MAXIMAL REDUCTION OF DETERMINISTIC SEMANTICS OF RULE-BASED MODELS: SYNTAX-INDEPENDENT SETUP

T. Petrov and H. Koeppl

School of Communication and Computer Sciences
Ecole Polytechnique Federale de Lausanne (EPFL)
1015 Lausanne, Switzerland
tatjana.petrov@epfl.ch, heinz.koeppl@epfl.ch

ABSTRACT

We develop a syntax-independent rule-based specification in a set-theoretical framework. The reactants of rules are usually defined as partially defined objects, called *patterns*, to denote that we can apply a reaction by substituting a pattern with any species that contains the context described by the pattern. We choose a very general data structure to represent the patterns: we write a pattern as a set of species which comply the pattern, and we call it a *macro-species*. We define the deterministic (ODE) semantics on such a rule set, and we define its reduction imposed by *aggregation* of species. We characterize for which aggregations we can correctly compute the reduced ODE semantics directly from the rule set, ie which aggregations are *self-consistent*. Finally, we propose an algorithm that, given a set of initial aggregates, computes the *least-refined* self-consistent aggregation which contains the initial aggregates.

1. INTRODUCTION

Rule-based models give a compact description of protein-protein interactions in spite of combinatorial number of protein complexes interacting in the cellular system [1, 2]. Still, the computation of ODE semantics of a rule set is hard or often infeasible. People have proposed reduction algorithms for ODE semantics of a rule set expressed in a rule-based language Kappa [3]. Reduction is done by introducing a formal framework which allows to compute the ODE semantics directly on typically far fewer *aggregates* of species rather then on concrete species. However, a particular syntax gives limitations to which sets of species can be aggregated, and this leads to limitations in both specification and the aggregation process. For those reasons, we develop a framework in which any macro-species or aggregate is represented as a set of species. We characterize the set of least-refined (minimal) self-consistent aggregations. The implication of our work is the following. It gives theoretical grounds to developing the self-consistent reduction of the ODE semantics in any concrete implementation of rule-based models. Moreover, since we characterize the least-refined aggregations, it can indicate improvements in current implementations.

2. PRELIMINARIES

We denote the set of species of the biochemical system by $S = \{s_1, ..., s_n\}$. For each observed species, more copies of the species may be occurring in the system ('reaction soup'). Any multi-set on species $\mathbf{c} : T \to {\mathbb{N}_0}^n$, where $T = \mathbb{R}_{\geq 0}$ is the continuous-time domain, we call a *configuration*. Instead of working with copy numbers, we rather scale it to the concentration units, and we denote it by $\mathbf{x} : T \to \mathbb{R}^n_{\geq 0}$. The i-th component of this vector we denote by $x_i(t)$ or $[s_i] := x_i(t)$. The dynamics is specified by a set of reactions. One reaction consists of the two configurations - left- and right-hand side (lhs and rhs in further text), and it means that if the left-hand side multi-set is a subset of the configuration of the whole system, it can be replaced by the multiset on the right-hand side.

Definition 1 *(Reaction-based biochemical system)* Given a finite set of species $S = \{s_1, ..., s_n\}$, a reaction-based biochemical system $\mathcal{S} = (S, \mathbf{x}_0, \{r_1, ..., r_N\})$ is given by an initial configuration $\mathbf{x}_0 \in \mathbb{R}^n_{\geq 0}$, and a set of reactions $\{r_1, ..., r_N\}$, such that

$$r_i : a_{1i}s_1 + ... + a_{ni}s_n \xrightarrow{k_i} b_{1i}s_1 + ... + b_{1n}s_n,$$

where $a_{1i}, ..., a_{ni}, b_{1i}, ..., b_{ni} \in \mathbb{N}_0$ are called stoichiometric coefficients. We will say that the reaction is of type (l, d), if there are l non-zero coefficients on the lhs of the rule, and if there are d non-zero coefficients on the rhs of the rule. Note that each reaction r_i is defined by a triple $(\mathbf{L}, \boldsymbol{\alpha}, k)_i$, where (i) $\mathbf{L} \in \mathbb{N}^n_0$ is the lhs of the reaction r_i, (ii) $\boldsymbol{\alpha} : \mathbf{L} \to \mathbb{N}^n_0$ is a bijection from lhs to the rhs, and (iii) $k \in \mathbb{R}_{\geq 0}$ is the reaction rate.

Definition 2 *(ODE semantics of a reaction-based biochemical system)* Given a reaction-based biochemical system $\mathcal{S} = (S, \mathbf{x}_0, \{r_1, ..., r_N\})$, we assume the mass-action law. Then, the derivatives of the mean change of concentrations of species over time is specified by the following system of differential maps

$$\mathbf{F}(\mathbf{x}(t)) = \frac{\mathrm{d}}{\mathrm{d}t}\mathbf{x}(t) = \sum_i^N \left(\hat{\nu}(\mathbf{x}(t), r_i) - \check{\nu}(\mathbf{x}(t), r_i)\right),$$

where the consumption and production terms of species s_j with respect to the reaction r_i are computed as $\check{\nu}(x_j(t), r_i) = k_i a_{ji} \prod_{l=1}^{n} \binom{x_l(t)}{a_{li}}$, and $\hat{\nu}(x_j(t), r_i) = k_i b_{ji} \prod_{l=1}^{n} \binom{x_l(t)}{a_{li}}$. The solution to this system of ODE equations, such that it satisfies the initial condition $\mathbf{x}_0$ we call the *ODE semantics* of $\mathcal{S}$.

3. RULE-BASED MODELING WITH SETS

Using a rule-based language enables grouping certain reactions which have the same rate. The reactants in rules are patterns, partially defined species. We can execute a rule by substituting a pattern with any species that contains the context described by the pattern. We here define a pattern as a set of species, *macro-species*, to specify that any species which belongs to that set can equally probably be chosen to take part in the rule application.

Definition 3 *(Rule-based biochemical system)* Given a finite set of species $S = \{s_1, ..., s_n\}$, and a finite set of macro-species $\mathcal{C} = \{C_1, ..., C_{\tilde{n}}\}$, $C_i \subseteq S$, a *rule-based* biochemical system $\tilde{\mathcal{S}} = (S, \mathcal{C}, \mathbf{x}_0, \{R_1, ..., R_M\})$ is given by an initial configuration $\mathbf{x}_0 \in \mathbb{N}_0^n$, and a set of rules $\{R_1, ..., R_M\}$. The rules are specified in the following form

$$R : a_1 L_1 + ... + a_l L_l \xrightarrow{k_i} b_1 D_1 + ... + b_d D_d,$$

where $a_1, ..., a_l, b_1, ..., b_d \in \mathbb{N}$ are positive integer coefficients, and $L_1, ..., L_l \in \mathcal{C}$ and $D_1, ..., D_d \in \mathcal{C}$. This rule is a rule of type (l, d). Note that each rule R_i can be seen as a triple $(\mathbf{L}, \boldsymbol{\beta}, k)_i$, where (i) $\mathbf{L}$ is a Cartesian set product $L_1 \times ... \times L_l$, (ii) $\boldsymbol{\beta}$ is an isomorphism between the Cartesian set product of the macro-species on lhs, and the macro-species on the rhs: $\boldsymbol{\beta} : L_1 \times ... \times L_l \to D_1 \times ... \times D_d$, and (iii) $k \in \mathbb{R}_{\geq 0}$ is a rule rate.

Any tuple of species $(c_1, ..., c_l)$ such that $(c_1, ..., c_l) \in L_1 \times ... \times L_l$ specifies one reaction. Thus, writing one rule $R = (\mathbf{L}, \boldsymbol{\beta}, k)$ is equivalent to writing $|L_1| \cdot ... \cdot |L_l|$ distinct reactions of type (l, d), all with the same reaction rate k.

Definition 4 *(Well-formed rules)* We will say that the rule is *well-formed*, if the sets $L_1 \times L_2 ... \times L_l$ and $D_1 \times ... \times D_d$ are isomorphic. We will restrict our attention to the rules where all coefficients on the left-hand side are equal to one: $a_1 = ... = a_l = 1$. Moreover, we assume that each two macro-species that appear on the lhs of the rule are mutually disjoint. We assume the same for the rhs. The map $\boldsymbol{\beta}$ being an isomorphism allows us to define its inverse $\boldsymbol{\beta}^{-1} : D_1 \times ... \times D_d \to L_1 \times ... \times L_l$. From now on, we will assume to be working only with well-formed rule-based biochemical systems.

Definition 5 *(ODE semantics of a rule-based biochemical system)* Given a rule-based biochemical system $\tilde{\mathcal{S}}$, we define its ODE semantics as the ODE semantics of its equivalent reaction-based biochemical system $\mathcal{S}$.

4. REDUCTIONS OF ODE SEMANTICS OF RULE-BASED MODELS

If we compute the ODE semantics of a rule-based biochemical system by translating it to reactions, we however again end up solving the system of differential equations, whose dimension is equal to the number of species. Since the rules are specified on macro-species, we search for a set of macro-species, that we call *aggregates*, such that instead of computing the ODE semantics over the tuple of species, we can compute it over the tuple of aggregates.

Definition 6 *(Aggregation)* Let $\mathcal{A} = \{A_1, ..., A_m\}$, where $A_i \subseteq S$ are macro-species on S. We call $\mathcal{A}$ an *aggregation* on S, and A_1,...,A_m we call *aggregates*. The set of all aggregations of S we denote by $\mathscr{A}$.

Given an aggregation $\mathcal{A} = \{A_1, ..., A_m\}$, we define the $m \times n$ matrix $\mathbf{\Pi} \in \{0, 1\}^{m \times n}$, such that

$$\Pi_{ij} = \begin{cases} 1 & \text{if } s_j \in A_i, \\ 0 & \text{otherwise.} \end{cases}$$

If we denote the multiset configuration over the aggregates by $\mathbf{c}_A$, and the same multiset configuration over the species by $\mathbf{c}_S$, then it holds that: $\mathbf{c}_A = \mathbf{\Pi} \cdot \mathbf{c}_S$ (vectors are written in column notation). The same applies for the configurations of concentrations, ie for any vector of concentrations over species $\mathbf{x}_S$, we can write it as a vector over the aggregates $\mathbf{x}_A = \mathbf{\Pi} \cdot \mathbf{x}_S$. In other words, the matrix $\mathbf{\Pi}$ defines a change of basis from the vector over species to the vector over aggregates.

Definition 7 *($\mathcal{A}$-reduction of ODE semantics of reaction-based biochemical system)* We consider a reaction-based biochemical system $\mathcal{S} = (S, \mathbf{x}_0, \{r_1, ..., r_N\})$, and an aggregation $\mathcal{A}$, which defines the (linear) transform $\mathbf{\Pi}$. Then, if $\mathbf{x}_S$ is the ODE semantics of the system $\mathcal{S}$, the function $\mathbf{x}_A = \mathbf{\Pi} \cdot \mathbf{x}_S$, is the $\mathcal{A}$-reduction of ODE semantics of $\mathcal{S}$.

Given a vector $\mathbf{x}_A$, we write the change of concentration of the aggregate A_i as $[A_i] := x_{Ai}(t) = \sum\{[s_j] | s_j \in A_i\}$. Note that, since the differentiation is a linear operator, we have that $\mathbf{F}(\mathbf{x}_A) = \mathbf{\Pi} \cdot \mathbf{F}(\mathbf{x}_S)$ defines the differential maps of the $\mathcal{A}$-reduction of the ODE semantics on the aggregates.

5. MAXIMAL SELF-CONSISTENT REDUCTIONS

Not all aggregations allow us to describe the reduced ODE semantics self-consistently. Let us assume that we have an aggregation $\mathcal{A} = \{A_1\}$, where $A_1 = \{s_1, s_2\}$ and the ODE maps are $\mathbf{F}([s_1]) = k_1[s_1]$, $\mathbf{F}([s_2]) = -k_1[s_1] + k_2[s_2]$. We can write the map $\mathbf{F}([A_1]) = \mathbf{F}([s_1]) + \mathbf{F}([s_2])$, and furthermore we compute $\mathbf{F}([A_1]) = k_2[s_2]$. However, this quantity cannot be expressed by using only the concentrations of the aggregates in $\mathcal{A}$, because we need to know the concentration of the macro-species $\{s_2\}$.

Definition 8 *(Self-consistent aggregation)* We consider a rule-based biochemical system $\tilde{\mathcal{S}}$, an aggregation $\mathcal{A} =$

$\{A_1, ..., A_m\}$, and the system of differential maps $\mathbf{F}(\mathbf{x})$ as specified in Dfn.2. If the quantity $\mathbf{F}(\mathbf{x}_A) = \mathbf{\Pi} \cdot \mathbf{F}(\mathbf{x}_S)$ can be written as a vector of polynomials $\mathbf{G}(\mathbf{x}_A)$ over the variables $[A_1]$,...,$[A_m]$, then we say that the aggregation $\mathcal{A}$ is *ODE self-consistent* with respect to the rule set $\tilde{\mathcal{S}}$.

Let us observe an aggregation $\mathcal{A} = \{A_1, A_2, A_3\}$, where $A_3 = A_1 \oplus A_2$. Then we can express the concentration of A_3 as $[A_3] = [A_1] + [A_2]$. Also, we can express $[A_1]$ as $[A_1] = [A_3] - [A_2]$. This tells us that, if $\mathcal{A}$ is self-consistent with respect to the rule set, we can also instead use the aggregation $\{A_1, A_2\}$, or $\{A_2, A_3\}$, and preserve the self-consistency.

Definition 9 *(Expressiveness of the aggregation)* The expressiveness of the aggregation $\mathcal{A} = \{A_1, ..., A_m\}$, denoted $\mathrm{span}(\mathcal{A})$, we define as a set of all macro-species over S that can be formed by finite application of disjoint union $\oplus$, or its inverse $\ominus$ over the aggregates from $\mathcal{A}$.

Let us observe again the aggregation $\mathcal{A} = \{A_1, A_2, A_3\}$, where $A_3 = A_1 \oplus A_2$. The aggregations $\{A_1, A_2\}$ and $\{A_2, A_3\}$ have the same expressiveness as $\{A_1, A_2, A_3\}$.

Definition 10 *(Refinement on aggregations)* Let $\mathcal{A}_1, \mathcal{A}_2$ be two fragmentations on S. We will say that $\mathcal{A}_1$ refines $\mathcal{A}_2$, written $\mathcal{A}_1 \preceq \mathcal{A}_2$, if $\mathrm{span}(\mathcal{A}_2) \subseteq \mathrm{span}(\mathcal{A}_1)$.

We observe that the expressiveness of the aggregation $\hat{\mathcal{A}} = \{\{s_1\}, ..., \{s_n\}\}$ are all subsets of S, ie $\mathrm{span}(\hat{\mathcal{A}}) = 2^S$. Moreover, any set of aggregates $\mathcal{A} = \{A_1, ..., A_m\}$ generates a group $\mathscr{G}(\mathcal{A}) = (\mathrm{span}(\mathcal{A}), \oplus)$. The set of all subgroups of $\mathscr{G}(\hat{\mathcal{A}})$ forms a complete lattice with respect to the subgroup relation [4]. Consequently, the set of all the aggregations on S, ordered by the refinement relation $\preceq$ forms a complete lattice $(\mathscr{A}, \preceq)$. Given a group $\mathscr{G}(\mathcal{A}) = (\mathrm{span}(\mathcal{A}), \oplus)$, there are many different representations (generating subsets of elements) of it. Those of them with minimal cardinality, we define as the *least-refined* aggregations that generate $\mathscr{G}(\mathcal{A})$.

6. ALGORITHM

We assume that we can use the following operations: (1) the intersection of two macro-species (written $C_1 \cap C_2$), (2) (disjoint) union and difference of macro-species (written $C_1 \oplus C_2$ and $C_1 \ominus C_2$ respectively), (3) rule application $\boldsymbol{\beta}$ to a tuple of macro-species (written $\boldsymbol{\beta}(C_1, ..., C_m)$), (4) inverse rule application $\boldsymbol{\beta}^{-1}$ on a tuple of macro-species (written $\boldsymbol{\beta}^{-1}(C_1, ..., C_m)$), (5) the i-th component of a tuple of macro-species (written $\pi_i(C_1, ..., C_m) = C_i$). In order to be able to compute the least-refined aggregation, we assume to have the operation of (6) membership check $C \in \mathrm{span}(\mathcal{A})$, ie we are able to decide whether a given macro-specie is expressive within the aggregation. We may now define the following problem.

INPUT. The initial aggregation $\mathcal{A}_0$, and a rule-based biochemical system $\tilde{\mathcal{S}} = (S, \mathcal{C}, \mathbf{x}_0, \{R_1, ..., R_M\})$;

OUTPUT. The least-refined aggregation $\mathcal{A}$ on S, such that (i) $\mathcal{A} \preceq \mathcal{A}_0$; and (ii) $\mathcal{A}$ is self-consistent with respect to $\tilde{\mathcal{S}}$.

We recall the complete lattice $(\mathscr{A}, \preceq)$ of all the aggregations on S ordered by the relation $\preceq$. The 'necessary' aggregates for ensuring self-consistency of an aggregation with respect to a set of rules are characterized by the following Theorem (Thm. 1).

Theorem 1 An aggregation $\mathcal{A} = \{A_1, ..., A_m\}$ is ODE self-consistent with respect to the rule-based biochemical system $\tilde{\mathcal{S}} = (S, \mathcal{C}, \mathbf{x}_0, \{R_1, ..., R_M\})$ if and only if for all $A \in \mathcal{A}$, and for all rules $R_j = (\mathbf{L}, \boldsymbol{\beta}, k)$, the following conditions are satisfied:

(i) If R_j is a rule of type $C \to D$, where $C, D \in \mathcal{C}$, then $(A_i \cap C) \ominus (\boldsymbol{\beta}^{-1}(A_i \cap D) \in \mathrm{span}(\mathcal{A})$, and $(\boldsymbol{\beta}^{-1}(A_i \cap D) \ominus (A_i \cap C)) \in \mathrm{span}(\mathcal{A})$;

(ii) Else, if R_j is of type $C_1, ..., C_l \to D_1, ..., D_d$, where $C_1, .., C_l, D_1, ..., D_d \in \mathcal{C}$, then

- **(a)** For all $i = 1, ..., l$, if $A \cap C_i \neq \emptyset$, then $C_1, ..., C_{i-1}, C_i \cap A, C_{i+1}, ..., C_m \in \mathrm{span}(\mathcal{A})$;
- **(b)** For all $i = 1, ..., d$, if $A_i \cap D_i \neq \emptyset$, then $\pi_j(\boldsymbol{\beta}^{-1}(D_1 \times ... \times (D_i \cap A_i) \times ... \times D_m))$ (where $j = 1, ..., l$), belongs to $\mathrm{span}(\mathcal{A})$.

Proof 1 *(Sketch)* We observe the production and the consumption terms for an aggregate A_i with respect to a rule R_j, ie $\hat{\nu}([A_i], R_j)$, and $\check{\nu}([A_i], R_j)$. We can write them as $\hat{\nu}([A_i], R_j) = \sum_{s \in A_i} \hat{\nu}([s], R_j)$ and $\check{\nu}([A_i], R_j) = \sum_{s \in A_i} \check{\nu}(s, R_j)$. We develop these expressions by using the Dfn.2, and we derive which are the aggregates that are necessary to be included in $\mathrm{span}(\mathcal{A})$, by using the operations of intersection and set difference over macro-species.

We may now define the function $\phi : \mathcal{A} \times \{R_1, ..., R_M\} \to \mathscr{A}$, such that $\phi(A_i, R_j)$ is the set of macro-species $\mathcal{A}'$ which is self-consistent on the aggregate A_i with respect to the rule R_j. Moreover, we define a monotone operator $\Phi : \mathscr{A} \to \mathscr{A}$, such that

$$\Phi(\mathcal{A}) = \mu\mathcal{A}'.(\mathrm{span}(\mathcal{A}') = \mathrm{span}(\mathcal{A}) \cup \bigcup_{A_i \in \mathcal{A}, 1 \leq j \leq M} \phi(A_i, R_j)).$$

The result of application of Φ to the aggregation $\mathcal{A}$ is an aggregation $\mathcal{A}'$, which is a minimal aggregation that is self-consistent for each of the aggregates in $\mathcal{A}$. Finally, the algorithm computes the least fixed point of Φ which contains $\mathcal{A}_0$, ie

$$\mathcal{A} = \mu\mathcal{A}.\,(\Phi(\mathcal{A}) = \mathcal{A} \text{ and } \mathcal{A} \preceq \mathcal{A}_0)\,.$$

We recall that the least-refined aggregation is not unique. Which one the algorithm is going to output, depends on the order in which the rules are enumerated in the rule-set, and on the initial aggregation.

6.1. Complexity remark

We think of an implementation of rule-based models, which uses some particular data structure for representing macro-species and aggregates. We assume that each of the operations $(1)-(6)$ takes a constant number of instructions. The presented algorithm finds one least-refined aggregation, ie one of the minimal representations of the group $\mathrm{span}(\mu\mathcal{A}.\Phi(\mathcal{A}_0))$, with complexity $\mathcal{O}(Mg^2(\hat{l}+\hat{d}))$, where M is the number of rules, g is the cardinality of the least-refined aggregation, $\hat{l}$ is the maximum number of macro-species on the lhs of the rules in $\tilde{\mathcal{S}}$, and $\hat{d}$ the maximal number of macro-species on rhs of the rules in $\tilde{\mathcal{S}}$.

7. EXAMPLE

We illustrate a small case study, specified in a rule-based syntax Kappa. We present the terminology, and how the self-consistency condition propagates the refinement to the current aggregation along the rules.

Example 1 *(A variation of the birth-death process)* Let us consider a biochemical system $\mathcal{S}$ where we have agents A and B. The interface of agent A contains two sites a and b, and each of them has two possible internal states. The interface of agent B has n sites a_1,...,a_n, each of them with two internal states. The internal states of both A and B species alternate with rates k_{1+}, k_{1-} for A, and k_{2+}, k_{2-} for B.

$$R_1 : \mathtt{A(a^u)} \underset{\mathtt{k_{1-}}}{\overset{\mathtt{k_{1+}}}{\rightleftharpoons}} \mathtt{A(a^p)} \qquad R_2 : \mathtt{A(b^u)} \underset{\mathtt{k_{1-}}}{\overset{\mathtt{k_{1+}}}{\rightleftharpoons}} \mathtt{A(b^p)}$$

$$R_3 : \mathtt{B(a_1^u)} \underset{\mathtt{k_{2-}}}{\overset{\mathtt{k_{2+}}}{\rightleftharpoons}} \mathtt{B(a_1^p)} \quad \cdots \quad R_{n+2} : \mathtt{B(a_n^u)} \underset{\mathtt{k_{2-}}}{\overset{\mathtt{k_{2+}}}{\rightleftharpoons}} \mathtt{B(a_n^p)}.$$

We set the initial condition to be n_{A_0} copy numbers of species $\mathtt{A(a^u, b^u)}$ and n_{B_0} copy numbers of the species $\mathtt{B(a_1^u, ..., a_n^u)}$. If we would translate the rules into reactions, we would have in total $2 \cdot (4 + n \cdot 2^{n-1})$ reactions instead of $2 \cdot (n+2)$ rules. The reachable set of species S contains all the modifications of both of the agents, and therefore we have the total number of reachable species to be $|S| = 4 + 2^n$. Let us denote some of the species: $s_1 = \mathtt{A(a^u, b^u)}$, $s_2 = \mathtt{A(a^u, b^p)}$, $s_3 = \mathtt{A(a^p, b^u)}$, $s_4 = \mathtt{A(a^p, b^p)}$. The macro-species that are used in the rule-based description are for example $C_1 = \{s_1, s_2\}$, which is the lhs of rule R_1, whose transformation function is given by $\beta((s_1, s_2)) = (s_3, s_4)$, and $C_2 = \{s_1, s_3\}$, the lhs of rule R_2, with transformation $\beta((s_1, s_3)) = (s_2, s_4)$. We illustrate the dependency propagation for the aggregation $\mathcal{A}_0 = \{\{s_1\}, \{s_1, s_2, s_3, s_4\}\}$. We compute that $\phi(\{s_1\}, R_1) = \{\{s_3\}\}$, and $\phi(\{s_1\}, R_2) = \{\{s_2\}\}$. Furthermore, for $\phi(\{s_2\}, R_1) = \{\{s_4\}\}$, but, since $\{s_4\} = \{s_1, s_2, s_3, s_4\} \ominus \{s_1\} \ominus \{s_2\} \ominus \{s_3\}$, the least-refined aggregation which will be output by the algorithm is $\mathcal{A} = \{\{s_1\}, \{s_2\}, \{s_3\}, \{s_1, s_2, s_3, s_4\}\}$. On the other hand, for the initial aggregation $\mathcal{A}_0 = \{\{s_1, s_2\}\}$, we get that $\phi(\{s_1, s_2\}, R_1) = \{\{s_3, s_4\}\}$, and $\phi(\{s_1, s_2\}, R_j) = \emptyset$, for $j = 2, ..., n+2$ (the rules are *silent* with respect to the aggregate). Therefore, the algorithm will output the aggregation $\mathcal{A} = \{\{s_1, s_2\}, \{s_3, s_4\}\}$. Note that this aggregation does not contain the agents of type B. Any reduction approach that does not consider the initial aggregation will involve the analysis of the rules $R_3, ..., R_{n+2}$.

Let us consider the case when the parameters are such that $k_1^+ = k_2^+$, $k_1^- = k_2^-$. Let $\mathcal{A}_0 = \{s_1, s_2\} \cup C_3$, where $C_3 = \{\mathtt{B(a_1^u, a_2^u, ..., a_n^u)}, ..., \mathtt{B(a_1^u, a_2^p, ..., a_n^p)}\}$. The algorithm outputs the solution $\mathcal{A} = \{\{s_1, s_2\} \cup C_3, \{s_3, s_4\} \cup C_3'\}$, where $C_3' = \{\mathtt{B(a_1^p, a_2^u, ..., a_n^u)}, ..., \mathtt{B(a_1^p, a_2^p, ..., a_n^p)}\}$. Moreover, note that in this case, the system could have been described with a single rule in the set notations, where $\mathbf{L} = \{s_1, s_2, s_3\} \cup C_3$. Since the current Kappa syntax does not support aggregating agents of different types, nor such a specification, nor the reduction are possible to express in Kappa. The more radical examples which illustrate the limitations of using a particular syntax for representing aggregates can be formed when rules include complexations, but we do not present them here due to the limited space.

8. CONCLUSIONS

This paper is the theoretical characterization of maximal reductions of ODE semantics of rule-based systems. The implication of the work is that it gives theoretical grounds to developing reductions of the ODE semantics in any concrete implementation of rule-based models. Efficient representation of the species in a graph structure, the aggregates as partially defined species, and the subset relation as the topological embedding relation between the two raises plenty of questions to research. Moreover, since we characterize the *maximal* reduction, it can indicate improvements in current implementations. Another extension is the analysis of reductions of stochastic semantics in the same framework, and combining the two.

9. ACKNOWLEDGEMENTS

The authors would like to thank to Vincent Danos and Jerome Feret, for the inspiration and many useful discussions on the topic. Heinz Koeppl acknowledges the support from the Swiss National Science Foundation, grant no. 200020-117975/1. Tatjana Petrov acknowledges the support from the Swiss Initiative in Systems Biology.

10. REFERENCES

[1] W. S. Hlavacek, J. R. Faeder, M. L. Blinov, R. G. Posner, M. Hucka, and W. Fontana, "Rules for Modeling Signal-Transduction Systems," *Science's STKE*, vol. 2006, no. 344, 2006.

[2] H. Conzelmann, D. Fey, and E. D. Gilles, "Exact model reduction of combinatorial reaction," *BMC Syst Biol*, vol. 2, no. 78, pp. 342–351, 2008.

[3] J. Feret, V. Danos, J. Krivine, R. Harmer, and W. Fontana, "Internal coarse-graining of molecular systems," *Proceedings of the National Academy of Sciences*, vol. 106, no. 16, pp. 6453–6458, April 2009.

[4] B. Davey, *Introduction to Lattices and Order*, University of Oxford, 2002.