)-model of size \(n = \# \mathcal{M} \geq 1/\eta\). There exists a finite set of functions \(\Psi^\otimes\) and a \((2\alpha, \Delta, \Omega \times \Omega, \Psi^\otimes, \Theta^\otimes)\)-model \(\mathcal{M}^\otimes\) with the following properties.

1. There is a bijection \(\mathcal{G}(\mathcal{M}) \to \mathcal{G}(\mathcal{M}^\otimes)\), \(G \to G^\otimes\).
2. Let \(U \subset \mathcal{G}(\mathcal{M})\) be an event such that \(\mathbb{P}[G \in U] > \varepsilon\). Then \(n^{-1} \ln \mathbb{E}[Z_{G}^2|U] \leq \mathbb{E}[Z_{G^\otimes}^2|U] + \varepsilon\).

**Proof.** Let \(\omega^\otimes = 2\alpha, \Delta^\otimes = \Delta, \Omega^\otimes = \Omega \times \Omega\) and \(\Theta^\otimes = \Theta\). Let us denote the pairs \((\omega, \omega') \in \Omega^\otimes\) by \(\omega \otimes \omega'\). Further, for any map \(\psi \in \Psi\) we let \(\psi^\otimes = \psi \otimes \psi\). Explicitly,

\[
\psi^\otimes : (\Omega^\otimes)^\# \to (0, \infty), \quad (\omega_1 \otimes \omega', \ldots, \omega_{\#_\psi} \otimes \omega'_{\#_\psi}) \mapsto \psi(\omega_1, \ldots, \omega_{\#_\psi}) \cdot \psi(\omega_1', \ldots, \omega'_{\#_\psi}).
\]

Moreover, \(V^\otimes = V\), \(F^\otimes = F\), \(d^\otimes = d\), \(t^\otimes = t\) and \(\sigma^\otimes(x) = \sigma_\ast(x) \otimes \sigma_\ast(x)\). Finally, for any \(a \in F\) we let \(\psi^\otimes_a = (\psi_a)^\otimes\). Hence, we obtain the \((\alpha^\otimes, \Delta^\otimes, \Omega^\otimes, \Psi^\otimes, \Theta^\otimes)\)-model \(\mathcal{M}^\otimes = (V^\otimes, F^\otimes, d^\otimes, t^\otimes, (\psi^\otimes_a)_{a \in F}, \sigma^\otimes_\ast)\). By construction, \(\mathcal{G}(\mathcal{M})\) is in one-to-one correspondence with \(\mathcal{G}(\mathcal{M}^\otimes)\). Let \(G \mapsto G^\otimes\) denote the canonical map \(\mathcal{G}(\mathcal{M}) \to \mathcal{G}(\mathcal{M}^\otimes)\). Since it is immediate from the construction that \(Z_G^2 \leq Z_{G^\otimes}\) for all \(G \in \mathcal{G}(\mathcal{M})\), the assertion follows from Theorem 4.1}

**Theorem 5.5.** Suppose that \((\mathcal{M}(n))_{n \geq 1}\) has high girth and converges locally to \(\vartheta \in \mathcal{P}(\mathcal{T})\). Furthermore, assume that there is a \(\vartheta\)-Belief Propagation fixed point \(p\) such that for any \(\varepsilon > 0\) we have

\[
\lim_{n \to \infty} \lim_{\ell \to \infty} \frac{1}{n} \ln \mathbb{E}[Z_{G(n)} \mathbb{1}_{\mathcal{B}_\vartheta^\otimes(n)}(G^\otimes(n)) \leq 2B_\vartheta(p) + \varepsilon] = 1 \quad \text{and} \quad \lim_{n \to \infty} \lim_{\ell \to \infty} \frac{1}{|T_\ell|} \ln \mathbb{E}[Z_{G(n)} \mathbb{1}_{\mathcal{B}_\vartheta^\otimes(n)}(G^\otimes(n)) \leq 2B_\vartheta(p) + \varepsilon|T_\ell]] = 0.
\]

Then \(\frac{1}{n} \ln Z_G(n)\) converges to \(B_\vartheta(p)\) in probability.
For a given $\phi$ the construction or, at least, identification of the $\phi$-Belief Propagation fixed point $\theta$ in Theorem 5.5 is similar to the computations done in the physics literature. However, to apply Theorem 5.5 it will generally be necessary to perform these calculations more thoroughly, e.g., by means of the contraction method [29]. Further, to verify condition (5.1) 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 $G^\otimes(n)$. This task may be far from trivial, but at least it is a well-defined combinatorial problem.

Finally, (5.2) provides that given that the local structure up to depth $\ell$ is “typical”, conditioning on the event that the second moment Bethe free energy is bounded by $2B_\phi(p) + \varepsilon$ 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 [7].

5.2. Proof of Theorem 5.5 The following lemma is the key ingredient to the proof of Theorem 5.5

Lemma 5.6. Let $\varepsilon > 0$ and let $\mathcal{U}_n(\varepsilon) = \{B_{\mathcal{M}_{n}^\otimes}(G^\otimes(n)) \leq 2B_\phi(p) + \varepsilon\}$. Assume that $\lim_{n \to \infty} P[\mathcal{U}_n(\varepsilon)] = 1$. Then

$$\liminf_{n \to \infty} n^{-1} \ln \mathbb{E}[Z_{G(n)}|\mathcal{U}_n(\varepsilon)] \geq B_\phi(p).$$

The proof of Lemma 5.6 rests on a generalisation of arguments used in [7, 15] for (plain) random graphs and the $k$-SAT problem, respectively. Before we carry out the details in Section 5.3 we show how it implies Theorem 5.5. In addition to Lemma 5.6 we employ the following simple observation.

Lemma 5.7. For any $\varepsilon > 0$ there is $\delta > 0$ such that $\limsup_{n \to \infty} \frac{1}{n} \ln \mathbb{P}\left[\left|\ln Z_{G(n)} - \mathbb{E}[\ln Z_{G(n)}]\right| > \varepsilon\right] \leq -\delta$.

Proof. There exists a number $\beta > 0$ that depends only on $\alpha, \Delta, \Omega, \Psi, \Theta$ such that the following is true. Assume that $G, G' \in G(n)$ are two factor graphs such that there exist two variable clones $(v, i), (v', i')$ such that

$$G_{w,j} = G'_{w,j} \quad \text{for all } (w, j) \in C_V \setminus \{(v, i), (v', i')\}.$$

Then $|\ln Z_G - \ln Z_{G'}| \leq \beta$. Therefore, the assertion follows from Azuma’s inequality. □

Proof of Theorem 5.5 Let $\varepsilon > 0$ and let $\mathcal{U}_n(\varepsilon) = \{B_{\mathcal{M}_{n}^\otimes}(G^\otimes(n)) \leq (2B_\phi(p) + \varepsilon)n\}$. Then Proposition 5.4 yields

$$\limsup_{n \to \infty} \frac{1}{n} \ln \mathbb{E}[Z_{G(n)}^2|\mathcal{U}_n(\varepsilon)] \leq 2B_\phi(p) + \varepsilon. \quad (5.3)$$

Together with Lemma 5.6 and the Paley-Zygmund inequality, (5.3) implies that

$$\liminf_{n \to \infty} \frac{1}{n} \ln \mathbb{P}\left[\frac{1}{n} \ln Z_{G(n)} \geq B_\phi(p) - \varepsilon|\mathcal{U}_n(\varepsilon)\right] \geq -3\varepsilon. \quad (5.4)$$

Since (5.4) holds for any $\varepsilon > 0$ and because $\lim_{n \to \infty} \mathbb{P}[\mathcal{U}_n(\varepsilon)] = 1$ for any $\varepsilon > 0$, Lemma 5.7 implies that

$$\liminf_{n \to \infty} \frac{1}{n} \mathbb{E}[\ln Z_{G(n)}] \geq B_\phi(p). \quad (5.5)$$

Conversely, (5.3) entail that for any $\varepsilon > 0$,

$$\limsup_{n \to \infty} \frac{1}{n} \ln \mathbb{E}[Z_{G(n)}|\mathcal{U}_n(\varepsilon)] \leq \frac{1}{2} \limsup_{n \to \infty} \frac{1}{n} \ln \mathbb{E}[Z_{G(n)}^2|\mathcal{U}_n(\varepsilon)] \leq B_\phi(p) + \varepsilon. \quad (5.6)$$

Because $\lim_{n \to \infty} \mathbb{P}[\mathcal{U}_n(\varepsilon)] = 1$ for any $\varepsilon > 0$, (5.6) yields

$$\limsup_{n \to \infty} \frac{1}{n} \mathbb{E}[\ln Z_{G(n)}] \leq B_\phi(p). \quad (5.7)$$

Finally, the assertion follows from (5.5), (5.7) and one further application of Lemma 5.7. □
5.3. Proof of Lemma 5.6  Let $\mathcal{T}_\ell = \{ \partial^\ell \Gamma : \Gamma \in \mathcal{T} \}$ be the set of all isomorphism classes of acyclic templates truncated after $\ell$ levels if the root is a variable clone and after $\ell + 1$ levels otherwise. Additionally, let $\mathcal{F}_\ell$ be the $\sigma$-algebra generated by the functions $\Gamma \mapsto 1\{ \partial^\ell \Gamma \cong \partial^\ell \Gamma' \}$ for $\Gamma' \in \mathcal{G}$. Let $\bar{p}^{(\ell)} = E[p|\mathcal{F}_\ell]$. Because the sequence $(\bar{p}^{(\ell)})_\ell$ is a martingale we have $\lim_{\ell \to \infty} \bar{p}^{(\ell)} = \bar{p}$ almost surely. More specifically, we may choose a version of the conditional expectation such that for any $T \in \mathcal{T}$ rooted at a factor node $(a, i)$ and any $j \in [d_T]$ we have

$$p^{(\ell)}(T_{ij}) = \bar{p}^{(\ell)}(T)_{ij}.$$

Let $\ell > 0$ be large, let $\varepsilon > 0$ be small and assume that $n$ is large enough. Consider a graph $G$ that is $100\ell$-acyclic such that $\lambda_n, G|\mathcal{F}_\ell$ is sufficiently close to $\partial|\mathcal{F}_\ell$ in total variation. Additionally, consider an assignment $\sigma : V_n \to \Omega$.

Let $\Theta_\ell \subset \mathcal{T}_\ell$ be the set of all trees rooted at variable clones. Define a map $t_{l,G} : C_V \cup C_F \to \Theta_\ell$ by letting $t_{l,G}(v, i) = \partial^l[G, v, i]$ and $t_{l,G}(a, j) = t_{l,G}(v, i)$ if $G_{v, i} = (a, j)$. Let $M(G, \ell)$ be the corresponding model. Then $G(M(G, \ell)) \subset G(n)$ because $t_{l,G}$ is a refinement of $t$. Hence, there is a canonical immersion $l : G(M(G, \ell)) \to G(n)$. Let us write $G \cong_{\ell} G'$ if $t_{l,G} = t_{l,G'}$.

Lemma 5.8. Assume that $G' \in G(M(G, \ell))$ is $2\ell + 2$-acyclic. Then $G \cong_{\ell} G'$.

Proof. We are going to show inductively for $l \in [\ell]$ that $G \cong_l G'$. The case $l = 1$ is immediate from the construction. Thus, assume that $l \geq 1$, let $(x, i) \in C_V$ and let $B$ be the set of all clones that have distance precisely $l - 1$ from $(x, i)$. Since $G'$ is $(2\ell + 2)$-acyclic, the pairwise distance of any two clones in $B$ is at least 2. Moreover, by induction we know that $t_{l,G}(w, j) = t_{l,G'}(w, j)$ for all $(w, j) \in B$. Therefore, $t_{l,G}(x, i) = t_{l,G'}(x, i)$.

We define a second model that accommodates $\sigma$. Its type set is $\Theta_\ell \times \Omega$. Define $t_{l,G,\sigma} : C_V \cup C_F \to \Theta_\ell \times \Omega$ by letting $t_{l,G,\sigma}(x, i) = (\partial[l][G, x, i], \sigma(x))$ for $(x, i) \in C_V$ and $t_{l,G,\sigma}(a, j) = t_{l,G}(x, i)$ if $G_{x, i} = (a, j)$. Let $M(G, \sigma, \ell)$ be the corresponding model. Clearly, $G(M(G, \sigma, \ell)) \subset G(M(G, \ell))$. For $G', G'' \in G(n)$ define

$$\text{dist}(G', G'') = \sum_{(v, i) \in C_V} 1 \{ G'_{v, i} \neq G''_{v, i} \}.$$

Lemma 5.9. For any $\varepsilon > 0$ there is $n_0$ such that for $n > n_0$ the following holds with probability at least $1 - \varepsilon$.

There exists a $4\ell$-acyclic $G' \in G(M(G, \ell))$ such that $\text{dist}(G', G(M(G, \sigma, \ell))) \leq n^{0.9}$.

The proof of Lemma 5.9 rests on the following two observations. Let $T = \{ t_{l,G,\sigma}(v, i) : (v, i) \in C_V \}$ be the set of all possible types. For each $\tau \in T$ let $n_\tau$ be the number of clones $(v, i)$ with $t_{l,G,\sigma}(v, i) = \tau$.

Lemma 5.10. There exists $\beta > 0$ such that for $n$ is large enough the following is true. For any $G$, there exists $3/4 < \gamma < 7/8$ such that for all $(v, i) \in C_V$ either

(i) \( \| (w, j) \in C_V : t_{l,G,\sigma}(w, j) = t_{l,G,\sigma}(v, i) \| \leq n^{\gamma} \), or

(ii) \( \| (w, j) \in C_V : t_{l,G,\sigma}(w, j) = t_{l,G,\sigma}(v, i) \| \geq n^{\gamma + \beta} \).

Proof. The number of possible types is bounded independently of $n$. Hence, choosing $\beta$ small enough, we can ensure that there exists an integer $j > 0$ such that $3/4 + j\beta < 7/8$ such that $|n^{3/4 + j\beta}|, n^{3/4 + (j + 1)\beta} \cap \{ n_\tau : \tau \in T \} = \emptyset$.

Fix $\beta, \gamma$ as in the previous lemma. Call $\tau$ rare if $n_\tau \leq n^{\gamma}$ and common otherwise. Let $Y$ be the number of variable clones that belong to cycles of length at most $10\ell$ in $G(M(G, \sigma, \ell))$.

Lemma 5.11. For large enough $n$ we have $E[Y] \leq n^{\gamma} \ln n$.

Proof. Let $R$ be the set of variable clones $(v, i)$ of a rare type and let $U$ be the set of all variable clones whose distance from $R$ in $G$ does not exceed $100\ell$. Since the maximum degree as well as the total number of types are bounded, we have $|U| \leq |R| \ln n \leq n^{\gamma} \sqrt{\ln n}$, provided that $n$ is big enough. Thus, to get the desired bound on $E[Y]$ we merely need to consider the set $W$ of common clones that are at distance more than $100\ell$ from $R$.

More specifically, let $(v, i)$ be a common clone. We are going to bound the probability that $(v, i) \in W$ and that $(v, i)$ lies on a cycle of length at most $10\ell$. To this end, we are going to explore the (random) factor graph from $(v, i)$ via the principle of deferred decisions. Let $i_1 = i, \ldots, i_l \in [\Delta]$ be a sequence of $l \leq 10\ell$ indices. If $(v, i)$ lies on a cycle of length at most $10\ell$, then there exists such a sequence $(i_1, \ldots, i_l)$ that corresponds to this cycle. Namely, with $v_0 = v$ the cycle comprises of the clones $(v_0, i_1), \ldots, (v_{l}, i_l)$ such that $G_{v_0, i_1}(M(G, \sigma, \ell)) = (v_{l+1}, i_{l+1})$. In particular, $v_0 = v_1$. Clearly, the total number of sequences $(i_1, \ldots, i_l)$ is bounded. Furthermore, given that $(v_1, i_1)$ is common, the probability that $v_0 = v_0$ is bounded by $2n^{-\gamma}$. Since $\gamma > 3/4$, the linearity of expectation implies that $E[Y] \leq |U| + 2n^{1-\gamma} \ln n \leq n^{\gamma} \ln n$. □
Lemma 5.12. Assume that $G''' \in \mathcal{G}(\mathcal{M}(\sigma, \ell))$ satisfies $Y(G''') \leq n^\gamma \ln^2 n$. Then there is a $4\ell$-acyclic $G' \in \mathcal{G}(\mathcal{M}(G, \ell))$ such that $\operatorname{dist}(G', G''') \leq n^{0.9}$.

Proof. Let $R$ be the set of variable clones $(v, i)$ of a rare type and let $U$ be the set of all variable clones whose distance from $R$ in $G$ does not exceed $10\ell$. Moreover, let $G'''' \in \mathcal{G}(\mathcal{M}(G, \ell))$ minimise $\operatorname{dist}(G'', G''')$ subject to the condition that $G''''_{v,i} = G_{v,i}$ for all $(v, i) \in U$. Then $\operatorname{dist}(G', G''') \leq n^\gamma \ln n$ because the total number of types is bounded. Therefore, the assumption $Y(G''') \leq n^\gamma \ln^2 n$ implies that $Y(G'') \leq n^\gamma \ln^3 n$, say. In addition, because $G$ is $100\ell$-acyclic, none of the clones in $R$ lies on a cycle of length at most $4\ell$ in $G''''$.

Altering only a bounded number of edges in each step, we are now going to remove the short cycles of $G''''$ one by one. Let $C$ be the set of common clones. The construction of $G''''$ ensures that only common clones lie on cycles of length at most $4\ell$. Consider one such clone $(v, i)$ and let $N$ be the set of all variable clones that are at distance two from $(v, i)$ in $G''''$. Once more by the construction of $G''''$ we have $N \subset C$. Furthermore, $|N| \leq \Delta^2$.

We claim that there exists $N' \subset C$ and a bijection $\xi : N \rightarrow N'$ such that the following conditions are satisfied.

(i) $t_{\ell,G}(w, j) = t_{\ell,G}(\xi(w), j)$ for all $(w, j) \in N$.
(ii) the pairwise distance in $G''''$ between any two clones in $N'$ is at least $100\ell$.
(iii) the distance in $G''''$ between $N \cup \{(v, i)\}$ and $N'$ is at least $100\ell$.
(iv) the distance between $R$ and $N'$ is at least $100\ell$.
(v) any $(w, j) \in N'$ is at distance at least $100\ell$ from any clone that belongs to a cycle of $G''''$ of length at most $4\ell$.

Since the maximum degree of $G''''$ is bounded by $\Delta$, there are no more than $n^\gamma \ln^4 n$ clones violate condition (iii), (iv) or (v). By comparison, there are at least $n^{\gamma+\beta}$ clones of any common type. Hence, the existence of $\xi$ follows.

Now, obtain $G''''$ from $G''''$ as follows.

- let $G''''_{\xi(w,j)} = G''_{w,j}$ and $G''''_{w,j} = G''_{\xi(w,j)}$ for all $(w, j) \in N$.
- let $G''''_{w,j} = G''_{w,j}$ for all $(w, j) \notin N \cup N'$.

It is immediate from the construction that any clone on a cycle of length at most $4\ell$ in $G'''''$ also lies on such a cycle of $G''''$. Moreover, $(v, i)$ does not lie on a cycle of length at most $4\ell$ in $G'''''$. Hence, $Y(G''') < Y(G''')$. In addition, all clones on cycles of length $4\ell$ and their neighbours are common. Hence, the construction can be repeated on $G''''$. Since $Y(G''') \leq n^\gamma \ln^3 n$, we ultimately obtain a $4\ell$-acyclic $G'''$ with $\operatorname{dist}(G', G''') \leq n^\gamma \ln^3 n < n^{0.9}$. \hfill \square

Proof of Lemma 5.9 The assertion is immediate from Lemmas 5.11 and 5.12 and Markov’s inequality. \hfill \square

Proof of Lemma 5.6 Let $\bar{p}^{(\ell)} = E[\bar{p}|F_{\ell}]$. Because $(\bar{p}^{(\ell)})_{\ell}$ is a martingale, we have $\lim_{\ell \rightarrow \infty} \bar{p}^{(\ell)} = \bar{p}$ almost surely. Further, $\bar{p}^{(\ell)}$ gives rise to a $\mathcal{M}(G, \ell)$-state $s^{(\ell)}$. Namely, $s^{(\ell)}$ simply partitions the variable nodes according to the types $t_{\ell,G}$, i.e., $x, y$ belong to the same class iff their degrees coincide and $t_{\ell,G}(x, i) = t_{\ell,G}(y, i)$ for all $i$. Now, for a variable node $x$ let $\bar{s}^{(x)} = \bar{p}^{(\ell)}(\tau)$ for $\tau \in \overline{\mathcal{S}}_{\ell}$ such that $\bar{\delta}^{(\ell)}_G(x, 1) \in \tau$. Similarly, for a constraint node $a$ let $\bar{\delta}^{(a)}_G(\tau)$ for $\tau \in \overline{\mathcal{S}}_{\ell}$ such that $\bar{\delta}^{(a)}_G(1, a) \in \tau$. Then the definition of $B_\varepsilon(p)$ ensures that for any $\varepsilon > 0$ there exist $\delta > 0$ and $\ell_0 > 0$ such that $\ell > \ell_0$ and $\|\mathcal{L}_G F_{\ell} - \bar{p} F_{\ell}\|_{TV} < \delta$ imply that

$$\left| B_{\mathcal{M}(G, \ell)}(s^{(\ell)}) - B_\varepsilon(p) \right| < \varepsilon, \quad (5.8)$$

where we remember the definition of the first term from (4.22).

Further, Lemma 4.12 implies that for large enough $n$ we have

$$n^{-1} \ln E[Z_{\mathcal{M}(G, \ell)}] \geq B_{\mathcal{M}(G, \ell)}(s^{(\ell)}) - \varepsilon. \quad (5.9)$$

Combining (5.9) with Lemmas 5.8 and 5.9 we conclude that for large enough $n$,

$$n^{-1} \ln E[Z_{\mathcal{M}(n)}(\mathcal{G}(\mathcal{M}))] \geq B_{\mathcal{M}(G, \ell)}(s^{(\ell)}) - 2\varepsilon. \quad (5.10)$$

Finally, the assertion follows from (5.8) and (5.10). \hfill \square

REFERENCES

[1] D. Achlioptas, A. Coja-Oghlan: Algorithmic barriers from phase transitions. Proc. 49th FOCS (2008) 793–802.
[2] D. Achlioptas, C. Moore: Random k-SAT: two moments suffice to cross a sharp threshold. SIAM Journal on Computing 36 (2006) 740–762.
[3] D. Achlioptas, A. Naor, Y. Peres: Rigorous location of phase transitions in hard optimization problems. Nature 435 (2005) 759–764.
[4] D. Achlioptas, A. Naor, Y. Peres: On the maximum satisfiability of random formulas. J. ACM 54 (2007).
[5] D. Achlioptas, Y. Peres: The threshold for random k-SAT is $2^k \ln 2 - O(k)$. Journal of the AMS 17 (2004) 947–973.
[6] H. Bethe: Statistical theory of superlattices. Proc. R. Soc. London A 150 (1935) 552–558.
[7] C. Bordenave, P. Caputo: Large deviations of empirical neighborhood distribution in sparse random graphs. arXiv:1308.5725 (2013).
[8] A. Coja-Oghlan, K. Panagiotou: The asymptotic $k$-SAT threshold. arXiv:1310.2728 (2014).
[9] A. Coja-Oghlan, L. Zdeborová: The condensation transition in random hypergraph 2-coloring. Proc. 23rd SODA (2012) 241–250.
[10] A. Dembo, O. Zeitouni: Large deviations techniques and applications (2nd ed.). Springer (1997).
[11] A. Dembo, A. Montanari, N. Sun: Factor models on locally tree-like graphs. Annals of Probability 41 (2013) 4162–4213.
[12] A. Dembo, A. Montanari, A. Sly, N. Sun: The replica symmetric solution for Potts models on $d$-regular graphs. Comm. Math. Phys. 327 (2014) 551–575.
[13] J. Ding, A. Sly, N. Sun: Satisfiability threshold for random regular NAE-SAT. Proc. 46th STOC (2014) 814–822.
[14] J. Ding, A. Sly, N. Sun: Maximum independent sets on random regular graphs. arXiv:1310.4787 (2013).
[15] J. Ding, A. Sly, N. Sun: Proof of the satisfiability conjecture for large $k$. arXiv:1411.0650 (2014).
[16] S. Franz, M. Leone: Replica bounds for optimization problems and diluted spin systems. J. Statist. Phys. 111 (2003) 535–564.
[17] A. Frieze, N. Wormald: Random $k$-Sat: a tight threshold for moderately growing $k$. Combinatorica 25 (2005) 297–305.
[18] A. Galanis, D. Stefankovic, E. Vigoda: Inapproximability for antiferromagnetic spin systems in the tree non-uniqueness region. Proc. 46th STOC (2014) 823–831.
[19] R. Gallager: Low-density parity check codes. MIT Press (1963).
[20] H. Georgii, H.: Gibbs measures and phase transitions. De Gruyter (2011).
[21] F. Guerra: Broken replica symmetry bounds in the mean field spin glass model. Comm. Math. Phys. 233 (2003) 1–12.
[22] S. Janson: Random regular graphs: asymptotic distributions and contiguity. Combinatorics, Probability and Computing 4 (1995) 369–405.
[23] F. Krzakala, A. Montanari, F. Ricci-Tersenghi, G. Semerjian, L. Zdeborova: Gibbs states and the set of solutions of random constraint satisfaction problems. Proc. National Academy of Sciences 104 (2007) 10318–10323.
[24] L. Lovász: Large networks and graph limits. Colloquium Publications 60 (2012), AMS.
[25] M. M´ ezard, A. Montanari: Information, physics and computation. Oxford University Press 2009.
[26] M. M´ ezard, G. Parisi, R. Zecchina: Analytic and algorithmic solution of random satisfiability problems. Science 297 (2002) 812–815.
[27] A. Montanari, D. Shah: Counting good truth assignments of random k-SAT formulae. Proc. 18th SODA (2007) 1255–1264.
[28] E. Mossel, D. Weitz, N. Wormald: On the hardness of sampling independent sets beyond the tree threshold. Probability Theory Related Fields 143 (2009) 401–439.
[29] R. Neininger, L. Rüschendorf: A general limit theorem for recursive algorithms and combinatorial structures. The Annals of Applied Probability 14 (2004) 378–418.
[30] D. Panchenko, M. Talagrand: Bounds for diluted mean-fields spin glass models. Probability Theory Related Fields 130 (2004) 319–336.
[31] J. Pearl: Probabilistic reasoning in intelligent systems: networks of plausible inference. Morgan Kaufmann, San Francisco, CA, USA, 1988.
[32] A. Sly, N. Sun: The computational hardness of counting in two-spin models on $d$-regular graphs. Proc. 53rd FOCS (2012) 361–369.
[33] E. Szemerédi: Regular partitions of graphs. Colloq. Internat. CNRS 260 (1978) 399–401.
[34] J. Yedidia, W. Freeman, Y. Weiss: Constructing free-energy approximations and generalized Belief Propagation algorithms. IEEE Transactions on Information Theory 51 (2005) 2262–2312.

Amin Coja-Oghlan, acoghan@math.uni-frankfurt.de, Goethe University, Mathematics Institute, 10 Robert Mayer St, Frankfurt 60325, Germany.

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