

Combinatorial approach to the interpolation method and scaling limits in sparse random graphs

Mohsen Bayati *

David Gamarnik †

Prasad Tetali ‡

December 29, 2009

Abstract

We establish the existence of free energy limits for several sparse random hypergraph models corresponding to certain combinatorial models on Erdős-Rényi graph $\mathbb{G}(N, c/N)$ and random r -regular graph $\mathbb{G}(N, r)$. For a variety of models, including independent sets, MAX-CUT, Coloring and K-SAT, we prove that the free energy both at a positive and zero temperature, appropriately rescaled, converges to a limit as the size of the underlying graph diverges to infinity. In the zero temperature case, this is interpreted as the existence of the scaling limit for the corresponding combinatorial optimization problem. For example, as a special case we prove that the size of a largest independent set in these graphs, normalized by the number of nodes converges to a limit w.h.p., thus resolving an open problem, (see Conjecture 2.20 in [Wor99], as well as [Ald],[BR],[JT08] and [AS03]).

Our approach is based on extending and simplifying the interpolation method developed by Guerra and Toninelli [GT02] and Franz and Leone [FL03],[FLT03]. Among other applications, this method was used to prove the existence of free energy limits for Viana-Bray and K-SAT models on Erdős-Rényi graphs. The case of zero temperature was treated by taking limits of positive temperature models. We provide instead a simpler combinatorial approach and work with the zero temperature case (optimization) directly both in the case of Erdős-Rényi graph $\mathbb{G}(N, c/N)$ and random regular graph $\mathbb{G}(N, r)$. In addition we establish the large deviations principle for the satisfiability property for constraint satisfaction problems such as Coloring, K-SAT and NAE-K-SAT. For example, let $p(c, q, N)$ and $p(r, q, N)$ denote, respectively, the probability that random graphs $\mathbb{G}(N, c/N)$ and $\mathbb{G}(N, r)$ are properly q -colorable. We prove the existence of limits of $N^{-1} \log p(c, q, N)$ and $N^{-1} \log p(r, q, N)$, as $N \rightarrow \infty$.

*Department of Electrical Engineering, Stanford University, Stanford, CA 94305; e-mail: bayati@stanford.edu.

†Operations Research Center and Sloan School of Management, MIT, Cambridge, MA 02139; e-mail: gamarnik@mit.edu

‡School of Mathematics and School of Computer Science, Georgia Tech, Atlanta, GA 30332-0160; e-mail: tetali@math.gatech.edu.

1 Introduction

Consider two random graph models on nodes $[N] \triangleq \{1, \dots, N\}$, the Erdős-Rényi graph $\mathbb{G}(N, c/N)$ and the random r -regular graph $\mathbb{G}(N, r)$. The first model is obtained by adding each edge of the $N(N-1)/2$ possible edges randomly independently with probability c/N , where $c > 0$ is a constant (does not grow with N). The second is a graph chosen uniformly at random from the space of all r -regular graphs on N nodes, where the integer r is a fixed constant. It is straightforward to see that $|\mathcal{I}_N|$ grows linearly with N . It was conjectured in several papers including Conjecture 2.20 in [Wor99], [GNS06], [BR], as well as [JT08] and [AS03] that the limit $\lim_N |\mathcal{I}_N|/N$ exists with probability one as $N \rightarrow \infty$. (Additionally, this problem was listed by D. Aldous as one of his six favorite open problems [Ald]. For a new collection of Aldous' favorite open problems see [AldNew]). The fact that the actual value of $|\mathcal{I}_N|$ concentrates around its mean follows from a standard Azuma-type inequality. However, a real challenge is to show that the expected value of $|\mathcal{I}_N|$ normalized by N does not fluctuate around different values, for large N .

This conjecture is in fact just one of a family of similar conjectures. Consider, for example, the random MAX-K-SAT problem - the problem of finding the largest number of satisfiable clauses of size K in a uniformly random instance of a K-SAT problem on N variables with cN clauses. This problem can be viewed as an optimization problem over a sparse random *hypergraph*. A straightforward argument shows that at least $1 - 2^{-K}$ fraction of the clauses can be satisfied with high probability (w.h.p.). It was conjectured in [CGHS04] that the proportion of the largest number of satisfiable clauses has a limit w.h.p. as $N \rightarrow \infty$. As a third example, consider the problem of partial q -coloring of a graph: finding a q -coloring of nodes which maximizes the total number of properly colored edges. It is natural to conjecture again that value of this maximum has a scaling limit w.h.p. (though we are not aware of any papers explicitly stating this conjecture).

Recently a powerful rigorous statistical physics method was introduced by Guerra and Toninelli [GT02] and further developed by Franz and Leone [FL03], Franz, Leone and Toninelli [FLT03], Panchenko and Talagrand [PT04], Montanari [Mon05], Kudekar and Macris [KM06]. The method is based on an ingenious interpolation between a random hypergraph (spin glass) model on N nodes on the one hand, and a disjoint union of random hypergraph models on N_1 and N_2 nodes, on the other hand, where $N = N_1 + N_2$. Using this method it is possible to show for certain random hypergraph models that when one considers the expected log-partition function, the derivative of the interpolation function has a definite sign at *every value* of the interpolation parameter. As a result the expected log-partition function of the N -node model is larger (or smaller depending on the details of the model) than the sum of the corresponding expected log-partition functions on N_1 and N_2 -node models. This super(sub)-additivity property is used to argue the existence of the (thermodynamic) limit of the expected log-partition function scaled by N . From this the existence of the scaling limits for the ground states (optimization problems described above) is also shown by taking a limit as positive temperature approaches zero temperature. In [FL03], the method was used to prove the scaling limit of log-partition functions corresponding to random K-SAT model for even K (and the so-called Viana-Bray models with random symmetric Hamiltonian functions.) After the publication of this paper, it was realized by Maneva and Montanari [FM] that the same argument applies to the case of odd K without much change. The idea of using super- and sub-additivity for proving the existence of scaling limits goes back to classical papers by Beardwood, Halton and Hammersley [BHH59] for random planar Traveling Salesman Problem, and Hammersley and Welsh [HW65] for showing the existence of limiting constants in the first passage percolation problem. It also used to show the existence of the limit of scaled log-partition functions for statistical mechanics models on lattices [Geo88],[Sim93].

Results and technical contributions. The goal of the present work is to simplify and extend the applicability of the interpolation method, and we do this in several important ways.

- First, we extend the interpolation method to a variety of models on Erdős-Rényi graphs not

considered before. Specifically, we consider independent set, MAX-CUT, Ising model, graph coloring (henceforth referred to as Coloring), K-SAT and Not-All-Equal K-SAT (NAE-K-SAT) models. The coloring model, in particular, is of special interest as it becomes the first non-binary model to which interpolation method is applied.

- Second, we provide a simpler and a more combinatorial interpolation scheme as well as analysis. Moreover, we treat the zero temperature case (optimization problem) directly and separately from the case of the log-partition function, and again the analysis turns out to be substantially simpler. As a result, we prove the existence of the limit of the appropriately rescaled value of the optimization problems in these models, including independent set problem, thus resolving an open problem earlier stated.
- Third, we extend the above results to the case of random regular graphs (and hypergraph ensembles, depending on the model). The case of random regular graphs has been considered before by Franz, Leone and Toninelli [FLT03] for the K-SAT and Viana-Bray models with even number of variables per clause, and Montanari [Mon05] in the context of bounds on the performance of certain *low density parity check* (LDPC) codes. In fact, both papers consider general degree distribution models. The second of these papers introduces a more complicated multi-phase interpolation scheme. In this paper we consider a modification of the interpolation scheme used in [FLT03] and apply it to the same six models we are focusing in the case of Erdős-Rényi graph.
- Finally, we prove the large deviation principle for the satisfiability property for Coloring, K-SAT and NAE-K-SAT models on Erdős-Rényi graph in the following sense. A well known satisfiability conjecture [Fri99] states that for each of these models there exists a (model dependent) critical value c^* such that for every $\epsilon > 0$, when the number of edges (or clauses for a SAT-type problem) is at most $(c^* - \epsilon)N$, the model is colorable (satisfiable) w.h.p. and when it is at least $(c^* + \epsilon)N$, it is not colorable (not satisfiable) w.h.p. as $N \rightarrow \infty$. Friedgut [Fri99],[AF99] came close to proving this conjecture by showing that these models exhibit sharp phase transition: there exists a sequence c_N^* such that for every ϵ , the model is colorable (satisfiable) w.h.p. as $N \rightarrow \infty$ when the number of edges (clauses) is at most $(c_N^* - \epsilon)N$ and is not colorable (satisfiable) w.h.p. when the number of edges (clauses) is at least $(c_N^* + \epsilon)N$. It is also reasonable to conjecture, and indeed is known [CGHS04] for the case $K = 2$, that not only the satisfiability conjecture is valid, but, moreover, the probability of satisfiability $p(c, N)$ decays to zero exponentially fast when $c > c^*$. In this paper we show that for these three models, namely Coloring, K-SAT and NAE-K-SAT, the limit $r(c) \triangleq \lim_{N \rightarrow \infty} N^{-1} \log p(c, N)$ exists for every c . Namely, while we do not prove the satisfiability conjecture and exponential rate of convergence to zero of the satisfiability probability above the critical threshold, we do prove that if the convergence to zero occurs exponentially fast, it does so at a well-defined rate. Assuming the validity of the satisfiability conjecture and the exponential rate of decay to zero above c^* , our result implies that $r(c) = 0$ when $c < c^*$ and $r(c) < 0$ when $c > c^*$. Moreover, our results would imply the satisfiability conjecture, if one could strengthen Friedgut's result as follows: for every $\epsilon > 0$, $p(c_N^* + \epsilon, N)$ converges to zero exponentially fast, where c_N^* is the same sequence as in Friedgut's theorem.

Organization of the paper. The remainder of the paper is organized as follows. In the following section we introduce the sparse random (Erdős-Rényi) and random regular (hyper)-graphs and introduce various combinatorial models of interest. Our main results are stated in Section 3. The proofs for the case of Erdős-Rényi graphs are presented in Sections 4 and 5. The proofs for the case of random regular graphs are in section 6. In the Appendix, we state and prove a simple modification of a classical super-additivity theorem - if a sequence is nearly super-additive, it has a limit after an appropriate normalization.

Notations. We close this section with a few notational conventions. $\mathbb{R}(\mathbb{R}_+)$ denotes the set of (non-negative) real values and $\mathbb{Z}(\mathbb{Z}_+)$ denotes the set of (non-negative) integer values. As before, $[N]$ denotes

the set of integers $\{1, \dots, N\}$. Throughout the paper, we treat $[N]$ as a set of nodes, and we consider splitting this into two sets of nodes, namely $[N_1] = \{1, \dots, N_1\}$ and $\{N_1 + 1, \dots, N\}$. For symmetry, with some abuse of notation, it is convenient to denote the second set by $[N_2]$ where $N_2 = N - N_1$. $\text{Bi}(N, \theta)$ denotes binomial distribution with N trials and success probability θ . $\text{Pois}(c)$ denotes a Poisson distribution with parameter c . A sequence of random variables X_N is said to converge to a random variable X with high probability (w.h.p.) if for every $\epsilon > 0$, $\lim_{N \rightarrow \infty} \mathbb{P}(|X_N - X| > \epsilon) = 0$. This is the usual convergence in probability.

2 Sparse random hypergraphs

Given a set of nodes $[N]$, and a positive integer K , a directed hyperedge is any ordered set of nodes $(i_1, \dots, i_K) \in [N]^K$. An undirected hyperedge is an unordered set of K nodes $i_1, \dots, i_K \in [N]$. A directed (undirected) K -uniform hypergraph on the node set $[N]$ is a pair $([N], E)$, where E is any set of directed (undirected) K -hyperedges $E = \{e_1, \dots, e_{|E|}\}$. A graph is called simple if the nodes within each edge $e_m, 1 \leq m \leq |E|$ are distinct and all the edges are distinct. A (directed or undirected) graph is called r -regular if each node $i \in [N]$ appears in exactly r edges. The necessary condition for such a graph to exist is $Nr/K \in \mathbb{Z}_+$. A degree $\Delta_i = \Delta_i(\mathbb{G})$ of a node i is the number of edges containing i . A matching is a set of hyperedges such that each node belongs to exactly one edge.

In order to address a variety of random models in a unified way, we introduce two random directed hypergraph models, namely Erdős-Rényi model $\mathbb{G}(N, M)$, $M \in \mathbb{Z}_+$ and random regular graph model $\mathbb{G}(N, r)$, $r \in \mathbb{Z}_+$. These two graph models, each consisting of N nodes, are described as follows. The first model $\mathbb{G}(N, M)$ is obtained by selecting M directed hyperedges uniformly at random with replacement from the space of all $[N]^K$ hyperedges. A variant of this is a *simple* Erdős-Rényi graph also denoted for convenience by $\mathbb{G}(N, M)$, which is obtained by selecting M edges uniformly at random without replacement from the set of all undirected hyperedges each consisting of *distinct* K nodes. In this paper we will consider exclusively the case when $M = \lceil cN \rceil$ and c is a positive constant (does not grow with N). In this case the probability distribution of the degree of a typical node is $\text{Pois}(c) + O(1/N)$. For this reason we will also call it a *sparse* random Erdős-Rényi graph.

The second model $\mathbb{G}(N, r)$ is defined to be an r -regular directed K -uniform hypergraph generated uniformly at random from the space of all such graphs. We assume $Nr/K \in \mathbb{Z}_+$, so that the set of such graphs is non-empty. A simple (directed or undirected) version of $\mathbb{G}(N, r)$ is defined similarly. In this paper we consider exclusively the case when r is a constant (as a function of N) and we call $\mathbb{G}(N, r)$ a *sparse* random regular graph.

From non-simple to simple graphs. While it is common to work with simple hypergraphs, for our purpose it is more convenient to establish results for directed non-simple graphs first. It is well-known, however, that both $\mathbb{G}(N, M)$ and $\mathbb{G}(N, r)$ graphs are simple with probability which remains at least a constant as $N \rightarrow \infty$, as long as c, r, K are constants. Since we prove statements which hold w.h.p., our results have immediate ramification for simple Erdős-Rényi and regular graphs.

It will be useful to recall the so-called configuration method of constructing the random regular graph [Bol85], [Bol80], [Gal63]. To each node i associate r nodes denoted j_1^i, \dots, j_r^i . We obtain a new set of Nr nodes. Consider a matching $e_1, \dots, e_{Nr/K}$ generated uniformly at random on this set of nodes. From this set of edges we generate a graph on the original N nodes by projecting each of the edge to their representative. Namely an edge (i_1, \dots, i_K) in the graph on N nodes is created iff there is an edge in this set of the form $(j_{k_1}^{i_1}, \dots, j_{k_K}^{i_K})$ for some $k_1, \dots, k_K \in [r]$. The resulting graph is random r -regular (not necessarily simple) graph, which we again denote by $\mathbb{G}(N, r)$. From now on when we talk about configuration graph, we have in mind the graph just described on Nr nodes. It is known [JLR00] that with probability bounded away from zero as $N \rightarrow \infty$ the resulting graph is in fact simple.

Given a hypergraph $\mathbb{G} = ([N], E)$ we will consider a variety of combinatorial structures on \mathbb{G} , which can be defined in a unified way using the notion of a Markov Random Field (MRF). The MRF is a hypergraph \mathbb{G} together with an alphabet $\chi = [q] = \{0, 1, \dots, q\}$ and a set of node and edge potentials $H_i, i \in [N], H_e, e \in E$. A node potential is a function $H_i : [q] \rightarrow \mathbb{R}$ and an edge potential is a function $H_e : [q]^K \rightarrow \{-\infty\} \cup \mathbb{R}$. Given a MRF $(\mathbb{G}, \chi, H_i, H_e, i \in [N], e \in E)$ and any $x \in [q]^N$, let

$$H(x) = \sum_{i \in [N]} H_i(x_i) + \sum_{e \in E} H_e(x_e), \quad H(\mathbb{G}) = \sup_{x \in [q]^N} H(x),$$

where $x_e = (x_i, i \in e)$. Namely, $H(x)$ is the value associated with a chosen assignment x and H is the optimal value, or the groundstate in the statistical physics terminology. In many cases the node and edge potentials will be random functions generated i.i.d. (see examples below).

Associated with an MRF is the Gibbs probability measure $\mu_{\mathbb{G}}$ on the set of node values $[q]^N$ defined as follows. Fix a parameter $\lambda > 0$ consider the probability measure

$$\mu_{\mathbb{G}}(x) = \frac{\lambda^{H(x)}}{Z_{\mathbb{G}}}$$

to every assignment $x \in [q]^N$, where $Z_{\mathbb{G}} = \sum_x \lambda^{H(x)}$ is the normalizing *partition function*. Observe that $\lim_{\lambda \rightarrow \infty} \log Z_{\mathbb{G}} / \log \lambda = H(\mathbb{G})$. Sometimes one considers $\lambda = \exp(1/T)$ where T is *temperature*. The case $T = 0$, namely $\lambda = \infty$ then corresponds to zero temperature, or equivalently the optimization. We distinguish this with a positive temperature case, namely $\lambda < \infty$.

We will consider in this paper a variety of MRF defined on sparse random graphs $\mathbb{G}(N, \lfloor cN \rfloor)$ and $\mathbb{G}(N, r)$. (In the statistical physics literature x_i are called spin values, and the corresponding MRF is called a *diluted spin glass model*.) We now describe some examples of concrete and well-known MRF and show that they fit the framework described above.

- **Independent set.** $K = 2$ and $q = 1$. Define $H_i(1) = 1, H_i(0) = 0$ for all $i \in [N]$. Define $H_e(1, 1) = -\infty, H_e(1, 0) = H_e(0, 1) = H_e(0, 0) = 0$ for every edge $e = (i_1, i_2)$. Then for every vector $x \in \{0, 1\}^N$ we have $H(x) = -\infty$ if there exists an edge $e_j = (i_1, i_2)$ such that $x_{i_1} = x_{i_2} = 1$ and $H(x) = |\{i : x_i = 1\}|$, otherwise. Equivalently, $H(x)$ takes finite value only on x corresponding to independent sets, and in this case it the cardinality of the independent set. $H(\mathbb{G})$ is the cardinality of the largest independent set.
- **MAX-CUT.** $K = 2$ and $q = 1$. Define $H_i(0) = H_i(1) = 0$. Define $H_e(1, 1) = H_e(0, 0) = 0, H_e(1, 0) = H_e(0, 1) = 1$. Every vector $x \in \{0, 1\}^N$ partitions nodes into two subsets of nodes taking values 0 and 1 respectively. $H(x)$ is the number of edges between the two subsets. $H(\mathbb{G})$ is the largest such number, also called maximum cut size.
- **Anti-ferromagnetic Ising model.** $K = 2$ and $q = 1$. Fix $\beta > 0, B \in \mathbb{R}$. Define $H_i(0) = -B, H_i(1) = B$. Define $H_e(1, 1) = H_e(0, 0) = -\beta, H_e(1, 0) = H_e(0, 1) = \beta$. It is more common to use alphabet $\{-1, 1\}$ instead of $\{0, 1\}$ for this model. We use the latter for consistency with the remaining models.
- **q-Coloring** $K = 2$ and q is arbitrary. $H_i(x) = 0, \forall x \in [q]$ and $H_e(x, y) = 0$ if $x = y$ and $H_e(x, y) = 1$ otherwise. Therefore for every $x \in [q]^N, H(x)$ is the number of properly colored edges and $H(\mathbb{G})$ is the maximum number of properly colored edges. This problem is also known as the max-q-cut problem - splitting nodes into q parts so that the number of edges between the parts is maximized.
- **Random K-SAT.** $K \geq 2$ is arbitrary, $q = 1$. $H_i = 0$ for all $i \in [N]$. The edge potentials are defined as follows. For each edge $e \in E$ generate $a_e = (a_1, \dots, a_K)$ uniformly at random from $\{0, 1\}$, independently for all edges. For each edge e set $H_e(a_1, \dots, a_K) = 0$ and $H_e(x) = 1$ for all

other $x = (x_1, \dots, x_K)$. Then for every $x \in \{0, 1\}^N$, $H(x)$ is the number of satisfied clauses and $H(\mathbb{G})$ is the largest number of satisfiable clauses. Often this model is called (random) MAX-K-SAT model. We drop the MAX prefix in the notation.

- **NAE-K-SAT (Not-All-Equal-K-SAT)**. The setting is as above except now we set $H_e(a_1, \dots, a_K) = H_e(1 - a_1, \dots, 1 - a_K) = 0$ and $H_e(x) = 1$ for all other x for each e .

It is for the K-SAT and NAE-K-SAT models that considering directed as opposed to undirected hypergraphs is convenient, as for these models the order of nodes in edges matters. For the remaining models, however, this is not the case.

In several examples considered above we have had only two possible values for the edge potential H_e and one value for the node potential. Specifically, for the cases of Coloring, K-SAT and NAE-K-SAT problems, H_e took only values 0 and 1. It makes sense to call instances of such problems “satisfiable” if $H(\mathbb{G}) = |E|$, namely every edge potential takes value 1. In the combinatorial optimization terminology this corresponds to finding a proper coloring, a satisfying assignment and a NAE satisfying assignment, respectively. We let $p(N, M) = \mathbb{P}(H(\mathbb{G}(N, M)) = M)$ denote the probability of satisfiability when the underlying graph is Erdős-Rényi graph $\mathbb{G}(N, M)$. We also let $p(N, r) = \mathbb{P}(H(\mathbb{G}(N, r)) = rNK^{-1})$ denote the satisfiability probability for a random regular graph $\mathbb{G}(N, r)$.

3 Main results

We now state our main results. Our first set of results concerns Erdős-Rényi graph $\mathbb{G}(N, \lfloor cN \rfloor)$.

Theorem 1. *For every $c > 0$, and for every one of the six models described in Section 2, there exists (model dependent) $H(c)$ such that*

$$\lim_{N \rightarrow \infty} N^{-1} H(\mathbb{G}(N, \lfloor cN \rfloor)) = H(c), \quad (1)$$

w.h.p. Also for every $c > 0$ there exists $p(c)$ such that

$$\lim_{N \rightarrow \infty} N^{-1} \log p(N, \lfloor cN \rfloor) = p(c), \quad (2)$$

for Coloring, K-SAT and NAE-K-SAT models. Moreover, $H(c), p(c)$ are continuous non-increasing functions of c , except for the MAX-CUT problem, in which case $H(c)$ is a continuous non-decreasing function of c .

As a corollary one obtains the following variant of the satisfiability conjecture.

Corollary 1. *For Coloring, K-SAT and NAE-K-SAT models there exists a critical value c_h^* such that $H(c) = c$ when $c < c_h^*$ and $H(c) < c$ when $c > c_h^*$. Similarly, there exists c_p^* , such that $p(c) = 0$ when $c < c_p^*$ and $p(c) < 0$ when $c > c_p^*$.*

Namely, there exists a threshold value c^* such that if the number of clauses is smaller than c^*N there exists a *nearly* satisfiable assignment (assignment satisfying all but $o(N)$ clauses), and if the number of clauses is larger than c^*N , then every assignment violates linearly in N many clauses. The interpretation for Coloring is similar. The result above was established earlier by the second author for randomly generated linear programming problem, using local weak convergence and martingale techniques [Gam04]. It would be interesting to see if the same result is obtainable using the interpolation method.

Can one use Corollary 1 to prove the satisfiability conjecture in the precise sense? The answer would be affirmative, provided that a stronger version of Friedgut’s result [Fri99] on the sharp thresholds for satisfiability properties holds.

Conjecture 1. For the Coloring, K-SAT and NAE-K-SAT models there exists a sequence M_N^* and $\gamma > 0$, such that for every $\epsilon > 0$,

$$\lim_{N \rightarrow \infty} p(N, \lfloor (1 - \epsilon)M_N^* \rfloor) = 1 \quad \text{and} \quad p(N, \lfloor (1 + \epsilon)M_N^* \rfloor) = O(\exp(-\gamma N)), \quad \text{for all } N.$$

In contrast, Friedgut's sharp phase transition result [Fri99] replaces the second part of this conjecture with (the weaker) $\lim_{N \rightarrow \infty} p(N, \lfloor (1 + \epsilon)M_N^* \rfloor) = 0$. Thus, we conjecture that beyond the phase transition region M_N^* , not only is the model *not satisfiable* w.h.p., but in fact the probability of satisfiability converges to zero exponentially fast. Conjecture 1 together with Theorem 1 implies the satisfiability conjecture using a simple counting argument which we omit. This conjecture is known and was mentioned on several occasions, but to the best of our knowledge, was never stated explicitly in any of the publications.

Let us now state our results for the existence of the scaling limit for the log-partition functions.

Theorem 2. For every $c > 0, \lambda \geq 1$, and for every one of the six models described in Section 2, there exists (model dependent) $z(c)$ such that

$$\lim_{N \rightarrow \infty} N^{-1} \log Z(\mathbb{G}(N, \lfloor cN \rfloor)) = z(c), \quad (3)$$

w.h.p., where $z(c)$ is continuous non-increasing functions of c , for all models except for MAX-CUT, in which case $z(c)$ is a continuous non-decreasing function of c .

Remark: The case $\lambda = 1$ is actually uninteresting as it corresponds to no interactions between the nodes leading to $Z(\mathbb{G}) = \prod_{i \in [N]} \lambda^{\sum_{x \in [q]} H_i(x)}$. In this case the limit of $N^{-1} \log Z(\mathbb{G}(N, \lfloor cN \rfloor))$ exists trivially when node potentials H_i are i.i.d. Unfortunately the proof technique based on interpolation method does not seem to extend to the case $\lambda < 1$. For Ising model this corresponds to a ferromagnetic case and the existence of the limit was established in [DM] using a local analysis technique.

We now turn to our results on random regular graphs.

Theorem 3. For every $r \in \mathbb{Z}_+$, and for every one of the six models described in the previous section, there exists (model dependent) $H(r)$ such that $\lim_{N \rightarrow \infty, N \in r^{-1}K\mathbb{Z}_+} N^{-1} H(\mathbb{G}(N, r)) = H(r)$ w.h.p. Also for every $r \in \mathbb{Z}_+$ there exists $p(r)$ such that $\lim_{N \rightarrow \infty, N \in r^{-1}K\mathbb{Z}_+} N^{-1} \log p(N, r) = p(r)$ for K-SAT and NAE-K-SAT models.

Note, that in the statement of the theorem we take limits along subsequence N such that NrK^{-1} is an integer, so that the resulting random hypergraph is well-defined. Unlike the case of Erdős-Rényi graph, we were unable to prove the existence of the large deviation rate $\lim_{N \rightarrow \infty, N \in r^{-1}K\mathbb{Z}_+} N^{-1} \log p(N, r)$ for the case of coloring on random regular graph. At the end of the proof of Theorem 3 we discuss challenges associated with obtaining such a result, which we still believe is true.

Finally, we state our results for the log-partition function limits for random regular graphs.

Theorem 4. For every $r \in \mathbb{Z}_+, \lambda \geq 1$, and for every one of the six models described in the previous section, there exists (model dependent) $z(r)$ such that w.h.p, we have

$$\lim_{N \rightarrow \infty} N^{-1} \log Z(\mathbb{G}(N, r)) = z(r). \quad (4)$$

4 Proofs: Optimization problems in Erdős-Rényi graphs

The following simple observation will be useful throughout the paper. Given two hypergraphs $\mathbb{G}_i = ([N], E_i)$, $i = 1, 2$ on the same set of nodes $[N]$ for each one of the six models in Section 2

$$|H(\mathbb{G}_1) - H(\mathbb{G}_2)| = O(|E_1 \Delta E_2|). \quad (5)$$

This follows from the fact that adding (deleting) an edge to (from) a graph changes the value of H by at most 1 for all models except for the Ising model, where the constant is β .

Our main technical result leading to the proof of Theorem 1 is as follows.

Theorem 5. *For every $1 \leq N_1, N_2 \leq N - 1$ such that $N_1 + N_2 = N$,*

$$\mathbb{E}[H(\mathbb{G}(N, \lfloor cN \rfloor))] \geq \mathbb{E}[H(\mathbb{G}(N_1, M_1))] + \mathbb{E}[H(\mathbb{G}(N_2, M_2))] \quad (6)$$

$$\log p(N, \lfloor cN \rfloor) \geq \log p(N, M_1) + \log p(N, M_2), \quad (7)$$

where $M_j \stackrel{d}{=} \text{Bi}(\lfloor cN \rfloor, N_j/N)$, $j = 1, 2$.

We remark that the randomness underlying the probability $p(N, M_j) = \mathbb{P}(H(\mathbb{G}(N_j, M_j)) = M_j)$ is both with respect to the randomness in the graph generation and the number of hyperedges M_j . Also in the theorem above, we do not assume independence of $M_j, j = 1, 2$.

Let us first show how this result implies Theorem 1.

Proof of Theorem 1. Since M_j have Bernoulli distribution, we have $\mathbb{E}[|M_j - \lfloor cN_j \rfloor|] = O(\sqrt{N})$. This together with observation (5) and Theorem 5 implies

$$\mathbb{E}[H(\mathbb{G}(N, \lfloor cN \rfloor))] \geq \mathbb{E}[H(\mathbb{G}(N_1, \lfloor cN_1 \rfloor))] + \mathbb{E}[H(\mathbb{G}(N_2, \lfloor cN_2 \rfloor))] - O(\sqrt{N}). \quad (8)$$

Namely the sequence $\mathbb{E}[H(\mathbb{G}(N, \lfloor cN \rfloor))]$ is “nearly” super-additive, short of the $O(\sqrt{N})$ correction term. Now we use Proposition 4 in the Appendix for the case $\alpha = 1/2$ to conclude that the limit

$$\lim_{N \rightarrow \infty} N^{-1} \mathbb{E}[H(\mathbb{G}(N, \lfloor cN \rfloor))] \triangleq H(c)$$

exists. Showing that this also implies convergence of $N^{-1}H(\mathbb{G}(N, \lfloor cN \rfloor))$ to $H(c)$ w.h.p. can be done using standard concentration results [JLR00] and we skip the details. It remains to show that $H(c)$ is a non-increasing (non-decreasing for MAX-CUT) continuous function. The first is an immediate consequence of the fact that $\mathbb{E}[H(\mathbb{G}(N, M_1))] \geq \mathbb{E}[H(\mathbb{G}(N, M_2))]$ ($\mathbb{E}[H(\mathbb{G}(N, M_1))] \leq \mathbb{E}[H(\mathbb{G}(N, M_2))]$ for MAX-CUT) when $M_1 \leq M_2$ - adding hyperedges can only decrease (increase for MAX-CUT) the objective value. The continuity follows from (5) which implies

$$\left| \mathbb{E}[H(\mathbb{G}(N, M_1))] - \mathbb{E}[H(\mathbb{G}(N, M_2))] \right| = O(|M_1 - M_2|).$$

In fact, this implies Lipschitz continuity of $H(c)$. This concludes the proof of (1).

We now turn to the proof of (2). We first establish the following claim for the three models of interest (Coloring, K-SAT, NAE-K-SAT).

Lemma 1.

$$|\log p(N_j, M_j) - \log p(N_j, \lfloor cN_j \rfloor)| = O(\sqrt{N}), \quad j = 1, 2. \quad (9)$$

Proof. We first assume K-SAT or NAE-K-SAT models. Note that for these models there exists a constant $\beta > 0$ such for every graph and potential realization $(\mathbb{G} = (V, E), H)$ such that the problem is satisfiable (namely $H(\mathbb{G}) = |E|$), if a randomly chosen hyperedge e is added with a potential chosen according to the model, then

$$\mathbb{P}(H(\mathbb{G} + e) = |E| + 1) \geq \beta.$$

In other words, if the current graph is satisfiable, the new graph obtained by adding a hyperedge remains satisfiable with at least a constant probability. Indeed, for example for the case of K-SAT, if the instance is satisfiable and x is a satisfying assignment, the added edge remains consistent with x with probability

at least $\beta \triangleq 1 - 1/2^K$. For the case of NAE-K-SAT it is $\beta = 1 - 1/2^{K-1}$. This observation implies that for every positive M, m ,

$$\mathbb{P}\left(H(\mathbb{G}(N, M+m)) = M+m \mid H(\mathbb{G}(N, M)) = M\right) \geq \beta^m,$$

further implying

$$\log \mathbb{P}(H(\mathbb{G}(N, M+m)) = M+m) \geq \log \mathbb{P}(H(\mathbb{G}(N, M)) = M) - m \log(1/\beta),$$

Now, suppose $M_j \stackrel{d}{=} \text{Bi}(\lfloor cN \rfloor, N_j/N)$. Using the concavity of log

$$\begin{aligned} \log \mathbb{P}(H(\mathbb{G}(N_j, M_j)) = M_j) &= \log \left(\sum_{m \geq 0} \mathbb{P}(H(\mathbb{G}(N_j, m)) = m) \mathbb{P}(M_j = m) \right) \\ &\geq \sum_{m \geq 0} \log(\mathbb{P}(H(\mathbb{G}(N_j, m)) = m)) \mathbb{P}(M_j = m). \end{aligned}$$

Again using the fact $\mathbb{E}[|M_j - \lfloor cN_j \rfloor|] = O(\sqrt{N})$ we obtain (9).

For the case of Coloring the proof is more involved. Given a constant $\delta > 0$ we call a graph \mathbb{G} on N nodes δ -unusual if it is colorable and in every coloring assignment there exists a color class with size at least $(1 - \delta)N$. Namely, for every x such that $H(x) = |E|$, there exists $k \in [q]$ such that the cardinality of the set $\{i \in [N] : x_i = k\}$ is at least $(1 - \delta)N$. We claim that if $M = \Theta(N)$ then

$$\mathbb{P}(\mathbb{G}(N, M) \text{ is } \delta\text{-unusual}) \leq \alpha^N(\delta), \tag{10}$$

for some $\alpha(\delta)$ such that $\alpha(\delta) \rightarrow 0$ as $\delta \rightarrow 0$. The claim is shown using the first moment method - the expected number of graphs with such a property is at most $\alpha^N(\delta)$. Indeed, given a subset $C \subset [N]$ such that $|C| \geq (1 - \delta)N$, the probability that the graph $\mathbb{G}(N, M)$ is consistent with coloring nodes in C with only one color is at most $(1 - (1 - \delta)^2)^{\Theta(N)}$, since we must have that no edge falls within the class C . There are at most $\binom{N}{\delta N} \approx \exp(H(\delta)N)$ choices for the subset C , where $H(\delta) = -\delta \log \delta - (1 - \delta) \log(1 - \delta)$ is the usual entropy function. It is easy to check that $\alpha(\delta) \triangleq \exp(-H(\delta))(1 - (1 - \delta)^2)^{\Theta(1)}$ satisfies $\alpha(\delta) \stackrel{\delta \rightarrow 0}{\rightarrow} 0$ and the claim is established.

Now observe that if a graph $\mathbb{G} = (V, E)$ is colorable but not δ -unusual, then adding a random edge e we obtain $\mathbb{P}(H(\mathbb{G} + e) = |E| + 1) \geq \delta(1 - \delta) \triangleq \beta$. Namely, in this case the probability goes down by at most a constant factor. This observation implies that when $M = \Theta(N)$, we have

$$\begin{aligned} \mathbb{P}(H(\mathbb{G}(N, M+1)) = M+1) &\geq \mathbb{P}(H(\mathbb{G}(N, M+1)) = M+1 \mid \mathbb{G}(N, M) \text{ is colorable and is not } \delta\text{-unusual}) \\ &\quad \times \mathbb{P}(\mathbb{G}(N, M) \text{ is colorable and is not } \delta\text{-unusual}) \\ &\geq \beta \mathbb{P}(\mathbb{G}(N, M) \text{ is colorable and is not } \delta\text{-unusual}) \\ &\geq \beta \mathbb{P}(\mathbb{G}(N, M) \text{ is colorable}) - \mathbb{P}(\mathbb{G}(N, M) \text{ is } \delta\text{-unusual}) \\ &= \beta \mathbb{P}(H(\mathbb{G}(N, M)) = M) - \alpha^N(\delta). \end{aligned}$$

From this we obtain that if $M = \Theta(N)$, then

$$\mathbb{P}(H(\mathbb{G}(N, M+m)) = M+m) \geq \beta^m \mathbb{P}(H(\mathbb{G}(N, M)) = M) - (1 - \beta)^{-1} \alpha^N(\delta)$$

Now observe that $\mathbb{P}(H(\mathbb{G}(N, M)) = M) \geq (1 - 1/q)^M$. Indeed split nodes into q equal groups. The likelihood that this creates a valid q -coloring is at least the claimed value. Now select δ small enough, so that $\alpha(\delta) < (1/2)\beta(1 - 1/q)$. Then for this choice of δ , we have obtain $\mathbb{P}(H(\mathbb{G}(N, M+m)) = M+m) \geq (\beta/2)^m \mathbb{P}(H(\mathbb{G}(N, M)) = M)$ for all large enough N . The remainder of the proof of the lemma is similar to the one for K-SAT and NAE-K-SAT models. \square

Lemma 1 in combination with Theorem 5 implies

$$\log p(N, \lfloor cN \rfloor) \geq \log p(N_1, \lfloor cN_1 \rfloor) + \log p(N_2, \lfloor cN_2 \rfloor) - O(\sqrt{N}). \quad (11)$$

To complete the proof of (2), we use Proposition 4 from the Appendix with α again set to $1/2$. Lipschitz continuity and monotonicity of $p(c)$ is proven similarly as for $H(c)$. \square

We now turn to the proof of Theorem 5 and in particular introduce the interpolation construction.

Proof of Theorem 5. We begin by constructing a sequence of graphs interpolating between $\mathbb{G}(N, \lfloor cN \rfloor)$ and a disjoint union of $\mathbb{G}(N_1, M_1)$ and $\mathbb{G}(N_2, M_2)$. Given N, N_1, N_2 s.t. $N_1 + N_2 = N$ and any $0 \leq r \leq \lfloor cN \rfloor$, let $\mathbb{G}(N, \lfloor cN \rfloor, r)$ be the random graph on nodes $[N]$ obtained as follows. It contains precisely $\lfloor cN \rfloor$ hyperedges. The first r hyperedges e_1, \dots, e_r are selected u.a.r. from all the possible directed hyperedges (namely they are generated as hyperedges of $\mathbb{G}(N, \lfloor cN \rfloor)$). The remaining $\lfloor cN \rfloor - r$ hyperedges $e_{r+1}, \dots, e_{\lfloor cN \rfloor}$ are generated as follows. For each $j = r+1, \dots, \lfloor cN \rfloor$, with probability N_1/N , e_j is generated independently u.a.r. from all the possible hyperedges on nodes $[N_1]$, and with probability N_2/N , it is generated u.a.r. from all the possible hyperedges on nodes $[N_2] (= \{N_1 + 1, \dots, N\})$. The choice of node and edge potentials H_v, H_e is done exactly according to the corresponding model, as for the case of graphs $\mathbb{G}(N, \lfloor cN \rfloor)$. Observe that when $r = \lfloor cN \rfloor$, $\mathbb{G}(N, \lfloor cN \rfloor, r) = \mathbb{G}(N, \lfloor cN \rfloor)$, and when $r = 0$, $\mathbb{G}(N, \lfloor cN \rfloor, r)$ is a union of disjoint graphs $\mathbb{G}(N_1, M_1), \mathbb{G}(N_2, M_2)$, where $M_j \stackrel{d}{=} \text{Bi}(\lfloor cN \rfloor, N_j/N)$.

Proposition 1. *For every $r = 1, \dots, \lfloor cN \rfloor$,*

$$\mathbb{E}[H(\mathbb{G}(N, \lfloor cN \rfloor, r))] \geq \mathbb{E}[H(\mathbb{G}(N, \lfloor cN \rfloor, r-1))]. \quad (12)$$

Also for Coloring, K-SAT and NAE-K-SAT models

$$\mathbb{P}\left(H(\mathbb{G}(N, \lfloor cN \rfloor, r)) = \lfloor cN \rfloor\right) \geq \mathbb{P}\left(H(\mathbb{G}(N, \lfloor cN \rfloor, r-1)) = \lfloor cN \rfloor\right). \quad (13)$$

Let us first show how Theorem 5 follows from this proposition. Observe that for a disjoint union of two graphs $\mathbb{G} = \mathbb{G}_1 + \mathbb{G}_2$, with $\mathbb{G} = (V, E), \mathbb{G}_1 = (V_1, E_1), \mathbb{G}_2 = (V_2, E_2)$, we always have $H(\mathbb{G}) = H(\mathbb{G}_1) + H(\mathbb{G}_2)$ and $\mathbb{P}(H(\mathbb{G}) = |E|) = \mathbb{P}(H(\mathbb{G}_1) = |E_1|)\mathbb{P}(H(\mathbb{G}_2) = |E_2|)$. The second observation implies further that $\log \mathbb{P}(H(\mathbb{G}) = |E|) = \log \mathbb{P}(H(\mathbb{G}_1) = |E_1|) + \log \mathbb{P}(H(\mathbb{G}_2) = |E_2|)$. Theorem 5 then follows from Proposition 1. \square

Proof of Proposition 1. Observe that $\mathbb{G}(N, \lfloor cN \rfloor, r-1)$ is obtained from $\mathbb{G}(N, \lfloor cN \rfloor, r)$ by deleting a hyperedge chosen u.a.r. independently from r hyperedges e_1, \dots, e_r and adding a hyperedge either to nodes $[N_1]$ or to $[N_2]$ with probabilities N_1/N and N_2/N respectively. Let \mathbb{G}_0 be the graph obtained after deleting but before adding a hyperedge. For the case of K-SAT and NAE-K-SAT (two models with random edge potentials), assume that \mathbb{G}_0 also encodes the underlying edge potentials of the instance. For the case of Coloring, K-SAT, NAE-K-SAT, note that the maximum value that H can achieve for the graph \mathbb{G}_0 is $\lfloor cN \rfloor - 1$ since exactly one hyperedge was deleted. We will establish a stronger result: conditional on *any* realization of the graph \mathbb{G}_0 (and random potentials), we claim that

$$\mathbb{E}[H(\mathbb{G}(N, \lfloor cN \rfloor, r)) | \mathbb{G}_0] \geq \mathbb{E}[H(\mathbb{G}(N, \lfloor cN \rfloor, r-1)) | \mathbb{G}_0]. \quad (14)$$

and

$$\mathbb{P}\left(H(\mathbb{G}(N, \lfloor cN \rfloor, r)) = \lfloor cN \rfloor \mid \mathbb{G}_0\right) \geq \mathbb{P}\left(H(\mathbb{G}_0(N, \lfloor cN \rfloor, r-1)) = \lfloor cN \rfloor \mid \mathbb{G}_0\right) \quad (15)$$

for Coloring, K-SAT, NAE-K-SAT. Proposition then follows immediately from these claims by averaging over \mathbb{G}_0 . Observe that *conditional on any realization* \mathbb{G}_0 , $\mathbb{G}(N, \lfloor cN \rfloor, r)$ is obtained from \mathbb{G}_0 by adding a hyperedge to $[N]$ u.a.r. That is the generation of this hyperedge is independent from the randomness of \mathbb{G}_0 . Similarly, conditional on any realization \mathbb{G}_0 , $\mathbb{G}(N, \lfloor cN \rfloor, r-1)$ is obtained from \mathbb{G}_0 by adding a hyperedge to $[N_1]$ or $[N_2]$ u.a.r. with probabilities N_1/N and N_2/N respectively.

We now prove properties (14) and (15) for each of the six models.

- **Independent sets.** Let $O^* \subset [N]$ be the set of nodes which belong to *every* largest independent set in \mathbb{G}_0 . Namely if $I \subset [N]$ is an i.s. such that $|I| = H(\mathbb{G}_0)$, then $O^* \subset I$. Then for every edge $e = (i, k)$, $H(\mathbb{G}_0 + e) = H(\mathbb{G}_0) - 1$ if $i, k \in O^*$ and $H(\mathbb{G}_0 + e) = H(\mathbb{G}_0)$ if either $i \notin O^*$ or $k \notin O^*$. Here $\mathbb{G}_0 + e$ denotes a graph obtained from \mathbb{G}_0 by adding e . When the edge e is generated u.a.r. from the all possible edges, we then obtain $\mathbb{E}[H(\mathbb{G}_0 + e)|\mathbb{G}_0] - H(\mathbb{G}_0) = -\left(\frac{|O^*|}{N}\right)^2$. Therefore, $\mathbb{E}[H(\mathbb{G}(N, \lfloor cN \rfloor, r)|\mathbb{G}_0)] - H(\mathbb{G}_0) = -\left(\frac{|O^*|}{N}\right)^2$. By a similar argument

$$\begin{aligned} \mathbb{E}\left[H(\mathbb{G}(N, \lfloor cN \rfloor, r-1)|\mathbb{G}_0)] - H(\mathbb{G}_0) &= -\frac{N_1}{N} \left(\frac{|O^* \cap [N_1]|}{N_1}\right)^2 - \frac{N_2}{N} \left(\frac{|O^* \cap [N_2]|}{N_2}\right)^2 \\ &\leq -\left(\frac{N_1}{N} \frac{|O^* \cap [N_1]|}{N_1} + \frac{N_2}{N} \frac{|O^* \cap [N_2]|}{N_2}\right)^2 \\ &= -\left(\frac{|O^*|}{N}\right)^2 = \mathbb{E}[H(\mathbb{G}(N, \lfloor cN \rfloor, r)|\mathbb{G}_0)] - H(\mathbb{G}_0), \end{aligned}$$

and (14) is established.

- **MAX-CUT.** Given \mathbb{G}_0 , let $\mathcal{C}^* \subset \{0, 1\}^{[N]}$ be the set of optimal solutions. Namely $H(x) = H(\mathbb{G}_0), \forall x \in \mathcal{C}^*$ and $H(x) < H(\mathbb{G}_0)$ otherwise. Introduce an equivalency relationship \sim on $[N]$. Given $i, k \in [N]$, define $i \sim k$ if for every $x \in \mathcal{C}^*$, $x_i = x_k$. Namely, in every optimal cut, nodes i and k have the same value. Let $O_j^* \subset [N], 1 \leq j \leq J$ be the corresponding equivalency classes. Given any edge $e = (i, k)$, observe that $H(\mathbb{G}_0 + e) = H(\mathbb{G}_0)$ if $i \sim k$ and $H(\mathbb{G}_0 + e) = H(\mathbb{G}_0) + 1$ otherwise. Thus

$$\mathbb{E}\left[H(\mathbb{G}(N, \lfloor cN \rfloor, r)|\mathbb{G}_0)] - H(\mathbb{G}_0) = 1 - \sum_{1 \leq j \leq J} \left(\frac{|O_j^*|}{N}\right)^2,$$

and

$$\mathbb{E}[H(\mathbb{G}(N, \lfloor cN \rfloor, r-1)|\mathbb{G}_0)] - H(\mathbb{G}_0) = 1 - \frac{N_1}{N} \sum_{1 \leq j \leq J} \left(\frac{|O_j^* \cap [N_1]|}{N_1}\right)^2 - \frac{N_2}{N} \sum_{1 \leq j \leq J} \left(\frac{|O_j^* \cap [N_2]|}{N_2}\right)^2.$$

Using $\frac{N_1}{N} \left(\frac{|O_j^* \cap [N_1]|}{N_1}\right)^2 + \frac{N_2}{N} \left(\frac{|O_j^* \cap [N_2]|}{N_2}\right)^2 \geq \left(\frac{|O_j^*|}{N}\right)^2$ we obtain (14).

- **Ising.** The proof is almost identical as for the MAX-CUT problem. The only difference is that $H(\mathbb{G}_0 + e) = H(\mathbb{G}_0) + \beta$ if $e = (i, j)$ and $i \not\sim j$ and $H(\mathbb{G}_0 + e) < H(\mathbb{G}_0) + \beta$ if $i \sim j$. Notice that the presence of the node potential (magnetic field B) does not affect the argument.
- **Coloring.** Let $\mathcal{C}^* \subset [q]^N$ be the set of optimal colorings. Namely $H(x) = H(\mathbb{G}_0), \forall x \in \mathcal{C}^*$. Given $i, k \in [N]$, define $i \sim k$ iff $x_i = x_k$ for every $x \in \mathcal{C}^*$. Namely, in every optimal coloring assignments, i and k receive the same color. Then for every edge e , $H(\mathbb{G}_0 + e) = H(\mathbb{G}_0) - 1$ if $i \sim k$ and $H(\mathbb{G}_0 + e) = H(\mathbb{G}_0)$ otherwise. The remainder of the proof of (14) is similar to the one for MAX-CUT.

Now let us show (15). Thus assume \mathbb{G}_0 is a colorable graph. Since it has $\lfloor cN \rfloor - 1$ edges it means $H(\mathbb{G}_0) = \lfloor cN \rfloor - 1$. Letting $O_j^* \subset [N], 1 \leq j \leq J$ denote the equivalence classes, we obtain that

$$\mathbb{P}\left(H(\mathbb{G}(N, \lfloor cN \rfloor, r)) = \lfloor cN \rfloor \mid \mathbb{G}_0\right) = 1 - \sum_{1 \leq j \leq J} \left(\frac{|O_j^*|}{N}\right)^2.$$

Similarly,

$$\mathbb{P}\left(H(\mathbb{G}(N, \lfloor cN \rfloor, r-1)) = \lfloor cN \rfloor \mid \mathbb{G}_0\right) = 1 - \frac{N_1}{N} \sum_{1 \leq j \leq J} \left(\frac{|O_j^* \cap [N_1]|}{N_1}\right)^2 - \frac{N_2}{N} \sum_{1 \leq j \leq J} \left(\frac{|O_j^* \cap [N_2]|}{N_2}\right)^2.$$

The relation (15) then again follows from convexity.

- **K-SAT.** Let $\mathcal{C}^* \subset \{0, 1\}^N$ be the set of optimal assignments. Define a node i (variable x_i) to be *frozen* if either $x_i = 0, \forall x \in \mathcal{C}^*$ or $x_i = 1, \forall x \in \mathcal{C}^*$. Namely, in every optimal assignment the value of i is always the same. Let O^* be the set of frozen variables. Let $e = (i_1, \dots, i_K) \subset [N]$ be a hyperedge and let $H_e : \{0, 1\}^K \rightarrow \{0, 1\}$ be some corresponding edge potential. Namely, for some $y_1, \dots, y_K \in \{0, 1\}$, $H_e(x_{i_1}, \dots, x_{i_K}) = 0$ if $x_{i_1} = y_1, \dots, x_{i_K} = y_K$ and $H_e = 1$ otherwise. Consider adding e with H_e to the graph \mathbb{G}_0 . Note that if $e \cap ([N] \setminus O^*) \neq \emptyset$ then $H(\mathbb{G}_0 + e) = H(\mathbb{G}_0) + 1$, as in this case at least one variable in e is non-frozen and can be adjusted to satisfy the clause. Otherwise, suppose $e \subset O^*$, and let $x_{i_1}^*, \dots, x_{i_K}^* \in \{0, 1\}$ be the corresponding frozen values of i_1, \dots, i_K . Then $H(\mathbb{G}_0 + e) = H(\mathbb{G}_0)$ if $x_{i_1}^* = y_1, \dots, x_{i_K}^* = y_K$, and $H(\mathbb{G}_0 + e) = H(\mathbb{G}_0) + 1$ otherwise. Moreover, for the random choice of H , the first event $H(\mathbb{G}_0 + e) = H(\mathbb{G}_0)$ occurs with probability $1/2^K$. We conclude that

$$\mathbb{E} \left[H(\mathbb{G}(N, \lfloor cN \rfloor, r) \mid \mathbb{G}_0) - H(\mathbb{G}_0) \right] = 1 - \frac{1}{2^K} \left(\frac{|O^*|}{N} \right)^K,$$

$$\mathbb{P} \left(H(\mathbb{G}(N, \lfloor cN \rfloor, r-1)) = \lfloor cN \rfloor \mid H(\mathbb{G}_0) = \lfloor cN \rfloor - 1 \right) = 1 - \frac{1}{2^K} \left(\frac{|O^*|}{N} \right)^K.$$

Similarly,

$$\mathbb{E}[H(\mathbb{G}(N, \lfloor cN \rfloor, r-1) \mid \mathbb{G}_0, H_0) - H(\mathbb{G}_0)] = 1 - \frac{1}{2^K} \frac{N_1}{N} \left(\frac{|O^* \cap [N_1]|}{N_1} \right)^K - \frac{1}{2^K} \frac{N_2}{N} \left(\frac{|O^* \cap [N_2]|}{N_2} \right)^K,$$

$$\mathbb{P} \left(H(\mathbb{G}(N, \lfloor cN \rfloor, r-1)) = \lfloor cN \rfloor \mid H(\mathbb{G}_0) = \lfloor cN \rfloor - 1 \right) = 1 - \frac{1}{2^K} \frac{N_1}{N} \left(\frac{|O^* \cap [N_1]|}{N_1} \right)^K - \frac{1}{2^K} \frac{N_2}{N} \left(\frac{|O^* \cap [N_2]|}{N_2} \right)^K.$$

Using the convexity of the function x^K on $x \in [0, \infty)$, we obtain the result.

- **NAE-K-SAT.** The proof is very similar. The estimate $1/2^K$ changes to $2/2^K$, but the rest of the proof is the same.

We have established (14) and (15). With this, the proof of Proposition 1 is complete. \square

5 Proofs: Log-partition function in Erdős-Rényi graphs

The following property serves as an analogue of (5). Given two hypergraphs $\mathbb{G}_i = ([N], E_i)$, $i = 1, 2$ on the same set of nodes $[N]$ for each one of the six models and each finite λ

$$|\log Z(\mathbb{G}_1) - \log Z(\mathbb{G}_2)| = O(|E_1 \Delta E_2|). \quad (16)$$

This follows from the fact that adding (deleting) an hyperedge to (from) a graph results in multiplying or dividing the partition function by at most λ for all models except for the Ising model, where the corresponding value is λ^β .

The analogue of Theorem 5 is the following result.

Theorem 6. *For every $1 \leq N_1, N_2 \leq N - 1$ such that $N_1 + N_2 = N$ and every $\lambda > 1$*

$$\mathbb{E}[\log Z(\mathbb{G}(N, cN))] \geq \mathbb{E}[\log Z(\mathbb{G}(N_1, M_1))] + \mathbb{E}[\log Z(\mathbb{G}(N_2, M_2))] \quad (17)$$

where $M_j \stackrel{d}{=} \text{Bi}(cN, N_j/N)$, $j = 1, 2$.

As before, we do not assume independence of $M_j, j = 1, 2$. Let us first show how this result implies Theorem 2.

Proof of Theorem 2. Since M_j have Bernoulli distribution, using observation (16) and Theorem 2 we obtain

$$\mathbb{E}[\log Z(\mathbb{G}(N, \lfloor cN \rfloor))] \geq \mathbb{E}[\log Z(\mathbb{G}(N_1, \lfloor cN_1 \rfloor))] + \mathbb{E}[H(\mathbb{G}(N_2, \lfloor cN_2 \rfloor))] - O(\sqrt{N}). \quad (18)$$

Now we use Proposition 4 in the Appendix for the case $\alpha = 1/2$ to conclude that the limit

$$\lim_{N \rightarrow \infty} N^{-1} \mathbb{E}[\log Z(\mathbb{G}(N, \lfloor cN \rfloor))] \triangleq z(c)$$

exists. Showing that this also implies convergence of $N^{-1}H(\mathbb{G}(N, \lfloor cN \rfloor))$ to $H(c)$ w.h.p. again can be done using standard concentration results [JLR00] by applying property (16) and we skip the details. The proof of continuity and monotonicity of $z(c)$ is similar to the one of $H(c)$. \square

Thus it remains to prove Theorem 6.

Proof of Theorem 6. Construct an interpolating graph $\mathbb{G}(N, \lfloor cN \rfloor, r)$, $0 \leq r \leq \lfloor cN \rfloor$ exactly as in the previous subsection. We now establish the following analogue of Proposition 1. \square

Proposition 2. For every $r = 1, \dots, cN$,

$$\mathbb{E}[\log Z(\mathbb{G}(N, cN, r))] \geq \mathbb{E}[\log Z(\mathbb{G}(N, cN, r-1))]. \quad (19)$$

Let us first show how Theorem 6 follows from this proposition. Observe that for disjoint union of two graphs $\mathbb{G} = \mathbb{G}_1 + \mathbb{G}_2$, with $\mathbb{G} = (V, E)$, $\mathbb{G}_1 = (V_1, E_1)$, $\mathbb{G}_2 = (V_2, E_2)$, we always have $\log Z(\mathbb{G}) = \log Z(\mathbb{G}_1) + \log Z(\mathbb{G}_2)$. Theorem 6 then follows from Proposition 2.

Proof of Proposition 2. Recall that $\mathbb{G}(N, cN, r-1)$ is obtained from $\mathbb{G}(N, cN, r)$ by deleting a hyperedge chosen u.a.r. independently from r hyperedges e_1, \dots, e_r and adding a hyperedge e either to nodes $[N_1]$ or to $[N_2]$ with probabilities N_1/N and N_2/N respectively. Let as before \mathbb{G}_0 be the graph obtained after deleting but before adding a hyperedge, and let Z_0 and μ_0 be the corresponding partition function and Gibbs measure respectively. In the case of K-SAT and NAE-K-SAT models we assume that \mathbb{G}_0 encodes the realizations of the random potentials as well. We now show that conditional on any realization of the graph \mathbb{G}_0

$$\mathbb{E}[\log Z(\mathbb{G}(N, cN, r)) | \mathbb{G}_0] \geq \mathbb{E}[\log Z(\mathbb{G}(N, cN, r-1)) | \mathbb{G}_0]. \quad (20)$$

The proof of (20) is done on a case by case basis and it is very similar to the proof of (14).

• **Independent sets.** We have

$$\begin{aligned} \mathbb{E}[\log Z(\mathbb{G}(N, cN, r)) | \mathbb{G}_0] - \log Z_0 &= \mathbb{E} \left[\log \frac{Z(\mathbb{G}(N, cN, r))}{Z_0} \Big| \mathbb{G}_0 \right] \\ &= \mathbb{E} \left[\log \frac{\sum_I \lambda^{|I|} - \sum_I 1_{\{e \subset I\}} \lambda^{|I|}}{\sum_I \lambda^{|I|}} \Big| \mathbb{G}_0 \right] \\ &= \mathbb{E} \left[\log \left(1 - \mu_0(e \subset I_0) \right) \Big| \mathbb{G}_0 \right], \end{aligned}$$

where the sums \sum_I are over independent sets only and I_0 denotes an independent set chosen randomly according to μ_0 . Notice, that since we are conditioning on graph \mathbb{G}_0 the only randomness

underlying the expectation operator is the randomness of the hyperedge e . Note that $\mu_0(e \subset I_0) < 1$ since $\mu_0(e \not\subset I_0) \geq \mu_0(I_0 = \emptyset) > 0$. Using the expansion $\log(1 - x) = -\sum_{m \geq 1} x^m/m$

$$\begin{aligned}
\mathbb{E}[\log Z(\mathbb{G}(N, cN, r)) | \mathbb{G}_0] - \log Z_0 &= -\mathbb{E} \left[\sum_{k=1}^{\infty} \frac{\mu_0(e \subset I_0)^k}{k} \middle| \mathbb{G}_0 \right] \\
&= -\sum_{k=1}^{\infty} \frac{1}{k} \mathbb{E} \left[\sum_{I^1, \dots, I^k} 1_{\{e \subset \cap_{j=1}^k I^j\}} \frac{\lambda^{\sum_{j=1}^k |I^j|}}{Z_0^k} \middle| \mathbb{G}_0 \right] \\
&= -\sum_{k=1}^{\infty} \frac{1}{k} \sum_{I^1, \dots, I^k} \frac{\lambda^{\sum_{j=1}^k |I^j|}}{Z_0^k} \mathbb{E} \left[1_{\{e \subset \cap_{j=1}^k I^j\}} \middle| \mathbb{G}_0 \right] \\
&= -\sum_{k=1}^{\infty} \frac{1}{k} \sum_{I^1, \dots, I^k} \frac{\lambda^{\sum_{j=1}^k |I^j|}}{Z_0^k} \left(\frac{|\cap_{j=1}^k I^j|}{N} \right)^2,
\end{aligned}$$

where in the last equality we have used the fact that e is distributed u.a.r. Similar calculation for $\log Z(\mathbb{G}(N, cN, r-1))$ that is obtained by adding an hyperedge to $G_0 \cap [N_1]$ with probability N_1/N or to $G_0 \cap [N_2]$ with probability N_2/N gives

$$\begin{aligned}
&\mathbb{E}[\log Z(\mathbb{G}(N, cN, r-1)) | \mathbb{G}_0] - \log Z_0 \\
&= -\sum_{k=1}^{\infty} \frac{1}{k} \sum_{I^1, \dots, I^k} \frac{\lambda^{\sum_{j=1}^k |I^j|}}{Z_0^k} \left[\frac{N_1}{N} \left(\frac{|\cap_{j=1}^k I^j \cap [N_1]|}{N_1} \right)^2 + \frac{N_2}{N_2} \left(\frac{|\cap_{j=1}^k I^j \cap [N_2]|}{N} \right)^2 \right].
\end{aligned}$$

Again using the convexity of $f(x) = x^2$ we obtain

$$\mathbb{E}[\log Z(\mathbb{G}(N, cN, r)) | \mathbb{G}_0] - \log Z_0 \geq \mathbb{E}[\log Z(\mathbb{G}(N, cN, r-1)) | \mathbb{G}_0] - \log Z_0$$

and (20) is established.

- **MAX-CUT.** Similarly to the case of independent sets we have

$$\begin{aligned}
&\mathbb{E}[\log Z(\mathbb{G}(N, cN, r)) | \mathbb{G}_0] - \log Z_0 \\
&= \mathbb{E} \left[\log \frac{Z(\mathbb{G}(N, cN, r))}{Z_0} \middle| \mathbb{G}_0 \right] \\
&= \mathbb{E} \left[\log \frac{\sum_{x \in \{0,1\}^N} 1_{\{x_i = x_j\}} \lambda^{H(x)} + \lambda \sum_{x \in \{0,1\}^N} 1_{\{x_i \neq x_j\}} \lambda^{H(x)}}{\sum_{x \in \{0,1\}^N} \lambda^{H(x)}} \middle| \mathbb{G}_0 \right] \\
&= \log \lambda + \mathbb{E} \left[\log \left(1 - (1 - \frac{1}{\lambda}) \mu_0(x_i = x_j) \right) \middle| \mathbb{G}_0 \right].
\end{aligned}$$

Since $\lambda > 1$ we have $0 < (1 - \lambda^{-1}) \mu_0(x_i = x_j) < 1$ (this is where the condition $\lambda > 1$ is used) implying

$$\begin{aligned}
&\mathbb{E}[\log Z(\mathbb{G}(N, cN, r)) | \mathbb{G}_0] - \log Z_0 - \log \lambda \\
&= -\mathbb{E} \left[\sum_{k=1}^{\infty} \frac{(1 - \lambda^{-1})^k \mu_0(x_i = x_j)^k}{k} \middle| \mathbb{G}_0 \right] \\
&= -\sum_{k=1}^{\infty} \frac{(1 - \lambda^{-1})^k}{k} \mathbb{E} \left[\sum_{x_1, \dots, x_k \in \{0,1\}^N} \frac{\lambda^{\sum_{\ell=1}^k H(x_\ell)}}{Z_0^k} 1_{\{x_i^\ell = x_j^\ell, \forall \ell\}} \middle| \mathbb{G}_0 \right] \\
&= -\sum_{k=1}^{\infty} \frac{(1 - \lambda^{-1})^k}{k} \sum_{x_1, \dots, x_k \in \{0,1\}^N} \frac{\lambda^{\sum_{\ell=1}^k H(x_\ell)}}{Z_0^k} \mathbb{E} \left[1_{\{x_i^\ell = x_j^\ell, \forall \ell\}} \middle| \mathbb{G}_0 \right].
\end{aligned}$$

Now for every sequence x_1, \dots, x_k introduce equivalency classes on $[N]$. Given $i, k \in [N]$, say $i \sim k$ if $x_i^\ell = x_k^\ell, \forall \ell = 1, \dots, k$. Namely, in every one of the cuts defined by $x_\ell, \ell = 1, \dots, k$, the nodes i and k belong to the same group. Let $O_s, 1 \leq s \leq J$ be the corresponding equivalency classes. For an edge $e = (i, k)$ generated u.a.r., observe that $\mathbb{E} \left[1_{\{x_i = x_j \forall \ell\}} | \mathbb{G}_0 \right] = \sum_{s=1}^J \left(\frac{|O_s|}{N} \right)^2$. Thus

$$\mathbb{E} [\log Z(\mathbb{G}(N, cN, r)) | \mathbb{G}_0] - \log Z_0 - \log \lambda = - \sum_{k=1}^{\infty} \frac{(1 - \lambda^{-1})^k}{k} \sum_{x_1, \dots, x_k} \frac{\lambda^{\sum_{\ell=1}^k H(\ell)}}{Z_0^k} \sum_{s=1}^J \left(\frac{|O_s|}{N} \right)^2,$$

and similarly,

$$\begin{aligned} & \mathbb{E} [\log Z(\mathbb{G}(N, cN, r-1)) | \mathbb{G}_0] - \log Z_0 - \log \lambda \\ &= - \sum_{k=1}^{\infty} \frac{(1 - \lambda^{-1})^k}{k} \sum_{x_1, \dots, x_k} \frac{\lambda^{\sum_{\ell=1}^k H(\ell)}}{Z_0^k} \sum_{s=1}^J \left(\frac{N_1}{N} \left(\frac{|O_s \cap [N_1]|}{N_1} \right)^2 + \frac{N_2}{N} \left(\frac{|O_s \cap [N_2]|}{N_2} \right)^2 \right). \end{aligned}$$

Using the convexity of the function $f(x) = x^2$, we obtain (20).

- **Ising, Coloring, K-SAT and NAE-K-SAT.** The proofs of the remaining cases are obtained similarly and is omitted. The condition $\lambda > 1$ is used to assert positivity of $1 - \lambda^{-1}$ in the logarithm expansion. □

6 Proofs: Random regular graphs

Our result leading to the proof of Theorem 3 is as follows.

Theorem 7. *For every N_1, N_2 such that $N = N_1 + N_2$ and $N_1 r / K, N_2 r / K$ are integers,*

$$\mathbb{E}[H(\mathbb{G}(N, r))] \geq \mathbb{E}[H(\mathbb{G}(N_1, r))] + \mathbb{E}[H(\mathbb{G}(N_2, r))] - O(N^{\frac{5}{6}}).$$

Proof. Fix N_1, N_2 such that $N_1 + N_2 = N$ and $N_1 r / K, N_2 r / K$ are integers. Let us first prove Theorem 7 for the simple case $\min_{j=1,2} N_j < 40N^{\frac{5}{6}}$. In this case starting from the graph $\mathbb{G}(N, r)$ we can obtain a disjoint union of graphs $\mathbb{G}(N_j, r)$ via at most $O(N^{\frac{5}{6}})$ hyperedge deletion and addition operations. Indeed, suppose without loss of generality that $N_1 < 40N^{\frac{5}{6}}$. Delete all the hyperedges inside $[N_1]$ as well as all the hyperedges connecting two parts. Then generate a random graph $\mathbb{G}(N_1, r)$ from scratch. Finally, complete a so obtained partial matching in the configuration model on $[N_2 r]$ and project. The total number of deleted and generated hyperedges is $O(N^{\frac{5}{6}})$ and indeed we obtain a disjoint union of graphs $\mathbb{G}(N_j, r), j = 1, 2$. The hyperedge deletion and generation operation changes the value of H by at most $O(N^{\frac{5}{6}})$, and the proof of Theorem 3 is complete.

Thus through the remainder of the section we assume $\min_{j=1,2} N_j \geq 40N^{\frac{5}{6}}$. Fix

$$T = Nr/K - (1/K)N^{\frac{2}{3}}. \tag{21}$$

Let $\mathbb{G}(N, r, T)$ denote the graph obtained by creating a size T matching on Nr nodes of the configuration model uniformly at random and then projecting. For example if T was Nr/K , then we would have obtained the usual $\mathbb{G}(N, r)$. In the current situation we have exactly $N^{\frac{2}{3}}$ isolated nodes in the configuration model.

We now describe an interpolation procedure which interpolates between $\mathbb{G}(N, r, T)$ and a union of two graphs on nodes $[N_1]$ and $[N_2]$. For every integer partition $K = K_1 + K_2$ such that $K_1, K_2 \geq 1$ let

$T_{K_1, K_2} \leq T$ be the (random) number of hyperedges which connect parts $[N_1]$ and $[N_2]$ in $\mathbb{G}(N, r, T)$ and such that each connecting hyperedge has exactly K_j nodes in part $[N_j r]$ in the configuration model. Let $T_0 = \sum_{K_1, K_2 \geq 1: K_1 + K_2 = K} T_{K_1, K_2}$. Observe that $T_0 \leq \min_{j=1,2} (N_j r)$.

Define $\mathbb{G}(N, T, 0) = \mathbb{G}(N, r, T)$ and define $\mathbb{G}(N, T, t)$, $1 \leq t \leq T_{1, K-1}$ recursively as follows. Assuming $\mathbb{G}(N, T, t-1)$ is already defined, consider the graph \mathbb{G}_0 obtained from $\mathbb{G}(N, T, t-1)$ by deleting an hyperedge connecting $[N_1]$ and $[N_2]$ chosen uniformly at random from the collection of hyperedges which have exactly 1 node in part $[N_1 r]$ and $K-1$ nodes in part $[N_2 r]$ (from the remaining $T_{1, K-1} - (t-1)$ such hyperedges). Then we construct $\mathbb{G}(N, T, t)$ by adding an hyperedge to the resulting graph as follows: with probability $1/K$ an hyperedge is added to connect K isolated nodes chosen uniformly at random among the all isolated nodes from the set $[N_1 r]$. With the remaining probability $(K-1)/K$ an hyperedge is added to connect K isolated nodes chosen uniformly at random among the all isolated nodes from the set $[N_2 r]$. It is possible that at some point there are no K isolated nodes available in $[N_j r]$. In this case we say that the interpolation procedure fails. In fact we say that the interpolation procedure fails if in either of the two parts the number of isolated nodes is strictly less than K , even if the attempt was made to add a hyperedge to a part where there is no shortage of such nodes.

Thus we have defined an interpolation procedure for $t \leq T_{1, K-1}$. We now define it for $T_{1, K-1} + 1 \leq t \leq T_{2, K-2}$ analogously : we delete a randomly chosen hyperedge connecting two parts such that the hyperedge has 2 nodes in part $j=1$, and $K-2$ nodes in part $j=2$. Then we add an hyperedge uniformly at random to part $j=1, 2$ to connect K isolated nodes with probability $2/K$ and $(K-2)/K$ respectively. The failure of the interpolation is defined similarly as above. We continue this for all partitions (K_1, K_2) until $(K-1, 1)$, including. For the (K_1, K_2) phase of the interpolation procedure the probabilities are K_1/K and K_2/K respectively. We have defined the set of interpolating graphs $\mathbb{G}(N, T, t)$, $0 \leq t \leq T_0$.

Let \mathcal{I} denote the event that the interpolation procedure succeeds. For simplicity, if the interpolation procedure fails in some step t' we still define $\mathbb{G}(N, T, t)$, $t' \leq t \leq T_0$ to be the same graph as the first graph at which the interpolation procedure fails: $\mathbb{G}(N, T, t) = \mathbb{G}(N, T, t')$. It will be convenient to define $\mathbb{G}(N, T, t) = \mathbb{G}(N, T, T_0)$ for $T_0 \leq t \leq \min_{j=1,2} (N_j r)$, whether the interpolation procedure fails or not.

Provided that the interpolation procedure succeeds, the graph $\mathbb{G}(N, T, \min_{j=1,2} N_j)$ is a disjoint union of two graphs on $[N_j]$, $j=1, 2$ each ‘‘close’’ to being an r -regular random graph, in some appropriate sense to be made precise later.

Our next goal is establishing the following analogue of Proposition 1. As in previous sections, let \mathbb{G}_0 denote the graph obtained from $\mathbb{G}(N, T, t-1)$ after deleting an hyperedge connecting two parts, but before an hyperedge is added to one of the parts, namely, before creating $\mathbb{G}(N, T, t)$, conditioned on the event that the interpolation process survives till t . If, on the other hand the interpolation procedure fails before t , let \mathbb{G}_0 be the graph obtained at the last successful interpolation step after the hyperedge deletion.

Proposition 3. *For every $t \leq \min_j N_j$*

$$\mathbb{E}[H(\mathbb{G}(N, T, t-1))] \geq \mathbb{E}[H(\mathbb{G}(N, T, t))] - O\left(\mathbb{E} \max_{j=1,2} \frac{1}{Z_j(t)}\right), \quad (22)$$

where $Z_j(t)$ denotes the number of isolated nodes in the j -th part of the configuration model of \mathbb{G}_0 .

Proof. Let \mathcal{I}_t be the event that the interpolation succeeds for t steps. Notice that $\mathbb{E}[H(\mathbb{G}(N, T, t-1)) | \mathcal{I}_{t-1}^c] = \mathbb{E}[H(\mathbb{G}(N, T, t)) | \mathcal{I}_{t-1}^c]$, since the two graphs are identical, and thus the statement of the proposition holds.

Now we will condition on the event \mathcal{I}_t and let \mathbb{G}_0 denote as before the graph obtained after edge deletion and before hyperedge addition. Moreover, condition on the event that \mathbb{G}_0 was obtained in phase (K_1, K_2) , $K_1 + K_2 = K$. Specifically, after deleting an hyperedge to obtain \mathbb{G}_0 , extra K_j isolated nodes were created in part $j=1, 2$, and the number of isolated nodes is at least K in both parts of

(configuration model) \mathbb{G}_0 . Let Δ_i denotes the degree of the node $i \in [N]$ in the graph \mathbb{G}_0 . Observe that then $Z_j(t) = \sum_{i \in [N_j]} r - \Delta_i$. We now establish a stronger result. Namely,

$$\mathbb{E}[H(\mathbb{G}(N, T, t-1)) | \mathbb{G}_0] \geq \mathbb{E}[H(\mathbb{G}(N, T, t)) | \mathbb{G}_0] - O\left(\max_{j=1,2} \frac{1}{\sum_{i \in [N_j]} r - \Delta_i}\right). \quad (23)$$

Observe that conditioned on obtaining graph \mathbb{G}_0 , the graph $\mathbb{G}(N, T, t-1)$ can be recovered from \mathbb{G}_0 in distributional sense by adding an hyperedge connecting K_1 isolated node from $[N_1r]$ to K_2 isolated node from $[N_2r]$, both chosen uniformly at random, and then projecting.

We now conduct model dependent case by case analysis.

- **Independent sets.** In this case $K = 2$ and the only possibility is $K_1 = K_2 = 1$. As in the previous section, O^* again denote the set of nodes in $[N]$ which belong to *every* largest independent set in \mathbb{G}_0 . Then in the case of creating graph $\mathbb{G}(N, T, t-1)$ from \mathbb{G}_0 , the newly added hyperedge e decreases H by one if both ends of e belong to O^* , and leaves it the same otherwise. The first event occurs with probability

$$\frac{\sum_{i_1 \in O^* \cap [N_1], i_2 \in O^* \cap [N_2]} (r - \Delta_{i_1})(r - \Delta_{i_2})}{\sum_{i_1 \in [N_1], i_2 \in [N_2]} (r - \Delta_{i_1})(r - \Delta_{i_2})} = \frac{\sum_{i \in O^* \cap [N_1]} (r - \Delta_i)}{\sum_{i \in [N_1]} (r - \Delta_i)} \frac{\sum_{i \in O^* \cap [N_2]} (r - \Delta_i)}{\sum_{i \in [N_2]} (r - \Delta_i)}$$

We now analyze the case of creating $\mathbb{G}(N, T, t)$. Conditioning on the event that e was added to part $[N_jr]$, the value of H decreases by one iff both ends of e fall into $O^* \cap [N_j]$. This occurs with probability

$$\frac{\left(\sum_{i \in O^* \cap [N_j]} (r - \Delta_i)\right)^2 - \sum_{i \in O^* \cap [N_j]} (r - \Delta_i)}{\left(\sum_{i \in [N_j]} (r - \Delta_i)\right)^2 - \sum_{i \in [N_j]} (r - \Delta_i)} = \frac{\left(\sum_{i \in O^* \cap [N_j]} (r - \Delta_i)\right)^2}{\left(\sum_{i \in [N_j]} (r - \Delta_i)\right)^2} - O\left(\frac{1}{\sum_{i \in [N_j]} (r - \Delta_i)}\right).$$

Therefore, the value of H decreases by one with probability

$$\frac{1}{2} \sum_{j=1,2} \frac{\left(\sum_{i \in O^* \cap [N_j]} (r - \Delta_i)\right)^2}{\left(\sum_{i \in [N_j]} (r - \Delta_i)\right)^2} - O\left(\max_{j=1,2} \frac{1}{\sum_{i \in [N_j]} (r - \Delta_i)}\right),$$

and stays the same with the remaining probability. Using the inequality $(1/2)(x^2 + y^2) \geq xy$ we obtain (23).

- **MAX-CUT, Ising, Coloring.** As in the proof of Theorem 1 we introduce equivalence classes $O_j^* \subset [N], 1 \leq j \leq J$ on the graph \mathbb{G}_0 . The rest of the proof is almost identical to the one for the Independent Set model and we skip the details. Notice, that in all of these cases we have $K = 2$ and the interpolation phase has only one stage corresponding to $(K_1, K_2) = (1, 1)$.
- **K-SAT.** This is the first model for which $K > 2$. Suppose the graph \mathbb{G}_0 was created in stage (K_1, K_2) . As in the previous section let $O^* \subset \{0, 1\}^N$ denote the set of optimal assignments in the graph \mathbb{G}_0 . Reasoning as in the previous section, when we reconstruct graph $\mathbb{G}(N, T, t-1)$ in the distributional sense by adding a random hyperedge connecting K_1 nodes in $[N_1r]$ with K_2 nodes in $[N_2r]$, the probability that the value of H decreases by one is precisely

$$\frac{1}{2^K} \left[\frac{\sum_{i \in O^* \cap [N_1]} (r - \Delta_i)}{\sum_{i \in [N_1]} (r - \Delta_i)} \right]^{K_1} \left[\frac{\sum_{i \in O^* \cap [N_2]} (r - \Delta_i)}{\sum_{i \in [N_2]} (r - \Delta_i)} \right]^{K_2}.$$

Similarly, creating $\mathbb{G}(N, T, t)$ from \mathbb{G}_0 decreases the value of H by one with probability

$$\frac{1}{2^K} \frac{K_1}{K} \left[\frac{\sum_{i \in O^* \cap [N_1]} (r - \Delta_i)}{\sum_{i \in [N_1]} (r - \Delta_i)} \right]^K + \frac{1}{2^K} \frac{K_2}{K} \left[\frac{\sum_{i \in O^* \cap [N_2]} (r - \Delta_i)}{\sum_{i \in [N_2]} (r - \Delta_i)} \right]^K - O\left(\max_{j=1,2} \frac{1}{\sum_{i \in [N_j]} (r - \Delta_i)}\right).$$

Applying Young's inequality, namely that $ab \leq pa^{\frac{1}{p}} + qb^{\frac{1}{q}}$ for every $a, b \geq 0, p + q = 1, p, q > 0$, with the choice $p = K_1/K, q = K_2/K$,

$$a = \left[\frac{\sum_{i \in O^* \cap [N_1]} (r - \Delta_i)}{\sum_{i \in [N_1]} (r - \Delta_i)} \right]^{K_1},$$

$$b = \left[\frac{\sum_{i \in O^* \cap [N_2]} (r - \Delta_i)}{\sum_{i \in [N_2]} (r - \Delta_i)} \right]^{K_2},$$

and canceling $1/2^K$ on both sides, we obtain the result.

- **NAE-K-SAT.** The proof is identical to the one for **K-SAT**. The only difference is factor $1/2^{K-1}$ replacing $1/2^K$.

□

Our next step is to control the error term in (22).

Lemma 2. *The following holds*

$$\mathbb{E} \left[\sum_{1 \leq t \leq T_0} \max_{j=1,2} \frac{1}{Z_j(t)} \right] = O(N^{\frac{2}{5}}). \quad (24)$$

Proof. Since \mathbb{G}_0 is obtained after deleting one hyperedge connecting two parts, but before adding a new hyperedge, then $Z_j(t) \geq 1$. A crude bound on the required expression is then $\mathbb{E}[T_0] = O(\min N_j)$. We have $\mathbb{E}[Z_j(0)] = N_j/N^{\frac{1}{3}} \geq 40N^{\frac{1}{2}}$ since the initial number of isolated nodes was $Nr/2 - T = N^{\frac{2}{3}}$ and $\min_j N_j \geq 40N^{\frac{5}{6}}$. Moreover, $\mathbb{P}(Z_j(0) < (1/2)N_j/N^{\frac{1}{3}}) = O(\exp(-N^{\delta_1}))$ for some $\delta_1 > 0$. Observe that $Z_j(t+1) - Z_j(t) = 0$ with probability one if the interpolation procedure failed for some $t' \leq t$. Otherwise, if t corresponds to phase (K_1, K_2) then $Z_j(t+1) - Z_j(t)$ takes values $-K_j + K$ with probability K_j/K and $-K_j$ with the remaining probability. This is because during the hyperedge deletion step $Z_j(t)$ decreases by K_j and during the hyperedge addition step it increases by K or by zero with probabilities K_j/K and $1 - K_j/K$ respectively. In particular, $\mathbb{E}[Z_j(t+1) - Z_j(t)] = 0$. The decision of whether to put the hyperedge into part 1 or 2 are done independently. Since $t \leq T_0 \leq N_j$, we conclude that for each $t \leq T_0$ we have $\mathbb{P}(Z_j(0) - Z_j(t) > N_j^{3/5}) = O(\exp(-N^{\delta_2}))$ for some $\delta_2 > 0$. Here any choice of exponent strictly larger than $1/2$ applies, but for our purposes $3/5$ suffices. It follows that, $Z_j(t) \geq (1/2)N_j/N^{\frac{1}{3}} - N_j^{3/5}$ for all t with probability at least $1 - O(N_j \exp(-N^\delta)) = 1 - O(N \exp(-N^\delta))$ for $\delta = \min(\delta_1, \delta_2)$. Using this and $T_0 \leq \min_j(N_j r)$ we obtain that with at least this probability the expression inside the expectation on the left hand-side of (24) is at most

$$\frac{N_j r}{(1/2)N_j N^{-\frac{1}{3}} - N_j^{3/5}} = \frac{N_j^{\frac{2}{5}} r}{(1/2)N_j^{\frac{2}{5}} N^{-\frac{1}{3}} - 1}$$

The assumption $\min N_j \geq 40N^{\frac{5}{6}}$ guarantees that the denominator is at least 1. The numerator is at most $N^{\frac{2}{5}} r$. We conclude that the expression inside the expectation is at most $N^{\frac{2}{5}} r$ with probability at least $1 - O(N \exp(-N^\delta))$. Since we also have $T_0 \leq Nr$ w.p.1, then using a very crude estimate $O(N \exp(-N^\delta)) = O(N^{-\frac{3}{5}})$, and $NN^{-\frac{3}{5}} = N^{\frac{2}{5}}$, we obtain the required result. □

As a corollary of Proposition 3 and Lemma 2 we obtain

Corollary 2.

$$\mathbb{E}[H(\mathbb{G}(N, T, 0))] \geq \mathbb{E}[H(\mathbb{G}(N, T, T_0))] - O(N^{\frac{2}{5}}).$$

Let us consider graph $\mathbb{G}(N, T, T_0)$. We further modify it by removing all hyperedges which connect two parts $[N_j]$ of the graph, if there are any such hyperedges left. Notice that if the event \mathcal{I} occurs, namely the interpolation procedure succeeds, no further hyperedges need to be removed. The resulting graph is a disjoint union of graphs obtained on nodes $[N_1r]$ and $[N_2r]$ by adding a random size partial matching uniformly at random. The actual size of these two matchings depends on in the initial size of the partial matching within each part, and also on how many of T_0 hyperedges go into each part during the interpolation steps, and how many were removed in the final part (if any). We now bounds on the sizes of these matchings.

Recall $\min_j N_j \geq 40N^{\frac{5}{6}}$. We showed in the proof of Lemma 2 that the interpolation procedure succeeds with probability $O(N \exp(-N^\delta))$ for some δ . This coupled with the fact that w.p.1, the number of hyperedges removed in the final stage is at most N , gives us that the expected number of hyperedges removed in the final stage is at most $O(N^2 \exp(-N^\delta)) = O(N^{\frac{2}{3}})$. Moreover, since the initial number of isolated nodes was $N^{\frac{2}{3}}$ and during the interpolation procedure the total number of isolated nodes never increases, then the total number of isolated nodes before the final removal of hyperedges in $\mathbb{G}(N, T, T_0)$ is at most $N^{\frac{2}{3}}$. We conclude that the expected number of isolated nodes in the end of the interpolation procedure is $O(N^{\frac{2}{3}})$. Then we can complete uniform random partial matchings on $[N_jr]$ to full uniform random matchings by adding at most that many hyperedges in expectation. The objective value of H changes by at most that much as well. The same applies to $\mathbb{G}(N, r, T)$ - we can complete this graph to a full matching on Nr nodes by adding at most $N^{\frac{2}{3}}$ hyperedges, since there are at most that many isolated nodes in this graph.

Coupled with Corollary 2 we obtain

$$\mathbb{E}[H(\mathbb{G}(N, r))] \geq \mathbb{E}[H(\mathbb{G}(N_1, r))] + \mathbb{E}[H(\mathbb{G}(N_2, r))] - O(N^{\frac{2}{3}}),$$

for the case $\min_j N_j \geq 40N^{\frac{5}{6}}$. This completes the proof of Theorem 7. \square

Proof of Theorem 3. The existence of the limit $\lim_{N \rightarrow \infty, N \in r^{-1}K\mathbb{Z}_+} N^{-1} \mathbb{E}[H(\mathbb{G}(N, r))] = H(r)$ follows immediately from Theorem 7 and Proposition 4 from the Appendix. Then the convergence w.h.p.

$$\lim_{N \rightarrow \infty, N \in r^{-1}K\mathbb{Z}_+} N^{-1} H(\mathbb{G}(N, r)) = H(r),$$

follows once again using standard concentration results [JLR00].

The proof of the existence of the large deviations limit $\lim_{N \rightarrow \infty, N \in r^{-1}K\mathbb{Z}_+} N^{-1} \log p(N, r) = p(r)$ for K-SAT and NAE-K-SAT models uses the same interpolation process and the same proof as the one used in the case of Erdős-Rényi graph. \square

We see that the case of Coloring presents additional difficulties in proving the large deviations result in the case of the random regular graph - the preprocessing and postprocessing such as deletion of hyperedges spanning both parts for the case $\min N_j < 40N^{\frac{5}{6}}$ makes the notion of δ -unusual graph not relevant here, since the deleted/added hyperedges are not selected uniformly at random. We have little doubt, however, that the large deviations result applies to the case of Coloring as well, which we leave as an open problem.

The proof of Theorem 4 uses the same interpolation as the one above and the proof itself mimics the one for Theorem 2. For this reason, we omit the details.

Acknowledgements

Authors are grateful for the insightful discussions with Silvio Franz, Andrea Montanari, Lenka Zdeborová and Florent Krzakala. Authors also thank Microsoft Research New England for the hospitality and the inspiring atmosphere, in which this work began.

References

- [AF99] D. Achlioptas and E. Friedgut, *A sharp threshold for k -colorability*, Random Structures and Algorithms **14** (1999), no. 1, 63–70.
- [Ald] D. Aldous, *Some open problems*, <http://stat-www.berkeley.edu/users/aldous/Research/problems.ps>.
- [AldNew] ———, *Some open problems*, <http://www.stat.berkeley.edu/~aldous/Research/OP/index.html>.
- [AS03] D. Aldous and J. M. Steele, *The objective method: Probabilistic combinatorial optimization and local weak convergence*, Discrete Combinatorial Probability, H. Kesten Ed., Springer-Verlag, 2003.
- [BHH59] J. Beardwood, J. H. Halton, and J. M. Hammersley, *The shortest path through many points*, Proc. Camb. Philos. Society **55** (1959), 299–327.
- [Bol80] B. Bollobas, *A probabilistic proof of an asymptotic formula for the number of regular graphs*, European J. Combinatorics **1** (1980), 311–316.
- [Bol85] ———, *Random graphs*, Academic Press, Inc., 1985.
- [BR] B. Bollobás and O. Riordan, *Sparse graphs: metrics and random models*, <http://arxiv.org/abs/0812.2656>.
- [CGHS04] D. Coppersmith, D. Gamarnik, M. Hajiaghayi, and G. Sorkin, *Random MAXSAT, random MAXCUT, and their phase transitions*, Random Structures and Algorithms **24** (2004), no. 4, 502–545.
- [DM] A. Dembo and A. Montanari, *Ising models on locally tree-like graphs*, Ann. of Appl. Prob., To appear.
- [FL03] S. Franz and M. Leone, *Replica bounds for optimization problems and diluted spin systems*, Journal of Statistical Physics **111** (2003), no. 3/4, 535–564.
- [FLT03] S. Franz, M. Leone, and F. L. Toninelli, *Replica bounds for diluted non-Poissonian spin systems*, J. Phys. A: Math. Gen. **36** (2003), 10967 – 10985.
- [KM06] S. Kudekar and N. Macris, *Sharp bounds for map decoding of general irregular ldpc codes*, Proceedings of ISIT, 2006.
- [FM] S. Franz, E. Maneva, and A. Montanari, *Private communication*.
- [Fri99] E. Friedgut, *Sharp thresholds of graph properties, and the k -SAT problem*, J. Amer. Math. Soc. **4** (1999), 1017–1054.
- [Gal63] R. G. Gallager, *Low density parity-check codes*, MIT Press, 1963.
- [Gam04] D. Gamarnik, *Linear phase transition in random linear constraint satisfaction problems*, Probability Theory and Related Fields. **129** (2004), no. 3, 410–440.
- [GNS06] D. Gamarnik, T. Nowicki, and G. Swirszcz, *Maximum weight independent sets and matchings in sparse random graphs. Exact results using the local weak convergence method*, Random Structures and Algorithms **28** (2006), no. 1, 76–106.
- [Geo88] H. O. Georgii, *Gibbs measures and phase transitions*, de Gruyter Studies in Mathematics 9, Walter de Gruyter & Co., Berlin, 1988.

- [GT02] F. Guerra and F.L. Toninelli, *The thermodynamic limit of diluted spin glass models*, Commun. Math. Phys. **230** (2002) no. 1, 71–79; also available at <http://arxiv.org/abs/cond-mat/0204280>
- [HW65] J. M. Hammersley and D. J. A. Welsh, *First-passage percolation, subadditive processes, stochastic networks, and generalized renewal theory*, Bernoulli and Bayse and Laplace Anniversary Volume (J. Neyman and L. Le Cam eds.), 1965, pp. 61–110.
- [JLR00] S. Janson, T. Luczak, and A. Rucinski, *Random graphs*, John Wiley and Sons, Inc., 2000.
- [JT08] S. Janson and A. Thomason, *Dismantling sparse random graphs*, Combinatorics, Probability and Computing **17** (2008), 259 – 264.
- [Mon05] A. Montanari, *Tight bounds for LDPC and LDGM codes under map decoding*, IEEE Transactions on Information Theory **51** (2005), no. 9, 3221–3246.
- [PT04] D. Panchenko and M. Talagrand, *Bounds for diluted mean-fields spin glass models*, Probability Theory and Related Fields **130** (2004), 312–336.
- [Sim93] B. Simon, *The statistical mechanics of lattice gases, Vol. I*, Princeton Series in Physics, Princeton University Press, Princeton, NJ, 1993.
- [Wor99] N. Wormald, *Models of random regular graphs*, in J.D. Lamb and D.A. Preece, editors, Surveys in Combinatorics, volume 267 ed., pp. 239–298, Cambridge University Press, 1999.

Appendix. Modified super-additivity theorem

Proposition 4. *Given $\alpha \in (0, 1)$, suppose a non-negative sequence $a_N, N \geq 1$ satisfies*

$$a_N \geq a_{N_1} + a_{N_2} - O(N^\alpha). \quad (25)$$

for every N_1, N_2 s.t. $N = N_1 + N_2$. Then the limit $\lim_{N \rightarrow \infty} (a_N/N)$ exists.

Proof. It is convenient to define $a_N = a_{\lfloor N \rfloor}$ for every real not necessarily integer value $N \geq 1$. It is then straightforward to check that the property (25) holds when extended to reals as well (thanks to the correction term $O(N^\alpha)$). Let

$$a^* = \limsup_{N \rightarrow \infty} \frac{a_N}{N}.$$

Fix $\epsilon > 0$ and find k such that $1/k < \epsilon \leq 1/(k-1)$. Find $N_0 = N_0(\epsilon)$ such that

$$\begin{aligned} N_0^{-1} a_{N_0} &\geq a^* - \epsilon, \\ k^\alpha N_0^{\alpha-1} &< \epsilon. \end{aligned}$$

Clearly, such N_0 exists. Consider any $N \geq kN_0$. Find r such that $kN_02^r \leq N \leq kN_02^{r+1}$. Applying (25) iteratively with $N_1 = N_2 = N/2$ we obtain

$$\begin{aligned} a_N &\geq 2^r a_{\frac{N}{2^r}} - \sum_{0 \leq l \leq r-1} O\left(2^l \left(\frac{N}{2^l}\right)^\alpha\right) \\ &= 2^r a_{\frac{N}{2^r}} - O\left(2^{(1-\alpha)r} N^\alpha\right), \end{aligned}$$

Now let us find i such that $(k+i)N_0 \leq N/2^r \leq (k+i+1)N_0$. Note $i \leq k$. Again using (25) successively with N_0 for N_1 and $N/2^r, (N/2^r) - N_0, (N/2^r) - 2N_0, \dots$ for N_2 , we obtain

$$\begin{aligned} a_{\frac{N}{2^r}} &\geq (k+i)a_{N_0} - O\left(k\left(\frac{N}{2^r}\right)^\alpha\right) \\ &\geq (k+i)a_{N_0} - O\left(k\left(\frac{N}{2^r}\right)^\alpha\right). \end{aligned}$$

Combining, we obtain

$$\begin{aligned} a_N &\geq 2^r(k+i)a_{N_0} - O\left(2^{(1-\alpha)r}N^\alpha\right) - O\left(k2^{r(1-\alpha)}N^\alpha\right) \\ &= 2^r(k+i)a_{N_0} - O\left(k2^{r(1-\alpha)}N^\alpha\right). \end{aligned}$$

Then

$$\begin{aligned} \frac{a_N}{N} &\geq \frac{2^r(k+i)}{2^r(k+i+1)} \frac{a_{N_0}}{N_0} - O\left(k2^{r(1-\alpha)}N^{\alpha-1}\right) \\ &\geq \left(1 - \frac{1}{(k+i+1)}\right)(a^* - \epsilon) - O\left(k2^{r(1-\alpha)}N^{\alpha-1}\right) \\ &\geq (1 - \epsilon)(a^* - \epsilon) - O\left(k2^{r(1-\alpha)}N^{\alpha-1}\right), \end{aligned}$$

where $1/k < \epsilon$ is used in the last inequality. Now

$$k2^{r(1-\alpha)}N^{\alpha-1} \leq k2^{r(1-\alpha)}(k2^rN_0)^{\alpha-1} = k^\alpha N_0^{\alpha-1} < \epsilon,$$

again by the choice of N_0 . We have obtained

$$\frac{a_N}{N} \geq (1 - \epsilon)(a^* - \epsilon) - \epsilon,$$

for all $N \geq N_0k$. Since ϵ was arbitrary the proof is complete. \square