Electronic Communications of the EASST
Volume 30 (2010)

International Colloquium on Graph and Model Transformation On the occasion of the 65th birthday of Hartmut Ehrig (GraMoT 2010)

A Note on Causalities in Reaction Systems

Robert Brijder, Andrzej Ehrenfeucht, and Grzegorz Rozenberg

9 pages

Guest Editors: Claudia Ermel, Hartmut Ehrig, Fernando Orejas, Gabriele Taentzer
Managing Editors: Tiziana Margaria, Julia Padberg, Gabriele Taentzer
ECEASST Home Page: http://www.easst.org/eceasst/ ISSN 1863-2122

A Note on Causalities in Reaction Systems

Robert Brijder[*1], **Andrzej Ehrenfeucht**[2], **and Grzegorz Rozenberg**[1 2]

[1] Leiden Institute of Advanced Computer Science,
Leiden Center for Natural Computing,
Leiden University, The Netherlands

[2] Department of Computer Science,
University of Colorado at Boulder, USA

Abstract: Reaction systems are a formal model of interactions between biochemical reactions. In this note we initiate an investigation of causalities in reaction systems which reflect the way that elements (entities) of a reaction system influence each other.

Keywords: natural computing; biochemical interactions; reaction systems; causal relationships

1 Introduction

Reaction systems are a formal model of interactions between biochemical reactions which is based on the idea that the underlying mechanisms of these interactions as well as the working of an individual reaction are: facilitation and inhibition.

Therefore a reaction is defined as a triplet of finite nonempty sets $a = (R, I, P)$, where R is the set of reactants needed for a to take place, I is the set of inhibitors each of which forbids a to take place, and P is the set of products produced by a when it takes place. The set $R \cup I$ forms the resources of a — these are all entities that directly influence a either as reactants or as inhibitors. Reactions (of a given biochemical system) influence each other through their products — they may contain entities which are reactants for some reactions (therefore facilitating these reactions) and they may contain entities which are inhibitors for some reactions (therefore inhibiting these reactions).

A reaction system $\mathscr{A} = (S, A)$ consists of a finite set of reactions A and a finite background set of entities used in reactions of A and entities needed to analyze the functioning of $\mathscr{A}$.

Research concerning reaction systems is quite broad. For example, it covers fundamental issues such as the notion of time in reaction systems ([ER09]), it is concerned with the dynamic processes in reaction systems and the way these processes guide the formations of compounds ([ER07a]), and it investigates the mathematical nature of functions (from states to states, and hence from finite sets into finite sets) definable by reaction systems ([EMR10]). In this note we initiate research on causalities in reaction systems, i.e., the ways that entities of a reaction system influence each other. We discuss here both static/structural causalities (i.e., embedded directly in the definition/specification of a reaction system) as well as dynamic causalities (i.e., the relationships formed through the dynamic runs of a reaction system).

* Supported by the Netherlands Organization for Scientific Research (NWO), project "Annotated graph mining".

2 Preliminaries

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

As usual, $\mathbb{Z}^+$ is the set of positive integers, and we let ω be the cardinality of $\mathbb{Z}^+$. The symmetric difference of sets Z_1 and Z_2, defined by $(Z_1 \setminus Z_2) \cup (Z_2 \setminus Z_1)$, is denoted by $Z_1 \oplus Z_2$. For a nonempty $U \subseteq \mathbb{Z}^+$, the minimal integer of U is denoted by $\min(U)$.

Let $\tau = W_0, \ldots, W_n$ be a sequence of sets. For a set S, we say that τ is an *S-sequence* if $W_i \subseteq S$ for all $i \in \{0, \ldots, n\}$. We denote the length of τ by $|\tau|$ (note that $|\tau| = n+1$). For a set Q, the *Q-projection of τ* is the Q-sequence of sets $\mathrm{proj}_Q(\tau) = W_0 \cap Q, \ldots, W_n \cap Q$.

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 *edges*. Note that we allow loops $(x,x) \in E$. For $x \in V$, $y \in V$ is an *outgoing (incoming, resp.) vertex of x* if $(x,y) \in E$ ($(y,x) \in E$, resp.). The set of outgoing (incoming, resp.) vertices of x is denoted by $\mathrm{out}_G(x)$ ($\mathrm{inc}_G(x)$, resp.). The *out-degree (in-degree, resp.) of x*, denoted by $\mathrm{od}_G(x)$ ($\mathrm{id}_G(x)$, resp.), is the number of outgoing (incoming, resp.) vertices of x, i.e., it equals $|\mathrm{out}_G(x)|$ ($|\mathrm{inc}_G(x)|$, resp.).

3 Reactions and Reaction Systems

In this and in the following section we recall the basic notions related to reaction systems (see, e.g., [ER07b]).

Definition 1 A *reaction* is a triplet $a = (R, I, P)$, where R, I, P are finite nonempty 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. Also, $M_a = R_a \cup I_a$ is the set of *resources of a*. If S is a set such that $R, I, P \subseteq S$, then a is a *reaction in S*, and $\mathrm{rac}(S)$ denotes the set of all reactions in S.

Definition 2 Let T be a finite set.

1. Let a be a reaction. Then a is *enabled by T*, denoted by *a en T*, if $R_a \subseteq T$ and $I_a \cap T = \varnothing$. The *result of a on T*, denoted by $\mathrm{res}_a(T)$, is defined by: $\mathrm{res}_a(T) = P_a$ if *a en T*, and $\mathrm{res}_a(T) = \varnothing$ otherwise.

2. Let A be a finite set of reactions. The *result of A on T*, denoted by $\mathrm{res}_A(T)$, is defined by: $\mathrm{res}_A(T) = \bigcup_{a \in A} \mathrm{res}_a(T)$.

The intuition behind the finite set T above is that it represents a state of a biochemical system (hence it is the set of biochemical entities present in this state). A reaction a is enabled in T (it will take place in T) if *all* of its reactants are present in T while *none* of its inhibitors are in T. This is the reason that we assume in Definition 1 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 T consists of all products of all reactions from

A that are enabled on T.

Example 1 *Let* $S = \{s_1, s_2, s_3, s_4\}$, $a_1 = (\{s_2\}, \{s_1, s_4\}, \{s_2\})$, *and* $a_2 = (\{s_2, s_3\}, \{s_1\}, \{s_3\})$. *Then* $a_1, a_2 \in \mathrm{rac}(S)$ *and, e.g.,* $M_{a_1} = \{s_1, s_2, s_4\}$ *and* $P_{a_1} = \{s_2\}$. *We have for* $A = \{a_1, a_2\}$, $\mathrm{res}_A(\{s_2, s_3\}) = \{s_2, s_3\}$.

We are ready now to recall the notion of a reaction system.

Definition 3 A *reaction system*, rs for short, is an ordered pair $\mathscr{A} = (S, A)$ such that S is a finite set, and $A \subseteq \mathrm{rac}(S)$.

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

The dynamic behavior of a rs is formalized through the notion of an interactive process.

Definition 4 Let $\mathscr{A} = (S, A)$ be a rs. An *(n-step) interactive process* in $\mathscr{A}$ is a pair $\pi = (\gamma, \delta)$ of finite equal length S-sequences $\gamma = C_0, \ldots, C_n$ and $\delta = D_0, \ldots, D_n$ for some $n \geq 1$, where $D_0 = \varnothing$ and $D_i = \mathrm{res}_{\mathscr{A}}(D_{i-1} \cup C_{i-1})$ for all $i \in \{1, \ldots, n\}$.

The sequence γ is the *context sequence of* π, denoted by $\mathrm{con}(\pi)$, and the sequence δ is the *result sequence of* π, denoted by $\mathrm{res}(\pi)$. Then the sequence $\tau = W_0, W_1, \ldots, W_n$ defined by $W_i = C_i \cup D_i$ for all $i \in \{0, \ldots, n\}$ is the *state sequence of* π, denoted by $\mathrm{st}(\pi)$, with $W_0 = C_0$ called the *initial state of* π (and of τ), denoted by $\mathrm{init}(\pi)$, and W_n called the *final state of* π (and of τ), denoted by $\mathrm{fst}(\pi)$.

If $C_i \subseteq D_i$ for all $i \in \{1, \ldots, n\}$, then we say that π and τ are *context-independent*. Note that a context-independent state sequence depends only on the initial state ($W_0 = C_0$) and its length ($n+1$). The set of all state sequences of $\mathscr{A}$ (i.e., all state sequences of all interactive processes in $\mathscr{A}$) is denoted by $STS(\mathscr{A})$, and the set of all context-independent state sequences of $\mathscr{A}$ is denoted by $CISTS(\mathscr{A})$.

Note that if W_i, W_{i+1} are two consecutive states in the state sequence of an interactive process π, then each entity in W_{i+1} is either produced by reactions from A enabled on W_i or it is provided by the corresponding context (C_{i+1}). Hence each entity in W_{i+1} is *created* through the state transitions from W_i to W_{i+1}. There is no *permanency* in reaction systems — each entity in a current state is there because it is sustained either by a reaction from A or by the context. This is a major difference with models of concurrent systems in computer science (such as Petri nets — see, e.g., [RE98]), where elements from the current state that are not involved in the local transformations performed by the system on this state just persist (go over to the successor state).

Example 2 *Let* $\mathscr{A} = (S, A)$ *be a rs with* $S = \{x, y, z_1, z_2\}$ *and*

$$\begin{aligned} A \;=\; & \{(\{x\}, \{y\}, \{z_1\}) \\ & (\{z_1\}, \{z_2\}, \{x, z_2\}) \\ & (\{z_2\}, \{z_1\}, \{x, y\})\}. \end{aligned}$$

Then the context-independent state sequence τ *with the initial state* $W_0 = \{x\}$ *and length* 5 *is*

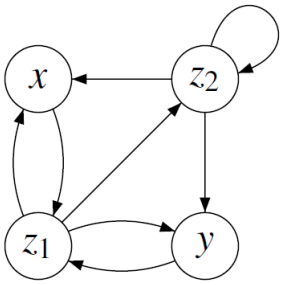

Figure 1: The influence graph from Example 3.

$\tau = W_0, W_1, W_2, W_3, W_4$ *where* $W_1 = \{z_1\}$, $W_2 = \{x, z_2\}$, $W_3 = \{x, y, z_1\}$, *and* $W_4 = \{x, z_2\}$. *We will use this* $\mathscr{A}$ *now as the running example of this paper.*

4 Resource Dependence and Product Influence

There are two basic ways that the entities of a reaction system can influence each other. If a reaction a produces an entity x (i.e., $x \in P_a$), then, for any entity $y \in M_a$, we say that x is resource dependent on y, and that y product-influences x. These dependencies are formally defined as follows.

Definition 5 Let $\mathscr{A} = (S, A)$ be a rs.

- Let $x \in S$.
  1. The *resource dependence set of* x, denoted by MD_x, is defined by $\bigcup\{M_a \mid a \in A, x \in P_a\}$.
  2. The *product influence set of* x, denoted by PI_x, is defined by $\bigcup\{P_a \mid a \in A, x \in M_a\}$.
- Let $q \in \mathbb{Z}^+$.
  1. $\mathscr{A}$ is a *rs with q-bounded resource dependence*, abbreviated by *q-MD rs*, if $|\mathrm{MD}_x| \leq q$ for each $x \in S$.
  2. $\mathscr{A}$ is a *rs with q-bounded product influence*, abbreviated by *q-PI rs*, if $|\mathrm{PI}_x| \leq q$ for each $x \in S$.

We introduce now the notion of the *influence graph* of a rs, which is a very convenient technical tool to investigate resource dependencies and product influences in reaction systems.

Definition 6 Let $\mathscr{A} = (S, A)$ be a rs. The *influence graph of* $\mathscr{A}$, denoted by $\inf_{\mathscr{A}}$, is the digraph (S, E), where for $x, y \in S$, $(x, y) \in E$ if and only if $x \in M_a$ and $y \in P_a$ for some $a \in A$.

Example 3 *The influence graph* $\inf_{\mathscr{A}}$ *of* $\mathscr{A}$ *of Example 2 is given in Figure 1.*

The usefulness of the influence graph of $\mathscr{A}$ in investigating resource dependencies and product influences in $\mathscr{A}$ stems from the fact that these parameters are directly expressible in $\inf_{\mathscr{A}}$ as standard graph-theoretical notions. Thus it is obvious that the following holds.

Lemma 1 *Let $\mathscr{A} = (S,A)$ be a rs. For each $x \in S$, $\mathrm{MD}_x = \mathrm{inc}_{\inf_{\mathscr{A}}}(x)$ and $\mathrm{PI}_x = \mathrm{out}_{\inf_{\mathscr{A}}}(x)$. Moreover, for $q \in \mathbb{Z}^+$, $\mathscr{A}$ is a q-MD rs if and only if $\mathrm{id}_{\inf_{\mathscr{A}}}(x) \leq q$ for all $x \in S$, and $\mathscr{A}$ is a q-PI rs if and only if $\mathrm{od}_{\inf_{\mathscr{A}}}(x) \leq q$ for all $x \in S$.*

Example 4 *From the influence graph of $\mathscr{A}$ in Figure 1 we find that $|\mathrm{PI}_{z_1}| = |\mathrm{PI}_{z_2}| = 3$, $|\mathrm{PI}_x| = |\mathrm{PI}_y| = 1$, and $|\mathrm{MD}_s| = 2$ for all $s \in \{x,y,z_1,z_2\}$. Hence $\mathscr{A}$ is a 3-PI rs and a 2-MD rs.*

We demonstrate now how to use the influence graph to obtain properties of resource dependencies and product influences.

Definition 7 Let $\mathscr{A} = (S,A)$ be a rs.

- The *average resource dependence of* $\mathscr{A}$, denoted by $\mathrm{avMD}(\mathscr{A})$, is defined as $\sum_{x\in S} \frac{|\mathrm{MD}_x|}{|S|}$.
- The *average product influence of* $\mathscr{A}$, denoted by $\mathrm{avPI}(\mathscr{A})$, is defined as $\sum_{x\in S} \frac{|\mathrm{PI}_x|}{|S|}$.

Theorem 1 *For every rs $\mathscr{A}$, $\mathrm{avMD}(\mathscr{A}) = \mathrm{avPI}(\mathscr{A})$.*

Proof. For every digraph $G = (V,E)$, $|E| = \sum_{x\in V} \mathrm{id}_G(x) = \sum_{x\in V} \mathrm{od}_G(x)$, as each edge incoming to some vertex x is outgoing from some vertex y. Hence, by Lemma 1, $\sum_{x\in S} |\mathrm{MD}_x| = \sum_{x\in S} |\mathrm{PI}_x|$, and the theorem holds. □

Note that, in general: *(i)* knowing that rs $\mathscr{A} = (S,A)$ is a q-MD rs does not yield a bound on $\max\{\mathrm{PI}_x \mid x \in S\}$, and symmetrically *(ii)* knowing that $\mathscr{A}$ is a q-PI rs does not yield a bound on $\max\{\mathrm{MD}_x \mid x \in S\}$. However, knowing that q bounds the size of resource dependence (product influence, resp.) of $\mathscr{A}$, by Theorem 1 we know that the average product influence (average resource dependence, resp.) of $\mathscr{A}$ is also bound by q (because the average does not exceed the maximum).

5 Causal Distances

In Section 4 we investigated *static* causalities in reaction systems, i.e., causalities "directly deducible" from the influence graph. In this section we investigate the way that entities influence each other within the *dynamics* of a reaction system, i.e., within interactive processes.

We begin with a useful technical result concerning symmetric differences of states of a rs. Considering symmetric differences allows us to single out the entities by which two states differ (and then to consider consequences of these differences).

Lemma 2 *Let $\mathscr{A} = (S,A)$ be a rs, and let $W,W' \subseteq S$. For each $y_2 \in \mathrm{res}_{\mathscr{A}}(W) \oplus \mathrm{res}_{\mathscr{A}}(W')$, we have that (y_1,y_2) is an edge of $\inf_{\mathscr{A}}$ for some $y_1 \in W \oplus W'$.*

Proof. Let $y_2 \in \mathrm{res}_{\mathscr{A}}(W) \oplus \mathrm{res}_{\mathscr{A}}(W')$. Then there is a reaction a of $\mathscr{A}$ with $y_2 \in P_a$ such that either (1) a is enabled by W and a is not enabled by W', or (2) a is enabled by W' and a is not enabled by W. Without loss of generality we assume case (1). As a is enabled by W and not enabled by W', either $R_a \cap (W \setminus W') \neq \varnothing$ or $I_a \cap (W' \setminus W) \neq \varnothing$. Hence there is a $y_1 \in W \oplus W'$ with $y_1 \in R_a \cup I_a = M_a$. Consequently, $y_2 \in \mathrm{PI}_{y_1}$, and therefore (y_1, y_2) is an edge of $\inf_{\mathscr{A}}$. □

The following lemma follows now from Lemma 2 by induction on n.

Lemma 3 *Let $\mathscr{A}$ be a rs. Let $\tau, \tau' \in CISTS(\mathscr{A})$ such that $\tau = W_0, W_1, \ldots, W_m$, $\tau' = W'_0, W'_1, \ldots, W'_m$ for some $m \geq 1$, and $W_0 \oplus W'_0 = \{x\}$. Then, for each $n \in \{1, \ldots, m\}$, if $y \in W_n \oplus W'_n$, then there is a path from x to y in $\inf_{\mathscr{A}}$ of length n.*

If we consider now q-PI reaction systems, then we obtain a bound on the cardinality of $W_n \oplus W'_n$.

Lemma 4 *Let $\mathscr{A}$ be a q-PI rs for some $q \geq 1$. Let $\tau, \tau' \in CISTS(\mathscr{A})$ be such that $\tau = W_0, W_1, \ldots, W_m$, $\tau' = W'_0, W'_1, \ldots, W'_m$ for some $m \geq 1$, and $|W_0 \oplus W'_0| = 1$. Then, for each $n \in \{1, \ldots, m\}$, $|W_n \oplus W'_n| \leq q^n$.*

Proof. By Lemma 3, for each $y \in W_n \oplus W'_n$, there is a path from x to y of length n. As $\mathscr{A}$ is a q-PI rs, by Lemma 1 there are at most q^n paths from x of length n, and so the result follows. □

Example 5 *We continue the running example. Recall that $\mathscr{A}$ is a 3-PI reaction system. The context-independent state sequence τ' with the initial state $W'_0 = \{x, z_1\}$ and length 5 is $\tau' = W'_0, W'_1, W'_2, W'_3, W'_4$ where $W'_1 = \{x, z_1, z_2\}$, $W'_2 = \{z_1\}$, $W'_3 = \{x, z_2\}$, and $W'_4 = \{x, y, z_1\}$. If we compare τ' with the context-independent state sequence τ with the initial state $W_0 = \{x\}$ in Example 2, then $W_0 \oplus W'_0 = \{z_1\}$, $W_1 \oplus W'_1 = \{x, z_2\}$, $W_2 \oplus W'_2 = \{x, z_1, z_2\}$, $W_3 \oplus W'_3 = W_4 \oplus W'_4 = \{y, z_1, z_2\}$. Hence, the upper bound of Lemma 4 indeed holds for τ and τ' as $|W_0 \oplus W'_0| = 1$, and $|W_1 \oplus W'_1| = 2 \leq 3$, $|W_2 \oplus W'_2| = 3 \leq 9$, etc.*

Definition 8 Let $\mathscr{A} = (S, A)$ be a rs, and $x, y \in S$.

- Let $\tau, \tau' \in CISTS(\mathscr{A})$ where $\tau = W_0, W_1, \ldots, W_m$, $\tau' = W'_0, W'_1, \ldots, W'_m$, and $W_0 \oplus W'_0 = \{x\}$. Let moreover $Z_{x,y}(\tau, \tau') = \{n \in \{0, \ldots, m\} \mid y \in W_n \oplus W'_n\}$. Then the *causal distance from x to y in τ, τ'* is defined by:

$$\delta_{x,y}(\tau, \tau') = \begin{cases} \min Z_{x,y}(\tau, \tau') & Z_{x,y}(\tau, \tau') \neq \varnothing \\ \omega & \text{otherwise} \end{cases}.$$

- The *causal distance from x to y* is defined by:

$$\mathrm{cd}_{x,y} = \min\{\delta_{x,y}(\tau, \tau') \mid \tau, \tau' \in CISTS(\mathscr{A}), |\tau| = |\tau'|, \text{ and } \mathrm{init}(\tau) \oplus \mathrm{init}(\tau') = \{x\}\}.$$

If the initial states of two state sequences τ and τ' (of equal length) differ by x only, then by comparing pairwise the corresponding states of τ and τ' one can reason about the causal influence, within the pair τ, τ', of x on an entity y. If y "appears" in the symmetric difference

of two corresponding states W_n and W'_n, then this appearance is caused by x. If n is the minimal such index (for τ and τ'), then it is the distance of causal influence of x on y within the pair τ, τ'. If on the other hand y never appears in the symmetric difference of two corresponding states of τ and τ', then x does not influence y within the pair τ, τ' and so the distance of causal influence of x on y is "infinite" (it is equal to ω). Obviously, the causal distance between x and y in the total dynamics of $\mathscr{A}$ is defined as the minimal distance over all pairs of state sequences τ, τ' as above. If this distance is n, for some $n \in \mathbb{Z}^+$, then in some situations (pairs τ, τ') x can causally influence y over the distance equal n.

Example 6 *We continue the running example. Recall the context-independent state sequences* τ *and* τ' *from Example 5 with* $\mathrm{init}(\tau) \oplus \mathrm{init}(\tau') = \{z_1\}$. *We have then (see Example 5):* $\delta_{z_1,z_1}(\tau,\tau') = 0$, $\delta_{z_1,x}(\tau,\tau') = \delta_{z_1,z_2}(\tau,\tau') = 1$, *and* $\delta_{z_1,y}(\tau,\tau') = 3$. *Note that we may thus have "gaps": there is no* $v \in S$ *with* $\delta_{z_1,v}(\tau,\tau') = 2$, *but, we have seen that* $\delta_{z_1,y}(\tau,\tau') = 3$. *Also note that* $\mathrm{cd}_{z_1,z_2} = 1$ *as the causal distance between different entities is (by definition) at least* 1, *and we have* $\delta_{z_1,z_2}(\tau,\tau') = 1$.

Using Lemma 3, the causal distance from x to y in state sequences τ, τ' has the following implication in terms of paths in $\inf_{\mathscr{A}}$.

Lemma 5 *Let* $\mathscr{A} = (S,A)$ *be a rs and let* $x \in S$. *Let* $\tau, \tau' \in CIST S(\mathscr{A})$ *such that* $|\tau| = |\tau'|$, *and* $\mathrm{init}(\tau) \oplus \mathrm{init}(\tau') = \{x\}$. *If* $\delta_{x,y}(\tau,\tau') = d$, *then there is a path from* x *to* y *in* $\inf_{\mathscr{A}}$ *of length* d.

Using Lemma 4, we can now bound the number of entities that have a causal distance d from a given entity x.

Theorem 2 *Let* $\mathscr{A} = (S,A)$ *be a rs and let* $x \in S$. *If* $\mathscr{A}$ *is a* q*-PI rs for* $q \geq 1$, *then for every* $d \in \mathbb{Z}^+$, $|\{y \in S \mid \mathrm{cd}_{x,y} = d\}| \leq q^d$.

As a corollary we prove that if, for a q-PI rs $\mathscr{A} = (S,A)$, there is a common (finite) bound on causal distances from x to y for all $x, y \in S$, then this common bound can be used to bound $|S|$.

Corollary 1 *Let* $\mathscr{A} = (S,A)$ *be a* q*-PI rs for some* $q \geq 1$. *Let* $x \in S$, *and let* $n_0 \geq 0$ *be such that* $\mathrm{cd}_{x,y} \leq n_0$ *for all* $y \in S$. *Then* $|S| \leq \sum_{d=0}^{n_0} q^d$.

Proof. Let $x \in S$. Then by Lemma 5, for every $d \in \mathbb{Z}^+$, $|\{y \in S \mid \mathrm{cd}_{x,y} = d\}| \leq q^d$. Hence $|\{y \in S \mid \mathrm{cd}_{x,y} \leq n_0\}| \leq \sum_{d=0}^{n_0} q^d$. Since $\mathrm{cd}_{x,y} \leq n_0$ for all $y \in S$, $|S| = |\{y \in S \mid \mathrm{cd}_{x,y} \leq n_0\}|$, and so $|S| \leq \sum_{d=0}^{n_0} q^d$. □

We note here that the causalities we investigate are *between entities* of a reaction system. This is quite different from the traditional research on causalities in models of concurrent systems (see, e.g., [RE98]), where the causal dependencies hold between events (actions of a system). We can do this, because (as pointed out in Section 3) each entity in a current state is *created* in the transition from the previous state. Hence our causal dependencies between entities x and y can be also seen as causal dependencies between the actions of creating x and y.

6 Predictability

Let $\mathscr{A} = (S,A)$ be a rs. Assume that we are interested in a specific $x \in S$, and we would like to know (to be able to predict) whether or not, for a specific $n \in \mathbb{Z}^+$, x will be present in the final state of a n-step process π. Since π is uniquely determined by $\mathrm{con}(\pi)$ (and A), knowing $\mathrm{con}(\pi)$ allows us to answer this query. However, since we are interested in a specific x and a specific n, perhaps to answer this query it suffices to know *only a part of* (each set of) $\mathrm{con}(\pi)$. More specifically, perhaps (for given x and n) there is a subset $Q \subseteq S$ which is the key to answering this query, meaning that if, for any two n-step interactive processes, the Q-projections of the context sequences of these processes are equal, then either x is in both final states (of these processes) or in none of them. Then Q is a subset of S which is a cause for x to be *uniformly* either present or absent in the final state of any n-step interactive process. Such predicting subsets of S are investigated in this section.

Definition 9 Let $\mathscr{A} = (S,A)$ be a rs. For $x \in S$, $n \geq 1$, and $Q \subseteq S$, we say that Q *n-predicts* x, if for arbitrary n-step interactive processes π_1 and π_2 the following holds:
if $\mathrm{proj}_Q(\mathrm{con}(\pi_1)) = \mathrm{proj}_Q(\mathrm{con}(\pi_2))$, then $x \in \mathrm{fst}(\pi_1)$ if and only if $x \in \mathrm{fst}(\pi_2)$.

Note that for all $x \in S$ and all $n \geq 1$, S n-predicts x.

Let, for $x \in S$ and $n \geq 1$, $\mathscr{P}_{x,n} = \{Q \subseteq S \mid Q \text{ } n\text{-predicts } x\}$. Since $S \in \mathscr{P}_{x,n}$, $\mathscr{P}_{x,n}$ is nonempty, and so it contains *minimal elements* (w.r.t. inclusion).

Theorem 3 *Let $\mathscr{A} = (S,A)$ be a rs, $x \in S$ and $n \geq 1$. Then $\mathscr{P}_{x,n}$ contains exactly one minimal element.*

Proof. Assume to the contrary that $\mathscr{P}_{x,n}$ contains two different minimal sets Q_1 and Q_2. Let $Z = Q_1 \cap Q_2$. As Z is strictly included in Q_1 (and Q_2), and Q_1 and Q_2 are minimal, Z does not n-predict x. Let thus π' and π'' be arbitrary n-step interactive processes such that $\mathrm{proj}_Z(\gamma') = \mathrm{proj}_Z(\gamma'')$ with $\gamma' = \mathrm{con}(\pi')$, $\gamma'' = \mathrm{con}(\pi'')$, and $x \in \mathrm{fst}(\pi')$, while $x \notin \mathrm{fst}(\pi'')$.

Let $\gamma' = C'_0,\ldots,C'_n$ and $\gamma'' = C''_0,\ldots,C''_n$. Consider now the S-sequence $\gamma = C_0,\ldots,C_n$, where $C_i = (Q_1 \cap C'_i) \cup (Q_2 \cap C''_i)$ for $i \in \{0,\ldots,n\}$.

We have $Q_1 \cap C_i = (Q_1 \cap C'_i) \cup (Q_1 \cap Q_2 \cap C''_i)$. Also, we have $Q_1 \cap Q_2 \cap C''_i = Z \cap C''_i$. Moreover, $Z \cap C''_i = Z \cap C'_i$, because $\mathrm{proj}_Z(\gamma'') = \mathrm{proj}_Z(\gamma')$. Since also $Z \subseteq Q_1$, we obtain $Q_1 \cap C_i = Q_1 \cap C'_i$. Consequently, $\mathrm{proj}_{Q_1}(\gamma) = \mathrm{proj}_{Q_1}(\gamma')$.

Analogously one proves that $\mathrm{proj}_{Q_2}(\gamma) = \mathrm{proj}_{Q_2}(\gamma'')$.

Let π be the interactive process corresponding to γ. Since Q_1 n-predicts x and $x \in \mathrm{fst}(\pi')$, we have $x \in \mathrm{fst}(\pi)$. Similarly, since Q_2 n-predicts x and $x \notin \mathrm{fst}(\pi'')$, $x \notin \mathrm{fst}(\pi)$ — a contradiction. Therefore $\mathscr{P}_{x,n}$ contains exactly one minimal element. □

We denote the (unique) minimal element of $\mathscr{P}_{x,n}$ for $x \in S$ and $n \geq 1$, by $\mathrm{prd}_{\mathscr{A}}(x,n)$, and refer to it as *the n-predictor of x (in $\mathscr{A}$)*.

Note that in a context-independent n-step interactive process $\pi = (\gamma,\delta)$, as far as the state sequence $\mathrm{st}(\pi)$ is concerned, we can assume that $\gamma = C_0,\ldots,C_n$ is such that $C_1 = \emptyset,\ldots,C_n = \emptyset$ (because all the contributions of $C_1,\ldots,C_n$ to $\mathrm{st}(\pi)$ are already included in $D_1,\ldots,D_n$ where $\delta = D_0,\ldots,D_n$). Therefore, if, for $x \in S$, we want to know whether or not x appears in the final

state of an arbitrary *context-independent* n-step interactive process π, it suffices to know which entities of $\mathrm{prd}_{\mathscr{A}}(x,n)$ are included in the initial state of π — all other entities of the initial state are irrelevant as far as this query is concerned!

Our next result bounds the size of n-predictors.

Theorem 4 *Let $\mathscr{A} = (S,A)$ be a q-MD rs for $q \geq 1$. For each $x \in S$ and each $n \geq 1$, $|\mathrm{prd}_{\mathscr{A}}(x,n)| \leq \sum_{k=0}^{n} q^k$.*

Proof. Each $y \in \mathrm{prd}_{\mathscr{A}}(x,n)$ product-influences x in at most n steps. By the definition of influence graph, for each entity y that product-influences x in k-step (for some k), there is a path of length k from y to x. Since $\mathscr{A}$ is a q-MD rs, by Lemma 1 there are at most $\sum_{k=0}^{n} q^k$ paths *to* x in $\inf_{\mathscr{A}}$ of length at most n. Therefore $|\mathrm{prd}_{\mathscr{A}}(x,n)| \leq \sum_{k=0}^{n} q^k$. □

Example 7 *Recall that $\mathscr{A}$ in the running example is a 2-MD rs (see Example 4). Hence we have, e.g., $|\mathrm{prd}_{\mathscr{A}}(z_1,1)| \leq 3$.*

Acknowledgements

The authors are indebted to Mike Main and the anonymous referees for useful comments.

Bibliography

[EMR10] A. Ehrenfeucht, M. G. Main, G. Rozenberg. Combinatorics of Life and Death for Reaction Systems. *International Journal of Foundations of Computer Science* 21(3):345–356, 2010.
doi:10.1142/S0129054110007295

[ER07a] A. Ehrenfeucht, G. Rozenberg. Events and modules in reaction systems. *Theoretical Computer Science* 376(1-2):3–16, 2007.
doi:10.1016/j.tcs.2007.01.008

[ER07b] A. Ehrenfeucht, G. Rozenberg. Reaction Systems. *Fundamenta Informaticae* 75(1-4):263–280, 2007.

[ER09] A. Ehrenfeucht, G. Rozenberg. Introducing time in reaction systems. *Theoretical Computer Science* 410(4-5):310–322, 2009.
doi:10.1016/j.tcs.2008.09.043

[RE98] G. Rozenberg, J. Engelfriet. Elementary Net Systems. In Reisig and Rozenberg (eds.), *Lectures on Petri Nets I: Basic Models*. Lecture Notes in Computer Science 1491, pp. 12–121. Springer, 1998.
doi:10.1007/3-540-65306-6_14