Journal of Statistical Mechanics: Theory and Experiment
An IOP and SISSA journal

The cavity method for large deviations

Olivier Rivoire

Laboratoire de Physique Théorique et Modèles Statistiques, Université Paris-Sud, Bâtiment 100, 91405 Orsay Cedex, France
E-mail: rivoire@lptms.u-psud.fr

Received 7 June 2005
Accepted 29 June 2005
Published 14 July 2005

Online at stacks.iop.org/JSTAT/2005/P07004
doi:10.1088/1742-5468/2005/07/P07004

Abstract. A method is introduced for studying large deviations in the context of statistical physics of disordered systems. The approach, based on an extension of the cavity method to atypical realizations of the quenched disorder, allows us to compute exponentially small probabilities (rate functions) over different classes of random graphs. It is illustrated with two combinatorial optimization problems, the vertex-cover and colouring problems, for which the presence of replica symmetry breaking phases is taken into account. Applications include the analysis of models on adaptive graph structures.

Keywords: cavity and replica method, disordered systems (theory), random graphs, networks, typical-case computational complexity

ArXiv ePrint: cond-mat/0506164

J. Stat. Mech. (2005) P07004

1742-5468/05/P07004+28$30.00

Contents

1. Introduction

The algorithmic complexity of a problem is traditionally measured on an ensemble of possible inputs (instances) by retaining the largest time it takes for an algorithm to solve one of the instances [1]. Statistical physics studies have, however, suggested a different characterization of hardness, based on the average case rather than the worst case [2]. This alternative approach is motivated by a generic phenomenon of concentration, according to which a particular instance behaves almost surely as the average case in the limit of infinite size. This self-averaging property is common to many disordered systems whose environment is specified by a quenched random variable from a prescribed ensemble: in the thermodynamical limit, the properties of a sample tend to be independent of the particular realization of the disorder. Due to this parallel, the methods first developed for physical disordered systems have been successfully applied to combinatorial optimization problems [2]. The interest in typical properties is, however, far from limited to physics or optimization; information [3] and graph theories [4] are two other major fields where they play a key role, as illustrated by the seminal works of Shannon [5] and Erdős–Rényi [6], respectively.

In any practical implementation, however, optimization or coding theories face large but finite system sizes. In such situations, controlling the deviations from the typical case becomes of primarily practical interest. The scope of large deviations theory [7] is precisely to evaluate the probability of the rare events associated with such finite size effects. In addition, and despite the tremendous number of their elements, large

doi:10.1088/1742-5468/2005/07/P07004

J. Stat. Mech. (2005) P07004

deviations theory also has a direct relevance for physical systems: as will be explained, it underlies the thermodynamics of systems whose configuration space is constituted of different realizations of the quenched disorder. Consequently, when the disorder consists of an ensemble of random graphs, it allows us to solve models with variable topologies. A large deviations analysis is thus useful to address the adaptability of constrained systems, with a physical example being covalent molecular networks subject to stress [8, 9].

A particularly powerful method for computing typical properties of disordered systems is the celebrated replica method [2]. The cavity method [2] provides an alternative tool that yields equivalent results but has two advantages over the replica method: it is based on assumptions formulated explicitly, and it applies to particular instances. In the present paper, we develop an approach of large deviations based on the cavity method, which we call the large deviations cavity method (LDCM). If large deviations have only recently raised an interest in the statistical physics community [10]–[15], they have a much longer history in probability theory [7], where they are notably used to rigorously solve statistical mechanics models [16]. At variance with this mathematical tradition, the method exposed here is non-rigorous and only provides a coherent heuristic framework for obtaining quantitative predictions, namely computing rate functions assuming a large deviations principle indeed holds. Nonetheless, as for the 'typical' cavity method of Mézard and Parisi [17, 18] that we will recover as a particular case, the LDCM is hoped to be amenable to rigorous studies.

The paper is organized as follows. The first section is devoted to introducing some basic elements of combinatorial optimization and large deviations theories, with an emphasis on their links with statistical mechanics. The second section presents the LDCM in its simplest, 'replica symmetric', form: we start by rederiving in a cavity-like fashion Cramér's theorem, the most elementary result in large deviations theory, and then discuss different graph ensembles, with explicit calculations on the vertex-cover problem. The third section deals with systems having a non-trivial internal structure: we notably generalize the large deviations approach to systems displaying a glassy 'replica symmetry breaking' phase, a situation that we illustrate in detail with the colouring problem. A conclusion closes the paper by suggesting some possible applications.

2. Optimization problems and large deviations

2.1. Optimization problems

The field of combinatorial optimization provides a broad class of disordered systems, which we use here to illustrate the potentialities of the LDCM. We therefore start with a brief introduction to this subject (see e.g. [1] for more details). Combinatorial optimization is primarily concerned with minimizing cost functions, $\mathcal{E} : \mathcal{C} \mapsto \mathbb{R}$, over some discrete configuration space $\mathcal{C}$. In view of quantifying their algorithmic complexity, optimization problems are defined over an ensemble $\mathcal{I}$ of instances I, each associated with a cost function $\mathcal{E}_I$. In particular, many combinatorial problems are defined over an ensemble of graphs [19], in which case an instance I is specified by a graph G, that is a set of N nodes $i = 1, \ldots, N$ associated with a subset of the pairs $\{(i, j)\}_{i \neq j}$, defining its edges.

Two prototypical examples will serve as an illustration. The first one is the vertex-cover problem [19], also known as independent set, which consists, given a graph G, in blackening as many of its nodes as possible while never blackening two connected nodes.

J. Stat. Mech. (2005) P07004

The second, quite similar, example is the colouring problem [19] which asks, given a graph G and q colours, whether it is possible to assign a colour to each node of G so that no two adjacent nodes have same colour. The colouring problem is strictly speaking a decision problem (the answer must be yes or no) but it is directly related to the optimization problem of minimizing the number of edges having two end-nodes sharing a same colour: if the minimum is zero, the graph is colourable; otherwise, it is uncolourable.

From the statistical physics viewpoint, the cost $\mathcal{E}_I[\sigma]$ represents the energy of a configuration $\sigma \in \mathcal{C}$, and the minimal cost $E_I = \min_\sigma \mathcal{E}_I[\sigma]$ corresponds to the ground-state energy of the disordered system having quenched disorder I. In this context, it is usual to introduce an inverse temperature β and a free-energy density $f_I(\beta)$ defined by $f_I(\beta) = -\ln[\sum_\sigma \exp(-\beta\mathcal{E}_I[\sigma])]/(\beta N)$, such that the ground-state energy density $\epsilon_I = E_I/N$ is given by the $\beta \to \infty$ limit, $\epsilon_I = \lim_{\beta\to\infty} f_I(\beta)$. For the colouring problem, the associated finite-temperature system is known as the antiferromagnetic Potts model [20], while it is called the hard-core model [21] for vertex-cover (with β representing a chemical potential).

For these two examples, the configuration space $\mathcal{C}$ is discrete: $\mathcal{C}$ can be taken as $\{0,1\}^N$ for vertex-cover, with $\sigma_i = 0$ and 1 corresponding respectively to uncovered (white) and covered (black) nodes, and as $\{1,\dots,q\}^N$ for colouring with $\sigma_i \in \{1,\dots,q\}$ now representing the colour assigned to i. Different ensembles $\mathcal{G}$ of graphs, defining different sets of instances $\mathcal{I}$, can be introduced. The cavity approach followed here applies to any ensemble of locally 'tree-like' graphs, that is, graphs whose degrees (the number of nodes to which a given node is connected) remain finite when $N \to \infty$. Three random graph ensembles will be specifically addressed here. The first one, denoted $\tilde{\mathcal{G}}_N^{(\gamma)}$, is the set of graphs with N nodes where each edge as a probability γ/N to be present, and is known as the binomial model or Erdős–Rényi ensemble [4]. The second one, denoted $\bar{\mathcal{G}}_N^{(\gamma)}$, and called the uniform model [4], is the set of graphs with N nodes and $M = \gamma N/2$ edges. Finally, the third one is defined through the degree distribution $p(k)$ of the nodes of its graphs [22]: each of the N nodes has degree k with independent probability $p(k)$ and the edges are drawn at random subject to that constraint. This last class notably includes random regular graphs [4], for which $p(k) = \delta_{r,k}$, and power-law distributed graphs, for which $p(k) \sim k^\tau$ (with an appropriate cut-off to insure normalization). Here we will also consider the Poissonian model denoted $\hat{\mathcal{G}}_N^{(\gamma)}$ and defined by $p(k) = \gamma^k \mathrm{e}^{-\gamma}/k!$. In the $N \to \infty$ limit, the binomial and uniform models $\tilde{\mathcal{G}}_N^{(\gamma)}$ and $\bar{\mathcal{G}}_N^{(\gamma)}$ share with $\hat{\mathcal{G}}_N^{(\gamma)}$ the same Poisson degree distribution. This equivalence between the three models extends to the typical properties of optimization problems defined on them but, as will be shown, does not hold for atypical features.

From the point of view of computational complexity, an important parameter is the size N of the instances, which, in the case of diluted graphs, is taken as the number of nodes. As seen on the vertex-cover and colouring problems, the size of the configuration space $\mathcal{C}$ over which optimization is to be performed increases exponentially with N, precluding any naive exhaustive search for large N and possibly making the problem highly non-trivial. In fact, both the vertex-cover and colouring problems are known to be NP-hard (NP stands for Non-deterministic Polynomial [1]) in the worst case, implying that no algorithm is known that can solve all instances of these problems in a time growing polynomially with N [1].

J. Stat. Mech. (2005) P07004

As stressed in the introduction, the focus on typical instances advocated by statistical physics is justified by the self-averaging property: when it holds for an ensemble $\mathcal{I}$ of instances, there is a typical value of the ground-state energy density $\bar{\epsilon}$ such that, for any $\varepsilon > 0$, the probability $\mathbb{P}[|E_I/N - \bar{\epsilon}| > \varepsilon]$ for the optimum E_I to deviate from $N\bar{\epsilon}$ goes to zero, $\mathbb{P}[|E_I/N - \bar{\epsilon}| > \varepsilon] \to 0$ as $N \to \infty$. Informally, large problems then typically all share a common optimum, which, physically, can often be traced back to the equivalence of their local properties. The 'typical' cavity method [17, 18] has been developed precisely to compute the most probable value $\bar{\epsilon}$ for problems on random graphs. The LDCM presented here is an extension of this approach that allows us to evaluate the N and ε dependences of vanishing probabilities such as $\mathbb{P}[|E_I/N - \bar{\epsilon}| > \varepsilon]$.

2.2. Large deviations

For finite N, an instance has always a finite probability to deviate from the typical case. The so-called large deviations [7] refer to the extensive deviations from $N\bar{\epsilon}$, of order $\mathrm{O}(N)$, as distinguished from the small, subextensive deviations from $N\bar{\epsilon}$, of order $\mathrm{o}(N)$, like for example $\mathrm{O}(\sqrt{N})$ fluctuations (see, however, below for a relation between the two). The present method is based on an ansatz, according to which large deviations are exponentially small in the size N of the instances; that is, the probability $\mathbb{P}_N[E_I]$ for an instance I taken out of the ensemble $\mathcal{I}$ to have an optimal cost E_I is supposed to satisfy

$$\mathbb{P}_N[E_I = N\epsilon] \asymp \mathrm{e}^{-NL(\epsilon)}, \tag{1}$$

where the symbol $\asymp$ stands here and in what follows for an exponential equivalence defined as $\lim_{N\to\infty} \ln(\mathbb{P}[E_I = N\epsilon])/N = -L(\epsilon)$. $L(\epsilon)$ is called a rate function, or large deviations function, and, in the simplest cases, is strictly positive, except for the typical value $\bar{\epsilon}$ where it achieves its zero minimum. The ansatz (1) is indeed known to hold in the solvable case where $\mathcal{E}_I$ consists of a sum of independent identically distributed variables (Cramér's theorem; see section 3.1), and this result is robust to the presence of weak correlations among the variables (Gärtner–Ellis theorem, to be stated below) [7]. More precisely, the relation (1) corresponds in the mathematical literature to the 'large deviations principle' [7], which, in its simplest form, can be stated as follows.

Large deviations principle. The sequence $\{\epsilon_N\}_{N\in\mathbb{N}}$ of real-valued random variables is said to satisfy the large deviations principle, with rate function $L : \mathbb{R} \to \mathbb{R}^+ \cup \{\infty\}$, if

(i) $\forall M \geq 0, \quad \{x \in \mathbb{R} : L(x) \leq M\}$ is compact,

(ii) for all closed subset F of $\mathbb{R}$, and all open subset O of $\mathbb{R}$,

$$\limsup_{N\to\infty} \frac{1}{N}\mathbb{P}_N[\epsilon_N \in F] \leq -L(F), \qquad \liminf_{N\to\infty} \frac{1}{N}\mathbb{P}_N[\epsilon_N \in O] \geq -L(O). \tag{2}$$

We point out right away that systems exist for which the previous ansatz does not hold. In the field of spin-glass models, an example is provided by the SK model [23] for which numerical studies [10] suggest different scalings on both sides of the typical value, that is, $\mathbb{P}_N(\epsilon) \asymp \exp[-N^a L(\epsilon)]$ with $a \simeq 1.2$ when $\epsilon < \bar{\epsilon}$, but $a \simeq 1.5$ when $\epsilon > \bar{\epsilon}$. However, in a variety of other spin-glass models, notably including models on diluted random graphs,

J. Stat. Mech. (2005) P07004

the ansatz (1) is supported by numerical evidence [10]. An exactly solvable case is the random energy model [24] where elementary calculations [10, 11] yield

$$\mathbb{P}_N(\epsilon) \asymp \begin{cases} \exp[-\exp(Ns(\epsilon))] & \text{if } \epsilon > \bar{\epsilon}, \\ \exp(Ns(\epsilon)) & \text{if } \epsilon < \bar{\epsilon}, \end{cases} \tag{3}$$

with $s(\epsilon) = \ln 2 - \epsilon^2$ and $\bar{\epsilon} = -\sqrt{\ln 2}$; the rate function of this model is therefore $L(\epsilon) = -s(\epsilon)$ for $\epsilon \leq \bar{\epsilon}$ and $L(\epsilon) = +\infty$ for $\epsilon > \bar{\epsilon}$.

From the analytical viewpoint, rate functions in the context of optimization problems have been studied by Montanari [12, 25], using the replica method. The results he obtained for the vertex-cover problem [12, 25] are strictly equivalent to the ones to be derived here from the cavity method. Yet, as for the typical case, the cavity approach has the advantages over the replica method to offer a more transparent derivation, and to open the way to algorithmic implementations on particular systems. For a model at finite temperature $1/\beta$, the replica method basically consists in inferring rate functions from the knowledge of the moments $\mathbb{E}[Z_I^n]$ of the partition function $Z_I(\beta) = \sum_\sigma \exp(-\beta \mathcal{E}_I[\sigma])$, with $\mathbb{E}[\cdot]$ referring to the average over the disorder, that is, the different instances I. As far as no replica symmetry breaking is involved, this procedure is motivated by the following rigorous result [7].

Gärtner–Ellis theorem. Let $\{\epsilon_N\}_{N\in\mathbb{N}}$ be a sequence of real-valued random variables and let $\mathcal{F} : \mathbb{R} \to \mathbb{R}$ be defined by

$$\mathcal{F}(y) = -\lim_{N\to\infty} \frac{1}{N} \ln \mathbb{E}[\mathrm{e}^{-yN\epsilon_N}]. \tag{4}$$

If $\mathcal{F}(y)$ exists, is finite and differentiable for every $y \in \mathbb{R}$, then the sequence $\{\epsilon_N\}_{N\in\mathbb{N}}$ satisfies the large deviations principle with rate function $L(\epsilon)$ given by the Legendre transform of $\mathcal{F}(y)$,

$$-L(\epsilon) = \inf_{y\in\mathbb{R}} [y\epsilon - \mathcal{F}(y)], \tag{5}$$

where minus signs are introduced to match usual conventions in statistical physics. We stress that, for sake of simplicity, this theorem is stated here with much stronger hypothesis than necessary; in particular the assumptions about the finiteness and differentiability of $\mathcal{F}$ can be relaxed [7].

To apply the replica method to optimization problems, the limit $\beta \to \infty$ has to be considered, and, to obtain non-trivial results in this limit, the replica number n must be rescaled with β, such that $\beta \to \infty$ and $n \to 0$ with $y = n/\beta$ finite. In this limit, the replica potential $\mathcal{F}(y)$ coincides with the function introduced in Gärtner–Ellis theorem, equation (4),

$$\mathrm{e}^{-N\mathcal{F}(y)} = \lim_{\beta\to\infty} \mathbb{E}[Z^{y/\beta}] = \mathbb{E}[\mathrm{e}^{-yE}] = \int \mathrm{e}^{-N[L(\epsilon)+y\epsilon]}\, \mathrm{d}\epsilon. \tag{6}$$

While proceeding differently, the cavity method to be presented will lead to the same rate function, again specified as the Legendre transform of the potential $\mathcal{F}(y)$. Although both the replica and the cavity methods, based on Legendre transformations, naturally yield convex functions, it should be stressed that convexity is not a necessary feature of rate

J. Stat. Mech. (2005) P07004

functions. In fact, non-convex rate functions are associated with phase transitions and are therefore encountered in many models of interest from the statistical mechanics point of view [16].

Large deviations deal with exponentially small probabilities and may appear as only an extreme feature of finite size effects, while a more refined description would consist in the complete probability distribution of E_I over $\mathcal{I}$. Interestingly, under the assumption that taking the $N \to \infty$ limit can be interchanged with taking the derivative with respect to y, small fluctuations can be extracted from the knowledge of the rate function near its typical minimum by expressing the cumulants $\langle (E_N)^k \rangle_c$ as

$$\langle (E_N)^k \rangle_c = (-1)^{k+1} N \frac{\partial^k \mathcal{F}}{\partial y^k}(y = 0), \tag{7}$$

where, as usual, the cumulants $\langle X^k \rangle_c$ of a random variable X are defined by $\ln \mathbb{E}[\mathrm{e}^{tX}] = \sum_{k=1}^{\infty} (t^k/k!) \langle X^k \rangle_c$. This relation holds for instance with sums of independent identically distributed variables, where it predicts the variance of the small fluctuations to be of order $\sqrt{N}$, in agreement with the central limit theorem. However, equation (7) may not be valid, as in cases where the rate function has multiple minima (physically corresponding to a phase transition).

2.3. Statistical mechanics interpretation

As well as having their own mathematical interest, large deviations are of direct relevance to statistical mechanics studies. In the context of optimization problems, rate functions can indeed be interpreted as defining an entropy on the space of the instances $\mathcal{I}$, corresponding to a thermodynamics over the quenched disorder. This relation, formalized by Sanov's theorem [16], is presented here in the restricted context where $\mathcal{I}$ is a class of graphs associated with a given optimization problem.

Viewing the ensemble of random graphs $\mathcal{G}_N$ as a phase space, each graph $G \in \mathcal{G}_N$ defines a configuration to which is associated the ground-state energy E_G, that is, the optimal cost for the optimization problem on G. If $|\mathcal{G}_N| \equiv \mathrm{e}^{N s_0}$ denotes the cardinality of $\mathcal{G}_N$, the microcanonical entropy $s(\epsilon)$ of the system is given by

$$\mathrm{e}^{N s(\epsilon)} = \#\{G \in \mathcal{G}_N; E_G = N\epsilon\} = \#\mathcal{G}_N \times \mathbb{P}_N[E_G = N\epsilon] \asymp \mathrm{e}^{N[s_0 - L(\epsilon)]}, \tag{8}$$

where $\#A$ denotes the cardinality of the set A. Thus, up to a linear transformation, the rate function $L(\epsilon)$ is nothing but the microcanonical entropy $s(\epsilon)$,

$$s(\epsilon) = s_0 - L(\epsilon). \tag{9}$$

Within this picture, the parameter y appearing in the replica method and Gärtner–Ellis theorem represents the external inverse temperature that allows us to study statistical mechanics on the configuration space spanned by the graphs, $y \equiv -\partial_\epsilon L(\epsilon) = \partial_\epsilon s(\epsilon)$ (y must be distinguished from the internal inverse temperature β which is set to infinity in the context of optimization). By construction, this space has no more quenched disorder, and a large deviations analysis appears as the statistical mechanics analysis of a pure system at finite inverse temperature y. From the opposite viewpoint, large deviations theory thus provides a meaning for negative temperatures, $y < 0$. Finally, the typical case is given by the infinite temperature limit, $y = 0$, as prescribed by replica theory.

J. Stat. Mech. (2005) P07004

We stress, however, that the possibility of deriving the thermodynamics of the system at inverse temperature y from the knowledge of its microcanonical entropy $s(\epsilon)$ is based on the equivalence between the microcanonical and canonical ensembles in the thermodynamical limit, which cannot always be taken for granted. In presence of non-convex rate functions, indeed, the two ensembles become inequivalent, and a first-order transition occurs, whose description requires a Maxwell construction; such a construction in the context of large deviations for a random graph has been recently described in [13].

We have restricted ourselves so far to the simplest case where the measure over the quenched disorder is an uniform measure over an ensemble of graphs, but more complicated structures can be considered as well. In particular, the disorder can have different origins, as with spin-glass models [2] or K-SAT optimization problems [1], where in addition to the graph structure, the quenched disorder comprises the specification of some random couplings between the variables. In this case, an instance I of the problem is first defined by selecting a graph G and then by choosing the couplings J. Large deviations can be taken with respect to J at fixed G: for typical graphs the effective system still contains a quenched disorder (the graph) which can be handled with the usual techniques of disordered systems, but if atypical graphs have to be addressed as well, a second temperature needs to be introduced. The two temperatures are in such a case associated with two levels of probability distributions, in a construction formally identical to Parisi's hierarchical scheme for handling replica symmetry breaking, as will be discussed in section 4.3. The same scheme also applies when going to lower levels to describe the internal structure of a given instance. This will be exemplified in section 4.1 where we discuss the implications of working with a finite temperature on the instances, or working with optimization problems displaying a replica symmetry breaking phase.

3. The large deviations cavity method

The 'typical' cavity method, as developed by Mézard and Parisi [17, 18], applies to a given instance I and addresses the structure of its phase space, that is the organization of the configurations $\sigma \in \mathcal{C}$ as a function of their energy density $\mathcal{E}[\sigma]/N$. The method can handle either a structure composed of a unique set (or a finite number of sets) of connected solutions, called a *replica symmetric* (RS) phase, or a structure composed of many disconnected clusters of configurations, called a *one-step replica symmetry breaking* (1RSB) phase [2]. In the latter case, the crucial assumption is made that the number of clusters with a given energy density ϵ is exponential in N,

$$\mathcal{N}_{\text{clusters}}(\epsilon) \asymp \mathrm{e}^{N\Sigma(\epsilon)}. \tag{10}$$

The 1RSB cavity method is specifically designed to compute the function $\Sigma(\epsilon)$, called the *complexity*, with the particular RS case corresponding to $\Sigma = 0$ [26]. The formal analogy between the 1RSB ansatz (10) defining the complexity $\Sigma(\epsilon)$ and the large deviations ansatz (1) defining the rate function $L(\epsilon)$ is at the root of the possibility to extend the typical cavity method yielding $\Sigma(\epsilon)$ to an atypical version yielding $L(\epsilon)$. To emphasize the parallel further, we introduce the function $\mathcal{L}(\epsilon)$ defined as $\mathcal{L}(\epsilon) \equiv -L(\epsilon)$, such that $\mathcal{L}(\epsilon)$ plays in the LDCM a role formally identical to the complexity in the typical cavity method:

$$\mathbb{P}_N(E = N\epsilon) \asymp \mathrm{e}^{N\mathcal{L}(\epsilon)}. \tag{11}$$

J. Stat. Mech. (2005) P07004

The analogy between the complexity $\Sigma(\epsilon)$ and the rate function $L(\epsilon)$ should not be taken as a coincidence: the complexity is fundamentally nothing but a rate function (or more accurately the entropy associated to it, as in equation (9)), which describes the large deviations of the energy over the different clusters of solutions. From this point of view, further elaborated in [27], the 1RSB cavity method is itself a large deviations method, acting on the self-generated (glassy) 'internal disorder' of a given sample. For glassy optimization problems, being able to address such large deviations is crucial since ground-state clusters are atypical, that is, exponentially less numerous than clusters with higher energies. These atypical ground-state clusters must be obtained by correctly tuning the 'internal inverse temperature', denoted μ in this context. Remarkably, while the LDCM to be presented will also apply to the typical case $y = 0$, the 1RSB cavity method is in general not able to describe the complete complexity curve $\Sigma(\epsilon)$, and notably fails to describe the most numerous, typical clusters, corresponding to $\mu = 0$.[1]

Our presentation of the LDCM will follow closely the presentation of Mézard and Parisi of their typical cavity method [17, 18], but major differences will show up in the way averages over the disorder are performed. To start with, we consider the simplest case where the underlying optimization problem is assumed to be itself RS; i.e., with no clustering induced by its internal disorder.

3.1. The cavity approach to Cramér's theorem

Although its most interesting applications involve random graphs, the cavity method is not restricted to this particular geometry. As an illustration of the ideas in their simplest setting, we consider the case, with no geometry, of a system made of N independent elements, each contributing to the total energy $\mathcal{E}_N$ by a random amount X_i. In other words, we consider here large deviations in the sum of independent identically distributed random variables. For such a system, the typical energy density follows from the law of large numbers, which, assuming the distribution $\rho(X)$ of the X_is to have a finite first moment, is $\bar{\epsilon} = \mathbb{E}[X] \equiv \int x\rho(x)\,\mathrm{d}x$. Large deviations are concerned with deviations from the prediction $\epsilon_N/N = \bar{\epsilon}$ and, for a sum of independent variables, are completely specified by Cramér's theorem, both a generalization of the law of large numbers and a corollary of Gärtner–Ellis theorem [7].

Cramér's theorem. Let the sequence $\{\epsilon_N\}_{N\in\mathbb{N}}$ of real random variables be given by $\epsilon_N = (\sum_{i=1}^N X_i)/N$, where the $\{X_i\}_i$ are independently identically distributed real random variables. If $\mathbb{E}[\mathrm{e}^{-yX}]$ is finite for all $y \in \mathbb{R}$, then $\{\epsilon_N\}_{N\in\mathbb{N}}$ satisfies the large deviations principle with rate function $L : \mathbb{R} \to \mathbb{R}$ defined as the Legendre transform of $\mathcal{F} : \mathbb{R} \to \mathbb{R}$ given by $\mathcal{F}(y) \equiv -\ln \mathbb{E}[\mathrm{e}^{-yX}]$, that is,

$$-L(\epsilon) \equiv \mathcal{L}(\epsilon) = \inf_{y\in\mathbb{R}} [y\epsilon - \mathcal{F}(y)]. \tag{12}$$

The basic idea behind the cavity approach is to estimate the change of the system upon the addition of a new variable (or, equivalently, upon the removal of a variable; hence the name 'cavity'). Let $\mathcal{E}_N$ be the extensive energy, $\mathcal{E}_N = \sum_{i=1}^N X_i$. By virtue of

[1] This disability originates from the presence of marginally stable states and can be bypassed by considering a generalization (supersymmetry breaking version) of the cavity method [42].

J. Stat. Mech. (2005) P07004

the assumed independence of the X_i, the probability distribution for $\mathcal{E}_{N+1} = \mathcal{E}_N + X_{N+1}$ is given by a convolution of those of $\mathcal{E}_N$ and X_{N+1}, which, with the ansatz (11), reads

$$\begin{aligned}\mathbb{P}_{N+1}(\mathcal{E}_{N+1} = E) &= \mathrm{e}^{(N+1)\mathcal{L}(E/(N+1))} = \mathbb{E}_X[\mathbb{P}_N(\mathcal{E}_N = E - X)] \\ &= \int \rho(\Delta E) \exp\left(N\mathcal{L}\left(\frac{E-\Delta E}{N}\right)\right) \mathrm{d}\Delta\, E. \end{aligned} \tag{13}$$

Assuming a smooth behaviour of $\mathcal{L}$, we write for large N,

$$\begin{aligned}(N+1)\mathcal{L}\left(\frac{E}{N+1}\right) &= N\mathcal{L}(\epsilon) + \mathcal{L}(\epsilon) - \epsilon\partial_\epsilon\mathcal{L}(\epsilon) + \mathrm{O}(1), \\ N\mathcal{L}\left(\frac{E-\Delta E}{N}\right) &= N\mathcal{L}(\epsilon) - \Delta E\partial_\epsilon\mathcal{L}(\epsilon) + \mathrm{O}(1), \end{aligned} \tag{14}$$

where $\epsilon \equiv E/N$. Setting $y \equiv \partial_\epsilon\mathcal{L}(\epsilon)$ thus yields

$$\mathcal{F}(y) \equiv y\epsilon - \mathcal{L}(\epsilon) = -\ln\mathbb{E}[\mathrm{e}^{-yX}]. \tag{15}$$

We conclude that $\mathcal{L}(\epsilon)$ is given by the Legendre transform of the potential $\mathcal{F}(y)$,

$$\begin{aligned}\mathcal{L}(\epsilon) &= \epsilon y - \mathcal{F}(y), \\ \epsilon &= \partial_y\mathcal{F}(y), \end{aligned} \tag{16}$$

as prescribed by Cramér's theorem.

3.2. Poissonian random graphs

We consider now models defined on random graphs, first under the assumption that the internal structure of an instance is replica symmetric (RS). As a further simplification (to be relaxed later on, as for the RS hypothesis), we assume that the only source of quenched disorder lies in the graph structure, as is the case for the vertex-cover and colouring problems. We consider here simultaneously the three ensembles of random graphs, $\tilde{\mathcal{G}}_N^{(\gamma)}$, $\bar{\mathcal{G}}_N^{(\gamma)}$ and $\hat{\mathcal{G}}_N^{(\gamma)}$ defined in section 2.1 and hereafter generically referred to as $\mathcal{G}_N^{(\gamma)}$. In the $N \to \infty$ limit, the degrees of graphs in $\mathcal{G}_N^{(\gamma)}$ have the same limiting distribution π_γ, where π_γ denotes the Poisson distribution with mean γ, $\pi_\gamma(k) = \gamma^k\mathrm{e}^{-\gamma}/k!$.

Taking a graph uniformly at random in $\mathcal{G}_N^{(\gamma)}$ defines the measure over the quenched disorder with respect to which large deviations are evaluated. Following the basic principles of the cavity method, we consider the changes in the system when its size is increased from N to $N+1$. The first idea would be to construct uniformly at random a graph G_{N+1} in $\mathcal{G}_{N+1}^{(\gamma)}$ from a graph G_N randomly chosen in $\mathcal{G}_N^{(\gamma)}$, by connecting a new node to k nodes of G_N, with k taken with the distribution π_γ. This construction is, however, too naive, since, if the initial graph was in $\bar{\mathcal{G}}_N^{(\gamma)}$, for example, its extension is in $\bar{\mathcal{G}}_{N+1}^{(\gamma')}$, with $\gamma'/2 = (M+k)/(N+1) \neq \gamma/2$, where $M = \gamma N/2$ denotes the number of edges in G_N. However, for the three models, it appears that this construction yields a graph uniformly at random in $\mathcal{G}_{N+1}^{(\gamma')}$, with $\gamma' = \gamma + \chi(\gamma,k)/N$, where we have obtained for $\bar{\mathcal{G}}_N^{(\gamma)}$ that $\bar{\chi}(\gamma,k) = 2k - \gamma$. For $\tilde{\mathcal{G}}_N^{(\gamma)}$, after addition of the new node the probability for an edge to be present is still γ/N and not $\gamma/(N+1)$, so that it is described by γ' satisfying $\gamma'/(N+1) = \gamma/N$, yielding $\tilde{\chi}(\gamma,k) = \gamma$. Finally, for $\hat{\mathcal{G}}_N^{(\gamma)}$, after addition of the new site,

J. Stat. Mech. (2005) P07004

the distribution of the degrees $\pi_\gamma(K)$ is modified to $(1-k/N)\pi_\gamma(K)+(k/N)\pi_\gamma(K-1)$, since k of the nodes of G_N receive an additional edge; this leads to an effective distribution $\pi_{\gamma'}(K)$ with $\gamma'=\gamma+k/N$, so that $\hat\chi(k,\gamma)=k$. The passage from N to $N+1$ is thus not one-to-one between ensembles and $\mathcal{G}^{(\gamma)}_{N+1}$ is obtained from different ensembles $\mathcal{G}_N^{(\gamma-\chi(\gamma,k)/N)}$, with

$$\begin{aligned}
&\bar\chi(\gamma,k)=2k-\gamma \quad \text{(Uniform model } \bar{\mathcal{G}}^{(\gamma)}_N\text{)},\\
&\tilde\chi(\gamma,k)=\gamma \quad \text{(Binomial model } \tilde{\mathcal{G}}^{(\gamma)}_N\text{)},\\
&\hat\chi(\gamma,k)=k \quad \text{(Poissonian model } \hat{\mathcal{G}}^{(\gamma)}_N\text{)},
\end{aligned}$$

and k taken with distribution $\pi_\gamma(k)$.[2] The fact that in all cases $\langle\chi(\gamma,k)\rangle=\gamma$, with the average $\langle\cdot\rangle$ taken with respect to π_γ, reflects the equivalence of the typical properties between the three models.

When a node is added, the ground-state energy is shifted by an amount ΔE. Conditioned to the fact that the new node is connected to k other nodes, this shift has distribution $P_n^{(k)}(\Delta E)$ from graph to graph (and from node to node on a given graph). Given $P_n^{(k)}(\Delta E)$, the argument followed in equation (13) can essentially be repeated,

$$\begin{aligned}
\mathrm{e}^{(N+1)\mathcal{L}(E/(N+1),\gamma)} &\asymp \mathrm{e}^{N\mathcal{L}(\epsilon)}\mathrm{e}^{\mathcal{L}(\epsilon,\gamma)-y\epsilon}=\mathrm{e}^{N\mathcal{L}(\epsilon)}\mathrm{e}^{-\mathcal{F}(y,\gamma)}\\
&=\sum_{k=0}^{\infty}\pi_\gamma(k)\int \mathrm{d}\Delta\, E P_n^{(k)}(\Delta E)\mathrm{e}^{N\mathcal{L}[(E-\Delta E)/N,\gamma-\chi(\gamma,k)/N]}\\
&\asymp \mathrm{e}^{N\mathcal{L}(\epsilon)}\sum_{k=0}^{\infty}\pi_\gamma(k)\int \mathrm{d}\Delta\, E P_n^{(k)}(\Delta E)\mathrm{e}^{-y\Delta E}\mathrm{e}^{-z\chi(\gamma,k)},
\end{aligned}\tag{17}$$

with the notations $\epsilon\equiv E/N$, $y\equiv\partial_\epsilon\mathcal{L}(\epsilon,\gamma)$, $z\equiv\partial_\gamma\mathcal{L}(\epsilon,\gamma)$ and $\mathcal{F}(y,\gamma)=y\epsilon-\mathcal{L}(\epsilon)$. Equation (17) gives the Legendre transform of the rate function, $\mathcal{F}(y,\gamma)$, as a function of y and z. To determine z, we need consider the energy shift ΔE due to a link addition, having distribution $P_\ell(\Delta E)$. More precisely, the average value of the energy shift when $\gamma\to\gamma+1/N$ at fixed number of nodes N is required, which is obtained by adding k new edges with an appropriate distribution $\sigma(k)$. For $\bar{\mathcal{G}}^{(\gamma)}_N$, adding a single edge results in $\gamma'/2=(M+1)/N=\gamma/2+1/N$, so we take formally $\sigma=\delta_{1/2}$, where $\delta_\theta(k)=\delta_{k,\theta}$ denotes the Kronecker symbol (this non-integer prescription could be avoided as in [18] by adding two nodes at once instead of one). For $\tilde{\mathcal{G}}^{(\gamma)}_N$, we take $\sigma=\pi_{1/2}$ because it corresponds to the distribution of the number of added edges when each of the $\sim N^2/2$ edges has a probability $1/N^2$ to be present in the $\tilde{\mathcal{G}}^{(\gamma+1/N)}_N$ graph, but absent in the $\tilde{\mathcal{G}}^{(\gamma)}_N$ one. Finally, for $\hat{\mathcal{G}}^{(\gamma)}_N$, the addition of one edge leads to $\gamma'=\gamma+2/N$, so that formally $\sigma=\delta_{1/2}$ as in the uniform model. A $1/N$ expansion of $\exp[N\mathcal{L}(\epsilon,\gamma+1/N)]$ then yields $z=\partial_\gamma\mathcal{L}(\epsilon,\gamma)$ as

$$\mathrm{e}^z=\sum_{k=0}^{\infty}\sigma(k)\left(\int \mathrm{d}\Delta\, E P_\ell(\Delta E)\mathrm{e}^{-y\Delta E}\right)^k,\tag{18}$$

[2] The distinction between γ and γ' at the level of the quantities $\chi(\gamma,k)$ and $\pi_\gamma(k)$ is irrelevant as it involves higher-order effects.

J. Stat. Mech. (2005) P07004

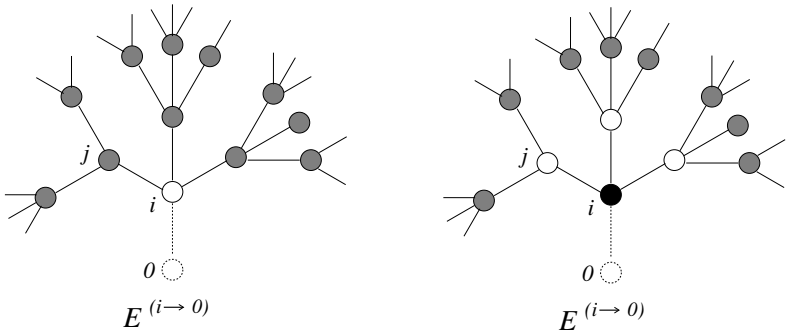

Figure 1. Rooted trees in the vertex-cover problem. This figure illustrates the recursion expressed by equation (19): on the left tree the root i (in the absence of 0) is constrained to be uncovered, while on the right tree it is constrained to be covered. The colouring in grey indicates that the nodes are neither constrained to be uncovered, nor to be covered.

with, as derived just above,

$$\bar{\sigma}(k) = \delta_{1/2}(k) \quad \text{(Uniform model } \bar{\mathcal{G}}_N^{(\gamma)}\text{)},$$
$$\tilde{\sigma}(k) = \pi_{1/2}(k) \quad \text{(Binomial model } \tilde{\mathcal{G}}_N^{(\gamma)}\text{)},$$
$$\hat{\sigma}(k) = \delta_{1/2}(k) \quad \text{(Poissonian model } \hat{\mathcal{G}}_N^{(\gamma)}\text{)}.$$

As in the typical cavity method [17, 18], the distributions $P_n^{(k)}(\Delta E)$ and $P_\ell(\Delta E)$ can be calculated by means of cavity fields. The fundamental assumption made at this stage is that the nodes to which a new node is added are independent in the absence of the added node. Under this assumption, the problem on a random graph is reduced to a problem on a tree with self-consistent boundary conditions (the so-called Bethe lattice). While the same procedure applies to other optimization problems, we restrict ourselves here for simplicity to the vertex-cover problem, for which we take the ground-state energy as the minimum of non-covered nodes under the constraint that neighbouring nodes cannot be both covered. A recursion is written for rooted trees with same degree distribution $\pi_\gamma(k)$ as the graphs (see figure 1). In general, if the degree distribution is $p(k)$, the root must be assigned the distribution $q(k) = (k+1)p(k+1)/\langle k\rangle$, which corresponds to the probability, when the edge $(i \to 0)$ is chosen at random, that i has k neighbours in addition to 0; the Poisson distribution has, however, the specificity that $q(k) = p(k) = \pi_\gamma(k)$. For the rooted tree with root-node i, let $E_0^{(i\to 0)}$ be the minimal energy with the root constrained to be non-covered (white), and $E_1^{(i\to 0)}$ the minimal energy with the root constrained to be covered (black). These quantities are related to those associated with the k rooted trees generated by deletion of the edges originating from i (see figure 1) by

$$\begin{aligned} E_0^{(i\to 0)} &= 1 + \sum_{j=1}^{k} \min(E_0^{(j\to i)}, E_1^{(j\to i)}), \\ E_1^{(i\to 0)} &= \sum_{j=1}^{k} E_0^{(j\to i)}. \end{aligned} \tag{19}$$

J. Stat. Mech. (2005) P07004

A cavity field is defined for each oriented edge as $h^{(i\to 0)} \equiv E_0^{(i\to 0)} - \min(E_0^{(i\to 0)}, E_1^{(i\to 0)})$; it satisfies the relation

$$h^{(i\to 0)} = \hat{h}^{(k)}(\{h^{(j\to i)}\}) = \max\left(0, 1 - \sum_{j=1}^{k} h^{(j\to i)}\right). \tag{20}$$

The addition of the new node i is associated with an energy shift given by

$$\begin{aligned}\Delta E_{\text{node}} &= \min(E_0^{(i\to 0)}, E_1^{(i\to 0)}) - \sum_{j=1}^{k} \min(E_0^{(j\to i)}, E_1^{(j\to i)}) \\ &= \Delta\hat{E}_n^{(k)}(\{h^{(j\to i)}\}) = \min\left(1, \sum_{j=1}^{k} h^{(j\to i)}\right).\end{aligned} \tag{21}$$

We will also need the energy shift corresponding to an edge addition, which reads

$$\begin{aligned}\Delta E_{\text{edge}} &= \min(E_0^{(1\to 2)} + E_0^{(2\to 1)}, E_0^{(1\to 2)} + E_1^{(2\to 1)}, E_1^{(1\to 2)} + E_0^{(2\to 1)}) \\ &\quad - \min(E_0^{(1\to 2)}, E_1^{(1\to 2)}) - \min(E_0^{(2\to 1)}, E_1^{(2\to 1)}) \\ &= \Delta\hat{E}_\ell(h^{(1\to 2)}, h^{(2\to 1)}) = \min(h^{(1\to 2)}, h^{(2\to 1)}).\end{aligned} \tag{22}$$

With the help of these equations, the distributions $P_n^{(k)}(\Delta E)$ and $P_\ell(\Delta E)$ are easily obtained once the distribution for the fields $P(h)$ is known.

Again similarly to the typical cavity method [17, 18], to derive the equation satisfied by $P(h)$, we introduce $R_{N+1}^{(\gamma)}(h, E)$, the probability to get an energy E and cavity field h when taking at random a graph in $\mathcal{G}_{N+1}^{(\gamma)}$ and choosing one of its nodes as a root. By definition, the marginalization over h gives $\mathbb{P}_{N+1}^{(\gamma)}(E/(N+1))$, the probability to get a graph in $\mathcal{G}_{N+1}^{(\gamma)}$ with energy E,

$$\int \mathrm{d}h\, R_{N+1}^{(\gamma)}(h, E) \equiv \mathrm{e}^{(N+1)\mathcal{L}(E/(N+1),\gamma)}. \tag{23}$$

Generalizing equation (17) slightly, we write

$$\begin{aligned}R_{N+1}^{(\gamma)}(h, E) &= \sum_{k=0}^{\infty} \pi_\gamma(k) \int \mathrm{d}\Delta\, E Q^{(k)}(h, \Delta E) \mathrm{e}^{N\mathcal{L}[(E-\Delta E)/N, \gamma - \chi(\gamma,k)/N]}, \\ &\asymp \mathrm{e}^{N\mathcal{L}(\epsilon)} \sum_{k=0}^{\infty} \pi_\gamma(k) \int \mathrm{d}\Delta\, E Q^{(k)}(h, \Delta E) \mathrm{e}^{-y\Delta E - z\chi(\gamma,k)},\end{aligned} \tag{24}$$

where $Q^{(k)}(h, \Delta E)$ denotes the joint distribution of the cavity field h and the energy shift ΔE. As in the typical cavity method, we verify that h is independent of E; more precisely,

$$\begin{aligned}&R_{N+1}^{(\gamma)}(h, E) = \mathrm{e}^{N\mathcal{L}(\epsilon)} \mathrm{e}^{-\mathcal{F}(y,\gamma)} P(h), \\ &P(h_0) = \frac{1}{Z} \sum_{k=0}^{\infty} \pi_\gamma(k) \int \prod_{i=1}^{k} \mathrm{d}h_i\, P(h_i) \delta(h_0 - \hat{h}^{(k)}(\{h_i\})) \mathrm{e}^{-y\Delta\hat{E}_n^{(k)}(\{h_i\}) - z\chi(\gamma,k)}, \\ &Z = \sum_{k=0}^{\infty} \pi_\gamma(k) \int \mathrm{d}\Delta\, E P_n^{(k)}(\Delta E) \mathrm{e}^{-y\Delta E} \mathrm{e}^{-z\chi(\gamma,k)}, \\ &\mathcal{F}(y, \gamma) = ye - \mathcal{L}(\epsilon, \gamma) = -\ln Z,\end{aligned} \tag{25}$$

J. Stat. Mech. (2005) P07004

where $P(h)$ also depends on γ and y. In the particular case where $\chi(\gamma, k)$ does not depend on k, the relation for $P(h)$ formally corresponds to what is known in the literature as a 1RSB factorized equation with 1RSB parameter y [18, 28]; this is the case with $\tilde{\mathcal{G}}_N^{(\gamma)}$ but not with $\bar{\mathcal{G}}_N^{(\gamma)}$ and $\hat{\mathcal{G}}_N^{(\gamma)}$.

Specializing now to the vertex-cover problem, the equations are simplified by the fact that $h \in \{0, 1\}$, so that the distribution $P(h)$ can be parameterized by a single real η, with $P(h) = \eta\delta(h-1) + (1-\eta)\delta(h)$. As is seen from equations (17) and (18), the distributions $P_n^{(k)}(\Delta E)$ and $P_\ell(\Delta E)$ are needed only through their Laplace transforms, which are given by

$$\begin{aligned}
\int \mathrm{d}\Delta\, E P_n^{(k)}(\Delta E)\mathrm{e}^{-y\Delta E} &= \int \prod_{i=1}^{k} \mathrm{d}h_i\, P(h_i)\mathrm{e}^{-y\Delta \hat{E}_n^{(k)}(\{h_i\})} = \mathrm{e}^{-y} + (1-\mathrm{e}^{-y})(1-\eta)^k, \\
\int \mathrm{d}\Delta\, E P_\ell(\Delta E)\mathrm{e}^{-y\Delta E} &= \int \prod_{i=1}^{2} \mathrm{d}h_i\, P(h_i)\mathrm{e}^{-y\Delta \hat{E}_\ell(h_1,h_2)} = 1 + (\mathrm{e}^{-y}-1)\eta^2.
\end{aligned} \tag{26}$$

As a first check, one verifies that, for $y = 0$, the typical RS ground-state energy density $\bar{\epsilon}^{(\gamma)}$ is reobtained [29], with the same value for the three ensembles $\tilde{\mathcal{G}}_N^{(\gamma)}$, $\bar{\mathcal{G}}_N^{(\gamma)}$ and $\hat{\mathcal{G}}_N^{(\gamma)}$,

$$\begin{aligned}
\bar{\epsilon}^{(\gamma)} &= 1 - \eta - \gamma\eta^2/2, \\
\eta &= \mathrm{e}^{-\gamma\eta}.
\end{aligned} \tag{27}$$

The equations for $y \neq 0$ can also be written explicitly. For the ensemble $\tilde{\mathcal{G}}_N^{(\gamma)}$, they read

$$\begin{aligned}
\eta &= \frac{1}{1+\mathrm{e}^{-y}(\mathrm{e}^{\gamma\eta}-1)}, \qquad z = \tfrac{1}{2}(\mathrm{e}^{-y}-1)\eta^2, \\
\mathcal{F}(y,\gamma) &= -\ln[\mathrm{e}^{-y} + (1-\mathrm{e}^{-y})\mathrm{e}^{-\gamma\eta}] + \frac{\gamma}{2}(\mathrm{e}^{-y}-1)\eta^2.
\end{aligned} \tag{28}$$

For the ensemble $\bar{\mathcal{G}}_N^{(\gamma)}$,

$$\begin{aligned}
\eta &= \frac{1}{1+\mathrm{e}^{-y}(\mathrm{e}^{\gamma\eta\mathrm{e}^{-z}}-1)}, \qquad z = \ln[1 + (\mathrm{e}^{-y}-1)\eta^2], \\
\mathcal{F}(y,\gamma) &= -\ln[\mathrm{e}^{-y} + (1-\mathrm{e}^{-y})\mathrm{e}^{-\gamma\eta\mathrm{e}^{-z}}] + \gamma(1-\mathrm{e}^{-z}) - \gamma z/2.
\end{aligned} \tag{29}$$

And for the ensemble $\hat{\mathcal{G}}_N^{(\gamma)}$,

$$\begin{aligned}
\eta &= \frac{1}{1+\mathrm{e}^{-y}(\mathrm{e}^{\gamma\eta\mathrm{e}^{-z}}-1)}, \qquad z = \tfrac{1}{2}\ln[1 + (\mathrm{e}^{-y}-1)\eta^2], \\
\mathcal{F}(y,\gamma) &= -\ln[\mathrm{e}^{-y} + (1-\mathrm{e}^{-y})\mathrm{e}^{-\gamma\eta\mathrm{e}^{-z}}] + \gamma(1-\mathrm{e}^{-z}).
\end{aligned} \tag{30}$$

The formulae (29) coincide with the result of the replica computation presented in [12]. The three corresponding rate functions are plotted in figure 2 for $\gamma = 2$.

A remarkable aspect of the vertex-cover problem is the presence, in the typical phase diagram, of a continuous phase transition at $\gamma_c = e \simeq 2.71$, from an RS phase at $\gamma < \gamma_c$ to a presumably full-RSB phase at $\gamma > \gamma_c$ [30]. Due to its continuous character, the phase transition can be located by analysing the stability of the RS ansatz. Extending the stability analysis from typical to atypical graphs thus provides, in the (γ, y) plane, a

J. Stat. Mech. (2005) P07004

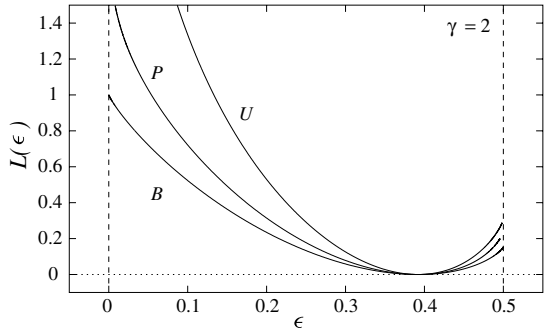

Figure 2. Rate functions for the optimal energy ϵ in the vertex-cover problem with the binomial model $\tilde{\mathcal{G}}_N^{(\gamma)}$ (label B), the Poissonian model $\hat{\mathcal{G}}_N^{(\gamma)}$ (label P) and the uniform model $\bar{\mathcal{G}}_N^{(\gamma)}$ (label U), all three for $\gamma = 2$. The common minimum at $\bar{\epsilon} \simeq 0.39$ corresponds to the prediction of the typical RS cavity method ($y = 0$) [29]. The larger curvature of the rate function for the uniform model with respect to the two other models can be interpreted by the fact that fixing the ratio of edges imposes more constraints on the graph than fixing the degree distribution or edge probability.

phase diagram with an 'AT line' [2] separating an RS phase from a full-RSB one. The three ensembles are not equivalent with respect to properties associated with atypical graphs, and we concentrate here on the binomial ensemble $\tilde{\mathcal{G}}_N^{(\gamma)}$.

RSB effects are much likely to appear first for negative values of y, corresponding to the most frustrated graphs. The failure of the RS approach can in fact be inferred from an asymptotic analysis of the $y \to -\infty$ limit: it yields $\eta \sim \mathrm{e}^{y/2}/\sqrt{\gamma}$, $\epsilon(y = -\infty) = 1/2$ and $L(y = -\infty) = (1 - \ln\gamma)/2$. Clearly, this is inconsistent as soon as $\gamma > e$, since then it predicts that $L(\epsilon = 1/2) < 0$. The value thus obtained coincides with the value of γ_c for the failure of the RS approach to typical graphs [29, 31] (the reason for this correspondence is elucidated below). The negativeness of the rate function is, however, a sufficient but not necessary sign of RSB. A more refined way to detect it consists in studying the stability of the RS large deviations ansatz. For the binomial model, it happens to be strictly equivalent to the stability analysis of a factorized 1RSB ansatz [32], and reads

$$(\gamma\eta)^2 \mathrm{e}^{-y} < 1. \tag{31}$$

It starts to be violated at $y = -\infty$ for $\gamma > 1$, while the typical graphs described with $y = 0$ are not concerned before $\gamma = e$. Indeed, for $\gamma < 1$, the RS ansatz is stable for all y: $(\gamma\eta)^2\mathrm{e}^{-y}$ is a decreasing function of y and for $y \to -\infty$ the asymptotic analysis yields $(\gamma\eta)^2\mathrm{e}^{-y} \sim \gamma$. At $\gamma = 1$, only the $y = -\infty$ point, corresponding to the maximum achievable energy $\epsilon = 1/2$, is marginally unstable. Finally, for $\gamma > 1$, there is a critical y_c such that RS is stable for $y > y_c$ but unstable for $y < y_c$; y_c increases when γ increases and reaches $y_c = 0$ for $\gamma_c = e$, the point where the typical problem undergoes the RSB transition. For $\gamma > \gamma_c$, while typical graphs are FRSB, some less frustrated graphs are still RS. The resulting phase diagram is shown in figure 3 in the plane (γ, y) and in figure 4 in the plane (γ, ϵ). The occurrence of RSB at $\gamma = 1$ is particularly interesting because this point corresponds to the percolation threshold of a giant connected component [4],

J. Stat. Mech. (2005) P07004

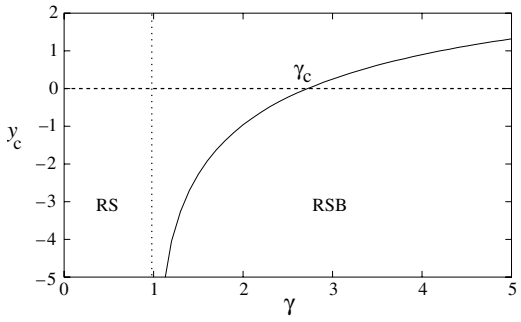

Figure 3. Phase diagram of the vertex-cover problem in the (γ, y) plane for the Erdős–Rényi ensemble $\tilde{\mathcal{G}}_N^{(\gamma)}$. The full line corresponds to the line $y_c(\gamma)$ where the RS solution becomes instable; there is no instability below the percolation threshold, $\gamma < 1$ [$y_c(\gamma) \to -\infty$ for $\gamma \to 1^+$]. The line $y = 0$ reproduces the phase diagram of the typical case with $\gamma_c = e \simeq 2.71$ being defined as the intersection of $y_c(\gamma)$ with $y = 0$.

which appeared totally irrelevant when restricting to typical graphs [29]. In contrast, when atypical graphs are included in the picture, the emergence of a giant component seems to be responsible for the onset of RSB, as shown in figure 3 (a similar analysis of the uniform model, however, reveals that in this case RSB appears only above an average connectivity $\gamma = 2$).

The opposite $y \to +\infty$ limit is also interesting since it is always correctly described by the RS ansatz, with $\epsilon(y = \infty, \gamma) = 0$ and $\mathcal{L}(y = \infty, \gamma) = -\gamma/2$. It can be interpreted as corresponding to graphs with no edge at all, which occurs with probability

$$\mathbb{P}_N^{(\gamma)}(\text{non frustrated}) \asymp \mathbb{P}_N^{(\gamma)}(\text{no edge}) \asymp \left(1 - \frac{\gamma}{N}\right)^{N^2/2} \asymp \mathrm{e}^{-N\gamma/2}. \tag{32}$$

Similar relations between the $y = \infty$ limit and the probability for non-frustrated samples have been reported in a variety of other models [11], providing consistent checks of the method.

3.3. Random graphs with given degree distributions

The LDCM applies as well to graph ensembles with non-Poissonian degrees, and an example is provided here with graph ensembles specified by their degree distribution $p(k)$, that is, with each node having an independent probability $p(k)$ of being of degree k. The reasoning for arbitrary $p(k)$ basically follows the procedure used for the Poissonian model $\hat{\mathcal{G}}_N^{(\gamma)}$ which was a particularly case where $p(k) = \gamma^k \mathrm{e}^{-\gamma}/k!$.

We thus first consider how the ensemble is modified when a new node is connected to k nodes of a graph made of N nodes having degree distribution $p(K)$. The degree distribution becomes $p'(K)$, with

$$p'(K) = \left(1 - \frac{k}{N}\right) p(K) + \frac{k}{N} p(K-1), \tag{33}$$

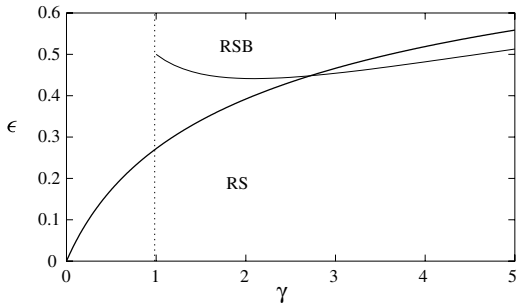

Figure 4. Phase diagram of the vertex-cover problem in the (γ, ϵ) plane for the Erdős–Rényi ensemble $\tilde{\mathcal{G}}_N^{(\gamma)}$. The full line represents the typical RS energy density, as given by equation (27), and the thin line, starting only from $\gamma = 1$, the energy density above which the RS approach fails, as given by equation (31). The two curves cross at $\gamma_c = e$, corresponding to the onset of RSB on typical graphs.

since a given node has a probability k/N to get its degree increased by one unit (the probability that it is increased by more than one unit is $\mathrm{O}(1/N^2)$ and is therefore neglected). Writing $p(K) = \sum_r p_r \delta_{K,r}$, we explicitly have $p'_r = p_r + (k/N)\delta p_r$ with $\delta p_r = p_{r-1} - p_r$. The set $\{p_r\}$ can serve as a characterization of the graph ensemble, and following the same scheme as for Poissonian graphs, we obtain

$$\mathrm{e}^{-\mathcal{F}(y,\{p_r\})} = \sum_{k=0}^{\infty} p_k \int \mathrm{d}\Delta\, E P_n^{(k)}(\Delta E) \mathrm{e}^{-y\Delta E} \mathrm{e}^{-kz}, \tag{34}$$

with as before $y = \partial_\epsilon \mathcal{L}(\epsilon, \{p_r\})$ and now

$$z \equiv \sum_r \delta p_r \frac{\partial \mathcal{L}(\epsilon, \{p_r\})}{\partial p_r}. \tag{35}$$

To get z, we notice that equation (33) with $k = 2$ describes the effect of an edge addition so that

$$\mathrm{e}^{2z} = \int \mathrm{d}\Delta\, E P_\ell(\Delta E) \mathrm{e}^{-y\Delta E}. \tag{36}$$

The potential whose Legendre transform yields the rate function can therefore be written:

$$\mathcal{F}(y, \{p_r\}) = -\ln \left[\sum_{k=0}^{\infty} p_k \int \mathrm{d}\Delta\, E P_n^{(k)}(\Delta E) \mathrm{e}^{-y\Delta E} \left(\int \mathrm{d}\Delta\, E' P_\ell(\Delta E') \mathrm{e}^{-y\Delta E'} \right)^{-k/2} \right]. \tag{37}$$

Similarly, the cavity equation for the fields reads

$$P(h_0) = \frac{1}{Z} \sum_{k=0}^{\infty} \frac{(k+1)p_{k+1}}{\langle k \rangle} \int \prod_{j=1}^{k} \mathrm{d}h_j\, P(h_j) \delta(h_0 - \hat{h}^{(k)}(\{h_j\})) \mathrm{e}^{-y\Delta \hat{E}_n^{(k)}(\{h_j\})} \mathrm{e}^{-kz} \tag{38}$$

where Z is the appropriate normalization and the presence of $(k+1)p_{k+1}/\langle k \rangle$ instead of p_k reflects the fact that an oriented edge is chosen at random, and not a node as in

J. Stat. Mech. (2005) P07004

equation (37) (only for Poissonian graphs do these two probabilities happen to be the same).

A case of particular interest is when $p(k) = \delta_{k,r}$, corresponding to random r-regular graphs. In such a case, the factor e^{-kz} in equation (38) is a constant which can be absorbed into the normalization $Z' = Z\mathrm{e}^{(r-1)z}$,

$$P(h_0) = \frac{1}{Z'} \int \prod_{j=1}^{r-1} \mathrm{d}h_j \, P(h_j) \delta(h_0 - \hat{h}^{(r-1)}(\{h_j\})) \mathrm{e}^{-y\Delta \hat{E}_n^{(r-1)}(\{h_j\})}. \tag{39}$$

This is formally identical to what is known as a 1RSB factorized cavity equation [18, 28]. The correspondence extends to the formula for the potential, equation (37), which becomes, for regular graphs,

$$\mathcal{F}(y) = -\ln \left[\int \mathrm{d}\Delta \, E P_n^{(r)}(\Delta E) \mathrm{e}^{-y\Delta E} \right] + \frac{r}{2} \ln \left[\int \mathrm{d}\Delta \, E P_\ell(\Delta E) \mathrm{e}^{-y\Delta E} \right]. \tag{40}$$

As a consequence, the 1RSB complexity of models defined on random regular graphs coincides with minus a rate function, as already noticed in [11]. Obviously, the correspondence holds only within the RS approximation that has been assumed throughout since, by nature of a 1RSB glassy phase, the complexity is positive while a rate function is necessarily non-negative; it will be shown below how the formalism needs to be extended to include the possibility of RSB. When the factorization does not hold, the correspondence between rate functions over the external disorder and negative complexities is only approximate; we have seen, however, that for vertex-cover on $\tilde{\mathcal{G}}_N^{(\gamma)}$ the rate function starts getting negative values precisely at the point $\gamma = e$ where typical graphs undergo an RSB transition, in agreement with the observation that the two quantities are approximatively related.

4. Multi-step large deviations

The LDCM can naturally be extended beyond the simple case of zero-temperature systems in an RS phase with disorder only specified by a random graph ensemble. We consider here successively finite-temperature systems, models with RSB phases, and external disorders including random couplings. In the three cases, a second temperature is needed to describe the large deviations with respect to the additional source of randomness. In each case also, the equations display a common 2RSB-like structure [2], which would be promoted to the nRSB type with $n > 2$ if n different sources of disorder were present.

4.1. Finite temperature

The simplest extension requiring multi-step large deviations consists in generalizing the description of a model on a given graph from zero to finite temperature. Two inverse temperatures are now required: β, for the thermodynamics on a given graph, and y for the large deviations in the graph ensemble. More precisely, large deviations now concern the density of free energy $f(\beta)$, with the limit $\beta \to \infty$ giving back to the large deviations for the ground-state density energy $\epsilon = \lim_{\beta\to\infty} f(\beta)$, as discussed so far. For any fixed value of β, the rate function $L(f, \beta) \equiv -\mathcal{L}(f, \beta)$ is calculated as before through the

J. Stat. Mech. (2005) P07004

Legendre transform of a potential $\mathcal{F}(y,\beta)$ satisfying

$$\begin{aligned} \mathrm{e}^{-N\mathcal{F}(y,\beta)} &= \int \mathrm{d}f\, \mathrm{e}^{N[\mathcal{L}(f,\beta)-yf]} = \frac{1}{\#\mathcal{G}} \sum_{G\in\mathcal{G}} Z_G(\beta)^{y/\beta}, \\ Z_G(\beta) &\equiv \mathrm{e}^{-\beta N f_G(\beta)} \equiv \sum_{C\in\mathcal{C}_G} \mathrm{e}^{-\beta E(C)}, \end{aligned} \tag{41}$$

where we introduced $Z_G(\beta)$, the partition function on the graph G at temperature β, $\mathcal{C}_G$, the set of configurations on the graph G, and $\#\mathcal{G}$, the cardinality of the ensemble of graphs $\mathcal{G}$. Note that the particular choice $y=\beta$ corresponds to the uniform measure over all configurations $\{C\in\mathcal{C}_G\}_{G\in\mathcal{G}}$:

$$\sum_{G\in\mathcal{G}}\sum_{C\in\mathcal{C}_G} \mathrm{e}^{-\beta E(C)} = \sum_{G\in\mathcal{G}} \mathrm{e}^{-\beta N f_G(\beta)} = (\#\mathcal{G})\mathrm{e}^{-\beta N\mathcal{F}(\beta,\beta)}. \tag{42}$$

From the technical point of view, we just have to replace in all formulae the functions $\hat{h}^{(k)}(\{h_i\})$, $\Delta\hat{E}_n^{(k)}(\{h_i\})$ and $\Delta\hat{E}_\ell(h_1,h_2)$ by their finite-temperature extensions $\hat{h}^{(k)}(\{h_i\};\beta)$, $\Delta\hat{F}_n^{(k)}(\{h_i\};\beta)$ and $\Delta\hat{F}_\ell(h_1,h_2;\beta)$. Taking the vertex-cover problem as an example, these quantities are derived by writing recursive equations for the conditional partition functions $Z_0^{(i\to 0)}(\beta)$ and $Z_1^{(i\to 0)}(\beta)$ instead of the conditional ground-state energies $E_0^{(i\to 0)}$ and $E_1^{(i\to 0)}$. More precisely, equations (19) are replaced with

$$\begin{aligned} Z_0^{(i\to 0)}(\beta) &= \mathrm{e}^{-\beta}\prod_{j=1}^{k}\left(Z_0^{(j\to i)}(\beta)+Z_1^{(j\to i)}(\beta)\right), \\ Z_1^{(i\to 0)}(\beta) &= \prod_{j=1}^{k} Z_0^{(j\to i)}(\beta). \end{aligned} \tag{43}$$

To get $\lim_{\beta\to\infty} h^{(i\to 0)}(\beta) = h^{(i\to 0)}$ with $h^{(i\to 0)}$ defined in equation (20), the cavity fields at finite temperature $h^{(i\to 0)}(\beta)$ are defined as

$$h^{(i\to 0)}(\beta) \equiv -\frac{1}{\beta}\ln\left(\frac{Z_0^{(i\to 0)}(\beta)}{Z_0^{(i\to 0)}(\beta)+Z_1^{(i\to 0)}(\beta)}\right). \tag{44}$$

With these definitions, the different functions required to compute the rate function $L(f,\beta)$ are

$$\begin{aligned} \hat{h}^{(k)}(\{h_j\},\beta) &= \frac{1}{\beta}\ln\left(1+\exp\left(\beta\left(1-\sum_{j=1}^{k}h_j\right)\right)\right), \\ \Delta\hat{F}_n^{(k)}(\{h_j\};\beta) &= -\frac{1}{\beta}\ln\left(\mathrm{e}^{-\beta}+\exp\left(-\beta\sum_{j=1}^{k}h_j\right)\right), \\ \Delta\hat{F}_\ell(h_1,h_2;\beta) &= -\frac{1}{\beta}\ln\left(\mathrm{e}^{-\beta h_1}+\mathrm{e}^{-\beta h_2}-\mathrm{e}^{-\beta(h_1+h_2)}\right), \end{aligned} \tag{45}$$

which all reduce as they should to $\hat{h}^{(k)}(\{h_j\})$, $\Delta\hat{E}_n^{(k)}(\{h_j\})$ and $\Delta\hat{E}_\ell(h_1,h_2)$ given in equation (20)–(22) when $\beta\to\infty$. The only practical difference with the $\beta=\infty$ case is that the distribution $P(h)$ has no more reason to be peaked on integers and therefore cannot be parameterized by a single real number.

J. Stat. Mech. (2005) P07004

4.2. Replica symmetry breaking

The phase space of a glassy 1RSB instance is structured into clusters whose energy is controlled by a parameter μ in exactly the same way that the finite inverse temperature β controls the equilibrium configurations according to their energy. Therefore, the extension of the LDCM from paramagnetic (RS) systems to glassy (1RSB) systems is formally similar to the extension from zero temperature to finite temperature. The counterpart of equation (41) reads

$$\begin{aligned} \mathrm{e}^{-N\mathcal{F}(y,\mu)} &= \frac{1}{\#\mathcal{G}} \sum_{G\in\mathcal{G}} \mathrm{e}^{-yN\phi_G(\mu)} = \int \mathrm{d}\phi\, \mathrm{e}^{N[\mathcal{L}(\phi,\mu)-y\phi]}, \\ \mathrm{e}^{-N\mu\phi_G(\mu)} &= \sum_{\alpha\in G} \mathrm{e}^{-\mu N\epsilon_\alpha} = \int \mathrm{d}\epsilon\, \mathrm{e}^{N[\Sigma_G(\epsilon)-\mu\epsilon]} \end{aligned} \tag{46}$$

where $\phi_G(\mu)$ is the 1RSB potential on graph G and $L(\phi,\mu) \equiv -\mathcal{L}(\phi,\mu)$ is the rate function describing the large deviations of $\phi(\mu)$ over the ensemble of random graphs; in these formulae, we reserve greek letters for quantities related to the internal structure and use α to index the clusters. The saddle points in equation (46) lead to the following Legendre transform relations:

$$\mathcal{F}(y,\mu) = y\phi - \mathcal{L}(\phi,\mu), \qquad y = \partial_\phi \mathcal{L}(\phi,\mu), \tag{47}$$

$$\mu\phi(\mu) = \mu\epsilon - \Sigma(\epsilon), \qquad \mu = \partial_\epsilon \Sigma(\epsilon). \tag{48}$$

These quantities are computed by applying the standard 1RSB cavity method [17] to a given set of atypical graphs characterized by their ground-state energy density ϵ_0. If $\rho_N(\epsilon|\epsilon_0)$ denotes the distribution of their clusters, the corresponding complexity is defined as $\rho_N(\epsilon|\epsilon_0) \asymp \exp[N\Sigma(\epsilon|\epsilon_0)]$; for an energy ϵ close to the ground-state reference energy ϵ_0, it becomes

$$\rho_N(\epsilon|\epsilon_0) \asymp \mathrm{e}^{\mu N(\epsilon-\epsilon_0)} \tag{49}$$

where $\mu \equiv \partial_\epsilon \Sigma(\epsilon = \epsilon_0|\epsilon_0)$ defines the 1RSB internal inverse temperature. The shift in energy ΔE induced by a node addition, which is needed in the recursion at the level of the graph average, is given by the shift in the reference energy, that is,

$$\rho_{N+1}(\epsilon|\epsilon_0) = \rho_N(\epsilon|\epsilon_0)\mathrm{e}^{-\mu\Delta E}. \tag{50}$$

The expression for the reweighting term $\Xi \equiv \mathrm{e}^{-\mu\Delta E}$ is read from the 1RSB cavity recursion which involves $\Pi(h)$, the distribution of cavity fields over the clusters, and $P[\Pi]$, the distribution of the Πs over the graphs; for a given (class of) graph, the connection of a new node to k other ones indeed yields

$\Pi_0 = \hat{\Pi}^{(k)}[\{\Pi_i\}]$, with

$$\begin{aligned} \hat{\Pi}^{(k)}[\{\Pi_i\}](h_0) &= \frac{1}{\Xi} \int \prod_{i=1}^{k} \Pi_i(h_i)\delta(h_0 - \hat{h}^{(k)}(\{h_i\}))\mathrm{e}^{-\mu\Delta\hat{E}_n^{(k)}(\{h_i\})}, \\ \Xi = \mathrm{e}^{-\mu\Delta E} &= \hat{\Xi}^{(k)}[\{\Pi_i\}] = \int \prod_{i=1}^{k} \Pi_i(h_i)\mathrm{e}^{-\mu\Delta\hat{E}_n^{(k)}(\{h_i\})}. \end{aligned} \tag{51}$$

J. Stat. Mech. (2005) P07004

Therefore, at the level of the graph average, we have for Poissonian graphs

$$P[\Pi_0] = \frac{1}{Z}\sum_{k=0}^{\infty}\pi_\gamma(k)\int\prod_{i=1}^{k}\mathcal{D}\Pi_i P[\Pi_i]\delta[\Pi_0 - \hat{\Pi}^{(k)}[\{\Pi_i\}]\hat{\Xi}^{(k)}(\{\Pi_i\})]^{y/\mu}\mathrm{e}^{-z\chi(k,\gamma)}, \tag{52}$$

and for graphs with fixed degree distribution

$$P[\Pi_0] = \frac{1}{Z'}\sum_{k=0}^{\infty}\frac{(k+1)p_{k+1}}{\langle k\rangle}\int\prod_{i=1}^{k}\mathcal{D}\Pi_i P[\Pi_i]\ \delta[\Pi_0 - \hat{\Pi}^{(k)}(\{\Pi_i\})]\hat{\Xi}^{(k)}[\{\Pi_i\}]^{y/\mu}\mathrm{e}^{-kz}. \tag{53}$$

The 1RSB large deviations equations thus have the structure of a typical factorized 2RSB theory, just as the RS large deviations equations resembled a typical factorized 1RSB theory. In particular, for $y = 0$, the non-factorized 1RSB formalism is exactly recovered.

Replica symmetry breaking (RSB) is relevant to many optimization problems, and the vertex-cover problem has already provided us with such an example. For this problem, studying the local stability of the RS ansatz was enough to locate the continuous transition to a RSB phase. However, other problems may display a different kind of glass transition, known as a discontinuous 1RSB transition, which, due to its discontinuous character, can only be correctly described by implementing a 1RSB formalism. Such a transition is found for instance in the colouring problem [33], which we take here as an illustrative example of the broader class of constraint satisfaction problems.

In the statistical physics point of view, a problem is satisfiable (SAT) if it has a zero ground-state energy density, $\epsilon = 0$.[3] In the presence of a clustered glassy phase, however, an alternative characterization is provided by the sign of the complexity Σ_0, giving, when positive, the number $\exp[N\Sigma_0]$ of clusters with $\epsilon = 0$. This complexity Σ_0 is obtained in the 1RSB formalism by taking the limit $\mu \to \infty$; for the three-colouring problem on Erdős–Rényi graphs $\tilde{\mathcal{G}}_N^{(\gamma)}$, Σ_0 is found to be positive only in a restricted interval $[\gamma_d, \gamma_c]$, with $\gamma_d = 4.42$ and $\gamma_c = 4.69$ [33], as schematically represented in figure 6. The threshold values γ_d and γ_c, which also appear in other constraint satisfaction problems such as K-SAT [34], locate two phase transitions, called respectively the clustering and SAT-UNSAT transitions: with probability one when $N \to \infty$, a graph with $\gamma < \gamma_d$ is colourable and RS, a graph with $\gamma_d < \gamma < \gamma_c$ is again colourable but RSB, and a graph with $\gamma > \gamma_c$ is uncolourable.

As an illustration, we consider here the three-colouring problem on the Erdős–Rényi ensemble $\tilde{\mathcal{G}}_N^{(\gamma)}$. Following [35], equation (53), the shift $\Delta\phi_\ell$ in the 1RSB potential due to a link addition is given by

$$\mathrm{e}^{-\mu\Delta\phi_\ell} = 1 + (\mathrm{e}^{-\mu} - 1)q\eta_1\eta_2, \tag{54}$$

where the η_js, with distribution $\rho(\eta)$, represent the 1RSB cavity fields for this problem [35]. In the LDCM, we need the Laplace transform of the distribution $P_\ell(\Delta\phi)$, which thus reads

$$\int \mathrm{d}\Delta\,\phi P_\ell(\Delta\phi)\mathrm{e}^{-y\Delta\phi} = \int\prod_{i=1,2}\mathrm{d}\eta_i\,\rho(\eta_i)\left(1 + (\mathrm{e}^{-\mu} - 1)q\eta_1\eta_2\right)^{y/\mu}. \tag{55}$$

[3] Strictly speaking, a problem is SAT if and only if $E = 0$, and it may be that $E > 0$ while $\epsilon = 0$ in the thermodynamical limit, as observed for the two-colouring problem [43]. Here, however, we take the viewpoint of defining a SAT graph by $\epsilon = 0$: this condition in fact turns out to be equivalent to $E = 0$ for the three-colouring problem.

J. Stat. Mech. (2005) P07004

If only SAT configurations are to be addressed, the general 1RSB-LDCM equations can be simplified by taking the limit $\mu \to \infty$. This limit enforces $\epsilon \to 0$ and $\mu\phi(\mu) \to -\Sigma_0$ and requires us to rescale y by taking $y \to \infty$ with $x = y/\mu$ fixed, such that $\mathcal{F}(y,\mu) \to \mathcal{F}(x)$ with

$$\mathcal{F}(x) = x\Sigma_0 - \mathcal{L}(\Sigma_0), \qquad x = \partial_{\Sigma_0}\mathcal{L}(\Sigma_0), \tag{56}$$

where $\mathcal{L}(\Sigma_0) = \lim_{\mu\to\infty} \mathcal{L}(\phi = -\Sigma_0/\mu, \mu)$. In this limit,

$$\int \mathrm{d}\Delta\, \phi P_\ell(\Delta\phi)\mathrm{e}^{-y\Delta\phi} \to \int \prod_{i=1,2} \mathrm{d}\eta_i\, \rho(\eta_i)\,(1 - q\eta_1\eta_2)^x\,. \tag{57}$$

Similarly for site addition, we have, again in the limit $\mu \to \infty$,

$$\int \mathrm{d}\Delta\, \phi P_n^{(k)}(\Delta\phi)\mathrm{e}^{-y\Delta\phi} \to \int \prod_{i=1}^{k} \mathrm{d}\eta_i\, \rho(\eta_i)\hat{\Xi}^{(k)}(\eta_1,\ldots,\eta_k)^x, \tag{58}$$

with

$$\Xi^{(k)} \equiv \lim_{\mu\to\infty} \mathrm{e}^{-\mu\Delta\phi_n^{(k)}} = \sum_{\ell=0}^{q-1}(-1)^\ell \binom{q}{\ell+1} \prod_{i=1}^{k}(1 - (\ell+1)\eta_i), \tag{59}$$

where $\Delta\phi_n^{(k)}$ refers to the shift in potential due to the connection of a new nodes to k old ones (see equation (56) in [35]).

The distribution $\rho(\eta)$ is determined, in the limit $\mu \to \infty$, by the self-consistent equation

$$\rho(\eta_0) = \frac{1}{Z}\sum_{k=0}^{\infty}\pi_\gamma(k)\int \prod_{i=1}^{k}\mathrm{d}\eta_i\, \rho(\eta_i)\delta(\eta_0 - \hat{\eta}^{(k)}(\{\eta_i\}))\hat{\Xi}^{(k)}(\{\eta_i\})^x\mathrm{e}^{-z\chi(k,\gamma)}, \tag{60}$$

with

$$\hat{\eta}^{(k)}(\{\eta_i\})) \equiv \frac{1}{\hat{\Xi}^{(k)}(\{\eta_i\})}\sum_{\ell=0}^{q-1}(-1)^\ell \binom{q-1}{\ell} \prod_{i=1}^{k}(1 - (\ell+1)\eta_i). \tag{61}$$

This equation can be solved numerically using a population dynamics algorithm [17]: as for the typical case [35], recovered here by taking $x = 0$, a peak at $\eta = 0$ is observed, so that $\rho(\eta)$ can be written $\rho(\eta) = t\delta(\eta) + (1-t)\tilde{\rho}(\eta)$ where $\tilde{\rho}$ represents a continuous part. Rate functions $L(\Sigma_0)$ obtained with this procedure are presented in figure 5.

Interestingly, for any value of γ, clustering and SAT-UNSAT transitions are found to occur within atypical graphs. These phase transitions are found by monitoring the parameter x, which, roughly speaking, characterizes the degree of frustration, with larger x corresponding to less frustrated graphs. For a given γ, we indeed find thresholds $x_c(\gamma)$ and $x_d(\gamma)$, with $x_c(\gamma) < x_d(\gamma)$, such that for $x < x_c(\gamma)$ the graphs are UNSAT ($\Sigma_0 < 0$) while for $x > x_d(\gamma)$ no further clustered solution is found ($\Sigma_0 = 0$). The global phase diagram in the (γ, x) plane is presented in figure 7. The typical phase diagram of figure 6 can be read on the line $x = 0$, with the thresholds γ_c and γ_d determined respectively by

J. Stat. Mech. (2005) P07004

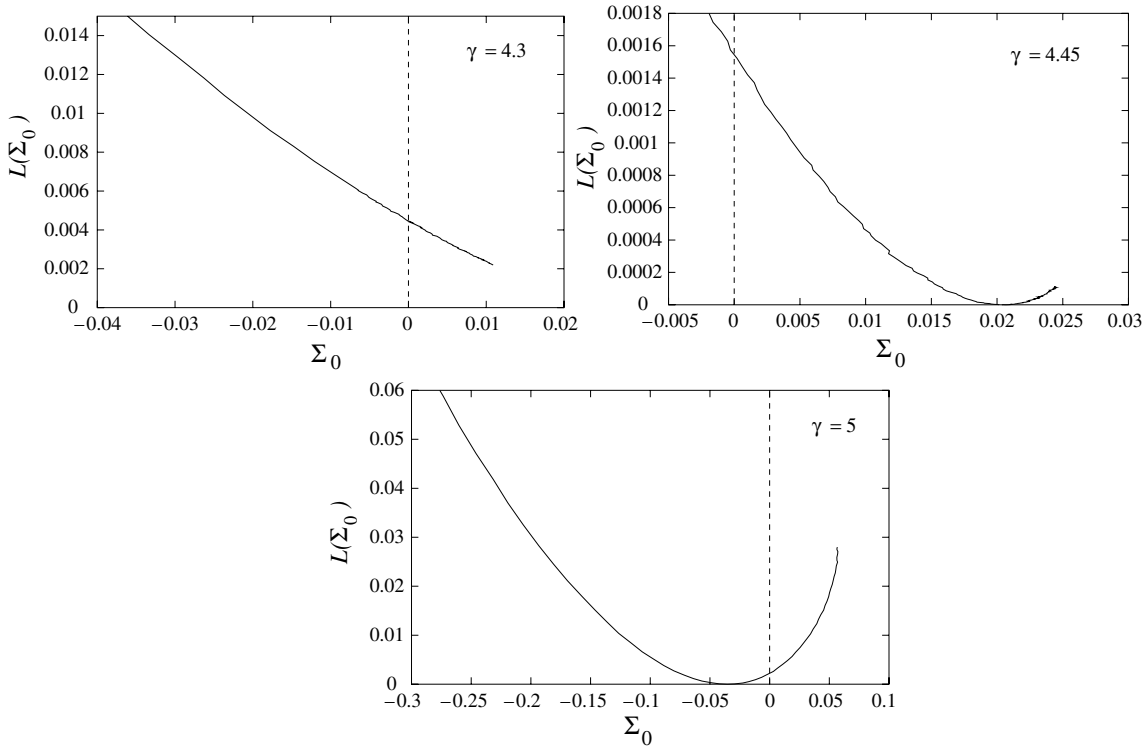

Figure 5. Rate functions $L(\Sigma_0)$ of the complexity Σ_0 in the three-colouring problem for the ensemble $\tilde{\mathcal{G}}_N^{(\gamma)}$ with $\gamma = 4.3 < \gamma_d$, $\gamma = 4.45 \in [\gamma_d, \gamma_c]$ and $\gamma = 5 > \gamma_c$, as obtained from the 1RSB-LDCM. A maximal value x_d of the slope $x = -\partial_{\Sigma_0} L(\Sigma_0)$, associated with a maximal value of Σ_0, is found above which a non-trivial $L(\Sigma_0)$ ceases to exist with, by definition of γ_d, $x_d(\gamma) < 0$ when $\gamma < \gamma_d$, and $x_d(\gamma) > 0$ when $\gamma > \gamma_d$. Note that for $\gamma = 4.3$, the curve displayed certainly does not describe the actual rate function which is expected to vanish at $\Sigma_0 = 0$: since the non-trivial solution shown coexists with the trivial solution reduced to the point ($\Sigma_0 = 0$, $L = 0$), the correct rate function may be obtained by a Maxwell construction, i.e., by drawing the supporting line originating from ($\Sigma_0 = 0$, $L = 0$) and tangent to the curve.

$x_c(\gamma_c) = 0$ and $x_d(\gamma_d) = 0$. We also expect that, for some values of x, the 1RSB ansatz does not hold, in analogy to what is found in the typical case [36]; however, we do not discuss this issue here; it could be handled by extending the stability analysis performed for typical instances [32].

In cases such as colouring on regular graphs where the cavity equations are factorized, an interpretation of $\mathcal{F}(y, \mu)$ can be given in terms of the probability $\mathbb{P}_N(\epsilon = 0)$ for a graph to be SAT while lying in a class of typically UNSAT graphs (or conversely for a graph in a typically SAT ensemble to be UNSAT). For an RS system first, it has been shown in section 3.3 that the rate function $L(\epsilon)$ is given by the negative 1RSB complexity $\Sigma(\epsilon)$, $L(\epsilon) = -\Sigma(\epsilon)$; in particular, the probability to be SAT for a graph in the typically UNSAT phase is $\mathbb{P}_N(\epsilon = 0) \asymp \mathrm{e}^{N\Sigma_0}$, where as before $\Sigma_0 \equiv \Sigma(\epsilon = 0)$. In the typically UNSAT

J. Stat. Mech. (2005) P07004

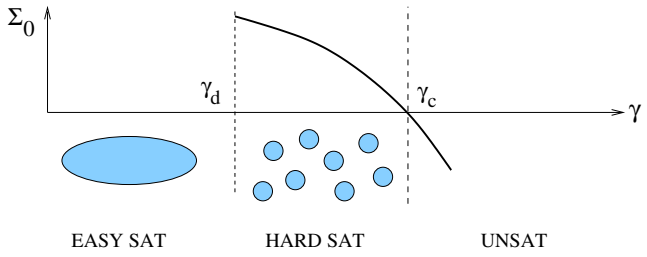

Figure 6. Pictorial representation of the phase diagram for typical instances of the three-colouring problem on Erdős–Rényi graphs $\tilde{\mathcal{G}}_N^{(\gamma)}$. For $\gamma < \gamma_d$, the set of solutions forms a single connected cluster. For $\gamma_d < \gamma < \gamma_c$, the set of solutions is organized into $\exp(N\Sigma_0)$ clusters, where Σ_0 is the complexity curve represented in the upper part. For $\gamma > \gamma_c$, $\Sigma_0 < 0$, indicating that there is (typically) no solution, i.e., the system is UNSAT.

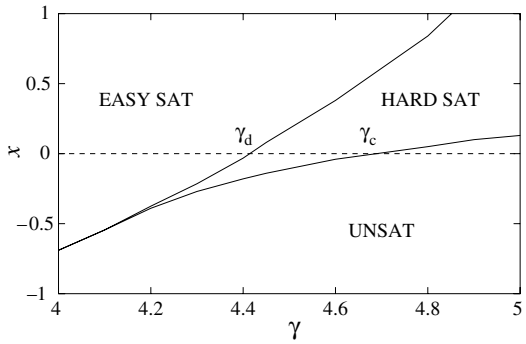

Figure 7. Phase diagram of the three-colouring problem on Erdős–Rényi graphs in the (γ, x) plane. The top curve gives $x_d(\gamma)$, the value of x at which a 1RSB solution appears, and the bottom curve gives $x_c(\gamma)$, the value of x at which the complexity vanishes. The intersection of these two curves with the line $x = 0$ gives back the typical phase diagram of figure 6 with the intersection points γ_d and γ_c. (Note that this diagram includes values from non-physical branches as discussed in figure 5 for $\gamma = 4.3$.)

phase of a 1RSB system one still has $\Sigma_0 < 0$, but the fluctuations of the complexity Σ_0 from graph to graph described by the rate function $L(\Sigma_0)$ have to be taken into account. The relation with $\mathbb{P}_N(\epsilon = 0)$ thus becomes

$$\mathbb{P}_N(\epsilon = 0) \asymp \int_{-\infty}^{0} \mathrm{d}\Sigma_0 \, \mathrm{e}^{-NL(\Sigma_0)} \mathrm{e}^{N\Sigma_0} \tag{62}$$

since $\mathrm{e}^{N\Sigma_0}$ must now be multiplied by the probability $\mathrm{e}^{-NL(\Sigma_0)}$ to actually get a complexity Σ_0. Two cases can then arise: the saddle point can lie at the boundary $\Sigma_0 = 0$, in which case $P_N(\epsilon = 0) \asymp \mathrm{e}^{NL(\Sigma_0=0)}$, or it can be strictly negative, $\Sigma_0 < 0$, in which case $P_N(\epsilon = 0) \asymp \mathrm{e}^{-N[L(\Sigma_0)-\Sigma_0]}$, with Σ_0 given by $1 = \partial_{\Sigma_0} L(\Sigma_0)$. An alternative formulation

J. Stat. Mech. (2005) P07004

can be given with $\mathcal{F}(x)$ related to $L(\Sigma_0)$ through

$$\mathrm{e}^{-N\mathcal{F}(x)} = \int \mathrm{d}\Sigma_0 \, \mathrm{e}^{-N[L(\Sigma_0)-x\Sigma_0]}. \tag{63}$$

Since $\Sigma_0 = \partial_x \mathcal{F}(x)$, one has to compute the x^* maximizing $\mathcal{F}(x)$, which is also associated to the saddle point $x^* = \partial_{\Sigma_0} L(\Sigma_0 = 0)$: $x^* < 1$ corresponds to the first case with $\mathbb{P}_N(\epsilon = 0) \asymp \mathrm{e}^{-N\mathcal{F}(x^*)}$, while $x^* > 1$ corresponds to the second case with $\mathbb{P}_N(\epsilon = 0) \asymp \mathrm{e}^{-N\mathcal{F}(1)}$ (the very same mechanism underlies the selection of the 1RSB parameter in the typical cavity method at finite temperature [17]). Physically, this second situation, $x^* > 1$, refers to very rare graphs with extremely small frustration, which are thus in an RS phase (it can be seen that $\mathcal{F}(x = 1)$ indeed gives back the RS rate function $L(\epsilon = 0)$). In the phase diagram of a problem like colouring, such a situation is expected only for the largest values of γ.

As such, the interpretation applies only for models where the factorization holds. Otherwise negative complexities $\Sigma(\epsilon) < 0$ should be interpreted as giving the probability $\mathrm{e}^{N\Sigma(\epsilon)}$ for a model with typical complexity to have a cluster with energy ϵ, which, because of the interference between the internal disorder and the local external disorder, is not associated with a rate function relative to the external disorder only. The estimation of $\mathbb{P}_N(\epsilon = 0)$ along the lines presented above is, however, still amenable, but one has to perform a two-step large deviations analysis involving a second rate function $L(L_0)$ over the first rate function L_0 estimated under the RS assumption; technically, this computation is quite similar to what has been done here for $L(\Sigma_0)$, which can be taken as a factorized approximation of $L(L_0)$.

4.3. Multiple sources of disorder

In problems other that vertex-cover or colouring, the definition of an instance can include some quenched values of the coupling constants, as happens for spin-glass models or K-SAT. In this case, the energy shifts include a dependence on the couplings J, with functions $\Delta\hat{E}_n^{(k;\{J_i\})}(\{h_i\})$ and $\Delta\hat{E}_\ell^{(J_{12})}(h_1, h_2)$. The situation is described by a two-step large deviations principle,

$$\mathbb{P}_G[E_{G,J} = N\epsilon] \asymp \mathrm{e}^{N\mathcal{L}_G(\epsilon)}, \qquad \mathbb{P}[\mathcal{L}_G] \asymp \mathrm{e}^{N\mathcal{K}(\mathcal{L}_G)}. \tag{64}$$

The rate functions are again computed through their Legendre transforms by considering two temperatures, y_J and y_G,

$$\mathcal{F}(y_J, y_G) = y_G f - \mathcal{K}(f, y_J), \qquad y_G = \partial_f \mathcal{K}(f, y_J), \tag{65}$$

$$y_J f(y_J) = y_J \epsilon - \mathcal{L}(\epsilon), \qquad y_J = \partial_\epsilon \mathcal{L}(\epsilon) \tag{66}$$

where the factor y_J in front of f in the second line is introduced as in equation (47) to conform with the traditional definition of free energies. Taking Poissonian graphs as an example, we have explicitly the following expressions, to be compared with equations (51)

J. Stat. Mech. (2005) P07004

and (52):

$$
\begin{aligned}
&\mathcal{P}[P_0] = \frac{1}{\mathcal{Z}} \sum_{k=0}^{\infty} \pi_\gamma(k) \int \prod_{i=1}^{k} \mathcal{D}P_i \mathcal{P}[P_i] \delta\left[P_0 - \hat{P}^{(k)}[\{P_i\}]\right] \hat{Z}^{(k)}[\{P_i\}]^{y_G/y_J} \mathrm{e}^{-z\chi(k,\gamma)}, \\
&P_0(h_0) = \hat{P}^{(k)}[\{P_i\}](h_0) \\
&\qquad = \frac{1}{Z} \mathbb{E}_J \int \prod_{i=1}^{k} \mathrm{d}h_i \, P_i(h_i) \delta(h_0 - \hat{h}^{(k,\{J_i\})}(\{h_i\})) \mathrm{e}^{-y_J \Delta \hat{E}_n^{(k;\{J_i\})}(\{h_i\})}, \\
&Z = \hat{Z}^{(k)}[\{P_i\}] = \mathbb{E}_J \int \prod_{i=1}^{k} \mathrm{d}h_i \, P_i(h_i) \mathrm{e}^{-y_J \Delta \hat{E}_n^{(k;\{J_i\})}(\{h_i\})}, \\
&\mathrm{e}^z = \sum_{k=0}^{\infty} \sigma(k) \left[\int \prod_{i=1}^{2} \mathcal{D}P_i \mathcal{P}[P_i] \left(\mathbb{E}_J \int \prod_{i=1}^{2} \mathrm{d}h_i \, P_i(h_i) \mathrm{e}^{-y_J \Delta \hat{E}_\ell^{(J_{12})}(h_1,h_2)} \right)^{y_G/y_J} \right]^k, \\
&\mathcal{F}(y_J, y_G, \gamma) = -\ln \mathcal{Z} = -\ln \left[\sum_{k=0}^{\infty} \pi_\gamma(k) \int \prod_{i=1}^{k} \mathcal{D}P_i \mathcal{P}[P_i] \delta[P_0 - \hat{P}^{(k)}[\{P_i\}]] \right. \\
&\qquad \left. \times \hat{Z}^{(k)}[\{P_i\}]^{y_G/y_J} \mathrm{e}^{-z\chi(k,\gamma)} \right],
\end{aligned}
\tag{67}
$$

where $\mathbb{E}_J$ denotes the average over the couplings J. Note that, with $y_G = 0$, we obtain large deviations with respect to the couplings on a typical graph, while with $y_J = 0$, we obtain large deviations with respect to the graphs for typical couplings. With $y_G = y_J$, the two sources of disorder are treated at a same level, in analogy with $y = \beta$ in section 4.1 and $\mu = y$ ($x = 1$) in section 4.2 (this prescription is sometimes referred to as the 'Nishimori temperature' [37]).

An interesting feature of the cavity equations is the possibility to implement them as a message-passing algorithm [34] to study, for example, here the large deviations with respect to the couplings on a given graph. The message passed along the oriented edge $(i \to j)$ is the distribution $P^{(i\to j)}$, which, in particular cases, can be parameterized by a finite set of real numbers, and the update rule is

$$
P^{(i\to 0)}(h_i) = \frac{1}{Z} \mathbb{E}_J \int \prod_{j \in i-0} \mathrm{d}h_j \, P^{(j\to i)}(h_j) \delta(h_0 - \hat{h}^{(k,\{J_{ji}\})}(\{h_j\})) \mathrm{e}^{-y_J \Delta \hat{E}_n^{(k;\{J_{ji}\})}(\{h_j\})} \tag{68}
$$

where the notation $j \in 0 - i$ means that, on the particular graph considered, j is a neighbour of 0 different from i. This algorithmic scheme could be used to design graphs with optimal properties, with for instance applications in coding theory, in the context of low-density parity-check codes [38].

5. Conclusion

While statistical physics of disordered systems has so far mostly focused on the thermodynamical properties of samples which are typical with respect to the source of quenched disorder, we have shown here that its methods can be extended to describe large deviations, that is, atypical samples. Large deviations are of foremost interest

J. Stat. Mech. (2005) P07004

in probability theory, and the approach followed here, though admittedly non-rigorous, is based on clearly formulated assumptions which should be amenable to mathematical justifications. In its simplest form, indeed, the LDCM we exposed assumes no replica symmetry breaking, a situation in which many of its typical predictions have been proved to be exact [39].

From the perspective of algorithmic complexity, the LDCM can be seen as a first step in an attempt to reconcile the worst case and typical case analysis, usually regarded as antagonistic. However, the scope of large deviations should not be regarded as restricted to optimization theory, as it notably allows us to work out the statistical mechanics of physical systems with adaptive structures. An example of such a system is constituted by random networks subject to mechanical constraints where the possibility to adapt leads to the occurrence of an intriguing intermediate phase, preceding a rigidity transition [8]; another example along the same lines is given by proteins whose structure is shaped by strong constraints [40]. Adaptative structures in random graphs are also of interest in the seemingly unrelated field of neural networks [41]. Finally, in the spirit of the most impressive achievements of its typical counterpart [34], it would also be interesting to implement the LDCM on particular ensembles of instances to analyse, and possibly design, graph structures with specific properties.

Acknowledgments

I am grateful to Andrea Montanari for providing me with his notes [25] on the analysis of large deviations by the replica method, and to Marc Mézard for his interest and encouragement.

References

[1] Papadimitriou C H and Steiglitz K, 1982 *Combinatorial Optimization: Algorithms and Complexity* (Englewood Cliffs, NJ: Prentice-Hall)
[2] Mézard M, Parisi G and Virasoro M A, 1987 *Spin-Glass Theory and Beyond* (*Lecture Notes in Physics* vol 9) (Singapore: World Scientific)
[3] Cover T M and Thomas J A, 1991 *Elements of Information Theory* (New York: Wiley)
[4] Bollobás B, 2001 *Random Graphs* 2nd edn (Cambridge: Cambridge University Press)
[5] Shannon C E, 1948 *Bell Syst. Tech. J.* **27** 379
[6] Erdős P and Rényi A, 1960 *Publ. Math. Inst. Hungar. Acad. Sci.* **5** 17
[7] Dembo A and Zeitouni O, 1993 *Large Deviations Techniques and Applications* (Boston: Jones and Bartlett)
[8] Barré J, Bishop A R, Lookman T and Saxena A, 2004 *Preprint* cond-mat/0408385
[9] Rivoire O, 2005 in preparation
[10] Andreanov A, Barbieri F and Martin O C, 2004 *Eur. Phys. J.* B **41** 365
[11] Rivoire O, 2004 *J. Stat. Phys.* **117** 453
[12] Montanari A and Zecchina R, 2002 *Phys. Rev. Lett.* **88** 178701
[13] Engel A, Monasson R and Hartmann A K, 2004 *J. Stat. Phys.* **117** 387
[14] Derrida B, Lebowitz J L and Speer E R, 2002 *J. Stat Phys.* **107** 599
[15] Barré J, Bouchet F, Dauxois T and Ruffo S, 2004 *Preprint* cond-mat/0406358
[16] Ellis R S, 1985 *Entropy, Large Deviations, and Statistical Mechanics* (New York: Springer)
[17] Mézard M and Parisi G, 2001 *Eur. Phys. J.* B **20** 217
[18] Mézard M and Parisi G, 2003 *J. Stat. Phys.* **111** 1
[19] Garey M and Johnson D, 1979 *Computers and Intractability: a Guide to the Theory of NP-Completeness* (San Francisco, CA: Freeman)
[20] Wu F Y, 1982 *Rev. Mod. Phys.* **54** 235
[21] Weigt M and Hartmann A K, 2001 *Phys. Rev.* E **63** 056127
[22] Newman M E J, Strogatz S H and Watts D J, 2001 *Phys. Rev.* E **64** 026118

J. Stat. Mech. (2005) P07004

[23] Sherrington D and Kirkpatrick S, 1975 *Phys. Rev. Lett.* **35** 1792
[24] Derrida B, 1980 *Phys. Rev. Lett.* **45** 79
[25] Montanari A, 2002 unpublished notes
[26] Monasson R, 1995 *Phys. Rev. Lett.* **75** 2847
[27] Rivoire O, 2005 *PhD Thesis* Université Paris-Sud
[28] Wong K Y M and Sherrington D, 1988 *J. Phys. A: Math. Gen.* **21** L459
[29] Weigt M and Hartmann A K, 2000 *Phys. Rev. Lett.* **84** 6118
[30] Weigt M and Hartmann A K, 2003 *J. Phys. A: Math. Gen.* **36** 11069
[31] Bauer M and Golinelli O, 2001 *Eur. Phys. J.* B **24** 339
[32] Montanari A and Ricci-Tersenghi F, 2003 *Eur. Phys. J.* B **33** 339
[33] Mulet R, Pagnani A, Weigt M and Zecchina R, 2002 *Phys. Rev. Lett.* **89** 268701
[34] Mézard M and Zecchina R, 2002 *Phys. Rev.* E **66** 056126
[35] Braunstein A, Mulet R, Pagnani A, Weigt M and Zecchina R, 2003 *Phys. Rev.* E **68** 036702
[36] Krzakala F, Pagnani A and Weigt M, 2004 *Phys. Rev.* E **70** 046705
[37] Nishimori H, 2001 *Statistical Physics of Spin Glasses and Information Processing: An Introduction* (Oxford: Oxford University Press)
[38] Richardson T J, Shokrollahi M A and Urbanke R L, 2001 *IEEE Trans. Inf. Theory* **47** 619
[39] Talagrand M, 2003 *Spin Glasses: a Challenge for Mathematicians. Cavity and Mean Field Models* (New York: Springer)
[40] Thorpe M F, Rader A, Lei M, Jacobs D J and Kuhn L A, 2001 *J. Mol. Graphics Modeling* **19** 60
[41] Wemmenhove B and Skantzos N S, 2004 *J. Phys. A: Math. Gen.* **37** 7843
[42] Rizzo T, 2004 *Preprint* cond-mat/0404729
[43] van Mourik J and Saad D, 2002 *Phys. Rev.* E **66** 056120

J. Stat. Mech. (2005) P07004