PHYSICAL REVIEW E **78**, 035101(R) (2008)

Flux distribution of metabolic networks close to optimal biomass production

Ginestra Bianconi
The Abdus Salam International Center for Theoretical Physics, Strada Costiera 11, 34014 Trieste, Italy
(Received 10 April 2008; published 4 September 2008)

We study a statistical model describing the steady-state distribution of the fluxes in a metabolic network. The resulting model defined on continuous variables can be solved by the cavity method. In particular, analytical tractability is possible, solving the cavity equation over an ensemble of networks with the same degree distribution as a real metabolic network. The flux distribution that optimizes production of biomass has a fat tail with a power-law exponent independent of the structural properties of the underlying network. These results are in complete agreement with the flux-balance-analysis outcome and in qualitative agreement with the experimental results.

DOI: 10.1103/PhysRevE.78.035101 PACS number(s): 89.75.Hc, 89.75.Fb, 89.75.Da

Recently, much attention has been addressed by the physics community to critical phenomena [1] in complex networks [2–7]. Complex topologies, usually characterized by non-Poisson degree distributions, have a large effect on the critical point and critical exponents of the dynamical models defining them. The Ising model [8–10], the epidemic spreading [11], and synchronization dynamics [12] are examples of dynamics models, where the complex structure has strong implications. Furthermore, in the last decade we have been witness to a big breakthrough in system biology, the interdisciplinary field that studies the biological problems going beyond the single-biomolecule framework, with a description of the intertwined reactions between the constituents of the cell in terms of networks. This has generated a theoretical framework in which previously unknown biological statistical findings have been formulated [13–15]. In system biology there was also the fast development of "*in silico*" biology in which experiments are simulated and the predictions are made to stimulate further experimental confirmations of the phenomena. A key example of a biological system in which the network picture is crucial and the *in silico* biology has made relevant advances is the prediction of the growth rate of single cells of different organisms and the study of the metabolic networks. Two key advances in this field have been the full characterization of the chemical reactions [17] for a series of model organisms, as different strains of *Escherichia coli* and *Saccaromyces cerevisiae* (see, for example, the BIGG database [16]), and the application of the techniques of linear programming for the study of the reactions' fluxes, an extension which goes under the name of flux-balance analysis [17].

The set of stoichiometric interactions in the cell can be represented as a network whose nodes are of two types: the metabolic substrates of metabolites and the nodes representing the reactions. This bipartite network goes under the name of factor graph. In a factor graph, to each reaction i is assigned a flux variable s_i and to each metabolite μ is assigned a steady-state condition for the production or consumption of the metabolites. The structure of the metabolic network has a projection on the metabolites which has a power-law degree distribution [15] and a hierarchical structure [18,19]. In the metabolic networks, to each reaction corresponds an enzyme which regulates the rate of each reaction and modulates the flux of the reactions. Consequently, the maximal flux rate is fixed by the maximal enzyme concentration inside the cell. Solving the nonlinear mass-law equations is a difficult problem in networks of thousands of nodes. To overcome this problem in flux-balance analysis for each reaction a variable, its flux, is introduced. Each flux includes all dynamical effects associated with each reaction of the organism. The flux-balance analysis [17,20] considers the steady state of the dynamics which optimizes the production of the biomass by linear programming.

The underlying assumption of flux-balance analysis—i.e., the assumption that the cell organisms optimize the biomass production—is very well confirmed by experimental results [20]. Different variations of the algorithm have been considered [21,22], relaxing the condition of optimization of the biomass, which are able to predict the state of the metabolic fluxes for knockout mutants. Nevertheless, flux-balance analysis has been proven to be consistent also with knockout experiments in which the cell organisms have had the chance to adapt [23].

In this paper we will study the flux distribution in metabolic networks that has a heavy tail as found in experiments [24] and in flux-balance-analysis [25] predictions in *Escherichia coli*. In particular, we add to the description of the metabolic networks some theoretical statistical mechanics insights using the cavity method [1,26,27]. Different theoretical models have been already proposed [28–30] for the flux distribution, but neither of them has been able to theoretically predict the outcome of the experiments or of the flux-balance-analysis calculations. Here we will relate the power-law exponent of the flux distribution with the steady distribution as an indication of criticality. In fact, the state of optimal biomass production can be studied as a critical state between a phase of suboptimal viable states and a phase of nonviable states in which the organism cannot survive. Thus we will characterize this optimized state doing an asymptotic expansion of the cavity equation close to the critical point and we will measure the critical exponents corresponding to this critical transition. The method which we formulate here is a method to solve the cavity equation on continuous variables defined on a compact interval of the real axis, and it can be extended to other critical phenomena on complex networks and continuous variables defined in a limited interval. We find that the distribution of the fluxes present in the optimized state develops a power-law tail with exponents

that are independent of the structure of the underlying network. The power-law exponent that we find is in full agreement with the results of flux-balance analysis [25] and only partially in agreement with the experimental results [24].

The metabolic network has a bow-tie structure [19]: the metabolites are divided into input metabolites, which are provided by the environment, and output metabolites, which provide the biomass and intermediate metabolites. The stoichiometric matrix is given by $(\xi_{\mu,i})$ where $\mu=1,\ldots,M$ indicates the number of metabolites and $i=1,\ldots,N$ the number of reactions and the sign of $\xi_{\mu,i}$ indicates if the metabolite μ is an input or output metabolite of the reaction i. In the flux-balance-analysis method we assume that each intermediate metabolite has a concentration c^{μ} at steady state—i.e., $\dot{c}_{\mu}=\Sigma_j \xi_{j,\mu} s_i=0$, where s_i is the flux of the metabolic reaction i. For the metabolites present in the environment and the metabolites giving rise to the biomass production, we can fix the rate at which they are, respectively, consumed $g_{\mu}^{in}<0$ and produced $g_{\mu}^{out}>0$—i.e., $\dot{c}_{\mu}=\Sigma_j \xi_{j,\mu} s_i=g_{\mu}^{in/out}$. We have already mentioned that the fluxes have some biological limitations. To describe these limitations we assume that the fluxes $s_i \in [0,L]$; i.e., the reactions occur at rates smaller than L.

The volume of solutions V of this problem is given by

$$V=\int_0^L \cdots \int_0^L \prod_{i=1}^{N} ds_i \prod_{\mu} \delta\left(\sum_j \xi_{j,\mu} s_i - g_{\mu}\right). \tag{1}$$

In the following we use belief propation (BP) equations in order to fix the probability distribution of the metabolic fluxes with the measure defined in (1). Belief propagation equations are defined on cavity graphs. The cavity graph C_{μ} is the factor graph of the metabolic network in the absence of metabolite μ. The equations of BP in particular are exact in the case in which the graph has local treelike structure, but are known to give good results also in the presence of loops in expanders and in general in networks in which mean-field arguments give good results.

In the cavity graph C_{μ} the flux s_i of a reaction i in which μ is reacting has a *cavity distribution* $p_{i\to\mu}(s_i)$. The cavity distribution $p_{i\to\mu}(s_i)$ is the distribution of the flux of reaction i the cavity graph C_{μ}—i.e., in absence of metabolite μ. Expressing $p_{i\to\mu}(s_i)$ in terms of the cavity distribution $p_{j\to\nu}(s_i)$ (where ν is a neighbor of and i different from μ and j is a neighbor of ν different from i), we get the BP equations

$$p_{i\to\mu}(s_i)=\frac{1}{C_{i,\mu}} \prod_{\nu\in N(i)\setminus\mu} \prod_{j\in N(\nu)\setminus i} \left[\int ds_j p_{j\to\nu}(s_j)\right] \times \prod_{\nu\in N(i)\setminus\mu} \delta\left(\sum_j \xi_{j,\nu} s_j + \xi_{i,\nu} s_i - g_{\nu}\right). \tag{2}$$

Solving the BP equations for the cavity distributions, the marginal probability of a flux s_i is given by

$$p_i(s_i)=\frac{1}{C_i} \prod_{\nu\in N(i)} \prod_{j\in N(\nu)\setminus i} \left[\int ds_j p_{j\to\nu}(s_j)\right] \times \prod_{\nu\in N(i)} \delta\left(\sum_j \xi_{j,\nu} s_j + \xi_{i,\nu} s_i - g_{\nu}\right).$$

The distribution of the fluxes producing or consuming the metabolite μ—i.e., $\underline{S}_{\mu}=\{s_i\}_{i\in N(\mu)}$—is given by

$$p_{\mu}(\underline{S}_{\mu})=\frac{1}{C_{\mu}} \delta\left(\sum_j \xi_{j,\mu} s_j - g_{\mu}\right) \prod_{j\in N(\mu)} p_{j\to\mu}(s_j). \tag{3}$$

The entropy of the metabolic network can be expressed as $\Sigma=-\Sigma_{\mu}\int \Pi_{i\in N(\mu)} ds_i p_{\mu}(\underline{S}_{\mu}) \ln p_{\mu}(\underline{S}_{\mu}) + \Sigma_i (k_i-1) \int ds_i p_i(s_i) \ln p_i(s_i)$.

In order to get some analytic results we assume that as long as we want to predict the statistical properties of the flux distribution, the metabolic network can be modelled as a random graph with M metabolites with degree distribution $p(k)$ and N reaction nodes with degree distribution $p(q)$. In this network the total number of links is given by $N\langle q\rangle = M\langle k\rangle$. In Fig. 1 we show the of $p(k)$ and $p(q)$ distributions for *Escherichia coli*. The $p(k)$ degree distribution for this organism has a fat tail with a degree distribution that can be fitted with a power law $p(k)\sim k^{-\gamma}$ with an exponent $\gamma\simeq 3$, while the $p(q)$ distribution is much more peaked. In different organisms the distribution of $p(q)$ and $p(k)$ do change, but the general scenario of a fat-tail $p(k)$ distribution and finite-scale $p(q)$ distribution remains unchanged. These kinds of networks have many more short loops than Erdös and Renyi networks, but mean-field arguments are shown to work very well. Therefore we apply to the BP equations (2) to a random network with given $p(r)$ and $p(q)$ with distributions. The links $\xi_{i,\mu}$ have a random sign indicating a uniquely defined direction of each reaction. The results are easily generalizable also to include distributions for reversible reactions. We mimic the environment and the biomass production by a random assignment of the $g_{\mu}^{in/out}$ to each metabolite μ of the network. In particular, we choose draw the g_{μ} form a distribution $\rho(g)$ defined as

$$\rho(g)=p_1\delta(g+g_1)+p_2\delta(g-g_2)+(1-p_1-p_2)\delta(g), \tag{4}$$

where p_1 indicates the fraction of input metabolites and g_1 the rate at which input metabolites are consumed, while p_2 indicates the fraction of output metabolites and g_2 is the rate of biomass production.

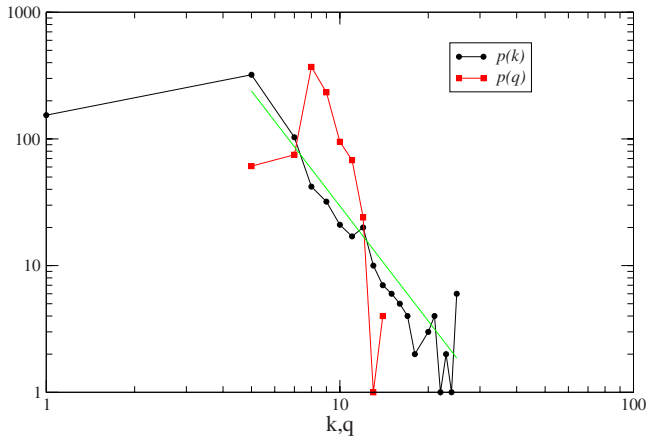

FIG. 1. (Color online) The degree distribution of $p(k)$ and $p(q)$ for *Escherichia coli*, data taken from the BIGG database [16]. The line indicates the power law $p(k)=k^{-\gamma}$ with $\gamma=3.0$.

To solve the BP equations (2) in this random ensemble of graphs, we introduce the function $P(s)$ indicating the average cavity distribution in the random ensemble

$$P(s) = \frac{1}{\langle q \rangle N} \sum_{i,\mu \in \partial i} \langle \overline{p_{i\to\mu}(s)} \rangle_{\rho(g)}, \tag{5}$$

where the overbar indicates the average over random sign of the $\xi_{i,\mu}$ and $\langle \cdots \rangle$ indicates the average over the probability distribution $\rho(g)$ of the g's. The Fourier transform of $P(s)$ is the function $\chi(w)$,

$$\chi(w) = \int ds\, e^{iws} P(s), \tag{6}$$

where $w = \frac{2\pi}{L} n$. Using Eqs. (2)–(5) we find that the function $\chi(w)$ satisfies the self-consistent equation

$$\chi(w) = \sum_q \frac{qp(q)}{\langle q \rangle} \prod_{\nu=1}^{q-1} \left\{ \sum_{k_\nu} \frac{k_\nu p(k_\nu)}{\langle k \rangle} \times \sum_{\{n_\nu\}} \sum_{\omega_\nu} \left[\frac{1}{2} [\chi(-\omega_\nu) + \chi(\omega_\nu)] \right]^{k_\nu - 1} \right\} \times \left\langle \frac{e^{-i\Sigma_\nu \omega_\nu g_\nu}}{C(\{k_\nu\})} \right\rangle_{\rho(g)} \delta\left(\sum_\nu \omega_\nu (-1)^{n_\nu} - w \right),$$

where $\langle \cdot \rangle$ indicates the average over the distribution of the incoming and outgoing fluxes g. The normalization constant $C(\{k_\nu\})$ is then given by

$$C(\{k_\nu\}) \simeq \prod_{\nu=1}^{q-1} \left\{ \sum_{\omega_\nu} \left[\frac{1}{2} [\chi(-\omega_\nu) + \chi(\omega_\nu)] \right]^{k_\nu - 1} \right\} \times \exp\left(-i \sum_\nu \omega_\nu g_\nu \right) \delta\left(\sum_\nu \omega_\nu (-1)^{n_\nu} \right). \tag{7}$$

The equation for $\chi(w)$ has solutions until the rate of biomass production $g_2 < G_c$, with G_c corresponding to the maximal allowed biomass production of the metabolic network. As a function of g_2, the metabolic network has a phase transition between viable values of biomass production and not viable values of the biomass production. The critical point is the point of optimal biomass production.

Close to the critical point of optimal biomass production $g_2 \simeq G_c$, we suppose that the distribution $P(s)$ develops a fat tail and that it can be expressed as

$$P(s) = s^{-\tau} \Phi(s |g_2 - G_c|^\sigma), \tag{8}$$

where Φ is a scaling function. The exponent σ determines the size of the critical window in which the flux distribution $P(s)$ preserves the fat tail consistently with the results of [25]. In the limit of large L we can assume that w (together with the ω_ν) is a continuous variable and we can perform an asymptotic expansion of $\chi(w)$:

$$\chi(w) = 1 - |w|^{\tau - 1} h(w/|g_2 - G_c|^\sigma). \tag{9}$$

We then solve the self-consistent equation for $\chi(w)$, Eq. (7), for the analytic distribution $p(q)$ and the distribution of the metabolites connectivity decaying like a power law $p(k) \sim k^{-\gamma}$. Close to the phase transition we have

$$C(\{k_\nu\}) = 1 + \sum_{\nu,\nu'} [A_1(g_\nu, g_{\nu'})(k_\nu - 1)(k_{\nu'} - 1) + A_2(g_\nu, g_{\nu'})(k_\nu - 1)(k_\nu - 2)(k_{\nu'} - 1)] - \sum_{\nu,nu',\nu''} A_3(g_\nu, g_{\nu'}, g_{\nu''})(k_\nu - 1)(k_{\nu'} - 1)(k_{\nu''} - 1), \tag{10}$$

where A_1, A_2, and A_3 are linear functions of g_ν, $g_{\nu'}$, and $g_{\nu''}$. If we develop (7) around the point $w = 0$, we get

$$\chi(w) = \sum_q \frac{qp(q)}{\langle q \rangle} \prod_{\nu=1}^{q-1} \left\{ \sum_{k_\nu} \frac{k_\nu p(k_\nu)}{\langle k \rangle} \sum_{\{n_\nu\}} \int d\omega_\nu \left[1 - (k_\nu - 1) \times |\omega_\nu|^{\tau - 1} (\mathrm{Re}\, h) + \frac{1}{2} k_\nu (k_\nu - 1) |\omega_\nu|^{2(\tau-1)} (\mathrm{Re}\, h)^2 \right] \right\} \times \left\langle \frac{\left(1 - i \sum_\nu \omega_\nu g_\nu\right)}{C(\{k_\nu\})} \right\rangle_{\rho(g)} \delta\left(\sum_\nu \omega_\nu (-1)^{n_\nu} - w \right). \tag{11}$$

Since the sums over the degrees k_ν are convergent, Eq. (11) can be written as

$$|w|^{\tau - 1} \{ h(x) - [\mathrm{Re}\, h(x)] \mathcal{C}_1(g_1, g_2) \} = w \mathcal{C}_2(g_1, g_2) + w^{2(\tau-1)} [\mathrm{Re}(h)]^2 \mathcal{C}_3(g_1, g_2), \tag{12}$$

with $x = w/|g_2 - G_c|^\sigma$ and with $\mathcal{C}_i$ linear functions of g_1 and g_2. Therefore proceeding as in other mean-field problems [31] we get the field exponents $\tau = 3/2$ and $\sigma = 2$ as long as the hypothesis of flux-balance analysis is satisfied. If the distribution $P(s)$ decays as a power law close to the optimal biomass production, also the distribution of the marginals $\hat{P}(s) = \sum_k \frac{p(k)}{\langle k \rangle} p_i(s)$ will decay with the same critical indices. The entropy goes like

$$\Sigma = |g_2 - G_c|^\alpha, \tag{13}$$

with $\alpha = \sigma(\tau - 1) = 1$.

Therefore in the physical range for each degree distribution $p(q)$ or $p(k_\nu)$ [15] the predicted power-law critical exponent for the flux distribution is $\tau = 3/2$, in good agreement with the flux-balance calculations [25]

In conclusion we have presented a statistical-mechanical approach to study the steady-state distribution of the fluxes in a metabolic network assuming optimization of the biomass. The analytic treatment finds a distribution of the fluxes which is a power law with a mean-field exponent $\tau = 3/2$ independent of the structure of the metabolic network. The method can be generalized to other critical phenomena defined on continuous variables on a finite interval, and work in this direction is in progress.

We acknowledge interesting discussions with Riccardo Zecchina. This work was supported by IST STREP GENNETEC Contract No. 034952.

[1] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, e-print arXiv:0705.0010.
[2] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[3] S. N. Dorogovtsev and J. F. F. Mendes, Adv. Phys. **51**, 1079 (2002).
[4] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).
[5] R. Pastor Satorras and V. Vespignani, *Evolution and Strucuture of the Internet: A Statistical Physics Approach* (Cambridge University Press, Cambridge, England, 2004).
[6] S. Boccaletti *et al.*, Phys. Rep. **424**, 175 (2006).
[7] G. Caldarelli, *Scale-Free Networks* (Oxford University Press, Oxford, 2007).
[8] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, Phys. Rev. E **66**, 016104 (2002).
[9] M. Leone, V. Vázquez, A. Vepsignani, and R. Zecchina, Eur. Phys. J. B **28**, 191 (2002).
[10] G. Bianconi, Phys. Lett. A **303**, 166 (2002).
[11] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001).
[12] T. Nishikawa, A. E. Motter, Y.-C. Lai, and F. C. Hoppensteadt, Phys. Rev. Lett. **91**, 014101 (2003).
[13] R. Sharan and T. Ideker, Nat. Biotechnol. **24**, 427 (2006).
[14] A. L. Barabasi and Z. N. Oltvai, Nat. Genet. **5**, 111 (2004).
[15] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabasi, Nature (London) **407**, 651 (2000).
[16] The BIGG database can be found at the webpage http://bigg.ucsd.edu/
[17] J. S. Edwards and B. O. Palsson, Proc. Natl. Acad. Sci. U.S.A. **97**, 5528 (2000).
[18] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A.-L. Barabási, Science **297**, 1551 (2002).
[19] R. Tanaka, Phys. Rev. Lett. **94**, 168101 (2005).
[20] J. S. Edwards, R. U. Ibarra, and B. O. Palsson, Nat. Biotechnol. **19**, 125 (2001).
[21] D. Segre, D. Vitkup, and G. M. Church, Proc. Natl. Acad. Sci. U.S.A. **99**, 15112 (2002).
[22] T. Shlomi, O. Berkman, and E. Ruppin, Proc. Natl. Acad. Sci. U.S.A. **102**, 7695 (2005).
[23] S. S. Fong and B. O. Palsson, Nat. Genet. **36**, 1056 (2004).
[24] M. Emmerling *et al.*, J. Bacteriol. **184**, 152 (2002).
[25] E. Almaas, B. Kovacs, T. Vicsek, Z. N. Oltvai, and A.-L. Barabási, Nature (London) **427**, 839 (2004).
[26] J. Yedida, W. T. Freeman, and Y. Weiss, in *Exploring Artificial Intelligence in the New Millennium*, edited by G. Lakemeyer and B. Nebel (Morgen Kaufmann, San Francisco, 2001), p. 239.
[27] M. Mézard and G. Parisi, Eur. Phys. J. B **20**, 217 (2001).
[28] A. De Martino, C. Martelli, R. Monasson, and I. P. del Castillo, J. Stat. Mech.: Theory Exp. (2007) P01006.
[29] G. Bianconi and R. Zecchina, Networks Heterog. Media **3**, 361 (2008).
[30] A. Braunstein, R. Mulet, and A. Pagnani, e-print arXiv:0709.0922.
[31] M. Marsili, P. De Los Rios, and S. Maslov, Phys. Rev. Lett. **80**, 1457 (1998).