Representing Reaction Systems by Trees

R. Brijder[1], A. Ehrenfeucht[2], and G. Rozenberg[2,3]

[1] Hasselt University and Transnational University of Limburg, Belgium
[2] Department of Computer Science,
University of Colorado at Boulder, USA
[3] Leiden Institute of Advanced Computer Science,
Leiden Center for Natural Computing,
Leiden University, The Netherlands

Abstract. Reaction systems formally model the functioning of the living cell. By representing sets of reactions by trees, we obtain a useful tool to investigate the state spaces of reaction systems. In particular, we give an upper bound on the fraction of inactive states within a subspace of the state space. This subspace represents partial knowledge of the (unknown) state under consideration.

1 Introduction

Reaction systems (see, e.g., [6] and [1]) are a formal model of the functioning of the living cell based on the idea/intuition that this functioning is determined by interactions of biochemical reactions (taking place in the cell) and these interactions are driven by two mechanisms: facilitation and inhibition.

Following the basic biochemical intuition, a *reaction* is formalized as a triplet $a = (R, I, P)$, where R, I, P are finite sets with R and I being disjoint. The sets R, I, P are called the set of *reactants*, the set of *inhibitors*, and the set of *products*, respectively. Then a *reaction system* is defined as an ordered pair $\mathcal{A} = (S, A)$, where A is a finite set of reactions, and S is a finite set such that, for each reaction in A, all three component sets are included in S. Hence a reaction system ($\mathcal{A}$) is basically a finite set of reactions (A) — we also specify the *background set* (S) which consists of *entities* needed to define the reactions and for reasoning about the system.

The behaviour of a reaction system $\mathcal{A} = (S, A)$ is formalized as follows. A state T of $\mathcal{A}$ is simply a set of entities, i.e., $T \subseteq S$. Then a reaction $a = (R, I, P) \in A$ is *enabled* by T, if all reactants of a are present in T (hence $R \subseteq T$) and none of the inhibitors of a is present in T (hence $I \cap T = \varnothing$). If a is enabled by T, then it produces its products (hence P will be included in the successor state of T). The effect of the whole set of reactions A on T (hence the effect of $\mathcal{A}$ on T) is cumulative: it is the union of the product sets of all reactions in A that are enabled by T.

Thus the behaviour of $\mathcal{A}$ is defined by its *state space* (the set of all subsets of S) together with all trajectories, i.e., all sequences of states such that each next

(successor) state is produced from a current state T by all reactions of $\mathcal{A}$ enabled by T.

Research topics concerning reaction systems are motivated either by biological considerations or by the need to understand the underlying computations. As a matter of fact, although originally motivated/inspired by the functioning of the living cell, by now reaction systems became a novel, elegant and challenging model of computation. Examples of research topics include: the studies of result functions that determine the trajectories/processes (see, e.g., [3] and [4]), causalities between entities ([2]), formation of (biological and biochemical) modules ([5]), and the issue of time in reaction systems ([7]).

In this paper we consider a representation of (the sets of reactions of) reaction systems. The representation we provide allows one to reason about the state spaces of reaction systems. The underlying intuition of this connection is the fact that the current state of a biochemical system (the cell) is often unknown, and one may only determine the existence and absence of some entities. Given sets U and V for which $U \subseteq W$ and $V \cap W = \varnothing$ for some unknown state W, we deduce an upper bound on the fraction of the states X, satisfying $U \subseteq X$ and $V \cap X = \varnothing$, for which no reaction is enabled. We efficiently obtain this result by representing sets of reactions as trees.

This paper is organized as follows. In Section 2 we settle/recall the basic notation and terminology concerning set families, and (labelled) graphs and trees. In Section 3 we discuss a representation of families of (pairwise incomparable) sets by trees, and then show that each such tree may be "optimally selected". The setup is generic, and does not depend on the notion of reaction system. Next, in Section 4, we consider (generic) states and substates in relation to trees. In Section 5 we formally recall the notion of reaction system and related notions, and in Section 6 we apply the results of Sections 3 and 4 to reaction systems.

2   Preliminaries

In this section we recall some basic notions concerning sets, graphs and trees in order to fix notation and terminology for this paper.

Two sets X and Y are called *incomparable* if both $X \nsubseteq Y$ and $Y \nsubseteq X$. Let $\mathcal{F}$ be a family of subsets of a finite set F. A *selector* (or *choice function*) of $\mathcal{F}$ is a function $c : \mathcal{F} \rightarrow F$, where $c(X) \in X$ for all $X \in \mathcal{F}$. We say that $c(\mathcal{F})$ is a *selection in $\mathcal{F}$*. Note that if S is a selection, then $S \cap X \neq \varnothing$ for all $X \in \mathcal{F}$. The term "smallest" means minimal w.r.t. cardinality. For example, a smallest selection S in $\mathcal{F}$ is a selection S in $\mathcal{F}$ which is minimal w.r.t. cardinality among all selections in $\mathcal{F}$. Since F is finite, a smallest selection exists.

A *directed graph* (*digraph*) is an ordered pair $G = (V, E)$, where V is a finite set of *vertices*, and $E \subseteq V \times V$ is the set of (directed) *edges*. A *labelled digraph G* is a 4-tuple (V, E, Σ, l), where (V, E) is a digraph, Σ is a finite alphabet (of *labels*), and $l : E \rightarrow \Sigma$ an *edge labelling*. A *path* in G is a sequence $\pi = e_1 e_2 \cdots e_n$ of edges of G such that there is a (unique) sequence of vertices $v_1 v_2 \cdots v_{n+1}$ with

$e_i = (v_i, v_{i+1})$ for all $i \in \{1, \ldots, n\}$. The *label set* of path π, denoted by $\mathrm{ls}(\pi)$, is the set $\{l(e) \mid e$ is an edge of $\pi\}$. The out-degree of a vertex v is denoted by $\deg(v)$.

A *tree* $T = (V, E)$ is a digraph, where $|E| = |V| - 1$, with a unique vertex $r \in V$, called the *root* of T, such that there is a (unique) path from r to any vertex of T. In this paper we consider mostly labelled trees. For $v \in V$, the subtree of T rooted in v is denoted by $T[v]$, and we let $\mathrm{hgt}(v)$ be the height (i.e., the maximal length among the paths from the root to a leaf) of $T[v]$. Finally, the set of leaves of T is denoted by $\mathrm{leav}(T)$.

3   Representing families of sets as trees

In this section we discuss how to represent families of sets by (unambiguously labelled) trees.

For a vertex v of a labelled tree T, we define the *support* of v (in T), denoted by $\sup_T(v)$, as the set of labels that appear in the (unique) path from the root to v. This is more formally defined as follows.

Definition 1. *Let $T = (V, E, \Sigma, l)$ be a labelled tree. The support function, $\sup_T$, is defined by:* $\sup_T : V \to 2^{\Sigma}$*, with* $\sup_T(v) = \mathrm{ls}(\pi_v)$*, for all* $v \in V$*, where π_v is the (unique) path in T from the root r to v.*

We write $\sup$ rather than $\sup_T$ whenever T is clear from the context. We say that T is *unambiguously labelled* if $\sup$ is injective, and we say that T is *ambiguously labelled* otherwise. Also, we set $\mathrm{supl}_T = \{\sup_T(v) \mid v \in \mathrm{leav}(T)\}$.
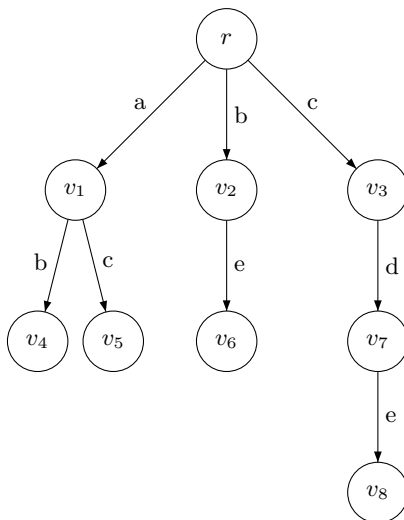

Fig. 1. An unambiguously labelled tree.

Example 2. Consider the tree T of Figure 1. We have, e.g., $\sup_T(v_2) = \{b\}$, $\sup_T(v_8) = \{c, d, e\}$, and $\sup_T(r) = \varnothing$. It is easy to verify that T is unambiguously labelled.

The following result states two basic properties of unambiguously labelled trees.

Lemma 3. *Let $T = (V, E, l)$ be a unambiguously labelled tree. The following conditions hold.*

1. *If π is a path in T, then the labels of any two distinct edges of π are distinct.*
2. *If e_1 and e_2 are two outgoing edges of any $v \in V$, then the labels of e_1 and e_2 are distinct.*

Proof. Assume to the contrary that a path $\pi = e_1 \ldots e_n$ is such that $l(e_i) = l(e_j)$ for some $1 \le i < j \le n$. If $e_j = (v_1, v_2)$, then $\sup(v_2) = \sup(v_1) \cup \{l(e_j)\} = \sup(v_1) \cup \{l(e_i)\} = \sup(v_1)$ (as $l(e_i) \in \sup(v_1)$). Hence T is ambiguously labelled — a contradiction.

Let $e_1 = (v, v_1)$ and $e_2 = (v, v_2)$ be outgoing edges of some $v \in V$. If $l(e_1) = l(e_2)$, then $\sup(v_1) = \sup(v) \cup \{l(e_1)\} = \sup(v) \cup \{l(e_2)\} = \sup(v_2)$ and therefore T is ambiguously labelled. A contradiction.   □

From now on we consider only unambiguously labelled tree and use the simple term "*tree*" rather than "unambiguously labelled tree". Also, we often simply write V to denote the vertex set of the tree under consideration.

For a vertex v, we let $O(v)$ to denote the set of labels of edges outgoing from v. Note that as T is unambiguously labelled, Condition 2 of Lemma 3 implies that $|O(v)| = \deg(v)$ for all $v \in V$.

For $v \in V$, we let $\mathcal{F}_T[v] = \mathrm{supl}_{T[v]}$, i.e., $\mathcal{F}_T[v]$ is the family of label sets of all paths from v to leaves of T. For the root r, $\mathcal{F}_T[r] = \mathrm{supl}_T$, and $\mathcal{F}_T[v] = \{\varnothing\}$ iff v is a leaf. Note that $O(v)$ is a selection in $\mathcal{F}_T[v]$. Alternatively, $\mathcal{F}_T[v]$ can be defined recursively. Indeed, if v is a leaf, then $\mathcal{F}_T[v] = \{\varnothing\}$, and if v is not a leaf, then $\mathcal{F}_T[v] = \{Z \cup \{l(e)\} \mid Z \in \mathcal{F}_T[v'] \text{ and } e = (v, v') \in E\}$.

Example 4. Consider again Example 2. We have, e.g., $O(r) = \{a, b, c\}$. Also, e.g., $\mathcal{F}_T[v_1] = \{\{b\}, \{c\}\}$ and $\mathcal{F}_T[r] = \mathrm{supl}_T = \{\{a, b\}, \{a, c\}, \{b, e\}, \{c, d, e\}\}$.

Thus, trees can be used to define families of sets: we say that a tree T *represents* a family $\mathcal{F}$ of sets if $\mathcal{F} = \mathcal{F}_T[r]$.

Definition 5. *Let $\mathcal{F}$ be a finite family of sets. A tree T is called* optimally selected *for $\mathcal{F}$ if T represents $\mathcal{F}$ and, for each $v \in V$, $O(v)$ is a smallest selection in $\mathcal{F}_T[v]$.*

Note that the out-degree of the root of an optimally selected tree T for $\mathcal{F}$ is minimal among all trees representing $\mathcal{F}$. Hence, from this point of view, a tree is optimally selected if the out-degree of each vertex is minimized using a greedy minimization approach starting from the root vertex. This is more precisely demonstrated in the proof of Theorem 7.
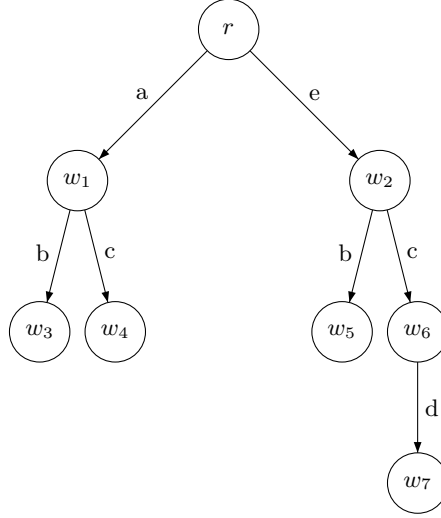
Fig. 2. An optimally selected tree, cf. Example 6.

Example 6. Consider again the tree T of Figure 1. Note that T is not optimally selected as $O(r) = \{a, b, c\}$, while $\{a, e\}$ and $\{b, c\}$ are (the smallest) selections in $\mathcal{F}_T[r]$. Let $\mathcal{G} = \mathcal{F}_T[r]$. An optimally selected tree for $\mathcal{G}$ is given in Figure 2.

Let $\mathcal{F}$ be a family of pairwise incomparable sets (hence for any distinct $X, Y \in \mathcal{F}$ we have $X \not\subseteq Y$ and $Y \not\subseteq X$). Theorem 7 shows that one can iteratively construct an optimally selected tree for $\mathcal{F}$. This is done by starting from the root and in each step introducing all outgoing edges and vertices from a vertex v according to a smallest selection in $\mathcal{F}_T[v]$.

Theorem 7. *Let $\mathcal{F}$ be a finite family of pairwise incomparable sets. There exists an optimally selected tree for $\mathcal{F}$.*

Proof. Assume that $\mathcal{F} \subseteq 2^Q$, i.e., the sets in $\mathcal{F}$ are subsets of a ground set Q. We recursively construct an optimally selected tree $k(\mathcal{F})$ for $\mathcal{F}$.

If $\mathcal{F} = \{\varnothing\}$, then we define $k(\mathcal{F})$ to be a tree having only a single vertex — clearly $k(\mathcal{F})$ is an optimally selected tree for $\mathcal{F}$.

Assume now $\mathcal{F} \neq \{\varnothing\}$. Let $\{l_1, \ldots, l_n\}$ be a smallest selection in $\mathcal{F}$ and let $\mathcal{F}_i = \{Z \setminus \{l_i\} \mid Z \in \mathcal{F} \text{ and } l_i \in Z\}$. Then each $\mathcal{F}_i$ is a family of pairwise incomparable sets. Now, define $k(\mathcal{F})$ to be a tree obtained from the trees $k(\mathcal{F}_1), \ldots, k(\mathcal{F}_n)$ by introducing a new (root) vertex r and adding for each $i \in \{1, \ldots, n\}$ an edge labelled by l_i from r to the root of $k(\mathcal{F}_i)$.

Recall that by definition $\mathcal{F}_T[v] = \{\varnothing\}$ if v is a leaf, and $\mathcal{F}_T[v] = \{Z \cup \{l(e)\} \mid Z \in \mathcal{F}_T[v'] \text{ and } e = (v, v') \in E\}$ if v is not a leaf. Hence, $T = k(\mathcal{F})$ satisfies $\mathcal{F} = \mathcal{F}_T[r]$, and therefore T represents $\mathcal{F}$. By construction, for each $v \in V$, $O(v)$ is a smallest selection in $\mathcal{F}_T[v]$, and hence T is an optimally selected tree for $\mathcal{F}$. □
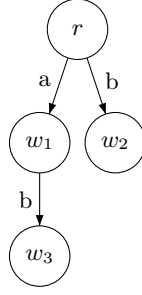
Fig. 3. An unambiguously labelled tree, cf. Example 8.

The following example shows that the condition requiring that the sets in $\mathcal{F}$ are pairwise incomparable is needed for Theorem 7 to hold.

Example 8. Let $\mathcal{F} = \{\{a, b\}, \{b\}\}$. Figure 3 depicts an unambiguously labelled tree that represents $\mathcal{F}$. However, there is no optimally selected tree for $\mathcal{F}$, because the unique smallest selection for $\mathcal{F}$ is $\{b\}$.

It is important to note that for a given family $\mathcal{F}$ of pairwise incomparable sets, an optimally selected tree for $\mathcal{F}$ may be not unique. For example, for $\mathcal{F} = \{\{a, b\}\}$ there are two trees (up to isomorphism) representing $\mathcal{F}$, and both are optimally selected. Indeed, both of these trees "are" a path π of length 2 from the root to the unique leaf, where the labels a and b appear in π either in the order a, b or in the order b, a.

Let $\mathcal{F}$ be a family of sets. Then $\mathcal{G} \subseteq \mathcal{F}$ is *maximally disjoint* w.r.t. $\mathcal{F}$ if the sets of $\mathcal{G}$ are pairwise disjoint and $\mathcal{G}$ is maximal with this property (i.e., each $Y \in \mathcal{F} \setminus \mathcal{G}$ properly intersects with some set in $\mathcal{G}$).

Lemma 9. *Let $\mathcal{F}$ be a finite nonempty family of sets such that $l = \max\{|X| \mid X \in \mathcal{F}\}$ and d is the cardinality of a smallest selection in $\mathcal{F}$. If $\mathcal{G} \subseteq \mathcal{F}$ is maximally disjoint w.r.t. $\mathcal{F}$, then $|\mathcal{G}| \geq \frac{d}{l}$.*

Proof. Let $\mathcal{G} \subseteq \mathcal{F}$ be maximally disjoint w.r.t. $\mathcal{F}$. Hence for each $Y \in \mathcal{F} \setminus \mathcal{G}$, $(\cup_{X \in \mathcal{G}} X) \cap Y \neq \varnothing$. Also, for each $Y \in \mathcal{G}$, $(\cup_{X \in \mathcal{G}} X) \cap Y \neq \varnothing$. Therefore there is a selection S in $\mathcal{F}$ with $S \subseteq \cup_{X \in \mathcal{G}} X$. Hence $d \leq |S| \leq |\cup_{X \in \mathcal{G}} X| \leq l|\mathcal{G}|$. Consequently, $|\mathcal{G}| \geq \frac{d}{l}$.   □

We now apply Lemma 9 to families of pairwise incomparable sets.

Corollary 10. *Let $\mathcal{F}$ be a finite nonempty family of pairwise incomparable sets, and let T represent $\mathcal{F}$. Let $v \in V$ be such that $O(v)$ is a smallest selection in $\mathcal{F}_T[v]$. Then there exists a subfamily $\mathcal{G}$ of $\mathcal{F}_T[v]$ consisting of disjoint sets such that $|\mathcal{G}| \geq \frac{\deg(v)}{\operatorname{hgt}(v)}$.*

Proof. Since $|O(v)| = \deg(u)$ is the cardinality of a smallest selection in $\mathcal{F}_T[v]$, and $|X| \leq \operatorname{hgt}(v)$ for all $X \in \mathcal{F}_T[v]$, by Lemma 9 we obtain $|\mathcal{G}| \geq \frac{\deg(v)}{\operatorname{hgt}(v)}$.   □

4   State Spaces in Trees

We now consider results concerning state spaces of reaction systems in relation to trees. As these results are quite generic and possibly applicable to other domains, we choose to first focus on the essential properties needed to obtain our results (without yet defining reaction systems), and then, in Section 6, relate the results in a precise way to the domain of reaction systems.

Let S be a finite set. We let $\bar{S} = \{\bar{x} \mid x \in S\}$ be a disjoint copy of S, i.e., $S \cap \bar{S} = \varnothing$. Moreover, we let $\bar{\bar{x}} = x$ for $x \in S$. For any subset $X \subseteq (S \cup \bar{S})$, we write $\bar{X} = \{\bar{x} \mid x \in X\}$.

Intuitively (this is made precise in Section 6), we consider S to be the set of all entities of a reaction system. A state W can be considered as a set $Q \cup \overline{(S \setminus Q)}$, where the entities of Q are present and the entities of $S \setminus Q$ are absent (in this state). A substate U is a subset of a state; for the entities in S that do not appear (with or without bar) in a substate it is not known whether or not they are present — hence we deal with incomplete knowledge here. We say that a state W is *compatible* with a substate U if $U \subseteq W$.

Define the *state space* (of S) as $\text{sspace}_S = \{Q \cup \overline{(S \setminus Q)} \mid Q \subseteq S\}$. The elements of sspace_S are called *states* (of S). We define the function $st : 2^S \to \text{sspace}_S$ as follows: for $Q \subseteq S$, $st(Q) = Q \cup \overline{(S \setminus Q)}$. Note that st is a bijection.

A *family of substates* $\mathcal{F}$ (of S) is a subset of $\{Z \mid Z \subseteq W \text{ and } W \in \text{sspace}_S\}$ such that the sets in $\mathcal{F}$ are pairwise incomparable. By Theorem 7 it is possible to represent $\mathcal{F}$ by an optimally selected tree T.

Example 11. Let $S = \{a, b, c\}$. Then $\text{sspace}_S = \{\{a, b, c\}, \{a, b, \bar{c}\}, \{a, \bar{b}, c\}, \{\bar{a}, b, c\}, \{a, \bar{b}, \bar{c}\}, \{\bar{a}, b, \bar{c}\}, \{\bar{a}, \bar{b}, c\}, \{\bar{a}, \bar{b}, \bar{c}\}\}$, and $\mathcal{F} = \{\{a, \bar{b}\}, \{\bar{b}, c\}, \{a, b, c\}\}$ is a family of substates (of S). On the other hand, e.g., $\mathcal{F} = \{\{a, \bar{b}\}, \{a, \bar{b}, c\}\}$ is *not* a family of substates (as $\{a, \bar{b}\} \subset \{a, \bar{b}, c\}$).

Let $\mathcal{F}$ be a family of substates and let T represent $\mathcal{F}$. For each vertex $v \in V$, $\sup(v)$ is a substate of S. The set of states compatible with $\sup(v)$ is denoted by $\mathcal{L}_v$, i.e., $\mathcal{L}_v = \{Q \in \text{sspace}_S \mid \sup(v) \subseteq Q\}$. The set $\mathcal{L}_v$ can be partitioned into the sets $\mathcal{L}_v^+$ and $\mathcal{L}_v^-$, where $\mathcal{L}_v^+$ consists of those states that are compatible with $\sup(u)$ where u is a leaf of $T[v]$. Formally, $\mathcal{L}_v^+ = \{Q \in \mathcal{L}_v \mid \sup(u) \subseteq Q \text{ for some } u \in \text{leav}(T[v])\}$, and $\mathcal{L}_v^- = \mathcal{L}_v \setminus \mathcal{L}_v^+$.

We will need the following technical lemma.

Lemma 12. *Let $\mathcal{F}$ be a family of substates, let T represent $\mathcal{F}$, and let $v \in V$. Let moreover $\mathcal{G} \subseteq \mathcal{F}_T[v]$ be such that the elements of $\mathcal{G}$ are pairwise disjoint, and let $\mathcal{L}_{\mathcal{G}}^- = \{Q \in \mathcal{L}_v \mid Z \not\subseteq Q \text{ for all } Z \in \mathcal{G}\}$. Then*

$$\frac{|\mathcal{L}_{\mathcal{G}}^-|}{|\mathcal{L}_v|} = \prod_{Z \in \mathcal{G}} \left(1 - \frac{1}{2^{|Z|}}\right).$$

Proof. For any $Z \in \mathcal{G}$, the ratio of all $Q \in \mathcal{L}_v$ such that $Z \subseteq Q$ to all $Q \in \mathcal{L}_v$ is $\frac{1}{2^{|Z|}}$. Hence the ratio of all $Q \in \mathcal{L}_v$ such that $Z \not\subseteq Q$ to all $Q \in \mathcal{L}_v$ is $1 - \frac{1}{2^{|Z|}}$. Now, $\frac{|\mathcal{L}_{\mathcal{G}}^{-}|}{|\mathcal{L}_v|}$ is the ratio of all $Q \in \mathcal{L}_v$ such that $Z \not\subseteq Q$ for *all* $Z \in \mathcal{G}$ to all $Q \in \mathcal{L}_v$. Consequently, $\frac{|\mathcal{L}_{\mathcal{G}}^{-}|}{|\mathcal{L}_v|} = \prod_{Z \in \mathcal{G}} \left(1 - \frac{1}{2^{|Z|}}\right)$. □

We now consider the ratio of $|\mathcal{L}_v^{-}|$ to $|\mathcal{L}_v|$, i.e., the fraction of states that are not compatible with $\sup(u)$ for any leaf u of $T[v]$ among all states that *are* compatible with $\sup(v)$.

Theorem 13. *Let $\mathcal{F}$ be a family of substates, let T represent $\mathcal{F}$, and let $v \in V$ be such that $O(v)$ is a smallest selection in $\mathcal{F}_T[v]$. Then*

$$\frac{|\mathcal{L}_v^{-}|}{|\mathcal{L}_v|} \leq \left(1 - \frac{1}{2^{\mathrm{hgt}(v)}}\right)^{\frac{\deg(v)}{\mathrm{hgt}(v)}}.$$

Proof. By Corollary 10 there is a subfamily $\mathcal{G}$ of $\mathcal{F}_T[v]$ consisting of disjoint sets such that $|\mathcal{G}| \geq \frac{\deg(v)}{\mathrm{hgt}(v)}$. Let again $\mathcal{L}_{\mathcal{G}}^{-} = \{Q \in \mathcal{L}_v \mid Z \not\subseteq Q \text{ for all } Z \in \mathcal{G}\}$. Then $\mathcal{L}_v^{-} \subseteq \mathcal{L}_{\mathcal{G}}^{-}$. We have now by Lemma 12

$$\frac{|\mathcal{L}_v^{-}|}{|\mathcal{L}_v|} \leq \frac{|\mathcal{L}_{\mathcal{G}}^{-}|}{|\mathcal{L}_v|} = \prod_{Z \in \mathcal{G}} \left(1 - \frac{1}{2^{|Z|}}\right).$$

Since $|Z| \leq \mathrm{hgt}(v)$ for all $Z \in \mathcal{G}$, we obtain

$$\frac{|\mathcal{L}_v^{-}|}{|\mathcal{L}_v|} \leq \prod_{Z \in \mathcal{G}} \left(1 - \frac{1}{2^{\mathrm{hgt}(v)}}\right).$$

Finally, $|\mathcal{G}| \geq \frac{\deg(v)}{\mathrm{hgt}(v)}$ and thus we obtain

$$\frac{|\mathcal{L}_v^{-}|}{|\mathcal{L}_v|} \leq \left(1 - \frac{1}{2^{\mathrm{hgt}(v)}}\right)^{\frac{\deg(v)}{\mathrm{hgt}(v)}}.$$

Consequently, the theorem holds. □

5 Reaction Systems

In this section we recall some basic notions related to reaction systems, see, e.g., [6] and [1].

Reaction systems is a formal model of the functioning of the living cell. The underlying idea is that this functioning is determined by the interactions between biochemical reactions and these interactions are driven by two mechanisms: facilitation and inhibition.

The formalization of a biochemical reaction follows the basic intuition that a biochemical reaction will take place if all of its reactants are present (in the current state of a biochemical system) and none of its inhibitors is present. When a reaction takes place, it creates its products. This leads to the following definition.

Definition 14. *A* reaction *is a triplet $a = (R, I, P)$, where R, I, P are finite sets such that $R \cap I = \varnothing$.*

The sets R, I, P are also denoted by R_a, I_a, P_a, and called the *reactant set of a*, the *inhibitor set of a*, and the *product set of a*, respectively. If S is a set such that $R, I, P \subseteq S$, then a is a *reaction in S*, and rac(S) denotes the set of all reactions in S.

Usually (see, e.g., [1]) one requires that, for each reaction (R, I, P), both R and I are nonempty. However, in this paper we use a tree representation of a set of reactions, and moving from the root to the leaf representing a reaction $a = (R_a, I_a, P_a)$ corresponds to gaining knowledge of a (from zero knowledge in the root to the full knowledge in the leaf). On the way along this path we represent the current knowledge by the currently known part of R_a and the currently known part of I_a, and either of these parts may be empty before we arrive at the leaf representing a. Thus, for the sake of simplicity, rather than to introduce the notion of a "pseudo reaction", in this paper we do allow the empty reactant set and the empty inhibitor set.

Example 15. Let $S = \{s_1, s_2, s_3, s_4\}$, $a_1 = (\{s_2, s_3\}, \{s_4\}, \{s_1\})$, and $a_2 = (\{s_2\}, \{s_1\}, \{s_1, s_4\})$. Then $a_1, a_2 \in$ rac(S) and, e.g., $P_{a_2} = \{s_1, s_4\}$.

A reaction system is a basic construct of the whole framework of reaction systems (see, e.g., [1]). It is essentially a finite set of reactions, however one also specifies the (background) set of all entities which are needed for specifying the reactions and for reasoning about the system.

Definition 16. *A* reaction system *is an ordered pair $\mathcal{A} = (S, A)$ such that S is a finite set, and $A \subseteq$ rac(S).*

The set S is called the *background set of $\mathcal{A}$*, its elements are called *entities*, and A is called the *set of reactions of $\mathcal{A}$* — note that since S is finite, so is A.

Definition 17. *Let W be a finite set, and let a be a reaction. Then a is* enabled *by W, denoted by a* en *W, if $R_a \subseteq W$ and $I_a \cap W = \varnothing$. The* result of a on W, *denoted by* res$_a(W)$, *is defined by:* res$_a(W) = P_a$ *if a* en *W, and* res$_a(W) = \varnothing$ *otherwise.*

A *state* of a reaction system is a subset $W \subseteq S$ of the background set. A reaction a is enabled in state W if *all* of its reactants are present in W while *none* of its inhibitors are in W. This is the reason that we assume in Definition 14 that,

for each reaction a, $R_a \cap I_a = \varnothing$, as otherwise a is never enabled. When a takes place it produces entities from P_a.

The effect of a *set* of reactions A is cumulative: the *result of A on W*, denoted by $\mathrm{res}_A(W)$, is defined by: $\mathrm{res}_A(W) = \bigcup_{a \in A} \mathrm{res}_a(W)$. For a reaction system $\mathcal{A} = (S, A)$, we write $\mathrm{res}_{\mathcal{A}}(W) = \mathrm{res}_A(W)$.

Example 18. Consider the reaction system $\mathcal{A} = (S, A)$ with S from Example 15 and $A = \{a_1, a_2\}$ with a_1 and a_2 from Example 15. Then for state $W = \{s_2, s_3\}$, we have $\mathrm{res}_{\mathcal{A}}(W) = \{s_1\} \cup \{s_1, s_4\} = \{s_1, s_4\}$, and for this successor state $W' = \{s_1, s_4\}$ we have $\mathrm{res}_{\mathcal{A}}(W') = \varnothing$.

In this paper we assume that the processes of reaction systems are so-called context-independent (see, e.g., [1]), i.e., we assume the behaviour of the reaction systems as a closed system (when there is no interference by the environment). Therefore the state transitions are determined only by the reactions of a reaction system (hence by the result function res_A).

The definition of the result function implies that the successor state consists of only the entities produced by the reactions in the current state. Thus there is *no permanency of entities*: an entity from a current state vanishes (in the transition to the successor state) unless it is produced/sustained by a reaction. This reflects the basic bioenergetics of the living cell, and forms a major difference with models of computation in computer science.

We also notice that the result of the set of reactions is cumulative and so we do not have the notion of conflict between reactions (if they share reactants). Thus we assume the *threshold nature of resources*: either an entity is available and then there is "enough of it" or it is not available. This reflects the level of abstraction adopted in the reaction systems model.

In this paper we consider the notion of enabling. In particular, we are interested, given a reaction system $\mathcal{A}$ and some partial knowledge of a state W, in the ratio of the states of $\mathcal{K}$ for which some reaction of $\mathcal{A}$ is enabled to all states of $\mathcal{K}$, where $\mathcal{K}$ is the set of states "compatible" with the partial knowledge of W.

Two reaction systems are called equivalent if they have the same behavior w.r.t. the $\mathrm{res}_{\mathcal{A}}$ function. This is formalized as follows.

Definition 19. *Reaction systems $\mathcal{A}$ and $\mathcal{A}'$ with common background set S are called* equivalent *if for all $W \subseteq S$, $\mathrm{res}_{\mathcal{A}}(W) = \mathrm{res}_{\mathcal{A}'}(W)$.*

Let a be a reaction, $\{P_1, P_2\}$ be a partition of P_a, $a_1 = (R_a, I_a, P_1)$, and $a_2 = (R_a, I_a, P_2)$. Then, as the result of a set A of reactions is cumulative, we have for all $W \subseteq S$, $\mathrm{res}_a(W) = \mathrm{res}_{\{a_1, a_2\}}(W)$. Hence, the reaction systems $(S, \{a\})$ and $(S, \{a_1, a_2\})$ are equivalent. Consequently, we say that $\mathcal{A}$ is in *singleton product normal form* if $|P_a| = 1$ for all $a \in A$.

Moreover, if $a, a' \in A$ with $P_a = P_{a'}$, $R_a \subseteq R_{a'}$ and $I_a \subseteq I_{a'}$, then for all $W \subseteq S$ $\mathrm{res}_{\{a, a'\}}(W) = \mathrm{res}_{\{a\}}(W)$. Therefore, reaction systems $(S, \{a\})$ and $(S, \{a, a'\})$ are equivalent. Hence, we may delete superfluous reactions: we say that $\mathcal{A}$ (or

A) is *reduced* if for all different $a, a' \in A$ with $P_a = P_{a'}$ we cannot have both $R_a \subseteq R_{a'}$ and $I_a \subseteq I_{a'}$.

Example 20. Reaction system $\mathcal{A} = (S, A)$ with $A = \{a_1, a_2, a_3\}$, $a_1 = (\{s_2, s_3\}, \{s_4\}, \{s_1\})$, $a_2 = (\{s_2\}, \{s_1\}, \{s_1\})\}$, and $a_3 = (\{s_2\}, \{s_1\}, \{s_4\})\}$ is both in singleton product normal form and reduced. As a matter of fact, $\mathcal{A}$ is equivalent to the reaction system from Example 18.

We assume from now on that each reaction system is both reduced and in singleton product normal form.

6   State Space of Reaction Systems

In this section we describe how reactions and states of reaction systems may be translated to states and pairwise incomparable sets as discussed in Sections 3 and 4. In this way, we can apply results of these sections to reaction systems — see Theorem 25.

Let $\mathcal{A} = (S, A)$ be a reaction system. As $\mathcal{A}$ is in singleton product normal form, the reactions of $\mathcal{A}$ can be partitioned according to the (unique) product p of each reaction. We fix now a $p \in S$, and consider the reaction system $\mathcal{A}_p = (S, A_p)$ with $A_p = \{a \in A \mid P_a = \{p\}\}$.

Note that A_p can be uniquely represented as the family of substates $\mathcal{F}_{A_p} = \{R_a \cup \bar{I}_a \mid a \in A_p\}$ of S. Indeed, as A is reduced, A_p is reduced as well, and so the sets in $\mathcal{F}_{A_p}$ are pairwise incomparable.

Example 21. Consider the reaction system $\mathcal{A} = (S, A)$ from Example 20. Then $\mathcal{A}_{s_1} = (S, A_{s_1})$ and $A_{s_1} = \{a_1, a_2\}$ with $a_1 = (\{s_2, s_3\}, \{s_4\}, \{s_1\})$ and $a_2 = (\{s_2\}, \{s_1\}, \{s_1\})\}$. We may represent $\mathcal{A}_{s_1}$ by the family $\mathcal{F}_{A_{s_1}} = \{\{s_2, s_3, \bar{s}_4\}, \{s_2, \bar{s}_1\}\}$ of substates. Note that the sets in $\mathcal{F}_{A_{s_1}}$ are indeed pairwise incomparable.

We may also move the other way around. Let $\mathcal{F}$ be a family of substates of S and fix a $p \in S$. For $Y \in \mathcal{F}$, we define $R_Y = Y \cap S$, $I_Y = \bar{Y} \cap S$, and $a_Y = (R_Y, I_Y, \{p\})$. Hence $A_{\mathcal{F}} = \{a_Y \mid Y \in \mathcal{F}\}$ is a set of reactions corresponding to $\mathcal{F}$. We say that $\mathcal{A}_{p,\mathcal{F}} = (S, A_{\mathcal{F}})$ is the *reaction system of $\mathcal{F}$ w.r.t. $p \in S$*.

Example 22. Consider again the family $\mathcal{F} = \{\{a, \bar{b}\}, \{\bar{b}, c\}, \{a, b, c\}\}$ of substates of $S = \{a, b, c\}$ from Example 11. The reaction system $\mathcal{A}_{p,\mathcal{F}}$ of $\mathcal{F}$ w.r.t. b is defined by: $\mathcal{A}_{p,\mathcal{F}} = (S, \{a_1, a_2, a_3\})$ with $a_1 = (\{a\}, \{b\}, \{b\})$, $a_2 = (\{c\}, \{b\}, \{b\})$, and $a_3 = (\{a, b, c\}, \varnothing, \{b\})$.

If a tree T represents $\mathcal{F}_{A_p}$, then we will also simply say that T *represents* A_p. For a tree T representing A_p and vertex v of T, we define $R_v = R_{\sup(v)}$, $I_v = I_{\sup(v)}$, $a_v = a_{\sup(v)}$, and $A_v = \{a_w \mid w \in \mathrm{leav}(T[v])\}$. Note that $\sup(v) = R_v \cup \overline{I_v}$. We define $\mathcal{K}_v = \{W \subseteq S \mid a_v \text{ en } W\}$, and then we let $\mathcal{K}_v^+ = \{W \in \mathcal{K}_v \mid a \text{ en } W, \text{ for some } a \in A_v\}$ and $\mathcal{K}_v^- = \mathcal{K}_v \setminus \mathcal{K}_v^+$.

The following result holds (recall the function st from Section 4).

Lemma 23. *Let a be a reaction in S and $W \subseteq S$. Then a en W iff $R_a \cup \bar{I}_a \subseteq st(W)$.*

Proof. We have a en W iff both $R_a \subseteq W$ and $I_a \cap W = \varnothing$ iff both $R_a \subseteq W$ and $\bar{I}_a \subseteq \overline{(S \setminus W)}$ iff $R_a \cup \bar{I}_a \subseteq W \cup \overline{(S \setminus W)} = st(W)$.   □

The following lemma will be used to transfer the results of Section 4 to reaction systems.

Lemma 24. *Let $\mathcal{A} = (S, A)$ be a reaction system, and $p \in S$. Let T be a tree representing A_p, and let $v \in V$. Then $st(\mathcal{K}_v) = \mathcal{L}_v$, $st(\mathcal{K}_v^+) = \mathcal{L}_v^+$, and $st(\mathcal{K}_v^-) = \mathcal{L}_v^-$.*

Proof. We have $\mathcal{K}_v = \{W \subseteq S \mid a_v \text{ en } W\}$. Hence by Lemma 23, $\mathcal{K}_v = \{W \subseteq S \mid R_v \cup \bar{I}_v \subseteq st(W)\} = \{W \subseteq S \mid \sup(v) \subseteq st(W)\}$. Thus, $st(\mathcal{K}_v) = \{Q \in \text{sspace}_S \mid \sup(v) \subseteq Q\} = \mathcal{L}_v$.

Next, we have $\mathcal{K}_v^+ = \{W \in \mathcal{K}_v \mid a \text{ en } W, \text{ for some } a \in A_v\}$. We obtain similarly, $\mathcal{K}_v^+ = \{W \in \mathcal{K}_v \mid \sup(w) \subseteq st(W) \text{ for some } w \in \text{leav}(T[v])\}$, and so $st(\mathcal{K}_v^+) = \{Q \in \mathcal{L}_v \mid \sup(w) \subseteq Q \text{ for some } w \in \text{leav}(T[v])\} = \mathcal{L}_v^+$.

Finally, $st(\mathcal{K}_v^-) = st(\mathcal{K}_v \setminus \mathcal{K}_v^+)$. As st is a bijection, $st(\mathcal{K}_v \setminus \mathcal{K}_v^+) = st(\mathcal{K}_v) \setminus st(\mathcal{K}_v^+) = \mathcal{L}_v \setminus \mathcal{L}_v^+ = \mathcal{L}_v^-$.   □

Given a set of reactions A_p and a vertex v of tree T representing A_p (such that T is optimally selected at v), the following result gives an upper bound on the fraction of states $W \subseteq S$ for which no reaction of A_p is enabled among the states that are "compatible" with v (i.e., the states in $\mathcal{K}_v$). In other words, this result gives an upper bound on the fraction of *dead states* (i.e., the states for which there is no reaction enabled) within the subspace of the state space determined by v.

As a consequence of Lemma 24 we obtain that $|\mathcal{K}_v| = |\mathcal{L}_v|$, $|\mathcal{K}_v^+| = |\mathcal{L}_v^+|$, and $|\mathcal{K}_v^-| = |\mathcal{L}_v^-|$. The following result follows then from Theorem 13.

Theorem 25. *Let $\mathcal{A} = (S, A)$ be a reaction system, and $p \in S$. Let T be a tree representing A_p, and let $v \in V$ such that $O(v)$ is a smallest selection in $\mathcal{F}_T[v]$. Then*

$$\frac{|\mathcal{K}_v^-|}{|\mathcal{K}_v|} \leq \left(1 - \frac{1}{2^{\text{hgt}(v)}}\right)^{\frac{\text{deg}(v)}{\text{hgt}(v)}}.$$

Note that the upper bound in Theorem 25 is determined *only* by the length $\text{hgt}(v)$ and the out-degree $\text{deg}(v)$ of v.

Theorem 25 yields the following corollary.

Corollary 26. *Let $\mathcal{A} = (S, A)$ be a reaction system, and $p \in S$. Let T be a tree representing A_p, and let $v \in V$ such that $O(v)$ is a smallest selection in $\mathcal{F}_T[v]$. Then*

$$\frac{|\mathcal{K}_v^-|}{|\mathcal{K}_v|} \leq e^{-\frac{\text{deg}(v)}{\text{hgt}(v) \cdot 2^{\text{hgt}(v)}}}.$$

Proof. Recall that $\left(1 - \frac{1}{m}\right)^m \leq e^{-1}$ for all positive integers m. By Theorem 25,

$$\frac{|\mathcal{K}_v^-|}{|\mathcal{K}_v|} \leq \left(1 - \frac{1}{2^{\mathrm{hgt}(v)}}\right)^{\frac{\deg(v)}{\mathrm{hgt}(v)}} = \left(\left(1 - \frac{1}{2^{\mathrm{hgt}(v)}}\right)^{2^{\mathrm{hgt}(v)}}\right)^{\frac{\deg(v)}{\mathrm{hgt}(v) \cdot 2^{\mathrm{hgt}(v)}}} \leq e^{-\frac{\deg(v)}{\mathrm{hgt}(v) \cdot 2^{\mathrm{hgt}(v)}}}.$$

$\square$

Acknowledgements

References

1. R. Brijder, A. Ehrenfeucht, M. G. Main, and G. Rozenberg. A tour of reaction systems. To appear in *Fundamenta Informaticae*, 2011.
2. R. Brijder, A. Ehrenfeucht, and G. Rozenberg. A note on causalities in reaction systems. *Electronic Communications of the EASST*, 30, 2010.
3. A. Ehrenfeucht, M. G. Main, and G. Rozenberg. Combinatorics of life and death for reaction systems. *International Journal of Foundations of Computer Science*, 21(3):345–356, 2010.
4. A. Ehrenfeucht, M. G. Main, and G. Rozenberg. Functions defined by reaction systems. *International Journal of Foundations of Computer Science*, 22(1):167–178, 2011.
5. A. Ehrenfeucht and G. Rozenberg. Events and modules in reaction systems. *Theoretical Computer Science*, 376(1-2):3–16, 2007.
6. A. Ehrenfeucht and G. Rozenberg. Reaction systems. *Fundamenta Informaticae*, 75(1-4):263–280, 2007.
7. A. Ehrenfeucht and G. Rozenberg. Introducing time in reaction systems. *Theoretical Computer Science*, 410(4-5):310–322, 2009.