Static Analysis of Boolean Networks Based on Interaction Graphs: A Survey

Loïc Paulevé
LUNAM Université, École Centrale de Nantes, IRCCyN UMR CNRS 6597
(Institut de Recherche en Communications et Cybernétique de Nantes)
1 rue de la Noé - B.P. 92101 - 44321 Nantes Cedex 3, France
loic.pauleve@irccyn.ec-nantes.fr

Adrien Richard
I3S, UMR 6070 CNRS & Université de Nice-Sophia Antipolis,
richard@i3s.unice.fr

Abstract

Boolean networks are discrete dynamical systems extensively used to model biological regulatory networks. The dynamical analysis of these networks suffers from the combinatorial explosion of the state space, which grows exponentially with the number $n$ of components. To face this problem, a classical approach consists in deducing from the interaction graph of the network, which only contains $n$ vertices, some information on the dynamics of the network. In this paper, we present results in this topic, mainly by focusing on the influence of positive and negative feedbacks.

Keywords: Boolean network, interaction graph, discrete dynamical system, regulatory network

1 Introduction

Biological regulatory networks model the influence of biological components between each other. While such relations can be deduced through experiments by biologists, such systems can raise complex global dynamical behaviours and motivate the use of formal verification techniques to validate a model, and predict and understand particular behaviours, such as proteins productions dynamics [9,8].

Boolean Networks (BNs) are a typical formalism used to model biological regulatory networks dynamics (see the seminal work of Stuart Kauffman [19,20] and René Thomas [46,48]). A BN associates to each component a Boolean value and a local logical function describing its evolution in front of the value of other compon-
ents of the network. The global dynamics is then a Boolean map applying the local functions on component values, with different settings of iteration schemes.

Because the size of the state space is exponential with the number of components, classical analysis, such as model-checking techniques, may be intractable with large systems. To cope with this combinatorial explosion, an amount of work establish relationships between the interaction graph of a BN and some properties on the dynamics. An interaction graph references the positive and negative influences between components into a signed directed graph having one vertex per component; and hence is a compact and static abstraction of BN dynamics.

This paper aims at giving an insight of the different results on the relationships between the interaction graph and the dynamical properties of BNs. These kinds of results are of particular interest in the modelling of regulatory networks: those interaction graphs are generally the most reliable information on the biological system, while few knowledge on their precise behaviours are available.

This paper is structured as follows. Sect. 2 presents the main definitions used in this paper. Sect. 3 details the major results extracting dynamical properties of a BN from the topology of its interaction graph. Finally, we discuss these state-of-the-art approaches, and mention other static analysis techniques for BNs in Sect. 4.

2 Definitions

Notations. Let $n$ be a positive integer, $[n] = \{1, \ldots, n\}$, and $\mathbb{B} = \{0, 1\}$. The $i$th component of a point $x \in \mathbb{B}^n$ is denoted $x_i$. If $I \subseteq [n]$, then $\overline{x}^I$ denotes the point $y \in \mathbb{B}^n$ such that $y_i = 1 - x_i$ if $i \in I$, and $y_i = x_i$ otherwise. $\overline{x}$ is an abbreviation for $\overline{x}^{[n]}$, and if $i \in [n]$, we write $\overline{x}^i$ instead of $\overline{x}^{(i)}$. For all $x, y \in \mathbb{B}^n$, we denote by $\Delta(x, y)$ the set of $i \in [n]$ such that $x_i \neq y_i$. Let $f : \mathbb{B}^n \to \mathbb{B}^n$. The $i$th component of $f(x)$ is denoted $f_i(x)$. Hence, $f$, may be seen as a function from $\mathbb{B}^n$ to $\mathbb{B}$. For all $x \in \mathbb{B}^n$, we set $\Delta f(x) = \Delta(x, f(x))$. Observe that $f(x) = \overline{x}^{\Delta f(x)}$.

2.1 Boolean networks

A Boolean network is a discrete dynamical system that consists in a collection of $n$ Boolean components evolving along a discrete time by mutual interactions. Dynamics of such a system are usually described by a directed graph on $\mathbb{B}^n$ that is constructed from a Boolean function $f : \mathbb{B}^n \to \mathbb{B}^n$. Several iteration graphs can be considered. In the scope of this paper, we focus on the synchronous, asynchronous and generalized iteration graphs.

Definition 2.1 [Iteration graphs]
- The synchronous iteration graph of $f$, denoted by $\Sigma(f)$, is the directed graph on $\mathbb{B}^n$ that contains an arc $x \to y$ if and only if $y = f(x) \neq x$.
- The asynchronous iteration graph of $f$, denoted by $\Lambda(f)$, is the directed graph on $\mathbb{B}^n$ that contains an arc $x \to y$ if and only if there exists $i \in \Delta f(x)$ such that $y = \overline{x}^i$. 
The generalized iteration graph of $f$, denoted by $\text{GIG}(f)$, is the directed graph on $\mathbb{B}^n$ that contains an arc $x \to y$ if and only if there exists a non-empty set $I \subseteq \Delta f(x)$ such that $y = \overline{x}^I$.

On the one hand, Kauffman [19,20] proposed to model gene (regulatory) networks with synchronous iteration graphs. One the other hand, Thomas [46,48,49] argued that it is unlikely that several components change simultaneously their value, and he proposed to model gene networks with asynchronous iteration graphs. In constrast with the synchronous description, the asynchronous description is indeterministic: a state $x$ may have several successors (the number of successors of $x$ is precisely $|\Delta f(x)|$). An other interesting feature of the asynchronous description is that it can be seen has a relevant approximation of continuous descriptions based on piece-wise linear differential systems [12,11,40,48,42]. Note that the generalized iteration graph contains, as subgraphs, both the synchronous and asynchronous iteration graphs.

In these three cases, the interpretation of $f : \mathbb{B}^n \to \mathbb{B}^n$ as a gene network is the same: $[n]$ corresponds to a set of genes, and at state $x \in \mathbb{B}^n$: the protein encoded by the $i$th gene is present if $x_i = 1$ and absent otherwise, and the $i$th gene is expressed (“on”) if $f_i(x) = 1$ and unexpressed (“off”) otherwise. Hence, the gene expressions depend on the “proteins concentration”, and the evolution of these concentrations depends on the gene expressions.

Example 2.2 [Running example] $n = 3$ and $f$ is defined by:

$$f_1(x) = x_3 \land (\overline{x_1} \lor x_2), \quad f_2(x) = x_3, \quad f_3(x) = x_1 \lor x_2 \lor x_3.$$ 

The synchronous, asynchronous and generalized iteration graphs of $f$ are:

\begin{figure}
\centering
\includegraphics[width=\textwidth]{iteration_graphs.png}
\end{figure}

2.2 Fixed points and attractors

A point $x \in \mathbb{B}^n$ is a fixed point of $f$ if $x = f(x)$. Fixed points are of particular interest, since they correspond to stable states (in each iteration graph, $x$ is a fixed point of $f$ if and only if it has no successor). In the context of gene networks, fixed points correspond to stable patterns of gene expressions and are often associated to particular biological functions. In the next definition, the notion of fixed point is extended to the one of attractor.
Definition 2.3 [Attractors] Let $\Gamma$ be the synchronous, asynchronous or generalized iteration graph of $f$. The attractors of $\Gamma$ are the smallest non empty subsets $A \subseteq \mathbb{B}^n$ without leaving arc, i.e. such that for every arc $x \rightarrow y$ in $\Gamma$, if $x \in A$ then $y \in A$. An attractor of cardinality at least two is said cyclic.

It derives that a cyclic attractor does not contain any fixed point. As a consequence, once a system is inside a cyclic attractor, it cannot reach a fixed point, and it necessarily describes sustained oscillations. Other easy observations on attractors follow.

Proposition 2.4 A point $x$ is a fixed point of $f$ if and only if $\{x\}$ is an attractor of $\Gamma$. The non-cyclic attractors of $\Gamma$ can thus be identified to the fixed points of $f$. Also, for every $x$ in $\mathbb{B}^n$, there exists at least one path starting from $x$ that reaches an attractor. Consequently, $\Gamma$ has always at least one attractor.

Example 2.5 [Running example continued] The attractors of $\text{AIG}(f)$ and $\text{GIG}(f)$ are the fixed point 000 and the cyclic attractor $\{001, 011, 101, 111\}$. The attractors of $\text{SIG}(f)$ are the fixed point 000 and the cyclic attractor $\{011, 101, 111\}$.

2.3 Interaction graphs

Inherent to the notion of Boolean network is the one of interaction graph. Such a graph depicts, in a very coarse way, the qualitative interactions between components, and is usually represented as a directed graph with $[n]$ as vertex set. An arc from $j$ to $i$ then means that the evolution of the $i$th component depends on the evolution of the $j$th one. Here, we consider two kinds of interactions: the positive and negative ones. This hence leads us to consider signed directed graphs.

Definition 2.6 [Signed directed graph] A signed directed graph on $[n]$ is a graph $G$ whose the set of vertices is $[n]$, and whose the set of arcs is a subset of $[n] \times \{+, -\} \times [n]$. If $(j, s, i)$ is an arc of $G$, we say that $G$ has an arc from $j$ to $i$ of sign $s$. $G$ is simple if, for every $i, j \in [n]$, there exists at most one arc from $j$ to $i$. A positive (resp. negative) cycle of $G$ is an elementary directed cycle that contains an even (resp. odd) number of negative arcs. The length of a cycle is the number of arcs it involves.

We are now in position to define, in a natural way, the interaction graph of a network whose the dynamics are described from $f : \mathbb{B}^n \rightarrow \mathbb{B}^n$.

Definition 2.7 [Interaction graph of $f$] The interaction graph of $f$, denoted by $G(f)$, is the signed directed graph on $[n]$ defined by: for all $i, j \in [n]$, there exists a positive (resp. negative) arc from $j$ to $i$ if and only if there exists $x \in \mathbb{B}^n$ with $x_j = 0$ such that $f_i(x) < f_i(\overline{x})$ (resp. $f_i(x) > f_i(\overline{x})$).

In other words, there exists a positive (resp. negative) arc from $j$ to $i$ if, in at least one state $x$, an increase of $x_j$ induces an increase (resp. decrease) of $f_i(x)$, i.e. the appearance of the $j$th protein turns on (resp. off) the $i$th gene.
Example 2.8 [Running example continued] The interaction graph of $f$ follows. It has a negative cycle of length one, and a positive cycle of length one, two and three. Note also that it is simple.

\[
\begin{array}{c}
1 \\
\rightarrow \\
2 \\
\rightarrow \\
3 \\
\end{array}
\]

3 Static analysis based on the interaction graph

3.1 Absence of cycle and convergence toward a unique fixed point

The following fundamental fixed point theorem of Robert [34,35] shows that the presence of a cycle in the interaction graph is necessary to prevent a simple synchronous convergence toward a unique fixed point.

**Theorem 3.1** If $G(f)$ has no cycle, then $f$ has a unique fixed point $x$, and every path of $\Sigma G(f)$ reaches this fixed point (in at most $n$ steps).

This convergence result was latter proved for asynchronous iteration graphs by Robert [36], and for generalized iteration graphs by Bahi and Michel [5] (actually, these authors proved the convergence for a much more general class of iteration schemes that include delays of communication between components).

**Theorem 3.2** If $G(f)$ has no cycle, then $f$ has a unique fixed point, and every path of $GIG(f)$ reaches this fixed point.

3.2 Absence of positive cycle and upper bound on the number of fixed points

In 1980, Thomas stated the following general rule: a necessary condition for a dynamical system to admit several stable states is the presence of a positive cycle in its interaction graph [47]. In the context of gene networks, one can interpret this rule as follow: positive cycles are necessary for differentiation processes [48]. This rule was first formally stated and proved in continuous frameworks [27,41,17,6,43,44]. A Boolean version of Thomas’ rule have then been stated and proved by Remy, Ruet and Thieffy [28].

**Theorem 3.3** If $G(f)$ has no positive cycle, then $f$ has at most one fixed point.

The following theorem, established in [33], shows that positive cycles are more generally necessary for the presence of several asynchronous attractors.

**Theorem 3.4** If $G(f)$ has no positive cycle, then $AIG(f)$ has at most one attractor.

This theorem is not valid in the synchronous case. Also, since $AIG(f)$ has always at least one attractor (cf. Proposition 2.4), the theorem remains valid if “at most one” is replaced by “a unique”. Finally, using it as a base case, one can obtain an
upper bound on the number of asynchronous attractors that only depends on the structure of connections between positive cycles \[3,2,30\].

**Theorem 3.5** Let \( I \) be a subset of \([n]\). If every positive cycle of \( G(f) \) has a vertex in \( I \), then \( \text{aig}(f) \) has at most \( 2^{|I|} \) attractors.

If \( G(f) \) has no positive cycle, then \( I = \emptyset \) satisfies the conditions of the statement, and the resulting upper bound is one (hence, Theorem 3.5 generalizes Theorem 3.4). The bound shows that it is not the number of positive cycles which is important to obtain several fixed points (or asynchronous attractors), but much more the structure of the connections between positive cycles. For instance, if \( G(f) \) has a thousand of positive cycles, and if all these cycles share a same vertex \( i \), then \( I = \{i\} \) satisfies the conditions of the statement, and the resulting upper bound is only two. In other words, in order for \( f \) to have a lot of fixed points, \( G(f) \) has to contain a lot of “rather disconnected” positive cycles.

We have seen that if \( G(f) \) has no positive cycle, then \( f \) has at most one fixed point. The following theorem of Aracena \[3,2\] shows that, under a very weak additional condition, \( f \) has actually no fixed point.

**Theorem 3.6** If \( G(f) \) has a minimal in-degree at least one and has no positive cycle, then \( f \) has no fixed point.

Aracena proved this theorem under the additional assumption that \( G(f) \) is simple, but this condition can be easily removed. Now, observe that this theorem implies quite directly Theorem 3.3. Indeed, suppose that \( G(f) \) has no positive cycle. If \( G(f) \) has a minimal in-degree at least one and has no positive cycle, then by Theorem 3.6, \( f \) has no fixed point. Otherwise, there exists a vertex, say \( n \) without loss generality, of in-degree zero in \( G(f) \). So \( f_n = \text{cst} = c \). Let \( h : B^{n-1} \rightarrow B^{n-1} \) defined by \( h_i(x) = f_i(x, c) \) for all \( x \in B^{n-1} \) and \( i \in [n-1] \). It is easy to see that \( G(h) \) is a subgraph of \( G(f) \). So \( G(h) \) has no positive cycle, and proceeding by induction on the number of components, one obtains that \( h \) has at most one fixed point. Since it is clear that \( h \) and \( f \) have the same number of fixed points, we deduce that \( f \) has at most one fixed point.

### 3.3 Absence of negative cycle and lower bound on the number of fixed points

Thomas also stated a general rule on negative cycles: a necessary condition for a dynamical system to produce sustained oscillations is the presence of a negative cycle in its interaction graph \[47\]. In the context of gene networks, one can interpret this rule as follow: negative cycles are necessary for homoeostasis phenomena \[48\]. The next theorem, proved in \[31\], may be seen as a Boolean version of second Thomas’ rule (and it is not valid in the synchronous case).

**Theorem 3.7** If \( G(f) \) has no negative cycle, then \( \text{aig}(f) \) has no cyclic attractor. Hence, if \( G(f) \) has no negative cycle, then \( f \) has at least one fixed point.

The second assertion is a consequence of the first one, since if \( G(f) \) has no negative cycle, then \( \text{aig}(f) \) has no cyclic attractor, and since \( \text{aig}(f) \) has always
at least one attractor, we deduce that $AIG(f)$ has an attractor of size one, which is a fixed point of $f$ (cf. Proposition 2.4). In [16], Goles and Salinas provide an algorithm with running time $O(n^2)$ to compute, under the absence of negative cycle in $G(f)$, a fixed point of $f$ (in the general case, the problem of finding a fixed point is NP-complete).

Note that Theorems 3.3 and 3.7 provides a nice “proof by dichotomy” of (the first assertion in) the theorem of Robert (Theorem 3.1): if $G(f)$ has no cycle, then following Theorem 3.3 (resp. 3.7), $f$ has at most (resp. at least) one fixed point, so $f$ has indeed a unique fixed point.

While Theorem 3.6 provides a sufficient condition for the absence of fixed point in term of absence of positive cycle, the following theorem, also established by Aracena [3, 2], provides a sufficient condition for the presence of multiple fixed points in term of absence of negative cycle.

**Theorem 3.8** Let $G(f)^*$ be the signed directed graph on $[n]$ that contains an arc from $j$ to $i$ of sign $s$ if $G(f)$ has an arc of sign $s$ from $j$ to $i$ or from $i$ to $j$. If $G(f)$ is simple, has a minimal in-degree at least one, and if $G(f)^*$ has no negative cycle, then there exists $x \in \mathbb{B}^n$ such that $x$ and $\overline{x}$ are fixed points of $f$.

It is easy to see that if $G(f)$ is strongly connected and has no negative cycle, then $G(f)$ is simple, has a minimal in-degree at least one, and $G(f)^*$ has no negative cycle. Hence, the next theorem is a consequence of the previous one.

**Theorem 3.9** If $G(f)$ is strongly connected and has no negative cycle, then $f$ has at least two fixed points.

The following theorem, established in [26], generalizes Theorem 3.8 by characterizing the number of topological fixed points of $f$, i.e. the number of fixed points of $f$ that only depend on $G(f)$.

**Theorem 3.10** Let $G(f)^\#$ be the signed directed graph on $[n]$ that contains an arc from $j$ to $i$ of sign $s$ if and only if $G(f)$ has an arc from $j$ to $i$ of sign $s$ and no arc from $j$ to $i$ of sign $-s$. Let $G(f)^*$ be as in Theorem 3.8. Let us say that $x$ is a topological fixed point of $f$ if it is a fixed point of every function $h : \mathbb{B}^n \to \mathbb{B}^n$ such that $G(h) = G(f)$. We have the following two properties:

- Let $p$ be the number of connected components in $G(f)^\#$. If the following three conditions are satisfied then $f$ has exactly $2^p$ topological fixed points, and otherwise, $f$ has 0 topological fixed point: (i) the minimal in-degree of $G(f)^\#$ is at least one; (ii) $G(f)^*$ has no negative cycle; (iii) for every $i \in [n]$, there exists at most one $j \in [n]$ such that $G(f)$ has both a positive and a negative arc from $j$ to $i$.
- If $x$ is a topological fixed point of $f$, then $x$ is a topological fixed point of $f$.

Note that if $G(f)$ is simple, then $G(f)^\# = G(f)$ and the third condition is trivially satisfied. We then recover Theorem 3.8.
3.4 Comparison of iteration graphs

As said above, there are two principal iteration graphs used to model BNs dynamics, the synchronous and asynchronous ones. It is then of great interest to try to compare them. There are very few results in this topic. However, Noual [23] recently established the following fundamental theorem, which shows that under some conditions on the cycles of \( G(f) \), every path of the generalized iteration graph, and in particular every path of the synchronous iteration graph, can be “imitated” by at least one path of the asynchronous iteration graph. Other results concerning comparison of iteration schemes are provided in [45,36,15,4].

**Theorem 3.11** Assume that \( G(f) \) is simple, has no positive cycle of even length, and has no negative cycle of odd length. If \( \text{GIG}(f) \) has an arc \( x \rightarrow y \), then \( \text{AIG}(f) \) has a path from \( x \) to \( y \) of length \( |\Delta(x,y)| \). Hence, the number of attractors in \( \text{GIG}(f) \) and \( \text{SIG}(f) \) is at least the number of attractors in \( \text{AIG}(f) \).

3.5 Network reduction

Since the number of possible states of a BN increases exponentially with the number \( n \) of components, it is interesting to try to reduce the system, typically by removing components, without losing too many dynamical properties. The following theorem, stated with other notations in [21], shows that if the \( n \)th component does not interact with itself, it can be suppressed in a natural way, without affecting the number of fixed points, the reachability by asynchronous paths, and the main features of the interaction graph. (For sake of simplicity, this theorem only deals with the suppression of the \( n \)th component, but it can be very easily adapted to the suppression of any component.)

**Theorem 3.12** Let \( \tilde{f} : B^{n-1} \rightarrow B^{n-1} \) be defined from \( f : B^n \rightarrow B^n \) by:

\[
\forall x \in B^{n-1}, \forall i \in [n-1], \quad \tilde{f}_i(x) = f_i(\tilde{x}), \quad \tilde{x} = (x, f_n(x,0)) \in B^n.
\]

If \( G(f) \) has no arc from \( n \) to itself, then the following three properties hold:

- A point \( x \) is a fixed point of \( \tilde{f} \) if and only if \( \tilde{x} \) is a fixed point of \( f \). As a consequence, \( \tilde{f} \) and \( f \) have the same number of fixed points.

- \( \text{AIG}(\tilde{f}) \) has a path from \( x \) to \( y \) if and only if \( \text{AIG}(f) \) has a path from \( \tilde{x} \) to \( \tilde{y} \). As a consequence, the number of attractors in \( \text{AIG}(\tilde{f}) \) is at least the number of attractors in \( \text{AIG}(f) \).

- If \( G(\tilde{f}) \) has a path from \( j \) to \( i \) with an even (resp. odd) number of negative arcs, then \( G(f) \) has a path from \( j \) to \( i \) with an even (resp. odd) number of negative arcs. As a consequence, if \( G(f) \) has no positive (resp. negative) cycle, then \( G(\tilde{f}) \) has no positive (resp. negative) cycle.

Note that this theorem gives another way to prove the (first assertion in the) theorem of Robert (Theorem 3.1). Indeed, suppose that \( G(f) \) has no cycle. Then it has no arc from \( n \) to itself, so \( \tilde{f} \) and \( f \) have the same number of fixed points,
and $G(\tilde{f})$ has no cycle (according to the third point). Using an induction on the number of components, we deduce that $\tilde{f}$ has a unique fixed point, so that $f$ has a unique fixed point.

4 Discussion

We have presented several results related to the static analysis of BNs through the topological analysis of their interaction graphs. First, we have seen that cycles are necessary to obtain “complex behaviours” (Theorems 3.1 and 3.2). Then, by focusing on positive and negative cycles, we have seen that necessary or sufficient conditions for the presence of several attractors or the presence of cyclic attractors can be obtained (Theorems 3.3-3.10). We have also seen that conditions on cycles permit to obtain results about the comparison of iteration schemes or the model reduction (Theorems 3.11 and 3.12).

For the sake of simplicity, we restricted ourself to the Boolean case and results concerning the global interaction graph. However, Theorems 3.1-3.5, 3.7 and 3.12 have extensions to the non-Boolean discrete case and/or generalizations in terms of local interaction graphs [34,36,5,38,37,28,33,29,30,31,32].

All the topics regarding the static analysis based on interaction graphs have not been covered in this short survey. In particular, a lot of works have been dedicated to the length of synchronous attractors [14,45,13,7,10,18,15].

With the aim at making tractable the computation of more precise dynamical properties from BNs, recent work investigate the use of other static analysis techniques. For instance, Naldi et al. [22] propose to represent each function $g_i$ defined by $g_i(x) = 1$ iff $f_i(x) = x_i$ with a binary decision diagram; leading to an efficient enumeration of all fixed points of $f$ through algebraic operations on the $n$ resulting decision diagrams. Paulevé et al. developed the Process Hitting framework into which over-approximations of asynchronous dynamics of BNs can be encoded. By using abstract interpretation techniques, very efficient over- and under-approximations of particular reachability properties within Process Hitting models have been built [24,25], having hence applications to reachability (static) analysis within BNs.

Overall, a large variety of properties can be statically derived from BNs. The extension of the presented topological analysis of the interaction graph to other modelling frameworks for biological regulatory networks is yet challenging. While, as mention above, several important results have already been extended to discrete networks, the new degree of freedom brought by discrete modelling prevents the straightforward extension of many results on BNs without imposing constraints on possible logical expressions (such as Theorems 3.6 and 3.8). Finally, the efficient quantification of dynamical properties within hybrid models extending discrete networks with time or stochastic features [39,1,50] calls for innovative static analysis techniques.
Acknowledgement

This work has been partially supported by the French National Agency for Research (ANR-10-BLANC-0218 BioTempo project).

References


Paulevé, L., M. Magnin and O. Roux, Static Analysis by Abstract Interpretation of Biological Regulatory Networks Dynamics, Research Report hal-00574353, IRCCyN (2011). URL http://hal.archives-ouvertes.fr/hal-00574353


