Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI 2015)

Membership Constraints in Formal Concept Analysis

Sebastian Rudolph† and **Christian Săcărea**‡ and **Diana Troancă**‡
†Technische Universität Dresden, Germany
sebastian.rudolph@tu-dresden.de
‡Universitatea Babes Bolyai, Romania
{csacarea,dianat}@cs.ubbcluj.ro

Abstract

Formal Concept Analysis (FCA) is a prominent field of applied mathematics using object-attribute relationships to define formal concepts – groups of objects with common attributes – which can be ordered into conceptual hierarchies, so-called concept lattices. We consider the problem of satisfiability of membership constraints, i.e., to determine if a formal concept exists whose object and attribute set include certain elements and exclude others. We analyze the computational complexity of this problem in general and for restricted forms of membership constraints. We perform the same analysis for generalizations of FCA to incidence structures of arity three (objects, attributes and conditions) and higher. We present a generic answer set programming (ASP) encoding of the membership constraint satisfaction problem, which allows for deploying available highly optimized ASP tools for its solution. Finally, we discuss the importance of membership constraints in the context of navigational approaches to data analysis.

1 Introduction

Conceptual Knowledge Processing and Representation is a particular approach to knowledge management, acknowledging the constitutive role of thinking, arguing and communicating human beings in dealing with knowledge and its processing. The term processing also underlines the fact that obtaining or approximating knowledge is a process which should always be conceptual in the above sense. The methods of Conceptual Knowledge Processing have been introduced and discussed by Rudolf Wille in [Wille, 2006], based on the pragmatic philosophy of Charles Sanders Peirce, continued by Karl-Otto Apel and Jürgen Habermas.

Wille defines Conceptual Knowledge Processing as an applied discipline dealing with knowledge which is constituted by conscious reflexion, discursive argumentation and human communication on the basis of cultural background, social conventions and personal experiences. Its main aim is to develop and maintain formal methods and instruments for processing information and knowledge which support rational thought, judgment and action of human beings and therewith promote critical discourse (see also [Wille, 1994; 1997; 2000]).

The mathematical theory underlying Conceptual Knowledge Processing is Formal Concept Analysis, providing a powerful and elegant mathematical tool for understanding and investigating knowledge, based on a set-theoretical semantics, comprising methods for representation, acquiring, and retrieval of knowledge, as well as for further theory building in several other domains of science.

Formal Concept Analysis (FCA) appeared at the end of the 1980's in order to restructure classical lattice theory into a form that is suitable for applications in data analysis. The fundamental data structure FCA uses is a *formal context*, which exploits the fact that data is quite often represented by incidence structures relating objects and attributes. FCA provides also a mathematization of the traditional, philosophical understanding of a *concept* as a unit of thought consisting of an extent (the set of objects falling under the concept) and an intent (the set of attributes characterizing the concept). Using mathematical operations, concepts are computed from the object-attribute data table. They can be naturally ordered, resulting in a conceptual hierarchy, called *concept lattice*. The entire information stored in a formal context is preserved by this operation and the concept lattice is the basis for further data analysis. It can be represented graphically in order to allow navigation among concepts, as well as to support communication. Different algebraic methods can be used in order to study its structure and to compute data dependencies. FCA also provides elegant methods to significantly reduce the effort of mining association rules.

Classical FCA was extended by Wille and Lehmann to the triadic case, featuring a ternary (objects vs. attributes vs. conditions) instead of a binary (objects vs. attributes) incidence relation [Lehmann and Wille, 1995], leading to the notions of *tricontext* and *triconcept*. This extension has been successfully used in inherently triadic scenarios such as collaborative tagging [Jäschke *et al.*, 2008].

Nevertheless, if the number of concepts is very large, a holistic graphical representation may become inefficient and unwieldy. Note that the number of concepts may be exponentially in the size of the underlying (tri)context.

Hence, a way of narrowing down the set of "interesting" concepts by specifying criteria appears as a crucial feature of conceptual knowledge management applications, in order to focus exactly on the data subset one is interested to explore or start exploration from. As a straightforward form of such criteria, we introduce *membership constraints* which specify that a formal concept's extent or intent must include certain elements and exclude others. The question of satisfiability of such membership constraints, i.e., to determine if there exists at all a formal concept is the starting point of our current research. In this paper, we analyze the computational complexity of this problem, both for the classical dyadic case and for higher arity generalizations of FCA, first for triadic data sets and then for the n-adic case. Moreover, we also discuss a generic answer set programming (ASP) encoding of membership constraint problems, which allows for deploying available highly optimized ASP tools for its solution. Finally, we turn our attention to the question wherefrom the entire problem setting started, namely we discuss the importance of membership constraints in the context of navigational approaches to data analysis and provide some conclusions of our work.

2 Preliminaries

2.1 Formal Concept Analysis

In the following, we briefly sketch some basic notions about FCA. For more, please refer to [Ganter and Wille, 1999].

Definition 1. *A* formal context *is a triple* $\mathbb{K} = (G, M, I)$ *with* G *and* M *being sets called* objects *and* attributes*, respectively, and* $I \subseteq G \times M$ *the binary* incidence relation *where* gIm *means that object* g *has attribute* m*.*

Finite formal contexts can be represented as cross-tables, the rows of which are representing objects, the columns attributes, while the incidence relation is represented by crosses in that table.

	m_1	m_2	m_3	m_4	m_5	m_6
g_1	×					
g_2		×		×		
g_3			×	×		
g_4				×		
g_5		×	×		×	
g_6						×

Figure 1: Formal context as a cross-table

Definition 2. *For a set* $A \subseteq G$ *of objects we define the derivation operator* $A^I := \{m \mid gIm \text{ for all } g \in A\}$ *and for a set* $B \subseteq M$ *of attributes, we analogously define* $B^I = \{g \mid gIm \text{ for all } m \in B\}$*. A* formal concept *of a context* $\mathbb{K}$ *is a pair* (A, B) *with* extent $A \subseteq G$ *and* intent $B \subseteq M$ *satisfying* $A^I = B$ *and* $B^I = A$*. We denote the set of formal concepts of the context* $\mathbb{K}$ *by* $\mathfrak{B}(\mathbb{K})$*.*

An alternative, useful way of characterizing formal concepts is that $A \times B \subseteq I$ and A, B are maximal w.r.t. this property, i.e., for every $C \supseteq A$ and $D \supseteq B$ with $C \times D \subseteq I$ must hold $C = A$ and $D = B$.

Definition 3. *If* $(A, B), (C, D) \in \mathfrak{B}(\mathbb{K})$*, we say that* (A, B) *is a* subconcept *of* (C, D) *(or equivalently,* (C, D) *is a* superconcept *of* (A, B)*), and we write* $(A, B) \leq (C, D)$ *if and only if* $A \subseteq C$ *(*$\Leftrightarrow D \subseteq B$*).*

The set $\mathfrak{B}(\mathbb{K})$ of formal concepts, ordered by the subconcept-superconcept relationship is a complete lattice and can be graphically represented as an order diagram.

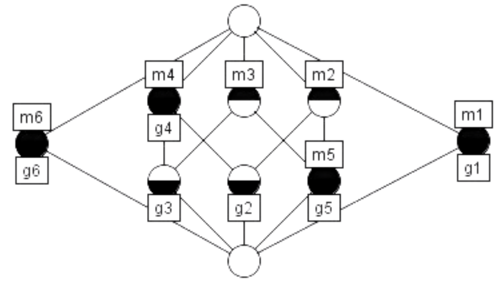

Figure 2: Concept lattice of the context in Figure 1

F. Lehmann and R. Wille extended in [Lehmann and Wille, 1995] the theory of FCA to deal with threedimensional data. This has been called Triadic FCA (3FCA), where objects are related to attributes and conditions.

Definition 4. *A* tricontext *is a quadruple* $\mathbb{K} = (G, M, B, Y)$ *with* G*,* M*, and* B *being sets called* objects*,* attributes*, and* conditions*, respectively, and* $Y \subseteq G \times M \times B$ *the ternary* incidence relation *where* $(g, m, b) \in Y$ *means that object* g *has attribute* m *under condition* b*.*

Finite tricontexts can be represented as three-dimensional cross-tables, which are typically displayed in "slices", e.g.:

b_1	m_1	m_2	m_3	m_4	b_2	m_1	m_2	m_3	m_4	b_3	m_1	m_2	m_3	m_4
g_1	×	×	×	×	g_1	×	×			g_1	×	×	×	
g_2	×		×	×	g_2		×	×	×	g_2		×	×	×
g_3	×	×		×	g_3	×	×	×	×	g_3		×	×	×
g_4	×	×	×		g_4	×	×	×	×	g_4	×	×	×	×

Definition 5. *A* triconcept *of a tricontext* $\mathbb{K}$ *is a triple* (A_1, A_2, A_3) *with* extent $A_1 \subseteq G$*,* intent $A_2 \subseteq M$*, and* modus $A_3 \subseteq B$ *satisfying* $A_1 \times A_2 \times A_3 \subseteq Y$ *and for every* $C_1 \supseteq A_1$*,* $C_2 \supseteq A_2$*,* $C_3 \supseteq A_3$ *that satisfy* $C_1 \times C_2 \times C_3 \subseteq Y$ *holds* $C_1 = A_1$*,* $C_2 = A_2$*, and* $C_3 = A_3$*. We denote by* $\mathfrak{T}(\mathbb{K})$ *the set of all triconcepts of* $\mathbb{K}$*.*

With the rise of folksonomies as data structure of social resource sharing systems, triadic FCA was directly applied in the study of folksonomies [Jäschke *et al.*, 2008]. Efficient algorithms to determine all (or all frequent) triconcepts of a tricontext have been developed. However, a visualization that would be as intuitive as concept lattices for classical FCA has remained elusive for the triadic case. Initial investigations into interactive ways of browsing the space of triconcepts have been made [Rudolph *et al.*, 2015].

2.2 Complexity Theory

We assume the reader to be familiar with complexity theory [Papadimitriou, 1994] and, in particular, the complexity classes AC_0 and NP.

We briefly recap that AC_0 (problems solvable by Boolean circuits of polynomial size and constant depth) coincides with expressibility by first-order formulae [Immerman, 1999]. It is worth noting that such problems can be solved in logarithmic space.

NP is the class of problems solvable by a nondeterministic Turing machine in polynomial time. We will provide here a traditional, prototypical NP-complete problem, which we will use later to show NP-hardness of certain problems.

> **problem**: 3SAT
> **input**: family $\mathcal{L} = \{L_1, \dots, L_n\}$ of 3-element sets L_i of literals of the form p or $\neg p$.
> **output**: YES in case the Boolean formula $\varphi_\mathcal{L} := \bigwedge_{\{\ell_1,\ell_2,\ell_3\}\in\mathcal{L}}(\ell_1 \vee \ell_2 \vee \ell_3)$ is satisfiable, NO otherwise.

Example 1. *Consider $\mathcal{L} = \{L_1, L_2, L_3\}$ with $L_1 = \{r, s, \neg q\}$, $L_2 = \{s, \neg q, \neg r\}$, and $L_3 = \{\neg q, \neg r, \neg s\}$. The corresponding 3SAT problem amounts to checking if $\varphi_\mathcal{L} = (\neg q \vee r \vee s) \wedge (\neg q \vee \neg r \vee s) \wedge (\neg q \vee \neg r \vee \neg s)$ is satisfiable. The answer is yes, since for the valuation $v = \{q \mapsto true, r \mapsto false, s \mapsto true\}$ the formula $\varphi_\mathcal{L}$ evaluates to true.*

3 Membership Constraints

In this section, we introduce membership constraints for formal concepts. These constraints express which objects and attributes should be part of the concept and which should not.

Definition 6. *A* membership constraint *on a formal context $\mathbb{K} = (G, M, I)$ is a quadruple $\mathbb{C} = (G^+, G^-, M^+, M^-)$ with $G^+ \subseteq G$ called* required objects*, $G^- \subseteq G$ called* forbidden objects*, $M^+ \subseteq M$ called* required attributes*, and $M^- \subseteq M$ called* forbidden attributes.

A formal concept (A, B) of $\mathbb{K}$ is said to satisfy *such a membership constraint if all the following conditions hold: $G^+ \subseteq A$, $G^- \cap A = \emptyset$, $M^+ \subseteq B$, $M^- \cap B = \emptyset$.*

A membership constraint is said to be satisfiable with respect to $\mathbb{K}$, if it is satisfied by one of its formal concepts.

Next we define the decision problem of membership constraint satisfiability (MCSAT).

> **problem** MCSAT
> **input**: formal context $\mathbb{K}$, membership constraint $\mathbb{C}$
> **output**: YES if $\mathbb{C}$ satisfiable w.r.t. $\mathbb{K}$, NO otherwise.

In the general case, the complexity of the (MCSAT) problem turns out to be intractable.

Theorem 7. *MCSAT is* NP*-complete, even when restricting to membership constraints of the form $(\emptyset, G^-, \emptyset, M^-)$*

Proof. NP membership is straightforward: after guessing a pair (A, B) from $2^G \times 2^M$, it can be checked in polynomial time if (A, B) is a formal concept of $\mathbb{K}$ and if it satisfies $\mathbb{C}$.

We prove NP hardness via a reduction from 3SAT. Given a set $\mathcal{L} = \{L_1, \dots, L_n\}$ of propositional literal sets over the set $\{p_1, \dots, p_k\}$ of propositional variables, define the formal context $\mathbb{K}_\mathcal{L} = (G, M, I)$ with

- $G = \mathcal{L} \cup P^+ \cup P^-$ with $P^+ = \{p_1, \dots, p_k\}$ and $P^- = \{\neg p_1, \dots, \neg p_k\}$,
- $M = P^+ \cup P^- \cup \tilde{P}$ with $\tilde{P} = \{\tilde{p}_1, \dots, \tilde{p}_k\}$,
- $I := \{(L_i, m) \mid L_i \in \mathcal{L}, m \in M \setminus L_i\}$
  $\cup\ \mathcal{L} \times \tilde{P}$
  $\cup\ \{(l_1, l_2) \mid l_1, l_2 \in P^+ \cup P^-, l_1 \neq l_2\}$
  $\cup\ \{(p_i, \tilde{p}_j) \mid i \neq j\}$
  $\cup\ \{(\neg p_i, \tilde{p}_j) \mid i \neq j\}$

Furthermore, let $\mathbb{C}_\mathcal{L}$ denote the membership constraint $(\emptyset, \mathcal{L}, \emptyset, \tilde{P})$.

Note that both $\mathbb{K}_\mathcal{L}$ and $\mathbb{C}_\mathcal{L}$ can be computed in polynomial time and are of polynomial size with respect to $\|\mathcal{L}\|$.

We will now show that $\mathcal{L}$ is satisfiable exactly if $\mathbb{C}$ is satisfiable w.r.t. $\mathbb{K}$.

"$\Rightarrow$": if $\mathcal{L}$ is satisfiable there must be a valuation $v : \{p_1, \dots p_k\} \to \{true, false\}$ under which $\mathcal{L}$ evaluates to true. Let L_v be the set of literals such that $p \in L_v$ whenever $v(p) = true$ and $\neg p \in L_v$ whenever $v(p) = false$. We next show that $((P^+ \cup P^-) \setminus L_v, L_v)$ is a formal concept of $\mathbb{K}_\mathcal{L}$. On one hand we have: $((P^+ \cup P^-) \setminus L_v)^I = \{m \in P^+ \cup P^- \mid m \notin (P^+ \cup P^-) \setminus L_v\} \cup \{\tilde{p} \in \tilde{P} \mid p, \neg p \notin (P^+ \cup P^-) \setminus L_v\} = L_v$. On the other hand $L_v^I = \{L_i \mid L_v \subseteq M \setminus L_i\} \cup \{l \mid l \notin L_v\} = \emptyset \cup (P^+ \cup P^-) \setminus L_v$.

Next, we observe that this formal concept satisfies $\mathbb{C}_\mathcal{L}$, since none of the L_i are contained in the extent and none of the $\tilde{p}_j$ are contained in its intent. Therefore, we have found a concept witnessing the satisfiability of $\mathbb{C}_\mathcal{L}$ w.r.t. $\mathbb{K}_\mathcal{L}$.

"$\Leftarrow$": Assume $\mathbb{C}_\mathcal{L}$ is satisfiable w.r.t. $\mathbb{K}_\mathcal{L}$. Then there must be a formal concept (A, B) of $\mathbb{K}_\mathcal{L}$ with $\mathcal{L} \cap A = \emptyset$ as well as $\tilde{P} \cap B = \emptyset$. Observe that A must contain one of p or $\neg p$ for each propositional variable p, since otherwise $\tilde{p} \in B$ would hold. Consequently B cannot contain both p and $\neg p$ for any propositional variable p. Moreover, for every $L_i \in \mathcal{L}$ there must be one $l \in B \cap (P^+ \cup P^-)$ with $l \in L_i$. Next, let $\hat{B} := P^+ \cap B \cup \{\neg p \mid p \notin B\}$. By our observation above, we know that $B \subseteq \hat{B}$ therefore $\hat{B}$ still contains at least one literal from every L_i. On the other hand, $\hat{B}$ directly corresponds to a valuation $v_{\hat{B}} : \{p_1, \dots p_k\} \to \{true, false\}$ mapping p_i to $true$ if $p_i \in \hat{B}$ and to $false$ if $\neg p_i \in \hat{B}$. Consequently, $v_{\hat{B}}$ is a valuation making φ true and hence showing that it is satisfiable. $\square$

The following example demonstrates the reduction of 3SAT to MCSAT deployed in the previous proof.

Example 2. *The 3SAT problem from Example 1 can be reduced to the question if the membership constraint $(\emptyset, \{L_1, L_2, L_3\}, \emptyset, \{\tilde{q}, \tilde{r}, \tilde{s}\})$ is satisfiable in the following context:*

	q	r	s	$\neg q$	$\neg r$	$\neg s$	$\tilde{q}$	$\tilde{r}$	$\tilde{s}$
$\{r, s, \neg q\}$	×				×	×	×	×	×
$\{s, \neg q, \neg r\}$	×	×				×	×	×	×
$\{\neg q, \neg r, \neg s\}$	×	×	×				×	×	×
q		×	×	×	×	×		×	×
r	×		×	×	×	×	×		×
s	×	×		×	×	×	×	×	
$\neg q$	×	×	×		×	×		×	×
$\neg r$	×	×	×	×		×	×		×
$\neg s$	×	×	×	×	×		×	×	

This is the case as witnessed by the formal concept $(\{r, \neg q, \neg s\}, \{q, s, \neg r\})$.

When analyzing the problem further it turns out that the simultaneous presence of forbidden objects and forbidden attributes in a membership constraint is the (only) reason for the established intractability. If one of the sets becomes empty, the complexity drops to a very pleasant level.

Theorem 8. *When restricted to membership constraints of the form* $(G^+, \emptyset, M^+, M^-)$ *or* $(G^+, G^-, M^+, \emptyset)$ *MCSAT is in* AC_0.

Proof. We show the claim for constraints of the form $\mathbb{C} = (G^+, \emptyset, M^+, M^-)$, the other case follows by duality. First observe that $({M^+}^I, {M^+}^{II})$ is a formal concept of $\mathbb{K}$ and it is subset-maximal w.r.t. its extent and subset-minimal w.r.t. its intent among all formal concepts whose intent contains M^+. Therefore, $(G^+, \emptyset, M^+, M^-)$ is satisfiable w.r.t. $\mathbb{K}$ if and only if it is satisfied by $({M^+}^I, {M^+}^{II})$. By definition, this is the case iff (1) $G^+ \subseteq {M^+}^I$ and (2) ${M^+}^{II} \cap M^- = \emptyset$. Statement (1) can be rephrased into the condition $G^+ \times M^+ \subseteq I$, while Statement (2) is equivalent to the condition that for every $m \in M^-$ there exists some $g \in (M^+)^I$ with $(g, m) \notin I$.

We now define a first-order-logic interpretation $\mathcal{I}_{\mathbb{K},\mathbb{C}} = (\Delta, \cdot^{\mathcal{I}})$ over the predicates p_G, p_M, p_{G^+}, p_{M^+}, p_{M^-} (all unary) and the binary predicate p_I as follows: $\Delta = G \cup M$, for every $X \in \{G, M, G^+, M^+, M^-\}$ we let $p_X^{\mathcal{I}} = X$, and $p_I^{\mathcal{I}} = I$. Obviously, $\mathcal{I}$ is an immediate representation of the MCSAT problem. Now the above formulated Statement (1) can be expressed by the first-order formula φ_1 defined as

$$\forall x, y.(p_{G^+}(x) \wedge p_{M^+}(y) \rightarrow p_I(x, y)),$$

while Statement (2) can be expressed by φ_2 defined as

$$\forall x.p_{M^-}(x) \rightarrow \exists y.(\forall z.(p_{M^+}(z) \rightarrow p_I(y, z)) \wedge \neg p_I(y, x)).$$

Consequently, satisfiability of $\mathbb{C}$ w.r.t. $\mathbb{K}$ coincides with the satisfaction of the fixed first-order-logic formula $\varphi = \varphi_1 \wedge \varphi_2$ in $\mathcal{I}_{\mathbb{K},\mathbb{C}}$. By the corresponding result from descriptive complexity theory [Immerman, 1999], we can conclude that the considered restricted version of MCSAT is in AC_0. $\square$

Finally, if only required objects or required attributes are given, MCSAT becomes trivial.

Theorem 9. *When restricted to membership constraints of the form* $(G^+, \emptyset, \emptyset, \emptyset)$ *or* $(\emptyset, \emptyset, M^+, \emptyset)$ *MCSAT is trivially true.*

Proof. For the form $(G^+, \emptyset, \emptyset, \emptyset)$, note that (G, G') is always a formal concept and always satisfies such a constraint. The other case follows by duality. $\square$

4 Membership Constraints in Triadic FCA

Next we define and investigate membership constraints and the corresponding satisfiability problem for the triadic case.

Definition 10. *A* triadic membership constraint *on a tricontext* $\mathbb{K} = (G, M, B, Y)$ *is a sextuple* $\mathbb{C} = (G^+, G^-, M^+, M^-, B^+, B^-)$ *with* $G^+ \subseteq G$ *called* required objects, $G^- \subseteq G$ *called* forbidden objects, $M^+ \subseteq M$ *called* required attributes, $M^- \subseteq M$ *called* forbidden attributes, $B^+ \subseteq B$ *called* required conditions, *and* $B^- \subseteq B$ *called* forbidden conditions

A triconcept (A_1, A_2, A_3) *of* $\mathbb{K}$ *is said to* satisfy *such a membership constraint if all the following conditions hold:* $G^+ \subseteq A_1$, $G^- \cap A_1 = \emptyset$, $M^+ \subseteq A_2$, $M^- \cap A_2 = \emptyset$, $B^+ \subseteq A_3$, $B^- \cap A_3 = \emptyset$.

A triadic membership constraint is said to be satisfiable with respect to $\mathbb{K}$, *if it is satisfied by one of its triconcepts.*

problem TMCSAT
input: tricontext $\mathbb{K}$, triadic membership constraint $\mathbb{C}$
output: YES if $\mathbb{C}$ satisfiable w.r.t. $\mathbb{K}$, NO otherwise.

Interestingly, in the triadic case, we find two possible sources for intractability. One of them (two nonempty forbidden sets) is tightly related to the case discussed in the previous section, while the other one (required and forbidden set of the same type) is tractable for classical FCA and becomes intractable only when going triadic.

Theorem 11. *TMCSAT is* NP*-complete, even when restricting to triadic membership constraints of the following forms:*

- $(\emptyset, G^-, \emptyset, M^-, \emptyset, \emptyset)$, $(\emptyset, G^-, \emptyset, \emptyset, \emptyset, B^-)$, $(\emptyset, \emptyset, \emptyset, M^-, \emptyset, B^-)$,
- $(G^+, G^-, \emptyset, \emptyset, \emptyset, \emptyset)$, $(\emptyset, \emptyset, M^+, M^-, \emptyset, \emptyset)$, $(\emptyset, \emptyset, \emptyset, \emptyset, B^+, B^-)$.

Proof. NP membership is straightforward: after guessing a triple (A_1, A_2, A_3) from $2^G \times 2^M \times 2^B$, it can be checked in polynomial time if (A_1, A_2, A_3) is a triconcept of $\mathbb{K}$ and if it satisfies $\mathbb{C}$.

We proceed by showing hardness for the restricted cases.

Given some (dyadic) formal context $\mathbb{K} = (G, M, I)$, we define its triadic version $\mathbb{T}(\mathbb{K}) = (G, M, \{*\}, I \times \{*\})$. Then, the set of all triconcepts of $\mathbb{T}(\mathbb{K})$ is $\{(G, M, \emptyset)\} \cup \{(A_1, A_2, \{*\}) \mid (A_1, A_2) \text{ concept of } \mathbb{K}\}$. Therefore, every MCSAT problem with context $\mathbb{K}$ and constraint $(\emptyset, G^-, \emptyset, M^-)$ can be reduced to the TMCSAT problem with tricontext $\mathbb{T}(\mathbb{K})$ and $(\emptyset, G^-, \emptyset, M^-, \emptyset, \emptyset)$. Since the former problem is NP-complete due to Theorem 7, the latter must be NP-hard. By symmetry, this argument carries over to constraints of the form $(\emptyset, G^-, \emptyset, \emptyset, \emptyset, B^-)$ and $(\emptyset, \emptyset, \emptyset, M^-, \emptyset, B^-)$.

Next, we show hardness for constraints of the form $(G^+, G^-, \emptyset, \emptyset, \emptyset, \emptyset)$. Again, we do so by a reduction from 3SAT. Given a set $\mathcal{L} = \{L_1, \ldots, L_n\}$ of propositional literal sets over the set $\{p_1, \ldots, p_k\}$ of propositional variables, define the tricontext $\mathbb{K}_{\mathcal{L}} = (G, M, B, Y)$ with

- $G = \{*\} \cup \mathcal{L}$,
- $M = \{*, p_1, \ldots, p_k\}$
- $B = \{*, \neg p_1, \ldots, \neg p_k\}$
- $Y = G \times M \times B \setminus \big(\{(*, p_i, \neg p_i) \mid 1 \leq i \leq k\}$ $\cup \{(L_j, p_i, *) \mid p_i \in L_j\} \cup \{(L_j, *, \neg p_i) \mid \neg p_i \in L_j\}\big)$

Furthermore, let $\mathbb{C}_{\mathcal{L}}$ denote the membership constraint $(\{*\}, \mathcal{L}, \emptyset, \emptyset, \emptyset, \emptyset)$.

Note that both $\mathbb{K}_{\mathcal{L}}$ and $\mathbb{C}_{\mathcal{L}}$ can be computed in polynomial time and are of polynomial size with respect to $\|\mathcal{L}\|$.

We will now show that $\mathcal{L}$ is satisfiable exactly if $\mathbb{C}$ is satisfiable w.r.t. $\mathbb{K}$.

"⇒": if $\mathcal{L}$ is satisfiable there must be a valuation $v : \{p_1, \ldots p_k\} \to \{true, false\}$ under which $\mathcal{L}$ evaluates to true. Let L_v be the set of literals such that $p \in L_v$ whenever $v(p) = true$ and $\neg p \in L_v$ whenever $v(p) = false$. We next show that $(A_1, A_2, A_3) = (\{*\}, \{*\} \cup (L_v \cap M), \{*\} \cup (L_v \cap B))$ is a triconcept of $\mathbb{K}_{\mathcal{L}}$: First, $A_1 \times A_2 \times A_3 \subseteq Y$, since L_v cannot both contain some p_i and $\neg p_i$ as it stems from a valuation. We now show that (A_1, A_2, A_3) is also maximal, i.e., no component can be extended while maintaining $A_1 \times A_2 \times A_3 \subseteq Y$. Since L_v already contains for every $i \in \{1, \ldots, n\}$ either p_i or $\neg p_i$, extending A_2 or A_3 would lead to some i satisfying $p_i \in A_2$ and $\neg p_i \in A_3$ which contradicts $(*, p_i, \neg p_i) \notin Y$. It remains to show that A_1 cannot be extended. Toward a contradiction, suppose it can, i.e., for some $L_j \in \mathcal{L}$ holds $\{L_j\} \times A_2 \times A_3 \subseteq Y$. However, by construction, we know that $L_j \cap L_v$ is non-empty. Assuming there is some $p_i \in L_j \cap L_v$, we conclude $p_i \in A_2$ and thus $(L_j, p_i, *) \in Y$ which is wrong by construction. Assuming there is some $\neg p_i \in L_j \cap L_v$, we conclude $\neg p_i \in A_3$ and thus $(L_j, *, \neg p_i) \in Y$ which, again, is wrong by construction. Hence A_1 cannot be extended either and (A_1, A_2, A_3) is indeed a triconcept, which obviously also satisfies $\mathbb{C}_{\mathcal{L}}$.

"⇐": Assume $\mathbb{C}_{\mathcal{L}}$ is satisfiable w.r.t. $\mathbb{K}_{\mathcal{L}}$. Then there must be a triconcept $(\{*\}, A_2, A_3)$ of $\mathbb{K}_{\mathcal{L}}$. Since $(*, p_i, \neg p_i) \notin Y$, we know that for no p_i holds $p_i \in A_2$ and $\neg p_i \in A_3$ at the same time. On the other hand, by maximality, for every p_i one of $p_i \in A_2$ and $\neg p_i \in A_3$ must hold. Therefore, we can define a valuation v by letting $v(p_i) = true$ whenever $p_i \in A_2$ and letting $v(p_i) = false$ whenever $\neg p_i \in A_3$. We now show that v is a valuation mapping $\mathcal{L}$ to $true$ and thus witnessing satisfiability of $\mathcal{L}$. By assumption, $(\{*\}, A_2, A_3)$ is maximal, thus – by maximality of the first component – for every $L_j \in \mathcal{L}$ must hold that $\{L_j\} \times A_2 \times A_3 \not\subseteq Y$. Then, by construction of $\mathbb{K}_{\mathcal{L}}$ there must be either some $p_i \in L_j$ with $p_i \in A_2$ or there must be some $\neg p_i \in L_j$ with $\neg p_i \in A_3$. In any case, this means that L_j is mapped to $true$ under v. Since the same argument applies to every $L_j \in \mathcal{L}$ we find that v is indeed a valuation witnessing the satisfiability of $\mathcal{L}$. □

The subsequent example demonstrates the 3SAT to 3MCSAT reduction for the new intractable case.

Example 3. *The 3SAT problem from Example 1 can be reduced to the question if the membership constraint* $(\{*\}, \{L_1, L_2, L_3\}, \emptyset, \emptyset, \emptyset, \emptyset)$ *is satisfiable in the following tricontext:*

$*$	$*$	q	r	s
$*$	×	×	×	×
$\neg q$	×		×	×
$\neg r$	×	×		×
$\neg s$	×	×	×	

L_1	$*$	q	r	s
$*$	×	×		
$\neg q$		×	×	×
$\neg r$	×	×	×	×
$\neg s$	×	×	×	×

L_2	$*$	q	r	s
$*$	×	×	×	
$\neg q$		×	×	×
$\neg r$		×	×	×
$\neg s$	×	×	×	×

L_3	$*$	q	r	s
$*$	×	×	×	×
$\neg q$		×	×	×
$\neg r$		×	×	×
$\neg s$		×	×	×

This is the case as witnessed by the triconcept $(\{*\}, \{*, q, s\}, \{*, \neg r\})$.

We finish the section by showing that excluding the critical cases discussed above, we regain tractability. We also identify the cases when the problem becomes trivial.

Theorem 12. *TMCSAT is in* AC_0 *when restricting to membership constraints of the forms* $(\emptyset, G^-, M^+, \emptyset, B^+, \emptyset)$, $(G^+, \emptyset, \emptyset, M^-, B^+, \emptyset)$, *and* $(G^+, \emptyset, M^+, \emptyset, \emptyset, B^-)$.

Proof. For $\mathbb{C}$ of the form $(\emptyset, G^-, M^+, \emptyset, B^+, \emptyset)$, note that (G_U, M, B) with $G_U = \{g \mid \{g\} \times M \times B \subseteq Y\}$ is a triconcept and for every triconcept (A_1, A_2, A_3) of $\mathbb{K}$ hold $G_U \subseteq A_1$ and (trivially) $A_2 \subseteq M$ as well as $A_3 \subseteq B$. Therefore $\mathbb{C}$ is satisfiable w.r.t. $\mathbb{K}$ if and only if (G_U, M, B) satisfies it. To check the latter, it suffices to check if $G_U \cap G^- = \emptyset$ which amounts to checking if for every $g \in G^-$ there are $m \in M$ and $b \in B$ with $(g, m, b) \notin Y$. This, in turn is equivalent to $\mathcal{I}_{\mathbb{K},\mathbb{C}}$ satisfying the first-order formula

$$\forall x. p_{G^-}(x) \to \exists y, z. (p_M(y) \land p_B(z) \land \neg p_Y(x, y, z)).$$

Again we can invoke descriptive complexity theory [Immerman, 1999] to conclude that the considered restricted version of TMCSAT is in AC_0.

AC_0 membership for the other forms of $\mathbb{C}$ follows by symmetry. □

Theorem 13. *TMCSAT is trivially true when restricting to membership constraints of the forms* $(\emptyset, \emptyset, M^+, \emptyset, B^+, \emptyset)$, $(G^+, \emptyset, \emptyset, \emptyset, B^+, \emptyset)$, *or* $(G^+, \emptyset, M^+, \emptyset, \emptyset, \emptyset)$.

Proof. For the form $(\emptyset, \emptyset, M^+, \emptyset, B^+, \emptyset)$, note that the triconcept (G_U, M, B) with $G_U = \{g \mid \{g\} \times M \times B \subseteq Y\}$ satisfies any constraint of this form, thus satisfiability is always ensured. The other cases follow by symmetry. □

5 Membership Constraints in n-adic FCA

Classical FCA and triadic FCA can be seen as two instances of a general framework that we call n-adic FCA. We provide the corresponding definitions and observe that the already identified causes of intractability are the only ones also when increasing the arity of the incidence relation further.

Definition 14. *An* n-context *is an* $(n{+}1)$*-tuple* $\mathbb{K} = (K_1, \ldots, K_n, R)$ *with* $K_1, \ldots, K_n$ *being sets, and* $R \subseteq K_1 \times \ldots \times K_n$ *the* n*-ary* incidence relation.

An n-concept *of an* n*-context* $\mathbb{K}$ *is an* n*-tuple* $(A_1, \ldots, A_n)$ *satisfying* $A_1 \times \ldots \times A_n \subseteq R$ *and for every* n*-tuple* $(C_1, \ldots, C_n)$ *with* $A_i \supseteq C_i$ *for all* $i \in \{1, \ldots, n\}$*, satisfying* $C_1 \times \ldots \times C_n \subseteq R$ *holds* $C_i = A_i$ *for all* $i \in \{1, \ldots, n\}$.

Definition 15. *A* n-adic membership constraint *on a* n*-context* $\mathbb{K} = (K_1, \ldots, K_n, R)$ *is a* $2n$*-tuple* $\mathbb{C} = (K_1^+, K_1^-, \ldots, K_n^+, K_n^-)$ *with* $K_i^+ \subseteq K_i$ *called* required sets *and* $K_i^- \subseteq K_i$ *called* forbidden sets.

An n*-concept* $(A_1, \ldots, A_n)$ *of* $\mathbb{K}$ *is said to* satisfy *such a membership constraint if* $K_i^+ \subseteq A_i$ *and* $K_i^- \cap A_i = \emptyset$ *hold for all* $i \in \{1, \ldots, n\}$.

An n*-adic membership constraint is said to be satisfiable with respect to* $\mathbb{K}$*, if it is satisfied by one of its* n*-concepts.*

problem nMCSAT
input: n-context $\mathbb{K}$, n-adic membership constraint $\mathbb{C}$
output: YES if $\mathbb{C}$ satisfiable w.r.t. $\mathbb{K}$, NO otherwise.

It turns out that the triadic case exhibits all necessary information needed to settle the general case, taking into account some straightforward adaptations, hence the following theorem is immediate.

Theorem 16. *For a fixed $n > 2$, the nMCSAT problem is*

- NP-*complete for any class of constraints that allows for*
  - *the arbitrary choice of at least two forbidden sets or*
  - *the arbitrary choice of at least one forbidden set and the corresponding required set,*
- *in* AC_0 *for the class of constraints with at most one forbidden set and the corresponding required set empty,*
- *trivially true for the class of constraints with all forbidden sets and at least one required set empty.*

6 Encoding in Answer Set Programming

Given that satisfiability of membership constraints can in general be NP-complete, it is nontrivial to find efficient algorithms. We note here that the problem can be nicely expressed with answer set programming (ASP, see for instance [Gebser *et al.*, 2012]). We will demonstrate this for the n-adic case. Assuming the specific problem is given by the following set of ground facts $F_{\mathbb{K},\mathbb{C}}$:

- $\texttt{set}_i(a)$ for all $a \in K_i$,
- $\texttt{rel}(a_1, \ldots, a_n)$ for all $(a_1, \ldots, a_n) \in R$,
- $\texttt{required}_i(a)$ for all $a \in K_i^+$, and
- $\texttt{forbidden}_i(a)$ for all $a \in K_i^-$.

Let P denote the following fixed answer set program (with rules for every $i \in \{1, \ldots, n\}$):

$$
\begin{aligned}
\texttt{in}_i(x) &\leftarrow \texttt{set}_i(x) \wedge \sim\texttt{out}_i(x)\\
\texttt{out}_i(x) &\leftarrow \texttt{set}_i(x) \wedge \sim\texttt{in}_i(x)\\
&\leftarrow \textstyle\bigwedge_{j\in\{1,\ldots,n\}} \texttt{in}_j(x_j) \wedge \sim\texttt{rel}(x_1, \ldots, x_n)\\
\texttt{exc}_i(x_i) &\leftarrow \textstyle\bigwedge_{j\in\{1,\ldots,n\}\setminus\{i\}} \texttt{in}_j(x_j) \wedge \sim\texttt{rel}(x_1, \ldots, x_n)\\
&\leftarrow \texttt{out}_i(x) \wedge \sim\texttt{exc}_i(x)\\
&\leftarrow \texttt{out}_i(x) \wedge \texttt{required}_i(x)\\
&\leftarrow \texttt{in}_i(x) \wedge \texttt{forbidden}_i(x)
\end{aligned}
$$

Intuitively, the first two lines "guess" an n-concept candidate by stipulating for each element of each K_i if they are in or out. The third rule eliminates a candidate if it violates the condition $A_1 \times \ldots \times A_n \subseteq R$, while the fourth and fifth rule ensure the maximality condition for n-concepts. Finally, the sixth and the seventh rule eliminate n-concepts violating the given membership constraint.

There is a one-to-one correspondence between the answer sets X of $F_{\mathbb{K},\mathbb{C}} \cup P$ and the n-concepts of $\mathbb{K}$ satisfying $\mathbb{C}$ obtained as $(\{a \mid \texttt{in}_1(a) \in X\}, \ldots, \{a \mid \texttt{in}_n(a) \in X\})$. Consequently, optimized off-the-shelf ASP tools can be used for checking satisfiability but also for enumerating all satisfying n-concepts.

Algorithm 1 interactive n-concept finding algorithm

```
function FINDNCONCEPTINTERACTIVE(𝕂)
 Input: n-context 𝕂 = (K_1, ..., K_n, R)
 Output: n-concept searched by user
 Data: membership constraint ℂ = (K_1^+, K_1^-, ..., K_n^+, K_n^-)

   ℂ := (∅, ..., ∅)
   ℂ := PROPAGATE(𝕂, ℂ)
   while K_i ≠ K_i^+ ∪ K_i^- for some i ∈ {1, ..., n} do
       have user pick some such i and a ∈ K_i \ (K_i^+ ∪ K_i^-)
       have user pick some decision ∈ {in, out}
       if decision = in then
           update ℂ by K_i^+ := K_i^+ ∪ {a}
       else
           update ℂ by K_i^- := K_i^- ∪ {a}
       end if
       ℂ := PROPAGATE(𝕂, ℂ)
   end while
   return (K_1^+, ..., K_n^+)
end function
```

Algorithm 2 propagation of user decisions

```
function PROPAGATE(𝕂, ℂ)
 Input: n-context 𝕂, membership constraint ℂ
 Output: updated membership constraint
 Data: membership constraint ℂ'

   for all i ∈ {1, ..., n} do
       for all a ∈ K_i \ (K_i^+ ∪ K_i^-) do
           obtain ℂ' from ℂ by adding a to K_i^+
           if NMCSAT(𝕂, ℂ') = NO then
               update ℂ by adding a to K_i^-
           end if
           obtain ℂ' from ℂ by adding a to K_i^-
           if NMCSAT(𝕂, ℂ') = NO then
               update ℂ by adding a to K_i^+
           end if
       end for
   end for
   return ℂ
end function
```

7 Navigation in Conceptual Spaces

In this section we briefly describe an interactive search scenario where membership constraints can be put to use to support a user in finding an n-concept with desired properties. This is particularly useful in cases where the number of n-concepts is very large. The method is formally specified in Algorithm 1 (which calls Algorithm 2, which in turn relies on an nMCSAT solving procedure NMCSAT). We next explain the intuition and the formal arguments behind this approach in more detail.

First, given an n-context $\mathbb{K} = (K_1, \ldots, K_n, R)$ and a corresponding membership constraint $\mathbb{C}$, let $[\mathbb{C}]_{\mathbb{K}}$ denote the set of all n-concepts of $\mathbb{K}$ that satisfy $\mathbb{C}$. Next, observe that for the "zero-constraint" $\mathbb{C}_\emptyset = (\emptyset, \ldots, \emptyset)$, the set $[\mathbb{C}_\emptyset]_{\mathbb{K}}$ contains all n-concepts of $\mathbb{K}$. Further, for two membership constraints $\mathbb{C}_1$ and $\mathbb{C}_2$ with $\mathbb{C}_1 \preceq \mathbb{C}_2$ (where we let $\preceq$ denote componentwise $\subseteq$ and read it as "more general than"), we observe $[\mathbb{C}_2]_{\mathbb{K}} \subseteq [\mathbb{C}_1]_{\mathbb{K}}$. Finally every n-concept $\mathcal{C} = (A_1, \ldots, A_n)$ of $\mathbb{K}$ gives rise to the characteristic membership constraint $\mathbb{C}_{\mathcal{C}} := (A_1, K_1 \setminus A_1, \ldots, A_n, K_n \setminus A_n)$ with $[\mathbb{C}_{\mathcal{C}}]_{\mathbb{K}} = \{\mathcal{C}\}$.

We now want to describe the identification of an n-concept by a user as an iterated approximation process starting from $\mathbb{C}_\emptyset$ and going along a chain of ever more specific (but satisfiable) membership constraints until $\mathbb{C}_{\mathcal{C}}$ is reached for some n-concept $\mathcal{C}$. Thereby, given a current constraint $\mathbb{C} = (K_1^+, K_1^-, \ldots, K_n^+, K_n^-)$, the next constraint is determined by the user by picking some $a \in K_i \setminus (K_i^+ \cup K_i^-)$ for some i and adding it either to K_i^+ or K_i^-. In words, for some element, whose participation in the looked-for n-concept is not yet determined, the user has to decide to include or exclude it. In order to avoid that the membership constraint turns unsatisfiable as a consequence of the user's refinement decision, we will perform constraint propagation on $\mathbb{C}$ before the interaction: for every $a \in K_i \setminus (K_i^+ \cup K_i^-)$ for some i where adding a to K_i^+ (respectively K_i^-) would result in an unsatisfiable constraint, we add it to K_i^- (respectively K_i^+). Note that not both can be the case at the same time, since otherwise $\mathbb{C}$ itself would be unsatisfiable.

Note that the interactive algorithm sketched here does not need to compute all (possibly exponentially many) n-concepts upfront, it merely relies on (polynomially many) subsequent nMCSAT checks. An appropriate user interface for this navigation method would consist in n labeled lists containing the elements of the K_i with elements from K_i^+ labeled with "in" (or an appropriate color), elements from K_i^- labeled with "out", and elements from $K_i \setminus (K_i^+ \cup K_i^-)$ labeled with "unknown". By clicking on one of the "unknown" elements, the user may switch it to "in" or "out". Subsequent constraint propagation as described above then will possibly turn other "unknown" labels into "in" or "out" as a ramification of the user's decision. When no more "unknown" labels are left, the target concept has been identified.

8 Conclusion

Motivated by requirements that arise naturally when applying conceptual analysis techniques to large knowledge sets, we have investigated a way of specifying selection criteria for "interesting" concepts. To this end, we defined membership constraints as collections of required, respectively forbidden objects and attributes. We have studied the computational complexity of the corresponding satisfiability problem (MCSAT) of determining if a formal concept exists that adheres to the given specification. We have proved that in its general form, the MCSAT problem is NP-complete even if we restrict only to "forbidden" objects and attributes. On the other hand, if we have no "forbidden" objects or no "forbidden" attributes, the complexity drops to AC_0.

When considering cases of arity three (objects, attributes, and conditions) or higher – with the notion of membership constraints appropriately adjusted – the corresponding problem is again NP-complete in general, but also here tractable special cases can be identified.

We presented a generic answer set programming (ASP) encoding for membership constraints, such that highly optimized ASP tools can be used to solve them. Finally we described an interactive search scenario in order to narrow down the search space for an n-concept with desired properties. This search paradigm relies on efficient methods for nMCSAT checking.

As an obvious and immediate avenue for future work, we will implement and evaluate our navigation framework based on the ASP-based satisfiability checker described.

Acknowledgements

The authors are grateful to the helpful comments from the anonymous reviewers of this paper.

References

[Ganter and Wille, 1999] Bernhard Ganter and Rudolf Wille. *Formal concept analysis - mathematical foundations*. Springer, 1999.

[Gebser *et al.*, 2012] Martin Gebser, Roland Kaminski, Benjamin Kaufmann, and Torsten Schaub. *Answer Set Solving in Practice*. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan and Claypool Publishers, 2012.

[Immerman, 1999] Neil Immerman. *Descriptive complexity*. Graduate texts in computer science. Springer, 1999.

[Jäschke *et al.*, 2008] Robert Jäschke, Andreas Hotho, Christoph Schmitz, Bernhard Ganter, and Gerd Stumme. Discovering shared conceptualizations in folksonomies. *Journal of Web Semantics*, 6(1):38–53, 2008.

[Lehmann and Wille, 1995] Fritz Lehmann and Rudolf Wille. A triadic approach to formal concept analysis. In Gerard Ellis, Robert Levinson, William Rich, and John F. Sowa, editors, *Proceedings of the Third International Conference on Conceptual Structures (ICCS 1995)*, volume 954 of *LNCS*, pages 32–43. Springer, 1995.

[Papadimitriou, 1994] Christos H. Papadimitriou. *Computational Complexity*. Theoretical computer science. Addison-Wesley, 1994.

[Rudolph *et al.*, 2015] Sebastian Rudolph, Christian Săcărea, and Diana Troancă. Towards a navigation paradigm for triadic concepts. In Jaume Baixeries, Christian Sacarea, and Manuel Ojeda-Aciego, editors, *Proceedings of the 13th International Conference on Formal Concept Analysis (ICFCA 2015)*, volume 9113 of *LNCS*, pages 232–248. Springer, 2015.

[Wille, 1994] Rudolf Wille. Plädoyer für eine philosophische Grundlegung der Begrifflichen Wissensverarbeitung. In *Begriffliche Wissensverarbeitung – Grundfragen und Aufgaben*, pages 11–25. B.I.-Wissenschaftsverlag, 1994.

[Wille, 1997] Rudolf Wille. Conceptual landscapes of knowledge: a pragmatic paradigm for knowledge processing. In Guy Mineau and Andrew Fall, editors, *Proceedings of the International Symposium on Knowledge Representation, Use, and Storage Efficiency*, pages 2–13. Simon Frasier University, Vancouver, 1997.

[Wille, 2000] Rudolf Wille. Begriffliche Wissensverarbeitung: Theorie und Praxis. *Informatik Spektrum*, (23):357–369, 2000.

[Wille, 2006] Rudolf Wille. Methods of conceptual knowledge processing. In Rokia Missaoui and Jürg Schmid, editors, *Proceedings of the 4th International Conference on Formal Concept Analysis (ICFCA 2006)*, volume 3874 of *LNCS*, pages 1–29. Springer, 2006.