Algorithmic Barriers from Phase Transitions

Dimitris Achlioptas[1⋆] and Amin Coja-Oghlan[2⋆⋆]

[1] UC Santa Cruz, Santa Cruz, CA 95064, USA, optas@cs.ucsc.edu
[2] University of Edinburgh, UK acoghlan@inf.ed.ac.uk

Abstract. For many random Constraint Satisfaction Problems, by now, we have asymptotically tight estimates of the largest constraint density for which they have solutions. At the same time, all known polynomial-time algorithms for many of these problems already completely fail to find solutions at much smaller densities. For example, it is well-known that it is easy to color a random graph using twice as many colors as its chromatic number. Indeed, some of the simplest possible coloring algorithms already achieve this goal. Given the simplicity of those algorithms, one would expect there is a lot of room for improvement. Yet, to date, no algorithm is known that uses $(2-\epsilon)\chi$ colors, in spite of efforts by numerous researchers over the years.

In view of the remarkable resilience of this factor of 2 against every algorithm hurled at it, we believe it is natural to inquire into its origin. We do so by analyzing the evolution of the set of k-colorings of a random graph, viewed as a subset of $\{1,\ldots,k\}^n$, as edges are added. We prove that the factor of 2 corresponds in a precise mathematical sense to a phase transition in the geometry of this set. Roughly, the set of k-colorings looks like a giant ball for $k \geq 2\chi$, but like an error-correcting code for $k \leq (2-\epsilon)\chi$. We prove that a completely analogous phase transition also occurs both in random k-SAT and in random hypergraph 2-coloring. And that for each problem, its location corresponds precisely with the point were all known polynomial-time algorithms fail. To prove our results we develop a general technique that allows us to prove rigorously much of the celebrated 1-step Replica-Symmetry-Breaking hypothesis of statistical physics for random CSPs.

Key words: algorithms, random structures, constraint satisfaction problems, phase transitions

⋆ Supported by NSF CAREER award CCF-0546900 and an Alfred P. Sloan Fellowship.
⋆⋆ Supported by the Deutsche Forschungsgemeinschaft (DFG CO 646)

1 Introduction

For many random Constraint Satisfaction Problems (CSP), such as random graph coloring, random k-SAT, random Max k-SAT, and hypergraph 2-coloring, by now, we have asymptotically tight estimates for the largest constraint density for which typical instances have solutions (see [5]). At the same time, all known efficient algorithms for each problem fair very poorly, i.e., they stop finding solutions at constraint densities *much* lower than those for which we can prove that solutions exist. Adding insult to injury, the best known algorithm for each problem asymptotically fairs no better than certain extremely naive algorithms for the problem.

For example, it has been known for nearly twenty years [10] that the following very simple algorithm will find a satisfying assignment of a random k-CNF formula with $m = rn$ clauses for $r = O(2^k/k)$: if there is a unit clause satisfy it; otherwise assign a random value to a random unassigned variable. While it is known that random k-CNF remain satisfiable for $r = \Theta(2^k)$, no polynomial-time algorithm is known to find satisfying assignments for $r = (2^k/k) \cdot \omega(k)$ for some function $\omega(k) \to \infty$.

Similarly, for all $k \geq 3$, the following algorithm [18, 2] will k-color a random graph with average degree $d \leq k \ln k$: select a random vertex with fewest available colors left and assign it a random available color. While it is known that random graphs remains k-colorable for $d \sim 2\, k \ln k$, no polynomial-time algorithm is known that can k-color a random graph of average degree $(1 + \epsilon)k \ln k$ for some fixed $\epsilon > 0$ and arbitrarily large k. Equivalently, while it is trivial to color a random graph using twice as many colors as its chromatic number, no polynomial-time algorithm is known that can get by with $(2 - \epsilon)\chi$ colors, for some fixed $\epsilon > 0$.

Random k-SAT and random graph coloring are not alone. In fact, for nearly every random CSP of interest, the known results establish a completely analogous state of the art:

1. There is a trivial upper bound on the largest constraint density for which solutions exist.
2. There is a non-constructive proof, usually via the second moment method, that the bound from (1) is essentially tight, i.e., that solutions do exist for densities nearly as high as the trivial upper bound.
3. Some simple algorithm finds solutions up to a constraint density much below the one from (2).
4. No polynomial-time algorithm is known to succeed for a density asymptotically greater than that in (3).

In this paper we prove that this is not a coincidence. Namely, for random graph coloring, random k-SAT, and random hypergraph 2-coloring, we prove that the point where all known algorithms stop is precisely the point where the geometry of the space of solutions undergoes a dramatic change. This is known as a "dynamical" phase transition in statistical physics and our results establish rigorously for random CSPs a large part of the "1-step Replica Symmetry Breaking" hypothesis [20]. Roughly speaking, this hypothesis asserts that while the set of solutions for low densities looks like a giant ball, at some critical point this ball shatters into exponentially many pieces that are far apart from one another and separated by huge "energy barriers". Algorithms (even extremely simple ones) have no problem finding solutions in the "ball" regime, but no algorithm is known that can find solutions in the "error-correcting code" regime.

We believe that the presence of dynamical phase transitions in random CSPs is a very general phenomenon, whose qualitative characteristics should be problem-independent, i.e., *universal*. The fact that we can establish the exact same qualitative picture for a problem with binary constraints over k-ary variables (random graph k-coloring) and a problem with k-ary constraints over binary variables (hypergraph 2-colorability) certainly lends support to this notion. That said, we wish to emphasize that determining for each random CSP the location of its dynamical phase transition (as we do in this paper for the three problems mentioned, in order to show that the transition coincides with the demise of all known algorithms) requires non-trivial, problem-specific ideas and computations.

Perhaps the following is an intuitive model of how a dynamical phase transition comes about. In random graph coloring, rather than thinking of the number of available colors as fixed and the constraint density (number of edges) as increasing, imagine that we keep the constraint density fixed, but we keep decreasing the number of available colors. If we start with q available colors where $q \gg \chi$, it is reasonable to imagine that the set of valid q-colorings, viewed as a subset of $\{1, 2, \ldots, q\}^n$, has a nice "round" shape, the rounder the greater q is relative to χ. By the same token, when we restrict our attention to the set of those q-colorings that only use colors $\{1, 2, \ldots, q-1\}$, we are taking a "slice' of the set of q-colorings. With each slicing the connectivity of the set at hands deteriorates, until at some point the set shatters. For example, slicing the 2-dimensional unit sphere through the origin yields a circle, but slicing the circle, yields a pair of points.

We conclude the introduction with a few words about the technical foundation for our work. To prove the existence (and determine the location) of a dynamical phase transition one needs access to statistical properties of the uniform measure over solutions. A geometric way of thinking about this is as follows. Given a CSP instance, say a k-CNF formula with m clauses chosen uniformly at random, consider the function H on $\{0,1\}^n$ that assigns to each truth assignment the number of clauses it violates. In this manner, H defines a "landscape" in which satisfying assignments correspond to valleys at sea-level. Understanding statistical properties of the uniform measure over solutions amounts to understanding "the view" one enjoys from such a valley, a probabilistically formidable task. As we discuss in Section 4, we can establish the following: the number of solutions of a random CSP is sufficiently concentrated around its exponentially large expectation for the view from a random sea-level valley to be "the same" as the view from an "artificial" valley. That is, from the valley that results by first selecting a random $\sigma \in \{0,1\}^n$ and then forming a random k-CNF formula, also with m clauses, but now chosen uniformly among the clauses satisfied by σ, i.e., the view from the *planted* satisfying assignment in the planted model. This is a *much* easier view to understand and we believe that the "transfer" theorems we establish in this paper will significantly aid in the analysis of random CSPs.

2 Statement of Results

To present our results in a uniform manner we need to introduce some common notions. Let V be a set of n variables, all with the same domain D, and let C be an arbitrary set of constraints over the variables in V. A CSP instance is a subset of C. We let $\mathrm{dist}(\sigma, \tau)$ denote the Hamming distance between $\sigma, \tau \in D^n$ and we turn D^n into a graph by saying that σ, τ are adjacent if $\mathrm{dist}(\sigma, \tau) = 1$. For a given instance I, we let $H = H_I : D^n \to \mathbb{N}$ be the function counting the number of constraints of I violated by each $\sigma \in D^n$.

Definition 1. *The **height** of a path $\sigma_0, \sigma_1, \ldots, \sigma_t \in D^n$ is $\max_i H(\sigma_i)$. We say that $\sigma \in D^n$ is a solution of an instance I, if $H(\sigma) = 0$. We will denote by $\mathcal{S}(I)$ the set of all solutions of an instance I. The **clusters** of an instance I are the connected components of $\mathcal{S}(I)$. A **region** is a non-empty union of clusters.*

Remark 1. The term cluster comes from physics. Requiring $\mathrm{dist}(\sigma, \tau) = 1$ to say that σ, τ are adjacent is somewhat arbitrary (but conceptually simplest) and a number of our results hold if one replaces 1 with $o(n)$.

We will be interested in distributions of CSP instances as the number of variables n grows. The set $C = C_n$ will typically consist of all possible constraints of a certain type, e.g., the set of all $\binom{n}{k}$ possible hyperedges in the problem of 2-coloring random k-uniform hypegraphs. We let $I_{n,m}$ denote the set of all CSP instances with precisely m distinct constraints from C_n and we let $\mathcal{I}_{n,m}$ denote the uniform distribution on the set of all instances $I_{n,m}$. We will say that a sequence of events $\mathcal{E}_n$ holds *with high probability* (w.h.p.) if $\lim_{n\to\infty} \Pr[\mathcal{E}_n] = 1$ and *with uniformly positive probability* (w.u.p.p.) if $\liminf_{n\to\infty} \Pr[\mathcal{E}_n] > 0$. As per standard practice in the study of random structures, we will take the liberty of writing $\mathcal{I}_{n,m}$ to denote the underlying random variable and, thus, write things like "The probability that $\mathcal{S}(\mathcal{I}_{n,m})$..."

2.1 Shattering

Definition 2. *We say that the set of solutions of $\mathcal{I}_{n,m}$ **shatters** if there exist constants $\beta, \gamma, \zeta, \theta > 0$ such that w.h.p. $\mathcal{S}(\mathcal{I}_{n,m})$ can be partitioned into regions so that:*

1. *The number of regions is at least $e^{\beta n}$.*
2. *Each region contains at most an $e^{-\gamma n}$ fraction of all solutions.*
3. *The Hamming distance between any two regions is at least ζn.*
4. *Every path between vertices in distinct regions has height at least θn.*

Our first main result asserts that the space of solutions for random graph coloring, random k-SAT, and random hypergraph 2-colorability shatters and that this shattering occurs just above the largest density for which any polynomial-time algorithm is known to find solutions for the corresponding problem. Moreover, we prove that the space remains shattered until, essentially, the CSP's satisfiability threshold. More precisely:

– A random graph with average degree d, i.e., $m = dn/2$, is w.h.p. k-colorable for $d \leq (2 - \gamma_k)k \ln k$, where $\gamma_k \to 0$. The best poly-time k-coloring algorithm w.h.p. fails for $d \geq (1 + \delta_k)k \ln k$, where $\delta_k \to 0$.

Theorem 1. *There exists a sequence $\epsilon_k \to 0$, such that the space of k-colorings of a random graph with average degree d shatters for all*

$$(1 + \epsilon_k)k \ln k \leq d \leq (2 - \epsilon_k)k \ln k \ . \tag{1}$$

– A random k-CNF formula with n variables and rn clauses is w.h.p. satisfiable for $r \leq 2^k \ln 2 - k$. The best poly-time satisfiability algorithm w.h.p. fails for $r > 2^{k+1}/k$. In [23], non-rigorous, but mathematically sophisticated evidence is given that a different algorithm succeeds for $r = \Theta((2^k/k) \ln k)$, but not higher.

Theorem 2. *There exists a sequence $\epsilon_k \to 0$ such that the space of satisfying assignments of a random k-CNF formula with rn clauses shatters for all*

$$(1 + \epsilon_k) \frac{2^k}{k} \ln k \leq r \leq (1 - \epsilon_k)2^k \ln 2 \ . \tag{2}$$

– A random k-uniform hypergraph with n variables and rn edges is w.h.p. 2-colorable for $r \leq 2^{k-1} \ln 2 - \frac{3}{2}$. The best poly-time 2-coloring algorithm w.h.p. fails for $r > 2^k/k$. In [23], non-rigorous, but mathematically sophisticated evidence is given that a different algorithm succeeds for $r = \Theta((2^k/k) \ln k)$, but not higher.

Theorem 3. *There exists a sequence $\epsilon_k \to 0$ such that the space of 2-colorings of a random k-uniform hypergraph with rn edges shatters for all*

$$(1 + \epsilon_k) \frac{2^{k-1}}{k} \ln k \leq r \leq (1 - \epsilon_k)2^{k-1} \ln 2 \ . \tag{3}$$

Remark 2. As the notation in Theorems 1,2,3 is asymptotic in k, the stated intervals may be empty for small values of k. In this extended abstract we have not optimized the proofs to deliver the smallest values of k for which the intervals are non-empty. Quick calculations suggest $k \geq 6$ for hypergraph 2-colorability, $k \geq 8$ for k-SAT, and $k \geq 20$ for k-coloring.

2.2 Rigidity

The regions mentioned in Theorems 1, 2 and 3 can be thought of as forming an error-correcting code in the solution-space of each problem. To make this precise we need to introduce the following definition and formalize the notion of "a random solution of a random instance".

Definition 3. *Given an instance I, a solution $\sigma \in \mathcal{S}(I)$ and a variable $v \in V$, we say that v in (I, σ):*

- *Is $f(n)$-**rigid**, if every $\tau \in \mathcal{S}(I)$ such that $\tau(v) \neq \sigma(v)$ has dist$(\sigma, \tau) \geq f(n)$.*
- *Is $f(n)$-**loose**, if for every $j \in D$, there exists $\tau \in \mathcal{S}(I)$ such that $\tau(v) = j$ and dist$(\sigma, \tau) \leq f(n)$.*

We will prove that while before the phase transition, in a typical solution, every variable is loose, after the phase transition nearly every variable is rigid. To formalize the notion of a random/typical solution, recall that $I_{n,m}$ denotes the set of all instances with m constraints over n variables and let $\Lambda = \Lambda_{n,m}$ denote the set of all instance–solution pairs, i.e., $\Lambda_{n,m} = \{(I, \sigma) : I \in I_{n,m}, \sigma \in \mathcal{S}(I)\}$. We let $\mathcal{U} = \mathcal{U}_{n,m}$ be the probability distribution induced on $\Lambda_{n,m}$ by the following:

Choose an instance $I \in I_{n,m}$ uniformly at random.
If $\mathcal{S}(I) \neq \emptyset$, select $\sigma \in \mathcal{S}(I)$ uniformly at random.

We will refer to instance-solution pairs generated according to $\mathcal{U}_{n,m}$ as **uniform** instance-solution pairs. We note that although the definition of uniform pairs allows for $\mathcal{S}(I)$ to be typically empty, i.e., to be in the typically unsatisfiable regime, we will only employ the definition for constraint densities such that w.h.p. $\mathcal{S}(I)$ contains exponentially many solutions. Hence, our liberty in also using the term a "typical" solution.

Theorem 4. *Let (I, σ) be a uniform instance-solution pair where:*

- *I is a graph with $dn/2$ edges, where d is as in* (1)*, and σ is a k-coloring of I, or,*
- *I is a k-CNF formula with rn clauses, where r is as in* (2)*, and σ is a satisfying assignment of I, or,*
- *I is a k-uniform hypergraph with rn edges, where r is as in* (3)*, and σ is a 2-coloring of I.*

W.h.p. the number of rigid variables in (I, σ) is at least $\gamma_k n$, for some sequence $\gamma_k \to 1$.

Remark 3. Theorem 4 is tight since for every finite constraint density, a random instance w.h.p. has $\Omega(n)$ variables that are not bound by any constraint.

The picture drawn by Theorem 4, whereby nearly all variables are rigid in typical solutions above the dynamical phase transition, is in sharp contrast with our results for densities below the transition for graph coloring and hypergraph 2-colorability. While we believe that an analogous picture holds for k-SAT, see Conjecture 1, for technical reasons we cannot establish this presently. (We discuss the additional difficulties imposed by random k-SAT in Section 4.)

Theorem 5. *Let (I, σ) be a uniform instance-solution pair where:*

- *I is a graph with $dn/2$ edges, where $d \leq (1 - \epsilon_k)k \ln k$, and σ is a k-coloring of I, or,*
- *I is a k-uniform hypergraph with rn edges, where $r \leq (1 - \epsilon_k)(2^{k-1}/k) \ln k$, and σ is a 2-coloring of I.*

There exists a sequence $\epsilon_k \to 0$ such that w.h.p. every variable in (I, σ) is $o(n)$-loose.

We note that in fact, for all d and r as in Theorem 5, w.u.p.p. (I, σ) is such that changing the color of any vertex to any color only requires changing the color of $O(\log n)$ other vertices.

Conjecture 1. Let (I, σ) be a uniform instance-solution pair where I is a k-CNF formula with rn clauses, where $r \leq (1 - \epsilon_k)(2^k/k) \ln k$, and σ is a satisfying assignment of I. There exists a sequence $\epsilon_k \to 0$ such that w.h.p. every variable in (I, σ) is $o(n)$-loose.

3 Background and Related Work

3.1 Algorithms

Attempts for a "quick improvement" upon either of the naive algorithms mentioned in the introduction for satisfiability/graph coloring, stumble upon the following general fact. Given a CSP instance, consider the bipartite graph in which every variable is adjacent to precisely those constraints in which it appears, known as the factor graph of the instance. For random formulas/graphs, factor graphs are locally tree-like, i.e., for any arbitrarily large constant D, the depth-D neighborhood of a random vertex is a tree w.h.p. In other words, locally, random CSPs are trivial, e.g., random graphs of any finite average degree are locally 2-colorable. Moreover, as the constraint density is increased, the factor graphs of random CSPs get closer and closer to being biregular, so that degree information is not useful either. Combined, these two facts render all known algorithms impotent, i.e., as the density is increased, their asymptotic performance matches that of trivial algorithms.

In [22], Mézard, Parisi, and Zecchina proposed a new satisfiability algorithm called Survey Propagation (SP) which performs extremely well experimentally on instances of random 3-SAT. This was very surprising at the time and allowed for optimism that, perhaps, random k-SAT instances might not be so hard. Moreover, SP was generalized to other problems, e.g., k-coloring [9] and Max k-SAT [8]. An experimental evaluation of SP for values of k even as small as 5 or 6 is already somewhat problematic, but to the extent it is reliable it strongly suggests that SP does not find solutions for densities as high as those for which solutions are known to exist. Perhaps more importantly, it can be shown that for densities at least as high as $2^k \ln 2 - k$, if SP can succeed at its main task (approximating the marginal probability distribution of the variables with respect to the uniform measure over satisfying assignments), so can a much simpler algorithm, namely Belief Propagation (BP), i.e., dynamic programming on trees.

The trouble is that to use either BP or SP to find satisfying assignments one sets variables iteratively. So, even if it is possible to compute approximately correct marginals at the beginning of the execution (for the entire formula), this can stop being the case after some variables are set. Concretely, in [23], Montanari et al. showed that (even within the relatively generous assumptions of statistical physics computations) the following Gibbs-sampling algorithm fails above the $(2^k/k) \ln k$ barrier, i.e., step 2 below fails to converge after only a small fraction of all variables have been assigned a value:

1. Select a variable v at random.
2. Compute the marginal distribution of v using Belief Propagation.
3. Set v to $\{0, 1\}$ according to the computed marginal distribution; simplify the formula; go to step 1.

3.2 Relating the Uniform and the Planted Model.

The idea of deterministically embedding a property inside a random structure is very old and, in general, the process of doing this is referred to as "planting" the property. In our case, we plant a solution σ in a random CSP, by only including constraints compatible with σ. Juels and Peinado [19] were perhaps the first to explore the relationship between the planted and the uniform model and they did so for the clique problem in dense random graphs $G_{n,1/2}$, i.e., where each edge appears independently with probability 1/2. They showed the distribution resulting from first choosing $G = G_{n,1/2}$ and then planting a clique of size $(1 + \varepsilon) \log_2 n$ is very close to $G_{n,1/2}$ and suggested this as a scheme to obtain a one-way-function. Since the planted clique has size only $(1 + \varepsilon) \log_2 n$, the basic argument in [19] is closely related to subgraph counting. In contrast, the objects under consideration in our work (k-colorings, satisfying assignments, etc.)

have an immediate impact on the *global* structure of the combinatorial object being considered, rather than just being local features, such as a clique on $O(\log n)$ vertices.

Coja-Oghlan, Krivelevich, and Vilenchik [12, 13] proved that for constraint densities well above the threshold for the existence of solutions, the planted model for k-coloring and k-SAT is equivalent to the uniform distribution *conditional* on the (exponentially unlikely) existence of at least one solution. In this conditional distribution as well as in the high-density planted model, the geometry of the solution space is very simple, as there is precisely one cluster of solutions.

3.3 Solution-space Geometry

In [7, 21] the first steps were made towards understanding the solution-space geometry of random k-CNF formulas by proving the existence of shattering and the presence of rigid variables for $r = \Theta(2^k)$. This was a far cry from the true $r \sim (2^k/k)\ln k$ threshold for the onset of both phenomena, as we establish here. Besides the quantitative aspect, there is also a fundamentally important difference in the methods employed in [7, 21] vs. those employed here. In those works, properties were established by taking a union bound over all satisfying assignments. It is not hard to show that the derived results are best possible using those methods and, in fact, there is good reason to believe that the results are genuinely tight, i.e., that for densities $o(2^k)$ the derived properties simply do not hold for *all* satisfying assignments. Here, we instead establish a systematic connection between the planted model and the process of sampling a random solution of a random instance. This argument allows us to analyze "typical" solutions while allowing for the possibility that a (relatively small, though exponential) number of "atypical" solutions exist. Therefore, we are for the first time in a position to analyze the extremely complex energy landscape of below-threshold instances of random CSPs, and to estimate quantities that appeared completely out of reach prior to this work.

4 Our Point of Departure: Symmetry, Randomness and Inversion

As mentioned, the results in this paper are enabled by a set of technical lemmas that allow one to reduce the study of "random solutions of random CSP instances" to the study of "planted CSP solutions". The conceptual origin of these lemmas can be traced to the following humble observation.

Let M be an arbitrary 0-1 matrix with the property that all its rows have the same number of 1s and all its columns have the same the number of 1s. A moment's reflection makes it clear that for such a matrix, both of the following methods select a uniformly random 1 from the entire matrix:

1. Select a uniformly random column and then a uniformly random 1 in that column.
2. Select a uniformly random row and then a uniformly random 1 in that row.

An example of how we employ this fact for random CSPs is as follows. Let $\mathcal{F}$ be the set of all k-CNF formulas with n variables and m distinct clauses (chosen among all $2^k\binom{n}{k}$ possible k-clauses). Say that $\sigma \in \{0,1\}^n$ NAE-satisfies a formula $F \in \mathcal{F}$ if under σ, every clause of F has at least one satisfied and at least one falsified literal. Let M be the $2^n \times |\mathcal{F}|$ matrix where$M_{\sigma,F} = 1$ iff $\sigma \in \{0,1\}^n$ NAE-satisfies F. By the symmetry of $\mathcal{F}$, it is clear that all rows of M have the same number of 1s. Imagine, for a moment, that the same was true for all columns. Then, a uniformly random solution of a uniformly random instance would be distributed *exactly* as a "planted" instance-solution pair: first select $\sigma \in \{0,1\}^n$ uniformly at random; then select m distinct clauses uniformly at random among all $2^{k-1}\binom{n}{k}$ clauses NAE-satisfied by σ.

Our contribution begins with the realization that exact row- and column-balance is not necessary. Rather, it is enough for the 1s in M to be "well-spread". More precisely, it is enough that the marginal distributions

induced on the rows and columns of M by selecting a uniformly random 1 from the entire matrix are both "reasonably close to" uniform. For example, assume we can prove that $\Omega(|\mathcal{F}|)$ columns of M have $\Theta(f(n))$ 1s, where $f(n)$ is the average number of 1s per column. Indeed, this is precisely the kind of property implied by the success of the second moment method for random NAE-k-SAT [3]. Under this assumption, proving that a property holds w.u.p.p. for a uniformly random solution of a uniformly random instance, reduces to proving that it holds w.h.p. for the planted solution of a planted instance, a dramatically simpler task.

There is a geometric intuition behind our transfer theorems which is more conveniently described when every constraint is included independently with the same probability p, i.e., we take $p = m/\left(2^k\binom{n}{k}\right)$. For all $k \geq 3$ and $m = rn$, it was shown in [3] that the resulting instances w.u.p.p. have exponentially many solutions for $r \leq 2^{k-1}\ln 2 - 3/2$. Consider now the following way of generating *planted* NAE k-SAT instances. First, select a formula F by including each clause with probability p, exactly as above. Then, select $\sigma \in \{0,1\}^n$ uniformly at random and remove from F all constraints violated by σ. Call the resulting instance F'. Our results say that as long as $q \equiv r(1-2^{-k+1}) \leq 2^{k-1}\ln 2 - 3/2$, the instance F' is "nearly indistinguishable" from a *uniform* instance created by including each clause with probability q. (We will make this statement precise shortly.)

To see how this happens, recall the function $H : \sigma \to \mathbb{N}$ counting the number of violated constraints under each assignment. Clearly, selecting F specifies such a function H_F, while selecting $\sigma \in \{0,1\}^n$ and removing all constraints violated by σ amounts to modifying H_F so that $H_F(\sigma) = 0$. One can imagine that such a modification creates a gradient in the vicinity of σ, a "crater" with σ at its bottom. What we prove is that as long as H_F already had an exponential number of craters and the number of craters is concentrated, adding one more crater does not make a big difference. Of course, if the density is increased further, the opened crater becomes increasingly obvious, as it takes a larger and larger cone to get from the typical values of H_F down to 0. Hence the ease with which algorithms solve planted instances of high density.

To prove our transfer theorems we instantiate this idea for random graph k-coloring, random k-uniform hypergraph 2-coloring, and random k-SAT. For this, a crucial step is deriving a lower bound on the number of solutions of a random instance. For example, in the case of random graph k-coloring, we prove that the number of k-colorings, $|\mathcal{S}(I_{n,m})|$, for a random graph with n vertices and m edges is "concentrated" around its expectation in the sense that w.h.p.

$$n^{-1}\ \left|\ln|\mathcal{S}(I_{n,m})| - \ln\mathbf{E}\left(|\mathcal{S}(I_{n,m})|\right)\right| = o(1)\ . \tag{4}$$

To prove this, we use the upper bound on the second moment $\mathbf{E}\left[|\mathcal{S}(I_{n,m})|^2\right]$ from [4] to show that w.u.p.p. $|\mathcal{S}(I_{n,m})| = \Omega(\mathbf{E}|\mathcal{S}(I_{n,m})|)$. Then, we perform a sharp threshold analysis, using theorems of Friedgut [15], to prove that (4) holds, in fact, with *high* probability. A similar approach applies to hypegraph 2-coloring.

The situation for random k-SAT is more involved. Indeed, we can prove that the number of satisfying assignments is *not* concentrated around its expectation in the sense of (4). This problem is mirrored by the fact that the second moment of the number of satisfying assignments exceeds the square of the first moment by an exponential factor (for any constraint density). Nonetheless, letting $F_k(n,m)$ denote a uniformly random k-CNF formula with n variables and m clauses, combining techniques from [6] with a sharp threshold analysis, we can derive a lower bound on the number of satisfying assignments that holds w.h.p., namely $n^{-1}\ln|\mathcal{S}(F_k(n,m))| \geq n^{-1}\ln\mathbf{E}|\mathcal{S}(F_k(n,m))| - \phi(k)$, where $\phi(k) \to 0$ exponentially with k. This estimates allows us to approximate the uniform model by the planted model sufficiently well in order to establish Theorems 2 and 4.

5 Proof sketches

Due to the space constraints, in the remaining pages we give proof sketches of our results for k-coloring, to offer a feel of the transfer theorems and of the style of the arguments one has to employ given those theorems (actual proofs appear in the Appendix). The proofs for hypergraph 2-coloring are relatively similar, as it is also a "symmetric" CSP and the second moment methods works on its number of solutions. For k-SAT, though, a significant amount of additional work is needed, as properties must be established with exponentially small error probability to overcome the large deviations in the number of satisfying assignments (proofs appear in the Appendix).

5.1 Transfer Theorem for Random Graph Coloring

We consider a fixed number $\varepsilon > 0$ and assume that $k \geq k_0$ for some sufficiently large $k_0 = k_0(\varepsilon)$. We denote $\{1, \ldots, k\}$ as $[k]$. We are interested in the probability distribution $\mathcal{U}_{n,m}$ on $\Lambda_{n,m}$ resulting from first choosing a random graph $G = G(n, m)$ and then a random k-coloring of G (if one exists). To analyze this distribution, we consider the distribution $\mathcal{P}_{n,m}$ on $\Lambda_{n,m}$ induced by following expermient.

P1. Generate a uniformly random k-partition $\sigma \in [k]^n$.
P2. Generate a graph G with m edges chosen uniformly at random among the edges bicolored under σ.
P3. Output the pair (G, σ).

The distribution $\mathcal{P}_{n,m}$ is known as the *planted model*.

Theorem 6. *Suppose that $d = 2m/n \leq (2 - \varepsilon)k \ln k$. There exists a function $f(n) = o(n)$ such that the following is true. Let $\mathcal{D}$ be any graph property such that $G(n, m)$ has $\mathcal{D}$ with probability $1 - o(1)$, and let $\mathcal{E}$ be any property of pairs $(G, \sigma) \in \Lambda_{n,m}$. If for all sufficiently large n*

$$\Pr_{\mathcal{P}_{n,m}}[(G, \sigma) \text{ has } \mathcal{E} | G \text{ has } \mathcal{D}] \geq 1 - \exp(-f(n)), \tag{5}$$

then $\Pr_{\mathcal{U}_{n,m}}[(G, \sigma) \text{ has } \mathcal{E}] = 1 - o(1)$.

5.2 Loose Variables Below the Transition

Suppose that $d \leq (1 - \varepsilon)k \ln k$. Recall that a graph with vertex set V is said to be ζ-choosable if for any assignments of color lists of length at least ζ to the elements of V, there is a proper coloring in which every vertex receives a color from its list. To prove Theorem 5, we consider the property $\mathcal{E}$ that all vertices are $o(n)$-loose and the following condition $\mathcal{D}$:

For any set $S \subset V$ of size $|S| \leq g(n)$ the subgraph induced on S is 4-choosable.

Here $g(n)$ is some function such that $f(n) \ll g(n) = o(n)$, where $f(n)$ is the function from Theorem 6. A standard argument shows that a random graph $G(n, m)$, where $m = O(n)$, satisfies $\mathcal{D}$ w.h.p.

By Theorem 6, we are thus left to establish (5). Let $\sigma \in [k]^n$ be a uniformly random k-partition, and let G be a random graph with m edges such that σ is a k-coloring of G. Since σ is uniformly random, we may assume that the color classes $V_i = \sigma^{-1}(i)$ satisfy $|V_i| \sim n/k$. Let $v_0 \in V$ be any vertex, and let $l \neq \sigma(v_0)$ be the "target color" for v_0. Our goal is to find a coloring τ such that $\tau(v_0) = l$ and $\text{dist}(\sigma, \tau) \leq g(n)$.

If v has no neighbor in V_l, then we can just assign this color to v_0. Otherwise, we run the following process. In the course of the process, every vertex is either *awake*, *dead*, or *asleep*. Initially, all the neighbors of v_0 in V_l are awake, v is dead, and all other vertices are asleep. In each step of the process, pick an awake

vertex w arbitrarily and declare it dead (if there is no awake vertex, terminate the process). If there are at least five colors $c_1(w), \dots, c_5(w)$ available such that w has no neighbor in $V_{c_i(w)}$, then we do nothing. Otherwise, we pick five colors $c_1(w), \dots, c_5(w)$ randomly and declare all asleep neighbors of w in $V_{c_j(w)}$ awake for $1 \leq j \leq 5$.

Lemma 1. *With probability at least $1 - \exp(-f(n))$ there are at most $g(n)$ dead vertices when the process terminates.*

The proof of Lemma 1 is based on relating our process to a subcritical branching process. The basic insight here is that when $d < (1-\varepsilon)k \ln k$ it is very likely that a vertex w has five immediately available colors. More precisely, for any w the number of neighbors in any class V_i with $i \neq \sigma(w)$ is asymptotically Poisson with mean $(1+o(1))\frac{2m}{(k-1)n} \leq (1-\varepsilon+o(1))\frac{k \ln k}{k-1}$. Hence, the probability that w does *not* have a neighbor in V_i is about $k^{\varepsilon-1}$. As there are k colors in total, we expect about $(k-1)^{\varepsilon}$ colors available for w, i.e., a lot.

To obtain a new coloring τ in which v_0 takes color l we consider the set D of all dead vertices. We let $\tau(u) = \sigma(u)$ for all $u \in V \setminus D$. Moreover, conditioning on the event $\mathcal{D}$, we can assign to each $w \in D$ a color from the list $\{c_1(w), \dots, c_5(w)\} \setminus \{l\}$. Thus, the new coloring τ differs from σ on at most $|D| \leq g(n) = o(n)$ vertices.

5.3 Rigid Variables Above the Transition

Suppose that $d \geq (1+\varepsilon)k \ln k$. To prove Theorem 4 for coloring we apply Theorem 7 as follows. We let $\alpha, \beta > 0$ be sufficiently small numbers and denote by $\mathcal{E}$ the following property of a pair $(G, \sigma) \in \Lambda_{n,m}$:

> There is a subgraph $G_* \subset G$ of size $|V(G_*)| \geq (1-\alpha)n$ such that for every vertex v of G_* and each color $i \neq \sigma(v)$ there are at least $\beta \ln k$ vertices vertex w in G_* that are adjacent to v such that $\sigma(w) = i$. (6)

Also, we let $\mathcal{D}$ be the property that the maximum degree is at most $(\ln n)^2$.

We shall prove that for a pair (G, σ) chosen from $\mathcal{U}_{n,m}$ a subgraph G_* as in (6) exists w.h.p. If that is so, then every vertex in G_* has at least one neighbor in every color class other than its own. Therefore, it is impossible to just assign a different color to any vertex in G_*. In fact, since all vertices in G_* have *a lot* (namely, at least $\beta \ln k$) of neighbors with every other color, the expansion properties of the random graph $G(n,m)$ imply that recoloring any vertex v in G_* necessitates the recoloring of at least $n/(k \ln k)$ further vertices. Loosely speaking, the conflicts resulting from recoloring v spread so rapidly that we necessarily end up recoloring a huge number of vertices. Thus, all vertices in G_* are $n/(k \ln k)$-rigid. Note that we can not hope for much better, as we can always recolor v by swapping two color classes, i.e., $\sim 2n/k$ vertices.

To prove the existence of the subgraph G_*, we establish the following.

Lemma 2. *Condition (5) holds for $\mathcal{D}$ and $\mathcal{E}$ as above.*

To obtain Lemma 2, let $(G, \sigma) \in \Lambda_{n,m}$ be a random pair chosen from the distribution $\mathcal{P}_{n,m}$. We may assume that $|\sigma^{-1}(i)| \sim n/k$ for all i. To obtain the graph G_*, we perform a "stripping process". As a first step, we obtain a subgraph H by removing from G all vertices that have fewer than $\gamma \ln k$ neighbors in any color class other than their own. If $\gamma = \gamma(\varepsilon)$ is sufficiently small, then the expected number of vertices removed in this way is less than $nk^{-\delta}$ for a $\delta > 0$, because for each vertex w the expected number of neighbors in another color class is bigger than $(1+\varepsilon)\ln k$. Then, we keep removing vertices from H that have "a lot" of neighbors outside of H. Given the event $\mathcal{D}$, we then show that with probabiltiy $1 - \exp(-\Omega(n))$ the final result of this process is a subgraph G_* that satisfies (6).

5.4 Proof of Theorem 1

Theorem 1 concerns the "view" from a random coloring σ of $G(n,m)$. Basically, our goal is to show that only a tiny fraction of all possible colorings are "visible" from σ, i.e., σ lives in a small, isolated valley. To establish the theorem, we need a way to measure how "close" two colorings σ, τ are. The Hamming distance is inappropriate here because two colorings σ, τ can be at Hamming distance n, although τ simply results from permuting the color classes of σ, i.e., although σ and τ are essentially identical. Instead, we shall use the following concept. Given two coloring σ, τ, we let $M_{\sigma,\tau} = (M^{ij}_{\sigma,\tau})_{1\leq i,j\leq k}$ be the matrix with entries

$$M^{ij}_{\sigma,\tau} = n^{-1}|\sigma^{-1}(i) \cap \tau^{-1}(j)|.$$

To measure how close τ is to σ we let

$$f_\sigma(\tau) = \|M_{\sigma,\tau}\|_F^2 = \sum_{i,j=1}^{k} (M^{ij}_{\sigma,\tau})^2 \ ,$$

be the squared Frobenius norm of $M_{\sigma,\tau}$. Observe that this quantity reflects the probability that a single random edge is monochromatic under both σ and τ, i.e., the correlation of σ and τ, precisely as desired. Hence, f_σ is a map from the set $[k]^n$ of k-partitions to the interval $\left[k^{-2}, f_\sigma(\sigma)\right]$, where $f_\sigma(\sigma) \geq k^{-1}$. Thus, the *larger* $f_\sigma(\tau)$, the more τ resembles σ. Furthermore, for a fixed $\sigma \in \mathcal{S}(G)$ and a number $\lambda > 0$ we let

$$g_{\sigma,G,\lambda}(x) = |\{\tau \in [k]^n : f_\sigma(\tau) = x \wedge H(\tau) \leq \lambda n\}|.$$

In order to show that $\mathcal{S}(G_{n,m})$ with $m = rn$ decomposes into exponentially many regions, we employ the following lemma.

Lemma 3. *Suppose that $r > (\frac{1}{2} + \varepsilon_k)k \ln k$. There are numbers $k^{-2} < y_1 < y_2 < k^{-1}$ and $\lambda, \gamma > 0$ such that with high probability a pair $(G,\sigma) \in \Lambda_{n,m}$ chosen from the distributoin $\mathcal{U}_{n,m}$ has the following two properties.*

1. *For all $x \in [y_1, y_2]$ we have $g_{\sigma,G,\lambda}(x) = 0$.*
2. *The number of colorings $\tau \in \mathcal{S}(G)$ such that $f_\sigma(\tau) > y_2$ is at most $\exp(-\gamma n) \cdot |\mathcal{S}(G)|$.*

Let $G = G_{n,m}$ be a random graph and call $\sigma \in \mathcal{S}(G)$ *good* if both (1) and (2) hold. Then Lemma 3 states that w.h.p. a $1 - o(1)$-fraction of all $\sigma \in \mathcal{S}(G)$ are good. Hence, to decompose $\mathcal{S}(G)$ into regions, we proceed as follows. For each $\sigma \in \mathcal{S}(G)$ we let $\mathcal{C}_\sigma = \{\tau \in \mathcal{S}(G) : f_\sigma(\tau) > y_2\}$. Then starting with the set $S = \mathcal{S}(G)$ and removing iteratively some $\mathcal{C}_\sigma$ for a good $\sigma \in S$ yields an exponential number of regions. Furthermore, each such region $\mathcal{C}_\sigma$ is separated by a linear Hamming distance from the set $\mathcal{S}(G)\setminus\mathcal{C}_\sigma$, because f_σ is "continuous" with respect to $n^{-1}\times$Hamming distance. Thus, Theorem 1 follows from Lemma 3.

Finally, by Theorem 6, to prove Lemma 3 it is sufficient to show the following.

Lemma 4. *Suppose that $r > (\frac{1}{2} + \varepsilon_k)k \ln k$. There are $k^{-2} < y_1 < y_2 < k^{-1}$ and $\lambda, \gamma > 0$ such that with probability at least $1 - \exp(-\Omega(n))$ a pair $(G,\sigma) \in \Lambda_{n,m}$ chosen from the distributoin $\mathcal{P}_{n,m}$ has the two properties stated in Lemma 3.*

The proof of Lemma 4 is based on the "first moment method". That is, for any $k^{-2} < y < k^{-1}$ we compute the *expected* number of assignments $\tau \in [k]^n$ such that $f_\sigma(\tau) = y$ and $H(\tau) \leq \lambda n$. This computation is feasible in the planted model and yields similar expressions as encountered in [4] in the course of computing the second moment of the number of k-colorings. Therefore, we can show that the expected number of such assignments τ is exponentially small for a regime $y_1 < y < y_2$, whence Lemma 21 follows from Markov's inequality.

References

1. D. Achlioptas, E. Friedgut. *A sharp threshold for k-colorability*, Random Struct. Algorithms **14**, 63–70, 1999.
2. D. Achlioptas and M. Molloy, *The analysis of a list-coloring algorithm on a random graph*, in Proc. of FOCS 1997, 204–212.
3. D. Achlioptas and C. Moore, *Random k-SAT: two moments suffice to cross a sharp threshold*, SIAM Journal on Computing, **36** (2006), 740–762.
4. D. Achlioptas and A. Naor, *The two possible values of the chromatic number of a random graph*, Annals of Mathematics, **162** (2005), 1333–1349.
5. D. Achlioptas, A. Naor, and Y. Peres, *Rigorous location of phase transitions in hard optimization problems*, Nature, **435** (2005), 759–764.
6. D. Achlioptas and Y. Peres, *The threshold for random k-SAT is $2^k \ln 2 - O(k)$*, Journal of the American Mathematical Society **17** (2004), 947–973.
7. D. Achlioptas, F. Ricci-Tersenghi, *On the solution space geometry of random constraint satisfaction problems*, in Proc. 38th ACM Symp. on Theory of Computing (2006), 130–139.
8. D. Battaglia, M. Kolar, R. Zecchina, *Minimizing energy below the glass thresholds*, Phys. Rev. E. **70** (2004), 036107.
9. A. Braunstein, R. Mulet, A. Pagnani, M. Weigt, R. Zecchina, *Polynomial iterative algorithms for coloring and analyzing random graphs*, Phys. Rev. E. **68** (2004), 036702.
10. M.-T. Chao and J. Franco, *Probabilistic analysis of two heuristics for the 3-satisfiability problem*, SIAM J. Comput. **15** (1986), 1106–1118.
11. V. Chvátal and B. Reed, *Mick gets some (the odds are on his side)*, in Proc. 33th Annual Symposium on Foundations of Computer Science (1992), 620–627.
12. A. Coja-Oghlan, M. Krivelevich, D. Vilenchik, *Why almost all k-colorable graphs are easy*, in Proc. 24th STACS (2007) 121–132.
13. A. Coja-Oghlan, M. Krivelevich, D. Vilenchik, *Why almost all k-CNF formulas are easy*, in Proc. 13th International Conference on Analysis of Algorithms.
14. E. Friedgut, *Sharp Thresholds of Graph Proprties, and the k-SAT Problem.* J. Amer. Math. Soc. **12** (1999), 1017–1054.
15. E. Friedgut, *Hunting for sharp thresholds.* Random Struct. Algorithms **26** (2005) 37–51
16. A. M. Frieze and S. Suen, *Analysis of two simple heuristics on a random instance of k-SAT*, Journal of Algorithms **20** (1996), 312–355.
17. A. Gerschenfeld, A. Montanari. *Reconstruction for models on random graphs.* in Proc. FOCS 2007, 194–204.
18. G.R. Grimmett, C.J.H. McDiarmid, *On colouring random graphs*, Math. Proc. Cambridge Philos. Soc., **77** (1975), 313–324.
19. A. Juels, M. Peinado: Hiding Cliques for Cryptographic Security. in Proc. SODA 1998, 678–684.
20. F. Krzakala, A. Montanari, F. Ricci-Tersenghi, G. Semerjianc, L. Zdeborova, *Gibbs states and the set of solutions of random constraint satisfaction problems.* Proc. National Academy of Sciences **104** (2007) 10318–10323.
21. M. Mézard, T. Mora, and R. Zecchina, *Clustering of Solutions in the Random Satisfiability Problem*, Phys. Rev. Lett. **94** (2005), 197205.
22. M. Mézard, G. Parisi, and R. Zecchina, *Analytic and Algorithmic Solution of Random Satisfiability Problems*, Science **297** (2002), 812–815.
23. A. Montanari, F. Ricci-Tersenghi, G. Semerjian. *Solving Constraint Satisfaction Problems through Belief Propagation-guided Decimation.* in Proc. 45th Allerton (2007).

In this appendix we present the proofs of our results for graph coloring and random k-SAT, thereby presenting the most important techniques. We omit the proofs for hypergraph 2-coloring, as these are similar to but simpler than the k-SAT proofs, due to the fact that the transfer theorem for hypergraph 2-coloring is as strong as that for k-coloring. We generally assume that n is sufficiently large.

A Graph coloring

A.1 The planted model

In this section we consider a fixed number $\varepsilon > 0$ and assume that $k \geq k_0$ for some sufficiently large $k_0 = k_0(\varepsilon)$. We are interested in the probability distribution $\mathcal{U}_{n,m}$ on $\Lambda_{n,m}$. To analyze this distribution, we consider the distribution $\mathcal{P}_{n,m}$ on $\Lambda_{n,m}$ induced by following expermient ("planted model").

P1. Generate a uniformly random k-partition $\sigma \in [k]^n$.
P2. Generate a graph G with m edges chosen uniformly at random among the edges bicolored under σ.
P3. Output the pair (G, σ).

Theorem 7. *Suppose that $d = 2m/n < (2-\varepsilon)k \ln k$. There exists a function $f(n) = o(n)$ such that the following is true. Let $\mathcal{D}$ be any graph property such that $G(n,m)$ has $\mathcal{D}$ with probability $1 - o(1)$, and let $\mathcal{E}$ be any property of pairs $(G,\sigma) \in \Lambda_{n,m}$. If for all sufficiently large n we have*

$$\Pr_{\mathcal{P}_{n,m}}\left[(G,\sigma) \text{ has } \mathcal{E} | G \text{ has } \mathcal{D}\right] \geq 1 - \exp(-f(n)), \tag{7}$$

then $\Pr_{\mathcal{U}_{n,m}}\left[(G,\sigma) \text{ has } \mathcal{E}\right] = 1 - o(1)$.

For a given assignment $\sigma \in [k]^n$ we let $G(\sigma)$ be the set of all graphs with m edges for which σ is a proper coloring. Then it is immediate that

$$|G(\sigma)| = \binom{\sum_{1 \leq i < j \leq k} |\sigma^{-1}(i)|}{m} = \binom{(n^2 - \sum_{i=1}^k |\sigma^{-1}(i)|^2)/2}{m}.$$

Hence,

$$\lambda = \max_\sigma |G(\sigma)| \leq \binom{(1-1/k)\binom{n}{2}}{m}.$$

Lemma 5. *There is a constant $\rho > 0$ such that the following is true. Let $\sigma \in [k]^n$ be chosen uniformly at random. Then* $\Pr\left[|G(\sigma)| \geq \rho\lambda\right] \geq \rho$.

Proof. Let $\gamma > 0$ be sufficiently small. Moreover, let $n_i = |\sigma^{-1}(i)|$, $\delta_i = n_i - n/k$, and $N = (1-k^{-1})\binom{n}{2}$. Then $\sum_i \delta_i = 0$. Therefore,

$$\sum_i n_i^2 = \frac{n^2}{k} + \sum_i \delta_i^2.$$

Since for a random σ the numbers n_i are multinomially distributed, with probability $\Omega(1)$ we have $|n_i - \frac{n}{k}| < \sqrt{\gamma n/k}$. Hence, letting $N(\sigma) = \sum_{i<j} n_i n_j$, we conclude that there is a constant $\rho > 0$ such that

$$\Pr\left[N(\sigma) \geq N - \gamma n\right] \geq \rho.$$

Thus, by Stirling's formula with probability at least ρ we have

$$\binom{N(\sigma)}{m}\binom{N}{m}^{-1} \geq \frac{1}{2}\cdot\left(\frac{N(\sigma)}{N}\right)^m\left(\frac{1-m/N}{1-m/N(\sigma)}\right)^{N(\sigma)-m}(1-m/N)^{N(\sigma)-N}.$$

Since $N=\Omega(n^2)$ and $m=O(n)$, in the case $N(\sigma)\geq N-\gamma n$ we have

$$\begin{aligned}\left(\frac{N(\sigma)}{N}\right)^m &\geq (1-\gamma n/N)^m \geq \Omega(1),\\ \frac{1-m/N}{1-m/N(\sigma)} &\geq 1,\\ (1-m/N)^{N(\sigma)-N} &\geq \Omega(1).\end{aligned}$$

Hence, choosing $\rho>0$ sufficiently small, we can ensure that $\Pr\left[|G(\sigma)|\geq\rho\lambda\right]\geq\rho$. □

Corollary 1. *We have* $|\Lambda_{n,m}|\geq\rho^2k^n\lambda$.

Lemma 6. *Suppose that there is a number* $\zeta>0$ *such that* $\Pr_{\mathcal{P}_{n,m}}[\mathcal{E}|\mathcal{C}]<\exp(-\zeta n)$. *Then*

$$|\{(G,\sigma)\in\Lambda_{n,m}\cap\mathcal{E}:G\in\mathcal{C}\}|\leq\rho^{-2}\exp(-\zeta n)|\Lambda_{n,m}|.$$

Proof. We have

$$\begin{aligned}\exp(-\zeta n)\geq\Pr_{\mathcal{P}_{n,m}}[\mathcal{E}|\mathcal{C}] &= \frac{\Pr_{\mathcal{P}_{n,m}}[\mathcal{E}\wedge\mathcal{C}]}{\Pr_{\mathcal{P}_{n,m}}[\mathcal{C}]}\\ &= k^{-n}\Pr_{\mathcal{P}_{n,m}}[\mathcal{C}]^{-1}\sum_{\sigma\in[k]^n}\frac{|\{G\in G(\sigma):(G,\sigma)\in\mathcal{E}\wedge G\in\mathcal{C})\}|}{|G(\sigma)|}\\ &\geq\lambda^{-1}k^{-n}\Pr_{\mathcal{P}_{n,m}}\left|\{(G,\sigma)\in\Lambda_{n,m}\cap\mathcal{E}:G\in\mathcal{C})\}\right|,\end{aligned}$$

because $|G(\sigma)|\leq\lambda$ for all σ. Hence, Corollary 1 yields

$$\exp(-\zeta n)\geq\frac{\rho^2}{|\Lambda_{n,m}|}\cdot\left|\{(G,\sigma)\in\Lambda_{n,m}\cap\mathcal{E}:G\in\mathcal{C})\}\right|,$$

as desired. □

Let $\mu=\mathrm{E}(|\mathcal{S}(G_{n,m})|)$ be the expected number of k-colorings of $G_{n,m}$. Combining the second moment argument from [4] with arguments from [1], we obtain the following result (see Appendix A.2).

Lemma 7. *There is a function* $f(n)=o(n)$ *such that* $|\mathcal{S}(G_{n,m})|\geq\exp(-f(n))\mu$ *with high probability.*

Proof of Theorem 7. Assume that a random pair (G,σ) chosen according to the uniform model has $\mathcal{E}$ with probability at least 2ζ, while $\Pr_{\mathcal{P}_{n,m}}[\mathcal{E}|\mathcal{C}]\leq\exp(-\zeta n)$ for an arbitrarily small $\zeta>0$. Since $G_{n,m}$ has the property $\mathcal{C}$ w.h.p., we conclude that

$$\Pr_{\mathcal{U}_{n,m}}[\mathcal{E}|\mathcal{C}]\geq\zeta.$$

Therefore, Lemma 7 entails that

$$b=|\{(G,\sigma)\in\Lambda_{n,m}\cap\mathcal{E}:G\in\mathcal{C}\}|\geq\zeta\exp(-f(n))|\Lambda_{n,m}|.$$

Since $f(n)=o(n)$, this contradicts Lemma 6. □

A.2 Proof of Lemma 7

To prove Lemma 7, we combine the second moment argument from [4] with a sharp threshold argument. Let $G = G(n, m)$ be a random graph and let X be the number of *balanced* colorings of G, i.e., colorings $\sigma \in [k]^n$ such that $|\sigma^{-1}(i) - n/k| \leq 1$ for all $1 \leq i \leq k$. Recall that $\mathcal{S}(G)$ denotes the set of all k-colorings of G. A direct computation involving Stirling's formula shows that

$$\mathrm{E}(X) \geq \Omega(n^{-k/2})\mathrm{E}|\mathcal{S}(G)|.$$

In addition, [4, Section 3] shows that there is a constant $C = C(k)$ such that

$$\mathrm{E}(X^2) \leq C(k)\mathrm{E}(X)^2.$$

Applying the Paley-Zigmund inequality, we thus conclude that there is a number $\alpha = \alpha(k) > 0$ such that

$$\Pr\left[X > \alpha\mathrm{E}(X)\right] \geq \alpha,$$

whence

$$\Pr\left[|\mathcal{S}(G)| \geq \Omega(n^{-k/2})\mathrm{E}|\mathcal{S}(G)|\right] \geq \alpha.$$

In addition, $\mathrm{E}|\mathcal{S}(G)|$ is easily computed: we have

$$n^{-1}\ln\mathrm{E}|\mathcal{S}(G)| = \ln k + r\ln(1-k^{-1}) + o(1),$$

where $r = m/n$. Thus, we obtain the following.

Lemma 8. *Let $\xi(k, r) = \ln k + r\ln(1 - k^{-1})$. Then $\Pr\left[n^{-1}\ln|\mathcal{S}(G(n, m))| \geq \xi(k, r) - o(1)\right] \geq \alpha$.*

To complete the proof of Lemma 7, we combine Lemma 8 with a sharp threshold result. Let $\mathcal{A}_\xi$ be the property that a graph G on n vertices has less than ξ^n k-colorings.

Lemma 9. *For any fixed $\xi > 0$ the property $\mathcal{A}_\xi$ has a sharp threshold. That is, there is a sequence r_n such that for any $\varepsilon > 0$ we have*

$$\lim_{n\to\infty}\Pr\left[G(n, (1-\varepsilon)\lceil r_n n\rceil) \text{ does not have } \mathcal{A}\right] = 1 - \lim_{n\to\infty}\Pr\left[G(n, (1+\varepsilon)\lceil r_n n\rceil) \text{ has } \mathcal{A}\right] = 0.$$

We shall prove Lemma 9 in Appendix A.3. Lemma 7 is an immediate consequence of Lemmas 8 and 9.

A.3 Proof of Lemma 9

The property $\mathcal{A}_\xi$ is monotone under the addition of edges. Therefore, it is sufficient to prove that $\mathcal{A}_\xi$ has a sharp threshold in the random graph $G(n, p)$, in which edges are added with probability p independently. Let $\mathcal{N} = \xi^n$. The argument builds upon [1]. We denote the set of k-colorings of a graph G by $\mathcal{S}(G)$.

Lemma 10. *Let $\mathcal{N}'$ be some number (that may depend on n), and let $0 < t < 1$ be fixed. Further, let $M = O(1)$ be an integer. Suppose that $\Pr\left[|\mathcal{S}(G(n, p))| \leq \mathcal{N}'\right] \leq 1 - \tau$. Moreover, assume that for a list of colors $c_1, \ldots, c_M \in \{1, \ldots, k\}$ the following is true: if we pick M vertices $v_1, \ldots, v_M$ indenpendently and uniformly at random, then with probability $\geq 1 - t/2$ the random graph $G(n, p)$ has at most $\mathcal{N}'$ k-colorings in which v_i receives a color different from c_i for all $1 \leq i \leq M$. Then with probability $\geq 1 - t/2 + o(1)$ the random graph $G(n, p)$ has at most $\mathcal{N}'$ k-colorings in which v_i receives a color different from c_i for all $1 \leq i \leq M - 1$.*

Proof. Let ω be a (very) slow growing function of n. Moreover, let $\mathcal{E}_i$ be the event that the first i constraints: "v_j must not receive color c_j" for $1 \leq j \leq i$, cause the number of k-colorings to be at most $\mathcal{N}'$. Then given that $G = G(n,p)$ has more than $\mathcal{N}'$ k-colorings (i.e., coditional on the event $\bar{\mathcal{E}}_0$), the probability of $\mathcal{E}_M$ is at least $\frac{1}{2}$. Hence, conditional on $\bar{\mathcal{E}}_0$, we have

$$\Pr\left[\mathcal{E}_{M-1}\right] + \Pr\left[\mathcal{E}_M | \bar{\mathcal{E}}_{M-1}\right] \Pr\left[\bar{\mathcal{E}}_{M-1}\right] \geq \frac{1}{2}.$$

We consider two cases: if $\Pr\left[\mathcal{E}_{M-1}\right] \geq \frac{1}{2} - \omega^{-1}$, then we are done. Hence, assume that $\Pr\left[\mathcal{E}_{M-1}\right] < \frac{1}{2} - \omega^{-1}$. Then $\Pr\left[\mathcal{E}_M | \bar{\mathcal{E}}_{M-1}\right] \geq \omega^{-1}$. Note that $\Pr\left[\mathcal{E}_M | \bar{\mathcal{E}}_{M-1}\right]$ is the fraction of vertices w such that forbidding color c_M at w reduces the number of colorings to at most $\mathcal{N}'$ (after the addition of the first $M-1$ constraints). We call such vertices w *good*, and denote the set of colorings that w spoils by Z_w and its size by $z_w = |Z_w|$. Let us consider two cases. Let y be the number of colorings of G that respect the first $M-1$ constraints. We consider two cases. In each of these two cases we shall prove that adding a small number of random edges to $G(n,p)$ reduces the number of colorings that respect the first $M-1$ constraints to at most $\mathcal{N}'$ with probability at least $1 - \omega^{-3}$.

Case 1: $y < \omega\mathcal{N}'$**.** Since for each of the y colorings the probability that a new random edge spoils this coloring is k^{-1}, we can reduce the number of colorings to at most $\mathcal{N}'$ by adding ω random edges (use Markov's inequality).

Case 2: $y \geq \omega\mathcal{N}'$**.** If w, w' are good, then $|Z_w \cap Z_{w'}| \geq y - 2\mathcal{N}'$. Therefore, adding an edge between two good vertices causes the number of colorings to drop to at most $2\mathcal{N}'$. Furthermore, the probability that a random edge joins two good vertices is $\Pr\left[\mathcal{E}_M | \bar{\mathcal{E}}_{M-1}\right]^2 \geq \omega^{-2}$. Therefore, after adding ω^{10} edges, we have reduced the number of proper colorings to at most $2\mathcal{N}'$ with probability $\geq 1 - \omega^{-4}$. Finally, adding an additional ω^{10} edges reduces the number of colorings to at most $\mathcal{N}'$ by the same argument as in Case 1.

Now, note that instead of *first* imposing the $M-1$ constraints $w_1, \ldots, w_{M-1}$ and *then* adding the random edges as in Cases 1 and 2 we could *first* add a set of $2\omega^{10}$ random edges to $G(n,p)$. As ω^{10} is of smaller order than the standard deviation of the number of edges of $G(n,p)$, the resulting distribution is within $o(1)$ from the originial distribution $G(n,p)$ in total variation distance. Therefore, we conclude that actually just imposing the $M-1$ constraints $w_1, \ldots, w_{M-1}$ suffices to increase the probability of having $\leq \mathcal{N}'$ k-colorings to $1 - \tau/2 + o(1)$. □

Corollary 2. *Let $\mathcal{N}'$ be some number (may depend on n), and let $0 < t < 1$ be fixed. Further, let $M = O(1)$ be an integer. Suppose that $\Pr\left[|\mathcal{S}(G(n,p))| \leq \mathcal{N}'\right] \leq 1 - \tau$. Then there is no list of colors $c_1, \ldots, c_M \in \{1, \ldots, k\}$ such that the following is true: if we pick M vertices $v_1, \ldots, v_M$ indenpendently and uniformly at random, then with probability $\geq 1 - t/2$ the random graph $G(n,p)$ has at most $\mathcal{N}'$ k-colorings in which v_i receives a color different from c_i for all $1 \leq i \leq M$.*

Proof. Applying the lemma M times, we can reduce the number of constraints that is necessary to reduce the number of colorings to $\leq \mathcal{N}'$ to 0. □

To prove that $\mathcal{A}_\xi$ has a sharp threshold, we assume for contratiction that this is not so. Hence, there exists an edge probability p^* such that the probability that G_{n,p^*} has $\mathcal{A}_\xi$ is exactly equal to $1-t$ for a small $t > 0$. Further, by [15, Theorem 2.4] there exists a fixed graph R on r vertices such that with probability $> 1 - t/3$ the following is true. If we first pick $G = G_{n,p^*}$ and then insert a random copy of R into G, then the resulting graph has $\mathcal{A}_\xi$. Furthermore, this graph R is k-colorable. In fact, by monotonicity we may assume that R is

uniquely k-colorable. The experiment of inserting a random copy of R into G_{n,p^*} is actually equivalent to the following (because G_{n,p^*} is symmetric with respect to vertex permutations). We let G_R denote a random graph obtained by first inserting a copy of R into the first r vertices $v_1, \ldots, v_r$, and then adding edges with probability p^* independently (among all n vertices $v_1, \ldots, v_n$). Then the probability that G_R has $\mathcal{A}_\xi$ is at least $1 - t/3$. Hence,

$$\Pr\left[|\mathcal{S}(G_R)| \leq \mathcal{N}\right] \geq 1 - t/3, \tag{8}$$

while by the choice of p^*

$$\Pr\left[|\mathcal{S}(G_{n,p^*})| > \mathcal{N}\right] \geq t > 0. \tag{9}$$

Let $\hat{G}$ signify the subraph of G_R induced on the $n-r$ vertices $v_{r+1}, \ldots, v_n$. Then $\hat{G} = G_{n-r,p^*}$, and (9) implies that

$$\Pr\left[|\mathcal{S}(\hat{G})| > k^{-r}\mathcal{N}\right] \geq t. \tag{10}$$

Furthermore, we can relate the k-colorings of G_R and the k-colorings of $\hat{G}$ as follows. Let Q be the set of edges from the set $\{v_1, \ldots, v_r\}$ to $\{v_{r+1}, \ldots, v_n\}$. Then w.h.p. $|Q| = O(1)$ and no vertex in $\{v_{r+1}, \ldots, v_n\}$ is incident to more than one edge in Q. Furthermore, since R admits a unique k-coloring, each edge in Q forbids its endpoint in $\{v_{r+1}, \ldots, v_n\}$ exactly one color. Hence, the edges in Q impose constraints $c_1, \ldots, c_M$ on $M = |Q|$ randomly chosen vertices $w_1, \ldots, w_M$ as in Lemma 10. Therefore, (8) implies that

$$\Pr\left[\hat{G} + M \text{ random constraints has at most } \mathcal{N} \ k\text{-colorings}\right] \geq 1 - t/3.$$

But then Corollary 2 implies that

$$\Pr\left[|\mathcal{S}(\hat{G})| \leq \mathcal{N}\right] \geq 1 - t/3.$$

Furthermore, as we may add another $\ln n$ random edges to $\hat{G}$ without shifting the distribution by more than $o(1)$ in total variation distance, and since each of these $\ln n$ edges reduces the expected number of colorings by k^{-1}, Markov's inequality entails that

$$\Pr\left[|\mathcal{S}(\hat{G})| \leq k^{-r}\mathcal{N}\right] \geq 1 - t/3 - o(1),$$

which contradicts (10).

A.4 Proof of Theorem 5

Suppose that $d \leq (1-\varepsilon)k \ln k$, and that $k \geq k_0(\varepsilon)$ for a sufficiently large $k_0(\varepsilon)$. Let $q = 5$ and recall that a graph is ζ-choosable if for any assignments of color lists of length at least ζ to the vertices of the graph there is a proper coloring such that each vertex receives a color from its list. To prove Theorem 5, we consider the property $\mathcal{E}$ that all vertices are loose and the following condition $\mathcal{D}$:

For any set $S \subset V$ of size $|S| \leq g(n)$ the subgraph induced on S is $(q-1)$-choosable.

Here $q > 0$ is a constant and $g(n) = \sqrt{nf(n)} = o(n)$, where $f(n)$ is the function from Theorem 7.

Lemma 11. *With high probability the random graph $G(n,m)$ satisfies $\mathcal{D}$.*

Proof. Since $m = O(n)$, this follows from a standard first moment argument. □

By Theorem 7, we just need to establish (7). Thus, let $\sigma \in [k]^n$ be a coloring such that the color classes $V_i = \sigma^{-1}(i)$ satisfy $|V_i| \sim n/k$, and let G be a random graph with m edges such that σ is a k-coloring of G. Let $v_0 \in V$ be any vertex; without loss of generality we may assume that $\sigma(v_0) = 1$. In addition, let $1 < l \leq k$ be the "target color" for v_0. If v_0 has no neighbor in V_l, then we can just assign this color to v_0.

Otherwise, we run the following process. In the course of the process, every vertex is either *awake*, *dead*, or *asleep*. Initially, all the neighbors of v_0 in V_l are awake, v is dead, and all other vertices are asleep. In each step of the process, pick an awake vertex w arbitrarily and declare it dead (if there is no awake vertex, terminate the process). If there are at least q colors $c_1(w), \ldots, c_q(w)$ such that w has no neighbor in $V_{c_i(w)}$, then we do nothing. Otherwise, we pick q colors $c_1(w), \ldots, c_q(w)$ randomly and declare all asleep neighbors of w in $V_{c_j(w)}$ awake for $1 \leq j \leq q$.

Lemma 12. *With probability at least $1-\exp(-f(n))$ there are at most $g(n)$ dead vertices when the process terminates.*

Proof. We show that the aforementioned process is dominated by a branching process in which the expected number of successors is less than one. Then the assertion follows from Chernoff bounds.

To set up the analogy, note that the expected number of neighbors of any $w \in V \setminus V_i$ in V_i is asymptotically $2m/k < \frac{1-\varepsilon}{1-k} \cdot \ln k$. Hence, the probability that w has no neighbor in V_i is at least $k^{\varepsilon/2-1}$. Therefore, the expected number of classes $i \neq \sigma(w)$ in which w has no neighbor is at least $(k-1)k^{\varepsilon/2-1} \geq k^{\varepsilon/3}$. Furthermore, the number of such classes is asymptotically binomially distributed. Therefore, assuming that k is sufficiently large, we conclude that the probability that there are less than q classes in which w has no neighbor is less than k^{-1}. Given that this is so, the number of neighbors of w in each of the q chosen classes $c_1(w), \ldots, c_q(w)$ has mean at most $2\ln k$. Therefore, the expected number of newly awake vertices resulting from w is at most $2k^{-1}\ln k$. Thus, the above process is dominated by a branching process with successor rate $2k^{-1}\ln k < 1$. Therefore, the assertion follows from stochastic domiance and Chernoff bounds. □

Proof of Theorem 5. Let S be the set of dead vertices left by the aforementioned process. By Lemma 12 we may assume that $|W| \leq g(n)$. Hence, conditioning on $\mathcal{D}$, we may assume that S is $(q-1)$-choosable. Now, we assign lists of colors to the vertices in S as follows. The list of v_0 just consists of its target color l. To any other $w \in W$ we assign the list $L_w = \{c_1(w), \ldots, c_q(w)\} \setminus \{l\}$. Now, we can color the subgraph $G[S]$ by assigning color l to v and a color from L_w to any other $w \in W$. We extend this to a coloring of G by assigning color $\sigma(u)$ to any $u \in V \setminus W$. Let τ signify the resulting coloring.

We claim that τ is a proper coloring of G. For both the subgraph induced on W and the subgraph induced on $V \setminus W$ are properly colored. Moreover, by construction no $w \in W \setminus \{v\}$ is adjacent to a vertex of color $c_j(w)$ in $V \setminus W$. Finally, σ and τ are at Hamming distance at most $|W| \leq g(n) = o(n)$. Hence, the assertion follows from Theorem 7. □

A.5 Rigid variables

Let $\alpha, \varepsilon > 0$, and assume that $k \geq k_0(\varepsilon)$ for a large enough $k_0(\varepsilon, \alpha) > 0$. Suppose that $(1+\varepsilon)k\ln k \leq d = 2m/n \leq (2-\varepsilon)k\ln k$. To prove Theorem 4 for coloring, we apply Theorem 7 as follows. We let $\beta = \beta(\varepsilon, \alpha) > 0$ be a sufficiently small number and denote by $\mathcal{E}$ the following property of a pair $(G, \sigma) \in \Lambda_{n,m}$.

> There is a subgraph $G_* \subset G$ of size $|V(G_*)| \geq (1-\alpha)n$ such that for every vertex v of G_* and each color $i \neq \sigma(v)$ there are at least $\beta \ln k$ vertices vertex w in G_* that are adjacent to v such that $\sigma(w) = i$. (11)

Also, we let $\mathcal{D}$ be the property that the maximum degree is at most $\ln^2 n$.

Lemma 13. *Condition (7) is satisfied with $\mathcal{D}$ and $\mathcal{E}$ as above.*

Proof of Theorem 4 for coloring. Given a random coloring σ of a random graph $G = G(n, m)$, Lemma 13 and Theorem 7 imply that w.h.p. there is a subgraph G_* satisfying (11). In addition, we assume that G has the following property.

There is no set $S \subset V$ of size $|S| \leq n/(k \ln k)$ that spans at least $|S|\frac{\beta}{2} \ln k$ edges. (12)

A standard 1st moment argument shows that (12) holds in $G(n, m)$ w.h.p.

Assume for contradiction that there is another coloring τ such that the set $U = \{v \in G_* : \sigma(v) \neq \tau(v)\}$ has size $|U| \leq n/(k \ln k)$. Let $U_i^+ = \{v \in G_* : \tau(v) = i \neq \sigma(v)\}$ and $U_i^- = \{v \in G_* : \sigma(v) = i \neq \tau(v)\}$. Then

$$|U| = \sum_{i=1}^{k} |U_i^+| = \sum_{i=1}^{k} |U_i^-|. \tag{13}$$

Every $v \in G_* \setminus V_i$ has at least $\beta \ln k$ neighbors in $G_* \cap \sigma^{-1}(i)$. Hence, if $v \in U_i^+$, then all of these neighbors lie inside of U_i^-. We claim that this implies that $|U_i^+| < |U_i^-|$. For assume that $U_i^+ \geq U_i^-$ and set $S = U_i^+ \cup U_i^-$. Then $|S| \leq |U| \leq n/(k \ln k)$, and S spans at least $|S|\frac{\beta}{2} \ln k$ edges, in contradiction to (12). Thus, we conclude that $|U_i^+| < |U_i^-|$ for *all* i, in contradiction to (13). Hence, all the vertices in G_* are $\Omega(n)$-rigid. □

A.6 Proof of Lemma 13

Let $(G, \sigma) \in \Lambda_{n,m}$ be a random pair chosen from the distribution $\mathcal{P}_{n,m}$. We may assume that $|\sigma^{-1}(i)| \sim n/k$ for all i and let $V_i = \sigma^{-1}(i)$. To simplify the analysis, we shall replace the random graph G, which has a *fixed* number m of edges, by a random graph G' in which is obtained by including each edge $\{v, w\}$ with $\sigma(v) \neq \sigma(w)$ with probability p independently. Here p is chosen so that the *expected* number $\sum_{1 \leq i < j \leq k} |V_i| \cdot |V_j| \cdot p$ of edges of G' equals m.

Lemma 14. *For any property $\mathcal{Q}$ we have* $\Pr[G \text{ has } \mathcal{Q} | \mathcal{D}] \leq O(\sqrt{n} \cdot \Pr[G' \text{ has } \mathcal{Q} | \mathcal{D}])$.

We defer the proof to Section A.7.

Thus, in the sequel we will work with G' rather than G. Let $\gamma = \gamma(\varepsilon) > 0$ be a sufficiently small number, and let $V_i = \sigma^{-1}(i)$. Moreover, for a vertex v and a set $Z \subset V$ let $e(v, Z)$ signify the number of v-Z-edges in G'. We construct a subgraph G_* of G' as follows.

1. Let $W_{ij} = \{v \in V_i : e(v, V_j) < \gamma \ln k\}$, $W_i = \bigcup_{j=1}^{k} W_{ij}$, and $W = \bigcup_{i=1}^{k} W_i$.
2. Let $U_{il} = \{v \in V_l : e(v, W_i) > \frac{\gamma}{2} \ln k\}$ and $U = \bigcup_{i \neq l} U_{il}$.
3. Let $Z = U$. While there is a vertex $v \in V \setminus Z$ that has at least 10 neighbors in Z, add v to Z.
4. Let $G_* = G' - \bigcup_{i=1}^{k} W_i - Z$.

For each vertex $v \in V_i$ and each color $j \neq i$ the expected number of neighbors of v with color i is $|V_i| \cdot \frac{2m}{n} \sim (1 + 2\varepsilon) \ln k$. Hence, the sets W_{ij} contain those vertices form V_i that have a lot fewer neighbors with color j than expected.

Lemma 15. *There is a number $\beta = \beta(\varepsilon) > 0$ such that with probability $\geq 1 - \exp(-\Omega(n))$ we have $|W_{ij}| < nk^{-2-\beta}$ for any i, j. Hence, $|W_i| \leq nk^{-1-\beta}$, and $|W| \leq nk^{-\beta}$.*

Proof. In the random graph G' for each $v \in V_i$ the number $e(v, V_j)$ is binomially distributed. Hence, the probability that $e(v, V_j) < \gamma \ln k$ is at most $\exp(-(1+\varepsilon')\ln k)$, where $\varepsilon' > 0$ depends only on ε and γ. Furthermore, as in G' edges occur independently, $|W_{ij}|$ is binomially distributed as well (with mean $\leq \frac{n}{k} \cdot \exp(-(1+\varepsilon')\ln k)$). Therefore, the assertion follows from Chernoff bounds. □

Each of the vertices in U has *a lot* of neighbors in the small set W. Therefore, since the random graph G' is a good expander, we expect U to be much smaller than W.

Lemma 16. *Given that $\mathcal{D}$ occurs, with probability at least $1 - \exp(-\Omega(n))$ the set U contains at most nk^{-7} vertices.*

We postpone the proof to Section A.8.

Lemma 17. *With probability $\geq 1 - \exp(-\Omega(n))$ the set Z contains at most nk^{-6} vertices.*

Proof. Assume that this is not the case. Let Y contain all vertices of U and the first $nk^{-6} - |U|$ vertices added to Z by step 3 of the construction of G_*. Then $|Y| \leq nk^{-6}$ and $e(Y) \geq 9|Y|$. However, a simple first moment argument shows that the probability that a set Y with these two properties is present in G' is at most $\leq \exp(-\Omega(n))$. □

Combining Lemma 15, 16, and 17, we conclude that G_* contains at least $n(1-\alpha)$ vertices (provided that k is sufficiently larger). Moreover, the construction of G_* ensures that this graph satisfies (11).

A.7 Proof of Lemma 14

Given that G' has exactly m edges, G' is just a uniformly random graph with planted coloring $V_1, \ldots, V_k$. That is, given that the number of edges is m, G' is identically distributed to G. Therefore,

$$\begin{aligned}\Pr\left[G' \text{ has } \mathcal{P}|\mathcal{D}\right] &\geq \frac{\Pr\left[G' \text{ has both } \mathcal{P} \text{ and } \mathcal{D} \text{ and has } m \text{ edges}\right]}{\Pr\left[G' \text{ has } \mathcal{D}\right]} \\ &\geq \Omega(n^{-\frac{1}{2}}) \cdot \frac{\Pr\left[G \text{ has both } \mathcal{P} \text{ and } \mathcal{D}\right]}{\Pr\left[G' \text{ has } \mathcal{D}\right]}. \end{aligned} \tag{14}$$

Furthermore, since $m = O(n)$, with probability $1 - o(1)$ the maximum degree of G as well as of G' is at most $\ln n$. Therefore, G, G' have $\mathcal{D}$ with probability $1 - o(1)$. Hence, (14) yields

$$\Pr\left[G' \text{ has } \mathcal{P}|\mathcal{D}\right] \geq \Omega(n^{-\frac{1}{2}}) \cdot \frac{\Pr\left[G \text{ has both } \mathcal{P} \text{ and } \mathcal{D}\right]}{\Pr\left[G \text{ has } \mathcal{D}\right]} = \Omega(n^{-\frac{1}{2}}) \cdot \Pr\left[G \text{ has } \mathcal{P}|\mathcal{D}\right],$$

as claimed.

A.8 Proof of Lemma 16

To analyze the sets U_{il} from the second step of the construction of G_*, consider

$$\begin{aligned} U'_{il} &= \{v \in V_l : e(v, W_i \setminus W_{il}) > \frac{\gamma}{4}\ln k\}, \\ U''_{il} &= \{v \in V_l : e(v, W_{il}) > \frac{\gamma}{4}\ln k\}. \end{aligned}$$

Lemma 18. *With probability* $\geq 1-\exp(-\Omega(n))$ *we have* $|U'_{il}| \leq nk^{-10}$.

Proof. The definition of the set $W_i \setminus W_{il}$ depends solely on the V_i-$V \setminus V_l$-edges. Therefore, the V_l-V_i-edges are indepenent of the random set $W_i \setminus W_{il}$, which with probability $\geq 1-\exp(-\Omega(n))$ has size $\leq nk^{-1-\beta}$ by the Lemma 15. Assuming that this is indeed the case, we conclude that for any vertex $v \in V_l$ the number $e(v, W_i \setminus W_{il})$ is binomially distributed with mean $npk^{-1-\beta} \leq 2k^{-\beta}\ln k$. Hence, the probability that v has $z = \frac{\gamma}{4}\ln k$ neighbors inside $W_i \setminus W_{il}$ is at most

$$\binom{nk^{-1-\beta}}{z}p^z \leq \left(\frac{8e}{\gamma k^{\beta}}\right)^z \leq k^{-10}/2.$$

Thus, $\mathrm{E}|U'_{il}| \leq nk^{-10}/2$. Finally, as $|U'_{il}|$ is binomially distributed, the assertion follows from Chernoff bounds. □

Lemma 19. *Conditional on the event* $\mathcal{D}$*, with probability* $\geq 1-\exp(-\Omega(n))$ *we have* $|U''_{il}| \leq nk^{-10}$.

Proof. We just need to analyze the bipartite subgraph $G'[V_i \cup V_l]$. The set W_{il} consists of all vertices $v \in V_i$ that have degree $< \gamma \ln k$ in this subgraph. To investigate $G'[V_i \cup V_l]$, we condition on the degree sequence $\boldsymbol{d}$ of this bipartite graph. Since we also condition on the event $\mathcal{D}$, the maximum degree is $\leq \ln^2 n$. Hence, we can generate the random bipartite graph with degree sequence $\boldsymbol{d}$ via the configuration model, and the probability that the resulting multigraph happens to be a simple graph is $\geq \exp(-O(\ln^4 n))$. Thus, we just need to study a random configuration.

Now, in a random configuration the probability that a vertex $v \in V_l$ has $\frac{\gamma}{4}\ln k$ neighbors in the set W_{il} is $\leq k^{-10}$, because the total number of edges of $G'[V_i \cup V_l]$ is concentrated about n^2pk^{-2}. Therefore, the (conditional) expected size of U''_{il} is $\leq nk^{-11}$. Consequently, Azuma's inequality yields that with probability $\geq 1-\exp(-\Omega(n))$ the size of U''_{il} is $\leq nk^{-10}$. □

Finally, Lemma 16 follows immedately from the fact that $U_{il} \subset U'_{il} \cup U''_{il}$.

A.9 Proof of Theorem 1

To prove the coloring part of Theorem 1, we need to come up with an appropriate way to measure how "similar" two k-colorings of a given graph are $G = G(n,m)$. A first idea may be to just use the Hamming distance. However, if we construct a coloring τ simply by permuting the color classes of another coloring σ, then σ and τ can have Hamming distance n, although they are essentially identical. Therefore, instead of the Hamming distance we shall use the following concept. Given two coloring σ, τ, we let $M_{\sigma,\tau} = (M^{ij}_{\sigma,\tau})_{1\leq i,j\leq k}$ be the matrix with entries

$$M^{ij}_{\sigma,\tau} = n^{-1}|\sigma^{-1}(i) \cap \tau^{-1}(j)|.$$

Then to measure how close τ is to σ we let

$$f_\sigma(\tau) = \|M_{\sigma,\tau}\|_F^2 = \sum_{i,j=1}^{k}(M^{ij}_{\sigma,\tau})^2$$

be the squared Frobenius norm of $M_{\sigma,\tau}$. Hence, f_σ is a map from the set $[k]^n$ of k-colorings to the interval $\left[k^{-2}, f_\sigma(\sigma)\right]$, where $f_\sigma(\sigma) \geq k^{-1}$. (Thus, the *larger* $f_\sigma(\tau)$, the more τ resembles σ.) Furthermore, for a fixed $\sigma \in \mathcal{S}(G)$ and a number $\lambda > 0$ we let

$$g_{\sigma,G,\lambda}(x) = |\{\tau \in [k]^n : f_\sigma(\tau) = x \wedge H(\tau) \leq \lambda n\}|.$$

In order to show that $\mathcal{S}(G_{n,m})$ with $m = rn$ decomposes into exponentially many regions, we employ the following lemma.

Lemma 20. *Suppose that $r > (\frac{1}{2} + \varepsilon_k)k \ln k$. There are numbers $k^{-2} < y_1 < y_2 < k^{-1}$ and $\lambda, \gamma > 0$ such that with high probability a pair $(G, \sigma) \in \Lambda_{n,m}$ chosen from the distributoin $\mathcal{U}_{n,m}$ has the following two properties.*

1. *For all $x \in [y_1, y_2]$ we have $g_{\sigma,G,\lambda}(x) = 0$.*
2. *The number of colorings $\tau \in \mathcal{S}(G)$ such that $f_\sigma(\tau) > y_2$ is at most $\exp(-\gamma n) \cdot |\mathcal{S}(G)|$.*

Let $G = G_{n,m}$ be a random graph and call $\sigma \in \mathcal{S}(G)$ *good* if 1. and 2. hold. Then Lemma 20 states that with high probability a $1 - o(1)$-fraction of all $\sigma \in \mathcal{S}(G)$ is good. Hence, to decompose $\mathcal{S}(G)$ into regions, we proceed as follows. For each $\sigma \in \mathcal{S}(G)$ we let

$$\mathcal{C}_\sigma = \{\tau \in \mathcal{S}(G) : f_\sigma(\tau) > y_2\}.$$

Then starting with the set $S = \mathcal{S}(G)$ and removing iteratively some $\mathcal{C}_\sigma$ for a good $\sigma \in S$ from S yields an exponential number of regions. Furthermore, each such region $\mathcal{C}_\sigma$ is separated by a linear Hamming distance from the set $\mathcal{S}(G) \setminus \mathcal{C}_\sigma$, because f_σ is continuous with respect to $n^{-1}\times$Hamming distance (that is, for any $\varepsilon > 0$ there is $\delta > 0$ such that $f_\sigma(\tau) < \varepsilon$ for all $\tau \in [k]^n$ satisfying $\mathrm{dist}(\sigma, \tau) < \delta n$). Thus, the property stated in Theorem 1 follows from Lemma 20.

To establish Lemma 20, we employ the planted model.

Lemma 21. *Suppose that $r > (\frac{1}{2} + \varepsilon_k)k \ln k$. There are $k^{-2} < y_1 < y_2 < k^{-1}$ and $\lambda, \gamma > 0$ such that with probability at least $1 - \exp(-\Omega(n))$ a pair $(G, \sigma) \in \Lambda_{n,m}$ chosen from the distributoin $\mathcal{P}_{n,m}$ the two properties stated in Lemma 20.*

Thus, Lemma 20 follows from Lemma 21 and Theorem 7.

Proof of Lemma 21. The proof is based on the first moment method. Let $\sigma \in [k]^n$ be a fixed assignment of colors to the vertices. We may assume that $|\sigma^{-1}(i)| \sim n/k$ for all $1 \le i \le k$, because all but an exponentially small fraction of all assignments in $[k]^n$ have this property. Further, let G be a graph with m edges such that σ is a k-coloring of G chosen uniformly at random from the set of all such graphs. A direct computation shows that for an assignment $\tau \in [k]^n$ the probability that $H(\tau) \le \lambda n$ is

$$\le \left(\frac{1 - 2k^{-1} + f_\sigma(\tau)}{1 - k^{-1}}\right)^{rn} \exp((\psi(\lambda) + o(1))n), \tag{15}$$

where $\lim_{\lambda\to 0} \psi(\lambda) = 0$. To prove the lemma, we shall compute the *expected* number of assignments τ such that $H(\tau) \le \lambda n$ and $f_\sigma(\tau) = x$ for a suitable $y_1 < x < y_2$.

To this end, we have to take into account the number of possible colorings τ. We parameterize the set of all possible τ by a matrix $A = (a_{ij})_{1\le i,j\le k}$, where $a_{ij} = n^{-1}|\sigma^{-1}(i) \cap \tau^{-1}(j)|$. Then by (15) the contribution of a matrix A to the first moment is at most

$$\mathcal{F}(A) = k^{-n}\binom{n}{(a_{ij}n)_{1\le i,j\le k}}\left(\frac{1 - 2k^{-1} + f_\sigma(\tau)}{1 - k^{-1}}\right)^{rn} \exp((\psi(\lambda) + o(1))n)$$

(the k^{-n} accounts for the fact that we consider the coloring σ fixed). Taking logarithms, we obtain

$$n^{-1} \ln \mathcal{F}(A) \sim -\ln k - \sum_{i,j=1}^{k} a_{ij} \ln(a_{ij}) + r \ln\left(\frac{1 - 2k^{-1} + \sum_{i,j=1}^{k} a_{ij}^2}{1 - k^{-1}}\right) + \psi(\lambda).$$

For a given number x we let $\mathcal{A}(x)$ be the set of all matrices $A = (a_{ij})_{1\le i,j\le k}$ such that $a_{ij} \ge 0$, $\sum_{i=1}^{k} a_{ij} \sim k^{-1}$, and $\sum_{i,j=1}^{k} a_{ij}^2 = x$. Since there are at most n^{k^2} possible matrices A, for any given x the expected number of colorings τ such that $f_\sigma(\tau) = x$ is at most

$$n^{k^2} \max_{A\in\mathcal{A}(x)} \mathcal{F}(A).$$

Hence, by continuity it suffices to show that for some $x \in [y_1, y_2]$ the expression $n^{-1} \max_{A\in\mathcal{A}(h)} \ln \mathcal{F}(A)$ is strictly negative for a small enough $\lambda > 0$.

Let $h = k^{-3/2}$ and $x = k^{-1} - 2h$. Then Theorem 9 from [4] shows that the maximum $\max_{A\in\mathcal{A}(x)} \ln \mathcal{F}(A)$ is attained for a matrix A with entries

$$a_{ii} = k^{-1} - h + o(h), \quad a_{ij} \sim h(k-1)^{-1} \quad (i \ne j)$$

asymptotically as k grows. An explicit computation shows that for this matrix A the value $\ln \mathcal{F}(A)$ is strictly negative, provided that λ is sufficiently small. Therefore, we can apply Markov's inequality to complete the proof. □

B Proofs for Random k-SAT

B.1 The planted model

Consider the distribution $\mathcal{U}_{n,m}$ on the set $\Lambda_{n,m}$ of pairs (F, σ), where F is a k-SAT formula with variables $x_1, \ldots, x_n$ and with m clauses, and σ is a satisfying assignment of F.

U1. Generate a random formula $F = F_k(n, m)$.
U2. Sample a satisfying assignment σ of F uniformly at random; if F is unsatisfiable, fail.
U3. Output the pair (F, σ).

To analyze this distribution, we consider the distribution $\mathcal{P}_{n,m}$ on $\Lambda_{n,m}$ induced by following expermient.

P1. Generate a random assignment $\sigma \in \{0, 1\}^n$.
P2. Generate a random k-CNF formula F with m clauses chosen uniformly among those satisfied by σ.
P3. Output the pair (F, σ).

Our goal is to establish the following connection between these two distributions.

Theorem 8. *There is a sequence $\varepsilon_k \to 0$ such that the following holds. Let $m = \lceil rn \rceil$ for some $r < (1 - \varepsilon_k)2^k \ln 2$, and let $f(n)$ be any function that such that $\lim_{n\to\infty} f(n) = \infty$. Let $\mathcal{D}$ be any property such that $F_k(n, m)$ has $\mathcal{D}$ with probability $1 - o(1)$, and let $\mathcal{E}$ be any property of pairs $(F, \sigma) \in \Lambda_{n,m}$. If for all sufficiently large n we have*

$$\Pr_{\mathcal{P}_{n,m}}[(F, \sigma) \text{ has } \mathcal{E} | F \text{ has } \mathcal{D}] \ge 1 - \exp(-k2^{3-k}n - f(n)), \tag{16}$$

then $\Pr_{\mathcal{U}_{n,m}}[(F, \sigma) \text{ has } \mathcal{E}] = 1 - o(1)$.

The proof of Theorem 8 is based on the following lemma, which we establish in Section B.2.

Lemma 22. *Let $\mu = 2^n(1 - 2^{-k})^m$ denote the expected number of satisfying assignments of a random k-CNF $F_k(n, m)$. Then for $k \ge 8$, w.h.p.*

$$|\mathcal{S}(F_k(n, m)| \ge \mu \exp(-k2^{3-k}n) \ .$$

Proof (Theorem 8). Assume for contradiction that there is a fixed $\alpha > 0$ such that $\Pr_{\mathcal{U}_{n,m}}[(F,\sigma) \text{ has } \mathcal{E}] \geq \alpha$ for infinitely many n. Then Lemma 22 implies that the set $L = \Lambda_{n,m} \setminus \mathcal{E}$ has size

$$|L| \geq \frac{\alpha}{2}\mu \exp(-k2^{3-k}n)\left[2\binom{n}{k}\right]^m = \frac{\alpha}{2}\exp(-k2^{3-k}n)|\Lambda_{n,m}|. \tag{17}$$

On the other hand, as $\mathcal{P}_{n,m}$ is just the uniform distribution on the set $\Lambda_{n,m}$, (16) implies that

$$|L| \leq \exp(-k2^{3-k}n - f(n))|\Lambda_{n,m}|.$$

As $f(n) \to \infty$, this contradicts (17) for sufficiently large n. □

B.2 Proof of Lemma 22

Let Λ_b be the function defined by

$$\Lambda_b(1/2,k,r) = 4\left[\frac{\left((1-\epsilon/2)^k - 2^{-k}\right)^2}{(1-\epsilon)^k}\right]^r \quad , \tag{18}$$

where ϵ satisfies

$$\epsilon(2-\epsilon)^{k-1} = 1 \ . \tag{19}$$

Lemma 23. *Suppose that $r < 2^k \ln 2 - k$. Then $F_k(n, rn)$ has at least $(\Lambda_b(1/2,k,r) - o(1))^{n/2}$ satisfying assignments w.h.p.*

Recall that $F_k(n,m)$ denotes a random k-SAT formula on n variables $x_1,\ldots,x_n$. For a fixed number $B > 1$ we let $\mathcal{A}_B$ denote the property that a k-SAT formula F on the variables $x_1,\ldots,x_n$ has less than $\frac{1}{2}B^n$ satisfying assignments. The following lemma shows that $\mathcal{A}_B$ has a sharp threshold.

Lemma 24. *For any $B > 1$ there is a sequence $(T_n^B)_{n\geq 1}$ of integers such that for any $\epsilon > 0$ we have*

$$\lim_{n\to\infty} \Pr(F_k(n,(1-\epsilon)T_n^B) \textit{ has property } \mathcal{A}_B) = 0, \textit{ and}$$
$$\lim_{n\to\infty} \Pr(F_k(n,(1+\epsilon)T_n^B) \textit{ has property } \mathcal{A}_B) = 1.$$

Assuming Lemma 24, we can infer Lemma 23 easily.

Proof (Lemma 23.). Let $r < 2^k \ln 2 - k$. Equations (18) and (19) show that $\rho \mapsto \Lambda_b(1/2,k,\rho)$ is a continuous function. Therefore, for any $\epsilon > 0$ there is a $0 < \delta < 2^k \ln 2 - k - r$ such that $r' = (1+\delta)^2 r$ satisfies

$$\Lambda_b(1/2,k,r') > \Lambda_b(1/2,k,r) - \epsilon.$$

Let $B = \sqrt{\Lambda_b(1/2,k,r')}$. Setting $t = \frac{1}{2}B^n$, the second moment argument from [6] shows in combination with the Paley-Zigmund inequality that

$$\liminf_{n\to\infty} \Pr[F_k(n,r'n) \text{ does not satisfy } \mathcal{A}_B] > 0.$$

Therefore, Lemma 24 implies that $r'n < (1+\delta)T_n^B$ for sufficiently large n. Consequently, for large n we have $rn = (1+\delta)^{-2}r'n < (1+\delta)^{-1}T_n^B$. Hence, Lemma 24 yields

$$\lim_{n\to\infty} \Pr[F_k(n,rn) \text{ does not satisfy } \mathcal{A}_B] = 1.$$

Thus, with probability $1 - o(1)$ the number Z of satisfying assignments of $F_k(n, rn)$ satisfies

$$Z \geq \frac{1}{2}B^n = \frac{1}{2}\Lambda_b(1/2, k, r')^{n/2} \geq \frac{1}{2}(\Lambda_b(1/2, k, r) - \epsilon)^{n/2}.$$

Since this is true for any $\epsilon > 0$, the assertion follows. □

Proof (Lemma 22.). As shown in [6], the solution ϵ to (19) satisfies

$$2^{1-k} + k4^{-k} < \epsilon < 2^{1-k} + 3k4^{-k} \ . \tag{20}$$

Plugging these bounds into (18) and performing a tedious but straightforward computation, we obtain that

$$\frac{1}{2}\ln \Lambda_b(1/2, k, r) \geq \ln 2 + r\left[\ln(1 - 2^{-k}) - k2^{3-2k}\right].$$

Since $r \leq 2^k$, the assertion thus follows from Lemma 23. □

To prove Lemma 24, we need a bit of notation. If ϕ is a formula on a set of variables $y_1, \ldots, y_l$ disjoint from $x_1, \ldots, x_n$, then we let $E_n(\phi)$ denote the set of all formulas that can be obtained from ϕ by substituting l distinct variables among $x_1, \ldots, x_n$ for $y_1, \ldots, y_l$. Moreover, for a formula F on $x_1, \ldots, x_n$ we let $F \oplus \phi = F \wedge \phi^*$, where ϕ^* is chosen uniformly at random from $E_n(\phi)$.

Note that $\mathcal{A}_B$ is a *monotone* property, i.e., if F has the property $\mathcal{A}_B$ and F' is another formula on the variables $x_1, \ldots, x_n$, then $F \wedge F'$ has the property $\mathcal{A}_B$ as well. Therefore, we can use the following theorem from Friedgut [15] to prove by contradiction that $\mathcal{A}_B$ has a sharp threshold. Let $\omega(n) = \lceil \log n \rceil$ for concreteness.

Theorem 9. *Suppose that $\mathcal{A}_B$ does* not *have a sharp threshold. Then there exist a number $\alpha > 0$, a formula ϕ, and for any $n_0 > 0$ numbers $N > n_0$, $M > 0$ and a formula F with variables $x_1, \ldots, x_N$ such that the following is true.*

T1. $\Pr(F \oplus \phi$ *has the property* $\mathcal{A}_B) > 1 - \alpha$.
T2. $\alpha < \Pr(F_k(N, M)$ *has the property* $\mathcal{A}_B) < 1 - 3\alpha$.
T3. *With probability at least α a random formula $F_k(N, M)$ contains an element of $E_N(\phi)$ as a subformula.*
T4. $\Pr(F \wedge F_k(N, 2\omega(N)))$ *has the property* $\mathcal{A}_B) < 1 - 2\alpha$.

In the sequel we assume the existence of α, ϕ, N, M, and F satisfying conditions **T1**–**T3**. To conclude that $\mathcal{A}_B$ has a sharp threshold, we shall show that then condition **T4** cannot hold. Clearly, we may assume that N is sufficiently large (by choosing n_0 appropriately). Let $V = \{x_1, \ldots, x_N\}$.

Lemma 25. *The formula ϕ is satisfiable.*

Proof. Any k-SAT formula that contains at most as many clauses as variables is satisfiable. Hence, to establish the lemma, we will show that the probability Q that $F_k(N, M)$ contains a subformula on l variables with at least l clauses is smaller than α; then the assertion follows from **T3**.

To prove this statement, we employ the union bound. There are $\binom{N}{l}$ ways to choose a set of l variables, and $\binom{M}{l}$ ways to choose slots for the l clauses of the subformula. Furthermore, the probability that the random clauses in these l slots contain only the chosen variables is at most $(l/N)^{kl}$. Hence, the probability that $F_k(N, M)$ has l variables that span a subformula with at least l clauses is at most

$$Q \leq \binom{N}{l}\binom{M}{l}(l/N)^{kl} \leq \left(\frac{e^2 M l^k}{N^2}\right)^l. \tag{21}$$

Further, **T2** implies that $M/N \leq 2^k$, because for $M/N > 2^k$ the expected number of satisfying assignments of $F_k(N, M)$ is less than 1. Thus, assuming that N is sufficiently large, we see that (21) implies $Q \leq (e^2(2l)^k/N)^l < \alpha$, as claimed. □

Thus, fix a satisfying assignment $\sigma : \{y_1, \ldots, y_l\} \to \{0, 1\}$ of ϕ. Then we say that a satisfying assignment χ of F is *compatible* with a tuple $(z_1, \ldots, z_l) \in V^l$ if $\chi(z_i) = \sigma(y_i)$ for all $1 \leq i \leq l$. Furthermore, we call a tuple $(z_1, \ldots, z_l) \in V^l$ *bad* if F has less than $\frac{1}{2}B^N$ satisfying assignments χ that are compatible with $(z_1, \ldots, z_l)$.

Lemma 26. *There are at least $(1 - \alpha)N^l$ bad tuples.*

Proof. The formula $F \oplus \phi$ is obtained by substituting l randomly chosen variables $(z_1, \ldots, z_l) \in V^l$ for the variables $y_1, \ldots, y_l$ of ϕ and adding the resulting clauses to F. Since by **T1** with probability at least $1 - \alpha$ the resulting formula has at most $\frac{1}{2}B^N$ satisfying assignments, a uniformly chosen tuple $(z_1, \ldots, z_l) \in V^l$ is bad with probability at least $1 - \alpha$. Thus, there are at least $(1 - \alpha)N^l$ bad tuples. □

Lemma 27. *With probability at least $1 - \alpha$ a random formula $F_k(N, \omega(N))$ contains l clauses $C_1, \ldots, C_l$ with the following two properties.*

B1. *For each $1 \leq i \leq l$ there is a k-tuple of variables $(v_i^1, \ldots, v_i^k) \in V^k$ such that $C_i = v_i^1 \vee \cdots \vee v_i^k$ if $\sigma(i) = 1$, and $C_i = \neg v_i^1 \vee \cdots \vee \neg v_i^k$ if $\sigma(i) = 0$.*

B2. *For any function $f : [l] \to [k]$ the l-tuple $(v_1^{f(1)}, \ldots, v_l^{f(l)})$ is bad.*

The proof of Lemma 27 is based on the following version of the Erdős-Simonovits theorem(cf. [15, Proposition 3.5]).

Theorem 10. *For any $\gamma > 0$ there are numbers $\gamma', \nu_0 > 0$ such that for any $\nu > \nu_0$ and any set $H \subset [\nu]^l$ of size $|H| \geq \gamma\nu^t$ the following is true. If l k-tuples $(w_1^1, \ldots, w_1^k), \ldots, (w_l^1, \ldots, w_l^k) \in [\nu]^k$ are chosen uniformly at random and independently, then with probability at least γ' for any function $f : [l] \to [k]$ the tuple $(w_1^{f(1)}, \ldots, w_l^{f(l)})$ belongs to H.*

Proof (Proof of Lemma 27). Assuming that N is sufficiently large, we apply Theorem 10 to $\gamma = 1 - \alpha$, $\nu = N$, and the set $H \subset [N]^l$ of bad l-tuples. Then by Lemma 26 we have $|H| \geq \gamma N^l$. Now, consider l random k-clauses $C_1, \ldots, C_l$ over the variable set V chosen uniformly and independently. Let $V_1, \ldots, V_l$ be the k-tuples of variables underlying $C_1, \ldots, C_l$. Then Theorem 10 entails that $V_1, \ldots, V_l$ satisfy condition **B2** with probability at least γ'. Moreover, given that this is the case, condition **B1** is satisfied with probability 2^{-kl}. Therefore, the clauses $C_1, \ldots, C_l$ satisfy both **B1** and **B2** with probability at least $\gamma' 2^{-kl}$. Hence, the probability that $F_k(N, \omega(N))$ does not feature an l-tuple of clauses satisfying **B1** and **B2** is at most $(1 - \gamma' 2^{-kl})^{\lfloor \omega(N)/l \rfloor}$. Since $\omega(N) = \lceil \log N \rceil$, we can ensure that this expression is less than α by choosing N large enough. □

Corollary 3. *With probability at least $1 - \alpha$ the formula $F \wedge F_k(N, \omega(N))$ has at most $\frac{1}{2}k^l \cdot B^N$ satisfying assignments.*

Proof. We will show that if $C_1, \ldots, C_l$ are clauses satisfying the two conditions from Lemma 27, then $F \wedge C_1 \wedge \cdots \wedge C_l$ has at most $\frac{1}{2}k^l B^N$ satisfying assignments. Then the assertion follows from Lemma 27.

Thus, let χ be a satisfying assignment of $F \wedge C_1 \wedge \cdots \wedge C_l$. Then by the **B1** for each $1 \leq i \leq l$ there is an index $f_\chi(i) \in [k]$ such that $\chi(v_i^{f_\chi(i)}) = \sigma(i)$. Moreover, by **B2** the tuple $(v_1^{f_\chi(1)}, \ldots, v_l^{f_\chi(l)})$

is bad. Hence, the map $\chi \mapsto f_\chi \in [k]^l$ yields a bad tuple $(v_i^{f_\chi(i)})_{1\le i\le l}$ for each satisfying assignment. Therefore, the number of satisfying assignments mapped to any tuple in $[k]^l$ is at most $\frac{1}{2}B^n$. Consequently, $F \wedge C_1 \wedge \cdots \wedge C_l$ has at most $\frac{1}{2}k^l \cdot B^n$ satisfying assignments in total. □

Corollary 4. *With probability at least* $1 - \frac{3}{2}\alpha$ *the formula* $F \wedge F_k(N, 2\omega)$ *satisfies* $\mathcal{A}_B$.

Proof. The formula $F^{**} = F \wedge F_k(N, 2\omega)$ is obtained from F by attaching $2\omega(N)$ random clauses. Let $F^* = F \wedge F_k(N, \omega(N))$ be the formula resulting by attaching the first $\omega(N)$ random clauses. Then by Corollary 3 with probability at least $1 - \alpha$ the formula F^* has at most $\frac{1}{2}k^l \cdot B^N$ satisfyng assignments. Conditioning on this event, we form F^{**} by attaching another $\omega(N)$ random clauses to F^*. Since for any satisfying assignment of F^* the probability that these additional $\omega(N)$ clauses are satisfied as well is $(1 - 2^{-k})^{\omega(N)}$, the expected number of satisfying assignments of F^{**} is at most

$$\frac{1}{2}k^l \cdot B^N \cdot (1 - 2^{-k})^{\omega(N)} \le \frac{\alpha}{4} \cdot B^N,$$

provided that N is sufficiently large. Therefore, Markov's inequality entails that

$$\Pr(F^{**} \text{ violates } \mathcal{A}_B | F^* \text{ has at most } \tfrac{1}{2}k^l \cdot B^N \text{ satisfying assignments}) \le \alpha/2.$$

Thus, we obtain

$$\begin{aligned}\Pr(F^{**} \text{ violates } \mathcal{A}_B) \le{}& \Pr(F^* \text{ has more than } \tfrac{1}{2}k^l \cdot B^N \text{ satisfying assignments})\\ &+ \Pr(F^{**} \text{ violates } \mathcal{A}_B | F^* \text{ has at most } \tfrac{1}{2}k^l \cdot B^N \text{ satisfying assignments}) \le 3\alpha/2,\end{aligned}$$

as desired. □

Combining Theorem 9 and Corollary 4, we conclude that $\mathcal{A}_B$ has a sharp threshold, thereby completing the proof of Lemma 24.

B.3 Proof of Theorem 2

Using Theorem 8, we shall establish the following lemma.

Lemma 28. *There exist numbers* $0 < \alpha_1 < \alpha_2 < \frac{1}{3}$, $\lambda > 0$, *and* $\gamma > 0$ *such that a random pair* (F, σ) *chosen from the distribution* $\mathcal{U}_{n,m}$ *has the following two properties w.h.p.*

1. *Any assignment* τ *such that* $\alpha_1 < n^{-1}\mathit{dist}(\sigma, \tau) < \alpha_2$ *satisfies* $H(\tau) \ge \lambda n$.
2. $|\{\tau \in \mathcal{S}(F) : \mathit{dist}(\sigma, \tau) < \beta_2 n\}| < 2^n(1 - 2^{-k})^m \exp(-\gamma n - k2^{3-k}n)$.

Proof (Theorem 2). Let $F = F_k(n, m)$ be a random k-SAT instance. To each assignment $\sigma \in \mathcal{S}(F)$ we assign the set

$$\mathcal{C}_\sigma = \{\tau \in \mathcal{S}(F) : \mathrm{dist}(\sigma, \tau) < \alpha_2 n\}.$$

Due to Lemma 22, a similar argument as in the proof of Theorem 1 in Section A.9 yields Theorem 2. □

Let $\lambda > 0$ be small but fixed. Let $F = F_k(n, m)$ be a random k-SAT formula with $m = rn$ clauses. Then for any $\sigma \in \{0, 1\}^n$ we have

$$n^{-1} \ln \Pr[\sigma \text{ is satisfying}] = r \ln(1 - 2^{-k}),$$

because of the independence of all m clauses. Furthermore, if $\tau \in \{0,1\}^n$ is a second assignment at Hamming distance αn from σ, then

$$n^{-1}\ln \Pr\left[\sigma,\tau \text{ are both satisfying}\right] = r\ln(1-2^{1-k}+2^{-k}(1-\alpha)^k).$$

Indeed, there is a function $\Psi(\lambda)$ such that $\lim_{\lambda\to 0}\Psi(\lambda)=0$ and

$$n^{-1}\ln \Pr\left[H(\sigma)=0 \wedge H(\tau)\leq \lambda n\right] = r\left[\ln(1-2^{1-k}+2^{-k}(1-\alpha)^k)+\Psi(\lambda)\right].$$

Let $X_{\alpha,\lambda}$ signify the number of assignments τ at Hamming distance αn from σ such that $H(\tau)\leq \lambda n$.

Lemma 29. *There is a number $0<\alpha^*<1/3$ such that for sufficiently small $\lambda>0$ we have*

$$n^{-1}\ln \mathrm{E}\left[X_{\alpha^*,\lambda}|\sigma \textit{ is satisfying}\right] < -k2^{3-k}.$$

Proof. There are $\binom{n}{\alpha n}$ ways to choose an assignment τ at Hamming distance αn from σ. Therefore, due to the formulas derived above, we have

$$n^{-1}\ln \mathrm{E}\left[X_{\alpha,\lambda}|\sigma \text{ is satisfying}\right] = -\alpha\ln\alpha-(1-\alpha)\ln(1-\alpha)+r\left[\ln\left(1-\frac{1-(1-\alpha)^k}{2^k-1}\right)+\Psi(\lambda)+o(1)\right].$$

Setting $\alpha^*=(k\ln k)^{-1}$ and simplifying, we obtain the assertion. □

Corollary 5. *There are numbers $\lambda>0$ and $0<\alpha_1<\alpha_2<\frac{1}{3}$ such that with probability at least $1-o(\exp(-k2^{3-k}n)$ in a pair $(F,\sigma)\in\Lambda_{n,m}$ chosen from the distribution $\mathcal{P}_{n,m}$ there is no assignment τ such that such that $H(\tau)<\lambda n$ and $\alpha_1<n^{-1}$dist$(\sigma,\tau)<\alpha_2$.*

Proof. If (F,σ) is chosen from $\mathcal{P}_{n,m}$, then F is distributed as a random formula $F_k(n,m)$ given that σ is a satisfying assignment. Therefore, the corollary follows from Lemma 29 and Markov's inequality, where we use the fact that $\alpha\mapsto n^{-1}\ln \mathrm{E}\left[X_{\alpha,\lambda}|\sigma \text{ is satisfying}\right]$ is a continuos function. □

Furthermore, the following estimate has been established in [7].

Lemma 30. *We have* $\max_{0\leq\alpha\leq\frac{1}{3}} n^{-1}\ln \mathrm{E}\left[X_{\alpha,\lambda}|\sigma \textit{ is satisfying}\right] < \ln 2 + r(1-2^{-k}) - 2\exp(k2^{3-k})$.

Finally, Lemma 28 follows from Theorem 8 in combination with Corollary 5 and Lemma 30.

B.4 Proof of Theorem 4 (*k*-SAT)

If F is a k-SAT formula and σ an assignment, then we say that a variable x *supports* a clause C if changing the value of x would render C unsatisfied. Suppose that k is sufficiently large and $(1+\varepsilon)2^k k^{-1}\ln k < r = m/n < (1-\varepsilon)2^k\ln 2$. Let $\gamma,\delta>0$ be sufficiently small numbers.

Lemma 31. *A pair (F,σ) chosen from $\mathcal{U}_{n,m}$ has the following property w.h.p.*

There is a set U of at least $(1-\delta)n$ variables such that each variable in U supports $\gamma\ln k$ clauses e that contain no variable from $V\setminus U$. (22)

Proof (Theorem 4). Let $\zeta > 0$ signify a sufficiently small constant. Let (F, σ) be chosen from the distribution $\mathcal{U}_{n,m}$. We may assume that the random pair (F, σ) satisfies (22). Moreover, a 1st moment computation shows that the random formula F has the following property w.h.p.

> There is no set Z of variables of size $|Z| \leq \zeta n$ such that F features at least $|Z|\gamma \ln k$ clauses e that contain at least two variables from Z. (23)

Now, assume for contradiction that there is a satisfying assignment τ of F such that the set $Z = \{v \in U : \tau(v) \neq \sigma(v)\}$ has size $1 \leq |Z| \leq \zeta n$. Each $v \in Z$ supports in σ at least $\gamma \ln k$ clauses e that contain no variable from $V \setminus U$. Since these clauses e are satisfied in τ, although $\tau(z) = 0$, each such e contains another variable $w \neq v$ from Z. Hence, F contains at least $|Z|\gamma \ln k$ clauses e containing at least two variables from Z, in contradiction to (23). □

Lemma 31 is an immediate consequence of Theorem 8 and the following lemma.

Lemma 32. *A pair (F, σ) chosen from $\mathcal{P}_{n,m}$ has the property (22) with probability $1 - o(\exp(-k2^{3-k}n))$.*

Proof. We may assume that $r = m/n = (1 + \varepsilon)2^k k^{-1} \ln k$ for a fixed $\varepsilon > 0$. Moreover, without loss of generality, we may assume that the assignment σ sets all variables $V = \{x_1, \ldots, x_n\}$ to true. Let F denote a random formlua with m clauses satisfied by σ, and let Ξ signify the set of all uniquely satisfied clauses of F. Consider the following process.

1. Let Z_0 be the set of all variables x that support fewer than $2\gamma \ln k$ clauses.
2. Let $Z = Z_0$. While there is a variable $x \in V \setminus Z$ that supports at least $\gamma \ln k$ clauses from Ξ that contain a variable from Z, add x to Z.

The *expected* number of uniquely satisfied clauses is at least $k2^{-k}m \geq (1 + \varepsilon)n \ln k$. Hence, each variable is expected to support at least $(1 + \varepsilon) \ln k$ clauses. Therefore, if $\gamma > 0$ is sufficiently small, then there is a contant $\beta > 0$ such that the probability that a variable x supports fewer than $2\gamma \ln k$ clauses is at most $k^{-1-\beta}$. Hence, by Chernoff bounds we have $|Z_0| \leq nk^{-1-\beta/2}$ with probability at least $1 - o(\exp(-k2^{3-k}n))$.

Thus, assume that $|Z_0| \leq nk^{-1-\beta/2}$. We claim that then the final set Z resulting from Step 2 has size at most $|Z| \leq 2nk^{-1-\beta/2}$. For assume that $|Z| > 2nk^{-1-\beta/2}$. Then Step 2 removed at least $|Z|/2$ variables, whence there are at least $\gamma \ln k|Z|/2$ clauses $e \in \Xi$ that contain two variables from Z. However, a standard 1st moment argument shows that the probability that there exists a set Z with this property is $o(\exp(-k2^{-k}n))$. Hence, with probability at least $1 - o(\exp(-k2^{-k}n)$ we have $|Z| \leq 2nk^{-1-\beta/2}$. Setting $U = V \setminus Z$ concludes the proof. □