



HAL
open science

Polyadic Relational Concept Analysis

Alexandre Bazin, Jessie Galasso-Carbonnel, Giacomo Kahn

► **To cite this version:**

Alexandre Bazin, Jessie Galasso-Carbonnel, Giacomo Kahn. Polyadic Relational Concept Analysis. 2022. hal-03908993v1

HAL Id: hal-03908993

<https://hal.science/hal-03908993v1>

Preprint submitted on 21 Dec 2022 (v1), last revised 27 Oct 2023 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Polyadic Relational Concept Analysis

Bazin, Alexandre¹, Galasso-Carbonnel, Jessie², and Kahn, Giacomo³

¹ LIRMM, CNRS, Université de Montpellier, FRANCE
alexandre.bazin@u-montpellier.fr

² DIRO, Université de Montréal, CANADA
jessie.galasso-carbonnel@umontreal.ca

³ Univ Lyon, Univ Lyon 2, Université Claude Bernard Lyon 1, Université Jean Monnet Saint-Etienne, INSA Lyon, DISP-UR4570, 69500 Bron, FRANCE
giacomo.kahn@univ-lyon2.fr

Abstract. Formal concept analysis is a mathematical framework based on lattice theory that aims at representing the information contained in binary object-attribute datasets (called formal contexts) in the form of a lattice of so-called formal concepts. Since its introduction, it has been extended to more complex types of data. In this paper, we are interested in two of those extensions: relational concept analysis and polyadic concept analysis that allow to process, respectively, relational data and n -ary relations. We present a framework for polyadic relational concept analysis that extends relational concept analysis to relational datasets that are made of n -ary relations. We show its basic properties and that it is a valid extension of relational concept analysis.

1 Introduction

Formal concept analysis (FCA [6]) is a mathematical framework based on lattice theory that aims at representing the information contained in binary object-attribute datasets (called formal contexts) in the form of a lattice of so-called formal concepts. Formal contexts being representations of the most basic type of data, direct applications of formal concept analysis on real, more complex data are limited but various extensions exist to exploit this powerful formalism. In this paper, we are interested in merging two of them : relational concept analysis and polyadic concept analysis.

Relational concept analysis (RCA [7]) is an extension of the formal concept analysis framework aimed at relational data that take the form of multiple formal contexts with relations between their objects. The idea is to represent through attributes the information contained in the relations, hereby enriching the descriptions of objects in the formal contexts. RCA has been used with great success in fields as diverse as software engineering, for variability modeling or feature location [2,8], or the analysis of ecological data about plants [12] or water quality [3].

Polyadic concept analysis (PCA [14]) is another extension of the formal analysis framework aimed at multidimensional data, i.e. n -ary relations between

elements of n dimensions. For instance, *objects* described by *attributes* under *conditions* in the tridimensional case. PCA and its tridimensional version, triadic concept analysis [11], have been used for data exploration with multiple experts [4], or tri-clustering in machine learning [9]. RDF datasets are also important examples of multidimensional data and, with them, knowledge graphs. Such data rarely come alone and are generally part of a larger dataset containing relations. It would thus seem natural to combine RCA and PCA. However, as they are, these two frameworks are not compatible. In a triadic object-attribute-condition setting, if we kept representing the information contained in relations by adding elements to the set of attributes, under which condition would a newly created attribute describe an object? To solve this problem, we propose a new framework, Polyadic Relational Concept Analysis (PRCA), that works on sets of n -dimensional contexts with relations between their objects. PRCA is different from Graph-FCA [5], that also allows one to manipulate a set of n -ary relations, since a graph-FCA concept is a graph structure that is not a formal concept of the input formal context, while PRCA outputs concept n -lattices.

This paper presents the PRCA framework and some of its properties, including the fact that applying PRCA to a bidimensional formal context yields the same results as RCA, proving PRCA is a proper generalisation of RCA. Section 2 recalls the necessary background notions of formal concept analysis, relational concept analysis and polyadic concept analysis. Section 3 contains the definition of PRCA and an example of its application to a small toy dataset. Section 4 presents a proof of the fact that PRCA produces concepts that are in a one-to-one correspondence with concepts of an n -context of the same dimensionality as the initial context on which it is applied, bounding the size of the output and implying that PRCA is a proper generalisation of RCA.

2 Background Notions

2.1 Formal Concept Analysis

Formal concept analysis [6] (FCA) is a mathematical framework that aims at using lattices to organise the units of knowledge that can be found in binary data.

Definition 1 (Formal context). *A formal context is a triple $(\mathcal{O}, \mathcal{A}, \mathcal{I})$ in which \mathcal{O} is a set of objects, \mathcal{A} is a set of attributes and $\mathcal{I} \subseteq \mathcal{O} \times \mathcal{A}$ is a binary relation between objects and attributes.*

Formal contexts (or just *contexts*) formalise binary datasets and can be represented by a crosstable, as illustrated in Figure 1, that has as objects plants, and attributes pests that are repelled by them.

Definition 2 (Derivation operators). *Let $(\mathcal{O}, \mathcal{A}, \mathcal{I})$ be a formal context. The operators*

$$.' : 2^{\mathcal{O}} \mapsto 2^{\mathcal{A}}$$

	Spodoptera (s)	Milax gagates (m)	Tetranychidae (t)	Locust (ℓ)
Citrus (C)	×	×		
Alpinia (A)			×	×
Pelargonium (P)		×	×	
Laphangium (L)			×	×

Fig. 1: A formal context describing four plants as objects (Citrus, Alpinia, Pelargonium and Laphangium) through four attributes (s, m, t, ℓ) representing the pest that are repelled by each of the plants. This particular context, while not physically accurate, was inspired by project Knomana (<https://agritrop.cirad.fr/591139/>) [10].

$$A' = \{o \in \mathcal{O} \mid \forall a \in A, (o, a) \in \mathcal{I}\}$$

and

$$.' : 2^{\mathcal{A}} \mapsto 2^{\mathcal{O}}$$

$$O' = \{a \in \mathcal{A} \mid \forall o \in \mathcal{O}, (o, a) \in \mathcal{I}\}$$

are called the derivation operators of the formal context.

The two derivation operators of a formal context form a *Galois connection* and, as such, their compositions $.'$ are closure operators, i.e. $X \subseteq X''$, $(X'')' = X'$ and if $X \subseteq Y$ then $X'' \subseteq Y''$. Sets X such that $X = X''$ are said to be *closed*.

Definition 3 (Formal concept). Let $\mathcal{C} = (\mathcal{O}, \mathcal{A}, \mathcal{I})$ be a formal context. A formal concept of \mathcal{C} is a pair $(E, I) \in 2^{\mathcal{O}} \times 2^{\mathcal{A}}$ such that $E = I'$ and $I = E'$. The set E is called the *extent* and the set I the *intent* of the concept.

Formal concepts are by definition pairs of closed sets. They represent interesting units of knowledge, or “classes”, found in the data. They correspond to maximal rectangles of crosses in the crosstable representation of the formal context (up to permutation of rows and columns).

Definition 4 (Formal concept, alternative definition). Let $\mathcal{C} = (\mathcal{O}, \mathcal{A}, \mathcal{I})$ be a formal context. A formal concept of \mathcal{C} is a pair $(E, I) \in 2^{\mathcal{O}} \times 2^{\mathcal{A}}$ such that $E \times I \subseteq \mathcal{I}$ and both E and I are maximal for this property.

The set of formal concepts of a formal context \mathcal{C} is denoted $\mathcal{T}(\mathcal{C})$. Formal concepts can be ordered by the inclusion relations on their extents or intents: $(E_1, I_1) \leq_{\mathcal{O}} (E_2, I_2) \Leftrightarrow E_1 \subseteq E_2$ or $I_2 \subseteq I_1$. The pair $(\mathcal{T}(\mathcal{C}), \leq_{\mathcal{O}})$ is a complete lattice [6] called the *concept lattice* of the formal context \mathcal{C} .

2.2 Relational Concept Analysis

Relational concept analysis (RCA [7]) is a generalisation of FCA to relational binary datasets.

Definition 5 (Relational context family (RCF)). A relational context family (RCF) is a pair (\mathbb{C}, \mathbb{R}) such that:

- $\mathbb{C} = \{\mathcal{C}_i = (\mathcal{O}_i, \mathcal{A}_i, \mathcal{I}_i)\}$ is a set of formal contexts (object-attribute relations)
- $\mathbb{R} = \{r_k \mid r_k \subseteq \mathcal{O}_i \times \mathcal{O}_j\}$ is a set of relations between objects of source contexts \mathcal{C}_i and those of target contexts \mathcal{C}_j .

We use $\text{src}(r)$ and $\text{tar}(r)$ to denote, respectively, the source and target context of a relation r .

Relational context families formalise relational binary datasets. Both the contexts in \mathbb{C} and the relations in \mathbb{R} can be represented as binary crosstables but, for the sake of clarity, we shall only depict contexts as crosstable and relations as sets of pairs, as illustrated in Figure 2. Figure 3 depicts the concept lattices obtained from the Figure 2 RCF.

Plants	Spodoptera (s)	Milax gagates (m)	Tetranychidae (t)	Locust (ℓ)
Citrus (C)	×	×		
Alpinia (A)			×	×
Pelargonium (P)		×	×	
Laphangium (L)			×	×

Areas	Forests (F)	Desert (D)	Grassland (G)
Europe (Eu)	×	×	
Africa (Af)	×	×	
Asia (As)	×		×
America (Am)	×	×	×

$$\text{GrowsIn} = \{(C, \text{Eu}), (C, \text{Af}), (A, \text{Eu}), (A, \text{Am}), (P, \text{As}), (L, \text{Eu})\}$$

Fig. 2: A RCF containing two contexts (our plant context and a context about areas and ecosystems), and a relation between the plants and the areas, stating which plant grows in which area.

RCA aims at finding units of knowledge in the data, in the same fashion as FCA, by taking into consideration not only objects and their attributes but also their relations with other objects. In order to do so, information on the relations is converted into attributes, enriching the descriptions of objects.

We use $r(o)$ to denote the *image* of an object $o \in \mathcal{O}_i$ by a relation $r \subseteq \mathcal{O}_i \times \mathcal{O}_j$.

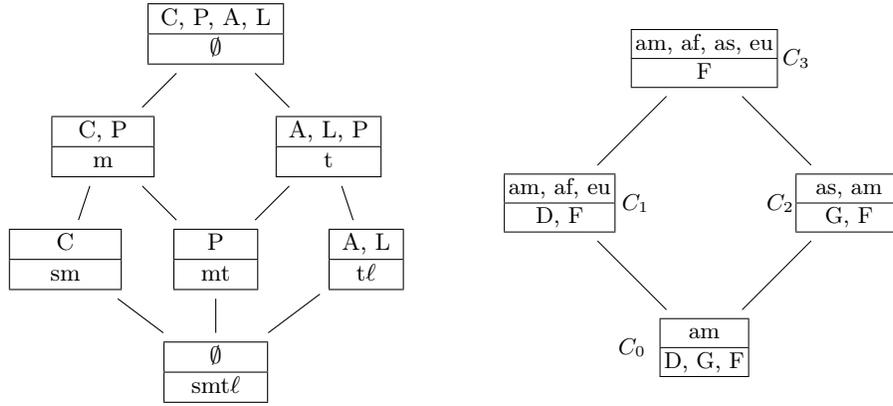


Fig. 3: Concepts lattices of the *Plants* (left) and *Areas* (right) contexts depicted in Figure 2.

Definition 6 (Relational scaling of a context). Let (\mathbb{C}, \mathbb{R}) be an RCF, $\mathcal{C}_i = (\mathcal{O}_i, \mathcal{A}_i, \mathcal{I}_i)$ a context in \mathbb{C} , R the set of relations in \mathbb{R} for which \mathcal{C}_i is the source and \mathcal{Q} a set of scaling operators. In this paper, we will only consider $\mathcal{Q} \subseteq \{\exists, \forall\}$. The formal context $\mathcal{C}_i^+ = (\mathcal{O}_i, \mathcal{A}_i^+, \mathcal{I}_i^+)$ is defined such that

$$\mathcal{A}_i^+ = \mathcal{A}_i \cup \{qr.X \mid q \in \mathcal{Q}, r \in R \text{ and } X \in \mathcal{T}(\text{tar}(r))\}$$

and

$$\mathcal{I}_i^+ = \mathcal{I}_i \cup \{(o, qr.X) \mid X = (E, I) \text{ and } E \text{ respects } q \text{ w.r.t. } r(o)\}$$

We write $\mathbb{C}^+ = \{\mathcal{C}_i^+ \mid \mathcal{C}_i \in \mathbb{C}\}$.

The computation of $(\mathbb{C}^+, \mathbb{R})$ requires the formal concepts of all the contexts in \mathbb{C} that are targets of relations. For this reason, the RCA process uses two steps iteratively:

- Compute all the concepts in all the contexts in \mathbb{C} that are targets of relations
- Construct $(\mathbb{C}^+, \mathbb{R})$ by using the concepts to generate new attributes.

We use $\mathbb{C}^0 = \mathbb{C}$ and $\mathbb{C}^m = \mathbb{C}^{(m-1)+}$ to denote the state of the contexts at step 0 and m , respectively. The process stops either when a user-defined number of iterations has been reached or when the concept lattices of every contexts do not change between \mathbb{C}^m and \mathbb{C}^{m+1} , which happens at some point [7].

Figs. 4 and 5 depict the result of the first (and final) step of the RCA process on the Figure 2 RCF. Four new relational attributes, corresponding to the four concepts in the lattice on the right-hand side of Fig 3, have been added to the *Plants* context. In the lattice, let us consider the concept $(\{A, P\}, \{t, \exists \text{Gr}C_2, \exists \text{Gr}C_3\})$. It expresses the fact that *Alpinia* (*A*) and *Pelargonium* (*P*) not only repel *tetranychidae* (*t*), but also *grow in* areas that correspond to C_2 and C_3 , that is areas that have either forests, or grasslands and forests as ecosystems.

Plants ¹	s	m	t	ℓ	$\exists\text{Gr}C_0$	$\exists\text{Gr}C_1$	$\exists\text{Gr}C_2$	$\exists\text{Gr}C_3$
Citrus (C)	×	×				×		×
Alpinia (A)		×	×	×		×	×	×
Pelargonium (P)	×	×					×	×
Laphangium (L)		×	×			×		×

Fig. 4: The context \mathcal{C}^+ , where \mathcal{C} is the plants context depicted in Figure 2 and $\mathcal{Q} = \{\exists\}$. The new attributes $\exists\text{GrowsIn}.C_i$ (abbreviated $\exists\text{Gr}C_i$) come from the four concepts $C_0 = (\{Am\}, \{D, G, F\})$, $C_1 = (\{Am, Af, Eu\}, \{D, F\})$, $C_2 = (\{As, Am\}, \{G, F\})$ and $C_3 = (\{Am, Af, As, Eu\}, \{F\})$.

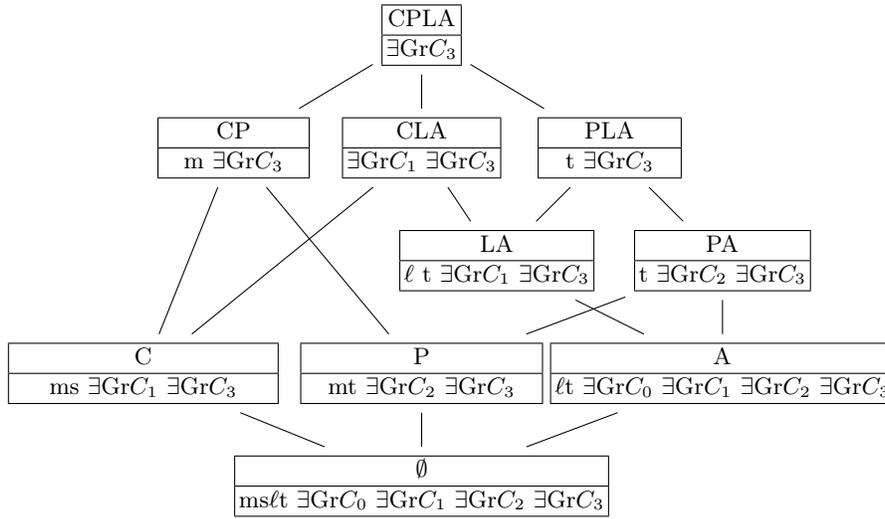


Fig. 5: Concept lattice of Plants^1 . Applying the same process for another step produces a concept lattice that is isomorphic to this one. Thus, the RCA process ends.

2.3 Polyadic Concept Analysis

Polyadic concept analysis [14] (PCA) is the generalisation of FCA to datasets that take the form of an n -ary relation.

Definition 7 (n -contexts). An n -context is an $(n + 1)$ -tuple $(\mathcal{D}_1, \dots, \mathcal{D}_n, \mathcal{I})$ in which all the \mathcal{D}_i are sets called dimensions and $\mathcal{I} \subseteq \prod_{i \in \{1, \dots, n\}} \mathcal{D}_i$ is an n -ary relation between dimensions.

Similarly to the bidimensional case, n -contexts formalise n -dimensional binary datasets. They can be represented as n -dimensional crosstables as illustrated in Figure 6.

Formal contexts are 2-contexts.

Plants	s	m	t	ℓ	s	m	t	ℓ
Citrus (C)	×	×			×	×		
Alpinia (A)			×	×			×	×
Pelagonium (P)	×	×					×	×
Laphangium (L)			×	×			×	×
				Root				Leaf

Fig. 6: A 3-context representing plants (Citrus, Alpinia, Pelagonium and Laphangium) that repel pests (s, m, t and ℓ), depending on their parts (leaf or root).

Definition 8 (*n*-concepts). Let $\mathcal{C} = (\mathcal{D}_1, \dots, \mathcal{D}_n, \mathcal{I})$ be an *n*-context. An *n*-concept of \mathcal{C} is an *n*-tuple (X_1, \dots, X_n) such that $\prod_{i \in \{1, \dots, n\}} X_i \subseteq \mathcal{I}$ and there is no $k \in \{1, \dots, n\}$ and $x_k \in \mathcal{D}_k \setminus X_k$ for which $(X_1, \dots, X_k \cup \{x_k\}, \dots, X_n)$ respects this property.

As in the bidimensional case, *n*-concepts are units of knowledge in the data. They correspond to maximal *n*-dimensional boxes of crosses in the *n*-context. We use $\mathcal{T}(\mathcal{C})$ to denote the set of *n*-concepts of an *n*-context \mathcal{C} . For a concept $C = (X_1, \dots, X_n)$, we use $\mathcal{D}_i(C) = X_i$ to denote the *i*th component of the tuple.

The 3-concepts from Figure 6 are:

$$\begin{array}{l}
 (\{L\}, \{Root, Leaf\}, \{t, \ell\}) \quad | \quad (\{A, L\}, \{Root, Leaf\}, \{t\}) \quad | \quad (\{A, P, L\}, \{Leaf\}, \{t\}) \\
 (\{C, A, P, L\}, \{Root, Leaf\}, \emptyset) \quad | \quad (\{A, P\}, \{Leaf\}, \{m, t\}) \quad | \quad (\{C, P\}, \{Root, Leaf\}, \{m\}) \\
 (\{C, A, P\}, \{Leaf\}, \{m\}) \quad | \quad (\{C\}, \{Root, Leaf\}, \{s, m\}) \quad | \quad (\emptyset, \{Root, Leaf\}, \{s, m, t, \ell\}) \\
 (\{A, L\}, \{Root\}, \{t, \ell\}) \quad | \quad (\{C, P\}, \{Root\}, \{s, m\}) \quad | \quad (\{C, A, P, L\}, \emptyset, \{s, m, t, \ell\}).
 \end{array}$$

The main theorem of PCA [14] states that, for any *n*-context \mathcal{C} , $(\mathcal{T}(\mathcal{C}), \subseteq_{\mathcal{D}_1}, \dots, \subseteq_{\mathcal{D}_n})$ is a complete *n*-lattice called the *concept n-lattice* of \mathcal{C} .

3 Polyadic Relational Concept Analysis

Relational concept analysis does not work with multidimensional contexts because of the way the information on relations is integrated into the contexts as attributes. Let us consider a 3-context describing OBJECTS through the ATTRIBUTES they possess under some CONDITIONS. The objects of this 3-contexts are in relation with objects of another context. Using the RCA approach, one would create attributes from the concepts of the other context. However, where should these attributes be placed? If added to the set of attributes, under which conditions are the objects described by them?

To solve this problem, we propose a new formalism, called Polyadic Relational Concept Analysis (PRCA), in which the new “attributes” constructed from a (quantifier,relation) pair at a given step form a new dimension in the *n*-context. More formally:

Definition 9 (Polyadic relational context family). Let a polyadic relational context family (PRCF) be a pair (\mathbb{C}, \mathbb{R}) in which

- $\mathbb{C} = \{C_i = (\mathcal{D}_1^i, \dots, \mathcal{D}_n^i, \mathcal{I}^i)\}$ is a set of n -contexts with various numbers of dimensions
- $\mathbb{R} = \{r_k \mid r_k \subseteq \mathcal{D}_1^i \times \mathcal{D}_1^j\}$ is a set of relations between the elements of the first dimensions (hereafter called objects without loss of generality) of source n -contexts C_i and those of target n -contexts C_j .

Just as in RCA, we use $\text{src}(r)$ and $\text{tar}(r)$ to denote, respectively, the source and target n -context of a relation r .

We use the polyadic relational context family that consists of the 3-context *Plants* from Figure 6, together with the context *Areas* and the relation *GrowsIn* from Figure 2 as a running example.

Similarly to RCA, we define the relational scaling of an n -context in order to take into account information on the relations involving its objects. We use the \cdot notation to denote the fusion of two tuples. For example, $(x_1, \dots, x_i, \dots, x_n) = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) \cdot (x_i)$.

Definition 10 (Relation scaling of an n -context). Let (\mathbb{C}, \mathbb{R}) be a PRCF, $C_i = (\mathcal{D}_1^i, \dots, \mathcal{D}_n^i, \mathcal{I}^i)$ an n -context in \mathbb{C} , R the set of relations in \mathbb{R} for which C_i is the source and \mathcal{Q} a set of scaling operators. We still only consider $\mathcal{Q} \subseteq \{\exists, \forall\}$. The formal context $C_i^+ = \{\mathcal{D}_1^i, \mathcal{D}_2^{i+}, \dots, \mathcal{D}_n^{i+}, \mathcal{D}_{n+1}^i, \dots, \mathcal{D}_{n+m}^i, \mathcal{I}^{i+}\}$ is defined such that

$$\mathcal{D}_k^{i+} = \mathcal{D}_k^i \cup \{\circ_k\}, \forall j \in \{2, \dots, n\}$$

$$\mathcal{D}_{n+k}^i = \{\circ_{n+k}\} \cup \{q_k r_k \cdot X \mid (q_k, r_k) \in \mathcal{Q} \times \mathbb{R} \text{ and } X \in \mathcal{T}(\text{tar}(r_k))\}$$

with $\mathcal{Q} \times \mathbb{R} = \{(q_1, r_1), \dots, (q_m, r_m)\}$,

$$Z = \{(d_1, \dots, d_n, q_1 r_1 \cdot X_1, \dots, q_m r_m \cdot X_m) \mid (d_1, \dots, d_n) \in \mathcal{I}_i\}$$

$$\text{and } \forall k \in \{1, \dots, m\}, \left\{ \begin{array}{l} r(d_1) \subseteq \mathcal{D}_1(X_k) \text{ if } q_k = \forall \\ r(d_1) \cap \mathcal{D}_1(X_k) \neq \emptyset \text{ if } q_k = \exists \end{array} \right\}$$

and

$$\mathcal{I}_i^+ = Z \cup \{x \cdot (\circ_{i_1}, \dots, \circ_{i_k}) \mid \{i_1, \dots, i_k\} \subseteq \{2, \dots, n+m\}\}$$

$$\text{and } \exists y \in \prod_{l=1}^k \mathcal{D}_{i_l}, x \cdot y \in Z.$$

We write $\mathbb{C}^+ = \{C_i^+ \mid C_i \in \mathbb{C}\}$.

Less formally, the relational scaling of the n -context consists in creating a new dimension for each quantifier and relation pair that contains elements constructed from concepts of the target context. Then, the whole descriptions of objects are projected on the elements of the new dimensions iff the image of the object by the relation satisfies the constraint of the quantifier w.r.t. the concept used to create the element. An additional element \circ_i is added to each dimension except the first and all the ‘‘crosses’’ of \mathcal{I} are projected on this new element.

Hence, \odot_i can be interpreted as a disjunction over all the elements of \mathcal{D}_i and is used to link the dimensions in case a set of objects is not described by any element \mathcal{D}_i .

Figure 7 shows the 4-context resulting from a step of this process for the polyadic RCF described above. The new dimension corresponding to the relational attributes is depicted on the bottom of the context. The layer of the neutral element \odot_4 contains the starting 3-context. Then, the layers that correspond to relational attributes are filled as described above, using the \exists quantifier. For instance, as (A, Am) is an element of the relation *GrowsIn*, and Am is in the extent of *Areas'* concept C_0 , the description of object A is copied on the new layer corresponding to $\exists \text{Gr}C_0$. The lines that correspond to Root and Leaf are copied as they are in the starting context, and the part that corresponds to \odot_3 gets their union. Since only the object A is related *via GrowsIn* to an element from C_0 's extent, only object A 's description is present in this layer. The next layer corresponds to a relation with C_1 's extent. Objects C , A , and L from *Plants* are related, in *GrowsIn* and with the quantifier \exists , to this concept. Thus, their descriptions are copied in the layers corresponding to Root and Leaf, and the union of their descriptions in \odot_3 .

4 Properties of PRCA's output

The PRCA process adds a new dimension to the context for every (quantifier, relation) pair, at each step. This can quickly result in an overwhelming number of dimensions. Even though we do not have tight upper bounds for the number of n -concepts in an n -context at the time of writing, it is generally admitted that this number quickly increases as dimensions are added which could render PRCA unusable on real data. In this section, we show that PRCA is indeed usable as most of the concepts it produces are in a one-to-one correspondence with the concepts of a context of the same dimensionality as the initial data. To do this, we prove the following Theorem 1.

Theorem 1. *Let $C = (\mathcal{D}_1, \dots, \mathcal{D}_n, \mathcal{I})$ be an n -context.*

Let $C^\square = (\mathcal{D}_1, \dots, \mathcal{D}_n, \mathcal{D}_{n+1}, \mathcal{I}^\square)$ be an $(n+1)$ -context such that

$$(x_1, \dots, x_n, x_{n+1}) \in \mathcal{I}^\square \Rightarrow (x_1, y_2, \dots, y_n, x_{n+1}) \in \mathcal{I}^\square, \forall (x_1, y_2, \dots, y_n) \in \mathcal{I}.$$

Let $C^\Delta = (\mathcal{D}_1, \mathcal{D}_2 \times \mathcal{D}_{n+1}, \dots, \mathcal{D}_n \times \mathcal{D}_{n+1}, \mathcal{I}^\Delta)$ be an n -context such that

$$(x_1, (x_2, y_2), \dots, (x_n, y_n)) \in \mathcal{I}^\Delta \Leftrightarrow (x_1, x_2, \dots, x_n, y_i) \in \mathcal{I}^\square, \forall i \in \{2, \dots, n\}.$$

We have that, if $C_{n+1} \neq \emptyset$,

$$(C_1, \dots, C_{n+1}) \in \mathcal{T}(C^\square) \Leftrightarrow (C_1, C_2 \times C_{n+1}, \dots, C_n \times C_{n+1}) \in \mathcal{T}(C^\Delta).$$

Plants ¹	\mathcal{O}_2 s m t ℓ	\mathcal{O}_2 s m t ℓ	\mathcal{O}_2 s m t ℓ
Citrus (C)	× × ×	× × ×	× × ×
Alpinia (A)	× × × ×	× × ×	× × ×
Pelagonium (P)	× × × ×	× × ×	× × ×
Laphangium (L)	× × × ×	× × ×	× × ×
	\mathcal{O}_3	Root	Leaf
\mathcal{O}_4			

Plants ¹	\mathcal{O}_2 s m t ℓ	\mathcal{O}_2 s m t ℓ	\mathcal{O}_2 s m t ℓ
Citrus (C)			
Alpinia (A)	× × × ×	× × ×	× × ×
Pelagonium (P)			
Laphangium (L)			
	\mathcal{O}_3	Root	Leaf
$\exists \text{Gr}C_0$			

Plants ¹	\mathcal{O}_2 s m t ℓ	\mathcal{O}_2 s m t ℓ	\mathcal{O}_2 s m t ℓ
Citrus (C)	× × ×	× × ×	× × ×
Alpinia (A)	× × × ×	× × ×	× × ×
Pelagonium (P)			
Laphangium (L)	× × × ×	× × ×	× × ×
	\mathcal{O}_3	Root	Leaf
$\exists \text{Gr}C_1$			

Plants ¹	\mathcal{O}_2 s m t ℓ	\mathcal{O}_2 s m t ℓ	\mathcal{O}_2 s m t ℓ
Citrus (C)			
Alpinia (A)	× × × ×	× × ×	× × ×
Pelagonium (P)	× × × ×	× × ×	× × ×
Laphangium (L)			
	\mathcal{O}_3	Root	Leaf
$\exists \text{Gr}C_2$			

Plants ¹	\mathcal{O}_2 s m t ℓ	\mathcal{O}_2 s m t ℓ	\mathcal{O}_2 s m t ℓ
Citrus (C)	× × ×	× × ×	× × ×
Alpinia (A)	× × × ×	× × ×	× × ×
Pelagonium (P)	× × × ×	× × ×	× × ×
Laphangium (L)	× × × ×	× × ×	× × ×
	\mathcal{O}_3	Root	Leaf
$\exists \text{Gr}C_3$			

Fig. 7: Context \mathcal{C}^+ extended with the relational attributes built from the concepts of *Areas*.

Proof. First, we show \Rightarrow . Let (C_1, \dots, C_{n+1}) be a concept in $\mathcal{T}(C^\square)$. We have that $C_1 \times \dots \times C_n \subseteq \mathcal{I}^\square$ so we know that $(x_1, (x_2, y_2), \dots, (x_n, y_n)) \in \mathcal{I}^\Delta$ for all y_i such that $(x_1, x_2, \dots, x_n, y_i) \in C_1 \times \dots \times C_n$. If $(C_1, C_2 \times C_{n+1}, \dots, C_n \times C_{n+1})$ is not a concept in $\mathcal{T}(C^\Delta)$, then there is a $c_1 \in C_1$ such that $\{c_1\} \times (C_2 \times C_{n+1}) \times \dots \times (C_n \times C_{n+1}) \subseteq \mathcal{I}^\Delta$ or there is a $c_i \in (C_i \times C_{n+1})$ such that $C_1 \times \dots \times (C_i \times C_{n+1}) \times \dots \times (C_n \times C_{n+1}) \subseteq \mathcal{I}^\Delta$. In both cases, it implies that (C_1, \dots, C_{n+1}) is not a concept in $\mathcal{T}(C^\square)$ as it is not a maximal box. Hence, $(C_1, C_2 \times C_{n+1}, \dots, C_n \times C_{n+1}) \in \mathcal{T}(C^\Delta)$.

Now, we show \Leftarrow . Let $(C_1, (C_2 \times C_{n+1}), \dots, (C_n \times C_{n+1}))$ be a concept in $\mathcal{T}(C^\Delta)$. We have that $C_1 \times (C_2 \times C_{n+1}) \times \dots \times (C_n \times C_{n+1}) \subseteq \mathcal{I}^\Delta$ so $C_1 \times C_2 \times \dots \times C_{n+1} \subseteq \mathcal{I}^\square$. If (C_1, \dots, C_{n+1}) is not a concept in $\mathcal{T}(C^\square)$, then there is a $c_i \in C_i$ such that $C_1 \times \dots \times \{c_i\} \times \dots \times C_{n+1} \subseteq \mathcal{I}^\square$ for all $i \in \{1, \dots, n+1\}$. Then, $(C_1, (C_2 \times C_{n+1}), \dots, (C_n \times C_{n+1}))$ is not a concept in $\mathcal{T}(C^\Delta)$ as it is not a maximal box. Hence, $(C_1, \dots, C_{n+1}) \in \mathcal{T}(C^\square)$.

The relational scaling of an n -context with a single (quantifier, relation) pair, as described in Definition 10, is an instance of the $(n+1)$ -context C^\square of the theorem. Hence, this theorem tells us that the number of concepts produced by PRCA is bounded by the maximal number of concepts in a context of the same dimensionality as the initial context (plus some number of n -concepts with empty components in the dimensions created by the process). Most importantly, it means that PRCA produces what amounts to the elements of a concept lattice when applied to a bidimensional context.

4.1 PRCA as a generalisation of RCA

In this section, we show that the concepts produced by PRCA when applied to a dyadic context are the same as the ones produced by RCA (Proposition 1). This validates PRCA as a generalisation of RCA.

Proposition 1. *Let $C = (\mathcal{D}_1, \mathcal{D}_2, \mathcal{I})$ be a dyadic context in a PRCF, $\mathcal{T}(C_{RCA}^+)$ the set of concepts of the context after an iteration of the RCA process and $\mathcal{T}(C_{PRCA}^+)$ the set of concepts of the context after an iteration of the PRCA process. We have that*

$$(X, Y, Z_1, \dots, Z_n) \in \mathcal{T}(C_{PRCA}^+) \Leftrightarrow (X, (Y \setminus \{\circ_2\}) \cup \bigcup_{i=1}^n (Z_i \setminus \{\circ_{i+2}\})) \in \mathcal{T}(C_{RCA}^+)$$

if $Y, Z_1, \dots, Z_n \neq \emptyset$.

Proof. First, we show \Rightarrow . If (X, Y, Z_1, \dots, Z_n) is a concept in $\mathcal{T}(C_{PRCA}^+)$, then for all $x \in X$ and $(y, z_1, \dots, z_n) \in Y \times Z_1 \times \dots \times Z_n$, we have that (x, y, z_1, \dots, z_n) is in the incidence relation \mathcal{I}_{PRCA}^+ . From the construction of \mathcal{I}_{PRCA}^+ , we have that $(x, y) \in \mathcal{I}$ and, for all $z_i \neq \circ_{i+2}$, either

$$\begin{cases} r(x) \cap A \neq \emptyset \text{ if } z_i = \exists r.(A, B) \\ r(x) \subseteq A \text{ if } z_i = \forall r.(A, B). \end{cases}$$

Thus, (x, z_i) is in \mathcal{I}_{RCA}^+ . As a consequence, if $(X, (Y \setminus \{\circ_2\}) \cup \bigcup_{i=1}^n Z_i \setminus \{\circ_{i+2}\})$ is not a concept in $\mathcal{T}(C_{RCA}^+)$, then either there is a v in $\mathcal{D}_1 \setminus X$ such that (v, y) is in \mathcal{I}_{RCA}^+ for all y in $(Y \setminus \{\circ_2\}) \cup \bigcup_{i=1}^n (Z_i \setminus \{\circ_{i+2}\})$, or there is a w in $((\mathcal{D}_2 \setminus \{\circ_2\}) \cup \bigcup_{i=1}^n (\mathcal{D}_{2+i} \setminus \{\circ_{2+i}\})) \setminus (Y \cup \bigcup_{i=1}^n Z_i)$ such that (x, w) is in \mathcal{I}_{RCA}^+ for all $x \in X$. By construction of \mathcal{I}_{PRCA}^+ , we have that both those cases imply that (X, Y, Z_1, \dots, Z_n) is not a concept in $\mathcal{T}(C_{PRCA}^+)$, which contradicts our hypothesis. Hence, $(X, (Y \setminus \{\circ_2\}) \cup \bigcup_{i=1}^n Z_i \setminus \{\circ_{i+2}\})$ is a concept in $\mathcal{T}(C_{RCA}^+)$.

Now, we show \Leftarrow . If $(X, (Y \setminus \{\circ_2\}) \cup \bigcup_{i=1}^n (Z_i \setminus \{\circ_{i+2}\}))$ is a concept in $\mathcal{T}(C_{RCA}^+)$, then for all $x \in X$ and $y \in Y \setminus \{\circ_2\}$, (x, y) is in \mathcal{I} . Additionally, for all z in $\bigcup_{i=1}^n Z_i \setminus \{\circ_{i+2}\}$, either

$$\begin{cases} r(x) \cap A \neq \emptyset \text{ if } z = \exists r.(A, B) \\ r(x) \subseteq A \text{ if } z = \forall r.(A, B) \end{cases}$$

Hence, by construction of \mathcal{I}_{PRCA}^+ , $X \times Y \times Z_1 \times \dots \times Z_n \subseteq \mathcal{I}_{PRCA}^+$. Then, if (X, Y, Z_1, \dots, Z_n) is not a concept in $\mathcal{T}(C_{PRCA}^+)$, it means either that there is a v in $\mathcal{D}_1 \setminus X$ such that $\{v\} \times Y \times Z_1 \times \dots \times Z_n \subseteq \mathcal{I}_{PRCA}^+$, that there is a w in $\mathcal{D}_2 \setminus Y$ such that $X \times \{w\} \times Z_1 \times \dots \times Z_n \subseteq \mathcal{I}_{PRCA}^+$, or that there is a z_i in $\mathcal{D}_{i+2} \setminus Z_i$ such that $X \times Y \times Z_1 \times \dots \times \{z_i\} \times \dots \times Z_n \subseteq \mathcal{I}_{PRCA}^+$. All three cases imply that $(X, (Y \setminus \{\circ_2\}) \cup \bigcup_{i=1}^n (Z_i \setminus \{\circ_{i+2}\}))$ is not a concept in $\mathcal{T}(C_{RCA}^+)$, which contradicts our hypothesis. Thus, (X, Y, Z_1, \dots, Z_n) is a concept in $\mathcal{T}(C_{PRCA}^+)$. \square

Proposition 1 shows that the knowledge generated by applying PRCA on a dyadic context or RCA is fundamentally the same, albeit through a reframing of the relational attributes as new dimensions.

We now provide an example of the application of PRCA, as described in Section 3, on the dyadic context depicted in Figure 2. The first step gives the same concept lattices as in Figure 3. The context *Plants* is then extended with a new dimension that contains the following elements

$$\{\circ_3, \exists \text{GrowsIn}C_0, \exists \text{GrowsIn}C_1, \exists \text{GrowsIn}C_2, \exists \text{GrowsIn}C_3\},$$

while its second dimension is extended with \circ_2 .

The incidence relation is built from the *Plants* context, by reporting the description of the objects on the corresponding layers: first in the layer corresponding to \circ_3 and then in the layers that correspond to concepts related to each object *via* the quantifier.

Let us list the concepts from the 3-context in Figure 8:

Plants ⁺	\mathcal{O}_2 s m t ℓ				
Citrus (C)	× × ×		× × ×		× × ×
Alpinia (A)	× × × ×	× × ×	× × × ×	× × ×	× × × ×
Pelargonium (P)	× × ×			× × ×	× × ×
Laphangium (L)	× × ×		× × ×		× × ×
	\mathcal{O}_3	$\exists \text{Gr}C_0$	$\exists \text{Gr}C_1$	$\exists \text{Gr}C_2$	$\exists \text{Gr}C_3$

Fig. 8: Our starting dyadic context is extended with a new dimension that contains its relational attributes.

$(\emptyset, \{\mathcal{O}_2, s, m, t, \ell\}, \{\mathcal{O}_3, \exists \text{GrowsIn}C_0, \exists \text{GrowsIn}C_1, \exists \text{GrowsIn}C_2, \exists \text{GrowsIn}C_3\})$
 $(\{A\}, \{\mathcal{O}_2, t, \ell\}, \{\mathcal{O}_3, \exists \text{GrowsIn}C_0, \exists \text{GrowsIn}C_1, \exists \text{GrowsIn}C_2, \exists \text{GrowsIn}C_3\})$
 $(\{C\}, \{\mathcal{O}_2, s, m\}, \{\mathcal{O}_3, \exists \text{GrowsIn}C_1, \exists \text{GrowsIn}C_3\})$
 $(\{P\}, \{\mathcal{O}_2, m, t\}, \{\mathcal{O}_3, \exists \text{GrowsIn}C_2, \exists \text{GrowsIn}C_3\})$
 $(\{A, P\}, \{\mathcal{O}_2, t\}, \{\mathcal{O}_3, \exists \text{GrowsIn}C_2, \exists \text{GrowsIn}C_3\})$
 $(\{C, P\}, \{\mathcal{O}_2, m\}, \{\mathcal{O}_3, \exists \text{GrowsIn}C_3\})$
 $(\{A, L\}, \{\mathcal{O}_2, t, \ell\}, \{\mathcal{O}_3, \exists \text{GrowsIn}C_1, \exists \text{GrowsIn}C_3\})$
 $(\{C, A, L\}, \{\mathcal{O}_2\}, \{\mathcal{O}_3, \exists \text{GrowsIn}C_1, \exists \text{GrowsIn}C_3\})$
 $(\{A, P, L\}, \{\mathcal{O}_2, t\}, \{\mathcal{O}_3, \exists \text{GrowsIn}C_3\})$
 $(\{C, A, P, L\}, \{\mathcal{O}_2\}, \{\mathcal{O}_3, \exists \text{GrowsIn}C_3\})$
 $(\{C, A, P, L\}, \emptyset, \{\mathcal{O}_3, \exists \text{GrowsIn}C_0, \exists \text{GrowsIn}C_1, \exists \text{GrowsIn}C_2, \exists \text{GrowsIn}C_3\})$
 $(\{C, A, P, L\}, \{\mathcal{O}_2, s, m, t, \ell\}, \emptyset)$

Let us consider the concept $(\{A, P\}, \{\mathcal{O}_2, t\}, \{\mathcal{O}_3, \exists \text{GrowsIn}C_2, \exists \text{GrowsIn}C_3\})$ from Figure 8. It provides the same information as the equivalent RCA concept $(\{A, P\}, \{t, \exists \text{GrowsIn}C_2, \exists \text{GrowsIn}C_3\})$ (see Figure 5). More generally, the RCA concepts can be obtained from the PRCA concepts by removing the elements \mathcal{O}_i and merging the last two components. Thus, one can check that the mapping described in Proposition 1 holds.

5 Discussion and Conclusion

Our proposed framework extends RCA to polyadic contexts. As such, it inherits the drawbacks of both relational and polyadic concept analysis. At each step, many new elements are created so dimensions can be big. As the number of n -concepts grows quickly with the size of the n -context, the size of the output can become overwhelming. In addition to that, efficient graphical representations of n -lattices are still open questions so human interpretation of the output of PRCA can be difficult. It would be interesting, as a future work, to study algorithms for the interactive exploration of the n -lattice in line with those that exist for the dyadic case [1]. However, the notion of *neighbours* of an n -concept in an n -lattice is blurrier than in traditional lattices so the generalisation of such algorithms is not straightforward.

The use cases for our methods are as wide as are RCA's. It can be used in place of RCA in dyadic relational datasets, and can take into account the eventual ternary or n -ary context in those datasets. These applications can range from plants datasets [13] to families of RDF datasets.

Lastly, allowing n -ary relations in contexts is a step forward but, in our framework, relations between objects of different contexts are still binary. As these relations can also be n -ary in real datasets (or just in real life), it would be interesting to improve the framework to take such relations into consideration.

Acknowledgements

Part of this work was supported by the French National Research Agency (ANR) through the project SmartFCA (ANR-21-CE23-0023).

References

1. Alexandre Bazin, Jessie Carbonnel, Marianne Huchard, Giacomo Kahn, Priscilla Keip, and Amirouche Ouzerdine. On-demand relational concept analysis. In *International Conference on Formal Concept Analysis*, pages 155–172. Springer, 2019.
2. Jessie Carbonnel, Marianne Huchard, and Clémentine Nebut. Exploring the variability of interconnected product families with relational concept analysis. In *Proceedings of the 23rd International Systems and Software Product Line Conference-Volume B*, pages 199–206, 2019.
3. Xavier Dolques, Agnes Braud, Corinne Grac, and Florence Le Ber. Analyzing water monitoring data with rca-based approaches. In *Workshop RealDataFCA '2021*, 2021.
4. Maximilian Felde and Gerd Stumme. Triadic exploration and exploration with multiple experts. In *International Conference on Formal Concept Analysis*, pages 175–191. Springer, 2021.
5. Sébastien Ferré and Peggy Cellier. Graph-fca: An extension of formal concept analysis to knowledge graphs. *Discrete applied mathematics*, 273:81–102, 2020.
6. Bernhard Ganter and Rudolf Wille. *Formal concept analysis: mathematical foundations*. Springer Science & Business Media, 2012.
7. Mohamed Rouane Hacene, Marianne Huchard, Amedeo Napoli, and Petko Valtchev. Relational concept analysis: mining concept lattices from multi-relational data. *Ann. Math. Artif. Intell.*, 67(1):81–108, 2013.
8. Nicolas Hlad, Bérénice Lemoine, Marianne Huchard, and Abdelhak-Djamel Seriai. Leveraging relational concept analysis for automated feature location in software product lines. In *Proceedings of the 20th ACM SIGPLAN International Conference on Generative Programming: Concepts and Experiences*, pages 170–183, 2021.
9. Dmitry I Ignatov, Dmitry V Gnatyshak, Sergei O Kuznetsov, and Boris G Mirkin. Triadic formal concept analysis and triclustering: searching for optimal patterns. *Machine Learning*, 101(1):271–302, 2015.
10. Priscilla Keip, Alain Gutierrez, Marianne Huchard, Florence Le Ber, Samira Sarter, Pierre Silvie, and Pierre Martin. Effects of input data formalisation in relational concept analysis for a data model with a ternary relation. In Diana Cristea, Florence Le Ber, and Baris Sertkaya, editors, *Formal Concept Analysis - 15th International Conference, ICFCA 2019, Frankfurt, Germany, June 25-28, 2019, Proceedings*, volume 11511 of *Lecture Notes in Computer Science*, pages 191–207. Springer, 2019.
11. Fritz Lehmann and Rudolf Wille. A triadic approach to formal concept analysis. In *International conference on conceptual structures*, pages 32–43. Springer, 1995.

12. Lina Mahrach, Alain Gutierrez, Marianne Huchard, Priscilla Keip, Pascal Marnotte, Pierre Silvie, and Pierre Martin. Combining implications and conceptual analysis to learn from a pesticidal plant knowledge base. In *International Conference on Conceptual Structures*, pages 57–72. Springer, 2021.
13. Pierre J Silvie, Pierre Martin, Marianne Huchard, Priscilla Keip, Alain Gutierrez, and Samira Sarter. Prototyping a knowledge-based system to identify botanical extracts for plant health in sub-saharan africa. *Plants*, 10(5):896, 2021.
14. George Voutsadakis. Polyadic concept analysis. *Order*, 19(3):295–304, 2002.