J Biosci (2022)47:17
DOI: 10.1007/s12038-022-00256-9

© Indian Academy of Sciences

A preference for link operator functions can drive Boolean biological networks towards critical dynamics

AJAY SUBBAROYAN[1,2], OLIVIER C MARTIN[3,4]* and AREEJIT SAMAL[1,2]*

[1]The Institute of Mathematical Sciences (IMSc), Chennai 600113, India

[2]Homi Bhabha National Institute (HBNI), Mumbai 400094, India

[3]Université Paris-Saclay, CNRS, INRAE, Univ Evry, Institute of Plant Sciences Paris-Saclay (IPS2), 91405 Orsay, France

[4]Université de Paris, CNRS, INRAE, Institute of Plant Sciences Paris-Saclay (IPS2), 91405 Orsay, France

*Corresponding authors (Emails, olivier.c.martin@inrae.fr; asamal@imsc.res.in)

MS received 18 October 2021; accepted 29 December 2021

Boolean modelling is a powerful framework to understand the operating principles of biological networks. The regulatory logic between biological entities in these networks is expressed as Boolean functions (BFs). There exist various types of BFs (and thus regulatory logic rules) which are meaningful in the biological context. In this contribution, we explore one such type, known as link operator functions (LOFs). We theoretically enumerate these functions and show that, among all BFs and even within the biologically consistent effective and unate functions (EUFs), the LOFs form a tiny subset. We then find that the AND-NOT LOFs are particularly abundant in reconstructed biological Boolean networks. By leveraging these facts, namely, the tiny representation of LOFs in the space of EUFs and their presence in the biological dataset, we show that the space of acceptable models can be shrunk by applying steady-state constraints to BFs, followed by the choice of LOFs which satisfy those constraints. Finally, we demonstrate that among a wide range of BFs, the LOFs drive biological network dynamics towards criticality.

Keywords. Boolean networks; criticality; gene regulatory networks; model selection; regulatory logic rules

1. Introduction

Kauffman (1969a, 1969b, 1993) and René Thomas (1973, 1979) pioneered the use of the discrete-state framework to model gene regulatory networks and demonstrated its potential to reproduce biological outcomes. In the past two decades, the use of the Boolean framework to reconstruct gene networks from experimental biological data has gained momentum. With the advent of sequencing technologies and boost in computational power, it has been possible not only to reconstruct gene networks but also to reproduce gene expression patterns (Mendoza *et al*. 1999; Albert and Othmer 2003; Kauffman *et al*. 2003; Fauré *et al*. 2006; Mendoza and Xenarios 2006).

The notion that complex biological systems are situated in the neighbourhood of a critical dynamical regime has been studied quite extensively both outside (Mora and Bialek 2011) and within the Boolean framework (Shmulevich and Kauffman 2004; Nykter *et al*. 2008; Villani *et al*. 2017, 2018; Daniels *et al*. 2018). The study of damage spreading in Boolean models of gene regulatory networks provides an insight into their dynamical 'regime'. One of the more frequently employed associated characteristics is

This article is part of the Topical Collection: Emergent dynamics of biological networks.

Supplementary Information: The online version contains supplementary material available at https://doi.org/10.1007/s12038-022-00256-9.

http://www.ias.ac.in/jbiosci
Published online: 15 March 2022

obtained by generating a Derrida plot (Derrida and Pomeau 1986). The Derrida plot is partitioned into 3 regions: ordered, critical and chaotic regimes. For models in the ordered regime, perturbations (small random changes in the state of the system) tend to remain small or disappear. In the case of models falling in the chaotic regime, perturbations spread out over many nodes in the network. In the critical regime, the dynamics is neither ordered nor chaotic. More recently, Daniels *et al*. (2018) considered a static measure as a proxy for damage spreading; specifically, the authors used the average sensitivity of a network and showed that most biological models largely fall in the critical regime for which the average sensitivity of the network is equal to 1.

In this work, we focus our attention on a certain type of Boolean function (BF) called link operator functions (LOFs) (Mendoza and Xenarios 2006; Zobolas *et al*. 2022). First, we show the relationship between the different LOFs, and subsequently enumerate the LOFs for different numbers of inputs. Thereafter, we show that LOFs represent an infinitesimal fraction of the space of all BFs, even if considering only the effective and unate functions (EUFs). Next, we show that a sizable proportion of BFs in biological systems are regulated by a specific LOF, namely, the AND-NOT logic. Following this, we present two case studies wherein we impose a given network structure but allow different BFs to examine the consequences of having to satisfy steady-state constraints corresponding to biological phenotypes (Henry *et al*. 2013; Zhou *et al*. 2016). In particular, we show that limiting the choice of BFs to LOFs during such model selection can dramatically shrink the size of the search space. Lastly, by computing the 'static' network sensitivity for a wide range of fixed biological network structures, but imposing different types of functions (effective functions, EUFs and LOFs), we find that the AND-NOT and OR-NOT logic in LOFs are closest to reproducing the average sensitivity distribution of biological regulatory logic.

We now briefly introduce the Boolean framework for modelling gene regulatory networks. A Boolean model is defined by a set of N nodes and L directed edges, where the N nodes correspond to the biological components such as genes or proteins, and the L directed edges correspond to the (oriented) interactions between them. Each node i is associated with a variable $x_i(t)$ which takes only binary values (0 or 1 for 'OFF' or 'ON', respectively), which defines the state of node i at time t. Furthermore, each node i has k_i incoming edges and is thus associated with a BF f_i of k_i variables. The BFs f_i (for all nodes i from 1 to N) along with an updating scheme, synchronous or asynchronous (Kauffman 1969a; Thomas 1991; Garg *et al*. 2008), determine the state of the system at the next time step, $t+1$. The above description is succinctly expressed by the equation

$$x_i(t+1) = f_i\left(x_i^1(t), x_i^2(t), \ldots, x_i^{k_i}(t)\right), \tag{1}$$

where x_i^j is the j^{th} input variable ($j \in [1, k_i]$) of the i^{th} node ($i \in [1, N]$). There are many types of functions f_i that have been defined in the literature that represent underlying molecular logic (see, for instance, Subbaroyan *et al*. 2021). Under the above-mentioned deterministic update rules, the system converges to a steady state (also called fixed point attractor) or into a cyclic attractor.

2. Link operator functions and their properties

2.1 *Motivation and definitions*

Mendoza and Xenarios (2006) defined a type of *veto* regulatory logic in Boolean networks which they used to model the differentiation of T-helper cells. Such a logic has been further studied by others (Ebadi and Klemm 2014). In the above-mentioned works, the *veto* logic operates as follows. If any inhibitor is present (ON), the regulated gene is turned OFF. If all inhibitors are absent and at least one activator is present, then the regulated gene is turned ON; otherwise, the gene is turned OFF.

In a recent contribution, Zobolas *et al.* (2022) used the *structure* of the logical expression of these *veto* BFs to explore a number of other BFs possessing similar logical structure and defined these as 'link operator functions' (LOFs). Their Boolean expression is constructed by *linking* a set of m activators (labelled as x_i) to a set of n inhibitors (labelled as y_j) by a logical operator shown as $\otimes$ in equation (2). We use the symbol k to denote the total number of regulators of the considered node, i.e., $k = m + n$. The general expression for these functions is given by:

$$(x_1, x_2, \ldots, x_m) \otimes (y_1, y_2, \ldots, y_n), \tag{2}$$

where the link operator $\otimes$ can be NOR, NAND, AND-NOT, OR-NOT, NOR-NOT, NAND-NOT, XOR, pairs, XNOR, among others. The activators (or inhibitors) $x_1, x_2, \ldots, x_m$ (or $y_1, y_2, \ldots, y_n$) are typically connected by only AND or OR operators. The LOFs are defined for functions which have at least one activator ($m \geq 1$) and one inhibitor ($n \geq 1$).

Notably, Zobolas *et al.* (2022) showed that only some link operators in equation (2) satisfy biologically relevant 'consistency' properties, namely, monotonicity and essentiality (or effectiveness). First, a BF exhibits unateness or monotonicity (Aracena 2008) if its output is an increasing or decreasing monotonic function of each of the k inputs. More explicitly, if an input is activatory (respectively, inhibitory), then, on increasing the value of that input from 0 to 1 and keeping all other inputs fixed, the output must never decrease (respectively, increase). BFs which are unate (or monotonic) in all inputs are known as unate functions (UFs). Second, a BF exhibits essentiality if each of the inputs of the function is essential. An input i of a BF with k inputs is said to be essential if it is used, i.e., if there exists at least one combination of the other $(k-1)$ inputs for which a change in input i causes a change in the output of the function. BFs which are essential (or effective) in all inputs are known as effective functions (EFs). Biological regulatory logic rules are typically expected to possess both of these 'consistency' properties (Raeymaekers 2002; Aracena 2008; Subbaroyan *et al.* 2021). BFs which possess both of the properties of unateness (or monotonicity) and essentiality (or effectiveness) in all inputs are known as effective and unate functions (EUFs).

In their recent work, Zobolas *et al.* (2022) focussed on 3 types of LOFs, namely, AND-NOT, OR-NOT and their function pairs (which in this article we call AND-pairs) that satisfy the above-mentioned two consistency properties (see table 1 for the exact definition). The AND-NOT, OR-NOT and AND-pairs are given by:

$$f_{\text{AND-NOT}} = (x_1 \vee x_2 \vee \ldots \vee x_m) \wedge \sim (y_1 \vee y_2 \vee \ldots \vee y_n) \tag{3}$$

$$f_{\text{OR-NOT}} = (x_1 \vee x_2 \vee \ldots \vee x_m) \vee \sim (y_1 \vee y_2 \vee \ldots \vee y_n) \tag{4}$$

$$f_{\text{AND-pairs}} = (x_1 \vee x_2 \vee \ldots \vee x_m) \wedge (\sim y_1 \vee \sim y_2 \vee \ldots \vee \sim y_n) \tag{5}$$

where $\vee$ is the OR operator, $\wedge$ is the AND operator and $\sim$ is the NOT operator. For an illustration of the LOFs, see figure 1.

We find that in addition to these three types of LOFs, another type of LOF can be constructed which satisfies the two consistency properties, and is complementary to the AND-pairs in a manner that the OR-NOT is complementary to the AND-NOT. (Note that if one complements an AND-NOT function, one gets an OR-NOT function but with the activators and inhibitors exchanged. Similarly, if one complements the AND-pairs, one gets the OR-pairs but with the activators and inhibitors exchanged.) We call this the OR-pairs and it is given by the expression:

$$f_{\text{OR-pairs}} = (x_1 \wedge x_2 \wedge \ldots \wedge x_m) \vee (\sim y_1 \wedge \sim y_2 \wedge \ldots \wedge \sim y_n), \tag{6}$$

where $\vee$ is the OR operator, $\wedge$ is the AND operator and $\sim$ is the NOT operator.

The biological interpretation for each of the LOFs is as follows:

- AND-NOT: The presence of a single inhibitor represses transcription independent of the presence of multiple activators. Thus, transcription takes place only in the absence of inhibitors and in the presence of at least one activator.
- OR-NOT: The presence of any activator guarantees transcription independent of the presence of inhibitors. In the absence of both inhibitors and activators, gene transcription takes place.
- AND-pairs: The presence of at least one activator *and* the absence of at least one inhibitor is sufficient to ensure transcription.
- OR-pairs: All activators must be present, *or* all inhibitors must be absent in order for transcription to take place.

Table 1 lists the four consistent types of LOFs, their expression and the additional types of BFs to which they belong, and figure 1 depicts the various LOFs. Henceforth, we reserve the word LOF to mean *only* the 4 consistent types, namely, AND-NOT, OR-NOT, AND-pairs and OR-pairs (table 1).

2.2 *Relationship between the different types of LOFs*

We note that there may be overlaps between two different types of LOFs, and between LOFs and other types of biologically meaningful BFs (Subbaroyan *et al.* 2021). Within the space of LOFs we observe that:

(a) AND-NOT and OR-NOT do not overlap.
(b) AND-pairs and OR-pairs do not overlap.

Table 1. The different types of consistent link operator functions (LOFs)

Type of LOF	Boolean expression	Effective	Unate	Canalyzing	Nested canalyzing	Collectively canalyzing
AND-NOT	$(x_1 \vee x_2 \vee \ldots \vee x_m) \wedge \sim (y_1 \vee y_2 \vee \ldots \vee y_n)$	Yes	Yes	Yes	Yes	No
OR-NOT	$(x_1 \vee x_2 \vee \ldots \vee x_m) \vee \sim (y_1 \vee y_2 \vee \ldots \vee y_n)$	Yes	Yes	Yes	Yes	No
AND-pairs ($n > 1$)	$(x_1 \vee x_2 \vee \ldots \vee x_m) \wedge (\sim y_1 \vee \sim y_2 \vee \ldots \vee \sim y_n)$	Yes	Yes	No	No	Yes
OR-pairs ($m > 1$)	$(x_1 \wedge x_2 \wedge \ldots \wedge x_m) \vee (\sim y_1 \wedge \sim y_2 \wedge \ldots \wedge \sim y_n)$	Yes	Yes	No	No	Yes

The 4 different types of LOFs are AND-NOT, OR-NOT, AND-pairs and OR-pairs. From this table, it can be ascertained that these 4 types of LOFs satisfy all the consistency properties considered in Zobolas *et al.* (2022). Note that Zobolas *et al.* (2022) have only considered the first 3 types in their work.

(c) The AND-NOT LOF is equivalent to the AND-pairs LOF if there is only one inhibitory input (n=1), for any value of k.
(d) The OR-NOT LOF is equivalent to the OR-pairs LOF if there is only one activatory input (m=1), for any value of k.

The above observations (c) and (d) serve as a motivation to construct a set of 4 non-overlapping types of LOFs (table 1). We first define two non-overlapping types of LOFs:

(i) AND-pairs (n>1) as the AND-pairs with more than one inhibitory input, and
(ii) OR-pairs (m>1) as the OR-pairs with more than one activatory input.

AND-pairs (n>1) and OR-pairs (m>1) do not overlap with the AND-NOT and OR-NOT LOFs, respectively.

Moreover, we observe that both AND-NOT and OR-NOT LOFs are 'nested canalyzing functions' (NCFs). A k-input BF is said to be 'nested canalyzing' if there exists a permutation of k input variables, such that setting the i^{th} variable to its 'canalyzing' input value fixes the output of the BF, under the condition that the previous ($i-1$) variables are not set to their canalyzing values (Kauffman *et al.* 2003; Szallasi and Liang 1998). The AND-pairs (n>1) and OR-pairs (m>1), on the other hand, are 'collectively canalyzing functions'. A k-input BF is said to be 'collectively canalyzing' if by fixing a certain subset of i inputs (such that $1< i < k$), the output of the function is determined (Reichhardt and Bassler 2007), while it is not when fixing fewer than i inputs.

2.3 *Cardinality of the different types of LOFs*

It is straightforward to count the number of LOFs. Consider the AND-NOT LOFs for instance. For a given number of inputs (k) and for a given number of activators (m) and inhibitors (n), there are C(k,m) (the binomial coefficient) ways to assign m activators and n inhibitors. Since all the activators are connected by an AND or an OR operator (see table 1), the permutations between them do not alter the BF. Hence, there are exactly C(k,m) BFs in the AND-NOT category. A similar argument holds for the number of functions in the OR-NOT category. For the AND-pairs (n>1) and OR-pairs (m>1), the number of functions for m activators and n inhibitors is C(k,m) − C(k,1) and C(k,n) − C(k,1), respectively. To calculate the total number of LOFs of a given type for k inputs, we sum over all the values of m. Hence, for both AND-NOT and OR-NOT, there are a total of 2^k-2 BFs each. We subtract '2' because LOFs do not include the cases where there are no activators or inhibitors, i.e., C(k,m=0) and C(k,n=0) are not counted.

Based on this exact counting of LOFs, it can be easily seen that LOFs form an extremely small subset of the space of all BFs and that their corresponding fraction decreases fast with increasing number of inputs (supplementary table 1). Furthermore, even within the space of EUFs, LOFs form a tiny subset. Figure 2 is a semi-log plot that shows this decrease in the fraction of LOFs with the increase in the number of inputs. Note that even if one pools the 4 classes of LOFs under consideration, the number of functions (for a given number of inputs) increases approximately by a factor

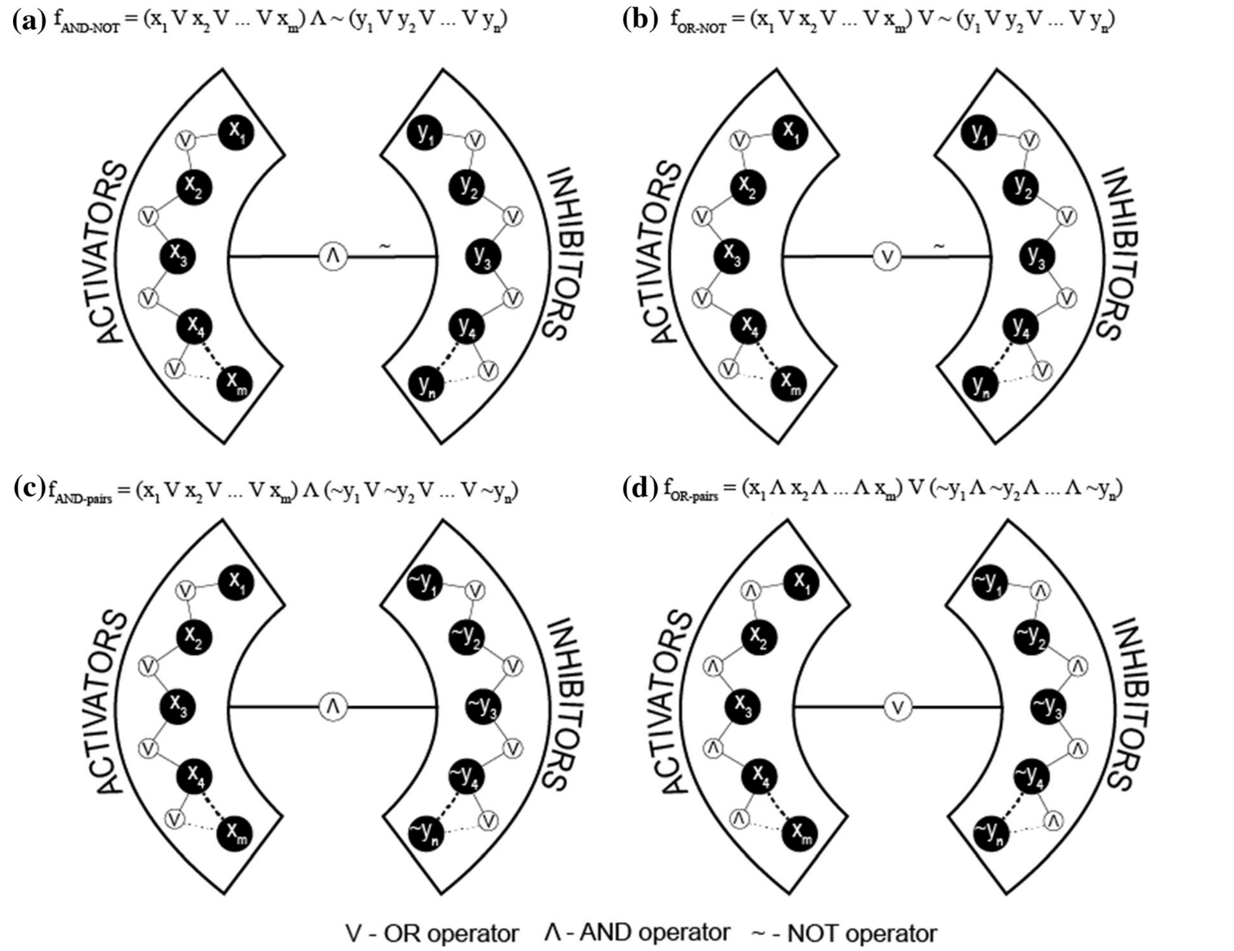

Figure 1. Illustrative figure for the various types of consistent LOFs. The inputs to LOF BFs are divided into two sets, namely, activators and inhibitors, denoted by the variables x_i and y_j, respectively. There are m activators and n inhibitors. The logical operators which connect the variables are the AND ($\wedge$), OR ($\vee$) and NOT ($\sim$) operators. The 4 types of LOFs shown are (**a**) AND-NOT logic, (**b**) OR-NOT logic, (**c**) AND-pairs logic and (**d**) OR-pairs logic.

of 4, which nevertheless does not affect our conclusion. Table 2 and figure 2 illustrate this point.

3. LOFs in biological networks

3.1 *AND-NOT LOFs are particularly abundant in Boolean models of biological systems*

Even though the LOFs are 'consistent' in terms of the effectiveness and monotonicity properties, it remains to be shown how frequently they arise in biological systems. To investigate this, we took as our dataset a collection of 57 Boolean models of biological systems from the Cell Collective database that are a result of the work of many authors, covering a wide variety of biological processes in a number of species spanning the multiple kingdoms of life. Only those models in the Cell Collective database where both the biological network and BFs were curated manually were considered in this study (supplementary table 2). It is clear from table 3 and figure 3 (see also supplementary table 3) that the AND-NOT are particularly abundant in reconstructed Boolean models, whereas the other types of LOFs such as OR-NOT, AND-pairs and OR-pairs are almost absent. Recall that BFs with at least one activator and one inhibitor can be LOFs. Hence, it is meaningful to calculate the fraction of LOFs in the biological dataset among those BFs with at least one activator and one inhibitor (supplementary table 4).

The dominance of AND-NOT LOFs in the dataset implies that regulatory logic is primarily governed by a special type of *veto* mechanism wherein the presence of a single inhibitor determines the output of the gene,

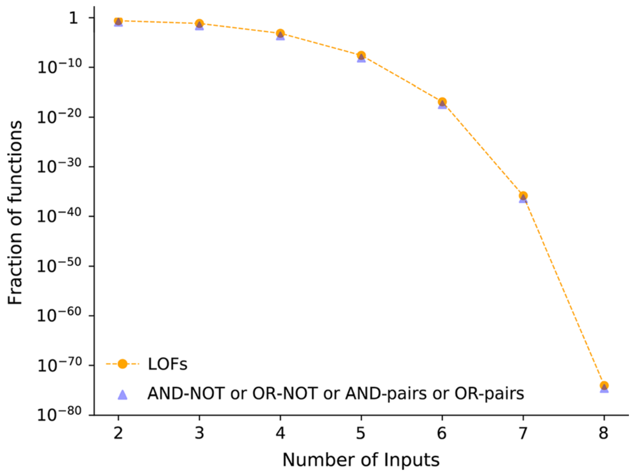

Figure 2. The reduction in the size of the space of consistent LOFs in comparison to the space of all BFs with increasing number of inputs. The decrease of the fraction of consistent LOFs with increasing number of inputs is extremely rapid. Here LOFs (orange circles) refer to the sum of the fractions of all 4 consistent LOFs, namely, the AND-NOT, OR-NOT, AND-pairs and OR-pairs (with any redundancies removed). The blue triangles represent any one of the aforementioned 4 types of LOFs, since each of them have the same number of functions.

independent of the presence of activators. In other words,

(i) the activators can function only in the absence of the inhibitors, and

(ii) the 'vetoing power' of all inhibitors is the same.

Thus, even though the activators are far more numerous than inhibitors in the biological dataset, the inhibitors generally control the logic output. Results inferred from empirical data are typically and rightly subject to scrutiny, in that they could be artefacts of a biased dataset. In the present case we believe that this is highly unlikely given the diversity of biological processes being modelled. Note that 54 out of 57 models in our dataset belong to the Eukaryota domain; the biological literature therein is abundant with cases where the repressor (or inhibitor) is able to suppress transcription even in the presence of many activators (Gaston and Jayaraman 2003).

3.2 *LOFs as facilitators of Boolean model reconstruction and selection: Two case studies*

Model selection is the problem of searching for models which exhibit high fidelity to behaviours identified in the biological data. In general there are many ways to satisfy such constraints (Laubenbacher and Stigler 2004; Cho *et al*. 2007; Dimitrova *et al*. 2011; Zhou *et al*. 2016). In this work, we follow the model selection procedure by Zhou *et al*. (2016): One begins by determining the network structure of the system via experimental data providing information on the regulatory interactions between the biological components.

Table 2. Number of link operator functions (LOFs) as a function of the number of activators (m), the number of inhibitors (n) and the total number of inputs (k)

				LOFs					
k	m	n	EUFs	AND-NOT	OR-NOT	AND-pairs ($n > 1$)	OR-pairs ($m > 1$)	Total	Fraction of EUFs that are LOFs
2	1	1	4	2	2	0	0	4	1
3	1	2	27	3	3	0	3	9	0.333
3	2	1	27	3	3	3	0	9	0.333
4	1	3	456	4	4	0	4	12	0.0263
4	2	2	684	6	6	6	6	24	0.0351
4	3	1	456	4	4	4	0	12	0.0263
5	1	4	34470	5	5	0	5	15	4.35×10^{-4}
5	2	3	68940	10	10	10	10	40	5.80×10^{-4}
5	3	2	68940	10	10	10	10	40	5.80×10^{-4}
5	4	1	34470	5	5	5	0	15	4.35×10^{-4}

Evidently, the total number of inputs (k) is equal to the sum of activators (m) and inhibitors (n), i.e., $k = m + n$. Importantly, a LOF should have at least one activating input ($m \geq 1$) and at least one inhibiting input ($n \geq 1$), and thus, LOFs can exist only for nodes with 2 or more inputs ($k \geq 2$). Here, we give the number of LOFs for different possible combinations of m activators and n inhibitors for a given number of inputs k. Moreover, we report separately the number of functions in the 4 different types of consistent LOFs, namely, AND-NOT, OR-NOT, AND-pairs ($n>1$) and OR-pairs ($m>1$). In addition, the table also gives the number of effective and unate functions (EUFs) for different possible combinations of m and n. As k increases, it can be seen that the LOFs become a tiny fraction of the EUFs.

Table 3. The abundance of link operator functions (LOFs) in the collection of BFs from reconstructed models of biological systems

					LOFs				
k	m	n	BFs in biological dataset	EUFs	AND-NOT	OR-NOT	AND-pairs ($n > 1$)	OR-pairs ($m > 1$)	Total
2	1	1	158	150	147	3	NA	NA	150
3	1	2	35	32	30	1	1	NA	32
	2	1	94	87	47	2	NA	0	49
4	1	3	16	16	13	1	0	NA	14
	2	2	38	35	17	0	0	0	17
	3	1	57	48	18	0	NA	0	18
5	1	4	4	4	1	0	0	NA	1
	2	3	16	15	10	0	0	0	10
	3	2	25	24	8	0	0	0	8
	4	1	20	17	4	0	NA	0	4

The dataset consists of BFs from 57 Boolean models compiled in the Cell Collective database (*https://cellcollective.org/*). Notably, a LOF should have at least one activating input ($m\geq1$) and at least one inhibiting input ($n\geq1$), and thus, LOFs can exist only for nodes with 2 or more inputs ($k\geq2$). Therefore, the dataset consists of the subset of BFs in the 57 reconstructed models that have at least one activating input ($m\geq1$) and at least one inhibiting input ($n\geq1$). The table classifies the BFs in the empirical dataset into effective and unate functions (EUFs) and different types of consistent LOFs. It is evident that EUFs, and moreover, the AND-NOT LOFs within EUFs, are abundant in the dataset regardless of k. In this table, we display the statistics for BFs in the biological dataset up to 5 inputs ($k\leq5$). In supplementary table 3, we display the statistics for all BFs in the biological dataset with $k\leq12$ inputs. 'NA' means 'not applicable', corresponding to values of m and n for which the LOF under consideration does not exist.

Next, dynamical models must reproduce the biological steady states, and in the Boolean framework, this corresponds to imposing constraints on the truth table or function assigned to every node of the network. Finally, among the various types of BFs, *biologically meaningful* functions can be chosen to ensure high biological relevance. Thus, by applying such successive constraints, we can zero in on a much smaller subset of models within the space of all possible models. We illustrate such a model selection procedure on two reconstructed gene regulatory networks (figure 4): a pancreas differentiation model (Zhou *et al*. 2016) and an epithelial–mesenchymal transition (EMT) model (Joo *et al*. 2018).

Table 4 illustrates the reduction in the number of possible models when imposing our successive constraints. Following Zhou *et al.* (2016), the network connectivity is imposed as well as the sign of each interaction when it is known. The problem is then to search the space of BFs at each node. The constraint of reproducing the steady states factorizes, and thus the number of models satisfying the constraints is given by the product of the number of BFs satisfying the constraints on each node. For instance, in the EMT model, there are a total of 268435456 ($=1\times256\times16\times256\times256$) models if one imposes neither steady state constraints nor constraints on the type of BFs, whereas there are 262144 ($=1\times64\times4\times32\times32$) models satisfying the steady-state constraints but ignoring further constraints on the type of BFs. These numbers also reflect the fact that even with a fixed network structure along with steady-state constraints on BFs, the number of models is astronomical.

By taking advantage of the tiny fraction of LOFs in the space of all BFs, we can tremendously shrink the

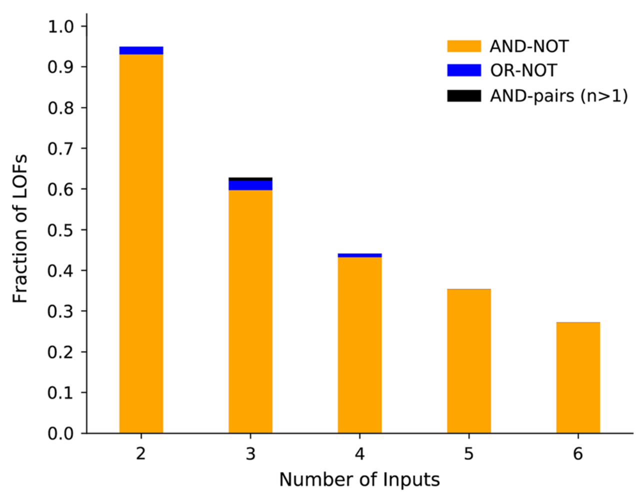

Figure 3. The fractions of the various types of consistent LOFs in the biological dataset. The AND-NOT LOFs are clearly abundant among the biological functions with at least one activator and one inhibitor, whereas the other types, although present, are not as abundant as the AND-NOT functions. Note that the fractions for each of the LOFs are calculated with respect to the number of BFs with at least one activator and one inhibitor as input.

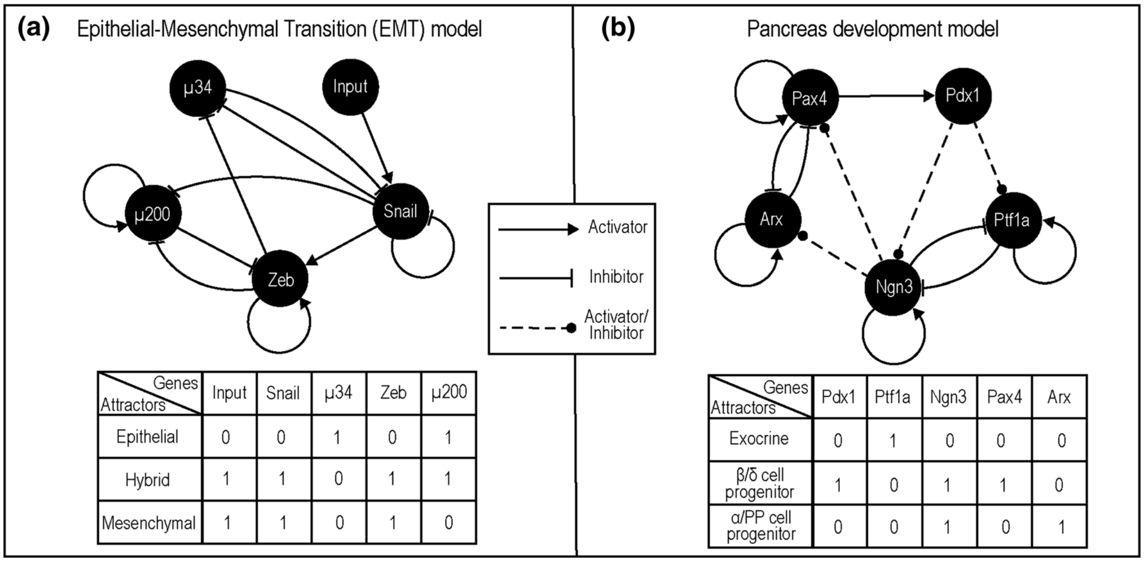

Figure 4. Schematic figure showing the two models of pancreas development and EMT, along with the attractors. Nodes are associated with genes and edges correspond to directed interactions. The biologically relevant attractors in both models are steady states. In the pancreas development network, the edges labelled 'Activator/Inhibitor' correspond to interactions whose signs were denoted as unknown in Zhou *et al.* (2016).

Table 4. Model selection by using different types of BFs with and without the steady-state constraints

	Pancreas development		EMT	
BF constraint	No constraint	Steady-state constraints	No constraint	Steady-state constraints
None	17179869184	1048576	268435456	262144
EUF	104976	7056	1458	140
NCF	65536	3600	1024	96
LOF	1296	100	54	8

The two Boolean models, pancreas development and EMT, with 5 nodes each, are used to illustrate the reduction in allowed models achieved by using various *biologically meaningful* BFs, both with and without the steady-state constraints.

number of models which are biologically relevant. More explicitly, in the case of the EMT model, the 262144 models obtained by applying only steady-state constraints are reduced to just 8 by demanding that the BFs are LOFs, whereas in the pancreatic development model, of 1048576 models which are obtained by imposing steady-state constraints, only 100 models satisfy the conditions of both reproducing the steady states and using LOFs for their regulatory logic. Note that the reduction factor is little less than 10^5 in the EMT model and about 10^4 in the pancreas model. We emphasize that both these models primarily serve as toy models to illustrate the procedure of model selection and consequently the shrinkage of the space of BFs which satisfy the attractor constraints. Although other biological constraints such as the relative stability (Zhou *et al*. 2016) could be used to further zero in on models, applying such constraints is beyond the scope of this study and hence will not be pursued further in this contribution. In essence, using LOFs can tremendously shrink the space of Boolean models to be explored.

3.3 *LOFs drive network dynamics towards 'criticality'*

Damage spreading (Derrida and Pomeau 1986) in discrete dynamical systems measures how two trajectories diverge and thus provides a measure of sensitivity to initial conditions, much like Lyapunov exponents do in continuous systems. Studies in Boolean models of biological gene regulatory networks suggest that these exhibit neither ordered nor chaotic

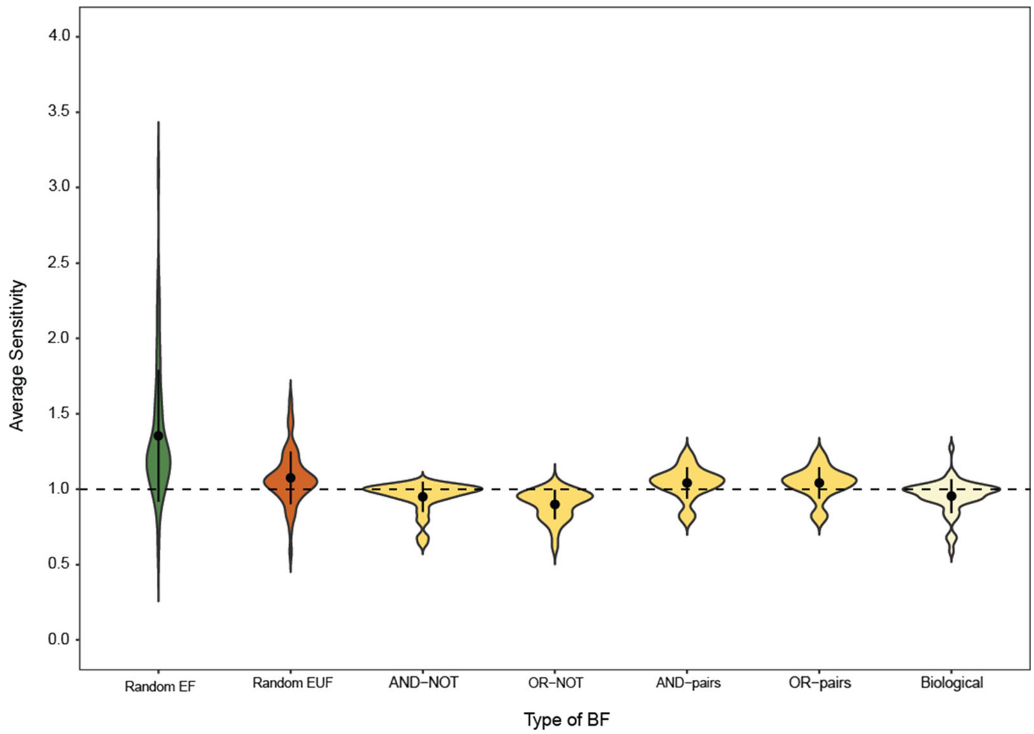

Figure 5. Sensitivity distribution of the various models in the biological dataset using various types of BFs. The sensitivity of models where the structure of the reconstructed biological network is preserved but with the BFs replaced by one of the following types: random EFs, random EUFs, AND-NOT, OR-NOT, AND-pairs, and OR-pairs LOFs. For comparison, we also include the case where the functions are as assigned originally in the reconstructed biological model. Since nodes with only activators or only inhibitors as inputs cannot be assigned LOFs, we assigned the biological functions to them and calculated the average sensitivity of the resulting network. This is done even in the case of EFs and EUFs so as to ensure a fair comparison between the distributions of the average sensitivities of the various BFs being considered.

behaviour, but rather an intermediate kind of behaviour known as 'critical'. Here we employ a static measure of damage spreading, as opposed to the one used to construct Derrida plots, namely, the *average sensitivity* of a Boolean network (Shmulevich and Kauffman 2004). First, the average sensitivity of a BF is given by the proportion of cases where changing one of the inputs at random changes the output value, averaged over all possible input combinations. The average sensitivity of the Boolean network is then the mean of the average sensitivity of all its BFs. Mathematically,

$$S = \frac{1}{N}\sum_{i=1}^{N}\left\langle \sum_{j=1}^{k_i} f(\mathbf{x}\oplus\mathbf{e}_j)\oplus f(\mathbf{x})\right\rangle_{\mathbf{x}}, \tag{7}$$

where N is the total number of nodes, k_i is the in-degree of node i, $\mathbf{e}_j$ is the unit vector corresponding to the j^{th} input, $\mathbf{x}$ labels the possible input k-tuples and $f(\mathbf{x})$ is the output of the BF when $\mathbf{x}$ is the input.

Shmulevich and Kauffman (2004) showed that under the synchronous update scheme, when using randomly drawn representatives of classes of functions, it is possible to infer the damage spreading regime of a Boolean network without resorting to dynamical simulations by simply determining the average sensitivity. Typically, networks with sensitivity $s\sim 1$ indicate that they fall in the critical regime, $s<1$ in the ordered regime and $s>1$ in the chaotic regime. Furthermore, by computing the sensitivity s of a wide range of biological Boolean models, Daniels *et al.* (2018) showed that most biological models fall in the 'critical' regime ($s\sim 1$).

In this work, we compared sensitivities of biological networks with fixed connectivity structure but varying functions, namely, EF, EUF, AND-NOT, OR-NOT, AND-pairs, OR-pairs LOFs and 'biological functions' (i.e., the functions as assigned by model builders). We performed this analysis on 57 models collected from the Cell Collective database (Helikar *et al.* 2012; *https://cellcollective.org*). In the case of EFs and EUFs, each node can be assigned BFs ranging over numerous values of average sensitivities, whereas for the LOFs of a given kind, there exists multiple functions, but all with the same value of average sensitivity.

In figure 5, we see that networks driven by LOF regulatory logic push the biological network dynamics towards criticality (s=1) (supplementary table 5). Based on the fraction of networks lying in the outliers of the biological distribution (supplementary table 6), AND-NOT and OR-NOT LOFs lead to more realistic behavior than other types of logic functions. Details of the procedure to generate EFs and EUFs and other assumptions used in these computations can be found in the supplementary material.

4. Discussion

A large-scale analysis to assess the abundance of LOFs in Boolean models of biological networks has not been carried out so far. The present analysis reveals the high preference for AND-NOT logic in the regulatory rules of genetic networks. This preference coupled with the fact that LOFs occupy a minute region in not only the space of all BFs but also within EUFs raises the question: why are LOFs, specifically the AND-NOT logic, preferred over other choices of BFs? We tackle this question by determining how the imposition of various types of regulatory rules affects damage spreading in such networks. Daniels *et al.* (2018), by using average sensitivity that is a static measure of damage spreading, showed that having canalyzing rules pushes Boolean models towards criticality. We go one step further to show that within both canalyzing functions and 'consistent' logic functions (i.e., EUFs), although LOF logic drives network dynamics towards criticality, eukaryotic mechanisms are predominantly driven by AND-NOT logic. Biological networks governed by OR-NOTs fall slightly in the ordered regime, in comparison with other types of BFs.

Given that there are multiple advantages to choosing LOFs as regulatory logic, it is worth noting that LOFs are also limited in their scope as they require at least one activator and one inhibitor. We observe that such nodes, all of whose inputs are either only activators or only inhibitors, are abundant in biological networks (supplementary table 4). Thus, it is the combined effect of those logic functions and the AND-NOT logic that shape the models we have studied here.

This work raises multiple questions for further investigations. First, is there a network structure–function relationship (Henry *et al.* 2013) which could give us a deeper insight into why certain logics are more preferable than others? Second, in tackling the problem of model selection, if we apply additional biological constraints such as the relative stability (Zhou *et al.* 2016) (in cellular differentiation processes) of attractors, how faithful will constructed models, whose functions are LOFs, be to the biology?

Acknowledgements

AS would like to thank Somdatta Sinha for encouraging him over the years to further scrutinize biological networks. AS would also like to thank the Max Planck Society, Germany, for the award of a Max Planck Partner Group in Mathematical Biology. IPS2 benefits from the support of Saclay Plant Sciences-SPS (ANR-17-EUR-0007).

Data Availability

All datasets including the BFs compiled from the 57 Boolean models of biological systems obtained from the Cell Collective database (*https://cellcollective.org/*) and the programs needed to reproduce the results of this study are available from the associated GitHub repository: *https://github.com/asamallab/LOF*.

References

Albert R and Othmer HG 2003 The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in *Drosophila melanogaster*. *J. Theor. Biol.* **223** 1–18

Aracena J 2008 Maximum number of fixed points in regulatory Boolean networks. *Bull. Math. Biol.* **70** 1398

Cho K-H, Choo S-M, Jung S, *et al.* 2007 Reverse engineering of gene regulatory networks. *IET Syst. Biol.* **1** 149–163

Daniels BC, Kim H, Moore D, *et al.* 2018 Criticality distinguishes the ensemble of biological regulatory networks. *Phys. Rev. Lett.* **121** 138102

Derrida B and Pomeau Y 1986 Random networks of automata: a simple annealed approximation. *Europhys. Lett.* **1** 45

Dimitrova E, García-Puente LD, Hinkelmann F, *et al.* 2011 Parameter estimation for Boolean models of biological networks. *Theor. Comput. Sci.* **412** 2816–2826

Ebadi H and Klemm K 2014 Boolean networks with veto functions. *Phys. Rev. E* **90** 022815

Fauré A, Naldi A, Chaouiya C and Thieffry D 2006 Dynamical analysis of a generic Boolean model for the control of the mammalian cell cycle. *Bioinformatics* **22** e124–e131

Garg A, Di Cara A, Xenarios I, Mendoza L and De Micheli G 2008 Synchronous versus asynchronous modeling of gene regulatory networks. *Bioinformatics* **24** 1917–1925

Gaston K and Jayaraman PS 2003 Transcriptional repression in eukaryotes: repressors and repression mechanisms. *Cell. Mol. Life Sci.* **60** 721–741

Helikar T, Kowal B, McClenathan S, *et al*. 2012 The cell collective: toward an open and collaborative approach to systems biology. *BMC Syst. Biol.* **6** 1–14

Henry A, Monéger F, Samal A and Martin OC 2013 Network function shapes network structure: the case of the *Arabidopsis* flower organ specification genetic network. *Mol. Biosyst.* **9** 1726–1735

Joo JI, Zhou JX, Huang S and Cho K-H 2018 Determining relative dynamic stability of cell states using Boolean network model. *Sci. Rep.* **8** 12077

Kauffman S, Peterson C, Samuelsson B and Troein C 2003 Random Boolean network models and the yeast transcriptional network. *Proc. Natl. Acad. Sci. USA* **100** 14796–14799

Kauffman SA 1993 *The Origins of Order: Self-organization and Selection in Evolution* (Oxford University Press, New York)

Kauffman SA 1969a Homeostasis and differentiation in random genetic control networks. *Nature* **224** 177–178

Kauffman SA 1969b Metabolic stability and epigenesis in randomly constructed genetic nets. *J. Theor. Biol.* **22** 437–467

Laubenbacher R and Stigler B 2004 A computational algebra approach to the reverse engineering of gene regulatory networks. *J. Theor. Biol.* **229** 523–537

Mendoza L, Thieffry D and Alvarez-Buylla ER 1999 Genetic control of flower morphogenesis in *Arabidopsis thaliana*: a logical analysis. *Bioinformatics* **15** 593–606

Mendoza L and Xenarios I 2006 A method for the generation of standardized qualitative dynamical systems of regulatory networks. *Theor. Biol. Med. Model.* **3** 13

Mora T and Bialek W 2011 Are biological systems poised at criticality? *J. Stat. Phys.* **144** 268–302

Nykter M, Price ND, Aldana M, *et al*. 2008 Gene expression dynamics in the macrophage exhibit criticality. *Proc. Natl. Acad. Sci. USA* **105** 1897

Raeymaekers L 2002 Dynamics of Boolean networks controlled by biologically meaningful functions. *J. Theor. Biol.* **218** 331–341

Reichhardt CJO and Bassler KE 2007 Canalization and symmetry in Boolean models for genetic regulatory networks. *J. Phys. Math. Theor.* **40** 4339–4350

Shmulevich I and Kauffman SA 2004 Activities and sensitivities in Boolean network models. *Phys. Rev. Lett.* **93** 48701

Subbaroyan A, Martin OC and Samal A 2021 Minimum complexity drives regulatory logic in Boolean models of living systems. *bioRxiv* 202109.20.461164

Szallasi Z and Liang S 1998 Modeling the normal and neoplastic cell cycle with "realistic Boolean genetic networks": their application for understanding carcinogenesis and assessing therapeutic strategies. *Pac. Symp. Biocomput.* **3** 66–67

Thomas R 1991 Regulatory networks seen as asynchronous automata: A logical description. *J. Theor. Biol.* **153** 1–23

Thomas R 1979 Kinetic logic: a Boolean approach to the analysis of complex regulatory systems, Proceedings of the EMBO course 'Formal analysis of genetic regulation', held in Brussels, September 6–16 1977 *Lecture Notes in Biomathematics* (Springer)

Thomas R 1973 Boolean formalization of genetic control circuits. *J. Theor. Biol.* **42** 563–585

Villani M, Campioli D, Damiani C, *et al.* 2017 Dynamical regimes in non-ergodic random Boolean networks. *Nat. Comput.* **16** 353–363

Villani M, La Rocca L, Kauffman SA and Serra R 2018 Dynamical criticality in gene regulatory networks. *Complexity* **2018** 5980636

Zhou JX, Samal A, d'Hérouël AF, Price ND and Huang S 2016 Relative stability of network states in Boolean network models of gene regulation in development. *Biosystems* **142** 15–24

Zobolas J, Monteiro PT, Kuiper M and Flobak Å 2022 Boolean function metrics can assist modelers to check and choose logical rules. *J. Theor. Biol.* **538** 111025

Corresponding editor: SUSMITA ROY