Combinatorics, Probability and Computing (2020), **29**, pp. 346–422
doi:10.1017/S0963548319000440

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

The replica symmetric phase of random constraint satisfaction problems

Amin Coja-Oghlan[†*], Tobias Kapetanopoulos[‡] and Noela Müller

Mathematics Institute, Goethe University, 10 Robert Mayer Str., Frankfurt 60325, Germany.
*Corresponding author. Email: acoghlan@math.uni-frankfurt.de

(Received 26 February 2018; revised 23 September 2019; first published online 3 December 2019)

Abstract

Random constraint satisfaction problems play an important role in computer science and combinatorics. For example, they provide challenging benchmark examples for algorithms, and they have been harnessed in probabilistic constructions of combinatorial structures with peculiar features. In an important contribution (Krzakala *et al.* 2007, *Proc. Nat. Acad. Sci.*), physicists made several predictions on the precise location and nature of phase transitions in random constraint satisfaction problems. Specifically, they predicted that their satisfiability thresholds are quite generally preceded by several other thresholds that have a substantial impact both combinatorially and computationally. These include the condensation phase transition, where long-range correlations between variables emerge, and the reconstruction threshold. In this paper we prove these physics predictions for a broad class of random constraint satisfaction problems. Additionally, we obtain contiguity results that have implications for Bayesian inference tasks, a subject that has received a great deal of interest recently (*e.g.* Banks *et al.* 2016, *Proc. 29th COLT*).

2010 MSC Codes: Primary: 05C80, 68Q87, 82B26; Secondary: 82B20

1. Introduction

1.1 Background and motivation

Random constraint satisfaction problems ('CSPs') have come to play a prominent role at the junction of combinatorics, computer science and statistical physics [7]. In combinatorics the study of random CSPs goes back to the seminal paper by Erdős and Rényi that started the theory of random graphs [40]. In modern language, they posed the problem of pinpointing the threshold for q-colourability in random graphs, a question that remains open to this day but that has nevertheless sparked ground-breaking contributions (*e.g.* [6, 72]). In computer science random CSPs are of fundamental interest as algorithmic benchmarks for computationally hard problems such as graph colouring or k-SAT and as gadgets for cryptographic constructions or reductions in complexity theory (*e.g.* [22, 41, 42, 45, 50]).

Random CSPs also occur as models of disordered systems in statistical physics. Specifically, whereas in classical models such as the Ising model on $\mathbb{Z}^d$ the interactions follow a regular lattice structure, geometries induced by sparse random graphs have been proposed as models of

[†]The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC Grant Agreement 278857–PTCC.
[‡]Supported by a Stiftung Polytechnische Gesellschaft PhD grant.

(spin-)glasses [61]. Over the past 20 years physicists have devised a non-rigorous but analytic technique for the study of these models, called the cavity method. The rigorous vindication of its 'predictions' has emerged as a challenging but fruitful endeavour, in the course of which novel proof techniques have been discovered (*e.g.* the interpolation method [20, 44, 53, 67]).

A fundamental question in the study of random CSPs concerns their *satisfiability thresholds*, which mark the largest density of constraints to variables up to which a solution likely exists. There has been tremendous progress over the past two decades (*e.g.* [5, 6, 8, 29, 36, 35]). But in an important paper [59] physicists predicted the existence of several further phase transitions preceding the satisfiability threshold. At these other transition points the geometry of the solution space and thus, probabilistically speaking, the Boltzmann distribution induced by the CSP instance undergo qualitative changes. These are expected to affect, for example, the performance of algorithms attempting to construct solutions or the mixing times of Markov chains [3, 47, 46, 64].

The most important of these phase transitions is called the *condensation phase transition*. Generally expected to occur at a constraint density within a whisker of the satisfiability threshold, it is thought to mark the onset of extensive long-range correlations. More precisely, for densities below condensation, the correlations between variables that are far apart in the hypergraph induced by the CSP instance are expected to decay. The regime of densities below the condensation phase transition is therefore called the *replica symmetric* phase. By contrast, long-range correlations are deemed to persist beyond the condensation threshold; in physics jargon, replica symmetry is broken. Furthermore, the *reconstruction threshold*, which in most examples occurs at a constraint density well below the condensation threshold, marks the onset of point-to-set correlations where the value assigned to a variable x remains correlated with the values assigned jointly to all the variables at distance ℓ from x even as $\ell \to \infty$. In the physics literature this has been associated with the shattering of the set of solutions into numerous tiny clusters [61, 62].

This paper contributes a systematic rigorous study of the replica symmetric phase for a broad family of random CSPs, for which we prove many of the conjectures from [59]. In particular, we pinpoint the precise condensation phase transition and we establish the absence of long-range correlations below this threshold. Concrete examples of CSPs covered by theses results include the random graph colouring problem, random hypergraph colouring and the random k-NAESAT problem. In all of these specific examples the generic approach developed here enables us to significantly strengthen prior results that were derived via problem-specific arguments.

In terms of techniques, the present paper builds upon [25, 27]. These papers almost exclusively dealt with models with soft constraints only (such as the Potts antiferromagnet), whereas here we extend those methods to the case of hard constraints that strictly forbid certain value combinations (such as graph colouring). While this difference may seem innocuous, the presence of hard constraints causes substantial technical complications. Before stating the main results about general CSPs in Section 2, in the following subsections we present some of their implications for two particularly well-studied examples, the random k-NAESAT problem and the random graph colouring problem.

1.2 Random k-NAESAT

Let $k \geqslant 3$ be an integer and consider the usual model $\mathbb{F}_k(n, m)$ of a random propositional formula over the Boolean variables $x_1, \ldots, x_n$. Thus, $\mathbb{F}_k(n, m)$ is obtained by inserting m independent random clauses of length k such that no variable appears twice in the same clause. We recall that a Boolean assignment σ of $x_1, \ldots, x_n$ is *NAE-satisfying* if, under both σ and its binary inverse $\bar{\sigma}$, all m clauses evaluate to 'true'. Here NAE stands for 'not-all-equal', because every clause must contain at least one literal that evaluates to true as well as at least one that evaluates to false. To parametrize the problem conveniently we will consider formulas with $m = \mathrm{Po}(dn/k)$ clauses for a fixed number $d > 0$. Thus, any variable occurs in d clauses on average. The problem of deciding whether a given k-CNF formula is NAE-satisfiable is NP-complete [71].

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

The random k-NAESAT problem is one of the standard examples of random CSPs and has received a great deal of attention. In particular, in an influential paper Achlioptas and Moore [5] pioneered the use of the second moment method for estimating the partition functions of random CSPs with the example of random k-NAESAT. To be precise, in the case of k-NAESAT the partition function $Z(\mathbb{F}_k(n,m))$ is simply the total number of NAE-satisfying assignments of the random formula. A straightforward first moment calculation shows that with high probability

$$\sqrt[n]{Z(\mathbb{F}_k(n,m))} \leqslant 2(1-2^{1-k})^{d/k+o(1)}. \tag{1.1}$$

Indeed, there are 2^n possible truth assignments. Moreover, the probability that any fixed truth assignment fails to NAE-satisfy one random k-clause is 2^{1-k}, because out of the 2^k possible assignments of k variables precisely two fail to be NAE-satisfying. In particular, (1.1) implies that $\mathbb{F}_k(n,m)$ fails to be NAE-satisfiable w.h.p. if

$$d > k2^{k-1}\ln 2 - k\ln 2/2.$$

The upper bound (1.1) is clearly tight for small densities d. For instance, if $d < 1/(k-1)$ is so small that the random hypergraph induced by $\mathbb{F}_k(n,m)$ does not contain a giant component w.h.p., then $Z(\mathbb{F}_k(n,m)) = \Theta(2^n(1-2^{1-k})^m)$ w.h.p., as is easily verified by counting NAE-solutions of acyclic formulas. But remarkably, Achlioptas and Moore showed via the second moment method that (1.1) remains tight for much larger densities, namely for $d < k2^{k-1}\ln 2 - k(1+\ln 2/2)$. Subsequently Coja-Oghlan and Zdeborová [32] improved this bound slightly and showed that (1.1) continues to be tight as long as

$$d < k2^{k-1}\ln 2 - k\left(\frac{\ln 2}{2} + \frac{1}{4}\right) + \varepsilon_k, \tag{1.2}$$

where ε_k hides an error term that tends to zero in the limit of large k. In fact, up to the precise value of ε_k the bound (1.2) matches the density up to which (1.1) has been predicted to be tight via the cavity method [59]. However, due to the ε_k the expression (1.2) is informative only for (very) large k.

By contrast, the following theorem establishes the *exact* physics prediction for every $k \geqslant 3$. To state the result we introduce $\Lambda(x) = x\ln x$ with the convention that $\Lambda(0) = 0$. Further, $\boldsymbol{\gamma}$ signifies a $\mathrm{Po}(d)$ random variable. Finally, let $\mathcal{P}_*[0,1]$ be the set of all probability measures π on $[0,1]$ with mean $1/2$ and let $(\boldsymbol{\rho}_i^{(\pi)})_{i\geqslant 1} \in [0,1]^\infty$ denote a family of samples from π, mutually independent and independent of $\boldsymbol{\gamma}$.

Theorem 1.1. *For $k \geqslant 3$, $d > 0$ and $\pi \in \mathcal{P}_*[0,1]$, let*

$$\mathcal{B}(d,\pi) = \mathbb{E}\Bigg[\frac{\Lambda(\prod_{i=1}^{\boldsymbol{\gamma}}(1-\prod_{j=1}^{k-1}\boldsymbol{\rho}_{ki+j}^{(\pi)}) + \prod_{i=1}^{\boldsymbol{\gamma}}(1-\prod_{j=1}^{k-1}(1-\boldsymbol{\rho}_{ki+j}^{(\pi)})))}{2(1-2^{1-k})^{\boldsymbol{\gamma}}}$$
$$- \frac{d(k-1)\Lambda(1-\prod_{j=1}^{k}\boldsymbol{\rho}_j^{(\pi)} - \prod_{j=1}^{k}(1-\boldsymbol{\rho}_j^{(\pi)}))}{k(1-2^{1-k})}\Bigg],$$

$$d_{\mathrm{cond}} = \inf\left\{d > 0 : \sup_{\pi\in\mathcal{P}_*[0,1]} \mathcal{B}(d,\pi) > \ln 2 + \frac{d}{k}\ln(1-2^{1-k})\right\}.$$

Then, for all $d < d_{\mathrm{cond}}$,

$$\sqrt[n]{Z(\mathbb{F}_k(n,m))} \xrightarrow{n\to\infty} 2(1-2^{1-k})^{d/k} \quad \textit{in probability.}$$

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

By contrast, for any $d > d_{\mathrm{cond}}$ there exists $\eta > 0$ such that

$$\limsup_{n\to\infty} \mathbb{P}\Big[\sqrt[n]{Z(\mathbb{F}_k(n,m))} > 2(1-2^{1-k})^{d/k} - \eta\Big]^{1/n} < 1. \tag{1.3}$$

Thus, d_{cond} marks the precise threshold up to which (1.1) is tight. Indeed, (1.3) shows that $\sqrt[n]{Z(\mathbb{F}_k(n,m))}$ takes a strictly smaller value with probability $1-\exp(-\Omega(n))$ for $d > d_{\mathrm{cond}}$. Admittedly, the formula for d_{cond}, involving an optimization problem over a probability measure on the unit interval, is not explicit, and it is potentially difficult to evaluate. But given the combinatorial intricacy of the (NP-hard) k-NAESAT problem we may not be entitled to a simple answer. More generally, the physics predictions typically take the form of distributional optimization problems. Yet it also seems clear that elementary techniques such as the combinatorial second moment method will hardly suffice to establish such predictions precisely.

Theorem 2.5 shows that d_{cond} is a genuine phase transition, called the condensation phase transition, since the functions $d \mapsto \mathbb{E}\sqrt[n]{Z(\mathbb{F}_k(n,m))}$ fail to converge to an analytic limit at the point d_{cond}. Indeed, the theorem implies that the limit exists and matches the entire function $2(1-2^{1-k})^{d/k}$ for $d < d_{\mathrm{cond}}$. By contrast, for $d > d_{\mathrm{cond}}$ the limit may not exist, and even if it does it is strictly smaller than $2(1-2^{1-k})^{d/k}$.

Additionally, up to d_{cond} there occurs an important decay of correlation phenomenon. Formally, let $\boldsymbol{\sigma}, \boldsymbol{\tau}$ signify two independently chosen random NAE-satisfying assignments of $\mathbb{F}_k(n,m)$ (given that the formula is NAE-satisfiable). Representing the Boolean values false and true by ± 1, we think of $\boldsymbol{\sigma}, \boldsymbol{\tau}$ as vectors in $\{\pm 1\}^n$. Let us denote the expectation with respect to the choice of $\boldsymbol{\sigma}, \boldsymbol{\tau}$ given the random formula $\mathbb{F}_k(n,m)$ by $\langle\cdot\rangle_{\mathbb{F}_k(n,m)}$, whereas we use the standard symbols $\mathbb{E}[\cdot]$, $\mathbb{P}[\cdot]$ to refer to the choice of $\mathbb{F}_k(n,m)$ itself. The second moment argument of Achlioptas and Moore [5] was based on showing by elementary calculations that for $d/k < 2^{k-1}\ln 2 - (1+\ln 2/2)$, the vectors $\boldsymbol{\sigma}, \boldsymbol{\tau}$ are nearly perpendicular w.h.p. Formally, their inner product satisfies $\boldsymbol{\sigma}\cdot\boldsymbol{\tau} = o(n)$ w.h.p. According to the cavity method, this property should extend right up to the condensation threshold d_{cond}. The following theorem verifies this conjecture.

Theorem 1.2. *Let $k \geqslant 3$. For all $0 < d < d_{\mathrm{cond}}$ we have*

$$\lim_{n\to\infty} \frac{1}{n}\mathbb{E}[\langle|\boldsymbol{\sigma}\cdot\boldsymbol{\tau}|\rangle_{\mathbb{F}_k(n,m)} \mid Z(\mathbb{F}_k(n,m)) > 0] = 0. \tag{1.4}$$

Due to standard results about probability measures on the cube $\{\pm 1\}^n$ we can express (1.4) in terms of pairwise correlations between the truth values assigned to variables [17]. Specifically, (1.4) is equivalent to the statement

$$\lim_{n\to\infty} \frac{1}{n^2}\sum_{i,j=1}^{n} \mathbb{E}[|\langle\boldsymbol{\sigma}(x_i)\cdot\boldsymbol{\sigma}(x_j)\rangle_{\mathbb{F}_k(n,m)}| \mid Z(\mathbb{F}_k(n,m)) > 0] = 0. \tag{1.5}$$

Hence, for $d < d_{\mathrm{cond}}$ the truth values $\boldsymbol{\sigma}(x_i), \boldsymbol{\sigma}(x_j)$ assigned to two randomly chosen variables x_i, x_j are asymptotically independent. Physics calculations predict that neither (1.4) nor (1.5) continue to hold for $d > d_{\mathrm{cond}}$.

Finally, let us refer to

$$d_{\mathrm{sat}} = \inf\Big\{d > 0 : \liminf_{n\to\infty} \mathbb{P}[Z(\mathbb{F}_k(n,m)) > 0] < 1\Big\}$$

as the *satisfiability threshold* of the random k-NAESAT problem. Coja-Oghlan and Panagiotou [28] determined the asymptotic value of d_{sat}, showing that

$$d_{\mathrm{sat}} = k2^{k-1}\ln 2 - k\left(\frac{\ln 2}{2} + \frac{1}{4}\right) + \varepsilon_k, \quad \text{where } \varepsilon_k \to 0 \text{ as } k \to \infty. \tag{1.6}$$

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

While (1.6) is asymptotically tight in the limit of large k, the condensation threshold d_{cond} from Theorem 1.1 yields a lower bound on d_{sat} for *every* $k \geqslant 3$. This is the best current lower bound for any specific k.

1.3 Random graph colouring

Let $\mathbb{G} = \mathbb{G}(n,p)$ denote the random graph on n vertices $\{1,\ldots,n\}$ where each of the $\binom{n}{2}$ possible edges is present with probability p independently. If we set $p = d/n$ for a fixed $d > 0$ and a large n, then the average degree of the random graph will be asymptotically equal to d. Let $q \geqslant 3$ be a number of colours and let $Z_q(\mathbb{G}(n,p))$ be the number of q-colourings of the random graph. Understanding the random variable $Z_q(\mathbb{G}(n,p))$ for given d, q is one of the longest-standing challenges in the theory of random graphs. In fact, the problem of identifying the *q-colourability threshold*, *i.e.* the largest value of d up to which $Z_q(\mathbb{G}(n,p)) > 0$ w.h.p., goes back to the seminal paper of Erdős and Rényi [40].

Like in the random k-NAESAT problem it is easy to determine the number of q-colourings for $d < 1$, where the there is no giant component yet. In this regime it is easily verified that

$$\sqrt[n]{Z_q(\mathbb{G}(n,p))} \xrightarrow{n\to\infty} q(1-1/q)^{d/2} \quad \text{in probability.} \tag{1.7}$$

In [27] the largest average degree d_{cond} up to which this convergence in probability occurs was determined. The precise formula involves a stochastic optimization problem akin to the one in Theorem 1.1. Asymptotically in the limit of large q we have $d_{\mathrm{cond}} = (2q-1)\ln q - 2\ln 2 + \varepsilon_q$. By comparison, for $d > (2q-1)\ln q - 1 + \varepsilon_q$ the random graph fails to be q-colourable w.h.p. [24].

Equation (1.7) gives a 'first-order' approximation to $Z_q(\mathbb{G}(n,p))$ up to errors of size $\exp(o(n))$. But how large might the fluctuations of $Z_q(\mathbb{G}(n,p))$ be? Clearly, adding, removing or rewiring a single edge is apt to change $Z_q(\mathbb{G}(n,p))$ by a constant factor (or even more). Consequently, since key variables such as the number of vertices and edges in the giant component have fluctuations of order $\Theta(\sqrt{n})$ even once we condition on the total number m of edges, one might expect $Z_q(\mathbb{G}(n,p))$ to exhibit multiplicative fluctuations of order at least $\exp(\Theta(\sqrt{n}))$. However, Bapst, Coja-Oghlan and Efthymiou [18] proved that for q exceeding a certain (undetermined but large) constant q_0 the random variable $Z_q(\mathbb{G}(n,p))$ is concentrated remarkably tightly for all $d < d_{\mathrm{cond}}$. More specifically, $Z_q(\mathbb{G}(n,p))$ has *bounded* multiplicative fluctuations once we condition on the number m of edges of the random graph. In fact, Raßmann [69] determined the precise limiting distribution of $Z_q(\mathbb{G}(n,p))$ given m for all $d < d_{\mathrm{cond}}$ under the assumption that $q > q_0$ is sufficiently large. As an application of our general results we obtain the limiting distribution of $\ln Z_q(\mathbb{G}(n,p))$ for $d < d_{\mathrm{cond}}$ for all $q \geqslant 3$, thereby closing the gap left by [18, 69].

Theorem 1.3. *Let $q \geqslant 3$ and $0 < d < d_{\mathrm{cond}}(q)$. With $(K_\ell)_{\ell\geqslant 3}$ a sequence of independent Poisson variables with means $\mathbb{E}[K_\ell] = d^\ell/(2\ell)$, let*

$$\mathcal{K} = \prod_{\ell=3}^{\infty} (1+\delta_\ell)^{K_\ell} \exp\left(-\frac{d^\ell \delta_\ell}{2\ell}\right), \quad \text{where } \delta_\ell = -(1-q)^{1-\ell}.$$

Then $\mathcal{K} > 0$ almost surely, and we have the following convergence in distribution:

$$\frac{Z_q(\mathbb{G}(n,p))}{q^n(1-1/q)^m} \xrightarrow{n\to\infty} \sqrt{q}\left(1+\frac{d}{q-1}\right)^{(1-q)/2} \exp\left(-\frac{d\delta_1}{2} - \frac{d^2\delta_2}{4}\right)\mathcal{K}.$$

As an application of Theorem 1.3 we obtain a result that characterizes the combinatorial structure of typical q-colourings of the random graph for all $d < d_{\mathrm{cond}}$ very accurately. A similar result was obtained previously in [26], but required the extraneous assumption that $q > q_0$ for some very large constant q_0. To formulate the result, let $\nabla_\ell(\mathbb{G}, v)$ denote the subgraph of $\mathbb{G}$ induced on the set

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

of vertices at distance at most ℓ from vertex v. For a fixed ℓ and large n this subgraph is a tree w.h.p. Furthermore, let $\mu_{\mathbb{G},\nabla_\ell(\mathbb{G},v)}$ denote the distribution on the set of q-colourings of $\nabla_\ell(\mathbb{G}, v)$ induced by a uniformly random q-colouring of the entire graph. For comparison, let $\mu_{\nabla_\ell(\mathbb{G},v)}$ be the uniform distribution on the set of all q-colourings of the subgraph $\nabla_\ell(\mathbb{G}, v)$ only. Clearly, *a priori* $\mu_{\mathbb{G},\nabla_\ell(\mathbb{G},v)}$ and $\mu_{\nabla_\ell(\mathbb{G},v)}$ could be quite different because the latter ignores the 'external' connections of the boundary vertices at distance ℓ from v via (long) paths through $\mathbb{G} - \nabla_\ell(\mathbb{G}, v)$. Yet the next theorem shows that for almost all vertices v the two distributions asymptotically coincide.

Theorem 1.4. *Let $q \geqslant 3$, $0 < d < d_{\mathrm{cond}}(q)$ and $\ell \geqslant 1$. Then*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{v=1}^{n} \mathbb{E}[d_{\mathrm{TV}}(\mu_{\mathbb{G},\nabla_\ell(\mathbb{G},v)}, \mu_{\nabla_\ell(\mathbb{G},v)})] = 0.$$

As an application of Theorem 1.4 we obtain a further result about the reconstruction problem. We will give a precise definition in Section 2 below, but intuitively reconstruction occurs when the colour of the vertex v remains correlated with the colours assigned to *all* the boundary vertices at distance precisely ℓ from v even for large values of ℓ. A well-known conjecture from [59] asserts that the threshold for reconstruction on the random graph coincides with the reconstruction threshold on the Galton–Watson tree that mimics the local structure of the random graph. Previously this was confirmed only under the assumption that q be large enough [18, 48, 64].

2. Main results

2.1 Random constraint satisfaction problems

In this section we present the main results of the paper for a general family of random CSPs. To set the stage we introduce a comprehensive model of random CSPs. The variables take values in a finite domain $\Omega \neq \emptyset$. They are bound by constraints that each involve precisely $k \geqslant 2$ variables and either discourage or outright forbid certain value combinations. The formal definition reads as follows.

Definition 2.1. Let $\Omega \neq \emptyset$ be a finite set and let Ψ be a finite set of functions $\Omega^k \to [0, 1]$. A *Ψ-constraint satisfaction problem* $G = (V, F, (\partial a)_{a\in F}, (\psi_a)_{a\in F})$ comprises

- a set V of *variables*,
- a set F of *constraints*,
- an ordered k-tuple $\partial a = (\partial_1 a, \ldots, \partial_k a) \in V^k$ for each $a \in F$ and
- a *constraint function* $\psi_a \in \Psi$ for each $a \in F$.

An assignment $\sigma \in \Omega^V$ *satisfies* G if $\psi_a(\sigma(\partial_1 a), \ldots, \sigma(\partial_k a)) > 0$ for all $a \in F$; in symbols, $\sigma \models G$.

A Ψ-CSP G induces a bipartite graph with vertex sets V and F where $a \in F$ is adjacent to $\partial_1 a, \ldots, \partial_k a$. We will therefore use graph-theoretic terminology and, for example, refer to $\partial_1 a, \ldots, \partial_k a$ as the neighbours of a. Moreover, the length of shortest paths in the bipartite graph induces a metric on the nodes of G.

For a Ψ-CSP G and an assignment $\sigma \in \Omega^V$, we let

$$\psi_G(\sigma) = \prod_{a\in F} \psi_a(\sigma(\partial_1 a), \ldots, \sigma(\partial_k a)).$$

Moreover, we introduce the *partition function* $Z(G) = \sum_{\sigma\in\Omega^V} \psi_G(\sigma)$ as well as the *Boltzmann distribution*

$$\mu_G(\sigma) = \psi_G(\sigma)/Z(G) \quad (\sigma \in \Omega^V),$$

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

providing that $Z(G) > 0$. Further, we let $S(G) = \{\sigma \in \Omega^V : \sigma \models G\}$ be the set of satisfying assignments. In many cases the functions $\psi \in \Psi$ are $\{0, 1\}$-valued. Then $Z(G) = |S(G)|$ is just the number of solutions. But as we will see in Section 3 there are interesting cases where the functions ψ take values strictly between 0 and 1.

Standard examples of CSPs fit the framework provided by Definition 2.1.

Example 2.2 (**hypergraph colouring**). Suppose that $k \geqslant 2$ is an integer, that $q \geqslant 2$ is a number of colours and that $g = (V, E)$ is a k-uniform hypergraph. Recall that a q-colouring of g is a map $\sigma : V \to \Omega = \{1, \ldots, q\}$ such that for every edge $e \in E$ there exist $v, w \in e$ with $\sigma(v) \neq \sigma(w)$ (*i.e.* no edge is monochromatic). Let $\Psi_{k,q} = \{\psi_{k,q}\}$ be the singleton containing the function

$$\psi_{k,q} : \Omega^k \to \{0, 1\}, \quad \sigma \mapsto 1 - \mathbf{1}\{\sigma_1 = \cdots = \sigma_k\}.$$

Then we can express the q-colourability problem on g as a $\Psi_{k,q}$-CSP G whose variables are the vertices V and whose constraints are the edges E of g. For each edge e the k-tuple ∂e simply contains the vertices incident with e in g (in any order) and $\psi_e = \psi_{k,q}$. Of course, the case $k = 2$ corresponds to the classical graph colouring problem.

Example 2.3 (***k*-NAESAT).** Suppose that $k \geqslant 2$ is an integer and that $g = a_1 \wedge \cdots \wedge a_m$ is a propositional formula over a set $V = \{x_1, \ldots, x_n\}$ of Boolean variables with clauses $a_1, \ldots, a_m$, each containing precisely k literals. Let $\Omega = \{-1, 1\}$ represent the Boolean values 'true' and 'false' and recall that an assignment $\sigma \in \Omega^V$ is *NAE-satisfying* for g if the expression evaluates to 'true' under both σ and its binary inverse $-\sigma$. This problem can be expressed as a CSP over the set $\Psi_{k-\mathrm{NAE}}$ containing the 2^k constraint functions

$$\psi_\tau : \Omega^k \to \{0, 1\}, \quad \sigma \mapsto 1 - \mathbf{1}\{\sigma = \tau\} - \mathbf{1}\{\sigma = -\tau\} \quad (\tau \in \Omega^k).$$

Indeed, we turn g into a $\Psi_{k-\mathrm{NAE}}$-CSP with variables V and constraints $F = \{a_1, \ldots, a_m\}$. We let ∂a_i be the k-tuple of variables occurring in the clause a_i. Moreover, letting $\tau_{i,j} = 1$ if the jth literal of a_i is negated and $\tau_{i,j} = -1$ otherwise, we let $\psi_{a_i} = \psi_{\tau_{i,1},\ldots,\tau_{i,k}}$.

We consider the following 'Erdős–Rényi-like' model of random CSP instances.

Definition 2.4. Suppose that Ψ is a finite set of functions $\Omega^k \to [0, 1]$ and that P is a probability distribution on Ψ. Then $\mathbb{G}(n, m, P)$ is the random Ψ-CSP with variables $V_n = \{x_1, \ldots, x_n\}$ and constraints $F_m = \{a_1, \ldots, a_m\}$ such that

- $\partial a_1, \ldots, \partial a_m \in V_n^k$ are chosen uniformly from the set of all $n(n-1)\cdots(n-k+1)$ tuples consisting of pairwise distinct variables, requiring the k-sets $(\{\partial_1 a_i, \ldots, \partial_k a_i\})_{i \leqslant m}$ to be pairwise distinct,
- the constraint functions $\psi_{a_1}, \ldots, \psi_{a_m} \in \Psi$ are chosen independently from the distribution P.

Thus, the constraints $a_1, \ldots, a_m$ are chosen nearly independently. The only condition is that the hypergraph induced on V_n with edges $\{\{\partial_1 a_i, \ldots, \partial_k a_i\} : i = 1, \ldots, m\}$ be k-uniform and simple. This condition is necessary to accommodate interesting examples such as the random graph colouring problem.

The main results of this paper apply to all CSPs that satisfy a few (relatively) easy-to-check assumptions. These come solely in terms of the distribution P on Ψ. Throughout the paper we always let $\boldsymbol{\psi}$ denote an element of Ψ drawn from P. Moreover, we let

$$q = |\Omega|, \quad \xi = q^{-k} \sum_{\sigma \in \Omega^k} \mathbb{E}[\boldsymbol{\psi}(\sigma)].$$

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

Furthermore, for $\psi : \Omega^k \to [0,1]$ and a permutation θ of $\{1,\ldots,k\}$ we let

$$\psi^\theta : \Omega^k \to [0,1], \quad \sigma \mapsto \psi(\sigma_{\theta(1)},\ldots,\sigma_{\theta(k)})$$

denote the function obtained by permuting the coordinates according to θ. *From here on we tacitly assume that the set Ψ is closed under permutations, that is, for every $\psi \in \Psi$ we have $\psi^\theta \in \Psi$. Moreover, we always assume that $P(\psi) > 0$ for all $\psi \in \Psi$ and that*

$$\min_{\psi\in\Psi,\sigma\in\Omega^k} \psi(\sigma) < \max_{\psi\in\Psi,\sigma\in\Omega^k} \psi(\sigma). \tag{2.1}$$

Let us write $\mathcal{P}(\Omega)$ for the set of all probability distributions on Ω. We identify $\mathcal{P}(\Omega)$ with the standard simplex in $\mathbb{R}^\Omega$. Moreover, we let $\mathcal{P}^2_*(\Omega)$ be the set of all probability distributions π on $\mathcal{P}(\Omega)$ such that $\int_{\mathcal{P}(\Omega)} \mu(\omega)\mathrm{d}\pi(\mu) = 1/q$ for all $\omega \in \Omega$. With these conventions the assumptions on P read as follows.

SYM. For all $i \in \{1,\ldots,k\}$, $\omega \in \Omega$ and $\psi \in \Psi$ we have

$$\sum_{\tau\in\Omega^k} \mathbf{1}\{\tau_i = \omega\}\psi(\tau) = q^{k-1}\xi$$

and for every permutation θ and every $\psi \in \Psi$ we have $P(\psi) = P(\psi^\theta)$.

BAL. The function

$$\phi : \mu \in \mathcal{P}(\Omega) \mapsto \sum_{\tau\in\Omega^k} \mathbb{E}[\boldsymbol{\psi}(\tau)] \prod_{i=1}^k \mu(\tau_i)$$

is concave and attains its maximum at the uniform distribution on Ω.

MIN. Let $\mathcal{R}(\Omega)$ be the set of all probability distribution $\rho = (\rho(s,t))_{s,t\in\Omega}$ on $\Omega\times\Omega$ such that

$$\sum_{s\in\Omega} \rho(s,t) = \sum_{s\in\Omega} \rho(t,s) = q^{-1} \quad \text{for all } t \in \Omega.$$

The function

$$\varphi : \rho \in \mathcal{R}(\Omega) \mapsto \sum_{\sigma,\tau\in\Omega^k} \mathbb{E}[\boldsymbol{\psi}(\sigma)\boldsymbol{\psi}(\tau)] \prod_{i=1}^k \rho(\sigma_i,\tau_i)$$

has the uniform distribution on $\Omega\times\Omega$ as its unique global minimizer.

POS. For all $\pi,\pi' \in \mathcal{P}^2_*(\Omega)$ the following is true. With $\boldsymbol{\rho}_1,\boldsymbol{\rho}_2,\ldots$ chosen from π, $\boldsymbol{\rho}'_1,\boldsymbol{\rho}'_2,\ldots$ chosen from π' and $\boldsymbol{\psi} \in \Psi$ chosen from P, all mutually independent, we have for every $\ell \geqslant 2$,

$$\mathbb{E}\Bigg[\bigg(1 - \sum_{\tau\in\Omega^k} \boldsymbol{\psi}(\tau)\prod_{i=1}^k \boldsymbol{\rho}_i(\tau_i)\bigg)^\ell + (k-1)\bigg(1 - \sum_{\tau\in\Omega^k} \boldsymbol{\psi}(\tau)\prod_{i=1}^k \boldsymbol{\rho}'_i(\tau_i)\bigg)^\ell$$
$$- k\bigg(1 - \sum_{\tau\in\Omega^k} \boldsymbol{\psi}(\tau)\boldsymbol{\rho}_1(\tau_1)\prod_{i=2}^k \boldsymbol{\rho}'_i(\tau_i)\bigg)^\ell\Bigg] \geqslant 0.$$

UNI. If G is a Ψ-CSP such that for every constraint a the variables $\partial_1 a,\ldots,\partial_k a$ are pairwise distinct and the bipartite graph induced by G is unicyclic, then G has a satisfying assignment.

Conditions **SYM** and **BAL** are symmetry assumptions. Specifically, **SYM** requires that no constraint exhibits an inherent 'preference' for any of the values $\omega \in \Omega$ if the values of the other variables are random. **BAL** is going to ensure that in a typical solution σ to a random CSP there

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

are about n/q variables that take each value $\omega \in \Omega$. Assumptions **MIN** and **POS** impose convexity conditions that are required for technical reasons. Finally, **UNI** is going to ensure that in the regime of constraint densities that we study, the probability of being satisfiable is either $1 - o(1)$ or $o(1)$. (In particular, the condition rules out the random graph 2-colouring problem.) Conditions **SYM**–**POS** occurred in earlier work on problems with soft constraints [25, 27].

Crucially, the above conditions *only* refer to the distribution P on the set Ψ of weight functions. They are usually (relatively) easy to check. Indeed, in Section 3 we will verify the conditions for several well-known examples. Not all of our results require all of the assumptions, and we shall always indicate in brackets which ones are needed.

2.2 The condensation phase transition

In order to state the main theorems in a unified way, we let m be a random variable with distribution $\mathrm{Po}(dn/k)$ and we introduce $\mathbb{G} = \mathbb{G}(n, m, P)$. In this way we are left with just the single parameter d. As in the examples in Section 1 we can easily calculate $Z(\mathbb{G})$ for small values of d. For instance, for $d < 1/(k-1)$ the bipartite graph induced by the random CSP does not feature a giant component. Therefore, **SYM** implies that $Z(\mathbb{G}) = q^n \xi^{m+o(n)}$ w.h.p. The following theorem determines the precise threshold up to which this identity holds, the *condensation threshold*. Recall that $\Lambda(x) = x \ln x$.

Theorem 2.5 (**SYM**, **BAL**, **MIN**, **UNI**). *Let $d > 0$. With $\boldsymbol{\gamma}$ a $\mathrm{Po}(d)$-random variable, $\boldsymbol{\rho}_1^{(\pi)}, \boldsymbol{\rho}_2^{(\pi)}, \ldots$ chosen from $\pi \in \mathcal{P}_*^2(\Omega)$ and $\boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \ldots \in \Psi$ chosen from P, all mutually independent, let*

$$\mathcal{B}(d, P, \pi) = \mathbb{E}\Bigg[q^{-1}\xi^{-\boldsymbol{\gamma}}\Lambda\Bigg(\sum_{\sigma\in\Omega}\prod_{i=1}^{\boldsymbol{\gamma}}\sum_{\tau\in\Omega^k}\mathbf{1}\{\tau_k = \sigma\}\boldsymbol{\psi}_i(\tau)\prod_{j=1}^{k-1}\boldsymbol{\rho}_{ki+j}^{(\pi)}(\tau_j)\Bigg) - \frac{d(k-1)}{k\xi}\Lambda\Bigg(\sum_{\tau\in\Omega^k}\boldsymbol{\psi}_1(\tau)\prod_{j=1}^{k}\boldsymbol{\rho}_j^{(\pi)}(\tau_j)\Bigg)\Bigg], \tag{2.2}$$

$$d_{\mathrm{cond}} = \inf\Bigg\{d > 0 : \sup_{\pi\in\mathcal{P}_*^2(\Omega)}\mathcal{B}(d, P, \pi) > \ln q + \frac{d}{k}\ln\xi\Bigg\}. \tag{2.3}$$

Then for all $d < d_{\mathrm{cond}}$ we have

$$\sqrt[n]{Z(\mathbb{G})} \xrightarrow{n\to\infty} q\xi^{d/k} \quad \text{in probability.} \tag{2.4}$$

By contrast, if P also satisfies **POS**, *then for any $d > d_{\mathrm{cond}}$ there exists $\varepsilon > 0$ such that*

$$\limsup_{n\to\infty}\mathbb{P}[\sqrt[n]{Z(\mathbb{G})} > q\xi^{d/k} - \varepsilon]^{1/n} < 1 - \varepsilon. \tag{2.5}$$

The functional $\mathcal{B}$ that is used to detect the condensation phase transition is a compact version of the so-called 'Bethe free entropy' from physics [61]. Generally, the Bethe free entropy provides a formula for the free energy of tree factor graphs in terms of local quantities, *i.e.* marginal distributions of variable and factor nodes. Alternatively, these marginal distributions can be described as fixed points of the Belief Propagation message passing scheme, and so equivalently, the Bethe free energy expresses the free energy of a finite tree factor graph in terms of the unique Belief Propagation fixed point. The intuitive explanation of the appearance of $\mathcal{B}$ in Theorem 2.5 is that asymptotically, up to a constraint density $d < d_{\mathrm{cond}}$, this line of computation remains true, when applied to the random (and possibly infinite) Galton–Watson tree that describes the local structure of the random graph.

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

Thus, for $d < d_{\mathrm{cond}}$ we have $Z(\mathbb{G}) = q^{n+o(n)}\xi^m$ with high probability. By contrast, $Z(\mathbb{G})$ is exponentially smaller than this expression for $d > d_{\mathrm{cond}}$. To be precise, $Z(\mathbb{G}) \leqslant q^{n-\Omega(n)}\xi^m$ with probability $1 - \exp(-\Omega(n))$ for $d > d_{\mathrm{cond}}$. Consequently, since $d \mapsto q\xi^{d/k}$ is an entire function, Theorem 2.5 shows that $\mathbb{E}\sqrt[n]{Z(\mathbb{G})}$, viewed as a function of d, fails to converge to an analytic limit at d_{cond} as $n \to \infty$. Therefore, d_{cond} marks a genuine phase transition.

Further, let us call

$$d_{\mathrm{sat}} = \inf\Big\{d > 0 : \liminf_{n\to\infty} \mathbb{P}[Z(\mathbb{G}) > 0] < 1\Big\}$$

the *satisfiability threshold* of the random CSP. Since (2.1) guarantees that $\xi > 0$, we have $q\xi^{d/k} > 0$ for all $d > 0$. Hence, (2.4) shows that $Z(\mathbb{G}) > 0$ w.h.p. for all $d < d_{\mathrm{cond}}$. In effect,

$$d_{\mathrm{cond}} \leqslant d_{\mathrm{sat}}. \tag{2.6}$$

Most of the prior contributions on lower-bounding satisfiability thresholds of various CSPs via the second moment method (*e.g.* [5, 6, 13, 39]) actually lower-bound the condensation threshold. To be precise, suppose that for some $d > 0$ the second moment bound

$$\mathbb{E}[Z(\mathbb{G})^2 \mid m] \leqslant O(\mathbb{E}[Z(\mathbb{G}) \mid m]^2)$$

holds with high probability over the choice of m. (For second moment calculations it is vital to condition on m.) Then the Paley–Zygmund inequality shows that there exists a constant $\delta > 0$ such that w.h.p. over the choice of m,

$$\mathbb{P}[Z(\mathbb{G}) \geqslant \delta q\xi^m \mid m] \geqslant \Omega(1).$$

Hence, (2.5) implies that $d \leqslant d_{\mathrm{cond}}$. In fact, in most examples of random CSPs (2.6) is strictly better than any previously known lower bound on the satisfiability threshold.

2.3 The Kesten–Stigum bound

While exact, the formula for d_{cond} from Theorem 2.5 may not be easy to evaluate. However, there is an important upper bound that is. For a function $\psi \in \Psi$, let $\Phi_\psi \in \mathbb{R}^{\Omega\times\Omega}$ be the matrix with entries

$$\Phi_\psi(\omega,\omega') = q^{1-k}\xi^{-1}\sum_{\tau\in\Omega^k} \mathbf{1}\{\tau_1 = \omega, \tau_2 = \omega'\}\psi(\tau) \quad (\omega,\omega' \in \Omega). \tag{2.7}$$

Further, let Ξ be the linear operator on the q^2-dimensional space $\mathbb{R}^\Omega \otimes \mathbb{R}^\Omega$ defined by

$$\Xi = \mathbb{E}[\Phi_{\boldsymbol{\psi}} \otimes \Phi_{\boldsymbol{\psi}}]. \tag{2.8}$$

Additionally, with $\mathbf{1}$ denoting the vector with all entries equal to one, let

$$\begin{aligned} \mathcal{E} &= \{z \in \mathbb{R}^q \otimes \mathbb{R}^q : \forall y \in \mathbb{R}^q : \langle z, \mathbf{1}\otimes y\rangle = \langle z, y\otimes\mathbf{1}\rangle = 0\} \quad \text{and} \\ d_{\mathrm{KS}} &= \Big((k-1)\max_{x\in\mathcal{E}\,:\,\|x\|=1}\langle \Xi x, x\rangle\Big)^{-1}, \end{aligned} \tag{2.9}$$

with the convention that $d_{\mathrm{KS}} = \infty$ if $\max_{x\in\mathcal{E}\,:\,\|x\|=1}\langle \Xi x, x\rangle = 0$.

Theorem 2.6 (**SYM**, **BAL**). *We have* $d_{\mathrm{cond}} \leqslant d_{\mathrm{KS}}$.

In the case of the random graph q-colouring problem (see Section 1.3 and Example 2.2) we calculate $d_{\mathrm{KS}} = (q-1)^2$. This expression matches the *Kesten–Stigum bound* which plays a role in broadcasting processes on random trees [56]. Moreover, for the graph colouring problem it was shown in [27] that $d_{\mathrm{cond}} \leqslant (q-1)^2$. Thus Theorem 2.6 extends the Kesten–Stigum bound

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

to general CSPs and shows that it always gives an upper bound on the condensation threshold. While the Kesten–Stigum bound is conjectured to be tight in a few cases (such as random graph 3-colouring), the bound fails to be tight in others (such as random graph 5-colouring) [73]. Generally the tightness of the Kesten–Stigum bound has implications for algorithmic problems, a point we discuss further below.

2.4 The number of solutions

Theorem 2.5 determines the leading exponential order of the partition function for $d < d_{\mathrm{cond}}$. The following theorem, which is the main result of the paper, takes a closer look and determines the precise limiting distribution of $Z(\mathbb{G})$ for $d < d_{\mathrm{cond}}$. Let

$$\Phi = \mathbb{E}[\Phi_{\boldsymbol{\psi}}] \in \mathbb{R}^{\Omega\times\Omega} \tag{2.10}$$

and let $\mathrm{Eig}(\Phi)$ be the multiset that contains the eigenvalues of Φ according to their geometric multiplicities.

Theorem 2.7 (**SYM**, **BAL**, **MIN**, **UNI**). *Suppose that $0 < d < d_{\mathrm{cond}}$. Let $(K_\ell)_{\ell\geqslant 1}$ be Poisson variables with means*

$$\mathbb{E}[K_\ell] = \frac{1}{2\ell}(d(k-1))^\ell$$

and let $(\boldsymbol{\psi}_{\ell,i,j})_{\ell,i,j\geqslant 1}$ be a sequence of samples from P, all mutually independent. Then

$$\mathcal{K} = \exp\left(\frac{d(k-1)(1-\mathrm{tr}(\Phi))}{2} + \mathbf{1}\{k=2\}\frac{d^2(1-\mathrm{tr}(\Phi^2))}{4}\right) \times \prod_{\ell=2+\mathbf{1}\{k=2\}}^{\infty}\exp\left(\frac{(d(k-1))^\ell}{2\ell}(1-\mathrm{tr}(\Phi^\ell))\right)\prod_{i=1}^{K_\ell}\mathrm{tr}\prod_{j=1}^{\ell}\Phi_{\boldsymbol{\psi}_{\ell,i,j}}$$

satisfies $\mathcal{K} > 0$ almost surely. Moreover, $\mathrm{Eig}(\Phi) \subset (-\infty, 0] \cup \{1\}$ *and*

$$\frac{Z(\mathbb{G})}{q^{n+1/2}\xi^m}\prod_{\lambda\in\mathrm{Eig}(\Phi)\setminus\{1\}}\sqrt{1-d(k-1)\lambda} \xrightarrow{n\to\infty} \mathcal{K} \tag{2.11}$$

in distribution.

Thus, Theorem 2.7 shows that $Z(\mathbb{G})$ is remarkably concentrated for $d < d_{\mathrm{cond}}$. Indeed, while one might *a priori* expect that fluctuations of variables such as the order and size of the giant component of $\mathbb{G}$ have a significant knock-on effect on $Z(\mathbb{G})$ and cause multiplicative fluctuations of order at least $\exp(\Omega(\sqrt{n}))$, Theorem 2.7 shows that $Z(\mathbb{G})$ merely has bounded multiplicative fluctuations. We are not aware of a general physics prediction as to the limiting distribution of the partition function of random CSPs, although there is a paper on the diluted version of the Sherrington–Kirkpatrick model [43] (which does not have hard constraints).

2.5 The overlap

One of the main predictions of the physics paper [59] is that for densities $d < d_{\mathrm{cond}}$ the Boltzmann distribution $\mu_{\mathbb{G}}$ does not exhibit extensive long-range correlations. The next theorem verifies this conjecture. Define the *overlap* of assignments $\sigma, \tau \in \Omega^{V_n}$ as the $\Omega\times\Omega$-matrix $\rho_{\sigma,\tau} = (\rho_{\sigma,\tau}(\omega,\omega'))_{s,t\in\Omega}$ with

$$\rho_{\sigma,\tau}(\omega,\omega') = |\sigma^{-1}(\omega)\cap\tau^{-1}(\omega')|/n.$$

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

Since $\sum_{\omega,\omega'} \rho_{\sigma,\tau}(\omega,\omega') = 1$, we can view $\rho_{\sigma,\tau}$ as a probability distribution on $\Omega \times \Omega$, namely the empirical distribution of the value combinations $(\sigma(x_i), \tau(x_i))_{i=1,\ldots,n}$. Let $\bar{\rho}$ be the uniform distribution on $\Omega \times \Omega$. Moreover, write $\boldsymbol{\sigma}, \boldsymbol{\tau}$ for two independent samples chosen from $\mu_{\mathbb{G}}$, $\langle \cdot \rangle_{\mathbb{G}}$ for the expectation with respect to $\boldsymbol{\sigma}, \boldsymbol{\tau}$ and $\mathbb{E}[\cdot]$ for the expectation with respect to the choice of $\mathbb{G}$.

Theorem 2.8 (**SYM**, **BAL**, **MIN**, **UNI**). *For all* $0 < d < d_{\mathrm{cond}}$ *we have*

$$\lim_{n\to\infty} \mathbb{E}[\langle \|\rho_{\boldsymbol{\sigma},\boldsymbol{\tau}} - \bar{\rho}\|_{\mathrm{TV}} \rangle_{\mathbb{G}} \mid Z(\mathbb{G}) > 0] = 0. \tag{2.12}$$

For $d < d_{\mathrm{cond}}$ the event $Z(\mathbb{G}) > 0$ occurs w.h.p. due to (2.6).

Theorem 2.8 shows that for $d < d_{\mathrm{cond}}$ the overlap of two random satisfying assignments $\boldsymbol{\sigma}, \boldsymbol{\tau}$ is roughly uniform, that is, there is no extensive correlation between $\boldsymbol{\sigma}, \boldsymbol{\tau}$. Using the general results from [17] regarding probability measures on discrete cubes, we can express this result in terms of pairwise correlations between variables. Specifically, for $1 \leqslant i < j \leqslant n$ let $\mu_{\mathbb{G},x_i,x_j}$ be the joint distribution of the values $\boldsymbol{\sigma}(x_i), \boldsymbol{\sigma}(x_j)$. Thus, $\mu_{\mathbb{G},x_i,x_j}$ is a probability distribution on $\Omega \times \Omega$. Then (2.12) can be rephrased equivalently as

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{1\leqslant i<j\leqslant n} \mathbb{E}[\|\mu_{\mathbb{G},x_i,x_j} - \bar{\rho}\|_{\mathrm{TV}} \mid Z(\mathbb{G}) > 0] = 0 \tag{2.13}$$

(see Appendix A for a proof). In other words, for most pairs i, j the values $\boldsymbol{\sigma}(x_i), \boldsymbol{\sigma}(x_j)$ are asymptotically independent. Equation (2.13) matches the precise definition of 'static replica symmetry' from [59, 61].

2.6 Local weak convergence

Since the expected distance between two uniform variables of $\mathbb{G}$ is $\Omega(\ln n)$, the correlation decay property (2.13) mostly concerns pairs of variables that are far apart. Complementing this result, the following theorem deals with the joint distribution of the values of variables in the vicinity of a specific reference variable. Formally, for a variable x of a CSP instance G let $\nabla_{2\ell}(G, x)$ be the CSP obtained from G by deleting all variables and constraints at a distance greater than 2ℓ from x. Of course, $\mu_{\nabla_{2\ell}(G,x)}$ denotes the Boltzmann distribution of this CSP. For comparison, let $\mu_{G,\nabla_{2\ell}(G,x)}$ denote the joint distribution of the variables in $\nabla_{2\ell}(G, x)$ under the Boltzmann distribution μ_G of the entire CSP G. Thus, if all functions ψ are $\{0, 1\}$-valued, then $\mu_{G,\nabla_{2\ell}(G,x)}(\sigma)$ is proportional to the number of possible ways of extending a satisfying assignment σ of $\nabla_{2\ell}(G, x)$ to a satisfying assignment of G.

A priori the two distributions $\mu_{\mathbb{G},\nabla_{2\ell}(\mathbb{G},x_i)}$ and $\mu_{\nabla_{2\ell}(\mathbb{G},x_i)}$ might be rather different. Indeed, under $\mu_{\nabla_{2\ell}(\mathbb{G},x_i)}$ the boundary variables at distance precisely 2ℓ from x_i are subject to the sub-CSP $\nabla_{2\ell}(\mathbb{G}, x_i)$ only, whereas in $\mu_{\mathbb{G},\nabla_{2\ell}(\mathbb{G},x_i)}$ they are connected to further constraints. These further constraints are apt to form longish chains (of a typical length of about $\Theta(\ln n)$) through which the boundary variables are connected with each other, at least if $d > 1/(k-1)$ exceeds the giant component threshold. Nevertheless, the following theorem shows that the correlations along these chains decay quickly enough so that the two distributions are close to each other for most variables x_i.

Theorem 2.9 (**SYM**, **BAL**, **MIN**). *Let* $0 < d < d_{\mathrm{cond}}$. *Then, for any* $\ell \geqslant 1$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[\|\mu_{\mathbb{G},\nabla_{2\ell}(\mathbb{G},x_i)} - \mu_{\nabla_{2\ell}(\mathbb{G},x_i)}\|_{\mathrm{TV}} \mid Z(\mathbb{G}) > 0] = 0. \tag{2.14}$$

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

2.7 Reconstruction

Theorem 2.9 allows us to prove a prediction from [59] regarding a 'point-to-set' decorrelation property called non-reconstruction. Recall that we let $\langle\cdot\rangle_{\mathbb{G}}$ denote the expectation with respect to samples $\boldsymbol{\sigma}$ from $\mu_{\mathbb{G}}$. Further, let $\langle\cdot\mid\overline{\nabla_{2\ell}(\mathbb{G},x_i)}\rangle_{\mathbb{G}}$ denote the conditional expectation given the values $\boldsymbol{\sigma}(x)$ of all variables x at a distance greater than 2ℓ from x_i. Then we define

$$\begin{aligned}\mathrm{corr}(d)=\limsup_{\ell\to\infty}\limsup_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\sum_{\omega\in\Omega}\mathbb{E}\Big[\mathbf{1}\{Z(\mathbb{G})>0\}\langle|\langle\mathbf{1}\{\boldsymbol{\sigma}(x_i)=\omega\}\mid\overline{\nabla_{2\ell}(\mathbb{G},x_i)}\rangle_{\mathbb{G}}-1/q|\rangle_{\mathbb{G}}\\+\mathbf{1}\{Z(\mathbb{G})=0\}\Big].\end{aligned}\tag{2.15}$$

In words, we choose a random variable x_i and a value $\omega\in\Omega$. Then we choose a random CSP $\mathbb{G}$ and check whether $\mathbb{G}$ is satisfiable. If so, we draw a sample $\boldsymbol{\tau}$ from the Boltzmann distribution $\mu_{\mathbb{G}}$ and fix the variables at a distance greater than 2ℓ from x_i to the values observed under $\boldsymbol{\tau}$ (the outer $\langle\cdot\rangle_{\mathbb{G}}$). Subsequently we draw a further sample $\boldsymbol{\sigma}$ from the Boltzmann distribution $\mu_{\mathbb{G}}$ given the boundary condition induced by $\boldsymbol{\tau}$ (the inner $\langle\cdot\rangle_{\mathbb{G}}$). The value that we record is by how much the conditional marginal probability differs from $1/q$. Additionally, unsatisfiable $\mathbb{G}$ contribute a value of one. Thus, if $\mathrm{corr}(d)=0$ then typically the value of x_i is independent of *all* the values at a large enough distance ℓ. The *reconstruction threshold*

$$d_{\mathrm{rec}}=\inf\{d>0:\mathrm{corr}(d)>0\}\wedge d_{\mathrm{cond}}\tag{2.16}$$

is defined as the smallest density where this decorrelation property fails (or at most d_{cond}).

A priori the reconstruction threshold seems extremely difficult to analyse because the definition of $\mathrm{corr}(d)$ involves the Boltzmann distribution induced by the random graph $\mathbb{G}$. However, verifying a prediction from [59], we prove that the Boltzmann distribution of the random graph can be replaced by that of a random Galton–Watson tree, which is conceptually far simpler. This multi-type Galton–Watson tree $\mathbb{T}(d,P)$ mimics the local structure of $\mathbb{G}$. Its types are either variables or constraints, which come with a weight function $\psi\in\Psi$. The root is a variable r, and the offspring of a variable is a $\mathrm{Po}(d)$ number of constraints whose weight functions are chosen from P independently. The parent variable occurs in a random position from $\{1,\ldots,k\}$ in each of these constraints; the positions are also chosen independently for each constraint. Moreover, each constraint has precisely $k-1$ children, which are variables. For an integer $\ell\geqslant 0$ we let $\mathbb{T}^{2\ell}(d,P)$ denote the top 2ℓ layers of this tree and we define

$$\mathrm{corr}^{\star}(d)=\lim_{\ell\to\infty}\sum_{\omega\in\Omega}\mathbb{E}\langle|\langle\mathbf{1}\{\boldsymbol{\sigma}(r)=\omega\}\mid\overline{\nabla_{2\ell}(\mathbb{T}^{2\ell}(d,P),r)}\rangle_{\mathbb{T}^{2\ell}(d,P)}-1/q|\rangle_{\mathbb{T}^{2\ell}(d,P)}.\tag{2.17}$$

Of course, the outer expectation $\mathbb{E}[\,\cdot\,]$ refers to the Galton–Watson process, the outer $\langle\cdot\rangle_{\mathbb{T}^{2\ell}(d,P)}$ represents the choice of a random boundary condition (*i.e.* the values of all variables at distance precisely 2ℓ from r), and the inner $\langle\cdot\rangle_{\mathbb{T}^{2\ell}(d,P)}$ stands for the conditional distribution of the value $\boldsymbol{\sigma}(r)$ given the boundary condition. The *tree reconstruction threshold* is defined as

$$d^{\star}_{\mathrm{rec}}=\inf\{d>0:\mathrm{corr}^{\star}(d)>0\}.$$

Theorem 2.10 (**SYM**, **BAL**, **MIN**, **POS**, **UNI**). *We have* $d^{\star}_{\mathrm{rec}}=d_{\mathrm{rec}}$.

Thus, Theorem 2.10 reduces the study of the reconstruction problem on $\mathbb{G}$ to the same problem on the random tree $\mathbb{T}(d,P)$, a task that can be tackled via a number of techniques (such as the 'contraction method' [15]).

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

2.8 Quiet planting

A random CSP organically gives rise to an associated distribution on inference problems called the *planted model*. This is a random CSP instance built around a given 'planted' solution. The algorithmic task is to detect and infer the planted solution from the CSP instance. This computational challenge, which has a remarkably long history, has been harnessed as a benchmark for algorithms based on a broad variety of paradigms, ranging from combinatorial to spectral methods to semidefinite programming (*e.g.* [9, 38, 57]). In addition, planted models have been put forward as one-way function candidates in cryptography [50].

To define the planted model, first draw an assignment $\boldsymbol{\sigma}^* \in \Omega^{V_n}$ uniformly at random. Given $\boldsymbol{\sigma}^*$, let $\mathbb{G}^*(n, m, P, \boldsymbol{\sigma}^*)$ be the random CSP instance drawn from the distribution

$$\mathbb{P}[\mathbb{G}^*(n, m, P, \boldsymbol{\sigma}^*) = G \mid \boldsymbol{\sigma}^*] = \frac{\psi_G(\boldsymbol{\sigma}^*)\mathbb{P}[\mathbb{G}(n, m, P) = G]}{\mathbb{E}[\psi_{\mathbb{G}(n,m,P)}(\boldsymbol{\sigma}^*)]}. \tag{2.18}$$

Thus, we reweigh the prior $\mathbb{G}(n, m, P)$ according to the weight $\psi_G(\boldsymbol{\sigma}^*)$ of the planted assignment. In the most common case where all functions $\psi \in \Psi$ are $\{0, 1\}$-valued, (2.18) can be stated equivalently as follows:

> Draw $\mathbb{G}^*(n, m, P, \boldsymbol{\sigma}^*)$ from the conditional distribution of $G(n, m, P)$ given the event $\{\boldsymbol{\sigma}^* \models G(n, m, P)\}$.

In other words, $\mathbb{G}^*(n, m, P, \boldsymbol{\sigma}^*)$ is chosen uniformly from the set of all CSP instances for which $\boldsymbol{\sigma}^*$ is satisfying.

In the event that

$$\mathbb{E}[\psi_{\mathbb{G}(n,m,P)}(\boldsymbol{\sigma}^*)] = 0,$$

the distribution $\mathbb{G}^*(n, m, P, \boldsymbol{\sigma}^*)$ is undefined. To deal with this technicality we let $\mathbb{G}^*$ be the conditional distribution of $\mathbb{G}^*(n, m, P, \boldsymbol{\sigma}^*)$ given $\mathbb{E}[\psi_{\mathbb{G}(n,m,P)}(\boldsymbol{\sigma}^*)] > 0$, where we recall that m has distribution $\mathrm{Po}(dn/k)$. Because in a random assignment $\boldsymbol{\sigma}^*$ each value $\omega \in \Omega$ very likely occurs about n/q times, condition **SYM** ensures that the event $\mathbb{E}[\psi_{\mathbb{G}(n,m,P)}(\boldsymbol{\sigma}^*)] > 0$ has probability $1 - \exp(-\Omega(n))$ for any fixed $d > 0$.

The most modest algorithmic question associated with the planted model is the *detection problem* (see [16, 33, 65]). It asks for an algorithm that can distinguish the planted model $\mathbb{G}^*$ from the null model $\mathbb{G}$. Formally, with probability $1/2$ the algorithm is given an input from the distribution $\mathbb{G}$, and with probability $1/2$ the input is drawn from $\mathbb{G}^*$. The task is to discern correctly with high probability from which distribution the input was chosen. The following theorem shows that d_{cond} marks the threshold from where such an algorithm exists. Recall that the two random graph models $\mathbb{G}, \mathbb{G}^*$ are *mutually contiguous* if, for any sequence $(\mathcal{E}_n)_{n\geqslant 1}$ of events, we have the equivalence

$$\lim_{n\to\infty} \mathbb{P}[\mathbb{G} \in \mathcal{E}_n] = 0 \quad \Leftrightarrow \quad \lim_{n\to\infty} \mathbb{P}[\mathbb{G}^* \in \mathcal{E}_n] = 0.$$

By contrast, we call the models *mutually orthogonal* if there exists $(\mathcal{E}_n)_{n\geqslant 1}$ such that

$$\lim_{n\to\infty} \mathbb{P}[\mathbb{G} \in \mathcal{E}_n] = 1, \quad \text{while } \lim_{n\to\infty} \mathbb{P}[\mathbb{G}^* \in \mathcal{E}_n] = 0.$$

Theorem 2.11 (**SYM**, **BAL**, **MIN**, **UNI**). *For all $d < d_{\mathrm{cond}}$ the models $\mathbb{G}$ and $\mathbb{G}^*$ are mutually contiguous. If* **POS** *is satisfied as well, then $\mathbb{G}$ and $\mathbb{G}^*$ are mutually orthogonal for all $d > d_{\mathrm{cond}}$.*

In particular, for $d < d_{\mathrm{cond}}$ no algorithm can tell with high probability whether its input stems from $\mathbb{G}$ or $\mathbb{G}^*$, regardless of the running time. By contrast, the proof of Theorem 2.11 yields an (exponential time) algorithm that distinguishes the two distributions w.h.p. for $d > d_{\mathrm{cond}}$.

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

The first part of Theorem 2.11 can be sharpened in an important way. Namely, the contiguity statement extends to the graph/satisfying assignment pairs $(\mathbb{G}^*, \boldsymbol{\sigma}^*)$ and $(\mathbb{G}, \boldsymbol{\sigma})$, where we recall that $\boldsymbol{\sigma}$ denotes a satisfying assignment drawn from the Boltzmann distribution $\mu_{\mathbb{G}}$.

Corollary 2.12 (**SYM**, **BAL**, **MIN**, **UNI**). *For every $d < d_{\mathrm{cond}}$ the pairs $(\mathbb{G}, \boldsymbol{\sigma})$ and $(\mathbb{G}^*, \boldsymbol{\sigma}^*)$ are mutually contiguous.*

Corollary 2.12 enables us to study typical properties of the pair $(\mathbb{G}, \boldsymbol{\sigma})$ by way of the planted model $(\mathbb{G}^*, \boldsymbol{\sigma}^*)$, a technique known as *quiet planting* [3, 60]. This method has proved vital for the analysis of many properties of specific examples of random CSPs (*e.g.* [10, 63]). Corollary 2.12 shows that quiet planting is a universal technique and establishes d_{cond} as the precise threshold up to which the method is applicable.

2.9 Discussion and related work

The results presented in this section vindicate and go in some ways beyond the predictions made in [59] on the basis of the non-rigorous cavity method for a broad class of random constraint satisfaction problems. In short, we obtain a very accurate description of the 'replica symmetric' phase of random CSPs, *i.e.* of the regime of densities up to the condensation threshold. Since in many prominent examples the condensation threshold is known to be quite close to the satisfiability threshold, these results typically cover most of the satisfiable regime. Furthermore, we expect that the 'quiet planting' result (Theorem 2.11 and Corollary 2.12) will pave the way for further detailed results on the evolution of random CSPs.

That said, a number of questions remain open. Specifically, we know very little about the regime $d > d_{\mathrm{cond}}$, *i.e.* beyond the replica symmetric phase. For instance, neither the decorrelation property (2.13) nor the local convergence property (2.14) are conjectured to extend beyond d_{cond}, but we do not currently have a proof. Furthermore, in [59] the reconstruction threshold is simply defined as $d_{\mathrm{rec}} = \inf\{d > 0 : \mathrm{corr}(d) > 0\}$, without taking the min with d_{cond} as in (2.16). We conjecture that these two definitions are equivalent, which would follow immediately if we knew that (2.12) does not hold for $d > d_{\mathrm{cond}}$. Also, apart from the example of the regular k-NAESAT problem for large k [74], the limit of $\sqrt[n]{Z(\mathbb{G})}$ is not known for $d > d_{\mathrm{cond}}$ for any random CSP.

An important feature of the results presented here is that they apply to CSPs with very small average degrees. In most previous work, particularly in work based on combinatorial second moment arguments [19, 28, 29, 36, 35], the assumption that the average variable degree is sufficiently large is endemic. The assumption is usually made implicitly by requiring, for example, that the number q of colours in the graph colouring problem or the clause length k in a random k-NAESAT problem is sufficiently large. Roughly speaking, these combinatorial arguments effectively use the notion that a sufficiently dense Erdős–Rényi graph is not very far from regular. By contrast, since here we avoid such asymptotic arguments, we are in a position to do away with implicit or explicit density assumptions.

One of the guiding themes in the theory of random CSPs is the quest for satisfiability thresholds. Despite considerable efforts, to this day the *exact* thresholds are known in only a handful of cases such as random 2-SAT, random 1-in-k-SAT, random k-XORSAT and random linear equations [4, 12, 23, 34, 37, 49, 68]. Additionally, a line of work on the second moment method [5, 8, 28, 29, 36] culminated in the exact computation of the k-SAT threshold for large k [35]. In other cases such as (hyper-) graph colouring, upper and lower bounds are known that differ by a small additive constant in the limit of large k and/or q [6, 13, 19, 24, 32, 39]. We observed that as a by-product Theorem 2.7 yields lower bounds on the satisfiability thresholds of several problems, particularly hypergraph colouring and random k-SAT for small k, which are at least as good as (and likely better than) those obtained in prior work [8, 13, 39].

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

While in Section 3 we will see many examples of random CSPs that satisfy the assumptions **SYM**, **BAL**, *etc.*, there are a few interesting ones that do not. For instance, the random k-SAT problem fails to satisfy **SYM**. At the same time, it is easy to prove that in random k-SAT the number of solutions is not as tightly concentrated as Theorem 2.7 shows it is in the case of problems that satisfy our assumptions. In fact, the random k-SAT partition function has multiplicative fluctuations of order $\exp(\Omega(\sqrt{n}))$. Thus, random k-SAT is materially different.

Theorems 2.6 and 2.11 can be seen as generalizations of results obtained in [16, 25] for the stochastic block model, a planted version of the Potts model that has become a prominent benchmark for Bayesian inference [1, 65]. In the stochastic block model the Kesten–Stigum bound marks the point from where an efficient algorithm is known to solve the detection problem [2]. But generally the Kesten–Stigum bound is strictly greater than the condensation threshold, and it has been conjectured that in the intermediate regime the detection problem can be solved in exponential but not in polynomial time [33]. In light of Theorems 2.6 and 2.11 it would be interesting to see if the detection problem can be solved efficiently for general random CSPs if $d > d_{\mathrm{KS}}$, and in fact if there are examples of (in the worst case NP-hard) random CSPs where efficient algorithms succeed for $d_{\mathrm{cond}} < d < d_{\mathrm{KS}}$.

With respect to proof techniques the present work builds strongly upon the methods developed in [25, 27]. The additional technical challenge that we need to confront is the presence of *hard* constraints that strictly forbid certain value combinations. In other words, we allow constraint functions ψ that may take the value 0, whereas [25] deals with soft constraints only, as does [27], apart from an *ad hoc* limiting result about the condensation threshold in the random graph colouring problem. We will discuss the difficulties that hard constraints cause in more detail as we proceed, but roughly speaking the matter is as follows. One of the main proof steps is to quantify precisely the evolution of the partition function of the random CSP if we add one random constraint after the other. While we can use the techniques from [25, 27] directly to analyse the *typical* effect of adding a hard constraint, there is an error probability that these estimates are off. In the case of soft constraints, this is not a very serious issue because the impact of a single soft constraint cannot be catastrophic. But in the presence of hard constraints it can. In fact, a single awkward constraint can wipe out all satisfying assignments in one stroke. In summary, we will still follow the strategy developed in [25, 27], but we have to come up with new ideas to cope with 'exceptional' cases more accurately. Hence, throughout Sections 6 and 7 we repeatedly adapt or apply arguments from [25, 27]. To avoid repetitions we put off those bits of the arguments that required only minute amendments to the appendix. Additionally, we will be able to extend several of the results from [25, 27] to the case of hard constraints directly by a limiting argument. More details can be found in Section 5, which contains a proof outline.

The proofs of Theorems 2.9 and 2.10 about local weak convergence and the reconstruction problem are based on a new argument that is somewhat more straightforward than prior ones from [25, 26, 64]. The basic proof idea, which goes back to the work of Gerschenfeld and Montanari [48], is to derive the desired properties of the Boltzmann distribution from the overlap result, Theorem 2.8 in our case. But the new insight here is that this implication can be obtained fairly directly from a key statement called the Nishimori identity (Lemma 5.1 below). A similar observation was made in [25, Section 11], but there the idea was used directly to deduce the reconstruction threshold, without considering local weak convergence explicitly. Here we first establish the local weak convergence result, from which we then derive the reconstruction statement. As it turns out, this line of argument allows for a shorter, more transparent proof. The details can be found in Section 8.

3. Examples

In the following we present several examples of well-studied CSPs that satisfy the assumptions of the main results.

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

3.1 Random k-NAESAT

In Example 2.3 we saw how the random k-NAESAT can be stated as a random CSP over $\Omega = \{\pm 1\}$ with $P_{k-\mathrm{NAE}}$ being the uniform distribution on the 2^k functions

$$\psi_\tau : \sigma \in \Omega^k \mapsto 1 - \mathbf{1}\{\sigma = \tau\} - \mathbf{1}\{\sigma = -\tau\} \quad \text{for } \tau \in \Omega^k.$$

Lemma 3.1. *For any $k \geqslant 3$ the distribution $P_{k-\mathrm{NAE}}$ satisfies* **SYM**, **BAL**, **MIN**, **POS** *and* **UNI**.

Proof. Clearly, $q = 2$ and $\xi = 1 - 2^{1-k}$, and it is immediate that $P_{k-\mathrm{NAE}}$ is permutation-invariant. Further, for either $\omega \in \Omega$ and any $\tau \in \Omega^k$ and any $i \in [k]$ the number of assignments $\sigma \in \Omega^k$ with $\sigma_i = \omega$ with $\psi_\tau(\sigma) = 1$ is equal to $2^{k-1} - 1$, which shows **SYM**. For **BAL** we observe that

$$\phi(\mu) = \sum_{\sigma \in \Omega^k} \mathbb{E}[\boldsymbol{\psi}(\sigma)] \prod_{i=1}^k \mu(\sigma_i) = 1 - 2^{-k} \sum_{\sigma,\tau \in \Omega^k} (\mathbf{1}\{\sigma = \tau\} + \mathbf{1}\{\sigma = -\tau\}) \prod_{i=1}^k \mu(\sigma_i) = 1 - 2^{1-k} \tag{3.1}$$

is constant. Further, regarding **MIN**, fix a probability distribution ρ on $\Omega \times \Omega$ such that $\rho(1,1) + \rho(1,-1) = \rho(1,1) + \rho(-1,1) = 1/2$ and let $r = \rho(1,1) + \rho(-1,-1)$. Then by (3.1),

$$\begin{aligned}
\varphi(\rho) &= \sum_{\sigma,\sigma' \in \Omega^k} \mathbb{E}[\boldsymbol{\psi}(\sigma)\boldsymbol{\psi}(\sigma')] \prod_{i=1}^k \rho(\sigma_i, \sigma_i') \\
&= 1 - 2^{2-k} + 2^{-k} \sum_{\sigma,\sigma',\tau \in \Omega^k} \mathbf{1}\{\sigma = \pm\tau,\ \sigma' = \pm\tau\} \prod_{i=1}^k \rho(\sigma_i, \sigma_i') \\
&= 1 - 2^{2-k} + 2^{1-k}(r^k + (1-r)^k).
\end{aligned}$$

This function is convex and attains its minimum at $r = 1/2$, corresponding to $\rho = \bar{\rho}$. Hence, $P_{k-\mathrm{NAE}}$ satisfies **MIN**.

Moving on to **POS**, fix two distributions $\pi, \pi' \in \mathcal{P}_*^2(\Omega)$ and an integer $\ell \geqslant 2$. Then

$$\begin{aligned}
\mathbb{E}\left[\left(1 - \sum_{\sigma \in \Omega^k} \boldsymbol{\psi}(\sigma) \prod_{i=1}^k \boldsymbol{\rho}_i(\sigma_i)\right)^\ell\right] &= 2^{-k} \sum_{\tau \in \Omega^k} \mathbb{E}\left[\left(\prod_{i=1}^k \boldsymbol{\rho}_i(\tau_i) + \prod_{i=1}^k \boldsymbol{\rho}_i(-\tau_i)\right)^\ell\right] \\
&= 2^{\ell-k} \prod_{i=1}^k \mathbb{E}[\boldsymbol{\rho}_i(1)^\ell + \boldsymbol{\rho}_i(-1)^\ell] \\
&= 2^{\ell-k} \mathbb{E}[\boldsymbol{\rho}_1(1)^\ell + \boldsymbol{\rho}_1(-1)^\ell]^k.
\end{aligned} \tag{3.2}$$

Analogously,

$$\mathbb{E}\left[\left(1 - \sum_{\sigma \in \Omega^k} \boldsymbol{\psi}(\sigma) \prod_{i=1}^k \boldsymbol{\rho}_i'(\sigma_i)\right)^\ell\right] = 2^{\ell-k} \mathbb{E}[\boldsymbol{\rho}_1'(1)^\ell + \boldsymbol{\rho}_1'(-1)^\ell]^k, \tag{3.3}$$

$$\mathbb{E}\left[\left(1 - \sum_{\tau \in \Omega^k} \boldsymbol{\psi}(\tau)\boldsymbol{\rho}_1(\tau_1) \prod_{i=2}^k \boldsymbol{\rho}_i'(\tau_i)\right)^\ell\right] = 2^{\ell-k} \mathbb{E}[\boldsymbol{\rho}_1(1)^\ell + \boldsymbol{\rho}_1(-1)^\ell] \mathbb{E}[\boldsymbol{\rho}_1'(1)^\ell + \boldsymbol{\rho}_1'(-1)^\ell]^{k-1}. \tag{3.4}$$

Due to the elementary inequality $X^k + (k-1)Y^k - kXY^{k-1} \geqslant 0$ for all $X, Y \geqslant 0$, **POS** follows from (3.2)–(3.4). Finally, condition **UNI** is satisfied for $k \geqslant 3$ because every k-clause contains a variable that does not belong to the cycle. □

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

Theorem 1.1 follows immediately by combining Lemma 3.1 with Theorem 2.5. Similarly, Theorem 1.2 follows from Lemma 3.1 and Theorem 2.8.

3.2 Random (hyper-) graph colouring

The random hypergraph colouring problem was defined as a CSP in Example 2.2. The following lemma shows that the problem satisfies all of our assumptions. Hence, Theorem 2.5 yields the exact condensation threshold of this problem for all values of the uniformity parameter k and the number q of colours, except naturally the trivial case $q = k = 2$. Additionally, Theorem 2.7 yields the limiting distribution of the number of colourings and Corollary 2.12 establishes quiet planting. An asymptotically tight quiet planting result was obtained prior to the present work by Ayre and Greenhill [14]. Specifically, for any fixed $k \geqslant 3$ they proved quiet planting for degrees $d < d_{\mathrm{cond}} - \varepsilon_k(q)$, where $\varepsilon_k(q) \to 0$ in the limit of large q. Additionally, Ayre and Greenhill obtain the precise rigidity threshold in the random hypergraph problem, a question that we do not deal with in the present work. Finally, for $k = 2$ we obtain Theorems 1.3 and 1.4 from Section 1.3.

Lemma 3.2. *For any $k \geqslant 2$, $q \geqslant 2$ with $k + q > 4$ the random hypergraph colouring problem satisfies* **SYM**, **BAL**, **MIN**, **POS** *and* **UNI**.

Proof. We have $\Omega = [q]$ and $\xi = 1 - q^{1-k}$ and the single constraint function $\psi_{k,q}$ is invariant under permutations of its coordinates. Furthermore, if we fix the colour of one vertex in a hyperedge, then there are $q^{k-1} - 1$ possible ways to colour the others so that the hyperedge is bichromatic. Hence, **SYM** is satisfied. With respect to **BAL** we have

$$\phi(\mu) = \sum_{\sigma \in \Omega^k} \psi_{k,q}(\sigma) \prod_{i=1}^{k} \mu(\sigma_i) = 1 - \sum_{\sigma \in \Omega} \mu(\sigma)^k. \tag{3.5}$$

This function is concave with its maximum attained at the uniform distribution, whence **BAL** follows. Coming to **MIN**, we fix a probability distribution ρ on Ω with uniform marginals. Then (3.5) implies that

$$\varphi(\rho) = \sum_{\sigma,\tau \in \Omega^k} \psi_{k,q}(\sigma)\psi_{k,q}(\tau) \prod_{i=1}^{k} \rho(\sigma_i, \tau_i) = 1 - 2q^{1-k} + \sum_{\sigma,\tau \in \Omega} \rho(\sigma, \tau)^k.$$

Clearly, the right-hand side is a convex function that attains its minimum at the uniform distribution, whence we obtain **MIN**.

To show **POS**, fix two $\pi, \pi' \in \mathcal{P}^2_*(\Omega)$ and $\ell \geqslant 2$. Then

$$\mathbb{E}\left[\left(1 - \sum_{\tau \in \Omega^k} \psi_{k,q}(\tau) \prod_{i=1}^{k} \boldsymbol{\rho}_i(\tau_i)\right)^{\ell}\right] = \sum_{\sigma_1,\ldots,\sigma_\ell \in \Omega} \mathbb{E}\left[\prod_{i=1}^{k}\prod_{j=1}^{\ell} \boldsymbol{\rho}_i(\sigma_j)\right] = \sum_{\sigma_1,\ldots,\sigma_\ell \in \Omega} \mathbb{E}\left[\prod_{j=1}^{\ell} \boldsymbol{\rho}_1(\sigma_j)\right]^k. \tag{3.6}$$

Similarly,

$$\mathbb{E}\left[\left(1 - \sum_{\tau \in \Omega^k} \psi_{k,q}(\tau) \prod_{i=1}^{k} \boldsymbol{\rho}'_i(\tau_i)\right)^{\ell}\right] = \sum_{\sigma_1,\ldots,\sigma_\ell \in \Omega} \mathbb{E}\left[\prod_{j=1}^{\ell} \boldsymbol{\rho}'_1(\sigma_j)\right]^k, \tag{3.7}$$

$$\mathbb{E}\left[\left(1 - \sum_{\tau \in \Omega^k} \psi_{k,q}(\tau)\boldsymbol{\rho}_1(\tau_1) \prod_{i=2}^{k} \boldsymbol{\rho}'_i(\tau_i)\right)^{\ell}\right] = \sum_{\sigma_1,\ldots,\sigma_\ell \in \Omega^k} \mathbb{E}\left[\prod_{j=1}^{\ell} \boldsymbol{\rho}_1(\sigma_j)\right]\mathbb{E}\left[\prod_{j=1}^{\ell} \boldsymbol{\rho}'_1(\sigma_j)\right]^{k-1}. \tag{3.8}$$

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

Thus, **POS** follows from (3.6)–(3.8) and the elementary inequality $X^k + (k-1)Y^k - kXY^{k-1} \geqslant 0$ for $X, Y \geqslant 0$. Finally, it is well known that condition **UNI** is satisfied for all $k, q \geqslant 2$ except $k = q = 2$. □

3.3 Balanced satisfiability

The following CSP was introduced in [8] to derive a lower bound on the satisfiability threshold for random k-SAT. Let $\Omega = \{\pm 1\}$, $k \geqslant 3$ and let $\lambda = \lambda(k) \in (0, 1)$ be the unique root of

$$(1-\lambda)(1+\lambda)^{k-1} - 1 = 0. \tag{3.9}$$

Further, for $\tau \in \Omega^k$ let

$$\psi_\tau(\sigma) = \lambda^{\sum_{j=1}^k \mathbf{1}\{\sigma_j = \tau_j\}} \left(1 - \prod_{i=1}^k \mathbf{1}\{\sigma_i = -\tau_i\}\right) \tag{3.10}$$

and let $P_{k-\mathrm{BAL}}$ be the uniform distribution on these 2^k functions.

If we omit the λ-factor in (3.10), then we recover the classical random k-SAT problem. Indeed, if we identify the Boolean values true and false with -1 and $+1$, then a constraint endowed with the function

$$\sigma \in \Omega^k \mapsto 1 - \prod_{i=1}^k \mathbf{1}\{\sigma_i = -\tau_i\} \in \{0, 1\} \tag{3.11}$$

represents a k-clause in which the ith variable appears positively if $\tau_i = 1$ and negatively if $\tau_i = -1$. However, as we explained in Section 2.9, the k-SAT problem fails to satisfy condition **SYM**, and thus the results of the present paper do not cover this example. In fact, for the same reason it is not possible to lower-bound the satisfiability threshold of random k-SAT by applying the second moment method to the number of satisfying assignments (see [5, 8]). Therefore, in order to lower-bound the k-SAT threshold, Achlioptas and Peres [8] introduced the weighted constraint functions (3.10). The λ-factor weighs each σ according to the number of true literals; more specifically, since $\lambda \in (0, 1)$ there is a penalty for 'over-satisfying' clauses. This penalty factor guarantees that **SYM** is satisfied in the resulting weighted CSP, which we call the *balanced satisfiability problem*. Achlioptas and Peres applied the second moment method to the corresponding partition function $Z(\mathbb{G}(n, m, P_{k-\mathrm{BAL}}))$, which yields a lower bound on the number of satisfying assignments as $\lambda \in (0, 1)$.

The following lemma shows that the balanced satisfiability problem meets all our conditions bar **POS**.

Lemma 3.3. *For any $k \geqslant 3$ the distribution $P_{k-\mathrm{BAL}}$ satisfies* **SYM**, **BAL**, **MIN** *and* **UNI**.

Theorem 2.5 and (2.6) therefore show that d_{cond} is a lower bound on the satisfiability threshold of the *balanced* satisfiability problem. In fact, because the ψ_τ are upper-bounded by the unweighted (3.11), d_{cond} is also a lower bound on the actual k-SAT threshold for every $k \geqslant 3$. This lower bound, although difficult to evaluate numerically, improves over the one that can be obtained via the second moment method. Furthermore, the contiguity result provided by Theorem 2.11 proves a statistical physics conjecture of Krzakala, Mézard and Zdeborová [58].

Proof of Lemma 3.3. For **SYM**, note that for any $\sigma \in \Omega^k$,

$$\mathbb{E}[\boldsymbol{\psi}(\sigma)] = 2^{-k} \sum_{\tau \in \Omega^k} \lambda^{\sum_{j=1}^k \mathbf{1}\{\sigma_j = \tau_j\}} \left(1 - \prod_{i=1}^k \mathbf{1}\{\sigma_i = -\tau_i\}\right) = 2^{-k}((1+\lambda)^k - 1) = 2^{1-k}\frac{\lambda}{1-\lambda}, \tag{3.12}$$

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

which directly implies $\xi = 2^{1-k}\lambda/(1-\lambda)$. Hence, for all $i \in \{1,\dots,k\}$, $\omega \in \Omega$ and $\tau \in \Omega^k$ we have

$$\begin{aligned}\sum_{\sigma\in\Omega^k} \mathbf{1}\{\sigma_i=\omega\}\psi_\tau(\sigma) &= \sum_{\sigma\in\Omega^k} \mathbf{1}\{\sigma_i=\omega\}\prod_{j=1}^{k}\lambda^{\mathbf{1}\{\sigma_j=\tau_j\}} - \sum_{\sigma\in\Omega^k}\prod_{j=1}^{k}\mathbf{1}\{\sigma_j=-\tau_j\}\mathbf{1}\{\sigma_i=\omega\}\\ &= (1+\lambda)^{k-1}\lambda^{\mathbf{1}\{\tau_i=\omega\}} - \mathbf{1}\{\tau_i\neq\omega\}\\ &= \frac{\lambda}{1-\lambda} = 2^{k-1}\xi.\end{aligned}$$

Thus, **SYM** is satisfied. As (3.12) implies that for any $\mu \in \mathcal{P}(\Omega)$, $\phi_{k-\mathrm{BAL}}(\mu) = \xi$, **BAL** is also satisfied.

We next turn to condition **MIN**. Fix a probability distribution ρ on $\Omega\times\Omega$ such that $\rho(1,1)+\rho(1,-1) = \rho(1,1)+\rho(-1,1) = 1/2$ and let $r = \rho(1,1)+\rho(-1,-1)$. Then $\varphi(\rho) = \varphi(r) = 2^{-k}f(r)$ with f from [8, equation (8)] and thus

$$\varphi(r) = \left(\frac{r}{2}(1-\lambda)^2+\lambda\right)^k - 2\left(\frac{r}{2}(1-\lambda)+\frac{\lambda}{2}\right)^k + \left(\frac{r}{2}\right)^k. \tag{3.13}$$

Using the definition of λ, we obtain

$$\varphi'(r) = \frac{k}{2}\sum_{j=1}^{k-1}\binom{k-1}{j}\left(\frac{r}{2}-\frac{1}{4}\right)^j\left(\frac{1}{4}\right)^{k-1-j}\left(1-\left(\frac{1-\lambda}{1+\lambda}\right)^j\right)^2. \tag{3.14}$$

It is immediate from (3.14) that $\varphi'(1/2) = 0$, while $\varphi'(r) > 0$ for $r \in (1/2, 1]$. For $r < 1/2$, all terms corresponding to odd j in (3.14) are negative, while those corresponding to even j are positive. Let

$$c_j = \binom{k-1}{j}\left(\frac{1}{4}\right)^{k-1-j}\left(1-\left(\frac{1-\lambda}{1+\lambda}\right)^j\right)^2 \quad \text{such that } \varphi'(r) = \frac{k}{2}\sum_{j=1}^{k-1} c_j\left(\frac{r}{2}-\frac{1}{4}\right)^j.$$

The ratio of an odd coefficient j and its even successor $j+1$ works out to be

$$\frac{c_j}{c_{j+1}} = \frac{(j+1)}{4(k-(j+1))}\left(1-\left(\frac{1-\lambda}{1+\lambda}\right)^j\right)^2\left(1-\left(\frac{1-\lambda}{1+\lambda}\right)^{j+1}\right)^{-2},$$

which is increasing in j. Thus, $\varphi'(r)$ is negative for all $r \in (0, 1/2)$ such that

$$\frac{1}{4}-\frac{r}{2} < \frac{c_1}{c_2},$$

which is the case for

$$r > \frac{1}{2} - \frac{(1+\lambda)^2}{4(k-2)} = r_k^*.$$

Unfortunately, only $r_3^* \leqslant 0$, and for $k \geqslant 4$ we upper-bound $\varphi'(r)$ by hand for all $r \in [0, r_k^*)$ to show that is is negative. By [8, Lemma 7] for all $k \geqslant 3$ the following bounds on λ hold:

$$2^{1-k} + k4^{-k} < 1-\lambda < 2^{1-k} + 3k4^{-k}. \tag{3.15}$$

Let

$$g(r) = (1-\lambda)2^{k-1} - 2\lambda^{k-1} + \frac{r^{k-1}}{1-\lambda}.$$

Using

$$\frac{r}{2}(1-\lambda)^2 < (1-\lambda) \quad \text{and} \quad \frac{r}{2}(1-\lambda) \geqslant 0,$$

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

we obtain

$$\begin{aligned}\varphi'(r) &= \frac{k}{2}\left((1-\lambda)^2\left(\frac{r}{2}(1-\lambda)^2+\lambda\right)^{k-1} - 2(1-\lambda)\left(\frac{r}{2}(1-\lambda)+\frac{\lambda}{2}\right)^{k-1} + \left(\frac{r}{2}\right)^{k-1}\right)\\ &< k2^{-k}(1-\lambda)\left((1-\lambda)2^{k-1} - 2\lambda^{k-1} + \frac{r^{k-1}}{1-\lambda}\right)\\ &= k2^{-k}(1-\lambda)g(r).\end{aligned}$$

As $g(r)$ is strictly increasing in r, finding a $\bar{r}_k \in [r_k^*, 1/2)$ such that $g(\bar{r}_k) \leqslant 0$ for all $k \geqslant 3$ suffices to establish **MIN**. To this end, for all $k \geqslant 5$, set

$$\bar{r}_k = \lambda\left((1-\lambda)\left(2 - \frac{1+3\cdot 2^{-(k+1)}k}{1-(k-1)2^{1-k}-3k(k-1)4^{-k}}\right)\right)^{1/(k-1)}.$$

Using (3.15) yields that

$$\begin{aligned}g(\bar{r}_k) &= (1-\lambda)2^{k-1} - \lambda^{k-1}\left(\frac{1+3\cdot 2^{-(k+1)}k}{1-(k-1)2^{1-k}-3k(k-1)4^{-k}}\right)\\ &\leqslant 1 - \frac{\lambda^{k-1}}{1-(k-1)2^{1-k}-3k(k-1)4^{-k}} + 3k2^{-(k+1)}\left(1 - \frac{\lambda^{k-1}}{1-(k-1)2^{1-k}-3k(k-1)4^{-k}}\right)\\ &\leqslant 0,\end{aligned}$$

because

$$\lambda^{k-1} \geqslant (1-2^{1-k}-3k4^{-k})^{k-1} \geqslant 1-(k-1)2^{1-k}-3k(k-1)4^{-k}.$$

To verify that $\bar{r}_k \geqslant r_k^*$, we have by (3.15) that for all $k \geqslant 6$

$$\begin{aligned}(1+k2^{-(k+1)})\left(2 - \frac{1+3\cdot 2^{-(k+1)}k}{1-(k-1)2^{1-k}-3k(k-1)4^{-k}}\right) &\geqslant \frac{1}{2},\\ (k-1)(2^{1-k}+3k4^{-k}) \leqslant 0.2 \quad \text{and} \quad (1+\lambda)^2 &\geqslant 3.8.\end{aligned} \tag{3.16}$$

Thus, combining (3.15) and (3.3) yields

$$\begin{aligned}\bar{r}_k &\geqslant \frac{\lambda}{2}\left(\exp\left(-\frac{\ln 2}{k-1}\right)\right)\\ &\geqslant \frac{1}{2}(1-2^{1-k}-3k4^{-k})\left(1-\frac{\ln 2}{k-1}\right)\\ &\geqslant \frac{1}{2} - \frac{2\ln 2 + 4(k-1)(2^{1-k}+3k4^{-k})}{4(k-1)}\\ &\geqslant \frac{1}{2} - \frac{2.2}{4(k-1)}\\ &\geqslant r_k^*.\end{aligned}$$

For $k=5$ we calculate that $r_5^* < 0.19$ whereas $\bar{r}_5 > 0.32$. We are left with the case $k=4$ where $r_4^* < 0.083$, but using g as an upper bound turns out to be too crude. We have $0.14 < 1-\lambda < 0.18$, and thus for all $r \in [0, 0.1]$ we calculate

$$\varphi'(r) \leqslant 2(0.18^2(0.1\cdot 0.18^2 + 0.18)^3 - 2\cdot 0.14\cdot 0.7^3 + 0.1^3) \leqslant -0.18.$$

Finally, **UNI** is satisfied for $k \geqslant 3$ because every k-clause contains a variable that does not belong to the cycle. □

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

3.4 Parity–majority

We consider the following compound CSP, which has been suggested as a device for constructing one-way functions in cryptography [11].[1] Each constraint function evaluates the XOR of two structurally different parts, namely a parity check and a majority function. Formally, let $\Omega = \{\pm 1\}$ and $k \geqslant 3$ be an odd integer. For $\tau \in \Omega^{2k}$ and a permutation θ of $[2k]$, define the constraint function $\psi_{\tau,\theta} : \Omega^{2k} \to \{0, 1\}$,

$$\psi_{\tau,\theta}(\sigma) = \mathbf{1}\left\{\prod_{i=1}^{k} \sigma_{\theta(i)}\tau_i = 1\right\}\mathbf{1}\left\{\sum_{i=k+1}^{2k} \sigma_{\theta(i)}\tau_i < 0\right\} + \mathbf{1}\left\{\prod_{i=1}^{k} \sigma_{\theta(i)}\tau_i = -1\right\}\mathbf{1}\left\{\sum_{i=k+1}^{2k} \sigma_{\theta(i)}\tau_i > 0\right\}.$$

Let

$$\Psi = \{\psi_{\tau,\theta} : \tau \in \Omega^{2k}, \theta \text{ permutation of } [2k]\}$$

and let P_{MAJ} be the uniform distribution over Ψ. In words, a sample from P_{MAJ} is generated by uniformly choosing a vector of 'signs' (determining for each position whether the corresponding input is negated) and k positions participating in the parity check and the majority function, respectively. Now, an assignment σ satisfies $\psi_{\tau,\theta}$ if either the parity of the literals $(\sigma_{\theta(i)}\tau_i)_{i=1,\ldots,k}$ equals 1 and the majority of literals $(\sigma_{\theta(i)}\tau_i)_{i=k+1,\ldots,2k}$ votes for -1, or *vice versa*.

Lemma 3.4. *For any $k \geqslant 3$, the parity–majority problem satisfies* **SYM**, **BAL**, **MIN** *and* **UNI**.

Permuting the inputs of the constraint functions is necessary for the second part of **SYM** to hold. However, as for the rest of the arguments the particular choice of θ does not make a structural difference, we may work with the identity map and lighten the notation to $\psi_{\tau,\mathrm{id}} = \psi_\tau$.

Claim 3.5. *For any $k \geqslant 3$, the parity–majority problem satisfies* **SYM**, **BAL** *and* **UNI**.

Proof. Let $\sigma \in \Omega^{2k}$ be arbitrary. The number of $\tau \in \Omega^{2k}$ with $\psi_\tau(\sigma) = 1$ equals 2^{2k-1}, as for any $(\tau_2, \ldots, \tau_{2k})$ there is exactly one choice of τ_1 which leads to $\psi_\tau(\sigma) = 1$. As each $\tau \in \Omega^{2k}$ is chosen with equal probability, this implies that $\mathbb{E}[\boldsymbol{\psi}(\sigma)] = 1/2$, irrespective of $\sigma \in \Omega^{2k}$. Thus, $\xi = 1/2$.

Similarly, for each $\tau \in \Omega^{2k}$, $i \in [2k]$, $\omega \in \Omega$, the number of $\sigma \in \Omega^{2k}$ with $\psi_\tau(\sigma) = 1$ and $\sigma_i = \omega$ is 2^{2k-2}, as $k \geqslant 3$ and *any* choice of $2k - 1$ components which satisfies $\sigma_i = \omega$ and does not fix one of the first k parity components (σ_j, say) can be extended to a satisfying assignment by choosing this variable σ_j in a unique way. Thus,

$$\sum_{\sigma \in \Omega^{2k}} \mathbf{1}\{\sigma_i = \omega\}\psi_\tau(\sigma) = 2^{2k-2} = 2^{2k-1}\xi,$$

and due to the construction of Ψ and the uniformity of P_{MAJ}, **SYM** is satisfied. Further, the above calculation shows that $\phi_{\mathrm{MAJ}}(\mu) = \xi$ for any $\mu \in \mathcal{P}(\Omega)$, and thus **BAL** is also satisfied as well. Finally, **UNI** is satisfied because, again, $k \geqslant 3$ and every k-clause contains a variable that does not belong to the cycle. □

To prove **MIN** we need to do a bit of calculus. Fix a probability distribution ρ on $\Omega \times \Omega$ such that $\rho(1, 1) + \rho(1, -1) = \rho(1, 1) + \rho(-1, 1) = 1/2$ and let $r = \rho(1, 1) + \rho(-1, -1)$.

[1] This problem was brought to our attention by Chris Brzuska.

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

Claim 3.6. *We have*

$$\varphi_{\mathrm{MAJ}}(\rho) = \sum_{\sigma,\sigma' \in \Omega^{2k}} \mathbb{E}[\boldsymbol{\psi}(\sigma)\boldsymbol{\psi}(\sigma')] \prod_{i=1}^{2k} \rho(\sigma_i, \sigma_i') = \sum_{\sigma,\sigma' \in \Omega^{2k}} \psi_{(1,\ldots,1)}(\sigma)\psi_{(1,\ldots,1)}(\sigma') \prod_{i=1}^{2k} \rho(\sigma_i, \sigma_i'). \tag{3.17}$$

Proof. Indeed, something stronger is true: for any $\tau, \tau' \in \Omega^{2k}$,

$$\begin{aligned}\sum_{\sigma,\sigma' \in \Omega^{2k}} \psi_\tau(\sigma)\psi_\tau(\sigma') \prod_{i=1}^{k} \rho(\sigma_i, \sigma_i') &= \sum_{\sigma,\sigma' \in \Omega^{2k}} \psi_{\tau'}(\sigma)\psi_{\tau'}(\sigma') \prod_{i=1}^{k} \rho(\tau_i'\tau_i\sigma_i, \tau_i'\tau_i\sigma_i') \\ &= \sum_{\sigma,\sigma' \in \Omega^{2k}} \psi_{\tau'}(\sigma)\psi_{\tau'}(\sigma') \prod_{i=1}^{k} \rho(\sigma_i, \sigma_i'),\end{aligned}$$

and the claim follows by applying the above to $\tau' = (1, \ldots, 1)$. □

Define

$$f : [0,1] \to \mathbb{R}, r \mapsto 2^{-k} \sum_{\sigma,\sigma' \in \Omega^{k}} \mathbf{1}\left\{\prod_{i=1}^{k} \sigma_i = 1\right\} \mathbf{1}\left\{\prod_{i=1}^{k} \sigma_i' = 1\right\} r^{\sum_{i=1}^{k} \mathbf{1}\{\sigma_i = \sigma_i'\}} (1-r)^{k - \sum_{i=1}^{k} \mathbf{1}\{\sigma_i = \sigma_i'\}}, \tag{3.18}$$

$$g : [0,1] \to \mathbb{R}, r \mapsto 2^{-k} \sum_{\sigma,\sigma' \in \Omega^{k}} \mathbf{1}\left\{\sum_{i=1}^{k} \sigma_i < 0\right\} \mathbf{1}\left\{\sum_{i=1}^{k} \sigma_i' < 0\right\} r^{\sum_{i=1}^{k} \mathbf{1}\{\sigma_i = \sigma_i'\}} (1-r)^{k - \sum_{i=1}^{k} \mathbf{1}\{\sigma_i = \sigma_i'\}}. \tag{3.19}$$

Claim 3.7. *With f and g defined in (3.18), (3.19), we have*

$$\varphi_{\mathrm{MAJ}}(r) = 2(f(r)g(r) + f(1-r)g(1-r)). \tag{3.20}$$

Proof. Using Claim 3.7, we rewrite

$$\begin{aligned}\varphi_{\mathrm{MAJ}}(r) = &\sum_{\sigma,\sigma' \in \Omega^{2k}} \left(\mathbf{1}\left\{\prod_{i=1}^{k} \sigma_i = 1\right\} \mathbf{1}\left\{\sum_{i=k+1}^{2k} \sigma_i < 0\right\} + \mathbf{1}\left\{\prod_{i=1}^{k} \sigma_i = -1\right\} \mathbf{1}\left\{\sum_{i=k+1}^{2k} \sigma_i > 0\right\}\right) \\ &\times \left(\mathbf{1}\left\{\prod_{i=1}^{k} \sigma_i' = 1\right\} \mathbf{1}\left\{\sum_{i=k+1}^{2k} \sigma_i' < 0\right\} + \mathbf{1}\left\{\prod_{i=1}^{k} \sigma_i' = -1\right\} \mathbf{1}\left\{\sum_{i=k+1}^{2k} \sigma_i' > 0\right\}\right) \\ &\times \prod_{i=1}^{2k} \rho(\sigma_i, \sigma_i') \\ = 2 &\sum_{\sigma,\sigma' \in \Omega^{2k}} \mathbf{1}\left\{\prod_{i=1}^{k} \sigma_i = 1\right\} \mathbf{1}\left\{\sum_{i=k+1}^{2k} \sigma_i < 0\right\} \mathbf{1}\left\{\prod_{i=1}^{k} \sigma_i' = 1\right\} \mathbf{1}\left\{\sum_{i=k+1}^{2k} \sigma_i' < 0\right\} \\ &\times \left(\prod_{i=1}^{2k} \rho(\sigma_i, \sigma_i') + \prod_{i=1}^{2k} \rho(\sigma_i, -\sigma_i')\right) \\ = 2&(f(r)g(r) + f(1-r)g(1-r)),\end{aligned} \tag{3.21}$$

as desired. □

We can easily write down an explicit expression for the parity component.

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

Claim 3.8. *For all $r \in [0,1]$ we have*

$$f(r) = \frac{1}{4}(1 + (1-2r)^k).$$

Proof. For odd k a pair $(\sigma, \sigma') \in \Omega^{2k}$ with exactly i common positions has the same parity, if and only if i is odd, thus

$$f(r) = 2^{-k} \sum_{\sigma \in \Omega^k} \mathbf{1}\left\{\prod_{i=1}^{k} \sigma_i = 1\right\} \sum_{i \in [k] : i \text{ is odd}} \binom{k}{i} r^i (1-r)^{k-i} = \frac{1}{2} \sum_{i \in [k] : i \text{ is odd}} \binom{k}{i} r^i (1-r)^{k-i}.$$

Now, since

$$1 + (1-2r)^k = (r + (1-r))^k - (1-(1-r))^k = 2 \sum_{i \in [k] : i \text{ is odd}} \binom{k}{i} r^i (1-r)^{k-i},$$

the assertion follows. □

Claim 3.9. *For all $r \in [0,1]$ we have $2f(r) + 2f(1-r) = 2g(r) + 2g(1-r) = 1$.*

Proof. Let $\bar{f}(r) = 1/2 - f(r)$ and $\bar{g}(r) = 1/2 - g(r)$, respectively. Rewriting $\varphi_{\mathrm{MAJ}}(r)$ in a slightly different fashion than before yields

$$\begin{aligned}
\varphi_{\mathrm{MAJ}}(r) = 2 \sum_{\sigma, \sigma' \in \Omega^{2k}} \Bigg(& \mathbf{1}\left\{\prod_{i=1}^{k} \sigma_i = 1\right\} \mathbf{1}\left\{\sum_{i=k+1}^{2k} \sigma_i < 0\right\} \mathbf{1}\left\{\prod_{i=1}^{k} \sigma'_i = 1\right\} \mathbf{1}\left\{\sum_{i=k+1}^{2k} \sigma'_i < 0\right\} \\
& + \mathbf{1}\left\{\prod_{i=1}^{k} \sigma_i = 1\right\} \mathbf{1}\left\{\sum_{i=k+1}^{2k} \sigma_i > 0\right\} \mathbf{1}\left\{\prod_{i=1}^{k} \sigma'_i = -1\right\} \mathbf{1}\left\{\sum_{i=k+1}^{2k} \sigma'_i < 0\right\} \Bigg) \prod_{i=1}^{2k} \rho(\sigma_i, \sigma'_i) \\
= 2(f(r)g(r) + & \bar{f}(r)\bar{g}(r)).
\end{aligned}$$

Thus, combining this with (3.21) we obtain

$$f(1-r)g(1-r) = \bar{f}(r)\bar{g}(r). \tag{3.22}$$

Since k is odd, Claim 3.8 yields

$$2\bar{f}(r) = 1 - \frac{1}{2}(1 + (1-2r)^k) = \frac{1}{2}(1 + (1-2(1-r))^k) = 2f(1-r). \tag{3.23}$$

The claim now readily follows from (3.22), (3.23) and the definitions of $\bar{f}, \bar{g}$. □

Claim 3.10. *The function f is strictly increasing on $(0,1) \setminus \{1/2\}$, while g is increasing on $[0,1)$.*

Proof. Given Claim 3.8 and recalling that k is odd, we see that f is strictly increasing on $[0, 1/2)$ and $(1/2, 1]$ with a saddle point at $r = 1/2$.

The function g, which corresponds to the majority part, is more complicated. For $j \in \{1, \ldots, k\}$, let $\mathcal{S}_j$ be the set of pairs of assignments with majority vote -1 which agree on exactly the first j components, and let $g_j = |\mathcal{S}_j|$ be the number of such pairs. Then

$$g(r) = 2^{-k} \sum_{j=1}^{k} \binom{k}{j} g_j r^j (1-r)^{k-j},$$

$$2^k g'(r) = g_1 (1-r)^{k-1} + \sum_{j=1}^{k-1} \binom{k-1}{j} (g_{j+1} - g_j) r^j (1-r)^{k-(j+1)}.$$

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

It is therefore sufficient to show that $g_{j+1} \geqslant g_j$ for all $j \in \{1, \ldots, k-1\}$. To this end, we consider the following injective map h from $\mathcal{S}_j$ to $\mathcal{S}_{j+1}$. Given a pair of solutions $(s^{(1)}, s^{(2)}) \in \mathcal{S}_j$, we let $(\bar{s}^{(1)}, \bar{s}^{(2)}) \in \Omega^{2k}$ denote the assignment pair obtained from $(s^{(1)}, s^{(2)})$ by swapping their $(j+1)$th component. There are two possible cases. If $(\bar{s}^{(1)}, \bar{s}^{(2)})$ is not in $\mathcal{S}_j$, we set the $(j+1)$th component of both $s^{(1)}$ and $s^{(2)}$ to -1 and obtain a valid solution pair in $\mathcal{S}_{j+1}$. On the other hand, if both $(s^{(1)}, s^{(2)}), (\bar{s}^{(1)}, \bar{s}^{(2)}) \in \mathcal{S}_j$, then in order for h to be injective, we assign 1 to the $(j+1)$th component of $\bar{s}^{(1)}, \bar{s}^{(2)}$. This gives a valid solution in $\mathcal{S}_{j+1}$, because the fact that both $(s^{(1)}, s^{(2)})$ and $(\bar{s}^{(1)}, \bar{s}^{(2)})$ are solutions implies that they have a majority vote of -1 irrespective of the value of their $(j+1)$th component. Thus g is increasing on $[0, 1)$. □

Proof of Lemma 3.4. Claim 3.5 establishes **SYM**, **BAL** and **UNI**. With respect to **MIN**, Claims 3.6–3.10 show that φ_{MAJ} has a unique minimum at $1/2$, as

$$\varphi'_{\mathrm{MAJ}}(r) = 2(f'(r)(2g(r) - 1/2) + g'(r)(2f(r) - 1/2)).$$ □

4. Preliminaries and notation

4.1 Basics

Throughout the paper we continue to use the notation introduced in Sections 2 and 5. In particular, we write $V_n = \{x_1, \ldots, x_n\}$ for a set of n variable nodes and $F_m = \{a_1, \ldots, a_m\}$ for a set of m constraint nodes. Further, $m(d, n)$ is a random variable with distribution $\mathrm{Po}(dn/k)$ and we just write $m(d)$ or m if n and/or d are apparent. Additionally, we let $\mathcal{M}(d)$ be the set of all sequences $m = m(n)$ such that $|m(n) - dn/k| \leqslant n^{3/5}$ for all n.

We write $\mathcal{P}(\mathcal{X})$ for the set of probability measures on a finite set $\mathcal{X}$. We identify $\mathcal{P}(\mathcal{X})$ with the standard simplex in $\mathbb{R}^{\mathcal{X}}$, thereby turning $\mathcal{P}(\mathcal{X})$ into a Polish space. Further, for $\sigma_1, \ldots, \sigma_l : V_n \to \Omega$ let $\rho_{\sigma_1,\ldots,\sigma_l} \in \mathcal{P}(\Omega^l)$ denote the *l-wise overlap*, defined by

$$\rho_{\sigma_1,\ldots,\sigma_l}(\omega_1, \ldots, \omega_l) = |\sigma_1^{-1}(\omega_1) \cap \cdots \cap \sigma_l^{-1}(\omega_l)|/n. \tag{4.1}$$

We use this notation also in the case $l = 1$, and then $\rho_{\sigma_1} \in \mathcal{P}(\Omega)$ is just the empirical distribution of the spins under σ_1. Further, we let $\bar{\rho}_l$ signify the uniform distribution on Ω^l. In particular, $\bar{\rho}_1$ is the uniform distribution on Ω. We usually omit the index l to ease the notation. An assignment $\sigma : V_n \to \Omega$ is *nearly balanced* if $\|\rho_\sigma - \bar{\rho}\|_{\mathrm{TV}} \leqslant n^{-2/5}$. In addition, for two spin assignments $\sigma, \tau : V \to \Omega$ we let $\sigma \triangle \tau = \{v \in V : \sigma(v) \neq \tau(v)\}$.

The entropy of a probability distribution $\mu \in \mathcal{P}(\mathcal{X})$ is always denoted by $\mathcal{H}(\mu)$. Thus, recalling that $\Lambda(z) = z \ln z$ for $z > 0$ and setting $\Lambda(0) = 0$, we have $\mathcal{H}(\mu) = -\sum_{x \in \mathcal{X}} \Lambda(\mu(x))$.

By default we use O notation to refer to the limit $n \to \infty$. On the few occasions where we refer to a different limit we say so.

4.2 Constraint satisfaction problems

In a few places we will need to look at a slightly more general class of constraint satisfaction problems than introduced in Section 2.1. Namely, let Ω be a finite set. By extension of Definition 2.1, a general *constraint satisfaction problem* $G = (V, F, (\partial a)_{a \in F}, (\psi_a)_{a \in F})$ consists of a finite set V of variables, a finite set F of constraints, a function $\psi_a : \Omega^{k_a} \to [0, 1]$ for some integer $k_a \geqslant 1$, and a tuple $\partial a \in V^{k_a}$. The difference here is that the ψ_a are not required to belong to a fixed finite set, and that the arities k_a of the constraints can be different. As before, we introduce

$$\psi_G(\sigma) = \prod_{a \in F} \psi_a(\sigma(\partial_1 a, \ldots, \partial_{k_a} a)) \quad (\sigma \in \Omega^V),$$

$$Z(G) = \sum_{\sigma \in \Omega^V} \psi_G(\sigma).$$

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

Further, if $Z(G) > 0$ we introduce the *Boltzmann distribution* by letting $\mu_G(\sigma) = \psi_G(\sigma)/Z(G)$ for $\sigma \in \Omega^V$.

We will need the following general observation about random CSPs.

Lemma 4.1 (**SYM**). *The function*

$$\phi : \mathbb{R}^\Omega \to \mathbb{R}, \quad \rho \mapsto \sum_{\tau \in \Omega^k} \mathbb{E}[\boldsymbol{\psi}(\tau)] \prod_{i=1}^{k} \rho(\tau_i)$$

satisfies $D\phi(\bar{\rho}) = k\xi \mathbf{1}$ *and* $D^2\phi(\bar{\rho}) = qk(k-1)\xi\Phi$. *Moreover,* ϕ *is strictly positive on the interior of* $\mathcal{P}(\Omega)$.

Proof. The first and second derivatives can be computed along the lines of the proof of [25, Lemma 4.4]. The positivity bit is immediate as the product $\prod_{i=1}^{k} \rho(\tau_i)$ is uniformly bounded below and $\sum_{\tau \in \Omega^k} \mathbb{E}[\boldsymbol{\psi}(\tau)] = q^k\xi > 0$. □

4.3 Boltzmann distributions

Suppose that $\mathcal{X}, V$ are finite sets and let $N = |V|$. For a measure $\mu \in \mathcal{P}(\mathcal{X}^V)$, a subset $U \subset V$ and $\sigma \in \mathcal{X}^U$ we let

$$\mu_U(\sigma) = \sum_{\tau \in \mathcal{X}^V} \mathbf{1}\{\forall i \in U : \tau_i = \sigma_i\}\mu(\tau).$$

Thus, μ_U is the marginal distribution that μ induces on U. Where the reference to U is evident we just write $\mu(\sigma)$. Additionally, we use the shorthand $\mu_{i_1,\ldots,i_h}$ for $\mu_{\{i_1,\ldots,i_h\}}$ if $i_1, \ldots, i_h \in V$.

If $\mu \in \mathcal{P}(\mathcal{X}^V)$, then $\boldsymbol{\sigma}_\mu, \boldsymbol{\tau}_\mu, \boldsymbol{\sigma}_{1,\mu}, \boldsymbol{\sigma}_{2,\mu}, \ldots \in \mathcal{X}^V$ denote mutually independent samples from μ. Where μ is apparent from the context we omit the index and just write $\boldsymbol{\sigma}, \boldsymbol{\tau}$, *etc.* If $X : (\mathcal{X}^V)^l \to \mathbb{R}$ is a random variable, then we write

$$\langle X \rangle_\mu = \langle X(\boldsymbol{\sigma}_1, \ldots, \boldsymbol{\sigma}_l) \rangle_\mu = \sum_{\sigma_1,\ldots,\sigma_l \in \Omega^{V_n}} X(\sigma_1, \ldots, \sigma_l) \prod_{j=1}^{l} \mu(\sigma_j).$$

Thus, $\langle X \rangle_\mu$ is the mean of X over independent samples from μ.

If $\mu = \mu_G$ is the Boltzmann distribution induced by a CSP instance G, we write $\boldsymbol{\sigma}_G$ *etc.* instead of $\boldsymbol{\sigma}_{\mu_G}$ and we also write $\langle \cdot \rangle_G$ rather than $\langle \cdot \rangle_{\mu_G}$. We use this notation to distinguish averages over μ_G from other sources of randomness (*e.g.* the choice of the random CSP), for which we reserve the symbols $\mathbb{E}[\cdot]$ and $\mathbb{P}[\cdot]$.

Let $\varepsilon > 0$ and $\ell \geqslant 2$. Following [17], we say that the probability measure $\mu \in \mathcal{P}(\mathcal{X}^V)$ is (ε, ℓ)-*symmetric* if

$$\sum_{1 \leqslant i_1 < \cdots < i_\ell \leqslant N} \|\mu_{i_1,\ldots,i_\ell} - \mu_{i_1} \otimes \cdots \otimes \mu_{i_\ell}\|_{\mathrm{TV}} < \varepsilon N^\ell.$$

(The idea is to express that the joint distribution of ℓ randomly chosen coordinates is likely to be close to a product distribution.) Further, an $(\varepsilon, 2)$-symmetric measure is simply called ε-*symmetric*. We need the following two results from [17].

Lemma 4.2 (**[17, Corollaries 2.3 and 2.4]**). *For any* $\mathcal{X} \neq \emptyset$, $l \geqslant 3$, $\delta > 0$ *there is* $\varepsilon > 0$ *such that for all* $N > 1/\varepsilon$ *the following is true:*

If $\mu \in \mathcal{P}(\mathcal{X}^I)$ *is* ε*-symmetric, then* μ *is* (δ, l)*-symmetric.*

Let $\mu^{\otimes\ell} \in \mathcal{P}((\mathcal{X}^V)^\ell)$ be the distribution $\mu^{\otimes\ell}(\sigma_1, \ldots, \sigma_\ell) = \prod_{j=1}^{\ell} \mu(\sigma_j)$.

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

Lemma 4.3 ([**17, Proposition 2.5**]). *For any $\varepsilon > 0$, $\ell \geqslant 1$, $\mathcal{X} \neq \emptyset$ there exists $\delta > 0$ such that for all $N > 1/\delta$ the following is true:*

If $\mu \in \mathcal{P}(\mathcal{X}^V)$ is δ-symmetric, then $\mu^{\otimes \ell}$ is ε-symmetric.

The following lemma relates ε-symmetry and the overlap.

Lemma 4.4. *For any $\varepsilon > 0$, $\mathcal{X} \neq \emptyset$ there exist $\delta > 0$, $n_0 > 0$ such that for all $n > n_0$ and all $\mu \in \mathcal{P}(\mathcal{X}^n)$ the following is true:*

If $\langle \|\rho_{\boldsymbol{\sigma},\boldsymbol{\tau}} - \bar{\rho}\|_{\mathrm{TV}} \rangle_\mu < \delta$, then μ is ε-symmetric and $\sum_{i=1}^n \|\mu_i - \bar{\rho}\|_{\mathrm{TV}} < \varepsilon n$.

Conversely, for any $\varepsilon > 0$, $\mathcal{X} \neq \emptyset$ there exist $\delta > 0$, $n_0 > 0$ such that for all $n > n_0$ and all $\mu \in \mathcal{P}(\mathcal{X}^n)$ the following is true:

If μ is δ-symmetric and $\sum_{i=1}^n \|\mu_i - \bar{\rho}\|_{\mathrm{TV}} < \delta n$, then $\langle \|\rho_{\boldsymbol{\sigma},\boldsymbol{\tau}} - \bar{\rho}\|_{\mathrm{TV}} \rangle_\mu < \varepsilon$.

Although Lemma 4.4 was known (and used) before, we are not aware of a convenient reference. We therefore prove the lemma in Appendix A.

Corollary 4.5. *For any finite set $\mathcal{X}$, any $\varepsilon > 0$ and any $l \geqslant 3$ there exist $\delta = \delta(\mathcal{X}, \varepsilon, l)$ and $n_0 = n_0(\mathcal{X}, \varepsilon, l)$ such that for all $n > n_0$ and all $\mu \in \mathcal{P}(\mathcal{X}^{V_n})$ the following is true:*

If $\langle \|\rho_{\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2} - \bar{\rho}\|_{\mathrm{TV}} \rangle < \delta$, then $\langle \|\rho_{\boldsymbol{\sigma}_1,\ldots,\boldsymbol{\sigma}_l} - \bar{\rho}_l\|_{\mathrm{TV}} \rangle < \varepsilon$.

Proof. This follows from Lemmas 4.3 and 4.4. □

The following lemma shows that there is a generic (randomized) way of perturbing a given measure in such a way that the outcome is likely ε-symmetric.

Lemma 4.6 ([**27, Lemma 5.3**]). *For any $\varepsilon > 0$ and any $\mathcal{X} \neq \emptyset$, there exists a bounded integer random variable $\boldsymbol{\theta}_\varepsilon \geqslant 0$ such that for all $\mu \in \mathcal{P}(\mathcal{X}^V)$ for sufficiently large N the following is true. Obtain a random probability measure $\check{\boldsymbol{\mu}} \in \mathcal{P}(\mathcal{X}^V)$ as follows.*

- *Choose a set $U \subset V$ of size $\boldsymbol{\theta}_\varepsilon$ uniformly at random.*
- *Independently draw $\check{\boldsymbol{\sigma}} \in \mathcal{X}^V$ from μ.*
- *Define the (random) probability measure*

$$\check{\boldsymbol{\mu}}(\sigma) = \frac{\mu(\sigma)\mathbf{1}\{\forall i \in U : \sigma_i = \check{\boldsymbol{\sigma}}_i\}}{\mu(\{\tau \in \mathcal{X}^V : \forall i \in U : \tau_i = \check{\boldsymbol{\sigma}}_i\})} \quad (\sigma \in \mathcal{X}^V).$$

Then $\check{\boldsymbol{\mu}}$ is ε-symmetric with probability at least $1 - \varepsilon$.

Thus, in order to obtain an ε-symmetric measure it suffices to peg a bounded number of randomly chosen coordinates to a 'reference configuration' $\check{\boldsymbol{\sigma}}$. Throughout the paper we let $\boldsymbol{\theta}_\varepsilon$ denote the random variable from Lemma 4.6. It will be convenient to use the convention that $\boldsymbol{\theta}_1 = 0$.

Finally, we need the following fact.

Lemma 4.7 ([**27, Lemma 4.7**]). *For any $\varepsilon > 0$ there is $\delta > 0$ such that for all sufficiently large N the following is true. If $\mu \in \mathcal{P}(\mathcal{X}^V)$ satisfies $\langle \|\rho_{\boldsymbol{\sigma},\boldsymbol{\tau}} - \bar{\rho}\|_{\mathrm{TV}} \rangle_\mu < \delta$, then for all nearly balanced τ we have $\langle \|\rho_{\boldsymbol{\sigma},\tau} - \bar{\rho}\|_{\mathrm{TV}} \rangle_\mu < \varepsilon$.*

Thus, if the overlap of two samples $\boldsymbol{\sigma}, \boldsymbol{\tau}$ is typically close to the uniform overlap $\bar{\rho}$, then in fact the overlap of a random $\boldsymbol{\sigma}$ with an *arbitrary* nearly balanced $\tau \in \mathcal{X}^V$ is likely close to uniform.

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

5. Proof strategy

In this section we outline the proofs of the main results presented in Section 2, deferring some of the details to later sections. Following [25] we approach the proofs of the main results by way of analysing the partition function of the planted model $\mathbb{G}^*$. This will enable us to construct a suitable random variable to which we can apply the small subgraph conditioning technique, originally developed by Robinson and Wormald [70] to count Hamilton cycles in random regular graphs, to prove Theorem 2.7. The other results then derive from Theorem 2.7.

5.1 The planted model revisited

Before we begin let us get a technical issue out of the way. The constraints of the random CSP $\mathbb{G}$ are not quite independent because we require the hypergraph underlying $\mathbb{G}$ to be simple. However, in the proofs this slight dependence becomes a nuisance. We therefore introduce a tweaked model $G(n, m, P)$ with variables $V_n = \{x_1, \ldots, x_n\}$ whose constraints $a_1, \ldots, a_m$ are chosen independently from the following distribution: for each a_i, the k-tuple $\partial a_i \in V_n^k$ is chosen uniformly at random, and the function $\psi_{a_i} \in \Psi$ is chosen from P independently of ∂a_i. Thus, it is possible that the same variable occurs twice in the constraint a_i.

Recall the function ϕ which appeared in **BAL** and Lemma 4.1. Due to independence of the constraints in $G(n, m, P)$, we have the identity

$$\mathbb{E}[\psi_{G(n,m)}(\sigma)] = \phi(\rho_\sigma)^m, \tag{5.1}$$

which will be used in various places below.

Naturally, there is a planted model that goes with $G(n, m, P)$. Namely, let Σ_n be the set of all $\sigma \in \Omega^{V_n}$ such that $\mathbb{E}[\psi_{G(n,m,P)}(\sigma)] > 0$. In other words, Σ_n is the set of assignments that may occur as satisfying assignments of some random CSP instance. By adaptation of (2.18), for $\sigma \in \Sigma_n$ we define the planted model $G^*(n, m, P, \sigma)$ by letting

$$\mathbb{P}[G^*(n, m, P, \sigma) = G] = \frac{\psi_G(\sigma)\mathbb{P}[G(n, m, P) = G]}{\mathbb{E}[\psi_{G(n,m,P)}(\sigma)]}, \tag{5.2}$$

for any possible CSP instance G. Equivalently, because the m constraints of $G(n, m, P)$ are drawn independently, (5.2) can be stated as follows: the constraints $a_1, \ldots, a_m$ are drawn independently from the distribution

$$\mathbb{P}[\partial a_i = (x_{i_1}, \ldots, x_{i_k}), \psi_{a_i} = \psi] = \frac{\psi(\sigma(x_{i_1}), \ldots, \sigma(x_{i_k}))P(\psi)}{\sum_{j_1,\ldots,j_k=1}^{n} \mathbb{E}[\boldsymbol{\psi}(\sigma(x_{j_1}), \ldots, \sigma(x_{j_k}))]}. \tag{5.3}$$

We continue to let $\boldsymbol{\sigma}^* = \boldsymbol{\sigma}_n^*$ denote a uniformly random assignment $V_n \to \Omega$. Suppose we first choose a random assignment $\boldsymbol{\sigma}^* \in \Sigma_n$ uniformly and then draw $G^*(n, m, P, \boldsymbol{\sigma}^*)$ from the planted model. What will be the resulting distribution on CSP instances? If we assume that all $\psi \in \Psi$ take values in $\{0, 1\}$, then this distribution on CSPs should roughly weigh each possible instance G according to its number $Z(G)$ of satisfying assignments; for G has one chance to come up as $G^*(n, m, P, \sigma)$ for each of its satisfying assignments σ. But of course this is only approximately right because the denominator in (5.2) may depend on σ. To correct for this, we introduce a distribution on assignments by letting

$$\mathbb{P}[\hat{\boldsymbol{\sigma}}_{n,m,P} = \sigma] = \frac{\mathbb{E}[\psi_{G(n,m,P)}(\sigma)]}{\mathbb{E}[Z(G(n, m, P))]} \quad \text{for } \sigma \in \Omega^{V_n}. \tag{5.4}$$

Condition **SYM** guarantees that the denominator $\mathbb{E}[Z(G(n, m, P))]$ is non-zero for all $n \geqslant q$. It will emerge in due course that the distributions $\boldsymbol{\sigma}^*$ and $\hat{\boldsymbol{\sigma}}_{n,m,P}$ are mutually contiguous (see Lemma 7.8 below). From now on we tacitly assume that $n \geqslant q$.

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

We claim that the random CSP $G^*(n, m, P, \hat{\boldsymbol{\sigma}}_{n,m,P})$ is distributed *exactly* as the distribution $G(n, m)$ reweighed according to the partition function. Formally, let $\hat{G}(n, m, P)$ be the random CSP with distribution

$$\mathbb{P}[\hat{G}(n, m, P) = G] = \frac{Z(G)\mathbb{P}[G(n, m, P) = G]}{\mathbb{E}[Z(G(n, m, P))]}. \tag{5.5}$$

Then we have the following.

Lemma 5.1. *For all σ, G we have*

$$\mathbb{P}[\hat{\boldsymbol{\sigma}}_{n,m,P} = \sigma] \cdot \mathbb{P}[G^*(n, m, P, \sigma) = G] = \mu_G(\sigma) \cdot \mathbb{P}[\hat{G}(n, m, P) = G]. \tag{5.6}$$

Proof. From the definitions (5.2), (5.4) and (5.5) it is immediate that

$$\begin{aligned}\mathbb{P}[\hat{\boldsymbol{\sigma}}_{n,m,P} = \sigma] \cdot \mathbb{P}[G^*(n, m, P, \sigma) = G] &= \frac{\mathbb{E}[\psi_{G(n,m,P)}(\sigma)]}{\mathbb{E}[Z(G(n, m, P))]} \cdot \frac{\psi_G(\sigma)\mathbb{P}[G(n, m, P) = G]}{\mathbb{E}[\psi_{G(n,m,P)}(\sigma)]} \\ &= \mu_G(\sigma) \cdot \frac{Z(G)\mathbb{P}[G(n, m, P) = G]}{\mathbb{E}[Z(G(n, m, P))]} \\ &= \mu_G(\sigma) \cdot \mathbb{P}[\hat{G}(n, m, P) = G],\end{aligned}$$

as claimed. □

Borrowing a term from the statistical physics literature [75], we call (5.6) the *Nishimori identity*. This identity will play a fundamental role because it allows us to analyse the partition function by way of the planted model. The definitions of the models $\hat{G}(n, m, P)$, $\hat{\boldsymbol{\sigma}}(n, m, P)$ and Lemma 5.1 already appeared in [27] for the case that all $\psi \in \Psi$ are strictly positive (soft constraints).

To unclutter the notation we will skip the reference to P where possible and just write $G(n, m)$, $\hat{G}(n, m)$, *etc.* Further, recalling that $m = m_d(n)$ is a random variable with distribution $\mathrm{Po}(dn/k)$, we introduce $\hat{G} = \hat{G}(n, m, P)$, $G^* = G^*(n, m, P, \boldsymbol{\sigma}^*)$ and $\hat{\boldsymbol{\sigma}} = \hat{\boldsymbol{\sigma}}_{n,m,P}$.

5.2 The heat is on

As mentioned earlier, the point of the present work is that we manage to accommodate hard constraints, *i.e.* functions ψ that may take the value 0. A natural first idea might be to deal with this case by softening the constraints so that the results from [25] apply and to deal with hard constraints by taking the 'softening parameter' to 0. Unfortunately, matters are not quite so simple. But we can still get some mileage out of this idea.

To be precise, for a parameter $\beta \geqslant 0$ and a function $\psi : \Omega^k \to [0, 1]$, define

$$\psi_\beta(\sigma) = \mathrm{e}^{-\beta} + (1 - \mathrm{e}^{-\beta})\psi(\sigma). \tag{5.7}$$

Thus, $\psi_\beta \geqslant \mathrm{e}^{-\beta}$ is a softened version of ψ, and we think of $\mathrm{e}^{-\beta}$ as the softening parameter. In physics jargon, (5.7) corresponds to a 'positive temperature' variant of the CSP, and β might be called the 'inverse temperature'. We let $\Psi_\beta = \{\psi_\beta : \psi \in \Psi\}$. Further, let P_β be the distribution of $\boldsymbol{\psi}_\beta$ and define

$$\xi_\beta = q^{-k} \sum_{\sigma \in \Omega^k} \mathbb{E}[\boldsymbol{\psi}_\beta(\sigma)] = \mathrm{e}^{-\beta} + (1 - \mathrm{e}^{-\beta})\xi.$$

Accordingly, we introduce the symbols $G_\beta(n, m) = G(n, m, P_\beta)$, $\hat{G}_\beta(n, m) = \hat{G}(n, m, P_\beta)$, *etc.*

In order to apply the results from [25] to the 'softened' CSP we observe that P_β satisfies our main assumptions; condition **UNI** is obsolete because all ψ_β are strictly positive.

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

Lemma 5.2. *If P satisfies any of the conditions* **SYM**, **BAL**, **MIN** *and* **POS**, *then so does P_β for any $\beta > 0$.*

Proof. Assuming that P satisfies **SYM**, we find

$$\begin{aligned}\sum_{\tau\in\Omega^k}\mathbf{1}\{\tau_i=\omega\}\psi_\beta(\tau) &= \mathrm{e}^{-\beta}q^{k-1}+(1-\mathrm{e}^{-\beta})\sum_{\tau\in\Omega^k}\mathbf{1}\{\tau_i=\omega\}\psi(\tau)\\ &= q^{k-1}(\mathrm{e}^{-\beta}+(1-\mathrm{e}^{-\beta})\xi)\\ &= q^{k-1}\xi_\beta.\end{aligned}$$

Similarly, if P satisfies **BAL**, then

$$\sum_{\tau\in\Omega^k}\mathbb{E}[\boldsymbol{\psi}_\beta(\tau)]\prod_{i=1}^k\mu(\tau_i)=\mathrm{e}^{-\beta}+(1-\mathrm{e}^{-\beta})\sum_{\tau\in\Omega^k}\mathbb{E}[\boldsymbol{\psi}(\tau)]\prod_{i=1}^k\mu(\tau_i)$$

is a concave function of μ that attains its maximum at the uniform distribution. Moving on to condition **MIN**, we observe that for any $\rho\in\mathcal{R}(\Omega)$,

$$\sum_{\sigma,\tau\in\Omega^k}\mathbb{E}[\boldsymbol{\psi}(\sigma)]\prod_{i=1}^k\rho(\sigma_i,\tau_i)=q^{-k}\sum_{\sigma\in\Omega^k}\mathbb{E}[\boldsymbol{\psi}(\sigma)]=\xi.$$

Hence,

$$\begin{aligned}&\sum_{\sigma,\tau\in\Omega^k}\mathbb{E}[\boldsymbol{\psi}_\beta(\sigma)\boldsymbol{\psi}_\beta(\tau)]\prod_{i=1}^k\rho(\sigma_i,\tau_i)\\ &\quad=\mathrm{e}^{-2\beta}+2\mathrm{e}^{-\beta}(1-\mathrm{e}^{-\beta})\xi+(1-\mathrm{e}^{-\beta})^2\sum_{\sigma,\tau\in\Omega^k}\mathbb{E}[\boldsymbol{\psi}(\sigma)\boldsymbol{\psi}(\tau)]\prod_{i=1}^k\rho(\sigma_i,\tau_i).\end{aligned}$$

Clearly, if P satisfies **MIN**, then the uniform distribution on $\Omega\times\Omega$ will be the unique global minimizer $\rho\in\mathcal{R}(\Omega)$ of the last expression. Finally, regarding **POS** we calculate

$$\mathbb{E}\Big[\Big(1-\sum_{\tau\in\Omega^k}\boldsymbol{\psi}_\beta(\tau)\prod_{i=1}^k\boldsymbol{\rho}_i(\tau_i)\Big)^\ell\Big]=(1-\mathrm{e}^{-\beta})^\ell\cdot\mathbb{E}\Big[\Big(1-\sum_{\tau\in\Omega^k}\boldsymbol{\psi}(\tau)\prod_{i=1}^k\boldsymbol{\rho}_i(\tau_i)\Big)^\ell\Big],$$

$$\mathbb{E}\Big[\Big(1-\sum_{\tau\in\Omega^k}\boldsymbol{\psi}_\beta(\tau)\prod_{i=1}^k\boldsymbol{\rho}'_i(\tau_i)\Big)^\ell\Big]=(1-\mathrm{e}^{-\beta})^\ell\cdot\mathbb{E}\Big[\Big(1-\sum_{\tau\in\Omega^k}\boldsymbol{\psi}(\tau)\prod_{i=1}^k\boldsymbol{\rho}'_i(\tau_i)\Big)^\ell\Big],$$

$$\mathbb{E}\Big[\Big(1-\sum_{\tau\in\Omega^k}\boldsymbol{\psi}_\beta(\tau)\boldsymbol{\rho}_1(\tau_1)\prod_{i=2}^k\boldsymbol{\rho}'_i(\tau_i)\Big)^\ell\Big]=(1-\mathrm{e}^{-\beta})^\ell\cdot\mathbb{E}\Big[\Big(1-\sum_{\tau\in\Omega^k}\boldsymbol{\psi}(\tau)\boldsymbol{\rho}_1(\tau_1)\prod_{i=2}^k\boldsymbol{\rho}'_i(\tau_i)\Big)^\ell\Big].$$

Hence, if P satisfies **POS**, then so does P_β. □

Can we use the softened CSP directly to, say, prove Theorem 2.5 about the condensation phase transition? Suppose we fix a CSP G with (hard) constraints from Ψ and let $Z_\beta(G)$ denote the partition function of the CSP with soft constraints obtained by replacing each ψ with the corresponding ψ_β. Then we verify immediately that $\lim_{\beta\to\infty}Z_\beta(G)=Z(G)$. In other words, G comes out as the 'zero temperature' limit of G_β. Consequently, we obtain

$$\lim_{\beta\to\infty}\mathbb{E}\sqrt[n]{Z(G_\beta)}=\mathbb{E}\sqrt[n]{Z(G)}$$

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

and therefore

$$\lim_{n\to\infty}\lim_{\beta\to\infty}\mathbb{E}\sqrt[n]{Z(G_\beta)}=\lim_{n\to\infty}\mathbb{E}\sqrt[n]{Z(G)}, \tag{5.8}$$

where the second line is conditional on the existence of limits. Furthermore, using the results from [25, 27], we can determine the condensation threshold of the softened CSP $Z(G(n,m,P_\beta))$. Hence, we should be able to compute

$$\lim_{\beta\to\infty}\lim_{n\to\infty}\mathbb{E}\sqrt[n]{Z(G(n,m,P_\beta))}, \tag{5.9}$$

at least for $d<d_{\mathrm{cond}}$.

Alas, the order of the limits in (5.8) and (5.9) is reversed. Whether the limits commute is arguably one of the most challenging open problems in the theory of random CSPs (see the discussion in [30]). The following result, which constitutes one of the main technical contributions of this paper, proves that in planted models the limits do indeed commute. Recall the expression $\mathcal{B}(d,P,\pi)$ from (2.2).

Theorem 5.3 (**SYM**, **BAL**). *For every $d>0$ we have*

$$\limsup_{n\to\infty}\frac{1}{n}\mathbb{E}[\ln Z(\hat{G})]\leqslant\lim_{\beta\to\infty}\sup_{\pi\in\mathcal{P}_*^2(\Omega)}\mathcal{B}(d,P_\beta,\pi)=\sup_{\pi\in\mathcal{P}_*^2(\Omega)}\mathcal{B}(d,P,\pi). \tag{5.10}$$

Furthermore, if **POS** *is satisfied as well, then*

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\mathbb{E}[\ln Z(\hat{G})]&=\lim_{\beta\to\infty}\lim_{n\to\infty}\frac{1}{n}\mathbb{E}[\ln Z(\hat{G}_\beta)]\\&=\lim_{\beta\to\infty}\sup_{\pi\in\mathcal{P}_*^2(\Omega)}\mathcal{B}(d,P_\beta,\pi)\\&=\sup_{\pi\in\mathcal{P}_*^2(\Omega)}\mathcal{B}(d,P,\pi).\end{aligned} \tag{5.11}$$

Apart from being a vital step toward the proofs of the main results, we believe that Theorem 5.3 may be of independent interest for the study of planted instances of CSPs. The proof of Theorem 5.3, which we carry out in Section 6, combines techniques from [25, 27] with new arguments required to deal with hard constraints.

5.3 The Kesten–Stigum bound

We are going to combine Theorem 5.3 with small subgraph conditioning to prove Theorem 2.7. To pave the way for this argument we need two preparations. First, because the eigenvalues of the operator Ξ from (2.8) will come up a lot, we need to investigate the spectrum of Ξ. Also recall the matrix Φ from (2.10) and the space $\mathcal{E}$ from (2.9). Additionally, let

$$\mathcal{E}'=\{x\in\mathbb{R}^q\otimes\mathbb{R}^q:\langle x,\mathbf{1}\otimes\mathbf{1}\rangle=0\}\supset\mathcal{E}. \tag{5.12}$$

Finally, let us introduce the matrices

$$\begin{aligned}\Phi_{\psi_\beta}(\omega,\omega')&=q^{1-k}\xi_\beta^{-1}\sum_{\tau\in\Omega^k}\mathbf{1}\{\tau_1=\omega,\tau_2=\omega'\}\psi_\beta(\tau)\quad\text{for }\omega,\omega'\in\Omega,\\\Phi_\beta&=\mathbb{E}[\Phi_{\boldsymbol{\psi}_\beta}],\\\Xi_\beta&=\mathbb{E}[\Phi_{\boldsymbol{\psi}_\beta}\otimes\Phi_{\boldsymbol{\psi}_\beta}].\end{aligned}$$

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

Lemma 5.4 (**SYM**, **BAL**). *The matrices* Φ, Ξ *enjoy the following properties.*

(i) Φ *is symmetric and doubly stochastic and* $\max_{x\perp\mathbf{1}}\langle\Phi x, x\rangle \leqslant 0$.
(ii) Ξ *is self-adjoint,* $\Xi(\mathbf{1}\otimes\mathbf{1}) = \mathbf{1}\otimes\mathbf{1}$ *and for every* x *we have* $\Xi(x\otimes\mathbf{1}) = (\Phi x)\otimes\mathbf{1}$, $\Xi(\mathbf{1}\otimes x) = \mathbf{1}\otimes(\Phi x)$ *and*

$$\langle\Xi(x\otimes\mathbf{1}), x\otimes\mathbf{1}\rangle \leqslant 0, \quad \langle\Xi(\mathbf{1}\otimes x), \mathbf{1}\otimes x\rangle \leqslant 0 \quad \text{if } x\perp\mathbf{1}. \tag{5.13}$$

Furthermore, $\Xi\mathcal{E}\subset\mathcal{E}$ *and* $\Xi\mathcal{E}'\subset\mathcal{E}'$.

Proof. Lemma 5.2 shows together with [25, Lemmas 3.5 and 3.6] that statements (i) and (ii) hold for Φ_β and Ξ_β for any $\beta > 0$. Since $\lim_{\beta\to\infty}\Phi_\beta = \Phi$ and $\lim_{\beta\to\infty}\Xi_\beta = \Xi$, the assertion follows. □

Since the self-adjoint operator Ξ induces an endomorphism of the subspace $\mathcal{E}$, we define the multi-set

$$\mathrm{Eig}^*(\Xi) = \{\lambda\in\mathbb{R} : \exists x\in\mathcal{E}\setminus\{0\} : \Xi x = \lambda x\} \tag{5.14}$$

that contains each eigenvalue according to its geometric multiplicity. To apply small subgraph conditioning we need the following bound on the spectral radius.

Proposition 5.5 (**SYM**, **BAL**). *We have* $d_{\mathrm{cond}}(k-1)\max_{\lambda\in\mathrm{Eig}^*(\Xi)}|\lambda| \leqslant 1$.

Proposition 5.5 is almost immediate from the following statement about the softened version of the random CSP. By extension of (2.3) and (2.9) we define

$$d_{\mathrm{cond}}(\beta) = \inf\left\{d > 0 : \sup_{\pi\in\mathcal{P}^2_*(\Omega)}\mathcal{B}(d, P_\beta, \pi) > \ln q + \frac{d}{k}\ln\xi\right\},$$
$$d_{\mathrm{KS}}(\beta) = \left((k-1)\max_{x\in\mathcal{E}:\|x\|=1}\langle\Xi_\beta x, x\rangle\right)^{-1}.$$

The following lemma paraphrases several results from [25, Section 5].

Lemma 5.6 (**SYM**, **BAL**). *We have*

$$d_{\mathrm{cond}}(\beta)(k-1)\max_{\lambda\in\mathrm{Eig}^*(\Xi_\beta)}|\lambda| \leqslant 1 \quad \text{for all } \beta > 0.$$

Moreover, if $d > 0$, $\beta_0 > 0$ *are such that* $d > d_{\mathrm{cond}}(\beta)$ *for all* $\beta > \beta_0$, *then there exists* $\varepsilon > 0$ *such that*

$$\sup_{\pi\in\mathcal{P}^*_2(\Omega)}\mathcal{B}(d, P_\beta, \pi) > \ln q + \frac{d}{k}\ln\xi_\beta + \varepsilon \quad \text{for all } \beta > \beta_0. \tag{5.15}$$

Proof of Proposition 5.5. Suppose that d is such that $d(k-1)\max_{\lambda\in\mathrm{Eig}^*(\Xi)}|\lambda| > 1$. Then for all sufficiently large β we have $d(k-1)\max_{\lambda\in\mathrm{Eig}^*(\Xi_\beta)}|\lambda| > 1$, because $\lim_{\beta\to\infty}\Xi_\beta = \Xi$. Therefore, Lemma 5.6 yields $\varepsilon > 0$ such that (5.15) is satisfied for all large enough β. Finally, since $\lim_{\beta\to\infty}\xi_\beta = \xi$, (5.10) yields $\sup_{\pi\in\mathcal{P}^2_*(\Omega)}\mathcal{B}(d, P, \pi) > \ln q + d\ln\xi/k$. Hence, $d > d_{\mathrm{cond}}$. □

Theorem 2.6 drops out as an immediate consequence of Lemma 5.4 and Proposition 5.5.

Proof of Theorem 2.6. We have $\max_{x\in\mathcal{E}:\|x\|=1}\langle\Xi x, x\rangle = \max_{\lambda\in\mathrm{Eig}^*(\Xi)}|\lambda|$ because Lemma 5.4 shows that Ξ is self-adjoint. Therefore, Theorem 2.6 follows from Proposition 5.5. □

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

5.4 The overlap

As a second preparation for the small subgraph conditioning we need to investigate the overlap of two randomly chosen satisfying assignments in the planted model.

Proposition 5.7 (**SYM**, **BAL**, **MIN**).

(1) *Suppose that* $d < d_{\mathrm{cond}}$. *There exists a sequence* $\zeta = \zeta(n) = o(1)$ *such that for all* $m \in \mathcal{M}(d)$ *we have*

$$\mathbb{E}\langle \|\rho_{\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2} - \bar{\rho}\|_{\mathrm{TV}}\rangle_{\hat{G}(n,m)} \leqslant \zeta. \tag{5.16}$$

(2) *Conversely, let* $D > 0$ *and assume that* **POS** *is satisfied as well. If for all* $d < D$ *we have*

$$\mathbb{E}\langle \|\rho_{\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2} - \bar{\rho}\|_{\mathrm{TV}}\rangle_{\hat{G}(n,m)} = o(1), \tag{5.17}$$

then $d_{\mathrm{cond}} \geqslant D$.

We defer the proof of Proposition 5.7 to Section 7. With ζ from Proposition 5.7 we define

$$\mathcal{Z}(G) = Z(G)\mathbf{1}\{\langle \|\rho_{\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2} - \bar{\rho}\|_{\mathrm{TV}}\rangle_G \leqslant \zeta\}. \tag{5.18}$$

Thus, $\mathcal{Z}(G)$ is a truncated version of the partition function $Z(G)$, where an instance G contributes only if its overlaps concentrate about $\bar{\rho}$. A similar truncated variable was used in [25] in the case of soft constraints and in [27] in the special case of the random graph colouring problem.

Corollary 5.8 (**SYM**, **BAL**, **MIN**). *If* $d < d_{\mathrm{cond}}$, *then*

$$\mathbb{E}[\mathcal{Z}(G(n,m))] \sim \mathbb{E}[Z(G(n,m))]$$

uniformly for all $m \in \mathcal{M}(d)$.

Proof. This is immediate from the first part of Proposition 5.7 and definition (5.5) of $\hat{G}(n,m)$. □

5.5 Small subgraph conditioning

We are ready to conduct small subgraph conditioning for the random variable $\mathcal{Z}(G(n,m))$. We begin by computing the first and second moments.

Proposition 5.9 (**SYM**, **BAL**). *Let* $d > 0$. *Then, uniformly for all* $m \in \mathcal{M}(d)$,

$$\mathbb{E}[Z(G(n,m))] \sim \frac{q^{n+1/2}\xi^m}{\prod_{\lambda\in\mathrm{Eig}(\Phi)\setminus\{1\}}\sqrt{1-d(k-1)\lambda}}. \tag{5.19}$$

Proposition 5.10 (**SYM**, **BAL**). *Let* $0 < d < d_{\mathrm{cond}}$. *Then uniformly for all* $m \in \mathcal{M}(d)$,

$$\mathbb{E}[\mathcal{Z}(G(n,m))^2] \leqslant \frac{(1+o(1))q^{2n+1}\xi^{2m}}{\prod_{\lambda\in\mathrm{Eig}^*(\Xi)}\sqrt{1-d(k-1)\lambda}}. \tag{5.20}$$

The expression on the right-hand side of (5.19) makes sense because $\mathrm{Eig}(\Phi)\setminus\{1\} \subset \mathbb{R}_{\leqslant 0}$ by Lemma 5.4. Similarly, Lemma 5.4 and Proposition 5.5 show that in (5.20) we only take square roots of positive numbers if $d < d_{\mathrm{cond}}$.

The proofs of Propositions 5.9 and 5.10 are virtually identical to the moment calculations performed in [25, Section 7]; we included them in Appendix B. Both are fairly straightforward, but

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

the calculation of the second moment hinges on the fact that only CSP instances whose overlap concentrates about $\bar{\rho}$ contribute to $\mathcal{Z}(G(n,m))$. In fact, the second moment of the original random variable $Z(G(n,m))$ is generally *much* bigger (by an exponential factor). In effect, we could not possibly base our small subgraph conditioning argument on the plain random variable $Z(G(n,m))$. Note, however, that up to d_{cond} the first moments of $\mathcal{Z}(G(n,m))$ and $Z(G(n,m))$ are asymptotically the same by Corollary 5.8.

Combining Corollary 5.8 with Propositions 5.9 and 5.10 and applying Lemma 5.4, we obtain

$$\frac{\mathbb{E}[\mathcal{Z}(G(n,m))^2]}{\mathbb{E}[\mathcal{Z}(G(n,m))]^2} \sim \prod_{\lambda\in\mathrm{Eig}^*(\Xi)} \frac{1}{\sqrt{1-d(k-1)\lambda}} \quad \text{if } d < d_{\mathrm{cond}},\ m \in \mathcal{M}(d). \tag{5.21}$$

Thus, Proposition 5.5 shows that the ratio of the second moment and the square of the first is bounded. However, the quotient does not generally converge to 1 as $n \to \infty$. Following the general small subgraph paradigm as set out in [55, 70], we will 'explain' the remaining variance in terms of the bounded-length cycles of the bipartite graph induced by the random CSP instance.

A similar strategy was used in [25] for problems with soft constraints, and we can re-use some of the terminology introduced there. A *signature of order* ℓ is a family

$$Y = (\psi_1, s_1, t_1, \psi_2, s_2, t_2, \ldots, \psi_\ell, s_\ell, t_\ell)$$

such that $\psi_1, \ldots, \psi_\ell \subset \Psi$, $s_1, t_1, \ldots, s_\ell, t_\ell \in [k]$ and $s_i \neq t_i$ for all $i \in [\ell]$ and $s_1 < t_1$ if $\ell = 1$. Let $\mathcal{Y}_\ell$ be the set of all signatures of order ℓ, let $\mathcal{Y}_{\leqslant\ell} = \bigcup_{l\leqslant\ell} \mathcal{Y}_l$ and let $\mathcal{Y} = \bigcup_{\ell\geqslant 1} \mathcal{Y}_\ell$. For a CSP G with variables V_n and constraints F_m, we call a family $(x_{i_1}, a_{h_1}, \ldots, x_{i_\ell}, a_{h_\ell})$ a *cycle of signature* Y *in* G if

CYC1. $i_1, \ldots, i_\ell \in [n]$ are pairwise distinct and $i_1 = \min\{i_1, \ldots, i_\ell\}$,

CYC2. $h_1, \ldots, h_\ell \in [m]$ are pairwise distinct and $h_1 < h_\ell$ if $\ell > 1$,

CYC3. $\psi_{a_{h_j}} = \psi_j$ and $\partial_{s_j} a_{h_j} = x_{i_j}$ for all $j \in \{1, \ldots, \ell\}$, $\partial_{t_j} a_{h_j} = x_{i_{j+1}}$ for all $j < \ell$ and $\partial_{t_\ell} a_{h_\ell} = x_{i_1}$.

Thus, the cycle, which, of course, alternates between variables and constraints, begins with the variable with the smallest index (**CYC1**). From there it is directed toward the constraint with the smaller index (**CYC2**). Furthermore, the constraint functions along the cycle are the ones prescribed by the signature, the cycle enters the jth constraint through its s_jth position and leaves through position number t_j (**CYC3**).

Let $C_Y(G)$ be the number of cycles of signature Y. Moreover, for an event $\psi \in \Psi$ and $h, h' \in \{1, \ldots, k\}$ define the $q \times q$ matrix $\Phi_{\psi,h,h'}$ by letting

$$\Phi_{\psi,h,h'}(\omega,\omega') = q^{1-k}\xi^{-1} \sum_{\tau\in\Omega^k} \mathbf{1}\{\tau_h = \omega, \tau_{h'} = \omega'\}\psi(\tau) \quad (\omega,\omega' \in \Omega). \tag{5.22}$$

In addition, for a signature $Y = (\psi_1, s_1, t_1, \ldots, \psi_\ell, s_\ell, t_\ell)$ define

$$\kappa_Y = \frac{1}{2\ell}\left(\frac{d}{k}\right)^\ell \prod_{i=1}^{\ell} P(\psi_i), \quad \Phi_Y = \prod_{i=1}^{\ell} \Phi_{\psi_i,s_i,t_i}, \quad \hat{\kappa}_Y = \kappa_Y \operatorname{tr}(\Phi_Y). \tag{5.23}$$

Finally, let $\mathfrak{S}$ be the event that the factor graph $G(n,m)$ is simple, that is, $\partial_1 a_i, \ldots, \partial_k a_i$ are pairwise distinct for every $i \in [m]$ and $\{\partial_1 a_i, \ldots, \partial_k a_i\} \neq \{\partial_1 a_j, \ldots, \partial_k a_j\}$ for all $1 \leqslant i < j \leqslant m$. The following proposition, whose proof we put off to Section 7, characterizes the joint distributions of the cycle counts in $G(n,m)$ and $\hat{G}(n,m)$.

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

Proposition 5.11 (**SYM**, **BAL**, **UNI**). *We have $\kappa_Y > 0$ for all $Y \in \mathcal{Y}$, and if $\hat{\kappa}_Y = 0$, then Y has order one and $C_Y(\hat{G}(n,m)) = 0$ deterministically for all n, m. Further, if $Y_1, Y_2, \ldots Y_l \in \mathcal{Y}$ are pairwise distinct and $y_1, \ldots, y_l \geqslant 0$, then for any $d > 0$,*

$$\mathbb{P}[\forall t \leqslant l : C_{Y_t}(G(n,m)) = y_t] \sim \prod_{t=1}^{l} \mathbb{P}[\mathrm{Po}(\kappa_{Y_t}) = y_t] \tag{5.24}$$

uniformly for all $m \in \mathcal{M}(d)$. If, in addition, $\hat{\kappa}_{Y_1}, \ldots, \hat{\kappa}_{Y_l} > 0$, then, uniformly for all $m \in \mathcal{M}(d)$,

$$\mathbb{P}[\forall t \leqslant l : C_{Y_t}(\hat{G}(n,m)) = y_t] \sim \prod_{t=1}^{l} \mathbb{P}[\mathrm{Po}(\hat{\kappa}_{Y_t}) = y_t]. \tag{5.25}$$

Finally,

$$\mathbb{P}[G(n,m) \in \mathfrak{S}] \sim \exp\left(-\frac{d(k-1)}{2} - \frac{\mathbf{1}\{k=2\}d^2}{4}\right),$$
$$\mathbb{P}[\hat{G}(n,m) \in \mathfrak{S}] \sim \exp\left(-\frac{d(k-1)}{2}\,\mathrm{tr}\,(\Phi) - \frac{\mathbf{1}\{k=2\}d^2}{4}\,\mathrm{tr}\,(\Phi^2)\right).$$

Based on Propositions 5.9, 5.10 and 5.11 the proof of Theorem 2.7 is fairly standard. We will carry out the details in Section 5.6. Then in Section 6 we will prove Theorem 5.3. Several of the proof ingredients will be re-used later in Section 7, where we establish Propositions 5.9, 5.10 and 5.11. With all the tools in place, in Section 7 we also complete the proofs of Theorems 2.5, 2.8 and 2.11. Finally, in Section 8 we prove Theorems 2.9 and 2.10.

5.6 Proof of Theorem 2.7

Fix $0 < d < d_{\mathrm{cond}}$ and let $m \in \mathcal{M}(d)$. Let $\mathfrak{F}_\ell = \mathfrak{F}_\ell(n,m)$ be the σ-algebra generated by the cycle counts $(C_Y)_{Y \in \mathcal{Y}_{\leqslant \ell}}$. The proof of Theorem 2.7 follows the original strategy from [70] by studying the conditional variance of $\mathcal{Z}(G(n,m))$ given $\mathfrak{F}_\ell$. Janson [55] stated a relatively general results that covers many applications of this strategy, but unfortunately not ours. The issue is that the number m of constraints in the statement of Theorem 2.7 is random. Therefore, we use a combinatorial argument that goes back to [31], which was also used in [25]. The proof here is similar to the one in [25], and actually considerably simpler because in the present paper the set Ψ of constraint functions is finite. Only the very last part of the proof requires a new argument to accommodate hard constraints.

We aim to prove that $\mathbb{E}[\mathrm{Var}(\mathcal{Z}(G(n,m))|\mathfrak{F}_\ell)]$ is much smaller than $\mathbb{E}[\mathcal{Z}(G(n,m))]$ for large enough ℓ. Then we will apply Chebyshev's inequality to $\mathbb{E}[\mathcal{Z}(G(n,m))|\mathfrak{F}_\ell]$ to derive that

$$\mathcal{Z}(G(n,m)) \sim \mathbb{E}[\mathcal{Z}(G(n,m))|\mathfrak{F}_\ell]$$

w.h.p. in the limit of large ℓ, n. Formally, we will prove the following.

Lemma 5.12 (**SYM**, **BAL**, **MIN**, **UNI**). *For any $\eta > 0$ there exists $\ell_0(\eta)$ such that for every $\ell > \ell_0(\eta)$ uniformly for all $m \in \mathcal{M}(d)$,*

$$\lim_{n\to\infty} \mathbb{P}[|\mathcal{Z}(G(n,m)) - \mathbb{E}[\mathcal{Z}(G(n,m))|\mathfrak{F}_\ell]| > \eta \mathbb{E}[Z(G(n,m))]] = 0.$$

We prove Lemma 5.12 by way of the basic identity

$$\mathrm{Var}[\mathcal{Z}(G(n,m))] = \mathrm{Var}(\mathbb{E}[\mathcal{Z}(G(n,m))|\mathfrak{F}_\ell]) + \mathbb{E}[\mathrm{Var}(\mathcal{Z}(G(n,m))|\mathfrak{F}_\ell)]. \tag{5.26}$$

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

Due to (5.26), to prove that $\mathbb{E}[\mathrm{Var}(\mathcal{Z}(G(n,m))|\mathfrak{F}_\ell)]$ is small it suffices to show that

$$\mathrm{Var}(\mathbb{E}[\mathcal{Z}(G(n,m))|\mathfrak{F}_\ell]) = \mathbb{E}[\mathbb{E}[\mathcal{Z}(G(n,m))|\mathfrak{F}_\ell]^2] - \mathbb{E}[\mathcal{Z}(G(n,m))]^2 \tag{5.27}$$

is nearly as big as $\mathrm{Var}[\mathcal{Z}(G(n,m))]$, so that will be our first intermediate goal. We begin with the following little calculation. Let $\delta_Y = \mathrm{tr}(\Phi_Y) - 1 = (\hat{\kappa}_Y - \kappa_Y)/\kappa_Y$.

Lemma 5.13 (**SYM**, **BAL**). *We have*

$$\sum_{\ell \geqslant 1} \sum_{Y \in \mathcal{Y}_{\leqslant \ell}} \delta_Y^2 \kappa_Y = -\frac{1}{2} \sum_{\lambda \in \mathrm{Eig}^*(\Xi)} \ln(1 - d(k-1)\lambda).$$

Proof. The proof is essentially identical to that of [25, Lemma 9.1]. Let

$$\boldsymbol{\Phi}_\ell = \prod_{i=1}^{\ell} \Phi_{\boldsymbol{\psi}_i}.$$

Then

$$\sum_{Y \in \mathcal{Y}_{\leqslant \ell}} \delta_Y^2 \kappa_Y = \sum_{Y \in \mathcal{Y}_{\leqslant \ell}} \frac{(\hat{\kappa}_Y - \kappa_Y)^2}{\kappa_Y} = \sum_{j=1}^{\ell} \frac{(d(k-1))^j}{2j} \mathbb{E}[(\mathrm{tr}\, \boldsymbol{\Phi}_j - 1)^2]. \tag{5.28}$$

Hence, applying (2.8), (2.10) and Lemma 5.4, we obtain

$$\mathbb{E}[(\mathrm{tr}\, \boldsymbol{\Phi}_j - 1)^2] = \mathrm{tr}\, \mathbb{E}[\boldsymbol{\Phi}_j \otimes \boldsymbol{\Phi}_j] - 2\, \mathrm{tr}\, \mathbb{E}[\boldsymbol{\Phi}_j] + 1 = \mathrm{tr}(\Xi^j) - 2\, \mathrm{tr}(\Phi^j) + 1. \tag{5.29}$$

Finally, since

$$\mathrm{tr}\,(\Xi^j) = \sum_{\lambda \in \mathrm{Eig}(\Xi)} \lambda^j = 1 + 2 \sum_{\lambda \in \mathrm{Eig}(\Phi) \setminus \{1\}} \lambda^j + \sum_{\lambda \in \mathrm{Eig}^*(\Xi)} \lambda^j = -1 + 2\, \mathrm{tr}(\Phi^j) + \sum_{\lambda \in \mathrm{Eig}^*(\Xi)} \lambda^j,$$

combining (5.28) and (5.29) gives

$$\sum_{Y \in \mathcal{Y}_{\leqslant \ell}} \frac{(\hat{\kappa}_Y - \kappa_Y)^2}{\kappa_Y} = \sum_{j=1}^{\ell} \sum_{\lambda \in \mathrm{Eig}^*(\Xi)} \frac{(d(k-1)\lambda)^j}{2j}. \tag{5.30}$$

Proposition 5.5 shows $d(k-1) \max_{\lambda \in \mathrm{Eig}^*(\Xi)} |\lambda| < 1$ for $d < d_{\mathrm{cond}}$, and thus we may take ℓ to infinity in (5.30). □

Lemma 5.14 (**SYM**, **BAL**, **MIN**, **UNI**). *Suppose that* $0 < d < d_{\mathrm{cond}}$, $\ell > 0$. *Then, uniformly for all* $m \in \mathcal{M}(d)$,

$$\mathbb{E}[\mathbb{E}[\mathcal{Z}(G(n,m))|\mathfrak{F}_\ell]^2] \geqslant (1 + o(1))\mathbb{E}[Z(G(n,m))]^2 \cdot \exp \sum_{Y \in \mathcal{Y}_{\leqslant \ell}} \delta_Y^2 \kappa_Y.$$

Proof. Fix a number $\alpha > 0$, pick $B = B(\alpha, \ell) > 0$ large, let

$$C = \{(c_Y)_{Y \in \mathcal{Y}_{\leqslant \ell}} \in \mathbb{Z}^{\mathcal{Y}_{\leqslant \ell}} : 0 \leqslant c_Y \leqslant B \text{ for all } Y \in \mathcal{Y}_{\leqslant \ell}\}$$

and let

$$\mathcal{C} = \{(C_Y(G(n,m)))_{Y \in \mathcal{Y}_{\leqslant \ell}} \in C\}.$$

Then (5.5) yield

$$\frac{\mathbb{E}[\mathbf{1}\mathcal{C} \cdot \mathbb{E}[Z(G(n,m))|\mathfrak{F}_\ell]^2]}{\mathbb{E}[Z(G(n,m))]^2} = \sum_{c \in \mathcal{C}} \frac{\mathbb{P}[\forall Y \in \mathcal{Y}_{\leqslant \ell} : C_Y(\hat{G}(n,m)) = c_Y]^2}{\mathbb{P}[\forall Y \in \mathcal{Y}_{\leqslant \ell} : C_Y(G(n,m)) = c_Y]}. \tag{5.31}$$

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

Proposition 5.11 yields

$$\mathbb{P}[\forall Y \in \mathcal{Y}_{\leqslant \ell} : C_Y(G(n,m)) = c_Y] \sim \prod_Y \mathbb{P}[\mathrm{Po}(\kappa_Y) = c_Y]$$

uniformly for all $c \in \mathcal{C}$. Similarly, if $c_Y = 0$ for all Y with $\hat{\kappa}_Y = 0$, then Proposition 5.11 yields

$$\mathbb{P}[\forall Y \in \mathcal{Y}_{\leqslant \ell} : C_Y(\hat{G}(n,m)) = c_Y] \sim \prod_Y \mathbb{P}[\mathrm{Po}(\hat{\kappa}_Y) = c_Y].$$

By contrast, if $c_Y > 0$ for some Y with $\hat{\kappa}_Y = 0$, then

$$\mathbb{P}[\forall Y \in \mathcal{Y}_{\leqslant \ell} : C_Y(\hat{G}(n,m)) = c_Y] = 0.$$

Thus, (5.31) gives

$$\begin{aligned}\frac{\mathbb{E}[\mathbf{1}\mathcal{C} \cdot \mathbb{E}[Z(G(n,m))|\mathfrak{F}_\ell]^2]}{\mathbb{E}[Z(G(n,m))]^2} &\sim \sum_{c \in \mathcal{C}} \prod_{Y \in \mathcal{Y}_{\leqslant \ell}} \frac{\mathbb{P}[\mathrm{Po}((1+\delta_Y)\kappa_Y) = c_Y]^2}{\mathbb{P}[\mathrm{Po}(\kappa_Y) = c_Y]} \\ &= \exp\Big[-\sum_{Y \in \mathcal{Y}_{\leqslant \ell}} (1+2\delta_Y)\kappa_Y\Big] \sum_{c \in \mathcal{C}} \prod_{Y \in \mathcal{Y}_{\leqslant \ell}} \frac{((1+\delta_Y)^2\kappa_Y)^{c_Y}}{c_Y!}. \end{aligned} \tag{5.32}$$

Choosing B sufficiently large, we can ensure that

$$\sum_{c \in \mathcal{C}} \prod_{Y \in \mathcal{Y}_{\leqslant \ell}} [((1+\delta_Y)^2\kappa_Y)^{c_Y}/c_Y!] \geqslant \exp\Big(-\alpha/2 + \sum_{Y \in \mathcal{Y}_\ell} (1+\delta_Y)^2\kappa_Y\Big).$$

Hence, (5.32) implies that for large n,

$$\frac{\mathbb{E}[\mathbf{1}\mathcal{C} \cdot \mathbb{E}[Z(G(n,m))|\mathfrak{F}_\ell]^2]}{\mathbb{E}[Z(G(n,m))]^2} \geqslant \exp\Big[-\alpha + \sum_{Y \in \mathcal{Y}_\ell} \delta_Y^2\kappa_Y\Big]. \tag{5.33}$$

Further, as $0 \leqslant \mathcal{Z}(G(n,m)) \leqslant Z(G(n,m))$,

$$\begin{aligned}&\mathbb{E}[\mathbf{1}\mathcal{C} \cdot (\mathbb{E}[Z(G(n,m))|\mathfrak{F}_\ell]^2 - \mathbb{E}[\mathcal{Z}(G(n,m))|\mathfrak{F}_\ell]^2)] \\ &\quad= \mathbb{E}[\mathbf{1}\mathcal{C} \cdot (\mathbb{E}[Z(G(n,m))|\mathfrak{F}_\ell] + \mathbb{E}[\mathcal{Z}(G(n,m))|\mathfrak{F}_\ell])(\mathbb{E}[Z(G(n,m))|\mathfrak{F}_\ell] - \mathbb{E}[\mathcal{Z}(G(n,m))|\mathfrak{F}_\ell])] \\ &\quad\leqslant 2\|\mathbf{1}\mathcal{C} \cdot \mathbb{E}[Z(G(n,m))|\mathfrak{F}_\ell]\|_\infty \mathbb{E}[\mathbb{E}[Z(G(n,m))|\mathfrak{F}_\ell] - \mathbb{E}[\mathcal{Z}(G(n,m))|\mathfrak{F}_\ell]]. \end{aligned} \tag{5.34}$$

Since B is (large but) fixed, Proposition 5.11 yields

$$\|\mathbf{1}\mathcal{C} \cdot \mathbb{E}[Z(G(n,m))|\mathfrak{F}_\ell]\|_\infty \leqslant O(\mathbb{E}[Z(G(n,m))]),$$

whereas Corollary 5.8 shows

$$\mathbb{E}[\mathbb{E}[Z(G(n,m))|\mathfrak{F}_\ell] - \mathbb{E}[\mathcal{Z}(G(n,m))|\mathfrak{F}_\ell]] = o(\mathbb{E}[Z(G(n,m))]).$$

Plugging these estimates into (5.34), we get

$$\mathbb{E}[\mathbf{1}\mathcal{C} \cdot (\mathbb{E}[Z(G(n,m))|\mathfrak{F}_\ell]^2 - \mathbb{E}[\mathcal{Z}(G(n,m))|\mathfrak{F}_\ell]^2)] = o(\mathbb{E}[Z(G(n,m))]).$$

Thus, the lemma follows from (5.33). □

Proof of Lemma 5.12. Given $\eta > 0$ choose $\alpha = \alpha(\eta) > 0$ small enough. We introduce the auxiliary random variable

$$\begin{aligned}X(G(n,m)) = &|\mathcal{Z}(G(n,m)) - \mathbb{E}[\mathcal{Z}(G(n,m))|\mathfrak{F}_\ell]| \\ &\times \mathbf{1}\{|\mathcal{Z}(G(n,m)) - \mathbb{E}[\mathcal{Z}(G(n,m))|\mathfrak{F}_\ell]| > \alpha^{1/3}\mathbb{E}[Z(G(n,m))]\}\end{aligned}$$

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

so that

$$X(G(n,m)) < \alpha^{1/3}\mathbb{E}[Z(G(n,m))]$$
$$\Rightarrow\ |\mathcal{Z}(G(n,m)) - \mathbb{E}[\mathcal{Z}(G(n,m))|\mathfrak{F}_\ell]| \leqslant \alpha^{1/3}\mathbb{E}[Z(G(n,m))]. \tag{5.35}$$

Combining (5.21), (5.27) and Lemmas 5.13 and 5.14, we obtain

$$\mathbb{E}[\mathrm{Var}[\mathcal{Z}(G(n,m))|\mathfrak{F}_\ell]] < \alpha\mathbb{E}[Z(G(n,m))]^2,$$

providing ℓ, n are large enough. Therefore, Chebyshev's inequality yields

$$\begin{aligned}
&\mathbb{E}[X(G(n,m))]\\
&\quad\leqslant \alpha^{1/3}\mathbb{E}[Z(G(n,m))]\sum_{j\geqslant 0} 2^{j+1}\mathbb{P}[X(G(n,m)) > 2^j\alpha^{1/3}\mathbb{E}[Z(G(n,m))]]\\
&\quad\leqslant \alpha^{1/3}\mathbb{E}[Z(G(n,m))]\sum_{j\geqslant 0} 2^{j+1}\mathbb{P}[|\mathcal{Z}(G(n,m)) - \mathbb{E}[\mathcal{Z}(G(n,m))|\mathfrak{F}_\ell]| > 2^j\alpha^{1/3}\mathbb{E}[Z(G(n,m))]]\\
&\quad\leqslant 4\alpha^{-1/3}\mathbb{E}[Z(G(n,m))]\cdot\mathbb{E}\left[\frac{\mathrm{Var}[\mathcal{Z}(G(n,m))|\mathfrak{F}_\ell]}{\mathbb{E}[Z(G(n,m))]^2}\right]\\
&\quad\leqslant 4\alpha^{2/3}\mathbb{E}[Z(G(n,m))].
\end{aligned} \tag{5.36}$$

Finally, the assertion follows from (5.35), (5.36) and Markov's inequality. □

Proof of Theorem 2.7. Let $(K_Y)_{Y\geqslant 1}$ be a family of mutually independent Poisson variables with means $\mathbb{E}[K_Y] = \kappa_Y$, let $(K_j)_{j\geqslant 1}$ be mutually independent Poisson variables with means $\mathbb{E}[K_j] = (d(k-1))^j/(2j)$ and let $(\boldsymbol{\psi}_{h,i,j})_{h,i,j\geqslant 1}$ be a family of samples from P, mutually independent and independent of the K_j. We first use an argument from [55] to show that the random variable $\mathcal{K}$ from Theorem 2.7 is well-defined. Let $\ell \geqslant 1$. Then (5.23) shows that the random variables

$$\mathcal{K}'_\ell = \prod_{Y\in\mathcal{Y}_\ell}\frac{(\mathrm{tr}\Phi_Y)^{K_Y}}{\exp(\kappa_Y\delta_Y)},\quad \mathcal{K}_\ell = \exp\left[\frac{(d(k-1))^\ell}{2\ell}(1-\mathrm{tr}(\Phi^\ell))\right]\prod_{i=1}^{K_\ell}\mathrm{tr}\prod_{j=1}^{\ell}\Phi_{\boldsymbol{\psi}_{\ell,i,j}}$$

are identically distributed. Further, since $\mathbb{E}[(\mathrm{tr}\Phi_Y)^{K_Y}] = \exp(\kappa_Y\delta_Y)$ and because the K_Y are mutually independent, we have $\mathbb{E}[\mathcal{K}_\ell] = \mathbb{E}[\mathcal{K}'_\ell] = 1$. Therefore, the random variables $\mathcal{K}_{\leqslant\ell} = \prod_{l\leqslant\ell}\mathcal{K}_l$ form a martingale. Additionally, since $\mathbb{E}[(\mathrm{tr}\Phi_Y)^{2K_Y}] = \exp(2\kappa_Y\delta_Y + \kappa_Y\delta_Y^2)$, Lemma 5.13 shows that the martingale is L_2-bounded. Therefore, $(\mathcal{K}_{\leqslant\ell})_{\ell\geqslant 1}$ converges to a limit $\mathcal{K}_*$ almost surely and in L_2. The random variable $\mathcal{K}$ is obtained from $\mathcal{K}_*$ by disregarding the factors $\ell = 1$ and $\ell = 2$ if $k = 2$.

As a next step we show that $\mathcal{K} > 0$ almost surely (this is where there is a significant difference between hard constraints and soft ones). There are two cases to consider. First, assume that $d < d_{\mathrm{cond}} \leqslant (k-1)^{-1}$. Then $\sum_{\ell\geqslant 1}\mathbb{E}[K_\ell] = O(1)$. Consequently, for any $\varepsilon > 0$ we can find $L > 0$ such that $\mathbb{P}[\forall\ell > L : K_\ell = 0] > 1-\varepsilon$. But given that $K_\ell = 0$ for all $\ell > L$, $\mathcal{K}$ is a finite product of positive terms, and thus $\mathcal{K}$ is positive. Next, suppose that $d_{\mathrm{cond}} > (k-1)^{-1}$. Then Lemma 5.13 implies that $\sum_{Y\in\mathcal{Y}}\delta_Y^2 < \infty$. Hence, there exists $\ell_0 > 1$ such that for all $\ell > \ell_0$ and all $Y\in\mathcal{Y}_\ell$ we have $|\delta_Y| \leqslant 1/2$. Thus, for $\ell > \ell_0$ we obtain

$$\mathbb{E}[\mathcal{K}_\ell^{-1}] = \prod_{Y\in\mathcal{Y}_\ell}\frac{\exp(\kappa_Y\delta_Y)}{(1+\delta_Y)^{K_Y}} = \exp\left[\sum_{Y\in\mathcal{Y}_\ell}\frac{\kappa_Y\delta_Y^2}{1+\delta_Y}\right] \leqslant \exp\left[4\sum_{Y\in\mathcal{Y}_\ell}\kappa_Y\delta_Y^2\right].$$

Consequently, Lemma 5.13 shows that the expected reciprocals $\mathbb{E}[\mathcal{K}_{\leqslant\ell}^{-1}]$ remain bounded for all ℓ, whence $\mathcal{K} > 0$ almost surely.

To complete the proof of Theorem 2.7, we recall that

$$\mathbb{E}|\mathcal{Z}(G(n,m)) - Z(G(n,m))| = o(\mathbb{E}[Z(G(n,m))])$$

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

by Corollary 5.8. Hence, Lemma 5.12 yields

$$\lim_{n\to\infty} \mathbb{P}[|Z(G(n,m)) - \mathbb{E}[Z(G(n,m))|\mathfrak{F}_\ell]| > \eta\mathbb{E}[Z(G(n,m))]] = 0 \quad \text{for any } \eta > 0. \tag{5.37}$$

Further, by Proposition 5.11 the conditional expectation $\mathbb{E}[Z(G(n,m))|\mathfrak{F}_\ell]$ is distributed as follows: for any non-negative integer vector $(c_Y)_{Y\in\mathcal{Y}_{\leqslant\ell}}$ such that $c_Y = 0$ if $\hat{\kappa}_Y = 0$ we have

$$\begin{aligned}
&\frac{\mathbb{E}[Z(G(n,m)) \mid \forall Y \in \mathcal{Y}_{\leqslant\ell} : C_Y(G(n,m)) = c_Y]}{\mathbb{E}[Z(G(n,m))]} \\
&= \frac{\mathbb{P}[\forall Y \in \mathcal{Y}_{\leqslant\ell} : C_Y(\hat{G}(n,m)) = c_Y]}{\mathbb{P}[\forall Y \in \mathcal{Y}_{\leqslant\ell} : C_Y(G(n,m)) = c_Y]} \quad \text{(by (5.5))} \\
&\sim \prod_{Y\in\mathcal{Y}_{\leqslant\ell}} \frac{\mathbb{P}[\mathrm{Po}(\hat{\kappa}_Y) = c_Y]}{\mathbb{P}[\mathrm{Po}(\kappa_Y) = c_Y]} \\
&= \prod_{Y\in\mathcal{Y}_{\leqslant\ell}} \frac{(\mathrm{tr}\Phi_Y)^{c_Y}}{\exp(\hat{\kappa}_Y - \kappa_Y)},
\end{aligned} \tag{5.38}$$

while

$$\mathbb{E}[Z(G(n,m)) \mid \forall Y \in \mathcal{Y}_{\leqslant\ell} : C_Y(G(n,m)) = c_Y] = 0$$

if $c_Y > 0$ for some signature Y with $\hat{\kappa}_Y = 0$. Indeed, Proposition 5.11 shows that $\hat{\kappa}_Y = 0$ can only occur for signatures of order one, and for such signatures we obtain $\mathrm{tr}\Phi_Y = 0$. Consequently, the conditional expectation is given by (5.38) in all cases. In other words, letting $Q_\ell(G(n,m)) = \mathbb{E}[Z(G(n,m))|\mathfrak{F}_\ell]/\mathbb{E}[Z(G(n,m))]$, we conclude that

$$Q_\ell(G(n,m)) \overset{n\to\infty}{\to} W_{\leqslant\ell}(G(n,m)) = \prod_{Y\in\mathcal{Y}_{\leqslant\ell}} \frac{(\mathrm{tr}\Phi_Y)^{C_Y(G(n,m))}}{\exp(\hat{\kappa}_Y - \kappa_Y)} \tag{5.39}$$

in probability. Therefore, Proposition 5.11 implies that $Q_\ell(G(n,m))$ converges to $\mathcal{K}_{\leqslant\ell}$ in distribution for every $\ell \geqslant 1$. Since $(\mathcal{K}_{\leqslant\ell})_\ell$ converges to $\mathcal{K}_*$ almost surely and in L_2, (5.37) shows that for any bounded continuous $g : \mathbb{R} \to \mathbb{R}$,

$$\begin{aligned}
&\forall \varepsilon > 0\, \exists \ell_0(\varepsilon)\, \forall \ell \geqslant \ell_0(\varepsilon) : \limsup_{n\to\infty} \mathbb{E}[g(\mathcal{K}_*)] - \mathbb{E}[g(\mathcal{K}_{\leqslant\ell})] < \varepsilon, \\
&\forall \varepsilon > 0\, \exists \ell_0'(\varepsilon)\, \forall \ell \geqslant \ell_0'(\varepsilon) : \limsup_{n\to\infty} \mathbb{E}[g(\mathcal{K}_{\leqslant\ell})] - \mathbb{E}\left[g\left(\frac{Z(G(n,m))}{\mathbb{E}[Z(G(n,m))]}\right)\right] < \varepsilon.
\end{aligned}$$

Combining these two statements, we conclude that $Z(G(n,m))/\mathbb{E}[Z(G(n,m))]$ converges to $\mathcal{K}_*$ in distribution. Further, as

$$\mathbb{P}[G(n,m) \in \mathfrak{S}\triangle\{C_1(G(n,m)) + \mathbf{1}\{k=2\}C_2(G(n,m)) = 0\}] = O(1/n),$$

we see that $Z(\mathbb{G}(n,m))/\mathbb{E}[Z(\mathbb{G}(n,m))]$ converges to $\mathcal{K}$ in distribution. Finally, plugging in the formula for the first moment from (5.19) yields (2.11). □

6. The planted model

In this section we prove Theorem 5.3. Specifically, in Section 6.1–6.4 we prove via an adaptation of the interpolation argument from [27] that the functional $\mathcal{B}$ provides a lower bound on $\mathbb{E}[\ln Z(\hat{G})]$. Some of the intermediate steps of this proof will be re-used in Section 7. Subsequently, in Section 6.5 we show how the results from [25] can be combined with a limiting argument to derive a matching upper bound on $\mathbb{E}[\ln Z(\hat{G})]$.

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

6.1 The interpolation method

We are going to prove the following lower bound on $\mathbb{E}[\ln Z(\hat{G})]$. It is worth mentioning that here, in contrast to other applications of the interpolation method, the scheme gives a *lower* bound on the free energy. This is because we apply the interpolation method to a planted model; see [27] for more comments.

Proposition 6.1 (**SYM**, **BAL**, **POS**). *If $\pi \in \mathcal{P}_*^2(\Omega)$ is supported on a finite set, then*

$$\liminf_{n\to\infty} \frac{1}{n}\mathbb{E}[\ln Z(\hat{G})] \geqslant \mathcal{B}(d, P, \pi).$$

We prove Proposition 6.1 via the interpolation method. Specifically, we adapt the interpolation argument developed in [27] for the case of soft constraints. The basic idea is to construct a family of random CSPs, parametrized by $t \in [0, 1]$, such that for $t = 1$ the model coincides with $\hat{G}$, while for $t = 0$ the CSP is so simple that we can calculate the partition function easily. Indeed, we will see that the logarithm of the partition function at $t = 0$ is asymptotically equal to $n\mathcal{B}(d, P, \pi)$ w.h.p. To obtain the desired lower bound on $\mathbb{E}[\ln Z(\hat{G})]$ we will prove that the mean of the logarithm of the partition function is a monotonically increasing function of t.

The intermediate models parametrized by $t \in [0, 1]$ comprise a blend of unary and k-ary constraints, and t governs the proportion of k-ary constraints. Thus, at $t = 0$ all constraints are unary, whereas at $t = 1$ there are k-ary constraints only. This interpolating family is best introduced by way of the following generalized random CSP. Suppose that $\pi \in \mathcal{P}_*^2(\Omega)$ has finite support. Moreover, let $\gamma = (\gamma_v)_{v\in[n]}$ be a sequence of integers, let $\theta \geqslant 0$ be an integer and let $U \subset [n]$. Define a random CSP $G(n, m, \gamma, \pi, U)$ with variables $V_n = \{x_1, \ldots, x_n\}$, k-ary constraints $a_1, \ldots, a_m$ and unary constraints $(b_{i,j})_{i\in[n],j\in[\gamma_i]}$, $(c_i)_{i\in U}$, all chosen mutually independently, as follows.

INT1. For $i \in [m]$ choose $\partial a_i \in V_n^k$ uniformly and independently pick $\psi_{a_i} \in \Psi$ from the distribution P.

INT2. For $i \in [n]$ and $j \in [\gamma_i]$ the constraint $b_{i,j}$ is adjacent to x_i only. The random function $\psi_{b_{i,j}}$ is defined as follows: with $(\boldsymbol{\rho}_{i,j,h})_{h\in[k-1]}$ drawn from π and $\boldsymbol{\psi}_{i,j}$ drawn from P mutually independently, let

$$\psi_{b_{i,j}}(\sigma) = \sum_{\tau_1,\ldots,\tau_{k-1}\in\Omega} \boldsymbol{\psi}_{i,j}(\tau_1, \ldots, \tau_{k-1}, \sigma) \prod_{h=1}^{k-1} \boldsymbol{\rho}_{i,j,h}(\tau_h) \quad (\sigma \in \Omega).$$

INT3. For $i \in U$ the unary constraint c_i is adjacent to x_i and for a uniformly random $\boldsymbol{\chi}_i \in \Omega$ we let

$$\psi_{c_i}(\sigma) = \mathbf{1}\{\sigma = \boldsymbol{\chi}_i\}.$$

Thus, $a_1, \ldots, a_m$ are chosen just as the constraints of $G(n, m)$. Moreover, the unary constraints $b_{i,j}$ acting on x_i come with random constraint functions $\boldsymbol{\psi}_{i,j}$ whose other $k - 1$ inputs are drawn independently from the distributions $\boldsymbol{\rho}_{i,j,1}, \ldots, \boldsymbol{\rho}_{i,j,k-1}$. Finally, the constraints c_i simply peg variable x_i to a specific value $\boldsymbol{\chi}_i$.

Like in Section 5.1 we consider several assorted random CSP models, such as a planted version of $G(n, m, \gamma, \pi, U)$. First, given an integer $0 \leqslant \theta \leqslant n$ let U denote a random subset of $[n]$ of size θ and let $G(n, m, \gamma, \pi, \theta) = G(n, m, \gamma, \pi, U)$. Thus, in $G(n, m, \gamma, \pi, \theta)$ we peg a random set of θ variables. Further, let $\hat{G}(n, m, \gamma, \pi, \theta)$ be the random CSP obtained by reweighing $G(n, m, \gamma, \pi, \theta)$ according to its partition function: for any possible outcome G of $G(n, m, \gamma, \pi, \theta)$ let

$$\mathbb{P}[\hat{G}(n, m, \gamma, \pi, \theta) = G] = \frac{Z(G) \cdot \mathbb{P}[G(n, m, \gamma, \pi, \theta) = G]}{\mathbb{E}[Z(G(n, m, \gamma, \pi, \theta))]}. \tag{6.1}$$

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

The denominator is positive for all $n \geqslant q$ because of **SYM** and because $\int_{\mathcal{P}(\Omega)} \rho \mathrm{d}\pi(\rho)$ is the uniform distribution on Ω. Further, by extension of (5.4) we define a distribution on assignments by letting

$$\mathbb{P}[\hat{\boldsymbol{\sigma}}_{n,m,\gamma,\pi,\theta} = \sigma] = \frac{\mathbb{E}[\psi_{G(n,m,\gamma,\pi,\theta)}(\sigma)]}{\mathbb{E}[Z(G(n,m,\gamma,\pi,\theta))]} \quad \text{for any } \sigma \in \Omega^{V_n}. \tag{6.2}$$

Additionally, let $\Sigma(n,m,\gamma,\pi,\theta) \subset \Omega^{V_n}$ be the support of $\hat{\boldsymbol{\sigma}}_{n,m,\gamma,\pi,\theta}$. Then for $\sigma \in \Sigma(n,m,\gamma,\pi,\theta)$ we define, by extension of (5.2), a planted random CSP by letting

$$\mathbb{P}[G^*(n,m,\gamma,\pi,\theta,\sigma) = G] = \frac{\psi_G(\sigma)\mathbb{P}[G(n,m,\gamma,\pi,\theta) = G]}{\mathbb{E}[\psi_{G(n,m,\gamma,\pi,\theta)}(\sigma)]} \tag{6.3}$$

for any possible outcome G of $G(n,m,\gamma,\pi,\theta)$.

We obtain the interpolating family of random CSPs by choosing the parameters m,γ,θ as appropriate random variables parametrized by t. Specifically, given $d > 0$ and $t \in [0,1]$ the number m_t of k-ary constraints has distribution $\mathrm{Po}(tdn/k)$. Moreover, for each $i \in [n]$ let $\boldsymbol{\gamma}_{t,i}$ have distribution $\mathrm{Po}((1-t)d)$ and let $\boldsymbol{\gamma}_t = (\boldsymbol{\gamma}_{t,i})_{i\in[n]}$. Additionally, let $\boldsymbol{\theta}_\varepsilon$ be distributed as the random variable from Lemma 4.6, with the convention that $\boldsymbol{\theta}_1 = 0$. All of these random variables are mutually independent. Finally, we let

$$\begin{aligned} G_{t,\varepsilon} &= G(n, m_t, \boldsymbol{\gamma}_t, \pi, \boldsymbol{\theta}_\varepsilon), & \hat{G}_{t,\varepsilon} &= \hat{G}(n, m_t, \boldsymbol{\gamma}_t, \pi, \boldsymbol{\theta}_\varepsilon), \\ \hat{\boldsymbol{\sigma}}_{t,\varepsilon} &= \hat{\boldsymbol{\sigma}}_{n,m_t,\boldsymbol{\gamma}_t,\pi,\boldsymbol{\theta}_\varepsilon}, & G^*_{t,\varepsilon} &= G^*(n, m_t, \boldsymbol{\gamma}_t, \pi, \boldsymbol{\theta}_\varepsilon, \hat{\boldsymbol{\sigma}}_{t,\varepsilon}). \end{aligned}$$

The following proposition provides the monotonicity in t that we alluded to above.

Proposition 6.2 (**SYM**, **BAL**, **POS**). *For every $\delta > 0$ there is $\varepsilon > 0$ such that, for large enough n, the following holds. Let*

$$\Gamma_t = \frac{td(k-1)}{k\xi}\mathbb{E}\left[\Lambda\left(\sum_{\tau\in\Omega^k} \boldsymbol{\psi}(\tau)\prod_{j=1}^{k} \boldsymbol{\rho}_j^{(\pi)}(\tau_j)\right)\right].$$

and define

$$\phi_\varepsilon(t) = \mathbb{E}[\ln Z(\hat{G}_{t,\varepsilon})]/n + \Gamma_t \quad \textit{for } t \in [0,1].$$

Then

$$\frac{\partial}{\partial t}\phi_\varepsilon(t) > -\delta \quad \textit{for all } t \in (0,1).$$

We observe that the random CSP $\hat{G}_{1,\varepsilon}$ at $t = 1$ contains $\mathrm{Po}(dn/k)$ k-ary constraints as well as a bounded number $\boldsymbol{\theta}_\varepsilon$ of unary constraints as per **INT3**. As we will see shortly, this implies that $\mathbb{E}[\ln Z(\hat{G}_{1,\varepsilon})] \leqslant \mathbb{E}[\ln Z(\hat{G})]$. Therefore, Proposition 6.2 shows that for any fixed $\delta > 0$ for large enough n,

$$\frac{1}{n}\mathbb{E}[\ln Z(\hat{G})] \geqslant \frac{1}{n}\mathbb{E}[\ln Z(\hat{G}_{0,\varepsilon})] - \Gamma_1 - \delta. \tag{6.4}$$

Further, $\hat{G}_{0,\varepsilon}$ consists of unary constraints only, and thus $\mathbb{E}[\ln Z(\hat{G}_{0,\varepsilon})]$ is going to be easy to compute. Hence, we will ultimately obtain Proposition 6.1 from (6.4).

But first we need to prove Proposition 6.2. In the special case of soft constraints (*i.e.* $\psi > 0$ for all $\psi \in \Psi$) the above construction of the interpolating family $\hat{G}_{t,\varepsilon}$ is identical to the one from [27], and Proposition 6.2 comes down to [27, Proposition 3.25]. In fact, the proof of Proposition 6.2 re-uses several of the steps and arguments from [27]. But the presence of hard constraints causes

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

subtle difficulties. This is because in order to calculate the derivative of $\phi_\varepsilon(t)$ we need to investigate the impact of adding a further random constraint to the random CSP instance $\hat{G}_{t,\varepsilon}$ on the logarithm of the partition function. Clearly, in the case of soft constraints the impact of a single constraint is bounded. But this need not be true in the case of hard constraints, and new arguments are required to deal with this issue. We will come to this in Section 6.3, just after establishing some basic facts about $\hat{G}_{t,\varepsilon}$. Then we will complete the proofs of Propositions 6.1 and 6.2 in Section 6.4.

6.2 Groundwork

Toward the proof of Proposition 6.2 we need a few basic observations regarding the probability distributions from the previous section. All of the following results are straightforward adaptations of the corresponding soft constraint versions from [27]. We begin with the following extension of the Nishimori identity.

Lemma 6.3. *For any G, σ we have*

$$\mathbb{P}[\hat{\boldsymbol{\sigma}}_{n,m,\gamma,\pi,\theta} = \sigma] \cdot \mathbb{P}[G^*(n,m,\gamma,\pi,\theta,\sigma) = G] = \mu_G(\sigma)\mathbb{P}[\hat{G}(n,m,\gamma,\pi,\theta) = G].$$

Proof. The proof is essentially identical to that of Lemma 5.1: (6.1), (6.2) and (6.3) yield

$$\begin{aligned}\mathbb{P}[\hat{\boldsymbol{\sigma}}_{n,m,\gamma,\pi,\theta} = \sigma] \cdot \mathbb{P}[G^*(n,m,\gamma,\pi,\theta,\sigma) = G] &= \frac{\psi_G(\sigma)\mathbb{P}[G(n,m,\gamma,\pi,\theta) = G]}{\mathbb{E}[Z(G(n,m,\gamma,\pi,\theta))]} \\ &= \mu_G(\sigma) \cdot \frac{Z(G)\mathbb{P}[G(n,m,\gamma,\pi,\theta) = G]}{\mathbb{E}[Z(G(n,m,\gamma,\pi,\theta))]} \\ &= \mu_G(\sigma)\mathbb{P}[\hat{G}(n,m,\gamma,\pi,\theta) = G],\end{aligned}$$

as desired. □

We are going to apply the Nishimori identity as follows. Suppose that $F(\sigma_0,\ldots,\sigma_\ell)$ is a function of $\ell+1$ assignments. Then Lemma 6.3 yields

$$\begin{aligned}\mathbb{E}\langle F(\boldsymbol{\sigma}_0,\ldots,\boldsymbol{\sigma}_\ell)\rangle_{\hat{G}(n,m,\gamma,\pi,\theta)} &= \sum_{\sigma_0\in\Omega^n} \mathbb{E}[\mu_{\hat{G}(n,m,\gamma,\pi,\theta)}(\sigma_0)\langle F(\sigma_0,\boldsymbol{\sigma}_1,\ldots,\boldsymbol{\sigma}_\ell)\rangle_{\hat{G}(n,m,\gamma,\pi,\theta)}] \\ &= \mathbb{E}\langle F(\hat{\boldsymbol{\sigma}}_{n,m,\gamma,\pi,\theta},\boldsymbol{\sigma}_1,\ldots,\boldsymbol{\sigma}_\ell)\rangle_{G^*(n,m,\gamma,\pi,\theta,\hat{\boldsymbol{\sigma}}_{n,m,\gamma,\pi,\theta})}. \qquad (6.5)\end{aligned}$$

Of course, in order to put (6.5) to work we need to get a handle on the distribution of $\hat{\boldsymbol{\sigma}}_{n,m,\gamma,\pi,\theta}$.

Lemma 6.4 (**SYM**). *For any assignment $\sigma \in \Omega^{V_n}$ we have*

$$\mathbb{E}[\psi_{G(n,m,\gamma,\pi,\theta)}(\sigma)] = q^{-\theta}\xi^{\sum_{v\in V}\gamma_v}\phi(\rho_\sigma)^m. \qquad (6.6)$$

In particular, $\hat{\boldsymbol{\sigma}}_{n,m,\gamma,\pi,\theta}$ and $\hat{\boldsymbol{\sigma}}_{n,m,\gamma',\pi,\theta'}$ are identically distributed for all $\gamma,\gamma',\theta,\theta'$.

Proof. The last factor in (6.6) emerges due to (5.1), because the k-ary constraints $a_1,\ldots,a_m$ are mutually independent and also the functions ψ_{a_i} and are independent of the neighbourhoods ∂a_i by **INT1**. Similarly, step **INT2** of the construction gives rise to the middle factor because the $\boldsymbol{\psi}_{i,j}$ are chosen independently of the $\boldsymbol{\rho}_{i,j,h}$ and $\mathbb{E}[\boldsymbol{\rho}_{i,j,h}(\tau)] = 1/q$ for every $\tau\in\Omega$. Hence, **SYM** yields $\mathbb{E}[\psi_{b_{i,j}}(\sigma)] = \xi$ for every $\sigma\in\Omega$. Finally, the factor $q^{-\theta}$ results from **INT3**. □

Corollary 6.5 (**SYM**, **BAL**). *Let $D>0$ and $\theta>0$. Then, uniformly for all $m \leqslant Dn/k$ and all γ, we have*

$$\mathbb{P}[\|\rho_{\hat{\boldsymbol{\sigma}}_{n,m,\gamma,\pi,\theta}} - \bar{\rho}\|_{\mathrm{TV}} > n^{-1/2}\ln n] \leqslant O(n^{-\ln n}).$$

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

Furthermore, for any $\eta > 0$ uniformly for all $m \leqslant Dn/k$ and all γ we have

$$\mathbb{P}[\|\rho_{\hat{\boldsymbol{\sigma}}_{n,m,\gamma,\pi,\theta}} - \bar{\rho}\|_{\mathrm{TV}} > \eta] \leqslant \exp(-\Omega(n)).$$

Proof. Let $\sigma \in \Omega^{V_n}$ and recall that $\rho_\sigma \in \mathcal{P}(\Omega)$ stands for the empirical distribution of σ. Lemma 6.4, (6.2) and (5.1) yield

$$\mathbb{P}[\hat{\boldsymbol{\sigma}}_{n,m,\gamma,\pi,\theta} = \sigma] = \mathbb{P}[\hat{\boldsymbol{\sigma}}_{n,m,0,\pi,0} = \sigma] = \frac{\phi(\rho_\sigma)^m}{\mathbb{E}[Z(G(n,m))]}$$

and **BAL** provides that the rightmost expression is concave in ρ_σ and attains its maximum at $\bar{\rho}$. □

Finally, we introduced the unary constraints from **INT3** in order to obtain the following.

Lemma 6.6. *For any $\varepsilon > 0$ there is $n_0 > 0$ such that for all $d > 0$, $t \in [0,1]$ we have*

$$\mathbb{P}[\mu_{\hat{G}_{t,\varepsilon}} \text{ is } \varepsilon\text{-symmetric}] \geqslant 1 - \varepsilon.$$

Proof. By Lemma 6.3 the random factor graph $\hat{G}_{t,\varepsilon}$ has the same distribution as $G^*_{t,\varepsilon}$. Spelling out (6.3) and using the second part of Lemma 6.3, we see that $G^*_{t,\varepsilon}$ is obtained by first drawing $G^*(n, m_t, \boldsymbol{\gamma}_t, \pi, 0, \hat{\boldsymbol{\sigma}}_{n,m_t,\boldsymbol{\gamma}_t,\pi,0})$ without pinning and subsequently pinning a random set U of $\boldsymbol{\theta}_\varepsilon$ variables to their planted values $\hat{\boldsymbol{\sigma}}_{n,m_t,\boldsymbol{\gamma}_t,\pi,0}$. Applying the first part of Lemma 6.3, we see that this experiment is equivalent to first generating a random factor graph $\hat{G}(n, m_t, \boldsymbol{\gamma}_t, \pi, 0)$, then drawing a sample $\boldsymbol{\sigma}$ from its Gibbs measure and subsequently pinning the variables in a random set U of size $\boldsymbol{\theta}_\varepsilon$ to the values $\boldsymbol{\sigma}(x_i)$, $i \in U$. This last experiment precisely matches the perturbation from Lemma 4.6, which therefore implies the assertion. □

6.3 Adding a constraint

As already mentioned, in order to prove Proposition 6.2 we basically need to study the impact of adding a single constraint to the random CSP $\hat{G}_{t,\varepsilon}$. The following proposition delivers this analysis. From here on we let $x_1, \ldots, x_k \in V_n$ denote a family of uniformly random variables, chosen mutually independently and independently of everything else.

Proposition 6.7 (**SYM**, **BAL**). *Let $D > 0$ and $\theta > 0$. Uniformly for all $m \leqslant Dn/k$ and all γ we have*

$$\begin{aligned}\mathbb{E}[\ln Z(\hat{G}(n, m+1, \gamma, \pi, \theta))] - \mathbb{E}[\ln Z(\hat{G}(n, m, \gamma, \pi, \theta))] \\ = o(1) + \xi^{-1}\mathbb{E}[\Lambda(\langle \boldsymbol{\psi}(\boldsymbol{\sigma}(x_1), \ldots, \boldsymbol{\sigma}(x_k))\rangle_{\hat{G}(n,m,\gamma,\pi,\theta)})].\end{aligned}$$

Proposition 6.7 extends [27, Proposition 3.30] from soft to hard constraints. To prove the proposition we need the following statement. The proof, although essentially identical to [27, Corollary 3.29], is included for the sake of completeness.

Lemma 6.8 (**SYM**, **BAL**). *Let $D > 0$ and $\theta > 0$. Uniformly for all $m \leqslant Dn/k$ and all γ the following is true. There is a coupling of $\hat{\boldsymbol{\sigma}}_{n,m,\gamma,\pi,\theta}$, $\hat{\boldsymbol{\sigma}}_{n,m+1,\gamma,\pi,\theta}$ such that*

$$\begin{aligned}\mathbb{P}[\hat{\boldsymbol{\sigma}}_{n,m,\gamma,\pi,\theta} \neq \hat{\boldsymbol{\sigma}}_{n,m+1,\gamma,\pi,\theta}] &= O(n^{-1}\ln^4 n), \\ \mathbb{P}[|\hat{\boldsymbol{\sigma}}_{n,m,\gamma,\pi,\theta} \triangle \hat{\boldsymbol{\sigma}}_{n,m+1,\gamma,\pi,\theta}| > \sqrt{n}\ln n] &= O(n^{-2}).\end{aligned}$$

Proof. The second bound is immediate from Corollary 6.5. To prove the first we bound the total variation distance of $\hat{\boldsymbol{\sigma}}_{n,m,\gamma,\pi,\theta}$, $\hat{\boldsymbol{\sigma}}_{n,m+1,\gamma,\pi,\theta}$. By Lemma 6.4 we may assume that $\theta = 0$, $\gamma = 0$, and thus $\hat{\boldsymbol{\sigma}}_{n,m,\gamma,\pi,\theta} = \hat{\boldsymbol{\sigma}}_{n,m}$. Moreover, due to Corollary 6.5 we may condition on the event that

$$\|\rho_{\hat{\boldsymbol{\sigma}}_{n,m}} - \bar{\rho}\|_{\mathrm{TV}} + \|\rho_{\hat{\boldsymbol{\sigma}}_{n,m+1}} - \bar{\rho}\|_{\mathrm{TV}} = O(n^{-1/2}\ln n).$$

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

Hence, consider σ such that $\|\rho_\sigma - \bar{\rho}\|_{\mathrm{TV}} = O(n^{-1/2}\ln n)$. By **SYM** and **BAL** the first derivative of the function

$$\phi(\rho) = \sum_{\tau\in\Omega^k} \mathbb{E}[\boldsymbol{\psi}(\tau_1,\ldots,\tau_k)] \prod_{j=1}^{k} \rho(\tau_j)$$

vanishes at $\bar{\rho}$ and thus $\phi(\rho) = \xi + O(\|\rho - \bar{\rho}\|_{\mathrm{TV}}^2)$. Therefore, by Lemma 6.4 and (5.1),

$$\frac{\mathbb{E}[\psi_{G(n,m+1)}(\sigma)]}{\mathbb{E}[\psi_{G(n,m)}(\sigma)]} = \phi(\rho_\sigma) = \xi + O(\ln^2 n/n). \tag{6.7}$$

Summing (6.7) on σ and applying **BAL** a second time, we obtain

$$\frac{\mathbb{E}[Z(G(n,m+1))]}{\mathbb{E}[Z(G(n,m))]} = \xi + O(\ln^2 n/n). \tag{6.8}$$

Plugging (6.7) and (6.8) into (6.2), we obtain $d_{\mathrm{TV}}(\hat{\boldsymbol{\sigma}}_{n,m}, \hat{\boldsymbol{\sigma}}_{n,m+1}) = O(\ln^4 n/n)$, as desired. □

The main difference between soft and hard constraints is that the addition of a single hard constraint can potentially have a dramatic impact on the partition function. In fact, a single hard constraint can diminish $\log Z$ by a linear amount $\Theta(n)$; one of the main technical challenges of this work is to cope with this possibility. However, the following crucial lemma shows that in the planted model such 'high-impact' constraints are unlikely to be present, and that even the collective impact of $n^{3/4}$ constraints is typically sublinear.

Lemma 6.9 (**SYM**, **BAL**). *For any $D > 0$ and $\theta > 0$ there is $n_0 > 0$ such that for all $n > n_0$ for all $m \leqslant Dn/k$ and all γ the following is true. With probability $1 - \exp(-n^{0.8})$ the random CSP $\hat{G}(n,m,\gamma,\pi,\theta)$ has the following property:*

> *If G' is obtained from $\hat{G}(n,m,\gamma,\pi,\theta)$ by deleting any set U of at most $n^{3/4}$ constraints, then $\ln Z(G') - \ln Z(\hat{G}(n,m,\gamma,\pi,\theta)) \leqslant n^{0.9}$.*

Proof. The proof is based on a double-counting argument; throughout we assume that n is sufficiently large. Let $\check{G} = \hat{G}(n,m,\gamma,\pi,\theta)$ for brevity. For a specific set U, let $\mathcal{E}(U)$ be the event that the factor graph G' satisfies $\ln Z(G') - \ln Z(\check{G}) > n^{0.9}$. Also, let $\mathcal{E}$ be the union of all the events $\mathcal{E}(U)$ with $|U| \leqslant n^{3/4}$. Additionally, let $\mathcal{I}$ be the event that $\check{G}$ has at least $n^{0.9}$ isolated variables and that no variable has degree larger than $n^{0.8}$. A standard balls-into-bins calculation shows that

$$\mathbb{P}[\mathcal{I}] \geqslant 1 - \exp(-2n^{0.8}). \tag{6.9}$$

Hence, it suffices to bound

$$\mathbb{P}[\mathcal{E}\cap\mathcal{I}] \leqslant \sum_{U:|U|\leqslant n^{3/4}} \mathbb{P}[\mathcal{E}(U)\cap\mathcal{I}]. \tag{6.10}$$

Let $\tilde{U}$ be the set of all k-ary constraints in U together with all the k-ary constraints that are adjacent to the unique variable appearing in a unary constraint from U. For a graph $\check{G} \in \mathcal{E}(U)\cap\mathcal{I}$ obtain $\tilde{G}$ by rewiring the constraints $a \in \tilde{U}$ such that in $\tilde{G}$ each is adjacent to distinct variables that are isolated in $\check{G}$. There is a sufficient supply of isolated variables because $\check{G} \in \mathcal{I}$ and $|U| \leqslant 2n^{0.8}$ on $\mathcal{I}$; the isolated vertices used and the rewiring protocol are deterministic given $\check{G}$. We claim that almost surely (with respect to choice of $\check{G}$),

$$Z(G') \leqslant \exp(O(|U|))Z(\tilde{G}). \tag{6.11}$$

Indeed, each k-ary constraint of $\tilde{G}$ not present in G' is connected with k variables that do not have any further neighbours. Hence, **SYM** ensures that the addition of these constraints decreases the

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

partition function by no more than a factor of $\xi^{|U|}$. Further, Lemma 6.3 ensures that each of the unary constraints contained in U is satisfiable (because we can think of $\check{G}$ as being obtained by first planting an assignment and then adding constraints that are satisfied under this assignment). Consequently, being connected in $\tilde{G}$ exclusively to variables that are adjacent to unary constraints only, the unary constraints in U have an impact of no more than $\exp(O(|U|))$ on the partition function. Thus, (6.11) follows.

Let $\check{\mathcal{G}} = \mathcal{E}(U) \cap \mathcal{I}$ and let $\tilde{\mathcal{G}}$ be the set of all possible graphs $\tilde{G}$ that can be obtained from some $\check{G} \in \check{\mathcal{G}}$. We define a bipartite graph structure on the (finite) sets $\check{\mathcal{G}}, \tilde{\mathcal{G}}$ by connecting each $\check{G}$ with the corresponding $\tilde{G}$. Thus, each vertex in $\check{\mathcal{G}}$ has degree one, but those in $\tilde{\mathcal{G}}$ may have many neighbours. However, we claim that for every $\tilde{G} \in \tilde{\mathcal{G}}$

$$\sum_{G \in \partial \tilde{G}} \mathbb{P}[G(n, m, \gamma, \pi, \theta) = G] \leqslant \exp(n^{0.81}) \mathbb{P}[G(n, m, \gamma, \pi, \theta) = \tilde{G}]. \tag{6.12}$$

Indeed, the only difference between $\tilde{G}$ and any neighbour $G \in \partial \tilde{G}$ is that $O(n^{0.8})$ constraints have different neighbours. Since in $G(n, m, \gamma, \pi, \theta)$ the neighbours are chosen uniformly, we obtain (6.12) from double-counting.

To complete the proof recall that $Z(\check{G}) \leqslant \exp(n^{0.9}) Z(G')$ for $\check{G} \in \mathcal{E}(U)$. Hence, (6.11) implies

$$Z(\tilde{G}) \geqslant Z(\check{G}) \exp(n^{0.9}/2).$$

Therefore, (6.12) gives

$$\begin{aligned} \mathbb{P}[\check{G} \in \mathcal{E}(U) \cap \mathcal{I}] &\leqslant \frac{\mathbb{E}[Z(G(n, m, \gamma, \pi, \theta)) \mathbf{1}\{G(n, m, \gamma, \pi, \theta) \in \mathcal{E}(U) \cap \mathcal{I}\}]}{\mathbb{E}[Z(G(n, m, \gamma, \pi, \theta))]} \\ &\leqslant \exp(n^{0.81}) \cdot \frac{\sum_{G \in \mathcal{G}} Z(G) \mathbb{P}[G(n, m, \gamma, \pi, \theta) = G]}{\sum_{G \in \mathcal{G}} Z(\tilde{G}) \mathbb{P}[G(n, m, \gamma, \pi, \theta) = G]} \\ &\leqslant \exp(-n^{0.9}/3). \end{aligned} \tag{6.13}$$

Finally, the assertion follows from (6.9), (6.10) and (6.13). □

Equipped with Lemma 6.9 we can complete the proof of Proposition 6.7. The argument is similar to the proof of [27, Lemma 3.32], except that we have to apply Lemma 6.9 to make coupling work.

Proof of Proposition 6.7. The proof is by way of a coupling of

$$\hat{G}(n, m, \gamma, \pi, \theta), \quad \hat{G}(n, m+1, \gamma, \pi, \theta).$$

By Lemma 6.8 we can couple

$$\hat{\boldsymbol{\sigma}}' = \hat{\boldsymbol{\sigma}}_{n, m, \boldsymbol{\gamma}_t, m_t, \boldsymbol{\theta}_\varepsilon}, \quad \hat{\boldsymbol{\sigma}}'' = \hat{\boldsymbol{\sigma}}_{n, m, \boldsymbol{\gamma}_t, m_t+1, \boldsymbol{\theta}_\varepsilon}$$

such that

$$\mathbb{P}[\hat{\boldsymbol{\sigma}}' = \hat{\boldsymbol{\sigma}}''] = 1 - O(\ln^4 n/n), \quad \mathbb{P}[|\hat{\boldsymbol{\sigma}}' \triangle \hat{\boldsymbol{\sigma}}''| > \sqrt{n} \ln n] = O(n^{-2}). \tag{6.14}$$

Further, given $\hat{\boldsymbol{\sigma}}', \hat{\boldsymbol{\sigma}}''$ we couple

$$G' \stackrel{\mathrm{d}}{=} G^*(n, m, \gamma, \pi, \theta, \hat{\boldsymbol{\sigma}}'), \quad G'' \stackrel{\mathrm{d}}{=} G^*(n, m+1, \gamma, \pi, \theta, \hat{\boldsymbol{\sigma}}'')$$

as follows.

Case 1. $\hat{\boldsymbol{\sigma}}' = \hat{\boldsymbol{\sigma}}''$. We couple so that all of their unary constraints as well as the first m k-ary constraints coincide. Additionally, G'' contains a single further random k-ary constraint a drawn according to (6.3) with respect to the planted assignment $\hat{\boldsymbol{\sigma}}'$. Hence,

$$\mathbb{E}\left[\ln \frac{Z(G'')}{Z(G')} \,\middle|\, \hat{\boldsymbol{\sigma}}' = \hat{\boldsymbol{\sigma}}''\right] = \mathbb{E}[\ln \langle \psi_a(\boldsymbol{\sigma}_{G'}) \rangle_{G'} \mid \hat{\boldsymbol{\sigma}}' = \hat{\boldsymbol{\sigma}}''] . \tag{6.15}$$

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

Case 2. $|\hat{\boldsymbol{\sigma}}'\triangle\hat{\boldsymbol{\sigma}}''| \leqslant \sqrt{n}\ln n$. The definition (6.3) of the planted distribution ensures that with probability $1-O(n^{-2})$ the total number X of constraints in either $G^*(n,m,\gamma,\pi,\theta,\hat{\boldsymbol{\sigma}}')$ or $G^*(n,m+1,\gamma,\pi,\theta,\hat{\boldsymbol{\sigma}}'')$ that are adjacent to a variable in $\hat{\boldsymbol{\sigma}}'\triangle\hat{\boldsymbol{\sigma}}''$ is bounded by $n^{2/3}$. Hence, we couple the first m constraints such that G', G'' coincide on those constraints that are not adjacent to any variable in $\hat{\boldsymbol{\sigma}}'\triangle\hat{\boldsymbol{\sigma}}''$, while the constraints that are adjacent to a variable in $\hat{\boldsymbol{\sigma}}'\triangle\hat{\boldsymbol{\sigma}}''$ are chosen independently. Additionally, G'' contains an $(m+1)$th constraint that is chosen independently of the rest. Thus, Lemma 6.9 implies that

$$\mathbb{E}\left[\ln\frac{Z(G'')}{Z(G')}\;\middle|\;|\hat{\boldsymbol{\sigma}}'\triangle\hat{\boldsymbol{\sigma}}''|\leqslant\sqrt{n}\ln n\right]=O(n^{0.9}). \tag{6.16}$$

Case 3. $|\hat{\boldsymbol{\sigma}}'\triangle\hat{\boldsymbol{\sigma}}''| > \sqrt{n}\ln n$. In this case we choose G', G'' independently from their respective distributions. The deterministic bound $|\ln Z(G')|,\ |\ln Z(G'')|\leqslant O(n+m)$ implies

$$\mathbb{E}\left[\ln\frac{Z(G'')}{Z(G')}\;\middle|\;|\hat{\boldsymbol{\sigma}}'\triangle\hat{\boldsymbol{\sigma}}''|>\sqrt{n}\ln n\right]=O(n). \tag{6.17}$$

Combining (6.14)–(6.17), Lemma 6.3 and applying Lemma 6.9 a second time, we conclude that

$$\begin{aligned}\mathbb{E}\left[\ln\frac{Z(\hat{G}(n,m+1,\gamma,\pi,\theta))}{Z(\hat{G}(n,m,\gamma,\pi,\theta))}\right]&=\mathbb{E}[\ln\langle\psi_a(\boldsymbol{\sigma}_{G'})\rangle_{G'}\mid\hat{\boldsymbol{\sigma}}'=\hat{\boldsymbol{\sigma}}'']+o(1)\\&=\mathbb{E}[\ln\langle\psi_a(\boldsymbol{\sigma}_{G'})\rangle_{G'}]+o(1).\end{aligned} \tag{6.18}$$

To compute $\mathbb{E}[\ln\langle\psi_a(\boldsymbol{\sigma}_{G'})\rangle_{G'}]$ we write $\boldsymbol{\sigma},\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2,\ldots$ for independent samples from $\mu_{G'}$. Thus, spelling out the definition of a, we find

$$\mathbb{E}[\ln\langle\psi_a(\boldsymbol{\sigma}_{G'})\rangle_{G'}]=\frac{\mathbb{E}[\boldsymbol{\psi}(\hat{\boldsymbol{\sigma}}'(y_1),\ldots,\hat{\boldsymbol{\sigma}}'(y_k))\ln\langle\boldsymbol{\psi}(\boldsymbol{\sigma}(y_1),\ldots,\boldsymbol{\sigma}(y_k))\rangle_{G'}]}{\mathbb{E}[\boldsymbol{\psi}(\hat{\boldsymbol{\sigma}}'(y_1),\ldots,\hat{\boldsymbol{\sigma}}'(y_k))]}.$$

Since by Corollary 6.5 the empirical distribution $\rho_{\hat{\boldsymbol{\sigma}}'}$ is asymptotically uniform with very high probability, the denominator equals $\xi+o(1)$ with probability $1-O(n^{-2})$. Thus,

$$\mathbb{E}[\ln\langle\psi_a(\boldsymbol{\sigma}_{G'})\rangle_{G'}]=(\xi^{-1}+o(1))\mathbb{E}[\boldsymbol{\psi}(\hat{\boldsymbol{\sigma}}'(y_1),\ldots,\hat{\boldsymbol{\sigma}}'(y_k))\ln\langle\boldsymbol{\psi}(\boldsymbol{\sigma}(y_1),\ldots,\boldsymbol{\sigma}(y_k))\rangle_{G'}]. \tag{6.19}$$

To proceed we are going to use the series expansion of the logarithm. This expansion applies because we may assume that the argument of the logarithm lies in the interval $(0,1]$. Indeed, to obtain the lower bound we simply observe that $\boldsymbol{\psi}(\hat{\boldsymbol{\sigma}}'(y_1),\ldots,\hat{\boldsymbol{\sigma}}'(y_k))>0$ because otherwise the prefactor vanishes, and $\mu_{G'}(\hat{\boldsymbol{\sigma}}')>0$ by Lemma 6.3. Moreover, $\boldsymbol{\psi}\leqslant 1$ by the definition of the constraint functions. Thus, expanding the logarithm we obtain

$$\mathbb{E}[\ln\langle\psi_a(\boldsymbol{\sigma}_{G'})\rangle_{G'}]=-(\xi^{-1}+o(1))\mathbb{E}\left[\sum_{\ell\geqslant1}\frac{\boldsymbol{\psi}(\hat{\boldsymbol{\sigma}}'(y_1),\ldots,\hat{\boldsymbol{\sigma}}'(y_k))}{\ell}\langle1-\boldsymbol{\psi}(\boldsymbol{\sigma}(y_1),\ldots,\boldsymbol{\sigma}(y_k))\rangle_{G'}^{\ell}\right].$$

Because the constraint functions are upper-bounded by 1, the sum is absolutely convergent. Hence, we may swap the sum and the expectation and obtain

$$\mathbb{E}[\ln\langle\psi_a(\boldsymbol{\sigma}_{G'})\rangle_{G'}]=-(\xi^{-1}+o(1))\sum_{\ell\geqslant1}\frac{1}{\ell}\mathbb{E}[\boldsymbol{\psi}(\hat{\boldsymbol{\sigma}}'(y_1),\ldots,\hat{\boldsymbol{\sigma}}'(y_k))\langle1-\boldsymbol{\psi}(\boldsymbol{\sigma}(y_1),\ldots,\boldsymbol{\sigma}(y_k))\rangle_{G'}^{\ell}].$$

Further, applying Lemma 6.3 once more we obtain

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

$$\begin{aligned}
&\mathbb{E}[\ln\langle\psi_a(\boldsymbol{\sigma}_{G'})\rangle_{G'}]\\
&\quad=-(\xi^{-1}+o(1))\\
&\qquad\times\sum_{\ell\geqslant1}\frac{1}{\ell}\mathbb{E}\Big[(1-(1-\boldsymbol{\psi}(\hat{\boldsymbol{\sigma}}'(y_1),\ldots,\hat{\boldsymbol{\sigma}}'(y_k))))\Big\langle\prod_{h=1}^{\ell}1-\boldsymbol{\psi}(\boldsymbol{\sigma}_h(y_1),\ldots,\boldsymbol{\sigma}_h(y_k))\Big\rangle_{G'}\Big]\\
&\quad=-(\xi^{-1}+o(1))\\
&\qquad\times\sum_{\ell\geqslant1}\frac{1}{\ell}\mathbb{E}\Big[\Big\langle\prod_{h=1}^{\ell}1-\boldsymbol{\psi}(\boldsymbol{\sigma}_h(y_1),\ldots,\boldsymbol{\sigma}_h(y_k))\Big\rangle_{G'}\Big]-\frac{1}{\ell}\mathbb{E}\Big[\Big\langle\prod_{h=1}^{\ell+1}1-\boldsymbol{\psi}(\boldsymbol{\sigma}_h(y_1),\ldots,\boldsymbol{\sigma}_h(y_k))\Big\rangle_{G'}\Big]\\
&\quad=-(\xi^{-1}+o(1))\\
&\qquad\times\Big[1-\mathbb{E}[\langle\boldsymbol{\psi}(\boldsymbol{\sigma}(y_1),\ldots,\boldsymbol{\sigma}(y_k))\rangle_{G'}]-\sum_{\ell\geqslant2}\frac{1}{\ell(\ell-1)}\mathbb{E}[\langle1-\boldsymbol{\psi}(\boldsymbol{\sigma}(y_1),\ldots,\boldsymbol{\sigma}(y_k))\rangle_{G'}^{\ell}]\Big].
\end{aligned}\tag{6.20}$$

Due to the series expansion

$$\Lambda(1-x)+x=\sum_{\ell\geqslant2}\frac{x^\ell}{\ell(\ell-1)},$$

the assertion follows by combining (6.18) and (6.20). □

6.4 The lower bound

Thanks to Proposition 6.7, the rest of the proof of Proposition 6.2 is almost identical to the proof of [27, Proposition 3.30], except that we have to pay a bit of attention to some convergence issues. Write $\langle\cdot\rangle_{t,\varepsilon}$ for the expectation with respect to the Gibbs measure of $\hat{G}_{t,\varepsilon}$. Unless specified otherwise, $\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2,\ldots$ denote independent samples from $\mu_{\hat{G}_{t,\varepsilon}}$. Moreover, we write $\boldsymbol{\psi}$ for a sample from P and $x_1,\ldots,x_k\in V_n$ for independently and uniformly chosen variables. Toward the proof of Proposition 6.2 we establish the following formula for the derivative of

$$\phi_\varepsilon(t)=\mathbb{E}[\ln Z(\hat{G}_{t,\varepsilon})]/n+\Gamma_t.$$

Lemma 6.10 (**SYM**, **BAL**). *Let $\boldsymbol{\rho}_1,\ldots,\boldsymbol{\rho}_k$ be chosen from π, mutually independently and independently of everything else. Set*

$$\begin{aligned}
\Xi_{t,\ell}=\mathbb{E}\Big[&\langle1-\boldsymbol{\psi}(\boldsymbol{\sigma}(y_1),\ldots,\boldsymbol{\sigma}(y_k))\rangle_{t,\varepsilon}^{\ell}-k\Big\langle1-\sum_{\tau\in\Omega^{k-1}}\boldsymbol{\psi}(\tau,\boldsymbol{\sigma}(y_1))\prod_{j<k}\boldsymbol{\rho}_j(\tau_j)\Big\rangle_{t,\varepsilon}^{\ell}\\
&+(k-1)\Big(1-\sum_{\tau\in\Omega^k}\boldsymbol{\psi}(\tau)\prod_{j=1}^{k}\boldsymbol{\rho}_j(\tau_j)\Big)^{\ell}\Big].
\end{aligned}$$

Then

$$\frac{\partial}{\partial t}\phi_\varepsilon(t)=o(1)+\frac{d}{k\xi}\sum_{\ell\geqslant2}\frac{\Xi_{t,\ell}}{\ell(\ell-1)}$$

uniformly for all $\varepsilon,t\in(0,1)$.

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

Let

$$\Delta_t = \mathbb{E}[\ln Z(\hat{G}_{t,\varepsilon}(m_t+1,\boldsymbol{\gamma}_t))] - \mathbb{E}[\ln Z(\hat{G}_{t,\varepsilon}(m_t,\boldsymbol{\gamma}_t))],$$
$$\Delta'_t = \mathbb{E}[\ln Z(\hat{G}_{t,\varepsilon}(m_t,\boldsymbol{\gamma}_t+\mathbf{1}_{x_1}))] - \mathbb{E}[\ln Z(\hat{G}_{t,\varepsilon}(m_t,\boldsymbol{\gamma}_t))].$$

Thus, Δ_t is the expected impact of adding one more k-ary constraint to $\hat{G}_{t,\varepsilon}$. Similarly, Δ'_t quantifies the average impact of adding a unary constraint as per **INT2**. The following standard calculation shows how

$$\frac{\partial}{\partial t}\mathbb{E}[\ln Z(\hat{G}_{t,\varepsilon})]$$

can be expressed in terms of Δ_t, Δ'_t.

Claim 6.11 (**SYM**, **BAL**). *We have*

$$\frac{1}{n}\frac{\partial}{\partial t}\mathbb{E}[\ln Z(\hat{G}_{t,\varepsilon})] = \frac{d}{k}\Delta_t - d\Delta'_t.$$

Proof. Let $P_\lambda(j) = \lambda^j \exp(-\lambda)/j!$. By the construction, the parameter t only affects the distribution of random variables $m_t, \boldsymbol{\gamma}_t$. Indeed,

$$\mathbb{E}[\ln Z(\hat{G}_{t,\varepsilon})] = \sum_{m,\gamma}\mathbb{E}[\ln Z(\hat{G}_{t,\varepsilon})|m_t=m,\boldsymbol{\gamma}_t=\gamma]P_{tdn/k}(m)\prod_{x\in V}P_{(1-t)d}(\gamma_x). \tag{6.21}$$

Since the derivatives of the Poisson densities come out as

$$\frac{\partial}{\partial t}P_{tdn/k}(m) = \frac{dn}{k}[\mathbf{1}\{m\geqslant 1\}P_{tdn/k}(m-1) - P_{tdn/k}(m)],$$
$$\frac{\partial}{\partial t}P_{(1-t)d}(\gamma_v) = -d[\mathbf{1}\{\gamma_v\geqslant 1\}P_{(1-t)d}(\gamma_v-1) - P_{(1-t)d}(\gamma_v)],$$

the assertion follows from (6.21) and the product rule. □

We proceed to calculate Δ_t, Δ'_t.

Claim 6.12 (**SYM**, **BAL**). *We have*

$$\Delta_t = o(1) - \frac{1-\xi}{\xi} + \sum_{\ell\geqslant 2}\frac{1}{\ell(\ell-1)\xi}\mathbb{E}[\langle 1-\boldsymbol{\psi}(\boldsymbol{\sigma}(y_1),\ldots,\boldsymbol{\sigma}(y_k))\rangle^{\ell}_{t,\varepsilon}].$$

Proof. Recalling the expansion

$$\Lambda(1-x)+x = \sum_{\ell\geqslant 2}\frac{x^\ell}{\ell(\ell-1)},$$

we obtain from Proposition 6.7 that

$$\begin{aligned}\Delta_t &= o(1) + \xi^{-1}\mathbb{E}[\Lambda(\langle\boldsymbol{\psi}(\boldsymbol{\sigma}(y_1),\ldots,\boldsymbol{\sigma}(y_k))\rangle_{t,\varepsilon})]\\ &= o(1) - \xi^{-1}(1-\mathbb{E}[\langle\boldsymbol{\psi}(\boldsymbol{\sigma}(y_1),\ldots,\boldsymbol{\sigma}(y_k))\rangle_{t,\varepsilon}])\\ &\quad + \sum_{\ell\geqslant 2}\frac{1}{\ell(\ell-1)\xi}\mathbb{E}[\langle 1-\boldsymbol{\psi}(\boldsymbol{\sigma}(y_1),\ldots,\boldsymbol{\sigma}(y_k))\rangle^{\ell}_{t,\varepsilon}].\end{aligned}$$

Further, Lemma 6.3, Corollary 6.5 and **SYM** yield

$$\mathbb{E}[\langle\boldsymbol{\psi}(\boldsymbol{\sigma}(y_1),\ldots,\boldsymbol{\sigma}(y_k))\rangle_{t,\varepsilon}] = \xi + o(1).$$ □

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

Claim 6.13 (**SYM**, **BAL**). *With $\boldsymbol{\rho}_1, \boldsymbol{\rho}_2, \ldots$ drawn from π mutually independently and independently of everything else,*

$$\Delta_t' = -\frac{1-\xi}{\xi} + \sum_{\ell \geqslant 2} \frac{1}{\ell(\ell-1)\xi} \mathbb{E}\left[\left\langle 1 - \sum_{\tau_1,\ldots,\tau_{k-1} \in \Omega} \boldsymbol{\psi}(\tau_1, \ldots, \tau_{k-1}, \boldsymbol{\sigma}(y_1)) \prod_{j=1}^{k-1} \boldsymbol{\rho}_j(\tau_j) \right\rangle_{t,\varepsilon}^{\ell}\right].$$

Proof. Lemma 6.3 shows that $\hat{\boldsymbol{\sigma}}_{n,m,\boldsymbol{\gamma}_t,m_t,\boldsymbol{\theta}_\varepsilon}, \hat{\boldsymbol{\sigma}}_{n,m,\boldsymbol{\gamma}_t+\mathbf{1}_x,m_t,\boldsymbol{\theta}_\varepsilon}$ are identically distributed and hence we can couple them identically. Let us write $\hat{\boldsymbol{\sigma}}$ for brevity. Further, we couple

$$G' \stackrel{\mathrm{d}}{=} \hat{G}_{t,\varepsilon}(m_t, \boldsymbol{\gamma}_t), \quad G'' \stackrel{\mathrm{d}}{=} \hat{G}_{t,\varepsilon}(m_t, \boldsymbol{\gamma}_t + \mathbf{1}_{y_1})$$

in the natural way: first choose G' from the distribution $\hat{G}_{t,\varepsilon}(m_t, \boldsymbol{\gamma}_t)$, then obtain G'' simply by adding one more unary constraint b with $\partial b = y_1$ according to step **G2** of our construction. Then

$$\mathbb{E}[\ln Z(\hat{G}_{t,\varepsilon}(m_t, \boldsymbol{\gamma}_t + \mathbf{1}_x))] - \mathbb{E}[\ln Z(\hat{G}_{t,\varepsilon}(m_t, \boldsymbol{\gamma}_t))] = \mathbb{E}\left[\ln \frac{Z(G'')}{Z(G')}\right] = \mathbb{E}[\ln \langle \psi_b(\boldsymbol{\sigma}(y_1))\rangle_{G'}]. \tag{6.22}$$

Since $\psi_b(\hat{\boldsymbol{\sigma}}(y_1)) > 0$ by construction, we see that $0 < \langle \psi_b(\boldsymbol{\sigma}(y_1))\rangle_{G'} \leqslant 1$ and therefore by Fubini's theorem

$$\mathbb{E}[\ln \langle \psi_b(\boldsymbol{\sigma}(y_1))\rangle_{G'}] = -\mathbb{E}\left[\sum_{\ell \geqslant 1} \frac{1}{\ell} \langle 1 - \psi_b(\boldsymbol{\sigma}(y_1))\rangle_{G'}^{\ell}\right] = -\sum_{\ell \geqslant 1} \frac{1}{\ell} \mathbb{E}[\langle 1 - \psi_b(\boldsymbol{\sigma}(y_1))\rangle_{G'}^{\ell}].$$

Hence, due to **INT2**, the upper bound $\psi_b \leqslant 1$, Lemma 6.3 and assumption **SYM**,

$$\begin{aligned}
&\mathbb{E}[\ln \langle \psi_b(\boldsymbol{\sigma}(y_1))\rangle_{G'}] \\
&= -\sum_{\ell \geqslant 1} \frac{1}{\xi\ell} \mathbb{E}\left[\left(\sum_{\tau \in \Omega^{k-1}} \boldsymbol{\psi}(\tau, \hat{\boldsymbol{\sigma}}(y_1)) \prod_{j<k} \boldsymbol{\rho}_j(\tau_j)\right) \left\langle \prod_{h=1}^{\ell} \left(1 - \sum_{\tau \in \Omega^{k-1}} \boldsymbol{\psi}(\boldsymbol{\sigma}_h(y_1)) \prod_{j<k} \boldsymbol{\rho}_j(\tau_j)\right)\right\rangle_{G'}\right] \\
&= -\sum_{\ell \geqslant 1} \frac{1}{\xi\ell} \mathbb{E}\left[\left\langle 1 - \sum_{\tau \in \Omega^{k-1}} \boldsymbol{\psi}(\boldsymbol{\sigma}(y_1)) \prod_{j=1}^{k-1} \boldsymbol{\rho}_j(\tau_j)\right\rangle_{G'}^{\ell} - \left\langle 1 - \sum_{\tau \in \Omega^{k-1}} \boldsymbol{\psi}(\boldsymbol{\sigma}(y_1)) \prod_{j=1}^{k-1} \boldsymbol{\rho}_j(\tau_j)\right\rangle_{G'}^{\ell+1}\right] \\
&= \xi^{-1}[\mathbb{E}[\boldsymbol{\psi}(\boldsymbol{\sigma}(y_1))] - 1 + \sum_{\ell \geqslant 2} \frac{1}{\ell(\ell-1)} \mathbb{E}[\langle 1 - \boldsymbol{\psi}(\boldsymbol{\sigma}(y_1))\rangle_{G'}^{\ell}]] \\
&= -\frac{1-\xi}{\xi} + \sum_{\ell \geqslant 2} \frac{1}{\xi\ell(\ell-1)} \mathbb{E}[\langle 1 - \boldsymbol{\psi}(\boldsymbol{\sigma}(y_1))\rangle_{G'}^{\ell}],
\end{aligned}$$

as claimed. □

Claim 6.14 (**SYM**, **BAL**). *With $\boldsymbol{\rho}_1, \boldsymbol{\rho}_2, \ldots$ drawn from π mutually independently and independently of everything else,*

$$\Delta_t'' = \frac{k}{d(k-1)} \frac{\partial}{\partial t} \Gamma_t = -\frac{1-\xi}{\xi} + \sum_{\ell \geqslant 2} \frac{1}{\ell(\ell-1)\xi} \mathbb{E}\left[\left(1 - \sum_{\tau \in \Omega^k} \boldsymbol{\psi}(\tau) \prod_{j=1}^{k} \boldsymbol{\rho}_j(\tau_j)\right)^{\ell}\right].$$

Proof. Since

$$\mathbb{E}\left[\sum_{\tau \in \Omega^k} \boldsymbol{\psi}(\tau) \prod_{j=1}^{k} \boldsymbol{\rho}_j(\tau_j)\right] = \xi,$$

this follows along the lines of the proof of Claim 6.12. □

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

Proof of Lemma 6.10. The assertion is immediate from Claims 6.11–6.14. □

Proof of Proposition 6.2. Let $\boldsymbol{\pi}_{t,\varepsilon}$ be the empirical distribution of the marginals of the random probability measure $\mu_{\hat{G}_{t,\varepsilon}}$. Write $\boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \ldots$ for independent samples drawn from $\boldsymbol{\pi}_{t,\varepsilon}$ and define

$$\Xi'_{t,\ell} = \mathbb{E}\Bigg[\bigg(1 - \sum_{\sigma\in\Omega^k} \boldsymbol{\psi}(\sigma)\prod_{j=1}^{k}\boldsymbol{\nu}_j(\sigma_j)\bigg)^{\ell} - k\bigg(1 - \sum_{\tau\in\Omega^k}\boldsymbol{\psi}(\tau)\boldsymbol{\nu}_1(\tau_k)\prod_{j<k}\boldsymbol{\rho}_j(\tau_j)\bigg)^{\ell} + (k-1)\bigg(1-\sum_{\tau\in\Omega^k}\boldsymbol{\psi}(\tau)\prod_{j=1}^{k}\boldsymbol{\rho}_j(\tau_j)\bigg)^{\ell}\Bigg].$$

Lemma 4.2 implies that for any $\eta > 0$, $\ell \geqslant 1$ there is $\varepsilon > 0$ such that, in the case that $\mu_{\hat{G}_{t,\varepsilon}}$ is ε-symmetric for all $\psi \in \Psi$ and all $t \in [0,1]$, we have

$$\frac{1}{n^k}\sum_{y_1,\ldots,y_k\in V}\Big|\langle 1-\psi(\boldsymbol{\sigma}(y_1),\ldots,\boldsymbol{\sigma}(y_k))\rangle^{\ell}_{\hat{G}_{t,\varepsilon}} - \bigg(1-\sum_{\sigma\in\Omega^k}\psi(\sigma)\prod_{j=1}^{k}\langle\mathbf{1}\{\boldsymbol{\sigma}(y_j)=\sigma_j\}\rangle_{\hat{G}_{t,\varepsilon}}\bigg)^{\ell}\Big| < \eta. \tag{6.23}$$

Further, $\mu_{\hat{G}_{t,\varepsilon}}$ is ε-symmetric with probability at least $1-\varepsilon$ by Lemma 4.6. Consequently, for any ℓ and any $\eta > 0$ we can pick $\varepsilon > 0$ small enough so that $|\Xi_{t,\ell} - \Xi'_{t,\ell}| < \eta$. Finally, since $|\Xi_{t,\ell}| \leqslant 2k$ for all t, ℓ and because the series $\sum_{\ell\geqslant 2} 1/(\ell(\ell-1))$ converges, the assertion follows from **POS**, (6.23) and Lemma 6.10. □

Proof of Proposition 6.1. By construction, $\hat{G}_{1,\varepsilon}$ is obtained from $\hat{G}$ by adding further constraints. Therefore, invoking Proposition 6.2 and the fundamental theorem of calculus, we find that for any $\delta > 0$ there is $\varepsilon > 0$ such that

$$\mathbb{E}[\ln Z(\hat{G})] \geqslant \mathbb{E}[\ln Z(\hat{G}_{1,\varepsilon})] \geqslant \mathbb{E}[\ln Z(\hat{G}_{0,\varepsilon})] - \Gamma_1 n - \delta n + o(n). \tag{6.24}$$

Furthermore, since $\hat{G}_{0,\varepsilon}$ consists of unary constraints only and since the number $\boldsymbol{\theta}_{\varepsilon}$ of pinned variables is bounded, we see that

$$\mathbb{E}[\ln Z(\hat{G}_{0,\varepsilon})] \geqslant \mathbb{E}[\ln Z(\hat{G}_{0,1})] - O(1).$$

Hence, taking $n \to \infty$ and then $\varepsilon \to 0$, we obtain from (6.24) that

$$\liminf_{n\to\infty}\frac{1}{n}\mathbb{E}[\ln Z(\hat{G})] \geqslant \liminf_{n\to\infty}\frac{1}{n}\mathbb{E}[\ln Z(\hat{G}_{0,1})] - \Gamma_1. \tag{6.25}$$

Thus, we are left to compute $\mathbb{E}[\ln Z(\hat{G}_{0,1})]$. We claim that with independent $\boldsymbol{\gamma} = \mathrm{Po}(d)$, $\boldsymbol{\psi}_i$ from P and $(\boldsymbol{\rho}_{h,i})_{h,i\geqslant 1}$ chosen from π,

$$\frac{1}{n}\mathbb{E}[\ln Z(\hat{G}_{0,1})] = \frac{1}{q}\mathbb{E}\Bigg[\xi^{-\boldsymbol{\gamma}}\Lambda\bigg(\sum_{\sigma\in\Omega}\prod_{h=1}^{\boldsymbol{\gamma}}\sum_{\tau\in\Omega^k}\mathbf{1}\{\tau_k=\sigma\}\boldsymbol{\psi}_h(\tau)\prod_{j=1}^{k-1}\boldsymbol{\rho}_{h,j}(\tau_j)\bigg)\Bigg]. \tag{6.26}$$

Indeed, since $\hat{G}_{0,1}$ has unary constraints only, $\mathbb{E}[\ln Z(\hat{G}_{0,1})]$ is equal to n times the contribution of just the component of $\hat{G}_{0,1}$ that contains the constraint x_1. Formally, we have

$$\mathbb{E}[\ln Z(\hat{G}_{0,1})] = \frac{n}{q}\mathbb{E}[\xi^{-\boldsymbol{\gamma}_{x_1}}\Lambda(z)], \quad \text{where } z = \sum_{\sigma\in\Omega}\prod_{j=1}^{\boldsymbol{\gamma}_{x_1}}\psi_{b_{1,j}}(\sigma), \tag{6.27}$$

because the constraints are chosen with a probability that is proportional to the partition function. Finally, the assertion follows from **INT2** and (6.24)–(6.27). □

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

6.5 The upper bound

To bound $\mathbb{E}[\ln Z(\hat{G}(n,m,P))]$ from above we use the formula for $\mathbb{E}[\ln Z(\hat{G}(n,m,P_\beta))]$ from [27] and take the limit $\beta\to\infty$. To this end we need to show that $\mathbb{E}[\ln Z(\hat{G}(n,m,P_\beta))]$ is an asymptotic upper bound on $\mathbb{E}[\ln Z(\hat{G}(n,m,P))]$ for large β.

Proposition 6.15 (**SYM**, **BAL**). *For any $d>0$ and any $\varepsilon>0$ there exists $\beta_0>0$ and $n_0>0$ such that, for all $m\in\mathcal{M}(d)$, $\beta>\beta_0$ and $n>n_0$, we have*

$$\mathbb{E}[\ln Z(\hat{G}(n,m,P))]\leqslant\mathbb{E}[\ln Z(\hat{G}(n,m,P_\beta))]+\varepsilon n.$$

To prove Proposition 6.15 we need the following basic fact about the random assignments $\hat{\boldsymbol{\sigma}}_{n,m}$, $\hat{\boldsymbol{\sigma}}_{n,m,P_\beta}$.

Lemma 6.16 (**BAL**). *For any $d>0$ and any $\varepsilon>0$ there exists $\beta_0>0$ and $n_0>0$ such that, for all $m\in\mathcal{M}(d)$, $\beta>\beta_0$ and $n>n_0$ for any nearly balanced σ, we have*

$$\mathbb{E}[\ln Z(G^*(n,m,P_\beta,\sigma))]\leqslant\mathbb{E}[\ln Z(G^*(n,m,P_\beta,\hat{\boldsymbol{\sigma}}_{n,m,P_\beta}))]+\varepsilon n.$$

Proof. Lemma 6.5 shows that $\hat{\boldsymbol{\sigma}}_{n,m,P}$ is nearly balanced with probability $1-O(n^{-2})$ and due to Lemma 5.2 the same holds for $\hat{\boldsymbol{\sigma}}_{n,m,P_\beta}$. Further, since $\psi_\beta(\tau)\leqslant 1$ for all $\psi\in\Psi$, we have the deterministic upper bound

$$\ln Z(G^*(n,m,\beta,\hat{\boldsymbol{\sigma}}_{n,m,\beta}))\leqslant n\ln q.$$

Therefore, it suffices to prove that

$$\mathbb{E}[\ln Z(G^*(n,m,P_\beta,\sigma))]\leqslant\mathbb{E}[\ln Z(G^*(n,m,P_\beta,\hat{\boldsymbol{\sigma}}_{n,m,P_\beta}))\mid\hat{\boldsymbol{\sigma}}_{n,m,P_\beta}\text{ is nearly balanced}]+\varepsilon n/2. \tag{6.28}$$

Hence, suppose that $\hat{\boldsymbol{\sigma}}_{n,m,P_\beta}$ is nearly balanced. Since σ is nearly balanced as well, there is a permutation π of $[n]$ such that the symmetric difference satisfies $|(\sigma\circ\pi)\triangle\hat{\boldsymbol{\sigma}}_{n,m,P_\beta}|\leqslant 2qn^{3/5}$. Indeed, because the value of the partition function is invariant under permutations of the variables, we may assume without loss that $\pi=\mathrm{id}$.

Letting $U=\sigma\triangle\hat{\boldsymbol{\sigma}}_{n,m,P_\beta}$, we couple $G^*(n,m,P_\beta,\sigma)$ and $G^*(n,m,P_\beta,\hat{\boldsymbol{\sigma}}_{n,m,P_\beta})$ as follows. Keeping in mind that the constraints are chosen independently according to (5.3), we first reveal for each $i=1,\ldots,m$ whether the corresponding constraint is adjacent to a variable in U in either $G^*(n,m,P_\beta,\sigma)$ or $G^*(n,m,P_\beta,\hat{\boldsymbol{\sigma}}_{n,m,P_\beta})$. If not, then the definition of the models ensures that the distribution of the constraint is identical in the two models and couple such that the ith constraints in the two factor graphs are identical. If, on the other hand, the ith constraint is adjacent to U in either instance, then we insert independently chosen constraints.

Let X be the number of constraints on which the two CSP instances differ under this coupling. Since the addition or removal of a single constraint can alter the partition function by at most a factor of $\exp(\pm\beta)$, we obtain

$$\mathbb{E}[\ln Z(G^*(n,m,P_\beta,\sigma))-\ln Z(G^*(n,m,P_\beta,\hat{\boldsymbol{\sigma}}_{n,m,P_\beta}))\mid X,\hat{\boldsymbol{\sigma}}_{n,m,P_\beta}]\leqslant 2\beta X. \tag{6.29}$$

Hence, we are left to bound X. Due to the independence of the constraints X is a binomial random variable. Moreover, since σ is nearly balanced and $|U|\leqslant 2qn^{3/5}$ assumption **SYM** yields

$$\sum_{h_1,\ldots,h_k\in[n]}\mathbb{E}[\boldsymbol{\psi}_\beta(\sigma(x_{h_1},\ldots,x_{h_k}))]=(\xi_\beta+o(1))n^k,$$

$$n^{-k}\sum_{h_1,\ldots,h_k\in[n]}\mathbb{E}[\boldsymbol{\psi}_\beta(\hat{\boldsymbol{\sigma}}_{n,m,P_\beta}(x_{h_1},\ldots,x_{h_k}))]=(\xi_\beta+o(1))n^k.$$

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

Thus, the bound $|U| \leqslant 2qn^{3/5}$ implies together with the construction (5.3) of the planted model that

$$\mathbb{E}[X \mid \hat{\boldsymbol{\sigma}}_{n,m,P_\beta}] \leqslant \frac{k|U|m}{(\xi_\beta + o(1))n} = O(m/n^{2/5}).$$

Therefore, the Chernoff bound yields $\mathbb{P}[X > n^{0.9} \mid \hat{\boldsymbol{\sigma}}_{n,m,P_\beta}] \leqslant O(n^{-2})$. Thus, (6.28) follows from (6.29) and the deterministic upper bound $\ln Z(G^*(n, m, P_\beta, \sigma)) \leqslant n \ln q$. □

Lemma 6.17 (**SYM**,**BAL**). *For any $d > 0$ and any $\varepsilon > 0$ there exists $\beta_0 > 0$ and $n_0 > 0$ such that for all $m \in \mathcal{M}(d)$, $\beta > \beta_0$ and $n > n_0$ for any nearly balanced σ we have*

$$\mathbb{E}[\ln Z(G^*(n, m, P, \sigma))] \leqslant \mathbb{E}[\ln Z(G^*(n, m, P_\beta, \sigma))] + \varepsilon n.$$

Proof. We use a coupling argument once more. We begin by calculating the total variation distance of the distributions from (5.3) according to which the constraints of $G^*(n, m, P, \sigma)$ and $G^*(n, m, P_\beta, \sigma)$ are drawn. First, because σ is nearly balanced, **SYM** shows that

$$\sum_{j_1,\ldots,j_k \in [n]} \mathbb{E}[\boldsymbol{\psi}(\sigma(x_{j_1}), \ldots, \sigma(x_{j_k}))] \sim \xi n^k, \qquad \sum_{j_1,\ldots,j_k \in [n]} \mathbb{E}[\boldsymbol{\psi}_\beta(\sigma(x_{j_1}), \ldots, \sigma(x_{j_k}))] \sim \xi_\beta n^k.$$

Hence, plugging in the definition (5.7) of the softened constraints, we obtain for any $\psi \in \Psi$ and any $i_1, \ldots, i_k \in [n]$

$$\begin{aligned}
&\left| \frac{\psi(\sigma(x_{i_1}), \ldots, \sigma(x_{i_k}))P(\psi)}{\sum_{j_1,\ldots,j_k \in [n]} \mathbb{E}[\boldsymbol{\psi}(\sigma(x_{j_1}), \ldots, \sigma(x_{j_k}))]} - \frac{\psi_\beta(\sigma(x_{i_1}), \ldots, \sigma(x_{i_k}))P_\beta(\psi_\beta)}{\sum_{j_1,\ldots,j_k \in [n]} \mathbb{E}[\boldsymbol{\psi}_\beta(\sigma(x_{j_1}), \ldots, \sigma(x_{j_k}))]} \right| \\
&\quad = o(n^{-k}) + \left| \frac{\psi(\sigma(x_{i_1}), \ldots, \sigma(x_{i_k}))P(\psi)}{\xi n^k} - \frac{\psi_\beta(\sigma(x_{i_1}), \ldots, \sigma(x_{i_k}))P_\beta(\psi_\beta)}{\xi_\beta n^k} \right| \\
&\quad \leqslant o(n^k) + \frac{P(\psi)}{n^k} \cdot \frac{1}{\xi(1 + (\mathrm{e}^\beta - 1)\xi)}.
\end{aligned}$$

Summing on $\psi, i_1, \ldots, i_k$, we conclude that the total variation distance of the distributions defined by (5.3) for P and P_β, respectively, is bounded by $O(\exp(-\beta))$ for large β. Hence, we can couple these distributions such that they coincide with probability $1 - O(\exp(-\beta))$. We then extend this coupling of the distribution of individual constraints to a coupling of $G^*(n, m, P, \sigma)$ and $G^*(n, m, P_\beta, \sigma)$ by drawing m times independently.

Letting X be the number of constraints in which $G^*(n, m, P_\beta, \sigma)$, $G^*(n, m, P, \sigma)$ differ, we thus obtain the estimate $\mathbb{E}[X] \leqslant O(\exp(-\beta))m$ for large β. Further, because the constraints are chosen independently, X is a binomial random variable. Thus, for large enough β the Chernoff bound shows that

$$\mathbb{P}[X > n/\beta^2] = O(n^{-2}). \tag{6.30}$$

Additionally, since $\psi_\beta(\sigma) \in [\exp(-\beta), 1]$ for all $\psi \in \Psi, \sigma \in \Omega^k$, we obtain the estimate

$$\mathbb{E}[\ln Z(G^*(n, m, P, \sigma)) - \ln Z(G^*(n, m, P_\beta, \sigma)) \mid X] \leqslant X\beta. \tag{6.31}$$

Finally, the assertion follows from (6.30), (6.31) and the deterministic bound

$$\ln Z(G^*(n, m, P_\beta, \sigma)) \leqslant n \ln q,$$

provided that $\beta = \beta(\varepsilon)$ is sufficiently large. □

Finally, Proposition 6.15 is immediate from Lemmas 6.16 and 6.17.

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

Proof of Theorem 5.3. To show the first part of the theorem assume that conditions **SYM** and **BAL** hold. Proposition 6.15 and [27, Proposition 3.6] readily imply that there exists β_0 such that for all $d > 0$ and $\beta > \beta_0$

$$\limsup_{n\to\infty} \frac{1}{n}\mathbb{E}[\ln Z(\hat{G})] \leqslant \sup_{\pi\in\mathcal{P}_*^2(\Omega)} \mathcal{B}(d, P_\beta, \pi).$$

Now, as Λ is bounded and continuous on $[0, 1]$ the convergence of the Bethe functional follows from the dominated convergence theorem.

Moving on to the second part, assume that additionally condition **POS** holds. In order to make use of Proposition 6.1, we need to show that every $\pi \in \mathcal{P}_*^2(\Omega)$ can be approximated arbitrarily well by distributions in $\mathcal{P}_*^2(\Omega)$ that have finite support. To this end, let S_q denote the standard simplex in $\mathbb{R}^\Omega$, let $\pi \in \mathcal{P}_*^2(\Omega)$ be a probability distribution that does *not* have finite support and let $B : \mathbb{N}_0 \times ([0, 1]^k)^\infty \times (S_q)^\infty \to \mathbb{R}$,

$$(\gamma, (\psi_i)_{i\geqslant 1}, (\rho_i)_{i\geqslant 1}) \mapsto q^{-1}\xi^{-\gamma}\Lambda\Bigg(\sum_{\sigma\in\Omega}\prod_{i=1}^{\gamma}\sum_{\tau\in\Omega^k}\mathbf{1}\{\tau_k = \sigma\}\psi_i(\tau)\prod_{j=1}^{k-1}\rho_{ki+j}(\tau_j)\Bigg) - \frac{d(k-1)}{k\xi}\Lambda\Bigg(\sum_{\tau\in\Omega^k}\psi_1(\tau)\prod_{j=1}^{k}\rho_j(\tau_j)\Bigg),$$

be as in the definition of $\mathcal{B}(d, P, \pi)$. We wish to approximate $\mathcal{B}(d, P, \pi)$ by $\mathcal{B}(d, P, \pi_N)$, where $\pi_N \in \mathcal{P}_*^2(\Omega)$ has finite support and $|\mathrm{supp}(\pi_N)| = N$. To this end, we proceed along the following lines.

(1) For every $N \in \mathbb{N}$, we find a discrete probability measure π_N on S_q, whose support consists of exactly N elements such that $\int_{\mathcal{P}(\Omega)} \mu(\omega)\mathrm{d}\pi(\mu) = 1/q$ for all $\omega \in \Omega$ and $(\pi_N)_{N\geqslant 1}$ converges weakly to π as $N \to \infty$.
(2) This implies that $B(\boldsymbol{\gamma}, (\boldsymbol{\psi}_i)_{i\geqslant 1}, (\boldsymbol{\rho}_i^{\pi_N})_{i\geqslant 1})$ converges weakly to $B(\boldsymbol{\gamma}, (\boldsymbol{\psi}_i)_{i\geqslant 1}, (\boldsymbol{\rho}_i^{\pi})_{i\geqslant 1})$. Here, all occurring random variables are independent.
(3) We then apply a variant of the dominated convergence theorem to show convergence of $\mathcal{B}(d, P, \pi_N)$ to $\mathcal{B}(d, P, \pi)$.

Step (1) is a quantization problem: fix $N \in \mathbb{N}$ and let $\mathcal{F}_N$ be the set of all Borel measurable maps $f : \mathbb{R}^\Omega \to \mathbb{R}^\Omega$ with $|f(\mathbb{R}^\Omega)| \leqslant N$. The standard theory on quantization for probability distributions, [51, Theorem 4.1 and Theorem 4.12], guarantees the existence of a function $f_N^* : \mathbb{R}^\Omega \to \mathbb{R}^\Omega$ with $|f_N^*(\mathbb{R}^\Omega)| = N$ and

$$\mathbb{E}[\|\boldsymbol{\rho}_1^\pi - f_N^*(\boldsymbol{\rho}_1^\pi)\|^2] = \inf_{f\in\mathcal{F}_N}\mathbb{E}[\|\boldsymbol{\rho}_1^\pi - f(\boldsymbol{\rho}_1^\pi)\|^2].$$

Here, $\|\cdot\|$ denotes the 2-norm on $\mathbb{R}^\Omega$. Moreover, the use of this norm implies [51, Remark 4.6] that for any such function f_N^*, $\mathbb{E}[f_N^*(\boldsymbol{\rho}_1^\pi)] = \mathbb{E}[\boldsymbol{\rho}_1^\pi]$. In order to see why $\mathbb{E}[\|\boldsymbol{\rho}_1^\pi - f_N^*(\boldsymbol{\rho}_1^\pi)\|^2] = o(1)$, we evoke the following almost sure approximation of $\boldsymbol{\rho}_1^\pi$ which does not fix the mean value, but provides an upper bound for $\mathbb{E}[\|\boldsymbol{\rho}_1^\pi - f_N^*(\boldsymbol{\rho}_1^\pi)\|^2]$. For any $L \in \mathbb{N}$, choose a cover of S_q by open balls of radius $1/L$. As S_q is compact, this cover has a finite sub-cover. By taking intersections of the balls in a finite sub-cover, we may assume that S_q is covered by a finite number of pairwise disjoint sets $B_1, \ldots, B_{j(L)}$, which have diameter at most $2/L$. In each such set B_i, we distinguish a point c_i. Setting $g_L^*(\boldsymbol{\rho}_1^\pi) = \sum_{i=1}^{j(L)} c_i\mathbf{1}\{\boldsymbol{\rho}_1^\pi \in B_i\}$, we have that almost surely, $\|\boldsymbol{\rho}_1^\pi - g_L^*(\boldsymbol{\rho}_1^\pi)\| \leqslant 2/L$ and the distribution of $g_L^*(\boldsymbol{\rho}^\pi)$ has finite support. We may thus find a sequence $(g_L^*)_L$ of functions which take only finitely many values each such that $g_L^*(\boldsymbol{\rho}_1^\pi)$ converges to $\boldsymbol{\rho}_1^\pi$ almost surely. Because both

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

$\boldsymbol{\rho}_1^{\pi}$ and $g_L^*(\boldsymbol{\rho}_1^{\pi})$ are bounded, $\mathbb{E}[\|\boldsymbol{\rho}_1^{\pi} - g_L^*(\boldsymbol{\rho}_1^{\pi})\|^2] = o(1)$ and thus also $\mathbb{E}[\|\boldsymbol{\rho}_1^{\pi} - f_N^*(\boldsymbol{\rho}_1^{\pi})\|^2] = o(1)$. This in turn implies that, if we denote the distribution of $f_N^*(\boldsymbol{\rho}_1^{\pi})$ by π_N, $(\pi_N)_{N\in\mathbb{N}}$ converges weakly to π.

We now turn to (2). Step (1) implies that $\bigotimes_{i=1}^{\infty}\pi_N$ converges weakly to $\bigotimes_{i=1}^{\infty}\pi$ as $N\to\infty$, as $\bigotimes_{i=1}^{\infty}\pi$ is determined by its finite-dimensional distributions. Due to independence, it is true that also

$$(\boldsymbol{\gamma}, (\boldsymbol{\psi}_i)_{i\geqslant 1}, (\boldsymbol{\rho}_i^{\pi_N})_{i\geqslant 1})_{N\geqslant 1} \overset{N\to\infty}{\longrightarrow} (\boldsymbol{\gamma}, (\boldsymbol{\psi}_i)_{i\geqslant 1}, (\boldsymbol{\rho}_i^{\pi})_{i\geqslant 1})$$

in distribution. Finally, B is a continuous function, and thus the continuous mapping theorem implies step (2).

Finally, $B(\boldsymbol{\gamma}, (\boldsymbol{\psi}_i)_{i\geqslant 1}, (\boldsymbol{\rho}_i^{\pi_N})_{i\geqslant 1})$ is integrable for any $N\in\mathbb{N}$ as well as dominated by the integrable random variable $q^{-1}\xi^{-\boldsymbol{\gamma}} + d(k-1)k^{-1}\xi^{-1}$. Hence, the dominated convergence theorem (say, in the version [54, Theorem A39]) yields (3).

Finally, Proposition 6.1 yields the second part of the theorem. □

7. Small subgraph conditioning

Having established Theorem 5.3 in the previous section, we move on to prove the remaining propositions required for the small subgraph conditioning argument outlined in Section 5. Subsequently we derive Theorems 2.5, 2.8 and 2.11 as well as Corollary 2.12. Most of the proofs in this section are based either on standard arguments (*e.g.* the Laplace method or the method of moments for convergence in distribution) or the arguments developed in [25, 27]. We continue to let $x_1, \ldots, x_k \in V_n$ denote variables drawn uniformly and independently.

7.1 Proof of Proposition 5.7

Proposition 6.7 provides a formula for the expected change of the logarithm of the partition function upon addition of a further constraint. We can use this formula to estimate the derivative of $\mathbb{E}[\ln Z(\hat{G})]$ with respect to d because

$$\begin{aligned}\frac{\partial}{\partial d}\mathbb{E}[\ln Z(\hat{G})] &= \sum_{m\geqslant 0}\mathbb{E}[\ln Z(\hat{G}(n,m))]\frac{\partial}{\partial d}\mathbb{P}[\mathrm{Po}(dn/k)=m]\\ &= \frac{1}{k}(\mathbb{E}[\ln Z(\hat{G}(n,m+1))] - \mathbb{E}[\ln Z(\hat{G}(n,m))]).\end{aligned} \tag{7.1}$$

The corresponding formula in the case of soft constraints was obtained in [25], and thanks to Proposition 6.7 the same argument extends to hard constraints with a little bit of care.

Lemma 7.1 (**SYM**, **BAL**, **MIN**). *Fix any $D > 0$.*

(1) *Uniformly for all $0 < d < D$ we have*

$$\frac{1}{n}\frac{\partial}{\partial d}\mathbb{E}[\ln Z(\hat{G})] \geqslant \frac{\ln\xi}{k} + o(1). \tag{7.2}$$

(2) *For any $\varepsilon > 0$ there is $\delta = \delta(\varepsilon, P) > 0$, independent of n or d, such that uniformly for all $0 < d < D$,*

$$\mathbb{E}\langle\|\rho_{\boldsymbol{\sigma},\boldsymbol{\tau}} - \bar{\rho}\|_{\mathrm{TV}}\rangle_{\hat{G}} > \varepsilon \;\Rightarrow\; \frac{1}{n}\frac{\partial}{\partial d}\mathbb{E}[\ln Z(\hat{G})] \geqslant \frac{\ln\xi}{k} + \delta + o(1). \tag{7.3}$$

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

(3) *Conversely, we have*

$$\mathbb{E}\langle\|\rho_{\boldsymbol{\sigma},\boldsymbol{\tau}}-\bar{\rho}\|_{\mathrm{TV}}\rangle_{\hat{G}}=o(1)\;\Rightarrow\;\frac{1}{n}\frac{\partial}{\partial d}\mathbb{E}[\ln Z(\hat{G})]=\frac{\ln\xi}{k}+o(1). \tag{7.4}$$

Proof. The first two assertions and their proofs are nearly identical to the soft constraint version [25, Corollary 6.3]; we still include the brief argument for completeness and because it leads up to the proof of the third assertion. Due to (7.1) we obtain from Proposition 6.7 that uniformly for all $d<D$,

$$\frac{k}{n}\frac{\partial}{\partial d}\mathbb{E}[\ln Z(\hat{G})]=o(1)+\xi^{-1}\mathbb{E}[\Lambda(\langle\boldsymbol{\psi}(\boldsymbol{\sigma}(y_1),\ldots,\boldsymbol{\sigma}(y_k))\rangle_{\hat{G}})]. \tag{7.5}$$

Further, (5.6), Corollary 6.5 and **SYM** yield

$$\mathbb{E}\langle\boldsymbol{\psi}(\boldsymbol{\sigma}(y_1),\ldots,\boldsymbol{\sigma}(y_k))\rangle_{\hat{G}}=\mathbb{E}[\boldsymbol{\psi}(\hat{\boldsymbol{\sigma}}(y_1),\ldots,\hat{\boldsymbol{\sigma}}(y_k))]=\xi+o(1). \tag{7.6}$$

Since $\langle\boldsymbol{\psi}(\boldsymbol{\sigma}(x_1),\ldots,\boldsymbol{\sigma}(x_k))\rangle_{\hat{G}}\in(0,1]$ and $\Lambda''(x)\geqslant 1/2$ for all $x\in(0,1]$, Taylor's formula gives

$$\begin{aligned}
&\mathbb{E}[\Lambda(\langle\boldsymbol{\psi}(\boldsymbol{\sigma}(x_1),\ldots,\boldsymbol{\sigma}(x_k))\rangle_{\hat{G}})]\\
&\quad\geqslant\Lambda(\xi)+\Lambda'(\xi)[\mathbb{E}\langle\boldsymbol{\psi}(\boldsymbol{\sigma}(x_1),\ldots,\boldsymbol{\sigma}(x_k))\rangle_{\hat{G}}-\xi]+\frac{1}{4}\mathbb{E}[(\langle\boldsymbol{\psi}(\boldsymbol{\sigma}(x_1),\ldots,\boldsymbol{\sigma}(x_k))\rangle_{\hat{G}}-\xi)^2]\\
&\quad=\Lambda(\xi)+\frac{1}{4}\mathbb{E}[\langle\boldsymbol{\psi}(\boldsymbol{\sigma}(x_1),\ldots,\boldsymbol{\sigma}(x_k))\rangle_{\hat{G}}^2]-\frac{\xi^2}{4}+o(1)\quad\text{(by (7.6))}.
\end{aligned} \tag{7.7}$$

Thus, (7.2) is immediate from (7.6), (7.7) and Jensen's inequality.

Now assume that $\mathbb{E}\langle\|\rho_{\boldsymbol{\sigma},\boldsymbol{\tau}}-\bar{\rho}\|_{\mathrm{TV}}\rangle_{\hat{G}}>\varepsilon$. Since Corollary 6.5 and (5.6) yield

$$\mathbb{E}\langle\|\rho_{\boldsymbol{\sigma}}-\bar{\rho}\|_{\mathrm{TV}}+\|\rho_{\boldsymbol{\tau}}-\bar{\rho}\|_{\mathrm{TV}}\rangle_{\hat{G}}=o(1),$$

assumptions **MIN** and **SYM** imply that there is $\delta=\delta(\varepsilon)>0$ such that

$$\sum_{\sigma,\tau\in\Omega^k}\mathbb{E}\langle\boldsymbol{\psi}(\sigma)\boldsymbol{\psi}(\tau)\prod_{i=1}^{k}\rho_{\boldsymbol{\sigma},\boldsymbol{\tau}}(\sigma_i,\tau_i)\rangle_{\hat{G}}>\delta+o(1)+q^{-2k}\sum_{\sigma,\tau\in\Omega^k}\mathbb{E}[\boldsymbol{\psi}(\sigma)\boldsymbol{\psi}(\tau)]=\xi^2+\delta+o(1). \tag{7.8}$$

Moreover,

$$\begin{aligned}
\mathbb{E}[\langle\boldsymbol{\psi}(\boldsymbol{\sigma}(x_1),\ldots,\boldsymbol{\sigma}(x_k))\rangle_{\hat{G}}^2]&=\mathbb{E}\langle\boldsymbol{\psi}(\boldsymbol{\sigma}(x_1),\ldots,\boldsymbol{\sigma}(x_k))\boldsymbol{\psi}(\boldsymbol{\tau}(x_1),\ldots,\boldsymbol{\tau}(x_k))\rangle_{\hat{G}}\\
&=\sum_{\sigma,\tau\in\Omega^k}\mathbb{E}\langle\boldsymbol{\psi}(\sigma)\boldsymbol{\psi}(\tau)\prod_{i=1}^{k}\rho_{\boldsymbol{\sigma},\boldsymbol{\tau}}(\sigma_i,\tau_i)\rangle_{\hat{G}}.
\end{aligned}$$

Thus, (7.3) follows from (7.7) and (7.8).

With respect to the last assertion, we apply the full Taylor expansion

$$\Lambda(1-x)=-x+\sum_{\ell\geqslant 2}x^\ell/(\ell(\ell-1))$$

to obtain, due to (7.6), that

$$\mathbb{E}[\Lambda(\langle\boldsymbol{\psi}(\boldsymbol{\sigma}(x_1),\ldots,\boldsymbol{\sigma}(x_k))\rangle_{\hat{G}})]=\xi-1+o(1)+\mathbb{E}\left[\sum_{\ell\geqslant 2}\frac{1}{\ell(\ell-1)}\langle 1-\boldsymbol{\psi}(\boldsymbol{\sigma}(x_1),\ldots,\boldsymbol{\sigma}(x_k))\rangle_{\hat{G}}^{\ell}\right]$$

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

Since $0 \leqslant \boldsymbol{\psi} \leqslant 1$, all terms of the last sum are in $[0,1]$. Hence, invoking Fubini's theorem and writing $\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2, \ldots$ for independent samples from $\mu_{\hat{G}}$, we obtain

$$\begin{aligned}&\mathbb{E}[\Lambda(\langle \boldsymbol{\psi}(\boldsymbol{\sigma}(x_1), \ldots, \boldsymbol{\sigma}(x_k))\rangle_{\hat{G}})] \\ &\quad = \xi - 1 + o(1) + \sum_{\ell \geqslant 2} \frac{1}{\ell(\ell-1)} \mathbb{E}\left\langle \prod_{h=1}^{\ell} 1 - \boldsymbol{\psi}(\boldsymbol{\sigma}_h(x_1), \ldots, \boldsymbol{\sigma}_h(x_k)) \right\rangle_{\hat{G}}. \end{aligned} \tag{7.9}$$

Moreover, since $\boldsymbol{\psi}$ is drawn independently of $\hat{G}$, we obtain

$$\begin{aligned}&\mathbb{E}\left\langle \prod_{h=1}^{\ell} 1 - \boldsymbol{\psi}(\boldsymbol{\sigma}_h(x_1), \ldots, \boldsymbol{\sigma}_h(x_k)) \right\rangle_{\hat{G}} \\ &\quad = \sum_{\chi \in \Omega^{\ell \times k}} \mathbb{E}\left[\prod_{h=1}^{\ell} (1 - \boldsymbol{\psi}(\chi_{h,1}, \ldots, \chi_{h,k})) \right] \\ &\qquad \times \mathbb{E}\left\langle \prod_{i=1}^{k} \mathbf{1}\{(\boldsymbol{\sigma}_1(x_i) = \chi_{1,i}, \ldots, \boldsymbol{\sigma}_\ell(x_i) = \chi_{\ell,i})\} \right\rangle_{\hat{G}}. \end{aligned} \tag{7.10}$$

We now claim that for any $\ell \geqslant 2$ and for any $\chi \in \Omega^{\ell \times k}$,

$$\mathbb{E}\left\langle \prod_{i=1}^{k} \mathbf{1}\{(\boldsymbol{\sigma}_1(x_i) = \chi_{1,i}, \ldots, \boldsymbol{\sigma}_\ell(x_i) = \chi_{\ell,i})\} \right\rangle_{\hat{G}} = q^{-k\ell} + o(1). \tag{7.11}$$

Indeed, by Lemma 4.4 the assumption $\mathbb{E}\langle \|\rho_{\boldsymbol{\sigma},\boldsymbol{\tau}} - \bar{\rho}\|_{\mathrm{TV}} \rangle_{\hat{G}} = o(1)$ implies that $\mu_{\hat{G}}$ is $o(1)$-symmetric w.h.p. and that its marginals satisfy $\sum_{i=1}^{n} \|\mu_{\hat{G},x_i} - \bar{\rho}\|_{\mathrm{TV}} = o(n)$. Hence, Lemma 4.3 shows that the ℓ-fold product measure $\mu_{\hat{G}}^{\otimes \ell}$ is $o(1)$-symmetric with asymptotically uniform marginals as well w.h.p. Thus, we obtain (7.11). Finally, plugging (7.11) into (7.10) and (7.10) into (7.9) and applying **SYM**, we obtain the assertion. □

Lemma 7.2 (**SYM**, **BAL**). *For any $\varepsilon > 0$, $d > 0$ there is $0 < \delta = \delta(\varepsilon, d, P) < \varepsilon$ such that the following holds. Assume that $m \in \mathcal{M}(d)$ is a sequence such that*

$$\limsup_{n \to \infty} \mathbb{E}\langle \|\rho_{\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2} - \bar{\rho}\|_{\mathrm{TV}} \rangle_{\hat{G}(n,m)} > \varepsilon. \tag{7.12}$$

Then

$$\limsup_{n \to \infty} \min\{\mathbb{E}\langle \|\rho_{\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2} - \bar{\rho}\|_{\mathrm{TV}} \rangle_{\hat{G}(n,m)} : \delta n < m - dn/k < 2\delta n\} > \delta.$$

Lemma 7.2 and its proof are syntactically identical to the soft constraint version [25, Lemma 6.1]. The proof is included in Appendix C for the sake of completeness.

Proof of Proposition 5.7. The proof of the first assertion is nearly identical to the soft constraint version [25, proof of Proposition 3.3]; we include the argument for completeness. Assume that there exist $D_0 < d_{\mathrm{cond}}$, $\varepsilon > 0$ such that

$$\limsup_{n \to \infty} \mathbb{E}\langle \|\rho_{\boldsymbol{\sigma},\boldsymbol{\tau}} - \bar{\rho}\|_{\mathrm{TV}} \rangle_{\hat{G}(n,m(D_0,n))} > \varepsilon.$$

Then Lemma 7.2 shows that there is $\delta > 0$ such that with $D_1 = D_0 + 3\delta/2 < d_{\mathrm{cond}}$ for infinitely many n we have

$$\mathbb{E}\langle \|\rho_{\boldsymbol{\sigma},\boldsymbol{\tau}} - \bar{\rho}\|_{\mathrm{TV}} \rangle_{\hat{G}(n,m)} > \delta + o(1) \quad \text{for all } D_0 + 4\delta/3 < d < D_1.$$

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

Hence, Lemma 7.1 implies that for infinitely many n,

$$\frac{1}{n}\mathbb{E}[\ln Z(\hat{G}(n, m(D_1, n)))] = \frac{1}{n}\mathbb{E}[\ln Z(\hat{G}(n, m(D_0, n)))] + \frac{1}{n}\int_{D_0}^{D_1} \frac{\partial}{\partial d}\mathbb{E}[\ln Z(\hat{G})]\mathrm{d}d$$
$$\geqslant \ln q + \frac{D_1}{k}\ln\xi + \Omega(1).$$

But then the second part of Theorem 5.3 yields

$$\sup_{\pi\in\mathcal{P}^2_*(\Omega)} \mathcal{B}(D_1, P, \pi) > \ln q + \frac{D_1}{k}\ln\xi,$$

in contradiction to $D_1 < d_{\mathrm{cond}}$.

Analogously, the second assertion follows from the third part of Lemma 7.1 by integrating on d. Specifically, assume that $D > 0$ is such that (5.16) is true for all $d < D$. Pick some $\Delta < D$. Then by the third part of Lemma 7.1 and dominated convergence,

$$\mathbb{E}[\ln Z(\hat{G}(n, m(\Delta, n))] = \ln q + \int_0^{\Delta} \frac{\partial}{\partial d}\mathbb{E}[\ln Z(\hat{G})]\mathrm{d}d = \ln q + \frac{\Delta}{k}\ln\xi + o(1).$$

Hence, Theorem 5.3 yields $\sup_{\pi\in\mathcal{P}^2_*(\Omega)} \mathcal{B}(\Delta, P, \pi) \leqslant \ln q + k^{-1}\Delta\ln\xi$. As this holds for all $\Delta \leqslant D$, we conclude that $d_{\mathrm{cond}} \geqslant D$. □

7.2 *Proof of Proposition 5.11*

The distribution of the random variables $C_Y(G(n, m))$ of the 'plain' random CSP can be calculated via a totally standard method of moments argument as set out in [21]. Our assumption that $P(\psi) > 0$ for all $\psi \in \Psi$ ensures that $\kappa_Y > 0$ for all signatures Y.

Lemma 7.3 [21]. *Let $d > 0$. For any $Y \in \mathcal{Y}$ we have $\mathbb{E}[C_Y(G(n, m))] \sim \kappa_Y$, uniformly for all $m \in \mathcal{M}(d)$. Moreover, if $Y_1, \ldots, Y_l \in \mathcal{Y}$ are pairwise disjoint and $y_1, \ldots, y_l \geqslant 0$, then, uniformly for all $m \in \mathcal{M}(d)$,*

$$\mathbb{P}[\forall i \leqslant l : C_{Y_i}(G(n, m)) = y_i] \sim \prod_{t=1}^{l} \mathbb{P}[\mathrm{Po}(\kappa_{Y_t}) = y_t]. \tag{7.13}$$

In order to determine the joint distribution of the random variables $C_Y(\hat{G}(n, m))$ we use the method of moments as well. More specifically, the argument is nearly identical to the one from [25], except that here it may be possible that $\hat{\kappa}_Y = 0$ for some signatures Y.

Lemma 7.4 (**SYM**, **BAL**). *Let $d > 0$. For any $Y \in \mathcal{Y}$ we have $\mathbb{E}[C_Y(\hat{G}(n, m))] = \hat{\kappa}_Y + o(1)$, uniformly for all $m \in \mathcal{M}(d)$. Moreover, if $Y_1, \ldots, Y_L \in \mathcal{Y}$ are pairwise distinct and $y_1, \ldots, y_L \geqslant 0$, then, uniformly for all $m \in \mathcal{M}(d)$,*

$$\mathbb{P}[\forall i \leqslant \ell : C_{Y_i}(\hat{G}(n, m)) = y_i] = o(1) + \prod_{l=1}^{L} \mathbb{P}[\mathrm{Po}(\hat{\kappa}_{Y_l}) = y_l].$$

The proof of Lemma 7.4 can be found in Appendix D.

Proof of Proposition 5.11. The fact that $P(\psi) > 0$ for all $\psi \in \Psi$ implies immediately that $\kappa_Y > 0$ for all signatures Y. Moreover, condition **UNI** implies that $\hat{\kappa}_Y = 0$ can hold only if Y has order one. Further, if indeed $\hat{\kappa}_Y = 0$, then the corresponding cycle is unsatisfiable deterministically and thus any factor graph G that contains such a cycle satisfies $Z(G) = 0$. Consequently, (5.5) ensures that $\mathbb{P}[C_Y(\hat{G}(n, m)) > 0] = 0$ for all n, m. The asymptotic identity (5.24) is immediate from Lemma 7.3. Moreover, Lemma 7.4 directly implies (5.25). □

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

7.3 Proof of Theorem 2.5

The first part of Theorem 2.5 readily follows from Theorem 2.7.

Lemma 7.5 (**SYM**, **BAL**, **MIN**, **UNI**). *If* $d < d_{\mathrm{cond}}$, *then* $\lim_{n\to\infty} \mathbb{E}\sqrt[n]{Z(\mathbb{G})} = q\xi^{d/k}$.

Proof. Since $\mathcal{K} > 0$ almost surely and because $\mathrm{Eig}(\Phi) \subset (-\infty, 0] \cup \{1\}$ by Lemma 5.4, the assertion is immediate from Theorem 2.7. □

To prove the second part of Theorem 2.5 concerning $d > d_{\mathrm{cond}}$ we generalize an argument for the random graph colouring problem from [27, Section 4] to the present broad class of random CSPs. We begin with the following general fact that essentially goes back to [3]. Let $\mathcal{M}_\varepsilon(d)$ be the set of all sequences $m = m(n)$ such that $|m(n) - dn/k| \leqslant \varepsilon n$ for all n.

Lemma 7.6 (**SYM**, **BAL**). *Let* $d > 0$. *For any* $\delta, \eta > 0$ *there is* $\varepsilon > 0$ *such that the following is true. Suppose that* $(\mathcal{E}_n)_n$ *is a sequence of events such that uniformly for all* $m \in \mathcal{M}_\eta(d)$,

$$\limsup_{n\to\infty} \mathbb{P}[G(n,m) \notin \mathcal{E}_n]^{1/n} < 1-\delta \quad \textit{while} \quad \limsup_{n\to\infty} \mathbb{P}[\hat{G}(n,m) \in \mathcal{E}_n]^{1/n} < 1-\delta.$$

Then, uniformly for all $m \in \mathcal{M}_\eta(d)$,

$$\limsup_{n\to\infty} \mathbb{P}[\sqrt[n]{Z(G(n,m))} \geqslant q\xi^{d/k} - \varepsilon]^{1/n} < 1-\varepsilon. \tag{7.14}$$

Proof. Pick $\varepsilon = \varepsilon(\delta) > 0$ sufficiently small, $\mathcal{U}_n = \{\sqrt[n]{Z} \geqslant q\xi^{d/k} - \varepsilon\}$ and assume that

$$\limsup \mathbb{P}[G(n,m) \in \mathcal{U}_n]^{1/n} = 1.$$

Then the assumption $\limsup \mathbb{P}[G(n,m) \notin \mathcal{E}_n]^{1/n} < 1-\delta$ implies that for infinitely many n,

$$\mathbb{P}[G(n,m) \notin \mathcal{E}_n \mid G(n,m) \in \mathcal{U}_n]^{1/n} \leqslant \left(\frac{\mathbb{P}[G(n,m) \notin \mathcal{E}_n]}{\mathbb{P}[G(n,m) \in \mathcal{U}_n]}\right)^{1/n} < 1-\delta+o(1).$$

Hence, Proposition 5.9 shows that for infinitely many n,

$$\begin{aligned}\mathbb{P}[\hat{G}(n,m) \in \mathcal{E}_n] &\geqslant \frac{\mathbb{E}[Z(G(n,m))\mathbf{1}\{Z(G(n,m)) \in \mathcal{U}_n \cap \mathcal{E}_n\}]}{\mathbb{E}[Z(G(n,m))]} \\ &\geqslant \left(\frac{q\xi^{d/k}-\varepsilon}{q\xi^{d/k}}\right)^{n+o(n)} \mathbb{P}[G(n,m) \in \mathcal{E}_n \mid G(n,m) \in \mathcal{U}_n]\mathbb{P}[G(n,m) \in \mathcal{U}_n] \\ &= \exp(o(n)),\end{aligned}$$

in contradiction to the assumption that

$$\limsup \mathbb{P}[\hat{G}(n,m) \in \mathcal{E}_n]^{1/n} < 1-\delta.$$

Thus,

$$\limsup \mathbb{P}[G(n,m) \in \mathcal{U}_n]^{1/n} < 1.$$

Choosing $\varepsilon > 0$ small enough we obtain (7.14). □

Lemma 7.7 (**SYM**, **BAL**, **POS**). *For any* $d > d_{\mathrm{cond}}$ *there exists* $\beta, \delta, \eta > 0$ *such that uniformly for all* $m \in \mathcal{M}_\eta(d)$,

$$\mathbb{P}\left[\frac{1}{n}\ln Z_\beta(\hat{G}(n,m)) < \ln q + \frac{d}{k}\ln\xi_\beta + \delta\right] < \exp(-\Omega(n)), \quad \textit{while} \tag{7.15}$$

$$\mathbb{P}\left[\frac{1}{n}\ln Z_\beta(G(n,m)) \geqslant \ln q + \frac{d}{k}\ln\xi_\beta + \delta\right] < \exp(-\Omega(n)). \tag{7.16}$$

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

Proof. Lemma 5.2 shows that Proposition 5.9 applies to $G(n, m, P_\beta)$. Thus,

$$\mathbb{E}[Z(G(n, m, P_\beta))] = O(q^n \xi_\beta^m)$$

and (7.16) is immediate from Markov's inequality.

We move on to the proof of (7.15). The definition of d_{cond} implies that for any $d > d_{\mathrm{cond}}$ there exist $d' < d$ and $\pi \in \mathcal{P}_*^2(\Omega)$ such that $\mathcal{B}(d', P, \pi) > \ln q + (d' \ln \xi)/k$. Hence, Theorem 5.3 shows that there is $\delta > 0$ and $n_0 > 0$ such that $\mathbb{E}[\ln Z(\hat{G})] > n(\ln q + (d \ln \xi)/k + 8\delta)$ for all $n > n_0$. Moreover, by construction we have $Z_\beta(\hat{G}) \geqslant Z(\hat{G})$ and $\lim_{\beta\to\infty} \xi_\beta = \xi$. Therefore, there exists $\beta_0 > 0$ such that

$$\mathbb{E}[\ln Z_\beta(\hat{G})] > n(\ln q + (d \ln \xi_\beta)/k + 7\delta) \quad \text{for all } n > n_0,\ \beta > \beta_0. \tag{7.17}$$

Due to the Nishimori identity Lemma 5.1 we can write (7.17) as

$$\mathbb{E}[\ln Z_\beta(G^*(n, m, \hat{\boldsymbol{\sigma}}_{n,m})] > n(\ln q + (d \ln \xi_\beta)/k + 7\delta) \quad \text{for all } n > n_0,\ \beta > \beta_0. \tag{7.18}$$

Now, fix $\beta > \beta_0$, pick a small enough $\eta = \eta(\beta, \delta) > 0$ and let $\mathcal{A}$ be the set of all assignments $\sigma : V_n \to \Omega$ such that $\|\rho_\sigma - \bar{\rho}\|_{\mathrm{TV}} < \eta$. Fix any $\sigma_0 \in \mathcal{A}$. Given $\hat{\boldsymbol{\sigma}}_{n,m} \in \mathcal{A}$ we can couple $G' = G^*(n, m, \hat{\boldsymbol{\sigma}}_{n,m})$ and $G'' = G^*(n, m, \sigma_0)$ such that

$$\mathbb{P}[|\ln Z_\beta(G') - \ln Z_\beta(G'')| > \delta n \mid \hat{\boldsymbol{\sigma}}_{n,m} \in \mathcal{A}] \leqslant \exp(-\Omega(n)). \tag{7.19}$$

Indeed, relabelling the variables if necessary, given $\hat{\boldsymbol{\sigma}}_{n,m} \in \mathcal{A}$ we may assume that $|\sigma_0 \triangle \hat{\boldsymbol{\sigma}}_{n,m}| \leqslant 2q\eta n$. Further, the planted model can alternatively be described as the result of adding constraints independently according to (5.3). Let X and Y be the number of constraints of G' and G'' respectively that are adjacent to a variable in $\sigma_0 \triangle \hat{\boldsymbol{\sigma}}_{n,m}$. Then X, Y are binomial random variables and (5.3) shows that $\mathbb{E}[X + Y] < \delta n/(4\beta)$ if $\eta > 0$ is chosen small enough. Now, we couple the constraints that are non-adjacent to $\sigma_0 \triangle \hat{\boldsymbol{\sigma}}_{n,m}$ in either random CSP instance identically, and the at most $X + Y$ constraints that are adjacent to $\sigma_0 \triangle \hat{\boldsymbol{\sigma}}_{n,m}$ independently. Hence, G', G'' differ in no more than $X + Y$ constraints. Since the construction of the softened constraints ψ_β ensures that the addition or removal of a single constraint can change the partition function by at most a factor of $\exp(\pm\beta)$, we conclude that $|\ln Z_\beta(G') - \ln Z_\beta(G'')| \leqslant \beta(X + Y)$. Since $\mathbb{E}[X + Y] < \delta n/(4\beta)$, (7.19) follows from the Chernoff bound. Furthermore, (7.19) implies together with Corollary 6.5 that

$$\mathbb{P}[|\ln Z_\beta(G') - \ln Z_\beta(G'')| > \delta n] \leqslant \exp(-\Omega(n)). \tag{7.20}$$

Let $m \in \mathcal{M}_\eta(d)$. We can couple $G'' = G^*(n, m, \sigma_0)$ and $G''' = G^*(n, m, \sigma_0)$ such that both CSP instances coincide on $m \wedge m$ constraints. Since m is a Poisson variable, it is therefore exponentially unlikely that G'', G''' differ on more than $\delta n/(2\beta)$ constraints, providing η is small enough. Consequently,

$$\mathbb{P}[|\ln Z_\beta(G'') - \ln Z_\beta(G''')| > \delta n] \leqslant \exp(-\Omega(n)) \tag{7.21}$$

uniformly for all $m \in \mathcal{M}_\eta(d)$. Combining (7.18), (7.20) and (7.21), we obtain

$$\mathbb{E}[\ln Z_\beta(G''')] > n(\ln q + (d \ln \xi_\beta)/k + 4\delta) \tag{7.22}$$

uniformly for all $m \in \mathcal{M}_\eta(d)$. Furthermore, since G''' consists of independent constraints drawn from the distribution (5.3) and because each of these constraints can shift $\ln Z_\beta(G''')$ by no more than $\pm\beta$, Azuma's inequality and (7.22) yield

$$\mathbb{P}[\ln Z_\beta(G''') \leqslant n(\ln q + (d \ln \xi_\beta)/k + 3\delta)] \leqslant \exp(-\Omega(n)) \tag{7.23}$$

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

uniformly for all $m \in \mathcal{M}_\eta(d)$. Finally, we couple G''' and $G'''' = G(n, m, \hat{\boldsymbol{\sigma}}_{n,m})$ just as in the proof of (7.18) to see that uniformly for all $m \in \mathcal{M}_\eta(d)$,

$$\mathbb{P}[|\ln Z_\beta(G''') - \ln Z_\beta(G'''')| > \delta n \mid \hat{\boldsymbol{\sigma}}_{n,m} \in \mathcal{A}] \leqslant \exp(-\Omega(n)). \tag{7.24}$$

Combining Corollary 6.5 with (7.23) and (7.24), we obtain (7.15). □

Proof of Theorem 2.5. The first part of the theorem is immediate from Lemma 7.5. With respect to the second assertion suppose that $d > d_{\mathrm{cond}}$ and fix β, δ, η as provided by Lemma 7.7. Then the events

$$\mathcal{E}_n = \left\{\frac{1}{n}\ln Z_\beta(G(n,m)) < \ln q + \frac{d}{k}\ln \xi_\beta + \delta\right\}. \tag{7.25}$$

satisfy the assumptions of Lemma 7.6. Since for any $\eta > 0$ we have $\mathbb{P}[|m - dn/k| > \eta] \leqslant \exp(-\Omega(n))$, Lemma 7.6 thus shows that $\mathbb{P}[Z(G) > q\xi^{d/k} - \varepsilon] \leqslant \exp(-\varepsilon n + o(n))$ for some $\varepsilon > 0$. Because $\mathbb{G}$ is distributed as G given $\mathfrak{S}$ and $\mathbb{P}[\mathfrak{S}] = \Omega(1)$ by Proposition 5.11, the second assertion follows. □

7.4 Proof of Theorem 2.11

To prove the contiguity statement we first show that $\hat{G}(n, m)$ and $G^*(n, m, \boldsymbol{\sigma}^*)$ are mutually contiguous. More specifically, we have the following.

Lemma 7.8 (**SYM**, **BAL**). *For any $D, \varepsilon > 0$ let there exist $\delta > 0$ and $n_0 > 0$ such that, for all $n > n_0$ and all $m \leqslant Dn/k$, the following two statements are true.*

(1) *If $\mathcal{E}$ is an event such that $\mathbb{P}[(G^*(n, m, \boldsymbol{\sigma}^*), \boldsymbol{\sigma}^*) \in \mathcal{E}] < \delta$, then*

$$\mathbb{P}[(G^*(n, m, \hat{\boldsymbol{\sigma}}_{n,m}), \hat{\boldsymbol{\sigma}}_{n,m}) \in \mathcal{E}] < \varepsilon.$$

(2) *If $\mathcal{E}$ is an event such that $\mathbb{P}[(G^*(n, m, \hat{\boldsymbol{\sigma}}_{n,m}), \hat{\boldsymbol{\sigma}}_{n,m}) \in \mathcal{E}] < \delta$, then*

$$\mathbb{P}[(G^*(n, m, \boldsymbol{\sigma}^*), \boldsymbol{\sigma}^*) \in \mathcal{E}] < \varepsilon.$$

The proof of Lemma 7.8 is identical to that of the soft constraint version [25, Corollary 4.8]. The details can be found in Appendix E.

Proof of Theorem 2.11. We use a similar argument as in [25], except that here we explicitly deal with the conditioning on $\mathfrak{S}$. With respect to the first assertion, suppose that $d < d_{\mathrm{cond}}$ and let $(\mathcal{E}_n)_n$ be a sequence of events. Let us first assume that $\mathbb{P}[\mathbb{G}^* \in \mathcal{E}_n] = o(1)$. Then Proposition 5.11 implies that $\mathbb{P}[G^* \in \mathcal{E}_n \cap \mathfrak{S}] = o(1)$. Thus, Lemmas 5.1 and 7.8 yield $\mathbb{P}[\hat{G} \in \mathcal{E}_n \cap \mathfrak{S}] = o(1)$. Furthermore, Theorem 2.7 shows that for any $\varepsilon > 0$ there is $\delta > 0$ such that $\mathbb{P}[Z(\mathbb{G}) < \delta q^n \xi^m] < \varepsilon$ for large enough n, because $\mathcal{K} > 0$ almost surely. Consequently,

$$\begin{aligned}
\mathbb{P}[\mathbb{G} \in \mathcal{E}_n] &\leqslant \varepsilon + \mathbb{P}[\mathbb{G} \in \mathcal{E}_n,\ Z(\mathbb{G}) \geqslant \delta q^n \xi^m] \\
&= \varepsilon + \mathbb{P}[G \in \mathcal{E}_n,\ Z(G) \geqslant \delta q^n \xi^m \mid \mathfrak{S}] \\
&= \varepsilon + \frac{\mathbb{P}[G \in \mathcal{E}_n \cap \mathfrak{S},\ Z(G) \geqslant \delta q^n \xi^m]}{\mathbb{P}[G \in \mathfrak{S}]} \\
&\leqslant \varepsilon + \frac{1}{\delta \mathbb{P}[G \in \mathfrak{S}]} \cdot \mathbb{E}\left[\frac{\mathbb{E}[Z(G)\mathbf{1}\{G \in \mathcal{E} \cap \mathfrak{S}\} \mid m]}{q^n \xi^m}\right].
\end{aligned}$$

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

Hence, combining (5.5), Propositions 5.9 and 5.11, we obtain a number $c = c(P, d) > 0$ such that for large n

$$\mathbb{P}[\mathbb{G} \in \mathcal{E}_n] \leqslant \varepsilon + c \cdot \mathbb{P}[\hat{G} \in \mathcal{E}_n \cap \mathfrak{S}].$$

Since this bound holds for any $\varepsilon > 0$ and because $\mathbb{P}[\hat{G} \in \mathcal{E}_n \cap \mathfrak{S}] = o(1)$, we conclude that $\mathbb{P}[\mathbb{G} \in \mathcal{E}_n] = o(1)$.

Conversely, assume that $\mathbb{P}[\mathbb{G} \in \mathcal{E}_n] = o(1)$. Then $\mathbb{P}[G \in \mathcal{E}_n \cap \mathfrak{S}] = o(1)$. Hence, as $\mathbb{P}[\mathcal{Z}(\hat{G}) = Z(\hat{G})] = 1 - o(1)$ by Proposition 5.7, we obtain

$$\begin{aligned}\mathbb{P}[\hat{G} \in \mathcal{E}_n \cap \mathfrak{S}] &= o(1) + \mathbb{P}[\hat{G} \in \mathcal{E}_n \cap \mathfrak{S},\ \mathcal{Z}(\hat{G}) = Z(\hat{G})] \\ &\leqslant o(1) + \mathbb{E}\left[\frac{\mathbb{E}[\mathcal{Z}(G)\mathbf{1}\{G \in \mathcal{E}_n \cap \mathfrak{S}\} \mid m]}{\mathbb{E}[Z(G) \mid m]} \cdot \mathbf{1}\{|m - dn/k| \leqslant \sqrt{n} \ln n\}\right]. \qquad (7.26)\end{aligned}$$

Further, the second moment bound from Proposition 5.10 shows together with the formula for the first moment from Proposition 5.9 that on the event $\{|m - dn/k| \leqslant \sqrt{n} \ln n\}$ the quotient $\mathbb{E}[\mathcal{Z}(G)^2|m]/\mathbb{E}[Z(G)|m]^2$ is bounded. Hence, for any $\varepsilon > 0$ there is $C = C(\varepsilon, P, d) > 0$ such that $\mathbb{E}[\mathbf{1}\{\mathcal{Z}(G) > C\mathbb{E}[Z(G)]\}] < \varepsilon$. Therefore, (7.26) yields

$$\mathbb{P}[\hat{G} \in \mathcal{E}_n \cap \mathfrak{S}] \leqslant o(1) + \varepsilon + C \cdot \mathbb{P}[G \in \mathcal{E}_n \cap \mathfrak{S}].$$

Since this bound holds for every fixed $\varepsilon > 0$ and $\mathbb{P}[G \in \mathcal{E}_n \cap \mathfrak{S}] = o(1)$, we obtain

$$\mathbb{P}[\hat{G} \in \mathcal{E}_n \cap \mathfrak{S}] = o(1).$$

Finally, since G^* and $\hat{G}$ are mutually contiguous by Lemma 7.8 and since $\mathbb{P}[\hat{G} \in \mathfrak{S}] = \Omega(1)$ by Proposition 5.11, we obtain

$$\mathbb{P}[\mathbb{G}^* \in \mathcal{E}_n \cap \mathfrak{S}] = \mathbb{P}[G^* \in \mathcal{E}_n \cap \mathfrak{S} \mid \mathfrak{S}] = o(1),$$

as desired.

Now assume that $d > d_{\mathrm{cond}}$. The events $\mathcal{E}_n$ from (7.25) satisfy the assumptions of Lemma 7.6. Thus

$$\mathbb{P}[G \in \mathcal{E}_n] = 1 - \exp(-\Omega(n)),$$

while $\mathbb{P}[\hat{G} \in \mathcal{E}_n] = \exp(-\Omega(n))$. Indeed, since $\mathbb{P}[G \in \mathfrak{S}], \mathbb{P}[\hat{G} \in \mathfrak{S}] = \Omega(1)$ by Proposition 5.11, we conclude that

$$\mathbb{P}[\mathbb{G} \in \mathcal{E}_n] = 1 - \exp(-\Omega(n)) = 1 - o(1),$$

while $\mathbb{P}[\mathbb{G}^* \in \mathcal{E}_n] = o(1)$ by Lemma 7.8. Thus, $\mathbb{G}$ and $\mathbb{G}^*$ are mutually orthogonal. □

7.5 Proof of Corollary 2.12

Assume that $(\mathcal{E}_n)_n$ is a sequence of events such that $\mathbb{P}[(\mathbb{G}, \boldsymbol{\sigma}) \in \mathcal{E}_n] = o(1)$. Then there exists a sequence $\varepsilon_n = o(1)$ such that the events $\mathcal{E}'_n = \{\langle \mathbf{1}\{(\mathbb{G}, \boldsymbol{\sigma}) \in \mathcal{E}_n\}\rangle_{\mathbb{G}} \geqslant \varepsilon_n\}$ satisfy $\mathbb{P}[\mathbb{G} \in \mathcal{E}'_n] = o(1)$. Hence, Theorem 2.11 yields

$$\mathbb{P}[\mathbb{G}^* \in \mathcal{E}'_n] = \mathbb{P}[G^* \in \mathcal{E}'_n \mid \mathfrak{S}] = o(1)$$

and thus

$$\mathbb{P}[G^* \in \mathcal{E}'_n \cap \mathfrak{S}] = o(1).$$

Therefore, applying Lemma 7.8, we obtain

$$\mathbb{P}[\hat{G} \in \mathcal{E}'_n \cap \mathfrak{S}] = o(1),$$

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

whence Lemma 5.1 yields

$$\mathbb{P}[(\hat{G}, \hat{\boldsymbol{\sigma}}) \in \mathcal{E}_n \cap \mathfrak{S}] = o(1).$$

Thus, applying Lemma 7.8 a second time, we obtain

$$\mathbb{P}[(G^*, \boldsymbol{\sigma}^*) \in \mathcal{E}_n \cap \mathfrak{S}] = o(1).$$

Since the probability of $\mathfrak{S}$ is bounded away from 0 by Proposition 5.11, we finally obtain

$$\mathbb{P}[(\mathbb{G}^*, \boldsymbol{\sigma}^*) \in \mathcal{E}_n] = \mathbb{P}[(G^*, \boldsymbol{\sigma}^*) \in \mathcal{E}_n \mid \mathfrak{S}] = o(1).$$

Conversely, assume that $\mathbb{P}[(\mathbb{G}^*, \boldsymbol{\sigma}^*) \in \mathcal{E}_n] = o(1)$. Then

$$\mathbb{P}[(G^*, \boldsymbol{\sigma}^*) \in \mathcal{E}_n \cap \mathfrak{S}] = o(1)$$

and thus Lemma 7.8 yields

$$\mathbb{P}[(\hat{G}, \hat{\boldsymbol{\sigma}}) \in \mathcal{E}_n \cap \mathfrak{S}] = o(1).$$

Hence, Lemma 5.1 shows that there exists a sequence $\varepsilon_n = o(1)$ such that for the event

$$\mathcal{E}'_n = \{\langle \mathbf{1}\{(\hat{G}, \boldsymbol{\sigma}) \in \mathcal{E}_n\}\rangle_{\hat{G}} \geqslant \varepsilon_n\}$$

we have

$$\mathbb{P}[\hat{G} \in \mathcal{E}'_n \cap \mathfrak{S}] = o(1).$$

Thus, Lemma 7.8 yields

$$\mathbb{P}[G^* \in \mathcal{E}'_n \cap \mathfrak{S}] = o(1)$$

and therefore

$$\mathbb{P}[\mathbb{G}^* \in \mathcal{E}'_n] = o(1)$$

by Proposition 5.11. Consequently, Theorem 2.11 yields

$$\mathbb{P}[\mathbb{G} \in \mathcal{E}'_n] = o(1).$$

Finally, unravelling the definition of $\mathcal{E}'_n$, we obtain

$$\mathbb{P}[(\mathbb{G}, \boldsymbol{\sigma}) \in \mathcal{E}_n] = o(1).$$

7.6 Proof of Theorem 2.8

Suppose that $d < d_{\mathrm{cond}}$. Then Proposition 5.7, Proposition 5.11 and Lemma 7.8 yield

$$\mathbb{E}\langle \|\rho_{\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2} - \bar{\rho}\|_{\mathrm{TV}}\rangle_{\mathbb{G}^*} = o(1).$$

Hence, Theorem 2.11 implies

$$\mathbb{E}\langle \|\rho_{\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2} - \bar{\rho}\|_{\mathrm{TV}}\rangle_{\mathbb{G}} = o(1),$$

as desired.

8. Reconstruction and local weak convergence

In this section we prove Theorems 2.9 and 2.10. Recall that for a variable x of a CSP instance G we let $\nabla_{2\ell}(G, x)$ denote the depth-2ℓ neighbourhood of x, rooted at x. Moreover, let $\partial^{2\ell}(G, x)$ be the set of variables at distance precisely 2ℓ from x. We drop G from the notation where the reference is apparent.

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

8.1 Proof of Theorem 2.9

We remember the random CSP $\mathbb{T} = \mathbb{T}(d, P)$ generated by a Galton–Watson process that describes the local neighbourhood structure and we continue to denote its root by r. Moreover, we write $\mathbb{T}^{2\ell} = \mathbb{T}^{2\ell}(d, P)$ for the CSP instance obtained from $\mathbb{T}$ by deleting all constraints and variables at a distance greater than 2ℓ from r. Due to condition **SYM** the partition function $Z(\mathbb{T}^{2\ell})$ is strictly positive. Hence, throughout this section we let $\boldsymbol{\chi}^{2\ell}$ denote a sample from the Boltzmann distribution $\mu_{\mathbb{T}^{2\ell}}$. The following lemma shows that the Galton–Watson process $\mathbb{T}$ also describes the local structure of the planted random CSP G^*.

Lemma 8.1 (**SYM**). *Let $\ell \geqslant 1$. For any possible outcome T of $\mathbb{T}^{2\ell}$ and for any assignment $\chi : V(T) \to \Omega$, the following is true. Let X be the number of variables of G^* for which there exists an isomorphism $\vartheta : T \to \nabla_{2\ell}(G^*, x)$ such that $\chi = \boldsymbol{\sigma}^* \circ \vartheta$. Then X/n converges to*

$$\mathbb{P}[\mathbb{T}^{2\ell} \cong T, \boldsymbol{\chi}^{2\ell} = \chi]$$

in probability.

Proof. Consider the following enhanced multi-type Galton–Watson process $(\hat{\mathbb{T}}, \hat{\boldsymbol{\chi}})$ whose types are variables x endowed with values $\boldsymbol{\chi}(x)$ and constraints endowed with weight functions $\psi \in \Psi$ and indices $h \in [k]$. The process starts from the tree $\hat{\mathbb{T}}^0$ consisting of the root r only, for which a value $\hat{\boldsymbol{\chi}}^0(r) \in \Omega$ is chosen uniformly at random. Then $\hat{\mathbb{T}}^{2\ell+2}, \hat{\boldsymbol{\chi}}^{2\ell+2}$ is obtained by appending two more layers to $\hat{\mathbb{T}}^{2\ell}, \hat{\boldsymbol{\chi}}^{2\ell}$ as follows. Each variable x of $\hat{\mathbb{T}}^{2\ell}$ at distance exactly 2ℓ from r independently generates $D = \mathrm{Po}(d)$ constraints $a_{x,1}, \ldots, a_{x,D}$ as offspring. The associated constraint functions $\psi_{a_{x,i}}$ are drawn independently from P, and the position $h_{x,i}$ where x appears in the constraint $a_{x,i}$ is drawn uniformly from $[k]$, independently for every i. Further, each constraint $a_{x,i}$ spawns $k-1$ variables $(y_{x,i,j})_{j \in [k] \setminus \{h_{x,i}\}}$. Their values $\hat{\boldsymbol{\chi}}^{2\ell+2}(y_{x,i,j})$ are jointly drawn from the distribution

$$\begin{aligned}&\mathbb{P}[\forall j \neq h_{x,i} : \hat{\boldsymbol{\chi}}^{2\ell+2}(y_{x,i,j}) = \sigma_j \mid \hat{\boldsymbol{\chi}}^{2\ell}(x)] \\ &\quad = q^{1-k}\xi^{-1}\psi_{a_{x,i}}(\sigma_1, \ldots, \sigma_{h_{x,i}-1}, \hat{\boldsymbol{\chi}}^{2\ell}(x), \sigma_{h_{x,i}+1}, \ldots, \sigma_k) \quad (\sigma_j \in \Omega). \end{aligned} \tag{8.1}$$

In other words, the $\hat{\boldsymbol{\chi}}^{2\ell+2}(y_{x,i,j})$ are chosen with probability proportional to the weight induced by $\psi_{a,i}$ given that x has value $\hat{\boldsymbol{\chi}}^{2\ell}(x)$.

Crucially, **SYM** guarantees that the distributions of $(\hat{\mathbb{T}}^{2\ell}, \hat{\boldsymbol{\chi}}^{\ell})$ and $(\mathbb{T}^{2\ell}, \boldsymbol{\chi}^{2\ell})$ coincide. Indeed, it is immediate from the construction that $\hat{\mathbb{T}}^{2\ell}$ is distributed precisely as $\mathbb{T}^{2\ell}$. Furthermore, **SYM** ensures that the marginal distribution $\mu_{\mathbb{T}^{2\ell},r}$ is uniform on Ω. Hence, induction on ℓ shows that $\boldsymbol{\chi}^{2\ell}$ satisfies the recurrence (8.1) that gives rise to $\hat{\boldsymbol{\chi}}^{2\ell}$. (One could say that the trees $\mathbb{T}, \hat{\mathbb{T}}$ satisfy a Nishimori identity.)

To complete the proof we set up a coupling of $\hat{\mathbb{T}}^{2\ell}, \hat{\boldsymbol{\chi}}^{2\ell}$ and the depth-2ℓ neighbourhood of variable x_1 of G^*. Because $\boldsymbol{\sigma}^*$ is chosen uniformly at random, $\boldsymbol{\sigma}^*(x_1)$ is uniformly distributed, just as $\hat{\boldsymbol{\chi}}^0(r)$. Furthermore, a standard random hypergraph argument shows that the degree of each variable x_i in G^* is asymptotically Poisson with mean d. Therefore, **SYM** and (5.3) show that the constraints $\psi_{r,1}, \ldots, \psi_{r,D}$ pending on r are distributed just like the constraints pending on r in the construction of $\mathbb{T}^1$, up to an error of $o(1)$ in total variation distance. This error stems from the fact that the degree D of r is asymptotically but not precisely a Poisson variable, and that some constraint may contain r twice; the latter occurs with probability $O(1/n)$. Further, (5.3) also shows that the values under $\boldsymbol{\sigma}^*$ of the variables at distance two from r are asymptotically distributed according to (8.1) because $\boldsymbol{\sigma}^*$ is nearly balanced w.h.p. The coupling extends to the higher levels $\ell \geqslant 1$ of the tree by induction. □

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

Consider the planted model $\hat{G}$ and a sample $\boldsymbol{\sigma}$ from its Boltzmann distribution. For any finite value of ℓ there will likely be substantial dependences between the values $(\boldsymbol{\sigma}(y))_{y\in\partial^{2\ell}x_1}$ of the variables at distance precisely 2ℓ from some reference variable, say x_1. Indeed, these variables are 'close' to x_1 and therefore their values are going to be correlated with the value $\boldsymbol{\sigma}(x_1)$, and thus with each other. In other words, the sub-CSP $\nabla^{2\ell}(\hat{G}, x_1)$ induces dependences between the variables $\partial^{2\ell}x_1$. But are there additional correlations between these variables? To answer this question we introduce the following notation. For a set U of variable/constraints of a CSP instance G we let $\mu_{G\to U}$ be the Gibbs measure of the CSP from G by deleting all constraints in U. Thus, in particular, $\hat{G}\to\nabla^{2\ell}x_1$ is the sub-CSP obtained by deleting all constraints within a radius 2ℓ of x_1. The following proposition, which constitutes the main step toward the proof of Theorem 2.8, shows that once we delete these constraints, the correlations between the variables $\partial^{2\ell}x_1$ disappear.

Proposition 8.2 (**SYM**, **BAL**, **MIN**). *If $0<d<d_{\mathrm{cond}}$, then for every $\ell\geqslant 1$ we have*

$$\lim_{n\to\infty}\mathbb{E}\Bigg[\sum_{\sigma:\partial^{2\ell}x_1\to\Omega}|\mu_{\hat{G}\to\nabla^{2\ell}x_1}(\sigma)-q^{-|\partial^{2\ell}x_1|}|\Bigg]=0. \tag{8.2}$$

To prove Proposition 8.2 we need a few preparations.

Lemma 8.3 (**SYM**, **BAL**, **MIN**). *Suppose that $0<d<d_{\mathrm{cond}}$ and let $m\in\mathcal{M}(d)$. There exists a sequence $\omega=\omega(n)\to\infty$ such that with probability at least $1-1/\omega$ the random factor graph $G^*(n,m,\boldsymbol{\sigma}^*)$ has the following two properties:*

(i) *$\mu_{G^*(n,m,\boldsymbol{\sigma}^*)}$ is $(1/\omega,\omega)$-symmetric,*
(ii) *the marginals of $\mu_{G^*(n,m,\boldsymbol{\sigma}^*)}$ satisfy*

$$\sum_{i=1}^{n}\|\mu_{G^*,x_i}-\bar{\rho}\|_{\mathrm{TV}}<n/\omega.$$

Proof. This is immediate from Lemma 4.4, Lemma 5.1 and Proposition 5.7. □

Corollary 8.4 (**SYM**, **BAL**, **MIN**). *Suppose that $0<d<d_{\mathrm{cond}}$, let $C>0$ and let $m\in\mathcal{M}(d)$. There exists a sequence $\omega=\omega(n)\to\infty$ such that with probability at least $1-1/\omega$ for all $\sigma\in\Omega^{V_n}$ with $\|\rho_\sigma-\bar{\rho}\|_{\mathrm{TV}}\leqslant Cn^{-1/2}$ the following two statements hold:*

(i) *$\mu_{G^*(n,m,\sigma)}$ is $(1/\omega,\omega)$-symmetric,*
(ii) *the Gibbs marginals satisfy*

$$\sum_{i=1}^{n}\|\mu_{G^*(n,m,\sigma),x_i}-\bar{\rho}\|_{\mathrm{TV}}<n/\omega.$$

Proof. Suppose that $\sigma,\tau\in\Omega^{V_n}$ both satisfy $\|\rho_\sigma-\bar{\rho}\|_{\mathrm{TV}}\leqslant Cn^{-1/2}$ and let $m_\psi(\sigma),m_\psi(\tau)$ be the number of constraints endowed with the constraint function $\psi\in\Psi$ in $G^*(n,m,\sigma)$ and $G^*(n,m,\tau)$, respectively. Then the vectors $(m_\psi(\sigma))_{\psi\in\Psi}$, $(m_\psi(\tau))_{\psi\in\Psi}$ are multinomially distributed. Furthermore, because $\|\rho_\sigma-\rho_\tau\|_{\mathrm{TV}}\leqslant 2Cn^{-1/2}$, condition **SYM** and the characterization (5.3) of the planted model imply that $|\mathbb{E}[m_\psi(\sigma)]-\mathbb{E}[m_\psi(\tau)]|=O(1)$. Therefore, the local limit theorem for the multinomial distribution shows that $(m_\psi(\sigma))_{\psi\in\Psi}$, $(m_\psi(\tau))_{\psi\in\Psi}$ have total variation distance $o(1)$. Consequently, there is a coupling such that these vectors coincide with probability $1-o(1)$.

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

Now, since $m_\psi(\sigma) = m_\psi(\tau)$ for all ψ, we claim that the isomorphism classes of $G^*(n,m,\sigma)$, $G^*(n,m,\tau)$ are mutually contiguous. Specifically, permuting the assignment τ suitably, we may assume that the symmetric difference $\sigma\triangle\tau$ contains no more than $2C\sqrt{n}$ variables. Let $\mathcal{I}$ be the event that all the variables in $\sigma\triangle\tau$ are isolated. Given $m_\psi(\sigma) = m_\psi(\tau)$ for all ψ and $\mathcal{I}$, the factor graphs $G^*(n,m,\sigma)$, $G^*(n,m,\tau)$ are identically distributed (due to (5.3)). Hence, it suffices to prove that the isomorphism classes of $G^*(n,m,\sigma)$ and of the conditional CSP $G^*(n,m,\sigma)$ given $\mathcal{I}$ are mutually contiguous; of course the same construction will apply to $G^*(n,m,\tau)$.

To derive this contiguity result let $\mathcal{J}$ be the event that $G^*(n,m,\sigma)$ has at least $n^{2/3}$ isolated variables in each of the sets $\sigma^{-1}(\chi)$, $\chi\in\Omega$. Because the constraints of $G^*(n,m,\sigma)$ are chosen independently and σ is nearly balanced, standard arguments show that $\mathcal{J}$ occurs with (very) high probability. Hence, let G' denote the random CSP $G^*(n,m,\sigma)$ given $\mathcal{J}$. Then we construct a random factor graph $G''\in\mathcal{I}\cap\mathcal{J}$ as follows: choose a one-to-one map ι from the set $\sigma\triangle\tau$ to the set of isolated variables of G' such that $\sigma(\iota(x)) = \sigma(x)$ for all x uniformly at random. Then obtain G'' from G' by swapping the variables x and $\iota(x)$ for all $x\in\sigma\triangle\tau$. Clearly, G' and G'' are isomorphic. Moreover, with I the number of isolated variables, we see that for every possible outcome G,

$$\mathbb{P}[G'' = G \mid I(G'') = I(G)] = \mathbb{P}[G^*(n,m,\sigma) = G \mid I(G^*(n,m,\sigma)) = I(G),\ \mathcal{I}].$$

Finally, $I(G'')$ and $I(G^*(n,m,\sigma))$ given $\mathcal{I}$ are mutually contiguous; for both satisfy a local limit theorem with standard deviation $\Theta(\sqrt{n})$, and their means differ by no more than $O(\sqrt{n})$. Since $\mathcal{J}$ occurs with high probability, we obtain the desired contiguity of the isomorphism classes of $G^*(n,m,\sigma)$ and $G^*(n,m,\sigma)$ given $\mathcal{I}$.

In summary, for any σ,τ with $\|\rho_\sigma - \bar\rho\|_{\mathrm{TV}}, \|\rho_\tau - \bar\rho\|_{\mathrm{TV}} \leqslant Cn^{-1/2}$, we can couple the $m_\psi(\sigma)$, $m_\psi(\tau)$ such that the isomorphism classes of $G^*(n,m,\sigma)$, $G^*(n,m,\tau)$ are mutually contiguous w.h.p. To complete the proof, we simply observe that the event $\|\rho_{\boldsymbol{\sigma}^*} - \bar\rho\|_{\mathrm{TV}} \leqslant Cn^{-1/2}$ occurs with a probability that is bounded away from zero by the central limit theorem. Therefore, the assertion follows from Lemma 8.3. □

Proof of Proposition 8.2. By Lemma 5.1 and Proposition 7.8 it suffices to prove (8.2) with $\hat G$ replaced by $G^* = G^*(n,m,\boldsymbol{\sigma}^*)$. Indeed, fix a large number $C>0$ and let $\mathcal{A}$ be the event that $\|\rho_{\boldsymbol{\sigma}^*} - \bar\rho\|_{\mathrm{TV}} \leqslant Cn^{-1/2}$. Then the probability of the event $\mathcal{A}$ is bounded away from 0 and, in fact, approaches 1 in the limit of large C. Further, let G' be the factor graph obtained from G^* by deleting all variables and constraints at a distance less than 2ℓ from x_1. Let n', m' be the number of variables and constraints of G' and let $\boldsymbol{\sigma}'$ be the assignment induced by $\boldsymbol{\sigma}^*$ on the set of variables of G'. Since ℓ is a fixed number, Lemma 8.1 implies that

$$\mathbb{P}[n+m-n'-m' \leqslant \ln n \mid \mathcal{A}] = 1 - o(1). \tag{8.3}$$

In particular, if $n - n' \leqslant \ln n$ and if $\mathcal{A}$ occurs, then $\|\rho_{\boldsymbol{\sigma}'} - \bar\rho\|_{\mathrm{TV}} \leqslant Cn^{-1/2}$ Moreover, since m is a Poisson variable with mean dn/k, (8.3) implies that

$$\mathbb{P}[m' - dn'/k \leqslant n^{3/5} \mid \mathcal{A}] = 1 - o(1).$$

Hence, recalling the definition of the set $\mathcal{M}(d)$, we see that on $\mathcal{A}$ we can apply Corollary 8.4 to G' with high probability. Consequently, with high probability on $\mathcal{A}$ the Gibbs measure

$$\mu_{G'} \text{ is } (1/\omega,\omega)\text{-symmetric and } \sum_{i=1}^{n} \|\mu_{G',x_i} - \bar\rho\|_{\mathrm{TV}} < n/\omega \tag{8.4}$$

for some $\omega\to\infty$.

To complete the proof let $\iota : \partial^{2\ell}_{G^*}x_1 \to V_n$ be a uniformly random map such that $\boldsymbol{\sigma}^*(\iota(x)) = \boldsymbol{\sigma}^*(x)$ for all x. Moreover, let G'' be the random factor graph obtained from G^* by connecting the constraints at distance $2\ell-1$ from x_1 with the images $\iota(x)$ instead of their original neighbours

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

$x \in \partial^{2\ell}_{G^*} x_1$. Then the distribution of G'' is identical to the distribution of G^*. Furthermore, since Lemma 8.1 implies that $|\partial^{2\ell}_{G^*} x_1| \leqslant \omega^{1/2}$ with high probability, (8.4) yields

$$\mathbb{E}[\|\mu_{G',\partial^{2\ell}_{G''} x_1} - \bar{\rho}\|_{\mathrm{TV}} \mid \mathcal{A}] = o(1).$$

Thus, on $\mathcal{A}$ with high probability the boundary condition $\mu_{G',\partial^{2\ell}_{G''} x_1}$ of the depth-2ℓ neighbourhood of x_1 is close in total variation distance to the free boundary condition, and therefore

$$\mathbb{E}[\|\mu_{G^*,\partial^{2\ell} x_1} - \mu_{\partial^{2\ell} x_1}\|_{\mathrm{TV}} \mid \mathcal{A}] = o(1). \tag{8.5}$$

Finally, since the probability of $\mathcal{A}$ converges to 1 as $C \to \infty$, the assertion follows from (8.5). □

Proof of Theorem 2.9. The theorem is immediate from Lemma 8.1 and Proposition 8.2. □

8.2 Proof of Theorem 2.10

Theorem 2.10 is almost an immediate consequence of Theorem 2.9, except that a bit of care is required to prove that $d^\star_{\mathrm{rec}} \leqslant d_{\mathrm{cond}}$. To this end, we first show that $\mu_{\hat{G}}$ is ε-symmetric with mostly uniform marginals for $d < d^\star_{\mathrm{rec}}$.

Lemma 8.5 (**SYM**, **BAL**). *Assume that $d < d^\star_{\mathrm{rec}}$ and let $\varepsilon > 0$. Then with probability at least $1 - \varepsilon + o(1)$ the Boltzmann distribution $\mu_{\hat{G}}$ is ε-symmetric and its marginals satisfy*

$$\sum_{i=1}^{n} \|\mu_{\hat{G},x_i} - \bar{\rho}\|_{\mathrm{TV}} < \varepsilon n.$$

Proof. Fix a small enough $\delta > 0$. If $d < d^\star_{\mathrm{rec}}$, then there exists a bounded $\ell = \ell(\delta) > 0$ such that

$$\mathbb{P}[\exists \omega \in \Omega : |\langle \mathbf{1}\{\boldsymbol{\sigma}(r) = \omega\} \mid \forall x \in \partial_{2\ell}(\mathbb{T}(d,P), r) : \boldsymbol{\sigma}(x) = \boldsymbol{\chi}^{2\ell}(x)\rangle - 1/q|_{\mathbb{T}^{2\ell}(d,P)} > \delta] \leqslant \delta.$$

Hence, Lemmas 7.8 and 8.1 show that for large enough n,

$$\mathbb{P}[\exists \omega \in \Omega : |\langle \mathbf{1}\{\boldsymbol{\sigma}(x_1) = \omega\} \mid \forall y \in \partial_{2\ell}(G^*(n,m,\hat{\boldsymbol{\sigma}}), x_1) : \boldsymbol{\sigma}(y) = \hat{\boldsymbol{\sigma}}(y)\rangle - 1/q|_{G^*(n,m,\hat{\boldsymbol{\sigma}})} > \delta] < \varepsilon/2.$$

Therefore, the Nishimori identity (5.6) yields

$$\mathbb{P}\big[\exists \omega \in \Omega : \big|\big\langle \mathbf{1}\{\boldsymbol{\sigma}(x_1) = \omega\} \mid \overline{\nabla_{2\ell}(\hat{G}, x_1)}\big\rangle_{\hat{G}} - 1/q\big|\big\rangle_{\hat{G}} > \delta\big] < \varepsilon/2. \tag{8.6}$$

Now let $\mathcal{E}$ be the event that x_1, x_2 have distance at least $2\ell + 2$ in $\hat{G}$ and that

$$\forall \omega \in \Omega : \big|\big\langle \mathbf{1}\{\boldsymbol{\sigma}(x_1) = \omega\} \mid \overline{\nabla_{2\ell}(\hat{G}, x_1)}\big\rangle_{\hat{G}} - 1/q\big|\big\rangle_{\hat{G}} \leqslant \delta. \tag{8.7}$$

Since the average degree of $\hat{G}$ is bounded w.h.p., (8.6) shows immediately that $\mathbb{P}[\mathcal{E}] \geqslant 1 - 2\varepsilon/3 + o(1)$.

But given $\mathcal{E}$ it is immediate that $\|\mu_{\hat{G},x_1,x_2} - \bar{\rho}\|_{\mathrm{TV}} < \varepsilon/4$, provided that δ is small enough, an observation that goes back to [66]. Indeed, since x_1, x_2 have distance greater than 2ℓ, conditioning on the values of *all* variables at distance 2ℓ from x_1 is stronger than just conditioning on the value of x_2. Thus, we conclude that

$$\mathbb{P}[\|\mu_{\hat{G},x_1,x_2} - \bar{\rho}\|_{\mathrm{TV}} < \varepsilon/4] \geqslant 1 - 2\varepsilon/3 + o(1). \tag{8.8}$$

Finally, because the distribution of $\hat{G}$ is invariant under permutations of the variables, (8.8) yields the assertion. □

Corollary 8.6 (**SYM**, **BAL**, **MIN**, **POS**). *We have $d^\star_{\mathrm{rec}} \leqslant d_{\mathrm{cond}}$.*

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

Proof. Let $0 < D < d^\star_{\mathrm{rec}}$ and let $x_1, \ldots, x_k$ denote uniformly and independently chosen variables. Due to (7.1) and Proposition 6.7 we have, uniformly for all $d \leqslant D$,

$$\frac{k\xi}{n}\frac{\partial}{\partial d}\mathbb{E}[\ln Z(\hat{G})] = f_n(d) + o(1) \quad \text{with } f_n(d) = \mathbb{E}[\Lambda(\langle \boldsymbol{\psi}(\boldsymbol{\sigma}(x_1), \ldots, \boldsymbol{\sigma}(x_k))\rangle_{\hat{G}})]. \tag{8.9}$$

We claim that, for every $d \leqslant D$,

$$\lim_{n\to\infty} f_n(d) = \Lambda(\xi). \tag{8.10}$$

Indeed, plugging in the expansion $\Lambda(1-x) = -x + \sum_{\ell\geqslant 2} x^\ell/(\ell(\ell-1))$, valid for all $x \in [-1,1]$, we obtain

$$\begin{aligned} f_n(d) &= 1 - \mathbb{E}[\langle \boldsymbol{\psi}(\boldsymbol{\sigma}(x_1), \ldots, \boldsymbol{\sigma}(x_k))\rangle_{\hat{G}}] + \sum_{\ell\geqslant 2}\frac{1}{\ell(\ell-1)}\mathbb{E}[\langle 1 - \boldsymbol{\psi}(\boldsymbol{\sigma}(x_1), \ldots, \boldsymbol{\sigma}(x_k))\rangle_{\hat{G}}^\ell] \\ &= 1 - \xi + o(1) + \sum_{\ell\geqslant 2}\frac{\mathbb{E}\langle\prod_{j=1}^\ell 1 - \boldsymbol{\psi}(\boldsymbol{\sigma}_j(x_1), \ldots, \boldsymbol{\sigma}_j(x_k))\rangle_{\hat{G}}}{\ell(\ell-1)} \qquad (8.11) \\ &\qquad \text{(by Lemma 5.1, Corollary 6.5 and \textbf{SYM}).} \end{aligned}$$

Further, by Lemma 8.5 there is a function $\varepsilon_n(d) = o(1)$ such that $\mu_{\hat{G}}$ is $\varepsilon_n(d)$-symmetric with probability at least $1 - \varepsilon_n(d)$. Therefore, Lemma 4.3 implies that the ℓ-fold product measure $\mu_{\hat{G}}^{\otimes\ell}$ is $o(1)$-symmetric w.h.p. for any fixed $\ell > 0$. Hence, every $\ell \geqslant 2$ we have w.h.p.

$$\begin{aligned} &\left\langle \prod_{j=1}^\ell 1 - \boldsymbol{\psi}(\boldsymbol{\sigma}_j(x_1), \ldots, \boldsymbol{\sigma}_j(x_k))\right\rangle_{\hat{G}} \\ &\quad = o(1) + \frac{1}{n^k}\sum_{\psi\in\Psi}\sum_{i_1,\ldots,i_k=1}^n\sum_{\sigma_1,\ldots,\sigma_\ell\in\Omega^k} P(\psi)\left(\prod_{j=1}^\ell 1 - \psi(\sigma_j)\right)\prod_{h=1}^k\prod_{j=1}^\ell \mu_{\hat{G},x_h}(\sigma_{j,h}). \end{aligned}$$

Thus, invoking the asymptotic uniformity of the Boltzmann marginals supplied by Lemma 8.5 and applying **SYM**, we see that w.h.p.

$$\left\langle \prod_{j=1}^\ell 1 - \boldsymbol{\psi}(\boldsymbol{\sigma}_j(x_1), \ldots, \boldsymbol{\sigma}_j(x_k))\right\rangle_{\hat{G}} = (1-\xi)^\ell + o(1). \tag{8.12}$$

Combining (8.11) and (8.12), we obtain (8.10).

Finally, (8.9), (8.10) and dominated convergence yield

$$\begin{aligned} \lim_{n\to\infty}\frac{1}{n}\mathbb{E}[\ln Z(\hat{G}(n, m_D))] &= \ln q + \lim_{n\to\infty}\frac{1}{n}\int_0^d \frac{\partial}{\partial D}\mathbb{E}[\ln Z(\hat{G})]\mathrm{d}d \\ &= \ln q + \frac{1}{k\xi}\lim_{n\to\infty}\int_0^D f_n(d)\mathrm{d}d = \ln q + D\ln(\xi)/k. \end{aligned}$$

Hence, Theorem 5.3 shows that $d_{\mathrm{cond}} \geqslant D$. Since this holds for any $D < d_{\mathrm{rec}}$, the assertion follows. □

Proof of Theorem 2.10. Corollary 8.6 shows that $d^\star_{\mathrm{rec}} \leqslant d_{\mathrm{cond}}$. Thus, we are left to show that $d^\star_{\mathrm{rec}} = d_{\mathrm{rec}}$. To prove that $d^\star_{\mathrm{rec}} \leqslant d_{\mathrm{rec}}$ suppose that $d < d^\star_{\mathrm{rec}}$. Then (2.17) ensures that for any $\varepsilon > 0$ there is ℓ such that w.h.p. we have

$$\mathbb{E}\Big(\Big|\langle \mathbf{1}\{\boldsymbol{\sigma}(r) = \omega\} \mid \overline{\nabla_{2\ell}(\mathbb{T}^{2\ell}(d,P), r)}\rangle_{\mathbb{T}^{2\ell}(d,P)} - 1/q\Big|\Big)_{\mathbb{T}^{2\ell}(d,P)} < \varepsilon. \tag{8.13}$$

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

Further, Theorem 2.9 shows that $\|\mu_{\mathbb{G},\nabla_{2\ell}(\mathbb{G},x_1)},\mu_{\mathbb{G},\nabla_{2\ell}(\mathbb{G},x_1)}\|_{\mathrm{TV}}=o(1)$ w.h.p. Moreover, by Theorem 2.11 and Lemma 8.1 the distribution of the neighbourhood $\nabla_{2\ell}(\mathbb{G},x_1)$ is at total variation distance $o(1)$ of the distribution of the random tree $\mathbb{T}^{\ell}(d,P)$. Therefore, (8.13) shows that $\mathrm{corr}(d)\leqslant\varepsilon$. Since this is true for any $\varepsilon>0$, we conclude that $d\leqslant d_{\mathrm{rec}}$.

Conversely, assume that $d<d_{\mathrm{rec}}\leqslant d_{\mathrm{cond}}$. Then we can just put the argument from the previous paragraph in reverse. Indeed, Theorem 2.11 and Lemma 8.1 show that the neighbourhood $\nabla_{2\ell}(\mathbb{G},x_1)$ is the distribution as $\mathbb{T}^{2\ell}(d,P)$, up to $o(1)$ in total variation. Further, for any $\varepsilon>0$ there exists ℓ such that

$$\sum_{\omega\in\Omega}\mathbb{E}\Big|\big\langle\mathbf{1}\{\boldsymbol{\sigma}(x_1)=\omega\}\mid\overline{\nabla_{2\ell}(\mathbb{G},x_1)}\big\rangle_{\mathbb{G}}-1/q\Big|_{\mathbb{G}}<\varepsilon+o(1).$$

because $\mathrm{corr}(d)=0$. Hence, Theorem 2.9 yields $\mathrm{corr}^{\star}(d)\leqslant\varepsilon$. Finally, because this bound holds for any $\varepsilon>0$ we obtain $\mathrm{corr}^{\star}(d)=0$. □

Appendix A: Proof of Lemma 4.4

To establish Lemma 4.4 we will utilize regularity results for discrete probability measures from [17]. For $\varepsilon>0$ choose $\eta=\eta(\varepsilon)>0$ and $n>1/\eta$ sufficiently large. By [17, Corollary 2.2], for any $\mu\in\mathcal{P}(\Omega^n)$, there exist $L\in\mathbb{N}$, $\mu^{(0)},\ldots,\mu^{(L)}\in\mathcal{P}(\Omega^n)$ and $w_0,w_1,\ldots,w_L$ such that we can decompose $\mu=\sum_{i=0}^{L}w_i\mu^{(i)}$ and

(i) $\mu^{(1)},\ldots,\mu^{(L)}$ are η-symmetric,
(ii) $w_0,\ldots,w_L\geqslant0$, $\sum_{i=0}^{L}w_i=1$, $\sum_{i=1}^{L}w_i\geqslant1-\eta$ and
(iii) $w_i\geqslant\eta/L$ for all $i\in[L]$.

Let us use the shorthand notation $\langle\cdot\rangle_i=\langle\cdot\rangle_{\mu^{(i)}}$ and note that $\|\cdot\|_{\mathrm{TV}}$ and $\|\cdot\|_2$ are equivalent norms in $\mathbb{R}^{q^2}$. For $i\in[L]$, we have

$$\begin{aligned}\langle\|\rho_{\boldsymbol{\sigma},\boldsymbol{\tau}}-\bar{\rho}\|_2^2\rangle_i&=\sum_{s,t\in\Omega}\frac{1}{n^2}\sum_{v,w\in[n]}\mu^{(i)}_{v,w}(s,s)\mu^{(i)}_{v,w}(t,t)-q^{-2}\\&=\sum_{s,t\in\Omega}\frac{1}{n^2}\Bigg[\sum_{v,w\in[n]}\mu^{(i)}_{v,w}(s,s)\mu^{(i)}_{v,w}(t,t)-\Bigg(\sum_{v\in[n]}\mu^{(i)}_{v}(s)\mu^{(i)}_{v}(t)\Bigg)^2\Bigg]\\&\quad+\Bigg[\sum_{s,t\in\Omega}\Bigg(\frac{1}{n}\sum_{v\in[n]}\mu^{(i)}_{v}(s)\mu^{(i)}_{v}(t)\Bigg)^2-q^{-2}\Bigg].\end{aligned}\tag{A.1}$$

Combining (i) and Lemma 4.3 yields that $\mu^{(i)}\otimes\mu^{(i)}$ is ζ-symmetric for a suitable $\zeta=\zeta(\eta)>0$. Thus, the first summand of (A.1) is $O(\zeta)$. Now assume that $\langle\|\rho_{\boldsymbol{\sigma},\boldsymbol{\tau}}-\bar{\rho}\|_{\mathrm{TV}}\rangle_\mu<\delta$ for $\delta(\eta,\zeta)>0$ sufficiently small. Due to (iii) and Jensen's inequality, $\langle\|\rho_{\boldsymbol{\sigma},\boldsymbol{\tau}}-\bar{\rho}\|_2^2\rangle_i<\sqrt{\delta/\eta}$ and consequently, (A.1) implies that for all $s,t\in\Omega$ we have

$$\Bigg|\frac{1}{n}\sum_{v\in[n]}\mu^{(i)}_{v}(s)\mu^{(i)}_{v}(t)-q^{-2}\Bigg|\leqslant O(\zeta^{1/2}).$$

Hence for all $s\in\Omega$ we have

$$\Bigg|\frac{1}{n}\sum_{v\in[n]}(\mu^{(i)}_{v}(s))^2-q^{-2}\Bigg|\leqslant O(\zeta^{1/2}),\quad\Bigg|\frac{1}{n}\sum_{v\in[n]}\mu^{(i)}_{v}(s)-q^{-1}\Bigg|\leqslant O(\zeta^{1/2}).\tag{A.2}$$

As the sum of squares is minimized by a uniform distribution, Taylor-expanding the function

$$f((\mu^{(i)}(s))_{s\in\Omega})=\frac{1}{n}\Bigg(\sum_{v\in[n]}(\mu^{(i)}(s))^2\Bigg)_{s\in\Omega}$$

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

around $q^{-1}\mathbf{1}_{q\times n}$ together with (A.2) yields

$$\left|\frac{1}{n}O(\|\mu^{(i)} - q^{-1}\mathbf{1}_{q\times n}\|_2^2)\right| \leqslant O(\zeta^{1/2}).$$

Thus, for all $i \in [L]$ we have

$$\frac{1}{n}\sum_{v\in[n]} \|\mu_v^{(i)}(\,\cdot\,) - q^{-1}\mathbf{1}\|_{\mathrm{TV}} < \zeta^{1/5}. \tag{A.3}$$

The ε-symmetry of μ now follows from (A.3) and [17, Lemma 2.8]. Moreover, equation (A.3) and (ii) imply

$$\frac{1}{n}\sum_{v\in[n]} \|\mu_v(\,\cdot\,) - q^{-1}\mathbf{1}\|_{\mathrm{TV}} < \varepsilon.$$

We now turn to the converse implication. First, a calculation as in (A.1) yields that

$$\langle\|\rho_{\sigma,\tau} - \bar{\rho}\|_2\rangle_\mu^2 \leqslant \langle\|\rho_{\sigma,\tau} - \bar{\rho}\|_2^2\rangle_\mu \leqslant \left(1 + \frac{1}{q}\right)\frac{2}{n^2}\sum_{v,w\in[n]} \|\mu_{v,w} - \bar{\rho}\|_{\mathrm{TV}}. \tag{A.4}$$

Secondly, we bound

$$\begin{aligned}
&\frac{1}{n^2}\sum_{v,w\in[n]} \|\mu_v \otimes \mu_w - \bar{\rho}\|_{\mathrm{TV}} \\
&\quad\leqslant \frac{1}{2n^2}\sum_{v,w\in[n]}\sum_{s,t\in\Omega}\left(\left|\mu_v(s) - \frac{1}{q}\right|\left|\mu_w(t) - \frac{1}{q}\right| + \frac{1}{q}\left(\left|\mu_v(s) - \frac{1}{q}\right| + \left|\mu_w(t) - \frac{1}{q}\right|\right)\right) \\
&\quad= 2\left(\frac{1}{n}\sum_{v\in[n]} \|\mu_v - \bar{\rho}\|_{\mathrm{TV}}\right)^2 + \frac{2}{n}\sum_{v\in[n]} \|\mu_v - \bar{\rho}\|_{\mathrm{TV}}.
\end{aligned} \tag{A.5}$$

Now, inequalities (A.4), (A.5) and the triangle inequality imply that by choosing $\delta > 0$ small enough, we have that any δ-symmetric μ with $1/n\sum_{v\in[n]} \|\mu_v - \bar{\rho}\|_{\mathrm{TV}} < \delta$ satisfies

$$\langle\|\rho_{\sigma,\tau} - \bar{\rho}\|_{\mathrm{TV}}\rangle_\mu < \varepsilon.$$

Appendix B: Moment calculations

The proofs of Propositions 5.9 and 5.10 are straightforward applications of the Laplace method; the calculations are identical to those performed in [25, Section 7].

Proof of Proposition 5.9. Let R_n be the set of all distributions $\rho \in \mathcal{P}(\Omega)$ such that the vector $n\rho \in \mathbb{R}^\Omega$ has integer entries. For $\rho \in R_n$ let $Z_\rho(G(n,m)) = Z(G(n,m))\langle\mathbf{1}\{\rho_\sigma = \rho\}\rangle_{G(n,m)}$ be the number of satisfying assignments $\sigma \in \mathcal{S}(G(n,m))$ with empirical distribution $\rho_\sigma = \rho$, so that

$$\mathbb{E}[Z(G(n,m))] = \sum_{\rho\in R_n} \mathbb{E}[Z_\rho(G(n,m))]. \tag{B.1}$$

Since the total number of assignments $\sigma \in \Omega^{V_n}$ with empirical distribution ρ is given by the multinomial coefficient $\binom{n}{n\rho}$ and because the m constraints of $G(n,m)$ are chosen independently, we can express the mean $\mathbb{E}[Z_\rho(G(n,m))]$ easily in terms of the function ϕ from condition **BAL**. Namely,

$$\mathbb{E}[Z_\rho(G(n,m))] = \binom{n}{\rho n}\phi(\rho)^m. \tag{B.2}$$

Further, because by **BAL** both the multinomial coefficient and the function $\phi(\rho)$ take their maximum at the uniform distribution $\bar{\rho}$, the contribution of the summands from the set $R_n' = \{\rho \in R_n : \|$

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

$\rho - \bar\rho\|_2 < n^{-1/2}\ln n\}$ dominates. Thus, approximating the multinomial coefficient in (B.2) via Stirling's formula, we obtain from (B.1)

$$\mathbb{E}[Z(G(n,m))] \sim \sum_{\rho\in R_n'} \mathbb{E}[Z_\rho(G(n,m))] \sim \sum_{\rho\in R_n'} \frac{\exp(nf_n(\rho))}{\sqrt{(2\pi n)^{q-1}\prod_{\omega\in\Omega}\rho(\omega)}}, \tag{B.3}$$

where

$$f_n(\rho) = \mathcal{H}(\rho) + \frac{m}{n}\ln\phi(\rho).$$

The gradient and the Hessian of the function $f_n(\rho)$ at $\rho = \bar\rho$ are computed easily. Indeed, using **SYM** and Lemma 4.1 we obtain

$$Df_n(\bar\rho) = (\ln(q) - 1 + km/n)\mathbf{1}, \quad D^2f_n(\bar\rho) = -q(\mathrm{id} - (k(k-1)m/n)\Phi) + (k^2m/n)\mathbf{1}, \tag{B.4}$$

and all third partial derivatives of f_n are uniformly bounded on R_n'. Furthermore, for all $\rho\in R_n'$ we have $\mathbf{1}\perp\rho-\bar\rho$ because $\bar\rho,\rho'$ are probability distributions on Ω. Hence,

$$f_n(\rho) = f_n(\bar\rho) - q\langle(\mathrm{id} - (k(k-1)m/n)\Phi)(\rho-\bar\rho),(\rho-\bar\rho)\rangle + O(n^{-3/2}\ln^3 n)$$

uniformly for all $\rho\in R_n'$. Thus, (B.3) boils down to

$$\begin{aligned}
&\mathbb{E}[Z(G(n,m))]\\
&\quad\sim \frac{q^{q/2}\exp(nf(\bar\rho))}{(2\pi n)^{(q-1)/2}}\sum_{\rho\in R_n'}\exp[-qn\langle(\mathrm{id}-(k(k-1)m/n)\Phi)(\rho-\bar\rho),(\rho-\bar\rho)\rangle]\\
&\quad= \frac{q^{n+1/2}\xi^m}{(2\pi n/q)^{(q-1)/2}}\sum_{\rho\in R_n'}\exp[-qn\langle(\mathrm{id}-(k(k-1)m/n)\Phi)(\rho-\bar\rho),(\rho-\bar\rho)\rangle] \quad (\text{by } \mathbf{SYM}).
\end{aligned} \tag{B.5}$$

By Lemma 5.4 the matrix Φ has precisely one positive eigenvalue, namely 1, with corresponding eigenvector $\mathbf{1}$. Since in (B.5) we sum only over ρ such that $\rho-\bar\rho\perp\mathbf{1}$, we can approximate the sum by a Gaussian integral over the $(q-1)$-dimensional orthogonal complement of $\mathbf{1}$ in $\mathbb{R}^q$ to obtain

$$\begin{aligned}
&\sum_{\rho\in R_n'}\exp[-qn\langle(\mathrm{id}-(k(k-1)m/n)\Phi)(\rho-\bar\rho),(\rho-\bar\rho)\rangle]\\
&\quad\sim \int_{\mathbb{R}^{q-1}}\exp\Bigg(-qn\sum_{\lambda\in\mathrm{Eig}[\Phi]\setminus\{1\}}(1-(k(k-1)m/n)\lambda)z_i^2\Bigg)\mathrm{d}z\\
&\quad\sim \frac{(2\pi n/q)^{(q-1)/2}}{\prod_{\lambda\in\mathrm{Eig}[\Phi]\setminus\{1\}}\sqrt{1-d(k-1)\lambda}}.
\end{aligned} \tag{B.6}$$

Combining (B.5) and (B.6) completes the proof. □

Proof of Proposition 5.10. Let R_n be the set of all distributions $\rho\in\mathcal{P}(\Omega\times\Omega)$ such that $n\rho\in\mathbb{R}^{\Omega\times\Omega}$ is integral and such that $\|\rho-\bar\rho\|_{\mathrm{TV}}\leqslant\zeta$. Let $\mathcal{Z}_\rho(G(n,m))$ be the number of pairs $(\sigma_1,\sigma_2)\in\mathcal{S}(G(n,m))$ with overlap $\rho_{\sigma_1,\sigma_2}=\rho$. Recalling the definition of $\mathcal{Z}$ from (5.18), we get

$$\begin{aligned}
\mathbb{E}[\mathcal{Z}(G(n,m))^2] &= \mathbb{E}[Z(G(n,m))^2\mathbf{1}\{\langle\|\rho_{\sigma_1,\sigma_2}-\bar\rho\|_{\mathrm{TV}}\rangle_{G(n,m)}\leqslant\zeta\}]\\
&= \sum_{\rho\in R_n}\mathbb{E}[\mathcal{Z}_\rho(G(n,m))].
\end{aligned} \tag{B.7}$$

Clearly, the total number of pairs $(\sigma_1,\sigma_2)\in\Omega^{V_n}$ with overlap ρ equals $\binom{n}{\rho n}$. Hence, recalling the function φ from condition **MIN**, using the independence of the constraints of $G(n,m)$ and applying Stirling's formula, we obtain

$$\mathbb{E}[\mathcal{Z}_\rho(G(n,m))] = \binom{n}{\rho n}\varphi(\rho)^m \sim \sum_{\rho\in R_n}\frac{\exp(nf_n(\rho))}{\sqrt{(2\pi n)^{q^2-1}\prod_{\omega,\omega'\in\Omega}\rho(\omega,\omega')}}, \tag{B.8}$$

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

where

$$f_n(\rho) = \mathcal{H}(\rho) + \frac{m}{n}\ln\varphi(\rho).$$

Once more it is straightforward to calculate the gradient and the Hessian of f_n at the point $\bar{\rho}$: condition **SYM** yields

$$Df_n(\bar{\rho}) = (2\ln(q) - 1 + km/n)\mathbf{1}, \quad D^2 f_n(\bar{\rho}) = -q^2(\mathrm{id} - (k(k-1)m/n)\Xi) + (k^2 m/n)\mathbf{1}, \tag{B.9}$$

and all third partial derivatives are uniformly bounded. Consequently, since $\mathbf{1} \perp \rho - \bar{\rho}$ for all $\rho \in R_n$, we obtain

$$f_n(\rho) = f_n(\bar{\rho}) - q^2\langle(\mathrm{id} - (k(k-1)m/n)\Xi)(\rho - \bar{\rho}), (\rho - \bar{\rho})\rangle + O(\|\rho - \bar{\rho}\|_{\mathrm{TV}}^3)$$

uniformly for all $\rho \in R_n$. Hence, (B.8) becomes

$$\begin{aligned}&\mathbb{E}[\mathcal{Z}(G(n,m))^2]\\ &\sim \frac{q^{2n+1}\xi^m}{(2\pi n/q^2)^{(q^2-1)/2}}\sum_{\rho\in R_n'}\exp\left[-q^2 n\langle(\mathrm{id} - (k(k-1)m/n)\Xi)(\rho - \bar{\rho}), (\rho - \bar{\rho})\rangle\right].\end{aligned} \tag{B.10}$$

Since $d < d_{\mathrm{cond}}$, Lemma 5.4 and Proposition 5.5 show that $\mathbf{1}$ is the only eigenvector of $\mathrm{id} - (k(k-1)m/n)\Xi$ with a non-negative eigenvalue. Consequently, because the sum only ranges over ρ such that $\mathbf{1} \perp \rho - \bar{\rho}$, we can approximate the sum by a Gaussian integral:

$$\begin{aligned}&\sum_{\rho\in R_n'}\exp\left[-q^2 n\langle(\mathrm{id} - (k(k-1)m/n)\Xi)(\rho - \bar{\rho}), (\rho - \bar{\rho})\rangle\right]\\ &\quad\sim \int_{\mathbb{R}^{q^2-1}}\exp\left(-q^2 n\sum_{\lambda\in\mathrm{Eig}'[\Xi]}(1 - (k(k-1)m/n)\lambda)z_i^2\right)\mathrm{d}z\\ &\quad\sim \frac{(2\pi n/q^2)^{(q^2-1)/2}}{\prod_{\lambda\in\mathrm{Eig}'[\Xi]}\sqrt{1 - d(k-1)\lambda}}.\end{aligned} \tag{B.11}$$

Thus, the assertion follows from (B.10) and (B.11). □

Appendix C: Proof of Lemma 7.2

In order to prove Lemma 7.2, we first establish a uniform upper bound on the total variation distance of $\hat{\boldsymbol{\sigma}}_{n,m}$ and $\hat{\boldsymbol{\sigma}}_{n,m'}$ for m and m' that are not too far from dn/k.

Lemma C.1 (**SYM**, **BAL**). For any $\eta > 0$, $d > 0$ there is $\delta > 0$ such that

$$\limsup_{n\to\infty}\max\{d_{\mathrm{TV}}\{\hat{\boldsymbol{\sigma}}_{n,m}, \hat{\boldsymbol{\sigma}}_{n,m'}\} : |m - dn/k| + |m' - dn/k| < \delta n\} < \eta. \tag{C.1}$$

Proof. Fix $\eta > 0$, $d > 0$ and recall the function ϕ from condition **BAL**. Lemma E.1 shows that there exists $c > 0$ such that, for all $0 < \delta < 1$ and all $m, m' \leqslant (d/k + \delta)n$, the bounds

$$c\left(\frac{\phi(\rho_\sigma)}{\xi}\right)^{m-m'} \leqslant \frac{\mathbb{P}[\hat{\boldsymbol{\sigma}}_{n,m} = \sigma]}{\mathbb{P}[\hat{\boldsymbol{\sigma}}_{n,m'} = \sigma]} \leqslant \frac{1}{c}\left(\frac{\phi(\rho_\sigma)}{\xi}\right)^{m-m'} \tag{C.2}$$

are valid. Moreover, Corollary 6.5 yields $C > 0$ such that for all $m, m' \leqslant (d/k + \delta)n$ we have

$$\mathbb{P}[\|\rho_{\hat{\boldsymbol{\sigma}}_{n,m}} - \bar{\rho}\|_{\mathrm{TV}} > C/\sqrt{n}] + \mathbb{P}[\|\rho_{\hat{\boldsymbol{\sigma}}_{n,m'}} - \bar{\rho}\|_{\mathrm{TV}} > C/\sqrt{n}] \leqslant \eta/4. \tag{C.3}$$

Further, suppose that $\sigma \in \Omega^{V_n}$ satisfies $\|\rho_\sigma - \bar{\rho}\|_{\mathrm{TV}} \leqslant C/\sqrt{n}$. Because **BAL** ensures that the first derivative of ϕ vanishes at $\bar{\rho}$, we have

$$\phi(\rho_\sigma) = \phi(\bar{\rho}) + O(\|\rho_\sigma - \bar{\rho}\|_{\mathrm{TV}}^2) = \xi + O(n^{-1}). \tag{C.4}$$

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

Combining (C.2) and (C.4), we obtain $c_1, c_2 > 0$ such that for all σ satisfying $\|\rho_\sigma - \bar{\rho}\|_{\mathrm{TV}} \leqslant C/\sqrt{n}$, for all $0 < \delta < 1$ and for all m, m' satisfying $|m - dn/k| + |m' - dn/k| < \delta n$ the estimates

$$c_1 \exp(-\delta c_2) \leqslant \frac{\mathbb{P}[\hat{\boldsymbol{\sigma}}_{n,m} = \sigma]}{\mathbb{P}[\hat{\boldsymbol{\sigma}}_{n,m'} = \sigma]} \leqslant \exp(\delta c_2)/c_1 \tag{C.5}$$

hold. Finally, the assertion follows from (C.3) and (C.5) by choosing $\delta > 0$ sufficiently small. □

Proof of Lemma 7.2. Assume that $d, \varepsilon > 0$ and $m \in \mathcal{M}(d)$ are such that

$$\limsup_{n\to\infty} \mathbb{E}\langle \|\rho_{\sigma_1,\sigma_2} - \bar{\rho}\|_{\mathrm{TV}}\rangle_{\hat{G}(n,m)} > \varepsilon.$$

Choose $\eta = \eta(\varepsilon) > 0$ small enough and then pick $\delta = \delta(\eta) > 0$ as in Lemma C.1. By assumption, there exist infinitely many n such that $|m - dn/k| < \delta n/2$ and

$$\mathbb{E}\langle \|\rho_{\sigma_1,\sigma_2} - \bar{\rho}\|_{\mathrm{TV}}\rangle_{\hat{G}(n,m)} > \varepsilon/2. \tag{C.6}$$

Fix a large enough n along with m' such that $m < m' < dn/k + 2\delta n$. We are going to argue that

$$\mathbb{E}\langle \|\rho_{\sigma_1,\sigma_2} - \bar{\rho}\|_{\mathrm{TV}}\rangle_{\hat{G}(n,m')} > \delta.$$

By Lemma C.1, there is a coupling of $\hat{\boldsymbol{\sigma}}_{n,m}, \hat{\boldsymbol{\sigma}}_{n,m'}$ such that

$$\mathbb{P}[\hat{\boldsymbol{\sigma}}_{n,m} = \hat{\boldsymbol{\sigma}}_{n,m'}] > 1 - \eta. \tag{C.7}$$

We extend this coupling to a coupling of $G' \stackrel{\mathrm{d}}{=} G^*(n, m, \hat{\boldsymbol{\sigma}}_{n,m})$ and $G'' \stackrel{\mathrm{d}}{=} G^*(n, m', \hat{\boldsymbol{\sigma}}_{n,m'})$ in the natural way. Specifically, given $\hat{\boldsymbol{\sigma}}_{n,m} = \hat{\boldsymbol{\sigma}}_{n,m'}$ we draw G'' with constraints $a_1, \ldots, a_{m'}$ from the distribution $G^*(n, m', \hat{\boldsymbol{\sigma}}_{n,m'})$ and we let G' simply be the factor graph comprising the first m constraints $a_1, \ldots, a_m$. Moreover, if $\hat{\boldsymbol{\sigma}}_{n,m} \neq \hat{\boldsymbol{\sigma}}_{n,m'}$, then we draw G', G'' independently from their respective marginal distributions.

Due to (C.6) and the Nishimori identity (5.4) we have $\mathbb{P}[\langle \|\rho_{\sigma,\tau} - \bar{\rho}\|_{\mathrm{TV}}\rangle_{G'} > \varepsilon/2] \geqslant \varepsilon/2$. Recalling the notion of *nearly balanced* below Lemma 4.5, we observe that because a random sample $\boldsymbol{\tau}$ from $\mu_{G'}$ and $\hat{\boldsymbol{\sigma}}_{n,m}$ are identically distributed, $\boldsymbol{\tau}$ is nearly balanced with probability $1 - o(1)$ by Corollary 6.5. Hence, provided that n is large enough, with probability at least $\varepsilon/3$ the random factor graph G' possesses a nearly balanced satisfying assignment τ_G such that $\langle \|\rho_{\sigma,\tau_{G'}} - \bar{\rho}\|_{\mathrm{TV}}\rangle_{G''} > \varepsilon/2$. In the event that there is no such event we just let $\tau_{G'}$ be an arbitrary nearly balanced assignment (not necessarily a satisfying one). This construction ensures that

$$\mathbb{E}[\langle \|\rho_{\sigma,\tau_{G'}} - \bar{\rho}\|_{\mathrm{TV}}\rangle_{G'}] \geqslant \varepsilon^2/6.$$

Hence, provided that η was chosen small enough, (C.7) and the Nishimori identity (5.4) yield

$$\mathbb{E}[\|\rho_{\hat{\sigma}_{n,m},\tau_{G'}} - \bar{\rho}\|_{\mathrm{TV}} \mid \hat{\boldsymbol{\sigma}}_{n,m} = \hat{\boldsymbol{\sigma}}_{n,m'}] \geqslant \varepsilon^2/7. \tag{C.8}$$

Finally, we also designate a nearly balanced assignment $\tilde{\tau}_{G''}$ for the factor graph G'' by simply letting $\tilde{\tau}_{G''}$ be the assignment $\tau_{G'''}$ of the factor graph G''' obtained from G'' by deleting the last $m' - m$ constraints $a_{m+1}, \ldots, a_{m'}$. Since given $\hat{\boldsymbol{\sigma}}_{n,m} = \hat{\boldsymbol{\sigma}}_{n,m'}$ we have $G''' = G'$, (C.8) yields

$$\mathbb{E}[\|\rho_{\hat{\sigma}_{n,m'},\tilde{\tau}_{G''}} - \bar{\rho}\|_{\mathrm{TV}} \mid \hat{\boldsymbol{\sigma}}_{n,m} = \hat{\boldsymbol{\sigma}}_{n,m'}] = \mathbb{E}[\|\rho_{\hat{\sigma}_{n,m},\tau_{G'}} - \bar{\rho}\|_{\mathrm{TV}} \mid \hat{\boldsymbol{\sigma}}_{n,m} = \hat{\boldsymbol{\sigma}}_{n,m'}] \geqslant \varepsilon^2/7.$$

Therefore, the Nishimori identity (5.4) and (C.7) imply

$$\mathbb{E}[\langle \|\rho_{\sigma,\tilde{\tau}_{G''}} - \bar{\rho}\|_{\mathrm{TV}}\rangle_{G''}] = \mathbb{E}[\|\rho_{\hat{\sigma}_{n,m'}} - \tilde{\tau}_{G''}\|_{\mathrm{TV}}] \geqslant \varepsilon^2/8. \tag{C.9}$$

Since $\tau_{G''}$ is nearly balanced, the assertion follows from (C.9) and Lemma 4.5. □

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

Appendix D: Proof of Lemma 7.4

The Nishimori identity (5.6) shows that $\hat{G} \stackrel{\mathrm{d}}{=} G^*(n, m, \hat{\boldsymbol{\sigma}}_{n,m})$. Moreover, Corollary 6.5 shows that $\hat{\boldsymbol{\sigma}}_{n,m}$ is nearly balanced with probability at least $1 - O(n^{-1})$. Hence, it suffices to prove that for any nearly balanced σ,

$$\begin{aligned} \mathbb{E}[C_Y(G^*(n, m, \sigma))] &= \hat{\kappa}_Y + o(1), \\ \mathbb{P}[\forall l \leqslant L : C_{Y_l}(G^*(n, m, \sigma) = y_l] &= o(1) + \prod_{l=1}^{L} \mathbb{P}[\mathrm{Po}(\hat{\kappa}_{Y_l}) = y_l]. \end{aligned} \tag{D.1}$$

We begin by calculating $\mathbb{E}[C_Y(G^*(n, m, \sigma))]$. Suppose that $i = (i_1, \ldots, i_\ell) \in [n]$ is a family of distinct indices such that $i_1 < \min\{i_2, \ldots, i_\ell\}$ and let $h = (h_1, \ldots, h_\ell) \in [m]$ be pairwise distinct such that $h_1 < h_\ell$ if $\ell > 1$. Set $i_{\ell+1} = i_1$. Let $\mathcal{C}_Y(i, h)$ be the event that $x_{i_1}, a_{h_1}, \ldots, x_{i_\ell}, a_{h_\ell}$ constitute a cycle with signature $Y = (\psi_1, s_1, t_1, \ldots, \psi_\ell, s_\ell, t_\ell)$. Then for any nearly balanced $\sigma \in \Omega^{V_n}$ we have

$$\begin{aligned} &\mathbb{P}[G^*(n, m, \sigma) \in \mathcal{C}_Y(i, h)] \\ &= \prod_{j=1}^{\ell} \frac{\sum_{u_1, \ldots, u_k \in [n]} \mathbf{1}\{u_{s_j} = i_j, u_{t_j} = i_{j+1}\} \psi_j(\sigma(x_{u_1}), \ldots, \sigma(x_{u_k})) P(\psi_j)}{\sum_{u_1, \ldots, u_k \in [n]} \mathbb{E}[\boldsymbol{\psi}(\sigma(x_{u_1}), \ldots, \sigma(x_{u_k}))]} \\ &= o(n^{-2\ell}) + \prod_{j=1}^{\ell} \frac{P(\psi_j)}{n^k \xi} \sum_{u_1, \ldots, u_k \in [n]} \mathbf{1}\{u_{s_j} = i_j, u_{t_j} = i_{j+1}\} \psi_j(\sigma(x_{u_1}), \ldots, \sigma(x_{u_k})) \quad \text{(by SYM)} \\ &= o(n^{-2\ell}) + n^{-2\ell} q^\ell \prod_{j=1}^{\ell} P(\psi_j) \Phi_{\psi_j, s_j, t_j}(\sigma(x_{i_h}), \sigma(x_{i_{h+1}})). \end{aligned} \tag{D.2}$$

Because σ is nearly balanced, summing (D.2) over i, h yields

$$\begin{aligned} \mathbb{E}[C_Y(G^*(n, m, \sigma))] &= \sum_{i,h} \mathbb{P}[G^*(n, m, \sigma) \in \mathcal{C}_Y(i, h)] \\ &= o(1) + \frac{1}{2\ell} \left(\frac{m}{n}\right)^\ell \operatorname{tr} \prod_{j=1}^{\ell} P(\psi_j) \Phi_{\psi_j, s_j, t_j} \\ &= o(1) + \hat{\kappa}_Y, \end{aligned}$$

which is the first part of (D.1).

The second part of (D.1) follows from the first part and a standard method of moments argument. More specifically, since $m = O(n)$ the random factor graph $G^*(n, m, \sigma)$ does not contain two overlapping cycles of bounded length w.h.p. Therefore, a straightforward extension of the above calculation shows that for any $j_1, \ldots, j_L \geqslant 2$ the joint factorial moment of the random variables $C_{Y_1}(G^*(n, m, \sigma)), \ldots, C_{Y_L}(G^*(n, m, \sigma))$ comes to

$$\mathbb{E}\left[\prod_{l=1}^{L} \left(\prod_{u=0}^{j_l - 1} C_{Y_l}(G^*(n, m, \sigma)) - u\right)\right] = o(1) + \prod_{l=1}^{L} \mathbb{E}[C_{Y_l}(G^*(n, m, \sigma))]^{j_l} = o(1) + \prod_{l=1}^{L} \hat{\kappa}_{Y_l}^{j_l}.$$

In effect, the number of cycles with signature Y is asymptotically Poisson with mean $\hat{\kappa}_Y$ by standard results on the joint convergence to asymptotic Poisson variables [21].

Appendix E: Proof of Lemma 7.8

To prove Lemma 7.8 we need the following rough but uniform estimate of the first moment.

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

Claim E.1 (**SYM**, **BAL**). *For any $D > 0$ there exists $c > 0$ such that*

$$cq^n\xi^m \leqslant \mathbb{E}[Z(G(n,m))] \leqslant q^n\xi^m \quad \text{for all } m \leqslant Dn/k.$$

Proof. Since constraints are chosen independently we have

$$\mathbb{E}[Z(G(n,m))] = \sum_{\sigma\in\Omega^{V_n}} \phi(\rho_\sigma)^m.$$

Because **SYM** and **BAL** yield $\phi(\rho_\sigma) \leqslant \xi$ for every σ, the upper bound $\mathbb{E}[Z(G(n,m))] \leqslant q^n\xi^m$ is immediate. With respect to the lower bound, we observe that there are $\Omega(q^n)$ assignments $\sigma : V_n \to \Omega$ with $\|\rho_\sigma - \bar{\rho}\|_{\mathrm{TV}} \leqslant n^{-1/2}$. Lemma 4.1 shows that for any such σ

$$\phi(\rho_\sigma) = \phi(\bar{\rho}) + k\xi\langle \mathbf{1}, \rho_\sigma - \bar{\rho}\rangle + O(\|\rho_\sigma - \bar{\rho}\|_{\mathrm{TV}}^2) = \phi(\bar{\rho}) + O(1/n).$$

Thus,

$$\mathbb{E}[Z(G(n,m))] \geqslant \Omega(q^n)(\phi(\bar{\rho}) + O(1/n))^m = \Omega(q^n\xi^m)$$

uniformly for all $m \leqslant Dn/k$. □

Claim E.2 (**SYM**, **BAL**). *Let $D > 0$. Uniformly for all $m \leqslant Dn/k$ and for all nearly balanced $\sigma \in \Omega^{V_n}$ we have*

$$\mathbb{P}[\hat{\boldsymbol{\sigma}}_{n,m} = \sigma] = \mathbb{P}[\boldsymbol{\sigma}^* = \sigma]\exp(nO(\|\rho_\sigma - \bar{\rho}\|_{\mathrm{TV}}^2) + O(1)). \tag{E.1}$$

Furthermore, for any $\varepsilon > 0$ there is $C > 0$ such that

$$\mathbb{P}[\|\rho_{\hat{\boldsymbol{\sigma}}_{n,m}} - \bar{\rho}\|_{\mathrm{TV}} > Cn^{-1/2}] < \varepsilon.$$

Proof. Recall ϕ from Lemma 4.1. Since constraints are chosen independently, we obtain

$$\mathbb{E}[\psi_{G(n,m)}(\sigma)] = \phi(\rho_\sigma)^m$$

for every σ. Due to Lemma 4.1, $\phi(\rho) = \xi + O(\|\rho - \bar{\rho}\|_{\mathrm{TV}}^2)$. Hence, Claim E.1 reveals that for a nearly balanced σ,

$$\begin{aligned}\mathbb{P}[\hat{\boldsymbol{\sigma}} = \sigma] &= \frac{\mathbb{E}[\psi_{G(n,m)}(\sigma)]}{\mathbb{E}[Z(G(n,m))]}\\ &= \Theta(q^{-n}\xi^{-m})\phi(\rho_\sigma)^m\\ &= \Theta(q^{-n})(1 + O(\|\rho - \bar{\rho}\|_{\mathrm{TV}}^2))^m\\ &= q^{-n}\exp\left(nO(\|\rho_\sigma - \bar{\rho}\|_{\mathrm{TV}}^2) + O(1)\right),\end{aligned}$$

which yields the first assertion. The second assertion follows from the estimate $\mathbb{P}[\hat{\boldsymbol{\sigma}} = \sigma] = \Theta(q^{-n}\xi^{-m})\phi(\rho_\sigma)^m$ and assumption **BAL**, which provides that $\bar{\rho}$ is the maximizer of ϕ. □

Proof of Lemma 7.8. Let $D, \varepsilon > 0$, pick $\delta > 0$ small enough and $n_0 > 0$ big enough. As a first step we observe that for any event $\mathcal{A}$ the following two implications are true:

$$\mathbb{P}[\boldsymbol{\sigma}^* \in \mathcal{A}] < \delta \Rightarrow \mathbb{P}[\hat{\boldsymbol{\sigma}}_{n,m} \in \mathcal{A}] < \varepsilon, \quad \mathbb{P}[\hat{\boldsymbol{\sigma}}_{n,m} \in \mathcal{A}] < \delta \Rightarrow \mathbb{P}[\boldsymbol{\sigma}^* \in \mathcal{A}] < \varepsilon. \tag{E.2}$$

These implications are immediate from Claim E.2. Indeed, assume that $\mathbb{P}[\boldsymbol{\sigma}^* \in \mathcal{A}] < \delta$. Then for a large $C > 0$,

$$\mathbb{P}[\hat{\boldsymbol{\sigma}}_{n,m} \in \mathcal{A}] \leqslant \mathbb{P}[\hat{\boldsymbol{\sigma}}_{n,m} \in \mathcal{A} \mid \|\rho_{\hat{\boldsymbol{\sigma}}_{n,m}} - \bar{\rho}\|_{\mathrm{TV}} \leqslant Cn^{-1/2}] + \varepsilon/2 \leqslant \exp(C^3)\mathbb{P}[\boldsymbol{\sigma}^* \in \mathcal{A}] + \varepsilon/2 < \varepsilon,$$

provided $\delta > 0$ was chosen small enough. The proof of the second implication is analogous.

To derive the assertion from (E.2), let $\mathcal{E}$ be an event and assume that

$$\mathbb{P}[(G^*(n,m,\boldsymbol{\sigma}^*), \boldsymbol{\sigma}^*) \in \mathcal{E}] < \delta.$$

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

Further, for an assignment σ let $\mathcal{E}_\sigma$ be the set of all pairs (G, σ) contained in $\mathcal{E}$. Assuming $\delta > 0$ is sufficiently small, we obtain

$$\begin{aligned}\mathbb{P}[(G^*(n,m,\hat{\boldsymbol{\sigma}}_{n,m}),\hat{\boldsymbol{\sigma}}_{n,m}) \in \mathcal{E}] &= \sum_{\sigma\in\Omega^{V_n}} \mathbb{P}[(G^*(n,m,\sigma),\sigma)\in\mathcal{E}]\mathbb{P}[\hat{\boldsymbol{\sigma}}_{n,m}=\sigma] \\ &\leqslant \varepsilon + \sum_{\sigma\in\Omega^{V_n}} \mathbb{P}[(G^*(n,m,\sigma),\sigma)\in\mathcal{E}]\mathbb{P}[\boldsymbol{\sigma}^*=\sigma] \\ &< 2\varepsilon.\end{aligned}$$

The proof of the reverse direction is analogous. □

Acknowledgement

We thank Chris Brzuska for bringing [11] and the parity–majority problem from Section 3.4 to our attention. We also thank Charilaos Efthymiou and Will Perkins for helpful discussions. Finally, we thank Peter Ayre and Catherine Greenhill for sharing [14].

References

[1] Abbe, E. (2017) Community detection and stochastic block models: Recent developments. *J. Mach. Learn. Res.* **18** 6446–6531.
[2] Abbe, E. and Sandon, C. (2018) Proof of the achievability conjectures for the general stochastic block model. *CPAM* **71** 1334–1406.
[3] Achlioptas, D. and Coja-Oghlan, A. (2008) Algorithmic barriers from phase transitions. In *Proc. 49th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 793–802.
[4] Achlioptas, D., Chtcherba, A., Istrate, G. and Moore, C. (2001) The phase transition in 1-in-k SAT and NAE 3-SAT. In *Proc. 12th Annual ACM–SIAM Symposium on Discrete Algorithms (SODA)*, pp. 721–722.
[5] Achlioptas, D. and Moore, C. (2006) Random k-SAT: Two moments suffice to cross a sharp threshold. *SIAM J. Comput.* **36** 740–762.
[6] Achlioptas, D. and Naor, A. (2005) The two possible values of the chromatic number of a random graph. *Ann. of Math.* **162** 1333–1349.
[7] Achlioptas, D., Naor, A. and Peres, Y. (2005) Rigorous location of phase transitions in hard optimization problems. *Nature* **435** 759–764.
[8] Achlioptas, D. and Peres, Y. (2004) The threshold for random k-SAT is $2^k \ln 2 - O(k)$. *J. Amer. Math. Soc.* **17** 947–973.
[9] Alon, N. and Kahale, N. (1997) A spectral technique for coloring random 3-colorable graphs. *SIAM J. Comput.* **26** 1733–1748.
[10] Anastos, M., Frieze, A. and Pegden, W. (2018) Constraining the clustering transition for colorings of sparse random graphs. *Electron. J. Combin.* **25** P1.72
[11] Applebaum, B. and Lovett, S. (2018) Algebraic attacks against random local functions and their countermeasures. *SIAM J. Comput.* **47** 52–79.
[12] Ayre, P., Coja-Oghlan, A., Gao P. and Müller, N (2019) The satisfiability threshold for random linear equations *Combinatorica*, in press.
[13] Ayre, P., Coja-Oghlan, A. and Greenhill, C. (2019) Hypergraph coloring up to condensation. *Random Struct. Alg.* **54** 615–652.
[14] Ayre, P. and Greenhill, C. (2019) Rigid colorings of hypergraphs and contiguity. *SIAM J. Discrete Math.* **33** 1575–1606.
[15] Bandyopadhyay, A. and Gamarnik, D. (2008) Counting without sampling: Asymptotics of the log-partition function for certain statistical physics models. *Random Struct. Alg.* **33** 452–479.
[16] Banks, J., Moore, C., Neeman, J. and Netrapalli, P. (2016) Information-theoretic thresholds for community detection in sparse networks. In *Proc. 29th Annual Conference on Learning Theory (COLT)*, pp. 383–416.
[17] Bapst, V. and Coja-Oghlan, A. (2016) Harnessing the Bethe free energy. *Random Struct. Alg.* **49** 694–741.
[18] Bapst, V., Coja-Oghlan, A. and Efthymiou, C. (2017) Planting colourings silently. *Combin. Probab. Comput.* **26** 338–366.
[19] Bapst, V., Coja-Oghlan, A., Hetterich, S., Rassmann, F. and Vilenchik, D. (2016) The condensation phase transition in random graph coloring. *Comm. Math. Phys.* **341** 543–606.
[20] Bayati, M., Gamarnik, D. and Tetali, P. (2013) Combinatorial approach to the interpolation method and scaling limits in sparse random graphs. *Ann. Probab.* **41** 4080–4115.
[21] Bollobás, B. (2001) *Random Graphs*, second edition. Cambridge University Press.

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

[22] Cheeseman, P., Kanefsky, B. and Taylor, W. (1991) Where the *really* hard problems are. In *Proc. 12th International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 331–337.
[23] Chvátal, V. and Reed, B. (1992) Mick gets some (the odds are on his side). In *Proc. 33rd Annual Symposium on Foundations of Computer Science (FOCS)*, IEEE, pp. 620–627.
[24] Coja-Oghlan, A. (2013) Upper-bounding the *k*-colorability threshold by counting covers. *Electron. J. Combin.* **20** P32.
[25] Coja-Oghlan, A., Efthymiou, C., Jaafari, N., Kang, M. and Kapetanopoulos, T. (2018) Charting the replica symmetric phase. *Comm. Math. Phys.* **359** 603–698.
[26] Coja-Oghlan, A. and Jaafari, N. (2016) On the Potts model on random graphs. *Electron. J. Combin.* **23** P4.3.
[27] Coja-Oghlan, A., Krzakala, F., Perkins, W. and Zdeborova, L. (2018) Information-theoretic thresholds from the cavity method. *Adv. Math.* **333** 694–795.
[28] Coja-Oghlan, A. and Panagiotou, K. (2012) Catching the *k*-NAESAT threshold. In *Proc. 44th Annual ACM Symposium on Theory of Computing (STOC)*, pp. 899–908.
[29] Coja-Oghlan, A. and Panagiotou, K. (2016) The asymptotic *k*-SAT threshold. *Adv. Math.* **288** 985–1068.
[30] Coja-Oghlan, A. and Reichman, D. (2013) Sharp thresholds and the partition function. *J. Statist. Phys. Conf. Ser.* **473** 012015.
[31] Coja-Oghlan, A. and Wormald, N. (2018) The number of satisfying assignments of random regular k-SAT formulas. *Combin. Probab. Comput.* **4** 496–530.
[32] Coja-Oghlan, A. and Zdeborová, L. (2012) The condensation transition in random hypergraph 2-coloring. In *Proc. 23rd Annual ACM–SIAM Symposium on Discrete Algorithms (SODA)*, pp. 241–250.
[33] Decelle, A., Krzakala, F., Moore, C. and Zdeborová, L. (2011) Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Phys. Rev. E* **84** 066106.
[34] Dietzfelbinger, M., Goerdt, A., Mitzenmacher, M., Montanari, A., Pagh, R. and Rink, M. (2010) Tight thresholds for cuckoo hashing via XORSAT. *International Colloquium on Automata, Languages, and Programming*, Springer, pp. 213–225.
[35] Ding, J., Sly, A. and Sun, N. (2015) Proof of the satisfiability conjecture for large *k*. In *Proc. 47th Annual ACM Symposium on Theory of Computing (STOC)*, pp. 59–68.
[36] Ding, J., Sly, A. and Sun, N. (2016) Satisfiability threshold for random regular NAE-SAT. *Comm. Math. Phys.* **341** 435–489.
[37] Dubois, O. and Mandler, J. (2002) The 3-XORSAT threshold. In *Proc. 43rd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 769–778.
[38] Dyer, M. and Frieze, A. (1989) The solution of some NP-hard problems in polynomial expected time. *J. Algorithms* **10** 451–489.
[39] Dyer, M., Frieze, A. and Greenhill, C. (2015) On the chromatic number of a random hypergraph. *J. Combin. Theory Ser. B* **113** 68–122.
[40] Erdős, P. and Rényi, A. (1960) On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl* **5** 17–61.
[41] Feige, U. (2002) Relations between average case complexity and approximation complexity. In *Proc. 24th Annual ACM Symposium on Theory of Computing (STOC)*, pp. 534–543.
[42] Feldman, V., Perkins, W. and Vempala, S. (2015) On the complexity of random satisfiability problems with planted solutions. In *Proc. 48th Annual ACM Symposium on Theory of Computing (STOC)*, pp. 77–86.
[43] Ferrari, U., Lucibello, C., Morone, F., Parisi, G., Ricci-Tersenghi, F. and Rizzo, T. (2013) Finite-size corrections to disordered systems on Erdős–Rényi random graphs. *Phys. Rev. B* **88** 184201.
[44] Franz, S. and Leone, M. (2003) Replica bounds for optimization problems and diluted spin systems. *J. Statist. Phys.* **111** 535–564.
[45] Galanis, A., Stefankovic, D. and Vigoda, E. (2015) Inapproximability for antiferromagnetic spin systems in the tree nonuniqueness region. *J. Assoc. Comput. Mach.* **62** 50.
[46] Gamarnik, D. and Sudan, M. (2014) Limits of local algorithms over sparse random graphs. In *Proc. 5th Conference on Innovations in Theoretical Computer Science (ITCS)*, pp. 369–376.
[47] Gamarnik, D. and Sudan, M. (2017) Performance of sequential local algorithms for the random NAE-*K*-SAT problem. *SIAM J. Comput.* **46** 590–619.
[48] Gerschenfeld, A. and Montanari, A. (2007) Reconstruction for models on random graphs. In *Proc. 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 194–204.
[49] Goerdt, A. (1996) A threshold for unsatisfiability. *J. Comput. Syst. Sci.* **53** 469–486
[50] Goldreich, O. (2000) Candidate one-way functions based on expander graphs. Cryptology ePrint Archive, report 2000/063. http://eprint.iacr.org/2000/063
[51] Graf, S. and Luschgy, H. (2007) *Foundations of Quantization for Probability Distributions*, Springer.
[52] Graf, S. and Luschgy, H. (2009) Quantization for probability measures in the Prokhorov metric. *SIAM Theory Probab. Appl.* **53** 216–241.
[53] Guerra, F. (2003) Broken replica symmetry bounds in the mean field spin glass model. *Comm. Math. Phys.* **233** 1–12.
[54] van der Hofstad, R. (2017) *Random Graphs and Complex Networks*, Cambridge University Press.
[55] Janson, S. (1995) Random regular graphs: Asymptotic distributions and contiguity. *Combin. Probab. Comput.* **4** 369–405.

https://doi.org/10.1017/S0963548319000440 Published online by Cambridge University Press

[56] Kesten, H. and Stigum, B. (1966) Additional limit theorem for indecomposable multidimensional Galton–Watson processes. *Ann. Math. Statist.* **37** 1463–1481.
[57] Krivelevich, M. and Vilenchik, D. (2006) Semirandom models as benchmarks for coloring algorithms. In *Proc. 3rd Workshop on Analytic Algorithmics and Combinatorics (ANALCO)*, pp. 211–221.
[58] Krzakala, F., Mézard, M. and Zdeborová, L. (2014) Reweighted belief propagation and quiet planting for random *K*-SAT. *J. Satisfiability, Boolean Modeling and Computation* **8** 149–171.
[59] Krzakala, F., Montanari, A., Ricci-Tersenghi, F., Semerjian, G. and Zdeborová, L. (2007) Gibbs states and the set of solutions of random constraint satisfaction problems. *Proc. Nat. Acad. Sci.* **104** 10318–10323.
[60] Krzakala, F. and Zdeborová, L. (2009) Hiding quiet solutions in random constraint satisfaction problems. *Phys. Rev. Lett.* **102** 238701.
[61] Mézard, M. and Montanari, A. (2009) *Information, Physics and Computation*, Oxford University Press.
[62] Mézard, M., Parisi, G. and Zecchina, R. (2002) Analytic and algorithmic solution of random satisfiability problems. *Science* **297** 812–815.
[63] Molloy, M. (2012) The freezing threshold for *k*-colourings of a random graph. In *Proc. 43rd Annual ACM Symposium on Theory of Computing (STOC)*, pp. 921–930.
[64] Montanari, A., Restrepo, R. and Tetali, P. (2011) Reconstruction and clustering in random constraint satisfaction problems. *SIAM J. Discrete Math.* **25** 771–808.
[65] Moore, C. (2017) The computer science and physics of community detection: Landscapes, phase transitions, and hardness. *Bulletin EATCS* **121**.
[66] Mossel, E., Neeman, J. and Sly, A. (2015) Reconstruction and estimation in the planted partition model. *Probab. Theory Rel. Fields* **162** 431–461.
[67] Panchenko, D. and Talagrand, M. (2004) Bounds for diluted mean-fields spin glass models. *Probab. Theory Rel. Fields* **130** 319–336.
[68] Pittel, B. and Sorkin, G. (2016) The satisfiability threshold for k-XORSAT. *Combin. Probab. Comput.* **25** 236–268.
[69] Rassmann, F. (2019) On the number of solutions in random graph k-colouring. *Combin. Probab. Comput.* **28** 130–158.
[70] Robinson, R. and Wormald, N. (1992) Almost all cubic graphs are Hamiltonian. *Random Struct. Alg.* **3** 117–125.
[71] Schaefer, T. (1978) The complexity of satisfiability problems. In *Proc. 10th Annual ACM Symposium on Theory of Computing (STOC)*, pp. 216–226.
[72] Shamir, E. and Spencer, J. (1987) Sharp concentration of the chromatic number of random graphs $G_{n,p}$. *Combinatorica* **7** 121–129
[73] Sly, A. (2011) Reconstruction for the Potts model. *Ann. Probab.* **39** 1365–1406.
[74] Sly, A., Sun, N. and Zhang, Y. (2016) The number of solutions for random regular NAE-SAT. In *Proc. 57th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 724–731.
[75] Zdeborová, L. and Krzakala, F. (2016) Statistical physics of inference: Thresholds and algorithms. *Adv. Phys.* **65** 453–552.

Cite this article: Coja-Oghlan A, Kapetanopoulos T and Müller N (2020). The replica symmetric phase of random constraint satisfaction problems. *Combinatorics, Probability and Computing* **29**, 346–422. https://doi.org/10.1017/S0963548319000440