Using alternated sums to express the number of occurrences of extended patterns in site-graphs [1]

Ferdinanda Camporesi[2]

Département d'informatique
École normale supérieure,
École normale supérieure, CNRS, PSL Research University, 75 005 Paris, France

Jérôme Feret[3]

Département d'informatique
École normale supérieure,
École normale supérieure, CNRS, PSL Research University, 75 005 Paris, France

Abstract

Site-graph rewriting languages as Kappa or BNGL supply a convenient way to describe models of signalling pathways. Unlike classical reaction networks, they emphasise on the biochemical structure of proteins. We use patterns to formalise properties about bio-molecular species. Intentionally, a pattern is a part of a species, but extensionally it denotes the multi-set of the species containing this pattern (with the multiplicity). Thus reasoning on patterns allows to handle symbolically arbitrarily big (if not infinite) multi-sets of species. This is a key point to design fast simulation algorithms or model reduction schemes. In this paper, we introduce the notion of extended patterns. Each extended pattern is made of a classical pattern and of a set of potential bonds between pairs of sites. Extended patterns have positive (when at least one of the potential bonds is realised) and negative (when none is realised) instances. They are important to express the consumption and the production of patterns by the rules that may break cycles in bio-molecular species by side-effects. We show that the number of positive (resp. negative) instances of extended patterns may be expressed as alternated sums of the number of occurrences of classical patterns.

Keywords: Rule-based modeling, structural invariants

1 Introduction

Site-graph rewriting languages as Kappa [9] or BNGL [1] supply a convenient way to describe models of signalling pathways. Unlike classical reaction networks, they

[1] This material is based upon works partially sponsored by the Defense Advanced Research Projects Agency (DARPA) and the U. S. Army Research Office under grant number W911NF-14-1-0367, and by the ITMO Plan Cancer 2014. The views, opinions, and/or findings contained in this article are those of the authors and should not be interpreted as representing the official views or policies, either expressed or implied, of DARPA, the U. S. Department of Defense, or ITMO.

[2] Email: campores@di.ens.fr

[3] Email: feret@di.ens.fr

emphasise on the biochemical structure of proteins. We use patterns to formalise properties about bio-molecular species. From an intentional perspective, a pattern is a part of a biochemical species. A pattern expresses some conditions over the states of sites in proteins. Interestingly, it permits to reason locally on a bio-molecular species. In reaction networks, we can say that a species is consumed or produced, with site-graph rewriting, we can say that a species is transformed by a local transformation. Beyond the capability to express large networks in a compact way, this also gives rise to more compact notions of causalities as expressed by stories [4] or influence maps [5]. Thanks to this, site-graph rewriting models may be simulated efficiently, by using data-structures that quickly maintain the number of potential embeddings [7,16,2], and this without ever compiling models into reaction networks. From an extensional perspective, a pattern is a (potentially infinite) linear combination of bio-molecular species (the ones that contain the pattern multiplied by the number of their occurrences). This opens the door to many algebraic relationships. One class of them comes from orthogonal refinement [11,13]. This consists in refining a given pattern, while considering all the potential states for the sites and the agents that are inserted. Gluing allows for the specialisation of a given rule to the consumption or to the production of a given pattern, so as to express the derivative of the concentration (in the differential semantics) of this pattern as an expression of the other patterns [6,14,10]. This opens the door to model reduction [11,6,14,3], where constraints, collected by static inspection of the rules (without ever considering the underlying reaction networks) are used to define sets of self-consistent patterns, the concentration of which may be defined by ODEs that only depend on these patterns. These constraints cope with the flow of information (how the state of some sites may control the modification of the state of other sites), backward compatibility [3] (that ensures that always more information is kept about a reactant than about the corresponding product), and also, quite inelegantly, about the ways cycles in bio-molecular species may be broken by side-effects [6,3].

In this paper, we discover a new class of algebraic equalities among the number occurrences of patterns. We introduce a class of patterns, called extended patterns. An extended pattern is made of a classical pattern and of a set of potential bonds between pairs of sites. An extended pattern may be seen as a symbolic representation of the set of the patterns that may be obtained by inserting in the initial pattern a subset of the potential bonds. Conversely, each instance of the initial pattern in another pattern, is associated with the subset of the potential bonds that are realised in the latter pattern. Extended patterns play an important role in expressing the consumption and the production of patterns by a degradation rule in case of cyclic bio-molecular species. An instance of a pattern may be consumed if it is connected to a protein that is degraded, no matter how many bonds it shares with this protein. Conversely, when a pattern is created by side-effects, we have to consider the potential original configurations of this pattern and focus on the instances of the initial pattern in which no other bonds are realised. To address this issue, we introduce the notions of positive (when at least one of the potential bond is realised) and negative (when none of the potential bond is realised) instances of an extended pattern. We show that the number of positive (resp. negative) instances of each extended pattern may be expressed as an alternated sum of the

number of occurrences of the patterns that are obtained by inserting a subset of the potential bonds in the initial pattern. Interestingly, the coefficient of each term in these combinations is either 1 or -1 depending on the evenness of the number of the potential bonds that have been inserted. This allows for expressing new relationships between the number of occurrences of patterns. In particular, we can get rid of the constraints related to cycles in bio-molecular species in model reduction [6,3]. This is important because any spurious constraint may have a huge impact by snowball effect, when the other constraints coming from the flow of information and backward compatibility are considered. As a result, we can get a more elegant formulation only relying on the flow of information and backward compatibility and we obtain a more compact model reduction.

Outline.

In Sect. 2, we introduce a toy example so as to motivate our framework. In Sect. 3, we recall the notion of embedding between patterns in site-graph rewriting. In Sec. 4, we introduce extended patterns, as patterns that may be refined by inserting some potential bonds, and relate the number of instances of the positive instances of an extended pattern (with at least one of the potential bonds inserted) and of the negative instances of an extended pattern (with none of the potential bonds inserted) to the number of instances of other patterns.

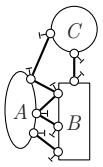
2 Case study

We introduce a toy example so as to motivate our framework.

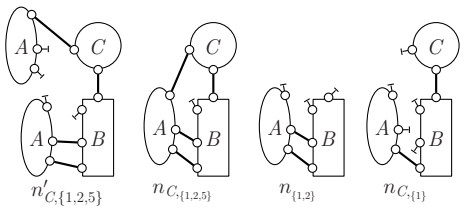
We consider three kinds of protein A, B, and C. Proteins of kind A have three binding sites identified with labels 1, 2, and 3; proteins of kind B have four binding sites identified with labels 1, 2, 3, and 4; and proteins of kind C have two binding sites identified with labels 1 and 2. Each binding site may be free, or bound to a single binding site in another protein. Not every bond is possible. We assume that only the following potential bonds are admitted: between sites 1 of proteins A and the sites 1 of proteins B; between sites 2 of proteins A and the sites 2 of proteins B; between sites 2 of proteins A and the sites 3 of proteins B; between sites 3 of proteins A and the sites 3 of proteins B; between sites 3 of proteins A and the sites 2 of proteins C; between sites 4 of proteins B and the sites 1 of proteins C.

These potential bonds are summarised in a graph, depicted in Fig. 1(a), that we call the contact map of the model. The contact map shows each kind of protein. Kinds of protein are drawn as big geometric shapes (ellipses, rectangles, circles). The sites of each kind of protein are drawn as small circles on the border of the protein in increasing order of their labels from the bottom up (site identifiers are omitted). Potential bonds between pairs of sites are denoted by undirected edges. Every site has a small '$\dashv$' symbol to denote the fact that it may either be free.

In Fig. 1(b), we show a potential state of the system. This state is made of several instances of protein; each instance documents its full set of binding sites. Some sites are free '$\dashv$' and some pairs of distinct sites are bound together. The state of the system may be split into connected components that we call bio-molecular species. Each bio-molecular species comes with a parametric quantity, written just

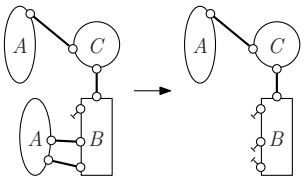


(a) Contact map.



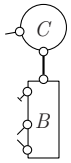
(b) A state.

(c) Degradation rule (for the proteins A).

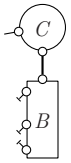


(d) One refinement of the degradation rule.

Fig. 1. The case study. In Fig. 1(a), the contact map summarises the different kinds of agent, their sites, and the potential bonds among them. In Fig. 1(b), we draw a potential state of the system. In Fig. 1(c), we draw a rule that may degradate proteins of kind A. In Fig. 1(d), we specialise this rule to the degradation of each protein A that is bound twice via their respective two lower sites to a protein B itself bond to a protein C bond to another protein A.

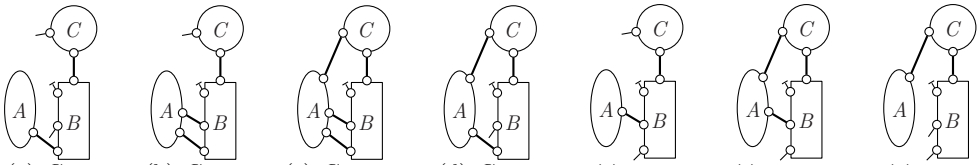


(a) Which amount of this pattern is consumed?



(b) Which amount of this pattern is produced?

Fig. 2. Production and consumption of patterns.



(a) $G_{\{1\}}$ (b) $G_{\{1,2\}}$ (c) $G_{\{1,2,5\}}$ (d) $G_{\{1,5\}}$ (e) $G_{\{2\}}$ (f) $G_{\{2,5\}}$ (g) $G_{\{5\}}$

Fig. 3. Patterns of interest.

bellow it, that may be, according to the choice of the semantics, either a number of occurrences (stochastic setting) or a non negative real number (differential setting).

We consider that each protein of kind A may be degraded at propensity or at rate 1 (according to the choice of semantics). In Fig. 1(c), we draw the corresponding degration rule. By side-effects, each degradation of a protein A will modify the states of the proteins of kinds B and C by releasing the sites that were bound to the degraded protein. In Fig. 1(d), we specialise this rule to the degradation of each protein A that is bound twice on its two lower sites to a protein B itself bond to a protein C bond to another protein A. This is not the only refinement of the rule in Fig. 1(c). We have picked a random one so to provide an example.

We consider two patterns in Fig. 2. We would like to express concisely the propensity or the rate (according to the semantics) of consumption of the pattern given in Fig. 2(a). Let us call N_{-} this quantity. We assume that the system is in the

state that is given in Fig. 1(b). Our pattern matches only the first two bio-molecular species. The first one contains two proteins of kind A. It follows that:

$$N_- = 2n'_{C,\{1,2,5\}} + n_{C,\{1,2,5\}}.$$

We notice that the previous expression may involve bio-molecular patterns with several instances of the protein A. This may raise issues by snowball effect in more combinatorial models. Indeed, it may be required to document many other sites in the proteins of kind A, which would make the computation unscalable. Instead we propose to focus on the patterns that are described in Fig. 3. Some sites are bound without specifying the site they are actually bound to (symbol '-'). This allows to include only one instance of A per pattern. The number of occurrences (or the concentration) of each pattern may be expressed as a linear combination of the concentration of each bio-molecular species, which gives the following equalities [4]:

(i) $n_{G_{\{1\}}} = n'_{C,\{1,2,5\}} + n_{C,\{1,2,5\}}$;
(ii) $n_{G_{\{2\}}} = n'_{C,\{1,2,5\}} + n_{C,\{1,2,5\}}$;
(iii) $n_{G_{\{5\}}} = n'_{C,\{1,2,5\}} + n_{C,\{1,2,5\}}$;
(iv) $n_{G_{\{1,2\}}} = n'_{C,\{1,2,5\}} + n_{C,\{1,2,5\}}$;
(v) $n_{G_{\{1,5\}}} = n_{C,\{1,2,5\}}$;
(vi) $n_{G_{\{2,5\}}} = n_{C,\{1,2,5\}}$;
(vii) $n_{G_{\{1,2,5\}}} = n_{C,\{1,2,5\}}$.

The following computation:

$$\begin{aligned}
N_- &= 2n'_{C,\{1,2,5\}} + n_{C,\{1,2,5\}} \\
N_- &= (3-1)n'_{C,\{1,2,5\}} + (3-3+1)n_{C,\{1,2,5\}} \\
N_- &= n_{G_{\{1\}}} + n_{G_{\{2\}}} + n_{G_{\{5\}}} - n_{G_{\{1,2\}}} - n_{G_{\{1,5\}}} - n_{G_{\{2,5\}}} + n_{G_{\{1,2,5\}}}
\end{aligned}$$

shows that the quantity N_- may be expressed as a linear combination of the number of the occurrences (or of the concentration) of the patterns in Fig. 3. We also notice that each pattern occurs with the coefficient 1 when the protein A has a odd number of bonds, and with the coefficient -1 when this number is even.

Now we would like to express concisely the propensity or the rate (according to the choice of the semantics) of production of the pattern given in Fig. 2(b). We restrict ourselves to the cases when the pattern is created due to the releasing of the two lower sites of the protein B whereas the third site was free already (so as to get the overall production of the pattern, we would have to consider all the combinations for the sites that have been released by side effects [10], we consider only this case for the sake of the example). Let us call N_+ this quantity. We assume that the system is in the state in Fig. 1(b). In this state, only the first bio-molecular species may produce our pattern in these conditions. It follows that:

$$N_+ = n'_{C,\{1,2,5\}}.$$

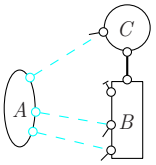
[4] We have neglected the bio-molecular species that do not occur in the state that we consider.

The following computation:

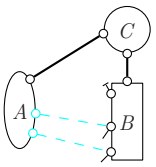
$$N_+ = n'_{C,\{1,2,5\}}$$
$$N_+ = (1-0)n'_{C,\{1,2,5\}} + (1-2+1)n_{C,\{1,2,5\}}$$
$$N_+ = n_{G_{\{5\}}} - n_{G_{\{1,5\}}} - n_{G_{\{2,5\}}} + n_{G_{\{1,2,5\}}}$$

shows that the quantity N_+ may be expressed as a linear combination of the number of the occurrences (or of the concentration) of the patterns in Fig. 3. To understand the multiplicative coefficient of each term, we have to consider the number of bonds that had to be inserted to the pattern $G_{\{5\}}$. The coefficient is equal to 1, whenever it is even, or equal to -1 otherwise.

In the rest of paper, we show that this approach can be generalised. Indeed the computation of N_- may be understood through the following pattern:



that we have annotated by a set of potential bonds. The quantity N_- is obtained by considering every pattern instance that realises at least one of the potential bonds. For understanding the computation of N_+, we shall take the following pattern:



also annotated by a set of potential bonds. We have considered every pattern instance that would realise none of the potential bonds. Such quantities are always equal to alternated sums of the number of occurrences (or of the concentration) of the patterns that are obtained by inserting a subset of the potential bonds, with a sign that depends on the evenness of the cardinal of this subset.

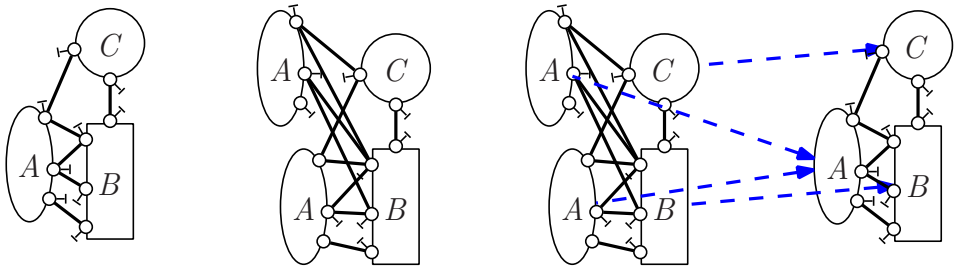
3 Site-graphs

In this section, we give some reminders about Kappa. Since we focus on counting some specific occurrences of patterns, we do not introduce the full semantics of Kappa. Instead, we introduce only the notions of site-graphs and of embeddings among them, and we omit the notions of rule and of rule applications. We also omit internal states, since dealing with them would raise no difficulty. We refer to [9,12] for a more complete description of Kappa.

3.1 Signature

Firstly we define the signature of a model.

Definition 3.1 (signature) *A signature is a triple* $\Sigma \triangleq (\Sigma_{ag}, \Sigma_{site}, \Sigma_{ag\text{-}st})$ *where:*



(a) The Σ-graph G_{CM}. (b) The Σ-graph G_Σ. (c) A morphism from G_Σ into G_{CM}

Fig. 4. Two Σ-graphs G_{CM} and G_Σ, and a morphism from G_{CM} to G_Σ. The Σ-graph G_{CM} is a contact map. It provides context-insensitive information about the potential state of each binding site. The Σ-graph G_Σ provides context-sensitive information. The potential states of the sites of an agent A may depend on wether the site on the bottom right is free or phosphorylated.

(i) Σ_{ag} *is a finite set of agent types,*

(ii) Σ_{site} *is a finite set of site identifiers;*

(iii) $\Sigma_{ag\text{-}st} : \Sigma_{ag} \to \wp(\Sigma_{site})$ *is a site map.*

Agent types in Σ_{ag} denote agents of interest, as kinds of proteins for instance. Site identifiers in Σ_{site} represent identified loci for capabilities of interactions. Agent types $A \in \Sigma_{ag}$ are associated with sets of sites $\Sigma_{ag\text{-}st}(A)$ which may be linked.

Example 1 (signature) *We define the signature for our case study:*

$$\Sigma \triangleq (\Sigma_{ag}, \Sigma_{site}, \Sigma_{ag\text{-}st})$$

where:

(i) $\Sigma_{ag} \triangleq \{A, B, C\}$;

(ii) $\Sigma_{site} \triangleq \{1, 2, 3, 4\}$;

(iii) $\Sigma_{ag\text{-}st} \triangleq [A \mapsto \{1, 2, 3\}, B \mapsto \{1, 2, 3, 4\}, C \mapsto \{1, 2\}]$

The agent types A, B, and C denote the three kinds of proteins. Each instance of the protein A has three sites the identifiers of which range from 1 *to* 3*; each instance of the protein B has four sites the identifiers of which range from* 1 *to* 4*; and each instance of the protein C has two sites the identifiers of which range from* 1 *to* 2*.*

3.2 Σ-graphs and morphisms among Σ-graphs

Σ-graphs are graphs the nodes of which are typed agents with some sites which may bear sets of binding states. In general, Σ-graphs encode some specific type disciplines [8]: they summarise the potential bonds and provide contextual conditions over them [3]. Patterns and bio-molecular species are specific kinds of Σ-graphs.

Definition 3.2 (Σ-graphs) *A Σ-graph is a tuple* $G \triangleq (\mathcal{A}_G, type_G, \mathcal{S}_G, \mathcal{L}_G)$ *where:*

(i) $\mathcal{A}_G \subseteq \mathbb{N}$ *is a finite set of agents,*

(ii) $type_G : \mathcal{A}_G \to \Sigma_{ag}$ *is a function mapping each agent to its type,*

(iii) $\mathcal{S}_G$ *is a subset of the set* $\{(n, i) \mid n \in \mathcal{A}_G, i \in \Sigma_{ag\text{-}st}(type_G(n))\}$,

(iv) $\mathcal{L}_G$ *is a function between the set* $\mathcal{S}_G$ *and the set* $\wp(\mathcal{S}_G \cup \{\dashv, -\})$ *such that*

for any two sites $(n,i),(n',i') \in \mathcal{S}_G$*, we have* $(n',i') \in \mathcal{L}_G(n,i)$ *if and only if* $(n,i) \in \mathcal{L}_G(n',i')$*.*

The set $\mathcal{S}_G$ denotes the set of binding sites. Whenever $\dashv \in \mathcal{L}_G(n,i)$, the site (n,i) may be free. Various levels of information may be given about the sites that are bound. Whenever $- \in \mathcal{L}_G(n,i)$, the site (n,i) may be bound to an unspecified site. Whenever $(n',i') \in \mathcal{L}_G(n,i)$ (and hence $(n,i) \in \mathcal{L}_G(n',i')$), the sites (n,i) and (n',i') may be bound together.

For a Σ-graph G, we write as $\mathcal{A}_G$ its set of agents, $type_G$ its typing function, $\mathcal{S}_G$ its set of sites, and $\mathcal{L}_G$ its set of links.

Example 2 (Σ-graph) *We give two examples of* Σ*-graph. We consider the* Σ*-graph* G_{CM} *that is defined as follows:*

(i) $\mathcal{A}_{G_{CM}} \triangleq \{1,2,3\}$*;*

(ii) $type_{G_{CM}} \triangleq [1 \mapsto A, 2 \mapsto B, 3 \mapsto C\,]$;

(iii) $\mathcal{S}_{G_{CM}} \triangleq \{(1,1),(1,2),(1,3),(2,1),(2,2),(2,3),(2,4),(3,1),(3,2)\}$*;*

(iv) $$\mathcal{L}_{G_{CM}} \triangleq \left[\begin{array}{l} (1,1) \mapsto \{\dashv,(2,1)\},(1,2) \mapsto \{\dashv,(2,2),(2,3)\}, \\ (1,3) \mapsto \{\dashv,(2,3),(3,2)\}, \\ (2,1) \mapsto \{\dashv,(1,1)\},(2,2) \mapsto \{\dashv,(1,2)\}, \\ (2,3) \mapsto \{\dashv,(1,2),(1,3)\},(2,4) \mapsto \{\dashv,(3,1)\}, \\ (3,1) \mapsto \{\dashv,(2,4)\},(3,2) \mapsto \{\dashv,(1,3)\} \end{array}\right].$$

and the Σ*-graph* G_Σ *that is defined as follows:*

(i) $\mathcal{A}_{G_\Sigma} \triangleq \{1,2,3,4\}$*;*

(ii) $type_{G_\Sigma} \triangleq [1 \mapsto A, 2 \mapsto A, 3 \mapsto B, 4 \mapsto C\,]$;

(iii) $$\mathcal{S}_{G_\Sigma} \triangleq \left\{\begin{array}{l} (1,1),(1,2),(1,3),(2,1),(2,2),(2,3), \\ (3,1),(3,2),(3,3),(3,4),(4,1),(4,2) \end{array}\right\};$$

(iv) $$\mathcal{L}_{G_\Sigma} \triangleq \left[\begin{array}{l} (1,1) \mapsto \{(3,1)\},(1,2) \mapsto \{\dashv,(3,2),(3,3)\},(1,3) \mapsto \{(3,3),(4,2)\}, \\ (2,1) \mapsto \{\dashv\},(2,2) \mapsto \{\dashv,(3,2),(3,3)\},(2,3) \mapsto \{\dashv,(3,3),(4,2)\}, \\ (3,1) \mapsto \{\dashv,(1,1)\},(3,2) \mapsto \{\dashv,(1,2),(2,2)\}, \\ (3,3) \mapsto \{\dashv,(1,2),(1,3),(2,2),(2,3)\},(3,4) \mapsto \{\dashv,(4,1)\}, \\ (4,1) \mapsto \{\dashv,(3,4)\},(4,2) \mapsto \{\dashv,(1,3),(2,3)\} \end{array}\right].$$

The Σ*-graphs* G_{CM} *and* G_Σ *are graphically described respectively in Figs. 4(a) and 4(b). We notice that agent identifiers are omitted (an agent is identified by its position). Site identifiers are omitted. Sites are depicted in increasing order of their identifiers from bottom up.*

The Σ*-graph* G_{CM} *plays a specific role: we call it the contact map of the model. In a contact map each agent type occurs exactly once and each agent documents its full set of sites. It can be interpreted as a context-insensitive description of the*

potential bindings between sites of agents.

Σ-graphs may be related by structure-preserving maps of agents, called morphisms. The definition of a morphism between two Σ-graphs is given as follows:

Definition 3.3 (morphisms) *A* morphism $h : G \to H$ *from the* Σ*-graph* G *into the* Σ*-graph* H *is a function of agents* $h : \mathcal{A}_G \to \mathcal{A}_H$ *satisfying, for all agent identifiers* n, $n' \in \mathcal{A}_G$, *for all site identifiers* $i \in \Sigma_{ag\text{-}st}(type_G(n))$, $i' \in \Sigma_{ag\text{-}st}(type_G(n'))$*:*

(i) $type_G(n) = type_H(h(n))$*;*

(ii) *if* $(n, i) \in \mathcal{S}_G$*, then* $(h(n), i) \in \mathcal{S}_H$*;*

(iii) *if* $(n', i') \in \mathcal{L}_G(n, i)$*, then* $(h(n'), i') \in \mathcal{L}_H(h(n), i)$*;*

(iv) *if* $\dashv \in \mathcal{L}_G(n, i)$*, then* $\dashv \in \mathcal{L}_H(h(n), i)$*;*

(v) *if* $- \in \mathcal{L}_G(n, i)$*, then* $\mathcal{L}_H(h(n), i) \cap \{-\} \cup \mathcal{S}_H \neq \emptyset$*.*

Morphisms preserve the type of agents. They also preserve each agent set of sites, but more sites may be documented in the image of the morphism. A site that may be free shall be mapped to a site that may be free. Two sites that may be bound together shall be mapped to two sites that may be bound together. Lastly, whenever a site may be bound to an unspecified site, it shall be mapped to a site that is bound to either an unspecified or a specified (or both) one.

Example 3 (morphisms) *The following function:* $[1 \mapsto 1, 2 \mapsto 1, 3 \mapsto 2, 4 \mapsto 3]$ *induces a morphism from the* Σ*-graph* G_Σ *into the* Σ*-graph* G_{CM}*. This morphism is graphically described in Fig. 4(c). We notice that both agents of type A have been merged into a single agent in the contact map, while merging the potential states of their sites. This way, the contact map provides a coarser (context-insensitive) summary of potential bonds in a model.*

Two morphisms from a Σ-graph E to a Σ-graph F, and from the Σ-graph F to a Σ-graph G respectively, compose in the usual way (and form a morphism from the Σ-graph E into the Σ-graph G).

3.3 Patterns and embeddings

Now we restrict the definition of Σ-graphs so as to focus on the ones that may express parts of the state of the system. These Σ-graphs, that we call patterns, are defined as follows:

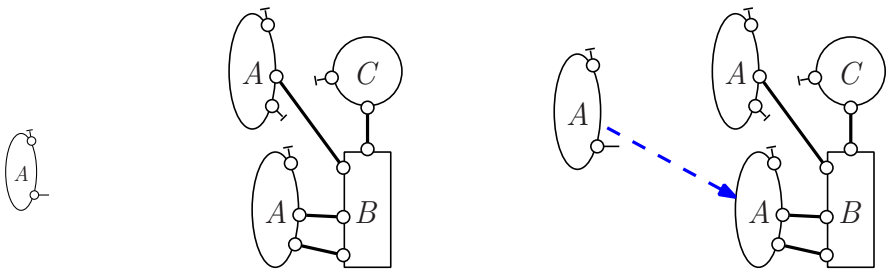
Definition 3.4 (patterns) *A pattern is a* Σ*-graph* P *such that, for every site* $s \in \mathcal{S}_P$ *both following conditions are satisfied:*

(i) *the set* $\mathcal{L}_P(s)$ *contains at most one element;*

(ii) *the set* $\mathcal{L}_P(s)$ *does not contain the element* s*.*

The first condition ensures that the state of every site is either unspecified, or free, or bound to an unspecified site, or bound to a single specific site. The second condition ensures that a site is never bound to itself.

Example 4 (patterns) *We give two examples of patterns. We consider the pattern* P *that is defined as follows:*



(a) The pattern P. (b) The species S. (c) An embedding from P into S.

Fig. 5. Two patterns P and S, and an embedding from the pattern P to the species S. The pattern S is a species: it forms a connected component and the state of each site in each agent is fully documented.

(i) $\mathcal{A}_P \triangleq \{1\}$;

(ii) $type_P \triangleq [1 \mapsto A]$;

(iii) $\mathcal{S}_P \triangleq \{(1,1),(1,3)\}$;

(iv) $\mathcal{L}_P \triangleq [(1,1) \mapsto \{-\}, (1,3) \mapsto \{\dashv\}]$;

and the pattern S that is defined as follows:

(i) $\mathcal{A}_S \triangleq \{1,2,3,4\}$;

(ii) $type_S \triangleq [1 \mapsto A, 2 \mapsto A, 3 \mapsto B, 4 \mapsto C]$;

(iii) $$\mathcal{S}_S \triangleq \left\{ \begin{array}{l} (1,1),(1,2),(1,3),(2,1),(2,2),(2,3), \\ (3,1),(3,2),(3,3),(3,4),(4,1),(4,2) \end{array} \right\};$$

(iv) $$\mathcal{L}_S \triangleq \left[\begin{array}{l} (1,1) \mapsto \{(3,1)\}, (1,2) \mapsto \{(3,2)\}, (1,3) \mapsto \{\dashv\}, \\ (2,1) \mapsto \{\dashv\}, (2,2) \mapsto \{(3,3)\}, (2,3) \mapsto \{\dashv\}, \\ (3,1) \mapsto \{(1,1)\}, (3,2) \mapsto \{(1,2)\}, (3,3) \mapsto \{(2,2)\}, (3,4) \mapsto \{(4,1)\}, \\ (4,1) \mapsto \{(3,4)\}, (4,2) \mapsto \{\dashv\} \end{array} \right].$$

The patterns P and S are graphically described respectively in Figs. 5(a) and 5(b).

A species is a connected pattern in which the state of each site is documented (no further information may be added). Depending on the choice of the semantics, the state of the system may be described either as a function from species to concentrations (differential setting), or as a multi-set of species (stochastic setting).

Patterns may be related by embeddings. Besides preserving the structure of patterns, embeddings map agents to agents injectively.

Definition 3.5 (embeddings) *An embedding is a morphism from a pattern into another one, that is induced by an injective agent function.*

We denote as $[P, P']$ the set of the embeddings from a pattern P to a pattern P'.

Example 5 (embeddings) *The function $[1 \mapsto 1]$ induces an embedding from the pattern P to the species S, as depicted in Fig. 5(c).*

As opposed to classical notions of embeddings between graphs, embeddings between patterns preserve the freeness of sites.

(a) The extended pattern EP. (b) The extended pattern EP'.

Fig. 6. Two extended patterns EP and EP'.

The composition of two embeddings is an embedding.

4 Extended patterns

Patterns may be used to specify constraints over bio-molecular species. They are very convenient to express conjunctions of positive assumptions, such the fact that this site in this agent is free and that those other two sites are bound together. But, they are not so convenient to express disjunctions and/or negative information about bio-molecular species such as the fact that these two sites are not bound together or such as the fact that either this site is free, or this other site is bound. In this section, we extend the definition of patterns so as to deal with specific kinds of disjunctive and negative information. Then, we express the number of instances of an extended pattern as an expression of the number of instances of some classical patterns.

4.1 Extended patterns and their semantics

Let us give the formal definition of an extended pattern.

Definition 4.1 (extended patterns) *An extended pattern is defined as a triple* $EP \triangleq (P, L, M)$ *where:*

(i) P *is a pattern;*

(ii) L *is a finite set of bond identifiers;*

(iii) M *is a map from the set* L *to the set* $\{(n, i) \mid n \in \mathcal{A}_P, i \in \Sigma_{ag\text{-}st}(type_P(n))\}^2$.

An extended pattern (P, L, M) stands for a pattern in which some potential bonds may be inserted. No requirement is done on the sites involved in these potential bonds, except that they must belong to agents of the pattern P. Some of them may be already specified in P (i.e. they belong to the set $\mathcal{S}_P$), and some others may be missing in P (i.e. they do not belong to the set $\mathcal{S}_P$).

Example 6 (extended patterns) *We give two examples of extended patterns. We consider the extended pattern* $EP \triangleq (P, L, M)$ *that is defined as follows:*

(i) $\mathcal{A}_P = \{1, 2, 3\}$;

(ii) $type_P = [\, 1 \mapsto A, 2 \mapsto B, 3 \mapsto C\,]$;

(iii) $\mathcal{S}_P = \{(2,1), (2,2), (2,3), (2,4), (3,1), (3,2)\}$;

(iv) $\mathcal{L}_P = \left[\begin{array}{l} (2,1) \mapsto \{-\}, (2,2) \mapsto \{-\}, (2,3) \mapsto \{-\}, (2,4) \mapsto \{(3,1)\}, \\ (3,1) \mapsto \{(2,4)\}, (3,2) \mapsto \{\dashv\} \end{array} \right]$;

(v) $L = \{1,2,3,4\}$;

(vi) $M = \left[\begin{array}{l} 1 \mapsto ((1,1),(2,1)), 2 \mapsto ((1,2),(2,2)), \\ 3 \mapsto ((1,2),(2,3)), 4 \mapsto ((1,3),(2,3)) \end{array} \right]$

and the extended pattern $EP' \triangleq (P', L', M')$ *that is defined as follows*

(i) $\mathcal{A}_{P'} = \{1,2,3\}$;

(ii) $type_{P'} = [\, 1 \mapsto A, 2 \mapsto B, 3 \mapsto C\,]$;

(iii) $\mathcal{S}_{P'} = \{(1,3),(2,1),(2,2),(2,3),(2,4),(3,1),(3,2)\}$;

(iv) $\mathcal{L}_{P'} = \left[\begin{array}{l} (1,3) \mapsto \{(3,2)\}, \\ (2,1) \mapsto \{-\}, (2,2) \mapsto \{-\}, (2,3) \mapsto \{\dashv\}, (2,4) \mapsto \{(3,1)\}, \\ (3,1) \mapsto \{(2,4)\}, (3,2) \mapsto \{(1,3)\} \end{array} \right]$;

(v) $L' = \{1,2\}$;

(vi) $M' = [1 \mapsto ((1,1),(2,1)), 2 \mapsto ((1,2),(2,2))]$.

The extended patterns EP *and* EP' *are graphically described respectively in Figs. 6(a) and 6(b).*

An extended pattern is a symbolic representation of a set of patterns, that are obtained by inserting some of the bonds in the image of the function M, without creating conflicts on any site. Each such pattern is called a valid annotation of the extended pattern, as defined as follows:

Definition 4.2 (valid annotations) *Let* $EP \triangleq (P, L, M)$ *be an extended pattern and* X *be a subset of* L*. We denote as* EP_X *the* Σ*-graph that is defined as follows:*

(i) $\mathcal{A}_{EP_X} \triangleq \mathcal{A}_P$;

(ii) $type_{EP_X} \triangleq type_P$;

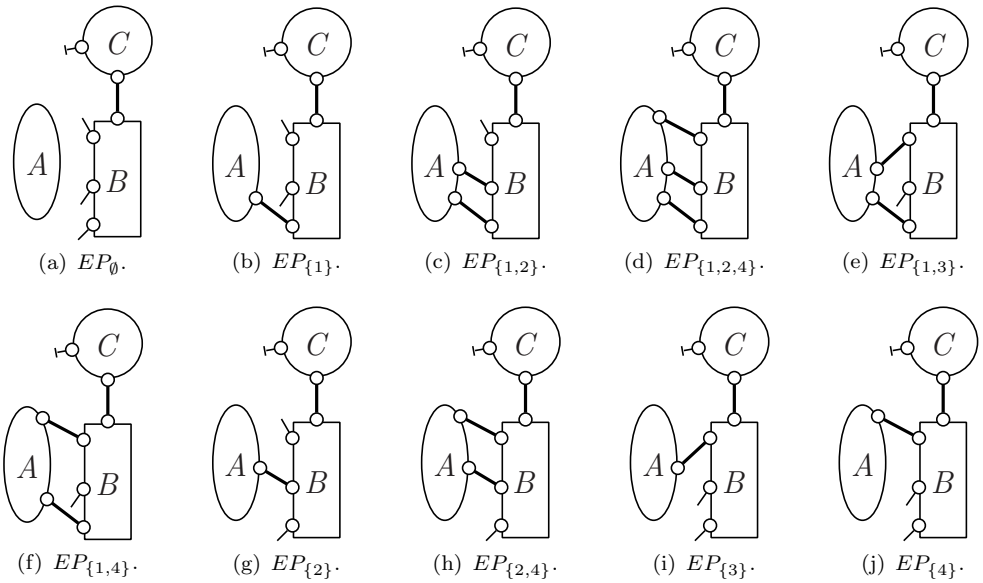
(iii) $\mathcal{S}_{EP_X} \triangleq \mathcal{S}_P \cup \bigcup (\{s, s'\} \mid \exists l \in X, M(l) = (s, s')\})$;

(iv) $\mathcal{L}_{EP_X} \triangleq \begin{cases} \mathcal{S}_{EP_X} \to \wp(\mathcal{S}_{EP_X} \cup \{\dashv, -\}) & \\ s \mapsto (\mathcal{L}_P(s) \cup \{s'\}) \setminus \{-\} & \text{if } \exists l \in X \text{ such that } M(l) = (s, s'), \\ s \mapsto (\mathcal{L}_P(s) \cup \{s'\}) \setminus \{-\} & \text{if } \exists l \in X \text{ such that } M(l) = (s', s), \\ s \mapsto \mathcal{L}_P(s) & \text{otherwise.} \end{cases}$

The set X *is a valid annotation of the extended pattern* EP *if and only if the* Σ*-graph* EP_X *is a pattern.*

We denote as $[\![EP]\!]$ *the set of the valid annotations of the extended pattern* EP*.*

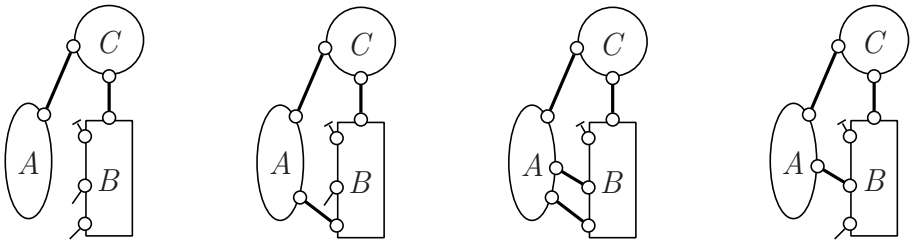
A valid annotation EP_X of an extended pattern $EP \triangleq (P, L, M)$ is obtained by inserting bonds $\{M(i) \mid i \in X\}$ in the initial pattern P. These bonds shall connect sites that are either missing, or that occur with an unspecified state, or that are bound to an unspecified site. Also it is not possible to insert several bonds on the



(a) $EP_{\emptyset}$. (b) $EP_{\{1\}}$. (c) $EP_{\{1,2\}}$. (d) $EP_{\{1,2,4\}}$. (e) $EP_{\{1,3\}}$.

(f) $EP_{\{1,4\}}$. (g) $EP_{\{2\}}$. (h) $EP_{\{2,4\}}$. (i) $EP_{\{3\}}$. (j) $EP_{\{4\}}$.

Fig. 7. Valid annotations of the extended pattern EP (e.g. see Fig. 6(a)).



(a) $EP'_{\emptyset}$. (b) $EP'_{\{1\}}$. (c) $EP'_{\{1,2\}}$. (d) $EP'_{\{2\}}$.

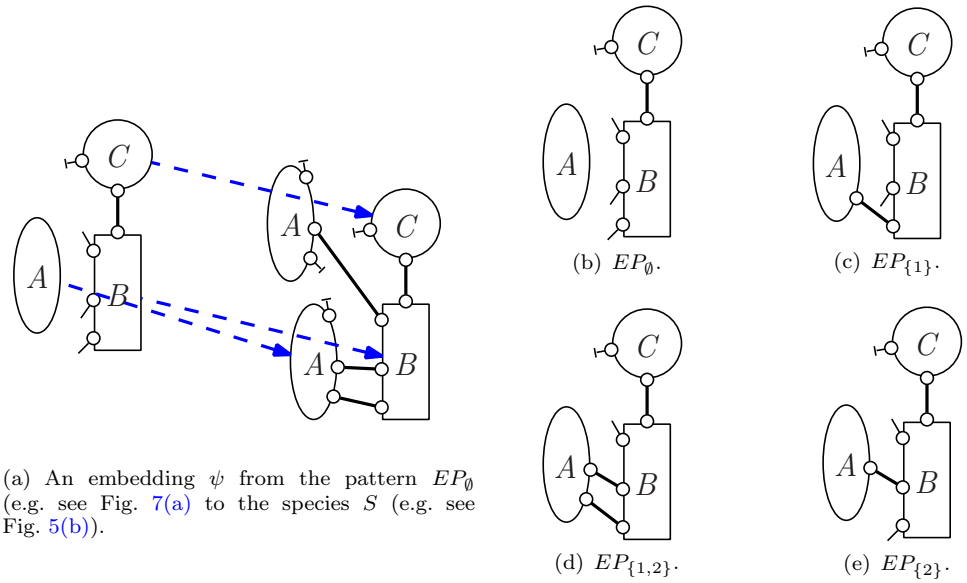
Fig. 8. Valid annotations of the extended pattern EP' (e.g. see Fig. 6(b)).

same site. Otherwise the insertion of the new bonds would induce a conflict: the annotation of the extended pattern would not produce a pattern. When a bond is inserted to a site that is bound to an unspecified site, its binding state is refined: we replace the binding state '−' with a pointer to the other site of the bond.

It is worth noticing that the set of the valid annotations of an extended pattern may not be stable upon set union. Moreover, the empty set $\emptyset$ is always a valid annotation and $EP_{\emptyset}$ is the pattern P.

Example 7 (valid annotations) *The extended pattern EP (e.g. see Fig. 6(a)) has exactely* 10 *valid annotations:* $\emptyset$, $\{1\}$, $\{1,2\}$, $\{1,2,4\}$, $\{1,3\}$, $\{1,4\}$, $\{2\}$, $\{2,4\}$, $\{3\}$, *and* $\{4\}$. *They are depicted in Fig. 7.*

We notice both sets $\{2\}$ *and* $\{3\}$ *are valid annotations of the extended pattern EP, whereas the set* $\{2,3\}$ *is not a valid annotation of it. This confirms that the set of the valid annotations of a given extended pattern may not be a valid annotation. The extended pattern EP' (e.g. see Fig. 6(b)) has exactely* 4 *valid annotations:* $\emptyset$, $\{1\}$, $\{1,2\}$, *and* $\{2\}$. *They are depicted in Fig. 8.*



(a) An embedding ψ from the pattern $EP_\emptyset$ (e.g. see Fig. 7(a)) to the species S (e.g. see Fig. 5(b)).

(b) $EP_\emptyset$.

(c) $EP_{\{1\}}$.

(d) $EP_{\{1,2\}}$.

(e) $EP_{\{2\}}$.

Fig. 9. Realisations of the extended pattern EP (e.g. see Fig. 6(a)) along an embedding ψ.

4.2 Extended pattern realisations

Given an extended pattern (P, L, M), every embedding ϕ from the pattern P into another pattern P' is associated with the set of the valid annotations that are made of bonds that are realised in the image of the embedding ϕ. This notion is defined as follows:

Definition 4.3 (realisations) *Let $EP \triangleq (P, L, M)$ be an extended pattern, P' be a pattern, ϕ be an embedding from the pattern P to the pattern P', and X be a subset of L.*

The set X is a realisation of the extended pattern EP along the embedding ϕ, if and only if the following conditions are satisfied:

(i) *the set X is a valid annotation of the extended pattern EP (i.e. $X \in [\![EP]\!]$);*

(ii) *there exist an embedding ϕ_1 from the pattern P to the pattern EP_X and an embedding ϕ_2 from the pattern EP_X to the pattern P' such that $\phi = \phi_2 \circ \phi_1$.*

We denote as $[\![\phi]\!]_{EP}$ the set of the realisations of the extended pattern EP along the embedding ϕ.

Example 8 (realisations) *We consider the embedding ψ from the pattern $EP_\emptyset$ (e.g. see Fig. 7(a)) to the species S (e.g. see Fig. 5(b)), that is induced by the agent map $[1 \mapsto 1]$. The embedding ψ is depicted in Fig. 9(a). The set $[\![\psi]\!]_{EP}$ of the realisations of the extended pattern EP (e.g. see to the pattern Fig. 6(a)) along the embedding ψ is equal to the set $\wp(\{1, 2\})$ of the subsets of the set $\{1, 2\}$. We show graphically the patterns corresponding to each of these realisations in Fig. 9.*

The set of the realisations of a given extended pattern along a given embedding enjoys the nice algebraic properties stated in the following proposition.

Proposition 4.4 *Let $EP \triangleq (P, L, M)$ be an extended pattern, P' be a pattern, ϕ*

be en embedding from the pattern P to the pattern P'.

Both following properties are satisfied:

(i) *the set $[\![\phi]\!]_{EP}$ is closed under set union (i.e. $\forall X, X' \in [\![\phi]\!]_{EP},\ X \cup X' \in [\![\phi]\!]_{EP}$);*

(ii) *the set $[\![\phi]\!]_{EP}$ is closed under subset (i.e $\forall X \in [\![\phi]\!]_{EP}, \forall X' \in \wp(X), X' \in [\![\phi]\!]_{EP}$).*

In particular, the empty set is always a realisation of an extended pattern along an embedding. The point that the set of the realisations of a given extended pattern along a given embedding is stable upon set union may sound counter-intuitive because of the fact that valid annotations are not stable per set union. Yet, it comes from the fact that only a subset of the bonds that occur in the image of the embedding may be added, and since this image is a pattern, its bonds are conflict-free.

Since the set of the realisations of an extended pattern EP along an embedding ϕ is finite and stable upon set union, the element $\bigcup [\![\phi]\!]_{EP}$ belongs to this set and it is a superset of any other element of this set.

4.3 Signed instances of extended patterns

Now we define the positive and the negative instances of an extended pattern (P, L, M). Intuitively, a positive instance of an extended pattern is an embedding stemming from the pattern P that realises at least one of the potential bonds (in the image of M). As seen in Sec. 2, the number of positive instances of extended patterns is useful to express the consumption of patterns due to side-effects. The negative instances of an extended pattern is the set of the embeddings stemming from the pattern P that realises none of the potential bonds (in the image of M). As seen in Sec. 2, negative instances of extended patterns are useful to express the production of patterns due to side-effects.

We give as follows the formal definition of positive and negative instances of an extended pattern.

Definition 4.5 (positive instances) *Let $EP \triangleq (P, L, M)$ be an extended pattern, P' be a pattern, and ϕ be an embedding from the pattern P to the pattern P'. We assume that, for any element $l \in L$, the pattern $P_{\{l\}}$ is connected.*

We say that ϕ is a positive instance of the extended pattern EP in the pattern P' if, and only if, the set $\bigcup [\![\phi]\!]_{EP} \neq \emptyset$.

The set of the positive instances of the extended pattern EP in the pattern P' is denoted as $[EP, P']_{\exists}$.

Definition 4.6 (negative instances) *Let $EP \triangleq (P, L, M)$ be an extended pattern, P' be a pattern, and ϕ be an embedding from the pattern P and the pattern P'. We assume that the pattern P is connected.*

We say that ϕ is a negative instance of the extended pattern EP in the pattern P' if, and only if, the set $\bigcup [\![\phi]\!]_{EP} = \emptyset$.

The set of the negative instances of the extended pattern EP in the pattern P' is denoted as $[EP, P']_{\nexists}$.

The negative and the positive instances of an extended pattern (P, L, M) induce a partition of the set of the embeddings that stem from the pattern P. This is

formalised in the following proposition:

Proposition 4.7 *Let $EP \triangleq (P, L, M)$ be an extended pattern and P' be a pattern. We assume that the pattern P is connected.*

Under this assumption, the sets $[EP, P']_{\exists}$ and $[EP, P']_{\nexists}$ are well-defined. Moreover, we have:

$$\sharp[EP, P']_{\nexists} + \sharp[EP, P']_{\exists} = \sharp[P, P'].$$

Proof. The function mapping each embedding ϕ from the pattern P to the pattern P' to the number 0 if $\bigcup[\![\phi]\!]_{EP}\phi$ is equal to the emptyset, or to the number 1 otherwise, induces a partition over the set $[P, P']$. □

4.4 Alternated sums

Given k and n two natural numbers, we denote as $\binom{n}{k}$ the number of parts of k elements in a set of n elements. In particular, $\binom{n}{k} = 0$ whenever $k > n$. Moreover, $\binom{n}{0} = 1$ and $\binom{n}{n} = 1$.

Lemma 4.8 *Let n be a non-zero natural number. The following equality:*

$$1 = \textstyle\sum_{k=1}^{n} (-1)^{k+1} \binom{n}{k}$$

is satisfied.

Proof. This is a direct consequence of the binomial expansion formula [15].

For each two real numbers a and b and each natural number n, the following equality is satisfied:

$$(a+b)^n = \sum_{k=0}^{n} a^k b^{n-k}.$$

Let us apply this formula with $a \triangleq (-1)$ and $b \triangleq 1$.

We get:

$$\begin{aligned} 0 &= (1 + (-1))^n \\ 0 &= \textstyle\sum_{k=0}^{n} (-1)^k 1^{n-k} \\ 0 &= (-1)^0 + \textstyle\sum_{k=1}^{n} (-1)^k \\ \textstyle\sum_{k=1}^{n} (-1)^{k+1} &= 1. \end{aligned}$$

□

Lemma 4.9 *Let $EP \triangleq (P, L, M)$ be an extended pattern, P' be a pattern, X be a valid annotation of the extended pattern EP. The following equality holds:*

$$\sharp[P_X, P'] = \textstyle\sum_{X' \in \wp(L),\, X \subseteq X'} \sharp\{\phi \in [P, P'] \mid \bigcup[\![\phi]\!]_{EP} = X'\}.$$

Proof. We introduce the embedding i from the pattern P to the pattern P_X that is induced by the identity function over the agents of the pattern P. We consider the function Φ mapping each embedding ϕ from the pattern P_X to the pattern P', to the top element $\bigcup[\![\phi \circ i]\!]_{EP}$ among the realisations $[\![\phi \circ i]\!]_{EP}$ of the extended pattern

EP along the embedding $\phi \circ i$. The function Φ induces a partition of the set of the embeddings from the pattern P_X to the pattern P', which proves the lemma. □

Now we can express the number of positive and negative instances of extended patterns as an alternated sums of the number of instances of its valid annotations.

Theorem 4.10 *Let EP be an extended pattern and P' be a pattern. Then the following equalities hold:*

(i) $\sharp[EP, P']_{\exists} = \sum_{X \in [\![EP]\!] \setminus \{\emptyset\}} (-1)^{\sharp X + 1} \sharp[EP_X, P']$.

(ii) $\sharp[EP, P']_{\nexists} = \sum_{X \in [\![EP]\!]} (-1)^{\sharp X} \sharp[EP_X, P']$.

Proof.

(i) The first equation follows from Lem. 4.8 and Lem. 4.9: [5]

$$
\begin{aligned}
&\sharp[EP, P']_{\exists} \\
&= \textstyle\sum_{X' \in [\![EP]\!] \setminus \{\emptyset\}} \sharp\{\phi \in [P, P'] \mid \bigcup [\![\phi]\!]_{EP} = X'\} \\
&= \textstyle\sum_{X' \in [\![EP]\!] \setminus \{\emptyset\}} \left(\sum_{k=1}^{\sharp X'} (-1)^{k+1} \binom{\sharp X'}{k} \right) \sharp\{\phi \in [P, P'] \mid \bigcup [\![\phi]\!]_{EP} = X'\} \\
&= \textstyle\sum_{X' \in [\![EP]\!] \setminus \{\emptyset\}} \left(\sum_{k=1}^{\sharp X'} (-1)^{k+1} \sum_{X \in \wp_k(X')} 1 \right) \sharp\{\phi \in [P, P'] \mid \bigcup [\![\phi]\!]_{EP} = X'\} \\
&= \textstyle\sum_{X' \in [\![EP]\!] \setminus \{\emptyset\}} \sum_{X \in \wp(X') \setminus \{\emptyset\}} (-1)^{\sharp X + 1} \sharp\{\phi \in [P, P'] \mid \bigcup [\![\phi]\!]_{EP} = X'\} \\
&= \textstyle\sum_{X \in \wp(L) \setminus \{\emptyset\}} (-1)^{\sharp X + 1} \sum_{X' \in \wp(L) \setminus \{\emptyset\},\, X \subseteq X'} \sharp\{\phi \in [P, P'] \mid \bigcup [\![\phi]\!]_{EP} = X'\} \\
&= \textstyle\sum_{X \in \wp(L) \setminus \{\emptyset\}} (-1)^{\sharp X + 1} \sum_{X' \in \wp(L) \setminus \{\emptyset\},\, X \subseteq X'} \sharp\{\phi \in [P, P'] \mid \bigcup [\![\phi]\!]_{EP} = X'\} \\
&= \textstyle\sum_{X \in \wp(L) \setminus \{\emptyset\}} (-1)^{\sharp X + 1} \sharp[P_X, P'].
\end{aligned}
$$

(ii) The second equation follows from the previous equation and Prop. 4.7:

$$
\begin{aligned}
\sharp[EP, P']_{\nexists} &= \sharp[P_\emptyset, P'] - \sharp[EP, P']_{\exists} \\
\sharp[EP, P']_{\nexists} &= 1^0 \sharp[P_\emptyset, P'] - \textstyle\sum_{X \in \wp(L) \setminus \{\emptyset\}} (-1)^{\sharp X + 1} \sharp[P_X, P'] \\
\sharp[EP, P']_{\nexists} &= 1^{\sharp \emptyset} + \textstyle\sum_{X \in \wp(L) \setminus \{\emptyset\}} (-1)^{\sharp X} \sharp[P_X, P'] \\
\sharp[EP, P']_{\nexists} &= \textstyle\sum_{X \in \wp(L)} (-1)^{\sharp X} \sharp[P_X, P'].
\end{aligned}
$$

□

Example 4.11 We consider the extended patterns $EP = (P, L, M)$ and $EP' = (P', L', M')$ that are depicted in Figs. 6(a) and 6(b) respectively. Let P'' be a pattern.

We have:

$$
\begin{aligned}
\sharp[EP, P'']_{\exists} = {} & \sharp[P_{\{1\}}, P''] + \sharp[P_{\{2\}}, P''] + \sharp[P_{\{3\}}, P''] + \sharp[P_{\{4\}}, P''] + \sharp[P_{\{1,2,4\}}, P''] \\
& - (\sharp[P_{\{1,2\}}, P''] + \sharp[P_{\{1,3\}}, P''] + \sharp[P_{\{1,4\}}, P''] + \sharp[P_{\{2,4\}}, P'']).
\end{aligned}
$$

and:

$$
\sharp[EP, P'']_{\nexists} = P'_\emptyset - \sharp[P_{\{1\}}, P''] - \sharp[P_{\{2\}}, P''] + \sharp[P_{\{1,2\}}, P''].
$$

[5] For every set X, we denote as $\wp_k(X)$ the set of the subsets of X with exactly k elements.

5 Conclusion

Counting occurrences of patterns is crucial in site-graph rewriting languages. We have proposed a new relationship among the number of occurrences of patterns. Given a pattern and a set of potential bonds to be inserted in this pattern, we have shown that the number of occurrences of this pattern, in which at least one (resp. none) of the potential bonds is realised, is always equal to an alternated sum of the number of occurrences of the patterns that are obtained by inserting subsets of potential bonds in the initial pattern.

This result has several applications. It simplifies the counting of potential rule applications during the simulation of a model by network free methods [7,16,2]. It also relaxes the requirements about cycles in the definition of the flow of information among sites, leading to more compact model reductions [6,3].

References

[1] Blinov, M., J. R. Faeder, B. Goldstein and W. S. Hlavacek, *Bionetgen: software for rule-based modeling of signal transduction based on the interactions of molecular domains.*, Bioinformatics (Oxford, England) **20** (2004).

[2] Boutillier, P., T. Ehrhard and J. Krivine, *Incremental update for graph rewriting*, in: H. Yang, editor, *Proc. ESOP'17*, LNCS **10201** (2017).

[3] Camporesi, F., J. Feret and J. Hayman, *Context-sensitive flow analyses: a hierarchy of model reductions.*, in: A. Gupta and T. Henzinger, editors, *Proc. CMSB'13*, number 8130 in LNCS (2013).

[4] Danos, V., J. Feret, W. Fontana, R. Harmer, J. Hayman, J. Krivine, C. D. Thompson-Walsh and G. Winskel, *Graphs, rewriting and pathway reconstruction for rule-based models*, in: D. D'Souza, T. Kavitha and J. Radhakrishnan, editors, *Proc. FSTTCS 2012*, LIPIcs **18** (2012).

[5] Danos, V., J. Feret, W. Fontana, R. Harmer and J. Krivine, *Rule-based modelling of cellular signalling, invited paper*, in: L. Caires and V. Vasconcelos, editors, *Proc. CONCUR'07*, LNCS **4703** (2007), pp. 17–41.

[6] Danos, V., J. Feret, W. Fontana, R. Harmer and J. Krivine, *Abstracting the differential semantics of rule-based models: exact and automated model reduction*, in: J. Jouannaud, editor, *Proc. LICS'10*, IEEE (2010).

[7] Danos, V., J. Feret, W. Fontana and J. Krivine, *Scalable simulation of cellular signaling networks, invited paper*, in: Z. Shao, editor, *Proc. APLAS'07*, LNCS **4807** (2007).

[8] Danos, V., R. Harmer and G. Winskel, *Constraining rule-based dynamics with types*, MSCS **23** (2013).

[9] Danos, V. and C. Laneve, *Formal molecular biology*, TCS **325** (2004).

[10] Feret, J., *Abstract interpretations of rule-based models*, habilitation, in preparation.

[11] Feret, J., V. Danos, J. Krivine, R. Harmer and W. Fontana, *Internal coarse-graining of molecular systems*, Proc. PNAS (2009).

[12] Feret, J., H. Koeppl and T. Petrov, *Stochastic fragments: A framework for the exact reduction of the stochastic semantics of rule-based models*, IJSI **7** (2013).

[13] Feret, J. and K. Q. Lý, *Reachability analysis via orthogonal sets of patterns.*, in: *Proc. SASB'16*, ENTCS, to appear.

[14] Harmer, R., V. Danos, J. Feret, J. Krivine and W. Fontana, *Intrinsic information carriers in combinatorial dynamical systems*, Chaos **20** (2010).

[15] O'Connor, J. J. and E. F. Robertson, "Abu Bekr ibn Muhammad ibn al-Husayn Al-Karaji," MacTutor History of Mathematics archive, University of St Andrews .

[16] Sneddon, M. W., J. R. Faeder and T. Emonet, *Efficient modeling, simulation and coarse-graining of biological complexity with nfsim*, Nat. meth. **8** (2011).