The influence of the basin structure of Boolean networks on their long range correlated dynamics

Peng Xu[1], Xianghong Wang[1], Wenbin Liu[1*]

[1] Department of Physics and Electronic information engineering, Wenzhou University

Wenzhou 325035, Zhejiang, China

* Corresponding author:wbliu6910@126.com

Abstract—**It has been known for quite some time that the $1/f$ dynamics play a vital role in living organisms. Recently we studied the long-range correlated dynamics of Boolean networks, and found that some networks could present the $1/f$ dynamics while others couldn't. An important question is what kind of networks can generate such dynamics? In this paper, we investigate this issue based on the attractor structure of Boolean networks. We find that multiple attractor networks prefer to generate the $1/f$ dynamics and systems with large basin entropy tend to sustain such dynamics in a wide noise range. Models for eight real genetic networks also partially support these observations.**

Key words—**Genetic Regulatory Network, Attractor, Basin Entropy.**

I. INTRODUCTION

Mathematical and computational modeling is becoming increasingly important for understanding the complex dynamical interactions in genetic regulatory systems. Since Kauffman's seminar inception, Boolean networks have been one of the most intensively studied models of discrete dynamical systems. Although such models are an over simplification of intracellular process, study of the relationships between structural organization and dynamical behaviors of Boolean networks have yielded important insights into the overall behaviors of genetic regulatory networks [1].

In physiology, free-running healthy systems typically generate complex output signals that have long-range correlations. Deviations from the $1/f$ pattern have been associated with disease or aging in various contexts [2, 3]. Generally, the long-range correlated dynamical behaviors include three regimes: Brownian noise, $1/f$ process and white noise. In 2004, Amaral et al. studied the long-range correlations of cellular automata (CA) model, and they found that this model could generate a variety of rich dynamic behaviors under some ordered rules, whereas random Boolean networks could only present white noise dynamics [4]. Recently, we studied the long-range correlated behaviors of Boolean networks with perturbations (BNp). We found the evolution of the correlated behaviors of this model generally follows one of the three types shown in Fig 1. [5]. As the noise η increases, type I refers to the exponent α goes through the

Brownian noise regime, the $1/f$ dynamics and the white noise regime regularly; type II refers to it starts from the white noise and then enter the $1/f$ dynamics and finally returns to the white noise regime again; while type III refers to it stays at the white noise regime no matter what level of noise. Obviously, both type I and II can sustain the $1/f$ dynamics in a continuous noise range while type III can't.
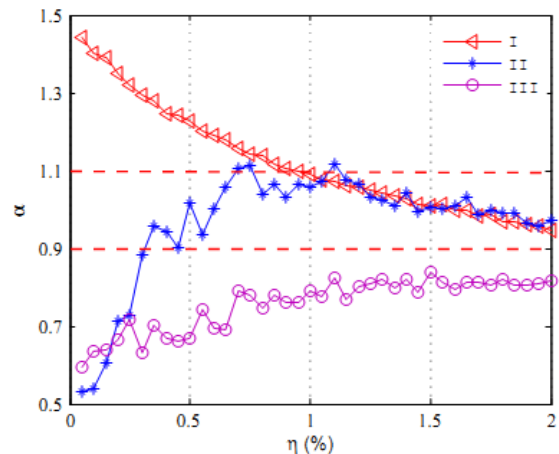

Fig. 1. Three evolution types of the exponent α with respect to the noise η

Despite its practical and fundamental interest, the origin of such correlated dynamics remains an unsolved problem. One important issue is what kind of networks can exhibit the $1/f$ dynamics show in the type I or type II. As the long-range correlated dynamics actually reflect the running smoothness of the system in its state space, it should be associated with the attractor structure of the networks. In this paper, we explore the presence of $1/f$ dynamics based on the partition of the state space (or the attractor structure). In Section 2, we briefly introduce Boolean networks, attractors and the basin of attraction, sensitivity and the long-range correlated behaviors of complex systems. Then we demonstrate the quantification of the state of Boolean networks in section 3, and present the simulation results and some discussions in section 4. Finally, we conclude in

Section 5.

II. BACKGROUND

A. Boolean networks

A Boolean network $G(V,F)$ is defined by a set of nodes $V = \{x_1, \cdots, x_n\}$, $x_i \in \{0,1\}$ and a set of Boolean functions $F = \{f_1, \cdots, f_n\}$, $f_i : \{0,1\}^{k_i} \to \{0,1\}$. Each node x_i represents the expression state of the gene x_i, where $x_i = 0$ means that the gene is OFF, and $x_i = 1$ means it is ON. For each node x_i, a Boolean function $f_i(x_1, \cdots, x_{k_i})$ with k_i specific input nodes is assigned to it and is used to update its value at the next time step. Under the synchronous updating scheme, all genes are updated simultaneously according to their corresponding update functions. The network's state at time t is represented by a binary vector $x(t) = (x_1(t), \ldots, x_n(t))$ and, in the absence of noise, the system transitions from state to state in a deterministic manner. The state space of Boolean networks can be partitioned into disjoint clusters of states. In each cluster, a randomly selected state will eventually attains a fixed state or some recurring states. Such state(s) are called attractors, and the states running into them comprise their basin of attraction. A network can have many attractors, but at least one must exist. It has been hypothesized that attractors in Boolean formalisms correspond to different cell types or cell fates of an organism. In other words, the phenotypic traits are encoded in the attractors.

From the information processing perspective, the partitioning of state space into disjoint basins of attraction can be view as a classifier. Within each basin, all states are in the same class in that they are associated with the same attractor. Krawitz proposed a parameter, the basin entropy, as a measure of the complexity of this classification process. The basin entropy h of a network B is defined as [6]

$$h(B) = -\sum_{\rho} w_{\rho} \log_2^{w_{\rho}}$$

where the weight w_{ρ} of an attractor ρ is the sum of all attractor states and its basin states normalized by the total state space 2^n. Obviously, the basin entropy actually reflects the uncertainty of dynamical behavior based on the distribution of the size of attractions.

B. The ordered behavior

The sensitivity S of a network, defined as the average sensitivity of the functions used in the network, is an ordered parameter that specifies how 1-bit perturbation spreads throughout the network on average. For $S < 1$, this perturbation tends to die out and the network runs in the ordered regime. For $S > 1$, this perturbation tends to diverge and the network runs in the chaotic regime. Networks operate along the boundary between order and chaos corresponding to $S = 1$ and are said to be critical. Given a Boolean network $G(V,F)$, its sensitivity can be calculated as[7]

$$S = \frac{1}{n}\sum_{i=1}^{n} s^{f_i}$$
$$= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k_i} \alpha_j^{f_i}$$
$$= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k_i}\frac{1}{2^{k_i}}\sum_{x\in\{0,1\}^{k_i}} \partial f_i(x)/\partial x_j$$

where $\alpha_j^{f_i}$ defines to the activity of gene x_j in function f_i, $\partial f_i(x)/\partial x_j = f_i(x^{(j,0)}) \oplus f_i(x^{(j,1)})$ denotes the partial derivative of f_i with respect to x_j, $\oplus$ is addition modulo 2 and $x^{(j,k)} = (x_1, \cdots, x_{j-1}, k, x_{j+1}, \cdots, x_k)$, $k = 0,1$.

C. The long-range correlated behaviour

The long-range correlated behaviors actually reflect the smoothness of a time series generated from a complex system. Applying the commonly used detrended fluctuation analysis (DFA) method, such long-range correlated dynamics can be quantified by a scaling exponent α, which is extracted with the linear regression in double-logarithmic coordinates by a least-squares algorithm. A DFA exponent of $\alpha = 1.5$ indicates the system exhibits the ideal Brownian noise. The totally unpredictability of white-noise yields as $\alpha = 0.5$. The $1/f$ process, corresponding to $\alpha = 1$, is a critical boundary between the Brownian noise regime and the white-noise. In practice, the $1/f$ process is relaxed to the range of $0.9 \le \alpha \le 1.1$.

III. METHODS

A. Quantification of the State of Boolean Networks

In reference [4], the state of CA models is quantified as the number of 1s in the corresponding binary vectors. However, it cannot reflect the basin structure of Boolean networks. Given a starting state s_0 under a noisy environment, it may either evolve toward its attractor states and run around them or jump out of its attraction basin and runs around other attractor states. In this paper, we always select the starting state $x(0) = (x_1(0), \ldots, x_n(0))$ from the largest basin B and its corresponding attractors A are defined as the zero point. Specifically, a state $x(t)$ is quantified as

$$u(t) = \begin{cases} q & x(t) \in B \\ 0 & x(t) \in A \\ c & x(t) \notin A \cup B \end{cases}$$

where q is the minimal steps to reach an attractor state $a \in A$ in the absence of noise, c is a large constant number which represents the system deviates from its original basin and runs into other cellular type or function (in this paper we set this value as 25). Our definition can explicitly indicate how far the system is from its original functional type.

B. Generation of Boolean Networks

In this paper, all simulations are performed on networks of size $n=10$ and $k=4$. For each sensitivity S, we randomly generate 300 networks. The sensitivity S ranges from 0.7 to 1.3 with an increment 0.1. Thus, we totally investigate 2100 networks. Concerning the noise environment, each gene can perturb independently from 0 to 1 or 1 to 0 with probability η. The noise η ranges from 0.05% to 2% with an increment 0.05%.

In order to eliminate the effect of starting state s_0, we first run a network 1000 steps then record its trajectory by 10000 steps at a specific noise η. The DFA exponent α of the trajectory is calculated by time scales $10 \le l \le 1000$. The final value of α for a network is an average over 20 independent runs.

C. Results and discussions

As mentioned in introduction, our intuition is that the basin structure of the state space should have important influence on the generation of the $1/f$ dynamics. First, the number of attractors is an important parameter to characterize the state space. Here, we generate two kinds of networks which have no constraint on the number of attractors and those with just two attractors. Fig.2. shows the percentage of networks generating the $1/f$ dynamics in the two kinds of networks. More than 80% of the networks with two attractors can generate the $1/f$ dynamics, whereas only 40~70% of the networks can in those without such constraint. Taking a further look at the latter networks, we find that more than 90% of those which can't present the $1/f$ dynamics have only one attractor. This suggests that single attractor networks tend to lose the $1/f$ dynamics. Concerning networks without networks on the number of attractors, the percentage of networks with the $1/f$ dynamics increases with sensitivity S. This may be explained as networks with higher sensitivity tend to have more attractors. Based on these observations, we may conclude that networks with multiple attractors prefer to generate the $1/f$ dynamics.
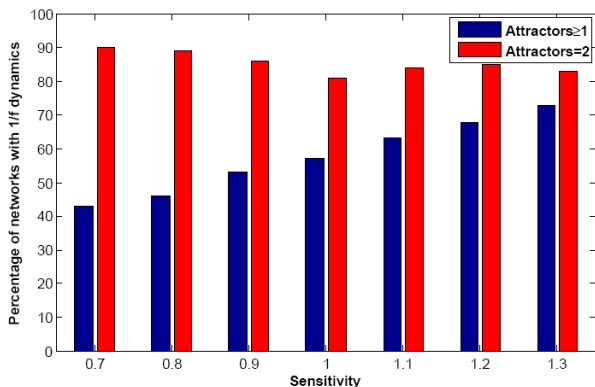
Fig. 2. The percentage of networks presents the $1/f$ dynamics at each sensitivity, blue colour for networks constructed randomly and red colour for networks with just two attractors.

Because some networks may sustain the $1/f$ dynamics in a wide noise range, while others may exhibit in a narrow noise range, we now study the influence of the basin size on the width of the $1/f$ dynamics based on networks with two attractors. As introduced in section 2.1, the size of basins can be characterized by the basin entropy. In this paper, we define the width of the $1/f$ dynamics as

$$d = (\eta_{0.9} - \eta_{1.1})/0.0005$$

where $\eta_{0.9}$ and $\eta_{1.1}$ denote the noise rate η corresponding the DFA exponent α is 0.9 and 1.1 respectively. Figure 3 shows the average width of the $1/f$ dynamics in each basin entropy range. It is easy to see that the average width increases with the basin entropy, and it reaches the maximal value when the entropy is larger than 0.5. This indicates that networks with larger basin entropy tend to exhibit the $1/f$ dynamics in a wider noise range than those with smaller basin entropy.
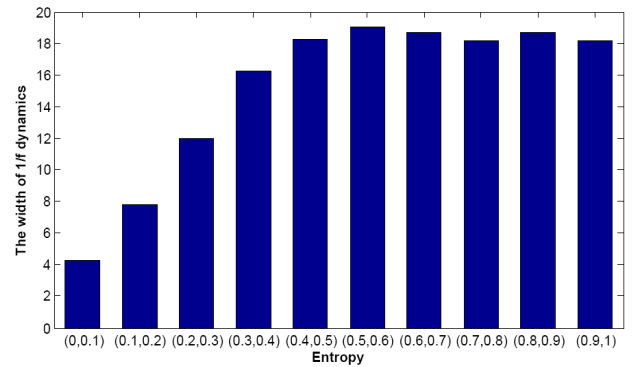
Fig. 3. The average width of the $1/f$ dynamics in each basin entropy for networks with two attractors.

From the definition of the basin entropy, small entropy generally indicates that the state space is partitioned heterogeneously by each basin. Specifically, one basin of attraction is very large while the other is very small. As such two basins degrade into one lager basin, the basin entropy becomes zero and the average width of the $1/f$ dynamics should also decrease to zero. That is the $1/f$ dynamics will disappear gradually. This induction in turn consists with that single attractor networks can hardly generate the $1/f$ dynamics.

TABLE I. The number of attractors and basin entropy for 8 real genetic networks

Real biological networks	Number of genes	Number of attractors	Basin entropy	Type of the 1/f dynamics
Melanoma cellular	7	4	0.79	I
Mammal cellular	10	2	1	I

Neuroglia lump cellular	14	4	0.62	I
Drosophila cellular	14	455	0.79	I
Fission yeast	14	2	0.93	I
Budding yeast	11	3	0.98	I
Drosophila segmentation	6	10	0.7	I
Arabidopsis	6	3	0.55	I I

Recently, our group studied the $1/f$ dynamics for eight real genetic systems: fission yeast cell cycle, budding yeast cell cycle, drosophila cell cycle, mammalian cell cycle, drosophila segment polarity, Arabidopsis flower morphogenesis, human neuroglioma cell and metastatic melanoma cell. The Boolean models of these systems are extracted from references [8-15]. As the maximal entropy is a function of the number of attractors, here we normalized the basin entropy of each system with their maximal entropy so that it lies between 0 and 1. Table 1. lists the number of genes, attractors, the relative basin entropy and the type of the $1/f$ dynamics of those networks. All these models have more than one attractor and their relative basin entropy is larger than 0.5. Therefore，these two aspects may be the partial reasons for the generation of the $1/f$ dynamics in these eight models.

IV. CONCLUSION

A hallmark of critical behaviour is the spontaneous emergence of complex and coordinated macroscopic behaviour in the form of long-range spatial or temporal correlations. Such coordination across many scales enables information to propagate over time from one part of the system to another in a high degree of specificity and sensitivity. Our previous study show the $1/f$ dynamics may be an important character of real genetic networks. In this paper, we investigate the mechanism of such dynamics based on the basin structure of the state space. First, we find that multiple attractor structure of state space prefers to generate the dynamics. Second, systems with larger basin entropy tend to sustain such dynamics in a wide noise range. Our findings are of great interest for understanding of the origin of such dynamics. However, there still exist about 10~20% networks can't generate the $1/f$ dynamics even for networks with two attractors. Therefore, there are some other factors, such as the number of relevant nodes and their corresponding basin entropy, which may affect the $1/f$ dynamics, and this will be our future work.

ACKNOWLEDGMENT

REFERENCES

[1] S.A.Kauffman. The Origins of Order: Self-Organization and Selection in Evolution (Oxford Univ.Press , NewYork) ,1993.

[2] I.Shmulevich, E.R.Dougherty, and W.Zhang. "From Boolean to probabilistic Boolean networks as models of genetic regulatory networks," Proceedings of the IEEE, 90(11): 1778–1792, 2002.

[3] I. Shmulevich, H.Lähdesmäki,E.R.Dougherty, J.Astola and W.Zhang. "The role of certain Post classes in Boolean network models of genetic networks," Proc Natl Acad Sci USA.100(19):10734-10739,2003.

[4] A.N.Amaral,D.G.Albert, A.Moreira, L.Ary. "Emergence of complex dynamics in a simple model of signaling networks," PNAS, 101(44):15551–15555, 2004.

[5] P.Xu,L. Liu,X. Zhu,X. Wang and W. Liu. "The study of complex fluctuations in Boolean networks," ICSPCC2011:9-12,2011

[6] P.Krawitz, I. Shmulevich. "Basin Entropy in Boolean Network Ensembles," PHYSICAL REVIEW LETTERS,98(4), 158701,2007.

[7] I.Shmulevich and E.R.Dougherty, *Genomic Signal Processing*, Princeton Series in Applied Mathematics, Princeton University Press, Princeton, NJ, USA, 2007.

[8] M. Davidich, S. Bornholdt. "The transition from differential equations to Boolean networks: A case study in simplifying a regulatory network model," Journal of Theoretical Biology 255(3): 269–277, 2008.

[9] P.Krawitz, I. Shmulevich. "Entropy of complex relevant components of Boolean networks," PHYSICAL REVIEW ,E 76, 036115,2007.

[10] Y. Xiao,E.R.Dougherty. "The impact of function perturbations in Boolean networks," BIOINFORMATICS, 23 (10):1265–1273, 2007.

[11] L.Mendoza,D.Theffry and E.R.Alvarez-Buylla. "Genetic control of flower morphogenesis in Arabidopsis thaliana: a logical analysis," BIOINFORMATICS, 15(7/8):593-606, 1999.

[12] S. Zhang,W. Ching,K. Michael and T.Akutsu. "Simulation study in Probabilistic Boolean Network models for genetic regulatory networks," Data Mining and Bioinformatics, 1(3):217-240, 2007.

[13] I. Maria,S. Borholdt. "Boolean Network Model Predicts Cell Cycle Sequence of Fission Yeast," Plos ONE,3(2),2008.

[14] A.Fauré,D.Thieffry. "Logical modelling of cell cycle control in eukaryotes: a comparative study," Mol Biosyst.5(12):1569-81, 2009.

[15] L. Sanchezi,C. Chaouiya,D. Thieffry. "Segmenting the fly embryo: logical analysis of the role of the Segment Polarity cross-regulatory module," Dev. Biol. 52: 1059-1075 ,2008.