Technische Universität Dresden
Institute for Theoretical Computer Science
Chair for Automata Theory

LTCS–Report

Decidability and Complexity of Threshold Description Logics Induced by Concept Similarity Measures

Franz Baader Oliver Fernández Gil

LTCS-Report 16-07

Postal Address:
Lehrstuhl für Automatentheorie
Institut für Theoretische Informatik
TU Dresden
01062 Dresden

http://lat.inf.tu-dresden.de

Visiting Address:
Nöthnitzer Str. 46
Dresden

Decidability and Complexity of Threshold Description Logics Induced by Concept Similarity Measures

Franz Baader[1] and Oliver Fernández Gil*[1]

[1]Theoretical Computer Science, TU Dresden, Germany

Abstract

In a recent research paper, we have proposed an extension of the light-weight Description Logic (DL) $\mathcal{EL}$ in which concepts can be defined in an approximate way. To this purpose, the notion of a graded membership function m, which instead of a Boolean membership value 0 or 1 yields a membership degree from the interval $[0, 1]$, was introduced. Threshold concepts can then, for example, require that an individual belongs to a concept C with degree at least 0.8. Reasoning in the threshold DL $\tau\mathcal{EL}(m)$ obtained this way of course depends on the employed graded membership function m. The paper defines a specific such function, called *deg*, and determines the exact complexity of reasoning in $\tau\mathcal{EL}(deg)$. In addition, it shows how concept similarity measures (CSMs) $\sim$ satisfying certain properties can be used to define graded membership functions $m_\sim$, but it does not investigate the complexity of reasoning in the induced threshold DLs $\tau\mathcal{EL}(m_\sim)$. In the present paper, we start filling this gap. In particular, we show that computability of $\sim$ implies decidability of $\tau\mathcal{EL}(m_\sim)$, and we introduce a class of CSMs for which reasoning in the induced threshold DLs has the same complexity as in $\tau\mathcal{EL}(deg)$.

*Supported by DFG Graduiertenkolleg 1763 (QuantLA).

Contents

1 Introduction

DLs are a well-investigated family of logic-based knowledge representation languages, which are frequently used to formalize ontologies for application domains such as biology and medicine. To define the important notions of such an application domain as formal concepts, DLs state necessary and sufficient conditions for an individual to belong to a concept. Once the relevant concepts of an application domain are formalized this way, they can be used in queries in order to retrieve new information from data. The DL $\mathcal{EL}$, in which concepts can be built using concept names as well as the concept constructors conjunction ($\sqcap$), existential restriction ($\exists r.C$), and the top concept ($\top$), has drawn considerable attention in the last decade since, on the one hand, important inference problems such as the subsumption problem are polynomial in $\mathcal{EL}$ [4, 1, 6]. On the other hand, though quite inexpressive, $\mathcal{EL}$ underlies the OWL 2 EL profile[1] and can be used to define biomedical ontologies, such as the large medical ontology SNOMED CT.[2]

Like all traditional DLs, $\mathcal{EL}$ is based on classical first-order logic, and thus its semantics is strict in the sense that all the stated properties need to be satisfied for an individual to belong to a concept. In applications where exact definitions are hard to come by, it would be useful to relax this strict requirement and allow for approximate definitions of concepts, where most, but not all, of the stated properties are required to hold. For example, in clinical diagnosis, diseases are often linked to a long list of medical signs and symptoms, but patients that have a certain disease rarely show all these signs and symptoms. Instead, one looks for the occurrence of sufficiently many of them. Similarly, people looking for a flat to rent or a bicycle to buy may have a long list of desired properties, but will also be satisfied if many, but not all, of them are met. In order to support defining concepts in such an approximate way, in [2] we have introduced a DL extending $\mathcal{EL}$ with threshold concept constructors of the form $C_{\bowtie t}$, where C is an $\mathcal{EL}$ concept, $\bowtie\, \in \{<, \leq, >, \geq\}$, and t is a *rational* number in $[0, 1]$. The semantics of these new concept constructors is defined using a graded membership function m that, given a (possibly complex) $\mathcal{EL}$ concept C and an individual d of an interpretation $\mathcal{I}$, returns a value from the interval [0,1], rather than a Boolean value from $\{0, 1\}$. The concept $C_{\bowtie t}$ then collects all the individuals that belong to C with degree $\bowtie t$, where this degree is computed using the function m. The DL $\tau\mathcal{EL}(m)$ is obtained from $\mathcal{EL}$ by adding these new constructors. There are, of course, different possibilities for how to define a graded membership function m, and the semantics of the obtained new logic $\tau\mathcal{EL}(m)$ depends on m.

In addition to introducing the family of DLs $\tau\mathcal{EL}(m)$, we have also defined a concrete graded membership function *deg*, which is obtained as a natural extension of the well-known homomorphism characterization of crisp membership and

[1] see http://www.w3.org/TR/owl2-profiles/

[2] see http://www.ihtsdo.org/snomed-ct/

subsumption in $\mathcal{EL}$ [4]. It is proved in [2] that concept satisfiability and ABox consistency are NP-complete in $\tau\mathcal{EL}(deg)$, whereas the subsumption and the instance checking problem are co-NP complete (the latter w.r.t. *data complexity*). In addition, it is shown how a CSM $\sim$ that is equivalence invariant, role-depth bounded and equivalence closed[3] (see [11]) can be used to define a graded membership function $m_\sim$. In particular, the graded membership function *deg* can be obtained in this way, i.e., there is a standard CSM $\sim^*$ such that $m_{\sim^*} = deg$. However, the complexity of reasoning in the DLs $\tau\mathcal{EL}(m_\sim)$ for $\sim \neq \sim^*$ has not been investigated in [2].

The goal of the present paper is to start filling this gap. Firstly, we will show that, for *computable* standard CSMs $\sim$, reasoning in $\tau\mathcal{EL}(m_\sim)$ can effectively be reduced to reasoning in the DL $\mathcal{ALC}$. Though the complexity of reasoning in $\mathcal{ALC}$ is known to be "only " PSpace [12], the complexity of the decision procedures for reasoning in $\tau\mathcal{EL}(m_\sim)$ obtained this way is non-elementary, due to the high complexity of the reduction function. Secondly, in order to obtain threshold DLs of lower complexity, we determine a class of standard CSMs definable using the *simi framework* of [11] such that reasoning in $\tau\mathcal{EL}(m_\sim)$ for a member $\sim$ of this class has the same complexity as reasoning in $\tau\mathcal{EL}(deg)$. Thirdly, we consider the problem of answering relaxed instance queries [7] using CSMs from this class. For the CSM $\sim^*$ corresponding to *deg*, it was shown in [2] that relaxed instance queries w.r.t. this CSM can be answered in polynomial time. We extend this result to all members of our class. This improves on the complexity upper bounds for answering relaxed instance queries in [7].

[3] In the following we will call a CSM satisfying these three properties a *standard* CSM.

2 The family of DLs $\tau\mathcal{EL}(m_\sim)$

First, we introduce the DL $\mathcal{EL}$ and show how, up to equivalence, all $\mathcal{EL}$ concept descriptions over a finite vocabulary and with a bounded role depth can be effectively computed. This will be used later to show the decidability result mentioned in the introduction. Second, we recall the definition of graded membership functions and the induced threshold DLs as well as some additional definitions and results from [2]. Third, we recall how concept similarity measures can be used to define graded membership functions.

2.1 The Description Logic $\mathcal{EL}$

Let $\mathsf{N_C}$ and $\mathsf{N_R}$ be finite sets of *concept* and *role* names, respectively. The set $\mathcal{C}_{\mathcal{EL}}(\mathsf{N_C}, \mathsf{N_R})$ of $\mathcal{EL}$ concept descriptions over $\mathsf{N_C}$ and $\mathsf{N_R}$ is inductively built from $\mathsf{N_C}$ using the concept constructors *conjunction* $(C \sqcap D)$, *existential restriction* $(\exists r.C)$, and *top* $(\top)$. The semantics of $\mathcal{EL}$ concept descriptions is defined using standard first-order logic interpretations. An *interpretation* $\mathcal{I} = (\Delta^\mathcal{I}, \cdot^\mathcal{I})$ consists of a non-empty domain $\Delta^\mathcal{I}$ and an interpretation function $\cdot^\mathcal{I}$ that interprets concept names in $\mathsf{N_C}$ as subsets of $\Delta^\mathcal{I}$ and assigns binary relations over $\Delta^\mathcal{I}$ to role names in $\mathsf{N_R}$. This function is inductively extended to complex concept descriptions as follows.

$$\begin{aligned}
\top^\mathcal{I} &:= \Delta^\mathcal{I}, \\
(C \sqcap D)^\mathcal{I} &:= C^\mathcal{I} \cap D^\mathcal{I}, \text{ and} \\
(\exists r.C)^\mathcal{I} &:= \{x \in \Delta^\mathcal{I} \mid \exists y.((x, y) \in r^\mathcal{I} \wedge y \in C^\mathcal{I})\}.
\end{aligned}$$

Given two $\mathcal{EL}$ concept descriptions C and D, we say that C is *subsumed* by D (in symbols $C \sqsubseteq D$) iff $C^\mathcal{I} \subseteq D^\mathcal{I}$ for all interpretations $\mathcal{I}$. These two concepts are equivalent (in symbols $C \equiv D$) iff $C \sqsubseteq D$ and $D \sqsubseteq C$. In addition, C is *satisfiable* iff $C^\mathcal{I} \neq \emptyset$ for some interpretation $\mathcal{I}$.[4]

Information about specific individuals (represented by a set of individual names $\mathsf{N_I}$) can be stated in an ABox, which is a finite set of *assertions* of the form $C(a)$ or $r(a, b)$, where $C \in \mathcal{C}_{\mathcal{EL}}(\mathsf{N_C}, \mathsf{N_R})$, $r \in \mathsf{N_R}$, and $a, b \in \mathsf{N_I}$. An interpretation $\mathcal{I}$ is then extended to assign domain elements $a^\mathcal{I}$ to individual names a. We say that $\mathcal{I}$ satisfies an assertion $C(a)$ iff $a^\mathcal{I} \in C^\mathcal{I}$, and $r(a, b)$ iff $(a^\mathcal{I}, b^\mathcal{I}) \in r^\mathcal{I}$. Furthermore, $\mathcal{I}$ is a model of the ABox $\mathcal{A}$ (denoted as $\mathcal{I} \models \mathcal{A}$) iff it satisfies all the assertions of $\mathcal{A}$. The ABox $\mathcal{A}$ is *consistent* iff $\mathcal{I} \models \mathcal{A}$ for some interpretation $\mathcal{I}$. Finally, an individual a is an *instance* of C in $\mathcal{A}$ iff $a^\mathcal{I} \in C^\mathcal{I}$ for all models $\mathcal{I}$ of $\mathcal{A}$.

As shown in [9], $\mathcal{EL}$ concept descriptions C can be transformed into an equivalent *reduced form* C^r by applying the rewrite rule $C \sqcap D \longrightarrow C$ if $C \sqsubseteq D$ modulo

[4] In $\mathcal{EL}$, all concept descriptions are satisfiable, but this is no longer the case for its extensions by threshold concepts introduced below.

associativity and commutativity of $\sqcap$ as long as possible, not only on the top-level conjunction of the description, but also under the scope of existential restrictions. Up to associativity and commutativity of $\sqcap$, equivalent $\mathcal{EL}$ concept descriptions have the same reduced form.

The size $\mathsf{s}(C)$ of an $\mathcal{EL}$ concept description C is the number of occurrences of symbols needed to write C. The *role depth* $\mathsf{rd}(C)$ of C is the maximal nesting of existential restrictions in C. More formally,

$$\begin{aligned}\mathsf{rd}(\top) = \mathsf{rd}(A) &:= 0,\\ \mathsf{rd}(C_1 \sqcap C_2) &:= \max(\mathsf{rd}(C_1), \mathsf{rd}(C_2)),\\ \mathsf{rd}(\exists r.C) &:= \mathsf{rd}(C) + 1.\end{aligned}$$

As shown in [5], for finite sets $\mathsf{N_C}$ and $\mathsf{N_R}$ and a fixed bound k on the role depth, $\mathcal{C}_{\mathcal{EL}}(\mathsf{N_C}, \mathsf{N_R})$ contains only finitely many equivalence classes of concept descriptions of role depth $\leq k$. The following lemma shows that finitely many representatives of these equivalence classes can be computed.

Lemma 1. *For all $k \geq 0$ there exists a* finite *set $\mathcal{R}^k \subseteq \mathcal{C}_{\mathcal{EL}}(\mathsf{N_C}, \mathsf{N_R})$ consisting of $\mathcal{EL}$ concept descriptions in reduced form and of role depth $\leq k$ such that $C^r \in \mathcal{R}^k$ holds for all $C \in \mathcal{C}_{\mathcal{EL}}(\mathsf{N_C}, \mathsf{N_R})$ with $\mathsf{rd}(C) \leq k$, and this set can effectively be computed.*

Proof. The lemma can be shown by induction on k. Concept descriptions of role depth $k = 0$ are conjunctions of concept names, where the empty conjunction corresponds to $\top$. The requirement to be reduced corresponds to the fact that each concept name occurs at most once in the conjunction. Thus,

$$\mathcal{R}^0 = \{ \bigsqcap_{A \in S} A \mid S \subseteq \mathsf{N_C} \},$$

which is obviously finite and, given $\mathsf{N_C}$, can easily be computed.

Up to equivalence, concept descriptions of role depth $\leq k$ for $k > 0$ are of the form

$$A_1 \sqcap \ldots \sqcap A_n \sqcap \exists s_1.D_1 \sqcap \ldots \sqcap \exists s_q.D_q$$

where $n \geq 0$, $q \geq 1$, $\{A_1, \ldots, A_n\} \subseteq \mathsf{N_C}$, and $D_i \in \mathcal{R}^{k-1}$ for all $1 \leq i \leq q$. The requirement to be reduced imposes the constraint that two different conjuncts $\exists r.D_i$ and $\exists r.D_j$ occurring in this conjunction satisfy that:

- D_i and D_j are concepts in reduced form, and
- $D_i \not\sqsubseteq D_j$ (and thus also $D_i \not\equiv D_j$).

Thus, for every role $r \in \mathsf{N_R}$ there are at most $|\mathcal{R}^{k-1}|$ conjuncts that are existential restrictions for r. Since by induction we know that $\mathcal{R}^{k-1}$ is finite, this implies that $\mathcal{R}^k$ is finite as well. In addition, starting from $\mathcal{R}^{k-1}$ the set $\mathcal{R}^k$ can be computed as follows:

```
$\mathcal{R}^k := \mathcal{R}^{k-1}$.
$\{r_1, \ldots, r_{|\mathsf{N_R}|}\}$ is a linear order of $\mathsf{N_R}$.
for all $(S_\epsilon, S_1, \ldots, S_{|\mathsf{N_R}|}) \in 2^{\mathsf{N_C}} \times \underbrace{2^{\mathcal{R}^{k-1}} \times \ldots \times 2^{\mathcal{R}^{k-1}}}_{|\mathsf{N_R}|}$ do
    if $(\forall S_i. [(C, D \in S_i \wedge C \neq D) \Rightarrow C \not\sqsubseteq D])$ and
       $(\mathsf{rd}(C) = k-1$ for at least one $C$ in $S_i)$ then
        construct the $\mathcal{EL}$ concept description $X$ as:
            $$X := \bigsqcap_{A \in S_\epsilon} A \sqcap \bigsqcap_{i=1}^{|\mathsf{N_R}|} \bigsqcap_{Y \in S_i} \exists r_i.Y$$
        $\mathcal{R}^k := \mathcal{R}^k \cup \{X\}$
    end if
end for
```

Since subsumption in $\mathcal{EL}$ is decidable and by induction $\mathcal{R}^{k-1}$ is computable, this procedure provides an effective way to compute $\mathcal{R}^k$. □

2.2 Extending $\mathcal{EL}$ with threshold concepts

In [2], $\mathcal{EL}$ is extended with *threshold concepts* $C_{\bowtie t}$, where C is an $\mathcal{EL}$ concept description, $\bowtie \in \{<, \leq, >, \geq\}$, and t is a rational number in $[0,1]$. These threshold concepts can then be used like concept names when building complex concept descriptions such as $(\exists r.A)_{<1} \sqcap \exists r.(A \sqcap B)_{\geq .8} \sqcap B$. Note that the concept C occurring within the threshold operator must be an $\mathcal{EL}$ concept description, and thus nesting of these operators is not allowed. The semantics of the threshold operators is defined using a *graded membership function*, which is defined as follows.[5]

Definition 2. A *graded membership function* m is a family of functions that contains for every interpretation $\mathcal{I}$ a function $m^{\mathcal{I}} : \Delta^{\mathcal{I}} \times \mathcal{C}_{\mathcal{EL}}(\mathsf{N_C}, \mathsf{N_R}) \to [0,1]$ satisfying the following conditions (for $C, D \in \mathcal{C}_{\mathcal{EL}}(\mathsf{N_C}, \mathsf{N_R})$):

$$\begin{aligned} &\textit{M1: } \forall \mathcal{I}\ \forall d \in \Delta^{\mathcal{I}} : d \in C^{\mathcal{I}} \Leftrightarrow m^{\mathcal{I}}(d, C) = 1, \\ &\textit{M2: } C \equiv D \Leftrightarrow \forall \mathcal{I}\ \forall d \in \Delta^{\mathcal{I}} : m^{\mathcal{I}}(d, C) = m^{\mathcal{I}}(d, D). \end{aligned}$$

Intuitively, given an interpretation $\mathcal{I}$ and $d \in \Delta^{\mathcal{I}}$, $m^{\mathcal{I}}(d, C) \in [0,1]$ represents the degree to which d belongs to C in $\mathcal{I}$. The concept $C_{\bowtie t}$ then collects all the

[5] Note that this definition corrects a typo in Def. 3 of [2].

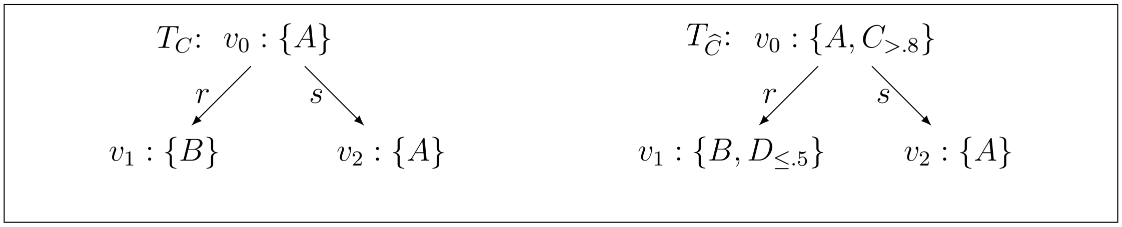


Figure 1: $\mathcal{EL}$ and $\tau\mathcal{EL}(m)$ description trees

elements of $\Delta^\mathcal{I}$ that belong to C with degree $\bowtie t$, as measured by m. To be more precise, the formal semantics of threshold concepts is then defined as follows:

$$(C_{\bowtie t})^\mathcal{I} := \{d \in \Delta^\mathcal{I} \mid m^\mathcal{I}(d, C) \bowtie t\}.$$

This way, a new family of DLs called $\tau\mathcal{EL}(m)$ is obtained, where m is a parameter indicating which function is used to obtain the semantics of threshold concepts.

In addition to this family of DLs, [2] introduces a concrete membership function *deg*, and investigates the computational properties of its corresponding DL $\tau\mathcal{EL}(\mathit{deg})$. We show that *satisfiability* and *consistency* are NP-complete, whereas *subsumption* and *instance checking* (w.r.t. data complexity) are coNP-complete in $\tau\mathcal{EL}(\mathit{deg})$ (Th. 5 and 6 in [2]). An important step towards obtaining these results was to characterize when an individual is an instance of a $\tau\mathcal{EL}(\mathit{deg})$ concept description in an interpretation. This characterization generalizes the corresponding one for *crisp* membership in $\mathcal{EL}$, which is based on the representation of concepts and interpretations as graphs, and the existence of homomorphisms between these graphs. Since it is needed in Section 3.3, we briefly describe the general ideas behind it. In fact, it turns out that this characterization works for $\tau\mathcal{EL}(m)$ regardless of which graded membership function m is used.

$\mathcal{EL}$ *description graphs* are graphs where the nodes are labeled with sets of concept names and the edges are labeled with role names. As shown in [1, 4], interpretations can be represented as (arbitrary) $\mathcal{EL}$ description graphs and $\mathcal{EL}$ concept descriptions as $\mathcal{EL}$ *description trees*, i.e., as description graphs that are trees (whose root we will always denote as v_0). Description trees can be extended to $\tau\mathcal{EL}(m)$ by allowing the node labels also to contain elements of the form $C_{\bowtie t}$. For instance, the left-hand side of Figure 1 depicts the $\mathcal{EL}$ description tree corresponding to the $\mathcal{EL}$ concept description $A \sqcap \exists r.B \sqcap \exists s.A$, whereas the right-hand side shows the $\tau\mathcal{EL}(m)$ description tree corresponding to the $\tau\mathcal{EL}(m)$ concept description $A \sqcap C_{>.8} \sqcap \exists r.(B \sqcap D_{\leq .5}) \sqcap \exists s.A$.

Based on the definition of homomorphisms between $\mathcal{EL}$ description trees in [4], the notion of a τ-homomorphism ϕ from a $\tau\mathcal{EL}(m)$ description tree $\widehat{H}$ into an $\mathcal{EL}$ description graph $G_\mathcal{I}$ representing an interpretation $\mathcal{I}$ is defined in [2] to be a mapping from the nodes of $\widehat{H}$ to the nodes of $G_\mathcal{I}$ such that

1. the concept names occurring in the label set of a node v of $\widehat{H}$ are contained in the label set of its image $\phi(v)$;
2. if (v, w) is an edge with label r in $\widehat{H}$, then there is an edge $(\phi(v), \phi(w))$ with label r in $G_{\mathcal{I}}$;
3. if the label set of a node v of $\widehat{H}$ contains $C_{\bowtie t}$, then $m^{\mathcal{I}}(\phi(v), C) \bowtie t$.

Conditions 1 and 2 correspond to the classical definition of homomorphisms between $\mathcal{EL}$ description graphs. From the results presented in [4], these classical homomorphisms can be used to characterize classical membership in $\mathcal{EL}$ concept descriptions.

Theorem 3. *Let $\mathcal{I}$ be an interpretation, $d \in \Delta^{\mathcal{I}}$, and C an $\mathcal{EL}$ concept description. Then, $d \in C^{\mathcal{I}}$ iff there exists a homomorphism φ from T_C to $G_{\mathcal{I}}$ such that $\varphi(v_0) = d$.*

Similarly, using τ-homomorphisms, membership in $\tau\mathcal{EL}(m)$ concept descriptions can be characterized as follows (the proof is very tedious, the details can be found in the Appendix).

Theorem 4. *Let $\mathcal{I}$ be an interpretation with associated $\mathcal{EL}$ description graph $G_{\mathcal{I}}$, $d \in \Delta^{\mathcal{I}}$, and $\widehat{C}$ a $\tau\mathcal{EL}(m)$ concept description with associated $\tau\mathcal{EL}(m)$ description tree $T_{\widehat{C}}$. Then, $d \in \widehat{C}^{\mathcal{I}}$ iff there exists a τ-homomorphism ϕ from $T_{\widehat{C}}$ to $G_{\mathcal{I}}$ such that $\phi(v_0) = d$.*

If the interpretation $\mathcal{I}$ is finite and m is computable in polynomial time, then the existence of a τ-homomorphism can be checked in polynomial time. For the case $m = deg$ this fact as well as Theorem 4 were already shown in [2].

2.3 CSMs and graded membership functions

A concept similarity measure (CSM) is a function that maps pairs of concept descriptions to values in $[0, 1]$. Intuitively, the larger this value is the more similar the concept descriptions are. More formally, a CSM for $\mathcal{EL}$ concept descriptions over $\mathsf{N_C}$ and $\mathsf{N_R}$ is a mapping $\sim : \mathcal{C}_{\mathcal{EL}}(\mathsf{N_C}, \mathsf{N_R}) \times \mathcal{C}_{\mathcal{EL}}(\mathsf{N_C}, \mathsf{N_R}) \to [0, 1]$. Examples of such measures as well as properties these measures should satisfy can, e.g., be found in [13, 7, 11].

We reproduce here the Definition 10 in [2], which shows how a CSM $\sim$ can be used to define an associated graded membership function $m_\sim$.

Definition 5. Let $\sim$ be a CSM. Then, for all interpretations $\mathcal{I}$ the function $m_\sim^{\mathcal{I}} : \Delta^{\mathcal{I}} \times \mathcal{C}_{\mathcal{EL}}(\mathsf{N_C}, \mathsf{N_R}) \to [0, 1]$ is defined as:

$$m_\sim^{\mathcal{I}}(d, C) := \max\{C \sim D \mid D \in \mathcal{C}_{\mathcal{EL}}(\mathsf{N_C}, \mathsf{N_R}) \text{ and } d \in D^{\mathcal{I}}\}.$$

To ensure that this definition yields a well-defined graded membership function, $\sim$ is required to be a *standard CSM*, which means that it needs to satisfy the following three properties:

- $\sim$ must be *equivalence invariant*, i.e., $C \equiv C'$ and $D \equiv D'$ implies $C \sim D = C' \sim D'$;
- $\sim$ must be *role-depth bounded*, i.e., $C \sim D = C_k \sim D_k$ where $k > \min\{\mathsf{rd}(C), \mathsf{rd}(D)\}$ and C_k, D_k are the *restrictions of* C, D *to role depth* k, which are obtained from C, D by removing all existential restrictions occurring at role depth k. More formally,
$$\begin{aligned} C_k &:= C && \text{if } C \in \mathsf{N_C} \text{ or } C = \top,\\ C_k &:= [C_1]_k \sqcap \ldots \sqcap [C_n]_k && \text{if } C = C_1 \sqcap \ldots \sqcap C_n,\\ [\exists r.C]_k &:= \begin{cases} \top & \text{if } k = 0,\\ \exists r.[C]_{k-1} & \text{otherwise;} \end{cases} \end{aligned}$$
- $\sim$ must be *equivalence closed*, i.e., the equivalence $C \equiv D$ iff $C \sim D = 1$ holds.

The first two conditions ensure that $m^{\mathcal{I}}_{\sim}(d, C)$ is well-defined, i.e., the maximum in the definition of this value really exists. In fact, these conditions imply that one can restrict the search for an appropriate $D \in \mathcal{C}_{\mathcal{EL}}(\mathsf{N_C}, \mathsf{N_R})$ to finitely many concept descriptions.

Lemma 6. *Let* $C \in \mathcal{C}_{\mathcal{EL}}(\mathsf{N_C}, \mathsf{N_R})$ *with* $\mathsf{rd}(C) = k$. *Then,*

$$m^{\mathcal{I}}_{\sim}(d, C) = \max\{C \sim D \mid D \in \mathcal{R}^{k+1} \textit{ and } d \in D^{\mathcal{I}}\}$$

Proof. Since $\sim$ is role-depth bounded, this means that to compute $m^{\mathcal{I}}_{\sim}$ as expressed in Definition 5, one can restrict the attention to concepts D of role depth at most $k + 1$. Moreover, for all such concept descriptions D we know that $D^r \in \mathcal{R}^{k+1}$ (see Lemma 1). Thus, being $\sim$ equivalence invariant, allows us to assume without loss of generality that $D \in \mathcal{R}^{k+1}$. □

Equivalence closedness is additionally needed to ensure that m satisfies the properties required in Definition 2.

3 Reasoning in $\tau\mathcal{EL}(m_\sim)$

We will now present a preliminary study of the complexity of reasoning in DLs $\tau\mathcal{EL}(m_\sim)$ for standard CSMs $\sim$. Obviously, there is a great variety of standard CSMs and not all of them are well-behaved from a computational point of view. In fact, we will start by showing that there are standard CSMs that are not computable. While non-computability of $\sim$ does not automatically imply that reasoning problems in $\tau\mathcal{EL}(m_\sim)$ are undecidable, we will see that there are non-computable CSMs $\sim$ such that the standard reasoning problems satisfiability, subsumption, consistency, and instance checking are undecidable in $\tau\mathcal{EL}(m_\sim)$. Afterwards, we will show that computability of $\sim$ implies decidability of these reasoning problems in $\tau\mathcal{EL}(m_\sim)$. Finally, we determine a class of standard CSMs such that reasoning in $\tau\mathcal{EL}(m_\sim)$ for a member $\sim$ of this class has the same complexity as reasoning in $\tau\mathcal{EL}(deg)$.

3.1 Undecidability

Despite the properties required for a CSM to be standard, the set of all such measures still exhibits a great diversity. In fact, it contains *infinitely* many CSMs that have very simple definitions but are, nevertheless, *non-computable* functions. We now define a particular set of standard CSMs, and we will see that it is not difficult to put it into a one-to-one correspondence with the *power set* of the natural numbers. This is the case even if the CSMs in such a set are defined w.r.t. $\mathcal{C}_{\mathcal{EL}}(\{A\}, \{r\})$.

Definition 7. Let $N \subseteq \mathbb{N}$ and $0 < a < 1$ a fixed rational number. Then, we define the concept similarity measure $\sim_N$ as follows:

$$C \sim_N D := \begin{cases} 1 & \text{if } C \equiv D \\ \mu(C, D) & \text{otherwise.} \end{cases}$$

where μ corresponds to the expression:

$$\mu(C, D) := \begin{cases} a & \text{if } \mathsf{rd}(C) = \mathsf{rd}(D) \text{ and } \mathsf{rd}(C) \in N \\ 0 & \text{otherwise.} \end{cases}$$

We now show that $\sim_N$ is a standard CSM.

Lemma 8. *Let $N \subseteq \mathbb{N}$ and $\sim_N$ defined as in Definition 7. Then, $\sim_N$ is a standard CSM.*

Proof. That $\sim_N$ is equivalence closed follows directly from its definition. Let us look at the other two properties.

1. *equivalence invariance*: let $C, C', D, D' \in \mathcal{C}_{\mathcal{EL}}(\{A\}, \{r\})$ such that $C \equiv C'$ and $D \equiv D'$. According to the definition of $\sim_N$ there are three possible values for $C \sim_N D$:

   - $C \sim_N D = 1$. This means that $C \equiv C' \equiv D \equiv D'$, and by definition $C \sim_N D = C' \sim_N D' = 1$.
   - $C \sim_N D = 0$. There are two possibilities:
     - $\mathsf{rd}(C) \neq \mathsf{rd}(D)$. Since $C \equiv C'$ and $D \equiv D'$, this means that $\mathsf{rd}(C') \neq \mathsf{rd}(D')$. Hence, $C' \sim_N D' = 0$.
     - $\mathsf{rd}(C) \notin N$. Then, $\mathsf{rd}(C') \notin N$ as well, and thus $C' \sim_N D' = 0$.
   - $C \sim_N D = a$. Then, $C \not\equiv D$, $\mathsf{rd}(C) = \mathsf{rd}(D)$ and $\mathsf{rd}(C) \in N$. Similarly as in the previous case, we obtain $C' \not\equiv D'$, $\mathsf{rd}(C') = \mathsf{rd}(D')$ and $\mathsf{rd}(C') \in N$. Thus, $C' \sim_N D' = a$.

2. *role-depth boundedness*: let $C, D \in \mathcal{C}_{\mathcal{EL}}(\{A\}, \{r\})$. Whenever $\mathsf{rd}(C) = \mathsf{rd}(D)$ the role-depth boundedness condition trivially holds for C and D, since for any $k > \mathsf{rd}(C)$ it is the case that $C = C_k$ and $D = D_k$. It remains to look at the case where $\mathsf{rd}(C) \neq \mathsf{rd}(D)$. It follows from the definition of $\sim_N$ that $C \sim_N D = 0$. Now, without loss of generality, let $\mathsf{rd}(C) < \mathsf{rd}(D)$. For any value $k > \mathsf{rd}(C)$ we have $\mathsf{rd}(C_k) < \mathsf{rd}(D_k)$. Then, $\mathsf{rd}(C_k) \neq \mathsf{rd}(D_k)$, and consequently $C_k \sim_N D_k = 0 = C \sim_N D$.

□

Hence, each subset N of the natural numbers induces a standard CSM $\sim_N$. More importantly, for all pairs of distinct subsets $N_1, N_2 \in \mathbb{N}$, the induced CSMs $\sim_{N_1}$ and $\sim_{N_2}$ are different. Just take a number n such that $n \in N_1$ and $n \notin N_2$ (or vice versa). Then, take two concepts C and D such that $\mathsf{rd}(C) = \mathsf{rd}(D) = n$ and $C \not\equiv D$ (the fixed signature $\{A\} \cup \{r\}$ ensures that this is always possible). By definition we will obtain $C \sim_{N_1} D = a$ and $C \sim_{N_2} D = 0$.

Thus, there are as many CSMs of this type as subsets of the natural numbers, namely, *uncountably* many. Since there are only countable many Turing Machines, there must be non-computable standard CSMs.

Proposition 9. *The set of standard CSMs on $\mathcal{EL}$ concept descriptions defined over $\mathcal{C}_{\mathcal{EL}}(\{A\}, \{r\})$, contains non-computable functions.*

As explained in Section 2.3, since $\sim_N$ is a standard CSM, the function $m_{\sim_N}$ is a well-defined graded membership function for all $N \subseteq \mathbb{N}$, and it induces the DL $\tau\mathcal{EL}(m_{\sim_N})$. Furthermore, the very simple definition of $\sim_N$ makes possible to use an algorithm deciding concept satisfiability in $\tau\mathcal{EL}(m_{\sim_N})$ as a component of an algorithm computing $\sim_N$. More precisely, given two $\mathcal{EL}$ concept descriptions C and D:

1. $C \equiv D \;\Rightarrow\; C \sim_N D = 1$.

2. $C \not\equiv D$ and $\mathsf{rd}(C) \neq \mathsf{rd}(D) \;\Rightarrow\; C \sim_N D = 0$.

3. Otherwise, the computation of $C \sim_N D$ solely depends on whether $\mathsf{rd}(C) \in N$. This could be alternatively solved by asking for satisfiability of the concept $C'_{\leq a} \sqcap C'_{\geq a}$ in $\tau\mathcal{EL}(m_{\sim_N})$, where C' is the following $\mathcal{EL}$ concept description:

   $$\exists \underbrace{r \ldots r}_{\mathsf{rd}(C)}.A$$

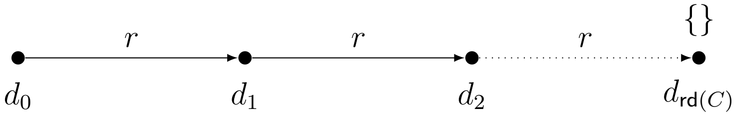
   A positive answer corresponds to $C \sim_N D = a$, while the opposite one yields $C \sim_N D = 0$. Let us see why this is true.

   - Satisfiability of $C'_{\leq a} \sqcap C'_{\geq a}$ implies that for some interpretation $\mathcal{I}$ and $d \in \Delta^{\mathcal{I}}$:

     $$m^{\mathcal{I}}_{\sim_N}(d, C') = a$$

     This means that for some concept F, $C' \sim_N F = a$ which by definition of $\sim_N$ implies $\mathsf{rd}(C') \in N$.

   - Conversely, let $C'_{\leq a} \sqcap C'_{\geq a}$ be unsatisfiable. Consider the interpretation $\mathcal{I}$ having the following description graph:

     One can observe that:

     $$d_0 \notin (C')^{\mathcal{I}} \qquad \text{and} \qquad d_0 \in (C^*)^{\mathcal{I}}, \text{ where } C^* := \exists \underbrace{r \ldots r}_{\mathsf{rd}(C)}.\top$$

     This means that $m^{\mathcal{I}}_{\sim_N}(d_0, C') < 1$. Since we are in the unsatisfiability case, it must be that $m^{\mathcal{I}}_{\sim_N}(d_0, C') = 0$. Moreover, since $d_0 \in (C^*)^{\mathcal{I}}$, such a concept is considered to compute $m^{\mathcal{I}}_{\sim_N}(d_0, C')$. Consequently, $C' \sim_N C^* = 0$. Thus, since $C' \not\equiv C^*$ and $\mathsf{rd}(C') = \mathsf{rd}(C^*)$, by definition of $\sim_N$ it follows that $\mathsf{rd}(C') \notin N$.

The first two steps of the previous algorithm consist of solving "fairly" easy tasks. Consequently, it becomes clear that decidability of the satisfiability problem in a DL $\tau\mathcal{EL}(m_{\sim_N})$ implies computability of the CSM $\sim_N$. Hence, the following undecidability result follows.

Proposition 10. *Let $N \subseteq \mathbb{N}$ and $\sim_N$ its corresponding concept similarity measure defined as in Definition 7. If $\sim_N$ is non-computable, then it induces an undecidable threshold DL $\tau\mathcal{EL}(m_{\sim_N})$.*

Summing up, on the one hand, we have shown that there are non-computable standard CSMs. This has been established by setting a one-to-one correspondence with the power set of the natural numbers. On the other hand, a subset of all non-computable standard CSMs $\sim$ induces a set of undecidable DLs $\tau\mathcal{EL}(m_\sim)$, where $m_\sim$ is constructed as described in Definition 5. Nevertheless, it is not yet clear to us whether non-computability of a standard CSM $\sim$ always implies undecidability of the induced DL $\tau\mathcal{EL}(m_\sim)$.

3.2 Decidability

Let $\sim$ be a computable standard CSM. We show decidability of reasoning in $\tau\mathcal{EL}(m_\sim)$ using an equivalence preserving and computable translation of $\tau\mathcal{EL}(m_\sim)$ concept descriptions into $\mathcal{ALC}$ concept descriptions. Since the standard reasoning problems are decidable in $\mathcal{ALC}$ such an effective translation obviously yields their decidability in $\tau\mathcal{EL}(m_\sim)$.

Recall that $\mathcal{ALC}$ [12] is obtained from $\mathcal{EL}$ by adding negation $\neg C$, whose semantics is defined in the usual way, i.e.,

$$(\neg C)^\mathcal{I} := \Delta^\mathcal{I} \setminus C^\mathcal{I}$$

Clearly, negation together with conjunction also yields disjunction $C \sqcup D$. Since $\mathcal{EL}$ is a fragment of $\mathcal{ALC}$, it suffices to show how to translate threshold concepts $C_{\bowtie t}$ into $\mathcal{ALC}$ concept descriptions. In addition, we can concentrate on the case where $\bowtie \in \{\geq, >\}$ since $C_{<t} \equiv \neg C_{\geq t}$ and $C_{\leq t} \equiv \neg C_{>t}$.

Lemma 11. *Let* $\bowtie \in \{\geq, >\}$*,* $t \in [0,1] \cap \mathbb{Q}$*, and* $C \in \mathcal{C}_\mathcal{EL}(\mathsf{N_C}, \mathsf{N_R})$ *with* $rd(C) = k$*. Then*

$$C_{\bowtie t} \equiv \bigsqcup \{D \mid D \in \mathcal{R}^{k+1} \text{ and } C \sim D \bowtie t\}.$$

Proof. Let $\mathcal{I}$ be an interpretation and $d \in \Delta^\mathcal{I}$. By the semantics of threshold concepts and Lemma 6, we know that $d \in (C_{\bowtie t})^\mathcal{I}$ iff

$$m_\sim^\mathcal{I}(d, C) = \max\{C \sim D \mid D \in \mathcal{R}^{k+1} \text{ and } d \in D^\mathcal{I}\} \bowtie t.$$

Since $\bowtie \in \{\geq, >\}$, this is equivalent to saying that there is a $D \in \mathcal{R}^{k+1}$ such that $C \sim D \bowtie t$ and $d \in D^\mathcal{I}$. This is in turn equivalent to $d \in \bigcup \{D^\mathcal{I} \mid D \in \mathcal{R}^{k+1} \text{ and } C \sim D \bowtie t\}$. □

Since $\mathcal{R}^{k+1}$ is finite, the disjunction on the right-hand side of the equivalence in the formulation of the lemma is finite, and thus this right-hand side is an admissible $\mathcal{ALC}$ concept description. This description can effectively be computed since $\mathcal{R}^{k+1}$ is computable by Lemma 1 and $\sim$ is computable by assumption.

Theorem 12. *If $\sim$ is a computable standard CSM, then satisfiability, subsumption, consistency and instance checking are decidable in $\tau\mathcal{EL}(m_\sim)$.*

Since the cardinality of $\mathcal{R}^k$ increases by one exponent with each increase of k, this approach provides only a *non-elementary* bound on the complexity of reasoning in $\tau\mathcal{EL}(m_\sim)$. We will now show that, for a restricted class of CSMs, one can obtain better complexity upper bounds.

3.3 Complexity

As shown in [2], there is a decidable standard CSM $\sim^*$ such that $deg = m_{\sim^*}$, and the complexity of reasoning in $\tau\mathcal{EL}(deg)$ is NP/coNP-complete for the standard reasoning problems. We will now identify a class of standard CSMs $\sim$ such that the complexity of reasoning in the induced threshold DLs $\tau\mathcal{EL}(m_\sim)$ is the same as in $\tau\mathcal{EL}(deg)$.

The CSM $\sim^*$ inducing *deg* is an instance of the *simi framework* introduced in [11]. This framework can be used to define a variety of similarity measures between $\mathcal{EL}$ concepts satisfying certain desirable properties. Here, we introduce a fragment of *simi* that is sufficient for our purposes.

To construct a CSM $\sim$ using *simi*, one first defines a directional measure $\sim_d$, and then uses a fuzzy connector $\otimes$ to combine the values obtained by comparing the reduced concepts in both directions with $\sim_d$:

$$C \sim D := (C^r \sim_d D^r) \otimes (D^r \sim_d C^r), \tag{1}$$

where the fuzzy connector $\otimes$ is a *commutative* binary operator $\otimes : [0,1] \times [0,1] \to [0,1]$ satisfying certain additional properties (see [11]). The definition of $C \sim_d D$ (see Def. 3 in [11]) depends on several parameters:

- A function g that assigns to every $\mathcal{EL}$ atom (i.e., concept name or existential restriction) a weight in $\mathbb{R}_{>0}$. This could be helpful, for instance, if one wants to express that some atom contributes more (is more important) to the similarity than others.

- A discounting factor $w \in [0,1)$[6]. The purpose of using this value is the following. Given two concept descriptions $\exists r.C$ and $\exists s.D$, if $C \sim_d D = 0$, having $w > 0$ allows to distinguish between the cases $r = s$ and $r \neq s$.

- A primitive measure between concept names and between role names: $pm : (\mathsf{N_C} \times \mathsf{N_C}) \cup (\mathsf{N_R} \times \mathsf{N_R}) \to [0,1]$, satisfying the following basic properties (different from [11] we do not deal with role inclusion axioms):

[6] The definition of w in [11] excludes the value 0. Nevertheless, all the properties shown in [11] to be satisfied by $\sim_d$ that are relevant to obtain our results also hold for $w = 0$.

- $pm(A, B) = 1$ iff $A = B$ for all $A, B \in \mathsf{N_C}$,
- $pm(r, s) = 1$ iff $r = s$ for all $r, s \in \mathsf{N_R}$.

In particular, the *default* primitive measure pm_d is defined as:

$$pm_d(A, B) := \begin{cases} 1 & \text{if } A = B \\ 0 & \textit{otherwise.} \end{cases}$$

and

$$pm_d(r, s) := \begin{cases} 1 & \text{if } r = s \\ 0 & \textit{otherwise.} \end{cases}$$

Once these parameters are fixed, the induced directional measure $\sim_d$ is defined as follows.

Definition 13 (extracted from Def. 3 in [11]). Let $C, D \in \mathcal{C}_{\mathcal{EL}}(\mathsf{N_C}, \mathsf{N_R})$. If $C \equiv \top$, then $C \sim_d D := 1$; if $C \not\equiv \top$ and $D \equiv \top$, then $C \sim_d D := 0$; otherwise, we use $top(C)$ and $top(D)$ to denote the set of $\mathcal{EL}$ atoms occurring in the top-level conjunction of C and D, and define

$$C \sim_d D := \frac{\sum\limits_{C' \in top(C)} \left[g(C') \times \max\limits_{D' \in top(D)} \left(simi_a(C', D') \right) \right]}{\sum\limits_{C' \in top(C)} g(C')}, \text{ where}$$

$simi_a(A, B) := pm(A, B)$ for all $A, B \in \mathsf{N_C}$,
$simi_a(\exists r.E, \exists s.F) := pm(r, s)[w + (1 - w)(E \sim_d F)]$, and
$simi_a(C', D') := 0$ in any other case.

The following two properties are satisfied by $\sim_d$ (see Lemma 1 in [11]). They will be useful later on to obtain our results. Let C, D and E be $\mathcal{EL}$ concept descriptions, then:

$$C \sim_d D \;=\; 1 \text{ iff } D \sqsubseteq C \tag{2}$$

$$D \sqsubseteq E \;\Rightarrow\; C \sim_d E \leq C \sim_d D \tag{3}$$

The proofs can be found in the extended version [10] of [11] (Lemma 14 and Lemma 15). They indicate that these properties hold regardless of whether the concepts C, D and E are in reduced form or not.

If $\otimes$, g and pm can be computed in polynomial time, then the induced CSM $\sim$ can also be computed in polynomial time (see [11], Lemma 2). Moreover, all the CSMs obtained as instances of *simi* where g assigns 1 to atoms of the form $\exists r.C$ are standard CSMs (see Lemma 30 in the Appendix). One such instance of *simi* is $\sim^*$, where $\otimes = \min$, $w = 0$, g assigns 1 to all atoms, and $pm = pm_d$. We now define a class of instances of *simi* containing $\sim^*$.

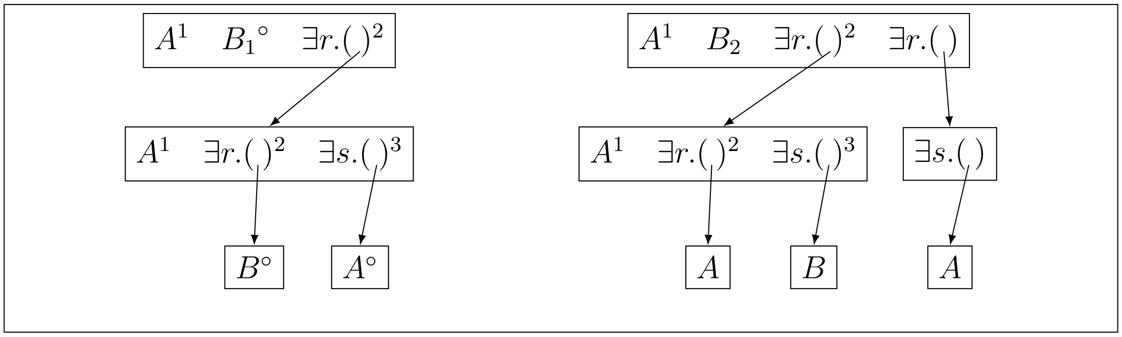

A^1 $B_1{}^\circ$ $\exists r.(\,)^2$
A^1 B_2 $\exists r.(\,)^2$ $\exists r.(\,)$
A^1 $\exists r.(\,)^2$ $\exists s.(\,)^3$
A^1 $\exists r.(\,)^2$ $\exists s.(\,)^3$
$\exists s.(\,)$
B°
A°
A
B
A

Figure 2: Computation of $simi_d$

Definition 14. The class *simi-mon* is obtained from *simi* by restricting the admissible parameters as follows:

- $\otimes$ is computable in polynomial time and monotonic w.r.t. $\geq$;[7]
- g is computable in polynomial time and assigns 1 to all atoms of the form $\exists r.C$;
- $pm = pm_d$ and w is arbitrary.

In the following we will show that, for all $\sim\, \in$ *simi-mon*, reasoning in $\tau\mathcal{EL}(m_\sim)$ is *not* harder than reasoning in $\tau\mathcal{EL}(deg)$. We start with illustrating some useful properties satisfied by CSMs in *simi-mon*.

3.3.1 Some properties satisfied by measures in *simi-m*$\ast$*n*

We first illustrate such properties through the following example.

Example 15. We consider a CSM $\sim$ whose definition deviates from the one of $\sim^*$ only in one place: we use $w = .5$. Consider

$$C := A \sqcap B_1 \sqcap \exists r.(A \sqcap \exists r.B \sqcap \exists s.A),$$
$$D := A \sqcap B_2 \sqcap \exists r.(A \sqcap \exists r.A \sqcap \exists s.B) \sqcap \exists r.\exists s.A.$$

Figure 2 basically shows the atoms in D chosen by max when computing $C \sim_d D$. The *superscripts* are used to denote the corresponding pairings for which the value is > 0. For instance, at the top level of C, A^1 means that A is paired with the top-level atom of D having the same superscript. The symbol $\circ$ on the left-hand side tells us that no match yielding a value > 0 exists. Now, removing the atoms without superscript in D yields the concept $Y := A \sqcap \exists r.(\exists r.\top \sqcap \exists s.\top)$. One can easily verify that $C \sim_d D = C \sim_d Y = 5/9$, and it is clear that C and D are both subsumed by Y.

[7] Examples are *average* and all polynomially computable bounded *t-norms*.

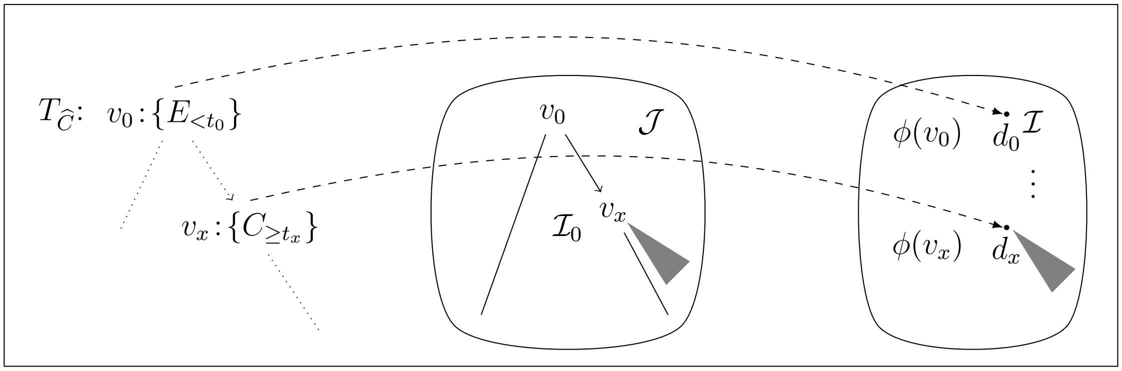


Figure 3: Polynomial bounded model construction

These properties can be generalized to all pair of concepts and measures in *simi-mon*, as stated in the following lemma whose proof is deferred to the Appendix.

Lemma 16. *Let $\sim \in$ simi-mon. For all $\mathcal{EL}$ concept descriptions C and D, there exists an $\mathcal{EL}$ concept description Y such that:*

1. *$D \sqsubseteq Y$ and $\mathsf{s}(Y) \leq \mathsf{s}(C)$,*
2. *$C \sim_d D = C \sim_d Y$,*
3. *$C \sqsubseteq Y$.*

3.3.2 Decidability in NP/coNP

We use the properties shown in Lemma 16 to prove that, like $\tau\mathcal{EL}(deg)$ (see Lemma 4 in [2]), $\tau\mathcal{EL}(m_\sim)$ enjoys a polynomial model property if $\sim \in$ *simi-mon.*

Lemma 17. *Let $\sim \in$ simi-mon and $\widehat{C}$ a $\tau\mathcal{EL}(m_\sim)$ concept description. If $\widehat{C}$ is satisfiable, then there is a tree-shaped interpretation $\mathcal{J}$ such that $\widehat{C}^{\mathcal{J}} \neq \emptyset$ and $|\Delta^{\mathcal{J}}| \leq \mathsf{s}(\widehat{C})$.*

Proof. Figure 3 outlines the description tree $T_{\widehat{C}}$ of a $\tau\mathcal{EL}(m_\sim)$ concept $\widehat{C}$, an interpretation $\mathcal{I}$ such that $d_0 \in \widehat{C}^{\mathcal{I}}$, and a corresponding τ-homomorphism ϕ obtained by applying Theorem 4. The tree in the middle represents the small interpretation $\mathcal{J}$ we want to build. The construction of $\mathcal{J}$ starts with a base interpretation $\mathcal{I}_0$ that corresponds to $T_{\widehat{C}}$ (first ignoring labels of the form $C_{\bowtie t}$). Consequently, the identity mapping ϕ_{id} from $T_{\widehat{C}}$ to $G_{\mathcal{I}_0}$ satisfies Conditions 1 and 2 required for τ-homomorphisms. However, the third condition need not be satisfied since, for instance, $m_\sim^{\mathcal{I}_0}(v_x, C)$ could well be smaller than t_x. To fix this, $\mathcal{I}_0$ is extended into $\mathcal{J}$ by attaching to v_x a *tree-shaped* interpretation (the gray triangle in the figure) such that $m_{\sim}^{\mathcal{J}}(v_x, C) \geq t_x$. This interpretation can be extracted from $\mathcal{I}$ using the fact that $\phi(v_x) = d_x$ implies that $m_{\sim}^{\mathcal{I}}(d_x, C) \geq t_x$ (because ϕ is a τ-homomorphism). To be more precise, consider an $\mathcal{EL}$ concept description D such that

$$d_x \in D^{\mathcal{I}} \text{ and } m_{\sim}^{\mathcal{I}}(d_x, C) = C \sim D.$$

In principle, we could use the interpretation $\mathcal{I}_D$ having the description tree T_D as the one to be attached to v_x. Note that $m_{\sim}^{\mathcal{I}}(d_x, C) \geq t_x$ implies that $C \sim D \geq t_x$. Moreover, $v_x \in D^{\mathcal{J}}$ would hold, and then by definition of $m_{\sim}$ we would obtain that $m_{\sim}^{\mathcal{J}}(v_x, C) \geq C \sim D \geq t_x$. However, we do not know anything about the size of D. This is where Lemma 16 comes into play. Instead of D it allows us to use the concept Y.

We now explain why it is possible to do that. First, we apply Lemma 16 with respect to the reduced form C^r of C. Then, Statement 1. tells us that $\mathsf{s}(Y) \leq \mathsf{s}(C^r)$ and that d_x also belongs to $Y^{\mathcal{I}}$ (since $D \sqsubseteq Y$). Statement 2. shows that Y yields the same value as D in the directional measure, i.e.,

$$C^r \sim_d D = C^r \sim_d Y$$

Now, using Property 3 and the fact that $D^r \sqsubseteq D$ and $D \sqsubseteq D^r$ (similarly for Y and Y^r, we also have:

$$C^r \sim_d D^r = C^r \sim_d Y^r \tag{4}$$

Finally, Statement 3. can be used to show that (4) also holds for $\sim$.

- $C \sqsubseteq Y$ implies that $Y^r \sim_d C^r = 1$ (by Property 2). Therefore, $C \sim Y = (C^r \sim_d Y^r) \otimes 1$. As $\otimes$ is monotonic and commutative, by definition of $\sim$ it holds that $C \sim D \leq C \sim Y^1$ (see (1)). But then, it must be the case that $C \sim D = C \sim Y$, for otherwise the *maximality* of D in Definition 5 would be contradicted.

This approach can be applied to all threshold concepts of the form $C'_{>t}$ or $C'_{\geq t}$ occurring in $\widehat{C}$. Let $\mathfrak{J}$ denote the set of interpretations used to extend $\mathcal{I}_0$ into $\mathcal{J}$, selected in the way we have just described. We remark that, except for nodes like v_x in $\mathcal{I}_0$ (where they will be plugged-in), their domain sets and $\Delta^{\mathcal{I}_0}$ are considered as pairwise disjoint. Then,

$$|\Delta^{\mathcal{J}}| = |\Delta^{\mathcal{I}_0}| + \sum_{\mathcal{I}_Y \in \mathfrak{J}} |\Delta^{\mathcal{I}_Y}| \tag{5}$$

For each threshold concept $C'_{>t}$ or $C'_{\geq t}$ occurring in $\widehat{C}$, the number of domain elements added to satisfy it (in the corresponding $\Delta^{\mathcal{I}_Y}$) is bounded by the size of C'. Notice, that Lemma 16 is applied to $(C')^r$ and $\mathsf{s}(Y) \leq \mathsf{s}((C')^r)$. Moreover, since as shown in [9] reduced forms are the smallest representatives of their equivalence classes, this confirms that $|\Delta^{\mathcal{I}_Y}| \leq \mathsf{s}(C')$. Finally, since the size of $\mathcal{I}_0$ is bounded by the size of $\widehat{C}$ (without counting the threshold concepts occurring in it), it thus holds that $|\Delta^{\mathcal{J}}| \leq \mathsf{s}(\widehat{C})$. The tree shape of $\mathcal{J}$ is guaranteed since $\mathcal{I}_0$ is tree-shaped and only *fresh* tree-shaped interpretations $\mathcal{I}_Y$ are attached to it.

Nevertheless, it remains to see why threshold concepts using $<$ or $\leq$, like $E_{<t_0}$ in the figure, are not violated. The reason is basically that they are satisfied in $\mathcal{I}$, and that everything occurring in the attached pieces T_Y also occurs in $\mathcal{I}$ (since $d_x \in Y^{\mathcal{I}}$). This intuition can be formally justified through the following observations:

- Since ϕ is a τ-homomorphism from $T_{\widehat{C}}$ to $G_{\mathcal{I}}$, it is also a classical homomorphism from $T_{\mathcal{I}_0}$ to $G_{\mathcal{I}}$ (i.e., it satisfies Conditions 1 and 2 required for τ-homomorphisms).
- Each T_Y added to a node v in $T_{\mathcal{I}_0}$ is such that $\phi(v) \in Y^{\mathcal{I}}$. Consequently, using Theorem 3, it is not hard to extend ϕ to a classical homomorphism φ from $G_{\mathcal{J}}$ to $G_{\mathcal{I}}$ such that $\varphi(v) = \phi(v)$ for all $v \in \Delta^{\mathcal{I}_0}$.
- In the particular case of v_0, let F be a concept such that $v_0 \in F^{\mathcal{J}}$ and $m^{\mathcal{J}}_{\sim}(v_0, E) = E \sim F$. By definition of $m_{\sim}$, this means that $v_0 \in F^{\mathcal{J}}$. Let w_0 be the root of the description tree T_F associated to F, by Theorem 3 there exists a homomorphism φ_0 from T_F to $G_{\mathcal{J}}$ such that $\varphi_0(w_0) = v_0$. Then, the composition $\varphi \circ \varphi_0$ is a homomorphism from T_F to $G_{\mathcal{I}}$ with $(\varphi \circ \varphi_0)(w_0) = d_0$. Hence, another application of Theorem 3 yields that $d_0 \in F^{\mathcal{I}}$. Consequently,

  $$m^{\mathcal{J}}_{\sim}(v_0, E) \leq m^{\mathcal{I}}_{\sim}(d_0, E)$$

  We know that $m^{\mathcal{I}}_{\sim}(d_0, E) < t_0$ (because $\phi(v_0) = d_0$ and ϕ is a τ-homomorphism). Thus, $m^{\mathcal{J}}_{\sim}(v_0, E) < t_0$. For any other occurrence of a threshold concept $E'_{<t}$ or $E'_{\leq t}$ in the label of a node v' of $T_{\widehat{C}}$, the same reasoning can be applied based on the fact that $\varphi(v') = \phi(v')$ as mentioned in the previous point.

Overall, we can then conclude that $\mathcal{J}$ satisfies $\widehat{C}$. □

Lemma 16 can also be used to show the following lemma (see the Appendix for the proof).

Lemma 18. *Let $\sim \in$ simi-mon. Additionally, let $\mathcal{I}$ be an interpretation, $d \in \Delta^{\mathcal{I}}$ and $C \in \mathcal{C}_{\mathcal{EL}}(\mathsf{N_C}, \mathsf{N_R})$ with $\mathsf{rd}(C) = k$. Then, there exists a concept $D \in \mathcal{R}^{k+1}$ such that:*

1. *$d \in D^{\mathcal{I}}$ and $C^r \sim_d D = \max\{C^r \sim_d D' \mid D' \in \mathcal{R}^{k+1}$ and $d \in (D')^{\mathcal{I}}\}$,*
2. *$m_{\sim}^{\mathcal{I}}(d, C) = (C^r \sim_d D) \otimes 1$.*

This lemma tells us, that to compute $m_{\sim}^{\mathcal{I}}(d, C)$, it is enough to compute the value:

$$\max\{C^r \sim_d D' \mid D' \in \mathcal{R}^{k+1} \text{ and } d \in (D')^{\mathcal{I}}\}$$

Based on this, we provide an algorithm (Algorithm 3 in the Appendix), which correctly computes the value $m_{\sim}^{\mathcal{I}}(d, C)$ for finite interpretations $\mathcal{I}$ in time polynomial in the size of $\mathcal{I}$ and C.

Proposition 19. *Let $\sim \in$ simi-mon. For every finite interpretation $\mathcal{I}$, $d \in \Delta^{\mathcal{I}}$, and $\mathcal{EL}$ concept description C, $m_{\sim}^{\mathcal{I}}(d, C)$ can be computed in time polynomial in the size of $\mathcal{I}$ and C.*

Together with this proposition, Lemma 17 yields a standard guess-and-check NP-procedure for satisfiability in $\tau\mathcal{EL}(m_{\sim})$. Regarding the other reasoning tasks, the constructions introduced in [2] for $\tau\mathcal{EL}(deg)$ to provide appropriate bounded model properties for them can also be applied for $\tau\mathcal{EL}(m_{\sim})$:

- *subsumption:* a polynomial bounded model property is shown in [2] for $\tau\mathcal{EL}(deg)$ concept descriptions of the form $\widehat{C} \sqcap \neg\widehat{D}$. This yields an NP decision procedure for the *non-subsumption* problem, hence the subsumption problem is in coNP. Such a construction starts with an interpretation $\mathcal{J}$ satisfying $\widehat{C}$, and proceeds to extend it into an interpretation $\mathcal{J}_p$ satisfying $\widehat{C} \sqcap \neg\widehat{D}$, by attaching to $\mathcal{J}$ interpretations of size polynomial in $\mathsf{s}(\widehat{D})$. The construction can then be adapted to $\tau\mathcal{EL}(m_{\sim})$, by using the interpretations associated to the concept Y obtained from the application of Lemma 16.
- *ABox consistency* and *instance checking* are generalizations of *satisfiability* and *subsumption*, respectively. The same technique can be used to replicate the constructions provided for them in the setting of $\tau\mathcal{EL}(deg)$ to $\tau\mathcal{EL}(m_{\sim})$.

Theorem 20. *Let $\sim \in$ simi-mon. In $\tau\mathcal{EL}(m_{\sim})$, satisfiability and consistency are in NP, whereas subsumption and instance checking (w.r.t. data complexity) are in coNP.*

3.3.3 NP-hardness

In [2], satisfiability in $\tau\mathcal{EL}(deg)$ is shown to be NP-hard by reducing an NP-complete variant $\mathcal{V}$ of propositional satisfiability to it. More precisely, such a variant $\mathcal{V}$ corresponds to the problem ALL-POS ONE-IN-THREE 3SAT (see [8], page 259), which we now introduce.

Definition 21 (ALL-POS ONE-IN-THREE 3SAT)**.** Let U be a set of propositional variables and $\mathcal{C}$ be a finite set of propositional clauses over U such that:

- each clause in $\mathcal{C}$ is a set of three literals over U, and
- no $c \in \mathcal{C}$ contains a negated literal.

ALL-POS ONE-IN-THREE 3SAT is the problem of deciding whether there exists a truth assignment to the variables in U, such that each clause in $\mathcal{C}$ has exactly one true literal.

Nevertheless, the reduction provided in [2] introduces a *fresh* concept name for each propositional variable occurring in an instance of $\mathcal{V}$. Since in the present paper we assume that concept descriptions in $\tau\mathcal{EL}(m_\sim)$ are defined over a fixed finite vocabulary $\mathsf{N_C} \cup \mathsf{N_R}$, it is thus not possible to use the same reduction. We will now present a new reduction that shows that satisfiability in $\tau\mathcal{EL}(m_\sim)$ is NP-hard, even if only one concept name and one role name is available. However, for this result to hold we need additional restrictions on $\sim$. Let *simi-smon* be the subset of *simi-mon* whose measures are defined using a fuzzy connector $\otimes$ that:

- is *strictly monotonic*, i.e.,

  $$x < y \Rightarrow x \otimes z < y \otimes z \text{ holds for all } x, y, z \in [0,1]$$

  , or
- has 1 as an *identity element*, i.e., $x \otimes 1 = x$ holds for all $x \in [0,1]$.

We know prove NP-hardness for the satisfiability problem in all threshold logics $\tau\mathcal{EL}(m_\sim)$ induced by a similarity measure $\sim$ in *simi-smon*.

In what follows we assume that the propositional variables occurring in any formula φ (as described in Definition 21) are ordered, and we will refer to them as $x_1, \ldots, x_n$. To cope with the fix number of concept names in $\mathsf{N_C}$, we use existential restrictions to simulate/represent the propositional variables occurring in a propositional formula φ. Let $c_1 \wedge \ldots \wedge c_q$ be the clauses of φ (as considered in Definition 21), and $x_1, \ldots, x_n$ the propositional variables occurring in φ. Truth assignments of the variables $x_1, \ldots, x_n$ can be identified in an interpretation $\mathcal{I}$ as follows. First, for each variable x_i occurring in φ, we associate to it the concept description $X^{\{i\}}$ which has the following definition:

$$X^{\{i\}} := \exists \underbrace{r \ldots r}_{i}.A$$

Second, let $d \in \Delta^{\mathcal{I}}$, then the pair $(\mathcal{I}, d)$ induces a truth assignment $\mathfrak{t}_d$ such that for all $1 \leq i \leq n$:

$$\mathfrak{t}_d(x_i) = \mathit{true} \text{ iff } d \in \left(X^{\{i\}}\right)^{\mathcal{I}} \tag{6}$$

The idea for our reduction is to construct a $\tau\mathcal{EL}(m_\sim)$ concept description $\widehat{C}_\varphi$ such that $d \in (\widehat{C}_\varphi)^\mathcal{I}$ iff $\mathfrak{t}_d$ satisfies exactly one literal in each clause of φ. To this end, we define two concepts $\widehat{D}_1$ and $\widehat{D}_2$. One expresses that no two variables are satisfied by $\mathfrak{t}_d$ in the same clause, while the other one states that at least one must be satisfied.

For each pair of variables x_i and x_j occurring in φ $(i \neq j)$, we define the concept description $X^{\{i,j\}}$ as:

$$X^{\{i\}} \sqcap X^{\{j\}}$$

Using these concepts, we construct the threshold concepts $\left(X^{\{i,j\}}\right)_{<1}$ to express that the variables x_i and x_j are not both mapped to *true* by an assignment $\mathfrak{t}_d$.

Lemma 22. *Let $\mathcal{I}$ be an interpretation and $d \in \Delta^\mathcal{I}$. Then, $d \in [(X^{\{i,j\}})_{<1}]^\mathcal{I}$ iff $\mathfrak{t}_d(x_i) = \mathit{false}$ or $\mathfrak{t}_d(x_j) = \mathit{false}$.*

Proof. Suppose that $d \in [(X^{\{i,j\}})_{<1}]^\mathcal{I}$. Since $m_\sim$ satisfies property *M1*, this means that $d \notin (X^{\{i,j\}})^\mathcal{I}$. Therefore,

$$d \notin \left(X^{\{i\}}\right)^\mathcal{I} \text{ or } d \notin \left(X^{\{j\}}\right)^\mathcal{I}$$

Hence, the claim holds by construction of $\mathfrak{t}_d$ in (6). The converse can be proved in a similar way. □

Thus, to enforce that $\mathfrak{t}_d$ does not satisfy two literals in a clause of φ, we define the following concept description $\widehat{D}_1$:

$$\widehat{D}_1 := \bigsqcap_{c \in \varphi} \bigsqcap_{\substack{x_i, x_j \in c \\ i \neq j}} \left(X^{\{i,j\}}\right)_{<1}$$

It remains to show how to express that $\mathfrak{t}_d$ must satisfy at least one. For each clause c_k $(1 \leq k \leq q)$ we define its corresponding $\mathcal{EL}$ concept description C_k as $\exists r.E_1^k$, where E_1^k is of the following form:

$$\begin{aligned}
&E_1^k := \gamma_1^k \sqcap \exists r.E_2^k \\
&\dots \\
&E_i^k := \gamma_i^k \sqcap \exists r.E_{i+1}^k \quad (1 \leq i < n) \\
&\dots \\
&E_n^k := \gamma_n^k
\end{aligned}$$

Here $\gamma_i^k = \top$ if x_i does not occur in c_k, otherwise $\gamma_i^k = A$.

Example 23. Let φ be the following propositional formula in CNF:

$$\{x_1, x_2, x_3\} \wedge \{x_1, x_4, x_3\} \wedge \{x_4, x_2, x_3\}$$

A total of four propositional variables and three clauses occur in φ. Then, the concept descriptions C_1, C_2 and C_3 are the ones having the following $\mathcal{EL}$ description trees:

$$T_{C_1}\colon \quad \{\} \overset{r}{\rule{2em}{0.4pt}} \{A\} \overset{r}{\rule{2em}{0.4pt}} \{A\} \overset{r}{\rule{2em}{0.4pt}} \{A\} \overset{r}{\rule{2em}{0.4pt}} \{\}$$

$$T_{C_2}\colon \quad \{\} \overset{r}{\rule{2em}{0.4pt}} \{A\} \overset{r}{\rule{2em}{0.4pt}} \{\} \overset{r}{\rule{2em}{0.4pt}} \{A\} \overset{r}{\rule{2em}{0.4pt}} \{A\}$$

$$T_{C_3}\colon \quad \{\} \overset{r}{\rule{2em}{0.4pt}} \{\} \overset{r}{\rule{2em}{0.4pt}} \{A\} \overset{r}{\rule{2em}{0.4pt}} \{A\} \overset{r}{\rule{2em}{0.4pt}} \{A\}$$

The nodes at the i^{th} level (except for the root) tell us whether the variable x_i occurs in a clause of φ. For x_2, the empty set (or $\top$) is used in T_{C_2} to represent that x_2 does not occur in c_2, while $\{A\}$ is used in the the other two trees to state that x_2 occurs in c_1 and c_3. The same idea applies for the rest of the variables occurring in φ.

One can easily verify that for all concepts C_k and variables x_i occurring in the clause c_k of φ, it holds that $C_k \sqsubseteq X^{\{i\}}$. The idea now is to use a threshold concept $(C_k)_{\geq t_k}$ to express that a domain element d is an instance of at least one $X^{\{i\}}$. For this to work, we have to show that t_k can be selected such that if that were not the case, then $m^{\mathcal{I}}_{\sim}(d, C_k)$ would always be smaller than t_k.

Let $X^{\{i\}}_n$ denote the following extension of $X^{\{i\}}$:

$$X^{\{i\}}_n := \exists \underbrace{r \ldots r}_{i}.(A \sqcap \exists \underbrace{r \ldots r}_{n-i}.\top) \qquad \Big(\text{vs. } X^{\{i\}} = \exists \underbrace{r \ldots r}_{i}.A\Big)$$

Again, it is easy to see that for all concepts C_k and variables x_i occurring in the clause c_k of φ, $C_k \sqsubseteq X^{\{i\}}_n \sqsubseteq X^{\{i\}}$ holds.

The following lemma tells us why such a value for t_k always exists.

Lemma 24. *Let c_k be a clause in φ and x_i, x_j, x_ℓ the variables occurring in it. Additionally, let D be an $\mathcal{EL}$ concept description such that:*

$$D \not\sqsubseteq X^{\{i\}}, \quad D \not\sqsubseteq X^{\{j\}} \quad \textit{and} \quad D \not\sqsubseteq X^{\{\ell\}}$$

Then, for all $\sim \in$ simi-mon, $C_k \sim_d D^r < C_k \sim_d X^{\{a\}}_n$, $a \in \{i, j, \ell\}$.

Proof. The proof can be found in the Appendix. □

Now, let $\bar{t}$ and $\underline{t}$ be the following values:

$$\bar{t} := \max\{(C_k \sim_d D^r) \otimes 1 \mid D \not\sqsubseteq X^{\{a\}}, \text{ for all } a \in \{i, j, \ell\}\}$$

$$\underline{t} := \min_{x_i \,\in\, c_k} \{(C_k \sim_d X_n^{\{i\}}) \otimes 1\}$$

From the proof of Lemma 24, one can see that the maximum defining $\bar{t}$ always exists. Moreover, for all $\sim \in$ *simi-smon*, whether is $\otimes$ *strictly monotonic* or it has 1 as *identity element*, by Lemma 24: $\bar{t} < \underline{t}$ holds. Then, we define the threshold concept $(C_k)_{\geq t_k}$ by selecting t_k as a rational number such that:

$$\bar{t} < t_k < \underline{t} \tag{7}$$

Lemma 25. *Let $\sim \in$ simi-smon, $\mathcal{I}$ be an interpretation and $d \in \Delta^{\mathcal{I}}$. If $m^{\mathcal{I}}_{\sim}(d, C_k) \geq t_k$, then $d \in \left(X^{\{i\}}\right)^{\mathcal{I}}$ for at least one variable x_i occurring in c_k.*

Proof. Suppose that $m^{\mathcal{I}}_{\sim}(d, C_k) \geq t_k$. Then, by Definition 5, there exists a concept description D such that $d \in D^{\mathcal{I}}$ and $m^{\mathcal{I}}_{\sim}(d, C_k) = C_k \sim D \geq t_k$. We want to show that $D \sqsubseteq X^{\{i\}}$ for some x_i occurring in c_k. Suppose that this is not true. Then, by Lemma 24 we know that:

$$C_k \sim_d D^r < C_k \sim_d X_n^{\{i\}}, \text{ for all } x_i \in c_k \tag{8}$$

Since D is maximal in Definition 5, by the properties shown in Lemma 16 we can assume that:

$$C_k \sim D = (C_k \sim_d D^r) \otimes 1$$

By the selection of t_k in (7), it follows that $C_k \sim D < t_k$ which is a contradiction. Therefore, $D \sqsubseteq X^{\{i\}}$ for at least one variable x_i occurring in c_k. Since $d \in D^{\mathcal{I}}$, this means that $d \in \left(X^{\{i\}}\right)^{\mathcal{I}}$. □

Finally, we define the concept $\widehat{D}_2$ as $\bigsqcap_{k=1}^{q} (C_k)_{\geq t_k}$, and then $\widehat{C}_\varphi$ corresponds to the conjunction $\widehat{D}_1 \sqcap \widehat{D}_2$. The following lemma shows that our reduction is correct.

Lemma 26. *Let $\sim \in$ simi-smon. Moreover, let φ be a propositional formula of the type considered in Definition 21. Then, there exists a truth assignment $\mathfrak{t}$ satisfying exactly one literal in each clause of φ iff $\widehat{C}_\varphi$ is satisfiable in $\tau\mathcal{EL}(m_\sim)$.*

Proof. ($\Leftarrow$) Assume that $\widehat{C}_\varphi$ is satisfiable in $\tau\mathcal{EL}(m_\sim)$. Then, there is an interpretation $\mathcal{I}$ and an element $d \in \Delta^{\mathcal{I}}$ such that $d \in (\widehat{C}_\varphi)^{\mathcal{I}}$. Consequently, $d \in (\widehat{D}_1 \sqcap \widehat{D}_2)^{\mathcal{I}}$. Then, for every clause c_k of φ $(1 \leq k \leq q)$ it hold:

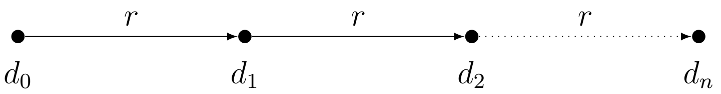
$$d \in [(C_k)_{\geq t_k}]^{\mathcal{I}} \quad \text{and} \quad d \in \left(\bigsqcap_{\substack{x_i, x_j \in\, c_k \\ x_i \neq x_j}} (X^{\{i,j\}})_{<1} \right)^{\mathcal{I}}$$

We take the truth assignment $\mathfrak{t}_d$ induced by the pair $(\mathcal{I}, d)$ as described in (6), and show that it satisfies exactly one literal in each clause of φ. Let c_k be an arbitrary clause of φ. Since $d \in [(C_k)_{\geq t_k}]^{\mathcal{I}}$ it follows that:

$$m_{\sim}^{\mathcal{I}}(d, C_k) \geq t_k$$

Hence, the application of Lemma 25 yields that $d \in \left(X^{\{i\}}\right)^{\mathcal{I}}$ for at least one variable x_i occurring in c_k. Therefore, by construction of $\mathfrak{t}_d$ we have that $\mathfrak{t}_d(x_i) =$ *true*. Now, let x_j and x_ℓ be the other two literals occurring in c_k. Since $d \in \left[(X^{\{i,j\}})_{<1}\right]^{\mathcal{I}}$ and $d \in \left[(X^{\{i,\ell\}})_{<1}\right]^{\mathcal{I}}$, the application of Lemma 22 yields that $\mathfrak{t}_d(x_j) =$ *false* and $\mathfrak{t}_d(x_\ell) =$ *false*. Thus, $\mathfrak{t}_d$ satisfies exactly one literal of c_k.

($\Rightarrow$) Assume that there is a truth assignment $\mathfrak{t}$ satisfying exactly one true literal in each clause of φ. We define the interpretation $\mathcal{I}$ having the following shape:



where $d_i \in A^{\mathcal{I}}$ iff $\mathfrak{t}(x_i) =$ *true*, for all $1 \leq i \leq n$. Note that by definition of $\mathfrak{t}_{d_0}$ in (6), it is the case that $\mathfrak{t} = \mathfrak{t}_{d_0}$. We show that $d_0 \in (\widehat{C}_\varphi)^{\mathcal{I}}$. Let us start with $\widehat{D}_1$. Assume that $d_0 \notin (\widehat{D}_1)^{\mathcal{I}}$, then by definition of $\widehat{D}_1$ there exist a clause c_k of φ and two variables x_i, x_j in c_k, such that $d_0 \notin [(X^{\{i,j\}})_{<1}]^{\mathcal{I}}$. Then, by Lemma 22 we obtain $\mathfrak{t}_{d_0}(x_i) = \mathfrak{t}_{d_0}(x_j) =$ *true*, which is a contradiction since $\mathfrak{t} = \mathfrak{t}_d$ and $\mathfrak{t}$ satisfies exactly one literal in c_k. Thus, $d_0 \in (\widehat{D}_1)^{\mathcal{I}}$.

Concerning $\widehat{D}_2$, let x_i be the variable in an arbitrary clause c_k satisfied by $\mathfrak{t}$. By construction of $\mathcal{I}$ we have that $d_0 \in \left(X_n^{\{i\}}\right)^{\mathcal{I}}$. Therefore,

$$m_{\sim}^{\mathcal{I}}(d_0, C_k) \geq C_k \sim X_n^{\{i\}} \tag{9}$$

Additionally, since $C_k \sqsubseteq X_n^{\{i\}}$, by Property (2) we have that $X_n^{\{i\}} \sim_d C_k = 1$. Consequently, $C_k \sim X_n^{\{i\}} = \left(C_k \sim_d X_n^{\{i\}}\right) \otimes 1$. Hence, combining (7) and (9) we obtain $m_{\sim}^{\mathcal{I}}(d_0, C_k) \geq t_k$ and $d_0 \in [(C_k)_{\geq t_k}]^{\mathcal{I}}$. Therefore, $d \in (\widehat{D}_2)^{\mathcal{I}}$.

Overall, we have shown that $d_0 \in (\widehat{D}_1 \sqcap \widehat{D}_2)^{\mathcal{I}}$. Thus, $\widehat{C}_\varphi$ is satisfiable in $\tau\mathcal{EL}(m_\sim)$. □

Since satisfiability can be reduced to the consistency, non-subsumption and non-instance problem, we thus obtain the following hardness results.

Proposition 27. *Let $\sim \in$ simi-smon. In $\tau\mathcal{EL}(m_\sim)$, satisfiability and consistency are NP-hard, whereas subsumption and instance checking are coNP-hard.*

3.4 Relaxed instance checking

In [2], it was shown for $\tau\mathcal{EL}(deg)$ that instance checking becomes polynomial if instead of arbitrary $\tau\mathcal{EL}(deg)$ concept descriptions one considers only threshold concepts of the form $C_{>t}$. We can show that this result holds not just for *deg*, but for all CSMs in our class *simi-mon*.

Proposition 28. *Let $\sim \in$ simi-mon. In $\tau\mathcal{EL}(m_\sim)$, the instance checking problem for threshold concepts of the form $C_{>t}$ can be decided in polynomial time.*

Proof. Let $\mathcal{A}$ be an $\mathcal{EL}$ ABox and a an individual of $\mathcal{A}$. Moreover, let $\mathcal{I}_\mathcal{A}$ be the canonical interpretation of $\mathcal{A}$. Since $\mathcal{I}_\mathcal{A} \models \mathcal{A}$, this means that if $a^\mathcal{I} \in (C_{>t})^\mathcal{I}$, then $m_\sim^{\mathcal{I}_\mathcal{A}}(a^{\mathcal{I}_\mathcal{A}}, C) > t$. Furthermore, it is well-known that for $\mathcal{EL}$, if $a^{\mathcal{I}_\mathcal{A}} \in D^{\mathcal{I}_\mathcal{A}}$, then $a^\mathcal{I} \in D^\mathcal{I}$ for all models $\mathcal{I}$ of $\mathcal{A}$. Hence, by Definition 5 it follows that $t < m_\sim^{\mathcal{I}_\mathcal{A}}(a^{\mathcal{I}_\mathcal{A}}, C) \leq m_\sim^\mathcal{I}(a^\mathcal{I}, C)$ for all $\mathcal{I}$.

Thus, to decide if $\mathcal{A} \models C_{>t}(a)$ holds, it suffices to verify whether $m_\sim^{\mathcal{I}_\mathcal{A}}(a^{\mathcal{I}_\mathcal{A}}, C) > t$ holds. This can be done in polynomial time, since $\mathcal{I}_\mathcal{A}$ is linear on the size of $\mathcal{A}$ and $m_\sim^{\mathcal{I}_\mathcal{A}}$ can be computed in polynomial time. □

Since it was shown in [2] (Proposition 5) that computing instances of threshold concepts of the form $C_{>t}$ in a logic $\tau\mathcal{EL}(m_\sim)$ corresponds to answering so-called relaxed instance queries w.r.t. $\sim$ (see [7]), this also yields a polynomiality result for answering relaxed instance queries w.r.t. CSMs in *simi-mon*.

4 Conclusions

We have shown that the complexity results for reasoning in the threshold logic $\tau\mathcal{EL}(deg)$ of [2] can be extended to a large class of logics $\tau\mathcal{EL}(m_\sim)$ that are induced by appropriate concept similarity measures. Like in [2], we do not consider terminological axioms (TBoxes) in the present paper. In [3], reasoning w.r.t. acyclic TBoxes in $\tau\mathcal{EL}(deg)$ was considered. It would be interesting to see whether the results of [3], which surprisingly show that acyclic TBoxes increase the complexity, can also be extended to our logics $\tau\mathcal{EL}(m_\sim)$ for $\sim \in$ *simi-mon*.

References

[1] Franz Baader. Terminological cycles in a description logic with existential restrictions. In *Proc. of the 18th Int. Joint Conf. on Artificial Intelligence (IJCAI 2003)*, pages 325–330, 2003.

[2] Franz Baader, Gerhard Brewka, and Oliver Fernández Gil. Adding threshold concepts to the description logic $\mathcal{EL}$. In *Proc. of the 10th Int. Symp. on Frontiers of Combining Systems (FroCoS 2015)*, volume 9322 of *LNCS*, pages 33–48. Springer, 2015.

[3] Franz Baader and Oliver Fernández Gil. Extending the description logic $\tau\mathcal{EL}(deg)$ with acyclic TBoxes. In *Proc. ECAI'16*, volume 285, pages 1096–1104. IOS Press, 2016.

[4] Franz Baader, Ralf Küsters, and Ralf Molitor. Computing least common subsumers in description logics with existential restrictions. In *Proc. of the 16th Int. Joint Conf. on Artificial Intelligence (IJCAI'99)*, pages 96–103, 1999.

[5] Franz Baader, Baris Sertkaya, and Anni-Yasmin Turhan. Computing the least common subsumer w.r.t. a background terminology. *J. of Applied Logic*, 5(3):392–420, 2007.

[6] Sebastian Brandt. Polynomial time reasoning in a description logic with existential restrictions, GCI axioms, and - what else? In *Proc. ECAI'04*, pages 298–302. IOS Press, 2004.

[7] Andreas Ecke, Rafael Peñaloza, and Anni-Yasmin Turhan. Similarity-based relaxed instance queries. *J. Applied Logic*, 13(4):480–508, 2015.

[8] M. R. Garey and David S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness.* W. H. Freeman, 1979.

[9] Ralf Küsters. *Non-Standard Inferences in Description Logics*, volume 2100 of *LNCS*. Springer, 2001.

[10] Karsten Lehmann. A framework for semantic invariant similarity measures for $\mathcal{ELH}$ concept descriptions. Master's thesis, TU Dresden, Germany, 2012. Available from: http://lat.inf.tu-dresden.de/research/mas.

[11] Karsten Lehmann and Anni-Yasmin Turhan. A framework for semantic-based similarity measures for $\mathcal{ELH}$-concepts. In *Proc. of the 13th Eur. Conf. on Logics in Artificial Intelligence (JELIA'2012)*, volume 7519 of *LNCS*, pages 307–319. Springer, 2012.

[12] Manfred Schmidt-Schauß and Gert Smolka. Attributive concept descriptions with complements. *Artif. Intell.*, 48(1):1–26, 1991.

[13] Boontawee Suntisrivaraporn. A similarity measure for the description logic $\mathcal{EL}$ with unfoldable terminologies. In *5th Int. Conf. on Intelligent Networking and Collaborative Systems*, pages 408–413. IEEE, 2013.

5 Appendix

Missing proofs of Section 2

The role depth of a $\tau\mathcal{EL}(m)$ concept description $\widehat{C}$ is defined as for $\mathcal{EL}$ concept descriptions, by considering threshold concepts as atomic concepts. Formally, we denote a $\tau\mathcal{EL}(m)$ description graph $\widehat{G}$ as a tuple $(V_G, E_G, \widehat{\ell}_G)$ where:

- V_G is a set of nodes,
- $E_G \subseteq V_G \times \mathsf{N_R} \times V_G$ is a set of edges labeled with role names, and
- $\widehat{\ell}_G$ is a function labeling the nodes in V_G.

We write $\ell_G(v)$ to denote the subset of $\widehat{\ell}_G(v)$ containing only labels from $\mathsf{N_C}$.

Theorem 4. Let $\mathcal{I}$ be an interpretation with associated $\mathcal{EL}$ description graph $G_{\mathcal{I}}$, $d \in \Delta^{\mathcal{I}}$, and $\widehat{C}$ a $\tau\mathcal{EL}(m)$ concept description with associated $\tau\mathcal{EL}(m)$ description tree $T_{\widehat{C}}$. Then, $d \in \widehat{C}^{\mathcal{I}}$ iff there exists a τ-homomorphism ϕ from $T_{\widehat{C}}$ to $G_{\mathcal{I}}$ such that $\phi(v_0) = d$.

Proof. Let $T_{\widehat{C}} = (V_T, E_T, v_0, \widehat{\ell}_T)$ be the description tree associated to $\widehat{C}$ and $\widehat{C}$ be of the form $\widehat{C}_1 \sqcap \ldots \sqcap \widehat{C}_q \sqcap \exists r_1.\widehat{D}_1 \sqcap \ldots \sqcap \exists r_n.\widehat{D}_n$, where each $\widehat{C}_i$ is either a concept name $A \in \mathsf{N_C}$ or a threshold concept $E_{\sim t}$.

($\Rightarrow$) Assume that $d \in \widehat{C}^{\mathcal{I}}$. Then, $d \in (\widehat{C}_i)^{\mathcal{I}}$ and $d \in (\exists r_j.\widehat{D}_j)^{\mathcal{I}}$ for all $1 \leq i \leq q$ and $1 \leq j \leq n$. We show by induction on the *role depth* of $\widehat{C}$ that there exists a τ-homomorphism ϕ from $T_{\widehat{C}}$ to $G_{\mathcal{I}}$ with $\phi(v_0) = d$.

Induction Base. $\mathsf{rd}(\widehat{C}) = 0$. Then, $n = 0$ and $T_{\widehat{C}}$ consists only of one node v_0 (the root), it has no edges, and $\left\{\widehat{C}_1, \ldots, \widehat{C}_q\right\}$ is the label set of v_0. The mapping $\phi(v_0) = d$ is a τ-homomorphism from $T_{\widehat{C}}$ to $G_{\mathcal{I}}$. For each $\widehat{C}_i$ of the form $A \in \mathsf{N_C}$ we know that $d \in A^{\mathcal{I}}$. Therefore, A is contained in the label set of d, and consequently ϕ satisfies Condition 1 required for τ-homomorphisms. In case $\widehat{C}_i$ is of the form $E_{\sim t}$, the fact that $d \in (\widehat{C}_i)^{\mathcal{I}}$ implies that ϕ satisfies the third condition.

Induction Step. Assume that the claim holds for all $\mathcal{EL}$ concept descriptions with role depth smaller than k. We show that it also holds for $\mathsf{rd}(\widehat{C}) = k$.

First, consider the concept $\widehat{D}_0 = \widehat{C}_1 \sqcap \ldots \sqcap \widehat{C}_q$. One can see that $T_{\widehat{D}_0} = (V_0, E_0, v_0, \widehat{\ell}_0)$ is exactly the description tree with $V_0 = \{v_0\}$, $E_0 = \emptyset$ and $\widehat{\ell}_0(v_0) = \widehat{\ell}_T(v_0)$. Since $d \in (\widehat{D}_0)^{\mathcal{I}}$ and $\mathsf{rd}(\widehat{D}_0) = 0$, by induction hypothesis there exists a τ-homomorphism ϕ_0 from $T_{\widehat{D}_0}$ to $G_{\mathcal{I}}$ with $\phi_0(v_0) = d$.

Now, consider any edge $v_0 r_j v_j$ in E_T. By the relationship between $T_{\widehat{C}}$ and $\widehat{C}$, there exists a *top level* atom $\exists r_j.\widehat{D}_j$ of $\widehat{C}$ such that $T_{\widehat{D}_j} = (V_j, E_j, v_j, \widehat{\ell}_j)$ is precisely the subtree of $T_{\widehat{C}}$ with root v_j. In addition, since $d \in (\exists r_j.\widehat{D}_j)^{\mathcal{I}}$ there exists $d_j \in \Delta^{\mathcal{I}}$ such that $d r_j d_j \in E_{\mathcal{I}}$ and $d_j \in (\widehat{D}_j)^{\mathcal{I}}$. Since $\mathsf{rd}(\widehat{D}_j) < k$, the application of the induction hypothesis to d_j and $\widehat{D}_j$ yields a τ-homomorphism ϕ_j from $T_{\widehat{D}_j}$ to $G_{\mathcal{I}}$ with $\phi_j(v_j) = d_j$.

It is not hard to see that for all nodes $v \in V_T$, there exists exactly one of these τ-homomorphisms ϕ_j $(0 \leq j \leq n)$ such that $v \in \mathsf{dom}(\phi_j)$. Based on this, we build a mapping ϕ from V_T to $V_{\mathcal{I}}$ as $\phi = \bigcup_{j=0}^{n} \phi_j$. Note that $\phi(v_0) = d$ by definition of ϕ_0. Hence, it remains to show that ϕ is τ-homomorphism.

1. **ϕ is a homomorphism from T_C to $G_{\mathcal{I}}$:** Let v be any node in V_T. We know that v is a node of one description tree $T_{\widehat{D}_j}$ and $\phi(v) = \phi_j(v)$ for the corresponding mapping ϕ_j. Since ϕ_j is a homomorphism, this means that $\ell_j(v) \subseteq \ell_{\mathcal{I}}(\phi_j(v))$. Therefore, $\ell(v) = \ell_j(v)$ implies $\ell(v) \subseteq \ell_{\mathcal{I}}(\phi(v))$. Now, let vrw be any edge from E_T. There are two possibilities:
   - vrw is of the form $v_0 r_j v_j$. As explained before we have $\phi(v_0) = d$, $\phi_j(v_j) = d_j$ and $d r_j d_j \in E_{\mathcal{I}}$. Hence, $\phi(v_0) r_j \phi(v_j) \in E_{\mathcal{I}}$.
   - $v, w \in \mathsf{dom}(\phi_j)$ for some $j \in \{1 \ldots n\}$. By construction of ϕ and the fact that ϕ_j is a homomorphism, it follows that $\phi(v) r \phi(w) \in E_{\mathcal{I}}$.
2. Condition 3 required for τ-homomorphisms follows from the fact that ϕ is constructed using τ-homomorphisms.

Thus, ϕ is a τ-homomorphism from $T_{\widehat{C}}$ to $G_{\mathcal{I}}$ with $\phi(v_0) = d$.

($\Leftarrow$) Assume that there exists a τ-homomorphism ϕ from $T_{\widehat{C}}$ to $G_{\mathcal{I}}$ with $\phi(v_0) = d$. We show by induction on the size of V_T that $d \in \widehat{C}^{\mathcal{I}}$.

Induction Base. $|V_T| = 1$. Then, $\widehat{C}$ is of the form $\widehat{C}_1 \sqcap \ldots \sqcap \widehat{C}_q$ and $\{\widehat{C}_1, \ldots, \widehat{C}_q\}$ is the label set of v_0. We distinguish two cases for all $\widehat{C}_i$:

- $\widehat{C}_i$ is of the form $A \in \mathsf{N_C}$. Since ϕ is τ-homomorphism, it is also a classical homomorphism, i.e., it satisfies Conditions 1 and 2. Hence, ignoring the labels of the form $E_{\sim t}$ we have $\ell_T(v_0) \subseteq \ell_{\mathcal{I}}(d)$. Thus, $d \in A^{\mathcal{I}}$.
- $\widehat{C}_i$ is of the form $E_{\sim t}$. By Condition 3 required for τ-homomorphisms we also have $d \in (E_{\sim t})^{\mathcal{I}}$.

Hence, $d \in (\widehat{C}_i)^{\mathcal{I}}$ for all conjuncts $\widehat{C}_i$ of $\widehat{C}$. Thus, $d \in \widehat{C}^{\mathcal{I}}$.

Induction Step. Assume that the claim holds for $|V_T| < k$. We show that it also holds for $|V_T| = k$. Since $k > 0$, there exist nodes $v_1, \ldots, v_n$ in V_T such that

$v_0 r_j v_j \in E_T$. This also means that $\widehat{C}$ is of the form $\widehat{C}_1 \sqcap \ldots \sqcap \widehat{C}_q \sqcap \exists r_1.\widehat{D}_1 \sqcap \ldots \sqcap \exists r_n.\widehat{D}_n$ with $n > 0$, and the description tree $T_{\widehat{D}_j} = (V_j, E_j, v_j, \widehat{\ell_j})$ associated to $\widehat{D}_j$ is the subtree of $T_{\widehat{C}}$ rooted at v_j. We consider the following two cases:

- $q > 0$. Then, $d \in (\widehat{C}_i)^{\mathcal{I}}$ can be shown in the same way as for the base case.
- Consider any $\exists r_j.\widehat{D}_j$, with $j \in \{1 \ldots n\}$. Since ϕ is also a homomorphism from $T_{\widehat{C}}$ to $G_{\mathcal{I}}$ and $v_0 r_j v_j \in E_T$, then there exists $e_j \in \Delta^{\mathcal{I}}$ such that $d r_j e_j \in E_{\mathcal{I}}$ and $\phi(v_j) = e_j$. Moreover, it is clear that $|V_j| < |V_T|$ and it is not difficult to see that the restriction of the domain of ϕ to V_j is also a τ-homomorphism from $T_{\widehat{D}_j}$ to $G_{\mathcal{I}}$ with $\phi(v_j) = e_j$. Hence, the application of induction hypothesis yields $e_j \in (\widehat{D}_j)^{\mathcal{I}}$, and this means that $d \in (\exists r_j.\widehat{D}_j)^{\mathcal{I}}$.

Thus, we have shown that $d \in \widehat{C}^{\mathcal{I}}$. □

Deciding the existence of a τ-homomorphism. If the interpretation $\mathcal{I}$ is finite and m is computable, then the existence of a τ-homomorphism can be decided. We present an algorithm that (under the previous conditions) can be used to decide the relation characterized by Theorem 4. As for *deg*, the starting point is the polynomial time algorithm (Algorithm 1 below) introduced in [4] to decide the existence of a homomorphism between two $\mathcal{EL}$ description trees.

Algorithm 1 Homomorphisms between $\mathcal{EL}$ description trees.

Input: Two $\mathcal{EL}$ description trees T_1 and T_2.
Output: "yes", if there exists a homomorphism from T_1 to T_2; "no", otherwise.

```
Let $T_1 = (V_1, E_1, v_0, \ell_1)$ and $T_2 = (V_2, E_2, w_0, \ell_2)$. Further, let $\{v_1, \ldots, v_n\}$ be
  a post-order sequence of $V_1$, i.e., $v_1$ is a leaf and $v_n = v_0$.
Define a labeling $\delta : V_2 \to 2^{V_1}$ as follows.
Initialize $\delta$ by $\delta(w) := \emptyset$ for all $w \in V_2$.
for all $1 \leq i \leq n$ do
    for all $w \in V_2$ do
        if ($\ell_1(v_i) \subseteq \ell_2(w)$ and for all $v_i r v \in E_1$ there is $w' \in V_2$ such that
           $v \in \delta(w')$ and $w r w' \in E_2$) then
            $\delta(w) := \delta(w) \cup \{v_i\}$
        end if
    end for
end for
If $v_0 \in \delta(w_0)$ then return "yes", else return "no".
```

Theorem 4 characterizes *membership* in $\tau\mathcal{EL}(m)$ concept descriptions via the existence of a τ-homomorphism from a $\tau\mathcal{EL}(m)$ description tree $T_{\widehat{C}}$ to an $\mathcal{EL}$

description graph $G_\mathcal{I}$ associated to an interpretation $\mathcal{I}$. If $\mathcal{I}$ is finite, then Algorithm 1 can be used to decide whether there exists a mapping satisfying the three conditions required for τ-homomorphisms. One needs only to replace the last line by $v_0 \in \delta(d)$ for some $d \in \Delta^\mathcal{I}$, since now T_2 becomes $G_\mathcal{I}$. In order to verify the third condition, we modify the test in line 6 to also consider whether $m^\mathcal{I}(d, E) \sim t$ for all $E_{\sim t} \in \widehat{\ell}_{T_{\widehat{C}}}(v_i)$. Algorithm 2 implements this modification.

Algorithm 2 τ-homomorphism from a $\tau\mathcal{EL}(m)$ description tree to $G_\mathcal{I}$.

Input: A $\tau\mathcal{EL}(m)$ description tree $\widehat{T}$ and a finite interpretation $\mathcal{I}$.
Output: "yes", if there exists a τ-homomorphism from $\widehat{T}$ to $G_\mathcal{I}$; "no", otherwise.

```
 1: Let $\widehat{T} = (V_T, E_T, v_0, \widehat{\ell}_T)$ and $G_\mathcal{I} = (V_\mathcal{I}, E_\mathcal{I}, \ell_\mathcal{I})$. Further, let $\{v_1, \ldots, v_n\}$ be a
    post-order sequence of $V_T$, i.e., $v_1$ is a leaf and $v_n = v_0$.
 2: Define a labeling $\delta : V_\mathcal{I} \to 2^{V_T}$ as follows.
 3: Initialize $\delta$ by $\delta(w) := \emptyset$ for all $w \in V_\mathcal{I}$.
 4: for all $1 \leq i \leq n$ do
 5:     for all $d \in \Delta^\mathcal{I}$ do
 6:         if $(\ell_T(v_i) \subseteq \ell_\mathcal{I}(d)$ and $[E_{\sim t} \in \widehat{\ell}_T(v_i) \Rightarrow m^\mathcal{I}(d, E) \sim t]$ and
 7:            $[v_i r v \in E_T \Rightarrow \exists d' \in \Delta^\mathcal{I} : v \in \delta(d')]$ and $d r d' \in E_\mathcal{I})$ then
 8:             $\delta(d) := \delta(d) \cup \{v_i\}$
 9:         end if
10:     end for
11: end for
12: If there exists $d \in \Delta^\mathcal{I}$ such that $v_0 \in \delta(d)$ then return "yes", else return "no".
```

Then, if one wants to know whether a precise element $e \in \Delta^\mathcal{I}$ belongs to $(\widehat{C})^\mathcal{I}$, Algorithm 2 shall be invoked on $T_{\widehat{C}}$ and $\mathcal{I}$. Note that a simple modification in line 12, namely testing whether $v_0 \in \delta(e)$, adapts the algorithm to answer the question for e.

Now, the main difference between Algorithms 1 and 2 is that the latter needs to compute $m^\mathcal{I}$ to verify whether $m^\mathcal{I}(d, E) \sim t$. Therefore, its computational complexity depends on how difficult is to compute $m^\mathcal{I}$ for a chosen m. In particular, if $m^\mathcal{I}$ can be computed in polynomial time as for the graded membership functions obtained from CSMs $\sim$ in *simi-mon*, Algorithm 2 will run in polynomial time.

Missing proofs of Section 3.3

The following theorem was shown in [9].

Theorem 29. *Let C, D be $\mathcal{EL}$ concept descriptions, C^r, D^r their reduced forms, and T_{C^r}, T_{D^r} the corresponding $\mathcal{EL}$ description trees. Then $C \equiv D$ iff there exists an isomorphism between T_{C^r} and T_{D^r}.*

Lemma 30. *Let $\sim$ be an instance of simi such that $g(C) = 1$ for all $\mathcal{EL}$ atoms C of the form $\exists r.C'$. Then, $\sim$ is a standard concept similarity measure.*

Proof. The equivalence closedness property has been already shown in [2] to hold for all instances of *simi*. Regarding equivalence invariance, it follows from the facts that $C \sim D$ is computed using the reduced forms of C and D, and that $C \equiv C'$ and $D \equiv D'$ imply that the structures of C^r and $(C')r$, respectively, D^r and $(D')^r$ are isomorphic (see Theorem 29). It remains to show that by restricting the function g as stated in our claim, the resulting instances of *simi* are also role-depth bounded.

To this end, we show that for every pair of $\mathcal{EL}$ concepts C, D and all $k > \min\{\mathsf{rd}(C), \mathsf{rd}(D)\}$ it holds that:

$$C \sim_d D = C_k \sim_d D_k$$

We proceed by induction on the role depth of C.

Induction base. $\mathsf{rd}(C) = 0$. Then, C is of the form $A \in \mathsf{N_C}$ or $\top$, and $C = C_k$. For $C = A$, the value $C \sim_d D$ is the result of the following expression:

$$\frac{g(A) \times \max\{simi_a(A, D') \mid D' \in top(D)\}}{g(A)}$$

As $C = C_k$, the value $C_k \sim_d D_k$ corresponds to:

$$\frac{g(A) \times \max\{simi_a(A, D') \mid D' \in top(D_k)\}}{g(A)}$$

By definition of $simi_a$, we know that $simi_a(A, D') = pm(A, D')$ if $D' \in \mathsf{N_C}$ and 0 otherwise. Since concept names occurring as top-level atoms in D also occur in D_k, this means that $C \sim_d D = C_k \sim_d D_k$. If $C \equiv \top$, the definition of $\sim_d$ implies that $C \sim_d D = 1$ and $C_k \sim_d D_k = 1$.

Induction step. Let C be an $\mathcal{EL}$ concept such that $\mathsf{rd}(C) = \rho$ with $\rho \geq 1$. Assuming our claim holds for all concepts of role depth smaller than ρ, we prove it also holds for C.

Since $\mathsf{rd}(C) \geq 1$, this means that C is of the form $C = C_1 \sqcap \ldots \sqcap C_n$, where $n \geq 1$ and $\mathsf{rd}(C_i) \leq \rho$ for all $1 \leq i \leq n$. For C and D we have:

$$C \sim_d D = \frac{\sum\limits_{i=1}^{n} \Big[g(C_i) \times \max\{simi_a(C_i, D') \mid D' \in top(D)\} \Big]}{\sum\limits_{i=1}^{n} g(C_i)}, \tag{10}$$

and for C_k and D_k:

$$C_k \sim_d D_k = \frac{\sum\limits_{i=1}^{n} \Big[g([C_i]_k) \times \max\{simi_a([C_i]_k, [D']_k) \mid [D']_k \in top(D_k)\} \Big]}{\sum\limits_{i=1}^{n} g([C_i]_k)} \tag{11}$$

Let us now use the induction hypothesis on all pairs of atoms (C', D') and $([C']_k, [D']_k)$, where $C' \in \mathit{top}(C)$ and $D' \in \mathit{top}(D)$. We distinguish two cases:

- $C' \in \mathsf{N_C}$. Obviously, it is still the case that $k > \min\{\mathsf{rd}(C'), \mathsf{rd}(D')\}$. Therefore, since $\mathsf{rd}(C') = 0 < \rho$, the application of induction yields $C' \sim_d D' = [C']_k \sim_d [D']_k$.
- C' is of the form $\exists r.E_c$. If $D' \in \mathsf{N_C}$, then $C' \sim_d D' = \mathit{simi}_a(C', D') = 0$. Additionally, we have that $[\exists r.E_c]_k = \exists r.[E_c]_{k-1}$ and $[D']_k = D'$. This again means that $[C']_k \sim_d [D']_k = \mathit{simi}_a(\exists r.[E_c]_{k-1}, [D']_k) = 0$.

  The other possibility corresponds to D' being of the form $\exists s.E_d$. Similarly to $\exists r.E_c$, we have $[\exists s.E_d]_k = \exists s.[E_d]_{k-1}$. Then, by definition of $\sim_d$ we have:

  $$C' \sim_d D' = \mathit{pm}(r, s)[w + (1 - w)E_c \sim_d E_d]$$

  $$[C']_k \sim_d [D']_k = \mathit{pm}(r, s)[w + (1 - w)[E_c]_{k-1} \sim_d [E_d]_{k-1}]$$

  Now, having $k > \min\{\mathsf{rd}(C), \mathsf{rd}(D)\}$ implies that $k - 1 > \min\{\mathsf{rd}(E_c), \mathsf{rd}(E_d)\}$. Hence, we can apply induction hypothesis to E_c (notice that $\mathsf{rd}(E_c) < \rho$) to obtain:

  $$E_c \sim_d E_d = [E_c]_{k-1} \sim [E_d]_{k-1}$$

  Thus, it follows that $C' \sim_d D' = [C']_k \sim_d [D']_k$.

Overall, we have just shown that for all pairs of atoms (C', D') where $C' \in \mathit{top}(C)$ and $D' \in \mathit{top}(D)$, it holds $C' \sim_d D' = [C']_k \sim_d [D']_k$.

Let us now take any top-level atom C' of C, and denote by D^* a top-level atom of D that maximizes the value $\mathit{simi}_a(C', D')$ among all $D' \in \mathit{top}(D)$. We make the following observations:

- $g(C') = g([C']_k)$. If $C' \in \mathsf{N_C}$, it is clear since $C' = [C']_k$. Otherwise, C' is an existential restriction as it is $[C']_k$ because $k > 0$.
- Suppose that $\mathit{simi}_a([C']_k, [D^*]_k)$ is not the maximum among all the values $\mathit{simi}_a([C']_k, [D']_k)$ in (11). Then, there exists $[D^x]_k \in \mathit{top}(D_k)$ such that $\mathit{simi}_a([C']_k, [D^*]_k) < \mathit{simi}_a([C']_k, [D^x]_k)$. From this we obtain:

  $$\begin{aligned}
  \mathit{simi}_a(C', D^*) &= C' \sim_d D^* && (C' \text{ and } D^* \text{ are atoms}) \\
  &= [C']_k \sim_d [D^*]_k \\
  &= \mathit{simi}_a([C']_k, [D^*]_k) && ([C']_k \text{ and } [D^*]_k \text{ are atoms}) \\
  &< \mathit{simi}_a([C']_k, [D^x]_k) \\
  &= [C']_k \sim_d [D^x]_k && ([C']_k, [D^x]_k \text{ are atoms}) \\
  &= C' \sim_d D^x \\
  &= \mathit{simi}_a(C', D^x) && (C', D^x \text{ are atoms})
  \end{aligned}$$

Hence, it follows that $simi_a(C', D^*) < simi_a(C', D^x)$ which contradicts the maximality of D^* with respect to $simi_a$ and C'. Therefore, the value $simi_a([C']_k, [D^*]_k)$ is the maximum among all the values $simi_a([C']_k, [D']_k)$ in (11).

Combining these two observations with the expressions in (10) and (11) we obtain that $C \sim_d D = C_k \sim_d D_k$. Once we have this, it follows that:

$$(C^r \sim_d D^r) \otimes (D^r \sim_d C^r) = ([C^r]_k \sim_d [D^r]_k) \otimes ([D^r]_k \sim_d [C^r]_k)$$

for all $k > \min\{\mathsf{rd}(C), \mathsf{rd}(D)\}$ (note that $\mathsf{rd}(C) = \mathsf{rd}(C^r)$). Thus, $C \sim D = C_k \sim D_k$ holds, and $\sim$ is role-depth bounded. □

Lemma 16. Let $\sim \in$ *simi-mon*. For all $\mathcal{EL}$ concept descriptions C and D, there exists an $\mathcal{EL}$ concept description Y such that:

1. $D \sqsubseteq Y$ and $\mathsf{s}(Y) \leq \mathsf{s}(C)$,
2. $C \sim_d D = C \sim_d Y$,
3. $C \sqsubseteq Y$.

Proof. We use induction on the structure of C to prove the claim.

- C is of the form $A \in \mathsf{N_C}$ or $\top$. If $C = A$, the value $C \sim_d D$ is the result of the following expression:

  $$\frac{g(A) \times \max\{simi_a(A, D') \mid D' \in top(D)\}}{g(A)}$$

  Since $pm = pm_d$, this means that $A \sim_d D = 1$ if $A \in top(D)$, otherwise $A \sim_d D = 0$. Choosing $Y := A$ or $Y := \top$, accordingly, ensures that our claims are true. Finally, if $C \equiv \top$, then the general definition of $\sim_d$ implies $C \sim_d X = 1$ for all concept descriptions X. Thus, setting $Y := \top$ satisfies the our claims.

- $C = C_1 \sqcap \ldots \sqcap C_n$ with $n > 1$. In this case we have:

  $$C \sim_d D = \frac{\sum\limits_{j=1}^{n} \Big[g(C_j) \times \max\{simi_a(C_j, D') \mid D' \in top(D)\} \Big]}{\sum\limits_{j=1}^{n} g(C_j)}$$

  Let D_j $(1 \leq j \leq n)$ be a top-level atom of D that maximizes the value $simi_a(C_j, D')$ among all $D' \in top(D)$. The application of the induction hypothesis to C_j and D_j yields a concept description Y_j such that:

- $D_j \sqsubseteq Y_j$ and $\mathsf{s}(Y_j) \leq \mathsf{s}(C_j)$,
- $C_j \sim_d D_j = C_j \sim_d Y_j$,
- $C_j \sqsubseteq Y_j$.

Obviously, $C_1 \sqcap \ldots \sqcap C_n \sqsubseteq Y_1 \sqcap \ldots \sqcap Y_n$ and $D_1 \sqcap \ldots \sqcap D_n \sqsubseteq Y_1 \sqcap \ldots \sqcap Y_n$. Therefore, the concept description $Y := Y_1 \sqcap \ldots \sqcap Y_n$ satisfy $C \sqsubseteq Y$, $D \sqsubseteq Y$ and $\mathsf{s}(Y) \leq \mathsf{s}(C)$.

Now, the value of $C \sim_d Y$ is computed by the following expression:

$$C \sim_d Y = \frac{\sum\limits_{j=1}^{n} \Big[g(C_j) \times \max\{ \mathit{simi}_a(C_j, Y') \mid Y' \in \mathit{top}(Y) \} \Big]}{\sum\limits_{j=1}^{n} g(C_j)} \tag{12}$$

Suppose that for some C_j $(1 \leq j \leq n)$, $\mathit{simi}_a(C_j, Y_j)$ is not the maximum among all the values $\mathit{simi}_a(C_j, Y')$. Then, there is $Y_\ell \in \mathit{top}(Y)$ such that $j \neq \ell$ and $\mathit{simi}_a(C_j, Y_j) < \mathit{simi}_a(C_j, Y_\ell)$. From this we obtain:

$$\begin{aligned}
\mathit{simi}_a(C_j, D_j) &= C_j \sim_d D_j && (C_j \text{ and } D_j \text{ are atoms}) \\
&= C_j \sim_d Y_j \\
&= \mathit{simi}_a(C_j, Y_j) && (C_j \text{ and } Y_j \text{ are atoms}) \\
&< \mathit{simi}_a(C_j, Y_\ell) \\
&\leq \mathit{simi}_a(C_j, D_\ell) && (D_\ell \sqsubseteq Y_\ell,\ (3) \text{ and } C_j, Y_\ell, D_\ell \text{ are atoms})
\end{aligned}$$

Hence, it follows that $\mathit{simi}_a(C_j, D_j) < \mathit{simi}_a(C_j, D_\ell)$ which contradicts the maximality of D_j with respect to simi_a and C_j. Hence, $\mathit{simi}_a(C_j, Y_j)$ is actually the maximum, and once this is true it is easy to see that $C \sim_d D = C \sim_d Y$.

- C is of the form $\exists r.C'$. Let D^* be the top-level atom of D maximizing the value $\mathit{simi}_a(C, D^*)$. If $D^* \in \mathsf{N_C}$, then $\mathit{simi}_a(C, D^*) = 0$ and $C \sim_d D = 0$. Then, choosing $Y = \top$ satisfies our claims.

  If D^* is of the form $\exists s.D'$, then $\sim$

  $$C \sim_d D = \mathit{pm}(r, s)[w + (1 - w) \times (C' \sim_d D')]$$

  For $r \neq s$ we have $\mathit{pm}(r, s) = 0$ and $C \sim_d D = 0$. Then, again choosing Y as $\top$ is enough. Otherwise, the application of induction hypothesis to C' w.r.t. D' yields a concept description Y' such that:

  - $D' \sqsubseteq Y'$ and $\mathsf{s}(Y') \leq \mathsf{s}(C')$,
  - $C' \sim_d D' = C' \sim_d Y'$,
  - $C' \sqsubseteq Y'$.

Then, for the concept descriptions $Y := \exists r.Y'$ we have that $C \sqsubseteq Y$ (recall that we already ruled out the case $r \neq s$), $D \sqsubseteq \exists r.D' \sqsubseteq Y$ and $\mathsf{s}(Y) \leq \mathsf{s}(C)$. Additionally,

$$C \sim_d Y = pm(r, r)[w + (1 - w) \times (C' \sim_d Y')]$$

□

Lemma 18 Let $\sim \in$ *simi-mon*. Additionally, let $\mathcal{I}$ be an interpretation, $d \in \Delta^{\mathcal{I}}$ and $C \in \mathcal{C}_{\mathcal{EL}}(\mathsf{N_C}, \mathsf{N_R})$ with $\mathsf{rd}(C) = k$. Then, there exists a concept $D \in \mathcal{R}^{k+1}$ such that:

1. $d \in D^{\mathcal{I}}$ and $C^r \sim_d D = \max\{C^r \sim_d D' \mid D' \in \mathcal{R}^{k+1} \text{ and } d \in (D')^{\mathcal{I}}\}$,
2. $m^{\mathcal{I}}_{\sim}(d, C) = (C^r \sim_d D) \otimes 1$.

Proof. Since $\mathcal{R}^{k+1}$ is a finite set, this means that there is at least one concept D such that $d \in D^{\mathcal{I}}$ and

$$C^r \sim_d D = \max\{C^r \sim_d D' \mid D' \in \mathcal{R}^{k+1} \text{ and } d \in (D')^{\mathcal{I}}\} \tag{13}$$

Suppose, however, that $m^{\mathcal{I}}(d, C) \neq (C^r \sim_d D) \otimes 1$. The application of Lemma 16 to C^r and D yields a concept description Y such that:

- $D \sqsubseteq Y$ and $C^r \sqsubseteq Y$,
- $C^r \sim_d D = C^r \sim_d Y$.

Hence, $D \sqsubseteq Y$ implies $d \in Y^{\mathcal{I}}$, and $C^r \sqsubseteq Y$ implies $Y \sim_d C^r = 1$. Moreover, since $Y \sqsubseteq Y^r$ and $Y^r \sqsubseteq Y$, by Property (3), it follows that $C^r \sim_d Y = C^r \sim_d Y^r$. Consequently, by definition of $\sim$ we have:

$$C \sim Y = (C^r \sim_d Y^r) \otimes 1$$

Thus, since $d \in Y^{\mathcal{I}}$ and $C^r \sim_d D = C^r \sim_d Y^r$, by definition of $m_{\sim}$ is must be the case that

$$m^{\mathcal{I}}_{\sim}(d, C) > C \sim Y \tag{14}$$

Now, let $D^* \in \mathcal{R}^{k+1}$ be a concept description such that $m^{\mathcal{I}}_{\sim}(d, C) = C \sim D^*$. By definition of $\sim$ we have that:

$$C \sim D^* = (C^r \sim_d (D^*)^r) \otimes ((D^*)^r \sim_d C^r)$$

Again, by Lemma 16 and the monotonicity of $\otimes$, one can assume that:

$$C \sim D^* = (C^r \sim_d (D^*)^r) \otimes 1$$

Moreover, by the maximality of D in (13) and the properties of Y, we know that:

$$(C^r \sim_d (D^*)^r) \leq C^r \sim_d Y^r$$

Hence, since $\otimes$ is commutative and monotonic it follows that $C \sim D^* \leq C \sim Y$. This is clearly in contradiction with the statement in (14). Thus, $m^{\mathcal{I}}_{\sim}(d, C) = C \sim Y$, and the lemma holds. □

Computation of $m^{\mathcal{I}}_{\sim}$. We now provide an algorithm that computes $m^{\mathcal{I}}_{\sim}$ for finite interpretations $\mathcal{I}$. First, we define the following notions which will be useful in the corresponding proofs.

Definition 31. Let C be an $\mathcal{EL}$ concept description and T_C its associated $\mathcal{EL}$ description tree. For all nodes $v \in V_{T_C}$ we denote by $T_C[v]$ the subtree of T_C rooted at v. Furthermore, the $\mathcal{EL}$ concept description $C[v]$ is the one having the description tree $T_C[v]$. Finally, the height $\eta(v)$ of a node v in T_C is the length of the longest path from v to a leaf of T_C.

We would like to point out that for all concept descriptions C^r in reduced form, the concepts $C^r[v]$ are also in reduced form (for all $v \in V_{T_{C^r}}$). This is a consequence of the fact that to obtain the reduced form of a concept C the rules are not only applied in the top-level conjunction of C, but also under the scope of existential restrictions (see Section 2.1).

Algorithm 3 considers each pair (v, e) with $v \in V_{T_{C^r}}$ and $e \in \Delta^{\mathcal{I}}$ only once. Therefore, since g and $\otimes$ are computable in polynomial time, it is easy to see that Algorithm 3 runs in time polynomial in the size of C and $\mathcal{I}$. The following lemma shows that it actually computes the value of $m^{\mathcal{I}}_{\sim}$, i.e., $S(v_0, d) \otimes 1 = m^{\mathcal{I}}_{\sim}(d, C)$.

Lemma 32. *Let $\sim \in$ simi-mon, C be an $\mathcal{EL}$ concept description, $\mathcal{I}$ a finite interpretation and $d \in \Delta^{\mathcal{I}}$. Then, Algorithm 3 outputs $m^{\mathcal{I}}_{\sim}(d, C)$, i.e., $S(v_0, d) \otimes 1 = m^{\mathcal{I}}_{\sim}(d, C)$.*

Proof. To show that $S(v_0, d) \otimes 1 = m^{\mathcal{I}}_{\sim}(d, C)$, we first prove the following claim:

$$S(v, e) = \max\{C^r[v] \sim_d D \mid e \in D^{\mathcal{I}}\}, \text{ for all } (v, e) \in V_{T_{C^r}} \times \Delta^{\mathcal{I}} \qquad (15)$$

Notice that for each pair (v, e) the value of $S(v, e)$ is assigned only once during a run of the algorithm. We prove the claim by induction on the height $\eta(v)$ of each node v in T_{C^r}.

Induction Base. $\eta(v) = 0$. Then, v is a leaf of T_{C^r}. This means that $C^r[v]$ is either $\top$ or a conjunction of different concept names. If it is $\top$, the algorithm

Algorithm 3 Computation of $m^{\mathcal{I}}_{\sim}$.

```
Input: A CSM $\sim \in$ simi-mon, an $\mathcal{EL}$ concept description $C$, a finite interpretation $\mathcal{I}$ and $d \in \Delta^{\mathcal{I}}$.
Output: $m^{\mathcal{I}}_{\sim}(d, C)$.
Let $C^r$ be the reduced form of $C$ and $G_{\mathcal{I}} = (V_{\mathcal{I}}, E_{\mathcal{I}}, \ell_{\mathcal{I}})$.
Let $\{w_1, \ldots, w_n\}$ be a post-order sequence of $V_{T_{C^r}}$ where $w_n = v_0$.
The assignment $S : V_{T_{C^r}} \times V_{\mathcal{I}} \to [0..1]$ is computed as follows:
for all $1 \leq i \leq n$ do
    $v = w_i$
    if $C^r[v] = \top$ then
        $S(v, e) := 1$ for all $e \in \Delta^{\mathcal{I}}$
    else
        Let $C^*$ be the following concept description:
            $$C^* := \bigsqcap_{A \in \ell_{T_{C^r}}(v)} A$$
        for all $e \in \Delta^{\mathcal{I}}$ do
            Let $D^*$ be the following concept description:
                $$D^* := \bigsqcap_{B \in \ell_{\mathcal{I}}(e)} B$$
            $c := \sum_{A \in top(C^*)} \left[ g(A) \times \max_{B \in top(D^*)} simi_a(A, B) \right]$
            for all $vs_iv_i \in E_{T_{C^r}}$ do
                $c := c + \max_{(e,e') \in s_i^{\mathcal{I}}} w + (1 - w) \times S(v_i, e')$
            end for
            $S(v, e) := \frac{c}{\sum_{A \in top(C^*)} g(A) + \sum_{vs_iv_i \in E_{T_{C^r}}} 1}$
        end for
    end if
end for
return $S(v_0, d) \otimes 1$
```

treats it properly by setting $S(v,e)=1$ in line 7. Otherwise, $C^r[v]$ is of the form $A_1 \sqcap \ldots \sqcap A_n$. Let D be any concept description such that $e \in D^{\mathcal{I}}$. Then,

$$C^r[v] \sim_d D = \frac{\sum\limits_{i=1}^{n} \left[g(A_i) \times \max\limits_{D' \in top(D)} (simi_a(A_i, D')) \right]}{\sum\limits_{i=1}^{n} g(A_i)}$$

By definition of $simi_a$, $simi_a(A_i, D') > 0$ iff $D' = A$. Therefore, existential restrictions in the top level of D are irrelevant to obtain the value $C^r[v] \sim_d D$. Therefore, one can restrict the attention to concepts D that are a conjunction of concept names B such that $e \in B^{\mathcal{I}}$. Consequently, by definition of D^* in Algorithm 3, it follows that $D^* \sqsubseteq D$. Since obviously $e \in (D^*)^{\mathcal{I}}$, by Property 3 we then have:

$$C^r[v] \sim_d D^* = \max\{C^r[v] \sim_d D \mid e \in D^{\mathcal{I}}\}$$

One can easily see that in this case, $S(v,e)$ gets assigned the value $C^r[v] \sim_d D^*$.

Induction Step. $\eta(v) > 0$. Let $v_1, \ldots, v_k$ be the children of v in T_{C^r} such that there exists at least one s_i successor of e in $\mathcal{I}$. The application of the max operator in line 14, selects for each s_i-successor v_i of v an s_i-successor e_i of e in $\Delta^{\mathcal{I}}$ that has the maximum value for $S(v_i, e_i)$. Such a value is then used in the computation of c. Let (v_i, e_i) be the pairs representing such a selection for all v_i. Two observations are in order:

- Since v_i is a child of v, it occurs first in the post-oder selected in line 2. Therefore, the value of $S(v_i, e_i)$ is computed before the computation of c for (v, e).
- The value of $S(v,e)$ as computed by Algorithm 3 corresponds to the following expression:

$$S(v,e) = \frac{\sum\limits_{A \in top(C^*)} \left[g(A) \times \max\limits_{B \in top(D^*)} simi_a(A,B) \right] + \sum\limits_{i=1}^{k} [w + (1-w) \times S(v_i, e_i)]}{\sum\limits_{A \in top(C^*)} g(A) + \sum\limits_{vs_iv_i \in E_{T_{C^r}}} 1} \tag{16}$$

- Since $\eta(v_i) < \eta(v)$, the application of the induction hypothesis yields

$$S(v_i, e_i) = \max\{C^r[v_i] \sim_d D \mid e_i \in D^{\mathcal{I}}\} \tag{17}$$

Let now D_i be a concept description such that $S(v_i, e_i) = C^r[v_i] \sim_d D_i$ and $e_i \in (D_i)^{\mathcal{I}}$, for all $i \in \{1 \ldots k\}$. We define the $\mathcal{EL}$ concept description D as:

$$D := D^* \sqcap \prod_{i=1}^{k} \exists s_i.D_i$$

Since $e \in (D^*)^{\mathcal{I}}$ and $(e, e_i) \in (s_i)^{\mathcal{I}}$ for all $1 \leq i \leq k$, this means that $e \in D^{\mathcal{I}}$. Again, it is not hard to see that Algorithm 3 assigns the value $C^r[v] \sim_d D$ to $S(v, e)$. Suppose, however, that there exists a concept description E such that $e \in E^{\mathcal{I}}$ and $C^r[v] \sim_d E > C^r[v] \sim_d D$. Then, by definition of $\sim_d$ there must exist top level atoms C^x and E^x of $C^r[v]$ and E^x, respectively, such that:

$$simi_a(C^x, E^x) > \max_{D' \in top(D)} simi_a(C^x, D') \tag{18}$$

We distinguish three cases regarding the form of C^x.

- C^x if of the form $A \in \mathsf{N_C}$. In such a case, the only possibility is $E^x = A$. But this means that $e \in A^{\mathcal{I}}$, and consequently A is also a top level atom of D, contradicting (18).
- C^x is of the form $\exists s_i.C^r[v_i]$ for $1 \leq i \leq k$. Then, E^x is of the form $\exists s_i.E'$. Since $e \in (E_x)^{\mathcal{I}}$, this means that there exists $e' \in \Delta^{\mathcal{I}}$ such that $(e, e') \in (s_i)^{\mathcal{I}}$ and $e' \in (E')^{\mathcal{I}}$. Clearly, $e' = e_i$ must hold, for otherwise e' would have been selected instead of e_i (in view of (18)).

  By definition of $simi_a$ we have:

  $$simi_a(\exists s_i.C^r[v_i], \exists s_i.E') = w + (1 - w) \times C^r[v_i] \sim_d E'$$

  and,

  $$simi_a(\exists s_i.C^r[v_i], \exists s_i.D_i) = w + (1 - w) \times C^r[v_i] \sim_d D_i$$

  Since $S(v_i, e_i) = C^r[v_i] \sim_d D_i$, the consequence (17) of applying the induction hypothesis tells us that $C^r[v_i] \sim_d E' \leq C^r[v_i] \sim_d D_i$. Thus, since $\exists s_i.D_i$ is a top level atom of D, we again obtain a contradiction with (18).
- C^x is any other existential restriction $\exists s.C'$ occurring in the top level of $C^r[v]$. This means that e has no s-successor in $\mathcal{I}$, and since $pm = \pm_d$, it follows that $simi_a(C^x, E^x) = 0$ always holds.

Overall, we have just shown that $S(v, e)$ satisfies our claim in (15). Furthermore, by Lemma 18 we know that $S(v_0, d) \otimes 1 = m^{\mathcal{I}}_{\sim}(d, C)$. Thus, Algorithm 3 is correct. □

Lemma 24. Let c_k be a clause in φ and x_i, x_j, x_ℓ the variables occurring in it. Additionally, let D be an $\mathcal{EL}$ concept description such that:

$$D \not\sqsubseteq X^{\{i\}}, \quad D \not\sqsubseteq X^{\{j\}} \quad \text{and} \quad D \not\sqsubseteq X^{\{\ell\}}$$

Then, for all $\sim \in$ *simi-mon*, $C_k \sim_d D^r < C_k \sim_d X^{\{a\}}_n$, $a \in \{i, j, \ell\}$.

Proof. First, notice that C_k and $X_n^{\{a\}}$ are already in reduced form. Since $D^r \not\sqsubseteq X^{\{a\}}$ for all $a \in \{i, j, \ell\}$, the inductive construction of the concept Y w.r.t. C_k and D^r in the proof of Lemma 16 tells us that such a Y can be of the form:

$$\exists \underbrace{r \ldots r}_{\rho}.\top, \text{ where } \rho = \min(\mathsf{rd}(C_k), \mathsf{rd}(D^r))$$

We also know that $C_k \sim_d D^r = C_k \sim_d Y$. In addition, it is clear that:

$$X_n^{\{a\}} \sqsubseteq \exists \underbrace{r \ldots r}_{\mathsf{rd}(C)}.\top \sqsubseteq \exists \underbrace{r \ldots r}_{\rho}.\top$$

Therefore, by using Property 3 of $\sim_d$ we obtain that:

$$C_k \sim_d X_n^{\{a\}} \geq C_k \sim_d \exists \underbrace{r \ldots r}_{\mathsf{rd}(C)}.\top \geq C_k \sim_d \exists \underbrace{r \ldots r}_{\rho}.\top$$

Furthermore, since both C_k and $X_n^{\{a\}}$ have an occurrence of the atom A at the same role depth, it is not difficult to see that by definition of $\sim_d$, $>$ also holds for the previous inequalities:

$$C_k \sim_d X_n^{\{a\}} > C_k \sim_d \exists \underbrace{r \ldots r}_{\mathsf{rd}(C)}.\top > C_k \sim_d \exists \underbrace{r \ldots r}_{\rho}.\top$$

Thus, it follows that $C_k \sim_d X_n^{\{a\}} > C_k \sim_d D^r$. □