



**HAL**  
open science

## Practical comparison of FCA extensions to model indeterminate value of ternary data

Priscilla Keip, Sébastien Ferré, Alain Gutierrez, Marianne Huchard, Pierre  
Silvie, Pierre Martin

► **To cite this version:**

Priscilla Keip, Sébastien Ferré, Alain Gutierrez, Marianne Huchard, Pierre Silvie, et al.. Practical comparison of FCA extensions to model indeterminate value of ternary data. CLA 2020 - 15th International Conference on Concept Lattices and Their Applications, Jun 2020, Tallinn, Estonia. pp.197-208. lirmm-03105801

**HAL Id: lirmm-03105801**

<https://hal-lirmm.ccsd.cnrs.fr/lirmm-03105801>

Submitted on 11 Jan 2021

**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.



Distributed under a Creative Commons Attribution 4.0 International License

# Practical Comparison of FCA Extensions to Model Indeterminate Value of Ternary Data <sup>\*</sup>

Priscilla Keip<sup>1</sup>, Sébastien Ferré<sup>2</sup>, Alain Gutierrez, Marianne Huchard<sup>3</sup>,  
Pierre Silvie<sup>1,4</sup>, and Pierre Martin<sup>1</sup>

<sup>1</sup> CIRAD, UPR AIDA, F-34398 Montpellier, France  
AIDA, Univ Montpellier, CIRAD, Montpellier, France  
{priscilla.keip,pierre.silvie,pierre.martin}@cirad.fr  
<sup>2</sup> Univ Rennes, CNRS, IRISA, 35042 Rennes, France  
ferre@irisa.fr

<sup>3</sup> LIRMM, Université de Montpellier, CNRS, Montpellier, France  
marianne.huchard@lirmm.fr

<sup>4</sup> IRD, UMR IPME, F-34394 Montpellier, France

**Abstract.** The Knomana knowledge base brings together knowledge from the scientific literature on the use of plants with pesticidal or antibiotic effects on animals, plants, and human beings to propose protection solutions using local plants. In this literature, the elements of the 3-tuple (protected organism, protecting plant, pest) are named using the binomial nomenclature consisting of the genus name followed by the species name. In some instances, authors use the abbreviation “sp.” in the singular or “spp.” in the plural, as species name, to indicate the indeterminate status of the species for a guaranteed genus. To suggest protection solutions, the indeterminacy of the species has to be hypothesized based on assigning the sp./spp. to the other species in the same genus and conversely. This paper discusses the classification of ternary data containing some indeterminate values generated by three extensions of Formal Concept Analysis.

**Keywords:** Graph-FCA · Relational Concept Analysis · Triadic Concept Analysis

## 1 Introduction

According to the Committee on World Food Security, governments have to improve small food producers’ access to knowledge on the use of biodiversity for crop protection to enhance food security and nutrition [7]. With the aim of proposing protection solutions based on the use of local plants, the Knomana Knowledge Base (KKB) brings together knowledge from the scientific literature on the use of plants with pesticidal or antibiotic effects on animals, plants, and

---

<sup>\*</sup> This work was supported by the French National Research Agency under the Investments for the Future Program, referred as ANR-16-CONV-0004.

human beings [13]. The methodology used to suggest protection solutions is based on Formal Concept Analysis [9].

In KKB, a plant use is characterized using 70 descriptors, for example, the part of plant used to prepare the bio-product, the method of preparation and its efficacy. In early January 2020, KKB contained 42,400 plant use descriptions from 630 documents, dated 1957 to 2019. These descriptions include a total of 1,782 plants to protect 64 (crop, animal and human) systems against 451 bio-aggressors (pests, diseases, bacteria, viruses, etc.). This 3-tuple (protected organism, protecting plant, pest<sup>5</sup>) is the core of the protection solution as its efficacy is linked to the interactions between these three elements.

In the literature, each element of the 3-tuple is named using the binomial nomenclature, which consists of the genus name followed by the species name. This designation is a widely accepted convention for the formal scientific naming of the biological organisms, and avoids the ambiguity associated to vernacular names that are not shared by all communities. For the name of some species within the 3-tuples, authors use the abbreviation “sp.” in the singular or “spp.” in the plural. The abbreviation sp. indicates that the genus name of the organism is known by the authors but its species name is not; the abbreviation spp. indicates that at least two organisms of the same genus are considered without their names being mentioned.

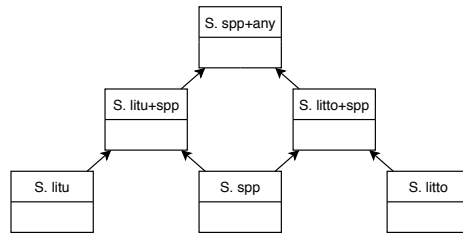
To classify 3-tuples using Formal Concept Analysis (FCA), the presence of the sp. or spp. abbreviation in a formal context corresponds to the absence of the name of the species, and thus to indeterminate value of data. This indeterminacy has to be hypothesized to suggest protection solutions. This paper deals with ternary relationships with FCA and discusses how to account for the indeterminate specification of species. Three extensions of FCA, Triadic Concept Analysis (TCA) [11], Relational Concept Analysis (RCA) [5], and Graph-FCA (GFCA) [3], were qualitatively and quantitatively evaluated on two reduced KKB datasets.

Section 2 presents a short excerpt of KKB to show the two issues addressed in this paper. Sections 3, 4 and 5 present solutions using, respectively, TCA, RCA, and GFCA. Section 6 compares the different solutions and Section 7 concludes the paper.

## 2 Indeterminate Information and Ternary Relationships in the Knomana Knowledge Base

For this work, two datasets were extracted from KKB. They describe known protection solutions against species of *Spodoptera*, including the pest *Spodoptera frugiperda* that is currently spreading very rapidly in Africa and is devastating maize, tomato and cabbage crops. The first dataset is composed of 12 3-tuples describing relations between six organisms, nine plants, and three pests (*S. littoralis*, *S. litura*, and *Spodoptera* spp.). This dataset enabled our qualitative evaluation that consisted in comparing the classifications provided by the

<sup>5</sup> Taken in this paper in the broadest sense to designate bio-aggressors.



**Fig. 1.** Aggregation hierarchy used to propagate the spp value in the first dataset.

FCA extensions to highlight similarities and differences among them. The second dataset was used for the quantitative evaluation of scalability. It is composed of 34 3-tuples describing relations between six organisms, 30 plants and four pests, *Spodoptera* spp., *S. frugiperda*, *S. littoralis*, and *S. litura*. This dataset includes the first one composed of three pests.

In both these datasets, the pests all belong to the genus *Spodoptera*. One is called *Spodoptera* spp., showing it refers to at least two species of the *Spodoptera* genus. It is not mentioned if the species are *S. littoralis*, *S. litura*, and *S. frugiperda* or others. To deal with this indeterminate value, we defined the three levels of aggregation presented in Fig. 1. The known species and the *Spodoptera* spp. indeterminate species are located at the bottom of the hierarchy. The hypothetical species, resulting from the aggregation of the ones located at the bottom, are located above: S. litu+spp (resp. S. litto+spp) gathers information on S.litu (resp. S. litto) and S.spp. That is, it encompasses known information on S. litu (resp. S. litto) or at least two *Spodoptera* species, possibly including S.litu (resp. S. litto). At the top of the hierarchy, S. spp+any gathers information on S. litto, S. litu, and S. spp. The use of this hierarchy differs depending on the FCA extension concerned and is described in the following sections.

Another characteristic of KKB is a central ternary relationship connecting the three main components: protected organisms, protecting plants and pests. FCA [4] and RCA consider binary relationships that require application of conceptual model transformations as proposed in the database domain [8]. The two other FCA extensions studied, TCA and Graph-FCA, are directly applicable to this kind of data: TCA because it was introduced for this purpose, Graph-FCA because it allows the definition of n-ary relationships of any dimension.

Both datasets and all files, corresponding to each evaluation of the three next sections, are available online [12].

### 3 Triadic Concept Analysis (TCA)

Triadic concept analysis (TCA) was introduced by Lehmann and Wille to deal with situations where “an object  $g$  has the attribute  $m$  under the condition  $b$ ” [11]. By extension, it can be considered to deal with any ternary relationship. In our case, the ternary relationship can be embedded in such a scheme. Pro-

tected organisms, protecting plants, and pests can be used respectively as objects, attributes and conditions. A triadic context is a 4-tuple  $K = (G, M, B, Y)$  where  $G$  is the set of objects,  $M$  the set of attributes,  $B$  the set of conditions, and  $Y \subseteq G \times M \times B$  associates an object and an attribute under a condition. For example, Table 1 (the three tables in the top row) is a triadic context in which the objects are protected organisms, the attributes are protecting plants, and the conditions are pests. It associates object A.escu with attribute A.indi under condition S.litto, meaning that *A.indi treats A.escu when attacked by S.litto*. Let us denote this relationship  $IT$ , and  $IT_{[Protected,Plant]}$  the mapping of  $IT$  on the first two components. Triplets inferred from this initial knowledge with the addition of S.litto+spp, S.litu+spp and S.spp+any, to deal with indeterminate value, are presented in the three tables in the bottom row in Table 1:  $\forall(po, pl, pe) \in IT_{[Protected,Plant]} \times \{S.litto, S.litu\}, (po, pl, pe + spp)$ ;  $\forall(po, pl, pe) \in IT_{[Protected,Plant]} \times \{S.spp\}, (po, pl, S.litto + spp)$  and  $(po, pl, S.litu + spp)$ ;  $\forall(po, pl) \in IT_{[Protected,Plant]}, (po, pl, S.spp + any)$ . It is easy to generalize this addition to species other than S.litto and S.litu.

A triadic concept  $(A1, A2, A3)$  of  $(G, M, B, Y)$  satisfies  $A1 \subseteq G$ ,  $A2 \subseteq M$  and  $A3 \subseteq B$  as well as:  $A1 \times A2 \times A3 \subseteq Y$  ( $(A1, A2, A3)$  is a rectangular parallelepiped full of  $\times$ ); and  $X1 \times X2 \times X3 \subseteq Y$ ,  $A1 \subseteq X1$ ,  $A2 \subseteq X2$  and  $A3 \subseteq X3$  implies that  $(A1, A2, A3) = (X1, X2, X3)$  (formal concept maximality).  $A1$ ,  $A2$  and  $A3$  are respectively called extent, intent and modus of the triadic concept  $(A1, A2, A3)$ . Table 1 highlights in red the triadic concept  $TC5 = (\{A.escu, B.ole\}, \{A.indi, C.papa\}, \{S.litto, S.litto + spp, S.spp + any\})$ . Selected triadic concepts for the triadic context of Table 1 are presented in Table 2 and were built using **FCA Tools Bundle**<sup>6</sup>. They reveal several facts on the *Spodoptera* excerpt. We chose them because they are non-trivial and provide valuable information on data. **TC2** indicates that no plant protects all organisms against *Spodoptera*. **TC5** groups A.escu and B.ole, which are protected against S.litto by A.indi and C.papa. This shows that A.indi and C.papa can replace one another to control S.litto. As shown by **TC6**, A. indi controls some *Spodoptera* species on the three plants A.escu, B.ole and Z.mays.

## 4 Relational Concept Analysis (RCA)

Relational Concept Analysis (RCA) [5] was developed to view a dataset (called a relational context family, or RCF) as a set of objects of several categories (e.g. protecting plants, protected organisms and pests), connected through several relationships (e.g. *an organism is protected by a plant* and *a plant treats a pest*). Three object-attribute contexts (**Plants** and **Protected**, with nominal scaling, and **Pests** with hierarchical scaling) describe the three object categories, while two object-object contexts (**protectedBy** and **treats**) encode the relationships. These relationships contain the same information as that listed in Table 1. RCA acts as an iterative process with several steps. Relations are considered through the addition of relational attributes based on quantifiers

<sup>6</sup> <https://fca-tools-bundle.com/>

**Table 1.** Triadic context, with highlighted concept TC5 (in red)

S. litu	W. pros	V. cane	V. fusc	V. parv	A. indi	C. spp	C. opul	C. papa	D. dumé	S. spp	W. pros	V. cane	V. fusc	V. parv	A. indi	C. spp	C. opul	C. papa	D. dumé	S. litto	W. pros	V. cane	V. fusc	V. parv	A. indi	C. spp	C. opul	C. papa	D. dumé	
Z. mays										Z. mays											Z. mays									
A. escu										A. escu											A. escu									
B. ole										B. ole											B. ole									
G. hirs										G. hirs											G. hirs									
S. lyco										S. lyco											S. lyco									
R. com	x									R. com											R. com									

S.litu+spp	W. pros	V. cane	V. fusc	V. parv	A. indi	C. spp	C. opul	C. papa	D. dumé	S.spp+any	W. pros	V. cane	V. fusc	V. parv	A. indi	C. spp	C. opul	C. papa	D. dumé	S.litto+spp	W. pros	V. cane	V. fusc	V. parv	A. indi	C. spp	C. opul	C. papa	D. dumé	
Z. mays										Z. mays											Z. mays									
A. escu										A. escu											A. escu									
B. oler										B. oler											B. oler									
G. hirsu										G. hirsu											G. hirsu									
S. lyco										S. lyco											S. lyco									
R. com	x									R. com	x										R. com									

**Table 2.** Three triadic concepts among the 9 concepts built using Table 1

TC2	Protected Plants Pests	A.escu B.ole G.hirsu R.com S.lyco Z.mays
TC5	Protected Plants Pests	S.litto S.litto+spp S.litu S.litu+spp S.spp S.spp-any
TC6	Protected Plants Pests	A.escu B.ole A.indi C.papa S.litto S.litto+spp S.spp-any

inspired by description logic operators [1]. The detailed process is described in [5] and for this paper, we used the RCAExplore tool [2] available online<sup>7</sup>. Figure 2 shows an excerpt of the result. The concept number in the lattices is respectively 12 (protected organisms), 14 (plants), and seven (pests). Concept *CPlant13* (Fig. 2 (middle)) groups plants (A. indi, C. papa) that treat at least one pest in *CPest1* (S.litto) (relational attribute  $\exists treats(CPest1)$ ). Concept *CProt8* (Fig. 2 (left)) groups organisms (A.escu, B.ole) that are protected by at least one plant in *CPlant2*, *CPlant1* and *CPlant13* (relational attributes  $\exists protectedBy(CPlant2)$ ,  $\exists protectedBy(CPlant1)$  and  $\exists protectedBy(CPlant13)$ , the latter being inherited). This concept chain, which crosses the three lattices, corresponds to Triadic concept **TC5**. It would also be found with quantifier  $\exists \supseteq$ . Similarly, the Concept chain *CProt9*, *CPlant1* and *CPest6* corresponds to Triadic concept **TC6**. Triadic concept **TC2** does not appear with quantifier  $\exists$ , because all protected organisms have at least one protecting plant against one pest: situation (*CPo*, *CPl*, *CPe*) such that  $\forall po \in CPo$ ,  $\exists pl \in CPl$ ,  $(po, pl) \in protectedBy$ , and  $\exists pe \in CPe$ ,  $(pl, pe) \in treats$ , that could be summarized as pattern  $\exists protectedBy \exists treats$ . **TC2** would appear with quantifier  $\exists \supseteq$ .

RCA enables various encodings of datasets. For example, two alternatives can be proposed for the above encoding named RCAC (for RCA with Chain). They

<sup>7</sup> <http://dataqual.engees.unistra.fr/logiciels/rcaExplore>

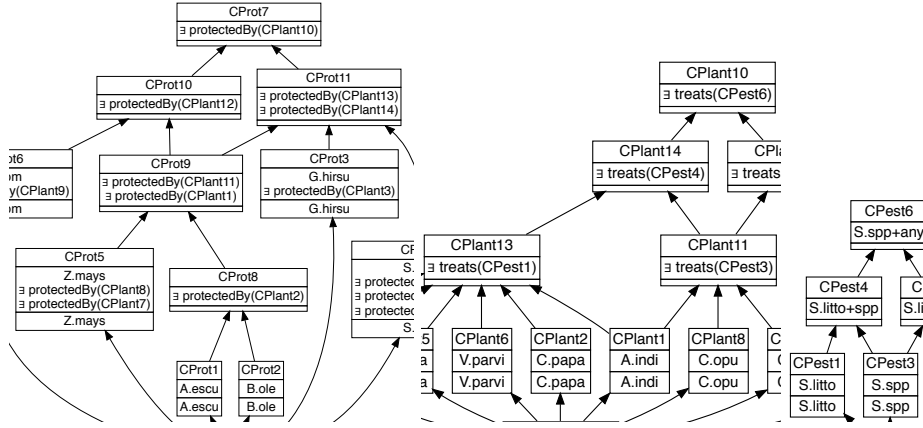


Fig. 2. RCAC Concept lattices: protected org. (left), plants (middle), pests (right)

are illustrated with quantifier  $\exists$ . The conceptual models corresponding to these variants are shown on the left in Figure 4. A first alternative (called RCAR for RCA with Reification) consists of reifying the ternary relation that is thus added as a fourth object-attribute context to the RCF, whose objects are the 3-tuples of the relation. This results in 4 lattices. Figure 3 shows an excerpt of the central lattice (24 concepts) where the objects are the 3-tuples. The other lattices are quite flat, or with a hierarchy (the case of pests). Concept  $CPion22$  groups the 3-tuples corresponding to Triadic Concept  $TC5$ , and the relational attribute  $\exists manages(CPest1)$  (S.litto). Their respective plants and protected organisms are separated into sub-concepts of  $CPion22$ . Triadic concept  $TC6$  corresponds to  $CPion19$ , and the relational attributes indicate that S.litto+spp is managed ( $\exists manages(CPest4)$ ) and the plant used is A.indi ( $\exists uses(CPlant1)$ ). A second alternative (called RCAG for RCA with Graph view) consists of an RCF composed of two categories that describe protected organisms and pests, respectively, and one relationship per protecting plant. This results in two lattices, with nine protected organism concepts (Fig. 4 (middle)), and seven pest concepts (Fig. 4 (right)). Triadic concept  $TC5$  (resp.  $TC6$ ) appears as  $CProt8$  (resp.  $CProt9$ ). The shared relational attributes describe protecting plants as roles pointing to controlled pests.

## 5 Graph-FCA (GFCA)

Graph-FCA (or GFCA) [3] was introduced as an extension of classical FCA to knowledge graphs, or to be more precise, to directed multi-hyper-graphs. A graph context is a 3-tuple  $K = (O, A, I)$  like in FCA except that objects play the role of nodes in the graph, and elements of the incidence relation  $I$  are directed hyper-edges connecting objects and labeled with attributes. Unary attributes connect a single node: they act as node labels and correspond to

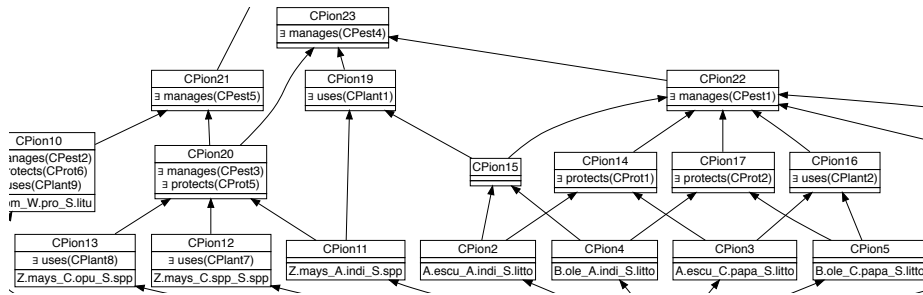


Fig. 3. Pests (RCAR)

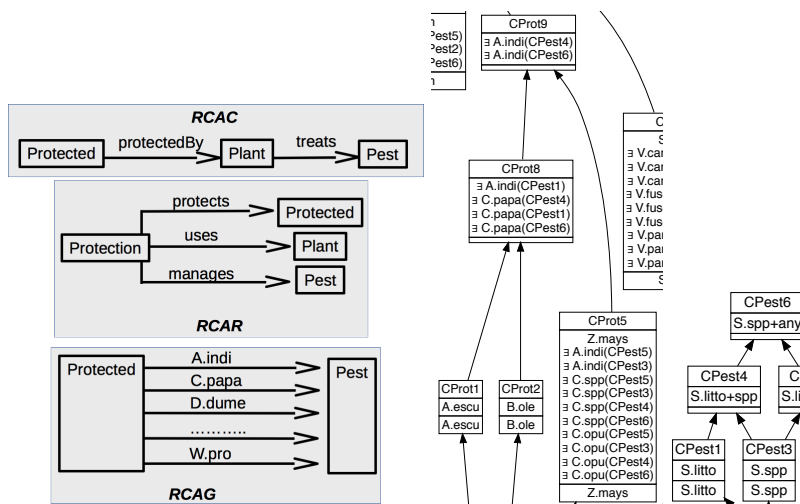
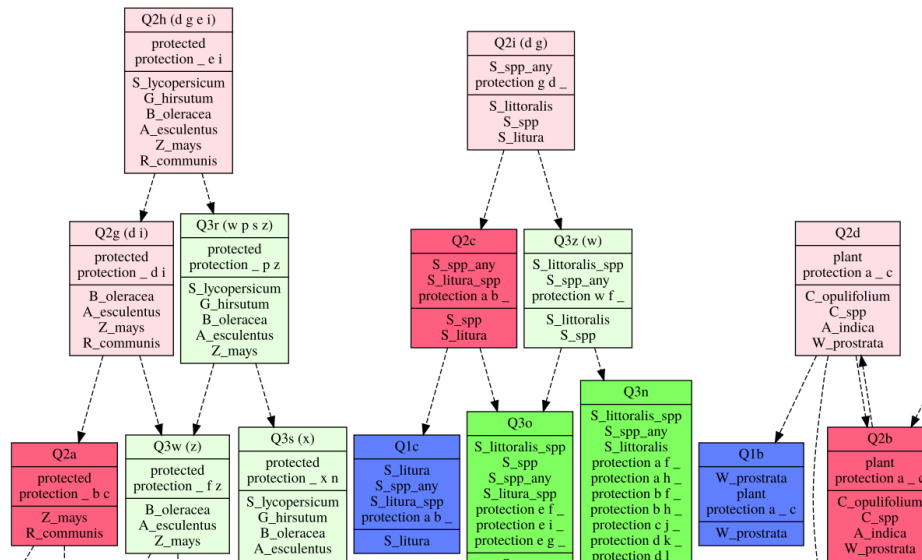


Fig. 4. Data models (left), RCAG lattices: Protected org. (middle), pests (right)

classical FCA attributes. Binary attributes connect one node to another, and therefore represent relationships or links between objects. N-ary attributes can encode more complex relationships involving more than two participants (e.g., “plant X protects organism Y against a pest Z”). The intents of *graph concepts* are like *conjunctive queries*, i.e. graph patterns with a distinguished node. GFCA also defines n-ary graph concepts, but here we only consider unary concepts.

Choosing a Graph-FCA representation of the 3-tuples (Protected,Plant,Pest) amounts to choosing what the objects and the attributes are. An immediate representation (GFCA-Ternary) would be to introduce a ternary attribute for the ternary relation between protected organisms, protecting plants, and pests: e.g., *protection(A.escu, A.indi, S.litto)*. Unary attributes are used with nominal scaling on organisms and plants, and with hierarchical scaling on pests: e.g., *A.indi(A.indi)*, *S.spp\_any(S.litto)*. As a result, we obtain 36 concepts (excluding bottom concepts): 13 concerning protected organisms, 17 concerning plants, and





**Fig. 5.** Concept lattices (excerpt) for GFCA-Ternary (edges as relational attributes)

6 concerning pests. They are grouped in three connected graph patterns: pattern Q1 is specific to the 3-tuple (R.com,W.pro,S.litu), and corresponds to triadic concept TC1; pattern Q3 contains all the other 3-tuples because they are all interconnected; the concepts in pattern Q2 generalize the concepts of the two other patterns.

Figure 5 shows an excerpt of GFCA output<sup>8</sup>. For instance, concept Q2g characterizes the four organisms (out of six) that are protected by plants (Q2d) that protect against S.litu+spp (Q2c). It contains two organisms, A.escu and B.ole, that are not known to be protected against S.litu+spp by any of those plants (they are not in concept Q2a, a sub-concept of Q2g). The concerned plants, i.e. C.opu, C.spp, A.indi, W.pro, are therefore candidates for protection of the two organisms against S.litu+spp. The triadic concepts can be found in the GFCA output. **TC2** corresponds to concept Q2e, which characterizes the set of protecting plants, showing that no plant protects all organisms because its intent contains no specific plant. **TC5** corresponds to concept Q3u, which states that A.indi and C.papa have in common to protect both A.escu (Concept Q3a) and B.ole (Concept Q3b) against S.litto (Concept Q3n). **TC6** corresponds to concept Q3w, which is the most general concept of the organisms that are specifically protected by A.indi. Its extent contains A.escu, B.ole, and Z.mays; and its intent states that the protection is against S.litto+spp.

A more compact representation of the 3-tuples (GFCA-Binary) would be to use plants as binary relations from organisms to pests: e.g. *A.indi(A.escu, S.litto)*. There are several possible justifications for this repre-

<sup>8</sup> Source code and user manual at <https://bitbucket.org/sebferre/graph-fca/>

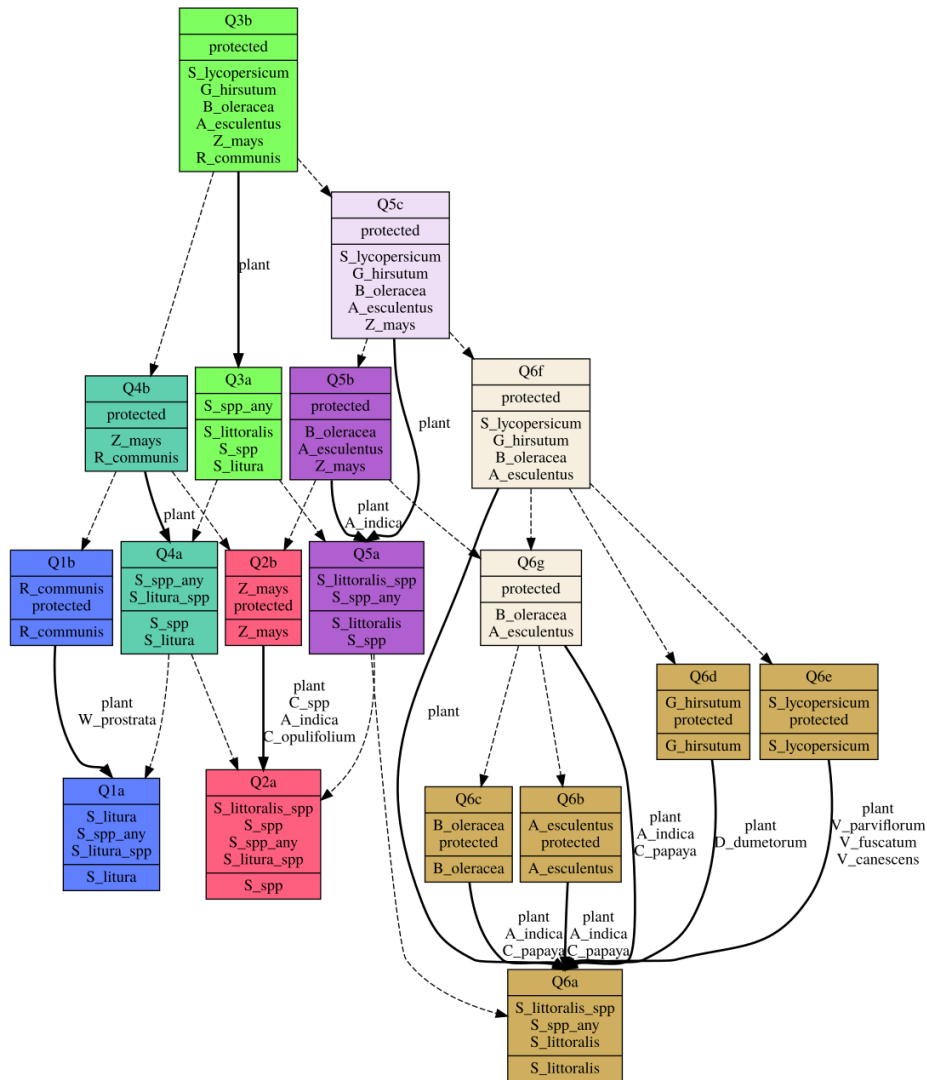


Fig. 6. Concept lattices and graph patterns for GFCA-Binary

sensation. First, protecting plants is the focus of the study, and we prefer concept intents that contain explicit references to plants: using them as attributes is one way to enforce this. Second, the plants concerned can be considered as agents intervening between organisms and pests. This representation resembles that of RDF triples where the middle element is used as an edge label [6]. We applied the same scaling as above for organisms and pests, and simple hierarchical scaling for plants with a binary edge  $plant(X, Z)$  for every 3-tuple  $(X, Y, Z)$ . Figure 6 shows the complete GFCA output, here with a graphical representa-

**Table 3.** Classifications obtained by FCA extensions (NA=Not applicable)

	TCA	RCAC	RCAR	RCAG	GFCa-Ter.	GFCa-Binary
#concepts 2 <sup>nd</sup> dataset	12	59	99	18	63	22
2 <sup>nd</sup> dataset content (organism-plant-pest)		(12-38-9)	protection 50 (8-32-9)	(10-NA-9) + 30 relations	(12-41-10)	(12-NA-10)
#concepts 1 <sup>st</sup> dataset	9	33	50	16	36	18

tion of intents: solid arrows represent the edges of the graph patterns that define concept intents. We obtained 18 concepts grouped in six patterns: 12 concerning organisms, and 6 concerning pests. Protecting plants are displayed as edge labels. The exclusive use of unary and binary attributes makes the output much easier to read. Despite the simplification, triadic concepts can still be found in the GFCa-Binary output, e.g.: **TC2** as the edge from Concept Q3b to Q3a; **TC5** as the edge from Q6g to Q6a; **TC6** as the edge from Q5b to Q5a.

## 6 Results of the Whole Use Case and Discussion

This section presents the results of the complete Spodoptera dataset, which is almost three times larger than the first one (an increase from 12 to 34 3-tuples). We then discuss the different methods.

*Results of the complete Spodoptera dataset.* The second dataset contains 34 3-tuples using six organisms, 30 plants, and four pests, while the first one has 12 3-tuples, using respectively, six, nine and three species. Table 3 presents the number of concepts per classification for each method. For all the classifications, more concepts are built by methods using the second dataset. The increase is almost 50% for TCA and RCAC and twice as high for RCAR and GFCa-Ternary. Despite the notable difference in the number of plants (30 vs 9), the classifications of RCAG and GFCa-Binary contain two and four additional concepts, respectively. This slight difference results from considering plants as relations between organisms and pests, which increases the number of relational attributes. Regarding the other methods, the additional concepts originate from the modeling of the 22 new 3-tuples, which implement the protection of *Z.mays* against the additional pest *S.frugi* using 22 new plants.

*Readability and ease of use.* TCA and GFCa both present their results as a single package. TCA lists the triadic concepts, and GFCa connects them in a single graph. RCA also connects the concepts but a switch between lattices is required to access the overall classification. When a classification is used as a support for exploration, classifying the objects per category, as provided by RCAG, is a good solution. For instance, RCAC and RCAG show that *A.escu* and *B.ole* are equivalent (CProt8) and share more with *Z.mays* than with the other species (CProt9). GFCa-Ternary is more difficult to read than GFCa-Binary because it has more concepts and uses ellipse nodes (not shown in this paper) in addition to concept nodes to represent ternary relationships.

*Modeling symmetry.* TCA, RCAR, and GFCA-Ternary modeling considers the three object categories symmetrically. Thus, 3-tuples sharing characteristics at a given position are recognizable at a glance. RCAG, RCAC, and GFCA-Binary modeling does not consider the three object categories symmetrically. For instance, RCAC classifies protecting plants in relation to targeted pests and protected organisms. Reverse and transitive relationships can be included in the formal contexts to achieve symmetry, but the cost is increased complexity. GFCA-Binary and RCAG handle plants as relation labels instead of objects, implying the absence of plant concepts in the lattice. Two other representations are also possible, where protected organisms or pests are used as relation labels.

*Visibility of concept hierarchies and relational patterns.* Triadic lattices are difficult to fully visualize and understand, even with the projection paradigm used by FCA Tools Bundle [10]. RCA clearly depicts the concept hierarchies, one for each object type, but not the relational patterns. GFCA has three output modes: only showing hierarchies such as RCA, only showing relational patterns, or showing the two combined. The latter is the richest representation but it is more difficult to grasp.

*Extensibility and limits.* All the methods allow the user to return to the original 3-tuples, provided a nominal scaling was used for RCA and GFCA. Like classical FCA, all the implementations of TCA, RCA (with  $\exists$ ,  $\exists\forall$  and  $\exists\supseteq$ ), and GFCA are sensitive to noisy data, in that missing 3-tuples increase the number of concepts. For RCA, noisy data can be processed with percentage quantifiers. TCA is restricted to ternary relations: objects, attributes and conditions cannot have their own description. These limitations reduce the modeling options. RCA and GFCA methods can apply different models depending on the expert's needs. Using  $\exists$  quantifier, RCA is less precise than the other FCA extensions, but this allows experts to hypothesize and suggest new protection solutions. As RCA and GFCA are relatively flexible, adding information on objects of the 3-tuples is possible, such as taxonomy for the species. GFCA is less flexible than RCA, as its patterns are exclusively existential, and the relations are two-way. But GFCA patterns can contain cycles while RCA patterns are (possibly infinite) trees, like class expressions in description logics [1].

## 7 Conclusion

The Knomana Knowledge Base (KKB) is a typical example of a dataset in the environmental domain, characterized by a lack of information concerning the name of some species and a main ternary relationship. In this paper, we address these two questions using a KKB excerpt. We introduce an aggregation hierarchy on species names and abbreviation spp. We define several representations of the ternary relationship to analyze the KKB excerpt with three FCA extensions. The discussion highlights the complementarity of the different approaches.

In a future study, we plan to deepen the qualitative analysis on pests, starting with *Spodoptera* species and to define an analytical process that takes advantage

of the three FCA extensions. We also plan to explore variants of the representations proposed in this paper. Our long-term objective is to provide domain experts with efficient tools to analyze their dataset, not only to extract existing knowledge in a deductive way, but also to assist them in setting up new hypotheses, and suggesting new experiments for bio-pesticides.

## References

1. Baader, F., Calvanese, D., McGuinness, D.L., Nardi, D., Patel-Schneider, P.F. (eds.): *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press (2003)
2. Dolques, X., Braud, A., Huchard, M., Ber, F.L.: Rcaexplore, a FCA based tool to explore relational data. In: *Supp. Proc. of ICFCA 2019*. pp. 55–59 (2019)
3. Ferré, S., Cellier, P.: Graph-FCA: An extension of formal concept analysis to knowledge graphs. *Discrete Applied Mathematics* **273**, 81–102 (2019)
4. Ganter, B., Wille, R.: *Formal Concept Analysis: Mathematical Foundations*. Springer (1999)
5. Hacene, M.R., Huchard, M., Napoli, A., Valtchev, P.: Relational concept analysis: mining concept lattices from multi-relational data. *Ann. Math. Artif. Intell.* **67**(1), 81–108 (2013)
6. Hitzler, P., Krötzsch, M., Rudolph, S.: *Foundations of Semantic Web Technologies*. Chapman & Hall/CRC (2009)
7. HLPE: Agroecological and other innovative approaches for sustainable agriculture and food systems that enhance food security and nutrition. Tech. rep., Food and Agriculture Organization of the United Nations, Rome (2019), High Level Panel of Experts on Food Security and Nutrition of the Committee on World Food Security
8. Jones, T.H., Song, I.: Binary equivalents of ternary relationships in entity-relationship modeling: A logical decomposition approach. *J. Database Manag.* **11**(2), 12–19 (2000)
9. Keip, P., Gutierrez, A., Huchard, M., Le Ber, F., Sarter, S., Silvie, P., Martin, P.: Effects of Input Data Formalisation in Relational Concept Analysis for a Data Model with a Ternary Relation. In: *ICFCA’19*. pp. 191–207 (2019)
10. Kis, L.L., Sacarea, C., Troanca, D.: FCA tools bundle - A tool that enables dyadic and triadic conceptual navigation. In: *Int. Work. "What can FCA do for Artificial Intelligence"?* (FCA4AI@ECAI). *CEUR Work. Proc.*, vol. 1703, pp. 42–50 (2016)
11. Lehmann, F., Wille, R.: A Triadic Approach to Formal Concept Analysis. In: *ICCS ’95*. pp. 32–43 (1995)
12. Martin, P., Ferré, S., Gutierrez, A., Huchard, M., Keip, P., Silvie, P.: Dataset on Spodoptera used to conduct the practical comparison of FCA extensions to Model Indeterminate Value of Ternary Data (2020). <https://doi.org/doi:10.18167/DVN1/VNCZYA>, CIRAD Dataverse, V1
13. Martin, P., Sarter, S., Tagne, A., Ilboudo, Z., Marnotte, P., Silvie, P.: Knowing the Useful Plants for Organic agriculture according to literature: Building and Exploring a Knowledge Base for Plant and Animal Health. In: *African organic conference*. pp. 137–141 (2018)