

# The Gene Modular Detection of Random Boolean Networks by Dynamic Characteristics Analysis

Xueyi Bai<sup>1,a</sup>, Binghui Guo<sup>\*1,b</sup>, Xiaohui Yang<sup>1</sup>, Wei Wei<sup>1</sup>, Zhiming Zheng<sup>1</sup>

<sup>1</sup> LMIB, BDBC and School of Mathematics and Systems Science,  
Beihang University, Beijing 100191, China

<sup>a</sup> email: grace.bai@buaa.edu.cn, <sup>b</sup> email: guobinghui@buaa.edu.cn

\*corresponding author

**Keywords:** Gene Network, Random Boolean Networks, Robustness, Modularity.

**Abstract.** In the last decades, random Boolean networks have been widely used in the sociology, biology and other fields. The study of the dynamic behaviour and modularity of random Boolean networks has a very important significance, both at the theoretical level and at the application level. In this paper, based on the classical random Boolean networks, we investigate the different dynamic characteristics between random and BA Boolean networks. By proposing new update rule with three operators, we study the relationship between modularity and the robustness with different parameters and calculate the phase diagram of given scale Boolean network instances.

## 1. Introduction

In application fields, if the entities in the system are abstracted as points, the relationships between the entities are abstracted as edges, many systems in the reality can be abstracted into network models [1]. Network model allows researchers to widely study the statistical characteristics, evolutionary mechanisms and the overall behavior of the network with the knowledge of graph theory and statistical physics.

In 1969, to study the process of gene networks and cell differentiation, Kauffman first proposed the famous Boolean network [2]. Boolean networks utilize Boolean logic value “0, 1” to construct the abstract models of gene networks, logical value “1” indicates that the gene is active; logical value “0” indicates that the gene is suppressed. The interaction between genes is expressed by the Boolean function of the nodes [3]. By simple logical relationship, Boolean network express the complex interactions between genes. Boolean networks have been widely used in a variety of biological research of gene regulatory networks [4].

Due to the number of nodes is fixed, the state space of the network is limited. And due to the transfer of node status in Boolean functions is determined, when the system reaches certain states, even continue to evolve, it will remain in the state or in states. If there are not external perturbations, the system is difficult to evolve from the state to the other state, for such a state we call the network attractor. The robustness is whether the system with a disturbance in the attractor can return to it [5]. This issue may have a more important consequence of biology and neuroscience [6].

Barabasi believes that understanding the topology of complex networks is the key to understand complex systems [7]. In nature, the interaction between the various communities constitutes the overall function of the network. So, the overall behavior of the network can be deduced through the nature of the community [8]. Individuals belonging to the same community have similar or identical functions.

## 2. Random Boolean Network Model

In this paper, we construct a random Boolean network model with two aspects: topological structure and Boolean function. The topological structure of the model in this paper is based on the ER network and BA network. The specific construction method is: first, use the classical algorithm to generate the classical ER or BA random network. The network is undirected, so we introduce a parameter  $\varrho$  to determine the direction of edges. In this paper,  $\varrho = 1/2$ . After that, for easy, we limit the number of indegree of all nodes in the network fixed. Experiments show that, although the number of indegree is fixed, the sum of indegree and outdegree of the nodes is still subject to the original distribution.

After the generation of topological structure, the update function of each node can be assigned [9]. When constructing update function, we use three operators: “ $\sim$ ”, “ $+$ ”, and “ $\times$ ”. “ $\sim$ ” indicates the reverse, “ $+$ ” indicates XOR, “ $\times$ ” indicates multiplication. When we construct the update function, we need to introduce two parameters  $P_1$  and  $P_2$ ,  $P_1$  is the probability of using “ $+$ ” (the probability of using “ $\times$ ” is  $1 - P_1$ ),  $P_2$  is the probability of using “ $\sim$ ”.

After getting the update functions of all nodes, we can study the dynamic characteristics. Constantly get different initial states to iterate, due to the number of nodes is small, so in theory all attractors of the random Boolean network can be obtained, as well as the robustness of the network.

## 3. Semi-supervised Non-Negative Matrix Factorization Algorithm

Given an undirected network  $G = (V, E)$ , the first step is to construct the similarity matrix  $K$  between nodes, the must connect matrix  $C_{ML}$  and the must be divided into matrix  $C_{CL}$ . Then introduce the composition of semi supervised, if  $(i, j) \in C_{ML}$  that node  $i$  and node  $j$  is similar, if  $(i, j) \in C_{CL}$  that node  $i$  and node  $j$  are in different classes. After all, the new similarity matrix is constructed:

$$\bar{K} = K + \alpha C_{ML} - \beta C_{CL} \quad (1)$$

The  $\alpha, \beta$  are the relative weights of the components of semi supervised.

Assuming the network  $G$  is divided into  $m$  communities, the  $N$  nodes of the network are in these communities. We use the non-negative  $m \times N$  matrix  $H$  to describe the relationship between nodes and communities, called as a group-point similarity matrix [10], where the greater the value, the closer relationship between the nodes and the community.

After theoretical deduction, we can get the iterative formula:

$$H_{ij} = H_{ij} \left( \frac{(H\bar{K})_{ij}}{(HH^T H)_{ij}} \right) \quad (2)$$

The algorithm can be described as follow:

Input:

$G$ : network, node set  $V$ , edge set  $E$

$C_{ML}/C_{CL}$ : must link/can't link matrixs

$m$ : classification number

$\tau$ : maximum iteration number

$\sigma$ : the algorithm convergence criteria

$\alpha$ 、 $\beta$ : the semi-supervised weight

Output:

$\{V_i\}_{i=1}^m$ : hard partition of node set  $V$

1: Construct the similar matrix  $\bar{K}$

2: Construct a nonnegative initial solution  $H^{(0)}$ , the element  $H_{ij}^{(0)}$  obeys the standard normal distribution. If  $H_{ij}^{(0)} < 0$ , get the absolute value of  $H_{ij}^{(0)}$

3: Update matrix  $H$ ;

4: Judge whether the algorithm converges, that is, whether  $|J_{SS-NMF} - J_{SS-NMF}^{(k-1)}| \leq \sigma$ , if converges, skip to step 6, otherwise the next step;

5: Determine whether to reach the largest number of iterations, if so, jump the next step, otherwise jump to step 3;

6: For each node  $i$ , find the location of the largest value in the column vector  $(H_{1i}, \dots, H_{mi})$ , which is to find out the  $l^* = \operatorname{argmax}_{1 \leq l \leq m} H_{li}$ , the node belongs to the community  $l^*$ , namely  $i \in V_{l^*}$ . If the maximum element in the column vector is more than one, then randomly select one from them;

7: Return:  $\{V_i\}_{i=1}^m$

#### 4. Results

We can get from the Figure 1 (a) and (b) that, the semi-supervised components  $\alpha, \beta$  have a weak influence on module density index  $D$ , but the number of indegree has a more obvious effect on modularity density index  $D$ . Both in ER random Boolean network and BA random Boolean network, modularity density index  $D$  increases with the increase of  $in$ , and the magnitude of the increasing decreases gradually. This shows that the less edges in the network, the less obvious of the modularity, and it can be predicted that there is an optimal number of edges for the network with  $N = 30$  nodes to make the modularity the most obvious.

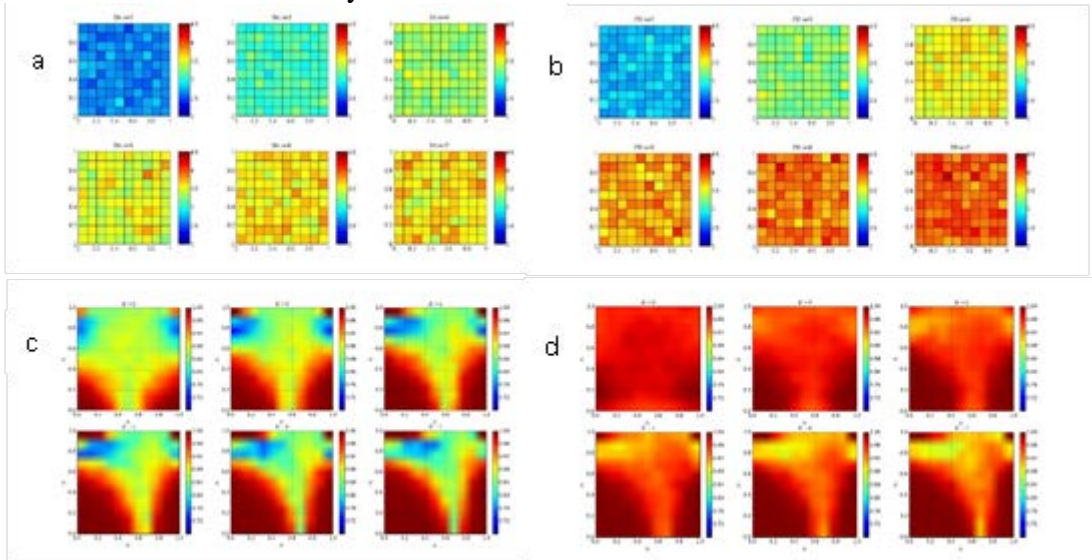


Figure 1 (a) and (b) are the relationship between the module density index  $D$  and the parameters  $(in, \alpha, \beta)$  of ER random Boolean network and BA random Boolean network respectively. (c) and (d) are respectively the relationship between the robustness index  $R$  and parameters  $(in, p_1, p_2)$  of ER random Boolean networks and BA random Boolean network.

So, the  $\mathcal{D}$  values of ER random Boolean network are generally higher than that of BA random Boolean network under the same parameters  $(m, a, b)$ , which shows that the ER random Boolean network is more modular than BA in the same condition.

What can get from (c) and (d) are: 1, whether it is ER random Boolean networks or BA random Boolean networks, the distribution of robustness index  $\mathcal{R}$  is basically consistent in different  $m$ , namely when the value of  $p_2$  is low, the value of  $\mathcal{R}$  is generally high, in addition to the  $p_1 = 0.6$ . When the value of  $p_2$  is high, the value of  $\mathcal{R}$  is generally low, especially when  $p_2 = 0.8$ ,  $\mathcal{R}$  value reaches the lowest point of the entire figure, showing that  $p_2$  has a large influence on the network robustness, namely the higher probability for taking “~”, the lower the network robustness; 2, with the increasing of the  $m$ , the number of cool color lattice in the two figure gradually increases, which shows that the network robustness decreases with the increase of the  $m$ ; 3, compared with (c) and (d), we can find that for the same set of parameters  $(m, p_1, p_2)$ , BA random Boolean network robustness index  $\mathcal{R}$  is generally higher than ER random Boolean network, indicating that the robustness of BA random Boolean network is better than ER random Boolean networks in the same condition.

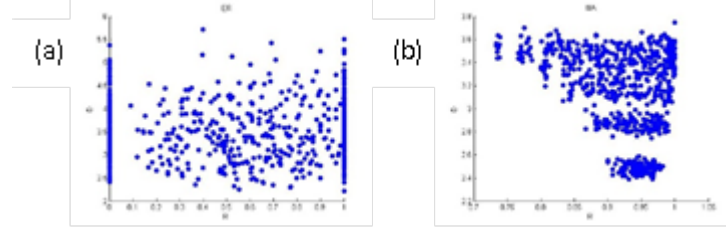


Figure 2 (a) and (b) are the relationship between modularity index  $\mathcal{D}$  and robustness index  $\mathcal{R}$  in ER and BA random Boolean network respectively.

From Figure 2, we can obtain that there is no obvious relationship between  $\mathcal{D}$  and  $\mathcal{R}$  in ER random Boolean network, but the relationship between  $\mathcal{D}$  and  $\mathcal{R}$  in BA random Boolean network is more obvious, that is, with the increase of the robustness index  $\mathcal{R}$ , the range of  $\mathcal{D}$  increases, which is from 3.6 to [2.4, 3.6]. This shows that in the BA random Boolean network, with the decrease of the robustness index  $\mathcal{R}$ , the modularity index  $\mathcal{D}$  will increase in the overall.

## 5. Conclusions

In this paper, we use the ER random Boolean network and BA random Boolean network in the study of random Boolean networks. The  $p_1, p_2$  are used to introduce the update function, and the semi-supervised non-negative matrix factorization algorithm is used for the modularity of the network. In the end, we show the relationship between the  $\mathcal{R}$ ,  $\mathcal{D}$  and the parameters, the frequency distribution of each of them, and the relationship between  $\mathcal{D}$  and  $\mathcal{R}$ .

Due to the limitation of the indegree in this paper, the number of independent variables of the update function is fixed, to achieve a more comprehensive understanding of the real network evolution rules, we can consider illimitation of Boolean networks indegree, that is, study the various characteristics of the network without restrictions on updating function.

## Acknowledgements

This work was supported by the Major Program of National Natural Science Foundation of China (11290141), National Natural Science Foundation of China (11401017, 11571028 and 11671025),

Fundamental Research Funds for the Central Universities, Fundamental Research of Civil Aircraft No. MJ-F-2012-04, and Innovation Foundation of BUAA.

## References

- [1] Ma XK. Models and Algorithms for Community Structure Detection in Complex Networks and Applications in Biological Networks[D]. Xidian University. 2014.
- [2] Luo C. Research on some problems in chaos and asynchronous Boolean networks[D]. Dalian University of Technology. 2013.
- [3] Luque B, Solé R V. Controlling chaos in random Boolean networks[J]. EPL (Europhysics Letters), 1997, 37(9): 597.
- [4] Wang XH, Wang X, Liu LL, Xu P, Liu WB. Research on dynamic behavior of Boolean networks[J]. Journal of Zhejiang Normal University(Nat. Sci). 2012. 35(1):47-52. DOI:10.3969/j.issn.1001-5051.2012.01.009.
- [5] Lloyd-Price J, Gupta A, Ribeiro A S. Robustness and Information Propagation in Attractors of Random Boolean Networks[J]. Plos One, 2012, 7(7).
- [6] Meng Y. The study on the robust of genetic regulatory network by boolean network model[D]. Central China Normal University. 2011. DOI:10.7666/d.y1897759.
- [7] Liu JX. The Research on Detecting Community Structure of Complex Networks and its Application in Financial Networks[D]. Lanzhou University of Technology. 2013.
- [8] Tian Y, Liu DY, Yang B. Application of Complex Networks Clustering Algorithm in Biological Networks[J]. Journal of Frontiers of Computer Science and Technology. 2010. 4(4):330-337. DOI:10.3778/j.issn.1673-9418.2010.04.005.
- [9] Chen Y. The Construction and Application Research of Complex Gene Logical Networks[D]. Beijing University of Technology. 2009.
- [10] Zhao K, Zhang SW, Pan Q. Identification of Overlapping Clusters in Complex Networks Using Symmetrical Nonnegative Matrix Factorization[J]. Computer Measurement and Control. 2010. 18(12):2872-2874.