



Analysis Tools for Interconnected Boolean Networks With Biological Applications

Madalena Chaves, Laurent Tournier

► To cite this version:

Madalena Chaves, Laurent Tournier. Analysis Tools for Interconnected Boolean Networks With Biological Applications. *Frontiers in Physiology*, 2018, 9, pp.1-18. 10.3389/fphys.2018.00586 . hal-01936199

HAL Id: hal-01936199

<https://hal.science/hal-01936199>

Submitted on 27 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



Analysis Tools for Interconnected Boolean Networks With Biological Applications

Madalena Chaves^{1*} and Laurent Tournier²

¹ Inria Sophia Antipolis - Méditerranée, Université Côte d'Azur, Valbonne, France, ² MaIAGE, INRA, Université Paris-Saclay, Jouy-en-Josas, France

OPEN ACCESS

Edited by:

Matteo Barberis,
University of Amsterdam, Netherlands

Reviewed by:

Laurence Calzone,
Institut Curie, France
Olaf Wolkenhauer,
University of Rostock, Germany
Aurélien Naldi,
École Normale Supérieure, France

*Correspondence:

Madalena Chaves
madalena.chaves@inria.fr

Specialty section:

This article was submitted to
Systems Biology,
a section of the journal
Frontiers in Physiology

Received: 01 February 2018

Accepted: 02 May 2018

Published: 29 May 2018

Citation:

Chaves M and Tournier L (2018)
Analysis Tools for Interconnected
Boolean Networks With Biological
Applications. *Front. Physiol.* 9:586.
doi: 10.3389/fphys.2018.00586

Boolean networks with asynchronous updates are a class of logical models particularly well adapted to describe the dynamics of biological networks with uncertain measures. The state space of these models can be described by an asynchronous state transition graph, which represents all the possible exits from every single state, and gives a global image of all the possible trajectories of the system. In addition, the asynchronous state transition graph can be associated with an absorbing Markov chain, further providing a semi-quantitative framework where it becomes possible to compute probabilities for the different trajectories. For large networks, however, such direct analyses become computationally untractable, given the exponential dimension of the graph. Exploiting the general modularity of biological systems, we have introduced the novel concept of *asymptotic graph*, computed as an interconnection of several asynchronous transition graphs and recovering all asymptotic behaviors of a large interconnected system from the behavior of its smaller modules. From a modeling point of view, the interconnection of networks is very useful to address for instance the interplay between known biological modules and to test different hypotheses on the nature of their mutual regulatory links. This paper develops two new features of this general methodology: a quantitative dimension is added to the asymptotic graph, through the computation of relative probabilities for each final attractor and a companion *cross-graph* is introduced to complement the method on a theoretical point of view.

Keywords: asynchronous Boolean networks, module interconnection, state transition graph, attractor computation, biological regulatory networks

1. INTRODUCTION

An intuitive representation of system interactions, an algorithmic description of state transitions, and the capacity to capture the global dynamics of the system, list some of the advantages of Boolean models, which remain a powerful tool in the modeling and analysis of biological networks (Wang et al., 2012; Abou-Jaoudé et al., 2016). Successfully predictive examples of Boolean models cover complex networks across many different organisms, from cell cycle (Li et al., 2004; Fauré et al., 2006), to fly or plant morphogenesis (Albert and Othmer, 2003; García-Gómez et al., 2017), and highly complex networks such as T-cell induction (Mendoza and Xenarios, 2006; Saez-Rodriguez et al., 2007), leukemia (Zhang et al., 2008) or apoptosis (Calzone et al., 2010).

In a modular view of a biological organism, each task is executed by a specific set of interactions among an ensemble of biological components; in other words, it can be said that there is a specific network, or module, for each specific task (signaling, metabolic, physiological, etc.). These modules often interact with each other, one task triggering the next in a chain of events or cyclic phenomena. Examples include chains of signaling networks such as MAPK cascades, genetic-metabolic interactions (Baldazzi et al., 2010), or coupled oscillations (Gérard and Goldbeter, 2012). However, in many cases, while experimental evidence supports the existence of links between two modules, their modes of interaction are still unclear (as in the case of mammalian cell cycle and circadian clock, see Feillet et al., 2015). In this context, mathematical tools are necessary to facilitate the analysis of the complex behavior obtained from the interconnection of two or more known modules.

One of the challenges in the analysis of Boolean networks is attractor computation, particularly for high-dimensional networks. For a network of dimension n , the size of the state transition graph is 2^n . A direct analysis of such a graph may become computationally costly, in terms of space and time, when $n \geq 20$. This is especially true with asynchronous updating, which includes numerous dynamical trajectories. Two very efficient methods have recently been developed: Zañudo and Albert (2013) compute all attractors of a network (up to $n \approx 100$), by isolating special properties of the state transition graph's components; Veliz-Cuba et al. (2014) compute all singletons (attractors containing a single state) for networks up to $n = 1,000$, by using a computational algebra approach.

In this paper, we propose a methodology aimed specifically at analyzing the interconnection between several known Boolean modules. The interconnection between two biological networks can be very hard to test *in vivo*: our methodology provides a platform for hypothesis testing, confirming or disproving assumptions regarding mutual regulatory effects, simulating and comparing various forms of interconnection schemes and corresponding emergent dynamical behavior. Our method relies on the construction of a new object, the *asymptotic graph*, introduced by Tournier and Chaves (2013), which is a directed graph constructed only from the set of attractors of each module and that captures all the asymptotic behaviors of the interconnected network.

After a brief review of Boolean network interconnections, two improvements to the asymptotic graph are introduced in this paper, to mitigate two of its known limitations. First, it was observed that the asymptotic graph may also recover spurious attractors, in addition to the true attractors of the full network (Tournier and Chaves, 2013); we introduce an extension, called the *cross graph* that solves this issue from a theoretical point of view. The cross graph is constructed from the set of strongly connected components of each separate module, while the asymptotic graph is constructed from *terminal* strongly connected components only. Second, to enrich the traditional ON/OFF representation inherent to Boolean models, we propose a method to assign probabilities to the edges of the asymptotic graph, thereby allowing a probabilistic representation

of the various possible trajectories of the composed network. Our methodology is applied first to a class of general randomly generated Boolean models and then to two state-of-the-art biological models in two different organisms: (i) to explore the interplay between mammalian cell cycle and circadian clock oscillators and (ii) to test hypotheses on the regulatory links between budding yeast cell cycle and cell size, where our analysis suggests that the START signal should come from mitosis phase.

2. INTERCONNECTIONS OF ASYNCHRONOUS BOOLEAN NETWORKS: A SHORT REVIEW

Throughout this paper, we will consider Boolean networks under asynchronous updates. An interconnected Boolean network is, briefly, the combined network formed by linking together, in an appropriately prescribed way, two or more separate Boolean modules. In previous works (Chaves and Tournier, 2011; Tournier and Chaves, 2013) we have introduced a new object, the *asymptotic graph*, that characterizes the attractors of the combined Boolean network in terms only of the attractors of the separate modules—hence with no need to compute the larger state transition graph. In the following, the definition of the main objects needed to introduce the asymptotic graph are briefly reviewed.

2.1. IO Asynchronous Boolean Networks and Their Interconnections

Let us start by a brief recall of the definition of an input-output asynchronous Boolean network (IO ABN), reprising the notation introduced by Tournier and Chaves (2013). An IO ABN Σ^A is characterized by three integers n_A, p_A, q_A ($n_A > 0$ is the dimension of the system, $p_A, q_A \geq 0$ are respectively the numbers of inputs and outputs) and by two Boolean maps: $f^A: \{0, 1\}^{p_A} \times \{0, 1\}^{n_A} \rightarrow \{0, 1\}^{n_A}$ (the transition function) and $h^A: \{0, 1\}^{n_A} \rightarrow \{0, 1\}^{q_A}$ (the output function). For any given input profile $u \in \{0, 1\}^{p_A}$, the asynchronous dynamics of the network are given by the *asynchronous transition graph* $G^{A,u}$, which is a digraph over the vertex set $\{0, 1\}^{n_A}$ defined as follows: for any state $x = (x_1, \dots, x_{n_A}) \in \{0, 1\}^{n_A}$, the set of its successors are the states $(x_1, \dots, \neg x_i, \dots, x_{n_A})$, for all $i \in \{1, \dots, n\}$ such that $f_i^A(u, x) \neq x_i$. The number of vertices of such a graph is 2^{n_A} and its number of arcs, denoted by m_A , verifies $0 \leq m_A \leq n_A 2^{n_A}$. It is therefore relatively sparse and can thus be efficiently stored by a $2^{n_A} \times 2^{n_A}$ adjacency matrix. In the following, we will consider that $G^{A,u}$ designates this matrix. Given two integers $i, j \in \{1, \dots, 2^{n_A}\}$, the (i, j) entry of the adjacency matrix equals 1 if state j is a successor of state i and 0 otherwise. In a classical abuse of notation, we associate each integer $i \in \{1, \dots, 2^{n_A}\}$ with its binary representation $x \in \{0, 1\}^{n_A}$ in lexicographic order, with the left-most bit being the most significant one; in other words: $i - 1 = \sum_{k=1}^{n_A} x_k 2^{n_A-k}$. Thus, we will indifferently call *state* either an integer $i \in \{1, \dots, 2^{n_A}\}$ or its Boolean representation $x \in \{0, 1\}^{n_A}$.

EXAMPLE 1. Consider the bidimensional single-input, single-output (SISO) network defined by: $f^A(u, x_1, x_2) = (u, x_1)$ and

$h^A(x_1, x_2) = x_2$. Graphically, this network can be represented as a simple cascade $u \rightarrow x_1 \rightarrow x_2$. Its dynamics are characterized by the two graphs $G^{A,0}$ and $G^{A,1}$, represented below in graphical and matricial forms:

$$G^{A,0}: \begin{array}{ccc} 10 & \rightarrow & 11 \\ \downarrow & & \downarrow \\ 00 & \leftarrow & 01 \end{array} \quad \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix},$$

$$G^{A,1}: \begin{array}{ccc} 01 & \rightarrow & 00 \\ \downarrow & & \downarrow \\ 11 & \leftarrow & 10 \end{array} \quad \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

In adjacency matrices, by convention the (i, j) entry equals 1 iff state j is a successor of state i . Here, the four states (rows and columns of the matrix) are intended in the following order: 00, 01, 10, 11. In $G^{A,0}$, state 00 does not have any successor, implying the first row of its adjacency matrix is zero: 00 is a steady state of the network. Similarly, 11 is a steady state of $G^{A,1}$. \square

Classically, an asynchronous transition graph $G^{A,u}$ is analyzed by first computing its decomposition into strongly connected components (SCCs), denoted by $A_u^1, \dots, A_u^{N_u^A}$, where $1 \leq N_u^A \leq 2^n$. The set of all SCCs forms a partition of the state space $\{0, 1\}^{n_A}$ and their computation can be efficiently achieved in $O(2^{n_A} + m_A)$. By contracting each SCC to a single vertex, a directed acyclic graph (dag) is constructed, sometimes called *condensation* graph or simply SCC graph. This dag provides a useful description of key dynamical behaviors of the network; in particular terminal SCCs (the leafs of the dag) correspond to the attractors of the network. More details about these graph theoretical tools can be found, for instance, in the textbook by Cormen et al. (2001).

Consider now two IO ABN Σ^A and Σ^B , of respective dimensions (n_A, p_A, q_A) and (n_B, p_B, q_B) and state variables $x \in \{0, 1\}^{n_A}$ and $y \in \{0, 1\}^{n_B}$. Note that all the methods presented in this paper generalize to more than two modules; however, in order to maintain a clear exposition of the results, the definitions are given for interconnections of two modules. An *interconnection scheme* of Σ^A and Σ^B consists in two interconnecting functions $\mu_A: \{0, 1\}^{q_B} \rightarrow \{0, 1\}^{p_A}$ and $\mu_B: \{0, 1\}^{q_A} \rightarrow \{0, 1\}^{p_B}$ mapping the outputs of each module to the inputs of the other module. For convenience, throughout this paper we will make the assumption that $q_B = p_A$ and $q_A = p_B$ and that the interconnecting functions are simply identity maps. Following Tournier and Chaves (2013), with this assumption the resulting interconnected network is the ABN of dimension $n_A + n_B$, with no input and no output, defined by the following transition function:

$$f: \{0, 1\}^{n_A} \times \{0, 1\}^{n_B} \longrightarrow \{0, 1\}^{n_A} \times \{0, 1\}^{n_B} \quad (1)$$

$$(x, y) \longmapsto (f^A(h^B(y), x), f^B(h^A(x), y)).$$

One can then consider the interconnection as a standalone network: its transition graph G can be constructed from this transition function f . Alternatively, one can also build the graph G directly from the set of transition graphs $G^{A,u}$, $u \in \{0, 1\}^{p_A}$ and

$G^{B,v}$, $v \in \{0, 1\}^{p_B}$ as follows. Let (x, y) and (x', y') be two Boolean vectors in $\{0, 1\}^{n_A} \times \{0, 1\}^{n_B}$, then (x', y') is a (asynchronous) successor of (x, y) if

- either $x = x'$ and y' is a successor of y in $G^{B,h^A(x)}$,
- or $y = y'$ and x' is a successor of x in $G^{A,h^B(y)}$.

It is possible to summarize this definition in a simple matricial form. First, for each $\alpha \in \{0, 1\}^{q_A}$, introduce the $2^{n_A} \times 2^{n_A}$ diagonal Boolean matrix $\Delta^{A,\alpha}$ such that $[\Delta^{A,\alpha}]_{ii} = 1$ if the output of state i is equal to α and 0 otherwise. Similarly, for module Σ^B introduce the $2^{n_B} \times 2^{n_B}$ diagonal Boolean matrices $\Delta^{B,\beta}$, with $\beta \in \{0, 1\}^{q_B}$. Then, G can be reconstructed by the formula:

$$G := \bigvee_{(\alpha, \beta) \in \{0, 1\}^{q_A} \times \{0, 1\}^{q_B}} (G^{A,\beta} \otimes \Delta^{B,\beta} \vee \Delta^{A,\alpha} \otimes G^{B,\alpha}), \quad (2)$$

where \otimes designates the classical Kronecker product. By replacing matrices Δ with identity matrices, one may recognize in this definition of G the notion of Cartesian product of graphs, first introduced by Sabidussi (1959). To be more precise, (2) generalizes the notion of Cartesian product to interconnections, by including only transitions that are consistent with the input-output scheme.

EXAMPLE 2. Consider module Σ^A defined in Example 1 and let the one-dimensional SISO module Σ^B defined by $f^B(v, y_1) = \neg v$ and $h^B(y_1) = y_1$. Its dynamics are given by

$$G^{B,0} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \text{ and } G^{B,1} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}.$$

The interconnected network can be reconstructed by using (1), leading to the 3-dimensional transition function $f(x_1, x_2, y_1) = (y_1, x_1, \neg x_2)$. Alternatively, the transition graph G can also be computed directly as the interconnection of the dynamics of the two separated modules by using (2):

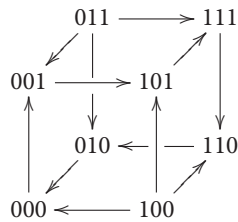
$$G = (G^{A,0} \otimes \Delta^{B,0}) \vee (G^{A,1} \otimes \Delta^{B,1}) \vee (\Delta^{A,0} \otimes G^{B,0}) \vee (\Delta^{A,1} \otimes G^{B,1}),$$

$$= \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix} \otimes \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \vee \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \vee$$

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \otimes \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \vee \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix},$$

$$= \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

In graphical form, this transition graph G of the interconnected network can be represented as:



This graph has a unique attractor, composed of six states: $\{001, 101, 111, 110, 010, 000\}$. \square

In the present paper, note that we assume the modules and the interconnection scheme are given. It is also possible to consider interconnections as a general *model reduction* technique, where a large network is first decomposed into *a priori* unknown sub-networks. The identification of an efficient decomposition, with the corresponding interconnecting scheme, would then become critical. This problem is related to the general problem of graph partitioning and is addressed elsewhere (Tournier and Chaves, 2013).

2.2. The Asymptotic Graph of an Interconnection

We can now give the definition of the asymptotic graph (Tournier and Chaves, 2013). First, list all the terminal SCCs of module Σ^A : $\{A_{u\alpha}^i, u \in \{0, 1\}^{p_A}, 1 \leq i \leq L_u^A\}$ and cut them with respect to their outputs, *ie.* define, for each output profile $\alpha \in \{0, 1\}^{q_A}$, the set $A_{u\alpha}^i := \{x \in A_{u\alpha}^i, h^A(x) = \alpha\}$. For some α such a set may be empty, in that case we will simply omit it. Similarly, define $\{B_{v\beta}^j, v \in \{0, 1\}^{p_B}, \alpha \in \{0, 1\}^{q_B}, 1 \leq j \leq L_v^B\}$ for module Σ^B . The *asymptotic graph* of the interconnection is then defined as the directed graph $G^{as} = (V^{as}, E^{as})$ such that the vertex set V^{as} is composed of all the cross products $A_{u\alpha}^i \times B_{v\beta}^j$ and the arc set E^{as} is constructed as follows:

- $A_{u\alpha}^i \times B_{v\beta}^j \rightarrow A_{u'\alpha'}^{i'} \times B_{v'\beta'}^{j'}$ iff there exist $x \in A_{u\alpha}^i, x' \in A_{u'\alpha'}^{i'}$ such that there exists a path from x to x' in $G^{A,\beta}$,
- $A_{u\alpha}^i \times B_{v\beta}^j \rightarrow A_{u\alpha}^i \times B_{v'\beta'}^{j'}$ iff there exist $y \in B_{v\beta}^j, y' \in B_{v'\beta'}^{j'}$ such that there exists a path from y to y' in $G^{B,\alpha}$.

Finally, introduce the function π as follows: if $V = A_{u\alpha}^i \times B_{v\beta}^j \in V^{as}$, $\pi(V) := \{(x, y), x \in A_{u\alpha}^i, y \in B_{v\beta}^j\}$ and if $R \subseteq V^{as}$, $\pi(R) := \bigcup_{V \in R} \pi(V)$. The interest of the asymptotic graph lies in the following theorem, a proof of which can be found in Tournier and Chaves (2013).

THEOREM 1. *If Q is an attractor of the interconnected network, then there exists a terminal SCC R of G^{as} such that $\pi(R) \subseteq Q$.*

EXAMPLE 3. Consider the interconnection of Example 2 above. The asymptotic graph is given by

$$\begin{array}{ccc} A_{00}^1 \times B_{01}^1 & \rightarrow & A_{11}^1 \times B_{01}^1 \\ \uparrow & & \downarrow \\ A_{00}^1 \times B_{10}^1 & \rightarrow & A_{11}^1 \times B_{10}^1 \end{array} \quad \text{with:} \quad \begin{cases} \pi(A_{00}^1 \times B_{01}^1) = \{001\}, \\ \pi(A_{11}^1 \times B_{01}^1) = \{111\}, \\ \pi(A_{11}^1 \times B_{10}^1) = \{000\}, \\ \pi(A_{00}^1 \times B_{10}^1) = \{110\}. \end{cases}$$

Therefore, G^{as} is composed of a single terminal SCC R , and $\pi(R) = \{001, 111, 000, 110\}$ is actually included into the (unique) attractor of the interconnected network. \square

Thanks to Theorem 1, the asymptotic graph is a powerful analytic tool as it recovers *all* the attractors of an interconnection (without missing any), by constructing a graph significantly smaller than the full interconnected graph G (section 4 below provides numerical results for random interconnections). However, it may happen that some terminal SCC of G^{as} does not correspond to an actual attractor of the interconnection. Such terminal SCCs, called *spurious* attractors, appear very rarely and there exist some sufficient conditions to detect *a priori* spurious attractors in certain cases. The most simple one, particularly useful for biological applications is the fact that when R is a singleton then it cannot be a spurious attractor. The proof, along with additional conditions are provided elsewhere (Tournier and Chaves, 2013; Chaves and Carta, 2015).

3. NEW ANALYSIS TOOLS

This section describes our new contributions. Our first goal is to improve the asymptotic graph construction to avoid the generation of spurious attractors (section 3.1) and our second goal is to update the asymptotic graph by adding quantitative information (probabilistic) on the state transitions (section 3.2).

3.1. A Theoretical Tool to Recover All the Dynamics of an Interconnection

The asymptotic graph of an interconnection is constructed only from the modules' attractors, generally implying a relatively manageable size allowing to analyze a wide range of practical examples of interconnections (see sections 4 and 5). Nevertheless, ignoring transient dynamical behaviors of the modules also implies two drawbacks for Theorem 1. First, spurious attractors may appear, although this phenomenon seems to be relatively rare as illustrated in section 4. Second, when a terminal SCC of G^{as} corresponds to an actual attractor, Theorem 1 only ensures an inclusion, meaning the predicted attractor may contain only a small proportion of states that are in the real attractor. We now propose a new graph, called the *cross-graph*, overcoming those two issues and ensuring, at the price of a higher computational cost, a one-to-one recovery of all the attractors of the interconnected network. Note that Tournier and Chaves (2013) already introduced a notion of cross-graph, however the cross-graph described in the following is significantly improved. In particular, its size is bounded by the size of the full interconnected graph, which was not the case for the older version.

Let Σ^A and Σ^B be two IO ABN of respective dimensions (n_A, p_A, q_A) and (n_B, p_B, q_B) . As before, suppose for convenience that $p_A = q_B, p_B = q_A$ and the interconnecting maps are simply identity maps. We also assume that each module has been separately analyzed: the transition graphs $G^{A,u}, u \in \{0, 1\}^{p_A}$ and $G^{B,v}, v \in \{0, 1\}^{p_B}$ have been constructed and decomposed into strongly connected components $\{A_{u\alpha}^i, 1 \leq i \leq N_u^A\}$ for each $u \in \{0, 1\}^{p_A}$ and $\{B_{v\beta}^j, 1 \leq j \leq N_v^B\}$ for each $v \in \{0, 1\}^{p_B}$. Let G

denote the full transition graph of the interconnected network, of size $2^{n_A+n_B}$. It can be computed thanks to (2), by interconnecting the modules' transition graphs. The idea behind the cross-graph is to generalize formula (2) in order to interconnect directly the SCCs of those graphs instead of the whole graphs themselves, thus potentially saving a significant amount of space when constructing the dynamics of the interconnection.

First, observe that the strongly connected components $\{A_u^i, 1 \leq i \leq N_u^A\}$ form a partition of the state space $\{0, 1\}^{n_A}$ of module Σ^A (N_u^A are integers verifying $1 \leq N_u^A \leq 2^{n_A}$). Therefore, for u varying in $\{0, 1\}^{p_A}$ we obtain 2^{p_A} partitions of the same finite set $\Omega = \{0, 1\}^{n_A}$. Let \mathfrak{P}_Ω denote the set of all partitions of Ω . Given two partitions $P_1, P_2 \in \mathfrak{P}_\Omega$, P_1 is said *finer than* P_2 , denoted by $P_1 \leq P_2$ if, for each element p in P_1 there is an element q in P_2 such that $p \subseteq q$ (in other words, partition P_1 is a *fragmentation* of partition P_2). The set $(\mathfrak{P}_\Omega, \leq)$ has the structure of a geometric lattice (see eg. Birkhoff, 1940). Consequently, for any set $S \subseteq \mathfrak{P}_\Omega$, there exists a (unique) greatest lower bound of S denoted by $\bigwedge S \in \mathfrak{P}_\Omega$. Coming back to the SCC decompositions, introduce the following partition:

$$\begin{aligned} Z^A &:= \bigwedge_{u \in \{0,1\}^{p_A}} \{A_u^i, 1 \leq i \leq N_u^A\}, \\ &= \{A^1, \dots, A^{N^A}\}, \end{aligned}$$

which is the coarsest partition of $\{0, 1\}^{n_A}$ that is finer than every SCC decomposition of all the transition graphs $G^{A,u}$. Once this partition is constructed, following the same idea as before it is further refined by cutting each set A^i according to their outputs: $A_\alpha^i := \{x \in A^i, h^A(x) = \alpha\}$, with the convention that such sets are simply omitted when they are empty. Therefore, we finally obtain a partition $Z_h^A = \{A_\alpha^i, 1 \leq i \leq N^A, \alpha \in \{0, 1\}^{q_A}\}$ of the state space $\{0, 1\}^{n_A}$ that is compatible with every SCC decompositions of the dynamics of modules Σ^A . By construction, the number of elements in this partition, denoted by M_A , verifies $1 \leq M_A \leq 2^{n_A}$. Applying the exact same procedure for module Σ^B , one obtains a similar partition $Z_h^B = \{B_\beta^j, 1 \leq j \leq N^B, \beta \in \{0, 1\}^{q_B}\}$ of the state space $\{0, 1\}^{n_B}$, containing M_B elements.

Once partitions Z_h^A and Z_h^B are defined, the construction of the cross graph closely resembles the one of the asymptotic graph. The cross graph is the digraph $G^{cr} = (V^{cr}, E^{cr})$, where the vertex set V^{cr} is composed of all cross-products $A_\alpha^i \times B_\beta^j$ and the arc set is constructed as follows:

- $A_\alpha^i \times B_\beta^j \rightarrow A_{\alpha'}^{i'} \times B_{\beta'}^{j'}$ iff there exist $a \in A_\alpha^i, a' \in A_{\alpha'}^{i'}$ such that there is a transition from a to a' in graph $G^{A,\beta}$,
- $A_\alpha^i \times B_\beta^j \rightarrow A_{\alpha'}^{i'} \times B_{\beta'}^{j'}$ iff there exist $b \in B_\beta^j, b' \in B_{\beta'}^{j'}$ such that there is a transition from b to b' in graph $G^{B,\alpha}$.

There is also a matricial form for the definition of G^{cr} . First, project each transition graph $G^{A,u}$ onto Z_h^A , leading to 2^{p_A} graphs, represented by their $M_A \times M_A$ adjacency matrices $H^{A,u}$, $u \in \{0, 1\}^{p_A}$. These projections can be rather straightforwardly achieved since Z_h^A is a fragmentation of the SCC decomposition of $G^{A,u}$. Second, for each $\alpha \in \{0, 1\}^{q_A}$, introduce the $M_A \times M_A$

diagonal matrix Δ_α^A such that entry $[\Delta_\alpha^A]_{ii} = 1$ if the output of the i -th element of Z_h^A is equal to α and 0 otherwise. Once similar objects $H^{B,u}$ and Δ_β^B have been constructed for module Σ^B , the cross-graph is simply defined by a generalization of formula (2):

$$G^{cr} := \bigvee_{(\alpha, \beta) \in \{0,1\}^{q_A} \times \{0,1\}^{q_B}} (H^{A,\beta} \otimes \Delta_\beta^B \vee \Delta_\alpha^A \otimes H^{B,\alpha}). \quad (3)$$

EXAMPLE 4. To illustrate this definition, let us consider two 2-dimensional, single-input single-output modules Σ^A and Σ^B , defined by their transition graphs given in **Figure 1A** and their output functions $h^A(x) = x_2$, $h^B(y) = y_1$. The full transition graph of the interconnection, built from (2), is depicted in **Figure 1B** and the cross-graph is depicted in **Figure 1C**: it is constructed from the two partitions $Z_h^A = \{\{00, 10\}, \{01, 11\}\} = \{\{0\}, \{1\}\}$ and $Z_h^B = \{\{00\}, \{10\}, \{01\}, \{11\}\}$. \square

The interest of the cross-graph lies in the following theorem, establishing the one-to-one correspondence between the terminal SCCs of G^{cr} and the attractors of the interconnected network.

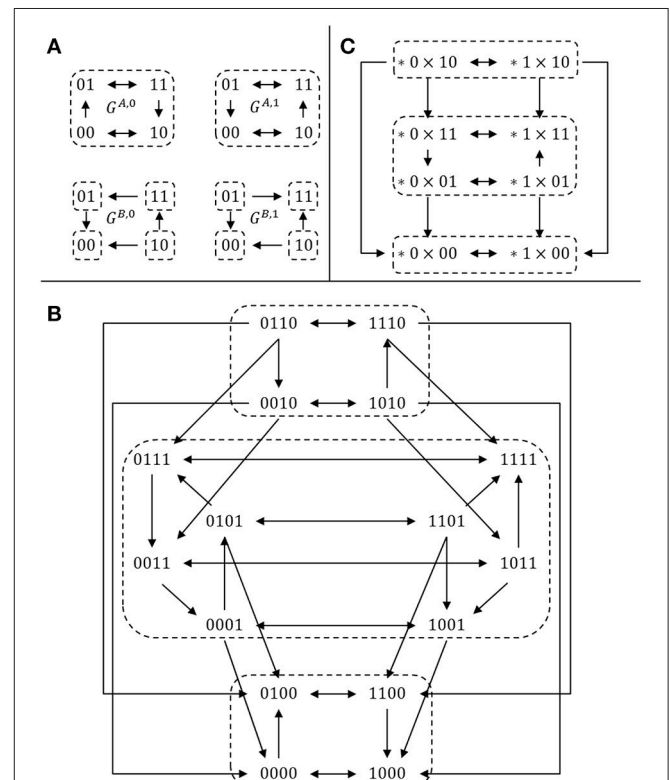


FIGURE 1 | Comparison between the cross graph of an interconnection and the full transition graph. **(A)** Transition graphs of two SISO modules (see Example 4); **(B)** full transition graph G of the interconnection; **(C)** cross graph G^{cr} of the interconnection. For each graph, dotted regions denote strongly connected components. There is a bijection between the SCC decomposition of the two graphs G (16 vertices) and G^{cr} (8 vertices), illustrating Theorem 2.

THEOREM 2. *Graphs G and G^{cr} have the same decomposition into strongly connected components. Furthermore, terminal SCCs of G^{cr} fully recover the attractors of the interconnected network.*

A proof of Theorem 2 is given in appendix. The size of the cross-graph is $M_A \times M_B$, which by construction is always less or equal than $2^{n_A+n_B}$, the size of the full interconnected graph G . The difference in size between the two graphs may vary greatly, and strongly depends on (i) the SCC decompositions of the two modules and (ii) as for the asymptotic graph, the numbers of inputs and outputs (and therefore the general modularity of the initial network). Part 4 proposes a brief evaluation of the performance of the method for a set of randomly generated interconnections. Although the interest of the cross-graph is mainly theoretical, in certain practical cases the full graph G can be too big to be stored easily while G^{cr} could.

Two possible extensions of the cross-graph method are noted here. First, Béranguier et al. (2013) proposed a compression of the SCC graph of a network, called the hierarchical transition graph (HTG). As the cross-graph is constructed from a combination of the modules' SCC decompositions, it would be possible to consider similarly a combination of the modules' HTG decompositions. Benefiting from the compactness of HTGs, such a construction would be even more compact than the cross-graph. Second, note that both the cross graph and the asymptotic graph methods require prior analysis of the modules' dynamics and the computation of their attractors, implicitly implying the dimensions of the modules are manageable. For a large network, Zañudo and Albert (2013) proposed an efficient characterization of attractors with the notion of "stable motifs," based on the network's interaction graph (see also Klarner et al., 2015). When considering interconnections of large modules, the investigation of the stable motifs of an interconnection would therefore constitute an interesting extension of Theorem 2.

3.2. A Probabilistic Asymptotic Graph

One of the limitations of Boolean models is the lack of quantitative details: while the state transition graph describes all possible dynamical behaviors, it gives no indication as to which trajectory is more likely to be observed under a given set of initial conditions. To circumvent this problem, Boolean models can be combined with probabilistic frameworks that account for biological perturbations and variability in the logical rules (Shmulevich et al., 2002; Mori et al., 2015). Another approach is to exploit the Markov chain description of the transition graph associated to the asynchronous Boolean model (Calzone et al., 2010; Stoll et al., 2017). Based on this description, Stoll et al. (2017) developed the MaBoSS software, which then applies Gillespie algorithm to produce continuous time trajectories.

We also use the Markov chain description to assign probabilities to the edges of the asymptotic graph, an approach which will lead to a more quantitative analysis of the interconnected network's dynamics. The output of our probabilistic asymptotic graph is thus the set of attractors of the full network, under a particular interconnection scheme, together with a *relative probability* for each of them (e.g., "there

is a probability p_1 that phenotype Q_1 is the outcome of this experiment").

The originality of our approach consists in assigning *incidence probabilities to the attractors of each separate module*, which can be obtained through the biological observations and measurements available for each module. The goal is to include biological information as an input and provide predictions that can be confronted to biological observations and therefore lead to validate or disprove the given interconnecting scheme.

3.2.1. Initializing Incidence Probabilities

Each transition in the asymptotic graph depends on two factors: which module is first "updated" (A or B) and, in response to an input change, how frequently does a switch occur from $A_{u\alpha}^i$ to $A_{u\alpha}^k$ (or from $B_{v\beta}^j$ to $B_{v\beta}^k$). These quantities may be represented by probabilities, defined *a priori*, from known data, experimental observations, or other modeling considerations.

Define

$$Q_A = P(\text{updating module } A \text{ first}).$$

Assume Boolean module Σ^A has a total of L_A same-output attractor-sets and Σ^B a total of L_B same-output attractor-sets,

$$\{A^{[i]} : A_{u_i\alpha_i}^i, i = 1, \dots, L_A\}, \{B^{[j]} : B_{v_j\beta_j}^j, j = 1, \dots, L_B\},$$

and each of these has a given *incidence probability* (meaning that it is observed with a certain frequency) defined as

$$P(A^{[i]}) = w_A^i, i = 1 \dots L_A, \quad P(B^{[j]}) = w_B^j, j = 1 \dots L_B.$$

The probabilities w_A^i and w_B^j may be assigned in different ways, for instance using experimental observations, or setting uniform probabilities ($w_A^i = 1/L_A$ for all i), or else from the size of their respective basin of attraction

$$w_A^i = \frac{\#\text{basin}^i}{\sum_i \#\text{basin}^i}, \quad (4)$$

but in any case they should satisfy $\sum_{i=1}^{L_A} w_A^i = 1$. Using these initial probabilities, a *joint incidence probability* may similarly be defined for each product of attractor-sets:

$$P(A^{[i]} \times B^{[j]}) = w_A^i w_B^j, \Rightarrow \sum_{i=1}^{L_A} \sum_{j=1}^{L_B} P(A^{[i]} \times B^{[j]}) = 1.$$

3.2.2. Transition Probabilities in the Asymptotic Graph

The probability of switching between two attractor-sets of the same module, but different inputs, can be defined in terms of conditional probabilities: define s_A^{ik} to be the probability that attractor $A^{[k]}$ is reached, conditional to the fact that the initial state is some $a^i \in A^{[i]}$. In other words, w_A^k must be weighted by the probability of a^i reaching any attractor in $G^{A,u,k}$:

$$s_A^{ik} = P(A^{[k]} | [a^i \in A^{[i]}]) = \frac{P(A^{[k]})}{\sum_{j \in \mathcal{J}} P(A^{[j]})} = \frac{w_A^k}{\sum_{j \in \mathcal{J}} w_A^j}, \quad (5)$$

where $\mathcal{J} = \{j: u_j = u_k \text{ and } a^i \rightsquigarrow A^{[j]}\}$ means that there exists a path in G^{A, u_k} leading from a^i to $A^{[j]}$, where $A^{[j]}$ is an attractor of G^{A, u_k} . A similar definition holds for s_B^{jk} .

Next, we can define the probability associated to an edge of V^{as} as:

$$\begin{aligned} P(A^{[i]} \times B^{[j]} \rightarrow A^{[k]} \times B^{[l]}) &= \bar{\varrho}_A s_A^{ik}, \\ P(A^{[i]} \times B^{[j]} \rightarrow A^{[i]} \times B^{[k]}) &= (1 - \bar{\varrho}_A) s_B^{jk}, \end{aligned} \quad (6)$$

with an “effective” probability $\bar{\varrho}_A$, computed based on the set of all outgoing edges from node $A^{[i]} \times B^{[j]}$:

$$\bar{\varrho}_A = \begin{cases} 0, & A^{[i]} \equiv A^{[k]} \\ 1, & B^{[j]} \equiv B^{[k]} \\ \varrho_A, & \text{otherwise.} \end{cases} \quad (7)$$

In other words, $\bar{\varrho}_A = 0$ if all outgoing edges have a fixed A -attractor, $A^{[i]} \times B^{[j]} \rightarrow A^{[i]} \times B^{[k]}$; $\bar{\varrho}_A = 1$, if all outgoing edges have a fixed B -attractor $A^{[i]} \times B^{[j]} \rightarrow A^{[k]} \times B^{[j]}$; $\bar{\varrho}_A = \varrho_A$ if outgoing edges may be of both types.

Note that these definitions ensure that the probabilistic asymptotic graph matrix has the property that all rows add up to 1:

$$\begin{aligned} &\sum_k P(A^{[i]} \times B^{[j]} \rightarrow A^{[k]} \times B^{[l]}) \\ &+ \sum_k P(A^{[i]} \times B^{[j]} \rightarrow A^{[i]} \times B^{[k]}) \\ &= \sum_k \bar{\varrho}_A s_A^{ik} + \sum_k (1 - \bar{\varrho}_A) s_B^{jk} = \bar{\varrho}_A + (1 - \bar{\varrho}_A) = 1 \end{aligned}$$

since both $\sum_k s_A^{ik} = 1$ and $\sum_k s_B^{jk} = 1$.

3.2.3. Relative Probabilities of the Attractors of an Interconnection

If the asymptotic graph G^{as} has two or more attractors, in addition to the transition probabilities, another useful information is the frequency of observing a given attractor, or in other words the *relative probability of each attractor* of the interconnection. This probability can be computed from the SCC graph $G^{Sd} = (V^{Sd}, E^{Sd})$ corresponding to G^{as} , which is an acyclic graph and can be represented by an absorbing Markov chain. By definition, V^{Sd} is composed of the strongly connected components of G^{as} . Let $C \in V^{Sd}$ contain L_C elements of V^{as} . Define the incidence probability of observing C as:

$$P(C) = \sum_{\ell=1}^{L_C} P(A^{[i(\ell)]} \times B^{[j(\ell)]}) = \sum_{\ell=1}^{L_C} w_A^{i(\ell)} w_B^{j(\ell)}.$$

Moreover, a probability of transition can also be associated to each edge of E^{Sd} , $P(C^i \rightarrow C^j)$, computed by adding all the probabilities of the edges in E^{as} that link elements of C^i to elements of C^j . Suppose there are m strongly connected components, $|V^{Sd}| = m$, and let the $m \times m$ matrix M with $M_{ij} = P(C^i \rightarrow C^j)$, be the absorbing Markov chain associated with the

graph G^{Sd} . Suppose M has r absorbing states, $\{C_a^k: k = 1, \dots, r\}$, these are also the attractors of G^{Sd} . Matrix M can be written in the following canonical form (Feller, 1970):

$$M = \begin{bmatrix} Q & R \\ 0 & I_r \end{bmatrix},$$

where I_r is the $r \times r$ identity matrix, Q is the $(m-r) \times (m-r)$ matrix of transitions between transient states and R is the $(m-r) \times r$ matrix of transitions from transient states to absorbing states. Since M is irreducible, it follows that $(I - Q)$ has an inverse (where I is the $(m-r) \times (m-r)$ identity matrix). Then the probability that there exists a path from a given state to one of the r absorbing states is given by the probability of being absorbed by r :

$$M_{absorp} = (I - Q)^{-1} R,$$

where $M_{absorp}(i, k)$ is the probability that transient state i converges to absorbing state k .

If, in addition, we wish to weigh these absorption probabilities by the incidence probabilities of observing C_a^k , we can define the *relative probability of an attractor of the asymptotic graph*:

$$P_{rel}(C_a^k) = P(C_a^k) + \sum_{i=1}^{m-r} M_{absorp}(i, k) P(C^i), \quad k = 1, \dots, r \quad (8)$$

where C_a^k denotes each attractor and $P(C_a^k)$ is the incidence probability of C_a^k .

4. PERFORMANCE ON RANDOM NETWORKS' INTERCONNECTIONS

In this part we propose a series of computational experiments to assess the efficiency of the asymptotic graph and the cross graph to recover the attractors of random interconnected Boolean networks. Following the general idea of inputs/outputs at the core of this paper, we start with a brief description of the algorithm used to generate random IO modules. We then present numerical results computed on random interconnections with varying connectivity, showing the respective advantages and limitations of the two methods in practice.

4.1. Generation of Random IO Networks With Varying Connectivity

The NK-model, introduced by Kauffman (1969), is a general statistical model to represent random Boolean networks by controlling their dimension N and their inner connectivity K . It is used for instance by Zañudo and Albert (2013) and Veliz-Cuba et al. (2014). Here it is slightly adapted to include inputs and outputs. Let Σ be an IO Boolean network of dimension (n, p, q) , of transition function $f: \{0, 1\}^p \times \{0, 1\}^n \rightarrow \{0, 1\}^n$ and output function $h: \{0, 1\}^n \rightarrow \{0, 1\}^q$. A usual way to depict such a network is by its wiring diagram, showing the dependencies between the different variables of the network. Equivalently, the wiring diagram can be represented by a $(n+q) \times (p+n)$ Boolean matrix

$$M = \begin{bmatrix} A & B \\ 0 & D \end{bmatrix},$$

where submatrices A ($n \times p$), B ($n \times n$) and D ($q \times n$) are defined as follows:

$$a_{ij} = \begin{cases} 1 & \text{if function } f_i \text{ depends explicitly of input variable } u_j, \\ 0 & \text{otherwise,} \end{cases}$$

$$b_{ij} = \begin{cases} 1 & \text{if function } f_i \text{ depends explicitly of variable } x_j, \\ 0 & \text{otherwise,} \end{cases}$$

$$d_{ij} = \begin{cases} 1 & \text{if output function } h_i \text{ depends explicitly of variable } x_j, \\ 0 & \text{otherwise.} \end{cases}$$

Let C designate the matrix $[A|B]$. The sum of the i -th row of C is the number of essential variables of logical function f_i , also called the *connectivity* of f_i . Given integers $n > 0$, $p, q \geq 0$ and a real number $K_{mean} \in [1, n]$, we construct a random IO network of dimension (n, p, q) and of average connectivity K_{mean} by applying the following procedure, which generates a dependency matrix M :

1. Let $D := 0$. For each $1 \leq i \leq q$, pick at random $j \in \{1, \dots, n\}$ and set $d_{ij} := 1$.
2. Generate n integers k_i in $\{0, \dots, n+p\}$ according to a binomial distribution of parameters $n+p$ (number of trials) and $\frac{K_{mean}}{n+p}$ (probability of success).
3. Let $C = [A|B] := 0$. For each $1 \leq i \leq n$, pick a random combination $(j_1, \dots, j_{k_i}) \in \{1, \dots, n+p\}^{k_i}$ (without replacement) and set $c_{i,j_l} := 1$ for all $1 \leq l \leq k_i$.
4. Check that each column of A is non-zero; while it is not the case, repeat step 3.
5. Set $M := \begin{bmatrix} C \\ 0|D \end{bmatrix}$.

Step 4 ensures the generated module actually depends of every inputs. Once the dependency matrix M is obtained, the last step consists in generating the $n+q$ Boolean functions according to M . A Boolean function of k variables is picked randomly among the 2^k possibilities; in case it is degenerate (i.e., at least one of the k variables is not essential), another one is chosen so as to ensure exact compatibility with M .

4.2. Complementarity of the Cross and Asymptotic Graph Methods

With this algorithm, it is possible to generate a IO module by controlling its inner connectivity, that is the number of actual dependencies in the wiring diagram. Thus, it becomes possible to generate random interconnections with varying degrees of *modularity*, according to the average connectivity of each module. We used this algorithm to generate 2,000 interconnections of two modules Σ^A and Σ^B of dimensions $(n_A, p_A, q_A) = (n_B, p_B, q_B) = (10, 2, 2)$:

$$\boxed{\Sigma^A} \begin{matrix} \Rightarrow \\ \Leftarrow \end{matrix} \boxed{\Sigma^B}, \quad (9)$$

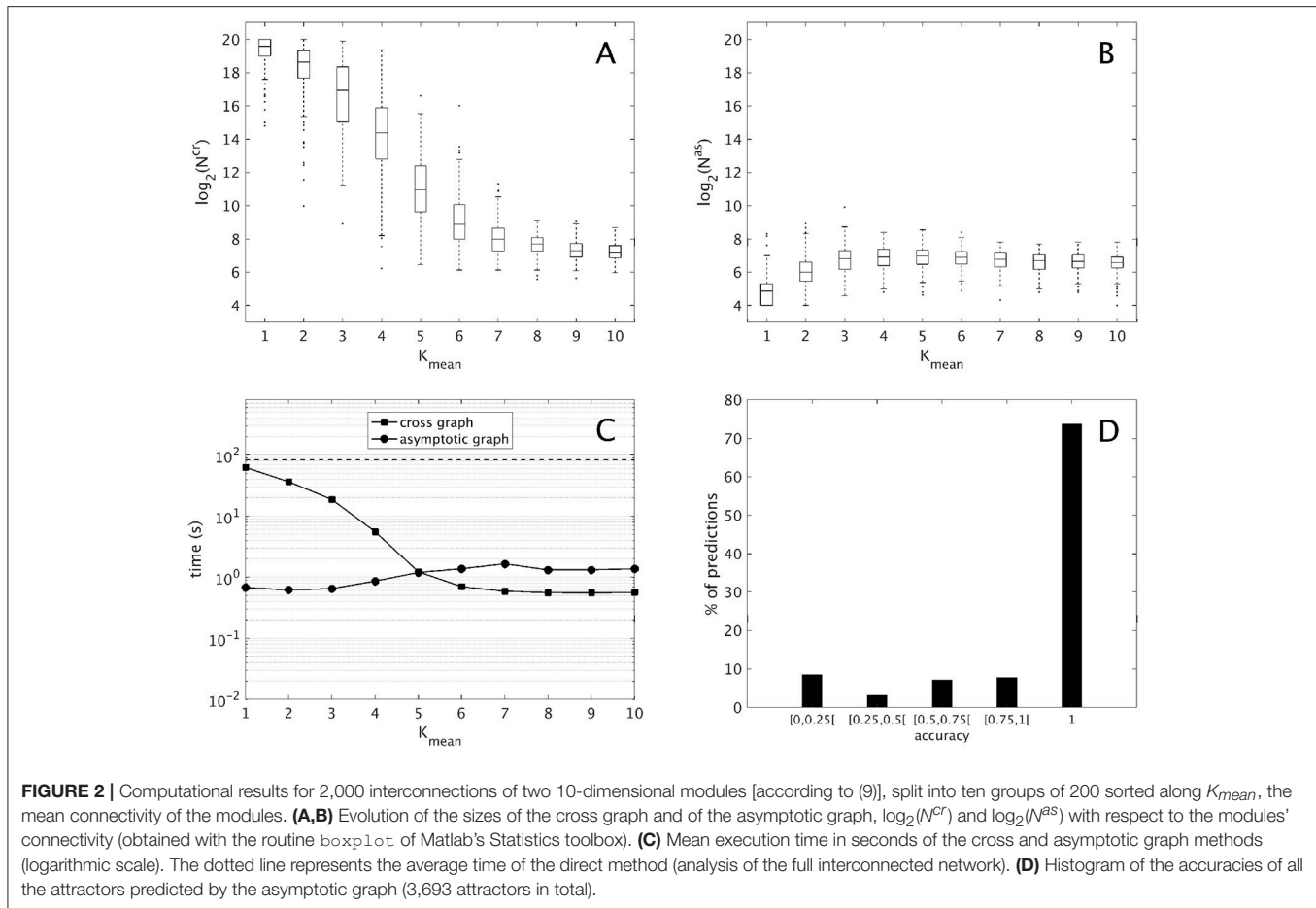
where the mean connectivity of Σ^A and Σ^B varies in $\{1, \dots, 10\}$. For each interconnection, both 10-dimensional modules were analyzed separately (including the computation of the transition graphs, their SCC decompositions and the computation of their attractors), then the cross graph and the asymptotic graph were

computed and compared. The main results are presented in **Figure 2** and summarized below. All computations were made with Matlab R2016b, The MathWorks, Inc.

First, we compare the respective sizes N^{cr} and N^{as} of the cross and the asymptotic graphs (i.e. their number of vertices). **Figures 2A,B** show respectively the evolution of $\log_2(N^{cr})$ and $\log_2(N^{as})$ with respect to the connectivity of the two modules. Obviously, both N^{cr} and N^{as} are below $N = 2^{20}$, which is the size of the full transition graph of the interconnected network. The cross graph, which captures both the transient and the asymptotic dynamics of the interconnection is relatively large, however its size seems to vary greatly with the modules' connectivity. When the connectivity increases, implying a highly modular interconnection, the ratio N^{cr}/N can reach very small values, emphasizing the interest of the cross graph to efficiently store the dynamics of large, modular interconnected networks. On the other hand, the asymptotic graph is always much smaller, several orders of magnitude under the size of the full transition graph. Contrary to the cross graph, it is particularly small when the modules have lower connectivity, making it particularly well adapted for biological networks. Interestingly, its size seems to reach a plateau when the mean connectivity is above $\frac{n}{2} = 5$.

Another way to compare the two approaches is by studying their average execution times. The times shown in **Figure 2C** include the analysis of the two 10-d modules and of the cross and asymptotic graph methods. The latter comprise the construction of G^{cr} (respectively, of G^{as}), the SCC decomposition of G^{cr} (respectively, of G^{as}) and the reconstruction of the attractors (respectively, of $\pi(R)$ for all terminal SCCs R of G^{as}). For the cross graph, the majority of the time is taken by the SCC decomposition of G^{cr} while for the asymptotic graph, the most time-consuming step is the construction of G^{as} itself (data not shown). For comparison, we also computed the complete dynamics of the 20-d interconnected network by using formula (2); on average, such direct method amounted to around 83 seconds (dotted line). Therefore, both methods are faster than the direct analysis of the full interconnected network. As before, the asymptotic graph is particularly efficient for low connectivity modules, while the cross graph is more efficient when the modules have high connectivity. Interestingly, for connectivity $K_{mean} = 5$ and higher, when both graphs have roughly the same size, the cross graph method becomes even more rapid than the asymptotic graph.

Finally, since both graphs were computed it was possible to evaluate the quality of the asymptotic graph predictions. Recall that according to Theorem 1, the asymptotic graph has two drawbacks. First, it may predict spurious attractors and second, when it identifies a true attractor it only predicts a subset $\pi(R)$ of the states lying in the attractor Q . The ratio $\frac{|\pi(R)|}{|Q|}$ is called the *accuracy* of the prediction. Among the 2,000 interconnections, 11 presented spurious attractors that is only 0.55% of the total. In all but one case, only one spurious attractor was detected. This result confirms the rarity of the appearance of spurious attractors. In total, we identified 3,693 true attractors. Among them more than 73% were completely recovered (see **Figure 2D**); overall, the mean accuracy is about 0.86, exhibiting the excellent predictive power of the asymptotic



graph when it comes to uncover the asymptotic behaviors of an interconnection.

4.3. A Powerful Tool to Analyze Large Interconnections of Biological Networks

According to the previous results, the asymptotic graph seems particularly well adapted when the mean connectivity of the modules is low (≤ 5), which is arguably where biological networks generally operate (Zañudo and Albert, 2013; Veliz-Cuba et al., 2014). Therefore we decided to test it further with higher dimensional interconnections, including four modules $\Sigma^A, \Sigma^B, \Sigma^C, \Sigma^D$ of dimension $n = 15$, with $K_{mean} \in \{1, \dots, 5\}$, $p_A = q_A = p_D = q_D = 1$ and $p_B = q_B = p_C = q_C = 2$:

$$\boxed{\Sigma^A} \begin{matrix} \rightarrow \\ \leftarrow \end{matrix} \boxed{\Sigma^B} \begin{matrix} \rightarrow \\ \leftarrow \end{matrix} \boxed{\Sigma^C} \begin{matrix} \rightarrow \\ \leftarrow \end{matrix} \boxed{\Sigma^D} \quad (10)$$

When $N^{cr} < 10^7$, the cross graph was also constructed and analyzed, in order to check the existence of spurious attractors. Since the global state space is $2^{60} > 10^{18}$, we skipped the last treatment (identification of the attractors in $\{0, 1\}^{60}$) to avoid possible explosions. Therefore, we only computed the terminal SCCs of G^{as} and, when available, the terminal SCCs of G^{cr} . The results are presented in **Table 1**. When G^{cr} could be analyzed,

we were able to detect spurious attractors in G^{as} : none were found. If the cross graph method is not practical for small K_{mean} , the asymptotic graph was always manageable, confirming its practical interest to analyze large biological networks, as long as they can be expressed as interconnections of modules with a reasonable number of inputs and outputs.

5. TWO BIOLOGICAL APPLICATIONS

The asymptotic graph construction and its probabilistic interpretation are now applied to two biological examples, centered on the mammalian and yeast cell cycles. Both cases illustrate the asymptotic graph concept, its informative description of a composite system, and its usefulness for testing biological hypotheses.

5.1. Mammalian Cell Cycle, Circadian Clock and Their Interconnection

There are two basic cellular oscillators in mammalian cells: cell cycle describes the different phases of cellular growth and division, while circadian clock describes the mechanism responsible for anticipating environmental changes and adapting the organism to deal with these changes (most notably, day-night differences). The interactions between these two oscillators are

TABLE 1 | Computational results for 200 interconnections of four 15-dimensional modules [according to (10)], split into five groups of 40, sorted along the mean connectivity of the modules.

K_{mean}	$\log_2(N^{cr})$		Time (s)		#treated/(#exp.)	$\log_2(N^{as})$		Time (s)		#spurious/(#treated)
	mean	std	mean	std		mean	std	mean	std	
1	57.3	2.3	—	—	0/40	8.5	1.4	9	2	—
2	52.2	4.1	—	—	0/40	9.8	1.1	9	7	—
3	42.4	5.7	—	—	0/40	11.0	1.4	63	185	—
4	29.6	5.9	493	361	6/40	11.3	1.1	40	51	0/6
5	20.9	4.7	176	223	28/40	11.0	1.0	27	38	0/28

The cross graph is treated (constructed and analyzed) only when $N^{cr} < 10^7$ ($\log_2(10^7) \approx 23.25$). The column #treated/(#exp.) indicates the number of times it was treated over the total number of experiments. When it is treated, we further verify the presence of spurious attractors in the asymptotic graph. The column #spurious/(#treated) indicates the number of times the asymptotic graph predicts a spurious attractor over the number of times the cross graph could be treated. Symbol — indicates that the corresponding value could not be computed.

still not fully understood, but recent works by Feillet et al. (2014) and Bieler et al. (2014) have uncovered unexpected bi-directional links between the two modules. Successful mathematical models for the cell cycle and clock have been developed, as well as some studies on their interactions (Gérard and Goldbeter, 2012), but many questions remain (Feillet et al., 2015).

5.1.1. Mammalian Boolean Modules

At the discrete level, a reference model of the cell cycle was developed and discussed by Fauré et al. (2006) (see **Figure 3**). It comprises 10 variables:

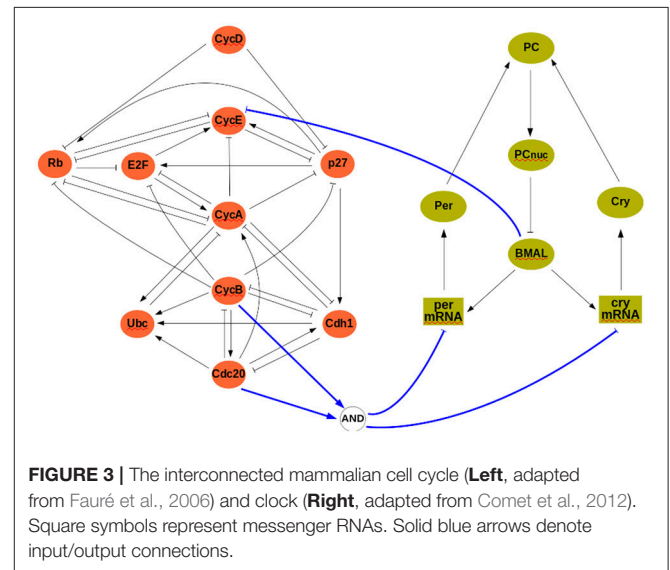
$$(CycD, Rb, E2F, CycE, CycA, p27, Cdc20, Cdh1, Ubc, CycB),$$

where $CycX$ ($X \in \{A, B, D, E\}$) represent four cyclins, each roughly corresponding to one of the four phases of the cell cycle. This constitutes our module Σ^A , and its rules can be found in the Supplementary Material. The clock model (module Σ^B) has 7 variables and is based on the work of Comet et al. (2012). To account for transcription shutdown during mitosis, the input v negatively affects all mRNAs:

$$\begin{aligned}
 BMAL^+ &= \neg PCnuc \\
 mPER^+ &= \neg v \wedge BMAL \\
 mCRY^+ &= \neg v \wedge BMAL \\
 pPER^+ &= mPER \\
 pCRY^+ &= mCRY \\
 PC^+ &= pPER \wedge pCRY \\
 PCnuc^+ &= PC.
 \end{aligned} \quad (11)$$

In the clock model, mX and pX denote mRNA and protein coded by gene X , while PC denotes the complex formed by the proteins PER and CRY, and $PCnuc$ denotes this complex in the nucleus.

A well established link between these two oscillators is that protein $BMAL$ acts on the cell cycle, possibly at different stages (Feillet et al., 2015). In our analysis, we will consider $BMAL$ acting during G1 phase. Although no conclusive evidence exists on how the cell cycle may affect the clock, we have considered that during cell division (or mitosis phase) gene expression is stopped (in the model, mitosis can be modeled as $Cdc20 \wedge CycB$,



see **Figure 3**). The interconnection between modules is thus given by:

$$u = h^B(b) = BMAL, \quad v = h^A(a) = Cdc20 \wedge CycB,$$

so that $u = 0$ (resp., $u = 1$) represents absence (resp., presence) of $BMAL$ and $v = 1$ represents mitosis. In the cell cycle model, $BMAL$ affects negatively the G1 phase, leading to a logical equation for cyclin E of the form $cycE^+ = \neg u \wedge (E2F \wedge \neg Rb)$ (see **Figure 3** and Supplementary Material).

Module Σ^A has a total of six, and module Σ^B has a total of three, same-output attractor sets. For algorithmic convenience, these are labeled using the lexicographic convention, that is $A_{\hat{u}\hat{\alpha}}^j$ for $\hat{u}, \hat{\alpha} \in \{1, 2\}$, where “decimal 1 = logical 0” and “decimal 2 = logical 1.” The attractors for both modules are as follows:

$$\begin{aligned}
 G^{A,u=0} : & A_{11}^1 = \{0100010100\}, A_{11}^2 (80 \text{ states}), A_{12}^3 (32 \text{ states}), \\
 G^{A,u=1} : & A_{21}^4 = \{0100010100\}, A_{21}^5 (40 \text{ states}), A_{22}^6 (16 \text{ states}), \\
 G^{B,v=0} : & B_{11}^1 (57 \text{ states}), B_{12}^2 (63 \text{ states}), \\
 G^{B,v=1} : & B_{22}^3 = \{1000000\}.
 \end{aligned}$$

In the case $u = 0$, module Σ^A becomes exactly the original model constructed by Fauré et al. (2006). Therefore, as expected, the attractors found for $G^{A,u=0}$ correspond exactly to those listed by Fauré et al. (2006). Attractors A_{11}^1 and A_{21}^4 correspond to a steady state where the only expressed proteins are *Rb*, *p27*, and *Cdh1*, hence representing the quiescent cell state. The (full) attractor $A_{11}^2 \cup A_{12}^3$ is a cyclic attractor containing 112 distinct states and corresponds to the known G1/S/G2/M cell cycle progression (Fauré et al., 2006). Similarly, $A_{21}^5 \cup A_{22}^6$ is a cyclic attractor of the graph $G^{A,u=1}$, with 56 states. It tends to describe the cell cycle progression, with the difference that $u = 1$ implies $CycE \equiv 0$. In either of the cyclic attractors, the attractor-sets A_{12}^3 and A_{22}^6 contain states representing mitosis, that is, the output of any state $a \in A_{22}^6 \cup A_{12}^3$ satisfies $h^A(a) = Cdc20 \wedge CycB = 1$.

The clock mechanism admits a cyclic attractor with 120 states, $B_{11}^1 \cup B_{12}^2$, which corresponds to regular circadian oscillations in the case $v = 0$. At mitosis, represented by $v = 1$, the clock network admits a single steady state attractor ($B_{22}^3 = \{1000000\}$), where all gene expression is arrested.

5.1.2. Asymptotic and Cross Graphs

The asymptotic graph for the interconnection of the two mammalian oscillators has 18 nodes and two attractors, with separate basins of attraction (Figure 4):

$$\begin{aligned} R_1 &= \{A_{11}^1 \times (B_{11}^1 \cup B_{12}^2), A_{21}^4 \times (B_{11}^1 \cup B_{12}^2)\} \\ R_2 &= \{(A_{11}^2 \cup A_{12}^3) \times B_{11}^1, A_{11}^1 \times B_{12}^2, A_{21}^5 \times (B_{11}^1 \cup B_{12}^2), \\ &\quad (A_{21}^5 \cup A_{22}^6) \times B_{22}^3, A_{22}^6 \times B_{12}^2, A_{12}^3 \times B_{22}^3\}. \end{aligned}$$

The cross graph contains 54,272 nodes (compare to the full size of the interconnection, $2^{17} = 131072$) and confirms the existence of exactly two cyclic attractors for the interconnected system and returns all their elements: attractor R_1 is composed of 120 states and R_2 is composed of 13,552 states.

Our methodology predicts two distinct operating modes for the coupled oscillators: R_1 corresponds to a quiescent cell with oscillatory clock, since it is the product of state 0100010100 representing a quiescent cell in module Σ^A and of cyclic attractor $B_{11}^1 \cup B_{12}^2$ representing regular clock oscillations. The attractor R_1 is thus in agreement with observations by Plikus et al. (2013) (hair cells in quiescent phase seem to have a running clock). In contrast, R_2 represents joint oscillations of the cell progression cycle ($A_{11}^2 \cup A_{12}^3$) and clock ($B_{11}^1 \cup B_{12}^2$) (see Figure 4 for the dynamics within R_2). The cell cycle and clock may jointly oscillate and alternate states with a regular cycle of cyclin E (which is present mostly through S phase and mitosis) or eventually switch to a joint cycle with absence of cyclin E ($A_{21}^5 \times B_{11}^1 \rightarrow A_{21}^5 \times B_{11}^1 \rightarrow A_{21}^5 \times B_{12}^2 \rightarrow A_{21}^5 \times B_{11}^1$). However, at mitosis (A_{12}^3), the clock may switch to its arrested steady state ($A_{12}^3 \times B_{11}^1 \rightarrow A_{12}^3 \times B_{22}^3$), which leads directly to a full degradation of cyclin E in the cell cycle ($A_{22}^6 \times B_{22}^3$).

To assign transition probabilities to the asymptotic graph, there are essentially two elements to define: ϱ_A which is the probability of updating first the component from module Σ^A ; and the incidence probability of each attractor from each module, w_A^i and w_B^j . To compute the incidence probabilities w_A^i and w_B^j , we have used the size of the original basins of attraction of $A_{u\alpha}^i$

in Σ^A and $B_{v\beta}^j$ in Σ^B , as in (4). However, for both modules, each attractor can be reached from any state, implying that the joint incidence probabilities, $P(A^{[i]} \times B^{[j]}) = w_A^i \times w_B^j$, are equal for all nodes of the asymptotic graph with: $w_A^i = 1/6$ ($i = 1, \dots, 6$) and $w_B^j = 1/3$ ($j = 1, \dots, 3$).

Figure 4 shows the transition probabilities obtained for two different values of the updating probability ϱ_A . These two graphs are very similar, differing only on the most frequent transitions (bold arrows, above 0.5). As should be expected, whenever the probability of first updating components from Σ^A is larger ($\varrho_A = 0.6$), the cell cycle oscillations dominate the global dynamics: most of the bold transitions in Figure 4 (bottom) concern switches between attractor-sets of Σ^A . In contrast, circadian clock oscillations are dominant for $\varrho_A = 0.2$ (Figure 4, top). The evolution from mitosis phase toward cell cycle progression ($A_{12}^3 \times B_{22}^3 \rightarrow A_{21}^5 \times B_{22}^3$ or $A_{12}^3 \times B_{22}^3 \rightarrow A_{22}^6 \times B_{22}^3$) is equally probable for either ϱ_A .

Computation of the relative probabilities (8) of reaching one of the attractors of the interconnected network yields

$$P_{rel}(R_1) = 0.333, \quad P_{rel}(R_2) = 0.667,$$

independently of the updating probability ϱ_A . An interpretation of these relative probabilities is that, in a typical population of cells, about one third are arrested in quiescent G0 state while the other two thirds follow the normal cell cycle progression G1/S/G2/M.

5.2. Budding Yeast Cell Growth and Cell Cycle START

Cell cycle and division is intimately linked with cell growth: a cell cannot divide into two daughter cells if its size is too small. There are many other factors that play a role in cell division (concentration of certain proteins, volume), but it remains unclear how a cell is able to perceive its own size and evaluate whether all conditions are in place for cell division (Turner et al., 2012).

In budding yeast, cell cycle is triggered by a START signal which is dependent on cell size. Li et al. (2004) propose a Boolean model that accurately describes cell cycle progression, taking START as an external input and stopping at a G1 phase steady state. One of the most important proteins involved in START is cyclin *Cln3*, which is involved in the G1-S phase transition and initiates cell cycle in the model of Li et al. (2004). Cyclin *Cln3* forms a complex with another protein *Whi3* but, in order to initiate cell cycle, *Cln3* must be folded and released from this complex, which is achieved with the help of a chaperon protein *Ydj1*. Recent work by Aldea et al. (2017) suggests that cell size is growth rate dependent and that *Ydj1* is one of the most important factors relating growth rate to cell size at START.

5.2.1. Budding Yeast Boolean Modules

A reference discrete model for the cell cycle was developed by Li et al. (2004). It comprises 11 variables:

$$(START, MBF, SBF, Cln1, Cdh1, Swi5, Cdc20, Clb5, Sic1, Clb1, Mcm)'$$

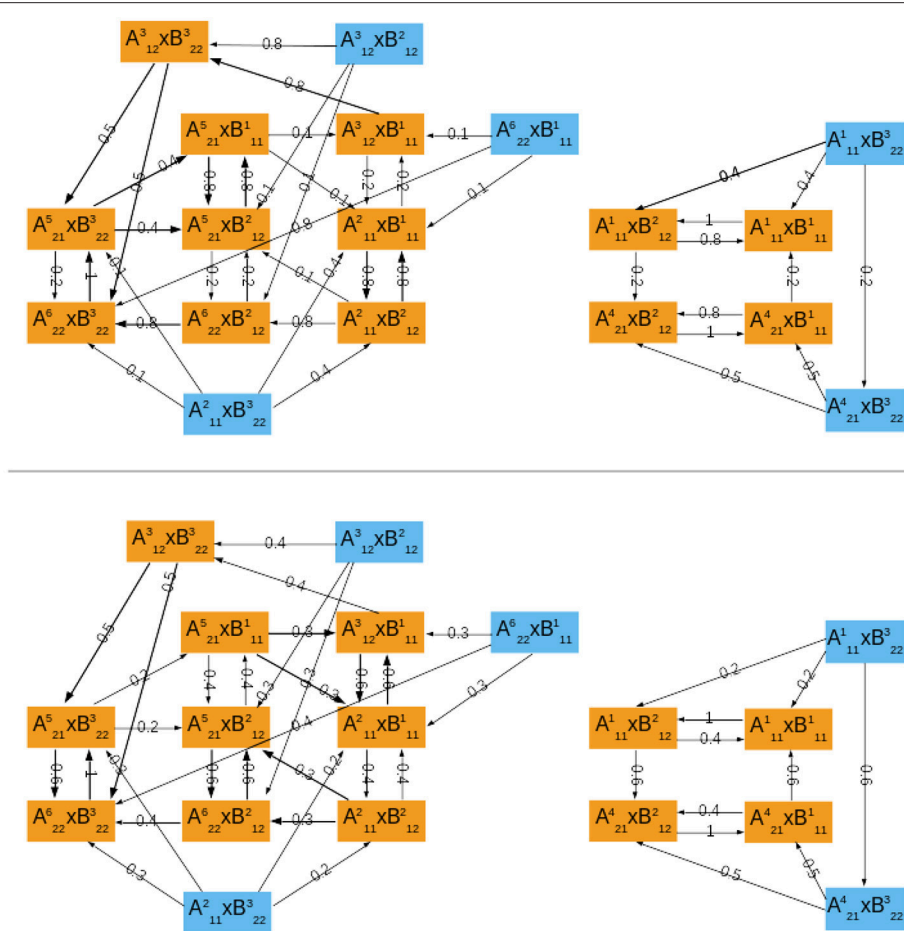
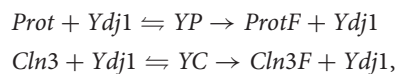


FIGURE 4 | The probabilistic asymptotic graph for the interconnected mammalian oscillators. Orange colored nodes belong to an attractor: R_1 at right and R_2 at left. Bold arrows represent transitions with probability ≥ 0.5 . **(Top)** $\epsilon_A = 0.2$. **(Bottom)** $\epsilon_A = 0.6$.

with *START* given by *Cln3* (see **Figure 5**; the Boolean rules can be found in the Supplementary Material).

To describe cell size dependence on growth rate Aldea et al. (2017) proposes a model where *Cln3* competes with a second hypothetical protein *Prot* for binding with *Ydj1* for folding:



and *Prot* would be a growth rate dependent protein. Here, we propose a basic Boolean network of this model, where the dependence on growth rate is modeled by an input v :

$$\begin{aligned} Ydj1^+ &= YP \vee YC \vee \neg(Prot \wedge Cln3) \\ YP^+ &= Ydj1 \wedge Prot \\ YC^+ &= Ydj1 \wedge Cln3 \\ Prot^+ &= v \\ ProtF^+ &= YP \\ Cln3^+ &= \neg Whi3 \end{aligned} \quad (12)$$

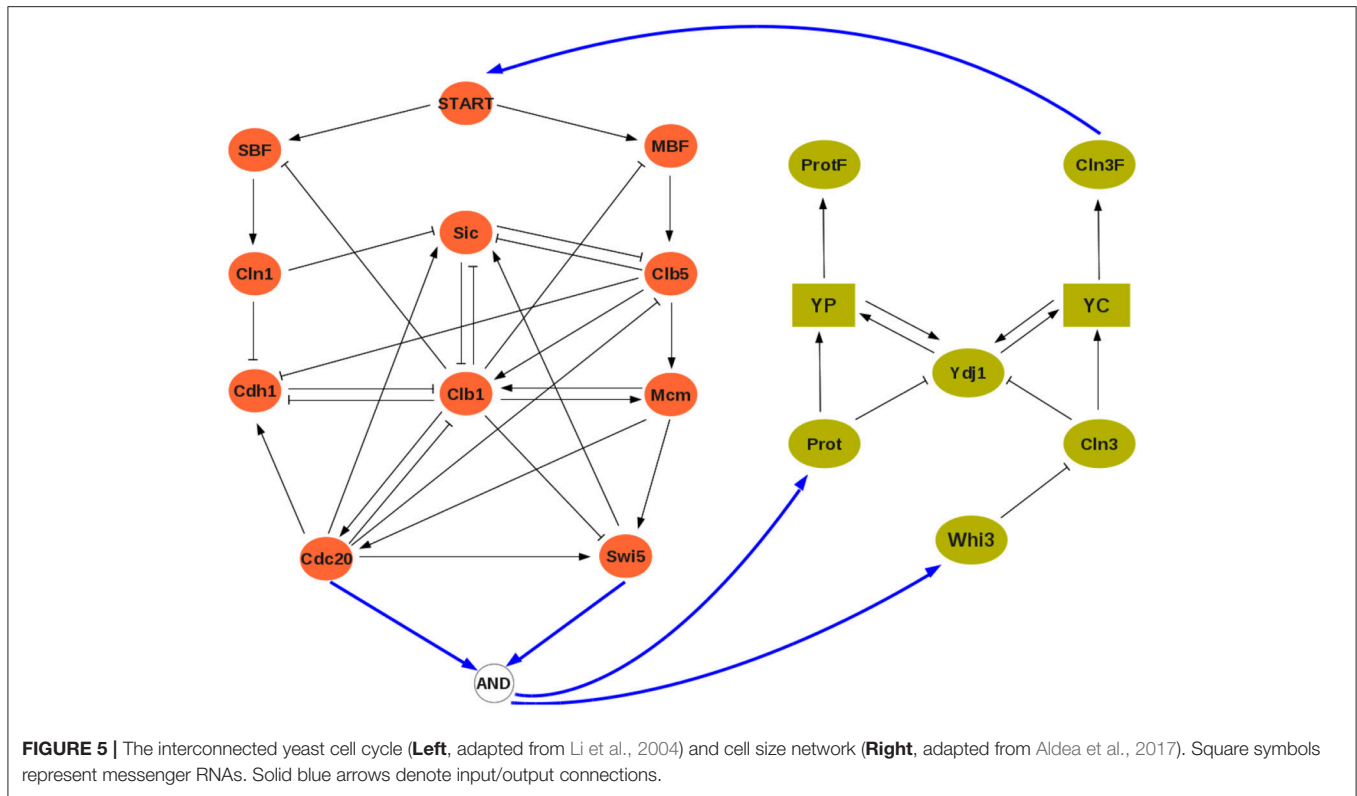
$$Cln3F^+ = YC$$

$$Whi3^+ = v.$$

The competition of *Prot* and *Cln3* for *Ydj1* is represented by the term $\neg(Prot \wedge Cln3)$ in the rule for *Ydj1* meaning that, in the absence of both *Prot* and *Cln3*, “free” protein *Ydj1* will be available. Both *Prot* and *Whi3* depend on growth rate, here given by input v . Later on, v will be computed as an output from the cell cycle model.

Computation of the graphs $G^{A,u}$ and $G^{B,v}$ yields the following attractors:

$$\begin{aligned} G^{A,u=0}: \quad & A_{11}^1 = \{0000000000\}, A_{11}^2 = \{00000000100\}, \\ & A_{11}^3 = \{00001000000\}, \\ & A_{1*}^4 = \{00001000100\}, A_{11}^5 = \{00110000000\}, \\ & A_{11}^6 = \{01000000100\}, \\ & A_{1*}^7 = \{01001000100\}, \\ G^{A,u=1}: \quad & A_{2*}^8 = \{10110110011\}, A_{2*}^9 = \{11000111011\}, \\ & A_{2*}^{10} = \{11110110011\}, \\ & A_{2*}^{11} = \{11110111011\}, \end{aligned}$$



$$G^{B,v=0}: B_{12}^1 = \{10100110\},$$

$$G^{B,v=1}: B_{21}^2 = \{11011001\}.$$

The symbol $*$ in A_{1*}^i or A_{2*}^i means that the output of this attractor depends on the function $h^A(a)$: three different forms for $h^A(a)$ will be tested (see 13–15 below). For instance, we have $h^A(A_{1*}^4) = 2$ whenever $h^A(a)$ is given by (15), so we should write A_{12}^4 ; but $h^A(A_{1*}^4) = 1$ in the other two cases, hence A_{11}^4 .

In the case $u = 0$, the yeast cell cycle model is exactly the one studied by Li et al. (2004) hence, as expected, the seven attractors A_{1*}^i of $G^{A,u=0}$ are those listed in **Table 1** of this reference. According to Li et al. (2004), attractor A_{1*}^4 represents the G1 steady state and has the largest attraction basin. Attractor A_{11}^2 is also close to G1 phase and has the second largest attraction basin. Using the size of the attractions basins, the incidence probabilities w_A^i have been computed according to Equation (4) and they are listed in **Table 2**.

5.2.2. Network Interconnection, Asymptotic and Cross Graphs

To establish a scheme of interconnection, observe that the cell size model acts on the cell cycle by triggering the start signal, that is START is given by (folded/free) protein *Cln3F*. Conversely, the input of the cell cycle to the cell size module is still unknown, the combination of variables and/or quantities used by the cell to detect its own size is a question for further analysis. As an hypothesis, we will assume that growth rate is detected through cell phase, since the cell cycle model provides this information.

TABLE 2 | Interconnection of yeast models.

Attractor	Boolean representation	w_A^i
A_{11}^1	0000000000	0.0802
A_{11}^2	0000000100	0.0882
A_{11}^3	0000100000	0.0792
A_{1*}^4	00001000100	0.0893
A_{11}^5	0011000000	0.0669
A_{11}^6	0100000100	0.0472
A_{1*}^7	01001000100	0.0490
A_{2*}^8	10110110011	0.0921
A_{2*}^9	11000111011	0.1290
A_{2*}^{10}	11110110011	0.0749
A_{2*}^{11}	11110111011	0.2039
Attractor	Boolean representation	w_B^i
B_{12}^1	10100110	0.5
B_{21}^2	11011001	0.5

Attractors and incidence probabilities, proportional to the size of each basin of attraction. For module Σ_B , there exists a unique attractor for each v , hence $w_B^1 = w_B^2$. The symbol $*$ means that the output of this attractor depends on the function $h^A(a)$: we have A_{12}^4 with a G1 indicator but A_{11}^4 in the other two cases.

To explore the plausibility of this hypothesis, we will thus consider three different indicators of the cell cycle phase (M, S, and G1 phases) and compare the asymptotic graphs of the three

corresponding interconnection schemes:

$$\text{M-phase: } u = h^B(b) = \text{Cln3F}, \quad v = h^A(a) = \text{Swi5} \wedge \text{Cdc20}, \quad (13)$$

$$\text{S-phase: } u = h^B(b) = \text{Cln3F}, \quad v = h^A(a) = \text{Clb5}, \quad (14)$$

$$\text{G1-phase: } u = h^B(b) = \text{Cln3F}, \quad v = h^A(a) = \text{Cdh1} \wedge \text{Sic}. \quad (15)$$

In the case of growth measured by M phase ($h^A(a) = \text{Swi5} \wedge \text{Cdc20}$), the asymptotic graph has a unique, cyclic, attractor (**Figure 6, top**):

$$R_1^M = \{A_{22}^{11} \times B_{12}^1, A_{22}^{11} \times B_{21}^2, A_{11}^2 \times B_{21}^1, A_{11}^2 \times B_{12}^2, A_{11}^4 \times B_{21}^2, A_{11}^4 \times B_{12}^1\}$$

This information is confirmed and complemented by computation of the cross graph, which has 524,288 nodes ($= 2^{19}$). Attractor R_1^M is composed of 116,520 states.

Interestingly, although neither Σ^A nor Σ^B have periodic orbits, in this case the interconnected network does exhibit an oscillatory orbit: at stationary G1 (A_{11}^4) the START signal (B_{12}^1) is received and the module Σ^A performs one cell cycle:

$$A_{11}^4 \times B_{21}^2 \rightarrow A_{11}^4 \times B_{12}^1 \rightarrow A_{22}^{11} \times B_{12}^1 \rightarrow A_{22}^{11} \times B_{21}^2,$$

setting *Cln3* back to its OFF state (B_{21}^2) and ending “near” M phase (A_{22}^{11}). At this point, the system returns to stationary G1 and repeats the cycle, waiting for cell to grow and again send

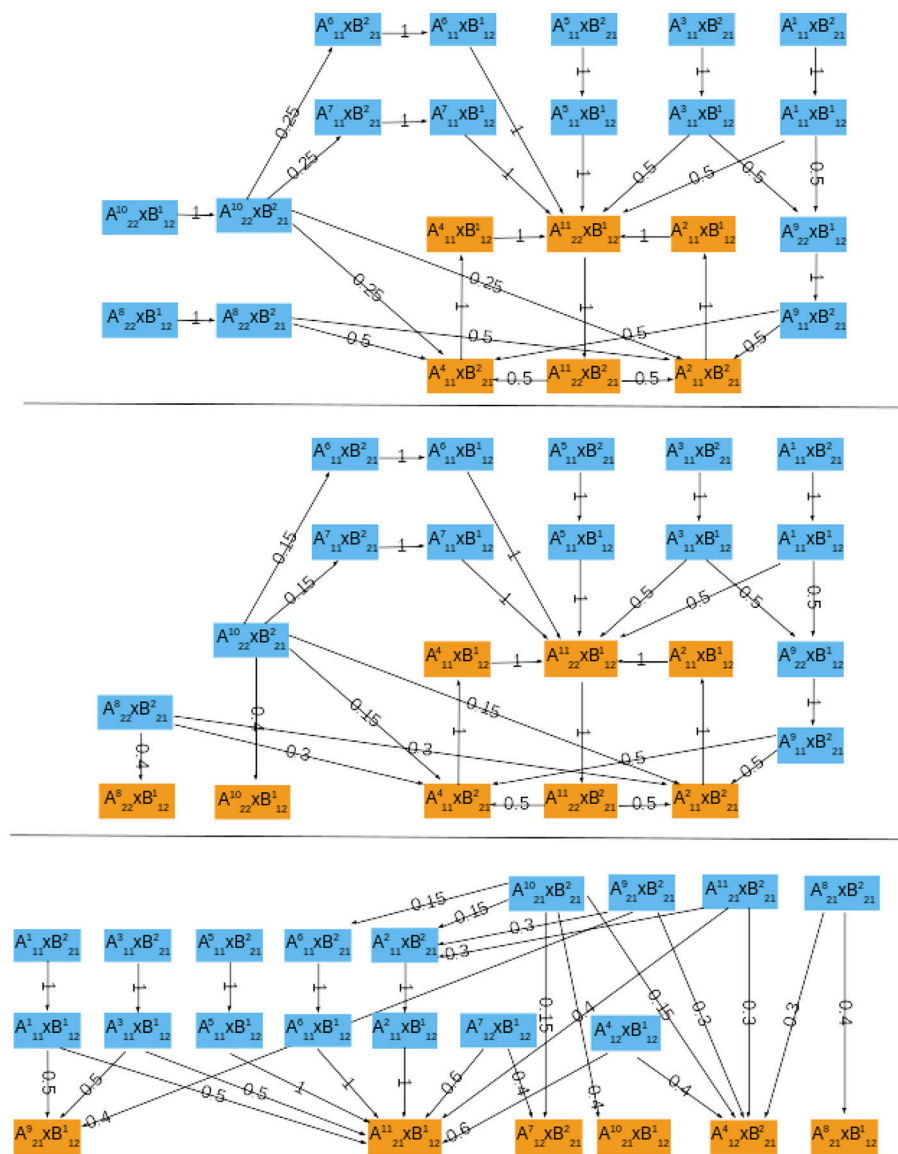


FIGURE 6 | The probabilistic asymptotic graphs for the interconnected yeast network, with growth rate measured by different indicators of the cell cycle. Orange colored nodes belong to an attractor. **(Top)** M phase indicator, there is exactly one (cyclic) attractor. **(Middle)** S/G2 phase indicator, there are two single state attractors and one cyclic attractor. **(Bottom)** G1 phase indicator, there are six single state attractors.

the start signal. Two alternative paths are proposed for the cell cycle, with G1-phase described either by A_{11}^4 or similar state A_{11}^2 . Since G^{as} contains a unique attractor, its relative probability P_{rel} is necessarily 1.

In the case of growth rate measured by S phase ($h^A(a) = Clb5$), the asymptotic graph (Figure 6, middle) has three attractors, two single steady state and one cyclic attractor:

$$\begin{aligned} R_1^S &= \{A_{22}^{11} \times B_{12}^1, A_{22}^{11} \times B_{21}^2, A_{11}^2 \times B_{21}^1, \\ &\quad A_{11}^2 \times B_{12}^1, A_{11}^4 \times B_{21}^2, A_{11}^4 \times B_{12}^1\} \\ R_2^S &= \{A_{21}^8 \times B_{12}^1\} \\ R_3^S &= \{A_{21}^{10} \times B_{12}^1\} \end{aligned}$$

In this case, however, computation of the cross graph shows that R_1^S is a spurious attractor, implying that the asymptotic graph has lost some information on transient pathways. In practice, the full graph contains pathways eventually leading from R_1^S to either R_2^S or R_3^S . This example shows the importance of verifying whether any of the asymptotic graph's attractors is spurious, and hence the usefulness of a complementary method as the cross graph. In this situation, the probabilistic interpretation of the asymptotic graph is unclear. The relative probabilities computed according to (8) yield equal probabilities for reaching attractors R_2^S and R_3^S (see Table 3). In contrast, R_1^S must now be interpreted as a transient set of states.

In the case G1 is used as measure of growth rate, we have $h^A(a) = Cdh1 \wedge Sic$ and the asymptotic graph (Figure 6, bottom) has six single state attractors but *no cyclic attractor*:

$$\begin{aligned} R_1^{G1} &= A_{12}^4 \times B_{21}^2, \quad R_2^{G1} = A_{12}^7 \times B_{21}^2, \\ R_3^{G1} &= A_{21}^8 \times B_{12}^1, \quad R_4^{G1} = A_{21}^9 \times B_{12}^1, \\ R_5^{G1} &= A_{21}^{10} \times B_{12}^1, \quad R_6^{G1} = A_{21}^{11} \times B_{12}^1. \end{aligned}$$

All these attractors are confirmed by the cross graph. Computation of relative probabilities shows that the single steady

state $A_{21}^{11} \times B_{12}^1$ is more frequently observed (with a percentage of around 54%, see Table 3). In this state all proteins of the cell cycle are expressed except for *Cdh1* and *Sic1*, which characterize stationary G1 phase. The cell growth module is in a state where *Cln3F* is available, thus setting START to 1. The interconnected system is thus locked in a steady state where the interaction links are fixed: $A_{21}^{11} \times B_{12}^1 = 11110111011 \times 10100110$, since the output of each attractor is equal to the input of the other.

5.2.3. Hypotheses Discrimination

These results appear to support a model for START signal of the form (12), as suggested by Aldea et al. (2017). Indeed, if cell size triggers START, then it can be assumed that there is a “critical size” which will be attained most probably at the end of G2 phase. And, in fact, the interconnected system exhibits an oscillatory cycle only in the case of M phase used as cell size indicator. This cycle is in agreement with cell cycle progression, meaning that the cell size module is able to trigger the START signal.

In contrast, when G1 or S phases are used as cell size indicator, the interconnected system has no oscillatory behavior. For the G1 case, the most frequent steady state ($A_{21}^{11} \times B_{12}^1$) represents a configuration where the cell size module permanently sets *Cln3F* = 1, and does not admit cell size to reset to zero. Note that G1 is the beginning of the cell cycle and a misleading indicator of “critical” size; in this case, the “critical” size is so small that the cell size module sets START permanently to 1 thus preventing the cell cycle to reset to zero and initiate a new cycle. Cells are locked in a steady state near mitosis and before early G1.

In conclusion, our analysis shows that neither G1 nor S phases are reliable cell growth indicators, but components from M phase are plausible candidates for detecting cell growth. We point out that the cell size Boolean network and the feedback interconnection points may admit several improvements, which are outside the scope of our paper. Nevertheless, we believe this first approach provides useful hints on how to further investigate and model the START signal in yeast.

6. DISCUSSION AND CONCLUSIONS

Our work illustrates a new concept for the analysis of an interconnection of Boolean networks: the goal is to study the coupled behavior of two or more modules, using only the dynamics of each separate module. A new methodology has been discussed, based on construction of the asymptotic and cross graphs both representative of the full network transition graph and guaranteed to compute all attractors of the interconnected network. The two graphs have different properties but also complement each other. The cross graph provides exact results, in the sense that it contains all transient and asymptotic behaviors of the interconnected network. The asymptotic graph is a lighter construction containing a minimal number of nodes while recovering all attractors. In contrast to the cross graph, no bijection with the full network transition graph is guaranteed, implying that spurious attractors may appear; however, this happens at an extremely low rate (less than 1%).

TABLE 3 | Attractors of the yeast interconnected system and their relative probabilities, $P_{rel}(R_i)$, for different updating probabilities ϱ_A .

Case S-phase output			
Attractor	$\varrho_A = 0.2$	$\varrho_A = 0.5$	$\varrho_A = 0.7$
$A_{21}^8 \times B_{12}^1$	0.1125	0.0938	0.0813
$A_{21}^{10} \times B_{12}^1$	0.1125	0.0938	0.0813
R_1^S	0.7750	0.8125	0.8375
Case G1-phase output			
Attractor	$\varrho_A = 0.2$	$\varrho_A = 0.5$	$\varrho_A = 0.7$
$A_{12}^4 \times B_{21}^2$	0.1042	0.1266	0.1415
$A_{12}^7 \times B_{21}^2$	0.0454	0.0401	0.0365
$A_{21}^8 \times B_{12}^1$	0.0829	0.0691	0.0599
$A_{21}^9 \times B_{12}^1$	0.1779	0.1585	0.1456
$A_{21}^{10} \times B_{12}^1$	0.0674	0.0562	0.0487
$A_{21}^{11} \times B_{12}^1$	0.5221	0.5495	0.5678

Construction of the two graphs for random input/output networks with varying connectivity reveals their complementarity in terms of modules' connectivity: for low connectivity ($K_{mean} \leq 5$), the asymptotic graph is much smaller (on average 0.01% of the full graph, against 28% for the cross graph; **Figure 2B**) and faster to compute; in contrast, for high connectivity ($K_{mean} > 5$), the size of the cross graph drastically reduces to 0.04% of the full graph (**Figure 2A**) becoming even faster to analyze than the asymptotic graph (**Figure 2C**). In addition, even though the asymptotic graph involves a drastic simplification of the state space, it has an unexpectedly high rate of accuracy, as shown in **Figure 2D**.

The practical advantages of our methodology are illustrated by the study of two well known biological networks. Among other useful characteristics, the asymptotic graph can greatly reduce the size of the state space, especially in the case of single-input single-output modules. For instance the mammalian and yeast interconnected networks, with an average connectivity of $K = 2.76$ and $K = 2.68$ respectively, have asymptotic graphs of only 18 and 22 nodes (compared to 2^{17} or 2^{19}).

The analysis of the coupling between cell cycle and circadian clock shows that, according to experimental observations (for instance by Plikus et al., 2013), the asymptotic graph predicts that mammalian cells in the quiescent state may have a working clock. Furthermore, under general hypotheses, the probabilistic approach predicts that one third of cells are arrested in the quiescent state but still have circadian oscillations, while the other two thirds follow a normal cell cycle progression intertwined with circadian oscillations. In the budding yeast example, we have explored a recent hypothesis by Aldea et al. (2017) for a mechanism to trigger the START signal and initiate cell cycle. The mechanism is based on cell size detection through cell growth

rate. Our analysis supports such a mechanism as a possible START trigger, and suggests that cell size indicator should come from an element during M phase.

The advantages of our analysis tools are multiple and particularly suited to the modeling of biological regulatory networks: by manipulating existing models as building blocks, the presented tools allow to rapidly simulate, compare, and test different coupling schemes or hypotheses on mutual regulatory effects, and therefore advance in the understanding of highly modular regulatory networks. The probabilistic interpretation and the analysis of transient behaviors emerge as two noteworthy directions for future developments in logical models.

AUTHOR CONTRIBUTIONS

MC and LT: equally contributed to conception, analysis and design of the study; MC and LT: wrote and revised the manuscript.

FUNDING

MC and LT are partly supported by the French agency for research through project ICycle ANR-16-CE33-0016-01. MC is partly funded by Labex Signallife ANR-11-LABX-0028-01.

SUPPLEMENTARY MATERIAL

The Boolean models used for mammalian and budding yeast cell cycles are provided as Supplementary Material.

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fphys.2018.00586/full#supplementary-material>

REFERENCES

- Abou-Jaoudé, W., Traynard, P., Monteiro, P. T., Saez-Rodriguez, J., Helikar, T., Thieffry, D., et al. (2016). Logical modeling and dynamical analysis of cellular networks. *Front. Genet.* 7:94. doi: 10.3389/fgene.2016.00094
- Albert, R., and Othmer, H. G. (2003). The topology of the regulatory interactions predicts the expression pattern of the *Drosophila* segment polarity genes. *J. Theor. Biol.* 223, 1–18. doi: 10.1016/S0022-5193(03)00035-3
- Aldea, M., Jenkins, K., and Csikász-Nagy, A. (2017). Growth rate as a direct regulator of the start network to set cell size. *Front. Cell Dev. Biol.* 5:57. doi: 10.3389/fcell.2017.00057
- Baldazzi, V., Ropers, D., Markowicz, Y., Kahn, D., Geiselman, J., and de Jong, H. (2010). The carbon assimilation network in *Escherichia coli* is densely connected and largely sign-determined by directions of metabolic fluxes. *PLoS Comput. Biol.* 6:e1000812. doi: 10.1371/journal.pcbi.1000812
- Béranguier, D., Chaouiya, C., Monteiro, P. T., Naldi, A., Remy, E., Thieffry, D., et al. (2013). Dynamical modeling and analysis of large cellular regulatory networks. *Chaos* 23:025114. doi: 10.1063/1.4809783
- Bieler, J., Cannavo, R., Gustafson, K., Gobet, C., Gatfield, D., and Naef, F. (2014). Robust synchronization of coupled circadian and cell cycle oscillators in single mammalian cells. *Mol. Syst. Biol.* 10:739. doi: 10.15252/msb.20145218
- Birkhoff, G. (1940). *Lattice Theory*, Vol. 25. Cambridge, MA: American Mathematical Society.
- Calzone, L., Tournier, L., Fourquet, S., Thieffry, D., Zhivotovsky, B., Barillot, E., et al. (2010). Mathematical modelling of cell-fate decision in response to death receptor engagement. *PLoS Comput. Biol.* 6:e1000702. doi: 10.1371/journal.pcbi.1000702
- Chaves, M., and Carta, A. (2015). Attractor computation using interconnected boolean networks: testing growth rate models in *E. coli*. *Theor. Comput. Sci.* 599, 47–63. doi: 10.1016/j.tcs.2014.06.021
- Chaves, M., and Tournier, L. (2011). "Predicting the asymptotic dynamics of large biological networks by interconnections of Boolean modules," in *Proceedings of the 50th conference Decision and Control and European Control conference* (Orlando, FL), 3026–3031.
- Comet, J.-P., Bernot, G., Das, A., Diener, F., Massot, C., and Cessieux, A. (2012). Simplified models for the mammalian circadian clock. *Proc. Comput. Sci.* 11, 127–138. doi: 10.1016/j.procs.2012.09.014
- Cormen, T., Leiserson, C., Rivest, R., and Stein, C. (2001). *Introduction to Algorithms*. Providence, RI: MIT Press, McGraw-Hill.
- Fauré, A., Naldi, A., Chaouiya, C., and Thieffry, D. (2006). Dynamical analysis of a generic boolean model for the control of the mammalian cell cycle. *Bioinformatics* 22, 124–131. doi: 10.1093/bioinformatics/btl210
- Feillet, C., Krusche, P., Tamanini, F., Janssens, R. C., Downey, M. J., Martin, P., et al. (2014). Phase locking and multiple oscillating attractors for the coupled mammalian clock and cell cycle. *Proc. Natl. Acad. Sci. U.S.A.* 111, 9828–9833. doi: 10.1073/pnas.1320474111
- Feillet, C., van der Horst, G. T., Lévi, F., Rand, D. A., and Delaunay, F. (2015). Coupling between the circadian clock and cell cycle oscillators: implication for healthy cells and malignant growth. *Front. Neurol.* 6:96. doi: 10.3389/fneur.2015.00096

- Feller, W. (1970). *An Introduction to Probability Theory and Its Applications*. New York, NY: Wiley.
- García-Gómez, M., Azpeitia, E., and Álvarez-Buylla, E. (2017). A dynamic genetic-hormonal regulatory network model explains multiple cellular behaviors of the root apical meristem of *Arabidopsis thaliana*. *PLoS Comput. Biol.* 13:e1005488. doi: 10.1371/journal.pcbi.1005488
- Gérard, C., and Goldbeter, A. (2012). Entrainment of the mammalian cell cycle by the circadian clock: modeling two coupled cellular rhythms. *PLoS Comput. Biol.* 8:e1002516. doi: 10.1371/journal.pcbi.1002516
- Kauffman, S. (1969). Metabolic stability and epigenesis in randomly constructed genetic nets. *J. Theor. Biol.* 22, 437–467. doi: 10.1016/0022-5193(69)90015-0
- Klarner, H., Bockmayr, A., and Siebert, H. (2015). Computing maximal and minimal trap spaces of Boolean networks. *Nat. Comput.* 14, 535–544. doi: 10.1007/s11047-015-9520-7
- Li, F., Long, T., Lu, Y., Ouyang, Q., and Tang, C. (2004). The yeast cell cycle is robustly designed. *Proc. Natl. Acad. Sci. U.S.A.* 101, 4781–4786. doi: 10.1073/pnas.0305937101
- Mendoza, L., and Xenarios, I. (2006). A method for the generation of standardized qualitative dynamical systems of regulatory networks. *Theor. Biol. Med. Model.* 3, 1–18. doi: 10.1186/1742-4682-3-13
- Mori, T., Flöttmann, M., Krantz, M., Akutsu, T., and Klipp, E. (2015). Stochastic simulation of boolean *rxncon* models: towards quantitative analysis of large signaling networks. *BMC Syst. Biol.* 9:45. doi: 10.1186/s12918-015-0193-8
- Plikus, M. V., Vollmers, C., de la Cruz, D., Chaix, A., Ramos, R., Panda, S., et al. (2013). Local circadian clock gates cell cycle progression of transient amplifying cells during regenerative hair cycling. *Proc. Natl. Acad. Sci. U.S.A.* 110, E2106–E2115. doi: 10.1073/pnas.1215935110
- Sabidussi, G. (1959). Graph multiplication. *Mathematische Zeitschrift* 72, 446–457. doi: 10.1007/BF01162967
- Saez-Rodriguez, J., Simeoni, L., Lindquist, J. A., Hemenway, R., Bommhardt, U., Arndt, B., et al. (2007). A logical model provides insights into T cell receptor signaling. *PLoS Comput. Biol.* 3:e163. doi: 10.1371/journal.pcbi.0030163
- Shmulevich, I., Dougherty, E. R., Kim, S., and Zhang, W. (2002). Probabilistic boolean networks: a rule-based uncertainty model for gene regulatory networks. *Bioinformatics* 18, 261–274. doi: 10.1093/bioinformatics/18.2.261
- Stoll, G., Caron, B., Viara, E., Dugourd, A., Zinovyev, A., Naldi, A., et al. (2017). Maboss 2.0: an environment for stochastic boolean modeling. *Bioinformatics* 33, 2226–2228. doi: 10.1093/bioinformatics/btx123
- Tournier, L., and Chaves, M. (2013). Interconnection of asynchronous Boolean networks, asymptotic and transient dynamics. *Automatica* 49, 884–893. doi: 10.1016/j.automatica.2013.01.015
- Turner, J., Ewald, J. C., and Skotheim, J. M. (2012). Cell size control in yeast. *Curr. Biol.* 22, R350–R359. doi: 10.1016/j.cub.2012.02.041
- Veliz-Cuba, A., Aguilar, B., Hinkelmann, F., and Laubenbacher, R. (2014). Steady state analysis of boolean molecular network models via model reduction and computational algebra. *BMC Bioinformatics* 15:221. doi: 10.1186/1471-2105-15-221
- Wang, R. S., Saadatpour, A., and Albert, R. (2012). Boolean modeling in systems biology: an overview of methodology and applications. *Phys. Biol.* 9:055001. doi: 10.1088/1478-3975/9/5/055001
- Zañudo, J., and Albert, R. (2013). An effective network reduction approach to find the dynamical repertoire of discrete dynamic networks. *Chaos* 23:025111. doi: 10.1063/1.4809777
- Zhang, R., Shah, M. V., Yang, J., Nyland, S. B., Liu, X., Yun, J. K., et al. (2008). Network model of survival signaling in large granular lymphocyte leukemia. *Proc. Natl. Acad. Sci. U.S.A.* 105, 16308–16313. doi: 10.1073/pnas.0806447105

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 Chaves and Tournier. 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.

APPENDIX

Proof of Theorem 2

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a digraph and let $v, v' \in \mathcal{V}$ be any two vertices of \mathcal{G} . Introduce the following notation:

- $v \rightarrow_{\mathcal{G}} v'$ means that there is an edge from v to v' in \mathcal{G} , i.e., $(v, v') \in \mathcal{E}$ (v' is a *successor* of v).
- $v \triangleright_{\mathcal{G}} v'$ means that there exists a path from v to v' in \mathcal{G} , i.e., there exist $k \geq 0$ vertices v_1, \dots, v_k such that $v = v_1 \rightarrow_{\mathcal{G}} v_2 \rightarrow_{\mathcal{G}} \dots \rightarrow_{\mathcal{G}} v_k = v'$ (v' is a *descendant* of v).
- $v \sim_{\mathcal{G}} v'$ means that there exists a path from v to v' and a path from v' to v in \mathcal{G} (v and v' are mutually reachable from each other). The relation $\sim_{\mathcal{G}}$ is an equivalence over $\mathcal{V} \times \mathcal{V}$.

Remark that according to the definition of partition Z_h^A , any A_α^i is included in a SCC of each graph $G^{A,u}$, in other words:

$$\forall a, a' \in A_\alpha^i, \forall u \in \{0, 1\}^{P_A}, a \sim_{G^{A,u}} a'.$$

For convenience, we introduce the two following maps π and ψ , establishing relationships between the two vertex sets V^{cr} and $\Omega = \{0, 1\}^{n_A + n_B}$.

- For $V = A_\alpha^i \times B_\beta^j \in V^{cr}$, let $\pi(V) := \{(a, b) | a \in A_\alpha^i, b \in B_\beta^j\} \subseteq \Omega$; and for $Q = \{V_1, \dots, V_k\} \subseteq V^{cr}$, define $\pi(Q) := \bigcup_{l=1}^k \pi(V_l) \subseteq \Omega$.
- For $x = (a, b) \in \Omega$, by definition of Z_h^A, Z_h^B there is a unique A_α^i and a unique B_β^j such that $A_\alpha^i \ni a, B_\beta^j \ni b$. Let $\psi(x) := A_\alpha^i \times B_\beta^j$; by extension, for $S \subseteq \Omega$, define $\psi(S) := \{\psi(x) | x \in S\} \subseteq V^{cr}$.

Theorem 2 is a consequence of the two following lemmas.

LEMMA 1. Let $x, y \in \Omega$ such that $x \triangleright_G y$, then either $\psi(x) = \psi(y)$, or $\psi(x) \triangleright_{G^{cr}} \psi(y)$.

PROOF: Suppose first that $x \rightarrow_G y$, that is to say either (i): $x = (a, b) \rightarrow_G (a', b) = y$ where $a \rightarrow_{G^{A, h^B(b)}} a'$ or (ii): $x = (a, b) \rightarrow_G (a, b') = y$ where $b \rightarrow_{G^{B, h^A(a)}} b'$. These two cases being perfectly symmetrical, consider for instance case (i). Let $A_\alpha^i, A_{\alpha'}^{i'}$ and B_β^j be respectively the (unique) sets such that $a \in A_\alpha^i, a' \in A_{\alpha'}^{i'}$ and $b \in B_\beta^j$. Two cases are to be considered.

Case 1: suppose $A_\alpha^i = A_{\alpha'}^{i'}$, then $\psi(x) = A_\alpha^i \times B_\beta^j = \psi(y)$.

Case 2: suppose $A_\alpha^i \neq A_{\alpha'}^{i'}$. Then according to the definition of

G^{cr} , from $a \rightarrow_{G^{A, \beta}} a'$ we deduce that $\psi(x) = A_\alpha^i \times B_\beta^j \rightarrow_{G^{cr}} A_{\alpha'}^{i'} \times B_\beta^j = \psi(y)$.

Suppose now that $x \triangleright_G y$, i.e., $x = x_1 \rightarrow_G x_2 \rightarrow_G \dots \rightarrow_G x_k = y$. By applying successively the previous result along that path, we deduce that either $\psi(x) = \psi(y)$ or $\psi(x) \triangleright_{G^{cr}} \psi(y)$, which concludes the proof. \square

LEMMA 2. Let $V, V' \in V^{cr}$ be two vertices of the cross graph.

- (i) $\forall x, y \in \pi(V), x \sim_G y$.
- (ii) If $V \triangleright_{G^{cr}} V'$, then for all $x \in \pi(V)$ and $y \in \pi(V')$, $x \triangleright_G y$.

PROOF: Let start with assertion (i). Let $V = A_\alpha^i \times B_\beta^j, x = (a, b) \in \pi(V)$ and $y = (a', b') \in \pi(V)$. Since a and a' both belong to the same $A_\alpha^i, a \sim_{G^{A, \beta}} a'$. In the same way, $b \sim_{G^{B, \alpha}} b'$. From there it is easy to verify that $(a, b) \sim_G (a', b') \sim_G (a', b')$, so $x \sim_G y$. Let us prove the second assertion. Suppose first that $V \rightarrow_{G^{cr}} V'$. For instance, let $V = A_\alpha^i \times B_\beta^j$ and $V' = A_{\alpha'}^{i'} \times B_{\beta'}^{j'}$ with $A_\alpha^i \ni a_1 \rightarrow_{G^{A, \beta}} a_2 \in A_{\alpha'}^{i'}$ (the symmetrical case can be treated completely analogously). Let $x = (a, b) \in \pi(V)$ and $y = (a', b') \in \pi(V')$. Since a and a_1 both belong to the same A_α^i , we have $a \sim_{G^{A, \beta}} a_1$. Similarly $a', a_2 \in A_{\alpha'}^{i'}$, implying $a' \sim_{G^{A, \beta}} a_2$. Therefore we have $a \sim_{G^{A, \beta}} a_1 \rightarrow_{G^{A, \beta}} a_2 \sim_{G^{A, \beta}} a'$, hence $(a, b) \triangleright_G (a', b)$. Now, since b and b' both belong to the same $B_\beta^j, b \sim_{G^{B, \alpha'}} b'$, which proves that $(a', b) \triangleright_G (a', b')$, therefore $x \triangleright_G y$.

Suppose now that $V \triangleright_{G^{cr}} V'$, i.e., $V = V_1 \rightarrow_{G^{cr}} V_2 \rightarrow_{G^{cr}} \dots \rightarrow_{G^{cr}} V_k = V'$. By applying successively the previous result along that path, we deduce that $x \triangleright_G y$ for any $x \in \pi(V)$ and $y \in \pi(V')$. \square

Lemmas 1 and 2 establish an exact correspondence between the paths in G and the paths in G^{cr} . The proof of the theorem becomes rather straightforward. Indeed, suppose $Q = \{V_1, \dots, V_k\}$ is a SCC of G^{cr} . Then Lemma 2 implies that $\pi(Q)$ is included in a SCC S of G . Suppose now that $\pi(Q) \subsetneq S$, i.e. there exists $y \in S \setminus \pi(Q)$ such that $\psi(y) \notin Q$. For any $x \in \pi(V) \subset S$, we have $x \sim_G y$ (with Lemma 1), implying $\psi(y) \in Q$ which is a contradiction. Therefore, $\pi(Q) = S$. Reciprocally, suppose $S \subseteq \Omega$ is a SCC of G . Then lemma 1 implies that $\psi(S)$ is included in a SCC Q of G^{cr} . Lemma 2 further yields $\psi(S) = Q$. By using a similar kind of reasoning, it is easy to show that there is an exact one-to-one correspondence between the terminal SCCs of G^{cr} and the attractors of G .