Hindawi
Complexity
Volume 2020, Article ID 3687961, 14 pages
https://doi.org/10.1155/2020/3687961

Research Article

The Dynamics of Canalizing Boolean Networks

Elijah Paul,[1] **Gleb Pogudin**,[2,3] **William Qin**,[4] **and Reinhard Laubenbacher**[5]

[1]*California Institute of Technology, Pasadena, CA, USA*
[2]*Courant Institute of Mathematical Sciences, New York University, New York, NY, USA*
[3]*Department of Computer Science, National Research University Higher School of Economics, Moscow, Russia*
[4]*Millburn High School, Millburn, NJ, USA*
[5]*Center for Quantitative Medicine, University of Connecticut Health Center and Jackson Laboratory for Genomic Medicine, Farmington, CT, USA*

Correspondence should be addressed to Reinhard Laubenbacher; laubenbacher@uchc.edu

Elijah Paul, Gleb Pogudin, and William Qin contributed equally to this work.

Received 13 September 2019; Accepted 6 November 2019; Published 20 January 2020

Guest Editor: George V. Popescu

Boolean networks are a popular modeling framework in computational biology to capture the dynamics of molecular networks, such as gene regulatory networks. It has been observed that many published models of such networks are defined by regulatory rules driving the dynamics that have certain so-called canalizing properties. In this paper, we investigate the dynamics of a random Boolean network with such properties using analytical methods and simulations. From our simulations, we observe that Boolean networks with higher canalizing depth have generally fewer attractors, the attractors are smaller, and the basins are larger, with implications for the stability and robustness of the models. These properties are relevant to many biological applications. Moreover, our results show that, from the standpoint of the attractor structure, high canalizing depth, compared to relatively small positive canalizing depth, has a very modest impact on dynamics. Motivated by these observations, we conduct mathematical study of the attractor structure of a random Boolean network of canalizing depth one (i.e., the smallest positive depth). For every positive integer ℓ, we give an explicit formula for the limit of the expected number of attractors of length ℓ in an n-state random Boolean network as n goes to infinity.

1. Introduction

Dynamic mathematical models are a key enabling technology in systems biology. Depending on the system to be modeled, the data and information available for their construction, the questions to be answered, and different modeling frameworks can be used. For kinetic models, systems of ordinary differential equations have a long tradition. Generally, they will have the very special structure of polynomial equations representing Michaelis–Menten kinetics, even in the case of systems, such as gene regulatory networks, that are not proper biochemical reaction networks. It is this special structure that gives models desirable properties and aids in model analysis. Besides continuous models, a range of discrete models are finding increasingly frequent use, in particular Boolean network models of a broad variety of biological systems, from intracellular molecular networks to population-level compartmental models (see e.g., [1–5]), going back to the work of Kauffman in the 1960s [6–8]. While Boolean network models, a collection of nodes, whose regulation by other nodes is described via a logical rule built from Boolean operators, are intuitive and mathematically simple to describe, their analysis is severely limited by the lack of mathematical tools. It generally consists of simulation results. Any set function on binary strings that takes on binary values can be represented as a Boolean function, so that the class of general Boolean networks is identical to the class of set functions on binary strings of a given length, making any general analysis impossible. The search for special classes of Boolean functions that are broad enough to cover all or most rules that occur in biology, but special enough to allow for mathematical approaches has a long history.

It was again Kauffman who proposed a class of functions [7] with properties inspired by the developmental biology

concept of *canalization*, going back to Waddington in the 1940s [9]. There is some evidence that canalizing Boolean functions do indeed appear disproportionately in published models and that the dynamics of Boolean network models consisting of canalizing functions has special properties, in particular a "small" number of attractors. This is important since, in the case of intracellular molecular network models, attractors correspond to the different phenotypes a cell is capable of. Here, again, the majority of available results are obtained by simulating large numbers of such networks. The main question of this paper is as follows: *What do the dynamics of a random canalizing Boolean network look like?* We approach this question using both computer simulations and analytical methods, with the main result of the paper being Theorem 2, which gives a provable formula for the number of expected attractors of a general Boolean network with a particular canalization property. In addition to providing important information about canalizing Boolean network models, this result can be viewed as a part of a growing body of mathematical results characterizing this class of networks that promises to be as rich as that for chemical reaction network models based on ordinary differential equations.

2. Background

The property of canalization for Boolean functions was introduced by Kauffman in [7], inspired by the concept of *canalization* from developmental biology [9]. A Boolean function is canalizing if there is a variable and a value of the variable such that if the variable takes the value, then the value of the function does not depend on other variables. It was shown that models defined by such functions often exhibit less chaotic and more stable behavior [10, 11]. Nested canalizing functions, obtained by applying the concept of canalization recursively, were introduced in [2]. They form a special subset of canalizing functions and have stable dynamics [11]. We note that there are other important properties shared by Boolean networks arising in modeling (for example, sparsity [7]). In this paper we focus only on canalization and its impact on the dynamics, and one of the natural future directions would be to consider several such properties simultaneously.

To cover more models arising in applications, the notion of nested canalizing function was relaxed by Layne et al. [12] by assigning to every Boolean function its *canalizing depth*. Noncanalizing functions have canalizing depth zero, and nested canalizing functions have the maximal possible canalizing depth equal to the number of variables. Canalizing depth of a Boolean network is defined as the minimum of the canalizing depths of the functions defining the network. In [12], activities and sensitivities of functions with different canalizing depths and stability and criticality of Boolean networks composed from such functions were investigated. It has been observed that Boolean networks of higher canalizing depth tend to be more stable and less sensitive. However, increasing the canalizing depth to the maximum does not improve the stability significantly compared to moderate positive canalizing depth. These observations give a strong indication of the biological utility of canalizing function, even with small canalizing depth.

Attractors in Boolean network models can be interpreted as distinct cell types [13, p. 202] and their lengths can be viewed as the variety of different gene expression patterns corresponding to the cell type. Thus, understanding the attractor structure of a random Boolean network defined by functions of a fixed canalizing depth is important for assessing biological relevance of such models. Analytic study of the attractor structure of nested canalizing Boolean networks has been carried out in [11]. For discussion about attractors of length one (i.e., steady state), we refer to [14].

3. Our Results

The main question of this paper is as follows: *What do the dynamics of a random canalizing Boolean network look like?* We approach this question using both computer simulations and analytical methods.

In our *computational experiments*, we generate approximately 30 million random Boolean networks of all possible canalizing depths with the number of variables ranging from 4 to 20. For each of these networks, we determine sizes of all the attractors and basins of attraction and analyze the obtained data. We discover the following:

(1) For a fixed number of variables, the sample mean of the number of attractors and average size of an attractor decrease when the canalizing depth increases

(2) The decrease of the average size of an attractor is much greater than the decrease of the number of attractors as the canalizing depth increases

(3) Both decreases from (8) are substantial when the canalizing depth changes from zero to small canalizing depths, but a further increase of the canalizing depth does not lead to a significant decrease for either the sample means or for the empirical distributions

(4) The relative decrease of the sample mean of the number of attractors and the average attractor size when the canalizing depth changes from zero to one becomes sharper when the number of variables increases

Observations (8) and (A.4) are consistent with the results obtained in [12] for sensitivity and stability. This provides new evidence that Boolean networks of small positive canalizing depth are almost as well-suited for modeling as those with nested canalizing functions, from the point of view of stability. Since there are many more canalizing functions of small positive canalizing depth than nested canalizing functions [15, Section 5], they provide a richer modeling toolbox.

Motivated by observation (A.4), we conduct a mathematical study of the attractor structure of a random Boolean network of canalizing depth one (that is, the minimal

positive depth). Our main theoretical result, Theorem 2, gives, for every positive integer ℓ, a formula for the limit of the expected number of attractors of length ℓ in a random Boolean network of depth one. The same formulas are valid for a random Boolean network defined by canalizing functions (see Remark 5). In particular, our formulas show that a large random network of depth one, on average, has more attractors of small sizes that an average Boolean network (Remark 6).

Formulas similar to the ones in our proofs (e.g., in Lemma A.4) have already appeared in the study of the average number of attractors of a given length in sparse Boolean networks, e.g., [16, equation (2)] and [17, equation (6)]. The results of [16, 17] are based on describing the asymptotic behavior of these formulas in terms of N, the number of nodes in the network, and the asymptotics is of the form $O(N^{\alpha})$. In our case, the average number of attractors of a given length simply approaches a constant as $N \longrightarrow \infty$ (that is, $O(1)$), but our methods allow us to find the exact value of this constant.

The source code we used for generating and analyzing data is available at https://github.com/MathTauAthogen/Canalizing-Depth-Dynamics. The raw data are available at https://github.com/MathTauAthogen/Canalizing-Depth-Dynamics/tree/master/data.

Structure of the Paper. The rest of the paper is organized as follows. Section 4 contains necessary definitions about canalizing functions and Boolean networks. Outlines of the algorithms used in our computational experiments are in Section 5. The main observations are summarized in Section 6. Our main theoretical result about attractors in a random Boolean network of canalizing depth one (Theorem 2) is presented in Section 7. Section 8 contains conclusions. The proofs are located in the Appendix.

4. Preliminaries

Definition 1. A *Boolean network* is a tuple $\mathbf{f} = (f_1, f_2, \ldots, f_n)$ of Boolean functions in n variables. For a state $\mathbf{a}_t = (a_{t,1}, a_{t,2}, \ldots, a_{t,n}) \in \{0,1\}^n$ at time t, we define the state $\mathbf{a}_{t+1} \coloneqq f(a_t) = (a_{t+1,1}, \ldots, a_{t+1,n}) \in \{0,1\}^n$ at time $t+1$ by

$$\begin{aligned} a_{t+1,1} &= f_1\left(a_{t,1}, \ldots, a_{t,n}\right), \\ &\vdots \\ a_{t+1,n} &= f_n\left(a_{t,1}, \ldots, a_{t,n}\right). \end{aligned} \tag{1}$$

Definition 2 (attractors and basins). Let $\mathbf{f} = (f_1, \ldots, f_n)$ be a Boolean network.

(i) A sequence $\mathbf{a}_1, \ldots, \mathbf{a}_\ell \in \{0,1\}^n$ of distinct states is called an *attractor* of $\mathbf{f}$ if $\mathbf{f}(\mathbf{a}_i) = \mathbf{a}_{i+1}$ for every $1 \le i < \ell$ and $\mathbf{f}(\mathbf{a}_\ell) = \mathbf{a}_1$.

(ii) An attractor $\mathbf{a}_1, \ldots, \mathbf{a}_\ell \in \{0,1\}^n$ is called a *steady state* if $\ell = 1$.

(iii) Let $A = (\mathbf{a}_1, \ldots, \mathbf{a}_\ell) \in (\{0,1\}^n)^\ell$ be an attractor of $\mathbf{f}$. The *basin* of A is the set

$$\{\mathbf{b} \in \{0,1\}^n \mid \exists N : \underbrace{\mathbf{f}(\mathbf{f}(\ldots(\mathbf{fb})\ldots))}_{N \text{ times}} \in A\}. \tag{2}$$

Definition 3. A nonconstant function $f(x_1, \ldots, x_n)$ is canalizing with respect to a variable x_i if there exists a canalizing value $a \in \{0,1\}$ such that

$$f\left(x_1, \ldots, x_{i-1}, a, x_{i+1}, \ldots, x_n\right) \equiv \text{const}. \tag{3}$$

Example 1. Consider $f(x_1, x_2) = x_1 \cdot x_2$ (the product is understood modulo 2, that is, logical AND). It is canalizing with respect to x_1 with canalizing value 0 because $f(0, x_2) = 0$ regardless of the value of x_2. Analogously, it is canalizing with respect to x_2 with canalizing value 0.

Consider $g(x_1, x_2) = x_1 + x_2$ (summation is understood modulo 2, that is, logical XOR). It is not canalizing with respect to x_1 because

$$\begin{aligned} g(0, x_2) &= x_2 \neq \text{const}, \\ g(1, x_2) &= \overline{x}_2 \neq \text{const}. \end{aligned} \tag{4}$$

The same argument works for x_2 as well.

Definition 4. $f(x_1, \ldots, x_n)$ has canalizing depth [15, Definition 2.3] k if it can be expressed as

$$f = \begin{cases} b_1, & x_{i_1} = a_1, \\ b_2, & x_{i_1} \neq a_1, x_{i_2} = a_2, \\ \vdots & \\ b_k, & x_{i_1} \neq a_1, x_{i_2} \neq a_2, \ldots, x_{i_{k-1}} \neq a_{k-1}, x_{i_k} = a_k, \\ g \neq b_k, & x_{i_1} \neq a_1, \ldots, x_{i_k} \neq a_k, \end{cases} \tag{5}$$

where

(i) $i_1, \ldots, i_k$ are distinct integers from 1 to n

(ii) $a_1, \ldots, a_k, b_1, \ldots, b_k \in \{0,1\}$

(iii) g is a noncanalizing function in the variables $\{x_1, \ldots, x_n\} / \{x_{i_1}, \ldots, x_{i_k}\}$

Example 2. For example, if $f(x_1, x_2, x_3) = (x_1 + x_2)x_3$,

$$f\left(x_1, x_2, x_3\right) = \begin{cases} 0, & x_3 = 0, \\ x_1 + x_2, & x_3 \neq 0, \end{cases} \tag{6}$$

and $x_1 + x_2$ is noncanalizing. Therefore, f has canalizing depth 1.

Remark 1. Since g in Definition 4 is noncanalizing, every function has a single well-defined canalizing depth. In particular, a function of depth two is not considered to have depth one.

Definition 5. We say that a canalizing Boolean function $f(x_1, \ldots, x_n)$ is nested if f has canalizing depth n, that is, $g = 0$ or $g = 1$ (see Definition 4). For example, $f(x_1, x_2, x_3) = x_1x_2x_3$ is nested canalizing because

$$f = \begin{cases} 0, & x_3 = 0, \\ 0, & x_3 \neq 0, x_2 = 0, \\ 0, & x_2, x_3 \neq 0, x_1 = 0, \\ 1, & x_1, x_2, x_3 \neq 0, \end{cases} \tag{7}$$

so the canalizing depth of f is 3, which is equal to $n = 3$.

Definition 6. We say that a Boolean network $\mathbf{f} = (f_1, \ldots, f_n)$ has canalizing depth k if $f_1, \ldots, f_n$ are Boolean functions of canalizing depth k.

5. Simulations: Outline of the Algorithms

In our computational experiment, we generated random Boolean networks of various canalizing depths. For each network, we store a list of pairs (a_i, b_i), where a_i is the size of the ith attractor of the network and b_i is the size of its basin. The generated data are available at https://github.com/MathTauAthogen/Canalizing-Depth-Dynamics/tree/master/data. To generate the data, we used two algorithms: one for generating a random Boolean network of a given canalizing depth and one for finding the sizes of attractors and their basins (Algorithm 1).

5.1. Generating Random Boolean Functions of a Given Canalizing Depth. [12, Section 5] contains a sketch of an algorithm for generating random Boolean functions that have canalizing depth at least k for a given k. Here, we generate functions of canalizing depth equal to k and take a different approach than [12]. In order to ensure that the probability distribution of possible outputs is uniform, we use the following structure theorem due to He and Macaulay [15].

Theorem 1 (see [15], Theorem 4.5). *Every Boolean function $f(x_1, \ldots, x_n) \not\equiv 0$ can be uniquely written as*

$$f(x_1, \ldots, x_n) = M_1(M_2(\cdots(M_{r-1}(M_r p_C + 1) + 1)\cdots) + 1) + b, \tag{8}$$

where $M_i = \prod_{j=1}^{k_i}(x_{i_j} + a_{i_j})$ for every $1 \leq i \leq r$, $p_C \not\equiv 0$ is a noncanalizing function, and $k = \sum_{i=1}^{r} k_i$ is the canalizing depth. Each x_i appears in exactly one of $\{M_1, \ldots, M_r, p_C\}$, and the only restrictions on equation (8) are the following "exceptional cases":

(E1) If $p_C \equiv 1$ and $r \neq 1$, then $k_r \geq 2$

(E2) If $p_C \equiv 1$ and $r = 1$ and $k_1 = 1$, then $b = 0$

Example 3. Consider $f(x_1, x_2, x_3, x_4) = x_1(x_2 + 1)(x_3x_4 + x_3 + x_4)$ can be represented as

$$f = ((x_1 + 0)(x_2 + 1))(((x_3 + 1)(x_4 + 1))(1) + 1) + 0, \tag{9}$$

so $M_1 = (x_1 + 0)(x_2 + 1)$, $M_2 = (x_3 + 1)(x_4 + 1)$, $b = 0$, $k = 4$, and $p_C = 1$. This can be verified by expanding the brackets in the original and new representations of f.

Consider $g(x_1, x_2, x_3, x_4, x_5) = 1 + x_5(x_1 + x_2)(x_3 + 1)x_4$. It can be represented as

$$g = (x_5 + 0)(((x_3 + 1)(x_4 + 0))(x_1 + x_2) + 1) + 1, \tag{10}$$

so $M_1 = (x_5 + 0)$, $M_2 = (x_4 + 0)(x_3 + 1)$, $b = 1$, $k = 3$, and $p_C = x_1 + x_2$.

Our algorithm is summarized in Algorithms 2 and 3 below. Correctness of Algorithm 2 follows from Theorem 1, and correctness of Algorithm 3 can be proved directly by induction on k.

Remark 2. The complexity of Algorithm 2 is $O(n2^n)$ (see Proposition B.2). Given that the size of the output is $O(2^n)$, and this is nearly optimal.

We measured the runtimes of our implementation of Algorithm 2 on a laptop with a Core i5 processor (1.60 GHz) and 8 Gb RAM. Generating a single function with 20 variables (the largest number we used in our simulations) takes 4.9 − 5.5 seconds (faster for smaller canalizing depth). On a laptop, our implementation can go up to 24 variables (∼2 minutes to generate a function), and then hits memory limits. One can go further by using a lower level language and more careful packing. However, already a Boolean function in 40 variables would require at least 128 Gb of memory.

Remark 3. We generate a random noncanalizing function as follows. We generate a random Boolean function and test for canalization until we generate a noncanalizing one. Then, we return it. Since canalizing functions are rare [15, Section 5], this algorithm is fast enough for our purposes (see Lemma B.1).

6. Simulations: Results

Notation 1. For a Boolean network $\mathbf{f} = (f_1, \ldots, f_n)$, let $N(\mathbf{f})$ and $S(\mathbf{f})$ denote the number of the attractors of $\mathbf{f}$ and the sum of the sizes of the attractors of $\mathbf{f}$, respectively. We define the average size of an attractor as $\mathrm{AS}(\mathbf{f}) := S(\mathbf{f})/N(\mathbf{f})$.

6.1. Sample Means of $N(\mathbf{f})$ and $AS(\mathbf{f})$. For every $n = 4, \ldots, 20$ and every $0 \leq k \leq n$, we generate random Boolean networks in n variables of canalizing depth k and compute the mean of $N(\mathbf{f})$ and $\mathrm{AS}(\mathbf{f})$. Figure 1 shows how these means depend on k for $n = 15$ (based on 50,000 samples for each k). The shape of the plots is similar for other values of n we did computation for (that is, $n = 4, \ldots, 20$). Note that although both means are decreasing, the decrease of the mean of $\mathrm{AS}(\mathbf{f})$ is more substantial.

In: A Boolean network $\mathbf{f} = (f_1, \ldots, f_n)$ in n variables
Out: A list of pairs (a_i, b_i), where a_i is the size of the ith attractor of $\mathbf{f}$ and b_i is the size of its basin
(1) (Network $\longrightarrow$ Graph) Build a directed graph G with 2^n vertices corresponding to possible states and a directed edge from $\mathbf{a}$ to $\mathbf{f}(\mathbf{a})$ for every $\mathbf{a} \in \{0, 1\}^n$.
(2) (Attractors) Perform a depth-first search [18, § 22.3] traversal on G viewed as an undirected graph to detect the unique cycle in each connected component, these cycles are the attractors.
(3) (Basins) For each cycle from Step 2, perform a depth-first search traversal on G with all the edges reversed. The dfs trees will be the basins.
(4) Return the sizes of the attractors and basins found on Steps 2 and 3.

Algorithm 1: Finding the sizes of the attractors and their basins.

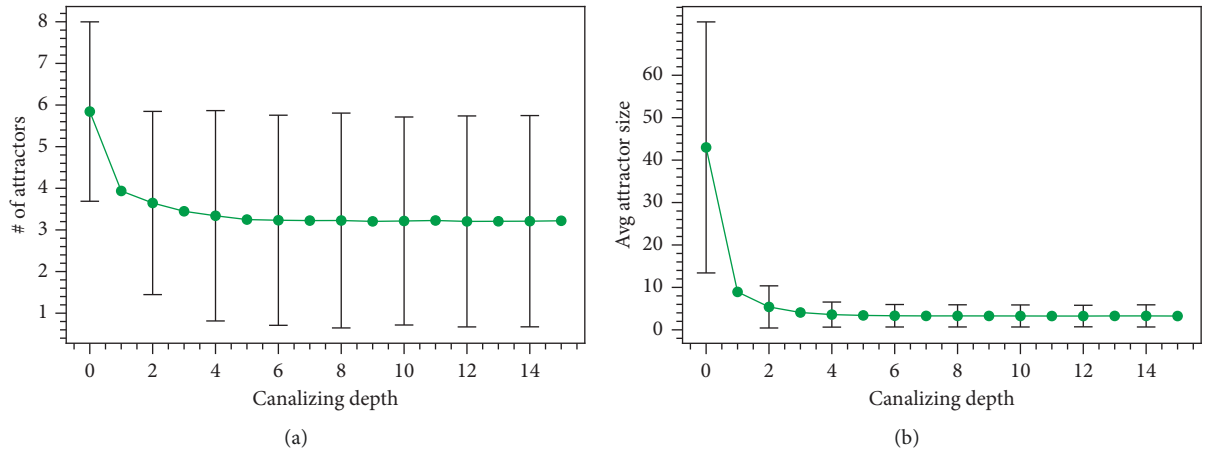

Figure 1: Dependencies of the sample means of $N(\mathbf{f})$ and $\mathrm{AS}(\mathbf{f})$ on the canalizing depth. (a) The number of attractors $N(\mathbf{f})$. (b) Average size of an attractor $\mathrm{AS}(\mathbf{f})$.

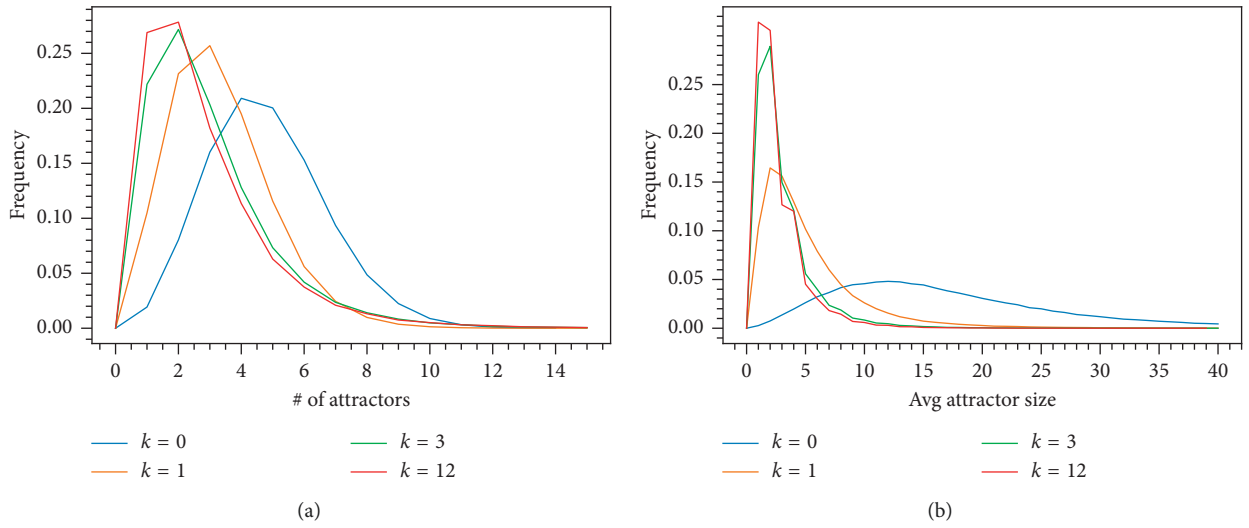

Figure 2: Empirical distributions of $N(\mathbf{f})$ and $\mathrm{AS}(\mathbf{f})$ for $n = 12$ and $k = 0, 1, 3, 12$. (a) Distribution of the number of attractors $N(\mathbf{f})$. (b) Distribution of the average size of an attractor $\mathrm{AS}(\mathbf{f})$.

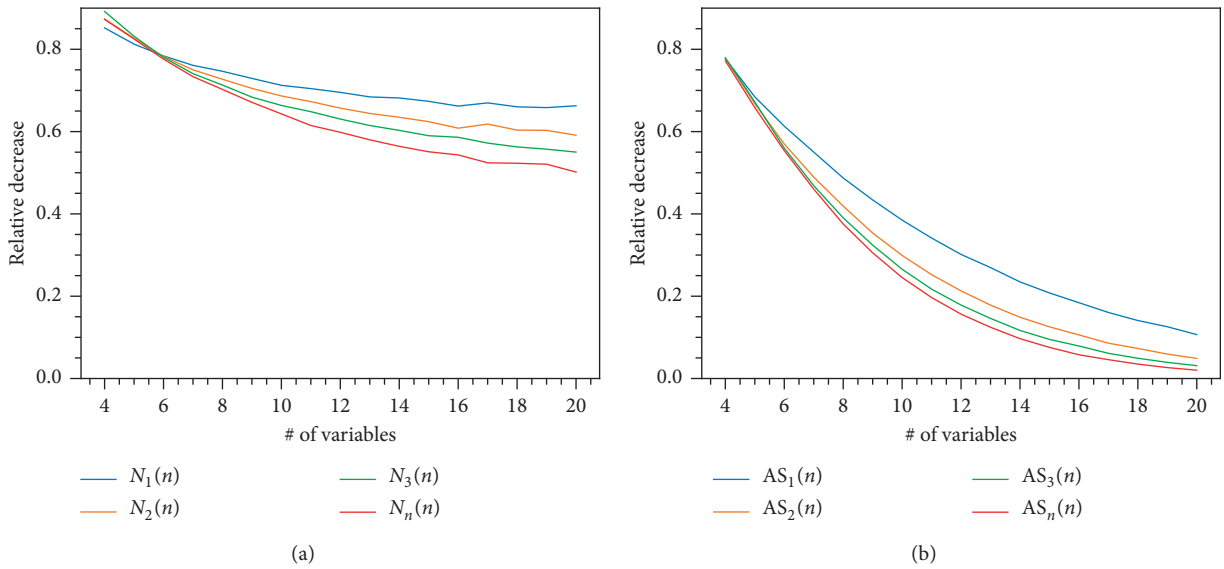

FIGURE 3: Dependence of the relative decreases of the sample means of $N(\mathbf{f})$ and $\mathrm{AS}(\mathbf{f})$ on the number of variables n. (a) Relative decrease of the number of attractors. (b) Relative decrease of the average size of an attractor.

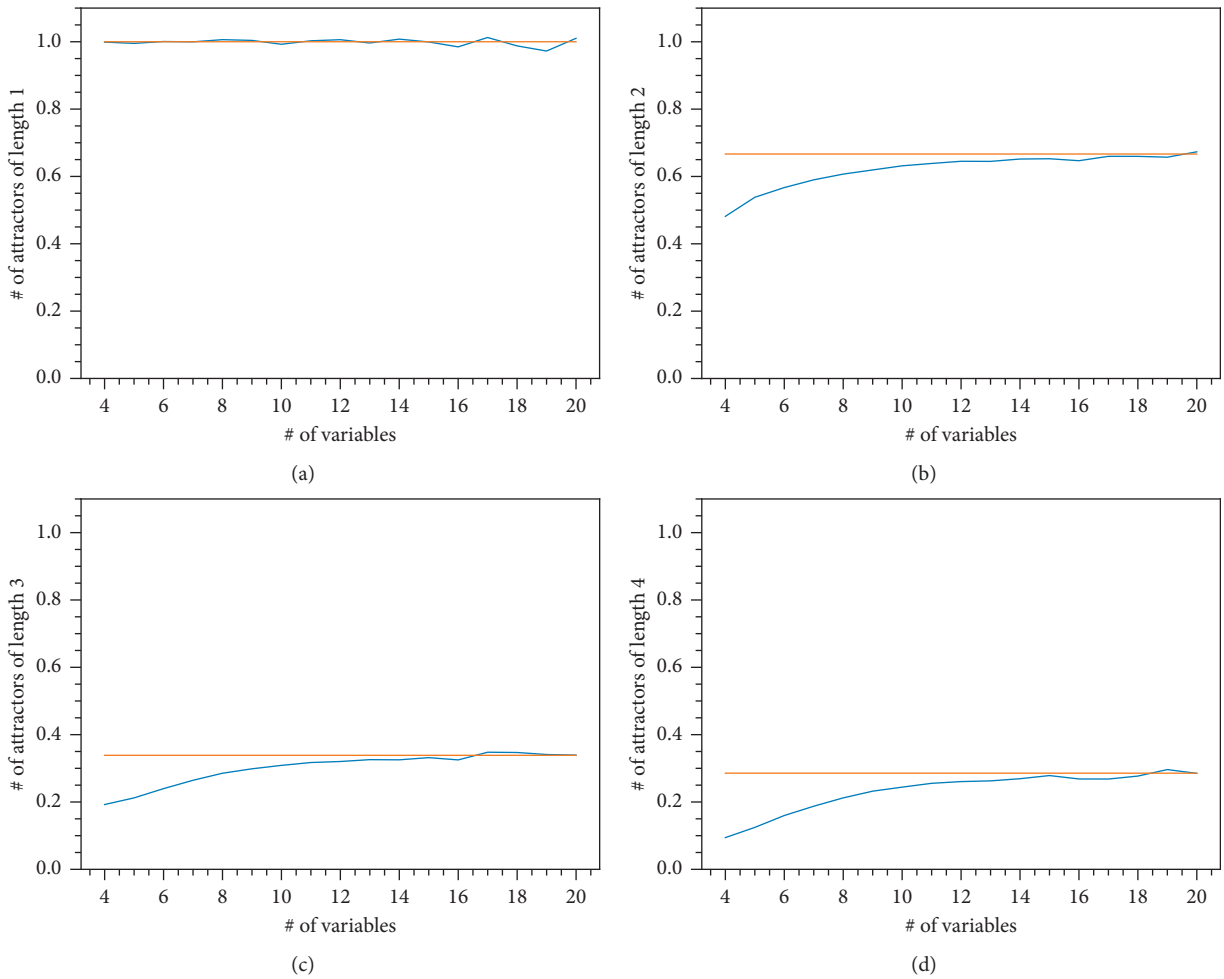

FIGURE 4: The average number of attractors of fixed length (blue plot) compared to the limiting value from Theorem 2 (orange plot). (a) Length 1. (b) Length 2. (c) Length 3. (d) Length 4.

In: Nonnegative integers k and n with $k \leq n$
Out: A Boolean function f in n variables of canalizing depth k such that, for fixed k and n, all possible outputs have the same probability
(1) In the notation of Theorem 1, generate the following:
(a) random bits $b, a_1, \ldots, a_n \in \{0, 1\}$;
(b) random subset $X \subset \{x_1, \ldots, x_n\}$ with $|X| = k$;
(c) random ordered partition $X = X_1 \sqcup \cdots \sqcup X_r$ of X (using Algorithm 2);
(d) random noncanalizing function $p_C \not\equiv 0$ in variables $\{x_1, \ldots, x_n\}/X$ (see Remark 3).
(2) Form a function $f(x_1, \ldots, x_n)$ using the data generated in Step 1 as in Theorem 1, where M_i involves exactly the variables from X_i for every $1 \leq i \leq r$.
(3) If f does not satisfy any of the conditions (E1) or (E2), discard it and run the algorithm again. Otherwise, return f.

Algorithm 2: Generating a random Boolean function of a given canalizing depth.

In: A finite set X with $|X| = k$
Out: An ordered partition $X = X_1 \sqcup \cdots \sqcup X_r$ into nonempty subsets $X_1, \ldots, X_r$ such that, for a fixed X, all possible outputs have the same probability
(1) Compute $p_0, \ldots, p_k$, where p_i is the number of ordered partitions of a set of size i, using the recurrence $p_j = \sum_{i=0}^{j-1} \binom{j}{i} p_{j-i}$, $p_0 = 1$ (see [19, equation (9)]).
(2) Generate an integer N uniformly at random from $[1, p_k]$.
(3) Find the minimum integer j between 1 and k such that $\sum_{i=0}^{j-1} \binom{k}{i} p_{k-i} \geq N$.
(4) Randomly select a subset $X_1 \subset X$ of size j.
(5) Generate an ordered partition $X_2 \sqcup \cdots \sqcup X_r$ of X/X_1 recursively.
(6) Return $X_1 \sqcup \cdots \sqcup X_r$.

Algorithm 3: Generating a random ordered partition of a given finite set.

6.2. Distributions of $N(\mathbf{f})$ and $AS(\mathbf{f})$. Figure 2 shows the empirical distributions of $N(\mathbf{f})$ and $\mathrm{AS}(\mathbf{f})$ for $n = 12$ and $k = 0, 1, 3, 12$ based on 300,000 samples for each k. From the plot, we can make the following observations:

(i) The distributions become more concentrated and the peak shifts towards zero when k increases
(ii) The distributions for nonzero canalizing depths (especially for larger depths) are much closer to each other that to the distribution for zero canalizing depth. This agrees with the plots on Figure 1.

6.3. Relative Decreases. From Figure 1, we can observe that, for both $N(\mathbf{f})$ and $\mathrm{AS}(\mathbf{f})$, the sample mean decreases rapidly for small canalizing depths. In order to understand how this decrease behaves for large n, we introduce

$$N_k(n) := \frac{\text{the sample mean of } N(f) \text{ for } n \text{ variables and canalizing depth } k}{\text{the sample mean of } N(f) \text{ for } n \text{ variables and canalizing depth } 0}. \tag{11}$$

$\mathrm{AS}_k(n)$ is defined analogously. Figure 3 plots $N_1(n)$, $N_2(n)$, $N_3(n)$, and $N_n(n)$ and $\mathrm{AS}_1(n)$, $\mathrm{AS}_2(n)$, $\mathrm{AS}_3(n)$, and $\mathrm{AS}_n(n)$ as functions of n. From the plots we see that

(i) The relative initial decrease from canalizing depth 0 to canalizing depth 1 becomes even more substantial when n increases
(ii) The relative decrease from canalizing depth 0 to canalizing depth 3 is already very close to the relative decrease from depth zero to the maximal depth (i.e., nested canalizing functions)

7. Theory: The Main Result

We will introduce notation needed to state the main theorem. Let us fix a positive integer ℓ. For a binary string $\alpha \in S := \{0, 1\}^\ell$, we define

(i) $|\alpha|$ denotes the number of ones
(ii) $\overline{\alpha}$ denotes component-wise negation
(iii) $s(\alpha)$ denotes a cyclic shift to the right

For binary strings $\alpha, \beta \in \{0, 1\}^\ell$, we define

$$f(\alpha,\beta) := \begin{cases} \frac{1}{2^{|\beta|}}, & \text{if } \alpha \vee \beta = \beta, \\ 0, & \text{otherwise,} \end{cases}$$

$$g(\alpha,\beta) := \frac{1}{4}\left(f(\alpha,\beta) + f(\overline{\alpha},\beta) + f(\alpha,\overline{\beta}) + f(\overline{\alpha},\overline{\beta})\right). \tag{12}$$

Then, we define a $2^\ell \times 2^\ell$ matrix G_ℓ by

$$(G_\ell)_{a,b} = g(a, s(b)), \tag{13}$$

where we interpret numbers $1 \le a$ and $b \le 2^\ell$ as binary sequences of length ℓ.

Theorem 2. *Let A_ℓ be the limit of the expected number of attractors of length ℓ in a random Boolean network of canalizing depth one (see Definition 6) when the number of variables n goes to infinity. Then,*

$$A_\ell = \frac{1}{\ell P'_{G_\ell}(1)}, \tag{14}$$

where P_{G_ℓ} is the characteristic polynomial of matrix G_ℓ introduced above. In particular, we have

$$\begin{aligned} A_1 &= 1, \\ A_2 &= \frac{2}{3} = 0.666\ldots, \\ A_3 &= \frac{64}{189} = 0.3386\ldots, \\ A_4 &= 0.2856\ldots, \\ A_5 &= 0.2004\ldots, \\ A_6 &= 0.1721\ldots. \end{aligned} \tag{15}$$

Remark 4. The plots below show that the result of Theorem 2 agrees with our simulations (Figure 4).

Remark 5. As explained in Remark A.1, Theorem 2 stills holds if we replace a random Boolean network of canalizing depth one with a random Boolean network defined by canalizing functions.

Example 4. Let $\ell = 2$. Then, for example, we have $f(0,2) = f(0,1) = 1/2$ and $g(0,1) = g(3,1) = 1/4$. In total, we have

$$G_2 = \begin{pmatrix} 3/8 & 1/4 & 1/4 & 3/8 \\ 1/8 & 1/4 & 1/4 & 1/8 \\ 1/8 & 1/4 & 1/4 & 1/8 \\ 3/8 & 1/4 & 1/4 & 3/8 \end{pmatrix}, \tag{16}$$

$$P_{G_2}(t) = t^4 - \frac{5}{4}t^3 + \frac{1}{2}t^2.$$

Remark 6. Theorem 2 and Corollary A.1 imply that $A_\ell > 1/\ell$ for every $\ell > 1$. On the other hand, a direct computation shows that the expected number of attractors of length ℓ in a random Boolean network (without any canalization requirements) is $1/\ell$. This is consistent with our observations from Section 6.1.

Remark 7. A sage script for computing numbers A_ℓ is available at https://github.com/MathTauAthogen/Canalizing-Depth-Dynamics/blob/master/core/theory.sage.

8. Conclusions

We conducted computational experiments to investigate the attractor structure of Boolean networks defined by functions of varying canalizing depth. We observed that networks with higher canalizing depth tend to have fewer attractors and the sizes of the attractors decrease dramatically when the canalizing depth increases moderately. As a consequence, the basins tend to grow when the canalizing depth increases. These properties are desirable in many biological applications of Boolean networks, so our results give new indications of the biological utility of Boolean networks defined by functions of positive canalizing depth.

We proved a theoretical result, Theorem 2, which complements the above observation as follows. The theorem implies that a large random Boolean network of canalizing depth one has on average more attractors of small size than a random Boolean network of the same size although it has less attractors in total. This also provides an explanation to the fact that the total size of attractors decreases faster than the number of attractors as the canalizing depth grows.

Furthermore, we observed that all the statistics we computed are almost the same in the case of the maximal possible canalizing depth (so-called nested canalizing Boolean networks) and in the case of moderate canalizing depth. This agrees with the results of Layne et al. [12]. This observation elucidates an interesting and powerful feature of canalization: even a very moderate canalizing influence in a Boolean network has a strong constraining influence on network dynamics. It would be of interest to explore the

prevalence of these features in published Boolean network models.

Finally, we provided evidence that the observed phenomena will occur for Boolean networks with larger numbers of state variables.

Appendix

A. Proofs

Notation A.1. We fix a positive integer ℓ.

(i) For every $1 \le i < j \le \ell$, we define a subset $S_{i,j} \subset S = \{0,1\}^\ell$ by

$$S_{i,j} \coloneqq \left\{(\alpha_1, \ldots, \alpha_\ell) \in S \,\middle|\, \alpha_i = \alpha_j\right\}. \tag{A.1}$$

(ii) For every $1 \le i < j \le \ell$, let $G_{\ell;i,j}$ be the submatrix of G_ℓ with rows and columns having indices from $S_{i,j}$.

Lemma A.1. *For every ℓ, we have*

(1) *G_ℓ^T is stochastic (see [20, § 8.5]), and G_ℓ has exactly one eigenvalue being equal to 1.*

(2) *For every $1 \le i < j \le \ell$, there exists a $2^{\ell-1} \times 2^{\ell-1}$ matrix $C_{\ell;i,j}$ with nonnegative entries such that $(2^{\ell+2}/2^{\ell+2} - 1)(G_{\ell;i,j} + C_{\ell;i,j})^T$ is stochastic and has exactly one of the eigenvalues being equal to 1.*

Proof. We will first show that G_ℓ^T is stochastic and irreducible (see [20, § 3.11]).

By definition, showing that G_ℓ^T is stochastic is equivalent to proving that, for every $\beta \in S \coloneqq \{0,1\}^\ell$:

$$\sum_{\alpha \in S} g(\alpha, s(\beta)) = 1. \tag{A.2}$$

Since shift just permutes binary strings, this sum is equal to $\sum_{\beta \in S} g(\alpha, \beta)$. For a fixed β and $k \le |\beta|$, the number of $\alpha \in S$ such $\alpha \vee \beta = \beta$ and $|\alpha| = k$ is equal to $\binom{|\beta|}{k}$. Thus,

$$\sum_{\alpha \in S} h(\alpha, \beta) = \sum_{k=0}^{|\beta|} \binom{|\beta|}{k} \frac{1}{2^{|\beta|}} = 1 \Longrightarrow \sum_{\beta \in S} g(\alpha, \beta) = 1. \tag{A.3}$$

To prove irreducibility, we observe that, if $0 \in S$ denotes a zero binary string, then $g(\alpha, 0) \neq 0$ and $g(0, \alpha) \neq 0$ for every $\alpha \in S$. Then, [20, § 3.11, Exercise 12a] implies that G_ℓ^T is irreducible.

Since G_ℓ^T is stochastic, its largest eigenvalue is equal to 1 [20, § 8.5, p.156]. Since G_ℓ^T is irreducible, [20, Theorem 8.2] implies that 1 is a simple eigenvalue.

To prove the second part of the lemma, we fix $1 \le i < j \le \ell$. We will show that for every $\beta \in S_{i,j}$

$$\sum_{\alpha \in S_{i,j}} g(\alpha, s(\beta)) \le \frac{2^{\ell+2} - 1}{2^{\ell+2}}. \tag{A.4}$$

Indeed, let γ be a binary string with all zeroes and one at the ith position. Then, since $g(\gamma, s(\beta)) \ge (1/2^{|\beta|+2}) \ge (1/2^{\ell+2})$, we have

$$\sum_{\alpha \in S_{i,j}} g(\alpha, s(\beta)) \le \left(\sum_{\alpha \in S} g(\alpha, s(\beta))\right) - g(\gamma, s(\beta)) \le 1 - \frac{1}{2^{\ell+2}}. \tag{A.5}$$

Inequality (A.4) implies that there exists a matrix $C_{\ell;i,j}$ with nonnegative entries such that $(2^{\ell+2}/2^{\ell+2} - 1)(G_{\ell;i,j} + C_{\ell;i,j})^T$ is stochastic.

Since $0 \in S_{i,j}$, the same argument as in the proof of the first part of the lemma shows that $(2^{\ell+2}/2^{\ell+2} - 1)(G_{\ell;i,j} + C_{\ell;i,j})^T$ is stochastic and has exactly one of the eigenvalues being equal to 1. □

Corollary A.1. *Let $P_\ell(t)$ be the charactersitic polynomial of G_ℓ. Then, for every $\ell > 1$, $|P'_\ell(1)| < 1$.*

Notation A.2. Fix a positive integer n. For vectors $\mathbf{a} = (a_1, \ldots, a_n) \in \mathbb{Z}_{\ge 0}^n$ and $\mathbf{b} = (b_1, \ldots, b_n) \in \mathbb{Z}_{\ge 0}^n$, we denote

$$\begin{aligned} \mathbf{a}! &\coloneqq a_1! \cdot \ldots \cdot a_n!, \\ \mathbf{a}^{\mathbf{b}} &\coloneqq a_1^{b_1} \cdot \ldots \cdot a_n^{b_n}, \\ |a| &\coloneqq a_1 + \cdots + a_n. \end{aligned} \tag{A.6}$$

Lemma A.2. *Let A be an $s \times s$ stochastic matrix with only one of the eigenvalues being one. We set*

$$C(A)_n \coloneqq \sum_{\substack{\mathbf{m} \in \mathbb{Z}_{\ge 0}^s \\ |\mathbf{m}| = n}} \frac{(A\mathbf{m})^{\mathbf{m}}}{n^n}. \tag{A.7}$$

Let $P_A(t)$ be the characteristic polynomial of A. Then, $\lim_{n \longrightarrow \infty} C(A)_n = 1/P'_A(1)$.

Proof. We recall that the Lambert W function [21] is the principal branch of the inverse of xe^x. We will use the notation $y(z) = -W(-z)$ from [22] so that $y(z) = ze^{y(z)}$. Function $y(z)$ has a singularity of the square-root type at $z = 1/e$ and has the following expansion around this point (see [22, p. 107]):

$$y(z) = 1 - \varepsilon + \frac{1}{3}\varepsilon^2 - \cdots, \quad \text{where } \varepsilon = \sqrt{2 - 2ez}. \tag{A.8}$$

From this, we obtain

$$\frac{1}{y(z)} = 1 + \varepsilon + \frac{2}{3}\varepsilon^2 - \cdots, \quad \text{where } \varepsilon = \sqrt{2 - 2ez}. \tag{A.9}$$

The main result of [23] implies that, for every complex $s \times s$ matrix A, we have

$$\sum_{\mathbf{m} \in \mathbb{Z}_{\ge 0}^s} \frac{(A\mathbf{m})^{\mathbf{m}}}{\mathbf{m}!} x^{|\mathbf{m}|} \exp\left(-x \sum_{i,j} m_j a_{i,j}\right) = \frac{1}{\det|E - xA|}. \tag{A.10}$$

Since A^T is stochastic, we have $\sum_{i=1}^{n} a_{i,j} = 1$, so

$$\sum_{\mathbf{m}\in\mathbb{Z}^s_{\geq 0}} \frac{(A\mathbf{m})^{\mathbf{m}}}{\mathbf{m}!} x^{|\mathbf{m}|} e^{-x|\mathbf{m}|} = \frac{1}{\det|E - xA|}. \tag{A.11}$$

If we perform a substitution $x = y(z)$ and use the definition of the Lambert W function, we obtain

$$\sum_{\mathbf{m}\in\mathbb{Z}^s_{\geq 0}} \frac{(A\mathbf{m})^{\mathbf{m}}}{\mathbf{m}!} z^{|\mathbf{m}|} = \frac{1}{\det|E - y(z)A|}. \tag{A.12}$$

From this, we obtain

$$\sum_{n=0}^{\infty} \frac{n^n C(A)_n}{n!} z^n = \sum_{\mathbf{m}\in\mathbb{Z}^s_{\geq 0}} \frac{(A\mathbf{m})^{\mathbf{m}}}{\mathbf{m}!} z^{|\mathbf{m}|} = \frac{1}{\det|E - y(z)A|} =: F(z). \tag{A.13}$$

$F(z)$ can be rewritten as

$$F(z) = \frac{1}{y(z)^s P_A(1/y(z))}. \tag{A.14}$$

Finding the asymptotic behavior of the Taylor coefficients of $F(z)$ would yield an asymptotic for $C(A)_n$. We will do this using singularity analysis [24, Chapter VI] (similarly to [22, Theorem 2]). Since $|y(z)| < 1$ for $|z| < 1/e$ (see [21, Figure 1]) and all roots of P_A lie in the unit circle due to the stochasticity of A, $1/e$ is the singularity of $F(z)$ with the smallest absolute value. Due to Lemma A.1, $P_A(t) = (1-t)Q_A(t)$, where $Q_A(1) \neq 0$. Using (A.8), we obtain the following expansion of $F(z)$ around $1/e$:

$$\begin{aligned} F(z) &= \frac{1}{(1-\varepsilon+\cdots)^s(-\varepsilon - 2/3\varepsilon^2 + \cdots)Q_A(1+\varepsilon+\cdots)} \\ &= \frac{-1}{Q_A(1)}\left(\frac{1}{\varepsilon} + \cdots\right), \text{ where } \varepsilon = \sqrt{2-2ez}. \end{aligned} \tag{A.15}$$

Singularity analysis [24, Corollary VI.1] implies that

$$\frac{n^n C(A)_n}{n!} \sim \frac{-e^n}{Q_A(1)\sqrt{2\pi n}}, \quad \text{as } n \longrightarrow \infty. \tag{A.16}$$

Using Stirling's formula, we get

$$C(A)_n \sim \frac{-n!e^n}{n^n Q_A(1)\sqrt{2\pi n}} \sim \frac{-1}{Q_A(1)}, \quad \text{as } n \longrightarrow \infty. \tag{A.17}$$

Using $P'_A = -Q'_A + (1-t)Q_\ell$, we deduce $P'_A(1) = -Q'_A(1)$, and this finishes the proof. □

Lemma A.3. *On the set of all Boolean networks with n states consider two probability distributions:*

(*A*) *All the networks with canalizing depth one have the same probability, and all others have probability zero*

(*B*) *ie probability assigned to each network is proportional to the product of the number of canalizing variables of the functions defining this network*

We fix a positive integer ℓ. By $A_{\ell,n}$ and $B_{\ell,n}$ we denote the average number of attractors of length ℓ in a random Boolean network with n states with respect to distributions (A) and (B), respectively. Then,

$$\lim_{n\longrightarrow\infty} A_{\ell,n} = \lim_{n\longrightarrow\infty} B_{\ell,n}. \tag{A.18}$$

Example A.1. We will illustrate the (*B*) distribution by an example. Consider the following three networks with two states:

$$\begin{aligned} \mathbf{f}_1 &= (x_1x_2 + 1, x_1 + x_2), \\ \mathbf{f}_2 &= (x_1x_2, x_1), \\ \mathbf{f}_3 &= (x_1x_2 + 1, x_1x_2). \end{aligned} \tag{A.19}$$

Since the canalizing depth of $x_1 + x_2$ is zero, $P_B(\mathbf{f}_1)$, the probability of $\mathbf{f}_1$ with respect to B, is zero. Since the canalizing depths of x_1x_2 and x_1 are 2 and 1, respectively, the ratio $P_B(\mathbf{f}_2)/P_B(\mathbf{f}_3)$ is equal to $(2 \cdot 1/2 \cdot 2) = (1/2)$.

Proof. Let F_n and F_n^* be the number of Boolean functions in n variables with canalizing depth exactly one and more than one, respectively. We will use the following bounds:

(1) $F_n^* \leq n^2 \cdot 4 \cdot 4 \cdot 2^{2^{n-2}}$: we look term-by-term. There are at most n^2 ways to choose first and second canalizing variables. There are at most 4 choices for the canalizing outputs and at most 4 choices for canalizing values for these two variables. There are at most $2^{2^{n-2}}$ core functions, since that is all possible functions, which may or may not be canalizing. Since redundant arrangements of canalizing variables are not accounted for, this must overcount.

(2) $F_n \geq 2^{2^{n-1}} - (n-1) \cdot 2 \cdot 2 \cdot 2^{2^{n-2}}$: this is a lower bound for the number of noncanalizing core function in $n-1$ variables because $(n-1) \cdot 2 \cdot 2 \cdot 2^{2^{n-2}}$ is an upper bound on the number of canalizing functions in $n-1$ variables (obtained in the same way as the bound above).

We also introduce

$$R_n := \frac{F_n^*}{F_n} \leq \frac{16n^2 2^{2^{n-2}}}{2^{2^{n-1}} - 4(n-1)2^{2^{n-2}}} = \frac{n^2}{2^{(2^{n-2})-4} - (1/4)(n-1)}. \tag{A.20}$$

For X being (*A*) or (*B*) and positive integer n, let $P_{X,n}$ denote the probability (it is always the same) of choosing a network from distribution X with all functions being of depth exactly one. Let P_n^* be the maximal probability of choosing a network from (*B*) with at least one function being of depth more than one, respectively. By S_n and S_n^* we denote the total number of attractors of length ℓ in networks with all functions being of depth exactly one and with at least one function being of depth more than one, respectively.

The statement of the lemma is equivalent to the statement that

$$\lim_{n\longrightarrow\infty} \left(A_{\ell,n} - B_{\ell,n}\right) = 0. \tag{A.21}$$

Using the notation introduced above, we can bound $A_{\ell,n} - B_{\ell,n}$ as

$$P_{n,A}S_n - P_{n,B}S_n - P_n^*S_n^* \le \left|A_{\ell,n} - B_{\ell,n}\right| \le P_{n,A}S_n + P_{n,B}S_n. \tag{A.22}$$

We set $U_n := S_n(P_{n,A} - P_{n,B})$ and $V_n := P_n^*S_n^*$. Then, (A.21) would follow from $\lim_{n\longrightarrow\infty}U_n = 0$ and $\lim_{n\longrightarrow\infty}V_n = 0$, so we will prove these two equalities.

Since any network has at most 2^n attractors of length ℓ, $S_n \le 2^nF_n^n$. Since the total sum of the products of canalizing depths over all Boolean networks does not exceed $(F_n + nF_n^*)^n$, we have $P_{n,B} \ge (1/(F_n + nF_n^*)^n)$. Since $P_{n,A} = 1/F_n^n$, we have

$$U_n \le 2^nF_n^n\left(\frac{1}{F_n^n} - \frac{1}{(F_n + nF_n^*)^n}\right) = 2^n\left(1 - \frac{1}{(1+nR_n)^n}\right)$$

$$= 2^n\frac{\binom{n}{1}nR_n + \binom{n}{2}(nR_n)^2 + \cdots + (nR_n)^n}{(1+nR_n)^n}. \tag{A.23}$$

(A.20) implies that $nR_n < 1$ for large enough n. Hence, for large enough n, we have

$$U_n \le 2^nnR_n\frac{2^n}{(1+nR_n)^n} \le 4^nnR_n \le \frac{4^nn^3}{2^{(2^{n-2})-4} - (1/4)(n-1)} \longrightarrow 0. \tag{A.24}$$

By similar arguments, $P_n^* \le n^n/F_n^n$ and $S_n^* \le 2^nn(F_n + F_n^*)^{n-1}F_n^*$, so

$$V_n \le 2^nn^{n+1}(F_n + F_n^*)^{n-1}F_n^*\frac{1}{F_n^n} \le 2^nn^{n+1}(1+R_n)^{n-1}R_n. \tag{A.25}$$

Since $R_n < 1$ for large enough n, using (A.20), we have

$$V_n \le 2^{2n-1}n^{n+1}R_n \le \frac{2^{2n-1}n^{n+3}}{2^{(2^{n-2})-4} - (1/4)(n-1)} \longrightarrow 0. \tag{A.26}$$

□

Remark 8. The proof of Lemma A.3 will be valid if we replace distribution (*B*) with any other distribution (*C*) such that, for every Boolean network $\mathbf{f} = (f_1, \ldots, f_n)$

(i) If at least one of f_i's is noncanalizing, $P_C(\mathbf{f}) = 0$

(ii) There exists a constant $P_{n,C}$ such that, if the canalizing depth of every f_i is one, then $P_C(\mathbf{f}) = P_{n,C}$

(iii) We have $(P_C(\mathbf{f})/P_{n,C}) \le (P_B(\mathbf{f})/P_{n,B})$ (using notation from the proof of Lemma A.3)

The above properties hold, for example, for the following distribution.

(C) All the networks defined by canalizing functions have the same probability, and all others have probability zero.

Using this distribution instead of (*A*), we see that Theorem 2 holds also for a random Boolean network defined by canalizing functions.

Lemma A.4. *We will use Notation A.1 and notation from Lemma A.2. Then, for every positive integers ℓ and n, we have*

$$C(G_\ell)_n - \sum_{1\le i<j\le\ell} C\left(G_{\ell;i,j}\right)_n \le \ell B_{\ell,n} \le C(G_\ell)_n. \tag{A.27}$$

Proof. We fix n. Consider a tuple $\mathbf{X} = (X_1, \ldots, X_\ell)$ of ℓ distinct elements of $\{0,1\}^n$. For $1 \le i \le n$, we denote $\mathbf{X}_i := (X_{1,i}, \ldots, X_{n,i})$. For $\alpha \in S$, let

$$n_\alpha := \left|\left\{i \mid 1 \le i \le n, X_i = \alpha\right\}\right|. \tag{A.28}$$

Then, $\sum_{\alpha\in S}n_\alpha = n$. First, we will show that

$$P(X_1, \ldots, X_\ell \text{ form an attractor in this order}) = \prod_{\alpha\in S}\left(\sum_{\beta\in S}g(\alpha, s(\beta))\frac{n_\beta}{n}\right)^{n_\alpha} = \frac{(G_\ell\mathbf{n})^{\mathbf{n}}}{n^n}, \tag{A.29}$$

where $\mathbf{n} = (n_0, n_1, \ldots, n_{2^\ell-1})$.

To prove (A.29), we will use that the functions f_i ($i = 1, \ldots, n$) in the network are chosen independently to decompose the left-hand side as

$$P(X_1, \ldots, X_\ell \text{ form an attractor in this order}) = \prod_{i=1}^{n}P\left(f_i(X_j) = X_{j+1,i} \text{ for every } 1 \le j \le n\right), \tag{A.30}$$

where we use notation $X_{n+1} = X_1$ and the probability of each Boolean function to be chosen is assumed to be proportional to the number of its canalizing variables. We show that, for every $1 \le i \le n$,

$$P\left(f_i(X_j) = X_{j+1,i} \quad \text{for every } 1 \le j \le n\right) = \sum_{\beta\in S}g(\mathbf{X}_i, s(\beta))\frac{n_\beta}{n}. \tag{A.31}$$

Then, (A.29) would follow from multiplying (A.31) for all *i*. To prove (A.31), without loss of generality, we consider $i = 1$. Consider a set

$$\Omega = \left\{(f,k) \mid f : \{0,1\}^n \longrightarrow \{0,1\}, 1 \le k \le n, x_k \text{ is canalizing for } f\right\}, \tag{A.32}$$

with a uniform probability distribution P_Ω. Observe that for a function f with canalizing variables $x_{k_1}, \ldots, x_{k_s}$, we have

$$P(f) = P_\Omega((f,k_1)) + \cdots + P_\Omega((f,k_s)). \tag{A.33}$$

If we can show that, for every $1 \le k \le n$,

$$P_\Omega\left(f(X_j) = X_{j+1,1} \quad \text{for every } 1 \le j \le n \mid (f,k) \in \Omega\right) = g(\mathbf{X}_1, s(\mathbf{X}_k)), \tag{A.34}$$

then (A.31) would follow by summing up (A.34) over all k and using the law of total probability.

We consider one of the canalizing variables of f, say, x_k. Let c be the canalizing value of x_{k_1}, and let v be the value taken by f when $x_{k_1} = c$. Then, $(c, v) \in \{0,1\}^2$, and all these

four cases have the same probability due to the symmetry. As $g(\alpha, s(\beta)) = (1/4)(h(\alpha,\beta) + h(\overline{\alpha},\beta) + h(\alpha,\overline{\beta}) + h(\overline{\alpha},\overline{\beta}))$, it is sufficient to show that

$$P_\Omega\left(f\left(X_j\right) = X_{j+1,1} \quad \text{for every } 1 \le j \le n \mid (f,k) \in \Omega \text{ and } c = v = 0\right) = h\left(\mathbf{X}_1, s\left(\mathbf{X}_k\right)\right), \tag{A.35}$$

and then sum for all $(c, v) \in \{0,1\}^2$.

To prove (A.35), consider any j, say $j = 1$. There are then 4 cases for the values of $X_{1,k}$ and $X_{2,1}$:

(1) $X_{1,k} = 1$ and $X_{2,1}$ is 0 or 1. With probability 1/2, we have $f(X_1) = X_{2,1}$. This is true due to symmetry, as for any f_1 which takes on the value w at X_1, we can produce another function g that is equal to 0 if $X_{1,k} = 0$ and $\overline{f}_1$ if $X_{1,k} = 1$. Then, $g(X_1) = \overline{w}$.

(2) $X_{1,k} = 0$ and $X_{2,1} = 1$. Since $X_{1,k} = c$, the probability of $f(X_1) = X_{2,1} \neq v = 0$ is zero.

(3) $X_{1,k} = X_{2,1} = 0$. Since $X_{1,k} = c$ and $X_{2,1} = v$, the canalization property implies that $f(X_1) = X_{2,1}$ with probability one.

The only case in which $\mathbf{X}_1 \vee s(\mathbf{X}_k) \neq s(\mathbf{X}_k)$ is where there is at least one j such that case 2 is realized. In this case, the probability in the left-hand side of (A.35) will be zero. Otherwise, each occurrence of case 1 will multiply the total probability by 1/2 and each occurrence of case 3 will multiply the total probability by 1. Thus, we show that the left-hand side of (A.35) is indeed equal to $h(\mathbf{X}_1, s(\mathbf{X}_k))$. This finishes the proof of (A.29).

To finish the proof of the lemma, we set

$$U := \left\{ \mathbf{n} \in \mathbb{Z}_{\ge 0}^S \,\middle|\, \sum_{\alpha \in S} n_\alpha = n \;\&\; \text{the support of } \mathbf{n} \text{ does not belong to } \bigcup_{1 \le i < j \le \ell} S_{i,j} \right\}. \tag{A.36}$$

Summing (A.29) over all ℓ-tuples $(X_1, \ldots, X_\ell)$ of distinct elements of $\{0,1\}^n$, we obtain (see (A.7))

$$\ell B_{\ell,n} = \sum_{\mathbf{n} \in U} \frac{(G_\ell \mathbf{n})^{\mathbf{n}}}{n^n} \le C(G_\ell)_n. \tag{A.37}$$

On the other hand, if $\mathbf{n}$ is supported on one some $S_{i,j}$, then $G_\ell \mathbf{n} = G_{\ell;i,j}\mathbf{n}|_{S_{i,j}}$, where $\mathbf{n}|_{S_{i,j}}$ denotes the restriction of $\mathbf{n}$ on the coordinates from $S_{i,j}$. This implies that

$$C(G_\ell)_n - \ell B_{\ell,n} \le \sum_{1 \le i < j \le \ell} C\left(G_{\ell;i,j}\right)_n. \tag{A.38}$$

This finishes the proof of the lemma. □

Proof of Theorem 2. We fix positive integer ℓ. In the notation of Lemma A.3, we have $A_\ell = \lim_{n \longrightarrow \infty} A_{\ell,n}$. Lemma A.3 implies that $A_\ell = \lim_{n \longrightarrow \infty} B_{\ell,n}$. We fix any $1 \le i < j \le \ell$, and let $C_{\ell;i,j}$ be the matrix given by Lemma A.1. We set $M := (2^{\ell+2}/2^{\ell+2} - 1)(G_{\ell;i,j} + C_{\ell;i,j})$. Then,

$$0 \le C\left(G_{\ell;i,j}\right)_n \le C\left(G_{\ell;i,j} + C_{\ell;i,j}\right)_n = \left(\frac{2^{\ell+2} - 1}{2^{\ell+2}}\right)^n C(M)_n. \tag{A.39}$$

Lemma A.2 implies that $\lim_{n \longrightarrow \infty} C(M)_n$ is finite, thus we have that $\lim_{n \longrightarrow \infty} C(G_{\ell;i,j})_n = 0$. We finish the proof of the theorem by considering the limit of (A.27) and applying Lemma A.2 to G_ℓ. □

B. Complexity analysis

Proposition B.1. *Complexity of Algorithm 3 is $O(k^3)$.*

Proof. First, we show the complexity of a single run the algorithm, i.e., not taking into account the recursive call, is $O(k^2)$. Since the first k rows of the Pascal's triangle can be precomputed in $O(k^2)$, the complexity of step 1 is also $O(k^2)$. Similarly, the complexity of step 3 is $O(k^2)$. It remains to observe that step 2 takes $O(1)$ and step 4 takes $O(k^2)$ (indeed, selecting a subset of size j amounts to selecting and removing j indices). In total, we obtain $O(k^2)$.

The depth of the recursion calls is at most k. Since the complexity of each single call is $O(k^2)$, so the total complexity is $O(k^3)$. □

Lemma B.1. *The average complexity of the algorithm for generating a function in $n > 0$ variables which is either 1 or noncanalizing described in Remark 3 is $O(n2^n)$.*

Proof. [25, p. 116] implies that the proportion of functions which are canalizing in n variables is bounded from above by $4n/2^{2^{n-1}}$. Note that [25] considers constant functions canalizing which we do not. Thus, the probability P_n of choosing a function which is either 1 or noncanalizing is bounded from above by

$$\frac{4n}{2^{2^{n-1}}} - \frac{1}{2^{2^n}} = \frac{4n - \left(1/2^{2^{n-1}}\right)}{2^{2^{n-1}}}. \tag{B.1}$$

This bound is less than 3/4 for all values of n except 1 and 2, but we can compute directly that $P_1 = 3/4$ and $P_2 = 13/16$. Therefore, the number of times the generation of a function needs to be repeated averages to $1/1 - P_n$, which does not exceed 4, so the average complexity of the whole procedure is the same as of a single generation step.

The complexity of a single step consists of generating a random function (which is $O(2^n)$) and checking whether it is canalizing or not. We perform this check by running linearly through the table for each variable, so the complexity is $O(n2^n)$ time. Thus, the total complexity is indeed $O(n2^n)$. □

Lemma B.2. *There is a constant $c < 1$ such that the probability that a function generated in steps 1 and 2 of Algorithm 2 does not satisfy one of the conditions (E1) or (E2) is bounded by c for every n.*

Proof. Notice that

$$P((E1) \text{ or } (E2) \text{ is false}) = P(r \neq 1)P((E1) \text{ is false} \mid r \neq 1) + P(r = 1)P((E2) \text{ is false} \mid r = 1). \tag{B.2}$$

We will show that there is a constant $c < 1$ such that $P((E1) \text{ is false} \mid r \neq 1)$ and $P(5.1 \text{ is false} \mid r = 1)$ do not exceed c.

(i) $P((E1) \text{ is false} \mid r \neq 1)$: the probability of having $k_r = 1$ (the only possible $k_r < 2$) is just the proportion of ordered partitions with a single element at the end. We can construct all of these by picking an element and then picking a partition of the remaining elements, so this creates $k \cdot p_{k-1}$ possibilities. Thus, the probability of this occurring is kp_{k-1}/p_k. [19, equation (5)] implies that this approaches $\ln(2) < 1$ as n goes to infinity. Thus, there exists such c.

(ii) $P((E2) \text{ is false} \mid r = 1)$: the probability of ever picking $b = 1$ is just 1/2, so we can take $c = 1/2$. □

Proposition B.2. *Complexity of Algorithm 2 is $O(n2^n)$.*

Proof. Lemma B.2 implies that the average number of reruns in step 3 is constant. Thus, the complexity of the algorithm is the same as of a single run.

Proposition B.1 and Lemma B.1 imply that the complexity of step 1 is $O(k^3 + (n-k)2^{n-k})$. Step 2 generates a truth table for the function. There are 2^n input-output pairs, and computing the function takes at most k steps, so this is $O(k2^n)$. In step 3, the conditions (E1) or (E2) are verified in $O(2^n)$ time.

Summing everything, we obtain $O(k^3 + (n-k)2^{n-k} + k2^n) = O(n2^n)$ □

Data Availability

Python/sage code and the results of simulations used to support the findings of this study have been deposited at https://github.com/MathTauAthogen/Canalizing-Depth-Dynamics.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors are grateful to Claus Kadelka, Christian Kratenthaler, and Doron Zeilberger for helpful discussions. GP was partially supported by NSF grants CCF-1564132, CCF-1563942, DMS-1760448, DMS-1853482, and DMS-1853650 by PSC-CUNY grants #69827-0047 and #60098-0048. RL was partially supported by Grants NIH 1U01EB024501-01 and NSF CBET-1750183. EP, GP, and WQ are grateful to the New York Math Circle, where their collaboration started.

References

[1] M. I. Davidich and S. Bornholdt, "Boolean network model predicts cell cycle sequence of fission yeast," *PLoS One*, vol. 3, no. 2, Article ID e1672, 2008.

[2] S. Kauffman, C. Peterson, B. Samuelsson, and C. Troein, "Random Boolean network models and the yeast transcriptional network," *Proceedings of the National Academy of Sciences*, vol. 100, no. 25, pp. 14796–14799, 2003.

[3] Z. Mai and H. Liu, "Boolean network-based analysis of the apoptosis network: irreversible apoptosis and stable surviving," *Journal of Theoretical Biology*, vol. 259, no. 4, pp. 760–769, 2009.

[4] J. Saez-Rodriguez, L. Simeoni, J. A. Lindquist et al., "A logical model provides insights into T cell receptor signaling," *PLoS Computational Biology*, vol. 3, no. 8, pp. 1–11, 2007.

[5] A. Veliz-Cuba and B. Stigler, "Boolean models can explain bistability in the lac operon," *Journal of Computational Biology*, vol. 18, no. 6, pp. 783–794, 2011.

[6] S. A. Kauffman, "Sequential DNA replication and the control of differences in gene activity between sister chromatids—a possible factor in cell differentiation," *Journal of Theoretical Biology*, vol. 17, no. 3, pp. 483–497, 1967.

[7] S. A. Kauffman, "Metabolic stability and epigenesis in randomly constructed genetic nets," *Journal of Theoretical Biology*, vol. 22, no. 3, pp. 437–467, 1969.

[8] S. Kauffman, "Homeostasis and differentiation in random genetic control networks," *Nature*, vol. 224, no. 5215, pp. 177-178, 1969.

[9] C. H. Waddington, "Canalization of development and the inheritance of acquired characters," *Nature*, vol. 150, no. 3811, pp. 563–565, 1942.

[10] F. Karlsson and M. Hörnquist, "Order or chaos in Boolean gene networks depends on the mean fraction of canalizing functions," *Physica A: Statistical Mechanics and Its Applications*, vol. 384, no. 2, pp. 747–757, 2007.

[11] S. Kauffman, C. Peterson, B. Samuelsson, and C. Troein, "Genetic networks with canalyzing Boolean rules are always stable," *Proceedings of the National Academy of Sciences*, vol. 101, no. 49, pp. 17102–17107, 2004.

[12] L. Layne, E. Dimitrova, and M. Macauley, "Nested canalyzing depth and network stability," *Bulletin of Mathematical Biology*, vol. 74, no. 2, pp. 422–433, 2012.

[13] S. Kauffman, *The Origins of Order: Self-Organization and Selection in Evolution*, Oxford University Press, Oxford, UK, 1993.

[14] A. Veliz-Cuba, B. Aguilar, F. Hinkelmann, and R. Laubenbacher, "Steady state analysis of Boolean molecular network models via model reduction and computational algebra," *BMC Bioinformatics*, vol. 15, no. 221, 2014.

[15] Q. He and M. Macauley, "Stratification and enumeration of boolean functions by canalizing depth," *Physica D: Nonlinear Phenomena*, vol. 314, no. 1–8, 2016.

[16] B. Samuelsson and C. Troein, "Superpolynomial growth in the number of attractors in kauffman networks," *Physical Review Letters*, vol. 90, no. 9, Article ID 098701, 2003.

[17] B. Drossel, "Number of attractors in random boolean networks," *Physical Review E*, vol. 72, no. 1, Article ID 016110, 2005.

[18] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, MIT Press, 3 edition, 2009.
[19] O. A. Gross, "Preferential arrangements," *The American Mathematical Monthly*, vol. 69, no. 1, pp. 4–8, 1962.
[20] D. Serre, *Matrices: Theory and Applications*, Springer, New York, NY, USA, 2 edition, 2010.
[21] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth, "On the Lambert W function," *Advances in Computational Mathematics*, vol. 5, no. 1, pp. 329–359, 1996.
[22] P. Flajolet, P. J. Grabner, P. Kirschenhofer, and H. Prodinger, "On Ramanujan's Q-function," *Journal of Computational and Applied Mathematics*, vol. 58, no. 1, pp. 103–116, 1995.
[23] L. Carlitz, "An application of MacMahon's Master theorem," *SIAM Journal on Applied Mathematics*, vol. 26, no. 2, pp. 431–436, 1974.
[24] P. Flajolet and R. Sedgewick, *Analytic Combinatorics*, Cambridge University Press, New York, NY, USA, 1 edition, 2009.
[25] C. C. Walker and A. E. Gelfand, "A system theoretic approach to the management of complex organizations: management by exception, priority, and input span in a class of fixed-structure models," *Behavioral Science*, vol. 24, no. 2, pp. 112–120, 1979.