Biological Regulatory Networks are Minimally Frustrated

Shubham Tripathi,[1, 2, 3] David A. Kessler,[4, *] and Herbert Levine[2, 3, †]

[1] *PhD Program in Systems, Synthetic, and Physical Biology, Rice University, Houston, Texas 77005, USA*
[2] *Center for Theoretical Biological Physics, Rice Univeristy, Houston, Texas 77005, USA*
[3] *Department of Physics, Northeastern Univeristy, Boston, Massachusetts 02115, USA*
[4] *Department of Physics, Bar-Ilan Univeristy, Ramat-Gan 52900, Israel*

(Dated: November 26, 2019)

Characterization of the differences between biological and random networks can reveal the design principles that enable the robust realization of crucial biological functions including the establishment of different cell types. Previous studies, focusing on identifying topological features that are present in biological networks but not in random networks, have, however, provided few functional insights. We use a Boolean modeling framework and ideas from spin glass literature to identify functional differences between five real biological networks and random networks with similar topological features. We show that minimal frustration is a fundamental property that allows biological networks to robustly establish cell types and regulate cell fate choice, and this property can emerge in complex networks via Darwinian evolution. The study also provides clues regarding how the regulation of cell fate choice can go awry in a disease like cancer and lead to the emergence of aberrant cell types.

arXiv:1911.10252v1 [q-bio.MN] 22 Nov 2019

Biological regulatory networks establish cell type-specific gene expression patterns [1] and regulate cell fate-choice in response to various signals. These networks present a contradiction analogous to the famed Levinthal paradox in protein folding [2]. Networks as large and complex as those regulating cell fate typically exhibit a huge number of stable states [3]. Each such stable state or collection of stable states with a reasonably shared pattern of gene expression represents a cell type [4, 5]. This relationship, however, predicts a number of cell types much larger than that seen in multicellular organisms. A smaller number of cell types can be attained via the evolutionary fine-tuning of network parameters [6] or by putting cells through a precise sequence of events during development [7]. In both scenarios, cell fate will be highly sensitive to intra- and extra-cellular perturbations, an undesirable property.

Features that distinguish biological regulatory networks from random networks may provide a clue regarding how these networks can robustly establish the smaller than expected number of cell types. Biological networks have been shown to often exhibit a scale-free degree distribution [8] which might allow these networks to define topologically stable cell types [9]. Regulatory networks in *Escherichia coli* and *Saccharomyces cerevisiae* have been shown to be hierarchically organized [10]. Certain network patterns, called motifs, are known to recur far more frequently in biological networks than in random networks [11], and often mediate cell fate choice [12]. However, these investigations of topological differences between biological and random networks have provided few insights into how the functional characteristics of biological networks differ from those of random networks and allow biological networks to establish cell types. In this Letter, we compare the dynamical behavior of biological networks with that of random networks which have similar topological features and observe some remarkable differences. These could hold the key to elucidating the design principles that allow biological regulatory networks to carry out their biological functions.

Boolean modeling of biological networks.— A Boolean modeling framework [13] has proven useful for characterizing the behavior of large networks in cases where the use of ordinary differential equations-based modeling frameworks becomes challenging due to the numerous and hard to estimate kinetic parameters involved. In this framework, the only knowledge required is whether each regulatory relationship between network nodes is activating or inhibitory. The state of a N-node network in such a framework may be specified via a sequence $\{s_i\}$ of N binary variables; $s_i = \pm 1$. When modeling a biological regulatory network, each network node represents a molecular species such as a transcription factor or micro-RNA. If species i (the molecular species represented by node i) is highly expressed, $s_i = +1$, otherwise $s_i = -1$. Regulatory relationships between molecular species are specified by a $N \times N$ matrix J where $J_{ij} = +1$ of species j promotes the expression of species i and $J_{ij} = -1$ if species j inhibits the expression of species i. The absence of any regulatory relationship between species i and species j is indicated by $J_{ij} = 0$. The discrete-time network dynamics can then be simulated using [14]

$$s_i(t+1) = \begin{cases} +1 & \sum_j J_{ij} s_j > 0 \\ -1 & \text{if } \sum_j J_{ij} s_j < 0 \\ s_i(t) & \sum_j J_{ij} s_j = 0 \end{cases} \quad (1)$$

The network state is updated asynchronously, i.e., at each discrete time point, a network node is chosen at random and its state updated using Eq. (1). Clearly, a state $\{s_i\}$ is a stable state of the network if s_i is a fixed point of Eq. (1) for all i.

Note that the dynamical behavior of a network in