

# Dynamics of Gene Regulation on Content-Based Networks

Duygu Balcan  
Department of Physics  
Istanbul Technical University  
Maslak 34469  
Istanbul, Turkey  
balcand@itu.edu.tr

Ayşe Erzan<sup>\*</sup>  
Department of Physics  
Istanbul Technical University  
Maslak 34469  
Istanbul, Turkey  
erzan@itu.edu.tr

## ABSTRACT

The content-based network model which we have proposed[2] offers possibilities to investigate processes based on molecular recognition and binding. In particular, it seems to be a promising model of gene regulation in its different respects such as topology and dynamics. Random Boolean Dynamics on our content-based network will be introduced and our results will be presented. The network has a scaling behaviour which is different from the scale-free networks investigated up to now.[5] The clustering coefficient and the degree distribution depend on the distribution of the recognition specificities required for binding, which can be tuned at will. The model exhibits a dynamical behaviour which is more ordered or close to critical than its scale-free counterparts. It also opens the way for investigating evolutionary pathways which could have led to existing biological networks.

## Categories and Subject Descriptors

I.6.0 [Simulation and Modelling]: General; G.2.1 [Discrete Mathematics]: Combinatorics; G.2.2 [Discrete Mathematics]: Graph Theory—*network problems*

## General Terms

Theory, Verification

## Keywords

networks, random Boolean dynamics, gene regulation networks

## 1. INTRODUCTION

We use the term “content-based” to denote network models where the interactions are established via linear codes associated with the nodes, the amount of information shared between the two nodes being a measure of the specificity of the pairwise interactions.

We have studied the topological properties of content-based networks in terms of simulations and analytical calculations. Topological features such as the degree distribu-

tions and the hierarchical organization of edges (k-core decomposition) of content-based networks are determined by the length distributions of the respective sequences and are able capture many properties of real complex networks. A concrete example is afforded by gene regulation networks.

Gene regulation networks consist of nodes each representing a gene and edges signifying biochemical interactions between them. These networks are examples of directed graphs where a link originating from gene A and terminating on gene B indicates that the expression of gene A contributes to regulation of gene B. The interaction between two genes is very specific, relying on a recognition and binding mechanism. To mimic such interactions we have considered genes as nodes each of which is associated with two linear codes whose characters have been chosen from a common alphabet. The interaction between a pair of nodes is based on a sequence matching rule: If the coding sequence associated with gene A is repeated as a subsequence in the promoting region of gene B, then a directed link from gene A to gene B is drawn.

In this paper we introduce an extended version of our model, incorporating random Boolean dynamics (RBD) at the nodes. We describe below the results for the structure of the attractors of this model in the very high-dimensional phase space formed by the states of expression of the genes, and offer a comparison with RBD on scale-free networks in the Mean Field approach.

## 2. MODEL

In our extended model each gene has been associated with two linear codes, one representing the promoting region (PR) and one for the transcription factor (TF) which it codes. The latter actually corresponds to the information (minimal sequence) that the transcription factor coded by the gene recognizes in the PR of the gene it regulates. The characters of PR and TF sequences are chosen from a common alphabet of size  $r$  with uniform probability  $1/r$ . Note that the length distributions of regulatory and coding regions have different forms, in general.

The interaction between a pair of nodes is again defined based on a sequence matching rule: If the TF of the  $i$ th gene  $\pi_i$  is repeated as a subsequence in the PR of  $j$ th gene  $\rho_j$  then the expression of the  $j$ th gene is regulated by the expression of the  $i$ th gene. So we may write down the element

<sup>\*</sup>Prof. Erzan also works at Feza Gürsey Institute P.O.B. 6, Çengelköy 34680, Istanbul, Turkey.

of interaction (adjacency) matrix as

$$w_{ij} = \begin{cases} 1 & \text{if } \pi_i \subset \rho_j \\ 0 & \text{otherwise} \end{cases} . \quad (1)$$

In general,  $w_{ij} \neq w_{ji}$ .

The length distribution of these two distinct regions (PR and TF) is the most important external input to our model and affects the topological properties of the constructed network dramatically. Here we will take these distributions to be the same, having either a truncated exponential or a Gaussian form, with  $l_{\min} \leq l \leq l_{\max}$ . From the above definition it follows that all the subsequences of length  $l_{\min} \leq l \leq l_{\rho_j}$  in the promoting region  $\rho_j$  constitute possible “binding sites” for the coding sequences  $\pi_i$ , of length  $l \leq l_{\rho_j}$ .

## 2.1 Dynamics

Random Boolean dynamics has been proposed by Kauffman[3, 4] as models of gene regulation. In his original model each gene is represented by a node and has  $K$  incoming edges from other randomly chosen nodes in the network. The dynamics of the system is defined by considering the nodes as logic gates, with random Boolean functions (RBF) assigned to each node, the input being the state of the neighbors communicated via the incoming edges. Despite the simplicity and highly abstract nature of the model it has been shown that it has a rich phase diagram; the system has “ordered” behaviour in a narrow region in the phase diagram of parameters  $K$  and  $p$ , the fraction of outputs of the RBF which take the value of unity.

RBD has been also studied on networks having scale-free in- or out-degree distributions by Aldana[1] and Mean Field results indicate that the system is in the ordered regime for the scale-free exponent greater than 2.5, for all values of  $p$ .

We have adapted the ideas of Kauffman to our content-based network whose construction has been defined above in the following way. Remembering that the regulation of a gene is determined via the possible binding sites of its promoting region, we have associated a random Boolean function with each PR. The binding state of each binding site in the regulatory region of the  $i$ th gene takes value 1 if there is an active gene whose coding sequence is equivalent to that of this binding site. The Boolean function associated with this PR takes the binding states of all its binding sites as input and gives 1 as output with probability  $p$  or 0 with probability  $1 - p$ . At a given time  $t$  and given the activation state of all nodes  $\Sigma(t)$ , we compute the state of the system at time  $t + 1$  under a randomly chosen set of Boolean functions.

The volume of the phase space of a system consisting of  $N$  nodes each having two states  $\{1, 0\}$  is  $\Omega = 2^N$ . Starting from all initial configurations of nodes  $\Sigma(0)$  (a list of 1s and 0s) it is possible to explore the phase space fully in principle. The phase space may also be considered as a directed graph where each node represent a configuration of system and a directed link from one node to the other represents the trajectory between configurations. In the following subsection we will summarize some of our simulation results.

## 2.2 Simulation Results

A system having a finite number of nodes has a phase space of finite size. If we start from an initial configuration we will eventually start to visit some already visited configurations, which means that we have fallen into a fixed point

or a cyclic orbit, i.e., an *attractor* of the system. The number of configurations constituting an attractor is called the *length of attractor* and the number of configurations that the system visits before falling into the attractor is called the *transient time*. We have observed that for sequence-length distributions of both types, the average values of the number and length of the attractors, as well as the average transient time, grow linearly with system size  $N$ . Nevertheless in all cases the growth rate is much bigger for the exponential sequence-length distribution than the Gaussian case. The scaling behaviour may still change for large  $N$  values (Here we have obtained the results for system size up to 16.). We have also computed the number of configurations falling into same attractor (namely the distribution of basin sizes), and the in-degree distribution of the flow diagram on the phase space (the *precursor* number distribution).

From our results we believe that the system is in the ordered or near critical phase although the scaling exponent for the out-degree distribution of our content-based network is smaller than 2.5 for both sequence-length distributions. Note that the simulation results have been obtained for  $p = 0.5$ , and therefore fall in the regime where the system is expected to be in the chaotic according to Mean Field arguments.

We have also investigated the behaviour of the overlap function  $\langle x(t) \rangle$  which shows whether the small perturbation in initial conditions spread over the system or die out,

$$\langle x(t) \rangle = \frac{1}{N} \sum_i \langle |\Sigma(t) - \Sigma^i(t)| \rangle , \quad (2)$$

where  $\Sigma(0)$  differs from  $\Sigma^i(0)$  at node  $i$ . The average has been taken over all different initial configurations and many different network realizations. We have observed that the overlap function converges to a narrow interval lying below 1 for all network sizes studied. The instability of the fixed point at 1 does not mean chaos, but rather reflects the fact that there is more than 1 attractor or periodic attractors. The complexity of the phase space does increase with  $N$ , allowing the formation of periodic patterns in both “space” and time, as well as the possibility of switching from one stable attractor to another under external perturbations.

## 3. ACKNOWLEDGMENTS

AE gratefully acknowledges partial support from the Turkish Academy of Sciences.

## 4. REFERENCES

- [1] M. Aldana. Boolean dynamics of networks with scale-free topology. *Physica D*, 185, 2003.
- [2] D. Balcan and A. Erzan. Random model for rna interference yields scale free network. *Eur. Phys. J. B*, 38, 2004.
- [3] S. A. Kauffman. Metabolic stability and epigenesis in randomly connected nets. *J. Theor. Biol.*, 22, 1969.
- [4] S. A. Kauffman. *The Origins of Order: Self-organization and Selection in Evolution*. Oxford University Press, N.Y., 1993.
- [5] M. Mungan, A. Kabakçioğlu, D. Balcan, and A. Erzan. Analytical solution of a stochastic content-based network model. *Phys A: Math Gen.*, 38, 2005.