2013 European Control Conference (ECC)
July 17-19, 2013, Zürich, Switzerland.

Approximate model reductions for combinatorial reaction systems

Tatjana Petrov and Heinz Koeppl

Abstract— **The paper considers a model reduction technique that is well-suited for biochemical reaction systems giving rise to the assembly of a large number of different molecular species. The reduction is performed by grouping species with common properties, directly from the model specification in terms of a rule-based language. In recent works, general algorithms for the exact reductions of rule-based models were established, but the state space often remains combinatorial. We extend this line of research by introducing approximate reductions, and an error measure which allows us to quantitatively study the effect of approximate model reductions.**

I. Introduction

During induction of signalling pathways transient complex formation of proteins and their post-translational modification give rise to a combinatorial number of distinct molecular species. By requiring the enumeration of all reachable species, standard chemical kinetics is impractical to describe such combinatorial assembly processes and alternative modeling paradigms need to be sought. One such alternative are rule-based languages such as Kappa [1] or BioNetGen [2], [3]. In these languages proteins instead of the molecular species take center stage; their modification as well as their binding configuration is explicitly traced, instead of abstracting every new bound protein into a new species. The mathematical objects underlying such a description are site-graphs, i.e. a generalization of standard graphs where nodes carry sites from which the edges of the graph emerge. Moreover, every such site may be assigned an internal state assuming values in a discrete set. The latter states can, for instance, encode post-translational modifications of protein residues.

By exploiting the limited context on which most binding and modification events are conditioned, recently attempts have been made to reduce the complexity of such systems [4], [5], [6]. Specifically, in [5], partitioning of the set of reachable species that are implicitly encoded in a rule-based model into macrospecies is performed yielding a self-consistent ordinary differential equation that is exact with respect to the projection of the original species-based differential equation. A similar partitioning was given in [6], [7] that also preserves the stochastic dynamics given by the continuous-time Markov chain (CTMC) underlying chemical kinetics. More specifically, the partitioning was shown to be equivalent to the notion of weak lumpability [8] or backward Markovian bisimulation. Importantly, it was also shown that the probability distribution over the species-based states can be recovered from the probability distribution over the states of the reduced model [9].

In this paper we extend this line of research by proposing how to construct an approximate reduction and bound the error induced by simulating the reduced model. Among the various metrics for stochastic processes (see [10] for an overview), we consider the Kullback-Leibler divergence. The formalism is presented in a comprehensive mathematical description of rule-based models in terms of site-graphs.

The remaining part of the paper is organized as follows. In Section II the basics of stochastic chemical kinetics and its thermodynamic limit are introduced whereas Section II-A lays out the mathematical framework of site-graphs and their modifications in terms of site-graph rewrite rules. The specific model reduction problem that we consider is defined in Section III. The results are given in Section IV.

II. Preliminaries and Formalism

For a well-mixed reaction system with molecular species $S_1, \dots, S_s$ the state of a system can be represented as a multiset of species. We denote their multiplicities in the multiset as $\vec{x} = (x_1, \dots, x_s) \in \mathcal{X} \subseteq \mathbb{N}_{\geq 0}^s$. The dynamics of such a system is determined by a set of r reactions. The k-th reaction reads

$$\nu_{1k} S_1, \dots, \nu_{sk} S_{sk} \xrightarrow{c_k} \nu'_{1k} S_1, \dots, \nu'_{sk} S_s, \tag{1}$$

where $\nu_{ik} \in \mathbb{N}_{\geq 0}$ and $\nu'_{ik} \in \mathbb{N}_{\geq 0}$ denote the substrate and product stoichiometric coefficients of species i, respectively. If the k-th reaction occurs, after being in the state $\vec{x}$, the next state will be $\vec{x} + (\nu'_k - \nu_k) = \vec{x} + \mu_k$, where μ_k is referred to as the stoichiometric change vector. Under the above physical assumption the species multiplicities follow a continuous-time Markov chain and we denote the state of the system as the t-indexed random vector $X(t) = (X_1(t), ..., X_s(t))$. Hence, the probability of moving to the state $\vec{x} + \mu_k$ from $\vec{x}$ after time Δ is

$$\mathsf{P}(X(t+\Delta) = \vec{x} + \mu_k | X(t) = \vec{x}) = \lambda_k(\vec{x})\Delta + o(\Delta), \tag{2}$$

with λ_k the propensity of reaction k, the functional form of which is often assumed to follow the principle of mass-action: $\lambda_k(\vec{x}) = c_k \prod_{i=1}^{s} \binom{x_i}{\nu_{ik}}$.

A. Site-graphs

A molecular species can be a protein, its post-translationally modified form or a protein complex that consists of proteins bound together. In order to reflect this internal structure of molecular species we represent them by *site-graphs*. Let $\mathcal{I}$ denote the set of all internal values that can be assigned to sites. Such internal states encode post-translational modifications. The function $I : \mathcal{S} \to \wp(\mathcal{I})$ denotes the set of internal values that a site s can take. The evaluation function $\psi : V \times \mathcal{S} \to \mathcal{I}$ assigns a value to a

node-site combination. In particular, given a site $s \in \Sigma(v)$, $\psi(v,s) \in I(s)$.

Definition 1: A **site-graph** $G = (V, \Sigma, E, \psi)$ is defined by a set of nodes V, an interface function $\Sigma : V \to \wp(\mathcal{S})$, a set of edges $E \subseteq \{((v,s),(v',s'))|v,v' \in V, v \neq v', s \in \Sigma(v), s' \in \Sigma(v')\}$, and an evaluation function $\psi : V \times \mathcal{S} \to \mathcal{I}$ that assigns an internal value to a node-site combination.

The function Σ in the above definition tracks the sites corresponding to a particular node of a site-graph. A site s can be evaluated only to a predefined set of values, $I(s)$. The set of predefined internal values is empty for sites which serve for creating bonds.

The following notions of sub-site-graph and site-graph renaming will be used for the formal description of the dynamics of rule-based models. More specifically, we use them formalize the embedding relation between a pattern and a reaction mixture when defining the rule application.

Definition 2: Let $G = (V, \Sigma, E, \psi)$ and $G' = (V', \Sigma', E', \psi')$ be two site-graphs. Then, G is a **sub-site-graph** of G', written $G \subseteq G'$, if $E \subseteq E'$, and for every node $v \in V$ and site $s \in \Sigma(v)$, we have that $v \in V'$, $s \in \Sigma'(v)$, $\psi(s) = \psi'(s)$.

Definition 3: Let $G = (V, \Sigma, E, \psi)$ be a site-graph and $\eta : V \to V'$ a renaming function. Then the **η-induced node-renamed site-graph**, G^η, is given by $G^\eta = (\eta(V), \Sigma^\eta, E^\eta, \psi^\eta)$, where $\Sigma^\eta(\eta(v)) = \Sigma(v)$, $E^\eta = \{\{(\eta(v_1), s_1), (\eta(v_2), s_2)\} \mid v_1, v_2 \in V\}$, and $\psi^\eta(\eta(v)) = \psi(v)$.

For a given set of nodes, interface function, and a set of edges, varying the evaluation of internal sites and omitting edges gives raise to different connected site-graph structures, which we call *patterns*. Such a set of patterns will be used in the analysis of projections and reductions rule-based models in Section III.

Definition 4: Given a site-graph $G = (V, \Sigma, E, \psi)$, a sequence of edges $(e_1, \dots e_k) \in E^k$, $e_i = ((v_i, s_i), (v_i', s_i'))$, such that $v_i' = v_{i+1}$ for $i = 1, \dots k-1$, is called a *path* between nodes v_1 and v_k. If there exists a path between every two nodes $v, v' \in V$, a site-graph $G = (V, \Sigma, E, \psi)$ is *connected*.

Definition 5: A **pattern** over a set of nodes V, interface function Σ and set of edges $\hat{E}$ is every connected site-graph $G = (V', \Sigma', E', \psi')$ such that $\Sigma' = \Sigma$, $V' \subseteq V$, $E' \subseteq E$.

In the further text, the set of nodes will be clear from context (practically– referring to the same rule-based model). The set of all patterns over the interface function Σ is denoted by $\mathcal{F}_\Sigma$.

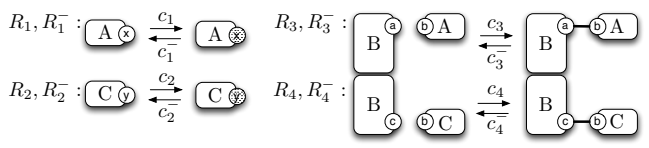


Fig. 1. An example of a rule-based model: $\hat{V} = \{A, B, C\}$, $\Sigma(A) = \{x, b\}$, $\Sigma(B) = \{a, c\}$, $\Sigma(C) = \{y, b\}$, $\mathcal{I}(x) = \mathcal{I}(y) = \{0, 1\}$, $\mathcal{I}(a) = \mathcal{I}(b) = \mathcal{I}(c) = \varnothing$, $\hat{E} = \{((A, b), (B, a)), ((C, b), (B, c))\}$.

B. Rule-based models

Definition 6: Let $G = (V, \Sigma, E, \psi)$ be a site-graph. We introduce three elementary site-graph transformations:

- adding an edge: $\delta_{ae}(G, e) = (V, \Sigma, E \cup \{e\}, \psi)$,
- deleting an edge: $\delta_{de}(G, e) = (V, \Sigma, E \setminus \{e\}, \psi)$,
- changing the internal state value $\delta_{ci}(G, v, s, i) = (V, \Sigma, E, \psi')$, such that $\psi'(v, s) = i$.

Note that the interface function Σ is unaltered under any of the transformations.

Let $c \in \mathbb{R}_{\geq 0}$ be a non-negative real number denoting the rate of the transformation. The triple (G, G', c), also denoted by $G \xrightarrow{c} G'$, is called a **site-graph-rewrite rule**, if $G' = (V', \Sigma, E', \psi')$ is a site-graph derived from $G = (V, \Sigma, E, \psi)$ by a finite number of applications of adding/deleting an edge and changing the internal state value: δ_{ae}, δ_{de}, δ_{ci}.

A **rule-based model** $\mathcal{R} \equiv \{R_i \equiv (G_i, G_i', c_i)_{i=1}^n\}$ is a collection of site-graph rewrite rules over the set of nodes $V \equiv \cup_i V_i$ total interface function Σ such that $\Sigma(v) = \cup_{i=1} \Sigma_i(v)$, and the set of edges $E \equiv \cup_i E_i$. The set of patterns $\mathcal{F}_\Sigma$ is exactly the set of molecular species appearing in a model, that is, $\mathcal{S} = \{S_1, \dots, S_s\} \equiv \mathcal{F}_\Sigma$.

For each node $v \in V$, we consider n_v copies or instances of the node v, denoted by $v^1, v^2, \dots, v^{n_v}$. A **reaction mixture** is a site-graph $\mathcal{G} = (\mathcal{V}, \hat{\Sigma}, \mathcal{E}, \psi)$, such that $\mathcal{V} = \{v^j \mid v \in V, j \in \{1, \dots, n_v\}\}$, $\hat{\Sigma}(v^j) = \Sigma(v)$, $\mathcal{E} \subseteq \{\{(v_1^i, s_1), (v_2^j, s_2)\} \mid \{(v_1, s_1), (v_2, s_2)\} \in \cup_i E_i\}$, and the evaluation function ψ is arbitrary.

In the further text, we are always referring to the model $\mathcal{R}$, and to a set of reaction mixtures $\mathbb{G}$. Moreover, we consider only the site-graph renamings induced by the injective functions from V to $\mathcal{V}$:

$$\Gamma := \{\eta \mid \eta : V \xrightarrow{1-1} \mathcal{V} \text{ and } \eta(v) \in \{v_1, \dots, v_{n_v}\}\}, \quad (3)$$

and we call them Γ-renamings. The notion of Γ-renaming formalizes the graph-embedding.

1) Stochastic semantics of rule-based models: We will now describe a Markov chain taking values in $\mathbb{G}$. Consider a reaction mixture $\mathcal{G} \in \mathbb{G}$, a rule $R_i = (G_i, G_i', c_i) \in \mathcal{R}$. Let $\eta \in \Gamma$ be such that $G_i^\eta \subseteq \mathcal{G}$. This implies that the rule R_i can be applied to a part of the reaction mixture $\mathcal{G}$. Let $\mathcal{G}'_{\eta,i}$ be the reaction mixture obtained after the application of the rule R_i. Note that $G^{\eta\prime} \subseteq \mathcal{G}'$. Let $\{\hat{X}_t\}$ be a CTMC with state-space $\mathbb{G}$, such that

$$\mathsf{P}(\hat{X}(t + \Delta) = \mathcal{G}'_{\eta,i} | \hat{X}(t) = \mathcal{G}) = c_i \Delta + o(\Delta). \quad (4)$$

Example 1: Scaffold protein B recruits independently the proteins A and C, and that the latter two are activated and deactivated spontaneously. These assumptions are captured by a set of rules depicted in Figure 1. The correponding set $\mathcal{F}_\Sigma$ contains 13 different patterns, which are exactly the set of molecular species formed by model $\mathcal{R}$. These are: five different species with no edges (two modifications of A, two modifications of C, and one B), four different species with one edge, and four different species with two edges.

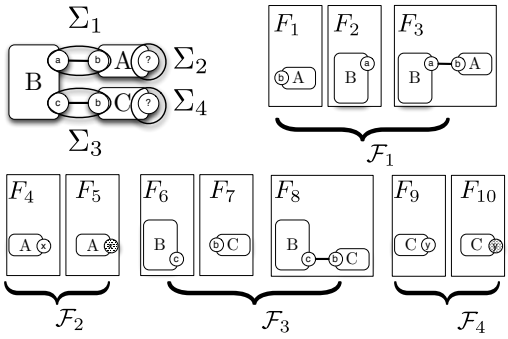
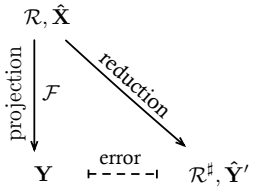


Fig. 2. Example 1: for a family of interfaces $\Sigma^1, \Sigma^2, \Sigma^3, \Sigma^4$, the set $\mathcal{F}$ contains 10 different patterns. The split of interfaces can be graphically represented by annotating the site-graph (V, Σ, E, ψ), where ψ function is not defined.

III. Problem statement

The CTMC assigned to a rule-based model can be reduced by projecting each reaction mixture to a multi-set of chosen patterns, $\mathcal{F}$. Some patterns are represented in Figure 2. The reduction occurs because many reaction mixtures project – in the sense to be formalized later – to the same state. We use the term *projection* for projecting the samples of $\hat{X} \in \mathbb{G}$ to a process $\hat{X}|_{\mathcal{F}} \in \mathcal{Y}_{\mathcal{F}} \subset \mathbb{N}_{\geq 0}^{|\mathcal{F}|}$. On the other hand, *reduction* stands for defining a process $Y_{\mathcal{F}}$ directly over the multisets of patterns. We call a reduction *exact*, if the processes $\hat{X}|_{\mathcal{F}}$ and $Y_{\mathcal{F}}$ are equivalent. Observe that $\hat{X}|_{\mathcal{F}}$ is not necessarily Markov, nor time-homogeneous, and it is therefore not always possible to define a corresponding CTMC over the projected state space, $\mathbb{N}_{\geq 0}^{M}$. In other words, for some choices of patterns, the reduction will necessarily be wrong. Conditions can be imposed on the generator matrix of the Markov chain $\hat{X}$ to ensure that the new process $\hat{X}|_{\mathcal{F}}$ is also Markov. The automatic procedure of how to construct the set of patterns which always give a sound reduction is given in [6]. Different to that, we discuss the following problem.

Problem 1: Given a rule-set $\mathcal{R}$, and a set of patterns $\mathcal{F}$,
(i) define a set of rules $\mathcal{R}^\sharp$ over patterns from $\mathcal{F}$, and
(ii) bound the error induced by simulating $\mathcal{R}^\sharp$.



The formal treatment of generating patterns and projecting the reaction mixture is provided in the Section IV.

IV. Results

We propose two ways of constructing the reduced chain, and we provide a bound on the error induced by simulating the reduced chains. The error is defined by means of information-theoretic cost function. We discuss both the error between transient probability distributions, as well as the error accumulated by simulating traces up to a given time limit. It is assumed that the generator matrix and the stationary distribution of the correct process are known. Finally, we provide detailed illustrations on an example.

Definition 7 (g-projection): Given a CTMC X over the alphabet $\mathcal{X} = \{1, \ldots, N\}$, $N \in \mathbb{N}$, and a partition function g: $\mathcal{X} \to \mathcal{Y}$, the g-projection of X is a process Y over the alphabet $\mathcal{Y} = \{1, \ldots, M\}$, $M < N$, whose samples are defined by projecting the samples of X through the function g: $Y_t = j$ iff $g(X_t) = j$. Given a state $j \in \mathcal{Y}$, we write $g^{-1}(j)$ for the preimage of j under g, that is, $g^{-1}(j) = \{i \in \mathcal{X} \mid g(i) = j\}$.

Given a set of patterns $\mathcal{F}$, our goal is to define a CTMC $Y_{\mathcal{F}}$ over the state space $\mathcal{Y} \subset \mathbb{N}_{\geq 0}^{M}$, such that it approximates well the corresponding projection, denoted by $\hat{X}|_{\mathcal{F}}$. Two reductions are proposed, termed *global* and *local*, depending on whether the stationary distribution of the original system is available (global) or not (local).

A. Local reductions

The idea behind a local reduction is to evaluate the transition rate between two abstract states $\vec{y}$ and $\vec{y}'$ with the assumption that all concrete states which are lumped to $\vec{y}$ are uniformly distributed. This reduction is preferable to the global one, in the sense that (i) it does not require knowing the stationary distribution of the original process, and (ii) it is 'executable', that is, possible to integrate in a rule-based modeling simulator by simply introducing dynamical reaction rates. However, we show by the experiments that the global reduction will generally provide better error bounds.

We first formalize a notion of a pattern.

Definition 8: Given a family of interface functions $\Sigma^1, \ldots, \Sigma^k$, such that given a node $v \in V$, $\Sigma(v) = \cup_{i=1}^{k} \Sigma^i(v)$, the set of patterns is defined by $\mathcal{F} = \mathcal{F}_{\Sigma_1} + \ldots + \mathcal{F}_{\Sigma_k} = \{F_1, \ldots, F_M\}$.

In particular, for $k = 1$, $\Sigma = \Sigma^1$, the set of patterns are called molecular species, and we denote them $\mathcal{F}_\Sigma = \mathcal{S}$. An example of inducing a set of patterns from a family of interface functions, accompanied by a graphical illustration is given in Figure 2.

Each reaction mixture can be projected to a multi-set of patterns from $\mathcal{F}$.

Definition 9: The projection function $\varphi_{\mathcal{F}} : \mathbb{G} \to \mathcal{Y}_{\mathcal{F}} \subset \mathbb{N}_{\geq 0}^{M}$ is defined by $\varphi_{\mathcal{F}}(\mathcal{G}) := (y_1, \ldots, y_M)$ iff $y_i = [\mathcal{G}, F_i]$, where the map $[\,] : \mathbb{G} \times \mathcal{F} \to \mathbb{N}_{\geq 0}$ denotes the number of sub-site-graphs of $\mathcal{G}$ which are an Γ-renaming of a pattern F:

$$[\mathcal{G}, F] := |\{\eta \in \Gamma \mid F^\eta \subseteq \mathcal{G}\}|. \tag{5}$$

The inverse transformation $\varphi^{-1} : \mathcal{Y} \to (\mathbb{G} \to [0, 1])$ maps a multiset of patterns to a probability distribution over the reaction mixtures projected to that pattern. With abuse of notation, let $\varphi^{-1}(\vec{y})$ also denote the set of reaction mixtures projected to $\vec{y}$: $\varphi^{-1}(\vec{y}) := \{\mathcal{G} \mid \varphi(\mathcal{G}) = \vec{y}\}$.

A φ-projection of $\hat{X}$ is also denoted by $\hat{X}|_{\mathcal{F}}$, taking values in the state space $\mathcal{Y}_{\mathcal{F}} \subset \mathbb{N}_{\geq 0}^{M}$. The projection $\varphi_{\mathcal{F}}$ acts as a partition function on the state space $\mathbb{G}$. The largest partitioning achieved by some $\mathcal{F}$-projection is for $k = 1$, $\Sigma^1 = \Sigma$.

Recall that the notion of Γ-renaming formalizes the graph-embedding. The number of different Γ-renamings of $\vec{y}$ to a subset of $\mathcal{G}$ is the number of embeddings of $\mathcal{G}$ to a multiset of patterns $\vec{y}$. The standard, species-based CTMC underlying the chemical reaction network is a projection to the set of patterns which coincide with the set of molecular species, that is, $\mathcal{Y}_{\mathcal{F}} = \mathcal{X}$. The effective dimension of the state space of any other $\mathcal{F}$-reduction is smaller or equal to the one of $\mathcal{X}$.

More concretely, the uniformity assumption sets that the conditional probability of being in a reaction mixture $\mathcal{G}$, given that the process is in any of the states $\varphi^{-1}(\vec{y})$, $\mathsf{P}(\hat{X} = \mathcal{G} \mid \hat{X} \in \varphi^{-1})$, equals to $|\varphi^{-1}(\vec{y})|^{-1}$. The concrete realization of the algorithm involves counting graph-embeddings, presented in [9]. We briefly explain how the distributions are derived on a concrete example in Section IV-E. The algorithm for an approximate $\mathcal{F}$-reduction is schemed as follows:

Input: A rule-based model $\mathcal{R} \equiv \{R_i \equiv (G_i, G_i', c_i)_{i=1}^n\}$ over the set of nodes V, interface function Σ and set of edges E; Set of patterns $\mathcal{F}$, such that $|\mathcal{F}| = M$; states $\vec{y}, \vec{y}' \in \mathcal{Y} \subset \mathbb{N}_{\geq 0}^M$;
Output: $f(\vec{y}, \vec{y}')$, such that $f(\vec{y}, \vec{y}') := E(\lambda(\mathcal{G}, \mathcal{G}') \mid \mathcal{G} \sim \mathcal{U}(\varphi_{\mathcal{F}}^{-1}(\vec{y})), \mathcal{G}' \sim \mathcal{U}(\varphi_{\mathcal{F}}(\vec{y}')))$.

B. Error measure: Relative entropy

If p and q are probability measures over a set S, and p is absolutely continuous with respect to q, then the relative entropy, or Kullback-Leibler divergence (KL divergence) from p to q, is defined as $D(p||q) = \int_S \ln\left(\frac{dp}{dq}\right) dp$, where $\frac{dp}{dq}$ is the Radon-Nikodym derivative of p with respect to q, and provided the expression on the right-hand side exists. The base of the logarithm is arbitrary.

The value of KL divergence is non-negative, it equals zero if the distributions match and it can be equal to infinity. It is not a metrics, since it is non-symmetric, and it does not satisfy the triangle inequality. A common technical interpretation is that KL divergence is the coding penalty associated with selecting the candidate q to approximate the correct distribution p.

The main reason why we employ KL-distance as a convenient error measure is that it has nice properties when applied to the probability space of traces generated by Markov sources: it can be computed as a function of solely the corresponding generator matrices and the transient distribution of the original process.

1) KL-distance over trace distributions – Discrete time: For a given $n \in \mathbb{N}$, denote by $X_{1:n}$(resp. $\tilde{X}_{1:n}$) the first n states of a discrete-time Markov chain (DTMC) with a transition matrix P (resp. $\tilde{P}$), over the state space $S = \{1, \ldots, \hat{s}\}$. Moreover, let $p_{1:n}(i_1, \ldots, i_n) = P(X_1 = i_1, \ldots, X_n = i_n)$ and $q_{1:n}(i_1, \ldots, i_n) = \tilde{P}(X_1 = i_1, \ldots, X_n = i_n)$, and assume that P is absolutely continuous wrt. $\tilde{P}$ (if $\tilde{P}(i,j) = 0$, then $P(i,j) = 0$), then $D(p_{1:n}||q_{1:n}) = \sum_{k=1}^{n-1} \sum_{i \in S} p_k(i)\theta(i) + \sum_{i \in S} p_1(i) \log \frac{p_1(i)}{q_1(i)}$, where $\theta(i) = \sum_{j \in S} P(i,j) \log \frac{P(i,j)}{\tilde{P}(i,j)}$.

Accordingly, the KL-divergence rate between two DTMCs is defined as a normalized asymptotic limit of the KL divergence. If X has a unique stationary distribution π, then $\bar{D}(p||q) = \lim_{n \to \infty} \frac{1}{n} D(p_{1:n}||q_{1:n}) = \sum_{i \in S} \pi(i)\theta(i)$, which can also be read as the expected value of θ under the distribution π.

2) KL-distance over trace distributions – Continuous time: For a given $T \in \mathbb{R}_{\geq 0}$, denote by $X_{[0,T)}$ (resp. $\tilde{X}_{[0,T)}$) the continuous random variable representing a trace of a CTMC X (resp. $\tilde{X}$) up to time T, with a generator Q (resp. $\tilde{Q}$), over the state space $S = \{1, \ldots, \hat{s}\}$, and let $\boldsymbol{\tau}_{[0,T)}$ and $\tilde{\boldsymbol{\tau}}_{[0,T)}$ denote the corresponding probability density functions. Assume that Q is absolutely continuous wrt. $\tilde{Q}$. The KL divergence between the distributions $\boldsymbol{\tau}_{[0,T)}$ and $\tilde{\boldsymbol{\tau}}_{[0,T)}$ is given by ([11])

$$D(\boldsymbol{\tau}_{[0,T]}||\tilde{\boldsymbol{\tau}}_{[0,T]}) = \int_0^T \sum_i p_t(i)\sigma(i)dt, \tag{6}$$

where $p_t(i) = \mathsf{P}(Y_t = i)$, and

$$\sigma(i) = \sum_{j \neq i} \left[Q(i,j) \log \frac{Q(i,j)}{\tilde{Q}i,j} + Q(i,j) - \tilde{Q}(i,j) \right]. \tag{7}$$

C. Error bounds

The error measure is defined in a general framework of Markov chain aggregation. Let X (resp. Y') be an irreducible, non-explosive CTMC, over the alphabet $\mathcal{X} = \{1, \ldots, N\}$ (resp. $\mathcal{Y} = \{1, \ldots, M\}$, $M < N$), with a generator matrix Q (resp. $\tilde{Q}'$), transient distribution p_t (resp. $\tilde{p}_t'$), a stationary distribution μ (resp. ν') and traces distribution $\boldsymbol{\tau}$ (resp. $\tilde{\tau}'$). The smaller process Y' is a reduction of X. Moreover, consider a partition function g: $\mathcal{X} \to \mathcal{Y}$, and denote the g-projection of X by Y, its transient distribution by $\tilde{p}_t$, a stationary distribution by ν and traces distribution by $\tilde{\tau}$ (the generator of Y does not necessarily exist).

Definition 10: (Error measure) The *transient aggregation error* of reduction Y' is defined by

$$\Delta_{X,Y'}(t) := D(\tilde{p}_t||\tilde{p}_t'), \tag{8}$$

and the *error rate* of reduction Y'

$$\bar{\Delta}_{X,Y'}(t) := \frac{d}{dt} D(\tilde{\boldsymbol{\tau}}_{[0,t]}||\tilde{\boldsymbol{\tau}}'_{[0,t]}). \tag{9}$$

Note that evaluating the transient aggregation error $\Delta_{X,Y'}$ is possible directly between Y and Y', because the distribution $\tilde{p}$ is a linear transformation of p, induced by lumping those states which are aggregated by partition g. More concretely, let $\mathbf{V}$ be an $N \times M$ matrix such that

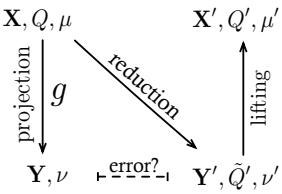


Fig. 3. The general framework: an upper bound on the error rate between Y and Y' is computed by lifting the process Y' to the state space of the original process.

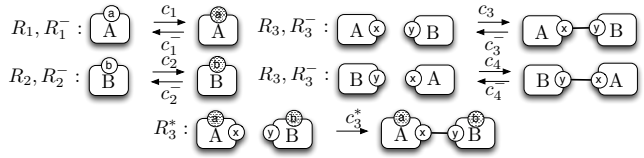


Fig. 4. Example 2: $V = \{A, B\}$, $\Sigma(A) = \{a, x, y\}$, $\Sigma(B) = \{b, x, y\}$, $\mathcal{I}(x) = \mathcal{I}(y) = \varnothing$, $E = \{((A, x), (B, y)), ((A, y), (B, x))\}$.

$V_{i,j} = \begin{cases} 1 & \text{if } g(i) = j, \\ 0 & \text{otherwise.} \end{cases}$ Then, $\Delta_{X,Y'}(t) = D(p_t\mathbf{V}\|\tilde{p}'_t)$. On the other hand, a similar direct evaluation of the error rate is not possible, since the equation (6) is applicable only if the process Y is also a CTMC. Inspired by the work of [12], we propose to lift the process Y' to the original state space, and to evaluate the KL-divergence of the lifted process wrt. the original one.

Let $\mathbf{U}^\pi$ be an $M \times N$ matrix, such that, for some probability vector π over $\mathcal{X}$, $U^\pi_{ij} = \begin{cases} \frac{\pi_j}{\sum_{k \in g^{-1}(i)} \pi_k} & \text{if } g^{-1}(i) = j, \\ 0 & \text{otherwise.} \end{cases}$ If the superscript is omitted, we let π be arbitrary.

Definition 11 (π-lifting): The π-lifting of Y' wrt. g, denoted by X'^π is a process over the alphabet $\mathcal{X}$, defined by the transition matrix $\mathbf{Q}' = \mathbf{V}\mathbf{Q}\mathbf{U}^\pi$, which is a matrix notation for $Q'_{ij} = \frac{\pi_j}{\sum_{k \in g^{-1}(g(j))} \pi_k} Q_{g(i)g(j)}, \ \ i, j \in \mathcal{X}$.

Clearly, X'^π permits an exact reduction w.r.t. g (and Y' is the resulting g-projection). Estimation of the transient distribution of X' can be done by simulating Y', and inverting the transient by $p' = \tilde{p}'\mathbf{U}^\pi$. The induced error we call the *transient de-aggregation error*: $\hat{\Delta}^\pi_{X,X'}(t) := D(p_t\|p'_t) = D(p_t\|\tilde{p}'_t\mathbf{U}^\pi)$. By using general result that the relative entropy is reduced under projection [13], the π-lifting ensures that the error rate between the processes in the projected state space is dominated by the error rate between the processes in the original state space. The case of discrete-time is discussed in more detail in [14]. In continuous time, the inequality is

$$\int_0^T \bar{\Delta}_{X,Y'}(t)dt \leq D(\boldsymbol{\tau}_{[0,T]}\|\boldsymbol{\tau}'_{[0,T]}).$$

D. Global reduction

The idea behind global reduction is to use the knowledge about the stationary distribution (invariant measure) ν of the CTMC X (thus, also Y). The generator of the reduction is defined by $\tilde{\mathbf{Q}}' = \mathbf{U}^\nu\mathbf{Q}\mathbf{V}$. The experiments show that the global reduction gives better bounds than the local reduction.

E. Experiments

Example 2: A rule-based model $\mathcal{R}_1 \equiv \{R_1, R_2, R_3, R^*_3, R_4, R^-_1, R^-_2, R^-_3, R^-_4\}$ (Figure 4) describes an alternating polymerization between nodes A and B. The rules R_3, R^-_3, R_4 and R^-_4 describe the binding and unbinding events. Moreover, each node has two activation levels, modelled by an internal state - a for node A and b for node B, which are regulated independently of the bindings (rules R_1, R^-_1, R_2, R^-_2). We assume that the binding between the nodes can be accelerated if both nodes are in the active mode. This is incorporated by rule R^*_3. The total rate of binding between an activated A and an activated B is $c_3 + c^*_3$.

The species which occur in a reaction mixture can be categorized into two types: chains, that have two free sites, and rings, that have no free sites. The total number of species depends on the total copy number of nodes, n_A and n_B, and for $n_A = n_B = n$, it becomes exponential in n.

We discuss three different reductions of this model, induced by the interface splits depicted in Figure 5, including the trivial reduction to the set of species. The estimation of the reduction of the dimension and size of the corresponding state spaces is given in Table 6. The transient transient aggregation and transient de-aggregation error are computed by numerically solving the Master equation for the set of patterns $\mathcal{S}$, $\mathcal{F}_1$ and $\mathcal{F}_2$. We restrict ourselves to the case when $c_4 = 0$, when no real polymerization is happening. This is because the binding specified by rule R_4 happens independently of the other reactions –it brings an additional dimension to the system, but it has no influence on the error which may occur between the approximate and correct dynamics. The results related to exact reductions are summarized in Figure 7 and Table 6.

Let A^1B^1 denote a species, which is a complex between nodes A and B where both nodes have an active internal state - a, resp. b, and denote the population of that species in the $\mathcal{S}$-projected state by x_{11}. Moreover, let A^1B, A^0B and AB^1 denote the patterns in $\mathcal{F}_2$, and y_{1-}, y_{0-}, y_{-1} respectively these patterns' population components. In $\mathcal{F}_2$-reduction, the value of x_{11} is approximated under the assumption listed in the algorithm scheme in Section IV to a value $\frac{y_{1-}y_{-1}}{y_{0-}+y_{1-}}$. In Figure 7a), the transient aggregation error is plotted as the KL distance between the corresponding distributions.

In Figure 8, we compare the error rate bounds provided for using local and global reduction. Global reduction shows a better estimate, but this advantage is less apparent when the volume increases. We can observe the trade-off between the size of the reduced state space and the error bound: for only 12 molecules reacting, the reduction $\mathcal{F}_2$ decreases the species-based state space from 1386 to 1092 states, while the reduction $\mathcal{F}_3$ decreases the species-based state space from 1386 to 343 states. As expected, the better reduction induces a bigger error.

V. Conclusion and fture work

The importance of executing models at different levels of detail is particularly notable when modeling protein interac-

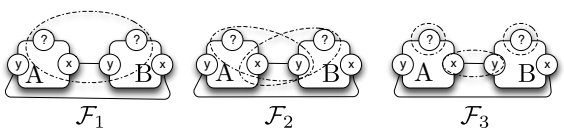


Fig. 5. Three different sets of patterns are induced by different spliting the interfaces of nodes A and B.

	$\mathcal{S}$	$\mathcal{F}_1$	$\mathcal{F}_2$
dim. of st. sp.	$O(4^n)$	7	6
size of st. sp.	$O(4^n)$	$O(n^7)$	$O(n^6)$
transient aggregation error			
$c_3^* = 0$, $\Omega = 10$	0	0	0
$c_3^* = c_3$, $\Omega = 10$	0	0([6])	Fig.7
$c_3^* = c_3$, $\Omega \to \infty$	0	0([6])	0([5])
transient de-aggregation error			
$c_3 = 0, \Omega = 10$	0	Fig.7	–

Fig. 6. A summary of properties of reductions.

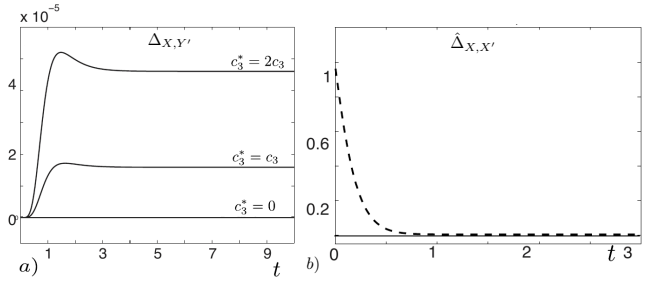


Fig. 7. Errors under $\mathcal{F}_2$-reduction. a) transient aggregation error for $\frac{c_3+c_3^*}{c_3} = 1, 2, 3$. By using a $\mathcal{F}_2$-reduction, under $c_3^* \neq 0$, the error occurs because the population of the species A^1B^1 is approximated under the assumption in Alg. **??**. b) transient de-aggregation error for correct and wrong initial distribution. The error decay happens due to the reversibility of rules in the rule-set, [7].

tions, which suffer the astronomical state space, but where the phenomena to be studied is combinatorial in itself (for example, studying phenotypic variety as in [15]). With this paper, we opened a perspective for studying quantitatively coarse-grained executions of a rule-based model, even when such executions are not consistent with the original model. More concretely, we proposed two constructions of approximate reductions of rule-based models, and we provided a bound on the error. The error is defined by means of a KL divergence. The presented approach relies on only the knowledge of the generator matrix of the original process, and it considers simulating exclusively the reduced system.

The proposed 'global' reduction technique is not limited to the application of rule-based models, and can be applied to any continuous-time Markov chain model. 'Local' reductions are specific to site-graph-rewrite grammars.

The implementation of the approximate reductions framework within the Kappa modeling framework is a work in progress. For that reason, we leave the analysis of large-scale case studies to future work. Another related problem to be considered is finding an optimal reduction for a given range of admissible sizes of the state space.

In a wider context of biological modeling, it is worth emphasizing that the here-presented work deals with providing approximate, but more efficient executions of a given rule-based model (taken as the 'ground truth'), while we do not address the problem of constructing or validating that model with respect to experimental data. Still, as biochemical models are already an approximation of reality, models obtained by approximate reductions can also be seen as

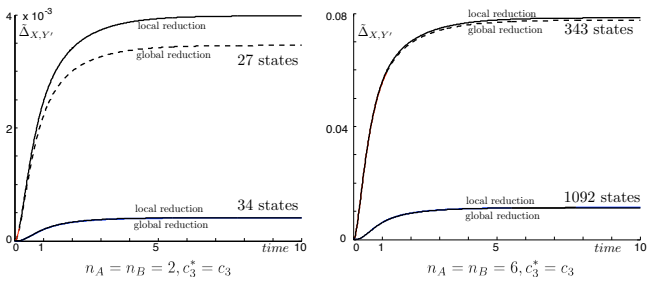


Fig. 8. Comparison of the error over the trace distributions, for a set of patterns $\mathcal{F}_2$ and $\mathcal{F}_3$, $\frac{c_3+c_3^*}{c_3} = 2$.

alternative model candidates which operate on a context that is more 'local' than some reference model; It would be interesting to investigate how the proposed error measure can serve to discriminate between such a set of candidate models.

REFERENCES

[1] V. Danos, J. Feret, W. Fontana, R. Harmer, and J. Krivine, "Rule-based modelling, symmetries, refinements," in *Proceedings of Formal Methods in Systems Biology (FMSB'08)*, LNBI, 2008.

[2] M. L. Blinov, J. R. Faeder, and W. S. Hlavacek, "BioNetGen: software for rule-based modeling of signal transduction based on the interactions of molecular domains," *Bioinformatics*, vol. 20, pp. 3289–3292, 2004.

[3] M. Sneddon, J. R. Feader, and T. Emonet, "Efficient modeling, simulation and coarse-graining of biological complexity with NFsim," *Nature Methods*, vol. 8, pp. 177–183, April 2011.

[4] N. Borisov, N. Markevich, B.N.Kholodenko, and E. Gilles, "Signaling through receptors and scaffolds: Independent interactions reduce combinatorial complexity," *Biophysical Journal*, vol. 89, 2005.

[5] 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, pp. 6453–6458, April 2009.

[6] J. Feret, T. Henzinger, H. Koeppl, and T. Petrov, "Lumpability abstractions of rule-based systems," *Theoretical Computer Science*, 2012.

[7] A. Ganguly, T. Petrov, and H. Koeppl, "Markov chain aggregation and its applications to combinatorial reaction networks." BISON group, ETH Zurich, in preparation.

[8] J. G. Kemeny, J. L. Snell, and A. W. Knapp, *Denumerable Markov Chains*. New York, NY, USA: Springer-Verlag, 1976.

[9] T. Petrov, J. Feret, and H. Koeppl, "Reconstructing Species-Based Dynamics from Reduced Stochastic Rule-Based Models.," in *WSC - Winter Simulation Conference - 2012* (C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, and A. M. Uhrmacher, eds.), (Berlin, Allemagne), Dec. 2012.

[10] A. Gibbs and F. Su, "On choosing and bounding probability metrics," *International Statistical Review (2002)*, vol. 70, no. 3, pp. 419–435, 2002.

[11] I. Cohn, T. El-Hay, N. Friedman, and R. Kupferman, "Mean field variational approximation for continuous-time bayesian networks," in *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, UAI '09, (Arlington, Virginia, United States), pp. 91–100, AUAI Press, 2009.

[12] K. Deng, P. G. Mehta, and S. P. Meyn, "Optimal Kullback-Leibler aggregation via spectral theory of Markov chains," vol. 56, pp. 2793–2808, Dec. 2011.

[13] R. M. Gray, *Entropy and information theory*. New York, NY, USA: Springer-Verlag New York, Inc., 1990.

[14] B. C. Geiger, T. Petrov, G. Kubin, and H. Koeppl, "Optimal kullback-leibler aggregation via information bottleneck," Apr. 2013. in preparation; preprint available: `arXiv:1304.6603 [cs.SY]`.

[15] J. Deeds, Eric, J. Krivine, J. Feret, V. Danos, and W. Fontana, "Combinatorial complexity and compositional drift in protein interaction networks.," *PLoS ONE*, vol. 7, no. 3, 2012.