

4-2012

## Mean-field Boolean network model of a signal transduction network

Naomi Kochi

*University of Nebraska Medical Center*

Mihaela Teodora Matache

*University of Nebraska at Omaha, dvelcsov@unomaha.edu*

Follow this and additional works at: <https://digitalcommons.unomaha.edu/mathfacpub>

 Part of the [Mathematics Commons](#)

Please take our feedback survey at: [https://unomaha.az1.qualtrics.com/jfe/form/SV\\_8cchtFmpDyGfBLE](https://unomaha.az1.qualtrics.com/jfe/form/SV_8cchtFmpDyGfBLE)

---

### Recommended Citation

Kochi, Naomi and Matache, Mihaela Teodora, "Mean-field Boolean network model of a signal transduction network" (2012). *Mathematics Faculty Publications*. 25.  
<https://digitalcommons.unomaha.edu/mathfacpub/25>

This Article is brought to you for free and open access by the Department of Mathematics at DigitalCommons@UNO. It has been accepted for inclusion in Mathematics Faculty Publications by an authorized administrator of DigitalCommons@UNO. For more information, please contact [unodigitalcommons@unomaha.edu](mailto:unodigitalcommons@unomaha.edu).

# MEAN-FIELD BOOLEAN NETWORK MODEL OF A SIGNAL TRANSDUCTION NETWORK

NAOMI KOCHI<sup>+</sup>, MIHAELA TEODORA MATACHE<sup>†\*</sup>

<sup>+</sup>Department of Genetics, Cell Biology, and Anatomy  
University of Nebraska Medical Center, Omaha NE 68198, USA

<sup>†</sup>Department of Mathematics  
University of Nebraska at Omaha, Omaha, NE 68182, USA  
dmatache@mail.unomaha.edu

\*corresponding author

**ABSTRACT.** In this paper we provide a mean-field Boolean network model for a signal transduction network of a generic fibroblast cell. The network consists of several main signaling pathways, including the receptor tyrosine kinase, the G-protein coupled receptor, and the Integrin signaling pathway. The network consists of 130 nodes, each representing a signaling molecule (mainly proteins). Nodes are governed by Boolean dynamics including canalizing functions as well as totalistic Boolean functions that depend only on the overall fraction of active nodes. We categorize the Boolean functions into several different classes. Using a mean-field approach we generate a mathematical formula for the probability of a node becoming active at any time step. The model is shown to be a good match for the actual network. This is done by iterating both the actual network and the model and comparing the results numerically. Using the Boolean model it is shown that the system is stable under a variety of parameter combinations. It is also shown that this model is suitable for assessing the dynamics of the network under protein mutations. Analytical results support the numerical observations that in the long-run at most half of the nodes of the network are active.

**Keywords:** Signal transduction network, mean-field approximation, canalizing functions, stability, bifurcations, noise, mutations, simulations.

## 1. INTRODUCTION

In this paper, we consider the fibroblast signal transduction network developed by Helikar et. al. [1] as a Boolean model of signal transduction in a generic fibroblast cell.

We study the dynamics of that protein network using a mean-field approximation formula for the density of ones at a given time step, for a combination of rules that come from [1].

It is known that Boolean network models have been used for modeling networks in which the node or cell activity can be described by two states, 1 and 0, “active and nonactive”, “up-regulated and down-regulated”, and in which each node is updated based on logical relationships with other nodes. Boolean networks can model a variety of real or artificial networks including among others: signal transduction networks (e.g. Helikar et. al. [1]), genetic regulatory networks or other biological networks (e.g. Kauffman [2], Shmulevich et. al. [3], [4], Klemm and Bornholdt [5], or Raeymaekers [6]), or neural networks (e.g. Huepe and Aldana [7]). Our goal is to provide a mathematical Boolean model (MF-model) that can be used to explore certain dynamical aspects of the Boolean representation of the fibroblast network in [1], which may be hard to explore by using that network or in laboratory experiments. In [1] the authors indicate that due to the highly interconnected nature of cytoplasmic protein networks their Boolean representation ignored a significant number of interactions with proteins outside their model. Signal transduction pathways are still being elucidated, and all of the nodes and connections between nodes in these pathways have not yet been determined. Therefore the network [1] is representative for only a fraction of the actual biological network whose size and connectivity are much larger. Moreover, nodes were not included in [1] unless they were generally expressed in a wide range of cell types, not only in a fibroblast cell. Thus, a mathematical Boolean model such as our MF-model of this type of network can be used as a starting point for exploring various types of cells, transcending the restrictions on both the topology and the dynamical rules of the fibroblast network.

In Boolean network modeling we are interested to understand what are the long-term dynamics of the network. A natural approach to answering these questions is to provide a mathematical model for certain numerical characteristics associated with the states of the network. For example, the density of ones, or equivalently, the fraction of active nodes at a given time step is a common measure that can capture certain aspects of the dynamics of the network. This numerical characteristic is explored under a mean-field approach by Huepe and Aldana [7], or Beck and Matache [8], who show that when the Boolean rules are threshold functions typical to neural networks and biological networks, an increase of the noise parameters may induce a phase transition. On the other hand Kauffman et. al. have shown in [9] that genetic networks with canalizing Boolean rules are stable for various connectivity distributions. At the same time Rämö et. al. [10]

show that Boolean networks constructed using one type of stabilizing functions (including canalizing) are always stable regardless of the average in-degree of network functions. Moreover, Nikolajewa et. al. have shown in [11] that many biologically relevant functions belong to the simplest hierarchically canalizing functions and show ordered behavior. All these papers have the property that they focus on networks with varying topologies but governed by rules from a single class of Boolean functions (threshold or other totalistic rules, or canalizing with one canalizing input, etc.). However, a real network such as the fibroblast signal transduction network aggregates various types of Boolean functions with various weights on the overall dynamics. Our goal is to provide a general mathematical mean-field aggregation procedure for the density of ones with several classes of Boolean functions that can be used to explore the importance of each type of Boolean function on the long-run behavior of the network. We will show that in the context of aggregation of several types of rules identified in the fibroblast network stability may prevail under various parameter modifications.

One rationale for constructing a MF-model based on the network model [1] is based on the fact that the network model has some structural restrictions while the MF-model does not. For example, consider a case where we want to change the connectivity for a certain node and investigate how such a change will affect the dynamics of the network. This task can be very complicated for the network model [1], and there are cases where it is impossible to carry it out. To add another connection to the node under consideration, we need to specify a new input node, its upstream nodes and regulation rules by its upstream nodes, and this process must be repeated for those upstream nodes and so forth. However, due to its complexity, this process can become intractable and too hard to continue. On the other hand, for the MF-model, as long as the parameters are clearly defined, all we need to do is to change some parameter values. Similarly, one might be interested in diminishing or increasing the presence of a protein of interest in the overall scheme of the cell. Again, for the network model [1] we would need to add/remove nodes and links and readjust the upstream and downstream regulatory mechanisms, which can be prohibitive. The parameters of the MF-model provide us with the possibility of altering the weight of each type of Boolean rule (protein), from an overwhelming presence of that type of rule in the network to the total absence of it. Hence, the benefit of having a mathematical model such as the MF-model for the network model is that once the MF-model is identified as a good fit for the network model, we can depart from the network model and explore its dynamics by only changing some parameter values in the MF-model. This way the

MF-model has the potential for providing new biological insights since through different combinations of parameters we can make assumptions which are connected to biological questions that may not be feasible to carry out in the laboratory or by using the network model [1] alone. For example, some types of nodes are known to be essential for the functionality of a biological network, while the role of other nodes is still unclear. As we will see, in the fibroblast network about 14% of the nodes obey the logical COPY function with only one input. A node with this function simply passes the state of its upstream node to its downstream node. This seems to be a redundancy, and raises the question if the presence or absence of these nodes will affect the network (in particular, the activity level of the network in this paper). This question cannot be easily answered by using the direct Boolean representation of the network, or in laboratory.

As mentioned earlier, another advantage of the MF-model proposed in this paper is that it allows extensions to networks that have some similarities with the fibroblast cell, but whose overall structures may be still be at least partially unknown. From a mathematical standpoint, one could increase the network size and the connectivity to any level to account for missing proteins and links in the fibroblast network. On the other hand, one could increase/decrease the presence of nodes obeying dynamical rules that stem from the fibroblast network but that may be under/over represented in the known part of the network. In this paper we are interested in creating a baseline for further study of individual types of nodes/dynamical rules that exhibit sensitivity to perturbations, thus having the potential to modify the dynamics of the whole network or other types of nodes. Such nodes could become potential targets in drug therapies.

This paper is structured as follows. The detailed fibroblast signal transduction network description is included in Section 2. Based on statistics obtained from the fibroblast network we classify the types of Boolean functions into several categories: (the logical) COPY, AND, and OR functions,  $\mathcal{C}$  representing canalizing, and OTHER. This classification corresponds to the actual rules observed in the fibroblast signal transduction network. The OTHER class consists of all Boolean functions occurring in the fibroblast network that are not included in the remaining categories. Using a mean-field approach and the Boolean function categories described above, we determine a formula for the probability of a protein being ON at a given time point assuming that the proteins obey the types of rules in our classification in various proportions. This is done in Section 3. In Section 4 we provide the aggregated MF-model and we compare the results obtained using the MF-model with the results obtained from actually evolving the fibroblast network and

collecting the fraction of active nodes at each iteration. We plot the results on the same graph for an easy comparison. The MF-model is identified as a good fit for the actual network. In Section 5 we use the MF-model to predict the network behavior under various scenarios, by varying the MF-model parameters. The MF-model is used to establish a baseline for comparison of the behavior of the network under various assumptions to the actual fibroblast network model [1]. We assess the impact of parameter variation on the overall dynamics. On the other hand, we consider protein mutations and compare the results obtained with our MF-model to those obtained by using the model in [1]. Thus one can assess the robustness or sensitivity of the network to this type of noise. Some analytical results regarding the fixed points of the aggregated MF-model are presented in Section 6, while Section 7 is dedicated to conclusions and further directions of research.

## 2. SIGNAL TRANSDUCTION NETWORK DESCRIPTION

The fibroblast signal transduction network consists of several main signaling pathways, including the receptor tyrosine kinase, the G-protein coupled receptor, and the Integrin signaling pathway. The network contains 130 nodes. Each node represents a signaling molecule (mainly protein). Nodes are governed by Boolean dynamics. In the original network of [1] there are also nine external input nodes which represent extracellular stimuli, adding a stochastic component to the network, or a "background noise". These nine nodes are external to the network and do not have any input nodes of their own, so they are not governed by some specific Boolean rules. Their states could be chosen in a deterministic or random fashion, depending on the type of "background noise" one wishes to consider. In this paper the "background noise" is chosen according to the procedure developed in [1]. However, the impact of the external inputs is not assessed in the MF-model of this paper.

We classify the types of Boolean rules of the fibroblast network according to the categories shown in Table 1 which include COPY, AND, and OR functions, canalizing functions  $\mathcal{C}$ , and OTHER. The first three functions can be expressed as  $\text{COPY}(x) = x$ ,  $\text{AND}(x_1, x_2, \dots, x_k) = x_1 \wedge x_2 \wedge \dots \wedge x_k$ ,  $\text{OR}(x_1, x_2, \dots, x_k) = x_1 \vee x_2 \vee \dots \vee x_k$ , where  $x, x_1, x_2, \dots, x_k \in \{0, 1\}$  are the inputs, and  $\wedge, \vee$  are the logical "and" and "or" symbols. Class  $\mathcal{C}$  incorporates canalizing rules with one or more canalizing inputs. A Boolean function  $f : \{0, 1\}^k \rightarrow \{0, 1\}$  is canalizing if at least for one value of one of the inputs the output is fixed, irrespective of the values of the other inputs. More precisely, there exists  $i \in \{1, 2, \dots, k\}$  and  $u, v \in \{0, 1\}$  such that for all  $x_1, x_2, \dots, x_k \in \{0, 1\}$ , if  $x_i = u$  then

k (connectivity)	Frequency	COPY	OR	AND	$\mathcal{C}$	OTHER
1	18 (13.8%)	18	0	0	0	0
2	19 (14.6%)	0	6	8	5	0
3	20 (15.4%)	0	2	1	15	2
4	22 (16.9%)	0	1	0	13	8
5	12 (9.2%)	0	1	0	4	7
6	10 (7.7%)	0	0	0	5	5
7	7 (5.4%)	0	0	0	3	4
8	10 (7.7%)	0	0	0	5	5
9	5 (3.8%)	0	0	0	3	2
10	4 (3.1%)	0	0	0	4	0
11	1 (0.8%)	0	0	0	1	0
12	1 (0.8%)	0	0	0	0	1
14	1 (0.8%)	0	0	0	1	0
Total	130 (100%)	18 (14%)	10 (8%)	9 (7%)	59 (45%)	34 (26%)

TABLE 1. Classification of the Boolean rules in the fibroblast network and corresponding frequencies.

$f(x_1, x_2, \dots, x_k) = v$ . In this case  $u$  is the canalizing value (of the canalizing input) and  $v$  is the canalized value (of the output). Observe that COPY, AND, OR are canalizing functions.

We collect statistics regarding the frequency of each type of rule by connectivity level  $k$ . We observe that the majority of the rules are canalizing (including COPY). To keep this approach to modeling the fibroblast network of [1] at a reasonable level of complexity, any other rules aside from the ones mentioned above are included in the last category, OTHER. Regarding the connectivity values, observe that more than 60% of the nodes have connectivity at most 6, and more than 90% have connectivity at most 8.

For each canalizing function in class  $\mathcal{C}$  we identify the number of canalizing variables as shown in Table 2. Note that 49 out of 59 canalizing functions have one or two canalizing inputs. For simplicity we will consider only these 49 canalizing functions in the MF-model. From preliminary simulations, we do not see a significant impact of this simplification; however, future work will allow for further refinements of the types of functions and connectivity values considered in modeling. We classify the functions in  $\mathcal{C}$  based on the canalizing  $\rightarrow$  canalized value relationship as shown in Table 3. This classification is obtained directly from the fibroblast network.

The classification of the types of Boolean rules of the fibroblast network according to the categories and statistics shown above is summarized in the pie charts of Figure 1.

Number of canalizing variables	1	2	3	4	5
Frequency	31 (52.5%)	18 (30.5%)	5 (8.5%)	3 (5.1%)	2 (3.4%)

TABLE 2. Frequency of number of canalizing variables for rules in class  $\mathcal{C}$ .

Canalizing $\rightarrow$ canalized values	Frequency
COPY	2
REVERSE	1
$0 \rightarrow 0$	16
$1 \rightarrow 0$	4
$1 \rightarrow 1$	8
$(0or0 \rightarrow 0)$	4
$(0or1 \rightarrow 0)$	4
$(0and1 \rightarrow 0)$	8
$(0or1 \rightarrow 1)$	2

TABLE 3. Classification of canalizing functions with one or two inputs. The first five functions have one canalizing input, while the last four have two canalizing inputs. The rules with one canalizing input are: COPY, meaning that the output state is the same as the input state; REVERSE, meaning that the output state is the reverse (opposite) of the input state;  $0 \rightarrow 0$ , that is if the canalizing input is on the canalizing value 0 then the output is 0;  $1 \rightarrow 0$  meaning that if the canalizing input has canalizing value 1 then the output is 0;  $1 \rightarrow 1$  which means that if the canalizing input is on its canalizing value 1 then the output is 1. The rules with two canalizing inputs are:  $(0or0 \rightarrow 0)$ , that is if at least one input is 0 then the output is 0;  $(0or1 \rightarrow 0)$  meaning that if at least one input is 1 then the output is 0;  $(0and1 \rightarrow 0)$  meaning that if one input is 0 and the other is 1 then the canalized value is 0; and  $(0or1 \rightarrow 1)$  meaning that if at least one input is 1 then the output is 1.

Given this particular classification of the fibroblast network rules, we are interested in exploring the impact of each type of rule on the dynamics of the network. More precisely, using a mean field approach we generate formulas for the quantity  $p(t)$ , the probability of any protein being ON at a given time step, using combinations of rules from the specified categories. Then we explore the impact of each type of rule on the overall dynamics. The quantity  $p(t)$  can be estimated by the fraction of active (ON) nodes at time  $t$ , whose magnitude can indicate the departure from the wild type (non-mutated) functionality of the network dynamics. We have explored the differences in overall network activity for the original fibroblast network [1] and mutated versions of the network, in which one or two nodes are frozen to either the ON or OFF state. We have noted that in some situations when two nodes are mutated ON, the overall network activity can increase by 20-25%, while the increase in the activity of individual types of nodes corresponding to Table 3 can be much higher. For example, if two nodes obeying the COPY rule (such

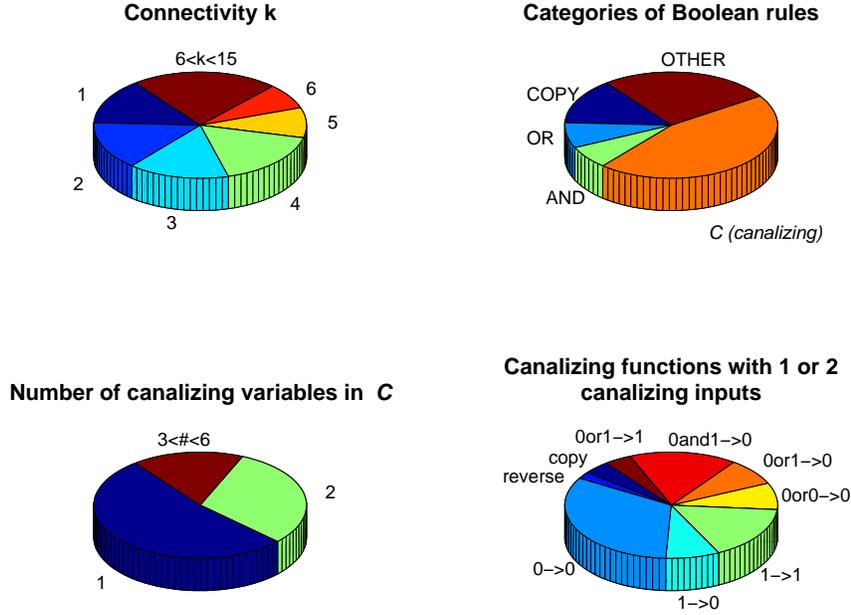


FIGURE 1. Summarizing pie charts for the various statistics of the fibroblast signal transduction network shown in Tables 1, 2, 3.

as proteins IL1\_TNFR, Trafs, SHP2 etc.) are frozen in the ON state, then the overall number of active nodes is increased by about 20%, while the corresponding increase in activity of the nodes obeying the  $1 \rightarrow 0$  rule is larger. This is to be expected given that in the fibroblast network, some of the nodes obeying the  $1 \rightarrow 0$  rule have canalizing inputs from among the nodes obeying the COPY rule. Thus, modifications in  $p(t)$  in comparison to the wild type behavior, may be an indication that mutations have occurred. In this paper we assess the impact of the individual rule categories specified in this section on the dynamics of the network, and establish a baseline for future comparison with the wild type dynamics.

### 3. INDIVIDUAL MEAN FIELD MODELS

Consider a Boolean network with  $N$  nodes  $V = \{x_1, x_2, \dots, x_N\}$  and corresponding connectivities  $\{k_1, k_2, \dots, k_N\}$ . Assume that the inputs are chosen randomly from the  $N$  nodes, thus generating a directed network with possible self-regulation of the nodes. In some cases, such as COPY, we will assume non self-regulation by default. The nodes are updated synchronously and we denote by  $p(t)$  the density of ones at time  $t$ , which is an estimate for the probability of finding a generic node ON at time  $t$ . We will determine  $p(t + 1)$  in terms of  $p(t)$ .

Now, if  $x$  is a generic node of the network, then  $p(t+1) = \mathcal{P}(x(t+1) = 1)$ , where the notation  $\mathcal{P}$  stands for the probability of an event. By the law of total probability this is given by  $\mathcal{P}(x(t+1) = 1|x(t) = 0) \cdot \mathcal{P}(x(t) = 0) + \mathcal{P}(x(t+1) = 1|x(t) = 1) \cdot \mathcal{P}(x(t) = 1)$ . But  $\mathcal{P}(x(t) = 1) = p(t)$  and thus  $\mathcal{P}(x(t) = 0) = 1 - p(t)$ , which implies

$$p(t+1) = (1 - p(t)) \cdot \mathcal{P}(x(t+1) = 1|x(t) = 0) + p(t) \cdot \mathcal{P}(x(t+1) = 1|x(t) = 1).$$

Thus to obtain  $p(t+1)$  we need to find out what are the expressions of  $\mathcal{P}(x(t+1) = 1|x(t) = 0)$  and  $\mathcal{P}(x(t+1) = 1|x(t) = 1)$  under the various types of Boolean rules in our classification.

We will consider the rules COPY, OR, AND,  $\mathcal{C}$ , and OTHER step by step.

**Case 1: COPY.** Node  $x$  has one input and the Boolean rule is COPY. In this case self-regulation is not allowed.

Then clearly

$$p(t+1)_{COPY} = p(t).$$

**Case 2: OR.** Node  $x$  has  $k$  inputs, and the Boolean rule is OR.

So if at least one input is ON then the output is ON. Then, if  $x$  is self-regulated,

$$\begin{aligned} \mathcal{P}(x(t+1) = 1|x(t) = 0) &= \mathcal{P}(\text{at least one of the remaining } k-1 \text{ inputs is } 1) = \\ &= 1 - \mathcal{P}(\text{all the remaining } k-1 \text{ inputs are } 0) = 1 - (1 - p(t))^{k-1}. \end{aligned}$$

On the other hand, if  $x$  is self-regulated then  $\mathcal{P}(x(t+1) = 1|x(t) = 1) = 1$ . Then

$$p(t+1)_{OR} = 1 - (1 - p(t))^k.$$

If the node  $x$  is not self-regulated, then its own state has no impact on its future state and  $\mathcal{P}(x(t+1) = 1|x(t) = 0) = \mathcal{P}(x(t+1) = 1|x(t) = 1) = 1 - (1 - p(t))^k$ , and consequently the equation above is valid also in this case.

**Case 3: AND.** Node  $x$  has  $k$  inputs, and the Boolean rule is AND.

Thus, in order for  $x$  to be ON, all its inputs have to be ON. We obtain immediately

$$p(t+1)_{AND} = p(t)^k.$$

**Case 4:  $\mathcal{C}$ .** Node  $x$  has  $k$  inputs, and the Boolean rule is of class  $\mathcal{C}$ .

From Table 2 we observe that canalizing functions with one or two canalizing variables make up more than 80% of the canalizing functions in class  $\mathcal{C}$ . For simplicity we will focus

only on these two cases. Moreover, the fibroblast data indicate that nodes obeying the functions of this case cannot be their own canalizing inputs.

**Case 4.a:** Node  $x$  has  $k$  inputs, one of which is canalizing. Denote the canalizing node by  $y \neq x$ . According to Table 3 we will consider three possibilities for the canalizing and canalized values:  $0 \rightarrow 0$ ,  $1 \rightarrow 0$ , and  $1 \rightarrow 1$ .

Under the  $\mathbf{0} \rightarrow \mathbf{0}$  rule, that is  $[y(t) = 0 \Rightarrow x(t+1) = 0]$ , we get  $\mathcal{P}(x(t+1) = 1|x(t) = 1) = \mathcal{P}(x(t+1) = 1|x(t) = 1, y(t) = 1) \cdot \mathcal{P}(y(t) = 1) = p(t) \cdot \mathcal{P}(x(t+1) = 1|x(t) = 1, y(t) = 1)$ , where this last probability is given by the rule used when the canalizing input is not on its canalizing value. Similarly,  $\mathcal{P}(x(t+1) = 1|x(t) = 0) = \mathcal{P}(x(t+1) = 1|x(t) = 0, y(t) = 1) \cdot \mathcal{P}(y(t) = 1) = p(t) \cdot \mathcal{P}(x(t+1) = 1|x(t) = 0, y(t) = 1)$ .

Now, we will assume that when the canalizing input is not on its canalizing value the Boolean rule is a biased function with probability  $q_m$  that the output value is 1 if  $m$  of the  $k$  inputs are 1. Then it is known (see for example [12]) that  $\mathcal{P}(x(t+1) = 1|x(t) = 1, y(t) = 1) = \sum_{m=0}^{k-b} \binom{k-b}{m} q_{m+b} p(t)^m (1-p(t))^{k-b-m}$  and  $\mathcal{P}(x(t+1) = 1|x(t) = 0, y(t) = 1) = \sum_{m=0}^{k-b} \binom{k-b}{m} q_{m+1} p(t)^m (1-p(t))^{k-b-m}$  where  $b = 1$  when  $x$  is not its own input and  $b = 2$  if  $x$  is self-regulated. Then

$$\begin{aligned} \mathcal{P}(x(t+1) = 1) &= p(t)^2 \sum_{m=0}^{k-b} \binom{k-b}{m} q_{m+b} p(t)^m (1-p(t))^{k-b-m} + \\ &\quad + p(t)(1-p(t)) \sum_{m=0}^{k-b} \binom{k-b}{m} q_{m+1} p(t)^m (1-p(t))^{k-b-m} = \\ &= \sum_{m=0}^{k-b} \binom{k-b}{m} p(t)^{m+1} (1-p(t))^{k-b-m} [q_{m+b} p(t) + q_{m+1} (1-p(t))] \end{aligned}$$

which implies

$$p(t+1)_{(0 \rightarrow 0)} = p(t) \sum_{m=0}^{k-b} P_{m, k-b}(p(t)) \cdot \gamma(p(t), q_{m+b}, q_{m+1})$$

where  $P_{m, k-b}(p(t)) = \binom{k-b}{m} p(t)^m (1-p(t))^{k-b-m}$  is obtained from the binomial distribution, and  $\gamma(p(t), q_{m+b}, q_{m+1}) = q_{m+b} p(t) + q_{m+1} (1-p(t))$ . This notation is similar to that used in [8] for a mean-field formula of  $p(t)$  for totalistic Boolean rules. We note that any biased rule is a totalistic rule, in other words it depends on the density of ones rather than the individual node states.

Under the  $\mathbf{1} \rightarrow \mathbf{0}$  rule, by an argument similar to the case  $\mathbf{0} \rightarrow \mathbf{0}$ , we obtain

$$p(t+1)_{(1 \rightarrow 0)} = (1-p(t)) \sum_{m=0}^{k-b} P_{m, k-b}(p(t)) \cdot \gamma(p(t), q_{m+b-1}, q_m).$$

Finally, under the  $\mathbf{1} \rightarrow \mathbf{1}$  rule, we get the same result as in the case  $\mathbf{1} \rightarrow \mathbf{0}$  to which we add  $p(t)$  due to the canalizing input and value. Thus

$$p(t+1)_{(\mathbf{1} \rightarrow \mathbf{1})} = p(t) + (1-p(t)) \sum_{m=0}^{k-b} P_{m,k-b}(p(t)) \cdot \gamma(p(t), q_{m+b-1}, q_m).$$

**Case 4.b:** Node  $x$  has  $k$  inputs, two of which are canalizing. Thus the connectivity of the node is at least 2. We will consider the four situations for the pairing of canalizing - canalized values:  $(0or0 \rightarrow 0)$ ,  $(0and1 \rightarrow 0)$ ,  $(0or1 \rightarrow 0)$ , and  $(0or1 \rightarrow 1)$ , based on the fibroblast network characteristics. Denote the canalizing inputs by  $y \neq x$  and  $z \neq x$ .

Under the  $(\mathbf{0or0} \rightarrow \mathbf{0})$  rule, meaning  $[y(t)z(t) = 0 \Rightarrow x(t+1) = 0]$ , we get  $\mathcal{P}(x(t+1) = 1|x(t) = a) = p(t)^2 \mathcal{P}(x(t+1) = 1|x(t) = a, y(t) = 1, z(t) = 1)$  where  $a = 0$  or  $1$ . But  $\mathcal{P}(x(t+1) = 1|x(t) = 1, y(t) = 1, z(t) = 1) = \sum_{m=0}^{k-b} \binom{k-b}{m} q_{m+b} p(t)^m (1-p(t))^{k-b-m}$  and  $\mathcal{P}(x(t+1) = 1|x(t) = 0, y(t) = 1, z(t) = 1) = \sum_{m=0}^{k-b} \binom{k-b}{m} q_{m+2} p(t)^m (1-p(t))^{k-b-m}$  where this time  $b = 2$  when  $x$  is not its own input and  $b = 3$  if  $x$  is self-regulated. Then

$$p(t+1)_{(0or0 \rightarrow 0)} = p(t)^2 \sum_{m=0}^{k-b} P_{m,k-b}(p(t)) \cdot \gamma(p(t), q_{m+b}, q_{m+2}).$$

Similarly, under the  $(\mathbf{0and1} \rightarrow \mathbf{0})$  rule, meaning  $[|y(t) - z(t)| = 1 \Rightarrow x(t+1) = 0]$ , we get

$$p(t+1)_{(0and1 \rightarrow 0)} = \sum_{m=0}^{k-b} P_{m,k-b}(p(t)) \cdot [p(t)^2 \gamma(p(t), q_{m+b}, q_{m+2}) + (1-p(t))^2 \gamma(p(t), q_{m+b-2}, q_m)].$$

Under the  $(\mathbf{0or1} \rightarrow \mathbf{0})$  rule, meaning  $[(1-y(t))(1-z(t)) = 0 \Rightarrow x(t+1) = 0]$ , we obtain

$$p(t+1)_{(0or1 \rightarrow 0)} = (1-p(t))^2 \sum_{m=0}^{k-b} P_{m,k-b}(p(t)) \cdot \gamma(p(t), q_{m+b-2}, q_m).$$

Finally, under the  $(\mathbf{0or1} \rightarrow \mathbf{1})$  rule, meaning  $[(1-y(t))(1-z(t)) = 0 \Rightarrow x(t+1) = 1]$ , we have

$$p(t+1)_{(0or1 \rightarrow 1)} = 3p(t)(1-p(t)) + (1-p(t))^2 \sum_{m=0}^{k-b} P_{m,k-b}(p(t)) \cdot \gamma(p(t), q_{m+b-2}, q_m).$$

**Case 5: OTHER.** Node  $x$  has  $k$  inputs, and the Boolean rule is chosen as desired, but not identical to the rules discussed in the previous cases. To simplify our approach we select again a biased function with probability  $q_m$  that the output value is 1 if  $m$  of the  $k$

inputs are 1. The parameters are chosen according to the fibroblast network, and generate Boolean functions that have not been incorporated in the previous considerations.

Following a procedure similar to the one used in the previous cases, we obtain

$$p(t+1)_{OTHER} = \sum_{m=0}^{k-b+2} P_{m,k-b+2}(p(t)) \cdot \gamma(p(t), q_{m+b-2}, q_m)$$

where  $b = 2$  when  $x$  is not its own input and  $b = 3$  if  $x$  is self-regulated.

To make the notation more uniform, we use  $b = 2$  when  $x$  is not its own input and  $b = 3$  if  $x$  is self-regulated for all the canalizing functions in the class  $\mathcal{C}$  contained in **Case 4**.

We summarize all the formulae below.

$$(1) \quad p(t+1)_{COPY} = p(t)$$

$$(2) \quad p(t+1)_{OR} = 1 - (1 - p(t))^k.$$

$$(3) \quad p(t+1)_{AND} = p(t)^k.$$

$$(4) \quad p(t+1)_{(0 \rightarrow 0)} = p(t) \sum_{m=0}^{k-b+1} P_{m,k-b+1}(p(t)) \cdot \gamma(p(t), q_{m+b-1}, q_{m+1})$$

$$(5) \quad p(t+1)_{(1 \rightarrow 0)} = (1 - p(t)) \sum_{m=0}^{k-b+1} P_{m,k-b+1}(p(t)) \cdot \gamma(p(t), q_{m+b-2}, q_m)$$

$$(6) \quad p(t+1)_{(1 \rightarrow 1)} = p(t) + (1 - p(t)) \sum_{m=0}^{k-b+1} P_{m,k-b+1}(p(t)) \cdot \gamma(p(t), q_{m+b-2}, q_m)$$

$$(7) \quad p(t+1)_{(0 \text{ or } 0 \rightarrow 0)} = p(t)^2 \sum_{m=0}^{k-b} P_{m,k-b}(p(t)) \cdot \gamma(p(t), q_{m+b}, q_{m+2})$$

$$(8) \quad p(t+1)_{(0 \text{ or } 1 \rightarrow 0)} = (1 - p(t))^2 \sum_{m=0}^{k-b} P_{m,k-b}(p(t)) \cdot \gamma(p(t), q_{m+b-2}, q_m)$$

$$(9) \quad p(t+1)_{(0 \text{ and } 1 \rightarrow 0)} = p(t)^2 \sum_{m=0}^{k-b} P_{m,k-b}(p(t)) \cdot \gamma(p(t), q_{m+b}, q_{m+2}) +$$

$$(1 - p(t))^2 \sum_{m=0}^{k-b} P_{m,k-b}(p(t)) \cdot \gamma(p(t), q_{m+b-2}, q_m)$$

$$(10) \quad p(t+1)_{(0or1 \rightarrow 1)} = 3p(t)(1-p(t)) + (1-p(t))^2 \sum_{m=0}^{k-b} P_{m,k-b}(p(t)) \cdot \gamma(p(t), q_{m+b-2}, q_m)$$

$$(11) \quad p(t+1)_{OTHER} = \sum_{m=0}^{k-b+2} P_{m,k-b+2}(p(t)) \cdot \gamma(p(t), q_{m+b-2}, q_m).$$

#### 4. AGGREGATED MEAN FIELD MODEL AND SIMULATIONS

Now we are able to aggregate the individual MF-models and provide a general formula for the density of ones under the Boolean rules analyzed above. This MF-model will be used to explore the dynamics of the fibroblast network under various scenarios.

First we consider the partition of the nodes in the classes defined in Section 3. Each class contains all the nodes of the network that obey the same Boolean rule. Thus there are 11 different classes in the partition:  $C_1 = \widehat{COPY}$ ,  $C_2 = \widehat{OR}$ ,  $C_3 = \widehat{AND}$ ,  $C_4 = \widehat{(0 \rightarrow 0)}$ ,  $C_5 = \widehat{(1 \rightarrow 0)}$ ,  $C_6 = \widehat{(1 \rightarrow 1)}$ ,  $C_7 = \widehat{(0or0 \rightarrow 0)}$ ,  $C_8 = \widehat{(0or1 \rightarrow 0)}$ ,  $C_9 = \widehat{(0and1 \rightarrow 0)}$ ,  $C_{10} = \widehat{(0or1 \rightarrow 1)}$ , and  $C_{11} = \widehat{OTHER}$ . Then  $V = \bigcup_{j=1}^{11} C_j$ . For simplicity, all the nodes in a class will be assumed to have the same number of inputs.

Now we can start aggregating the density of ones for all these classes. The density of ones corresponding to a class  $C_j$  is  $p_{C_j}(t+1)$  given by one of the formulas in Section 3. If we denote by  $\alpha_j$  the fraction of nodes of the network in class  $C_j$ , then the aggregated density of ones, or the probability that any node of the network is 1 at time  $t+1$ , is given by

$$(12) \quad p(t+1) = \sum_{j=1}^{11} \alpha_j p_{C_j}(t+1) \quad \text{where} \quad \sum_{j=1}^{11} \alpha_j = 1.$$

To test the accuracy of the aggregated MF-model (12), we simulate the density of ones and plot the results for both the MF-model and the actual fibroblast network for comparison. We use parameter values obtained from the actual fibroblast signal transduction network as described below.

To determine the connectivity associated with each class  $C_j$ , we select the most frequent connectivity that corresponds to each class according to the fibroblast network model. In case the largest frequency corresponds to more than one connectivity level, we consider the average connectivity in the MF-model. The most frequent connectivity values  $k$  and their frequencies are included in Table 4. For example, in class  $C_3 = \widehat{AND}$  the value  $k_3 = 2$  is the most frequent and therefore this one is used in the simulations. On the other hand, in class  $C_7 = \widehat{(0or0 \rightarrow 0)}$  there are four different values of  $k$  with the highest frequency,

namely 3, 4, 6, 14. Thus in the simulation we use the (rounded up) average connectivity  $k = 7$ . In parentheses we provide the corresponding percentages for each class, which are used as estimates for the parameters  $\alpha_j$  in the MF-model (12). We note here that the actual chosen connectivity values in this situation do not make a big difference in the simulation results. Therefore we choose not to increase the computational costs by varying the connectivity. We use the simplified version where all nodes in a class have the same number of inputs given by the average connectivity for that class in the fibroblast network. However, later we will vary the connectivity of each class to understand its impact on the network behavior.

Class	most frequent $k$	frequency ( $\alpha_j$ )
$C_1 = \widehat{COPY}$	1	18 (13.8%)
$C_2 = \widehat{OR}$	2	10 (7.7%)
$C_3 = \widehat{AND}$	2	9 (6.9%)
$C_4 = \widehat{(0 \rightarrow 0)}$	3	16 (12.3%)
$C_5 = \widehat{(1 \rightarrow 0)}$	3	4 (3.1%)
$C_6 = \widehat{(1 \rightarrow 1)}$	3	8 (6.2%)
$C_7 = \widehat{(0or0 \rightarrow 0)}$	3,4,6,14 (average = 7)	4 (3.1%)
$C_8 = \widehat{(0or1 \rightarrow 0)}$	4,6,8,10 (average = 7)	4 (3.1%)
$C_9 = \widehat{(0and1 \rightarrow 0)}$	2	8 (6.2%)
$C_{10} = \widehat{(0or1 \rightarrow 1)}$	3,4 (average = 4)	2 (1.5%)
$C_{11} = \widehat{OTHER}$	4	47 (36.2%)

TABLE 4. Most frequent connectivity values in each class and the corresponding proportions with respect to all Boolean functions. When multiple  $k$  values have the highest frequency in a class, a rounded up average is used in the MF-model. The proportions  $\alpha_j$  to be used in the MF-model 12 are also provided.

Furthermore, the fibroblast network information reveals that all the classes except  $C_7$  are non-self-regulatory. Therefore, we set  $b = 3$  for  $C_7$  and  $b = 2$  for the rest of the classes in the aggregated MF-model (12).

Finally, we determine  $q_m$ , that is the probability that the output value is 1 if  $m$  of the  $k$  inputs of a function that belongs to either  $C$  or OTHER class are 1. We note that there is a clear distinction between the biased functions used in  $\mathcal{C}$  and the ones used in the OTHER class (Figure 2). This is because of the essential difference in the nature of the functions in these two classes: most of the canalizing functions in  $\mathcal{C}$  have inhibitory

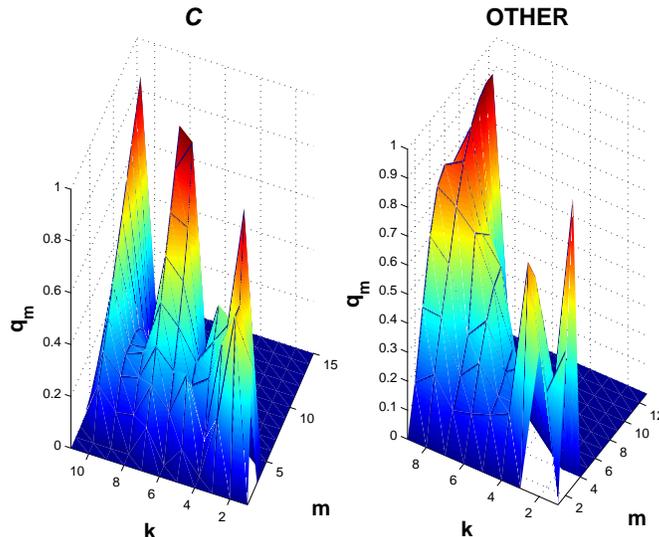


FIGURE 2. The probability that the output value is 1 if  $m$  of the  $k$  inputs of a function in  $\mathcal{C}$  (right) and OTHER (left) are 1.

effects on a node while functions in OTHER have more activating effects. Therefore in simulations we use different values for the parameter  $q_m$  for these two classes. To obtain  $q_m$ , the nodes obeying rules in  $\mathcal{C}$  are separated from the nodes with rules in OTHER class. After that the  $q_m$  values for the nodes in each class are estimated as frequencies from the logical tables, and sorted according to the connectivity level  $k$ . Finally, the average  $q_m$  for each  $k$  is calculated. The left graph of Figure 2 shows the average  $q_m$  values for the corresponding  $k$  and  $m$  values associated with functions in  $\mathcal{C}$ , while the right graph of Figure 2 shows the average  $q_m$  values for  $k$  and  $m$  values associated with functions in the OTHER class.

Now that all parameters have been identified from the fibroblast network, we generate simulations that allow us to check the accuracy of the MF-model. To this aim, In Figure 3 we plot  $p(t+1), p(t+2), p(t+3), p(t+64)$  versus  $p(t)$  for the MF-model (red line) and the actual network (green dots). We select a large number of possible initial values  $p(t)$  in  $[0, 1]$ . When iterating the network, 100 different initial states are chosen such that the frequency of active nodes is the desired value of  $p(t)$ . In this context, the active nodes are selected randomly from 130 nodes within the network. However, since in the original fibroblast network 9 external nodes are inputs to some of these 130 nodes, their states must be defined to update the network. Therefore, the average of  $p(t+1), p(t+2), p(t+3)$  and  $p(t+64)$  is calculated over all  $2^9 = 512$  different states of the external inputs respectively. The states of the external inputs are dynamically changed according to the procedures

used in [1], which generate a “background noise” that exists in all biological systems. In this paper we focus on the dynamics of the molecules internal to the fibroblast cell, so we average the effect of the external noise.

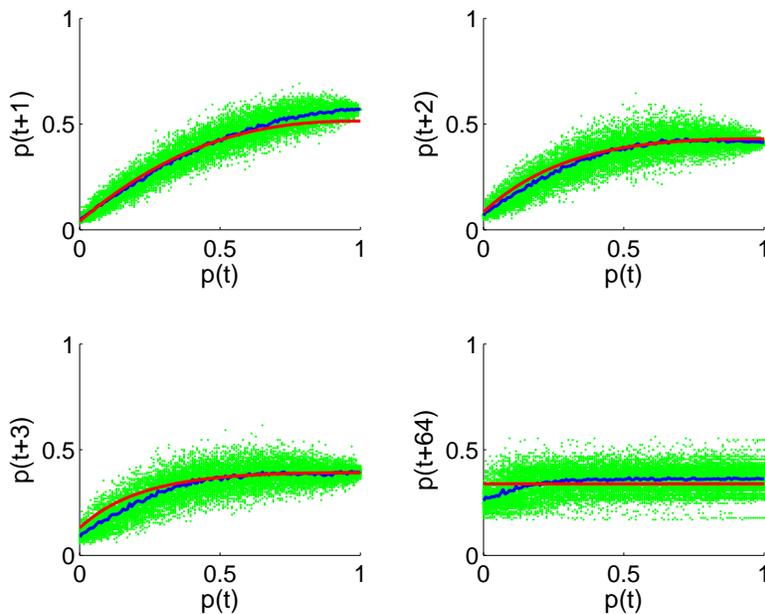


FIGURE 3.  $p(t+1), p(t+2), p(t+3)$  and  $p(t+64)$  versus  $p(t)$ . The green dots indicate the results of 100 simulations, each of which is obtained by averaging over 512 different initial conditions of external inputs with a fixed initial condition of fixed density of ones. The blue line indicates the overall average of these simulations. The aggregated MF-model (12) is indicated by the red line.

Thus the results of the 100 simulations are represented as green dots and their average as the blue line. We observe right away that the aggregated MF-model shown in red is a very good approximation of the (average) results obtained by iterating the fibroblast network. We have actually simulated a larger number of iterations, but the results are similar. The density of ones converges to a horizontal line. Despite the relatively small network size in comparison to the assumption of a large network for the mean-field approach, the MF-model captures the actual long term behavior of the network. We supplement the visual match with a computation of correlation coefficients between the MF-model and the network values of the density of ones for all 64 iterations related to Figure 3. We also compute the corresponding Euclidean distances between the vectors of points generating the two curves in each case. The results are contained in Figure 4. Note that the correlation coefficient is very high meaning that there is a very strong correlation between the MF-model and the network results, while the Euclidean distance fluctuates

around 0.3, which is a low value in comparison to the maximum possible Euclidean distance of  $\sqrt{130} \sim 11.4018$  (since the vector length is 130 and the vectors consist of values in  $[0, 1]$ ).

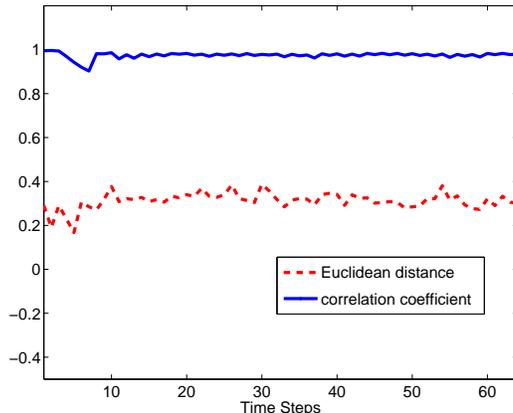


FIGURE 4. Correlation coefficients and Euclidean distances between the MF-model and network results for the 64 iterations corresponding to Figure 3. The correlation coefficient is very high, while the Euclidean distance between the vectors generating the two curves of Figure 3 is about 0.3, much smaller than the maximum possible Euclidean distance of about 11.4.

The MF-model in this paper relies on the classification of the Boolean rules into various classes. The authors of [1] show that both the topology and the logic are important for the dynamics of the network. However, to emphasize the importance of the Boolean rules we perform simulations of the network similar to those in Figure 3, in which we preserve the topology of the network, that is the nodes and the connectivity, but we destroy the logic by using randomly chosen biased functions with the actual true bias of the network, 0.4368. As expected, the plots in Figure 5 show stability around the bias starting with the first iteration, which is the only one shown. Higher order iterations are almost identical, so the first iteration is a good representation of the long run behavior. We plot the actual MF-model from Figure 3 for both  $p(t + 1)$  (red curve) and  $p(t + 64)$  (red dotted line) for comparison. Note that the fact that in the long run the activity level settles is valid for both the fibroblast and the biased network, but the usage of different types of rules in the MF-model makes it a significantly better approximation for the network's density of ones from the first iteration as seen in Figure 3. Therefore the logic is significant. We note here that a MF-model for the biased network yields the trivial formula  $p(t + 1) = p(t)$ , a horizontal line located at the level of the initial condition for the density of ones.

In conclusion, we can use the MF-model as a baseline for further investigations of the behavior of the system under variation of parameters and noise procedures. We keep in

mind that the MF-model is a good representation of the average density of ones obtained over all possible external input combinations.

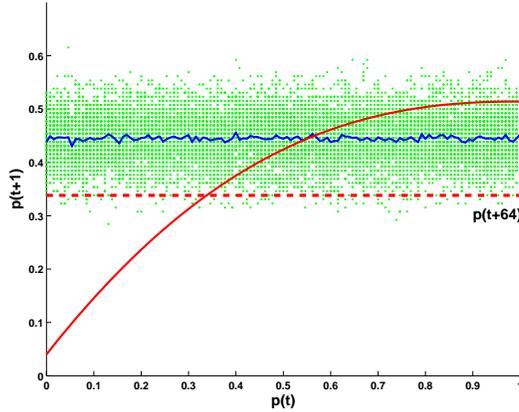


FIGURE 5.  $p(t+1)$  versus  $p(t)$  for a biased network preserving the topology of the fibroblast network. The bias is set equal to the actual bias of the fibroblast network of 0.4368. Higher order iterations yield similar results. This figure is an analog of Figure 3. The red curve represents  $p(t+1)$ , and the red dotted line represents  $p(t+64)$  from the MF-model of Figure 3 for comparison. Note the mismatch which is due to ignoring the variety of Boolean rules in the network.

## 5. NETWORK DYNAMICS AND PROTEIN MUTATIONS

Since the aggregated MF-model (12) is a good approximation of the fibroblast network, we use the MF-model to explore the dynamics of the network under various scenarios for the parameter combinations. In this paper we are mainly interested in exploring the overall dynamics of (12) and compare them with existing conclusions on the dynamics of the fibroblast model in [1]. Therefore the following numerical results refer to the aggregated density of ones (12), and we use this MF-model to set a baseline for future research. However, due to the biological significance of the activity of certain types of nodes as opposed to the fraction of active nodes in the entire network, our future work will assess the importance of the individual node types and their impact not only on the aggregated networks, but also on other types of nodes. The MF-model presented here will be our baseline for studies on the types of nodes/dynamical rules that exhibit sensitivity to perturbations, thus having the potential to modify the dynamics of the whole network or other types of nodes. Such nodes could become potential targets in drug therapies. At the end of this section we provide one example in which we focus on the activity level of one individual class of nodes.

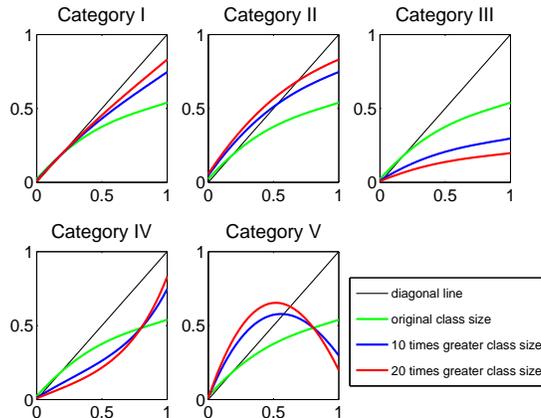


FIGURE 6. Typical behavior of classes in each of the five categories. The green line represents the plot of  $p(t+1)$  vs.  $p(t)$  for the MF-model (12) with equal class weights. The blue (red) line indicates the behavior of MF-model (12) when the weight of a class belonging to the indicated category is increased 10 (20) times.

To our knowledge the Boolean network in [1] is among the largest biological networks in Boolean representation in existence. To reduce the computational costs associated with a network of this size and complexity, we note that for the purpose of assessing the dynamics of the fraction of active nodes we can group the 11 different types of functions in categories based on the overall shape of the density of ones. To this aim we investigate the impact of a given class by first setting the size of all the classes to be the same and then increasing the weight of the class of interest to observe its impact on the MF-model. We find that some classes can be grouped together since they behave somewhat similarly, due to their intrinsic nature. Thus we regroup the original 11 different types of classes into 5 categories as shown in Figure 6 and Table 5. The size of each category is determined by the sum of the frequencies of its classes. Observe that the five categories can be viewed as follows:  $C_1$  - slightly inhibitory functions,  $C_2$  - functions that are slightly activating for small activity levels and inhibitory for high activity levels; these functions tend to keep the overall activity level at moderate values,  $C_3$  - strongly inhibitory functions,  $C_4$  - functions that are strongly inhibitory for small activity levels and mildly inhibitory for large activity levels,  $C_5$  - functions that are strongly activating for small activity levels and strongly inhibitory for high activity levels. In preliminary simulations we did not observe a significant impact of this simplification at least from the point of view of the density of ones. However, future explorations with the original classification are likely to indicate the necessity to consider more refined classes of functions that go beyond the original 11 class grouping.

Category	Class	Frequency
I	$C_1$	18
II	$C_2, C_6$	18
III	$C_5, C_8, C_9$	16
IV	$C_3, C_4, C_7$	29
V	$C_{10}, C_{11}$	49

TABLE 5. The 11 different types of functions are separated into 5 categories. The frequency of each category is determined by the sum of the frequencies of classes included in that category.

Next we select a representative class in each category and construct the MF-model including only those classes. Among all possible 36 different combinations of representative classes,  $C_1, C_2, C_9, C_4$  and  $C_{11}$  are selected as representatives for the categories I - V, since they yield the smallest Euclidean distance between the simplified MF-model with 5 categories and the original MF-model (12) with 11 classes of functions. The Euclidean distance is computed between vectors used to generate Figure 3.

Thus we will focus on these 5 different types of functions. Future work will include a more in depth analysis of the MF-model by refining the types of functions back to 11 categories for an exhaustive look at the dynamics of the network.

The advantage of having a mathematical model such as the MF-model is that one could use it to make predictions on the behavior of the original system. Given that the MF-model proposed here is based on a fairly large number of parameters, it would be interesting to see what is the impact of some of these parameters on the dynamics. The parameters are: the connectivity  $k$  for each type of Boolean function, representing the number of inputs (regulators) of the given nodes; the probabilities  $q_m$  for the canalizing functions as well as the OTHER functions representing the probability that a node becomes active if  $m$  of its inputs are active; the parameter  $b$  for each class of functions, which indicates allowable self-regulatory properties that are directly observed in the fibroblast network for all types of functions; the fractions  $\alpha$  of the classes of functions, indicating the weight of each class of functions among all the functions in the network, in other words the presence of each type of regulatory functions in the biological network.

So let us explore some aspects of the dynamics of the network. We start with bifurcation diagrams for the representative classes in the five categories. All parameters are set to their values obtained from the fibroblast model, except the connectivity of each of the five classes which is varied for a few selected values of the weight  $\alpha$  of that class, including the

values that correspond to the fibroblast network. The remaining weights are re-scaled by adding/subtracting an equal amount. We note that in all situations, the system exhibits order with a single attracting fixed point for each value of  $k$ . The fixed points are situated approximately in the interval  $[0, 1/2]$ . The results are shown in Figure 7. One important observation that cannot be made in a laboratory experiment or even by using the Boolean representation of the fibroblast network [1] which is subject to structural restrictions, is that for  $\alpha_{11} = 0$  shown in the  $C_{11}$  subplot, the origin is a stable fixed point for all connectivity levels. This suggests that the functions incorporated in class  $C_{11}$  have a significant impact on the dynamics of the system. Once they are turned off, basically the entire system is turned off. Recall now that  $C_{11}$  consists of all Boolean functions that did not fit any of the other ten classes of functions, and that we modeled them by using biased functions. Future work will focus on a deeper understanding of the functions in this class, and their role in the system evolution. The MF-model has identified this class as a potential key factor for cell death, in comparison to the other classes of nodes whose suppression from the network does not stop the cell activity.

In conclusion, when starting with parameters as in the fibroblast model, a modification of the connectivity level or the weight of the class does not produce a significant qualitative change in the long-run behavior of the system. We note here that according to the findings of [1], the fibroblast network exhibits stability with a moderate activity level similar to what is seen in Figure 3. However, according to the bifurcation diagrams, quantitatively the situation can change under a modified scenario. For example an increase in the presence of class  $C_2$  (category II) nodes can lead to a rather aggressive increase in the activity level. On the other hand, an increase in the presence of class  $C_4$  (category IV) nodes could lead to a less aggressive but still significant decrease of the activity level. This would mean that for instance networks that are somewhat similar to the fibroblast network but in which class  $C_2$  nodes are more prevalent, are expected to exhibit a higher overall activity level. On the other hand, networks in which the presence of class  $C_{11}$  is minimal, exhibit lack of activity.

As a matter of fact, even if we apply more variation on the parameters  $q_m$  which occur in the canalizing functions as well as in class  $C_{11}$ , the system exhibits stability in the long run. Such modifications can be looked upon as mutations of the node since they imply changes in the outputs of the truth tables. In simulations we explore the following types of distributions for  $q_m$ : uniform, normal, power law, and chi square. These produce various possibilities for the values of  $q_m$ . For example mainly small values of  $q_m$  for small

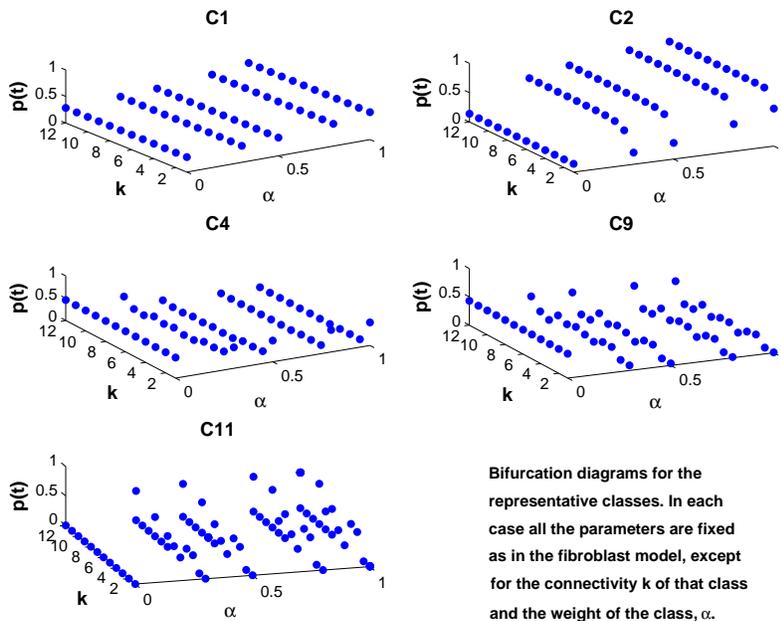


FIGURE 7. Bifurcation diagram for the representative classes for the five categories, whose connectivity is varied from 1 to 12 which is the maximum value obtained from the fibroblast network. A few values for the weight of each class are considered. All the other parameters are fixed according to the statistics obtained from the fibroblast model. Observe that the system exhibits stability for all connectivity values. The points are plotted after iterating many initial values for 200 time steps.

$m$  would mean that a node under consideration is inhibited by a lack of activity in its neighborhood, while large values of  $q_m$  for small  $m$  would mean that a few active inputs are sufficient to activate the node, so there is a bias toward the activation of the node. On the other hand, small values of  $q_m$  for large  $m$  corresponds to the case where a large number of active inputs has an inhibitory effect on the output so there is a bias towards inhibition, while large values of  $q_m$  for large  $m$  means that many active inputs have a significant chance of activating the node under consideration. All of the distributions used for  $q_m$  yield stability with fixed points or periodic orbits. The bifurcation diagrams are similar to those in Figure 7, and therefore are not included here.

The study of the robustness of a Boolean network to various types of perturbations is an important aspect of the evolution of systems under Boolean models. These systems have to respond and adapt to interior and exterior disturbances. In this paper we investigate briefly the effect of simple mutations of nodes belonging to the five representative classes  $C_1$ ,  $C_2$ ,  $C_4$ ,  $C_9$ , and  $C_{11}$  respectively. We show that the MF-model (12) can make good predictions of the behavior of the fibroblast network [1]. We consider the following two cases of mutations: (1) Nodes in the target class are mutated and keep signaling, so

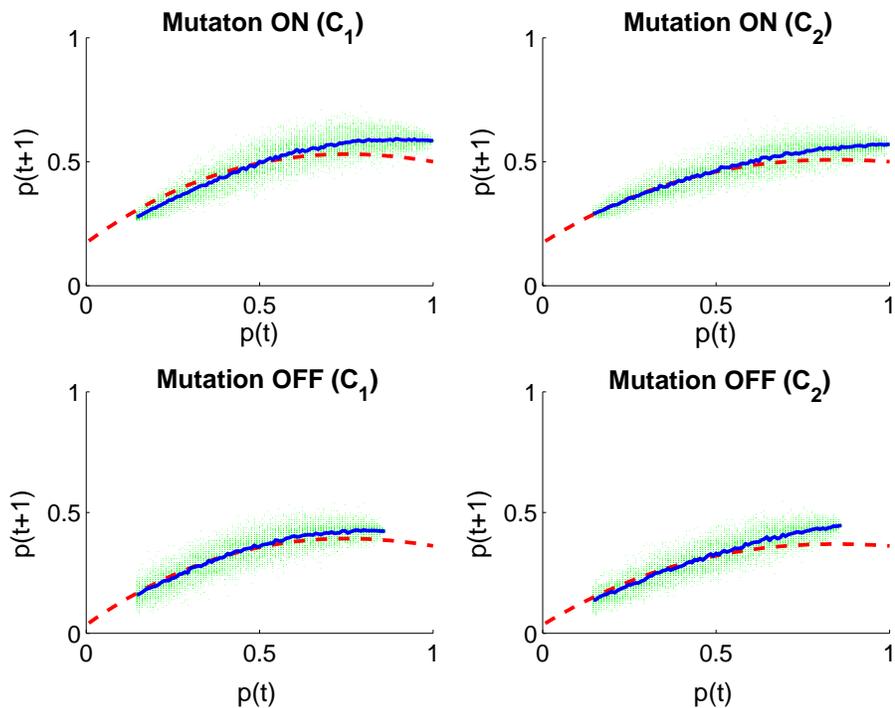


FIGURE 8.  $p(t + 1)$  versus  $p(t)$  in case of protein mutations. For each value of  $p(t)$ , 100 different selections of external inputs in the fibroblast network are considered and the green dots indicate the results of these 100 simulations. The blue line indicates the average of these simulations. The aggregated MF-model with mutation is indicated by the dotted red lines. Top left: the nodes belonging to  $C_1$  are mutated ON. Top right: the nodes belonging to  $C_2$  are mutated ON. Bottom left: the nodes belonging to  $C_1$  are mutated OFF. Bottom right: the nodes belonging to  $C_2$  are mutated OFF.

their state is always 1 or ON; in the fibroblast network the corresponding outputs of the truth tables are set to 1, while in the MF-model the corresponding density of ones is set equal to 1; and (2) Nodes in the target class never turn on, so their state is always 0 or OFF; in the fibroblast network the corresponding outputs of the truth tables are set to 0, while in the MF-model the corresponding density of ones is set equal to 0. Now, nodes belonging to a given class of Boolean functions may have a common behavior in the network. For example, both proteins Grb2 and Nck belong to class  $C_{10}$ , and they are both adaptor proteins (that is they are accessory to main proteins in a signal transduction pathway mediating specific protein-protein interactions that drive the formation of protein complexes). By mutating an entire class of nodes we may be able to associate biological and mathematical aspects of a particular class of nodes which may not be observed in laboratory.

The weights  $\alpha_j$  are selected as in the fibroblast network. Then we iterate the Boolean network and the fibroblast model of [1] and obtain  $p(t+1)$  for a large number of possible initial values  $p(t)$  in  $[0, 1]$ . The mutations on classes  $C_1$  and  $C_2$  are shown in Figure 8. The other cases are similar, and are not included here. For each fixed initial value  $p(t)$ , we consider 100 different combinations of the 9 external inputs in the fibroblast network. We plot  $p(t+1)$  versus  $p(t)$  for all 100 simulations with green dots together with their average in blue. The dotted red line represents the aggregated MF-model with mutation. We can see that the red and blue curves are close. The MF-model tends to underestimate the density of ones for larger initial values  $p(t)$ . However, these results confirm that the aggregated MF-model serves as a good prediction tool for the fibroblast signal transduction network under mutations. Observe also that in comparison to Figure 3, top left graph, the overall activity level is slightly increased under mutation ON and slightly decreased under mutation OFF, which is to be expected for classes  $C_1$  and  $C_2$ . In particular class  $C_2$  is the OR function which is by its nature an activator so its mutation ON would not result in a significant increase in the activity level, while its mutation OFF would result in a more clear decrease of activity. The actual weight of these two classes in the original fibroblast network is about 13% for  $C_1$  and 7% for  $C_2$ . Given the results of the bifurcation diagrams of Figure 7, an increased presence of these classes in the network would increase the baseline activity level and the impact of their mutations on the overall network activity.

The results in this section support previous research results on the stability of biologically meaningful Boolean functions and networks [1], [9], [10], [11].

As specified in the beginning of this section, due to the biological significance of the activity level of certain types of nodes as opposed to the overall fraction of active nodes of the network, it is of interest to identify the types of nodes that exhibit sensitivity to perturbations, thus having the potential to modify their own behavior and the dynamics of the whole network or other types of nodes. Such nodes could become targets in drug therapies. We provide one example. In Figure 9 we present bifurcation diagrams along the connectivity parameter  $k$ , paired with Lyapunov exponent computations for a network consisting of nodes of class  $C_{10}$  only. The left graphs show that this type of nodes have an intrinsic chaotic nature for a significant number of connectivity values. However, when applying a mutation on the parameters  $q_m$ , more precisely when  $q_m$  obeys a different type of function than the one specific for class  $C_{10}$  (in particular a function that is typical for class  $C_{11}$ ), the network exhibits order for an increased number of connectivity values

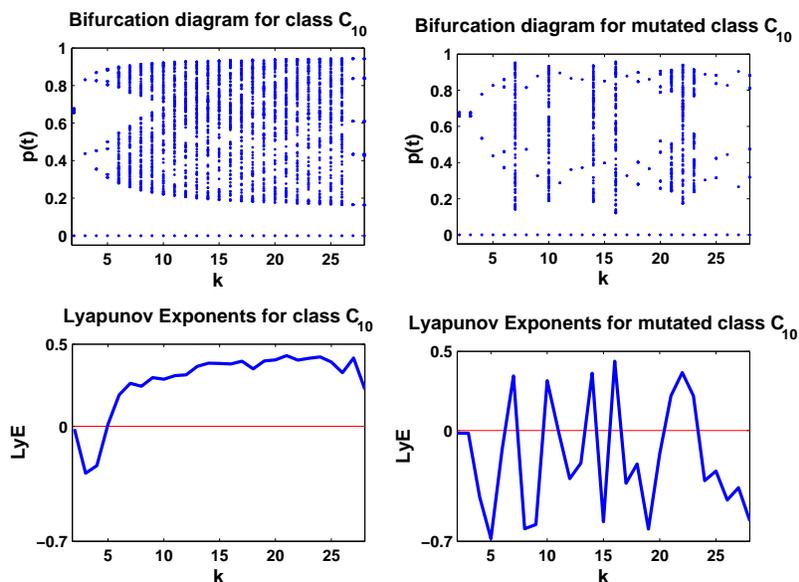


FIGURE 9. Bifurcation diagrams and corresponding Lyapunov exponents (LyE) for a network consisting solely of  $C_{10}$  nodes. The left graphs show that this type of nodes exhibit a chaotic nature for various connectivity values. However, under a mutation on the parameters  $q_m$ , the results shown in the right graphs indicate that the network can be stabilized for a significant number of connectivity values.

(the right graphs). Thus, one can use this MF-model to find for example, types of nodes, as well as  $q_m$  and connectivity values that will lead to a desired long-term behavior. A comprehensive study of the roles of each individual class of nodes on the activity level of other nodes is still to be performed. However, this example suggests that the MF-model proposed in this paper can be used to identify types of nodes, parameters, and mutations that yield a desired dynamical regime.

We will finalize our analysis with a discussion on fixed points in the next section. However, before that, we summarize the main steps of the modeling and simulation procedures in this paper in Table 6.

## 6. ON FIXED POINTS

We finalize our analysis by focusing on the fixed points of the maps included in the aggregated MF-model (12). Basically this reduces to solving the system:

$$p_{C_j}(t+1) = p_{C_j}(t) \quad j = 1, 2, \dots, 11$$

<b>Modeling steps and assumptions</b>
Identify the classes $C_1 - C_{10}$ of Boolean functions in the fibroblast network.
Class $C_{11}$ incorporates the (yet) non-identified functions. These are assumed to be biased rules.
All nodes in a class are assumed to have the same connectivity.
Individual MF-models are generated for each Boolean class. For canalizing functions it is assumed that biased rules are used when the canalizing input is not on its canalizing value.
The weighted aggregation of the models is constructed, where the weights represent the proportion of classes of functions in the entire network.
<b>Simulation steps</b>
Estimate the average connectivity by classes of functions of the actual network.
Estimate the weights of the classes using the proportions from the actual network.
Estimate the biases using empirical bias distributions of the actual network for each class involving biased functions.
With these estimates run the MF-model against the actual network iterations to confirm the accuracy of the MF-model.
Once the MF-model is shown to be a good approximation of the original network, various other scenarios can be explored: Example: Vary parameters to assess their impact on the overall evolution of the density of ones. Example: Freeze some nodes or classes of nodes to identify the role of mutations. Example: Apply variation of parameters for one class of nodes and assess the impact on activity levels of other classes.

TABLE 6. The main steps of modeling and simulations, together with the basic assumptions.

where  $p_{C_j}(t+1)$  are given by the formulae (1)-(11). Note that in these formulae,  $p(t)$  is given by (12). The system may be solved numerically for specific values of the parameters. However, we are interested in providing some theoretical results.

Instead of giving a full description of all the computations, we only show a few examples. All the others are done in a similar way. For instance, if the map under consideration is the AND map in (3), then we have to solve the equation  $p^k = p_{C_3}$ , where  $p = \sum_{j=1}^{11} \alpha_j p_{C_j}$ . But note that if  $k \rightarrow \infty$  then  $p^k \rightarrow 0$  and therefore  $p_{C_3} \rightarrow 0$ . Thus for the AND map the fixed points approach zero as the connectivity increases. Similarly, under the OR rule,  $p_{C_2} \rightarrow 1$  as  $k$  increases. On the other hand, the COPY map (1) yields the fixed point equation  $(1 - \alpha_1)p_{C_1} = \sum_{j=2}^{11} \alpha_j p_{C_j}$ .

Now let us look at a canalizing function. Under the  $(0 \rightarrow 0)$  map (4), we obtain the fixed point condition  $p_{C_4} = p \sum_{m=0}^{k-1} \binom{k-1}{m} p^m (1-p)^{k-1-m} q_{m+1}$ . Recall that in this case the parameter  $b = 2$  in (4). To understand what happens to the fixed points as the connectivity increases, consider the generic binomial sum  $\sum_{m=0}^k \binom{k}{m} p^m (1-p)^{k-m} \cdot q_m$ ,

where  $m, k$  are integers,  $p \in [0, 1]$  and  $q_m \in [0, 1], \forall m$ . We would like to see what happens with this sum as  $k \rightarrow \infty$ . Asymptotic results are common practice in dynamical systems and chaos theory. Despite the fact that in the fibroblast network the connectivity level is rather small, the true protein molecule networks have a much larger number of nodes and connectivity levels. As a matter of fact, it is estimated that some cells such as a typical eukaryotic cell like a hepatocyte in the liver have close to a million protein molecules in a cell with hundreds of interactions between cells [13]-[14]. Thus, asymptotic results are a feasible approach to understanding the dynamics of the true networks.

So, consider the sum

$$S_k = \sum_{m=0}^k \binom{k}{m} p^m (1-p)^{k-m} \cdot q_m$$

where  $m, k$  are integers, and  $p \in [0, 1]$ , and let  $\{q_m\}$  be any bounded sequence of nonnegative numbers.

**Proposition 1.** *Under the assumptions above, one has that*

$$(13) \quad \liminf_{k \rightarrow +\infty} q_k \leq \liminf_{k \rightarrow +\infty} S_k \leq \limsup_{k \rightarrow +\infty} S_k \leq \limsup_{k \rightarrow +\infty} q_k.$$

Hence, if  $q_m \rightarrow q$  when  $m \rightarrow +\infty$ , then

$$\lim_{k \rightarrow +\infty} S_k = q.$$

Denote

$$q := \liminf_{k \rightarrow +\infty} q_k, \quad Q := \limsup_{k \rightarrow +\infty} q_k, \quad l := \liminf_{k \rightarrow +\infty} S_k, \quad \text{and} \quad L := \limsup_{k \rightarrow +\infty} S_k.$$

Since  $0 \leq q_m$  for all  $m$ , clearly one has that  $0 \leq q$ . First, assume  $q > 0$ .

Note that, for all fixed positive integers  $m_0 > 1$

$$(14) \quad \lim_{k \rightarrow +\infty} \sum_{m=0}^{m_0-1} \binom{k}{m} p^m (1-p)^{k-m} \cdot q_m = 0$$

because power growth is weaker than exponential growth at  $+\infty$ , and  $\{q_m\}$  is a bounded sequence.

Now, for all fixed positive integers  $m_0 > 1$ ,

$$(15) \quad \lim_{k \rightarrow +\infty} \sum_{m=m_0}^k \binom{k}{m} p^m (1-p)^{k-m} = 1.$$

Denote

$$S_k(m_0) := \sum_{m=m_0}^k \binom{k}{m} p^m (1-p)^{k-m} \cdot q_m$$

and note that, by (14),

$$(16) \quad \liminf_{k \rightarrow +\infty} S_k(m_0) = l \quad \text{and} \quad \limsup_{k \rightarrow +\infty} S_k(m_0) = L.$$

Now choose  $0 < \epsilon < q$  arbitrary. There is some  $m_0 > 1$  so that

$$0 < q - \epsilon < q_m < Q + \epsilon \quad m \geq m_0.$$

The consequence is

$$(q - \epsilon) \sum_{m=m_0}^k \binom{k}{m} p^m (1-p)^{k-m} \leq S_k(m_0) \leq (Q + \epsilon) \sum_{m=m_0}^k \binom{k}{m} p^m (1-p)^{k-m}.$$

Let  $k \rightarrow +\infty$  and note that, by (15) and (16), one gets

$$(q - \epsilon) \leq l \leq L \leq (Q + \epsilon) \quad 0 < \epsilon < q.$$

Now let  $\epsilon \rightarrow 0$ . One gets

$$q \leq l \leq L \leq Q,$$

which proves (13) if  $q > 0$ .

If  $q = 0$ , one can repeat the argument above, establishing that

$$0 \leq S_k(m_0) \leq (Q + \epsilon) \sum_{m=m_0}^k \binom{k}{m} p^m (1-p)^{k-m},$$

for some  $m_0$ , then let  $k \rightarrow +\infty$ , and after that  $\epsilon \rightarrow 0$ , thus obtaining the inequalities

$$0 \leq l \leq L \leq Q.$$

□

Using equation (13) one can obtain the following bounds for  $p_{C_j}, j = 1, 2, \dots, 11$ , when we allow  $k \rightarrow \infty$ . We make the simplifying assumption that the sequence of probabilities  $q_m$  is the same for all classes of functions. Recall that  $q = \liminf_{m \rightarrow +\infty} q_m$  and  $Q = \limsup_{m \rightarrow +\infty} q_m$ . Then

$$(17) \quad \begin{aligned} p_{C_1} &= p; & p_{C_2} &= 1; & p_{C_3} &= 0; \\ p \cdot q &\leq p_{C_4} \leq p \cdot Q; \\ (1-p) \cdot q &\leq p_{C_5} \leq (1-p) \cdot Q; \end{aligned}$$

$$\begin{aligned}
 p + (1 - p) \cdot q &\leq p_{C_6} \leq p + (1 - p) \cdot Q; \\
 p^2 \cdot q &\leq p_{C_7} \leq p^2 \cdot Q; \\
 (1 - p)^2 \cdot q &\leq p_{C_8} \leq (1 - p)^2 \cdot Q; \\
 [p^2 + (1 - p)^2] \cdot q &\leq p_{C_9} \leq [p^2 + (1 - p)^2] \cdot Q; \\
 [3p(1 - p) + (1 - p)^2 \cdot q] &\leq p_{C_{10}} \leq [3p(1 - p) + (1 - p)^2 \cdot Q]; \\
 q &\leq p_{C_{11}} \leq Q.
 \end{aligned}$$

After multiplying each inequality with its corresponding weight  $\alpha$  and adding all inequalities, we get that  $p$  is bounded below by

$$\begin{aligned}
 \alpha_1 p + \alpha_2 + \alpha_4 p q + \alpha_5 (1 - p) q + \alpha_6 [p + (1 - p) q] + \alpha_7 p^2 q + \alpha_8 (1 - p)^2 q + \\
 \alpha_9 [p^2 + (1 - p)^2] q + \alpha_{10} [3p(1 - p) + (1 - p)^2 q] + \alpha_{11} q
 \end{aligned}$$

and above by

$$\begin{aligned}
 \alpha_1 p + \alpha_2 + \alpha_4 p Q + \alpha_5 (1 - p) Q + \alpha_6 [p + (1 - p) Q] + \alpha_7 p^2 Q + \alpha_8 (1 - p)^2 Q + \\
 \alpha_9 [p^2 + (1 - p)^2] Q + \alpha_{10} [3p(1 - p) + (1 - p)^2 Q] + \alpha_{11} Q.
 \end{aligned}$$

Since  $0 \leq q \leq Q \leq 1$ , if we let  $q = 0$  and  $Q = 1$  in the above inequality, and set all weights equal to  $\alpha$ , we obtain the double inequality

$$\alpha(1 + 5p - 3p^2) \leq p \leq \alpha(7 - 2p + 2p^2).$$

From here, we get that

$$(18) \quad -1 + \frac{2}{3}\sqrt{3} \leq p \leq \frac{13}{4} - \frac{\sqrt{113}}{4} \Rightarrow 0.1547 \leq p \leq 0.5925.$$

Although this result is for the particular case of equal weights, it confirms the previous observations that the (stable) fixed points are situated approximately in the interval  $[0, 1/2]$ . Thus, in the long run, at most half of the nodes are active at any given time point. Of course, if  $q > 0$  and  $Q < 1$  then this interval becomes narrower. On the other hand, if  $q = Q$ , that is the sequence  $q_m$  converges, then the inequalities become equalities and we obtain that the fixed point is given by

$$p = \frac{7q + 6 \pm \sqrt{-71q^2 + 136q + 48}}{2(5q - 3)}.$$

Of course, if we let  $q = 0$  or  $1$  in this formula, we obtain the previous numerical values of (18).

## 7. CONCLUSIONS AND FURTHER DIRECTIONS OF INVESTIGATION

In this paper we provide a mean-field Boolean network model for a signal transduction network of a fibroblast cell, using various categories of Boolean functions, including totalistic rules and canalizing functions. The MF-model is shown to be a good fit for the average dynamical behavior of the actual network for a combination of parameters that matches those of the fibroblast system. Using the MF-model it is shown that a simple change in connectivity or in the weight of each class of Boolean functions considered in the MF-model does not lead to a significant change in the overall behavior of the system. Basically the system shows stability. It is also shown that this MF-model is a good approximation for a network in which protein mutations take place. Analytical results support the numerical observations that the (stable) fixed points are mostly located in the interval  $[0, 1/2]$ . This kind of insight is hard to obtain by considering the source network model [1] for the MF-model in this paper, due to its inherent structural constraints. As opposed to [1], with the current MF-model one can alter the connectivity or the presence of any type of Boolean function (protein) in the network and understand the impact on the long term dynamics, thus transcending the restrictions on both the topology and the dynamical rules of the fibroblast network. This way the MF-model can be used as a starting point for exploring other types of cells with a core similar to that of the fibroblast network, and to test biological questions that may not be feasible to carry out in the laboratory. It also has the potential to identify types of nodes that exhibit increased sensitivity to perturbations and that could be identified as potential targets in drug therapies, or to clarify the role of nodes whose role in the functionality of a biological network may still be uncertain. At the same time, the MF-model can be used to identify types of nodes, parameters, and mutations that yield a desired dynamical regime.

Future work will consist of an exhaustive analysis of possible dynamical behaviors under parameter variations and different types of perturbations of the network. Special consideration will be given to identifying the impact of other types of rules, currently incorporated into class  $C_{11}$  of the MF-model. Various types of mutations of the internal nodes of the network, or alteration of the external inputs will be considered. The robustness to such disturbances will be quantified, and the level of impact of each individual type of rule on the other rules will be assessed. Refinements of the node classes will be performed to incorporate nested canalizing functions or other types of functions that have not yet been identified. Moreover, the synchronization of networks built using the MF-model will be under consideration. We will also consider a more relaxed version in which the nodes are

allowed to obey more than one rule, thus extending the analysis to probabilistic Boolean networks [4].

Although the density of ones is the focus of this paper, it is only one numerical measure associated to the network, and the main goal of this work is to understand what is the impact of modifications of connectivity levels, weights of certain types of nodes in the overall structure of the network, or biases of Boolean rules on the activity level of the network. Thus we can assess the combinations of types nodes whose modification can generate the most significant disruption of the activity level of the network. A comprehensive account of these combinations is yet to be performed. On the other hand, the density of ones would not be able to distinguish between combinations of individual node activities that yield similar overall network activity. The density of ones has to be used further in conjunction with other measures to assess the importance of the individual types of nodes on the overall network dynamics in order to identify potential individual targets for therapies. A couple of numerical measures to be used in the future are: approximate entropy for quantifying the regularity of system or node evolutions [15], and average influences of certain nodes or node types on the dynamics of other nodes or the entire network which is a measure of sensitivity of the network to small disturbances [3]. These can shed further light on what types of nodes are more likely to bring the system to a desired regime.

Networks such as the fibroblast signal transduction network tend to be extremely large. However, only a small fraction of the network is actually known in detail. In order for biologists to be able to understand the entire functionality of the network, all the nodes and the regulatory mechanisms would have to be known. This task may be accomplished sometime in the future. However, it will require significant research efforts, together with a global system that can integrate the individual research results, paired with vast data processing tools. Meanwhile, we could use the MF-model in this paper to grow a larger network that incorporates even further types of nodes and regulatory mechanisms. For example, one could consider a modular model in which we consider multiple copies of the same network and link the copies in a deterministic or random fashion. Moreover, one could induce noise in the network by varying the number of active modules. Thus the network becomes asynchronous and provides a more realistic view of the dynamics. Alternative ways of growing the network will be under consideration.

## 8. ACKNOWLEDGEMENTS

We would like to thank J. Rogers and T. Helikar for discussions on the fibroblast network and access to the fibroblast model and the software used in [1].

## REFERENCES

- [1] Helikar T., Konvalina J., Heidel J., Rogers J.A., *Emergent decision-making in biological signal transduction networks*, PNAS, 105(6), 2008, p. 1913-1918.
- [2] Kauffman S.A., *The origins of order*, Oxford University Press, 1993, p. 173-235.
- [3] Shmulevich I., Kauffman S.A., *Activities and sensitivities in Boolean network models*, Physical Review Letters, 93(4), 2004, 048701.
- [4] Shmulevich I., Dougherty E.R., Zhang W., *From Boolean to probabilistic Boolean networks as models for genetic regulatory networks*, Proceedings of the IEEE, 90(11), 2002, p. 1778-1792.
- [5] Klemm K., Bornholdt S., *Stable and unstable attractors in Boolean networks*, Phys. Rev. E 72, 2000, 055101.
- [6] Raeymaekers L., *Dynamics of Boolean networks controlled by biologically meaningful functions*, Journal of Theoretical Biology, 218, 2002, p. 331-341.
- [7] Huepe C., Aldana-González M., *Dynamical phase transition in a neural network model with noise: an exact solution*, Journal of Statistical Physics, 108(3-4), 2002, p. 527-540.
- [8] Beck G.L., Matache M.T., *Dynamical behavior and influence of stochastic noise on certain generalized Boolean networks*, Physica A, 387(19-20), 2008, p. 4947-4958.
- [9] Kauffman S., Peterson C., Samuelsson B., Troein C., *Genetic networks with canalizing Boolean rules are always stable*, PNAS, 101(49), 2004, p. 17102-17107.
- [10] Rämö P., Kesseli J., Yli-Harja O., *Stability of functions in Boolean models of gene regulatory networks*, Chaos, 15(3), 2005, 34101.
- [11] Nikolajewa S., Friedel M., Wilhelm T., *Boolean networks with biologically relevant rules show ordered behavior*, Biosystems, 90(1), 2007, p. 40-47.
- [12] Drossel B., *Random Boolean Networks*, Review article, in Reviews of Nonlinear Dynamics and Complexity, Vol. 1, Ed. HG Schuster, Wiley, 2008.
- [13] Lodish H., Berk A., Zipursky S.L., Matsudaira P., Baltimore D., Darnell J., *Molecular Cell Biology*, New York: W. H. Freeman, 2000.
- [14] Schmidt M.H., Dikic I., *The Cbl interactome and its functions*, Nat. Rev. Mol. Cell. Biol., 6(12), 2005, p. 907-18.
- [15] Pincus S.M., *Approximate entropy as a measure of system complexity*, PNAS(88), 1991, p. 2297-2301.