LA-UR- 09 - 0115

Title:	Quiet Planting in the Locked Constraints Satisfaction Problems
Author(s):	Lenka Zdeborova/226997/T-4/LANL Florent Krzakala/229590/T-4/LANL
Intended for:	SIAM Journal on Discrede Mathematics

Los Alamos
NATIONAL LABORATORY
——— EST.1943 ———

QUIET PLANTING IN THE LOCKED CONSTRAINTS SATISFACTION PROBLEMS

LENKA ZDEBOROVA* AND FLORENT KRZAKALA†‚*

Abstract. We study the planted ensemble of locked constraint satisfaction problems. We describe the connection between the random and planted ensembles. The use of the cavity method is combined with arguments from reconstruction on trees and first and second moment considerations; in particular the connection with the reconstruction on trees appears to be crucial. Our main result is the location of the hard region in the planted ensemble, thus providing hard satisfiable benchmarks. In a part of that hard region instances have with high probability a single satisfying assignment.

Key words. Constraint Satisfaction Problems, Planted Ensemble, Belief Propagation, Reconstruction on Trees, Instances with a Unique Single Assignment.

AMS subject classifications. 90C27 68Q25 05C80

Constraint Satisfaction Problems (CSP) are very general in nature: Consider a set of N discrete variable and a set of M Boolean constraints; the problem consists is finding a configuration of variables that satisfies all the constraints or in proving that no such configuration exists. As such, CSPs are subject of interests in many different fields such as computer science, discrete mathematics, physics, engineering and computational biology. Random ensembles of CSPs have proved to be a fertile source of research activity; as hard benchmarks they serve for testing new algorithmic ideas [4, 27], they are used to create efficient coding schemes [10, 11], to model complex glass forming liquids [3, 19], or to understand the origin of average computational hardness [26, 34]. Combining know-how from many branches of mathematics, computer science and statistical physics seems to be fruitful for understanding of these stunning objects with very rich behavior.

The most commonly studied random ensembles of CSPs are created by choosing the graph of variables and constraints as a random bipartite graph with a certain left and right degree distributions. Another natural way of creating a random instance, called planting, is to first assign a configuration to variables and then to choose only constraints compatible with such a configuration. Both these ensembles can be useful to mimic instances created in some practical applications. In particular planted instances maybe be created in adaptive situations when only constraints satisfied by the current state of variables can b added.

By planting we create by definition a satisfiable instance. Such instances are in particular useful as benchmarks to evaluate the performance of incomplete solvers, such as stochastic local search [30]. Based on the example of the planted K-satisfiability problem it is often anticipated that the planted ensemble is algorithmically easier than the random one as a bias towards the planted assignment is created in the graph. Also, for most of the studied problems it was proven that at large density of constraints is it indeed easy to find an satisfying assignment near to the planted one, see e.g. [2, 5, 9]. On the other hand if the planted ensemble would be algorithmically hard in some region of the parameters than these instances could serve as one-way functions and have application in cryptography. Yet, compared to the random ensemble relatively

1

little is known about the existence, size and properties of algorithmically hard regions in the planted ensemble.

In this paper we study a way of planting an assignment which changes only in a minimal way the properties of the random ensemble. We call this a "quiet planting". The concept of quiet planting was introduced in [20], but some of its consequences were already used as a tool for proofs in [1]. Both these works were mainly concentrated on the coloring problems (and the hyper-graph bi-coloring). In this work we will focus on quiet planting in the so-called locked CSP, introduced recently in [34, 33]. The locked CSPs have very interesting phase diagram which description-wise is much simpler than the one of graph coloring or K-satisfiability. On the other hand algorithmically they are much harder and the boundaries between the easy and hard regions are, unlike in the coloring or K-satisfiability, well understood (at least on the heuristic level of the cavity method). This special behavior stems from the fact that in the locked problems the space of solutions consists of separated points.

Here, we combine the idea of quiet planting with the special behavior of the locked CSPs and obtain random CSPs ensembles with very interesting properties. The paper is organized as follows: In Sec. 1 we summarize our main results in the context of related works, in Sec. 2 we introduce the necessary definitions and notations and in Sec. 3 we summarize the phase diagram of the locked problems derived in [34, 33]. In Sec. 4 we argue about the equivalence between the random and planted ensembles based on the cavity equations and on a second moment argument. In Sec. 5 we describe the phase where instances of our problems have with large probability a single solution. Finally, in Sec. 6 we discuss the algorithmical hardness of the planted instances.

1. Main results and related works. The results of this paper apply to the *factorized locked* CSP, see Defs. 2.2, 2.4. We list in five points the most important contributions of the present article:

(i) The idea of *quiet* planting is to plant a configuration which will have all the properties of a typical satisfying assignment on the resulting graph. Such a problem is very closely related to the reconstruction on trees [22] where we broadcast an assignment taken uniformly at random from all the satisfying ones. The quiet planting is always possible on a tree, as exact marginal probabilities can be obtained via the belief propagation algorithm. On random graphs this can be achieved asymptotically if the belief propagation has a stable uniform fixed point, which this is the case on the factorized CSPs, see Def. 2.4.

(ii) All the properties of the planted ensemble created via quiet planting can be deduced from the properties on the purely random ensemble, keeping in mind that in the planted ensemble of the locked problems there is the planted assignment on top of the other satisfying ones, which we would found on the random ensemble. Thus, among others, in the satisfiable phase the random and planted ensembles are asymptotically equivalent, see Def. 4.1. Such equivalence can also be established rigorously based on a second moment argument as in [1], and the result that in the factorized locked models the second moment is able to pin the satisfiability threshold sharply.

(iii) Next to the interesting conceptual results, the most important practical result is establishing the region where the instances from the planted ensemble are computationally hard. It is anticipated that a easy-hard-easy pattern appears as the constraint density is increased. We conjecture that the two boundaries

on the hard phase correspond to two different reconstruction thresholds – the onset of hardness coincides with the *small noise* reconstruction with threshold [33], called the dynamical transition in the physics literature [18], and the end of the hard region is given by the threshold for the *robust* reconstruction [12]. This last point point also corresponds to the Kesten-Stigum bound for the canonical reconstruction on trees [15, 16] and to the spin glass local instability in the purely random ensemble [25]. We also show that outside the hard region algorithms based on belief propagation are able to find solutions efficiently. In particular in the high average degree easy region, the Belief Propagation (BP) algorithm converges directly to the planted solution.

(iv) Given we have located the values of parameters where the instances of the planted ensemble are hard, these can serve as very challenging satisfiable benchmarks. Such benchmarks are in particular interesting for evaluation of the incomplete solvers, as most of the complete solvers use a variant of the unit clause propagation and as locked constraints produce relatively many implications they might not seem to be that hard for branch and bound based solvers.

(v) As the hard region in the planted ensemble extends to the phase which is unsatisfiable in the random ensemble, we show that in that phase the planted instances have with high probability a single satisfying assignment (or a pair of them in case a global symmetry is present). Moreover depending on the constraint density these unique satisfying assignment (USA) instances can be found in the hard or in the easy region. Some USA instances are extensively used in evaluation of quantum algorithms, see e.g. [32, 8]. In the current works these instances are, however, generated with exponential cost, and thus their classical computational hardness is almost impossible to be evaluated.

Large part of our results are based on the heuristic cavity method approach [24]. Based on computations of the second moment we were able to prove part of our results for the R-in-K SAT problems on random regular graphs. This includes the equivalence between the planted and random ensembles in the satisfiable phase. Location of the satisfiability transition. And uniqueness on the satisfying assignment in the unsatisfiable phase. Extending these proofs to the other locked factorized CSP should be possible, however, more involved.

TABLE 1.1

Sketchy summary of the properties of the different phases in the random ensemble of the factorized locked problems, the parameter l is the average number of constraints in which a variable appears. The three thresholds l_d, l_s and l_l are defined in detail later in the paper.

RANDOM	$l < l_d$	$l_d < l < l_s$	$l_s < l < l_l$	$l_l < l$
BP, any init.	converges	converges	converges	does not
fixed point	uniform	uniform	uniform	×
solutions	exponentially	exponentially	none	none
finding solution	easy	hard	×	×
reconstruction	not possible	possible	×	×

2. Definitions and notations. In this section we specify the class of constraint satisfaction problems to which our results apply. We state several necessary definitions, the crucial notions will be the definition of a *locked* [34] and *factorized* constraint satisfaction problem. It is only on the factorized problems where there is a very close

TABLE 1.2
The same as Tab. 1.1 for the random planted ensemble, its definition is given in Sec. 4.

LOCKED	$l < l_d$	$l_d < l < l_s$	$l_s < l < l_l$	$l_l < l$
BP, random init.	converges	converges	converges	converges
fixed point	uniform	uniform	uniform	planted
BP, planted init.	converges	converges	converges	converges
fixed point	uniform	planted	planted	planted
solutions	exponentially	exponentially	one/two	one/two
finding solution	easy	hard	hard	easy
reconstruction	not possible	possible	possible	possible
robust recons.	not possible	not possible	not possible	possible

relation between the usual random and the planted ensemble, as discussed in [20]. It is also the fact that in the locked problems solutions are far from each other [34] which makes them particularly interesting for considerations in this context.

DEFINITION 2.1. *A constraint a containing K variables, the domain of each variable being X, is a function from X^K to $\{0, 1\}$. If the function evaluates to 1 (0) we say that constraint a is satisfied (not satisfied). A constraint is* locked *if and only if there are no two satisfying assignments of variables which would differ in a single value (out of the K ones).* In this paper we will consider for concreteness binary variables, that is $X = \{0, 1\}$. But the results are generalizable to general domain size.

DEFINITION 2.2. *A constraint satisfaction problem consists in deciding if there exists a configuration of N variables which would satisfy simultaneously a set of M constraints. A constraint satisfaction problem is called* locked *if and only if all the M constraint are locked and each of the N variables belong to at least two different constraints.*

We shall illustrate our findings on the so called occupation constraint satisfaction problems [28, 33].

DEFINITION 2.3. *In* occupation problems *every constraint a depends only on the sum of variables it contains. Thus every occupation constraint containing K_a variables can be characterized by a binary $K_a + 1$ component vector A_a such that the constraint is satisfied if and only if the sum r of the K_a variables is such that $A_a(r) = 1$.*

An occupation constraint a is locked if and only if for all $i = 0, \ldots, K_a - 1$ we have $A_a(i)A_a(i+1) = 0$. We will consider occupation problems where every constraint contains K variables and is given by the same vector A. To give and example for the notation vector $A = 0100$ would correspond to the 1-in-3 SAT problem (exact cover), which is locked, or $A = 0110$ would correspond to the hyper-graph bi-coloring problem, which is not locked. More examples can be found in [33]. For problems which do not have other name established in the literature we use the notation i-or-j-...-in-K SAT for vector A with non-zero components A_i, A_j, etc.

Let us now write the belief propagation (BP) equations [29, 21, 23] for the occupation constraint satisfaction problems. The basic quantities in BP are messages. We define $\psi_{s_i}^{a \to i}$ as the probability that the constraint a is satisfied, conditioned to the fact that the value of variable i is s_i. Belief propagation equations approximate these probabilities by assuming that the factor graph [21] underlying the CSP is a tree

$$\psi_{s_i}^{a \to i} = \frac{1}{Z^{a \to i}} \sum_{\{s_j\}} \delta_{A(s_i + \sum_j s_j), 1} \prod_{j \in \partial a - i} \prod_{b \in \partial j - a} \psi_{s_j}^{b \to j}, \qquad (2.1)$$

where $Z^{a \to i}$ is a normalization constant assuring $\psi_1^{a \to i} + \psi_0^{a \to i} = 1$. Fig. 2.1 shows
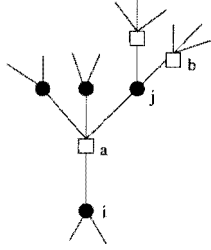

FIG. 2.1. *Part of the factor graph to illustrate the meaning of indices in the belief propagation equations (2.1).*

the corresponding part of the factor graph. The belief propagation estimate of the probability that a variables i is occupied is then

$$\chi^i = \frac{\prod_{a \in \partial i} \psi_1^{a \to i}}{\prod_{a \in \partial i} \psi_1^{a \to i} + \prod_{a \in \partial i} \psi_0^{a \to i}}, \tag{2.2}$$

Note that if an assignment $\{\sigma\}$ is a solution of the problem then $\psi_{\sigma_i}^{a \to i} = 1$, $\psi_{\neg \sigma_i}^{a \to i} = 0$ is a fixed point of the BP equations (2.1).

DEFINITION 2.4. *A given instance on a constraint satisfaction problem is factorized if and only if the belief propagation equations initialized randomly converge almost surely (with probability approaching one as the number of variables $N \to \infty$) to a uniform fixed point, i.e., the value of $\psi^{a \to i}$ does not depend on the indexes a and i.* Note that it is a non-trivial task to check if a problem satisfies this definition, and it depends on the degree distribution. In what follows we will state two large groups of locked problems which are factorized. Evaluation of the condition in Def. 2.4 is based on the cavity solution of the problem.

Our derivations and proofs will hold only asymptotically ($N \to \infty$) on graphs which are locally tree like, that is the shortest loop going trough a random variables has diverging length as $N \to \infty$. Families of sparse random graphs, i.e. the degree distribution of variables $Q(l)$ does not depend on N, are locally tree like. Examples of the variable degree distribution we will be using are:

- Regular $Q(l) = \delta_{L,l}$.
- Truncated Poisson $Q(0) = Q(1) = 0$, $Q(l) = c^l/[(e^c - 1 - c)l!]$ for $l \geq 3$. The average degree in this case is $\bar{l} = c(1 - e^{-c})/[1 - (1 + c)e^{-c}]$.
- Poisson $Q(l) = e^{-c}c^l/l!$.

To generate random graphs with a given variable degree distribution we generate first a degree sequence from $Q(l)$ such that it is compatible with number KM and then consider a random permutation of KM numbers (which does not create double edges).

As follows from the analysis of the locked problems in [34], a locked CSP can be factorized in two different cases

(a) On the regular graphs, that is every variable is contained in L constraints.

On the regular graphs the fixed point of the BP equations then satisfies

$$\psi_0 = \frac{1}{Z} \sum_{r=0}^{K-1} \delta_{A(r),1} \binom{K-1}{r} \psi_1^{(L-1)r} \psi_0^{(L-1)(K-1-r)} , \qquad (2.3)$$

$$\psi_1 = \frac{1}{Z} \sum_{r=0}^{K-1} \delta_{A(r+1),1} \binom{K-1}{r} \psi_1^{(L-1)r} \psi_0^{(L-1)(K-1-r)} , \qquad (2.4)$$

where Z is the normalization. For the probability that a variable in occupied one has in this case

$$\chi = \frac{\psi_1^L}{\psi_1^L + \psi_0^L} , \qquad (2.5)$$

Let us call x_r the probability that a constraint contains r occupied variables then

$$x_r = \frac{\binom{K}{r} \delta_{A(r),1} \psi_1^{r(L-1)} \psi_0^{(K-r)(L-1)}}{\sum_{t=0}^{K} \binom{K}{t} \delta_{A(t),1} \psi_1^{t(L-1)} \psi_0^{(K-t)(L-1)}} . \qquad (2.6)$$

(b) For the balanced locked problems [33], that is when the vector A is symmetric, $A(i) = A(K - i)$ for all $i = 0, \ldots, K$ and this 0-1 symmetry is not spontaneously broken (no ferromagnet-like transition). In this case the belief are $\psi_1 = \psi_0 = \chi = 1/2$. For the probability that a constraint contains r occupied variables we have here

$$x_r = \frac{\binom{K}{r} \delta_{A(r),1}}{\sum_{t=0}^{K} \binom{K}{t} \delta_{A(t),1}} . \qquad (2.7)$$

In particularly simple case of (a) is the R-in-K SAT where $1 \le R \le K/2$. If every variable has L connections and every constraint have to contain exactly R occupied variables then the number of occupied variables is exactly MR/L, and thus $\chi = R/K$.

3. Basic properties of the random factorized locked problems. As was shown in [34, 33] on the locked problems the belief propagation gives asymptotically exact values of beliefs ψ. And thus also the Bethe entropy [21, 23] gives the asymptotically correct logarithm of the number of solutions. It reads for the balanced locked problems

$$s\left(\bar{l}\right) = \log 2 + \frac{\bar{l}}{K} \log \left[2^{-K} \sum_{r=0}^{K} \delta_{A(r),1} \binom{K}{r} \right] \qquad (3.1)$$

and for the fixed degree of variables

$$s(L) = \frac{L}{K} \log \left[\sum_{r=0}^{K} \delta_{A(r),1} \binom{K}{r} \psi_1^{(L-1)r} \psi_0^{(L-1)(K-r)} \right] - (L - 1) \log \left[\psi_0^L + \psi_1^L \right] , \quad (3.2)$$

where ψ_1, ψ_0 is a fixed point of eqs. (2.3-2.4). This entropy simplifies further for the R-in-K SAT on regular graphs where we get and explicit formula

$$s(L) = \frac{L}{K} \log \binom{K}{R} - (L - 1) H \left(\frac{R}{K} \right) , \qquad (3.3)$$

where $H(x) = -x \log x - (1-x) \log (1-x)$ is the entropy function. The satisfiability transition l_s is then defined by

$$\text{satisfiability threshold } l_s: \quad s(l_s) = 0. \tag{3.4}$$

For $\bar{l} < l_s$ the problem has almost surely exponentially many solution, the exponent given by $s(\bar{l})$, whereas for $\bar{l} > l_s$ the problem almost surely does not have any solution.

Authors of [34, 33] also argued about existence of a second critical point in the locked problems, l_d. This critical point separates regions where for $\{\sigma\}$ being a satisfying assignment the $\psi_{\sigma_i}^{a \to i} = 1$, $\psi_{\neg \sigma_i}^{a \to i} = 0$ is a *stable* fixed point of the BP equations (2.1) if and only if $\bar{l} > l_d$. That is if an infinitesimal perturbation is introduced to these messages, the iteration of (2.1) goes back to the solution-related fixed point. Equivalently only for $\bar{l} > l_d$ each solutions has a corresponding fixed point not only at zero temperature, eq. (2.1), but also at infinitesimal temperature, that is if constraint can be violated with infinitesimal probability. Authors of [34, 33] also conjectured that for $\bar{l} > l_d$ a typical solution does not have solutions up to an extensive Hamming distance, whereas for $\bar{l} < l_d$ there are solutions at non-extensive Hamming distance.

For the locked problems on regular graphs it is always $2 < l_d < 3$, in other words at $L = 2$ the systems are in the non-separated phase, and for $L \geq 3$ the solutions are always separated and solution-corresponding fixed point is stable etc. For the balanced locked problems the expression for l_d follows from [33]

$$\frac{l_d}{Q(2)} = 2(K-1) - 2 \frac{\sum_{r=1}^{K-2} r \binom{K-1}{r} \delta_{A(r+1),1} \delta_{A(r),0} \delta_{A(r-1),0}}{\sum_{r=0}^{K-2} \delta_{A(r+1),1} \binom{K-1}{r}}, \tag{3.5}$$

where $Q(2)$ is the fraction of variables of degree two. It thus follows that whenever the degree of every variable is larger or equal to three the system is in the phase where solutions are separated.

There is a deep connection between this dynamical threshold and the reconstruction problem. In the reconstruction problem one creates a tree with the same degree properties as the graph. Then we broadcast a satisfying assignment chosen uniformly at random from all the possible ones. Finally the reconstruction problem consist of deciding whether the assignment on leaves of the tree contains some information about the value of the root. In the locked problems the values of the root is always uniquely implied by the values of the leaves. However, if an infinitesimal noise is introduced on the leaves then there is no information left if an only if $\bar{l} < l_d$. Thus l_d was called the *small noise* reconstruction threshold in [33].

Note also at this point that the survey propagation equations [26] in the locked problems have a solution given by the solution of the belief propagation, see [33]. This is a reason why in this paper we do not discuss the survey propagation.

To summarize the random locked factorized problems are in the non-clustered phase for $\bar{l} \leq l_d$, which was shown to be algorithmically easy in [34, 33]. For $l_d \leq \bar{l} \leq l_s$ the space of solutions is clustered and it is hard to find any of the solutions. For $\bar{l} \geq l_s$ no more solutions exist.

4. Equivalence of the random and planted ensembles for $\bar{l} < l_s$. The planted ensemble of graphs, which is the main subject of the present paper, is created in the following way:
   (i) Make each of N variables occupied with probability χ (2.5), call the number of occupied variables N_1.

TABLE 3.1

The critical values for all the balanced locked problems up to $K = 8$ on the regular and truncated Poissonian ensemble. The integer value L_s (resp. L_l) is defined as the first larger or equal to l_s (resp. l_l), the stars denote that $L_s = l_s$ (resp. $L_l = l_l$). The sign 'x' means that the problem ceases to be balanced before the instability arises. We remind here that the vector A codes for what are the allowed sums of variables around a constraint.

A	L_s	L_l	c_d	c_s	c_l	l_d	l_s	l_l
00100	3	4*	1.256	1.853	2.821	2.513	2.827	3.434
0001000	4	6*	1.904	3.023	4.965	2.856	3.576	5.144
000010000	5	8*	2.337	3.942	6.994	3.116	4.276	7.039
5-in-10	5	10*	2.660	4.794	8.999	3.325	4.944	9.009
6-in-12	6	12*	2.918	5.455	11.00	3.502	5.586	11.00
01010	4*	∞	1.904	3.594	∞	2.856	4	∞
0101010	6*	∞	2.660	5.903	∞	3.325	6	∞
010101010	8*	∞	3.132	7.978	∞	3.654	8	∞
0010100	6	46*	2.561	5.349	45.00	3.260	5.489	45.00
000101000	7	29*	2.975	6.650	28.00	3.542	6.708	28.00
001010100	8	> 100	3.110	7.797	> 100	3.638	7.822	> 100
010010010	6	x	2.173	4.896	x	3.014	5.083	x

TABLE 3.2

The critical values for all the regular (non-balanced) locked problems up to $K = 6$. The integer value L_s (resp. L_l) is defined as the first larger or equal to l_s (resp. l_l), the stars denote that $L_s = l_s$ (resp. $L_l = l_l$).

A	L_s	L_l
0100	3	3*
01000	3	4*
010000	3	5*
0100000	3	6*
001000	4	5*
0010000	4	6*

A	L_s	L_l
010100	5	> 50
0101000	6	> 50
010010	4	10
0100100	4	14
0100010	4	7

(ii) Choose a degree sequence from the probability distribution $Q(l)$ in such a way that $KM = \sum_{i=1}^{N} l_i$.

(iii) For each constraint, according to probabilities x_r (2.6,2.6), choose the number r_a of occupied variables to which it is connected. Repeat until $\sum_{a=1}^{M} r_a = \sum_{i=1}^{N_1} l_i$, here i are indexes of the occupied variables.

(iv) Choose a random permutation of $\sum_{a=1}^{M} r_a$ numbers, choose a random permutation of $MK - \sum_{a=1}^{M} r_a$ numbers, and connect the constraints to variables according to that permutations. Repeat until there are no double edges.

Note that there are several models how to plant a solution, we could have chosen any other which behaves the same as the above one in the thermodynamic limit.

DEFINITION 4.1. *Two ensembles of random graphs are* asymptotically equivalent *if and only if in the thermodynamic limit every property which is almost surely true on a graph from one ensemble is also almost surely true on a graph from the other ensemble.*

DEFINITION 4.2. *The planting is called* quiet *if the corresponding planted and random ensembles are equivalent.* In this section we argue about the asymptotic

equivalence of the planted and the random ensembles for the factorized locked problems in the range of parameters corresponding to the satisfiable phase on the random ensemble. We will use two different approaches, the first one based on analogy between planting and the reconstruction on tress [22] and the other one on the computing of the second moment method and using theorems from [1].

4.1. Planting and the reconstruction on trees. The formulas created by planting a solution are locally tree-like, just as the random formulas. Thus the planted problems can be solved via the cavity method [24], the first step in that solution is writing the belief propagation equations. The BP equations on a planted formulas are the same as on a random formula, thus given by eq. (2.1). The difference comes when one wants to find the typical fixed point without resorting to the single formula experiments. A standard way to do this is the population dynamics technique [24], where the distribution of messages over edges $P(\psi)$ is represented by a pool drawn from the distribution. On the planted formulas one has to distinguish if the value planted on the corresponding variable was 0 or 1. In the population dynamics one thus needs to keep two populations, one $P_1(\psi)$ (second $P_0(\psi)$) representing the messages $\psi^{a \to i}$ for variables i where we planted value 1 (0). The closed iterative equation for the two distributions read

$$P_s(\psi) = \sum_{\{l_i\}} q(\{l_i\}) \sum_{r=0}^{K} \frac{y_r(s)}{\binom{K-1}{r-s}} \sum_{\{s_i\}} \delta_{r, \sum_i s_i + s} \int \prod_{i=1}^{K-1} \prod_{j=1}^{l_i} \mathrm{d}P_{s_i}(\psi^j)\, \delta(\psi - \mathcal{F}(\{\psi^j\}))\,,$$

$$(4.1)$$

where the $y_r(s)$ is the conditional probability that in the planted solution there is $r - s$ occupied variables within $K - 1$ ones, conditioning on the fact that the Kth variable has value s.

$$y_r(0) = \frac{(K - r)\, x_r}{\sum_{t=0}^{K}(K - t)\, x_t}\,, \qquad y_r(1) = \frac{r\, x_r}{\sum_{t=0}^{K} t\, x_t}\,, \qquad (4.2)$$

where the x_r are the probabilities that a constraint contains r variables occupied in the planted solutions, eqs. (2.6,2.7). The $q(\{l_i\}) = \prod_{i=1}^{K-1}(l_i + 1)Q(l_i + 1)/\bar{l}$ represents the distribution of $K - 1$ excess degree, i.e. number of edges coming from a size of a random edge. And the sum over $\{s_i\}$ is over all values of variables such that $r - s$ of them are occupied. The function $\mathcal{F}(\{\psi^j\})$ is defined by the belief propagation eq. (2.1).

After a detailed inspection of these equations we realize that the very same equations would be obtained from analysis of reconstruction on trees [22]. Reconstruction on trees is defined in the following way: First construct a rooted tree factor-graphs with the degree distribution of variables $Q(l)$ and degree of constraints K. Consider all the constraints to be given by the vector A (nonzero elements of A are possible numbers of occupied variables around the constraint). Second consider a random satisfying configuration chosen uniformly from all the possible ones, call it broadcasted configuration. The problem of reconstruction consists in deciding what is the information the values of the leaves of this tree carry about the value of the root? If this information is nonzero then we say that reconstruction is possible if the information is zero that the reconstruction is not possible.

To generate a configuration uniformly at random over all possible ones we choose a values of the root according to (2.5) and then broadcast a configuration towards leaves, choosing the number of occupied variables according to probabilities (4.2).

Following the derivation of [22] we obtain eq. (4.1). In the reconstruction context the $P_s(\psi)$ is interpreted as the probability distribution of messages ψ going from clauses to variables which had value s in the broadcasted configuration. Authors of [22] also pointed out the equivalence between eq. (4.1) and the formalism of one-step replica symmetry breaking. A lot is known about the solutions of this equation.

In the case of factorized locked problems eq. (4.1) can have only two possible fixed points. The first possible fixed point is equivalent to the solution of the BP equations and reads

$$P_0(\psi) = P_1(\psi) = \delta\left[\psi - \begin{pmatrix} \psi_1 \\ \psi_0 \end{pmatrix}\right], \tag{4.3}$$

where ψ_0, ψ_0 is the factorized fixed point of the BP equations (2.1). The second fixed point is given by the planted configuration

$$P_0(\psi) = \delta\left[\psi - \begin{pmatrix} 0 \\ 1 \end{pmatrix}\right], \qquad P_1(\psi) = \delta\left[\psi - \begin{pmatrix} 1 \\ 0 \end{pmatrix}\right]. \tag{4.4}$$

The reconstruction on trees is possible if and only if this second fixed point is stable under perturbations, according to [33] this happens for $\bar{l} > l_d$, where l_d was defined in eq. 3.5. Both these fixed points have physical meaning also on the planted ensemble and the consequence of their existence is that all the properties of the space of solution in the planted ensemble can be induced from the fact that is looks like union of the space of solution in the purely random ensemble and the planted configuration.

Thus in particular in the region $\bar{l} < l_s$ the planted ensemble is asymptotically equivalent to the purely random ensemble, as for $\bar{l} < l_s$ there are exponentially many solutions and the existence of the planted configuration does not change their properties. For $\bar{l} > l_s$, on the other hand, this means that there exists only the single planted solution (pair of solutions in case of balanced problem a couple of solutions).

4.2. Second moment argument. The conclusion of the previous paragraph can also be obtained in a different and more rigorous, but maybe less intuitive, way. The asymptotic equivalence of the planted and purely random ensembles follows from the theorem of [1]

THEOREM 4.3 (Achlioptas and Coja-Oghlan [1]). *If for a given set of parameters (vector A, constraint degree K, variables degree distribution $Q(l)$, $N \to \infty$) the second moment of the number of solutions in the purely random ensemble is smaller than some constant C times the square of the first moment, i.e. $\mathbb{E}(\mathcal{N}^2) > C\,[\mathbb{E}(\mathcal{N})]^2$, then the planted and random ensembles are asymptotically equivalent for that set of parameters.* Proof of this theorem can be found in [1], and its consequences are discussed for the coloring and bi-coloring of factor-graphs problems.

Note that in the factorized locked problems the following proposition holds

PROPOSITION 4.4. *The relation $\mathbb{E}(\mathcal{N}^2) > C\,[\mathbb{E}(\mathcal{N})]^2$ holds in all the factorized locked problems on the purely random ensemble as long as $\bar{l} \leq l_s$.* The first and second moment of the number of solutions in the occupation problems has been computed for a general degree distribution in [33]. It has been also argued in [33] that the above proposition holds in the balanced locked problems. Here we illustrate that it also holds in the R-in-K SAT on random L-regular graphs for $L < l_s$. The first moment entropy defined as $s_{1st} = \log\mathbb{E}(\mathcal{N})/N$ is in the R-in-K SAT on random L-regular graphs given by eq. (3.3). The second moment entropy $s_{2nd} = \log\mathbb{E}(\mathcal{N}^2)/N$ is given

by $s_{2\mathrm{dn}} = \max_t s_{2\mathrm{nd}}(t)$ where

$$s_{2\mathrm{nd}}(t) = \frac{L}{K} \log \left\{ K! \sum_{s=0}^{R} \frac{\left[\left(\frac{tR}{K}\right)^s \left[\frac{(1-t)R}{K}\right]^{2(R-s)} \left[1 + \frac{(t-2)R}{K}\right]^{K-2R+s}\right]^{1-\frac{1}{L}}}{(R-s)!\,(R-s)!\,s!\,(K-2R-s)!} \right\} . \quad (4.5)$$

The interpretation of the parameter $0 \le t \le 1$ follows from expression

$$\mathbb{E}(\mathcal{N}_G^2) = \sum_{\sigma_1,\sigma_2} P(\sigma_1\,SAT, \sigma_2\,SAT), \quad (4.6)$$

where σ_1 and σ_2 are configurations and $P(\cdot)$ is a probability over the graph ensemble. Parameter t in (4.5) is then number of sites occupied in both σ_1 and σ_2 divided by number of sites occupied in one of the solutions, RN/K. We remind that in the R-in-K SAT the satisfiability threshold is given by cancellation of th entropy (3.3)

$$l_s = \left[1 - \frac{\log\binom{K}{R}}{KH\left(\frac{R}{K}\right)}\right]^{-1} . \quad (4.7)$$

PROPOSITION 4.5. *For all the regular R-in-K SAT problems holds: For $L < l_s$ we have $2s_{1\mathrm{st}} = s_{2\mathrm{nd}}$, and for $L > l_s$ we have $s_{1\mathrm{st}} = s_{2\mathrm{nd}}$.* We were not able to prove this statement using formal algebra, but as $s_{2\mathrm{nd}}$ is a maximum of a function of one variable t we can persuade ourself by plotting $s_{1\mathrm{st}}$ and $s_{2\mathrm{nd}}(t)$ that the statement is correct, see e.g. Fig. 4.1.

It is easy to persuade ourself that in the ratio $\mathbb{E}(\mathcal{N}^2)/[\mathbb{E}(\mathcal{N})]^2$ the non-exponential term do not depend on the system size N. Thus from the fact Proposition 4.4 follows from the relation $2s_{1\mathrm{st}} = s_{2\mathrm{nd}}$.

We also investigated numerically general formulas for the second moments method presented in [33] and concluded that Proposition 4.5, and thus the asymptotic equivalence of the planted and random ensemble, holds also for all the other locked factorized problems.
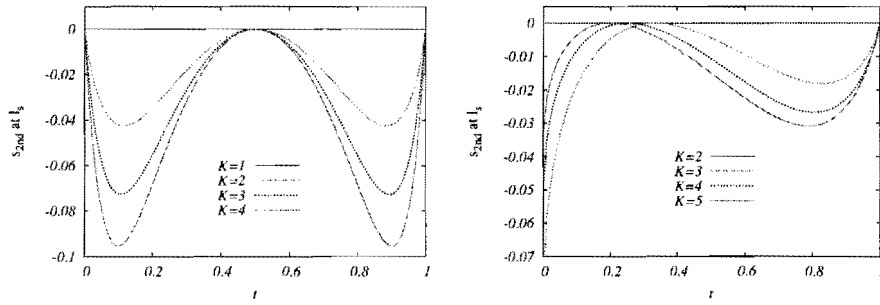

FIG. 4.1. *The second moment entropy function $s_{2\mathrm{nd}}(t)$ (4.5) at l_s for several values of K, in the K-in-$2K$ SAT problem on the left, and 1-in-K SAT on the right.*

5. Single solution instances for $\bar{l} > l_s$. As we argued in the introduction there it is very important to be able to create hard instances which have a single solution

with large probability. Based on the relation between planting and reconstruction on trees we concluded that in the region $\bar{l} > l_s$ with high probability there is a single solution to the problem on the large typical instances from the planted ensemble (or a couple in case of balanced problems). This statement can also be supported by more rigorous arguments.

In this section we go trough this argumentation for the R-in-K SAT on random regular graphs. We believe that generalization of the proofs from this section is possible also to the other factorized locked problems.

First note that the first moment in the planted R-in-K SAT is related in a simple way to the first and second moment in the purely random ensemble. It holds for the entropies

$$s_{1\text{st,pl}}(t) = s_{2\text{nd}}(t) + s_{1\text{st}}. \tag{5.1}$$

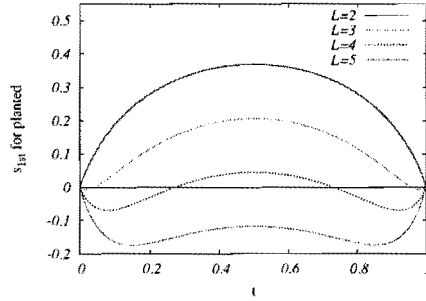See an example of the function $s_{1\text{st,pl}}(t)$ in Fig. 5.1. From Proposition 4.5 it follows

FIG. 5.1. *The first moment entropy in the 4-in-8 SAT on L regular planted ensemble.*

that for $L > l_s$ the first moment entropy in the planted ensemble is a negative function for all $0 < t < 1$. The parameter t is in the planted ensemble interpreted as the distance from the planted solution. Thus for $L > l_s$ there are no solution at an extensive distance from the planted solution (except the solution at distance one in the balanced problems).

THEOREM 5.1. *Consider a large instance of the R-in-K SAT problem drawn from the planted ensemble, degree of variables be $L > 2$. Then there exist an $\epsilon > 0$ such that with high probability there is no solution at distance smaller than ϵN from the planted solution* We will use the expander properties of regular bipartite graphs. The following theorem is well known in the theory of expanders [31]

THEOREM 5.2. *[Sipser and Splielman [31]] Consider a random factor-graph with degree of variables L and degree of constraints K. Then, for any $\delta < L-1$, there exists a constant $\epsilon > 0$, such that with high probability for every set of $\tilde{N} \leq \epsilon N$ variables the number of neighboring constraints is larger than $\delta \tilde{N}$, in other words the factor graph is a (ϵ, δ) expander.*

Proof. [of Theorem 5.1] Now consider the factor-graph and the planted solution. In order to find another solution a certain number N_1 of variables have to be changed. The property defining a locked constraint is that if a variable is changes at least one other have to be changed in order to satisfy the constraint again. If $N_1 < \epsilon N$ where ϵ is the value from Theorem 5.2 then variables have been changed in at least δN_1 clauses. Thus the total degree of changed variables $LN_1 > 2\delta N_1$. But as δ can be

as near to $L - 1$ as we wish this inequality is impossible. This proofs that with high probability there is no solution at distance smaller than ϵ from the planted one. $\Box$

The properties of the first moment in the planted ensemble together with Theorem 5.1 imply that in the planted R-in-K SAT on random regular graphs there is almost surely a single solution (or a pair of solutions for $R = K/2$).
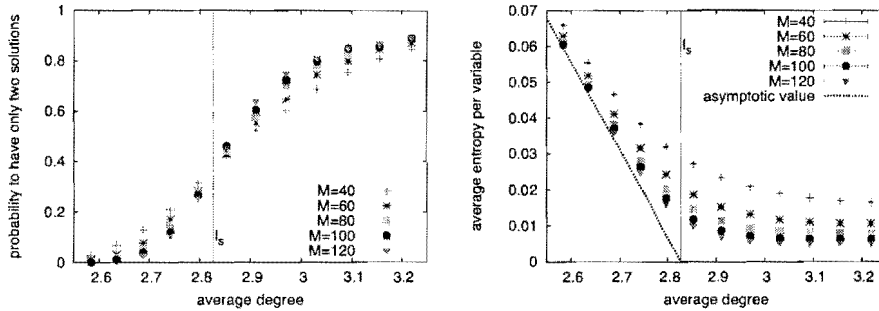

FIG. 5.2. *Left: Probability (over 5000 instances) that there is a single pair of solutions in the 2-in-4 SAT as a function of the average degree and the size of the graph. Right: Data are the average entropy density (logarithm of the number of solutions per variables) of the instances. The line represents the entropy density in the thermodynamical limit, eq. (3.2). The data are obtained with the* relsat *algorithm [14]. In both parts we marked the threshold* $l_s = 2.827$.

6. Average computational hardness.

One of the most interesting aspects of the study of random constraint satisfaction problems is the average computational hardness of a given ensemble. This has been discussed extensively in both the computer science and physics literature in particular for the K-satisfiability or coloring problems. The situation in the planted factorized locked problems is very interesting, as the average degree is changed the problem undergoes easy-hard-easy transition and moreover the first threshold can be pined more sharply than in the more commonly studied problems.

We argued that in the satisfiable phase $\bar{l} < l_s$ the planted and random ensemble are asymptotically equivalent, this includes average behavior of algorithms. It was argued in [34, 33] that for average degree $\bar{l} < l_d$ the locked problems are algorithmically easy whereas for $l_d < \bar{l} < l_s$ they are on average hard.

The second hard-easy transition is particular to the planted ensemble and happens in the phase $\bar{l} > l_s$. We study the behavior of the BP equations initialized randomly to locate this transition.

6.1. The spinodal point.

By definition of the factorized locked problems the belief propagation equations (2.1) initialized randomly converge to a uniform fixed point. But as the average degree is growing this ceases to be true. In the problems that we are studying here, it actually exists a critical average degree l_l beyond which belief propagation converges spontaneously towards the planted solution. This yields a clear hard-easy transition in the algorithmic complexity. In statistical physics terms this threshold l_l corresponds to a spinodal point of the liquid state. The spinodal point also corresponds to the Kesten-Stigum bound [15, 16], and to the robust reconstruction threshold on trees [12]. This is yet another important connection between the reconstruction and the planted ensemble.

In order to compute the spinodal point let us first define matrix $z(s'|s)$. Consider

a variable and one of its neighbors, $z(s'|s)$ is then the probability that in the planted configuration the variable was assigned s' given that its neighbors was s. In the terms on reconstruction on trees $z(s'|s)$ is the probability that in the broadcasting a variable was assigned s' given its parent was s. Components of $z(s'|s)$ can be computed using the quantities $y_r(s)$ from eq. (4.2) and reads

$$z(0|0) = \sum_{r=0}^{K} \left(1 - \frac{r}{K-1}\right) y_r(0), \qquad z(1|0) = 1 - z(0|0), \tag{6.1}$$

$$z(1|1) = \sum_{r=0}^{K} \frac{r-1}{K-1} y_r(1), \qquad z(0|1) = 1 - z(1|1) \tag{6.2}$$

explicit formulas for the regular problems

$$z(0|0) = \frac{\sum_{r=0}^{K-2} \binom{K-2}{r} \delta_{A(r),1} \psi_1^{r(L-1)} \psi_0^{(K-r-1)(L-1)}}{\sum_{r=0}^{K-1} \binom{K-1}{r} \delta_{A(r),1} \psi_1^{r(L-1)} \psi_0^{(K-r-1)(L-1)}}, \tag{6.3}$$

$$z(1|1) = \frac{\sum_{r=2}^{K} \binom{K-2}{r-2} \delta_{A(r),1} \psi_1^{(r-1)(L-1)} \psi_0^{(K-r)(L-1)}}{\sum_{r=1}^{K} \binom{K-1}{r-1} \delta_{A(r),1} \psi_1^{(r-1)(L-1)} \psi_0^{(K-r)(L-1)}}. \tag{6.4}$$

The first eigenvalue of this matrix is equal to one, and is associated with a trivial homogeneous eigenvector. The second eigenvalue of the matrix z is given by

$$\lambda = z(0|0) + z(1|1) - 1. \tag{6.5}$$

A well-known property of the reconstruction on a tree is that reconstruction is always possible beyond the so called Kesten-Stigum (KS) threshold [15, 16]. In our notation the KS condition says that if $(L-1)(K-1)\lambda^2 > 1$ then the reconstruction is possible, i.e., the leaves asymptotically contain some information on the variable sent by the root. On random graphs the Kesten-Stigum condition is equivalent to the spin-glass local instability [22, 18], that is for $(L-1)(K-1)\lambda^2 > 1$ the belief propagation equation 2.1 do not converge. This can be seen from the fact that

$$\lambda = \frac{\partial \psi_1^{a \to i}}{\partial \psi_1^{b \to j}}, \tag{6.6}$$

where $j \in \partial a \setminus i$, and $b \in \partial j \setminus a$.

The eigenvalue λ and the condition for solvability $(L-1)(K-1)\lambda^2 > 1$ also appear in the problem of robust reconstruction on trees [12]. In the problem of robust reconstruction it is required that even in arbitrary large fraction of the values on the leaves is erased there is still an information about the root left.

The analysis of the instability o the uniform BP fixed point towards the planted solution then goes as follows. Consider a part of the factor-graph as depicted in Fig. 2.1. Denote the values of the messages in the uniform fixed BP fixed point by over-bars. Consider the incoming message to be perturbed from the uniform value as

$$\left(\begin{array}{c} \psi_1^{b \to j} = \overline{\psi_1} + \epsilon \\ \psi_0^{b \to j} = \overline{\psi_0} - \epsilon \end{array} \right). \tag{6.7}$$

Note that ϵ can be both negative or positive. The equation 6.6 then implies that the outgoing message will be

$$\left(\begin{array}{c} \psi_1^{a \to i} = \overline{\psi_1} + \lambda\epsilon \\ \psi_0^{a \to i} = \overline{\psi_0} - \lambda\epsilon \end{array} \right). \tag{6.8}$$

In other words, any infinitesimal noise in one of the incoming message is multiplied by λ in the recursion.

We call the perturbation of the incoming message ϵ_+ is j was occupied in the planted configuration, and ϵ_- otherwise. If the variable i was planted in the occupied state, then j was planted occupied with probability $z(1|1)$, and empty with probability $z(0|1)$. Similarly, if the variable i was planted in the empty state, then j was planted empty with probability $z(0|0)$ and occupied with probability $z(1|0)$. Thus the evolution of the perturbation is governed by the equations:

$$\begin{pmatrix} \epsilon_+^{a \to i} \\ \epsilon_-^{a \to i} \end{pmatrix} = \lambda \begin{pmatrix} z(1|1) & z(0|1) \\ z(1|0) & z(0|0) \end{pmatrix} \begin{pmatrix} \epsilon_+^{b \to j} \\ \epsilon_-^{b \to j} \end{pmatrix}. \tag{6.9}$$

Moreover there are $(K-1)(L-1)$ of possible incoming messages in the regular graphs, thus the criterion $(K-1)(L-1)\lambda^2 = 1$. If $(K-1)(L-1)\lambda^2 < 1$ then the perturbation decreases and we find only the uniform BP fixed point, if on contrary $(K-1)(L-1)\lambda^2 > 1$ the the uniform BP fixed point is unstable and BP converges instead to the planted solutions. Fig. 6.1 confirms that this is true even on rather small graphs. On the balanced locked problems, where we are not restricted to the regular graphs, the correct condition is $(K-1)\gamma\lambda^2 = 1$, where γ is the mean of the excess degree distribution $q(l) = (l+1)Q(l+1)/\bar{l}$. The spinodal point l_l, see Tabs. 3.1,3.2, is then define by

$$(K-1)(l_l - 1)\lambda^2 = 1 \tag{6.10}$$

for the regular graphs, and

$$(K-1)\lambda^2 = \frac{1 - e^{-c_l}}{c_l} \tag{6.11}$$

for the truncated Poissonian distribution.

6.2. Belief propagation as a solver. Belief propagation reinforcement is a good solver in the region $\bar{l} < l_d$ as shown empirically in [34, 33] in the random ensemble, as the two ensemble are equivalent in that region nothing changes for the the planted ensemble. We indeed verified this numerically.

Based on the above arguments, the belief propagation equations converge to the uniform fixed point for $\bar{l} < l_l$ and directly to the planted solution for $\bar{l} > l_l$. In order to verify that on finite size instances, we have performed the following numerical experiment: we have generated many planted instances for different sizes and average degree (5000 instances for each set of parameters). We then iterated the BP equations (2.1) starting from random initial condition. For numerical stability reasons we used dumping in the iterations, i.e. each time we computed a new message we kept one half of the sum of the new and old message. As a convergence criterion we used that the messages should not change more that 2.10^{-3} per message (we checked that a smaller criterion does not change the quality of results, only slows down the computation). This way every iteration converged either to a configuration where the bias of each variable pointed towards the planted solution (or its negation) or a point very near to the uniform fixed point. Fig. 6.1 shows in what fraction of the runs we were able to find the planted solution and in particular it confirms that for $\bar{l} > l_l$ is is easy to find it in linear time. On the right of the same figure we plot the average convergence time (given the criterion 2.10^{-3} per message). We see that around the spinodal point l_l the convergence time diverges from both the sides (slightly faster from the large degree side).
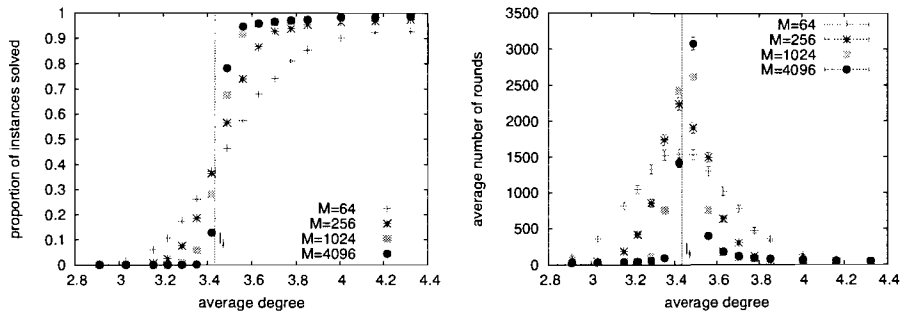
FIG. 6.1. *Belief propagation on the 2-in-4 SAT problem. Left: Probability that the belief propagation algorithm finds the planted configuration when initialized randomly plotted as a function of the average degree for several system sizes. Right: The convergence time dependence on the average degree. In both cases, we have stopped the BP iterations when the average change per message was less than 2.10^{-3}. In both parts we marked the spinodal threshold $l_l = 3.434$.*

7. Conclusions and perspectives. In this work we have studied a class of constraint satisfaction problems on a planted ensemble. The solution is planted in a *quiet* way, i.e. the planted configuration is one of the typical solutions of the resulting instance. We know how to realize such planting only on the factorized problems. We describe several connections between this quiet planting and the problem of reconstruction on trees.

We study the locked problems because of the simple structure on the space of their solutions — solutions are isolated points instead of clusters. This property makes the locked problems, however, very hard algorithmically. We focused on the class of occupation locked problems in this manuscript, all our result generalize to any factorized locked problem, on non-binary variables for example.

On the non-locked but factorized problems, as e.g. the graph coloring, the concept of quiet planting stays valid [20], however the random and planted ensembles are not equivalent up to the satisfiability threshold. And in the unsatisfiable phase the planted ensemble has exponentially many solutions, instead of a single one as is the case in the locked problems. Also the non-locked problems are much less friendly for first and second moment considerations. The phase diagram of the locked but non-factorized problem will not be very different from the one presented here. However, the thresholds will be different in the planted and random ensembles and the two ensembles are not equivalent.

One of the most important result of our work is the location of the algorithmically hard region, between $l_d \leq \bar{l} \leq l_l$, in the problems under investigation. It would be in particular interesting to design an algorithm which would provably find solutions in the region $\bar{l} > l_l$, we have only heuristic and numerical arguments. This is also challenging in the non-locked problems, as e.g. the graphs coloring, where we predicted the planted Poisson ensemble to be easy above $l_l = (q-1)^2$ (on planted regular graphs $L_l = (q-1)^2 + 1$), where q is the number of colors. Results establishing that the planted ensemble on coloring is easy above Cq^2, where C is some constant quite larger that one, are already known [17, 6].

Another interesting aspect of the planted ensemble is to compare it with the ensemble of satisfying instances from the random ensemble. This issue has been addressed e.g. in [5, 7] and certain relations of the two ensembles were discovered.

Some remaining open questions are: Is there a region of average degree, and if yes what are its boundaries, where these two ensembles are asymptotically equivalent? And how does that property depend on the problem in question?

Finally, another consequence of our work worth discussing is that we know how to generate unique satisfying assignment instances - both in the hard and easy regions. Such instances are often used for evaluating the performance of the quantum annealing algorithm, but so far they have been generated with an exponential cost from an ensemble with unknown classical average computational complexity [32, 8]. In our opinion, these works should be repeated on instances on the locked problems. We conjecture that in the classically hard region also the quantum annealing will be exponential (this is because we anticipate a first order phase transition in the transverse magnetic field, as in [13]).

REFERENCES

[1] DIMITRIS ACHLIOPTAS AND AMIN COJA-OGHLAN, *Algorithmic barriers from phase transitions*. arXiv:0803.2122v2, 2008.

[2] S. BEN-SHIMON AND D. VILENCHIK, *Message passing for the coloring problem: Gallager meets alon and kahale*, in Proceedings of the 13th International Conference on Analysis of Algorithms, DMTCS proc., 2007, pp. 217–226.

[3] G. BIROLI AND M. MÉZARD, *Lattice glass models*, Phys. Rev. Lett., 88 (2002), p. 025501.

[4] PETER CHEESEMAN, BOB KANEFSKY, AND WILLIAM M. TAYLOR, *Where the Really Hard Problems Are*, in Proc. 12th IJCAI, San Mateo, CA, USA, 1991, Morgan Kaufmann, pp. 331–337.

[5] A. COJA-OGHLAN, M. KRIVELEVICH, AND D. VILENCHIK, *Why almost all k-colorable graphs are easy to color*, in Proceedings of the 24th International Symposium on Theoretical Aspects of Computer Science (STACS), LNCS 4393, 2007, pp. 121–132.

[6] AMIN COJA-OGHLAN, ELCHANAN MOSSEL, AND DAN VILENCHIK, *A spectral approach to analyzing belief propagation for 3-coloring*. Preprint: arXiv:0712.0171v1, 2007.

[7] R. MONASSON F. ALTARELLI AND F. ZAMPONI, *Can rare sat formulas be easily recognized? on the efficiency of message passing algorithms for k-sat at large clause-to-variable ratios*, J. Phys. A: Math. Theor., 40 (2007), pp. 867–886.

[8] E. FARHI, J. GOLDSTONE, S. GUTMANN, J. LAPAN, A. LUNDGREN, AND D. PREDA, *A quantum adiabatic evolution algorithm applied to random instances of an np-complete problem*, Science, 292 (2001), p. 472.

[9] U. FEIGE, E. MOSSEL, AND D. VILENCHIK, *Complete convergence of message passing algorithms for some satisfiability problems*, in Proc. of Random 2006, LNCS 4410, 2006, p. 339350.

[10] ROBERT G. GALLAGER, *Low-density parity check codes*, IEEE Trans. Inform. Theory, 8 (1962), pp. 21–28.

[11] R. G. GALLAGER, *Information theory and reliable communication*, John Wiley and Sons, New York, 1968.

[12] SVANTE JANSON AND ELCHANAN MOSSEL, *Robust reconstruction on trees is determined by the second eigenvalue*, Ann. Probab., 32 (2004), pp. 2630–2649.

[13] THOMAS JOERG, FLORENT KRZAKALA, JORGE KURCHAN, AND A. C. MAGGS, *Simple glass models and their quantum annealing*, Phys. Rev. Lett., 101 (2008), p. 147204.

[14] R. J. BAYARDO JR. AND J. D. PEHOUSEK, *Counting models using connected components*, in Proc. 17th AAAI, Menlo Park, California, 2000, AAAI Press, pp. 157–162.

[15] H. KESTEN AND B. P. STIGUM, *Additional limit theorems for indecomposable multidimensional galton-watson processes*, The Annals of Mathematical Statistics, 37 (1966), p. 1463.

[16] ———, *Limit theorems for decomposable multi-dimensional galton-watson processes*, J. Math. Anal. Appl., 17 (1966), p. 309.

[17] M. KRIVELEVICH AND D. VILENCHIK, *Semi-random models as benchmarks for coloring algo-*

rithms, in Proceedings of the Third Workshop on Analytic Algorithmics and Combinatorics (ANALCO), 2006, pp. 211–221.

[18] FLORENT KRZAKALA, ANDREA MONTANARI, FEDERICO RICCI-TERSENGHI, GUILHEM SEMERJIAN, AND LENKA ZDEBOROVÁ, *Gibbs states and the set of solutions of random constraint satisfaction problems*, Proc. Natl. Acad. Sci. U.S.A, 104 (2007), p. 10318.

[19] F. KRZAKALA, M. TARZIA, AND L. ZDEBOROVÁ, *A Lattice Model for Colloidal Gels and Glasses* , Phys. Rev. Lett., 101 (2008), p. 165702.

[20] FLORENT KRZAKALA AND LENKA ZDEBOROVÁ, *Hiding quiet solutions in random constraint satisfaction problems*. arXiv:0901.2130v1, 2009.

[21] F. R. KSCHISCHANG, B. FREY, AND H.-A. LOELIGER, *Factor graphs and the sum-product algorithm*, IEEE Trans. Inform. Theory, 47 (2001), pp. 498–519.

[22] MARC MÉZARD AND ANDREA MONTANARI, *Reconstruction on trees and spin glass transition*, J. Stat. Phys., 124 (2006), pp. 1317–1350.

[23] M. MÉZARD AND A. MONTANARI, *Information, Physics, Computation: Probabilistic approaches*, Cambridge University Press, Cambridge, 2008. In preparation: www.lptms.u-psud.fr/membres/mezard/.

[24] M. MÉZARD AND G. PARISI, *The bethe lattice spin glass revisited*, Eur. Phys. J. B, 20 (2001), p. 217.

[25] M. MÉZARD, G. PARISI, AND M. A. VIRASORO, *Spin-Glass Theory and Beyond*, vol. 9 of Lecture Notes in Physics, World Scientific, Singapore, 1987.

[26] M. MÉZARD, G. PARISI, AND R. ZECCHINA, *Analytic and algorithmic solution of random satisfiability problems*, Science, 297 (2002), pp. 812–815.

[27] DAVID G. MITCHELL, BART SELMAN, AND HECTOR J. LEVESQUE, *Hard and easy distributions for SAT problems*, in Proc. 10th AAAI, Menlo Park, California, 1992, AAAI Press, pp. 459–465.

[28] T. MORA, *Géométrie et inférence dans l'optimisation et en théorie de l'information*, PhD thesis, Université Paris-Sud, 2007. http://tel.archives-ouvertes.fr/tel-00175221/en/.

[29] J. PEARL, *Reverend bayes on inference engines: A distributed hierarchical approach*, in Proceedings American Association of Artificial Intelligence National Conference on AI, Pittsburgh, PA, USA, 1982, pp. 133–136.

[30] BART SELMAN, HENRY A. KAUTZ, AND BRAM COHEN, *Local search strategies for satisfiability testing*, in Proceedings of the Second DIMACS Challange on Cliques, Coloring, and Satisfiability, Michael Trick and David Stifler Johnson, eds., Providence RI, 1996.

[31] MICHAEL SIPSER AND DANIEL A. SPIELMAN, *Expander codes*, IEEE Transactions on Information Theory, 42 (1996), pp. 1710–1722.

[32] A. P. YOUNG, S. KNYSH., AND V.N. SMELYANSKIY, *Size dependence of the minimum excitation gap in the quantum adiabatic algorithm*, Phys. Rev. Lett., 101 (2008), p. 170503.

[33] L. ZDEBOROVÁ AND M. MÉZARD, *Constraint satisfaction problems with isolated solutions are hard*, J. Stat. Mech., (2008), p. P12004.

[34] ———, *Locked constraint satisfaction problems*, Phys. Rev. Lett., 101 (2008), p. 078702.