Griffin: A Tool for Symbolic Inference of Synchronous Boolean Molecular Networks
Stalin Munoz, Miguel Carrillo, Eugenio Azpeitia, David A. Rosenblueth

To cite this version:
Stalin Munoz, Miguel Carrillo, Eugenio Azpeitia, David A. Rosenblueth. Griffin: A Tool for Symbolic Inference of Synchronous Boolean Molecular Networks. Frontiers in Genetics, Frontiers, 2018, 9, 10.3389/fgene.2018.00039. hal-02621445

HAL Id: hal-02621445
https://hal.inrae.fr/hal-02621445
Submitted on 26 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License
Griffin: A Tool for Symbolic Inference of Synchronous Boolean Molecular Networks

Stalin Muñoz1,2,3, Miguel Carrillo1, Eugenio Azpeitia4,5 and David A. Rosenblueth1,6*

1Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, Universidad Nacional Autónoma de México, Mexico City, Mexico, 2Facultad de Ingeniería, Universidad Nacional Autónoma de México, Mexico City, Mexico, 3Maestría en Ciencias de la Complejidad, Universidad Autónoma de la Ciudad de México, Mexico City, Mexico, 4Institut National de Recherche en Informatique et en Automatique Project-Team Virtual Plants, Inria, CIRAD, INRA, Montpellier, France, 5Department of Evolutionary Biology and Environmental Studies, University of Zurich, Zurich, Switzerland, 6Centro de Ciencias de la Complejidad, Universidad Nacional Autónoma de Mexico, Mexico City, Mexico

Boolean networks are important models of biochemical systems, located at the high end of the abstraction spectrum. A number of Boolean gene networks have been inferred following essentially the same method. Such a method first considers experimental data for a typically underdetermined “regulation” graph. Next, Boolean networks are inferred by using biological constraints to narrow the search space, such as a desired set of (fixed-point or cyclic) attractors. We describe Griffin, a computer tool enhancing this method. Griffin incorporates a number of well-established algorithms, such as Dubrova and Teslenko’s algorithm for finding attractors in synchronous Boolean networks. In addition, a formal definition of regulation allows Griffin to employ “symbolic” techniques, able to represent both large sets of network states and Boolean constraints. We observe that when the set of attractors is required to be an exact set, prohibiting additional attractors, a naive Boolean coding of this constraint may be unfeasible. Such cases may be intractable even with symbolic methods, as the number of Boolean constraints may be astronomically large. To overcome this problem, we employ an Artificial Intelligence technique known as “clause learning” considerably increasing Griffin’s scalability. Without clause learning only toy examples prohibiting additional attractors are solvable: only one out of seven queries reported here is answered. With clause learning, by contrast, all seven queries are answered. We illustrate Griffin with three case studies drawn from the Arabidopsis thaliana literature. Griffin is available at: http://turing.iimas.unam.mx/griffin.

Keywords: molecular networks, Boolean networks, model inference, Boolean satisfiability problem, clause learning, biological constraints, attractors

1. INTRODUCTION

Synchronous Boolean networks (Kauffman, 1969), as simple as they are, can encode meaningful biological information (Huang, 1999; Bornholdt, 2001, 2005, 2008; Fisher and Henzinger, 2007). Hence, such models have emerged as valuable candidates for representing dynamics of molecular networks. At the same time, the inference of network dynamics from experimental data (sometimes called the “inverse problem”) has become increasingly relevant with the advent of high-throughput technologies. Because of their simplicity, synchronous Boolean networks could become excellent...
network-inference models. Nevertheless, the problem of synchronous Boolean network inference is neglected, as efforts in Boolean-network research have mainly been devoted to analysis. Our objective is to present and illustrate the practical use of Griffin, a computer tool for the inference of synchronous Boolean networks.

1.1. Inference of Boolean Networks

The inference of Boolean networks from biological data is at present coming of age. On the one hand, a methodology that has been employed to infer a number of Boolean gene regulatory networks has established itself. This methodology follows the next three steps. The first one is determining the set of genes regulating each gene, in addition to the “sign” of each regulation based on experimental results. Second, a number of constraints are considered. These constraints can be, for example, (a) a set of input-output pairs representing time-series data (Liang et al., 1998; Akutsu et al., 1999; Lähte äsmäki et al., 2003; Chueh and Lu, 2012; Haider and Pal, 2012; Berestovsky and Nakleh, 2013; Han et al., 2014; Barman and Kwon, 2017), (b) a desired set of fixed-point attractors (Albert and Othmer, 2003; Li et al., 2004; Pal et al., 2005; Mendoza, 2006; Davidich and Bornholdt, 2008; Tarissan et al., 2008; La Rota et al., 2011; Azpeitia et al., 2013, 2014), or (c) a set of temporal-logic constraints (Bernot et al., 2004; Chabrier-Rivier et al., 2004; Fages et al., 2004; Calzone et al., 2006a,b; Mateus et al., 2007; Khalis et al., 2009; Richard et al., 2012). Third, search is performed to find Boolean networks consistent with the available information.

On the other hand, there are now many methods and algorithms for the analysis of Boolean networks that have demonstrated their effectiveness. We noticed that some of these algorithms can also be used as auxiliary techniques for the inference of network dynamics. In particular, we have selected Dubrova and Teslenko’s (2011) algorithm for finding cyclic attractors in synchronous Boolean networks, Garg et al.’s (2008) method for finding basins of attraction, and Naldi et al.’s (2007) formula for finding fixed-point attractors. (We use the terms “steady state dynamics” and “fixed-point attractors” interchangeably.) Griffin has combined these algorithms with novel techniques, offering a tool automating many steps of the inference of molecular network dynamics.

1.2. Synchronous Boolean Networks

Boolean networks have variables representing the genes, mRNA, proteins, or any other type of molecule included in the network. Variables have only two values, and time is discrete. A state is a tuple (i.e., ordered list) of values, one for each variable. A Boolean network consists of a (finite) set of states together with a “transition” relation (sometimes also called “update” or “successor” relation).

In asynchronous Boolean networks, states may have more than one successor. In synchronous Boolean networks, by contrast, states have exactly one successor. Previous studies suggest that having multiple successors can provide a closer description of the biological phenomena and can eliminate dynamical artifacts (Garg et al., 2008). There may be advantages, nevertheless, in considering synchronous networks.

1.3. Outline of Griffin’s Methodology

In a nutshell, Griffin’s methodology is the following. Griffin has as input a (generalization of a) “regulation” graph (sometimes also called “interaction” or “influence” graph), in addition to biological constraints, such as an expected set of either fixed-point or cyclic attractors of both the wild-type and of mutants of an organism. The output is typically a set of Boolean networks satisfying such constraints.

Because in Griffin the regulation graph plays a prominent role, the definition of a regulation is fundamental. Griffin uses the definition of Naldi et al. (2007), Richard et al. (2012), and Mori and Mochizuki (2017).
A regulation graph is normally an underdetermined specification of a Boolean network because even with the formal definition of regulation and sometimes also due to lack of information the regulation graph may be satisfied by many Boolean networks.

Griffin then combines the regulation graph with the biological constraints, to reduce the number of possible networks. This is essentially done by constructing a typically large Boolean formula representing: (1) a formal definition of the regulation graph and (2) the biological constraints. This formula is then given to a "SAT solver." These solvers employ algorithms finding value assignments to the variables occurring in the formula that make such a formula true. Each such assignment represents a possible network.

Notice that both the regulation graph and many of the biological constraints rely on biological information. Thus, Griffin mainly produces biological meaningful solutions (i.e., networks that are coherent with the available experimental information).

Griffin’s methodology required the development of multiple innovations, including the generalization of the concept of a regulation graph that we call R-graph, the gradual application of biological constraints at runtime using clause learning, and the handling of both mutations and partially known fixed-point attractors. Some of Griffin’s innovations will be described below.

1.4. Symbolic Techniques

Griffin relies heavily on “symbolic” methods. Such methods do not represent sets explicitly, but rather codified as Boolean expressions. This allows for the representation of large sets of states, as well as the pruning of large fragments of the search space. It is important to observe that numerous off-the-shelf symbolic computational tools exist.

1.5. Background of Griffin

The original prototype of Griffin appeared in Rosenblueth et al. (2014), where the authors presented a formal framework for the inference of Boolean networks from a standard regulation graph by direct application of a SAT solver to Boolean formulas codifying regulations, fixed-point attractors, and single-point mutations.

Examples illustrating the use of Griffin already appeared in: Rosenblueth et al. (2014), employing sets of desired fixed-points for mutations; Weinstein et al. (2015), looking for interactions that are necessary for the existence of a cyclic attractor; Garcia-Gómez et al. (2017), verifying if a set of regulations was enough to obtain a desired set of expected attractors; and Azpeitia et al. (2017), making extensive use of hypotheses and partially defined sets of attractors for multi-point mutations.

Features illustrated in previous articles mentioning Griffin have been left out from this work. We demonstrate in this article Griffin’s new functionalities through three cases studies in section 2.2.

1.6. Organization

The rest of this article is organized as follows. The section “Results” has two parts: first we give a description of Griffin, and then we turn our attention to three case studies. Next follows the section “Discussion,” relating our work with other approaches and summarizing our results. Section 4 gives formal definitions appearing in the pseudo-code of the algorithms. Finally, the Supplementary Material includes all query files for the case studies, and a detailed explanation of the syntax used to formulate partially defined fixed-point constraints.

2. RESULTS

This section first gives an account of Griffin, and then presents three case studies.

2.1. Griffin

2.1.1. R-Graphs

A first contribution of Griffin is a generalization of the ordinary regulation graph that we call "R-graph" (for “regulation”). In an R-graph, vertices are, as in an ordinary regulation graph, molecular species. However, instead of having edges denoting only two kinds of regulation (positive or negative), an R-graph has 16 kinds of labels for edges, allowing the user to express a range of nuances of what is known about a regulation. Table 1 shows these 16 possible labels. In addition to having the ordinary positive and negative regulations (first two lines), we have ambiguous regulations (third line). Ambiguous regulations are regulations that depending on the “context” are negative or positive. As in ordinary regulation graphs, it is also possible to indicate the fact that we know that there is no regulation from one molecule to another (fourth line). We, however, view labels of edges as Boolean functions over the set of positive and negative regulations. This allows us to represent additional

<table>
<thead>
<tr>
<th>Depiction</th>
<th>Meaning</th>
<th>Acronym</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ ]</td>
<td>None</td>
<td>NR</td>
<td>¬R+ ∧ ¬R-</td>
</tr>
<tr>
<td>[ ]</td>
<td>Positive, unambiguous</td>
<td>MPU</td>
<td>R+ ∧ ¬R-</td>
</tr>
<tr>
<td>[ ]</td>
<td>Negative, unambiguous</td>
<td>MNU</td>
<td>¬R+ ∧ R-</td>
</tr>
<tr>
<td>[ ]</td>
<td>Manditory, ambiguous</td>
<td>MA</td>
<td>R+ ∨ ¬R-</td>
</tr>
<tr>
<td>[ ]</td>
<td>No regulation</td>
<td>NR</td>
<td>¬R+ ∧ ¬R-</td>
</tr>
<tr>
<td>[ ]</td>
<td>Positive, possibly ambiguous</td>
<td>MPPA</td>
<td>R+</td>
</tr>
<tr>
<td>[ ]</td>
<td>Negative, possibly ambiguous</td>
<td>MNPA</td>
<td>¬R-</td>
</tr>
<tr>
<td>[ ]</td>
<td>Mandatory, unknown sign, ambiguous</td>
<td>MUSPA</td>
<td>R+ ∨ ¬R-</td>
</tr>
<tr>
<td>[ ]</td>
<td>Optional, unambiguous</td>
<td>OPU</td>
<td>¬R-</td>
</tr>
<tr>
<td>[ ]</td>
<td>Optional, negative, unambiguous</td>
<td>ONU</td>
<td>¬R+</td>
</tr>
<tr>
<td>[ ]</td>
<td>Optional, ambiguous</td>
<td>OA</td>
<td>(R+ ↔ R-)</td>
</tr>
<tr>
<td>[ ]</td>
<td>Optional, possibly ambiguous</td>
<td>OPPA</td>
<td>R+ ∨ ¬R-</td>
</tr>
<tr>
<td>[ ]</td>
<td>Optional, negative, possibly ambiguous</td>
<td>ONPA</td>
<td>¬R+ ∨ R-</td>
</tr>
<tr>
<td>[ ]</td>
<td>Optional, unknown sign, unambiguous</td>
<td>OUSPA</td>
<td>¬(R+ ∧ R-)</td>
</tr>
<tr>
<td>[ ]</td>
<td>Contradiction</td>
<td>False</td>
<td>0</td>
</tr>
<tr>
<td>[ ]</td>
<td>Tautology</td>
<td>True</td>
<td>1</td>
</tr>
</tbody>
</table>

Each R-regulation is denoted by an acronym. Leaving out trivial regulations false, NR, and true, R-regulations can be classified as optional or mandatory. In the case of no regulation (NR), edges are not depicted. The last column shows the constraint added by Griffin in each case (R+) denotes a positive regulation from u to v for the Boolean network f (R+ (u,v), f ((u,v), v), R- a negative regulation (R- (u,v), v), see definition 3).
possibilities in a uniform manner. For example, we can represent the contingency in which we know that in certain contexts a regulation is positive, but we do not know whether or not there are other contexts where such a regulation is negative. This case would correspond to the fifth line (“Mandatory, positive, possibly ambiguous”).

Other interesting situations are the optional regulations. Such regulations can be used to test or predict putative regulations. Just as in mandatory regulations, the sign of optional regulations can vary. For instance, the first optional line (“Optional, positive, unambiguous”) represents a situation where we wish to allow for the possibility of no regulation, but in case there is one it should be positive and cannot be negative. In line 14 (last line before “Contradiction”), by contrast, we allow for the possibility of a positive regulation, a negative regulation, or no regulation.

The last column of Table 1 shows the corresponding Boolean function over positive and negative regulations ($R^+$ and $R^-$), defined in section 4.

We can thus represent any combination of necessity and possibility for positive and negative regulations. This generalization includes two trivial functions (Contradiction and Tautology) listed in the last two lines of Table 1.

### 2.1.2. Meaning of a Regulation

The next question we have to address is the meaning of a (positive or negative) “functional regulation” (i.e., $R^+$ and $R^-$). This meaning is important in the sense that it establishes a connection between regulation graphs (or R-graphs) and state graphs. We take the meaning proposed by Naldi et al. (2007) (and used also by Richard et al., 2012 and Mori and Mochizuki, 2017). According to this definition, a regulation is functional if it is sufficient to modify the activity of the regulated variable in a non-empty set of (molecular) contexts. A context is a specific combination of values of all other variables. Observe that it is the formal definition of regulation (Defn. 3) that allows us to define R-graphs. Some previous works (e.g., Espinosa-Soto et al., 2004) have inferred gene network dynamics manually, and without employing a formal definition of regulation. As a result, it is not clear in such inferences whether or not all possibilities of networks that are coherent with the experimental information have been explored.

### 2.1.3. Tractability of Boolean-Network Inference

Apart from an R-graph, an important input to Griffin can be a desired set of attractors (be they fixed-point or cyclic). Writing a Boolean formula representing a desired fixed-point, for instance, is straightforward (Rosenblueth et al., 2014). Hence, a set of such fixed-points can be represented as the conjunction (the “AND” function) of such formulas. The resulting formula, however, does not say anything about the unwanted fixed-points, and the solution networks might have additional fixed-points. Although such solutions, with additional fixed-points, might be acceptable, it may also be the case that we do not wish additional fixed-points. We must then explicitly express so in the Boolean formula (a “clause”), so as to block each unwanted fixed-point from solutions. This situation is analogous to the “frame problem” (Shanahan, 1997) in Artificial Intelligence. As a result, a direct approach to coding an exact set of fixed-point attractors (not allowing additional ones) produces a formula whose size is proportional to the number of states (i.e., exponential in the number of molecular species). Such an approach is not scalable. Moreover, if we are looking for solutions with no cyclic attractors, then coding this into a Boolean formula results in an intractable combinatorial explosion even for small networks.

### Algorithm 1: Griffin strategy for network iteration

**Input:** $Q$, a griffin query  
**Output:** writes all satisfying Boolean networks

```latex
\begin{algorithm}
\caption{Griffin strategy for network iteration}
\begin{algorithmic}[1]
\State $i \leftarrow 0$, $n \leftarrow \infty$
\If{$Q$.limit-Boolean-networks} $n \leftarrow Q$.limit
\State $\psi \leftarrow \text{equisatisfiable-CNF} (\text{encode}(Q))$
\State $\mathcal{E} \leftarrow \text{SAT-engine.instance}(\psi)$
\While{$\mathcal{E}$.satisfiable() and $i < n$}
\State $N \leftarrow \text{decode-network}(\mathcal{E}.\text{next-assignment}())$
\If{not $Q$.allow-additional-cycles}
\State $S \leftarrow \text{Dubrova-Teslenko}(N)$
\For{$A \in S$}
\If{$|A| > 1$}
\If{$A \notin \text{allowed-fixed-points}(Q)$}
\State block-attractor($\mathcal{E}, A$)
\EndIf
\Else
\If{$Q$.block-steady-a-posteriori}
\If{$A \notin \text{allowed-fixed-points}(Q)$}
\State block-attractor($\mathcal{E}, A$)
\EndIf
\EndIf
\EndIf
\EndFor
\EndIf
\If{$Q$.block-steady-a-posteriori}
\State $S \leftarrow Naldi-Thieffy-Chaouiya(N)$
\For{$A \in S$}
\If{$A \notin \text{allowed-fixed-points}(Q)$}
\State block-attractor($\mathcal{E}, A$)
\EndIf
\EndIf
\EndIf
\State write($N$)
\State $i \leftarrow i + 1$
\EndWhile
\State \textbf{Function} block-attractor($\text{engine, attractor}$):
\State $\phi \leftarrow \text{encode-attractor}(\text{attractor})$
\State $\text{engine.add-clauses} (\text{equisatisfiable-CNF}(\neg \phi))$
\State \textbf{return}
\EndFunction
\end{algorithmic}
\end{algorithm}
```
Armed with Dubrova and Teslenko’s (2011) method for detecting attractors in synchronous Boolean networks, Griffin does not block all unwanted attractors in the initial formula, but performs a “lazy” blocking, gradually adding the subformulas of unwanted attractors that appear in potential solutions. This technique, called “clause learning” (Franco and Martin, 2009), is used in SAT solvers, but at a lower level than Griffin does.

Table 2 shows how the gradual application of biological constraints at runtime is able to tackle all queries in our case studies. Griffin exhibits a significant increase in scalability as compared with the 2014 prototype of Griffin that did not have this feature.

2.1.4. Partially Known State Transitions
Griffin is able to represent partially known state transitions, in particular those corresponding to fixed-point attractors where the value of some species is unknown.

2.1.5. Whole Queries and Query Splitting
It is also possible to have certain user control in the search for Boolean networks through an operating mode called query splitting. In the query-splitting mode, Griffin can (1) first instantiate an R-graph into all possible ordinary regulation graphs, (2) next obtain answers for each such instance, and (3) finally combine all resulting answers. It may be useful to partition a problem this way, as the time required to solve all instances might be less than the time required for solving the original query. Figure 1 shows an example of query splitting using the “radial” exploration strategy (see Algorithm 2).

In the radial exploration strategy, an ordinary regulation graph resulting from removing all optional regulations of an R-graph, called a center, is obtained. Next, combinations of optional regulations are added to the obtained center. The number of optional regulations added to the center is called the radius.

2.1.6. Griffin’s Structure
Algorithm 1 shows the exploration strategy used by Griffin to find the set of satisfying solutions to a query. Figure 2 depicts the processing of a query by Griffin.

2.2. Case Studies
We demonstrate Griffin’s functionalities with three different Boolean network models taken from the literature. The first example allows us to illustrate queries for (a) finding Boolean networks given an R-graph, (b) finding the sets of fixed-point attractors of the Boolean networks of a given R-graph, and (c) finding Boolean networks given an R-graph and a desired set of fixed-point attractors.

The second example is devoted to hypothetical regulations. Such regulations will enable us to exemplify Griffin’s whole-query vs. query-splitting approaches.

The third example shows how uncertainty in the steady state behavior of a system can be effectively expressed by combining partially known fixed-point attractors and explicit exclusions of fixed-point attractors.

Algorithm 2: Radial exploration strategy for query splitting

<table>
<thead>
<tr>
<th>Input:</th>
<th>Q, a griffin query</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>r, the exploration radius</td>
</tr>
<tr>
<td>Output:</td>
<td>writes all satisfying networks</td>
</tr>
<tr>
<td>1 begin</td>
<td></td>
</tr>
<tr>
<td>2 core ← ∅, V_m ← ∅, V_o ← ∅</td>
<td></td>
</tr>
<tr>
<td>3 J ← Q.get-R-interactions()</td>
<td></td>
</tr>
<tr>
<td>4 Q.allow-hypotheses ← false</td>
<td></td>
</tr>
<tr>
<td>for i ∈ J −{NR, true, false} do</td>
<td></td>
</tr>
<tr>
<td>6 if</td>
<td>options(i)</td>
</tr>
<tr>
<td>7</td>
<td>core ← core ∪ options(i)</td>
</tr>
<tr>
<td>8 else if ∅ ∉ options(i) then</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>V_m ← V_m ∪ options(i)</td>
</tr>
<tr>
<td>10 else</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>V_o ← V_o ∪ (options(i) − ∅)</td>
</tr>
<tr>
<td>12 end</td>
<td></td>
</tr>
<tr>
<td>13 for mantle ∈ one-of-each(V_m) do</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>center ← mantle ∪ one-of-each(core)</td>
</tr>
<tr>
<td>15 for n ∈ 0..r do</td>
<td></td>
</tr>
<tr>
<td>16 for crust ∈ one-of-each(combinations(</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>V_o, n)) do</td>
</tr>
<tr>
<td>18</td>
<td>Q.set-as-known-regulations(center ∪ crust)</td>
</tr>
<tr>
<td>19</td>
<td>write(Griffin.find-all-networks(Q))</td>
</tr>
<tr>
<td>20 end</td>
<td></td>
</tr>
<tr>
<td>21 end</td>
<td></td>
</tr>
<tr>
<td>22 for each target(i) do</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>S ← ∅</td>
</tr>
<tr>
<td>24 if i ∈ {MPPA, MMUSPA, MU, OPUSPA, ONUSPA, OUSPA} then</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>S ← {source(i) → target(i)}</td>
</tr>
<tr>
<td>26 end</td>
<td></td>
</tr>
<tr>
<td>27 if i ∈ {MA, MPPA, MUSPA, MUSPA, OPPA, OPLA, OPUSPA} then</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>S ← S ∪ {source(i) → target(i)}</td>
</tr>
<tr>
<td>29 end</td>
<td></td>
</tr>
<tr>
<td>30 if i ∈ {MA, MPPA, MUSPA, OPPA} then</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>S ← S ∪ {source(i) → target(i), source(i) → target(i)}</td>
</tr>
<tr>
<td>32 end</td>
<td></td>
</tr>
<tr>
<td>33 if i ∈ {OA, ONUSPA, ONUSPA, OPUSPA, OUSPA} then</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>S ← S ∪ ∅</td>
</tr>
<tr>
<td>35 end</td>
<td></td>
</tr>
<tr>
<td>36 return S</td>
<td></td>
</tr>
<tr>
<td>37 end</td>
<td></td>
</tr>
<tr>
<td>38 end</td>
<td></td>
</tr>
<tr>
<td>39 Function one-of-each(A):</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>return {x</td>
</tr>
</tbody>
</table>
TABLE 2 | Scalability increase through clause learning in Griffin.

<table>
<thead>
<tr>
<th>R-graph size</th>
<th>No. of Boolean networks</th>
<th>Exclusion of attractors</th>
</tr>
</thead>
<tbody>
<tr>
<td>n m indegree</td>
<td>2^n \prod_{i=1}^{d} (2^{d_i})</td>
<td>\text{No. of exclusionary clauses}</td>
</tr>
<tr>
<td></td>
<td>\sum_{k=1}^{2^n} (2^n!)/((2^n - k)!k!)</td>
<td>NE</td>
</tr>
<tr>
<td>Small R-graphs with constant indegree</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 1 1 1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>2 2 1 2</td>
<td>16</td>
<td>24</td>
</tr>
<tr>
<td>3 4 2 2</td>
<td>256</td>
<td>16,072</td>
</tr>
<tr>
<td>3 6 2 3</td>
<td>4,096</td>
<td>16,777,216</td>
</tr>
<tr>
<td>3 9 3 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. thaliana root (questions 5 and 6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11 50 4</td>
<td>2.2 \times 10^{5,891}</td>
<td>\times</td>
</tr>
<tr>
<td>A. thaliana modified root (question 7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11 90 5</td>
<td>2.2 \times 10^{5,891}</td>
<td>\times</td>
</tr>
<tr>
<td>A. thaliana flower (questions 1, 2, and 3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13 45 4</td>
<td>4.2 \times 10^{28,499}</td>
<td>\times</td>
</tr>
<tr>
<td>A. thaliana sepal (question 4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21 32 2</td>
<td>8.6 \times 10^{12,346,634}</td>
<td>\times</td>
</tr>
</tbody>
</table>

The first four columns (from the left) are the R-graph parameters used in a query. The first column shows the number of vertices, n; the second column shows the number of edges, m; the third column, labeled d, is an indegree, while the fourth column, labeled #, is the number of vertices in the R-graph having indegree d. The fifth column, no. of Boolean networks, is an upper bound of the number of satisfying Boolean networks. The last three columns give an idea of the difficulty of achieving an exclusion of attractors in Griffin's solutions through two different strategies. These columns show an estimate of the number of clauses required to exclude all possible attractors (column labeled no. of exclusionary clauses) and whether adding such constraint is tractable using naive encoding (column labeled NE) or using clause learning (column labeled CL). Naive encoding is the strategy used in early versions of Griffin and consists of adding in advance all constraints to the formula that prohibit the unwanted attractors. With naive encoding, the size of the formula required to enforce an exact set of desired attractors is too big in all case studies presented here (although question 4 do not ask for such condition), while clause learning allows Griffin to provide answers to all queries. Clause learning endows Griffin with a significant improvement in scalability.

Some features, such as the use of constraints on mutants and specification of desired cyclic attractors, have been left out from this article. However, we have illustrated the use of these capabilities in previous articles (Rosenblueth et al., 2014; Weinstein et al., 2015; Azpeitia et al., 2017; García-Gómez et al., 2017).
Algorithm 3: Exploration of sets of fixed point attractors

Input: $Q$, a griffin query
Output: $S$, a sample of collections of fixed point attractors

begin
  $S \leftarrow \emptyset$
  $\varphi \leftarrow\text{griffin.encode}(Q)$
  while $\text{griffin.has-more-models}(\varphi)$ do
    $f \leftarrow\text{griffin.next-Boolean-network}(\varphi)$
    $A \leftarrow\text{griffin.compute-attractors}(f)$
    $\psi \leftarrow\text{griffin.encode}(A)$
    $\varphi \leftarrow \varphi \land \neg \psi$
    $S \leftarrow S \cup A$
  end
return $S$
end

2.2.1. First Case Study: Arabidopsis thaliana Flower Model

We start by taking the $A.\ thaliana$ floral organ determination gene regulatory network model developed by Alvarez-Buylla et al. (2010) (henceforth referred to as $A.\ thaliana$ flower model). In their work (Mendoza and Alvarez-Buylla, 1998; Mendoza et al., 1999; Espinosa-Soto et al., 2004; Alvarez-Buylla et al., 2010), these authors report that this network is capable of reproducing the stable gene expression observed during the development of the flower organs, plus some inflorescence stages. Observe that these authors do not use a formal definition of regulation. Moreover, the search for networks satisfying the regulation graph (built from the literature) is done manually and without employing the expected set of fixed points to guide the search. Normally, these authors report only one Boolean network consistent with the regulation graph and recovering a set of fixed points coherent with the experimental information. We will show our attempts at reproducing their last results (Alvarez-Buylla et al., 2010) with Griffin.

2.2.1.1. Regulation graph as input

We first illustrate a direct attempt at trying to recover the 10 known fixed-points (i.e., steady states) for the model of $A.\ thaliana$ without using the desired set of fixed-points as input. This approach assumes that genes do not oscillate. This assumption can be justified by observing that even with such a constraint we did not recover the desired set of steady states. Hence, if we allowed for the possibility of cyclic attractors, thus...
FIGURE 2 | Schematic of Griffin’s information flow. The query at the top-left corner is first parsed. Depending on the options it is either directly encoded into a Boolean formula or split into simpler queries. The Boolean formula is fed into a SAT solver. Each satisfying assignment returned by the SAT solver is decoded into a synchronous Boolean network, a candidate solution. It is also possible that the SAT solver concludes that the formula is a contradiction, that is, there are no synchronous Boolean networks satisfying the query. If a candidate solution is found, blocking constraints prevent finding the same solution twice. Further properties of the candidate solution are verified, possibly resulting in the generation of additional constraints. The runtime blocking analysis concludes with the rejection or acceptance of the candidate solution. If no candidate solution passes the runtime blocking analysis, then Griffin concludes there are no satisfying solutions to the query.
enlarging the size of the search space, we would be even less likely to recover the desired set of fixed-point attractors.

Question 1. The first question we asked Griffin was: What are the Boolean networks satisfying the known regulation graph of the A. thaliana flower model (Figure 3A) and that do not have cyclic attractors?

Listing 1 in Supplementary Material shows the corresponding query as posed to Griffin. Griffin's answer was positive, giving a large number of satisfying Boolean networks before exhausting the available resources. We hence limited to 100 the number of answers. These answers, however, were similar to each other, and none of them had the desired set of fixed-point attractors. Consequently, we devised search strategies to look for diversity in the answers.

Question 2. What are the possible distinct fixed-point attractor sets (FPASs) for Boolean networks compatible with the known regulatory interactions of A. thaliana flower model and that do not have cyclic dynamics?

Normally, Griffin can be used through a query language. However, if such a language is not expressive enough for the desired query, it is possible to run Griffin through its application programming interface (API). Algorithm 3 exemplifies the use of Griffin via its API. This algorithm computes distinct FPASs. To prevent the possibility of finding the same combination more than once, there is a blocking constraint representing the negation of that particular FPAS.

It is important to note that this search algorithm is not complete, in the sense that some FPASs may be excluded from the answer as a result of the blocking strategy. Adding a blocking clause not only prevents a particular FPAS from being found again; it also blocks all FPASs that are a superset of the FPAS of a satisfying solution found by Griffin. If the whole set of FPASs is required, this algorithm would have to be modified with more sophisticated blocking techniques. In our case, this was not deemed necessary because the computation of the sample of FPASs already exhausted the computational resources.

Griffin found 2,896 Boolean networks. The sets of attractors were not only distinct but some of them also had distinct cardinalities. The minimum number of fixed-point attractors for the networks in the sample was two; the maximum was 15. We found 133 different fixed-point attractors in such a sample. Of the 10 desired fixed-point attractors that are compatible with the regulation graph for A. thaliana flower (Figure 3A), Griffin was only able to find nine, as explained in Figure 4.

To assess the diversity of the sample, we analyzed the similarity of the FPASs between each pair of found networks. Figure 5 is a visualization of the similarity matrix computed for the 2,896 networks in the sample. The visualization shows not only that many FPASs share common elements, but also that there are significant number of them that do not share any fixed-point attractor (2,896² × 0.46 entries in the matrix are zeros; that is, 46% of pairs are completely different in their fixed-point attractors). Our conclusion is that results produced by Griffin show that there is a significant diversity of fixed-point attractors in Boolean networks with dynamics compatible with a given regulation graph (see Figure 6).

Note that a regulation graph (without biological constraints) can be satisfied by a vast number of Boolean networks corresponding to different dynamic behaviors. Therefore, in addition to the regulation graph, biological constraints on Griffin's input (as the set of desired attractors) are important for producing more accurate Boolean networks.

We show next that Griffin was able to find fully compliant Boolean networks once steady-state constraints were added to the query.

2.2.1.2. Regulation graph plus expected set of attractors as input

Question 3. The third question we asked Griffin was: Are there any Boolean networks satisfying the known regulatory interactions of the A. thaliana flower model that have exactly the known set of fixed-point attractors?

Listing 2 in Supplementary Material has a new constraint with respect to the previous query, asking Griffin to limit the search to those Boolean networks having exactly the set of known fixed-point attractors of Figure 4. As with the previous question, Griffin's answer was positive, giving a large number of satisfying Boolean networks. Before exhausting the resources, Griffin found 328,565 Boolean networks satisfying the constraints. We observe that were it not for the fact that Griffin blocks attractors on the fly (as opposed to blocking unwanted attractors in advance), the resources would be exhausted when building the formula (given the length of the formula required for blocking any additional fixed-point or cyclic attractor). After adding to the input the set of desired attractors, the number of solutions was still vast. Griffin allows adding constrains that could reduce even more the number of solutions. An example of such constraints is the effect of mutations on the set of attractors (Rosenblueth et al., 2014; Azpeitia et al., 2017).

As well as inferring Boolean networks, Griffin can be used to perform analysis, as we show next. Figure 7 depicts a state space diagram for the 10 basins of attraction for the network having the combination reported in the sixth row of Table 3. Our results would suggest that gene regulatory networks are robust allowing them to be stable in developmental and evolutionary scales, as observed for flowers (Drinnan et al., 1994).

2.2.2. Second Case Study: A. thaliana Sepal Primordium Polarity

La Rota et al. (2011) developed a model of sepal primordium polarity in the young flower of A. thaliana. These authors defined a set of expected attractors by analyzing the expression patterns of the most important genes during sepal development. Then, using published data, they defined the set of known regulatory interactions and looked for putative binding sites indicating putative missing regulatory interactions. Finally, using mathematical programming, they built Boolean networks that conformed with the known and missing interactions to verify if any of such networks could produce the set of expected attractors. We chose this example as it enables us to illustrate how
incomplete knowledge about the regulators of specific genes can be captured in a Griffin query. In the following subsection, we also illustrate how to set a query-splitting strategy.

2.2.2.1. Hypothetical regulations

The model of sepal primordium polarity for *A. thaliana* of La Rota et al. (2011) (henceforth referred to as *A. thaliana* sepal) was integrated considering incomplete knowledge and uncertainty in the data. After analyzing the data, these workers concluded that the model based on experimental data was unable to obtain the set of expected attractors. Then, based on bioinformatic data, they included a number of hypothetical missing regulations that could be necessary to obtain the desired attractors. The computational methods used by La Rota et al. (2011) were based on mixed integer linear programming representations and iterative procedures. They used their algorithm to reproduce steady states and regulatory interactions to be functional, as well as other biologically meaningful constraints. We were interested in testing Griffin’s ability to find satisfying solutions to this problem.
Figure 4 | A sample of 133 fixed-point attractors that are compatible with the interaction graph of Figure 3A. Black squares represent expressed genes; gray squares represent unexpressed genes. From bottom-left to top-right, there are seven groups of 19 rows and 13 columns each. Each column is labeled with a particular gene. Each row is an expression profile for the 13 genes and corresponds to a compatible fixed-point attractor for the regulations present in the interaction graph. At the top-right corner the 10 known fixed-point attractors corresponding to the cell types: inflorescence shoot apical meristem I1, I2, I3, I4; sepal SE; petal PE1, PE2; stamen ST1, ST2; and carpel CA. Nine of the 10 fixed-point attractors present in the sample have been highlighted in green. The algorithm used to generate the sample did not find the I4 profile highlighted in red.

Question 4. The question we asked Griffin was: Are there any Boolean networks satisfying the known and hypothetical regulations of A. thaliana sepal model that have the known set of fixed-point attractors with no constraints on possible additional attractors?

Listing 3 in Supplementary Material shows the Griffin query used to answer the previous question. It is simple to list hypothetical regulations, such as the 19 hypotheses formulated by La Rota et al. (2011) in their regulation graph (Figure 3B [based on Figure 3D of La Rota et al., 2011]). By construction, the regulations of satisfying networks found by Griffin, if any, are functional (Rosenblueth et al., 2014). Therefore, no special instruction is needed to specify this requirement. We included the expected fixed-point attractors as a constraint. Additionally, we asked Griffin to perform query splitting (see definition 16, below) on the question using the radial exploration strategy of Algorithm 2. By varying the query option corresponding to the radius, Griffin limited the number of simultaneous hypotheses considered at a time for each query $q \in s(P)$ in the splitting (see definition 16).

We asked several queries varying the radius (paragraph “Whole Queries and Query Splitting”). As there are 19 hypotheses, this is also the maximum possible radius. Table 4 summarizes the results. It can be seen that having control on the radius provided valuable information about the hypotheses. In particular, Griffin found all combinations of hypothetical regulations that reproduced the expected attractors, including the smallest set of required hypothetical regulations to obtain the attractors. In this case, there were no satisfying networks below
FIGURE 5 | Similarity matrix of sets of fixed-point attractors for a sample of 2,896 Boolean networks satisfying the interaction graph of Arabidopsis thaliana flower model shown in Figure 3A and that do not have cyclic attractors. The similarity matrix shows the degree to which different sets of fixed-point attractors share the same elements. Each row and column represents a particular witness. The entry \((i,j)\) represents the similarity between the fixed-point attractor sets of witness \(i\) and witness \(j\). The maximum similarity between two networks equals 1 if they have the same set of fixed-points. Algorithm 3 ensures that the witnesses have different sets of fixed-point attractors. Therefore, the only entries in the similarity matrix that are equal to one are in the diagonal. The minimum similarity between two networks is zero. This happens when the intersection of their sets of fixed-point attractors is empty. High values of similarity are colored red while low values of similarity are colored blue. Zero similarity corresponds to black. The similarity of two sets was found using the cosine similarity given by
\[
sim(x,y) = \frac{x \cdot y}{\|x\| \|y\|}
\]
where \(x\) and \(y\) are binary vectors encoding the presence or absence of different fixed-point attractors.

a radius eight. In other words, the simplest models able to satisfy the constraints must include the right combination of at least eight hypotheses. There were 32 solutions at this radius, and only eight different right combinations (same number of associated interaction graphs). Griffin was able to compute all the solutions for the query, a total of 439,296. Notice that in the R-graph of this example, there are 19 hypothetical regulations, which is equivalent to \(2^{19} = 524,288\) ordinary regulation graphs. Griffin analyzed all these possible ordinary regulation graphs in a single question. We can also see the difference between the cardinality of the query splitting \(|s(P)| = 524,288\) and the cardinality of the instantiations of the corresponding R-graph \(|G_G| = 6,912\). That is, only about 1.32% of the ordinary regulation graphs produced solutions. To measure the computational cost of the query-splitting strategy, we compared the computing time for a single query (with no query-splitting strategy). This example shows that making 524,288 questions needed fewer computational resources than a single complex question did. Figure 8 shows graphically...
the information of the table. It can be seen that the computing time correlates well with the number of solutions found.

### 2.2.3. Third Case Study: *A. thaliana* Root Stem Cell Niche

The *A. thaliana* root stem cell niche is a model system for the study of plant development and stem cell niche dynamics. We consider here the work of Azpeitia et al. (2013). These authors have used multiple strategies for finding missing gene interactions able to explain the observed experimental steady state dynamics. Their study faces the computational complexity of searching within the space of potential Boolean functions, which is large and grows exponentially in the number of regulator genes. Here, we review some of their proposed procedures for predicting missing interactions, expressing such procedures as symbolic constraints. In fact, we have previously used Griffin (Rosenblueth et al., 2014) to reproduce and extend (by finding other satisfying Boolean networks) a previously published Boolean network for *A. thaliana* root stem cell niche (Azpeitia et al., 2010). In this subsection, we illustrate how Griffin can deal with uncertainty in the gene interactions as well as in the attractors.

#### 2.2.3.1. Using R-graphs to express hypotheses

In the next question, we explore the root stem cell niche model of Azpeitia et al. (2013) (henceforth referred to as *A. thaliana* root model). From the point of view of the Boolean network inference problem, this model is interesting because it is rich in hypotheses, that is, it incorporates uncertainties about the existence or sign of its regulatory interactions.

**Question 5.** Does there exist a Boolean network satisfying the *A. thaliana* root model?

The *A. thaliana* root model includes 30 hypothetical regulations as well as 20 known regulations. Figure 3C shows the set of regulations and Listing 4 in Supplementary Material exhibits the query file for the corresponding question. Griffin’s answer to the query was negative. This is compatible with the results reported by Azpeitia et al. (2013). Griffin’s conclusion, however, is important for the following reasons:

1. The search space of Boolean functions is vast. The number of regulators for each gene is as follows (see Figure 3C): ACR4, 1; AUX_IAA, 4; AUXIN, 3; CLE40, 3; JKD, 5; MGP, 5; MR165, 4; PHB, 8; SCR, 6; SHR, 4; WOX5, 7. The total number of possible Boolean networks is then given by: $2^{3} \times 2^{4} \times 2^{3} \times \ldots \times 2^{7} \approx 4.9 \times 10^{174}$.

2. Azpeitia et al. (2013) performed extensive exploration of the search space, applying a set of procedures designed to reduce the complexity of the search and to focus on solutions with biological relevance. Their technique incorporated a heuristic search that tested combinations of hypothetical regulations adding each regulation one at a time. After three months of testing no solution was found. Because only a portion of the search space was explored, it was not possible to conclude whether or not there were satisfying Boolean networks.

3. When Griffin’s answer is negative, i.e., when it concludes that the query is unsatisfiable, its conclusion is often found as a consequence of a trivial contradiction in the Boolean formula given to the SAT solver. If that is the case, the answer is found immediately. In our case, however, Griffin took 6.8 ms to give the answer.

4. Despite the size of the search space, this example illustrates that Griffin’s runtime blocking strategy is effective. Inspecting Griffin’s log files, we learned that there is no trivial contradiction in the query formula. Griffin iterates over 22 candidate Boolean networks, rejecting all because they exhibit additional attractors. Griffin blocks 15 fixed-point attractors, five cyclic attractors of length two, and two cyclic attractors of length three. After blocking one of the incompatible attractors of the 22nd network, a contradiction is generated, allowing Griffin to conclude the nonexistence of a satisfying solution.

The fact that no network satisfied the constraints suggested that such constraints might be incorrect. After analyzing the assumptions of the model, we learned that there existed uncertainty about the specification of the fixed-point attractors of the model. In particular, the model contains (1) an hormone, auxin, whose distribution is graded along the *A. thaliana* root, and (2) a mobile peptide, CLE40, that diffuses from its expression domain. Neither the exact region where auxin is active nor the diffusion coefficient of CLE40 are known. Thus, their Boolean values in the attractors is not clear. Griffin is able to express such degree of uncertainty by the use of partially known fixed-point attractors. This is explored in the following subsection.

#### 2.2.3.2. Partially known fixed-point attractors

Uncertainty in stationary gene expression profiles when inferring the dynamics of a molecular network from the literature occurs...
FIGURE 7 | Depiction of basins of attraction for the ten known fixed-point attractors corresponding to the Boolean network having the combination of basin sizes reported in the sixth row in Table 3. The biggest component with 2,488 states corresponding to ST2 is colored red, the second biggest with 2,432 states, corresponds to CA and is colored light blue, I1 is colored purple, I2 dark green, I3 dark blue, I4 orange, SE brown, PE1, the smallest component with only four states, is colored yellow, PE2 pink and ST1 light green. The fixed-point attractors are represented by bigger circles than those of non stationary states. Griffin use the algorithm of Dubrova and Teslenko (2011) to compute the attractors of the Boolean networks. Binary Decision Diagrams and the backward reachable set algorithm of Garg et al. (2008) are applied to each attractor to compute their corresponding basins.

TABLE 3 | Frequency of sizes of basins of attraction for the 10 known fixed-point attractors of Figure 4 in a sample of 328,565 Boolean networks found by Griffin that recover the exact set of fixed-point attractors.

<table>
<thead>
<tr>
<th>I1</th>
<th>I2</th>
<th>I3</th>
<th>I4</th>
<th>SE</th>
<th>PE1</th>
<th>PE2</th>
<th>CA</th>
<th>ST1</th>
<th>ST2</th>
<th>Num. of networks</th>
</tr>
</thead>
<tbody>
<tr>
<td>896</td>
<td>896</td>
<td>512</td>
<td>512</td>
<td>2,596</td>
<td>4</td>
<td>2,592</td>
<td>48</td>
<td>40</td>
<td>96</td>
<td>156,227</td>
</tr>
<tr>
<td>884</td>
<td>884</td>
<td>524</td>
<td>524</td>
<td>2,596</td>
<td>4</td>
<td>2,592</td>
<td>48</td>
<td>40</td>
<td>96</td>
<td>95,352</td>
</tr>
<tr>
<td>908</td>
<td>908</td>
<td>500</td>
<td>500</td>
<td>2,596</td>
<td>4</td>
<td>2,592</td>
<td>48</td>
<td>40</td>
<td>96</td>
<td>57,725</td>
</tr>
<tr>
<td>968</td>
<td>968</td>
<td>440</td>
<td>440</td>
<td>2,596</td>
<td>4</td>
<td>2,592</td>
<td>48</td>
<td>40</td>
<td>96</td>
<td>19,259</td>
</tr>
<tr>
<td>896</td>
<td>896</td>
<td>512</td>
<td>512</td>
<td>2,588</td>
<td>4</td>
<td>2,584</td>
<td>52</td>
<td>44</td>
<td>104</td>
<td>1</td>
</tr>
<tr>
<td>916</td>
<td>916</td>
<td>524</td>
<td>524</td>
<td>164</td>
<td>4</td>
<td>168</td>
<td>2,432</td>
<td>56</td>
<td>2,488</td>
<td>1</td>
</tr>
</tbody>
</table>

Cyclic attractors were prohibited in the query. There are only six different combinations of basin sizes. The last column shows the number of Boolean networks having the corresponding combinations of basin sizes.
Question 7. What are the Boolean network satisfying the complexity of the query.

The modified definition of its steady-state behavior?

following question that partially known fixed-point attractors,
using partially known fixed-point attractors. We show in the
the attractors could also be important when inferring a molecular
being able to express uncertainty (due to lack of information ) in
found in Table S1 in Supplementary Material.

combined with explicit exclusion of certain fixed-point attractors,
network, and that
find a satisfying solution after 2.9 s. This example shows that
known fixed-point attractors as well as prohibitions. A detailed
of active hypotheses at each radius. The third column has the cumulati
graphs. Not all R-graphs are consistent with the specified con straints. As we can see in the fourth
column that gives the total number of satisfying R-graphs for each run , it is necessary to add at least eight simultaneous hypotheses to find satisfying networks. Each satisfying R-graph can have one or more satisfying Boolean networks. The total number of satisfying networks is shown in the fifth column. The processing time for each query is shown in the last column.

As a comparison of performance, we also timed the query without using the center-radius strategy. The elapsed time was of 5:14 h. This illustrates that, in this case, splitting the query
into several simpler ones (the third column gives the exact number of queries run), does not add time. The center-radius strategy finished over 4 min before the single query did.

frequently. To an extent, Griffin can deal with uncertainty by
using partially known fixed-point attractors. We show in the
following question that partially known fixed-point attractors, combined with explicit exclusion of certain fixed-point attractors, can be used to formulate complex hypotheses.

Question 6. Does there exist a Boolean network satisfying the
A. thaliana root model having uncertainty in the definition of its
steady-state behavior?

Listing 5 in Supplementary Material shows a modified query of the previous question. The constraints on the fixed-point attractors have been changed. Instead of just asking for nine well-defined fixed-point attractors, some uncertainty in the specification of them has been introduced by including partially known fixed-point attractors as well as prohibitions. A detailed explanation on the syntax and meaning of the constraint can be found in Table S1 in Supplementary Material. Griffin was able to find a satisfying solution after 2.9 s. This example shows that being able to express uncertainty (due to lack of information) in the attractors could also be important when inferring a molecular network, and that Griffin can easily express such uncertainty.

In the following example, we would like to significantly extend the number of hypothetical regulations of the A. thaliana root model (henceforth referred to as the modified A. thaliana root model), and test whether Griffin is able to cope with the complexity of the query.

Question 7. What are the Boolean network satisfying the modified A. thaliana root model having uncertainty in the definition of its steady-state behavior?

The modified A. thaliana root model included the 20 known interactions of the original model, as well as 70 hypothetical interactions, 66 of which were of unknown sign and four were positive. The corresponding R-regulation graph is shown in Figure 3D.

The significance of this example is twofold:

1. The search space of Boolean functions is significantly larger than that of the original A. thaliana root model. The number of regulators for each gene is as follows (see Figure 3D): ACR4, 8; AUX_IAA, 10; AUXIN, 9; CLE40, 7; JD, 8; MGP, 8; MR165, 4; PHB, 11; SCR, 9; SHR, 8; WOX5, 8. The total number of possible Boolean networks is then given by: $2^8 \times 2^{10} \times 2^{29} \times \cdots \times 2^8 \approx 4.8 \times 10^{1,661}$.

2. Griffin required 5,520 variables to represent the update function in this problem, plus 160 regulation variables (the $R^+$s and $R^-$s) plus an unidentified number of switching variables required to represent the formula in an “equisatisfiable” conjunctive normal form (Tseitin, 1968).

The query for question 7 is shown in Listing 6 in Supplementary Material. A satisfying Boolean network was found after 24 min., which has a total of 83 regulations, as can be seen in Figure 9.

For questions 1, 2, 3, and 7 we ran Griffin on a Dell PowerEdge T320 Server with 80 GB of RAM. For the rest of the questions we used a laptop computer with Intel(R) Core(TM) i7-4710HQ CPU @ 2.50GHz and 16 GB of RAM.

3. DISCUSSION

In this section, we first give a summary of Griffin. Next, we review methods for Boolean-network inference and related formalisms, so as to place Griffin in context within this area of research. Finally, we discuss possible directions for future work.

### TABLE 4 | Satisfying R-graphs and Boolean networks found by Griffin for the center-radius exploration strategy applied to A. thaliana sepal model.

<table>
<thead>
<tr>
<th>Radius</th>
<th>Regulation graphs at radius</th>
<th>Cumulative regulation graphs</th>
<th>Cumulative satisfying Regulation graphs</th>
<th>Cumulative satisfying networks</th>
<th>Cumulative computing time [h:m]</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>75,582</td>
<td>1.69,766</td>
<td>8</td>
<td>32</td>
<td>0:01</td>
</tr>
<tr>
<td>9</td>
<td>92,378</td>
<td>2.62,144</td>
<td>84</td>
<td>448</td>
<td>0:03</td>
</tr>
<tr>
<td>10</td>
<td>92,378</td>
<td>3.54,522</td>
<td>410</td>
<td>3,120</td>
<td>0:05</td>
</tr>
<tr>
<td>11</td>
<td>75,582</td>
<td>4.30,104</td>
<td>1,243</td>
<td>14,048</td>
<td>0:11</td>
</tr>
<tr>
<td>12</td>
<td>50,388</td>
<td>4.80,492</td>
<td>2,661</td>
<td>45,120</td>
<td>0:28</td>
</tr>
<tr>
<td>13</td>
<td>27,132</td>
<td>5.07,624</td>
<td>4,303</td>
<td>1,08,912</td>
<td>1:06</td>
</tr>
<tr>
<td>14</td>
<td>11,628</td>
<td>5.19,252</td>
<td>5,675</td>
<td>2,04,496</td>
<td>2:08</td>
</tr>
<tr>
<td>15</td>
<td>3,876</td>
<td>5.23,128</td>
<td>6,481</td>
<td>3,07,968</td>
<td>3:22</td>
</tr>
<tr>
<td>16</td>
<td>969</td>
<td>5.24,097</td>
<td>6,809</td>
<td>3,86,496</td>
<td>4:22</td>
</tr>
<tr>
<td>17</td>
<td>171</td>
<td>5.24,268</td>
<td>6,897</td>
<td>4,25,968</td>
<td>4:54</td>
</tr>
<tr>
<td>18</td>
<td>19</td>
<td>5.24,287</td>
<td>6,911</td>
<td>4,37,728</td>
<td>5:08</td>
</tr>
<tr>
<td>19</td>
<td>1</td>
<td>5.24,288</td>
<td>6,912</td>
<td>4,39,296</td>
<td>5:10</td>
</tr>
</tbody>
</table>

Each row summarizes the results of individual runs. The first column shows the radius. The second column has the corresponding binomial coefficient giving the number of combinations of active hypotheses at each radius. The third column has the cumulative number of R-graphs. Not all R-graphs are consistent with the specified constraints. As we can see in the fourth column that gives the total number of satisfying R-graphs for each run, it is necessary to add at least eight simultaneous hypotheses to find satisfying networks. Each satisfying R-graph can have one or more satisfying Boolean networks. The total number of satisfying networks is shown in the fifth column. The processing time for each query is shown in the last column.

As a comparison of performance, we also timed the query without using the center-radius strategy. The elapsed time was of 5:14 h. This illustrates that, in this case, splitting the query into several simpler ones (the third column gives the exact number of queries run), does not add time. The center-radius strategy finished over 4 min before the single query did.
In the inference of Boolean networks form regulation graphs, the fact that multiple networks might be consistent with the same (given) regulation graph is sometimes neglected. Often, publications reporting the inference of Boolean networks consider a single network. Even if several such networks are analyzed, only a fragment of the possibly vast number of consistent networks is usually considered. Griffin, by contrast, employs a formal definition of regulation (Naldi et al., 2007; Richard et al., 2012; Mori and Mochizuki, 2017) to build an implicit representation of all possible networks satisfying the regulation graph. Griffin then augments such a representation with biological constraints, resulting in a uniform representation of all potential networks and at the same time reducing such potential number of solution networks. This representation is then given to a powerful, symbolic search mechanism.

A key contribution to Model Checking (Clarke et al., 1999) was the incorporation of symbolic representations of data (Burch et al., 1992). Such representations allowed Model Checking to verify systems, first with $10^{20}$ states, and subsequently with many more. As with model checkers, Griffin is able to handle large systems because of also using symbolic data structures.

There exist two main symbolic techniques: binary-decision diagrams (BDDs) and SAT solvers. Griffin employs both, for solving different subproblems. Griffin finds cyclic attractors with a SAT solver (Dubrova and Teslenko, 2011), whereas it computes basins of attraction (Garg et al., 2008) with BDDs. The top level in Griffin, which builds a Boolean formula representing all biological information and constraints, also employs a SAT solver. Griffin thus capitalizes on the phenomenal recent developments of such solvers. This dramatic progress has “resulted in speed-ups of many orders of magnitude that have turned many problems that were considered intractable in the 1980s into trivially solved problems now.” (Franco and Martin, 2009).

Another strong point of our tool is the concept of an R-graph, a generalization of ordinary regulation graphs. R-graphs represent advantages from both the user and the computational viewpoints. The user has available the whole repertoire of combinations involving the sign of a regulation and whether a regulation is compulsory or simply hypothetical. For example, the user can specify a hypothetical regulation that, if present, should be negative and cannot be positive. From a computational point of view, an R-graph encompasses all hypothetical regulations in one formula, thus avoiding having...
FIGURE 9 | Modified A. thaliana root Boolean network found by Griffin satisfying question 7. (A) Interaction graph of the solution has 83 regulations of the 90 possible (see Figure 3D). (B) There are 11 fixed-point attractors in the solution. For each expression profile active genes are shown in green while inactive genes are shown in gray. The number corresponds to the decimal notation used in Table S1 in Supplementary Material.

to perform many separate tests. Computer tools employing ordinary regulation graphs, by contrast, usually test all possible combinations of presence and absence of all hypothetical regulations, resulting in numerous analyses.

Griffin is hence not only an innovative but also powerful computer tool for the inference of Boolean networks. It is possible, for example, to detect erroneous inferences due to the experimental, error-prone nature of biological data. Regulations that were believed to be positive might in fact turn out to be negative or vice versa. Griffin has detected these situations, as reported in Azpeitia et al. (2017).

3.2. Related Work

As a first comparison between our work and related articles, it is important to point out a difference in the type of input data. In this work, Griffin input is composed of (partial) information about the network topology (R-graphs) along with other data representing biological constraints. R-graphs contain information on genetic connectivity that was inferred from data obtained by direct measurement of gene expression or protein interaction. By contrast, inference methods proposed by other authors (e.g., Laubenbacher and Stigler, 2004) have an input composed of time series together with other data that capture information on the network dynamics. Time series, unlike the R-graphs, represent experimental data obtained by direct measurement of gene expression or protein interaction.

There are a number of tutorials on Boolean-network inference (D’haeseleer et al., 2000; Markowetz and Spang, 2007; Karlebach and Shamir, 2008; Hecker et al., 2009; Hickman and Hodgman, 2009; Berestovsky and Nakhleh, 2013). From these tutorials we can classify algorithms according to (1) the expected input, (2) the kind of model inferred, and (3) the search strategy.

Much of the effort in Boolean-network inference has been aimed at having a binarized time-series data as input. Hence, multiple methods have been proposed and some of them have been compared with each other (Berestovsky and Nakhleh, 2013). An influential method in this category is REVEAL (Liang et al., 1998) (employing Shannon’s mutual information between all pairs of molecules to extract an influence graph; the truth tables of the update functions for each molecule are simply taken from the time series). The use of mutual information and time-series for the inference of Boolean networks continues to develop (Barman and Kwon, 2017), as well as alternative methods based on time series. Example are Han et al. (2014) (using a Bayesian approximation), Shmulevich et al. (2003), Lähdesmäki et al. (2003), Akutsu et al. (1999), Laubenbacher and Stigler (2004), and Layek et al. (2011) (using a generate-and-test method, generating all possible update functions for one gene and testing with the input data). Extra information can be included in addition to time-series data. For example, an expected set of stable states (Layek et al., 2011), previously known regulations (Haider and Pal, 2012) and gene expression data (Chueh and Lu, 2012) are used as an aid to curtail the number of possible solutions.

Griffin belongs to a second family of methods taking as input a possibly partial regulation graph (perhaps obtained form the literature). There are also approaches employing both time-series data and a regulation graph, such as Ostrowski et al. (2016).
A third important area of research is the development of algorithms taking as input temporal-logic specifications, based on Model Checking (Clarke et al., 1999). Works following this approach are: Calzone et al. (2006b), Mateus et al. (2007), and Streck and Siebert (2015).

As for the kind of model inferred, here we would be concerned with Boolean networks and similar formalisms. Among the nonprobabilistic approaches, we find synchronous Boolean networks, asynchronous networks based on Thomas’s formalism (Bernot et al., 2004; Khalis et al., 2009; Corblin et al., 2012; Richard et al., 2012), and polynomial dynamical systems (Laubenbacher and Stigler, 2004). Typically, methods based on temporal logic infer Kripke structures, which are closely related to Boolean networks. Probabilistic models, one of the other hand, include Bayesian networks, and have the advantage of being able to deal with noise and uncertainty.

From the search-strategy point of view, Boolean-network inference methods may employ a simple random-value assignment (Pal et al., 2005), exhaustive search (Akutsu et al., 1999) or more elaborate algorithms. The work of Chueh and Lu (2012) is based on p-scores and that of Liang et al. (1998) guides search with Shannon’s mutual information. Genetic algorithms are used by Saez-Rodriguez et al. (2009) and Ghaffarizadeh et al. (2017). Linear programming is the basis of Tarissan et al. (2008). The methods of Ostrowski et al. (2016) and Corblin et al. (2012), are based on Answer Set Programming (Brewka et al., 2011). Algebraic methods use reductions (of polynomials over a finite field) modulo an ideal of vanishing polynomials.

Approaches based on temporal logic (sometimes augmented with constraints), such as Calzone et al. (2006b) and Mateus et al. (2007), normally employ Model Checking (Clarke et al., 1999). Model Checking, in turn, is often based on symbolic approaches: BDDs and SAT solvers. Biocham (Calzone et al., 2006b) and SMBioNet (Bernot et al., 2004; Khalis et al., 2009; Richard et al., 2012) use a model checker as part of a generate-and-test method.

Having classified various approaches, we now mention a work similar to ours in spirit: Pal et al. (2005). These authors also propagate fixed-point constraints onto the truth tables of the update functions of each variable (molecule species). There is, however, no search technique, other than randomly giving values to the remaining entries of such truth tables. There is a random assignment of values to such entries, and a check for unwanted attractors. Neither is there a formal definition of regulation.

In contrast with the logical approach of our work, we devote our attention now to an algebraic approach to the problem of inference (reverse engineering) of Boolean networks. Instead of using a Boolean network to model the dynamics of a gene network, the algebraic approach (Laubenbacher and Stigler, 2004; Jarrah et al., 2007; Veliz-Cuba, 2012) uses a polynomial dynamical system \( F: S^n \rightarrow S^n \) where \( S \) has the structure of a finite field \( \mathbb{Z}_p \) for an appropriate prime number \( p \). The first benefit of this algebraic approach is that each component of \( F \) can be expressed by a polynomial (in \( n \) variables with coefficients in \( S \)) such that the degree of each variable is at most equal to the cardinality of the field \( S \).

Following a computational algebra approach, in a framework of modeling with polynomial dynamical systems, Laubenbacher and Stigler (2004) propose a reverse-engineering algorithm. This algorithm takes as input a time series \( s_1, \ldots, s_m \in S^n \) of network states (where \( S \) is a finite field), and produces as output a polynomial dynamical system \( F = (F_1, \ldots, F_n): S^n \rightarrow S^n \) such that \( \forall i \in [1, \ldots, n]: F_i \in S[x_1, \ldots, x_n] \) and \( F_i(s) = s_{j+1,i} \) for \( j \in [1, \ldots, m] \). An advantage of the algebraic approach of this algorithm is that there exists a well-developed theory of algorithmic polynomial algebra, with a variety of procedures already implemented, supporting the implementation task. The time complexity of this algorithm is quadratic in the number of variables \( (n) \) and exponential in the number of time points \( (m) \).

Comparing Griffin with the reverse-engineering algebraic algorithms proposed by Laubenbacher and Stigler (2004) and Veliz-Cuba (2012), we found three basic differences. (1) Algebraic algorithms can handle discrete multi-valued variables, while Griffin only handles Boolean (two-valued) variables. Multi-valued variables give more flexibility and detail to the modeling process, but Boolean variables (of Boolean networks) lead to simpler models (see section 1). (2) The input of the algebraic algorithms, typically time series, provides simple information coming directly from experimental measurements, while input of Griffin, R-graphs and Griffin queries, provides structured information allowing a more precise specification of the required Boolean network. (3) The algebraic algorithm of Veliz-Cuba (2012) uses a formal definition of regulation, but this definition does not match the definition of regulation used by Griffin. While Griffin allows for R-regulations based on Boolean combinations of positive and negative regulations, Veliz-Cuba (2012) uses regulations restricted to unate functions \( h \) such that, for all variable \( x \): \( h \) does not depend on \( x \), or \( h \) depends positively on \( x \), or \( h \) depends negatively on \( x \).

Finally, we observe that sometimes results might have been reported overoptimistically. There have been some doubts cast upon the effectiveness of a number of methods of inference of network dynamics (Wimburly et al., 2003), especially those based on more general-purpose learning methods. It is therefore important to establish tests such as the DREAM challenges (Stolovitzky et al., 2007) emphasizing reproducibility.

3.3. Future Work

Both Griffin and model checkers are based on symbolic search-algorithms, either SAT solvers or BDDs. Hence, a natural possibility to consider is the use of an off-the-shelf model checker as a search mechanism for Griffin. Model checkers take as input a temporal-logic formula. We did not find any advantage, however, in using a temporal logic, such as computation-tree logic (CTL) or linear-time logic (LTL), for expressing regulation over ordinary Boolean logic. Once a problem is expressed in Boolean logic, a SAT solver is precisely a method for finding a solution to such a problem. Hence, we disposed of temporal logic and directly expressed the regulation graph in Boolean logic.

Lacking a temporal logic makes the regulation graph the main means of communication of the user with Griffin. This suggested
enriching the concept of a regulation graph as much as possible, which resulted in our generalization: the R-graph. In addition, the value of a number of parameters can be established with Griffin’s query language. Nevertheless, as shown in our case studies, from a Biology point of view, it is easy to fall into situations where Griffin’s query language is not expressive enough. In such situations, the user can employ Griffin’s API, thus bypassing the query language. Moreover, Griffin’s query language (apart from the R-graph, which can be viewed as a shorthand of a Boolean-logic formula) does not have a logical basis. Therefore, compared with a model checker, it would be desirable that Griffin have a temporal logic.

Temporal logic has been shown to be effective for the analysis of Boolean networks (Arellano et al., 2011; Carrillo et al., 2012; Klarner and Siebert, 2015). At the same time, our case studies have shown that clause learning could be effective in pruning the search space. This suggests combining Model Checking with clause learning. Much more complex queries than are currently possible could be formulated to Griffin in this manner. This is, therefore, an important avenue of research.

Another possibility for future research would explore the combination of Griffin with different approaches of Boolean network inference. If the search space is too large for Griffin to give useful solutions because of exhausting available resources, Griffin could incorporate algorithms based on other approaches, such as genetic algorithms, through its API.

Yet another direction of future work would extend Griffin to include either (a) biological information relating species with each other, such as protein complexes, thus reducing the size of the search space, or (b) partially defined Boolean functions for a particular gene.

Finally, we mention an important improvement that would allow the computation and the output of sets of Boolean networks instead of a single network at a time. Currently, the SAT solver employed by Griffin returns one assignment of all the variables appearing in the Boolean formula representing the R-graph together with the constraints (i.e., a “minterm”) at a time. By contrast, it would be more useful to have a SAT Solver returning a set of assignments represented by a partial assignment (i.e., a term that is not necessarily a minterm). This would allow Griffin, in turn, to output sets of Boolean networks.

4. MATERIALS AND METHODS

This section gives definitions and fixes the notation employed in Griffin’s algorithms.

4.1. General Notation

\(N^+\) is the set of positive natural numbers. Unless differently stated, we assume that \(n \in N^+\). \(N^+\) is an initial segment of \(N^+\), \(N^+\) = \{x \in N^+ | x \leq n\}. \(B = \{0, 1\}\) is a set of Boolean values. If \(b \in B\), then \(b^\prime\) is the complement of \(b\). If \(x \in B^n\), we say that \(x\) is a state and \(x\) the \(i\)-th component of \(x\). We sometimes omit parentheses and commas when writing a state. If \(f : X \rightarrow X\), and \(k \in N^+\), we write \(f^k\) to denote the iterated composition of \(f\) with itself, i.e., \(f^1 = f\) and \(f^{i+1} = f \circ f^i\). If \(v \in B^k\), \(p \in (N^+_n)^k\), and \(p_i \neq p_j\) for \(i \neq j\), then \(x[v/p]\) is the state resulting from replacing, for all \(i \in N^+_k\), the \(p_i\)-th component of \(x\) by \(v_i\). We denote as \(x^{-i}\) the vector resulting from replacing the \(i\)-th component of \(x\) with the complement of \(x_i\), that is, \(x^{-i} = x[x_i/i]\). Given \(k < n\), \(v \in B^k\), and \(p \in (N^+_n)^k\), \(B^n[v/p]\) is a subspace of \(B^n\) defined by \(B^n[v/p] = \{x[v/p] | x \in B^n\}\). Alternatively, we describe subspaces by means of strings of zeros, ones, and asterisks used as wildcards indicating the free (unknown value) components. For example, \(B^5[(0,1,0)/(1,4,6)]\) can be written simply as \(0^*1^*0^*\)

4.2. Synchronous Boolean Networks, Regulation Graphs, and State Graphs

Definition 1. We define a synchronous Boolean network with \(n\) components as a function \(f : B^n \rightarrow B^n\). The \(i\)-th component of \(f\) is a function \(f_i : B^n \rightarrow B\). The set of positive natural numbers \(N^+\) is an ordered set of \(n\) variables, \(B\) is a set of \(n\) Boolean values, \(B^n\) is a set of \(n\)-tuples of \(B\), \(B^n[v/p]\) is a subspace of \(B^n\) defined by \(B^n[v/p] = \{x[v/p] | x \in B^n\}\). Alternatively, we describe subspaces by means of strings of zeros, ones, and asterisks used as wildcards indicating the free (unknown value) components. For example, \(B^5[(0,1,0)/(1,4,6)]\) can be written simply as \(0^*1^*0^*\).

To relate a synchronous Boolean network with a molecular network dynamics, we interpret each component of a state \(x\) as representing the activation state of a particular variable denoting a molecule included in the network (which can be a gene, a protein, a hormone, for example). Given a set of variables \(v\), we use a bijective function \(v : N^+_n \rightarrow V\) to relate the components of a state \(x\) with their respective variables. The molecule represented by the variable \(v(i)\), also denoted \(v_i\), is said to be active if \(x_i = 1\) and inactive otherwise.

Definition 2. If \(p \in (N^+_n)^k\) and \(v \in B^k\) with \(k \in N^+_n\), the mutation of \(f\) by \(p = v\) is defined as \(f^{p=v} : B^n \rightarrow B^n\), where, for all \(x \in B^n\) and \(j \in N^+_n\), \(f^{p=v}_j(x) = v_j\) and \(f^{p=v}_j(x) = f_j(x)\) if \(i \in N^+_n - \{p_1, \ldots, p_k\}\). We say that \(f^{p=v}\) is a single-point mutation of \(f\) if \(k = 1\), and a multi-point mutation if \(k > 1\).

Definition 3. (Naldi et al., 2007; Richard et al., 2012; Mori and Mochizuki, 2017). A positive regulation between variables \(v_i\) and \(v_j\) is a function \(R^+ : V \times BN \times V \rightarrow B\) such that

\[
R^+(v_i, f, v_j) = \begin{cases} 
1 & \text{if } \exists x \in B^n : f_j(x) \neq f_j(x^{-i}) \text{ and } x_i = f_j(x) \\
0 & \text{otherwise}
\end{cases}
\]

Similarly, a negative regulation between \(v_i\) and \(v_j\) is a function \(R^- : V \times BN \times V \rightarrow B\) such that

\[
R^-(v_i, f, v_j) = \begin{cases} 
1 & \text{if } \exists x \in B^n : f_j(x) \neq f_j(x^{-i}) \text{ and } x_i \neq f_j(x) \\
0 & \text{otherwise}
\end{cases}
\]

Definition 4. We say that \(G\) is a regulation graph if \(G = (V, I^+, I^-)\), where \(V\) is the set of vertices, \(I^+ \subseteq V \times V\) is the set of positive regulations, and \(I^- \subseteq V \times V\) is the set of negative regulations. If \(i \in I^+ \cap I^-\), we say that \(i\) is an ambiguous regulation.

Definition 5. The regulation graph of \(f\) is \(G_f = (V, I^+, I^-)\) where:

a. \((v_i, v_j) \in I^+ \text{ iff } R^+(v_i, f, v_j) = 1\).

b. \((v_i, v_j) \in I^- \text{ iff } R^-(v_i, f, v_j) = 1\).
Intuitively, the regulation graph of \( f, G_f \), describes the structure of \( f \) and its interpretation as a molecular network where edges represent molecular regulations. Note that \( G_f \) may have both a positive and a negative regulation from \( j \) to \( i \). Observe that whereas a Boolean network \( f \) has a unique regulation graph, a regulation graph may have more than one Boolean network.

**Definition 6.** Given a regulation graph \( G \), we say that \( f \) is a satisfying Boolean network of \( G \), denoted by \( f \models G \), if \( G = G_f \). The set of all satisfying Boolean networks of \( G \) is denoted by \( B(G) = \{ f \mid f \models G \} \).

**Definition 7.** The state graph of a Boolean network \( f \) is the graph \( \hat{G}_f = (\mathbb{B}^n, \{(x,f(x)) \mid x \in \mathbb{B}^n\}) \).

**Definition 8.** We say that \( \omega \subseteq \mathbb{B}^n \) is an attractor of \( f \) if \( \omega \) is a terminal strongly connected component of \( \hat{G}_f \) (Ruet, 2017). That is, \( \omega \) is an attractor if \( \omega \) is strongly connected (each state in \( \omega \) is reachable from any other state in \( \omega \)) and no state or edge of \( \hat{G}_f \) can be added to \( \omega \) without causing \( \omega \) to be no longer strongly connected. In particular, \( \omega \) is a fixed point (or stationary state) of \( f \) if \( |\omega| = 1 \), it is a cyclic attractor otherwise. The size \( k \) of an attractor is defined as its cardinality. The set of attractors of \( f \) is denoted as \( \mathcal{A}(f) \). The set of fixed points of \( f \) is denoted as \( \mathcal{F}(f) \). The set of cyclic attractors of \( f \) is denoted as \( \mathcal{C}(f) \). If \( \omega \in \mathcal{A}(f) \), the basin of attraction of \( \omega \) is \( \mathcal{B}(\omega) = \{ x \mid x \in \mathbb{B}^n \land \exists k \in \mathbb{N}_{\geq 1} : f^k(x) \in \omega \} \).

### 4.3. Boolean-Network Inference

#### 4.3.1. Constrained-Regulation Graphs and R-Graphs

Before formally introducing a Boolean network query, we provide additional definitions that will allow us to express partial knowledge of regulation graphs.

**Definition 9.** A constrained-regulation graph \( G = (V,I) \) is a labeled graph over the set of variables \( V \), where \( I \) is a set of constrained regulations \( I \subseteq V \times L \times V \). Labels \( L = \{ g \mid g : V \times BN \times V \rightarrow \mathbb{B} \} \) are regulation constraints limiting the acceptable Boolean networks.

**Definition 10.** We say that \( f \) satisfies a constrained-regulation graph \( G = (V,I,f) \), if \( \forall u,v : (u,g,v) \in I \) implies \( g(u,f,v) = 1 \). The set of all satisfying Boolean networks of \( G \) is denoted by \( B(G) = \{ f \mid f \models G \} \).

It can be proved that given a regulation graph \( G = (V,I^+,I^-) \), if \( G' = (V,I) \) is defined in such a way that \((u,g,v) \in I \) iff

a. \( (u,v) \in I^+ \) implies \( g(u,f,v) = R^+(u,f,v) \) and

b. \( (u,v) \in I^- \) implies \( g(u,f,v) = R^-(u,f,v) \),

then \( \forall f \in BN : f \models G \iff f \models G' \).

**Definition 11.** An R-graph is a constrained-regulation graph \( G = (V,I) \) where the constraints are Boolean combinations of positive and negative regulations, that is, \( V(u,g,v) \in I : \exists h : \mathbb{B}^2 \rightarrow \mathbb{B} \) such that \( g(u,f,v) = h(R^+(u,f,v),R^-(u,f,v)) \) with \( f \in BN \). Edges of an R-graph are called R-regulations.

**Table 1** shows the list of possible R-regulations for an R-graph. An example of an R-graph is shown in Figure 1. R-graphs are of special interest because they can be viewed as generalizations of ordinary regulation graphs, that are able to express hypotheses and incomplete knowledge about the dependencies between network variables.

**Definition 12.** Given a regulation graph \( G_1 = (V,I^+,I^-) \) and a constrained-regulation graph \( G_2 = (V,I) \), we say that \( G_1 \) is an instantiation of \( G_2 \) if there exists \( f \in BN \) such that \( f \models G_1 \) and \( f \models G_2 \). The set of all instantiations of \( G \) is denoted as \( G_G \).

Instantiations of constrained-regulation graphs are regulations graphs that share the same satisfying networks, Figure 1 shows an example of the set of all instantiations of an R-graph.

### 4.3.2. Network Constraint Problems and Queries

**Definition 13.** A network constraint \( c \) is a Boolean function on the class of Boolean networks, \( c : BN \rightarrow \mathbb{B} \).

The previous definition is a particular case of a Boolean constraint of arity \( k \) as defined in Creignou et al. (2001) and Dantsin and Hirsch (2009). In our case \( k = n^2 \), given that we have \( 2^n \) states in \( \mathbb{B}^n \) and for each state \( x \) there are \( n \) components specifying the next state \( F(x) \).

Examples of meaningful network constraints include, but are not limited to:

1. Network equality: \( c(f) = 1 \iff f = f_0 \) for \( f_0 \in BN \).
2. Network membership: \( c(f) = 1 \iff f \in F_0 \) for \( F_0 \subset BN \).
3. Attractor equality: \( c(f) = 1 \iff A(f) = S_0 \) for \( S_0 \subset P(\mathbb{B}) \).
4. Steady-state inclusion: \( c(f) = 1 \iff F_P(f) \supseteq B_0 \) for \( B_0 \subset \mathbb{B}^n \).
5. Steady-state prohibitions: \( c(f) = 1 \iff x_0 \notin F_P(f) \) for \( x_0 \in \mathbb{B}^n \).
6. Fixed-point equality after mutation: \( c(f) = 1 \iff F_P(f^{p=\omega}) = B_0 \)
   for \( p \in (N_n^+)^k, v \in B^k, k \in N_n^+ \) and \( B_0 \subset \mathbb{B}^n \).
7. Fixed-point inclusion after mutation: \( c(f) = 1 \iff F_P(f^{p=\omega}) \subseteq B_0 \)
   for \( p \in (N_n^+)^k, v \in B^k, k \in N_n^+ \) and \( B_0 \subset \mathbb{B}^n \).
8. State transition presence: \( c(f) = 1 \iff f(x_0) = y_0 \) for \( x_0, y_0 \in \mathbb{B}^n \).

Note that network constraints are different from the constraint regulations of definition 9. As can be seen in the previous listing, the former refer to general properties of \( f \), the Boolean network to be inferred, whereas the latter refer to specific properties of the regulations between pairs of variables, for example restricting a regulation to be a positive regulation.

**Definition 14.** Given a constrained-regulation graph \( G = (V,I) \) and a set of network constraints \( C \), we say that \( P \) is a network constraint problem (NCP) \( P = (G,C) \), where \( C = \{ c_1, \ldots, c_m \} \) is a set of network constraints. We say that \( f \in BN \) is a solution of \( P \), denoted as \( f \models P \), if \( f \models G \) and for all \( c \in C \), \( c(f) = 1 \). The set of all solutions of \( P \) is \( F_P \).
Definition 15. Given a network-constraint problem \( P = (G, C) \), we consider two possible queries over the set of solutions of \( P \):

a. k-search queries, denoted search\(_k\)(\( P \)): find any \( F_0 \subseteq BN \) such that \( \forall f \in F_0: f \models P \) and \( |f| = \min(k, |F_0|) \),

b. All-search queries, denoted search(\( P \)): find all \( f \) such that \( f \models P \).

A query \( Q \) over a set of solutions of \( P \) is denoted as \( Q \). The class of queries is symbolized by \( \mathcal{Q} \). The set of solutions satisfying a query \( Q \) is denoted as \( F_Q \). We call Griffin queries the queries implemented in Griffin, a subset of the class \( \mathcal{Q} \) (see http://turing. iimas.unam.mx/griffin/guide.html#examplesfile).

Definition 16. Given a query \( Q(P) \) with \( P = (G, C) \), a query splitting is a function \( s: Q \rightarrow \mathcal{P}(Q) \) such that

a. \( \forall q(g, C) \in s(Q): f \models g \rightarrow f \models G \) and

b. \( f \models G \rightarrow \exists q(g, C) \in s(Q): f \models g \).

AUTHOR CONTRIBUTIONS

DR conceived the project. DR, EA, and SM designed the research. SM developed Griffin and performed the research. EA designed and directed the case studies. MC, DR, and SM developed the formal definitions and theoretical results. SM and MC prepared tables and figures. All authors wrote the manuscript.

FUNDING

We are grateful to Conacyt grant 221341 for support.

ACKNOWLEDGMENTS

We gratefully acknowledge the facilities provided by IIMAS, UNAM. We should like to thank Wassim Abou-Jaoudé, François Fages, Mariana Esther Martínez-Sánchez, Sylvain Soliman, Denis Thieffry, and Nathan Weinstein for fruitful discussions. We greatly appreciate the feedback provided by all participants of the course Griffin, una herramienta de model checking para el análisis de redes booleanas. Finally, we are very grateful to the reviewers for their valuable observations.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2018.00039/full#supplementary-material

REFERENCES


**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Muñoz, Carrillo, Azpeitia and Rosenblueth. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.