ACTA PHYSICA DEBRECINA XLVI, 47 (2012)

MODELLING GENE REGULATION WITH BOOLEAN NETWORKS

E. Fenyvesi[1], G. Palla[2]

[1] University of Debrecen, Department of Experimental Physics, 4032 Debrecen, Egyetem 1, Hungary
[2] Statistical and Biological Physics Research Group of HAS (Pázmány P. stny. 1/A, H-1117 Budapest, Hungary)

Abstract

Due to the changing environment, living cells must continously adapt their protein synthesis to react to the various inputs coming from their surrounding. A key element in the protein synthesis is the transcription of DNA to RNA. The so called regulatory proteins can catalyze or block the transcription of other proteins, thus help the cell in sensing the environment and regulating the rate of transcription of structural genes. A very simple model of the above process is provided by Boolean networks, where the state of each protein (gene) is either "on" or "off", and the interaction between proteins is represented by a directed graph. Here we overview the motivation and settings of this model and highlight possibilities for further research with a special focus on applying the latest results of control theory to Boolean networks.

I. Introduction

Proteins are essentially important ingredients of cell structure, catalising biochemical reactions, transporting different types of molecules, and taking part in a number of other vital processes. Twenty kinds of amino acids are

available for building up proteins. Each protein is consisting of hundreds of amino acids. All information that is needed to create a protein, is encoded in the DNA with the help of four bases (A, T, C, G). These bases form two complementer pairs (A-T, C-G), which follow each other along the double helix of DNA. Each amino acid is encoded with a basis triplet. (A doublet wouldn't be enough, since only $4^2 = 16$ different pairs can be constituted from two bases.) All the triplets that encode a given amino acid constitute a gene.

Transcription is the process of creating a complementary RNA copy of a sequence of DNA. During transcription, a DNA sequence is read by an RNA polymerase, which produces a complementary, antiparallel RNA strand. As opposed to DNA replication, transcription is resulting in an RNA complement that includes uracil (U) in all instances where thymine (T) would have occurred in a DNA complement. If the gene transcribed encodes a protein, the result of transcription is messenger RNA (mRNA), which will then be used to create that protein via the process of translation [6].

Regulatory proteins help the cell sense the environment and regulate the rate of transcription of structural genes (in this case regulatory proteins are called transcription factors). The genes encoding the regulatory proteins are referred to as regulatory genes. There are two types of regulatory proteins, negative-acting and positive-acting proteins. Transcription factors bind to either enhancer or promoter regions of DNA adjacent to the genes that they regulate, but they also use a variety of other mechanisms for the regulation of gene expression [6].

The interaction between genes through transcription factors can be represented by a network, in which nodes correspond to genes, and a directed link between a pair of genes represents interaction [1]. In this paper we overwiev the simplest model for gene regulation proposed by Boolean networks. In Sect. II. we discuss the motivation for representing gene regulatory networks with Boolean networks, whereas in Sect. III we overwiev of general properties of Boolean networks. Finally in Sect. IV we depict further ideas for research in this direction.

II. Modelling gene regulatory networks

There were many efforts to describe the topology and dynamics of gene regulatory networks. One of the possible methods is examining the cell in action through measurment. Then we can make statements of the gene regulatory network by using methods of statistical physics [1]. Another method is modelling the topology and dynamics of the network, testing the model by computer simulations, and then comparing the results of the simulations with available measurement data.

We usually model gene regulatory networks with directed graphs. Nodes represent genes, and directed edges show how a given gene acts on another gene. Topology has a large effect on the behaviour of the network. The quantities used to describe the topology a network in general are the distribution of in- and out-degrees, in- and out-degree correlations, assortativity, occurence of loops and different kinds of small subgraphs, community structure [3, 4]. In order to study the effect of the topology on the dynamics we usually have to examine these features one by one.

In order to take into account the time-dependent nature of the gene transcription, we assign a time-dependent state, $x_i(t)$ to each node. The value of this variable can correspond to, e.g., the transcription factor concentration [1]. In this case, the quantity changes continously in time.

Linear dynamics is the simplest continous dynamics [2]. Most real systems are driven by nonlinear processes, but the controllability of most nonlinear systems is in many aspects structurally similar to that of nonlinear systems, prompting us to start our study using the canonical linear, time-invariant dynamics [3]. To simplify the discussion further, we suppose that a gene can be in one of two states: allowed (turned on), when $x_i(t) = 1$, or blocked (turned off), when $x_i(t) = 0$. In this case we have a Boolean network [5]. A Boolean network can be generalized by introducing more than two discrete states ($x_i(t) \in 0, 1, 2, \ldots n$) [4]. We denote the state wector of the entire graph by $x(t) = (x_1(t), x_2(t), \ldots, x_N(t))$, where N is the number of nodes in the graph.

In the next step we choose the dynamical rules governing the behaviour of the network in time. A cell interacts with its environment, so the behaviour

of the gene regulatory network can be controlled from outside of the cell. Our goal is to find how to control a gene regulatory network. Moreover, small fluctuations can occur in gene expression pattern and in chemical reactions inside the cell. We would also like to learn if these fluctuations die out, or grow quickly. If these die out, the cell remains in its original state, but if these grow, the cell will turn into another state. We aim at understanding how stable is the behaviour of a gene regulatory network against small fluctuations [4].

We can see that examining the dynamical stability and controlability of a model network is essentially important in order to have more knowledge of real gene regulatory networks. Additionally, this knowledge can be applied to other systems, which can be described with the same network model and are governed by similar dynamical rules.

III. Boolean networks

Deterministic Boolean networks are formally defined by a state vector $x(t) = (x_1(t), x_2(t), \dots x_N(t))^T$, where $x_i \in \{0, 1\}$, and a set of update functions f_i such that $x_i(t) = f_i(x_{j_{i;1}}(t-1), x_{j_{i;2}}(t-1), \dots)$, where $j_{i;k}$ denote the indices of the K_i^{in} nodes that input to node i [4]. The update function f_i is defined for each node i by specifying a $2^{K_i^{in}}$-entry truth table in the following way: we assign 0 with probability p_i or 1 with probability $1-p_i$ for every possible input signal. We define "sensitivity" q_i as the probability that the output f_i changes when given two different input strings [4]. If the Boolean functions are completely random, $q_i = 2p_i(1-p_i)$.

As a special case we can create a regular $N-K$ network whose nodes have uniformly $K_i^{in} = K$ entries [4] . If the graph structure and the update functions are fixed, we can set up the system from a given initial state. The network evolves in descrete timesteps. Since there are finite possible states (2^N) of the network, it will return to a previous state after at most 2^N timesteps, and then it will show periodical behaviour.

To examine the stability of a Boolean network, consider two close initial states, $x(t)$ and $\tilde{x}(t)$. Their divergence can be quantified by the Hamming distance: $h(t) = \sum_{i=1}^{N} |x_i(t) - \tilde{x_i}(t)|$. If $h(t) \to 0$ as $t \to \infty$, the network is

stable. If the network is instable, $h(t)$ quickly increases to $O(N)$, signifying chaotic behaviour. A "critical" network is at the border separating stability and chaos [4]. There is a critical number of entries $K_c = \frac{1}{2p(1-p)}$. If $K < K_c$, the network is stable. If $K > K_c$, it is unstable, and if $K = K_c$, it is critical [4]. E.g., with $K = 10$ entries the network is unstable for $q = 0.215$, $q = 0.3$, $q = 0.4$ and $q = 0.5$. We can see this behaviour for various values of sensitivity on Fig.1., showing our simulation results. We generated a regular $N-K$ network with $N = 10000$ nodes and $K = 10$ connections, and started the simulations from 100 different initially close state-pairs (their initial Hamming distance was 100), and averaged their Hamming distances at each timesteps. At the beginning the average Hamming distance quickly groves showing chaotic behaviour. But after some timesteps it will saturate at a constant value, because the network will get to an attractor and remains there. So the Hamming distance of a state-pair will saturate at a constant value, and because of this, the average of all Haming distances will saturate at a constant value, too.

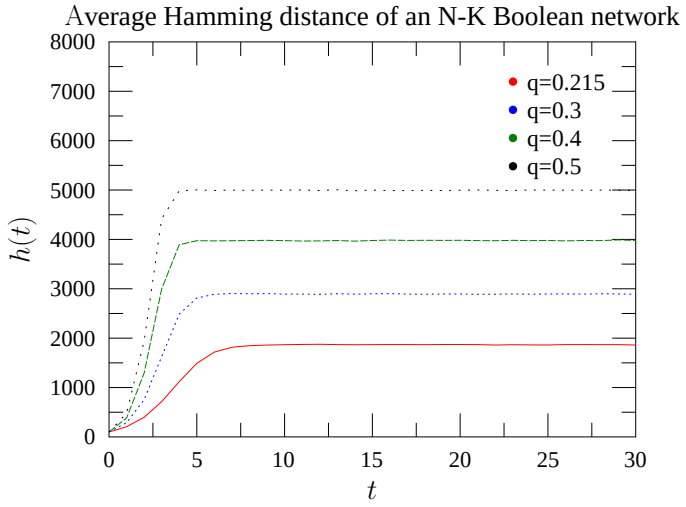

Figure 1: Evolution of the average Hamming distance for a Boolean network of $N = 10000$ and $K = 10$, for various values of the sensitivity.

We can create a directed graph that shows the transition between all

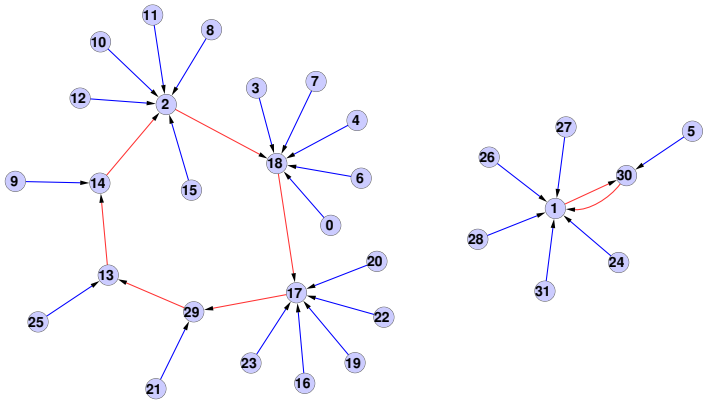

Figure 2: A graph of states of an $N-K$ Boolean network of $N=5$, $K=2$ and $p=0.5$. Attractors are cycles, marked with red.

possible states of a given Boolean network (with given update functions) if we assign a node index for every state in the following way: imagine the state vector of the network as a binary number, then convert this binary number to a decimal number. This decimal number will be a node index. Directed rings and loops correspond to attractors (cycles and fixpoints). This method works only for small graphs, where $N=5$ to $N=20$. Our hope is that examinig these small graphs can reveal interesting effects, which may be generalised for larger graphs as well, where the order of the number of nodes is as large as in real gene regulatory networks.

The networks on figure 2 and 3 shows the transiton between states of a reguar $N-K$ network, where $N=10$, $K=2$ and $p_i=p=0.5$. The most relevant difference between them is that the graph on fig. 3 is connected, while the other on fig. 2 is not connected. On fig. 2 two attractors can be seen, marked with red. On fig. 3 a loop marked also with red shows the fixpoint of the system. It is visible, that by the same N, K and p parameters we get topologically different networks of states.

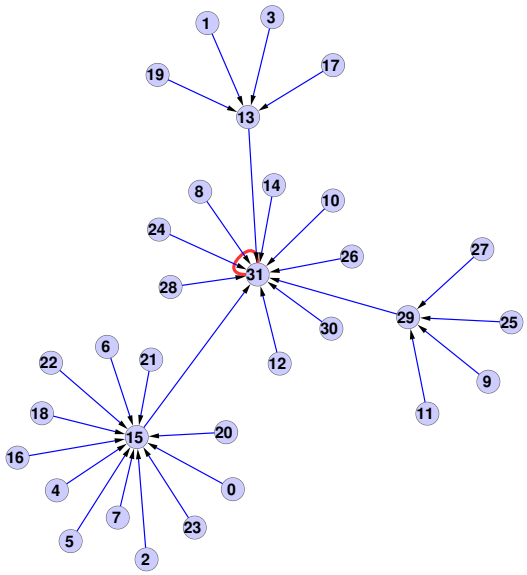

Figure 3: Another graph of states of an $N-K$ Boolean network of $N=5$, $K=2$ and $p=0.5$. The attractor is a fixpoint.

IV. Discussion and outlook

In summary we have shown that Boolean networks offer a simple and intuitive model for gene regulatory networks. The attractors of the Boolean network can be regarded as periodical cycles and steady states of the modelled gene regulatory network. A cell adapts to changes in its environment with the help of regulatory proteins and gene regulatory network. Therefore, if we assume that the environmental conditions are steady for a long time and the cell has adopted to these conditions, the regulatory network is in one of its attractors. However, if there is a change in the environment, after some time the network will get to another attractor driving the cell to its new steady state or periodical cycle. If the graph of the states of gene regulatory network is not connected, there is no transition between separate subgraphs. (On fig. 2 we can see two separate subgraphs.) Each connected subgraph has exactly one attractor, so there has to be a controller outside of the regulatory network that can drive the system from one subgraph to another. These outside controllers could correspond to e.g., special regulatory proteins. These controller regulatory proteins are able to change the state of the whole network. Another natural assumption is that the optimal way of control while changing from one attractor to another happens with the help of as few controller nodes as possible. The aim of our further work is examining the dependence of the minimum number of controller nodes on different parameters of a given Boolean network.

References

[1] Lezon, T. R., Banavar, J. R., Cieplak, M., Maritan A. and Fedoroff, N. V., Proc. Natl. Acad. Sci. **103**, 19033–19038 (2006).

[2] Kalman, R. E., J. Soc. Indus. Appl. Math. Ser. A **1**, 152–192 (1963).

[3] Liu YY, Slotine JJ, Barabási AL., Nature. **473(7346)**, 167–73 (2011)

[4] Andrew Pomerance, Edward Ott, Michelle Girvan and Wolfgang Losert, PNAS **106**, 20 (8209–8214).

[5] Kauffman S., J Theor Biol. **22**, 437–467 (1969).

[6] James E. Darnell, Harvey F. Lodish, David Baltimore, *Molecular cell biology* (Scientific American Books, 1986).