Introduction to the Process Hitting and inference of its underlying Biological Regulatory Network

Maxime Folschette[1]
Joint work: Loïc Paulevé, Katsumi Inoue, Morgan Magnin, Olivier Roux

[1] LUNAM Université, École Centrale de Nantes, IRCCyN UMR CNRS 6597
(Institut de Recherche en Communications et Cybernétique de Nantes)
1 rue de la Noë – B.P. 92101 – 44321 Nantes Cedex 3, France.
Maxime.Folschette@irccyn.ec-nantes.fr

Abstract

In this paper, the Process Hitting (PH), a recently introduced framework to model concurrent processes, is introduced. It is notably suitable to model Biological Regulatory Networks (BRNs) with partial knowledge of cooperations by defining the most permissive dynamics. On the other hand, the qualitative modeling of BRNs has been widely addressed using René Thomas' formalism, which is also depicted. A translation from PH to Thomas' representation of BRNs is finally presented. It relies on an analysis of all regulations to infer the Interaction Graph, then the possible parametrizations.

1 Introduction

As regulatory phenomena play a crucial role in biological systems, they need to be studied accurately. Biological Regulatory Networks (BRNs) consist in sets of either positive or negative mutual effects between the components. Besides continuous models of physicists, often designed through systems of ordinary differential equations, a discrete modeling approach was initiated by René Thomas in 1973 [16] allowing the representation of the different levels of a component, such as concentration or expression levels, as integer values. Nevertheless, these dynamics can be precisely established only with regard to some kind of "focal points", related to as Thomas' parameters, indicating the evolutionary tendency of each component. This modeling has motivated numerous works (see [12, 9, 15, 1]), and other approaches related to our work, which rely on temporal logic [7] and constraint programming [4, 5], aim at determining models consistent with partial data on the regulatory structure and dynamics. While the formal checking of dynamical properties is often limited to small networks because of the state graph explosion, the main drawback of this framework is the difficulty to specify Thomas' parameters, especially for large networks.

In order to address the formal checking of dynamical properties within very large BRNs, we recently introduced in [10] a new formalism, named the *"Process Hitting"* (PH), to model concurrent systems having components with a few qualitative levels. A PH describes, in an atomic manner, the possible evolutions of a "process" (representing one component at one level) triggered by the hit of at most one other "process" in the system. This particular structure makes the formal analysis of BRNs with hundreds of components tractable [11]. PH is suitable, according to the precision of this information, to model BRNs with different levels of abstraction by capturing the most general dynamics.

In [6] we showed that starting from one PH model, it is possible to find the underlying interactions, then the underlying Thomas' parameters. This method relies on an exhaustive search of the interactions between components of the PH model, and an enumeration of the (possibly large) nesting set of valid parameters, so that the resulting dynamics are ensured to respect the PH dynamics, i.e. no spurious transitions are made possible. The first benefit of this approach is that it makes possible the construction refining of BRNs with a partial and progressively brought knowledge in PH, while being able to export such models in the Thomas' framework. Our second contribution is to enhance the knowledge of the formal links between both modelings. The method can be applied to large BRNs (up to 40 components).

2 Frameworks

2.1 The Process Hitting framework

A Process Hitting (PH) (Def. 1) gathers a finite number of concurrent *processes* grouped into a finite set of *sorts*. A sort stands for a component of the system while a process, which belongs to a unique sort, stands for one of its expression levels. A process is noted a_i where a is the sort and i is the process identifier within the sort a. At any time, exactly one process of each sort is present; a *state* of the PH corresponds to such a set of processes.

The concurrent interactions between processes are defined by a set of *actions*. Actions describe the replacement of a process by another of the same sort conditioned by the presence of at most one other process in the current state. An action is denoted by $a_i \rightarrow b_j \nearrow b_k$, which is read as "$a_i$ *hits* b_j to make it bounce to b_k", where a_i, b_j, b_k are processes of sorts a and b, called respectively *hitter*, *target* and *bounce* of the action.

Definition 1 (Process Hitting) *A* Process Hitting *is a triple* $(\Sigma, L, \mathcal{H})$*, where:*

- $\Sigma = \{a, b, \dots\}$ *is the finite set of* sorts*;*

- $L = \prod_{a \in \Sigma} L_a$ *is the set of states with* $L_a = \{a_0, \ldots, a_{l_a}\}$ *the finite set of* processes *of sort* $a \in \Sigma$ *and* l_a *a positive integer, with* $a \neq b \Rightarrow L_a \cap L_b = \emptyset$*;*
- $\mathcal{H} = \{a_i \to b_j \Rsh b_k \in L_a \times L_b \times L_b \mid (a, b) \in \Sigma^2 \wedge b_j \neq b_k \wedge a = b \Rightarrow a_i = b_j\}$ *is the finite set of* actions.

Given a state $s \in L$, the process of sort $a \in \Sigma$ present in s is denoted by $s[a]$. An action $h = a_i \to b_j \Rsh b_k \in \mathcal{H}$ is *playable* in $s \in L$ if and only if $s[a] = a_i$ and $s[b] = b_j$. In such a case, $(s \cdot h)$ stands for the state resulting from the play of the action h in s, with $(s \cdot h)[b] = b_k$ and $\forall c \in \Sigma, c \neq b, (s \cdot h)[c] = s[c]$.

Modeling cooperation. As described in [10], the cooperation between processes to make another process bounce can be expressed in PH by building a *cooperative sort*. Fig. 1 shows an example of a cooperative sort bc between sorts b and c, defined with 4 processes (one for each sub-state of the presence of processes b_1 and c_1). For the sake of clarity, processes of bc are indexed using the sub-state they represent. Hence, bc_{01} represents the sub-state $\langle b_0, c_1 \rangle$, and so on. Each process of sort b and c hit bc, which makes it bounce to the process reflecting the status of the sorts b and c (e.g., $b_1 \to bc_{00} \Rsh bc_{10}$ and $b_1 \to bc_{01} \Rsh bc_{11}$). Then, to represent the cooperation between processes b_1 and c_1, the process bc_{11} hits a_1 to make it bounce to a_2 instead of independent hits from b_1 and c_1. The same cooperative sort is used to make b_0 and c_0 cooperate to hit a_1 and make it bounce to a_0.

Example 1 *Fig. 1 represents a PH* $(\Sigma, L, \mathcal{H})$ *with* $\Sigma = \{a, b, c, bc\}$*, and:*

$$L_a = \{a_0, a_1, a_2\}, \qquad L_b = \{b_0, b_1\},$$
$$L_{bc} = \{bc_{00}, bc_{01}, bc_{10}, bc_{11}\}, \qquad L_c = \{c_0, c_1\}.$$

This example models a BRN where the component a *has three qualitative levels, components* b *and* c *are Boolean and* bc *is a cooperative sort. In this BRN,* a *inhibits* b *at level* 2 *while* b *and* c *activate* a *with independent actions (e.g.* $b_0 \to a_2 \Rsh a_1$*) or through the cooperative sort* bc *(e.g.* $bc_{11} \to a_1 \Rsh a_2$*). Indeed, the reachability of* a_2 *and* a_0 *is conditioned by a cooperation of* b *and* c*, as explained above.*

A Process Hitting model can be obtained from the literature or from a BRN as described in [10]. In both methods, the identification of interactions allows to define the set of actions leading to the desired dynamics, but an under- or over-approximation can also be built if the interactions are not precisely known (by adding or removing all actions allowing a given behavior). This can be used especially in cases where a cooperative sort cannot be built because of a lack of information.

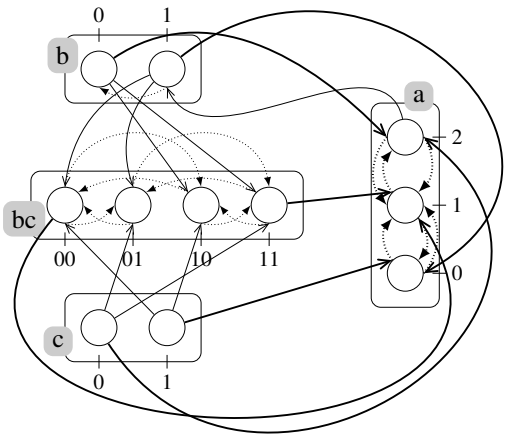


Figure 1: A PH example with four sorts: three components (a, b and c) and a cooperative sort (bc). Actions targeting processes of a are in thick lines.

2.2 Thomas' modeling

Thomas' formalism, here inspired by [13, 3], lies on two complementary descriptions of the system. First, the *Interaction Graph* (IG) models the structure of the system by defining the components' mutual influences. Its nodes represent components, while its edges labeled with a threshold stand for either positive or negative interactions (Def. 2); l_a denotes the maximum level of a component a.

Definition 2 (Interaction Graph) *An* Interaction Graph *(IG)* (Γ, E_+, E_-) *is a triple where:*

- Γ *is a finite number of* components*,*
- E_+ *(resp.* E_-*)* $\subset \{a \xrightarrow{t} b \mid a, b \in \Gamma \wedge t \in [1; l_a]\}$ *is the set of positive (resp. negative)* regulations *between two nodes, labeled with a* threshold*.*

A regulation from a *to* b *is unique, i.e. if* $a \xrightarrow{t} b \in E_+$ *(resp.* E_-*), then there is no regulation* $a \xrightarrow{t'} b$ *in* E_- *(resp.* E_+*), and no other regulation* $a \xrightarrow{t''} b$ *in* E_+ *(resp.* E_-*) with* $t'' \neq t$*.*

For an interaction of the IG to take place, the expression level of its head component has to be higher than its threshold; otherwise, the opposite influence is expressed. For any component $a \in \Gamma$, $\Gamma^{-1}(a)$ is the set of its regulators:

$$\Gamma^{-1}(a) = \{b \in \Gamma \mid \exists b \xrightarrow{t} a \in E_+ \cup E_-\} \ .$$

A *state* s of an IG (Γ, E_+, E_-) is an element in $\prod_{a\in\Gamma}[0; l_a]$ and $s[a]$ refers to the level of component a in s.

The specificity of Thomas' approach lies in the use of discrete *parameters* to represent focal level intervals (Def. 3). The use of intervals instead of single values for parameters allows a wider range of expressiveness, by allowing behaviors impossible to define with single values.

Definition 3 (Discrete parameter $K_{x,A,B}$ and Parametrization K) *Let $x \in \Gamma$ be a given component and A (resp. B) $\subset \Gamma^{-1}(x)$ a set of its* activators *(resp.* inhibitors*), such that $A \cup B = \Gamma^{-1}(x)$ and $A \cap B = \emptyset$. The discrete* parameter *$K_{x,A,B} = [i; j]$ is a non-empty interval so that $0 \leq i \leq j \leq l_x$. With regard to the dynamics, x will tend towards $K_{x,A,B}$ in the states where its activators (resp. inhibitors) are the regulators in set A (resp. B), except in the case where $x \in K_{x,A,B}$ for which it does not evolve.*
The complete map $K = (K_{x,A,B})_{x,A,B}$ of discrete parameters for an IG is called a parametrization *of this IG.*

At last, dynamics are defined in BRN in a unitary and asynchronous way: from a given state s, a transition to another state s' is possible provided that only one component a will evolve of exactly one level towards $K_{a,A,B}$, where A (resp. B) is the set of activators (resp. inhibitors) of a in s, provided that $a \notin K_{a,A,B}$ in s.

Example 2 *Fig. 2(left) represents the Interaction Graph (Γ, E_+, E_-) with $\Gamma = \{a, b, c\}$, and:*

$$E_+ = \{b \xrightarrow{1} a, c \xrightarrow{1} a\} \qquad\qquad E_- = \{a \xrightarrow{2} b\} \ .$$

In particular, $\Gamma^{-1}(a) = \{b, c\}$. Fig. 2(right) gives a possible parametrization of this IG. In this BRN, the following transitions are possible:

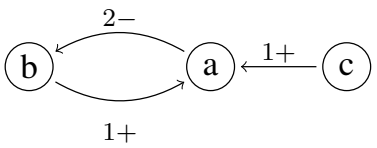
$$\langle a_0, b_1, c_1\rangle \to \langle a_1, b_1, c_1\rangle \to \langle a_2, b_1, c_1\rangle \to \langle a_2, b_0, c_1\rangle \to \langle a_1, b_0, c_1\rangle,$$

where a_i is the component a at level i.

3 BRN Inference

This section focuses on the inference of a complete BRN with Thomas' parameters from a given PH.

In order to infer a BRN, one has to find the Interaction Graph (IG) first, as some constraints on the parametrization rely on it. Inferring the IG is an



$$K_{a,\{b,c\},\emptyset} = [2;2] \qquad K_{b,\{a\},\emptyset} = [0;1]$$
$$K_{a,\{b\},\{c\}} = [1;1] \qquad K_{b,\emptyset,\{a\}} = [0;0]$$
$$K_{a,\{c\},\{b\}} = [1;1]$$
$$K_{a,\emptyset,\{b,c\}} = [0;0] \qquad K_{c,\emptyset,\emptyset} = [0;1]$$

Figure 2: (left) IG example. Regulations are represented by the edges labeled with their sign and threshold. For instance, the edge from b to a is labeled "1+", which stands for: $b \xrightarrow{1} a \in E_+$. (right) One admissible parametrization of the left IG.

abstraction step which consists, from atomistic actions of a PH, in determining the global influence of every component on each of its successors.

Then, given the IG inferred from a PH, one can find the discrete parameters that model the behavior of the studied PH. As some parameters may remain undetermined, another step allows to enumerate all parametrizations compatible with the inferred parameters.

3.1 Interaction Graph inference

This step assumes that the studied PH defines two types of sorts: the sorts corresponding to BRN components, which will appear in the IG, and the cooperative sorts, as defined in Subsect. 2.1. The identification of these two sets of sorts relies on the observation of their possible behavior, which in both cases observe some rules, and can be automated. For instance, given the definition of cooperative sorts, if the actions hitting a sort lead to a unique fixed point for any configuration of its predecessors, then we can deduce that this sort is a cooperative sort. Conversely, because of the BRN dynamics explained in Subsect. 2.2, if all actions hitting a sort make its processes bounce at most one level away (e.g. if a_1 can bounce to a_0 or a_2 but not to a_3), then this sort is likely to correspond to a BRN component.

Inferring global influences of a predecessor b on a component a requires to find "local influences" from this predecessor first, by considering a given state of the PH and changing only the active process of b. The aim is to compare the set of processes towards which the component a will evolve, for each active process of b, leaving the active process of all the other sorts unchanged. Indeed, if after increasing the level of b, i.e. activating a higher process of b, we notice that a tends to reach a higher (resp. lower) level, we can then deduce that b activates (resp. inhibits) a in this selected state. Of course, only predecessors of a have to be considered.

This has to be observed on every possible state in order to infer a local influence. Indeed, if all local influences of b on a are the same (activations or inhibitions) we can deduce that the global influence of b on a is also the same, and the related threshold is the lowest level of b for which we can observe such an influence. An unsigned edge with no threshold is inferred if two different local influences are found, or in other particular cases (when a behavior cannot be represented as a BRN).

Example 3 *Consider, in the PH of Fig. 1, the sub-state $\sigma = \langle b_0, c_0, bc_{00}\rangle$ of predecessors of a. In this sub-state, a can be hit by the following actions:*

$$\mathcal{H}_a^\sigma = \{b_0 \to a_2 \Rsh a_1, c_0 \to a_2 \Rsh a_1, bc_{00} \to a_1 \Rsh a_0\} \ .$$

Thus, if a evolves, it will eventually reach process a_0. But if a higher process of b is activated, that is, b_1 instead of b_0, thus considering the sub-state $\sigma' = \langle b_1, c_0, bc_{10}\rangle$, then a can be hit by the two following actions:

$$\mathcal{H}_a^{\sigma'} = \{b_1 \to a_0 \Rsh a_1, c_0 \to a_2 \Rsh a_1\} \ ,$$

and will eventually reach process a_1.

Therefore, in this sub-state of predecessors of a, b locally activates a. Furthermore, if this analysis is carried for all possible sub-states of predecessors of a, only local activations are found, thus giving: $b \xrightarrow{1} a \in E_+$.

After applying this method to all pairs of influence, the IG given in Fig. 2 is inferred.

3.2 Parameters inference

This subsection presents some results related to the inference of independent discrete parameters from a given PH, equivalent to those presented in [10]. We suppose in the following that the considered PH is well-formed for parameters inference, i.e. its inferred IG does not contain any unsigned edge, and in each sort, all processes activating (resp. inhibiting) another component share the same behavior. Let $K_{a,A,B}$ be the parameter we want to infer for a given component $a \in \Gamma$, and $A \subset \Gamma^{-1}(a)$ (resp. $B \subset \Gamma^{-1}(a)$) a set of its activators (resp. inhibitors). This inference, as for the IG inference, relies on the search of processes of a towards which it will eventually evolve for the given configuration A, B of its regulators.

For each sort $b \in \Gamma^{-1}(a)$, we define a context that contains all processes of b activating (resp. inhibiting) a if $b \in A$ (resp. B). From all contexts of all predecessors of a, we create a global context $C_{A,B}$ that represents the configuration A, B (including the cooperative sorts involved). The parameter $K_{a,A,B}$ specifies towards which values a eventually evolves as long as the

configuration A, B holds, which can now be computed by considering the dynamics of a in the global context $C_{A,B}$.

Example 4 *Consider the PH of Fig. 1, from which the IG of Fig. 2 is inferred. Inferring the parameter $K_{a,\{b,c\},\emptyset}$ requires to understand the behavior of a in the sub-state $\langle b_1, c_1, bc_{11}\rangle$. In this sub-state, a tends to eventually reach process a_2; thus, we can deduce the parameter: $K_{a,\{b,c\},\emptyset} = [2; 2]$. Inferring all parameters leads to the complete parametrization given in Fig. 2.*

3.3 Admissible parametrizations enumeration

The previous inference step may leave several parameters undetermined, due to missing cooperations or behaviors impossible to represent in a BRN. If it is not possible to change the PH model in order to remove these inconclusive cases, one can perform a last step to enumerate all valid values for each parameter that could not be inferred given the above results. We consider that a parameter is valid if any transition it involves in the resulting BRN is allowed by the studied PH by actions that represent this behavior. We also add some biological constraints on the whole parametrizations, given in [3]. These constraints lead to a family of admissible parametrizations which we can enumerate and are ensured to observe a coherent behavior that is included in the original PH.

Answer Set Programming (ASP) [2] turns out to be effective for the enumerative searches developed in this paper, as it efficiently tackles the inherent complexity of the models we use, thus allowing an efficient execution of the formal tools developed. Furthermore, ASP finds a particularly interesting application in the research of admissible parametrizations regarding the properties presented above, as this enumeration can be naturally formulated by using of aggregates and constraints.

3.4 Implementation

The inference method described in this paper has been implemented as a tool named `ph2thomas`, as part of PINT[1], a library gathering PH related tools. Our implementation mainly consists of ASP programs that are solved using Clingo[2].

In the previous sections, the methods and results are illustrated on a toy example considered as a very small network containing 3 components (a, b and c). But our approach can also successfully handle large PH models of BRNs found in the literature such as an ERBB receptor-regulated G1/S transition model from [14] which contains 20 components, and a T-cells receptor model

[1] Available at `http://process.hitting.free.fr`
[2] Available at `http://potassco.sourceforge.net`

from [8] which contains 40 components[3]. For each model, IG and parameters inferences are performed together in less than a second on a standard desktop computer.

4 Conclusion

This work establishes the abstraction relationship between PH, which is more abstract and allows incomplete knowledge on cooperations, and Thomas' approach for qualitative BRN modeling. This motivates the concretization of PH models into a set of compatible Thomas' models in order to benefit from the complementary advantages of these two formal frameworks and extract some global information about the influences between components.

As an extension of the present work, we plan to explore new semantics of BRNs to be able to tackle influences currently represented by unsigned edges.

Acknowledgment. This work was partially supported by the Fondation Centrale Initiatives.

References

[1] Jamil Ahmad, Olivier Roux, Gilles Bernot, Jean-Paul Comet, and Adrien Richard. Analysing formal models of genetic regulatory networks with delays. *International Journal of Bioinformatics Research and Applications (IJBRA)*, 4(2), 2008.

[2] Chitta Baral. *Knowledge Representation, Reasoning and Declarative Problem Solving*. Cambridge University Press, 2003.

[3] Gilles Bernot, Franck Cassez, Jean-Paul Comet, Franck Delaplace, Céline Müller, and Olivier Roux. Semantics of biological regulatory networks. *Electronic Notes in Theoretical Computer Science*, 180(3):3 – 14, 2007.

[4] Fabien Corblin, Eric Fanchon, and Laurent Trilling. Applications of a formal approach to decipher discrete genetic networks. *BMC Bioinformatics*, 11(1):385, 2010.

[5] Fabien Corblin, Eric Fanchon, Laurent Trilling, Claudine Chaouiya, and Denis Thieffry. Automatic inference of regulatory and dynamical properties from incomplete gene interaction and expression data. In *IPCAT*, volume 7223 of *LNCS*, pages 25–30. Springer, 2012.

[3] Both models are available as examples distributed with PINT.

[6] Maxime Folschette, Loïc Paulevé, Katsumi Inoue, Morgan Magnin, and Olivier Roux. Concretizing the process hitting into biological regulatory networks. In David Gilbert and Monika Heiner, editors, *Computational Methods in Systems Biology*, Lecture Notes in Computer Science, pages 166–186. Springer Berlin Heidelberg, 2012.

[7] Z. Khalis, J.-P. Comet, A. Richard, and G. Bernot. The SMBioNet method for discovering models of gene regulatory networks. *Genes, Genomes and Genomics*, 3(special issue 1):15–22, 2009.

[8] Steffen Klamt, Julio Saez-Rodriguez, Jonathan Lindquist, Luca Simeoni, and Ernst Gilles. A methodology for the structural and functional analysis of signaling and regulatory networks. *BMC Bioinformatics*, 7(1):56, 2006.

[9] Aurélien Naldi, Elisabeth Remy, Denis Thieffry, and Claudine Chaouiya. A reduction of logical regulatory graphs preserving essential dynamical properties. In *Computational Methods in Systems Biology*, volume 5688 of *LNCS*, pages 266–280. Springer, 2009.

[10] Loïc Paulevé, Morgan Magnin, and Olivier Roux. Refining dynamics of gene regulatory networks in a stochastic π-calculus framework. In *Transactions on Computational Systems Biology XIII*, pages 171–191. Springer, 2011.

[11] Loïc. Paulevé, Morgan Magnin, and Olivier Roux. Static analysis of biological regulatory networks dynamics using abstract interpretation. *Mathematical Structures in Computer Science*, in press, 2012. Preprint: `http://loicpauleve.name/mscs.pdf`.

[12] Adrien Richard and Jean-Paul Comet. Necessary conditions for multistationarity in discrete dynamical systems. *Discrete Applied Mathematics*, 155(18):2403 – 2413, 2007.

[13] Adrien Richard, Jean-Paul Comet, and Gilles Bernot. *Modern Formal Methods and App.*, chapter Formal Methods for Modeling Biological Regulatory Networks, pages 83–122. 2006.

[14] Ozgur Sahin, Holger Frohlich, Christian Lobke, Ulrike Korf, Sara Burmester, Meher Majety, Jens Mattern, Ingo Schupp, Claudine Chaouiya, Denis Thieffry, Annemarie Poustka, Stefan Wiemann, Tim Beissbarth, and Dorit Arlt. Modeling ERBB receptor-regulated G1/S transition to find novel targets for de novo trastuzumab resistance. *BMC Systems Biology*, 3(1), 2009.

[15] Heike Siebert and Alexander Bockmayr. Incorporating time delays into the logical analysis of gene regulatory networks. In *Computational Methods in Systems Biology*, volume 4210 of *LNCS*, pages 169–183. Springer, 2006.

[16] René Thomas. Boolean formalization of genetic control circuits. *Journal of Theoretical Biology*, 42(3):563 – 585, 1973.