



# Approximate Reaction Systems Based on Rough Set Theory

Andrea Campagner<sup>1</sup>(✉), Davide Ciucci<sup>1</sup>, and Valentina Dorigatti<sup>2</sup>

<sup>1</sup> Dipartimento di Informatica, Sistemistica e Comunicazione,  
University of Milano–Bicocca, Viale Sarca 336/14, 20126 Milan, Italy  
a.campagner@campus.unimib.it

<sup>2</sup> Dipartimento di Scienze Teoriche e Applicate, University of Insubria,  
Via J.H. Dunant 3, 21100 Varese, Italy

**Abstract.** In this work we investigate how Rough Set Theory could be employed to model uncertainty and information incompleteness about a Reaction System. The approach that we propose is inspired by the idea of an abstract scientific experiment: we define the notion of *test*, which defines an approximation space on the states of a Reaction System, and *observation*, to represent the interactive process of knowledge building that is typical of complex systems. We then define appropriate notions of reducts and study their characterization in terms of both computational complexity and relationships with standard definitions of reducts in terms of Information Tables.

**Keywords:** Complex systems · Reaction Systems · Rough sets

## 1 Introduction

Complex systems, that are characterized by the mutual interaction of basic components, represent currently one of the topics of major interest in many disciplines. This interest has been fostered both by the potential impact that these systems have in the real world and also by the difficulty that they pose with respect to the modeling and formalization point of view. Indeed, as interaction represents one of the main features of complex systems, there has been increasing attention towards developing mathematical and formal models that are explicitly based on the notion of interaction: some prominent examples are cellular automata [5], membrane computing [14], formalisms to describe concurrent processes [4, 16], reaction systems [8]. This latter class of models has recently been proposed as a simple and abstract formalization of biochemical processes involving *substances* and *reactions*, by which the states (i.e., collections of substances) are transformed. While interesting from a computational or purely mathematical point of view, one of the major limitations of this framework (and, more in general, of abstract *idealized* models of complex systems), as recently acknowledged in [6], relates to the fact that these models *ignore* the realistic aspects that are *intrinsic* in complexity, in particular with respect to the fact that information

available about these systems is usually only *partial*, *uncertain* and *incomplete* and acquired through *interaction* with the system.

Rough Set theory [15] has originally been proposed to explicitly deal with this type of information: both with respect to the representation of uncertain and potentially incomplete information [13] (through the notion of lower and upper approximations) and also with respect to knowledge acquisition [3, 10, 22] (through the notion of reducts and rule extraction). Indeed, the relationship between these two mathematical frameworks have been investigated, under the perspective of Interactive Granular Computing [19], in [6, 18] where Rough Set Theory is integrated with Reaction Systems in order to be able to account for uncertainty and incomplete knowledge in the latter formalism.

In this work, we also discuss how to relate these two modeling frameworks, though under a different perspective. Indeed, the main purpose of this article is to investigate how Rough Set Theory can be used to study Reaction Systems, both from the modeling point of view and from the uncertainty representation and management one. More specifically, we will consider the case where states of a Reaction System are not directly perceived as is, but only through the observation of the results of some *experiments* or *tests* that have been performed on those states, as would be the case in a realistic scientific experiment. As such, the reaction system is intrinsically built on information that can be affected by different forms of uncertainty. Notably, while we will focus on the specific case of Reaction Systems, the methodology that we propose mainly considers the *graph of the dynamics* that underlies the model and thus, at least in principle, should be easily generalizable to any class of discrete dynamical systems.

The rest of this paper will be structured as follows: in Sect. 2 we recall the necessary background concerning both Reactions Systems, Rough Sets and their linking; while in Sect. 3 we present the mathematical framework that we propose. Finally, in Sect. 4 we discuss the obtained results and possible future research directions.

## 2 Mathematical Background

In this section, the basic notions on both reaction systems and rough sets are given.

### 2.1 Introduction to Reaction Systems

Reaction Systems are a model of computation inspired by biochemical reactions involving reactants, inhibitors and products from a finite background set.

**Definition 1.** *A Reaction System is an ordered pair  $(S, A)$  such that  $S$  is a finite set of substances or entities, and  $A$  is a set of reactions in  $S$ .*

A reaction can be formally defined as follows.

**Definition 2.** A reaction, in a reaction system  $(S, A)$ , is a triplet  $a = (R_a, I_a, P_a)$  where  $R_a \subseteq S$  is the set of reactants,  $I_a \subseteq S$  is the set of inhibitors, and  $P_a \subseteq S$  is the set of products.

The result of applying reaction  $a$  to a set  $X \subseteq S$ , denoted by  $res_a(X)$ , is conditional: if  $R_a$  is included in  $X$  and  $I_a$  is disjoint with  $X$ , then  $a$  is enabled on  $X$ , otherwise  $a$  is not enabled on  $X$ . If  $a$  is enabled on  $X$ , then  $a$  transforms the set of reactants into the product set. Thus, formally:

$$res_a(X) = \begin{cases} P_a & R_a \subseteq X \text{ and } I_a \cap X = \emptyset \\ \emptyset & \text{otherwise} \end{cases} \quad (1)$$

For a reaction system  $(S, A)$ , the result function of  $A$  is  $res_A : 2^S \rightarrow 2^S$ , and for each  $T \subseteq S$  it is defined as:

$$res_A(T) = \bigcup_{a \in en_A(T)} P_T \quad (2)$$

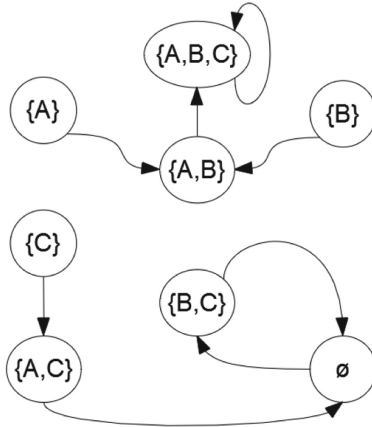
where  $en_A(T)$  is the set of reactions of  $A$  enabled in  $T$ .

Given a RS  $R = (S, A)$  the associated *graph of the dynamics* is the graph  $G[R] = (V, E)$  where  $V = 2^S$  and  $(v_1, v_2) \in E$  if  $res_A(v_1) = v_2$ . An example of a Reaction System is illustrated in Example 1.

**Example 1.** Let  $R = (S, A)$  be a Reaction System where:

- $S = \{A, B, C\}$ ;
- $A = \{(\emptyset, ABC, BC), (A, C, AB), (B, C, AB), (C, AB, AC), (AB, \emptyset, ABC)\}$ .

The graph of the dynamics of  $R$  is shown in Fig. 1.



**Fig. 1.** Graph of the dynamics for the Reaction System described in Example 1.

We refer the reader to [7] for a recent overview and tutorial on Reaction Systems.

## 2.2 Introduction to Rough Sets

Rough sets are an approach to imperfect knowledge proposed by Zdzilaw Pawlak to model uncertain and incomplete knowledge [15]. For recent overviews on Rough Set Theory and applications we refer the reader to [1,23]. The basic notion of Rough Set Theory is that of an information table.

**Definition 3.** An Information Table is an ordered pair  $IT = (U, Att)$  such that  $U$  is a finite non-empty set of objects and  $Att$  is a finite non-empty set of attributes, where each  $a \in Att$  is a function  $a : U \mapsto V_a$  and  $V_a$  is the set of possible values of  $a$ .

Given an  $IT$ , we say that two objects  $u, u'$  are *indiscernible* w.r.t.  $B \subseteq Att$  if  $\forall b \in B, b(u) = b(u')$ . Indiscernibility defines an equivalence relation where the equivalence class of an object  $u$  is denoted as  $[u]_B$ .

Given an Information Table  $IT = (U, Att)$  and  $B \subseteq Att$ , we can define for  $X \subseteq U$  its *rough approximation* (or, *rough set*) as  $B(X) = \langle l_B(X), u_B(X) \rangle$ , where  $l_B(X) = \bigcup_{[u]_B \subseteq X} [u]_B$  is the *lower approximation* of  $X$  and  $u_B(X) = \bigcup_{[u]_B \cap X \neq \emptyset} [u]_B$  is the respective *upper approximation*. We denote with  $\mathcal{R}_B(U)$  the set of rough sets on  $U$  determined by  $B \subseteq Att$ .

The lower approximation of a set consists of all the elements that *surely* belong to that set, while the upper approximation of a set is made of all the element that *possibly* belong to the set. The boundary region can be defined as  $Bnd(X) = u(X) \setminus l(X)$  and can be understood as the collection of elements whose belonging to the set is not certain.

Given an Information Table  $IT = (U, Att)$ , a *super-reduct* [21] is a subset of attributes  $R \subseteq Att$  such that  $\forall x, [x]_R = [x]_{Att}$ . A super-reduct  $R$  is a *reduct* if no subset of  $R$  is also a super-reduct. We denote by  $RED(IT)$  the set of reducts of  $IT$ , the *core* of an  $IT$  is defined as  $Co(IT) = \bigcap_{R \in RED(IT)} R$ .

Finally, we notice that sometimes the starting point for defining rough sets is a so-called approximation space  $(U, R)$ , with  $U$  a set of instances and  $R$  an equivalence relation (or, equivalently, a partition of  $U$ ). Thus, any Information Table induces an approximation space, which is a more general notion. The lower and upper approximations are, then, defined exactly as above.

## 2.3 Related Work on Linking Rough Sets and Reaction Systems

The importance of linking Rough Set Theory and Reaction Systems, with the goal of augmenting the formalism of Reaction Systems with notions of partial information and incompleteness, has been recognized in [6,18]. Intuitively, in these studies, the basic concept is that of a situation that could be understood as a state of the system under observation. Situations can only be perceived through attributes (that could represent physical experiments or other properties) and for the observed situations (which represents the objects in an Information Table) we are able to precisely tell whether a given substance was present or not in that situation. However, we can give a lower and an upper approximation of the present substances.

Formally, in this framework, the authors start from the substances  $s$  of a Reaction System  $R = (S, A)$  and, for each such substance, they define a Decision Table  $DT(s) = (U, Att_s, d_s)$ , that is an Information Table  $(U, Att_s)$  plus a decision  $d_s : U \mapsto \{0, 1\}$ , where  $U$  is a set of physical situations,  $Att_s$  are attributes through which the physical situation is perceived and  $d_s(u) = 1$  iff substance  $s$  is present in situation  $u$ . Then, the set of situations in which  $s$  is present is represented by the decision class  $D(s) = \{u | d_s(u) = 1\}$ . Since it can happen that the attributes  $Att_s$  do not carry enough information to take a clear decision, the decision class can be approximated via the information given by the attributes  $Att_s$  using the standard Rough Set notions of lower and upper approximations, thus defining,  $L(D(s))$  and  $U(D(s))$ . Then, the authors define how a state  $\hat{S}$  could be represented by aggregation of the decision systems  $DT(s)$  for  $s \in \hat{S}$ .

The approach that we take in the following is similar in spirit, in that we also take states as the basic notion of our framework and we assume that, in general, these states are not completely recognizable but only perceived via tests that affirm whether some substances are present or not in the current situation. A fundamental difference, however, relates to the fact that the decision attribute in the framework of [6, 18] can be seen as an a-priori notion that is independent of the attributes, in that it is already represented in the decision system. As we will see in the following sections, in the approach that we propose the decision w.r.t. a substance being present or not in a situation is only an a-posteriori notion that is entirely defined by the values of the attributes or, as we will call them, *tests*. Indeed, the result of the tests is the only information that we have about a state and we are able to state that a given substance  $s$  is present in a given situation only inasmuch the result of the tests is able to do so.

The notion of test that we will introduce resembles the notion of a *sensor* in complex dynamical systems [11]: both represent available information about the state of a complex system and, in both cases, one of the most interesting problem is related to finding a minimal and sufficient set of tests (resp. sensors) that are able to accurately *describe* the dynamics of the whole, partially unobservable system. The main differences between these two notions relate to the fact that: sensors are defined in the context of *classical* (i.e. based on dynamical systems theory), typically continuous, complex systems while the notion of test that we will introduce is based on *discrete* dynamical systems; furthermore, the underlying theory for minimal set of sensors are based on ideas from statistical mechanics, control theory and related disciplines, while the theory that we develop for tests is based on Rough Sets and graph theory.

### 3 Methods

As argued in Sect. 1, one of the main features of real complex systems which is lacking in the formalism of Reaction Systems is the ability to model partial or uncertain information about the states of the system. Further, a Reaction System is fully specified in terms of the reactions, while in reality the model

is usually construed via gradual observation of the behavior of the system. In this section we will formalize both concepts through application of Rough Set Theory to Reaction Systems. We assume that the dynamics of the complex system that we observe is fully described by an underlying Reaction System which, however, may be unknown. The goal is then to understand, given a certain set of experimental tests that we may perform, whether these tests are sufficient to accurately describe the dynamics of the system. In order to do so, in Sect. 3.1 we will formalize the notion of partial observability of a Reaction System through the notion of *Approximate Reaction System* and *tests*. Further we will consider the issue of dynamic acquisition of knowledge about a Reaction System, formalized via *observations*, that is states of partial knowledge about the graph of the dynamics of a Reaction System. In Sect. 3.2 we will describe reducts for Approximate Reaction Systems, their existence conditions and characterization.

### 3.1 Approximate Reaction Systems

**Definition 4.** An Approximate Reaction System (ARS) is a triple  $R = (S, A, T)$ , where  $S$  is the set of substances,  $A$  is the set of reactions and  $T$  is the set of tests. A test  $t \in T$  is a function  $t : S \mapsto \{\perp, \top\}$ , we denote with  $\text{supp}(t) = \{s \in S \mid t(s) = \top\}$  the support of  $t$ . The result of test  $t$  on state

$$X \subseteq S \text{ is } r_t(X) = \bigvee_{s \in X} t(s) = \begin{cases} \top & \text{supp}(t) \cap X \neq \emptyset \\ \perp & \text{otherwise} \end{cases} \quad (3)$$

**Definition 5.** We say that a test  $t$  identifies a substance  $s \in S$  if  $\text{supp}(t) = \{s\}$ . As all tests  $t$  that identify a given substance  $s$  are isomorphic, we will denote any such test as  $t_s$ .

Intuitively, a test represents a piece of information about the state of a Reaction System that tells an observer whether some given substances are present, or not, in the state. In particular a test is given a *disjunctive interpretation*, it is only able to tell us whether at least one (but not necessarily all) of the substances it tests for are present in the given state. The intuition for this definition derives from the concept of a *chemical test*, that is a qualitative or quantitative procedure designed to identify, quantify, or characterise a chemical compound or chemical group: so, a test that identifies a substance represents a chemical test that is able to precisely detect a single chemical compound (e.g. a test for blood sugar), while a chemical test for recognizing chemical groups, e.g. acids, can be represented by a general test. Then, an Approximate Reaction System represents the uncertain and partial knowledge that we have on the behaviour of a real underlying reaction system given that we are only able to observe its states through the tests specified by  $T$ .

We observe that a set of tests  $T$  defines an indiscernibility partition of the states:

$$X \sim_T Y \text{ iff } \forall t \in T, t(X) = t(Y) \quad (4)$$

We denote by  $[X]_T$  the equivalence class of state  $X \subseteq 2^S$  determined by the set of tests  $T$ . Thus, it follows that the set of test determines an approximation space  $(2^S, \sim_T)$ . The rough approximations of the states are formally defined as follows: let  $X \subseteq 2^S$  be a state, then, its *rough approximation* determined by  $T$  is given by  $r(X) = \langle l(X), u(X) \rangle$  where

$$l(X) = \bigcap_{Y \in [X]_T} Y \quad (5)$$

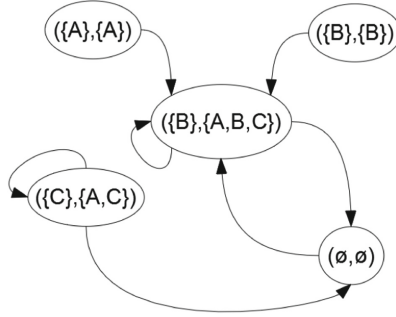
$$u(X) = \bigcup_{Y \in [X]_T} Y \quad (6)$$

Given an ARS  $R = (S, A, T)$ , the associated graph of the dynamics is the graph  $G[R] = (V_T, E_T)$  where:

- $V_T = \mathcal{R}_T(2^S)$ ;
- $(v_1 = \langle l(X), u(X) \rangle, v_2 = \langle l(Y), u(Y) \rangle) \in E_T$  iff  $\exists l(X) \subseteq W \subseteq u(X), l(Y) \subseteq Z \subseteq u(Y)$  s.t.  $(W, Z)$  is an edge in the graph of the non-approximated reaction system.

An example of an ARS and its associated graph of the dynamics is shown in Example 2.

**Example 2.** Let  $R$  be a ARS  $R = (S, A, T)$ : where  $S$  and  $A$  are as detailed in Example 1, while  $T = \{t_1, t_2, t_3\}$  where  $\text{supp}(t_1) = \{B\}$ ,  $\text{supp}(t_2) = \{A, C\}$  and  $\text{supp}(t_3) = \{B, C\}$ . Figure 2 illustrates the related graph of the dynamics.



**Fig. 2.** Graph of the dynamics for the ARS in Example 2.

We notice that this graph features a form of non-determinism as, for example, there are multiple outgoing arcs from the node labeled  $\langle B, ABC \rangle$ . We notice, furthermore, that the state where only substance  $B$  is distinct from the state  $\langle B, ABC \rangle$  even though the lower approximation of the latter one is exactly  $B$ .

As illustrated in Example 2, one can observe that, in general, the incompleteness and uncertainty determined by the fact that the result of tests is the

only information available about the states, the resulting graph of the dynamics could feature a form of *non-determinism*, while the graph of the dynamics of a standard Reaction System is necessarily deterministic. This suggests that, having fixed a set of tests, if we observe the evolution of a system and we derive that the resulting graph of the dynamics is non-deterministic then, the employed test are not sufficient to properly describe the system (at least, if we assume that the underlying phenomenon *could* be modeled as a Reaction System).

While tests formalize the notion of partial observability in terms of the substances, they do not provide a formalization of the idea that, in general, knowledge about a complex system is acquired iteratively by repeatedly observing its evolution over time from an initial state. We formalize this other notion via *observations*:

**Definition 6.** Given the graph  $G[R] = (V_T, E_T)$  of an ARS  $R$ , we denote as  $G[R]_x = (V_x, E_x)$  the set of all maximal paths starting from  $x$ .

We say that an observation of an ARS is a collection  $\mathcal{O}(R) = \{G[R]_{x_1}, \dots, G[R]_{x_n}\}$  for  $x_1, \dots, x_n \in V_T$ . We denote with  $V_{\mathcal{O}(R)}, E_{\mathcal{O}(R)}$ , respectively, the set of nodes and edges in  $\mathcal{O}(R)$ .

Given an observation of an ARS we can define the respective Information Table as:

**Definition 7.** An ARS Information Table  $I[\mathcal{O}(R)]$  for an observation  $\mathcal{O}(R)$  of an ARS  $R = (S, A, T)$  is an ordered pair  $(U, T)$ , where  $U = \bigcup_{G[R]_x \in \mathcal{O}(R)} V_x$ .

Thus an ARS Information Table represents two different types of partial, incomplete information about an underlying Reaction System: first, the incompleteness of information w.r.t. the global dynamics of the Reaction System as only the dynamics involving the states under observation is known; second, the incompleteness of information w.r.t. the states, as these are only observed through the set of tests that are performed.

We notice that the definition of *identifies* that we previously defined applies only to single tests. In order to generalize this notion we would need to consider set of tests. Intuitively a set of tests  $F$  identifies  $s$  if we know with certainty that, given a state  $X$ , if  $\exists t \in F, t(X) = \top$  then  $s \in X$ .

Formally,

**Definition 8.** A set of tests  $F \subseteq T$  identifies  $s \in S$  if

$$\begin{aligned} \forall X \subseteq 2^S, \exists t \in F \text{ s.t. } s \in \text{supp}(t) \text{ and } t(X) = \top \wedge \\ \forall s' \in \text{supp}(t) \setminus \{s\}, \exists t' \neq t \in F \text{ s.t. } s' \in \text{supp}(t') \wedge t'(X) = \perp. \end{aligned} \quad (7)$$

This notion allows to define an alternative formulation of lower and upper approximations that is not explicitly based on the equivalence relation on the states:

**Definition 9.** The lower and upper approximation defined by the relation  $T$  identifies  $s$  are, respectively:

$$l'(X) = \{s \in S \mid \forall t \in T \text{ s.t. } s \in \text{supp}(t), t(X) = \top \wedge T \text{ identifies } s\} \quad (8)$$

$$u'(X) = \{s \in S \mid \forall t \in T \text{ s.t. } s \in \text{supp}(t), t(X) = \top\} \quad (9)$$



Then, we can prove the following result, that states that the two alternative formulations are equivalent:

**Lemma 1.**  $l(X) = l'(X)$  and  $u(X) = u'(X)$ .

*Proof.* Consider first the upper approximation  $u(X)$ : by Definition 6 a substance  $s$  is in  $u(X)$  iff  $\exists Y \in [X]_T. s \in Y$ . Thus,  $\forall t \in T. s \in \text{supp}(t), t(Y) = \top$ ; but, by definition of  $\sim_T$  this also means that  $t(X) = \top$ . Since  $s$  was arbitrary we can see that Definition 6 implies Definition 8. For the converse we can consider two cases:

1.  $s \in X$ , then obviously Definition 6 holds;
2.  $s \notin X$  but  $\forall t \in T$  s.t.  $s \in \text{supp}(t), t(X) = \top$ . Let  $Y = X \cup \{s\}$ , then evidently  $Y \in [X]_T$  but this means that Definition 6 follows.

As regards the lower approximation, we showed that the first part of the Definition characterizes the substances that are in the upper approximation, then we must show that the condition that  $T$  identifies  $s$  is necessary and sufficient for saying that  $s$  is also in the lower approximation. Let us assume that  $T$  identifies  $s$ , and  $t$  be the test that satisfies the condition for state  $X$ . Similarly for each substance  $s'$  let  $t^{s'}$  be the test s.t.  $s' \in \text{supp}(t^{s'})$  and  $t^{s'}(X) = \perp$ . Then if  $\text{supp}(t) = \{s\}$  the implication obviously follows, so let us focus on the case where  $\{s\} \subset \text{supp}(t)$ . Consider the equivalence class  $[X]_T$ , then evidently  $\forall Y \in [X]_T, t(Y) = \top$  and  $t^{s'}(Y) = \perp$  which means that  $s' \notin Y$  and since this holds  $\forall s' \neq s \in \text{supp}(t)$  is must hold that  $s \in Y$ , so Definition 5 follows.

For the converse, notice that if Definition 5 holds then if  $s \in l(X)$ , then  $\exists t \in T. s \in \text{supp}(t) \wedge t(Y) = \top$ , otherwise there would be a state  $Z = Y \setminus \{s\}$  with both  $Z, Y \in [X]_T$ . If  $\{s\} = \text{supp}(t)$  then Definition 8 follows. On the contrary, consider  $s' \neq s \in \text{supp}(t)$  such that  $\exists t' \neq t$  with  $s' \in \text{supp}(t')$ . If  $\exists Y$  s.t.  $t'(Y) = \perp$  then we are done. Otherwise we can notice that such a couple  $t, t'$  must exist otherwise it must exist  $Z, Y \in [X]_T$  s.t.  $Z = (Y \setminus \{s\}) \cup \{s'\}$  but this is an absurd as we assumed that  $s \in l(X)$ . Thus Definition 8 really is a characterization of lower approximations.

In the following section we will define the concept of reduct for an ARS.

### 3.2 Reducts

Given an  $ARS = (S, A, T)$ , we may ask whether the given set of tests is sufficient to describe the dynamics that we could observe, if we had been able to fully observe the states of the Reaction System. More in general, one may assume that the provided set of tests is all we can have (e.g., for a certain situation the provided tests are the most precise and powerful that are known) to describe the Reaction System: in this case we can ask whether all the tests available are necessary or there exists some test that is redundant.

Both these two concepts correspond to the idea of a reduct in Rough Set Theory.

**Definition 10.** Given an ARS  $R_T = (S, A, T)$  and an observation  $\mathcal{O}(R_T)$ , for a given  $F \subseteq T$ , we define  $R_F = (S, A, F)$ . Then, we say that  $F$  is a:

- complete super-reduct if  $G[R_F] = G[(S, A)]$ ;
- relative super-reduct if  $G[R_F] = G[R_T]$ ;
- weak super-reduct if  $G[R_F]_{|\mathcal{O}(R_T)} = G[R_T]_{|\mathcal{O}(R_T)}$

where  $G[R_F]_{|\mathcal{O}(R_T)}$  is the restriction of  $G[R_F]$  to  $\mathcal{O}(R_T)$ . We say that  $F$  is a complete (resp. relative, weak) reduct if it is a complete (resp. relative, weak) super-reduct and it is minimal w.r.t. this property.

The following result characterizes (complete, relative, weak) reducts in terms of Information Tables:

**Proposition 1.** Let  $R_T = (S, A, T)$  be an ARS and  $\mathcal{O}(R_T)$  an observation. Then  $F \subseteq T$ :

- is a weak reduct iff it is a reduct for  $I[\mathcal{O}(R_T)]$ ;
- is a relative reduct iff it is a reduct for  $I[R_T] = (2^S, T)$ ;
- is a complete reduct iff it is a reduct for  $I^*[R_T] = (2^S, T \cup \{t_s | s \in S\})$ , where  $t_s$  is a test that identifies  $s$ .

*Proof.* The case of weak reducts follows directly from Definition 10.

The condition for relative reducts is equivalent to saying that  $\mathcal{R}_T(2^S) = \mathcal{R}_F(2^S)$ , that is the set of rough sets of states are the same when considering the full set of tests or the reduct  $F$ . We can notice that an equivalent condition for  $F$  being a relative reduct would be being a reduct for the ARS information system  $I[R_T] = (\{[X]_T : X \subseteq 2^S\}, T)$  in which the equivalence classes determined by  $T$  are made explicit.

On the other hand, the condition for complete reducts states that  $F$  must be able to identify all the substances  $s \in S$ .  $\square$

**Corollary 1.** The smallest complete reduct  $RED_{min}$  of an ARS where  $\forall s \in S \exists t_s \in T$  has  $|RED_{min}| = |S|$ .

We notice that while the definition of (complete, relative, weak) reducts suggests an algorithm for checking whether  $F \subseteq T$  is a reduct (e.g. by constructing the discernibility matrix for the corresponding Information Table), the time complexity of this algorithm is linear in the size of the ARS Information Table but, in general, exponential in  $|S|$ . We can see from Corollary 1 that, at least for complete reducts, a simple algorithm for finding reducts (and hence for testing them) when we restrict to the case where  $\forall s \in S, t_s \in T$  and that operates in time linear in  $|T|$  can be given.

A different, but equivalent, characterization of reducts can be formulated in terms of the *identifies* relation defined in Sect. 3.1.

**Theorem 1.** Let ARS  $= (S, A, T)$  be an ARS and let  $S_T = \{s \in S | T \text{ identifies } s\}$ . Then  $F \subseteq T$ :

- is a complete super-reduct iff  $S = S_F$ ;
- is a relative super-reduct iff  $S_T = S_F$ .

*Proof.* This follows from the fact that if the condition holds then  $\forall X$  the lower and upper approximations remain equal.

Notice that while Theorem 1 and Proposition 1 are equivalent characterizations, the former result suggests an algorithm for testing reducts whose runtime is  $O(|S|^2|T|^2)$ . Algorithm 1 describes the algorithm for the case of complete reducts, the case for relative reducts is equivalent. The consequence of this result is that the problem of finding complete and relative reducts is in *NP* not only when considering the graph of the dynamics as the size of the problem, but also when considering the size of the Reaction System. As finding reducts in general Information Tables is *NP*-complete [20], we conjecture that the problem of finding (complete, relative) reducts lies in the same complexity class.

---

**Algorithm 1.** A polynomial-time algorithm for the verification of complete reducts.

---

**procedure** CHECK-COMplete-REDUCT

**Require:**  $R = (S, A)$  Reaction System,  $F$  a reduct

$check \leftarrow \top$

**for all**  $s \in S$  **do**

**for all**  $f \in F : s \in \text{supp}(f)$  **do**

$temp \leftarrow \top$

**for all**  $s' \neq s \in \text{supp}(f)$  **do**

$temp \leftarrow temp \wedge \exists f' \neq f. s' \in \text{supp}(f') \wedge s \notin \text{supp}(f')$

**end for**

**if**  $temp = \top$  **then**

$check \leftarrow \top$

        Break

**else**

$check \leftarrow \perp$

**end if**

**end for**

**end for**

**Return**  $check$

**end procedure**

---

Notice that while a similar characterization could be given also for checking weak reducts, in that case the complexity would still be polynomial w.r.t. the number of states in the *observation*, hence, in the worst case, exponential in the number of substances.

## 4 Conclusion

In this paper we considered the study of mathematical methods to model complex systems, focusing on the formalism of Reaction Systems, in particular with

respect to their ability to model incomplete and partial information. As these characteristics are commonly represented through Rough Set Theory, and also acknowledging a recent research direction towards the linking of Reaction Systems and Rough Sets, we developed a mathematical framework, based on core Rough Set theoretic concept to study these issues. We introduced the notion of partial observability of the states of a Reaction System, through the notion of tests, and after observing that this induces an approximation space we applied ideas from Rough Set Theory to define lower and upper approximations; reducts that could be used to automatically model Reaction Systems based on (potentially uncertain and incomplete) observations. In order to further the applications of Rough Set Theory to the study of complex systems, we think that the following open problems may be of interest:

- We provided a characterization of complete and relative reducts based on tests and their ability to identify the substances. This characterization suggests that the problem of finding (complete, relative) reducts is in  $NP$  not only w.r.t. the size of the graph of the dynamics (which is in general exponential in the number of substances) but also w.r.t. the size of the Reaction System. Similar characterizations for weak reducts would be interesting;
- We considered reducts as sets of tests that are able to represent, without loss of information, the graph of the dynamics of the Reaction System (or Approximate Reaction System). It is not hard, however, to observe that this definition may be too restrictive: indeed, if one's interest only concerns the general dynamics of a system, then an approximated graph may be tolerable as long as it has the same properties of the original graph (e.g., w.r.t. the reachability of states). It would then be interesting to give a definition of reducts that characterizes this property of invariance w.r.t. the satisfaction of properties expressed in a given logic [2];
- While in this work we considered approximations and reducts, Rough Set Theory also encompasses methods for *rule induction* [9, 17] in order to *explain* a Decision Table via sets of rules. Applying these approaches in the context of Approximate Reaction Systems and observations (and, more in general, complex systems) could enable the interactive and iterative learning and updating of Reaction System models [12] based on observed dynamics;
- Finally, while the present work applies to Reaction Systems, we argued that, as the proposed methods mainly use the graph of the dynamics, these notions could also be extended to other discrete complex systems formalisms: in order to do so, appropriate definitions of tests should be considered.

## References

1. Akama, S.: Topics in Rough Set Theory: Current Applications to Granular Computing. Intelligent Systems Reference Library, vol. 168. Springer, Cham (2020). <https://doi.org/10.1007/978-3-030-29566-0>
2. Azimi, S., Gratie, C., Ivanov, S., Manzoni, L., Petre, I., Porreca, A.E.: Complexity of model checking for reaction systems. Theoret. Comput. Sci. **623**, 103–113 (2016)

3. Bello, R., Falcon, R.: Rough sets in machine learning: a review. In: Wang, G., Skowron, A., Yao, Y., Ślęzak, D., Polkowski, L. (eds.) *Thriving Rough Sets*. SCI, vol. 708, pp. 87–118. Springer, Cham (2017). [https://doi.org/10.1007/978-3-319-54966-8\\_5](https://doi.org/10.1007/978-3-319-54966-8_5)
4. Bergstra, J.A., Ponse, A., Smolka, S.A.: *Handbook of Process Algebra*. Elsevier, Amsterdam (2001)
5. Chopard, B., Droz, M.: *Cellular Automata*. Springer, Dordrecht (1998). <https://doi.org/10.1007/978-94-015-9153-9>
6. Dutta, S., Jankowski, A., Rozenberg, G., Skowron, A.: Linking reaction systems with rough sets. *Fundamenta Informaticae* **165**(3–4), 283–302 (2019)
7. Ehrenfeucht, A., Petre, I., Rozenberg, G.: Reaction systems: a model of computation inspired by the functioning of the living cell. In: Konstantinidis, S., Moreira, N., Reis, R., Shallit, J. (eds.) *The Role of Theory in Computer Science: Essays Dedicated to Janusz Brzozowski*, pp. 1–32. World Scientific, Singapore (2017)
8. Ehrenfeucht, A., Rozenberg, G.: Reaction systems. *Fundamenta informaticae* **75**(1–4), 263–280 (2007)
9. Hu, M., Yao, Y.: Structured approximations as a basis for three-way decisions in rough set theory. *Knowl. Based Syst.* **165**, 92–109 (2019)
10. Lin, T.Y., Cercone, N.: *Rough Sets and Data Mining: Analysis of Imprecise Data*. Springer, Boston (2012). <https://doi.org/10.1007/978-1-4613-1461-5>
11. Liu, Y.Y., Slotine, J.J., Barabási, A.L.: Observability of complex systems. *Proc. Natl. Acad. Sci.* **110**(7), 2460–2465 (2013)
12. Męski, A., Koutny, M., Penczek, W.: Reaction mining for reaction systems. In: Stepney, S., Verlan, S. (eds.) *UCNC 2018*. LNCS, vol. 10867, pp. 131–144. Springer, Cham (2018). [https://doi.org/10.1007/978-3-319-92435-9\\_10](https://doi.org/10.1007/978-3-319-92435-9_10)
13. Orłowska, E.: *Incomplete Information: Rough Set Analysis*. Studies in Fuzziness and Soft Computing, vol. 13. Physica, Heidelberg (2013). <https://doi.org/10.1007/978-3-7908-1888-8>
14. Păun, G., Rozenberg, G.: A guide to membrane computing. *Theoret. Comput. Sci.* **287**(1), 73–100 (2002)
15. Pawlak, Z.: Rough sets. *Int. J. Comput. Inf. Sci.* **11**(5), 341–356 (1982)
16. Reisig, W.: *Petri Nets: An Introduction*, vol. 4. Springer, Heidelberg (2012)
17. Sakai, H., Nakata, M.: Rough set-based rule generation and apriori-based rule generation from table data sets: a survey and a combination. *CAAI Trans. Intell. Technol.* **4**(4), 203–213 (2019)
18. Skowron, A., Dutta, S.: Rough sets: past, present, and future. *Nat. Comput.* **17**(4), 855–876 (2018). <https://doi.org/10.1007/s11047-018-9700-3>
19. Skowron, A., Jankowski, A., Dutta, S.: Interactive granular computing. *Granul. Comput.* **1**(2), 95–113 (2016)
20. Skowron, A., Rauszer, C.: The discernibility matrices and functions in information systems. In: Słowiński, R. (ed.) *Intelligent Decision Support*, pp. 331–362. Springer, Dordrecht (1992). [https://doi.org/10.1007/978-94-015-7975-9\\_21](https://doi.org/10.1007/978-94-015-7975-9_21)
21. Ślęzak, D., Dutta, S.: Dynamic and discernibility characteristics of different attribute reduction criteria. In: Nguyen, H.S., Ha, Q.-T., Li, T., Przybyła-Kasperek, M. (eds.) *IJCRS 2018*. LNCS (LNAI), vol. 11103, pp. 628–643. Springer, Cham (2018). [https://doi.org/10.1007/978-3-319-99368-3\\_49](https://doi.org/10.1007/978-3-319-99368-3_49)
22. Thangavel, K., Pethalakshmi, A.: Dimensionality reduction based on rough set theory: a review. *Appl. Soft Comput.* **9**(1), 1–12 (2009)
23. Wang, G., Skowron, A., Yao, Y., Ślęzak, D., Polkowski, L. (eds.): *Thriving Rough Sets*. SCI, vol. 708. Springer, Cham (2017). <https://doi.org/10.1007/978-3-319-54966-8>