# Reconstruction of Boolean regulatory models of flower development exploiting an evolution strategy

Gonzalo A. Ruz [*] Eric Goles [†] Sylvain Sené [‡]

**Abstract**

One of the first popular applications of Boolean networks for gene regulatory networks corresponds to the Mendoza & Alvarez-Buylla network of flower development. In this paper, we consider this model and a reduced version to reconstruct synthetic threshold Boolean networks that have the same asymptotic behavior as these base models. For this, we employ an evolution strategy to search for neighboring solutions. We were able to find solutions with fewer edges as well as networks with more balanced distributions of basins of attractions. Overall, our results show the effectiveness of using evolutionary computation in this application to explore alternative solutions with desired properties.

## 1 Introduction

Boolean networks are relatively high-level models for gene regulatory networks (GRNs). They were introduced by Stuart Kauffman [3] in the late '60s. Under this model, nodes can be either switched on (value 1) or off (value 0), and the edges represent direct relations between nodes. The dynamics of the network (how the node values change through time) is given by a set of updating rules (Boolean functions, one for each node) and an updating scheme. Typically, the synchronous or parallel updating scheme is used due to its simplicity: all the nodes are updated, in each time step, at the same time. Although, there are many deterministic and non-deterministic updating schemes (like the fully asynchronous: in each time step a randomly selected node is updated) that can be used. For more details on different updating schemes, please refer to [1].

Since Boolean networks only consider two states for each node, then for a network with $n$ nodes there are $2^n$ possible configurations from where the network can start from. Given the deterministic nature of this model, all the possible configurations will end up eventually after successive updates in two types of steady states, known as attractors. One of them is called a *fixed point* which is an invariant state that remains fixed regardless of the updating scheme used. The other type of attractors are known as *limit cycles*, these are a set of states that are revisited with a certain periodicity. In the context of GRN, the attractors (most commonly the fixed points) are associated with different cell types. We can define the *basin of attraction*, which consists in the set of states that converge to a respective attractor. Usually one refers to the size of the basin of attraction, of an attractor, as the number of states that converge to that particular attractor.

[*] Facultad de Ingeniería y Ciencias, Universidad Adolfo Ibáñez, Santiago, Chile; Center of Applied Ecology and Sustainability (CAPES), Santiago, Chile; gonzalo.ruz@uai.cl

[†] Facultad de Ingeniería y Ciencias, Universidad Adolfo Ibáñez, Santiago, Chile; eric.chacc@uai.cl

[‡] Université d'Aix-Marseille, Université de Toulon, CNRS, LIS, Marseille, France; sylvain.sene@lis-lab.fr

Well known techniques to reconstruct Boolean network models from binary gene expression data include REVEAL [5] and the Best-Fit extension algorithm [4]. Also, metaheuristics have been considered such as simulated annealing [8], swarm intelligence [10, 13, 9], genetic algorithms [14], differential evolution [7], and evolution strategy [11, 12, 15], amongst others.

One of the first biological applications of Boolean networks that caught the attention of GRN modelers was the work by Mendoza & Alvarez-Buylla [6] which proposed a Boolean network model to represent the dynamical behavior of the flower development process in *Arabidopsis thaliana* plants. Their model belongs to a particular class of Boolean networks called threshold Boolean networks. In this Boolean model, the edges have weights to represent the strength of the relation (positive or negative) between nodes. Also, each node has a threshold value. Then, nodes update their values using a Boolean Heaviside function that depends linearly on its inputs.

The Mendoza & Alvarez-Buylla network, if updated in parallel, has thirteen attractors, composed of seven limit cycles of length two each, with no biological meaning, i.e., they do not represent any cell types. The remaining six attractors are fixed points, each one representing different cell types, in particular, four of them represent tissues of the flower: sepals, petals, carpels and stamens. Whereas for the remaining two fixed points, one represents inflorescence meristematic cells, and the other, unobserved cells in nature (but could be potentially experimentally induced) called mutant in [6]. For other updating schemes, the limit cycles in this model, tend to disappear, and only remain the six fixed points (fixed points are invariant with respect to changes in the updating scheme).

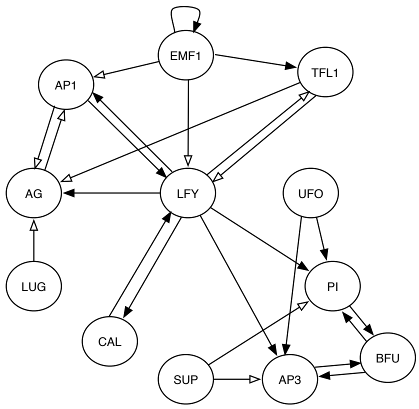
In [2] a reduced version (fewer edges) of the Mendoza & Alvarez-Buylla network was constructed, that preserves the same asymptotic behavior (attractors) of the original network. Nevertheless, it is not clear if this reduced network is a minimal network, i.e., no more edges can be removed without affecting the asymptotic behavior.

In this paper, we propose a method to reconstruct synthetic gene regulatory networks of flower development in *Arabidopsis thaliana* under the threshold Boolean network formalism, using as a starting point the Mendoza & Alvarez-Buylla network and the reduced model, employing an evolutionary strategy to find the different network parameters (weight matrix and threshold vector) that yield synthetic networks with the same attractors as the original model. We will analyze topological (wiring) and dynamical characteristics of the resulting networks. Also, we are interested to see if networks with fewer edges than the reduced model can be found. Overall, the possibility to explore neighboring solutions around the original and reduced model will allow us to shed light on how robust, in the sense of the network structure, are these models.

## 2 Background

### 2.1 Original Mendoza & Alvarez-Buylla network

In [6], Mendoza & Alvarez-Buylla proposed a threshold Boolean network that captured the dynamics of the floral development in *Arabidopsis thaliana*. The model consists of 12 interacting chemical species, designated by EMF1, TFL1, LFY, AP1, CAL, LUG, UFO, BFU, AG, AP3, PI, and SUP. The BFU species, is a dimer of the AP3 and PI proteins, all the rest are proteins as well. So in this model we have $n = 12$, and each node $x_i$ from



$$W = \left(\begin{array}{lcccccccccccc} & EMF1 & TFL1 & LFY & AP1 & CAL & LUG & UFO & BFU & AG & AP3 & PI & SUP \\ EMF1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ TFL1 & 1 & 0 & -2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ LFY & -2 & -1 & 0 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ AP1 & -1 & 0 & 5 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ CAL & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ LUG & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ UFO & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ BFU & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ AG & 0 & -2 & 1 & -2 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ AP3 & 0 & 0 & 3 & 0 & 0 & 0 & 2 & 1 & 0 & 0 & 0 & -2 \\ PI & 0 & 0 & 4 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & -1 \\ SUP & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right) \quad \Theta = \begin{pmatrix} 0 \\ 0 \\ 3 \\ -1 \\ 1 \\ 0 \\ 0 \\ 1 \\ -1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Figure 1: Original Mendoza & Alvarez-Buylla network. Activations (resp. repressions) are represented by full arrows (resp. empty arrows). Below, the matrix $W$ of size $12 \times 12$ contains the interaction weights between genes and $\Theta$ is the thresholds vector.

$i = 1, \ldots, n$ will update its value using the following rule:

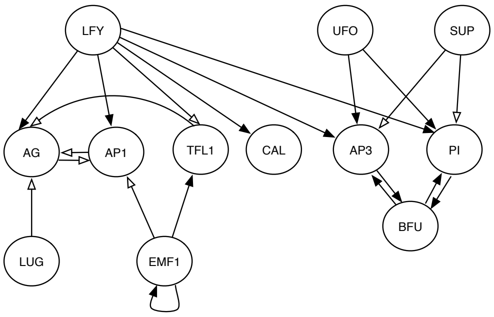
$$x_i(t+1) = H\left(\sum_{j=1}^{n} w_{ij} x_j(t) - \theta_i\right) \tag{1}$$

$$H(z) = \begin{cases} 1, & \text{if } z > 0 \\ 0, & \text{if } z \leqslant 0 \end{cases}$$

with $w_{ij}$ the weight of the edge coming from node $j$ into the node $i$, and $\theta_i$ the activation threshold of node $i$. The weights and thresholds are the network's parameters (see Fig. 1).

If we start in any of the $2^{12} = 4096$ possible configurations, and use the parallel updating scheme, then the network will converge to one of the possible thirteen attractors. As mentioned in the introduction, seven of these, are limit cycles of length two each, which have no biological meaning. The remaining six are the following fixed points: 1) {000100000000}, 2) {000100010110}, 3) {000000001000}, 4) {000000011110}, 5) {110000000000}, 6) {110000010110}. Each one has associated a cell type: 1) sepal, 2) petal, 3) carpel, 4) stamen, 5) inflorescence, 6) mutant (unobserved cell). In [6], a block-sequential updating scheme is proposed. A sequential updating scheme consists, in every time step, each node is updated following a predefined order. In the case of block-sequential, the set of nodes, for a given sequence, is partitioned into blocks. The nodes in a same block are updated in parallel, but blocks follow each other sequentially. In this case, the proposed block-sequential updating scheme is as follows: (EMF1, TFL1)(LFY, AP1, CAL)(LUG, UFO, BFU)(AG, AP3, PI)(SUP). With this updating scheme, the seven limit cycles no longer exist, and all the configurations converge to one of the six fixed points.

For simplicity we will refer to this network from now on as the original network.



$$W = \left(\begin{array}{ccccccccccccc} & EMF1 & TFL1 & LFY & AP1 & CAL & LUG & UFO & BFU & AG & AP3 & PI & SUP \\ EMF1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ TFL1 & 1 & 0 & -2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ LFY & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ AP1 & -2 & 0 & 5 & 0 & 0 & 0 & 0 & 0 & -2 & 0 & 0 & 0 \\ CAL & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ LUG & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ UFO & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ BFU & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ AG & 0 & -2 & 1 & -2 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ AP3 & 0 & 0 & 3 & 0 & 0 & 0 & 2 & 1 & 0 & 0 & 0 & -2 \\ PI & 0 & 0 & 4 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & -1 \\ SUP & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right) \Theta = \begin{pmatrix} 0 \\ 0 \\ 0 \\ -2 \\ 1 \\ 0 \\ 0 \\ 1 \\ -1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Figure 2: Reduced Mendoza & Alvarez-Buylla network.

### 2.2 Reduced Mendoza & Alvarez-Buylla network

In [6], a reduced version is generated that exhibits the same asymptotic behavior (attractors) as the original model. This version has 21 edges as shown in Fig. 2, instead of the 25 edges in the original network.

From now on we will refer to this network as the reduced network.

## 3 Methods

To reconstruct synthetic networks starting from the two *A. thaliana* networks described previously we will use an evolution strategy (ES) developed in [11] and used recently in [15]. A flow chart of the ES is shown in Fig. 3, where it can be seen that the main variation operator is mutation. The initial candidate solutions (networks) are generated using as a seed one of the two *A. thaliana* networks, depending on the simulation. Edges are removed or added from the original (reduced) network *ngh* times (a user defined parameter) to generate a candidate solution, as well as the respective threshold vector is changed. The fitness function is the mean squared error between the dynamics (output) of the candidate network and the dynamics of the original (reduced) network, given the same input. Therefore, it is a minimization problem, where we want to find networks with the least error. After the fitness value is computed for each candidate solution, these are ranked in a descending order. Then, the top $m\%$ are selected (another user defined parameter) to perform mutation, in a similar way as the candidate networks were generated, but now using the top ranked networks as seeds to generate new solutions. These new solutions plus the top $m\%$ are completed with random candidate networks, using as seed the original (reduced) network, to generate the new population. More details of the algorithm can be found in [12].

For all the simulations described ahead, the following parameters were used. The *ngh* parameter is selected randomly between 1 and 30 for each candidate network. The elements of the weight matrices and the threshold vectors were constrained to the following integer range $[-5, -4, \ldots, 4, 5]$. Also, $popSize = 20$, $m\% = 30\%$ and max iterations =100.

In what follows, we describe the different simulations to be conducted.

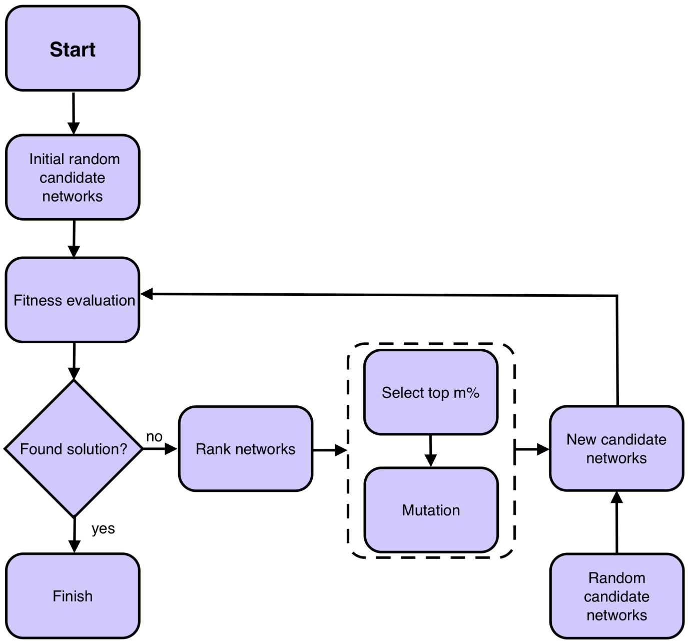


Figure 3: Evolution strategy (ES) flow chart to search for synthetic networks.

### 3.1 Simulation 1

Using the ES and the parameters defined previously, we will proceed to infer 1000 synthetic networks that contain only the six fixed points of the original network. It is important to point out that when the proposed fitness function reaches 0, we can assure that the candidate solution will in fact have the six fixed points, but we cannot assure that these will be the only ones existing, since there might be additional fixed points present. In order to avoid this situation, whenever a candidate solution obtains a fitness value of 0, we check how many fixed points the networks has. If the network has more than six, than we penalize the solution by adding 0.1 to the fitness value. By this way, we can assure that, when the fitness value is 0, the network will only have the desired six fixed points and no others. We use the original network as the seed to generate candidate solutions. For the resulting networks, we will compute the distribution of the total, positive, and negative number of edges.

### 3.2 Simulation 2

The same as Simulation 1, but now using the reduced network as the seed.

### 3.3 Simulation 3

The same as Simulation 1, but now we will reconstruct synthetic networks that have the first four fixed points of the original Mendoza & Alvarez-Buylla network, that are associated to specific cell types of the flower: sepal, petal, carpel, and stamen.

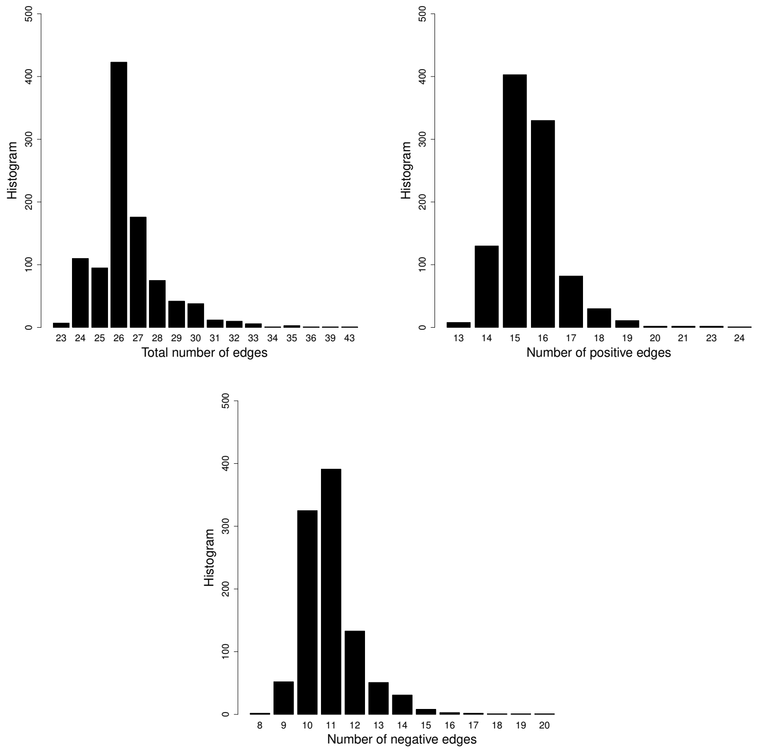


Figure 4: Edge distributions of the resulting synthetic networks found by the ES that contained only the six fixed points using the original network as the starting point.

### 3.4 Simulation 4

The same as Simulation 3, but now using the reduced Mendoza & Alvarez-Buylla network as the wildtype network.

## 4 Results and discussion

### 4.1 Results from simulation 1

Histograms showing the distributions of the total number of edges, number of positive edges, and number of negative edges, are shown in Fig. 4.

We can see that the resulting synthetic networks have a total number of edges that ranges from 23 to 43, with the mode at 26. The original network has 25 edges. The number of positive edges ranges from 13 to 24 (15 in the original network) with the mode at 15. Finally, the number of negative edges ranges from 8 to 20 (10 in the original model), with the mode at 11. It is interesting to point out, that we were not able to find networks with a total number of edges of 21 like the reduced network. This provides an insight on the fact that the reduced network developed in [2], is not found in an easy way, and it could be a minimal network.

### 4.2 Results from simulation 2

The distributions of the number of edges (total, positive, and negative) are shown in Fig. 5.

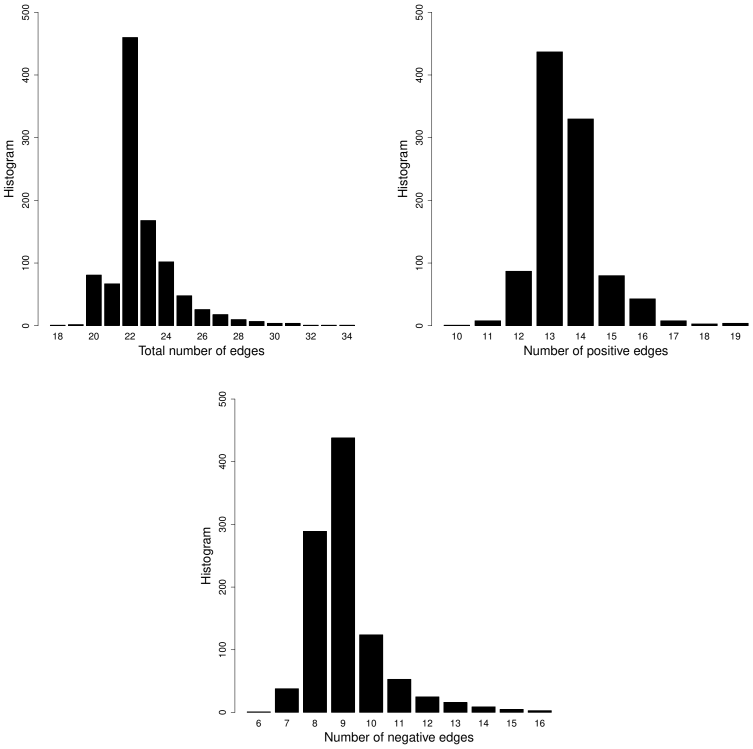


Figure 5: Edge distributions of the resulting synthetic networks found by the ES that contained only the six fixed points using the reduced network as the starting point.

Table 1: Basin of attraction for the six fixed points

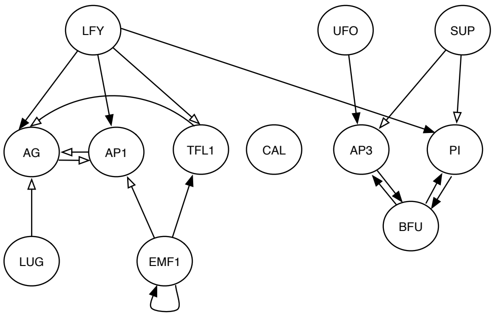
| Attractors | Parallel original | Parallel reduced | Parallel net18 | BS original | BS reduced | BS net18 | Cell types |
|---|---|---|---|---|---|---|---|
| Fixed point 1 | 168 | 168 | 540 | 1344 | 1344 | 1344 | Sep |
| Fixed point 2 | 248 | 248 | 124 | 192 | 192 | 192 | Pet |
| Fixed point 3 | 24 | 24 | 36 | 448 | 448 | 448 | Car |
| Fixed point 4 | 8 | 8 | 4 | 64 | 64 | 64 | Sta |
| Fixed point 5 | 384 | 384 | 960 | 1792 | 1792 | 1792 | Inf |
| Fixed point 6 | 384 | 384 | 192 | 256 | 256 | 256 | Mut |

What first caught our attention is that a network with 18 edges was found. This is 3 edges fewer than the reduced network. This means that the reduced network is not minimal, in the sense that we have found a synthetic network with the same asymptotic behavior, with fewer edges. Fig. 5 shows the resulting network with 18 edges. By comparing with the reduced network (Fig. 2) we noticed that the 3 edges that have been omitted are: the incoming edge from LFY to CAL, the incoming edge from LFY to AP3, and the incoming edge from UFO to PI.

Overall, the mode for the total number of edges is 22, the mode for positive edges is 13, and the mode for negative edges is 9.

Additionally, we compared the basin of attraction of the six fixed points for the original, reduced, and the network with 18 edges (named net18 from now on), using the parallel and the block-sequential (BS) updating scheme, described previously. The results are shown in Table 1.

We notice that for the parallel updating scheme, the original and the reduced network



$$W = \left(\begin{array}{ccccccccccccc} & EMF1 & TFL1 & LFY & AP1 & CAL & LUG & UFO & BFU & AG & AP3 & PI & SUP \\ EMF1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ TFL1 & 1 & 0 & -2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ LFY & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ AP1 & -2 & 0 & 5 & 0 & 0 & 0 & 0 & 0 & -2 & 0 & 0 & 0 \\ CAL & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ LUG & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ UFO & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ BFU & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ AG & 0 & -2 & 1 & -2 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ AP3 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 1 & 0 & 0 & 0 & -2 \\ PI & 0 & 0 & 4 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 \\ SUP & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right) \Theta = \begin{pmatrix} 0 \\ 0 \\ 0 \\ -2 \\ 1 \\ 0 \\ 4 \\ 1 \\ -1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Figure 6: Synthetic network found with 18 edges (net18).

have the same size of basin of attraction for each fixed points. For these two networks, almost 30% of the total 4096 states are part of the basin of attraction of the fixed points. The remaining 70% form part of the basin of attractions of the limit cycles. For net18 updated in parallel we can appreciate that size of the basin of attraction for each fixed point is different from the original (and reduced) network. Nevertheless, there are some similarities. In the three networks, fixed point 4 has the smallest basin of attraction, whereas fixed point 5 has the largest basin of attraction. In the case of net18, 45% of the configurations are part of the basin of attractions of the fixed points, the remaining converge to the limit cycles. The size of the basin of attraction plays an interesting and important role in GRN modeling, since attractors (cell types) with small basin of attractions, means that the network has little chances to converge to that attractor. On the other hand, large basin of attractions means that the network converges most of the time to the attractor with the largest basin of attraction associated to it. Here we notice that the original and reduced models do not perform well when we use the parallel updating scheme, net18 improves a bit, but still does not pass the 50% barrier. When we use the block-sequential updating scheme (more biologically meaningful) the limit cycles disappear, thus, obtaining 100% of the possible configurations converging to one of the possible 6 fixed points. An interesting fact is that the basin of attraction sizes are the same for each fixed points, for the three models. But we can evidence another problem, the largest basin of attraction is for fixed point 5 which is not a tissue of the flower. The next simulation explores a solution for this issue.

### 4.3 Results from simulation 3

By using the original network as the seed, the ES was capable of finding solutions that contained only the first four fixed points and no other. The resulting topology distributions are shown in Fig. 7.

An example of a resulting network with 23 edges (net23) in shown in Fig. 8.

The basin of attraction of each fixed point using the parallel and the block-sequential updating scheme appears in Table 2.

We observed that fixed point 1 and 2 had significantly larger basin of attractions than

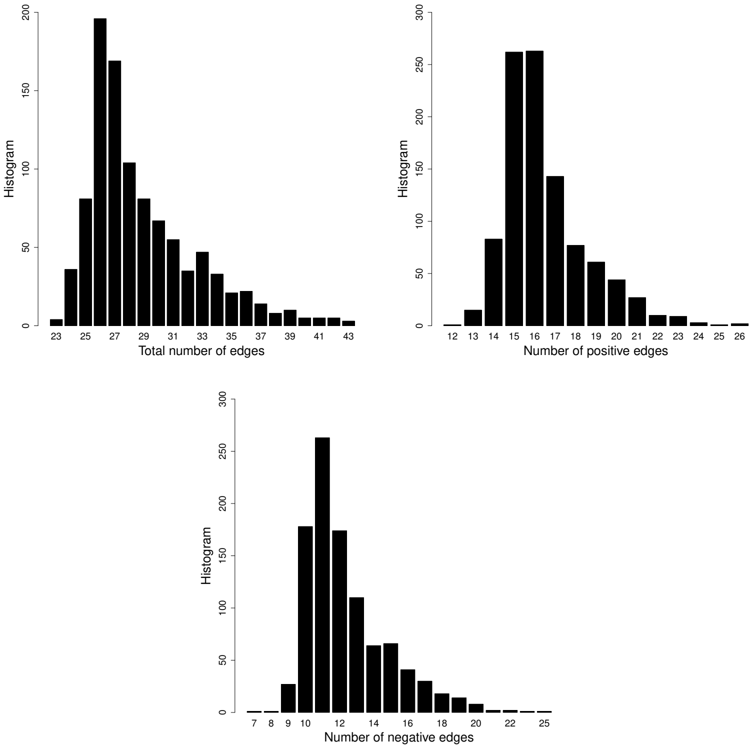
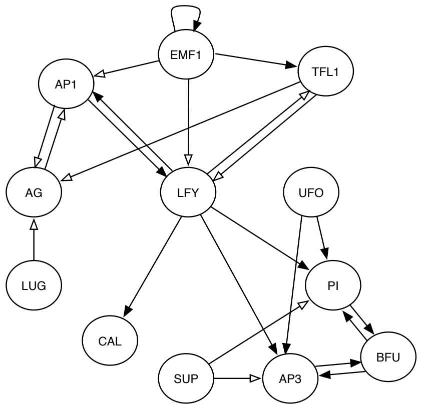


Figure 7: Edge distributions of the resulting synthetic networks found by the ES that contained only the first four fixed points using the original network as the starting point.

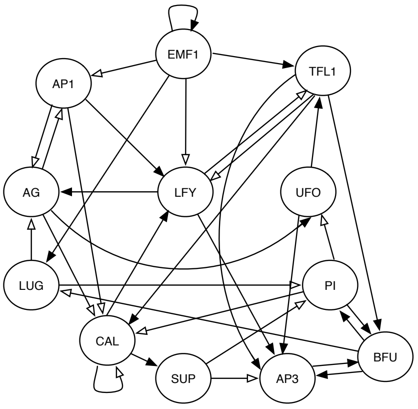


$$W = \left(\begin{array}{c|cccccccccccc} & EMF1 & TFL1 & LFY & AP1 & CAL & LUG & UFO & BFU & AG & AP3 & PI & SUP \\ EMF1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ TFL1 & 1 & 0 & -2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ LFY & -2 & -1 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ AP1 & -1 & 0 & 5 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ CAL & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ LUG & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ UFO & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ BFU & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ AG & 0 & -2 & 0 & -2 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ AP3 & 0 & 0 & 3 & 0 & 0 & 0 & 2 & 1 & 0 & 0 & 0 & -2 \\ PI & 0 & 0 & 4 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & -1 \\ SUP & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right) \Theta = \begin{pmatrix} 3 \\ 0 \\ 3 \\ -1 \\ 1 \\ 0 \\ 0 \\ 1 \\ -1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Figure 8: An example of a resulting network obtained with 23 edges (net23) that contains only the first four fixed points of the original network.

Table 2: Basin of attraction for the four fixed points using net23 and net34

| Attractors | Parallel net23 | BS net23 | BS net34 | Cell types |
|---|---|---|---|---|
| Fixed point 1 | 504 | 3136 | 1184 | Sep |
| Fixed point 2 | 616 | 448 | 864 | Pet |
| Fixed point 3 | 24 | 448 | 1184 | Car |
| Fixed point 4 | 8 | 64 | 864 | Sta |



$$W = \left(\begin{array}{ccccccccccccc}
 & EMF1 & TFL1 & LFY & AP1 & CAL & LUG & UFO & BFU & AG & AP3 & PI & SUP \\
EMF1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
TFL1 & 1 & 0 & -2 & 0 & 0 & 0 & 3 & 0 & 0 & 0 & 0 & 0 \\
LFY & -2 & -1 & 0 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
AP1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\
CAL & 0 & 1 & 0 & -3 & -4 & 0 & 0 & 0 & -3 & 0 & -3 & 0 \\
LUG & 3 & 0 & 0 & 0 & 0 & 0 & 0 & -5 & 0 & 0 & 0 & 0 \\
UFO & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & -4 & 0 \\
BFU & 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\
AG & 0 & 0 & 1 & -4 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
AP3 & 0 & 1 & 3 & 0 & 0 & 0 & 2 & 1 & 0 & 0 & 0 & -2 \\
PI & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & -1 \\
SUP & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}\right) \quad \Theta = \begin{pmatrix} 3 \\ 1 \\ 3 \\ -1 \\ 2 \\ 4 \\ 4 \\ 0 \\ -4 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

Figure 9: A resulting network obtained with 34 edges (net34) with more evenly distributed basin of attraction per fixed point.

the other two fixed points. When changing to the block-sequential updating scheme, this issue did not improve. The ideal case is that each cell type (fixed point) should have more or less the same chance of developing, this means that the basin of attraction should be approximately evenly distributed per fixed point. Within the 1000 networks we searched for a network that had the most evenly distributed basin of attraction per fixed points. The network that best satisfied this restriction is shown in Fig. 9 which has 34 edges (net34).

We see from Table 2 that basin of attractions for this network are more evenly spread in the four fixed points.

## 4.4 Results from simulation 4

If we consider the reduced network to search for synthetic networks with only the first four fixed points, then the distribution of the total, positive, and negative number of edges of the solutions found are shown in Fig. 10.

We see that in this case, we can find a network with 17 edges (net17) (see Fig. 11).

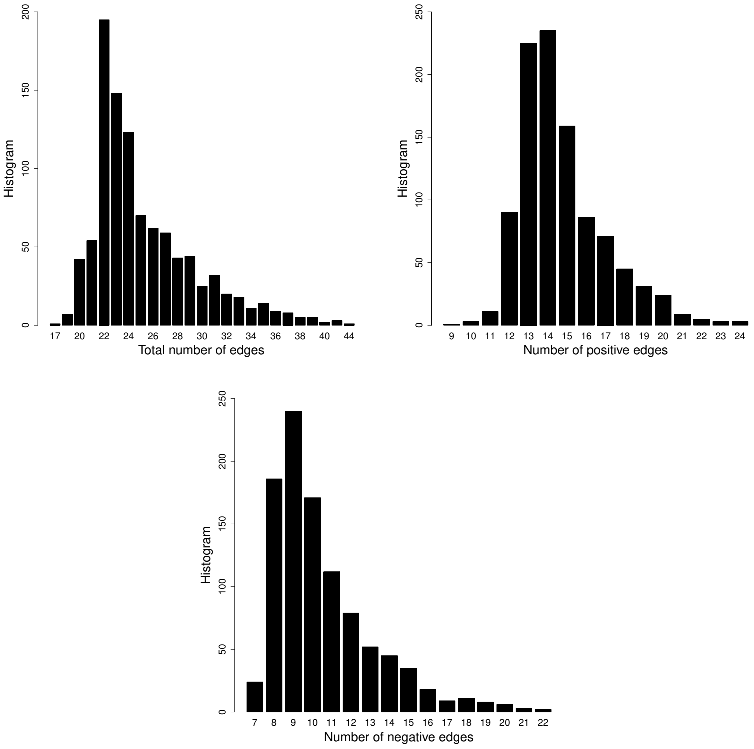
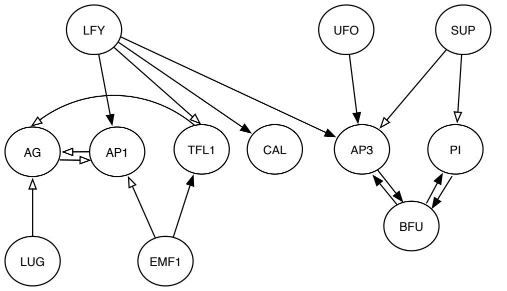


Figure 10: Edge distributions of the resulting synthetic networks found by the ES that contained only the first four fixed points using the reduced network as the starting point.

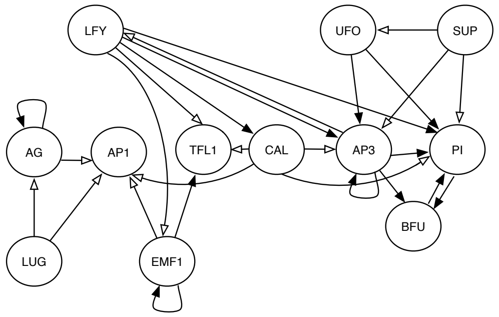


$$W = \left(\begin{array}{ccccccccccccc} & EMF1 & TFL1 & LFY & AP1 & CAL & LUG & UFO & BFU & AG & AP3 & PI & SUP \\ EMF1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ TFL1 & 1 & 0 & -2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ LFY & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ AP1 & -2 & 0 & 5 & 0 & 0 & 0 & 0 & 0 & -2 & 0 & 0 & 0 \\ CAL & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ LUG & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ UFO & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ BFU & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ AG & 0 & -2 & 0 & -2 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ AP3 & 0 & 0 & 3 & 0 & 0 & 0 & 2 & 1 & 0 & 0 & 0 & -2 \\ PI & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 \\ SUP & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right) \Theta = \begin{pmatrix} 0 \\ 5 \\ 0 \\ -2 \\ 1 \\ 0 \\ 0 \\ 1 \\ -1 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

Figure 11: An example of a resulting network obtained with 17 edges (net17) that contains only the first four fixed points of the reduced network.

Table 3: Basin of attraction for the four fixed points using net17 and net27

| Attractors | Parallel net17 | BS net17 | BS net27 | Cell types |
|---|---|---|---|---|
| Fixed point 1 | 1260 | 2688 | 1024 | Sep |
| Fixed point 2 | 140 | 384 | 1024 | Pet |
| Fixed point 3 | 108 | 896 | 1024 | Car |
| Fixed point 4 | 12 | 128 | 1024 | Sta |



$$W = \left(\begin{array}{c|cccccccccccc} & EMF1 & TFL1 & LFY & AP1 & CAL & LUG & UFO & BFU & AG & AP3 & PI & SUP \\ EMF1 & 1 & 0 & -4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ TFL1 & 1 & 0 & -2 & 0 & -3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ LFY & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ AP1 & -2 & 0 & 0 & 0 & -3 & -1 & 0 & 0 & -2 & 0 & 0 & 0 \\ CAL & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ LUG & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ UFO & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -5 \\ BFU & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ AG & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 5 & 0 & 0 & 0 \\ AP3 & 0 & 0 & 3 & 0 & -1 & 0 & 2 & 0 & 0 & 5 & 0 & -2 \\ PI & 0 & 0 & 4 & 0 & -2 & 0 & 1 & 1 & 0 & 3 & 0 & -1 \\ SUP & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right) \Theta = \begin{pmatrix} 1 \\ 0 \\ 4 \\ -1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 4 \\ 0 \\ 1 \\ 0 \end{pmatrix}$$

Figure 12: A resulting network obtained with 27 edges (net27) with evenly distributed basin of attraction per fixed point.

This network omits the following edges from the reduced network: the incoming edge from UFO to PI, the incoming edge from LFY to AG, the incoming edge from LFY to PI, and the loop of EMF1.

The basin of attraction per fixed point using the parallel and block-sequential updating scheme for net17 is shown in Table 3.

Here we noticed that fixed point 1 had the largest basin of attraction when using the parallel and the block-sequential. Similar to before, we searched within the 1000 solutions for a network that had the most evenly spread basin of attraction per fixed point. The resulting network that best satisfied this restriction is shown in Fig. 12 that has 27 edges (net27).

Table 3 shows that this network when updated using the block-sequential updating scheme, partitions the state space evenly in four, i.e., the size of the basin of attraction of each fixed point is 1024.

# 5 Conclusion

We have shown that an evolutionary computation approach can effectively reconstruct alternative solutions based on existing gene regulatory models under the threshold Boolean network paradigm. For the particular application presented in this work, we were able to find interesting solutions, such as networks with fewer edges than the existing models. Also, it was found that in order to change the distribution of the sizes of the basin of attractions, so that each fixed point had more or less the same basin of attraction size, this was achieved by increasing the complexity (number of edges) of the base models used. Future research will consider more recent gene regulatory models as base models such as the one developed in [16].

## Acknowledgment

This work was supported by ECOS-CONICYT C16E01 (GR, SS, EG), FONDECYT 1180706 (GR), FONDECYT 1140090 (EG), Basal Project CMM (GR, EG), "Investissement d'avenir" program ANR-16-CONV-00001 (SS) and PACA Fri Project 2015 01134 (SS).

## References

[1] J. Aracena, E. Goles, A. Moreira, and L. Salinas. On the robustness of update schedules in Boolean networks. *Biosystems*, 97:1–8, 2009.

[2] Jacques Demongeot, Eric Goles, Michel Morvan, Mathilde Noual, and Sylvain Sené. Attraction basins as gauges of robustness against boundary conditions in biological complex systems. *PLOS ONE*, 5(8):1–18, 08 2010.

[3] S. A. Kauffman. Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of Theoretical Biology*, 22:437–467, 1969.

[4] H. Lahdesmaki, I. Shmulevich, and O. Yli-Harja. On learning gene regulatory networks under the boolean network model. *Machine Learning*, 25:147–167, 2003.

[5] S. Liang, S. Fuhrman, and R. Somogyi. Reveal, a general reverse engineering algorithm for inference of genetic network architectures. In *Pac. Symp. Biocomput*, pages 18–29, 1998.

[6] L. Mendoza and E. R. Alvarez-Buylla. Dynamics of the genetic regulatory network for arabidopsis thaliana flower morphogenesis. *Journal of Theoretical Biology*, 193:307–319, 1998.

[7] G. A. Ruz, D. Ashlock, T. Ledger, and E. Goles. Inferring bistable lac operon Boolean regulatory networks using evolutionary computation. In *The 2017 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB 2017)*, pages 1–8, 2017.

[8] G. A. Ruz and E. Goles. Learning gene regulatory networks with predefined attractors for sequential updating schemes using simulated annealing. In *Proc. of IEEE the Ninth International Conference on Machine Learning and Applications (ICMLA 2010)*, pages 889–894, 2010.

[9] G. A. Ruz and E. Goles. Reconstruction and update robustness of the mammalian cell cycle network. In *2012 IEEE Symposium on Computational Intelligence and Computational Biology, CIBCB 2012*, pages 397–403, 2012.

[10] G. A. Ruz and E. Goles. Learning gene regulatory networks using the bees algorithm. *Neural Computing and Applications*, 22:63–70, 2013.

[11] G. A. Ruz and E. Goles. Neutral graph of regulatory Boolean networks using evolutionary computation. In *The 2014 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB 2014)*, pages 1–8, 2014.

[12] G. A. Ruz, T. Timmermann, J. Barrera, and E. Goles. Neutral space analysis for a boolean network model of the fission yeast cell cycle network. *Biological Research*, 47:64, 2014.

[13] G. A. Ruz, T. Timmermann, and E. Goles. Building synthetic networks of the budding yeast cell-cycle using swarm intelligence. In *Proceedings - 2012 11th International Conference on Machine Learning and Applications, ICMLA 2012*, volume 1, pages 120–125, 2012.

[14] G. A. Ruz, T. Timmermann, and E. Goles. Reconstruction of a GRN model of salt stress response in Arabidopsis using genetic algorithms. In *The 2015 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB 2015)*, pages 1–8, 2015.

[15] G. A. Ruz, T. Timmermann, and E. Goles. Neutral space analysis of gene regulatory network models of salt stress response in arabidopsis using evolutionary computation. In *The 2016 IEEE Congress on Evolutionary Computation (IEEE CEC 2016)*, pages 4281–4288, 2016.

[16] Yara-Elena Sánchez-Corrales, Elena R. Álvarez-Buylla, and Luis Mendoza. The arabidopsis thaliana flower organ specification gene regulatory network determines a robust differentiation process. *Journal of Theoretical Biology*, 264(3):971–983, 2010.