Searching for feasible stationary states in reaction networks by solving a Boolean constraint satisfaction problem

A. Seganti,[1] A. De Martino,[1,2,3] and F. Ricci-Tersenghi[1,2,4]
[1]*Dipartimento di Fisica, Sapienza Università di Roma, p.le A. Moro 2, 00185 Rome, Italy*
[2]*IPCF-CNR, UOS di Roma, Dipartimento di Fisica, Sapienza Università di Roma, Rome, Italy*
[3]*Center for Life Nano Science@Sapienza, Istituto Italiano di Tecnologia, Viale Regina Elena 291, 00161 Rome, Italy*
[4]*INFN, Sezione di Roma 1, Dipartimento di Fisica, Sapienza Università di Roma, Rome, Italy*
(Received 6 November 2013; published 26 February 2014)

We analyze the solutions, on single network instances, of a recently introduced class of constraint-satisfaction problems (CSPs), describing feasible steady states of chemical reaction networks. First, we show that the CSPs generalize the scheme known as network expansion, which is recovered in a specific limit. Next, a full statistical mechanics characterization (including the phase diagram and a discussion of the physical origin of the phase transitions) for network expansion is obtained. Finally, we provide a message-passing algorithm to solve the original CSPs in the most general form.

DOI: 10.1103/PhysRevE.89.022139 PACS number(s): 02.50.Tt, 05.10.−a, 87.16.Yc

I. INTRODUCTION

The mathematical theory of chemical reaction networks is mainly concerned with issues such as the existence, number, and stability of fixed points for realistic (e.g., mass-action-based) dynamics of continuous state variables representing the concentrations of chemical compounds. In many real-world cases, however, full-fledged dynamical approaches are prevented either by a lack of knowledge about kinetic constants or by sheer size considerations. This is the case, for instance, for cellular metabolic networks at genome-scale [1,2]. On the other hand, being able to describe viable steady states of the network in terms of coarse-grained, discrete variables might be a useful (albeit less comprehensive) alternative.

At the simplest level, the operation of networks of chemical reactions can be thought to depend on the availability of reaction substrates and of the enzymes required to process them. In particular, one can think that reactions may occur whenever all of the required substrates are available, which in turn makes the reaction products available, and so on. Likewise, compound availability can be assumed to depend at least on there being an active reaction producing it. Using these ideas, it is possible to associate to a given reaction network (i.e., to a given bipartite graph encoding for reaction-compounds interactions) a set of logical constraints to be satisfied by Boolean state variables indicating whether a reaction is active or inactive and whether a compound is available or not. This type of reasoning has led to the formulation of the network expansion (NE) framework [3–6], which is aimed at quantifying the amount of activity in the network bulk that is generated upon assuming the availability of a seed of compounds. For cellular metabolic networks, the seed usually includes both external species (e.g., nutrients) and internal ones (e.g., water, currency metabolites such as ATP, etc.). In concrete terms, given a seed, one would like to retrieve the pattern(s) of reaction-activation–compound availability that are induced via the topology (or, more properly, the stoichiometry) of the reaction network.

Network expansion is basically a Boolean constraint satisfaction problem (CSP), possibly one of several types that can be reasonably defined to describe the operation of chemical systems, in analogy with those used in the past to describe other biological mechanisms such as transcriptional regulation [7–9]. Despite their "natural" appeal, however, they have received little attention from a statistical mechanics perspective. Besides the general theoretical interest, extracting biologically significant information from them requires being able to explore (in a controlled way) their very large and rich space of solutions. Devising algorithms that are able to carry out this task, even for the basic NE scheme, however, is far from trivial.

Recently [10], we have proposed a class of CSPs inspired by the so-called constraint-based models for flux analysis [11], which is aimed at describing the space of configurations of a chemical reaction network through minimal Boolean feasibility constraints. These CSPs have been defined on random reaction networks (RRN), i.e., bipartite graphs (the two classes of nodes corresponding to "reactions" and "reagents," respectively) characterized by the parameters λ, representing the mean of the Poisson distributed degrees of metabolites, and q $(1-q)$, giving the probability that a reaction has two (one) input or output compounds. The structure of a RRN is described by an $M \times N$ connectivity matrix $\widehat{\xi}$, with $\xi_i^m \in \{1,0,-1\}$ depending on whether compound $m \in \{1,\ldots,M\}$ is a substrate ($\xi_i^m = -1$), a product ($\xi_i^m = 1$), or is not involved ($\xi_i^m = 0$) in reaction $i \in \{1,\ldots,N\}$. In this kind of network, a *nutrient* is a compound with in-degree 0, while a *sink* has out-degree 0. Denoting by $\nu_i \in \{0,1\}$ (inactive or active) the state associated to reaction i and by $\mu_m \in \{0,1\}$ (unavailable or available) that associated to compound m, for a given RRN, feasible assignments ($\boldsymbol{\mu} = \{\mu_m\}, \boldsymbol{\nu} = \{\nu_i\}$) are defined to be such that $\Gamma_m = 1 \forall\, m$ (except for nutrients that are externally fixed) and $\Delta_i = 1 \forall\, i$, where

$$\Gamma_m = \delta_{\mu_m,0}\delta_{x_m,0}(\delta_{y_m,0})^{\alpha} + \delta_{\mu_m,1}(1-\delta_{x_m,0})(1-\delta_{y_m,0})^{\alpha} \quad (1)$$

$$\Delta_i = \delta_{\nu_i,0} + \delta_{\nu_i,1}\prod_{m\in\partial i_{\text{in}}}\mu_m, \quad (2)$$

$\partial i_{\rm in}$ is the set of substrates of reaction i, $\alpha \in \{0,1\}$ is a fixed parameter, and

$$x_m \equiv \sum_{i\in\partial m_{\rm in}} \nu_i \quad \text{and} \quad y_m \equiv \sum_{i\in\partial m_{\rm out}} \nu_i, \tag{3}$$

with $\partial m_{\rm in}$ ($\partial m_{\rm out}$) the set of reactions producing (consuming) chemical species m. Condition (2) says that reactions can always be inactive, while they can activate only if all input compounds are available; likewise, condition (1) allows for a compound m to be available if at least one reaction produces it (for $\alpha = 0$) or if at least one reaction produces it and one consumes it (for $\alpha = 1$). As explained in detail in [10], the case $\alpha = 0$ ("soft mass balance" or soft-MB) describes steady states that allow for a net production of compounds, while the case $\alpha = 1$ ("hard mass balance" or hard-MB) corresponds, in this coarse-grained view, to a fully mass-balanced scenario.

Once the topology of the reaction network, encoded in an adjacency matrix $\widehat{\xi}$, is given, the setup presented in [10] aims to retrieve Boolean patterns of activity of reactions (or of metabolite availabilities) induced by the fact that a certain seed of metabolites (in our case, formed by nutrients only) is available from the outset. The cavity-based population dynamics technique developed in [10] allows in particular to sample configurations $(\boldsymbol{\mu},\boldsymbol{\nu})$ with a probability given by

$$P(\boldsymbol{\mu},\boldsymbol{\nu}) \propto \prod_{m=1}^{M} \Gamma_m \prod_{i=1}^{N} \Delta_i e^{\theta\nu_i}, \tag{4}$$

with a "chemical potential" θ that allows us to select states according to the overall number of active processes and, in turn, compute network ensemble-averaged quantities. This study has revealed a rich phase structure characterized by hysteresis, which might potentially hinder the retrieval of individual solutions.

Here we extend the previous analysis in a direction hopefully more useful for applications to quantitative biology (which will be our next step) by searching for solutions to the above CSP on a *given* reaction network. In this context, we are going to discuss the statistical properties of the solutions found by several search methods (including NE) and introduce an improved decimation-based technique. Furthermore, we will compare the statistical properties of the solutions found in the single instance with the ensemble-averaged results obtained in [10]. The details of the cavity theory on which our algorithms are based, as well as of the algorithms themselves (belief propagation complemented by decimation), are reported in the Appendixes.

II. NETWORK EXPANSION REVISITED

A. The problem

The basic idea behind NE is that, given a seed compound (e.g., a nutrient), a reaction can (and will) activate when all its substrates are available (AND-like constraint), whereas a compound will be available if at least one of the reactions that produce it is active (OR-like constraint). The numerical procedure of NE transfers the information about the availability of certain metabolites across the network links, as explained pictorially in Fig. 1. We shall term this type of process a propagation of external inputs (PEI).

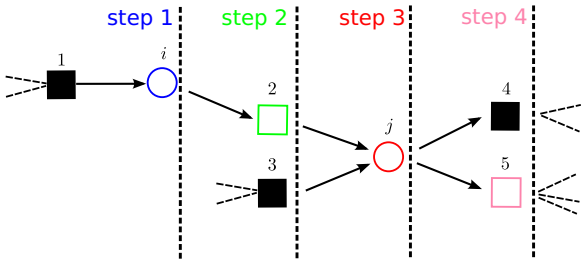

FIG. 1. (Color online) Sketch of four steps of the propagation of external inputs (PEI) algorithm (serving as the basis of the network expansion method [3]). Black squares represent compounds initially available. In step 1, reaction i is activated by virtue of the availability of compound 1; in step 2, metabolite 2 becomes available by virtue of the activation of reaction i; in step 3, reaction j activates as both 2 and 3 are available; and so on.

It is simple to understand that, as soon as the reaction network departs from a linear topological structure, the propagation will likely stop after a small number of steps unless the availability of additional compounds is invoked. Indeed, in network expansion PEI is aided by the assumption that highly connected metabolites such as water are abundant. Because of its intuitive appeal, it is useful to analyze briefly the properties of PEI in somewhat more detail.

One can write down equations for the probability $\langle\nu_i\rangle$ that reaction i will be active and for the probability $\langle\mu_m\rangle$ that metabolite m will be available by simply considering that, under PEI in a given network, a reaction can activate when all of its inputs are available and a metabolite becomes available when at least one reaction is producing it. This implies that

$$\langle\nu_i\rangle = \prod_{n\in\partial i_{\rm in}} \langle\mu_n\rangle\,, \tag{5}$$

$$1 - \langle\mu_m\rangle = \prod_{k\in\partial m_{\rm in}} (1 - \langle\nu_k\rangle). \tag{6}$$

To prove the link between PEI and the CSPs defined above, note that, using the definition (4), one can easily compute the mean values

$$\langle\nu_j\rangle = \frac{\sum_{\boldsymbol{\mu},\boldsymbol{\nu}} \nu_j \prod_{m=1}^{M} \Gamma_m \prod_{i=1}^{N} \Delta_i e^{\theta\nu_i}}{\sum_{\boldsymbol{\mu},\boldsymbol{\nu}} \prod_{m=1}^{M} \Gamma_m \prod_{i=1}^{N} \Delta_i e^{\theta\nu_i}}, \tag{7}$$

$$\langle\mu_n\rangle = \frac{\sum_{\boldsymbol{\mu},\boldsymbol{\nu}} \mu_n \prod_{m=1}^{M} \Gamma_m \prod_{i=1}^{N} \Delta_i e^{\theta\nu_i}}{\sum_{\boldsymbol{\mu},\boldsymbol{\nu}} \prod_{m=1}^{M} \Gamma_m \prod_{i=1}^{N} \Delta_i e^{\theta\nu_i}}. \tag{8}$$

Under the mean-field approximation, we can set

$$\Gamma_m(\mu_m,\{\nu_i\}) = \Gamma_m(\mu_m,\{\langle\nu_i\rangle\}), \tag{9}$$

$$\Delta_i(\nu_i,\{\mu_m\}) = \Delta_i(\nu_i,\{\langle\mu_m\rangle\}), \tag{10}$$

which in turn implies

$$\langle\nu_i\rangle = \frac{e^{\theta}\prod_{n\in\partial i_{\rm in}}\langle\mu_n\rangle}{1 + e^{\theta}\prod_{n\in\partial i_{\rm in}}\langle\mu_n\rangle}, \tag{11}$$

$$\langle\mu_m\rangle = 1 - \prod_{k\in\partial m_{\rm in}} (1 - \langle\nu_j\rangle). \tag{12}$$

In the limit $\theta \to \infty$, we have

$$\langle \nu_i \rangle = \begin{cases} 1 & \text{if } \prod_{n\in\partial i_{\text{in}}} \langle \mu_n \rangle = 1 \\ 0 & \text{if } \prod_{n\in\partial i_{\text{in}}} \langle \mu_n \rangle = 0 \end{cases} \tag{13}$$

so that Eqs. (5) and (6) are recovered. In other terms, PEI is the mean-field approximation at $\theta \to \infty$ of the CSPs considered in [10].

It is simple to derive analytically the phase diagram of PEI in the ensemble of RRN defined in [10]. The probability that a metabolite is available is

$$\gamma = \overline{\langle \mu_m \rangle}, \tag{14}$$

where the overbar denotes an average over the network realizations. Using (5) and (6), one sees that

$$\gamma = e^{-\lambda}\rho_{\text{in}} + \sum_{k_m \geqslant 1} D_M(k_m) \left(1 - \prod_{j=1}^{k_m} (1 - \overline{\langle \nu_j \rangle}) \right),$$

where we have assumed that nutrients (fractionally given by roughly $e^{-\lambda}$ nodes) have a fixed probability ρ_{in} of being available and where $D_M(k) = e^{-\lambda}\lambda^k/k!$ is the distribution of metabolite in- (and out-) degrees. In turn, this gives

$$\gamma = e^{-\lambda}\rho_{\text{in}} + 1 - e^{-\lambda\tau}, \tag{15}$$

where $\tau = \overline{\langle \nu_i \rangle}$ is the probability that a reaction is active, which, recalling that the in- and out-degrees of reactions are distributed according to $D_R(d) = q\delta_{d,2} + (1-q)\delta_{d,1}$, satisfies (within a mean-field approximation)

$$\tau = \overline{\prod_{b\in\partial i_{\text{in}}} \langle \mu_n \rangle} = (1-q)\gamma + q\gamma^2. \tag{16}$$

Putting things together, γ is seen to satisfy the condition

$$\gamma = e^{-\lambda}\rho_{\text{in}} + 1 - \exp\{-\lambda[(1-q)\gamma + q\gamma^2]\}, \tag{17}$$

which can be solved for γ upon changing the values of ρ_{in}, q, and λ. The resulting phase diagram in the (q,λ) plane, based on the behavior of the solution $\gamma^\star(\rho_{\text{in}})$, is displayed in Fig. 2.

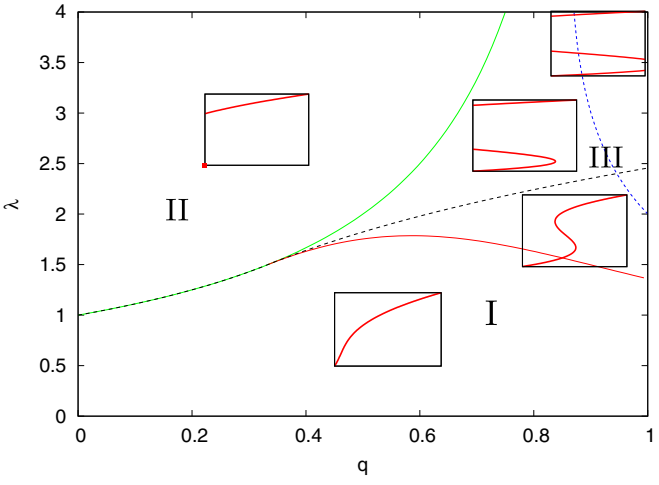

FIG. 2. (Color online) Phase diagram obtained by propagation of external inputs (PEI) on RRN in the (q,λ) plane. The insets display the curves $\gamma^\star$ vs ρ_{in} obtained in the different sectors; all lines are analytical. See text for details.

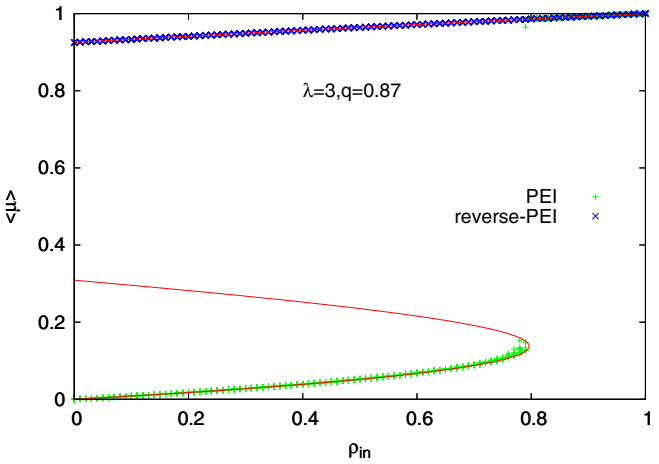

FIG. 3. (Color online) Theoretical solution of Eq. (17) (solid line) vs ρ_{in}, together with the results obtained by PEI and reverse-PEI.

Three regions can be distinguished. In region I, Eq. (17) has a unique solution and $\gamma^\star$ is a monotonously increasing function of ρ_{in} (note that $\gamma^\star = 0$ is always a solution when $\rho_{\text{in}} = 0$). Outside region I, the curve $\gamma^\star$ versus ρ_{in} displays an inflection point. If the point lies outside the interval $[0,1]$ (for both γ and ρ_{in}), then (17) has a unique nonzero solution for $\rho_{\text{in}} > 0$ and two different solutions at $\rho_{\text{in}} = 0$ (region II). In region III, instead, a range of values of ρ_{in} exists where three distinct solutions (with different values of γ) of (17) occur. This sector can be further divided according to the number of solutions found for $\rho_{\text{in}} = 0$ and 1. The black dashed line marks the boundary between phases with, respectively, one and three solutions for $\rho_{\text{in}} = 0$, while the dashed blue line separates the region with one and three solutions for $\rho_{\text{in}} = 1$.

For any fixed ρ_{in}, whenever solutions with different values of $\langle \mu \rangle$ coexist, those with the smallest $\langle \mu \rangle$ can be retrieved by straightforward PEI starting from a configuration where no metabolite is available except for nutrients. Solutions with larger $\langle \mu \rangle$, on the other hand, can be found by "reverse-PEI." In this procedure, a configuration where internal metabolites are all available and nutrients are fixed with probability ρ_{in} is initially selected, and then a solution is found by enforcing the constraints in an iterative way. The results for both procedures are presented in Fig. 3 for $\lambda = 3$ and $q = 0.87$ (deep into region III in Fig. 2).

B. Origin of the phase transition within the mean-field approximation in PEI

We show here that, as might have been guessed, the phase transitions occurring in PEI (see Fig. 2) are, from a physical viewpoint, of a percolation type.

To analyze the effectiveness of PEI, we start by identifying the so-called propagation of external regulation (PER) core of the system [12], which is the subnetwork obtained by fixing the nutrient availability (with probability ρ_{in}) and then propagating this information inside the network. In this way, some variables will be assigned a definite value (either 1 or 0). At convergence, a fraction γ_1 (γ_0) of metabolites will be fixed to 1 (0), while a fraction τ_1 (τ_0) of reactions will be fixed to 1 (0). One easily

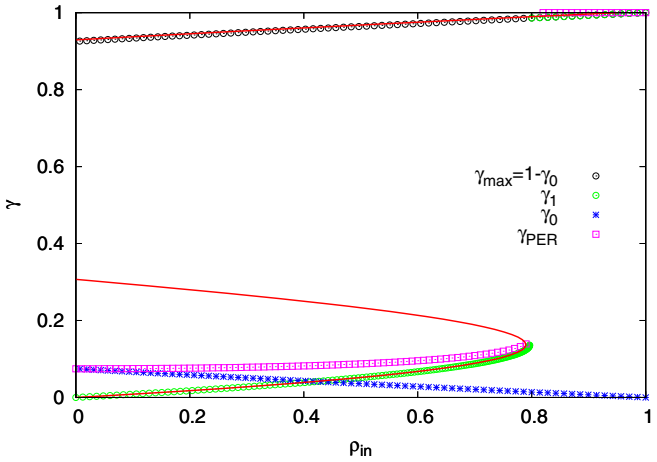

FIG. 4. (Color online) Weights of the different components of the PER core for a graph with $\lambda = 3$ and $q = 0.87$. The red line corresponds to the solution of Eq. (17).

sees that, at the fixed point, the following equations hold:

$$1 - \tau_0 = q(1 - \gamma_0)^2 + (1 - q)(1 - \gamma_0), \tag{18}$$

$$\gamma_0 = \sum_{k\neq 0} D_M(k)\tau_0^k + (1 - \rho_{\text{in}})D_M(0), \tag{19}$$

$$\tau_1 = q\gamma_1^2 + (1 - q)\gamma_1, \tag{20}$$

$$1 - \gamma_1 = \sum_{k\neq 0} D_M(k)(1 - \tau_1)^k + (1 - \rho_{\text{in}})D_M(0). \tag{21}$$

In turn, one obtains

$$\tau_0 = 1 - q(1 - \gamma_0)^2 - (1 - q)(1 - \gamma_0), \tag{22}$$

$$\gamma_0 = e^{-\lambda}e^{\lambda\tau_0} - \rho_{\text{in}}e^{-\lambda}, \tag{23}$$

$$\tau_1 = q\gamma_1^2 + (1 - q)\gamma_1, \tag{24}$$

$$\gamma_1 = 1 - e^{-\lambda\tau_1} + \rho_{\text{in}}e^{-\lambda}. \tag{25}$$

Unsurprisingly, the equations for γ_1 and τ_1 take us back to (17). On the other hand, the fraction of metabolites in the PER core is given by

$$\gamma_{\text{PER}} = \gamma_1 + \gamma_0. \tag{26}$$

Hence the fraction of metabolites that are not fixed by propagating nutrient availability is given by $1 - \gamma_{\text{PER}}$, and the maximum achievable availability for metabolites (that we will often call "magnetization" in the following using statistical physics jargon) is given by $\gamma_{\max} = 1 - \gamma_0$. Figure 4 displays the different contributions for a specific choice of the parameters, together with the corresponding solution of Eq. (17).

The excellent agreement of γ_1 with the analytical line for the feasible values of the magnetization suggests that straightforward PEI will be able to recover solutions with lower magnetization when the latter coexist with high-magnetization solutions. On the other hand, the highest magnetizations coincide, expectedly, with the largest achievable average metabolite availability. Finally, depending on the value of λ and q, one obtains a single solution when no PER core exists, and two solutions (with magnetizations $\gamma_{\max}$ and γ_1) in the presence of a PER core. Hence the transition is a typical percolation transition between a phase in which the internal variables are trivially determined by the nutrients (in the absence of a PER core) to one in which the internal variables are not univocally determined (in the presence of a PER core).

III. SOLUTIONS ON INDIVIDUAL NETWORKS BY BELIEF PROPAGATION AND DECIMATION

We turn now to the analysis of the soft-MB and hard-MB CSPs (1) and (2) for general θ. In essence, we have derived the cavity equations for the CSPs, presented in Appendix A1, and used the belief propagation (BP) algorithm discussed in Appendix A2a to compute the statistics of solutions on single instances of RRNs. Next, in order to obtain *individual configurations* of variables that satisfy our CSPs, we resorted to the decimation scheme presented in Appendix A2b. Results are presented in Figs. 5 and 6 for soft-MB ($\alpha = 0$) and in Figs. 7 and 8 for hard-MB ($\alpha = 1$). Results from belief propagation, labeled "BP," are compared with results retrieved by the population dynamics algorithm developed in [10] (labeled "pop" and corresponding to the ensemble average) and with

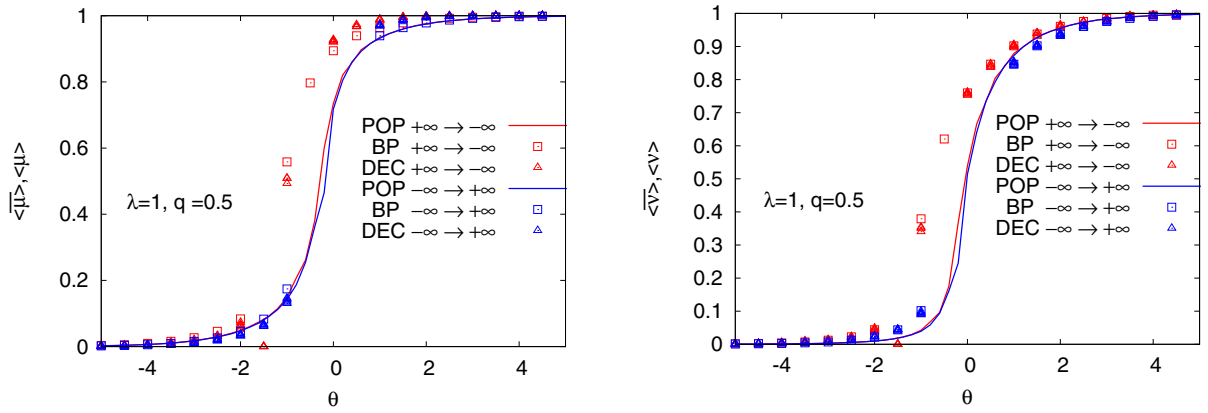

FIG. 5. (Color online) Soft-MB for $\lambda = 1$, $q = 0.5$, and $\rho_{\text{in}} = 1$. Left: average fraction of available metabolites, $\langle\mu\rangle$ ($\overline{\langle\mu\rangle}$ for population dynamics) vs θ. Right: average fraction of active reactions, $\langle\nu\rangle$ ($\overline{\langle\nu\rangle}$ for population dynamics) vs θ.

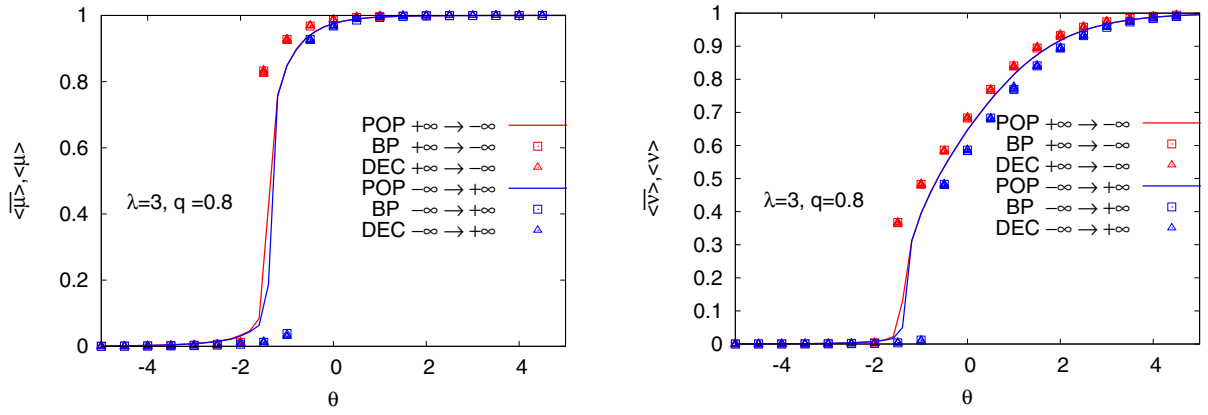

FIG. 6. (Color online) Soft-MB for $\lambda = 3$, $q = 0.8$, and $\rho_{\text{in}} = 1$. Left: average fraction of available metabolites, $\langle\mu\rangle$ ($\overline{\langle\mu\rangle}$ for population dynamics) vs θ. Right: average fraction of active reactions, $\langle\nu\rangle$ ($\overline{\langle\nu\rangle}$ for population dynamics) vs θ.

the decimation results (labeled "dec"). In [10], the solution space was explored by two different protocols, which we also use here: by reducing θ starting from a large positive value ($+\infty \to -\infty$ in the figure legends) and by doing the reverse ($-\infty \to +\infty$ in the figure legends). If the decimation scheme does not converge, the corresponding point is absent.

It is clear that decimation generically fails to converge close to the transitions both in the soft-MB and, more severely, in the hard-MB case. Apart from this, the three methods give results that are in remarkable qualitative agreement, including the ability to describe discontinuities in $\langle\mu\rangle$ and $\langle\nu\rangle$ upon varying θ. It is noteworthy that many different configurations appear to be feasible. These configurations are spread over a broad range of densities, especially in the soft-MB case. So our method based on BP and decimation is able to sample the solution space by just varying a single parameter (the chemical potential θ in the present case), even in cases when only "extremal" solutions seem to satisfy the CSP for metabolite nodes (as, e.g., in the left panel in Fig. 8) while the density of active reaction is varying in a more continuous manner (see the right panel in the same figure).

As detailed in Appendix A1, during decimation nutrients must be treated with special care. This is because the prior assignment of availability for each nutrient (which, as mentioned above, follows a probabilistic rule with parameter ρ_{in}) does not always coincide, after decimation, with the fraction $\langle\mu\rangle_{\text{EXT}}$ of nutrients available in the actual solution retrieved. We analyze the relation between the average magnetization of all metabolites $\langle\mu\rangle$, the average magnetization of nutrients $\langle\mu\rangle_{\text{EXT}}$, and the parameter ρ_{in} in Figs. 9 and 10. We first note that the decimation algorithm is able to obtain solutions with very different values of $\langle\mu\rangle$ and $\langle\mu\rangle_{\text{EXT}}$. The quantity $\langle\mu\rangle$ mainly depends on θ and is rather insensitive to the value of ρ_{in} (see Fig. 9), while the quantity $\langle\mu\rangle_{\text{EXT}}$ is correlated to ρ_{in} and takes in general a value larger than ρ_{in} (see Fig. 10). We note a difference between the data shown in the two panels of Fig. 10: the data in the left panel show a clear dependence of $\langle\mu\rangle$ on ρ_{in} that is missing in the right panel data. This can be ascribed to the different topological structure of the underlying graphs, the one with parameters $\lambda = 1$ and $q = 0.5$ being richer in linear or quasilinear pathways, through which the information about nutrients availability may propagate in a straightforward way. We are not showing these correlations for the hard-MB case, since the solutions found on random reaction networks are typically concentrated around $\langle\mu\rangle = 0$ and 1, and computing such correlations is impossible.

Finally, we would like to compare the solutions of the complete problem to the solutions obtained using the mean-field approximation (MF) presented in the preceding section. Indeed, in Sec. II A we showed how to obtain solutions for

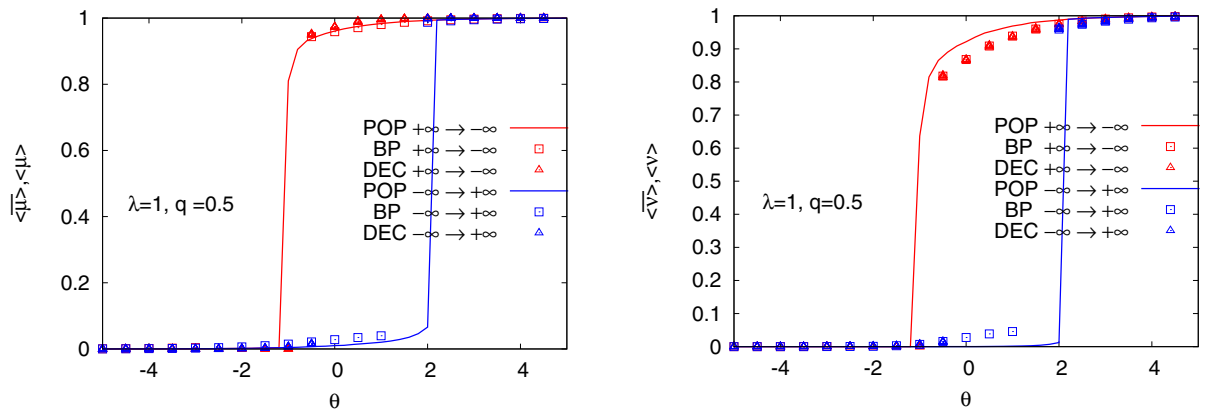

FIG. 7. (Color online) Hard-MB for $\lambda = 1$, $q = 0.5$, and $\rho_{\text{in}} = 1$. Left: average fraction of available metabolites, $\langle\mu\rangle$ ($\overline{\langle\mu\rangle}$ for population dynamics) vs θ. Right: average fraction of active reactions, $\langle\nu\rangle$ ($\overline{\langle\nu\rangle}$ for population dynamics) vs θ.

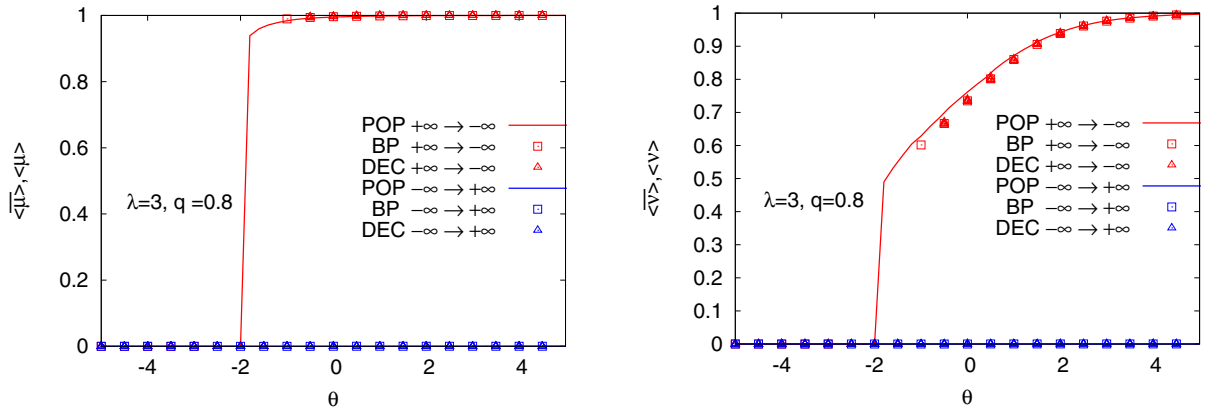

FIG. 8. (Color online) Hard-MB for $\lambda = 3$, $q = 0.8$, and $\rho_{\text{in}} = 1$. Left: average fraction of available metabolites, $\langle\mu\rangle$ ($\overline{\langle\mu\rangle}$ for population dynamics) vs θ. Right: average fraction of active reactions, $\langle\nu\rangle$ ($\overline{\langle\nu\rangle}$ for population dynamics) vs θ.

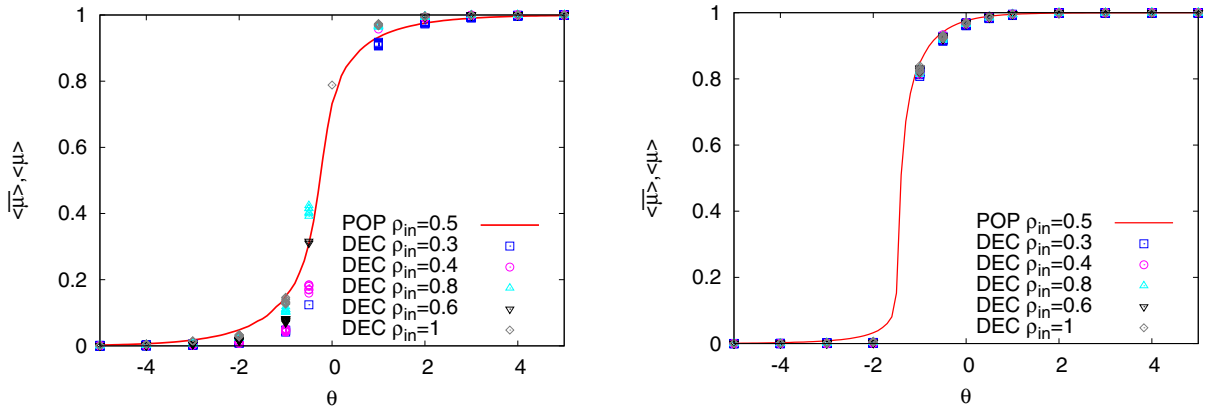

FIG. 9. (Color online) Soft-MB: the behavior of the average fraction of available metabolites, $\langle\mu\rangle$ ($\overline{\langle\mu\rangle}$ in population dynamics) for $\lambda = 1$ and $q = 0.5$ (left panel) and $\lambda = 3$ and $q = 0.8$ (right panel), seems to depend weakly on the value of ρ_{in} (the probability that nutrients are present at the first decimation step).

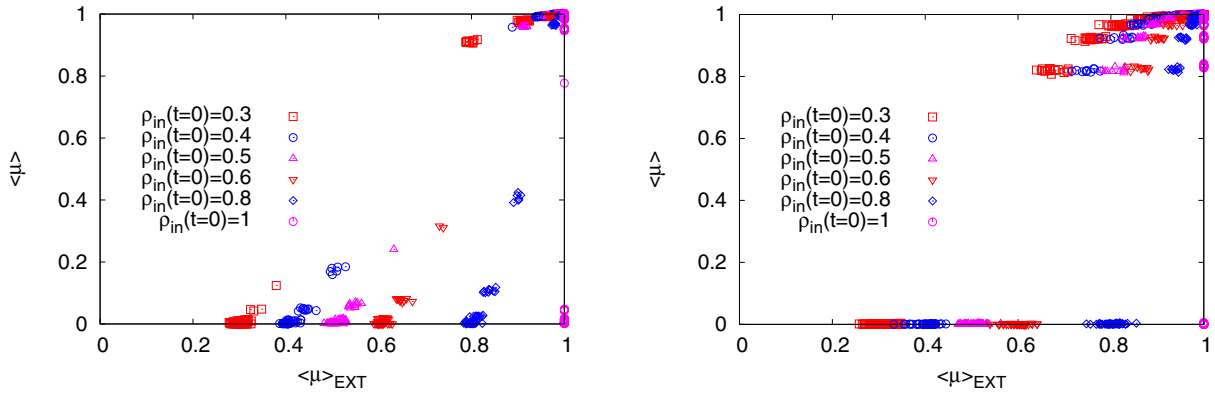

FIG. 10. (Color online) Soft-MB: Behavior of $\langle\mu\rangle$ vs $\langle\mu\rangle_{\text{EXT}}$ for various ρ_{in}, for $\theta = (-5, -4.5, \ldots, 4.5, 5)$ and for $\lambda = 1$ and $q = 0.5$ (left) and $\lambda = 3$ and $q = 0.8$ (right). ρ_{in} is the probability that nutrients are available before starting the decimation process, while $\langle\mu\rangle_{\text{EXT}}$ is the fraction of available nutrients in the solution actually found by the decimation process. We see that $\langle\mu\rangle_{\text{EXT}}$ and ρ_{in} are well correlated, especially for small $\langle\mu\rangle$.

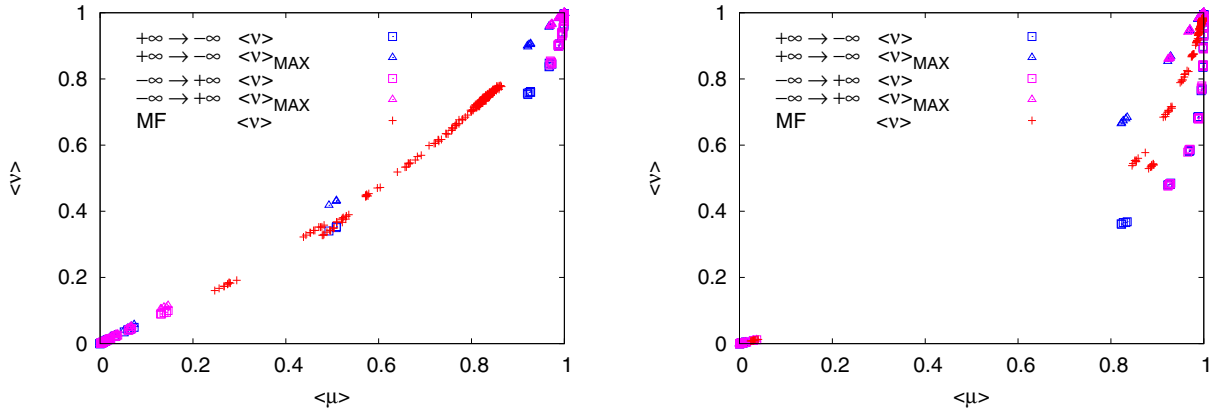

FIG. 11. (Color online) Plot of $\langle\nu\rangle$ vs $\langle\mu\rangle$ for $\lambda = 1$ and $q = 0.5$ (left) and $\lambda = 3$ and $q = 0.8$ (right). Data have been obtained for both the mean field (MF) and the complete problem solutions. $\langle\nu\rangle_{\max}$ is the largest reaction magnetization achievable in the complete problem, without changing the state of metabolites.

the MF problem at $\theta \to \infty$ using the PEI or reverse-PEI procedure. However, in order to compare the two approaches, MF solutions at all θ have to be computed. This can be done by searching for solutions to the MF equations (11) at finite θ, and then using the same decimation procedure presented in Appendix A2b. It is important to notice, however, that in the MF case both BP and the decimation algorithm can be written in a simpler form as only one message per variable is needed; furthermore, for $\theta \to \infty$, BP and decimation together behave exactly as a warning propagation algorithm [13].

In Fig. 11 we present the magnetizations $\langle\mu\rangle$ and $\langle\nu\rangle$ of solutions obtained by the decimation procedure both for MF and for the complete problem, with different values of $\theta \in (-\infty,\infty)$. Here MF solutions are represented by crosses, while squares represent the magnetizations of the solutions obtained for the complete CSP. In the latter case, $\langle\nu\rangle$ seems to be always smaller than in MF. However, we have to keep in mind that our CSP allows for configurations where a reaction is inactive even if all its neighboring metabolites are present. Such reactions could be switched on without violating any constraint. The data marked $\langle\nu\rangle_{\max}$ (triangles in Fig. 11) have been obtained by switching on all possible reactions without changing the configuration of metabolites. This is the upper bound for the reaction activity in the complete problem, and it always lies above the MF result.

From data shown in Fig. 11 it is clear that the solutions to the complete problem span a wider range in both $\langle\mu\rangle$ and $\langle\nu\rangle$. Moreover, the solutions sampled at the MF level are a subset of the solutions found in the complete problem. Nevertheless, the MF equations are simpler to solve and so they can be useful in the analysis of larger and more structured networks.

IV. CONCLUSIONS

Stationary states of chemical reaction networks can often be described in a compact way through the information regarding reaction activity or inactivity and reagent availability or unavailability. In these conditions, Boolean CSPs provide a framework to describe feasible operation states of chemical reaction networks. The problem posed by sampling their solution space (even for an individual network, as discussed here) is, however, substantial. We have presented an efficient computational method to generate solutions for a class of CSPs inspired by constraint-based models of cell metabolism. Extending previous work concerned with ensemble properties, we have focused here on characterizing the solution space for single instances of RRNs, and on clarifying the connection between the CSPs discussed in [10] and the network expansion scheme [3]. Concerning the latter point, we have shown that NE is recovered as a limiting case of the present CSPs, and that our method permits a thorough exploration of its solution space, much beyond the computational approaches employed previously. Moreover, after computing the exact phase diagram of NE, we have quantitatively connected the macroscopic changes in the solutions to percolation phenomena.

Regarding the general CSPs, we have presented results obtained by a decimation algorithm, which is able to find many different solutions in a wide range of $\langle\mu\rangle$ and $\langle\nu\rangle$. Measurements made on single instances turn out to be in remarkable agreement with the population dynamics study of [10]. In addition, we have quantified the relation between the initial nutrient availability (ρ_{in}) and the final one ($\langle\mu\rangle_{\text{EXT}}$).

The method presented here can be generalized to include a certain fraction of reversible reactions. Applicability to more realistic network topologies is potentially limited by convergence issues. Future work will explore this aspect and, more importantly, the emerging picture of the solution space on bacterial metabolic networks. The type of approach discussed here (a simplified Boolean CSP), by providing in a quick way information about reaction activity and metabolites availability, can be of valuable help in improving standard algorithms for the sampling of feasible solutions of linear constraint-based models such as FBA.

ACKNOWLEDGMENTS

This work is supported by the Italian Research Minister through the FIRB Project No. RBFR086NN1 and by the

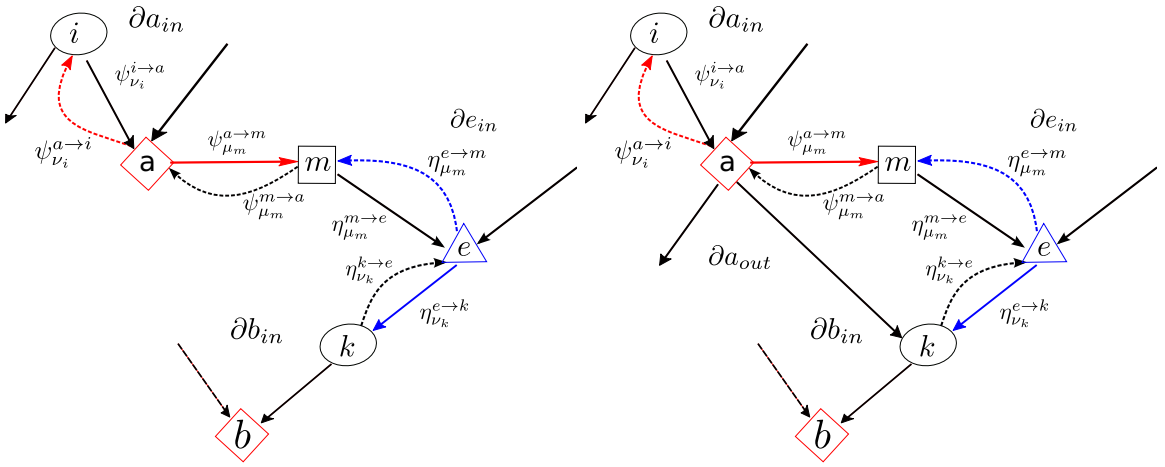

FIG. 12. (Color online) Schema of the cavity method for soft-MB (left) and hard-MB (right) constraints.

DREAM Seed Project of the Italian Institute of Technology (IIT). The IIT Platform Computation is gratefully acknowledged.

APPENDIX

1. Cavity equations

In general a CSP, such as soft-MB or hard-MB, can be solved efficiently in random networks using the belief propagation algorithm [14] or equivalently the replica symmetric cavity method [15]. In this method, the marginal of a variable is computed by creating a "cavity" inside the system, removing a subpart of the network. Thus it is possible to obtain a "cavity marginal" and then reintroduce the variables removed. Finally, the complete marginal of the variables follows directly from the cavity marginals.

In this kind of approach, the system is divided between "variable" and "function" nodes. In the RRN this amounts to adding two types of function nodes (constraint Γ and Δ) as presented in Fig. 12. In the following, we will use the letters $a,b,\ldots$ for the metabolite constraint, and $e,f,\ldots$ for the reaction constraint. Furthermore, we introduce the condensed notations $\partial a^R = \partial a \backslash m$, as the reaction neighbors of the metabolite constraint a; $\partial e^M = \partial e \backslash i$, as the metabolites neighbors of the reaction constraint e; ∂a_i^R represents the reaction neighbors of a that are in the same group as i without i, and $\partial a_{\neg i}^R$ represents the reaction neighbors of a that are in the opposite group from i.

The resulting equations for the system are (for a full derivation, refer to [10])

$$\psi_{\mu_m}^{m\to a} = \prod_{f\in\partial m^R} \eta_{\mu_m}^{f\to m}/Z^{m\to a},$$

$$\psi_{\mu_m}^{a\to m} = \left[\delta_{\mu_m,0}\prod_{j\in\partial a_{\text{in}}^R}\psi_0^{j\to a}\left(\prod_{j\in\partial a_{\text{out}}^R}\psi_0^{j\to a}\right)^{\alpha} + \delta_{\mu_m,1}\left(1-\prod_{j\in\partial a_{\text{in}}^R}\psi_0^{j\to a}\right)\right.$$
$$\left.\times\left(1-\prod_{j\in\partial a_{\text{out}}^R}\psi_0^{j\to a}\right)^{\alpha}\right]/Z^{a\to m},$$

$$Z^{a\to m} = \left(1-\prod_{j\in\partial a_{\text{in}}^R}\psi_0^{j\to a}\right)\left(1-\prod_{j\in\partial a_{\text{out}}^R}\psi_0^{j\to a}\right)^{\alpha} + \prod_{j\in\partial a_{\text{in}}^R}\psi_0^{j\to a}\left(\prod_{j\in\partial a_{\text{out}}^R}\psi_0^{j\to a}\right)^{\alpha},$$

$$\psi_{\nu_i}^{i\to a} = \eta_{\nu_i}^{e\to i}\left(\prod_{b\in\partial i_{\text{in}}^M\backslash a}\psi_{\nu_i}^{b\to i}\right)^{\alpha}\prod_{b\in\partial i_{\text{out}}^M\backslash a}\psi_{\nu_i}^{b\to i}/Z^{i\to a},$$

$$\psi_{\nu_i}^{a\to i}Z^{a\to i} = \psi_0^{m\to a}(1-\nu_i)\prod_{j\in\partial a_{\text{in}}^R\backslash i}\psi_0^{j\to a}\left(\prod_{j\in\partial a_{\text{out}}^R\backslash i}\psi_0^{j\to a}\right)^{\alpha} + \psi_1^{m\to a}\left(1-\prod_{j\in\partial a_{\neg i}^R}\psi_0^{j\to a}\right)^{\alpha}$$
$$\times\left[\left(1-\prod_{j\in\partial a_i^R}\psi_0^{j\to a}\right)+\nu_i\prod_{j\in\partial a_i^R}\psi_0^{j\to a}\right],$$

$$Z^{a\to i} = \psi_0^{m\to a}\prod_{j\in\partial a_i^R}\psi_0^{j\to a}\left(\prod_{j\in\partial a_{\neg i}^R}\psi_0^{j\to a}\right)^{\alpha} + \psi_1^{m\to a}\left(1-\prod_{j\in\partial a_{\neg i}^R}\psi_0^{j\to a}\right)^{\alpha}$$
$$\times\left(2-\prod_{j\in\partial a_i^R}\psi_0^{j\to a}\right),$$

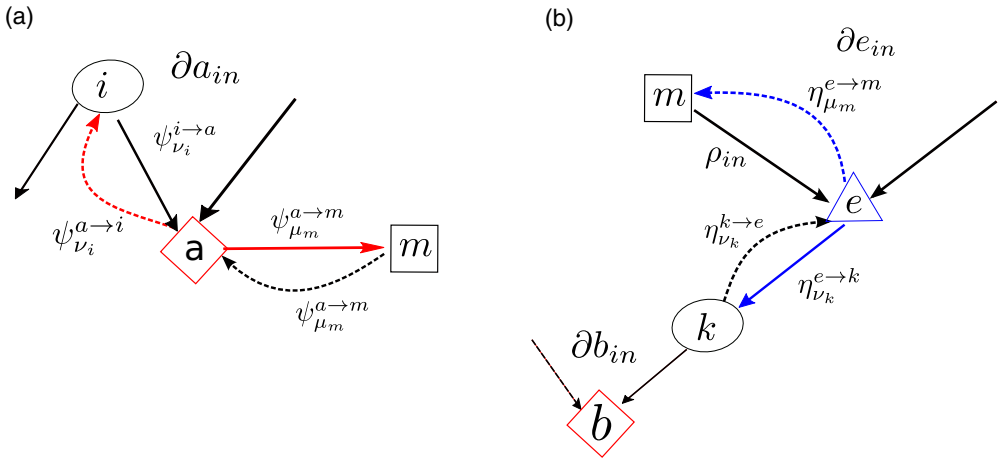

FIG. 13. (Color online) Representation of the external metabolites in our network. **A** is the product while **B** is the nutrient.

and, for the reaction constraints,

$$\eta^{i\to e}_{\nu_i} = \left(\prod_{b\in\partial i^M_{\rm in}} \psi^{b\to i}_{\nu_i}\right)^{\alpha} \prod_{b\in\partial i^M_{\rm out}} \psi^{b\to i}_{\nu_i}/Z^{i\to e},$$

$$\eta^{e\to i}_{\nu_i} = \left[\delta_{\nu_i,0} + e^{\theta}\delta_{\nu_i,1}\prod_{n\in\partial e^M}\eta^{n\to e}_1\right]/Z^{e\to i},$$

$$Z^{e\to i} = 1 + e^{\theta}\prod_{m\in\partial e^M}\eta^{m\to e}_1,$$

$$\eta^{m\to e}_{\mu_m} = \psi^{a\to m}_{\mu_m}\prod_{f\in\partial m^R\setminus e}\eta^{f\to m}_{\mu_m}/Z^{m\to e},$$

$$\eta^{e\to m}_{\mu_m} = \left[\eta^{i\to e}_0 + e^{\theta}\eta^{i\to e}_1\mu_m\prod_{n\in\partial e^M\setminus m}\eta^{n\to e}_1\right]/Z^{e\to m},$$

$$Z^{e\to m} = 2\eta^{i\to e}_0 + e^{\theta}\eta^{i\to e}_1\prod_{n\in\partial e^M\setminus m}\eta^{n\to e}_1.$$

It is important to note that nutrients are treated differently from the rest of the network. In fact, in Ref. [10] we decided that nutrients should not have an associated metabolite constraint while sinks have it (Fig. 13). Hence looking at constraint (2), it is immediately clear that in this setting if a nutrient is present, its neighboring reactions can be either active or not, whereas if a nutrient is absent, no neighboring reaction can function. Indeed, this is how nutrients are used in real networks.

2. Methods

a. Belief propagation algorithm

Belief propagation is an algorithm for efficient unbiased sampling of the solutions of a set of equations [14]. In a nutshell, in BP it is considered that each variable sends a message to its neighbors. This message represents the belief that the variable has about the state of its neighbors. The outcome of this algorithm is the BP-marginal for variables μ and ν.

It is worth noting that while in the complete case many different messages exist between the variables (see Appendix A1), in a mean-field approximation, the messages are the same for all neighbors and correspond to $\langle\mu_m\rangle$ and $\langle\nu_i\rangle$. Nevertheless, the functioning of the algorithm is similar in the two cases: first we generate a RRN with a given q and λ, then we initialize the messages (to a random value or to the last value computed) and we iterate the equations until convergence. Finally, for the complete problem (in mean field the BP marginal is equal to the marginal) at convergence it is possible to recover the marginals as

$$p(\mu_m) = \psi^{a\to m}_{\mu_m}\prod_{f\in\partial m^R}\eta^{f\to m}_{\mu_m}/Z^m,$$
$$p(\nu_i) = \eta^{e\to i}_{\nu_i}\left(\prod_{b\in\partial i^M_{\rm in}}\psi^{b\to i}_{\nu_i}\right)^{\alpha}\prod_{b\in\partial i^M_{\rm out}}\psi^{b\to i}_{\nu_i}/Z^i, \tag{A1}$$

where

$$Z^m = \sum_{\mu_m}\psi^{a\to m}_{\mu_m}\prod_{f\in\partial m^R}\eta^{f\to m}_{\mu_m},$$
$$Z^i = \sum_{\nu_i}\eta^{e\to i}_{\nu_i}\left(\prod_{b\in\partial i^M_{\rm in}}\psi^{b\to i}_{\nu_i}\right)^{\alpha}\prod_{b\in\partial i^M_{\rm out}}\psi^{b\to i}_{\nu_i}. \tag{A2}$$

All networks in this paper have $M = 10^4$ while $N = \lambda M/(1+q)$.

The simplest way to sample the solutions is by fixing one of the two free variables remaining: θ or $\rho_{\rm in}$. By changing $\rho_{\rm in}$ we can see how the configuration of the solutions changes when the nutrients have a probability $\rho_{\rm in}$ of functioning. Whereas by changing θ we can observe what happens if we constrain the system to switch on (or off) the reactions. Each behavior is important to understanding how the system is organized. In each case, the mean over the metabolites, $\langle\mu\rangle$, and the reactions, $\langle\nu\rangle$ [$\langle x\rangle$ is the average over the measure $P(\mu,\nu)$, (4)], has been computed.

b. Decimation procedure

The BP algorithm is an efficient way to obtain the *probability* that a variable will take a certain state. Nevertheless, one is generally confronted with the problem of obtaining actual *configurations* of variables that satisfy a CSP. To to find it, we resorted to a decimation procedure already used with success in other cases [16,17].

In decimation, first BP is run and then the BP marginal is used as the real marginal of the variable, thus setting the variable to 0 or 1 *according to the marginal*. Hence during decimation, variables are set one at a time, starting from the most polarized (with the BP-marginal near 0 or 1) and then running BP to make sure that the constraints are satisfied and that no contradiction occurs. This procedure is then iterated until all variables are decimated or until some constraint is violated.

Using this procedure, it is thus possible to obtain a Boolean configuration that is a solution of the CSP problem under study. It is important to note that while BP is an unbiased way of sampling the solution space (at least for problems on random graphs), the decimation process is highly dependent on the procedure used to decimate. Nevertheless, if the procedure converges, the configuration found will be a solution of the CSP. Furthermore, assuming BP marginals are unbiased for a RRN, it is possible to understand whether we are sampling fairly well the solution space with decimation.

To reproduce the behavior already observed in [10], the algorithm that we used to obtain the results presented in Figs. 5–8 is an extension of the standard decimation procedure presented above. In our algorithm, for a given θ, first a BP solution is found and stored, then the system is decimated N_{dec} times, each time starting from the same BP solution stored. Finally, the BP solution for the next θ is obtained by initializing the messages with the last stored BP solution. For each system under study, we applied this procedure following the two protocols ($+\infty \to -\infty$ or $-\infty \to +\infty$) presented in [10]. All the results in this article have been obtained with $N_{\mathrm{dec}} = 5$ for the complete problem and $N_{\mathrm{dec}} = 10$ for MF.

[1] D. A. Beard and H. Qian, *Chemical Biophysics: Quantitative Analysis of Cellular Systems* (Cambridge University Press, New York, 2008).

[2] B. O. Palsson, *Systems Biology: Simulation of Dynamics Network States* (Cambridge University Press, New York, 2011).

[3] T. Handorf, O. Ebenhöh, and R. Heinrich, J. Mol. Evol. **61**, 498 (2005).

[4] O. Ebenhöh, T. Handorf, and R. Heinrich, Genome Inf. **15**, 35 (2004).

[5] K. Kruse and O. Ebenhöh, Genome Inf. **20**, 91 (2008).

[6] T. Handorf and O. Ebenhöh, Nucl. Acids Res. **35**, W613 (2007).

[7] Z. Burda, A. Krzywicki, O. C. Martin, and M. Zagorski, Proc. Natl. Acad. Sci. (USA) **108**, 17263 (2011).

[8] P. François and V. Hakim, Proc. Natl. Acad. Sci. (USA) **101**, 580 (2004).

[9] L. Correale, M. Leone, A. Pagnani, M. Weigt, and R. Zecchina, Phys. Rev. Lett. **96**, 018101 (2006).

[10] A. Seganti, A. De. Martino, and F. Ricci-Tersenghi, J. Stat. Mech. (2013) P09009.

[11] B. O. Palsson, *Systems Biology: Properties of Reconstructed Networks* (Cambridge University Press, New York, 2006).

[12] L. Correale, M. Leone, A. Pagnani, M. Weigt, and R. Zecchina, J. Stat. Mech. (2006) P03002.

[13] A. Braunstein, M. Mézard, and R. Zecchina, Random Struct. Algorithms **27**, 201 (2005).

[14] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, 2nd ed. (Morgan Kaufmann, San Francisco, 1988).

[15] M. Mézard and G. Parisi, J. Stat. Phys. **111**, 1 (2003).

[16] F. Ricci-Tersenghi and G. Semerjian, J. Stat. Mech. (2009) P09001.

[17] A. Decelle and F. Ricci-Tersenghi, Phys. Rev. Lett. **112**, 070603 (2014).