Statistical mechanics of the hitting set problem

Marc Mézard and Marco Tarzia
CNRS; Laboratoire de Physique Théorique et Modèles Statistiques, Université Paris-Sud, UMR 8626, Orsay CEDEX 91405, France
(Received 2 July 2007; published 18 October 2007)

In this paper we present a detailed study of the hitting set (HS) problem. This problem is a generalization of the standard vertex cover to hypergraphs: one seeks a configuration of particles with minimal density such that every hyperedge of the hypergraph contains at least one particle. It can also be used in important practical tasks, such as the group testing procedures where one wants to detect defective items in a large group by pool testing. Using a statistical mechanics approach based on the cavity method, we study the phase diagram of the HS problem, in the case of random regular hypergraphs. Depending on the values of the variables and tests degrees different situations can occur: The HS problem can be either in a replica symmetric phase, or in a one-step replica symmetry breaking one. In these two cases, we give explicit results on the minimal density of particles, and the structure of the phase space. These problems are thus in some sense simpler than the original vertex cover problem, where the need for a full replica symmetry breaking has prevented the derivation of exact results so far. Finally, we show that decimation procedures based on the belief propagation and the survey propagation algorithms provide very efficient strategies to solve large individual instances of the hitting set problem.

DOI: 10.1103/PhysRevE.76.041124 PACS number(s): 05.20.−y, 05.40.−a, 02.50.−r, 75.10.Nr

I. INTRODUCTION AND MOTIVATION

In this paper we discuss the statistical physics of the hitting set (HS) problem, which is a NP-complete problem in set theory [1]: Given a collection of subsets $\mathcal{S}$ of a universe $\mathcal{U}$, one is asked to find a subset $\mathcal{H}$ of $\mathcal{U}$ that intersects ("hits") every set in $\mathcal{S}$. In other words, every set in $\mathcal{S}$ must contain at least one element of $\mathcal{H}$. The HS is an interesting problem in itself, and it is closely related to several other well known optimization and decision problems. Interestingly enough, it turns out that the HS problem is the generalization to hypergraphs of the celebrated vertex cover (VC) problem, which is intimately related to spin glass problems in statistical physics, and has received a lot of attention in recent years from the physics community [2–5].

The problem can be easily stated in terms of Boolean variables on a bipartite graph: We consider a large population of N variables x_i, which can be either *active* ($x_i=1$) or *inactive* ($x_i=0$), and M function nodes (or *tests*), t_a, and build up a random regular hypergraph (or *factor graph*), i.e., a bipartite graph where each variable is connected exactly to L tests, and each test is connected exactly to K variables (with $NL=MK$). The function nodes enforce the constraint that at least one of the variables to which they are connected must be active. The case $K=2$ reduces to the standard VC problem. In this paper we will be mainly interested in the *minimal* HS, i.e., a cover of the hypergraph with the minimal possible number of active variables.

The model can be also interpreted as a system of M Boolean clauses over the N variables which should be simultaneously satisfied: each clause is an OR function involving K randomly chosen variables, (e.g., $x_{i_1}\vee\cdots\vee x_{i_K}=1$) and they are such that each variable appears exactly in L clauses. Thus, the minimal HS configuration corresponds to a pattern of the variables x_i which satisfies all the M clauses with the minimal number of ones.

The interest of the present study is twofold. Statistical physics studies of the VC problem on random graphs have shown that, depending on the average degree of a variable in the graph, the problem is either simple (meaning that replica symmetry is unbroken), or very difficult [meaning that a full replica symmetry breaking (RSB) scheme is necessary]. As the theory of full RSB is not well under control for "finite connectivity" graphs (where the degrees of variables are finite), the difficult region is not fully understood, notwithstanding the recent progress made in Refs. [2–4]. As we shall see, the HS displays, for certain families of random graphs, an intermediate situation which is both nontrivial, because RSB is needed, but under control, because the solution is given by a first order RSB (1RSB), as developed for finite connectivity problems in [6]. Therefore this model joins the family of well controlled hard combinatorial optimization problems in which 1RSB is supposed to give exact results, a family which includes K satisfiability [7–10], graph coloring [11], random Boolean equations [12], 1-in-K satisfability [13], and lattice glasses [14,15].

On the other hand, there are several applications of HS to important practical "real-world" tasks. In particular, HS is closely related to the *group testing* (or *pool testing*) procedures [16,17]. The object of the group testing is to identify an *a priori* unknown subset of a large population of N variables, called the set of *active* (or *defective*) items, using as few queries as possible. Each query (or test) is connected to a certain subset of K items, and informs the tester about whether or not the subset contains at least one active item. A negative answer implies that all the items of the subset are inactive. This approach is used in many different applications, beginning with an efficient blood testing procedure [18]. Other applications include quality control in product testing [19], searching files in a storage systems [20], efficient accessing of computer memories, sequential screening of experimental variables [21], and many others, such as the basic problem of DNA libraries screening, which is very important in modern biological applications such as monolocal antibody generation. Objectives of group testing range from finding an optimal strategy with the minimal number of tests,

to devising an efficient algorithm able to reconstruct the pattern of active items.

In the group testing problem, it is easy to first identify the variables which are *sure zeros* (the ones which are connected at least to one negative test). If one now considers the reduced graph obtained from the original GT problem by removing these sure zero variables, as well as all the negative tests, one obtains a reduced graph (with fluctuating degrees). The problem of identifying the active items in this graph is exactly the HS problem. Therefore the study of the phase diagram of the HS could give useful insights, for example, to understand which is the best reconstruction algorithm for the pattern of active items to be used depending on the topological properties of the factor graph.

In this paper we first present the study of the phase diagram of the random regular HS problem, where each variable appears in L tests and each test involves K variables, based on the cavity method [6] (these results could also be obtained in the framework of the replica approach [9,22]). We show that depending on the values of the variables and tests degree different situations can occur: The minimal HS problem can be either in a replica symmetric (RS) phase or in a one-step replica symmetry breaking (1RSB) phase. In these two cases we give explicit results on the minimal density of active items, and the structure of the phase space. On the other hand, there are also cases (like, e.g., the ordinary VC with $K=2$ and $L\geq 3$), where a higher order RSB pattern is needed; these are more difficult problems that we do not address in this paper. The summary of the situation for the various values of K and L is contained in Fig. 4. We then introduce the survey propagation (SP) and the belief propagation (BP) type algorithms, the analog for this problem of the ones which have turned out to be so efficient in K satisfiability [10]. We show that a decimation procedure based on the surveys turns out to be an efficient way to solve large instances of the HS problem.

The paper is organized as follows: In the next section we define the problem and we introduce the statistical mechanics formulation; in Sec. III we develop the cavity approach and work out the RS solution; in Sec. IV we focus on the 1RSB solution for the phase diagram, on the minimal HS limit, and on the stability of the 1RSB approach with respect to further breaking of the replica symmetry; Sec. V explains the use of the BP and SP algorithms, and their application with a decimation procedure, to solve individual large instances of the problem; Finally, in Sec. VI we discuss the results found and conclude this work.

II. HITTING SET: DEFINITION AND STATISTICAL PHYSICS FORMULATION

We consider a factor graph containing $N+M$ vertices: N of them are associated with variables, we shall label them by $i,j,\ldots\in\{1,\ldots,N\}$. The other M are associated with function nodes (or tests), denoted by $a,b\ldots\in\{1,\ldots,M\}$. An edge $i-a$ between variable i and vertex a is present in the factor graph whenever variable i appears in test a. The factor graph is bipartite.

The HS problem is the following: each variable i can be either inactive ($x_i=0$) or active ($x_i=1$). We request that, for each of the M tests, at least one out of the K variables connected to it be equal to 1. The optimization problem (minimal HS) consists of finding a configuration that satisfies all these constraints and has the smallest value $A_{\min}$ of the "weight,"

$$A(\{x_i\})=\sum_{i=1}^{N} x_i. \tag{1}$$

One is also interested in knowing the number of configurations which satisfy all constraints, when A has a fixed value $\geq A_{\min}$.

We shall use the following statistical mechanics formulation of the problem. Given an instance of the problem, characterized by a factor graph, we introduce the set of admissible configurations $\mathcal{C}$, which are all configurations of the N variables such that, for each clause a, at least one variable connected to a takes value 1. Denoting by ∂a the variables which enter in clause a (i.e, the set of neighbors of a in the factor graph), we thus impose $\Pi_{i\in\partial a}(1-x_i)=0$. The Boltzmann-Gibbs measure (canonical ensemble) is defined as $P(\{x_i\})=(1/Z)e^{-\mu A(\{x_i\})}$. The chemical potential μ controls the overall density of ones $\rho=\Sigma_i x_i/N$. The minimal HS problem is recovered in the $\mu\to\infty$ limit, where the Boltzmann-Gibbs measure concentrates on configurations with the smallest number of active variables. We are also interested in understanding the properties of the microcanonical measure P_A^{mc} which is the uniform measure on the admissible configurations $\{x_i\}$ with a fixed weight $A(\{x_i\})=A$, whenever such configurations exist. These properties will be studied hereafter through a detour to the canonical ensemble, using ensemble equivalence.

The partition function of the model reads

$$Z=\sum_{\{x_i\}} e^{-\mathcal{H}(\{x_i\})}=e^{-\mu F}=e^{S-\mu N\rho}, \tag{2}$$

where F is the free energy and S is the entropy.

In the following we shall be particularly interested in the random L,K HS problem, in which the factor graph is a random regular (L,K) bipartite graph, uniformly chosen from all the possible graphs where each variable is connected to L tests and each test is connected to K variables.

The thermodynamic limit is taken by letting the number of variables N and the number of constraints M go to infinity with a fixed ratio $\alpha=M/N$, keeping the degrees K and L fixed. In the following we use the cavity method [6], which allows us to write down iterative self-consistent relations for local expectation values, which are exact on a loopless factorized graph (i.e., a tree). For any finite values of N and M the graph is only locally treelike, and it has loops whose average length is expected to scale as $\ln N$. Therefore the cavity method is expected to provide good approximations for large samples and to become exact in the thermodynamic limit.

III. CAVITY APPROACH AND REPLICA SYMMETRIC SOLUTION

Given a graph and a variable i, we consider a subgraph rooted in i obtained by removing the edge between i and one

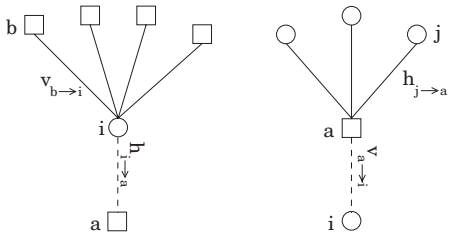

FIG. 1. Subgraph rooted in i in absence of the neighboring test a (on the left). The function nodes b belong to the neighborhood of i ($b \in \partial i \backslash a$) and send the messages $v_{b\to i}$ to the site i. The site i sends the message $h_{i\to a}$ to the function node a, according to Eqs. (5). Analogously, on the left we have sketched the subgraph rooted in the function node a in absence of the edge with the variable i. The sites j belong to the neighborhood of a ($j \in \partial a \backslash i$) and send the messages $h_{j\to a}$ to the function node a. According to Eqs. (5), a sends the message $v_{a\to i}$ to the site i.

of its neighboring tests, a. Define $Z_0^{(i\to a)}$ and $Z_1^{(i\to a)}$ as the partition functions of this subgraph restricted to configurations where the variable x_i is respectively inactive ($x_i=0$) or active ($x_i=1$). Assuming that the subgraph is a tree (which is generically correct, when one takes the large N limit, up to any finite depth), these restricted partition functions can be written recursively (see Fig. 1): consider the function nodes b belonging to the neighborhood ∂i of i. For each $b \in \partial i \backslash a$, consider the restricted partition functions $Y^{(b\to i)}$ and $Y_1^{(b\to i)}$ on the rooted branches of the graph starting from b, which are defined, respectively, as the total partition function of the branch, and the partition function of the branch restricted to configurations in which at least one of the remaining $K-1$ variables connected to the test is active.

One obtains the following recursion relations:

$$Z_1^{(i\to a)} = e^{-\mu} \prod_{b\in\partial i\backslash a} Y^{(b\to i)},$$

$$Z_0^{(i\to a)} = \prod_{b\in\partial i\backslash a} Y_1^{(b\to i)}. \tag{3}$$

Analogously, one can express $Y^{(a\to i)}$ and $Y_1^{(a\to i)}$ in terms of the restricted partition functions $Z_0^{(j\to a)}$ and $Z_1^{(j\to a)}$ for $j \in \partial a \backslash i$:

$$Y_1^{(a\to i)} = \prod_{j\in\partial a\backslash i} (Z_0^{(j\to a)} + Z_1^{(j\to a)}) - \prod_{j\in\partial a\backslash i} Z_0^{(j\to a)},$$

$$Y^{(a\to i)} = \prod_{j\in\partial a\backslash i} (Z_0^{(j\to a)} + Z_1^{(j\to a)}). \tag{4}$$

It is now convenient to introduce two local cavity fields on each edge of the graph, defined as $e^{\mu h_{i\to a}} = Z_1^{(i\to a)}/(Z_0^{(i\to a)} + Z_1^{(i\to a)})$, and $e^{\mu v_{a\to i}} = Y_1^{(a\to i)}/Y^{(a\to i)}$. Basically, $e^{\mu h_{i\to a}}$ measures the local probability that the variable i is active in absence of the link with the test a. In terms of the local cavity fields, the recursion relations, Eqs. (3) and (4), read

$$e^{\mu h_{i\to a}} = \frac{\exp(-\mu)}{\exp(-\mu) + \exp\left(\sum_{b\in\partial i\backslash a} \mu v_{b\to i}\right)},$$

$$e^{\mu v_{a\to i}} = 1 - \prod_{j\in\partial a\backslash i} (1 - e^{\mu h_{j\to a}}). \tag{5}$$

For any given finite graph, Eqs. (5) provide a set of $2MK$ (two equations for each edge of the graph) coupled algebraic equations, the so-called *belief propagation* (BP) equations, which can in principle be solved on any individual instance. From these local fields one can compute the system free energy density $f=F/N$ in terms of *variable*, *test*, and *edge* contributions [6]. In order to do that, let us consider an intermediate object, a factor graph made up by N variables and M tests where KL "defective" variables are connected only with $L-1$ tests and KL defective tests are connected only to $K-1$ variables, while all other variables and tests have their natural degree (respectively L and K). We can now go from this intermediate graph to a well defined regular factor graph where each variable is connected to L tests and each test to K variables in two ways:

(i) We can either add K new items and L new tests and connect each of the item to L out of the LK defective tests and each new test to K out of the LK defective variables. In this way we obtain a regular hypergraph made up by $N+K$ variables and $M+L$ tests, all with their natural degree;

(ii) or we can add LK new edges between pairs of defective variables and tests. In this way we obtain a regular hypergraph made up by N variables and M function nodes, all with their natural degree.

In formulas we get

$$F(N+K) - F(N) = F_0 + \sum_{i=1}^{K} \Delta F_V^i + \sum_{a=1}^{L} \Delta F_T^a - \left(F_0 + \sum_{(j,b)} \Delta F_e^{(j,b)}\right), \tag{6}$$

where F_0 is the free energy of the intermediate graph, ΔF_V^i is the free energy shift due to the addition of a new variable i, ΔF_T^a is the free energy shift due to the addition of a new test a, and $\Delta F_e^{(j,b)}$ is the free energy shift due to the addition of a new edge between the item j and the test b. Supposing that the free energy scales linearly with the number of items, the previous relation allows us to determine the free energy density.

The free energy shifts appearing in Eq. (6) can be written in terms of the restricted partition functions defined above:

$$e^{-\mu\Delta F_V^i} = \frac{\prod_{b\in\partial i} Y_1^{(b\to i)} + e^{-\mu} \prod_{b\in\partial i} Y^{(b\to i)}}{\prod_{b\in\partial i} Y^{(b\to i)}},$$

$$e^{-\mu\Delta F_T^a} = \frac{\prod_{j\in\partial a} (Z_0^{(j\to a)} + Z_1^{(j\to a)}) - \prod_{j\in\partial a} Z_0^{(j\to a)}}{\prod_{j\in\partial a} (Z_0^{(j\to a)} + Z_1^{(j\to a)})},$$

$$e^{-\mu\Delta F_e^{(j,b)}} = \frac{Z_0^{(j\to b)} Y_1^{(b\to j)} + Z_1^{(j\to b)} Y^{(b\to j)}}{Y^{(b\to j)}(Z_0^{(j\to b)} + Z_1^{(j\to b)})}. \tag{7}$$

The free energy shifts can be finally rewritten in the following way in terms of the local cavity fields:

$$-\mu\Delta F_V^{i\cup\partial i} = \ln\left[\exp(-\mu) + \exp\left(\sum_{b\in\partial i} \mu v_{b\to i}\right)\right],$$

$$-\mu\Delta F_T^{a\cup\partial a} = \ln\left[1 - \prod_{j\in\partial a}(1 - e^{\mu h_{j\to a}})\right],$$

$$-\mu\Delta F_e^{(j,b)} = \ln[e^{\mu h_{j\to b}} + e^{\mu v_{b\to j}} - e^{\mu h_{j\to b}+\mu v_{b\to j}}]. \tag{8}$$

The expectation value of the number of active variables on each site i (also called *marginals* of the BP equations) can be obtained by deriving the free energy with respect to the chemical potential leading to

$$\rho = \left\langle \frac{1}{N}\sum_i x_i \right\rangle = \frac{1}{N}\frac{\partial \mu F}{\partial \mu} = \frac{1}{N}\sum_i \frac{\exp(-\mu)}{\exp(-\mu) + \exp\left(\sum_{b\in\partial i} \mu v_{b\to i}\right)}. \tag{9}$$

A. Replica symmetric solution, entropy crisis, and stability

Equations (5) can be written for arbitrary graphs and they provide exact marginal probability distributions (and thus exact densities of active variables) only for loopless trees. They are particularly suited for very large random hypergraphs, where, due to the local treelike structure, they are expected to provide exact results in the RS phase.

The simplest hypothesis one can make is the so-called *replica symmetric* (RS) ansatz. Assuming that there is a single state describing the equilibrium behavior of the system, one can look for *factorized replica symmetric* solutions of the cavity equations, where all the local fields are equal on all the edges of the graph, i.e., $h_{i\to a}=h_{RS}$ and $v_{a\to i}=v_{RS}$, $\forall(i,a)$. Within this approximation, Eqs. (5) reduce to

$$\mu v_{RS} = \ln\left\{1 - \left[\frac{e^{\mu(L-1)v_{RS}}}{e^{-\mu} + e^{\mu(L-1)v_{RS}}}\right]^{K-1}\right\}. \tag{10}$$

The free energy shifts reduce to node and edge independent quantities and can be easily evaluated, along with all the thermodynamic observables.

In Fig. 2 the density of active variables ρ and the entropy density $s=S/N=-\mu F/N+\mu\rho$ are plotted as functions of the chemical potential μ for $L=2$ and $K=6$ (left panel) and for $L=6$ and $K=12$ (right panel). In this RS solution, ρ goes to $1/K$ as μ goes to infinity. One readily notes that while in the first case the entropy density stays finite even in the $\mu\to\infty$ limit (meaning that there is an extensive number of RS states with density of active variables equal to $1/K$ satisfying the minimal covering of the hypergraph) in the second case the entropy density becomes negative for chemical potentials larger than a certain threshold $\mu_{s=0}$ (or, equivalently, for densities of active variables smaller that a certain value).

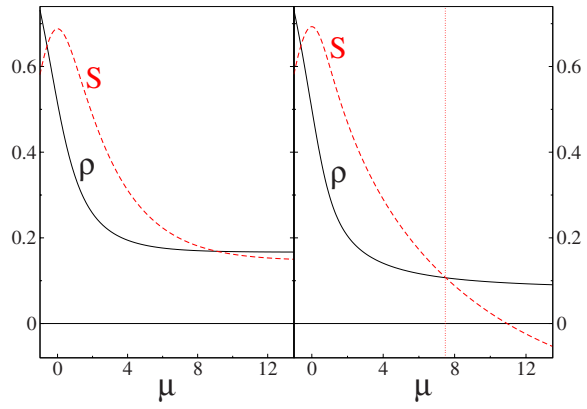

FIG. 2. (Color online) Density of active items ρ and entropy density $s=S/N$ as a function of the chemical potential μ in the RS solution of the HS for $L=2$ and $K=6$ (left panel) and for $L=6$ and $K=12$ (right panel). The red dotted vertical line in the right panel corresponds to the point where the RS solution becomes unstable

These results can be easily understood by noting that in the $\mu\to\infty$ limit the RS fields are simply given by

$$\mu h_{RS} = -\frac{\mu}{L} - \frac{L-1}{L}\ln(K-1),$$

$$\mu v_{RS} = -\frac{\mu}{L} + \frac{1}{L}\ln(K-1). \tag{11}$$

Therefore according to Eq. (9), $\rho=1/K$, whereas the free energy density reads

$$\mu f(\mu\to\infty) = \frac{\mu}{K} + \frac{(L-1)(K-1)-1}{K}\ln K - \frac{(L-1)(K-1)}{K}\ln(K-1), \tag{12}$$

which immediately leads to

$$s(\mu\to\infty) = \frac{1}{K}\{(L-1)(K-1)\ln(K-1) - [(L-1)(K-1)-1]\ln K\}. \tag{13}$$

In the large connectivity limit $(K,L\gg 1)$, the entropy density scales as

$$s(\mu\to\infty; L,K\gg 1) \simeq \frac{\ln K - L}{K}. \tag{14}$$

Thus $s(\mu\to\infty)$ is positive if $\ln K > L$, while it is negative for $L>\ln K$.

Clearly, the results presented above imply that the RS solution is wrong for μ large enough, at least in the case $L=6$ and $K=12$. The RS solution turns out to be incorrect if the assumption of the existence of a single equilibrium state fails, meaning that fields incoming to a given node become correlated. To gain further insight, one can test the stability of this RS solution by computing the nonlinear susceptibility, defined as

$$\chi_2 = \frac{1}{N}\sum_{i,j} \langle x_i x_j \rangle_c^2. \tag{15}$$

If χ_2 diverges, the incoming cavity fields are strongly correlated and the RS assumption is inconsistent. (Notice that the divergence of the linear susceptibility corresponds to a modulation instability which is not compatible with the random nature of the graph.) By using the fluctuation-dissipation theorem, one can relate the connected correlation functions between nodes i and j to the local cavity fields [15]. This finally leads to the following stability criterion:

$$\sum_{j\in\partial a\backslash i; b\in\partial j\backslash a} \left(\frac{\partial v_{a\to i}}{\partial v_{b\to j}}\right)^2 \leq 1. \tag{16}$$

For the RS solution this criterion yields

$$\sqrt{(L-1)(K-1)}\left|\frac{e^{\mu h_{RS}}(1-e^{\mu v_{RS}})}{e^{\mu v_{RS}}(1-e^{\mu h_{RS}})}\right| \leq 1. \tag{17}$$

Using the previous equation, we find that for several values of L and K the RS solution becomes indeed unstable above a certain chemical potential. For $L=6$ and $K=12$ the point where the instability appears is marked on the right plot as a vertical dotted line. However, it is well known from the physics of glassy systems that the RS solution can be wrong because of a first order transition to a replica symmetry breaking solution, not detected by the stability argument.

IV. 1RSB CAVITY APPROACH

Figure 2 clearly shows that the RS solution fails at high chemical potential and low density of active items (at least for $L=6$ and $K=12$), due to the fact that the hypothesis of the existence of a single state becomes inconsistent. Therefore in this case other solutions must be found. In the following we employ a one-step replica symmetry breaking (1RSB) approach [6] within the cavity method described in the previous section. More precisely, we assume that exponentially many (*pure*) states (in the size of the system) exist and that the neighbors of a given node, in the absence of the node itself, are uncorrelated only within each of these states.

Local fields on a given edge can now fluctuate from pure state to pure state. The cavity method provides a statistical description of the local fields in each state α, which must be weighted according to their Boltzmann weight $e^{-\mu F_\alpha}$ [6]. In order to deal with this situation, on each edge of the graph one has to introduce two probability distribution functions, $P_{i\to a}(h)$, and $Q_{a\to i}(v)$: $P_{i\to a}(h)$ gives the probability density of finding the fields $h_{i\to a}$ equal to h for a randomly chosen state [respectively, $Q(v)$ gives the probability density of finding the field $v_{a\to i}=v$]. By using the cavity method, the following self-consistent integral relations are found [6,14]:

$$P_{i\to a}(h) = \mathcal{Z}_1 \int \prod_{b\in\partial i\backslash a} [dv_{b\to i} Q(v_{b\to i})] \times \delta[h - \bar{h}(\{v_{b\to i}\})] e^{-\mu m \Delta F_{V,iter}^{i\cup\partial i\backslash a}},$$

$$Q_{a\to i}(v) = \mathcal{Z}_2 \int \prod_{j\in\partial a\backslash i} [dh_{j\to a} P(h_{j\to a})] \times \delta[v - \bar{v}(\{h_{j\to a}\})] e^{-\mu m \Delta F_{T,iter}^{a\cup\partial a\backslash i}}, \tag{18}$$

where $\bar{h}(\{v_{b\to i}\})$ and $\bar{v}(\{h_{j\to a}\})$ enforce the local cavity equations, Eqs. (5), m is the usual 1RSB parameter [22] (fixed by the maximization of the free energy with respect to it), $\mathcal{Z}_1$ and $\mathcal{Z}_2$ are normalization factors, and $\Delta F_{V,iter}^{i\cup\partial i\backslash a}$ and $\Delta F_{T,iter}^{a\cup\partial a\backslash i}$ are the free energy shifts involved in the iteration processes, which take into account the reweighting factors of the different pure states. Using the relations

$$-\mu\Delta F_{V,iter}^{i\cup\partial i\backslash a} = \ln\frac{Z_0^{(i\to a)} + Z_1^{(i\to a)}}{\prod_{b\in\partial i\backslash a} Y^{(b\to i)}} = -\mu\Delta F_V^{i\cup\partial i} + \mu\Delta F_e^{(i,a)},$$

$$-\mu\Delta F_{T,iter}^{a\cup\partial a\backslash i} = \ln\frac{Y^{a\to i}}{\prod_{j\in\partial a\backslash i}(Z_0^{(j\to a)} + Z_1^{(j\to a)})} = -\mu\Delta F_T^{a\cup\partial i} + \mu\Delta F_e^{(i,a)}, \tag{19}$$

one obtains the expressions of the free energy shifts involved in the iteration processes in terms of the local cavity fields

$$-\mu\Delta F_{V,iter}^{i\cup\partial i\backslash a} = \ln\left[\exp(-\mu) + \exp\left(\sum_{b\in\partial i\backslash a}\mu v_{b\to i}\right)\right],$$

$$-\mu\Delta F_{T,iter}^{a\cup\partial a\backslash i} = 0. \tag{20}$$

In analogy with Eq. (6), from the local fields probability distribution, one can also compute the 1RSB free energy of the system, as a sum of contributions due to the addition of a new item $\Delta\phi_V^i$, addition of a new test $\Delta\phi_T^a$, and addition of a new edge between a variable and a test, $\Delta\phi_e^{(i,a)}$:

$$K\phi = \sum_{i=1}^{K}\Delta\phi_V^i + \sum_{a=1}^{L}\Delta\phi_T^a - \sum_{(j,b)}\Delta\phi_e^{(j,b)}, \tag{21}$$

with

$$\Delta\phi_V^i = -\frac{1}{m\mu}\ln\left\{\int\prod_{b\in\partial i}[dv_{b\to i}Q(v_{b\to i})]e^{-\mu m\Delta F_V^i}\right\},$$

$$\Delta\phi_T^a = -\frac{1}{m\mu}\ln\left\{\int\prod_{j\in\partial a}[dh_{j\to a}P(h_{j\to a})]e^{-\mu m\Delta F_T^a}\right\},$$

$$\Delta\phi_e^{(i,a)} = -\frac{1}{m\mu}\ln\left\{\int[dh_{i\to a}P(h_{i\to a})] \times [dv_{a\to i}Q(v_{a\to i})]e^{-\mu m\Delta F_e^{(i,a)}}\right\}, \tag{22}$$

where the free energy shifts $\Delta F_V^{i\cup\partial i}$, $\Delta F_T^{a\cup\partial a}$, and $\Delta F_e^{(i,a)}$ are defined in Eq. (8).

In the 1RSB formalism [6], the 1RSB free energy $\phi(\mu,m)$ is given by (in the thermodynamic limit)

$$-\mu m\phi(\mu,m) = -\mu m f(\mu) + \Sigma(f). \tag{23}$$

Therefore by Legendre transforming the 1RSB free energy, we obtain the complexity $\Sigma(f)$ (i.e., the logarithm of the number of states with free energy density equal to f):

$$\Sigma = \mu m^2 \frac{\partial \phi(\mu,m)}{\partial m}, \quad f = \frac{\partial [m\phi(\mu,m)]}{\partial m}, \tag{24}$$

with $\mu m = \partial\Sigma(f)/\partial f$.

On infinite random regular hypergraphs one can assume that the 1RSB glassy phase is translationally invariant, i.e., that the probability distribution of the local cavity fields are edge independent: for all edges $a-i$, $P_{i\to a}(h)=P(h)$, and $Q_{a\to i}(v)=Q(v)$. In this case, Eqs. (18) become two coupled integral equations for the probability distributions P and Q, and can be solved for any value of μ and m by means of a population dynamics algorithm [6].

At low μ one finds with this procedure that $P(h)$ and $Q(v)$ always converge toward the RS solution: starting from populations of fields with an arbitrary distribution, they converge to populations of identical fields such that $P(h)=\delta(h-h_{RS})$ and $Q(v)=\delta(v-v_{RS})$, where h_{RS} and v_{RS} are the values of the local cavity fields which satisfy the factorized RS equation, Eq. (10).

When μ is increased, a first nontrivial distribution is found at $\mu=\mu_d$ for $m=1$. At this point many states appear. The phase space splits into many clusters of solutions and, even though they are only metastable (the equilibrium state is still given by the RS solution at this point, since the maximum of the 1RSB free energy occurs at $m>1$), they could trap most of the algorithms and dynamical procedures which look for a covering pattern of a given hypergraph. Thus for a density of active items smaller than this "dynamical" threshold a *survey propagation* algorithm [7] should be used to find solutions of the HS for finite instances.

A static phase transition (which is only relevant at equilibrium) appears at higher chemical potential, $\mu=\mu_c$, where the maximum of the 1RSB free energy is located in $m=1$, the complexity vanishes [6,14,15], and a thermodynamic transition from the RS phase to a 1RSB glassy one takes place (for $L=6$ and $K=12$ we find that $\mu_d \simeq 6.2$ and $\mu_c \simeq 7$).

A. Minimal hitting set

Now we consider the minimal HS problem. Namely, we still request that all the clauses are satisfied, but using the minimum possible number of active variables. This corresponds to the $\mu\to\infty$ limit in our statistical physics formulation.

We thus consider the $\mu\to\infty$ limit of the 1RSB equations, Eqs. (18). In this limit, according to Eqs. (5), it is self-consistent to assume that the local cavity fields $h_{i\to a}$ and $v_{a\to i}$ can be either equal to minus one or to zero. In particular, the field $v_{a\to i}$ is equal to minus one if all the incoming fields $h_{j\to a}$ are equal to -1 too, while it equals zero if at least one of the $h_{j\to a}$ is zero. On the other hand, $h_{i\to a}$ turns out to be equal to -1 if all the incoming fields $v_{b\to i}$ are equal to zero, and it equals zero if at least one of the $v_{b\to i}$ is -1. Therefore the probability distribution functions reduce to

$$P_{i\to a}(h) = (1-g)^{i\to a}\delta(h+1) + g^{i\to a}\delta(h),$$

$$Q_{a\to i}(v) = (1-u)^{a\to i}\delta(v+1) + u^{a\to i}\delta(v) \tag{25}$$

and the 1RSB integral equations reduce to algebraic equations for the coefficients. For instance, the second of Eqs. (18) becomes

$$u^{a\to i} = \left[1 - \prod_{j\in\partial a\setminus i}(1-g^{j\to a})\right], \tag{26}$$

where $y=\mu m$. Analogously, for the first equation we have

$$g^{i\to a} = \frac{\left(1 - \prod_{b\in\partial i\setminus a} u^{b\to i}\right)e^{-y}}{e^{-y} + (1-e^{-y})\prod_{b\in\partial i\setminus a} u^{b\to i}}. \tag{27}$$

Finally, after some algebra, the following system of equations for the coefficients $u^{a\to i}$ can be obtained:

$$u^{a\to i} = 1 - \prod_{j\in\partial a\setminus i}\left(\frac{\prod_{b\in\partial j\setminus a} u^{b\to j}}{e^{-y} + (1-e^{-y})\prod_{b\in\partial j\setminus a} u^{b\to j}}\right). \tag{28}$$

These equations are the ("zero temperature") survey propagation (SP) equations [7,10] for the minimal HS problem. In the case $K=2$, one recovers the result found for the VC in Refs. [2,3]. We will see in the next section how to use them algorithmically in order to solve single instances. Assuming that the coefficients $u^{a\to i}$ solving Eq. (28) have been determined, the 1RSB free energy can be computed, according to Eqs. (22):

$$\Delta\phi_V^i = -\frac{1}{y}\ln\left[1 + (1-e^{-y})\prod_{a\in\partial i} u^{a\to i}\right],$$

$$\Delta\phi_T^a = -\frac{1}{y}\ln\left[1 - (1-e^{-y})\prod_{i\in\partial a}\frac{\prod_{b\in\partial i\setminus a} u^{b\to i}}{e^{-y} + (1-e^{-y})\prod_{b\in\partial i\setminus a} u^{b\to i}}\right],$$

$$\Delta\phi_e^{(i,a)} = -\frac{1}{y}\ln\left[1 - (1-e^{-y})(1-u^{a\to i})\times\frac{\prod_{b\in\partial i\setminus a} u^{b\to i}}{e^{-y} + (1-e^{-y})\prod_{b\in\partial i\setminus a} u^{b\to i}}\right]. \tag{29}$$

In the minimal HS limit one has that $-y\phi=\Sigma-y\rho$. According to Eq. (24), the complexity Σ is recovered by Legendre transforming the function ϕ via the relation

$$\Sigma = y^2\frac{\partial\phi}{\partial y}, \tag{30}$$

which leads to the following equation for the density of active items:

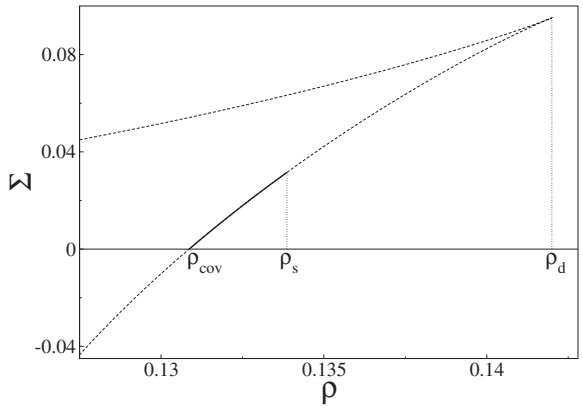

FIG. 3. Complexity Σ as a function of the density of active variables ρ for $L=4$ and $K=8$. For $\rho_{cov} \leq \rho \leq \rho_s$ the 1RSB ansatz is stable. ρ_{cov} (where $\Sigma=0$) is the minimal covering density. Below ρ_{cov} the complexity becomes negative: it is no longer possible to find solutions with smaller densities where all the clauses are satisfied and a covering-uncovering (COV-UNCOV) transition takes place. The 1RSB ansatz is no longer stable for $\rho > \rho_s$, where further breaking of the replica symmetry is expected to occur.

$$\rho = \phi + y\frac{\partial \phi}{\partial y}. \tag{31}$$

On infinite random-regular hypergraphs, the coefficients $u^{a\to i}$ can be assumed to be edge independent: $u^{a\to i}=u$. In this case the problem reduces to solving the single algebraic equation for the coefficient u:

$$u = 1 - \left(\frac{u^{L-1}}{e^{-y} + (1-e^{-y})u^{L-1}}\right)^{K-1}. \tag{32}$$

Once one has found the solution u to this equation, the 1RSB free energy simplifies to

$$\phi(y) = -\frac{1}{y}\ln\frac{1+(1-e^{-y})u^L}{(1-u)^{L/K}}. \tag{33}$$

The complexity $\Sigma(\rho)$, and the density ρ are then easily evaluated using Eqs. (30) and (31).

In Fig. 3 the complexity Σ is plotted as a function of ρ for $L=4$ and $K=8$. The complexity vanishes at ρ_{cov}, corresponding to the minimal HS density. At lower densities Σ is negative, implying that for $\rho < \rho_{cov}$ no solutions satisfying all the clauses can be found and a COV-UNCOV transition takes place. The complexity has a maximum at ρ_d (given by $\partial^2[y\phi(y)]=0$). The curve also displays a nonconcave part for $y < y_d$, the physical interpretation of which has not yet been understood.

B. Stability of the 1RSB solution

To determine whether the equilibrium state is really described by a 1RSB solution or whether further replica symmetry breakings occur, one has to study the stability of the 1RSB solution. The stability analysis of the RS ansatz investigates if the RS state tends to split into exponentially many states. Since in the 1RSB phase the Gibbs measure is decomposed in a cluster of different thermodynamic pure states [22], there are two different kinds of instabilities that might show up [15,23]: (i) either the states can aggregate into different clusters (in order to study this first instability one has to compute the interstate susceptibility), or (ii) each state can fragment in different states (in order to study this second instability the intrastate susceptibility must be computed).

In the minimal HS limit, the instability of the first kind can be easily studied by computing the eigenvalue of the (1×1) Jacobian matrix associated with Eq. (28) [15,23]. Since the linear susceptibility is related to a modulation instability incompatible with the underlying structure of the lattice, the nonlinear (spin glass) susceptibility must be considered. Therefore the criterion for the stability of the 1RSB ansatz simply reads

$$\sum_{j\in\partial a\backslash i; b\in\partial j\backslash a}\left(\frac{\partial u^{a\to i}}{\partial u^{b\to j}}\right)^2 \leq 1, \tag{34}$$

which yields

$$\sqrt{(L-1)(K-1)}\left|\frac{(1-u)e^{-y}}{u[e^{-y}+(1-e^{-y})u^{L-1}]}\right| \leq 1. \tag{35}$$

To study the instability of the second kind, instead, we consider a two-step RSB-like ansatz of the form $\mathcal{Q}[Q] \simeq \Sigma_l u_l \hat{\delta}[Q(u)-\delta(u-l)]$, and $\mathcal{P}[P] \simeq \Sigma_l g_l \hat{\delta}[P(h)-\delta(h-l)]$, where the 1RSB states coincide with the 2RSB clusters and the 2RSB states reduce to single configurations. We want to compute the widening of $\delta Q = \Sigma_{m\neq l}\epsilon_m[\delta(u-m)-\delta(u-l)]$, which can be written in terms of the widening of $\delta P = \Sigma_{q\neq r}\gamma_q[\delta(h-q)-\delta(h-r)]$, and check whether or not it grows under iteration. In order to do that we have to sum over all the perturbations of the cavity fields on the neighboring sites, which change the configuration of a given site from l to m, according to their Boltzmann weight (for a general explanation see [15,23]):

$$\begin{aligned} u_l^{a\to i}\langle\epsilon_m\rangle_l = \frac{1}{\mathcal{Z}_1} & \sum_{\{j_1,\ldots,j_{K-1}\}\in\partial a\backslash i;(g^{j_1\to a},\ldots,g^{j_{K-1}\to a})\to l} \\ & \times g^{j_1\to a}\cdots g^{j_{K-1}\to a} e^{(y_2-y_1)\Delta\phi_{T,iter}^{a\cup\partial a\backslash i}(g^{j_1\to a},\ldots,g^{j_{K-1}\to a})} \\ & \times \sum_{w,q\neq g^{j_w\to a},(g^{j_1\to a},\ldots,q,\ldots g^{j_{K-1}\to a})\to m} \\ & \times e^{y_2\Delta\phi_{T,iter}^{a\cup\partial a\backslash i}(g^{j_1\to a},\ldots,q,\ldots,g^{j_{K-1}\to a})}\langle\gamma_{g^{j_w\to a}}\rangle_q. \end{aligned} \tag{36}$$

A similar equation, which gives the $\langle\gamma_q\rangle_r$ in terms of the $\langle\epsilon_m\rangle_l$, can be derived. The problem can then be rewritten solely in terms of the coefficient u. Using a transfer matrix representation we find $\langle\epsilon_0\rangle_{-1}=(L-1)\times(K-1)\Sigma_{b,c=0,-1;b\neq c}T_{(0,-1)(b,c)}\langle\epsilon_b\rangle_c$. Therefore the 1RSB solution is stable, provided that the eigenvalue of T of largest modulus λ is such that $(L-1)(K-1)\lambda \leq 1$. The transfer matrix T reads (without losing generality we can set $y_1=y_2=y$)

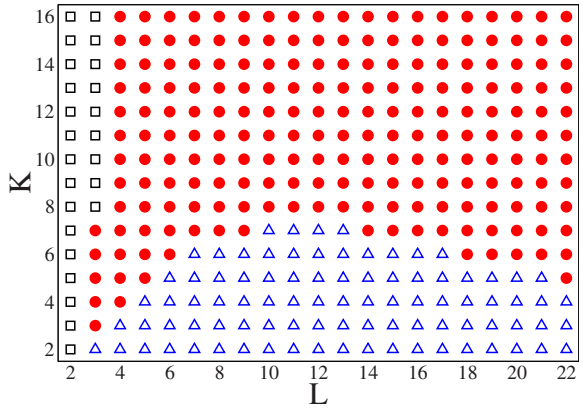

FIG. 4. (Color online) Phase diagram of the minimal hitting set problem. Black squares, red circles, and blue triangles correspond, respectively, to the values of L and K for which the minimal HS configurations are described by a replica symmetric ansatz, a one-step replica symmetry breaking ansatz, and a higher order replica symmetry breaking ansatz. As already shown in [2], for $K=2$ the solution of VC exhibits higher order RSB for every value of L.

$$T=\frac{1}{\mathcal{Z}}\begin{pmatrix} 0 & u^{(L-1)(K-1)}e^{-y} \\ u^{LK-L-K}(1-u) & 0 \end{pmatrix}, \qquad (37)$$

where the normalization is given by $\mathcal{Z}=[e^{-y}+(1+e^{-y})u^{L-1}]^{K-1}$. Therefore the stability criterion of the 1RSB solution is given by

$$(L-1)(K-1)\frac{\sqrt{e^{-y}u^{2(L-1)(K-1)-1}(1-u)}}{[e^{-y}+(1-e^{-y})u^{L-1}]^{K-1}} \leq 1. \qquad (38)$$

For some values of L and K, according to the stability criterions given in Eqs. (34) and (38), we find that the 1RSB solution is stable around the COV-UNCOV transition, and therefore the threshold is likely to be exact. On the contrary, for other values of L and K the 1RSB approach is unstable, and a higher order replica symmetry breaking transition is expected to occur. In this case a more involved analysis would be required to locate the COV-UNCOV threshold (the 1RSB result is expected to provide a lower bound). This happens, for instance, for $K=2$, where the results already found for the standard VC problem are fully recovered [2].

The phase diagram of the system as a function of the item and test degrees L and K is presented in Fig. 4, showing the relative positions of the different phases in the minimal covering limit.

V. SURVEY PROPAGATION AND SURVEY INSPIRED DECIMATION

We have already mentioned the possibility that, in analogy with K satisfiability and other optimization problems, the BP and SP equations, Eqs. (5), (26), and (27), may provide efficient algorithmic tools to find solutions of the HS problem on large instances [7,10]. In the present section, taking advantage of the analytical investigations presented above, we test numerically the efficiency of these message passing algorithms on large samples [24] (for several values of the connectivity), and compare their performance to a simple heuristic covering algorithm. We show that BP and SP provide an efficient way to solve the HS, being able to find covering patterns whose sizes are very close to the minimal one. Moreover, in the region of K and L where a one-step RSB occurs, we show that an SP algorithm allows us to improve the BP results, and to find solutions of the HS extremely close to the COV-UNCOV threshold predicted by the theoretical analysis.

In the following, we briefly describe the three algorithms used, and finally we present the results found.

A. Belief propagation (BP) algorithm

The BP algorithm simply consists in finding by iteration a solution of Eqs. (5) on a given factor graph, and then applying an iterative *decimation* procedure which, at each step, polarizes the most biased variables, until all the variables are fixed. Thus the BP algorithm works by iterating the three following steps:

(i) Solve the BP equations, Eqs. (5), on the graph, until all the messages converge to a fixed point.

(ii) Compute the marginals acting on each variable (i.e., the probability of being 0 or 1), and we polarize the most biased variable, by assigning to it the most probable value.

(iii) Generate the new *reduced graph*, by removing the variable that has been polarized and all the edges incoming on it. In the case where the variable has been polarized to 1, we also remove all the tests to which it is connected (since they are automatically satisfied) and all the edges incoming on those tests.

There is a problem concerning the convergence of BP that has to be mentioned. In particular, as discussed in the previous sections, depending on the value of K, L, and μ, the entropy of the RS solution of the (i.e., the number of solutions of the BP equations) can be either positive or negative. In order to make the BP equations converge, we fix the value of μ in such a way that RS entropy of the problem is positive. As a matter of fact, while the BP algorithm is iterated, the decimated graph modifies and consequently the entropy associated to the problem defined on the reduced graph changes. As a consequence, it can occur that the BP equations do not converge on the reduced graph, after that some variables have been fixed. In order to overcome this problem, during the decimation procedure we tune the chemical potential in such a way that the RS entropy is always kept positive. (Another possible and equivalent strategy consists in choosing at each decimation step the largest value of μ for which the BP equations converge to a fixed point.)

B. Survey propagation (SP) algorithm

In some regions of the phase space the BP equations possess a high number of solutions (corresponding to different thermodynamic states), and none can be found using a local iterative updating scheme. BP works well if a single cluster of minimal HS exists. However, a breaking of the replica symmetry implies the emergence of clustering in the solution space. This effect is captured by the SP algorithm, as first proposed in [10], which describes the statistics over all the

solutions of the BP equation, by taking into account their thermodynamic weight. Basically, the SP algorithm is very much like the BP one, and it consists of the same three steps as before. The only difference is that the SP messages, Eqs. (26) and (27), must now be used instead of the BP ones. In order to have a minimal HS covering of the factor graph, we would like to set the value of y to that corresponding to the COV-UNCOV transition, where the 1-RSB free energy has a maximum and the complexity vanishes. However, it can occur that, once some variables have been fixed, the SP equations stop to converge on the reduced graph for that value of y. This is due to the fact that while the decimation is carried on, the graph and, consequently, the complexity change, and y may now fall into the uncoverable region of the decimated problem. Therefore, in order to overcome this problem, after some decimation steps we recompute the complexity of the decimated problem defined on the reduced graph, and we tune the value of y to that corresponding to the new COV-UNCOV threshold.

C. Greedy algorithm

Here we present a very simple heuristic algorithm which allows us to find a covering pattern of the hypergraph. The algorithm consists of the following steps:

(i) We pick up the variable with highest degree and we set it to 1 (if more than one variable has the highest degree, we pick up a variable at random from among all those with the highest degree). We then remove from the graph this variable and all the edges incoming on it. We also remove all the tests connected to that variable (since they are satisfied) and all the edges incoming on those tests. We repeat this procedure until there are no variables left with degree larger than zero, and no more tests.

(ii) We fix to zero all the remaining variables, which are all isolated.

D. Results

We have tested these three algorithms on large instances for several values of L and K. It turns out that in general both BP and SP perform much better than the greedy algorithm, and are able to find efficiently solutions of the HS very close to the minimal COV-UNCOV threshold predicted by the analytical calculations (in general the size of the solutions provided by BP and SP is always less than a few per cent larger than the minimal one). In particular, for values of L and K for which a one-step RSB transition occurs (see Fig. 4), SP is able to improve the BP result. Just to give a more quantitative idea of the performance of the different algorithms, in Fig. 5 we present the results of the numerical tests of the three algorithms for $L=4$ and $K=6$.

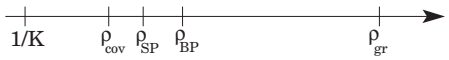

FIG. 5. Sizes of the covering of the hypergraph obtained by the BP, the SP, and the greedy algorithm for a HS problem with $L=4$ and $K=6$, for 12 288 variables and 8192 tests. The COV-UNCOV threshold in this case is equal to $\rho_{cov} \simeq 0.178$; the SP algorithm allows us to find solutions of size $\rho_{SP} \simeq 0.182$; the BP algorithm generates coverings of size $\rho_{BP} \simeq 0.186$. Finally, the greedy algorithm gives solutions of size $\rho_{gr} \simeq 0.212$.

VI. DISCUSSION AND CONCLUSION

In this paper we have studied the statistical mechanics of a generalized vertex covering problem on the hypergraph, which might have many practical applications. As an example, the problem of group testing is deeply related to the HS, and the knowledge of the phase diagram of the latter could give important insights on how to devise an efficient reconstruction algorithm for the former, depending on L, K, and on the density of active items.

The minimal HS has been studied in great detail. For a low enough degree of the items ($L=2$, $L=3$, $K>7$, $L=4$, and $K>21$, ...) we find that the RS solution describes correctly the minimal covering configurations. For bigger values of L, the RS ansatz becomes inconsistent, whereas the 1RSB solution is stable and is likely to provide the correct solution around the COV-UNCOV threshold. Both in the RS and in the 1RSB region we have found explicit results on the minimal density of active items required to cover the factor graph, and on the structure of the phase space. On the other hand, there are also cases (e.g., the ordinary VC problem [2,3], $K=2$ and $L \geq 3$) where further RSB steps are required. In these cases the 1RSB approach becomes inconsistent and higher RSB patterns are required.

Finally, we have shown that a decimation procedure based on the BP and SP equations turns out to be a very efficient strategy to solve large individual instances of the HS problem.

ACKNOWLEDGMENTS

We warmly thank L. Zdeborová for kind and fruitful discussions, and P. Tetali for pointing out to us the literature on hitting sets. This work was partially supported by EVERGROW (EU Consortium FP6 IST).

[1] S. Cook, *Proceedings of the Third Annual ACM Symposium on Theory of Computing* (AMC, New York, 1971), pp. 151–158; M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness* (Freeman, New York, 1979).

[2] M. Weigt and A. K. Hartmann, Phys. Rev. E **63**, 056127 (2001); Phys. Rev. Lett. **84**, 6118 (2000).

[3] M. Weigt and H. Zhou, Phys. Rev. E **74**, 046110 (2006).

[4] M. Bauer and O. Golinelli, Eur. Phys. J. B **24**, 339 (2001); H. Zhou, *ibid.* **32**, 265 (2003); Phys. Rev. Lett. **94**, 217203 (2005).

[5] A. Frieze, Discrete Math. **81**, 171 (1990); P. Gazmuri, Net-

works **14**, 367 (1984).
[6] M. Mézard and G. Parisi, Eur. Phys. J. B **20**, 217 (2001); J. Stat. Phys. **111**, 1 (2003).
[7] M. Mézard, G. Parisi, and R. Zecchina, Science **297**, 812 (2002).
[8] R. Monasson and R. Zecchina, Phys. Rev. Lett. **76**, 3881 (1996).
[9] G. Biroli, R. Monasson, and M. Weigt, Eur. Phys. J. B **14**, 551 (2000).
[10] M. Mézard and R. Zecchina, Phys. Rev. E **66**, 056126 (2002).
[11] F. Krzakala, A. Pagnani, and M. Weigt, Phys. Rev. E **70**, 046705 (2004); R. Mulet, A. Pagnani, M. Weigt, and R. Zecchina, Phys. Rev. Lett. **89**, 268701 (2002); L. Zdeborová and F. Krzakala, Phys. Rev. E **76**, 031131 (2007).
[12] M. Mézard, F. Ricci-Tersenghi, and R. Zecchina, J. Stat. Phys. **111**, 505 (2003).
[13] J. Raymond, A. Sportiello, and L. Zdeborová, Phys. Rev. E **76**, 011101 (2007).
[14] G. Biroli and M. Mézard, Phys. Rev. Lett. **88**, 025501 (2001); A. Hartmann and M. Weigt, Europhys. Lett. **62**, 533 (2003); M. P. Ciamarra, M. Tarzia, A. de Candia, and A. Coniglio, Phys. Rev. E **67**, 057105 (2003).
[15] O. Rivoire, G. Biroli, O. Martin, and M. Mézard, Eur. Phys. J. B **37**, 55 (2004).
[16] T. Berger, IEEE Trans. Inf. Theory **48**, 1741 (2002); H. Q. Ngo and D.-Z. Du, Discrete Mathematical Problems with Medical Applications **55** (2000).
[17] M. Mézard and C. Toninelli, e-print arXiv:0706.3104.
[18] D. Dorfman, Ann. Math. Stat. **14**, 436 (1943).
[19] M. Sobel and P. A. Groll, Bell Syst. Tech. J. **38**, 1179 (1959).
[20] W. H. Kautz and R. R. Singleton, IEEE Trans. Inf. Theory **10**, 363 (1964).
[21] C. H. Li, J. Am. Stat. Assoc. **57**, 455 (1962).
[22] M. Mézard, G. Parisi, and M. A. Virasoro, *Spin-glass Theory and Beyond*, Lecture notes in Physics Vol. 9 (World Scientific, Singapore, 1987).
[23] A. Montanari and F. Ricci-Tersenghi, Eur. Phys. J. B **33**, 339 (2003).
[24] In order to generate random regular hypergraphs, we employ the fast fully polynomial randomized approximation scheme introduced in M. Bayati, J. Han, and A. Saberi, *Proceedings of the International Workshop on Randomization and Computation (RANDOM)*, 2007, e-print arXiv:cs/0702124v2. Basically, the algorithm is as follows: one starts with an empty graph and sequentially adds edges between variables i and function nodes a with probability proportional to $\hat{d}_i\hat{d}_a$, where $\hat{d}_i$ and $\hat{d}_a$ denote, respectively, the remaining degrees of the variable i and the function node a.