On the Number of *NK*-Kauffman Networks Mapped into a Functional Graph

Federico Zertuche
Instituto de Matemáticas
Unidad Cuernavaca, Universidad Nacional Autónoma de México
Avenida Universidad 1001, Cuernavaca, Morelos 62210, Mexico

NK-Kauffman networks $\mathcal{L}_K^N$, where N is the number of Boolean variables and K the average number of connections, are studied. The K connections are random and chosen with equal probability, while the Boolean functions are randomly chosen with a bias p. The injectivity of the map $\Psi : \mathcal{L}_K^N \to \mathsf{G}_{2^N}$, where G_{2^N} is the set of functional graphs with 2^N nodes, is studied. In the asymptotic regime $N \gg 1$, it is found that a critical connectivity $K_c \sim O(\ln \ln N)$ exists such that Ψ is many-to-one for $K < K_c$ and injective for $K > K_c$. The analysis is extended when the tautology and contradiction Boolean functions are excluded from the construction of $\mathcal{L}_K^N$. For such a case, it is found that Ψ always remains injective.

1. Introduction

NK-Kauffman networks were proposed by Stuart A. Kauffman in 1969 as a starting point for a mechanism that mimics cell metabolic behavior, and also as a way to study the transition from disorder to order in living organisms [1, 2]. There exists an extensive literature about the subject, with analytical and computational calculations that have been dedicated to the subject and many extensions of it; see [2–6] and references therein.

Since the impact of the modern synthesis of evolutionary theory, numerous researchers have been expanding the structural elements to comprehend and theorize about evolutionary processes and then enhance abilities that test and measure them [7–9]. They have studied the role of epigenetics in speciation, and particularly, how they could improve variations and promote or repress fitness: in brief, they have addressed the question of the processes of development and evolutionary change at genomic levels [10, 11]. Of the utmost importance is the mapping of a set of genotypes to a set of phenotypes, called the *genotype-phenotype map*, which we will refer to as Ψ throughout the paper. It is difficult to comprehend the entire structure of Ψ; however, many properties are well known [12–16]. One of the most remark-

https://doi.org/10.25088/ComplexSystems.25.4.329

able is that Ψ should be a many-to-one map, due to the robustness of living organisms against mutations and/or replication errors during mating [13–16]. Nevertheless, Kauffman networks are very simplified models of real metabolic processes and give good insight into them. See, for example, the case studied by Kauffman about cellular differentiation [1, 2]. For these reasons, and also for the pure mathematical understanding of the dynamics of Kauffman networks, it is important to study the map Ψ, which is the main subject of this paper. Two approximations of Ψ follow from the structure of NK-Kauffman networks: (*i*) Ψ is not an stochastic map; and (*ii*) the map Ψ is a function in the mathematical sense. Let us go into the mathematical model.

Let $[N] = \{1, 2, \ldots, N\}$ denote the set of the first N natural numbers. A Kauffman network consists of a set of N Boolean variables $S_i(t) \in \mathbb{Z}_2$, with $i \in [N]$, which evolve deterministically and synchronously in a discrete time $t = 0, 1, 2, \ldots$ according to N different K-Boolean functions

$$b_K^{(i)} : \mathbb{Z}_2^K \to \mathbb{Z}_2, \quad i = 1, \ldots, N. \tag{1}$$

The evolution rule, at each time step t, given by

$$S_i(t+1) = b_K^{(i)}\left(S_{i_1}(t), S_{i_2}(t), \ldots, S_{i_K}(t)\right), \quad i = 1, \ldots, N, \tag{2}$$

where

$$C_K^{(\alpha_i)} = \{i_1, \ldots, i_K\} \subseteq [N]\alpha_i = 1, \ldots, \binom{N}{K},$$

constitutes the *K-connection set* of inputs at site i. The elements of $C_K^{(\alpha_i)}$ are selected randomly, with equal probability and without repetition. Also, each K-Boolean function $b_K^{(i)}$ is chosen randomly and independently, with a bias

$$0 < p < 1,$$

such that $b_K^{(i)} = +1$ with probability p, and $b_K^{(i)} = 0$ with probability $1 - p$, for each of its 2^K arguments.

Once this random selection is done, a Kauffman network has been constructed, and its subsequent dynamics is deterministically given by iterations of equation (2).

Let us rewrite the dynamical rule of equation (2) in a notation more suitable for counting purposes: to each K-connection set $C_K^{(\alpha_i)}$, we assign a K-connection function

$$C_K^{*(\alpha_i)} : \mathbb{Z}_2^N \longrightarrow \mathbb{Z}_2^K,$$

defined by

$$C_K^{*(\alpha_i)}(\mathbf{S}) = C_K^{*(\alpha_i)}(S_1, \ldots, S_N) = (S_{i_1}, \ldots, S_{i_K}) \; \forall \; S \in \mathbb{Z}_2^N.$$

Then equation (2) may be recast through the composition

$$\mathbb{Z}_2^N \xrightarrow{C_K^{*(\alpha_i)}} \mathbb{Z}_2^K \xrightarrow{b_K^{(i)}} \mathbb{Z}_2, \quad i = 1, \ldots, N,$$

which defines the *star functions*

$$b_K^{*(\alpha_i)} \equiv b_K^{(i)} \circ C_K^{*(\alpha_i)} : \mathbb{Z}_2^N \to \mathbb{Z}_2. \tag{3}$$

So, equation (2) is equivalent to

$$S_i(t+1) = b_K^{*(\alpha_i)}(S(t)), \quad i = 1, \ldots, N.$$

Let us denote by $\mathcal{L}_K^N$ the set of NK-Kauffman networks so defined. They constitute a subset of the Boolean endomorphisms

$$\mathcal{B}_N = \{\mathbf{F} : \mathbb{Z}_2^N \longrightarrow \mathbb{Z}_2^N\}$$

that is isomorphic to the set $\mathcal{G}_{2^N}$ of functional graphs with 2^N nodes, as shown in [5].

In the mathematical language of NK-Kauffman networks, Ψ is represented by

$$\Psi : \mathcal{L}_K^N \longrightarrow \mathcal{G}_{2^N}. \tag{4}$$

It associates its dynamics to each Kauffman network through the functional graph $g \in \mathcal{G}_{2^N}$ [6]. It is to be noted that only for the case $K = N$, Ψ is a bijection [3–6]. In Kauffman's models, $S_i = +1$ represents an active gene, while $S_i = 0$ represents a passive one, so Ψ represents the way in which Kauffman networks (genotypes) are translated into functional graphs (phenotypes) [6, 16].

In addition to biological applications, the mathematical understanding of Kauffman networks is important, as long as they are constructed with an elementary cellular automaton that has been well studied, and whose behavior is not a simple function of its Wolfram number [see equation (6b) in the next section].

In [6] a classification of Boolean functions was introduced as a way to study the injective properties of Ψ for the case $p = 1/2$ [5]. The purpose of this paper is to extend the analysis to the cases of general p and $p = 1/2$, but with the tautology and contradiction Boolean functions excluded from the construction of $\mathcal{L}_K^N$.

Once the exact combinatorial formulas have been obtained, the analysis is done in the asymptotic regime $N \gg 1$, the so-called thermodynamical limit. Calculations show that there is a critical connectivity that is not a constant but grows slowly as $K_c \sim O(\ln \ln N)$, with a width decreasing like $O(1/\ln \mathrm{N})$.

The paper is organized as follows: in Section 2, we introduce the formal mathematical concepts to be used, as well as the concept of *Boolean irreducibility* as a tool for making the combinatorial calculations. In Section 3, the equiprobable case $p = 1/2$, for the extraction of the K-Boolean functions in equation (1), is reviewed and introduced as a starting point for the generalization to the case $0 < p < 1$. In Section 4, the injective properties of Ψ for general p are studied. In Section 5, the case $p = 1/2$ with the tautology and contradiction Boolean functions excluded is studied. In Section 6, the conclusions are drawn, with implications for the context of theoretical biology. The Appendices are intended to provide mathematical support for the interested reader.

2. General Concepts

For the sake of simplicity, we will omit subindexes or superindexes, such as (i) or others, when not necessary. In what follows, we denote by $\oplus$ the addition modulo 2 for the elements of $\mathbb{Z}_2$, and by

$$\Gamma_K^N = \left\{ C_K^{(\alpha)} \subseteq [N] \,\middle|\, \sharp C_K^{(\alpha)} = K \right\}_{\alpha=1,\ldots,\binom{N}{K}}$$

the set of K-connections $C_K^{(\alpha)}$. For any K-Boolean function in equation (1),

$$b_K : \mathbb{Z}_2^K \to \mathbb{Z}_2,$$

its negation $\neg\, b_K$ is defined by

$$\neg\, b_K = b_K \oplus 1.$$

A well-known fact is that any K-Boolean function is completely determined by its *truth table*

$$\mathfrak{B}(b_K) = [\sigma_1, \sigma_2, \ldots, \sigma_{2^K}], \tag{5}$$

where $\sigma_s \in \mathbb{Z}_2$ is the s^{th} image of b_K given by

$$s = s(\mathbf{S}) = 1 + \sum_{i=1}^{K} S_i 2^{i-1}, \quad 1 \le s \le 2^K, \tag{6a}$$

and defines a *total order* among the 2^K possible arguments of b_K. The K-Boolean functions also may be classified and *totally ordered* according to Wolfram's notation, by a natural number n given by [17, 18]

$$n = \sum_{s=1}^{2^K} 2^{s-1} \sigma_s, \quad 0 \le n \le 2^{2^K} - 1. \tag{6b}$$

Of particular importance are the tautology $\tau_K \equiv b_K^{\left(2^{2^K}-1\right)}$ and contradiction $\neg\, \tau_K \equiv b_K^{(0)}$ K-Boolean functions, with images:

$$\mathcal{B}(\tau_K) = [\underbrace{1, 1, \ldots, 1}_{2^K}] \tag{7}$$

and

$$\mathcal{B}(\neg\, \tau_K) = [\underbrace{0, 0, \ldots, 0}_{2^K}]. \tag{8}$$

Let us add, when appropriate, a superscript n to each of the K-Boolean functions and denote by

$$\Xi_K = \left\{ b_K^{(n)} : \mathbb{Z}_2^K \longrightarrow \mathbb{Z}_2 \right\}_{n=0,1,\ldots,2^{2^K}-1} \tag{9}$$

its set.

Table 1 gives an example of the truth table for the 16 possible $K = 2$ Boolean functions listed according to their Wolfram number from equation (6b).

s	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
2	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
3	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
4	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1

Table 1. The truth table for the 16 $K = 2$ Boolean functions. In the left column are its inputs s as given by equation (6a). The following columns contain the outputs ordered by Wolfram's number n as given by equation (6b).

The average number of Kauffman networks $\vartheta(N, K, p)$ that Ψ maps into the same functional graph g is given from equation (4) by

$$\vartheta(N, K, p) = \frac{1}{\sharp\Psi\left(\mathcal{L}_K^N\right)} \sum_{g \in \Psi\left(\mathcal{L}_K^N\right)} \sharp\Psi^{-1}(g), \tag{10}$$

where

$$\Psi^{-1}(g) = \left\{ f \in \mathcal{L}_K^N \,\middle|\, \Psi(f) = g \right\}.$$

This implies

$$\Psi^{-1}(g) \cap \Psi^{-1}(g') = \emptyset \;\forall\; g \neq g'.$$

https://doi.org/10.25088/ComplexSystems.25.4.329

So, $\mathcal{L}_K^N$ may be disjointedly decomposed as

$$\mathcal{L}_K^N = \bigsqcup_{g \in \Psi(\mathcal{L}_K^N)} \Psi^{-1}(g),$$

so that

$$\sharp\mathcal{L}_K^N = \sum_{g \in \Psi(\mathcal{L}_K^N)} \sharp\Psi^{-1}(g).$$

Then, from equation (10)

$$\vartheta(N, K, p) = \frac{\sharp\mathcal{L}_K^N}{\sharp\Psi(\mathcal{L}_K^N)}. \tag{11}$$

The value of $\sharp\mathcal{L}_K^N$ is easily calculated, noting that for each site i, there are 2^{2^K} K-Boolean functions b_K and $\binom{N}{K}$ different K-connections. So, the number of different inputs $\mathfrak{I}$ at any site is

$$\mathfrak{I} = 2^{2^K}\binom{N}{K}, \tag{12a}$$

and then

$$\sharp\mathcal{L}_K^N = \left[2^{2^K}\binom{N}{K}\right]^N, \tag{12b}$$

for the N sites.

For the calculation of $\sharp\Psi(\mathcal{L}_K^N)$, we must take into account that for each site, not all the inputs of equation (12) give a different output, because there may be some repetitions $\mathfrak{R}$ due to the noninjectivity of Ψ. This is because not all of the b_K depend completely on their K arguments, as we will explain soon. So, from equation (12a), at any site the number of outputs Ω is

$$\Omega = \mathfrak{I} - \mathfrak{R} = 2^{2^K}\binom{N}{K} - \mathfrak{R}. \tag{13}$$

In order to calculate $\mathfrak{R}$, the concepts of *Boolean irreducibility* and *degree of irreducibility* were introduced in [6] as follows:

1. A K-Boolean function b_K is *irreducible* in its m^{th} argument S_m $(1 \leq m \leq K)$, if and only if there exists an input $(S_1, \ldots, S_m, \ldots, S_K) \in \mathbb{Z}_2^K$ such that
   $b_K(S_1, \ldots, S_m \oplus 1, \ldots, S_K) = b_K(S_1, \ldots, S_m, \ldots, S_K) \oplus 1.$
2. Otherwise, the K-Boolean function b_K is *reducible* in the argument S_m.

3. If b_K is irreducible on λ of its arguments and reducible in the remaining $K - \lambda$, then we say that it has a *degree of irreducibility* λ $(0 \leq \lambda \leq K)$. We denote by $\lambda(b_K)$ the function

   $\lambda : \Xi_K \to [K] \cup \{0\},$

   and by λ its value.

4. If $\lambda(b_K) = K$, b_K is called *totally irreducible*.

Table 2 shows the values of λ for the $K = 2$ Boolean functions.

s	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
λ	0	2	2	1	2	1	2	2	2	2	1	2	1	2	2	0
ω	0	1	1	2	1	2	2	3	1	2	2	3	2	3	3	4

Table 2. The degree of irreducibility λ and weight ω for the 16 $K = 2$ Boolean functions.

We define the sets of λ-irreducible functions by

$$\Upsilon_K(\lambda) = \{b_K \in \Xi_K \mid \lambda(b_K) = \lambda\}, \tag{14}$$

and denote by $\beta_K(\lambda)$ their cardinalities

$$\beta_K(\lambda) = \#\Upsilon_K(\lambda).$$

The values of $\beta_K(\lambda)$ were calculated in [19, 20]. In Appendix A, we include the calculation of them for the interested reader. Of particular importance are the values:

$$\beta_K(0) = 2 \tag{15}$$

corresponding to the set

$$\Upsilon_K(0) = \{\tau, \neg\,\tau\},$$

and

$$\beta_K(1) = 2K \tag{16}$$

corresponding to the set

$$\Upsilon_K(1) = \{b_K \in \Xi_K \mid b_K = S_i \vee b_K = S_i \oplus 1, \forall\, i \in [K]\}.$$

As an example, for the case $K = 2$ in Table 2, it can be seen that $\Upsilon_K(0)$ has as its elements the functions **0** $(\neg\,\tau)$ and **15** (τ), while $\Upsilon_K(1)$ has the functions **3** $(\neg\, S_2)$, **5** $(\neg\, S_1)$, **10** (S_1) and **12** (S_2).

In terms of equation (14), equation (9) may be disjointedly decomposed as

$$\Xi_K = \bigsqcup_{\lambda=0}^{K} \Upsilon_K(\lambda). \tag{17}$$

3. The Case $p = 1/2$

Now we have all the concepts to calculate $\mathfrak{R}$ in equation (13) for the case $p = 1/2$. For that scope, note that for any site, the star functions $b_K^{*(\alpha)} \equiv b_K \circ C_K^{*(\alpha)}$ in equation (3) may give the same output if the connection set $C_K^{(\alpha)}$ is changed to another connection set, depending on the degree of irreducibility of the b_K associated to $b_K^{*(\alpha)}$. Let $b_K^{*(\alpha)}$ be such that $\lambda(b_K) = \lambda$, and let $\mathcal{I}_\lambda^{(\alpha)} \equiv \{i_1^{(\alpha)}, \ldots, i_\lambda^{(\alpha)}\}$ be the set of subindexes of the irreducible arguments of b_K. Then

$$\Lambda_K^N(b_K^{*(\alpha)}, \lambda) = \{C_K^{(\beta)} \in \Gamma_K^N \mid \mathcal{I}_\lambda^{(\alpha)} \subseteq C_K^{(\beta)} \wedge C_K^{(\beta)} \neq C_K^{(\alpha)}\}$$

is the set of connections, apart from $C_K^{(\alpha)}$, such that $b_K^{*(\alpha)} : \mathbb{Z}_2^N \to \mathbb{Z}_2$ gives the same output $\forall\, \mathbf{S} \in \mathbb{Z}_2^N$. Note that any $C_K^{(\beta)} \in \Lambda_K^N(b_K^{*(\alpha)}, \lambda)$ has λ indexes fixed and $K - \lambda$ indexes free, which are elements in $[N] \setminus \mathcal{I}_\lambda^{(\alpha)}$. So $\sharp\Lambda_K^N(b_K^{*(\alpha)}, \lambda)$ is equal to the number of subsets of $[N] \setminus \mathcal{I}_\lambda^{(\alpha)}$ that have cardinality $K - \lambda$, minus one; then

$$\sharp\Lambda_K^N(b_K^{*(\alpha)}, \lambda) = \binom{N-\lambda}{K-\lambda} - 1.$$

So, the number of repetitions that a function with a λ degree of irreducibility produces is

$$\beta_K(\lambda)\left[\binom{N-\lambda}{K-\lambda} - 1\right],$$

giving in equation (13):

$$\Omega = 2^{2^K}\binom{N}{K} - \sum_{\lambda=0}^{K} \beta_K(\lambda)\left[\binom{N-\lambda}{K-\lambda} - 1\right]. \tag{18}$$

Then, we obtain

$$\sharp\Psi(\mathcal{L}_K^N) = \left\{2^{2^K}\binom{N}{K} - \sum_{\lambda=0}^{K} \beta_K(\lambda)\left[\binom{N-\lambda}{K-\lambda} - 1\right]\right\}^N, \tag{19}$$

and from equation (11)

$$\vartheta^{-1}\left(N, K, \frac{1}{2}\right) = \left\{\frac{2^{2^K}\binom{N}{K} - \sum_{\lambda=0}^{K} \beta_K(\lambda)\left[\binom{N-\lambda}{K-\lambda} - 1\right]}{2^{2^K}\binom{N}{K}}\right\}^N. \tag{20}$$

https://doi.org/10.25088/ComplexSystems.25.4.329

In the asymptotic regime $N \gg 1$, $K \sim O(\ln \ln N)$, we may apply Stirling's approximation for the factorial [21]

$$(N-K)! \approx \sqrt{2\pi}\, e^{-N} N^{N-K+1/2}.$$

For the combinatorial coefficients, this gives

$$\binom{N}{K} \approx \frac{N^K}{K!}. \tag{21}$$

Since

$$\binom{N-1}{K-1} = \frac{K}{N}\binom{N}{K}, \tag{22}$$

and from equation (14) or (A.1) $\beta_K(\lambda) < 2^{2^K}\ \forall\ \lambda$, we obtain the following asymptotic formula using equations (15) and (16):

$$\vartheta^{-1}\left(N, K, \frac{1}{2}\right) \approx \left\{1 - \frac{2}{2^{2^K}} - O\left(\frac{K^2}{N^2}\right)\right\}^N.$$

Then

$$\vartheta^{-1}\left(N, K, \frac{1}{2}\right) \approx e^{-N\varphi(K)}\left[1 + O\left(\frac{1}{N}\right)\right], \tag{23}$$

with

$$\varphi(K) \equiv \frac{2}{2^{2^K}}.$$

As we may see from equation (23), for $N \to \infty$, $\vartheta^{-1}(N, K, 1/2)$ goes to zero exponentially for $K =$ constant. However, if K grows with N in such a way that the product $N\varphi(K)$ remains constant as N grows, we may obtain values for equation (23) throughout the range $0 < \vartheta^{-1}(N, K, 1/2) \leq 1$. From equation (10), Ψ is a many-to-one function in the case where $\vartheta^{-1}(N, K, 1/2) \approx 0$, and an injective function if and only if $\vartheta^{-1}(N, K, 1/2) = 1$. From equation (20), injectivity is only obtained in the extreme case in which $K = N$, where $\mathcal{L}_K^N \cong \mathcal{G}_{2^N}$. As we are going to see, there is a critical connectivity K_c that grows with N, such that for $K < K_c$, $\vartheta^{-1}(N, K, 1/2) \approx 0$, making Ψ a many-to-one function. While for $K > K_c$, $\vartheta^{-1}(N, K, 1/2) \approx 1$, making Ψ approach an injective function. The critical connectivity K_c is defined by the equation

$$\vartheta^{-1}\left(N, K_c, \frac{1}{2}\right) = \frac{1}{2}$$

https://doi.org/10.25088/ComplexSystems.25.4.329

with the result [5, 6]

$$K_c \approx \log_2 \log_2\left(\frac{2N}{\ln 2}\right). \tag{24}$$

The width of the transition ΔK_c at K_c may be estimated by expanding $\vartheta^{-1}(N, K, 1/2)$ in Taylor series up to the first order in $K - K_c$:

$$\vartheta^{-1}\left(N, K, \frac{1}{2}\right) \approx \frac{1}{2}\left[1 - N\varphi'(K_c)(K - K_c)\right],$$

where $\varphi'(K) = d\varphi(K)/dK$. Then we may define

$$\Delta K_c \equiv K_1 - K_0,$$

where K_0 and K_1 are such that $\vartheta^{-1}(N, K_0, 1/2) = 0$ and $\vartheta^{-1}(N, K_1, 1/2) = 1$ in the first-order approximation. This gives

$$\Delta K_c = -\frac{2}{N\varphi'(K_c)} = \frac{2}{(\ln 2)^3 \log_2\left(\frac{2N}{\ln 2}\right)} \sim O\left(\frac{1}{\ln N}\right).$$

Equation (24) is in good agreement with the results of theoretical biology when N represents the number of genes of living organisms on Earth that have a typical value of $N \sim 10^4$, predicting a robust genotype-phenotype map Ψ for $K \lesssim 4$, where Ψ is a many-to-one map as expected in genetics [2, 12–16]. Moreover, the connectivity K corresponds in genetics to the average number of epistatic interactions that are known to be of order $K \sim 4$ [2, 13–15]. So Kauffman networks, notwithstanding their simplicity, give good insight into biological processes.

4. The Case for General *p*

Now we are going to extend the analysis to when the outputs of the K-Boolean functions $b_K^{(i)}$ are extracted with probability p that $b_K^{(i)} = +1$ for a given input. Then, the extraction probability of b_K is given by

$$\Pi_p(b_K) = \Pi_p \circ \omega(b_K) = p^{\omega}(1-p)^{2^K - \omega}, \tag{25}$$

where $\omega = 0, 1, \ldots, 2^K$ is the *weight* of b_K, defined by the *weight function*

$$\omega(b_K) = \sum_{s=1}^{2^K} \sigma_s,$$

with σ_s given by equation (5). The last row of Table 2 gives the values of ω for the $K = 2$ Boolean functions. Now $\sharp\Psi(\mathcal{L}_K^N)$ may be calculated by taking averages in equation (18). Multiplying equation (18) by equation (25), we have

$$\Omega\,\Pi_p(b_K) = 2^{2^K}\binom{N}{K}\Pi_p(b_K) - 2^{2^K}\sum_{\lambda=0}^{K}\Pi_p(b_K\,|\,\lambda)\left[\binom{N-\lambda}{K-\lambda} - 1\right],$$

where

$$\Pi_p(b_K\,|\,\lambda) = \frac{\beta_K(\lambda)}{2^{2^K}}\Pi_p(b_K)$$

is the conditional probability of extracting a b_K such that it has a λ degree of irreducibility. Taking the average, we obtain

$$\langle\Omega\rangle = 2^{2^K}\binom{N}{K} - 2^{2^K}\sum_{\lambda=0}^{K}\Pi_p(\Upsilon_K(\lambda))\left[\binom{N-\lambda}{K-\lambda} - 1\right],$$

where

$$\Pi_p(\Upsilon_K(\lambda)) = \sum_{b_K\in\Xi_K}\Pi_p(b_K\,|\,\lambda) \equiv \sum_{b_K\in\Upsilon_K(\lambda)}\Pi_p(b_K)$$

is the probability that $b_K \in \Upsilon_K(\lambda)$. So for $\sharp\Psi(\mathcal{L}_K^N)$, we have:

$$\sharp\Psi(\mathcal{L}_K^N) = \left\{2^{2^K}\binom{N}{K} - 2^{2^K}\sum_{\lambda=0}^{K}\Pi_p(\Upsilon_K(\lambda))\left[\binom{N-\lambda}{K-\lambda} - 1\right]\right\}^N.$$

Note that because of the uncorrelated random extraction of $C_K^{(\alpha_i)}$ and $b_K^{(i)}$, we may do a mean field approximation and take the average before taking the N^{th} power in order to obtain $\sharp\Psi(\mathcal{L}_K^N)$. Then from equation (11), by applying equation (21) we obtain in the asymptotic regime $N \gg 1$, $K \sim O(\ln\ln N)$:

$$\vartheta^{-1}(N, K, p) \approx \{1 - \varphi(K, p)\}^N \approx e^{-\varphi N}, \tag{26a}$$

where

$$\varphi = \varphi(K, p) = \Pi_p(\Upsilon_K(0)).$$

The probabilities $\Pi_p(\Upsilon_K(\lambda))$ were calculated in [23] and we report the result in Appendix B for the interested reader. However, $\Pi_p(\Upsilon_K(0))$ may be easily calculated by noting that $\Upsilon_K(0) = \{\tau, \neg\,\tau\}$. So, from equations (8) and (25)

$$\varphi(K, p) = p^{2^K} + (1-p)^{2^K}. \tag{26b}$$

https://doi.org/10.25088/ComplexSystems.25.4.329

As for the equiprobable case $p = 1/2$, the transition from a many-to-one to an injective function is obtained from equations (26a) and (26b) by setting $e^{-\varphi N} = 1/2$, which now leads to the transcendental equation

$$p^{2^{K_c}} + (1-p)^{2^{K_c}} \approx \frac{\ln 2}{N}. \tag{27}$$

Calculations using equations (26a) and (26b) show that $\Delta K_c \sim O(1/\ln N)$ for the width of the transition. Clearly, equation (27) reduces to equation (24) at $p = 1/2$ and is symmetrical in that point, as it should be. Equation (27) may be solved numerically for K_c versus p, for a fixed N. This is shown in Figure 1 for $N = 10^4$. Below the curve, Ψ is a many-to-one function, while above, it is injective. Note also from the graph that deviations from the equiprobable case $p = 1/2$ may conform to the expected values of the epistatic connections for living organisms on Earth, which go from 6×10^3 in yeast to less than 4×10^4 for *H. sapiens* [12–15]. This makes Kauffman networks a robust model for mathematical biology, even in the case of adding a bias p.

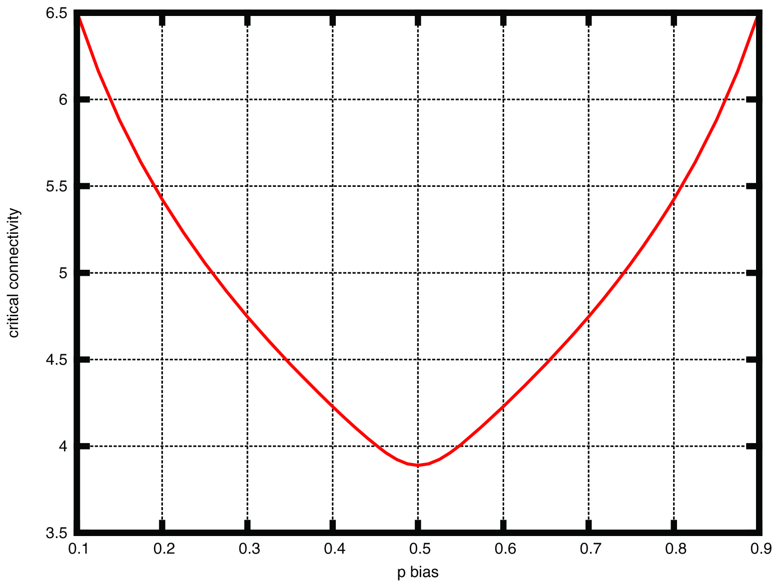

Figure 1. Graph of K_c versus p, for fixed $N = 10^4$.

https://doi.org/10.25088/ComplexSystems.25.4.329

5. Suppression of Tautology and Contradiction

Let us see what happens if the K-Boolean functions tautology τ and contradiction $\neg\tau$ are excluded from the construction of the NK-Kauffman networks. For simplicity, we consider the equiprobable case $p = 1/2$. As we have seen, $\mathcal{T}_K(0) = \{\tau, \neg\tau\}$. Let us denote by $\tilde{\mathcal{L}}_K^N$ the set of NK-Kauffman networks constructed by excluding the functions in $\mathcal{T}_K(0)$. Then, the number of Boolean K-functions in the construction is $2^{2^K} - 2$, and we obtain instead of equations (12b) and (19)

$$\#\tilde{\mathcal{L}}_K^N = \left[\left(2^{2^K} - 2\right)\binom{N}{K}\right]^N$$

and

$$\#\Psi(\tilde{\mathcal{L}}_K^N) = \left\{\left(2^{2^K} - 2\right)\binom{N}{K} - \sum_{\lambda=1}^{K} \beta_K(\lambda)\left[\binom{N-\lambda}{K-\lambda} - 1\right]\right\}^N,$$

respectively. Once again

$$\tilde{\vartheta}\left(N, K, \frac{1}{2}\right) = \frac{\#\tilde{\mathcal{L}}_K^N}{\#\Psi(\tilde{\mathcal{L}}_K^N)}.$$

Now, in the asymptotic regime, the leading term for the repetitions is the one with $\lambda = 1$. From equation (16) and from equations (21) and (22), for $N \gg 1$ and $K \sim O(\ln \ln N)$, it follows that

$$\tilde{\vartheta}\left(N, K, \frac{1}{2}\right) \approx \exp\left(\frac{2K^2}{2^{2^K} - 2}\right)\left[1 - O\left(\frac{1}{N}\right)\right], \tag{28}$$

which goes exponentially to one. So we see that the presence or absence of the tautology and contradiction Boolean functions plays a crucial role in the robustness of Kauffman networks. Now there is no transition from many-to-one to injection and Ψ remaining injective. This is in agreement with Kimura's neutral theory, with τ and $\neg\tau$ playing the role of the random drift [2].

6. Conclusion

We have calculated the asymptotic equations (26a) and (26b), which give the average number $\vartheta(N, K, p)$ of NK-Kauffman networks that the function Ψ maps onto the same functional graph for general values of the bias p. The asymptotic expression of the transition from a many-to-one to the injective function equation (27) was calculated and solved numerically for N fixed. We have seen that while NK-Kauffman networks are simplified models of biological systems, they give good insight into theoretical biology, since they have concordance with biological theories. In particular, we have seen that the suppression of tautology and contradiction Boolean functions from the constructions of Kauffman networks makes them a non-robust model, as long as Ψ remains injective. This is in good agreement with Kimura's neutral theory [2].

It is worth saying that the understanding of Ψ for Kauffman networks is also important from the strictly mathematical point of view, since the dynamical behaviors of the elements in $\mathcal{L}_K^N$ (the Kauffman networks) are related by means of Ψ through functional graphs [3–6].

The relation with percolation and information flow was not addressed in this paper, though it is an open and important field for future research.

Acknowledgments

This work was supported in part by PAPIIT projects IN103815 and IN401513. The author thanks Abigail Meza for fruitful discussions and Pilar López Rico for informatics services. Last but not least, Rubén González Zainez and Patricia Peláez Flores for computer advice.

Appendix

A. The Coefficients $\beta_K(\lambda)$

From equation (17), since the union is disjoint and $\sharp\Xi_K = 2^{2^K}$, it follows that

$$2^{2^K} = \sum_{\lambda=0}^{K} \beta_K(\lambda), \tag{A.1}$$

but $\beta_K(\lambda)$ is equal to the number of ways to form λ irreducible arguments from K arguments; that is,

$$\beta_K(\lambda) = \binom{K}{\lambda} \mathfrak{S}_\lambda, \tag{A.2}$$

where $\mathfrak{S}_\lambda \equiv \beta_\lambda(\lambda)$. Then

$$2^{2^K} = \sum_{\lambda=0}^{K} \binom{K}{\lambda} \mathfrak{S}_\lambda. \tag{A.3}$$

Using the inversion formula of [24], which asserts that for any two sequences of real numbers $\{f_r\}_{r=0}^n$ and $\{g_r\}_{r=0}^n$, such that

$$f_n = \sum_{r=0}^{n} \binom{n}{r} g_r,$$

it follows that

$$g_n = \sum_{r=0}^{n} (-1)^{n-r} \binom{n}{r} f_r,$$

we may invert equation (A.3) to obtain for equation (A.2) the closed expression

$$\beta_K(\lambda) = \binom{K}{\lambda} \sum_{m=0}^{\lambda} (-1)^{\lambda-m} \binom{\lambda}{m} 2^{2^m}. \tag{A.4}$$

B. The Expression for $\Pi_p(\Upsilon_K(\lambda))$

From equation (25), the probability $\Pi_p(\Upsilon_K(\lambda))$ that b_K is in $\Upsilon_K(\lambda)$ is given by

$$\Pi_p(\Upsilon_K(\lambda)) = \sum_{b_K \in \Upsilon_K(\lambda)} \Pi_p(b_K).$$

A closed expression for $\Pi_p(\Upsilon_K(\lambda))$ was calculated in [23], with the result

$$\Pi_p(\Upsilon_K(\lambda)) = \binom{K}{\lambda} \sum_{m=0}^{\lambda} (-1)^{\lambda-m} \binom{\lambda}{m} \left[p^{2^{K-m}} + (1-p)^{2^{K-m}}\right]^{2^m}.$$

Note that for the case $p = 1/2$, by using equation (A.4) we obtain $\Pi_{1/2}(\Upsilon_K(\lambda)) = \beta_K(\lambda) / 2^{2^K}$, as it should be.

References

[1] S. A. Kauffman, "Metabolic Stability and Epigenesis in Randomly Constructed Nets," *Journal of Theoretical Biology*, **22**(3), 1969 pp. 437–467. doi:10.1016/0022-5193(69)90015-0.

[2] S. A. Kauffman, *The Origins of Order: Self-Organization and Selection in Evolution*, New York: Oxford University Press, 1993.

[3] D. Romero and F. Zertuche, "The Asymptotic Number of Attractors in the Random Map Model," *Journal of Physics A: Mathematical and General*, **36**(13), 2003 3691. doi:10.1088/0305-4470/36/13/304.

[4] D. Romero and F. Zertuche, "Grasping the Connectivity of Random Functional Graphs," *Studia Scientiarum Mathematicarum Hungarica*, **42**(1), 2005 pp. 1–19. doi:10.1556/SScMath.42.2005.1.1.

[5] D. Romero and F. Zertuche, "Number of Different Binary Functions Generated by *NK*-Kauffman Networks and the Emergence of Genetic Robustness," *Journal of Mathematical Physics*, **48**(8), 2007 083506. doi:10.1063/1.2768747.

[6] F. Zertuche, "On the Robustness of *NK*-Kauffman Networks against Changes in Their Connections and Boolean Functions," *Journal of Mathematical Physics*, **50**(4), 2009 043513. doi:10.1063/1.3116166.

[7] C. H. Waddington, "The Epigenotype," *International Journal of Epidemiology*, **41**(1), 2012 pp. 10–13. doi:10.1093/ije/dyr184.

[8] P. R. Grant and H. S. Horn, eds., *Molds, Molecules and Metazoa: Growing Points in Evolutionary Biology*, Princeton, NJ: Princeton University Press, 1992.

[9] J. L. Fierst and P. C. Phillips, "Modeling the Evolution of Complex Genetic Systems: The Gene Network Family Tree," *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, **324**(1), 2015 pp. 1–12. doi:10.1002/jez.b.22597.

[10] B. Hallgrímsson and B. K. Hall, eds., *Epigenetics: Linking Genotype and Phenotype in Development and Evolution*, Berkeley: University of California Press, 2011.

[11] E. Jablonka and M. J. Lamb (A. Zeligowski, illus.), *Evolution in Four Dimensions: Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life*, Cambridge, MA: MIT Press, 2005.

[12] J. Krebs, S. Kilpatrick and E. Goldstein (eds.), *Lewin's Genes XI*, 11th ed., Burlington, MA: Jones and Bartlett Learning, 2014.

[13] A. Wagner, "Evolution of Gene Networks by Gene Duplications: A Mathematical Model and Its Implications on Genome Organization," *Proceedings of the National Academy of Sciences*, **91**(10), 1994 pp. 4387–4391. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC43790.

[14] A. Wagner, "Does Evolutionary Plasticity Evolve?," *Evolution*, **50**(3), 1996 pp. 1008–1023. doi:10.2307/2410642.

[15] A. Wagner, *Robustness and Evolvability in Living Systems*, Princeton, NJ: Princeton University Press, 2005.

[16] J. M. Vargas, C. R. Stephens, H. Waelbroeck and F. Zertuche, "Symmetry Breaking and Adaptation: Evidence from a 'Toy Model' of a Virus," *BioSystems*, **51**(1), 1999 pp. 1–14. doi:10.1016/S0303-2647(98)00093-8.

[17] G. Weisbuch, *Complex Systems Dynamics: An Introduction to Automata Networks* (S. Ryckebusch, trans.), Reading, MA: Addison-Wesley, Advanced Book Program, 1991.

[18] S. Wolfram, "Universality and Complexity in Cellular Automata," *Physica D*, **10**(1–2), 1984 pp. 1–35. doi:10.1016/0167-2789(84)90245-8.

[19] F. Zertuche, "Phase Transition in *NK*-Kauffman Networks and Its Correction for Boolean Irreducibility," *Physica D: Nonlinear Phenomena*, **275**, 2014 pp. 35–42. doi:10.1016/j.physd.2014.02.006.

[20] M. Takane and F. Zertuche, "$\mathbb{Z}_2$-Algebras in the Boolean Function Irreducible Decomposition," *Journal of Mathematical Physics*, **53**(2), 2012 023516. doi:10.1063/1.3685444.

[21] G. B. Arfken and H. J. Weber, *Mathematical Methods for Physicists*, 6th ed., Boston: Elsevier, 2005.

[22] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, New York: Dover Publications, 1972.

[23] F. Zertuche, "Canalization and the Stability of *NK*-Kauffman Networks," *Physica D: Nonlinear Phenomena*, **306**, 2015 pp. 1–5. doi:10.1016/j.physd.2015.05.004.

[24] L. Comtet, *Advanced Combinatorics: The Art of Finite and Infinite Expansions*, rev. and enl. ed. (J. W. Nienhuys, trans.), Boston: D. Reidel, 1974 pp. 143–144.