

Formal Quantitative Analysis of Reaction Networks Using Chemical Organisation Theory

Chunyan Mu¹, Peter Dittrich², David Parker^{1(✉)}, and Jonathan E. Rowe¹

¹ School of Computer Science, University of Birmingham, Birmingham, UK
d.a.parker@cs.bham.ac.uk

² Institute of Computer Science, Friedrich-Schiller-University Jena, Jena, Germany

Abstract. Chemical organisation theory is a framework developed to simplify the analysis of long-term behaviour of chemical systems. An organisation is a set of objects which are closed and self-maintaining. In this paper, we build on these ideas to develop novel techniques for formal quantitative analysis of chemical reaction networks, using discrete stochastic models represented as continuous-time Markov chains. We propose methods to identify organisations, to study quantitative properties regarding movement between these organisations and to construct an organisation-based coarse graining of the model that can be used to approximate and predict the behaviour of the original reaction network.

1 Introduction

In this paper, we study reaction networks and chemical organisation theory, in particular, investigating the applicability of probabilistic model checking to their analysis. Reaction networks are widely used in modelling chemical phenomena. They describe the dynamical interaction between processes of living systems in a formal way. Reaction networks can be difficult to understand and analyse since they can represent complex interaction behaviour over large state spaces.

Chemical organisation theory [7, 9] provides a way to analyse complex dynamical networks and reason about the long-term behaviour of chemical systems. The complex network is decomposed into a set of sub-networks called “organisations”. An *organisation* is a set of objects (for example, the species or molecules in a reaction system) which are closed and self-maintaining. Informally, closed means that no new object can be produced by the interactions within the set, and self-maintaining means that no object of the set disappears from the system, i.e., every consumed object of the set can be generated within the set. The concept of organisation allows us to lift the complex reaction network to a hierarchic structure including all stable states and states depicting accumulating molecules regarding to the organisations. The dynamics of the complex state space of the reaction network can then be mapped to movements among the set of organisations. Building a chemical organisation-based model thus helps us to model the structure and behaviour of complex reaction networks, and to simplify the dynamical analysis of the overall system.

In order to study the evolution of reaction networks, we apply probabilistic model checking, a formal verification technique for modelling and analysis of

systems with stochastic behaviour. It has been used to study models across a wide range of application domains, including chemical and biological systems. Probabilistic model checking is based on the exhaustive construction and analysis of a state-based probabilistic model, typically a Markov chain or variant. In this work, we model the reaction networks as continuous-time Markov chains. Quantitative properties of interest about the system being analysed are formally specified using temporal logic. Here we use CSL (Continuous Stochastic Logic) [2] with rewards, a quantitative extension of the temporal logic CTL.

Specifically, in this work, we use CSL model checking of continuous-time Markov chains to investigate connections between chemical organisations using model decompositions into strongly connected components (SCCs). We develop an algorithm to automatically find organisations, and then perform a quantitative dynamical analysis in terms of organisations, asking, for example, “what is the probability of moving from one organisation to another?” or “what is the expected time to leave an organisation?” A coarse grained Markov chain model of hierarchic organisations for a given reaction network is then constructed as a result. We implement our techniques as an extension of the probabilistic model checking tool PRISM [15], and illustrate the approach on a set of example reaction networks. Approximating and predicting the system behaviour over time evolution is a direct application of our coarse grained model.

Related work. There are various approaches to modelling the dynamics of reaction networks. Feinberg and Horn [8] proposed methods to identify *positive* stationary states in which *all* molecular species are present in a network. Heinrich and Schuster [12] studied network structure based on *flux modes*, each of which specifies a set of reaction rules that can take place at a steady state and thus implies a set of species participating in those reactions. Species regarding to a flux modes were not required to be *self-maintaining* or *closed* however. We are more interested in the stationary states in which a *subset* of species are present, which is formalised in organisation theory [7]. In that area, the focus was typically on *qualitative* properties, and ODEs [6], approximating the evolution of reaction networks in *continuous* dynamical systems. Kreyssig et al. [14] studied the effects of small particle numbers on long-term behaviours in *discrete* biochemical systems. We build on their notion of discrete organisation but focus on *quantitative* analysis of the transitive dynamics among the organisations, which was not considered in [14]. Other approaches for approximate analysis of discrete models of reaction networks include the use of Linear Noise Approximation [4], the Central Limit Approximation [3] and “sliding window” abstractions [17].

2 Probabilistic Model Checking

Probabilistic model checking is a variant of *model checking* [5], a well-established formal method to automatically verify the correctness of real-life systems. Classical model checking answers the question of whether the behaviour of a given system satisfies a property or not. It thus requires two inputs: a description of

the system and a specification of one or more required properties of that system, normally in temporal logic (such as CTL or LTL).

In probabilistic model checking, the models are extended with information about the likelihood that transitions take place. In practice, these models are usually Markov chains or Markov decision processes. In this work, we model the reaction systems as continuous-time Markov chains (CTMCs), which are widely used in fields such as performance analysis or systems biology to model systems with stochastic real-time behaviour. Formally, we define them as follows.

Definition 1 (CTMC). A CTMC is a tuple $\mathcal{A} = (Q, Q_0, \Delta, L)$, where: Q is a finite set of states; $Q_0 \subseteq Q$ is the set of initial states; $\Delta : Q \times Q \rightarrow \mathbb{R}_{\geq 0}$ is the transition rate matrix; $L : Q \rightarrow 2^{AP}$ is a labelling function assigning, to each state $q \in Q$, a set of atomic propositions, from a set AP , that are true in q .

The transition rate matrix Δ assigns a *rate* to each pair of states in the CTMC, which is used as the parameter of an exponential distribution.

In this work, the probabilistic temporal logic CSL (Continuous Stochastic Logic) is used to formally represent properties of reaction networks. It was originally introduced by Aziz et al. [1] and extended by Baier et al. [2]. The extended version allows for the specification of reward (or cost) properties, to reason about rewards (or costs) that have been attached to a CTMC. The extended version of CSL that we use allows us to represent properties such as “the probability of all of species A degrading within t time units is at most 0.1” or “the expected time elapsed before a B molecule first appears is at most 10”.

Definition 2 (CSL syntax). An (extended) CSL formula is an expression Ψ derived from the grammar:

$$\Psi ::= \text{true} \mid p \mid \neg\Psi \mid \Psi \wedge \Psi \mid P_{\bowtie\lambda}(\Psi \ U^I \ \Psi) \mid R_{\bowtie r}[\diamond\Psi]$$

where $p \in AP$ an atomic proposition, $\lambda \in [0, 1]$ is a probability threshold, $r \in \mathbb{R}_{\geq 0}$ is a reward threshold, $\bowtie \in \{<, \leq, \geq, >\}$ and I is an interval of $\mathbb{R}_{\geq 0}$.

CSL formulas are described over the states of a Markov chain. A state q satisfies $P_{\bowtie\lambda}(\psi)$ if the probability of taking a path from q satisfying ψ is in the interval specified by $\bowtie \lambda$. Here, the path formula ψ is an “until” operator: $\Psi \ U^I \ \Psi'$ asserts that Ψ' is satisfied at some future time point within interval I , and that Ψ is true up until that point. We omit the interval I when $I = [0, \infty)$. Common derived operators include: “eventually” $\diamond^I\Psi := \text{true} \ U^I \ \Psi$ and “always” $\square^I\Psi := \neg\diamond^I\neg\Psi$. For example, $P_{\leq\lambda}(\square^I\Psi) \equiv P_{\geq 1-\lambda}(\diamond^I\neg\Psi)$. The R operator is used for reward properties: $R_{\bowtie r}[\diamond\Psi]$ is true from state q if the expected reward cumulated before a state satisfying Ψ is reached meets the bound $\bowtie r$. We also use *numerical queries*, e.g., $R_{=?}[\diamond\Psi]$, which return the actual expected reward (or probability), rather than check it against a bound. Rewards and costs are treated identically: here, we will use the R operator to formalise properties about the expected time elapsing before an event’s occurrence. We omit a full definition of the semantics of CSL with respect to a Markov chain. Full details can be found in, for example, [2].

3 Modelling Reaction Networks with CTMCs

A reaction network consists of a set of molecules (or, molecular species to be more precise) and a set of reaction rules.

Definition 3. A reaction network is a pair $(\mathcal{M}, \mathcal{R})$ consisting of a set of possible molecular species \mathcal{M} , and a set $\mathcal{R} \subseteq \mathcal{P}_{\mathcal{M}}(\mathcal{M}) \times \mathcal{P}_{\mathcal{M}}(\mathcal{M})$ of possible reactions among those species, where $\mathcal{P}_{\mathcal{M}}(C)$ denotes the set of all multisets of elements over the set C . For a reaction $(R, P) \in \mathcal{R}$, the multisets R and P denote the reactants and products of the reaction, respectively, and we write $R(s)$ and $P(s)$ for the number of molecules of species s consumed by (reactants) and produced by (products) the reaction, respectively.

For simplicity, we write $s_1 + s_2 + \dots + s_n \rightarrow s'_1 + s'_2 + \dots + s'_n$, instead of $(\{s_1, s_2, \dots, s_n\}, \{s'_1, s'_2, \dots, s'_n\}) \in \mathcal{R}$ to denote the existence of a reaction.

There are multiple ways in which we can obtain a dynamical model given a reaction network. One way is to consider (real-valued) concentrations of each molecular species and then represent the (deterministic) behaviour of the reactions as a set of ordinary differential equations. Here, we take a *discrete, stochastic* view of the network, modelling the (integer-valued) population count of each species and considering its evolution as a stochastic process, and in particular as a continuous-time Markov chain [11]. The latter is particularly appropriate when the numbers of molecules can be assumed to be relatively small in practice, and is the approach that we take in this work.

Furthermore, we will assume also that the reaction network is executing within a finite volume, which is modelled by limiting the total number $N_{\max} \in \mathbb{N}$ of molecules that can be present at any given time [14]. We also need to define the rates at which reaction events occur in the CTMC. To retain a general approach, we allow an arbitrary function rate_r from reactant populations to rate values for each reaction r .

Definition 4 (CTMC for reaction network). Given a reaction network $(\mathcal{M}, \mathcal{R})$, a volume limit $N_{\max} \in \mathbb{N}$ and a rate function $\text{rate}_r : \mathbb{N}^{\mathcal{M}} \rightarrow \mathbb{R}_{\geq 0}$ for each $r \in \mathcal{R}$, we define the corresponding CTMC $\mathcal{A} = (Q, Q_0, \Delta, L)$ where:

- $Q = \{q : \mathcal{M} \rightarrow \mathbb{N} \mid \sum_{s \in \mathcal{M}} q(s) \leq N_{\max}\}$

is the set of population counts of \mathcal{M} and Δ is defined as follows. For states $q, q' \in Q$ and reaction $(R, P) \in \mathcal{R}$, we write $q \xrightarrow{(R, P)} q'$ if and only if, for each species $s \in \mathcal{M}$, we have $q(s) \geq R(s)$ and $q'(s) = q(s) - R(s) + P(s)$, and $\sum_{s \in \mathcal{M}} q'(s) \leq N_{\max}$. Then, for any $q, q' \in Q$, we have:

- $\Delta(q, q') = \sum \{|\text{rate}_r(q) \mid r \in \mathcal{R} \text{ and } q \xrightarrow{r} q'\}$, and we call r the transition label of $q \xrightarrow{r} q'$.

Q_0 can be any subset of Q representing initial configurations of interest, and L can be any labelling function over Q that identifies states with relevant properties.

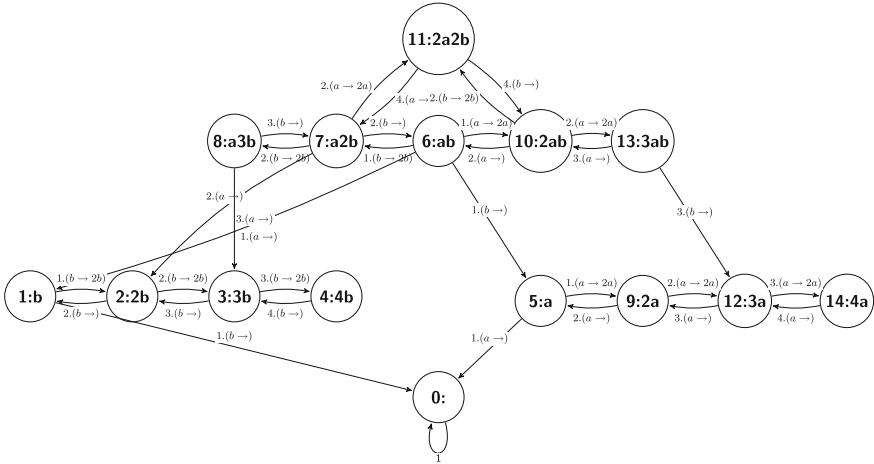


Fig. 1. State transition graph of the CTMC for Example 1. State labels show index and population count, e.g., 11 : 2a2b denotes that there are 2a and 2b in state 11.

For our examples we usually follow the general law of mass-action by setting $\text{rate}_r(q) = \lambda_r \cdot \prod_{s \in R} q(s)$ with λ_r being a kinetic rate constant for reaction r (and assuming the stoichiometric coefficient of each reactant is at most one).

Each state $q \in Q$ of the CTMC gives the number $q(s)$ of molecules of each species $s \in \mathcal{M}$ that are currently present. For a state q , we also write $\phi(q)$ to denote the set of molecular species that are present, i.e., $\phi(q) = \{s \mid q(s) > 0\}$, and define $\phi(Q') = \cup_{q \in Q'} \phi(q)$ for a set of states $Q' \subseteq Q$. We let $\text{Acc}(q) \subseteq Q$ denote the set of states that are reachable from q .

Example 1. Consider the reaction network \mathcal{A} with species $\mathcal{M} = \{a, b\}$ and reactions $\mathcal{R} = \{a + b \rightarrow a + 2b, a \rightarrow 2a, b \rightarrow 2b, a \rightarrow \emptyset, b \rightarrow \emptyset\}$. Assume the volume of the system is $N_{\max} = 4$, and the rate of each reaction is of second order, i.e., $\#a \cdot \#b$, $\#a \cdot \#a$, $\#b \cdot \#b$, $\#a \cdot \#a$, $\#b \cdot \#b$, respectively, where $\#a$ denotes the number of molecules of species a . We obtain a CTMC with 15 states (Fig. 1). Note that throughout this work, without loss of generality, we use second-order kinetic laws in order to avoid any complicated issues regarding units and scaling.

4 Chemical Organisation Theory and SCC Decomposition

Chemical organisation theory [7] provides a way to cope with the complex “constructive” dynamics of a reaction network by deriving a set of organisations [10], and then mapping the movement through state space to a movement between organisations. Such an abstract view allows us to analyse and predict the dynamical behaviour of a complex reaction network more easily. An organisation is a set of molecules that is algebraically *closed* and *self-maintaining*. A subset $C \subseteq \mathcal{M}$ is called “closed” if no molecules outside C can be produced by applying all reactions possible in C to multisets over C ; a subset $S \subseteq \mathcal{M}$ is “self-maintaining”

if all reactions that are able to fire in S can occur at certain strictly positive rates without reducing the amount of any species of S . We say that a reaction $(R, P) \in \mathcal{R}$ can *fire* in a set of species S , if S contains each reactant species from R .

Definition 5 (Organisation [7]). *A subset of $O \subseteq \mathcal{M}$ is a chemical organisation if it is closed and self-maintaining, that is, if for all $(R, P) \in \mathcal{R}$, $R \subseteq O$ implies $P \subseteq O$ (closure), and there exists a strictly positive flux vector $v > 0$ such that $N_O \cdot v \geq 0$ with N_O being the stoichiometric matrix of the reactions that can fire in O (self-maintenance). An entry $n_{i,r}$ of the stoichiometric matrix $N_O = (n_{i,r})$ denotes the number of molecules of species $i \in \mathcal{M}$ produced when firing reaction r once. The product with a flux vector, $N \cdot v$, results in a vector of the net-production rates for each species for the respective reaction rates v .*

Note that the set of organizations is defined with respect to a reaction network and thus independent from an initial state. However, given an initial state, there is in general only a subset of organizations reachable.

As discussed above, we model the dynamics of a reaction network as a Markov chain. A state is defined as the number of each molecular species and, with a limited total number of molecules, cases of both too few or too many molecules can prevent reaction rules being fired. As a consequence, we need to define *discrete organisations*, and the states contributing to generate them. From now on, \mathcal{R}_q denotes the reactions firing in any state reachable from a state q .

Definition 6 (Discrete organisation and internal generator [14]). *Let $(\mathcal{M}, \mathcal{R})$ be a reaction network. A subset of species $D \subseteq \mathcal{M}$ is called a discrete organisation if there is a state $q \in Q$ such that: (i) $\phi(\text{Acc}(q)) = D$ (closure); and (ii) there is a sequence of transition labels (r_1, \dots, r_k) where $r_i \in \mathcal{R}$ such that $\cup_{i=1}^k \{r_i\} = \mathcal{R}_q$ and $q' = (r_k \circ \dots \circ r_1)(q)$ satisfies $\forall s \in D : q'(s) \geq q(s)$ (self-maintenance), where \circ denotes a composition operator, i.e., $r_j \circ r_i(q_i) = r_j(q_j)$ for $q_j = r_i(q_i)$. Such a state q is called an internal generator of the discrete organisation.*

Definition 7 (Generator). *A state $q' \in Q$ is called a generator of organisation D iff $\exists q \in \text{Acc}(q')$ such that q is an internal generator of D .*

Note that, in general, the organisation D generated by a state q' is not unique. However, if q is an internal generator, there is only one organisation it generates. Unless specifically stated otherwise, we say *organisation* rather than *discrete organisation* in the rest of the paper.

Example 2. The discrete organisations for Example 1 are: $\{a, b\}$, $\{a\}$, $\{b\}$, $\{\}$ and the corresponding generators are, respectively (cf. Fig. 1):

$\{6, 7, 8, 10, 11, 13\}$, $\{5, 6, 7, 8, 9, 10, 11, 12, 13, 14\}$, $\{1, 2, 3, 4, 6, 7, 8, 10, 11, 13\}$, $\{0, 1, \dots, 14\}$.

In order to analyse the system behaviour and perform an organisation-based quantitative analysis of the reaction network, we study the connections between chemical organisations and the decompositions into strongly connected components (SCCs) of the Markov chain.

Definition 8 (SCC [13]). A strongly connected component (SCC) of a Markov chain is a maximal set of states T such that, for every pair of states q and q' , there is a path from q to q' .

Intuitively, in the Markov chain for a reaction network, SCCs are important for an organisation-based analysis. However, some but not all SCCs correspond to organisations. In the next section, we will describe an algorithm to find organisations based on a decomposition into SCCs and then identifying those self-maintaining a set of species. We first note that *bottom* strongly connected components do relate to organisations.

Definition 9 (BSCC). A bottom strongly connected component (BSCC) is an SCC T from which no state outside T is reachable from T .

Proposition 1. Each BSCC corresponds to a (unique) organisation, which is generated (uniquely) by any state of that BSCC.

However, there are organisations whose internal generators are *not* contained in any BSCC. In order to also include such organisations, we call SCCs that correspond to an organisation *good* SCCs.

Definition 10 (Good SCC). An SCC T is called good if it contains a cycle of the firing of every “possible” reaction rule, i.e., those whose reactants R appear in the SCC ($R \subseteq \{\phi(q) \mid q \in T\}$).

Example 3. All SCCs are good in Example 1.

Clearly, some generators can contribute to multiple organisations. This makes it more difficult to decompose the Markov into its sets of generators. However, internal generators located in good SCCs contribute uniquely to an organisation.

Proposition 2. A generator g is an internal generator of organisation D iff it is located in a good SCC T such that: $g \in T \wedge \bigcup_{q \in T} \phi(q) = D$.

Proposition 3. Given a good SCC T , let $A = \phi(T)$, if A is closed, then A is a discrete organisation, then $\{q \mid q \in T\}$ is the set of internal generators of A .

Example 4. In Example 1, the internal generators of organisations $\{a, b\}$, $\{a\}$, $\{b\}$ and $\{\}$ are $\{6, 7, 8, 10, 11, 13\}$, $\{5, 9, 12, 14\}$, $\{1, 2, 3, 4\}$ and $\{0\}$, respectively.

5 Organisation-Based Analysis of Reaction Networks

In this section, we propose techniques for quantitative organisation-based analysis of reaction networks. We first introduce an algorithm to find the set of organisations for a specific reaction network. We then use probabilistic model checking to analyse quantitative properties regarding the dynamics of the network with respect to its organisations. Such organisation-based quantitative analyses can be used to construct the structure of organisation-based coarse-grained model, and provide a framework to approximate the complex dynamical behaviours of the original reaction networks in our next step.

5.1 Finding Organisations

Computing the organisations of a reaction network requires an analysis of the strongly connected components of its Markov chain’s underlying transition graph. Since every state in a good SCC is an internal generator of an organisation, we identify good SCCs to find the organisations of the reaction network. Algorithm 1 presents the procedure for finding organisations of a given reaction network modelled as a CTMC. It is based on the following procedures:

- $\text{Tarjan}(\mathcal{A})$ returns the set of strongly connected components of the Markov chain \mathcal{A} , using Tarjan’s SCC algorithm [16] on the underlying digraph;
- $\text{findGoodSCCs}(\text{SCC})$ returns the “good” part SCC_G of \mathcal{A} in which each possible reaction rule is able to be fired;
- find a set of closed molecules appearing in each $\text{scc} \in \text{SCC}_G$, and its relevant internal generators i.e., states in scc which generate the organisation.

Algorithm 1. Finding organisations of a reaction network

```

Data: CTMC  $\mathcal{A}$  of reaction network  $(\mathcal{M}, \mathcal{R})$ 
Result:  $\mathcal{O}$  as a set of organisations,  $\mathcal{G} : \mathcal{O} \rightarrow \mathcal{P}(Q)$  as a mapping from
organisations to sets of internal generators

 $\mathcal{O} = \{\}$ ;
 $\mathcal{G} = \{\}$ ;
 $\text{SCC} \leftarrow \text{Tarjan}(\mathcal{A})$ ;
 $\text{SCC}_G \leftarrow \text{findGoodSCCs}(\text{SCC}) \cup \text{BSCC}$ ;
for  $\text{scc} \in \text{SCC}_G$  do
     $M_g \leftarrow \{\phi(q) \mid q \in \text{scc}\}$  ;
    if  $M_g$  is closed then
        if  $M_g \notin \mathcal{O}$  then
             $\mathcal{O} \leftarrow \mathcal{O} \cup M_g$           /* add new organisation */ ;
             $\mathcal{G}(M_g) \leftarrow \{q \mid q \in \text{scc}\}$  /* add new internal generators */ ;
        else
             $\mathcal{G}(M_g) \leftarrow \mathcal{G}(M_g) \cup \{q \mid q \in \text{scc}\}$  /* update generators */ ;
        end
    end
end
return  $\mathcal{O}, \mathcal{G}$ .
    
```

5.2 Organisation-Based Probabilistic Analysis

We now illustrate, via several examples, how we derive quantitative organisation-based properties of reaction networks. We implemented the organisation and generator detection process described above in the PRISM model checker, along with

a translator that converts descriptions of reaction networks into the PRISM modelling language to allow construction of the corresponding CTMC. Organisation-based properties of the network, such as probabilities (bounds or average) of the movements among organisations, or the expected time to leave or stay at an organisation, are computed using CSL formulae.

Example 5. Consider the reaction network with molecular species $\mathcal{M} = \{a, b\}$ and reactions rules include: $\{a + b \rightarrow a, a \rightarrow 2a, b \rightarrow 2b, a \rightarrow \emptyset, b \rightarrow \emptyset\}$ with stochastic rates: $\#a \cdot \#b, (\#a)^2, (\#b)^2, (\#a)^2, (\#b)^2$ respectively.

Letting $N_{\max} = 10$, the resulting CTMC has 66 states and 201 transitions, and there are 4 SCCs ($\{a > 0, b > 0\}, \{a > 0, b = 0\}, \{a = 0, b > 0\}, \{a = b = 0\}$) with 1 BSCC ($\{a = b = 0\}$). For reference, we show both the PRISM language model and the CTMC for this example in the Appendix (Figs. 5 and 6).

The first property we consider is the probability of moving between organisations. Specifically, the probability of moving from O_1 to O_2 can be specified in CSL as: $P_{=?}[o_1 U o_2]$, where o_1 and o_2 are atomic propositions labelling states which represent internal generators of organisations O_1 and O_2 . In this example, all SCCs are good and each (good) SCC generates exactly one organisation. To visualise the movement between organisation, we analyse the property above for each pair of organisations and construct the abstract transition graph shown in Fig. 2 (left). Blocks are labelled with organisations and, for each possible transition between organisations, we show the range of probabilities (over all states in the source organisation) and the average value (over the same set of states).

We also consider the expected time to leave (the generators of) each organisation. The CSL property to specify this, for some organisation O_i , is: $R_{=?}[\diamond \neg o_i]$, where o_i is an atomic proposition as above, \neg denotes negation and we assign a state reward of 1 to every state of the CTMC, indicating the amount of reward that is accumulated per unit time until $\neg o_i$ is satisfied. This value is also shown for each organisation in Fig. 2 (left), inside the block for the corresponding organisation.

Finally, we consider the effect of making some constructive perturbation to the reaction network, by adding rules to create species with a small rate. Figure 2 (right) shows the results of the same analysis described above for the following constructive perturbation: $\{\emptyset \rightarrow a, \emptyset \rightarrow b\}$ both with reaction rate $\gamma = 0.01$. The result shows that, generating a and b with a small rate can cause an upward movement and slightly affect the system's behaviour. Note that the upward flow introduced by the constructive perturbation leads to a smoother flow.

Example 6. Consider now the reaction network with $\mathcal{M} = \{a, b, c, d\}$ and $\mathcal{R} = \{a + b \rightarrow a + 2b, a + d \rightarrow a + 2d, b + c \rightarrow 2c, c \rightarrow b, b + d \rightarrow c, b \rightarrow \emptyset, c \rightarrow \emptyset, d \rightarrow \emptyset\}$. We will consider two groups of rates for a purpose of comparison in Sect. 7: R2:: $\#a * \#b, \#b * \#c, \#c * \#c, \#b * \#d, \#b * \#b, \#c * \#c, \#d * \#e$ and R1:: $\#a * \#b, \#b * \#c, \#c, \#b * \#d, \#b * \#b, \#c, \#d$. We only use R1 in this section. Figure 7 (in the Appendix) shows the structure of the CTMC for $N_{\max} = 5$. Even for a small volume $N_{\max} = 5$, the structure is quite complex: 126 states, 386 transitions, 28 SCCs and 6 BSCCs.

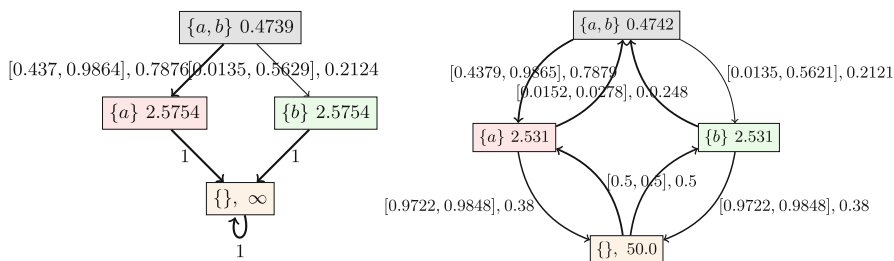


Fig. 2. Organisation-based transition model without/with constructive perturbation

Figure 8 (in the Appendix) illustrates, in the same fashion as above, the transition probabilities between *all* SCCs of the CTMC, and the expected time to leave them. Note that not all SCCs are good SCCs in this example: we highlight good SCCs in colour in Fig. 8. For instance, the SCC labelled as (99, 105; 0.25) is not a good one. There are two states in this SCC: state 99 ($a = 2, b = 0, c = 1, d = 1$) and state 105 ($a = 2, b = 1, c = 1, d = 1$). The set of molecules appearing in this node is closed, but reaction rules such as $c \rightarrow \emptyset$ and $d \rightarrow \emptyset$ cannot be fired within the SCC and it is therefore not good. In addition, the SCC labelled as (12, 27; 0.25) is also not a good one. It contains state 12 ($a = 0, b = 0, c = 2, d = 1$) and state 27 ($a = 0, b = 1, c = 1, d = 1$). The set of molecules appeared in this node is closed, but reaction rule $c \rightarrow \emptyset$ is unable to be fired locally, i.e., this decay will only introduce transitions to other SCCs. Similar cases can happen for some of the other reaction rules.

Figure 9 (in the Appendix) presents the transition probabilities between good SCCs only, and the expected time to leave them. Note that multiple good SCCs can contribute to the generation of one organisation. For instance, both good SCCs labelled 65... and 98... contribute to organisation $\{a, b, c\}$. Based on this graph, we can build up the transition graph over organisations. Figure 3 presents the transition probabilities (internal generators of) organisations, and the expected time to leave each of them. It helps us to understand the movement between organisations and can be viewed as an abstract model capturing the behaviour of the reaction network at the level of organisations.

In addition, we also present transition graphs over the lattice of molecules (states in which a set of molecules in the lattice with positive numbers) for a quantitative analysis for organisations from a different point of view, see Fig. 10

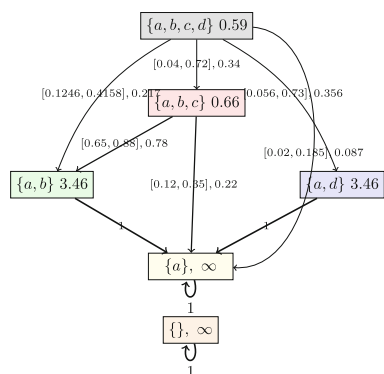


Fig. 3. Transition probabilities (bounds and average) between generators of organisations for Example 6 with $N_{\max} = 5$ and the expected time to leave them.

in the Appendix. The transition probabilities are given in bound. Specifically, the probability of movement from $\{a, b, c\}$ to $\{a, b\}$ can be specified as: $P_{=?}[(a > 0 \wedge b > 0 \wedge c > 0) \cup (a > 0 \wedge b > 0 \wedge c = 0)]$. Note that Figs. 3, 10 can be used to build coarse-grained model from different view.

6 Organisation-Oriented Interval Markov Chain

The organisation-oriented transition graph generated by the quantitative analysis can be used to build a coarse-grained model (with either interval based or average based probabilistic transitions). Such a coarse-grained model can mimic the complex reaction behaviours of the reaction network in an abstract way, whose state space and movement structure are much smaller. We can then perform approximation, prediction, and quantitative analysis upon the coarse-grained model instead of the complex concrete model. This section focuses on formalising the interval-based organisation coarse-grained model. Specifically, our quantitative analysis computes an organisation-oriented interval Markov chain, in which each abstract state is specified by a set of internal generators of an organisation, and the abstract transition provides the information about the uncertainty of the abstract behaviours of the system. Probabilities of moving from one abstract state to another are given by the lower and upper bounds, which provides the under and over approximation of the concrete probabilities.

Definition 11 (Organisation-oriented interval Markov chain). *An organisation-oriented interval Markov chain is a tuple $\mathcal{A}_I^\sharp = (Q^\sharp, Q_0^\sharp, \Delta^\sharp, L)$, where*

- Q^\sharp is a finite set of abstract states, each of which $q^\sharp \in Q^\sharp$ is a set of internal generators of an organisation o : $q^\sharp \subseteq \mathcal{G}_I(o)$;
- $Q_0^\sharp \subseteq Q^\sharp$ is the set of initial abstract states;
- $\Delta^\sharp : Q^\sharp \times Q^\sharp \rightarrow [lb, ub]$ is the abstract transition matrix, s.t. $\Delta^\sharp(q^\sharp, q^{\sharp'}) = [lb, ub]$, where lb and ub are the lower and upper bound of a set of concrete probabilistic transitions: $\{\Delta(q, q') \mid q \in q^\sharp, q' \in q^{\sharp'}\}$ specified in the relevant concrete model \mathcal{A} respectively;
- $L : Q^\sharp \rightarrow 2^{AP}$ is a labelling function over Q^\sharp that identifies properties of interest.

An *abstract path* is an execution of the organisation-oriented interval Markov chain.

Definition 12 (Abstract path). *An abstract path ω^\sharp is a non-empty sequence of states $q_0^\sharp q_1^\sharp \dots$, where $q_i^\sharp \in Q^\sharp$ and $\forall i. \Delta^\sharp(q_i^\sharp, q_{i+1}^\sharp) \subseteq (0, k]$ where $0 < k \leq 1$. The set of all finite and infinite paths of \mathcal{A}_I^\sharp starting in state q^\sharp are denoted as: $\text{Path}_{\text{fin}}^{\mathcal{A}_I^\sharp}(q^\sharp)$ and $\text{Path}^{\mathcal{A}_I^\sharp}(q^\sharp)$ respectively.*

Definition 13 (Probability bounds of abstract paths). *The lower (Prob^-) and upper bound (Prob^+) of the probability of a finite abstract path $\omega_{\text{fin}}^\sharp$ starting from state q^\sharp are respectively:*

$$\text{Prob}_{q^\sharp}^-(\omega_{\text{fin}}^\sharp) \triangleq \begin{cases} 1 & \text{if } n = 0 \\ \text{Prob}_{q^\sharp}^-(\omega_0^\sharp, \omega_1^\sharp) \times \cdots \times \text{Prob}_{q^\sharp}^-(\omega_{n-1}^\sharp, \omega_n^\sharp) & \text{otherwise} \end{cases}$$

$$\text{Prob}_{q^\sharp}^+(\omega_{\text{fin}}^\sharp) \triangleq \begin{cases} 1 & \text{if } n = 0 \\ \text{Prob}_{q^\sharp}^+(\omega_0^\sharp, \omega_1^\sharp) \times \cdots \times \text{Prob}_{q^\sharp}^+(\omega_{n-1}^\sharp, \omega_n^\sharp) & \text{otherwise} \end{cases}$$

where n denotes the length of the abstract path, ω_i^\sharp denotes the i^{th} element of ω^\sharp .

We focus on the reachability properties, for instance, the probability bounds of reaching or moving to an organisation of interests from another.

Definition 14 (Reachability properties). *Let \mathcal{A}_I^\sharp be an organisation-based interval Markov chain. The lower and upper bound of the probability of reaching an abstract state $q^{\sharp'}$ from q^\sharp are computed by:*

$$\text{Reach}_{\mathcal{A}_I^\sharp}^-(q^\sharp, q^{\sharp'})$$

$$\triangleq \min \left\{ \sum_{\omega^\sharp \in \text{Path}_{\text{fin}}^{\mathcal{A}_I^\sharp}(q^\sharp)} \{ \text{Prob}_{q^\sharp}^-(\omega^\sharp) \mid \omega_0^\sharp = q^\sharp \wedge \exists i \geq 0. \omega_i^\sharp = q^{\sharp'} \}, 1 \right\}$$

$$\text{Reach}_{\mathcal{A}_I^\sharp}^+(q^\sharp, q^{\sharp'})$$

$$\triangleq \min \left\{ \sum_{\omega^\sharp \in \text{Path}_{\text{fin}}^{\mathcal{A}_I^\sharp}(q^\sharp)} \{ \text{Prob}_{q^\sharp}^+(\omega^\sharp) \mid \omega_0^\sharp = q^\sharp \wedge \exists i \geq 0. \omega_i^\sharp = q^{\sharp'} \}, 1 \right\}.$$

Our organisation-oriented interval Markov chain should safely approximate the concrete CTMC describing the probabilistic behaviours of the system.

Theorem 1 (Soundness of the abstract semantics). *Let \mathcal{A}_I^\sharp and \mathcal{A} be the coarse-grained model and the relevant concrete model of a reaction network respectively, $\forall q^\sharp = Q, q^{\sharp'} = Q' \in Q^\sharp \subseteq Q$, we have:*

$$\text{Reach}_{\mathcal{A}_I^\sharp}^-(q_1^\sharp, q_2^\sharp) \leq \text{Reach}_{\mathcal{A}}^-(Q, Q'), \quad \text{Reach}_{\mathcal{A}_I^\sharp}^+(q_1^\sharp, q_2^\sharp) \geq \text{Reach}_{\mathcal{A}}^+(Q, Q').$$

Proof. Let ω^\sharp denote an abstract path starting from q^\sharp and reaching $q^{\sharp'}$. For any $\omega^\sharp \in \text{Path}_{\mathcal{A}_I^\sharp}(q^\sharp, q^{\sharp'})$, such as $\omega_0^\sharp = q^\sharp, \omega_{|\omega^\sharp|}^\sharp = q^{\sharp'}$, assume $|\omega^\sharp| = n \in \mathbb{N}$, and let $\omega \in \text{Path}_{\mathcal{A}}(q, q')$ denote a concrete path starting from a state in Q and reaching a state Q' , we have:

$$\begin{aligned}
 \text{Reach}^-(q^\sharp, q^{\sharp'}) &= \sum_{\omega^\sharp} \text{Prob}_{q^\sharp}^-(\omega^\sharp) = \sum_{\omega^\sharp} \left(\text{Prob}^-(\omega_0^\sharp, \omega_1^\sharp) \times \cdots \times \text{Prob}^-(\omega_{n-1}^\sharp, \omega_n^\sharp) \right) \\
 &= \sum_{\omega^\sharp} \left(\prod_{i=0}^{n-1} \inf\{\text{Prob}(q_i, q_{i+1}) \mid q_i \in \omega_i^\sharp, q_{i+1} \in \omega_{i+1}^\sharp\} \right) \\
 &\leq \sum_{\omega^\sharp} \inf\{\text{Prob}(q_0, q_n) \mid q_0 \in \omega_0^\sharp, q_n \in \omega_n^\sharp\} \\
 &= \sum_{\omega} \inf\{\text{Prob}(\omega_0, \omega_n) \mid \omega_0 \in Q, \omega_n \in Q'\} = \text{Reach}^-(Q, Q').
 \end{aligned}$$

Similarly, we have $\text{Reach}^+(q^\sharp, q^{\sharp'}) \geq \text{Reach}^+(Q, Q')$. □

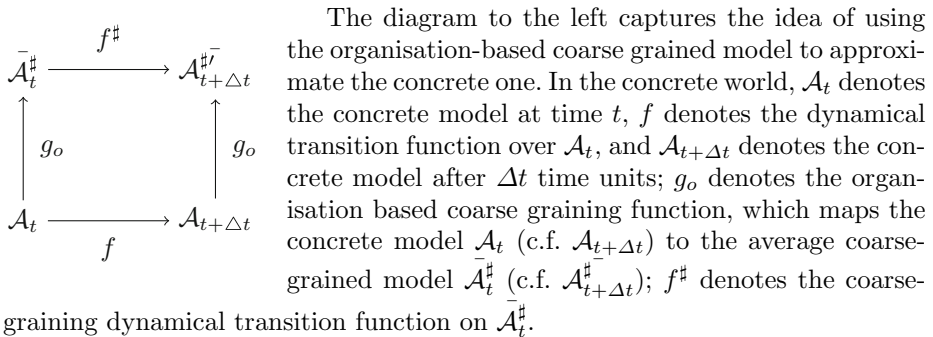
Example 7. Consider again the reaction network described in Example 6:

- by applying the coarse-grained model shown in Fig. 3, we can calculate the probability of movement from $q_{\{a,b,c,d\}}^\sharp$ to $q_{\{a,b\}}^\sharp$ is: $[0.1506, 1]$; the probability of movement from $q_{\{a,b,c,d\}}^\sharp$ to $q_{\{a\}}^\sharp$ is: $[0.2314, 1]$;
- by applying the concrete model shown in Fig. 7, we obtain the probability of movement from $Q_{\{a,b,c,d\}}$ to $Q_{\{a,b\}}$ is: $[0.1776, 0.8268]$; the probability of movement from $Q_{\{a,b,c,d\}}$ to $Q_{\{a\}}$ is: $[1, 1]$.

Note that our abstract model safely approximates the concrete one.

7 An Application of the Coarse-Grained Model

This section presents an application of our organisation-oriented coarse grained model. We address the following problem: given a reaction network and a fixed number of the maximum population of the system, construct the average-based organisation coarse-grained model $\bar{\mathcal{A}}^\sharp$ (focus on the average transition probabilities between abstract states for simplicity and intuition, this can be replaced by interval-based transitions directly), can we predict the behaviour of the system at any future time using $\bar{\mathcal{A}}^\sharp$?



We apply the traditional “master equation” approach to calculate the stochastic time evolution of the reaction network. We briefly review the main features of the master equation formalism for our purpose of calculating the prediction of an reaction network at any future time. The probability function

$P(X_1, X_2, \dots, X_n; t)$ defines the probability of number of X_i molecules of species S_i for $i \in \{1, \dots, n\}$ at time t . This function thus describes the “stochastic state” of the system at time t . The master equation is the time-evolution equation for the function $P(t)$. Function $P(X_1, \dots, X_n; t+dt)$ can be viewed as the sum of the probabilities of different ways that the system can reach the state X_1, \dots, X_n at time $t+dt$: $P(X_1, \dots, X_n; t+dt) = P(X_1, \dots, X_n; t)(1 - \sum_{i=1}^m \alpha_i dt) + \sum_{j=1}^n \beta_j dt$, where the quantity $\beta_j dt$ denotes the probability that the system is entering the state (X_1, \dots, X_n) at time $t+dt$, and the quantity $\alpha_i dt$ denotes the probability that is leaving (X_1, \dots, X_n) at time t . To avoid confusion, we use $P(t)$ as a short notation of $P(X_1, \dots, X_n; t)$. Consider a coarse-grained model \mathcal{A}^\sharp , and any abstract state q_i^\sharp , let α_i denote the average rate of leaving state q_i^\sharp , i.e., $\frac{dP_i(t)}{dt} = -\alpha_i P_i(t)$, E_i denote the expected time to leave state q_i^\sharp , we have: $E_i = \int_0^\infty P_i(t) dt = \int_0^\infty e^{-\alpha_i t} dt = \frac{1}{\alpha_i}$, i.e., $\alpha_i = \frac{1}{E_i}$ is the rate of leaving q_i^\sharp . In addition, for any $j \neq i$ and $\Delta^\sharp(q_j^\sharp, q_i^\sharp) > 0$, similarly, $\beta_j = \frac{1}{E_j}$ is the rate of coming to q_i^\sharp from q_j^\sharp . Therefore, for all $i \in \{1, \dots, n\}$, we have:

$$\frac{dP_i(t)}{dt} = -\frac{1}{E_i} P_i(t) + \sum_{j=0, j \neq i, \Delta^\sharp(q_j^\sharp, q_i^\sharp) > 0}^n \frac{1}{E_j} P_j(t).$$

We therefore build a set of equations for all i . By solving the set of equations, we can obtain the distributions of the system at any future time.

Example 8. Consider again Example 6. Due to the coarse-grained model shown in Fig. 3 ($N_{\max} = 5$), we construct the master equations as follows:

$$\left\{ \begin{array}{l} \frac{dP_{\{a,b,c,d\}}(t)}{dt} = -\frac{1}{0.59} P_{\{a,b,c,d\}}(t) \\ \frac{dP_{\{a,b,c\}}(t)}{dt} = -\frac{1}{0.66} P_{\{a,b,c\}}(t) + 0.34 * \frac{1}{0.59} P_{\{a,b,c,d\}}(t) \\ \frac{dP_{\{a,b\}}(t)}{dt} = -\frac{1}{3.46} P_{\{a,b\}}(t) + 0.217 * \frac{1}{0.59} P_{\{a,b,c,d\}}(t) + 0.78 * \frac{1}{0.66} P_{\{a,b,c\}}(t) \\ \frac{dP_{\{a,d\}}(t)}{dt} = -\frac{1}{3.46} P_{\{a,d\}}(t) + 0.356 * \frac{1}{0.59} P_{\{a,b,c,d\}}(t) \\ \frac{dP_{\{a\}}(t)}{dt} = \frac{0.087}{0.59} P_{\{a,b,c,d\}}(t) + \frac{0.22}{0.66} P_{\{a,b,c\}}(t) + \frac{1}{3.46} P_{\{a,b\}}(t) + \frac{1}{3.46} P_{\{a,d\}}(t) \end{array} \right.$$

By solving the above equation systems, Fig. 4 presents a comparison between the time evolution of the reaction network via master equation simulation through the organisation-based average coarse-grained model (left) and the exact evolution of the system through the original concrete model (right).

Figures 11 and 12 (in Appendix) present experimental results for the case of $N_{\max} = 10$ with rates R1 and R2 respectively. Note that our prediction produces a similar pattern of the concrete behaviours with time evolution for this case. We focus on the average-based approximation here for the purpose of presenting and comparing the pattern of the system behaviours with time evolution more clearly and intuitively. Our further experiments also demonstrate that the interval-based prediction can safely approximate the concrete model. The precision of the results varies regarding to different models and rates of reaction rules, however the basic pattern of behaviours can be captured. Further note that the

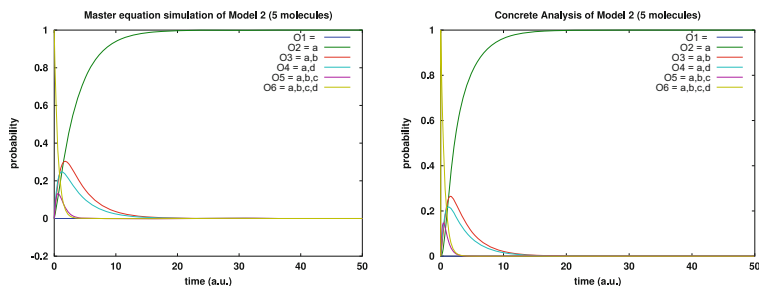


Fig. 4. Organisation dynamics prediction via the average coarse-grained model(left) and the concrete model(right) of Example 6, for $N_{max} = 5$. (Color figure online)

coarse-grained model can be used to predict qualitative dynamical properties of the original model, like the absence (or presence) of asymptotically stable attractors inside organizations that have small (or large) time to leave probabilities.

8 Conclusions

This paper investigates the combination of chemical organisation theory and probabilistic model checking for the analysis of reaction networks modelled as continuous-time Markov chains. We use model decompositions into strongly connected components (SCCs), and study the problem of how to analyse the model in terms of organisations. We have presented an algorithm to compute a coarse-grained Markov chain model of hierarchic organisations for a given reaction network. The algorithm computes chemical organisations by identifying a set of good SCCs which can contribute to generating organisations, and building an interval Markov chain based on the organisation-based quantitative analysis.

Experiments with our method on a set of example reaction network models demonstrate that the movements between organisations and the expected time spent in them can approximate the concrete long-term behaviour of the reaction network. The organisation-based coarse grained model helps to summarise and reason about the structure and behaviour of the complex model by focusing on stable states featuring accumulating species.

We also demonstrate how our model can be used to approximate the system behaviour with time evolution. The experiments show that our prediction can mimic the main pattern of concrete behaviour in the long run, but the interval-based organisation coarse graining may suffer from over-estimation. We apply an average-based organisation coarse graining and compute its stochastic time evolution. Our experiments show that the precision of the prediction and approximation varies regarding to different models and rates of their reaction rules. As future work, to improve the precision of the approximation and predictions, we plan to develop algorithms to selectively refine the coarse-grained models.

Acknowledgements. The authors acknowledge support from the European Union through funding under FP7-ICT-2011-8 project HIERATIC (316705). We also thank the anonymous reviewers for their helpful and detailed comments.

APPENDIX: Supplementary Details for Examples 5 and 6

```

ctmc

const int N_MAX = 10;
const double rA = N_MAX; // rA
const double rB = N_MAX; // rB
formula total = a + b;
init total <= N_MAX endinit

module RN
  a : [0..N_MAX]; // range value of species a
  b : [0..N_MAX]; // range value of species b
  c : [0..N_MAX]; // range value of species c

  // r1: a+b -> a
  [r1] (a*b > 0) & (a > 0) & (b > 0) & (total<= N_MAX) -> a*b : (a'=a-1) & (b'=b);

  // r2: a -> 2a
  [r2] (rA*a > 0) & (a > 0) & (total+1<= N_MAX) -> rA*a : (a'=a+1) ;

  // r3: b -> 2b
  [r3] (rB*b > 0) & (b > 0) & (total<= N_MAX) -> rB*b : (b'=b+1) ;

  // r4: a -> 0
  [r4] (a*a > 0) & (total<= N_MAX) -> a*a : (a'=a-1);

  // r5: b -> 0
  [r5] (a*b > 0) & (total<= N_MAX) -> a*b : (b'=b-1);

endmodule

```

Fig. 5. Example 5 in the PRISM modelling language

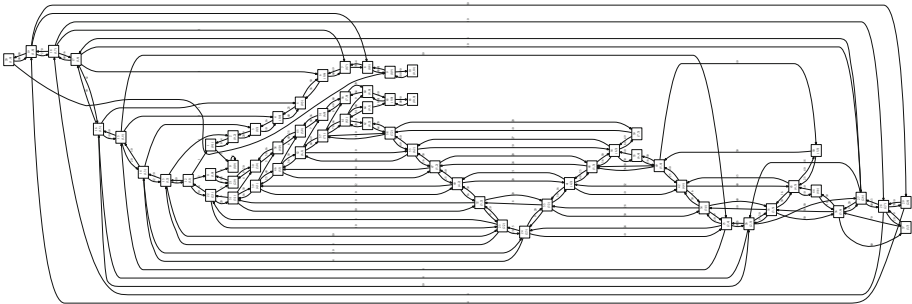


Fig. 6. Example 5: CTMC model with 4 SCCs and 1 BSCC

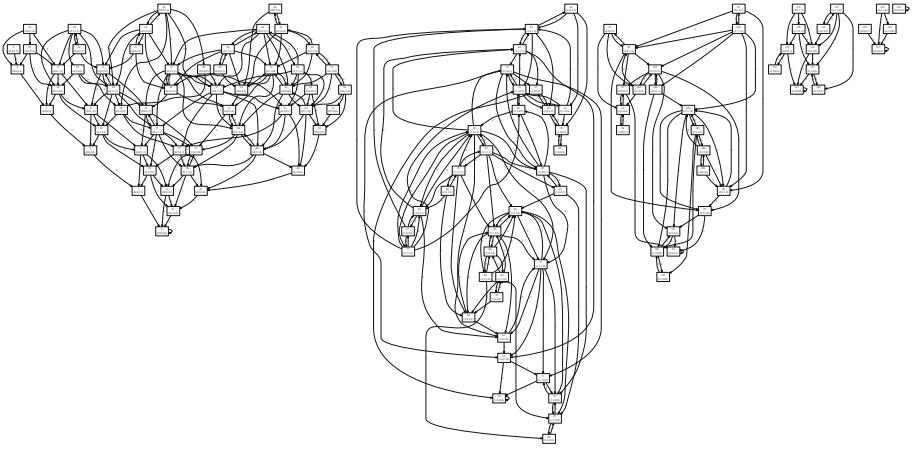


Fig. 7. CTMC for the reaction network from Example 6, with 28 SCCs and 6 BSCCs.

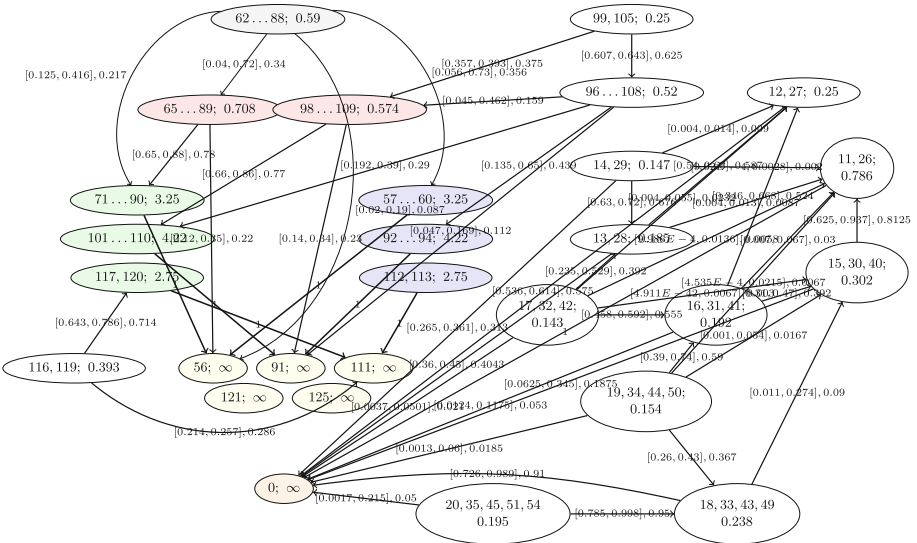


Fig. 8. Transition probabilities (bounds/averages) between all SCCs of the CTMC for Example 6 and expected leaving times.

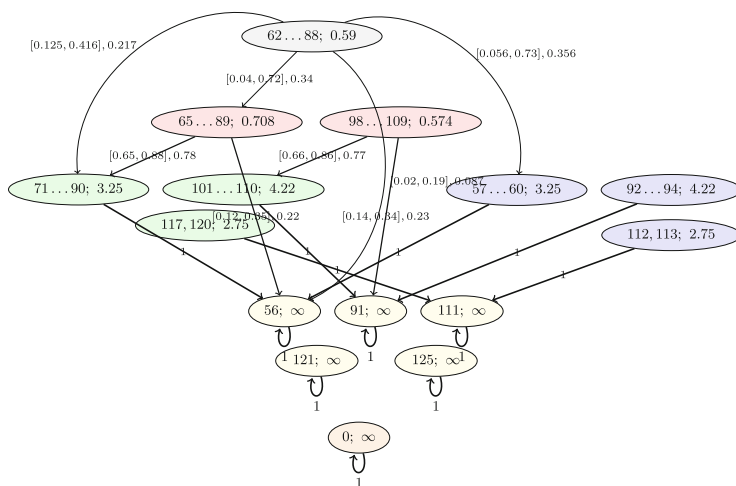


Fig. 9. Transition probabilities (bounds and average) between good SCCs for Example 6 and the expected time to leave them.

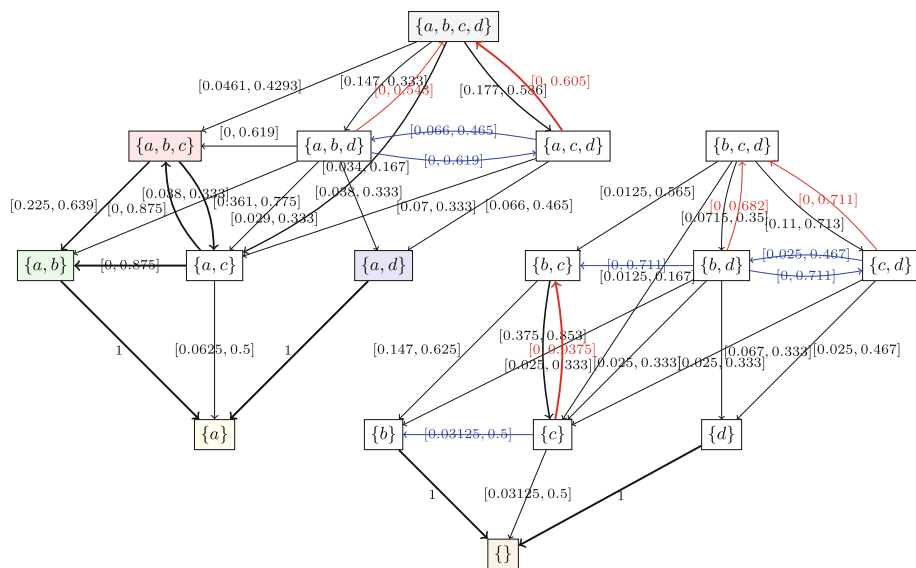


Fig. 10. Example 6: transition probabilities in bounds among the lattice of molecules

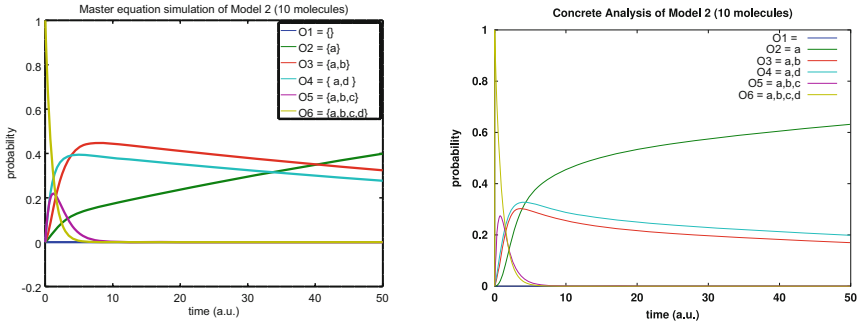


Fig. 11. Organisation dynamics predication via master equation simulation over the average coarse-grained model(left) and the original model(right) of Example 6, for $N_{\max} = 10$ with rates: $\#a * \#b$, $\#b * \#c$, $\#c$, $\#b * \#d$, $\#b * \#b$, $\#c$, $\#d$ for each reaction rule respectively. (Color figure online)

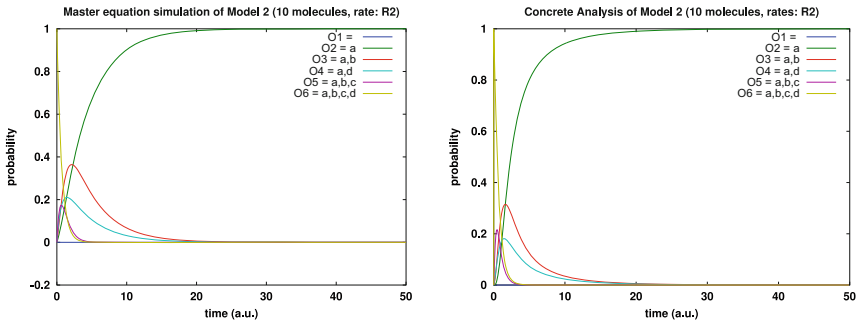


Fig. 12. Organisation dynamics predication via master equation simulation over the average coarse-grained model(left) and the original model(right) of Example 6, for $N_{\max} = 10$ with rates: $\#a * \#b$, $\#b * \#c$, $\#c * \#c$, $\#b * \#d$, $\#b * \#b$, $\#c * \#c$, $\#d * \#d$ for each reaction rule respectively. (Color figure online)

References

1. Aziz, A., Sanwal, K., Singhal, V., Brayton, R.: Verifying continuous time Markov chains. In: Alur, R., Henzinger, T.A. (eds.) CAV 1996. LNCS, vol. 1102, pp. 269–276. Springer, Heidelberg (1996)
2. Baier, C., Haverkort, B., Hermanns, H., Katoen, J.-P.: Model-checking algorithms for continuous-time Markov chains. IEEE TSE **29**(6), 524–541 (2003)
3. Bortolussi, L., Lanciani, R.: Model checking Markov population models by central limit approximation. In: Joshi, K., Siegle, M., Stoelinga, M., D’Argenio, P.R. (eds.) QEST 2013. LNCS, vol. 8054, pp. 123–138. Springer, Heidelberg (2013)
4. Cardelli, L., Kwiatkowska, M., Laurenti, L.: Stochastic analysis of chemical reaction networks using linear noise approximation. In: Roux, O., Bourdon, J. (eds.) CMSB 2015. LNCS, vol. 9308, pp. 64–76. Springer, Heidelberg (2015)
5. Clarke, E., Grumberg, O., Peled, D.: Model Checking. The MIT Press, Cambridge (2000)

6. Dittrich, P., Ziegler, J., Banzhaf, W.: Artificial chemistries-a review. *Artif. Life* **7**(3), 225–275 (2001)
7. Dittrich, P., di Fenizio, P.: Chemical organisation theory. *Bull. Math. Biol.* **69**(4), 1199–1231 (2007)
8. Feinberg, M., Horn, F.J.M.: Dynamics of open chemical systems and the algebraic structure of the underlying reaction network. *Chem. Eng. Sci.* **29**(3), 775–787 (1973)
9. Fontana, W.: Algorithmic chemistry. In: *Artificial Life II*. Addison Wesley (1992)
10. Fontana, W., Buss, L.W.: “The arrival of the fittest”: toward a theory of biological organization. *Bull. Math. Biol.* **56**(1), 1–64 (1994)
11. Gillespie, D.: Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **81**(25), 2340–2361 (1977)
12. Heinrich, R., Schuster, S.: *The Regulation of Cellular Systems*. Chapman and Hall, New York (1996)
13. Kleinberg, J., Tardos, É.: *Algorithm Design*. Addison-Wesley, Boston (2006)
14. Kreyssig, P., Wozar, C., Peter, S., Veloz, T., Ibrahim, B., Dittrich, P.: Effects of small particle numbers on long-term behaviour in discrete biochemical systems. *Bioinformatics* **30**(17), 475–481 (2014)
15. Kwiatkowska, M., Norman, G., Parker, D.: PRISM 4.0: verification of probabilistic real-time systems. In: Gopalakrishnan, G., Qadeer, S. (eds.) *CAV 2011*. LNCS, vol. 6806, pp. 585–591. Springer, Heidelberg (2011)
16. Tarjan, R.: Depth first search and linear graph algorithms. *SIAM J. Comput.* **1**(2), 146–160 (1972)
17. Wolf, V., Goel, R., Mateescu, M., Henzinger, T.: Solving the chemical master equation using sliding windows. *BMC Syst. Biol. J.* **4**, 42 (2010)