Monotonicity, frustration, and ordered response: an analysis of the energy landscape of perturbed large-scale biological networks

G. Iacono, C. Altafini*

October 17, 2009

Abstract

For large-scale biological networks represented as signed graphs, the index of frustration measures how far a network is from a monotone system, i.e., how incoherently the system responds to perturbations. In this paper we find that the frustration is systematically lower in transcriptional networks than in signaling and metabolic networks. Interpreting this result in terms of energetic cost of an interaction, an erroneous or contradictory transcriptional action costs much more than a signaling/metabolic error, and therefore must be avoided as much as possible. Averaging over all possible perturbations, however, we also find that signaling/metabolic networks tend to undergo a phase transition to order in an energetic regime lower than for transcriptional networks, meaning that, in spite of the higher frustration, they can achieve a globally ordered response to perturbations even for moderate values of the strength of the interactions. Furthermore, an analysis of the energy landscape shows that signaling and metabolic networks lack energetic barriers around their global optima, a property also favouring global order. In conclusion, transcriptional and signaling/metabolic networks appear to have systematic differences in both the index of frustration and the transition to global order. These differences are interpretable in terms of the different functions of the various classes of networks.

*The authors are with SISSA-ISAS, International School for Advanced Studies, via Beirut 2-4, 34014 Trieste, Italy. Corresponding author: altafini@sissa.it

Author's summary Studying the properties of a dynamical system is fundamental for understanding which functions/tasks it is optimized for. Biological networks, such as transcriptional, signaling and metabolic networks, are essentially large scale dynamical systems with highly complex behavior. The majority of the kinetic details needed for a classical dynamical system investigation, e.g. for an analysis of the stability and robustness properties, are not accessible for such biological networks. However, by using a minimal representation of these networks, like a graph of signed interactions, it becomes possible to study other important dynamical properties, like the degree of frustration, which represents how coherently the system responds to perturbations. The aim of this paper is to use algorithms and methods inspired from statistical physics (such as the notions of gauge invariance, of mean field theory for heterogeneous networks, and of phase transition), in order to characterize several classes of large-scale biological networks in terms of their monotonicity, frustration level, and degree of ordered response. Representing biological networks as graphs of interactions and subsequently studying their properties is a first step towards a less descriptive and more quantitative approach to the study of large-scale complex dynamical systems.

1 Introduction

For complex systems such as biological networks, rather than a precise description of the dynamics, which requires a quantity of kinetic details rarely accessible in large scale systems, it is often more reasonable to use a minimal representation, such as a graph of interactions between the molecular variables of interest [12, 13, 14, 23] and perhaps a sign describing the functional mode of the pairwise interaction. Such graphical approaches have been extensively used in recent years to model transcriptional [20, 28], signaling [19, 21, 22] and metabolic networks [12]. Signed adjacency graphs have been investigated in several different contexts, such as economics [10, 24], social balance [1], and in the theory of frustrated spin systems [5, 8], see [34] for a survey. In particular, in [30] signed graphs are linked to the theory of monotone dynamical systems [29] and this last is used as a paradigm to explain the highly predictable and ordered response of biological systems in many cases. In a biological network, a response to a perturbation propagating incoherently through the network may result in an unpredictable or contradictory behavior of the system as well as in a waste of resources. From a dynamical point of view, a system never showing any such contradiction is called monotone [29, 30], see Supplementary Notes for a more rigorous definition. From a statistical physics perspective, the problem of determining when a dynamical system is monotone (or near monotone) is equivalent to checking when an Ising model with signed couplings has no (or little) frustration [30, 6, 18]. We know from the theory of Ising models that it is energetically favorable for neighbouring spins to be aligned when the coupling constant is positive and to be antialigned when it is negative. Frustration then appears when in a cyclic subgraph or in a fan-in node it is impossible to satisfy all of these pairwise relationships [30].

We have tested a number of large scale biological networks, and the frustration index we observe varies considerably according to the type of network analyzed: it is very low for gene regulatory networks and much higher for signaling and metabolic networks. In this paper we propose an interpretation of this different behavior based on the characteristic "energy" associated to the interactions of a graph, namely that the higher energetic content of a transcription event strongly disfavors incoherent or contradictory transcriptional orders. For the "cheaper" signaling and metabolic events, instead, such a tight control may not be required, especially since a higher frustration may induce a richer and more complex dynamical behavior.

If rather than focusing only on the energy of the optimal configuration (i.e., on the ground state of the system) we average the state of the system over all possible perturbations, we can observe that the more frustrated signaling/metabolic networks achieve "order" (i.e, tend to populate their global minimum of energy) in a range of coupling strengths which is lower than for the transcriptional networks, meaning that these networks tend to respond to perturbations as coherently as they can even for moderate values of energy. This behavior partially compensates for the higher frustration, which, as already mentioned, might be instrumental to the achievement of more complex dynamics than those required for the transcriptional networks. These last networks, on the other hand, only contain strong interactions and are therefore not concerned with the lower energetic regime. Coherently, they show a topological structure richer in tree-like subgraphs which disfavor the phase transition, and which are absent in the other classes of networks.

That signaling and metabolic networks may require a lower energetic content to experience a transition to ordered behavior is also confirmed by the structure of their energy landscapes which, unlike for the transctiptional networks, lack high and neat energetic barriers around the global optima, meaning that reconfiguration to the ground state can be easily achieved even at modest energies.

2 Results

The representation of a biological network as an n-dimensional signed adjacency matrix is given by a matrix $\mathcal{J}$ of elements $J_{ij} \in \{\pm J, 0\}$, $i, j = 1, \ldots, n$. $\mathcal{J}$ is assumed symmetric (i.e., the frustrated cycles we seek are in the underlying nonoriented graph), and with zero diagonal (i.e., no self-loops), see [30] and the Supplementary Notes for details on the formulation of the problem. Coherently with our choice of model, we assume that also the perturbations affecting the system are of unit magnitude in each component, $s_i \in \{\pm 1\}$, $i = 1, \ldots, n$. In correspondence of a vector $\mathbf{s} = \begin{bmatrix} s_1 & \ldots & s_n \end{bmatrix}$ of such signed perturbations, or "spin" variables, let us consider the "energy" function

$$h(\mathbf{s}) = -\frac{1}{2} \sum_{i,j=1}^{n} J_{ij} s_i s_j = -\frac{1}{2} \mathbf{s}^T \mathcal{J} \mathbf{s}, \tag{1}$$

which expresses the total cost associated to the perturbation $\mathbf{s}$. Assuming that all couplings of a network have the same strength, $|J_{ij}| = J$, the cost of each interaction depends on the sign of J_{ij}: for $J_{ij} > 0$ (activator) the aligned s_i, s_j spin configuration is more energetically favorable ($-J_{ij} s_i s_j = -J < 0$) than the antialigned one ($-J_{ij} s_i s_j = J > 0$) and viceversa for $J_{ij} < 0$. Of all 2^n possible spin assignments, those respecting monotonicity will be such that $s_i J_{ij} s_i > 0$ on each edge of the graph, i.e., those contributing to minimizing $h(\mathbf{s})$. A spin system is said frustrated when not all these constraints $s_i J_{ij} s_j > 0$ can be satisfied simultaneously by any assignment. Computing how far a given network is from being monotone corresponds to computing $\mathbf{s}$ that globally minimizes (1). It has been shown [6] that this is an NP-hard problem, equivalent to the MAX-CUT problem or, in terms of the Ising model, to computing the exact frustration index of the network [30, 2], call it δ. In [11] (see also Supplementary Notes for a quick recap), we proposed efficient heuristic algorithms providing fairly tight upper and lower bounds for δ of biological networks of the size of the thousands nodes. From the theory of monotone systems [29, 30], $\mathcal{J}$ is monotone if and only if there exists a diagonal signature matrix P_σ (i.e., a matrix having on the diagonal the vector σ of elements $\sigma_i \in \{\pm 1\}$) such that $\mathcal{J}_\sigma = P_\sigma \mathcal{J} P_\sigma$ has all nonnegative entries, see Lemma 2.1 in [29]. $\mathcal{J}_\sigma$ and $\mathcal{J}$ have different sign patterns but the same frustration index δ, as P_σ is a change of sign through a cut set of the graph of $\mathcal{J}$ and such "gauge transformations" [32] leave the sign of each cycle of the graph (and hence δ) unaltered.

Let us consider first as an illustrative example the yeast cell cycle network of [16], see Fig. 1. The number of negative signs on the original (symmetrized, without self-loops) adjacency matrix $\mathcal{J}$ is 10. However, a gauge transformation on the three nodes Cib1,2 Clb5,6 and Cln1,2 yields a $\mathcal{J}_\sigma$ with only 4 negative edges, which is a global optimum for the frustration index δ, see Fig. 1 (a). The presence of frustrated cycles in a network leads to a lack of coherence in the response of the system to perturbations. This can be observed in Fig. 1 (b), where the response of the yeast cell cycle and of a monotone network built on the same graph are compared. The behavior of the non-monotone cell cycle network is less predictable and potentially contradictory (see also Fig. S1 for analogous considerations on the simpler feedforward loop example [20]). It is then important to have an estimate of how close a true network is to being monotone i.e., frustration-free. Our algorithms allow to obtain a $\mathcal{J}_\sigma$ with as low as possible negative signs also for large scale networks. This number is typically close to δ, meaning that it is now much easier to localize on the graph of $\mathcal{J}_\sigma$ the potentially frustrated bonds (or, more properly, the frustrated cycles). Another consequence is that the candidate ground state for $\mathcal{J}_\sigma$ is now straightforward to identify, as it corresponds to the "all spins up" configuration, call it $\mathbf{1}$. Hence, the candidate ground state for the original $\mathcal{J}$ can be found reversing the gauge transformation: $\mathbf{s}_{ground} = P_\sigma \mathbf{1}$. Approximate values for the frustration index δ and for the corresponding energy minimum $h(\mathbf{s}_{ground}) = -\mathbf{s}_{ground}^T \mathcal{J} \mathbf{s}_{ground}$ not very far from the true ones

can therefore be computed.

Frustration in large scale biological networks For 8 large-scale biological networks, 4 transcriptional (*E.coli*, *Yeast*, *B.subtilis* and *Corynebacterium*), 2 signaling (*EGFR* and *Toll-like*) and two metabolic (*E.coli* and *Yeast* networks), we considered the corresponding signed graphs (see Tables S1-S2 for details on the networks and the Supplementary Notes for the construction procedure followed) and estimated δ through the algorithms mentioned above. The theory of signed graphs provides us with an upper bound on the frustration index (see Supplementary Notes), call it δ_{max}, which is a function only of the number of nodes, edges and cycles of the networks. The ratio $\delta/\delta_{\max}$, Fig. 2 (a), shows a marked difference between transcriptional networks and signaling/metabolic networks, with the former exhibiting a consistently lower level of frustration than the latter. The upper bound δ_{max}, however, disregards completely the topological structure and the sign arrangements of a network. To take into account also these parameters, we constructed a null-model of the networks, obtained by randomly reshuffling the signs of the edges, while maintaining the same number of positive and negative edges of the original graph, see Supplementary Notes for details. For the Z-score of this null model, a negative value means that the edges are arranged in order to decrease frustration. We can observe in Fig. 2 (b) that all the transcriptional networks have a negative Z-score, and only them (p-values of the Z-score in Table S5). The characteristic property of the transcriptional networks that enhances monotonicity is the tendency of many nodes to have a skewed distribution of signs in their edges, see Fig. 3. Up to a gauge transformation, in fact, highly asymmetric sign distributions correspond to highly positive sign concentrations, hence closer to monotone than random sign distributions. The "packing" of signs on certain nodes is primarily due to the mode of action of the transcription factors. Although dual role (i.e., both activator and repressor) transcription factors exist in both prokaryotes and eukaryotes [26, 17], most transcription factors seem to be playing only one role on their target genes. The nature of this single role is sometimes associated to the regulatory domains found on the proteins, especially for activator domains, which are usually enriched in proline, glutamine or acidic amino acid residues [9, 25, 31]. The dual role transcription factors are usually able to perform opposite functions according to possible different positions of their binding sequence with respect to the gene sequence, or according to different cellular contexts, or simply enhancing only the formation of the closed complex DNA-RNA polymerase [26]. For example, 71% of the *E.coli* transcription factors function only as activators or repressors, Fig. 3 (b). The ontological analysis of the dual role transcription factors is significantly enriched for categories such as interfacing the cell with its extracellular environment and for the elaboration of external stimuli (see Table S6). Hence mixing role transcription factors are more often mediating signaling events than their single role counterpart. It is shown in Fig. 3 (a) that all transcriptional networks (and only them) have sign arrangements on the edges that are more skewed than expected (with respect to a binomial distribution model, see Supplementary Notes and Table S7) and also this property contributes to their monotonicity (Fig. S3).

Another structural difference between transcriptional and signaling/metabolic networks is the overrepresentation in these last classes of short frustrated cycles. As explained in the Supplementary Notes, this characteristic is encoded in the level of detail (stoichiometric) that we choose to represent our networks, and expresses the lack of global monotonicity of a biochemical reaction involving multiple reagents, see also [30, 6, 18, 11].

Average frustration and ordered response The values of δ and $h(\mathbf{s}_{ground})$ alone are not enough to characterize how monotonically the system behaves *in average*. In fact, the energy landscapes of frustrated Ising spin systems are known to be usually rugged [4, 27], and the

presence of a single deep minimum in (1) is not enough to guarantee that the energy averaged over all configurations $\mathbf{s}$ (corresponding to all possible multinode perturbations) is indeed more negative than in other systems whose energy landscape is characterized by valleys which are maybe less deep but with larger basins. In other words, to characterize how monotone is the response of the system to arbitrarily complex perturbations we have to consider the average value that $h(\mathbf{s})$ assumes over all possible spin assignments, weighted by the probability of each $\mathbf{s}$. This "internal energy", call it $\langle h \rangle$, is an indicator of how coherently the system is behaving in average: the more negative $\langle h \rangle$ is, the less the responses of the system to perturbations are "contradictory" at some fan-in node or along oriented cycles. Denote with $Z(\beta) = \sum_{s_i \in \pm 1} e^{-\beta h(\mathbf{s})}$ the partition function of the system, $\beta \in \mathbb{R}$. For spin systems, β has the meaning of an inverse temperature. In the context of biological networks, the temperature is taken as $\sim 298\, K$ and it is not a varying parameter. However, we can use β to describe the strength of the coupling of a network. Recall that in forming the energy (1), $\mathcal{J}$ was taken as a signed adjacency matrix with all couplings equal to J, regardless of the nature of the network studied. As a matter of fact, metabolic, signaling and transcriptional interactions are characterized by widely different energetic costs. In particular, if a metabolic reaction or a signaling event might have a comparable energetic content, a link in a gene regulatory network describes the entire cascade of events in which the transcription of a gene can be broken down and overall its cost is much higher than in the other networks[1]. Hence, in our fixed temperature context, taking into account the coupling strength β (dimensionally still an inverse temperature) rescales $h(\mathbf{s})$ to the "absolute" energy $\beta h(\mathbf{s})$. The probability of a given configuration $\mathbf{s}$, $p(\mathbf{s}) = e^{-\beta h(\mathbf{s})} Z(\beta)^{-1}$, is a function of β and is maximized in the (usually degenerate) ground state $\mathbf{s}_{ground}$. Using β as a Lagrange multiplier, the expectation value of $h(\mathbf{s})$, defined as

$$\langle h \rangle = -\frac{\partial \ln Z(\beta)}{\partial \beta},$$

expresses this simultaneous weighting of the configurations by their degeneracy and energetic content. The more negative $\langle h \rangle$ is, the more we expect the system to respond coherently to a generic perturbation. For any $\beta > 0$, $\langle h \rangle > 0$ and, as β increases, $\langle h \rangle$ reaches a stationary value, see Fig. 1 (c). For spin systems, small values of β represent a regimen where thermal fluctuations are dominant and all states tend to be equally populated. As β increases, a spin system usually undergoes a phase transition characterized by the appearance of long range correlations. For our biological networks, when β (i.e., the energetic content of an edge of the network) is too small, the behavior of the network tends to be random (and all states $\mathbf{s}$ equiprobable) regardless of the monotonicity of the network, a clear obstacle to carrying out any meaningful task. On the other hand, when $\beta \to \infty$, the probability concentrates exclusively on the ground states ($Z(\beta)$ becomes a Dirac delta function) and the behavior of the system becomes as ordered as its frustration index allows, i.e., the system response is as coherent and coordinated as possible, regardless of the type of perturbation, see Fig. 1 (c). It is then important to see how the probabilities of the various states and the internal energy vary as a function of β on the various categories of networks under exam. Computing $p(\mathbf{s})$ and $\langle h \rangle$ exactly is impossible for systems larger than a few tens of nodes. For larger networks we shall make use of a mean field approximation for heterogeneous networks [15, 3]. This approximation, see the "Methods" Section for details, allows to estimate the mean field "magnetization" $\langle \mathbf{s}_\sigma \rangle$ in the gauge transformed basis, and the corresponding mean field energy h_{mf} (a gauge invariant). Fig. 4 shows the behavior of $\langle \mathbf{s}_\sigma \rangle$ and h_{mf} for a transcriptional, a signaling and a metabolic network as function of β. In all three cases, the population concentrates in the ground state

[1]The corresponding time constants are even more widely separated: $\sim$ *msec* for metabolic and signaling events, *sec* $\div$ *min* for transcription [7, 23].

when β grows, and, correspondingly, h_{mf} achieves its minimum, which results as negative as the frustration of the network allows. The true characteristic value of β at which each of the classes of networks should be computed is unknown, except for $\beta_{transcr} \gg \beta_{signal} \sim \beta_{metab}$. Interestingly, as β grows, the transcriptional network is slower to reach its energetic minimum than the other two networks, and likewise for the other 5 networks, see Table S4 and Fig. S5. We can interpret this result by saying that since $\beta_{transcr}$ is high, it is much less plausible for a transcriptional network to be operating in a regimen of low β than it is for signaling/metabolic networks. On the contrary, for these last two classes of networks, it is not unlikely to have couplings of medium-low strength. Hence it gets much more important that $\langle \mathbf{s}_\sigma \rangle \to 1$ even in correspondence of moderate values of β, because this helps in maintaining a coherent behavior in response to perturbations, as required in order to carry out correctly a biological task. This shift of the coherence barrier towards the low energetic regions is a consequence of the topology of the networks. In fact, as can be seen on Fig. 4, also the completely monotone networks built on the same graphs (blue curves) as well as other networks with random sign assignment to the same edges as our $\mathcal{J}$ (green curves) present the same characteristic patterns in spite of different δ. A feature behind this difference is the already mentioned overrepresentation of closed nonoriented cycles of short length in the structure of metabolic and signaling networks. Also the lower dispersion in the number of degree classes k in these networks contributes to the quick convergence of $\langle \mathbf{s}_\sigma \rangle$ to 1. However, the main reason behind the different thresholds for β is the presence or less of leaves in the graph. For example, the *E.coli* transcriptional network has 38% of the nodes that are not involved in any (nonoriented) cycle. Dropping these nodes, we obtain mean field plots in which the threshold for order is lower, and similar to those of the signaling/metabolic networks. All of our transcriptional networks have a high percentage of nodes that are leaves, much higher that the signaling/metabolic networks, see Table S2. The complete lack of feedback, characteristic of tree-like subnetworks, disfavours the achievement of a globally ordered behavior, which is instead maximized by short cycles, like the 3-node motifs of signaling/metabolic networks. This is expected from the theory of spin systems, where long-range correlations are more easily achieved in dense graphs than in sparse ones. Of course adding leaves to a graph does not change its monotonicity properties (a tree is always monotone).

Sampling the energy landscape Further information, from a different perspective, can be obtained studying the structure of the energy landscape of the different networks [4]. In order to have a picture of how this landscape looks like, we have applied our frustration minimization algorithms to uniformly distributed initial conditions and registered the local and global minima achieved in the process (see Fig. S2 and Table S3). Fig. 5 shows these distributions as a function of the relative Hamming distance. For the transcriptional network of *E.coli* and the *Yeast* metabolic network, the global minima are localized in a small region, while *EGFR* has two broader valleys of global minima. In all three cases, the global minima are surrounded by many local minima, thus confirming the ruggedness of the landscapes. As can be seen on Fig. S8, unlike *EGFR* and the metabolic network, the local minima of the transcriptional network of *E.coli* tend to have an energetic difference from the global ones which grows linearly with the distance. In addition, the separation between the well of global minima and its surroundings is much more neat in *E.coli* than in the other two networks, as can be seen on the Montecarlo trajectories of Fig. 6 and even more clearly on their average gradient. See also Figs. S7-S10 for analogous consideration on the remaining 5 networks. Overall, it appears that global and local minima in the transcriptional networks are separated by high and steep energetic barriers, while on the other networks there always exist low-energy routes between random spin configurations and global minima, possibly passing through low-energy local minima. This of course facilitates

the achievement of the ground state and the creation of global order even in a regime of moderate values of β.

3 Discussion

For a gene regulatory network, the couplings represent the cost of the entire action of transcription of a gene. This includes many important molecular steps, from the binding of a transcription factor to the promoter region of a target gene to the final release of the newly synthesized mRNA molecule. Energetically, such complex process is various times more relevant than a signaling event or a metabolic reaction. Also the corresponding time scales differ by several orders of magnitude [7, 23]. Hence it is natural to expect that in a transcriptional network the genes behave in concert and that the fraction of the gene-gene interactions that contribute to decreasing the energy is substantially larger than in a metabolic or signaling network, as a frustrated bond costs much more to the cell and its effect lasts much longer. In particular, frustrations manifest themselves on the cycles of the underlying nonoriented graph of the network as contradictory transcriptional orders. While changing the transcriptional commands is necessary to cope with e.g. different environmental conditions, encoding them as frustrated cycles can easily lead to unpredictable or erroneous dynamical behavior. Both the topology and the sign assignments to the nodes of the transcriptional networks contribute to achieve a degree of monotonicity which is higher than expected from null models. On the contrary, incoherent signaling or metabolic actions are energetically much less relevant than a single transcriptional event and can be easily tolerated by the cell, especially since nonmonotone patterns favour a richer dynamical behavior. While the level of detail at which we model our networks (functional for transcriptional networks, stoichiometric for signaling and metabolic networks, see Supplementary Notes) contributes to the systematic differences in the frustration index, other factors such as the tendency of the transcriptional network to have skewed sign distributions are also crucial in attaining a low frustration. It is interesting, then, to notice that the transcription factors violating this rule are primarily involved in the mediation of external signaling, rather than in regulatory or structural functions.

For spin systems, the tendency to satisfy pairwise all bonds grows when the temperature decreases, meaning that even distant spins become correlated, although in a frustrated Ising spin system all the conditions can never be satisfied simultaneously. In this paper, we consider the strength of the couplings as the key factor that determines the onset of "long range correlations". If we parametrize the networks by the coupling strength and study the probability of finding the system response in the ground state (i.e., in the state that maximizes monotonicity and minimizes the energy), as a function of this coupling strength, we observe that for signaling/metabolic networks it is higher than for transcriptional networks in the region of medium/low values of the couplings. This behavior, which is due to the topological structure of the networks and to the energy landscape it determines, could reflect the tendency of signaling/metabolic networks to attain long-range correlations and a globally ordered response in spite of the weaker energetic content of their interactions. As such, it helps maintaining coherence of the response in spite of the higher level of frustration of these networks (which, again, favors a richer dynamical behavior). For transcriptional networks, on the other hand, owing to the strong couplings, the regime of low energies is less important, hence tree-like motifs, which hinder the establishment of long-range correlations are abundant.

A Montecarlo investigation of the energy landscape of the networks [4, 33] suggests that transcriptional networks have a more funneled landscape than the other networks (at least around the global optima), with a single deep well of global minima delimited by high barriers, while in signaling and metabolic networks the optima are surrounded by local minima of comparable

energy. Order in these classes of networks is favored also by the lack of neat energetic barriers separating local and global optima, which enables the reconfiguration to the global optimum through low-energy paths.

In conclusion, we have observed that distinct classes of biological networks seem to be characterizable by different features when responding to perturbations. Transcriptional networks are less frustrated than expected and much less frustrated than signaling and metabolic networks, meaning that they admit highly coherent responses to perturbations. On the other hand, the signaling/metabolic networks have the ability to achieve an average ordered response in a lower range of coupling strengths than the transcriptional networks. We explain the first feature as the need to avoid as much as possible erroneous or contradictory transcriptional actions which would cost much more to the cell than analogous incoherent signaling/metabolic events. The second feature partially compensates for the higher frustration of these last networks, by lowering the interaction strength needed for a phase transition to order, and thereby ensuring the effectiveness of this reduced coherent behavior in an energetic range more critical for these classes of networks.

4 Methods

Mean field approximation in heterogeneous signed networks For a given signed network $\mathcal{J}$, apply first the gauge transformation P_σ required to minimize the overall number of negative signs on the edges, while maintaining the frustration index δ invariant. Denote then $k(1), \ldots, k(\xi)$ the ξ different degrees of the nodes of $\mathcal{J}_\sigma$, of probabilities $p_k(1), \ldots, p_k(\xi)$, and $\langle k \rangle$ the average degree of $\mathcal{J}$. The nodes having degree k will have a certain distribution of positive and negative edges. Let $k_{pn}(1), \ldots, k_{pn}(\xi)$ be the differences between positive and negative edges averaged over all nodes of each degree class. As after preprocessing with P_σ each node has more positive than negative edges, we are guaranteed that $k_{pn} \geqslant k/2$. Following [15], the self-consistency equation on each degree class k is given by

$$\langle \mathbf{s}_\sigma \rangle_k = \tanh\left(\beta k_{pn} J \langle u \rangle\right) \tag{2}$$

where

$$\langle u \rangle = \sum_k \frac{k_{pn} p_k}{\langle k \rangle} \tanh\left(\beta k_{pn} J \langle u \rangle\right) \tag{3}$$

is the effective "field magnetization" of each node from its neighboring nodes and the subindex σ in $\mathbf{s}$ indicates that the value is computed in the gauge transformed basis. The use of k_{pn} instead of the degree k corrects the equations (2)-(3) for the frustration of the system. In practice, for our preprocessed networks the number of negative signs is at most 20% (often much less), meaning that $k_{pn} \sim k$ for most degree classes. From (2) and (3), we have an expression for the mean field weighted with respect to the degree classes:

$$\langle \mathbf{s}_\sigma \rangle = \sum_k p_k \langle \mathbf{s}_\sigma \rangle_k$$

and, neglecting fluctuations around each $\langle s \rangle_k$, the mean field Hamiltonian is

$$h_{mf} = -n \sum_k k_{pn} p_k \langle \mathbf{s}_\sigma \rangle_k^2. \tag{4}$$

As $P_\sigma^2 = I$, the energy is invariant to the gauge transformation P_σ. In fact, from $\mathbf{s}_\sigma = P_\sigma \mathbf{s}$, we have $-\mathbf{s}^T \mathcal{J} \mathbf{s} = -(P_\sigma \mathbf{s})^T P_\sigma \mathcal{J} P_\sigma (P_\sigma \mathbf{s}) = -\mathbf{s}_\sigma^T \mathcal{J}_\sigma \mathbf{s}_\sigma$. Hence the mean field calculations for $\mathcal{J}_\sigma$ are valid also in the original $\mathcal{J}$. In addition, however, as $\mathbf{s}_{\sigma,ground} = \mathbf{1}$, in the gauge transformed system we have that, as $\beta \to \infty$, $\langle \mathbf{s}_\sigma \rangle \to 1$, a property which is in general not verified in the original basis (which will concentrate at its own ground state $\mathbf{s}_{ground} = P_\sigma \mathbf{1}$).

Acknowledgments The authors would like to thank K. Nakai for providing the transcriptional network of *B.subtilis*.

References

[1] T. Antal, P. L. Krapivsky, and S. Redner. Dynamics of social balance on networks. *Phys. Rev. E*, 72(3):036121, 2005.

[2] F. Barahona. On the computational complexity of Ising spin glass models. *J. Phys. A. Math. Gen.*, 15:3241–3253, 1982.

[3] G. Bianconi. Mean field solution of the Ising model on a Barabási-Albert network. *Phys. Lett. A*, 303(2-3):166 – 168, 2002.

[4] C. L. Brooks, J. N. Onuchic, and D. J. Wales. Statistical thermodynamics. Taking a walk on a landscape. *Science*, 293:612–613, Jul 2001.

[5] D. Chowdhury. *Spin Glasses and other frustrated systems*. Princeton University Press, 1986.

[6] B. DasGupta, G. A. Enciso, E. Sontag, and Y. Zhang. Algorithmic and complexity results for decompositions of biological networks into monotone subsystems. *Biosystems*, 90(1):161–178, 2007.

[7] D. Fell. *Understanding the control of metabolism*. Portland Press, 1997.

[8] K. Fischer and H. J.A. *Spin Glasses*. Cambridge University Press, 1991.

[9] K. Han and J. L. Manley. Transcriptional repression by the drosophila even-skipped protein: definition of a minimal repression domain. *Genes Dev*, 7(3):491–503, Mar 1993.

[10] F. Harary. Graph theoretic methods in the management sciences. *Management Sci.*, 5:387–403, 1959.

[11] G. Iacono, F. Ramezani, N. Soranzo, and C. Altafini. Determining the distance to monotonicity of a biological network: a graph-theoretical approach. *preprint*, 2009.

[12] N. Jamshidi and B. Palsson. Formulating genome-scale kinetic models in the post-genome era. *Mol. Syst. Biol.*, 4:171, 2008.

[13] B. N. Kholodenko. Cell-signalling dynamics in time and space. *Nat. Rev. Mol. Cell Biol.*, 7:165–176, Mar 2006.

[14] Y. K. Kwon and K. H. Cho. Coherent coupling of feedback loops: a design principle of cell signaling networks. *Bioinformatics*, 24(17):1926–1932, 2008.

[15] M. Leone, A. Vázquez, A. Vespignani, and R. Zecchina. Ferromagnetic ordering in graphs with arbitrary degree distribution. *Eur. Phys. J. B*, 28(2):191–197, 2002.

[16] F. Li, T. Long, Y. Lu, Q. Ouyang, and C. Tang. The yeast cell-cycle network is robustly designed. *Proc Natl Acad Sci U S A*, 101(14):4781–4786, Apr 2004.

[17] J. Ma. Crossing the line between activation and repression. *Trends Genet*, 21(1):54–59, Jan 2005.

[18] A. Ma'ayan, R. Iyengar, and E. Sontag. Proximity of intracellular regulatory networks to monotone. *IET Systems Biology*, 2:103–112, 2008.

[19] A. Ma'ayan, S. L. Jenkins, S. Neves, A. Hasseldine, E. Grace, B. Dubin-Thaler, N. J. Eungdamrong, G. Weng, P. T. Ram, J. J. Rice, A. Kershenbaum, G. A. Stolovitzky, R. D. Blitzer, and R. Iyengar. Formation of regulatory patterns during signal propagation in a Mammalian cellular network. *Science*, 309:1078–1083, Aug 2005.

[20] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs: simple building blocks of complex networks. *Science*, 298(5594):824–827, 2002.

[21] K. Oda and H. Kitano. A comprehensive map of the toll-like receptor signaling network. *Mol. Syst. Biol.*, 2:2006.0015, 2006.

[22] K. Oda, Y. Matsuoka, A. Funahashi, and H. Kitano. A comprehensive pathway map of epidermal growth factor receptor signaling. *Mol Syst Biol*, 1:2005, 2005.

[23] J. A. Papin, T. Hunter, B. O. Palsson, and S. Subramaniam. Reconstruction of cellular signalling networks and analysis of their properties. *Nat Rev Mol Cell Biol*, 6(2):99–111, Feb 2005.

[24] J. Quirk and R. Ruppert. Qualitative economics and the stability of equilibrium. *Rev. Econ. Stud.*, 32:311–326, 1965.

[25] V. A. Rhodius and S. J. Busby. Positive activation of gene expression. *Curr Opin Microbiol*, 1(2):152–159, Apr 1998.

[26] S. Roy, S. Garges, and S. Adhya. Activation and repression of transcription by differential contact: two sides of a coin. *J Biol Chem*, 273(23):14059–14062, Jun 1998.

[27] H. Seyed-allaei, H. Seyed-allaei, and M. R. Ejtehadi. Energy-landscape networks of spin glasses. *Physical Review E*, 77(3):031105, 2008.

[28] S. S. Shen-Orr, R. Milo, S. Mangan, and U. Alon. Network motifs in the transcriptional regulation network of Escherichia coli. *Nat. Genet.*, 31(1):64–68, 2002.

[29] H. L. Smith. Systems of ordinary differential equations which generate an order preserving flow. A survey of results. *SIAM Review*, 30(1):87–113, 1988.

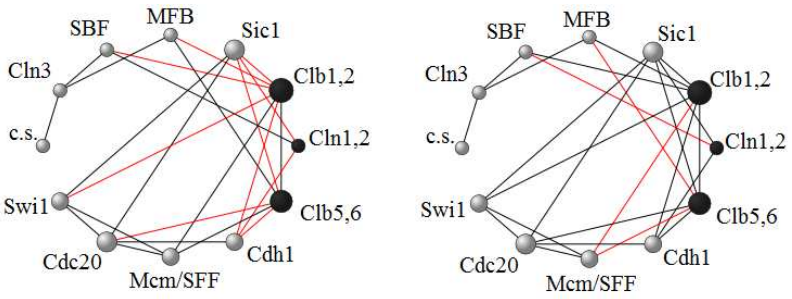
[30] E. D. Sontag. Monotone and near-monotone biochemical networks. *Systems and Synthetic Biology*, 1:59–87, 2007.

[31] A. Stepchenko and M. Nirenberg. Mapping activation and repression domains of the vnd/nk-2 homeodomain protein. *Proc Natl Acad Sci U S A*, 101(36):13180–13185, Sep 2004.

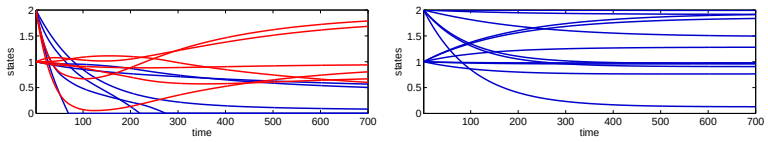
[32] G. Toulouse. Theory of the frustration effect in spin glasses : I. *Communications on Physics*, 2:115, 1977.

[33] J. Wang, B. Huang, X. Xia, and Z. Sun. Funneled landscape leads to robustness of cell networks: yeast cell cycle. *PLoS Comput. Biol.*, 2:e147, Nov 2006.

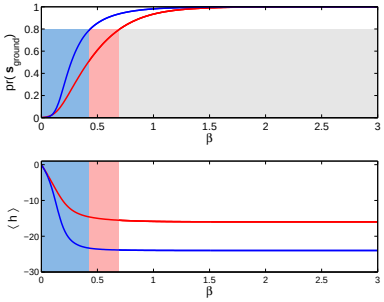
[34] T. Zaslavsky. Bibliography of signed and gain graphs. *Electr. J. Combinatorics*, DS8, 1998.



(a) original and gauge transformed network

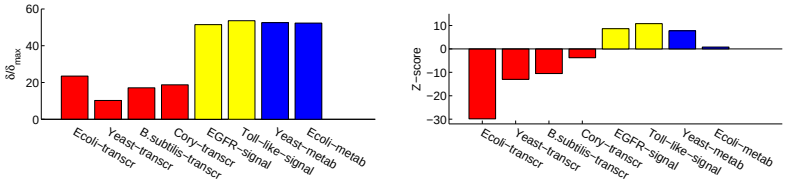


(b) simulation for original and monotone network



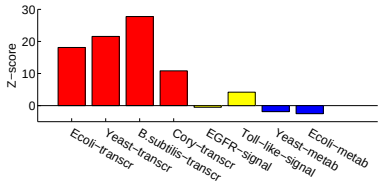
(c) ground state probability and average internal energy

Figure 1: Yeast cell cycle signed network of [16]. In (a) the application of a gauge transformation to the three nodes in black reduces the number of negative edges (in red) and $\mathbf{s}_\sigma = \mathbf{1}$ becomes a ground state. In (b) a typical response to a perturbation is shown for the yeast cell cycle network and for a monotone network on the same graph: in the second system the order is always maintained in the response (blue trajectories are monotone states). In (c) the probability of being in a ground state (upper plot) and the average internal energy $\langle h \rangle$ (lower plot) are shown for the yeast cell cycle network (red) and for the monotone network (blue) as a function of β. The monotone network achieves order (here $p(\mathbf{s}_{ground}) > 0.8$) earlier with growing β and the energy minimum reached is lower.

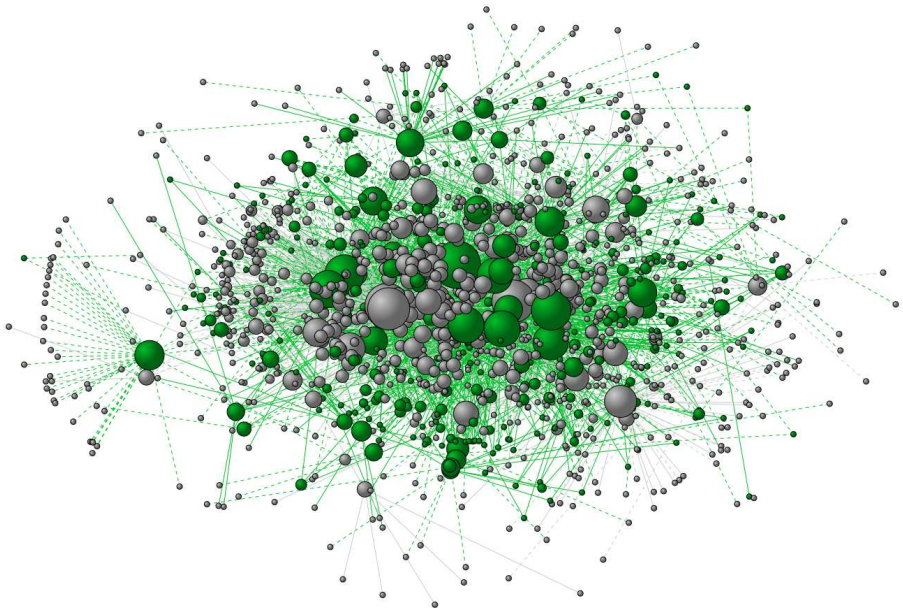


(a) original network (b) Z-score

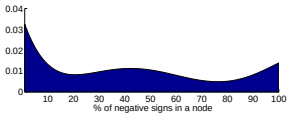
Figure 2: Frustration index of the 8 biological networks used in this study. In (a) the ratio δ/δ_{max} is based only on the number of nodes and edges of a network and shows that the frustration index is much lower for transcriptional than for signaling/metabolic networks. The Z-score in (b) takes into account also the topology of a network. Again, the transcriptional networks are more monotone (i.e., less frustrated) than expected from a null model, while metabolic and in particular signaling are less monotone (i.e., more frustrated) than expected.



(a) Z-score for the sign packing index

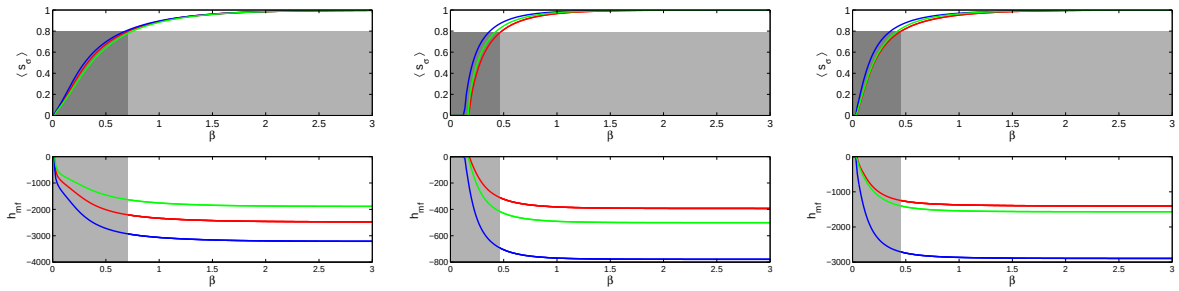


(b) Sign arrangement in *E.coli-transcr*



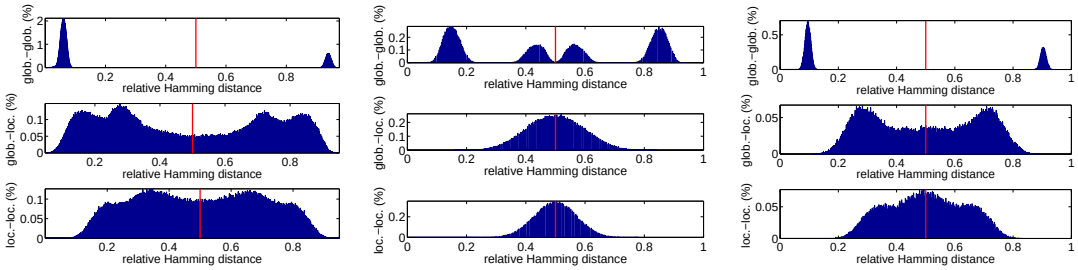
(c) Distribution of signs in *E.coli-transcr*

Figure 3: (a) Z-score for the sign packing index (see Supplementary Notes for a definition). The 4 transcriptional networks have sign arrangements on the nodes that are significantly asymmetric, hence improving their frustration index. (b) Representation of the sign packing property on the *E.coli* transcriptional network. The nodes significantly enriched in either positive or negative edges are shown in green (the size is proportional to their connectivity). The distribution of negative edges (dashed) is shown in (c). This graph should be compared with the random sign assignment of Fig. S4.



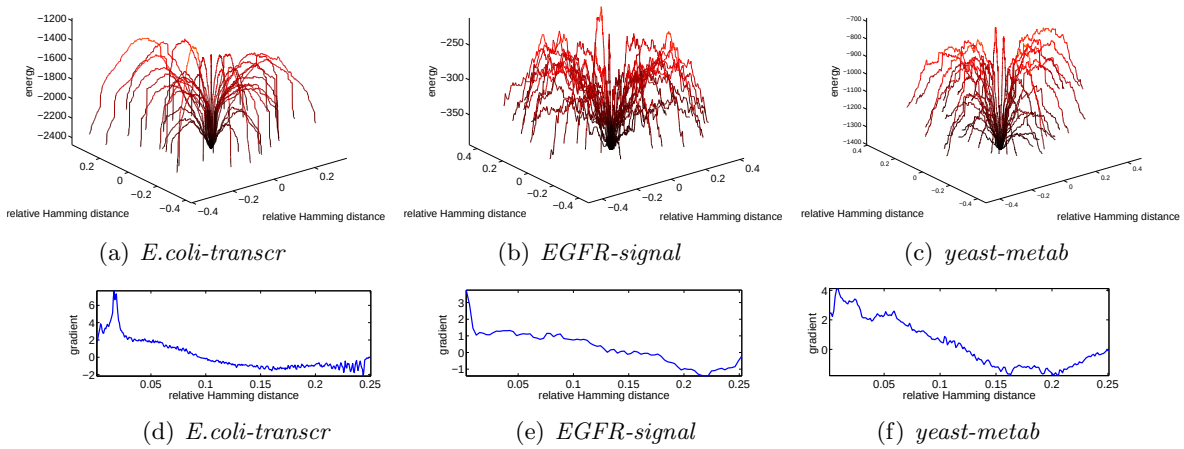
(a) *E.coli-transcr* (b) *EGFR-signal* (c) *yeast-metab*

Figure 4: Computation of the mean field "magnetization" $\langle \mathbf{s}_\sigma \rangle$ (in the gauge transformed basis) and energy h_{mf} for a transcriptional (*E.coli-transcr*, left panel), a signaling (*EGFR-signal*, middle), and a metabolic (*Yeast-metab*, right) network as a function of β (coupling strength). The values for the three true networks are depicted in red. In blue and green the same $\langle \mathbf{s}_\sigma \rangle$ and h_{mf} for two alternative networks built on the same graph: the exactly monotone network (i.e., with all $J_{ij} > 0$), and a network with random sign assignments. The gray shaded areas in the upper plot delimit the region $\langle \mathbf{s}_\sigma \rangle \leqslant 0.8$, i.e., the region in which the response of the system to a generic perturbation results in a low-medium $\langle \mathbf{s}_\sigma \rangle$. $\langle \mathbf{s}_\sigma \rangle \geqslant 0.8$ means that in the gauge transformed basis the state $\mathbf{s}_\sigma$ "concentrates" sufficiently well at the ground state, and, correspondingly, the energy is in average sufficiently close to the minimum (lower plots). For the *E.coli-transc* network the threshold $\langle \mathbf{s}_\sigma \rangle \geqslant 0.8$ is achieved in correspondence of $\beta = 0.71$, higher than the $\beta = 0.46$, 0.45 of *EGFR-signal* and *Yeast-metab*. Similar differences are observed in the other networks, see Table S4 and Fig. S5, and are also confirmed in Metropolis-Montecarlo simulations, see S6. The difference between transcriptional and signaling/metabolic networks expresses a tendency of these last to attain the highest possible $\langle \mathbf{s}_\sigma \rangle$ even for moderate values of β, motivated by the necessity of these networks to maintain order in the response to perturbations in a region of coupling strengths lower than for transcriptional networks (where the "true" β is typically much higher).



(a) *E.coli-transcr* (b) *EGFR-signal* (c) *yeast-metab*

Figure 5: Relative Hamming distance (number of spin flips over the number of nodes) between pairs of minima found by the algorithms for a transcriptional, *E.coli-transc*, (a), a signaling, *EGFR-signal*, (b), and a metabolic, *Yeast-metab*, (c), network. The top plots refer to pairs of global minima; the middle row to pairs global-local minima and the bottom row to pairs of local minima. In all three networks the red line delimits the global symmetry axis of the spin assignment (the locations of the minima have a global spin flip symmetry; the different height of the peaks means that an area has been explored less by the random searches of the algorithm, not that they have different "probabilities"). While for *E.coli-transcr* and *Yeast-metab* the minima are concentrated in a single well, which is quite tight and located near the right margin of the histograms (i.e., short interminimum distances), in *EGFR-signal* there are two such wells and they are disjoint and quite broad. In all 3 networks, adding the local minima, the landscape of minima becomes diffuse, with many different local minima located at varying distances from the global ones.



(a) *E.coli-transcr* (b) *EGFR-signal* (c) *yeast-metab*

(d) *E.coli-transcr* (e) *EGFR-signal* (f) *yeast-metab*

Figure 6: Montecarlo trajectories connecting a global minimum to its surrounding local minima. The spin configurations of a global and a local minimum are randomly chosen among those provided by our minimization procedure. The first is mapped in the second by a number of moves (single spin flips) equal to the Hamming distance between the two minima. For visualization purposes, the trajectories are depicted as emanating from a unique point and radially distributed according to a polar coordinate. The vertical axis (and color code) represents the energy, the two remaining axes a relative Hamming distance between spin configurations. The three plots essentially confirm the landscape described in Fig. 5. For *E.coli-transcr*, global and local minima seem to be always separated by a high and steep barrier. In *EGFR-signal* and *Yeast-metab*, the landscape is scattered with different local minima, many of which have energies similar to the global ones, see Fig. S8. This results in some trajectories never emerging from the ordered phase while moving from a minimum to the optimal frustration. The lower row shows the average gradient over 1000 Montecarlo trajectories originating in a global minimum. For *E.coli-transcr* the barriers of the well of global optima is precisely observable in correspondence of the peak of the gradient. For *Yeast-metab* such kinetic traps are less steep. For *EGFR-signal* no clear boundary at all is observable. This, together with Fig. S8, suggests that in the last two networks also spin configurations that are distant from the optimum have cheap routes to converge to the optimal frustration through intermediate low-energy local minima.